8-2008

Interval Estimation for the Ratio of Percentiles from Two Independent Populations.

Pius Matheka Muindi

East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd

Part of the Statistical Theory Commons

Recommended Citation

This Thesis - Open Access is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.
Interval Estimation for the Ratio of Percentiles from Two Independent Populations

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

by

Pius Muindi

August 2008

Robert Price, Ph.D., Chair
Robert Gardner, Ph.D.
Edith Seier, Ph.D.
Yali Liu, Ph.D

Keywords: confidence intervals (CI), ratio of percentiles, percentiles.
ABSTRACT

Interval Estimation for the Ratio of Percentiles from two Independent Populations

by

Pius Muindi

Percentiles are used everyday in descriptive statistics and data analysis. In real life, many quantities are normally distributed and normal percentiles are often used to describe those quantities. In life sciences, distributions like exponential, uniform, Weibull and many others are used to model rates, claims, pensions etc. The need to compare two or more independent populations can arise in data analysis. The ratio of percentiles is just one of the many ways of comparing populations. This thesis constructs a large sample confidence interval for the ratio of percentiles whose underlying distributions are known. A simulation study is conducted to evaluate the coverage probability of the proposed interval method. The distributions that are considered in this thesis are the normal, uniform and exponential distributions.
DEDICATION

To my wife Susan.
ACKNOWLEDGMENTS

I wish to acknowledge all the graduate students and staff members of the Department of Mathematics at East Tennessee State University. I would like to thank Dr. Edith Seier and Dr. Yali Liu for their input and support. Special thanks goes to Dr. Robert Gardner who has made my study at ETSU successful and enjoyable.

Most importantly I wish to acknowledge Dr. Robert Price, my advisor, who has greatly contributed to my interest and knowledge in statistics and without whom this project would have been almost impossible.

Lastly I wish to thank my wife and my family for their moral and financial support.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>4</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>5</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>10</td>
</tr>
<tr>
<td>1.1 Maximum Likelihood Estimators (MLEs)</td>
<td>12</td>
</tr>
<tr>
<td>2 NORMAL DISTRIBUTION</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Unbiased Estimators of $\mu$ and $\sigma$</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Approximate Confidence Interval for the Ratio of Percentiles from Two Independent Normal Distributions</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Simulation Results</td>
<td>21</td>
</tr>
<tr>
<td>3 EXPONENTIAL DISTRIBUTION</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Unbiased Estimator of $\theta$</td>
<td>24</td>
</tr>
<tr>
<td>3.2 Approximate Confidence Interval for the Ratio of Percentiles from Two Independent Exponential Distributions</td>
<td>25</td>
</tr>
<tr>
<td>3.3 Simulation Results</td>
<td>28</td>
</tr>
<tr>
<td>4 UNIFORM DISTRIBUTION</td>
<td>31</td>
</tr>
<tr>
<td>4.1 Unbiased Estimators of $a$ and $b$</td>
<td>31</td>
</tr>
<tr>
<td>4.2 Approximate Confidence Interval for the Ratio of Percentiles from Two Independent Uniform Distributions</td>
<td>38</td>
</tr>
<tr>
<td>4.3 Simulation Results</td>
<td>42</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1  Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Normal Populations. . . . . . . 22
2  Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Exponential populations. . . . . 29
3  Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Uniform populations. . . . . . . 43
LIST OF FIGURES

1 Coverage Rates vs Sample Size for Ratio of Normal Percentiles at $p = 0.10$ 23

2 Coverage Rates vs Sample Size for Ratio of Exponential Percentiles at $p = 0.10$ 30

3 Coverage Rates vs Sample Size for Ratio of Uniform Percentiles at $p = 0.10$ 44
1 INTRODUCTION

Percentiles or quantiles are very common in statistics. Percentiles are used to determine the value of an observation with a given percentage below it. Percentiles are used as a measure of central tendency as well as a measure of spread. The median, which is one of the most used measures of central tendency, refers to the 50th percentile. The lower and upper quartiles which refer to the 25th and the 75th percentiles, respectively, are commonly used as measures of dispersion. The five-number summary used in making the famous box plot is basically a summary of specific percentiles of a distribution namely 1st, 25th, 50th, 75th and 100th. Quantiles are also used in measuring reliability by determining the time to failure, survival and hazard functions [3]. Percentiles are also used in hydrology and in statistical process control [3].

The need to compare two distributions arises and the researcher will require some particular parameter as a point of comparison. In many instances researchers are compelled to use the mean as the point of comparison with the assumption that the mean is the most reliable parameter for describing the population. This is not always the case and sometimes the median may be more reliable when the distribution is strongly skewed [7]. The researcher may also be interested in other percentiles of the population and if there are two independent populations a comparison of the percentiles may be of value.

Approximate distribution-free intervals for both the difference and ratio of medians have been developed [7]. We consider interval estimation for the ratio of percentiles when the underlying distributions are known.
We begin by defining a few of the terminologies used in this thesis. A percentile or a quantile is the value of a variable below which a certain percentage of observations fall. A percentile can also be defined as one of the 99 point scores that divide a ranked distribution into groups, each of which contain 1/100 of the scores. A Parameter is a number that describes the population. In statistical practice, the value of a parameter is usually not known because it is difficult to examine the entire population. A Statistic is a number that can be computed from a sample without making use of any unknown parameters. In practice, we often use a statistic to estimate an unknown parameter. If the random variables $X_1, X_2, \ldots, X_n$ are arranged in ascending order of magnitude and then written as $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, we call $X_{(i)}$ the $i$th order statistic $[2]$ ($i = 1, 2, \ldots, n$). If $E[u(X_1, X_2, ..., X_n)] = \theta$, the statistic $u(X_1, X_2, ..., X_n)$ is called an unbiased estimator of $\theta$ $[5]$.

**Theorem 1.1** $[6]$ Draw a simple random sample of size $n$ from any population with mean $\mu$ and standard deviation $\sigma$. When $n$ is large, the sampling distribution of the sample mean $\bar{x}$ is approximately normal with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$. This is referred to as the central limit theorem.

**Theorem 1.2** $[4]$ Let $X_n$ be a sequence of random variables such that $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Suppose the function $f(x)$ is differentiable at $\theta$ and $f'(\theta) \neq 0$. Then $\sqrt{n}(f(X_n) - f(\theta)) \xrightarrow{D} N(0, \sigma^2(f'(\theta))^2)$ where $\xrightarrow{D}$ denotes converge in distribution. This is referred to as the Delta Method.
Theorem 1.3 [4] Let \( \{X_n\} \) be a sequence of \( p \)-dimensional vectors. Suppose 
\[
\sqrt{n}(X_n - \mu_0) \xrightarrow{D} N_p(0, \Sigma).
\]
Let \( g \) be a transformation \( g(x) = (g_1(x), \ldots, g_k(x))' \) such that \( 1 \leq k \leq p \) and the \( k \times p \) matrix of partial derivatives,
\[
B = \left[ \frac{d g_i}{d \mu_j} \right] \quad i = 1, \ldots, k; \ j = 1, \ldots, p,
\]
are continuous and do not vanish in a neighborhood of \( \mu_0 \). Let \( B_0 = B \) at \( \mu_0 \). Then
\[
\sqrt{n}(g(X_n) - g(\mu_0)) \xrightarrow{D} N_k(0, B_0 \Sigma B_0').
\]
This is an extension of the Delta Method.

1.1 Maximum Likelihood Estimators (MLEs)

Let \( X_1, X_2, \ldots, X_n \) be a random sample from a distribution that depends on one or more unknown parameters \( \theta_1, \theta_2, \ldots, \theta_m \) with p.m.f. or p.d.f. denoted by \( f(x; \theta_1, \theta_2, \ldots, \theta_m) \). Suppose that \( \theta_1, \theta_2, \ldots, \theta_m \) is restricted to a given parameter space \( \Omega \). Then the joint p.m.f. or p.d.f. of \( X_1, X_2, \ldots, X_n \) namely
\[
L(\theta_1, \theta_2, \ldots, \theta_m) = f(x_1; \theta_1, \ldots, \theta_m) f(x_2; \theta_1, \ldots, \theta_m) \cdots f(x_n; \theta_1, \ldots, \theta_m),
\]
(\( \theta_1, \theta_2, \ldots, \theta_m \) \( \in \) \( \Omega \)), when regarded as a function of \( \theta_1, \theta_2, \ldots, \theta_m \) is called the likelihood function. Say 
\[
[u_1(X_1, X_2, \ldots, X_n), u_2(X_1, X_2, \ldots, X_n), \ldots, u_m(X_1, X_2, \ldots, X_n)]
\]
is that \( m \)-tuple in \( \Omega \) that maximizes \( L(\theta_1, \theta_2, \ldots, \theta_m) \). Then
\[ \hat{\theta}_1 = u_1(X_1, X_2, \ldots, X_n) \]
\[ \hat{\theta}_2 = u_2(X_1, X_2, \ldots, X_n) \]
\[ \vdots \]
\[ \hat{\theta}_m = u_1(X_1, X_2, \ldots, X_n) \]

are **maximum likelihood estimators** of \( \theta_1, \theta_2, \ldots, \theta_m \) respectively; and the corresponding values of these statistics, namely
\[ u_1(X_1, X_2, \ldots, X_n), u_2(X_1, X_2, \ldots, X_n), \ldots u_m(X_1, X_2, \ldots, X_n) \]

are called **maximum likelihood estimates** [5]. In many practical cases these estimators (and estimates) are unique.

