Exploring the variance of the sample variance through estimation and simulation

Christina Stradwick
christina.stradwick@gmail.com

EXPLORING THE VARIANCE OF THE SAMPLE VARIANCE THROUGH ESTIMATION AND SIMULATION

A thesis submitted to
the Graduate College of
Marshall University
In partial fulfillment of
the requirements for the degree of
Master of Arts
in
Mathematics
by
Christina Stradwick
Approved by
Dr. Laura Adkins, Committee Chairperson
Dr. Avishek Mallick
Dr. Alaa Elkadry

Marshall University
May 2019

APPROVAL OF THESIS/DISSERTATION

We, the faculty supervising the work of Christina Stradwick, affirm that the thesis, *Exploring the Variance of the Sample Variance Through Estimation and Simulation*, meets the high academic standards for original scholarship and creative work established by the Department of Mathematics and the College of Science. This work also conforms to the editorial standards of our discipline and the Graduate College of Marshall University. With our signatures, we approve the manuscript for publication.

Laura Adkins

4-15-2019

Dr. Laura Adkins, Department of Mathematics | Committee Chairperson | Date

Avishek Mallick

4-16-2019

Dr. Avishek Mallick, Department of Mathematics | Committee Member | Date

4-16-2019

Dr. Alaa Elkadry, Department of Mathematics | Committee Member | Date

ACKNOWLEDGEMENTS

I would like to give thanks, first and foremost, to God for blessing me with the opportunities and abilities that led me to where I am, today.

I would also like to thank my parents for their unwavering love and support which have undergirded me every step of the way. A special thanks goes to my mother for recognizing the importance of mathematics and teaching me the concepts at an early age. She made certain never to let on that she, herself, had an aversion to the subject so she could watch my love for mathematics grow.

Thank you to my advisor Dr. Laura Adkins for her direction and encouragement throughout the research and writing processes. She was my first statistics professor, and my statistics education at Marshall has come full circle by culminating under her supervision.

I would like to extend my heartfelt gratitude to my thesis committee members Dr. Alaa Elkadry and Dr. Avishek Mallick for all of their advice and guidance concerning my thesis and beyond. Their knowledge and wisdom have been invaluable resources.

Finally, I must say a sincere thank you to every professor at Marshall University who has sown into my life over the years. Each has had a hand in shaping who I am, and I am grateful for their contributions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In this thesis, we examine properties of the variance of the sample variance, which we will denote $V(S^2)$. We derive a formula for this variance and show that it only depends on the sample size, variance, and kurtosis of the underlying distribution. We also derive the maximum likelihood estimators for this parameter, $\hat{V}(S^2)$, under the normal, exponential, Bernoulli, and Poisson distributions and end the thesis with simulations demonstrating the distributions of these estimators.

CHAPTER 1

INTRODUCTION

In statistics, the sampling distribution of the sample variance, S^2, is known under the assumption of sampling from a normal distribution. While there are many instances in which normality of the underlying distribution is a reasonable assumption, this is not always the case. In this thesis we examine the variance of the sampling distribution of S^2 when the sample is not necessarily drawn from a normal distribution.

The goal of this thesis is to provide a formula for $V(S^2)$ in terms of the moments of the underlying distribution. The derivation of this formula for a general distribution is shown in Chapter 2 and then the formula is used to find expressions for $V(S^2)$ under the assumptions of normal, Bernoulli, exponential, and Poisson distributions. Although $V(S^2)$ for the normal distribution is a known result, we have included the derivation to show we arrive at the same conclusion.

In practice, population parameters may be unknown and collecting all samples of size n is not likely to be feasible. In these cases, $V(S^2)$ cannot be observed and calculating a theoretical value of $V(S^2)$ using the formula is not possible. These limitations necessitate a method of estimating $V(S^2)$ based on sample data. In Chapter 3 we derive maximum likelihood estimators for $V(S^2)$ and then discuss the asymptotic distributions of these MLEs.

To end, we use simulated samples from the normal, Bernoulli, exponential, and Poisson distributions to demonstrate our results. Samples of varying sizes and population parameters are used to determine how changes in each affect both the value of $V(S^2)$ and the distribution of the estimators.

CHAPTER 2

VARIANCE FORMULA

2.1 STATEMENT OF FORMULA FOR $V(S^2)$

In general, collecting all samples of size n from a population to compute $V(S^2)$ is not feasible. It is useful to have a formula for $V(S^2)$ in terms of moments that are more easily observed or calculated. It can be shown that $V(S^2)$ is dependent on the sample size, the 2nd central moment, and the 4th standardized moment (known as the variance and kurtosis, respectively) of the underlying distribution. For most common distributions, these moments can be easily derived.

Theorem 2.1.1. *If $X_1, X_2, ..., X_n$ is a random sample from a distribution with mean μ_X and variance σ_X^2, then the variance of the sampling distribution of S^2, denoted $V(S^2)$, is given by*

$$V(S_X^2) = \sigma_X^4 \left\{ \frac{1}{n} E\left[\left(\frac{X - \mu_X}{\sigma_X} \right)^4 \right] + \frac{3-n}{n(n-1)} \right\}.$$

2.2 PROOF

Assume a random sample $X_1, X_2, ..., X_n$ from a distribution with mean μ_X and variance σ_X^2. To simplify our calculations, we will work with the standardized random variables and note how the linear transformation affects the variance. We begin by laying some groundwork concerning the standardized random variables.

Let Y_i be the standardization of X_i given by $Y_i = \frac{X_i - \mu_X}{\sigma_X}$ (or alternatively, $Y_i = \frac{1}{\sigma_X} X_i - \frac{\mu_X}{\sigma_X}$). It will be useful to know the moments of Y_i. Since the Y_is are identically distributed, the kth moments of all Y_is are equivalent. The first moment, $E(Y_i)$, is 0 since Y_i is a standardization of X_i. Similarly, since the variance of a standardization is 1 and $V(Y_i) = E(Y_i^2) - [E(Y_i)]^2$, we have that $E(Y_i^2) = 1$.

It can be shown that if $Y_i = aX_i + b$ for $i = 1, 2, ..., n$ then $S_Y^2 = a^2 \cdot S_X^2$, so it is clear that $S_Y^2 = \frac{1}{\sigma_X^2} S_X^2$. Furthermore, since sample variance is an unbiased estimator of population variance, we have that $E(S_Y^2) = \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = 1$. This can also be seen directly since the sample variance, S_Y^2, is an unbiased estimator of population variance, and the variance of a standardized random variable is always 1. Now, $V(S_Y^2) = E[(S_Y^2)^2] - [E(S_Y^2)]^2$. Since $E(S_Y^2) = 1$, we have that

$$V(S_Y^2) = E[(S_Y^2)^2] - 1. \tag{2.1}$$

Of particular interest is the first term on the right-hand side. Using the computational formula for sample variance, we have that

$$E[(S_Y^2)^2] = E\left(\left[\frac{\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}}{n-1}\right]^2\right).$$

After some algebraic manipulation and using the properties of linear operators, we have

$$\begin{aligned}
E[(S_Y^2)^2] &= \frac{1}{(n-1)^2} E\left[\left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}\right)^2\right] \\
&= \frac{1}{(n-1)^2} E\left[\left(\sum_{i=1}^n Y_i^2\right)^2 - \frac{2}{n}\left(\sum_{i=1}^n Y_i^2\right)\left(\sum_{i=1}^n Y_i\right)^2 + \frac{1}{n^2}\left(\sum_{i=1}^n Y_i\right)^4\right] \\
&= \frac{1}{(n-1)^2} E\left[\sum_i \sum_j (Y_i^2 Y_j^2) - \frac{2}{n} \sum_i \sum_j \sum_k (Y_i^2 Y_j Y_k) + \frac{1}{n^2} \sum_i \sum_j \sum_k \sum_l (Y_i Y_j Y_k Y_l)\right] \\
&= \frac{1}{(n-1)^2} \left[\sum_i \sum_j E(Y_i^2 Y_j^2) - \frac{2}{n} \sum_i \sum_j \sum_k E(Y_i^2 Y_j Y_k) + \frac{1}{n^2} \sum_i \sum_j \sum_k \sum_l E(Y_i Y_j Y_k Y_l)\right].
\end{aligned}$$

We will refer to $\sum_i \sum_j E(Y_i^2 Y_j^2)$, $-\frac{2}{n} \sum_i \sum_j \sum_k E(Y_i^2 Y_j Y_k)$, and $\frac{1}{n^2} \sum_i \sum_j \sum_k \sum_l E(Y_i Y_j Y_k Y_l)$ as Term A, Term B, and Term C, respectively, and examine each term in turn.

