Some Aspects of Estimators for Variance of Normally Distributed Data

N. Hemachandra and Puja Sahu

nh@iitb.ac.in; puja.sahu@iitb.ac.in

Industrial Engineering and Operations Research
Indian Institute of Technology Bombay, Mumbai, India

First version: November, 2013; Second version: March, 2014

Abstract

Normally distributed data arise in various contexts and often one is interested in estimating its variance. We limit ourselves to the class of estimators that are (positive) multiples of sample variances. Two important qualities of the estimators are bias, which captures the accuracy of the estimator, and variance, which captures the estimator's precision. Apart from the two standard estimators for variance of such normally distributed data, we also consider the one that minimizes the mean square error and another which minimizes the maximum of the square of the bias and variance, the minmax estimator. This minmax estimator can be identified as fixed point of a suitable function. Our observation is that, for moderate to large sample sizes, all these estimators have the same order of the mean square error, (which is the inverse of the sample size). We next consider another criterion for quality of an estimator – the fraction of square of bias in mean square error of the estimator. For UMVUE, this fraction is zero; while for the minmax estimator this fraction is half. So, while these estimators are quite similar on mean square error aspect, they differ in the contribution of bias to their mean square error. Another framework to compare the estimators is the Pareto efficient frontier with the components, squared bias and variance of the estimators. The classical estimators, UMVUE, optimal MSE and minmax estimators are non-dominated feasible points in this space i.e., lie on the Pareto frontier which we also identify.

1 Introduction

It is well known that normally distributed data sets are observed in numerous situations. One major reason for this is the following situation. Suppose one has a random error which is the aggregate of a large collection of errors. Then, under mild conditions, by the classical Central Limit Theorem and its variants [Billingsley, 1995], [Chung, 2001], [Fristedt and Gray, 1997], [Kallenberg, 2002], [Wasserman, 2005], etc., the standardized sum of the collections of errors (and hence suitable scaling of the centered random error) has approximately the distribution of a zero mean normal random variable.

A central theme in statistical inference is that given a sample from a parametric distribution, one is interested in finding a suitable 'best' estimator for a parameter of the distribution. Most of such inference procedures concentrate on the unbiased estimators and finding the 'best' (i.e., the one having the minimum variance) amongst them. These are the classical Uniformly Minimum Variance Unbiased Estimators, (UMVUE) [Casella and Berger, 2002], [DeGroot and Schervish, 2012], etc. However, one can trade-off the bias of the estimator to achieve lower variance and hence find a better estimator in terms of Mean Squared Error (MSE) as MSE is the sum of squared bias and variance of the estimator, leading to the optimal MSE estimator. Further, one can view both squared bias and variance of an estimator as equally important and hence search for an estimator that minimizes the maximum of these two (undesirable) quantities, the minmax estimator.

In fact, one can view MSE as capturing a quality of an estimator and hence compare various estimators on the basis of their MSEs. Also, one can compare estimators in terms of the percentage of squared bias in MSE. Yet an another way to compare estimators is to view this comparison as a multi-criteria problem involving squared bias and variance and then search for those estimators that are Pareto optimal: the set of estimators such that reducing of one of these quantities leads to increase in the other quantity.

In this chapter, we illustrate the above aspects of estimators and various measures of quality of estimators when the underlying data is normally distributed and the parameter we are interested is the variance of this normal random variable.

Consider a random sample $(X_1, X_2, \ldots, X_n)$ of size n from a $N(\mu, \sigma^2)$ distribution and consider the two cases for estimation of population variance (σ^2): μ known and μ unknown. For μ known case, the classical unbiased estimator is

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

But one can also consider the following estimators of σ^2:

$$\overline{S}_c^2 = c\sum_{i=1}^{n}(X_i - \mu)^2, \tag{2}$$

parametrized by coefficients, $c > 0$. $\overline{S}_c^2$ can be viewed as scalings of $\hat{\sigma}^2$.

Similarly for the μ unknown case, we decided to look for estimator for σ^2 of the form

$$S_c^2 = c\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad c > 0. \tag{3}$$

It is assumed that the sample size, n, is at least two. Also, we can restrict ourselves to $c > 0$ as estimators of this nature dominate the zero estimator corresponding to $c = 0$, both on MSE as well as minmax criteria. Details on this and related points are in given in the appendix.

With $c > 0$, it is known [Casella and Berger, 2002], since χ_n^2 is a random variable of location-scale exponential family, that

$$\frac{\overline{S}_c^2}{c\sigma^2} \sim \chi_n^2, \tag{4}$$

$$\frac{S_c^2}{c\sigma^2} \sim \chi_{n-1}^2. \tag{5}$$

2 Different risk criteria and estimators based on them

A loss function measures the quality of an estimator. Bias of an estimator can be interpreted as capturing its *accuracy* while variance can be interpreted as measuring its *precision*, [Casella and Berger, 2002], [Alpaydin, 2010]. One of the most common methods is minimizing the variance over the class of unbiased estimators. This gives us what is called the Uniform Minimum Variance Unbiased Estimator (UMVUE). Another popular loss function is mean squared error (MSE) which is the sum of square of the bias and variance, [Casella and Berger, 2002]. Minimizing MSE can be interpreted as minimizing the weighted average of square of bias and variance where weights are equal. Treating each of the squared bias and the variance of an estimator as dissatisfactions associated with that estimator, minimizing the maximum of them can be viewed as attempt towards achieving a certain notion of fairness (towards squared bias and variance); thus, a minmax estimator can be viewed as an estimator with this property. Also, both the optimal MSE estimator and minmax estimator are *biased* estimators, unlike UMVUE.

2.1 Uniformly Minimum Variance Unbiased Estimator (UMVUE)

The UMVUE is, as the name suggests, the estimator that has the minimum variance among the unbiased estimators for the parameter of interest. Among the two basic measures of the quality of an estimator, the bias is more important factor for UMVUE than the variance. Hence, first the bias is brought down to its minimum possible value, that is zero, and then the 'best' is picked from this class of estimators with the minimum value of bias.

However, the existence of an unbiased estimator can not always be guaranteed. As pointed out by [Doss and Sethuraman, 1989], when an unbiased estimator does not exist, any attempt to reduce the bias below a given value can result in substantial increment in the variance, thereby providing an even worse estimator on the MSE grounds.

The UMVUE can be obtained analytically by identifying the coefficient, c, as a function of sample size, n, (for a given n) at which squared bias, B_c^2, becomes zero. Note that we could uniquely pin down the coefficient c (and hence the UMVUE) because the set of unbiased estimators for σ^2 is a singleton for both the cases: μ known and μ unknown.

(i) **Case: μ known**

$$
\begin{aligned}
& B_c = 0 \\
\Rightarrow\ & E(\overline{S}_c^2) - \sigma^2 = 0 \\
\Rightarrow\ & c\sigma^2 E\left(\frac{\overline{S}_c^2}{c\sigma^2}\right) - \sigma^2 = 0 \\
\Rightarrow\ & (nc-1)\sigma^2 = 0 \qquad \text{(using (4))} \\
\Rightarrow\ & c = \frac{1}{n}
\end{aligned} \tag{6}
$$

Thus, the UMVUE for σ^2 in the μ known case is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

(ii) **Case: μ unknown**

$$
\begin{aligned}
& B_c = 0 \\
\Rightarrow\ & E(S_c^2) - \sigma^2 = 0 \\
\Rightarrow\ & c\sigma^2 E\left(\frac{S_c^2}{c\sigma^2}\right) - \sigma^2 = 0 \\
\Rightarrow\ & [(n-1)c - 1]\sigma^2 = 0 \qquad \text{(using (5))} \\
\Rightarrow\ & c = \frac{1}{n-1}
\end{aligned} \tag{7}
$$

The UMVUE for σ^2 in the μ unknown case is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Suppose, for a given sample size, we plot squared bias values on the horizontal axis and variance values on the vertical axis. Then the UMVUE can be determined on the graph as the point where the graph touches the vertical axis, i.e., where the squared bias (and hence the bias) becomes zero. The graphical method for finding the UMVUE for σ^2 for the case of known mean, μ, gives us the estimators shown in Figure 1.

2.2 The optimal MSE estimator

The UMVUE gives the best estimator in the class of unbiased estimators. But an optimal MSE estimator tries to minimize the Mean Squared Error (MSE) which is the sum of squared bias and variance of the estimator. Thus, the optimal MSE estimator has lower MSE value than the UMVUE since it has significant decrease in the variance as compared to the increment in the bias value.

