Methods of Evaluating Estimators

Let $X_1, X_2, \cdots, X_n$ be n i.i.d. random variables, i.e., a random sample from $f(x|\theta)$, where θ is unknown. An estimator of θ is a function of (only) the n random variables, i.e., a statistic $\hat{\theta} = r(X_1, \cdots, X_n)$. There are several method to obtain an estimator for θ, such as the MLE, method of moment, and Bayesian method.

A difficulty that arises is that since we can usually apply more than one of these methods in a particular situation, we are often face with the task of choosing between estimators. Of course, it is possible that different methods of finding estimators will yield the same answer (as we have see in the MLE handout), which makes the evaluation a bit easier, but, in many cases, different methods will lead to different estimators. We need, therefore, some criteria to choose among them.

We will study several measures of the quality of an estimator, so that we can choose the best. Some of these measures tell us the quality of the estimator with small samples, while other measures tell us the quality of the estimator with large samples. The latter are also known as asymptotic properties of estimators.

1 Mean Square Error (MSE) of an Estimator

Let $\hat{\theta}$ be the estimator of the unknown parameter θ from the random sample $X_1, X_2, \cdots, X_n$. Then clearly the deviation from $\hat{\theta}$ to the true value of θ, $|\hat{\theta} - \theta|$, measures the quality of the estimator, or equivalently, we can use $(\hat{\theta} - \theta)^2$ for the ease of computation. Since $\hat{\theta}$ is a random variable, we should take average to evaluation the quality of the estimator. Thus, we introduce the following

Definition: The mean square error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is the function of θ defined by $E(\hat{\theta} - \theta)^2$, and this is denoted as $MSE_{\hat{\theta}}$.

This is also called the risk function of an estimator, with $(\hat{\theta} - \theta)^2$ called the quadratic loss function. The expectation is with respect to the random variables $X_1, \cdots, X_n$ since they are the only random components in the expression.

Notice that the MSE measures the average squared difference between the estimator $\hat{\theta}$ and the parameter θ, a somewhat reasonable measure of performance for an estimator. In general, any increasing function of the absolute distance $|\hat{\theta} - \theta|$ would serve to measure the goodness of an estimator (mean absolute error, $E(|\hat{\theta}-\theta|)$, is a reasonable alternative. But MSE has at least two advantages over other distance measures: First, it is analytically tractable and, secondly, it has the interpretation

$$MSE_{\hat{\theta}} = E(\hat{\theta}-\theta)^2 = Var(\hat{\theta}) + (E(\hat{\theta})-\theta)^2 = Var(\hat{\theta}) + (Bias\ of\ \hat{\theta})^2$$

This is so because

$$\begin{aligned} E(\hat{\theta}-\theta)^2 &= E(\hat{\theta}^2) + E(\theta^2) - 2\theta E(\hat{\theta}) \\ &= Var(\hat{\theta}) + [E(\hat{\theta})]^2 + \theta^2 - 2\theta E(\hat{\theta}) \\ &= Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \end{aligned}$$

Definition: The bias of an estimator $\hat{\theta}$ of a parameter θ is the difference between the expected value of $\hat{\theta}$ and θ; that is, $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$. An estimator whose bias is identically equal to 0 is called unbiased estimator and satisfies $E(\hat{\theta}) = \theta$ for all θ.

Thus, MSE has two components, one measures the variability of the estimator (precision) and the other measures the its bias (accuracy). An estimator that has good MSE properties has small combined variance and bias. To find an estimator with good MSE properties, we need to find estimators that control both variance and bias.

For an unbiased estimator $\hat{\theta}$, we have

$$MSE_{\hat{\theta}} = E(\hat{\theta}-\theta)^2 = Var(\hat{\theta})$$

and so, if an estimator is unbiased, its MSE is equal to its variance.

Example 1: Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with density function $f(x|\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)$, the maximum likelihood estimator for σ

$$\hat{\sigma} = \frac{\sum_{i=1}^n |X_i|}{n}$$

is unbiased.

