Estimating Population Parameters

Prerequisites

(1) You should be familiar with the concept of a statistic and the distinction between a sample and a population. You should understand the following diagram

$$\text{population } (X) \longrightarrow \text{observation } (X_i) \longrightarrow \text{sample } (Y) \longrightarrow \text{statistic } (Z)$$

You should grasp the ideas behind the following

A *statistic* is drawn from a *sample*.
A *parameter* is a property of a *population*.
Statistics are used to estimate *parameters*.

Example (1)

The diameter of gaskets manufactured by a company are distributed with mean μ and standard deviation σ. A sample of 20 gaskets was taken and their diameters measured. The sample is represented by the random variables $X_1, X_2, ..., X_{20}$. Which of the following are statistics?

(*a*) $Z = X_1 + X_2 + ... + X_{20}$

(*b*) $\bar{X} = \frac{X_1 + X_2 + ... + X_{20}}{20}$

(*c*) $X_1 + X_2 + ... + X_{20} - 20\mu$

(*d*) $\sigma^2 (X_1 + X_2 + ... + X_{20})$

Solution

The rule here is that a statistic must be a function of the observations only and cannot involve population parameters. Hence (*a*) and (*b*) are statistics; (*c*) and (*d*), which involve the population parameters μ and σ are not.

Example (2)

Several well-known probability distributions are described in terms of standard parameters. State the parameters and their interpretation for each of the following probability distributions.

(*a*) Binomial

(*b*) Geometric

(*c*) Poisson

(*d*) Normal

(*c*) Uniform continuous

Solution

(*a*) The binomial distribution is $X \sim B(n,p)$ where *n* is the number of trials and *p* is the probability of success in any one trial.

(*b*) The geometric distribution is $X \sim Geo(p)$ where *p* is the probability of a success.

(*c*) The Poisson distribution is $X \sim Po(\lambda)$ where λ is the mean (and variance) of the distribution.

(*d*) The normal distribution is $X \sim N(\mu,\sigma^2)$ where μ is the mean and σ^2 is the variance.

(*e*) The uniform (or rectangular) continuous distribution is $X \sim R(a,b)$ where *a* and *b* are the end-points of the interval over which *X* is uniform.

Example (3)

Define the terms *population* and *statistic*.

Solution

A population is a set of all possible observations of a certain phenomenon. The phenomenon may be based on an event or an object, and collectively these are often referred to as individuals. With each individual there is associated a property, and the association of this property with the individual is the phenomenon that is being observed. A sample is a subset of a population.

In this definition the terms *population* and *individual* may refer to actual people but they may refer also refer to inanimate objects, events or phenomena of any kind. Examples of populations are (1) The height (property) of all men living on a certain island (individuals); (2) The position given relative to a set of coordinates (property) of all electrons (individuals) bound to a uranium atom; (3) The poker value (property) of a hand of five cards dealt at random from a pack of 52 cards (individuals).

(2) You should be familiar with the algebra of expectation and variance.

Expectation algebra

If *X* and *Y* are independent random variables and *a* and *b* are scalars, then

$$E(aX + bY) = aE(X) + bE(Y)$$

$$E(aX - bY) = aE(X) - bE(Y)$$

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y)$$

$$\text{var}(aX - bY) = a^2 \text{var}(X) + b^2 \text{var}(Y)$$

Example (4)

X is a random variable with mean μ and variance σ^2. Independently, Y is a random variable with mean 3μ and variance $2\sigma^2$. Find the expectation and variance of the random variable $U = 2X - 3Y$.

Solution

$$\begin{aligned} E(U) &= E(2X - 3Y) \\ &= 2E(X) - 3E(Y) \\ &= 2\mu - 3 \times 3\mu \\ &= -7\mu \end{aligned}$$

$$\begin{aligned} \text{var}(U) &= \text{var}(2X - 3Y) \\ &= \text{var}(2X) + \text{var}(3Y) \\ &= 2^2 \text{var}(X) + 3^2 \text{var}(Y) \\ &= 4\sigma^2 + 9 \times 2\sigma^2 \\ &= 22\sigma^2 \end{aligned}$$

Example (5)

Let *2X* represent twice (i.e. a scalar multiple) the random variable *X*. Let $X + X$ represent the sum of two independent observations of the random variable *X*. Determine in terms of $\text{var}(X)$ the variance of

(*a*) $2X$

(*b*) $X + X$

Explain why the expression $X + X$ is ambiguous and should be avoided if possible. Propose an alternative notation.

Solution

(a) This the scalar multiple by 2 of a single observation of the random variable X.

$$\text{var}(2X) = 2^2\,\text{var}(X) = 4\,\text{var}(X).$$

(b) This is the sum of two independent observations of the random variable X.

$$\text{var}(X+X) = \text{var}(X) + \text{var}(X) = 2\,\text{var}(X)$$

The expression $X + X$ is ambiguous because it can be easily confused with the expression $2X$, which stands for a totally different concept. It is best to distinguish two independent observations of a single population by using subscripts, and it is better to write $X_1 + X_2$. Then we can see clearly that we cannot add the two variables to obtain $2X$.

