Review: Random variables and sampling theory

0.12 Overview

The textbook and this subject guide assume that you have previously studied basic statistical theory and have a sound understanding of the following topics:

- descriptive statistics (mean, median, quartile, variance, etc.)
- random variables and probability
- expectations and expected value rules
- population variance, covariance, and correlation
- sampling theory and estimation
- unbiasedness and efficiency
- loss functions and mean square error
- normal distribution
- hypothesis testing, including:
  - t tests
  - Type I and Type II error
  - the significance level and power of a t test
  - one-sided versus two-sided t tests
- confidence intervals
- convergence in probability, consistency, and plim rules
- convergence in distribution and central limit theorems.

There are many excellent textbooks that offer a first course in statistics. The Review chapter of my textbook is not a substitute. It has the much more limited objective of providing an opportunity for revising some key statistical concepts and results that will be used time and time again in the course. They are central to econometric analysis and if you have not encountered them before, you should postpone your study of econometrics and study statistics first.

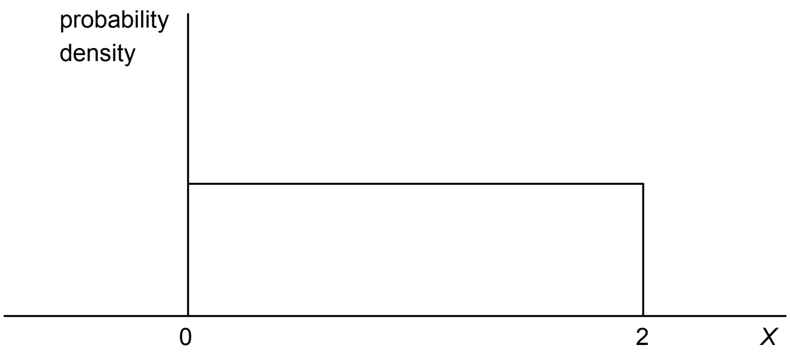
0.13 Learning outcomes

After working through the corresponding chapter in the textbook, studying the corresponding slideshows, and doing the starred exercises in the textbook and the additional exercises in this subject guide, you should be able to explain what is meant by all of the items listed in the Overview. You should also be able to explain why they are important. The concepts of efficiency, consistency, and power are often misunderstood by students taking an introductory econometrics course, so make sure that you aware of their precise meanings.

0.14 Additional exercises

[**Note:** Each chapter has a set of additional exercises. The answers to them are provided at the end of the chapter after the answers to the starred exercises in the text.]

AR.1 A random variable X has a continuous uniform distribution from 0 to 2. Define its probability density function.



AR.2 Find the expected value of X in Exercise AR.1, using the expression given in Box R.1 in the text.

AR.3 Derive $E(X^2)$ for X defined in Exercise AR.1, using the expression given in Box R.1.

AR.4 Derive the population variance and the standard deviation of X as defined in Exercise AR.1, using the expression given in Box R.1.

AR.5 Using equation (R.9), find the variance of the random variable X defined in Exercise AR.1 and show that the answer is the same as that obtained in Exercise AR.4. (Note: You have already calculated $E(X)$ in Exercise AR.2 and $E(X^2)$ in Exercise AR.3.)

AR.6 In Table R.6, μ_0 and μ_1 were three standard deviations apart. Construct a similar table for the case where they are two standard deviations apart.

AR.7 Suppose that a random variable X has a normal distribution with unknown mean μ and variance σ^2. To simplify the analysis, we shall assume that σ^2 is known. Given a sample of observations, an estimator of μ is the sample mean, $\overline{X}$. An investigator wishes to test $H_0 : \mu = 0$ and believes that the true value cannot be negative. The appropriate alternative hypothesis is therefore $H_1 : \mu > 0$ and the investigator decides to perform a one-sided test. However, the investigator is mistaken because μ could in fact be negative. What are the consequences of erroneously performing a one-sided test when a two-sided test would have been appropriate?

AR.8 Suppose that a random variable X has a normal distribution with mean μ and variance σ^2. Given a sample of n independent observations, it can be shown that:

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2$$

is an unbiased estimator of σ^2. Is $\sqrt{\widehat{\sigma}^2}$ either an unbiased or a consistent estimator of σ?

0.15 Answers to the starred exercises in the textbook

R.2 A random variable X is defined to be the larger of the two values when two dice are thrown, or the value if the values are the same. Find the probability distribution for X.

Answer:

The table shows the 36 possible outcomes. The probability distribution is derived by counting the number of times each outcome occurs and dividing by 36. The probabilities have been written as fractions, but they could equally well have been written as decimals.

red / green	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Value of X	1	2	3	4	5	6
Frequency	1	3	5	7	9	11
Probability	1/36	3/36	5/36	7/36	9/36	11/36

R.4 Find the expected value of X in Exercise R.2.

Answer:

The table is based on Table R.2 in the text. It is a good idea to guess the outcome before doing the arithmetic. In this case, since the higher numbers have the largest probabilities, the expected value should clearly lie between 4 and 5. If the calculated value does not conform with the guess, it is possible that this is because the guess was poor. However, it may be because there is an error in the arithmetic, and this is one way of catching such errors.

X	p	Xp
1	1/36	1/36
2	3/36	6/36
3	5/36	15/36
4	7/36	28/36
5	9/36	45/36
6	11/36	66/36
Total		161/36 = 4.4722

R.7 Calculate $E(X^2)$ for X defined in Exercise R.2.

