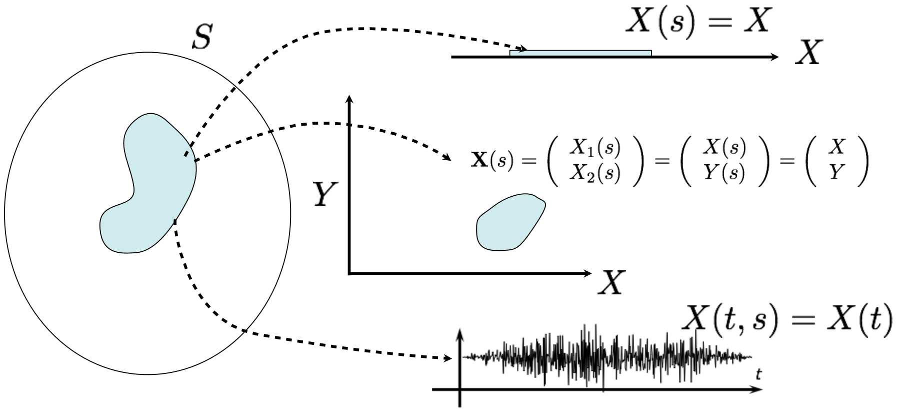
Signal Processing (EE2S31)

Stochastic Processes for EE

Lecture 6: Prediction/Estimation

TU Delft
Delft University of Technology

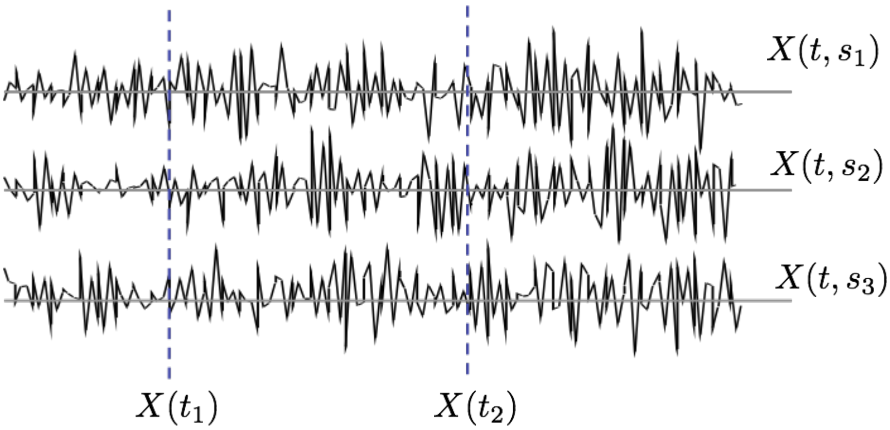
Summarizing



One experiment → one outcome or realization of the stochastic/random process

TU Delft
Delft University of Technology

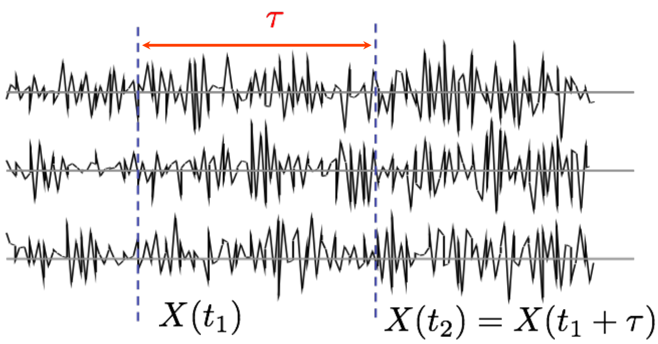
Description random process



TU Delft
Delft University of Technology

Autocovariance

- Joint behavior of X(t) at time t_1 and t_2 can be described by the autocovariance function
  the covariance of $X(t_1)$ and $X(t_2)$ for all t_1 and t_2