MSE takes into account both the squared bias and variance and weighs them equally. Hence, it makes sure that the performance of the estimator does not overlook the effects of any one of these components. This helps us extend the domain of comparison to the biased estimators as well, unlike the case of UMVUE, where by definition, we look only for unbiased estimators and choose the optimal amongst them.

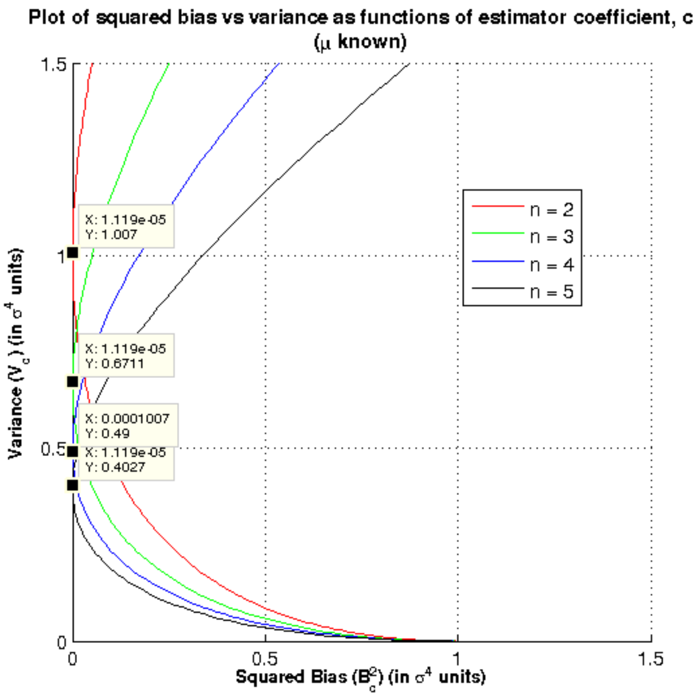


Figure 1: UMVUE for normal variance (σ^2) for different sample sizes (n)

We use the MSE value as criterion for comparing different estimators in a later section. Hence, as a baseline case, we now derive the optimal MSE estimator here. But before that we define the bias and the variance of an estimator, denoted by B_c and V_c, respectively. First consider the μ known case. The bias B_c for an estimator $\overline{S}_c^2, (c > 0)$ is obtained as:

$$
\begin{aligned}
B_c &= E(\overline{S}_c^2) - \sigma^2 \\
\Rightarrow B_c &= c\sigma^2 E\left(\frac{\overline{S}_c^2}{c\sigma^2}\right) - \sigma^2 \\
\Rightarrow B_c &= (nc-1)\sigma^2. \qquad (8)
\end{aligned}
$$

And the variance of $\overline{S}_c^2$ is given by:

$$
\begin{aligned}
V_c &= \mathrm{Var}(\overline{S}_c^2) \\
\Rightarrow V_c &= c^2\sigma^4 \,\mathrm{Var}\left(\frac{\overline{S}_c^2}{c\sigma^2}\right) \\
\Rightarrow V_c &= 2nc^2\sigma^4. \qquad (9)
\end{aligned}
$$

Similarly for an estimator $S_c^2, (c > 0)$ of σ^2 in the μ unknown case, we obtain the bias and variance values as:

$$
\begin{aligned}
B_c &= [(n-1)c - 1]\sigma^2, \qquad (10) \\
V_c &= 2(n-1)c^2\sigma^4. \qquad (11)
\end{aligned}
$$

Using the mean and variance of chi-squared distribution from (4) and (5), MSE can be calculated as:

(i) **Case: μ known**

$$\begin{aligned}
\mathrm{MSE}(\overline{S}_c^2) &= (\mathrm{bias}(\overline{S}_c^2))^2 + \mathrm{Var}(\overline{S}_c^2) \\
&= \left[nc\sigma^2 - \sigma^2\right]^2 + 2nc^2\sigma^4 \\
&= \sigma^4[(n^2+2n)c^2 - 2nc + 1]. \qquad (12)
\end{aligned}$$

The minimum MSE estimator can be found by setting

$$\begin{aligned}
&\frac{\partial}{\partial c}\mathrm{MSE}(\overline{S}_c^2) = 0 \\
&\Rightarrow \sigma^4[2c(n^2+2n) - 2n] = 0 \\
&\Rightarrow c^*_{\mathrm{MSE}} = \frac{1}{n+2}. \quad (\because n \geq 1, \sigma^4 \neq 0) \qquad (13)
\end{aligned}$$

c^*_{MSE} is a point of minimum since

$$\frac{\partial^2}{\partial c^2}\mathrm{MSE}(\overline{S}_c^2) = \sigma^4(n^2+2n) > 0 \quad (\because n \geq 1, \sigma^2 > 0).$$

Thus the minimum MSE estimator for variance of normal distribution when mean is known is

$$\overline{S}^2_{c^*_{\mathrm{MSE}}} = \frac{1}{n+2}\sum_{i=1}^{n}(X_i - \mu)^2. \qquad (14)$$

(ii) **Case: μ unknown** By similar analysis as in previous case,

$$\begin{aligned}
\mathrm{MSE}(S_c^2) &= [E(S_c^2) - \sigma^2]^2 + \mathrm{Var}(S_c^2) \\
&= \left[(n-1)c\sigma^2 - \sigma^2\right]^2 + 2(n-1)c^2\sigma^4 \\
&= \sigma^4[(n^2-1)c^2 - 2(n-1)c + 1]. \qquad (15)
\end{aligned}$$

The minimum MSE estimator in this case is obtained as:

$$\begin{aligned}
&\frac{\partial}{\partial c}\mathrm{MSE}(S_c^2) = 0 \\
&\Rightarrow \sigma^4[2c(n^2-1) - 2(n-1)] = 0 \\
&\Rightarrow c^*_{\mathrm{MSE}} = \frac{1}{n+1} \quad (\because n > 1, \sigma^4 \neq 0). \qquad (16)
\end{aligned}$$

Thus the minimum MSE estimator for variance of normal distribution when mean is unknown is

$$S^2_{c^*_{\mathrm{MSE}}} = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2. \qquad (17)$$

2.2.1 Graphical Illustration

We can identify the optimal MSE estimator for a given sample size, n, graphically by plotting the MSE values (in σ^4 units) versus the coefficient values and picking the coefficient value that has the minimum MSE value on this plot.

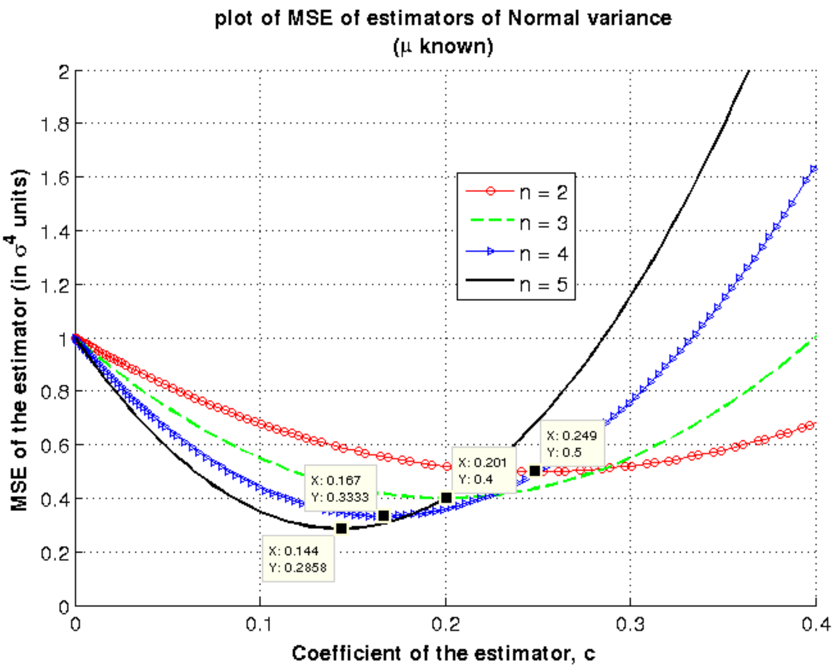


Figure 2: Coefficient of optimal MSE estimator of σ^2 for different sample sizes (n)

We consider the case where the mean, μ, is known and identify the optimal MSE estimator as described above (See Figure 2). The marked points are the coordinates corresponding to coefficient of optimal MSE estimator, c^*_{MSE}, and its MSE value, $\text{MSE}(\overline{S}^2_{c^*_{\text{MSE}}})$, which are:

$$c^*_{\text{MSE}} = \frac{1}{n+2} \quad \text{and} \quad \text{MSE}\left(\overline{S}^2_{c^*_{\text{MSE}}}\right) = \frac{2\sigma^4}{n+2}.$$

Hence, as n increases, c^*_{MSE} and $\text{MSE}(\overline{S}^2_{c^*_{\text{MSE}}})$ decrease in value, which is also verified from the graph.

