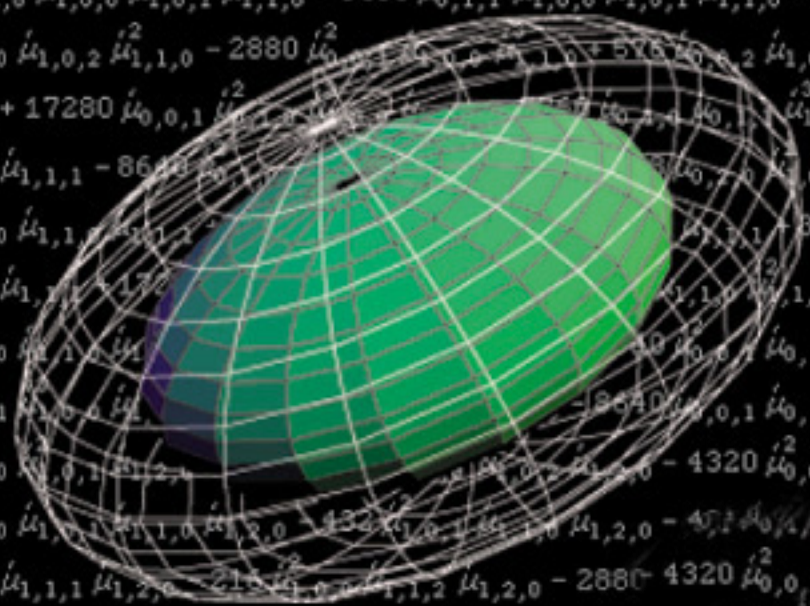
SPRINGER TEXTS IN STATISTICS

MATHEMATICAL STATISTICS *with* *Mathematica*®



COLIN ROSE
MURRAY D. SMITH

Mathematical Statistics with *Mathematica*

Chapter 9 – Statistical Decision Theory

Please reference this 2002 edition as:

Rose, C. and Smith, M.D. (2002)
Mathematical Statistics with Mathematica, Springer-Verlag, New York.

Latest edition

For the latest up-to-date edition, please visit: www.mathStatica.com

Chapter 9

Statistical Decision Theory

9.1 Introduction

Statistical decision theory is an approach to decision making for problems involving random variables. For any given problem, we use the notation D to denote the set that contains all the different decisions that can be made. There could be as few as two elements in D, or even an uncountably large number of possibilities. The aim is to select a particular decision from D that is, in some sense, optimal. A wide range of statistical problems can be tackled using the tools of decision theory, including estimator selection, hypothesis testing, forecasting, and prediction. For discussion ranging across different types of problems, see, amongst others, Silvey (1975), Gourieroux and Monfort (1995), and Judge *et al*. (1985). In this chapter, emphasis focuses on using decision theory for estimator selection.

Because the decision problem involves random variables, the impact of any particular choice will be uncertain. We represent uncertainty by assuming the existence of a parameter $\theta \in \Theta$ whose true value θ_0 is unknown. The decision problem is then to select an estimator from the set D whose elements are the estimators proposed for a given problem. Our goal, therefore, is to select an estimator from D in an optimal fashion.

9.2 Loss and Risk

Optimality in decision theory is defined according to a *loss structure*, the latter being a function that applies without variation to all estimators in the decision set D. The *loss function*, denoted by $L = L(\hat{\theta}, \theta)$, measures the disadvantages of selecting an estimator $\hat{\theta} \in D$. Loss takes a single, non-negative value for each and every combination of values of $\hat{\theta} \in D$ and $\theta \in \Theta$, but, apart from that, its mathematical form is *discretionary*. This, for example, means that two individuals tackling the same decision problem, can reach different least-loss outcomes, the most likely reason being that their chosen loss functions differ. Moreover, since L is a function of the random variable $\hat{\theta}$, L is itself a random variable, so that the criterion of minimisation of loss is not meaningful. Although we cannot minimise loss L, we can minimise the *expected loss*. The expected loss of $\hat{\theta}$ is also known as the *risk* of $\hat{\theta}$, where the risk function is defined as

$$R_{\hat{\theta}}(\theta) = E\left[L(\hat{\theta}, \theta)\right] \tag{9.1}$$

where $\hat{\theta}$ is a random variable with density $g(\hat{\theta};\theta)$. Because the expectation is with respect to the density of $\hat{\theta}$, risk is a non-random function of θ. Notice that because the loss L is non-negative, risk must also be non-negative. Given a particular estimator chosen from D, we solve (9.1) to obtain its risk. As its name would suggest, the smaller the value of risk, the better off we are — the decision criterion is to *minimise risk*.[1]

With the aid of risk, we now return to the basic question of how to choose amongst the estimators in the decision set. Consider two estimators of θ_0, namely, $\hat{\theta}$ and $\tilde{\theta}$, both of which are members of a decision set D. We say that $\hat{\theta}$ *dominates* $\tilde{\theta}$ if the risk of the former is no greater than the risk of the latter throughout the entire parameter space, with the added proviso that the risk of $\hat{\theta}$ be strictly smaller in some part of the parameter space; that is, $\hat{\theta}$ dominates $\tilde{\theta}$ if

$$R_{\hat{\theta}}(\theta) \;\le\; R_{\tilde{\theta}}(\theta), \quad \text{for all } \theta \in \Theta$$

along with

$$R_{\hat{\theta}}(\theta) \;<\; R_{\tilde{\theta}}(\theta), \quad \text{for some } \theta \in \Theta^* \subset \Theta$$

where Θ^* is a non-null set. Notice that dominance is a *binary* relationship between estimators in D. This means that if there are d estimators in D, then there are $d(d-1)/2$ dominance relations that can be tested. Once an estimator is shown to be dominated, then we may rule it out of our decision process, for we can always do better by using the estimator(s) that dominate it; a dominated estimator is termed *inadmissible*. Finally, if an estimator is not dominated by any of the other estimators in D, then it is deemed to be *admissible*; an admissible estimator is eligible to be selected to estimate θ_0. Figure 1 illustrates these concepts.

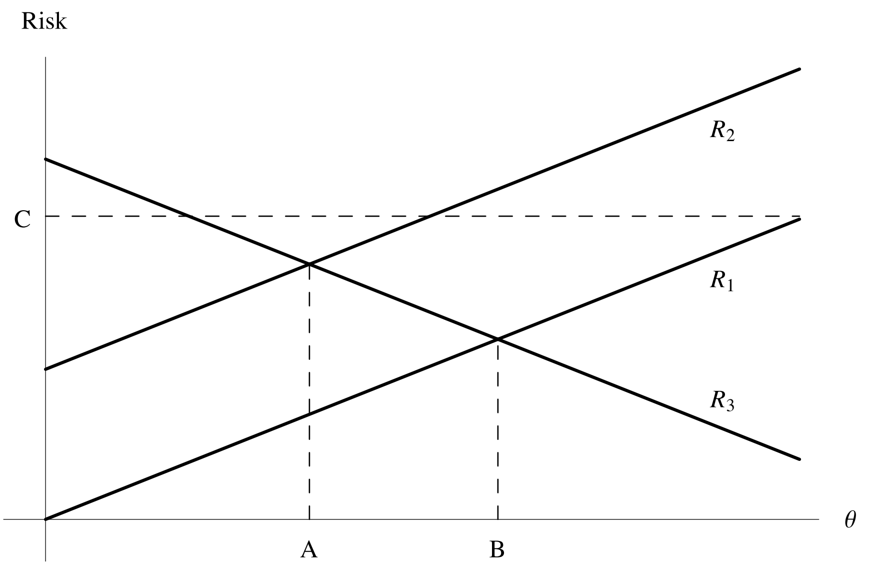


Fig. 1: Risk comparison

The decision set here is $D = \{\hat\theta_1, \hat\theta_2, \hat\theta_3\}$, and the risk of each estimator is plotted as a function of θ, and is labelled on each (continuous) line; the diagram is plotted over the entire parameter space Θ. The first feature to observe is that the risk of estimator $\hat\theta_1$ (denoted by R_1) is everywhere below the risk of $\hat\theta_2$ (denoted by R_2). Thus, $\hat\theta_1$ dominates $\hat\theta_2$ (therefore $\hat\theta_2$ is inadmissible). The next feature concerns the risk functions of $\hat\theta_1$ and $\hat\theta_3$ (denoted by R_3); they cross at B, and therefore neither estimator dominates the other. To the left of B, $\hat\theta_1$ has smaller risk and is preferred to $\hat\theta_3$, whereas to the right of B the ranking is reversed. It follows that both $\hat\theta_1$ and $\hat\theta_3$ are admissible estimators. Of course, if we knew (for example) that the true parameter value θ_0 lay in the region to the left of B, then $\hat\theta_1$ is dominant in D and therefore preferred. However, generally this type of knowledge is not available. The following example, based on Silvey (1975, Example 11.2), illustrates some of these ideas.