Solution: Let us first calculate $E(|X|)$ and $E(|X|^2)$ as

$$\begin{aligned} E(|X|) &= \int_{-\infty}^{\infty} |x| f(x|\sigma)dx = \int_{-\infty}^{\infty} |x| \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)dx \\ &= \sigma\int_0^{\infty} \frac{x}{\sigma}\exp\left(-\frac{x}{\sigma}\right)d\frac{x}{\sigma} = \sigma\int_0^{\infty} ye^{-y}dy = \sigma\Gamma(2) = \sigma \end{aligned}$$

and

$$\begin{aligned} E(|X|^2) &= \int_{-\infty}^{\infty} |x|^2 f(x|\sigma)dx = \int_{-\infty}^{\infty} |x|^2 \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)dx \\ &= \sigma^2\int_0^{\infty} \frac{x^2}{\sigma^2}\exp\left(-\frac{x}{\sigma}\right)d\frac{x}{\sigma} = \sigma^2\int_0^{\infty} y^2e^{-y}dy = \sigma\Gamma(3) = 2\sigma^2 \end{aligned}$$

Therefore,

$$E(\hat{\sigma}) = E\left(\frac{|X_1| + \cdots + |X_n|}{n}\right) = \frac{E(|X_1|) + \cdots + E(|X_n|)}{n} = \sigma$$

So $\hat{\sigma}$ is an unbiased estimator for σ.

Thus the MSE of $\hat{\sigma}$ is equal to its variance, i.e.

$$\begin{aligned} MSE_{\hat{\sigma}} &= E(\hat{\sigma} - \sigma)^2 = Var(\hat{\sigma}) = Var\left(\frac{|X_1| + \cdots + |X_n|}{n}\right) \\ &= \frac{Var(|X_1|) + \cdots + Var(|X_n|)}{n^2} = \frac{Var(|X|)}{n} \\ &= \frac{E(|X|^2) - (E(|X|))^2}{n} = \frac{2\sigma^2 - \sigma^2}{n} = \frac{\sigma^2}{n} \end{aligned}$$

The Statistic S^2: Recall that if $X_1, \cdots, X_n$ come from a normal distribution with variance σ^2, then the sample variance S^2 is defined as

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

It can be shown that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. From the properties of χ^2 distribution, we have

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

and

$$Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1) \Rightarrow Var(S^2) = \frac{2\sigma^4}{n-1}$$

Example 2: Let $X_1, X_2, \cdots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with expected value μ and variance σ^2, then $\bar{X}$ is an unbiased estimator for μ, and S^2 is an unbiased estimator for σ^2.

Solution: We have

$$E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{E(X_1) + \cdots + E(X_n)}{n} = \mu$$

Therefore, $\bar{X}$ is an unbiased estimator. The MSE of $\bar{X}$ is

$$MSE_{\bar{X}} = E(\bar{X} - \mu)^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$$

This is because

$$Var(\bar{X}) = Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{Var(X_1) + \cdots + Var(X_n)}{n^2} = \frac{\sigma^2}{n}$$

Similarly, as we showed above, $E(S^2) = \sigma^2$, S^2 is an unbiased estimator for σ^2, and the MSE of S^2 is given by

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

Although many unbiased estimators are also reasonable from the standpoint of MSE, be aware that controlling bias does not guarantee that MSE is controlled. In particular, it is sometimes the case that a trade-off occurs between variance and bias in such a way that a small increase in bias can be traded for a larger decrease in variance, resulting in an improvement in MSE.

Example 3: An alternative estimator for σ^2 of a normal population is the maximum likelihood or method of moment estimator

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

It is straightforward to calculate

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

so $\hat{\sigma^2}$ is a biased estimator for σ^2. The variance of $\hat{\sigma^2}$ can also be calculated as

$$Var(\hat{\sigma^2}) = Var\left(\frac{n-1}{n}S^2\right) = \frac{(n-1)^2}{n^2}Var(S^2) = \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}.$$

Hence the MSE of $\hat{\sigma^2}$ is given by

$$\begin{aligned} E(\hat{\sigma^2} - \sigma^2)^2 &= Var(\hat{\sigma^2}) + (Bias)^2 \\ &= \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \frac{2n-1}{n^2}\sigma^4 \end{aligned}$$

We thus have (using the conclusion from Example 2)

$$MSE_{\hat{\sigma^2}} = \frac{2n-1}{n^2}\sigma^4 < \frac{2n}{n^2}\sigma^4 = \frac{2\sigma^4}{n} < \frac{2\sigma^4}{n-1} = MSE_{S^2}.$$

This shows that $\hat{\sigma^2}$ has smaller MSE than S^2. Thus, by trading off variance for bias, the MSE is improved.

