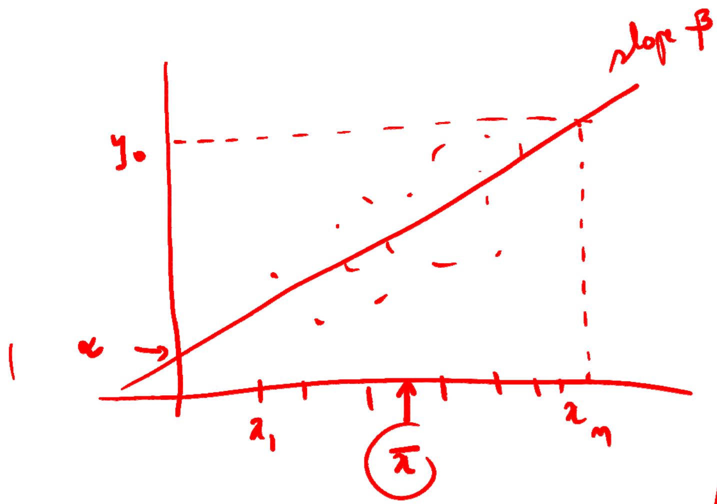


$$y = \alpha_1 + \beta x$$

$x_0 = 0$

use MLE

$\hat{\alpha}$ and $\hat{\beta}$

$$\alpha_1 = \alpha - \beta \bar{x}$$

$$\hat{\alpha}_1 = \hat{\alpha} - \hat{\beta} \bar{x}$$

$$= \bar{y} - \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})} \cdot \bar{x}$$

$y =$

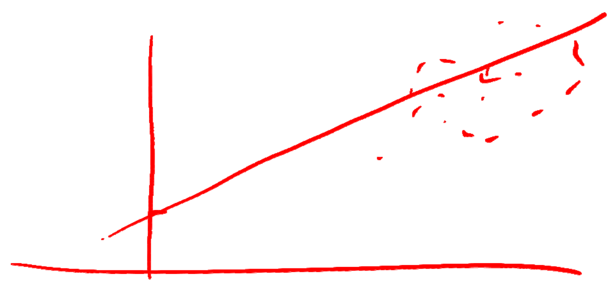
or

$$y - \underbrace{y_0}_{\alpha} = \beta(x - x_0)$$

$$y = \alpha + \beta(x - x_0)$$

$$y = \alpha + \beta(x - \bar{x})$$

$$\hat{\alpha} = \bar{y}$$



Math 4501 - Probability and Statistics II

lower-bound on variance of unbiased estimators

6.6 - $\underbrace{\text{Cramér-Rao}}$ inequality and the efficiency on an unbiased estimator

$X_1, \dots, X_n$ random sample (iid)

estimator $\hat{\theta} = u(X_1, \dots, X_n)$ $\leftarrow$ estimator $\hat{\theta}$ is itself a r.v

$\hat{\theta}$ is unbiased if $E[\hat{\theta}] = \theta$

if $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators for θ, how to choose?

Some mild regularity conditions

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf

$$f(x;\theta)\ , \quad x \in S\ , \quad \theta \in \Omega = \{\theta : c < \theta < d\} \subseteq \mathbb{R}\ ,$$

with support S not depending on the unknown parameter θ. ← IMPORTANT !!

Let $Y = u(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ. ⇝ $\hat{\theta} = Y = u(X_1, \ldots, X_n)$

In what follows, we assume that the following *regularity conditions* hold:

1. $\dfrac{\partial}{\partial\theta} \ln f(x;\theta)$ exists for all $x \in S$ and all $\theta \in \Omega$. ⇝ f is differentiable w.r.t θ : $\dfrac{\partial}{\partial\theta} \ln(f(x,\theta)) = \dfrac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)}$
2. $\dfrac{\partial}{\partial\theta} \displaystyle\int \cdots \int L(\bar{x};\theta)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int \cdots \int \frac{\partial}{\partial\theta} L(\bar{x};\theta)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$
3. $\dfrac{\partial}{\partial\theta} \displaystyle\int \cdots \int u(\bar{x}) L(\bar{x};\theta)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int \cdots \int u(\bar{x}) \frac{\partial}{\partial\theta} L(\bar{x};\theta)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$
4. $0 < E\left[\left(\dfrac{\partial}{\partial\theta} \ln f(X;\theta)\right)^2\right] < \infty$ for all $\theta \in \Omega$.

where we use the notation $\bar{x} = (x_1, \ldots, x_n)$, $L(\bar{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ and $u(\bar{x}) = u(x_1, \ldots, x_n)$.