Answer:

The table is based on Table R.3 in the text. Given that the largest values of X^2 have the highest probabilities, it is reasonable to suppose that the answer lies somewhere in the range 15–30. The actual figure is 21.97.

X	X^2	p	X^2p
1	1	1/36	1/36
2	4	3/36	12/36
3	9	5/36	45/36
4	16	7/36	112/36
5	25	9/36	225/36
6	36	11/36	396/36
Total			791/36 = 21.9722

R.10 Calculate the population variance and the standard deviation of X as defined in Exercise R.2, using the definition given by equation (R.8).

Answer:

The table is based on Table R.4 in the textbook. In this case it is not easy to make a guess. The population variance is 1.97, and the standard deviation, its square root, is 1.40. Note that four decimal places have been used in the working, even though the estimate is reported to only two. This is to eliminate the possibility of the estimate being affected by rounding error.

X	p	$X-\mu_X$	$(X-\mu_X)^2$	$(X-\mu_X)^2p$
1	1/36	−3.4722	12.0563	0.3349
2	3/36	−2.4722	6.1119	0.5093
3	5/36	−1.4722	2.1674	0.3010
4	7/36	−0.4722	0.2230	0.0434
5	9/36	0.5278	0.2785	0.0696
6	11/36	1.5278	2.3341	0.7132
Total				1.9715

R.12 Using equation (R.9), find the variance of the random variable X defined in Exercise R.2 and show that the answer is the same as that obtained in Exercise R.10. (Note: You have already calculated μ_X in Exercise R.4 and $E(X^2)$ in Exercise R.7.)

Answer:

$E(X^2)$ is 21.9722 (Exercise R.7). $E(X)$ is 4.4722 (Exercise R.4), so μ_X^2 is 20.0006. Thus the variance is 21.9722 – 20.0006 = 1.9716. The last-digit discrepancy between this figure and that in Exercise R.10 is due to rounding error.

R.14 Suppose a variable Y is an exact linear function of X:

$$Y = \lambda + \mu X$$

where λ and μ are constants, and suppose that Z is a third variable. Show that $\rho_{XZ} = \rho_{YZ}$

Answer:

We start by noting that $\left(Y_i - \overline{Y}\right) = \mu\left(X_i - \overline{X}\right)$. Then:

$$\begin{aligned}
\rho_{YZ} &= \frac{E\left[\left(Y_i-\overline{Y}\right)\left(Z_i-\overline{Z}\right)\right]}{\sqrt{E\left[\left(Y_i-\overline{Y}\right)^2\right]E\left[\left(Z_i-\overline{Z}\right)^2\right]}} \\
&= \frac{E\left[\mu\left(X_i-\overline{X}\right)\left(Z_i-\overline{Z}\right)\right]}{\sqrt{E\left[\mu^2\left(X_i-\overline{X}\right)^2\right]E\left[\mu^2\left(Z_i-\overline{Z}\right)^2\right]}} \\
&= \frac{\mu E\left[\left(X_i-\overline{X}\right)\left(Z_i-\overline{Z}\right)\right]}{\sqrt{\mu^2 E\left[\left(X_i-\overline{X}\right)^2\right]E\left[\left(Z_i-\overline{Z}\right)^2\right]}} \\
&= \rho_{XZ}.
\end{aligned}$$

R.16 Show that, when you have n observations, the condition that the generalised estimator $(\lambda_1 X_1 + \cdots + \lambda_n X_n)$ should be an unbiased estimator of μ_X is $\lambda_1 + \cdots + \lambda_n = 1$.

Answer:

$$\begin{aligned} E(Z) &= E(\lambda_1 X_1 + \cdots + \lambda_n X_n) \\ &= E(\lambda_1 X_1) + \cdots + E(\lambda_n X_n) \\ &= \lambda_1 E(X_1) + \cdots + \lambda_n E(X_n) \\ &= \lambda_1 \mu_X + \cdots + \lambda_n \mu_X \\ &= (\lambda_1 + \cdots + \lambda_n)\mu_X. \end{aligned}$$

Thus $E(Z) = \mu_X$ requires $\lambda_1 + \cdots + \lambda_n = 1$.

R.19 In general, the variance of the distribution of an estimator decreases when the sample size is increased. Is it correct to describe the estimator as becoming more efficient?

Answer:

No, it is incorrect. When the sample size increases, the variance of the estimator decreases, and as a consequence it is more likely to give accurate results. Because it is improving in this important sense, it is very tempting to describe the estimator as becoming more efficient. But it is the wrong use of the term. Efficiency is a comparative concept that is used when you are comparing two or more alternative estimators, all of them being applied to the same data set with the same sample size. The estimator with the smallest variance is said to be the most efficient. You cannot use efficiency as suggested in the question because you are comparing the variances of the *same* estimator with *different* sample sizes.

R.21 Suppose that you have observations on three variables X, Y, and Z, and suppose that Y is an exact linear function of Z:

$$Y = \lambda + \mu Z$$

where λ and μ are constants. Show that $\widehat{\rho}_{XZ} = \widehat{\rho}_{XY}$. (This is the counterpart of Exercise R.14.)