For many applications there is just one unknown parameter. In these cases, the likelihood function is given by

\[ L(\theta) = \prod_{i=1}^{n} f(x_i; \theta). \]

Let \( X_1, X_2, \ldots, X_n \) be continuous random variables from a population whose underlying distribution is known and parameters unknown. Let \( k_p \) denote the 100\( p \)th percentile. \( k_p \) can be expressed in terms of the population parameters.

By way of maximum likelihood estimation, we can approximate the unknown population parameters using statistics obtained from a sample drawn from the population whose underlying distribution is known. Then \( k_p \) can be approximated using those estimates.
Let $k_{p,x}$ be the 100pth percentile from the first population ($X$) and let $k_{p,y}$ be
the 100pth percentile from the second population ($Y$). Denote the percentile ratio
as $k_p \text{ ratio}$. Then $k_p \text{ ratio}$ is given by $k_{p,x}/k_{p,y}$. We next discuss an approximate
$(1 - \alpha)100\%$ confidence interval for the $k_p \text{ ratio}$. To construct a confidence interval for
$k_p \text{ ratio}$ we will estimate the natural log of $k_p \text{ ratio}$ and the variance of the estimated
natural log of $k_p \text{ ratio}$ and then exponentiate the end points of the interval. This is
based on the delta method.
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with mean $\mu$ and variance $\sigma^2$. Then the p.d.f. of $X$ is given by

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (1)$$

Let $k_p$ denote the $100p$th percentile, then

$$k_p = \mu + Z_p \sigma \quad (2)$$

where $Z_p$ denotes the $100p$th percentile of the standard normal distribution $N(0,1)$ [1]. Since $\mu$ and $\sigma$ are unknown, we need to estimate $k_p$. This can be done using the minimum variance unbiased estimators of $\mu$ and $\sigma$.

### 2.1 Unbiased Estimators of $\mu$ and $\sigma$

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population $N(\mu, \sigma^2)$ and let $\bar{X}$ denote the sample mean and $S^2$ denote the sample variance where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i , \quad (3)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 , \quad (4)$$

and

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} \quad (i = 1, 2, \ldots, n). \quad (5)$$

$\bar{X}$ is the maximum likelihood estimator of $\mu$ and by the central limit theorem $\bar{X}$ is an unbiased estimator of $\mu$ and $E(\bar{X}) = \mu$. 

15
Lemma 2.1 [5] Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a normal distribution \( N(\mu, \sigma^2) \). Then the distribution of \( (n-1)S^2/\sigma^2 \) is \( \chi^2(n-1) \), where \( \chi^2(n-1) \) is a Chi-square distribution with \( n-1 \) degrees of freedom.

Proposition 2.2 \( S^2 \) is an unbiased estimator of \( \sigma^2 \).

Proof

We can use Lemma 2.1 to show that \( S^2 \) is an unbiased estimator of \( \sigma^2 \). By Lemma 2.1, the distribution of \( (n-1)S^2/\sigma^2 \) is \( \chi^2(n-1) \). Therefore

\[
E \left( \frac{(n-1)S^2}{\sigma^2} \right) = E \left( \chi^2(n-1) \right)
\]

\[
\frac{n-1}{\sigma^2} E(S^2) = n - 1
\]

\[
E(S^2) = \sigma^2.
\]

However, \( S \) is not an unbiased estimator of \( \sigma \). Thus, we need to find a constant \( C \) that can be used to remove the biasness such that \( E(CS) = CE(S) = \sigma \). We know from Lemma 2.1 \( \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \). So, \( \frac{\sqrt{n-1}S}{\sigma} \sim \sqrt{\chi^2(n-1)} \). Next we need to find the p.d.f. of \( Y = \sqrt{\chi^2(n-1)} \). Suppose \( f(x) \) and \( g(y) \) are p.d.f.’s of \( \chi^2(n-1) \) and \( \sqrt{\chi^2(n-1)} \) respectively. Then

\[
f(x) = \frac{1}{\Gamma \left( \frac{n-1}{2} \right) 2^{\frac{n-1}{2}}} x^{\frac{n-1}{2} - 1} e^{-\frac{x}{2}}, \quad 0 \leq x < \infty.
\]

By change of variable technique we have \( y = \sqrt{x} \Rightarrow x = y^2 \) and \( \frac{dx}{dy} = 2y \). Thus

\[
g(y) = \frac{1}{\Gamma \left( \frac{n-1}{2} \right) 2^{\frac{n-1}{2}}} (y^2)^{\frac{n-1}{2} - 1} e^{-\frac{(y^2)}{2}} 2y, \quad 0 \leq y < \infty.
\]
Now,

\[ E(Y) = \int_0^\infty \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} y^{n-1} e^{-\frac{y^2}{2}} \, dy \]

\[ = \int_0^\infty (y^2)^\frac{1}{2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} y^{n/2} e^{-\frac{y^2}{2}} \, dy \]

\[ = \int_0^\infty \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} y^{n/2} e^{-\frac{y^2}{2}} \, dy . \]

Letting \( t = y^2, \, dt = 2y \, dy \) and we obtain,

\[ E(Y) = \int_0^\infty \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} t^{n/2} e^{-\frac{t}{2}} \, dt \]

\[ = \int_0^\infty \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} t^{n/2-1} e^{-\frac{t}{2}} \, dt \]

\[ = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} \int_0^\infty t^{n/2-1} e^{-\frac{t}{2}} \, dt \]

\[ = \frac{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}{\Gamma\left(\frac{n-1}{2}\right)2^{n/2}} \int_0^\infty \frac{t^{n/2-1} e^{-\frac{t}{2}} \, dt}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} . \]

Hence,

\[ E(Y) = \frac{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)} . \] (8)

Since \( Y = \sqrt{\chi^2(n-1)} \sim \frac{\sqrt{n-1}}{\sigma} S \), \( E(Y) = \frac{\sqrt{n-1}E(S)}{\sigma} \) and therefore,

\[ E(S) = \frac{\sigma E(Y)}{\sqrt{n-1}} \]

\[ = \frac{\sigma}{\sqrt{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)} \]

\[ = \frac{\sigma}{\sqrt{\frac{n-1}{2}}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \]

\[ = \frac{\sigma}{\sqrt{\frac{n-1}{2}}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \]. (9)
Hence, we have
\[
\frac{\sqrt{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} E(S) = \sigma \quad \Rightarrow \quad C = \frac{\sqrt{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \quad (10)
\]
and
\[
E(CS) = \sigma . \quad (11)
\]
Thus an unbiased estimator of \( k_p \) is
\[
\hat{k}_p = \bar{X} + Z_p CS \quad [1]. \quad (12)
\]
To compute confidence intervals we need the variance and the estimated variance of the statistic of interest. In this case we require the variance of \( \hat{k}_p \). Therefore
\[
Var(\hat{k}_p) = Var(\bar{X} + Z_p CS)
\]
\[
= Var(\bar{X}) + (CZ_p)^2 Var(S)
\]
\[
= \frac{\sigma^2}{n} + C^2 Z_p^2 \left[ E(S^2) - (E(S))^2 \right]
\]
\[
= \frac{\sigma^2}{n} + C^2 Z_p^2 \left[ \sigma^2 - \frac{\sigma^2}{C^2} \right] \quad \text{by proposition 2.2 and (11)}
\]
\[
= \frac{\sigma^2}{n} \left( 1 + nC^2 Z_p^2 (1 - \frac{1}{C^2}) \right)
\]
\[
= \frac{\sigma^2}{n} \left( 1 + nZ_p^2 (C^2 - 1) \right) \quad [1]. \quad (13)
\]
Thus the estimated variance of \( \hat{k}_p \) denoted \( \widehat{Var}(\hat{k}_p) \) is
\[
\widehat{Var}(\hat{k}_p) = \left( \frac{S^2}{n} \left( 1 + nZ_p^2 (C^2 - 1) \right) \right). \quad (14)
\]
2.2 Approximate Confidence Interval for the Ratio of Percentiles from Two Independent Normal Distributions

In this section we are going to develop a method of computing the \((1 - \alpha)100\%\) confidence intervals for the ratio of two normal percentiles. These two percentiles must come from two independent normal distributions and they must be the same 100\textit{pth} percentile. For example, if we use the 45\textit{th} percentile from the first distribution then we must use the 45\textit{th} percentile from the second distribution. The means, variances, and sample sizes of the two samples need not be equal.

