CONFIDENCE INTERVALS WITH THE EFFICIENT BOOTSTRAP SAMPLE DESIGN (EBSD): PERFORMANCE AND COMPARISON

A DISSERTATION

SUBMITTED ON THE FIFTH DAY OF APRIL 2023

TO THE DEPARTMENT OF BIOSTATISTICS

IN PARTIAL FUFILLMENT OF THE REQUIREMENTS OF THE SCHOOL OF PUBLIC HEALTH AND TROPICAL MEDICINE

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY BY

NICHOLAS DAGINCOURT, MS

APPROVED: ________________ 4/10/2023

Sudesh Srivastav [Committee Chair], PhD; date

________________ 4/10/2023

John Lefante, PhD; date

________________ 4/5/2023

Tumulesh Solanky, PhD; date

Abstract

The Monte Carlo Bootstrap resampling method is among the most useful tools for accurate confidence interval computation. An inherent flaw of the method though, is its use of Monte Carlo resampling. Monte Carlo resampling relies on random resampling from the original sample in order to generate a confidence interval. Using random resampling, however, causes a method to yield different results nearly every time the method is performed on the same data. Further Monte Carlo resampling introduces simulation error. Simulation error occurs because for each draw each sample point has a $\frac{1}{n}$ probability of being chosen, and inevitably, some sample points randomly contribute to the sampling distribution more frequently than others.

The Efficient Bootstrap Sample Design method for a sample of size n (EBSD(n)) has been created to address these inefficiencies inherent to the Monte Carlo Bootstrap[1]. EBSD(n) eliminates simulation error using principles of BIBD. The construction of this design allows for a fixed, systematic approach of constructing replicable confidence interval results.

The motivation of this work was to compare the accuracy of confidence intervals applied on EBSD(n) to the accuracy of confidence intervals applied on the Monte Carlo Bootstrap. In order to do this, confidence interval methods type-1 error rate was computed for methods commonly applied on the Monte Carlo Bootstrap and for methods applied on EBSD(n). Two types of methods applied on EBSD(n) were tested for accuracy. (i) A new confidence interval method called E-skew and (ii) Confidence interval methods that are applied commonly on the Monte Carlo Bootstrap but instead were applied on EBSD(n).

Both (i) and (ii) performed relatively accurately for specific types of probability distributions and statistics studied. Further, the new method E-skew was measured to be in statistically significant agreement with the BC_a and Bootstrap-t algorithms using the Kappa statistic. This suggested E-skew could provide similar accuracy to these methods in a real data context while also benefitting from the advantages of EBSD(n).

Table of Contents

Chapter 1: Background and Significance 1
1.1 Background 1
1.2 Introduction to the Problem 2
Chapter 2: Literature Review 5
2.1 Bootstrap Theory and the non-Parametric Bootstrap Sampling Distribution 6
2.2 The Parametric Bootstrap 7
2.3 Bias 8
2.4 Variance Estimation 17
2.5 Confidence Bounds 21
2.6 Research Questions 30
Chapter 3: Methods 32
3.1 Bootstrap Sample Construction using EBSD(n) 32
3.2 Confidence Intervals using EBSD(n) 40
3.3 Workflow 52
3.4 Simulation Study Procedure 55
3.5 Criteria for Confidence Interval Method Comparison 58
Chapter 4: Main Results 60
4.1 Sample Mean 60
4.2 Ratio of Independent Sample Means 132
4.3 Pearson Correlation Coefficient 184
4.4 Trimmed Sample Mean 214
4.5 Sample Median 255
4.6 Varied Monte Carlo Bootstrap Iterations Levels 294
4.7 Application of EBSD(n) in Microarray Data 304
Chapter 5: Conclusion 335
REFERENCES 346

ACKNOWLEDGEMENTS

I am very thankful to the many who were integral to the completion of this work. First and foremost, thank you to Dr. Sudesh Srivastav for his mentorship, guidance and brilliance throughout this process. I first took a course that Dr. Srivastav taught in 2012, Design of Experiments, I found his approach to the course very engaging and satisfying. Simultaneously in 2012 he guided me in completing my Master's Thesis. Then starting in 2018 he became my advisor for my PhD dissertation and constantly taught, inspired, and challenged me to perform the best research I possibly could. His passion for the material I was working on was evident throughout my research. He was patient with me throughout and allowed me to develop a familiarity and eventually a mastery of the material so that I could perform my best possible work.

Thank you to Dr. John Lefante for helping to guide me in the microarray data analysis portion of this dissertation. Additionally, although I never had an in-person class with Dr. Lefante, he agreed to provide an independent study for me in Mediation and Moderation analysis that allowed me to complete my course work. I appreciate that he was an advocate for me throughout this process.

Thank you to Dr. Tumulesh Solanky, for agreeing to be a member of my committee, participating and taking his time out to go through the voluminous amount of work here. The number of results here is a bit overwhelming and his ability to engage with it is much appreciated.

Additionally, thank you to my wife Clare for her patience and emotional support throughout this process.

Last but not least, thank you to my parents for their support and always being amazing cheerleaders. Their belief in me never seems to wane and their support was crucial in this process.

Notation

This section references notation that will appear throughout the course of the dissertation below; the notation will also be defined when it first appears in the section where it is introduced.

Generic Statistical Inference Notation

- n references the size the sample for data points collected from the population of interest
- $\hat{\theta}$ is the estimate of the true population parameter θ from the sample. For example, $\hat{\theta}$ could be $\bar{X}$ computed from the sample where $\theta = \mu$.
- $s^2 = (n - 1)^{-1} \sum (X_i - \bar{X})^2$ is the usual sample variance, and $\hat{\sigma}^2 = (n)^{-1} \sum (X_i - \bar{X})^2 = (n - 1)\, s^2/n$ is the population variance.
- $t_{1-\frac{\alpha}{2}, n-1}$ denotes the t-statistic for the $1 - \frac{\alpha}{2}$ level confidence interval on $n - 1$ degrees of freedom.
- $O\left(\frac{1}{\sqrt{n}}\right)$ signifies first order convergence.
- $O\left(\frac{1}{n}\right)$ signifies second order convergence.

The Monte Carlo Bootstrap

- B is the number of resamples in a bootstrap or permutation distribution.
- $\hat{\theta}_i^*$ is the estimate of the population parameter from the i^{th} bootstrap resample.
- $\hat{\boldsymbol{\theta}}_B^*$ is the bootstrap sampling distribution of $\hat{\theta}$

- $\bar{\theta}_B^*$ is the mean of the B bootstrap estimates from the bootstrap sampling distribution, i.e. $\bar{\theta}_B^* = \frac{1}{B}\sum \hat{\theta}_i^*$.
- The mean of the bootstrap distribution is $\bar{\theta}_B^*$ or $\bar{X}^*$
- the standard deviation of the bootstrap distribution (the bootstrap standard error) is $s_B = \sqrt{(B - 1)^{-1} \sum(\hat{\theta}^*{}_i - \overline{\overline{\theta^*}})^2}$

Monte Carlo Bootstrap Probability Vector Definitions

- $\sum_{i=1}^{B} \bar{\boldsymbol{P}}^{*i} / B$ represents the average probability for each sample point across the B bootstrap resamples.
- $T^{QUAD}(P^*) = c_0 + (P^* - P^0)^T U + (\frac{1}{2})(P^* - P^0)^T V (P^* - P^0)$ is the quadratic statistic based on the weighted P^* vector, where c_0 is the original $\hat{\theta}$ estimate from the sample, U is an n-vector satisfying $\sum_1^n U_i = 0$ and V is an $n \times n$ symmetric matrix satisfying $\sum_i V_{ij} = \sum_j V_{ij} = 0$ for all i, j. Further P^* is defined as vector of probabilities $(P_1^*, \ldots P_n^*)^T$ satisfying $0 \le P_i^* \le 1$ and $\sum_1^n P_i^* = 1$ and where P^0 is defined as the probability vector of each element being equally likely to be chosen from the original sample: $P^0 = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})^T$.

The Monte Carlo Bootstrap Bias Estimates

- $b(\theta)$ is the bias of the estimate $\hat{\theta}$ from the true population parameter θ
- While $b(\hat{\theta})$, is an estimate of the true bias $b(\theta)$ using information from the sample
- $\hat{\theta}_{br} = \hat{\theta} - b(\hat{\theta})$ is the bias reduced estimate of $\hat{\theta}$ where $\hat{\theta}$ is reduced by an estimate of the bias $b(\hat{\theta})$.

- $T(\overline{\boldsymbol{P}}^{*})$ is the statistic computed on the average weighted proportion of elements $\overline{\boldsymbol{P}}^{*}$ where $\overline{\boldsymbol{P}}^{*}$ = $\overline{bias_B}$ = $\bar{\theta}_B^{*} - T(\overline{\boldsymbol{P}}^{*})$, is defined as the bootstrap estimate of bias for the estimator $\hat{\theta}$

BC_a Algorithm

- $\hat{z}_0 = \Phi^{-1}((\sum_{i=1}^{B} \theta_i < \hat{\theta})/B)$ is the bias adjustment component for the Monte Carlo bootstrap BC_a interval
- $\hat{a} = \frac{\sum_{i=1}^{n}(\hat{\boldsymbol{\theta}}_{(.)} - \hat{\boldsymbol{\theta}}_i)^3}{6(\sum_{i=1}^{n}((\hat{\boldsymbol{\theta}}_{(.)} - \hat{\boldsymbol{\theta}}_i)^2)^{3/2}}$ is the estimate of acceleration for the bootstrap BC_a interval

Bootstrap-t Algorithm

- $t^{*}{}_B = \frac{\hat{\theta}_B^{*} - \hat{\theta}}{\hat{s}_{\widehat{\theta^*}_B}}$ is the t-statistic computed for the b^th bootstrap resample. $\hat{s}_{\widehat{\theta^*}_B}$ is the sample standard deviation for the bth bootstrap resample for a statistic $\hat{\theta}_B^{*}$.
- $t^{*}{}_{(1-\frac{\alpha}{2})}$ is the element of the $1 - \frac{\alpha}{2}$ percentile from the ordered bootstrap sampling distribution

The Jackknife

- $\hat{\theta}_{J(.)} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\theta_{J_i}^{*}}$, is the jackknife estimate of the population parameter θ, where it is the average of the $\widehat{\theta_{J_i}^{*}}$ jackknife samples.
- $\hat{\theta}_{j_i}^{*}$ is the i^th jackknife estimate where the i^th element is removed from the sample.
- $\hat{\sigma}_{\hat{\theta}_j}{}^2 = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(-i)} - \hat{\theta}_{(.)})^2$ is defined as the jackknife estimate of variance for the estimator $\hat{\theta}$

- $\widehat{bias}_{jack} = (n-1) * (\hat{\theta}_{J(.)} - \hat{\theta})$ is defined as the jackknife estimate of bias for the estimator $\hat{\theta}$

EBSD($\boldsymbol{n}$)

- Correspondingly, from the EBSD method we have $\hat{\theta}_i^*$ as before where $\bar{\theta}_E^* = \frac{1}{4n^2+1}\sum \hat{\theta}_i^*$.

List of Tables

Table 1. Random Permutation Design 14
Table 2. Randomized Block Design 15
Table 3. Survival Time (Days) Treatment group Mouse Data 20
Table 3.1 Survival Time (Days) Treatment group Mouse Data 37
Table 3.2 Efron and Tibshirani Estimated Standard Error for the Sample Mean: Mouse Data: Treatment Group 37
Table 3.3 Comparison of Relative Design Sizes 39
Table 3.4 Confidence Interval Method's Used in the Simulation Study 53
Table 3.5 Parameter Specifications for each Statistic for each distribution of interest 53
Table 4.1 Simulation Parameter Specifications for Sample Mean 62
Table: Sample Mean - N1U99 Upper limit error rate (α = 0.01), Normal Distribution, $N(\mu = 4$, σ =1), Bootstraps=10000 68
Table: Sample Mean - N1L99 Lower limit error rate (α = 0.01), Normal Distribution, $N(\mu = 4$, σ =1), Bootstraps=10000 69
Table: Sample Mean - E1U99 Upper limit error rate (α = 0.01), Exponential Distribution Exp(λ = 0.10), Bootstraps=10000 83
Table: Sample Mean - E1L99 Lower limit error rate (α = 0.01), Exponential Distribution, Exp(λ = 0.10), Bootstraps=10000 84
Table: Sample Mean - G1U99 Upper limit error rate (α = 0.01), Gamma Distribution, gamma($\alpha = 2,\ \lambda = 2$), Bootstraps = 10000 96
Table: Sample Mean - G1L99 Lower limit error rate (α = 0.01), Gamma Distribution, gamma($\alpha = 2,\ \lambda = 2$), Bootstraps = 10000 97
Table: Sample Mean - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distribution, log-normal($\mu = 4$, σ = 0.2), Bootstraps=10000 109
Table: Sample Mean - LN1L99 Lower limit error rate (α = 0.01), Log-Normal Distribution, log-normal(μ = 4, σ = 0.2), Bootstraps=10000 110
Table: Sample Mean - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.6*$N(\mu1 = 4, \sigma1 = 4)$ + 0.4*$N(\mu2 = 8,\ \sigma2 = 8)$, Bootstraps=10000 121
Table: Sample Mean - MN1L99 Lower limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.6*$N(\mu1 = 4, \sigma1 = 4)$ + 0.4*$N(\mu2 = 8,\ \sigma2 = 8)$, Bootstraps=10000 122
Table 4.2 Simulation Parameter Specifications for the Ratio of Sample Means 134
Table: Ratio of Sample Means - N1U99 Upper limit error rate (α = 0.01), Normal Distributions, $N1(\mu1 = 100, \sigma1 = 1), N2(\ \mu2 = 50,\ \sigma2 = 1)$, Bootstraps = 10000 138
Table: Ratio of Sample Means - N1L99 Lower limit error rate (α = 0.01), Normal Distribution, $N1\mu1 = 100, \sigma1 = 1, N2(\ \mu2 = 50,\ \sigma2 = 1)$, Bootstraps = 10000 139
Table: Ratio of Sample Means - E1U99 Upper limit error rate (α = 0.01), Exponential Distributions, $Exp1$($\lambda1$ = 0.1), $Exp2$($\lambda2$ =0.2), Bootstraps=10000 151
Table: Ratio of Sample Means - E1L99 Lower limit error rate (α = 0.01), Exponential Distributions, $Exp1$($\lambda1$ = 0.1), $Exp2$($\lambda2$ =0.2), Bootstraps=10000 152
Table: Ratio of Sample Means - G1U99 Upper limit error rate (α = 0.01), Gamma Distributions, $gamma1\alpha1 = 4, \lambda1 = 1, gamma2(\alpha2 = 3, \lambda2 = 1)$, Bootstraps=10000 164
Table: Ratio of Sample Means - G1L99 Lower limit error rate (α = 0.01) Gamma Distributions, $gamma1\alpha1 = 4, \lambda1 = 1, gamma2(\alpha2 = 3, \lambda2 = 1)$, Bootstraps=10000 165

Table: Ratio of Sample Means - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distributions, $log-normal1\mu1 = 4, \sigma1 = 0.2, log-normal2(\mu2 \approx 3.3, \sigma2 = 0.2)$, Bootstraps=10000 ... 171
Table: Ratio of Sample Means - LN1L99 Lower Limit error rate (α = 0.01), Log-Normal Distributions, $log-normal1\mu1 = 4, \sigma1 = 0.2, log-normal2(\mu2 \approx 3.3, \sigma2 = 0.2)$, Bootstraps=10000 ... 172
Table: Ratio of Sample Means - MN1U99 Upper limit error rate (α = 0.01), Mixture Distribution: $MN1(0.6*N(\mu1 = 50, \sigma1 = 1) + 0.4*N(\mu2 = 100, \sigma2 = 1))$, $MN2(0.6*N(\mu1 = 25, \sigma1 = 1) + 0.4*N(\mu2 = 50, \sigma2 = 1))$, Bootstraps=10000 ... 178
Table: Ratio of Sample Means - MN1L99 Lower limit error rate (α = 0.01), Mixture Distribution: $MN1(0.6*N(\mu1 = 50, \sigma1 = 1) + 0.4*N(\mu2 = 100, \sigma2 = 1))$, $MN2(0.6*N(\mu1 = 25, \sigma1 = 1) + 0.4*N(\mu2 = 50, \sigma2 = 1))$, Bootstraps=10000 ... 179
Table 4.3 Simulation Parameter Specifications for the Pearson Correlation Coefficient ... 186
Table: Pearson Correlation Coefficient - N1U99 Upper limit error rate (α = 0.01), Bivariate Normal Distributions, $MVN(\mu1 = 4, \sigma1 = 1, \mu2 = 4, \sigma2 = 1, \rho = 0.1)$, Bootstraps=10000 ... 190
Table: Pearson Correlation Coefficient - N1L99 Lower limit error rate (α = 0.01), Bivariate Normal Distributions, $MVN(\mu1 = 4, \sigma1 = 1, \mu2 = 4, \sigma2 = 1, \rho = 0.1)$, Bootstraps=10000 ... 191
Table: Pearson Correlation Coefficient - NNU99 Upper limit error rate (α = 0.01), Bivariate non-Normal Distribution, Bivariate Non-Normal(Skew=3, Kurtosis=61, ρ = 0.1), Bootstraps=10000 ... 203
Table: Pearson Correlation Coefficient - NNL99 Lower limit error rate (α = 0.01), Bivariate non-Normal Distribution, Bivariate Non-Normal(Skew=3, Kurtosis=61, ρ = 0.1), Bootstraps=10000 ... 204
Table 4.4 Simulation Parameter Specifications for the Trimmed Sample Mean Statistic ... 215
Table: Trimmed Sample Mean - N1U99 Upper limit error rate (α = 0.01), Normal Distribution, $N(\mu = 50, \sigma = 1)$, Bootstraps=10000 ... 218
Table: Trimmed Sample Mean - N1L99 Lower limit error rate (α = 0.01), Normal Distribution, $N(\mu = 50, \sigma = 1)$, Bootstraps=10000 ... 219
Table: Trimmed Sample Mean - E1U99 Upper limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda = 0.10)$, Bootstraps=10000 ... 226
Table: Trimmed Sample Mean - E1L99 Lower limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda = 0.10)$, Bootstraps=10000 ... 227
Table: Trimmed Sample Mean - G1U99 Upper limit error rate (α = 0.01), Gamma Distribution, gamma(α =2, λ=2), Bootstraps=10000 ... 234
Table: Trimmed Sample Mean - G1L99 Lower limit error rate (α = 0.01), Gamma Distribution, gamma(α =2, λ=2), Bootstraps=10000 ... 235
Table: Trimmed Sample Mean - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distribution, log-normal($\mu = 4$, σ =0.2), Bootstraps=10000 ... 242
Table: Trimmed Sample Mean - LN1L99 Lower limit error rate (α = 0.01), Log-Normal Distribution, log-normal($\mu = 4$, σ =0.2), Bootstraps=10000 ... 243
Table: Trimmed Sample Mean - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions: $0.6*N1(\mu1 = 4, \sigma1 = 4) + 0.4*N2(\mu2 = 8, \sigma2 = 8)$, Bootstraps = 10000 ... 250
Table: Trimmed Sample Mean - MN1L99 Lower limit error rate (α = 0.01), Mixture of two Normal Distributions: $0.6*N1(\mu1 = 4, \sigma1 = 4) + 0.4*N2(\mu2 = 8, \sigma2 = 8)$, Bootstraps = 10000 ... 251

Table 4.5 Simulation Parameter Specifications for the Median Statistic....................................258
Table: Sample Median - N1U99 Upper limit error rate (α = 0.01), Normal distributions, $N(\mu$=4, σ=1), Bootstraps=5000 260
Table: Sample Median - N1L99 Lower limit error rate (α = 0.01), Normal distribution, $N(\mu$=4, σ=1), Bootstraps=5000 261
Table: Sample Median - E1U99 Upper limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda$=0.10), Bootstraps=5000 267
Table: Sample Median - E1L99 Lower limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda$=0.10), Bootstraps=5000 268
Table: Sample Median - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distribution, log-normal(μ=4, σ=0.2), Bootstraps=5000..274
Table: Sample Median - LN1L99 Lower limit error rate (α = 0.01), Log-Normal Distribution, log-normal(μ=4, σ=0.2), Bootstraps=5000..275
Table: Sample Median - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.50*$N1(\mu1$ =4, $\sigma1$=1) + 0.50*$N2(\mu2$ =8, $\sigma2$=1), Bootstraps=5000..................281
Table: Sample Median - MN1L99 Lower limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.50*$N1(\mu1$ =4, $\sigma1$=1) + 0.50*$N2(\mu2$ =8, $\sigma2$=1), Bootstraps=5000..................282
Table: Sample Median - C1U99 Upper limit sample median error rate (α = 0.01), Cauchy Distribution, Cauchy($x0$ =0, γ=1), Bootstraps=5000 288
Table: Sample Median - C1L99 Lower Limit Error rate (α = 0.01), Cauchy distribution, Cauchy($x0$ =0, γ=1), Bootstraps=5000 289
Table: Mean - N1U99 Advantage of EBSD(n) when comparing across Bootstrap Iteration Levels,...297
α = 0.01 significance level, upper limit 297
Table Mean - E1U99 Advantage of EBSD(n) when comparing across Bootstrap Iteration Levels,...299
α = 0.01 significance level, upper limit, α = 0.01 significance level, upper limit.......................299
Table Ratio of Sample Means - NU199 Advantage of EBSD(n) when comparing across Bootstrap Iteration Levels, α = 0.01 significance level, upper limit, α = 0.01 significance level, upper limit...301
Table: Pearson Correlation Coefficient - NU199 Advantage of EBSD(n) when comparing across Bootstrap Iteration Levels, α = 0.01 significance level, upper limit.................................303
Table DE599 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5..310
Table DE1099 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10..311
Table DE1499 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14..311
Table DE595 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5..315

Table DE1095 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10 315
Table DE1495 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14 316
Table DE590 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5 318
Table DE1090 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10 318
Table DE1490 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14 319
Table NDE599 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5 322
Table NDE1099 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10 322
Table NDE1499 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14 323
Table NDE595 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5 326
Table NDE1095 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10 326
Table NDE1495 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14 327
Table NDE590 Comparison of E-skew and Monte Carlo Bootstrap classification at 0.10 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5 329
Table NDE1090 Comparison of E-skew and Monte Carlo Bootstrap classification at 0.10 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10 329
Table NDE1490 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level among the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14 330

List of Figures

Figure 3.1. Comparison of Sampling Distributions for Mouse Data 38
Figure 3.2 Flow Chart of Simulation Study 56
Figure: Sample Mean - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Normal Distribution 74
Figure: Sample Mean - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Normal Distribution 75
Figure: Sample Mean - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Normal Distribution 76
Figure: Sample Mean - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Normal Distribution 77
Figure: Sample Mean - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Normal Distribution 78
Figure: Sample Mean - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Normal Distribution 79
Figure: Sample Mean - EU99 - One-Sided Upperlimit Error Rates for 99% CI for the Exponential Distribution 88
Figure: Sample Mean - EL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Exponential Distribution 89
Figure: Sample Mean - EU95 - One-Sided Upperlimit Error Rates for 95% CI for the Exponential Distribution 90
Figure: Sample Mean - EL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Exponential Distribution 91
Figure: Sample Mean - EU90 - One-Sided Upperlimit Error Rates for 90% CI for the Exponential Distribution 92
Figure: Sample Mean - EL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Exponential Distribution 93
Figure: Sample Mean - GU99 - One-Sided Upperlimit Error Rates for 99% CI for the Gamma Distribution 101
Figure: Sample Mean - GL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Gamma Distribution 102
Figure: Sample Mean - GU95 - One-Sided Upperlimit Error Rates for 95% CI for the Gamma Distribution 103
Figure: Sample Mean - GL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Gamma Distribution 104
Figure: Sample Mean - GU90 - One-Sided Upperlimit Error Rates for 90% CI for the Gamma Distribution 105
Figure: Sample Mean - GL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Gamma Distribution 106
Figure: Sample Mean - LNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Log-Normal Distribution 113
Figure: Sample Mean - LNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Log-Normal Distribution 114

Figure: Sample Mean - LNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Log-Normal Distribution 115
Figure: Sample Mean - LNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Log-Normal Distribution 116
Figure: Sample Mean - LNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Log-Normal Distribution 117
Figure: Sample Mean - LNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Log-Normal Distribution 118
Figure: Sample Mean - MNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Mixture of two Normals 125
Figure: Sample Mean - MNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Mixture of two Normals 126
Figure: Sample Mean - MNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Mixture of two Normals 127
Figure: Sample Mean - MNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Mixture of two Normals 128
Figure: Sample Mean - MNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Mixture of two Normals 129
Figure: Sample Mean - MNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Mixture of two Normals 130
Figure: Ratio of Sample Means - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Normal Distribution 142
Figure: Ratio of Sample Means - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Normal Distribution 143
Figure: Ratio of Sample Means - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Normal Distribution 144
Figure: Ratio of Sample Means - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Normal Distribution 145
Figure: Ratio of Sample Means - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Normal Distribution 146
Figure: Ratio of Sample Means - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Normal Distribution 147
Figure: Ratio of Sample Means - EU99 - One-Sided Upperlimit Error Rates for 99% CI for the Exponential Distribution 155
Figure: Ratio of Sample Means - EL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Exponential Distribution 156
Figure: Ratio of Sample Means - EU95 - One-Sided Upperlimit Error Rates for 95% CI for the Exponential Distribution 157
Figure: Ratio of Sample Means: EL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Exponential Distribution 158
Figure: Ratio of Sample Means: EU90 - One-Sided Upperlimit Error Rates for 90% CI for the Exponential Distribution 159
Figure: Ratio of Sample Means: EL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Exponential Distribution 160

Figure: Ratio of Sample Means: G99 - One-Sided Error Rates for 99% CI for the Gamma Distribution........166
Figure: Ratio of Sample Means: G95 - One-Sided Error Rates for 95% CI for the Gamma Distribution........167
Figure: Ratio of Sample Means: G90 - One-Sided Error Rates for 90% CI for the Gamma Distribution........168
Figure: Ratio of Sample Means: LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution........173
Figure: Ratio of Sample Means: LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution........174
Figure: Ratio of Sample Means: LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution........175
Figure: Ratio of Sample Means: MN99 - One-Sided Error Rates for 99% CI for the Mixture of two Normals 180
Figure: Ratio of Sample Means: MN95 - One-Sided Error Rates for 95% CI for the Mixture of two Normals 181
Figure: Ratio of Sample Means: MN90 - One-Sided Error Rates for 90% CI for the Mixture of two Normals 182
Figure: Pearson Correlation Coefficient - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Bivariate Normal Distribution........194
Figure: Pearson Correlation Coefficient - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Bivariate Normal Distribution........195
Figure: Pearson Correlation Coefficient - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Bivariate Normal Distribution........196
Figure: Pearson Correlation Coefficient - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Bivariate Normal Distribution........197
Figure: Pearson Correlation Coefficient - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Bivariate Normal Distribution........198
Figure: Pearson Correlation Coefficient - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Bivariate Normal Distribution........199
Figure: Pearson Correlation Coefficient - NNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Bivariate Normal Distribution........207
Figure: Pearson Correlation Coefficient - NNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Bivariate Normal Distribution........208
Figure: Pearson Correlation Coefficient - NNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Bivariate Normal Distribution........209
Figure: Pearson Correlation Coefficient - NNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Bivariate Normal Distribution........210
Figure: Pearson Correlation Coefficient - NNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Bivariate Normal Distribution........211
Figure: Pearson Correlation Coefficient - NNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Bivariate Normal Distribution........212
Figure: Trimmed Sample Mean - N99 - One-Sided Error Rates for 99% CI for the Normal Distribution........220

Figure: Trimmed Sample Mean - N95 - Error Rates for 95% CI for the Normal Distribution........ 221
Figure: Trimmed Sample Mean: N90 - One-Sided Error Rates for 90% CI for the Normal Distribution........ 222
Figure: Trimmed Sample Mean - E99 - One-Sided Error Rates for 99% CI for the Exponential Distribution........ 228
Figure: Trimmed Sample Mean - E95 - One-Sided Error Rates for 95% CI for the Exponential Distribution........ 229
Figure: Trimmed Sample Mean - E90 - One-Sided Error Rates for 90% CI for the Exponential Distribution........ 230
Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 99% CI for the Gamma Distribution........ 236
Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 95% CI for the Gamma Distribution........ 237
Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 90% CI for the Gamma Distribution........ 238
Figure: Trimmed Sample Mean - LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution........ 244
Figure: Trimmed Sample Mean - LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution........ 245
Figure: Trimmed Sample Mean - LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution........ 246
Figure: Trimmed Sample Mean - MN99 - One-Sided Error Rates for 99% CI for the Mixure of two Normals........ 252
Figure: Trimmed Sample Mean - MN95 - One-Sided Error Rates for 95% CI for the Mixure of two Normals........ 253
Figure: Trimmed Sample Mean - MN90 - One-Sided Error Rates for 90% CI for the Mixure of two Normals........ 254
Figure: Sample Median - N99 - One-Sided Error Rates for 99% CI for the Normal Distribution........ 262
Figure: Sample Median - N95 - One-Sided Error Rates for 95% CI for the Normal Distribution........ 263
Figure: Sample Median - N90 - One-Sided Error Rates for 90% CI for the Normal Distribution........ 264
Figure: Sample Median - E99 - One-Sided Error Rates for 99% CI for the Exponential Distribution........ 269
Figure: Sample Median - E95 - One-Sided Error Rates for 95% CI for the Exponential Distribution........ 270
Figure: Sample Median - E90 - One-Sided Error Rates for 90% CI for the Exponential Distribution........ 271
Figure: Sample Median - LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution........ 276
Figure: Sample Median - LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution........ 277

Figure: Sample Median - LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution.......... 278
Figure: Sample Median - MN99 - One-Sided Error Rates for 99% CI for the Mixture of two Normal Distributions.......... 283
Figure: Sample Median - MN95 - One-Sided Error Rates for 95% CI for the Mixture of two Normal Distributions.......... 284
Figure: Sample Median - MN90 - One-Sided Error Rates for 90% CI for the Mixture of two Normal Distributions.......... 285
Figure: Sample Median - C99 - One-Sided Error Rates for 99% CI for the Cauchy Distribution.......... 290
Figure: Sample Median - C95 - One-Sided Error Rates for 95% CI for the Cauchy Distribution.......... 291
Figure: Sample Median - C90 - One-Sided Error Rates for 90% CI for the Cauchy Distribution.......... 292
Figure DE99 Kappa Agreements for the Differentially Expressed Data Set at the 0.01 alpha level.......... 312
Figure DE95 Kappa Agreements for the Differentially Expressed Data Set at the 0.05 alpha level.......... 317
Figure DE90 Kappa Agreements for the Differentially Expressed Data Set at the 0.10 alpha level.......... 320
Figure NDE99 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.01 alpha level.......... 324
Figure NDE95 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.05 alpha level.......... 328
Figure NDE90 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.10 alpha level.......... 331

Chapter 1: Background and Significance

1.1 Background

Resampling methods are a subset of procedures used in statistics which involve drawing multiple new samples called resamples from the original data. Resampling methods are computer intensive and have become easier to implement as computing power has increased. These methods have advantages as they are useful in computing confidence intervals in many non-standard cases. For example, if the size of the sample is small and the population of interest is non-normal, non-parametric resampling methods can produce more accurate confidence intervals than normal theory methods do[2]. Further, many sample statistics have statistically intractable properties that resampling methods are equipped well to deal with. Although many resampling methods exist the non-parametric Monte Carlo Bootstrap method is perhaps the most widely used.

In the standard Monte Carlo Bootstrap although there is only one sample, resampling on this original sample with replacement is performed B times. For each of these B resamples the statistic of interest is computed to generate a Bootstrap sampling distribution. The Bootstrap sampling distribution is then used to compute a confidence interval for the statistic of interest.

One may ask, how many Bootstrap resamples are enough to provide accurate parameter estimation? Commonly the number of Bootstrap resamples varies from 200 to 10,000. More than 5,000 bootstrap resamples are recommended to prevent undue variation in confidence interval's implied p-values[3].

1.2 Introduction to the Problem

The Monte Carlo Bootstrap has two flaws. First the method has an inherent level of random repetition because each sample is generated by sampling with replacement from the original sample. This causes the original sample points to not be evenly represented in the resulting Bootstrap sampling distribution. Second, one user may choose to specify a different number of resamples than another. Consequently, both of these issues can cause different users to yield different results using the same data.

Therefore, a method that provides the benefits of the Bootstrap, eliminates simulation error and fixes the size of the sampling distribution should be of benefit to the statistician. The Efficient Bootstrap Sample Design method for a sample of size n (EBSD(n)) simultaneously provides these qualities[1].

The construction of EBSD(n) is a systematic approach for Bootstrap sample creation. This subset of all Bootstrap samples has a resulting variance of $\bar{X}$'s that is the same as the variance of $\bar{X}$'s from the complete enumeration of all possible samples. This subset of Bootstrap samples and the resulting sampling distribution for the statistic of interest can be used to estimate unknown parameters and is particularly helpful in the case of small sample sizes.

EBSD(n) although advantageous in eliminating previous issues inherent with the Monte Carlo Bootstrap, also introduces a new issue: What confidence interval method should be used on the resulting EBSD(n) sampling distribution? Confidence interval methods, previously derived to be used on a Bootstrap sampling distribution, may not work as well using a EBSD(n) sampling distribution.

In considering an algorithm that may perform well using $\text{EBSD}(n)$ one must understand problems Bootstrap methods grapple with. Common Bootstrap confidence interval methods are typically categorized into one of two confidence interval method convergence types: first order convergence methods and second order convergence methods. When analyzing non-normal data, first order convergence methods require larger sample sizes to converge to the correct pre-specified nominal error rates than do second order convergence methods[4]. Another issue, and this issue can be dealt with in a variety of ways is the issue of estimation of non-linear statistics. Second order Bootstrap methods like the Bootstrap-t perform can be less computationally efficient when the statistic computed is not the sample mean. For example, if there is no closed form solution for the sample variance of the statistic of interest the Bootstrap-t requires double bootstrapping which increases the total number of Bootstrap resamples to be performed. Therefore, a method that can adjust for non-normality and perform accurately and efficiently for non-linear statistics using the $\text{EBSD}(n)$ sampling distribution would be of interest and is the motivation for this work.

This dissertation will be broken into five chapters. In chapter 2, a literature review of the most used resampling method for confidence interval computation, the Bootstrap, will be presented. As a part of the literature review estimator bias, Monte Carlo bias, variance estimation, and transformation invariance are discussed. After discussing the issues raised by using the Monte Carlo Bootstrap method, confidence interval techniques that correct for these issues are presented.

Chapter 3 proposes the new approach, the E-skew method, for dealing with non-normal data and non-linear statistics when using $\text{EBSD}(n)$. Chapter 4 discusses the results and how the E-skew method and other methods commonly used with the Monte Carlo Bootstrap but applied on EBSD(n) compare to common first and second order Monte Carlo Bootstrap methods. The focus of the results is confidence interval performance measured by one-sided error rate. Further a real data example is presented to compare the performance of the E-skew method to common Bootstrap methods in a practical context. Finally, in chapter 5 the most important results from this work will be summarized. Further in chapter 5, motivation for additional research will be discussed with the possibility of further improving the methods laid out here.

Chapter 2: Literature Review

This chapter will discuss the concepts of parameter estimation and confidence interval computation mainly through the lens of resampling methods. Section 2.1 contains a discussion on the theory of sampling distributions and how the population sampling distribution is related to the non-parametric Bootstrap sampling distribution. In section 2.2, the parametric Bootstrap is discussed as basis of comparison with the non-parametric Bootstrap, though the non-parametric Bootstrap is of primary focus in this dissertation. Next in section 2.3 statistical bias is discussed. Here two forms of bias are of focus:

- Bias introduced by a biased estimator; if the estimator produces a biased estimate for the population parameter, the Bootstrap can be used to estimate the amount of bias via the bias function $b(\theta)$.
- Random bias introduced by the Monte Carlo Bootstrap; each element is not equally represented due to random resampling; some elements are chosen over others by chance from the original sample.

In the remainder of section 2.3, methods to measure and mitigate the bias introduced by the Bootstrap are discussed. In section 2.4 variance estimation is discussed both in terms of computation of the variance estimate and in terms of measuring the quality of the estimate. In section 2.5 many commonly used confidence interval methods are presented along with a description of each method; a discussion of the strengths and weaknesses of each method are also detailed.

2.1 Bootstrap Theory and the non-Parametric Bootstrap Sampling Distribution

Among resampling techniques, the Bootstrap is one of the most common and has prompted many types of confidence intervals. As mentioned in the previous chapter in the Monte Carlo Bootstrap, we draw n observations with replacement from the original data and perform this resampling B times to generate a non-parametric Bootstrap sampling distribution for the statistic of interest. We compute the statistic of interest $\hat{\theta}$ and use the non-parametric Bootstrap sampling distribution to generate a confidence interval around $\hat{\theta}$ at the desired $1-\frac{\alpha}{2}$ level.

The non-parametric Bootstrap sampling distribution is used as an estimate of the true population sampling distribution. This principle is at the core of our general approach in statistics, we estimate what we do not know from the information we have and often do this using the plug-in-principle[5].

A common form of the plug-in-principle is in computation of the standard error from the mean of the sample; the formula for the standard error is of course $\frac{\sigma}{\sqrt{n}}$ but in practice we do not know the true value for σ so instead we estimate σ by computing s^2 the unbiased estimator of σ^2. In the Bootstrap instead of estimating an individual population parameter using an unbiased estimator we estimate the entire sampling distribution by plugging in an estimate for it, the non-parametric Bootstrap sampling distribution[3]. In other words, the sampling distribution of $\hat{\theta}$ can be estimated by the non-parametric Bootstrap sampling distribution and then can be used to yield an accurate confidence interval for $\hat{\theta}$.

In fact, the central limit theorem dictates as $n \rightarrow \infty$ the sampling distribution of $\hat{\theta}$ and the non-parametric Bootstrap sampling distribution $\hat{\boldsymbol{\theta}}_{\boldsymbol{B}}^{*}$ in limit are bell shaped and approximately normally distributed. Further, if the distribution of $\hat{\theta}$ is not dependent on other unknown parameters the non-parametric Bootstrap sampling distribution offers a better approximation to the true population sampling distribution than does the approximation from the central limit theorem[6].

For many statistics $\hat{\theta}$, the sample standard deviation of $\hat{\theta}$ is difficult to compute. We use the standard deviation to measure the uncertainty of the estimate $\hat{\theta}$; this allows us to compute a confidence interval for $\hat{\theta}$. When computation of the standard deviation is intractable the non-parametric Bootstrap distribution allows us to bypass this problem. In this way the non-parametric Bootstrap resampling method facilitates confidence interval computation and has prompted the derivation of several confidence interval methods.

2.2 The Parametric Bootstrap

In the non-parametric Bootstrap, we generate B samples from $\boldsymbol{X}$ and define $P_{\hat{\theta}}$, the probability distribution function (PDF) of $\hat{\theta}$, to be the Bootstrap sampling distribution $\hat{\theta}_{B}^{*}$. Instead, with the parametric Bootstrap $P_{\hat{\theta}}$ is defined by a PDF based on a parametric assumption of the population the data was drawn from. Most commonly $P_{\hat{\theta}}$ is assumed to be from a normal model $N(\hat{\theta}, \frac{\hat{\sigma}^2}{n})$ where $\hat{\theta}$ and $\frac{\hat{\sigma}^2}{n}$ are estimated via maximum likelihood estimation. Once the maximum likelihood parameters are estimated B samples are drawn from this parametric model based on our underlying assumption of normality and the same techniques are then applied to these B samples as they would in the non-parametric case[7]. Of course, one does not need to assume a normal model to perform a parametric Bootstrap. Any statistical distribution can be chosen if it properly models the data of interest.

2.3 Bias

Choosing the right estimator is integral to confidence interval accuracy. The more bias introduced by the estimator the less accurate the interval will be. For resampling methods, bias can be introduced by using a biased estimator, by the resampling method (i.e. from Monte Carlo random error), or both. Further resampling methods can be used to estimate the amount of bias introduced by the estimator. In this section bias is introduced first by considering the case where no resampling is performed. In this case bias is measured in the original estimator by making distributional assumptions; distributional assumptions allow bias to be defined by a bias function. The bias function can then be used to negate the bias introduced in the original estimator. Although it may seem like removing bias would always be desirable, depending on the estimator it can do more harm than good.

Although we may be able to define the bias introduced, trying to eliminate it using a bias function can be problematic. Following a discussion of the issue with implementing bias reduction, methods to assess bias from estimators and resampling methods are listed. Methods better at estimating bias converge to the bias in limit as $B \to \infty$ at a smaller number of resamples and perform well for a greater number of estimators.

Bias Adjustment and the Effect on Mean Square Error

Estimating and negating bias from an estimator can lead to a more accurate estimator however this is not always the case. Before discussing methods for estimating bias using resampling techniques, the theoretical bias of an estimator when no resampling performed is considered. Bias and mean square error are both important to consider for an estimator $\hat{\theta}$ and are defined below.

By making distributional assumptions on the population from which $\hat{\theta}$ is drawn, the bias of the estimator can be defined functionally. If $b(\theta)$, the bias of the estimate $\hat{\theta}$ from the true population parameter, is non-zero then bias is defined as[6]:

$$b(\theta) = E_{\theta}(\hat{\theta}) - \theta$$

This means if $b(\theta) > 0$, it implies $E_{\theta}(\hat{\theta}) \not\equiv \theta$ and the $\hat{\theta}$ distribution is not centered on the unknown value of θ. Instead, it is biased by the amount $b(\theta)$ where $b(\theta)$ is defined as an unknown function of θ. In theory, if we desire to eliminate bias, although $b(\theta)$ is unknown, we can estimate $b(\theta)$ with $b(\hat{\theta})$. Then, the estimate $\hat{\theta}$ can simply have it's bias removed by subtracting $b(\hat{\theta})$ seen below where $\hat{\theta}_{br}$ is the bias corrected estimate[6]:

$$\hat{\theta}_{br} = \hat{\theta} - b(\hat{\theta})$$

Bias correction can lead to additional variability in the estimate of the parameter due to additional variability introduced by $b(\hat{\theta})$, i.e., $b(\hat{\theta})$ is not necessarily equal to $b(\theta)$. Although the increased variability impacts the mean square error, the direction of the impact on the mean square error varies depending on the estimator; it can cause the mean square error to be larger or smaller depending on the variability introduced by $b(\hat{\theta})$[6]. If using $b(\hat{\theta})$ in the computation of $\hat{\theta}_{br}$ leads to an increase in the mean square error of the estimate of θ, it trades one problem in for another.

In the following sections below we attempt to estimate the bias function $b(\hat{\theta})$ using resampling methods as bias reduction techniques are rampant in parameter estimation. However, that discussion is preempted with this section to note this does not always lead to greater confidence interval accuracy. Theoretically, functional effective bias reduction is dependent on making proper assumptions about the distribution of the population parameter which may not be known.

Bootstrap Bias Corrected Estimate

When using resampling methods there are multiple approaches to bias estimation and bias adjustment. In practice when choosing to use the Bootstrap we do not know the bias function $b(\theta)$ but we can still use the Bootstrap to estimate the bias function without the knowledge of $b(\theta)$. As we already know we can get a sampling distribution of estimates $\widehat{\theta_1^*}$, . . ., $\widehat{\theta_B^*}$ from the Bootstrap and we can then calculate the average of these estimates[6]:

$$\bar{\theta}_B^* = \frac{1}{B}\sum_{i=1}^{B} \hat{\theta}_i^*$$

Therefore an accurate approximation of $b(\hat{\theta})$ should be $2\hat{\theta}$ - $\bar{\theta}_B^*$. The greater the number of iterations (B) we perform the more accurate this approximation will be and since computing power is cheap attaining sufficiently large B should be readily attainable[4]. The Bootstrap bias corrected estimate above is defined as[6]:

$$\hat{\theta}_{br}^* = \hat{\theta} - (\bar{\theta}_B^* - \hat{\theta})) = 2\hat{\theta} - \bar{\theta}_B^* .$$

For large enough B this will be indistinguishable from $\hat{\theta}_{br}$ and be as if the estimate of the bias when the function $b(\theta)$ is known. This of course is dependent on the sample taken and the estimator used. If the sample taken is not representative or the population assumption is incorrect this correction and the others discussed will only provide limited parameter estimation improvement[6].

Better Bootstrap Bias Estimate

A bias corrected Bootstrap estimate that minimizes bias at smaller sample sizes faster than the Bootstrap bias corrected estimate is called the better Bootstrap bias estimate; defined as[8]:

$$\overline{bias}_B = \bar{\theta}_B^* - T(\bar{\boldsymbol{P}}^*) \ ,$$

$\boldsymbol{P^0}$**,** is a vector with a uniform probability, $1/n$, of picking each sample point if resampling from the original data. Further $\boldsymbol{P}^{*\mathbf{1}}$ is the proportion of times each sample point is selected from the first resample. Thus $\bar{P}^* = \sum_{i=1}^{B} \bar{\boldsymbol{P}}^{*i} / B$ represents the average probability for each sample point across the B resamples. This bias estimate converges faster than the Bootstrap bias corrected estimate and therefore is considered a preferable method[8]. Because computing power is cheap the inferiority of the Bootstrap bias corrected estimate can simply be made by increasing the number of resamples B[8]. It is worth mentioning here using a confidence interval method requiring double bootstrapping (bootstraps of each Bootstrap resample), like the Bootstrap-t method, could add more than a minimal amount to the run time[8].

The Jackknife estimate of Bias

The Jackknife was the original computer-based method for estimating bias and standard errors. To perform the Jackknife from a sample of size n, from a data set $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the Jackknife sample $\boldsymbol{x}_i$ is defined as $\boldsymbol{x}$ with data point i removed. Hence

$$\boldsymbol{x}_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

The Jackknife estimate of bias is then[8]:

$$\widehat{bias}_{jack} = (n-1) * (\hat{\theta}_{J(.)} - \hat{\theta}), \text{ where } \hat{\theta}_{J(.)} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\theta_i^*} \ .$$

This formula breaks down when the statistic is unsmooth like the sample median, but it works well for smooth plug-in statistics like the mean or the ratio of means, more technically if $\hat{\theta} = T(P^*)$ is twice differentiable[8].

This formula is derived based on the quadratic statistic:

$$T^{QUAD}(P^*) = c_0 + (P^* - P^0)^T U + (\tfrac{1}{2})(P^* - P^0)^T V (P^* - P^0)$$ [18]

Where c_0 is the original $\hat{\theta}$ estimate from the sample, U is an n-vector satisfying $\sum_1^n U_i = 0$ and V is an $n \times n$ symmetric matrix satisfying $\sum_i V_{ij} = \sum_j V_{ij} = 0$ for all i, j. Further P^* is defined as vector of probabilities $(P_1^*, \ldots\ P_n^*)^T$ satisfying $0 \le P_i^* \le 1$ and $\sum_1^n P_i^* = 1$ and where P^0 is defined as the probability vector of each element being equally likely to be chosen from the original sample[9]:

$$P^0 = (\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n})^T.$$

This approximation agrees closely with the ideal Bootstrap estimate $\widehat{bias}_\infty$ and it does so with far fewer samples than is required for the Bootstrap. Ultimately, the method in this section and the previous two are all trying to approximate the ideal $bias_\infty$[8].

Balanced Bootstrap

In multiple sections above, the Bootstrap was used to estimate the bias function $b(\theta)$ for an estimator of the true population parameter. In some cases, we know the estimator is an unbiased estimate of the true population parameter, for example we know $\bar{X}$ is an unbiased estimator of μ i.e. $E(\bar{X})$ = μ. Although this is the case, bias is still introduced from Monte Carlo resampling when generating the Bootstrap sampling distribution. Random resampling does not guarantee each sample point is given equal weight and this random error creates bias.

One approach to eliminating this issue is eliminating this Monte Carlo error through the balance Bootstrap procedure[10]. Instead of randomly resampling, B copies of the original sample are created, and this string of $n * B$ samples is permuted. The permutation is meant to mimic the variation created by using Monte Carlo Bootstrap resampling while retaining the property that each element is represented an equal number of times. After the permutation, the string is split into B successive samples which can be used to generate a Bootstrap sampling distribution[10]. The Monte Carlo bias elimination referred to here is considered first order bias reduction or first order balance.

A second approach to generating a balanced design is performed by generating n copies each of size n from the original sample[10]. The resulting matrix is a randomized block design where not only is each element represented n-times but also column balance is obtained; each element is represented only once in each column[10]. Then to achieve B resamples we create k randomized block designs, where $B = n * k$. This column balancing is useful for complex non-linear statistics[10]. This is also considered first order balance.

Example First Order Balance

Take the original sample to be the integers 1, 2, 3, 4, 5. Now say we generate 10-copies of this original sample and concatenate the resamples in a string hence our string is:

12345123451234512345123451234512345123451234512345

In the first first-order balanced Bootstrap method discussed, we randomly permute this string and results in the string:

43542143415454313555113313422412515431232125432252

We then have the resulting matrix of resamples where i denotes the i^{th} column and B denotes the B^{th} resample:

Table 1. Random Permutation Design

	i				
B	**1**	**2**	**3**	**4**	**5**
1	4	3	5	4	2
2	1	4	3	4	1
3	5	4	5	4	3
4	1	3	5	5	5
5	1	1	3	3	1
6	3	4	2	2	4
7	1	2	5	1	5
8	4	3	1	2	3
9	2	1	2	5	4
10	3	2	2	5	2

In the second first-order balanced Bootstrap method discussed again we again have the original string:

12345123451234512345123451234512345123451234512345

But we first partition this string into two identical strings of length 25:

- 1234512345123451234512345
- 1234512345123451234512345

Because the original sample is of length 5 and we create n-copies as mentioned above. We split the original string into two strings of length 25 because the string length is now the square of the original sample size. This allows two randomized block designs to be created, one example of this is displayed below:

Table 2. Randomized Block Design

Design	i					
1	**B**	**1**	**2**	**3**	**4**	**5**
	1	3	2	4	5	3
	2	4	3	3	2	4
	3	2	4	5	1	2
	4	1	1	2	3	5
	5	5	5	1	4	1
2						
	6	5	2	4	4	5
	7	4	3	5	2	2
	8	2	5	3	5	3
	9	3	4	2	1	4
	10	1	2	1	3	1

Generally, a design is said to be r ordered balanced if all n^r possible sequences are represented an equal number of times[10]. I.e., if r=2 second order balance means all possible pairs (1, 2), (1, 3) etc... occur an equal number of times for any r=2 columns[10]. Second order balance is discussed further below.

Second-Order Balanced Bootstrap

Second order balance requires all n^2 values occur with equal frequency for every possible pair of columns[10]. Two separate approaches can be used to obtain second order balance (1) using orthogonal Latin squares and (2) using balanced incomplete block design.

A Latin square is a randomized block design of size n^2 where each element from the sample occurs exactly one time for each row and each column. For the Latin square approach take the first column to be $(1, . . , 1, 2, ... ,2, n, . . , n)$ and the second column to be $(1, 2, 3, 4, 5, ... , 1, 2, ... , n)$ then for any other column j, the equal pair condition is obtained with each of columns one or two if the elements in the remaining columns correspond to those in a Latin square. If the same can be said for columns, $3, ... , n$ then each of the remaining possible pairs of columns must also correspond to a Latin square. Further this also implies the successive Latin squares must be orthogonal. A complete design must have $n - 2$ orthogonal Latin squares and is only available if n is a power of a prime number. In this case to obtain second order balance the number of bootstraps must be B $\geq n^2$. A proof yielding exact second order correctness is shown in Hinkley et all[10].

In the balanced incomplete block design method, we have $B = kn$ blocks where each block retains the size of the original sample. Two primary conditions must be met for the transpose of the incidence matrix are (1) the diagonal elements of the concordance matrix NN' are all equal to $k(2n - 1)$ and (2) the off diagonal elements are all equal to $k(n - 1)$. In deriving the design initially k blocks are chosen. Then to create the second set of k blocks, the index of the initial set of k blocks is modified by adding 1. This process is repeated cyclically for index modifications of size 2 through $(n - 1)$ to generate the kn total blocks[10].

Both the Latin square and BIBD methods yield variance approximations correct to the order $(1/n)$ as if performing Taylor Expansion[10]. The BIBD method is a good primer before talking about EBSD(n) as it is used to derive second order balance for EBSD(n).

2.4 Variance Estimation

After generating resamples using either the Monte Carlo Bootstrap or another resampling process the next step is to determine the range of values at a significance level α the parameter may take. The uncertainty of the parameter can be obtained by estimating the variance of the statistic of interest. Suppose we obtain an estimate $\hat{\theta}$ of the true parameter; the variance of the estimator $\hat{\theta}$ is defined as $\sigma_{\hat{\theta}}^2$. Variance estimates assess the quality of the estimate $\hat{\theta}$ and are used to construct a confidence interval around $\hat{\theta}$. Particularly if the distribution of $\hat{\theta}$ is approximately normally distributed the estimated variance should provide an accurate assessment on the certainty of the estimate. A commonly used method the Bootstrap-t method can be used to optimize the effectiveness of variance in cases where the data is not approximately normally distributed. In this section three types of variance estimation procedures are mentioned, the variance substitution method, the Jackknife method, and the Bootstrap implementation of the substitution method. Following the discussion of these variance estimation methods, the Jackknife-after-Bootstrap is discussed as a tool for measuring the quality of the variance estimate.

Substitution Variance Estimation

An obvious variance estimation approach is to calculate the variance of the statistic simply by using the formula for the variance of the statistic based on the data from the original sample with no resampling used[11].

An obvious example of this is estimating the variance of the mean in a one sample problem. In this case the variance of the mean statistic would be calculated by estimating the variance of the sample mean using the traditional formula $(n - 1)^{-1} \sum(X_i - \bar{X})^2$.

Jackknife Variance Estimation

When $\boldsymbol{X}$ represents a random sample of size n, i.e., $\boldsymbol{X} = (X_1, \ldots, X_n)$, the jack knife method can be used to provide such variance estimation. Let $\hat{\theta}_{(-\mathrm{i})}$ denote the estimate $\hat{\theta}$ and

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(-\mathrm{i})} \ .$$

The variance $\sigma_{\hat{\theta}_J}{}^2$ is then estimated by the Jackknife method via[11]

$$\hat{\sigma}_{\hat{\theta}_J}{}^2 = \frac{n-1}{n} \sum_{i=1}^{n} (\widehat{\theta^*_{-\iota_\iota}} - \hat{\theta}_{(.)})^2$$

This approach to variance estimation provides an accurate estimate for the sample mean statistic when the data is approximately normally distributed, but it does not for all statistics and distributional types. For example, if θ and $\hat{\theta}$ are the population and sample median, the Jackknife estimate is inaccurate for large samples[11]. Also, for vectors generated from more complex probability structures Jackknife variance estimates are not as accurate or effective[12].

Bootstrap Variance Estimation

Although a functional form for $\sigma^2_{\hat{\theta}}$, the sample variance, would be ideal for any sample statistic, often a functional form for the variance of the sample statistic is not available. The Bootstrap method provides a very simple algorithm for getting accurate approximations to $\hat{\sigma}^2_{\hat{\theta}}$[11].

Let the statistic of interest be $\hat{\theta}$ with variance $\sigma^2_{\hat{\theta}}$, and let $\boldsymbol{\hat{\theta}^*_B}$ be the Bootstrap sampling distribution with variance $\sigma^2_{\hat{\theta}^*{}_B}$. For each Bootstrap generated sample there is a $\hat{\theta}^*{}_B$ generated by calculating the statistic from the Bootstrap sample. Thus, we yield Bootstrap sample estimates $\hat{\theta}^*{}_1,\ ,\ldots,\hat{\theta}^*{}_B,$ and can compute the variance of these Bootstrap estimates as

$$s^2{}_B = \hat{\sigma}^2_{\hat{\theta}_B} = \frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i{}^* - \bar{\theta}^*)^2 \text{ where } \bar{\theta}^*_B = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i{}^*.$$

Hence an unbiased estimate of $\sigma^2_{\hat{\theta}}$ is the sample variance of the Bootstrap sampling distribution whose accuracy can be controlled by selecting large B[11]. Thus $\hat{\sigma}^2_{\hat{\theta}_B}$ can be considered a substitution variance estimate of $\hat{\sigma}^2_{\hat{\theta}}$. Take for example estimating the sample mean using $\hat{\theta} = \bar{X}$. We know:

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n},$$

where σ is the standard deviation of $\bar{X}$. The substitution principle would estimate σ^2/n by $\hat{\sigma}^2/n$, where:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - X)^2 \ .$$

This $\hat{\sigma}^2$ is the variance of $\bar{X}$, which places probability $1/n$ on each of the X_i. Instead of using the sample formula to compute $\sigma^2_{\bar{X}}$, the Bootstrap variance estimation method generates B samples, of size n each, and computes the B sample averages $\bar{X}_1^*, \ldots, \bar{X}_B^*$ of these samples[11]. For large B the sample variance

$$\hat{\sigma}^2_{\bar{X}_B} = \frac{1}{B-1}\sum_{i=1}^{B}(\bar{X}_i^* - \bar{\bar{X}}^*)^2 \text{ , where } \bar{\bar{X}}^* = \frac{1}{B}\sum_{i=1}^{B}\bar{X}_i^*$$

will then be an accurate approximation of $\hat{\sigma}^2/n$.

It is not necessary to conduct this many simulations and compute the sample variance for the mean from such a large number of Bootstrap sample. We can just compute the sample variance using the formula for the variance of the sample mean. However, for statistics other than the sample mean this approach is not always available, whereas the Bootstrap method is applicable universally for the variance of any sample statistic of interest[11].

Jackknife-after-Bootstrap

The Jackknife-after-Bootstrap method can also be an efficient estimate of the uncertainty of the standard error generated from the Bootstrap.

Consider an example from Efron and Tibshirani's *An Introduction to Bootstrap* on page 11 of the textbook are 7 sample points for survival time in days of mice following a test surgery which is considered the treatment group.

Table 3. Survival Time (Days) Treatment group Mouse Data

X_1	X_2	X_3	X_4	X_5	X_6	X_7
94	197	16	38	99	141	23

The Jackknife-after-Bootstrap estimation for the standard error of the sample statistic is systematically achieved from each sample point by only considering those resamples from the Bootstrap that exclude that sample point. For example, for X_1 =94 only those resamples from the Monte Carlo Bootstrap that do not include this sample point are considered and then for each resample considered the statistic of interest is calculated. Thus $\widehat{se}_{b(1)}$ is the sample standard deviation of $\hat{\theta}_{(i_1)}, \hat{\theta}_{(j_1)}, \dots, \hat{\theta}_{(k_1)}$ for the first element from the sample where resamples i, j, k do not include the first sample element (94). Then, in this case, $\widehat{se}_{b(i_e)}, \dots, \widehat{se}_{b(k_e)}$, where $\widehat{se}_{b(.)} = \sqrt{\frac{\hat{\sigma}^2_{\hat{\theta}}}{n-1}}$ are computed in a similar way for each of the remaining elements, respectively[13].

Once these seven standard errors have been computed we take the sample variance of these 7 values to obtain $\widehat{se}_{jack}(\widehat{se}_b)$. This resulting standard error can be measured against the standard error generated from the Monte Carlo Bootstrap. Particularly at a smaller number of Bootstrap resamples, the Jackknife-after-Bootstrap is an overestimate[13].

2.5 Confidence Bounds

Below, several Bootstrap confidence interval methods will be discussed along with the strengths and weaknesses of each. The t-interval with Bootstrap estimated standard error, the percentile method and the basic percentile method will be introduced first. These commonly used methods each are first order accurate and have inefficiencies when compared to more accurate second order methods.

After introducing these methods, the second order accelerated bias corrected method and the Bootstrap-t method will be discussed. The theory used in the accelerated bias corrected method is used as motivation for the confidence intervals proposed for EBSD(n). The Bootstrap-t produces the most accurate upper limit one-sided error rate intervals for positively skewed data when the statistic is the sample mean[3].

Monte Carlo Bootstrap Confidence Interval Methods

Percentile Method

The percentile method simply uses the $\alpha/2$ and $(1 - \frac{\alpha}{2})$ percentiles of the Monte Carlo Bootstrap sampling distribution for the $1 - \alpha$ two-sided confidence interval. The percentile method is a first order method. Percentile intervals can be truncated at small sample sizes based on normal theory assumptions[3]. When the statistic computed is biased in nature the percentile interval is affected by bias twice in computation. First from the computation of the original statistic and second by a biased Monte Carlo Bootstrap distribution derived from a biased statistic[3]. If skewness exists in the underlying data, the percentile interval fails to properly compensate for the asymmetry. When skewness in the data exists, the interval needs to reach many standard errors to the right or left depending on the asymmetry[3]. Even though the percentile method has a level of asymmetry it has only approximately 1/3 the level of asymmetry provided by the skewness corrected t-interval method (derived from asymptotic means)[3].

Another issue in using the percentile method is if one is calculating an interval for the sample mean $\bar{X}$ at small sample sizes. The percentile interval does not extend far enough out on either end, i.e. it is too narrow[3]. This is because for the sample mean $\bar{X}$ the percentile interval is using normal theory and the implied $\hat{\sigma}^2$ is $\left(\frac{1}{n}\right) * \sum (X_i - \bar{X})^2$ rather than s^2. Therefore, on this fact alone the percentile interval is short by a factor $\sqrt{(n-1)/(n)}$. Further because the algorithm assumes $\bar{X}$ is drawn from symmetric data this implies the use of $z_{\alpha/2} * \hat{\sigma}/\sqrt{n}$ in place of $t_{\frac{\alpha}{2},n-1} * \hat{s}/\sqrt{n}$ and further truncates the size of the interval[3].

Despite these issues of bias and skew the percentile method does have one advantage; it is transformation invariant. This is an advantageous property because it says no matter the method used the results will be the same, i.e. the results are independent of the way the probability model for the data is parametrized.

T-interval with Monte Carlo Bootstrap Estimated Standard Error Method

This method estimates the variance of the statistic by calculating the variance of the Monte Carlo Bootstrap distribution. The degree of the Monte Carlo error effects the resulting confidence interval and will vary depending upon the samples selected from the Monte Carlo Bootstrap.

$$C.I. = \hat{\theta} \pm t_{1-\frac{\alpha}{2},n-1} * \hat{s}_B,\ \hat{s}_B = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i^* - \bar{\theta}^*)^2},\ \text{where } \bar{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^{*}$$ [3]

The t-interval with Bootstrap standard error is a first order accurate method. In the case of the sample mean this interval offers little advantage over non-resampling methods and the use of the standard t-interval[3]. This first order method also has a few additional issues worth mentioning. First the method is not transformation invariant; if one applies a strictly increasing transformation $\hat{\theta}$, $g(\theta)$, one will yield at least slightly different coverage probabilities for $g^{-1}(g(\theta_U))$ than for θ_U[3]. This could lead to a difference in interpretation when indirectly inferring from the interval! Second the confidence interval is shortened by a factor of $\sqrt{(n-1)/(n)}$ because the empirical distribution has theoretical variance $\hat{\sigma}^2 = \left(\frac{1}{n}\right) * \sum\ (x_i - \bar{x})^2$ rather than s^2. Finally, if skew exists in the data the t-interval requires a large sample size to converge to the correct coverage probabilities and it converges even more slowly than the percentile method mentioned in the previous section[3].

One advantage however is the amount of bias introduced by the interval. As discussed previously, if bias exists in the underlying statistic there is no mechanism to correct for this bias. Although this is the case, the method only applies one copy of bias in the computation of the confidence interval: the bias inherent in the computation of the original statistic $\hat{\theta}$[3].

Basic Method

The basic percentile confidence interval method is like the percentile method except it corrects for simple bias in the underlying statistic. It is the mirror image of the percentile confidence interval in that it reaches as far above θ as the percentile reaches below[3]. The basic confidence interval is of the form[3]:

$$[F^{*-1}(\Phi(2z_0 - z_1)),\ F^{*-1}(\Phi(2z_0 + z_1))]$$ [3]

where,

$$z_0 = \Phi^{-1}(F^*(\hat{\theta})),\ z_1 = \Phi^{-1}(1-\alpha)) \text{ and}$$

$F^*(.)$ = the Monte Carlo Bootstrap sampling distribution for the parameter of interest.

If the only underlying issue in the data is simple bias the basic percentile method will appropriately fix the issue and provide more accurate confidence bounds than the percentile method. However, there are two major problems with this approach. First the Monte Carlo Bootstrap sampling distribution of $\hat{\theta}^* - \hat{\theta}$ is highly dependent on the distribution of $\hat{\theta}$. This implies $\hat{\theta}^*$ is a good approximation for $\hat{\theta}$ - θ, precisely when $\hat{\theta}$ = θ[3]. This is problematic because it requires our sample to perfectly represent the population parameter based on the sample statistic. Second, if any skewness exists in the data or the statistic is non-linear in nature the bias correction causes the interval to be asymmetric in the wrong direction[3].

The basic percentile interval is considered to use the wrong pivot, $\hat{\theta}^* - \hat{\theta}$, "forward" whereas Efron's percentile uses the wrong pivot "backwards"[3]. The issue with both approaches of course is one is choosing a statistic that is not pivotal. A pivotal statistic is one whose distribution is independent of the population parameter[3]. In the following sections we will look at methods attempting to correct these problems mentioned so far.

Bias-Corrected Accelerated Adjusted Percentile Method

When our estimate $\hat{\theta}$ consistently underestimates or overestimates the target θ a bias correction might help matters in computing confidence intervals. This led Efron to propose the following bias corrected accelerated adjusted percentile Bootstrap method. It is as easily implemented as the ordinary percentile method and it generally improves accuracy[14]. Transformation invariance is maintained, but there is a somewhat arbitrary link to the normal distribution[14]. However, for not so small samples a case can be made the normal approximation is appropriate when dealing with properly transformed estimates. Below is the general definition of the bias corrected accelerated percentile method, and it is noted exact coverage can be obtained when the distribution of $\hat{\theta}$ only depends on θ assuming some other normalizing conditions apply[14].

The bias adjustment is a median bias adjustment derived from the Monte Carlo Bootstrap distribution and the Jackknife procedure is often used to estimate the acceleration. The bias-corrected, accelerated adjusted (BC_a) confidence interval is computed in the form[15]:

$$[\hat{\theta}_{(\alpha_1)}{}^{*}, \hat{\theta}_{(\alpha_2)}{}^{*}]$$

where

$$\alpha_1 = \Phi(\widehat{z_0} + \frac{\widehat{z_0} + z_{(\alpha/2)}}{1-\hat{a}(\widehat{z_0} + z_{(\alpha/2)})}),\ \alpha_2 = \Phi(\widehat{z_0} + \frac{\widehat{z_0} + z_{(1-\alpha)/2}}{1-\hat{a}(\widehat{z_0} + z_{(1-\alpha)/2})}),\ \hat{a} = \frac{\sum_{i=1}^{n}(\hat{\boldsymbol{\theta}}_{(.)} - \hat{\boldsymbol{\theta}}_{i})^{3}}{6(\sum_{i=1}^{n}((\hat{\boldsymbol{\theta}}_{(.)} - \hat{\boldsymbol{\theta}}_{i})^{2})^{3/2}} \text{ and,}$$

$$\widehat{z_0} = \Phi^{-1}((\sum_{i=1}^{B} \hat{\theta}_i^* < \hat{\theta})/B)$$

The estimate $\hat{a}$ is derived as an estimate of the skew in the data to ensure accurate confidence intervals when the sample is not approximately normally distributed. The median bias is the $\widehat{z_0}$ statistic calculated above. In choosing the percentile to be taken from the Monte Carlo Bootstrap distribution α_1 and α_2 percentiles are chosen to reflect this adjustment for bias and skewness. Under regularity conditions, the BC_a coverage errors can be shown to converge on the order $1/n$, implying the confidence interval is second order accurate[15].

Bias-Corrected Accelerated Adjusted Percentile Method Continued: ABC Method

A minor disadvantage of using the Bootstrap as a means of generating confidence intervals is the computational burden of generating tens of thousands of bootstraps and sub-bootstraps. The ABC method for computing confidence intervals does not depend on generating resamples and rather approximates the BC_a interval by computing numerical derivatives[16].

The ABC confidence limit for θ, denoted $\hat{\theta}_{ABC}[1-\alpha]$, is constructed as follows:

$$w \equiv \hat{z}_0 + z^{(1-\alpha)},\ \lambda \equiv \frac{w}{(1-\hat{a}w)^2},\ \hat{\delta} \equiv \dot{T}(\boldsymbol{P}^0),\ \hat{\theta}_{ABC}[1-\alpha] = T(\boldsymbol{P}^0 + \frac{\lambda\hat{\delta}}{\hat{\sigma}})$$ [20]

where $\boldsymbol{P}^0 = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})^T$, and the function is defined as[16]

$$\dot{T}_i = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)\boldsymbol{P}^0 + \epsilon \boldsymbol{e}_i\right) - T(\boldsymbol{P}^0)}{\epsilon}$$

The ABC procedure computes the constants $\hat{z}_0$ and $\hat{a}$ generated with the BC_a but rather than doing so using Bootstrap replications it does so through numerical second derivatives[16].

The acceleration constant $\hat{a}$ is 1/6 times the standardized skewness of the empirical influence components[20]:

$$\hat{a} = \frac{\sum_{i=1}^{n} \dot{T}_i^3}{6(\sum_{i=1}^{n} \dot{T}_i^2)^{3/2}} \text{ where, } \dot{T} = \frac{\lim_{\epsilon \to 0} T((1-\epsilon)\boldsymbol{P}^0 + \epsilon \boldsymbol{e}_i) - T(\boldsymbol{P}^0)}{\epsilon}$$

Bootstrap-t Method

The Bootstrap-t method also provides better confidence intervals in the case of skewed data for the mean statistic[3]. For each resample of the Monte Carlo Bootstrap method a t-statistic is calculated. This collection of t-statistics is the Monte Carlo Bootstrap sampling distribution of interest and is used in deriving the confidence interval. The t-statistic for each resample is calculated as:

$$t^*{}_B = \frac{\hat{\theta}_B^* - \hat{\theta}}{\hat{s}_{\widehat{\theta^*}_B}}$$

$\hat{\theta}_B^*$ is the estimate of the population parameter for each resample, $\hat{\theta}$ is the estimate of the parameter from the original sample, and $\hat{s}_{\widehat{\theta^*}_B}$ is the estimate of the standard error of the population parameter for each resample. Then for the computation of the confidence interval the (1- $\frac{\alpha}{2}$) and $\frac{\alpha}{2}$ percentiles of the ordered Monte Carlo Bootstrap sampling distribution are taken to determine the t-statistics used for the confidence interval computation. The confidence interval takes the form[3]:

$$(\hat{\theta} - t^*_{(1-\frac{\alpha}{2})}\frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}},\ \hat{\theta} - t^*_{\frac{\alpha}{2}}\frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}})$$

Although the Bootstrap-t method is second order accurate, like the t-interval with Bootstrap standard error method, it is not transformation invariant[3]. Also, computing the t-statistic makes most sense when θ is a location parameter (i.e., mean, median), but that is not always the statistic of interest[14]. If instead we are estimating the correlation with $\theta = \rho$, we cannot treat ρ as a location parameter and methods using the t-statistic perform poorly. The Fisher-z transform below has been suggested as an alternative approach[8]:

$$z = \frac{1}{2} log\{(1 + \rho)/(1 - \rho)\}$$

where after computing the confidence bounds for z, the bounds are then back transformed to confidence bounds for ρ.

This approach is helpful because it normalizes the correlation statistic and has been shown to improve coverage properties. This transformation is implemented in the simulation work in chapter 4 by applying the Fisher-z transformation to both Bootstrap and EBSD(n) methods for the correlation statistic.

A further disadvantage of the Bootstrap-t method is the source of its better coverage properties, namely the requirement of an appropriate scale estimate (standard error) for the method's confidence interval[14]. If the formula for the sample variance of the statistic is not known, we must calculate Bootstrap variance estimates on each Bootstrap resample to achieve the appropriate standard error estimate for each ${t^*}_B$. Performing bootstraps of each Bootstrap resample can become computationally intensive, relative to other confidence interval methods, especially when many confidence interval computations are required.

Non-resampling Second Order Confidence Interval Methods

T-skew Method

The t-skew method is not a resampling method but rather a confidence interval method that adjusts the t-statistic for skewness. The t-statistic adjustment is based on skewness corrected t-statistic. The t-skew corrected confidence interval can be implemented in the following form for the sample mean[3]:

$$\bar{X} + \left(\gamma * \frac{1}{6*\sqrt{n}} * (1 + 2 * {t_{\frac{\alpha}{2},n-1}}^2) \pm t_{\frac{\alpha}{2},n-1} \right) * \frac{\hat{s}}{\sqrt{n}},\ where\ \gamma = \frac{\left(\frac{1}{n}\right) * \sum_{i=1}^{n} (X_i - \bar{X})^3}{\hat{s}^3}$$

The motivation for this adjustment is the t-statistic has been shown to be twice as skewed as $\bar{X}$ in the opposite direction of the skew of $\bar{X}$[3]. Further if drawing a sample from exponentially distributed data; data with an inherent level of skew (γ =2), means there is a strong dependent relationship between $\bar{X}$ and the standard error computed from the sample. This dependent relationship causes the acceleration we are adjusting for in the BC_a method. Thus, adjusting for skewness in the t-statistic is necessary to produce the most accurate confidence intervals[3]. For many statistics, the sample variance is unknown, thus a logical approach would be to estimate it using a resampling method. This estimate could then be plugged into the formula above.

2.6 Research Questions

The primary research question is how the accuracy of the methods using EBSD(n) compare to methods commonly used with the Monte Carlo Bootstrap. Confidence interval accuracy was evaluated using the type 1 one-sided error-rate for each side of the confidence interval. Once the one-sided error rate for both the upper and lower limit were computed, the percent error from the theoretically true one-sided error rate was computed.

Confidence interval methods do not perform uniformly accurately for all significance levels, statistics, and distributional types. Thus, secondary research questions investigated were:

- How did method accuracy vary across alpha significance level?
- How did method accuracy vary for different types of probability distributions?
- How did method accuracy vary when the statistic was linear vs. non-linear?

Four specific comparisons of interest in comparing confidence interval methods for the EBSD(n) method to confidence interval methods for the Monte Carlo Bootstrap were:

- Performance for the mean statistic when the distribution was skewed in nature
- Performance for the Pearson correlation coefficient

- Performance for the ratio of two independent means statistic
- Performance for data drawn from the mixture of two normal distributions

Additionally, accuracy of confidence interval methods applied on EBSD(n) compared specifically to the same method applied on the Monte Carlo Bootstrap. This comparison will also be performed when a larger or smaller number of Bootstraps was employed for the Monte Carlo Bootstrap.

Chapter 3: Methods

In this chapter the construction of the five most common Bootstrap confidence interval methods (T-interval with Bootstrapped Standard Error, Percentile, Basic, Bootstrap-t and BC_a) using EBSD(n) is described. A new method E-skew of constructing confidence intervals using EBSD(n) is introduced in section 3.2. The outlines for comparing Bootstrap Confidence Intervals (BCI) using the EBSD(n) (including E-skew) with Monte Carlo methods are displayed in sections 3.3, 3.4 and 3.5.

3.1 Bootstrap Sample Construction using EBSD($\boldsymbol{n}$)

As described in the introduction, an EBSD(n) design can be used to select an optimal subset of all possible Bootstrap samples. This subset of samples is optimal because the resulting Bootstrap sampling distribution for the sample mean has the same mean and variance as if from the complete enumeration of all possible samples. The following theorem below states this property for an EBSD(n) design:

Theorem 3.1.1: If Sample Average, $T(X)$, is the statistic of interest then the Bootstrap mean ($\overline{T_B}$) and variance (s_B^2) estimate of T for Bootstrap samples, generated by an EBSD(n) method, is exact[1] i.e.,

$$\overline{T_B} = E[T(\boldsymbol{X}^*)] = \bar{X} = T(\boldsymbol{X}), \text{ and}$$

$$s_B^2 = Var(T(X^*)) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n^2} = \frac{s_\infty^2}{n}$$

for $T(X^*) = \frac{\sum_{j=1}^{n} X_j^*}{n}$, where $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ denotes the Bootstrap sample of $\boldsymbol{X}$ generated by an EBSD(n) method for a random sample $X = (X_1, X_2, \dots, X_n)$ of size n and $s_\infty^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n^2}$.

Next we will discuss specifically how the EBSD(n) design is constructed. The EBSD(n) design is constructed using Balanced incomplete block design (BIBD) principles. Below are the definitions of the BIBD principles that underlie EBSD(n).

BIBD Principles Underlying EBSD($\boldsymbol{n}$)

A BIBD design with parameters (v, b, r, k, λ) is an arrangement of v symbols in b sets each of size k ($<v$) such that:

- Every symbol appears in each set at most once,
- Every symbol occurs in exactly r sets,
- Every pair of distinct symbols occurs in exactly λ sets.

The symbol and the set are called respectively the treatment and the block, in experimental settings. BIBD (v, b, r, k, λ) denotes a BIBD with parameters (v, b, r, k, λ). Further if a balanced incomplete block design exists it implies the existence of a balanced ternary block design which is used in the construction of EBSD(n).

Theorem 3.1.2: Resultantly, the existence of balanced incomplete block design $B(v', b', r', k', \lambda')$ with parameters[1]:

$v' = 2n, b' = 2(2n-1), r' = 2n - 1, k' = n$ and $\lambda' = n - 1$, for $n \geq 2$ implies the existence of balanced ternary block design $BTD(v, b, r, \rho_1, \rho_2, k, \lambda)$ with parameters

$v = n = k, b = 2(2n-1) = r,\ \rho_1 = 2n, \rho_2 = n - 1$ and $\lambda = 4(n-1)$,

where ρ_i denotes the number of blocks each element appears only i-times, i= 1, 2.

EBSD(n) leverages these principles mentioned above to yield its design. Construction of the EBSD(n) design is accomplished by following the procedure below.

Procedure to Produce EBSD($\boldsymbol{n}$)

Step 1: From the existence of a balanced incomplete block design $B(2n, 4n-2, 2n-1, n, n-1)$ suggests the existence of a balanced ternary block design[1] $BTD(n, 4n - 2, 4n - 2, 2n, n-1, 4(n-1))$.

Step 2: Generate a cyclic group of auto-morphism of order n of the $BTD(n, 4n - 2, 4n - 2, 2n, n - 1, n, 4(n - 1))$ design of Step 1 with n elements i.e., permute the elements of each block with the residual modulo n cyclically. Here the elements of each block B_i of the BTD design taken as the set $B_i\ (mod\ n)$ and mapping $i \rightarrow i + 1$ $(mod\ n)$ for each element i maps B_i into n blocks. It is easy to attest the parameters of the derived $BTD(v', b', r', \rho_1', \rho_2', k', \lambda')$ design as below[1]:

$v' = n = k', b' = r' = 2n(2n-1), \rho_1' = 2n^2, \rho_2' = n(n - 1)$ and $\lambda' = 4n(n - 1)$.

Step 3: Adjoin 2 copies of n blocks of $\{(1,1, \ldots 1), (2,2, \ldots 2), \ldots, (n, n, \ldots .)\}$ and a single block with the complete set of elements $(1, 2, \ldots, n)$ to the $BTD(\tilde{v}, \tilde{b}, \tilde{r}, \widetilde{\rho_1}, \widetilde{\rho_2}, \tilde{k}, \tilde{\lambda})$ design of Step 2.

The resulting $4n^2$+1 blocks can then in turn be thought of as Bootstrap samples. These samples can be used to construct a sampling distribution for the sample mean. As stated above the mean and variance of this sampling distribution matches the exact sample mean ($\bar{X}$) and exact variance of the mean (${\hat{\sigma}_{\bar{X}}}^2$) as if from the complete enumeration of all possible samples.

In fact, there have been many past attempts to generate a design with the property stated above. For example, the second order balanced bootstrap attempted to achieve these properties by using Latin squares. However, despite the complex derivation used, the design is only able to achieve the properties provided by EBSD(n) if size of the original sample is equal to the power of a prime number[10]. Therefore, EBSD(n) finally is able to achieve the property of exact first and second moments that many other methods had endeavored but failed to achieve. Therefore, naturally this sampling distribution is preferable over sampling distributions such as the balanced bootstrap or the Monte Carlo Bootstrap.

Further, the elimination of simulation error for the sampling distribution of the sample mean should facilitate accurate confidence interval construction. A Bootstrap sampling distribution with no simulation error yields a better estimate of the distribution of the data and should yield a better estimate of the distributions skew and thus producing more accurate confidence intervals.

Also if one is interested in constructing a confidence interval for any statistic other than the sample mean constructing a sampling distribution which matches on the first two moments for the mean in theory should provide a better Bootstrap sampling distribution for the statistic of interest. This is because using Monte Carlo resampling introduces simulation error which leads to error bias. This error bias leads to less accurate sampling distributions and consequently less accurate confidence intervals. True population parameters can be excluded from inaccurate confidence intervals more frequently. Thus sampling distributions with error bias lead to worse parameter estimation.

As the sampling distribution is integral to accurate confidence interval construction, below is a definition for the Bootstrap sampling distribution constructed from EBSD(n).

Bootstrap Sampling Distribution Constructed from EBSD($\boldsymbol{n}$)

An EBSD(n) design can be used to construct a Bootstrap sampling distribution with ($4n^2$+1) elements. Similar to the concept of the Monte Carlo Bootstrap sampling distribution in Chapter 2, the Bootstrap sampling distribution constructed from EBSD(n) is generated by calculating the statistic of interest for each sample from EBSD(n). This collection of statistics $\hat{\theta}^*_{E_1}, \ldots, \hat{\theta}^*_{E_{4n^2+1}}$ is termed the Bootstrap sampling distribution constructed from EBSD(n) and can be used for confidence interval construction.

To compare the Bootstrap sampling distribution constructed from EBSD(n) to the Monte Carlo Bootstrap sampling distribution we compare the standard error generated from each using the data from Table 4 below. As stated above the variance of the sample mean sampling distribution is the same as if from the complete enumeration of all possible samples. If the exact variance of the sample mean is achieved then the exact standard error of the mean is also achieved as the standard error is simply a one-to-one function of ${\hat{\sigma}_{\bar{X}}}^2$ ($Standard\ Error_{\bar{X}}$= $\sqrt{{\hat{\sigma}_{\bar{X}}}^2/n}$ where n is the size of the original sample).

Thus below we refer back to the example mentioned in section 2.4 and also presented below. It is seen that from the results published by Efron also in Table 5 below, that trials using 50, 100, 250, 500 and 1,000 Monte Carlo Bootstrap samples all resulted in standard errors that failed to match the exact standard error. According to Efron, only if approximately an infinite number of Monte Carlo Bootstraps resamples were performed may the resulting standard error be exact. The Bootstrap sampling distribution constructed using EBSD(n) yields a standard error for the mean of 23.36 using the data Efron used (Table 4) and this is accomplished using only 197 samples. Therefore the advantage in the sampling distribution generated from EBSD(n) is that it identifies a small subset of samples that guarantee exact first two moments for the sample mean, while conversely, the Monte Carlo Bootstrap cannot guarantee this even when generating many more resamples to construct it's sampling distribution.

Table 3.1 Survival Time (Days) Treatment group Mouse Data

X_1	X_2	X_3	X_4	X_5	X_6	X_7
94	197	16	38	99	141	23

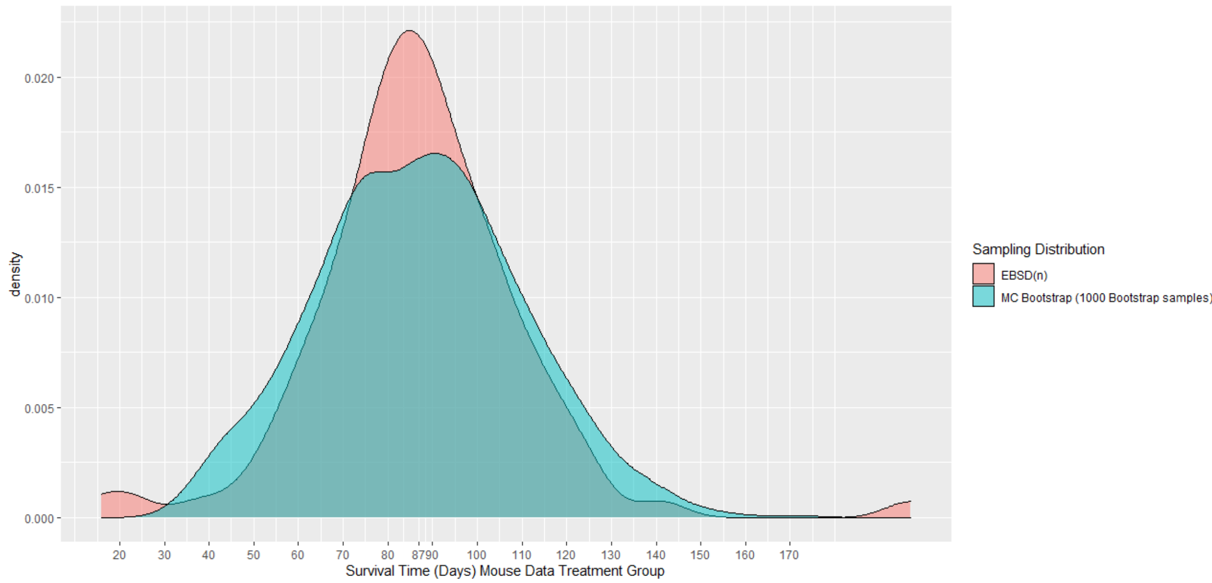
Table 3.2 Efron and Tibshirani Estimated Standard Error for the Sample Mean: Mouse Data: Treatment Group

B^*	50	100	250	500	1000	∞
Standard Error of the Sample Mean	19.72	23.63	22.32	23.79	23.02	23.36

*B = # of Bootstrap samples.

Below we see a comparison of the sampling distributions in Figure 1 below. In the figure it is shown that the Bootstrap sampling distribution constructed from EBSD(n) is centered at the mean value of the Mouse data from the original sample. Conversely, for the Monte Carlo Bootstrap simulation error causes the sampling distribution to be centered above the mean value from the original sample.

Figure 3.1. Comparison of Sampling Distributions for Mouse Data



In the Efron example above we see why it would be impractical to use the Monte Carlo Bootstrap to generate a sampling distribution for the mean that are exact for the first two moments. One may then argue why not just enumerate out all possible samples in order to generate a design with exact moments? The issue with this approach is that complete enumeration of all possible samples is unfeasible even for relatively small sample sizes. The number of samples required when performing complete enumeration is n^n and the number of samples required for complete unique enumeration is $\binom{2n-1}{n}$ where n is the size of the original sample. Therefore, the design size exponentially increases as the original sample size increases. In this way, we see the relative benefit of EBSD(n) because the required number of samples for EBSD(n) ($4n^2$+1) is much smaller as sample size increases. The size of the complete enumeration of all possible samples and the complete enumeration of all possible unique samples is compared to the EBSD(n) design size in the table below.

Table 3.3 Comparison of Relative Design Sizes			
Sample Size (n)	**Number of Samples When Enumerating All Possible Samples**	**Number of Samples When Enumerating All Possible Unique Samples**	**Number of EBSD(n) ($4n^2$+1) Required Samples**
3	27	10	37
4	256	35	65
5	3,125	126	101
6	46,656	462	145
7	823,543	1716	197
8	16,777,216	6435	257
9	387,420,489	24310	325
10	10 $*10^{10}$	92378	401
11	28.5 $*10^{11}$	352716	485
13	3.02 $*10^{15}$	5,200,300	677
15	4.4 $*10^{18}$	77,558,760	901
19	1.97 $*10^{24}$	1.7 e11	1445
23	2.08 $*10^{32}$	4.1 $*10^{13}$	2117
25	8.88 $*10^{34}$	6.3 $*10^{14}$	2501
30	2.06 $*10^{44}$	5.9 $*10^{16}$	3601
40	1.21 $*10^{64}$	5.4 $*10^{22}$	6401
70	1.44 $*10^{129}$	4.7 $*10^{40}$	19601
100	1 $*10^{200}$	4.5 $*10^{58}$	40001
∞	∞	∞	∞

The complete enumeration of all possible samples can be done up to sample size 15 using high performance computing. For sample sizes greater than 15 limitations in computational capability make complete enumeration impossible to achieve. Thus EBSD(n) is needed for the construction of a design that yields a sampling distribution with exact first and second moments for the sample mean for sample sizes larger than 15. The design sizes required in the table above highlight how impractical it can be to perform complete enumeration as sample size increases.

3.2 Confidence Intervals using EBSD(n)

As stated in section 3.1 confidence intervals can be computed using the Bootstrap sampling distribution constructed from EBSD(n). A number of different algorithmic approaches are possible. Below is a proposed method for a new confidence interval method called E-skew. In this section specific techniques employed for each statistic studied in the simulation are detailed. After proposal of the E-skew confidence interval, algorithmic approaches commonly used with the Monte Carlo Bootstrap are proposed using EBSD(n).

E-skew Method for Constructing a Confidence Interval for a Statistic $\hat{\boldsymbol{\theta}}$

For the case of constructing a confidence interval for a statistic $\hat{\theta}$ the following method adjusting for skew in the Bootstrap sampling distribution constructed from EBSD(n) is proposed:

Step 1: A modification of the t-skew method mentioned in section 2.5 is used to construct a confidence interval lower limit and upper limit for each sample from EBSD(n). Consequently, there are two sampling distributions, a lower limit sampling distribution and an upper limit sampling distribution. For a statistic $\hat{\theta}$ the confidence interval formula for each sample is:

$$\hat{\theta}^*_{E_i} + \frac{\hat{s}_{\hat{\theta}^*_{E_i}}}{\sqrt{n}} * \left(\left(\frac{\widehat{skew}_{\hat{\theta}^*_{E_i}}}{6\sqrt{n}}\right) * (1 + 2 * t_{\frac{\alpha}{2},n-1}{}^2) \pm t_{\frac{\alpha}{2},n-1}\right)$$

- $\hat{\theta}^*_{E_i}$ is defined as the sample statistic for the ith sample from EBSD(n)
- The Jackknife method is used to estimate the sample variance for each EBSD(n) sample and is computed as:

$$\hat{s}^2_{\hat{\theta}^*_{E_i}} = \frac{1}{n-1}\sum_{j=1}^{n}(\hat{\theta}_{E_i(-j)} - \hat{\theta}_{E_i(.)})^2$$

- n is the size of the original sample
- E_i is the ith sample from EBSD(n)
- E_{ij} is the ith sample and jth element from EBSD(n)
- Lastly the skew for each sample is estimated as:

$$\widehat{skew}_{\hat{\theta}^*_{E_i}} = \frac{\left(\frac{1}{n}\right)*\sum_{j=1}^{n}(\hat{\theta}_{E_i(-j)} - \hat{\theta}_{E_i(.)})^3}{(\hat{s}_{\hat{\theta}^*_{E_i}})^3}$$

In the case where $\widehat{skew}_{\hat{\theta}^*_{E_i}}$ is undefined as would be the case when s_{E_i} =0 (this occurs for the samples with the repeated elements), $\widehat{skew}_{\hat{\theta}^*_{E_i}}$ is assigned the value of 0.

These variance and skew estimates from the Jackknife are used in conjunction with the t-skew adjustment to generate an EBSD(n) sampling distribution for both the upper and lower limit for any statistic of interest.

Step 2: The t-skew adjusted upper and lower limit is calculated for the statistic of interest ($\hat{\theta}$) from the original sample.

Step 2A: For the upper limit $\hat{\theta}_U$ is the t-skew adjusted upper limit from the original sample is computed as:

$$\hat{\theta}_U = \hat{\theta} + \frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}} * (\left(\frac{\widehat{skew}_{\hat{\theta}}}{6\sqrt{n}}\right) * \left(1 + 2 * t_{\frac{\alpha}{2},n-1}{}^2\right) - t_{\frac{\alpha}{2},n-1})$$

The $\widehat{skew}_{\hat{\theta}}$ and $\hat{s}_{\hat{\theta}}$ for $\hat{\theta}_U$ are estimated using the Jackknife just like they are for each sample from EBSD(n):

where $\widehat{skew}_{\hat{\theta}} = \frac{\left(\frac{1}{n}\right) * \sum_{j=1}^{n} (\hat{\theta}_{-j} - \hat{\theta}_{(.)})^3}{(\hat{s}_{\hat{\theta}})^3}$ where $\hat{s}_{\hat{\theta}}$ is the sample standard deviation of the j Jackknifed samples from the original sample: $\hat{s}_{\hat{\theta}} = \left(\frac{1}{n-1}\right) * \sum_{j=1}^{n} (\hat{\theta}_{-j} - \hat{\theta}_{(.)})^2$ and where $\hat{\theta}_{(.)} = \frac{1}{n} \sum_{j=1}^{n} \hat{\theta}_{(-j)}$ and $\hat{\theta}_{-j}$ is the statistic calculated with the jth element removed from the original sample.

Step 2B: Similarly, for the lower limit $\hat{\theta}_L$ is the t-skew adjusted lower limit from the original sample and is computed as:

$$\hat{\theta}_L = \hat{\theta} + \frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}} * \left(\left(\frac{\widehat{skew}_{\hat{\theta}}}{6\sqrt{n}}\right) * \left(1 + 2 * t_{\frac{\alpha}{2}, n-1}{}^2\right) + t_{\frac{\alpha}{2}, n-1}\right)$$

where $\widehat{skew}_{\hat{\theta}}$ and $\hat{s}_{\hat{\theta}}$ are the same estimates as what was used for the upper limit.

Step 3: The median bias ($\widehat{z_0}_U$ and $\widehat{z_0}_L$) are calculated for both the upper and lower limit as is done with the BC_a method.

Step 3A: $\hat{\theta}_U$ is compared to the t-skew adjusted upper limit computed from each EBSD(n) sample E_i and $\widehat{z_0}_U$ is computed as below:

$$\widehat{z_0}_U : \phi^{-1}\left(\left(\sum_{i=1}^{4n^2+1} \hat{\theta}_{U_{E_i}} < \hat{\theta}_U\right) / (4n^2 + 1)\right)$$

- $\hat{\theta}_{U_{E_i}}$ is defined as the t-skew adjusted upper limit of sample E_i from EBSD(n)

Step 3B: $\hat{\theta}_L$ is compared to the t-skew adjusted lower limit computed from each EBSD(n) sample E_i and $\widehat{z_0}_L$ is computed as below:

$$\widehat{z_0}_L : \phi^{-1}\left(\left(\sum_{i=1}^{4n^2+1} \hat{\theta}_{L_{E_i}} > \hat{\theta}_L\right) / (4n^2 + 1)\right)$$

Where the inequality sign is flipped when computing for the lower limit.

- $\hat{\theta}_{L_{E_i}}$ is defined as the t-skew adjusted lower limit of sample E_i from EBSD(n)

Step 4: The quantile to be taken from each sampling distribution (upper limit and lower limit) is calculated. The modified BC_a adjustment determines which limit from the sampling distribution is chosen for either side of the interval.

Step 4A: The alpha percentile to be chosen from the upper limit sampling distribution is calculated as:

$$\alpha_U = \phi\left(\frac{\widehat{z_0}_U}{\left(1 - \hat{a}_U * \widehat{z_0}_U\right)}\right)$$

where $\hat{a}_{\boldsymbol{U}}$ is estimated using the Jackknife approach like what is done with the BC_a method:

$$\hat{a}_{\boldsymbol{U}} = \frac{\sum_{j=1}^{n}(\hat{\theta}_{U_{(.)}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{U}_{-j}})^3}{6\sum_{j=1}^{n}((\hat{\theta}_{U_{(.)}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{U}_{-j}})^2)^{3/2}},$$

$\hat{\theta}_{U_{-j}}$ is the t-skew adjusted upper limit for the jackknifed sample with the jth element removed from the original sample.

$\hat{\theta}_{U_{(.)}}$ is the average t-skew adjusted upper limit for the j Jackknifed samples:

$$\hat{\theta}_{U_{(.)}} = \frac{1}{n}\sum_{j=1}^{n}\hat{\theta}_{U_{(-j)}} .$$

Step 4B: The alpha percentile to be chosen from the lower limit sampling distribution is calculated where the opposite sign is imposed on the denominator of the formula for α_L:

$$\alpha_L = \phi\left(\frac{\widehat{z_0}_L}{\left(\hat{a}_L * \widehat{z_0}_L - 1\right)}\right)$$

where $\hat{a}_L$ is estimated using the Jackknife approach like $\hat{a}_U$:

$$\hat{a}_L = \frac{\sum_{j=1}^{n}(\hat{\theta}_{L-j} - \hat{\theta}_{L(.)})^3}{6\sum_{j=1}^{n}((\hat{\theta}_{L-j} - \hat{\theta}_{L_{(.)}})^2)^{3/2}}$$

$\hat{\theta}_{L-j}$ is the t-skew adjusted lower limit for the jackknifed sample with the jth element removed from the original sample

$\hat{\theta}_{L_{(.)}}$ is the average t-skew adjusted lower limit for the j Jackknifed samples:

$$\hat{\theta}_{L_{(.)}} = \frac{1}{n}\sum_{j=1}^{n}\hat{\theta}_{L_{(-j)}} \; .$$

Each of $\hat{a}_L$ and $\hat{a}_U$ are calculated using a Jackknife approach, the same approach as what is used for the BC_a method. Using the Jackknife, n upper limits and n lower limits are calculated. Each limit is calculated from a sample of size $n - 1$, where for each limit a different unique element has been removed prior to calculation.

Step 5: The α_L and α_U quantiles are applied to each EBSD(n) sample respectively and confidence interval limits are generated. The quantiles α_L and α_U are picked from limit's sampling distribution. No matter which limit is picked for each end of the interval a skew adjusted limit will be selected from a random resample of the original sample.

a. Illustration of E-skew Confidence Interval Construction for the Sample Mean $\bar{X}$

For the case of constructing a confidence interval for the sample mean statistic $\bar{X}$, an approach that is nearly the same as the general case $\hat{\theta}$ is proposed. The difference with the mean is how the sample skew and sample variance of the mean are calculated for each EBSD(n) sample. For the sample mean both statistics do not need to be estimated using the Jackknife for each EBSD(n) sample. Therefore, the skew adjusted t-statistic confidence interval calculated on each sample is the same as what is used for the t-skew method mentioned in section 2.5. The confidence interval formula for each sample is:

$$\bar{X}^*_{E_i} + \frac{\hat{s}^*{}_{\bar{X}_{E_i}}}{\sqrt{n}} * \left(\left(\frac{\widehat{skew}_{\bar{X}^*_{E_i}}}{6\sqrt{n}}\right) * (1 + 2 * t_{\frac{\alpha}{2},n-1}{}^2) \pm t_{\frac{\alpha}{2},n-1}\right)$$

- $\widehat{skew}_{\bar{X}^*_{E_i}} = \frac{\left(\frac{1}{n}\right) * \sum_{j=1}^{n}(X_{E_{ij}} - \bar{X}^*_{E_i})^3}{\hat{s}^*{}_{\bar{X}_{E_i}}{}^3}$
- $\bar{X}^*_{E_i}$ is defined as the sample mean for a given sample from EBSD(n).
- $\hat{s}^*{}_{\bar{X}_{E_i}}$ is the estimated sample standard deviation for each sample:

$$\sqrt{\left(\frac{1}{n-1}\right) * \sum_{j=1}^{n}(X_{E_{ij}} - \bar{X}^*_{E_i})^2}$$

- n is the size of the original sample
- E_i is the ith sample from EBSD(n)
- $X_{E_{ij}}$ is the ith sample and jth element from EBSD(n)

In the case where $\widehat{skew}_{\bar{X}^*_{E_i}}$ is undefined as would be the case when s_{E_i} =0 (this occurs for the samples with the repeated elements), $\widehat{skew}_{\bar{X}^*_{E_i}}$ is assigned the value of 0.

Once the sampling distribution for the upper and lower limits is generated the remaining algorithm steps are the same as they are for a statistic $\hat{\theta}$. Correspondingly not only does $\widehat{skew}_{\bar{X}^*_{E_i}}$ and $\hat{s}^*{}_{\bar{X}_{E_i}}$ not need to be estimated using the Jackknife, $\widehat{skew}_{\bar{X}}$ and $\hat{s}_{\bar{X}}$ for calculating the t-skew adjusted limits from the original sample also can be estimated directly as the same concept applies.

b. Confidence Interval for the Ratio of Means

For the case of constructing a confidence interval for the ratio of means statistic $\hat{r}$ the following method adjusting for skew in the Bootstrap sampling distribution constructed from EBSD(n) is proposed:

Step 1: The approach for computing a confidence interval for $\hat{r}$ is the same as the approach for $\hat{\theta}$, except E-skew is computed for the natural log of $\hat{r}$ and then the resulting limits are back transformed to generate the confidence interval for $\hat{r}$ where $\hat{r}$ is defined as $\frac{\bar{X}}{\bar{Y}}$, where $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$.

$$\ln(\hat{r})_{E_i}^{*} + \frac{\hat{s}_{\ln(\hat{r})_{E_i}^{*}}}{\sqrt{n}} * (\left(\frac{\widehat{skew}_{\ln(\hat{r})_{E_i}^{*}}}{6\sqrt{n}}\right) * (1 + 2 * t_{\frac{\alpha}{2},n-1}{}^{2}) \pm t_{\frac{\alpha}{2},n-1})$$

- $\ln(\hat{r})_{E_i}^{*}$ is defined as the natural log of the ratio statistic for the ith sample from EBSD(n)
- The Jackknife method is used to estimate the sample variance for each EBSD(n) sample and is computed as:

  $\hat{s}_{\ln(\hat{r})_{E_i}^{*}}^{2} = \frac{1}{n-1} \sum_{j=1}^{n} (\ln(\hat{r})_{E_i(-\mathrm{j})} - \ln(\hat{r})_{E_i(.)})^2$
- n is the size of the original sample
- E_i is the ith sample from EBSD(n)
- E_{ij} is the ith sample and jth element from EBSD(n)
- Lastly the skew for each sample is estimated as:

  $\widehat{skew}_{\ln(\hat{r})_{E_i}^{*}} = \frac{\left(\frac{1}{n}\right) * \sum_{j=1}^{n} (\ln(\hat{r})_{E_i(-\mathrm{j})} - \ln(\hat{r})_{E_i(.)})^3}{(\hat{s}_{\ln(\hat{r})_{E_i}^{*}})^3}$

In the case where $\widehat{skew}_{\ln(\hat{r})^*_{E_i}}$ is undefined as would be the case when s_{E_i} =0 (this occurs for the samples with the repeated elements), $\widehat{skew}_{\ln(\hat{r})^*_{E_i}}$ is assigned the value of 0.

