Estimation theory

- ✓ Parametric estimation
- ✓ Properties of estimators
- ✓ Minimum variance estimator
- ✓ Cramer-Rao bound
- ✓ Maximum likelihood estimators
- ✓ Confidence intervals
- ✓ Bayesian estimation

Random Variables

Let X be a scalar random variable (rv)

$$X : \Omega \to \mathbb{R}$$

defined over the set of elementary events Ω.
The notation

$$X \sim F_X(x),\ f_X(x)$$

denotes that:

- $F_X(x)$ is the *cumulative distribution function (cdf)* of X

$$F_X(x) = P\left\{X \leq x\right\}, \quad \forall x \in \mathbb{R}$$

- $f_X(x)$ is the *probability density function (pdf)* of X

$$F_X(x) = \int_{-\infty}^{x} f_X(\sigma)\, d\sigma, \quad \forall x \in \mathbb{R}$$

Multivariate distributions

Let $X = (X_1, \dots, X_n)$ be a vector of rvs

$$X : \Omega \to \mathbb{R}^n$$

defined over Ω.

The notation

$$X \sim F_X(x),\ f_X(x)$$

denotes that:

- $F_X(x)$ is the *joint cumulative distribution function (cdf)* of X

$$F_X(x) = P\left\{X_1 \leq x_1, \dots, X_n \leq x_n\right\}, \quad \forall x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

- $f_X(x)$ is the *joint probability density function (pdf)* of X

$$F_X(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(\sigma_1, \dots, \sigma_n)\, d\sigma_1 \dots d\sigma_n, \quad \forall x \in \mathbb{R}^n$$

Moments of a rv

- First order moment (*mean*)

$$m_X = E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

- Second order moment (*variance*)

$$\sigma_X^2 = \mathrm{Var}(X) = E\left[(X - m_X)^2\right] = \int_{-\infty}^{+\infty} (x - m_X)^2\, f_X(x)\, dx$$

Example The *normal* or *Gaussian pdf*, denoted by $N(m, \sigma^2)$, is defined as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-m)^2}{2\sigma^2}}.$$

It turns out that $E[X] = m$ and $\mathrm{Var}(X) = \sigma^2$.

Conditional distribution

Bayes formula

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

One has:

$\Rightarrow \; f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy$

$\Rightarrow$ If X and Y are independent: $f_{X|Y}(x|y) = f_X(x)$

Definitions:

- conditional mean: $E[X|Y] = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\, dx$
- conditional variance: $P_{X|Y} = \int_{-\infty}^{+\infty} (x - E[X|Y])^2\, f_{X|Y}(x|y)\, dx$

Gaussian conditional distribution

Let X and Y Gaussian rvs such that:

$$E[X] = m_X \qquad E[Y] = m_Y$$

$$E\left[\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}'\right] = \begin{pmatrix} R_X & R_{XY} \\ R'_{XY} & R_Y \end{pmatrix}$$

It turns out that:

$$\begin{aligned} E[X|Y] &= m_x + R_{XY} R_Y^{-1}(Y - m_Y) \\ P_{X|Y} &= R_X - R_{XY} R_Y^{-1} R'_{XY} \end{aligned}$$

Estimation problems

Problem. Estimate the value of $\theta \in \mathbb{R}^p$, using an observation y of the rv $Y \in \mathbb{R}^n$.

Two different settings:

a. *Parametric estimation*

The pdf of Y depends on the unknown parameter θ

b. *Bayesian estimation*

The unknown θ is a random variable

Parametric estimation problem

- The cdf and pdf of Y depend on the unknown parameter vector θ,

$$Y \sim F_Y^{\theta}(x),\ f_Y^{\theta}(x)$$

- $\Theta \subseteq \mathbb{R}^p$ denotes the *parameter space*, i.e., the set of values which θ can take
- $\mathcal{Y} \subseteq \mathbb{R}^n$ denotes the *observation space*, to which belongs the rv Y

Parametric estimator

The parametric estimation problem consists in finding θ on the basis of an observation y of the rv Y.

Definition 1 *An* estimator *of the parameter* θ *is a function*

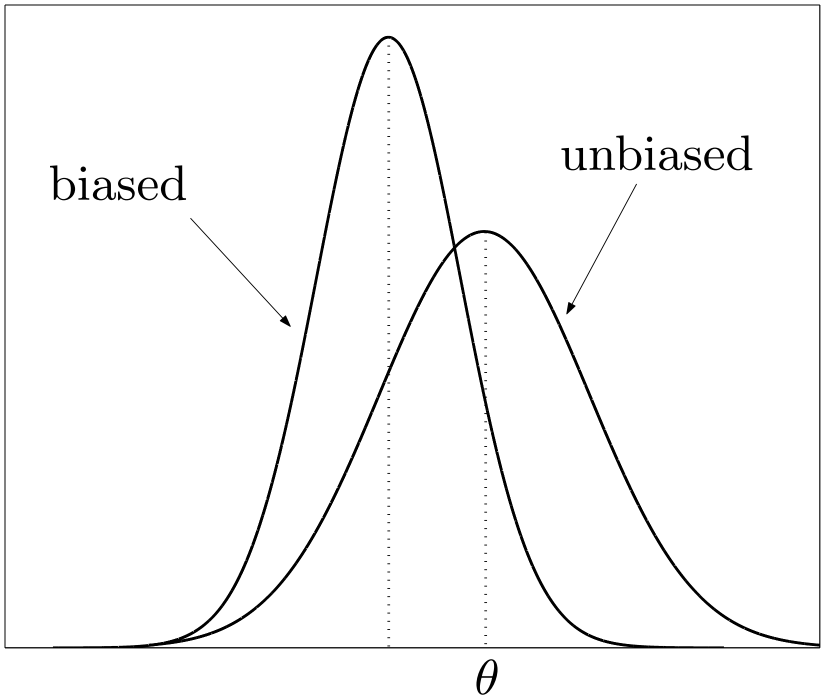
$$T : \mathcal{Y} \longrightarrow \Theta$$