**Theorem 2.3** Let \(X_1, X_2, \ldots, X_n\) be a random sample of size \(n\) from a normal population \(X\) with mean \(\mu_x\) and variance \(\sigma_x^2\) and let \(Y_1, Y_2, \ldots, Y_m\) be a random sample of size \(m\) from another normal population \(Y\) with mean \(\mu_y\) and variance \(\sigma_y^2\) where \(X\) and \(Y\) are independent. Let \(k_{p,x}\) and \(k_{p,y}\) be the 100\textit{pth} percentiles from populations \(X\) and \(Y\), respectively. Then an approximate \((1 - \alpha)100\%\) confidence interval for the ratio of the percentiles \(k_{p,x}\) and \(k_{p,y}\) is given by

\[
\frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \exp \left( \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \right). \tag{15}
\]

**Proof**

We know that

\[
k_{p,x} = \mu_x + Z_p \sigma_x \quad \text{and} \quad k_{p,y} = \mu_y + Z_p \sigma_y
\]

and the unbiased estimators of \(k_{p,x}\) and \(k_{p,y}\) are

\[
\hat{k}_{p,x} = \bar{X} + Z_p C_x S_x \quad \text{and} \quad \hat{k}_{p,y} = \bar{Y} + Z_p C_y S_y
\]

(16)
respectively, where

\[ C_x = \sqrt{\frac{n-1}{2} \Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \quad \text{and} \quad C_y = \sqrt{\frac{m-1}{2} \Gamma\left(\frac{m-1}{2}\right)} \frac{1}{\Gamma\left(\frac{m}{2}\right)}. \]  

(18)

The variances of these unbiased estimators \( \hat{k}_{p,x} \) and \( \hat{k}_{p,y} \) are

\[ \text{Var}(\hat{k}_{p,x}) = \frac{\sigma_x^2}{n} \left(1 + nZ_p^2(C_x^2 - 1)\right) \quad \text{and} \quad \text{Var}(\hat{k}_{p,y}) = \frac{\sigma_y^2}{m} \left(1 + mZ_p^2(C_y^2 - 1)\right) \]  

(19)

and their respective estimated variances are

\[ \hat{\text{Var}}(\hat{k}_{p,x}) = \frac{S_x^2}{n} \left(1 + nZ_p^2(C_x^2 - 1)\right) \quad \text{and} \quad \hat{\text{Var}}(\hat{k}_{p,y}) = \frac{S_y^2}{m} \left(1 + mZ_p^2(C_y^2 - 1)\right). \]  

(20)

The estimated \( k_p \) ratio denoted by \( \hat{k}_p \) ratio is given by

\[ \hat{k}_p \text{ ratio} = \frac{\hat{k}_{p,x}}{\hat{k}_{p,y}}. \]  

(21)

Now we need to find the variance of the \( \hat{k}_p \) ratio. This can be simplified by introducing natural logarithms and using the delta method to compute the variance.

\[ \ln(\hat{k}_p \text{ ratio}) = \ln \left(\frac{\hat{k}_{p,x}}{\hat{k}_{p,y}}\right) = \ln(\hat{k}_{p,x}) - \ln(\hat{k}_{p,y}) \]  

(22)

The next few equations employ the delta method to compute the variance of \( \ln(\hat{k}_p \text{ ratio}) \).

\[ \text{Var}(\ln(\hat{k}_p \text{ ratio})) = \]  

\[ = \begin{bmatrix} \frac{d \ln(\hat{k}_p \text{ ratio})}{d \hat{k}_{p,x}} & \frac{d \ln(\hat{k}_p \text{ ratio})}{d \hat{k}_{p,y}} \end{bmatrix} \begin{bmatrix} \text{Var}(\hat{k}_{p,x}) & 0 \\ 0 & \text{Var}(\hat{k}_{p,y}) \end{bmatrix} \begin{bmatrix} \frac{d \ln(\hat{k}_p \text{ ratio})}{d \hat{k}_{p,x}} \\ \frac{d \ln(\hat{k}_p \text{ ratio})}{d \hat{k}_{p,y}} \end{bmatrix}. \]

The variance covariance matrix has zero entries because X and Y are independent and their covariance is zero.

\[ \text{Var}(\ln(\hat{k}_p \text{ ratio})) = \begin{bmatrix} \frac{1}{\hat{k}_{p,x}} & -\frac{1}{\hat{k}_{p,y}} \end{bmatrix} \begin{bmatrix} \text{Var}(\hat{k}_{p,x}) & 0 \\ 0 & \text{Var}(\hat{k}_{p,y}) \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{k}_{p,x}} \\ -\frac{1}{\hat{k}_{p,y}} \end{bmatrix}. \]  

(23)
Hence

\[ \text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2} \]  \hspace{1cm} (24) 

and the estimated variance of the ln(\(\hat{k}_p \text{ ratio}\)) is

\[ \text{Var} \ln(\hat{k}_p \text{ ratio}) = \frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2} \]  \hspace{1cm} (25) 

The (\(1 - \alpha\))\% confidence interval for the ln(\(k_p \text{ ratio}\)) can be computed as

\[ \ln(\hat{k}_{p,x}) - \ln(\hat{k}_{p,y}) \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \]  \hspace{1cm} (26) 

and exponentiating the expression we have the (\(1 - \alpha\))\% confidence interval for the \(k_p \text{ ratio}\) as

\[ \frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \exp \left( \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \right) \]  \hspace{1cm} □

2.3 Simulation Results

To test this method, a simulation was done. This simulation involved a random generation of two normally distributed samples using the R statistical software [8]. The parameters \(\mu\) and \(\sigma^2\) used to generate first sample were fixed at 30 and 2, respectively, and those of the second sample were fixed at 20 and 3 respectively. The simulation process involved 50000 runs for each different set of sample sizes and the resulting empirical coverages rates were recorded. Table 1 on next page shows some of the simulation results. The empirical coverage can be seen to converge to 0.95 in Figure 1.
Table 1: Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Normal Populations.

<table>
<thead>
<tr>
<th>percentiles</th>
<th>n</th>
<th>m</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.8789</td>
<td>0.9169</td>
<td>0.9568</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8972</td>
<td>0.9455</td>
<td>0.9842</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8826</td>
<td>0.9292</td>
<td>0.9682</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8973</td>
<td>0.9467</td>
<td>0.9855</td>
</tr>
<tr>
<td>$p = 0.10$</td>
<td>100</td>
<td>100</td>
<td>0.8987</td>
<td>0.9483</td>
<td>0.9875</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8995</td>
<td>0.9492</td>
<td>0.9892</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.8997</td>
<td>0.9499</td>
<td>0.9902</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8994</td>
<td>0.9496</td>
<td>0.9897</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8996</td>
<td>0.9497</td>
<td>0.9899</td>
</tr>
<tr>
<td>$p = 0.40$</td>
<td>10</td>
<td>50</td>
<td>0.8710</td>
<td>0.9220</td>
<td>0.9718</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8964</td>
<td>0.9443</td>
<td>0.9870</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8793</td>
<td>0.9290</td>
<td>0.9781</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8963</td>
<td>0.9453</td>
<td>0.9882</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8996</td>
<td>0.9471</td>
<td>0.9887</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8997</td>
<td>0.9475</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9000</td>
<td>0.9506</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8921</td>
<td>0.9440</td>
<td>0.9864</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8994</td>
<td>0.9472</td>
<td>0.9888</td>
</tr>
<tr>
<td>$p = 0.70$</td>
<td>10</td>
<td>50</td>
<td>0.8717</td>
<td>0.9212</td>
<td>0.9732</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8963</td>
<td>0.9461</td>
<td>0.9871</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8785</td>
<td>0.9302</td>
<td>0.9770</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8944</td>
<td>0.9477</td>
<td>0.9875</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8997</td>
<td>0.9485</td>
<td>0.9895</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8999</td>
<td>0.9493</td>
<td>0.9898</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.8999</td>
<td>0.9500</td>
<td>0.9901</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8906</td>
<td>0.9416</td>
<td>0.9849</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8992</td>
<td>0.9494</td>
<td>0.9897</td>
</tr>
<tr>
<td>$p = 0.90$</td>
<td>10</td>
<td>50</td>
<td>0.8732</td>
<td>0.9224</td>
<td>0.9664</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8993</td>
<td>0.9453</td>
<td>0.9874</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8799</td>
<td>0.9310</td>
<td>0.9760</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8961</td>
<td>0.9455</td>
<td>0.9873</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8995</td>
<td>0.9497</td>
<td>0.9891</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8996</td>
<td>0.9499</td>
<td>0.9895</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.8997</td>
<td>0.9499</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8942</td>
<td>0.9462</td>
<td>0.9870</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8985</td>
<td>0.9482</td>
<td>0.9891</td>
</tr>
</tbody>
</table>
In Figure 1, the values 1−5 on the $x$−axis have been allocated to sets of samples sizes in increasing order. On the $y$−axis we have the empirical coverage rates for 95% confidence interval at $p = 0.10$. For the first case where $n < m$ we have 1, 2, 3, 4 and 5 corresponding to the (10, 50), (50, 100), (100, 200), (150, 250) and (200, 500) sample size combinations. For the second case where $n = m$ we have 1, 2, 3, 4 and 5 corresponding to the (10, 10), (50, 50), (100, 100), (200, 200) and (500, 500) sample size combinations. For the third case where $n > m$ we have 1, 2, 3, 4 and 5 corresponding to the (50, 10), (100, 50), (200, 100), (250, 150) and (500, 200) sample size combinations.