Once the sampling distribution for the upper and lower limits is generated the remaining the steps are the same as they are for a statistic $\hat{\theta}$. After the remaining E-skew steps are completed for $\ln(\hat{r})$ the resulting confidence interval for $\ln(\hat{r})$ is back transformed in order to yield a confidence interval for $\hat{r}$.

The natural log transformation can yield undefined estimates if an element from the sampling distribution is less than 0 prior to transformation. In practice negative values are not typically collected from nature and this is not a problem. However, for example from the normal distribution a negative element could be randomly generated if the center of the normal distribution where values are being simulated from is near 0. Enough negative elements in an individual sample will result in a sampling distribution with a negative element. If a negative element is generated for $\hat{r}$ prior to using the natural log transformation the resulting sampling distribution of $\ln(\hat{r})$ will have an undefined element. When measuring the performance of E-skew for the ratio of means statistic, distributional parameters were set to be centered sufficiently far above 0 for such distributional types. Therefore, the chance a negative element would contribute to the sampling distribution should not occur.

In the very rare cases the estimate for $\frac{\widehat{z_0}_L}{(\hat{a}_L * \widehat{z_0}_L - 1)}$ and $\frac{\widehat{z_0}_U}{(1 - \hat{a}_U * \widehat{z_0}_U)}$ would diverge to infinity. In these cases, the element taken from the given confidence interval limit sampling distribution was the maximum value from that sampling distribution.

c. Confidence Interval for the Pearson Correlation Coefficient

For the case of constructing a confidence interval for the Pearson correlation coefficient $\hat{\rho}$ the following method adjusting for skew in the Bootstrap sampling distribution constructed from EBSD(n) is proposed:

Step 1: A similar approach to computing the confidence interval for $\hat{\theta}$ is used for $\hat{\rho}$, except E-skew is computed for the Fisher-z transformation (Fisher 1915) of $\hat{\rho}$ and then the resulting limits are back transformed to generate the confidence interval for $\hat{\rho}$ where $\hat{\rho}$ is defined as $\hat{\rho} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$, where $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$, and the Fisher-z transformation is defined as $\frac{1}{2} log\{\frac{1 + \hat{\rho}}{1 - \hat{\rho}}\}$. Because the Fisher-z transformed Pearson correlation coefficient uses standard normal quantile $z_{1-(\frac{\alpha}{2})}$ in confidence interval construction no skew adjustment for the t-statistic is computed for each sample of EBSD(n).

Therefore the confidence interval computed for each sample E_i is:

$$\frac{1}{2} log\{\frac{1 + \hat{\rho}_{E_i}{}^*}{1 - \hat{\rho}_{E_i}{}^*}\} \pm z_{1-(\frac{\alpha}{2})}\sqrt{\frac{1}{n-3}}$$

- $\hat{\rho}_{E_i}$ is defined as the Fisher-z transformed Pearson correlation coefficient for sample i from EBSD(n)
- n is the size of the original sample
- E_i is the ith sample from EBSD(n)
- E_{ij} is the ith sample and jth element from EBSD(n)

Furthermore, $\widehat{skew}_{\hat{\rho}^*_{E_i}}$ and $\hat{s}^2_{\hat{\rho}^*_{E_i}}$ did not need to be estimated for each sample of EBSD(n) because neither of these statistics contribute to the Fisher-z transformed confidence interval. Although individual skew corrections are not applied on each sample the method below still calculates the upper and lower limit for each sample of the EBSD(n) method using the $\alpha_U = \phi\,(\frac{\widehat{z_0}_U}{(1-\hat{a}_U*\widehat{z_0}_U)})$ percentile of the upper and $\alpha_L =$ $\phi\,(\frac{\widehat{z_0}_L}{(\hat{a}_L*\widehat{z_0}_L-1)})$ percentile of the lower limits. These corresponding α_u and α_L percentiles are chosen from the EBSD(n) sampling distribution as was done for the mean and ratio of means statistics. The z_{0_U}, z_{0_L}, $\hat{a}_U$, $\hat{a}_L$ would be computed the same as how was described previously for E-skew.

After the remaining E-skew steps are completed for the Fisher-z transformation of $\hat{\rho}$ the resulting confidence interval is back transformed in order to yield a confidence interval for $\hat{\rho}$.

Again in the rare cases like with $\ln(\hat{r})$ where $\frac{\widehat{z_0}_L}{(\hat{a}_L*\widehat{z_0}_L-1)}$ and $\frac{\widehat{z_0}_U}{(1-\hat{a}_U*\widehat{z_0}_U)}$ would diverge to infinity, the element taken from the given confidence interval limit sampling distribution was set to the maximum value from that sampling distribution.

Computation of Pearson Correlation Coefficient Confidence Interval

A further modification is implemented for the purposes of yielding more accurate bounds for the correlation parameter when using the EBSD(n) sampling distribution for this simulation work. In generating the EBSD(n) $2n$ paired samples must be generated in which a single element from the original sample is repeated n times. In the case of the correlation parameter the repeated paired samples take the form of repeated pairs n times.

Since for **X** = {(($U_{\mathbf{1}}, V_{\mathbf{1}}$) , …., ($U_{\boldsymbol{n}}, V_{\boldsymbol{n}}$))}, in the case of each repeated pair $\mathbf{X}^* = (U_1^*, \ldots,$ U_n^* **)** is the same element repeated n times the variance of ($U_1^*, \ldots, U_n^*$) (this also goes for V^*) is 0 and the corresponding estimate of the correlation for this pair of samples is undefined. Rather than throwing away these important paired samples that allow us to achieve the second order balance the correlation value is defined to be -1 or 1 depending on the elements in question. If the repeated element from U is less than the repeated element from V the correlation is assigned the value of -1, otherwise if the element from V is greater than or equal to U the correlation is assigned the value of 1.

d. Confidence Interval for the Trimmed Mean

The trimmed mean statistic is used as a location statistic in the one sample case when estimating the mean or the median is suboptimal. Typically, when outliers exist in the original sample and are believed to be compromising the estimate produced by the mean; the trimmed mean can be used as an alternative. The trimmed mean removes a designated percentage from either end of the sample.

The k-times trimmed mean is represented as:

$$\bar{X}_{tk} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X_i$$

The trimmed mean is a compromise between the mean and median in that it is not as subject to misrepresenting the population mean in the case of extreme outliers and yet can still represent the existence of skewness in the resulting confidence interval if skewness can be observed in the remaining 80-90% of the data. Once samples were trimmed, mean confidence interval computation was implemented on the remaining data as is described for the proposed E-skew confidence interval for the sample mean. For this work, each end of the ordered sample was trimmed by 10% prior to confidence interval computation.

In cases where estimates $\hat{a}_L$, $\hat{a}_U$ and $\widehat{skew}_{\tilde{X}_i}$ for the ith Jackknifed sample of the original sample diverged these estimates were set to 0 in the computation of the E-skew method.

Monte Carlo Bootstrap Confidence Interval Methods applied on EBSD($\boldsymbol{n}$)

The t-interval with bootstrap estimated standard error, percentile, basic and BC_a methods were applied on the Bootstrap sampling distribution constructed from EBSD(n). In each case these algorithms were identical to how each method is applied on the Monte Carlo Bootstrap Sampling distribution. In cases where a transformation was used for E-skew as detailed above (independent ratio of sample means statistic and Pearson correlation coefficient), the same transformation was used for each of these methods applied on EBSD(n). The Bootstrap-t method was also applied on EBSD(n) but required a modification detailed below.

Bootstrap-t Method with EBSD($\boldsymbol{n}$) (ES)

The Bootstrap-t method applied on EBSD(n) is computed similar to how it is computed for the Monte Carlo Bootstrap with a slight modification. For each sample of EBSD(n) a t-statistic is calculated. This collection of t-statistics is the Bootstrap sampling distribution constructed from EBSD(n) and is used in deriving the confidence interval. The t-statistic for each resample is calculated as:

$$t^*{}_{E_i} = \frac{\hat{\theta}^*_{E_i} - \hat{\theta}}{\hat{s}_{\widehat{\theta^*}_{E_i}}}$$

$\hat{\theta}^*_{E_i}$ is the statistic of interest calculated for each resample, $\hat{\theta}$ is the statistic from the original sample, and $\hat{s}_{\widehat{\theta^*}_{E_i}}$ is the estimate of the standard error of the statistic for each resample. Then for the computation of the confidence interval the (1- $\frac{\alpha}{2}$) and $\frac{\alpha}{2}$ percentiles of the EBSD(n) sampling distribution are taken to determine the t-statistics used for the confidence interval computation. The confidence interval takes the form:

$$(\hat{\theta} - t^*_{\left(1-\frac{\alpha}{2}\right)} \frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}},\ \hat{\theta} - t^*_{\frac{\alpha}{2}} \frac{\hat{s}_{\hat{\theta}}}{\sqrt{n}})$$

The problem with this approach is what to do with the t-statistics for the samples of repeated elements. One approach is to assign the t-statistic the value of 0 similar to what is done for $\widehat{skew}_{\bar{X}^*_{E_i}}$ in E-skew. The sample of repeated elements end up concentrated near the 50th percentile of the resulting EBSD(n) sampling distribution and information is lost as they are not distributed across the sampling distribution. For this simulation study the approach of assigning the t-statistic the value 0 was taken in order to generate error rate results.

3.3 Workflow

In the simulation study done for this dissertation detailed below, one-sided error rates were computed for both the upper and lower limit for the E-skew, ET, EP, EBC, EBC_a, and ES methods as well as for Monte Carlo Bootstrap methods: BT, BP, BC, BC_a, BS, ABC. Each of these methods were studied for multiple statistics, distributional types, sample sizes and confidence levels. For Monte Carlo Bootstrap methods, the number of Bootstrap samples specified ranged from 200, 500, 1,000, 5,000, or 10,000 depending on the simulation in question. Each simulation was repeated 10,000 times for each sample size, statistic and distributional type tested. The sample sizes used were n = 5, 10, 15, 20, 30 and 40. Specifically, the methods tested using the Bootstrap constructed from EBSD(n) and using the Monte Carlo Bootstrap were:

Table 3.4 Confidence Interval Method's Used in the Simulation Study

Bootstrap constructed from EBSD(n) Confidence Interval method	Monte Carlo Bootstrap Confidence Interval method
E-skew	
T-interval with EBSD(n) standard error (ET)	T-interval with Bootstrap standard error method (BT)
Percentile method using EBSD(n) (EP)	Percentile method (BP)
Basic/reverse method using EBSD(n) (EBC)	Basic/Reverse method (BC)
Bias-Corrected accelerated method using EBSD(n) (EBC_a)	Bias-Corrected accelerated method (BC_a)/ABC method
Bootstrap-t method using EBSD(n) (ES)	Bootstrap-t method (BS)

Below table 8 describes the parameters specified for each distributional type. In total the table is split into 8 different distributional types by statistic. The distributions are: The Normal distribution, the Exponential distribution, the Gamma distribution, the Log-Normal distribution, the Mixture of Two Normal distributions, the Bivariate Normal distribution, the Bivariate Non-Normal distribution, and the Cauchy distribution.

Table 3.5 Parameter Specifications for each Statistic for each distribution of interest

	Parameters specified
Normal Distribution	
Sample Mean	
(μ, σ)	(4, 1), (4, 4), (4, 8)
Sample Median	
(μ, σ)	(4, 1)
Trimmed Mean	
(μ, σ)	(4, 1)
Ratio of Sample Means	

$(\mu_1, \sigma_1, \mu_2, \sigma_2)$	(100, 1, 50, 1), (50, 1, 100, 1), (100, 1, 100, 1)
Exponential Distribution	
Sample Mean	
λ	0.10, 0.01, 1
Sample Median	
λ	0.10
Trimmed Mean	
λ	0.10
Ratio of Sample Means	
(λ_1, λ_2)	(0.10, 0.20), (0.20, 0.05), (0.10, 0.10)
Gamma Distribution	
Sample Mean	
(α, λ)	(2, 2), (2, 3)
Trimmed Mean	
(α, λ)	(2, 2)
Ratio of Sample Means	
$(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$	(4, 1, 3, 1)
Log-Normal Distribution	
Sample Mean	
(μ, σ)	(4, 0.2), (4, 2), (4, 3)
Trimmed Mean	
(μ, σ)	(4, 0.2)
Ratio of Sample Means	
$(\mu_1, \sigma_1, \mu_2, \sigma_2)$	(4, 0.2, 3.3, 0.2)
Mixture of Two Normal Distributions	
Sample Mean	
$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	(4, 4, 0.5, 8, 8, 0.5), (4, 4, 0.6, 8, 8, 0.4), (4, 4, 0.8, 8, 8, 0.2)
Trimmed Mean	
$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	(4, 4, 0.6, 8, 8, 0.4)
Median	
$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	(4, 1, 0.5, 8, 1, 0.5)
Ratio of Sample Means	
$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2, \mu_3, \sigma_3, p_3, \mu_4, \sigma_4, p_4)$	(50, 1, 0.6, 100, 1, 0.4, 25, 1, 0.6, 50, 1, 0.4)
Bivariate Normal Distribution	
Pearson Correlation Coefficient	
$(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$	(4, 1, 4, 1, 0.1), (4, 1, 4, 1, 0.5), (4, 1, 4, 1, 0.9)
Bivariate Non-Normal Distributions	
Pearson Correlation Coefficient	
(Skew, Kurtosis, ρ)	(3, 61, 0.1), (3, 61, 0.5)

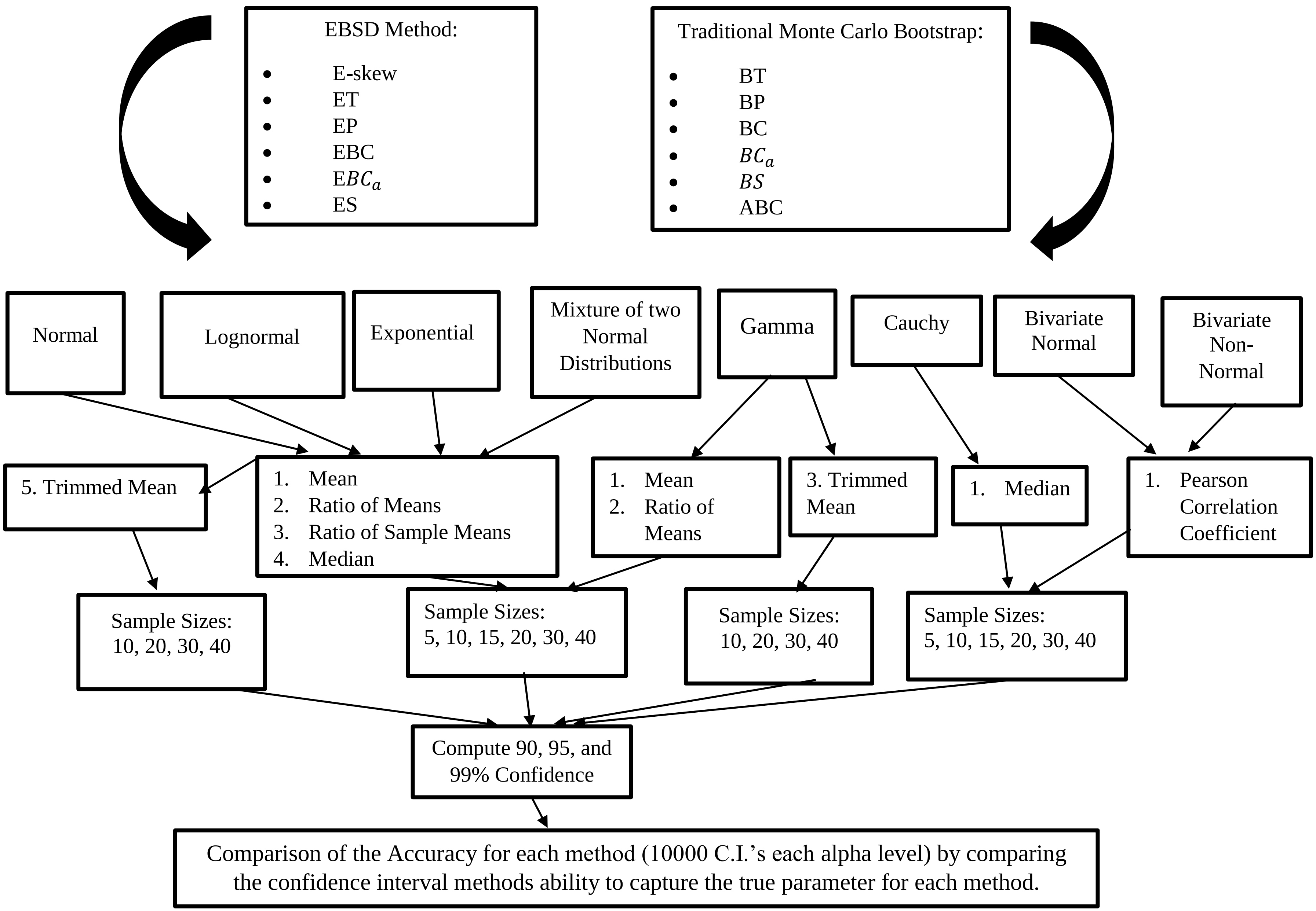
Cauchy Distribution	
Median	
(x_0, γ)	(0, 1)

In total the E-skew method and the other methods mentioned in section 3.2 using EBSD(n) were compared against the six Monte Carlo Bootstrap methods. For each method when a transformation was applied to the E-skew method it was also applied to every other method studied; both to methods applied on EBSD(n) and on the Monte Carlo Bootstrap. For the median statistic EBSD(n) methods and Monte Carlo Bootstrap methods were compared but the E-skew method was not a part of the comparison as skew adjustment was not deemed necessary for the median statistic.

3.4 Simulation Study Procedure

First Figure 2 below displays the simulation method used in this dissertation graphically via a flow chart. Following the figure, the simulation procedure is described step by step in order to give the reader an overview of the simulation process.

Figure 3.2 Flow Chart of Simulation Study



Step 1: Selection of the Sample Size Configuration for the Sample

Six different sample size configurations were used, “small” sample sizes n = 5 and n = 10, “Moderately small” sample sizes n = 15 and n = 20 and “Marginally small” sample sizes n = 30, and n = 40.

Step 2: Selection of the Probability Distribution

Eight probability distribution types: Exponential, Normal, Cauchy, Gamma, Log-Normal, Mixture of two Normal distributions, Bivariate Normal, and Bivariate Non-Normal. The parameters specified for each of these distributions is detailed in Table 8 above.

Step 3: Specification of Types of Confidence Intervals on the Same Data set

90, 95 and 99% confidence intervals were generated. Thus 3 confidence intervals were generated for each probability distribution, sample size, and statistic combination for each method.

Step 4: Confidence Intervals for True Parameter Capture Across Significance Level

The ability to capture the true population parameter was compared between Bootstrap confidence intervals constructed from EBSD(n) and Monte Carlo Bootstrap confidence intervals. For Monte Carlo Bootstrap confidence intervals 200, 500, 1,000 and 10,000 samples were used for the mean, ratio of means, and Pearson correlation coefficient. For the trimmed mean statistic Monte Carlo Bootstrap samples of 500 and 10,000 were used. For the median statistic 500, 1,000 and 5,000 Monte Carlo Bootstrap samples were used. The upper and lower limit of the 90, 95, and 99% confidence interval for each statistic was compared to the true population parameter. If the upper limit was greater or lower limit less than the true population parameter one-sided coverage was achieved for that specific end of the confidence interval. If the upper limit was less than or the lower limit greater than the true population parameter one-sided coverage was not achieved for that end of the confidence interval.

This process was repeated 10,000 times in order to compute an error rate. This error rate was then compared to the true nominal error rate in order to compute a percent error. The smaller the percent error the more accurate the method. Graphics analysis (one-sided error rate plots) were also used to compare each of the test methods. Next is a section describing how error rates were compared to the true nominal error rate using percent error.

3.5 Criteria for Confidence Interval Method Comparison

Accuracy was measured by computing the error rate for the upper limit and lower limit separately and then computing each limit's percent error against the true nominal error rate. The percent error was defined as:

$$\left|100 * \frac{(\hat{\varepsilon}_s - \varepsilon_n)}{\varepsilon_n}\right|$$

$\hat{\varepsilon}_s$ = error rate computed from the 10,000 generated samples.

ε_n = theoretical true nominal error rate based on the alpha level specified.

As shown above the absolute value of the percent error was taken and the percent error reported was always be greater than or equal to zero. Therefore, if two error rates, one of which was smaller and the other larger, were equidistant from the nominal error rate they would correspondingly have the same reported percent error.

These percent errors were computed separately for the upper and lower limit, the true nominal error rate for an $\alpha = 0.05$ confidence interval for each end of the interval would be 0.025.

Justification for Reporting Percent Error Separately for Each Limit End

An accurate 95% confidence interval upper limit is less than the true population parameter 2.5% of the time while it's lower limit is greater than the true population parameter 2.5% of the time. As a counter example, consider a 95% confidence interval that misses 5% for the upper limit and 0% for the lower limit. Although the total error rate is the same as the total theoretical nominal error rate this example implies both limits would be smaller than what they should be. This theoretical interval would then present a clearly biased picture of the true population parameter; we may think the true population parameter is smaller than it is.

Chapter 4: Main Results

This chapter is divided into two parts. First results from the simulation study will be reported and discussed. Results from the simulation study are reported in sections 4.1-4.6. In sections 4.1-4.5, error rate results for each statistic are reported for the sample mean, the ratio of sample means, the Pearson correlation coefficient, the trimmed mean and the median. In section 4.6 error rate results are compared between EBSD(n) and Monte Carlo Bootstrap methods, where for Monte Carlo Bootstrap methods, error rate results using multiple Bootstrap iterations sizes are compared.

Then in section 4.7 results from a real data example using microarray data are reported and discussed. In this real data example, the E-skew method is compared to Monte Carlo Bootstrap methods using the Kappa agreement statistic.

4.1 Sample Mean

For the sample mean portion of the simulation study, results for six different sample sizes are reported ($n =$ 5, 10, 15, 20, 30, and 40). For each of these sample sizes, confidence interval error rates are reported at the α = 0.01, 0.05, and 0.10 significance levels.

The probability distributions used in the simulation study for the sample mean were the normal, exponential, gamma, log-normal, and mixture of two normal distributions. For each distribution, the population parameters specified are displayed below in Table 4.1. These parameter specifications are the same as the specifications for the sample mean in Table 3.5 in Chapter 3.

For each sample size, population parameter specification, and probability distribution combination 10,000 separate samples were generated. For the Monte Carlo Bootstrap confidence interval methods each of the 10,000 samples used 10,000 Monte Carlo Bootstrap resamples to create its Bootstrap sampling distribution. The comparisons discussed in this section are made between EBSD(n) methods and Bootstrap methods that use 10,000 Bootstrap resamples. In addition, confidence interval method error rate results were measured on the same 10,000 unique samples using 200, 500, and 1,000 Bootstrap resamples. These alternative Bootstrap resampling levels were performed for the normal distribution and exponential distribution simulations. The error rate results at these additional Bootstrap resampling levels are reported in the Appendix. Each generated unique sample had confidence intervals computed using the confidence interval methods listed below.

- For methods using EBSD(n) this included: E-skew, ET, EBC, EP, EBC_a, and ES.
- For methods using the Monte Carlo Bootstrap this includes: BT, BC, BP, BC_a/ABC, and BS.

Below in Table 4.1 is a description of the parameter specifications used for the sample mean statistic in this simulation study:

Table 4.1 Simulation Parameter Specifications for Sample Mean		
Probability distribution	Population Parameter	Parameter code: Specified Parameter Values
Normal distribution	(μ, σ)	N1: (4, 1) N2: (4, 4) N3: (4, 8)
Exponential distribution	λ	E1: 0.10 E2: 0.01 E3: 1
Gamma distribution	(α, λ)	G1: (2, 2) G2: (2, 3)
Log-Normal distribution	(μ, σ)	LN1: (4, 0.2) LN2: (4, 3)
Mixture of two normal distributions	$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	MN1: (4, 4, 0.5, 8, 8, 0.5) MN2: (4, 4, 0.8, 8, 8, 0.2)

For each of these specified distributions an error rate was computed using the 10,000 unique samples for each confidence interval method. The percent error of each confidence interval's error rate was computed by comparing the error rate result from the 10,000 unique samples against the theoretical nominal error rate. Because three different significance levels were studied the theoretical nominal error rate was different depending on the level specified. Error rates were also computed separately for the upper limit and the lower limit. Therefore, the theoretical nominal error rate in this study for the α = 0.01 significance level is 0.005, for the α = 0.05 significance level it is 0.025, and for the 0.10 significance level it is 0.05. Percent errors of these error rates were computed as discussed in Section 3.5. The smaller the percent error the closer the actual error rate is to the true theoretical nominal error rate. The closer the error rate is to the nominal error rate the more accurate the confidence interval.

Each error rate results table included in sections 4.1-4.5 is assigned a combination of letters and numbers as it's table "number". Each table "number" begins with a probability distribution code that is assigned using capitalized letters to refer to the probability distribution the table pertains to. Each corresponding parameter specification is assigned the value "1", "2" or "3" as displayed in Table 4.1. Specification numbering is also performed in tables 4.2, 4.3, 4.4 and 4.5 for each respective section. Probability distribution codes for the sample mean are:

- N = the normal distribution
- E = the exponential distribution
- G = the gamma distribution
- LN = the log-normal distribution
- MN = the mixture of two normal distributions

Following the statistical distribution code and parameter specification number, a capitalized "U" or "L" is assigned following the parameter specification number to indicate whether the table refers to an error rate for the lower limit or upper limit of the confidence interval. Lastly, following the confidence interval limit letter, a confidence interval number is assigned that corresponds to the α significance level measured in that table. For confidence intervals reported at the α=0.01 significance level, "99" is assigned after the interval limit letter. Similarly, a table reporting error rate results at the α = 0.05 significance level has a "95" assigned and table reporting at the α = 0.10 significance level has a "90" assigned.

As an example, the table "N1U99" indicates the results were computed using data generated from a normal distribution for the first parameter specification considered. The specification may change depending on the statistic studied. The "U" indicates the results are specific to the upper limit of the confidence interval. Further the "99" indicates the results are computed for the α = 0.01 significance level.

For figure "numbering" the parameter specification number is removed as all parameter specifications for a given statistic, distribution, significance level and limit end are presented in one figure. Therefore, Figure NU99 on page 74 displays error rates at all three specifications: $N(\mu = 4, \sigma = 1)$, $N(\mu = 4, \sigma = 4)$, and $N(\mu = 4, \sigma = 8)$ for the sample mean statistic.

As a final note, for all error rate result tables in section 4.1-4.5 the method that achieved the error rate with the smallest percent error has its error rate and percent error bolded at each sample size. The method that yielded the error rate with the smallest percent error was the most accurate at that sample size for that set of 10,000 unique samples.

a. Normal Distribution

This sub section has two purposes. The first is to compare the accuracy of E-skew to the accuracy of all other methods studied for the sample mean statistic when data is normally distributed. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods again for the sample mean statistic when data is normally distributed. Below mean error rates are compared for data generated from the normal distribution. First the results for data generated from $N(\mu = 4, \sigma = 1)$ distribution at the α = 0.01 significance level are considered. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables N1U99 and N1L99 on pages 68 and 69 below. For the sample mean statistic three separate normal distributions at three α significance levels were studied. However, in this section because of the volume of error rate results, only these error rate results are displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the $N(\mu = 4, \sigma = 1)$ specification, the $N(\mu = 4, \sigma = 4)$ and $N(\mu = 4, \sigma = 8)$ results can be viewed visually in figures NU99, NL99, NU95, NL95, NU90 and NL90 on pages 74-79. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

Compared to the other methods studied, E-skew performed relatively less accurately at the $\alpha = 0.01$ significance level and relatively more accurately at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels. For the $N(\mu = 4, \sigma = 1)$ parameter specification at the specified $\alpha = 0.01$ significance level for the upper limit, E-skew had the error rate with the smallest percent error at sample size 15. E-skew also had the error rate with a percent error as small as any method using EBSD(n) at sample size 30 for both the upper and lower limit as well.

Other methods using EBSD(n) were relatively more accurate compared to E-skew and Monte Carlo Bootstrap methods at the $\alpha = 0.01$ significance level than when they were compared to these methods at the α = 0.05 and α = 0.10 significance levels. For sample size 40 at the α = 0.01 significance level for the upper limit, the EP method had the error rate with the smallest percent error. At sample size 10 for the upper limit, the ET, BT and BS methods all had the error rate with the equally smallest percent error. For the lower limit at this specified α significance level, at sample sizes 20 and 40, the EP method had the error rate with the smallest percent error. These results are shown below in N1U99 and N1L99. These results can also be viewed visually in figures NU99 and NL99.

Although several methods performed accurately using EBSD(n) at the α = 0.01 significance level, the strength of the E-skew method is demonstrated when comparing the error rates across α significance level. For example in the case of the normal distribution $N(\mu = 4, \sigma = 1)$, the E-skew method maintained or improved accuracy comparatively as the α significance level increased from α = 0.01 to α = 0.10. At the α = 0.01 significance level for the upper limit, E-skew attained the error rate with the smallest percent error for one sample size. At the α = 0.05 significance level, E-skew only attained the error rate with the smallest percent error at one sample size for the upper limit and one sample size for the lower limit. However, at this α significance level for the upper limit, E-skew's percent error was also smaller at every sample size when compared to any other method applied on EBSD(n). Further, for the upper limit at this significance level, it attained an error rate with a smaller percent error compared to any Monte Carlo Bootstrap method other than BS. Similarly for the lower limit at the α = 0.05 significance level, it attained an error rate with a smaller percent error compared to any other method implemented on EBSD(n) at every sample size measured. At the α =

0.10 significance level E-skew attained the error rate with the *smallest* percent error for three sample sizes for the upper limit and one sample size for the lower limit even when compared to BS.

By comparison the EBC/EP/EBC_a/ET all were more accurate at the α = 0.01 significance level and became less accurate at larger α levels. At the α = 0.01 significance level for the upper limit, one of these four methods attained the error rate with the smallest percent error for two separate sample sizes (EP at sample size 40 and ET at sample size 10). At the α = 0.01 significance level for the lower limit, one of these methods attained the error rate with the smallest percent error at two separate sample sizes (EP at sample sizes 20 and 40). However, at the α = 0.05 and α = 0.10 significance levels, none of the four methods attained the error rate with the smallest percent error for any sample size for either the upper or lower limit.

Table: Sample Mean - N1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Normal Distribution, $\boldsymbol{N}(\boldsymbol{\mu} = 4, \boldsymbol{\sigma} = 1)$**, Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0111 (122%)	0.0089 (78%)	**0.005** **(0%)**	0.0055 (10%)	0.0052 (4%)	0.0056 (12%)
BT	0.0065 (30%)	**0.0072** **(44%)**	0.0053 (6%)	**0.005** **(0%)**	0.0053 (6%)	0.0059 (18%)
ET	0.0067 (34%)	**0.0072** **(44%)**	0.0053 (6%)	0.0051 (2%)	0.0053 (6%)	0.006 (20%)
BC	0.0506 (912%)	0.0224 (348%)	0.0123 (146%)	0.0096 (92%)	0.008 (60%)	0.0082 (64%)
EBC	0.0318 (536%)	0.0101 (102%)	0.0027 (46%)	0.0031 (38%)	0.0028 (44%)	0.0052 (4%)
BP	0.0521 (942%)	0.0227 (354%)	0.0119 (138%)	0.0098 (96%)	0.008 (60%)	0.0081 (62%)
EP	0.0318 (536%)	0.0096 (92%)	0.0029 (42%)	0.0035 (30%)	0.0022 (56%)	**0.005** **(0%)**
BS	**0.0059** **(18%)**	**0.0072** **(44%)**	0.0047 (6%)	0.0043 (14%)	**0.0051** **(2%)**	0.0058 (16%)
ES	0.1586 (3072%)	0.2649 (5198%)	0.2931 (5762%)	0.3249 (6398%)	0.3186 (6272%)	0.3434 (6768%)
$\boldsymbol{BC_a}$	0.0565 (1030%)	0.0235 (370%)	0.0124 (148%)	0.0101 (102%)	0.0086 (72%)	0.0081 (62%)
$\boldsymbol{EBC_a}$	0.0494 (888%)	0.0098 (96%)	0.006 (20%)	0.0042 (16%)	0.0505 (910%)	0.0687 (1274%)
ABC	0.047 (840%)	0.0226 (352%)	0.0123 (146%)	0.01 (100%)	0.0086 (72%)	0.0082 (64%)

Table: Sample Mean - N1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Normal Distribution, $\boldsymbol{N}(\boldsymbol{\mu} = 4, \boldsymbol{\sigma} = 1)$**, Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0113 (126%)	0.0057 (14%)	0.0068 (36%)	0.0069 (38%)	0.0049 (2%)	0.0056 (12%)
BT	0.0069 (38%)	**0.005 (0%)**	0.006 (20%)	0.0073 (46%)	0.0049 (2%)	0.0056 (12%)
ET	0.0066 (32%)	0.0048 (4%)	0.0058 (16%)	0.0072 (44%)	0.0049 (2%)	0.0055 (10%)
BC	0.05 (900%)	0.02 (300%)	0.0145 (190%)	0.0129 (158%)	0.0081 (62%)	0.0086 (72%)
EBC	0.0287 (474%)	0.0107 (114%)	0.0033 (34%)	0.0044 (12%)	0.0027 (46%)	0.0055 (10%)
BP	0.0497 (894%)	0.0187 (274%)	0.0141 (182%)	0.0127 (154%)	0.008 (60%)	0.008 (60%)
EP	0.0318 (536%)	0.0082 (64%)	0.0027 (46%)	**0.0047 (6%)**	0.0025 (50%)	**0.0053 (6%)**
BS	**0.0052 (4%)**	0.0043 (14%)	**0.0053 (6%)**	0.0056 (12%)	**0.005 (0%)**	0.0056 (12%)
ES	0.1581 (3062%)	0.2597 (5094%)	0.3037 (5974%)	0.3258 (6416%)	0.3173 (6246%)	0.3462 (6824%)
$\boldsymbol{BC_a}$	0.0488 (876%)	0.0198 (296%)	0.0153 (206%)	0.013 (160%)	0.0075 (50%)	0.0081 (62%)
$\boldsymbol{EBC_a}$	0.032 (540%)	0.0069 (38%)	0.0057 (14%)	0.0054 (8%)	0 (100%)	0 (100%)
ABC	0.0452 (804%)	0.019 (280%)	0.0154 (208%)	0.0133 (166%)	0.0076 (52%)	0.008 (60%)

Simulations were not only performed for the normal distribution with parameters $N(\mu = 4, \sigma = 1)$. Simulations were also performed for parameter specifications $N(\mu = 4,$

σ = 4), and $N(\mu = 4$, σ = 8). When the value of σ was increased, E-skew performed relatively less accurately compared to the other methods studied at the α = 0.05 and α = 0.10 significance levels than it did at when σ was set to 1.

When σ was increased from 1 to 4, for the upper limit, E-skew had the error rate with the smallest percent error at only one sample size for the upper limit at all three α significance levels. When σ was increased further to 8, E-skew did not have the smallest percent at any sample size for the upper limit at the α = 0.01 and α = 0.05 significance levels. However, at the α = 0.10 significance level for the upper limit, E-skew had the error rate with the smallest percent error at three sample sizes. For the lower limit at the α = 0.01 significance level, E-skew attained the error rate with the smallest percent error only at one sample size. E-skew also attained the smallest percent error for the lower limit at the α = 0.05 and α = 0.10 significance levels for one and two sample sizes respectively.

Similarly, changing the parameter specification yielded a similar error rate pattern across α significance level for the other methods applied on EBSD(n). When σ was increased from 1 to 4, BS had the error rate with the smallest percent error at each of sample sizes 5, 10, 15 and 20 for each α significance level for the upper limit. For the lower limit at the α = 0.01 significance level, one of the EBC/EP/EBC_a/ET methods had the error rate with the smallest percent error at one of the six sample sizes. For the lower limit at the α = 0.05 and α = 0.10 significance levels, ET attained the error rate with the smallest percent error for one and two sample sizes respectively. In most cases for a given significance level and sample size combination when a method using EBSD(n) did not attain the error rate with the smallest percent error BS attained the smallest percent error.

Again when σ was increased to 8, at the α=0.01 significance level for the upper limit, ET had the error rate with the smallest percent error at sample sizes 10 and 40. In no other case did a method other than E-skew using EBSD(n) attain the error rate with the smallest percent error for the upper limit at the α=0.05 or α=0.10 significance levels. For the lower limit at the α=0.01 significance level, EP and ET both attained the error rate with the smallest percent error at sample size 40 and EBC_a attained the error rate with the smallest percent error at sample size 15. For the lower limit at the α=0.05 significance level, ET attained the error rate with the smallest percent error at sample size 40. In no other case did a method other than E-skew using EBSD(n) attain the error rate with the smallest percent error for the lower limit at any other sample size or α significance level. Again, BS attained the error rate with the smallest percent error most frequently when a method using EBSD(n) did not.

Another comparison of interest is comparing how the same algorithm performed using the Monte Carlo Bootstrap to how it performed using the EBSD(n) (i.e. how did ET compare to BT etc.). The specific comparisons below are related to the $N(\mu = 4, \sigma = 1)$ distribution but similar results occurred for each parameter specification.

ET and BT performed approximately equally accurately for both the upper and lower limit at all three α significance levels. The two methods had error rates with either equally small, slightly smaller, or slightly larger percent errors depending on the simulation and sample size.

Although there was not a marked difference between the ET and BT methods, when comparing the BC method to the EBC method and the BP method to the EP method at the α = 0.01 significance level there was a marked difference from the comparison at the α = 0.05 and the α = 0.10 significance levels. At the α = 0.01 significance level, EBC and EP had error rates with smaller percent errors at each sample size for both the upper and lower limit. This is intuitively logical when considering the EBSD(n) sampling distribution of $\bar{X}$. The sampling distribution from EBSD(n) has larger minimum and maximum values because of the repeated blocks that are automatically assigned. Thus at the α = 0.01 significance level, both percentile methods EP and EBC were relatively more accurate compared to their Monte Carlo Bootstrap counterpart because the range of the EBSD(n) sampling distribution was larger and thus the confidence intervals using EBSD(n) were wider.

Although EP and EBC performed more accurately compared to their Monte Carlo Bootstrap counterparts at the α = 0.01 significance level, they did not at the α = 0.05 and α = 0.10 significance levels. At each sample size for both the α = 0.05 and α = 0.10 significance levels, for both the upper and lower limit, the EBC and EP methods had error rates with larger percent errors compared to BC and BP respectively.

EBC_a performed better at sample sizes 5, 10, 15 and 20 for the upper limit at the α = 0.01 significance level compared to BC_a. EBC_a also had an error rate with a smaller percent error compared to BC_a at sample size 5-20 for the lower limit at the α = 0.01 significance level. Additionally, EBC_a had an error rate with a smaller percent error at sample size 5 for the lower limit at both the α = 0.05 and α = 0.10 significance levels. In the remaining cases for these two significance levels, the error rate for BC_a had a percent error that was smaller at each sample size and significance level in comparison to EBC_a.

Figure: Sample Mean - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Normal Distribution

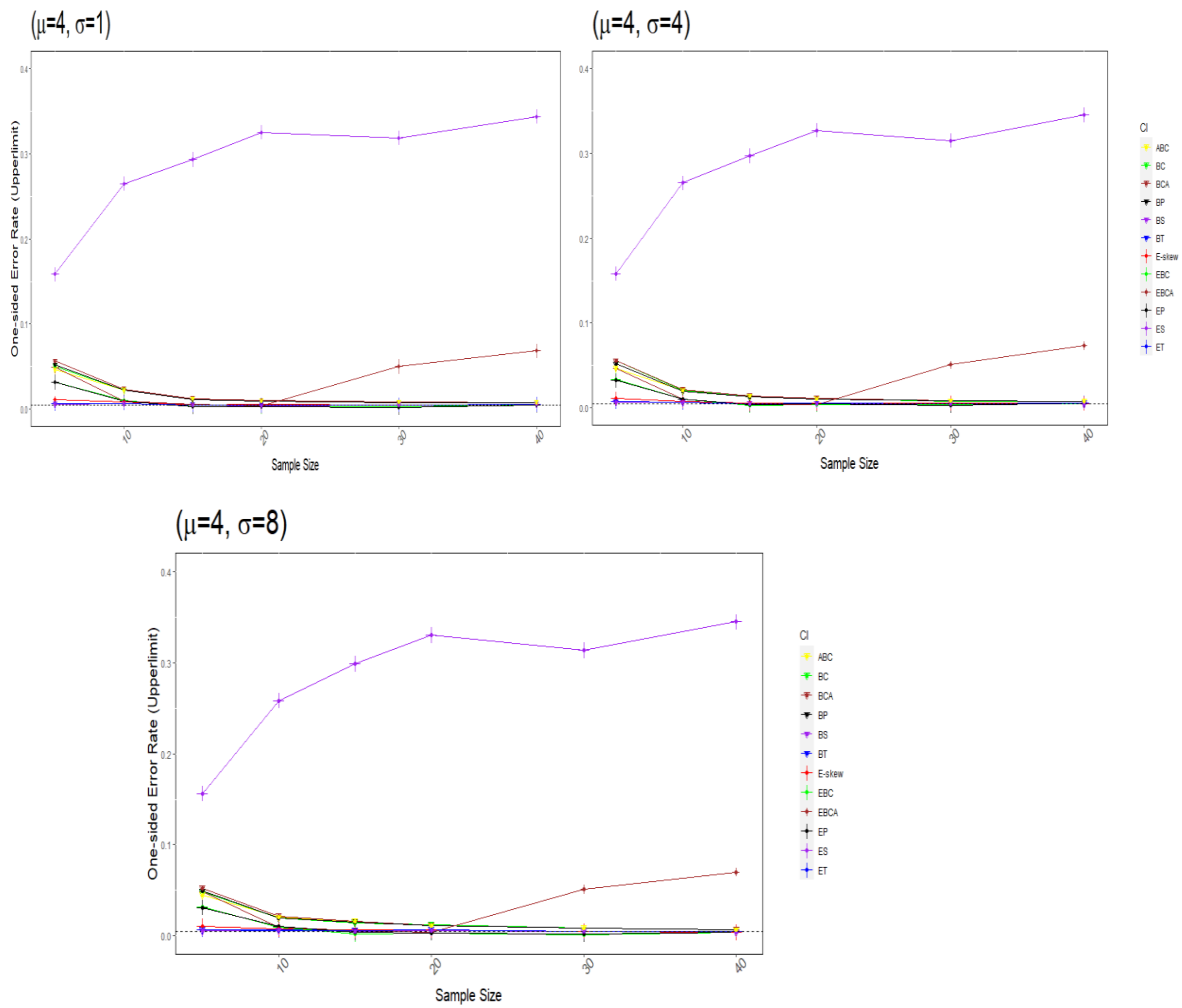


Figure: Sample Mean - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Normal Distribution

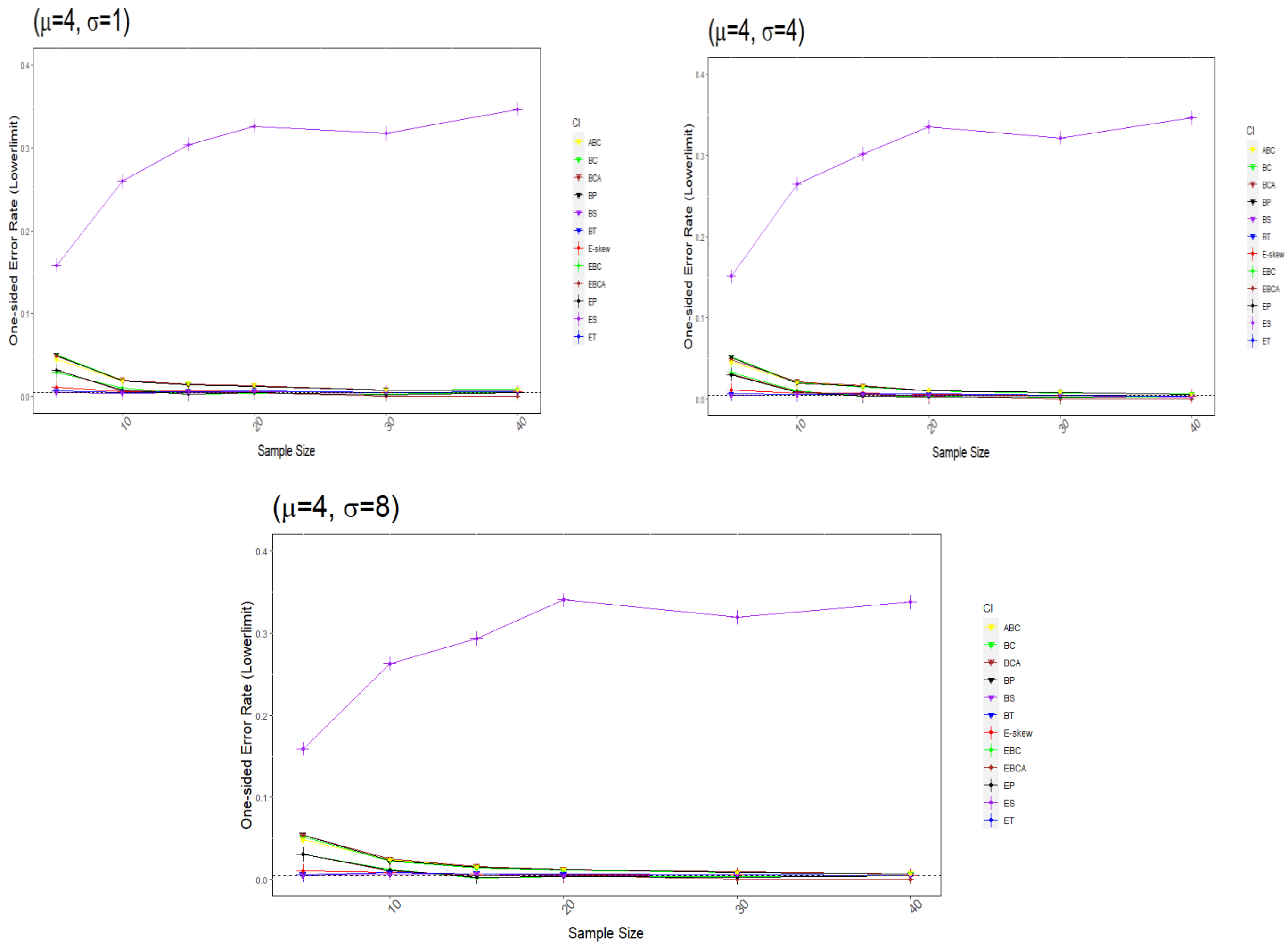


Figure: Sample Mean - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Normal Distribution

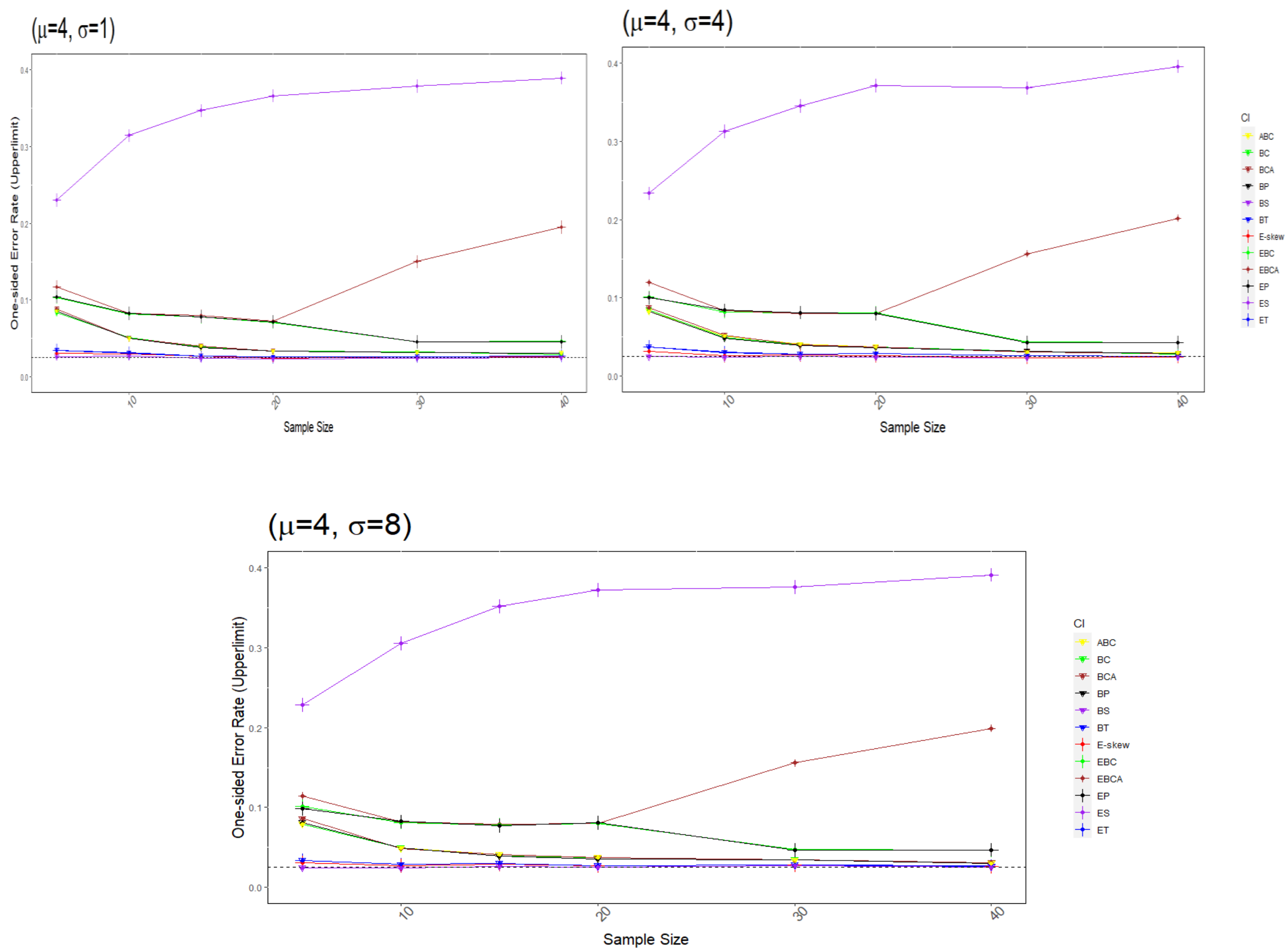


Figure: Sample Mean - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Normal Distribution

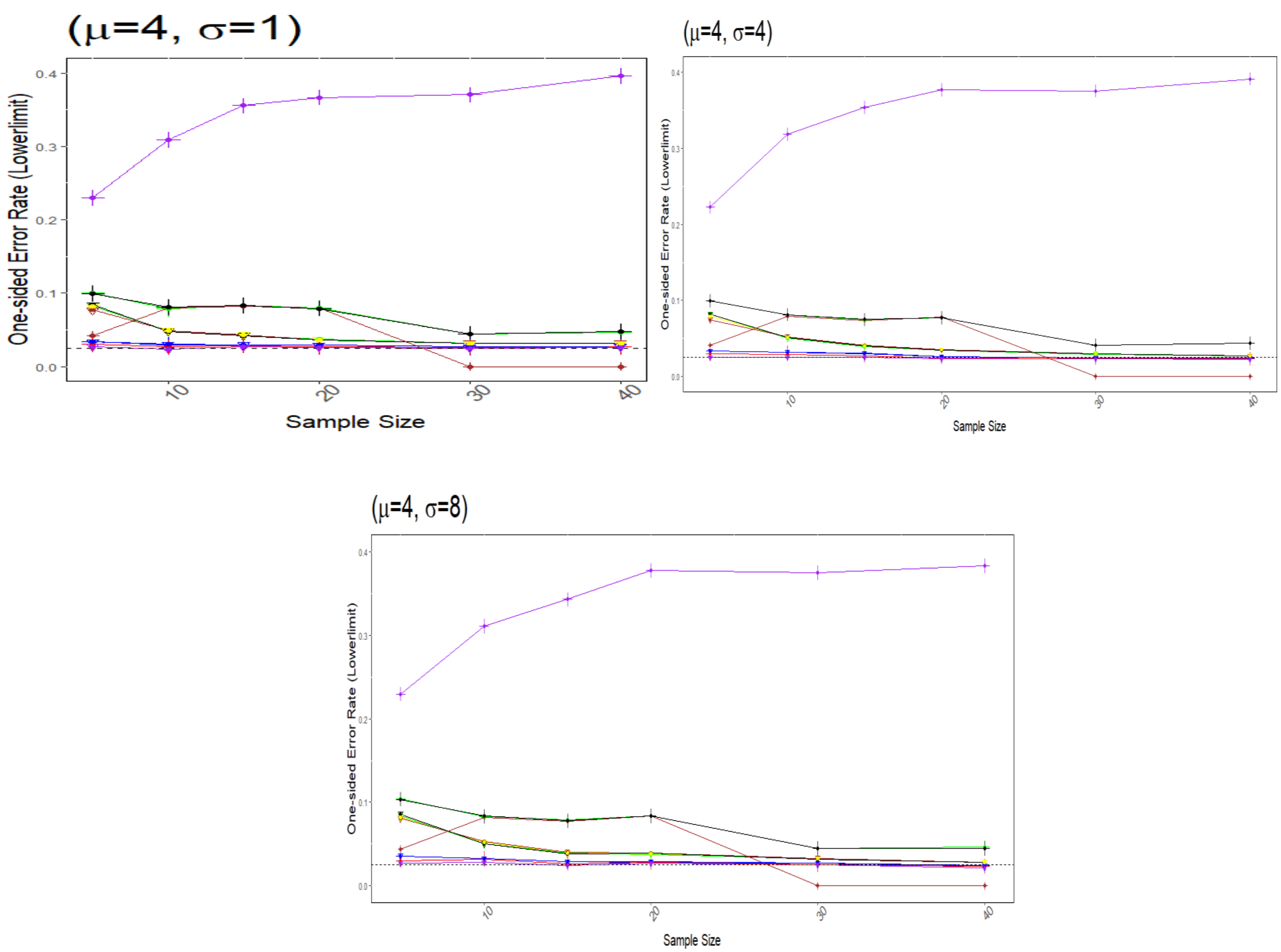


Figure: Sample Mean - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Normal Distribution

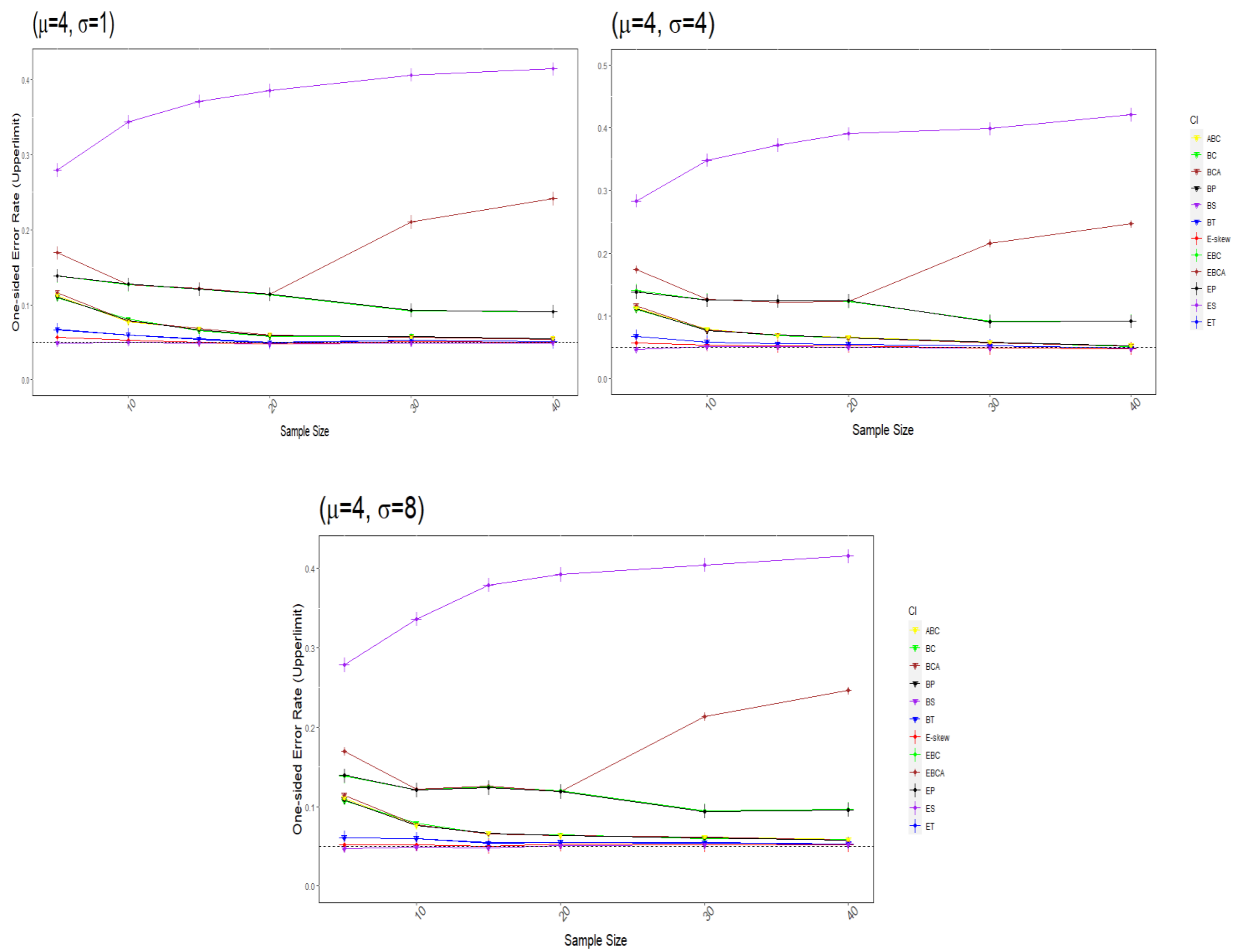
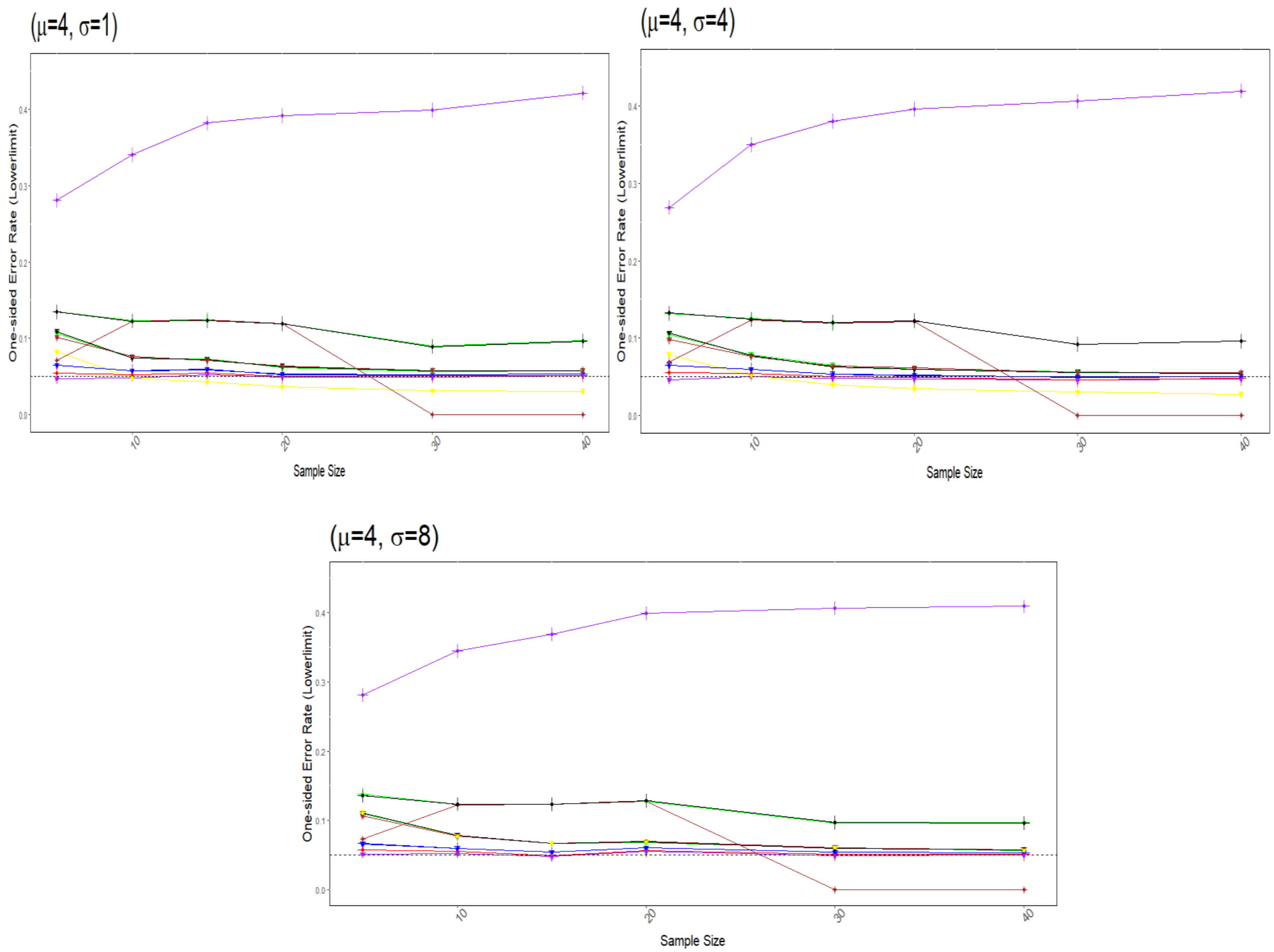


Figure: Sample Mean - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Normal Distribution



b. Exponential Distribution

The purpose of this sub section is to compare E-skew error rates and error rates for other methods using EBSD(n) to Monte Carlo Bootstrap method error rates for the sample mean statistic when data is exponentially distributed. Below mean error rates are compared for data generated from an exponential distribution with λ=0.10. The error results at the α = 0.01 significance level are displayed in tables E1U99, and E1L99 on pages 83 and 84 below. Three separate exponential distributions were studied, however because of the volume of error rate results, only the results for the exponential distribution with Exp(λ=0.10) at the α = 0.01 significance level are displayed in the tables. Detailed numerical results for simulations not included here can be viewed in Appendix tables. Although these tables only report results for the Exp(λ=0.10) specification, the Exp(λ=1) and Exp(λ=0.01) results can be viewed visually in figures EU99, EL99, EU95, EL95, EU90 and EL90 on pages 88-93.

For data generated from an exponential distribution with λ=0.10, at the specified α = 0.01 significance level, E-skew performed relatively less accurately than the other methods applied on EBSD(n). For the upper limit at sample size 40, E-skew did have the error rate with the smallest percent error compared to any method using EBSD(n). Although for the upper limit E-skew had the smallest percent error among methods applied on EBSD(n), E-skew did not have the error rate with the smallest percent error when compared to all Monte Carlo Bootstrap methods. However, for the lower limit at the α = 0.01 significance level, E-skew did have the error rate with the smallest percent error at sample size 15 and 20 compared to all other methods.

Other methods applied on EBSD(n) did perform relatively more accurately compared to E-skew and Monte Carlo Bootstrap methods at the α = 0.01 significance level. For the upper limit at the α = 0.01 significance level, the EBC_a method did perform most accurately for sample sizes 10, 15 and 20. For the lower limit at this significance level, the EBC_a, EP, and EBC methods all tied for the error rate with the smallest percent error at sample size 5. For methods that used the Monte Carlo Bootstrap, the BS method had the error rate with the smallest percent error at sample sizes 5, 30 and 40 for the upper limit but did not have the error rate with the smallest percent error at any sample size for the lower limit. The ABC method had the error rate with the smallest percent error at sample size 30 and 40 for the lower limit but did not have the error rate with the smallest percent error for any sample size for the upper limit.

Once again, the strength of the E-skew method is demonstrated when comparing the error rates across α significance level. The E-skew method improved accuracy relative to methods applied on EBSD(n) as the α significance level increased from α = 0.01 to α = 0.10. At the α = 0.01 significance level, E-skew attained the error rate with the second smallest percent error at sample size 40 for the upper limit and the error rate with the smallest percent error for two sample sizes for the lower limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew also attained the error rate with the second smallest percent error for every sample size. At the α = 0.05 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at four sample sizes. At the α = 0.10 significance level for the lower limit, it achieved the error rate with the smallest percent error for one sample size and the error rate with the smallest other than ABC/BC_a methods for three other sample sizes.

The EBC_a method was more accurate compared to the other methods studied at the α = 0.01 significance level and was relatively less accurate compared to the other methods studied at larger α levels. At the α = 0.01 significance level for the upper limit, EBC_a had the error rate with the smallest percent error for three separate sample sizes. At the α = 0.05 and α = 0.10 significance levels for the upper limit, it did not have the error rate with the smallest percent error for any sample size. Further EBC_a's percent error was larger for every sample size when compared to E-skew for both significance levels. For the lower limit as stated above EBC_a/EP/EBC all tied for the error rate with the smallest percent error at sample size 5 for the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels, no other method using EBSD(n) other than E-skew attained an error rate with the smallest percent error at any sample size.

Table: Sample Mean - E1U99 Upper limit error rate (α = 0.01), Exponential Distribution Exp(λ = **0.10**), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0413 (726%)	0.0317 (534%)	0.0224 (348%)	0.021 (320%)	0.0128 (156%)	0.0116 (132%)
BT	0.061 (1120%)	0.0546 (992%)	0.0414 (728%)	0.0397 (694%)	0.0295 (490%)	0.0242 (384%)
ET	0.0607 (1114%)	0.0545 (990%)	0.0413 (726%)	0.0399 (698%)	0.0293 (486%)	0.0244 (388%)
BC	0.1827 (3554%)	0.1109 (2118%)	0.0826 (1552%)	0.0713 (1326%)	0.0502 (904%)	0.0397 (694%)
EBC	0.1548 (2996%)	0.0639 (1178%)	0.0287 (474%)	0.0206 (312%)	0.0117 (134%)	0.0136 (172%)
BP	0.1374 (2648%)	0.0745 (1390%)	0.0499 (898%)	0.0438 (776%)	0.0287 (474%)	0.0222 (344%)
EP	0.0985 (1870%)	0.0576 (1052%)	0.0359 (618%)	0.0473 (846%)	0.0331 (562%)	0.0298 (496%)
BS	**0.0177 (254%)**	0.0166 (232%)	0.014 (180%)	0.0148 (196%)	**0.0084 (68%)**	**0.0087 (74%)**
ES	0.1723 (3346%)	0.2455 (4810%)	0.2676 (5252%)	0.3115 (6130%)	0.2649 (5198%)	0.3064 (6028%)
BC_a	0.1267 (2434%)	0.0557 (1014%)	0.0339 (578%)	0.0279 (458%)	0.0166 (232%)	0.0137 (174%)
EBC_a	0.1129 (2158%)	**0.0147 (194%)**	**0.0061 (22%)**	**0.0032 (36%)**	0.1935 (3770%)	0.1975 (3850%)
ABC	0.1113 (2126%)	0.0532 (964%)	0.0325 (550%)	0.0277 (454%)	0.0162 (224%)	0.0137 (174%)

Table: Sample Mean - E1L99 Lower limit error rate (α = 0.01), Exponential Distribution, Exp($\lambda = 0.10$), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0026 (48%)	0.0042 (16%)	**0.0056 (12%)**	**0.005 (0%)**	0.0071 (42%)	0.0063 (26%)
BT	6e-04 (88%)	3e-04 (94%)	2e-04 (96%)	5e-04 (90%)	6e-04 (88%)	5e-04 (90%)
ET	6e-04 (88%)	3e-04 (94%)	2e-04 (96%)	5e-04 (90%)	6e-04 (88%)	5e-04 (90%)
BC	0.0105 (110%)	0.0026 (48%)	0.0011 (78%)	0.0011 (78%)	8e-04 (84%)	5e-04 (90%)
EBC	**0.0056 (12%)**	0.0018 (64%)	0.001 (80%)	0.0014 (72%)	4e-04 (92%)	8e-04 (84%)
BP	0.011 (120%)	**0.0047 (6%)**	0.0039 (22%)	0.003 (40%)	0.0024 (52%)	0.0018 (64%)
EP	**0.0056 (12%)**	0.001 (80%)	2e-04 (96%)	1e-04 (98%)	1e-04 (98%)	2e-04 (96%)
BS	9e-04 (82%)	9e-04 (82%)	0.0013 (74%)	0.0019 (62%)	0.0024 (52%)	0.0023 (54%)
ES	0.1492 (2884%)	0.2828 (5556%)	0.3176 (6252%)	0.3354 (6608%)	0.3396 (6692%)	0.355 (7000%)
BC_a	0.0126 (152%)	0.0088 (76%)	0.0091 (82%)	0.006 (20%)	0.0062 (24%)	0.006 (20%)
EBC_a	**0.0056 (12%)**	0.0025 (50%)	0.002 (60%)	0.0019 (62%)	0 (100%)	0 (100%)
ABC	0.0115 (130%)	0.008 (60%)	0.0082 (64%)	0.0061 (22%)	**0.0061 (22%)**	**0.0055 (10%)**

Simulation were not only performed for the exponential distribution with parameter Exp($\lambda = 0.10$). Simulations were also performed for the parameter specifications $\lambda = 0.01$, and $\lambda = 1$. When increasing the value of λ, a similar pattern of results was seen across α significance level for E-skew. When λ was increased from 0.10 to 1, for the upper limit at the α = 0.01 significance level, E-skew attained the error rate with the second smallest percent error at sample size 40 for the upper limit and the error rate with the smallest percent error for four separate sample sizes for the lower limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew attained the error rate with the second smallest percent error for every sample size. At the α = 0.05 and α = 0.10 significance levels for the lower limit, E-skew attained the error rate with the smallest percent error at one sample size and an error rate with a smaller percent error compared to any other method other than ABC/BC_a at three other sample sizes for each significance level. When λ was decreased from 0.10 to 0.01, a similar pattern across α level occurred for E-skew as well.

Similarly, changing the parameter specification yielded a similar error rate pattern across α level for each method studied as it did when λ =0.10 was specified. When λ was increased from 0.10 to 1, the EBC_a method was most accurate for sample sizes 10, 15 and 20 at the α = 0.01 significance level for the upper limit. However, when α significance level was modified, EBC_a was relatively less accurate compared to the other methods studied at these sample sizes for the upper limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, it did not have the error rate with the smallest percent error for any sample size. Further EBC_a's percent error was larger for every sample size when compared to E-skew for both significance levels. At the α = 0.01

significance level for the lower limit, the EBC method attained the error rate with the smallest percent error at sample size 5. No other method using EBSD(n) other than E-skew attained the error rate with the smallest percent error at any other sample size or α level.

Instead when λ was decreased to 0.01, at the α=0.01 significance level for the upper limit, the EBC_a method attained the error rate with the smallest percent error at the same three sample sizes, but then was unable to achieve the error rate with the smallest percent error at any sample size for the α = 0.05 and α = 0.10 significance levels. For the lower limit at the α=0.01 significance level, EBC attained the error rate with the smallest percent error at sample size 5 but then no other method using EBSD(n) attained the error rate with the smallest percent error at any other sample size or α level.

Again a comparison of how the same algorithm performed using the Monte Carlo Bootstrap and EBSD(n) was made. The specific comparisons below are related to the Exp($\lambda = 0.10$) distribution but similar results occurred for each exponential parameter specification.

Like in the normal distribution case, although there was not a marked difference between the ET and BT methods, when comparing the BC method to the EBC method and the BP method to the EP method there was a marked difference between the comparison at the α = 0.01 significance level and the comparison at the α = 0.05, 0.10 significance levels. At the α = 0.01 significance level for the upper limit, the EBC method had an error rate with a smaller percent error at each sample size compared to the

BC method. The EP method also had an error rate with a smaller percent error at sample sizes 5, 10 and 15 compared to the BP method for the upper limit.

For exponentially distributed data, EBC_a also performed better at sample sizes 10, 15 and 20 for the upper limit at the α = 0.01 significance level compared to BC_a. EBC_a also had an error rate with a smaller percent error compared to BC_a at sample size 5 for the lower limit at the α = 0.01 significance level. In the remaining cases the error rate for BC_a had a percent error that was smaller at each sample size and significance level in comparison to EBC_a.

Again for exponentially distributed data the ES method had an error rate with a larger percent error at each sample size compared to the BS method for both the upper and lower limit.

Figure: Sample Mean - EU99 - One-Sided Upperlimit Error Rates for 99% CI for the Exponential Distribution

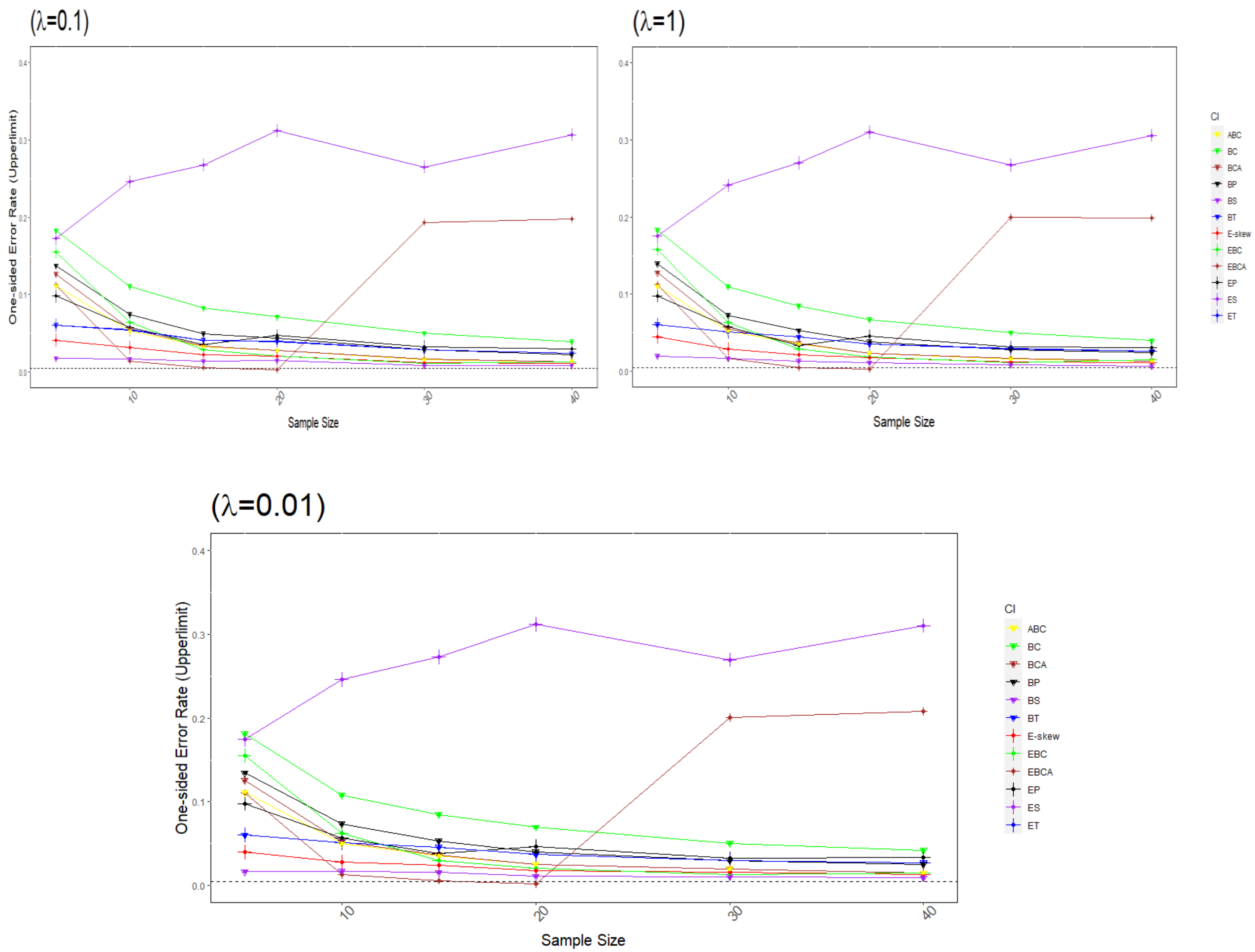


Figure: Sample Mean - EL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Exponential Distribution

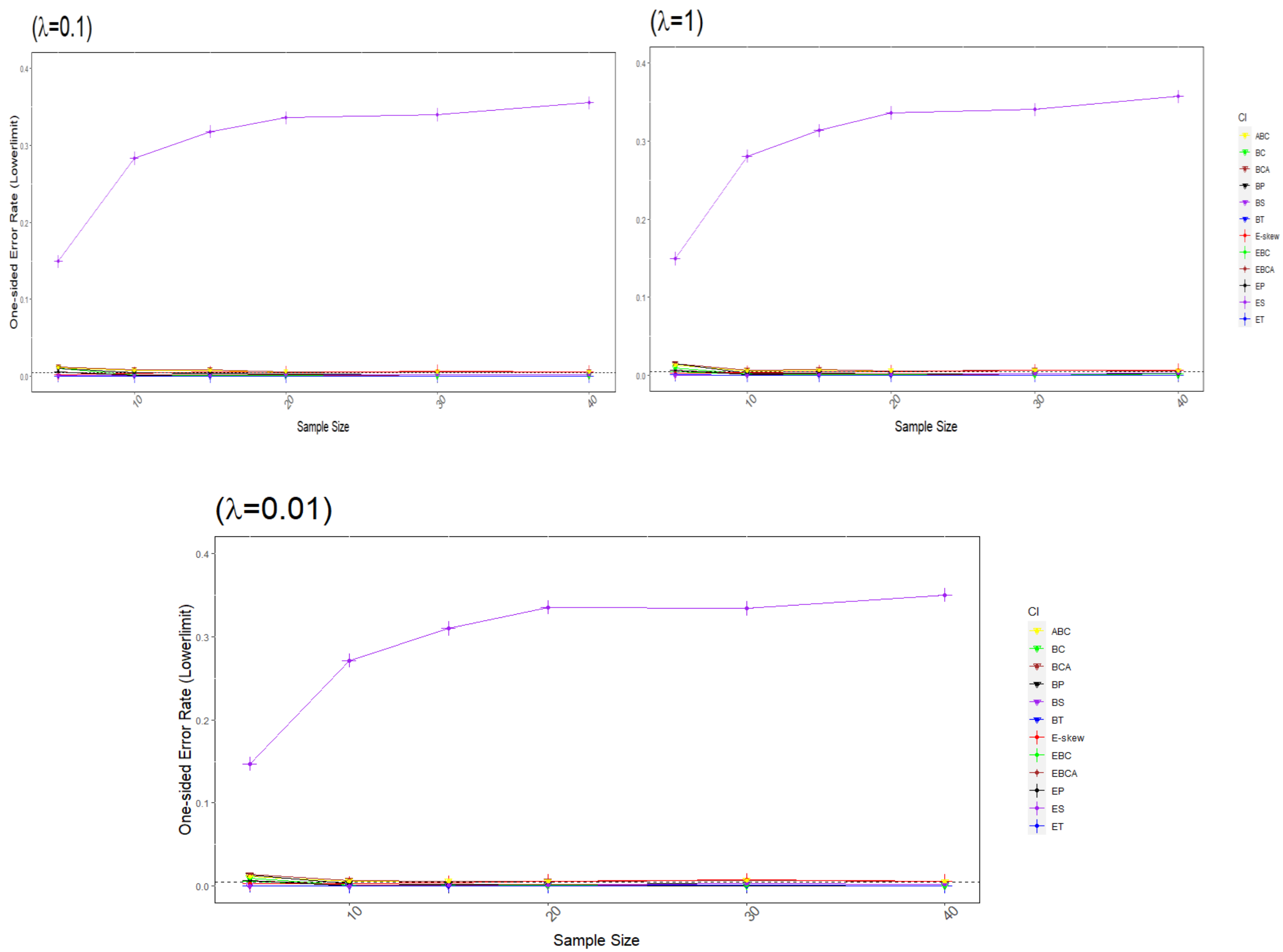


Figure: Sample Mean - EU95 - One-Sided Upperlimit Error Rates for 95% CI for the Exponential Distribution

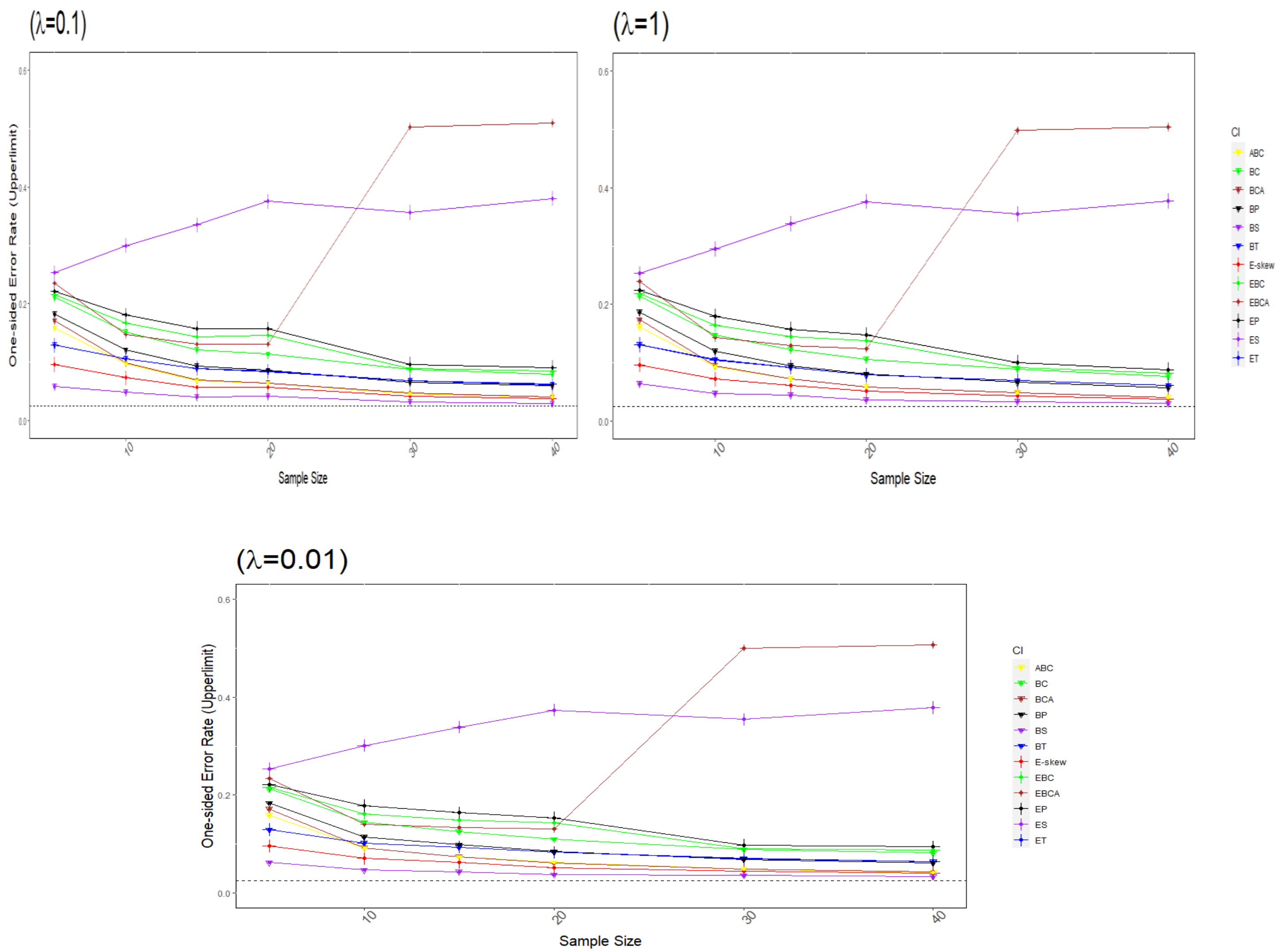


Figure: Sample Mean - EL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Exponential Distribution

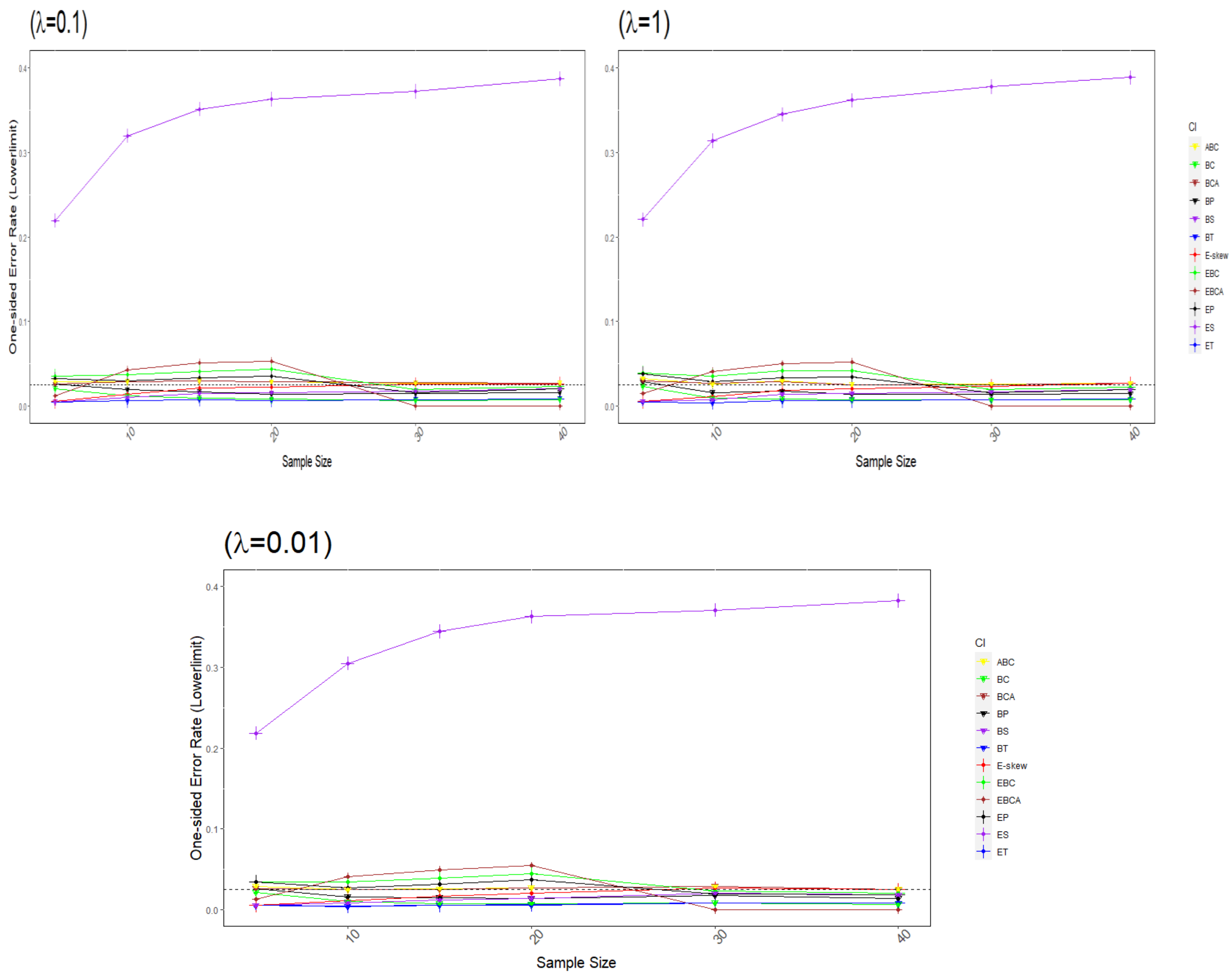


Figure: Sample Mean - EU90 - One-Sided Upperlimit Error Rates for 90% CI for the Exponential Distribution

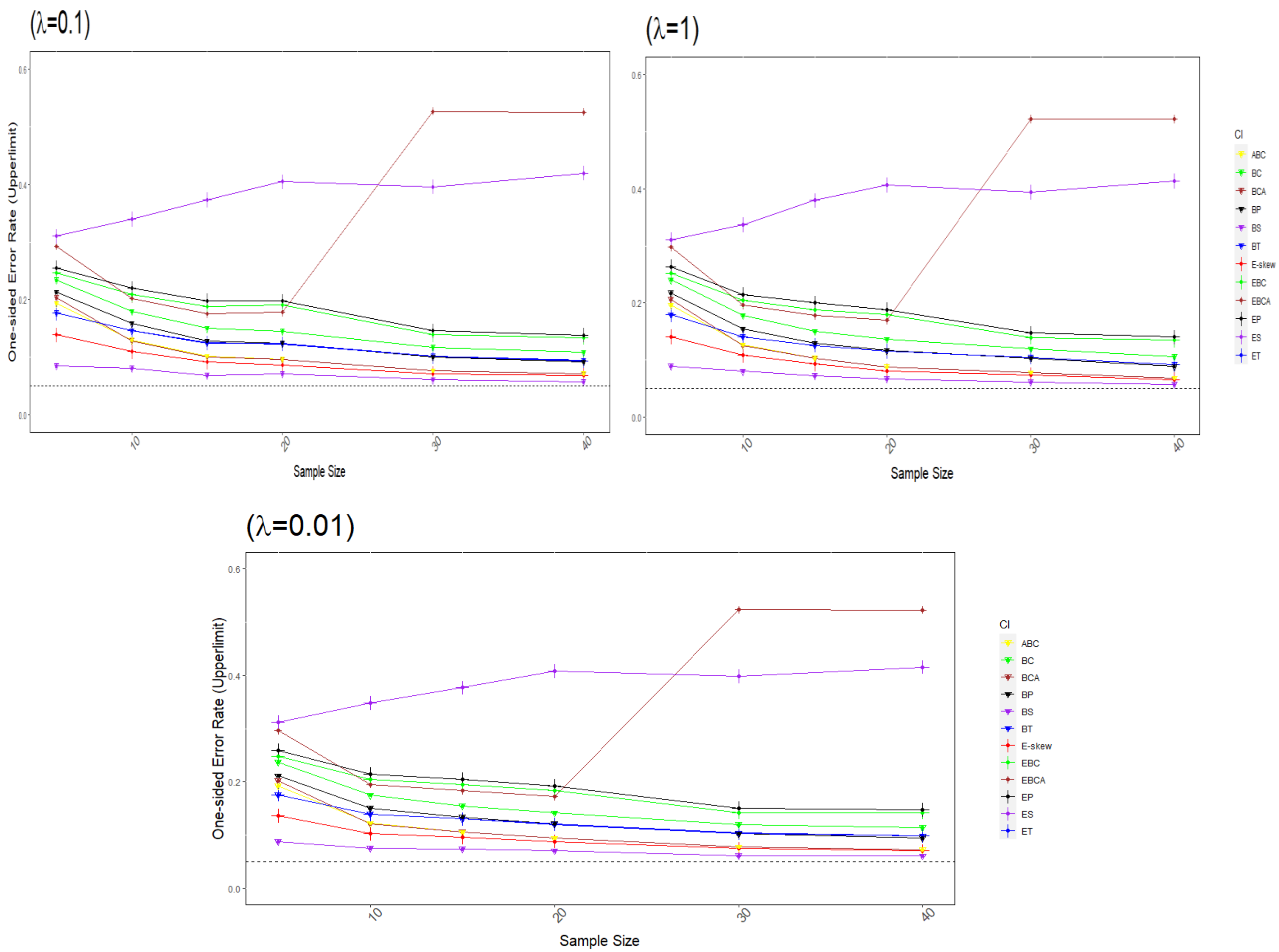
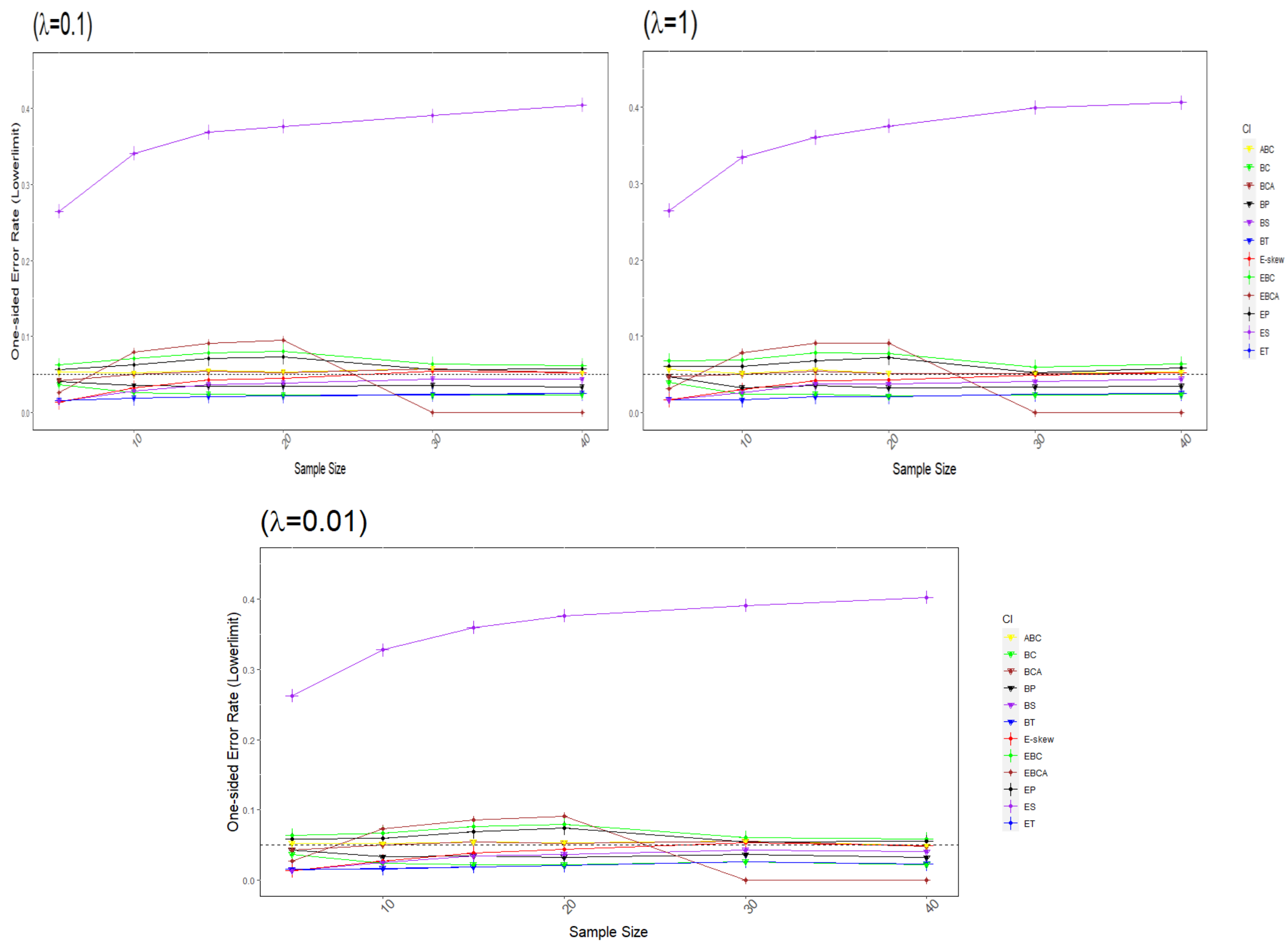


Figure: Sample Mean - EL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Exponential Distribution



c. Gamma Distribution

The purpose of this sub section is to compare E-skew error rates and error rates for other methods using EBSD(n) to Monte Carlo Bootstrap method error rates for the sample mean statistic when data is generated from a gamma distribution. The first comparison performed is for data generated from an gamma distribution with $\alpha = 2$ and $\lambda = 2$. The error results at the $\alpha = 0.01$ significance level are displayed in tables G1U99, and G1L99 on pages 96 and 97 below.

Two separate gamma distributions were studied, however because of the volume of error rate results in this section, only the results for the gamma($\alpha = 2$, $\lambda = 2$) at the $\alpha = 0.01$ significance level are displayed in the tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables. Although the tables only report results for the gamma($\alpha = 2$, $\lambda = 2$) specification, the gamma($\alpha = 2$, $\lambda = 3$) specification results can be viewed visually in figures GU99, GL99, GU95, GL95, GU90 and GL90 on pages 101-106. The ABC method for this sub section and the remaining sub sections was not included for these simulations. It was found that the ABC and BC_a results were very similar thus repeatedly reporting both results was considered redundant.

For the gamma($\alpha = 2$, $\lambda = 2$) parameter specification at the specified $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$ significance levels, error rate results for E-skew were very similar to what they were for data generated from the exponential distribution. For the upper limit at the $\alpha = 0.01$ significance level, E-skew had the error rate with the second smallest percent error at sample size 40. For the lower limit at the $\alpha = 0.01$ significance level, E-skew had the error rate with the smallest percent error at three sample sizes. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew had the error rate with the second smallest percenter error at each sample size and a smaller percent error than any other method used on EBSD(n) at each sample size. For the lower limit, E-skew had the error rate with the smallest percent error at sample size 40 at each α significance level. Additionally, at the α = 0.01 significance level for the lower limit, E-skew also had the error rate with the smallest percent error at sample size 5.

By comparison the EBC/EP/EBC_a/ET all were relatively more accurate compared to the other methods studied at the α = 0.01 significance level and relatively less accurate compared to other methods at larger α significance levels. At the α = 0.01 significance level for the upper limit, EBC_a had the error rate with the smallest percent error at sample sizes 10, 15 and 20. Additionally, at the α = 0.05 significance for the lower limit, EBC_a had the error rate with the smallest percent error at sample size 5. For the α = 0.10 significance level for the lower limit, EBC had the error rate with the smallest percent error at sample size 30. Otherwise no method other than E-skew using EBSD(n) had the error rate with the smallest percent error at any sample size or significance level.

Table: Sample Mean - G1U99 Upper limit error rate (α = 0.01), Gamma Distribution, gamma($\alpha = 2,\ \lambda\ =\ 2$), **Bootstraps = 10000**

Sample size	5	10	15	20	30	40
E-skew	0.0281 (462%)	0.0223 (346%)	0.0147 (194%)	0.012 (140%)	0.0096 (92%)	0.0093 (86%)
BT	0.0353 (606%)	0.035 (600%)	0.025 (400%)	0.0248 (396%)	0.0182 (264%)	0.0191 (282%)
ET	0.0353 (606%)	0.0348 (596%)	0.0253 (406%)	0.0245 (390%)	0.0184 (268%)	0.0189 (278%)
BC	0.1313 (2526%)	0.0759 (1418%)	0.0525 (950%)	0.0444 (788%)	0.0328 (556%)	0.0293 (486%)
EBC	0.0999 (1898%)	0.0405 (710%)	0.0139 (178%)	0.0091 (82%)	0.0081 (62%)	0.0123 (146%)
BP	0.1035 (1970%)	0.0571 (1042%)	0.034 (580%)	0.0293 (486%)	0.0194 (288%)	0.0193 (286%)
EP	0.0712 (1324%)	0.0386 (672%)	0.0188 (276%)	0.0251 (402%)	0.0181 (262%)	0.0218 (336%)
BS	**0.0155** **(210%)**	0.0166 (232%)	0.0116 (132%)	0.0089 (78%)	**0.0076** **(52%)**	**0.0074** **(48%)**
ES	0.1687 (3274%)	0.2457 (4814%)	0.2776 (5452%)	0.3184 (6268%)	0.2874 (5648%)	0.3222 (6344%)
BC_a	0.0986 (1872%)	0.0462 (824%)	0.0258 (416%)	0.0197 (294%)	0.013 (160%)	0.0122 (144%)
EBC_a	0.0894 (1688%)	**0.0132** **(164%)**	**0.0054** **(8%)**	**0.0029** **(42%)**	0.2011 (3922%)	0.2067 (4034%)

Table: Sample Mean - G1L99 Lower limit error rate (α = 0.01), Gamma Distribution, gamma($\alpha = 2,\ \lambda\ =\ 2$), $\mathbf{Bootstraps = 10000}$

Sample size	5	10	15	20	30	40
E-skew	**0.0047** **(6%)**	0.0043 (14%)	0.0035 (30%)	0.0034 (32%)	**0.0048** **(4%)**	**0.0052** **(4%)**
BT	0.0017 (66%)	0.0014 (72%)	0.0011 (78%)	0.0011 (78%)	0.0015 (70%)	0.0016 (68%)
ET	0.0017 (66%)	0.0014 (72%)	0.001 (80%)	0.001 (80%)	0.0015 (70%)	0.0015 (70%)
BC	0.0198 (296%)	**0.0053** **(6%)**	0.0029 (42%)	0.0021 (58%)	0.002 (60%)	0.0015 (70%)
EBC	0.01 (100%)	0.0038 (24%)	0.0014 (72%)	0.0014 (72%)	8e-04 (84%)	0.0012 (76%)
BP	0.0211 (322%)	0.0071 (42%)	**0.0057** **(14%)**	**0.004** **(20%)**	0.0036 (28%)	0.0039 (22%)
EP	0.0113 (126%)	0.0027 (46%)	2e-04 (96%)	2e-04 (96%)	2e-04 (96%)	6e-04 (88%)
BS	0.0012 (76%)	0.002 (60%)	0.0016 (68%)	0.002 (60%)	0.0027 (46%)	0.0039 (22%)
ES	0.1444 (2788%)	0.2734 (5368%)	0.3131 (6162%)	0.3291 (6482%)	0.338 (6660%)	0.3558 (7016%)
$\boldsymbol{BC_a}$	0.0215 (330%)	0.0107 (114%)	0.01 (100%)	0.0065 (30%)	0.0064 (28%)	0.0069 (38%)
$\boldsymbol{EBC_a}$	0.0113 (126%)	0.0032 (36%)	0.0023 (54%)	0.0021 (58%)	0 (100%)	0 (100%)

Simulation were not only performed for the gamma distribution with parameters: gamma($\alpha = 2$, $\lambda = 2$). Simulations were also performed for the parameter specification: gamma($\alpha = 2$, $\lambda = 3$). When considering a change in parameter specification (i.e. increasing the value of λ) E-skew performed approximately as accurately at the $\alpha = 0.01$

significance level and relatively more accurately at the α = 0.05 and α = 0.10 significance levels.