Answer:

We start by noting that $\left(Y_i - \overline{Y}\right) = \mu\left(Z_i - \overline{Z}\right)$. Then:

$$\begin{aligned} \widehat{\rho}_{XY} &= \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2}} \\ &= \frac{\sum\left(X_i - \overline{X}\right)\mu\left(Z_i - \overline{Z}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\mu^2\left(Z_i - \overline{Z}\right)^2}} \\ &= \frac{\sum\left(X_i - \overline{X}\right)\left(Z_i - \overline{Z}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\left(Z_i - \overline{Z}\right)^2}} \\ &= \widehat{\rho}_{XZ} \end{aligned}$$

R.26 Show that, in Figures R.18 and R.22, the probabilities of a Type II error are 0.15 in the case of a 5 per cent significance test and 0.34 in the case of a 1 per cent test. Note that the distance between μ_0 and μ_1 is three standard deviations. Hence the right-hand 5 per cent rejection region begins 1.96 standard deviations to the right of μ_0. This means that it is located 1.04 standard deviations to the left of μ_1. Similarly, for a 1 per cent test, the right-hand rejection region starts 2.58 standard deviations to the right of μ_0, which is 0.42 standard deviations to the left of μ_1.

Answer:

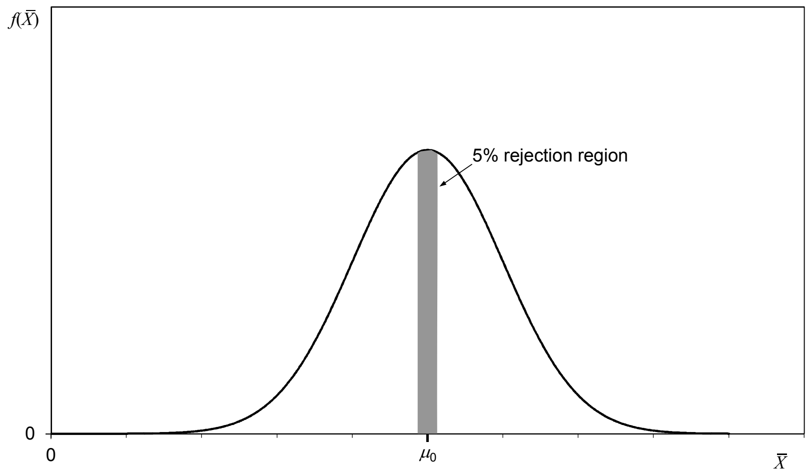
For the 5 per cent test, the rejection region starts $3 - 1.96 = 1.04$ standard deviations below μ_1, given that the distance between μ_1 and μ_0 is 3 standard deviations. See Figure R.18. According to the standard normal distribution table, the cumulative probability of a random variable lying 1.04 standard deviations (or less) above the mean is 0.8508. This implies that the probability of it lying 1.04 standard deviations below the mean is 0.1492. For the 1 per cent test, the rejection region starts $3 - 2.58 = 0.42$ standard deviations below the mean. See Figure R.22. The cumulative probability for 0.42 in the standard normal distribution table is 0.6628, so the probability of a Type II error is 0.3372.

R.27 Explain why the difference in the power of a 5 per cent test and a 1 per cent test becomes small when the distance between μ_0 and μ_1 becomes large.

Answer:

The powers of both tests tend to one as the distance between μ_0 and μ_1 becomes large. The difference in their powers must therefore tend to zero.

R.28 A random variable X has unknown population mean μ. A researcher has a sample of observations with sample mean $\overline{X}$. He wishes to test the null hypothesis $H_0 : \mu = \mu_0$. The figure shows the potential distribution of $\overline{X}$ conditional on H_0 being true. It may be assumed that the distribution is known to have variance equal to one.



The researcher decides to implement an unorthodox (and unwise) decision rule. He decides to reject H_0 if $\overline{X}$ lies in the central 5 per cent of the distribution (the tinted area in the figure).

(a) Explain why his test is a 5 per cent significance test.

(b) Explain in intuitive terms why his test is unwise.

(c) Explain in technical terms why his test is unwise.

Answer:

The following discussion assumes that you are performing a 5 per cent significance test, but it applies to any significance level.

If the null hypothesis is true, it does not matter how you define the 5 per cent rejection region. By construction, the risk of making a Type I error will be 5 per cent. Issues relating to Type II errors are irrelevant when the null hypothesis is true.

The reason that the central part of the conditional distribution is not used as a rejection region is that it leads to problems when the null hypothesis is false. The probability of not rejecting H_0 when it is false will be lower. To use the obvious technical term, the power of the test will be lower.

The figure shows the power functions for the test using the conventional upper and lower 2.5 per cent tails and the test using the central region. The horizontal axis is the difference between the true value and the hypothetical value μ_0 in terms of standard deviations. The vertical axis is the power of the test. The first figure has been drawn for the case where the true value is greater than the hypothetical value. The second figure is for the case where the true value is lower than the hypothetical value. It is the same, but reflected horizontally.

The greater the difference between the true value and the hypothetical mean, the more likely is it that the sample mean will lie in the right tail of the distribution conditional on H_0 being true, and so the more likely is it that the null hypothesis will be rejected by the conventional test. The figure shows that the power of the test approaches one asymptotically. However, if the central region of the distribution is used as the rejection region, the probability of the sample mean lying in it will *diminish* as the difference between the true and hypothetical values increases, and the power of the test approaches zero asymptotically. This is an extreme example of a very bad test procedure.