⊕ ***Example 1:*** The Risk of a Normally Distributed Estimator

Suppose that a random variable $X \sim N(\theta, 1)$, where $\theta \in \mathbb{R}$ is an unknown parameter. The random variable $\hat\theta = X + k$ is proposed as an estimator of θ, where constant $k \in \mathbb{R}$. Thus, the estimate of θ is formed by adding k to a single realisation of the random variable X. The decision set, in this case, consists of all possible choices for k. Thus, $D = \{k : k \in \mathbb{R}\}$ is a set with an uncountably infinite number of elements. By the linearity property of the Normal distribution, it follows that estimator $\hat\theta$ is Normally distributed; that is, $\hat\theta \sim N(\theta + k, 1)$ with pdf $f(\hat\theta; \theta)$:

$$\mathbf{f} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\hat\theta - (\theta + k)\right)^2};$$

$$\mathbf{domain[f]} = \{\hat\theta, -\infty, \infty\} \;\&\&\; \{-\infty < \theta < \infty, -\infty < k < \infty\};$$

Let $c_1 \in \mathbb{R}_+$ and $c_2 \in \mathbb{R}_+$ be two constants chosen by an individual, and suppose that the loss structure specified for this problem is

$$L(\hat\theta, \theta) = \begin{cases} c_1(\hat\theta - \theta) & \text{if } \hat\theta \geq \theta \\ c_2(\theta - \hat\theta) & \text{if } \hat\theta < \theta. \end{cases} \tag{9.2}$$

In **mathStatica**, we enter this as:

$$\mathbf{L} = \mathbf{If}\left[\hat\theta \geq \theta,\ c_1\left(\hat\theta - \theta\right),\ c_2\left(\theta - \hat\theta\right)\right];$$

Figure 2 plots the loss function when $c_1 = 2$, $c_2 = 1$ and $\theta = 0$. The asymmetry in the loss function leads to differing magnitudes of loss depending on whether the estimate is larger or smaller than $\theta = 0$. In Fig. 2, an over-estimate of θ causes a greater loss than an under-estimate of the same size. In this case, intuition suggests that we search for a $k < 0$, for if we are to err, we will do better if the error results from an under-estimate. In a similar vein, if $c_1 < c_2$, then over-estimates are preferred to under-estimates, so we would expect to choose a $k > 0$; and when the loss is symmetric $c_1 = c_2$, no correction would be necessary and we would choose $k = 0$. We now show that it is possible to identify the unique value of k that minimises risk. Naturally, it will depend on the values of c_1 and c_2.

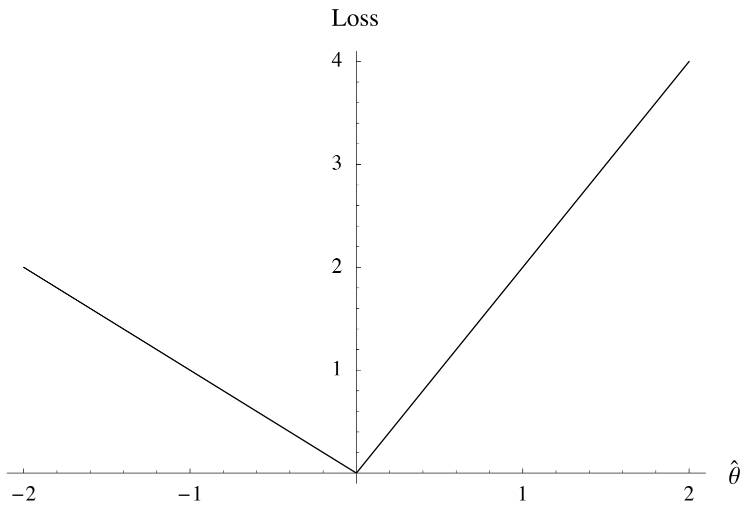


Fig. 2: An asymmetric loss function ($c_1 = 2$ and $c_2 = 1$)

Risk is expected loss:

```
Risk = Expect[L, f]
```

$$\frac{e^{-\frac{k^2}{2}}\,(c_1 + c_2)}{\sqrt{2\,\pi}} + \frac{1}{2}\,k\,\left(\left(1 + \mathrm{Erf}\left[\frac{k}{\sqrt{2}}\right]\right)\,c_1 + \left(-1 + \mathrm{Erf}\left[\frac{k}{\sqrt{2}}\right]\right)\,c_2\right)$$

from which we see that risk is dependent on factors that are under our control; that is, it does not depend on values of θ. For given values of c_1 and c_2, the value of k which minimises risk can be found in the usual way. Here is the first derivative of risk with respect to k:

```
d1 = D[Risk, k] // Simplify
```

$$\frac{1}{2}\,\left(\left(1 + \mathrm{Erf}\left[\frac{k}{\sqrt{2}}\right]\right)\,c_1 + \left(-1 + \mathrm{Erf}\left[\frac{k}{\sqrt{2}}\right]\right)\,c_2\right)$$

and here is the second derivative:

```
d2 = D[Risk, {k, 2}] // Simplify
```

$$\frac{e^{-\frac{k^2}{2}}\,(c_1 + c_2)}{\sqrt{2\,\pi}}$$

Notice that the second derivative `d2` is positive for all k.

Next, set the first derivative to zero and solve for k:

```
solk = Solve[d1 == 0, k]
```

$$\left\{\left\{k \to \sqrt{2}\ \text{InverseErf}\left[0,\ \frac{-c_1 + c_2}{c_1 + c_2}\right]\right\}\right\}$$

This value for k must globally minimise risk, because the second derivative `d2` is positive for all k. Let us calculate some optimal values for k for differing choices of c_1 and c_2:

```
solk /. {c1 → 2, c2 → 1} // N
```

```
{{k → -0.430727}}
```

```
solk /. {c1 → 2, c2 → 3} // N
```

```
{{k → 0.253347}}
```

```
solk /. {c1 → 1, c2 → 1} // N
```

```
{{k → 0.}}
```

For example, the first output shows that the estimator that minimises risk when $c_1 = 2$ and $c_2 = 1$ is

$$\hat{\theta} = X - 0.430727\,.$$

This is the only admissible estimator in D for it dominates all others with respect to the loss structure (9.2). In each of the three previous outputs, notice that the optimal value of k depends on the asymmetry of the loss function as induced by the values of c_1 and c_2, and that its sign varies in accord with the intuition given earlier. ■

Of course, all decision theory outcomes are conditional upon the assumed loss structure, and as such may alter if a different loss structure is specified. Consider, for example, the *minimax* decision rule: the particular estimator $\hat{\theta}$ is preferred if

$$\hat{\theta} = \arg\min\left(\max_{\theta\in\Theta}\ R_{\hat{\Theta}}(\theta)\right), \quad \text{for all } \hat{\Theta} \in D.$$

In other words, $\hat{\theta}$ is preferred over all other estimators in the decision set if its maximum risk is no greater than the maximum risk of all other estimators. If two estimators have the same maximum risk (which may not necessarily occur at the same points in the parameter space), then we would be indifferent between them under this criterion. The minimax criterion is conservative in the sense that it selects the estimator with the least worst risk. To illustrate, consider Fig. 1 once again. We see that for the admissible estimators, $\hat{\theta}_1$ and $\hat{\theta}_3$, maximum risk occurs at the extremes of the parameter space. The value C corresponds to the maximum risk of $\hat{\theta}_1$. Since the maximum risk of $\hat{\theta}_3$ is greater than C, it follows that $\hat{\theta}_1$ is the minimax estimator.