(3) You should be familiar with sequences of numbers and aware that such sequences may or may not converge to a limit.

A sequence is any string of numbers in a given order. It is usual to denote successive members of the sequence by letters with numerical subscripts.

$u_0,\ u_1,\ u_2,\ u_3, \ldots, u_n, \ldots$

If the sequence tends towards a single value, then it is said to converge, or to be convergent. Rules for proving that a given sequence converges is a whole topic in itself. Here we will permit informal arguments, and allow that if a sequence is obviously converging then it is converging. Nothing will be lost by this loose approach since questions shall not be set directly on this concept. It is a background concept.

Example (6)

Write down the first four terms and the nth term of the sequence whose rth term is $\frac{r}{2r+1}$. Assume that this sequence converges. To what limit does this sequence converge?

Solution

$$u_r = \frac{r}{2r+1}$$

$$u_1 = \frac{1}{2\times 1+1} = \frac{1}{3} \qquad u_2 = \frac{2}{2\times 2+1} = \frac{2}{5}$$

$$u_3 = \frac{3}{2\times 3+1} = \frac{3}{7} \qquad u_4 = \frac{4}{2\times 4+1} = \frac{4}{9}$$

$$u_n = \frac{n}{2n+1}$$

When n is very large this gets closer and closer to $\frac{n}{2n} = \frac{1}{2}$. It converges to $\frac{1}{2}$.

In this chapter we shall be assembling elements from each of these background topics to deal with the process of estimating population parameters by means of statistics drawn from samples.

Example (7)

A fair cubical die may be assumed to follow a discrete uniform distribution with the probability of each number coming up equal to $\frac{1}{6}$. Let X stand for the random variable that is the result of throwing one die.

(*a*) Find the expectation of X.

(*b*) A sample of 10 die throws was taken and the results recorded as follows.

$X_1 = 5 \quad X_2 = 4 \quad X_3 = 1 \quad X_4 = 3 \quad X_5 = 2$
$X_6 = 6 \quad X_7 = 1 \quad X_8 = 4 \quad X_9 = 5 \quad X_{10} = 4$

The mean of the sample (a statistic) is calculated at each successive throw of the die. Determine at each stage the statistical mean of the sample and thus generate a sequence of 10 numbers.

(*c*) Form a conjecture as follow: (1) conjecture whether sequence is convergent or divergent; (2) if it is convergent state the limit to which you think it is converging, otherwise explain why you think it is not convergent.

Solution

(*a*) $E(X) = \sum x\, P(X = x) = \frac{1+2+3+4+5+6}{6} = 3.5$

(*b*) The statistical means are computed as follows

$\bar{X}_1 = \frac{5}{1} = 5 \qquad \bar{X}_2 = \frac{5+4}{2} = 4.5$

$\bar{X}_3 = \frac{5+4+1}{3} = 3.33 \qquad \bar{X}_4 = \frac{5+4+1+3}{4} = 3.25$

$\bar{X}_5 = \frac{5+4+1+3+2}{5} = 3 \qquad \bar{X}_6 = \frac{5+4+1+3+2+6}{6} = 3.5$

$\bar{X}_7 = \frac{5+4+1+3+2+6+1}{7} = 3.14 \qquad \bar{X}_8 = \frac{5+4+1+3+2+6+1+4}{8} = 3.25$

$\bar{X}_9 = \frac{5+4+1+3+2+6+1+4+5}{9} = 3.44 \qquad \bar{X}_{10} = \frac{5+4+1+3+2+6+1+4+5+4}{10} = 3.5$

(*c*) Whilst the sequence is oscillating it appears to be convergent. We expect to converge on the theoretical expectation of the population, that is, on 3.5.

We will in fact prove below that this conjecture is true. Here the mean of the sample, given by

$$\bar{X} = \frac{\sum \bar{X}_i}{n} \qquad \bar{X} = \frac{\text{sum all the values}}{\text{total number of values}}$$

is a statistic. It is also a variable since it takes different values at each successive stage as the number of observations in the sample increases (or as different samples of the same size are taken). Therefore we should distinguish between the variable and the values that it takes. We denote the variable by a capital letter, $\bar{X}$ and the values that it takes by a lower case letter $\bar{x}$. Thus $\bar{X} = 3.25$ is interpreted as "the statistical sample mean takes the value 3.25". We have seen that as the sample size is increased this statistic takes different values. The values themselves form a sequence, and we have conjectured that the sequence is convergent on the true (theoretical) mean of the distribution.