Further notes:

Condition (1) is a requirement on the differentiability of $f(x;\theta)$ with respect to θ.

Conditions (2) and (3) ensure that the operations of integration with respect to $x_1, \ldots, x_n$ and differentiation with respect to θ can be interchanged when evaluating the derivatives of the expectations on the corresponding left hand sides.

- This can often be confirmed when either of the following cases hold:
    1) The function $f(x;\theta)$ has bounded support in x (and we are assuming that the bounds do not depend on θ);
    2) The function $f(x;\theta)$ has unbounded support, is continuously differentiable in x, and the integral for the expectation of the estimator Y converges uniformly for all θ.

The case of discrete distributions:

- the development for discrete random variables is completely analogous, with the obvious replacement of integrals by summations.

Cramér-Rao inequality

Theorem (Cramér-Rao inequality for scalar unbiased estimators)

Proof is not on the book BUT, it is on the lecture notes

Suppose the regularity conditions discussed above hold.

If $Y = u(X_1, X_2, \ldots, X_n)$ *is an unbiased estimator of* θ, *then*

$$\boxed{\mathrm{Var}(Y) \geq \frac{1}{I(\theta)}},$$

where $I(\theta)$, *called the Fisher information, is defined as*

$$\boxed{I(\theta) = nE\left[\left(\frac{\partial}{\partial\theta}\ln(f(X;\theta))\right)^2\right]},$$

Condition 4 says that $I(\theta) > 0$ AND $I(\theta) < \infty$

with the expected value taken with respect to the pdf/pmf of X.

INTERPRETATION:

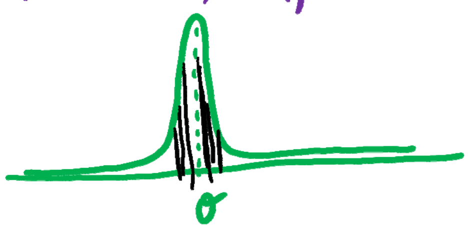
1. There is no unbiased estimator of θ with variance lower than $1/I(\theta)$
2. If we have found an estimator Y with variance equal to $1/I(\theta)$, then this is the unbiased estimator of θ with lowest variance.

What is the meaning of the Fisher information

$$I(\theta) = n\,E\left[\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2\right]$$

$I(\theta)$ measures the amount of information that the r.v. X can give about θ

For instance, if θ is the mean of X and $f(x,\theta)$ is very concentrated about θ

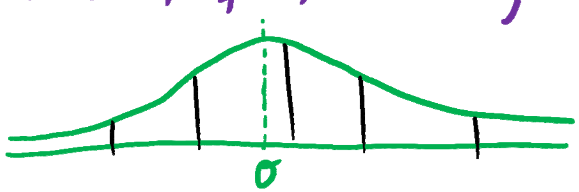


then $I(\theta)$ is large

(because we don't need as many samples to infer a good estimate for θ)

If instead, $f(x;\theta)$ is very wide, then $I(\theta)$ is small

($f(x,\theta)$ has relatively less information about θ)



Simplified formula for $I(\theta) = n\, E\left[\left(\frac{\partial}{\partial\theta} \ln f(X,\theta)\right)^2\right]$ (using an additional assumption)

Continuous-variable case:

First note that: $E\left[\frac{\partial}{\partial\theta} \ln f(X,\theta)\right] = 0$

$E[u(X)] = \int u(x) \cdot f(x)\, dx$

Proof: $E\left[\frac{\partial}{\partial\theta} \ln f(X,\theta)\right] = \int_S \frac{\partial}{\partial\theta} \ln f(x,\theta) \cdot f(x,\theta)\, dx$

$$= \int_S \frac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)} \cdot f(x,\theta)\, dx$$

$$= \int_S \frac{\partial}{\partial\theta} f(x,\theta)\, dx = \frac{\partial}{\partial\theta} \underbrace{\int f(x,\theta)\, dx}_{1} =$$

$$= \frac{\partial}{\partial\theta}(1) = 0$$

Since we've seen that

$$\int \frac{\partial}{\partial \theta} \ln f(x,\theta) \cdot f(x,\theta)\, dx = 0$$

if we are allowed to differentiate the integral once more ← ADDITIONAL ASSUMPTION

$$\frac{\partial}{\partial \theta} \int \frac{\partial}{\partial \theta} \ln f(x,\theta) \cdot f(x,\theta)\, dx = \frac{\partial}{\partial \theta} 0$$

assume we can interchange derivative and integral once more

$$\int \frac{\partial}{\partial \theta} \left[\frac{\partial}{\partial \theta} \ln f(x,\theta) \cdot f(x,\theta) \right] dx = 0$$