9.3 Mean Square Error as Risk

The *Mean Square Error* (MSE) of an estimator $\hat{\theta}$ is defined as

$$\text{MSE}(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right].$$

Thus, if a quadratic loss function $L\left(\hat{\theta}, \theta\right) = \left(\hat{\theta} - \theta\right)^2$ is specified, $\text{MSE}(\hat{\theta})$ is equivalent to risk; that is, MSE is risk under quadratic loss. MSE can be expressed in terms of the first two moments of $\hat{\theta}$. To see this, let $\overline{\theta} = E[\hat{\theta}]$ for notational convenience, and write

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E\left[\left(\left(\hat{\theta} - \overline{\theta}\right) - (\theta - \overline{\theta})\right)^2\right] \\ &= E\left[\left(\hat{\theta} - \overline{\theta}\right)^2\right] + E\left[(\theta - \overline{\theta})^2\right] - 2\,E\left[\left(\hat{\theta} - \overline{\theta}\right)(\theta - \overline{\theta})\right] \qquad (9.3) \\ &= \text{Var}(\hat{\theta}) + \left(\text{Bias}(\hat{\theta})\right)^2. \end{aligned}$$

Bias is defined as $E[\hat{\theta}] - \theta = \overline{\theta} - \theta$. Thus, the first term in the second line defines the variance of $\hat{\theta}$; the second term is the squared bias of $\hat{\theta}$, and as it is non-stochastic, the outer expectation is superfluous; the third term is zero because

$$E\left[\left(\hat{\theta} - \overline{\theta}\right)(\theta - \overline{\theta})\right] = (\theta - \overline{\theta})\,E\left[\hat{\theta} - \overline{\theta}\right] = (\theta - \overline{\theta})\left(E[\hat{\theta}] - \overline{\theta}\right) = 0.$$

As the last line of (9.3) shows, estimator choice under quadratic loss depends on both variance and bias. If the decision set D consists only of unbiased estimators, then choosing the estimator with the smallest risk coincides with choosing the estimator with least variance. But should the decision set also include biased estimators, then choice based on risk is no longer as straightforward, as there is now potential to trade off variance against bias. The following diagram illustrates.

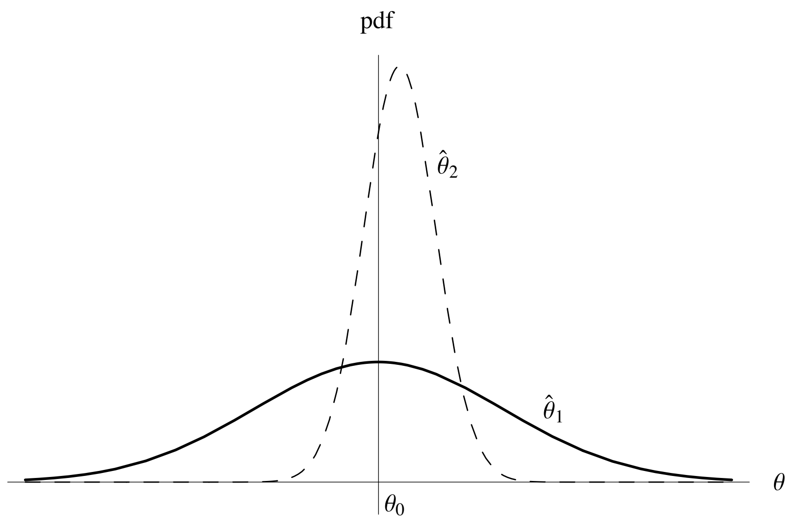


Fig. 3: Estimator densities: $\hat{\theta}_1$ has large variance (——), $\hat{\theta}_2$ is biased (– – –)

Figure 3 depicts the (scaled) pdf of two estimators of θ_0, labelled $\hat{\theta}_1$ and $\hat{\theta}_2$. Here, $\hat{\theta}_1$ is unbiased for θ_0, whereas $\hat{\theta}_2$ has a slight bias. On the other hand, the variance, or spread, of $\hat{\theta}_1$ is far greater than it is for $\hat{\theta}_2$. On computing MSE for each estimator, it would not be surprising to find $\text{MSE}(\hat{\theta}_1) > \text{MSE}(\hat{\theta}_2)$, meaning that $\hat{\theta}_2$ is preferred to $\hat{\theta}_1$ under quadratic loss. The trade-off between bias and variance favours the biased estimator in this case. However, if we envisage the pdf of $\hat{\theta}_2$ (the dashed curve) shifting further and further to the right, the cost of increasing bias would soon become overwhelming, until eventually $\text{MSE}(\hat{\theta}_2)$ would exceed $\text{MSE}(\hat{\theta}_1)$.

⊕ ***Example 2:*** Estimators for the Normal Variance

Consider a random variable $X \sim N(\mu, \theta)$, and let $(X_1, \ldots, X_n)$ denote a random sample of size n drawn on X. Asymptotic arguments may be used to justify estimating the variance parameter θ using the statistic $T = \sum_{i=1}^{n} (X_i - \overline{X})^2$, because, for example, the estimator

$$\hat{\theta} = \frac{T}{n} \xrightarrow{p} \theta.$$

That is, $\hat{\theta}$ is a consistent estimator of θ. However, the estimator remains consistent if the denominator n is replaced by, for example, $n-1$. Doing so yields the estimator $\tilde{\theta} = T/(n-1)$, for which $\tilde{\theta} \xrightarrow{p} \theta$ as the subtraction of 1 from n in the denominator becomes insignificant as n becomes larger. We therefore cannot distinguish between $\hat{\theta}$ and $\tilde{\theta}$ using asymptotic theory. As we have seen in *Example 1* of Chapter 7, estimator $\tilde{\theta}$ is an unbiased estimator of θ; consequently, given that $\hat{\theta} < \tilde{\theta}$, it follows that $\hat{\theta}$ must be biased downward for θ (*i.e.* $E[\hat{\theta}] < \theta$). On the other hand, the variance of $\tilde{\theta}$ is larger than that of $\hat{\theta}$. To summarise the situation: both estimators are asymptotically equivalent, but in finite samples there is a bias–variance trade-off between them. Proceeding along decision theoretic lines, we impose a quadratic loss structure $L(\Theta, \theta) = (\Theta - \theta)^2$ on the estimators in the decision set $D = \{\Theta : \Theta = \hat{\theta} \text{ or } \tilde{\theta}\}$.

From *Example 27* of Chapter 4, we know that $T/\theta \sim$ Chi-squared$(n-1)$. Therefore, the pdf of T, say $f(t)$, is:

$$\mathbf{f} = \frac{t^{\frac{n-1}{2}-1}\, e^{-t/(2\theta)}}{(2\theta)^{\frac{n-1}{2}}\, \Gamma[\frac{n-1}{2}]};$$

```
domain[f] = {t, 0, ∞} && {n > 0, θ > 0};
```

The MSE of each estimator can be derived by:

$$\mathbf{M\hat{S}E = Expect}\left[\left(\frac{t}{n} - \theta\right)^2, \mathbf{f}\right]$$

— This further assumes that: {n > 1}

$$\frac{(-1 + 2\,n)\,\theta^2}{n^2}$$

```
MS̃E = Expect[(t/(n - 1) - θ)^2, f]
```

— This further assumes that: {n > 1}

$$\frac{2\,\theta^2}{-1+n}$$

Both MSEs depend upon θ and sample size n. However, in this example, it is easy to rank the two estimators, because $\mathrm{M\hat{S}E}$ is strictly smaller than $\mathrm{M\tilde{S}E}$ for any value of θ:

```
MŜE - MS̃E // Simplify
```

$$\frac{(1-3\,n)\,\theta^2}{(-1+n)\,n^2}$$

Therefore, $\hat{\theta}$ dominates $\tilde{\theta}$ given quadratic loss, so $\tilde{\theta}$ is inadmissible given quadratic loss.