In this example we relate the sample mean to an estimate of the true population mean (a parameter). The process can be generalised to the estimation of other parameters of varying probability distributions. So in general we use the term *estimator* to stand for the statistic that is the variable used to estimate a parameter and the term *estimate* to stand for the value that the statistic takes. An *unbiased estimator* is a statistic that homes in on the true value of a population parameter.

Example (8)

A die is suspected of being weighted so that it is an unfair die and hence biased. It is thrown 20 times and the sample mean of the 20 throws is 5. Conjecture whether this proves that the die is biased. Assume that the sample mean is an *unbiased estimator* of the true population mean.

Solution

The point of this (open ended) question is that in this case the true population mean is **unknown**. We suspect that it is not 3.5, which is what it would be if the die were fair. Therefore, the sample mean may be used as an *estimator* for the true population mean. After 20 throws of the die the *estimate* is 5.

The question actually asks you to evaluate a hypothesis. Hypothesis testing is not the subject of this chapter as such, and therefore we will give only an informal answer pr conjecture. As the estimate for a sample size of 20 differs "significantly" from the expected mean we conclude that the hypothesis is not true, that the die is biased and its true population mean is in the region of 5 rather than 3.5. In order to make this argument rigorous we would have to give a precise determination of the term "significantly" as we used it above.

Estimators

An estimator is a statistic (variable) used to estimate a population parameter. The actual value the estimator takes is called an estimate. It is sometimes useful to adopt a distinct symbol for an estimator. For example, suppose we wish to estimate the mean μ and variance σ^2 of a population. We shall denote estimators by the hat symbol. Thus $\hat{\mu}$ and $\hat{\sigma}^2$ stand for the estimators of μ and σ^2 respectively.

What is required of an estimator in general?

(1) Firstly, we require that the estimator converges on a limit as the sample size increases. We call this condition *consistency*.

Consistent estimator

A statistic (estimator) $\hat{T}$ is said to be consistent if $\mathrm{var}\left(\hat{T}\right) \to 0$ as $n \to \infty$ where n is the sample size. (A sample size can approach ∞ if the population is sampled with replacement.)

If the variance of the estimator gets smaller and smaller as the sample size gets larger and larger this means that the estimator converges on a unique limit. Obviously, a statistic that does not converge is a hopeless candidate for an estimator of a population parameter.

Example (9)

A population X takes discrete values greater than 0 and is sampled n times. Let $X_1, X_2, ..., X_n$ stand for the n observations. Explain why the statistic

$\hat{Z} = X_1 + X_2 + ... + X_n$

cannot be used as an estimator of any population parameter.

Solution

It is not consistent. Since each observation is a real number greater than 0, the sum of all these observations must be divergent. The variance of $\hat{Z}$

$\mathrm{var}\left(\hat{Z}\right) = \mathrm{var}\left(X_1\right) + \mathrm{var}\left(X_2\right) + ... + \mathrm{var}\left(X_n\right) = n\,\mathrm{var}\left(X\right)$

is also divergent.

(2) This shows that all estimators are statistics, but not all statistics are estimators. Yet although an estimator must converge on a number (limit), being convergent is not sufficient to define an appropriate estimator. We also require that an estimator converges on the true value of the parameter. The point is that an estimator may converge, but **not** on the true value of the parameter. An estimator that converges on the true value is said to be an *unbiased estimator*. An estimator that converges but not on the true value is said to be a *biased estimator*.

Unbiased estimator

A statistic $\hat{T}$ is said to be an unbiased estimator of a population parameter p if $E\left(\hat{T}\right) = p$.

This says that the estimator not only does converge, but also is expected to converge on the real value of the population parameter. Note that owing to chance factors it is always possible than an estimator does not *actually* converge on a true population parameter in a given practical application. All we can say is that theoretically it should converge on the parameter and state what the probability is that it in fact has not.[1]

Unbiased estimator of the population mean

The sample mean, $\overline{X}$, is an unbiased and consistent estimate of the population mean, μ.

<u>Proof</u>

Let a population X be sampled n times. Let $X_1, X_2, ..., X_n$ stand for the n observations in the sample. Then the sample mean is

$$\overline{X} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Then

$$E\left(\overline{X}\right) = E\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n}\left\{E\left(X_1\right) + E\left(X_2\right) + ... + E\left(X_n\right)\right\} = \frac{1}{n}\left\{\mu + \mu + ... + \mu\left(n \text{ times}\right)\right\} = \mu$$

This proves that it is unbiased. To prove that $\overline{X}$ is consistent we shall first show that $\text{var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$.