When λ was increased from 2 to 3, for the upper limit at the α = 0.01 significance level, E-skew attained the error rate with the second smallest percent error at sample size 40 for the upper limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew attained the error rate with the second smallest percent error for every sample size. At the α = 0.01 significance level, E-skew achieved the error rate with the smallest percent error for four sample sizes for the lower limit. At the α = 0.05 significance level for the lower limit, E-skew attained the smallest percent error at one sample size and a smaller percent error than any other method other than BC_a for one other sample size. At the α = 0.10 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at two sample sizes and a smaller percent error than any method other than BC_a for two other sample sizes.

Similar results were observed for the other methods implemented on EBSD(n) when changing the parameter specification of λ from 2 to 3 across α significance level. When λ was increased from 2 to 3, the EBC_a method was again relatively more accurate compared to the other methods studied at the α = 0.01 significance level than it was at larger α levels. At the α = 0.01 significance level for the upper limit EBC_a had the error rate with the smallest percent error for two separate sample sizes. At the α = 0.05 and α = 0.10 significance levels for the upper limit it did not have the error rate with the smallest percent error for any sample size. Further EBC_a's percent error was larger for every sample size when compared to E-skew for both significance levels. At the α = 0.05

significance level for the lower limit, the EBC method attained the error rate with the smallest percent error at sample size 30, and EP attained the smallest percent error at sample size 40. No other method using EBSD(n) other than E-skew attained the error rate with smallest percent error at any other sample size or α level.

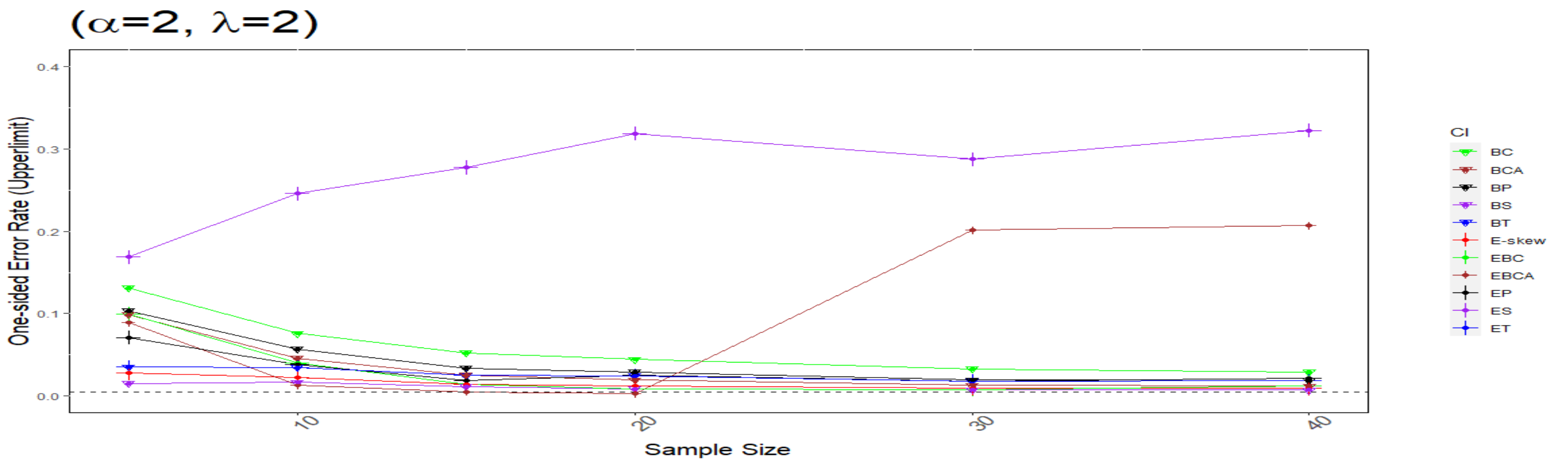
Similar patterns were again observed as they were for simulations from the exponential distribution when comparing methods using EBSD(n) to the same method using the Monte Carlo Bootstrap at each α level and sample size. The specific comparisons below are related to the gamma($\alpha = 2$, $\lambda = 2$) specification but similar results occurred for the gamma($\alpha = 2$, $\lambda = 3$) specification.

EBC had an error rate with a smaller percent error for every sample size when compared to BC for the upper limit at the $\alpha = 0.01$ significance level. EP had an error rate with a smaller percent error for every sample size except at sample size 40 for the upper limit at the $\alpha = 0.01$ significance level. Both methods had error rates with larger percent errors for the upper limit at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels for the upper limit at each sample size. For the lower limit at the $\alpha = 0.01$ significance level, EBC and EP each had error rates with smaller percent errors at sample size 5 compared to their BC and BP counter parts. For the lower limit at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels, EBC and EP also had error rates with smaller percent errors at sample size 30 and 40 compared to BC and BP respectively. For the remaining cases BC and BP had error rates with smaller percent errors.

Again EBC_a performed better at sample sizes 10, 15 and 20 for the upper limit at the $\alpha = 0.01$ significance level compared to BC_a. Also at sample size 5 for the lower limit at the α = 0.01 significance level, EBC_a had an error rate with a smaller percent error compared to BC_a. In the remaining cases the error rate for BC_a had a percent error that was smaller at each sample size and significance level in comparison to EBC_a.

Again ET and BT had very similar error rates at each sample size for both the upper and lower limit at each significance level. ES had a larger error rate at every sample size for both the upper and lower limit at each significance level when compared to BS.

Figure: Sample Mean - GU99 - One-Sided Upperlimit Error Rates for 99% CI for the Gamma Distribution



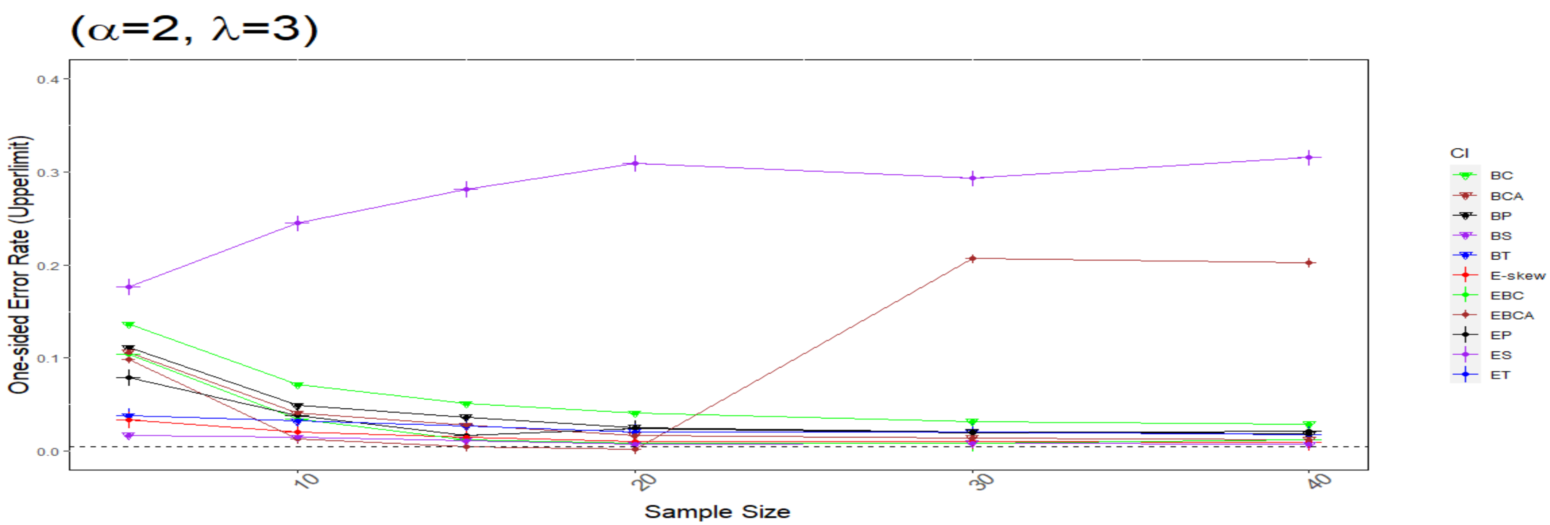
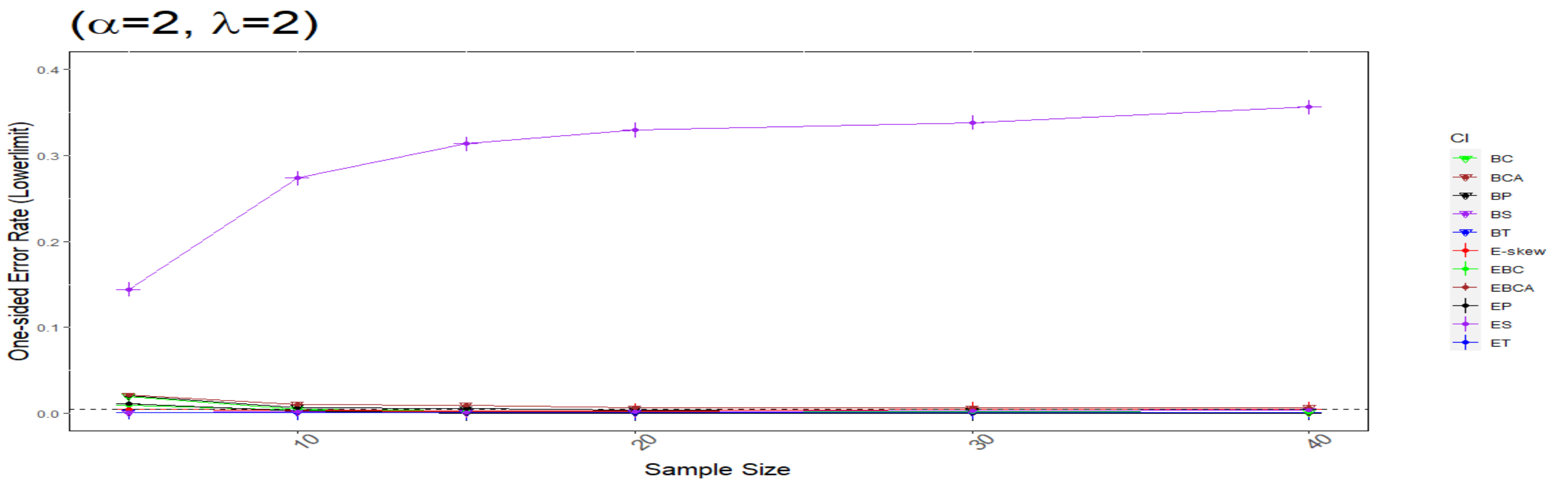


Figure: Sample Mean - GL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Gamma Distribution



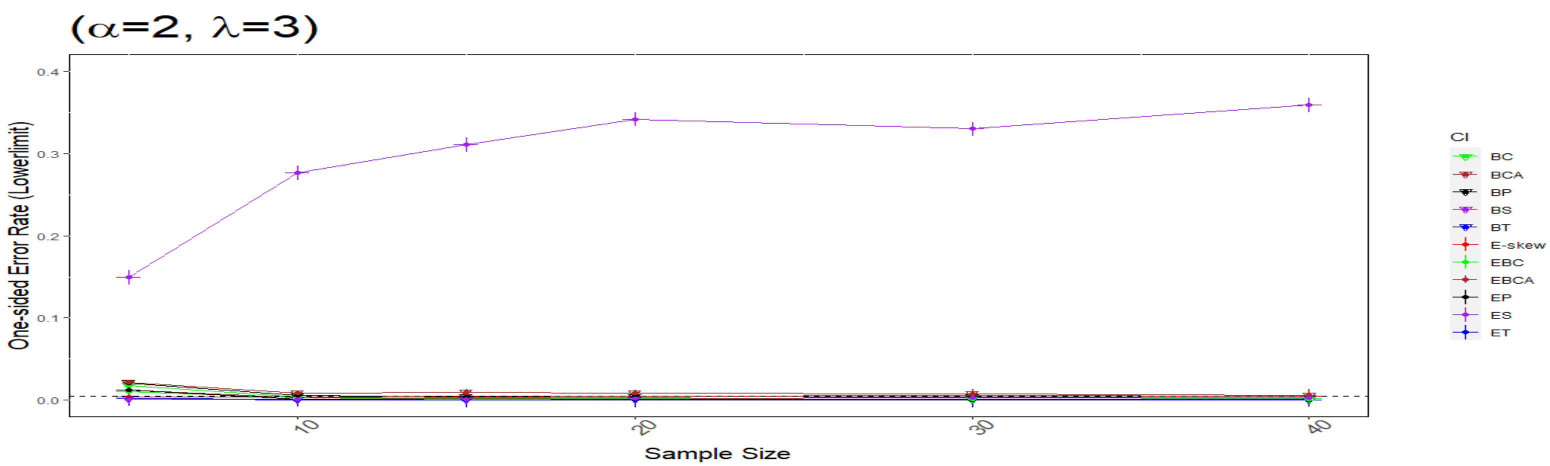


Figure: Sample Mean - GU95 - One-Sided Upperlimit Error Rates for 95% CI for the Gamma Distribution

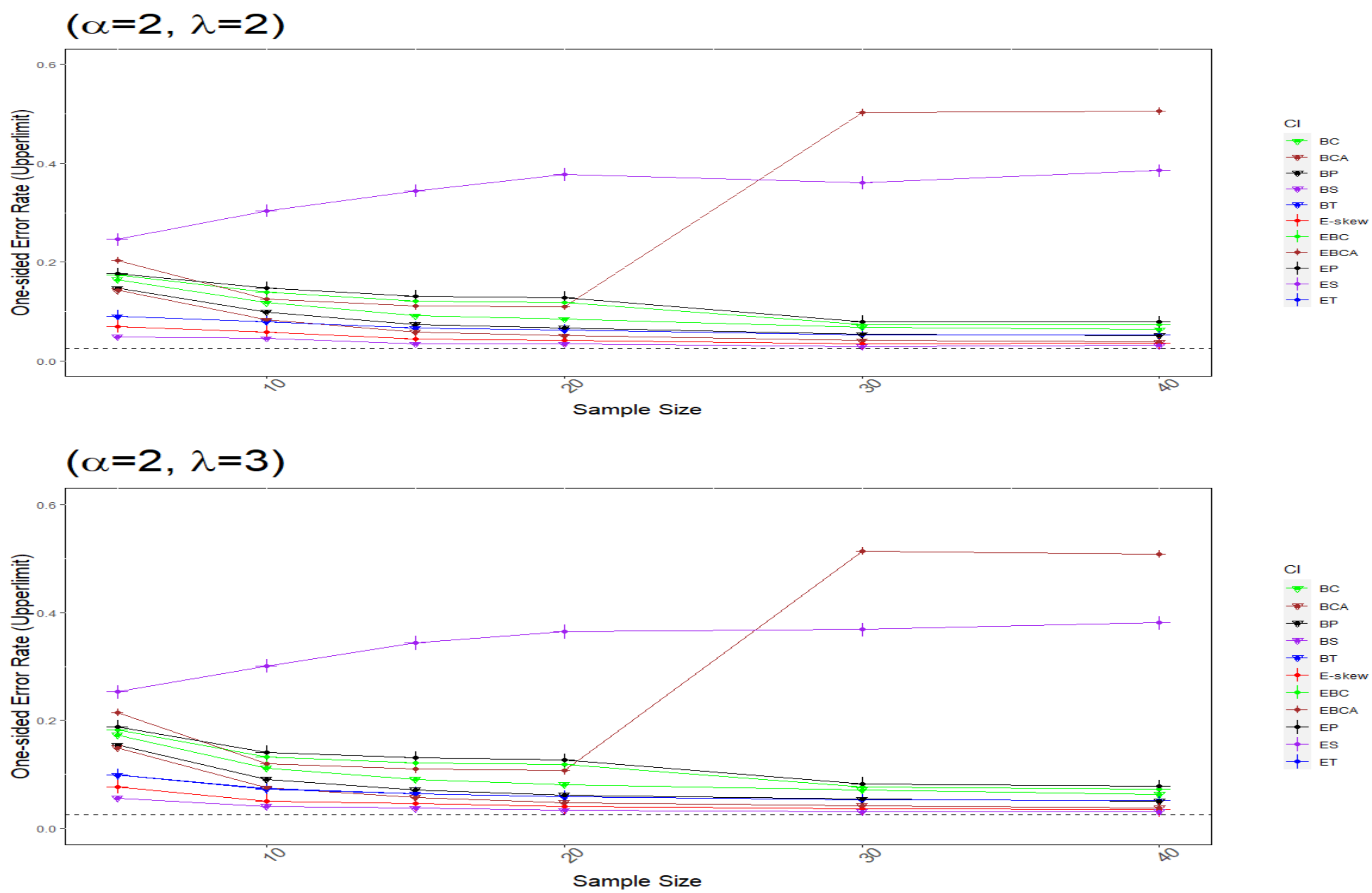
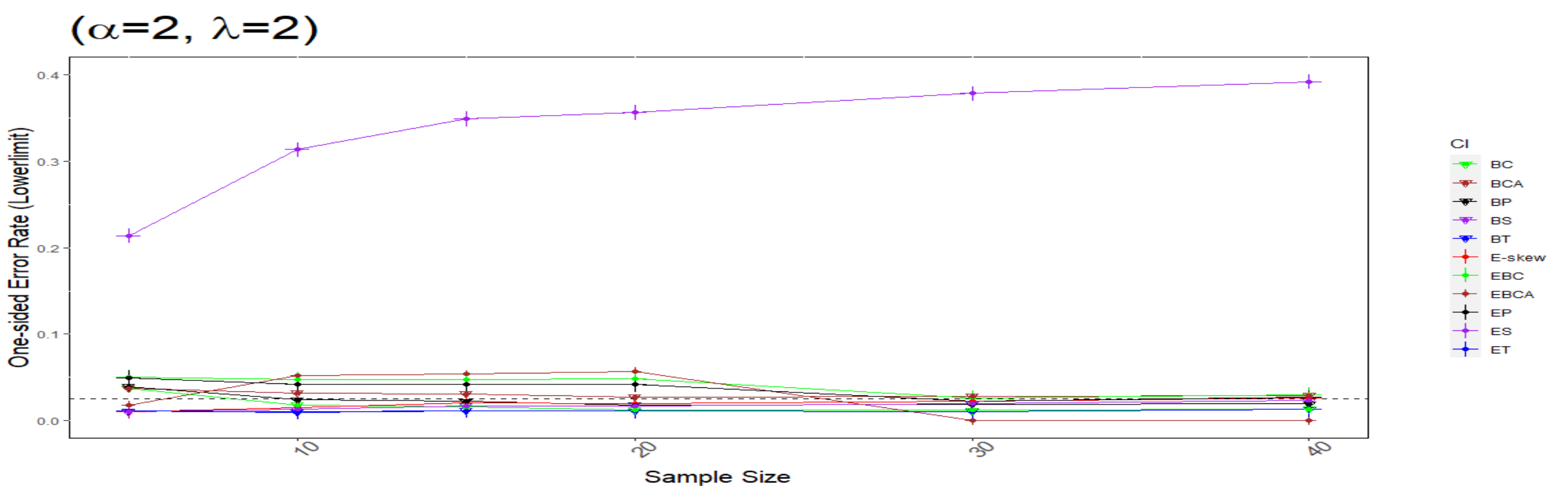


Figure: Sample Mean - GL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Gamma Distribution



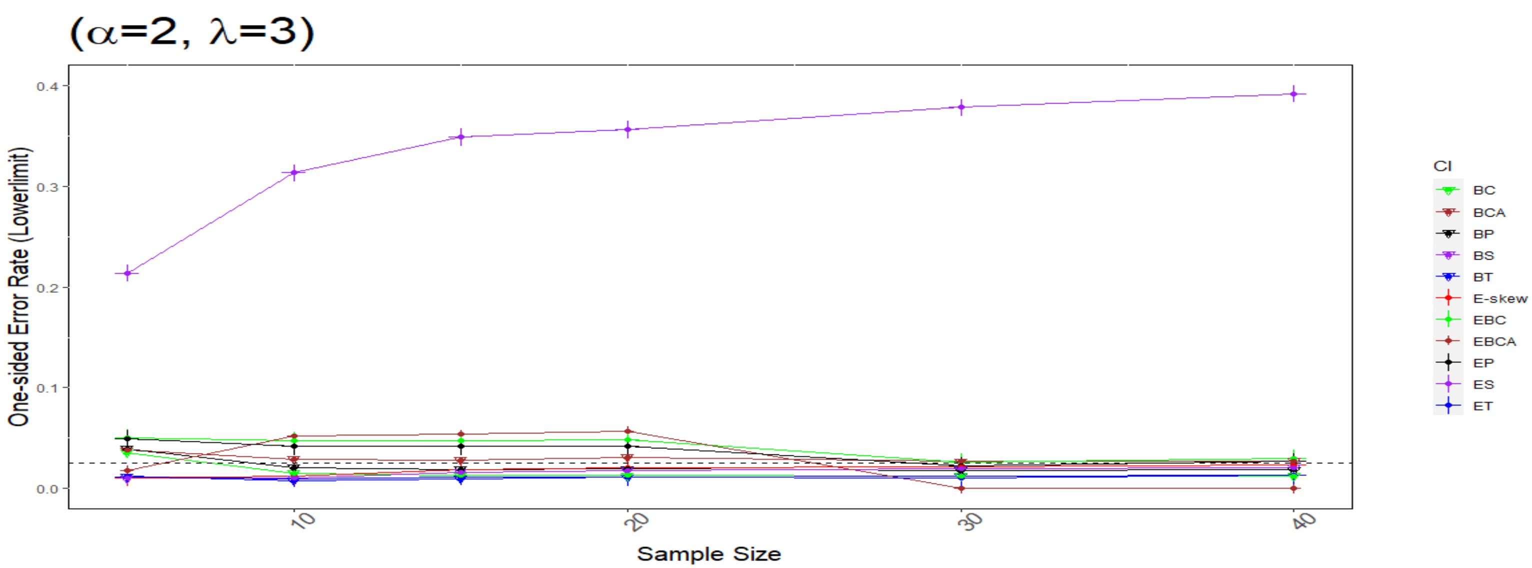
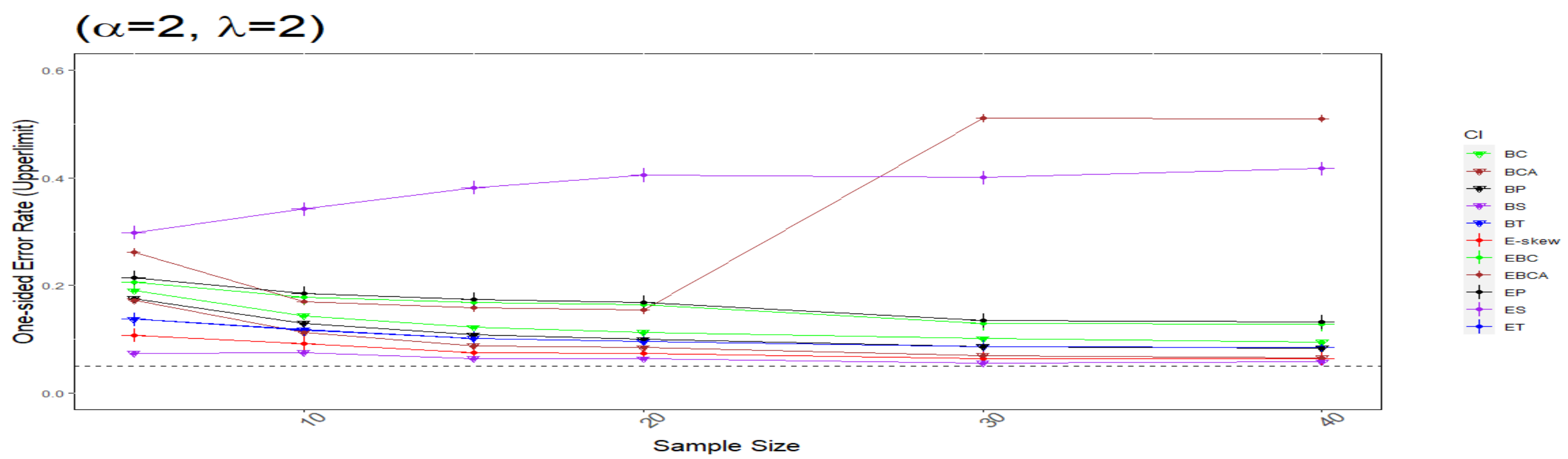


Figure: Sample Mean - GU90 - One-Sided Upperlimit Error Rates for 90% CI for the Gamma Distribution



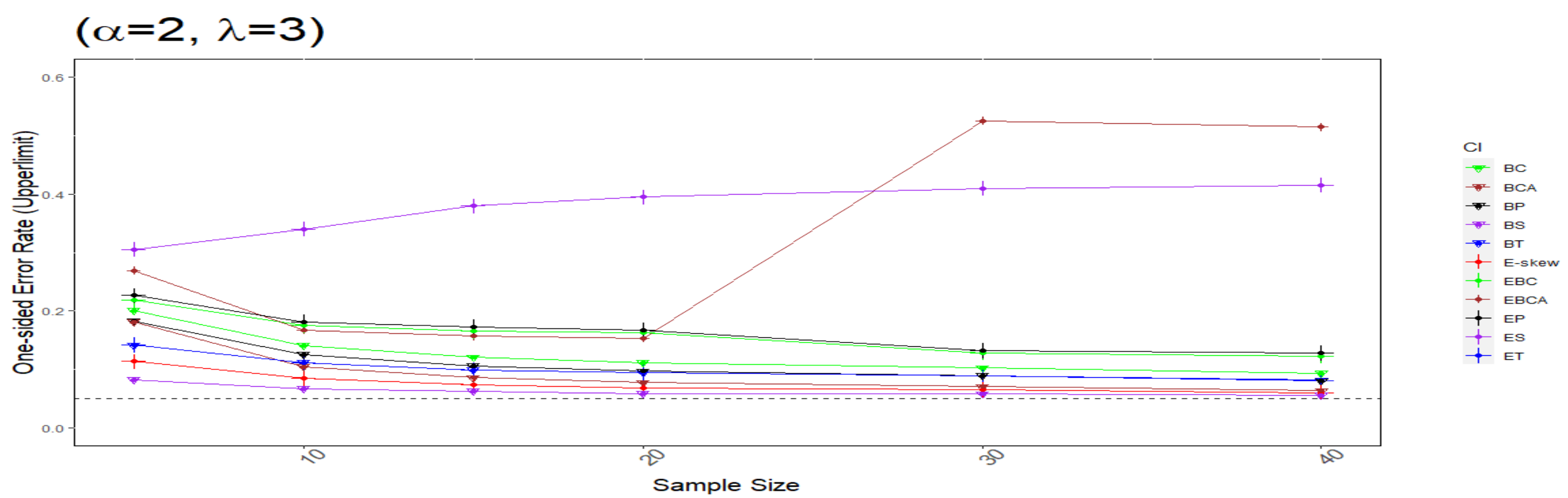
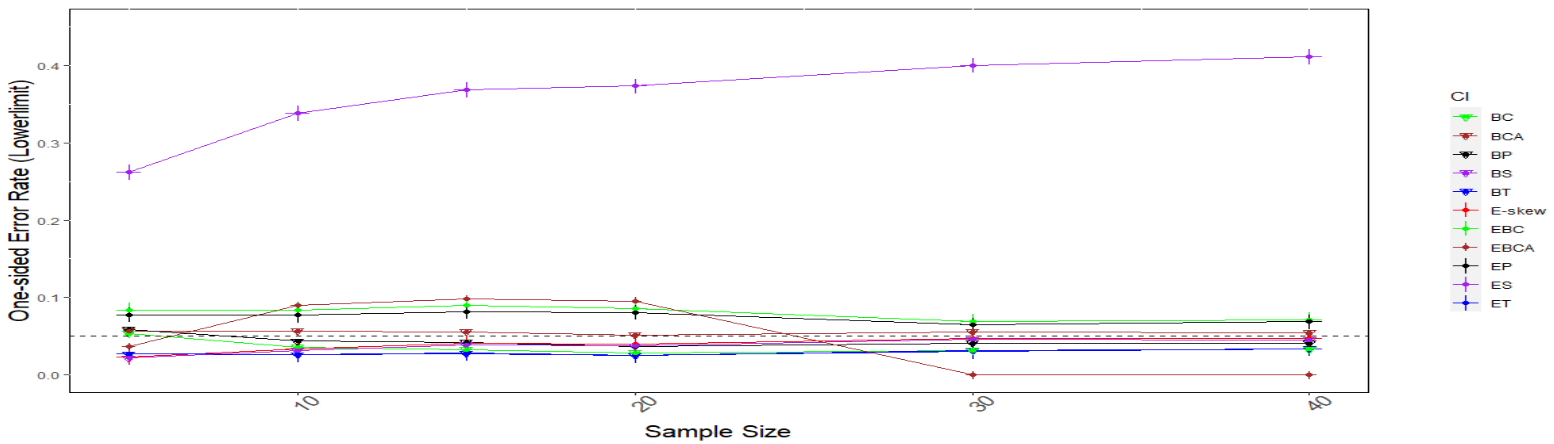
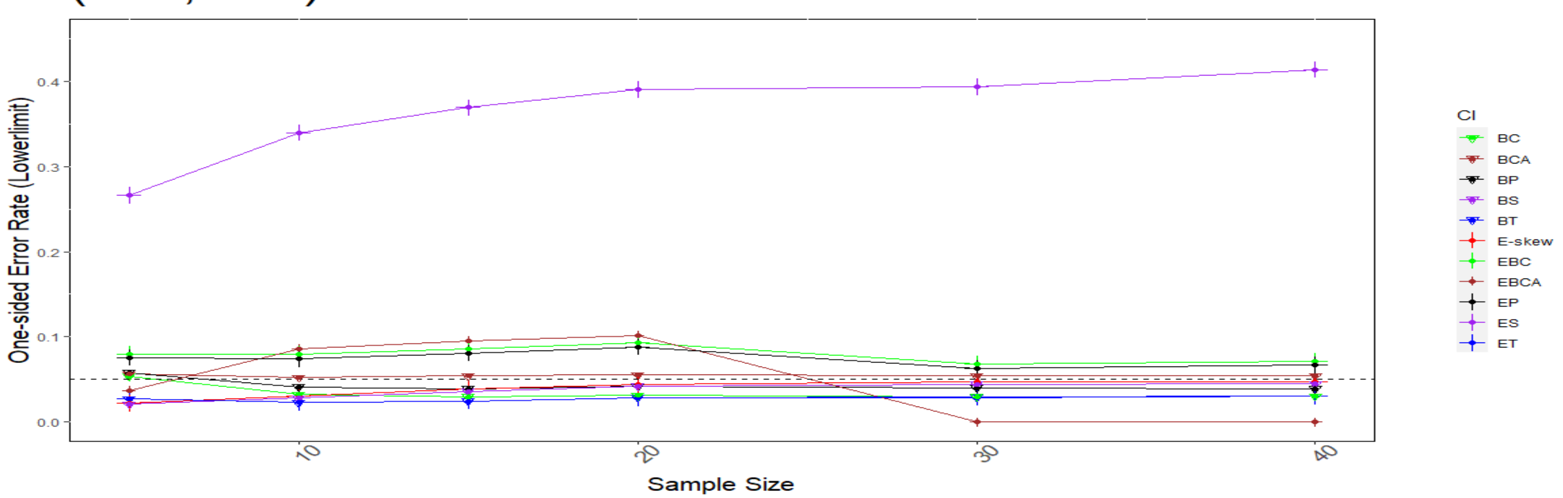


Figure: Sample Mean - GL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Gamma Distribution

(α=2, λ=2)



(α=2, λ=3)



d. Log-normal Distribution

The purpose in this sub section is to compare E-skew error rates and error rates for other methods using EBSD(n) to Monte Carlo Bootstrap method error rates for the sample mean statistic when data is log-normally distributed. The first comparison is performed using data generated from a log-normal distribution with $\mu = 4$ and σ = 0.2. Two separate log-normal distributions were studied, however because of the volume of error rate results in this section, only the results for the log-normal($\mu = 4$, σ = 0.2) at the α = 0.01 significance level are displayed in the tables LN1U99 and LN1L99 below on pages 109 and 110. Detailed numerical results for simulations not included in the tables can be viewed in Appendix. Although the tables only report results for one specifications, the log-normal($\mu = 4$, σ = 3) results can be viewed visually in figures LNU99, LNL99, LNU95, LNL95, LNU90 and LNL90 on pages 113-118.

For the log-normal($\mu = 4$, σ = 0.2) parameter specification at the specified α = 0.01, α = 0.05 and α = 0.10 significance levels, error rate results for E-skew were very similar to what they were for data generated from the exponential distribution. E-skew performed relatively less accurately at the α = 0.01 compared to other methods applied on EBSD(n) than it did at larger α significance levels. For the upper limit at the α = 0.01 significance level, E-skew had the error rate with the second smallest percent error at sample size 40. For the lower limit at the α = 0.01 significance level, E-skew had the error rate with the smallest percent error at sample size 15 and sample size 20 and the second smallest percent error at sample sizes 10 and 40.

At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew had the error rate with the second smallest percenter error at each sample size and a smaller percent error compared to any other method used on EBSD(n) at each sample size. For the lower limit at the α = 0.05 significance level, it had the error rate with the smallest percent error at sample size 40. For the lower limit at the α = 0.10 significance level, it had the error rate with the smallest percent error at sample size 20. Also at the α = 0.05 significance level for the lower limit, E-skew had the error rate with the second smallest percent error at four other sample sizes and was the method with smallest percent error among all methods using EBSD(n) at each of these sample sizes.

By comparison the EBC/EP/ EBC_a/ET all were relatively more accurate compared to E-skew and Monte Carlo Bootstrap methods at the α = 0.01 significance level than they were at larger α levels. At the α = 0.01 significance level for the upper limit, EBC_a had the error rate with the smallest percent error among any method at sample size 20. Also at this significance level for the upper limit, EBC had the error rate with the smallest percent error at sample sizes 15 and 30. Additionally, at the α = 0.01 significance level for the lower limit, EBC_a had the error rate with the smallest percent error at sample size 10. Otherwise no method other than E-skew using EBSD(n) had and error rate with the smallest percent error at any sample size larger than sample size 5 at any significance level for either limit end.

Table: Sample Mean - LN1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Log-Normal Distribution, log-normal($\boldsymbol{\mu}$ = 4, $\boldsymbol{\sigma}$ = 0.2)**, Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0155 (210%)	0.0113 (126%)	0.0115 (130%)	0.0073 (46%)	0.0074 (48%)	0.0065 (30%)
BT	0.0113 (126%)	0.0127 (154%)	0.0138 (176%)	0.0104 (108%)	0.0092 (84%)	0.0093 (86%)
ET	0.0112 (124%)	0.0127 (154%)	0.0141 (182%)	0.0105 (110%)	0.009 (80%)	0.0098 (96%)
BC	0.076 (1420%)	0.0372 (644%)	0.0289 (478%)	0.0185 (270%)	0.0135 (170%)	0.0135 (170%)
EBC	0.0503 (906%)	0.0169 (238%)	**0.0065** **(30%)**	0.006 (20%)	**0.0042** **(16%)**	0.0077 (54%)
BP	0.0682 (1264%)	0.0324 (548%)	0.0239 (378%)	0.0148 (196%)	0.0106 (112%)	0.0113 (126%)
EP	0.0438 (776%)	0.0175 (250%)	0.008 (60%)	0.0087 (74%)	0.0063 (26%)	0.0098 (96%)
BS	**0.0091** **(82%)**	**0.0089** **(78%)**	0.0096 (92%)	0.0066 (32%)	0.0068 (36%)	**0.0058** **(16%)**
ES	0.1554 (3008%)	0.2707 (5314%)	0.3017 (5934%)	0.3202 (6304%)	0.3052 (6004%)	0.333 (6560%)
$\boldsymbol{BC_a}$	0.0691 (1282%)	0.0301 (502%)	0.0229 (358%)	0.0127 (154%)	0.0102 (104%)	0.0094 (88%)
$\boldsymbol{EBC_a}$	0.0702 (1304%)	0.0105 (110%)	0.0072 (44%)	**0.0041** **(18%)**	0.2014 (3928%)	0.2005 (3910%)

Table: Sample Mean - LN1L99 Lower limit error rate (**α** = 0.01), Log-Normal Distribution, log-normal(**μ** = 4, **σ** = 0.2)**, Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0073 (46%)	0.0055 (10%)	**0.005 (0%)**	**0.0056 (12%)**	0.004 (20%)	0.0049 (2%)
BT	0.0036 (28%)	0.0035 (30%)	0.0025 (50%)	0.0035 (30%)	0.002 (60%)	0.0035 (30%)
ET	**0.0037 (26%)**	0.0036 (28%)	0.0024 (52%)	0.0033 (34%)	0.002 (60%)	0.0034 (32%)
BC	0.0364 (628%)	0.0113 (126%)	0.0065 (30%)	0.0058 (16%)	0.0033 (34%)	0.0043 (14%)
EBC	0.0225 (350%)	0.0059 (18%)	0.0016 (68%)	0.0033 (34%)	0.0016 (68%)	0.0034 (32%)
BP	0.0374 (648%)	0.0132 (164%)	0.0084 (68%)	0.0066 (32%)	**0.0046 (8%)**	**0.005 (0%)**
EP	0.0221 (342%)	0.0066 (32%)	9e-04 (82%)	0.0017 (66%)	0.001 (80%)	0.003 (40%)
BS	0.0035 (30%)	0.0043 (14%)	0.0036 (28%)	0.0043 (14%)	0.0031 (38%)	0.0047 (6%)
ES	0.1535 (2970%)	0.2572 (5044%)	0.3037 (5974%)	0.3394 (6688%)	0.3231 (6362%)	0.3499 (6898%)
BC_a	0.0364 (628%)	0.0153 (206%)	0.0112 (124%)	0.0095 (90%)	0.0071 (42%)	0.0066 (32%)
EBC_a	0.0221 (342%)	**0.0052 (4%)**	0.004 (20%)	0.0038 (24%)	0 (100%)	0 (100%)

Another set of results are presented for the log-normal distribution where σ is increased from 0.2 to 3. The reason to discuss another set of results for this distribution is the theoretical skewness of the log-normal distribution is determined as a function of σ. The function is increasing; skewness = $(e^{\sigma^2} + 2) * \sqrt{e^{\sigma^2} - 1}$, this increases the role of skewness in this data simulation from approximately 0.6 to more than 720,000. The purpose of this example is to observe an extreme theoretical situation that would likely not be observed in practice.

For the upper limit at the α = 0.01 significance level, E-skew attained the error rate with the third smallest percent error at sample sizes 40 for the upper limit behind both BS and BC_a when σ was increased from 0.2 to 3. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew attained the error rate with the third smallest percent error at sample sizes 20, 30 and 40. Despite the adjustments used for E-skew, the method could not adjust for skew as well as either second order Monte Carlo Bootstrap method. E-skew was still able to provide an error rate with a smaller percent error than any method implemented on EBSD(n) at these sample sizes and at the α = 0.05 and α = 0.10 significance levels.

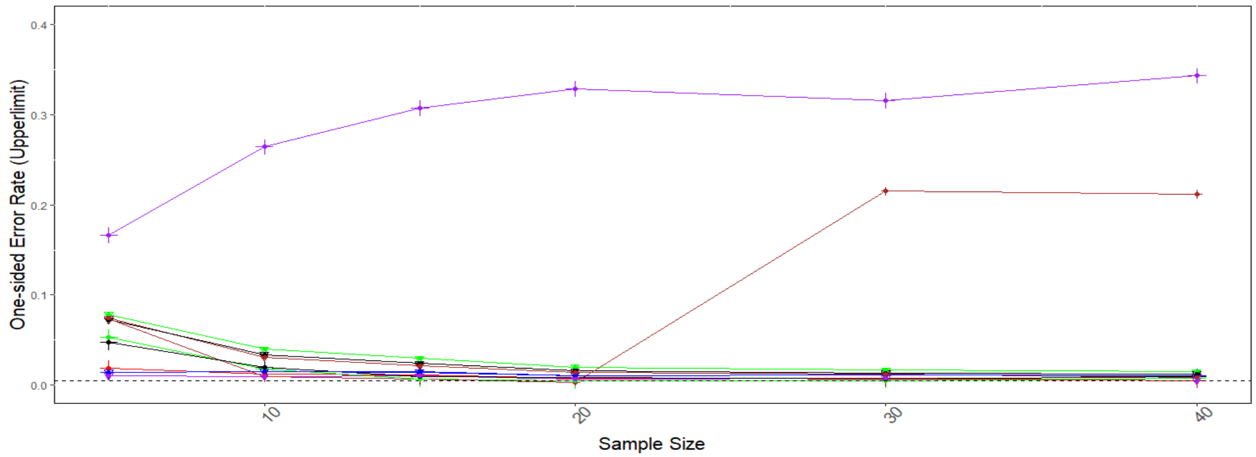
At the α = 0.01 significance level, E-skew performed inaccurately for the lower limit. E-skew had the error rate with the largest percent error rate at two separate sample sizes. However, at the α = 0.05 significance level for the lower limit, E-skew performed more accurately, attaining the error rate with the smallest percent error at sample sizes 10 and 15. Further, at the α = 0.10 significance level for the lower limit, E-skew performed very accurately attaining the error rate with the smallest percent error at five sample size.

When σ was increased from 0.2 to 3, the EBC_a method was again relatively more accurate at the α = 0.01 significance level than it was at larger α levels. At the α = 0.01 significance level for the upper limit EBC_a had the error rate with the second smallest percent error to BS for three separate sample sizes. At the α = 0.05 for the upper limit EBC_a method had the error rate with the smallest percent error of any method implemented on EBSD(n) at sample size 10. EBC_a's percent error was larger for every other sample size at each significance level considered for the upper limit when compared to E-skew.

Surprisingly, ES performed better for the upper limit at the α = 0.05 and α = 0.10 significance levels when compared to E-skew at sample sizes 5 and 10. Additionally, at the α = 0.01 level for the upper limit, ES had an error rate with a smaller percent error than any other method implemented on EBSD(n) at sample sizes 5 and 30. For the lower limit no method other than E-skew performed relatively accurately at any significance level for any of the sample size greater than sample size 5 considered.

Figure: Sample Mean - LNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Log-Normal Distribution

(μ=4, σ=0.2)



(μ=4, σ=3)

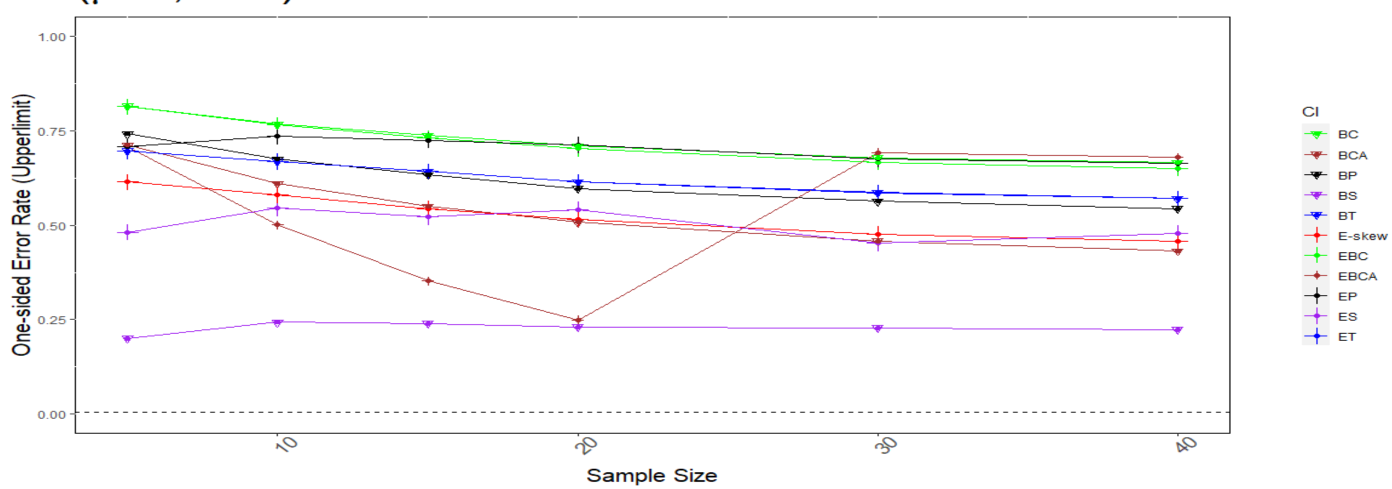


Figure: Sample Mean - LNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Log-Normal Distribution

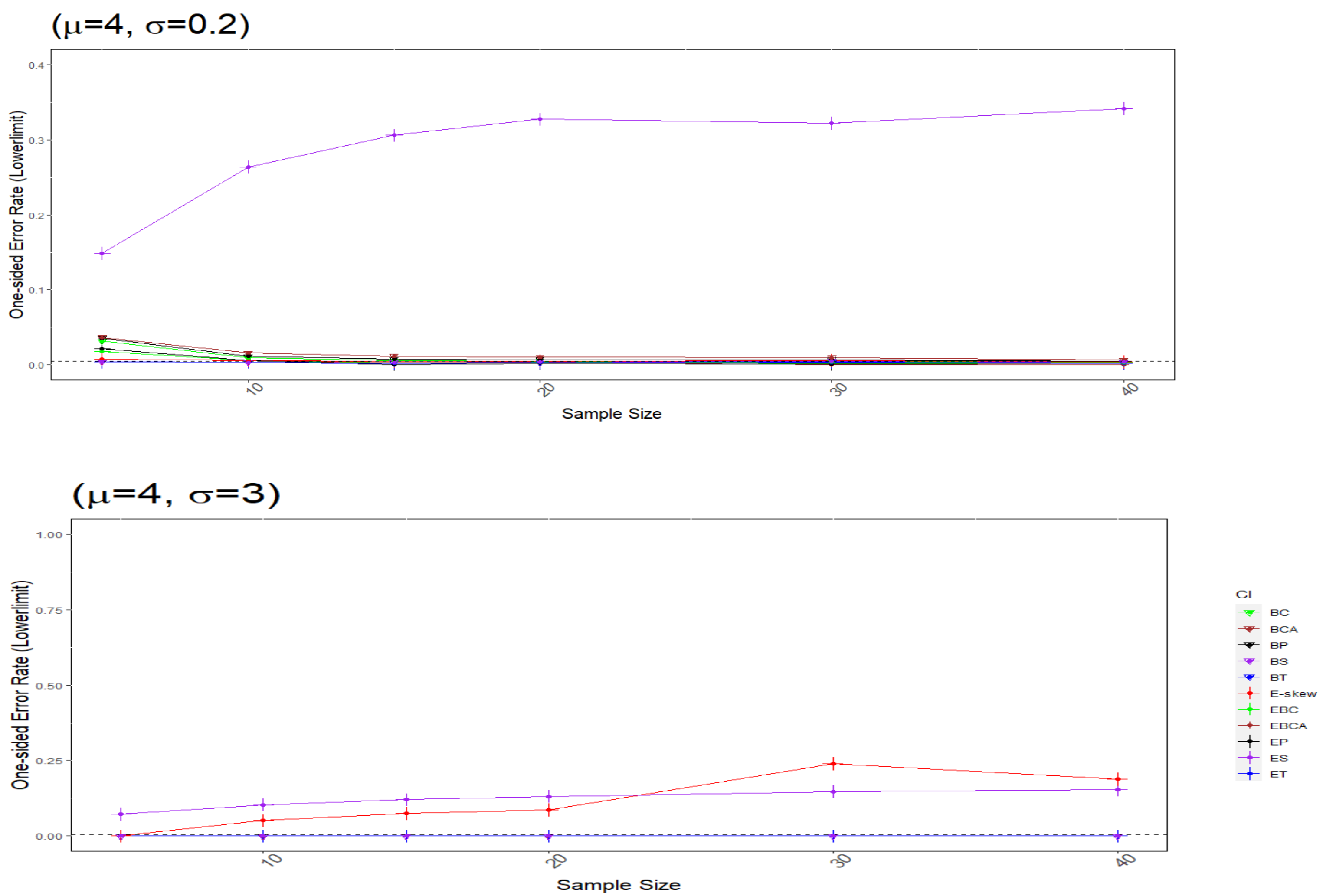


Figure: Sample Mean - LNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Log-Normal Distribution

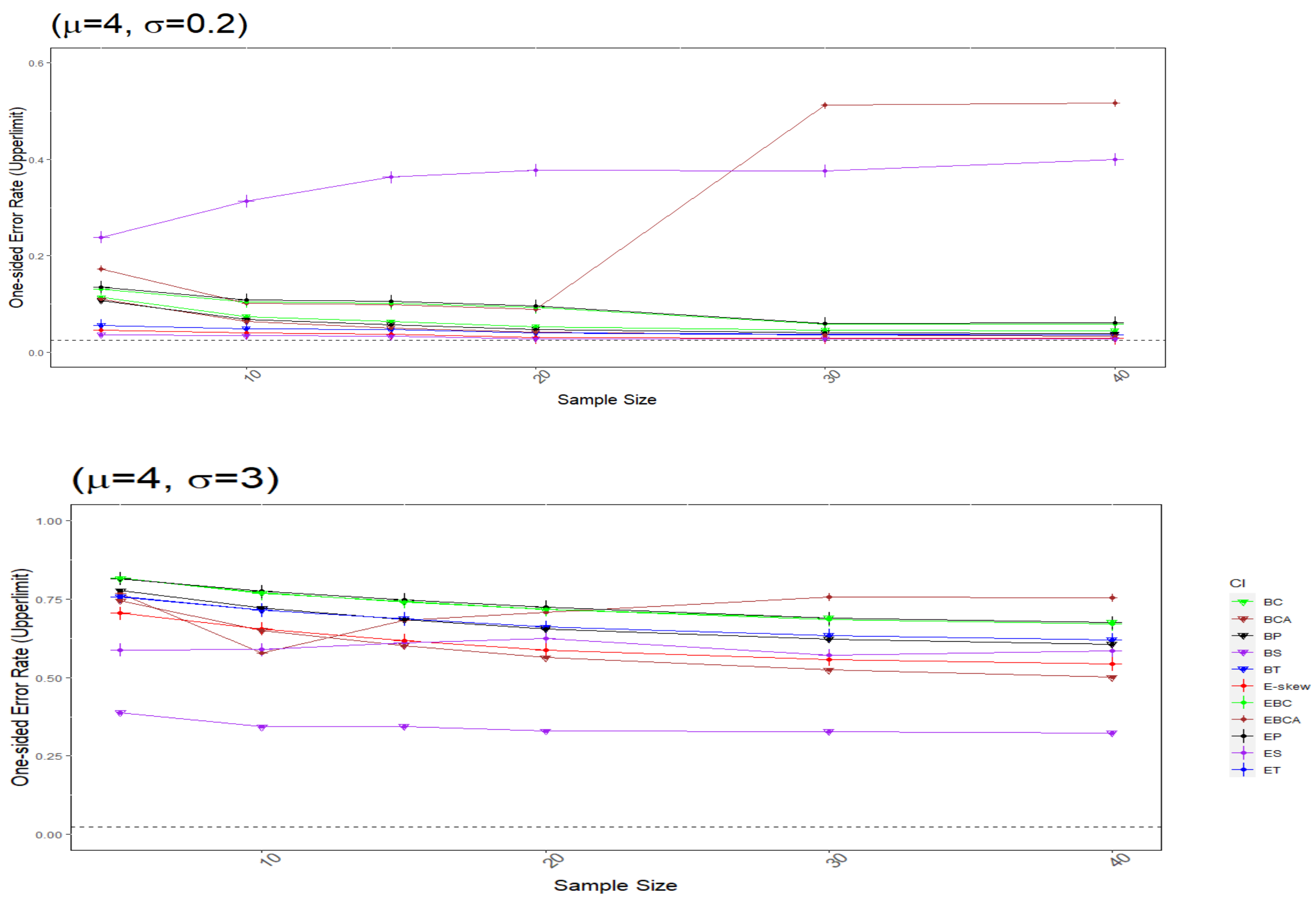


Figure: Sample Mean - LNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Log-Normal Distribution

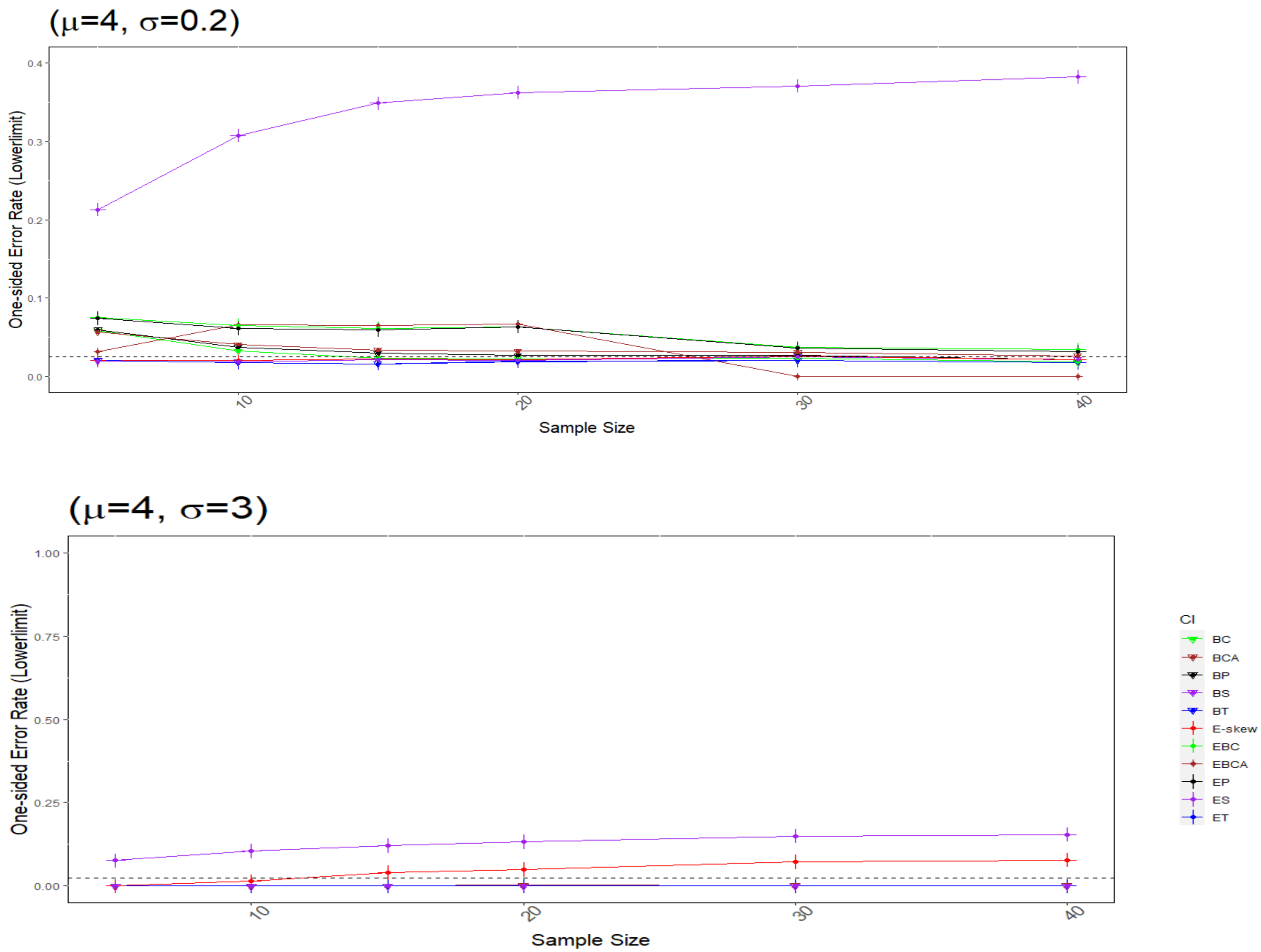
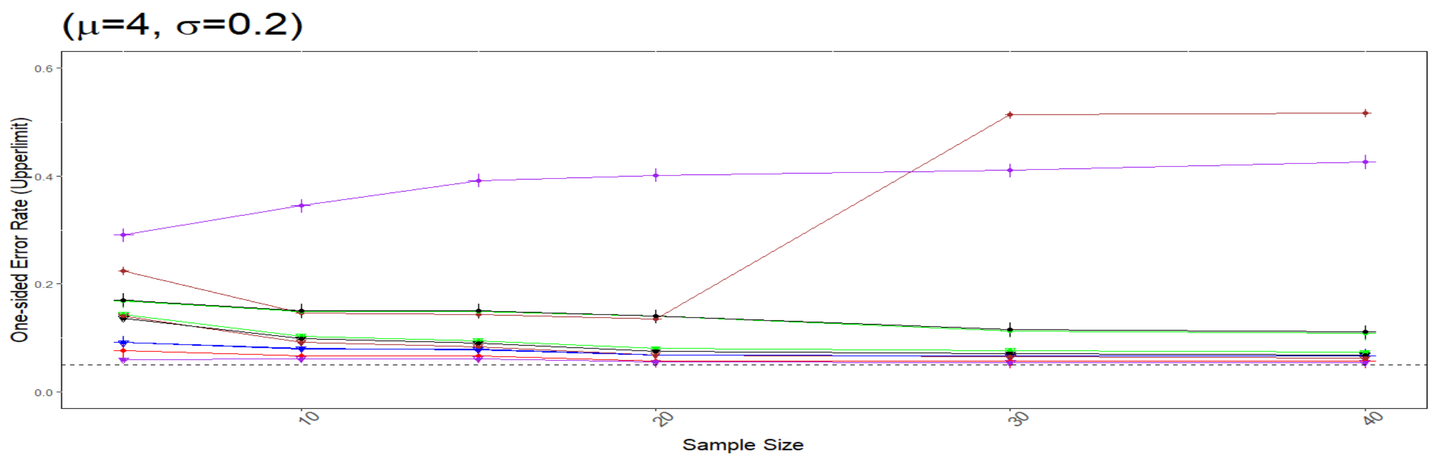


Figure: Sample Mean - LNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Log-Normal Distribution



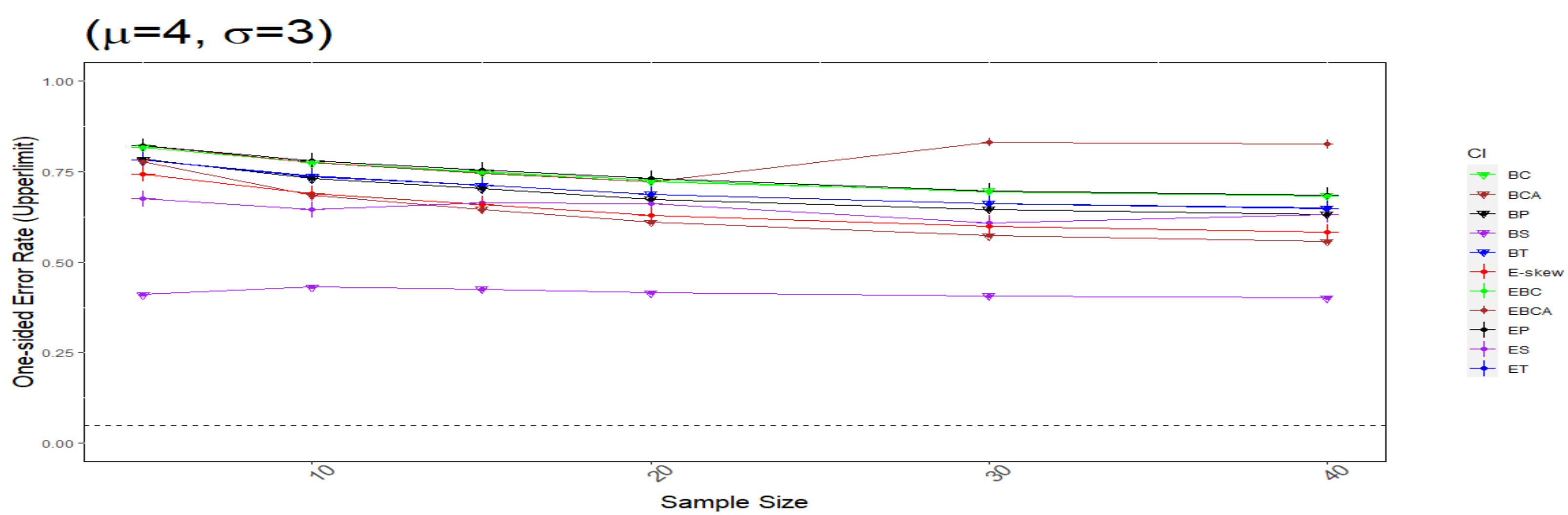
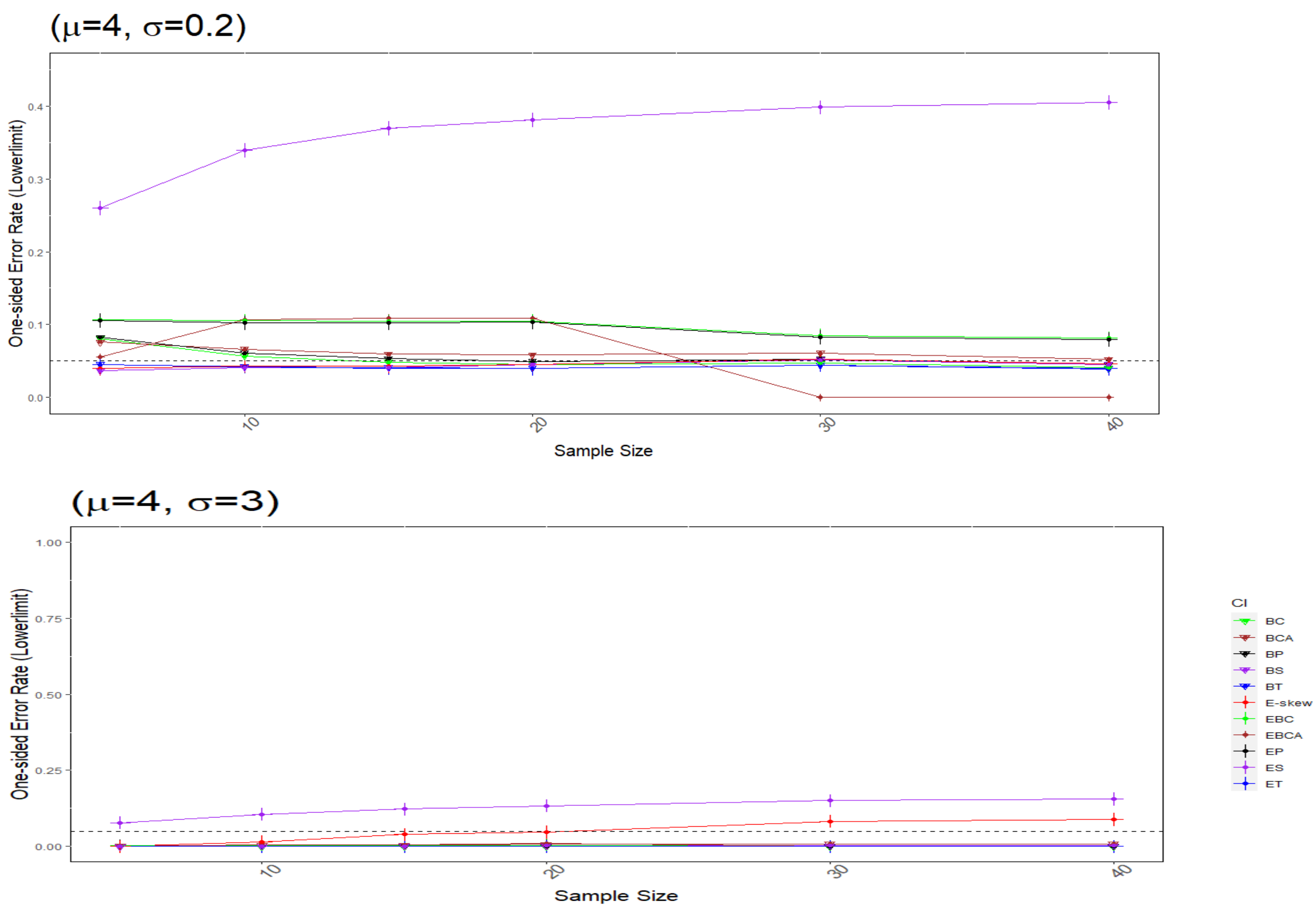


Figure: Sample Mean - LNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Log-Normal Distribution



e. Mixture of Two Normal Distributions

The purpose here is to compare E-skew error rates and error rates for other methods using EBSD(n) to Monte Carlo Bootstrap method error rates for the sample mean statistic when data is generated from a mixture of two normal distributions. For the first comparison below, data was generated from a mixture of normal distributions with specification 0.6*$N(\mu_1 = 4$, $\sigma_1 = 4)$ + 0.4*$N(\mu_1 = 8$, $\sigma_1 = 8)$. The error results at the α = 0.01 significance level are displayed in tables MN1U99, and MN1L99 in pages 121 and 122 below. Two separate mixture of normal distributions were studied, however because of the volume of error rate results, only the results for the 0.6*$N(\mu_1 = 4$, $\sigma_1 = 4)$ + 0.4*$N(\mu_1 = 8$, $\sigma_1 = 8)$ specification at the α = 0.01 significance level are displayed in the tables. Detailed numerical results for simulations not included here can be viewed in Appendix tables. Although these tables only report results for the 0.6*$N(\mu_1 = 4$, $\sigma_1 = 4)$ + 0.4*$N(\mu_1 = 8$, $\sigma_1 = 8)$ specification, the 0.8*$N(\mu_1 = 4$, $\sigma_1 = 4)$ + 0.2*$N(\mu_1 = 8$, $\sigma_1 = 8)$ specification results can be viewed visually in figures MNU99, MNL99, MNU95, MNL95, MNU90 and MNL90 on pages 125-130.

For the 0.6*$N(\mu_1 = 4$, $\sigma_1 = 4)$ + 0.4*$N(\mu_1 = 8$, $\sigma_1 = 8)$ parameter specification at the specified α = 0.01, α = 0.05 and α = 0.10 significance levels, error rate results for E-skew were very similar to what they were for data generated from the normal distribution. E-skew performed relatively less accurately at the α = 0.01 significance level, and relatively more accurately at the α = 0.05 and α = 0.10 significance levels when compared to both Monte Carlo Bootstrap and other EBSD(n) methods. For the upper limit at the α = 0.01 significance level, E-skew and BS both had the error rate with the smallest percent error at sample size 40. For the lower limit at the α = 0.01 significance level, E-skew had the error rate with the smallest percent error at sample sizes 5 and 15. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew had the error rate with the second smallest percent error at each sample size and a smaller percent error than any other method used on EBSD(n) at each sample size. For the lower limit, it had the error rate with the smallest percent error compared to any other method for four sample sizes at the α = 0.05 significance level and three sample sizes at the α = 0.10 significance level.

Several methods performed relatively more accurately using EBSD(n) at the α = 0.01 significance level than at larger α significance levels. For the upper limit at the α = 0.01 significance level, EBC had the error rate with the smallest percent error for two sample sizes and EP had the smallest at one sample size. However, at the α = 0.05 and α = 0.10 significance levels for the upper limit, neither method had the error rate with the smallest percent error at any sample size and neither method had the smallest percent error among methods applied on EBSD(n). For the lower limit, a similar finding occurred. For the lower limit at the α = 0.01 significance level, EBC_a was found to have the smallest percent error at sample size 20 and EP the smallest at sample size 10. However, none of EBC/EP/EBC_a/ET had an error rate with the smallest percent error for the lower limit for any sample size at either the α = 0.05 or α = 0.10 significance levels.

Table: Sample Mean - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.6*$N(\mu_1 = 4, \sigma_1 = 4)$ + 0.4*$N(\mu_2 = 8,$ $\sigma_2 = 8)$, **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0206 (312%)	0.0157 (214%)	0.0122 (144%)	0.0123 (146%)	0.0112 (124%)	**0.0073 (46%)**
BT	0.0138 (176%)	0.0132 (164%)	0.0102 (104%)	0.0119 (138%)	0.0107 (114%)	0.009 (80%)
ET	0.0141 (182%)	0.013 (160%)	0.0103 (106%)	0.0121 (142%)	0.0108 (116%)	0.0088 (76%)
BC	0.0711 (1322%)	0.0318 (536%)	0.0226 (352%)	0.0202 (304%)	0.0144 (188%)	0.0119 (138%)
EBC	0.0442 (784%)	0.0199 (298%)	**0.0075 (50%)**	**0.009 (80%)**	0.008 (60%)	0.0085 (70%)
BP	0.0751 (1402%)	0.0346 (592%)	0.0228 (356%)	0.0199 (298%)	0.0147 (194%)	0.0111 (122%)
EP	0.0506 (912%)	0.02 (300%)	0.0078 (56%)	0.0112 (124%)	**0.0079 (58%)**	0.0095 (90%)
BS	**0.0094 (88%)**	**0.0122 (144%)**	0.0105 (110%)	0.011 (120%)	0.011 (120%)	**0.0073 (46%)**
ES	0.1905 (3710%)	0.2743 (5386%)	0.2904 (5708%)	0.3186 (6272%)	0.3084 (6068%)	0.3313 (6526%)
BC_a	0.08 (1500%)	0.038 (660%)	0.024 (380%)	0.0197 (294%)	0.0164 (228%)	0.0106 (112%)
EBC_a	0.0829 (1558%)	0.0193 (286%)	0.0106 (112%)	0.0103 (106%)	0.2058 (4016%)	0.2081 (4062%)

Table: Sample Mean - MN1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Mixture of two Normal Distributions, 0.6*$\boldsymbol{N(\mu_1 = 4, \sigma_1 = 4)}$ + 0.4*$\boldsymbol{N(\mu_2 = 8, \sigma_2 = 8)}$, **Bootstraps=10000**						
Sample size	5	10	15	20	30	40
E-skew	**0.0073** **(46%)**	0.0059 (18%)	**0.0055** **(10%)**	0.0057 (14%)	0.0064 (28%)	0.0055 (10%)
BT	0.0025 (50%)	0.0026 (48%)	0.0013 (74%)	0.0017 (66%)	0.0024 (52%)	0.0029 (42%)
ET	0.0025 (50%)	0.0027 (46%)	0.0013 (74%)	0.0017 (66%)	0.0023 (54%)	0.0028 (44%)
BC	0.0264 (428%)	0.0093 (86%)	0.0037 (26%)	0.0043 (14%)	0.0036 (28%)	0.0034 (32%)
EBC	0.015 (200%)	0.0061 (22%)	0.0028 (44%)	0.0039 (22%)	0.0021 (58%)	0.0034 (32%)
BP	0.0352 (604%)	0.0132 (164%)	0.0069 (38%)	0.0079 (58%)	0.0066 (32%)	**0.0053** **(6%)**
EP	0.0185 (270%)	**0.0046** **(8%)**	9e-04 (82%)	9e-04 (82%)	8e-04 (84%)	0.0021 (58%)
BS	0.0022 (56%)	0.0036 (28%)	0.0029 (42%)	0.0044 (12%)	**0.0055** **(10%)**	**0.0053** **(6%)**
ES	0.1784 (3468%)	0.275 (5400%)	0.3087 (6074%)	0.3439 (6778%)	0.3343 (6586%)	0.3521 (6942%)
$\boldsymbol{BC_a}$	0.0375 (650%)	0.0182 (264%)	0.0114 (128%)	0.0123 (146%)	0.0104 (108%)	0.008 (60%)
$\boldsymbol{EBC_a}$	0.0185 (270%)	0.0056 (12%)	0.0036 (28%)	**0.0048** **(4%)**	0 (100%)	0 (100%)

Simulation were not only performed for the mixture of normal distribution with parameters: 0.6*$N(\mu_1 = 4, \sigma_1 = 4)$ + 0.4*$N(\mu_1 = 8, \sigma_1 = 8)$. Simulations were also performed for parameter specifications: 0.8*$N(\mu_1 = 4, \sigma_1 = 4)$ + 0.2*$N(\mu_1 = 8, \sigma_1 = 8)$. When considering a change in parameter specification (changing the values of p_1 and

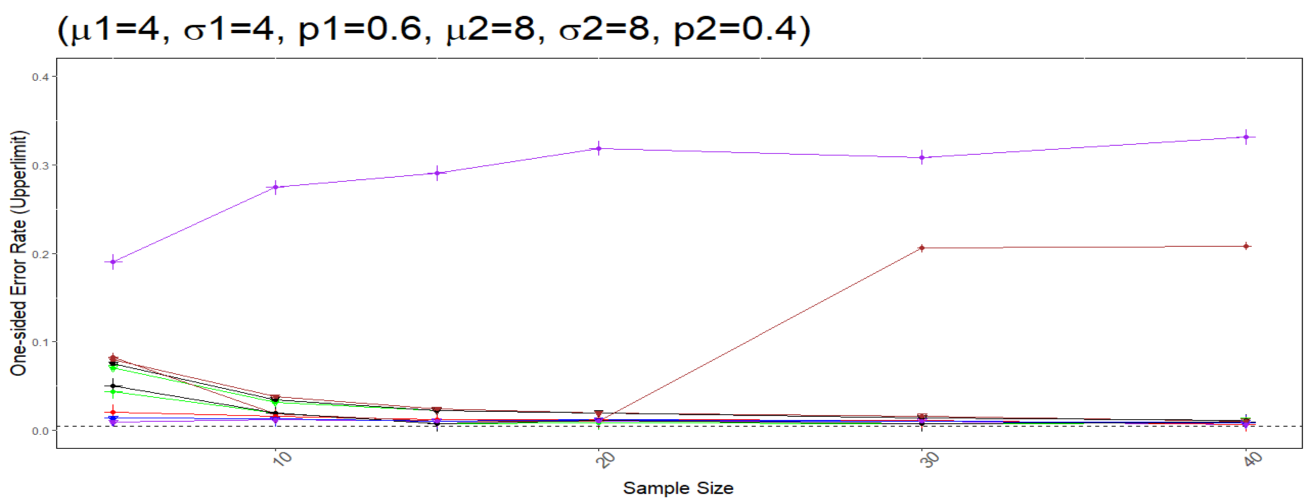
p_2) E-skew did not perform relatively as accurately compared to the remaining methods studied at the α = 0.01 significance level for the upper limit as it did for the first specification studied. When p_1 was increased from 0.6 to 0.8, for the upper limit, E-skew did not have the error rate with the smallest percent error at any sample size for the upper limit at the α = 0.01 significance level. It also did not have the error rate with the smallest percent error among methods applied on EBSD(n) at any sample size.

However, for the upper limit at the α = 0.05 significance level, E-skew did perform relatively more accurately relative to other methods applied on EBSD(n). E-skew had the error rate with the smallest percent error among methods applied on EBSD(n) and the error rate with the second smallest percent error overall at four of the six sample sizes considered. At the α = 0.10 significance level for the upper limit, it had the error rate with the smallest percent error among all methods applied on EBSD(n) and the second smallest percent error overall for all six sample sizes.

E-skew also did not perform as accurately relative to the Monte Carlo Bootstrap methods studied for this parameter specification for the lower limit at the α = 0.01 and α = 0.05 significance levels. E-skew did not attain the error rate with the smallest percent error at any sample size and only attained the second smallest at sample size 10 at the α = 0.05 significance level. At the α = 0.10 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at four sample sizes among methods applied on EBSD(n).

The EP method performed more accurately at the α = 0.01 significance level for the upper limit relative to the first parameter specification studied. At this significance level for the upper limit, EP had the error rate with the smallest percent error at four sample sizes. For the lower limit, EP also had the smallest percent error among methods applied on EBSD(n) at this significance level. Further for the α = 0.01 significance level for the lower limit, EBC_a had the error rate with the smallest percent error among methods applied on EBSD(n) at one sample size. At the α = 0.05 significance levels for the upper limit, ET had the error rate with the smallest percent error among EBSD(n) methods at sample size 10. For the lower limit, ET had the error rate with the smallest percent error among EBSD(n) methods at sample size 5 at the α = 0.05 and α = 0.10 significance levels. No method other than E-skew applied on EBSD(n) achieved an error rate with the smallest percent error at any other sample size for either the upper or lower limit at either the α = 0.05 or α = 0.10 significance levels.

Figure: Sample Mean - MNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Mixture of two Normals



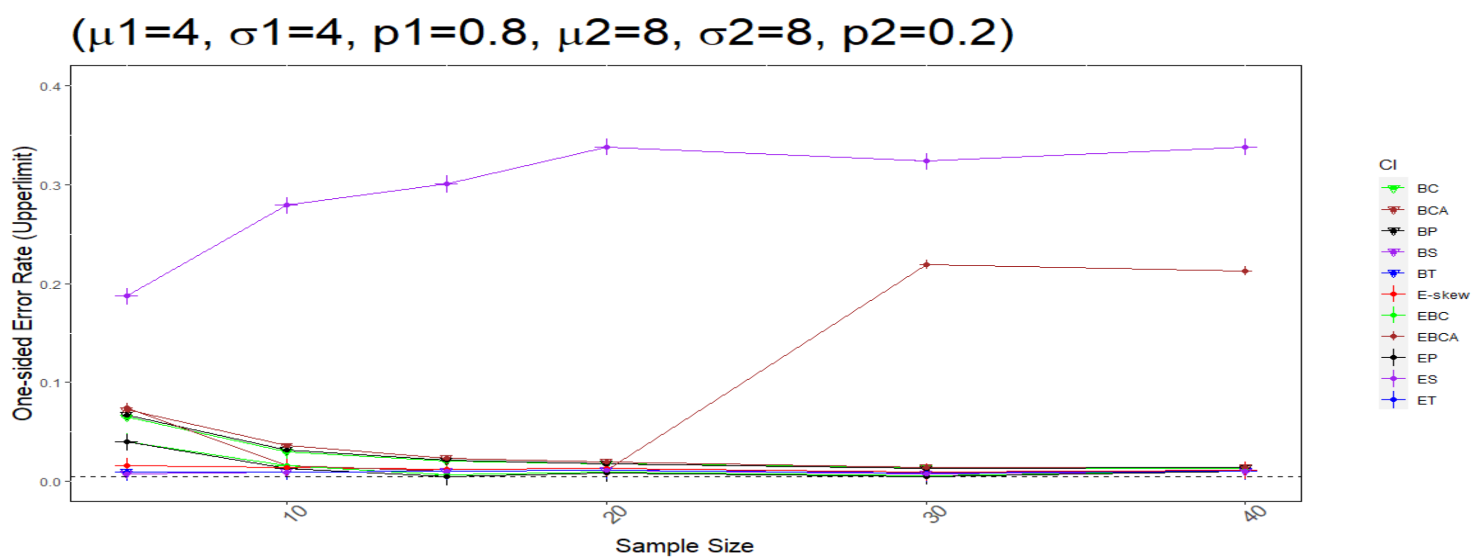
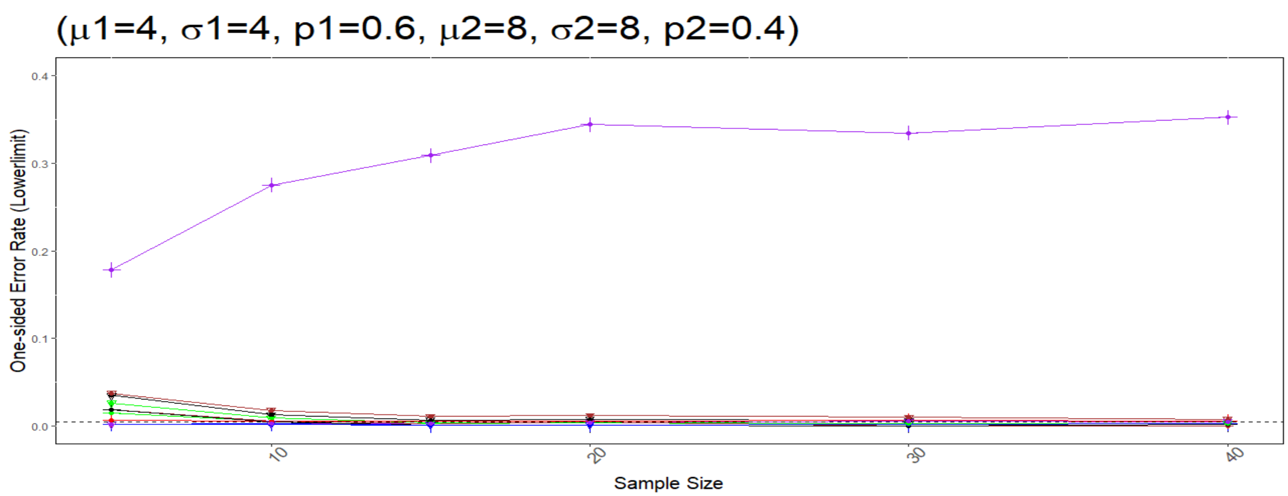


Figure: Sample Mean - MNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Mixture of two Normals



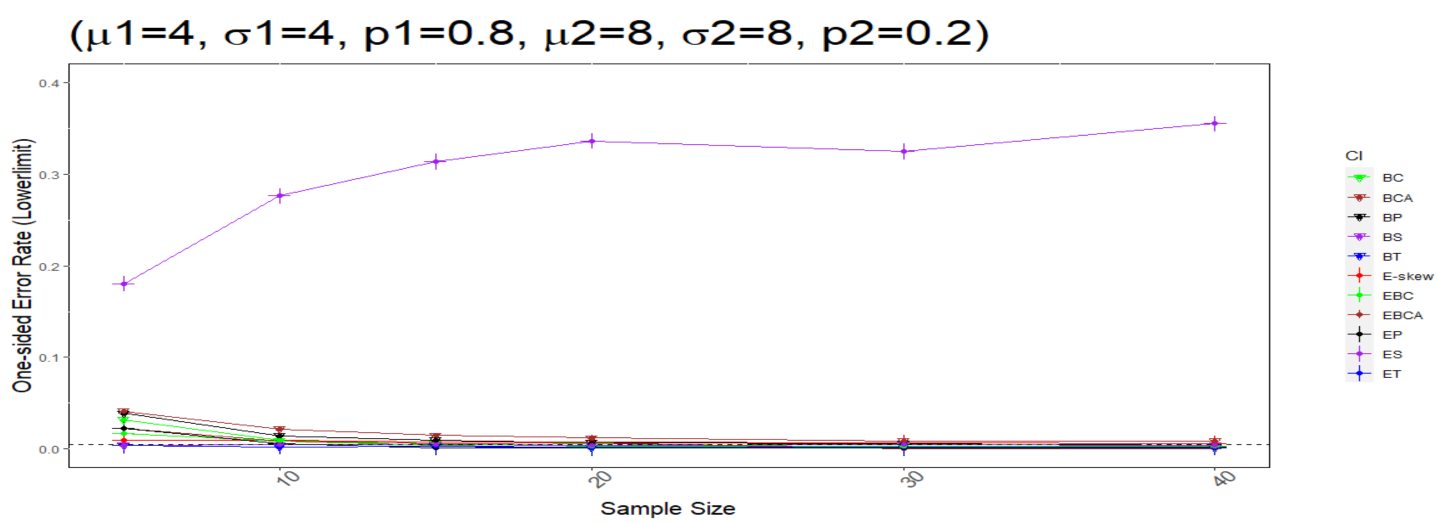
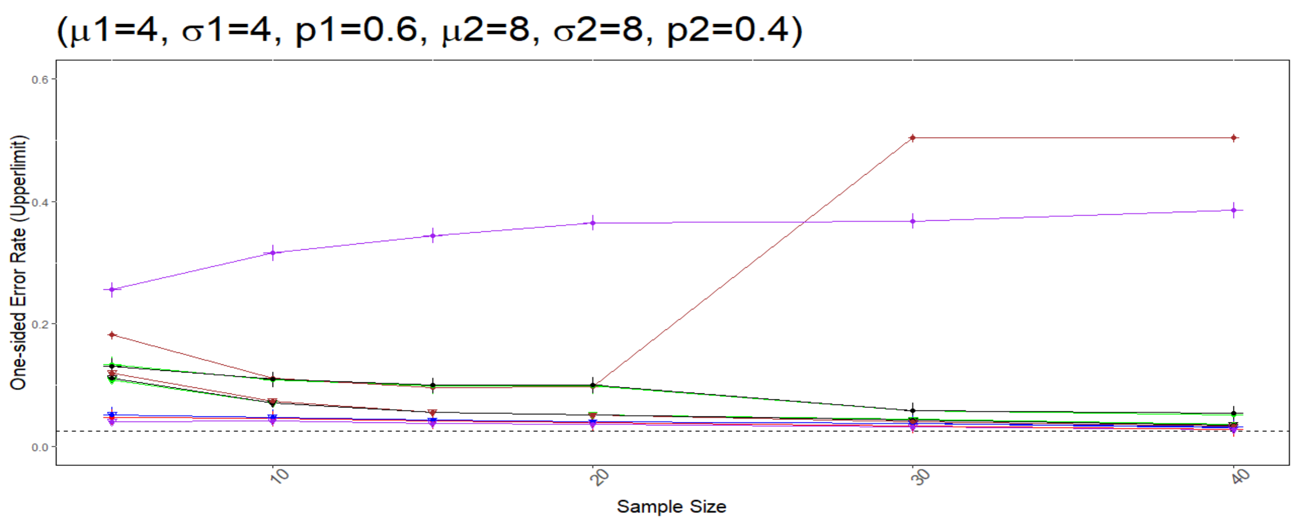


Figure: Sample Mean - MNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Mixture of two Normals



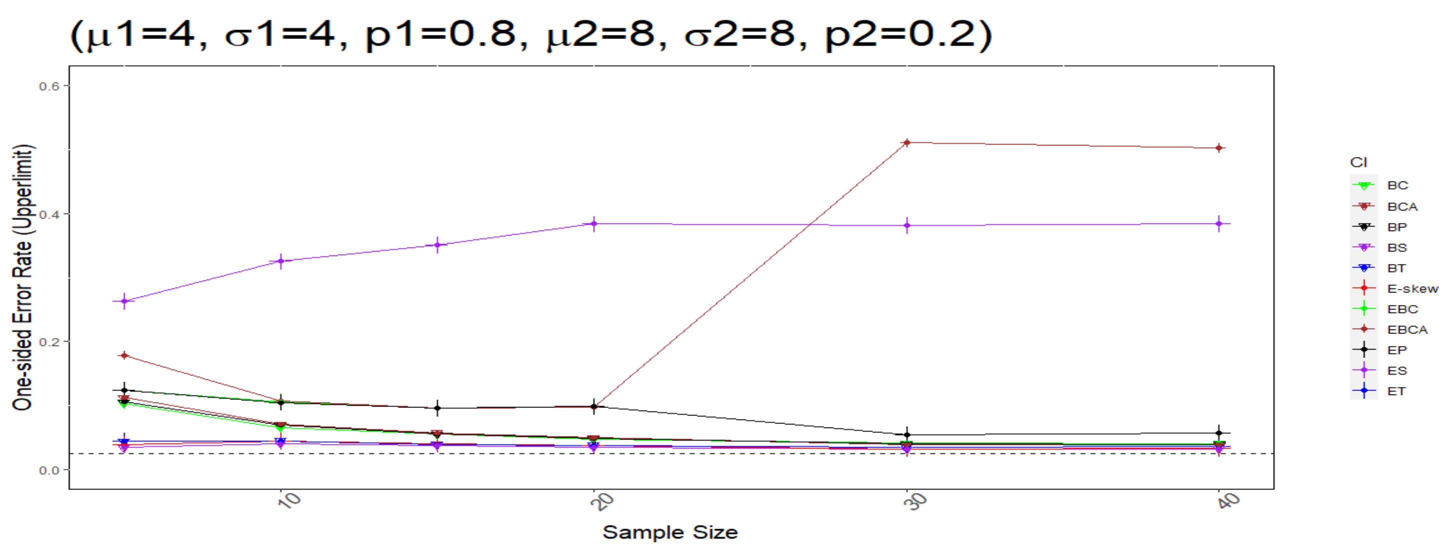


Figure: Sample Mean - MNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Mixture of two Normals

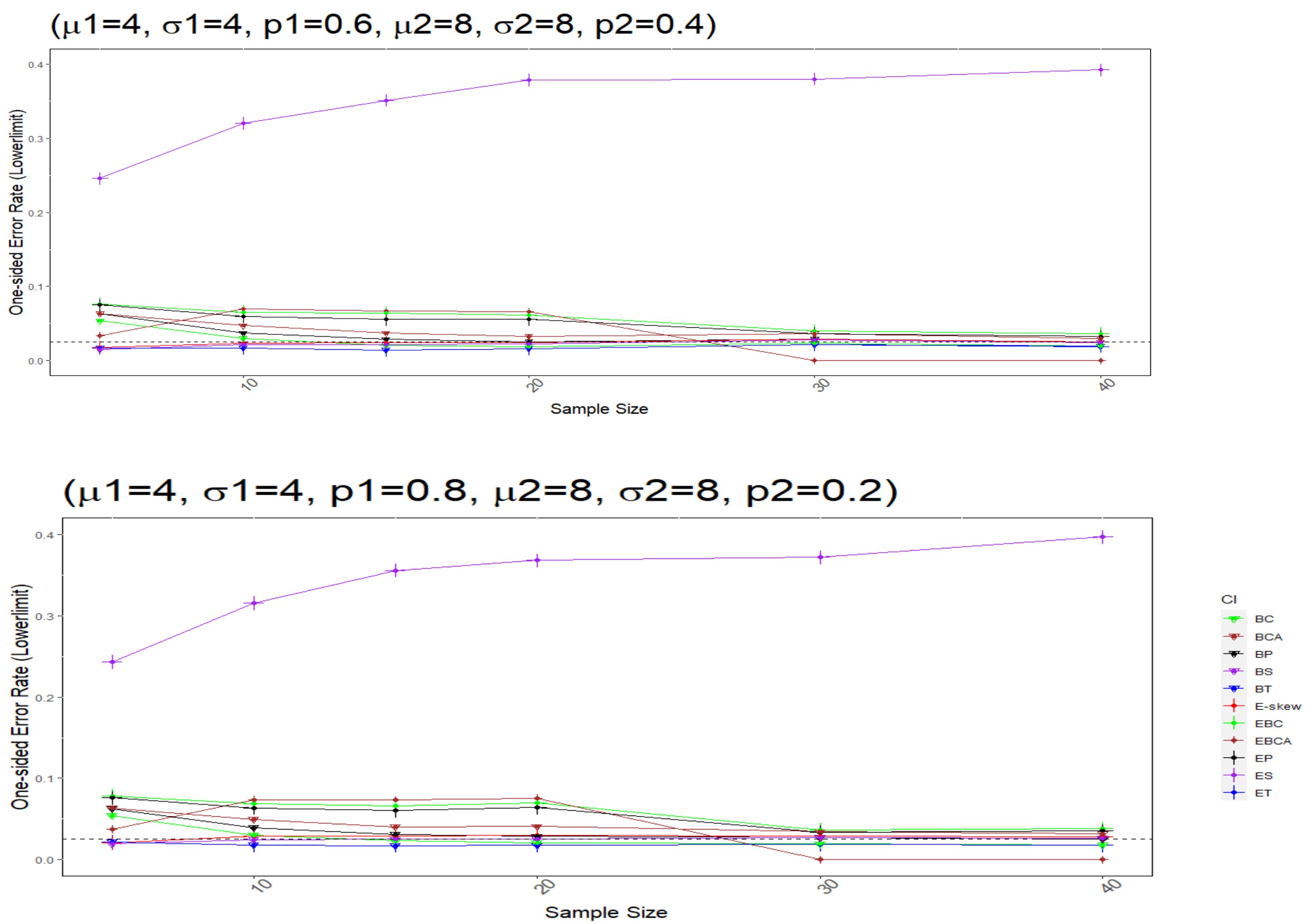
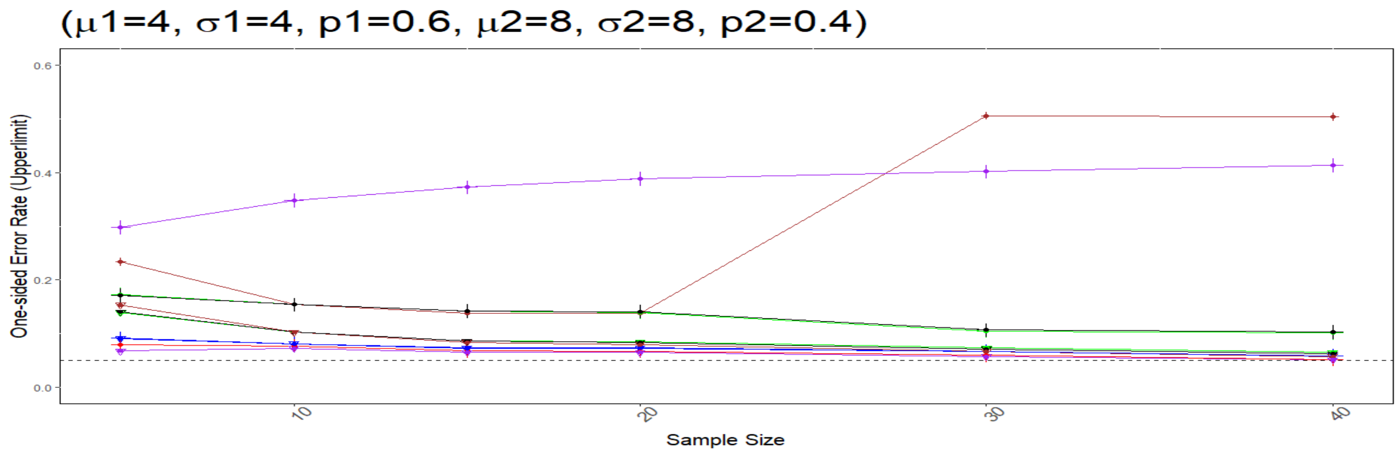


Figure: Sample Mean - MNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Mixture of two Normals



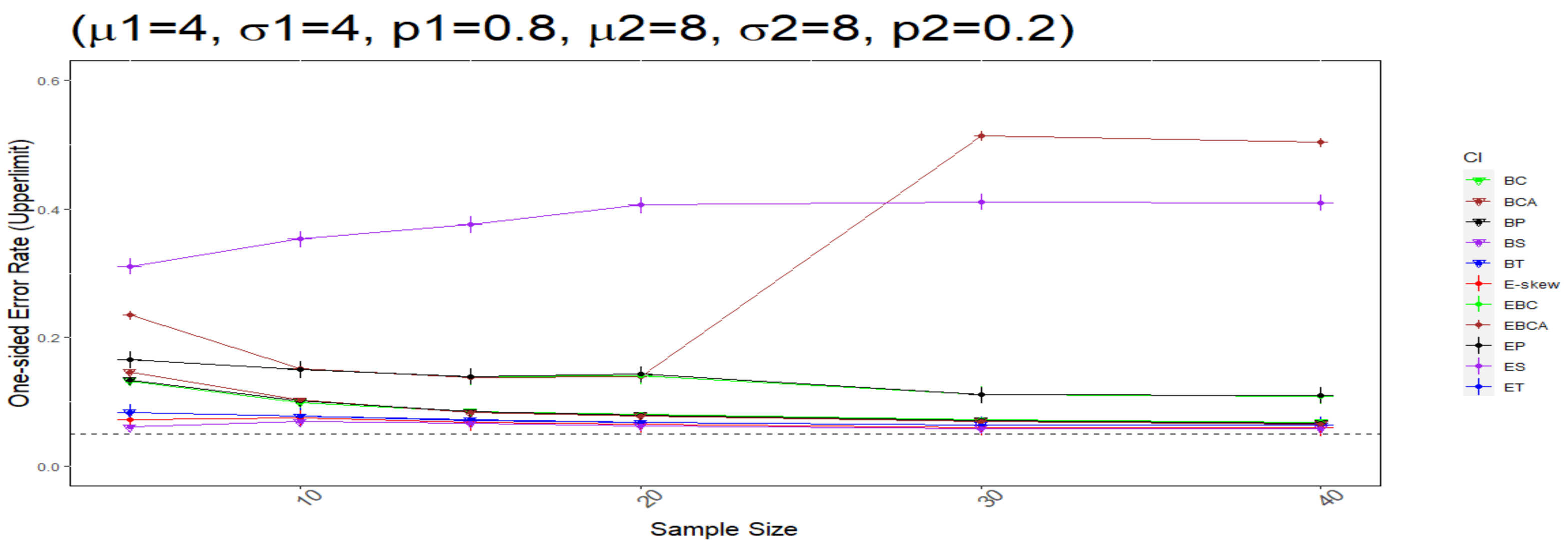
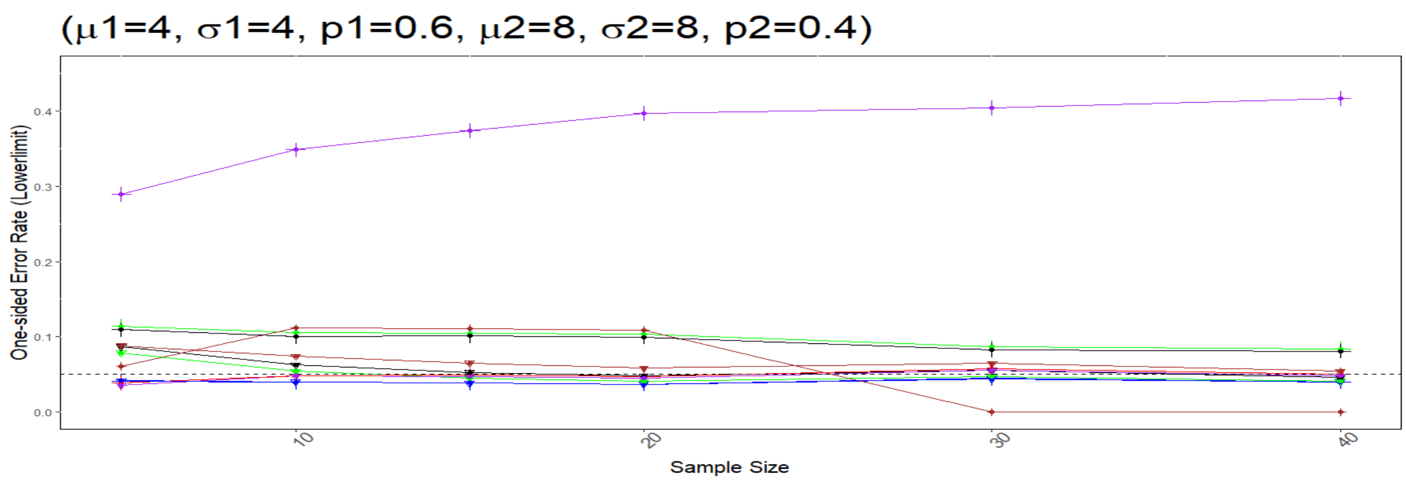
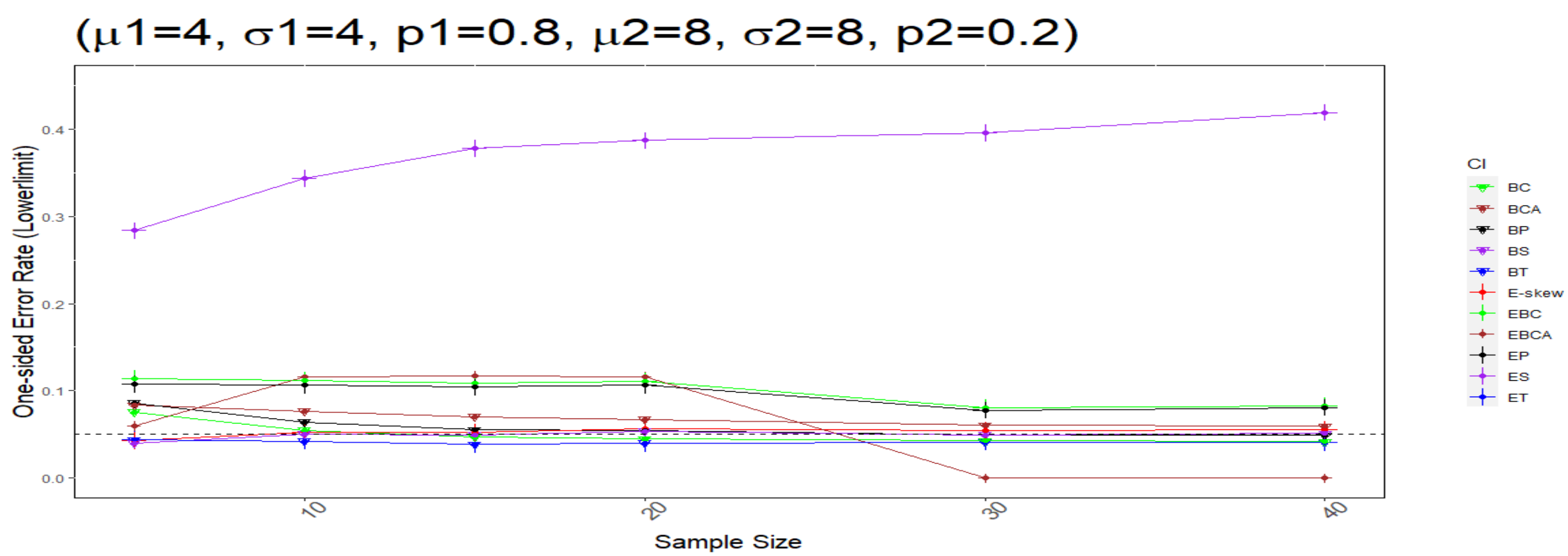


Figure: Sample Mean - MNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Mixture of two Normals





Sample Mean Results Discussion for E-skew and other methods using EBSD(n)

In general, the E-skew method yielded error rates for the sample mean with smaller percent errors compared to any other method applied on EBSD(n) when considering all three significance levels studied. Additionally, E-skew was more accurate compared to all Monte Carlo Bootstrap and EBSD(n) methods in some instances for normally distributed data for both the upper and lower limit. When data was generated from skewed distributions, E-skew performed relatively more accurately for the upper limit compared to all other methods applied on EBSD(n) at the α = 0.05 and α = 0.10 significance levels. In general, E-skew performed the most accurately in comparison to all other methods at the α = 0.10 significance level.