Consider the graph of squared bias versus variance values for estimators with coefficients, c in Figure 3 and Figure 4. To spot the optimal MSE estimator on this graph, we draw a line with slope -1 and translate it so at we get a tangent to the curve. At this point of intersection, sits the optimal MSE estimator.

2.3 The min max estimator

Now, instead of using MSE criterion which takes sum of the squared bias and the variance of the estimator, we propose to look at these quantities as vector $\begin{pmatrix} B_c^2 \\ V_c \end{pmatrix}$ and try to find an estimator that compares the two components on the same scale by using min max criterion: $\min_c \max\{B_c^2, V_c\}$. Note that we are considering the estimators only with $c > 0$.

Squared bias and variance are two different criteria for comparing the estimators; yet it is difficult to weigh one against the other. Even though MSE, as a sum of squared bias and variance of the estimator, gives equal weightage to both of them, the minmax criterion explicitly treats squared bias and variance as equally important and competing risks. Such a minmax estimator minimizes the maximum of squared bias and variance among all the estimators of form given by (2) or (3), as the case may be. The minmax estimator views both squared bias and variance in a more equitable manner. We also interpret the minmax estimator as a fixed point for a suitable function.

(i) **Case: μ known**

$$\begin{aligned}\min_c \max_{B_c^2, V_c}(\overline{S}_c^2) &= \min_c \max\{(nc-1)^2\sigma^4, 2nc^2\sigma^4\}\\ &= \sigma^4 \min_c \max\{(nc-1)^2, 2nc^2\}.\end{aligned}$$

since σ^4 is just a positive scaling factor. Now,

$$\begin{aligned}&B_c^2 \geq V_c\\ \Leftrightarrow\ &(nc-1)^2 \geq 2nc^2\\ \Leftrightarrow\ &(n^2-2n)c^2 - 2nc + 1 \geq 0.\end{aligned}$$

For a fixed n, the RHS term is a quadratic in c, whose sign can be deduced by looking at its roots. The roots of this quadratic are given as:

$$c = \frac{\sqrt{n} \pm \sqrt{2}}{\sqrt{n}(n-2)}.$$

Thus, from the definitions of B_c^2 and V_c, one has that

$$\begin{aligned}B_c^2 \geq V_c \quad &\text{when} \quad 0 \leq c \leq \frac{\sqrt{n}-\sqrt{2}}{\sqrt{n}(n-2)} \quad \text{or} \quad c \geq \frac{\sqrt{n}+\sqrt{2}}{\sqrt{n}(n-2)},\\ \text{and, } B_c^2 < V_c \quad &\text{when} \quad \frac{\sqrt{n}-\sqrt{2}}{\sqrt{n}(n-2)} < c < \frac{\sqrt{n}+\sqrt{2}}{\sqrt{n}(n-2)}.\end{aligned}$$

The roots $\frac{\sqrt{n}-\sqrt{2}}{\sqrt{n}(n-2)}$ and $\frac{\sqrt{n}+\sqrt{2}}{\sqrt{n}(n-2)}$ are the points where $B_c^2 = V_c$. So the minmax estimator is chosen by comparing the variance values at these roots. (We chose variance values over squared bias values for the ease of comparison.)

Say the two roots are c_1 and c_2, then

$$\begin{aligned}V_{c_1} &= 2nc_1^2 = \frac{2(n+2-2\sqrt{2n})}{(n-2)^2},\\ V_{c_2} &= 2nc_2^2 = \frac{2(n+2+2\sqrt{2n})}{(n-2)^2}.\end{aligned}$$

Clearly, $V_{c_1} < V_{c_2}$ for $n \geq 1$ except $n = 2$.

Hence, the coefficient of the min max estimator is

$$c^*_{\text{mm}} = \frac{\sqrt{n}-\sqrt{2}}{\sqrt{n}(n-2)} = \frac{1}{\sqrt{n}(\sqrt{n}+\sqrt{2})}, \quad \text{for } n \neq 2. \tag{18}$$

Special case: $n = 2$

$$\begin{aligned}
&B_c^2 \geq V_c \\
&\Leftrightarrow (2c-1)^2 \geq 4c^2 \\
&\Leftrightarrow 1 - 4c \geq 0 \\
&\Rightarrow c^*_{\text{mm}} = \frac{1}{4}.
\end{aligned}$$

This is same as that given by $c^*_{\text{mm}} = \dfrac{\sqrt{n}-\sqrt{2}}{\sqrt{n}(n-2)} = \dfrac{1}{\sqrt{n}(\sqrt{n}+\sqrt{2})}$ for $n = 2$. Also, $c^*_{\text{MSE}} = \dfrac{1}{n+2} = \dfrac{1}{4}$. Thus, the optimal MSE estimator and the min max estimator for variance when mean is known are same when sample size is 2.

(ii) **Case: μ unknown**

$$\min_c \max_{B_c^2, V_c}(S_c^2) = \sigma^4 \min_c \max\{[(n-1)c-1]^2, 2(n-1)c^2\}.$$

$$\begin{aligned}
&B_c^2 \geq V_c \\
&\Leftrightarrow [(n-1)c-1]^2 \geq 2(n-1)c^2 \\
&\Leftrightarrow (n^2-4n+3)c^2 - 2(n-1)c + 1 \geq 0
\end{aligned}$$

The roots of the above quadratic equation are:

$$c = \frac{(\sqrt{n-1}) \pm \sqrt{2}}{(\sqrt{n-1})(n-3)}.$$

As in the previous case, we compare the variance values at these two roots to choose the coefficient of the minmax estimator. Say the two roots are c_1 and c_2, then

$$\begin{aligned}
V_{c_1} &= 2(n-1)c_1^2 = \frac{2(n+1-2\sqrt{2(n-1)})}{(n-3)^2}, \\
V_{c_2} &= 2(n-1)c_2^2 = \frac{2(n+1+2\sqrt{2(n-1)})}{(n-3)^2}.
\end{aligned}$$

Comparing the numerators we can say that $V_{c_1} < V_{c_2}$ for all $n \geq 2$ except $n = 3$. Hence,

$$c^*_{\text{mm}} = \frac{(\sqrt{n-1})-\sqrt{2}}{(\sqrt{n-1})(n-3)} = \frac{1}{(\sqrt{n-1})(\sqrt{n-1}+\sqrt{2})}, \quad \text{for } n \neq 3. \tag{19}$$

Special case: $n = 3$

$$\begin{aligned}
&B_c^2 \geq V_c \\
&\Leftrightarrow 1 - 4c \geq 0 \\
&\Rightarrow c^*_{\text{mm}} = \frac{1}{4}.
\end{aligned}$$

Again, this is same as that given by $c^*_{\text{mm}} = \dfrac{(\sqrt{n-1})-\sqrt{2}}{(\sqrt{n-1})(n-3)} = \dfrac{1}{(\sqrt{n-1})(\sqrt{n-1}+\sqrt{2})}$ for $n = 3$. Also, $c^*_{\text{MSE}} = \dfrac{1}{n+1} = \dfrac{1}{4}$.

2.3.1 min max **estimator as a fixed point**

The min max estimator for a given sample size, n, can be obtained as the fixed point of a function, $f(c)$.

(i) For μ known case, the function f is

$$f(c) = \frac{nc - 1}{\sqrt{2n}}.$$

One can observe that, for a min max estimator, squared bias and variance values are equal and obtain f from here.

$$\begin{aligned} &B_c^2 = V_c \\ &\Rightarrow 2nc^2 = (nc-1)^2 \\ &\Rightarrow c = \frac{nc-1}{\sqrt{2n}}. \end{aligned}$$

(ii) For μ unknown case, the desired function can be similarly obtained as

$$\tilde{f}(c) = \frac{(n-1)c - 1}{\sqrt{2(n-1)}}.$$

2.3.2 Graphical Illustration

The min max estimator for σ^2 is obtained by looking for the intersection point of the curve with a line whose slope is +1 and passes through origin of the graph since this line will have equal values for squared bias and variance.