[1] The statement of probability is not dealt with in this chapter, but taken up in subsequent chapters when we deal with confidence intervals for population parameters.

$$\begin{aligned}\mathrm{var}\left(\bar{X}\right) &= \mathrm{var}\left(\frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)\right) \\ &= \frac{1}{n^2}\mathrm{var}\left(X_1 + X_2 + \ldots + X_n\right) \\ &= \frac{1}{n^2}\left(\mathrm{var}\left(X_1\right) + \mathrm{var}\left(X_2\right) + \ldots + \mathrm{var}\left(X_n\right)\right) \\ &= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \ldots + \sigma^2\left(n \text{ times}\right)\right) \\ &= \frac{1}{n^2} \times n\sigma^2 \\ &= \frac{\sigma^2}{n}\end{aligned}$$

Then as $n \to \infty$ then $\frac{\sigma^2}{n} \to 0$. Hence $\overline{X}$ is consistent. Note that this statement is part of the content of the central limit theorem.

The central limit theorem

If $X_1, X_2, X_3, \ldots, X_n$ is a random sample of size n from **any** distribution with mean μ and variance σ^2 then, for large n, the distribution of the sample mean $\bar{X}$ is *approximately normal* and $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ where $\bar{X} = \frac{1}{n}\left(X_1, X_2, X_3, \ldots, X_n\right)$. As $n \to \infty$ the approximation becomes better and better.

Regarding this theorem, we have shown here that $E\left(\overline{X}\right) = \mu$ and that $\mathrm{var}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. What we have not shown is that $\overline{X}$ is normally distributed.

Example (10)

The weights (in kg) of ten randomly chosen chickens from a farm were as follows.

2.3 2.8 2.6 1.9 2.5 2.4 3.1 2.5 2.6 2.0

Assume that these weights can be regarded as a random sample from a $N\left(\mu, \sigma^2\right)$ distribution. Calculate an unbiased estimate of μ.

Solution

$$\hat{\mu} = \frac{\sum x_i}{n} = \frac{2.3 + 2.8 + 2.6 + 1.9 + 2.5 + 2.4 + 3.1 + 2.5 + 2.6 + 2.0}{10} = 2.47 \text{ kg}$$

Biased and unbiased estimators of the population variance

The *sample variance* is the statistic $S^2 = \frac{\sum X^2}{n} - (\overline{X})^2$.

In fact this statistic is a consistent but **biased** estimator of the population variance σ^2. This means that whilst it does converge on a limit, the limit is not the same as the true population variance. What we shall do first is prove that it is biased. To do this we show $E(S^2) \neq \sigma^2$. First note that since $\text{var}(X) = E(X^2) - [E(X)]^2$, we have

$$\sigma^2 = E(X^2) - \mu^2$$
$$E(X^2) = \sigma^2 + \mu^2 \qquad (1)$$

From the central limit theorem we also have $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$, hence

$$E(\overline{X}^2) - E\left[(\overline{X})\right]^2 = \frac{\sigma^2}{n}$$

$$E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2 \qquad (2)$$

Then

$$\begin{aligned}
E(S^2) &= E\left(\frac{\sum X^2}{n} - (\overline{X})^2\right) \\
&= \frac{1}{n}E\left(\sum X^2\right) - E\left[(\overline{X})^2\right] \\
&= \frac{1}{n}\left(\sum E(X^2) - nE\left[(\overline{X})^2\right]\right) \\
&= \frac{1}{n}\left(\sum E(X^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) && \text{By result (2) above} \\
&= \frac{1}{n}\left(nE(X^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) && \sum E(X^2) = E(X^2) + E(X^2) + \ldots_E(X^2)\ (n \text{ times}) \\
&= \frac{1}{n}\left(n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) && \text{By result (1) above} \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

So S^2 does not converge on the true value of the population variance σ^2. (We will assume that it does converge). It would, therefore, be a mistake to adopt $\hat{\sigma}^2 = S^2$.

Unbiased estimator of the population variance

$s^2 = \frac{n}{n-1}S^2$ is an unbiased estimator of the population variance.

This follows since $E\left(S^2\right)=\frac{n-1}{n}\sigma^2$ then $E\left(s^2\right)=\left(\frac{n}{n-1}\right)\times\left(\frac{n-1}{n}\right)\sigma^2=\sigma^2$.

Remarks

(1) Keep a clear distinction between S^2 (upper case) used to denote the sample variance and s^2 (lower case) used to denote the unbiased estimator of the population variance, both being related by the equation $s^2=\frac{n}{n-1}S^2$. Most calculators carry functions for both of these statistics.

(2) It is again assumed that the sample is taken with replacement. This means that an observation does not remove an object from the population. This is why the sample size is potentially infinite. The population may contain a finite number of items or events, but as these are never reduced, the sample may be as large as one likes and can always be made larger.

(3) To prove that s^2 is a consistent estimator we should show that $\operatorname{var}\left(s^2\right)\to 0$ as $n\to\infty$. It can in fact be proven that $\operatorname{var}\left(s^2\right)=\frac{(n-1)E\left[(X-\mu)^4\right]-(n-3)\sigma^4}{n(n-1)}$. We omit this proof. From this it follows that $\operatorname{var}\left(s^2\right)\to 0$ as $n\to\infty$.