For moderately skewed data, like data generated from the exponential distribution, E-skew outperformed every other method implemented on EBSD(n) at the α = 0.05 and α = 0.10 significance levels for the upper limit. E-skew consistently performed more accurately for the upper limit at these significance levels compared to every method studied other than BS. At the α = 0.01 significance level for the upper limit EBC_a performed more accurately than E-skew at small to moderate sample sizes ($n =$ 10, 15, 20) but then it's error rate percent error was much larger than E-skew's at sample sizes 30 and 40. For the lower limit for moderately skewed data E-skew and BC_a/ABC both obtained the error rate with the smallest percent error depending on the simulation run and the sample size at each significance level.

For substantially skewed data like in the case of the second log-normal distribution example, E-skew did not perform as accurately. Particularly at the α = 0.01

significance level for the lower limit, E-skew was among the least accurate methods tested. However even in this case, E-skew performed accurately at the α = 0.05 and α = 0.10 significance levels for the lower limit, when compared not only to other methods applied on EBSD(n) but to Monte Carlo Bootstrap methods as well.

For data generated from a mixture of two normal distributions E-skew performed slightly less accurately than the most accurate Monte Carlo Bootstrap method for the upper limit across α level. For the lower limit it performed the most accurately relative to the other methods studied at the α = 0.10 significance level. Otherwise for the lower limit, it performed comparatively as accurately as it did for most of the distributions studied.

Other methods applied on EBSD(n) performed more accurately at the α = 0.01 significance level. At this significance level the method applied on EBSD(n) outperformed its Monte Carlo Bootstrap counterpart frequently for both the upper and lower limit. However, at the α = 0.05 and α = 0.10 significance levels, methods applied on EBSD(n) performed less accurately other than their Monte Carlo Bootstrap counterpart. In general at these significance levels E-skew performed relatively accurately compared to the other method applied on EBSD(n).

4.2 Ratio of Independent Sample Means

For the ratio of sample means portion of the simulation study, results for six different sample sizes are reported ($n =$ 5, 10, 15, 20, 30, and 40). For each of these sample sizes, confidence interval error rates are reported at the α = 0.01, 0.05, and 0.10 significance levels.

The probability distributions used in the simulation study for the ratio of sample means statistic were the normal, exponential, gamma, log-normal, and mixture of two normal distributions. For each distribution, the population parameters specified are displayed below in Table 4.2. These parameter specifications are the same as the specifications for the sample mean in Table 3.5 in Chapter 3.

For each sample size, population parameter specification, and probability distribution combination 10,000 separate samples were generated. For the Monte Carlo Bootstrap confidence interval methods each of the 10,000 samples used 10,000 Monte Carlo Bootstrap resamples to create its Bootstrap sampling distribution. The comparisons discussed in this section are made between EBSD(n) methods and Bootstrap methods that use 10,000 Bootstrap resamples. In addition, confidence interval method error rate results were measured on the same 10,000 unique samples using 200, 500, and 1,000 Bootstrap resamples. These alternative Bootstrap resampling levels were performed for the normal distribution and exponential distribution simulations. The error rate results at these additional Bootstrap resampling levels are reported in the Appendix.

For the remaining statistics studied, including the ratio of sample means statistic, results for ES and EBC_a are not reported. Both methods performed very irregularly for the sample mean. For ES at nearly every sample size, probability distribution and significance level combination ES had an error rate with a larger percent error than every other method. Not only was it larger, in most case EBC_a's percent error was many multiples larger compared to every other method. Additionally, EBC_a's percent error did not show substantial improvement at the larger sample sizes studied. EBC_a was also not reported below. EBC_a was not reported because the method did not demonstrate characteristics typical of a method that would be considered valid. At sample sizes 30 and 40 for the upper limit EBC_a's error rate would increase dramatically and for the lower limit EBC_a's error rate would be 0. This would occur even when the distributions studied were standard cases like data generated from the normal distribution. Therefore, these two methods were omitted from further study. Below are the methods that were studied for the ratio of means statistic:

- For methods using EBSD(n) this included: E-skew, ET, EBC, and EP
- For methods using the Monte Carlo Bootstrap this includes: BT, BC, BP, BC_a, and BS.

Each method in this section used the natural log transformation. A sampling distribution was computed on the natural log transformation of the ratio statistic and then the resulting confidence interval limits were back transformed. Below in Table 4.2 is a description of the parameter specifications used for the sample mean statistic in this simulation study:

Table 4.2 Simulation Parameter Specifications for the Ratio of Sample Means

Probability distribution	Population Parameter	Parameter code: Specified Parameter Values
Normal distribution	(μ_1, σ_1)	N1: (100, 1, 50, 1) N2: (50, 1, 100, 1) N3: (100, 1, 100, 1)
Exponential distribution	(λ_1, λ_2)	E1: (0.10, 0.20) E2: (0.20, 0.05)
Gamma distribution	$(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$	G1: (4, 1, 3, 1)
Log-Normal distribution	$(\mu_1, \sigma_1, \mu_2, \sigma_2)$	L1: (4, 0.2, 3.3, 0.2)
Mixture of two normal distributions	$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2, \mu_3, \sigma_3, p_3, \mu_4, \sigma_4, p_4)$	M1: (50, 1, 0.6, 100, 1, 0.4, 25, 1, 0.6, 50, 1, 0.4)

a. Normal Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the ratio of means statistic when data is normally distributed. The second purpose is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for the ratio of means statistic when data is normally distributed. For the ratio of means statistic two independent samples are generated for each simulation. Then the ratio of means statistic was computed as the mean of sample 1 divided by the mean of sample 2.

For the normal distribution three pairs of independent samples were generated. First the results for data generated with sample 1 distributed as: $N(\mu = 100, \sigma = 1)$ and sample 2 distributed as: $N(\mu = 50, \sigma = 1)$ at the $\alpha = 0.01$ significance level are discussed. The results of this simulation at the $\alpha = 0.01$ significance level can be viewed in tables N1U99 and N1L99 on pages 138 and 139 below. For the ratio of sample means statistic, three separate pairs of independent samples were generated from the normal distributions at three different α significance levels. However, in this section because of the volume of error rate results, only the results for the $N(\mu = 100, \sigma = 1)$, $N(\mu = 50, \sigma = 1)$ specification pair at the $\alpha = 0.01$ significance level are displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the $N(\mu = 100, \sigma = 1)$, $N(\mu = 50, \sigma = 1)$ specification pair, the $N(\mu = 50, \sigma = 1)$, $N(\mu = 100, \sigma = 1)$ and $N(\mu = 100, \sigma = 1)$, $N(\mu = 100, \sigma = 1)$ results can be viewed visually in figures NU99, NL99, NU95, NL95, NU90 and NL90 on pages 142-147. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the $N(\mu = 100, \sigma = 1)$, $N(\mu = 50, \sigma = 1)$ parameter specification pair at the specified $\alpha = 0.01$ significance level, E-skew performed relatively accurately compared to the other methods studied at moderate sample sizes. For the upper and lower limit at the $\alpha = 0.01$ significance level, E-skew had the error rate with the smallest percent error at sample sizes 30 and 40. E-skew also had the error rate with the smallest percent error compared to any method that used EBSD(n) at every sample size for the lower limit. Also for the upper limit when compared to every other method applied on EBSD(n), E-skew had an error rate with a smaller percent error at every sample size but sample size 20. For the upper limit at sample size 20, the EP method had the error rate with the smallest percent error. Otherwise, either BT or BS had the error rate with the smallest percent error at the remaining sample sizes for both the upper and lower limit. These results are shown below in N1U99 and N1L99. These results can also be viewed visually in figures NU99 and NL99.

E-skew performed even more accurately at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels than it did at the $\alpha = 0.01$ significance level. At the $\alpha = 0.01$ significance level E-skew attained the error rate with the smallest percent error for two sample sizes for both the upper and lower limit. At the $\alpha = 0.05$ significance level, E-skew attained the error rate with the smallest percent error at five sample sizes for both the upper and lower limit. At the $\alpha = 0.10$ significance level, E-skew attained the error rate with the smallest percent error for three sample sizes for the upper limit and all six sample size for the lower limit.

By comparison percentile methods applied on EBSD(n) performed relatively more accurately at the α = 0.01 significance level and relatively less accurately at the α = 0.05 and α = 0.10 significance levels. For the upper limit at the α = 0.01 significance level, EP did achieve the error rate with the smallest percent error at sample size 20 when compared to every other method studied. However, at the α = 0.05 and α = 0.10 significance levels, no method other than E-skew achieved an error rate with the smallest percent error at any sample size for the upper or lower limit. Additionally, when comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors for each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart. However, at the α = 0.05 and α = 0.10 significance levels the EBSD(n) percentile algorithms had a larger percent error at each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart.

Additionally, ET performed relatively less accurately compared to BT at every sample size and significance level. ET had an error rate with a larger percent error for both the upper and lower limit at each sample size and significance level when compared to BT.

Table: Ratio of Sample Means - N1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Normal Distributions, $\boldsymbol{N_1(\mu_1 = 100, \sigma_1 = 1), N_2(\ \mu_2 = 50,\ \ \sigma_2 = 1), \mathbf{Bootstraps} = 10000}$

Sample size	5	10	15	20	30	40
E-skew	0.0098 (96%)	0.0071 (42%)	0.0058 (16%)	0.0074 (48%)	**0.0047 (6%)**	**0.0048 (4%)**
BT	**0.0065 (30%)**	**0.0055 (10%)**	**0.0055 (10%)**	0.0062 (24%)	0.0043 (14%)	0.0056 (12%)
ET	0.038 (660%)	0.019 (280%)	0.0123 (146%)	0.0123 (146%)	0.0081 (62%)	0.0072 (44%)
BC	0.0486 (872%)	0.0219 (338%)	0.0132 (164%)	0.0122 (144%)	0.0082 (64%)	0.0071 (42%)
EBC	0.029 (480%)	0.0101 (102%)	0.0037 (26%)	0.0039 (22%)	0.0027 (46%)	0.0046 (8%)
BP	0.0489 (878%)	0.0216 (332%)	0.013 (160%)	0.0116 (132%)	0.0084 (68%)	0.007 (40%)
EP	0.0285 (470%)	0.0084 (68%)	0.0031 (38%)	**0.0058 (16%)**	0.0025 (50%)	0.0041 (18%)
BS	0.0076 (52%)	0.0079 (58%)	0.0065 (30%)	0.007 (40%)	0.0057 (14%)	0.0062 (24%)
$\boldsymbol{BC_a}$	0.0535 (970%)	0.023 (360%)	0.0138 (176%)	0.0118 (136%)	0.0082 (64%)	0.0072 (44%)

Table: Ratio of Sample Means - N1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Normal Distribution, $\boldsymbol{N_1(\mu_1 = 100, \sigma_1 = 1), N_2(\mu_2 = 50,\ \sigma_2 = 1), Bootstraps = 10000}$

Sample size	5	10	15	20	30	40
E-skew	0.0124 (148%)	0.007 (40%)	0.0058 (16%)	0.0048 (4%)	**0.0054** **(8%)**	**0.0054** **(8%)**
BT	**0.0064** **(28%)**	**0.006** **(20%)**	**0.0055** **(10%)**	0.0046 (8%)	0.006 (20%)	0.0061 (22%)
ET	0.0385 (670%)	0.0164 (228%)	0.0113 (126%)	0.0082 (64%)	0.0081 (62%)	0.0077 (54%)
BC	0.0486 (872%)	0.0179 (258%)	0.0122 (144%)	0.0083 (66%)	0.0082 (64%)	0.0077 (54%)
EBC	0.0301 (502%)	0.0087 (74%)	0.0023 (54%)	0.0032 (36%)	0.0034 (32%)	0.0058 (16%)
BP	0.0524 (948%)	0.0198 (296%)	0.0124 (148%)	0.0091 (82%)	0.0093 (86%)	0.0083 (66%)
EP	0.0313 (526%)	0.0102 (104%)	0.0031 (38%)	0.0033 (34%)	0.0036 (28%)	0.0055 (10%)
BS	0.0075 (50%)	0.0066 (32%)	0.0062 (24%)	**0.0051** **(2%)**	0.0059 (18%)	0.0061 (22%)
$\boldsymbol{BC_a}$	0.0506 (912%)	0.0225 (350%)	0.0148 (196%)	0.0107 (114%)	0.0106 (112%)	0.009 (80%)

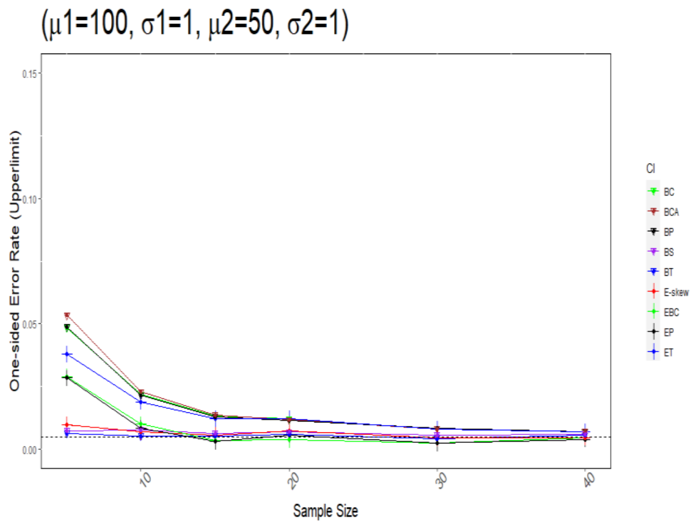
Simulation were not only performed for the $N(\mu = 100, \sigma = 1)$, $N(\mu = 50, \sigma = 1)$ specification pair. Simulations were also performed for the following parameter specification pairs: $N(\mu = 50, \sigma = 1)$, $N(\mu = 100, \sigma = 1)$ and $N(\mu = 100, \sigma = 1)$, $N(\mu = 100, \sigma = 1)$. When considering a change in parameter specification (i.e. decreasing the expected value of the ratio of means from 2 to 1) E-skew performed relatively as accurately across α significance level as it did for the first parameter specification tested. For the upper limit at the $\alpha = 0.01$ significance level, E-skew did not have the error rate with the smallest percent error at any sample size. However, for the lower limit at the α = 0.01 significance level, it did have the error rate with the smallest percent error at two sample sizes. Additionally, E-skew did have the error rate with the smallest percent error among all methods applied on EBSD(n) at every sample size but sample sizes 20 and 40 for the upper limit and every sample size but sample size 40 for the lower limit at the α = 0.01 significance level.

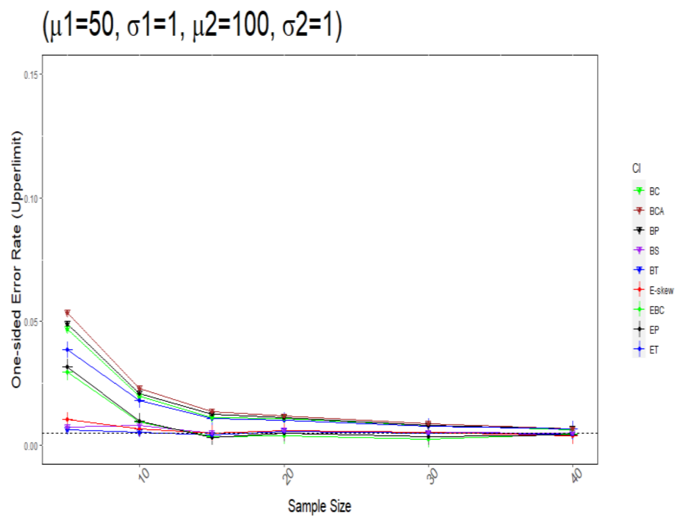
Further as seen with the first specification studied, E-skew's accuracy improved at the α = 0.05 and α = 0.10 significance levels. At the α = 0.05 significance level for the upper limit, E-skew had the error rate with the smallest percent error at all six sample sizes. For the lower limit at this significance level, E-skew had the error rate with the smallest percent error at five of six sample sizes. At the α = 0.10 significance level, E-skew attained the error rate with the smallest percent error at five sample sizes for the upper limit and three sample sizes for the lower limit. No other method applied on EBSD(n) attained the error rate with the smallest percent error for any sample size for the upper or lower limit at the α = 0.05 and α = 0.10 significance levels.

When the expected value of the ratio of sample means from the two independent samples was decreased to 0.5, E-skew was relatively as accurate as it was for the other two parameter specifications studied. E-skew attained the error rate with the smallest percent error at the α = 0.01 significance level for one sample size for the upper limit and two for the lower limit. Also the EP method performed relatively accurately at the α = 0.01 significance level, attaining the error rate with the smallest percent error at sample size 40 for the lower limit. No other method using EBSD(n) attained the error rate with the smallest percent error at the α=0.05 and α=0.10 significance levels. At the α=0.05

significance level for the upper limit, E-skew attained the error rate with the smallest percent error for four sample sizes for the upper limit and five sample sizes for the lower limit. Additionally, at the α=0.10 significance level, E-skew attained the error rate with the smallest percent at five of six sample sizes for both the upper and lower limit.

Figure: Ratio of Sample Means - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Normal Distribution





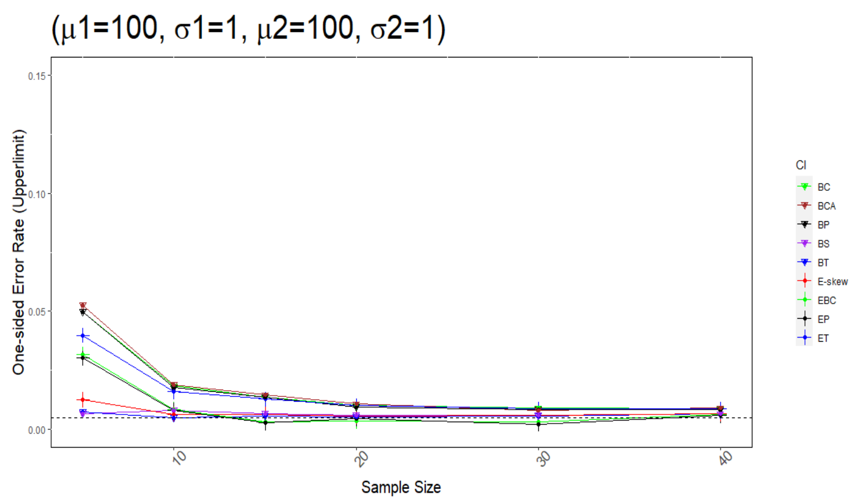


Figure: Ratio of Sample Means - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Normal Distribution

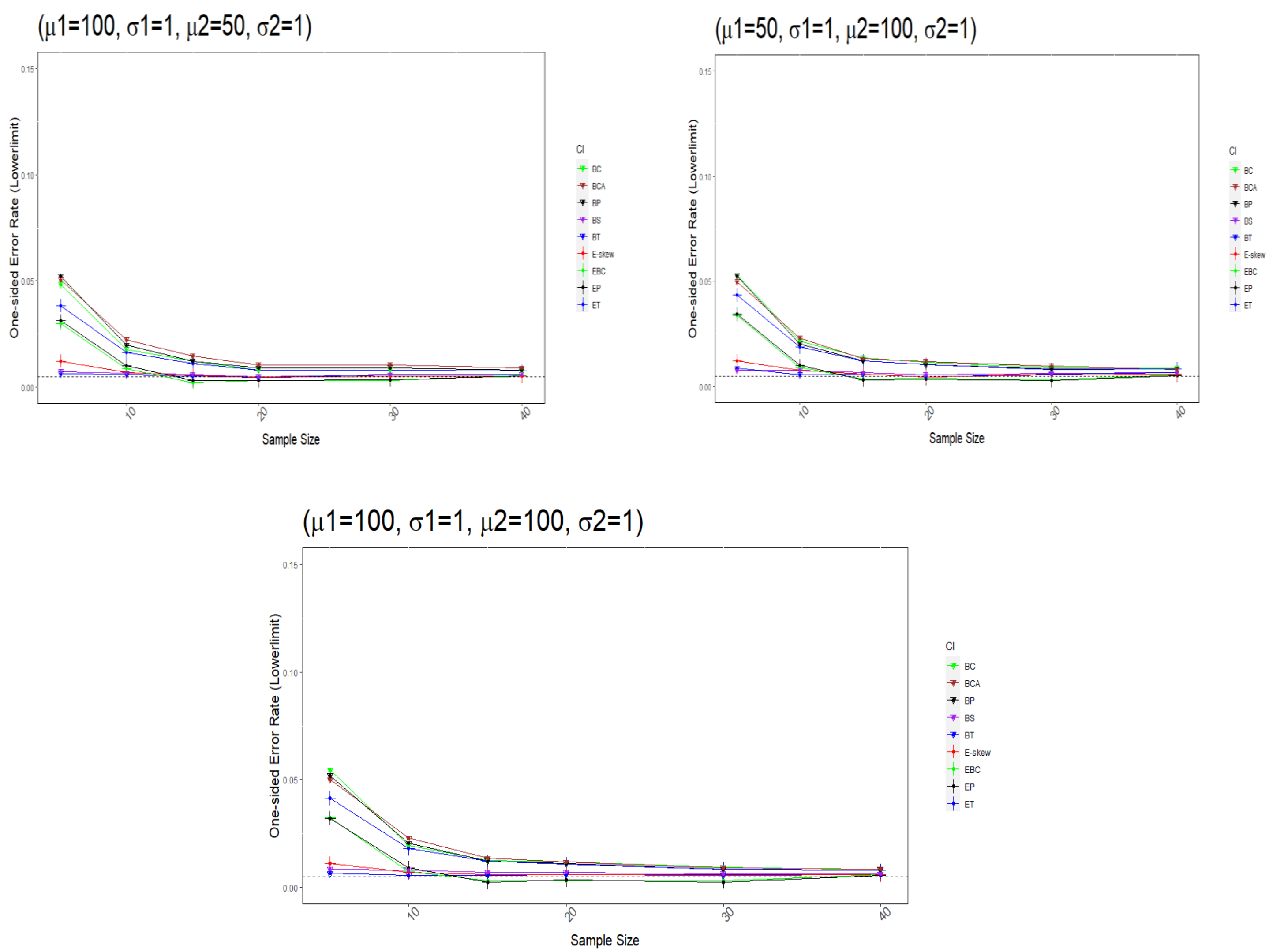
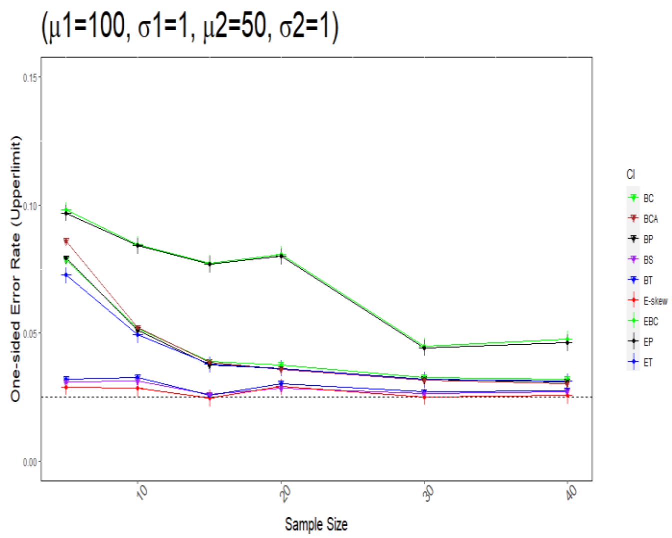
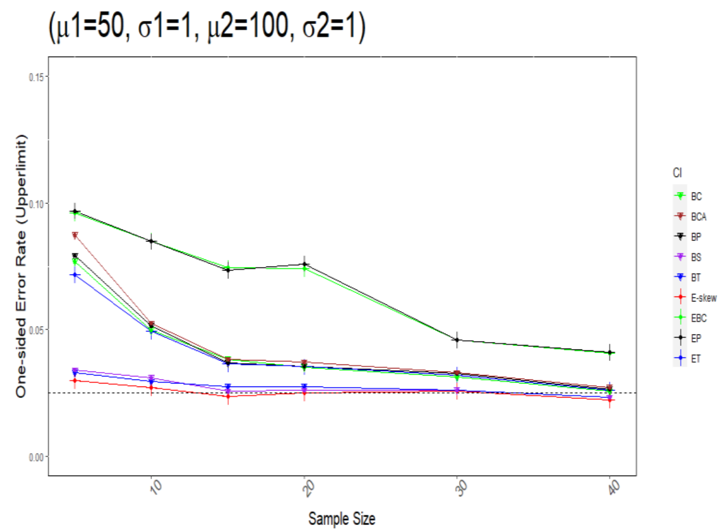


Figure: Ratio of Sample Means - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Normal Distribution





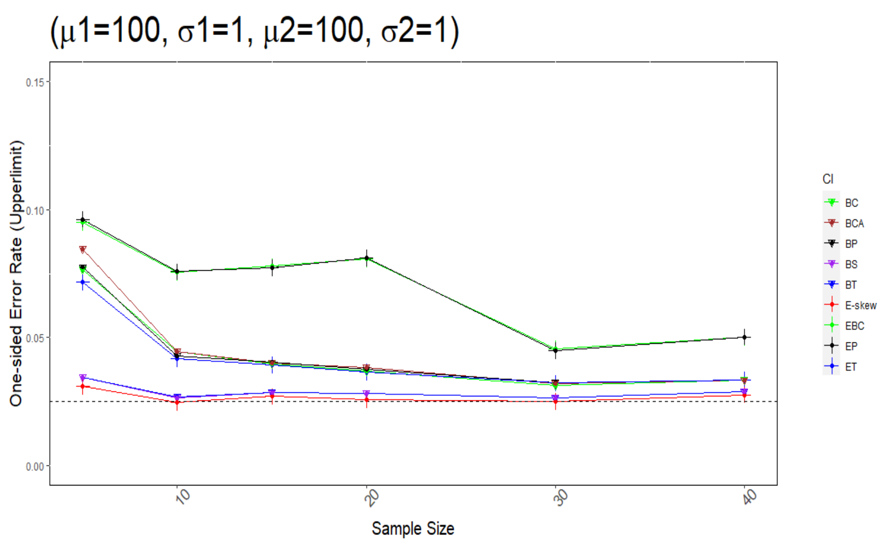
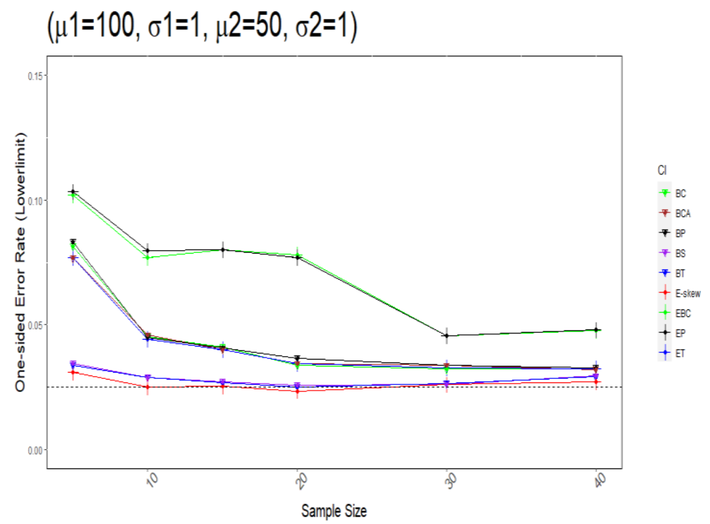
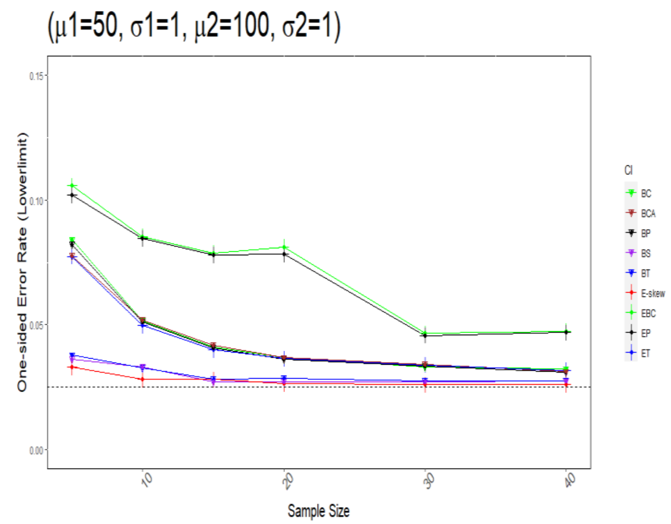


Figure: Ratio of Sample Means - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Normal Distribution





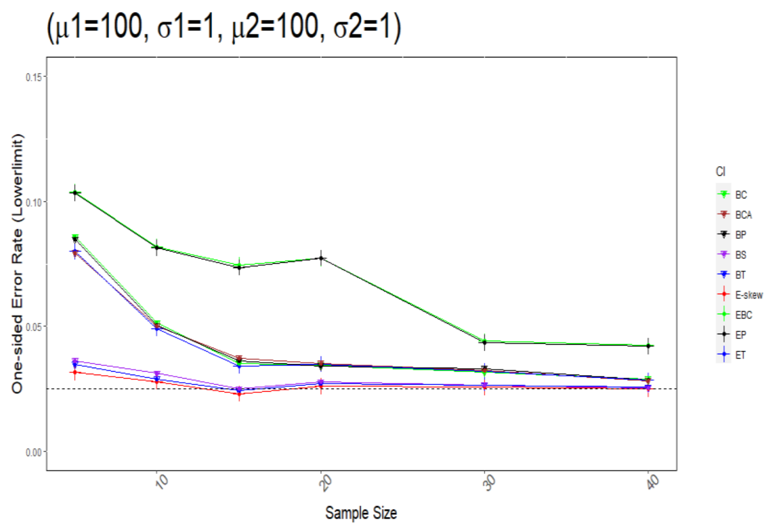


Figure: Ratio of Sample Means - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Normal Distribution

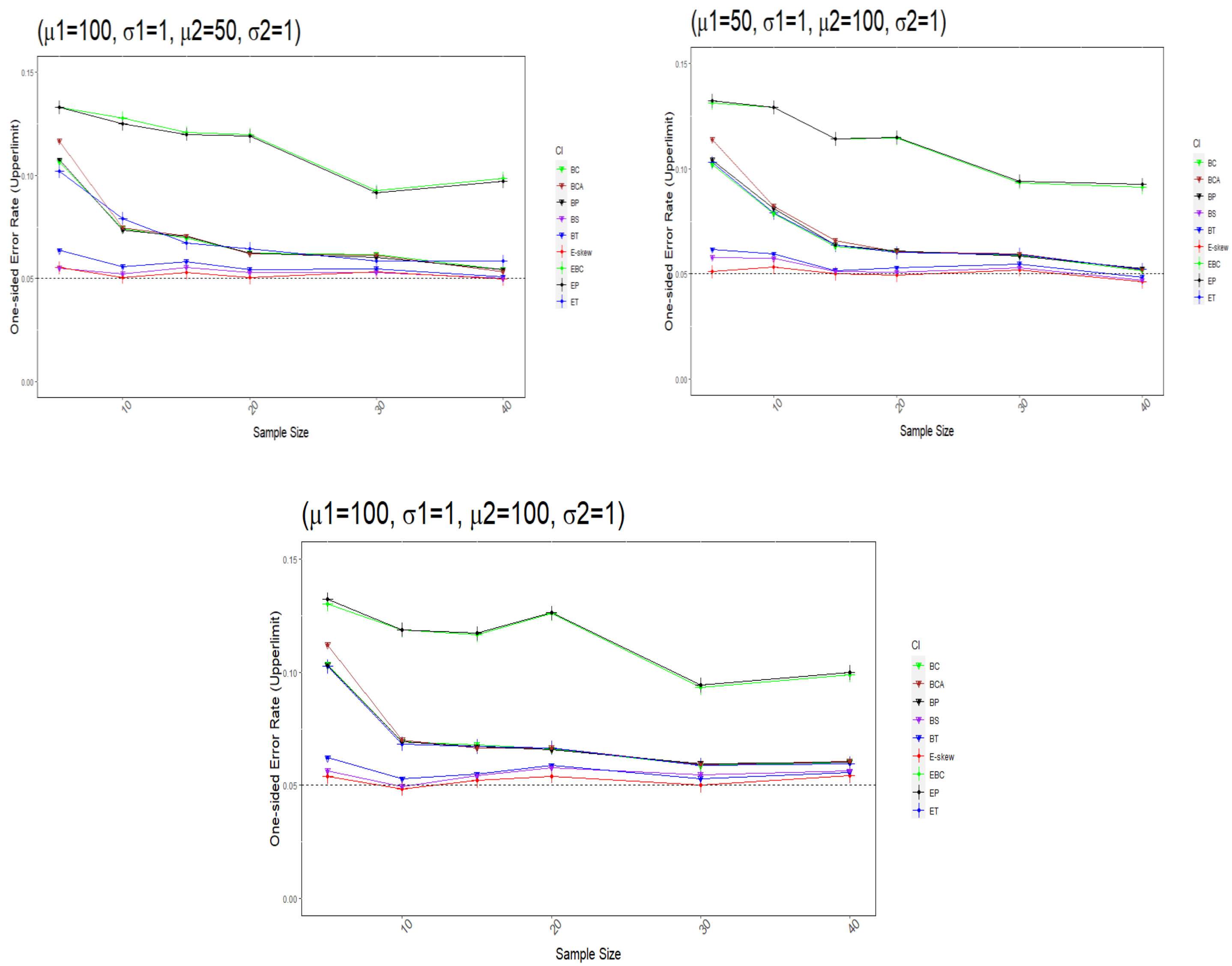
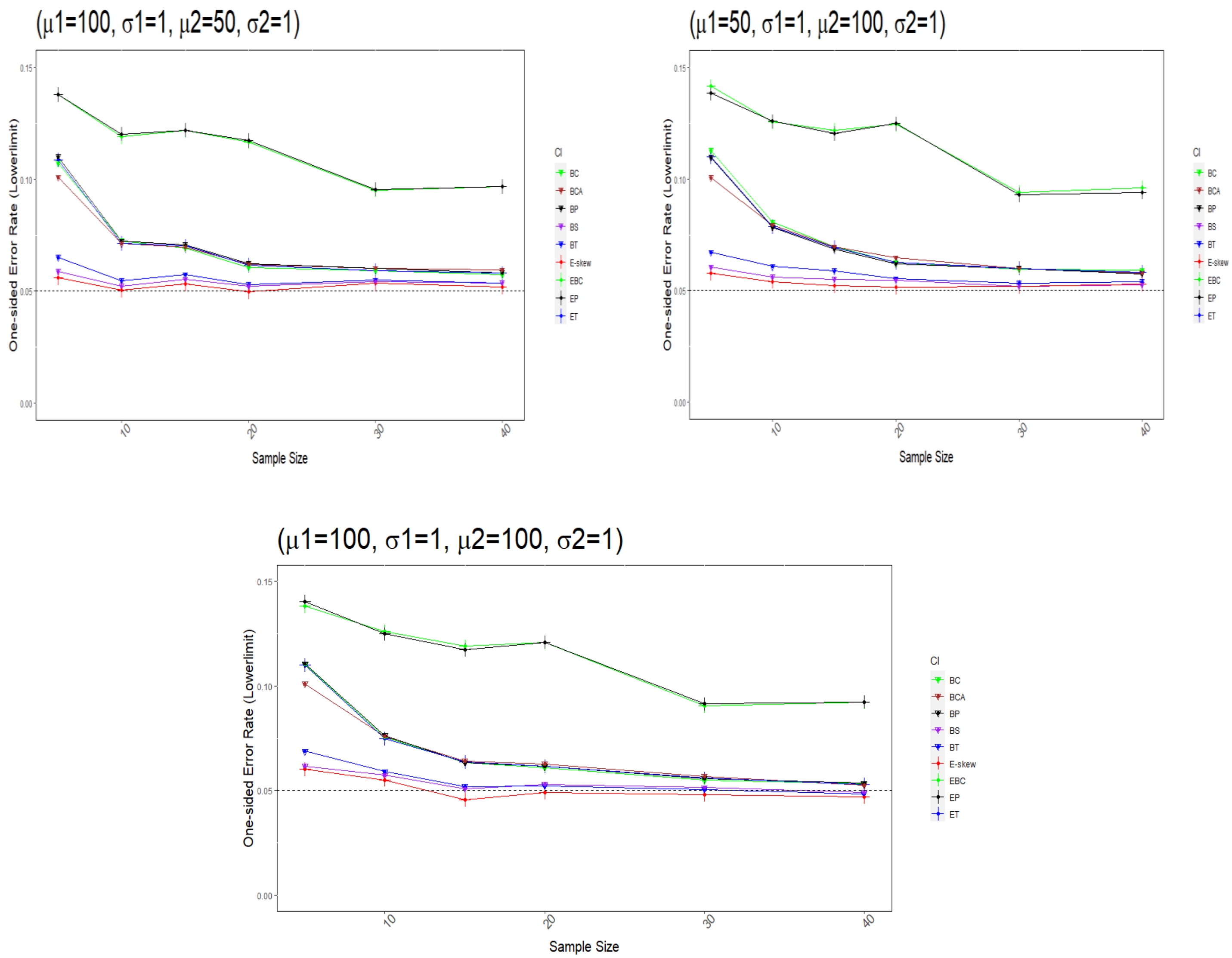


Figure: Ratio of Sample Means - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Normal Distribution



b. Exponential Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the ratio of means statistic when data is exponentially distributed. The second purpose is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for the ratio of means statistic when data is exponentially distributed. E-skew error rates and error rates for other methods using EBSD(n) are compared to Monte Carlo Bootstrap method error rates for data generated from a pair of independent exponential distributions Exp(λ=0.10) and Exp(λ=0.20).

The error rate results at the α = 0.01 significance level are displayed in tables E1U99, and E1L99 on pages 151 and 152 below. For the exponential distribution, two separate pairs of independent samples were studied, however because of the volume of error rate results, only the Exp(λ=0.10), Exp(λ=0.20) pair at the α = 0.01 significance level are displayed in the tables. Detailed numerical results for simulations not included here can be viewed in Appendix tables. Although these tables only report results for the Exp(λ=0.10), Exp(λ=0.20) pair, the Exp(λ=0.20), Exp(λ=0.05) specification pair can be viewed visually in figures EU99, EL99, EU95, EL95, EU90 and EL90 on pages 155-160.

For the Exp(λ=0.10), Exp(λ=0.20) parameter specification at the specified α = 0.01 significance level, for the upper limit, E-skew performed relatively less accurately compared to BS and the other methods applied on EBSD(n) that were studied. E-skew did not have the error rate with the smallest percent error at any sample size for the upper limit and for the lower limit had the error rate with the smallest percent error for only one sample size (n=30). However, for the upper and lower limit, E-skew did have the error rate with the smallest percent error at sample size 5 compared to any method using EBSD(n).

Other methods applied on EBSD(n) performed relatively accurately compared to E-skew and Monte Carlo Bootstrap methods at the α = 0.01 significance level. At this significance level for the upper limit, ET and EP had the error rate with the smallest percent error at two separate sample sizes and EP had the error rate with the smallest percent error at three separate sample sizes. For the lower limit at this significance level, ET had the error rate with the smallest percent error at sample sizes 20 and 30. EP had the error rate with the smallest percent error at sizes 10 and 40. Also for the lower limit at this significance level, EBC had the error rate with the smallest percent at sample size 15.

Once again, the strength of the E-skew method was demonstrated when comparing the error rates across α significance level. E-skew maintained or improved relative accuracy compared to Monte Carlo Bootstrap and EBSD(n) methods, across sample size, as the α significance level increased from α = 0.01 to α = 0.10. At the α = 0.05 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at four separate sample sizes among methods applied on EBSD(n). At the α = 0.10 significance level for the lower limit, it achieved the error rate with the smallest percent error among all methods for four sample sizes (n =5, 20, 30 and 40). Additionally, for the upper limit at the α = 0.05 significance level, E-skew attained the smallest percent error at sample size 40 and additionally the smallest among all methods applied on EBSD(n) at sample sizes 5 and 30. For the upper limit at the α = 0.10

significance level, E-skew attained the error rate with the smallest percent error at the same four sample sizes studied as for the lower limit.

By comparison the EP method was relatively more accurate at the α = 0.01 significance level and became relatively less accurate compared to the other methods studied at larger α levels. At the α = 0.01 significance level for the upper limit, EP had the error rate with the smallest percent error at sample size 40 with a percent error of 0%. However at the α = 0.05 and α = 0.10 significance levels for the upper limit, the error rate percent errors increased to 98.8% and 102.4% respectively at sample size 40. The percent error increased similarly for the lower limit from the α = 0.01 to the α = 0.05, and α = 0.10 significance levels. ET did perform relatively similarly across significance level. At the α = 0.01 significance level for the upper limit, it had the error rate with the smallest percent error at two sample sizes. Then at the α = 0.05 and 0.10 significance levels for the upper limit respectively, ET again attained the error rate with the smallest percent error at two sample sizes in each case. A similar pattern was found for the lower limit. Although ET performed relatively as accurately across α significance level, the sample size that ET attained the smallest percent error were sample sizes 10, 15, 20 or 30. At each significance level for the upper limit, ET's percent error was larger at sample size 40 than it was at the sample size it attained the smallest percent error for a given α significance level and limit end.

Table: Ratio of Sample Means - E1U99 Upper limit error rate (α = 0.01), Exponential Distributions, Exp_1(λ_1 = 0.1), Exp_2(λ_2 =0.2), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0169 (238%)	0.0135 (170%)	0.0127 (154%)	0.0088 (76%)	0.0093 (86%)	0.0071 (42%)
BT	**0.007** **(40%)**	0.0079 (58%)	0.01 (100%)	0.0072 (44%)	0.0079 (58%)	0.0065 (30%)
ET	0.0313 (526%)	0.0095 (90%)	0.0081 (62%)	**0.0049** **(2%)**	**0.004** **(20%)**	0.0025 (50%)
BC	0.053 (960%)	0.0219 (338%)	0.0171 (242%)	0.0121 (142%)	0.0116 (132%)	0.0085 (70%)
EBC	0.0297 (494%)	0.0094 (88%)	0.0029 (42%)	0.0036 (28%)	0.0038 (24%)	0.0054 (8%)
BP	0.0561 (1022%)	0.0209 (318%)	0.0171 (242%)	0.0123 (146%)	0.0106 (112%)	0.0095 (90%)
EP	0.0338 (576%)	**0.007** **(40%)**	**0.0038** **(24%)**	0.0043 (14%)	0.0029 (42%)	**0.005** **(0%)**
BS	0.0127 (154%)	0.0107 (114%)	0.0105 (110%)	0.0077 (54%)	0.0084 (68%)	0.0065 (30%)
BC_a	0.0619 (1138%)	0.0222 (344%)	0.0182 (264%)	0.013 (160%)	0.0115 (130%)	0.0097 (94%)

Table: Ratio of Sample Means - E1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Exponential Distributions, $\boldsymbol{Exp_1}(\boldsymbol{\lambda_1}$ = 0.1), $\boldsymbol{Exp_2}(\boldsymbol{\lambda_2}$ =0.2), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0187 (274%)	0.0136 (172%)	0.0109 (118%)	0.0077 (54%)	**0.0069** **(38%)**	0.0073 (46%)
BT	**0.0084** **(68%)**	0.0096 (92%)	0.0078 (56%)	0.0063 (26%)	0.007 (40%)	0.0062 (24%)
ET	0.0305 (510%)	0.0111 (122%)	0.0073 (46%)	**0.0041** **(18%)**	**0.0031** **(38%)**	0.003 (40%)
BC	0.0542 (984%)	0.0244 (388%)	0.0156 (212%)	0.0106 (112%)	0.0102 (104%)	0.0082 (64%)
EBC	0.0327 (554%)	0.0101 (102%)	**0.0039** **(22%)**	0.0026 (48%)	0.0028 (44%)	0.0045 (10%)
BP	0.0543 (986%)	0.0228 (356%)	0.0165 (230%)	0.0117 (134%)	0.0099 (98%)	0.0085 (70%)
EP	0.0303 (506%)	**0.0094** **(88%)**	0.0027 (46%)	0.0036 (28%)	0.0022 (56%)	**0.0049** **(2%)**
BS	0.0128 (156%)	0.0136 (172%)	0.0104 (108%)	0.0062 (24%)	0.0072 (44%)	0.006 (20%)
$\boldsymbol{BC_a}$	0.0536 (972%)	0.0243 (386%)	0.017 (240%)	0.0115 (130%)	0.0103 (106%)	0.0092 (84%)

Simulation were not only performed for the exponential distribution for the parameter pair Exp(λ=0.10), Exp(λ=0.20). Simulations were also performed for Exp(λ=0.20), Exp(λ=0.05) parameter specifications pair. When the expected ratio of means was decreased from 2 to 0.25, E-skew performed most accurately at the α = 0.10 significance level. At the α = 0.10 significance level for both the upper and lower limit, E-skew had the error rate with the smallest percent error compared to all methods measured at five of six sample sizes. At the α = 0.05 significance level for both the upper and lower limit E-skew attained the error rate with the smallest percent error at sample size 40 when compared to any other method. It also had the error rate with the smallest percent error for three sample sizes at this significance level, for both the upper and lower limit, when compared to any other method using EBSD(n). At the α = 0.01 significance level, E-skew did not perform as accurately. At no sample size for either the upper or lower limit did E-skew attain the error rate with the smallest percent error. Additionally, E-skew only had the error-rate with the smallest percent error among methods applied on EBSD(n) at sample size 5 for both the upper and lower limit.

For the other methods using EBSD(n), changing the parameter specification yielded similar error rate pattern across α significance level as it did for the previous parameter specification. When the expected value for the ratio of means was decreased from 2 to 0.25, ET performed relatively as accurately across α level. At the α = 0.01 significance level for both the upper and lower limit, ET had the error rate with the smallest percent error at two sample sizes. At the α = 0.05 and α = 0.10 significance levels, for both interval ends, it had the error rate with the smallest percent error at three sample sizes and one sample size respectively. Again ET had the smallest percent errors at sample sizes 10, 15, 20 or 30 at each α significance level. A consistent pattern for error rates across α significance level was also seen for the percentile methods EBC and EP. EBC and EP performed relatively more accurately at the α = 0.01 significance level, and then relatively less accurately at the other two significance levels. At the α = 0.01 significance level, EBC had the error rate with the smallest percent error for two sample sizes for the upper limit, while EP had the smallest at one sample size for both the upper and lower limit. However at the α = 0.05 and α = 0.10 significance levels, both EBC and

EP had error rates with percent errors that were larger when compared to their Monte Carlo Bootstrap counterpart at each sample size.

Figure: Ratio of Sample Means - EU99 - One-Sided Upperlimit Error Rates for 99% CI for the Exponential Distribution

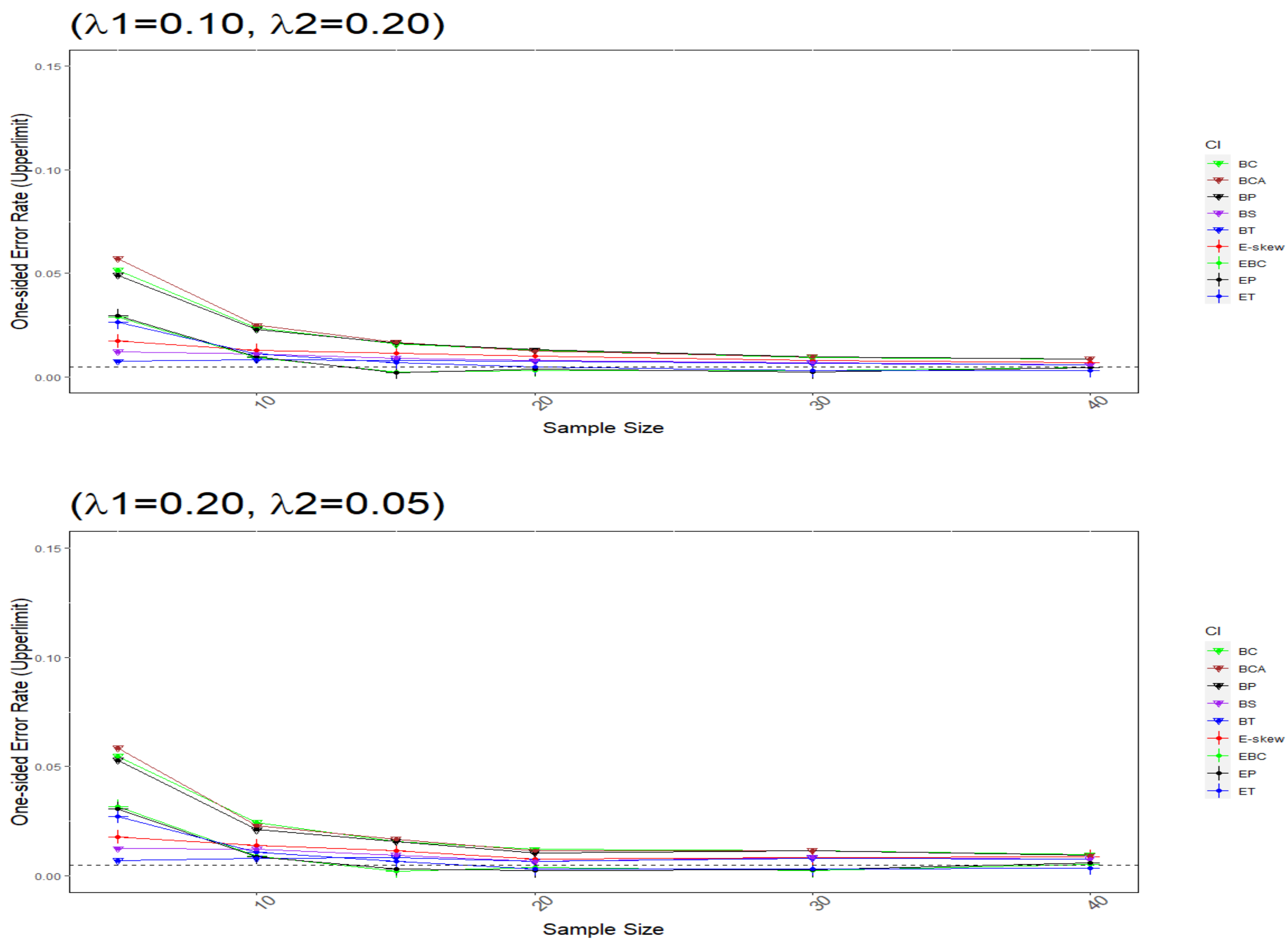
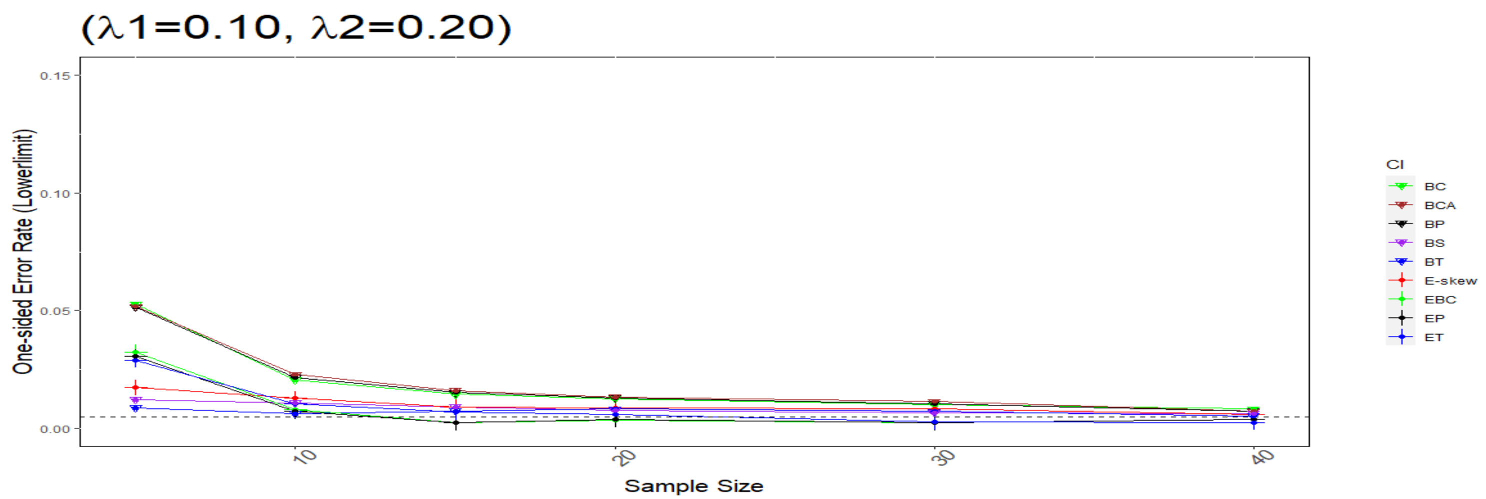


Figure: Ratio of Sample Means - EL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Exponential Distribution



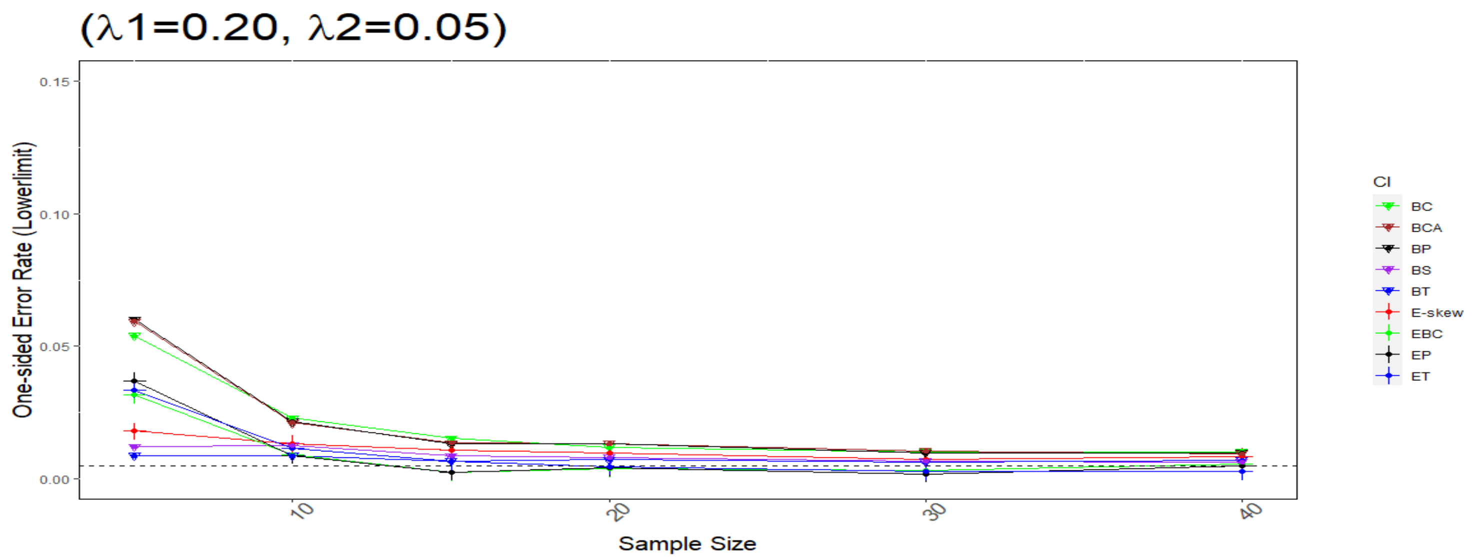
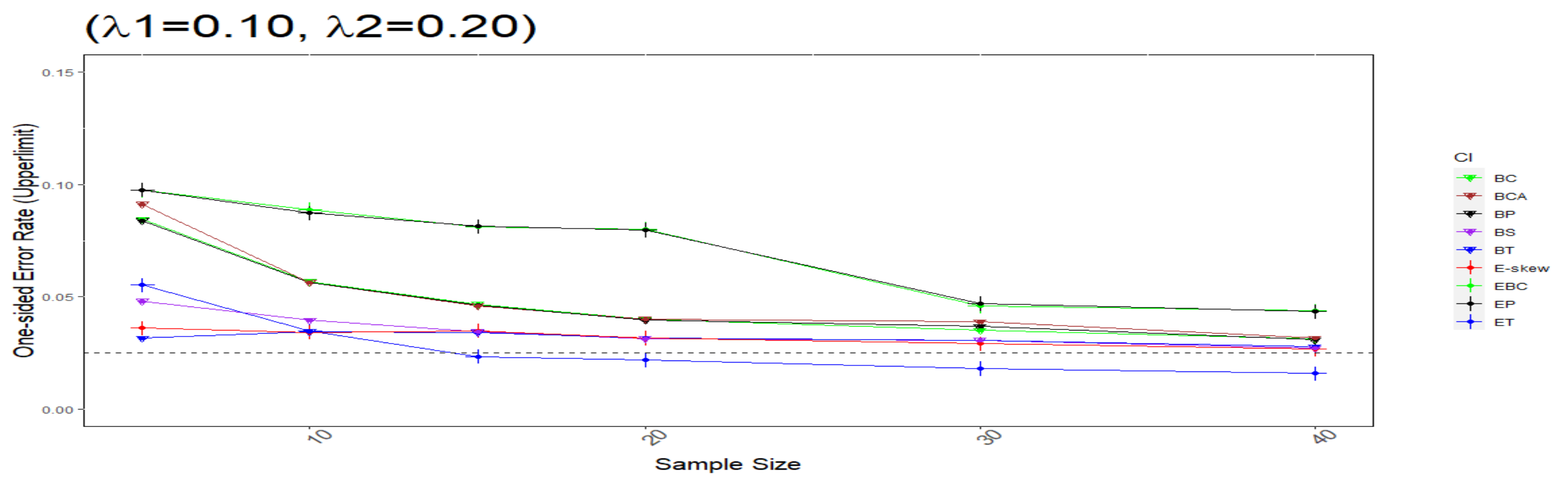


Figure: Ratio of Sample Means - EU95 - One-Sided Upperlimit Error Rates for 95% CI for the Exponential Distribution



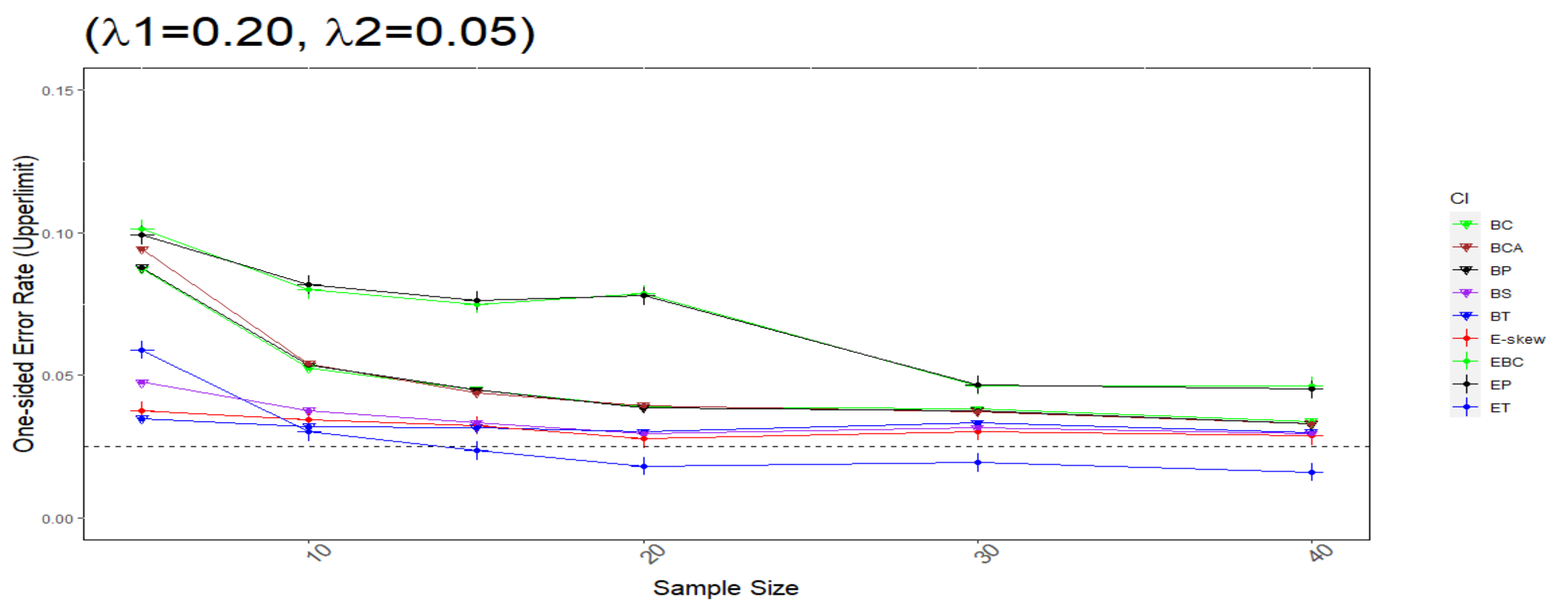
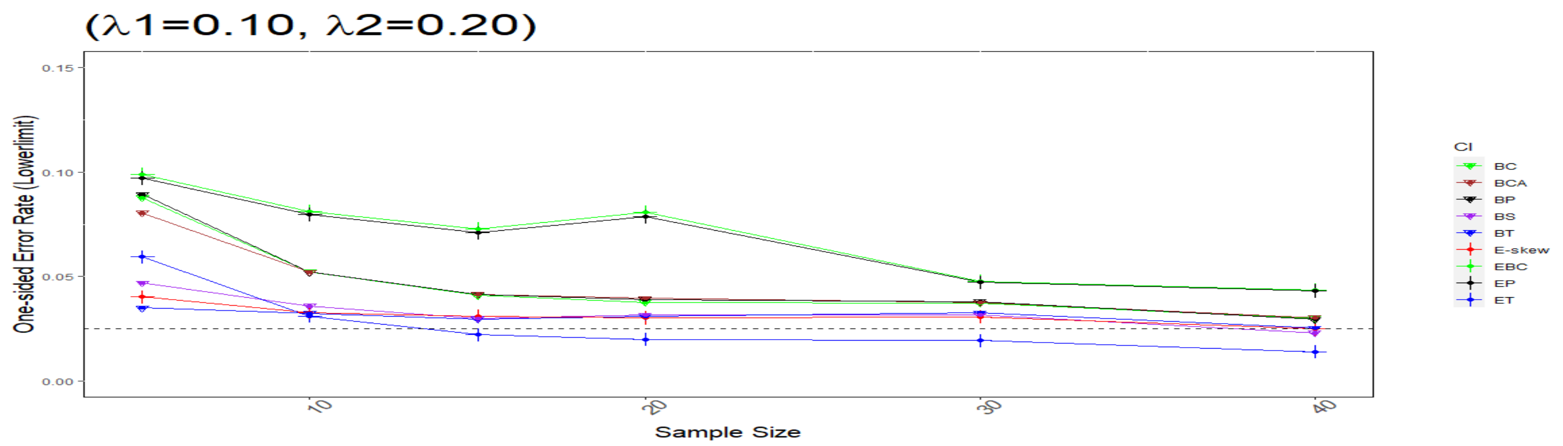


Figure: Ratio of Sample Means: EL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Exponential Distribution



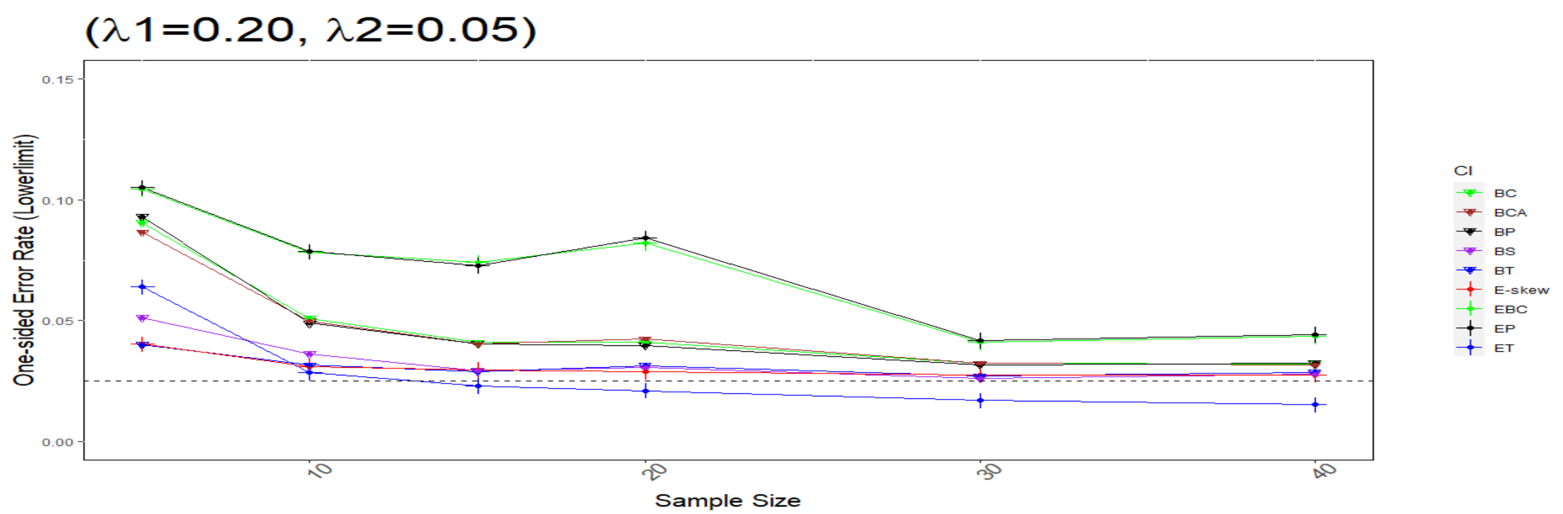
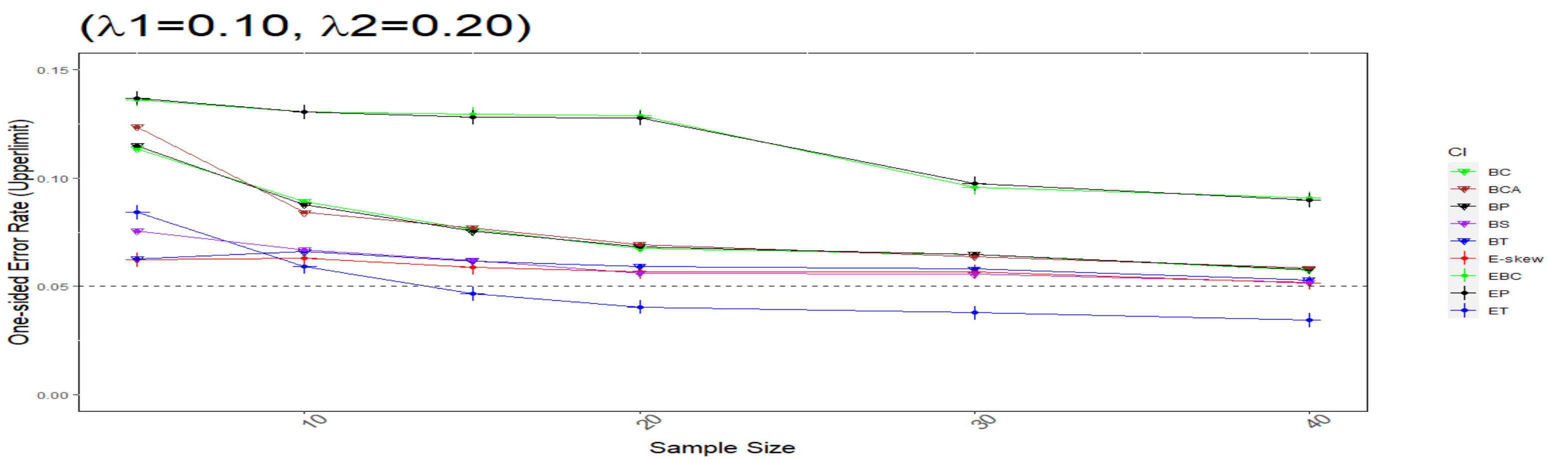


Figure: Ratio of Sample Means: EU90 - One-Sided Upperlimit Error Rates for 90% CI for the Exponential Distribution



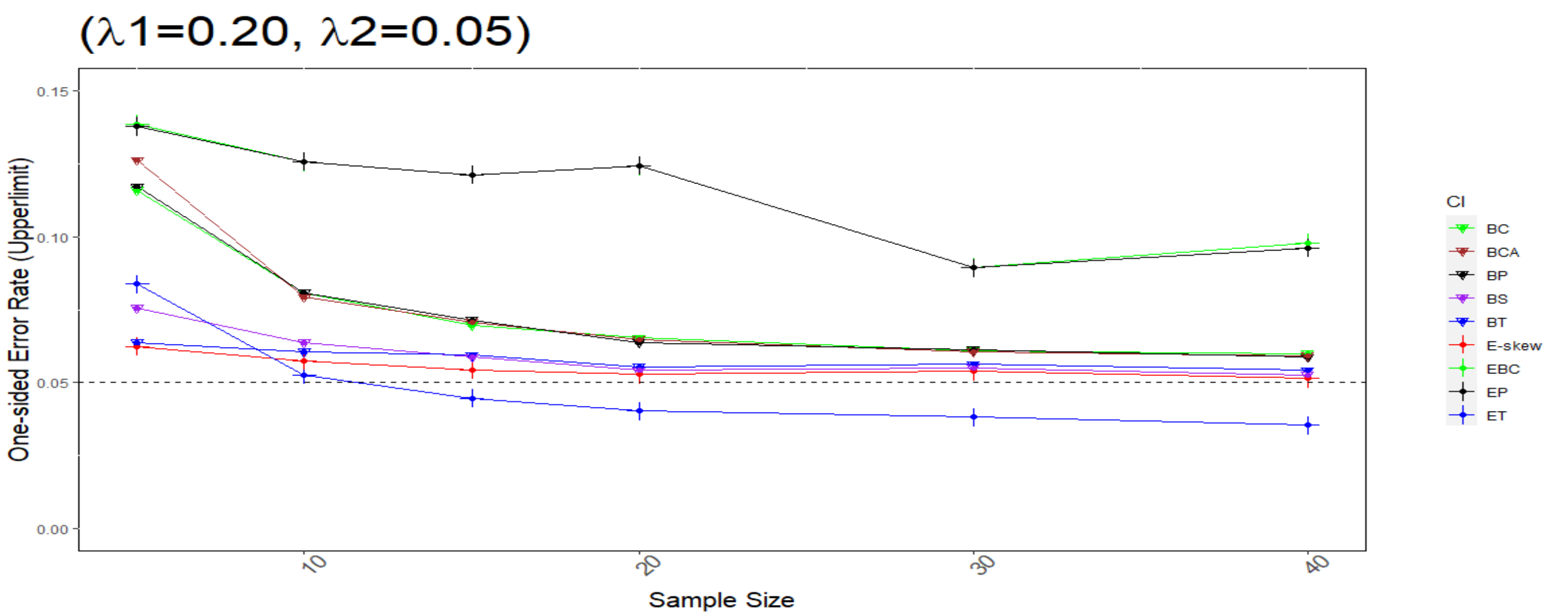
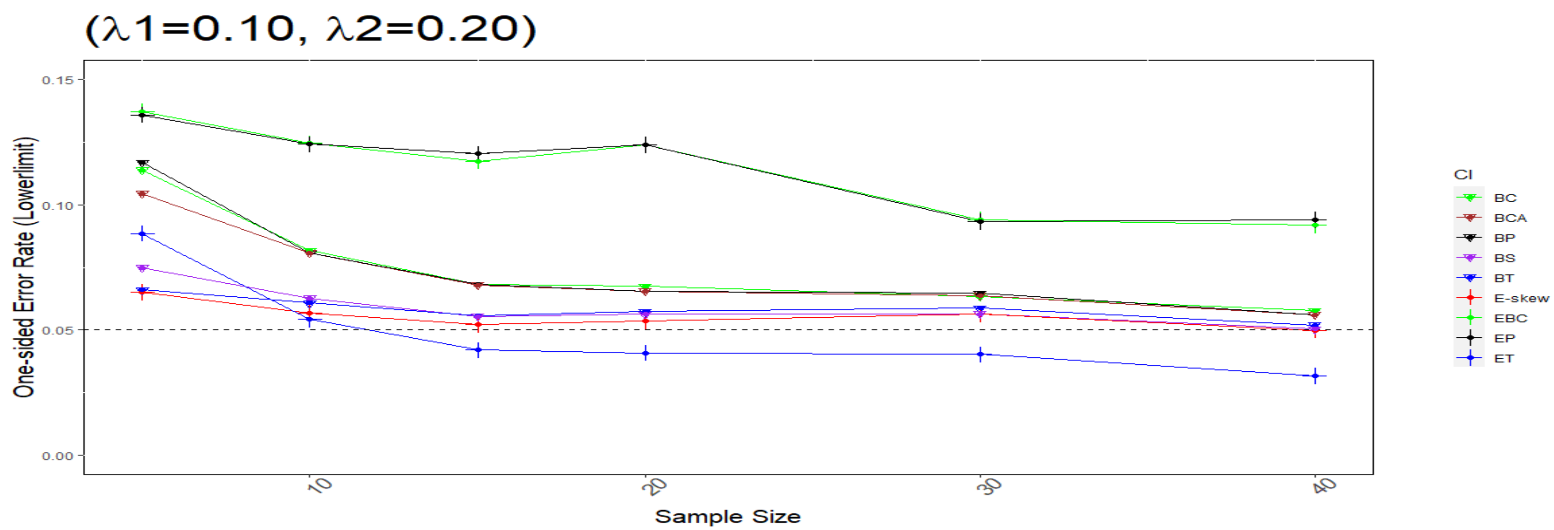
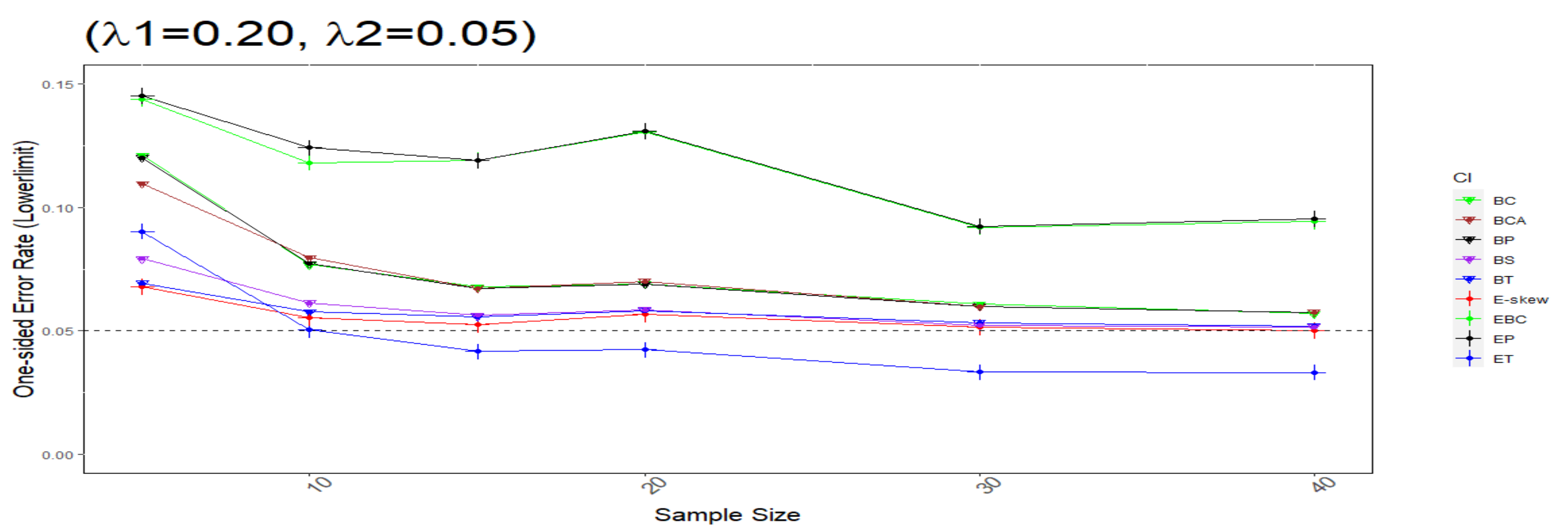


Figure: Ratio of Sample Means: EL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Exponential Distribution





c. Gamma Distribution

The first purpose of this sub section is to compare the accuracy of the E-skew method to the accuracy of all other methods studied for the ratio of means statistic when data is generated from a gamma distribution. The second purpose is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for the ratio of means statistic when data is generated from a gamma distribution. E-skew error rates and error rates for other methods using EBSD(n) are compared to Monte Carlo Bootstrap method error rates for data generated from a pair of independent gamma distributions Gamma(α=4, λ=1) and Gamma(α=3, λ=1). The error results at the α = 0.01 significance level are displayed in tables G1U99, and G1L99 on pages 164 and 165 below. For the gamma distribution, only one pair of independent samples was studied. Error rate results at the α = 0.05 and α = 0.10 specification pair is displayed in Appendix tables. Although these tables only report results at the α = 0.01 significance level, results for the other two significance levels can be viewed visually in figures G99, G95, and G90 on pages 166-168.

For the Gamma(α=4, λ=1), Gamma(α=3, λ=1) parameter specification at the specified α = 0.01 significance level for the upper limit, E-skew performed relatively accurately at moderate sample sizes compared to the other methods studied. For the upper limit, E-skew had the error rate with the smallest percent error at sample size 40 for the upper limit. Additionally, for the upper limit at this significance level, E-skew had the error rate with the smallest percent error at four of six sample sizes compared to any other method using EBSD(n). However, for the lower limit, E-skew mostly performed relatively less accurately compared to every other method applied on EBSD(n) studied at the α = 0.01 significance level.

Percentile methods using EBSD(n) performed relatively accurately at the α = 0.01 significance level compared to Monte Carlo Bootstrap methods. At this significance level for the upper limit, EBC had the error rate with the smallest percent error at one sample size for both the upper and lower limit. Additionally, for the lower limit at this significance level, EP had the error rate with the smallest percent error at three separate sample sizes. At this significance level, both EBC and EP had error rates with a smaller percent error compared to their Monte Carlo Bootstrap counterpart at five of six sample sizes for the upper limit and smaller at all six sample sizes for the lower limit. Conversely, ET had an error rate with an equal or larger percent error at each sample size for both the upper and lower limit.

Once again, the strength of the E-skew method is demonstrated when comparing the error rates across α significance level. E-skew maintained or improved relative accuracy as the α significance level increased from α = 0.01 to α = 0.10. At the α = 0.01 significance level, E-skew attained the error rate with the smallest percent error at sample size 40 for the upper limit. At the α = 0.05 significance level for the upper limit, E-skew attained the error rate with the smallest percent error at five of six sample sizes studied. At the α = 0.05 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at one sample size and the smallest among all methods applied on EBSD(n) at two other sample sizes. At the α = 0.10 significance level for the upper limit, E-skew attained the error rate with the smallest percent error at four of six sample sizes studied. At the α = 0.10 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at one sample size and the smallest among all methods applied on EBSD(n) at three other sample sizes.

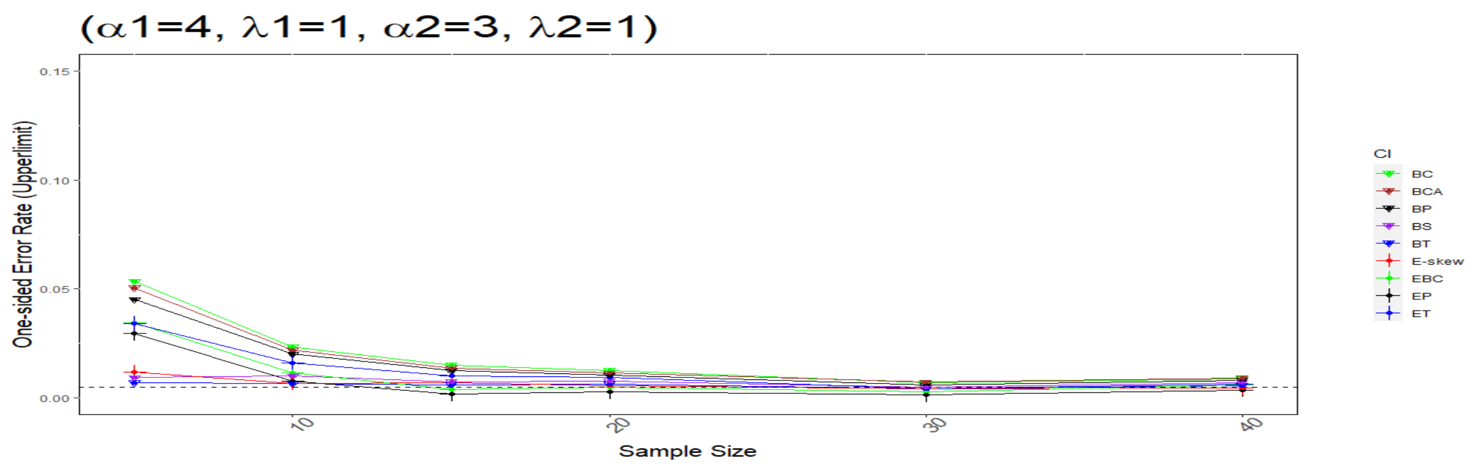
By comparison the EP and EBC methods which each attained the error rate with the smallest percent error at one sample at the α = 0.01 significance level, did not at any sample size for the other two significance levels. In addition, EBC and EP had error rates with percent errors that were larger when compared to their Monte Carlo Bootstrap counterpart at each sample size. ET's relative accuracy improved as significance level increased. At the α = 0.01 significance level for the upper and lower limit, ET had an error rate with a percent error that was equal to or larger at each sample size compared to BT. However, at the α = 0.05 and 0.10 significance levels for the lower limit, it attained the error rate with the smallest percent error at two sample sizes. Additionally, for the upper limit, it attained the error rate with the smallest percent error at one sample size for the α = 0.10 significance level. Therefore, ET performed relatively more accurately at the α = 0.05 and 0.10 significance levels and relatively less accurately at the α = 0.01 significance level unlike EBC and EP.

Table: Ratio of Sample Means - G1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Gamma Distributions, $\boldsymbol{gamma_1(\alpha_1 = 4, \lambda_1 = 1)}$, $\boldsymbol{gamma_2(\alpha_2} = 3,$ $\boldsymbol{\lambda_2 = 1)}$, **Bootstraps=10000**						
Sample size	5	10	15	20	30	40
E-skew	0.0121 (142%)	0.0069 (38%)	0.007 (40%)	0.0057 (14%)	0.0044 (12%)	**0.0047** **(6%)**
BT	**0.0072** **(44%)**	**0.0067** **(34%)**	**0.0061** **(22%)**	0.0064 (28%)	0.0042 (16%)	0.0057 (14%)
ET	0.0342 (584%)	0.0162 (224%)	0.0103 (106%)	0.0094 (88%)	0.0042 (16%)	0.0065 (30%)
BC	0.0536 (972%)	0.0235 (370%)	0.0151 (202%)	0.0128 (156%)	0.007 (40%)	0.0088 (76%)
EBC	0.0346 (592%)	0.0117 (134%)	0.0038 (24%)	**0.005** **(0%)**	0.0028 (44%)	0.0061 (22%)
BP	0.0452 (804%)	0.0202 (304%)	0.0125 (150%)	0.0105 (110%)	0.006 (20%)	0.0082 (64%)
EP	0.0297 (494%)	0.0078 (56%)	0.0019 (62%)	0.0029 (42%)	0.0016 (68%)	0.0037 (26%)
BS	0.0094 (88%)	0.0101 (102%)	0.0075 (50%)	0.0078 (56%)	**0.0049** **(2%)**	0.0071 (42%)
$\boldsymbol{BC_a}$	0.0506 (912%)	0.022 (340%)	0.0138 (176%)	0.0115 (130%)	0.0074 (48%)	0.0093 (86%)

Table: Ratio of Sample Means - G1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01) Gamma Distributions, $\boldsymbol{gamma_1(\alpha_1 = 4, \lambda_1 = 1)}$, $\boldsymbol{gamma_2(\alpha_2} = 3,$ $\boldsymbol{\lambda_2 = 1)}$, **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0145 (190%)	0.0105 (110%)	0.0099 (98%)	0.008 (60%)	0.0086 (72%)	0.0082 (64%)
BT	**0.0079 (58%)**	**0.0065 (30%)**	0.0061 (22%)	0.0075 (50%)	0.008 (60%)	0.0071 (42%)
ET	0.0372 (644%)	0.0153 (206%)	0.0116 (132%)	0.0101 (102%)	0.008 (60%)	0.0074 (48%)
BC	0.0482 (864%)	0.0201 (302%)	0.0144 (188%)	0.0121 (142%)	0.0101 (102%)	0.009 (80%)
EBC	0.0302 (504%)	0.0071 (42%)	0.0026 (48%)	0.003 (40%)	0.0031 (38%)	**0.0045 (10%)**
BP	0.0564 (1028%)	0.0216 (332%)	0.0162 (224%)	0.0144 (188%)	0.0117 (134%)	0.0105 (110%)
EP	0.0347 (594%)	0.0107 (114%)	**0.0048 (4%)**	**0.0052 (4%)**	**0.0049 (2%)**	0.006 (20%)
BS	0.0083 (66%)	0.0096 (92%)	0.0084 (68%)	0.0084 (68%)	0.0078 (56%)	0.0069 (38%)
$\boldsymbol{BC_a}$	0.0529 (958%)	0.0229 (358%)	0.0163 (226%)	0.0133 (166%)	0.0112 (124%)	0.0103 (106%)

Figure: Ratio of Sample Means: G99 - One-Sided Error Rates for 99% CI for the Gamma Distribution



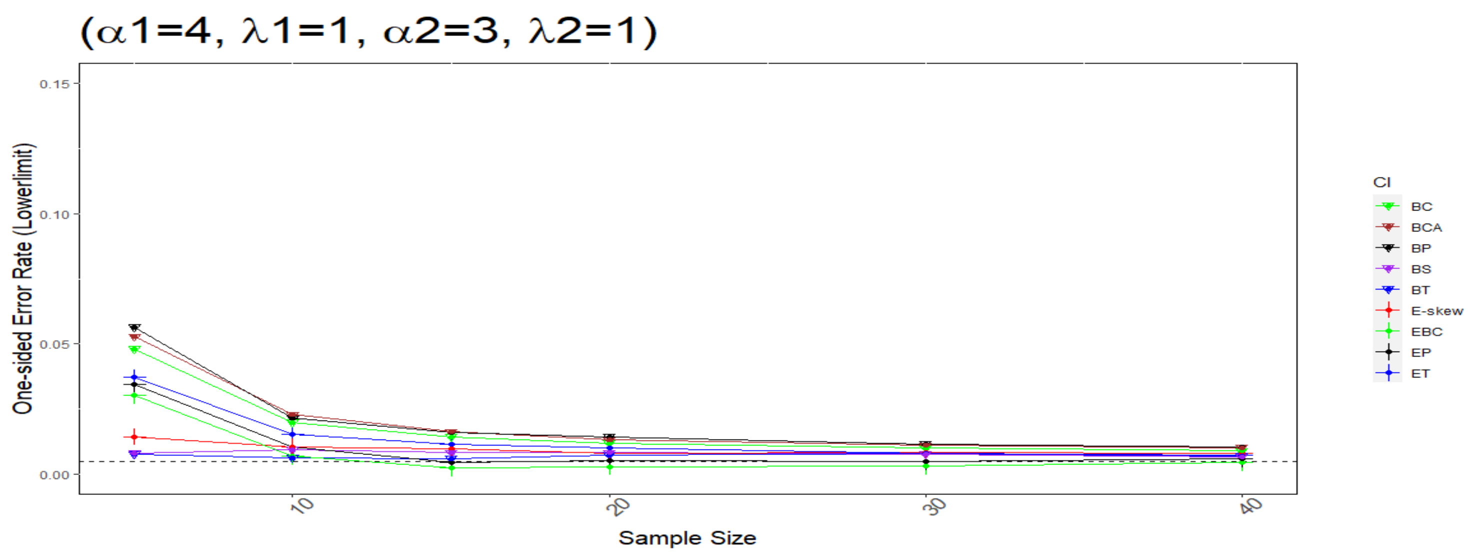
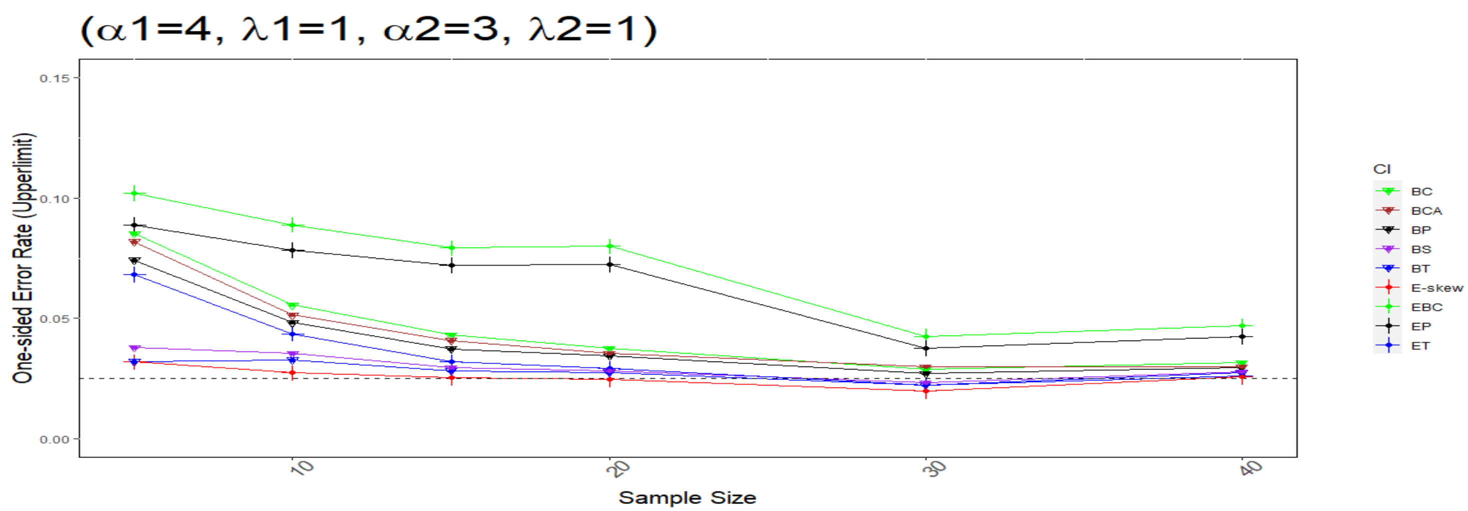


Figure: Ratio of Sample Means: G95 - One-Sided Error Rates for 95% CI for the Gamma Distribution



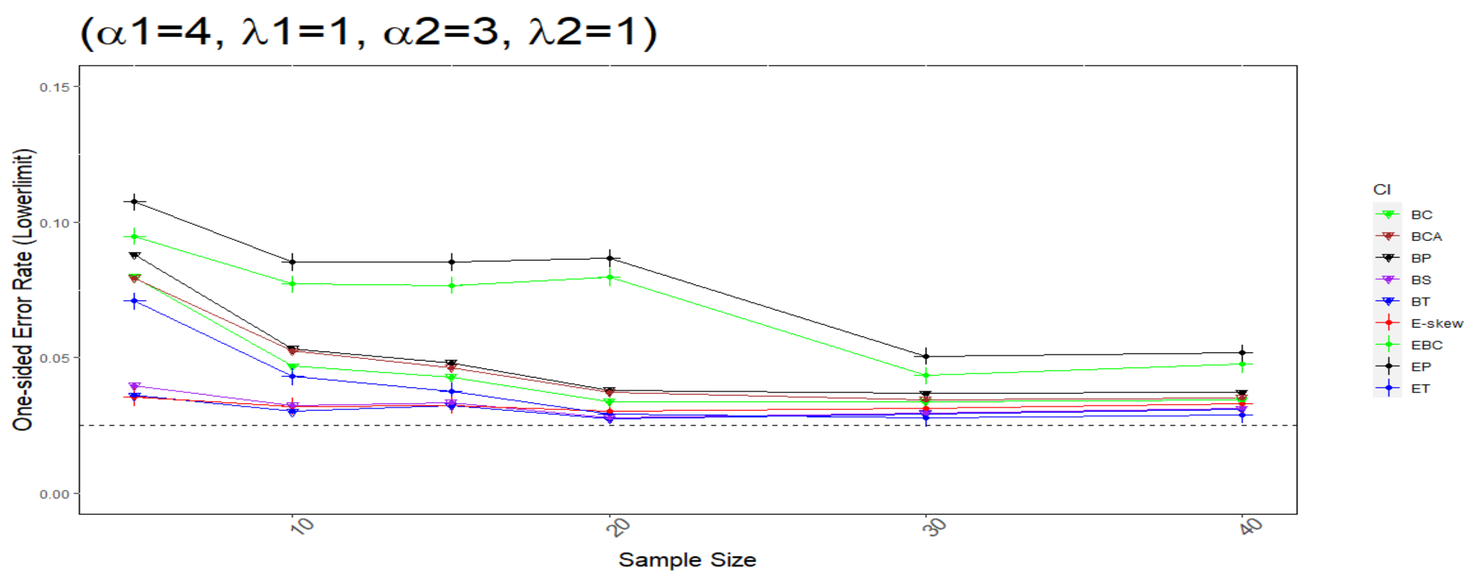
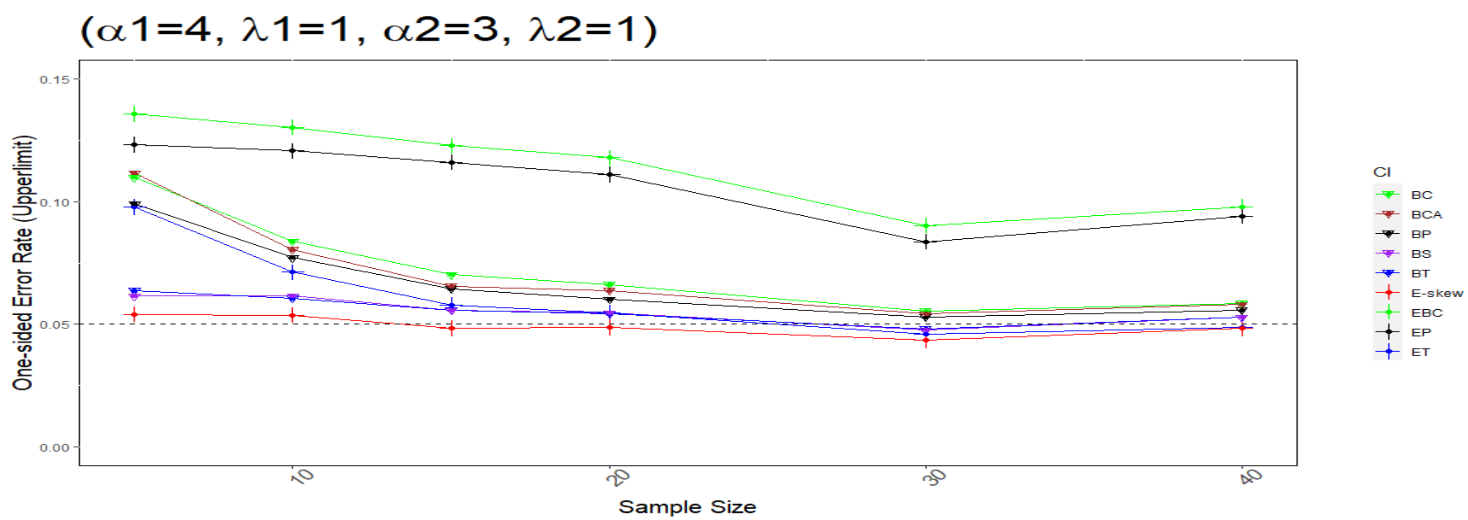
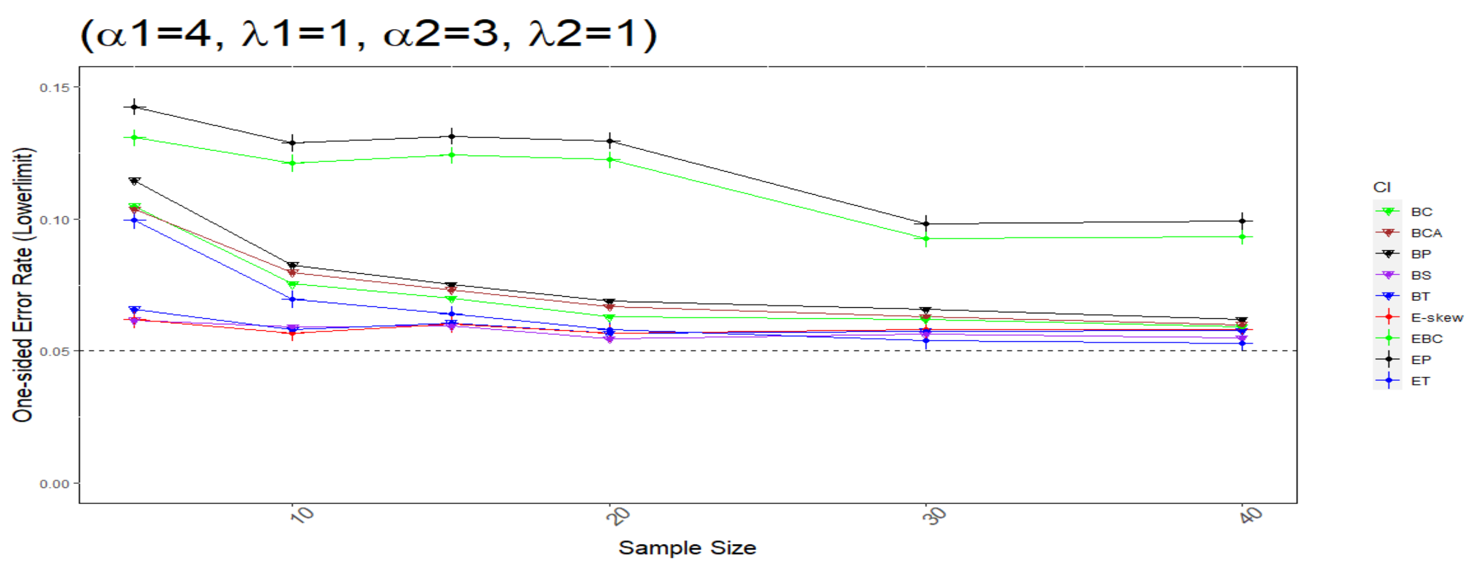


Figure: Ratio of Sample Means: G90 - One-Sided Error Rates for 90% CI for the Gamma Distribution





d. Log-Normal Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the ratio of means statistic when data is log-normally distributed. The second purpose is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for the ratio of means statistic when data is log-normally distributed. Data was generated from a pair of independent samples with parameter specifications: lognormal(μ=4, σ=0.2), lognormal(μ=3.3, σ=0.2).

The error results at the α = 0.01 significance level are displayed in tables LN1U99, and LN1L99 on pages 171 and 172 below. For the log-normal distribution, only one pair of independent samples was studied, error rate results at the α = 0.05 and α = 0.10 specification pair is displayed in Appendix tables. Although these tables only report results at the α = 0.01 significance level, results for the other two significance levels can be viewed visually in figures LN99, LN95, and LN90 on pages 173-175.

For the Lognormal(μ=4, σ=0.2), Lognormal(μ=3.3, σ=0.2) parameter specification pair at the specified α = 0.01 significance level, E-skew performed relatively more accurately when compared to other methods applied on EBSD(n). For both the upper and lower limit, E-skew had the error rate with the smallest percent error among methods applied on EBSD(n) at three sample sizes.