Use product rule for derivative inside integral:

$$\int \frac{\partial^2}{\partial\theta^2}\left[\ln f(x,\theta)\right] \cdot f(x,\theta) + \underbrace{\frac{\partial}{\partial\theta}\ln f(x,\theta) \cdot \frac{\partial}{\partial\theta} f(x,\theta)}_{\frac{\partial}{\partial\theta}\ln f(x,\theta) \cdot f(x,\theta)} \, dx = 0$$

NOTE: since $\frac{\partial}{\partial\theta}\ln f(x,\theta) = \dfrac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)}$, then $\frac{\partial}{\partial\theta} f(x,\theta) = \left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right) \cdot f(x,\theta)$

$$\int \left\{ \frac{\partial^2}{\partial\theta^2}\left[\ln f(x,\theta)\right] + \left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2 \right\} f(x,\theta)\, dx = 0$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta) + \left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2\right] = 0$$

linearity of expected value

$$E\left[\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta)\right]$$

CONCLUSION:

$$I(\theta) = n\,E\left[\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2\right] = -n\,E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta)\right]$$

ANOTHER CONSEQUENCE

$$I(\theta) = n\left\{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right] - \Big(\underbrace{E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right]}_{0}\Big)^2\right\}$$

$$= n\,\mathrm{Var}\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)$$

Remark (proved in the previous slides!)

Under additional regularity assumptions involving the existence of second derivatives and the validity of interchanging the order of certain differentiations and integrations, it holds that:

$$E\left[\left(\frac{\partial}{\partial\theta}\ln(f(X;\theta))\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln(f(X;\theta))\right]$$

Note: the identity above is useful whenever the expectation on the right hand side is easier to evaluate than the one on the left hand side (appearing on the definition of the Fisher information $I(\theta)$).

Definition (Efficiency)

Let $Y = u(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ. The estimator *efficiency* is defined as

$$e(Y) = \frac{I(\theta)^{-1}}{\mathrm{var}(Y)}$$

Cramer-Rao lower bound

Notes:

1) The efficiency of an unbiased estimator $\hat{\theta}$ measures how close this estimator's variance comes to the Cramér-Rao lower bound.

2) The Cramér-Rao lower bound gives

$$e(Y) \leq 1 \ .$$

also $e(Y) \geq 0$

3) If the variance of an unbiased estimator equals the Cramér-Rao lower bound (i.e. its efficiency is 1), it is called a *minimum-variance unbiased estimator*.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution

does not depend on θ

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad \underbrace{0 < x < \infty}, \quad \theta \in \Omega = \{\theta : 0 < \theta < \infty\} .$$

We have seen previously that the maximum likelihood estimator of θ, $\boxed{\hat{\theta} = \bar{X}}$, is unbiased. ← earlier example from Sec. 6.4

1. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ for a general n. ← sample size
2. What is the efficiency of $\hat{\theta}$?

(1) Recall that the Cramér-Rao inequality states that

$$\boxed{\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}}$$

lower bound!

where the Fisher information $I(\theta)$ is given by

$$I(\theta) = n\, E\left[\left(\frac{\partial}{\partial\theta} \ln f(X,\theta)\right)^2\right]$$

We know that (under mild assumptions which are satisfied in this case):

$$E\left[\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta)\right]$$

Let us evaluate the expectation on the RHS:

$$\ln f(x,\theta) = \ln\left(\frac{1}{\theta}e^{-x/\theta}\right) = -\ln\theta - \frac{x}{\theta}$$

$$\frac{\partial}{\partial\theta}\ln f(x,\theta) = -\frac{1}{\theta} + \frac{x}{\theta^2} \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2}\ln f(x,\theta) = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

Then

$$E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X,\theta)\right] = E\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right] \overset{\text{linearity}}{=} \frac{1}{\theta^2} - \frac{2}{\theta^3}\overbrace{E[X]}^{\theta} =$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^3} \cdot \theta = -\frac{1}{\theta^2}$$

Then, we conclude that

$$I(\theta) = -n\, E\left[\frac{\partial^2}{\partial \theta^2} \ln f(x,\theta) \right] = (-n) \cdot \left(-\frac{1}{\theta^2} \right) = \frac{n}{\theta^2}$$

The Cramer-Rao lower bound for the variance of an unbiased estimator is

$$\frac{1}{I(\theta)} = \frac{1}{n/\theta^2} = \frac{\theta^2}{n}$$

this is the smallest possible value for the variance of an unbiased estimator of θ

(2) we want to compute the efficiency:

$$e(\hat{\theta}) = \frac{1/I(\theta)}{Var(\hat{\theta})}$$ ← determined already on item 1.