The above example does not imply that S^2 should be abandoned as an estimator of σ^2. The above argument shows that, on average, $\hat{\sigma^2}$ will be closer to σ^2 than S^2 if MSE is used as a measure. However, $\hat{\sigma^2}$ is biased and will, on the average, underestimate σ^2. This fact alone may make us uncomfortable about using $\hat{\sigma^2}$ as an estimator for σ^2.

In general, since MSE is a function of the parameter, there will not be one "best" estimator in terms of MSE. Often, the MSE of two estimators will cross each other, that is, for some parameter values, one is better, for other values, the other is better. However, even this partial information can sometimes provide guidelines for choosing between estimators.

One way to make the problem of finding a “best” estimator tractable is to limit the class of estimators. A popular way of restricting the class of estimators, is to consider only unbiased estimators and choose the estimator with the lowest variance.

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of a parameter θ, that is, $E(\hat{\theta}_1) = \theta$ and $E(\hat{\theta}_2) = \theta$, then their mean squared errors are equal to their variances, so we should choose the estimator with the smallest variance.

A property of Unbiased estimator: Suppose both A and B are unbiased estimator for an unknown parameter θ, then the linear combination of A and B: $W = aA + (1-a)B$, for any a is also an unbiased estimator.

Example 4: This problem is connected with the estimation of the variance of a normal distribution with unknown mean from a sample $X_1, X_2, \cdots, X_n$ of i.i.d. normal random variables. For what value of ρ does $\rho \sum_{i=1}^{n}(X_i - \bar{X})^2$ have the minimal MSE?

Please note that if $\rho = \frac{1}{n-1}$, we get S^2 in example 2; when $\rho = \frac{1}{n}$, we get $\hat{\sigma^2}$ in example 3.

Solution:

As in above examples, we define

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

Then,

$$E(S^2) = \sigma^2 \quad \text{and} \quad \text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

Let

$$e_\rho = \rho \sum_{i=1}^{n}(X_i - \bar{X})^2 = \rho(n-1)S^2$$

and let $t = \rho(n-1)$ Then

$$E(e_\rho) = \rho(n-1)E(S^2) = \rho(n-1)\sigma^2 = t\sigma^2$$

and

$$Var(e_\rho) = \rho^2(n-1)^2 Var(S^2) = \frac{2t^2}{n-1}\sigma^4$$

We can Calculate the MSE of e_ρ as

$$\begin{aligned} MSE(e_\rho) &= Var(e_\rho) + [Bias]^2 = Var(e_\rho) + \left[E(e_\rho) - \sigma^2\right]^2 \\ &= Var(e_\rho) + (t\sigma^2 - \sigma^2)^2 = Var(e_\rho) + (t-1)^2\sigma^4. \end{aligned}$$

Plug in the results before, we have

$$MSE(e_\rho) = \frac{2t^2}{n-1}\sigma^4 + (t-1)^2\sigma^4 = f(t)\sigma^4$$

where

$$f(t) = \frac{2t^2}{n-1} + (t-1)^2 = \left(\frac{n+1}{n-1}t^2 - 2t + 1\right)$$

when $t = \frac{n-1}{n+1}$, $f(t)$ achieves its minimal value, which is $\frac{2}{n+1}$. That is the minimal value of $MSE(e_\rho) = \frac{2\sigma^4}{n+1}$, with $(n-1)\rho = t = \frac{n-1}{n+1}$, i.e. $\rho = \frac{1}{n+1}$.

From the conclusion in example 3, we have

$$MSE_{\hat{\sigma^2}} = \frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1} = MSE_{S^2}.$$

It is straightforward to verify that

$$MSE_{\hat{\sigma^2}} = \frac{2n-1}{n^2}\sigma^4 \geq \frac{2\sigma^4}{n+1} = MSE(e_\rho)$$

when $\rho = \frac{1}{n+1}$.

2 Efficiency of an Estimator

As we pointed out earlier, Fisher information can be used to bound the variance of an estimator. In this section, we will define some quantity measures for an estimator using Fisher information.

