Estimation Theory

Fredrik Rusek

Introduction

Chapters 1-3.4

Schedule, tentative

- Lectures: Tuesdays 10-12 in E:3139
  - Easter break?
  - 27/4 cancelled

- Seminars: (about) every second Monday, 10-12 in E:3139. 1st on 9/3
  - Purpose is to discuss homework problems
  - Recommended to go through the problems in advance

Contents

- Chapters 1-13 in the book (+ some extra minor material)
- I don’t know how many lectures that we need, depends on the amount of discussion at the lectures
- 8-12 lectures is a good guess
- My plan is to finish before or around midsummer

Examination

- Examination is done via hand in assignements.
  - One set per chapter in the book
  - Problems are discussed at seminars

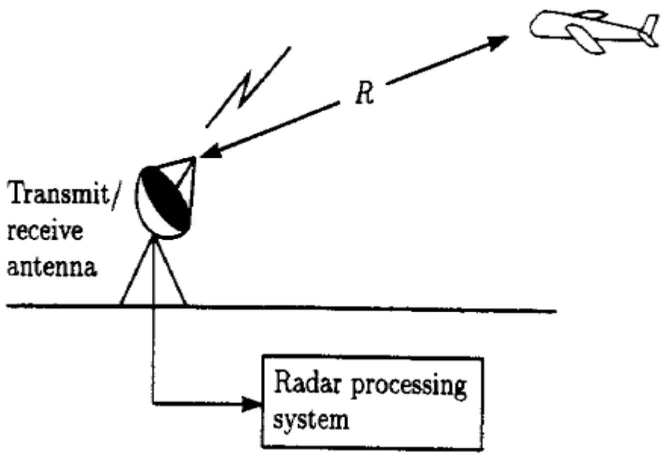
- Reqs for passing degree
  - 80% attendance at lectures
  - 80% attendance at seminars
  - Hand in all home-assignements

- 9 ECTS

Estimation vs. Detection theory

In estimation theory, we try to estimate the value of a continuous variable

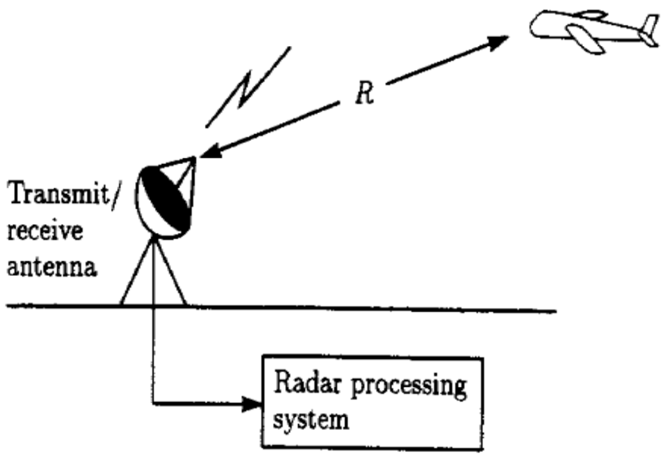
In this example, we try to estimate the distance to the airplane ***R***



Estimation vs. Detection theory

In estimation theory, we try to estimate the value of a continuous variable

In this example, we try to estimate the distance to the airplane ***R***



If it was a detection problem, we would have tried to estimate the ***presence*** of an airplane (0/1)

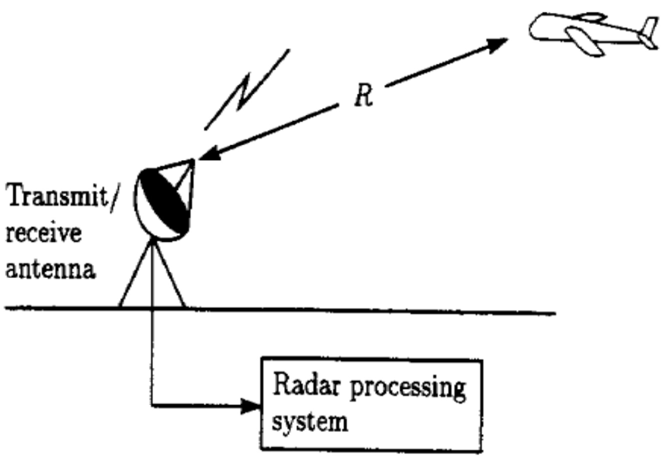
In Detection Theory, we try to estimate the value of a discrete variable

Estimation vs. Detection theory

Simply put,

In detection theory, we are either right or wrong

In estimation theory, we are always wrong



Estimation vs. Detection theory

Remark: When the cardinality of the set is large, discrete problems are usually classified as estimation theory problems

Two famous examples:

- *German tank problem*
  - ***Estimate number of produced german tanks per month based on the number of tanks you observe at battlefield***
- *Doomsday problem*
  - ***Estimate how many more humans that will be born in the future from the number of humans born so far (around 10^{11})***

Some words about the book

- Simple to read
- For engineers, not for mathematicians
- More than half of the book is examples
- Many examples are used throughout the book
- Almost no proofs
- 99% based on discrete time. The reader is assumed to be able to convert continuous time to discrete on his own. Possibly, we will add one lecture at the end dealing with this shortcoming
- First part is assuming deterministic parameters to estimate, second part is assuming random parameters

Some words about the book

- Simple to read
- For engineers, not for mathemticians
- More than half of the book is examples
- Many examples are used throughout the book
- Almost no proofs
- 99% based on discrete time. The reader is assumed to be able to convert continuous time to discrete on his own. Possibly, we will add one lecture at the end dealing with this shortcoming
- First part is assuming **deterministic parameters** to estimate, second part is assuming random parameters

May be confusing if you have background on the subject, for example knowledge of the MMSE estimator for MIMO communications

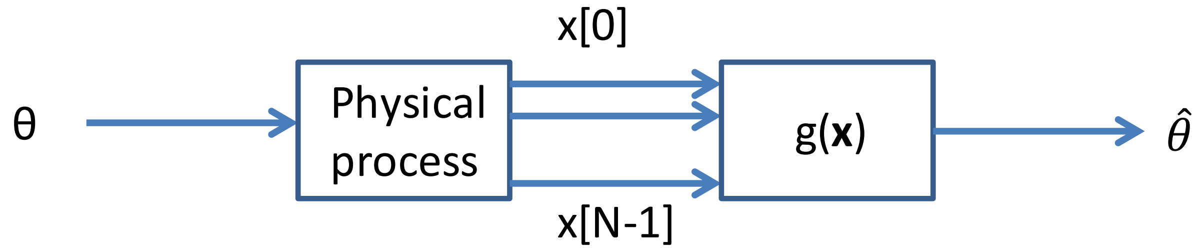
Chapter 1

Mathematical formulation of the problem we'll study

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

- x[n] is a sequence of, possibly dependent, observations. These observations carry information about a parameter θ that we would like to estimate

- We do this by constructing a (deterministic) function g(**x**) that produces an estimate of θ



Chapter 1

Likelihood functions vs conditional probabilities

- The data **x**=x[0]…x[N-1] is of course dependent on the parameter we would like to estimate, θ, in some way

- We denote by p(**x**;θ) a family of PDFs parameterized by θ. In words, *"This is the pdf that **x** will abide if the the unknown parameter is θ"*

- **Note that θ is, on the most basic level, not a random variable**

- If θ was the realization of a random process, then we have the conditional pdf p(**x**|θ)

Chapter 1

Likelihood functions vs conditional probabilities

- The data **x**=x[0]…x[N-1] is of course dependent on the parameter we would like to estimate, θ, in some way

