STA261 Lecture 6 — 2017-07-26

Neil Montgomery

Last edited: 2017-07-26 19:01

unfinished business from Lecture 5

Theorem 5.2 had a typo in its statement

Theorem 5.2
$$\mathcal{I}(\theta) = \text{Var}(S(\theta))$$
 and $\mathcal{I}(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \mathbf{X})\right)$

The proof was fine.

Proposition 5.4

The definition of information uses a sample:

$$\mathcal{I}(heta) = \mathcal{I}(heta; \mathbf{X}) = Eigg(igg(rac{\partial}{\partial heta} \log L(heta; \mathbf{X})igg)^2igg)$$

Rice defines something also called $I(\theta)$ which is for a single random variable X **Proposition 5.4a** $\mathcal{I}(\theta) = nI(\theta)$.

Proposition 5.4

The definition of information uses a sample:

$$\mathcal{I}(\theta) = \mathcal{I}(\theta; \mathbf{X}) = E\left(\left(\frac{\partial}{\partial \theta} \log L(\theta; \mathbf{X})\right)^2\right)$$

Rice defines something also called $I(\theta)$ which is for a single random variable X

Proposition 5.4a $\mathcal{I}(\theta) = nI(\theta)$.

Proposition 5.4 Under suitable conditions, $\sqrt{\mathcal{I}(\theta)}(\hat{\theta} - \theta) = \sqrt{nI(\theta)}(\hat{\theta} - \theta)$ converges (in distribution) to a standard normal distribution.

a few MLE extras

Asymptotic normality implies (under the same "regularity") conditions asymptotic unbiasedness.

a few MLE extras

Asymptotic normality implies (under the same "regularity") conditions asymptotic unbiasedness.

In many cases $Var(\hat{\theta}) = \mathcal{I}(\theta)^{-1}$. For large samples we can still often use $Var(\hat{\theta}) \approx \mathcal{I}(\theta)^{-1}$.

example where things fall apart

Example 6.0: Consider Uniform[0, θ]. Which nice properties does $\hat{\theta}$ have?

Under suitable conditions:

smoothness of density

Under suitable conditions:

- smoothness of density
- \blacktriangleright non-zero part of density does not depend on θ

Under suitable conditions:

- smoothness of density
- \blacktriangleright non-zero part of density does not depend on θ

Under suitable conditions:

- smoothness of density
- lacktriangleright non-zero part of density does not depend on heta

an MLE is:

consistent (not proven)

Under suitable conditions:

- smoothness of density
- lacktriangleright non-zero part of density does not depend on heta

an MLE is:

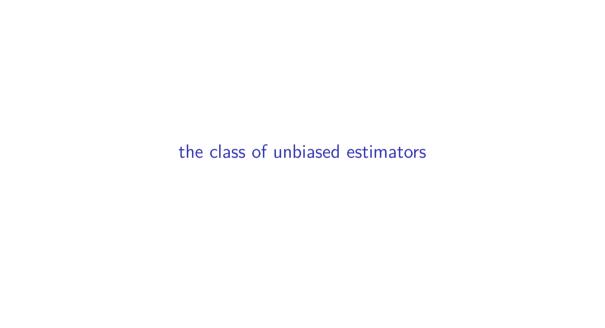
- consistent (not proven)
- asymptotically normal (proven) (and therefore asymptotically unbiased)

Under suitable conditions:

- smoothness of density
- lacktriangleright non-zero part of density does not depend on heta

an MLE is:

- consistent (not proven)
- asymptotically normal (proven) (and therefore asymptotically unbiased)
- ▶ invariant, $\widehat{h(\theta)} = h(\widehat{\theta})$ (not proven tedious in general but very easy when, say h is differentiable and monotone increasing. Try it!)



a possible "gold standard" for estimators

If we restrict ourselves to unbiased estimators only, a plausible standard would be to choose the one that has the smallest variance.

The purpose of the next part of the course is to see when it is possible to find such estimators.

a possible "gold standard" for estimators

If we restrict ourselves to unbiased estimators only, a plausible standard would be to choose the one that has the smallest variance.

The purpose of the next part of the course is to see when it is possible to find such estimators.

This will cause us to define and investigate the notion of "sufficiency". But not yet.

a lower bound on the variance of an unbiased estimator

If you somehow knew the lowest possible variance among unbiased estimators, and you have an unbiased estimator with that variance, then you're done.

Reminder 6.1: The correlation coefficient ρ is a property of a joint distribution, and $-1 \leqslant \rho \leqslant 1$.

Theorem 6.2: Suppose X_1, \ldots, X_n is i.i.d. with density $f(x; \theta)$ and $T = t(X_1, \ldots, X_n)$ is unbiased for θ . Then (under the usual conditions on f):

$$\mathsf{Var}(T)\geqslant rac{1}{\mathcal{I}(heta)}$$

a lower bound on the variance of an unbiased estimator

If you somehow knew the lowest possible variance among unbiased estimators, and you have an unbiased estimator with that variance, then you're done.

Reminder 6.1: The correlation coefficient ρ is a property of a joint distribution, and $-1 \leqslant \rho \leqslant 1$.

Theorem 6.2: Suppose X_1, \ldots, X_n is i.i.d. with density $f(x; \theta)$ and $T = t(X_1, \ldots, X_n)$ is unbiased for θ . Then (under the usual conditions on f):

$$\mathsf{Var}(T)\geqslant rac{1}{\mathcal{I}(heta)}$$

This is called the "Cramer-Rao" lower bound.

examples

Example 6.3: $N(\mu, 1)$

Example 6.4: Poisson(λ)

Example 6.5: Uniform $[0, \theta]$