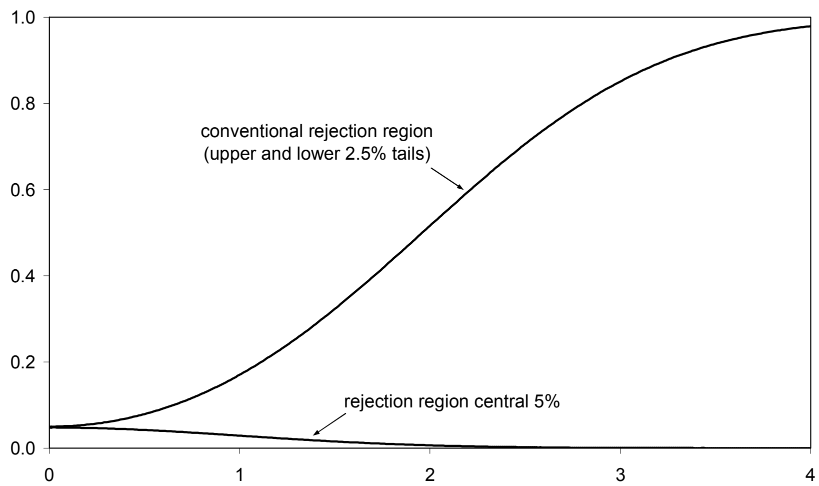


Figure 1: Power functions of a conventional 5 per cent test and one using the central region (true value > μ_0).

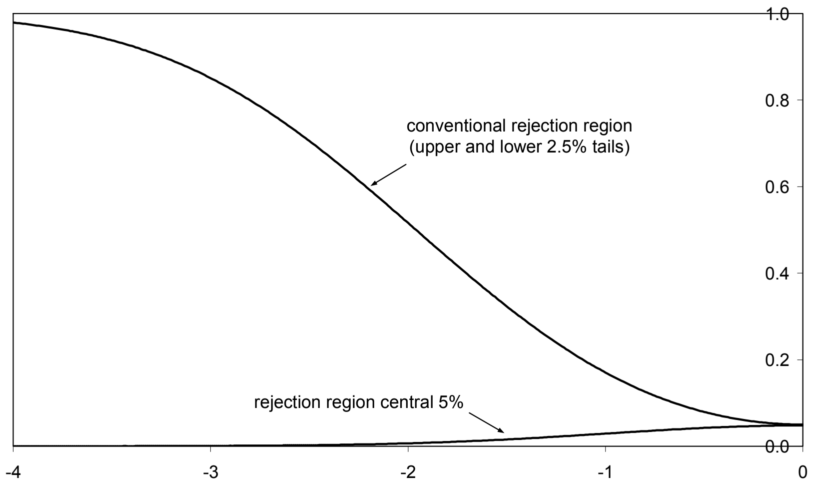


Figure 2: Power functions of a conventional 5 per cent test and one using the central region (true value $< \mu_0$).

R.29 A researcher is evaluating whether an increase in the minimum hourly wage has had an effect on employment in the manufacturing industry in the following three months. Taking a sample of 25 firms, what should she conclude if:

(a) the mean decrease in employment is 9 per cent, and the standard error of the mean is 5 per cent

(b) the mean decrease is 12 per cent, and the standard error is 5 per cent

(c) the mean decrease is 20 per cent, and the standard error is 5 per cent

(d) there is a mean *increase* of 11 per cent, and the standard error is 5 per cent?

Answer:

There are 24 degrees of freedom, and hence the critical values of t at the 5 per cent, 1 per cent, and 0.1 per cent levels are 2.06, 2.80, and 3.75, respectively.

(a) The t statistic is -1.80. Fail to reject H_0 at the 5 per cent level.

(b) $t = -2.40$. Reject H_0 at the 5 per cent level but not the 1 per cent level.

(c) $t = -4.00$. Reject H_0 at the 1 per cent level. Better, reject at the 0.1 per cent level.

(d) $t = 2.20$. This would be a surprising outcome, but if one is performing a two-sided test, then reject H_0 at the 5 per cent level but not the 1 per cent level.

R3.33 Demonstrate that the 95 per cent confidence interval defined by equation (R.89) has a 95 per cent probability of capturing μ_0 if H_0 is true.

Answer:

If H_0 is true, there is 95 per cent probability that:

$$\left| \frac{\overline{X} - \mu_0}{\text{s.e.}(\overline{X})} \right| < t_{\text{crit}}.$$

Hence there is 95 per cent probability that $|\overline{X} - \mu_0| \leq t_{\text{crit}} \times \text{s.e.}(\overline{X})$. Hence there is 95 per cent probability that (a) $\overline{X} - \mu_0 < t_{\text{crit}} \times \text{s.e.}(\overline{X})$ and (b) $\mu_0 - \overline{X} < t_{\text{crit}} \times \text{s.e.}(\overline{X})$.

(a) can be rewritten $\overline{X} - t_{\text{crit}} \times \text{s.e.}(\overline{X}) < \mu_0$, giving the lower limit of the confidence interval.

(b) can be rewritten $\overline{X} - \mu_0 > -t_{\text{crit}} \times \text{s.e.}(\overline{X})$ and hence $\overline{X} + t_{\text{crit}} \times \text{s.e.}(\overline{X}) > \mu_0$, giving the upper limit of the confidence interval.