(4) Observe also that $\frac{n}{n-1}\to 1$ as $n\to\infty$. Hence $S^2\to s^2$ as $n\to\infty$. So in fact the practical difference between the biased and unbiased estimate is only significant when the sample size is small. It is important theory to recognise that not all estimators are unbiased.

Questions involving estimation are straightforward. The main point is that when estimating variance it is the unbiased sample variance s^2 that is used, not the biased one S^2.

Example (11)

Calculate the unbiased estimate for the population variance for the chickens sampled in example (10).

Solution

$$E\left(X^2\right)=\frac{\sum(x_i)^2}{n}$$
$$=\frac{(2.3)^2+(2.8)^2+(2.6)^2+(1.9)^2+(2.5)^2+(2.4)^2+(3.1)^2+(2.5)^2+(2.6)^2+(2.0)^2}{10}$$
$$=6.213$$

$$S^2 = \text{var}(X) = E(X^2) - [E(X)]^2 = 6.213 - (2.47)^2 = 0.1121$$
$$s^2 = \frac{n}{n-1} S^2 = \frac{10}{9} \times 0.1121 = 0.125 \ \ (3 \text{ s.f.})$$

Example (12)

The waiting time for a certain bus running during weekends was sampled on 50 weekdays. The waiting time, *t*, for the sample was found to have

$\sum t = 1631 \qquad \sum t^2 = 58{,}127$

Calculate an unbiased estimate for the variance of the waiting time of the bus. Give your answer to 3 significant figures.

Solution

$n = 50 \qquad \sum t = 1631 \qquad \sum t^2 = 58{,}127$

$$\bar{t} = \frac{\sum t}{n} = \frac{1631}{50} = 32.62$$
$$S^2 = \frac{\sum t^2}{n} - (\bar{t})^2 = \frac{58127}{50} - (32.62)^2 = 98.4756$$
$$\begin{aligned} s^2 &= \frac{n}{n-1} S^2 \\ &= \frac{50}{49} \times 98.4756 \\ &= 100.435 \\ &= 100 \ \ (3 \text{ s.f.}) \end{aligned}$$

Estimating population parameters from linear combinations of samples

We have observed above that the unbiased estimator of the population mean is the sample mean

$$\bar{X} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Note that the sample mean is itself a linear sum of the n observations in the sample, followed by a scaling with scale factor $\frac{1}{n}$.

$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ Scaling by $\frac{1}{n} \times$ Sum of X_1, $X_2, \ldots, X_n$ observations.

This raises the possibility of estimators that arise from other linear combinations together with scalings and transformations of independent random variables.

Example (13)

In example (4) we defined X to be a random variable with mean μ and variance σ^2; and independently, Y to be a random variable with mean 3μ and variance $2\sigma^2$. Let U be the random variable $U = a(2X - 3Y)$ where a is a constant.

(*a*) Find the value of a that makes U an unbiased estimator for μ.

(*b*) Find the variance of this unbiased estimator.

Solution

(*a*) In the solution to example (4) we showed that

$$E(2X - 3Y) = 2E(X) - 3E(Y) = 2\mu - 3 \times 3\mu = -7\mu$$

We require $E(U) = \mu$. Hence

$$E\left[a(2X - 3Y)\right] = \mu$$
$$-7a \times \mu = \mu$$
$$a = -\frac{1}{7}$$

(*b*) We employ the rule $\operatorname{var}(aX \pm bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y)$. Hence

$$\begin{aligned}
\operatorname{var}(U) &= \operatorname{var}\left(\frac{1}{7}(2X - 3Y)\right) \\
&= \frac{1}{7^2}\left(\operatorname{var}(2X) + \operatorname{var}(3Y)\right) \\
&= \frac{1}{49}\left(2^2 \operatorname{var}(X) + 3^2 \operatorname{var}(Y)\right) \\
&= \frac{1}{49}\left(4\sigma^2 + 9 \times 2\sigma^2\right) \\
&= \frac{22}{49}\sigma^2
\end{aligned}$$

Most efficient estimator

In the last section we saw that a linear combination of two random variables can be an estimator for a population parameter. This means that a given population parameter might have several different estimators. The question therefore arises, which of these estimators is the best? The term "best" here is vague and requires clarification. Best in the sense of "most efficient" means in this context the estimator that converges quickest on the true population parameter as the sample size n increases. The speed of convergence is linked to the variance of the estimator. The smaller the variance the faster the estimator converges. Therefore, the best estimator, in the sense of "most efficient", is the estimator with the smallest variance.

Most efficient estimator

The most efficient estimator of a population parameter p is the estimator with the smallest variance.