Figure 1: Coverage Rates vs Sample Size for Ratio of Normal Percentiles at $p = 0.10$
3  EXPONENTIAL DISTRIBUTION

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential population with mean $\theta$ and variance $\theta^2$. Then the p.d.f. of $X$ is given by

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 \leq x < \infty. \quad (27)$$

Since this is a continuous distribution the cumulative distribution function (c.d.f) can be used to determine the value that lies on a given percentile

$$F(x) = P(X \leq k_p) = \int_0^{k_p} \frac{1}{\theta} e^{-x/\theta} dx = 1 - e^{-k_p/\theta} = p \quad (28)$$

and solving for $k_p$ we have

$$k_p = -\theta \ln(1 - p) \quad (29)$$

where $k_p$ is the 100$pt$ percentile. Since $\theta$ is not known we need to find an unbiased estimator of $\theta$ that can be used to estimate $k_p$.

3.1  Unbiased Estimator of $\theta$

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential population with mean $\theta$ and let $\bar{X}$ denote the sample mean. $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is an unbiased estimator of $\theta$ [5]. Also by the Central Limit Theorem the sample mean is an unbiased estimator of the population mean and $E(\bar{X}) = \theta$. Therefore the estimated $k_p$ denoted $\hat{k}_p$ is given by

$$\hat{k}_p = -\bar{X} \ln(1 - p). \quad (30)$$
Next we need to find the variance of $\hat{k}_p$ and the estimated variance of $\hat{k}_p$, denoted by $\hat{\text{Var}}(\hat{k}_p)$.

$$\begin{align*}
\text{Var}(\hat{k}_p) &= \text{Var}(-\bar{X} \ln(1-p)) \\
&= (\ln(1-p))^2 \text{Var}(\bar{X}) \\
&= (\ln(1-p))^2 \frac{\sigma^2}{n} \\
&= (\ln(1-p))^2 \frac{\theta^2}{n} .
\end{align*}$$

(31)

The above follows from the fact that $\bar{X}$ is normally distributed with mean $\mu$ and variance $\frac{\sigma^2}{n}$ and for the exponential distribution $\sigma^2 = \theta^2$ and so

$$\hat{\text{Var}}(\hat{k}_p) = (\ln(1-p))^2 \bar{X}^2 \frac{n}{n} .$$

(32)

3.2 Approximate Confidence Interval for the Ratio of Percentiles from Two Independent Exponential Distributions

In this section, we are going to develop a method of computing the $(1 - \alpha)100\%$ confidence intervals for the ratio of two exponential percentiles. These two percentiles must come from two independent exponential distributions and they must be the same $100pth$ percentile. For example, if we use the $45th$ percentile from the first distribution then we must use the $45th$ percentile from the second distribution. The means, variances, and sample sizes of the two samples need not be equal.

**Theorem 3.1** Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from an exponential population $X$ with mean $\theta_x$ and variance $\theta_x^2$ and let $Y_1, Y_2, \ldots, Y_m$ be a random sample of size $m$ from another exponential population $Y$ with mean $\theta_y$ and variance $\theta_y^2$ where
X and Y are independent. Let \( k_{p,x} \) and \( k_{p,y} \) be the 100\( p \)th percentiles from populations X and Y respectively. Then an approximate \((1 - \alpha)100\%\) confidence interval for the ratio of the percentiles \( k_{p,x} \) and \( k_{p,y} \) is given by

\[
\frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \exp \left( \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{1}{n} + \frac{1}{m}} \right).
\]  

(33)

Proof

We know that

\[
k_{p,x} = -\theta_x \ln(1 - p) \quad \text{and} \quad k_{p,y} = -\theta_y \ln(1 - p)
\]  

(34)

and the unbiased estimators of \( k_{p,x} \) and \( k_{p,y} \) are

\[
\hat{k}_{p,x} = -\bar{X} \ln(1 - p) \quad \text{and} \quad \hat{k}_{p,y} = -\bar{Y} \ln(1 - p).
\]  

(35)

respectively. The unbiased estimator of \( k_p \) ratio denoted by \( \hat{k}_p \) ratio is given by

\[
\hat{k}_p \ \text{ratio} \quad = \quad \frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \\
\quad = \quad \frac{-\bar{X} \ln(1 - p)}{-\bar{Y} \ln(1 - p)} \\
\quad = \quad \frac{\bar{X}}{\bar{Y}}.
\]  

(36)

Now we need to find the variance of the \( \hat{k}_p \) ratio. This can be done by introducing natural logarithms and using the delta method to compute the variance.

\[
\ln(\hat{k}_p \ \text{ratio}) \quad = \quad \ln \left( \frac{\bar{X}}{\bar{Y}} \right) \quad = \quad \ln(\bar{X}) - \ln(\bar{Y}).
\]  

(37)

The next few equations employ the delta method to compute the variance of \( \ln(\hat{k}_p \ \text{ratio}) \).

\[
\text{Var}(\ln(\hat{k}_p \ \text{ratio})) =
\]
\[ \begin{bmatrix} \frac{d \ln(k_p \text{ ratio})}{dX} & \frac{d \ln(k_p \text{ ratio})}{dY} \end{bmatrix} \begin{bmatrix} \text{Var}(X) & 0 \\ 0 & \text{Var}(Y) \end{bmatrix} \begin{bmatrix} \frac{d \ln(k_p \text{ ratio})}{dX} \\ \frac{d \ln(k_p \text{ ratio})}{dY} \end{bmatrix}. \]

The variance covariance matrix has zero entries because \( X \) and \( Y \) are independent and their covariance is zero.

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \begin{bmatrix} \frac{1}{X} & -\frac{1}{Y} \end{bmatrix} \begin{bmatrix} \frac{\sigma_x^2}{n} & 0 \\ 0 & \frac{\sigma_y^2}{m} \end{bmatrix} \begin{bmatrix} \frac{1}{X} \\ -\frac{1}{Y} \end{bmatrix}. \tag{38}
\]

Hence

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{\sigma_x^2}{nX^2} + \frac{\sigma_y^2}{mY^2}, \tag{39}
\]

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{\theta_x^2}{nX^2} + \frac{\theta_y^2}{mY^2} \tag{40}
\]

and the estimated variance of the \( \ln(\hat{k}_p \text{ ratio}) \) is

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{\bar{X}^2}{nX^2} + \frac{\bar{Y}^2}{mY^2}, \tag{41}
\]

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{1}{n} + \frac{1}{m}. \tag{42}
\]

The \((1 - \alpha)100\%\) confidence interval for the \( \ln(k_p \text{ ratio}) \) can be computed as

\[
\ln(\bar{X}) - \ln(\bar{Y}) \pm Z_{(1 - \frac{\alpha}{2})} \sqrt{\frac{1}{n} + \frac{1}{m}}, \tag{43}
\]

and exponentiating the expression we have the \((1 - \alpha)100\%\) confidence interval for the \( k_p \text{ ratio} \) as

\[
\frac{\bar{X}}{\bar{Y}} \exp \left( \pm Z_{(1 - \frac{\alpha}{2})} \sqrt{\frac{1}{n} + \frac{1}{m}} \right). \]
3.3 Simulation Results

To test this method, 50000 simulations were done. Two random samples were generated using the R statistical software [8]. The first sample was generated from an exponential distribution with mean fixed at 0.1 and the second sample was generated from an exponential distribution with mean fixed at 0.05. The simulation process involved 50000 runs for each different set of sample sizes and the resulting empirical coverage rates were recorded. Table 2 on the following page shows the empirical coverage rates for a few sets of sample sizes. The empirical coverage can be seen to converge to 0.95 in Figure 2.
Table 2: Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Exponential populations.