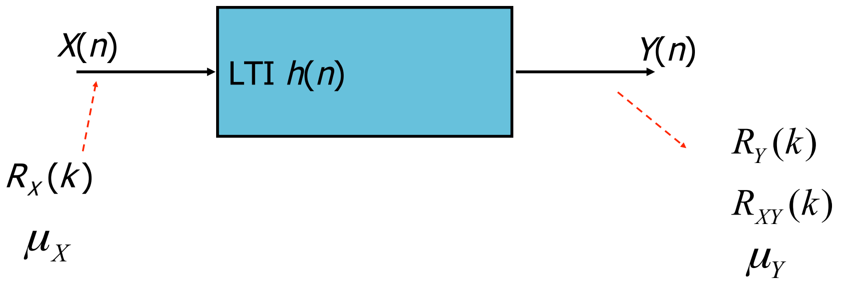
TU Delft
Delft University of Technology

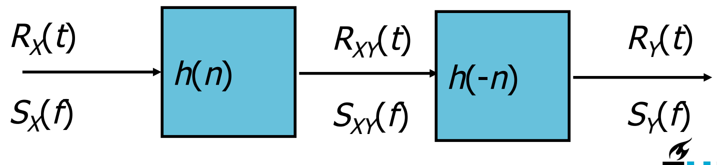
Do understand:

- Stationary and wide-sense stationary processes
- Independence
- Uncorrelated, orthogonal
- Expected value
- Autocorrelation and autocovariance function
- Cross-correlation
- iid
- Standard normal distribution, Gaussian distribution
- Uniform distribution

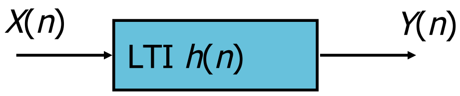
Summarizing

For wide sense stationary input and LTI systems:





Example (1)

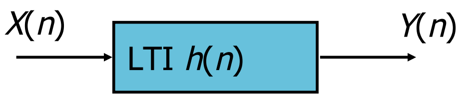


Input $X(n)$ is zero-mean,iid, var$[X] = \sigma^2$ and continuously present at the input of a digital filter.

What is $R_y[n]$ when the input/ouput relation is given by

$$Y[n] = \frac{1}{2}(X[n] + Y[n-1])$$

Example (2)



$$R_X(n) = \sigma^2\delta(n)$$

$S_Y = \sigma^2|H(z)|^2$ System function: (use Z- or Fourier-transforms)

$$H(z) = \frac{1/2}{1 - 1/2z^{-1}} \quad \text{ROC}: \ |z| > \frac{1}{2}$$

Hence, anti-causal system!!

$$R_Y[n] = h(n) * h(-n) * R_x(n) \Rightarrow$$

$$S_Y(z) = H(z)H(z^{-1})\sigma^2 = \frac{1/2}{1 - 1/2z^{-1}} \frac{1/2}{1 - 1/2z}\sigma^2, \quad \text{ROC}: \ \frac{1}{2} < |z| < 2$$

Example (3)

$$S_Y(z) = \frac{1/2}{1-1/2z^{-1}}\frac{1/2}{1-1/2z}\sigma^2 = \frac{-z}{1-2z}\,\frac{1/2}{1-1/2z}\sigma^2$$

$$S_Y(z) = \frac{A}{1-2z} + \frac{B}{1-1/2z}$$

with $A = -\sigma^2/3$ and $B = \sigma^2/3$.

$$S_Y(z) = \frac{-\sigma^2/3}{1-2z} + \frac{\sigma^2/3}{1-1/2z}$$

TU Delft

Example (4)

$$S_Y(z) = \frac{-\sigma^2/3}{1-2z} + \frac{\sigma^2/3}{1-1/2z}$$

Use:

$$a^n u[n] \Rightarrow \frac{1}{1-az^{-1}} \quad |z| > |a|$$

$$-a^n u[-n-1] \Rightarrow \frac{1}{1-az^{-1}} \quad |z| < |a|$$

$$x[-n] \Rightarrow X(z^{-1})$$

$$R_Y[n] = \frac{\sigma^2}{3}\left(\left(\frac{1}{2}\right)^n u(n-1) + \left(\frac{1}{2}\right)^{-n} u(-n)\right)$$