Given the estimator $T(\cdot)$, if one observes, y, then the estimate of θ is $\hat{\theta} = T(y)$.

There are infinite possible estimators (all the functions of y!). Therefore, it is crucial to establish a *criterion* to assess the quality of an estimator.

Unbiased estimator

Definition 2 *An estimator $T(\cdot)$ of the parameter θ is* unbiased *(or* correct*) if $E^{\theta}[T(\cdot)] = \theta$, $\forall \theta \in \Theta$.*



Pdf of two estimators $T(\cdot)$

Examples

- Let $Y_1, \ldots, Y_n$ be identically distributed rvs, with mean m.

  The *sample mean*

  $$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

  is an unbiased estimator of m. Indeed,

  $$E\left[\bar{Y}\right] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i] = m$$

- Let $Y_1, \ldots, Y_n$ be independent identically distributed (i.i.d.) rvs, with variance σ^2.

  The *sample variance*

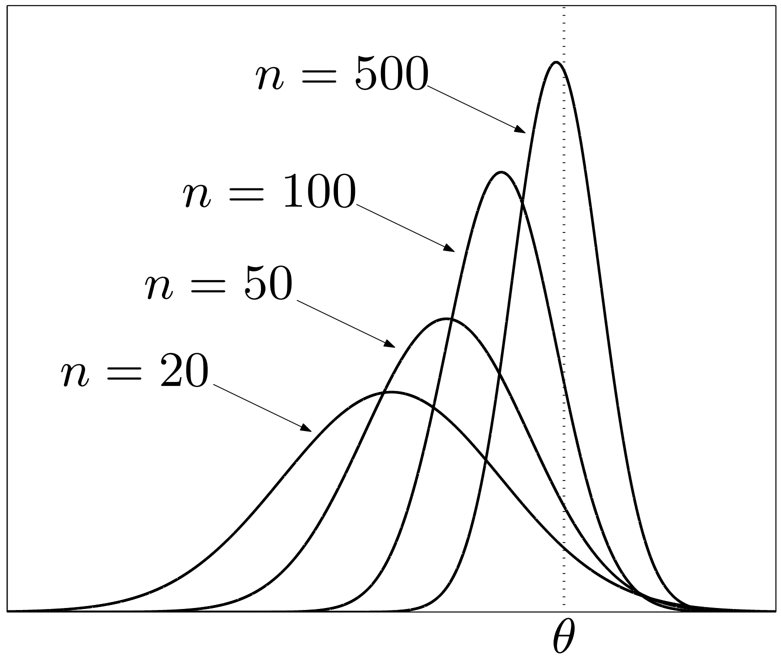
  $$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

  is an unbiased estimator of σ^2.

Consistent estimator

Definition 3 *Let* $\{Y_i\}_{i=1}^{\infty}$ *be a sequence of rvs. The sequence of estimators* $T_n{=}T_n(Y_1,\ldots,Y_n)$ *is said to be* consistent *if* T_n *converges to* θ *in probability for all* $\theta \in \Theta$, *i.e.*

$$\lim_{n\to\infty} P\left\{\|T_n - \theta\| > \varepsilon\right\} = 0 \quad , \quad \forall \varepsilon > 0 \quad , \quad \forall \theta \in \Theta$$



A sequence of consistent estimators $T_n(\cdot)$

Example

Let $Y_1, \ldots, Y_n$ be independent rvs with mean m and finite variance. The *sample mean*

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

is a consistent estimator of m, thanks to the next result.

Theorem 1 (Law of large numbers)
Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of independent rvs with mean m and finite variance. Then, the sample mean $\bar{Y}$ converges to m in probability.

A suffcient condition for consistency

Theorem 2 *Let $\hat{\theta}_n = T_n(y)$ be a sequence of unbiased estimators of $\theta \in \mathbb{R}$, based on the realization $y \in \mathbb{R}^n$ of the n-dimensional rv Y, i.e.:*

$$E^\theta[T_n(y)] = \theta, \quad \forall n, \quad \forall \theta \in \Theta.$$

If

$$\lim_{n \to +\infty} E^\theta\left[(T_n(y) - \theta)^2\right] = 0,$$

then, the sequence of estimators $T_n(\cdot)$ is consistent.

