

LARGE-SAMPLE PROPERTIES OF ESTIMATORS

We want an estimator $\hat{\theta}_n$ of a parameter $\theta \in \Theta \subset \mathbb{R}$ to tend, in some sense, to be closer to θ when the sample size n is increased.

The following definition provides a formalization of this property of an estimator.

\downarrow a sequence indexed by the sample size n

Defn: An estimator (really a sequence of estimators) $\hat{\theta}_n$ of $\theta \in \Theta \subset \mathbb{R}$ is called consistent if for every $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1.$$

Remark: This means that for any distance $\varepsilon > 0$, $\hat{\theta}_n \in (\theta - \varepsilon, \theta + \varepsilon)$ with probability tending to 1 as $n \rightarrow \infty$, for any value of the unknown parameter θ .

So consistency describes a sense in which $\hat{\theta}_n$ tends to be closer to θ as the sample size n is increased.

Example: Let X_1, \dots, X_n iid Uniform(0, θ), θ unknown.

Show that $\hat{\theta}_n = X_{(n)}$ is a consistent estimator of $\theta \in (0, \infty)$.

Recall

$$X_{(n)} \sim f_{X_{(n)}}(x) = \frac{n}{\theta^n} x^{n-1} \mathbf{1}(0 < x < \theta).$$

Choose any $\varepsilon > 0$. Then

$$P(|\hat{\theta}_n - \theta| < \varepsilon) = P(\theta - \varepsilon < X_{(n)} < \theta + \varepsilon)$$

$$= \int_{\theta - \varepsilon}^{\theta + \varepsilon} \frac{n}{\theta^n} x^{n-1} \mathbf{1}(0 < x < \theta) dx$$

$$= \int_{\theta - \varepsilon}^{\theta} \frac{n}{\theta^n} x^{n-1} \mathbf{1}(0 < x < \theta) dx$$

$$= \begin{cases} 1 - \left[\frac{\theta - \varepsilon}{\theta} \right]^n, & 0 < \varepsilon < \theta \\ 1, & \varepsilon \geq \theta, \end{cases}$$

and, for $0 < \varepsilon < \theta$, $\lim_{n \rightarrow \infty} 1 - \left[\frac{\theta - \varepsilon}{\theta} \right]^n = 1.$
 $\underbrace{(\theta - \varepsilon)/\theta < 1}$

So, $\hat{\theta}_n = X_{(n)}$ is a consistent estimator of θ .

The following theorem, called the Weak Law of Large Numbers (WLLN), says that the sample mean is a consistent estimator of the population mean, provided the population variance is finite.

Theorem: Let X_1, \dots, X_n be a random sample from a dist. with mean μ and variance $\sigma^2 < \infty$. Then $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ is a consistent estimator of μ ; that is, for any value of $\mu \in \mathbb{R}$ and every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

Proof: We make use of Chebychev's inequality:

Since $\text{Var } \bar{X}_n = \sigma^2/n$, we may write, for any $\varepsilon > 0$,

$$P(|\bar{X}_n - \mu| < \varepsilon) = P\left(|\bar{X}_n - \mu| < \left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)\sigma/\sqrt{n}\right)$$

Chebychev's inequality

For any r.v. X with $\mathbb{E}X = \mu$ and $\text{Var } X = \sigma^2 < \infty$,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

$$\geq 1 - \frac{1}{\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)^2}$$

$$= 1 - \frac{\sigma^2}{n} \frac{1}{\varepsilon^2}.$$

and $\lim_{n \rightarrow \infty} 1 - \frac{\sigma^2}{n} \frac{1}{\varepsilon^2} = 1.$

We do not often prove consistency by using its definition directly.
 The following theorem provides sufficient conditions for consistency.

Theorem: A sequence of estimators $\hat{\theta}_n$ of θ is consistent if

$$(i) \quad \lim_{n \rightarrow \infty} \text{Var } \hat{\theta}_n = 0$$

$$(ii) \quad \lim_{n \rightarrow \infty} \text{Bias } \hat{\theta}_n = 0$$

Remark: This means that we can establish the consistency of an estimator by showing that its MSE goes to zero as $n \rightarrow \infty$.

We can very easily prove the WLLN by invoking this theorem.

WLLN (again): Let X_1, \dots, X_n be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Then $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ is a consistent estimator of μ .