2.1 Efficient Estimator

Suppose $\hat{\theta} = r(X_1, \cdots, X_n)$ is an estimator for θ, and suppose $E(\hat{\theta}) = m(\theta)$, a function of θ, then T is an unbiased estimator of $m(\theta)$. By information inequality,

$$\mathrm{Var}(\hat{\theta}) \geq \frac{[m'(\theta)]^2}{nI(\theta)}$$

when the equality holds, the estimator $\hat{\theta}$ is said to be an efficient estimator of its expectation $m(\theta)$. Of course, if $m(\theta) = \theta$, then T is an unbiased estimator for θ.

Example 5: Suppose that $X_1, \cdots, X_n$ form a random sample from a Bernoulli distribution for which the parameter p is unknown. Show that $\bar{X}$ is an efficient estimator of p.

Proof: If $X_1, \cdots, X_n \sim$ Bernoulli(p), then $E(\bar{X}) = p$, and $\text{Var}(\bar{X}) = p(1-p)/n$. By example 3 from the fisher information lecture note, the fisher information is $I(p) = 1/[p(1-p)]$. Therefore the variance of $\bar{X}$ is equal to the lower bound $1/[nI(p)]$ provided by the information inequality, and $\bar{X}$ is an efficient estimator of p.

Recall that in the proof of information inequality, we used the Cauchy-Schwartz inequality,

$$\left\{\text{Cov}_\theta[\hat{\theta}, l_n'(\mathbf{X}|\theta)]\right\}^2 \leq \text{Var}_\theta[\hat{\theta}]\text{Var}_\theta[l_n'(\mathbf{X}|\theta)].$$

From the proof procedure, we know that if the equality holds in Cauchy-Schwartz inequality, then the equality will hold in information inequality. We know that if and only if there is a linear relation between $\hat{\theta}$ and $l_n'(\mathbf{X}|\theta)$, the Cauchy-Schwartz inequality will become an equality, and hence the information inequality will become an equality. In other words, $\hat{\theta}$ will be an efficient estimator if and only if there exist functions $u(\theta)$ and $v(\theta)$ such that

$$\hat{\theta} = u(\theta)l_n'(\mathbf{X}|\theta) + v(\theta).$$

The functions $u(\theta)$ and $v(\theta)$ may depend on θ but not depend on the observations $X_1, \cdots, X_n$.

Because $\hat{\theta}$ is an estimator, it cannot involve the parameter θ. Therefore, in order for $\hat{\theta}$ to be efficient, it must be possible to find functions $u(\theta)$ and $v(\theta)$ such that the parameter θ will actually be canceled from the right side of the above equation, and the value of $\hat{\theta}$ will depend on the observations $X_1, \cdots, X_n$ and not on θ.

Example 6: Suppose that $X_1, \cdots, X_n$ form a random sample from a Poisson distribution for which the parameter θ is unknown. Show that $\bar{X}$ is an efficient estimator of θ.

Proof: The joint p.m.f. of $X_1, \cdots, X_n$ is

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^n x_i!}.$$

Then

$$l_n(\mathbf{X}|\theta) = -n\theta + n\bar{X}\log\theta - \sum_{i=1}^n \log(X_i!),$$

and

$$l_n'(\mathbf{X}|\theta) = -n + \frac{n\bar{X}}{\theta}.$$

If we now let $u(\theta) = \theta/n$ and $v(\theta) = \theta$, then

$$\bar{X} = u(\theta)l_n'(\mathbf{X}|\theta) + v(\theta).$$

Since the statistic $\bar{X}$ has been represented as a linear function of $l_n'(\mathbf{X}|\theta)$, it follows that $\bar{X}$ will be an efficient estimator of its expectation θ. In other words, the variance of $\bar{X}$ will attain the lower bound given by the information inequality.

Suppose $\hat{\theta}$ is an efficient estimator for its expectation $E(\hat{\theta}) = m(\theta)$. Let a statistic T be a linear function of $\hat{\theta}$, i.e. $T = a\hat{\theta} + b$, where a and b are constants. Then T is an efficient estimator for $E(T)$, i.e., a linear function of an efficient estimator is an efficient estimator for its expectation.