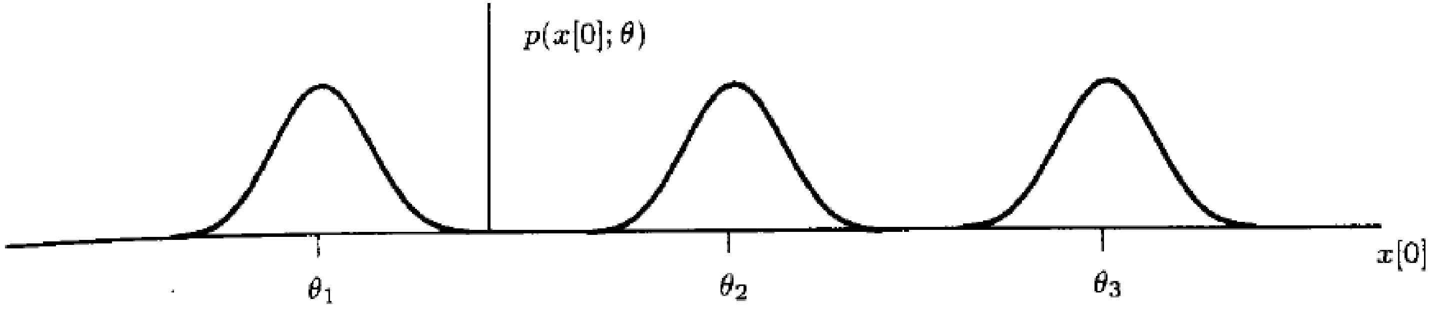
- We denote by p(**x**;θ) a family of PDFs parameterized by θ. In words, *"This is the pdf that **x** will abide if the the unknown parameter is θ"*

- **Note that θ is, on the most basic level, not a random variable**

- If θ was the realization of a random process, then we have the conditional pdf p(**x**|θ)
- **Note that a likelihood and a conditional pdf have the same formulas, it is only the interpretation of them that differ**

Chapter 1

Likelihood functions vs conditional probabilities



Given the functional form of the family of likelihoods, $p(x[0];\theta)$, we can infer the value of θ from an observation $x[0]$

For example, if $x[0]<0$, then it is unlikely that $\theta = \theta_2$

Chapter 2 – Minimum variance unbiased estimators

Unbiased estimators

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

Recall that the estimator is

- A function only of **x**
- Random, since **x** is random

Since $\hat{\theta}$ is random, it has an expectation E($\hat{\theta}$).

The estimator is unbiased if

$$\mathrm{E}\left(\hat{\theta}\right) = \theta, \quad \forall\theta$$

Chapter 2 – Minimum variance unbiased estimators

Unbiased estimators

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

Recall that the estimator is

- A function only of **x**
- Random, since **x** is random

Since $\hat{\theta}$ is random, it has an expectation E($\hat{\theta}$).

The estimator is unbiased if

$$\mathrm{E}(\hat{\theta}) = \theta, \quad \forall\theta$$

$$E(\hat{\theta}) = \int g(\mathbf{x})p(\mathbf{x};\theta)\,d\mathbf{x} = \theta \qquad \text{for all } \theta.$$

Chapter 2 – Minimum variance unbiased estimators

Unbiased estimators

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

Recall that the estimator is

- A function only of **x**
- Random, since **x** is random

Since $\hat{\theta}$ is random, it has an expectation E($\hat{\theta}$).

The estimator is unbiased if

$$\mathrm{E}\left(\hat{\theta}\right) = \theta, \quad \forall\theta$$

It also has a variance

$$var\left(\hat{\theta}\right) = E(\hat{\theta} - \theta)^2$$

Chapter 2 – Minimum variance unbiased estimators

Unbiased estimators

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

Recall that the estimator is

- A function only of **x**
- Random, since **x** is random

Since $\hat{\theta}$ is random, it has an expectation E($\hat{\theta}$).

The estimator is unbiased if

$$\mathrm{E}(\hat{\theta}) = \theta, \quad \forall\theta$$

It also has a variance

$$var(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

Note: The variance of the estimator depends on θ

Chapter 2 – Minimum variance unbiased estimators

Quality of estimator

A natural criterion may be the mean square error

$$\mathrm{mse}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

Chapter 2 – Minimum variance unbiased estimators

Quality of estimator

A natural criterion may be the mean square error

$$\mathrm{mse}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

However,

$$\begin{aligned}\mathrm{mse}(\hat{\theta}) &= E\left\{\left[\left(\hat{\theta} - E(\hat{\theta})\right) + \left(E(\hat{\theta}) - \theta\right)\right]^2\right\} \\ &= \mathrm{var}(\hat{\theta}) + \left[E(\hat{\theta}) - \theta\right]^2 \\ &= \mathrm{var}(\hat{\theta}) + b^2(\theta)\end{aligned}$$

where *b(*θ*)* is the *bias:* $b(\theta) = E(\hat{\theta}) - \theta$

Chapter 2 – Minimum variance unbiased estimators

Quality of estimator

A natural criterion may be the mean square error

$$\text{mse}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

However,

$$\begin{aligned}\text{mse}(\hat{\theta}) &= E\left\{\left[\left(\hat{\theta} - E(\hat{\theta})\right) + \left(E(\hat{\theta}) - \theta\right)\right]^2\right\} \\ &= \text{var}(\hat{\theta}) + \left[E(\hat{\theta}) - \theta\right]^2 \quad +2E([\hat{\theta} - E(\hat{\theta})]b(\theta)) \\ &= \text{var}(\hat{\theta}) + b^2(\theta)\end{aligned}$$

=0

where *b(*θ*)* is the *bias:* $b(\theta) = E(\hat{\theta}) - \theta$

Chapter 2 – Minimum variance unbiased estimators

The dependence of the bias on θ is ***bad news***, as is shown with an example next

Chapter 2 – Minimum variance unbiased estimators

Example 2.1 - Unbiased Estimator for DC Level in White Gaussian Noise

$$x[n] = A + w[n] \qquad n = 0, 1, \ldots, N-1$$

– A: DC level to be estimated

– w[n], zero mean white Gaussian noise

Proposed estimator: $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ **(proven optimal later)**

Unbiased

$$E(\hat{A}) = E\left[\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right] = A$$

Variance

$$\mathrm{var}(\hat{A}) = \mathrm{var}\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{var}(x[n]) = \frac{\sigma^2}{N}$$

Chapter 2 – Minimum variance unbiased estimators

Example 2.1 - Unbiased Estimator for DC Level in White Gaussian Noise

$$x[n] = A + w[n] \qquad n = 0, 1, \ldots, N-1$$

- A: DC level to be estimated
- w[n], zero mean white Gaussian noise

Modified estimator: $$\check{A} = a\frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

bias $$(a-1)A$$

$$\text{mse}(\check{A}) = \text{var}(\hat{\theta}) + b^2(\theta) = \frac{a^2\sigma^2}{N} + (a-1)^2A^2.$$

Chapter 2 – Minimum variance unbiased estimators

Example 2.1 - Unbiased Estimator for DC Level in White Gaussian Noise

$$x[n] = A + w[n] \qquad n = 0, 1, \ldots, N-1$$

- A: DC level to be estimated
- w[n], zero mean white Gaussian noise

Modified estimator: $$\check{A} = a\frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

$$\frac{d\,\mathrm{mse}(\check{A})}{da} = \frac{2a\sigma^2}{N} + 2(a-1)A^2 \qquad a_{\mathrm{opt}} = \frac{A^2}{A^2 + \sigma^2/N}$$

$$\mathrm{mse}(\check{A}) = \mathrm{var}(\hat{\theta}) + b^2(\theta) = \frac{a^2\sigma^2}{N} + (a-1)^2A^2.$$

The optimal estimator depends on A -> Not realizable!!