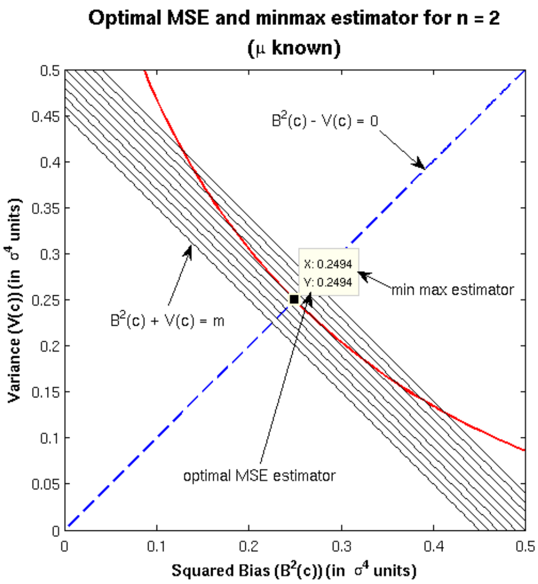
Figure 3 shows the min max estimator for a few sample sizes when μ is known. Note that, there may be more than one (ideally, two) intersection points of the curve with the line. We pick the one that is closest to the origin, since it corresponds to the minimum value amongst all the estimators with equal squared bias and variance values. Similarly, we have plotted the mimmax estimator for σ^2 when μ is unknown in the Figure 4.

3 Comparison of the estimators

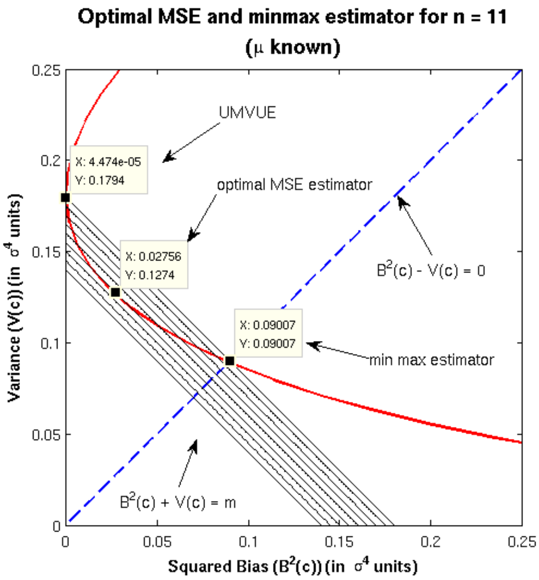
Since we have more than one estimator for our parameter σ^2, we would like to test their performance on different grounds. For comparison of these estimators, we look at some popular criteria such as Mean Squared Error (MSE). Although MSE is a good enough criterion, it does not take into account the individual contributions of squared bias and variance components. So we devise some other benchmarks for comparison such as ratio of squared bias to the MSE of an estimator, which identifies relative contribution of squared bias towards the MSE, and Pareto efficiency, which treats the comparison as a multi-criteria optimization problem.

3.1 Comparison based on MSE

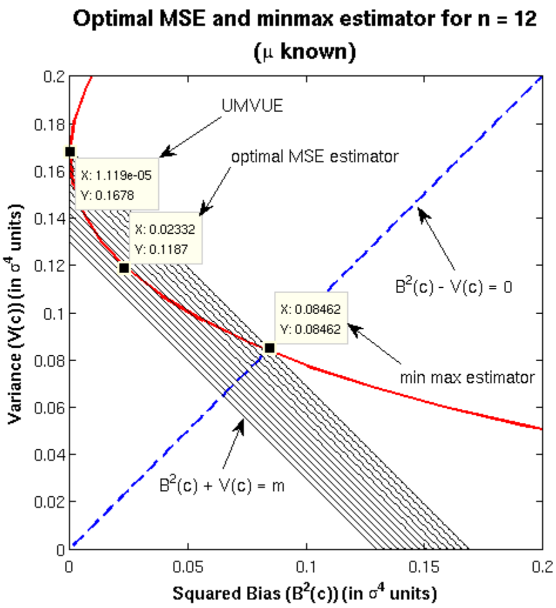
The first measure that we use to compare these estimators is the MSE of each estimator.



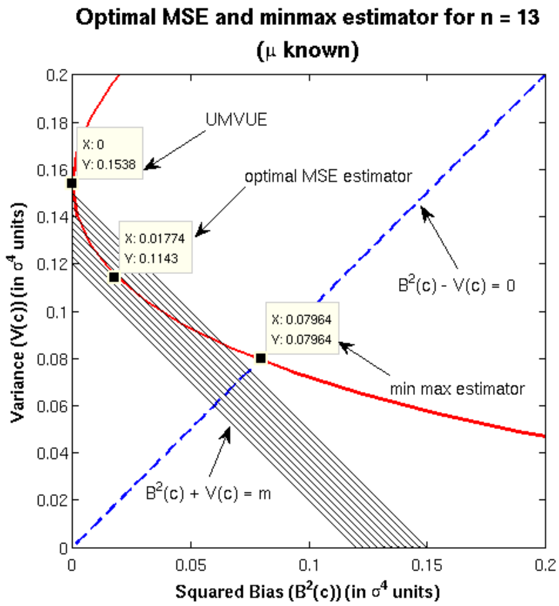
(a) Sample size $n = 2$



(b) Sample size $n = 11$



(c) Sample size $n = 12$



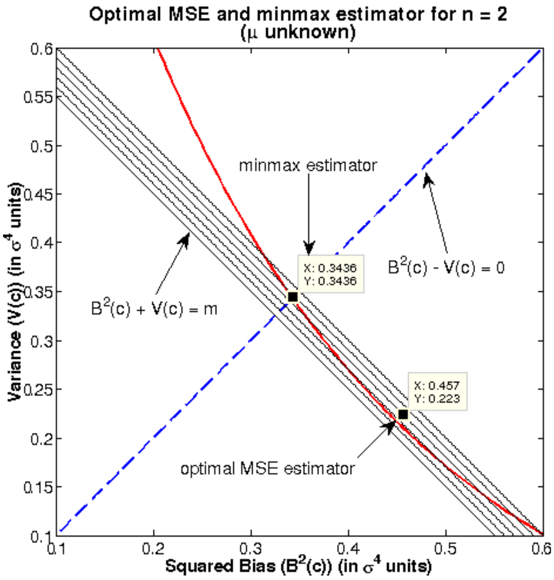
(d) Sample size $n = 13$

Figure 3: Optimal MSE, classical and min max estimator of σ^2 with known μ

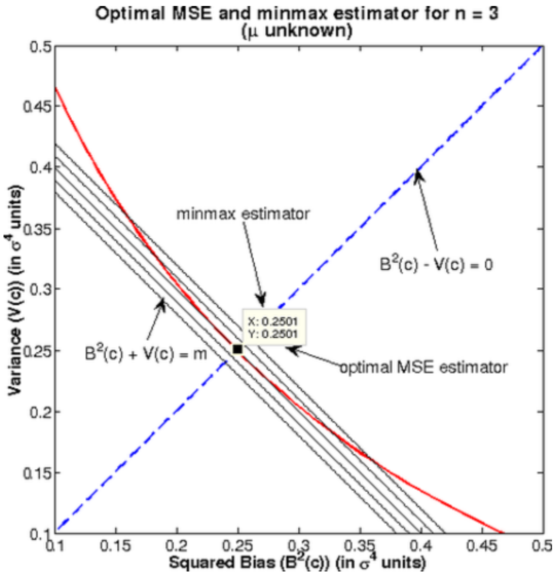
3.1.1 Case: μ known

(i) Optimal MSE estimator $\left(\overline{S}^2_{c^*_{\text{MSE}}}\right)$

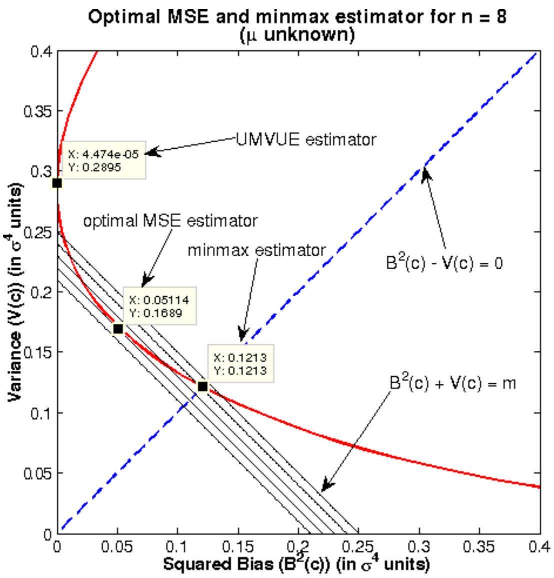
$$\begin{aligned}
\text{MSE}(\overline{S}^2_{c^*_{\text{MSE}}}) &= \sigma^4[(n^2+2n)c^{*2}_{\text{MSE}} - 2nc^*_{\text{MSE}} + 1] \\
&= \sigma^4\left[(n^2+2n)\frac{1}{(n+2)^2} - \frac{2n}{n+2} + 1\right] \\
&= \frac{2\sigma^4}{n+2}
\end{aligned}$$



