

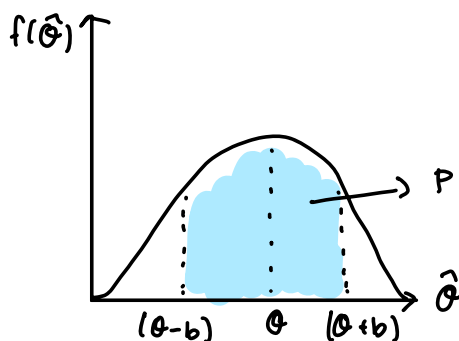
Evaluating the Goodness of a point estimation

Error of Estimation $\left\{ \begin{array}{l} \text{distance between an estimator } \hat{\theta} \\ \text{its target parameter } \theta \end{array} \right.$

$$E = |\hat{\theta} - \theta|$$

Since $\hat{\theta}$ is a random variable, E is also a random variable

Although we cannot say how large or small it will be for a particular estimate, we can make probability statements about it.



$b \rightarrow$ is a probabilistic bound on error of estimation

$$P(E < b)$$

probability that E is less than b is represented by the shaded area

$$\begin{aligned} P(|\hat{\theta} - \theta| < b) &= P[-b < (\hat{\theta} - \theta) < b] \\ &= P(\theta - b < \hat{\theta} < \theta + b) \end{aligned}$$

$P(E < b)$ provides a measure of goodness of a single estimator

$$\begin{aligned} b = 0.01 & \quad P(\theta - 0.01 < \hat{\theta} < \theta + 0.01) \\ b = 0.1 & \quad P(\theta - 0.1 < \hat{\theta} < \theta + 0.1) \end{aligned}$$

if we chose a higher value of b we increase the prob. This probability identifies fraction of times in repeated sampling that estimator $\hat{\theta}$ falls within b units of target parameter θ

Suppose we know prob density of $\hat{\theta}$ & want to find value of b such that

$$\begin{aligned} &P(E < b) = 0.95 \\ &\int_{\theta-b}^{\theta+b} f(\hat{\theta}) \cdot d\hat{\theta} = 0.95 \end{aligned}$$

However, whether we know probability density of $\hat{\theta}$ or not, if $\hat{\theta}$ is unbiased we can find an approximate bound on E by expressing b as a multiple of standard error of $\hat{\theta}$

For instance for $k \geq 1$ let $b = k \sigma_{\hat{\theta}}$ from Tchebysheff's theorem that E will be less than $k \sigma_{\hat{\theta}}$ with prob. at least $1 - 1/k^2$

for $k=2 \Rightarrow E$ will be less than $b = 2 \sigma_{\hat{\theta}}$ with prob at least 0.75

$$b = \sigma_{\hat{\theta}} \quad \text{standard error} \quad b = 2 \sigma_{\hat{\theta}}$$

Consistency

It ensures that the estimation rule will produce an estimate that is close to true parameter with high probability if sample is large enough.

Let $\hat{\theta}$ be an estimator of θ based on sample size n . Then, $\hat{\theta}$ is a consistent estimator if

$$\lim_{n \rightarrow \infty} P[|\hat{\theta} - \theta| < c] = 1$$

where c is an arbitrarily chosen small positive number. This means that prob. that value of $\hat{\theta}$ falls in interval $[\theta - c, \theta + c]$ can be made arbitrarily close to 1 given a large enough sample size no matter how small c is.

If this is true then sequence of random variables $\hat{\theta}$ is said to converge in probability to constant θ . θ is said to be prob. limit of sequence $\hat{\theta}$.

$$p\lim \hat{\theta} = \theta$$

This estimator $\hat{\theta}$ is a consistent estimator for θ if $p\lim \hat{\theta} = \theta$

In other words if $\hat{\theta}$ is unbiased estimator of parameter θ and $\text{Var}(\hat{\theta}) \rightarrow 0$ and $n \rightarrow \infty$, then $\hat{\theta}$ is a consistent estimator of θ .

Ex: Show that for a random sample from a normal population, sample variance S^2 is a consistent estimator of σ^2 (sigma squared)

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

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

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

$\text{Var}(S^2) \rightarrow 0$ when $n \rightarrow \infty$

Thus, S^2 is a consistent estimator of variance of a normal population

Convergence in Distribution

Probability distribution of an estimation as sample size becomes increasingly large

Estimators whose distribution are unknown in small samples can have a particular known distributions in large samples

↳ A well known example is $\rightarrow$ CLT

Let $x_1, x_2, \dots, x_n$ be independently & identically distributed random variables with

$$E(x_i) = \beta, \quad \text{Var}(x_i) = \sigma^2, \quad i = 1, \dots, n$$

$$\text{also let } \hat{\beta}_n = \sum_{i=1}^n x_i / n$$

$$\sqrt{n} (\hat{\beta}_n - \beta) \Rightarrow \text{converges in distribution to a } N(0, \sigma^2)$$

It has a limiting distribution or asymptotic distr
 $\sqrt{n} (\hat{\beta}_n - \beta) \sim N(0, \sigma^2)$

Therefore we can make probability statements about an estimator when exact small sample distribution of estimator cannot be determined

If a limiting distribution exists, we can use that limiting distribution as an approximation to true but

unknown distribution If sample size is sufficiently large.

