Stat 452
4-1

Properties of Estimators

Let θ represent an unknown parameter.

Let $\hat{\theta}$ be an estimator of θ.

Defn: If $E[\hat{\theta}] = \theta$ then

$\hat{\theta}$ is an unbiased estimator of θ.

Defn: $E[\hat{\theta}] - \theta$ is the bias of the estimator.

Example: Suppose that μ is a population mean. Let $\bar{x}$ be the sample mean based on n observations. Find the bias of $\bar{x}$.

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu$$

$$= \mu$$

Bias $= E[\bar{x}] - \mu = \mu - \mu = 0$

So $\bar{x}$ is an unbiased estimator of μ.

③

We would like estimators to have small variances.

Find $Var(\bar{X})$

$$= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V[X_i]$$

(assuming independence of $X_1, \ldots, X_n$)

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2)$$

$$= \frac{\sigma^2}{n}$$

Note: $Var(\bar{x}) \to 0$ as $n \to \infty$

This property implies that $\bar{x}$ is a consistent estimator of μ.

④

Suppose that the sample variance, s^2, is used as an estimator of the population variance σ^2.

$$E(s^2) = E\left(\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}\right)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i-\bar{X})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i^2+\bar{X}^2-2X_i\bar{X})$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\Big[\underbrace{E(X_i^2)}_{①}+\underbrace{E(\bar{X}^2)}_{②}-2\underbrace{E(X_i\bar{X})}_{③}\Big]$$

①: $V(X_i) = E(X_i^2) - \mu^2$

$E(X_i^2) = \sigma^2 + \mu^2$

(2) $$V(\bar{X}) = E[\bar{X}^2] - (E[\bar{X}])^2$$

$$\frac{\sigma^2}{n} = E[\bar{X}^2] - \mu^2$$

$$E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$$

(3) $$E[X_i \bar{X}] = E\left[X_i \frac{1}{n} \sum_{j=1}^{n} X_j\right]$$

$$= \frac{1}{n}\left[\sum_{j \neq i} E[X_i X_j] + E[X_i^2]\right]$$

$$= \frac{1}{n}\left[(n-1)\mu^2 + \sigma^2 + \mu^2\right]$$

$$= \frac{1}{n}\left[n\mu^2 + \sigma^2\right] = \mu^2 + \frac{\sigma^2}{n}$$

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

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left[\sigma^2 - \frac{\sigma^2}{n}\right] = \frac{1}{n-1} n\left(\sigma^2 - \frac{\sigma^2}{n}\right)$$

$$= \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \sigma^2$$

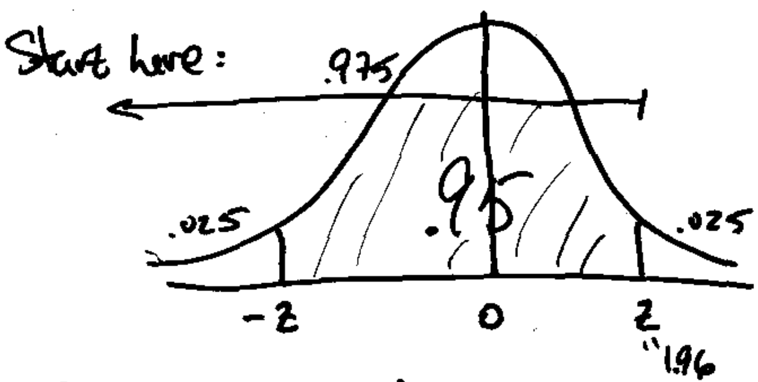
Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables, each with mean μ and variance σ^2, then

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$ will be

approximately normally distributed with $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$, for large enough n.

How to use this:



Standard normal distribution

$$P(-1.96 < Z < 1.96) = .95$$

$$\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$ is approx. std. normal for large n.

$$P\left(-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = .95$$

$$-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96$$ is true with 95% confidence

$$-1.96 \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < 1.96 \frac{\sigma}{\sqrt{n}}$$

$$-\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} > \mu > \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\boxed{\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}}$$ 95% Confidence Interval

(9)

$$\bar{X} \pm \underbrace{1.96\frac{\sigma}{\sqrt{n}}}_{\text{margin of error}}$$

standard error

point estimate

σ is the population standard deviation