Example (14)

Let X be a random variable with mean μ and variance σ^2; and independently, Y be a random variable with mean 3μ and variance $2\sigma^2$. Let U and V be estimators for μ such that

$U = \frac{1}{4}(X+Y) \qquad V = b(2X+Y)$

(a) Show that U is an unbiased estimator for μ and find the value of b that makes V an unbiased estimator of μ.

(b) Find the variances of both these unbiased estimators and hence determine which of the two is the more efficient.

Solution

(a) $E(U) = E\left[\frac{1}{4}(X+Y)\right] = \frac{1}{4}\left[E(X)+E(Y)\right] = \frac{1}{4}(\mu+3\mu) = \mu$

$E(V) = \mu$

$b\,E(2X+Y) = \mu$

$b(2\mu+3\mu) = \mu$

$b = \frac{1}{5}$

(b) $\text{var}(U) = \text{var}\left(\frac{1}{4}(X+Y)\right) = \frac{1}{16}\left(\text{var}(X)+\text{var}(Y)\right) = \frac{1}{16}\left(\sigma^2+2\sigma^2\right) = \frac{3}{16}\sigma^2 = 0.1875\sigma^2$

$$\text{var}(V) = \text{var}\left(\frac{1}{5}(2X+Y)\right)$$

$$= \frac{1}{25}\left(\text{var}(2X)+\text{var}(Y)\right)$$

$$= \frac{1}{25}\left(4\,\text{var}(X)+\text{var}(Y)\right) = \frac{1}{25}\left(4\sigma^2+2\sigma^2\right) = \frac{6}{25}\sigma^2 = 0.24\sigma^2$$

Therefore U is the more efficient estimator.

Most efficient estimator for a population mean

The sample mean, $\overline{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(X_1+X_2+\ldots+X_n)$, is the most efficient linear estimator of the population mean, μ.

Proof

This proof is optional. The proof will be by contradiction. We have shown above that $\text{var}\left(\overline{X}\right)=\frac{\sigma^2}{n}$. Now assume that there is another linear unbiased estimator of μ that is more efficient than $\overline{X}$. Let this alternative estimator be $\hat{T}$. Then $\hat{T}$ is given by

$$\hat{T}=a_1X_1+a_2X_2+...+a_nX_n$$

for some constants $a_1, a_2,...,a_n$. Since $\hat{T}$ is unbiased we have $E\left(\hat{T}\right)=\mu$. Now

$$\begin{aligned}E\left(\hat{T}\right)&=E\left(a_1X_1+a_2X_2+...+a_nX_n\right)\\&=a_1E\left(X_1\right)+a_2E\left(X_2\right)+...+a_nE\left(X_n\right)\\&=a_1\mu+a_2\mu+...+a_n\mu\\&=\left(a_1+a_2+...+a_n\right)\mu\end{aligned}$$

Therefore, since $E\left(\hat{T}\right)=\mu$, then

$$a_1+a_2+...+a_n=1 \qquad (1)$$

The variance of $\hat{T}$ is given by

$$\begin{aligned}\text{var}\left(\hat{T}\right)&=\text{var}\left(a_1X_1+a_2X_2+...+a_nX_n\right)\\&=\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)\text{var}\left(X\right)\\&=\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)\sigma^2\end{aligned}$$

We have assumed that $\hat{T}$ is more efficient than $\overline{X}$. Hence $\text{var}\left(\hat{T}\right)<\text{var}\left(\overline{X}\right)$. This implies

$$\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)\sigma^2<\frac{\sigma^2}{n}$$

$$\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2<\frac{1}{n} \qquad (2)$$

Now

$$\left(a_1-\frac{1}{n}\right)^2+\left(a_2-\frac{1}{n}\right)^2+...+\left(a_n-\frac{1}{n}\right)^2\geq 0 \qquad \left[\text{This is a sum of squares, and squares are always >0}\right]$$

$$\left(a_1^{\,2}-2\frac{a_1}{n}+\frac{1}{n^2}\right)+\left(a_2^{\,2}-2\frac{a_2}{n}+\frac{1}{n^2}\right)+...+\left(a_2^{\,2}-2\frac{a_2}{n}+\frac{1}{n^2}\right)\geq 0$$

$$\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)-\frac{2}{n}\left(a_1+a_2+...+a_n\right)+n\left(\frac{1}{n^2}\right)\geq 0$$

$$\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)-\frac{2}{n}+\frac{1}{n}\geq 0 \qquad \left[\text{By } (1): a_1+a_2+...+a_n=1\right]$$

$$\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)-\frac{1}{n}\geq 0$$

$$\left(\left(a_1\right)^2+\left(a_2\right)^2+...+\left(a_n\right)^2\right)\geq\frac{1}{n}$$

This statement contradicts statement (2). Therefore the assumption that $\hat{T}$ is a more efficient estimator than $\overline{X}$ must be false.

We state also (but without proof) the following.