Example. Let $Y_1, \ldots, Y_n$ be independent rvs with mean m and variance σ^2. We know that the *sample mean* $\bar{Y}$ is an unbiased estimate of m. Moreover, it turns out that

$$\mathrm{Var}(\bar{Y}) = \frac{\sigma^2}{n}$$

Therefore, the sample mean is a consistent estimator of the mean.

Mean square error

Consider an estimator $T(\cdot)$ of the scalar parameter θ.

Definition 4 *We define* mean square error (MSE) *of* $T(\cdot)$,

$$E^{\theta}\left[(T(Y)-\theta)^2\right]$$

If the estimator $T(\cdot)$ is unbiased, the mean square error corresponds to the variance of the estimation error $T(Y)-\theta$.

Definition 5 *Given two estimators* $T_1(\cdot)$ *and* $T_2(\cdot)$ *of* θ*,* $T_1(\cdot)$ *is* better *than* $T_2(\cdot)$ *if*

$$E^{\theta}\left[(T_1(Y)-\theta)^2\right] \leq E^{\theta}\left[(T_2(Y)-\theta)^2\right] \quad , \quad \forall \theta \in \Theta$$

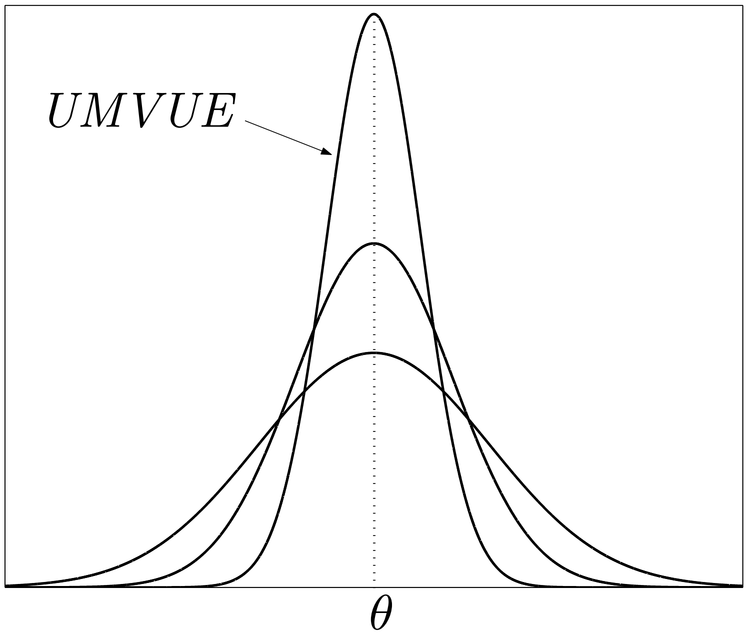
If we restrict our attention to unbiased estimators, we are interested to the one with the least MSE for any value of θ (notice that it may not exist).

Minimum variance unbiased estimator

Definition 6 *An unbiased estimator $T^*(\cdot)$ of θ is* UMVUE (Uniformly Minimum Variance Unbiased Estimator) *if*

$$E^\theta\left[(T^*(Y)-\theta)^2\right] \leq E^\theta\left[(T(Y)-\theta)^2\right], \quad \forall\theta\in\Theta$$

for any unbiased estimator $T(\cdot)$ of θ.



Minimum variance linear estimator

Let us restrict our attention to the class of *linear estimators*

$$T(x) = \sum_{i=1}^{n} a_i x_i \quad , \quad a_i \in \mathbb{R}$$

Definition 7 *A linear unbiased estimator $T^*(\cdot)$ of the scalar parameter θ is said to be* BLUE *(Best Linear Unbiased Estimator) if*

$$E^\theta\left[(T^*(Y)-\theta)^2\right] \leq E^\theta\left[(T(Y)-\theta)^2\right], \quad \forall\theta \in \Theta$$

for any linear unbiased estimator $T(\cdot)$ di θ.

Example Let Y_i be independent rvs with mean m and variance σ_i^2, $i = 1, \ldots, n$.

$$\hat{Y} = \frac{1}{\displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} Y_i$$

is the BLUE estimator of m.

Cramer-Rao bound

The *Cramer-Rao bound* is a lower bound to the variance of any unbiased estimator of the parameter θ.

Theorem 3 *Let $T(\cdot)$ be an unbiased estimator of the scalar parameter θ, and let the observation space $\mathcal{Y}$ be independent on θ. Then (under some technical assumptions),*

$$E^{\theta}\left[(T(Y)-\theta)^2\right] \geq [I_n(\theta)]^{-1}$$

where $I_n(\theta)=E^{\theta}\left[\left(\frac{\partial \ln f_Y^{\theta}(Y)}{\partial\theta}\right)^2\right]$ *(*Fisher information*).*

Remark To compute $I_n(\theta)$ one must know the actual value of θ; therefore, the Cramer-Rao bound is usually unknown in practice.

