14.381 STATISTICS
FALL 2004

12/01/2004
12/03/2004
TWO RECITATIONS

HANDOUT #9

TA: José Tessada
(tessada@mit.edu)

CRAMER-RAO LOWER BOUND: Let $X_1, \ldots, X_n$ be a sample with pdf $f(x|\theta)$, and let $W(X) = W(X_1, \ldots, X_n)$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta[W(X)] = \int_X \frac{\partial}{\partial \theta} [W(x) \; f(x|\theta)] \, dx$$

and

$$Var_\theta[W(x)] < \infty$$

Then,

$$Var_\theta(W(X)) \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(x)]\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial \theta} \log f(x|\theta)\right)^2\right)}$$

The approach in this case is completely different to the one based on complete, sufficient statistics. We try to find a bound for the variance and then see whether we can find an unbiased estimator that achieves this lower bound.
Be careful with the fact that we need to be able to take derivatives inside of the integrals. Michael stated some implications of the regularity conditions.
Be aware of some differences between what is written in the book and what Michael says in the lectures.

Special cases of the C-R formula:

- iid sample
- exponential family (and some other pdfs → see Lemma 7.3.11, p. 338)

Corollary 7.3.15 (Attainment): Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, where $f(x|\theta)$ satisfies the conditions of the Cramer-Rao Theorem. Let $L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$ denote the likelihood function. If $W(X) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(X)$ attains the Cramer-Rao lower bound iff

$$a(\theta)\left[W(x) - \tau(\theta)\right] = \frac{\partial}{\partial\theta} \log L(\theta|x)$$

for some function $a(\theta)$.

Using some results stated in the book we can also write the following corollary.

Corollary: If $X_1, \ldots, X_n$ are iid from an exponential family, with pdf $f(x|\theta)$, and if $W(X) = W(X_1, \ldots, X_n)$ with

$$\frac{d}{d\theta} E_\theta[W(X)] = \int \frac{\partial}{\partial\theta}\left[W(x) f(x|\theta)\right] dx$$

$$\mathrm{Var}_\theta[W(X)] < \infty$$

then

$$\mathrm{Var}_\theta[W(X)] \geq \frac{\left[\frac{d}{d\theta} E_\theta(W(X))\right]^2}{-n E_\theta\left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)\right]}$$

Notes:

(a) It is not easy to show that the conditions of the CR thm. hold; you can try to compute the bound and then show that there is an estimator that beats it.

(b) Quick guide: in general, if the range of the pdf depends on the parameter(s), then the theorem will not be applicable.

(c) Nothing guarantees that if the C-R lower bound exists, there is an estimator that attains it.

Example: $N(\mu, \sigma^2)$ $\quad X_i \sim iid$

$$L(\mu, \sigma^2 | X) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right\}$$

$$\frac{\partial}{\partial\sigma^2}\log L(\mu, \sigma^2 | X) = \frac{n}{2\sigma^4}\left(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

with $a(\sigma^2) = n/(2\sigma^4)$, we have that the best unbiased estimator of σ^2 is $\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n}$.

The problem is that you can compute iff μ is known.
$\Rightarrow$ in the general case when μ is unknown this expression cannot be computed.

What can we do? We can try to find an estimator with a variance "close" to the bound.

Normal distribution: C-R lower bound

$$L(\mu, \sigma^2 | X) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right\}$$

Compute the CR lower bound using the formula for random sample from exponential family:

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log L = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

$$-E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\log L\right) = -E\left(\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}\right) = \frac{1}{2\sigma^4}$$

So any unbiased estimator of σ^2 satisfies

$$\mathrm{Var}(W) \geq \frac{2\sigma^4}{n}$$

Now, we know that

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

Consider an alternative estimator:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n+1} = \frac{1}{n+1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

We know that

$$Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2_{n-1}$$

$$\mathrm{Var}(Y) = 2(n-1)$$

$$\hat{\sigma}^2 = \frac{\sigma^2}{n+1} \cdot Y \quad \Rightarrow \quad \mathrm{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{(n+1)^2}\mathrm{Var}(Y)$$

$$= 2\sigma^4 \frac{n-1}{(n+1)^2}$$

$\mathrm{Var}(\hat{\sigma}^2) <$ C-R lower bound !

$$\underbrace{\frac{n-1}{n+1}}_{<1} \cdot \underbrace{\frac{1}{n+1}}_{<\frac{1}{n}} \quad \Rightarrow \quad \frac{n-1}{(n+1)^2} < \frac{1}{n}$$

Why is this happening?
We are looking at estimators with different expected values.

HYPOTHESIS TESTING

- Null hypothesis: what we are assuming to be true.
- Alternative hypothesis: the alternative; literally...
- p-value: the probability of obtaining the test statistic we computed or one more extreme given our null is true.
- significance level: before we decide to reject or accept, we set up a significance level (normally 5%). This is the probability of rejecting the null when the null is true. We reject H_0 if p-value < Sig. level
- power: $P(\text{rejecting } H_0 \mid \theta \in \Theta)$. The power is a function of θ. At $\theta = \theta_0$ (if Θ_0 is a singleton), the power is just the significance level. As θ moves away from θ_0 the power should increase (greater chance of rejecting $\theta = \theta_0$).
- Type I error: Rejecting H_0 when H_0 is true.
- Type II error: ✓ H_1 when H_1 is true.

$$\text{Power} = \begin{cases} P(\text{type I}) & \text{when } \theta \in \Theta_0 \\ 1 - P(\text{type II}) & \text{when } \theta \in \Theta_1 \end{cases}$$