2.2.1 TERM A

There are two possible cases to consider for Term A: either $i = j$ or $i \neq j$. There are n ways that $i = j$ and $\frac{n!}{(n-2)!}$ ways that $i \neq j$. If $i = j$, then $E(Y_i^2 Y_j^2) = E(Y_i^4)$, and if $i \neq j$ then $E(Y_i^2 Y_j^2) = E(Y_i^2)E(Y_j^2) = 1$ because Y_i and Y_j are independent. Note that $E(Y_i^4)$ is the 4th moment of Y_i and $E(Y_i^2)$ is the 2nd moment of Y_i. This gives us that

$$\sum_i \sum_j E(Y_i^2 Y_j^2) = \left[nE(Y_i^4) + \frac{n!}{(n-2)!} \right]$$
$$= n\left[E(Y_i^4) + (n-1) \right].$$

2.2.2 TERM B

In the middle term there are five cases: either i,j, and k all have distinct values; the value of $i = j$ which is distinct from k; the value of $i = k$ which is distinct from j; the value of $k = j$ which is distinct from i; or the value of $i = j = k$.

There are $n(n-1)(n-2)$ ways to get three distinct values for i,j, and k. The contribution to the sum from this case is

$$n(n-1)(n-2)E(Y_i^2)E(Y_j)E(Y_k) = n(n-1)(n-2)(1)(0)(0)$$
$$= 0.$$

There are $n(n-1)$ ways for i to equal j but be distinct from k. In this case the contribution to the sum is

$$n(n-1)E(Y_i^3)E(Y_k) = n(n-1)E(Y_i^3)(0)$$
$$= 0.$$

Similarly, if i is equal to k but distinct from j, there is no contribution to the sum.

There are $n(n-1)$ ways for k to equal j but be distinct from i. In this case the contribution to the sum is

$$\begin{aligned} n(n-1)E(Y_i^2)E(Y_j^2) &= n(n-1)(1)(1) \\ &= n(n-1). \end{aligned}$$

Finally, there are n ways for all 3 values to be equal, and in this case the contribution to the sum is $nE(Y_i^4)$.

Since only the terms from the last two cases are nonzero, we have that

$$\begin{aligned} -\frac{2}{n}\sum_i\sum_j\sum_k E(Y_i^2Y_jY_k) &= -\frac{2}{n}[nE(Y_i^4) + n(n-1)] \\ &= -2[E(Y_i^4) + (n-1)]. \end{aligned}$$

2.2.3 TERM C

For the last term in the equation for $E[(S_Y^2)^2]$, we will again consider 5 cases in terms of equivalence classes: either i, j, k, and l all fall into the same equivalence class; they all fall into distinct equivalence classes; there are 3 equivalence classes of order 2, 1, and 1; there are 2 equivalence classes of order 2 and 2; or there are 2 equivalence classes of order 3 and 1. For any case that contains an equivalence class of order 1, the independence of the random variables can be used to split $E(Y_iY_jY_kY_l)$ into a product of expected values containing a factor of $E(Y_i)$, $E(Y_j)$, $E(Y_k)$, or $E(Y_l)$. Since each of these expected values is 0, there is no contribution to the sum from these cases.

There are two nonzero cases for Term C: either i, j, k, and l all fall into the same equivalence class or there are 2 equivalence classes of order 2 and 2. There are n ways for $i = j = k = l$ to hold. In this case, $E(Y_iY_jY_kY_l) = E(Y_i^4)$, so the contribution to the sum from this case is $nE(Y_i^4)$. For the case of 2 equivalence classes both of order 2, $E(Y_iY_jY_kY_l) = E(Y_i^2)E(Y_j^2) = 1$. There are $3n(n-1)$ ways for this case to occur, so the contribution to the sum from this case is $3n(n-1)$. Putting the two nonzero cases together, we have that

$$\frac{1}{n^2}\sum_i\sum_j\sum_k\sum_l E(Y_iY_jY_kY_l) = \frac{1}{n^2}\left[nE(Y_i^4)+3n(n-1)\right]$$
$$= \frac{1}{n}\left[E(Y_i^4)+3(n-1)\right].$$

2.2.4 FINAL DERIVATION

Combining the results from section 2.2.1, 2.2.2, and 2.2.3, we have that

$$E[(S_Y^2)^2] = \frac{1}{(n-1)^2}\left\{n\left[E(Y_i^4)+(n-1)\right]-2[E(Y_i^4)+(n-1)]+\frac{1}{n}[E(Y_i^4)+3(n-1)]\right\}.$$

Using some algebra to simplify, we have

$$\begin{aligned}
E[(S_Y^2)^2] &= \frac{1}{(n-1)^2}\left[\left(n-2+\frac{1}{n}\right)E(Y_i^4)+\left(n-2+\frac{3}{n}\right)(n-1)\right] \\
&= \frac{1}{n(n-1)^2}\left[(n^2-2n+1)E(Y_i^4)+(n^2-2n+3)(n-1)\right] \\
&= \frac{1}{n(n-1)^2}\left[(n-1)^2E(Y_i^4)+(n^2-2n+3)(n-1)\right] \\
&= \frac{1}{n}E(Y_i^4)+\frac{n^2-2n+3}{n(n-1)}.
\end{aligned}$$

Substituting into Equation 2.1, gives us

$$\begin{aligned}
V(S_Y^2) &= \frac{1}{n}E(Y_i^4)+\frac{n^2-2n+3}{n(n-1)}-1. \\
&= \frac{1}{n}E(Y_i^4)+\frac{3-n}{n(n-1)} \\
&= \frac{1}{n}E\left[\left(\frac{X_i-\mu_X}{\sigma_X}\right)^4\right]+\frac{3-n}{n(n-1)}
\end{aligned}$$

Recall that $S_Y^2 = \frac{1}{\sigma_X^2} S_X^2$, thus giving us that $V(S_Y^2) = V(\frac{1}{\sigma_X^2} S_X^2) = \frac{1}{\sigma_X^4} V(S_X^2)$. This implies $V(S_X^2) = \sigma_X^4 V(S_Y^2)$. So we have

$$\begin{aligned} V(S_X^2) &= \sigma_X^4 V(S_Y^2) \\ &= \sigma_X^4 \left\{ \frac{1}{n} E\left[\left(\frac{X_i - \mu_X}{\sigma_X} \right)^4 \right] + \frac{3-n}{n(n-1)} \right\} \end{aligned} \tag{2.2}$$

Since this formula involves moments of X, only, we will henceforth drop the subscripts and refer only to the moments of X.

2.3 DERIVING $V(S^2)$ FOR SOME COMMON DISTRIBUTIONS

Using the formula to calculate the value of $V(S^2)$ for any distribution whose variance and kurtosis are known is straightforward. We have derived these formulas for the normal, Bernoulli, exponential, and Poisson distributions.