Cramer-Rao bound

For a parameter vector θ and any unbiased estimator $T(\cdot)$, one has

$$E^{\theta}\left[(T(Y)-\theta)\,(T(Y)-\theta)'\right] \geq [I_n(\theta)]^{-1} \tag{1}$$

where

$$I_n(\theta) = E^{\theta}\left[\left(\frac{\partial \ln f_Y^{\theta}(Y)}{\partial\theta}\right)'\left(\frac{\partial \ln f_Y^{\theta}(Y)}{\partial\theta}\right)\right]$$

is the *Fisher information matrix*.

The inequality in (**??**) is in matricial sense ($A \geq B$ means that $A - B$ is positive semidefinite).

Definition 8 *An unbiased estimator $T(\cdot)$ such that equality holds in* (**??**) *is said to be* efficient*.*

Cramer-Rao bound

If the rvs $Y_1, \ldots, Y_n$ are *i.i.d.*, it turns out that

$$I_n(\theta) = nI_1(\theta)$$

Hence, for fixed θ the Cramer-Rao bound decreases as $\dfrac{1}{n}$ with the size n of the data sample.

Example Let $Y_1, \ldots, Y_n$ be *i.i.d.* rvs with mean m and variance σ^2. Then

$$E\left[\left(\bar{Y} - m\right)^2\right] = \frac{\sigma^2}{n} \geq [I_n(\theta)]^{-1} = \frac{[I_1(\theta)]^{-1}}{n}$$

where $\bar{Y}$ denotes the sample mean. Moreover, if the rvs $Y_1, \ldots, Y_n$ are normally distributed, one has also $I_1(\theta) = \dfrac{1}{\sigma^2}$.

Since the Cramer-Rao bound is achieved, *in the case of normal i.i.d rvs, the sample mean is an efficient estimator of the mean.*

Maximum likelihood estimators

Consider a rv $Y \sim f_Y^\theta(y)$, and let y be an observation of Y. We define *likelihood function*, the function of θ (for fixed y)

$$L(\theta|y) = f_Y^\theta(y)$$

We choose as estimate of θ the value of the parameter which maximises the likelihood of the observed event (this value depends on y!).

Definition 9 *A* maximum likelihood estimator *of the parameter θ is the estimator*

$$T_{ML}(x) = \arg\max_{\theta\in\Theta} L(\theta|x)$$

Remark The functions $L(\theta|x)$ and $\ln L(\theta|x)$ achieve their maximum values for the same θ. In some cases is easier to find the maximum of $\ln L(\theta|x)$ (exponential distributions).

Properties of the maximum likelihood estimators

Theorem 4 *Under the assumptions for the existence of the Cramer-Rao bound, if there exists an efficient estimator* $T^*(\cdot)$*, then it is a maximum likelihood estimator* $T_{ML}(\cdot)$*.*

Example Let $Y_i \sim N(m, \sigma_i^2)$ be independent, with known σ_i^2, $i = 1, \ldots, n$. The estimator

$$\hat{Y} = \frac{1}{\displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} Y_i$$

of m is unbiased and such that $\mathrm{Var}(\hat{Y}) = \dfrac{1}{\displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$, while $I_n(m) = \displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2}$.

Hence, $\hat{Y}$ is efficient, end therefore it s a maximum likelihood estimator of m.

The maximum likelihood estimator has several nice asymptotic properties.

Theorem 5 *If the rvs* $Y_1, \ldots, Y_n$ *are* i.i.d., *then (under suitable technical assumptions)*

$$\lim_{n \to +\infty} \sqrt{I_n(\theta)}\,(T_{ML}(Y) - \theta)$$

is a random variable with standard normal distribution N(0,1).

Theorem **??** states that the maximum likelihood estimator

- asymptotically unbiased
- consistent
- asymptotically efficient
- asymptotically normal

Example Let $Y_1, \ldots, Y_n$ be normal rvs with mean m and variance σ^2. The sample mean

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

is a maximum likelihood estimator of m.
Moreover, $\sqrt{I_n(m)}(\bar{Y} - m) \sim N(0, 1)$, being $I_n(m)=\dfrac{n}{\sigma^2}$.

Remark *The maximum likelihood estimator may be biased.* Let $Y_1, \ldots, Y_n$ be independent normal rvs with variance σ^2. The maximum likelihood estimator of σ^2 is

$$\hat{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

which is biased, being $E[\hat{S}^2] = \dfrac{n-1}{n}\sigma^2$.

Confidence intervals

In many estimation problems, it is important to establish a set to which the parameter to be estimated belongs with a known probability.

Definition 10 *A* confidence interval *with confidence level* $1 - \alpha$*,* $0 < \alpha < 1$*, for the scalar parameter* θ *is a function that maps any observation* $y \in \mathcal{Y}$ *into an interval* $\mathcal{B}(y) \subseteq \Theta$ *such that*

$$P^{\theta}\left\{\theta \in \mathcal{B}(y)\right\} \geq 1 - \alpha \quad , \quad \forall \theta \in \Theta$$

Hence, a confidence interval of level $1 - \alpha$ for θ is a subset of Θ such that, if we observe y, then $\theta \in \mathcal{B}(y)$ with probability at least $1 - \alpha$, whatever is the true value $\theta \in \Theta$.