In this problem, it is possible to broaden the decision set from two estimators to an uncountably infinite number of estimators (all of which retain the asymptotic property of consistency) and then determine the (unique) dominant estimator; that is, the estimator that minimises MSE. To do so, we need to suppose that all estimators in the decision set have general form $\hat{\Theta} = T/(n+k)$, for some real value of k that is independent of n. The estimators that we have already examined are special cases of $\hat{\Theta}$, corresponding to $k=-1$ (for $\tilde{\theta}$) and 0 (for $\hat{\theta}$). For arbitrary k, the MSE is:

```
MSEk = Expect[(t/(n + k) - θ)^2, f]
```

— This further assumes that: {n > 1}

$$\frac{(-1+k\,(2+k)+2\,n)\,\theta^2}{(k+n)^2}$$

The minimum MSE can be obtained in the usual way by solving the first-order condition:

```
Solve[D[MSEk, k] == 0, k]
```

{{k → 1}}

… because the sign of the second derivative when evaluated at the solution is positive:

```
D[MSEk, {k, 2}] /. k → 1 // Simplify
```

$$\frac{2\,(-1+n)\,\theta^2}{(1+n)^3}$$

We conclude that θ^* dominates all other estimators with respect to quadratic loss, where

$$\theta^* = \frac{T}{n+1} = \frac{1}{n+1}\sum_{i=1}^{n}(X_i-\overline{X})^2. \qquad ■$$

⊕ ***Example 3:*** Sample Mean Versus Sample Median for Bernoulli Trials

Suppose that $Y \sim \text{Bernoulli}(\theta)$; that is, Y is a Bernoulli random variable such that $P(Y=1)=\theta$ and $P(Y=0)=1-\theta$, where θ is an unknown parameter taking real values within the unit interval, $0<\theta<1$. Suppose that a random sample of size n is drawn on Y, denoted by $(Y_1, \ldots, Y_n)$. We shall consider two estimators, namely, the sample mean $\hat{\theta}$, and the sample median $\tilde{\theta}$, and attempt to decide between them on the basis of quadratic loss. The decision set is $D = \{\hat{\theta}, \tilde{\theta}\}$.

The sample mean

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

is clearly a function of the sample sum S. In *Example 21* of Chapter 4, we established $S = \sum_{i=1}^{n} Y_i \sim \text{Binomial}(n, \theta)$, the Binomial distribution with index n and parameter θ. Therefore, $\hat{\theta}$ is a discrete random variable that may take values in the sample space $\hat{\Omega} = \{0, n^{-1}, 2\,n^{-1}, \ldots, 1\}$. Let $f(s)$ denote the pmf of S:

```
f = Binomial[n, s] θ^s (1 - θ)^(n-s);

domain[f] = {s, 0, n} &&
        {0 < θ < 1, n > 0, n ∈ Integers} && {Discrete};
```

The MSE of $\hat{\theta}$, the sample mean, is given by

```
MŜE = Expect[(s/n - θ)^2, f]
```

$$-\frac{(-1+\theta)\,\theta}{n}$$

The sample space of the sample median also depends upon the sample sum, but it is also important to identify whether the sample size is odd- or even-valued. To see this, consider first when n is odd: $\tilde{\theta}$, the sample median, will take values from $\tilde{\Omega}_{\text{odd}} = \{0, 1\}$. If the estimate is zero, then there have to be more zeroes than ones in the observed sample; this occurs when $S \le (n-1)/2$. The reverse occurs if $S \ge (n+1)/2$, for then there must be more ones than zeroes: hence the sample median must be 1. The next case is when n is even: now $\tilde{\theta}$ can take values from $\tilde{\Omega}_{\text{even}} = \{0, \frac{1}{2}, 1\}$. The outcome of $\frac{1}{2}$ exists (by convention) in even-sized samples because the number of zeroes can match exactly the number of ones.

Let us assume the sample size n is even. Then

$P(\tilde{\Theta} = \tilde{\theta})$:	$P(S \le \frac{n}{2} - 1)$	$P(S = \frac{n}{2})$	$P(S \ge \frac{n}{2} + 1)$
$\tilde{\theta}$:	0	$\frac{1}{2}$	1

Table 1: The pmf of $\tilde{\theta}$ when n is even

Let $g(\tilde{\theta})$ denote the pmf of $\tilde{\theta}$. We enter this using List Form:

$$g = \left\{\text{Prob}\left[\frac{n}{2} - 1,\ f\right],\quad f\ /.\ s \to \frac{n}{2},\quad 1 - \text{Prob}\left[\frac{n}{2},\ f\right]\right\};$$
$$\text{domain}[g] = \left\{\tilde{\theta},\ \left\{0,\ \frac{1}{2},\ 1\right\}\right\}\ \&\&\ \left\{n > 0,\ \frac{n}{2} \in \text{Integers}\right\}\ \&\&\ \{\text{Discrete}\};$$

Then, the MSE of $\tilde{\theta}$, the sample median, is:

$$\tilde{\text{MSE}} = \text{Expect}\left[(\tilde{\theta} - \theta)^2,\ g\right]$$

$$\theta^2 + \left(-\frac{1}{2} + \theta\right)^2 (-(-1+\theta)\,\theta)^{n/2}\,\text{Binomial}\left[n,\ \frac{n}{2}\right] - \frac{1}{\Gamma[2 + \frac{n}{2}]\,\Gamma[\frac{n}{2}]}\left((-1+\theta)^{1+n}\,\theta\left(\frac{\theta}{1-\theta}\right)^{n/2}\Gamma[1+n]\ \text{Hypergeometric2F1}\left[1,\ 1 - \frac{n}{2},\ 2 + \frac{n}{2},\ \frac{\theta}{-1+\theta}\right]\right) - \frac{1}{\Gamma[1+\frac{n}{2}]^2}\left((1-\theta)^{-n/2}\,(-1+\theta)^n\,\theta^{\frac{4+n}{2}}\,\Gamma[1+n]\ \text{Hypergeometric2F1}\left[1,\ -\frac{n}{2},\ 1 + \frac{n}{2},\ \frac{\theta}{-1+\theta}\right]\right)$$

The complicated nature of this expression rules out the possibility of a simple analytical procedure to compare MSEs for arbitrary n. However, by selecting a specific value of n, say $n = 4$, we can compare the estimators by plotting their MSEs, as illustrated in Fig. 4.

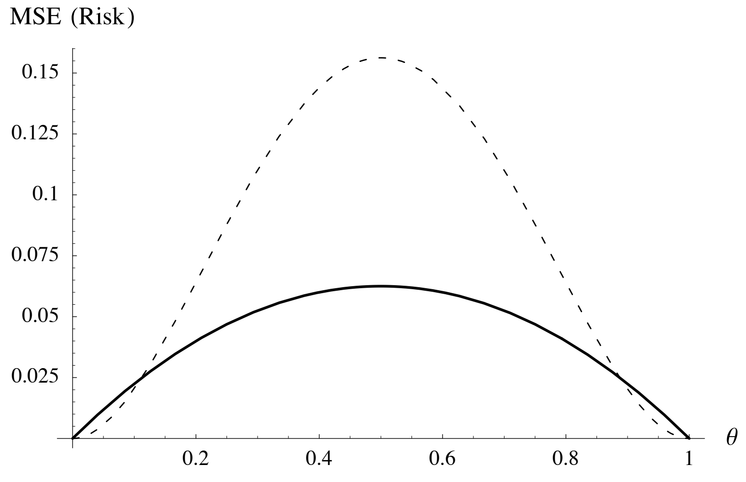


Fig. 4: $\hat{\text{MSE}}$ (——) and $\tilde{\text{MSE}}$ (– – –) when $n = 4$

Evidently, the risk of the sample mean ($\hat{\text{MSE}}$) is *nearly* everywhere below the risk of the sample median ($\tilde{\text{MSE}}$). Nevertheless, the plot shows that there exist values towards the edges of the parameter space where the sample median has lower risk than the sample mean; consequently, when $n = 4$, both estimators are admissible with respect to quadratic loss. Thus, to make a decision, we need to know θ_0, the true value of θ. Of course, it is precisely because θ_0 is unknown that we started this investigation. So, our decision-theoretic approach has left us in a situation of needing to know θ_0 in order to decide how to estimate it! Clearly, further information is required in order to reach a decision. To progress, we might do the following:

(i) Experiment with increasing n: that is, replace '4' with larger even numbers in the above analysis. On doing so, we find that the parameter region for which the sample mean is preferred also increases. This procedure motivates a formal asymptotic analysis of each MSE for $n \to \infty$, returning us to the type of analysis developed in Chapter 8.

(ii) Alter the decision criterion: for example, suppose the decision criterion was to select the minimax estimator — the estimator that has the smaller maximum risk. From the diagram, we see that the maximum risk of $\hat{\theta}$ (occurring at $\theta = \frac{1}{2}$) is smaller than the maximum risk of $\tilde{\theta}$ (also occurring at $\theta = \frac{1}{2}$). Hence, the sample mean is the minimax estimator.