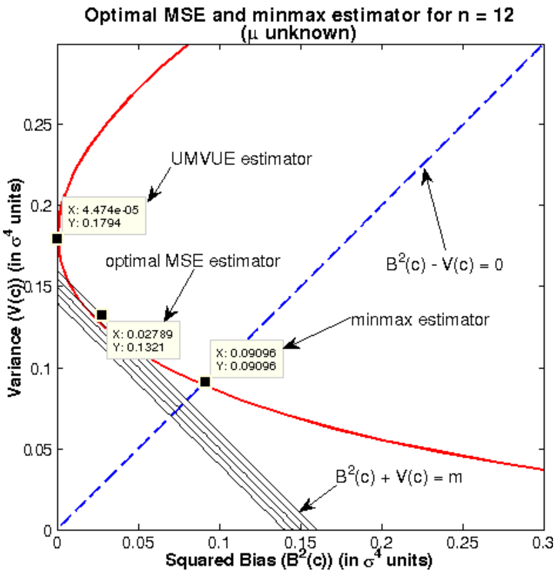
(a) Sample size $n = 2$



(b) Sample size $n = 3$



(c) Sample size $n = 8$



(d) Sample size $n = 12$

Figure 4: Optimal MSE, classical and min max estimator of σ^2 with unknown μ

(ii) min max estimator $\left(\overline{S}^2_{c^*_{\text{mm}}}\right)$

$$\begin{aligned}
\text{MSE}(\overline{S}^2_{c^*_{\text{mm}}}) &= \sigma^4[(n^2+2n)c^{*2}_{\text{mm}} - 2nc^*_{\text{mm}} + 1] \\
&= \sigma^4\left[(n^2+2n)\frac{1}{n(\sqrt{n}+\sqrt{2})^2} - \frac{2n}{\sqrt{n}(\sqrt{n}+\sqrt{2})} + 1\right] \\
&= \frac{4\sigma^4}{(\sqrt{n}+\sqrt{2})^2}
\end{aligned}$$

(iii) Classical unbiased estimator ($\hat{\sigma}^2$)

$$\begin{aligned}\text{MSE}(\hat{\sigma}^2) &= \sigma^4[(n^2+2n)\frac{1}{n^2} - 2n\frac{1}{n} + 1] \\ &= \frac{2\sigma^4}{n}\end{aligned}$$

Comparison Results:

(a) Clearly, $\dfrac{1}{n+2} < \dfrac{1}{n} \quad \forall n = 1, 2, \ldots$

$\therefore$ MSE $\left(\overline{S}^2_{c^*_{\text{MSE}}}\right) <$ MSE$(\hat{\sigma}^2)$ for $n = 2, 3, \ldots$.

(b) Consider, $n+2$ and $\dfrac{(\sqrt{n}+\sqrt{2})^2}{2}$, i.e., $2n+4$ and $n+2+2\sqrt{2n}$.

$$\begin{aligned}\text{Now, } \ 2n+4 &\geq n+2+2\sqrt{2n} \\ \Rightarrow n+2 &\geq 2\sqrt{2n}\end{aligned}$$

which is true for $n = 2, 3, \ldots$

$\therefore$ MSE $\left(\overline{S}^2_{c^*_{\text{MSE}}}\right) \leq$ MSE$(\overline{S}^2_{c^*_{\text{mm}}})$ for $n = 2, 3, \ldots$.

(c) Next consider, n and $\dfrac{(\sqrt{n}+\sqrt{2})^2}{2}$, i.e., $2n$ and $n+2+2\sqrt{2n}$.

$$\begin{aligned}\text{Now, } \ 2n &\geq n+2+2\sqrt{2n} \\ \Rightarrow n-2 &\geq 2\sqrt{2n} \qquad (20)\end{aligned}$$

which holds true only for $n = 12, 13, \ldots$

For $n \in \{2, \ldots, 11\}$, $n-2 < 2\sqrt{2n}$. Thus,

$$\begin{aligned}\text{MSE}(\overline{S}^2_{c^*_{\text{mm}}}) &< \text{MSE}(\hat{\sigma}^2) \quad \forall n \in \{2, \ldots, 11\}, \\ \text{and } \ \text{MSE}(\overline{S}^2_{c^*_{\text{mm}}}) &> \text{MSE}(\hat{\sigma}^2) \quad \forall n = 12, 13, \ldots\end{aligned}$$

Note that the inequality holds strictly in this case, because there is no integral value of n such that (20) holds with equality.

3.1.2 Case: μ unknown

(i) Optimal MSE estimator $\left(S^2_{c^*_{\text{MSE}}}\right)$

$$\begin{aligned}\text{MSE}(S^2_{c^*_{\text{MSE}}}) &= \sigma^4[(n^2-1)c^{*2}_{\text{MSE}} - 2(n-1)c^*_{\text{MSE}} + 1] \\ &= \frac{2\sigma^4}{n+1}\end{aligned}$$

(ii) min max estimator $\left(S^2_{c^*_{\text{mm}}}\right)$

$$\begin{aligned}\text{MSE}(S^2_{c^*_{\text{mm}}}) &= [(n^2-1)c^{*2}_{\text{mm}} - 2(n-1)c^*_{\text{mm}} + 1] \\ &= \frac{4\sigma^4}{(\sqrt{n-1}+\sqrt{2})^2} \\ &= \frac{4\sigma^4}{n+1+2\sqrt{2(n-1)}}\end{aligned}$$

(iii) Classical unbiased estimator (S^2)

$$\begin{aligned}\text{MSE}(S^2) &= \text{Var}(S^2) \qquad (\because S^2 \text{ is unbiased}) \\ &= \frac{2(n-1)\sigma^4}{(n-1)^2} \\ &= \frac{2\sigma^4}{n-1}\end{aligned}$$

(iv) Classical biased estimator (s^2)

$$\begin{aligned}\text{MSE}(s^2) &= \sigma^4[(n^2-1)\frac{1}{n^2} - 2(n-1)\frac{1}{n} + 1] \\ &= \frac{(2n-1)\sigma^4}{n^2}\end{aligned}$$

Comparison Results:

(a) Clearly, $\dfrac{1}{n+1} < \dfrac{1}{n-1} \quad \forall n = 2, 3, \dots$

$\therefore \text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right) < \text{MSE}(S^2)$ for $n = 2, 3, \dots$.

(b) $\dfrac{1}{n^2} < \dfrac{1}{(n-1)^2} \quad \forall n = 2, 3, \dots$

$\therefore \text{MSE}\,(s^2) < \text{MSE}(S^2)$ for $n = 2, 3, \dots$.

(c) $n - 1 > 0 \quad \forall n = 2, 3, \dots$

$\Rightarrow 2n^2 + n - 1 > 2n^2$

$\Rightarrow (2n-1)(n+1) > 2n^2$

$\Rightarrow \dfrac{(2n-1)}{n^2} > \dfrac{2}{n+1}$

$\Rightarrow \text{MSE}(s^2) > \text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right)$

Thus, $\text{MSE}(S^2) > \text{MSE}(s^2) > \text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right)$ for $n = 2, 3, \dots$.

(d) To compare $\text{MSE}\left(S^2_{c^*_{\text{mm}}}\right)$ and $\text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right)$, lets assume

$$\begin{aligned}&\text{MSE}\left(S^2_{c^*_{\text{mm}}}\right) \leq \text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right) \\ \Rightarrow\ &\frac{4\sigma^4}{n+1+2\sqrt{2(n-1)}} \leq \frac{2\sigma^4}{n+1} \\ \Rightarrow\ &\frac{2}{n+1+2\sqrt{2(n-1)}} \leq \frac{1}{n+1} \\ \Rightarrow\ &n+1 \leq 2\sqrt{2(n-1)} \\ \Rightarrow\ &(n+1)^2 \leq 8(n-1) \quad (\because n+1>0, 2\sqrt{2(n-1)} > 0, \text{hence squaring preserves the inequality}) \\ \Rightarrow\ &n^2 - 6n + 9 \leq 0 \\ \Rightarrow\ &(n-3)^2 \leq 0\end{aligned}$$

which is contradictory for all $n = 2, \ldots$ except for $n = 3$ where equality holds.
$\therefore \text{MSE}\left(S^2_{c^*_{\text{mm}}}\right) \geq \text{MSE}\left(S^2_{c^*_{\text{MSE}}}\right)$ for $n = 2, 3, \ldots$.