Example Let $Y_1, \ldots, Y_n$ be normal rvs with unknown mean m and known variance σ^2. Then, $\dfrac{\sqrt{n}}{\sigma}(\bar{Y} - m) \sim N(0,1)$, where $\bar{Y}$ is the sample mean.

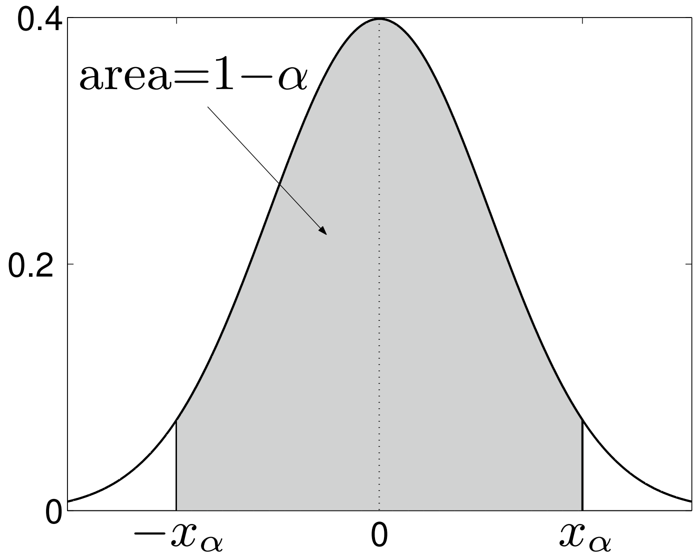
Let y_α be such that $\displaystyle\int_{-y_\alpha}^{y_\alpha} \frac{1}{\sqrt{2\pi}} e^{-y^2} dy = 1 - \alpha$. Being,

$$1 - \alpha = P\left\{ \left| \frac{\sqrt{n}}{\sigma}(\bar{Y} - m) \right| \leq y_\alpha \right\} = P\left\{ \bar{Y} - \frac{\sigma}{\sqrt{n}}\, y_\alpha \leq m \leq \bar{Y} + \frac{\sigma}{\sqrt{n}}\, y_\alpha \right\}$$

one has that

$$\left[\bar{Y} - \frac{\sigma}{\sqrt{n}}\, y_\alpha \, , \, \bar{Y} + \frac{\sigma}{\sqrt{n}}\, y_\alpha \right]$$

is a confidence interval of level $1 - \alpha$ for m.



Nonlinear ML estimation problems

Let $Y \in \mathbb{R}^n$ be a vector of rvs such that

$$Y = U(\theta) + \varepsilon$$

where

- $\theta \in \mathbb{R}^p$ is the unknown parameter vector
- $U(\cdot) : \mathbb{R}^p \to \mathbb{R}^n$ is a known function
- $\varepsilon \in \mathbb{R}^n$ is a vector of rvs, for which we assume

$$\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

<u>Problem</u>: find a maximum likelihood estimator of θ

$$\hat{\theta}_{ML} \;=\; T_{ML}(Y)$$

Least squares estimate

The pdf of the *data* Y is

$$f_Y(y) \;=\; f_\varepsilon(y - U(\theta)) \;=\; L(\theta|y)$$

Therefore,

$$\begin{aligned} \hat{\theta}_{ML} &= \arg\max_\theta \ln L(\theta|y) \\ &= \arg\min_\theta (y - U(\theta))' \Sigma_\varepsilon^{-1} (y - U(\theta)) \end{aligned}$$

If the covariance matrix Σ_ε is known, we obtain the *weighted least squares estimate.*

If $U(\theta)$ is a generic nonlinear function, the solution must be computed numerically MATLAB Optimization Toolbox $\rightarrow$ `>> help optim`
This problem can be computationally intractable!

Linear estimation problems

If the function $U(\cdot)$ is *linear*, i.e., $U(\theta) = U\theta$ with $U \in \mathbb{R}^{n\times p}$ known matrix, one has

$$Y = U\theta + \varepsilon$$

and the maximum likelihood estimator is the so-called *Gauss-Markov estimator*

$$\hat{\theta}_{ML} = \hat{\theta}_{GM} = (U'\Sigma_\varepsilon^{-1}U)^{-1}U'\Sigma_\varepsilon^{-1}y$$

In the special case $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ (the rvs ε_i are independent!), one has the celebrated *least squares estimator*

$$\hat{\theta}_{LS} = (U'U)^{-1}U'y$$

A special case: biased measurement error

How to treat the case in which $E[\varepsilon_i] = m_\epsilon \neq 0$, $\forall i = 1, \ldots, n$?

1) If m_ϵ is known, just use the "unbiased" measurements $Y - m_\varepsilon \mathbb{1}$:

$$\hat{\theta}_{ML} = \hat{\theta}_{GM} = (U'\Sigma_\varepsilon^{-1}U)^{-1}U'\Sigma_\varepsilon^{-1}(y - m_\epsilon \mathbb{1})$$

where $\mathbb{1} = [1 \quad 1 \quad \ldots \quad 1]'$.