(iii) Finally, comparing the number of points in $\tilde{\Omega}_{\text{odd}}$ or $\tilde{\Omega}_{\text{even}}$ (the sample space of the sample median), relative to the number of points in $\hat{\Omega}$ (the sample space of the sample mean) is probably sufficient argument to motivate selecting the sample mean over the sample median for Bernoulli trials, because the parameter space is the (0, 1) interval of the real line. ■

9.4 Order Statistics

9.4 A Definition and `OrderStat`

Let X denote a continuous random variable with pdf $f(x)$ and cdf $F(x)$, and let $(X_1, X_2, \ldots, X_n)$ denote a random sample of size n drawn on X. Suppose that we place the variables in the random sample in ascending order. The re-ordered variables, which we shall label $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, are known as *order statistics*. By construction, the order statistics are such that $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$; for example, $X_{(1)} = \min(X_1, \ldots, X_n)$ is the smallest order statistic and corresponds to the sample minimum, and $X_{(n)}$ is the largest order statistic and corresponds to the sample maximum. Each order statistic is a continuous random variable (this is inherited from X), and each has domain of support equivalent to that of X. For example, the pdf of $X_{(r)}$, the r^{th} order statistic ($r \in \{1, \ldots, n\}$), is given by[2]

$$\frac{n!}{(r-1)!\,(n-r)!}\, F(x)^{r-1} \left(1 - F(x)\right)^{n-r} f(x) \tag{9.4}$$

where x represents values assigned to $X_{(r)}$. Finally, because X is continuous, any ties (*i.e.* two identical outcomes) between the order statistics can be disregarded as ties occur with probability zero. For further discussion of order statistics, see David (1981), Balakrishnan and Rao (1998a, 1998b) and Hogg and Craig (1995).

mathStatica's `OrderStat` function automates the construction of the pdf of order statistics for a size n random sample drawn on a random variable X with pdf f. In particular, `OrderStat[r, f]` finds the pdf of the r^{th} order statistic, while `OrderStat[{r, s, …, t}, f]` finds the joint pdf of the order statistics indicated in the list. An optional third argument, `OrderStat[r, f, m]`, sets the sample size to m.

⊕ ***Example 4:*** Order Statistics for the Uniform Distribution

Let $X \sim$ Uniform(0, 1) with pdf $f(x)$:

```
f = 1;      domain[f] = {x, 0, 1};
```

Let $(X_1, \ldots, X_n)$ denote a random sample of size n drawn on X, and let $(X_{(1)}, \ldots, X_{(n)})$ denote the corresponding order statistics. Then, the pdf of the smallest order statistic $X_{(1)}$ is given by:

```
OrderStat[1, f]
```

$$n\,(1-x)^{-1+n}$$

The pdf of the largest order statistic $X_{(n)}$ is given by:

```
OrderStat[n, f]
```

$$n\,x^{-1+n}$$

and the pdf of the r^{th} order statistic is given by:

```
OrderStat[r, f]
```

$$\frac{(1-x)^{n-r}\,x^{-1+r}\,n!}{(n-r)!\,(-1+r)!}$$

Note that `OrderStat` assumes an arbitrary sample size n. If a specific value for sample size is required, or if you wish to use your own notation for 'n', then this may be conveyed using a third argument to `OrderStat`. For example, if $n = 5$, the pdf of the r^{th} order statistic is:

```
OrderStat[r, f, 5]
```

$$\frac{120\,(1-x)^{5-r}\,x^{-1+r}}{(5-r)!\,(-1+r)!}$$

In each case, the domain of support of $X_{(r)} = x \in (0, 1)$. ■

⊕ ***Example 5:*** Operating on Order Statistics

Let $X \sim \text{Exponential}(\lambda)$ with pdf $f(x)$:

$$f = \frac{1}{\lambda} e^{-x/\lambda}; \qquad \text{domain}[f] = \{x, 0, \infty\} \&\& \{\lambda > 0\};$$

Let $(X_1, \ldots, X_n)$ denote a random sample of size n drawn on X, and let $(X_{(1)}, \ldots, X_{(n)})$ denote the corresponding order statistics. Here is $g(x)$, the pdf of the r^{th} order statistic:

```
g = OrderStat[r, f]
```

$$\frac{e^{-\frac{(1+n-r)\,x}{\lambda}}\,\left(1 - e^{-\frac{x}{\lambda}}\right)^{-1+r}\,n!}{\lambda\,(n-r)!\,(-1+r)!}$$

The `domain` statement to accompany `g` may be found using **mathStatica**'s `OrderStatDomain` function:

```
domain[g] = OrderStatDomain[r, f]
```

{x, 0, ∞} && {n ∈ Integers, r ∈ Integers, λ > 0, 1 ≤ r ≤ n}

Figure 5 plots the pdf of $X_{(r)}$ as r increases.

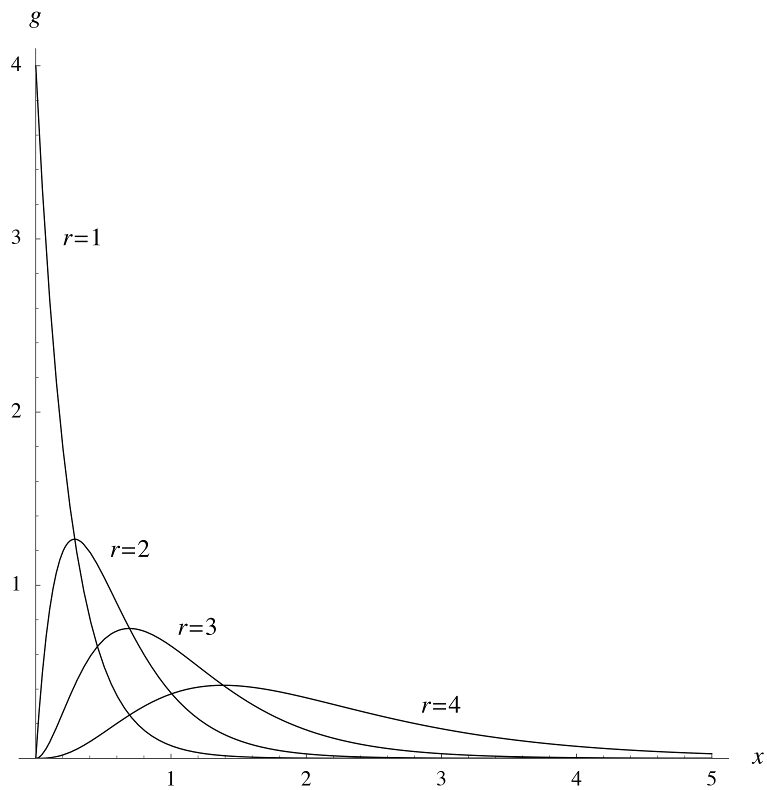


Fig. 5: The pdf of the r^{th} order statistic, as r increases (with $n = 4$, $\lambda = 1$)

We can now operate on $X_{(r)}$ using **mathStatica** functions. For example, here is the mean of $X_{(r)}$ as a function of n, r and λ:

```
Expect[x, g]
```

```
λ (HarmonicNumber[n] - HarmonicNumber[n - r])
```

and here is the variance:

```
Var[x, g]
```

```
λ² (-PolyGamma[1, 1 + n] + PolyGamma[1, 1 + n - r])
```

⊕ ***Example 6:*** Joint Distributions of Order Statistics

Once again, let $X \sim$ Exponential(λ) with pdf:

$$f = \frac{1}{\lambda} e^{-x/\lambda}; \qquad \text{domain}[f] = \{x, 0, \infty\} \&\& \{\lambda > 0\};$$

For a size n random sample drawn on X, the joint pdf of the order statistics $(X_{(1)}, X_{(2)})$ is:

```
g = OrderStat[{1, 2}, f]
```

$$\frac{e^{-\frac{x_1+(-1+n)\,x_2}{\lambda}}\,\Gamma[1+n]}{\lambda^2\,\Gamma[-1+n]}$$

```
domain[g] = OrderStatDomain[{1, 2}, f]
```