Hence there is 95 per cent probability that μ_0 will lie in the confidence interval.

R.34 In Exercise R.29, a researcher was evaluating whether an increase in the minimum hourly wage has had an effect on employment in the manufacturing industry. Explain whether she might have been justified in performing one-sided tests in cases (a) – (d), and determine whether her conclusions would have been different.

Answer:

First, there should be a discussion of whether the effect of an increase in the minimum wage could have a positive effect on employment. If it is decided that it cannot, we can use a one-sided test and the critical values of t at the 5 per cent, 1 per cent, and 0.1 per cent levels become 1.71, 2.49, and 3.47, respectively.

1. The t statistic is -1.80. We can now reject H_0 at the 5 per cent level.
2. $t = -2.40$. No change, but much closer to rejecting at the 1 per cent level.
3. $t = -4.00$. No change. Reject at the 1 per cent level (and 0.1 per cent level).
4. $t = 2.20$. Here there is a problem because the coefficient has the unexpected sign. In principle we should stick to our guns and fail to reject H_0. However, we should consider two further possibilities. One is that the justification for a one-sided test is incorrect (not very likely in this case). The other is that the model is misspecified in some way and the misspecification is responsible for the unexpected sign. For example, the coefficient might be distorted by omitted variable bias, to be discussed in Chapter 6.

R.37 A random variable X has population mean μ_X and population variance σ_X^2. A sample of n observations $\{X_1, \ldots, X_n\}$ is generated. Using the plim rules, demonstrate that, subject to a certain condition that should be stated:

$$\text{plim}\left(\frac{1}{\overline{X}}\right) = \frac{1}{\mu_X}.$$

Answer:

plim $\overline{X} = \mu_X$ by the weak law of large numbers. Provided that $\mu_X \neq 0$, we are entitled to use the plim quotient rule, so:

$$\text{plim}\left(\frac{1}{\overline{X}}\right) = \frac{\text{plim } 1}{\text{plim } \overline{X}} = \frac{1}{\mu_X}.$$

R.39 A random variable X has unknown population mean μ_X and population variance σ_X^2. A sample of n observations $\{X_1, \ldots, X_n\}$ is generated. Show that:

$$Z = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \cdots + \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}}X_n$$

is an unbiased estimator of μ_X. Show that the variance of Z does not tend to zero as n tends to infinity and that therefore Z is an inconsistent estimator, despite being unbiased.

Answer:

The weights sum to unity, so the estimator is unbiased. However, its variance is:

$$\sigma_Z^2 = \left(\frac{1}{4} + \frac{1}{16} + \cdots + \frac{1}{4^{n-1}} + \frac{1}{4^{n-1}}\right)\sigma_X^2.$$

This tends to $\sigma_X^2/3$ as n becomes large, not zero, so the estimator is inconsistent.

Note: the sum of a geometric progression is given by:

$$1 + a + a^2 + \cdots + a^n = \frac{1 - a^{n+1}}{1 - a}.$$

Hence:

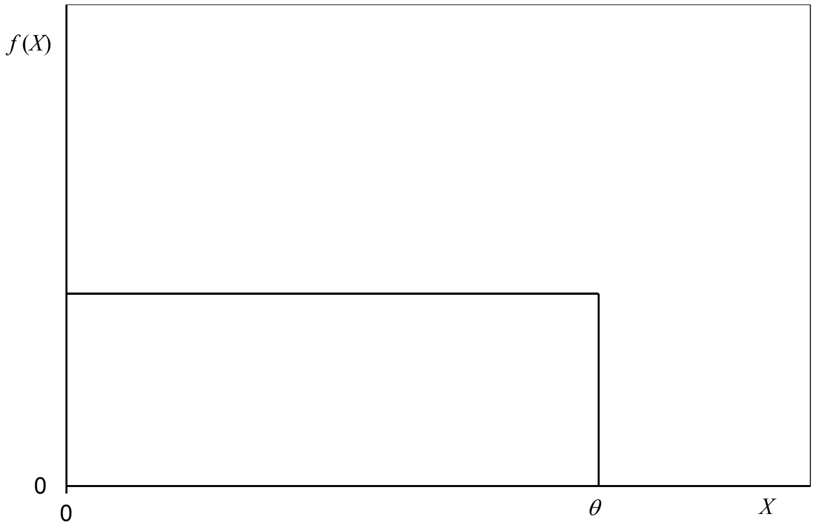
$$\begin{aligned}
\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}} &= \frac{1}{2}\left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{n-2}}\right) + \frac{1}{2^{n-1}} \\
&= \frac{1}{2} \times \frac{1 - \left(\frac{1}{2}\right)^{n-1}}{1 - \frac{1}{2}} + \frac{1}{2^{n-1}} \\
&= 1 - \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}} = 1
\end{aligned}$$

and:

$$\begin{aligned}
\frac{1}{4} + \frac{1}{16} + \cdots + \frac{1}{4^{n-1}} + \frac{1}{4^{n-1}} &= \frac{1}{4}\left(1 + \frac{1}{4} + \cdots + \frac{1}{4^{n-2}}\right) + \frac{1}{4^{n-1}} \\
&= \frac{1}{4} \times \frac{1 - \left(\frac{1}{4}\right)^{n-1}}{1 - \frac{1}{4}} + \frac{1}{4^{n-1}} \\
&= \frac{1}{3}\left(1 - \left(\frac{1}{4}\right)^{n-1}\right) + \frac{1}{4^{n-1}} \to \frac{1}{3}
\end{aligned}$$

as n becomes large.