2) If m_ε is unknown, estimate it!
Let $\bar{\theta} = [\theta' \quad m_\varepsilon]' \in \mathbb{R}^{p+1}$ and then

$$Y = [U \quad \mathbb{1}]\bar{\theta} + \varepsilon$$

Then, apply the Gauss-Markov estimator with $\bar{U} = [U \quad \mathbb{1}]$ to obtain an estimate of $\bar{\theta}$ (simultaneous estimate of θ and m_ε).
Clearly, the variance of the estimation error of θ will be higher wrt case 1)

Gauss-Markov estimator

The estimates $\hat{\theta}_{GM}$ and $\hat{\theta}_{LS}$ are widely used in practice, also if some of the assumptions on ε do not hold or cannot be validated. In particular, the following result holds.

Theorem 6 *Let $Y = U\theta + \varepsilon$ with ε a vector of random variables with zero mean and covariance matrix Σ. Then, the Gauss-Markov estimator is the BLUE estimator of the parameter θ,*

$$\hat{\theta}_{BLUE} = \hat{\theta}_{GM}$$

and the corresponding covariance of the estimation error is equal to

$$E\left[\left(\hat{\theta}_{GM} - \theta\right)\left(\hat{\theta}_{GM} - \theta\right)'\right] = \left(U'\Sigma^{-1}U\right)^{-1}.$$

Examples of least squares estimate

Example 1.

$Y_i = \theta + \varepsilon_i, \;\; i = 1, \ldots, n$

ε_i independent rvs with zero mean and variance σ^2

$\Rightarrow \;\; E[Y_i] = \theta$

We want to estimate θ using observations of $Y_i, \;\; i = 1, \ldots, n$

One has $Y = U\theta + \varepsilon$ with $U = (1 \; 1 \; \ldots \; 1)'$ and

$$\hat{\theta}_{LS} = (U'U)^{-1}U'y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The least squares estimator is equal to the sample mean (and it is also the maximum likelihood estimate if the rvs ε_i are normal).

Example 2.

Same setting of Example 1, with $E[\varepsilon_i^2] = \sigma_i^2$, $i = 1, \ldots, n$

In this case, $E[\varepsilon\varepsilon'] = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix}$

$\Rightarrow$ The least squares estimator is still the sample mean

$\Rightarrow$ The Gauss-Markov estimator is

$$\hat{\theta}_{GM} = (U'\Sigma_\varepsilon^{-1}U)^{-1}U'\Sigma_\varepsilon^{-1}y = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\, y_i$$

and is equal to the maximum likelihood estimate if the rvs ε_i are normal.

Summary of GM and LS estimator properties

Assumption on ε	**GM estimator**	**LS estimator**
none	$\arg\min_\theta (y - U\theta)^T \Sigma_\varepsilon^{-1} (y - U\theta)$ with given weight matrix Σ_ε	$\arg\min_\theta \|\|y - U\theta\|\|^2$
$\mathrm{E}[\varepsilon]$ known	unbiased	unbiased
$\mathrm{E}[\varepsilon] = m_\varepsilon$ $\mathrm{Var}\{\varepsilon\} = \Sigma_\varepsilon$	BLUE estimator	BLUE estimator if $\Sigma_\varepsilon = \sigma_\varepsilon^2 I_n$
$\varepsilon \sim N(m_\varepsilon, \Sigma_\varepsilon)$	ML estimator efficient, UMVUE	ML estimator if $\Sigma_\varepsilon = \sigma_\varepsilon^2 I_n$

Bayesian estimation

Estimate an unknown rv X, using observations of the rv Y

Key tool: joint pdf $f_{X,Y}(x, y)$

$\Rightarrow$ least mean square error estimator

$\Rightarrow$ optimal linear estimator

Bayesian estimation: problem formulation

Problem:
Given observations y of the rv $Y \in \mathbb{R}^n$, find an estimator of the rv X based on the data y.

Solution: an estimator $\hat{X} = T(Y)$, where $T(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^p$

To assess the quality of the estimator we must define a suitable criterion: in general, we consider the *risk function*

$$J_r = E[d(X, T(Y))] \; = \int\!\!\int d(x, T(y))\, f_{X,Y}(x, y)\, dx\, dy$$

and we *minimize* J_r with respect to all possible estimators $T(\cdot)$

$d(X, T(Y)) \rightarrow$ "distance" between the unknown X and its estimate $T(Y)$

Least mean square error estimator

Let $d(X, T(Y)) = \|X - T(Y)\|^2$.
One gets the least *mean square error (MSE)* estimator

$$\hat{X}_{MSE} = T^*(Y)$$

where

$$T^*(\cdot) \;=\; \arg\min_{T(\cdot)} \; E[\|X - T(Y)\|^2]$$

Theorem

$$\hat{X}_{MSE} \;=\; E[X|Y] \;.$$

The *conditional mean* of X given Y is the least MSE estimate of X based on the observation of Y

Let $Q(X, T(Y)) = E[(X - T(Y))(X - T(Y))']$. Then:
$Q(X, \hat{X}_{MSE}) \leq Q(X, T(Y))$, for any $T(Y)$.

