Wed: talk about quiz, start Notes 6
OH today 11:20-12:20, Wed 2:30-4
HW 4 due on Wed night

05: CONSISTENCY AND INVARIANCE

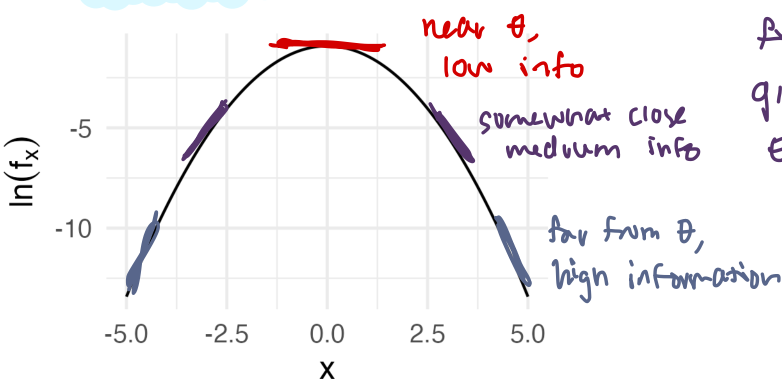
Larsen & Marx 5.7

Prof Amanda Luby

Cramer-Rao lower bound $V(\hat{\theta}) \geq \frac{1}{nI(\theta)}$

1 Fisher Information Follow Up

$$I(\theta) = E[(\frac{\partial \ln f_y(y;\theta)}{\partial \theta})^2] = [-E(\frac{\partial^2 \ln f_y(y;\theta)}{\partial \theta^2})]$$



$X \sim N(0,1)$

$I(\theta)$ is a general formula for how "informative" any given data point is for estimating θ, based on the pdf

info = "unexpected"

2 Consistent Estimators

When we've considered bias and efficiency, we've mostly assumed that our data has a fixed sample size. This makes sense in the context of historical statistics: data was time-consuming and expensive to gather, and so experiments were very rigorously designed with a lot of consideration for sample sizes. For any given dataset, we're generally working with a fixed sample size. As data has become easier and cheaper to gather, the *asymptotic* behavior of estimators has also become an important consideration. We may find, for example, that an estimator has a desired behavior *in the limit* that it fails to have for any fixed sample size.

Example: Recall the MLE for a Unif$(0, \theta)$ distribution is $\hat{\theta} = X_{\text{max}}$. In Notes02, we showed that $E(X_{\text{max}}) = \frac{n}{n+1}\theta$.

$n=3 \rightarrow E(X_{max}) = \frac{3}{4}\theta$

$n=100 \quad = \frac{100}{101}\theta$

$n=10,000 \quad = \frac{10000}{10,001}\theta$

$\lim_{n\to\infty} E(X_{max}) = \lim_{n\to\infty} \frac{n}{n+1}\theta = \theta$

$\rightarrow X_{max}$ is asymptotically unbiased for θ

Consistency

As estimator $\hat{\theta}_n = h(W_1, ..., W_n)$ is said to be *consistent* if it converges in probability to to θ – that is, for all $\epsilon > 0$:

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1$$

equivalent strategies

Note: To solve certain kinds of problems, it can be helpful to think of this definition in an epsilon/delta way: $\hat{\theta}_n$ is consistent if for all $\epsilon > 0$ and $\delta > 0$, there exists $n(\epsilon, \delta)$ such that:

$$P(|\hat{\theta}_n - \theta| < \epsilon) > 1 - \delta \quad \text{for some } n(\epsilon, \delta)$$

If we set ϵ, δ to be constant, can we find an n that makes this true?

Example: Is the MLE for a Unif$(0, \theta)$ distribution consistent? $\hat{\theta} = X_{max}$ $\quad f_{Y_{max}} = \frac{ny^{n-1}}{\theta^n}$

$$\begin{aligned} P(|\hat{\theta}_n - \theta| < \epsilon) &= P(-\epsilon < \hat{\theta}_n - \theta < \epsilon) \\ &= P(\theta - \epsilon < \hat{\theta}_n < \theta + \epsilon) \\ &= P(\theta - \epsilon < \hat{\theta}_n < \theta) \\ &= \int_{\theta-\epsilon}^{\theta} \frac{ny^{n-1}}{\theta^n} dy \\ &= \frac{y^n}{\theta^n} \Big|_{\theta-\epsilon}^{\theta} \\ &= 1 - \left(\frac{(\theta-\epsilon)^n}{\theta^n}\right) \\ &= 1 - \left(\frac{\theta-\epsilon}{\theta}\right)^n \end{aligned}$$

since $\frac{\theta - \epsilon}{\theta} < 1$, $\lim_{n \to \infty} 1 - \left(\frac{\theta-\epsilon}{\theta}\right)^n = 1$