2.3.1 NORMAL

Suppose that a random sample is collected from $N(\mu, \sigma^2)$. The variance of the distribution is σ^2 and the kurtosis is equal to 3 regardless of the values of μ and σ^2. Substituting into Equation 2.2, we have

$$\begin{aligned} V(S^2) &= \sigma^4 \left[\frac{1}{n} \cdot 3 + \frac{3-n}{n(n-1)} \right] \\ &= \sigma^4 \left[\frac{3(n-1) + 3 - n}{n(n-1)} \right] \\ &= \sigma^4 \left[\frac{3n - 3 + 3 - n}{n(n-1)} \right] \\ &= \sigma^4 \left[\frac{2n}{n(n-1)} \right] \\ &= \frac{2\sigma^4}{n-1}. \end{aligned}$$

As stated previously, the distribution of S^2 under normality is a known result. The ratio $\frac{(n-1)S^2}{\sigma^2}$ follows a chi-square distribution with $(n-1)$ degrees of freedom. Thus

$V\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$ which implies

$$\begin{aligned}
\frac{(n-1)^2}{\sigma^4}V(S^2) &= 2(n-1) \\
V(S^2) &= 2(n-1)\cdot\frac{\sigma^4}{(n-1)^2} \\
V(S^2) &= \frac{2\sigma^4}{n-1}.
\end{aligned}$$

This is the same result given by our derived formula for $V(S^2)$.

2.3.2 BERNOULLI

For a random sample $X_1, X_2, ..., X_n$ of Bernoulli random variables with success probability p, the variance is pq and the kurtosis is $\frac{1-3pq}{pq}$ where $q = 1-p$. Substituting into Equation 2.2 gives

$$\begin{aligned}
V(S^2) &= (pq)^2\left[\frac{1}{n}\cdot\frac{1-3pq}{pq} + \frac{3-n}{n(n-1)}\right] \\
&= (pq)^2\left[\frac{1-3pq}{npq} + \frac{3-n}{n(n-1)}\right] \\
&= (pq)^2\left[\frac{(1-3pq)(n-1)+(3-n)pq}{npq(n-1)}\right] \\
&= (pq)^2\left[\frac{n-1-3npq+3pq+3pq-npq}{npq(n-1)}\right] \\
&= pq\left[\frac{n-1-4npq+6pq}{n(n-1)}\right] \\
&= \frac{pq}{n} + \frac{(pq)^2(6-4n)}{n(n-1)}.
\end{aligned}$$

2.3.3 EXPONENTIAL

If a random sample $X_1, X_2, .., X_n$ is taken from $Exp(\beta)$, then the variance is β^2 and the kurtosis is equal to 9. Substituting into Equation 2.2 gives us

$$\begin{aligned} V(S^2) &= \beta^4 \left[\frac{1}{n} \cdot 9 + \frac{3-n}{n(n-1)} \right] \\ &= \beta^4 \left[\frac{9(n-1)+3-n}{n(n-1)} \right] \\ &= \beta^4 \left[\frac{9n-9+3-n}{n(n-1)} \right] \\ &= \beta^4 \left[\frac{8n-6}{n(n-1)} \right]. \end{aligned}$$

2.3.4 POISSON

A random sample of size n with $X_i \sim Poisson(\lambda)$ will have variance λ and kurtosis $\frac{1+3\lambda}{\lambda}$. Substituting into Equation 2.2 yields

$$\begin{aligned} V(S^2) &= \lambda^2 \left[\frac{1}{n} \cdot \frac{1+3\lambda}{\lambda} + \frac{3-n}{n(n-1)} \right] \\ &= \lambda^2 \left[\frac{(1+3\lambda)(n-1)+(3-n)\lambda}{\lambda n(n-1)} \right] \\ &= \lambda^2 \left[\frac{n-1+3\lambda n-3\lambda+3\lambda-\lambda n}{\lambda n(n-1)} \right] \\ &= \lambda^2 \left[\frac{n-1+2\lambda n}{\lambda n(n-1)} \right] \\ &= \lambda \left[\frac{n-1+2\lambda n}{n(n-1)} \right] \\ &= \lambda \left[\frac{1}{n} + \frac{2\lambda}{(n-1)} \right] \\ &= \frac{\lambda}{n} + \frac{2\lambda^2}{n-1} \end{aligned}$$

CHAPTER 3

ESTIMATION

3.1 MAXIMUM LIKELIHOOD ESTIMATION

In practice, the true values of the parameters of the underlying distribution will be unknown which renders our equation for $V(S^2)$ impractical. Instead, we will need a method of estimating the value of $V(S^2)$ from a collected sample. In practice the invariance property of maximum likelihood estimators can be used to find the MLE of $V(S^2)$ [1].

Lemma 3.1.1 (Invariance Property of MLEs)**.** *If $\hat{\theta}$ is the maximum likelihood estimator of θ and $t(\theta)$ is any function of θ, then the maximum likelihood estimator of $t(\theta)$ is given by $t(\hat{\theta})$.*

We have used Lemma 3.1.1 to derive the MLEs for $V(S^2)$ under the assumption of sampling from the normal, Bernoulli, exponential, and Poisson distributions.

3.1.1 NORMAL

Recall that for a random sample of size n from $N(\mu, \sigma^2)$,

$$V(S^2) = \frac{2\sigma^4}{n-1}.$$

Since the MLE for σ^2 is known to be the *biased* estimator of variance, $\frac{\sum(X_i - \bar{X})^2}{n}$, we have that the MLE of $V(S^2)$ is given by

$$\hat{V}(S^2) = \frac{2}{n-1} \cdot \left[\frac{\sum(X_i - \bar{X})^2}{n}\right]^2.$$

Equivalently, we can write that

$$\hat{V}(S^2) = \frac{2(n-1)}{n^2}\left(S^2\right)^2.$$

3.1.2 BERNOULLI

For a random sample of size n from $Bernoulli(p)$,

$$V(S^2) = \frac{p(1-p)}{n} + \frac{p^2(1-p)^2(6-4n)}{n(n-1)}.$$

Using the fact that the MLE of p is the sample proportion, $\hat{p}$, we have

$$\hat{V}(S^2) = \frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{p}^2(1-\hat{p})^2(6-4n)}{n(n-1)}.$$

3.1.3 EXPONENTIAL

If our random sample of size n is from $\text{Exp}(\beta)$, we have

$$V(S^2) = \beta^4 \left[\frac{8n-6}{n(n-1)}\right].$$

Substituting the MLE of β, $\bar{X}$, gives us

$$\hat{V}(S^2) = \bar{X}^4 \left[\frac{8n-6}{n(n-1)}\right].$$

3.1.4 POISSON

If our random sample of size n comes from $Poisson(\lambda)$, then

$$V(S^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}.$$

Substituting in the MLE, $\bar{X}$, gives

$$\hat{V}(S^2) = \frac{\bar{X}}{n} + \frac{2\bar{X}^2}{n-1}.$$

3.2 DISTRIBUTION OF THE MLES

The asymptotic distribution of the MLEs is worth noting. Let $\hat{\theta}$ be the MLE of θ and let $t(\theta)$ be a differentiable function of θ. For sufficiently large n,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Big/ nE\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2}\right]}} \tag{3.1}$$

converges to a standard normal distribution under some conditions of regularity [2]. So for sufficiently large n, $t(\hat{\theta})$ is approximately normal with mean $t(\theta)$ and variance

$$\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Big/ nE\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2}\right].$$

In our case, we will take θ to be the population parameter of the underlying distribution and $t(\theta)$ to be the formula for $V(S^2)$. Also, $f(X|\theta)$ will be the density function of the underlying distribution. A similar result holds utilizing the probability mass function for discrete distributions. We will calculate Z for the Bernoulli, exponential, and Poisson distributions, first, then look at the normal distribution last.