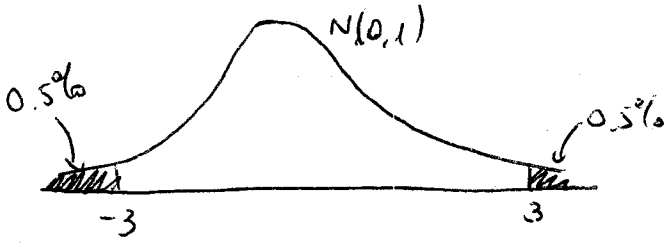
Example — Intuition:
The nature of testing is to use our sample to ask questions about how likely something is in the population, and to draw conclusions based on that likelihood.
→ You have a population of men. You <u>believe</u> their mean height is $\theta_0 = 68"$. The variance is known to be $\sigma^2 = 16"$. You sample 100 men, they have an average height of $69.2"$

(a) Test if the population mean $\neq 68$
(b) Test if the population mean > 68

Under the null hypothesis (of $\theta_0 = 68"$),

$$Z = \frac{69.2 - 68}{(4/10)} = 3 \quad \text{has} \quad Z \sim N(0,1)$$

because $\bar{X} \sim N(68, \sigma^2/n)$



(a) If the null is true, there is a 1% chance of observing what we observed;
(b) If the null is true, there is a 0.5% chance of observing what we observed.

Both probabilities are small - given that we observed $\bar{X} = 69.2$, it is unlikely that $\theta_0 = 68$.

Likelihood ratio test

$$T_{LR} = 2\left[\ell(\hat{\theta}_{ML} \mid x_1, \ldots, x_n) - \ell(\theta_0 \mid x_1, \ldots, x_n)\right]$$

→ Neyman - Pearson lemma: if this exists, it's the best.

example: $x_1, \ldots, x_n$ iid $f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta \end{cases}$

$$L(\theta|x) = \begin{cases} e^{-\sum x_i + n\theta} & \theta \leq x_{(1)} \\ 0 & \theta > x_{(1)} \end{cases}$$

$H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$

$$\log L(\theta|x) = \begin{cases} -\Sigma x_i + n\theta & \theta \leq x_{(1)} \\ -\infty & \theta > x_{(1)} \end{cases}$$

$$T_{LR} = \begin{cases} 2\left[-\Sigma x_i + n x_{(1)} - (-\Sigma x_i + n\theta_0) \right] & \text{if } x_{(1)} > \theta_0 \\ 0 & \text{if } x_{(1)} \leq \theta_0 \end{cases}$$

$$T_{LR} = 2(n(x_{(1)} - \theta_0))$$

$$T_{LR} = 2n(x_{(1)} - \theta_0)$$

Intuitively you will reject for $x_{(1)}$ "sufficiently" far to the right ... $x_{(1)} >> \theta_0$ implies that you reject H_0

Lagrange multiplier test

$$T_{LM}(x_1, \ldots, x_n) = \frac{\left[\frac{\partial}{\partial\theta} \ell(\theta_0 | x_1, \ldots, x_n)\right]^2}{-\frac{\partial^2}{\partial\theta^2} \ell(\theta_0 | x_1, \ldots, x_n)}$$

Why is this useful?
- No need to actually compute any estimator
- Has a clear interpretation in terms of a constrained optimization problem.

$$\frac{\partial}{\partial\theta} \ell(\theta_0 | x_1, \ldots, x_n) = \lambda$$

where λ is the Lagrange multiplier in the constrained maximization problem.

Wald test statistic

$$T_W(X_1, \ldots, X_n) = \frac{(\hat{\theta}_{ML} - \theta_0)^2}{\left[-\frac{\partial^2}{\partial \theta^2} \ell(\hat{\theta}_{ML} | X_1, \ldots, X_n)\right]^{-1}}$$

where $\left[-\frac{\partial^2}{\partial \theta^2} \ell(\hat{\theta}_{ML} | X_1, \ldots, X_n)\right]^{-1} \approx \text{Var}(\hat{\theta}_{ML})$

This T_W is a different approximation to the LRT. The Wald principle is far more general and applies in many other situations.

Wald principle: reject H_0 when $|\hat{\theta} - \theta_0|$ is "large", where $\hat{\theta}$ is some estimator of θ.

Notes:

- if $X_i \sim \text{iid } N(\mu, \sigma^2)$ σ^2 known, all the 3 tests are numerically equivalent!

Generalization of the LRT:

$$H_0: \theta \in \Theta_0 \quad \text{vs} \quad H_1: \theta \in \Theta_1$$

LR test statistic:

$$T_{LR}(X_1, \ldots, X_n) = 2\left[\ell(\hat{\theta}_{ML} | X_1, \ldots, X_n) - \ell(\hat{\theta}_0 | X_1, \ldots, X_n)\right]$$

where

$$\hat{\theta}_{ML} = \underset{\theta \in \Theta_0 \cup \Theta_1}{\text{argmax}} \; \ell(\theta | X_1, \ldots, X_n)$$

$$\hat{\theta}_0 = \underset{\theta \in \Theta_0}{\text{argmax}} \; \ell(\theta | X_1, \ldots, X_n)$$

Exercises

① $Y_i = \beta X_i + \varepsilon_i$ $\quad$ X_i constants (known)

$\varepsilon_i \sim N(0, \sigma^2)$ $\quad$ σ^2 known

iid

Want to test $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$.

- Construct the LRT
- Show it is numerically equivalent to the traditional Z-test. (the same as a t-test but it has a $N(0,1)$ distribution because σ^2 is known).