Other methods using EBSD(n) performed relatively accurately at the α = 0.01 significance level. At this significance level for the upper limit, EBC had the error rate with the smallest percent error at three sample sizes for the upper limit and two sample sizes for the lower limit. Also, at this significance level, EP had the error rate with the smallest percent error at sample size 40 for the upper and lower limit. EP's percent error was tied for the smallest with EBC. At the α = 0.01 significance level, EP also had the error rate with smallest percent error at sample size 15 for the lower limit. Conversely, ET did not perform accurately at the α = 0.01 significance. For both the upper and lower limit ET had an error rate with a larger percent error when compared to BT at each sample size.

Once again, the strength of the E-skew method is demonstrated when comparing the error rates across α significance level. E-skew improved relative accuracy as the α significance level increased from α = 0.01 to α = 0.10. In fact, E-skew was the relatively more accurate than every other method studied across sample size. At both the α = 0.05 and α = 0.10 significance levels for both the upper and lower limit E-skew attained the smallest percent error at each sample size compared to every other method studied.

By comparison, clearly the EP and EBC methods did not attain the error rate with the smallest percent error compared to every other method studied at any sample size for the α = 0.05 or α = 0.10 significance levels. In addition, EBC and EP had error rates with percent errors that were larger when compared to their Monte Carlo Bootstrap counterpart at each sample size. In addition, ET had error rates with percent errors that were larger when compared to their Monte Carlo Bootstrap counterpart at each sample size for all three significance levels.

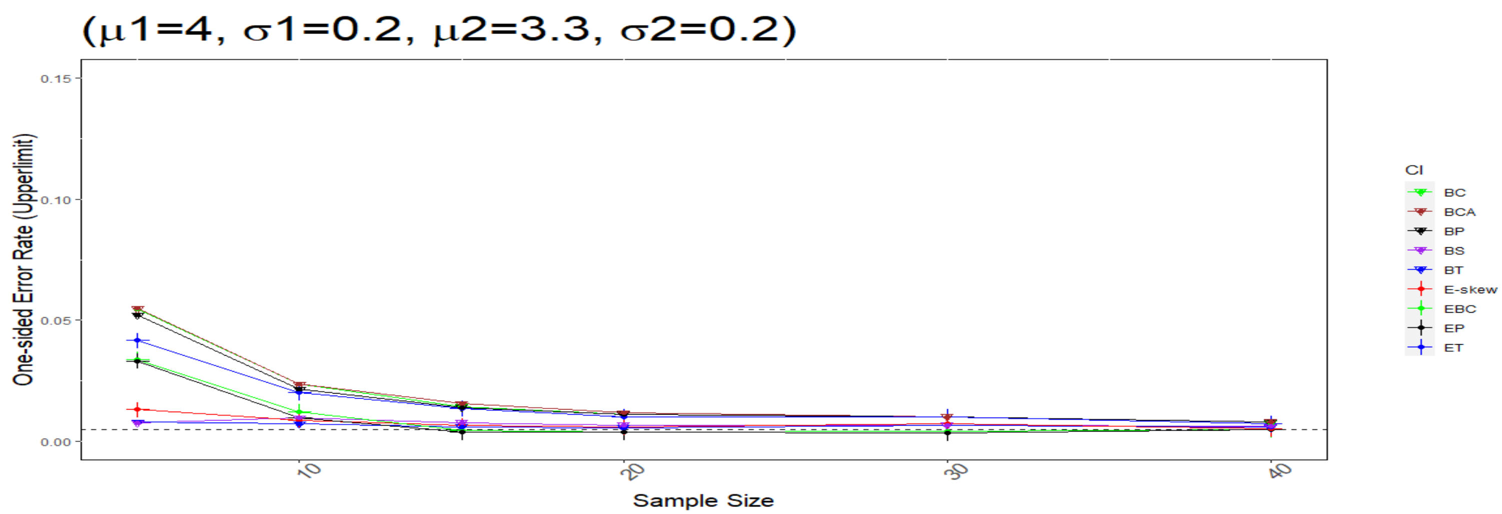
Table: Ratio of Sample Means - LN1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Log-Normal Distributions, $\boldsymbol{log-normal_1(\mu_1 = 4, \sigma_1 = 0.2)}$, $\boldsymbol{log-normal_2}(\boldsymbol{\mu_2} \approx 3.3$, $\boldsymbol{\sigma_2 = 0.2}$), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0132 (164%)	0.0089 (78%)	0.0066 (32%)	0.0059 (18%)	0.0073 (46%)	0.0052 (4%)
BT	0.008 (60%)	**0.0074** **(48%)**	0.0062 (24%)	**0.0057** **(14%)**	0.0068 (36%)	0.0059 (18%)
ET	0.0417 (734%)	0.0202 (304%)	0.0137 (174%)	0.0102 (104%)	0.0102 (104%)	0.0076 (52%)
BC	0.0548 (996%)	0.0236 (372%)	0.0145 (190%)	0.0113 (126%)	0.0102 (104%)	0.0076 (52%)
EBC	0.0339 (578%)	0.0123 (146%)	**0.0048** **(4%)**	0.0041 (18%)	**0.0043** **(14%)**	**0.0049** **(2%)**
BP	0.0522 (944%)	0.0216 (332%)	0.0141 (182%)	0.0114 (128%)	0.0102 (104%)	0.008 (60%)
EP	0.0333 (566%)	0.01 (100%)	0.0039 (22%)	0.0039 (22%)	0.0035 (30%)	**0.0051** **(2%)**
BS	**0.0077** **(54%)**	0.0097 (94%)	0.0079 (58%)	0.0069 (38%)	0.0068 (36%)	0.0063 (26%)
$\boldsymbol{BC_a}$	0.0549 (998%)	0.0236 (372%)	0.0157 (214%)	0.0119 (138%)	0.0103 (106%)	0.0082 (64%)

Table: Ratio of Sample Means - LN1L99 Lower Limit error rate (α = 0.01), Log-Normal Distributions, $\boldsymbol{log - normal_1(\mu_1 = 4, \sigma_1 = 0.2)}$, $\boldsymbol{log - normal_2}(\mu_2 \approx 3.3,$ $\boldsymbol{\sigma_2 = 0.2}$), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.011 (120%)	0.0078 (56%)	0.0071 (42%)	0.0061 (22%)	0.0044 (12%)	0.0056 (12%)
BT	0.0077 (54%)	**0.0059** **(18%)**	0.0066 (32%)	0.0059 (18%)	**0.0053** **(6%)**	0.0057 (14%)
ET	0.0413 (726%)	0.0185 (270%)	0.0129 (158%)	0.0108 (116%)	0.0084 (68%)	0.0082 (64%)
BC	0.053 (960%)	0.0219 (338%)	0.0147 (194%)	0.0114 (128%)	0.0083 (66%)	0.0076 (52%)
EBC	0.0347 (594%)	0.0106 (112%)	0.0027 (46%)	**0.0051** **(2%)**	0.0031 (38%)	**0.0052** **(4%)**
BP	0.0543 (986%)	0.0202 (304%)	0.0141 (182%)	0.0116 (132%)	0.0087 (74%)	0.0082 (64%)
EP	0.0353 (606%)	0.0085 (70%)	**0.0044** **(12%)**	0.0038 (24%)	0.002 (60%)	**0.0048** **(4%)**
BS	**0.0076** **(52%)**	0.0082 (64%)	0.008 (60%)	0.0071 (42%)	0.0067 (34%)	0.0057 (14%)
$\boldsymbol{BC_a}$	0.0526 (952%)	0.0214 (328%)	0.0152 (204%)	0.0122 (144%)	0.0097 (94%)	0.0085 (70%)

Figure: Ratio of Sample Means: LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution



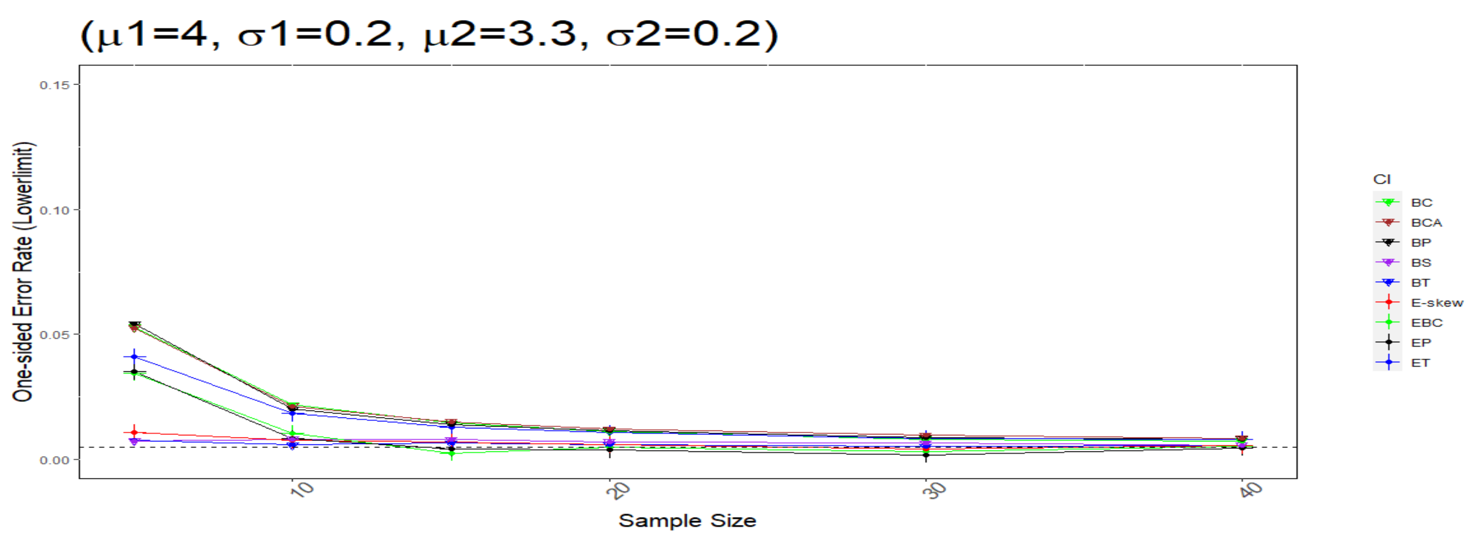
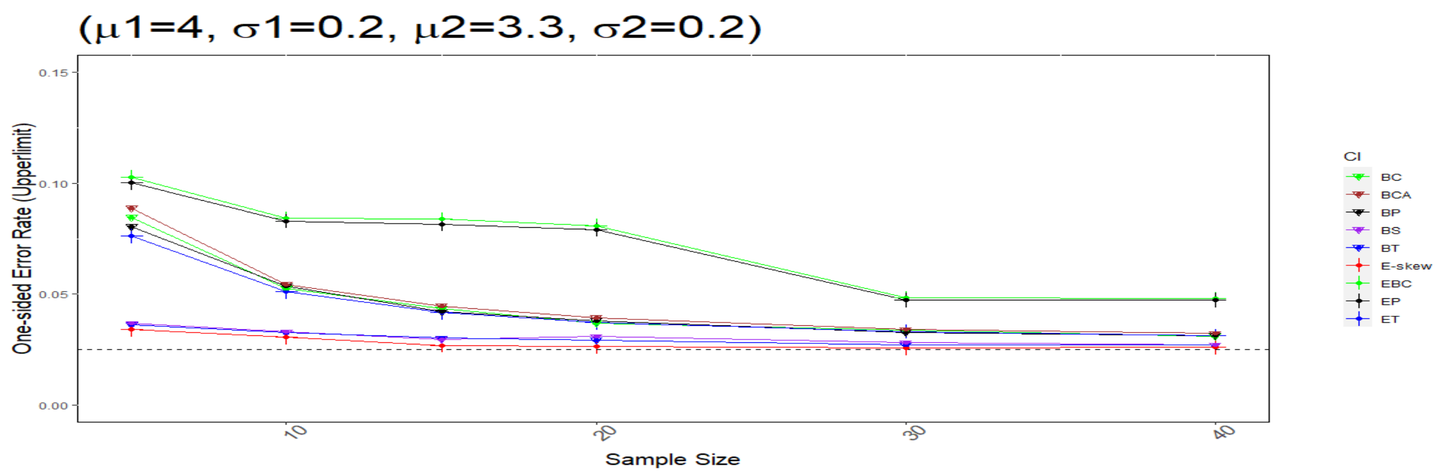


Figure: Ratio of Sample Means: LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution



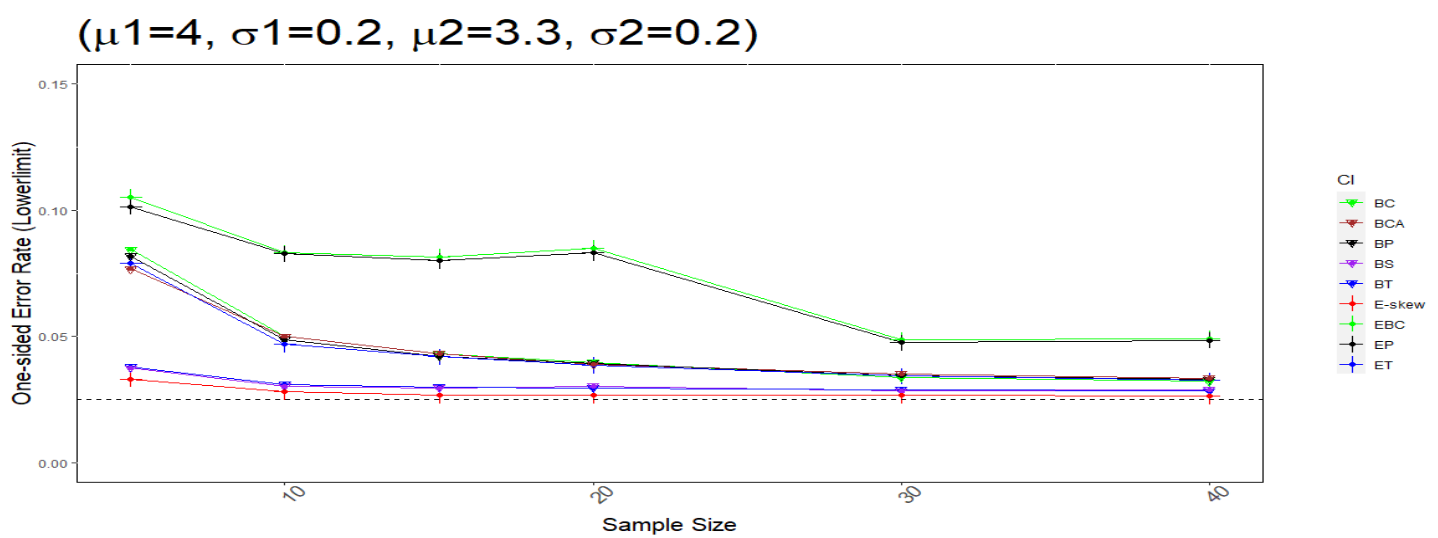
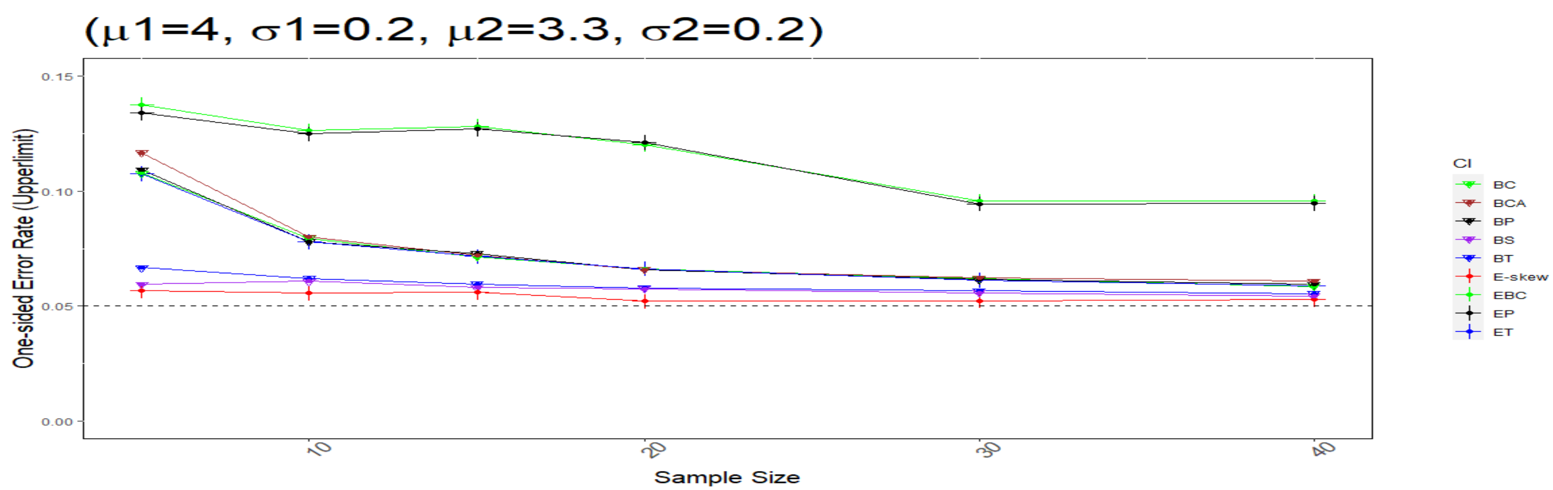
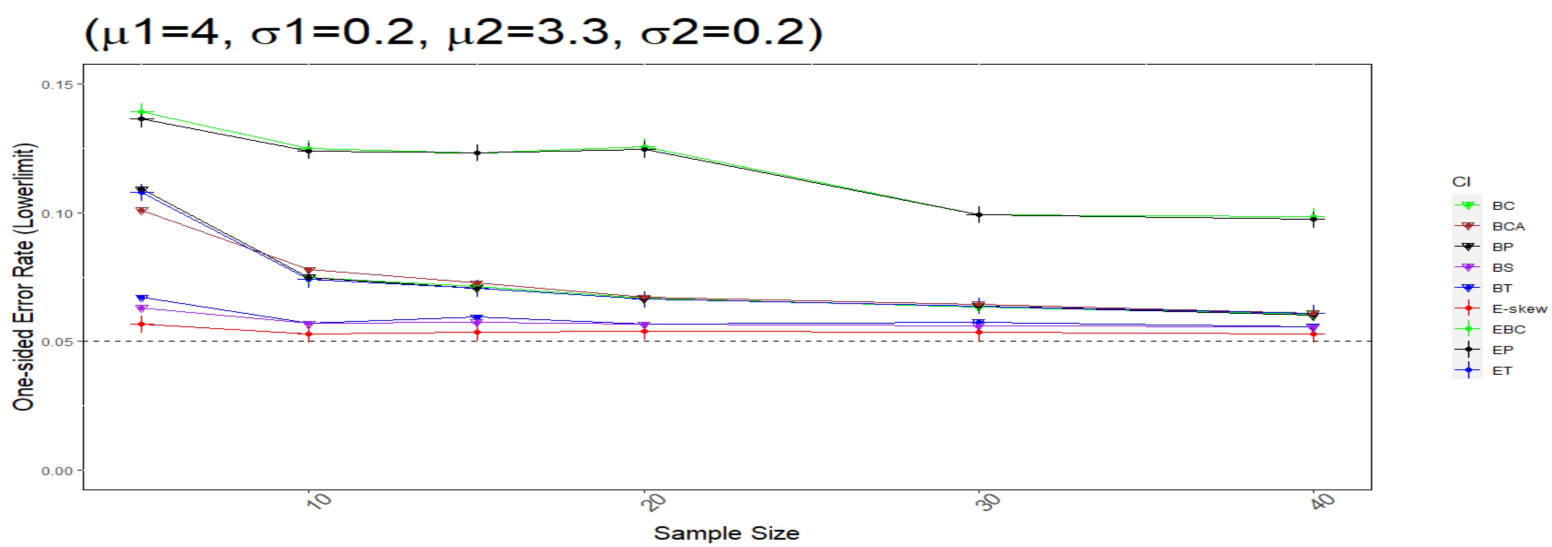


Figure: Ratio of Sample Means: LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution





e. Mixture Distribution

This sub-section compares E-skew error rates to the error rates of every other method studied for the ratio of means statistic when data is generated from a mixture of two normal distributions. Additionally, error rate comparisons for this statistic and distribution are made between other methods using EBSD(n) and Monte Carlo Bootstrap methods. Error rate comparisons are made from data generated from a pair of independent mixture of normal distributions 0.6*$N(\mu_1 = 50$, $\sigma_1 = 1)$ + 0.4*$N(\mu_1 = 100$, $\sigma_1 = 1)$ and 0.6*$N(\mu_1 = 25$, $\sigma_1 = 1)$ + 0.4*$N(\mu_1 = 50$, $\sigma_1 = 1)$. The error results at the α = 0.01 significance level are displayed in tables MN1U99, and MN1L99 on pages 178 and 179 below. For the mixture of normal distributions, only one pair of independent samples was studied. Error rate results for this specification pair at the α = 0.05 and α = 0.10 are displayed in Appendix tables. Although these tables only report results at the α = 0.01 significance level, results for the other two significance levels can be viewed visually in figures MN99, MN95, and MN90 on pages 180-182.

For the 0.6*$N(\mu_1 = 50$, $\sigma_1 = 1)$ + 0.4*$N(\mu_1 = 100$, $\sigma_1 = 1)$, 0.6*$N(\mu_1 = 25$, $\sigma_1 = 1)$ + 0.4*$N(\mu_1 = 50$, $\sigma_1 = 1)$ parameter specification pair at the specified α = 0.01 significance level, for the upper and lower limit, E-skew did not perform relatively as accurately as other methods applied on EBSD(n). Other than at sample size 5, E-skew had a larger percent error at each sample size for both the upper and lower limit at the α = 0.01 significance level compared to at least one EBSD(n) method.

Percentile methods using EBSD(n) performed relatively accurately compared to Monte Carlo Bootstrap methods at the α = 0.01 significance level. At this significance level for the upper limit, EBC had the error rate with the smallest percent error at sample size 15 for the upper limit. At this significance for the upper limit, EP had the error rate with the smallest percent error at sample size 20 and for the lower limit at both sample sizes 15 and 20. Conversely, ET did not perform as accurately relative to BT at the α = 0.01 significance level. For both the upper and lower limit ET had an error rate with a larger percent error when compared to BT at each sample size.

Once again, the strength of the E-skew method is demonstrated when comparing the error rates across α significance level. E-skew improved accuracy comparatively as the α significance level increased from α = 0.01 to α = 0.10. At the α = 0.05 significance level, for the upper limit E-skew attained the smallest percent error at sample size 10. Additionally, at the α = 0.05 significance level, for both the upper and lower limit, E-skew attained the error rate with the smallest percent error of any method applied on EBSD(n) at each sample size. At the α = 0.10 significance level for the upper limit, E-skew attained the error rate with the smallest percent error at sample size 5 and 10 and the smallest at each sample size among methods applied on EBSD(n). At the α = 0.10 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at sample size 5, 10 and 40 and the smallest at each sample size among methods applied on EBSD(n).

By comparison, EP and EBC methods did not attain the error rate with the smallest percent error at any sample size for the α = 0.05 or α = 0.10 significance levels. In addition, for the α = 0.10 significance level for both the upper and lower limit, EBC and EP had error rates with percent errors that were larger when compared to their Monte Carlo Bootstrap counterpart at each sample size. At the α = 0.05 significance level, the same was true except at sample size 5. Therefore, in these simulations, EP and EBC performed relatively accurately at the α = 0.01 significance level but performed relatively less accurately at the α = 0.05 and α = 0.10 significance levels.

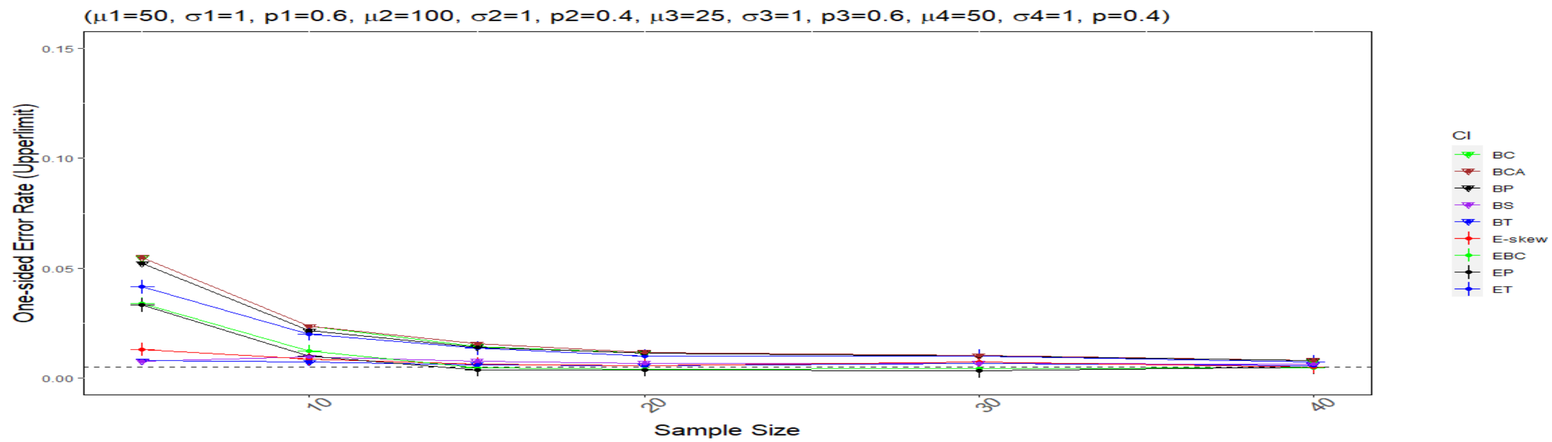
Table: Ratio of Sample Means - MN1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Mixture Distribution: $\boldsymbol{MN_1}$(0.6*$\boldsymbol{N}(\boldsymbol{\mu_1 = 50, \sigma_1 = 1})$ + 0.4*$\boldsymbol{N}(\boldsymbol{\mu_2 = 100,\ \sigma_2 = 1})$), $\boldsymbol{MN_2}$(0.6*$\boldsymbol{N}(\boldsymbol{\mu_1 = 25, \sigma_1 = 1})$ + 0.4*$\boldsymbol{N}(\boldsymbol{\mu_2 = 50,\ \sigma_2 = 1})$), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0139 (178%)	0.0118 (136%)	0.0089 (78%)	0.0115 (130%)	0.0085 (70%)	0.0071 (42%)
BT	**0.0072** **(44%)**	**0.0071** **(42%)**	0.008 (60%)	0.0079 (58%)	**0.0064** **(28%)**	**0.0052** **(4%)**
ET	0.0358 (616%)	0.0214 (328%)	0.015 (200%)	0.0148 (196%)	0.0098 (96%)	0.0083 (66%)
BC	0.0577 (1054%)	0.0239 (378%)	0.0186 (272%)	0.0159 (218%)	0.0111 (122%)	0.0085 (70%)
EBC	0.053 (960%)	0.0115 (130%)	**0.0028** **(44%)**	0.002 (60%)	2e-04 (96%)	0.0027 (46%)
BP	0.0557 (1014%)	0.0274 (448%)	0.0162 (224%)	0.0123 (146%)	0.009 (80%)	0.007 (40%)
EP	0.0309 (518%)	0.0084 (68%)	0.0021 (58%)	**0.0044** **(12%)**	0.0022 (56%)	0.0039 (22%)
BS	6e-04 (88%)	0 (100%)	2e-04 (96%)	0.0015 (70%)	0.0026 (48%)	0.0032 (36%)
$\boldsymbol{BC_a}$	0.0607 (1114%)	0.035 (600%)	0.0156 (212%)	0.0103 (106%)	0.0072 (44%)	0.0054 (8%)

Table: Ratio of Sample Means - MN1L99 Lower limit error rate (α = 0.01), Mixture Distribution: MN_1(0.6*$N(\mu_1 = 50, \sigma_1 = 1)$ + 0.4*$N(\mu_2 = 100, \ \sigma_2 = 1)$), MN_2(0.6*$N(\mu_1 = 25, \sigma_1 = 1)$ + 0.4*$N(\mu_2 = 50, \ \sigma_2 = 1)$), **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0138 (176%)	0.0127 (154%)	0.0104 (108%)	0.009 (80%)	0.0087 (74%)	0.0072 (44%)
BT	**0.0072** **(44%)**	**0.0081** **(62%)**	0.0086 (72%)	**0.0066** **(32%)**	0.0066 (32%)	**0.0049** **(2%)**
ET	0.0366 (632%)	0.021 (320%)	0.0147 (194%)	0.0129 (158%)	0.0105 (110%)	0.0085 (70%)
BC	0.0565 (1030%)	0.0253 (406%)	0.0178 (256%)	0.0141 (182%)	0.0103 (106%)	0.0083 (66%)
EBC	0.0515 (930%)	0.012 (140%)	0.0033 (34%)	7e-04 (86%)	3e-04 (94%)	0.0024 (52%)
BP	0.0583 (1066%)	0.0268 (436%)	0.0151 (202%)	0.0109 (118%)	0.0088 (76%)	0.0067 (34%)
EP	0.0322 (544%)	0.0084 (68%)	**0.0035** **(30%)**	**0.0034** **(32%)**	0.0023 (54%)	0.004 (20%)
BS	7e-04 (86%)	0 (100%)	5e-04 (90%)	7e-04 (86%)	0.0025 (50%)	0.0027 (46%)
BC_a	0.0552 (1004%)	0.0339 (578%)	0.0142 (184%)	0.0074 (48%)	**0.0063** **(26%)**	0.0057 (14%)

Figure: Ratio of Sample Means: MN99 - One-Sided Error Rates for 99% CI for the Mixture of two Normals



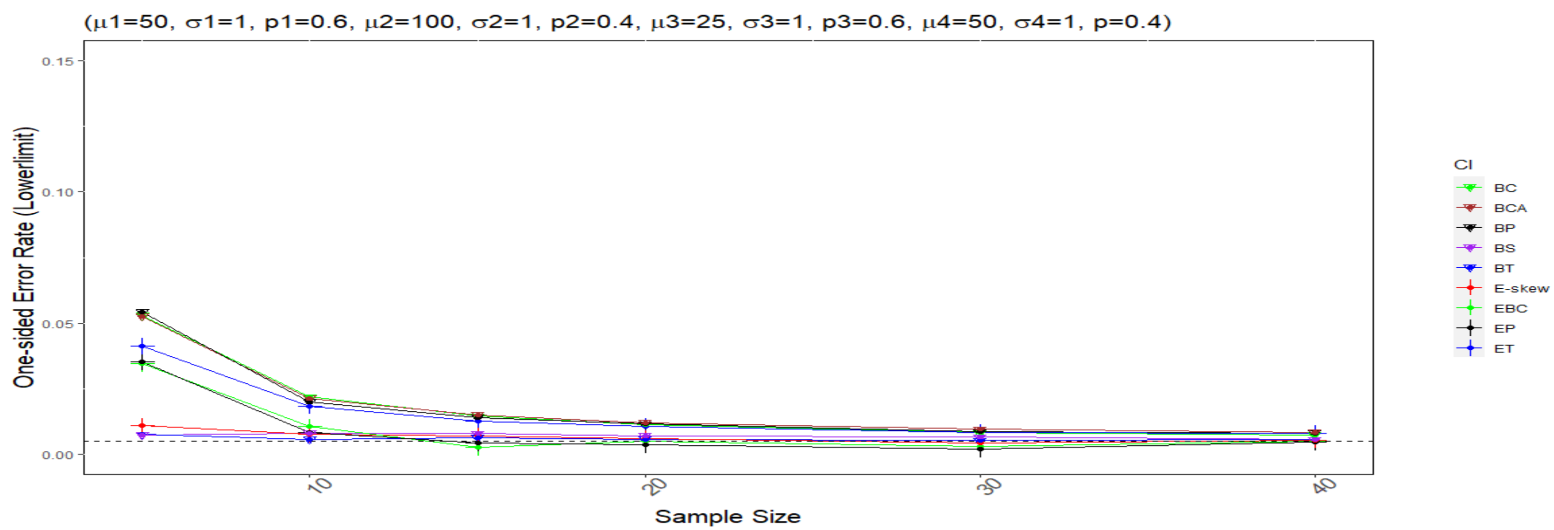
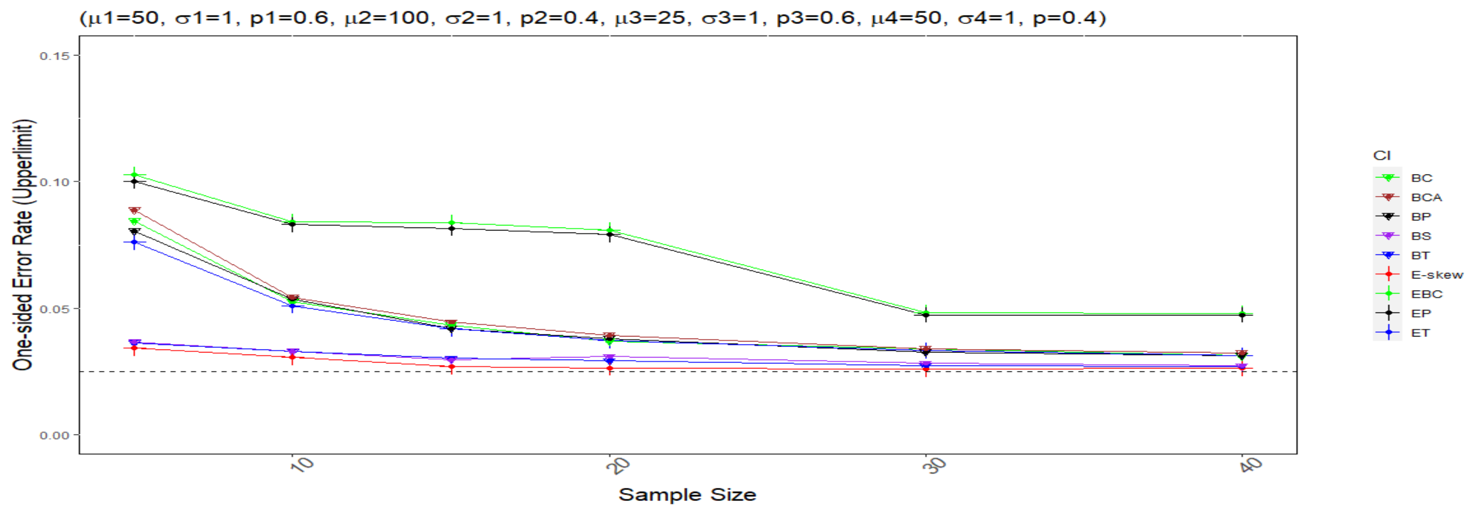


Figure: Ratio of Sample Means: MN95 - One-Sided Error Rates for 95% CI for the Mixture of two Normals



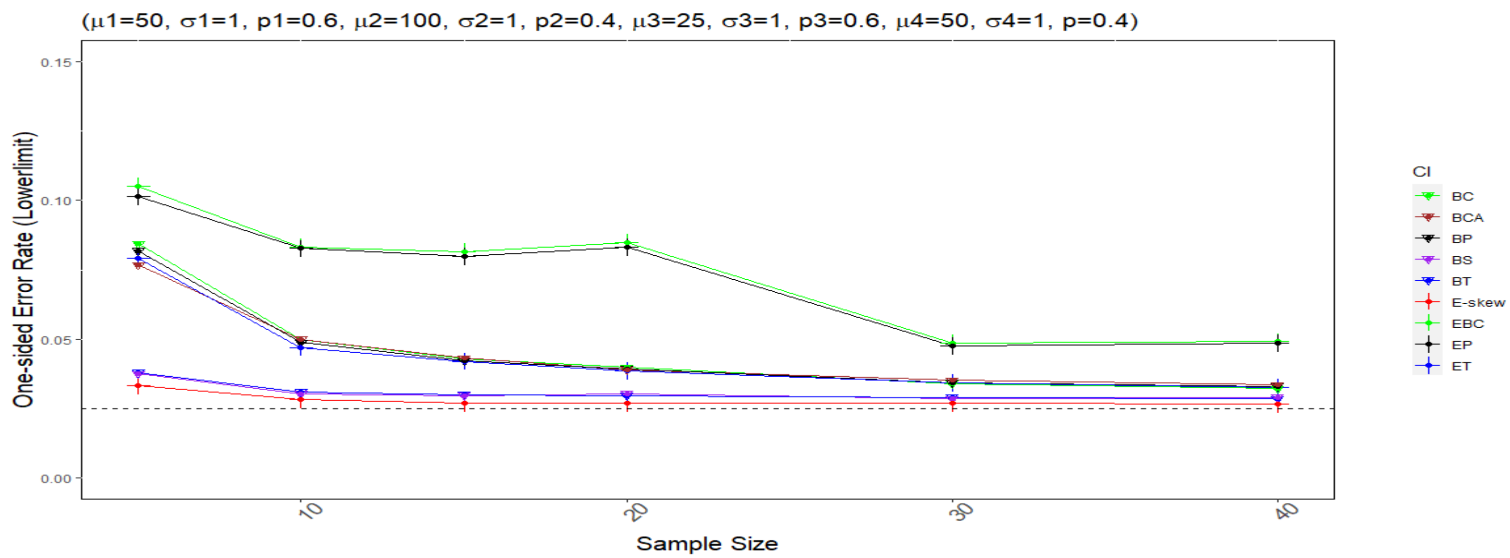
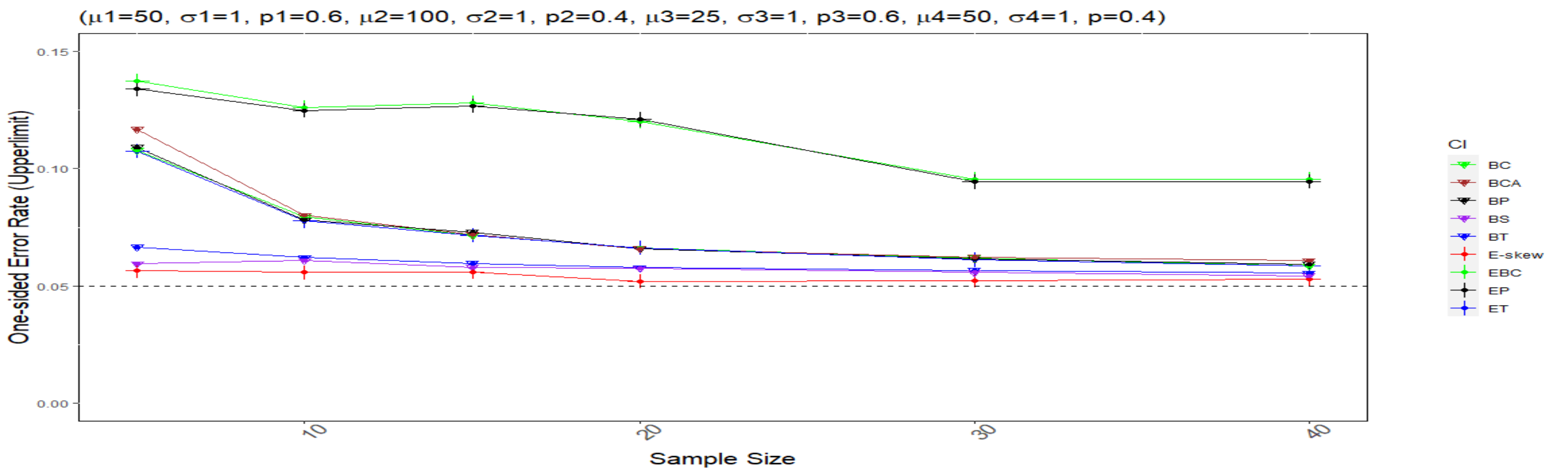
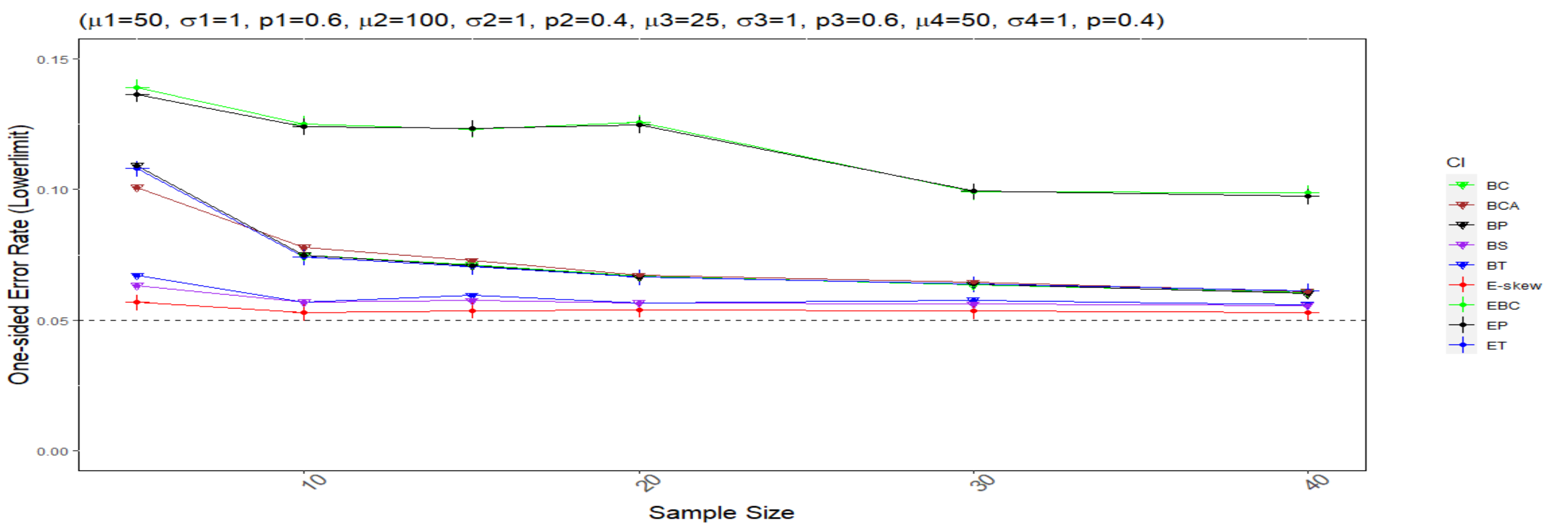


Figure: Ratio of Sample Means: MN90 - One-Sided Error Rates for 90% CI for the Mixture of two Normals





Ratio of Sample Means Results for E-skew and other methods using EBSD(n)

In general, the E-skew method was relatively as accurate, if not more accurate, for the ratio of sample means statistic than it was for the sample mean statistic. For example, when data was generated from the normal distribution with an expected ratio of means value of 2, when compared to every method studied for this statistic of the 36 sample size cases (6 sample sizes*2 limit ends*3 significance levels) studied for data generated from the normal distribution, E-skew had the error rate with the smallest percent error for 23 of them. E-skew in particular, performed more accurately at the α = 0.05 and α = 0.10 significance levels.

For data that was exponentially distributed, E-skew again performed relatively more accurately at the α = 0.05 and α = 0.10 significance levels than at the α = 0.01 significance level. At the α = 0.01 significance level, E-skew performed relatively less accurately compared to at least one other method using EBSD(n). At the α = 0.05 and α = 0.10 significance levels, E-skew's accuracy improved relative to other methods it was compared to. At the α = 0.05 and α = 0.10 significance levels E-skew was relatively more accurate than every other methods applied on EBSD(n). Further at larger sample sizes at these significance levels, it was also relatively more accurate than both methods applied on EBSD(n) and Monte Carlo Bootstrap methods.

For data that was log-normally distributed α = 0.05 and α = 0.10 significance levels, E-skew performed relatively more accurately at each sample size for both the upper and lower limit than every other method it was compared to.

For data generated from a mixture of two normal distributions E-skew performed the most accurately among methods applied on EBSD(n) at the α = 0.05 and α = 0.10 significance levels. Additionally, in a few instances it was more accurate not only for all methods applied on EBSD(n) but all Monte Carlo Bootstrap methods as well at these significance levels. At the α = 0.01 significance level, at least one method applied on EBSD(n) was more accurate at nearly every sample size.

Other percentile methods applied on EBSD(n) performed better at the α = 0.01 significance level. At this significance level the percentile methods applied on EBSD(n) outperformed their Monte Carlo Bootstrap counterpart frequently for both the upper and lower limit. However, at the α = 0.05 and α = 0.10 significance levels, like in the case of the sample mean, methods applied on EBSD(n) other than E-skew performed less accurately than Monte Carlo Bootstrap confidence intervals. The ET method performed relatively as accurately across α level for exponential and gamma distributed data. For data generated both from the exponential and gamma distribution it performed accurately in comparison to other methods applied on EBSD(n) at each significance level. For data generated from the normal, log-normal, and mixture of two normal ET performed relatively less accurate across sample size and significance level compared to BT.

4.3 Pearson Correlation Coefficient

For the Pearson correlation coefficient portion of the simulation study, results for six different sample sizes are reported (n = 5, 10, 15, 20, 30, and 40). For each of these sample sizes, confidence interval error rates are reported at the α = 0.01, 0.05, and 0.10 significance levels.

The probability distributions used in the simulation study for the Pearson correlation coefficient were three bivariate normal distributions and two bivariate non-normal skewed distributions. For each distribution, the population parameters specified are displayed below in Table 4.3. These parameter specifications are the same as the specifications for the sample mean in Table 3.5 in Chapter 3.

For each sample size, population parameter specification, and probability distribution combination 10,000 separate samples were generated. For the Monte Carlo Bootstrap confidence interval methods each of the 10,000 samples used 10,000 Monte Carlo Bootstrap resamples to create its Bootstrap sampling distribution. The comparisons discussed in this section are made between EBSD(n) methods and Bootstrap methods that use 10,000 Bootstrap resamples. In addition, confidence interval method error rate results were measured on the same 10,000 unique samples using 200, 500, and 1,000 Bootstrap resamples. These alternative Bootstrap resampling levels were performed for each distribution studied here. The error rate results at these additional Bootstrap resampling levels are reported in the Appendix. Each generated unique sample had confidence intervals computed using the confidence interval methods listed below.

- For methods using EBSD(n) this included: E-skew, ET, EBC, and EP
- For methods using the Monte Carlo Bootstrap this includes: BT, BC, BP, BC_a, and BS.

Each method in this section used the fisher transformation. A sampling distribution was computed on the fisher transformation of the Pearson correlation coefficient and then the resulting confidence interval limits were back transformed.

Below in Table 4.3 is a description of the parameter specifications used for the sample mean statistic in this simulation study:

Table 4.3 Simulation Parameter Specifications for the Pearson Correlation Coefficient		
Probability distribution	Population Parameter	Parameter code: Specified Parameter Values
Normal distribution	$(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$	N1: (4, 1, 4, 1, 0.1) N2: (4, 1, 4, 1, 0.5) N3: (4, 1, 4, 1, 0.9)
Non-Normal distribution	(Skew, Kurtosis, ρ)	NN1: (3, 61, 0.1) NN2: (3, 61, 0.5)

It can be noted in the tables below and in the tables in the appendix, the error rate results for the BC and BS methods are identical. This is because the fisher transformation was used for the Pearson correlation coefficient and hence a constant value was specified for the formula's sample variance and therefore the Boot package in R yields the same result for each method. This is further because for example the numerator of the t-statistic for the $\frac{\alpha}{2}$ end of the sample distribution is: ($\hat{\theta}^*_{\alpha/2}$-$\hat{\theta}$). Then when this quantity is applied to the t-statistic formula, the constant value in the denominator is eliminated by multiplying by the standard error and the lower end point becomes: $\hat{\theta}$ - ($\hat{\theta}^*_{\alpha/2}$-$\hat{\theta}$) = 2$\hat{\theta}$ - ($\hat{\theta}^*_{\alpha/2}$). This is the same as the $\alpha/2$ percentile from the Basic sampling distribution.

a. Bivariate Normal Distribution

The first purpose of this section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the Pearson correlation coefficient on bivariate normally distributed data. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods again for the Pearson correlation coefficient for bivariate normally distributed data.

For the bivariate normal distribution, simulations using three different specifications were performed. First the results for data generated from a bivariate normal distribution with specification $N(\mu = \binom{4}{4}, \Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix})$ at the α = 0.01 significance level are discussed. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables N1U99 and N1L99 on pages 190 and 191 below. For the Pearson correlation coefficient three separate pairs of independent samples were generated from the bivariate normal distribution at three α significance levels. However, in this section because of the volume of error rate results, only the error rates reported at the α = 0.01 significance level for the bivariate normal distribution with covariance matrix $\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$ are displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the $N(\mu = \binom{4}{4}, \Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix})$ specification, the $N(\mu = \binom{4}{4}, \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix})$ and the $N(\mu = \binom{4}{4}, \Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix})$ results can be viewed visually in figures NU99, NL99, NU95, NL95, NU90 and NL90 on pages 194-199. In these figures, the dashed horizontal line represents the target nominal one-sided error rate based on the α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the ($\mu = \binom{4}{4}$, $\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$) parameter specification at the specified α = 0.01 significance level, for the upper and lower limit, E-skew was relatively more accurate compared to both Monte Carlo Bootstrap and EBSD(n) methods across sample size. E-skew had the error rate with the smallest percent error at sample sizes 20, 30 and 40 for the upper limit. For the lower limit E-skew had the error rate with the smallest percent error at sample sizes 5, 20, 30 and 40. E-skew also had the error rate with the smallest percent error of any method using EBSD(n) studied at every sample size for both the upper limit and lower limit. Otherwise, either BC_a or BS/BC had the error rate with the smallest percent error at the remaining sample sizes greater than sample size 5, for both the upper and lower limit. These results are shown below in N1U99 and N1L99. These results can also be viewed visually in figures NU99 and NL99.

E-skew performed even more accurately at the α = 0.05 and α = 0.10 significance level than it did at the α = 0.01 significance level. The E-skew method was even more accurate relative the other methods studied as the α significance level was modified from α = 0.01 to α = 0.10. At the α = 0.01 significance level for both the upper and lower limit, E-skew attained the error rate with the smallest percent error for four sample sizes. At both the α = 0.05 and α = 0.10 significance level for the lower limit, E-skew attained the error rate with the smallest percent error at five sample sizes. At the α = 0.05 significance level for the upper limit, E-skew attained the error rate with the smallest percent error for five sample sizes. At the α = 0.10 significance level for the upper limit, E-skew attained the error rate with the smallest percent error for four sample sizes.

Other than EBC for sample size 10 at the α = 0.01 significance level for the upper limit, no method other than E-skew using EBSD(n) attained the error rate with the smallest percent error at any other sample size for any of the three significance levels for either the upper or lower limit. Additionally, EP was relatively more accurate at the α = 0.01 significance level and relatively less accurate at the α = 0.05 or α = 0.10 significance levels when compared to BP. At the α = 0.01 significance level for the lower limit, EP attained an error rate with a smaller percent error at five sample sizes compared to BP. However at the α = 0.05 significance level EP accomplished this for only two sample sizes and at the α = 0.10 significance level accomplished this for one sample size. EBC's relatively accuracy to BC to did not change across α significance level. For both the α = 0.01 and α = 0.10 significance levels for both the upper and lower limit, EBC only attained the error rate with a smaller percent error at one sample size. The ET method was ineffective as the resulting confidence interval was too large at each significance level leading to error rates of 0 for both the upper and lower limit at every sample size but sample size 5.

Table: Pearson Correlation Coefficient - N1U99 Upper limit error rate (α = 0.01), Bivariate Normal Distributions, $MVN(\mu_1 = 4, \sigma_1 = 1, \mu_2 = 4, \sigma_2 = 1, \rho = 0.1)$, **Bootstraps=10000**

Sample size	5	10	15	20	30	40
E-skew	0.0065 (30%)	0.0068 (36%)	0.0043 (14%)	**0.0062 (24%)**	**0.0059 (18%)**	**0.0053 (6%)**
BT	0 (100%)	0.0111 (122%)	0.0102 (104%)	0.0108 (116%)	0.0082 (64%)	0.0086 (72%)
ET	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BC	1e-04 (98%)	0.0018 (64%)	**0.0045 (10%)**	0.0074 (48%)	0.0068 (36%)	0.0065 (30%)
EBC	8e-04 (84%)	1e-04 (98%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BP	**0.0062 (24%)**	0.0132 (164%)	0.0115 (130%)	0.0108 (116%)	0.0085 (70%)	0.009 (80%)
EP	3e-04 (94%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BS	1e-04 (98%)	0.0018 (64%)	**0.0045 (10%)**	0.0074 (48%)	0.0068 (36%)	0.0065 (30%)
BC_a	0.0014 (72%)	**0.006 (20%)**	0.0071 (42%)	0.0087 (74%)	0.0074 (48%)	0.007 (40%)

Table: Pearson Correlation Coefficient - N1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Bivariate Normal Distributions, $\boldsymbol{MVN}(\boldsymbol{\mu_1} = 4, \boldsymbol{\sigma_1} = 1, \boldsymbol{\mu_2} = 4, \boldsymbol{\sigma_2} = 1, \rho = 0.1)$, **Bootstraps=10000**						
Sample size	5	10	15	20	30	40
E-skew	**0.0068** (**36%**)	0.006 (20%)	0.0058 (16%)	**0.0048** (**4%**)	**0.005** (**0%**)	**0.0056** (**12%**)
BT	0 (100%)	0.0088 (76%)	0.0111 (122%)	0.0108 (116%)	0.0083 (66%)	0.009 (80%)
ET	4e-04 (92%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BC	0 (100%)	0.0015 (70%)	**0.0045** (**10%**)	0.0058 (16%)	0.0065 (30%)	0.0079 (58%)
EBC	0.0016 (68%)	0.0014 (72%)	2e-04 (96%)	2e-04 (96%)	0 (100%)	0 (100%)
BP	0.0161 (222%)	0.0131 (162%)	0.0131 (162%)	0.0118 (136%)	0.0092 (84%)	0.0103 (106%)
EP	0.0074 (48%)	0.0025 (50%)	4e-04 (92%)	5e-04 (90%)	1e-04 (98%)	5e-04 (90%)
BS	0 (100%)	0.0015 (70%)	**0.0045** (**10%**)	0.0058 (16%)	0.0065 (30%)	0.0079 (58%)
$\boldsymbol{BC_a}$	0.0031 (38%)	**0.0048** (**4%**)	0.0073 (46%)	0.008 (60%)	0.0072 (44%)	0.0085 (70%)

Simulation were not only performed for the ($\mu = \binom{4}{4}$, $\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$) parameter specification. Simulations were also performed for the ($\mu = \binom{4}{4}$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$) and ($\mu = \binom{4}{4}$, $\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$) specifications. When considering a change in parameter specification (i.e. increasing the correlation between the paired samples from 0.1 to 0.5) E-skew did not perform relatively as accurately compared as it did when ρ = 0.10. When ρ was set to 0.50, at the α = 0.01 significance level for the upper limit, E-skew had the error rate with the smallest percent error at five of the six sample sizes measured. However, for the lower limit at this significance level, it attained the error rate with the smallest percent error only at three sample sizes. At the α = 0.05 significance level, for the upper and lower limit respectively, E-skew had the error rate with the smallest percent error at four and two sample sizes. At the α = 0.10 significance level, E-skew performed even less relatively accurate than when ρ = 0.10, attaining the error rate with the smallest percent error at one sample size for the upper limit and three for the lower limit.

When the correlation was increased even further to 0.90, for the upper limit, E-skew performed even less relatively accurate compare to the other methods studied at each significance level than it did when the correlation was increased to 0.50. At the α = 0.01 significance level for the upper limit, E-skew did attain the error rate with the smallest percent error at four of six sample sizes. However at the α = 0.05 and α = 0.10 significance levels, E-skew did not attain the error rate with the smallest percent error for any sample size for the upper limit. Further for the lower limit, at all three significance levels, E-skew did not attain the error rate with the smallest percent error at any sample size.

Other methods applied on EBSD(n) performed more accurately relative to E-skew and Monte Carlo Bootstrap methods when the correlation was increased to 0.90. At the α = 0.01 significance level for the lower limit, EBC attained an error rate with a smaller percent error at one sample size. At the α = 0.05 significance level, for the upper and lower limit respectively, EBC attained this for one and three sample sizes respectively. Additionally, at the α = 0.05 and α = 0.10 significance levels for the upper limit, EP attained the error rate with the smallest percent error at one sample size.

When comparing EBSD(n) methods to their Monte Carlo Bootstrap counterpart at the specified correlation value of ρ=0.90, the EBSD(n) methods relative accuracy varied depending on the significance level. The EP method performed more accurately relative to BP for the lower limit at the α = 0.01 significance level, attaining an error rate with a smaller percent error at each sample size. At the α = 0.05 and α = 0.10 significance levels EP performed relatively less accurately compared to BP for the majority of sample sizes for the lower limit.

EBC performed relatively less accurate compared to BC at the α = 0.01 significance level for the upper limit. EBC attained an error rate with a smaller percent error at three sample sizes compared to BC for the upper limit at the α = 0.01 significance level. For the lower limit, EBC performed relatively more accurately than BC at the α = 0.05 significance level, and relatively less accurately at the α = 0.01 significance level. At the α = 0.01 significance level for the lower limit, EBC had an error rate with a smaller percent error compared to BC only at one sample size. By comparison at the α = 0.05 significance level for the lower limit, it had an error rate with a smaller percent error than BC for three sample sizes. ET had an error rate of 0 at every significance level for every sample size for the upper limit except at sample size 5 at the α = 0.10 significance level. For the lower limit it had a significance level that was non-zero for one sample size at the α = 0.01 significance level, two sample sizes at the α = 0.05 significance level, and three sample sizes at the α = 0.10 significance level.

Figure: Pearson Correlation Coefficient - NU99 - One-Sided Upperlimit Error Rates for 99% CI for the Bivariate Normal Distribution

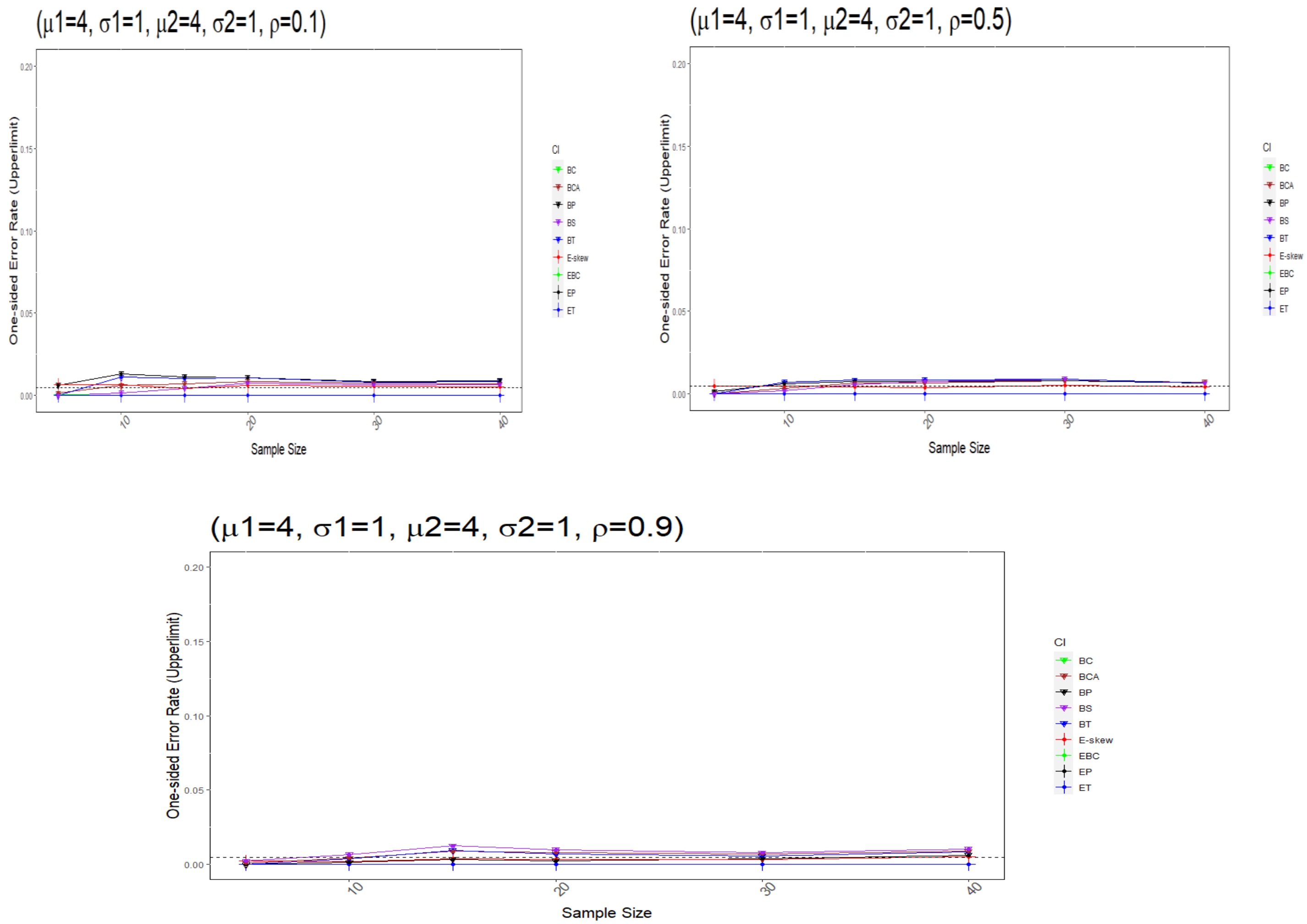


Figure: Pearson Correlation Coefficient - NL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Bivariate Normal Distribution

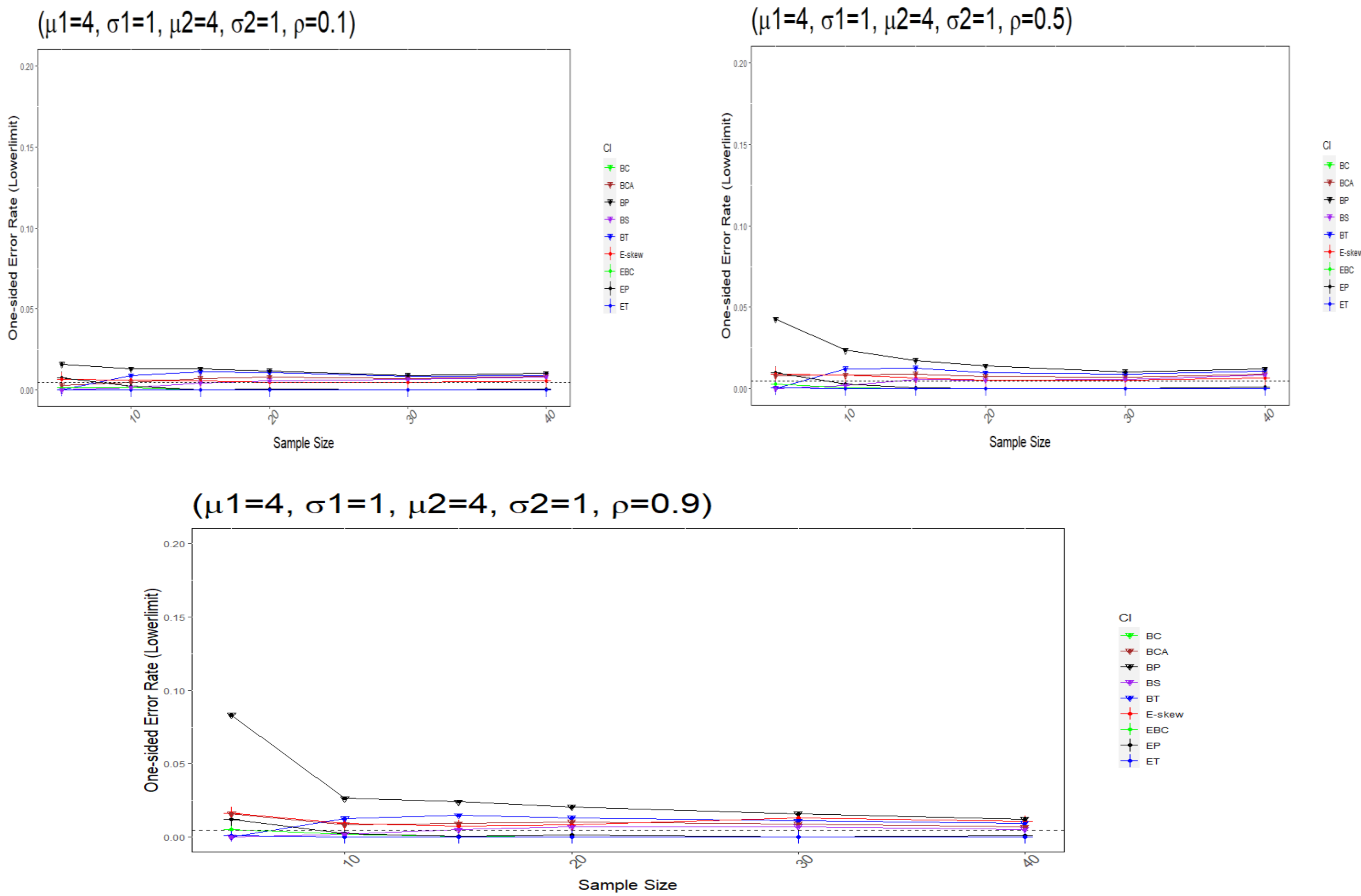


Figure: Pearson Correlation Coefficient - NU95 - One-Sided Upperlimit Error Rates for 95% CI for the Bivariate Normal Distribution

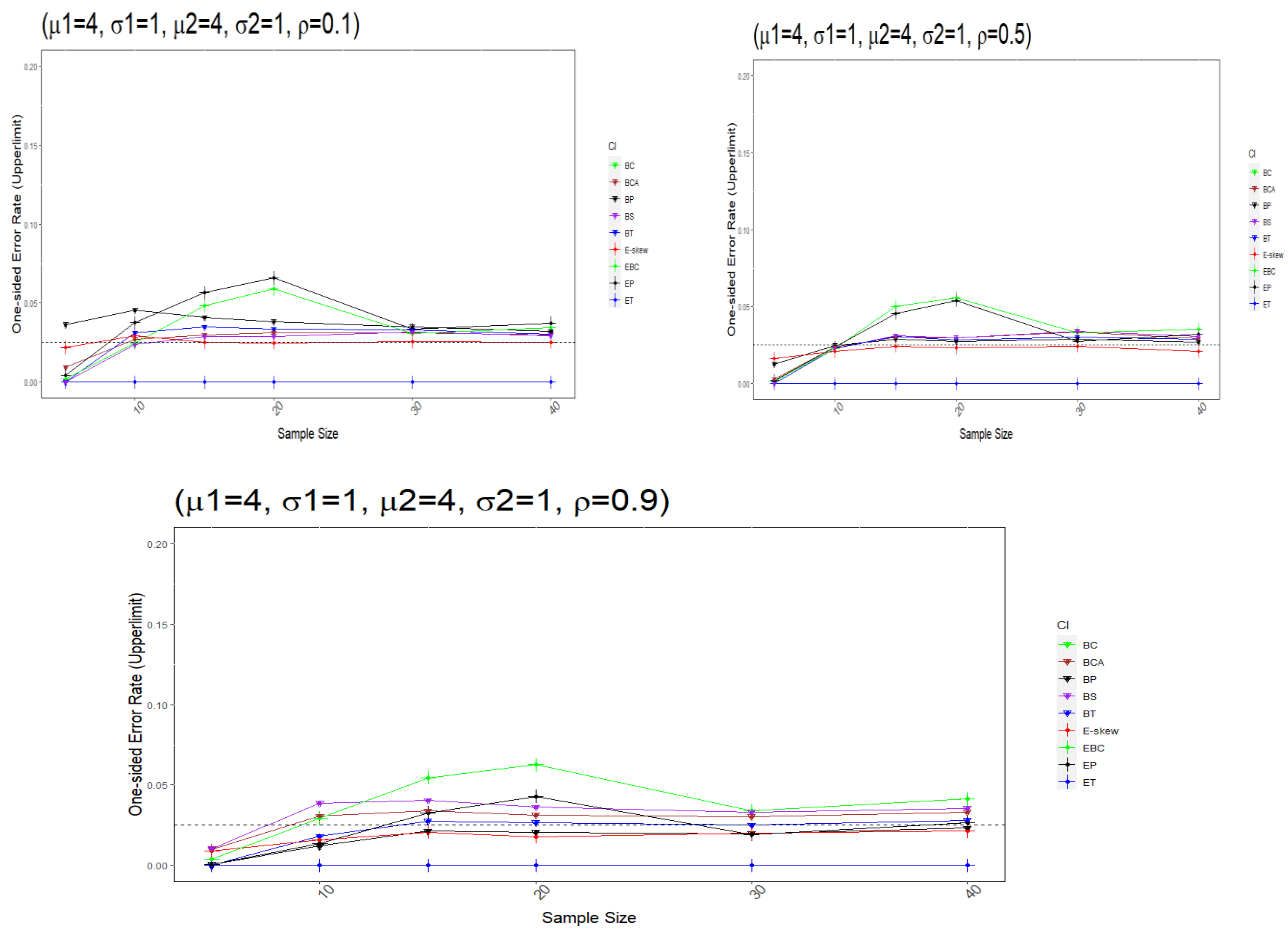


Figure: Pearson Correlation Coefficient - NL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Bivariate Normal Distribution

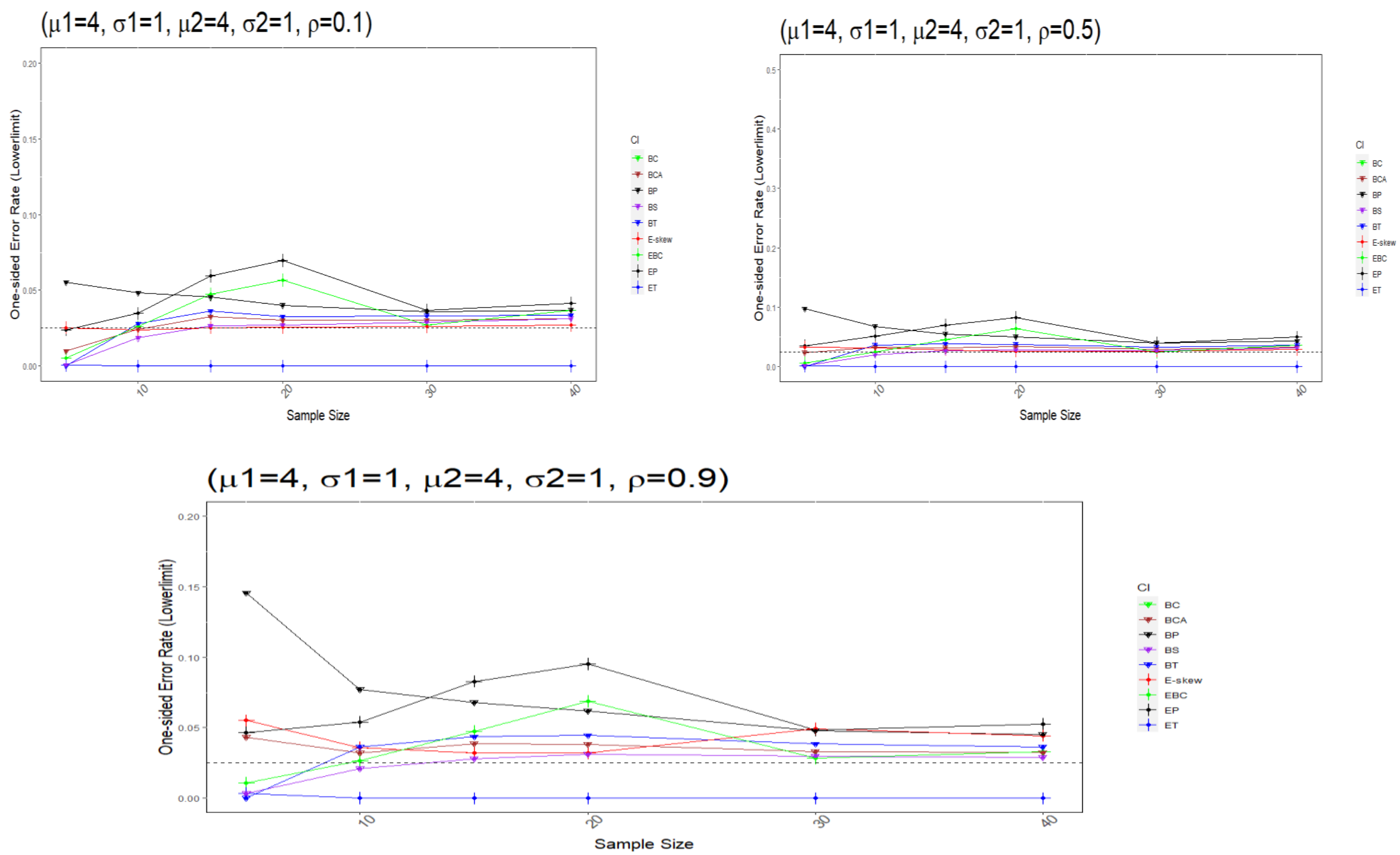


Figure: Pearson Correlation Coefficient - NU90 - One-Sided Upperlimit Error Rates for 90% CI for the Bivariate Normal Distribution

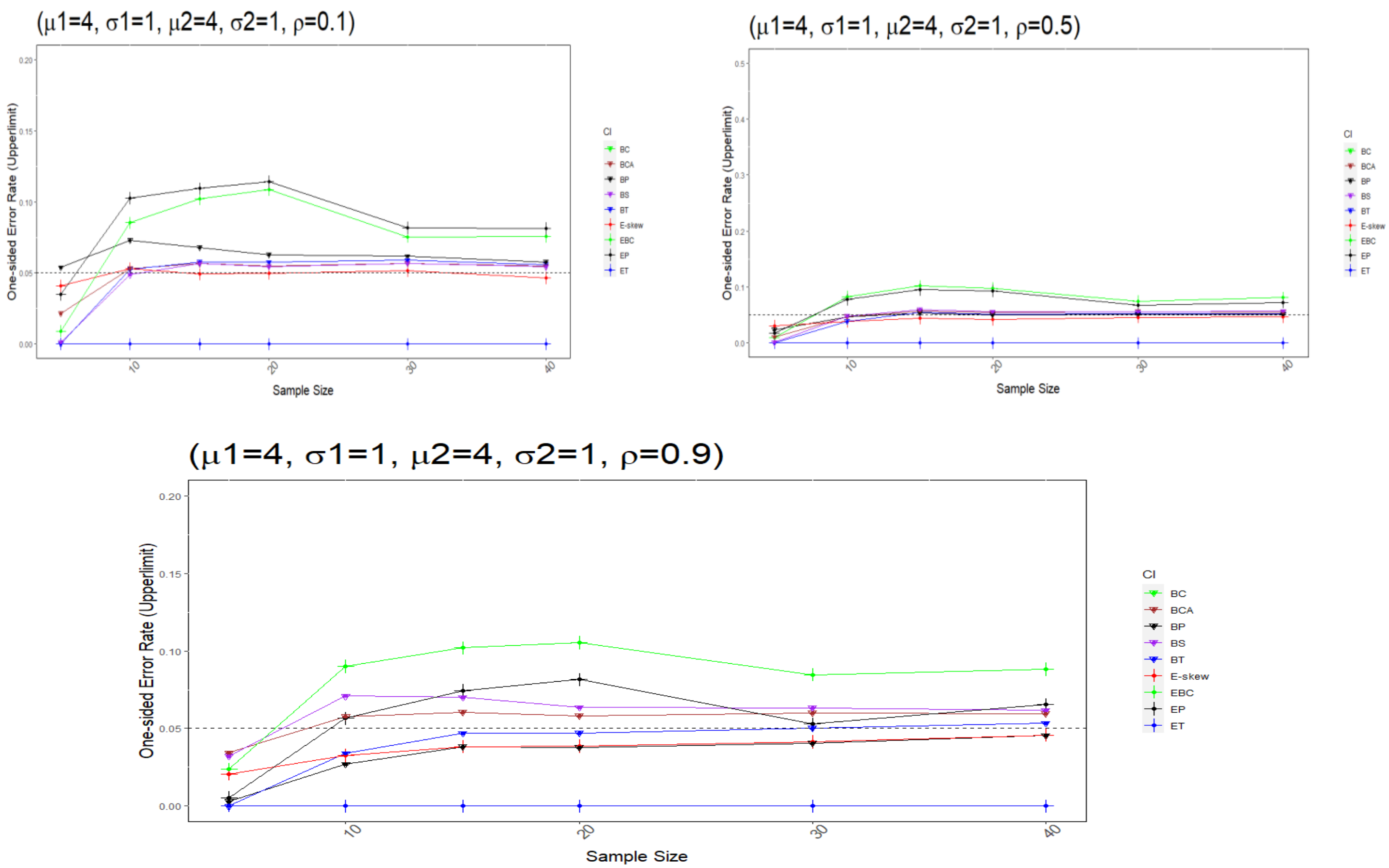
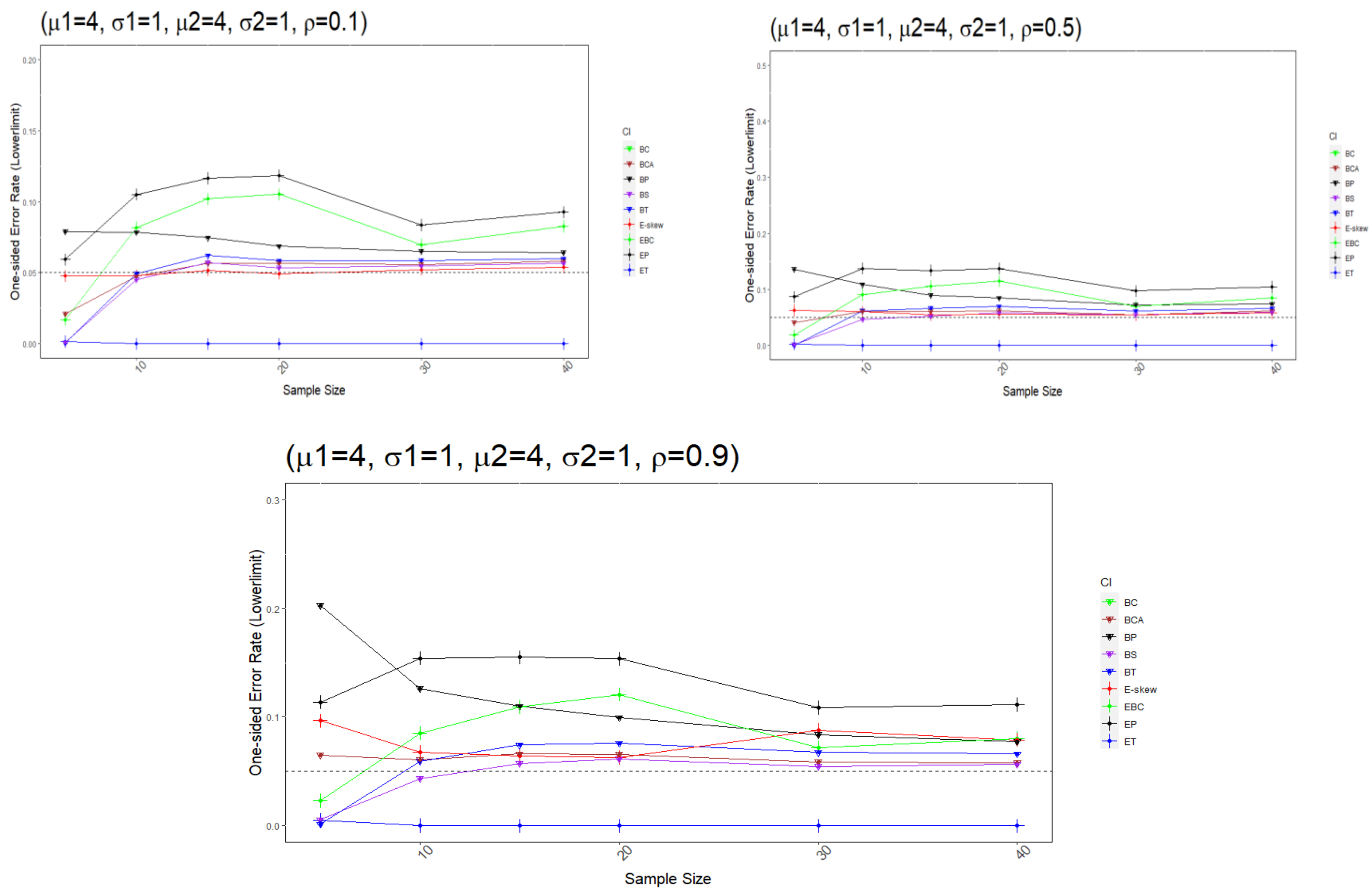


Figure: Pearson Correlation Coefficient - NL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Bivariate Normal Distribution



b. Bivariate Non-Normal Distribution

The first purpose of this section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the Pearson correlation coefficient on bivariate non-normally distributed data. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for the Pearson correlation coefficient again for bivariate non-normally distributed data.

For the bivariate non-normal distribution two sample specifications were generated. First the results for data using bivariate non-normal specifications: NN(skew=3, kurtosis=61, ρ=0.1) at the α = 0.01 significance level are considered. Confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables NN1U99 and NN1L99 on pages 203 and 204 below. For the Pearson correlation coefficient two separate pairs of bivariate samples were generated from the bivariate non-normal distributions at three different α significance levels. However, in this section because of the volume of error rate results, only the error rates reported at the α = 0.01 significance level with ρ=0.1 are displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables in this sub section only report results for the NN(skew=3, kurtosis=61, ρ=0.1) specification, the NN(skew=3, kurtosis=61, ρ=0.5) results can be viewed visually in figures NNU99, NNL99, NNU95, NNL95, NNU90 and NNL90 on pages 207-212. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the α significance level. Each colored line represents a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the NN(skew=3, kurtosis=61, ρ=0.1) parameter specification at the specified α = 0.01 significance level, for the upper and lower limit, E-skew performed relatively less accurately compared to the other methods studied. E-skew did not have the error rate with the smallest percent error at any sample size for either the upper or lower limit. E-skew also had an error rate with a larger percent error at each sample size compared to every other method applied on EBSD(n) for both the upper and lower limit.

At the α = 0.10 significance level for the upper limit, E-skew was relatively more accurate compared to every other method studied for a few sample sizes. E-skew achieved the error rate with the smallest percent error at both sample size 5 and 40 at this significance level. Additionally, at the α = 0.05 significance level for the upper limit, E-skew achieved the error rate with the smallest percent error at sample size 5. E-skew also performed relatively more accurately than every other method applied on EBSD(n) for a few sample sizes at the α = 0.05 and α = 0.10 significance levels for the lower limit. For this end of the confidence interval, E-skew had the error rate with the smallest percent error among methods applied on EBSD(n) for one sample size at the α = 0.05 significance level and three sample sizes at the α = 0.10 significance level.

The EP method performed relatively more accurately than BP at the α = 0.01 significance level, and relatively less accurately than BP at the α = 0.05 and α = 0.10 significance levels. At the α = 0.01 significance level for both the upper and lower limit, EP had an error rate with an equal or smaller percent error at each sample size when compared to BP. However, at the α = 0.05 significance level for both the upper and lower limit, EP had an error rate with a smaller percent error at only four sample sizes. Further, at the α = 0.10 significance level, EP had an error rate with a smaller percent error compared to BP at one sample size for the lower limit and no sample sizes for the upper limit. Conversely from what was seen for the simulation from the bivariate normal distribution, EBC performed relatively less accurate to BC at both the α = 0.01 and α = 0.10 significance levels for both limit ends. The error rate for ET was 0 or near 0 at each sample size for both the upper and lower limit at each significance level.

Table: Pearson Correlation Coefficient - NNU99 Upper limit error rate (α = 0.01), Bivariate non-Normal Distribution, Bivariate Non-Normal(Skew=3, Kurtosis=61, ρ = 0.1)**, Bootstraps=10000**						
Sample size	5	10	15	20	30	40
E-skew	0.0115 (130%)	0.0104 (108%)	0.0125 (150%)	0.0113 (126%)	0.0121 (142%)	0.0101 (102%)
BT	0 (100%)	**0.0025** **(50%)**	0.0039 (22%)	**0.0059** **(18%)**	**0.0037** **(26%)**	**0.0057** **(14%)**
ET	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BC	0 (100%)	0.0019 (62%)	**0.0042** **(16%)**	0.0038 (24%)	0.0032 (36%)	0.004 (20%)
EBC	**0.0011** **(78%)**	0 (100%)	0 (100%)	0 (100%)	1e-04 (98%)	1e-04 (98%)
BP	0.3094 (6088%)	0.0315 (530%)	0.014 (180%)	0.012 (140%)	0.01 (100%)	0.0104 (108%)
EP	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BS	0 (100%)	0.0019 (62%)	**0.0042** **(16%)**	0.0038 (24%)	0.0032 (36%)	0.004 (20%)
BC_a	0.4912 (9724%)	0.0152 (204%)	0.0111 (122%)	0.0136 (172%)	0.0149 (198%)	0.0163 (226%)

Table: Pearson Correlation Coefficient - NNL99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Bivariate non-Normal Distribution, Bivariate Non-Normal(Skew=3, Kurtosis=61, ρ = 0.1), **Bootstraps=10000**						
Sample size	5	10	15	20	30	40
E-skew	0.0179 (258%)	0.0224 (348%)	0.0178 (256%)	0.0184 (268%)	0.0178 (256%)	0.0181 (262%)
BT	0 (100%)	0.0059 (18%)	0.0073 (46%)	**0.0053** **(6%)**	**0.0044** **(12%)**	**0.0057** **(14%)**
ET	1e-04 (98%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)
BC	3e-04 (94%)	0.0057 (14%)	0.0063 (26%)	0.0042 (16%)	0.0034 (32%)	0.0039 (22%)
EBC	0.0028 (44%)	0.001 (80%)	2e-04 (96%)	2e-04 (96%)	0 (100%)	3e-04 (94%)
BP	0.0211 (322%)	0.0145 (190%)	0.0181 (262%)	0.0176 (252%)	0.0175 (250%)	0.0179 (258%)
EP	0.0077 (54%)	0.0011 (78%)	1e-04 (98%)	1e-04 (98%)	2e-04 (96%)	1e-04 (98%)
BS	3e-04 (94%)	**0.0057** **(14%)**	**0.0063** **(26%)**	0.0042 (16%)	0.0034 (32%)	0.0039 (22%)
$\boldsymbol{BC_a}$	**0.0029** **(42%)**	0.0086 (72%)	0.0115 (130%)	0.013 (160%)	0.0179 (258%)	0.0184 (268%)

Simulation were not only performed for the NN(skew=3, kurtosis=61, ρ=0.1) parameter specification. Simulations were also performed for the NN(skew=3, kurtosis=61, ρ=0.5) specification. When considering a change in parameter specification (i.e. increasing the correlation between the paired samples from 0.1 to 0.5) E-skew performed relatively less accurately than the other methods studied across α significance level. At the α = 0.01 significance level for the upper limit, E-skew had the error rate with a larger percent error compared to every method but BS/BC. For the lower limit at this significance level, E-skew had an error rate with a larger percent error compared to every other method. When increasing the significance level, E-skew performed relatively more accurately for the upper limit compared to the other methods studied than it did at the $\alpha = 0.01$ significance level. E-skew attained the error rate with the smallest percent at one sample for each of the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels. However, for the lower limit at these significance levels, E-skew was still among the least accurate methods. E-skew was the method with the largest percent error at each sample size.

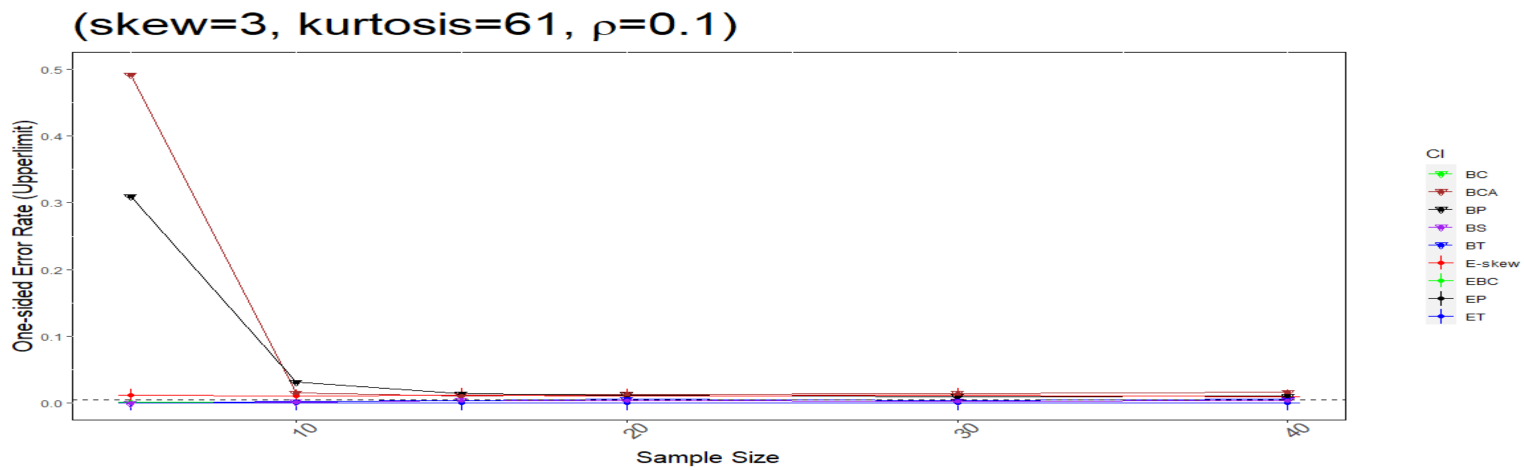
EP performed relatively more accurately compared to BP at the $\alpha = 0.01$ and $\alpha = 0.05$ significance levels for the lower limit. At the $\alpha = 0.01$ significance level for the lower limit, EP attained an error rate with a smaller percent error at each sample size. For the $\alpha = 0.05$ significance level for the lower limit, EP attained an error rate with a smaller percent error for five of six sample sizes. However, at the $\alpha = 0.10$ significance level for the lower limit, EP had an error rate with a larger percent error at five of six sample sizes. For the upper limit, EP performed relatively less accurately compared to BP. EP had an error rate with a larger percent error for most sample sizes at each significance level compared to BP.

At the $\alpha = 0.01$ significance level for the lower limit, EBC performed relatively more accurately than BC. EBC achieved the error rate with the smallest percent error compared to all other methods at five of six sample sizes for the lower limit at this significance level. In addition, for the upper limit at this significance level, it achieved the error rate with the smallest percent error compared to every other method at sample size 40. Then when α significance level was modified, EBC performed less accurately compared to BC at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels. At the $\alpha = 0.05$ significance level, EBC had an error rate with a larger percent error for half the sample sizes when compared to BC for both the upper and lower limit. At the $\alpha = 0.10$

significance level, EBC had an error rate with a larger percent error for five of six sample sizes for both the upper and lower limit.

The error rate for ET was again at or near 0 at each sample size and significance level when the correlation was increased from 0.1 to 0.5.

