STA261 Lecture 7 — 2017-07-31

Neil Montgomery

Last edited: 2017-07-31 18:58

unbiased estimators with smallest variance

In Lecture 6 we found a lower bound on the variance of an unbiased estimator (Cramer-Rao), under some conditions.

unbiased estimators with smallest variance

In Lecture 6 we found a lower bound on the variance of an unbiased estimator (Cramer-Rao), under some conditions.

There were two examples where the bound was acheived (Poisson and Normal), and one example where the bound was beaten (because the conditions were violated).

unbiased estimators with smallest variance

In Lecture 6 we found a lower bound on the variance of an unbiased estimator (Cramer-Rao), under some conditions.

There were two examples where the bound was acheived (Poisson and Normal), and one example where the bound was beaten (because the conditions were violated).

In today's tutorial you found an unbiased estimator where the bound was not acheived. So perhaps there is a better unbiased estimator (in fact there is not) for λ from a Gamma(α_0, λ) distribution.

sufficiency

There is one last property of an estimator that we'll look at, called *sufficiency*, which captures the idea that there is no information about θ left behind in the sample.

sufficiency

There is one last property of an estimator that we'll look at, called *sufficiency*, which captures the idea that there is no information about θ left behind in the sample.

Consider a sample $X_1, \ldots, X_n$ from a Bernoulli(p) distribution and let $T_k = \sum_{i=1}^{k} X_i$ be the sum of the first k items in the sample.

sufficiency

There is one last property of an estimator that we'll look at, called *sufficiency*, which captures the idea that there is no information about θ left behind in the sample.

Consider a sample $X_1, \ldots, X_n$ from a Bernoulli(p) distribution and let $T_k = \sum_{i=1}^{k} X_i$ be the sum of the first k items in the sample.

Note 6.0: Say $n = 20$ and $k = 10$. The conditional distribution of $(X_1, X_2, \ldots, X_{20})$ given $T_{10} = t$ depends on the value of p.

T_{10} leaves behind some information about p in the sample.

sufficiency

Note 6.1: But if $n = 20$ and $k = 20$, the conditional distribution of $(X_1, X_2, \ldots, X_{20})$ given $T_{20} = t$ does not depend on the value of p.

sufficiency

Note 6.1: But if $n = 20$ and $k = 20$, the conditional distribution of $(X_1, X_2, \ldots, X_{20})$ given $T_{20} = t$ does not depend on the value of p.

Definition: A statistic $T(X_1, \ldots, X_n)$ is *sufficient* for θ if the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $T = t$ does not depend on θ.

Note: the definition works for vector-valued T and vector θ.

sufficiency via factorization of pmf or pdf

The definition doesn't help so much for actually finding sufficient statistics.

Proposition 6.2: $T(X_1, \ldots, X_n)$ is sufficient for θ if and only if there are two functions g and h such that:

$$L(\theta; \mathbf{x}) = g(T(\mathbf{x}), \theta)h(\mathbf{x}).$$

factorization examples, and notes

Example 6.3: Sample is from Uniform$(0,\theta)$. $X_{(n)}$ is sufficient for θ.

factorization examples, and notes

Example 6.3: Sample is from Uniform(0,θ). $X_{(n)}$ is sufficient for θ.

Example 6.4: Sample is from $N(\mu, \sigma^2)$. The vector $(\overline{X}, S^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

factorization examples, and notes

Example 6.3: Sample is from Uniform($0,\theta$). $X_{(n)}$ is sufficient for θ.

Example 6.4: Sample is from $N(\mu, \sigma^2)$. The vector $(\overline{X}, S^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

Note 6.5: Any 1-1 function of a sufficient statistic is also sufficient.

factorization examples, and notes

Example 6.3: Sample is from Uniform($0,\theta$). $X_{(n)}$ is sufficient for θ.

Example 6.4: Sample is from $N(\mu, \sigma^2)$. The vector $(\overline{X}, S^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

Note 6.5: Any 1-1 function of a sufficient statistic is also sufficient.

Example 6.6: Sample is from $N(\mu, \sigma^2)$. The vector $(\sum X_i, \sum X_i^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

recall "conditional expectation"

For random variables X and Y, the expected value of Y given a particular value $X = x$ is (continuous case):

$$E(Y|X = x) = \int y f(y|x)\, dy$$

where $f(y|x) = f(x, y)/f_X(x)$ is the conditional density of Y given X *with that particular value of* x *plugged in.*

recall "conditional expectation"

For random variables X and Y, the expected value of Y given a particular value $X = x$ is (continuous case):

$$E(Y|X = x) = \int yf(y|x)\,dy$$

where $f(y|x) = f(x, y)/f_X(x)$ is the conditional density of Y given X *with that particular value of x plugged in.*

This is a constant, often dependent on the value of x. So you can think of $E(Y|X = x)$ as a function of x, say $g(x)$.

recall "conditional expectation"

For random variables X and Y, the expected value of Y given a particular value $X = x$ is (continuous case):

$$E(Y|X = x) = \int yf(y|x)\,dy$$

where $f(y|x) = f(x, y)/f_X(x)$ is the conditional density of Y given X *with that particular value of* x *plugged in.*

This is a constant, often dependent on the value of x. So you can think of $E(Y|X = x)$ as a function of x, say $g(x)$.

The *random variable* $g(X)$ is called the "conditional expectation of Y given X", or $E(Y|X)$.

properties of conditional expectation | “conditional variance”

$$E(g(X)) = E(E(Y|X)) = E(Y)$$

properties of conditional expectation | "conditional variance"

$$E(g(X)) = E(E(Y|X)) = E(Y)$$

The conditional variance of Y given X is the random variable:

$$\mathrm{Var}(Y|X) = E\left(Y^2|X\right) - [E(Y|X)]^2$$

properties of conditional expectation | “conditional variance”

$$E(g(X)) = E(E(Y|X)) = E(Y)$$

The conditional variance of Y given X is the random variable:

$$\mathrm{Var}(Y|X) = E\left(Y^2|X\right) - [E(Y|X)]^2$$

Fact:

$$\mathrm{Var}(Y) = \mathrm{Var}(E(Y|X)) + E(\mathrm{Var}(Y|X))$$

Rao-Blackwellization

Under certain conditions, if you find an unbiased estimator that is a function of a sufficient statistic, *it will be the best unbiased estimator*.

The reason is mostly because of what is called the Rao-Blackwell theorem.

Proposition 6.7: Suppose W is unbiased for θ, and suppose T is sufficient for θ. Then the random variable $E(W|T)$ is unbiased with smaller variance than W.