— The domain is: $\{0 < x_1 < x_2 < \infty\}$, which we enter into mathStatica as :

```
{{x1, 0, x2}, {x2, x1, ∞}} && {n ∈ Integers, λ > 0, 2 ≤ n}
```

where x_1 denotes values assigned to $X_{(1)}$, and x_2 denotes values assigned to $X_{(2)}$. In this bivariate case, the domain of support of $(X_{(1)}, X_{(2)})$ is given by the non-rectangular region $\Omega = \{(x_1, x_2) : 0 < x_1 < x_2 < \infty\}$. At present, **mathStatica** does not support non-rectangular regions (see §6.1 B). However, **mathStatica** functions such as `Expect`, `Var`, `Cov` and `Corr` do know how to operate on triangular regions which have general form $a < x < y < z < \cdots < b$, where a and b are constants. Here, for example, is the correlation coefficient between $X_{(1)}$ and $X_{(2)}$:

```
Corr[{x1, x2}, g]
```

$$\frac{-1+n}{\sqrt{1-2n+2n^2}}$$

The non-zero correlation coefficient and (especially) the non-rectangular domain of support of $(X_{(1)}, X_{(2)})$ illustrate a general property of order statistics — they are mutually dependent. ■

⊕ ***Example 7:*** Order Statistics for the Laplace Distribution

The `OrderStat` function also supports pdf's which take a piecewise form. For example, let random variable $X \sim \text{Laplace}(\mu, \sigma)$ with piecewise pdf:

$$\textbf{f = If}\Big[\mathbf{x < \mu,\ \frac{e^{\frac{x-\mu}{\sigma}}}{2\,\sigma},\ \frac{e^{-\frac{x-\mu}{\sigma}}}{2\,\sigma}}\Big];$$
$$\textbf{domain[f] = \{x, -}\infty\textbf{, }\infty\textbf{\} \&\& \{}\mu \in \textbf{Reals, } \sigma > 0\textbf{\};}$$

The pdf of the r^{th} order statistic, $X_{(r)}$, is given by:[3]

OrderStat[r, f]

$$\text{If}\Big[x < \mu,\ \frac{2^{-r}\, e^{\frac{r\,(x-\mu)}{\sigma}}\,\big(1-\frac{1}{2}\,e^{\frac{x-\mu}{\sigma}}\big)^{n-r}\, n!}{\sigma\,(n-r)!\,(-1+r)!},\ \frac{2^{-1-n+r}\, e^{\frac{(1+n-r)\,(-x+\mu)}{\sigma}}\,\big(1-\frac{1}{2}\,e^{\frac{-x+\mu}{\sigma}}\big)^{-1+r}\, n!}{\sigma\,(n-r)!\,(-1+r)!}\Big]$$

Notice that **mathStatica**'s output is in piecewise form too.

As a special case, let $X_{(1)}$ denote the smallest order statistic from a random sample of size n drawn on the standardised Laplace distribution (*i.e.* the Laplace(0, 1) distribution). The pdf of $X_{(1)}$ is given by:

$$\mathbf{g_1 = OrderStat[1, f\ /.\ \{\mu \to 0,\ \sigma \to 1\}]}$$

$$\text{If}\Big[x < 0,\ \frac{1}{2}\, e^{x}\,\Big(1-\frac{e^{x}}{2}\Big)^{-1+n}\, n,\ 2^{-n}\, e^{-n\,x}\, n\Big]$$

$$\mathbf{domain[g_1] = OrderStatDomain[1, f\ /.\ \{\mu \to 0,\ \sigma \to 1\}]}$$

{x, -∞, ∞} && {n ∈ Integers, 1 ≤ n}

Figure 6 shows how the pdf of $X_{(1)}$ varies as n increases. It is evident that the bulk of the mass of the pdf of $X_{(1)}$ shifts to the left, as n increases.

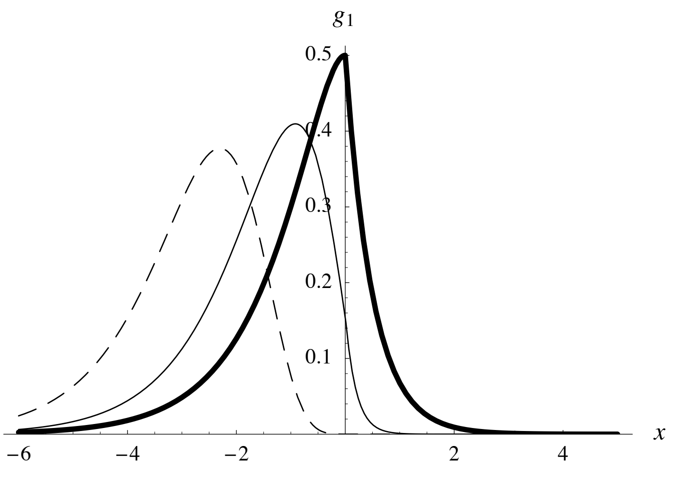


Fig. 6: pdf of $X_{(1)}$: $n = 2$ (▬), $n = 5$ (—), $n = 20$ (– – –)

As a final illustration, consider the joint pdf of the two smallest order statistics, $X_{(1)}$ and $X_{(2)}$, when the sample size is $n = 5$:

```
g12 = OrderStat[{1, 2}, f /. {μ → 0, σ → 1}, 5]
```

$$20\,\text{If}\left[x_1 < 0,\ \frac{e^{x_1}}{2},\ \frac{e^{-x_1}}{2}\right]\ \text{If}\left[x_2 < 0,\ -\frac{1}{16}\,e^{x_2}\,(-2 + e^{x_2})^3,\ \frac{1}{16}\,e^{-4\,x_2}\right]$$

The joint pdf is illustrated from differing perspectives in Fig. 7 and Fig. 8.

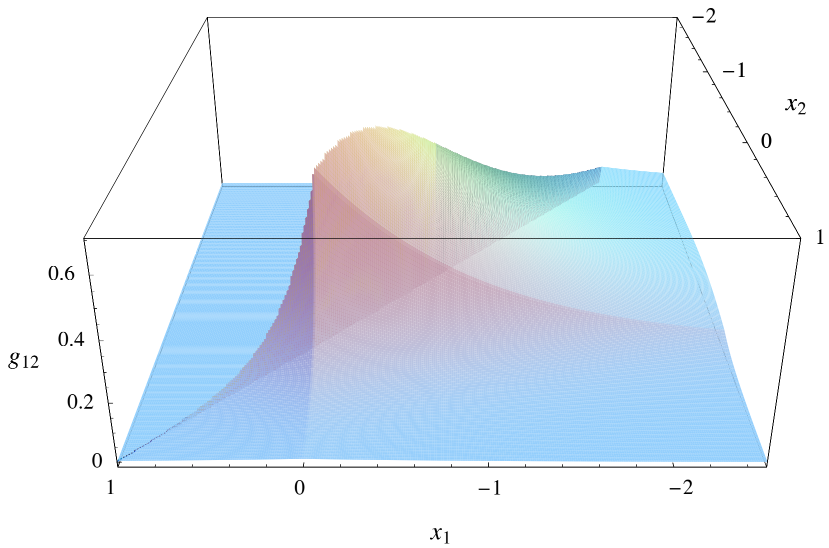


Fig. 7: pdf of g_{12} ('front' view)

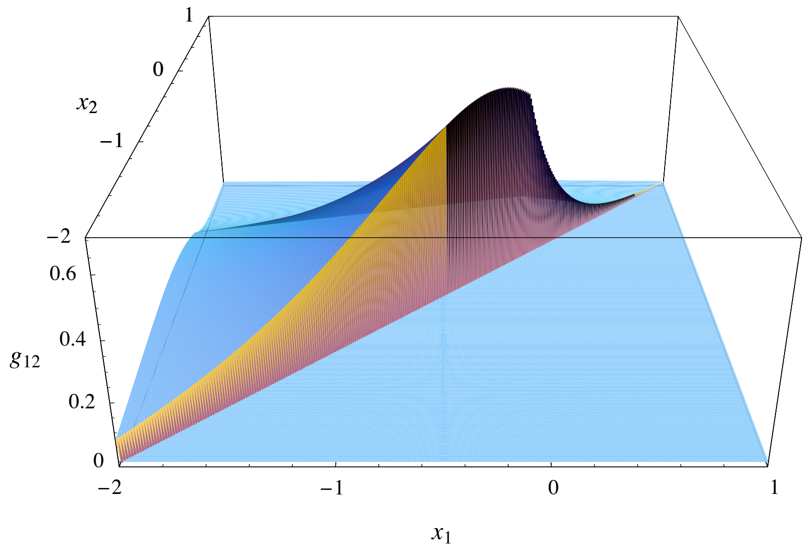


Fig. 8: pdf of g_{12} ('rear' view)

In Fig. 7, ridges are evident along the lines $x_1 = 0$ and $x_2 = 0$; this is consistent with the piecewise nature of g_{12}. In Fig. 8, the face of the plane $x_1 = x_2$ is prominent; this neatly illustrates the domain of support of $X_{(1)}$ and $X_{(2)}$ (*viz.* the triangular region $\{(x_1, x_2) : -\infty < x_1 < x_2 < \infty\}$). The domain of support can also be illustrated as the shaded region in Fig. 9.