Proof: We can see that $E(T) = aE(\hat{\theta}) + b = am(\theta) + b$, by information inequality

$$\mathrm{Var}(T) \geq \frac{a^2[m'(\theta)]^2}{nI(\theta)}.$$

We also have

$$\mathrm{Var}(T) = \mathrm{Var}(a\hat{\theta} + b) = a^2\mathrm{Var}(\hat{\theta}) = a^2\frac{[m'(\theta)]^2}{nI(\theta)},$$

since $\hat{\theta}$ is an efficient estimator for $m(\theta)$, $\mathrm{Var}(\hat{\theta})$ attains its lower bound. Our computation shows that the variance of T can attain its lower bound, which implies that T is an efficient estimator for $E(T)$.

Now, let us consider the exponential family distribution

$$f(x|\theta) = \exp[c(\theta)T(x) + d(\theta) + S(x)],$$

and we suppose there is a random sample $X_1, \cdots, X_n$ from this distribution. We will show that the sufficient statistic $\sum_{i=1}^n T(X_i)$ is an efficient estimator of its expectation.

Clearly,

$$l_n(\mathbf{X}|\theta) = \sum_{i=1}^n \log f(X_i|\theta) = \sum_{i=1}^n [c(\theta)T(X_i) + d(\theta) + S(X_i)] = c(\theta)\sum_{i=1}^n T(X_i) + nd(\theta) + \sum_{i=1}^n S(X_i),$$

and

$$l_n'(\mathbf{X}|\theta) = c'(\theta)\sum_{i=1}^n T(X_i) + nd'(\theta).$$

Therefore, there is a linear relation between $\sum_{i=1}^n T(X_i)$ and $l_n'(\mathbf{X}|\theta)$:

$$\sum_{i=1}^n T(X_i) = \frac{1}{c'(\theta)}l_n'(\mathbf{X}|\theta) - \frac{nd'(\theta)}{c'(\theta)}.$$

Thus, the sufficient statistic $\sum_{i=1}^n T(X_i)$ is an efficient estimator of its expectation. Any linear function of $\sum_{i=1}^n T(X_i)$ is a sufficient statistic and is an efficient estimator of its expectation. Specifically, if the MLE of θ is a linear function of sufficient statistic, then MLE is efficient estimator of θ.

Example 7. Suppose that $X_1, \cdots, X_n$ form a random sample from a normal distribution for which the mean μ is known and the variance σ^2 is unknown. Construct an efficient estimator for σ^2.

Solution: Let $\theta = \sigma^2$ be the unknown variance. Then the p.d.f. is

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta}(x-\mu)^2\right\},$$

which can be recognized as a member of exponential family with $T(x) = (x-\mu)^2$. So $\sum_{i=1}^n (X_i - \mu)^2$ is an efficient estimator for its expectation. Since $E[(X_i - \mu)^2] = \sigma^2$, $E[\sum_{i=1}^n (X_i - \mu)^2] = n\sigma^2$. Therefore, $\sum_{i=1}^n (X_i - \mu)^2/n$ is an efficient estimator for σ^2.

2.2 Efficiency and Relative Efficiency

For an estimator $\hat{\theta}$, if $E(\hat{\theta}) = m(\theta)$, then the ratio between the CR lower bound and $\text{Var}(\hat{\theta})$ is called the efficiency of the estimator $\hat{\theta}$, denoted as $e(\hat{\theta})$, i.e.

$$e(\hat{\theta}) = \frac{[m'(\theta)]^2/[nI(\theta)]}{\text{Var}(\hat{\theta})}.$$

By the information inequality, we have $e(\hat{\theta}) \leq 1$ for any estimator $\hat{\theta}$.

Note: some textbooks or materials define efficient estimator and efficiency of an estimator only for unbiased estimator, which is a special case of $m(\theta) = \theta$ in our definitions.

If an estimator is unbiased and and its variance attains the Cramér-Rao lower bound, then it is called the minimum variance unbiased estimator (MVUE).

To evaluate an estimator $\hat{\theta}$, we defined the mean squared error as

$$MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

If the estimator is unbiased, then $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$. When two estimators are both unbiased, comparison of their MSEs reduces to comparison of their variances.