(e) For comparing $\text{MSE}\left(S^2_{c^*_{\text{mm}}}\right)$ and $\text{MSE}\left(s^2\right)$, lets assume

$$\begin{aligned}
&\text{MSE}\left(S^2_{c^*_{\text{mm}}}\right) \geq \text{MSE}\left(s^2\right) \\
\Rightarrow\ &\frac{4\sigma^4}{n+1+2\sqrt{2(n-1)}} \geq \frac{(2n-1)\sigma^4}{n^2} \\
\Rightarrow\ &4n^2 \geq (2n-1)(n+1+2\sqrt{2(n-1)}) \\
\Rightarrow\ &2n^2 - 4n\sqrt{2(n-1)} - n + 2\sqrt{2(n-1)} + 1 \geq 0
\end{aligned}$$

By explicit computation, the above is true for $n \in \{7, 8, \ldots\}$,
while for $n \in \{2, \ldots, 6\}$, $2n^2 - 4n\sqrt{2(n-1)} - n + 2\sqrt{2(n-1)} + 1 \leq 0$

$$\begin{aligned}
\text{Thus,}\ \ &\text{MSE}(S^2_{c^*_{\text{mm}}}) < \text{MSE}(s^2) \quad \forall n \in \{2, \ldots, 6\}, \\
\text{and}\ \ &\text{MSE}(S^2_{c^*_{\text{mm}}}) > \text{MSE}(s^2) \quad \forall n = 7, 8, \ldots
\end{aligned}$$

(f) Now comparing $\text{MSE}(S^2_{c^*_{\text{mm}}})$ and $\text{MSE}(S^2)$, we assume

$$\begin{aligned}
&\text{MSE}(S^2_{c^*_{\text{mm}}}) < \text{MSE}(S^2) \\
\Rightarrow\ &\frac{4\sigma^4}{n+1+2\sqrt{2(n-1)}} < \frac{2\sigma^4}{n-1} \\
\Rightarrow\ &\frac{2}{n+1+2\sqrt{2(n-1)}} < \frac{1}{n-1} \\
\Rightarrow\ &2(n-1) < n+1+2\sqrt{2(n-1)} \\
\Rightarrow\ &n-3 < 2\sqrt{2(n-1)} \qquad (21)
\end{aligned}$$

which holds true only for $n \in \{2, \ldots, 12\}$. The inequality reverses for $n = 13, 14, \ldots$

$$\begin{aligned}
\text{Thus,}\ \ &\text{MSE}(S^2_{c^*_{\text{mm}}}) < \text{MSE}(S^2) \quad \forall n \in \{2, \ldots, 12\}, \\
\text{and}\ \ &\text{MSE}(S^2_{c^*_{\text{mm}}}) > \text{MSE}(S^2) \quad \forall n = 13, 14, \ldots
\end{aligned}$$

Note that the inequality holds strictly in this case, because there is no integral value of n such that (21) holds with equality.

3.2 Ratio of squared bias to MSE

While MSE is a good way of simultaneously capturing both accuracy and precision of an estimator, it can not capture the relative contributions of square of bias and variance. One simple way to capture this is to consider the percentage of square of bias in MSE. So, for an estimator T, we consider a quantity f_c, the ratio of square of the bias of the estimator T to its MSE, defined as:

$$f_c(S^2_c) = \frac{B^2_c}{\text{MSE}(S^2_c)} \qquad (22)$$

In general, $f_c : (0, \infty) \to [0, 1]$. For the above class of estimators that we are considering, we have $f_c : (0, \infty) \to [0, 1)$ (we do not consider $c = 0$, as such an estimator gives inadmissable estimate of zero for the variance σ^2).

3.2.1 Case: μ known

For a sample of size n, $f_c(.)$ for a general estimator is

$$f_c(\overline{S}_c^2) = \frac{(nc-1)^2}{(nc-1)^2 + 2nc^2} \tag{23}$$

The ratios for the estimators of interest are listed below:

(i) $f_c(\overline{S}^2_{c^*_{\text{mm}}}) = 0.5$

(ii) $f_c(\overline{S}^2_{c^*_{\text{MSE}}}) = \left(\frac{-2}{n+2}\right)^2 \div \frac{2}{(n+2)} = \frac{2}{n+2}$

(iii) $f_c(\sigma^2) = 0$

3.2.2 Case: μ unknown

For a sample of size n, $f_c(.)$ for a general estimator is

$$f_c(S_c^2) = \frac{((n-1)c-1)^2}{((n-1)c-1)^2 + 2(n-1)c^2} \tag{24}$$

The ratios for the estimators of interest are listed below:

(i) $f_c(S^2_{c^*_{\text{mm}}}) = 0.5$

(ii) $f_c(S^2_{c^*_{\text{MSE}}}) = \left(\frac{-2}{n+1}\right)^2 \div \frac{2}{(n+1)} = \frac{2}{n+1}$

(iii) $f_c(s^2) = \left(\frac{-1}{n}\right)^2 \div \frac{2n-1}{n^2} = \frac{1}{2n-1}$

(iv) $f_c(S^2) = 0$

Estimators with f_c close to 0 or 1 have either square of the bias or variance dominating the other. Notice that the ratio f_c is constant for minmax estimator and UMVUE. But for the optimal MSE estimator (and the classical estimator in the μ unknown case), the ratio is a decreasing function of sample size n and is bounded on both sides by the ratios for minmax and UMVUE.

3.3 Pareto efficient estimators

We now treat the squared bias-variance aspects of the linear estimators as a multi-criteria problem and plot the Pareto frontier in squared bias and variance space. We first show below that all the estimators that we have considered in the previous sections, are Pareto optimal. Then we identify the Pareto frontier.

3.3.1 Pareto optimality of the above estimators

We show below that the UMVUE, optimal MSE and minmax estimators are Pareto optimal on the squared bias versus variance frontier.

minmax estimator: For a given sample size, n, consider the minmax estimator with squared bias, $B^2_{c^*_{\text{mm}}}$ and variance, $V_{c^*_{\text{mm}}}$; where $B^2_{c^*_{\text{mm}}} = V_{c^*_{\text{mm}}}$. Any estimator with (squared bias, variance) $= (a, a)$, $a < B^2_{c^*_{\text{mm}}}$ cannot lie on the curve of squared bias versus variance, by definition and uniqueness of minmax estimator. Therefore, consider an estimator with (B^2_c, V_c). Let $B^2_c < B^2_{c^*_{\text{mm}}}$, then V_c has to be greater than $V_{c^*_{\text{mm}}}$, otherwise it contradicts the fact that $(B^2_{c^*_{\text{mm}}}, V_{c^*_{\text{mm}}})$ is minmax estimator. Similarly, if $V_c < V_{c^*_{\text{mm}}}$, then B^2_c will have to be more than $B^2_{c^*_{\text{mm}}}$ to preserve the definition of $(B^2_{c^*_{\text{mm}}}, V_{c^*_{\text{mm}}})$. Thus, one cannot reduce both squared bias and variance of the minmax estimator, by considering any other estimator in the considered class, parametrized by the coefficient, c. Therefore, minmax estimator is a Pareto point.

Optimal MSE estimator: When n is fixed, there cannot be an estimator with (B^2_c, V_c) such that $(B^2_c + V_c) < (B^2_{c^*_{\text{MSE}}} + V_{c^*_{\text{MSE}}})$ since $(B^2_{c^*_{\text{MSE}}}, V_{c^*_{\text{MSE}}})$ denotes the optimal MSE estimator. As the optimal MSE estimator is unique, consider an estimator (B^2_c, V_c) with $(B^2_c + V_c) > (B^2_{c^*_{\text{MSE}}} + V_{c^*_{\text{MSE}}})$. Suppose $B^2_c < B^2_{c^*_{\text{MSE}}}$. This implies that $V_c > V_{c^*_{\text{MSE}}}$. On the other hand, if for this estimator, $V_c < V_{c^*_{\text{MSE}}}$, then $B^2_c > B^2_{c^*_{\text{MSE}}}$ to preserve the optimality of $(B^2_{c^*_{\text{MSE}}}, V_{c^*_{\text{MSE}}})$. Hence, the optimal MSE estimator is also a Pareto point.

Uniformly Minimum Variance Unbiased Estimator (UMVUE): UMVUE is characterized by $(0, V_{c^*_{\text{UMVUE}}})$, since by definition its bias has to be zero. Since UMVUE is unique, we cannot have an estimator (B^2_c, V_c) with $B^2_c = 0$ and $V_c < V_{c^*_{\text{UMVUE}}}$ for a given n. However if we insist on reduced variance, i.e., $V_c < V_{c^*_{\text{UMVUE}}}$, then B^2_c for this estimator has to be positive, by definition of UMVUE. Thus, UMVUE is another Pareto point on the squared bias versus variance curve for a fixed n.