TU Delft

This week: prediction/estimation

Consider random variables: Estimating a RV X

- Estimating X: blind estimate
- Minimum mean squared error
- Linear estimation

Estimating the mean: The sample mean

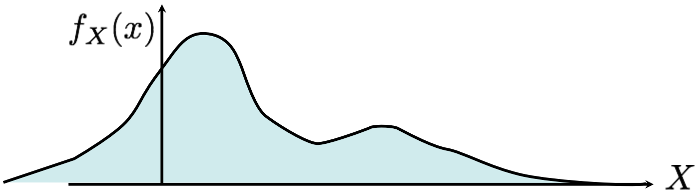
- Characteristics of the sample mean
- Markov and Chebychev inequality
- Weak law of large numbers

Estimation of a random variable

Without doing any measurements, can we estimate the random variable X? What would be our best gamble? (blind estimate)

Estimation of a random variable

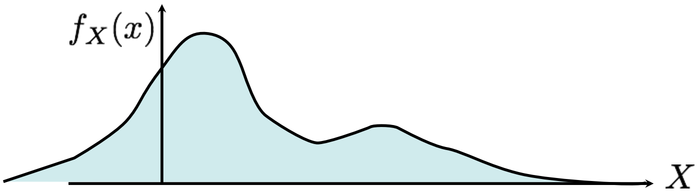
- Without doing any measurements, can we estimate what X is?
- Ok, we know $f_X(x)$, now what?



Estimation of a random variable

- Without doing any measurements, can we estimate what X is?
- We need to define an error for an estimate $\hat{x}_B$
- Typically, the Mean Square Error (MSE):

$$e = E\left[(X - \hat{x}_B)^2\right]$$



Blind estimate

- Minimize the MSE

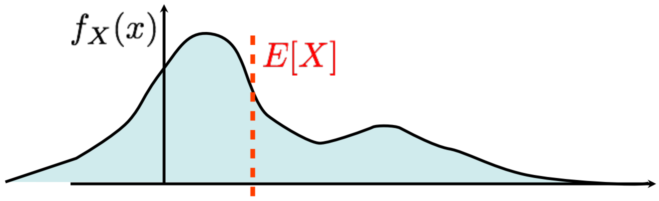
$$e = E\left[(X - \hat{x}_B)^2\right]$$
$$= E[X^2] - 2\hat{x}_B E[X] + \hat{x}_B^2$$

- Set derivative to zero, and solve:

$$\frac{\partial e}{\partial \hat{x}_B} = -2E[X] + 2\hat{x}_B = 0$$

- So

$$\hat{x}_B = E[X]$$



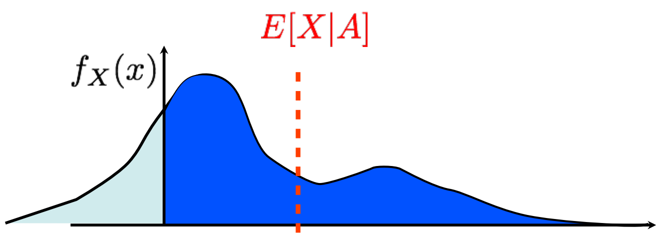
TU Delft
Delft University of Technology

Estimation given an event

- Now assume we know that X comes from an event A

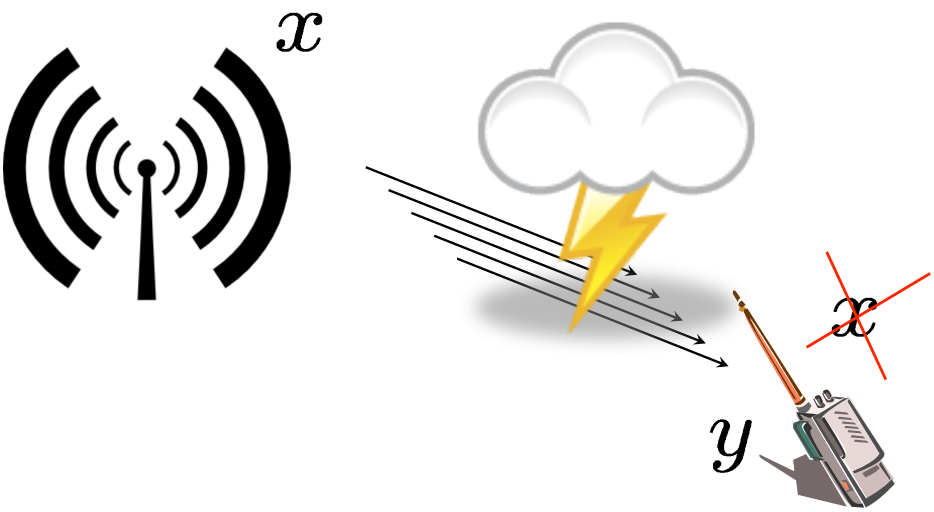
$$A : \{X \geq 0\}$$

- In an identical manner: $\hat{x}_A = E[X|A]$



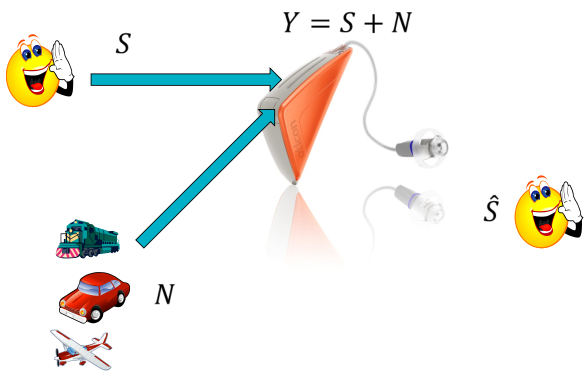
TU Delft
Delft University of Technology

Estimate Given a Realization/Event



TU Delft

Estimate Given a Realization



TU Delft

Estimate Given a Realization

Problem formulation:

- We are interested in the value of a certain random variable X.
- Somehow we can only observe a distorted version of X, say $Y(X)$.

How to estimate X using the observable random variable $Y(X)$?

Bayesian Estimation!

TU Delft 9

Bayesian estimation of a RV

- Define a non-negative cost function $\mathcal{C}(X, \hat{X}(Y))$, e.g., $\mathcal{C}(X, \hat{X}(Y)) = (X - \hat{X}(Y))^2$.
- Minimize expected costs: $\mathcal{R} = E\left[\mathcal{C}(X, \hat{X}(Y))\right]$

Since both X and $\hat{X}$ are random random variables we can express $\mathcal{R}$ as

$$\mathcal{R} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathcal{C}(X, \hat{X}(Y)) f_{X,Y}(x, y)\,dx\,dy$$

TU Delft 20

Bayesian estimation of a RV

Remember Bayes rule:

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

Using Bayes rule we can write:

$$\mathcal{R} = \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} \mathcal{C}(x, \hat{x}(y)) f_{X|Y}(x|y) dx}_{I} f_Y(y) dy$$

Notice that

- $\mathcal{C}(x, \hat{x}(y)) \geq 0$
- $f_{X|Y}(x|y) \geq 0$
- $f_Y(y) \geq 0$

TU Delft

Bayesian estimation of a RV

We can simplify our problem:

In order to minimize $\mathcal{R}$, it is sufficient to minimize

$$\mathcal{I} = \int_{-\infty}^{+\infty} \mathcal{C}(x, \hat{x}(y)) f_{X|Y}(x|y) dx$$

for each realization y.

$$\hat{x}(y) = \arg \min_{\hat{x}(y)} \int_{-\infty}^{+\infty} \mathcal{C}(x, \hat{x}(y)) f_{X|Y}(x|y) dx$$