Most efficient estimator for a population variance

The unbiased estimator of a population variance given by $s^2 = \frac{n}{n-1}S^2$ where S^2 is the sample variance is also the most efficient estimator for the population variance.

Questions may be set requiring you to find in a given context the most efficient estimator for a population parameter.

Example (15)

The random variable *X* follows a Poisson distribution with parameter λ. Independently, the random variable *Y* follows a Poisson distribution with parameter 2λ. Consider the estimator

$$W = \frac{kX + Y}{k+2} \qquad k \neq -2$$

(*a*) Show that *W* is an unbiased estimator for λ for all possible values of *k*.

(*b*) Find the variance of *W* in terms of *k* and λ.

(*c*) Hence find the value of *k* that gives the best estimator of this form.

Solution

(*a*)

$$\begin{aligned} E(W) &= E\left(\frac{kX+Y}{k+2}\right) \\ &= \frac{1}{k+2}E(kX+Y) \\ &= \frac{1}{k+2}\left(kE(X)+E(Y)\right) \\ &= \frac{1}{k+2}(k\lambda+2\lambda) = \frac{k+2}{k+2}\lambda = \lambda \end{aligned}$$

(*b*)

$$\begin{aligned} \operatorname{var}(W) &= \operatorname{var}\left(\frac{kX+Y}{k+2}\right) \\ &= \frac{1}{(k+2)^2}\left(\operatorname{var}(kX)+\operatorname{var}(Y)\right) \\ &= \frac{1}{(k+2)^2}\left(k^2\operatorname{var}(X)+\operatorname{var}(Y)\right) \\ &= \frac{k^2+2}{(k+2)^2}\lambda \end{aligned}$$

We have to find the value of k that makes the function $f(k) = \frac{k^2+2}{(k+2)^2}$ a minimum.

Differentiating

$$f'(k) = \frac{d}{dk}\left\{\frac{k^2+2}{(k+2)^2}\right\}$$
$$= \frac{2k(k+2)^2 - (k^2+2)2(k+2)}{(k+2)^4}$$
$$= \frac{2k(k+2) - 2(k^2+2)}{(k+2)^3} \qquad k \neq -2$$

For turning points

$$2k(k+2) - 2(k^2+2) = 0$$
$$2k^2 + 4k - 2k^2 - 4 = 0$$
$$4k = 4$$
$$k = 1$$

We should show that this is a minimum. To do this, examine values of $f'(k)$ around $k = 1$. When $k < 1$, $f'(k) = \frac{4k-4}{(k+2)^3} < 0$. When $k > 1$, $f'(k) = \frac{4k-4}{(k+2)^3} > 0$, so this is a minimum.

Estimators of probability

Consider a population where the probability of a certain event is unknown. Let a sample be drawn from this population. We wish to derive an estimator for the unknown probability from the sample.

Example (16)

The discrete random variable X takes the values "success" and "failure". The probability of a success, which is unknown, is θ. A random sample of n observations on X is taken. N denotes the number of times the value "success" occurs in the sample.

(*a*) State the probability of a failure.

(*b*) State the probability distribution of N.

(*c*) Prove that the statistic $\hat{\theta} = \frac{N}{n}$ is an unbiased estimator for θ.

(*d*) Find the variance of $\hat{\theta}$ and deduce that $\hat{\theta}$ is an unbiased, consistent estimator for θ.

Solution

(a) $P(\text{success}) = \theta$

By the law of total probability

$P(\text{failure}) = 1 - \theta$

(b) $N \sim B(n, \theta)$

(c) $E(\hat{\theta}) = E\left(\frac{N}{n}\right) = \frac{1}{n}E(N) = \frac{1}{n} \times n\theta = \theta$ $\quad$ $\left[\text{if } X \sim B(n,p) \text{ then } E(X) = np\right]$

Hence, since $E(\hat{\theta}) = \theta$, $\hat{\theta}$ is unbiased.

(d) $\text{var}(\hat{\theta}) = \text{var}\left(\frac{N}{n}\right) = \frac{1}{n^2}\text{var}(N) = \frac{1}{n^2} \times n\theta(1-\theta) = \frac{1}{n}\theta(1-\theta)$

$\left[\text{if } X \sim B(n,p) \text{ then } \text{var}(X) = np(1-p)\right]$

$\text{var}(\hat{\theta}) = \frac{1}{n}\theta(1-\theta) \to 0 \text{ as } n \to \infty$

Hence $\hat{\theta}$ is an unbiased, consistent estimator for θ.

This example deals with the essential theory that underlies questions where an unknown probability is estimated.

Most efficient estimator for a population probability

Let X be a discrete random variable and let P be an event. Let the probability that P occurs be θ. If N is the number of times P occurs in a sample of size n, then the most efficient estimator for θ is $\hat{\theta} = \frac{N}{n}$.