3.2.1 BERNOULLI

Assuming we take a random sample of size n from $Bernoulli(p)$,

$$\begin{aligned}
\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 &= \left\{\frac{\partial}{\partial p}\left[\frac{p-p^2}{n} - \frac{(p^2-2p^3+p^4)(6-4n)}{n(n-1)}\right]\right\}^2 \\
&= \left\{\left[\frac{1-2p}{n} - \frac{(2p-6p^2+4p^3)(6-4n)}{n(n-1)}\right]\right\}^2 \\
&= \left\{\left[\frac{1-2p}{n} - \frac{2p(1-3p+2p^2)(6-4n)}{n(n-1)}\right]\right\}^2 \\
&= \left\{\left[\frac{1-2p}{n} - \frac{2p(1-2p)(1-p)(6-4n)}{n(n-1)}\right]\right\}^2 \\
&= \left\{(1-2p)\left[\frac{1}{n} + \frac{2p(1-p)(6-4n)}{n(n-1)}\right]\right\}^2 \\
&= (1-2p)^2\left[\frac{1}{n} + \frac{2p(1-p)(6-4n)}{n(n-1)}\right]^2.
\end{aligned}$$

Furthermore, the pmf is given by $f(x|p) = p^x(1-p)^{1-x}$ so that

$$\begin{aligned}
\ln f(x|p) &= \ln p^x + \ln(1-p)^{1-x} \\
&= x\ln p + (1-x)\ln(1-p).
\end{aligned}$$

The first partial derivative with respect to p is

$$\frac{\partial \ln f(x|p)}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p}$$

leaving the second partial derivative with respect to p to be

$$\frac{\partial^2 \ln f(x|p)}{\partial p^2} = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.$$

Negating and taking the expected value gives us

$$
\begin{aligned}
E\left[-\frac{\partial^2 \ln f(X|p)}{\partial p^2}\right] &= \frac{E(X)}{p^2} + \frac{1-E(X)}{(1-p)^2} \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} \\
&= \frac{1}{p} + \frac{1}{1-p} \\
&= \frac{1-p+p}{p(1-p)} \\
&= \frac{1}{p(1-p)}.
\end{aligned}
$$

When substituted into (3.1) we have that

$$
Z = \frac{\left[\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{p}^2(1-\hat{p})^2(6-4n)}{n(n-1)}\right] - \left[\frac{p(1-p)}{n} + \frac{p^2(1-p)^2(6-4n)}{n(n-1)}\right]}{\sqrt{\left\{(1-2p)\left[\frac{1}{n} + \frac{2p(1-p)(6-4n)}{n(n-1)}\right]\right\}^2 \Big/ \frac{n}{p(1-p)}}}
$$

and after simplification we find that

$$
Z = \frac{\left[\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{p}^2(1-\hat{p})^2(6-4n)}{n(n-1)}\right] - \left[\frac{p(1-p)}{n} + \frac{p^2(1-p)^2(6-4n)}{n(n-1)}\right]}{\left|(1-2p)\left[\frac{1}{n} + \frac{2p(1-p)(6-4n)}{n(n-1)}\right]\right| \sqrt{\frac{p(1-p)}{n}}}
$$

follows an approximate standard normal distribution for sufficiently large n.

3.2.2 EXPONENTIAL

Under the assumption of a random sample from Exp(β), $\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 = \left\{4\beta^3\left[\frac{8n-6}{n(n-1)}\right]\right\}^2$ which gives us $\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 = 16\beta^6\left[\frac{8n-6}{n(n-1)}\right]^2$. The pdf for Exp($\beta$) is $f(x|\beta) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$. So then,

$$\begin{aligned}
\ln f(x|\beta) &= \ln\left(\frac{1}{\beta}e^{-\frac{x}{\beta}}\right) \\
&= \ln\left(\frac{1}{\beta}\right) + \ln e^{\frac{x}{\beta}} \\
&= -\ln\beta - x\beta^{-1} \\
\frac{\partial \ln f(x|\beta)}{\partial \beta} &= -\beta^{-1} + x\beta^{-2} \\
\frac{\partial^2 \ln f(x|\beta)}{\partial \beta^2} &= \beta^{-2} - 2x\beta^{-3}.
\end{aligned}$$

Negating and taking the expected value on both sides yields

$$\begin{aligned}
E\left[-\frac{\partial^2 \ln f(X|\beta)}{\partial \beta^2}\right] &= -\beta^{-2} + 2\beta^{-3}E(X) \\
&= -\beta^{-2} + 2\beta^{-3}(\beta) \\
&= -\frac{1}{\beta^2} + \frac{2}{\beta^2} \\
&= \frac{1}{\beta^2}.
\end{aligned}$$

So for an *Exponential*(β) distribution,

$$\begin{aligned}
Z &= \frac{\left[\frac{8n-6}{n(n-1)}\right](\bar{X}^4 - \beta^4)}{\sqrt{16\beta^6\left[\frac{8n-6}{n(n-1)}\right]^2 \Big/ \frac{n}{\beta^2}}} \\
&= \frac{\bar{X}^4 - \beta^4}{\sqrt{\frac{16\beta^8}{n}}} \\
&= \frac{\sqrt{n}\left(\bar{X}^4 - \beta^4\right)}{4\beta^4}
\end{aligned}$$

follows an approximate standard normal distribution when n is sufficiently large.

3.2.3 POISSON

Assuming our sample is composed of iid random variables that follow $Poisson(\lambda)$,

$$\begin{aligned}\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 &= \left\{\frac{\partial}{\partial \lambda}\left[\frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\right]\right\}^2 \\ &= \left[\frac{1}{n} + \frac{4\lambda}{n-1}\right]^2.\end{aligned}$$

The mass function for a Poisson distribution is given by $f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$, so

$$\begin{aligned}\ln f(x|\lambda) &= \ln \frac{\lambda^x}{x!}e^{-\lambda} \\ &= \ln e^{-\lambda} + \ln\left(\frac{\lambda^x}{x!}\right) \\ &= -\lambda + \ln \lambda^x - \ln x! \\ &= -\lambda + x\ln\lambda - \ln x! \\ \frac{\partial \ln f(x|\lambda)}{\partial \lambda} &= -1 + \frac{x}{\lambda} \\ \frac{\partial^2 \ln f(x|\lambda)}{\partial \lambda^2} &= -\frac{x}{\lambda^2}.\end{aligned}$$

Negating and taking the expected value on both sides yields

$$\begin{aligned}E\left[-\frac{\partial^2 \ln f(X|\lambda)}{\partial \lambda^2}\right] &= E\left[\frac{X}{\lambda^2}\right] \\ &= \frac{1}{\lambda^2}E[X] \\ &= \frac{1}{\lambda^2}\cdot\lambda \\ &= \frac{1}{\lambda}.\end{aligned}$$

The Z value is then given by

$$Z = \frac{\frac{\bar{x}}{n} + \frac{2(\bar{x})^2}{n-1} - \left[\frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\right]}{\sqrt{\left[\frac{1}{n} + \frac{4\lambda}{n-1}\right]^2 \Big/ \frac{n}{\lambda}}}$$

$$= \frac{\frac{\bar{x}}{n} + \frac{2(\bar{x})^2}{n-1} - \left[\frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\right]}{\sqrt{\frac{\lambda}{n}\left[\frac{1}{n} + \frac{4\lambda}{n-1}\right]^2}}$$

$$= \frac{\frac{\bar{x}}{n} + \frac{2(\bar{x})^2}{n-1} - \left[\frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\right]}{\left[\frac{1}{n} + \frac{4\lambda}{n-1}\right]\sqrt{\frac{\lambda}{n}}}$$

where Z follows an approximate standard normal distribution for n sufficiently large.