TU Delft

Bayesian Estimation: MMSE

$$\begin{aligned}\frac{dI(\hat{x}(y))}{d\hat{x}(y)} &= \frac{d}{d\hat{x}(y)}\int_{-\infty}^{+\infty}(x-\hat{x}(y))^2 f_{X|Y}(x|y)dx \\ &= -2\int_{-\infty}^{+\infty}(x-\hat{x}(y)) f_{X|Y}(x|y)dx \\ &= 0\end{aligned}$$

$$\int_{-\infty}^{+\infty}\hat{x}(y) f_{X|Y}(x|y)dx = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx$$

TU Delft 23

Bayesian Estimation: MMSE

$$\hat{x}(y)\underbrace{\int_{-\infty}^{+\infty} f_{X|Y}(x|y)dx}_{=1} = E[X|y]$$

$$\hat{x}_{\text{MMSE}}(y) = E[X|y]$$

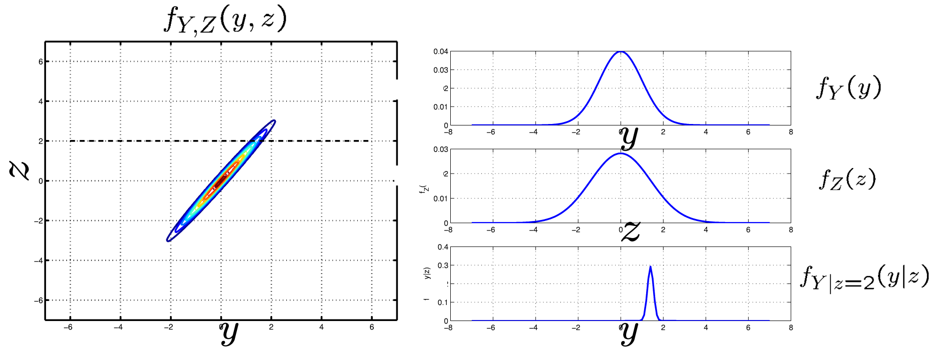
The optimal estimator under squared error criterion is thus the conditional expectation!

TU Delft 24

Bayesian Estimation: MMSE

Example: $Z = Y + N$

with high correlation between random variables Y and Z



With observation $z = 2$, the posterior density $f_{Y|Z}(y|z)$ is very concentrated around $\hat{y} = E\{Y|z = 2\} \approx 1.5$.

TU Delft

Minimum mean square estimation

- When you know the joint pdf

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

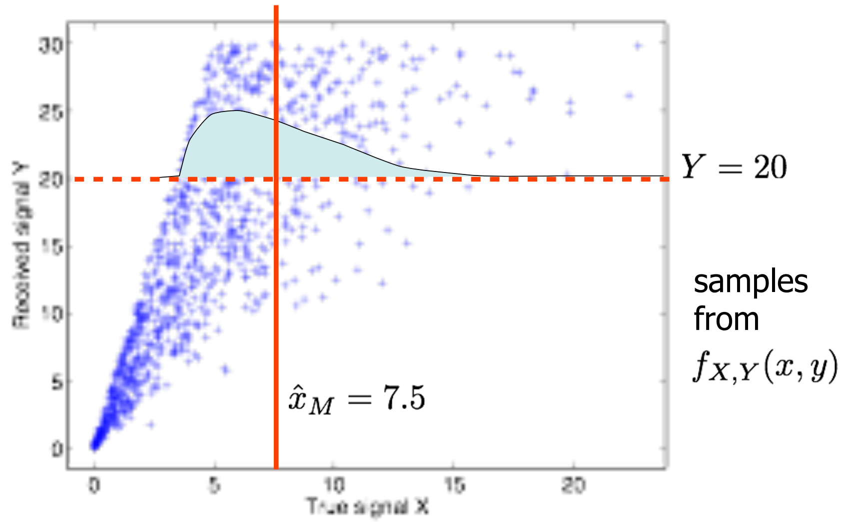
- then the minimum mean square error estimate of X given the observation Y=y is

$$\begin{aligned}\hat{x}_M &= E[X|Y=y] \\ &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx \\ &= \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)}dx\end{aligned}$$