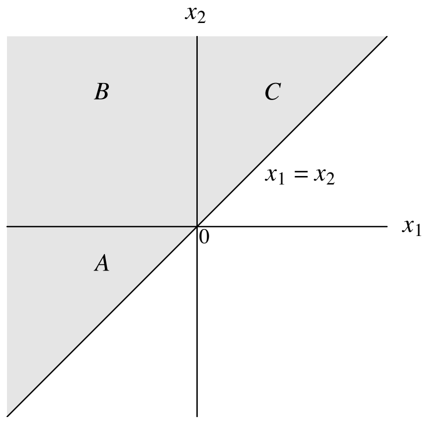


Fig. 9: Domain of support of $X_{(1)}$ and $X_{(2)}$ (shaded)

mathStatica cannot operate on the pdf, because the pdf has a multiple `If` structure. However, we may proceed by separating the domain of support into three distinct regions, labelled *A*, *B* and *C* in Fig. 9. In the triangular region *A*, the pdf of $X_{(1)}$ and $X_{(2)}$ is given by:

```
Ag12 = Simplify[g12, {x1 < 0, x2 < 0}]
```

$$-\frac{5}{8}\, e^{x_1+x_2}\ (-2 + e^{x_2})^3$$

… while in the rectangular region *B*, the pdf is given by:

```
Bg12 = Simplify[g12, {x1 < 0, x2 > 0}]
```

$$\frac{5}{8}\, e^{x_1-4\,x_2}$$

… and finally, in region *C*, the pdf is:

```
Cg12 = Simplify[g12, {x1 > 0, x2 > 0}]
```

$$\frac{5}{8}\, e^{-x_1-4\,x_2}$$

In this way, we can verify that the pdf integrates to unity over its domain of support:

$$\int_{-\infty}^{0}\int_{-\infty}^{x_2} \mathrm{Ag}_{12}\, dx_1\, dx_2 + \int_{0}^{\infty}\int_{-\infty}^{0} \mathrm{Bg}_{12}\, dx_1\, dx_2 + \int_{0}^{\infty}\int_{0}^{x_2} \mathrm{Cg}_{12}\, dx_1\, dx_2$$

1

9.4 B Applications

Estimators such as the sample median (used to estimate location) and the sample interquartile range (to estimate scale) may be constructed from the order statistics of a random sample. In *Example 8*, we derive the MSE of the sample median, while in *Example 9* we derive the MSE of the sample range (a function of two order statistics).

⊕ ***Example 8:*** Sample Median versus Sample Mean

Two estimators of location are the sample median and the sample mean. In this example, we compare the MSE performance of each estimator when $X \sim \text{Logistic}(\theta)$, the location-shifted Logistic distribution with pdf $f(x)$:

$$\mathbf{f} = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}; \qquad \textbf{domain[f] = \{x, -∞, ∞\} \&\& \{θ ∈ Reals\};}$$

where $\theta \in \mathbb{R}$ is the location parameter (the mean of X). For simplicity, we assume that a random sample of size n drawn on X is odd-sized (*i.e.* n is odd), and so we shall write $n = 2r + 1$, for $r \in \{1, 2, \ldots\}$. Therefore, the sample median, which we denote by M, corresponds to the middle order statistic $X_{(r+1)}$. Thus, the pdf of M is given by:

g = OrderStat[r + 1, f, 2 r + 1] /. x → m

$$\frac{e^{(1+r)(m+\theta)}\,(e^m + e^\theta)^{-2(1+r)}\,(1+2r)!}{r!^2}$$

Here is the domain of support of M:

domain[g] = {m, -∞, ∞} && {θ ∈ Reals, r > 0};

The MSE of the sample median is given by $E[(M-\theta)^2]$. Unfortunately, if we evaluate `Expect[(m - θ)^2, g]`, an unsolved integral is returned. There are two possible reasons for this: (i) either *Mathematica* does not know how to solve this integral, or (ii) *Mathematica* can solve the integral, but needs a bit of help![4] In this case, we can help out by expressing the integrand in a simpler form. Since we want $E[(M-\theta)^2] = E[U^2]$, consider transforming M to the new variable $U = M - \theta$. The pdf of U, say g_u, is obtained using **mathStatica**'s `Transform` function:

g_u = Transform[u == m - θ, g]

$$\frac{e^{(1+r)u}\,(1+e^u)^{-2(1+r)}\,(1+2r)!}{r!^2}$$

domain[g_u] = TransformExtremum[u == m - θ, g]

```
{u, -∞, ∞} && {r > 0}
```

Since the functional form of the pdf of U does not depend upon θ, it follows that the MSE cannot depend on the value of θ. To make things even simpler, we make the further transformation $V = e^U$. Then, the pdf of V, denoted g_v, is:

```
gv = Transform[v == ⅇ^u, gu]
```

$$\frac{v^r\,(1+v)^{-2\,(1+r)}\,(1+2\,r)!}{r!^2}$$

```
domain[gv] = TransformExtremum[v == ⅇ^u, gu]
```

{v, 0, ∞} && {r > 0}

Since $V = \exp(U)$, it follows that $E[U^2] = E[(\log V)^2]$. Therefore, the MSE of the sample median is:

```
MSEmed = Expect[Log[v]^2, gv]
```

2 PolyGamma[1, 1 + r]

Our other estimator of location is the sample mean $\overline{X}$. To obtain its MSE, we must evaluate $E\left[(\overline{X}-\theta)^2\right]$. Because $\overline{X} = s_1/n$, where $s_1 = \sum_{i=1}^{n} X_i$ is the sample sum, the MSE is an expression involving power sums, and we can therefore use **mathStatica**'s Moments of Moments toolset (see §7.3) to solve the expectation. The MSE corresponds to the 1[st] raw moment of $(\frac{1}{n}\,s_1-\theta)^2$, and so we shall present the answer in terms of raw population moments of X (hence `ToRaw`):

```
sol = RawMomentToRaw[1, (s1/n - θ)^2]
```

$$-2\,\theta\,\acute{\mu}_1 + \frac{(-1+n)\,\acute{\mu}_1^2}{n} + \frac{n\,\theta^2 + \acute{\mu}_2}{n}$$

We now find $\acute{\mu}_1$ and $\acute{\mu}_2$, and substitute these values into the solution:

```
MSEmean =
  sol /. Table[μ́i → Expect[x^i, f], {i, 2}] // Simplify
```

$$\frac{\pi^2}{3\,n}$$

where $n = 2\,r+1$.

Both MSE_{med} and MSE_{mean} are independent of θ, but vary with sample size. We can compare the performance of each estimator by plotting their respective MSE for various values of r, see Fig. 10.