Chapter 2 – Minimum variance unbiased estimators

Summary

We seek a function g(**x**) that estimates a parameter θ well

Chapter 2 – Minimum variance unbiased estimators

Summary

We seek a function g(**x**) that estimates a parameter θ well

- The natural performance metric is mse $\text{mse}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$
  - But this does not work as the bias term b(θ) makes the optimal function g(**x**) dependent on θ

- Restrict to the class of unbiased estimators, i.e.,

$$E(\hat{\theta}) = \int g(\mathbf{x}) p(\mathbf{x}; \theta)\, d\mathbf{x} = \theta \qquad \text{for all } \theta.$$

  - The MSE is now the variance $\text{var}(\hat{\theta})$ of the estimator, which is taken as performance metric

Chapter 2 – Minimum variance unbiased estimators

Summary

We seek a function g(**x**) that estimates a parameter θ well

- The natural performance metric is mse $\text{mse}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$
  - But this does not work as the bias term b(θ) makes the optimal function g(**x**) dependent on θ

- Restrict to the class of unbiased estimators, i.e.,

$$E(\hat{\theta}) = \int g(\mathbf{x})p(\mathbf{x};\theta)\,d\mathbf{x} = \theta \qquad \text{for all } \theta.$$

  - The MSE is now the variance $\text{var}(\hat{\theta})$ of the estimator, which is taken as performance metric

*GOAL: Find unbiased g(**x**) with as small* $\text{var}(\hat{\theta})$ *as possible*

Chapter 2 – Minimum variance unbiased estimators

Check point

It is easy to find g(**x**) with a small variance, $\text{var}(\hat{\theta})$, for example g(**x**)=0 has zero variance

Chapter 2 – Minimum variance unbiased estimators

Check point

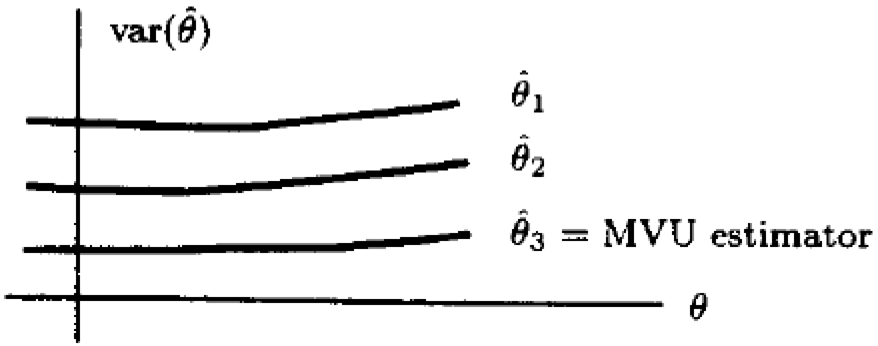
It is easy to find g(**x**) with a small variance, $\mathrm{var}(\hat{\theta})$, for example g(**x**)=0 has zero variance

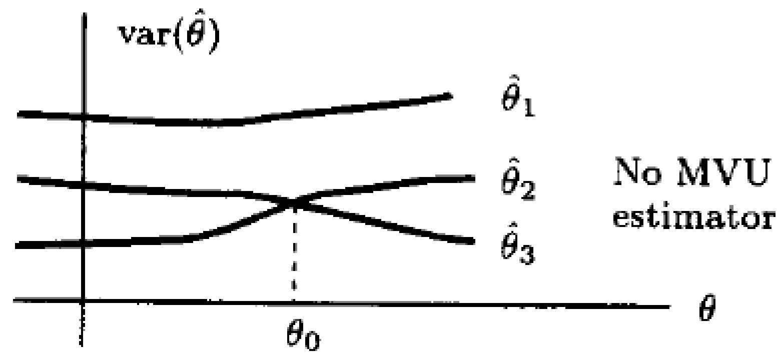
But g(**x**)=0 is not unbiased

Chapter 2 – Minimum variance unbiased estimators

Minimum variance unbiased estimator (MVU)

Recall, the variance of the estimator, $\mathrm{var}(\hat{\theta})$, depends on θ, and to be *the* optimal estimator, it must provide the smallest variance for *all* θ

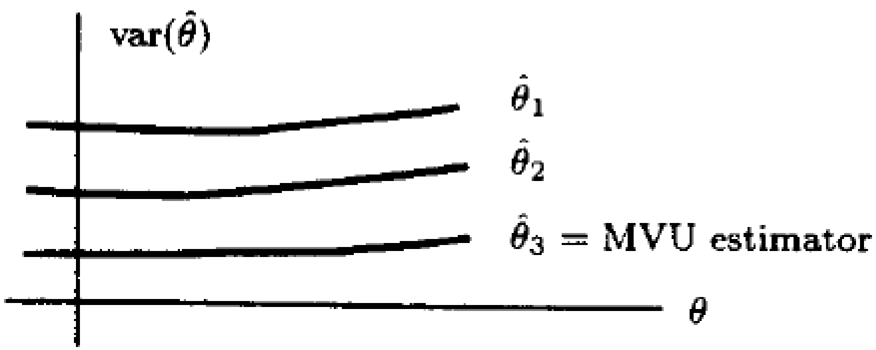


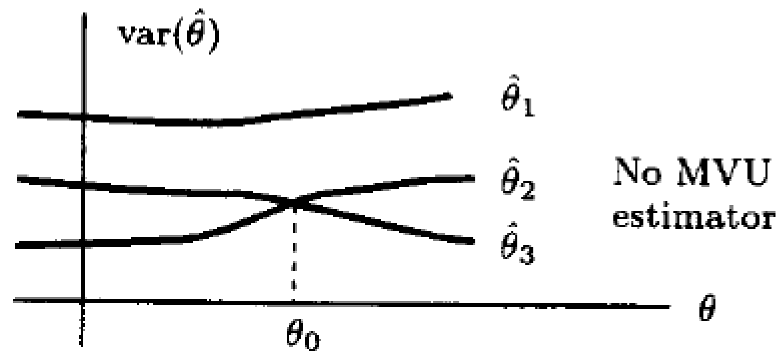


Chapter 2 – Minimum variance unbiased estimators

Minimum variance unbiased estimator (MVU)

Recall, the variance of the estimator, $\text{var}(\hat{\theta})$, depends on θ, and to be *the* optimal estimator, it must provide the smallest variance for *all* θ





The MVU exists only in some cases, see example 2.3 for a case where it does not exist
Sometimes, there is not even any unbiased estimator at all

Chapter 2 – Minimum variance unbiased estimators

Unbiasedness for Vector parameters $\boldsymbol{\theta} = [\theta_1\, \theta_2 \ldots \theta_p]^T$

$$E(\hat{\theta}_i) = \theta_i \qquad \text{or} \qquad E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$$

MVU for Vector parameters

$\text{var}(\hat{\theta}_i)$ for $i = 1, 2, \ldots, p$ is minimum among all unbiased estimators.

Chapter 3 – Cramer-Rao lower bound

The CRLB is useful in the following ways:

- It provides a lower bound on the variance of any unbiased estimator
- If an MVU exists, the function g(**x**) will fall out as a side result

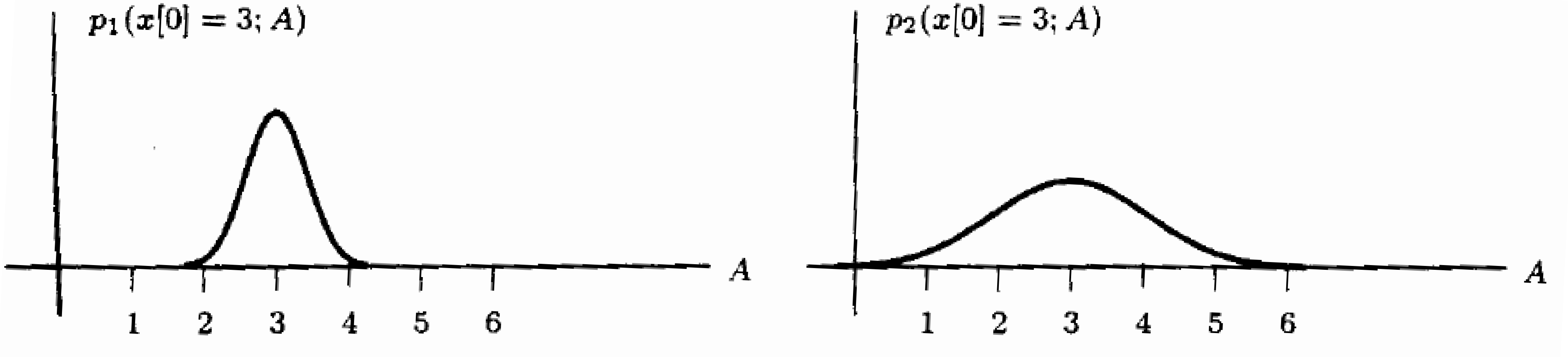
Some properties of CRLB

- Derived for unbiased estimators (can easily be extended to biased, but not in this book)
- The bound is not always reachable (it is only a lower bound)
- A regularity condition must hold, so the CRLB cannot always be applied
- Better bounds exist (more about this later)
- Cramér and Rao proved it independently in the mid 40s
- French mathematician Frechét proved it earlier, but never published

Chapter 3 – Cramer-Rao lower bound

WARNING Some hand waving in these arguments

Basic idea of CRLB:



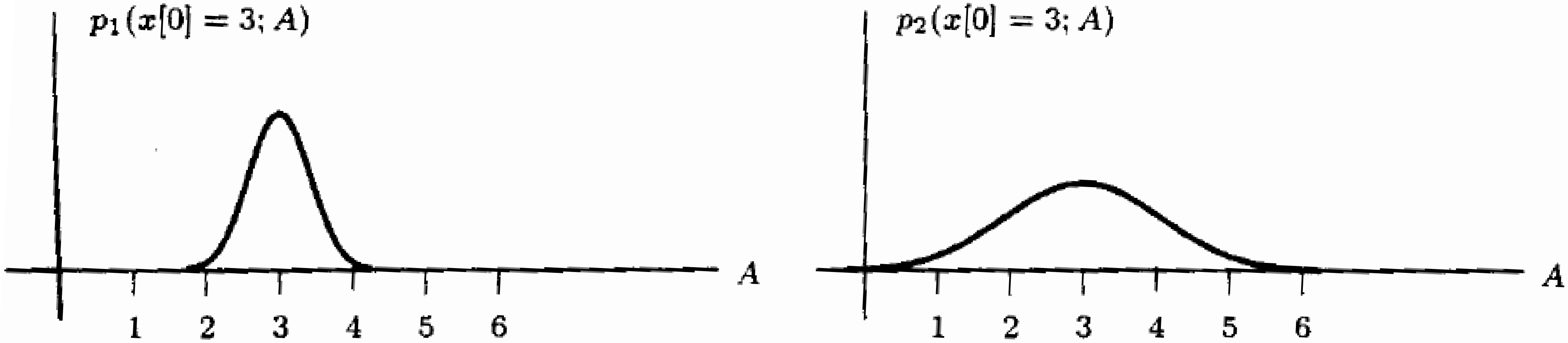
(a) $\sigma_1 = 1/3$ (b) $\sigma_2 = 1$

For case (a): the value A=5 is not very likely
For case (b): A=5 cannot be ruled out

Chapter 3 – Cramer-Rao lower bound

Some hand waving in these arguments

Basic idea of CRLB:

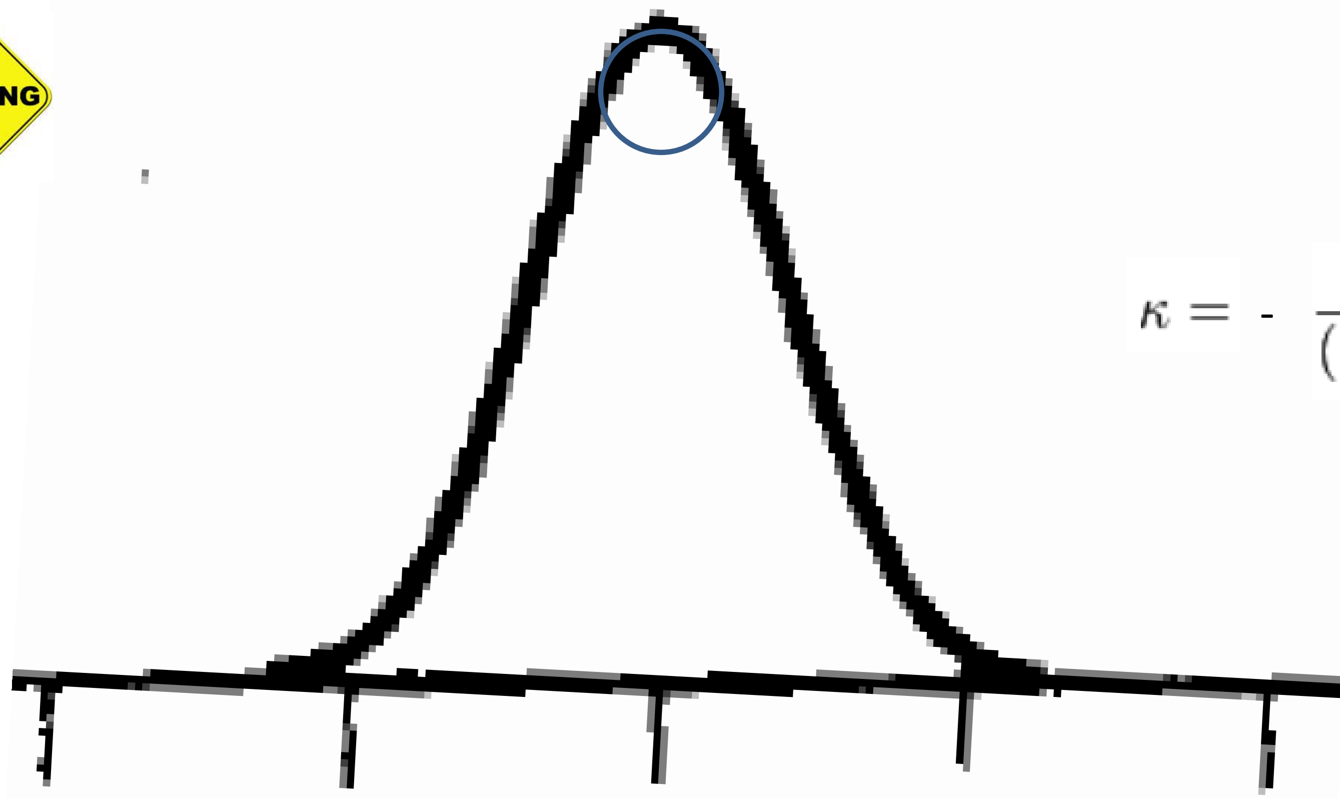


(a) $\sigma_1 = 1/3$ **(b)** $\sigma_2 = 1$

For case (a): the value A=5 is not very likely
For case (b): A=5 cannot be ruled out

In fact, the *sharpness* around x[0]=3 is important
If p(x[0];A) is very sharp around A=x[0], then we can expect to be able to infer the value of A with a much better precision

Chapter 3 – Cramer-Rao lower bound

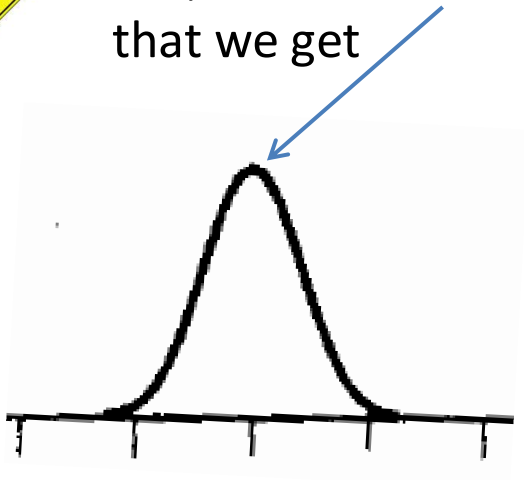


$$\kappa = - \frac{|y''|}{(1+y'^2)^{3/2}}$$

In mathematics, ”sharpness” has the name curvature
The curvature is measured by the reciprocal of the radius of the circle

Chapter 3 – Cramer-Rao lower bound

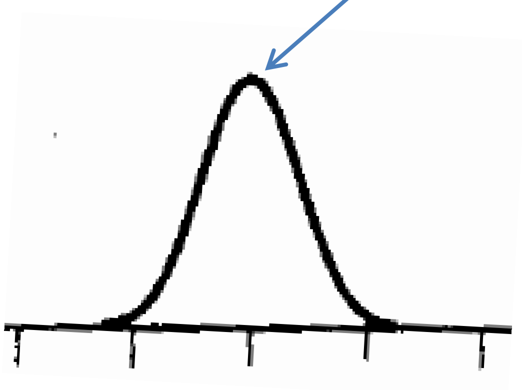
But, at the correct value, an unbiased estimator has 0 slope, so that we get



$$\kappa = -|y''|$$

Chapter 3 – Cramer-Rao lower bound

But, at the correct value, an unbiased estimator has 0 slope, so that we get



$$\kappa = -|y''|$$

Further, multiplicative constants should not affect the precision, e.g., $-x^2$ should be as good as the likelihood $-2x^2$