R.41 A random variable X has a continuous uniform distribution over the interval from 0 to θ, where θ is an unknown parameter.



The following three estimators are used to estimate θ, given a sample of n observations on X:

(a) twice the sample mean

(b) the largest value of X in the sample

(c) the sum of the largest and smallest values of X in the sample.

Explain verbally whether or not each estimator is (1) unbiased, and (2) consistent.

Answer:

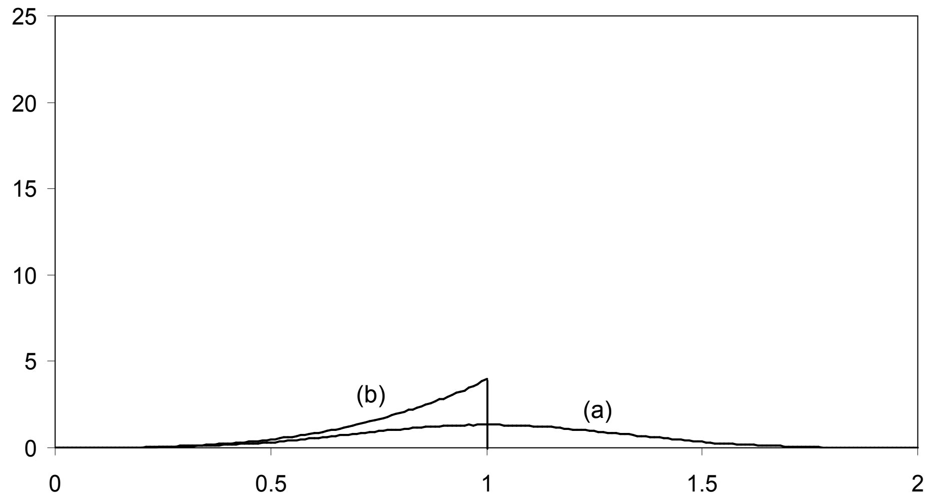
(a) It is evident that $E(X) = E(\overline{X}) = \theta/2$. Hence $2\overline{X}$ is an unbiased estimator of θ. The variance of $\overline{X}$ is σ_X^2/n. The variance of $2\overline{X}$ is therefore $4\sigma_X^2/n$. This will tend to zero as n tends to infinity. Thus the distribution of $2\overline{X}$ will collapse to a spike at θ and the estimator is consistent.

(b) The estimator will be biased downwards since the highest value of X in the sample will always be less than θ. However, as n increases, the distribution of the estimator will be increasingly concentrated in a narrow range just below θ. To put it formally, the probability of the highest value being more than ϵ below θ will be $\left(1 - \frac{\epsilon}{\theta}\right)^n$ and this will tend to zero, no matter how small ϵ is, as n tends to infinity. The estimator is therefore consistent. It can in fact be shown that the expected value of the estimator is $\frac{n}{n+1}\theta$ and this tends to θ as n becomes large.

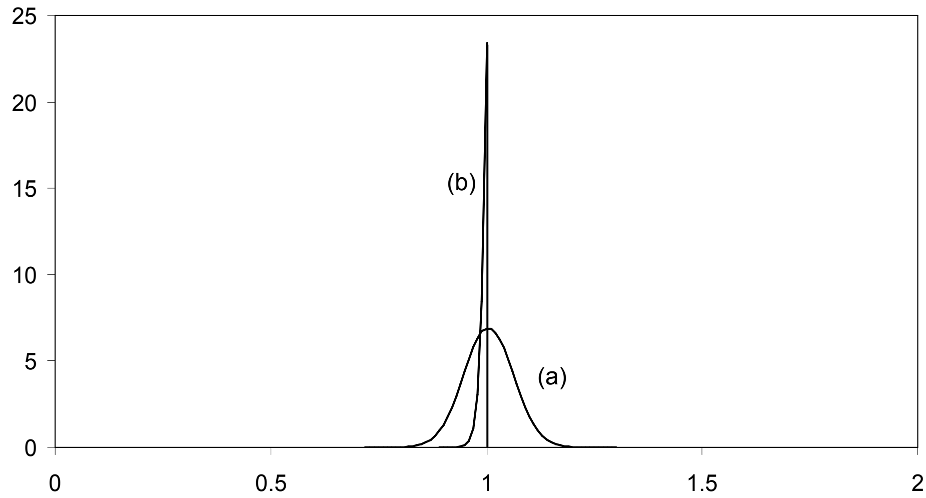
(c) The estimator will be unbiased. Call the maximum value of X in the sample X_{max} and the minimum value X_{min}. Given the symmetry of the distribution of X, the distributions of X_{max} and X_{min} will be identical, except that that of X_{min} will be to the right of 0 and that of X_{max} will be to the left of θ. Hence, for any n, $E(X_{\text{min}}) - 0 = \theta - E(X_{\text{max}})$ and the expected value of their sum is equal to θ. The estimator will be consistent for the same reason as explained in (b).