3.3.2 Pareto frontier

For a given sample size, n, all the estimators with coefficient c lying on the lower arm of the squared bias versus variance plot are Pareto estimators (refer to Figure 3 and 4). This can be argued as follows:

Consider the μ known case. Let the sample size, n, be fixed. Then the bias and variance of an estimator for σ^2 with coefficient c are,

$$B_c = nc - 1 \tag{25}$$

$$V_c = 2nc^2 \tag{26}$$

From the above two equations, bias and variance of an estimator can be related as:

$$V_c = 2n\left(\frac{B_c + 1}{n}\right)^2$$

$$\Rightarrow V_c = \frac{2}{n}B^2_c + \frac{4}{n}B_c + \frac{2}{n} \tag{27}$$

Comparing the above equation with the general form of a conic section

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

we see that it satisfies the following restriction:

$$B^2 - 4AC = 0.$$

Hence the equation (27) represents a parabola with B_c and V_c as the variables.

Note that, B_c^2 itself is a parabola in variable c, with vertex at $c = \frac{1}{n}$. Hence, B_c^2 decreases as c varies over $(0, \frac{1}{n}]$ and increases over $(\frac{1}{n}, \infty)$. On the other hand, V_c, which is also a parabola in c with vertex at $c = 0$, increases monotonically over $(0, \infty)$. Therefore, on the interval $(0, \frac{1}{n}]$, the values of B_c^2 and V_c cannot be increased simultaneously. Thus, this arm of the parabola represents the class of Pareto optimal estimators for σ^2.

Similarly for the μ unknown case, when the sample size n is fixed, the bias and variance of an estimator with coefficient c are:

$$\begin{aligned} B_c &= (n-1)c - 1 \\ V_c &= 2(n-1)c^2 \end{aligned}$$

As in the μ known case, it can be established in the μ unknown case also that B_c and V_c for the estimators trace a parabola. Also, B_c^2 is a parabola in variable c, with vertex at $c = \frac{1}{n-1}$ as well as V_c is a parabola in c with vertex at $c = 0$. Thus, B_c^2 decreases monotonically on $c \in (0, \frac{1}{n-1}]$, whereas V_c is ever increasing on this interval. Hence, the estimators with coefficients $c \in (0, \frac{1}{n-1}]$ are Pareto optimal since we cannot simultaneously improve their squared bias and variance.

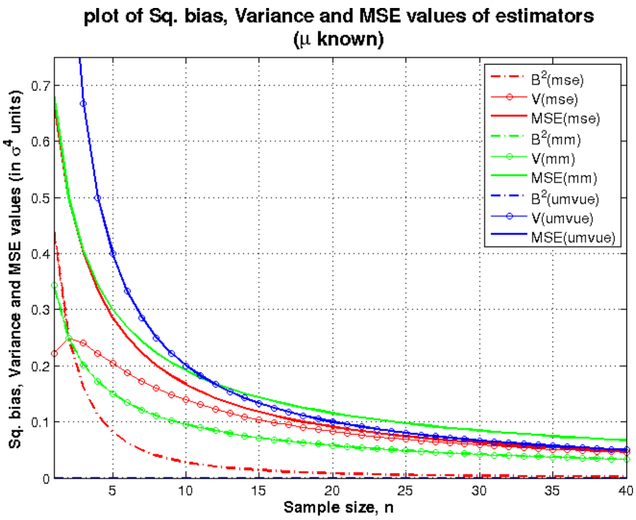
In fact, it can be seen from Figures 3 and 4 that the optimal MSE estimator, minmax estimator, UMVUE and MLE – all lie on this Pareto frontier, since their coefficients belong to the specified interval that captures the Pareto optimal estimators. The c^*_{MSE}, S^2 and s^2 estimators are sandwiched between minmax estimator and UMVUE estimators in the variance vs squared bias plot. Another observation is that the c^*_{MSE}, S^2 (and s^2, when mean of the data is also to be estimated) cluster near the UMVUE estimator in the variance vs squared-bias plot showing that the variance is a predominant component in the MSE of these estimators. This is also brought out in the f_c measure of these estimators.

4 Comparison of MSE and f_c values of different estimators as functions of n

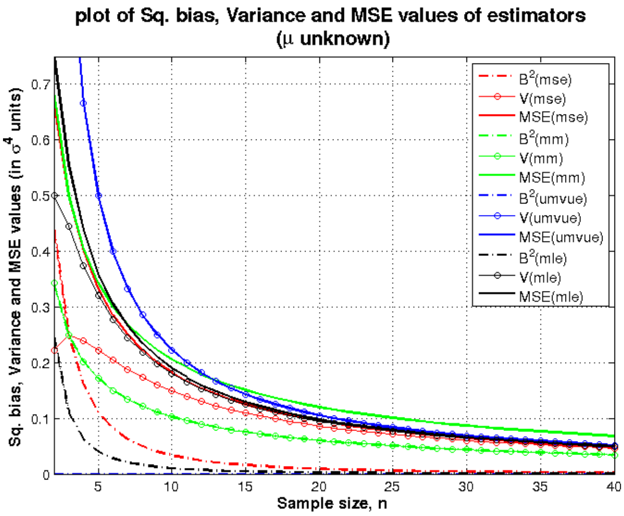
We have summarized in Table 1 the values of squared bias, variance, MSE and fc for different estimators under both the cases in Table 1. It can be easily observed that the MSE values of the considered estimators are $O(\frac{1}{n})$ since the bias can be 0 or $O(\frac{1}{n^2})$ and the variance is $O(\frac{1}{n})$.

Considering the μ known case, from the Figure 3, we see that for a fixed sample size, n, optimal MSE estimator has the lowest MSE value in the graphs; while among the remaining two estimators, $\min\max$ estimator performs better for small sample sizes ($n \in \{2, \ldots, 11\}$), whereas $\hat{\sigma}^2$ performs better for $n \in \{12, 13, \ldots\}$ in terms of MSE values. Parallel results can be observed in Figure 4 for the case of unknown μ.

Figure 5 plots the squared bias, variance and MSE values for the considered estimators of σ^2 for a range of sample sizes under both the cases: μ known and μ unknown. Figure 6 compares the ratio of the squared bias to the MSE value of the mentioned estimators for a range of sample sizes under both the cases: μ known and μ unknown.