$\rightarrow \hat{\theta}_n = X_{max}$ is consistent for θ

Note 1: "CDF approach" is uglier if ϵ is in both limits of integration

Note: consistent $\Rightarrow$ asymptotically unbiased but the reverse is not always true

$\frac{1}{n}\sum X_i \to \theta \qquad E(\frac{1}{n}\sum X_i) \to \theta$

There are a number of useful *inequalities* in probability theory that make proving consistency easier. I'm going to give a quick overview of some of these inequalities here, but they can also be found in Blitzstein & Hwang Ch 10.1. The proofs are extremely short and sweet, and I highly recommend reading this subsection of the book if you didn't cover it in Stat51.

Cauchy-Schwarz inequality

For any random variables X and Y with finite variances,

$$|E(XY)| \le \sqrt{E(X^2)\,E(Y^2)}$$

Example: Let $X = X \quad Y = 1$

$$|E(X \cdot 1)| \le \sqrt{E(X^2)\,E(1)}$$

$$|E(X)| \le \sqrt{E(X^2)}$$

$$E(X)^2 \le E(X^2)$$

$\cup$ $f'' > 0$ $\qquad$ $\cap$ $f'' < 0$

Jensen's Inequality

Let W be a random variable, and let g be a convex function and h be a concave function:

$$E(g(X)) \geq g(E(X)) \qquad E(h(X)) \leq h(E(X))$$

Example: $\cdot\ x^2$ convex $\rightarrow E(X^2) \geq E(X)^2$ $\qquad$ $\cdot\ 1/x$ convex $x>0$ $\quad E(1/X) \geq 1/E(X)$

$\cdot\ |x|$ convex $\rightarrow E|X| \geq |E(X)|$ $\qquad$ $\cdot\ \ln x$ concave $x>0$ $\quad E(\ln X) \leq \ln E(X)$

tail behavior/ probabilities

Markov's Inequality

For any random variable W and any constant a,

$$P(|W| > a) \leq \frac{E(|W|)}{a}$$

Chebyshev's inequality

Let W be any random variable with mean μ and variance σ^2. For any $\epsilon > 0$,

$$P(|W - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Chernoff's inequality

Let W be any random variable and constants a and t,

$$P(W \geq a) \leq \frac{E(e^{tW})}{e^{ta}}$$

MGF

- Most useful if we know MGF of W
- minimize WRT t to get a tighter bound

Example: Let $X_1, ..., X_n$ be a random sample from a discrete pdf $p_x(k; \mu)$, where $E(X) = \mu$ and $V(X) = \sigma^2 < \infty$. Let $\hat{\mu}_n = \frac{1}{n}\sum X_i$. Is $\hat{\mu}$ a consistent estimator for μ?

By Chebyshev's inequality,

let $W = \hat{\mu}_n$

$\rightarrow E(W) = \mu$

$\rightarrow V(W) = \frac{1}{n^2}\sum V(X_i)$

$= \frac{1}{n^2} \cdot n \cdot \sigma^2$

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

$$P(|\hat{\mu}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0 \Rightarrow \hat{\mu}_n = \frac{1}{n}\sum X_i$ is consistent for μ regardless of distribution of X!

Note: This is the weak law of large numbers (WLLN) and it was first proved by Chebyshev in 1866

Example: Let $X_1, ..., X_n \sim$ Unif$(0, \theta)$. Recall $\hat{\theta}_{MoM} = 2\bar{X}$, and $E(\hat{\theta}_{MoM}) = \theta$ and $V(\hat{\theta}_{MoM}) = \frac{\theta^2}{3n}$ (Notes02). Is $\hat{\theta}_{MoM}$ consistent for θ? * will do on Wed

3 Invariant Estimators

We're not going to go as in-depth with this property right now, but we'll come back to it over the next few weeks. Hopefully it is intuitive why it is desirable.

Invariance Property of consistent estimators

Any continuous function of a consistent estimator is consistent.

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1 \Rightarrow \lim_{n\to\infty} P(|g(\hat{\theta}) - g(\theta)| < \epsilon) = 1$$

Invariance Property of MLE's

Let $W_1, ..., W_n$ be a random sample from some distribution $f_w(\theta)$, and let $\hat{\theta} = h(W_1, ..., W_n)$ be the maximum likelihood estimator for θ. Suppose we want to find the estimator for $g(\theta)$, where g is any function.

$g(\hat{\theta})$ is the MLE for $g(\theta)$

This is also sometimes called "plug-in principle"