Figure: Pearson Correlation Coefficient - NNU99 - One-Sided Upperlimit Error Rates for 99% CI for the Bivariate Normal Distribution



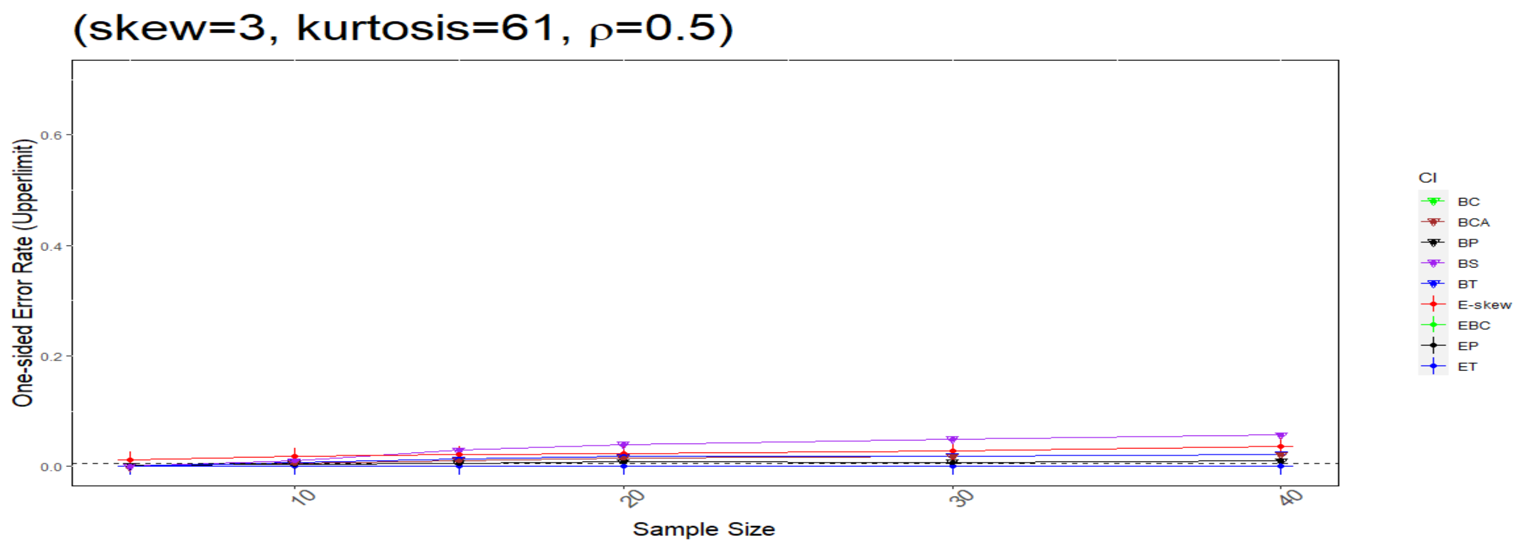
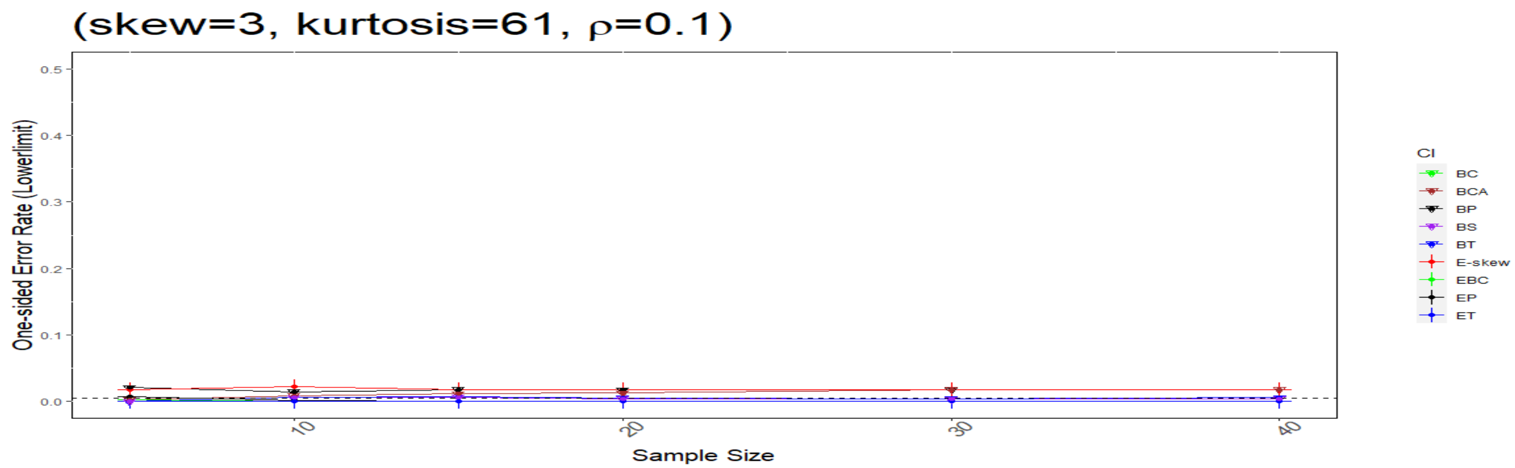


Figure: Pearson Correlation Coefficient - NNL99 - One-Sided Lowerlimit Error Rates for 99% CI for the Bivariate Normal Distribution



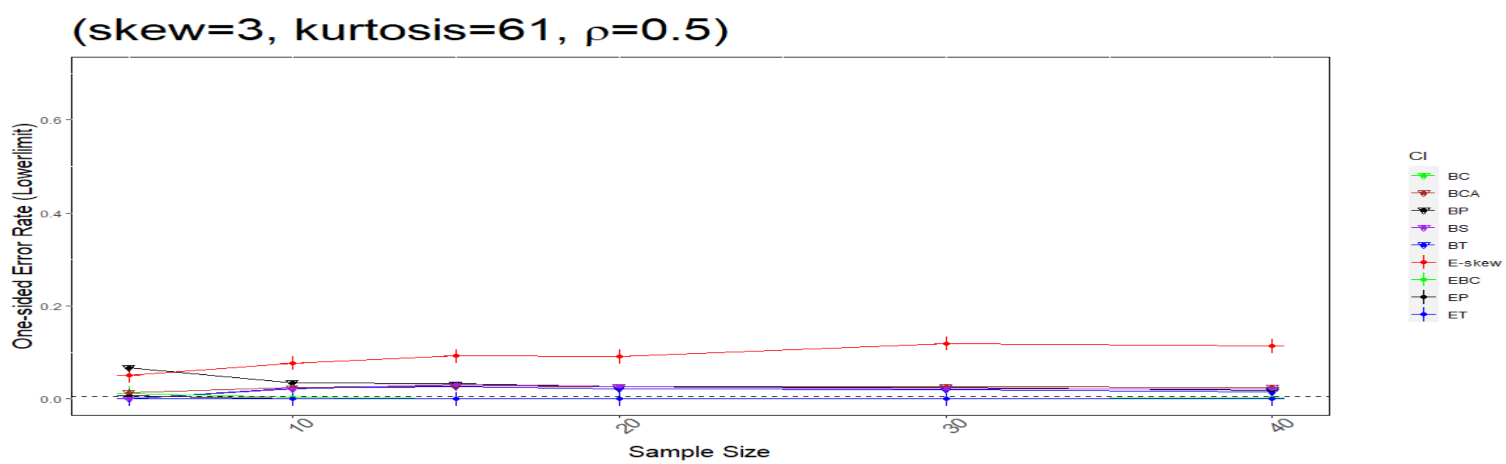
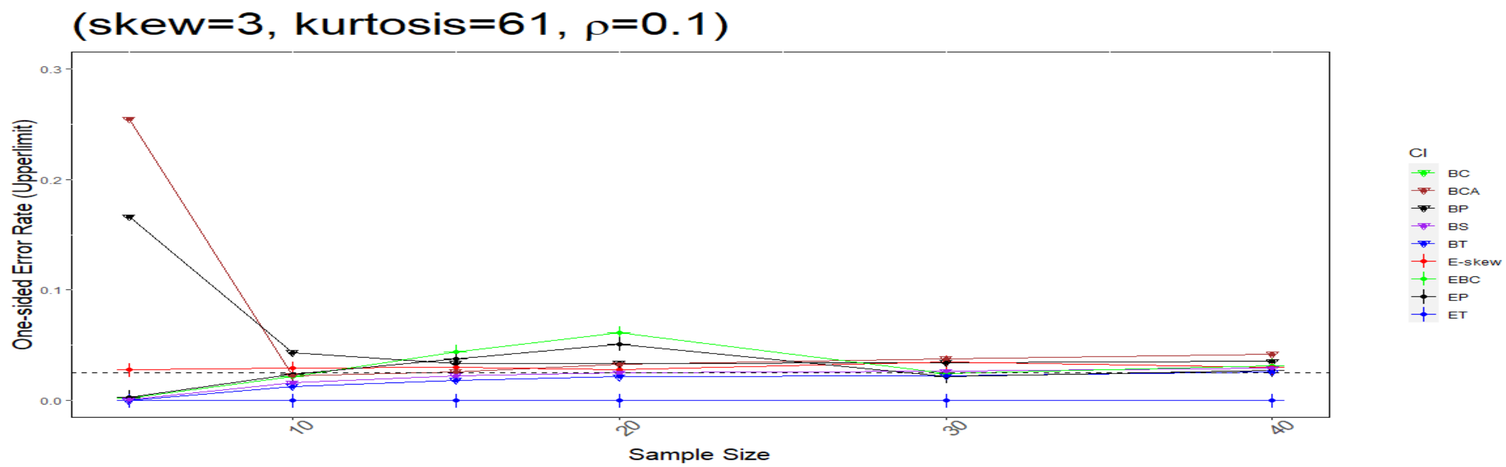


Figure: Pearson Correlation Coefficient - NNU95 - One-Sided Upperlimit Error Rates for 95% CI for the Bivariate Normal Distribution



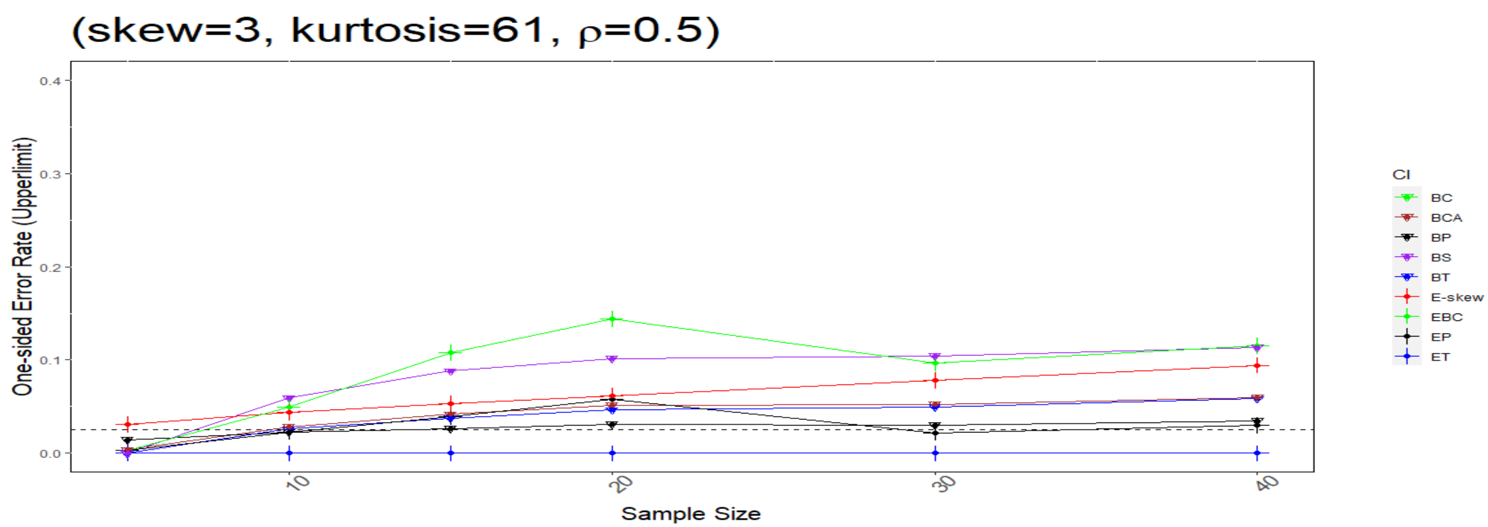
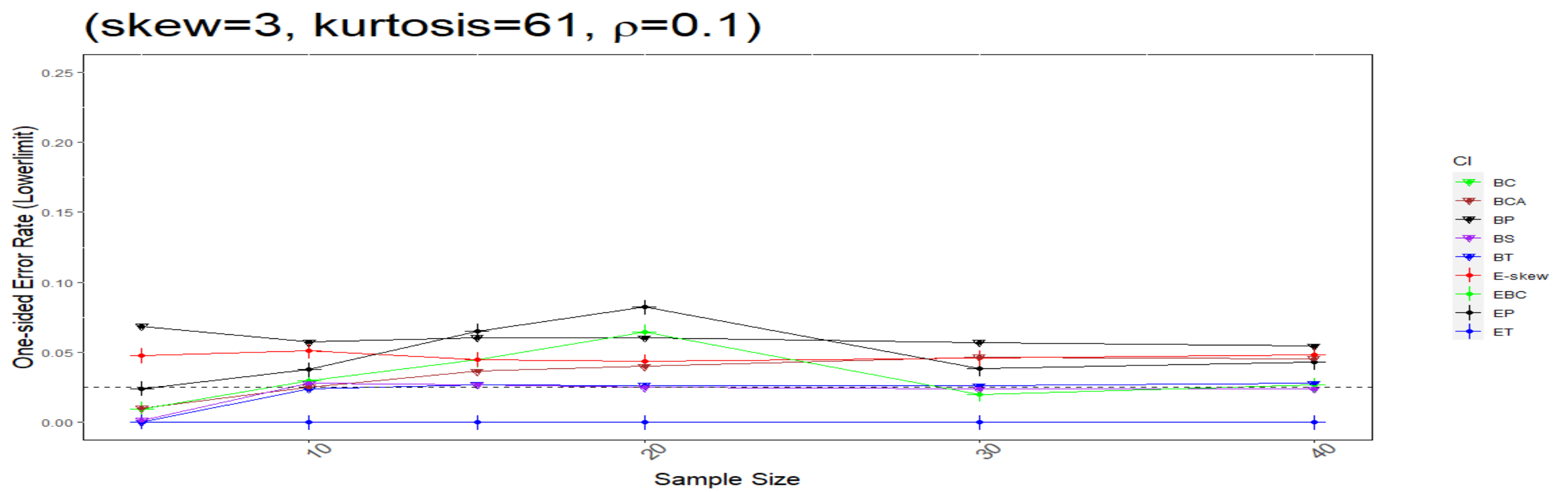


Figure: Pearson Correlation Coefficient - NNL95 - One-Sided Lowerlimit Error Rates for 95% CI for the Bivariate Normal Distribution



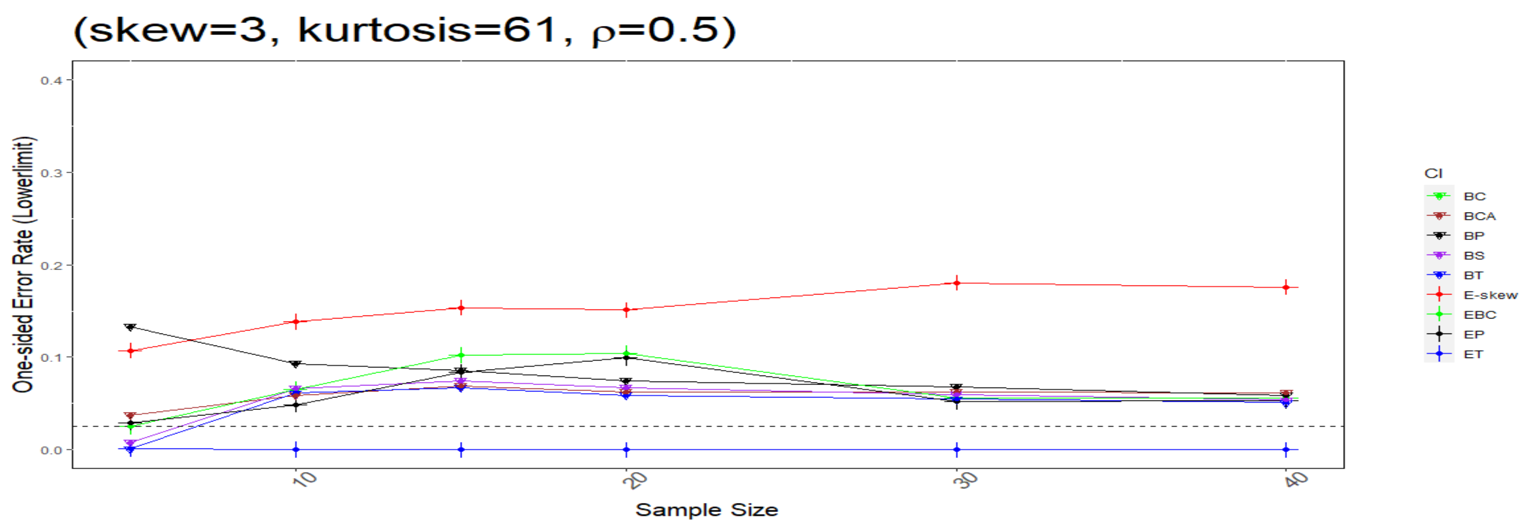
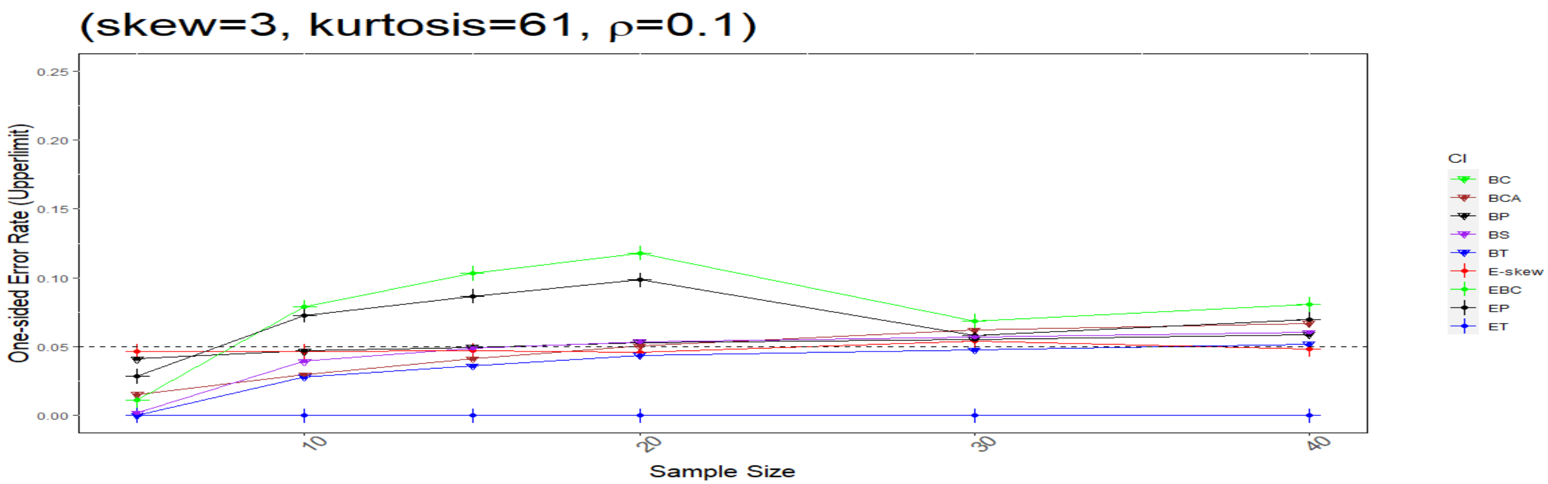


Figure: Pearson Correlation Coefficient - NNU90 - One-Sided Upperlimit Error Rates for 90% CI for the Bivariate Normal Distribution



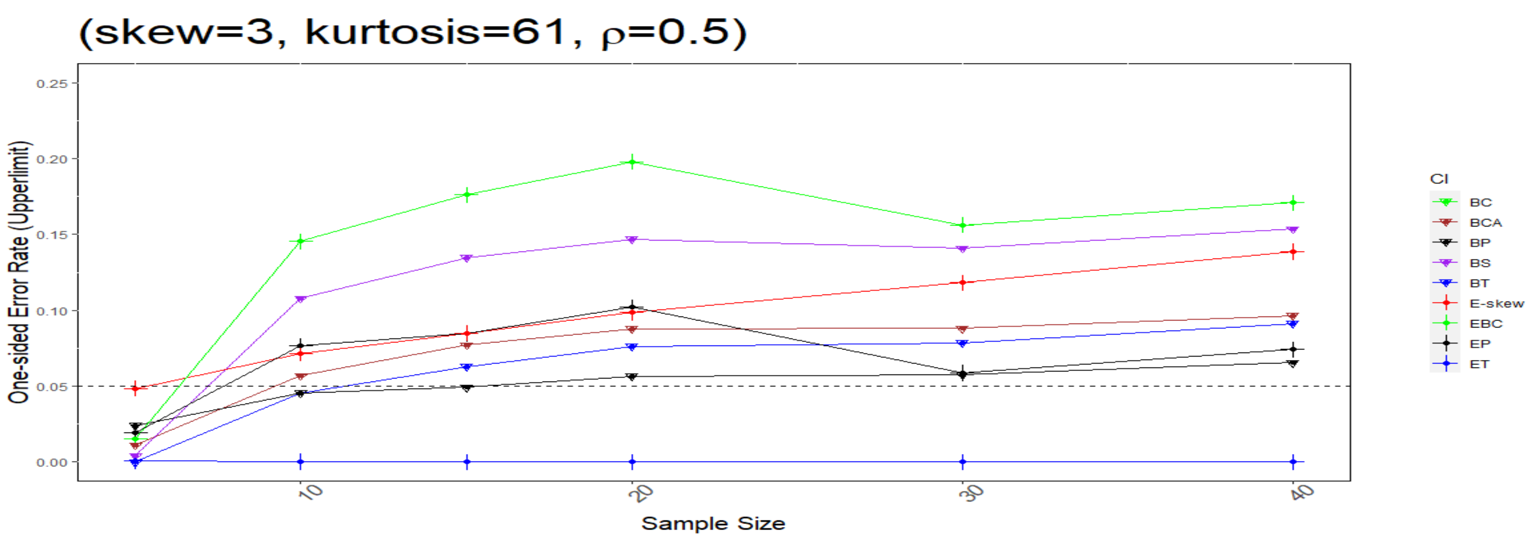
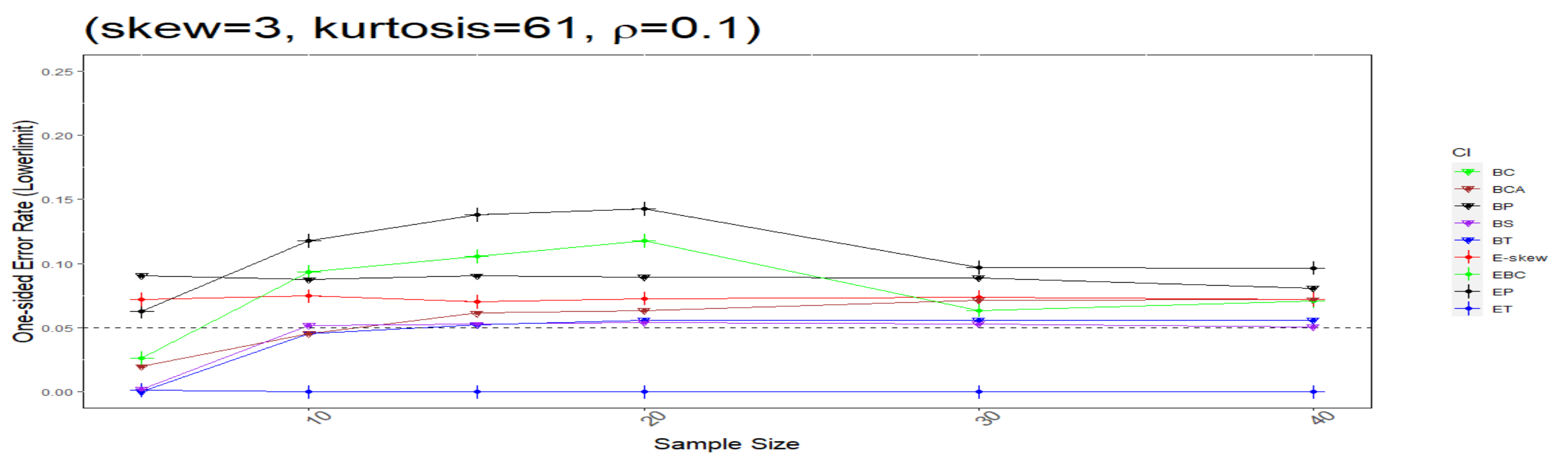
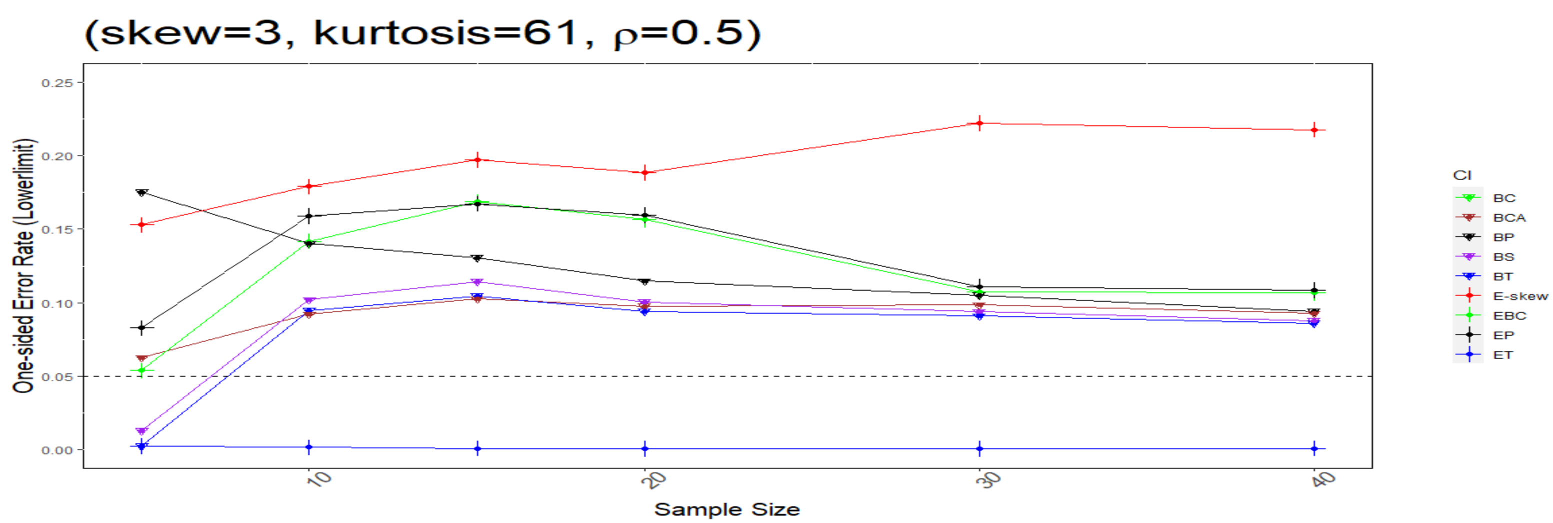


Figure: Pearson Correlation Coefficient - NNL90 - One-Sided Lowerlimit Error Rates for 90% CI for the Bivariate Normal Distribution





Pearson Correlation Coefficient Results for methods using EBSD(n)

In general, the E-skew method was relatively accurate compared to other methods for the Pearson correlation coefficient when data was generated from a bivariate normal distribution. E-skew performed relatively most accurately when the correlation specified was ρ=0.1. As the Pearson coefficient was increased close to the boundary (the coefficient must be bounded between -1, and 1) E-skew performed relatively less accurately. In general, E-skew achieved the error rate with the smallest percent error in comparison to all other methods the most frequently at the α = 0.05 significance level when ρ=0.1 and the data generated was bivariate normal.

For moderately skewed data, like data generated from the bivariate non-normal distribution, E-skew did not perform as accurately. When ρ=0.1, E-skew performed relatively less accurately than both Monte Carlo Bootstrap and other EBSD(n) methods at the α = 0.01 significance level. Additionally, when the significance level was modified, E-skew frequently had an error rate with a larger percent error compared to at least one Monte Carlo Bootstrap method. Although this was the case, E-skew did perform relatively more accurately compared to other methods applied on EBSD(n) at the α = 0.05 and α = 0.10 significance levels. When the Pearson correlation coefficient was increased to ρ=0.5 E-skew also performed relatively less accurately compared to the other methods studied. For the lower limit, E-skew performed the least accurately compared to any other method. For the upper limit it generally had error rates with larger percent errors than other methods applied on EBSD(n) and Monte Carlo Bootstrap methods.

Regardless of the parameter specification and distribution drawn from, other methods applied on EBSD(n) varied in relative accuracy at each significance level. EP performed relatively more accurately at the α = 0.01 significance level, and relatively less accurately when significance level was modified. EBC performed relatively more accurately at the α = 0.10 significance level and less accurately as significance level decreased. ET had an error rate near 0 regardless of the sample size, significance level and distribution the data was drawn from consequently indicating the method was generating intervals that were far too large.

4.4 Trimmed Sample Mean

For the trimmed sample mean portion of the simulation study, results for four different sample sizes are reported ($n =$ 10, 20, 30, and 40). For each of these sample sizes, confidence interval error rates are reported at the α = 0.01, 0.05, and 0.10 significance levels.

The probability distributions used in the simulation study for the trimmed sample mean were the normal, exponential, gamma, log-normal, and mixture of two normal distributions. For each distribution, the population parameters specified are displayed below in Table 4.4. These parameter specifications are the same as the specifications for the sample mean in Table 3.5 in Chapter 3.

For each sample size, population parameter specification, and probability distribution combination 10,000 separate samples were generated. For the Monte Carlo Bootstrap confidence interval methods each of the 10,000 samples used 10,000 Monte Carlo Bootstrap resamples to create its Bootstrap sampling distribution. The comparisons discussed in this section are made between EBSD(n) methods and Bootstrap methods that use 10,000 Bootstrap resamples. In addition, confidence interval method error rate results were measured on the same 10,000 unique samples using 500 Bootstrap resamples. This alternative Bootstrap resampling level was performed for each distribution tested. The error rate results at these additional Bootstrap resampling levels are reported in the Appendix. Each generated unique sample had confidence intervals computed using the confidence interval methods listed below.

- For methods using EBSD(n) this included: E-skew, ET, EBC, EP, EBC_a, and ES.
- For methods using the Monte Carlo Bootstrap this includes: BT, BC, BP, BC_a/ABC, and BS.

Below in Table 4.4 is a description of the parameter specifications used for the trimmed sample mean statistic in this simulation study:

Table 4.4 Simulation Parameter Specifications for the Trimmed Sample Mean Statistic

Probability distribution	Population Parameter	Parameter code: Specified Parameter Values
Normal distribution	(μ,σ)	N1: (50, 1)
Exponential distribution	(λ)	E1: (0.1)
Gamma distribution	(α, λ)	G1: (2, 2)
Log-normal distribution	(μ,σ)	LN1: (4, 0.2)
Mixture of two normal distributions	$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	M1: (4, 4, 0.6, 8, 8, 0.4)

a. Normal Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the trimmed sample mean when data is normally distributed. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods again for the trimmed sample mean when data is normally distributed. First the results for data generated from $N(\mu = 50, \sigma = 1)$ distribution at the $\alpha = 0.01$ significance level are discussed. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables N1U99 and N1L99 on pages 218 and 219 below. Detailed numerical results for error rate results tested at the $\alpha = 0.05$ and $\alpha = 0.10$ significance levels, as well as error rate results for Monte Carlo Bootstrap methods where 500 bootstrap resamples were specified can be viewed in Appendix tables. Results can also be viewed visually in figures N99, N95, and N90 on pages 220-222.

In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 10, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the $N(\mu = 50, \sigma = 1)$ parameter specification at the specified $\alpha = 0.01$ significance level, for the upper limit, E-skew did not have the error rate with the smallest percent error at any sample size. E-skew also did not have the error rate with the smallest percent error of any method using EBSD(n) at any sample size at this significance level. The percentile methods, EP and EBC, were more accurate at the $\alpha = 0.01$ significance level relative to E-skew and Monte Carlo Bootstrap methods. At sample sizes 10, 20, and 30, EP had the error rate with the smallest percent error for the upper and lower limit.

Additionally, the EBC method had the error rate with the smallest percent error at sample size 40 for the upper and lower limit. These results are shown below in N1U99 and N1L99. These results can also be viewed visually in figure N99.

The E-skew method demonstrates its relative value when comparing the error rates across α significance level. The E-skew method was more accurate relative to the other methods applied on EBSD(n) when the α significance level modified from α = 0.01 to α = 0.10. At the α = 0.05 and α = 0.10 significance levels for the upper and lower limit, E-skew achieved the error rate with the smallest percent error among all methods applied on EBSD(n) at all four sample sizes. Additionally, when comparing E-skew to Monte Carlo Bootstrap methods it had an error rate with a smaller percent error than any method except BS.

Compared to their Monte Carlo Bootstrap counterpart, the percentile methods EBC and EP, performed relatively more accurately compared at the α = 0.01 significance level and relatively less accurately at the α = 0.05 and α = 0.10 significance levels. EP and EBC both had error rates with smaller percent errors at every sample size for both the upper and lower limit at the α = 0.01 significance level compared to BP and BC respectively. Then when the α significance level was modified to α = 0.05, both methods had error rates with larger percent errors at three of four sample sizes for both the upper and lower limit. At the α = 0.10 significance level for both limits, EP and EBC performed even relatively less accurately compared to BP and BC, having error rates with a larger percent error at every sample size for each comparison respectively.

ET did not attain the error rate with the smallest percent error at any α significance level. It did, however, perform about equally well to BT; ET achieved an error rate with a smaller percent error than BT for about half the sample sizes studied for both the upper and lower limit.

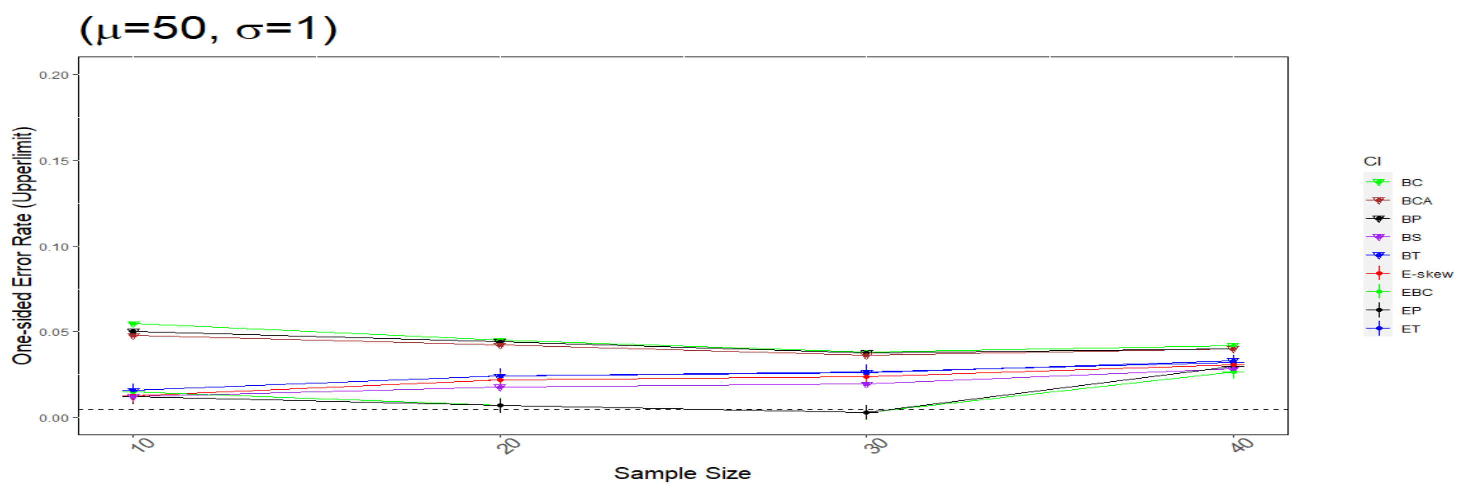
Table: Trimmed Sample Mean - N1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Normal Distribution, $\boldsymbol{N}(\boldsymbol{\mu} = 50, \boldsymbol{\sigma} = 1)$, **Bootstraps=10000**

Sample size	10	20	30	40
E-skew	0.0128 (156%)	0.022 (340%)	0.0237 (374%)	0.0307 (514%)
BT	0.0158 (216%)	0.0241 (382%)	0.0262 (424%)	0.0332 (564%)
ET	0.0158 (216%)	0.0244 (388%)	0.0265 (430%)	0.0323 (546%)
BC	0.0547 (994%)	0.045 (800%)	0.038 (660%)	0.0421 (742%)
EBC	0.0148 (196%)	0.0072 (44%)	0.0029 (42%)	**0.0267** **(434%)**
BP	0.0502 (904%)	0.0444 (788%)	0.0377 (654%)	0.04 (700%)
EP	**0.0123** **(146%)**	**0.0069** **(38%)**	**0.0031** **(38%)**	0.0298 (496%)
BS	0.0124 (148%)	0.018 (260%)	0.0196 (292%)	0.0283 (466%)
$\boldsymbol{BC_a}$	0.0478 (856%)	0.0423 (746%)	0.0363 (626%)	0.0402 (704%)

Table: Trimmed Sample Mean - N1L99 Lower limit error rate (α = 0.01), Normal Distribution, $N(\mu = 50, \sigma = 1)$, **Bootstraps=10000**

Sample size	10	20	30	40
E-skew	0.0147 (194%)	0.0244 (388%)	0.0251 (402%)	0.0248 (396%)
BT	0.0185 (270%)	0.0279 (458%)	0.0282 (464%)	0.0267 (434%)
ET	0.0185 (270%)	0.0283 (466%)	0.0283 (466%)	0.0265 (430%)
BC	0.0572 (1044%)	0.0456 (812%)	0.0396 (692%)	0.0358 (616%)
EBC	0.0153 (206%)	0.008 (60%)	0.0044 (12%)	**0.021 (320%)**
BP	0.0544 (988%)	0.043 (760%)	0.0388 (676%)	0.0346 (592%)
EP	**0.0146 (192%)**	**0.0068 (36%)**	**0.0046 (8%)**	0.0253 (406%)
BS	0.0148 (196%)	0.0214 (328%)	0.0222 (344%)	0.0222 (344%)
BC_a	0.0555 (1010%)	0.0414 (728%)	0.0377 (654%)	0.0339 (578%)

Figure: Trimmed Sample Mean - N99 - One-Sided Error Rates for 99% CI for the Normal Distribution



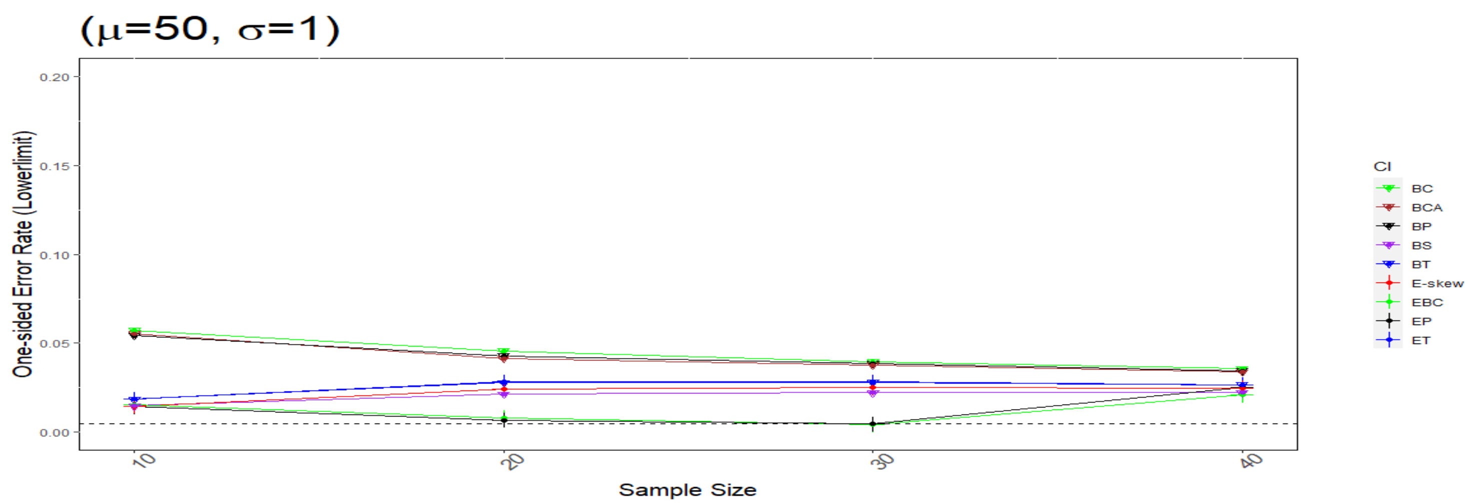
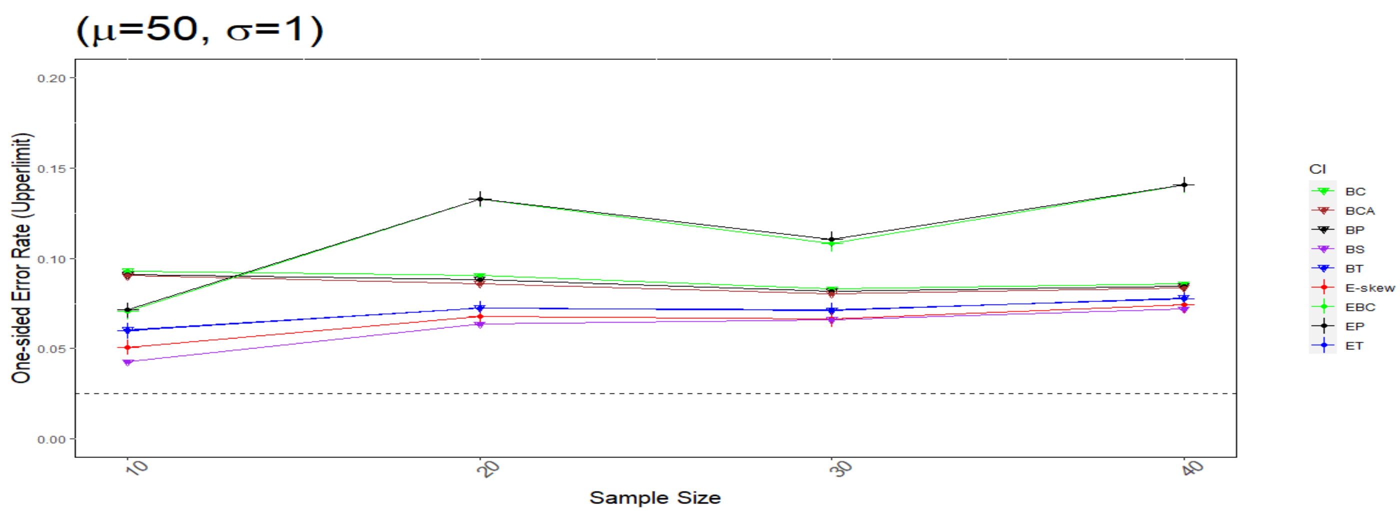


Figure: Trimmed Sample Mean - N95 - Error Rates for 95% CI for the Normal Distribution



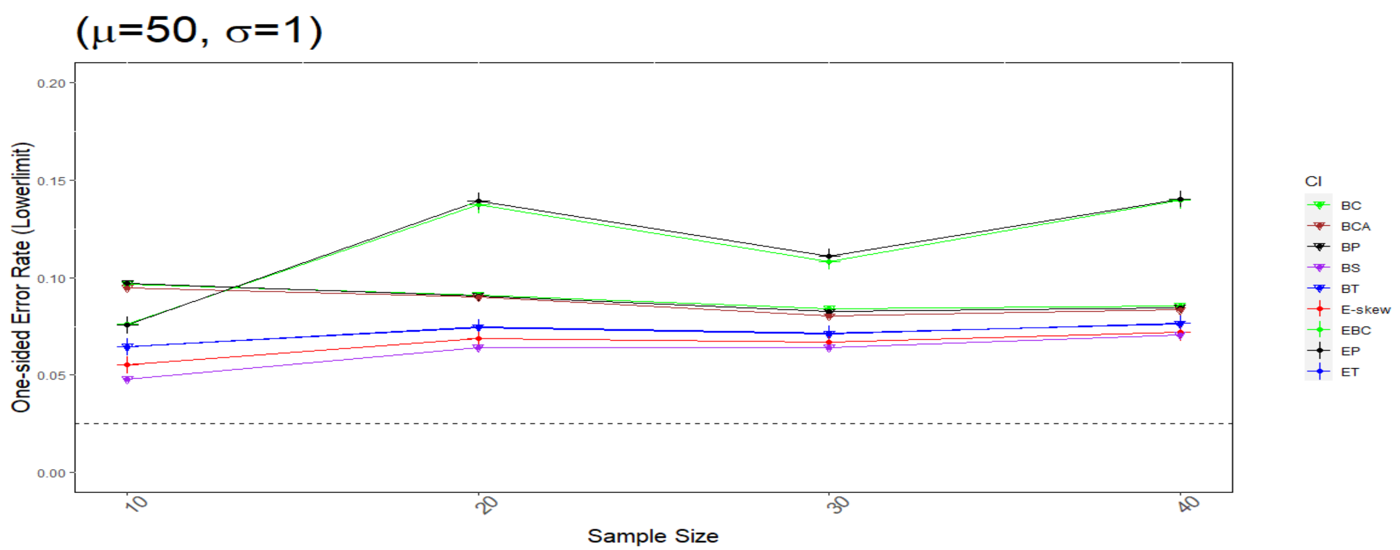
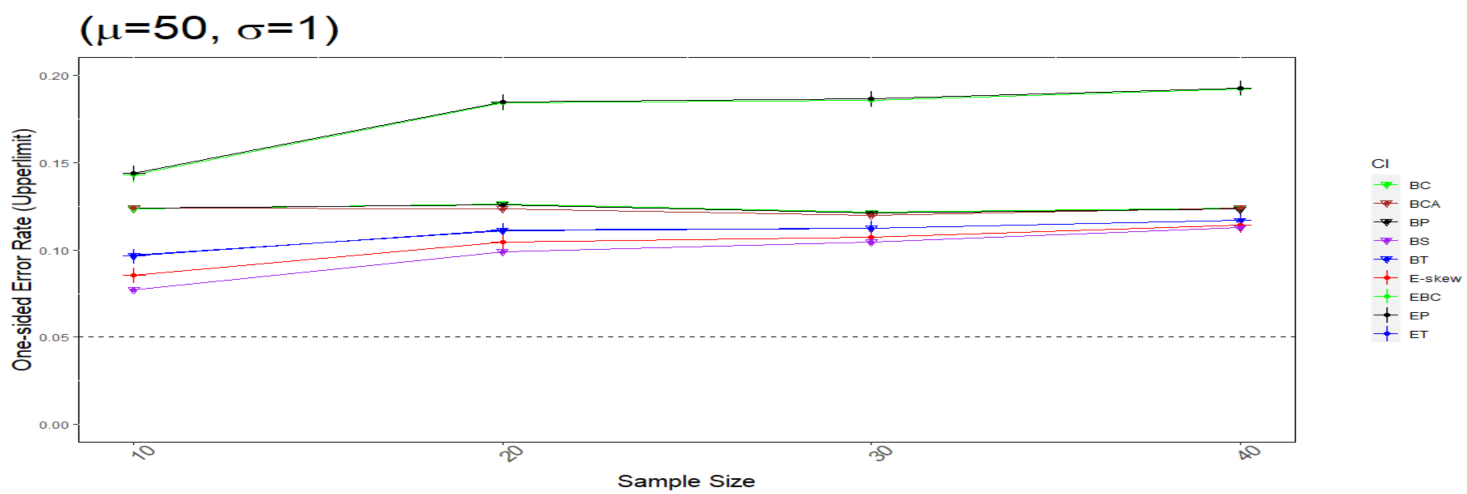
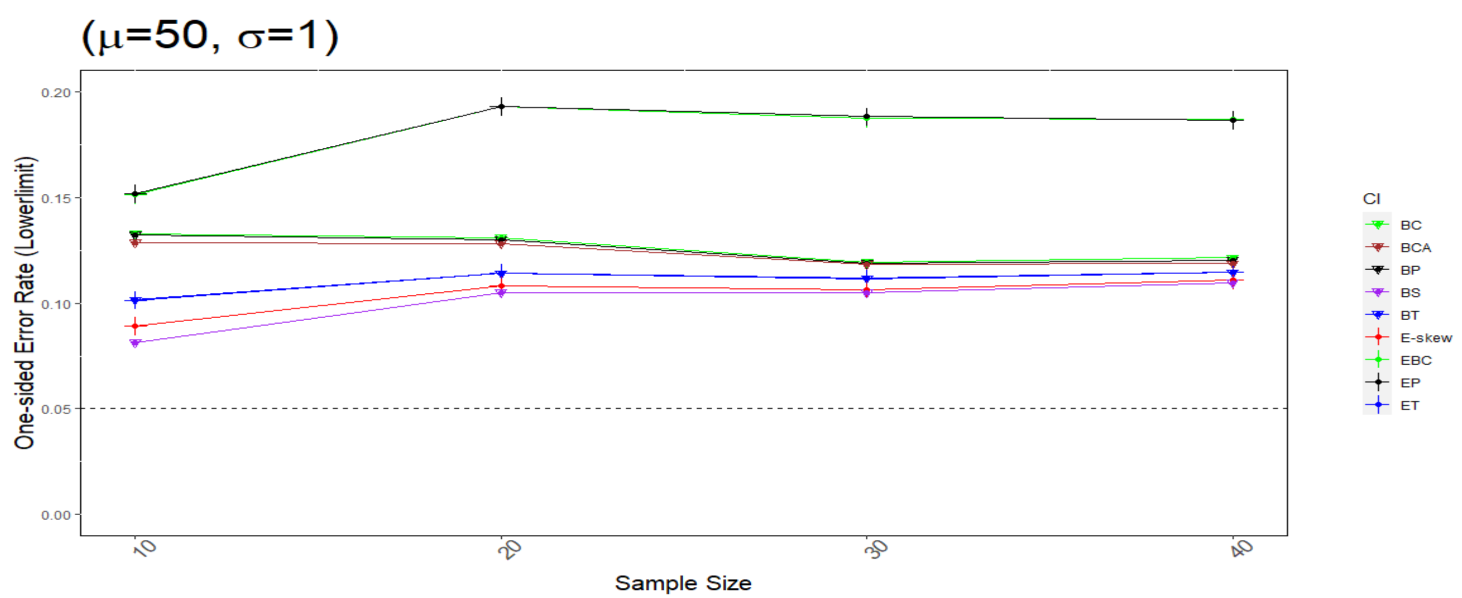


Figure: Trimmed Sample Mean: N90 - One-Sided Error Rates for 90% CI for the Normal Distribution





b. Exponential Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the trimmed sample mean statistic on data drawn from an exponential distribution. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for this same statistic and distribution. First the results for data generated from Exp(λ = 0.10) distribution at the α = 0.01 significance level are considered. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables E1U99 and E1L99 on pages 226 and 227 below. Detailed numerical results for error rate results tested at the α = 0.05 and α = 0.10 significance levels, as well as error rate results for Monte Carlo Bootstrap methods where 500 bootstrap resamples were specified can be viewed in Appendix tables. Results can also be viewed visually in figures E99, E95, and E90 on pages 228-230.

In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 10, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the Exp(λ = 0.10) parameter specification at the specified α = 0.01 significance level, for the upper limit, E-skew did not have the error rate with the smallest percent error at any sample size. E-skew also did not have the error rate with the smallest percent error of any method using EBSD(n) at any sample size at this significance level. EP performed relatively accurately compared to the other methods applied on EBSD(n) for the upper limit at this α = 0.01 significance level. Namely, at sample sizes 10, 20, and 30 EP had the error rate with the smallest percent error for the upper and lower limit among all methods applied on EBSD(n). Additionally, the EBC method had the error rate with the smallest percent error at sample size 40 for both the upper and lower limit among all methods applied on EBSD(n) and had the smallest among every method compared for the lower limit at sample size 40. These results are shown below in E1U99 and E1L99. These results can also be viewed visually in figure E99.

The E-skew method demonstrates its relative value when comparing the error rates across α significance level. In the case of the Exp(λ = 0.10) distribution, the E-skew method's accuracy relative to percentile methods applied on EBSD(n) when the α significance level was modified from α = 0.01 to α = 0.05 and α = 0.10 for the upper limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew achieved the error rate with the smallest percent error among all methods applied on EBSD(n) at all four sample sizes.

EP performed relatively more accurately compared to the other methods applied on EBSD(n) at the α = 0.01 significance level than it did at the α = 0.05 and α = 0.10 significance levels. At the α = 0.01 significance level for the upper limit, EP attained the error rate with the smallest percent error among methods applied on EBSD(n) at sample sizes 10 and 20. At the α = 0.05 and α = 0.10 significance levels for the upper limit however, EP failed to attain the error rate with the smallest percent error among methods applied on EBSD(n) at any sample size. For the upper limit the same could be said of the EBC. However for the lower limit, EBC did perform relatively more accurately compared to the other methods studied across α significance level. At the α = 0.05 significance level, EBC attained the error rate with the smallest percent error compared to every other method at three of four sample sizes. Further at the α = 0.10 significance level, EBC attained the error rate with the smallest percent error compared to every other method at every sample size studied.

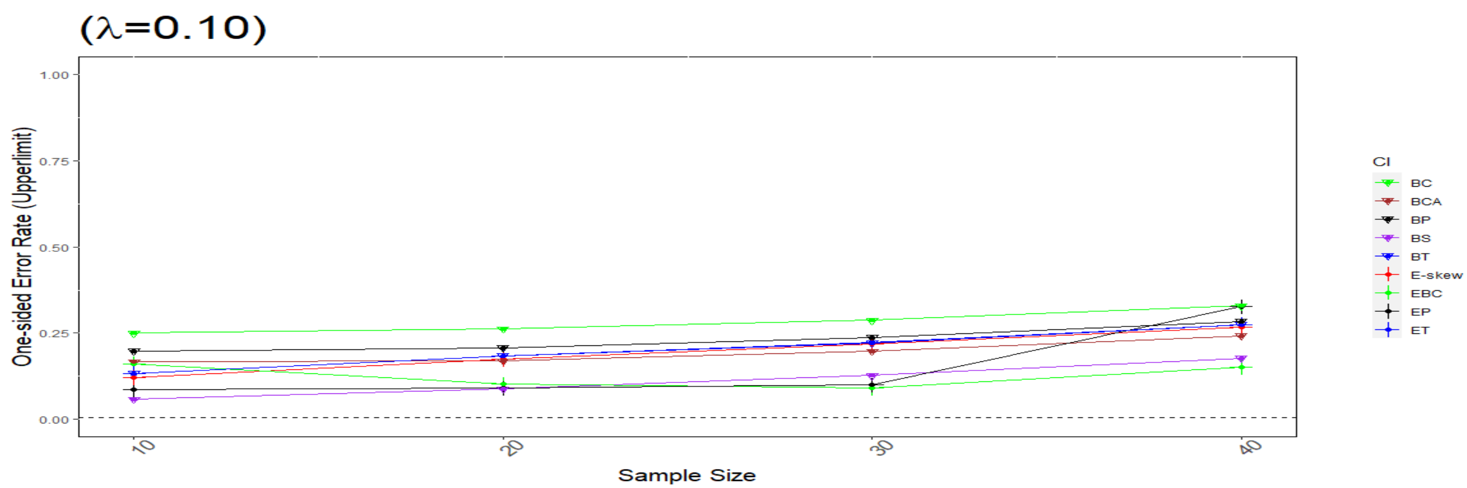
Similarly, for the upper limit, the percentile methods applied on EBSD(n) performed relatively accurately compared to their Monte Carlo Bootstrap counterparts at the α = 0.01 significance level. EP had an error rate with a smaller percent error for three of four sample sizes when compared to BP and EBC did for all four sample sizes when compared to BC. Then when significance level was modified to α = 0.05 and α = 0.10, EBC had an error rate with a larger percent compared to BC for three of the four sample sizes studied. Similarly, EP had an error rate that was larger for three of the four sample sizes at the α = 0.05 significance level, and for all four sample sizes at the α = 0.10 significance level.

ET did not attain the error rate with the smallest percent error at any α significance level. It did, however, perform about equally well to BT; ET achieved an error rate with a smaller percent error than BT for about half the sample sizes studied for both the upper and lower limit. The two methods alternated which had a higher and which a lower percent error across sample size, and their percent errors were approximately equivalent.

Table: Trimmed Sample Mean - E1U99 Upper limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda$ = 0.10)**, Bootstraps=10000**				
Sample size	10	20	30	40
E-skew	0.1211 (2322%)	0.1747 (3394%)	0.2176 (4252%)	0.2679 (5258%)
BT	0.1338 (2576%)	0.1838 (3576%)	0.2219 (4338%)	0.2742 (5384%)
ET	0.1337 (2574%)	0.1835 (3570%)	0.2228 (4356%)	0.2748 (5396%)
BC	0.2501 (4902%)	0.2621 (5142%)	0.287 (5640%)	0.329 (6480%)
EBC	0.1606 (3112%)	0.1023 (1946%)	0.0906 (1712%)	0.1515 (2930%)
BP	0.1982 (3864%)	0.2076 (4052%)	0.238 (4660%)	0.2846 (5592%)
EP	0.0858 (1616%)	0.0907 (1714%)	0.0999 (1898%)	0.3264 (6428%)
BS	**0.0585** **(1070%)**	**0.0882** **(1664%)**	**0.1273** **(2446%)**	**0.1779** **(3458%)**
BC_a	0.167 (3240%)	0.1695 (3290%)	0.1972 (3844%)	0.2424 (4748%)

Table: Trimmed Sample Mean - E1L99 Lower limit error rate (α = 0.01), Exponential Distribution, $Exp(\lambda$ = 0.10)**, Bootstraps=10000**				
Sample size	10	20	30	40
E-skew	0.0012 (76%)	0.001 (80%)	5e-04 (90%)	1e-04 (98%)
BT	0.0015 (70%)	0.0011 (78%)	6e-04 (88%)	3e-04 (94%)
ET	0.0015 (70%)	0.001 (80%)	6e-04 (88%)	3e-04 (94%)
BC	0.0079 (58%)	0.0021 (58%)	0.001 (80%)	6e-04 (88%)
EBC	0.0012 (76%)	4e-04 (92%)	0 (100%)	**0.0011** **(78%)**
BP	0.0106 (112%)	0.0029 (42%)	0.0018 (64%)	8e-04 (84%)
EP	0.0013 (74%)	1e-04 (98%)	0 (100%)	1e-04 (98%)
BS	**0.0022** **(56%)**	0.0014 (72%)	0.0012 (76%)	4e-04 (92%)
BC_a	0.0128 (156%)	**0.004** **(20%)**	**0.0021** **(58%)**	**0.0011** **(78%)**

Figure: Trimmed Sample Mean - E99 - One-Sided Error Rates for 99% CI for the Exponential Distribution



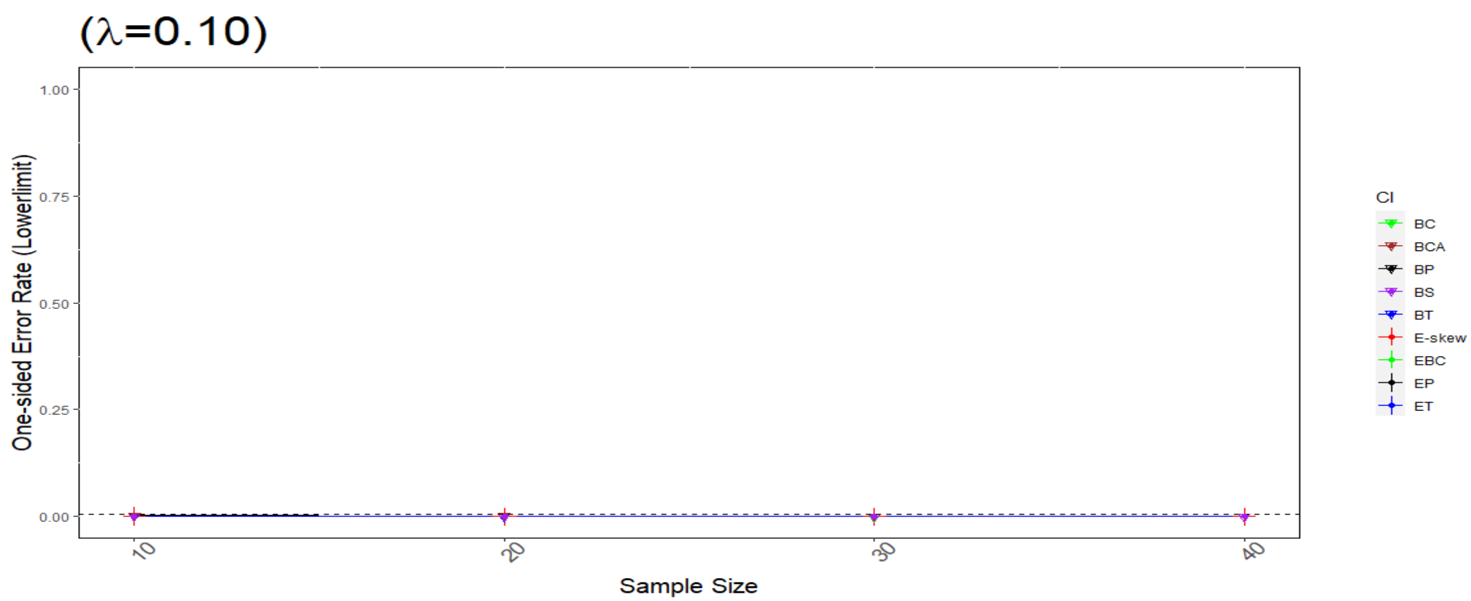
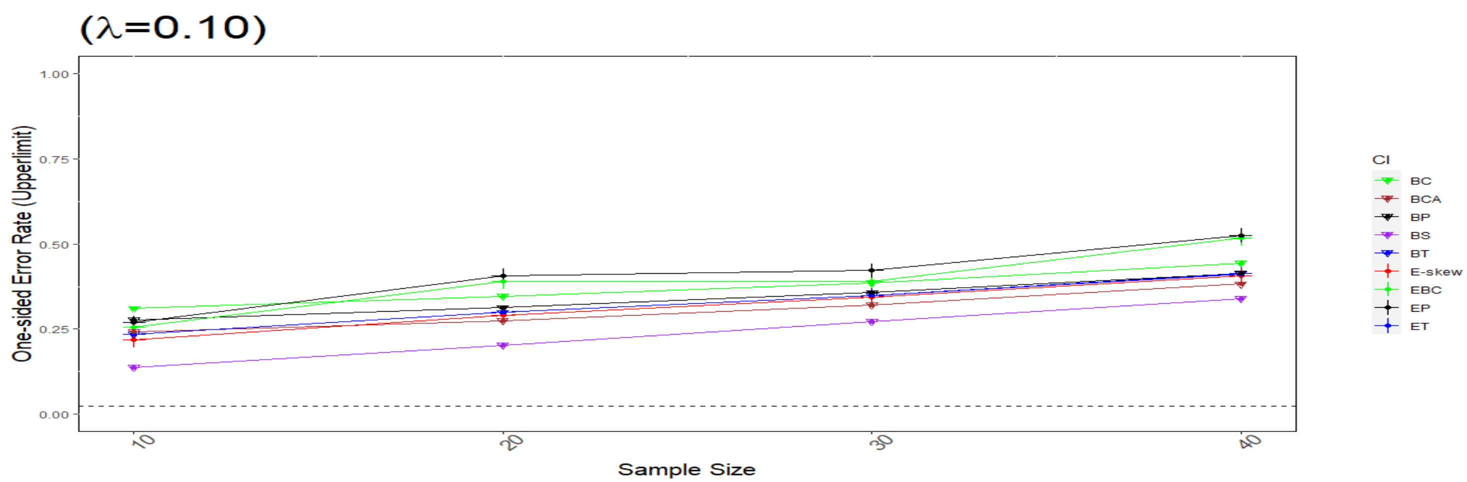


Figure: Trimmed Sample Mean - E95 - One-Sided Error Rates for 95% CI for the Exponential Distribution



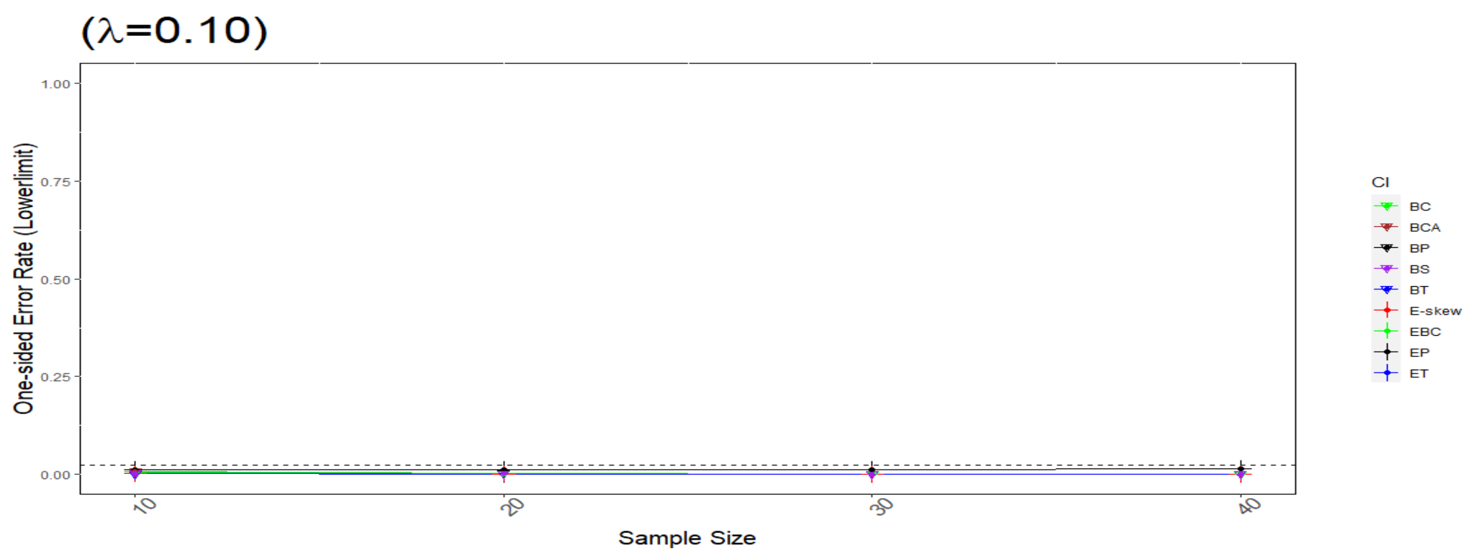
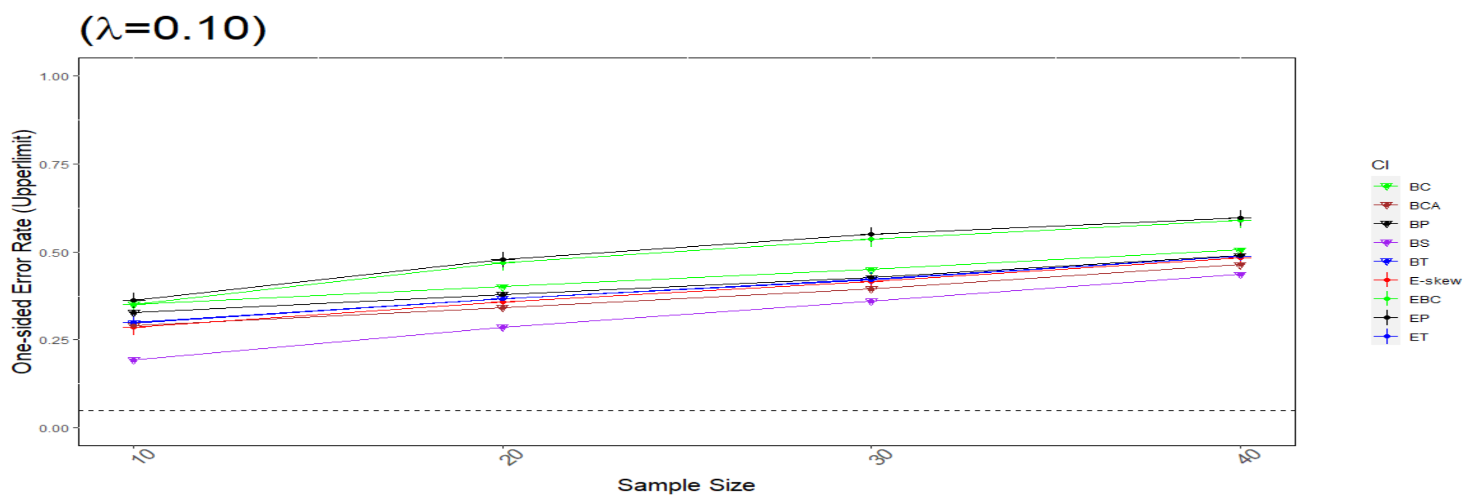
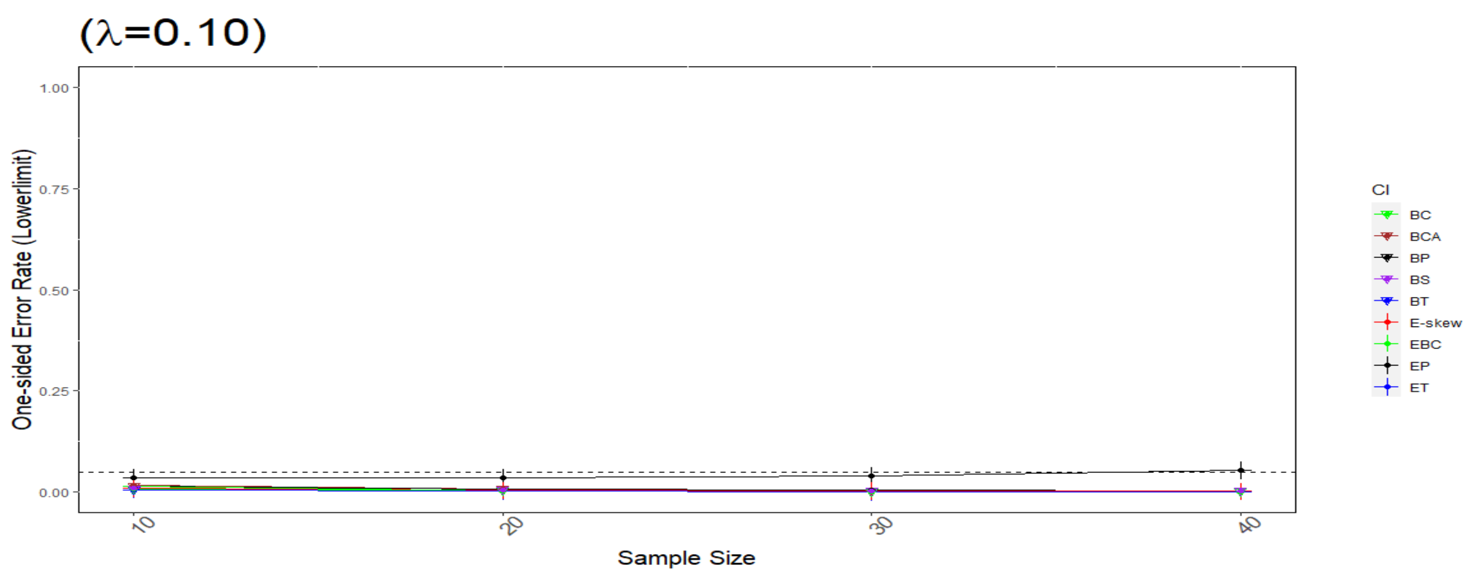


Figure: Trimmed Sample Mean - E90 - One-Sided Error Rates for 90% CI for the Exponential Distribution





c. Gamma Distribution

The first purpose of this sub section is to compare the performance of E-skew to the accuracy of all other methods studied for the trimmed sample mean statistic on data drawn from a gamma distribution. The second is to compare the accuracy of other methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for this same statistic and distribution. First the results for data generated from gamma(α=2, λ = 2) distribution at the α = 0.01 significance level are considered. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables G1U99 and G1L99 on pages 234 and 235 below. Detailed numerical results for error rate results tested at the α = 0.05 and α = 0.10 significance levels, as well as error rate results for Monte Carlo Bootstrap methods where 500 bootstrap resamples were specified can be viewed in Appendix tables. Results can also be viewed visually in figures G99, G95, and G90 on pages 236-238.

In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 10, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the gamma(α=2, λ = 2) parameter specification at the specified α = 0.01 significance level, for both the upper and lower limit, E-skew did not have the error rate with the smallest percent error at any sample size. E-skew also did not have the error rate with the smallest percent error of any method using EBSD(n) at any sample size at this significance level. EP performed relatively accurately for the upper limit at this α = 0.01 significance level compared to the other methods studied. Namely, at sample sizes 10 and 20 EP had the error rate with the smallest percent error for the upper limit among all methods considered. Additionally, the EBC method had the error rate with the smallest percent error at sample sizes 30 and 40 for the upper limit among all methods considered. These results are shown below in G1U99 and G1L99. These results can also be viewed visually in figure G99.

The E-skew method demonstrates its relative value when comparing the error rates across α significance level. In the case of the gamma(α=2, λ = 2) distribution, the E-skew method's accuracy improved relative to the percentile methods applied on EBSD(n) when the α significance level was modified from α = 0.01 to α = 0.05 and α = 0.10 for the upper limit. At the α = 0.05 and α = 0.10 significance levels for the upper limit, E-skew achieved the error rate with the smallest percent error among all methods applied on EBSD(n) at all four sample sizes.

Compared to the other methods studied, EP performed relatively more accurately at the α = 0.01 significance level and relatively less accurately at the α = 0.05 and α = 0.10 significance levels. At the α = 0.01 significance level for the upper limit, EP attained the error rate with the smallest percent error among methods applied on EBSD(n) at sample sizes 10, and 20. Then at the α = 0.05 and α = 0.10 significance levels for the upper limit, EP failed to attain the error rate with the smallest percent error at any sample size. For the upper limit the same could be said of the EBC. However for the lower limit, EBC did perform relatively accurately compared to the other methods studied across α significance level. At the α = 0.05 significance level for the lower limit, EBC attained the error rate with the smallest percent error compared to all other methods at two of four sample sizes. Further at the α = 0.10 significance level for the lower limit, EBC attained the error rate with the smallest percent error at sample size 40. For the lower limit EP also attained the error rate with the smallest percent error at the α = 0.05 significance level at sample size 40, and the smallest percent error at the α = 0.10 significance level at sample size 30.

At the α = 0.01 significance level for the upper limit, the percentile methods applied on EBSD(n) performed relatively accurately compared to their Monte Carlo Bootstrap counterparts. EP had an error rate with a smaller percent error for three of four sample sizes when compared to BP and EBC did for all four sample sizes when compared to BC. Then when significance level was modified to α = 0.05 and α = 0.10, EBC and EP were relatively less accurate than BC and BP. EP and EBC had an error rates that were larger for three of the four sample sizes for the upper limit at the α = 0.05 significance level, and for all four sample sizes at the α = 0.10 significance level when compared to their Monte Carlo Bootstrap counterpart. At the α = 0.05 and α = 0.10 significance level for the lower limit, both EP and EBC attained an error rate with a smaller percent error at three of four sample sizes considered compared to their Monte Carlo Bootstrap counterpart.

Again similar results were found when comparing ET to BT. At each significance level and sample size studied ET and BT had error rates with approximately equivalent percent errors.

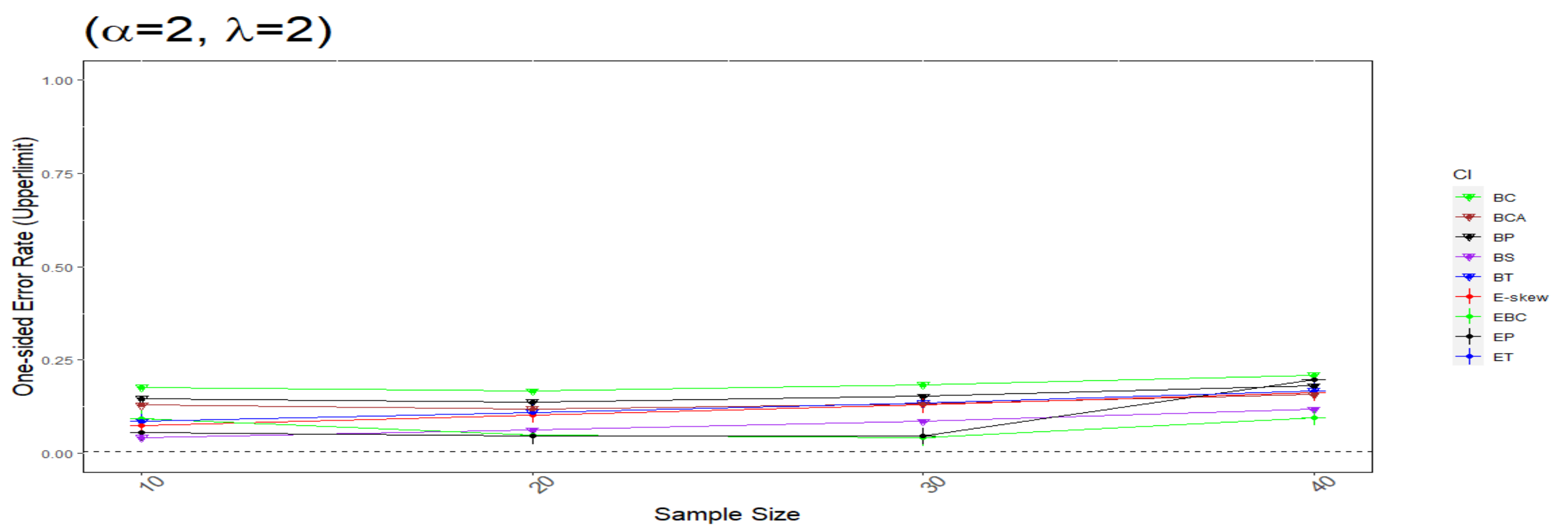
Table: Trimmed Sample Mean - G1U99 Upper limit error rate (α = 0.01), Gamma Distribution, gamma(α =2, λ=2)**, Bootstraps=10000**

Sample size	10	20	30	40
E-skew	0.0757 (1414%)	0.1039 (1978%)	0.1306 (2512%)	0.162 (3140%)
BT	0.0861 (1622%)	0.1101 (2102%)	0.1364 (2628%)	0.1686 (3272%)
ET	0.0859 (1618%)	0.1107 (2114%)	0.1364 (2628%)	0.168 (3260%)
BC	0.178 (3460%)	0.1674 (3248%)	0.184 (3580%)	0.2092 (4084%)
EBC	0.0946 (1792%)	0.0489 (878%)	**0.0422** **(744%)**	**0.0966** **(1832%)**
BP	0.1471 (2842%)	0.1381 (2662%)	0.1532 (2964%)	0.1814 (3528%)
EP	**0.0559** **(1018%)**	**0.0468** **(836%)**	0.0469 (838%)	0.1971 (3842%)
BS	0.0432 (764%)	0.0627 (1154%)	0.0864 (1628%)	0.1184 (2268%)
BC_a	0.1298 (2496%)	0.1189 (2278%)	0.1325 (2550%)	0.1581 (3062%)

Table: Trimmed Sample Mean - G1L99 Lower limit error rate (α = 0.01), Gamma Distribution, gamma(α =2, λ=2)**, Bootstraps=10000**

Sample size	10	20	30	40
E-skew	0.0028 (44%)	0.0027 (46%)	0.0023 (54%)	0.0021 (58%)
BT	0.0036 (28%)	0.0035 (30%)	0.0026 (48%)	0.0023 (54%)
ET	**0.0036** **(28%)**	0.0035 (30%)	0.0026 (48%)	0.0022 (56%)
BC	0.0147 (194%)	**0.0063** **(26%)**	0.004 (20%)	0.0032 (36%)
EBC	0.003 (40%)	6e-04 (88%)	1e-04 (98%)	0.003 (40%)
BP	0.0148 (196%)	0.0073 (46%)	0.0048 (4%)	0.0036 (28%)
EP	0.0035 (30%)	8e-04 (84%)	3e-04 (94%)	0.0015 (70%)
BS	0.0034 (32%)	0.0036 (28%)	0.0024 (52%)	0.0028 (44%)
BC_a	0.0167 (234%)	0.0091 (82%)	**0.0056** **(12%)**	**0.0042** **(16%)**

Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 99% CI for the Gamma Distribution



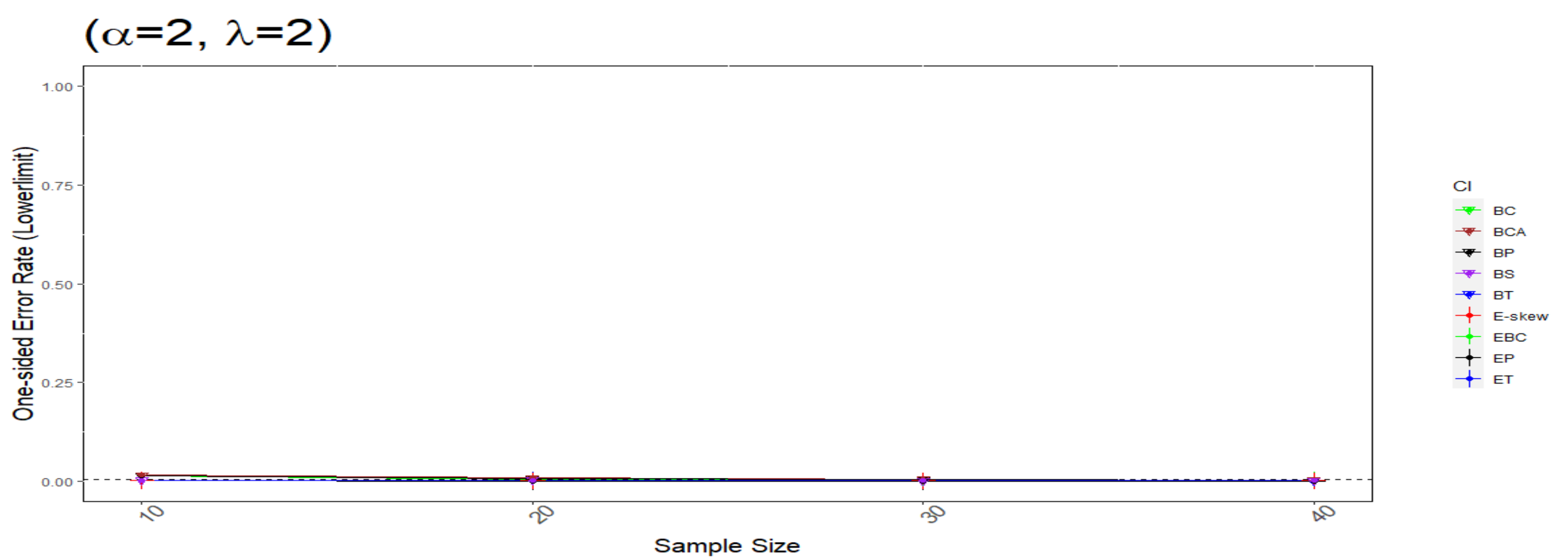
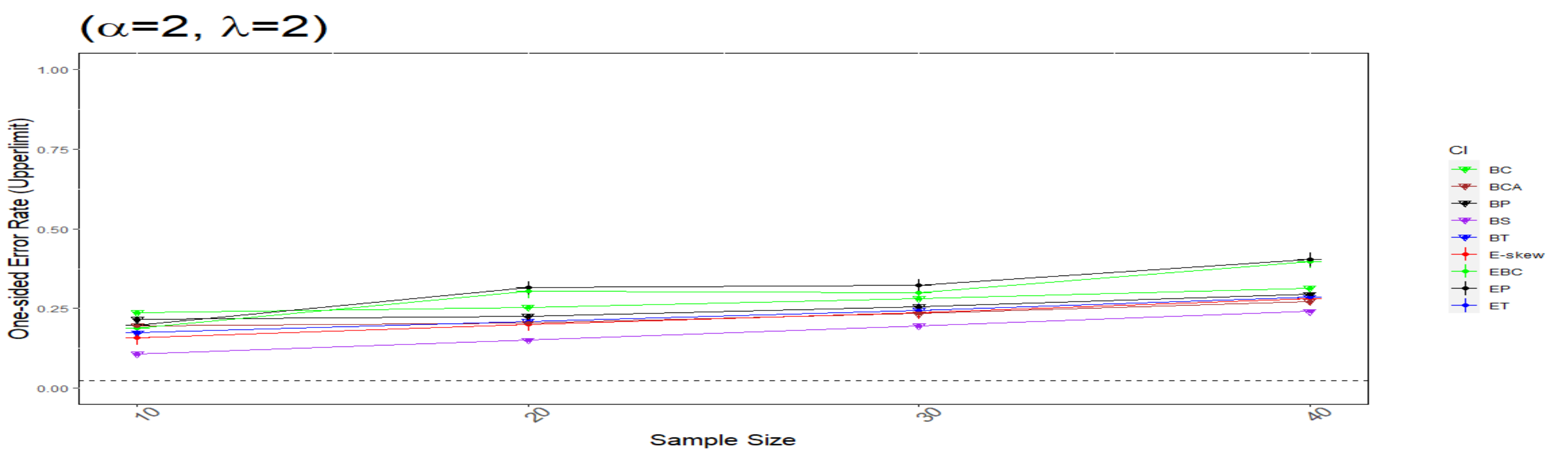


Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 95% CI for the Gamma Distribution



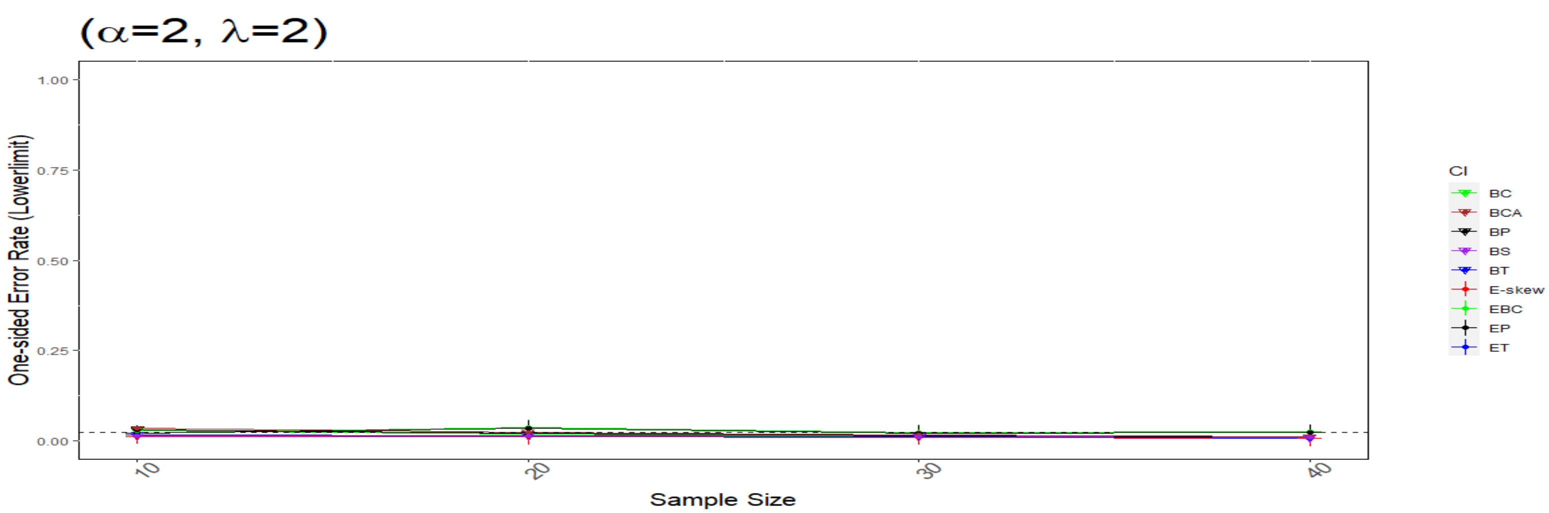
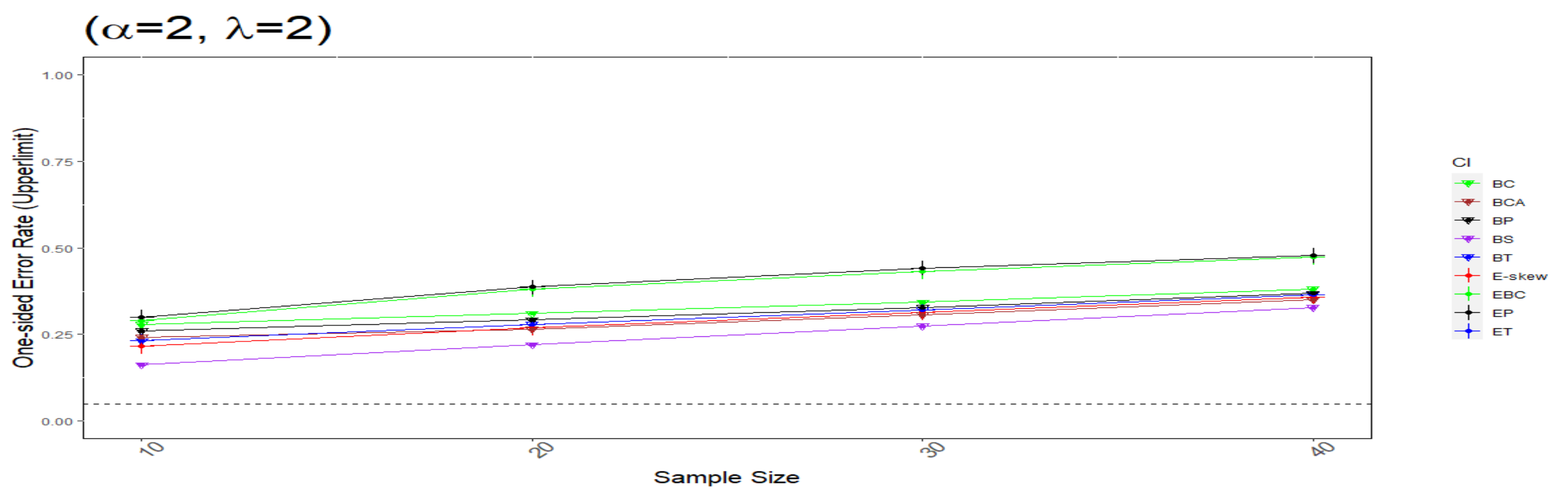
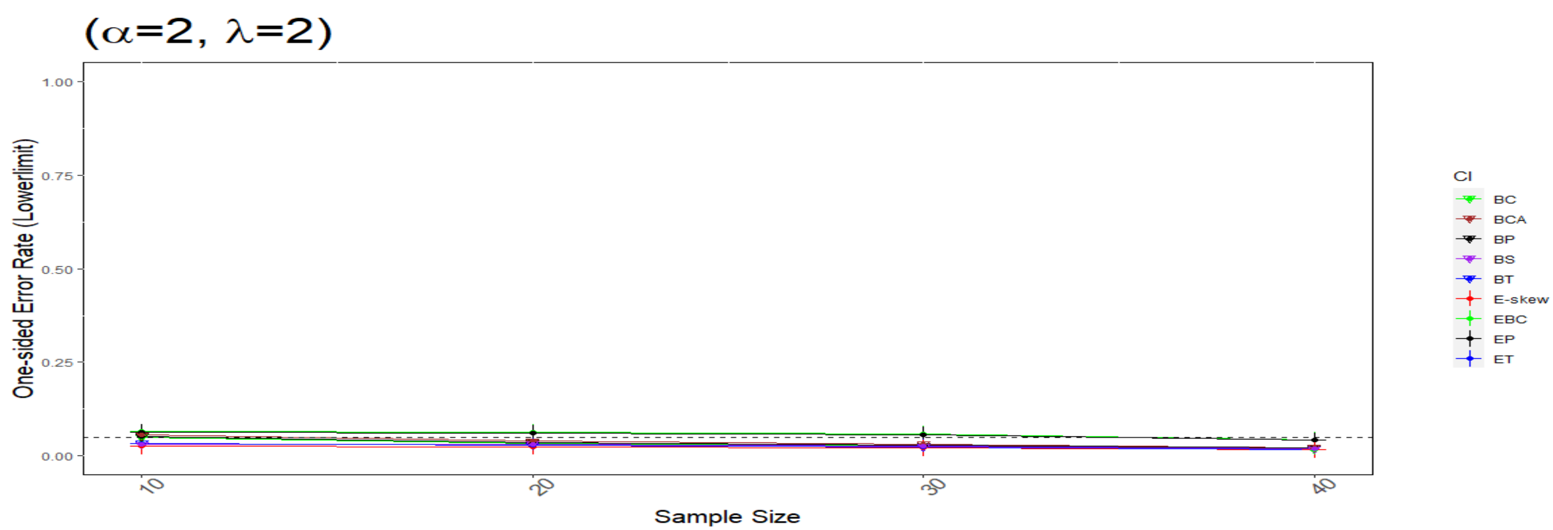


Figure: Trimmed Sample Mean - G99 - One-Sided Error Rates for 90% CI for the Gamma Distribution





d. Log-Normal Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the trimmed sample mean statistic on data drawn from a log-normal distribution. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for this same statistic and distribution. First the results for data generated from a log-normal(μ=4, σ = 0.2) distribution at the α = 0.01 significance level are considered. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables LN1U99 and LN1L99 on pages 242 and 243 below. Detailed numerical results for error rate results tested at the α = 0.05 and α = 0.10 significance levels, as well as error rate results for Monte Carlo Bootstrap methods where 500 bootstrap resamples were specified can be viewed in Appendix tables. Results can also be viewed visually in figures LN99, LN95, and LN90 on pages 244-246.

In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 10, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the log-normal(μ=4, σ = 0.2) parameter specification at the specified α = 0.01 significance level, for the upper limit, E-skew did not have the error rate with the smallest percent error at any sample size. E-skew also did not have the error rate with the smallest percent error of any method using EBSD(n) at any sample size at this significance level for the upper limit. However, for the lower limit at this significance level, E-skew did attain the error rate with the smallest percent error at one sample size.

EBC was relatively more accurate compared to the other methods studied for the upper limit at this α = 0.01 significance level. At sample sizes 20, 30, and 40 EBC had the error rate with the smallest percent error for the upper limit among all methods considered. For the lower limit EBC attained the error rate with the smallest percent error at sample size 10. These results are shown below in LN1U99 and 4LN1L99. These results can also be viewed visually in figure LN99.

The E-skew method demonstrates its value when comparing the error rates across α significance level. In the case of the log-normal(μ=4, σ = 0.2) distribution, the E-skew method's accuracy improved relative to percentile methods applied on EBSD(n) when the α significance level was modified from α = 0.01 to α = 0.05 and α = 0.10 for the lower limit. At the α = 0.05 significance level for the lower limit, E-skew achieved the error rate with the smallest percent error among all methods at all four sample sizes. At the and α = 0.10 significance level for the lower limit, E-skew achieved the error rate with the smallest percent error at three of four sample sizes. For the upper limit at both significance levels, E-skew achieved the error rate with the smallest percent error among all methods applied on EBSD(n) and second smallest overall to BS.

EBC performed relatively accurately compared to the other methods studied at the α = 0.01 significance level. At the α = 0.01 significance level for the upper limit, EBC attained the error rate with the smallest percent error among methods applied on EBSD(n) at sample sizes 20, 30 and 40. Then at the α = 0.05 and α = 0.10 significance levels EBC failed to attain the error rate with the smallest percent error at any sample size.

At the α = 0.01 significance level for both the upper and lower limit, the percentile methods applied on EBSD(n) performed relatively accurately compared to their Monte Carlo Bootstrap counterparts. EP had an error rate with a smaller percent error for all four sample sizes when compared to BP and EBC did for all four sample sizes when compared to BC. Then when significance level was modified to α = 0.05, EBC had an error rate with a larger percent error compared to BC for three of the four sample sizes studied. When significance level was modified to α = 0.10, EBC had an error rate with a larger percent compared to BC for all four sample sizes studied. Similarly, EP had an error rates with a larger percent error for three of the four sample sizes for the upper and lower limit at the α = 0.05 significance level when compared to their Monte Carlo Bootstrap counterpart. At the α = 0.10 significance level, for the upper limit and lower limit, EP had a larger percent error at each sample size when compared to their Monte Carlo Bootstrap counterpart.

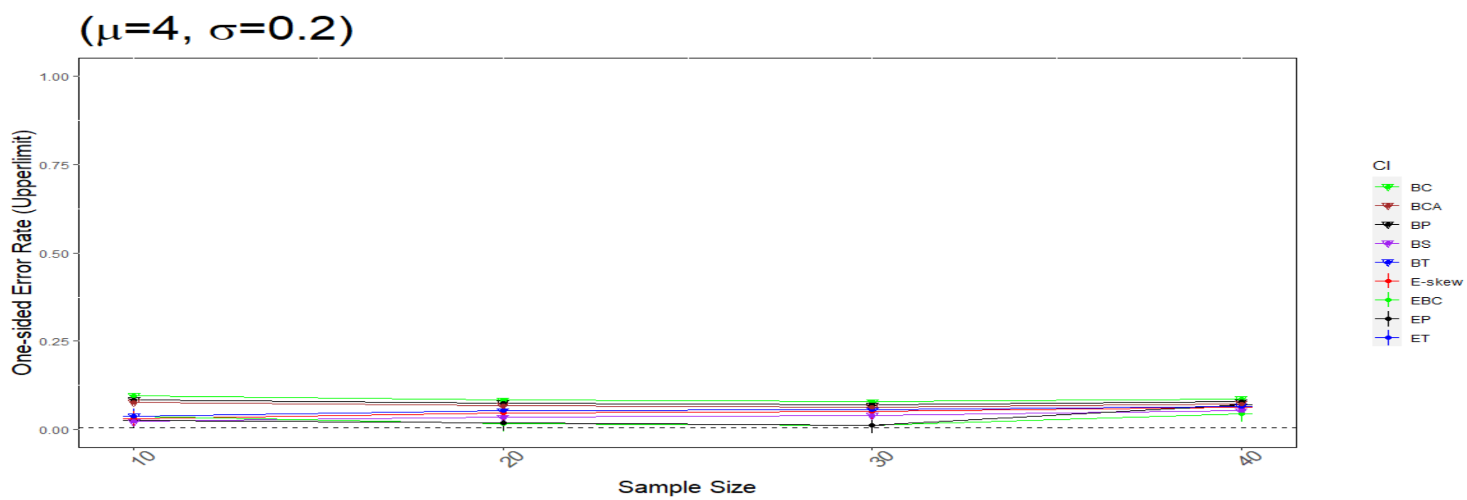
Again similar results were found when comparing ET to BT. At each significance level and sample size studied ET and BT had error rates with approximately equivalent percent errors.

Table: Trimmed Sample Mean - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distribution, log-normal($\mu = 4$, σ =0.2), **Bootstraps=10000**

Sample size	10	20	30	40
E-skew	0.031 (520%)	0.0481 (862%)	0.0522 (944%)	0.0634 (1168%)
BT	0.0385 (670%)	0.0537 (974%)	0.0558 (1016%)	0.0663 (1226%)
ET	0.0387 (674%)	0.0541 (982%)	0.0558 (1016%)	0.0666 (1232%)
BC	0.097 (1840%)	0.0841 (1582%)	0.0791 (1482%)	0.0869 (1638%)
EBC	0.0376 (652%)	**0.0166 (232%)**	**0.0114 (128%)**	**0.0448 (796%)**
BP	0.0854 (1608%)	0.0752 (1404%)	0.0696 (1292%)	0.0794 (1488%)
EP	0.0262 (424%)	0.0198 (296%)	0.0123 (146%)	0.0701 (1302%)
BS	**0.023 (360%)**	0.0351 (602%)	0.0404 (708%)	0.0536 (972%)
BC_a	0.078 (1460%)	0.068 (1260%)	0.0641 (1182%)	0.0738 (1376%)

Table: Trimmed Sample Mean - LN1L99 Lower limit error rate (α = 0.01), Log-Normal Distribution, log-normal($\mu = 4$, σ =0.2)**, Bootstraps=10000**				
Sample size	10	20	30	40
E-skew	0.0071 (42%)	0.0085 (70%)	0.0073 (46%)	**0.007** **(40%)**
BT	0.009 (80%)	0.0103 (106%)	0.008 (60%)	0.0079 (58%)
ET	0.0091 (82%)	0.0103 (106%)	0.008 (60%)	0.0081 (62%)
BC	0.0324 (548%)	0.0192 (284%)	0.0134 (168%)	0.0107 (114%)
EBC	**0.0068** **(36%)**	0.0014 (72%)	7e-04 (86%)	0.009 (80%)
BP	0.0321 (542%)	0.0204 (308%)	0.0137 (174%)	0.012 (140%)
EP	0.0079 (58%)	0.0022 (56%)	9e-04 (82%)	0.0082 (64%)
BS	0.0081 (62%)	**0.0082** **(64%)**	**0.0072** **(44%)**	0.0079 (58%)
BC_a	0.0324 (548%)	0.0216 (332%)	0.0143 (186%)	0.0129 (158%)

Figure: Trimmed Sample Mean - LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution



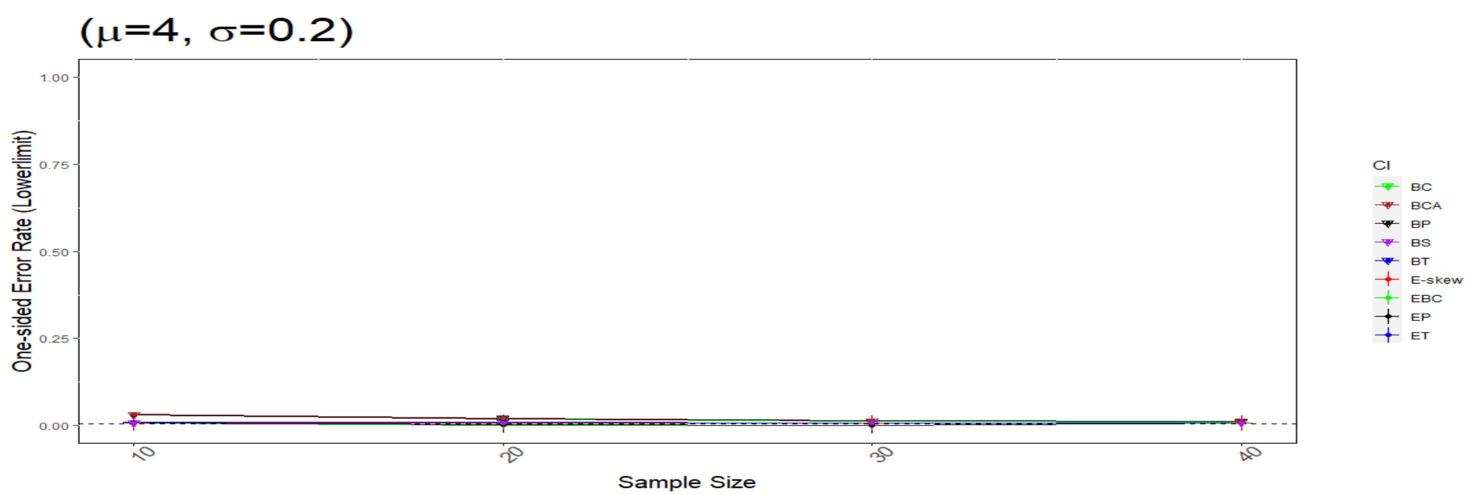
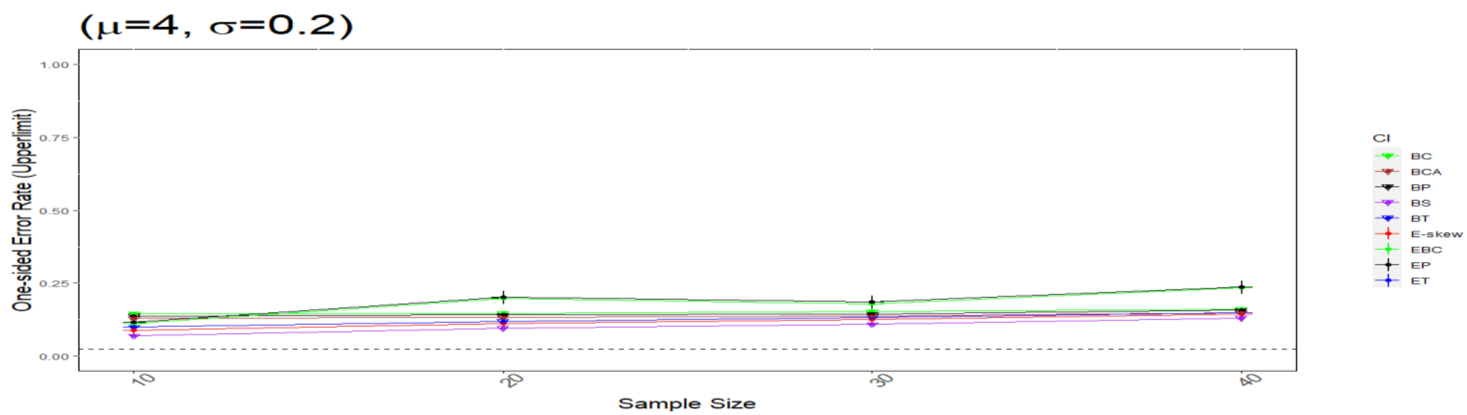


Figure: Trimmed Sample Mean - LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution



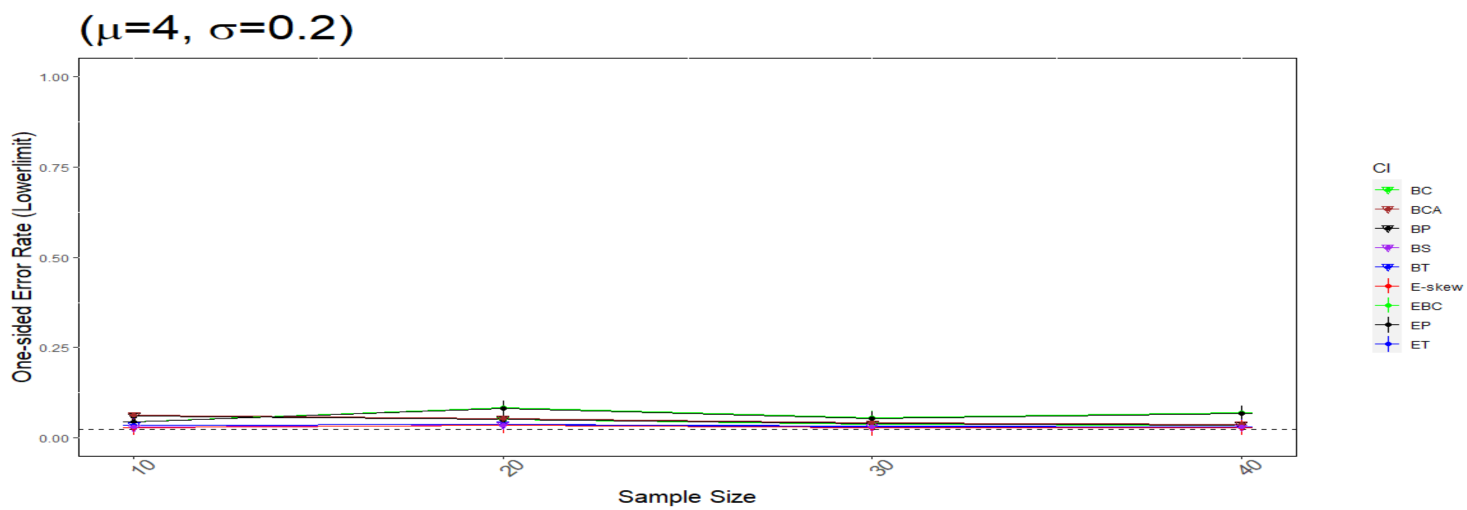
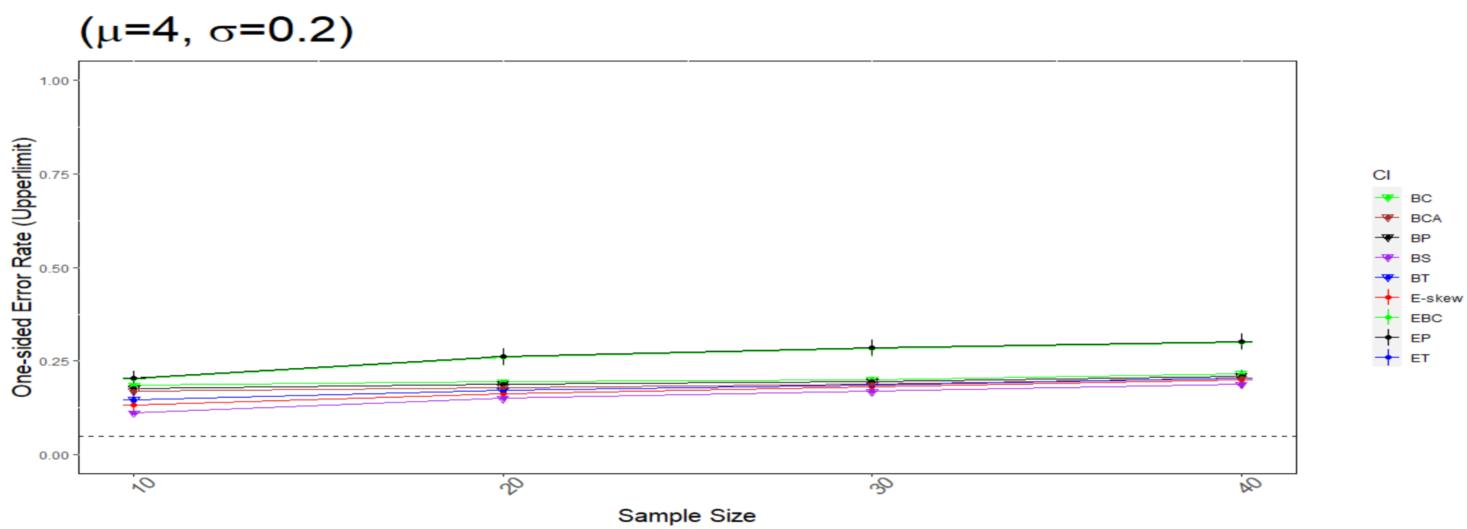
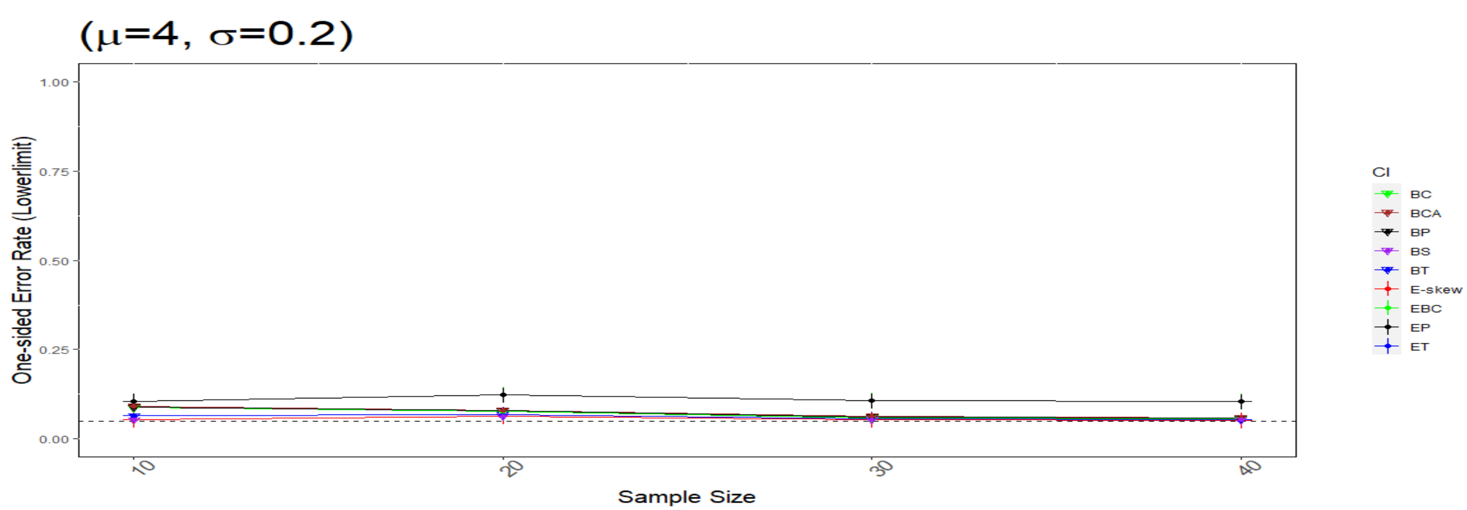


Figure: Trimmed Sample Mean - LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution





e. Mixture Distribution

The first purpose of this sub section is to compare the accuracy of E-skew to the accuracy of all other methods studied for the trimmed sample mean statistic on data drawn from the mixture of two normal distributions. The second is to compare the accuracy of other methods that use the EBSD(n) method to the accuracy of Monte Carlo Bootstrap methods for this same statistic and distribution. First the results for data generated from a distribution at the α = 0.01 significance level are considered. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables MN1U99 and MN1L99 on pages 250 and 251 below. Detailed numerical results for error rate results tested at the α = 0.05 and α = 0.10 significance levels, as well as error rate results for Monte Carlo Bootstrap methods where 500 bootstrap resamples were specified can be viewed in Appendix tables. Results can also be viewed visually in figures MN99, MN95, and MN90 on pages 252-254.