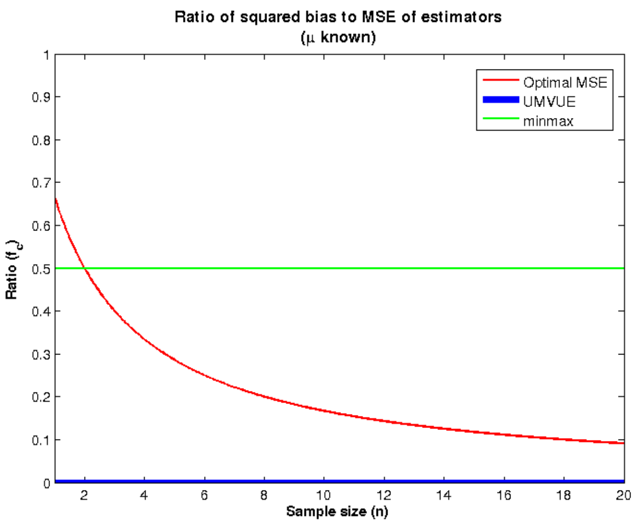


(a) when μ is known

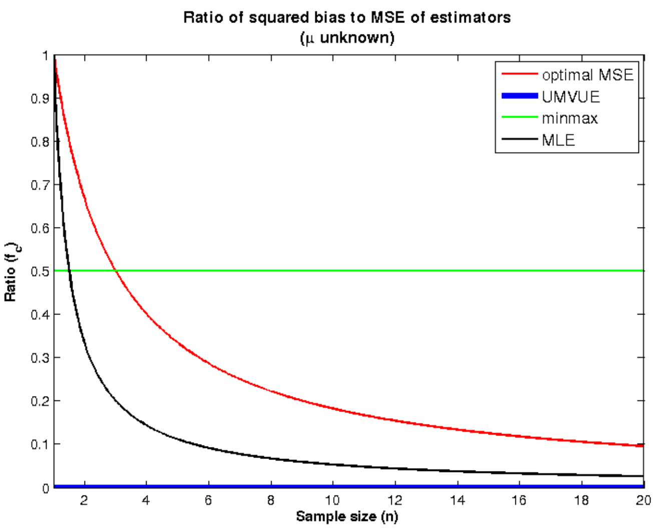


(b) when μ is unknown

Figure 5: Comparison of MSE values for the estimators of σ^2



(a) when μ is known



(b) when μ is unknown

Figure 6: Comparison of ratios f_c for the estimators of σ^2

	Estimator	**Definition**	$(B^2(c), V(c))$ in σ^4 units	MSE in σ^4 units	Ratio f_c
μ known	Optimal MSE, $\overline{S}^2_{c^*_{\mathrm{MSE}}}$	$\frac{1}{n+2}\sum_{i=1}^n (X_i-\mu)^2$	$\left(\frac{4}{(n+2)^2}, \frac{2n}{(n+2)^2}\right)$	$\frac{2}{n+2}$	$\frac{2}{n+2}$
	min max, $\overline{S}^2_{c^*_{\mathrm{mm}}}$	$\frac{1}{\sqrt{n}(\sqrt{n}+\sqrt{2})}\sum_{i=1}^n (X_i-\mu)^2$	$\left(\frac{2}{(\sqrt{n}+\sqrt{2})^2}, \frac{2}{(\sqrt{n}+\sqrt{2})^2}\right)$	$\frac{4}{(\sqrt{n}+\sqrt{2})^2}$	0.5
	UMVUE/MLE, $\hat{\sigma}^2$	$\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$	$\left(0, \frac{2}{n}\right)$	$\frac{2}{n}$	0
μ unknown	Optimal MSE, $S^2_{c^*_{\mathrm{MSE}}}$	$\frac{1}{n+1}\sum_{i=1}^n (X_i-\overline{X})^2$	$\left(\frac{4}{(n+1)^2}, \frac{2(n-1)}{(n+1)^2}\right)$	$\frac{2}{n+1}$	$\frac{2}{n+1}$
	min max, $S^2_{c^*_{\mathrm{mm}}}$	$\frac{1}{(\sqrt{n-1})(\sqrt{n-1}+\sqrt{2})}\sum_{i=1}^n (X_i-\overline{X})^2$	$\left(\frac{2}{(\sqrt{n-1}+\sqrt{2})^2}, \frac{2}{(\sqrt{n-1}+\sqrt{2})^2}\right)$	$\frac{4}{(\sqrt{n-1}+\sqrt{2})^2}$	0.5
	UMVUE, S^2	$\frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$	$\left(0, \frac{2}{n-1}\right)$	$\frac{2}{n-1}$	0
	MLE, s^2	$\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2$	$\left(\frac{1}{n^2}, \frac{2(n-1)}{n^2}\right)$	$\frac{2n-1}{n^2}$	$\frac{1}{2n-1}$

Table 1: MSE values and ratios for different estimators of σ^2 for a sample size n

5 Discussion

In both the cases (μ known and μ unknown), the distribution of the estimators was known[Casella and Berger, 2002] to be chi-squared distribution (with n and $n-1$ degrees of freedom for μ known and μ unknown cases, respectively) and hence computation of bias, variacne and MSE was easy. We proposed a minmax estimator by using min max criterion over squared bias and variance of the estimator. We saw that, though the optimal MSE estimator always had the minimum MSE value for a given sample size; the min max estimator performed better than the classical estimator(s) on the MSE value for small sample sizes. That all these estimators are Pareto optimal shows that the improvement in one of the two competing risks, the variance and squared bias, comes only at the cost of increase of the other risk in these estimators.

Our first conclusion is that as sample size increases, i.e., in the context of 'Big Data', these estimators are nearly same when MSE as a measure of quality of an estimator. When the measure of quality of an estimator is the fraction of squared bias in MSE of an estimator, it is not more than half (for minmax estimator) and zero for UMVUE estimators. Our finding that all these estimators are Pareto optimal in the space of squared bias and variance of estimators complies with the above conclusion by pointing out that for these estimators an attempt to decrease the squared bias (variance) leads to increase in variance (squared bias).

Note that our observations and analysis were restricted to above class of estimators. As a part of further investigation, one aspect could be to identify a 'better' class of estimators. Another aspect is to consider non-normal data. Similar ideas could be used for estimating the parameters of a linear regression model.

Appendix

We elaborate on some of the points we mentioned in the beginning of the chapter about the sample size, n and the range of the coefficient, c.

1. All the estimators from the class that we have defined, will take value zero if we have $n = 0$ or 1. Therefore, the sample should have at least two data points, i.e., $n \geq 2$.

2. One can look for the 'best' estimator for μ in the class of estimators denoted by

$$T_d = d \sum_{i=1}^{n} X_i, \quad d \geq 0.$$

   But the search for an optimal estimator in this class leads to an unrealizable estimator since the coefficient, c^*_{MSE}, of this estimator itself depends on μ and σ^2 values. So, the classical unbiased estimator, $\overline{X}$, for μ is used to define the class of estimators for σ^2 when μ is unknown.

3. It should be noted that our class of estimators,

$$S_c^2 = c \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

consists of estimators with coefficients, $c > 0$. The case $c = 0$ can be safely ignored, since our optimal estimators for different risk criteria can be obtained from the open interval $c \in (0, \infty)$, as shown below.
When $c = 0$, the estimator (S_0^2) becomes a constant estimator taking value zero, under both the cases: μ known and μ unknown. The bias, variance and MSE values (in σ^4 units) for such an estimator are:

$$\begin{aligned} B_0 &= -1 \text{ and } V_0 = 0, \\ \Rightarrow \text{MSE}(S_0^2) &= B_0^2 + V_0 = 1 \qquad (28) \\ \text{and } \max\{B_0^2, V_0\} &= \max\{1, 0\} = 1. \qquad (29) \end{aligned}$$

We will now show that in both the cases when μ known or unknown, we have at least one 'better' estimator.

- **Case: μ known** For a general estimator, $\overline{S}_c^2$, using (8) and (9), we have:

  $$\begin{aligned} \Rightarrow \text{MSE}(\overline{S}_c^2) &= (nc-1)^2 + 2nc^2 \\ \text{and } \max\{B_c^2, V_c\} &= \max\{(nc-1)^2, 2nc^2\}. \end{aligned}$$

  Given n, consider $c = \frac{1}{2n}$. Then

  $$\text{MSE}(\overline{S}_c^2) = \left(\frac{1}{4} + \frac{1}{2n}\right) \leq \frac{1}{2} \quad \forall n \geq 2 \qquad (30)$$

  $$\text{and } \max\{B_c^2, V_c\} = \max\left\{\frac{1}{4}, \frac{1}{2n}\right\} \leq \frac{1}{4} \quad \forall n \geq 2. \qquad (31)$$

  Comparing (28) with (30) and (29) with (31), we can see that the estimator with $c = \frac{1}{2n}$ is better than that with $c = 0$ on MSE value and minmax criterion.

- **Case: μ unknown** As in the above case, for $c = \frac{1}{2(n-1)}$, using (10) and (11)

  $$\text{MSE}(S_c^2) = \left(\frac{1}{4} + \frac{1}{2(n-1)}\right) \leq \frac{3}{4} \quad \forall n \geq 2 \qquad (32)$$

  $$\text{and } \max\{B_c^2, V_c\} = \max\left\{\frac{1}{4}, \frac{1}{2(n-1)}\right\} \leq \frac{1}{2} \quad \forall n \geq 2. \qquad (33)$$

  Comparing (28) with (32) and (29) with (33), we can see that the estimator with $c = \frac{1}{2(n-1)}$ is better than that with $c = 0$.

Note that all the above computations and comparisons are in σ^4 units.

References

[Alpaydin, 2010] Alpaydin, E. (2010). *Introduction to Machine Learning.* MIT Press, Cambridge, 2nd edition.

[Billingsley, 1995] Billingsley, P. (1995). *Probability and Measure.* John Wiley, New York, 3rd edition.

[Casella and Berger, 2002] Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Duxbury Thomson Learning, 2nd edition.

[Chung, 2001] Chung, K. L. (2001). *A Course in Probability Theory*. Academic Press, New York, 3rd edition.

[DeGroot and Schervish, 2012] DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison Wesley Publishing Company Incorporated, 4th edition.

[Doss and Sethuraman, 1989] Doss, H. and Sethuraman, J. (1989). The price of bias reduction when there is no unbiased estimate. *The Annals of Statistics*, 17(1):440–442.

[Friedman, 1997] Friedman, J. H. (1997). On bias, variance, 0/1 - loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77.

[Fristedt and Gray, 1997] Fristedt, B. E. and Gray, L. F. (1997). *A Modern Approach to Probability Theory*. Probability and Its Applications. Birkhäuser, Boston.

[Kallenberg, 2002] Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer-Verlag, New York, 2nd edition.

[Wasserman, 2005] Wasserman, L. (2005). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science, New York, 2nd edition.