<table>
<thead>
<tr>
<th>percentiles</th>
<th>n</th>
<th>m</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.8932</td>
<td>0.9442</td>
<td>0.9854</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8996</td>
<td>0.9461</td>
<td>0.9894</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8925</td>
<td>0.9431</td>
<td>0.9875</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.9023</td>
<td>0.9502</td>
<td>0.9888</td>
</tr>
<tr>
<td>$p = 0.10$</td>
<td>100</td>
<td>100</td>
<td>0.8986</td>
<td>0.9492</td>
<td>0.9902</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.9003</td>
<td>0.9486</td>
<td>0.9893</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9003</td>
<td>0.9504</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8915</td>
<td>0.9415</td>
<td>0.9860</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8991</td>
<td>0.9472</td>
<td>0.9894</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.8903</td>
<td>0.9432</td>
<td>0.9855</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8987</td>
<td>0.9480</td>
<td>0.9897</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8891</td>
<td>0.9424</td>
<td>0.9868</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8963</td>
<td>0.9494</td>
<td>0.9894</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.9008</td>
<td>0.9460</td>
<td>0.9904</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8964</td>
<td>0.9497</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9009</td>
<td>0.9499</td>
<td>0.9897</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8903</td>
<td>0.9428</td>
<td>0.9848</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8988</td>
<td>0.9503</td>
<td>0.9895</td>
</tr>
<tr>
<td>$p = 0.40$</td>
<td>10</td>
<td>50</td>
<td>0.8909</td>
<td>0.9440</td>
<td>0.9853</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.9002</td>
<td>0.9511</td>
<td>0.9898</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8913</td>
<td>0.9430</td>
<td>0.9861</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8980</td>
<td>0.9473</td>
<td>0.9894</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8988</td>
<td>0.9486</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.9017</td>
<td>0.9527</td>
<td>0.9898</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9006</td>
<td>0.9485</td>
<td>0.9903</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8929</td>
<td>0.9410</td>
<td>0.9854</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8982</td>
<td>0.9486</td>
<td>0.9895</td>
</tr>
<tr>
<td>$p = 0.70$</td>
<td>10</td>
<td>50</td>
<td>0.8910</td>
<td>0.9446</td>
<td>0.9855</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8989</td>
<td>0.9492</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8935</td>
<td>0.9451</td>
<td>0.9855</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8991</td>
<td>0.9494</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8980</td>
<td>0.9493</td>
<td>0.9896</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.9002</td>
<td>0.9508</td>
<td>0.9895</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.8995</td>
<td>0.9493</td>
<td>0.9896</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8926</td>
<td>0.9445</td>
<td>0.9853</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8994</td>
<td>0.9493</td>
<td>0.9889</td>
</tr>
</tbody>
</table>
In Figure 2, the values 1 - 5 on the $x$ - axis have been allocated to sets of samples sizes in increasing order. On the $y$ - axis we have the empirical coverage rates for 95% confidence interval at $p = 0.10$. For the first case where $n < m$ we have 1, 2, 3, 4 and 5 corresponding to the (10, 50), (50, 100), (100, 200), (150, 250) and (200, 500) sample size combinations. For the second case where $n = m$ we have 1, 2, 3, 4 and 5 corresponding to the (10, 10), (50, 50), (100, 100), (200, 200) and (500, 500) sample size combinations. For the third case where $n > m$ we have 1, 2, 3, 4 and 5 corresponding to the (50, 10), (100, 50), (200, 100), (250, 150) and (500, 200) sample size combinations.

Figure 2: Coverage Rates vs Sample Size for Ratio of Exponential Percentiles at $p = 0.10$
4 UNIFORM DISTRIBUTION

Let \( X_1, X_2, \ldots, X_n \) be a random sample from a uniform population with minimum \( a \) and maximum \( b \). Then the p.d.f. of \( X \) is given by

\[
f(x) = \frac{1}{b-a} \quad a \leq x \leq b.
\]  \hspace{1cm} (44)

Just like any other continuous distribution, the cumulative distribution function (c.d.f.) can be easily manipulated to come up with a formula for computing the percentiles

\[
F(x) = P(X \leq k_p) = \int_a^{k_p} \frac{1}{b-a} \, dx = \frac{k_p - a}{b-a} = p.
\]  \hspace{1cm} (45)

Solving for \( k_p \) we have

\[
k_p = a + (b - a)p
\]  \hspace{1cm} (46)

where \( k_p \) denotes the 100\( p \)th percentile. For us to estimate the unknown parameters \( a \) and \( b \), we need to find unbiased estimators of both parameters.

4.1 Unbiased Estimators of \( a \) and \( b \)

Since \( a \) and \( b \) are the population minimum and population maximum, the first and last order statistics are the most likelihood estimators of \( a \) and \( b \) respectively. Order statistics are observations of a random sample arranged, or ordered in magnitude from the smallest to the largest [5]. Let \( X_1, X_2, \ldots, X_n \) be a random sample from a uniform population with minimum \( a \) and maximum \( b \) and let \( X_{(1)}, X_{(2)}, \ldots, X_{(n)} \) be the order statistics associated with the random sample where \( X_{(1)} \) and \( X_{(n)} \) are the
minimum and maximum of $X_i$, respectively. Then $X_{(1)}$ and $X_{(n)}$ are the maximum likelihood estimators of $a$ and $b$, respectively. Though $X_{(1)}$ and $X_{(n)}$ are not exactly unbiased estimators of $a$ and $b$ they are asymptotic unbiased estimators. That is $\lim_{n \to \infty} E(X_{(1)}) = a$ and $\lim_{n \to \infty} E(X_{(n)}) = b$. This is shown below. The p.d.f. of the $r$th order statistics denoted $f_r(x)$ is given by

$$f_r(x) = \frac{n!}{(r-1)!1!(n-r)!} [F(x)]^{r-1} [1 - F(x)]^{n-r} f(x) \quad a < x < b \quad 1 \leq r \leq n \quad [5].$$ \hfill (47)

Substituting $r=1$ and $r=n$ for the first and last order statistics we have

$$f_1(x) = n[1 - F(x)]^{n-1} f(x) = \frac{n(b-x)^{n-1}}{(b-a)^n} \quad a < x < b \quad \hfill (48)$$

and

$$f_n(x) = n[F(x)]^{n-1} f(x) = \frac{n(x-a)^{n-1}}{(b-a)^n} \quad a < x < b \quad [5]. \hfill (49)$$

Now we need to find $E(X_{(1)})$ and $E(X_{(n)})$.

$$E(X_{(1)}) = \int_a^b x \frac{n(b-x)^{n-1}}{(b-a)^n} dx$$

$$= \frac{n}{(b-a)^n} \int_a^b x(b-x)^{n-1} dx. \hfill (50)$$

Applying integration by parts we have

$$u = x, \ du = dx, dv = (b-x)^{n-1}, v = \frac{-(b-x)^n}{n}$$

$$E(X_{(1)}) = \frac{n}{(b-a)^n} \left[ -\frac{x(b-x)^n}{n} \bigg|_a^b - \int_a^b \frac{-(b-x)^n}{n} dx \right]$$

$$= \frac{n}{(b-a)^n} \left[ \frac{-(b-x)^n}{n} - \frac{(b-x)^{n+1}}{n(n+1)} \right]_a^b$$

$$= \left[ \frac{-x(b-x)^n}{(b-a)^n} - \frac{(b-x)^{n+1}}{n(n+1)} \right]_a^b$$

$$= a + \frac{b-a}{n+1}. \hfill (50)$$
\[ E(X_{(n)}) = \int_a^b x \frac{n(x-a)^{n-1}}{(b-a)^n} dx \]
\[ = \frac{n}{(b-a)^n} \int_a^b x(x-a)^{n-1} dx. \]

Applying integration by parts we have

\[ u = x, \quad du = dx, \quad dv = (x-a)^{n-1}, \quad v = \frac{(x-a)^n}{n} \]

\[ E(X_{(n)}) = \frac{n}{(b-a)^n} \left[ \frac{x(x-a)^n}{n} - \frac{(x-a)^{n+1}}{n(n+1)} \right]_a^b \]
\[ = \frac{n}{(b-a)^n} \left[ \frac{x(x-a)^n}{n} - \frac{(x-a)^{n+1}}{n(n+1)} \right]_a^b \]
\[ = b - \frac{b-a}{n+1}. \] (51)

From above \( \lim_{n \to \infty} E(X_{(1)}) = a \) and \( \lim_{n \to \infty} E(X_{(n)}) = b \) and the two are the best minimum variance unbiased estimators of \( a \) and \( b \) respectively. Now we can estimate \( k_p \) denoted \( \hat{k}_p \) as follows,

\[ \hat{k}_p = X_{(1)} + (X_{(n)} - X_{(1)})p = (1-p)X_{(1)} + pX_{(n)}. \] (52)

Since we are building confidence intervals we need the variance of \( \hat{k}_p \).

\[ Var(\hat{k}_p) = (1-p)^2 Var(X_{(1)}) + p^2 Var(X_{(n)}) + 2p(1-p) Cov(X_{(1)}X_{(n)}). \] (53)