Optimal linear estimator

The least MSE estimator needs the knowledge of the conditional distribution of X given Y
$\rightarrow$ Simpler estimators

Linear estimators:

$$T(Y) = AY + b$$

$A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^{p \times 1}$: estimator coefficients (to be determined)

The *Linear Mean Square Error (LMSE)* estimate is given by

$$\hat{X}_{LMSE} \;=\; A^* Y + b^*$$

where

$$A^*, b^* \;=\; \arg \min_{A,b} \; E[\|X - AY - b\|^2]$$

LMSE estimator

Theorem

Let X and Y be rvs such that:

$$E[X] = m_X \qquad E[Y] = m_Y$$

$$E\left[\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}'\right] = \begin{pmatrix} R_X & R_{XY} \\ R'_{XY} & R_Y \end{pmatrix}$$

Then

$$\hat{X}_{LMSE} \;=\; m_X + R_{XY}R_Y^{-1}(Y - m_Y)$$

i.e,

$$A^* = R_{XY}R_Y^{-1} \quad , \qquad b^* = m_X - R_{XY}R_Y^{-1}m_Y$$

Moreover,

$$E[(X - \hat{X}_{LMSE})(X - \hat{X}_{LMSE})'] = R_X - R_{XY}R_Y^{-1}R'_{XY}$$

Properties of the LMSE estimator

- The LMSE estimator does not require knowledge of the joint pdf of X e Y, but only of the covariance matrices R_{XY}, R_Y (second order statistics)
- The LMSE estimate satisfies

$$\begin{aligned} E[(X - \hat{X}_{LMSE})Y'] &= E[\{X - m_X - R_{XY}R_Y^{-1}(Y - m_Y)\}Y'] \\ &= R_{XY} - R_{XY}R_Y^{-1}R_Y = 0 \end{aligned}$$

$\Rightarrow$ The optimal linear estimator is *uncorrelated* with data Y

- If X and Y are jointly Gaussian

$$E[X|Y] = m_X + R_{XY}R_Y^{-1}(Y - m_Y)$$

hence

$$\hat{X}_{LMSE} = \hat{X}_{MSE}$$

$\Rightarrow$ In the Gaussian setting, the MSE estimate is a *linear* function of the observed variables Y, and therefore is equal to the LMSE estimate

Sample mean and covariances

In many estimation problems, 1st and 2nd order moments are not known

What if only a set of data x_i, y_i, $i = 1, \ldots, N$, is available? Use the *sample means* and *sample covariances* as estimates of the moments

- Sample means

$$\hat{m}_X^N = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \hat{m}_Y^N = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- Sample covariances

$$\begin{aligned}
\hat{R}_X^N &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{m}_X^N)(x_i - \hat{m}_X^N)' \\
\hat{R}_Y^N &= \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{m}_Y^N)(y_i - \hat{m}_y^N)' \\
\hat{R}_{XY}^N &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{m}_X^N)(y_i - \hat{m}_y^N)'
\end{aligned}$$

Example of LMSE estimation (1/2)

$Y_i, \;\; i = 1, \ldots, n$, rvs such that

$$Y_i = u_i X + \varepsilon_i$$

where

- X rv with mean m_X and variance σ_X^2;
- u_i known coefficients;
- ε_i independent rvs wih zero mean and variance σ_i^2.

One has

$$Y = UX + \varepsilon$$

where $U = (u_1 \;\, u_2 \;\; \ldots \;\; u_n)'$ and $E[\varepsilon\varepsilon'] = \Sigma_\varepsilon = \mathrm{diag}\{\sigma_i^2\}$.

We want to compute the LMSE estimate

$$\hat{X}_{LMSE} \;=\; m_X + R_{XY} R_Y^{-1} (Y - m_Y)$$

Example of LMSE estimation (2/2)

- $m_Y = E[Y] = Um_X$
- $R_{XY} = E[(X - m_X)(Y - Um_X)'] = \sigma_X^2 U'$
- $R_Y = E[(Y - Um_X)(Y - Um_X)'] = U\sigma_X^2 U' + \Sigma_\varepsilon$

Being $\left(U\sigma_X^2 U' + \Sigma_\varepsilon\right)^{-1} = \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1} U \left(U'\Sigma_\varepsilon^{-1} U + \frac{1}{\sigma_X^2}\right)^{-1} U'\Sigma_\varepsilon^{-1}$, one gets

$$\hat{X}_{LMSE} = \frac{U'\Sigma_\varepsilon^{-1} Y + \frac{1}{\sigma_X^2} m_X}{U'\Sigma_\varepsilon^{-1} U + \frac{1}{\sigma_X^2}}$$

Special case: $U = (1\ 1\ \ldots\ 1)'$ (i.e., $Y_i = X + \varepsilon_i$)

$$\hat{X}_{LMSE} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} Y_i + \frac{1}{\sigma_X^2} m_X}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} + \frac{1}{\sigma_X^2}}$$

Remark: the a priori info on X is treated as additional data.