3.2.4 NORMAL

Because the normal distribution has two parameters, the asymptotic distribution of the MLE will be an extension of the property described in Section 3.2 to the multiparameter case [3]. Under the assumption of normality, θ is the vector

$$\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$$

and although the formula for $V(S^2)$ under normality is not dependent on μ, this parameter must be considered in the calculation of Z. To standardize the MLE $\hat{V}(S^2)$ we must know the (asymptotic) mean and variance. We know $\hat{V}(S^2)$ is a consistent estimator, so the asymptotic mean is the true value of $V(S^2)$. The asymptotic variance in the multivariate case is given by the quadratic form $\frac{1}{n}\left[\frac{\partial t(\theta)}{\partial\theta}\right]^T E\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial\theta}\right]^{-1}\left[\frac{\partial t(\theta)}{\partial\theta}\right]$. The vector in this expression is

$$\left[\frac{\partial t(\theta)}{\partial\theta}\right] = \begin{bmatrix} \frac{\partial}{\partial\mu}\left(\frac{2\sigma^4}{n-1}\right) \\ \frac{\partial}{\partial\sigma^2}\left(\frac{2\sigma^4}{n-1}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{4\sigma^2}{n-1} \end{bmatrix}$$

and the matrix $E\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial\theta}\right]$ will be a 2x2 symmetric Fisher information. For the normal distribution,

$$\begin{aligned} \ln f(x|\theta) &= \ln f(x|\mu, \sigma^2) \\ &= -\frac{1}{2}\ln\sigma^2 - \ln\sqrt{2\pi} - \frac{(x-\mu)^2}{2\sigma^2} \end{aligned}$$

and the first partial derivatives are given by

$$\frac{\partial \ln f(x|\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}(x-\mu) \tag{3.2}$$

$$\frac{\partial \ln f(x|\mu,\sigma^2)}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}. \tag{3.3}$$

The partial derivative of (3.2) with respect to μ

$$\frac{\partial^2 \ln f(x|\mu,\sigma^2)}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

and the partial derivative of (3.3) with respect to σ^2

$$\frac{\partial^2 \ln f(x|\mu,\sigma^2)}{\partial(\sigma^2)^2} = \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3}$$

form the diagonal elements of the matrix $\left[\frac{\partial^2 \ln f(x|\theta)}{\partial\theta}\right]$. The off-diagonal elements are equal and are given by

$$\frac{\partial^2 \ln f(x|\mu,\sigma^2)}{\partial\sigma^2\partial\mu} = \frac{\partial^2 \ln f(x|\mu,\sigma^2)}{\partial\mu\partial\sigma^2} = -\frac{(x-\mu)}{(\sigma^2)^2}.$$

We must negate this matrix and take its expected value to arrive at the Fisher information. Since $E(X-\mu)=0$ and $E[(X-\mu)^2]=\sigma^2$, we have

$$\begin{aligned}
E\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial\theta}\right] &= E\begin{bmatrix} \frac{1}{\sigma^2} & \frac{(X-\mu)}{(\sigma^2)^2} \\ \frac{(X-\mu)}{(\sigma^2)^2} & \frac{(X-\mu)^2}{(\sigma^2)^3} - \frac{1}{2(\sigma^2)^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{\sigma^2}{(\sigma^2)^3} - \frac{1}{2(\sigma^2)^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^4} - \frac{1}{2\sigma^4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}
\end{aligned}$$

and the inverse of this matrix is easily shown to be

$$E\left[-\frac{\partial^2 \ln f(X|\theta)}{\partial\theta}\right]^{-1} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}.$$

Note that none of the elements of the vectors or matrices contain μ. This implies that, as is expected, the asymptotic variance will not depend on μ. In fact,

$$\begin{aligned}
\frac{1}{n}\left[\frac{\partial t(\theta)}{\partial\theta}\right]^T E\left[-\frac{\partial^2 \ln f(x|\theta)}{\partial\theta}\right]^{-1}\left[\frac{\partial t(\theta)}{\partial\theta}\right] &= \frac{1}{n}\begin{bmatrix} 0 & \frac{4\sigma^2}{n-1} \end{bmatrix}\begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}\begin{bmatrix} 0 \\ \frac{4\sigma^2}{n-1} \end{bmatrix} \\
&= \frac{1}{n}\cdot\frac{32\sigma^8}{(n-1)^2} \\
&= \frac{32\sigma^8}{n(n-1)^2}
\end{aligned}$$

so that in our final derivation,

$$
\begin{aligned}
Z &= \frac{\frac{2}{n-1} \cdot \left[\frac{\sum(X_i - \bar{X})^2}{n}\right]^2 - \frac{2\sigma^4}{n-1}}{\sqrt{\frac{32\sigma^8}{n(n-1)^2}}} \\
&= \frac{\frac{2}{n-1}\left\{\left[\frac{\sum(X_i - \bar{X})^2}{n}\right]^2 - \sigma^4\right\}}{\frac{4\sigma^4}{n-1}\sqrt{\frac{2}{n}}} \\
&= \frac{\left[\frac{\sum(X_i - \bar{X})^2}{n}\right]^2 - \sigma^4}{2\sigma^4\sqrt{\frac{2}{n}}}
\end{aligned}
$$

does not depend on μ. Again, Z follows an approximate standard normal distribution for n sufficiently large.

CHAPTER 4

SIMULATION

Our goal with this simulation was to see how closely simulated results lined up with our theoretical findings as well as examine how varying values of the sample size and parameters from the underlying distribution affect $V(S^2)$.

4.1 PROCEDURE

For each of the common distributions considered, the following procedure was followed:

1. Simulate 1,000 samples of size 100 each from the distribution with given values of population parameters.
2. Using our derived formula, calculate the true value of $V(S^2)$ with given parameters of the underlying distribution.
3. Calculate the 1,000 MLEs of the population parameters and use these MLEs to calculate the MLEs for $V(S^2)$.
4. Create a histogram of the MLEs of $V(S^2)$.
5. Repeat steps 1-4 two more times changing the population parameter values each time.

Once this process is completed with three different population parameter values and samples of size 100, repeat the entire process using samples of size 1,000 and then 10,000. Use the same three values of the population parameters for each value of the sample size. In total, each distribution should have 9 sets of 1,000 samples each.

4.2 NORMAL

For the normal distribution, we used the values $\sigma = 1, 10, 100$ and held $\mu = 0$ as the mean did not affect $V(S^2)$. Table 1 gives the true values of $V(S^2)$ produced when the formula derived in Chapter 2 is applied to the listed values of n and σ.

Normal	$\sigma = 1$	$\sigma = 10$	$\sigma = 100$
n=100	0.020202	202.02	2.0202e+06
n=1000	0.002002	20.02	200200
n=10000	0.00020002	2.0002	20002

Table 1: Selected Computations - Normal
True values of $V(S^2)$ with varying sample sizes and values of σ. Mean parameter, μ, assumed to be 0.

Notice that as you move across the rows of the table from left to right, the value of $V(S^2)$ increases by a factor of 10^4 which is expected since the value of σ increases by a factor of 10 and the formula for $V(S^2)$ contains a factor of σ^4. Similarly, moving down the rows of the table, the value of $V(S^2)$ decreases by a factor of 10, approximately, as n increases by a factor of 10. This comes from the division by $n - 1$ in the formula.