The first figure shows the distributions of the estimators (a) and (b) for 1,000,000 samples with only four observations in each sample, with $\theta = 1$. The second figure shows the distributions when the number of observations in each sample is equal to

100. The table gives the means and variances of the distributions as computed from the results of the simulations. If the mean square error is used to compare the estimators, which should be preferred for sample size 4? For sample size 100?



Sample size = 4



Sample size = 100

	Sample size 4		Sample size 100	
	(a)	(b)	(a)	(b)
Mean	1.0000	0.8001	1.0000	0.9901
Variance	0.0833	0.0267	0.0033	0.0001
Estimated bias	0.0000	−0.1999	0.0000	−0.0099
Estimated mean square error	0.0833	0.0667	0.0033	0.0002

It can be shown (Larsen and Marx, *An Introduction to Mathematical Statistics and Its Applications*, p.382, that estimator (b) is biased downwards by an amount $\theta/(n+1)$ and that its variance is:

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

while estimator (a) has variance $\theta^2/3n$. How large does n have to be for (b) to be preferred to (a) using the mean square error criterion?

The crushing superiority of (b) over (a) may come as a surprise, so accustomed are we to finding that the sample mean in the best estimator of a parameter. The underlying reason in this case is that we are estimating a boundary parameter, which, as its name implies, defines the limit of a distribution. In such a case the optimal properties of the sample mean are no longer guaranteed and it may be eclipsed by a score statistic such as the largest observation in the sample. Note that the standard deviation of the sample mean is inversely proportional to $\sqrt{n}$, while that of (b) is inversely proportional to n (disregarding the differences between n, $n+1$, and $n+2$). (b) therefore approaches its limiting (asymptotically unbiased) value much faster than (a) and is said to be superconsistent. We will encounter superconsistent estimators again when we come to cointegration in Chapter 13. Note that if we multiply (b) by $(n+1)/n$, it is unbiased for finite samples as well as superconsistent.

0.16 Answers to the additional exercises

AR.1 The total area under the function over the interval [0, 2] must be equal to 1. Since the length of the rectangle is 2, its height must be 0.5. Hence $f(X) = 0.5$ for $0 \le X \le 2$, and $f(X) = 0$ for $X < 0$ and $X > 2$.

AR.2 Obviously, since the distribution is uniform, the expected value of X is 1. However we will derive this formally.

$$E(X) = \int_0^2 X f(X)\,\mathrm{d}X = \int_0^2 0.5X\,\mathrm{d}X = \left[\frac{X^2}{4}\right]_0^2 = \left[\frac{2^2}{4}\right] - \left[\frac{0^2}{4}\right] = 1.$$

AR.3 The expected value of X^2 is given by:

$$E(X^2) = \int_0^2 X^2 f(X)\,\mathrm{d}X = \int_0^2 0.5X^2\,\mathrm{d}X = \left[\frac{X^3}{6}\right]_0^2 = \left[\frac{2^3}{6}\right] - \left[\frac{0^3}{6}\right] = 1.3333.$$

AR.4 The variance of X is given by:

$$\begin{aligned} E\left([X-\mu_X]^2\right) = \int_0^2 [X-\mu_X]^2 f(X)\,\mathrm{d}X &= \int_0^2 0.5[X-1]^2\,\mathrm{d}X \\ &= \int_0^2 (0.5X^2 - X + 0.5)\,\mathrm{d}X \\ &= \left[\frac{X^3}{6} - \frac{X^2}{2} + \frac{X}{2}\right]_0^2 \\ &= \left[\frac{8}{6} - 2 + 1\right] - [0] = 0.3333. \end{aligned}$$

The standard deviation is equal to the square root, 0.5774.

AR.5 From Exercise AR.3, $E(X^2) = 1.3333$. From Exercise AR.2, the square of $E(X)$ is 1. Hence the variance is 0.3333, as in Exercise AR.4.

AR.6 Table R.6 is reproduced for reference:

Table R.6 Trade-off between Type I and Type II errors, one-sided and two-sided tests

	Probability of Type II error if $\mu = \mu_1$	
	One-sided test	Two-sided test
5 per cent significance test	0.09	0.15
2.5 per cent significance test	0.15	(not investigated)
1 per cent significance test	0.25	0.34

Note: The distance between μ_1 and μ_0 in this example was 3 standard deviations.

Two-sided tests

Under the (false) $H_0 : \mu = \mu_0$, the right rejection region for a two-sided 5 per cent significance test starts 1.96 standard deviations above μ_0, which is 0.04 standard deviations below μ_1. A Type II error therefore occurs if $\bar{X}$ is more than 0.04 standard deviations to the left of μ_1. Under $H_1 : \mu = \mu_1$, the probability is 0.48.

Under H_0, the right rejection region for a two-sided 1 per cent significance test starts 2.58 standard deviations above μ_0, which is 0.58 standard deviations *above* μ_1. A Type II error therefore occurs if $\bar{X}$ is less than 0.58 standard deviations to the right of μ_1. Under $H_1 : \mu = \mu_1$, the probability is 0.72.

One-sided tests

Under $H_0 : \mu = \mu_0$, the right rejection region for a one-sided 5 per cent significance test starts 1.65 standard deviations above μ_0, which is 0.35 standard deviations below μ_1. A Type II error therefore occurs if $\bar{X}$ is more than 0.35 standard deviations to the left of μ_1. Under $H_1 : \mu = \mu_1$, the probability is 0.36.