Example (17)

The discrete random variable X takes the values 1, 2, 3 with probabilities $\theta, 2\theta, \lambda$ respectively, where θ is a constant $0 < \theta < \frac{1}{3}$. In order to estimate θ a random sample of n observations of X was taken.

(a) Find λ and write down the probability distribution of X in terms of θ alone. Determine $E(X)$ and $\text{var}(X)$.

(b) Let $\bar{X}$ denote the mean of the n observations in the random sample of X. Consider the estimator

$\hat{\theta}_1 = \frac{a - \bar{X}}{4}$

where a is a constant.

(*i*) Find the value of a that makes $\hat{\theta}_1$ an unbiased estimator for θ.

(*ii*) Find the variance of $\hat{\theta}_1$.

(*c*) Let N denote the number of occurrences of the value 2 in the sample. Another possible estimator is

$$\hat{\theta}_2 = \frac{N}{2n}$$

(*i*) Show that is an unbiased estimator for θ.

(*ii*) Find the variance of $\hat{\theta}_2$.

(*d*) (*i*) Show that $\text{var}(\hat{\theta}_2) - \text{var}(\hat{\theta}_1) = \frac{\theta}{8n}$.

(*ii*) State, with a reason, which is the better estimator.

Solution

(*a*) By the law of total probability $\lambda = 1 - \theta - 2\theta = 1 - 3\theta$

The probability distribution of X is

x	1	2	3
$P(X = x)$	θ	2θ	$1-3\theta$

$$E(X) = \sum x\,P(X = x)$$
$$= 1 \times \theta + 2 \times 2\theta + 3 \times (1 - 3\theta)$$
$$= 3 - 4\theta$$

$$E(X^2) = \sum x^2\,P(X = x)$$
$$= 1^2 \times \theta + 2^2 \times 2\theta + 3^2 \times (1 - 3\theta)$$
$$= 9 - 18\theta$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$
$$= 9 - 18\theta - (3 - 4\theta)^2$$
$$= 9 - 18\theta - (9 - 24\theta + 16\theta^2)$$
$$= 6\theta - 16\theta^2$$

(*b*) (*i*)

$$E(\hat{\theta}_1) = \theta$$
$$E\left(\frac{a - \bar{X}}{4}\right) = \theta$$
$$\frac{1}{4}\left(a - E(\bar{X})\right) = \theta$$
$$\frac{1}{4}\left(a - (3 - 4\theta)\right) = \theta \qquad \left[E(\bar{X}) = E(X) = \theta\right]$$
$$\frac{a}{4} - \frac{3}{4} + \theta = \theta$$
$$a = 3$$

(*ii*) $\text{var}\left(\hat{\theta}_1\right) = \text{var}\left(\frac{a - \bar{X}}{4}\right)$

$$= \frac{1}{16}\text{var}\left(a - \bar{X}\right)$$
$$= \frac{1}{16}\text{var}\left(\bar{X}\right)$$
$$= \frac{1}{16}\text{var}\left(\frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)\right)$$
$$= \frac{1}{16n^2}\left(\text{var}\left(X_1\right) + \text{var}\left(X_2\right) + \ldots + \text{var}\left(X_n\right)\right)$$
$$= \frac{1}{16n^2} \times n\,\text{var}\left(X\right)$$
$$= \frac{1}{16n}\left(6\theta - 16\theta^2\right)$$
$$= \frac{1}{8n}\theta\left(3 - 8\theta\right)$$

(*c*) $N \sim B\left(n, 2\theta\right)$

(*i*) $E\left(\hat{\theta}_2\right) = E\left(\frac{N}{2n}\right) = \frac{1}{2n}E\left(N\right) = \frac{1}{2n} \times 2n\theta = \theta$

(*ii*) $\text{var}\left(\hat{\theta}_2\right) = \text{var}\left(\frac{N}{2n}\right) = \frac{1}{4n^2}\text{var}\left(N\right) = \frac{1}{4n^2} \times n \times 2\theta \times \left(1 - 2\theta\right) = \frac{1}{2n}\theta\left(1 - 2\theta\right)$

(*d*) (*i*) $\text{var}\left(\hat{\theta}_2\right) - \text{var}\left(\hat{\theta}_1\right) = \frac{1}{2n}\theta\left(1 - 2\theta\right) - \frac{1}{8n}\theta\left(3 - 8\theta\right)$

$$= \frac{\theta}{8n}\left(4 - 8\theta - 3 + 8\theta\right)$$
$$= \frac{\theta}{8n}$$

(*i*) We have

$\text{var}\left(\hat{\theta}_2\right) - \text{var}\left(\hat{\theta}_1\right) = \frac{\theta}{8n}$

Since θ and $8n$ are positive

$\text{var}\left(\hat{\theta}_2\right) > \text{var}\left(\hat{\theta}_1\right)$

Hence $\hat{\theta}_1$ is the better estimator