Note that

$$Var(\hat{\theta}) = Var(\bar{X}) = Var\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) \overset{X_1,\dots,X_m \text{ independent}}{=} \frac{1}{m^2}\sum_{i=1}^{n} \underbrace{Var(X_i)}_{\theta^2}$$ ← we formula sheet

$$= \frac{1}{m^2}\underbrace{\sum_{i=1}^{m}\theta^2}_{m\theta^2} = \frac{1}{m^2}\cdot m\cdot\theta^2 = \frac{\theta^2}{m}$$

← lowest possible variance!!!

CONCLUSION: $\hat{\theta} = \bar{X}$ is a minimum-variance unbiased estimator of θ

and $$e(\hat{\theta}) = \frac{1/I(\theta)}{Var(\hat{\theta})} = \frac{\theta^2/m}{\theta^2/m} = 1$$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution

pmf

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \quad , \quad \lambda \in \Omega = \{\lambda : 0 < \lambda < \infty\} .$$

support does not depend on λ

~~We have seen previously that the maximum likelihood estimator of θ, $\hat{\theta} = \bar{X}$, is unbiased.~~

1. Determine the maximum likelihood estimator of λ.
2. Show that $\hat{\lambda}$ is unbiased. $\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sec. 6.4 content

Do items 1+2 as HW (solutions are on lecture notes)

3. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of λ for a general n.
4. Determine the efficiency of $\hat{\lambda}$.

(3) Recall that the Cramér-Rao lower bound is:

$$\mathrm{Var}(\hat{\lambda}) \geq \frac{1}{I(\lambda)}$$

where $I(\lambda)$ is the Fisher information

$$I(\lambda) = n\,E\left[\left(\frac{\partial}{\partial\lambda}\ln f(X,\lambda)\right)^2\right] = -n\,E\left[\frac{\partial^2}{\partial\lambda^2}\ln f(X,\lambda)\right]$$

Since:

$$\ln f(x,\lambda) = \ln\left(\frac{\lambda^x e^{-\lambda}}{x!}\right) = x\ln\lambda - \lambda - \ln(x!)$$

$$\frac{\partial}{\partial\lambda}\ln f(x,\lambda) = \frac{x}{\lambda} - 1 \quad \text{and} \quad \frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda) = -\frac{x}{\lambda^2}$$

$$E\left[\frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda)\right] = E\left[-\frac{X}{\lambda^2}\right] \underset{\text{linearity}}{=} -\frac{1}{\lambda^2}\underbrace{E[X]}_{\lambda} = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

Then

$$I(\lambda) = -n\,E\left[\frac{\partial^2}{\partial\lambda^2}\ln f(x,\theta)\right] = -n\cdot\left(-\frac{1}{\lambda}\right) = \frac{n}{\lambda}$$

and so, the Cramer - Rao lower bound for an unbiased estimator of λ is

$$\frac{1}{I(\lambda)} = \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

(4) To find the efficiency of $\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

we need to find $Var(\hat{\lambda})$

We have:

$$Var(\hat{\lambda}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{X_1, \ldots, X_n \text{ independent}} = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var(X_i)}_{\lambda} = \frac{1}{n^2}\cdot n\lambda = \frac{\lambda}{n}$$

Since $Var(\hat{\lambda}) = \dfrac{1}{I(\lambda)} = 1$, then:

a) $\hat{\lambda}$ is a minimum variance unbiased estimate of λ

b) $e(\hat{\lambda}) = \dfrac{1/I(\lambda)}{Var(\hat{\lambda})} = 1$

Concluding remarks:

- In general, the Cramér-Rao lower bound is not an attainable lower bound:
  - there often exists a lower bound for variance that is greater than the Cramér-Rao lower bound.
  - the two previous examples were too optimistic...

- Roughly, an unbiased estimator whose variance coincides with the Cramér-Rao lower bound exists if and only if sampling is from a member of the exponential class of distributions (yet to be introduced).
  - Some elements of the exponential class of distributions:
    - Normal; Gamma (exponential and chi-square are particular cases)
    - Bernoulli; Poisson; Geometric