Proof: Since $\mathbb{E} \bar{X}_n = \mu$, $\text{Bias } \bar{X}_n = 0$, and since $\text{Var } \bar{X}_n = \sigma^2/n$, $\lim_{n \rightarrow \infty} \text{Var } \bar{X}_n = 0$, so \bar{X}_n is a consistent estimator of μ .

E.g. Let X_1, \dots, X_n be independent r.v.s with the same distribution as $X \sim \text{Uniform}(0, \theta)$.

Consider two possible estimators of θ :

$$\hat{\theta}_n = X_{(n)}$$

$$\tilde{\theta}_n = 2\bar{X}_n$$

In the previous notes we found

$$\text{MSE } \hat{\theta}_n = \theta^2 \left[\frac{2}{(n+2)(n+1)} \right], \quad \text{so} \quad \lim_{n \rightarrow \infty} \text{MSE } \hat{\theta}_n = 0$$

$$\text{MSE } \tilde{\theta}_n = \frac{\theta^2}{3n}, \quad \text{so} \quad \lim_{n \rightarrow \infty} \text{MSE } \tilde{\theta}_n = 0.$$

Therefore, both $\hat{\theta}_n$ and $\tilde{\theta}_n$ are consistent estimators of θ .

In the following example, we see an estimator which is unbiased, but not consistent:

E.g. Let Y_1, \dots, Y_n be a random sample from the Exponential (λ) distribution.

Let $\hat{\lambda}_1 = n Y_{(1)}$.

$$\text{Now } Y_{(1)} \sim f_{Y_{(1)}}(y) = n \left[1 - (1 - e^{-\frac{y}{\lambda}}) \right]^{n-1} \frac{1}{\lambda} e^{-\frac{y}{\lambda}} \mathbf{1}(y > 0)$$

$$= \frac{n}{\lambda} e^{-\frac{y(n-1)}{\lambda}} - \frac{y}{\lambda} \mathbf{1}(y > 0)$$

$$= \frac{1}{(\lambda/n)} e^{\frac{y/(n-1)}{\lambda}} \mathbf{1}(y > 0).$$

pdf of Exponential (n/λ)

$$\text{So } E \hat{\lambda}_1 = n E Y_{(1)} = n \frac{\lambda}{n} = \lambda, \text{ giving Bias } \hat{\lambda}_1 = 0.$$

Variance does not go

$$\text{But } \text{Var } \hat{\lambda}_1 = \text{Var} [n Y_{(1)}] = n^2 \text{Var } Y_{(1)} = n^2 \left(\frac{\lambda}{n} \right)^2 = \lambda. \quad \text{to zero as } n \text{ increases}$$

$$\text{So } \lim_{n \rightarrow \infty} \text{Var } \hat{\lambda}_1 = \lambda, \text{ so } \hat{\lambda}_1 \text{ is not consistent for } \lambda.$$

The following is an example in which one of two consistent estimators is preferred because of its finite-sample properties:

E.g. let X_1, \dots, X_n be independent Bernoulli(p) r.v.s, with p unknown.

Set $Y = X_1 + \dots + X_n$ and consider two estimators of p :

$$\hat{p}_n = Y/n$$

$$\tilde{p}_n = \frac{Y+2}{n+4} \quad \left(\begin{array}{l} \text{add two successes and} \\ \text{two failures to the r.s.} \end{array} \right)$$

In the previous lecture, we found

$$\text{MSE } \hat{p}_n = \frac{p(1-p)}{n}$$

$$\text{MSE } \tilde{p}_n = \left(\frac{n}{n+4}\right)^2 \frac{p(1-p)}{n} + \left(\frac{2-4p}{n+4}\right)^2.$$

We have $\lim_{n \rightarrow \infty} \text{MSE } \hat{p}_n = 0$ and $\lim_{n \rightarrow \infty} \text{MSE } \tilde{p}_n = 0$,

so both \hat{p}_n and \tilde{p}_n are consistent estimators of p .

However, $\text{MSE } \tilde{p}_n \leq \text{MSE } \hat{p}_n$ for some values of p , as discussed in the previous lecture.