Sufficiency

An estimated $\hat{\theta}$ is said to be sufficient if it utilizes all information in a sample relevant to estimation of θ

$$f(x_1, x_2, \dots, x_n | \hat{\theta}) = f(x_1, x_2, \dots, x_n) / g(\hat{\theta})$$

The statistic $\hat{\theta}$ is a sufficient estimator of parameter θ iff for each value of $\hat{\theta}$, conditional probability distribution or density of random sample $x_1, x_2, \dots, x_n$ given is independent of θ

Ex:

If $x_1, x_2, \dots, x_n$ constitute a random sample of size n from a Bernoulli Population
Show that

$$\hat{\theta} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

is a sufficient estimator of parameter θ

$$f(x_i; \theta) = \theta^{x_i} (1-\theta)^{1-x_i} \quad \text{for } x_i = 0, 1$$

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} \\ &= \theta^{\sum x_i} (1-\theta)^{n - \sum x_i} \\ &= \theta^x (1-\theta)^{n-x} \\ &= \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}, \quad x_i = 0, 1 \\ &\quad i = 1, 2, \dots, n \end{aligned}$$

Since $x = x_1 + x_2 + \dots + x_n$ is binomial random variable with parameter θ & n
Its distribution is

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Transformation of variable technique gives us

$$g(\hat{\theta}) = \binom{n}{n\hat{\theta}} \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}} \text{ for } \hat{\theta} = 0, \frac{1}{n}, \dots, 1$$

$$f(x_1, x_2, \dots, x_n | \hat{\theta}) = f(x_1, x_2, \dots, x_n) / g(\hat{\theta})$$

$$= \frac{\theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}}{\binom{n}{n\hat{\theta}} \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}} = \frac{1}{\binom{n}{n\hat{\theta}}} = \frac{1}{\binom{n}{x}} = \frac{1}{\binom{n}{x_1 + x_2 + \dots + x_n}}$$

for $x_i = 0, 1$
 $i = 1, 2, \dots, n$

This does not depend on θ , $\hat{\theta} = \frac{x}{n}$ is a sufficient estimator

Since it is tedious to check whether a statistic is a sufficient estimator of a given parameter, it can easily be done by factorization theorem.

Statistic $\hat{\theta}$ is a sufficient estimator of parameter θ iff joint prob. distr of density of random sample can be factored so that $f(x_1, x_2, \dots, x_n; \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \dots, x_n)$ where $g(\hat{\theta}, \theta)$ depend on $\hat{\theta}$ & θ
 $h(x_1, x_2, \dots, x_n)$ does not depend on θ

Ex: show that $\bar{x}$ is a sufficient estimator of mean μ of a normal population with var, σ^2

$$f(x_1, x_2, \dots, x_n; \mu) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2}$$

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n [(x_i - \bar{x}) - (\mu - \bar{x})]^2$$

$$= \sum (x_i - \bar{x})^2 + \sum (\bar{x} - \mu)^2$$

$$= \sum (x_i - \bar{x})^2 + n \cdot (\bar{x} - \mu)^2$$

$$f(x_1, x_2, \dots, x_n; \mu) = \left\{ \frac{\sqrt{n}}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2} \right\}$$

$$\frac{1}{\sqrt{n}} \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^{n-1} e^{-\frac{1}{2} \sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^2}$$

1st factor $\Rightarrow$ depends only on a estimate $\bar{x}$ & μ

2nd factor $\rightarrow$ not include μ

$\bar{x}$ is a sufficient estimator

If $\hat{\theta}$ is a sufficient estimator of θ , then any single valued function of $y = U(\hat{\theta})$ not including θ is also a sufficient estimator of θ & $U(\hat{\theta})$

$$U(\hat{\theta})$$

plus, provided $y = U(\hat{\theta})$ can be solved to give single valued inverse $\hat{\theta} = w(y)$. This follows from factorization theorem since we can write

$$\{ (x_1, x_2, \dots, x_n; \theta) = g[w(y), \theta] h(x_1, x_2, \dots, x_n)$$

where $g[w(y), \theta] \rightarrow$ depends on y & θ

$h(\dots) \rightarrow$ depends on sampling information

Ex: old exam question

If $\hat{\theta}_1$ & $\hat{\theta}_2$ are unbiased estimators of some parameter θ , what condition must be imposed on the constant k_1 & k_2 so that

$$k_1 \hat{\theta}_1 + k_2 \hat{\theta}_2$$

is also an unbiased estimator?