Under H_0, the right rejection region for a one-sided 1 per cent significance test starts 2.33 standard deviations above μ_0, which is 0.33 standard deviations *above* μ_1. A Type II error therefore occurs if $\bar{X}$ is less than 0.33 standard deviations to the right of μ_1. Under $H_1 : \mu = \mu_1$, the probability is 0.63.

Hence the table is:

Trade-off between Type I and Type II errors, one-sided and two-sided tests

	Probability of Type II error if $\mu = \mu_1$	
	One-sided test	Two-sided test
5 per cent significance test	0.36	0.48
1 per cent significance test	0.63	0.72

AR.7 We will assume for sake of argument that the investigator is performing a 5 per cent significance test, but the conclusions apply to all significance levels.

If the true value is 0, the null hypothesis is true. The risk of a Type I error is, by construction, 5 per cent for both one-sided and two-sided tests. Issues relating to Type II error do not arise because the null hypothesis is true.

If the true value is positive, the investigator is lucky and makes the gain associated with a one-sided test. Namely, the power of the test is uniformly higher than that for a two-sided test for all positive values of μ. The power functions for one-sided and two-sided tests are shown in the first figure below.

If the true value is negative, the power functions are as shown in the second figure. That for the two-sided test is the same as that in the first figure, but reflected horizontally. The larger (negatively) is the true value of μ, the greater will be the probability of rejecting H_0 and the power approaches 1 asymptotically. However, with a one-sided test, the power function will decrease from its already very low value. The power is not automatically zero for true values that are negative because even for these it is possible that a sample might have a mean that lies in the right tail of the distribution under the null hypothesis. But the probability rapidly falls to zero as the (negative) size of μ grows.

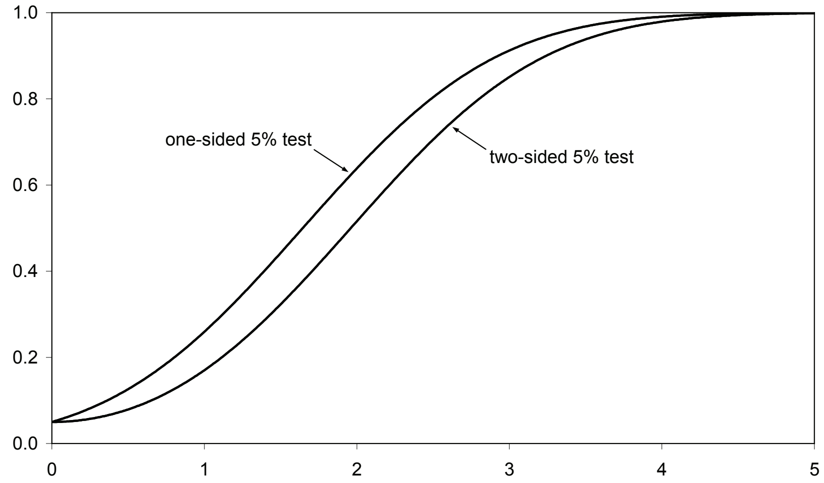


Figure 3: Power functions of one-sided and two-sided 5 per cent tests (true value > 0).

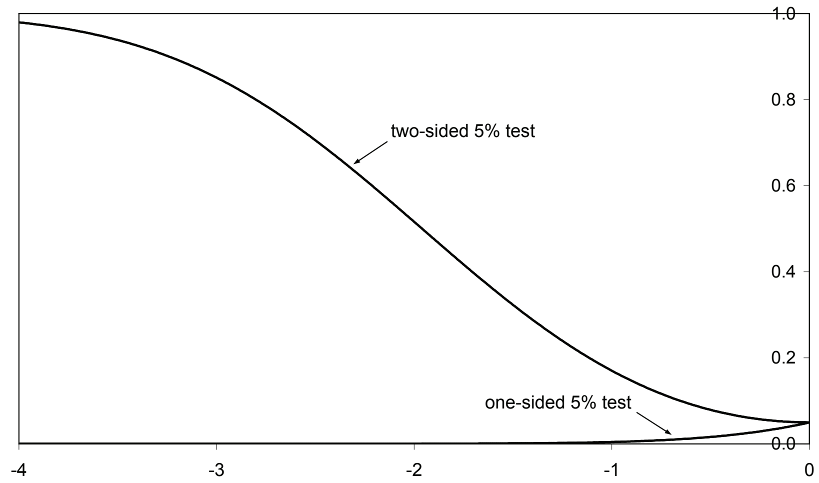


Figure 4: Power functions of one-sided and two-sided 5 per cent tests (true value < 0).

AR.8 We will refute the unbiasedness proposition by considering the more general case where Z^2 is an unbiased estimator of θ^2. We know that:

$$E\left[(Z-\theta)^2\right] = E(Z^2) - 2\theta E(Z) + \theta^2 = 2\theta^2 - 2\theta E(Z).$$

Hence:

$$E(Z) = \theta - \frac{1}{2\theta} E\left[(Z-\theta)^2\right].$$

Z is therefore a biased estimator of θ except for the special case where Z is equal to θ for all samples, that is, in the trivial case where there is no sampling error.

Nevertheless, since a function of a consistent estimator will, under quite general conditions, be a consistent estimator of the function of the parameter, $\sqrt{\widehat{\sigma}^2}$ will be a consistent estimator of σ.