The histograms in Figure 1 correspond to the values of $V(S^2)$ in the same relative locations in Table 1. That is, the top leftmost histogram represents the distribution of the MLEs from samples of size 100 taken from $N(0, 1)$ and the bottom rightmost histogram represents the distribution of the MLEs from samples of size 10,000 taken from $N(0, 100, 000)$. The histograms corresponding to samples of size 100 are centered around the true value of $V(S^2)$ but are highly right skewed. Increasing the sample size to 1,000 significantly reduces the skewness apparent in the histograms, but there is still an obvious right skew. At sample sizes of 10,000 the histograms very nearly resemble a normal distribution.

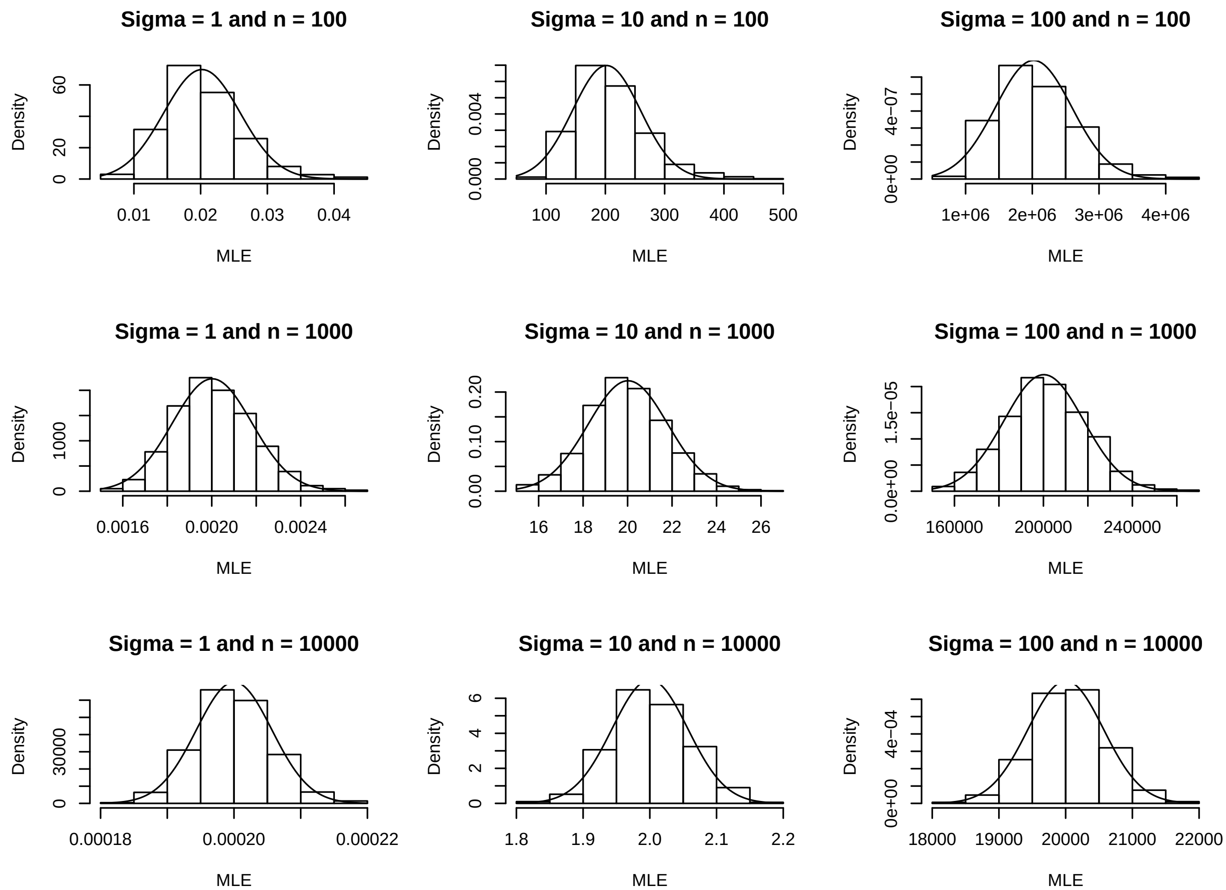

Figure 1: Histogram of Estimators - Normal
Histograms of maximum likelihood estimators for $V(S^2)$ given varying sample sizes and true values of σ while holding $\mu = 0$.

4.3 BERNOULLI

When sampling from the Bernoulli distribution, parameter values were taken to be $p = 0.1, 0.3, 0.7$. Table 2 gives the true values of $V(S^2)$ for the chosen values of p and samples of size 100, 1000, and 10000. It is worth noting that $Bernoulli(p)$ and $Bernoulli(1-p)$ produce the same value of $V(S^2)$. This can be seen in the columns of Table 2 corresponding to $p = 0.3$ and $p = 0.7$. With sample sizes as small as $n = 100$, the value of $V(S^2)$ is already less than 0.001. With sample sizes at $n = 10000$, the value of $V(S^2)$ is within 10^{-6} of being equal to 0 meaning the sampling standard deviation of S^2 is less than 0.001. This implies there will be little variation in the observed values of S^2 for large values of n.

Bernoulli	p = 0.1	p = 0.3	p = 0.7
n=100	0.000578	0.000345	0.000345
n=1000	0.000058	.000034	0.000034
n=10000	0.000006	0.000003	0.000003

Table 2: Selected Computations - Bernoulli
True values of $V(S^2)$ for given values of n and p.

Just as was the case for the normal distribution, the relative positions of the histograms in Figure 2 match those of the true values of $V(S^2)$ in Table 2. Recall that in the histograms corresponding to smaller sample sizes for the normal distribution, the distribution was right skewed. For the Bernoulli distribution, however, the distributions appear to be left skewed. Even sample sizes as large as $n = 10000$ do not appear to be large enough to overcome this skewness. As mentioned earlier, the true values of $V(S^2)$ are equal for the histograms in columns 2 and 3 for each row. The histograms for $p = 0.3$ and $p = 0.7$ while do not fully resemble a normal distribution when $n = 100$, but the shapes do resemble each other.