In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 10, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

For the 0.6*$N(\mu_1 = 4, \sigma_1 = 4)$ + 0.4*$N(\mu_2 = 8, \sigma_2 = 8)$ parameter specification at the specified α = 0.01 significance level, for the upper limit, E-skew did not have the error rate with the smallest percent error at any sample size. E-skew did have the error rate with the smallest percent error for three of four sample sizes at this significance level for the lower limit.

EP performed relatively accurately compared to the other methods studied for the upper limit at this α = 0.01 significance level. Namely, at sample size 10 EP and EBC both had the error rate with the smallest percent error for the upper limit among all methods considered. For the lower limit EP attained the error rate with the smallest percent error at sample size 40. Additionally, the EBC method had the error rate with the smallest percent error at sample sizes 20, 30 and 40 for the upper limit among all methods considered. These results are shown below in MN1U99 and MN1L99. These results can also be viewed visually in figure MN99.

The E-skew method demonstrates its value when comparing error rates across α significance level. Once again E-skew performed relatively more accurately at the α = 0.05 and α = 0.10 significance levels compared to other methods applied on EBSD(n). At both significance levels for the upper limit, E-skew attained the error rate with the smallest percent error at each sample size among methods applied on EBSD(n). At the α = 0.05 significance level for the lower limit, E-skew had the error rate with the smallest percent error for three of four sample sizes. At the α = 0.10 significance level for the lower limit, E-skew had the error rate with the smallest percent error at all four sample sizes studied.

Again for this distributional case both methods, EP and EBC, at the α = 0.05 and α = 0.10 significance levels for both limit ends, performed relatively less accurately than their Monte Carlo Bootstrap counterpart. At the $\alpha = 0.05$ significance level for both the upper and lower limit, EP had an error rate with a larger percent error for three of four sample sizes compared to BP. At the $\alpha = 0.10$ significance level for the upper and lower limit, EP had an error rate with a larger percent error for all four sample sizes compared to BP. For EBC similar results were found as what was found for EP. EBC had an error rate with a larger percent error compared to BC at each significance level and for both limit ends for every sample size except sample size 5 at the $\alpha = 0.01$ significance level.

Again similar results were found when comparing ET to BT to what was found when comparing ET to BT for the previous comparisons in this section. At each significance level and sample size studied ET and BT had error rates with approximately equivalent percent errors.

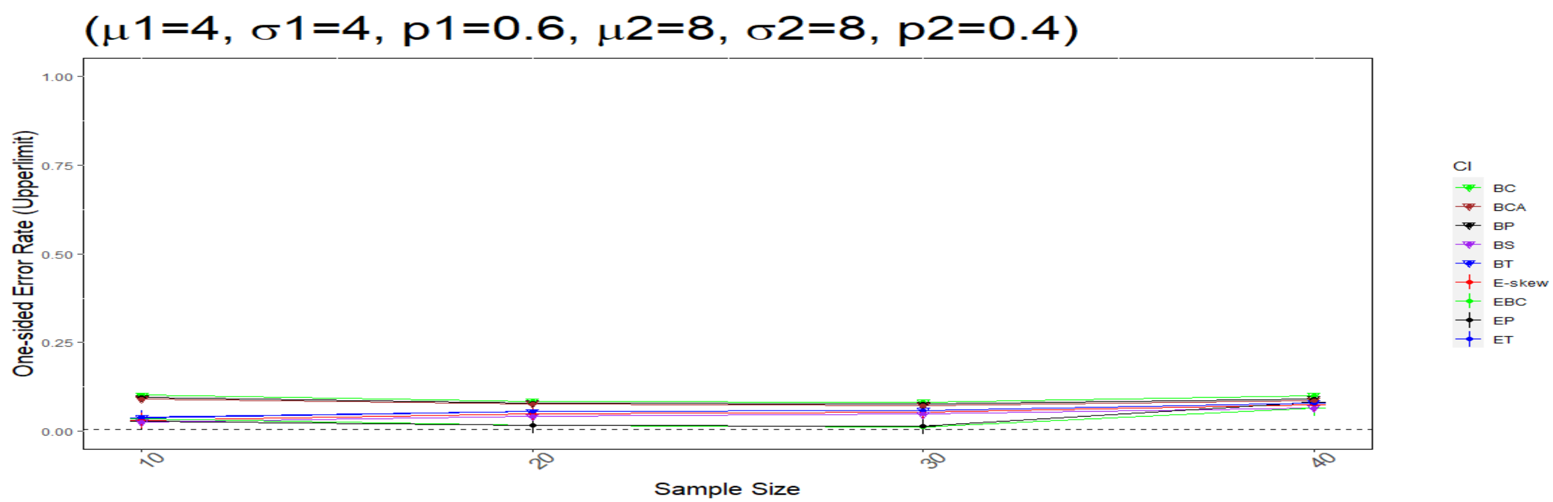
Table: Trimmed Sample Mean - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions: 0.6*$N_1(\mu_1 = 4, \sigma_1 = 4)$ + 0.4*$N_2(\mu_2 = 8,\ \sigma_2 = 8)$, $\mathbf{Bootstraps = 10000}$

Sample size	10	20	30	40
E-skew	0.0309 (518%)	0.0503 (906%)	0.0553 (1006%)	0.076 (1420%)
BT	0.0394 (688%)	0.0557 (1014%)	0.0589 (1078%)	0.0795 (1490%)
ET	0.039 (680%)	0.0557 (1014%)	0.0587 (1074%)	0.0792 (1484%)
BC	0.1021 (1942%)	0.0843 (1586%)	0.0825 (1550%)	0.1005 (1910%)
EBC	0.0349 (598%)	**0.0176** **(252%)**	**0.0136** **(172%)**	**0.0647** **(1194%)**
BP	0.0953 (1806%)	0.0801 (1502%)	0.0771 (1442%)	0.0922 (1744%)
EP	0.0285 (470%)	**0.0176** **(252%)**	0.0141 (182%)	0.0816 (1532%)
BS	**0.0274** **(448%)**	0.0423 (746%)	0.0486 (872%)	0.0657 (1214%)
BC_a	0.0914 (1728%)	0.0776 (1452%)	0.0736 (1372%)	0.0871 (1642%)

Table: Trimmed Sample Mean - MN1L99 Lower limit error rate (α = 0.01), Mixture of two Normal Distributions: 0.6*$N_1(\mu_1 = 4, \sigma_1 = 4)$ + 0.4*$N_2(\mu_2 = 8,\ \sigma_2 = 8)$, $\mathbf{Bootstraps = 10000}$

Sample size	10	20	30	40
E-skew	**0.0046** (**8%**)	**0.0068** (**36%**)	**0.0092** (**84%**)	0.0084 (68%)
BT	0.0059 (18%)	0.0086 (72%)	0.0106 (112%)	0.009 (80%)
ET	0.0059 (18%)	0.0087 (74%)	0.0106 (112%)	0.0091 (82%)
BC	0.0248 (396%)	0.0153 (206%)	0.0154 (208%)	0.0122 (144%)
EBC	0.0043 (14%)	0.0024 (52%)	7e-04 (86%)	0.0139 (178%)
BP	0.0301 (502%)	0.019 (280%)	0.0181 (262%)	0.0139 (178%)
EP	0.0055 (10%)	0.0018 (64%)	6e-04 (88%)	**0.0063** (**26%**)
BS	0.0068 (36%)	0.0089 (78%)	0.0117 (134%)	0.0111 (122%)
BC_a	0.0339 (578%)	0.0222 (344%)	0.0203 (306%)	0.0169 (238%)

Figure: Trimmed Sample Mean - MN99 - One-Sided Error Rates for 99% CI for the Mixure of two Normals



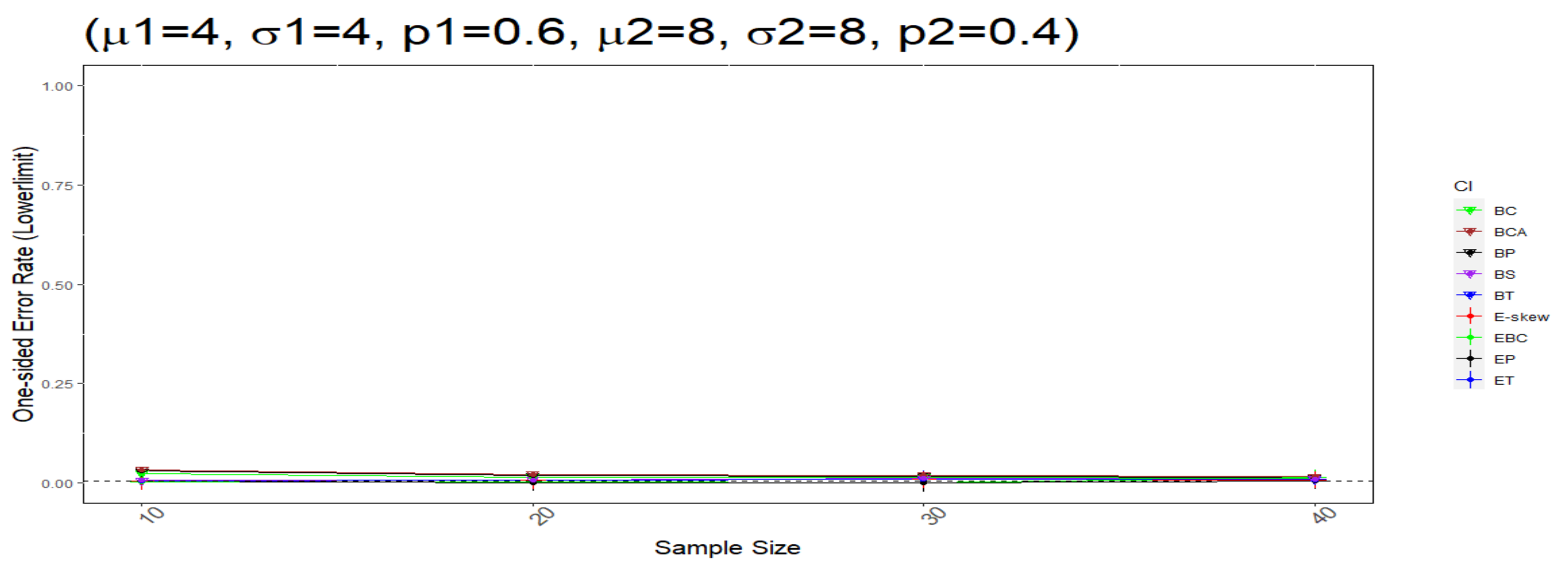
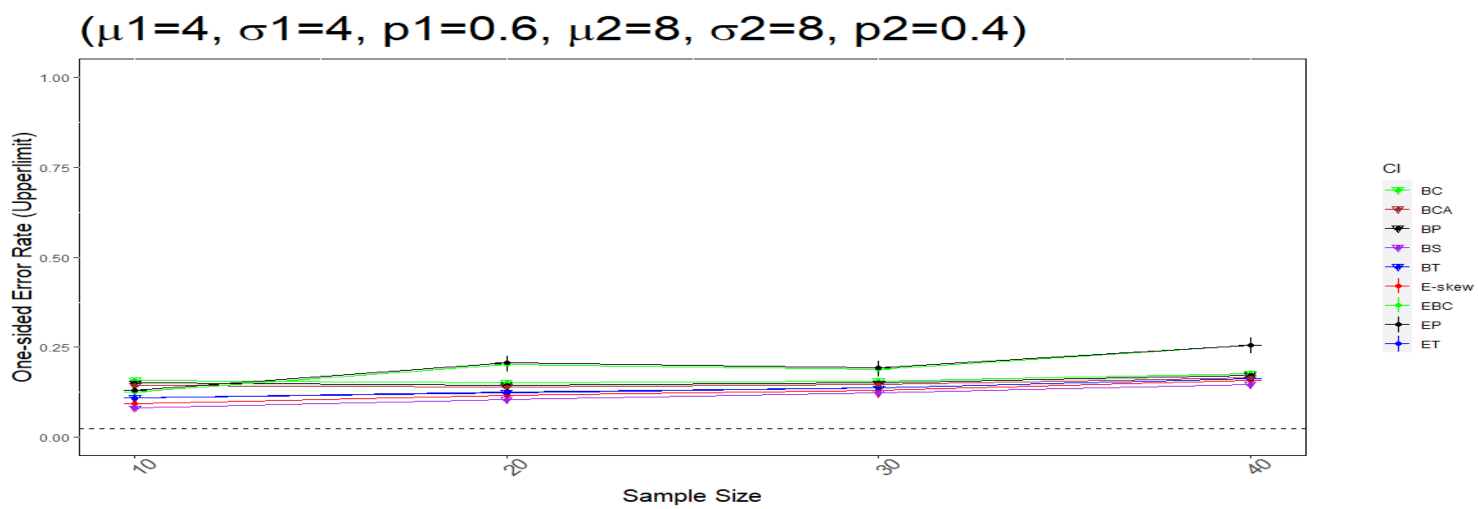


Figure: Trimmed Sample Mean - MN95 - One-Sided Error Rates for 95% CI for the Mixure of two Normals



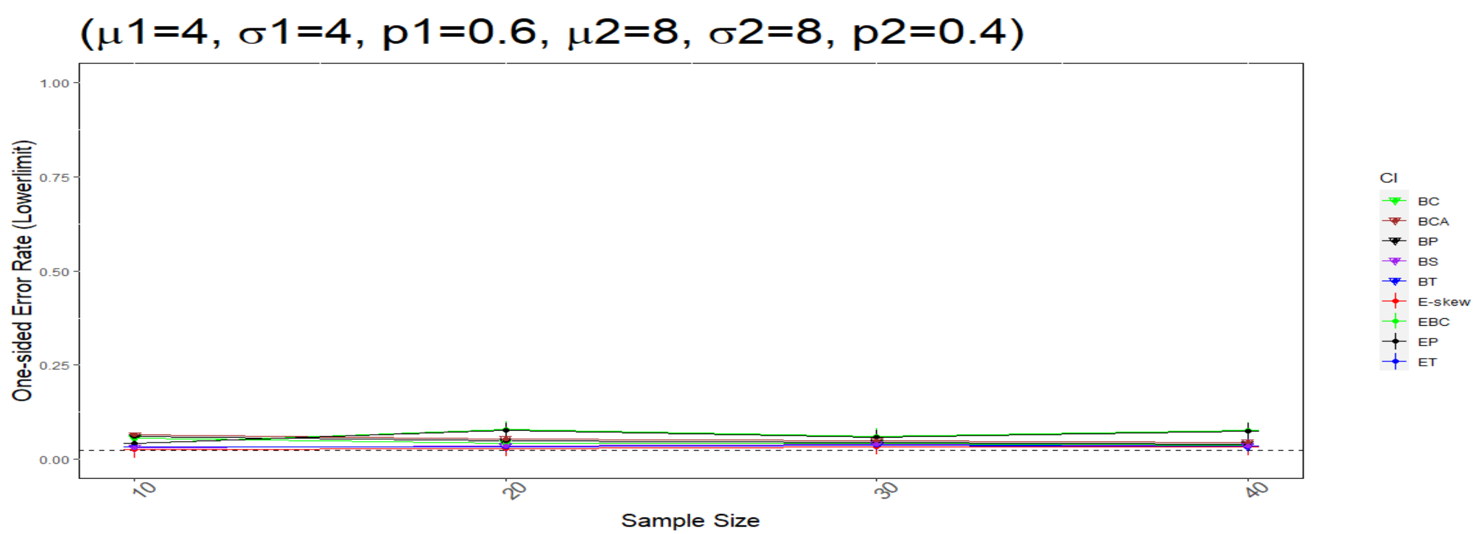
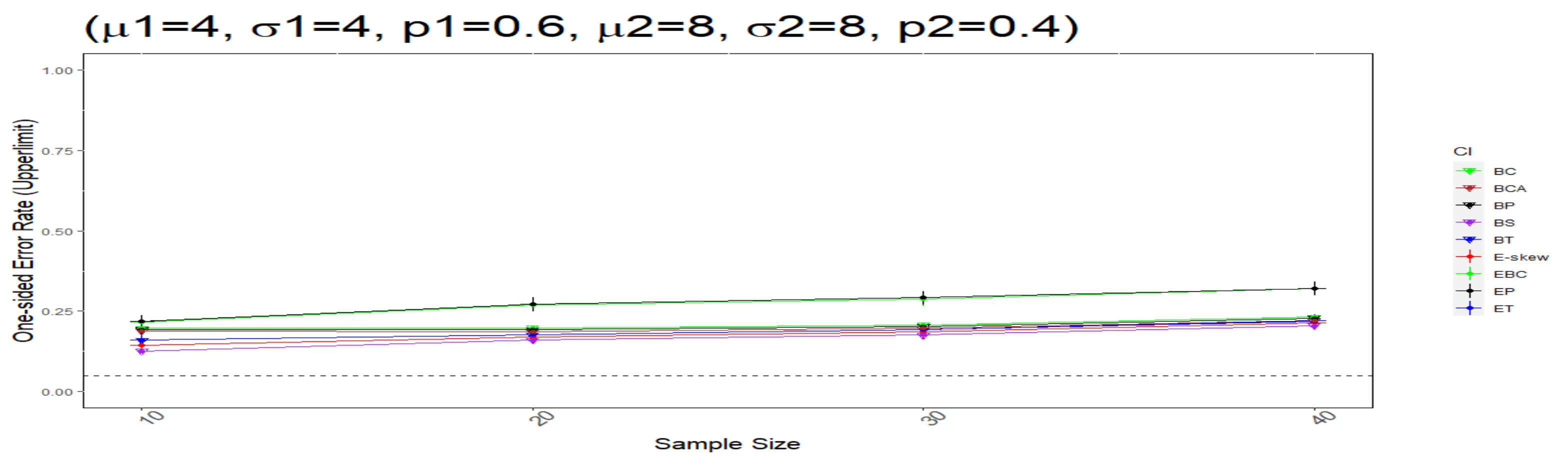
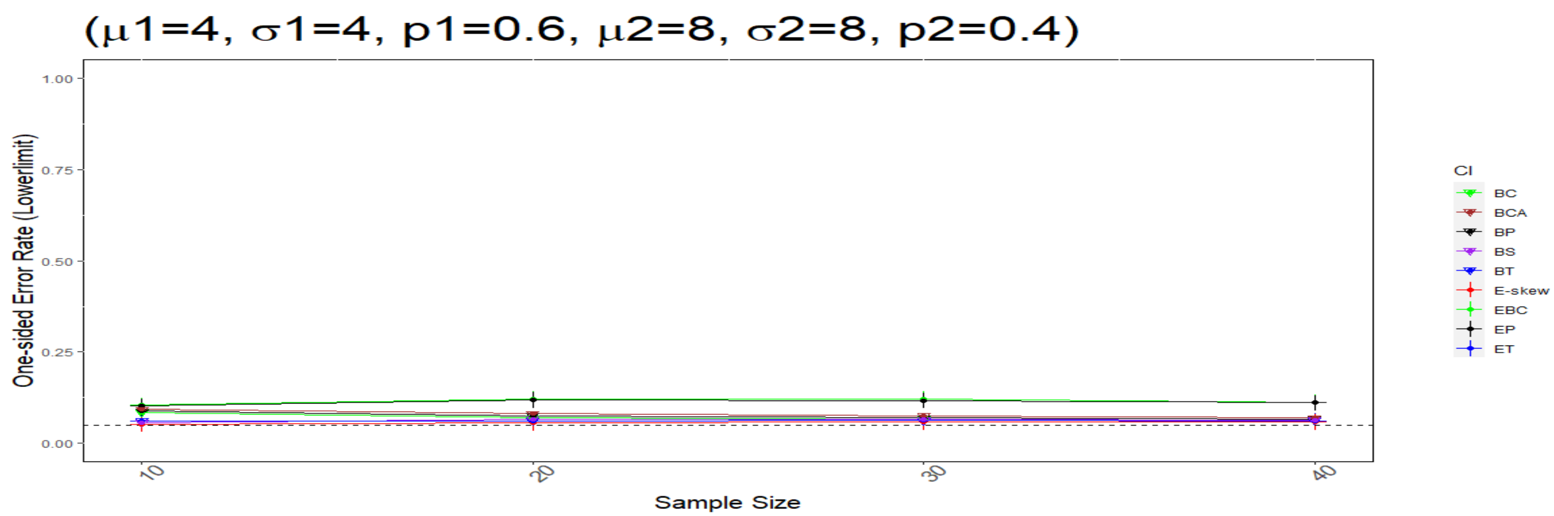


Figure: Trimmed Sample Mean - MN90 - One-Sided Error Rates for 90% CI for the Mixure of two Normals





Trimmed Sample Mean Results Discussion for methods using EBSD(n)

In general, the E-skew method yielded error rates for the trimmed sample mean with smaller percent errors than any other method applied on EBSD(n) for normally distributed data at the α = 0.05 and α = 0.10 significance levels. In general, E-skew performed relatively more accurately in comparison to other methods at the α = 0.05 and α = 0.10 significance levels than it did at the α = 0.01 significance level.

For moderately skewed data, like data generated from the exponential distribution, E-skew outperformed every other method implemented on EBSD(n) at the α = 0.05 and α = 0.10 significance levels for the upper limit. E-skew consistently performed better for the upper limit at these significance levels compared to any other method applied on EBSD(n).

Percentile methods applied on EBSD(n) performed better at the α = 0.01 significance level. At this significance level the method applied on EBSD(n) outperformed their Monte Carlo Bootstrap counterpart frequently for both the upper and lower limit. However, at the α = 0.05 and α = 0.10 significance levels, percentile methods applied on EBSD(n) performed less accurately than their Monte Carlo Bootstrap counterpart. ET performed approximately equivalently well as BT for each distribution studied.

4.5 Sample Median

For the sample median portion of the simulation study, results for six different sample sizes are reported (n = 5, 10, 15, 20, 30, and 40). For each of these sample sizes, confidence interval error rates are reported at the α = 0.01, 0.05, and 0.10 significance levels.

The probability distributions used in the simulation study for the sample mean were the normal, exponential, Cauchy, log-normal, and mixture of two normal distributions. For each distribution, the population parameters specified are displayed below in Table 4.5. These parameter specifications are the same as the specifications for the sample mean in Table 3.5 in Chapter 3.

For each sample size, population parameter specification, and probability distribution combination 10,000 separate samples were generated. For the Monte Carlo Bootstrap confidence interval methods each of the 10,000 samples used 5,000 Monte Carlo Bootstrap resamples to create its Bootstrap sampling distribution. The comparisons discussed in this section are made between EBSD(n) methods and Bootstrap methods that use 5,000 first level Bootstrap resamples. In addition, confidence interval method error rate results were measured on the same 10,000 unique samples using 1000 and 500 Bootstrap resamples. This alternative Bootstrap resampling level was performed for each distribution tested. The error rate results at these additional Bootstrap resampling levels are reported in the Appendix.

For the BS method second level Bootstrap iterations were required. This is because a closed form solution for the variance of the sample median is not available for every sample size case studied. Therefore, for the sake of consistency, second level Bootstraps were used for the BS method for every sample size studied in this section. Due to programming limitations only 5 second level Bootstrap iterations were able to be performed computationally. When using 5000 first level Bootstrap iterations BS still performed relatively accurately despite using only 5 second level Bootstrap iterations.

E-skew confidence intervals were not generated for the median statistic. E-skew confidence intervals were not generated because the median statistic is not effected by skewed data the same way the mean statistic is. The purpose of E-skew is to provide adjusted confidence intervals in the case of data that is non-normally distributed so that the resulting confidence interval has an implied error rate that matches the theoretical nominal error rate. The implied error rates for confidence intervals generated for the median statistic are not impacted by skew. Therefore, using E-skew confidence intervals for the median statistic was deemed unnecessary. Instead ET, EBC and EP confidence intervals alone should suffice in evaluating the viability of using EBSD(n) to generate confidence intervals for the median statistic. Each generated unique sample had confidence intervals computed using the confidence interval methods listed below.

- For methods using EBSD(n) this included: ET, EBC, and EP.
- For methods using the Monte Carlo Bootstrap this includes: BT, BC, BP, BC_a/ABC, and BS.

Below in Table 4.5 s a description of the parameter specifications used for the trimmed sample mean statistic in this simulation study:

Table 4.5 Simulation Parameter Specifications for the Median Statistic		
Probability distribution	Population Parameter	Parameter code: Specified Parameter Values
Normal distribution	(μ, σ)	N1: (4, 1)
Exponential distribution	(λ)	E1: (0.10)
Cauchy distribution	(x_0, γ)	C1: (0, 1)
Log-Normal distribution	(μ, σ)	L1: (4, 0.2)
Mixture of two normal distributions	$(\mu_1, \sigma_1, p_1, \mu_2, \sigma_2, p_2)$	M1: (4, 1, 0.5, 8, 1, 0.5)

a. Normal Distribution

The purpose of this sub section is to compare the performance of methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for the sample median statistic. For the normal distribution one parameter specification type was generated. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables N1U and N1L on pages 260 and 261 below. For the sample median statistic the one specification simulated was studied at three different α significance levels. However, in this section because of the volume of error rate results, only the error rate results at the $\alpha = 0.01$ significance level is displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the $N(\mu = 4, \sigma = 1)$ at the $\alpha = 0.01$ significance level results can be viewed visually in figures N99, N95, and N90 on pages 262-264. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use

EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

Percentile type algorithms that used EBSD(n) performed relatively more accurately than Monte Carlo Bootstrap methods at the α = 0.01 significance level for the upper limit. For the upper limit, EP did achieve the error rate with the smallest percent error when compared to any other method at sample sizes 15 and 30. Additionally, for the lower limit EP achieved the error rate with the smallest percent error at sample size 15.

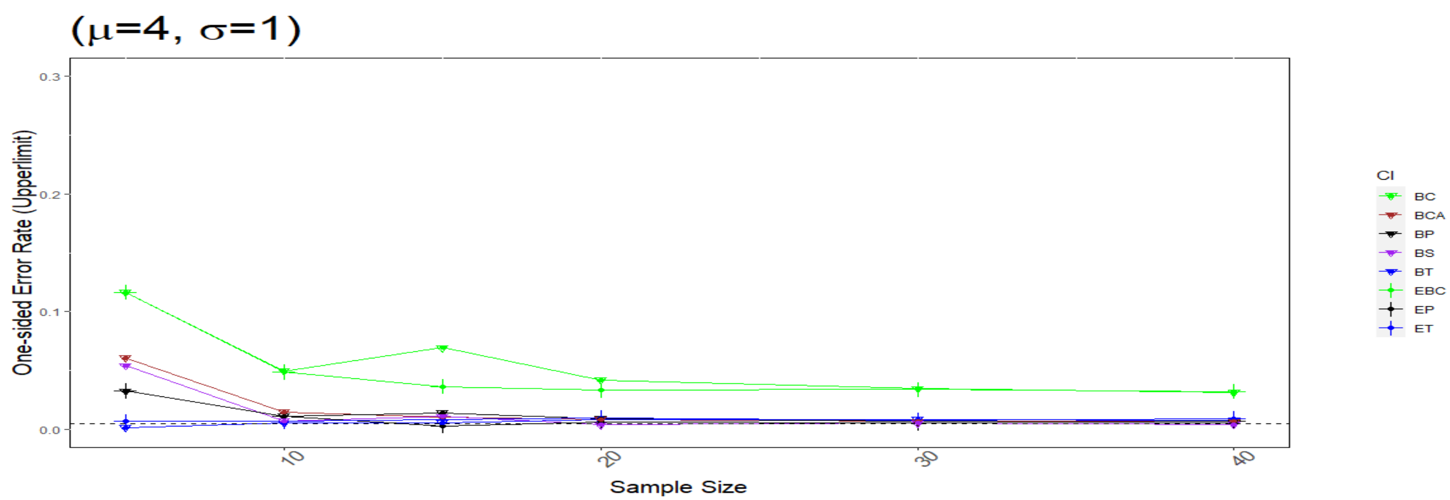
However, at the α = 0.05 and α = 0.10 significance levels, EP did not achieve an error rate with the smallest percent error at any sample size for the upper or lower limit. Further, EBC did not achieve an error rate with the smallest percent error at any of the three significance levels. When comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors at multiple sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels though, the percentile algorithms applied on EBSD(n) had an error rate with a larger percent error at each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart. Additionally, ET had an error rate with a larger percent error compared to BT for both the upper and lower limit at each significance level for every sample size except sample size 5. At sample size 5, ET had the error rate with the smallest percent error compared to every other method studied for each significance level for both the upper and lower limit except for the upper limit at the α = 0.05 level.

Table: Sample Median - N1U99 Upper limit error rate (α = 0.01), Normal distributions, N(μ=4, σ=1), Bootstraps=5000

Sample size	5	10	15	20	30	40
BT	0.0018 (64%)	0.0061 (22%)	0.0074 (48%)	0.0082 (64%)	0.0077 (54%)	0.007 (40%)
ET	**0.007** **(40%)**	0.0073 (46%)	0.0087 (74%)	0.0099 (98%)	0.008 (60%)	0.0093 (86%)
BC	0.112 (2140%)	0.0574 (1048%)	0.0686 (1272%)	0.0448 (796%)	0.0381 (662%)	0.0317 (534%)
EBC	0.1166 (2232%)	0.0489 (878%)	0.0367 (634%)	0.0333 (566%)	0.034 (580%)	0.0325 (550%)
BP	0.03 (500%)	0.0122 (144%)	0.0165 (230%)	0.0091 (82%)	0.0082 (64%)	**0.0053** **(6%)**
EP	0.0328 (556%)	0.0112 (124%)	**0.0033** **(34%)**	0.0064 (28%)	**0.0055** **(10%)**	0.0071 (42%)
BS	0.0514 (928%)	**0.0058** **(16%)**	0.0104 (108%)	**0.0054** **(8%)**	0.0044 (12%)	0.0037 (26%)
BC_a	0.0539 (978%)	0.0155 (210%)	0.0132 (164%)	0.0096 (92%)	0.0089 (78%)	0.0065 (30%)

Table: Sample Median - N1L99 Lower limit error rate (α = 0.01), Normal distribution, N(μ=4, σ=1), Bootstraps=5000						
Sample size	5	10	15	20	30	40
BT	0.0018 (64%)	0.0061 (22%)	0.0074 (48%)	0.0082 (64%)	0.0077 (54%)	0.007 (40%)
ET	**0.006** **(20%)**	0.0084 (68%)	0.0088 (76%)	0.0089 (78%)	0.0081 (62%)	0.0089 (78%)
BC	0.112 (2140%)	0.0574 (1048%)	0.0686 (1272%)	0.0448 (796%)	0.0381 (662%)	0.0317 (534%)
EBC	0.112 (2140%)	0.0559 (1018%)	0.0381 (662%)	0.0376 (652%)	0.0372 (644%)	0.0323 (546%)
BP	0.03 (500%)	0.0122 (144%)	0.0165 (230%)	0.0091 (82%)	0.0082 (64%)	**0.0053** **(6%)**
EP	0.03 (500%)	0.0121 (142%)	**0.0048** **(4%)**	0.0074 (48%)	0.0082 (64%)	0.0054 (8%)
BS	0.0514 (928%)	**0.0058** **(16%)**	0.0104 (108%)	**0.0054** **(8%)**	**0.0044** **(12%)**	0.0037 (26%)
BC_a	0.0539 (978%)	0.0155 (210%)	0.0132 (164%)	0.0096 (92%)	0.0089 (78%)	0.0065 (30%)

Figure: Sample Median - N99 - One-Sided Error Rates for 99% CI for the Normal Distribution



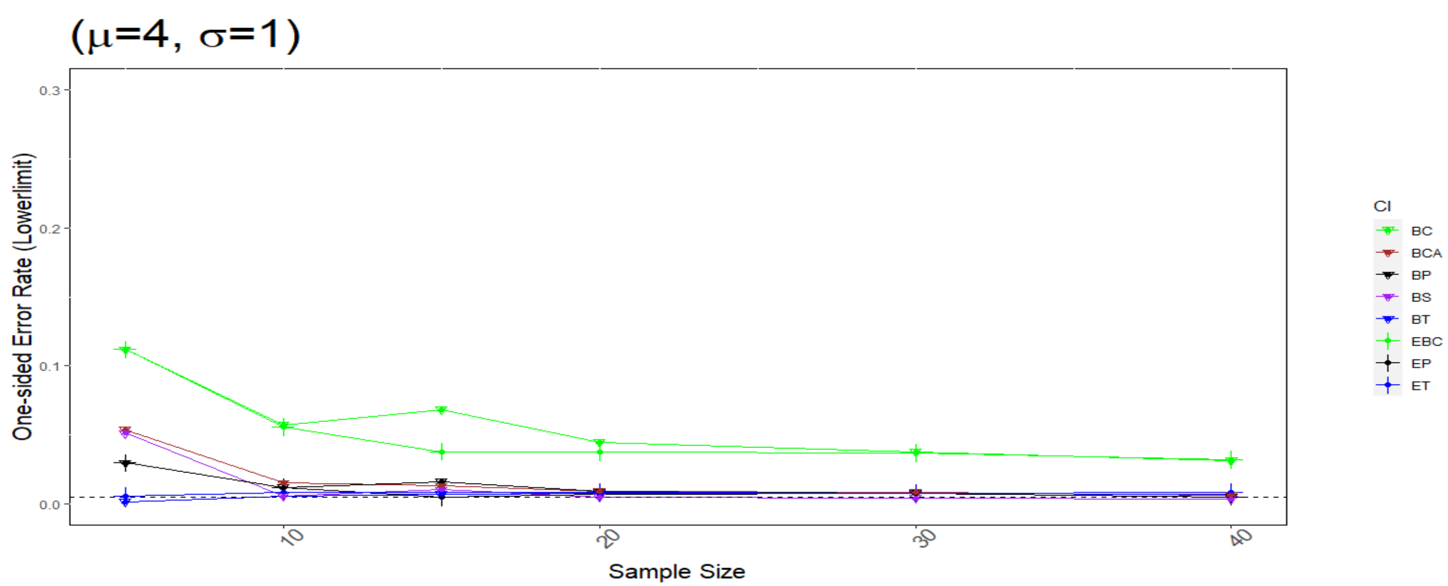
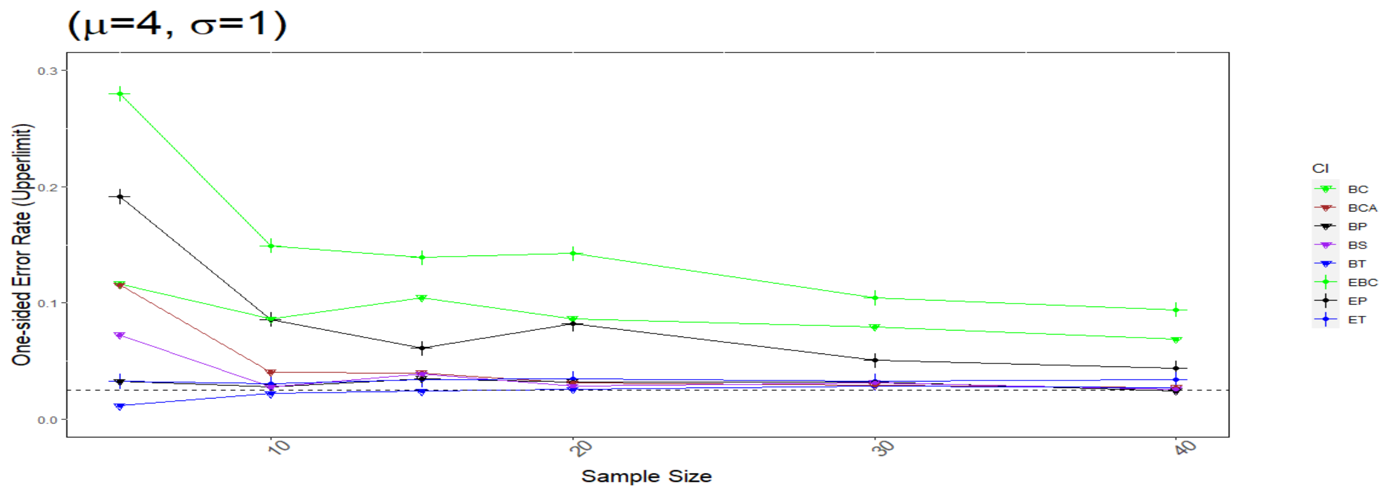


Figure: Sample Median - N95 - One-Sided Error Rates for 95% CI for the Normal Distribution



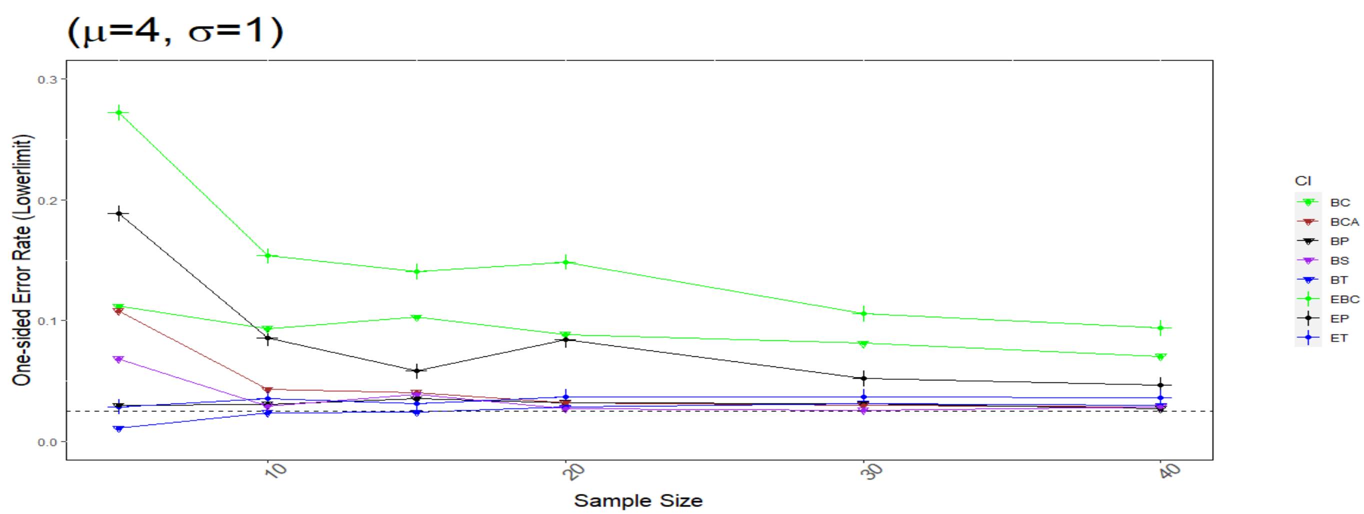
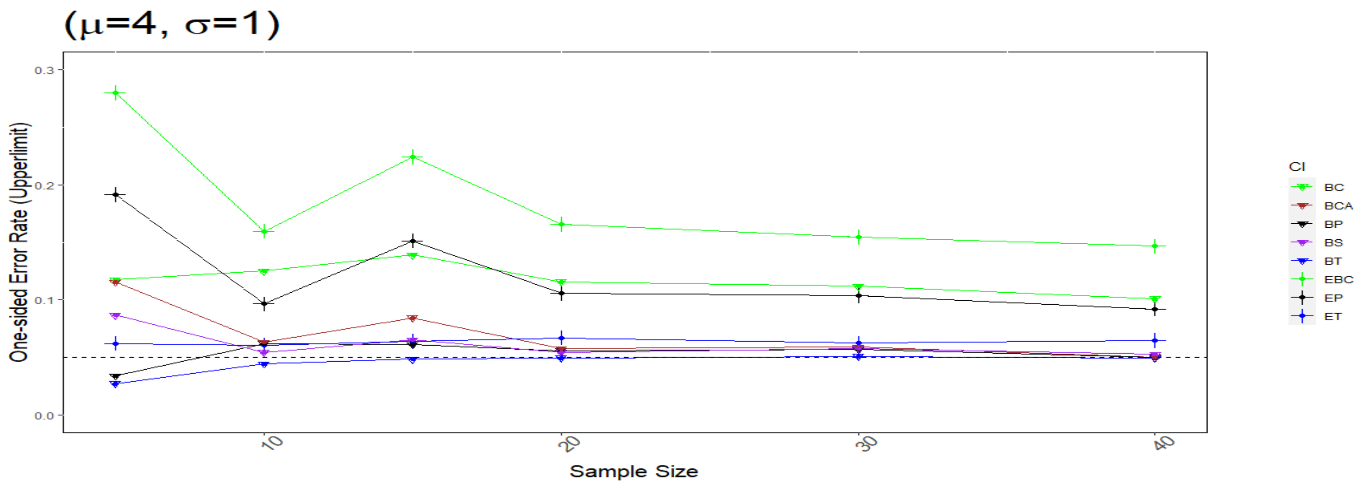
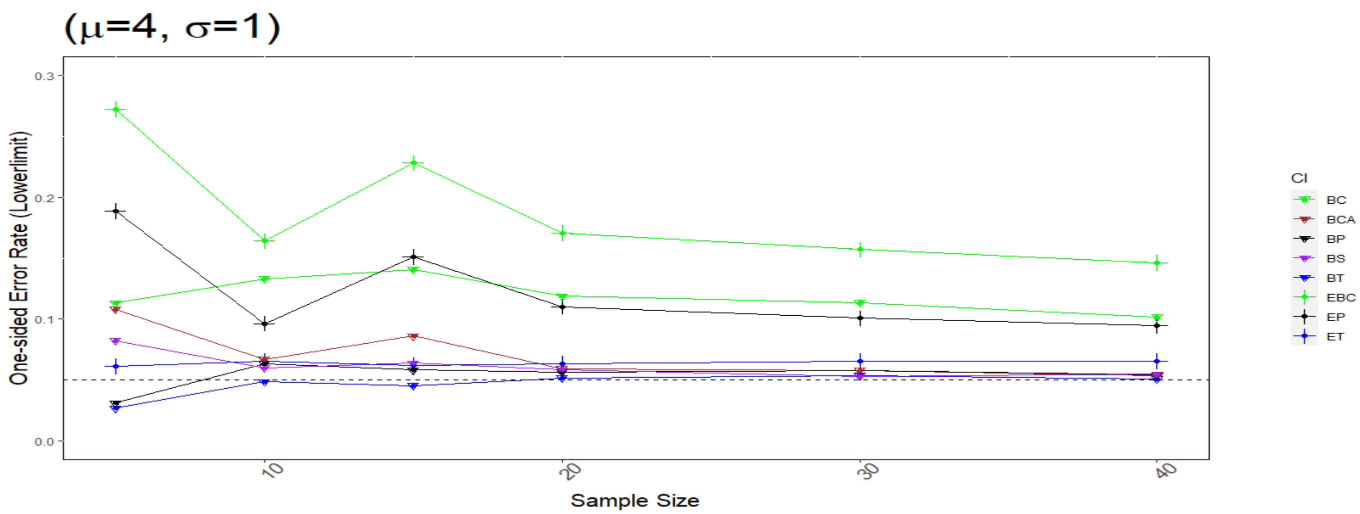


Figure: Sample Median - N90 - One-Sided Error Rates for 90% CI for the Normal Distribution





b. Exponential Distribution

The purpose of this sub section is to compare the performance of methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for the sample median statistic for data generated from an exponential distribution. For the exponential distribution one parameter specification type was generated. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables E1U99 and E1L99 on pages 267 and 268 below. For the sample median statistic, the one specification simulated was studied at three different α significance levels. However, in this section because of the volume of error rate results, only the error rate results at the $\alpha = 0.01$ significance level is displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the Exp(λ=0.10) at the $\alpha = 0.01$ significance level results can be viewed visually in figures E99, E95, and E90 on pages 269-271. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

Compared to the other methods studied, EP was relatively more accurate at the α = 0.01 and relatively less accurate at the α = 0.05 and α = 0.10 significance levels. For the upper and lower limit at the α = 0.01 significance level, EP achieved the error rate with the smallest percent error when compared to any other method at sample sizes 15.

However, at the α = 0.05 and α = 0.10 significance levels EP did not achieved an error rate with the smallest percent error at any sample size for the upper or lower limit. Further EBC did not achieve an error rate with the smallest percent error at any of the three significance levels. When comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors at many sample sizes for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels though, the EBSD(n) percentile algorithm had an error rate with a larger percent error at each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart.

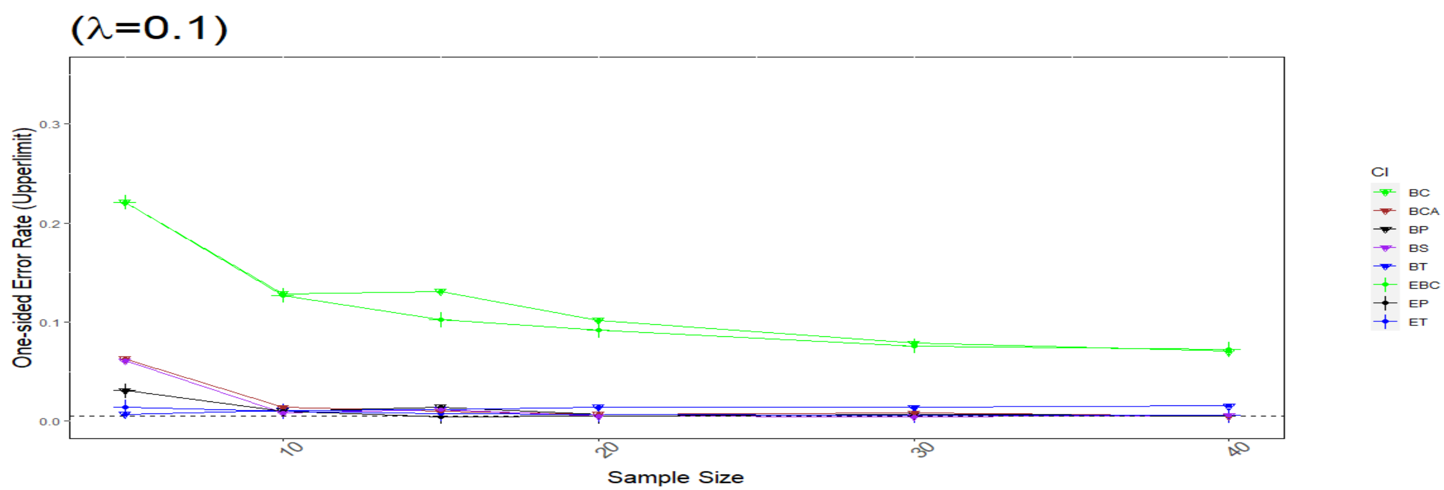
ET performed relatively accurately across significant level in comparison to BT. ET had an error rate with a smaller percent error when compared to the other methods studied for the majority of sample sizes for the upper and lower limit at each significance level. At α = 0.05 significance level for the upper limit, ET had the error rate with the smallest percent error at four sample sizes compared to any other method. Also at the α = 0.10 significance level for the upper limit ET had the error rate with the smallest percent at error at two sample sizes. Similar results were seen for the lower limit for ET as well.

Table: Sample Median - E1U99 Upper limit error rate ($\boldsymbol{\alpha}$ = 0.01), Exponential Distribution, $\boldsymbol{Exp}(\boldsymbol{\lambda}$=0.10), Bootstraps=5000

Sample size	5	10	15	20	30	40
BT	**0.007** **(40%)**	0.0112 (124%)	0.0118 (136%)	0.0144 (188%)	0.0139 (178%)	0.0154 (208%)
ET	0.014 (180%)	0.0101 (102%)	0.0077 (54%)	0.0067 (34%)	0.0058 (16%)	0.0057 (14%)
BC	0.2212 (4324%)	0.1286 (2472%)	0.1311 (2522%)	0.102 (1940%)	0.0787 (1474%)	0.0705 (1310%)
EBC	0.2212 (4324%)	0.1272 (2444%)	0.1026 (1952%)	0.0921 (1742%)	0.0761 (1422%)	0.0726 (1352%)
BP	0.0309 (518%)	0.0103 (106%)	0.0143 (186%)	0.0072 (44%)	0.0061 (22%)	**0.0054** **(8%)**
EP	0.0309 (518%)	0.0103 (106%)	**0.0046** **(8%)**	0.0053 (6%)	0.0068 (36%)	0.0057 (14%)
BS	0.0609 (1118%)	**0.0082** **(64%)**	0.0116 (132%)	**0.005** **(0%)**	**0.0044** **(12%)**	0.0058 (16%)
$\boldsymbol{BC_a}$	0.0631 (1162%)	0.014 (180%)	0.0099 (98%)	0.0069 (38%)	0.0081 (62%)	0.0055 (10%)

Table: Sample Median - E1L99 Lower limit error rate ($\boldsymbol{\alpha}$ = 0.01), Exponential Distribution, $\boldsymbol{Exp}(\boldsymbol{\lambda}$=0.10), Bootstraps=5000						
Sample size	5	10	15	20	30	40
BT	0.0016 (68%)	0.0029 (42%)	0.0031 (38%)	**0.0049** **(2%)**	0.0033 (34%)	**0.0049** **(2%)**
ET	**0.004** **(20%)**	**0.0045** **(10%)**	0.0042 (16%)	0.0066 (32%)	0.0046 (8%)	0.0047 (6%)
BC	0.0713 (1326%)	0.0261 (422%)	0.0383 (666%)	0.0205 (310%)	0.0182 (264%)	0.0162 (224%)
EBC	0.0713 (1326%)	0.0257 (414%)	0.0157 (214%)	0.0165 (230%)	0.0169 (238%)	0.0152 (204%)
BP	0.032 (540%)	0.0108 (116%)	0.0152 (204%)	0.0074 (48%)	**0.0047** **(6%)**	0.0077 (54%)
EP	0.032 (540%)	0.01 (100%)	**0.005** **(0%)**	0.0045 (10%)	0.0042 (16%)	0.0072 (44%)
BS	0.0483 (866%)	0.0055 (10%)	0.0103 (106%)	0.0036 (28%)	0.0044 (12%)	0.0055 (10%)
$\boldsymbol{BC_a}$	0.0525 (950%)	0.0137 (174%)	0.0114 (128%)	0.0081 (62%)	0.0066 (32%)	0.0079 (58%)

Figure: Sample Median - E99 - One-Sided Error Rates for 99% CI for the Exponential Distribution



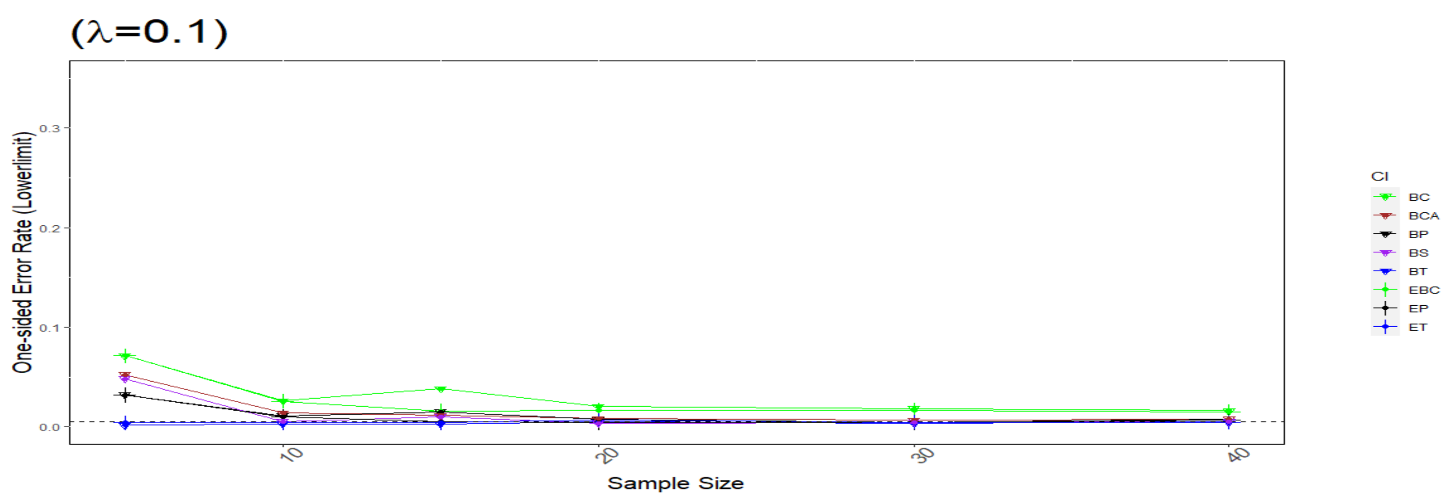
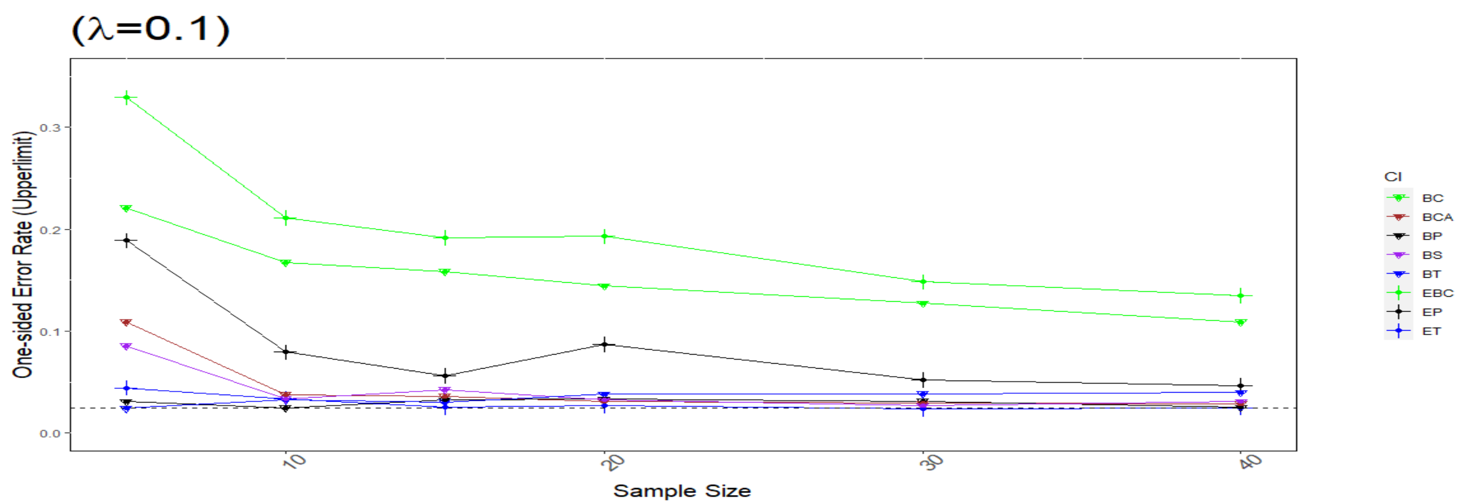


Figure: Sample Median - E95 - One-Sided Error Rates for 95% CI for the Exponential Distribution



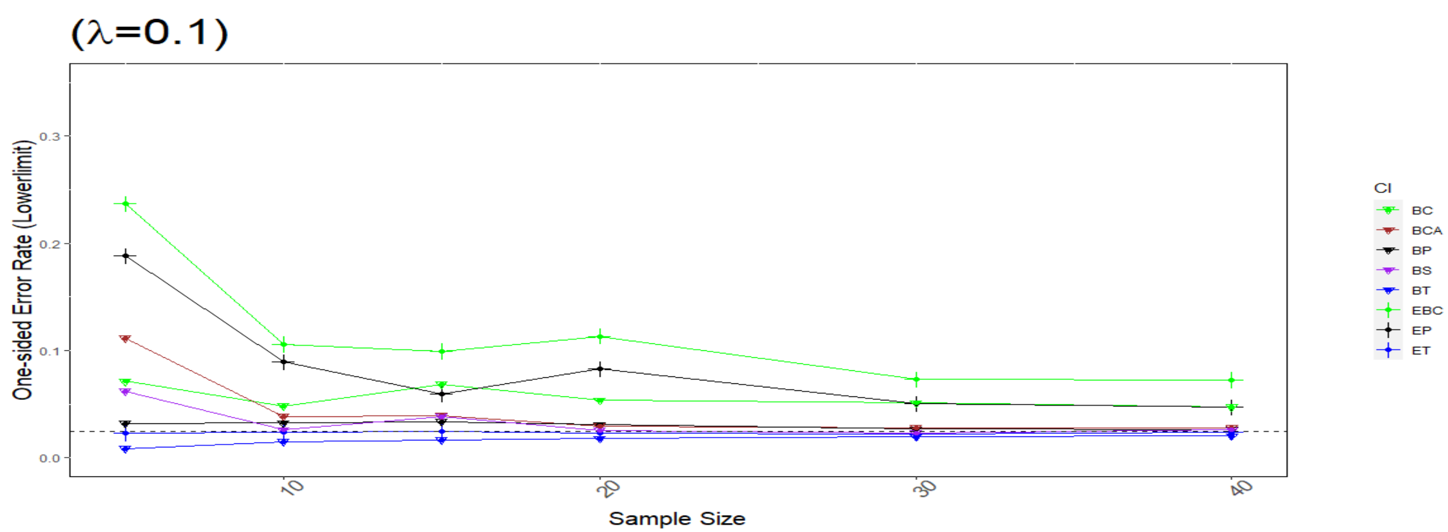
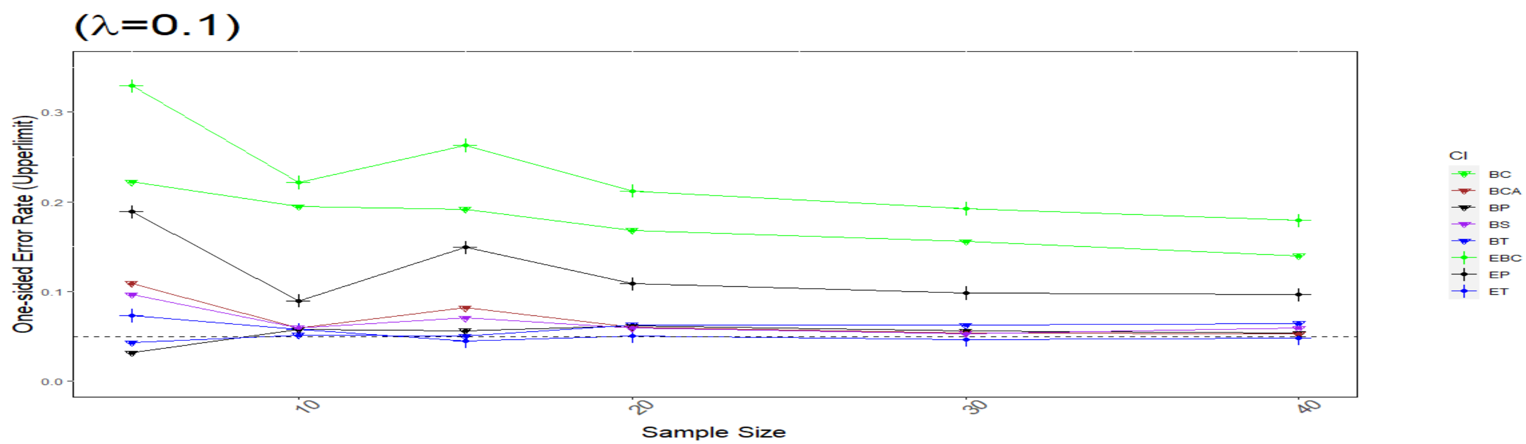
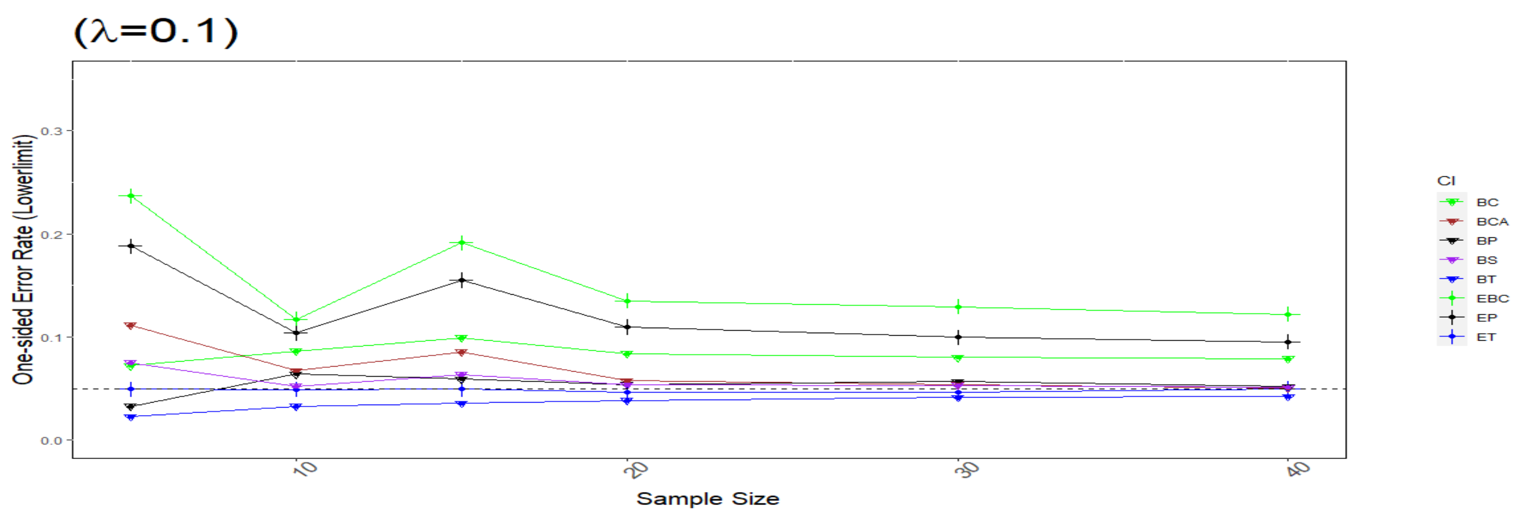


Figure: Sample Median - E90 - One-Sided Error Rates for 90% CI for the Exponential Distribution





c. Log-Normal Distribution

The purpose of this sub section is to compare the performance of methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for the sample median statistic for data from a log-normal distribution. For the log-normal distribution one parameter specification type was generated. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables LN1U99 and LN1L99 on pages 274 and 275 below. For the sample median statistic the one specification simulated was studied at three different α significance levels. However, in this section because of the volume of error rate results, only the error rate results at the $\alpha = 0.01$ significance level is displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the log-normal(μ =4, σ =0.2) at the $\alpha = 0.01$ significance level results can be viewed visually in figures LN99, LN95, and LN90 on pages 276-278. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

EP was again relatively more accurate at the α = 0.01 significance level compared to the other methods studied. For the upper limit, the EP did achieve the error rate with the smallest percent error when compared to any other method at sample sizes 15 and 40, it also achieved the error rate with the smallest percent error at sample size 20 for the lower limit.

However, at the α = 0.05 and α = 0.10 significance levels EP did not achieved an error rate with the smallest percent error at any sample size for the upper or lower limit. Further EBC did not achieve an error rate with the smallest percent error at any of the three significance levels. When comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors at many sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels though, the EBSD(n) percentile algorithms had an error rate with a larger percent error at each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart.

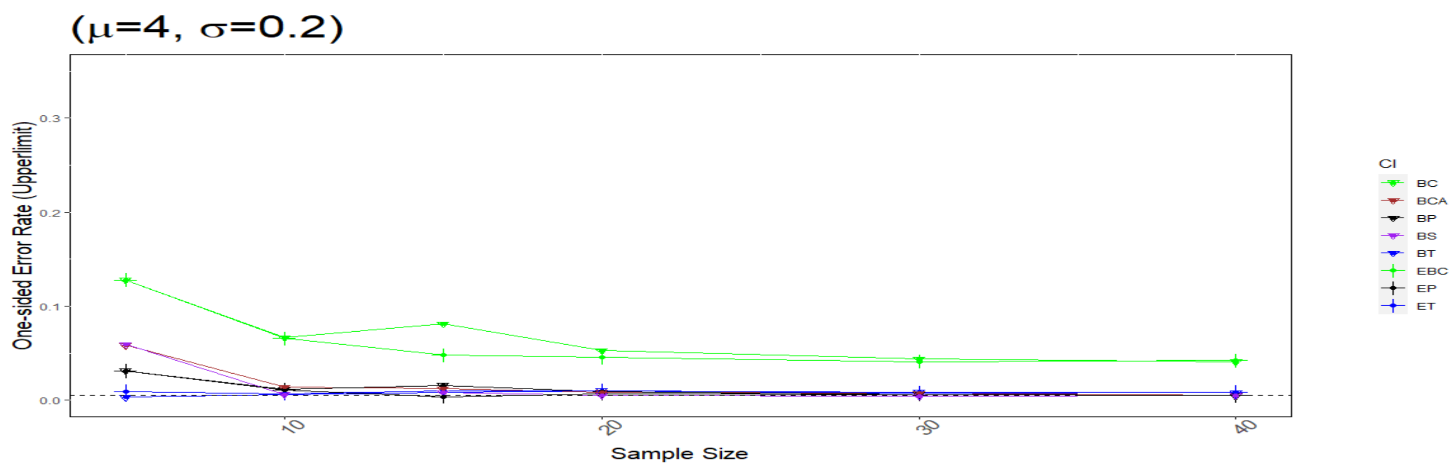
ET performed relatively less accurately in comparison to BT across significant level. ET had an error rate with a larger percent error for the most sample sizes for the upper and lower limit at each significance level. Further other than sample sizes 5 and 15, ET did not have an error rate with the smallest percent error when compared to all other methods for the other sample sizes.

Table: Sample Median - LN1U99 Upper limit error rate (α = 0.01), Log-Normal Distribution, log-normal(μ=4, σ=0.2), Bootstraps=5000						
Sample size	5	10	15	20	30	40
BT	**0.0032** **(36%)**	**0.006** **(20%)**	0.0082 (64%)	0.0102 (104%)	0.0086 (72%)	0.0081 (62%)
ET	0.0091 (82%)	0.0071 (42%)	0.0098 (96%)	0.0104 (108%)	0.0075 (50%)	0.0086 (72%)
BC	0.128 (2460%)	0.0667 (1234%)	0.0812 (1524%)	0.0529 (958%)	0.0439 (778%)	0.0408 (716%)
EBC	0.128 (2460%)	0.0657 (1214%)	0.0478 (856%)	0.0458 (816%)	0.0411 (722%)	0.0421 (742%)
BP	0.0311 (522%)	0.0114 (128%)	0.0159 (218%)	0.0091 (82%)	0.0062 (24%)	0.0048 (4%)
EP	0.0311 (522%)	0.0112 (124%)	**0.0039** **(22%)**	0.0072 (44%)	0.0062 (24%)	**0.005** **(0%)**
BS	0.0594 (1088%)	0.0061 (22%)	0.0082 (64%)	**0.0055** **(10%)**	**0.0041** **(18%)**	0.0054 (8%)
BC_a	0.0584 (1068%)	0.0138 (176%)	0.0123 (146%)	0.0085 (70%)	0.0074 (48%)	0.0055 (10%)

Table: Sample Median - LN1L99 Lower limit error rate (α = 0.01), Log-Normal Distribution, log-normal(μ=4, σ=0.2), Bootstraps=5000

Sample size	5	10	15	20	30	40
BT	0.0014 (72%)	0.0035 (30%)	**0.0044** **(12%)**	0.0067 (34%)	**0.0049** **(2%)**	0.0069 (38%)
ET	**0.0048** **(4%)**	0.0061 (22%)	0.0065 (30%)	0.009 (80%)	0.006 (20%)	0.0092 (84%)
BC	0.0929 (1758%)	0.0435 (770%)	0.0514 (928%)	0.038 (660%)	0.0293 (486%)	0.027 (440%)
EBC	0.0929 (1758%)	0.0423 (746%)	0.0257 (414%)	0.0315 (530%)	0.0274 (448%)	0.0268 (436%)
BP	0.0328 (556%)	0.0121 (142%)	0.014 (180%)	0.0084 (68%)	0.0064 (28%)	**0.0054** **(8%)**
EP	0.0328 (556%)	0.0118 (136%)	0.0034 (32%)	**0.0064** **(28%)**	0.0062 (24%)	0.0056 (12%)
BS	0.054 (980%)	**0.0048** **(4%)**	0.0085 (70%)	**0.0036** **(28%)**	0.0041 (18%)	**0.0046** **(8%)**
BC_a	0.0566 (1032%)	0.0156 (212%)	0.0106 (112%)	0.0081 (62%)	0.0067 (34%)	0.0059 (18%)

Figure: Sample Median - LN99 - One-Sided Error Rates for 99% CI for the Log-Normal Distribution



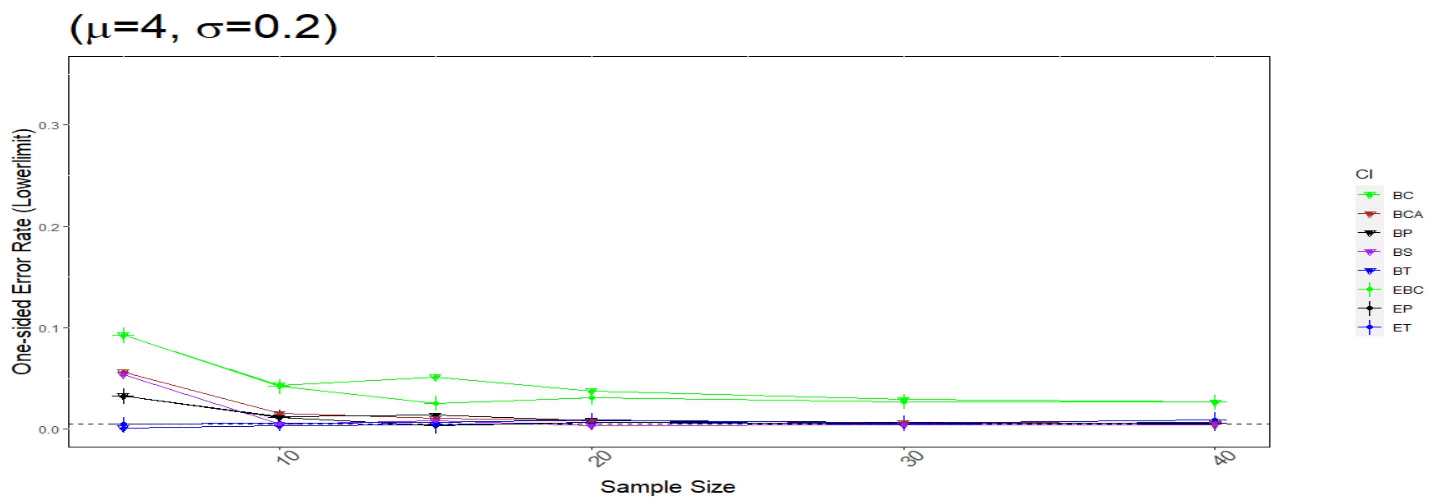
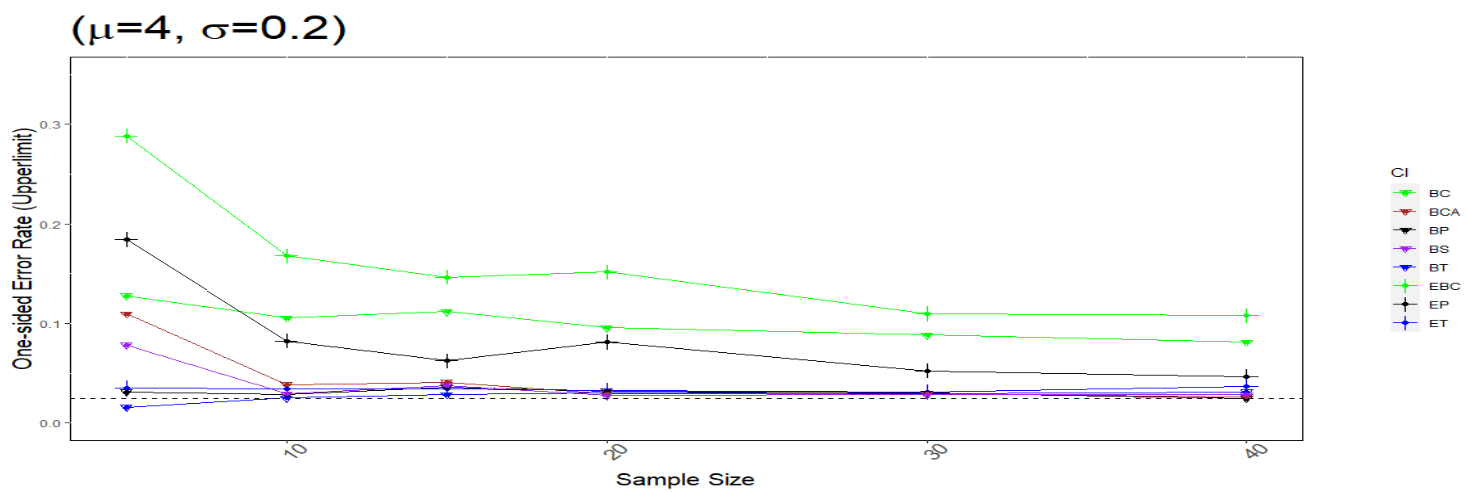


Figure: Sample Median - LN95 - One-Sided Error Rates for 95% CI for the Log-Normal Distribution



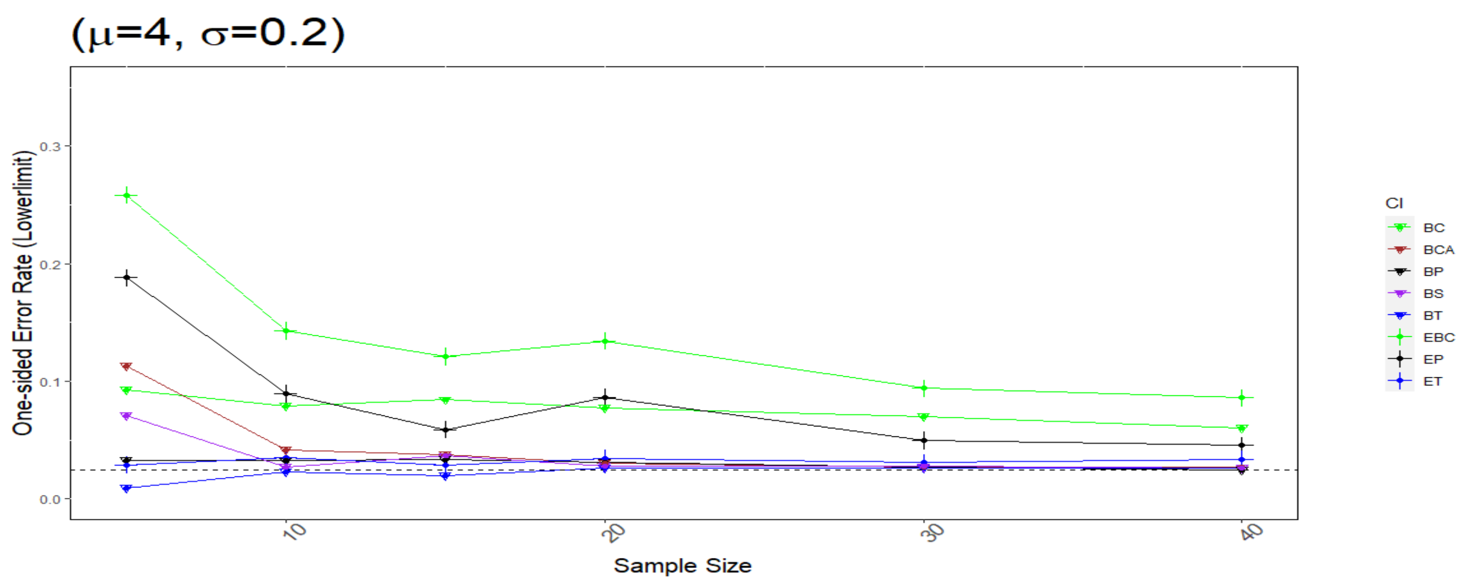
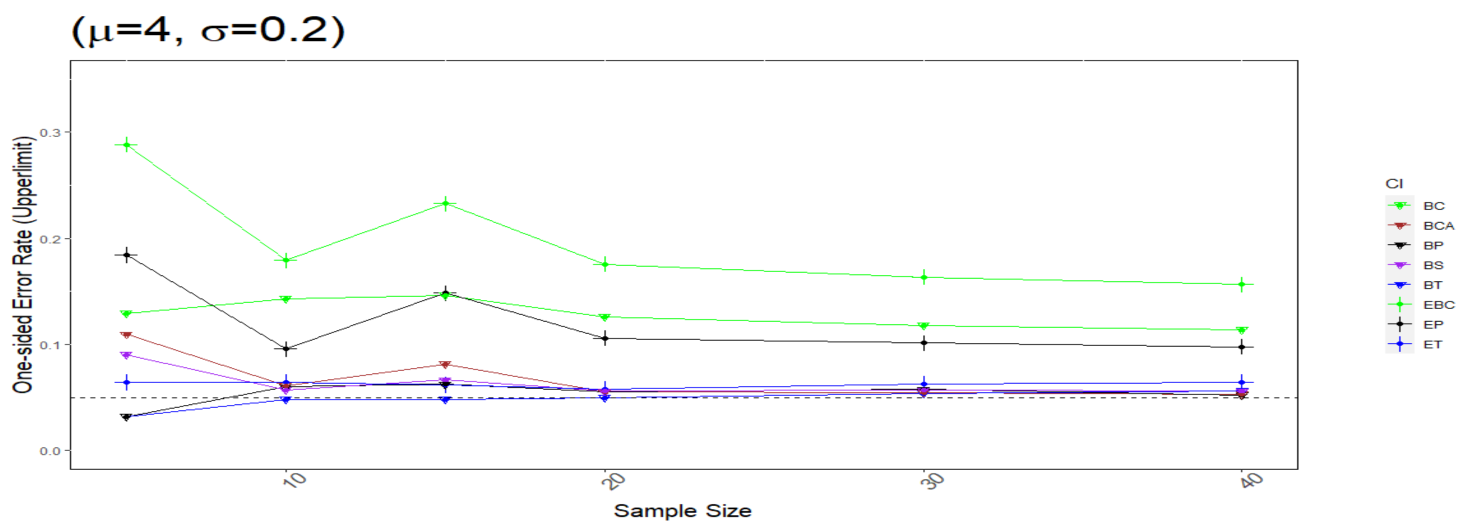
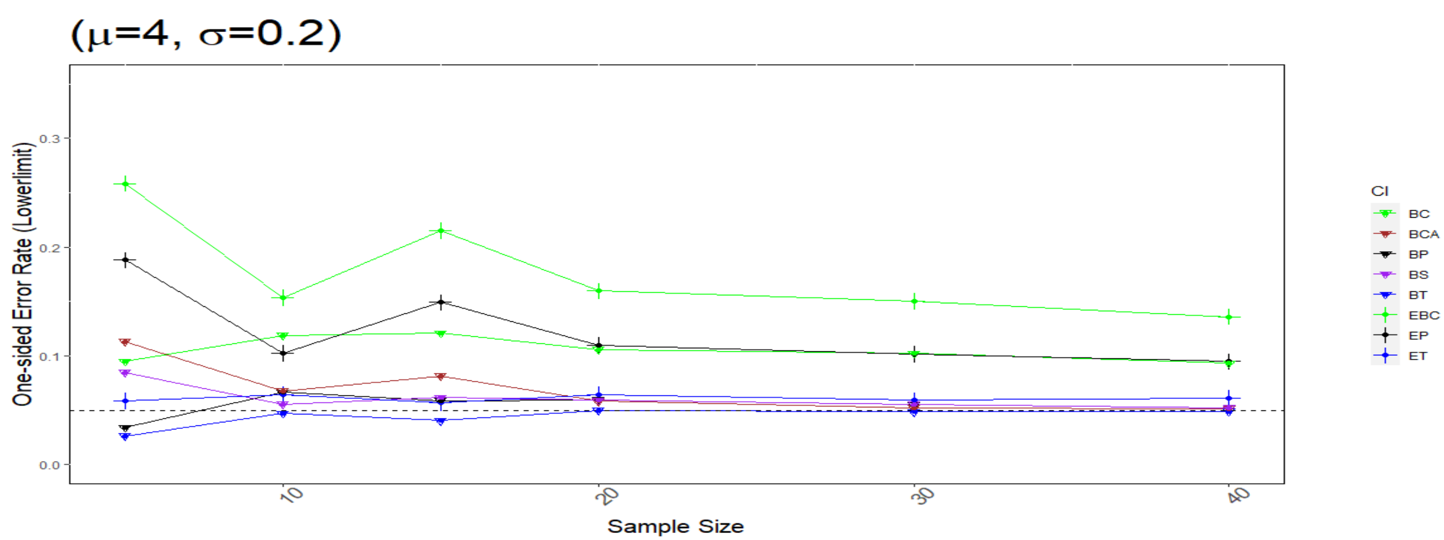


Figure: Sample Median - LN90 - One-Sided Error Rates for 90% CI for the Log-Normal Distribution





d. Mixture Distribution

The purpose of this sub section is to compare the performance of methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for the sample median statistic for data from a mixture of two normal distribution. For the mixture of two normal distributions one parameter specification type was generated. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables MN1U99 and MN1L99 on pages 281 and 282 below. For the sample median statistic the one specification simulated was studied at three different α significance levels. However, in this section because of the volume of error rate results, only the error rate results at the $\alpha = 0.01$ significance level is displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the mixture of two normal distributions (μ_1= 4, σ_1= 1, p_1= 0.5, μ_2= 8, σ_2 =1, p_2= 0.5) at the $\alpha = 0.01$ significance level, results can be viewed visually in figures MN99, MN95, and MN90 on pages 283-285. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

EP was again relatively more accurate at the α = 0.01 significance level in comparison to the other methods studied. For both the upper and lower limit, the EP method achieved the error rate with the smallest percent error when compared to any other method at four of six sample sizes considered.

However, at the α = 0.05 and α = 0.10 significance levels EP did not achieved an error rate with the smallest percent error at any sample size for the upper or lower limit. Further EBC did not achieve an error rate with the smallest percent error at any of the three significance levels. When comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors at many sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels though, the EBSD(n) percentile algorithms had an error rate with a larger percent error at each sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart.

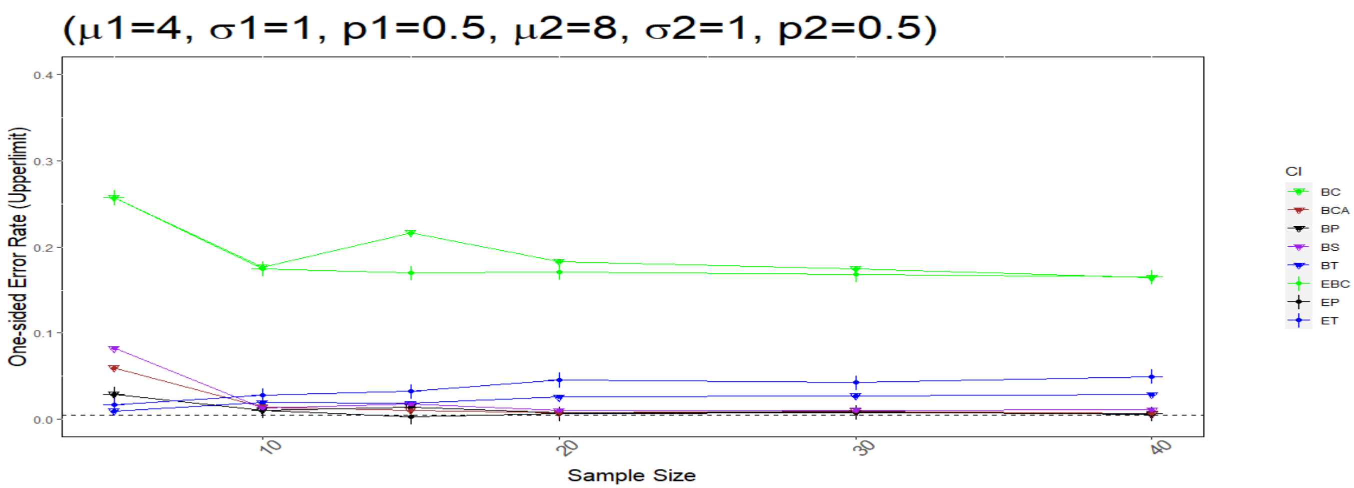
ET performed relatively less accurately in comparison to BT across significant level. For most sample sizes studied, at each significance level, ET had an error rate with a larger percent error for the upper and lower limit. Further ET did not achieve the error rate with the smallest percent error at any sample size for any significance level when compared to all other methods studied.

Table: Sample Median - MN1U99 Upper limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.50*$N_1(\mu_1$ =4, σ_1=1) + 0.50*$N_2(\mu_2$ =8, σ_2=1), Bootstraps=5000						
Sample size	5	10	15	20	30	40
BT	**0.0098** **(96%)**	0.0197 (294%)	0.0188 (276%)	0.0259 (418%)	0.0272 (444%)	0.0289 (478%)
ET	0.0175 (250%)	0.0278 (456%)	0.0325 (550%)	0.0457 (814%)	0.0426 (752%)	0.0498 (896%)
BC	0.257 (5040%)	0.1765 (3430%)	0.2167 (4234%)	0.1832 (3564%)	0.1745 (3390%)	0.1647 (3194%)
EBC	0.257 (5040%)	0.1746 (3392%)	0.1696 (3292%)	0.1709 (3318%)	0.1682 (3264%)	0.1649 (3198%)
BP	0.0293 (486%)	0.0105 (110%)	0.0142 (184%)	0.0082 (64%)	0.0084 (68%)	**0.0061** **(22%)**
EP	0.0293 (486%)	**0.0104** **(108%)**	**0.0031** **(38%)**	**0.0065** **(30%)**	**0.0082** **(64%)**	0.0065 (30%)
BS	0.0824 (1548%)	0.0134 (168%)	0.0178 (256%)	0.0102 (104%)	0.0106 (112%)	0.0114 (128%)
BC_a	0.0601 (1102%)	0.0142 (184%)	0.0108 (116%)	0.0082 (64%)	0.0097 (94%)	0.0073 (46%)

Table: Sample Median - MN1L99 Lower limit error rate (α = 0.01), Mixture of two Normal Distributions, 0.50*$N_1(\mu_1$ =4, σ_1=1) + 0.50*$N_2(\mu_2$ =8, σ_2=1), Bootstraps=5000

Sample size	5	10	15	20	30	40
BT	**0.0114** **(128%)**	0.0198 (296%)	0.0215 (330%)	0.0294 (488%)	0.0269 (438%)	0.0275 (450%)
ET	0.0208 (316%)	0.0285 (470%)	0.0369 (638%)	0.0494 (888%)	0.0403 (706%)	0.0474 (848%)
BC	0.2573 (5046%)	0.179 (3480%)	0.2169 (4238%)	0.1804 (3508%)	0.1701 (3302%)	0.1631 (3162%)
EBC	0.2573 (5046%)	0.1773 (3446%)	0.1735 (3370%)	0.168 (3260%)	0.1646 (3192%)	0.1638 (3176%)
BP	0.0331 (562%)	0.0098 (96%)	0.0168 (236%)	0.0085 (70%)	**0.006** **(20%)**	**0.0055** **(10%)**
EP	0.0331 (562%)	**0.0097** **(94%)**	**0.005** **(0%)**	**0.0064** **(28%)**	0.0062 (24%)	**0.0055** **(10%)**
BS	0.0841 (1582%)	0.0123 (146%)	0.0187 (274%)	0.0122 (144%)	0.0099 (98%)	0.011 (120%)
BC_a	0.0629 (1158%)	0.0144 (188%)	0.0129 (158%)	0.0086 (72%)	0.0072 (44%)	0.0062 (24%)

Figure: Sample Median - MN99 - One-Sided Error Rates for 99% CI for the Mixture of two Normal Distributions



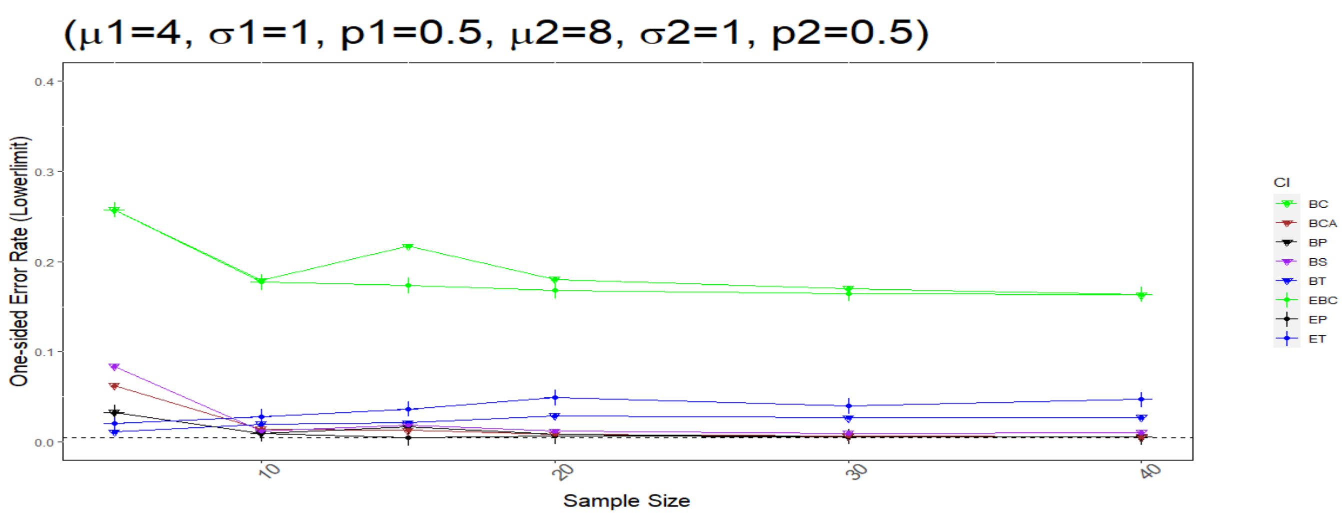
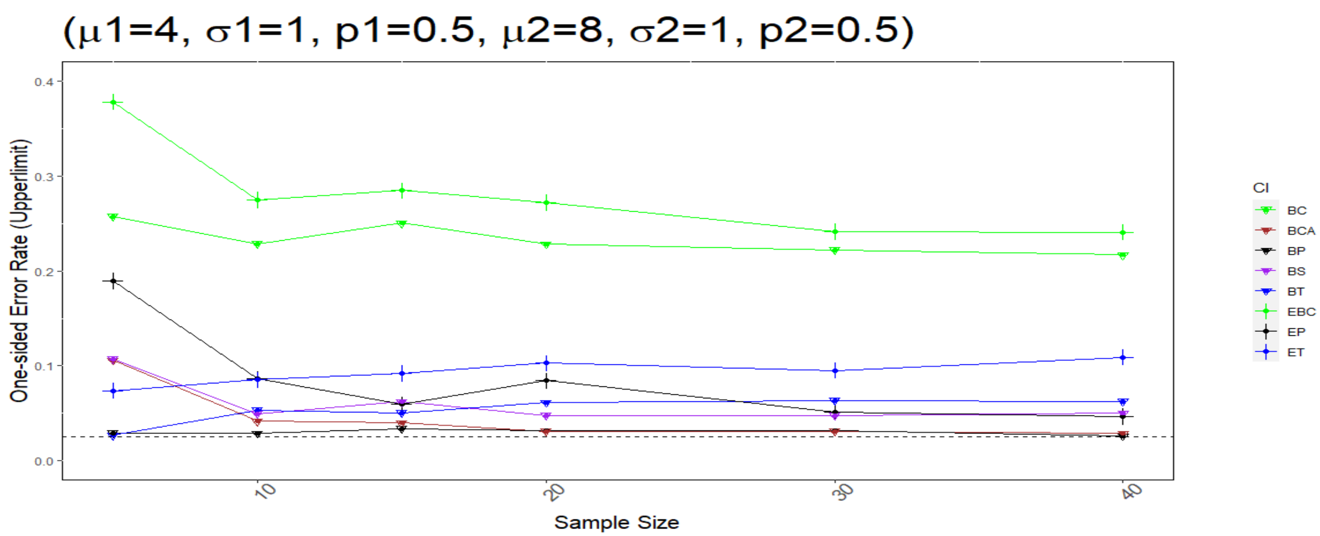


Figure: Sample Median - MN95 - One-Sided Error Rates for 95% CI for the Mixture of two Normal Distributions



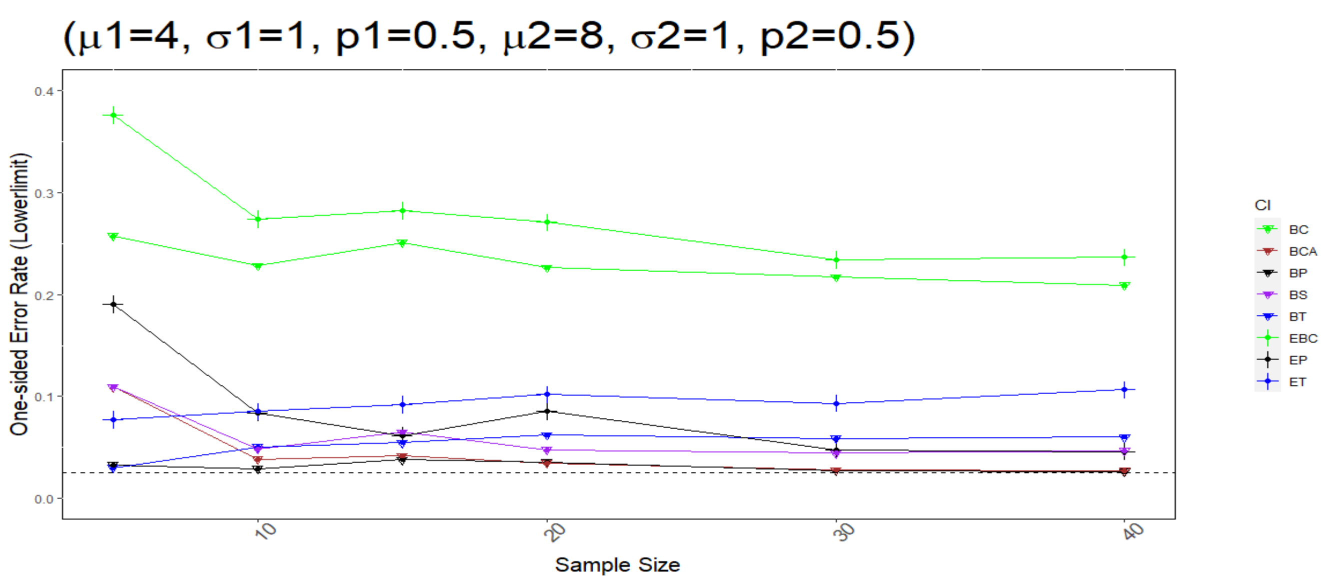
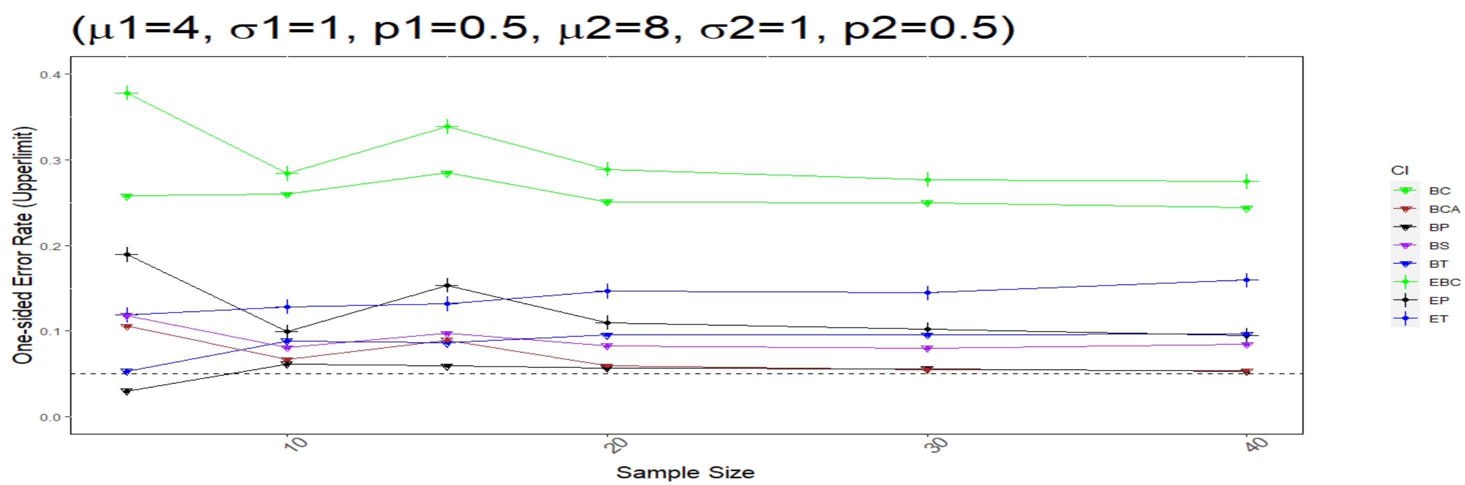
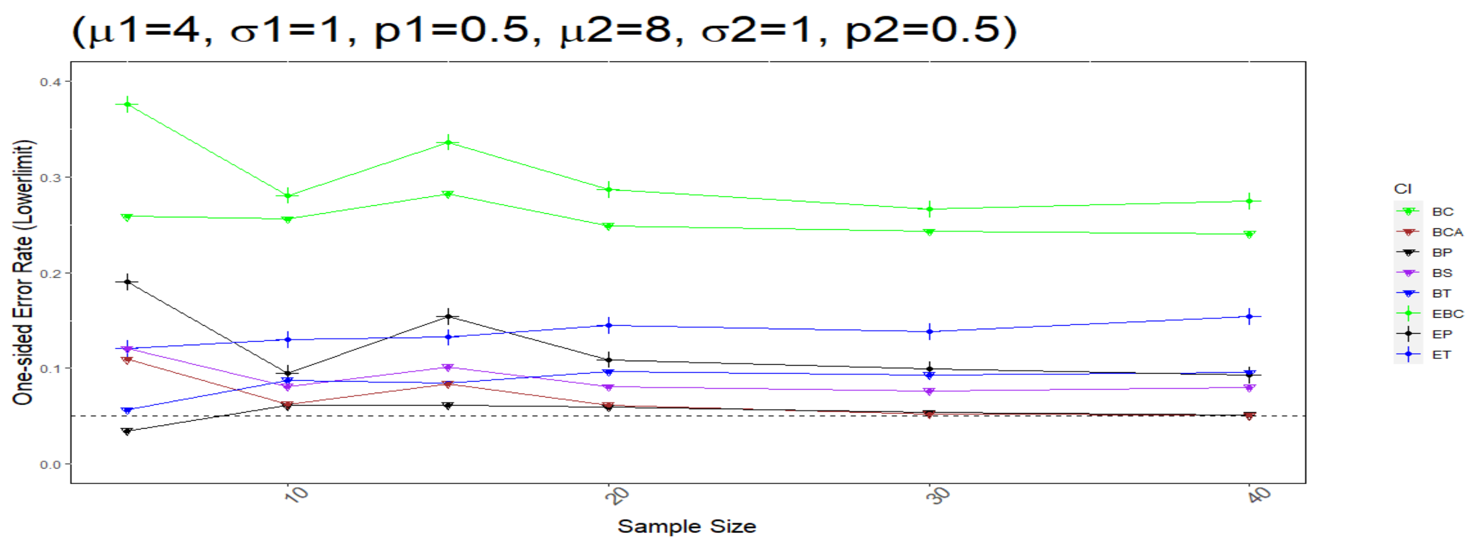


Figure: Sample Median - MN90 - One-Sided Error Rates for 90% CI for the Mixture of two Normal Distributions





e. Cauchy Distribution

The purpose of this sub section is to compare the performance of methods that use the EBSD(n) method to the performance of Monte Carlo Bootstrap methods for the sample median statistic for data from a Cauchy distribution. For the Cauchy distribution one parameter specification type was generated. These confidence interval method error rates and their corresponding percent errors can be viewed and compared to one another in each of tables C1U99 and C1L99 on pages 288-289 below. For the sample median statistic the one specification simulated was studied at three different α significance levels. However, in this section because of the volume of error rate results, only the error rate results at the $\alpha = 0.01$ significance level is displayed in tables. Detailed numerical results for simulations not included in these tables can be viewed in Appendix tables.