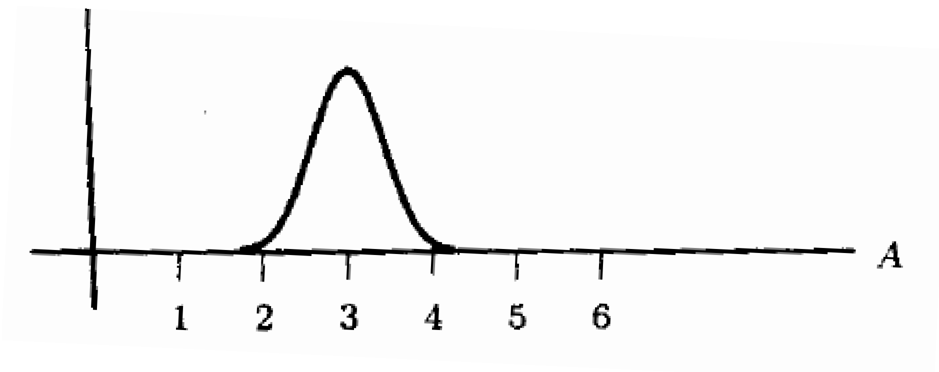
Taking the log of the likelihood removes these constants. Hence, the "sharpness" can be effectively measured by

$$-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}$$

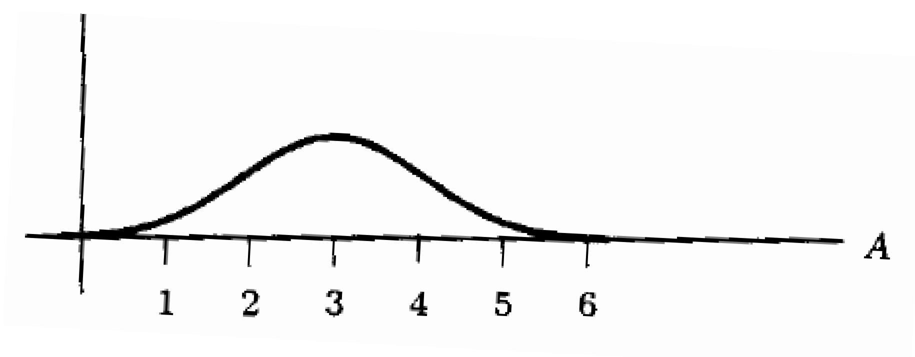
Chapter 3 – Cramer-Rao lower bound

If the ”sharpness” of the log-likelihood function Is very big, we have a plot according to

$$-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}$$

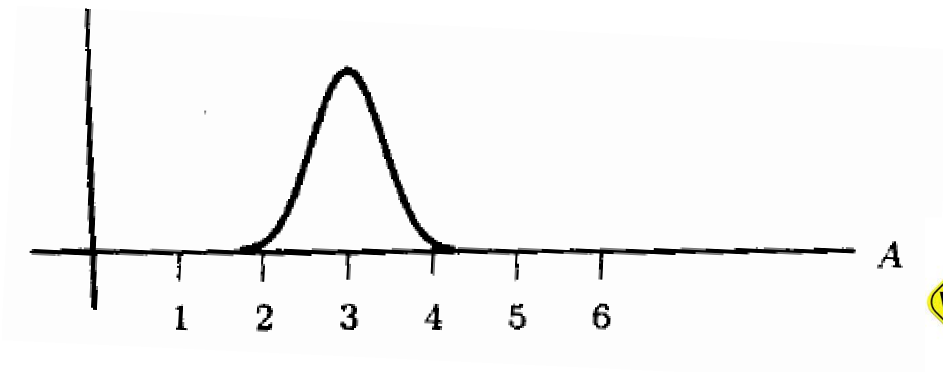


while we get the below plot if it is small

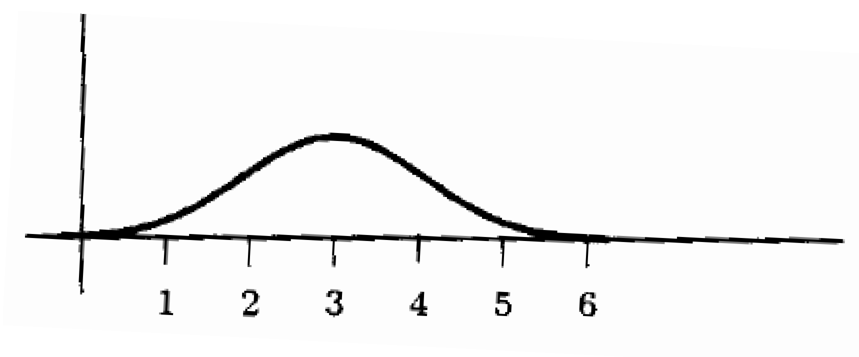


Chapter 3 – Cramer-Rao lower bound

If the ”sharpness” of the log-likelihood function $-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}$ Is very big, we have a plot according to



while we get the below plot if it is small

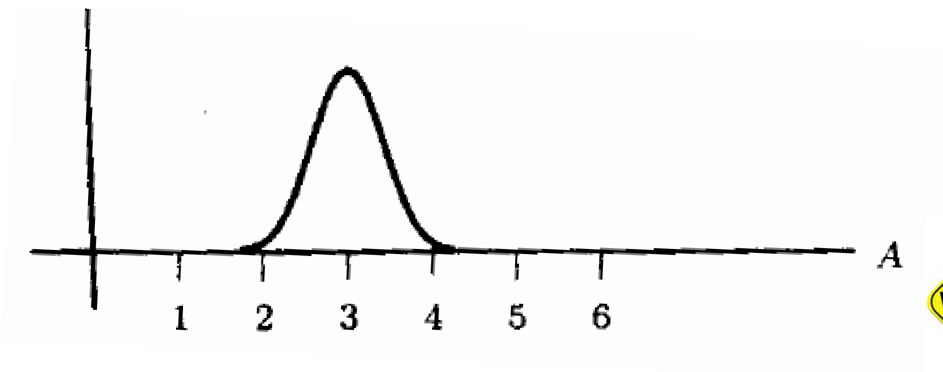


Clearly, we expect much better estimation precision for the first case than for the second, which ”proves” that the curvature is most relevant for determining bounds on the variance

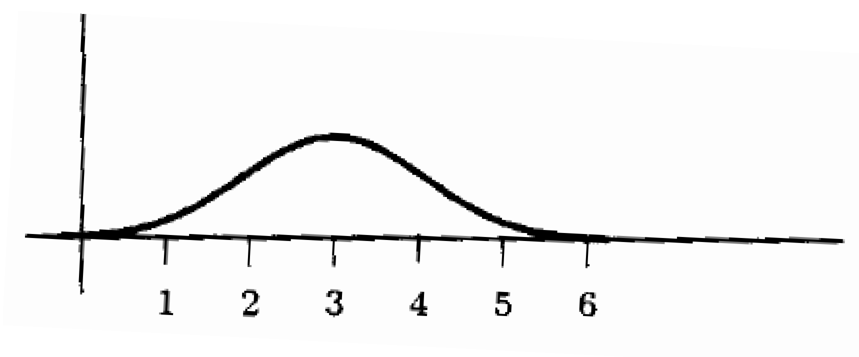
Chapter 3 – Cramer-Rao lower bound

If the "sharpness" of the log-likelihood function
Is very big, we have a plot according to

$$-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}$$



while we get the below plot if it is small

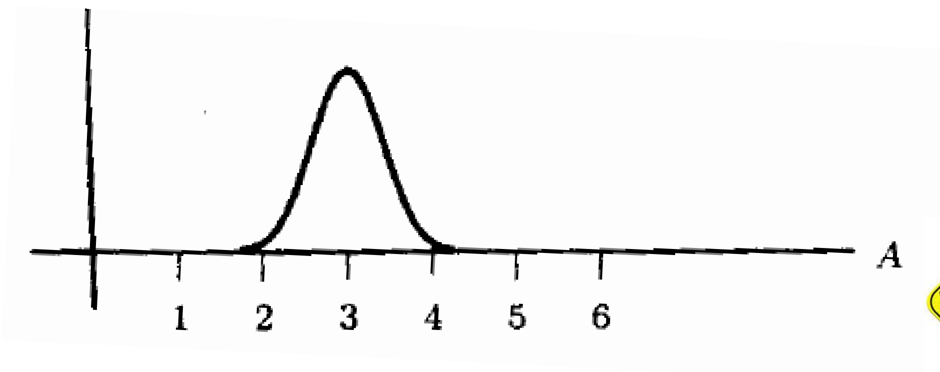


Guess:
"Precision = 1/sharpness"