The last term of (53) exists because \( X_{(1)} \) and \( X_{(n)} \) are not independent. We begin by computing the variance of \( X_{(1)} \) as

\[ Var(X_{(1)}) = E(X_{(1)}^2) - (E(X_{(1)}))^2 \]
where

\[
E(X_{(1)}^2) = \int_a^b x^2 \frac{n(b-x)^{n-1}}{(b-a)^n} \, dx \quad \text{(integration by parts)}
\]

\[
= \frac{n}{(b-a)^n} \int_a^b x^2 (b-x)^{n-1} \, dx
\]

\[
u = x^2, \; du = 2x \, dx, \; dv = (b-x)^{n-1}, v = \frac{-(b-x)^n}{n}
\]

\[
E(X_{(1)}^2) = \frac{n}{(b-a)^n} \left[ \frac{-x^2(b-x)^n}{n} - \int_a^b \frac{-2x(b-x)^n}{n} \, dx \right]_a^b
\]

\[
s = 2x, \; ds = 2, \; dt = (b-x)^n, t = \frac{-(b-x)^{n+1}}{n+1}
\]

\[
E(X_{(1)}^2) = \frac{n}{(b-a)^n} \left[ \frac{-x^2(b-x)^n}{n} - \int_a^b \frac{-2x(b-x)^n}{n} \, dx \right]_a^b
\]

\[
= a^2 + \frac{2a(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} \quad (54)
\]

and therefore

\[
Var(X_{(1)}) = a^2 + \frac{2a(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} - \left( a + \frac{b-a}{n+1} \right)^2
\]

\[
= \frac{n(b-a)^2}{(n+1)^2(n+2)} \quad (55)
\]

Next we compute the variance of \( X_{(n)} \) as

\[
Var(X_{(n)}) = E(X_{(n)}^2) - (E(X_{(n)}))^2
\]
where

\[
E(X_{(n)}^2) = \int_a^b x^2 \frac{n(x-a)^{n-1}}{(b-a)^n} \, dx \quad \text{(integration by parts)}
\]

\[
= \frac{n}{(b-a)^n} \int_a^b x^2(x-a)^{n-1} \, dx
\]

\[
u = x^2, \quad du = 2xdx, \quad dv = (x-a)^{n-1}, \quad v = \frac{(x-a)^n}{n}
\]

\[
E(X_{(n)}^2) = \frac{n}{(b-a)^n} \left[ \frac{x^2(x-a)^n}{n} - \frac{2(x-a)^{n+1}}{n+1} \right]_a^b
\]

\[
s = 2x, \quad ds = 2, \quad dt = (x-a)^n, \quad t = \frac{(x-a)^{n+1}}{n+1}
\]

\[
E(X_{(n)}^2) = \frac{n}{(b-a)^n} \left[ \frac{x^2(x-a)^n}{n} - \frac{2(x-a)^{n+1}}{n+1} \right]_a^b
\]

\[
= b^2 - \frac{2b(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} \quad \text{(56)}
\]

and so

\[
Var(X_{(n)}) = b^2 - \frac{2b(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} - \left( \frac{b-a}{n+1} \right)^2
\]

\[
= \frac{n(b-a)^2}{(n+1)^2(n+2)}. \quad \text{(57)}
\]

Lastly we need to compute the covariance of \(X_{(1)}\) and \(X_{(n)}\) as

\[
Cov(X_{(1)}, X_{(n)}) = E(X_{(1)}X_{(n)}) - E(X_{(1)})E(X_{(n)}).
\]

\text{(58)}
The joint p.d.f of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{1,n}(x_{(1)}, x_{(n)}) = \frac{1}{(b-a)^n} n(n-1)(x_{(n)} - x_{(1)})^{n-2} \quad a < x_{(1)} \leq x_{(n)} < b \quad [2]. \quad (59)$$

Let $X_{(1)} = x$ and $X_{(n)} = y$

$$E(X_{(1)}X_{(n)}) = \frac{1}{(b-a)^n} \int_{a}^{b} \int_{a}^{y} xyn(n-1)(y-x)^{n-2}dxdy$$

$$= \frac{n(n-1)}{(b-a)^n} \int_{a}^{b} \int_{a}^{y} xy(y-x)^{n-2}dxdy$$

$$= \frac{n(n-1)}{(b-a)^n} \int_{a}^{b} y \left[ \int_{a}^{y} x(y-x)^{n-2}dx \right] dy .$$

First we integrate 1 using integration by parts

$$(1) = \int_{a}^{y} x(y-x)^{n-2}dx \quad u = x, du = dx, dv = (y-x)^{n-2}, v = \frac{-(y-x)^{n-1}}{n-1}$$

$$= \frac{-x(y-x)^{n-1}}{n-1} + \int_{a}^{y} \frac{(y-x)^{n-1}}{n-1}dx$$

$$= \left[ \frac{-x(y-x)^{n-1}}{n-1} - \frac{(y-x)^{n}}{n(n-1)} \right]_{a}^{y}$$

$$= \frac{a(y-a)^{n-1}}{n-1} + \frac{(y-a)^{n}}{n(n-1)} .$$

Plugging (1) back to our equation and using integration by parts we have

$$E(X_{(1)}X_{(n)}) = \frac{n(n-1)}{(b-a)^n} \int_{a}^{b} y \left[ \frac{a(y-a)^{n-1}}{n-1} + \frac{(y-a)^{n}}{n(n-1)} \right] dy$$

$$= \frac{1}{(b-a)^n} \left[ n \int_{a}^{b} ay(y-a)^{n-1}dy + \int_{a}^{b} y(y-a)^{n}dy \right] .$$
where
\[ (2) = na \int_a^b y(y-a)^{n-1} dy, \quad u = y, du = dy, dv = (y-a)^{n-1}, v = \frac{(y-a)^n}{n} \]
\[ = na \left[ \frac{y(y-a)^n}{n} - \int \frac{(y-a)^n}{n} dy \right]_a^b \]
\[ = na \left[ \frac{y(y-a)^n}{n} - \frac{(y-a)^{n+1}}{n(n+1)} \right]_a^b \]
\[ = ab(b-a)^n - \frac{a(b-a)^{n+1}}{n+1} \]

and
\[ (3) = \int_a^b y(y-a)^n dy, \quad u = y, du = dy, dv = (y-a), v = \frac{(y-a)^{n+1}}{n+1} \]
\[ = \left[ \frac{y(y-a)^{n+1}}{n+1} - \int \frac{(y-a)^{n+1}}{n+1} dx \right]_a^b \]
\[ = \left[ \frac{y(y-a)^{n+1}}{n+1} - \frac{(y-a)^{n+2}}{(n+1)(n+2)} \right]_a^b \]
\[ = \frac{b(b-a)^{n+1}}{n+1} - \frac{(b-a)^{n+2}}{(n+1)(n+2)}. \]

Putting (2) and (3) together we have
\[
E(X_1 X_{(n)}) = \frac{1}{(b-a)^n} \left[ \left( ab(b-a)^n - \frac{a(b-a)^{n+1}}{n+1} \right) + \left[ \frac{b(b-a)^{n+1}}{n+1} - \frac{(b-a)^{n+2}}{(n+1)(n+2)} \right] \right]
\]
\[
E(X_1 X_{(n)}) = \frac{1}{(b-a)^n} \left[ \left( ab(b-a)^n - \frac{a(b-a)^{n+1}}{n+1} + \frac{b(b-a)^{n+1}}{n+1} - \frac{(b-a)^{n+2}}{(n+1)(n+2)} \right) \right]
\]

hence
\[
E(X_1 X_{(n)}) = ab - \frac{a(b-a)}{n+1} + \frac{b(b-a)}{n+1} - \frac{(b-a)^2}{(n+1)(n+2)}
\]
\[ = ab + \frac{(b-a)^2}{n+1} - \frac{(b-a)^2}{(n+1)(n+2)}. \]

(60)
Now we have

\[
\text{Cov}(X_{(1)}X_{(n)}) = E(X_{(1)}X_{(n)}) - E(X_{(1)})E(X_{(n)})
\]

\[
= ab + \frac{(b-a)^2}{n+1} - \frac{(b-a)^2}{(n+1)(n+2)} - \left( a + \frac{b-a}{n+1} \right) \left( b - \frac{b-a}{n+1} \right)
\]

\[
= \frac{(b-a)^2}{(n+1)^2(n+2)}
\]

(61)

and

\[
\text{Var}( \hat{k}_p ) = (1-p)^2 \text{Var}(X_{(1)}) + p^2 \text{Var}(X_{(n)}) + 2p(1-p) \text{Cov}(X_{(1)}X_{(n)})
\]

\[
= (1-p)^2 \left( \frac{n(b-a)^2}{(n+1)^2(n+2)} \right) + p^2 \left( \frac{n(b-a)^2}{(n+1)^2(n+2)} \right)
\]

\[
+ 2p(1-p) \left( \frac{(b-a)^2}{(n+1)^2(n+2)} \right)
\]

\[
\text{Var}( \hat{k}_p ) = \left( \frac{(b-a)^2}{(n+1)^2(n+2)} \right) (n(1-p)^2 + np^2 + 2p(1-p))
\]

(62)

and the estimated variance of \( \hat{k}_p \) denoted \( \hat{\text{Var}}( \hat{k}_p ) \) is

\[
\hat{\text{Var}}( \hat{k}_p ) = \left( \frac{(X_{(n)} - X_{(1)})^2}{(n+1)^2(n+2)} \right) (n(1-p)^2 + np^2 + 2p(1-p))
\]

(63)

4.2 Approximate Confidence Interval for the Ratio of Percentiles from Two
Independent Uniform Distributions

In this section, we are going to develop a method of computing the \((1-\alpha)100\%\) confidence intervals for the ratio of two uniform percentiles. Just like the previous two cases, these two percentiles must come from two independent exponential distributions and they must be the same 100\(p\)th percentile. For example, if we use the 45\(th\) percentile from the first distribution then we must use the 45\(th\) percentile from the
second distribution. The population parameters and sample sizes of the two samples need not be equal.