Example of Bayesian estimation (1/2)

Let X and Y be two rvs whose joint pdf is

$$f_{X,Y}(x,y) = \begin{cases} -\frac{3}{2}x^2 + 2xy & 0 \le x \le 1,\ 1 \le y \le 2 \\ 0 & \text{else} \end{cases}$$

We want to find the estimates $\hat{X}_{MSE}$ and $\hat{X}_{LMSE}$ of X, based on one observation of the rv Y.

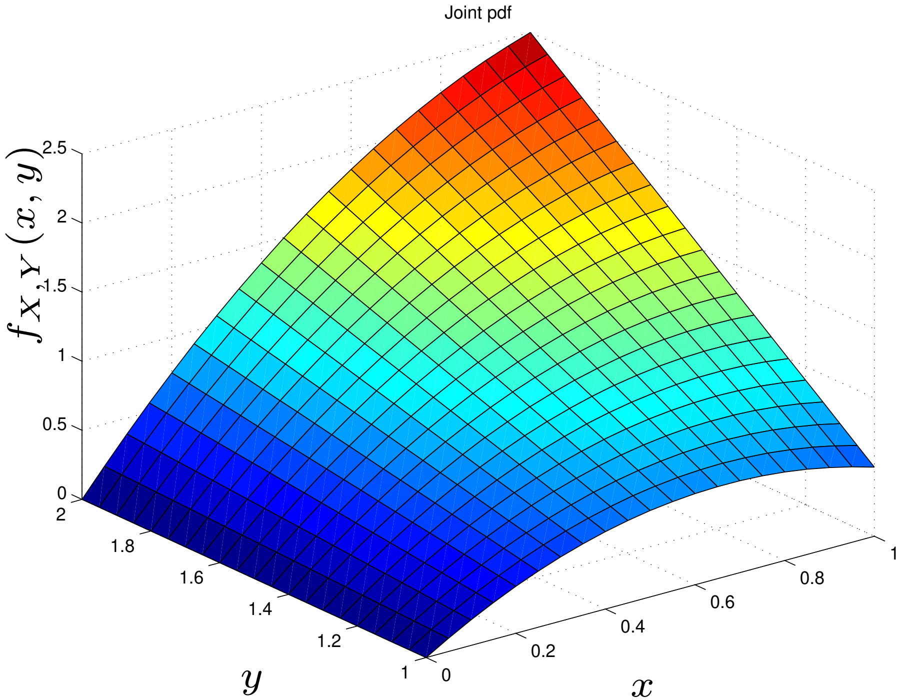
Solutions:

- $\hat{X}_{MSE} = \dfrac{\frac{2}{3}y - \frac{3}{8}}{y - \frac{1}{2}}$

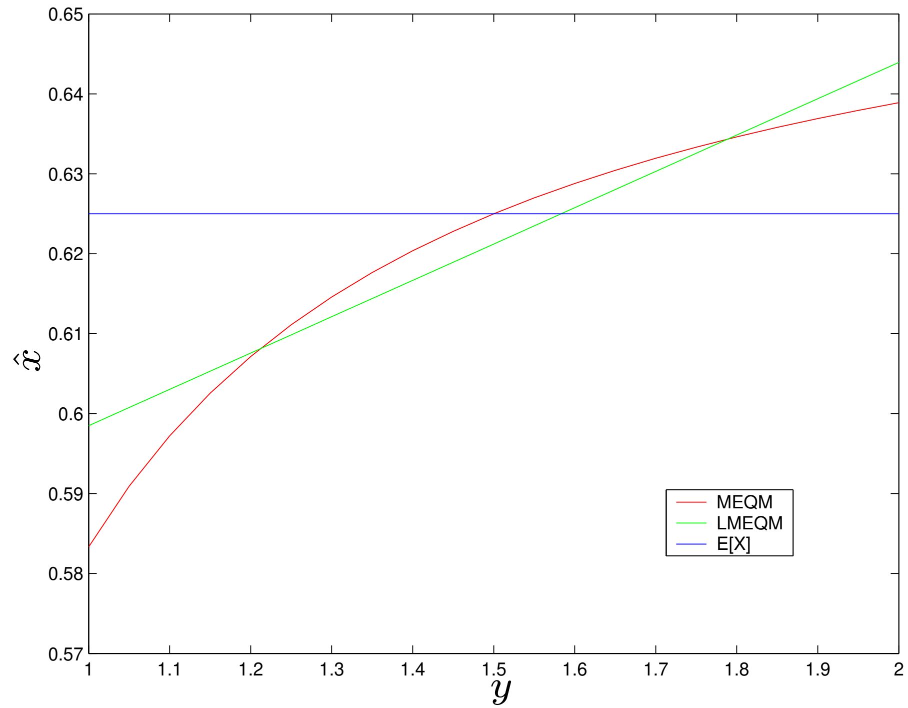
- $\hat{X}_{LMSE} = \dfrac{1}{22}y + \dfrac{73}{132}$

See MATLAB file: `Es_bayes.m`

Example of Bayesian estimation (2/2)



Joint pdf $f_{X,Y}(x,y)$



$\hat{X}_{MSE}(y)$ (red)

$\hat{X}_{LMSE}(y)$ (green)

$E[X]$ (blue)