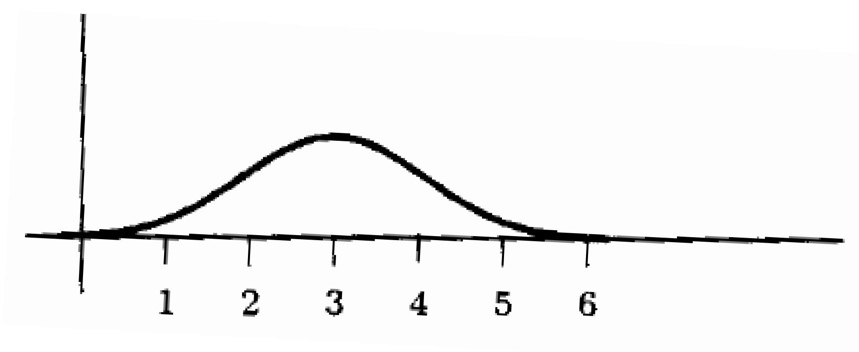
Chapter 3 – Cramer-Rao lower bound

If the ”sharpness” of the log-likelihood function Is very big, we have a plot according to

$$-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}$$



while we get the below plot if it is small



Guess:
”Precision = 1/sharpness”

We next discuss an example where the guess is correct.
Then we prove it by the Cramer-Rao theorem

Chapter 3 – Cramer-Rao lower bound

Example 3.1 - PDF Dependence on Unknown Parameter

$$x[0] = A + w[0]$$ Estimate A

$$w[0] \sim \mathcal{N}(0, \sigma^2)$$

$$\hat{A} = x[0] \qquad \operatorname{var}(\hat{A}) = \sigma^2$$

Likelihood $$p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0] - A)^2\right]$$

Chapter 3 – Cramer-Rao lower bound

Example 3.1 - PDF Dependence on Unknown Parameter

$x[0] = A + w[0]$ **Estimate A**

$w[0] \sim \mathcal{N}(0, \sigma^2)$

$\hat{A} = x[0]$ $\quad\quad \mathrm{var}(\hat{A}) = \sigma^2$

Likelihood $p\,(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0] - A)^2\right]$

Find curvature: $\ln p(x[0]; A) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x[0] - A)^2$

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

$$-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}$$

Chapter 3 – Cramer-Rao lower bound

Example 3.1 - PDF Dependence on Unknown Parameter

$x[0] = A + w[0]$ Estimate A

$w[0] \sim \mathcal{N}(0, \sigma^2)$

$\hat{A} = x[0]$ $\quad \mathrm{var}(\hat{A}) = \sigma^2$

Likelihood $p\,(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0] - A)^2\right]$

Find curvature:

$$\ln p(x[0]; A) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x[0] - A)^2$$

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

$$-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}$$

So…

$$\mathrm{var}(\hat{A}) = \frac{1}{-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}}$$

The guess is correct in this case!!!

Chapter 3 – Cramer-Rao lower bound

Interlude

In the previous example, the second derivative did not depend on x[0], but it will in general

$$-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2} = \frac{1}{\sigma^2}$$

Curvature is a measure of precision, but we cannot have measure of precision that depends on the particular realization of **x**

Solution: Find expected curvature

$$-E\left[\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}\right]$$

This is independent of ***x***

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF* $p(\mathbf{x};\theta)$ *satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0 \qquad \textit{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. *Then, the variance of any unbiased estimator* $\hat{\theta}$ *must satisfy*

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ *and the expectation is taken with respect to* $p(\mathbf{x};\theta)$. *Furthermore, an unbiased estimator may be found that attains the bound for all* θ *if and only if*

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g *and* I. *That estimator, which is the MVU estimator, is* $\hat{\theta} = g(\mathbf{x})$, *and the minimum variance is* $1/I(\theta)$.

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF $p(\mathbf{x};\theta)$ satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0 \qquad \text{for all } \theta$$

Some condition must hold. Must check

where the expectation is taken with respect to $p(\mathbf{x};\theta)$.

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF $p(\mathbf{x};\theta)$ satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \qquad \textit{for all } \theta$$

Some condition must hold. Must check

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

Reciprocal of curvature bounds precision

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ and the expectation is taken with respect to $p(\mathbf{x};\theta)$.

The CRLB *only* takes into account what the likelihood looks like around the correct value

This is fine, since it is a lower bound

A complicated pdf with many peaks will lead to estimators very far away from the CRLB

Examples of better bounds: Weiss-Weinstein, Bobrovsky-Zakai etc etc

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF* $p(\mathbf{x};\theta)$ *satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \qquad \textit{for all } \theta$$

Some condition must hold. Must check

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. *Then, the variance of any unbiased estimator* $\hat{\theta}$ *must satisfy*

Reciprocal of curvature bounds precision

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ *and the expectation is taken with respect to* $p(\mathbf{x};\theta)$. *Furthermore, an unbiased estimator may be found that attains the bound for all* θ *if and only if*

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g *and* I. *That estimator, which is the MVU estimator, is* $\hat{\theta} = g(\mathbf{x})$, *and the minimum variance is* $1/I(\theta)$.

Derivative of log-likelihood produces the estimator

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF* $p(\mathbf{x};\theta)$ *satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \qquad \text{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. *Then, the variance of any unbiased estimator* $\hat{\theta}$ *must satisfy*

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ *and the expectation is taken with respect to* $p(\mathbf{x};\theta)$. *Furthermore, an unbiased estimator may be found that attains the bound for all* θ *if and only if*

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g *and* I. *That estimator, which is the MVU estimator, is* $\hat{\theta} = g(\mathbf{x})$, *and the minimum variance is* $1/I(\theta)$.

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x}$$

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x}$$

$$= \int \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\, d\mathbf{x}$$

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0$

$$
\begin{aligned}
E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] &= \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} p(\mathbf{x};\theta)\,d\mathbf{x} \\
&= \int \frac{\partial p(\mathbf{x};\theta)}{\partial\theta}\,d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\int p(\mathbf{x};\theta)\,d\mathbf{x}
\end{aligned}
$$

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} p(\mathbf{x};\theta)\,d\mathbf{x}$$

$$= \int \frac{\partial p(\mathbf{x};\theta)}{\partial\theta}\,d\mathbf{x}$$

$$= \frac{\partial}{\partial\theta}\int p(\mathbf{x};\theta)\,d\mathbf{x}$$

The regularity condition is in most books that this equality must hold (derivation and integration can be interchanged)

Alternative regularity condition (plus some differentiability conditions that I don't mention)

$B = \{x : p(x;\theta) > 0\}$

Not dependent on θ

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x}$$

$$= \int \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\, d\mathbf{x}$$

$$= \frac{\partial}{\partial \theta} \int p(\mathbf{x};\theta)\, d\mathbf{x}$$

The regularity condition is in most books that this equality must hold (derivation and integration can be interchanged)

Alternative regularity condition (plus some differentiability conditions that I don't mention)

$B = \{x : p(x;\theta) > 0\}$

Not dependent on θ

Example of likelihood that violates the regularity condition:

$$p(x;\theta) = \frac{1}{\theta}, \qquad 0 \le x \le \theta$$

Chapter 3 – Cramer-Rao lower bound

Proof.

Check the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$

$$
\begin{aligned}
E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] &= \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} \\
&= \int \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\, d\mathbf{x} \\
&= \frac{\partial}{\partial \theta} \int p(\mathbf{x};\theta)\, d\mathbf{x} \\
&= \frac{\partial 1}{\partial \theta} \\
&= 0.
\end{aligned}
$$

Chapter 3 – Cramer-Rao lower bound

Proof.(interlude)

Interchanging derivation and integration allows us (at least me) to actually understand the regularity condition

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = \frac{\partial}{\partial \theta} \underbrace{E\left[\ln p(\mathbf{x};\theta)\right]} = 0$$

This is independent of x.