$$E[k_1 \hat{\theta}_1 + k_2 \hat{\theta}_2] = k_1 \theta + k_2 \theta = (k_1 + k_2) \theta$$

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$$

$$E(\hat{\theta}) = \theta$$

$$k_1 + k_2 = 1$$

Ex: If $x_1, x_2, \dots, x_n$ constitute a random sample from a normal population with $\mu = 0$, show that

$\sum_{i=1}^n \frac{x_i^2}{n}$ is an unbiased estimator of σ^2

$$E\left[\sum_{i=1}^n \frac{x_i^2}{n}\right] = \frac{1}{n} \sum E(x_i^2) = \frac{1}{n} \sum (\sigma^2 + \mu^2) = \frac{1}{n} \sum \sigma^2 = \sigma^2$$

Ex: Show that sample proportion $\frac{x}{n}$ is a minimum variance unbiased estimator of binomial parameter θ

Hint: x/n as mean of a random sample of size n from a bernoulli population with parameter θ

$$f(x, \theta) = \theta^x (1-\theta)^{1-x}$$

$$E(x) = \theta, \quad E(x^2) = \theta$$

$$\ln f(x, \theta) = x \ln \theta + (1-x) \ln(1-\theta)$$

$$\frac{\partial \ln f(x, \theta)}{\partial \theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

$$E\left[\frac{\partial \ln f(x, \theta)}{\partial \theta}\right] = \frac{1}{\theta^2(1-\theta)^2} E(x-\theta)^2 = \frac{1}{\theta(1-\theta)}$$

$$\frac{1}{n E\left[\frac{\partial \ln f(x, \theta)}{\partial \theta}\right]} = \frac{\theta(1-\theta)}{n} = \text{Var}\left(\frac{x}{n}\right)$$

$$E\left(\frac{x}{n}\right) = \frac{n\theta}{n} = \theta \Rightarrow \text{unbiased}$$

$\frac{x}{n}$ is a minimum variance unbiased estimation

Ex: Show that mean of a random sample of size n is a minimum variance unbiased estimator of parameter λ of a Poisson population.

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \mu = \lambda \quad \sigma^2 = \lambda$$

$$\text{Var}(\bar{x}) = \lambda/n$$

$$\ln f(x; \lambda) = x \ln \lambda - \lambda - \ln x!$$

$$\frac{\partial \ln f(x; \lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1$$

$$E\left[\frac{\partial \ln f(x; \lambda)}{\partial \lambda}\right] = \frac{E(x^2)}{\lambda^2} = \frac{2}{\lambda} E(x) + 1$$

$$= \frac{\lambda + \lambda^2}{\lambda^2} - \frac{2}{\lambda} \cdot \lambda + 1 = \frac{1}{\lambda}$$

$$\frac{1}{nE[\dots]} = \frac{1}{n} = \text{Var}(\bar{x})$$

$\bar{x} \Rightarrow$ minimum variance unbiased estimator

Ex: Show that $\frac{x+1}{n+2}$ is a consistent estimator of binomial parameter θ

$$E\left(\frac{x+1}{n+2}\right) = \frac{1}{n+2} n\theta + \frac{1}{n+2} = \frac{n}{n+2} \theta + \frac{1}{n+2}$$

$\Rightarrow \theta$ as $n \rightarrow \infty$

It is asymptotically unbiased

$$\text{Var}\left(\frac{x+1}{n+2}\right) = \frac{1}{(n+2)^2} \quad \text{Var}(x) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n} \xrightarrow[n \rightarrow \infty]{\text{as}} 0$$

consistent estimator

Ex: If x_1 & x_2 are independent random variables having binomial distributions with parameters θ & n_1 , θ & n_2 . Show that $\frac{x_1+x_2}{n_1+n_2}$ is a sufficient estimator of θ

$$f(x_1, x_2) = \binom{n_1}{x_1} \binom{n_2}{x_2} \theta^{x_1+x_2} (1-\theta)^{(n_1+n_2)-(x_1+x_2)}$$

$$\hat{\theta} = \frac{x_1+x_2}{n_1+n_2} \Rightarrow x_1+x_2 = \hat{\theta} (n_1+n_2)$$

By factorization theorem

$$= \underbrace{\binom{n_1}{x_1} \binom{n_2}{x_2}}_{h(x_1, x_2)} \underbrace{\theta^{(n_1+n_2)\hat{\theta}} (1-\theta)^{(n_1+n_2)(1-\hat{\theta})}}_{g(\theta, \hat{\theta})}$$

Estimator is sufficient

unbiasedness, efficiency, consistency, sufficiency } robustness