For a function U to be an unbiased estimator for θ, then

$$E(U) = \theta$$

When choosing between two unbiased estimators for θ, the better one is the one with the smaller variance.

For $Y = \frac{X}{n}$, where X is the number of successes in n independent Bernoulli trials,

$$E(Y) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p \text{ and}$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

For a random sample of n independent observations from a distribution having mean μ and variance σ^2, an unbiased estimator of σ^2 (population variance) is

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

Although a more helpful and practical formula is

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$$

Distribution of a sample mean from a Normal distribution with population mean μ and variance σ^2 is

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$E(\bar{X}) = \mu,\ \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n},\ SE(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})}$$

Other useful facts to know include:

- Mean and variance of a binomial distribution are np and $np(1-p)$ respectively
- Mean and variance of a Poisson distribution are both λ
- Mean and variance of a pdf $f(x)$ are given by
  $E(X) = \int x f(x)\mathrm{d}x$ and
  $\mathrm{Var}(X) = \int x^2 f(x)\mathrm{d}x - (E(X))^2$

Example

A survey of 800 people shows that 352 people prefer dark chocolate to milk chocolate. Find unbiased estimators of the mean and standard error of the proportion of people who prefer dark chocolate to milk chocolate.

$$\hat{p} = \frac{352}{800} = 0{\cdot}44 \qquad \hat{\sigma} = \sqrt{\frac{0{\cdot}44 \times 0{\cdot}56}{800}} = 0{\cdot}0175$$

Example

A manufacturer of certain types of batteries tests 10 batteries to see how long they will last with reasonable use. The results, in hours, are shown below:

3·5, 3·7, 3·6, 3·9, 3·0,
3·5, 3·5, 3·6, 3·9, 3·2

You may assume that this is a random sample from a Normal distribution with mean μ and variance σ^2.

Calculate unbiased estimates of μ and σ^2.

$$\sum x = 35{\cdot}4 \text{ and } \sum x^2 = 126{\cdot}02$$

$$\hat{\mu} = \frac{35{\cdot}4}{10} = 3{\cdot}54,\ s^2 = \frac{1}{9}\left(126{\cdot}02 - \frac{35{\cdot}4^2}{10}\right)$$

$$= 0{\cdot}078\dot{2}$$

Example

The continuous random variable X has probability density function f given by

$$f(x) = \frac{1}{2}x^2 + \frac{1}{3}\theta x + \frac{1}{3}, \text{ for } -1 < x < 1.$$
$$f(x) = 0 \quad \text{otherwise,}$$
where $0 < \theta < 2$

(a) Find $E(X)$.

$$E(X) = \int_{-1}^{1} x\left(\frac{1}{2}x^2 + \frac{1}{3}\theta x + \frac{1}{3}\right)\mathrm{d}x = \frac{2}{9}\theta$$

(b) Express $P(X > 0)$ in terms of θ.

$$P(X > 0) = \int_0^1 \left(\frac{1}{2}x^2 + \frac{1}{3}\theta x + \frac{1}{3}\right)\mathrm{d}x = \frac{1}{2} + \frac{1}{6}\theta$$

(c) For a random sample of n observations of X, let $\bar{X}$ denote the sample mean and let Y denote the number of positive observations. Show that

$$T_1 = \frac{9}{2}\bar{X} \text{ and } T_2 = \frac{6}{n}Y - 3 \text{ are both unbiased estimators of } \theta.$$

$$E(T_1) = \frac{9}{2}E(X) = \theta \qquad Y \sim B(n,p) \text{ so } E(T_2) = \frac{6}{n}E(Y) - 3$$

$$E(T_2) = \frac{6}{n}n\left(\frac{1}{2} + \frac{1}{6}\theta\right) - 3 = \theta$$

Since $E(T_1) = \theta$ and $E(T_2) = \theta$ they are unbiased estimators of θ.

(d) Determine which of T_1 and T_2 is the better estimate for θ.

$$\mathrm{Var}(T_1) = \left(\frac{9}{2}\right)^2 \mathrm{Var}(\bar{X}) = \frac{81}{4n}\mathrm{Var}(X) \qquad E(X^2) = \int_{-1}^{1} x^2\left(\frac{1}{2}x^2 + \frac{1}{3}\theta x + \frac{1}{3}\right)\mathrm{d}x = \frac{19}{45}$$

$$\mathrm{Var}(T_1) = \frac{81}{4n}\left(\frac{19}{45} - \frac{4}{81}\theta^2\right) = \frac{171 - 20\theta^2}{20n}$$

$$\mathrm{Var}(T_2) = \frac{36}{n}p(1-p) = \frac{9 - \theta^2}{n} = \frac{180 - 20\theta^2}{20n}$$

Since $\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$ T_1 is the better estimator of θ