**Theorem 4.1** Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a uniform population $X$ with minimum $a$ and maximum $b$ and let $Y_1, Y_2, \ldots, Y_m$ be another random sample of size $m$ from another exponential population $Y$ with minimum $c$ and maximum $d$ where $X$ and $Y$ are independent. Let $k_{p,x}$ and $k_{p,y}$ be the $100$th percentiles from populations $X$ and $Y$ respectively. Then an approximate $(1 - \alpha)100\%$ confidence interval for the ratio of the percentiles $k_{p,x}$ and $k_{p,y}$ denoted $\frac{k_{p,x}}{k_{p,y}}$ is given by

$$\frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \exp \left( \pm Z_{(1 - \alpha/2)} \sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \right). \quad (64)$$

**Proof**

We know that

$$k_{p,x} = a + (b - a)p \quad \text{and} \quad k_{p,y} = c + (d - c)p \quad (65)$$

and the respective unbiased estimators of $k_{p,x}$ and $k_{p,y}$ are

$$\hat{k}_{p,x} = (1 - p)X_{(1)} + pX_{(n)} \quad \text{and} \quad \hat{k}_{p,y} = (1 - p)Y_{(1)} + pY_{(m)}. \quad (66)$$

The variances of these unbiased estimators $\hat{k}_{p,x}$ and $\hat{k}_{p,y}$ are

$$\text{Var}(\hat{k}_{p,x}) = \left( \frac{(b - a)^2}{(n + 1)^2(n + 2)} \right) \left( n(1 - p)^2 + np^2 + 2p(1 - p) \right)$$

and

$$\text{Var}(\hat{k}_{p,y}) = \left( \frac{(d - c)^2}{(m + 1)^2(m + 2)} \right) \left( m(1 - p)^2 + mp^2 + 2p(1 - p) \right). \quad (67)$$
and their estimated variances are

\[
\text{Var}(\hat{k}_{p,x}) = \left( \frac{(X(n) - X(1))^2}{(n + 1)^2(n + 2)} \right) \left( n(1 - p)^2 + np^2 + 2p(1 - p) \right)
\]

and

\[
\text{Var}(\hat{k}_{p,y}) = \left( \frac{(Y(n) - Y(1))^2}{(m + 1)^2(m + 2)} \right) \left( m(1 - p)^2 + mp^2 + 2p(1 - p) \right) . \tag{68}
\]

The estimated \( k_p \) ratio denoted by \( \hat{k}_p \) ratio is given by

\[
\hat{k}_p \text{ ratio} = \frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} . \tag{69}
\]

Now we need to find the variance of the \( \hat{k}_p \) ratio. This can be simplified by introducing natural logarithms and using the delta method to compute the variance.

\[
\ln(\hat{k}_p \text{ ratio}) = \ln \left( \frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \right) = \ln(\hat{k}_{p,x}) - \ln(\hat{k}_{p,y}) . \tag{70}
\]

The next few equations employ the delta method to compute the variance of \( \ln(\hat{k}_p \text{ ratio}) \).

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) =
\]

\[
\begin{bmatrix}
\frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,x}} & \frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,y}} \\
\frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,x}} & \frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,y}}
\end{bmatrix}
\begin{bmatrix}
\text{Var}(\hat{k}_{p,x}) & 0 \\
0 & \text{Var}(\hat{k}_{p,y})
\end{bmatrix}
\begin{bmatrix}
\frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,x}} \\
\frac{d\ln(\hat{k}_p \text{ ratio})}{dk_{p,y}}
\end{bmatrix}.
\]

The variance covariance matrix has zero entries because X and Y are independent and their covariance is zero.

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \begin{bmatrix}
\frac{1}{k_{p,x}} & -\frac{1}{k_{p,y}} \\
\frac{1}{k_{p,x}} & -\frac{1}{k_{p,y}}
\end{bmatrix}
\begin{bmatrix}
\text{Var}(\hat{k}_{p,x}) & 0 \\
0 & \text{Var}(\hat{k}_{p,y})
\end{bmatrix}
\begin{bmatrix}
\frac{1}{k_{p,x}} \\
\frac{1}{k_{p,y}}
\end{bmatrix} . \tag{71}
\]

Hence

\[
\text{Var}(\ln(\hat{k}_p \text{ ratio})) = \frac{\text{Var}(\hat{k}_{p,x})}{k_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{k_{p,y}^2} . \tag{72}
\]
and the estimated variance of the $\ln(\hat{k}_p \text{ ratio})$ is

$$\text{Var} \ln(\hat{k}_p \text{ ratio}) = \frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}. \quad (73)$$

The $(1 - \alpha)100\%$ confidence interval for the $\ln(k_p \text{ ratio})$ can be computed as

$$\ln(\hat{k}_{p,x}) - \ln(\hat{k}_{p,y}) \pm Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \quad (74)$$

and exponentiating the expression we have the $(1 - \alpha)100\%$ confidence interval for the $k_p \text{ ratio}$ as

$$\frac{\hat{k}_{p,x}}{\hat{k}_{p,y}} \exp \left( \pm Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\text{Var}(\hat{k}_{p,x})}{\hat{k}_{p,x}^2} + \frac{\text{Var}(\hat{k}_{p,y})}{\hat{k}_{p,y}^2}} \right). \quad \blacksquare$$
4.3 Simulation Results

To test this method of building confidence intervals for the ratio of uniform percentiles, simulations were done. Two random random samples were generated using the R statistical software [8]. The first random sample was generated from a uniform distribution with minimum and maximum fixed at 1 and 3 respectively and the second sample was generated from a uniform distribution with minimum and maximum fixed at 2 and 5 respectively. The simulation process involved 100000 runs for each set of sample sizes and the resulting empirical coverage rates were recorded. Table 3 shows some of the empirical coverages rates from the simulations. The empirical coverage can be seen to converge to 0.95 in Figure 3.
Table 3: Empirical Coverage Rates of 90%, 95% and 99% Confidence Intervals for the Ratio of Percentiles from Two Uniform populations.

<table>
<thead>
<tr>
<th>percentiles</th>
<th>n</th>
<th>m</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.7412</td>
<td>0.7857</td>
<td>0.8523</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8643</td>
<td>0.9005</td>
<td>0.9449</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8301</td>
<td>0.8796</td>
<td>0.9401</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8852</td>
<td>0.9261</td>
<td>0.9664</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8940</td>
<td>0.9305</td>
<td>0.9693</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8979</td>
<td>0.9328</td>
<td>0.9717</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.8981</td>
<td>0.9356</td>
<td>0.9723</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.7756</td>
<td>0.8143</td>
<td>0.8773</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8828</td>
<td>0.9219</td>
<td>0.9636</td>
</tr>
<tr>
<td>$p = 0.10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.8093</td>
<td>0.8566</td>
<td>0.9129</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8872</td>
<td>0.9259</td>
<td>0.9672</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8204</td>
<td>0.8768</td>
<td>0.9412</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8865</td>
<td>0.9297</td>
<td>0.9714</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8946</td>
<td>0.9349</td>
<td>0.9741</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8984</td>
<td>0.9372</td>
<td>0.9758</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9015</td>
<td>0.9395</td>
<td>0.9775</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8139</td>
<td>0.8616</td>
<td>0.9190</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8906</td>
<td>0.9285</td>
<td>0.9714</td>
</tr>
<tr>
<td>$p = 0.40$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.8089</td>
<td>0.8577</td>
<td>0.9132</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8880</td>
<td>0.9230</td>
<td>0.9640</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8226</td>
<td>0.8737</td>
<td>0.9353</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8870</td>
<td>0.9293</td>
<td>0.9712</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8947</td>
<td>0.9330</td>
<td>0.9741</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8995</td>
<td>0.9363</td>
<td>0.9758</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9017</td>
<td>0.9391</td>
<td>0.9752</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.8165</td>
<td>0.8619</td>
<td>0.9177</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8884</td>
<td>0.9264</td>
<td>0.9671</td>
</tr>
<tr>
<td>$p = 0.70$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>0.7761</td>
<td>0.8238</td>
<td>0.8860</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>0.8734</td>
<td>0.9114</td>
<td>0.9537</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>0.8241</td>
<td>0.8746</td>
<td>0.9351</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.8871</td>
<td>0.9267</td>
<td>0.9674</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.8962</td>
<td>0.9328</td>
<td>0.9697</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.8983</td>
<td>0.9358</td>
<td>0.9727</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>0.9016</td>
<td>0.9361</td>
<td>0.9733</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
<td>0.7814</td>
<td>0.8302</td>
<td>0.8923</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.8825</td>
<td>0.9188</td>
<td>0.9606</td>
</tr>
</tbody>
</table>

$p = 0.10$  |     |     |       |       |       |

$p = 0.40$  |     |     |       |       |       |

$p = 0.70$  |     |     |       |       |       |

$p = 0.90$  |     |     |       |       |       |
In Figure 3, the values 1 – 5 on the $x$–axis have been allocated to sets of samples sizes in increasing order. On the $y$–axis we have the empirical coverage rates for 95% confidence interval at $p = 0.10$. For the first case where $n < m$ we have 1, 2, 3, 4 and 5 corresponding to the (10,50), (50,100), (100,200), (150,250) and (200,500) sample size combinations. For the second case where $n = m$ we have 1, 2, 3, 4 and 5 corresponding to the (10,10), (50,50), (100,100), (200,200) and (500,500) sample size combinations. For the third case where $n > m$ we have 1, 2, 3, 4 and 5 corresponding to the (50,10), (100,50), (200,100), (250,150) and (500,200) sample size combinations.