Given two unbiased estimators, $\hat{\theta}$ and $\tilde{\theta}$, of a parameter θ, the **relative efficiency** of $\hat{\theta}$ relative to $\tilde{\theta}$ is defined as

$$\text{eff}(\hat{\theta}, \tilde{\theta}) = \frac{\text{Var}(\tilde{\theta})}{\text{Var}(\hat{\theta})}.$$

Thus, if the efficiency is smaller than 1, $\hat{\theta}$ has a larger variance than $\tilde{\theta}$ has. This comparison is most meaningful when both $\hat{\theta}$ and $\tilde{\theta}$ are unbiased or when both have the same bias. Frequently, the variances of $\hat{\theta}$ and $\tilde{\theta}$ are of the form

$$\text{var}(\hat{\theta}) = \frac{c_1}{n} \quad \text{and} \quad \text{var}(\tilde{\theta}) = \frac{c_2}{n}$$

where n is the sample size. If this is the case, the efficiency can be interpreted as the ratio of sample sizes necessary to obtain the same variance for both $\hat{\theta}$ and $\tilde{\theta}$.

Example 8: Let $Y_1, \cdots, Y_n$ denote a random sample from the uniform distribution on the interval $(0, \theta)$. Consider two estimators,

$$\hat{\theta}_1 = 2\bar{Y} \quad \text{and} \quad \hat{\theta}_2 = \frac{n+1}{n} Y_{(n)},$$

where $Y_{(n)} = \max(Y_1, \cdots, Y_n)$. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

Solution: Because each Y_i follows a uniform distribution on the interval $(0, \theta)$, $\mu = E(Y_i) = \theta/2$, and $\sigma^2 = \text{Var}(Y_i) = \theta^2/12$. Therefore,

$$E(\hat{\theta}_1) = E(2\bar{Y}) = 2E(\bar{Y}) = 2\mu = \theta,$$

so $\hat{\theta}_1$ is unbiased. Furthermore

$$\text{Var}(\hat{\theta}_1) = \text{Var}(2\bar{Y}) = 4\text{Var}(\bar{Y}) = 4\frac{\sigma^2}{n} = \frac{\theta^2}{3n}.$$

To find the mean and variance of $\hat{\theta}_2$, recall that the density function of $Y_{(n)}$ is given by

$$g_{(n)}(y) = n[F_Y(y)]^{n-1} f_Y(y) = \begin{cases} n\left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} & \text{for} \quad 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$E(Y_{(n)}) = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{n+1}\theta,$$

it follows that $E(\hat{\theta}_2) = E\{[(n+1)/n]Y_{(n)}\} = \theta$, i.e. $\hat{\theta}_2$ is an unbiased estimator for θ.

$$E(Y_{(n)}^2) = \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{n+2}\theta^2,$$

therefore, the variance of $Y_{(n)}$ is

$$\text{Var}(Y_{(n)}) = E(Y_{(n)}^2) - E(Y_{(n)})^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2.$$

Thus,

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\frac{n+1}{n}Y_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 \left[\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right] = \frac{\theta^2}{n(n+2)}.$$

Finally, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is given by

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = \frac{\theta^2/[n(n+2)]}{\theta^2/(3n)} = \frac{3}{n+2}.$$

3 Exercises

Exercise 1. X, the cosine of the angle at which electrons are emitted in muon decay has a density

$$f(x) = \frac{1 + \alpha X}{2} \quad -1 \le x \le 1 \quad -1 \le \alpha \le 1$$

The parameter α is related to polarization. Show that $E(X) = \frac{\alpha}{3}$. Consider an estimator for the parameter α, $\hat{\alpha} = 3\bar{X}$. Computer the variance, the bias, and the mean square error of this estimator.

Exercise 2. Suppose that $X_1, \cdots, X_n$ form a random sample from a normal distribution for which the mean μ is unknown and the variance σ^2 is known. Show that $\bar{X}$ is an efficient estimator of μ.

Exercise 3. Suppose that $X_1, \cdots, X_n$ form a random sample of size n from a Poisson distribution with mean λ. Consider $\hat{\lambda}_1 = (X_1 + X_2)/2$ and $\hat{\lambda}_2 = \bar{X}$. Find the efficiency of $\hat{\lambda}_1$ relative to $\hat{\lambda}_2$.

Exercise 4. Suppose that $Y_1, \cdots, Y_n$ denote a random sample of size n form an exponential distribution with density function given by

$$f(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta} & \text{for} y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Consider two estimators $\hat{\theta}_1 = nY_{(1)}$, and $\hat{\theta}_2 = \bar{Y}$, where $Y_{(1)} = \min(Y_1, \cdots, Y_n)$. Please show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimator of θ, find their MSE, and find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.