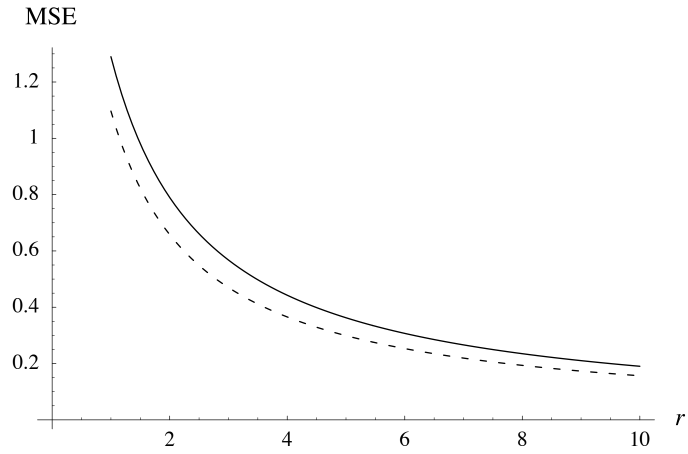


Fig. 10: MSE of sample mean (– – –) and sample median (—)

We see that the MSE of the sample mean (the dashed line) is everywhere below the MSE of the sample median (the unbroken line), and that this persists for all r. Hence, the sample mean dominates the sample median in mean square error (risk under quadratic loss) when estimating θ. We conclude that the sample median is inadmissible in this situation. However, this does not imply that the sample mean is admissible, for there may exist another estimator that dominates the sample mean under quadratic loss. ■

⊕ ***Example 9:*** Sample Range versus Largest Order Statistic

Let $X \sim \text{Uniform}(0, \theta)$, where $\theta \in \mathbb{R}_+$ is an unknown parameter, with pdf:

$$\mathbf{f} = \frac{1}{\theta}; \qquad \mathbf{domain[f] = \{x, 0, \theta\} \&\& \{\theta > 0\};}$$

The sample range R is defined as the distance between the smallest and largest order statistics; that is, $R = X_{(n)} - X_{(1)}$. It may be used to estimate θ. Another estimator is the sample maximum, corresponding to the largest order statistic $X_{(n)}$. In this example, we compare the performance of both estimators on the basis of their respective MSE.

To derive the distribution of R, we first obtain the joint pdf of $X_{(1)}$ and $X_{(n)}$:

```
g = OrderStat[{1, n}, f] // FunctionExpand
```

$$\frac{(-1+n)\, n \left(\frac{-x_1+x_n}{\theta}\right)^n}{(-x_1+x_n)^2}$$

with non-rectangular domain of support:

```
domain[g] = OrderStatDomain[{1, n}, f]
```

— The domain is: $\{0 < x_1 < x_n < \theta\}$, which we enter into mathStatica as :

$\{\{x_1, 0, x_n\}, \{x_n, x_1, \theta\}\} \&\& \{n \in \text{Integers}, \theta > 0, 1 < n\}$

We use **mathStatica**'s `Transform` function to perform the transformation from $(X_{(1)}, X_{(n)})$ to (R, S), where $S = X_{(1)}$. Here is the joint pdf of (R, S):

```
grs = Transform[{r == xn - x1, s == x1}, g]
```

$$\frac{(-1+n)\, n\, \left(\frac{r}{\theta}\right)^n}{r^2}$$

with non-rectangular support $\{(r, s) : 0 < r < \theta,\ 0 < s < \theta - r\}$. Integrating out S yields the pdf of R:

$$g_r = \int_0^{\theta - r} g_{rs}\, ds$$

$$\frac{(-1+n)\, n\, \left(\frac{r}{\theta}\right)^n (-r+\theta)}{r^2}$$

```
domain[gr] = {r, 0, θ} && {θ > 0, n > 1, n ∈ Integers};
```

The MSE for the sample range is:

```
MSErange = Expect[(r - θ)^2, gr]
```

$$\frac{6\,\theta^2}{2 + 3\,n + n^2}$$

Our other estimator of θ is the sample maximum $X_{(n)}$. The pdf of $X_{(n)}$ is:

```
gn = OrderStat[n, f]
```

$$\frac{n\, \left(\frac{x}{\theta}\right)^n}{x}$$

```
domain[gn] = OrderStatDomain[n, f]
```

$\{x, 0, \theta\}$ && $\{n \in \text{Integers}, \theta > 0, 1 \le n\}$

The MSE of $X_{(n)}$ is:

```
MSEmax = Expect[(x - θ)^2, gn]
```

$$\frac{2\,\theta^2}{2 + 3\,n + n^2}$$

so $\text{MSE}_{\text{range}} = 3\ \text{MSE}_{\text{max}}$ for all permissible values of θ and n. Therefore, the sample range is inadmissible.

Inadmissibility of the sample range does not imply that the sample maximum is admissible. Indeed, consider the following estimator that scales the sample maximum:

$$X^*_{(n)} = \frac{n+1}{n} X_{(n)}.$$

The MSE of the scaled estimator is:

```
MSEscaled = Expect[((n + 1)/n x - θ)^2, gn]
```

$$\frac{\theta^2}{n\,(2+n)}$$

Dividing by the MSE of $X_{(n)}$ finds:

```
MSEscaled/MSEmax // Simplify
```

$$\frac{1+n}{2\,n}$$

which is strictly less than unity for all $n > 1$, implying that the sample maximum $X_{(n)}$ is inadmissible too! ■

9.5 Exercises

1. Let $\hat{\theta}$ denote an estimator of an unknown parameter θ, and let $a > 0$ and $b > 0$ $(a \neq b)$ denote constants. Consider the asymmetric quadratic loss function

$$L(\hat{\theta}, \theta) = \begin{cases} a(\hat{\theta} - \theta)^2 & \text{if } \hat{\theta} > \theta \\ b(\hat{\theta} - \theta)^2 & \text{if } \hat{\theta} \leq \theta. \end{cases}$$

Plot the loss function against values of $(\hat{\theta} - \theta)$, when $a = 1$ and $b = 2$.

2. Varian (1975) introduced the linex (linear–exponential) loss function

$$L(\hat{\theta}, \theta) = e^{c(\hat{\theta} - \theta)} - c(\hat{\theta} - \theta) - 1$$

where $\hat{\theta}$ denotes an estimator of an unknown parameter θ, and constant $c \neq 0$.

(i) Investigate the loss function by plotting L against $(\hat{\theta} - \theta)$ for various values of c.

(ii) Using linear–exponential loss in the context of *Example 1* (*i.e.* $X \sim N(\theta, 1)$ and $\hat{\theta} = X + k$), determine the value of k which minimises risk.

3. Suppose that $X \sim$ Exponential(θ), where $\theta > 0$ is an unknown parameter. The random variable $\hat{\theta} = X / k$ is proposed as an estimator of θ, where constant $k > 0$. Obtain the risk, and the value of k which minimises risk, when the loss function is:

(i) symmetric quadratic $L_1(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.

(ii) linear–exponential $L_2(\hat{\theta}, \theta) = e^{\hat{\theta} - \theta} - (\hat{\theta} - \theta) - 1$.

4. Let random variable T have the same pdf $f(t)$ as used in *Example 2*. For estimators of θ of general form $\hat{\Theta} = T/(n+k)$, where real $k > -n$, consider the asymmetric quadratic loss function

$$L(\hat{\Theta}, \theta) = \begin{cases} (\hat{\Theta} - \theta)^2 & \text{if } \hat{\Theta} > \theta \\ b(\hat{\Theta} - \theta)^2 & \text{if } \hat{\Theta} \le \theta. \end{cases}$$

(i) After transforming from T to $\hat{\Theta}$, derive the risk of $\hat{\Theta}$ as a function of θ, n, k and b (the solution takes about 140 seconds to compute on our reference machine).

(ii) Explain why the minimum risk estimator does not depend on θ.

(iii) Setting $n = 10$, use numerical methods to determine the value of k which yields the minimum risk estimator when (a) $b = \frac{1}{2}$ and (b) $b = 2$. Do your results make sense?

5. Let $X_{(n)}$ denote the largest order statistic of a random sample of size n from $X \sim \text{Beta}(a, b)$.

(i) Derive the pdf of $X_{(n)}$.

(ii) Use `PlotDensity` to plot (on a single diagram) the pdf of $X_{(n)}$ when $a = 2$, $b = 3$ and $n = 2$, 4 and 6.

6. Let $X_{(1)}$, $X_{(2)}$ and $X_{(3)}$ denote the order statistics of a random sample of size $n = 3$ from $X \sim N(0, 1)$.

(i) Derive the pdf and cdf of each order statistic.

(ii) Use `PlotDensity` to plot (on a single diagram) the pdf of each order statistic (use the interval $(-3, 3)$).

(iii) Determine $E[X_{(r)}]$ for $r = 1, 2, 3$.

(iv) The pdf of $X_{(1)}$ and the pdf of $X_{(3)}$ appear to be similar—perhaps they differ by a simple mean shift? Test this assertion by plotting (on a single diagram) the pdf of $X_{(3)}$ and Y, where the random variable $Y = X_{(1)} + 3/\sqrt{\pi}$.

7. Apply the loss function $L_k(\hat{\Theta}, \theta) = |\hat{\Theta} - \theta|^k$ in the context of *Example 9* (note: symmetric quadratic loss corresponds to the special case $k = 2$). Find the values of k for which the sample maximum dominates the sample range.