Can be interpreted as the expected likelihood if the parameter of interest has a certain value θ

Chapter 3 – Cramer-Rao lower bound

Proof.(interlude)

Interchanging derivation and integration allows us (at least me) to actually understand the regularity condition

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$$

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = \frac{\partial}{\partial \theta} \; E\left[\ln p(\mathbf{x};\theta)\right] = 0$$

Or, no matter what θ is, you must get the same value of the log-likelihood on average

Chapter 3 – Cramer-Rao lower bound

Proof.(generalization)

The proof in the book is more general than the statement of the theorem. However, the generalization in the proof will be used later in the book

We assume some underlying parameter θ, but that we would like to estimate a function thereof $\alpha=g(\theta)$.

For example, θ can be the DC level, but we are more interested in the power of the DC level, so that $\alpha=\theta^2$, i.e., $g(x)=x^2$

Warning: the book uses g(x) both for the estimator function

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1])$$

and for

$$\alpha=g(\theta)$$

Chapter 3 – Cramer-Rao lower bound

Proof

The unbiased condition now reads

$$E(\hat{\alpha}) = \alpha = g(\theta)$$

or

$$\int \hat{\alpha} p(\mathbf{x};\theta)\, d\mathbf{x} = g(\theta)$$

Take differential (and change order of differentiation and integration)

$$\int \hat{\alpha} \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\, d\mathbf{x} = \int \hat{\alpha} \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial \theta}$$

Chapter 3 – Cramer-Rao lower bound

Proof

The unbiased condition now reads

$$E(\hat{\alpha}) = \alpha = g(\theta)$$

or

$$\int \hat{\alpha} p(\mathbf{x};\theta)\, d\mathbf{x} = g(\theta)$$

Take differential (and change order of differentiation and integration)

$$\int \hat{\alpha} \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\, d\mathbf{x} = \int \hat{\alpha} \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial \theta}$$

Use regularity condition (I added a ”0” term to the l.h.s.)

$$\int (\hat{\alpha} - \alpha) \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial \theta}$$

Chapter 3 – Cramer-Rao lower bound

Proof (Interlude: 3-term Cauchy-Schwarz)

$$\left[\int w(\mathbf{x})g(\mathbf{x})h(\mathbf{x})\,d\mathbf{x}\right]^2 \leq \int w(\mathbf{x})g^2(\mathbf{x})\,d\mathbf{x}\int w(\mathbf{x})h^2(\mathbf{x})\,d\mathbf{x}$$

equality if and only if $g(\mathbf{x}) = ch(\mathbf{x})$ for c some constant not dependent on $\mathbf{x}$

Chapter 3 – Cramer-Rao lower bound

Proof

$$\left[\int w(\mathbf{x})g(\mathbf{x})h(\mathbf{x})\,d\mathbf{x}\right]^2 \leq \int w(\mathbf{x})g^2(\mathbf{x})\,d\mathbf{x}\int w(\mathbf{x})h^2(\mathbf{x})\,d\mathbf{x}$$

equality if and only if $g(\mathbf{x}) = ch(\mathbf{x})$ for c some constant not dependent on $\mathbf{x}$

Compare now with what we had before

$$\int(\hat{\alpha}-\alpha)\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}p(\mathbf{x};\theta)\,d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta}$$

$$\begin{aligned} w(\mathbf{x}) &= p(\mathbf{x};\theta) \\ g(\mathbf{x}) &= \hat{\alpha}-\alpha \\ h(\mathbf{x}) &= \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} \end{aligned}$$

Chapter 3 – Cramer-Rao lower bound

Proof

$$\left[\int w(\mathbf{x})g(\mathbf{x})h(\mathbf{x})\,d\mathbf{x}\right]^2 \leq \int w(\mathbf{x})g^2(\mathbf{x})\,d\mathbf{x}\int w(\mathbf{x})h^2(\mathbf{x})\,d\mathbf{x}$$

equality if and only if $g(\mathbf{x}) = ch(\mathbf{x})$ for c some constant not dependent on $\mathbf{x}$

Compare now with what we had before

$$\int(\hat{\alpha}-\alpha)\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}p(\mathbf{x};\theta)\,d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta}$$

$$\begin{aligned} w(\mathbf{x}) &= p(\mathbf{x};\theta) \\ g(\mathbf{x}) &= \hat{\alpha}-\alpha \\ h(\mathbf{x}) &= \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} \end{aligned}$$

$$\left(\frac{\partial g(\theta)}{\partial\theta}\right)^2 \leq \int(\hat{\alpha}-\alpha)^2 p(\mathbf{x};\theta)\,d\mathbf{x}\int\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2 p(\mathbf{x};\theta)\,d\mathbf{x}$$

Chapter 3 – Cramer-Rao lower bound

Proof

$$\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \leq \int (\hat{\alpha} - \alpha)^2 p(\mathbf{x};\theta)\, d\mathbf{x} \int \left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2 p(\mathbf{x};\theta)\, d\mathbf{x}$$

But this can be written as

$$\mathrm{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]}$$

Chapter 3 – Cramer-Rao lower bound

Proof

$$\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \leq \int (\hat{\alpha}-\alpha)^2 p(\mathbf{x};\theta)\, d\mathbf{x} \int \left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2 p(\mathbf{x};\theta)\, d\mathbf{x}$$

But this can be written as

$$\operatorname{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]}$$

Should be

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

This is not the same as is stated in the theorem, so we are not done yet

Chapter 3 – Cramer-Rao lower bound

Proof

We have

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$$

$$\int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0$$

$$\operatorname{var}(\hat{\alpha}) \geq \frac{\left(\dfrac{\partial g(\theta)}{\partial \theta}\right)^2}{E\left[\left(\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]}$$

Should be

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

Chapter 3 – Cramer-Rao lower bound

Proof

We have

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$$

$$\int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0$$

$$\operatorname{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]}$$

Should be

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

Take one more differential, and change order

$$\frac{\partial}{\partial \theta} \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0$$

$$\int \left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) + \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \frac{\partial p(\mathbf{x};\theta)}{\partial \theta}\right] d\mathbf{x} = 0$$

Chapter 3 – Cramer-Rao lower bound

Proof

Rearranging gives

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right] = \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} p(\mathbf{x};\theta)\,d\mathbf{x} = E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right].$$

$$\operatorname{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial\theta}\right)^2}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right]}$$

Should be

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]$$

Take one more differential, and change order

$$\frac{\partial}{\partial\theta}\int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} p(\mathbf{x};\theta)\,d\mathbf{x} = 0$$

$$\int\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2} p(\mathbf{x};\theta) + \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\frac{\partial p(\mathbf{x};\theta)}{\partial\theta}\right] d\mathbf{x} = 0$$

2nd part done….

Chapter 3 – Cramer-Rao lower bound

Proof

$$\mathrm{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]}$$

If $\alpha = g(\theta) = \theta$, the statement of the theorem follows

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF $p(\mathbf{x};\theta)$ satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0 \qquad \textit{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ and the expectation is taken with respect to $p(\mathbf{x};\theta)$.

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF $p(\mathbf{x};\theta)$ satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \qquad \text{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ and the expectation is taken with respect to $p(\mathbf{x};\theta)$. Furthermore, an unbiased estimator may be found that attains the bound for all θ if and only if

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g and I. That estimator, which is the MVU estimator, is $\hat{\theta} = g(\mathbf{x})$, and the minimum variance is $1/I(\theta)$.