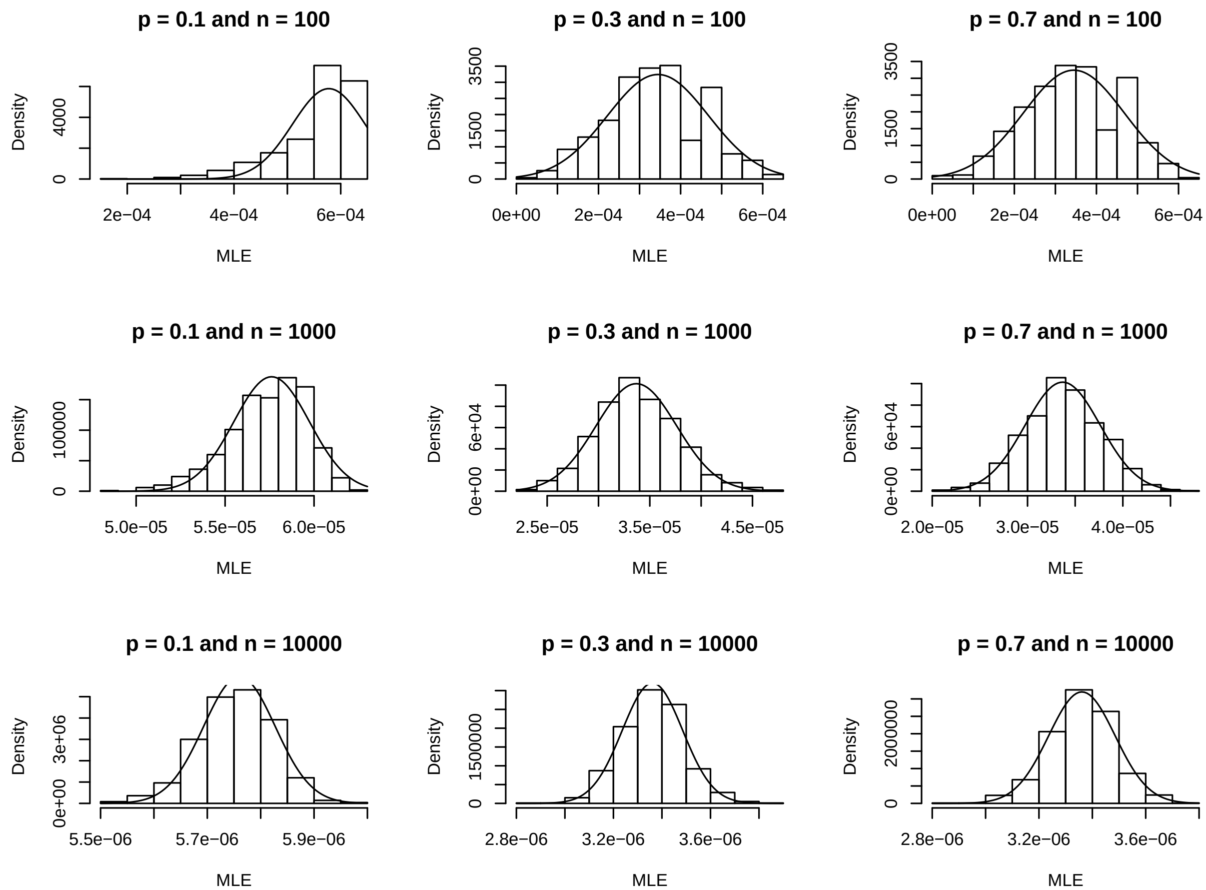

Figure 2: Histogram of Estimators - Bernoulli
Histograms of maximum likelihood estimators for $V(S^2)$ given varying sample sizes and true values of p. (Note: The histograms are labeled as binomial since each of the samples drawn from $Bernoulli(p)$ makes up a binomial sample with n trials.)

4.4 EXPONENTIAL

Samples for the exponential were taken from distributions with mean values $\beta = 1, 10, 100$. Using our derived formula, $V(S^2)$ was calculated for these distributions. These values can be found in Table 3. Moving across the rows of the table, $V(S^2)$ increases by a factor of 10^4 as β increases by a factor of 10. As n increases by a factor of 10 down the columns of the table, the value of $V(S^2)$ decreases by approximately the same factor.

Because of the exponential distribution's applications in modelling the time before an event occurs, it is possible to have large values of the mean. The column corresponding to $\beta = 100$

Exponential	$\beta = 1$	$\beta = 10$	$\beta = 100$
n=100	0.080202	802.02	8020200
n=1000	0.008002	80.02	800200
n=10000	0.0008	8.0002	80002

Table 3: Selected Computations - Exponential
True values of $V(S^2)$ for samples of size n taken from $Exponential(\beta)$.

shows that even with large values of n, $V(S^2)$ is still quite large. When μ is large, different samples will have widely varying sample variances.

The histograms in Figure 3 correspond to the true values of $V(S^2)$ in the same relative position in Table 3. The histograms of the MLEs when $n = 100$ are highly right skewed. As is expected, the distributions become less skewed as n increases. At $n = 10000$, the histograms appear to be approximately normal.

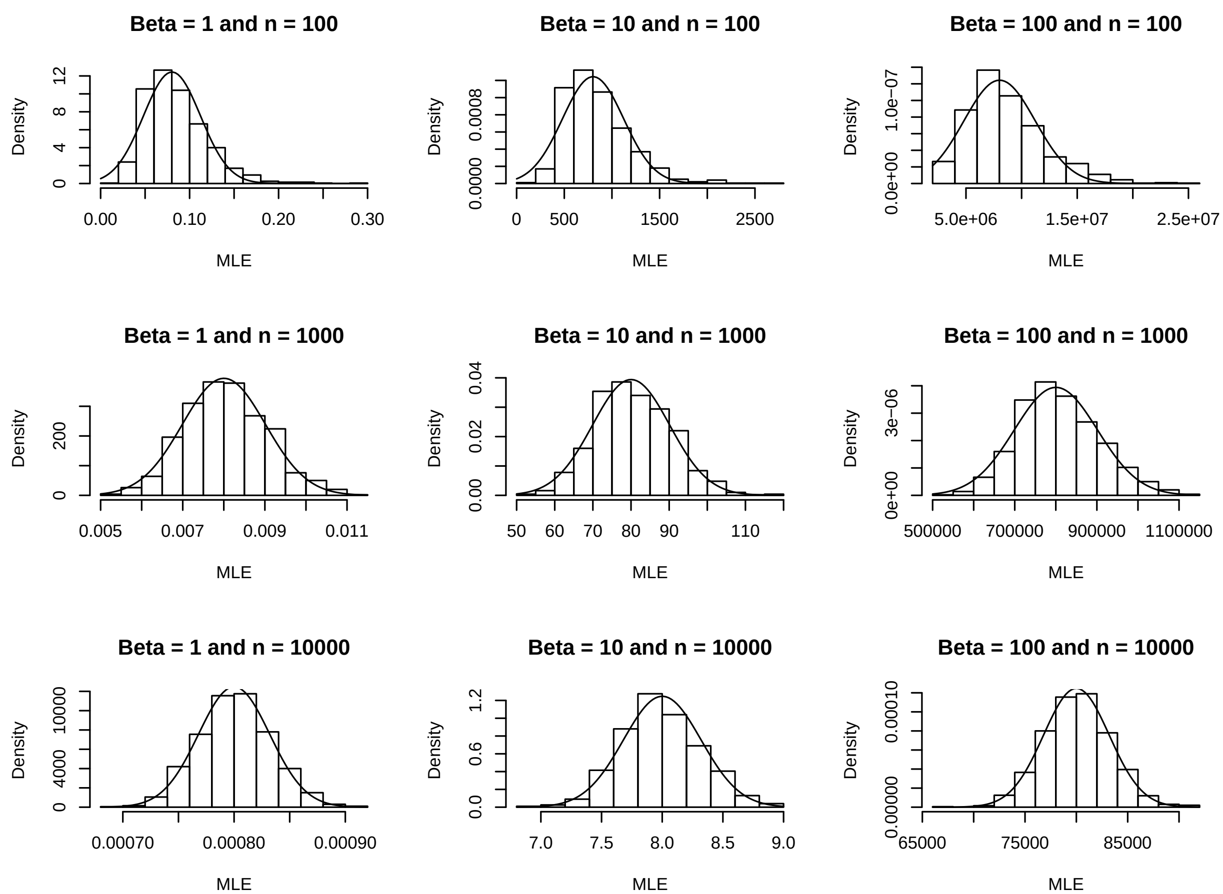

Figure 3: Histogram of Estimators - Exponential
Histograms of maximum likelihood estimators for $V(S^2)$ given varying sample sizes and true values of β.

4.5 POISSON

The three chosen parameter values for the Poisson distribution were $\lambda = 10, 30, 100$. The true values of $V(S^2)$ are given in Table 4. As λ increases, $V(S^2)$ appears to increase by approximately the square of the same factor. For example, when λ increases by a factor of 3 from $\lambda = 10$ to $\lambda = 30$, $V(S^2)$ increases by approximately a factor of 9. On the other hand, just as we have seen with each of the other distributions, the value of $V(S^2)$ decreases by approximately the same factor by which the sample size increases. It is expected that increasing the sample size will decrease $V(S^2)$ because S^2 should approach the population variance as n gets larger.