Although the tables only report results for the Cauchy (x_0=0, γ=1) at the $\alpha = 0.01$ significance level results can be viewed visually in figures C99, C95, and C90 on pages 290-292. In these figures the dashed horizontal line represents the target nominal one-sided error rate based on the confidence interval α significance level. Each colored line represent a different confidence interval method with error rates plotted at sample sizes 5, 10, 15, 20, 30 and 40. Plot points marked with cross symbols represent methods that use EBSD(n). Plot points marked with triangles represent methods that use the Monte Carlo Bootstrap.

EP again performed relatively accurately compared to the other methods studied at the α = 0.01 significance level. For both the upper and lower limit, the EP method achieved the error rate with the smallest percent error at sample size 15 when compared to any other method. Additionally, for the lower limit, EP achieved the error rate with the smallest percent error at sample size 20.

However, at the α = 0.05 and α = 0.10 significance levels EP did not achieved an error rate with the smallest percent error at any sample size for the upper or lower limit. Further, EBC did not achieve an error rate with the smallest percent error at any of the three significance levels. When comparing EBC and EP to BC and BP respectively, both methods had error rates with smaller percent errors at many sample size for the upper and lower limit compared to their Monte Carlo Bootstrap counterpart at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels though, the percentile methods applied on EBSD(n) had error rates with larger percent errors at each sample size for the upper and lower limit when compared to their Monte Carlo Bootstrap counterpart.

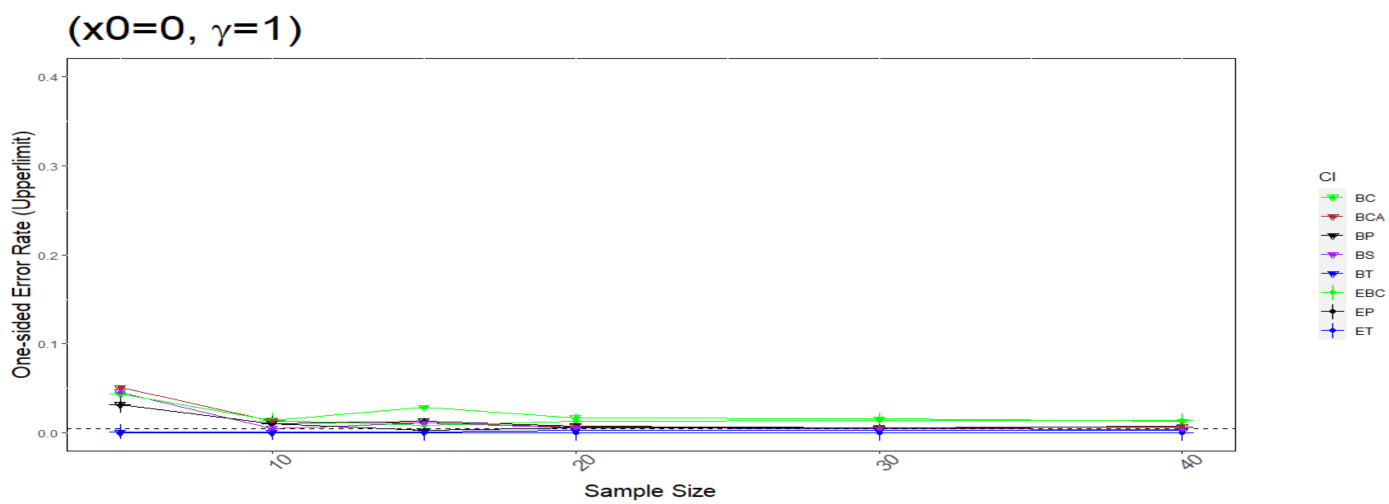
Additionally, ET performed relatively less accurately in comparison to BT across significant level. ET had an error rate with a larger percent error for the majority of sample sizes for both the upper and lower limit at each significance level. With this being the case, at each significance level ET did not achieve the error rate with the smallest percent error at any sample size greater than 5 when compared to the other methods studied.

Table: Sample Median - C1U99 Upper limit sample median error rate (α = 0.01), Cauchy Distribution, Cauchy(x_0 =0, γ=1), Bootstraps=5000						
Sample size	5	10	15	20	30	40
BT	4e-04 (92%)	5e-04 (90%)	0.0014 (72%)	0.0032 (36%)	0.0029 (42%)	0.003 (40%)
ET	**0.0017** **(66%)**	0.001 (80%)	3e-04 (94%)	3e-04 (94%)	1e-04 (98%)	0 (100%)
BC	0.0436 (772%)	0.0146 (192%)	0.029 (480%)	0.0174 (248%)	0.0157 (214%)	0.0132 (164%)
EBC	0.0436 (772%)	0.0144 (188%)	0.0091 (82%)	0.0132 (164%)	0.0143 (186%)	0.0139 (178%)
BP	0.0318 (536%)	0.0106 (112%)	0.0135 (170%)	0.008 (60%)	0.0055 (10%)	0.0069 (38%)
EP	0.0318 (536%)	0.0104 (108%)	**0.0027** **(46%)**	0.0058 (16%)	0.0055 (10%)	0.0069 (38%)
BS	0.0458 (816%)	**0.0052** **(4%)**	0.0119 (138%)	**0.0055** **(10%)**	**0.0047** **(6%)**	**0.0043** **(14%)**
BC_a	0.0509 (918%)	0.0134 (168%)	0.0113 (126%)	0.0073 (46%)	0.0059 (18%)	0.0075 (50%)

Table: Sample Median - C1L99 Lower Limit Error rate (α = 0.01), Cauchy distribution, Cauchy(x_0 =0, γ=1), Bootstraps=5000

Sample size	5	10	15	20	30	40
BT	4e-04 (92%)	8e-04 (84%)	0.0017 (66%)	0.0018 (64%)	0.0027 (46%)	0.004 (20%)
ET	**0.002** **(60%)**	9e-04 (82%)	1e-04 (98%)	3e-04 (94%)	2e-04 (96%)	0 (100%)
BC	0.0481 (862%)	0.0175 (250%)	0.0281 (462%)	0.0172 (244%)	0.0152 (204%)	0.0152 (204%)
EBC	0.0481 (862%)	0.0173 (246%)	0.008 (60%)	0.0138 (176%)	0.0142 (184%)	0.0161 (222%)
BP	0.0309 (518%)	0.0118 (136%)	0.0144 (188%)	0.0073 (46%)	**0.0047** **(6%)**	**0.0055** **(10%)**
EP	0.0309 (518%)	0.0113 (126%)	**0.0041** **(18%)**	**0.0051** **(2%)**	0.0041 (18%)	0.0062 (24%)
BS	0.0484 (868%)	**0.0058** **(16%)**	0.009 (80%)	0.0042 (16%)	0.0035 (30%)	**0.0055** **(10%)**
BC_a	0.0516 (932%)	0.0155 (210%)	0.0122 (144%)	0.0073 (46%)	0.0058 (16%)	0.0063 (26%)

Figure: Sample Median - C99 - One-Sided Error Rates for 99% CI for the Cauchy Distribution



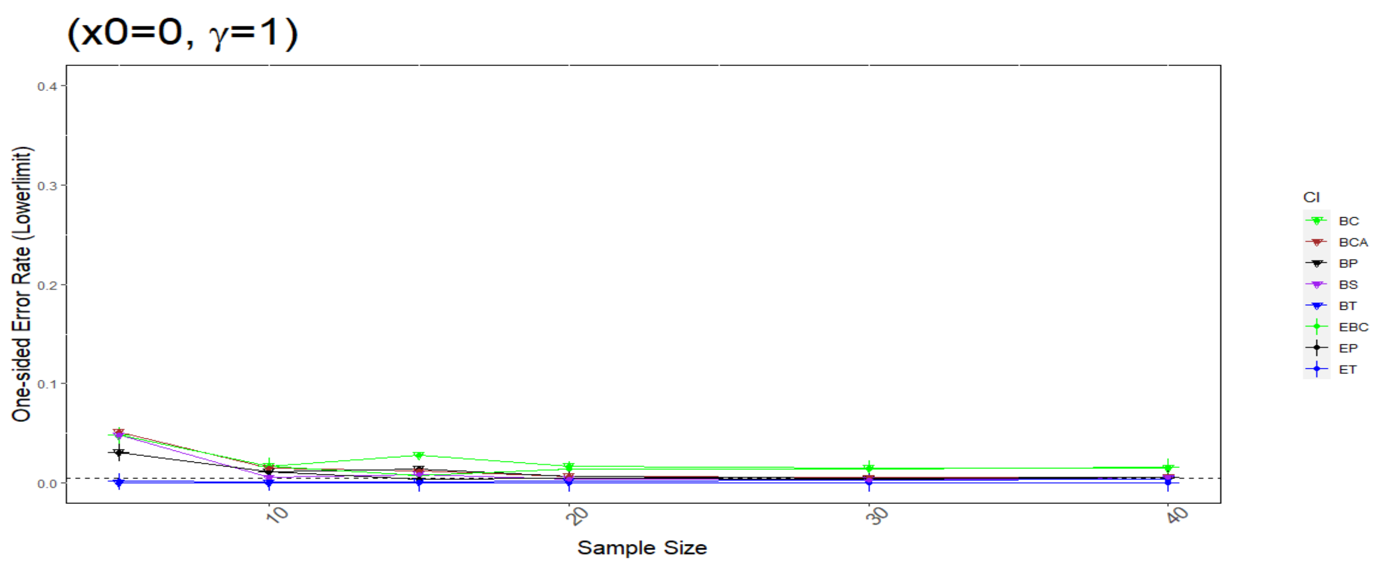
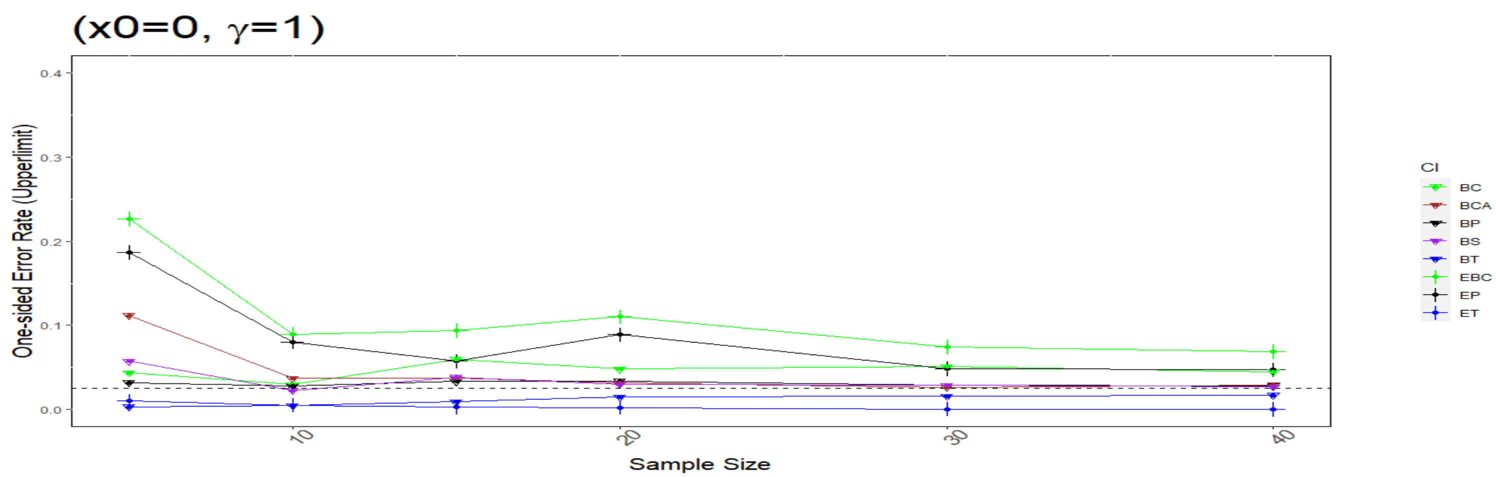


Figure: Sample Median - C95 - One-Sided Error Rates for 95% CI for the Cauchy Distribution



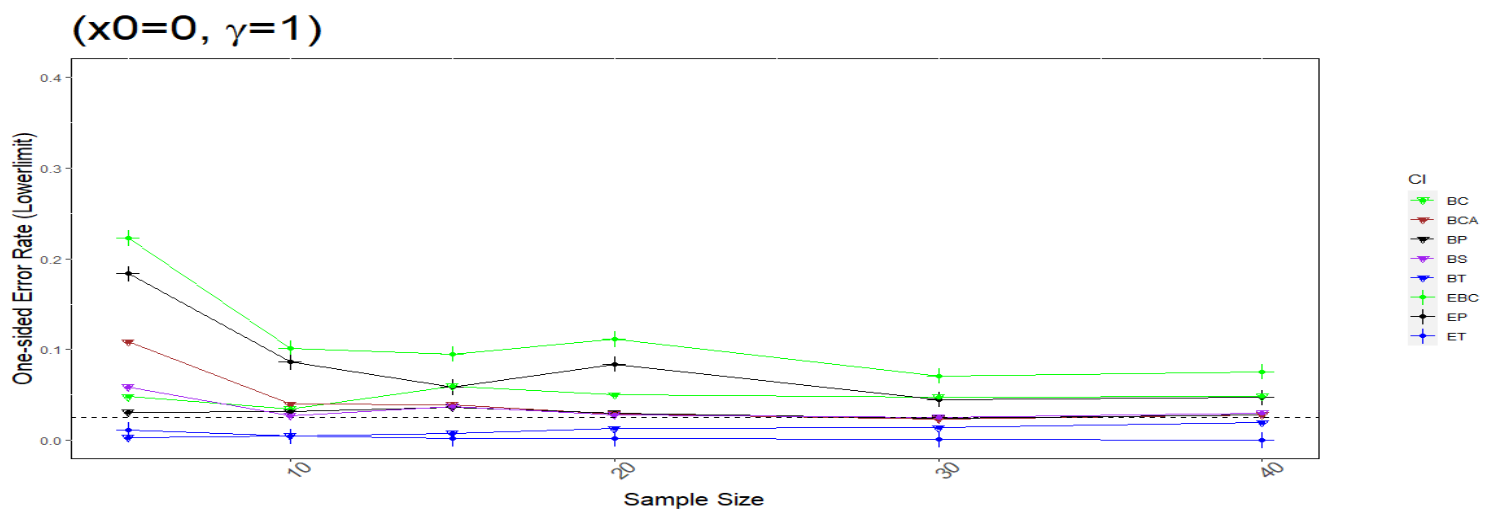
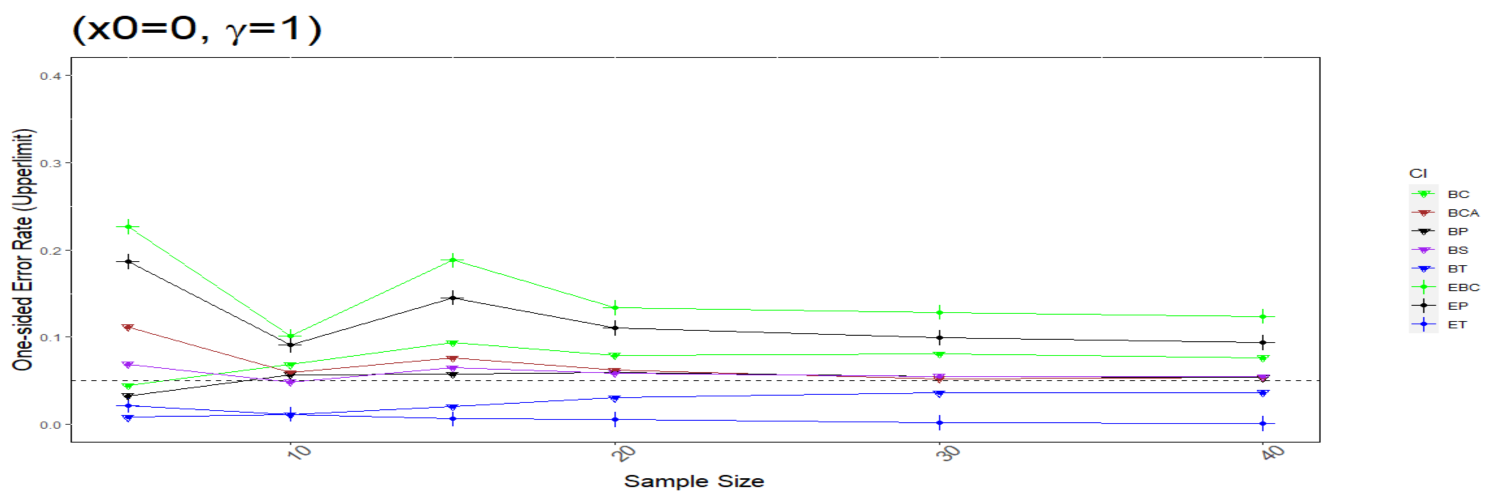
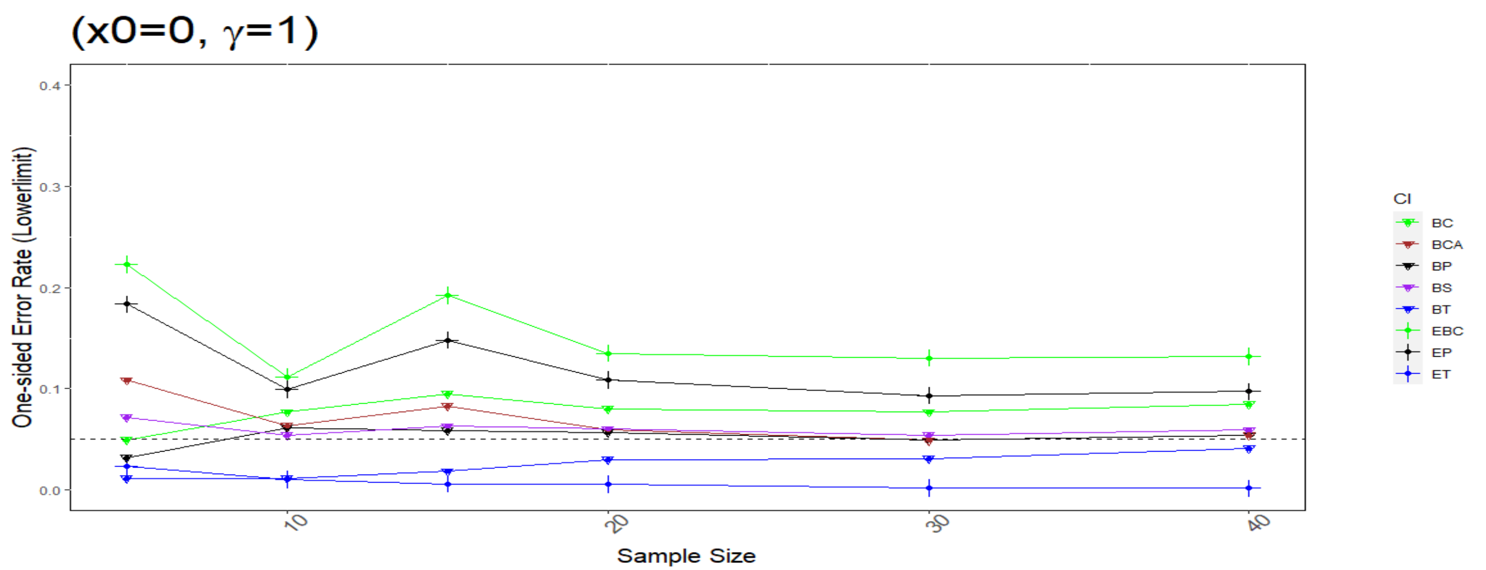


Figure: Sample Median - C90 - One-Sided Error Rates for 90% CI for the Cauchy Distribution





Median Results for Methods using EBSD(n)

Regardless of the parameter specification and distribution drawn from the same pattern was seen for the percentile methods applied on EBSD(n). EP and EBC performed relatively more accurately at the α = 0.01 significance level, and relatively less accurately as significance level was modified.

ET however did have varied results depending on the distribution studied. For the normal, log-normal and Cauchy distributions, ET performed relatively more accurately in multiple instances at small sample sizes (n=5, 10, 15), than BT and in many instances achieved error rates with smaller percent errors across sample size. For the mixture of two normal distributions ET performed relatively less accurately compared to BT for small, medium and large sample sizes. For the exponential distribution which was the distribution with the greatest degree of skew studied for the median statistic, ET performed relatively more accurately than BT for both the α = 0.05 and α = 0.10 significance levels. These results indicate further investigation into ET's relative accuracy for the median statistic could be warranted. ET may be a preferable method for the median statistic if the data being analyzed is skewed.

4.6 Varied Monte Carlo Bootstrap Iterations Levels

The purpose of this section is to compare EBSD(n) error rate results to error rate results from the Monte Carlo Bootstrap at multiple bootstrap resample levels. One advantage of EBSD(n) methods relative to bootstrap methods is the consistency of the algorithm. With methods applied on EBSD(n) there will always be $4n^2 + 1$ bootstrap samples. With methods applied on the Monte Carlo Bootstrap the number of resamples specified will depend on the preference of the user. At smaller numbers of Bootstrap resamples the variance of the error rate will be larger. If there is greater variance in error rate results, this could lead Monte Carlo Bootstrap methods to yield error rates that vary from the theoretical nominal error rate more than EBSD(n) methods.

Sample Mean

For the first example below, results are viewed for data generated from a normal distribution with parameters $N(\mu$=4, σ=1) at the α = 0.01 significance level for the upper limit. The tables below list the error rates for Monte Carlo Bootstrap resamples of 200, 500, 1000 and 10,000 for the mean statistic and compares these results to what is found using EBSD(n). The error rate closest to the true nominal error rate among the methods considered is bolded for each sample size.

When 10,000 Bootstrap resamples were used, the BT method had an error rate with corresponding percent error that was less than or equal to ET's at each sample size. Instead, when 200 resamples were specified, ET had an error rate with corresponding percent error that was less than or equal to BT's at four of the five sample sizes considered. Therefore, increasing the number of Bootstrap resamples did impact the resulting comparison of relative accuracy of ET and BT. When a smaller number of Bootstrap resamples were used, ET was found to be more accurate than BT. When a larger number of Bootstrap resamples were used BT was found to be more accurate than ET.

BC and BP at the α = 0.01 significance level, performed less accurately compared to EBC and EP at each sample size even when 10,000 Bootstrap resamples were specified. Further, at each sample size BC and BP had error rates with corresponding percent errors that were larger when 10,000 Bootstrap resamples were specified relative to when 200 Bootstrap resamples were specified. Therefore, for these two methods decreasing Bootstrap iteration level did not impact each methods comparison to their EBSD(n) counterpart.

EBC_a and ES were not included in this section, as these methods were generally found to be unreliable. However, the error rate variation in BC_a and BS is measured against the error rate accuracy of E-skew when the number of Bootstrap resamples used was varied. When 10,000 Bootstrap resamples were used BC_a had an error rate with a corresponding percent error that was smaller at each sample size compared to when 200 Bootstrap resamples were used. However even when 10,000 Bootstrap resamples were used, BC_a still had error rates with corresponding percent errors that were larger compared to E-skew at each sample size studied. Finally, the BS error rate was not found to be impacted by the number of Bootstrap resamples used. The BS error rate using 10,000 Bootstrap resamples had a corresponding percent error that was smaller than the BS error rate using 200 Bootstrap at sample sizes 5 and 30. Conversely, the BS error rate using 200 Bootstrap resamples had error rates with corresponding percent errors at 10, 15 and 40. At each Bootstrap resample level, BS had the error rate with the smallest percent error for the majority of sample sizes among Monte Carlo Bootstrap methods.

Table: Mean - N1U99 Advantage of EBSD($\boldsymbol{n}$) when comparing across Bootstrap Iteration Levels, $\boldsymbol{\alpha}$ = 0.01 significance level, upper limit

	Sample Size				
	N=5	N=10	N=15	N=30	N=40
E-skew	0.0111	0.0089	**0.005**	0.0052	0.0056
ET	0.0067	**0.0072**	0.0053	0.0053	0.006
EP	0.0318	0.0096	0.0029	0.0022	**0.005**
EBC	0.0318	0.0101	0.0027	0.0028	0.0052

#Boots	**10000**					**1000**					**500**					**200**				
Sample Size	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**
BT	0.0065	**0.0072**	0.0053	0.0053	0.0059	0.0069	0.0075	**0.0051**	0.0053	0.0064	0.0066	0.008	0.0053	**0.0051**	0.0064	0.007	0.0079	0.0057	**0.0052**	0.006
BC	0.0506	0.0224	0.0123	0.008	0.0082	0.0508	0.0216	0.0117	0.0077	0.0081	0.0499	0.0208	0.0113	0.0093	0.0079	0.0494	0.0193	0.0108	0.0082	0.0069
BP	0.0521	0.0227	0.0119	0.008	0.0081	0.0526	0.0221	0.0113	0.0085	0.0075	0.0521	0.0209	0.0112	0.008	0.0082	0.0504	0.0206	0.0113	0.008	0.0074
$\boldsymbol{BC_a}$	0.0565	0.0235	0.0124	0.0086	0.0081	0.0576	0.0235	0.0119	0.0088	0.0073	0.0561	0.0234	0.0129	0.0084	0.0084	0.0588	0.0254	0.0151	0.01	0.0093
BS	**0.0059**	0.0072	0.0047	**0.0051**	0.0058	**0.006**	**0.007**	0.0043	**0.0051**	**0.0051**	**0.0054**	**0.0063**	**0.0049**	0.0052	**0.0054**	**0.0067**	**0.0071**	**0.005**	**0.0052**	**0.0049**

Below in table E1U99 are the error rate results at the α = 0.01 significance level for the upper limit for the mean statistic for data drawn from an exponential distribution with parameter λ = 0.10. Again similar results were found as what was found from the normal distribution for the sample mean. Increasing the number of Bootstrap resamples did impact the resulting comparison of relative accuracy of ET and BT. When a small number of Bootstrap resamples were used, ET was found to be more accurate than BT. When a larger number of Bootstrap resamples were used BT was found to be more accurate than ET. BC and BP at the α = 0.01 significance level, performed less accurately compared to EBC and EP at each sample size even when 10,000 Bootstrap resamples were specified. Further, at each sample size BC and BP had error rates with corresponding percent errors that were larger when 10,000 Bootstrap resamples were specified relative to when 200 Bootstrap resamples were specified. When 10,000 Bootstrap resamples were used BC_a had an error rate with corresponding percent error that was smaller at each sample size compared to when 200 Bootstrap resamples were used. Finally, the BS error rate was not found to be impacted by the number of Bootstrap resamples used.

Table Mean - E1U99 Advantage of EBSD($\boldsymbol{n}$) when comparing across Bootstrap Iteration Levels,
$\boldsymbol{\alpha}$ = 0.01 significance level, upper limit, $\boldsymbol{\alpha}$ = 0.01 significance level, upper limit

	Sample Size				
	N=5	N=10	N=15	N=30	N=40
E-skew	0.0413	0.0317	0.0224	0.0128	0.0116
ET	0.0607	0.0545	0.0413	0.0293	0.0244
EP	0.0985	0.0576	0.0359	0.0331	0.0298
EBC	0.1548	0.0639	0.0287	0.0117	0.0136

#Boots	**10000**					**1000**					**500**					**200**				
Sample Size	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**
BT	0.061	0.0546	0.0414	0.0295	0.0242	0.0609	0.0546	0.041	0.0299	0.0243	0.0609	0.0554	0.0424	0.0299	0.0247	0.0611	0.0548	0.0422	0.0308	0.0248
BC	0.1827	0.1109	0.0826	0.0502	0.0397	0.1817	0.1107	0.082	0.0488	0.0393	0.1814	0.1084	0.0799	0.0475	0.0388	0.1805	0.1064	0.0772	0.0465	0.037
BP	0.1374	0.0745	0.0499	0.0287	0.0222	0.1388	0.073	0.049	0.0272	0.0214	0.137	0.0729	0.0486	0.0282	0.0211	0.1352	0.0691	0.0454	0.0261	0.0206
$\boldsymbol{BC_a}$	0.1267	0.0557	0.0339	0.0166	0.0137	0.1279	0.0566	0.0349	0.0155	0.0132	0.1312	0.0595	0.0379	0.0172	0.0141	0.1388	0.07	0.0456	0.0261	0.0209
BS	**0.0177**	**0.0166**	**0.014**	**0.0084**	**0.0087**	**0.0186**	**0.0159**	**0.0133**	**0.0084**	**0.0089**	**0.0186**	**0.0173**	**0.0144**	**0.0078**	**0.009**	**0.0196**	**0.0143**	**0.0135**	**0.0074**	**0.0076**

Ratio of Means

Below in NU199 are the error rate results at the α = 0.01 significance level for the upper limit for the ratio of means statistic for the $N(\mu = 100, \sigma = 1)$, $N(\mu = 50, \sigma = 1)$ specification pair. Again similar results were found as what was found from the normal distribution for the sample mean. Increasing the number of Bootstrap resamples did impact the resulting comparison of relative accuracy of ET and BT. When a small number of Bootstrap resamples were used, ET was found to be more accurate than BT. When a larger number of Bootstrap resamples in the majority of cases ET was more accurate than BT but not always. BC and BP at the α = 0.01 significance level, performed less accurately compared to EBC and EP at each sample size even when 10,000 Bootstrap resamples were specified. Further, at each sample size BC and BP had error rates with corresponding percent errors that were larger when 10,000 Bootstrap resamples were specified relative to when 200 Bootstrap resamples were specified. When 10,000 Bootstrap resamples were used BC_a had an error rate with corresponding percent error that was smaller at each sample size compared to when 200 Bootstrap resamples were used. Finally, the BS error rate was not found to be impacted by the number of Bootstrap resamples used.

Table Ratio of Sample Means - NU199 Advantage of EBSD($\boldsymbol{n}$) when comparing across Bootstrap Iteration Levels, $\boldsymbol{\alpha}$ = 0.01 significance level, upper limit, $\boldsymbol{\alpha}$ = 0.01 significance level, upper limit

	Sample Size				
	N=5	**N=10**	**N=15**	**N=30**	**N=40**
E-skew	0.0098	0.0071	0.0058	**0.0047**	**0.0048**
ET	0.038	0.019	0.0123	0.0081	0.0072
EP	0.0285	0.0084	0.0031	0.0025	0.0041
EBC	0.029	0.0101	0.0037	0.0027	0.0046

#Boots	**10000**					**1000**					**500**					**200**				
Sample Size	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**
BT	**0.0065**	**0.0055**	**0.0055**	0.0043	0.0056	**0.0065**	**0.0056**	**0.0057**	0.0044	0.0056	**0.0067**	**0.0053**	**0.0057**	0.0047	**0.0056**	0.0069	**0.0056**	0.0061	**0.0051**	0.0058
BC	0.0486	0.0219	0.0132	0.0082	0.0071	0.0481	0.0216	0.0133	0.0085	0.0072	0.0488	0.0219	0.0125	0.0077	0.0069	0.0478	0.0196	0.0114	0.0077	0.0066
BP	0.0489	0.0216	0.013	0.0084	0.007	0.0482	0.0207	0.0128	0.0084	0.0081	0.0471	0.0204	0.0124	0.0065	0.007	0.0457	0.0192	0.0121	0.0082	0.0066
BC_a	0.0535	0.023	0.0138	0.0082	0.0072	0.0524	0.0222	0.0136	0.0085	0.0082	0.0522	0.0227	0.0126	0.0077	0.007	0.0548	0.0246	0.0145	0.0114	0.0086
BS	0.0076	0.0079	0.0065	0.0057	0.0062	0.0074	0.0075	0.0067	**0.0055**	**0.0052**	**0.0067**	0.0074	0.0068	**0.0051**	0.0059	0.0078	0.0072	**0.0057**	**0.0051**	**0.0054**

Pearson Correlation Coefficient

Below in NU199 are the error rate results at the α = 0.01 significance level for the upper limit for the Pearson correlation coefficient for data drawn from an bivariate normal distribution with parameters $N(\mu = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix})$. In this case increasing the number of Bootstrap resamples did not impact the resulting comparison of relative accuracy of ET and BT. BT using 200 resamples, had an error rate which was equal to or smaller compared to ET at each sample size. The same was found to be the case for BP when compared to EP. When 10,000 Bootstrap resamples were used BC_a had an error rate with corresponding percent error that was smaller at four of five sample size compared to when 200 Bootstrap resamples were used. Additionally, BC_a with 10,000 Bootstrap resamples had an error rate with a smaller percent error when compared to E-skew at sample size 10. However, BC_a with 200 Bootstrap resamples had an error rate with a larger percent error when compared to E-skew at every sample size considered. Therefore, increasing the number of Bootstrap resamples did impact the relative accuracy of BC_a and E-skew. For BS the number of resamples did not impact the relative accuracy when compared to E-skew. At one of five sample sizes BS using 10,000 resamples had an error rate with a smaller percent error when compared to E-skew. However, when 200 resamples were specified at this sample size, BS had an error rate with an even smaller percent error than when 10,000 resamples were specified. Therefore, BS was found to be more accurate than E-skew at sample size 15 regardless whether 10,000 or 200 resamples were specified.

Table: Pearson Correlation Coefficient - NU199 Advantage of EBSD($\boldsymbol{n}$) when comparing across Bootstrap Iteration Levels, $\boldsymbol{\alpha}$ = 0.01 significance level, upper limit

	Sample Size				
	N=5	N=10	N=15	N=30	N=40
E-skew	0.0065	0.0068	0.0043	**0.0059**	**0.0053**
ET	0	0	0	0	0
EP	3e-04	0	0	0	0
EBC	8e-04	1e-04	0	0	0

#Boots	**10000**					**1000**					**500**					**200**				
Sample Size	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**	**N=5**	**N=10**	**N=15**	**N=30**	**N=40**
BT	0	0.0111	0.0102	0.0082	0.0086	0	0.0109	0.0105	0.0082	0.0081	0	0.0112	0.0111	0.0086	0.0087	1e-04	0.011	0.0107	0.008	0.0088
BC	1e-04	0.0018	0.0045	0.0068	0.0065	1e-04	0.0019	**0.0043**	**0.0072**	**0.0073**	1e-04	0.0018	**0.0047**	**0.0059**	**0.0068**	0	**0.0029**	**0.0048**	**0.0065**	**0.0062**
BP	**0.0062**	0.0132	0.0115	0.0085	0.009	**0.0062**	0.0128	0.0112	0.0087	0.0081	0.0062	0.0121	0.0106	0.0081	0.0089	**0.0075**	0.0128	0.0108	0.0081	0.0083
$\boldsymbol{BC_a}$	0.0014	**0.006**	0.0071	0.0074	0.007	0.003	**0.0067**	0.0065	0.0074	**0.0073**	**0.0044**	**0.0079**	0.0085	0.0073	0.008	0.0077	0.0138	0.0128	0.0095	0.0097
BS	1e-04	0.0018	**0.0045**	0.0068	0.0065	1e-04	0.0019	**0.0043**	**0.0072**	**0.0073**	1e-04	0.0018	**0.0047**	**0.0059**	**0.0068**	0	**0.0029**	**0.0048**	**0.0065**	**0.0062**

4.7 Application of EBSD(n) in Microarray Data

The purpose of this section is to test the level of agreement between the E-skew method and Monte Carlo Bootstrap methods using real data. Agreement between the E-skew method and Bootstrap methods is tested using Cohen's Kappa statistic defined in detail below. The real data example included in the Gene expression Omnibus (GEO) project with accession code GSE9574 was originally used for a study by Tripathi A, King C, de la Morenas A, et al. The available data consists of two cohorts and 22,283 genes for 29 available patients. The first cohort consists of patients with epithelium adjacent to a breast tumor and the second consists of patients undergoing reduction mammoplasty without apparent breast cancer.

For this work the complete set of genes is divided into two data sets analyzed separately. The first data set, the Differentially Expressed Data Set, includes the top 250 differentially expressed genes. Differential expression between the two cohorts is determined by the GEO2R software. The second data set, the non-Differentially Expressed Data Set, includes the remaining genes not among the top 250 most differentially expressed. For this work all available data from the GSE9574 study are used to rank differential expression.

To assess the agreement between the E-skew method and the Monte Carlo Bootstrap, the Cohen's Kappa statistic is used as mentioned above. The Kappa algorithm is derived as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where $p_0 = \frac{a+d}{a+b+c+d}$ and $p_e = \frac{(a+b)}{(a+b+c+d)} * \frac{(a+c)}{(a+b+c+d)} + \frac{(d+c)}{(a+b+c+d)} * \frac{(d+b)}{(a+b+c+d)}$, and where a, b, c and d represent the cell counts in the 2X2 contingency table with an example table shown below.

Example of two by two contingency table used in the work below for a given alpha level and sample size.			
	E-skew		
	Positive	**Negative**	**Total**
Bootstrap Positive	a	b	$a + c$
Bootstrap Negative	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

The cell counts $a, b, c,$ and d are defined as follows:

- The a cell refers to the number genes positively classified by both the E-skew and Bootstrap methods. For the Differentially Expressed Data Set the a cell is the number genes classified by both the E-skew method and Bootstrap methods as differentially expressed. Conversely, for the non-Differentially Expressed Data Set the a cell is the number genes classified as not significantly differentially expressed by both the E-skew method and Bootstrap method.
- The d cell refers to the number of genes negatively classified for both the E-skew method and Bootstrap method. For the Differentially Expressed Data Set the d cell is the number of genes classified by both the E-skew method and Bootstrap methods as not differentially expressed. Conversely, for the non-Differentially Expressed Data Set the d cell is the number genes classified as significantly differentially expressed by both the E-skew method and Bootstrap method.

- The b cell refers to the number of genes negatively classified by the E-skew method that the Bootstrap method positively classified. For the Differentially Expressed Data Set the b cell is the number of genes classified by the E-skew method as not differentially expressed that the Bootstrap method classified as differentially expressed. Conversely, for the non-Differentially Expressed Data Set the b cell is the number genes classified as significantly differentially expressed by the E-skew method that the Bootstrap method classified as not differentially expressed.
- The c cell refers to the number of genes negatively classified by the Bootstrap method that the E-skew positively classified. For the Differentially Expressed Data Set the c cell is the number of genes classified by the E-skew method as differentially expressed that the Bootstrap method classified as not differentially expressed. Conversely, for the non-Differentially Expressed Data Set the c cell is the number genes classified as not significantly differentially expressed by the E-skew method that the Bootstrap method classified as differentially expressed.

The Kappa statistic range of possible values is from -1 to 1. A Kappa agreement of 1 means two tests yield identical results for every data point considered. Conversely a Kappa statistic of -1 means two tests yield opposite results for every data point considered.

Cohen's Kappa can also have many different qualitative interpretations. The below will be used to define the level of agreement between E-skew and the Monte Carlo Bootstrap method in the results section. This approach for interpreting Kappa can be found in Koch and Landis (1977).

- A Kappa value less than 0.00 means poor agreement between two methods
- A Kappa value ranging from 0.00-0.20 means slight agreement between two methods
- A Kappa value ranging from 0.21-0.40 means fair agreement between two methods
- A Kappa value ranging from 0.41-0.60 means moderate agreement between two methods
- A Kappa value ranging from 0.61-0.80 means substantial agreement between two methods
- A Kappa value greater than 0.80 means near perfect agreement between two methods

The standard error of the Kappa statistic can be computed as:

$$SE_{\kappa} = \sqrt{\frac{p_0(1-p_0)}{N(1-p_e)^2}}$$

The standard error for the Kappa statistic can be computed to create a confidence interval for the Kappa statistic. The corresponding confidence interval for the Kappa statistic is computed as:

$$CI_{\kappa} = \kappa \pm z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{N(1-p_e)^2}}$$

The confidence interval can be used to test whether the agreement between two tests (in this case two intervals) is statistically significantly different from 0.

E-skew is tested for agreement with five Monte Carlo Bootstrap methods: the BS, BC_a, BP, BC and BT. The process flow of how the E-skew is compared to the Bootstrap is as follows:

First the complete set of genes are ranked by differential expression between the two cohorts using all available data. Then EBSD(n) is used to generate an EBSD(n) sampling distribution for each gene and for each cohort and simultaneously the Monte Carlo Bootstrap is used to generate a Monte Carlo Bootstrap sampling distribution for each of the same. To compute each Monte Carlo confidence interval 10,000 Bootstrap resamples are drawn consistent with the number of Bootstraps specified in sections 4.1-4.5.

Confidence intervals are generated at each of three sample size levels 5, 10 and 14 for each of three significance level (α =0.01, α =0.05, α =0.10). Each EBSD(n) sampling distribution is then used to generate an E-skew confidence interval and each Monte Carlo Bootstrap sampling distribution is used to generate five different Monte Carlo Bootstrap confidence intervals: a BP, BT, BC, BC_a, and BS interval. For each data set the confidence interval generated for each gene in each cohort is compared. Then 2X2 contingency tables are created comparing E-skew to each Monte Carlo Bootstrap method. Contingency tables are reported for the Differentially Expressed Data Set and the non-Differentially Expressed Data Set separately. These contingency tables are then used to generate a Cohen's Kappa statistic of agreement and corresponding Kappa confidence interval comparing E-skew to each Monte Carlo Bootstrap method.

Application of E-skew in Microarray Data: Differentially Expressed Data Set

At the α = 0.01 significance level in table DE599 below E-skew had slight or poor agreement with each Monte Carlo Bootstrap method considered at sample size 5. As sample size increased in tables DE1099 and DE1499 the Kappa agreement increased between E-skew and each Monte Carlo Bootstrap method. At sample size 10 E-skew had slight or fair agreement with each Monte Carlo Bootstrap method. At sample size 14 E-skew had fair or moderate agreement with each Monte Carlo Bootstrap method. At each sample size none of the Kappa agreements were statistically significantly different from one another.

Further, at sample size 5 none of the Kappa coefficients were statistically significantly different from 0 indicating any agreement found between E-skew and a Monte Carlo Bootstrap method was by chance. At sample size 10 and 14 however the agreement between E-skew and each Monte Carlo Bootstrap method other than the BC at sample size 10 were found to be statistically significantly greater than 0. The E-skew had highest agreement at sample size 10 with the BP method. The agreement with BP at sample size 10 was not significantly higher than it was with BT, BC_{a} and BS. The E-skew also had highest agreement at sample size 14 with the BS method. This agreement was also not statistically significantly higher than the agreement between E-skew and any of the other methods.

Kappa agreements were lower at the α = 0.01 significance level for the Differentially Expressed Data Set. In general Kappa agreements were lower at sample size 5 at the α = 0.01 significance level than they were for any combination of significance level and sample size for both sets of genes. However, at each successive sample size Kappa agreements increased between E-skew and each Monte Carlo

Bootstrap method. These patterns can also be viewed in Figure DE99 below. In the figure each Kappa was plotted at each sample size for the Differentially Expressed Data Set at the α = 0.01 significance level.

Table DE599 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5

	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	0	38	-0.008	(-0.39 , 0.374)
BC Negative	1	211		
BT Positive	0	1	0	(-1 , 1)
BT Negative	0	249		
BP Positive	1	35	0.047	(-0.338 , 0.432)
BP Negative	0	214		
BC_a **Positive**	1	37	0.044	(-0.330 , 0.418)
BC_a **Negative**	0	212		
BS Positive	0	1	0	(-1 , 1)
BS Negative	0	249		

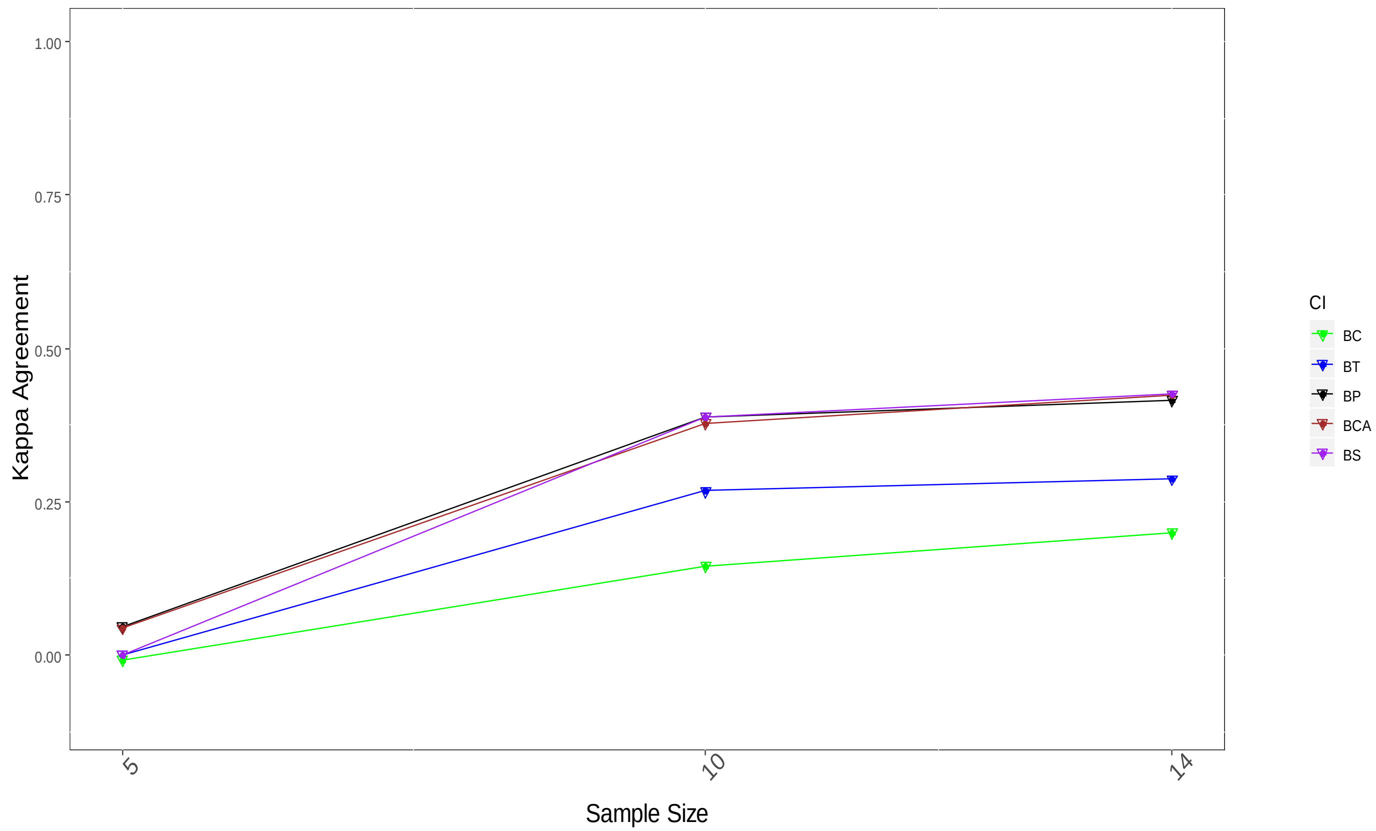
Table DE1099 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10

	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	18	47	0.145	(-0.069 , 0.36)
BC Negative	27	158		
BT Positive	10	5	0.267	(-0.006 , 0.541)
BT Negative	35	200		
BP Positive	35	51	0.389	(0.214 , 0.564)
BP Negative	10	154		
BC_a **Positive**	38	60	0.378	(0.211 , 0.545)
BC_a **Negative**	7	145		
BS Positive	13	1	0.388	(0.133 , 0.644)
BS Negative	32	204		

Table DE1499 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14

	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	117	45	0.20	(0.028 , 0.372)
BC Negative	46	42		
BT Positive	83	15	0.287	(0.138 , 0.435)
BT Negative	80	72		
BP Positive	154	50	0.416	(0.244 , 0.587)
BP Negative	9	37		
BC_a **Positive**	155	50	0.424	(0.253 , 0.595)
BC_a **Negative**	8	37		
BS Positive	84	79	0.425	(0.288 , 0.563)
BS Negative	0	87		

Figure DE99 Kappa Agreements for the Differentially Expressed Data Set at the 0.01 alpha level



At the α = 0.05 and α = 0.10 significance levels in tables DE595, DE1095, DE1495, DE590, DE1090, and DE1490 below, Kappa agreements were higher between E-skew and each Monte Carlo Bootstrap method than they were at the α = 0.01 significance level for a given sample size. At sample size 5 for both significance levels the E-skew method was in highest agreement with the BT method. Further at sample size 5, agreements ranged from fair to substantial at the α = 0.05 significance level, and moderate to substantial at the α = 0.10 significance level. At the α = 0.05 and α = 0.10 significance levels the agreement between E-skew and BT was statistically significantly higher than the BC, BP and BC_{a} methods but not higher than the BS method.

At sample size 10 for both alpha levels, E-skew had highest agreement with the BS method. The agreements ranged from fair to substantial. No one method had a statistically significantly higher Kappa than every other method. However, the BC Kappa was statistically significantly lower than the BC_{a} and BS Kappa's. All of the Kappa's were statistically significantly greater than 0.

At sample size 14 in table DE1495 and DE1490 below, for both alpha levels, E-skew had highest agreement with the BC_{a} method. Agreements ranged from fair to substantial at the α = 0.05 significance level, and fair to perfect at the α = 0.10 significance level. Only the BC_{a} Kappa was statistically significantly greater than 0 at both the α = 0.05 and α = 0.10 significance levels. Further, at the α = 0.10 significance level the E-skew had perfect agreement with the BC_{a} and BP methods. All three methods the BC_{a}, BP and E-skew found 248 of the 250 differentially expressed genes to be differentially expressed.

As can be viewed in Figures DE95 and DE90, only the BC_a Kappa increased at each successive sample size at both the α = 0.05 and α = 0.10 significance levels. Further at the α = 0.10 significance level the BP Kappa increased at each successive sample size. At the α = 0.05 significance level all other method's Kappa decreased from sample size 10 to sample size 14. At the α = 0.10 significance level the BS and BT method's Kappa were lower at sample size 14 than they were at sample size 10.

The reason there was lower agreement at sample size 14 at these two alpha levels for the BS and BT methods was because BS and BT found a lower count of differentially expressed genes. Table E1L99 in section 4.1 demonstrated at sample size 15 BS and BT's lower limit percent error was greater than for the E-skew and BC_a methods. In turn their confidence interval's length was larger than E-skew and BC_a. The larger each cohort's interval the more likely the intervals will overlap. Therefore, because BT and BS's interval length were longer they were classifying fewer genes as differentially expressed and generating less agreement with E-skew.

Table DE595 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5				
	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	16	60	0.262	(0.101 , 0.423)
BC Negative	1	173		
BT Positive	12	6	0.662	(0.467 , 0.857)
BT Negative	5	227		
BP Positive	17	74	0.226	(0.078 , 0.374)
BP Negative	0	159		
BC_a **Positive**	17	79	0.21	(0.065 , 0.354)
BC_a **Negative**	0	154		
BS Positive	9	1	0.649	(0.424 , 0.874)
BS Negative	8	232		

Table DE1095 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10				
	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	101	56	0.275	(0.154 , 0.396)
BC Negative	33	60		
BT Positive	91	17	0.525	(0.42 , 0.63)
BT Negative	43	99		
BP Positive	129	55	0.503	(0.394 , 0.613)
BP Negative	5	61		
BC_a **Positive**	134	47	0.611	(0.511 , 0.712)
BC_a **Negative**	0	69		
BS Positive	97	2	0.693	(0.604 , 0.781)
BS Negative	37	114		

Table DE1495 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14

	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	219	7	0.228	(-0.052 , 0.509)
BC Negative	19	5		
BT Positive	216	6	0.25	(-0.012 , 0.512)
BT Negative	22	6		
BP Positive	238	9	0.388	(-0.004 , 0.781)
BP Negative	0	3		
BC_a **Positive**	235	4	0.681	(0.448 , 0.914)
BC_a **Negative**	3	8		
BS Positive	214	0	0.461	(0.256 , 0.666)
BS Negative	24	12		

Figure DE95 Kappa Agreements for the Differentially Expressed Data Set at the 0.05 alpha level

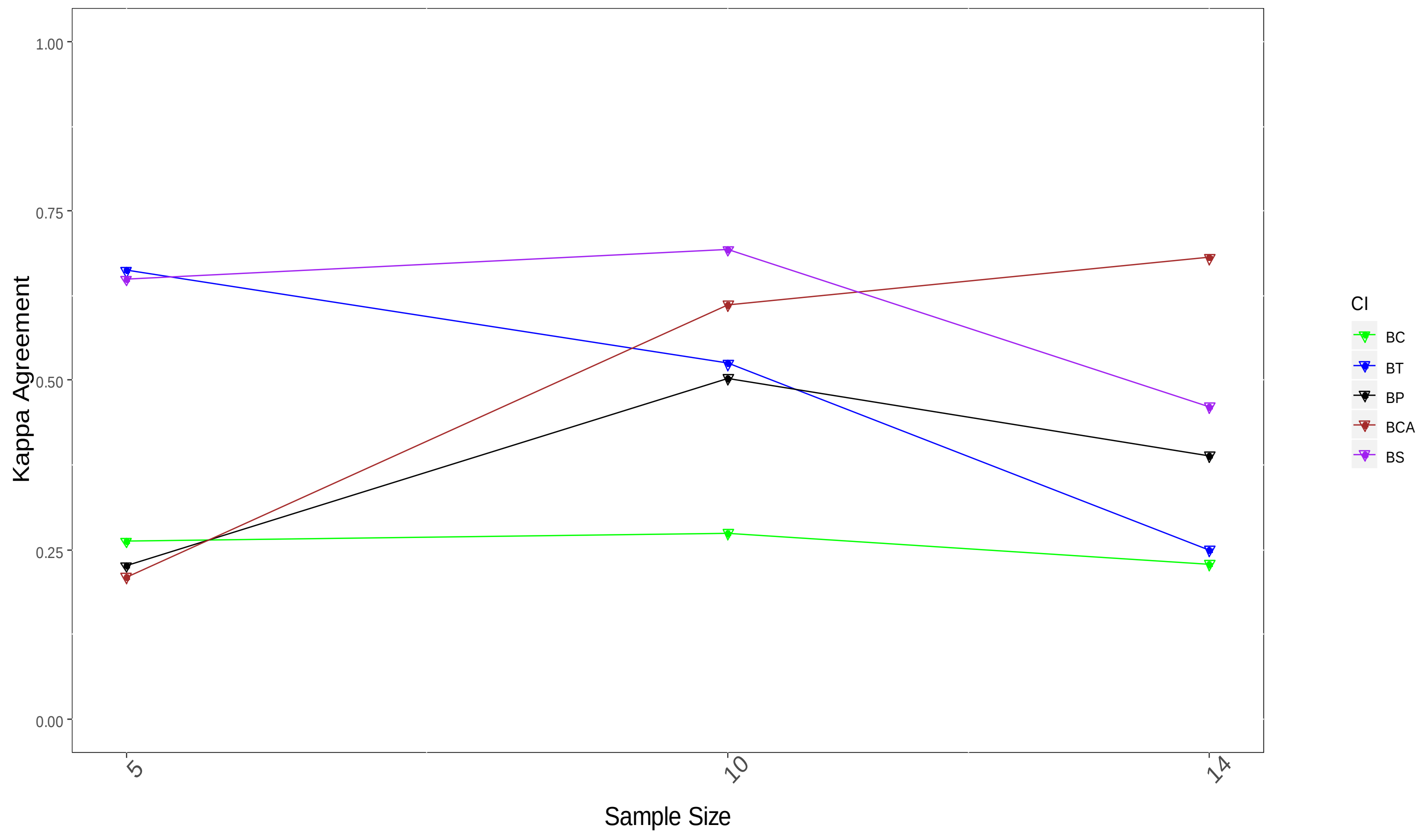


Table DE590 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5

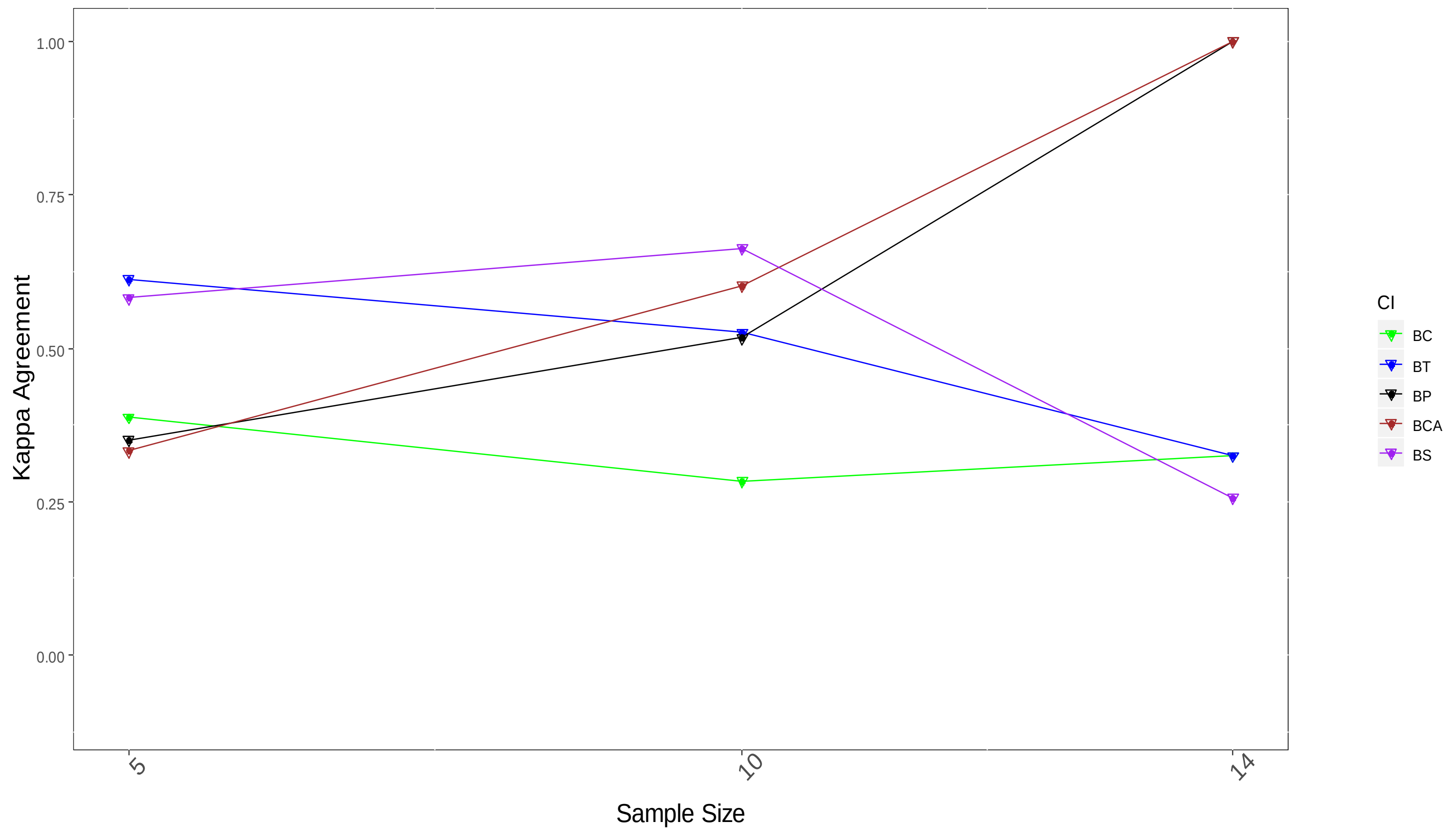
	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	40	69	0.387	(0.285 , 0.490)
BC Negative	1	140		
BT Positive	34	24	0.612	(0.505 , 0.720)
BT Negative	7	185		
BP Positive	41	79	0.351	(0.251 , 0.450)
BP Negative	0	130		
BC_a **Positive**	41	83	0.332	(0.234 , 0.431)
BC_a **Negative**	0	126		
BS Positive	22	6	0.582	(0.452 , 0.712)
BS Negative	19	203		

Table DE1090 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10

	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	172	32	0.283	(0.146 , 0.420)
BC Negative	25	21		
BT Positive	167	14	0.526	(0.419 , 0.632)
BT Negative	30	39		
BP Positive	196	31	0.517	(0.386 , 0.648)
BP Negative	0	22		
BC_a **Positive**	197	27	0.603	(0.484 , 0.722)
BC_a **Negative**	0	26		
BS Positive	162	0	0.662	(0.575 , 0.749)
BS Negative	35	53		

Table DE1490 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level for the Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14				
	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	240	0	0.324	(-0.062 , 0.711)
BC Negative	8	2		
BT Positive	240	0	0.324	(-0.062 , 0.711)
BT Negative	8	2		
BP Positive	248	0	1.00	(1 , 1)
BP Negative	0	2		
$\boldsymbol{BC_a}$ **Positive**	248	0	1.00	(1 , 1)
$\boldsymbol{BC_a}$ **Negative**	0	2		
BS Positive	237	0	0.256	(-0.104 , 0.617)
BS Negative	11	2		

Figure DE90 Kappa Agreements for the Differentially Expressed Data Set at the 0.10 alpha level



Application of E-skew in Microarray Data: non-Differentially Expressed Data Set

At the α = 0.01 significance level in Table NDE599 below E-skew had slight agreement with each Monte Carlo Bootstrap method considered at sample size 5. At sample size 10 in Table NDE1099 below E-skew had slight or fair agreement with each method. At sample size 14 in Table NDE1499 E-skew had fair or moderate agreement with each method.

At sample size 5 the method with highest agreement with E-skew was BS. However, none of the Kappa's were significantly different from 0. As sample size increased in tables NDE1099 and NDE1499 the Kappa agreement generally increased but not always. For every method, the Kappa at sample size 14 was greater than it was at sample size 5 except in the case of the BS method. Further at sample sizes 10 and 14 the BC_{a} had highest agreement with E-skew. At both sample sizes the BC_{a} Kappa's were significantly different from 0. Also at sample size 14 the BP Kappa was significantly different from 0.

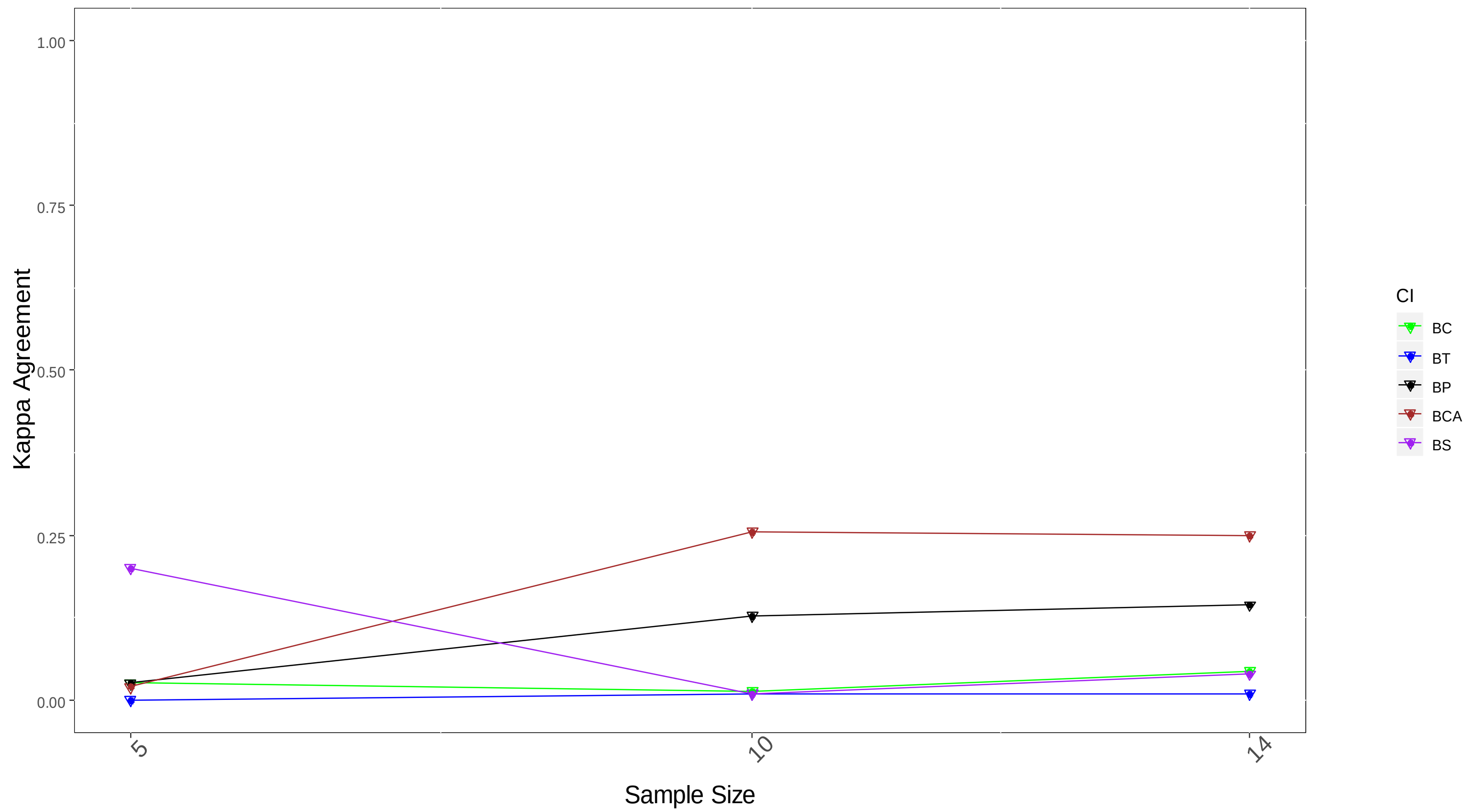
As seen in Figure NDE99, the Kappa agreement between E-skew and each method is plotted for the non-Differentially Expressed Data Set at the α = 0.01 significance level.

Table NDE599 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5				
	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	21502	2	0.025	(-0.083 , 0.134)
BC Negative	522	7		
BT Positive	22024	9	0	(-0.858 , 0.858)
BT Negative	0	0		
BP Positive	21352	0	0.025	(-0.070 , 0.121)
BP Negative	672	9		
$\boldsymbol{BC_a}$ **Positive**	21158	0	0.02	(-0.065 , 0.104)
$\boldsymbol{BC_a}$ **Negative**	866	9		
BS Positive	21997	5	0.199	(-0.165 , 0.564)
BS Negative	27	4		

Table NDE1099 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10				
	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	21785	208	0.013	(-0.148 , 0.174)
BC Negative	38	2		
BT Positive	21822	209	0.009	(-0.166 , 0.185)
BT Negative	1	1		
BP Positive	21777	192	0.128	(-0.017 , 0.272)
BP Negative	46	18		
$\boldsymbol{BC_a}$ **Positive**	21740	166	0.256	(0.135 , 0.377)
$\boldsymbol{BC_a}$ **Negative**	83	44		
BS Positive	21823	209	0.009	(-0.166 , 0.185)
BS Negative	0	1		

Table NDE1499 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.01 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14				
	E-skew			
	Positive	**Negative**	κ	99% CI for κ
BC Positive	20851	1066	0.044	(-0.027 , 0.114)
BC Negative	84	32		
BT Positive	20930	1092	0.01	(-0.065 , 0.085)
BT Negative	5	6		
BP Positive	20820	994	0.144	(0.079 , 0.208)
BP Negative	115	104		
$\boldsymbol{BC}_{\mathrm{a}}$ **Positive**	20738	896	0.250	(0.193 , 0.307)
$\boldsymbol{BC}_{\mathrm{a}}$ **Negative**	197	202		
BS Positive	20934	1075	0.039	(-0.035 , 0.113)
BS Negative	1	23		

Figure NDE99 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.01 alpha level



At the α = 0.05 significance level in tables NDE595, NDE1095, and NDE1495 below E-skew had slight to substantial agreement at sample sizes 5 and 14 while fair to moderate agreement at sample size 10. At this alpha level all the Kappa's at each sample size were statistically significantly greater than 0. At sample size 14 the BC_a Kappa agreement was statistically significantly higher than any other agreement.

At the α = 0.10 significance level in tables NDE590, NDE1090, and NDE1490 below, E-skew had slight to moderate agreement at sample sizes 5, and moderate to substantial agreement at sample size 10. At sample size 14 each method's Kappa agreement was substantial or near perfect. At this alpha level all the Kappa's at each sample size were also statistically significantly greater than 0. At sample size 14 the BC_a Kappa agreement was statistically significantly higher than any other agreement.

Figure NDE95 and NDE90 below display the Kappa agreement between E-skew and each method for the α = 0.05 and α = 0.10 significance levels respectively. As is displayed in the figures, the Kappa agreement's increased at each successive sample size for the BC_a, BP and BC methods.

Table NDE595 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5				
	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	21497	2	0.188	(0.116 , 0.260)
BC Negative	477	57		
BT Positive	21945	26	0.544	(0.424 , 0.665)
BT Negative	29	33		
BP Positive	21324	0	0.149	(0.085 , 0.214)
BP Negative	650	59		
$\boldsymbol{BC}_{\mathrm{a}}$ **Positive**	21183	0	0.125	(0.066 , 0.185)
$\boldsymbol{BC}_{\mathrm{a}}$ **Negative**	791	59		
BS Positive	21965	22	0.704	(0.600 , 0.808)
BS Negative	9	37		

Table NDE1095 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10				
	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	21489	231	0.334	(0.272 , 0.396)
BC Negative	200	113		
BT Positive	21653	266	0.335	(0.261 , 0.41)
BT Negative	36	78		
BP Positive	21443	117	0.548	(0.501 , 0.594)
BP Negative	246	227		
$\boldsymbol{BC}_{\mathrm{a}}$ **Positive**	21312	35	0.592	(0.552 , 0.631)
$\boldsymbol{BC}_{\mathrm{a}}$ **Negative**	377	309		
BS Positive	21686	222	0.516	(0.453 , 0.579)
BS Negative	3	122		

Table NDE1495 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.05 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14

	E-skew			
	Positive	**Negative**	κ	95% CI for κ
BC Positive	20226	877	0.504	(0.476 , 0.532)
BC Negative	278	652		
BT Positive	20453	1023	0.465	(0.434 , 0.497)
BT Negative	51	506		
BP Positive	20177	520	0.684	(0.663 , 0.705)
BP Negative	327	1009		
BC_a **Positive**	20045	319	0.738	(0.720 , 0.756)
BC_a **Negative**	459	1210		
BS Positive	20502	808	0.624	(0.598 , 0.649)
BS Negative	2	721		

Figure NDE95 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.05 alpha level

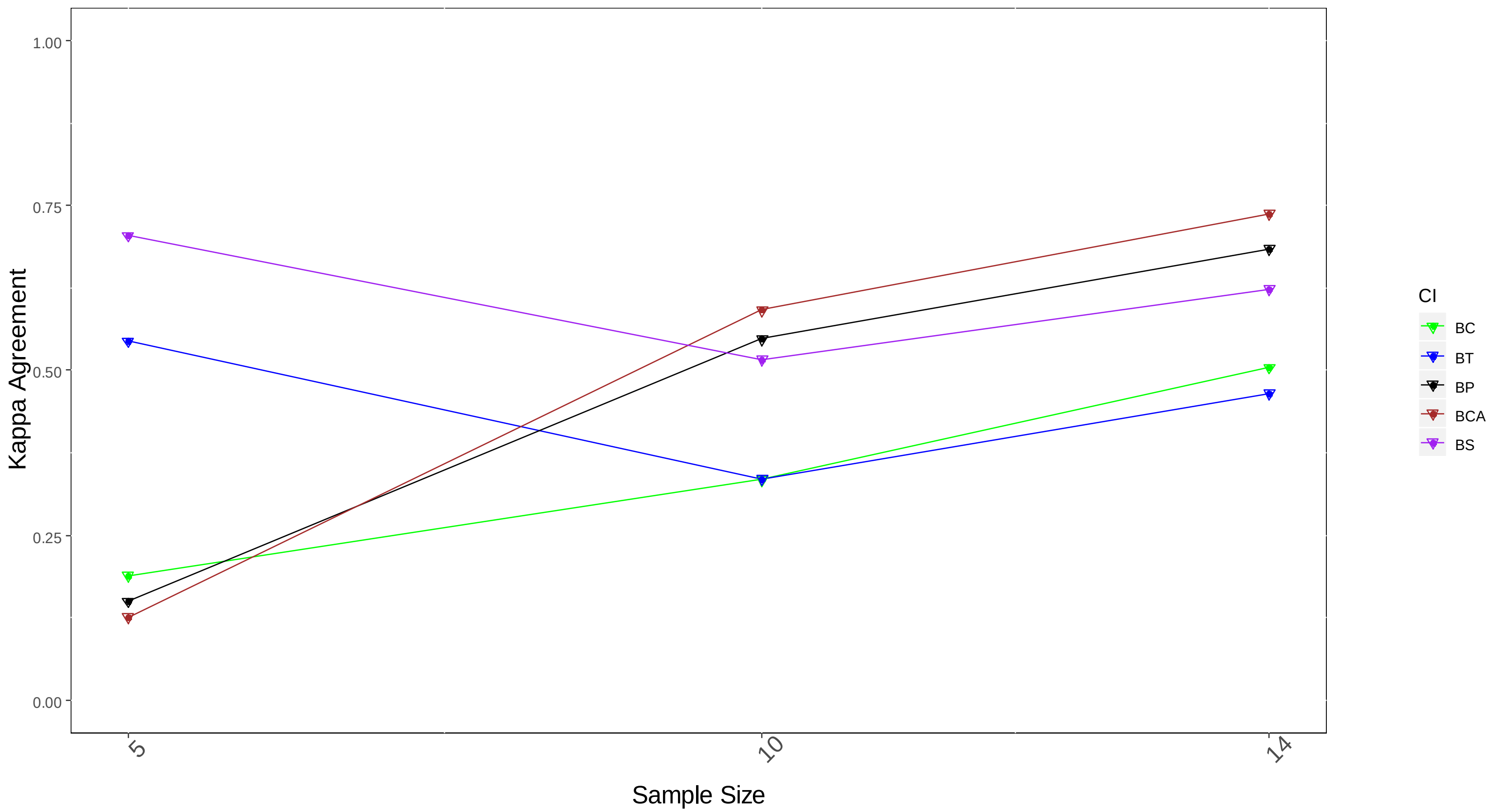


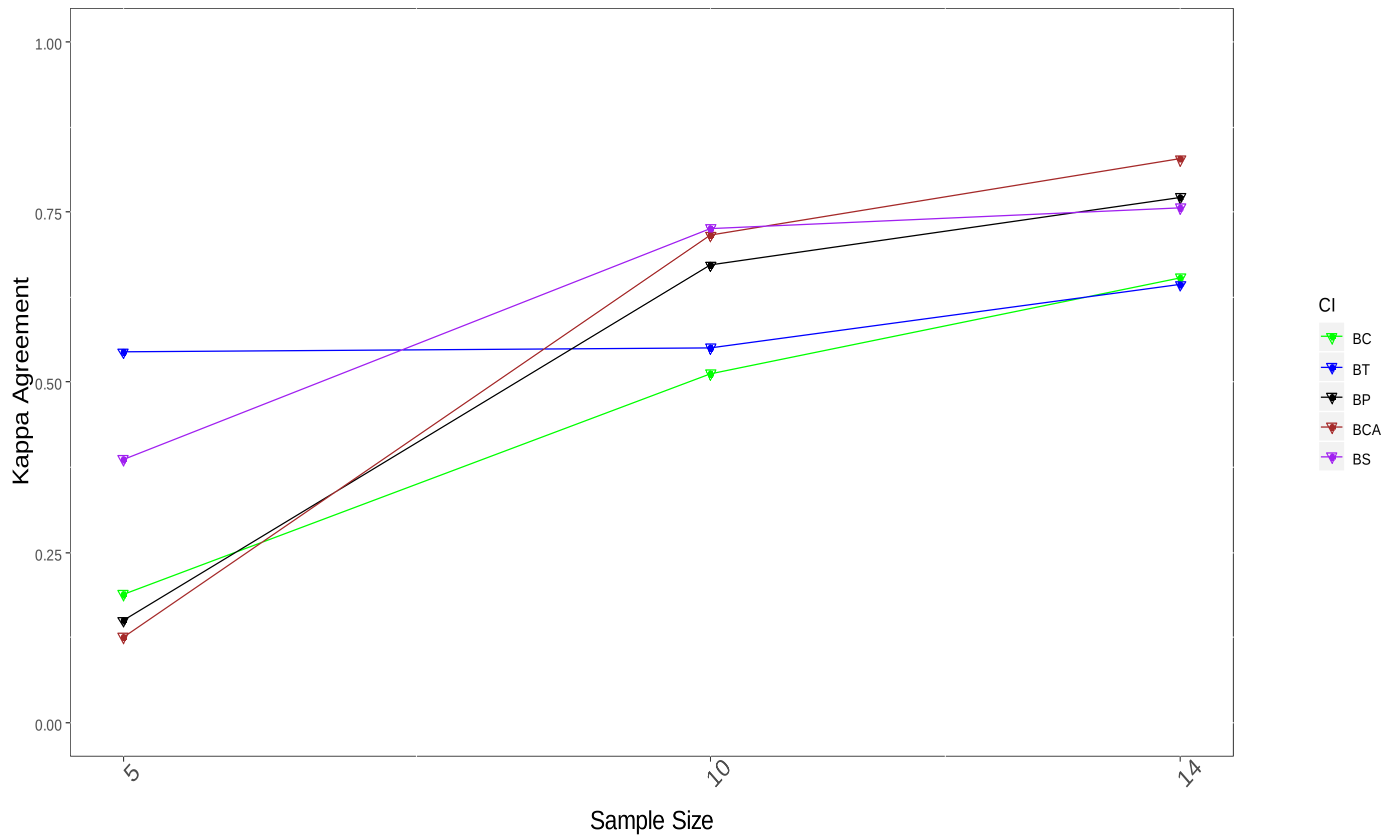
Table NDE590 Comparison of E-skew and Monte Carlo Bootstrap classification at 0.10 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =5				
	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	21497	2	0.188	(0.128 , 0.249)
BC Negative	477	57		
BT Positive	21945	26	0.544	(0.443 , 0.645)
BT Negative	29	33		
BP Positive	21324	0	0.149	(0.095 , 0.204)
BP Negative	650	59		
$\boldsymbol{BC_a}$ **Positive**	21183	0	0.125	(0.075 , 0.176)
$\boldsymbol{BC_a}$ **Negative**	791	59		
BS Positive	21965	22	0.387	(0.294 , 0.481)
BS Negative	9	37		

Table NDE1090 Comparison of E-skew and Monte Carlo Bootstrap classification at 0.10 alpha level for the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =10				
	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	20649	468	0.513	(0.487 , 0.540)
BC Negative	412	504		
BT Positive	20962	551	0.551	(0.522 , 0.579)
BT Negative	99	421		
BP Positive	20586	213	0.672	(0.652 , 0.692)
BP Negative	475	759		
$\boldsymbol{BC_a}$ **Positive**	20401	31	0.716	(0.698 , 0.733)
$\boldsymbol{BC_a}$ **Negative**	660	941		
BS Positive	21051	402	0.725	(0.703 , 0.748)
BS Negative	10	570		

Table NDE1490 Comparison of E-skew and Monte Carlo Bootstrap classification at the 0.10 alpha level among the non-Differentially Expressed Data Set. Significance measured indirectly by comparing confidence intervals in each cohort, sample size =14

	E-skew			
	Positive	**Negative**	κ	90% CI for κ
BC Positive	18767	1203	0.653	(0.639 , 0.667)
BC Negative	337	1726		
BT Positive	19000	1375	0.643	(0.629 , 0.658)
BT Negative	104	1554		
BP Positive	18725	724	0.771	(0.760 , 0.782)
BP Negative	379	2205		
BC_a **Positive**	18607	391	0.828	(0.818 , 0.837)
BC_a **Negative**	497	2538		
BS Positive	19099	1045	0.757	(0.745 , 0.769)
BS Negative	5	1884		

Figure NDE90 Kappa Agreements for the non-Differentially Expressed Data Set at the 0.10 alpha level



Application of E-skew in Microarray Data Results Discussion

When comparing the Kappa agreements between the non-Differentially Expressed and Differentially Expressed data sets, there were varied results between the two data sets depending on the alpha level and sample size. For the non-Differentially Expressed Data Set each Kappa agreement was statistically significantly greater than 0 for both the α = 0.05 and α = 0.10 significance levels. This could not be said for every method for the Differentially Expressed Data Set. Although, this was the case at the α = 0.05 and α = 0.10 significance levels, at the α = 0.01 significance level, the Kappa's for the Differentially Expressed Data Set were higher at sample sizes 10 and 14. Further, these Kappa's were statistically significantly greater than 0. This could not be said for the genes from the non-Differentially Expressed Data Set at this significance level at these sample sizes. Therefore, which data set yielded higher Kappa agreements depended on the significance level considered.

Although there were varied agreements depending on the significance level considered, these results were encouraging when referring to the simulation results in section 4.1. Table E199U reported upper limit confidence interval error rates at successive sample sizes 5, 10, and 15 at the 0.01 alpha level for data drawn from the exponential distribution. Data drawn from an exponential distribution with fixed α = 1, are considered moderately skewed with skew approximately equal to 2 as mentioned previously. The distribution of gene fold expression data studied in this section were also right skewed. In Table E199U the method with the smallest deviation from the true nominal error rate for the upper limit level was the BS method. Simulation results at the α = 0.05 and α = 0.10 significance levels are also reported in the Appendix. These

Appendix tables also found BS to have the smallest deviation from the true error rate at the α = 0.05 and α = 0.10 significance levels. Similarly, among the methods considered in section 4.7 the BC_{a} method had the smallest deviation from the nominal one-sided error rate for the lower limit. This can be seen in tables E199L and Appendix tables.

These results are not surprising because at small sample sizes the BC_{a} and BS methods are more accurate as second order methods and have error rates at small sample sizes closer to target nominal error rates compared to first order methods. For sample sizes 10 and 14, other than for the Differentially Expressed Data Set at the α = 0.01 significance level at sample size 10, the method with the highest Kappa was either the BS or BC_{a}. At sample size 5 the method with highest agreement with E-skew varied but at such a small sample size intervals are inherently less reliable because so little data is available. Consequently, having such high agreement with the most accurate Monte Carlo Bootstrap methods indicates the E-skew method is performing appropriately in this real data example.

Further the E-skew uses EBSD(n) which has inherent advantages. One disadvantage of relying on Monte Carlo resampling in testing individual gene expression is every time resampling is performed this changes the significance or non-significance of a given gene. Although most genes will repeatedly be found to be significant or non-significant for each 10,000 resamples performed, a path forward remains unclear for those genes that alternate between significant and not depending on the trial run. In other words, if a second set of 10,000 Bootstrap resamples are performed, some genes which are found to be non-significant in the first set could then be found to be significant in the second because of random Monte Carlo variation. Further there is not only Monte Carlo variation but user variation. The selection of the number Monte Carlo Bootstrap resamples as mentioned before will vary depending on the user. The selection of 10,000 Monte Carlo Bootstrap resamples was made to limit variation. In fact, a user could choose to only perform 200 or 500 resamples, as is often done, and this would increase the variation in classification from one trial to the next. This highlights an advantage of using the E-skew method. The E-skew method will find the same genes to be significant with repeated runs of the algorithm any time the method is performed on the same data because it relies on the same underlying fixed design of EBSD(n).

Although the results varied depending on the significance level, sample size and data set under consideration frequently E-skew was found to be in statistically significant agreement with many of the Monte Carlo Bootstrap methods. Further the results show the E-skew method had high agreement with second order methods BC_{a} and BS at sample size 10 and 14 in nearly every case. EBSD(n) provides a framework from which E-skew can derive consistent results. Thus E-skew provided consistent results with second order methods at small sample sizes based on Cohen's Kappa while also providing consistent results by using EBSD(n).

Chapter 5: Conclusion

The Monte Carlo Bootstrap has been a common but inefficient technique for confidence interval computation. The inefficiency in the technique, namely, is that it introduces simulation error. Hence, the author was motivated to compare confidence interval methods using the Efficient Bootstrap Sample Design for a sample of size n (EBSD(n)) to those using the Monte Carlo Bootstrap. The purpose of this comparison was to highlight the added advantage of using EBSD(n) to eliminate Monte Carlo Bootstrap simulation error in confidence interval computation. Some aspects of this simulation study highlighted the benefit of using EBSD(n) more so than others.

In general the relative accuracy of methods using EBSD(n) varied depending on the sample size, significance level, statistic and distribution studied. Relative improvement in accuracy using EBSD(n) was found frequently at the α = 0.01 significance level for each statistic studied. Additionally, E-skew provided relative accuracy for non-linear statistics across α significance level. Noteworthy results for specific sample sizes, significance levels, statistics and statistical distributions are described below.

For the sample mean when the data was approximately normally distributed at the α = 0.01 significance level, percentile methods such as EP and EBC, performed relatively more accurately compared to E-skew and Monte Carlo Bootstrap methods. Although EP and EBC performed relatively accurately at the α = 0.01 significance level, they performed relatively less accurately at the α = 0.05 and α = 0.10 significance levels. Although EP and EBC performed relatively less accurately in the latter case, E-skew performed relatively accurately at the α = 0.05 and α = 0.10 significance levels, providing a reasonable facsimile to the most accurate Monte Carlo Bootstrap methods, BS and BC_a.

For the sample mean at the α = 0.01 significance level, EBC_a performed relatively accurately at small sample sizes when data was generated from a skewed distribution such as data generated from the exponential distribution. EBC_a, in particular, performed relatively accurately at sample sizes 10, 15 and 20 as it had the error rate with the smallest percent error at each of these sample sizes. This is a significant finding as often BS performs most accurately for the long tailed end of a distribution for the sample mean across sample size. This suggests EBC_a may be able to provide relatively accurate results at the α = 0.01 significance level when the data is skewed for small sample sizes. However, at sample sizes 30 and 40 at the α = 0.01 significance level for the upper limit, EBC_a had an error rate with one of the largest percent errors compared to all of the other methods studied. Although EBC_a did perform relatively accurately at these small sample sizes at α = 0.01 significance level, it was unable to maintain the same degree of relative accuracy at the α = 0.05 and α = 0.10 significance levels.

For the ratio of sample means statistic when data was generated from the normal distribution with an expected ratio of means value of 2, E-skew performed relatively accurately. In particular, E-skew performed more accurately at nearly every sample size and significance level when compared to other methods applied on EBSD(n) for this statistic. When comparing E-skew to both Monte Carlo Bootstrap and EBSD(n) for this specification, of the 36 cases studied (6 sample sizes*2 limit ends*3 significance levels), E-skew had the error rate with the smallest percent error for 23 of them. Similar results for E-skew were found for the other two normal distribution specifications studied. This suggests E-skew provides relatively accurate confidence intervals for the ratio of means statistic when the data from each independent sample is normally distributed.

The E-skew error rate results for data generated from the log-normal distribution were found to be similar to the error rate results from the normal distribution, in fact for the simulation performed on the log-normally distributed data, E-skew had the error rate with the smallest percent error at every sample size for the α = 0.05 and α = 0.10 significance levels.

For the ratio of sample means statistic when data was generated from the exponential distribution with an expected ratio of means value of 2, E-skew performed relatively more accurately than other methods applied on EBSD(n) at the α = 0.05 and α = 0.10 significance levels. Further at the α = 0.10 significance level, E-skew performed relatively more accurately than all methods it was compared to. In this case, E-skew achieved the error rate with the smallest percent error for 3 of 6 sample sizes for the upper limit and 4 of 6 sample sizes for the lower limit. At the α = 0.01 significance level, EP performed relatively more accurately than E-skew. Similar E-skew results were found for the ratio of means statistic for both exponential distribution parameter specifications. Additionally, similar E-skew results were also found for data generated from the gamma distribution and the mixture of two normal distributions. For the gamma distribution, E-skew performed relatively accurately for the upper limit at the α = 0.05 and α = 0.10 significance levels, attaining the error rate with the smallest percent error in 10 of 12 cases. For the mixture of two normal distributions, at the α = 0.05 and α = 0.10

significance levels E-skew attained the error rate with the smallest percent error among all methods applied on EBSD(n) for both the upper and lower limit.

For the ratio of sample means statistic, EBC and EP performed more accurately at the α = 0.01 significance level at each sample size compared to their Monte Carlo Bootstrap counterpart. They also performed less accurately compared to their Monte Carlo Bootstrap counterpart at each sample size for the α = 0.05 and α = 0.10 significance levels for each distribution studied.

For the Pearson correlation coefficient, E-skew performed relatively accurately when the data was generated from the bivariate normal distribution. For ρ = 0.10, E-skew had the error rate with the smallest percent error compared to any other method at 3 of 6 sample sizes for the upper limit and 4 of 6 sample sizes for the lower limit at the α = 0.01 significance level. At the α = 0.05 and α = 0.10 significance levels, E-skew attained the error rate with the smallest percent for an even higher percentage of sample sizes. As ρ was increased E-skew attained the error rate with the smallest percent error less frequently but still did so at many sample sizes studied. This suggests when computing confidence intervals for the Pearson correlation coefficient, E-skew may be relatively accurate compared to Monte Carlo Bootstrap methods when the data is generated from a distribution that is approximately bivariate normal.

When compared to other methods applied on EBSD(n), E-skew performed more accurately at every sample size at the α = 0.05 and α = 0.10 significance levels. EBC, EP and ET did not perform as accurately at each sample size and significance level compared to their Monte Carlo Bootstrap counterpart.

When the data was non-normally distributed, E-skew did not perform as accurately for the Pearson correlation coefficient as the Monte Carlo Bootstrap methods for most sample sizes. In a few cases at the α = 0.05 and α = 0.10 significance levels, E-skew did attain the error rate with the smallest percent error when the data was non-normally distributed and ρ = 0.10.

For the trimmed sample mean for data that was normally distributed, EP or EBC attained the error rate with the smallest percent error at every sample size compared to the other methods studied including Monte Carlo Bootstrap methods at the α = 0.01 significance level. However, at the α = 0.05 and α = 0.10 significance levels both methods performed less accurately than their Monte Carlo Bootstrap counterpart. This suggests when computing confidence intervals at the α = 0.01 significance level on data where outliers have been removed, it may be preferable to use EP or EBC over any Monte Carlo Bootstrap method. However, when the data was slightly skewed as in the log-normal and gamma cases EP or EBC attained the error rate with the smallest percent error for the upper limit at three of four sample sizes. When the data was moderately skewed as in the case of the exponential distribution neither EP or EBC attained the error rate with the smallest percent error for the upper limit at any sample size. This suggests as the data becomes more skewed EP and EBC become relatively less accurate compared to Monte Carlo Bootstrap methods for the trimmed sample mean. At the α = 0.05 and α = 0.10 significance levels E-skew performed relatively more accurately than every other method applied on $\text{EBSD}(n)$ and second most accurately to BS in nearly every case. Again in this instance, E-skew provided a reasonable facsimile to BS in terms of accuracy while also having the advantages inherent in using $\text{EBSD}(n)$.

For the sample median, percentile methods applied on EBSD(n) performed more accurately than their Monte Carlo Bootstrap counterparts at the α = 0.01 significance level, and less accurately at the α = 0.05 and α = 0.10 significance levels. For some sample sizes at the α = 0.01 significance level, percentile methods applied on EBSD(n) had error rates with smaller percent errors than every Monte Carlo Bootstrap method.

The main simulation results were studied using 10,000 Bootstrap resamples. Some methods using EBSD(n) that performed less accurately compared to Monte Carlo Bootstrap methods could be found to be more accurate if the number of Bootstrap iterations specified was decreased. For example, for the sample mean examples in section 4.6, BT which was found to have an error rate with a smaller percent error compared to ET for the upper limit for 10,000 Bootstrap iterations, frequently had a larger percent error when fewer Bootstrap iterations were used. Therefore, in the simulation study the Monte Carlo Bootstrap methods were compared to EBSD(n) using a best case approach. In practice fewer Monte Carlo Bootstrap iterations are often specified and this would lead to a further comparative advantage for EBSD(n) methods.

Lastly, in the real data example in section 4.7, although the results varied depending on the alpha level, sample size and data set under consideration frequently E-skew was found to be in statistically significant agreement with many of the Bootstrap methods. Further the results show E-skew had high agreement with second order methods BC_a and BS at sample size 10 and 14 in nearly every case. As the microarray data was skewed, one can refer back to the results for data simulated from the exponential distribution as a basis for comparison. For this part of the simulation study, the BC_a and BS methods provided the most accurate results for the lower limit and upper limit respectively. Since E-skew is in highest agreement with these methods at sample sizes 10 and 14 this is an indication E-skew is performing accurately in a real data context. Thus by using EBSD(n), E-skew can derive consistent accurate results which is motivation to choose the method when analyzing skewed data.

Limitations of E-skew

For the sample mean, when the data was highly skewed as in the second log-normal specification studied in section 4.1, E-skew was not able to adjust for skew as successfully as the BC_{a} and BS methods. For moderately skewed distributions like the exponential, E-skew adjusted for skew as well or nearly as well as these methods. The more dramatic the skew, the less relatively accurate E-skew was compared to these methods.

Additionally, for non-linear statistics, E-skew is dependent on using the Jackknife method to estimate the skew for each Bootstrap sample that makes up EBSD(n). As described in section 2.4 the Jackknife is inaccurate for large sample sizes for the sample median. Therefore, E-skew will also not perform accurately for the sample median at large sample sizes as it is dependent on the Jackknife method. The same, in theory, could be said for other statistics the Jackknife performs inaccurately for. Additionally, in section 2.4 it is mentioned that the Jackknife does not perform as accurately for more complex probability structures. Therefore, this would also be a limitation for E-skew as well.

Limitations of Other Monte Carlo Bootstrap methods Applied on EBSD($\boldsymbol{n}$)

Generally speaking, EBC, EP and EBC_{a} performed relatively less accurately when the α significance level was greater than 0.01 for the sample sizes studied. Additionally, at the α = 0.01 significance level, EBC_{a} generally performed relatively accurately for the upper limit at sample sizes 10, 15 and 20 and then relatively inaccurately for the upper limit at sample sizes 30 and 40. For the BC_a method it is uncommon for the upper limit error rate's percent error to be larger at sample sizes 30 and 40 than it is at sample size 20 when data is drawn from the same population across sample size. However, this was found to happen consistently for the EBC_a method.

In the construction of EBSD(n), as sample size increased, the proportion of samples from EBSD(n) with $\bar{X}_{E_i}$ equal to $\bar{X}$ from the original sample increased. At sample sizes 30 and 40, approximately half of the samples EBSD(n) produced had a $\bar{X}_{E_i}$ equal to the $\bar{X}$ from the original sample.

This property causes the percent error of BC_a's upper limit error rate to increase greatly. The $\widehat{z_0}$ value in the BC_a algorithm is overly sensitive to samples whose $\bar{X}_{E_i}$ is equal to the $\bar{X}$ from the original sample because of the $\bar{X}_{E_i} < \bar{X}$ inequality. Therefore when the BC_a algorithm is applied on EBSD(n), EBC_a, the strictly less than inequality causes the $\widehat{z_0}$ value and consequently the percentile taken from the EBSD(n) sampling distribution to be much smaller for the upper limit than the percentile that would be taken using the Monte Carlo Bootstrap method at sample size 30 and 40. This causes the resulting confidence interval to be narrower and therefore excludes the true population mean at a higher frequency compared to other methods at these sample sizes.

The ES method did not compare favorably to the BS method. For each sample size at each significance level the ES method had an error rate with a larger percent error than BS for both the upper and lower limit. Both methods used the Bootstrap-t algorithm described in section 2.5. This algorithm used t-statistics generated from the Bootstrap sampling distribution to compute a confidence interval. The reason ES did not perform as well as BS is because of the repeated blocks inherent to the EBSD(n) design. In the Monte Carlo Bootstrap, the probability of the same element being picked n times with replacement in an individual resample is $n * (\frac{1}{n})^n$. For example, for a sample of size 30 one way a resample would have a sample variance of 0 would be if the first element were picked 30 consecutive times when sampling with replacement. Of course, this is not exclusive to the first element, if any of the n elements are chosen every time for a resample this will cause the variance to be 0. If the sample variance is 0, the t-statistic is undefined in the resulting Monte Carlo Bootstrap sampling distribution. For the Monte Carlo Bootstrap, the probability of this occurring once in 10,000 Bootstrap resamples is unlikely even at sample sizes as small as 10 (This probability can be computed as: (1-(1-(10*(($\frac{1}{10}$)^{10})))^{10,000})≈0.00001). Conversely, EBSD(n) guarantees $\frac{2n}{4n^2+1}$ samples will have this condition. Accordingly, a high percentage of the t-statistic sampling distribution is undefined. At sample size 10 approximately 5% of the sample sizes will not be usable for establishing a distribution for the t-statistic. Further at sample size 40, still a little more than 1% of the samples will not be usable. This had an impact on the percentile to be used from the EBSD(n) sampling distribution for a significance level α. It is unclear how to proceed using the Bootstrap-t because although the samples were unusable by removing them from the sampling distribution information would be lost.

The problems inherent with the Bootstrap-t method might cause one to wonder why doesn't the E-skew method suffer from the same problem when implemented on EBSD(n)? If the t-statistic cannot be computed for the repeated samples and the E-skew method uses the skew corrected t-statistic, then this should cause these samples to be thrown out of the E-skew sampling distribution? The difference was because upper and lower limits were being computed rather than t-statistics. The E-skew method automatically assigned the value of the repeated element as the upper and lower limits because there is no variance in these samples. These elements were distributed across the sampling distribution rather than being concentrated at the center of the sampling distribution.

An automatic assignment was applied for the Bootstrap-t, the t-statistic was assigned the value of 0. However, because the repeated elements samples end up concentrated near the 50^{th} percentile of the resulting EBSD(n) sampling distribution, information was lost as these t-statistics were not distributed across the sampling distribution. Therefore, the results for the Bootstrap-t algorithm using EBSD(n) was quite inaccurate relative to the other methods considered.

Future Work

There are many interesting potential investigations that could be performed for the E-skew and EBSD(n). First, further investigation would be merited in finding a way to adjust the E-skew algorithm to perform relatively as accurately as BS and BC_a in the case of highly skewed data. The component of interest that could be investigated for adjustment is the computation of the α_U or α_L parameters. The reason this is specifically of interest is because there is potentially a modification that could be made here so that the percentile chosen from the upper limit (or lower limit) sampling distribution would be higher (or lower) and thus provide better coverage for data generated from highly skewed distributions.

Further investigation would also be merited into EBSD(n) and specifically the application of EBC_a on EBSD(n). The fact that EBC_a performed relatively accurately compared to BS at the α = 0.01 significance level for the upper limit for sample sizes 10, 15 and 20 is quite promising. Potentially a slight modification to the BC_a algorithm when applied on EBSD(n) could lead to highly accurate results at sample size 30 and 40.

REFERENCES

1. Srivastav SS, Srivastav AS. Efficiency balanced designs for bootstrap simulations. Communications in Statistics-Theory and Methods. (2020) February.

2. Hall, P. (1988a). "Theoretical comparison of bootstrap confidence intervals." (With discussion) Ann. Statist. 16, 927-985.

3. Hesterberg, T. "What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum." arXiv:1411.5279. 2014.

4. Davidson A., Hinkley D., "Improved Calculation" "Bootstrap Methods and their Applications." Cambridge University Press, (1997) 438-446.

5. Efron B., Tibshirani R., "The empirical distribution function and the plug-in principle". "An Introduction to the Bootstrap". Chapman and Hall, New York, NY (1993) 31-37.

6. Scholz F., "The Bootstrap as a Bias Reduction Tool". "The Bootstrap Small Sample Properties." bcstech-93-051, Boeing Computer Services, Research and Technology (2007) 10 -15.

7. Efron B., Tibshirani R., "An overview of nonparametric and parametric inference". "An Introduction to the Bootstrap". Chapman and Hall, New York, NY (1993) 296-320.

8. Efron B., Tibshirani R., "Estimates of Bias". "An Introduction to the Bootstrap." Chapman and Hall, New York, NY (1993) 124-139.

9. Efron B., Tibshirani R., "A geometrical representation for the bootstrap and jackknife". "An Introduction to the Bootstrap". Chapman and Hall, New York, NY (1993) 283-295.

10. Graham R.L., Hinkley D.V., John P. M. W., Shi S. (1990). "Balanced Design of Bootstrap Simulations." Journal of the Royal Statistical Society: Series B (Methodological), Volume 52, Issue 1, 185–202.

11. Scholz F., “Variance Estimation”. “The Bootstrap Small Sample Properties.” bcstech-93-051, Boeing Computer Services, Research and Technology (2007) 16 - 18.

12. Efron B., Tibshirani R., “The jackknife”. “An Introduction to the Bootstrap”. Chapman and Hall, New York, NY (1993) 141-150.

13. Efron B., Tibshirani R., “Assessing the error in bootstrap estimates”. “An Introduction to the Bootstrap”. Chapman and Hall, New York, NY (1993) 271-281.

14. Scholz F., “Bootstrap Confidence Bounds “The Bootstrap Small Sample Properties.” bcstech-93-051, Boeing Computer Services, Research and Technology, (2007) 41 -46.

15. Efron, B. (1987). “Better bootstrap confidence intervals.” (With discussion) J. Amer. Statist. Assoc. 82, 171-200.

16. Efron B., Tibshirani R., “Better bootstrap confidence intervals”. “An Introduction to the Bootstrap”. Chapman and Hall, New York, NY (1993) 178-199.