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Lets go back and check the condition for equlity in the C-S inequality

$$\left[\int w(\mathbf{x})g(\mathbf{x})h(\mathbf{x})\,d\mathbf{x}\right]^2 \leq \int w(\mathbf{x})g^2(\mathbf{x})\,d\mathbf{x}\int w(\mathbf{x})h^2(\mathbf{x})\,d\mathbf{x}$$

equality if and only if $g(\mathbf{x}) = ch(\mathbf{x})$ for c some constant not dependent on $\mathbf{x}$

$$\begin{aligned} w(\mathbf{x}) &= p(\mathbf{x};\theta) \\ g(\mathbf{x}) &= \hat{\alpha} - \alpha \\ h(\mathbf{x}) &= \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \end{aligned}$$

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Lets go back and check the condition for equlity in the C-S inequality

$$\left[\int w(\mathbf{x})g(\mathbf{x})h(\mathbf{x})\,d\mathbf{x}\right]^2 \leq \int w(\mathbf{x})g^2(\mathbf{x})\,d\mathbf{x}\int w(\mathbf{x})h^2(\mathbf{x})\,d\mathbf{x}$$

equality if and only if $g(\mathbf{x}) = ch(\mathbf{x})$ for c some constant not dependent on $\mathbf{x}$

$$\begin{aligned} w(\mathbf{x}) &= p(\mathbf{x};\theta) \\ g(\mathbf{x}) &= \hat{\alpha} - \alpha \\ h(\mathbf{x}) &= \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} \end{aligned}$$

So, for equality we must have

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = \frac{1}{c}(\hat{\alpha} - \alpha)$$

Where c can depend on θ, but *not* on **x**

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Now, assume $\alpha = g(\theta) = \theta$ (other g() are discussed in later sections) which yields

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c}(\hat{\alpha} - \alpha) \longrightarrow \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c(\theta)}(\hat{\theta} - \theta)$$

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Now, assume $\alpha = g(\theta) = \theta$ (other g() are discussed in later sections) which yields

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c}(\hat{\alpha} - \alpha) \quad \longrightarrow \quad \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c(\theta)}(\hat{\theta} - \theta)$$

Take one more differential:

$$\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} = -\frac{1}{c(\theta)} + \frac{\partial \left(\frac{1}{c(\theta)}\right)}{\partial \theta}(\hat{\theta} - \theta)$$

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Now, assume $\alpha = g(\theta) = \theta$ (other g() are discussed in later sections) which yields

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = \frac{1}{c}(\hat{\alpha} - \alpha) \longrightarrow \frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = \frac{1}{c(\theta)}(\hat{\theta} - \theta)$$

Take one more differential:

$$\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2} = -\frac{1}{c(\theta)} + \frac{\partial\left(\frac{1}{c(\theta)}\right)}{\partial\theta}(\hat{\theta} - \theta)$$

Take expectation with respect to **x** and use the unbiasedness of the estimator

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right] = \frac{1}{c(\theta)}$$

Chapter 3 – Cramer-Rao lower bound

Proof (last part)

Now, assume $\alpha = g(\theta) = \theta$ (other g() are discussed in later sections) which yields

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c}(\hat{\alpha} - \alpha) \quad \longrightarrow \quad \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{1}{c(\theta)}(\hat{\theta} - \theta)$$

Take one more differential:

$$\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} = -\frac{1}{c(\theta)} + \frac{\partial \left(\frac{1}{c(\theta)}\right)}{\partial \theta}(\hat{\theta} - \theta)$$

Take expectation with respect to **x** and use the unbiasedness of the estimator

$$-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right] = \frac{1}{c(\theta)} \qquad c(\theta) = \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} = \frac{1}{I(\theta)}$$

Chapter 3 – Cramer-Rao lower bound

Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter) *It is assumed that the PDF $p(\mathbf{x};\theta)$ satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right] = 0 \qquad \textit{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]} \tag{3.6}$$

where the derivative is evaluated at the true value of θ and the expectation is taken with respect to $p(\mathbf{x};\theta)$. Furthermore, an unbiased estimator may be found that attains the bound for all θ if and only if

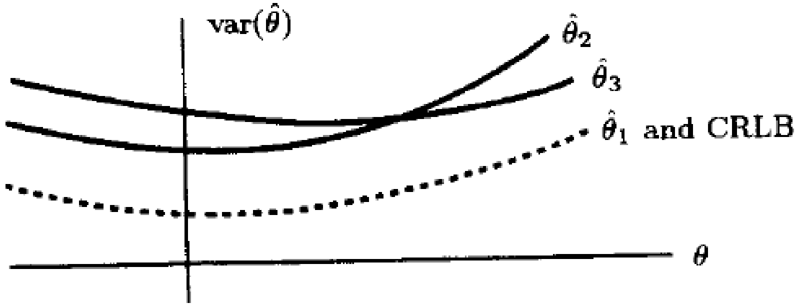
$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g and I. That estimator, which is the MVU estimator, is $\hat{\theta} = g(\mathbf{x})$, and the minimum variance is $1/I(\theta)$.

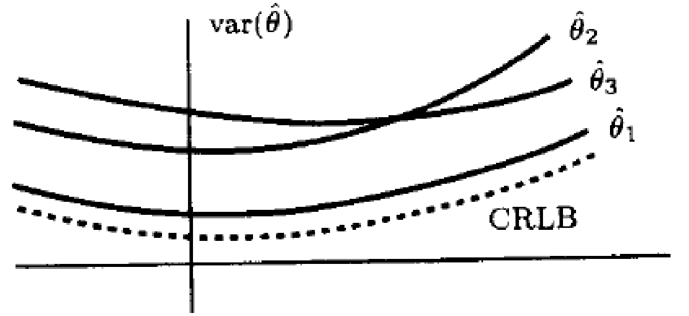
Chapter 3 – Cramer-Rao lower bound

Some remarks

- The CRLB depends on θ
- An estimator that meets the CRLB is said to be *efficient*
- An efficient estimator is **not the same** as an MVU estimator



$\hat{\theta}_1$ efficient and MVU



$\hat{\theta}_1$ MVU but not efficient

Chapter 3 – Cramer-Rao lower bound

Example (3.2, a look at 3.1 again)

Example 3.1 - PDF Dependence on Unknown Parameter

$$x[0] = A + w[0]$$ **Estimate A**

$$w[0] \sim \mathcal{N}(0, \sigma^2)$$

In this example, we just guessed that x[0] would be a good estimator for A.

$$\hat{A} = x[0] \qquad \mathrm{var}(\hat{A}) = \sigma^2$$

But what if we cannot make such guess?

Chapter 3 – Cramer-Rao lower bound

Example (3.2, a look at 3.1 again)

Example 3.1 - PDF Dependence on Unknown Parameter

$x[0] = A + w[0]$ **Estimate A**

$w[0] \sim \mathcal{N}(0, \sigma^2)$

In this example, we just guessed that x[0] would be a good estimator for A.

$\hat{A} = x[0]$ $\quad \text{var}(\hat{A}) = \sigma^2$.

But what if we cannot make such guess?
Use the CRLB theorem

Furthermore, an unbiased estimator may be found that attains the bound for all θ if and only if

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

for some functions g and I. That estimator, which is the MVU estimator, is $\hat{\theta} = g(\mathbf{x})$, and the minimum variance is $1/I(\theta)$.

Chapter 3 – Cramer-Rao lower bound

Example (3.2, a look at 3.1 again)

Example 3.1 - PDF Dependence on Unknown Parameter

$$x[0] = A + w[0]$$ Estimate A

$$w[0] \sim \mathcal{N}(0, \sigma^2)$$

In this example, we just guessed that x[0] would be a good estimator for A.

$$\hat{A} = x[0] \qquad \mathrm{var}(\hat{A}) = \sigma^2.$$

But what if we cannot make such guess?
Use the CRLB theorem

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

Chapter 3 – Cramer-Rao lower bound

Example (3.2, a look at 3.1 again)

Example 3.1 - PDF Dependence on Unknown Parameter

$$x[0] = A + w[0]$$ **Estimate A**

$$w[0] \sim \mathcal{N}(0, \sigma^2)$$

In this example, we just guessed that x[0] would be a good estimator for A.

$$\hat{A} = x[0] \qquad \mathrm{var}(\hat{A}) = \sigma^2.$$

But what if we cannot make such guess?
Use the CRLB theorem

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

Now identify

$$\begin{aligned} \theta &= A \\ I(\theta) &= \frac{1}{\sigma^2} \\ g(x[0]) &= x[0] \end{aligned}$$

Chapter 3 – Cramer-Rao lower bound

Example (3.2, a look at 3.1 again)

Example 3.1 - PDF Dependence on Unknown Parameter

$$x[0] = A + w[0]$$ Estimate A

$$w[0] \sim \mathcal{N}(0, \sigma^2)$$

In this example, we just guessed that x[0] would be a good estimator for A.

$$\hat{A} = x[0] \qquad \mathrm{var}(\hat{A}) = \sigma^2.$$

But what if we cannot make such guess?
Use the CRLB theorem

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

Now identify

$$\begin{aligned} \theta &= A \\ I(\theta) &= \frac{1}{\sigma^2} \\ g(x[0]) &= x[0] \end{aligned}$$

In most cases we cannot express the derivative in this way, and an efficient estimator does not exist in those cases