Poisson	$\lambda = 10$	$\lambda = 30$	$\lambda = 100$
n=100	2.1202	18.4818	203.02
n=1000	0.2102	1.8318	20.12
n=10000	0.021002	0.183018	2.0102

Table 4: Selected Computations - Poisson
True values of $V(S^2)$ for given sample sizes and values of λ.

The histograms in Figure 4 correspond to the values of $V(S^2)$ in the same relative positions in Table 4. Of all the histograms produced in our simulation, those for the Poisson distributions appear to be the least skewed. Interestingly, the histograms from samples sizes of $n = 10000$ are very close to a normal curve, but each histogram has two or three observations that fall into one of the tails giving it a slight skew. Each histogram does appear to be centered at the true value of $V(S^2)$ regardless of parameter values or sample size.

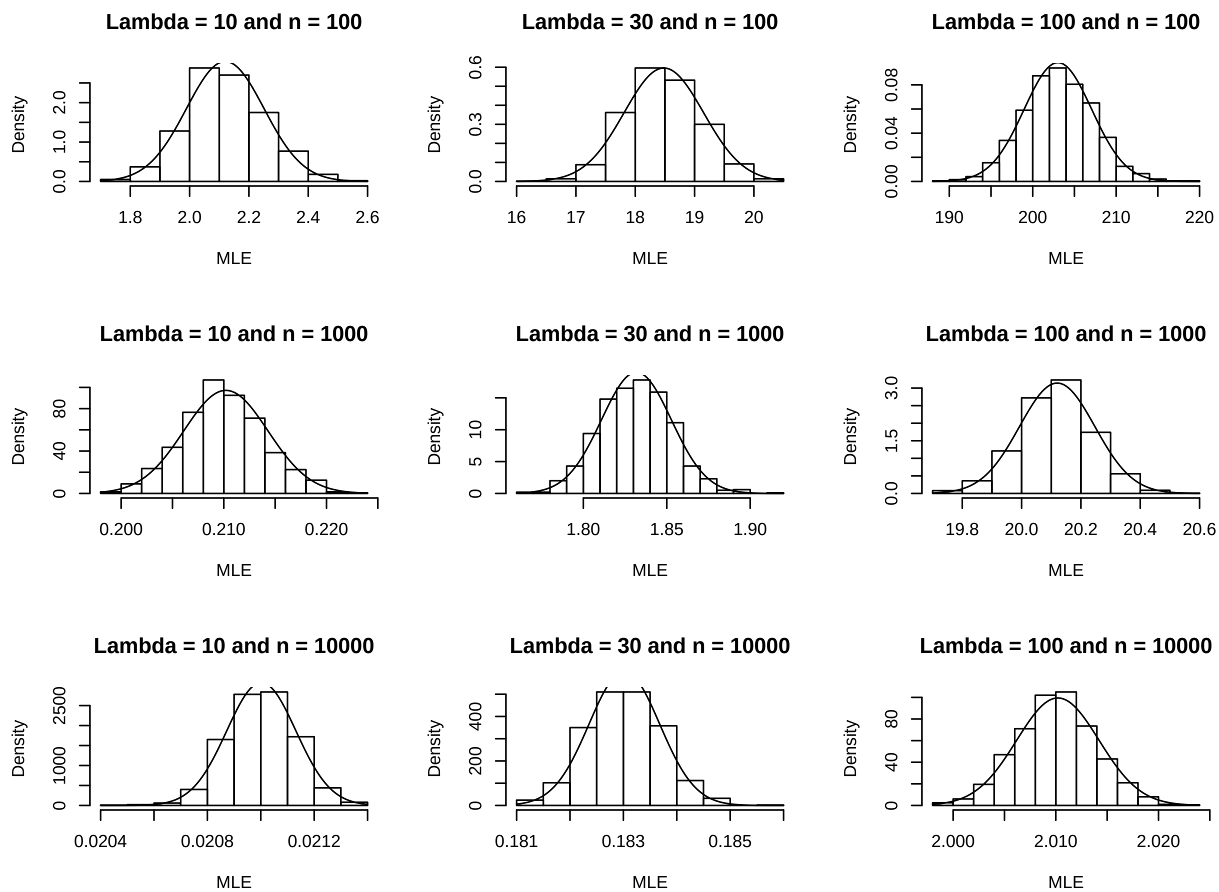

Figure 4: Histogram of Estimators - Poisson
Histograms of maximum likelihood estimators of for $V(S^2)$ given varying sample sizes and true values of λ.

CHAPTER 5

CONCLUSION

In this work it is shown that the variance of the sample variance can be calculated without the assumption of sampling from the normal distribution. This value, $V(S^2)$, is dependent only on the sample size and the variance and kurtosis of the distribution from which the samples are drawn. When maximum likelihood estimators of the parameters of the underlying distribution can be derived, the invariance property of MLEs can be used to come up with MLEs for $V(S^2)$. The asymptotic distributions of these MLEs are discussed as well.

Simulation from the normal, Bernoulli, exponential, and Poisson distributions was used to show that the MLEs behaved in the way we expected. Each of the histograms shown in Figures 1, 2, 3, and 4 is centered around the true value of $V(S^2)$ recorded in Table 1, 2, 3, or 4. The value of n has a significant affect on the value of $V(S^2)$. It is well known that as n increases, the variance of a sample approaches the population variance. Intuitively, this would lead $V(S^2)$ to decrease as the sample size gets large.

REFERENCES

[1] R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, 222(594–604):309–368, 1922.

[2] Dennis D. Wackerly, William Mendenhall III, and Richard L. Scheaffer. *Mathematical Statistics with Applications*. Brooks/Cole, Belmont, California, 2008.

[3] E.L. Lehmann. *Theory of Point Estimation*. John Wiley & Sons, Inc., New York, New York, 1983.

[4] Catherine Forbes, Merran Evans, Nicholas Hastings, and Brian Peacock. *Statistical Distributions*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2011.

APPENDIX A

LETTER FROM INSTITUTIONAL RESEARCH BOARD

MARSHALL UNIVERSITY.
www.marshall.edu

Office of Research Integrity

February 13, 2019

Christina Stradwick
1707 Doulton Avenue
Huntington, WV 25701

Dear Ms. Stradwick:

This letter is in response to the submitted thesis abstract entitled *"Exploring the Variance of the Sample Variance through Estimation and Simulation."* After assessing the abstract, it has been deemed not to be human subject research and therefore exempt from oversight of the Marshall University Institutional Review Board (IRB). The Code of Federal Regulations (45CFR46) has set forth the criteria utilized in making this determination. Since the information in this study does not involve human subjects as defined in the above referenced instruction, it is not considered human subject research. If there are any changes to the abstract you provided then you would need to resubmit that information to the Office of Research Integrity for review and a determination.

I appreciate your willingness to submit the abstract for determination. Please feel free to contact the Office of Research Integrity if you have any questions regarding future protocols that may require IRB review.

Sincerely,

Bruce F. Day, ThD, CIP
Director

WE ARE... MARSHALL.

One John Marshall Drive • Huntington, West Virginia 25755 • Tel 304/696-4303
A State University of West Virginia • An Affirmative Action/Equal Opportunity Employer

APPENDIX B

TABLE OF MOMENTS

	$E(X) = \mu$	$E\left[(X-\mu)^2\right] = \sigma^2$	$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$	$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$
Normal	μ	σ^2	0	3
Bernoulli	p	pq	$\frac{1-2p}{\sqrt{pq}}$	$\frac{1-3pq}{pq}$
Exponential	β	β^2	2	9
Poisson	λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1+3\lambda}{\lambda}$

Table B.1: Moments of Common Distributions
Mean, variance, and 3rd and 4th standardized moments for common distributions. More information regarding the moments of these and other distributions can be found in [4].