We may wish to establish the consistency of a function of estimators for a function of parameters as the following theorem allows:

Theorem: Suppose $\{\hat{\theta}_{1,n}\}_{n \geq 1}$ and $\{\hat{\theta}_{2,n}\}_{n \geq 1}$ are consistent sequences of estimators for the parameters θ_1 and θ_2 , respectively. Then

- (i) $\hat{\theta}_{1,n} \pm \hat{\theta}_{2,n}$ is consistent for $\theta_1 \pm \theta_2$
- (ii) $\hat{\theta}_{1,n} \hat{\theta}_{2,n}$ is consistent for $\theta_1 \theta_2$
- (iii) $\hat{\theta}_{1,n} / \hat{\theta}_{2,n}$ is consistent for θ_1 / θ_2 , so long as $\theta_2 \neq 0$.
- (iv) For any continuous function $T: \mathbb{R} \rightarrow \mathbb{R}$, $T(\hat{\theta}_{1,n})$ is consistent for $T(\theta_1)$.
- (v) For any sequences $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$ such that $\lim_{n \rightarrow \infty} a_n = 1$ and $\lim_{n \rightarrow \infty} b_n = 0$, $a_n \hat{\theta}_{1,n} + b_n$ is consistent for θ_1 .

This theorem allows us to prove consistency of S_n^2 for σ^2 , a fact of which we will later make much use.

Result (Consistency of S_n^2): Let X_1, \dots, X_n be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$ and 4th moment $\mu_4 < \infty$. Then S_n^2 is a consistent estimator of σ^2 .

Proof: First, note that $\sigma^2 = \mu_2 - \mu_1^2$.

Recall the notation of moments:

$$\begin{aligned}\mu_1 &= \mathbb{E}X^1 \\ \mu_2 &= \mathbb{E}X^2 \\ \mu_3 &= \mathbb{E}X^3 \\ \mu_4 &= \mathbb{E}X^4\end{aligned}$$

Now decompose S_n^2 as follows:

$$\begin{aligned}S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right] \\ &= \left(\frac{n}{n-1} \right) \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{n}{n-1} \right) \bar{X}_n^2\end{aligned}$$

(a) By the WLLN, $\frac{1}{n}(X_1^2 + \dots + X_n^2)$ is a consistent estimator of μ_2 ; it is the sample mean of X_1^2, \dots, X_n^2 , where $\mathbb{E}X_i^2 = \mu_2$ and

$$\text{Var } X_i^2 = \mathbb{E}[X_i^2]^2 - (\mathbb{E}X_i^2)^2 = \mu_4 - \mu_2^2 < \infty.$$

(b) By (iv) of the preceding theorem, \bar{X}_n^2 is consistent for μ_1^2 , since \bar{X}_n is consistent for μ and $\mathcal{I}(\cdot) = (\cdot)^2$ is a continuous function.

(c) By (v) of the preceding theorem and (a), since $\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1$,

$$\left(\frac{n}{n-1} \right) \frac{1}{n} \sum_{i=1}^n X_i^2 \text{ is consistent for } \mu_2.$$

(d) By (v) of the preceding theorem and (b), since $\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1$,

$$\left(\frac{n}{n-1} \right) \bar{X}_n^2 \text{ is consistent for } \mu_1^2.$$

(e) By (i) of the preceding theorem, (c) and (d) give

$$S_n^2 = \left(\frac{n}{n-1} \right) \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{n}{n-1} \right) \bar{X}_n^2 \text{ is consistent for } \mu_2 - \mu_1^2 = \sigma^2. \quad \square$$

The following result is also very important:

Result (Consistency of $\hat{p}(1-\hat{p})$): Let $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p)$, and let $\hat{p} = \frac{1}{n}(X_1 + \dots + X_n)$. Then $\hat{p}(1-\hat{p})$ is a consistent estimator of $\text{Var } X_i = p(1-p)$.

Proof: (a) By the WLLN, \hat{p} is consistent for $E X_i = p$.
(b) By (iv) in the preceding theorem, since \hat{p} is consistent for p and $T(\cdot) = (\cdot)(1-\cdot)$ is a continuous function, $\hat{p}(1-\hat{p})$ is consistent for $p(1-p)$. \square

NOTATION FOR CONSISTENCY

If $\hat{\theta}_n$ is consistent for θ , we often write

$$\hat{\theta}_n \xrightarrow{P} \theta,$$

where the " \xrightarrow{P} " denotes something called convergence in probability. So we sometimes say " $\hat{\theta}_n$ converges in probability to θ ".