Figure 3: Coverage Rates vs Sample Size for Ratio of Uniform Percentiles at $p = 0.10$
5 CONCLUSION

The method used to compute an approximate confidence interval for ratio of percentiles from the three distributions can be extended to other distributions. Typically, the approximate coverage probability for a ratio can be improved by first applying the delta method to the natural log function. The interval for the ratio can be found by exponentiating the end points of the interval found based on the natural log transformation.

5.1 Results

From the simulations results in Tables 1 – 3, it can be seen that the empirical coverage rates are largely dependent on the sample sizes. Obviously, it can be seen that the larger the sample size the better the coverage. There are a few variations which correspond to whether the samples sizes in question are equal or not equal and other variations may occur as a result of changing the population parameters. When the sample sizes are unequal, the coverage can be low if the larger of the two sample sizes corresponded to the population that had the larger variability. For the uniform distribution, the coverage is slow to converge to the specified level of confidence. This may be due to the fact that the order statistics are biased estimators of the parameters or it may be caused by the kurtosis present in the uniform distribution.

5.2 Limitations

There are two main limitations associated with this method of computing confidence intervals.
First is the fact that this method only works for populations whose underlying distribution is known. If one intends to apply this method on any collected data, then one must determine what kind of distribution the data comes from otherwise the method cannot be used. A possible distribution-free method known as bootstrapping may be a solution.

Secondly the method is dependent on sample size and may not be reliable for very small sizes. For the uniform distribution this method may be only reliable for sample sizes greater than or equal to 50.

This method also relies on the central limit theorem thus the reason for the use of the Z-statistic. Other minor problems may arise from the difficulty to determine the maximum likelihood estimators and the unbiased estimators for the population parameters. This can be seen in the case of the uniform distribution where order statistics were used and this made the whole process quite lengthy and cumbersome.
BIBLIOGRAPHY


confidence_interval = function(n1,n2,mu1,sig1,mu2,sig2,alpha,p){
  z=qnorm(1-alpha/2)
  zp=qnorm(p) #z for the pth percentile
  #sample1=(insert sample 1 here)
  #sample2=(insert sample 2 here)
  sample1=rnorm(n1,mu1,sig1) #random normal sample 1
  sample2=rnorm(n2,mu2,sig2) #random normal sample 2
  c1 = (sqrt((n1-1)/2)*gamma((n1-1)/2))/gamma(n1/2) #constant for removing std dev bias
  c2 = (sqrt((n2-1)/2)*gamma((n2-1)/2))/gamma(n2/2) #constant for removing std dev bias
  #lnc1 = log(sqrt((n1-1)/2)) + lgamma((n1-1)/2) - lgamma(n1/2)
  # if n1 is large (outside gamma function)
  #c1 = exp(lnc1)
  #lnc2 = log(sqrt((n2-1)/2)) + lgamma((n2-1)/2) - lgamma(n2/2)
  # if n2 is large (outside gamma function)
  #c2 = exp(lnc2)
  mean1=mean(sample1) #mean of sample 1
  mean2=mean(sample2) #mean of sample 2
  sd1=sd(sample1) #standard deviation of sample 1
sd2 = sd(sample2)  # standard deviation of sample 2
kp1 = mean1 + (c1*zp*sd1)  # MVUE of the pth percentile 1
kp2 = mean2 + (c2*zp*sd2)  # MVUE of the pth percentile 2
var_kp1_hat = (sd1^2/n1)*(1 + n1*zp^2*(c1^2 - 1))  # estimated variance of kp1
var_kp2_hat = (sd2^2/n2)*(1 + n2*zp^2*(c2^2 - 1))  # estimated variance of kp2
kp_ratio = kp1/kp2  # estimated percentile ratio
var_lnratio = var_kp1_hat/kp1^2 + var_kp2_hat/kp2^2  # variance of the natural log of kp_ratio
lb = kp_ratio*exp(-z*sqrt(var_lnratio))  # lower limit of the percentile ratio confidence interval
ub = kp_ratio*exp(z*sqrt(var_lnratio))  # upper limit of the percentile ratio confidence interval
list(c1, c2, mean1, mean2, sd1, sd2, kp1, kp2, var_kp1_hat, var_kp2_hat, lb, ub)
# list all the statistics defined above that you desire to compute }
confidence_interval(200, 100, 20, 3, 30, 2, .05, .1)  # insert your values here
Appendix B: R Code for the CI of Exponential Percentiles Ratio

confidence_interval = function(n1,n2,mu1,mu2,alpha,p){
  z=qnorm(1-alpha/2) #100pth percentile of the standard normal N(0,1)
  sample1=(insert sample 1)
  sample2=(insert sample 2)
  sample1=rexp(n1,mu1) #randomly generated exponential sample 1
  sample2=rexp(n2,mu2) #randomly generated exponential sample 2
  mean1=mean(sample1) #mean of sample 1
  mean2=mean(sample2) #mean of sample 2
  kp1 = -mean1*log(1-p) #100pth percentile of sample 1
  kp2 = -mean2*log(1-p) #100pth percentile of sample 2
  ratio_hat=kp1/kp2 #ratio of the two 100pth percentiles
  var_lnratio = 1/n1 + 1/n2 #variance of the natural log of ratio_hat
  lb = ratio_hat*exp(-z*sqrt(var_lnratio)) #lower bound of confidence interval
  ub = ratio_hat*exp(z*sqrt(var_lnratio)) #upper bound of confidence interval
  list(mean1,mean2,kp1,kp2,ratio_hat,lb,ub) #list all the statistics required
}

confidence_interval(100,200,1/10,1/20,.05,.1) #Insert your values here
Appendix C: R Code for the CI of Uniform Percentiles Ratio [8]

```r
confidence_interval= function(n1,n2,a,b,c,d,alpha,p){
  z=qnorm(1-alpha/2)
  #sample1=(insert sample 1 here)
  #sample2=(insert sample 2 here)
  sample1=runif(n1,a,b) #random uniform sample 1
  sample2=runif(n2,c,d) #random uniform sample 2
  x=sort(sample1) # sort sample 1 ascending
  y=sort(sample2) # sort sample 2 ascending
  kp1=x[1]+(x[n1]-x[1])*p # MVUE of the 100pth percentile 1
  kp2=y[1]+(y[n2]-y[1])*p # MVUE of the 100pth percentile 2
  var_kp1=(((x[n1]-x[1])^ 2)*((n1*(1-p)^ 2)+(n1*p^ 2)+(2*p*(1-p))))/
  ((((n1+1)^ 2)*
  (n1+2)))*(kp1^ 2)) #variance of kp1
  var_kp2=(((y[n2]-y[1])^ 2)*((n2*(1-p)^ 2)+(n2*p^ 2)+(2*p*(1-p))))/
  ((((n2+1)^ 2)*
  (n2+2)))*(kp2^ 2)) #variance of kp2
  kp_ratio=kp1/kp2 #estimated percentile ratio
  var_lnratio = (var_kp1)+(var_kp2) #variance of natural log of
  percentile ratio
  lb = kp_ratio*exp(-z*sqrt(var_lnratio)) #lower limit of percentile
  ub = kp_ratio*exp(z*sqrt(var_lnratio)) #upper limit of percentile
}
```
ratio confidence interval
list(mean(sample1),mean(sample2),kp1,kp2,kp_ratio,lb,ub) #list desired statistics
}

confidence_interval(200,100,1,3,2,5,.05,.1) # insert your values here
VITA

PIUS MUINDI

Personal Data:  Place of Birth: Machakos, Kenya
                Marital Status: Married

Education:  M.S. Mathematics (Statistics),
            East Tennessee State University,
            Johnson City, Tennessee 2008
            B.A. Mathematics and Economics,
            University of Nairobi,
            Nairobi, Kenya 1999

Professional Experience:  Graduate Assistant,
                          East Tennessee State University,
                          Johnson City, Tennessee, 2006–2008
                          Teacher, Khalsa Girls Secondary School,
                          Research Assistant, Research Solutions,
                          Nairobi, Kenya, 2001–2004