26

TU Delft
Delft University of Technology

Example



Linear estimation of X given Y

- But but... we often do not have $f_{X,Y}(x,y)$?!
- What can we do, if we only have

$$E[X],\, E[Y],\, Var[X],\, Var[Y],\, Cov[X,Y] \text{ ??}$$

- Well, we can do linear estimation:

$$\hat{x}_L(y) = ay + b$$

Linear estimation of X given Y

- We want to minimize the mean square error:

$$e_L = E\left[(X - \hat{x}_L(Y))^2\right]$$

- Let's be brave, and expand:

$$= E\left[(X - aY - b)^2\right]$$
$$= E[X^2] - 2aE[XY] - 2bE[X] + a^2E[Y^2] + 2abE[Y] + b^2$$

- When is the error minimal?
  Set the derivative to 0, solve a and b.

Linear estimation of X given Y

$$e_L = E[X^2] - 2aE[XY] - 2bE[X] + a^2E[Y^2] + 2abE[Y] + b^2$$

- Set derivative to zero:

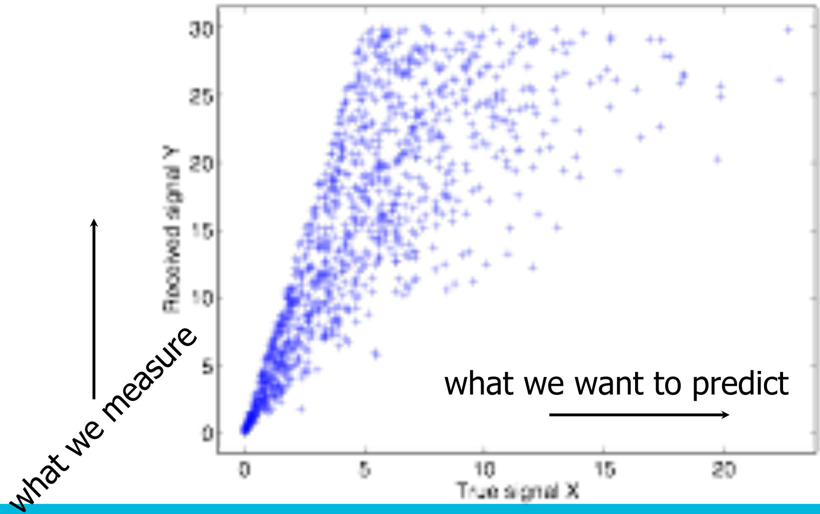
$$\frac{\partial e_L}{\partial a} = -2E[XY] + 2aE[Y^2] + 2bE[Y] = 0$$
$$\frac{\partial e_L}{\partial b} = -2E[X] + 2aE[Y] + 2b = 0$$

- Solve:

$$a^* = \frac{Cov[X,Y]}{Var[Y]} \qquad b^* = E[X] - a^*E[Y]$$

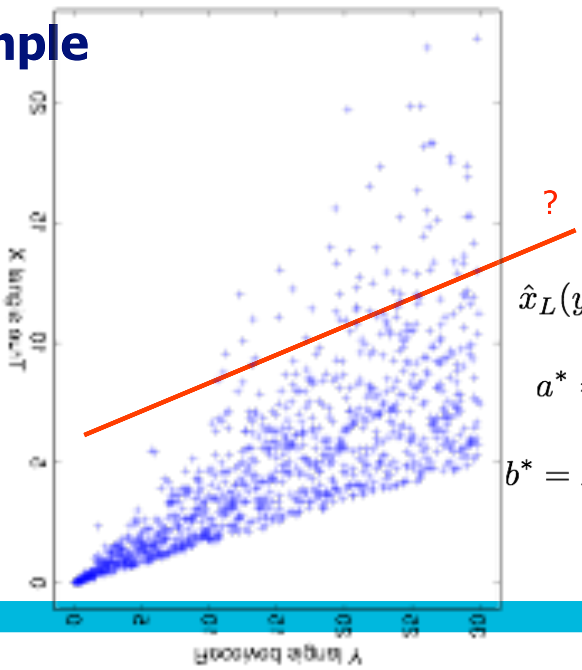
Example

$$\hat{x}_L(y) = ay + b$$



TU Delft
Delft University of Technology

Example



$\hat{x}_M$

$$\hat{x}_L(y) = a^* y + b^*$$

$$a^* = \frac{Cov[X,Y]}{Var[Y]}$$

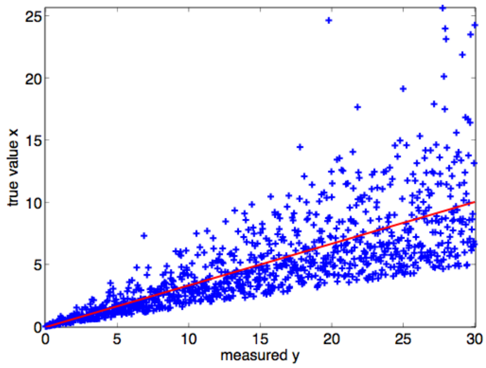
$$b^* = E[X] - a^* E[Y]$$

TU Delft
Delft University of Technology

Example $\hat{x}_L(y) = a^* y + b^*$

- From the given examples, I estimated:

$$Cov[X,Y] = 25.2$$
$$Var[Y] = 75.2$$
$$E[X] = 5.0$$
$$E[Y] = 15.0$$



$$a^* = \frac{Cov[X,Y]}{Var[Y]} = 0.33$$
$$b^* = E[X] - a^* E[Y] = -0.05$$

33

TUDelft
Delft University of Technology

Bayesian Estimation: Example

Assume that we observe realizations of y and that we want to estimate x.

The joint density of x and y is given by:

$$f_{X,Y}(x,y) = \begin{cases} 10x & 0 \leq x \leq y^2 \quad 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We need the conditional density $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$.

Therefore, we first compute marginal $f_Y(y)$.

34

TUDelft

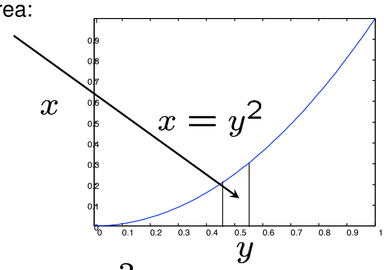
Bayesian Estimation: Example

$$f_{X,Y}(x,y) = \begin{cases} 10x & 0 \le x \le y^2 \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

computing the marginal $f_Y(y)$:

To compute $f_Y(y)$, integrate $f_{X,Y}(x,y)$ over this area:

$$f_Y(y) = \int_0^{y^2} 10x\,dx = 5y^4$$



$$f_{X|Y}(x|y) = \frac{10x}{5y^4} = \frac{2x}{y^4} \quad 0 \le x \le y^2$$

TU Delft 35

Bayesian Estimation: Example

MMSE estimator:

For the MMSE estimator we need to compute the consitional expectation $E_X[X|y]$.

$$\begin{aligned} E_X[X|y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx \\ &= \int_0^{y^2} \frac{2x^2}{y^4}\,dx = \frac{2}{3}y^2. \end{aligned}$$

Thus, $\hat{x}_{\text{MMSE}}(y) = \frac{2}{3}y^2$.

TU Delft 36

Linear MMSE estimation of a RV

An easier way to derive an estimator for a specific RV is to constrain the estimator to have a linear dependence on the observable random variable:

$$\widehat{X}_{\text{lin}}(Y) = aY + b$$

TU Delft 37

Linear MMSE estimation of a RV

Setting

$$\frac{\partial E\left[(X - \widehat{X}_{\text{lin}}(Y))^2\right]}{\partial a} = 0$$

and

$$\frac{\partial E\left[(X - \widehat{X}_{\text{lin}}(Y))^2\right]}{\partial b} = 0$$

and solving for a and b then leads to

$$\widehat{X}_{\text{lin}(Y)} = \frac{E[XY] - E[Y]E[X]}{E\left[Y^2\right] - E[Y]^2}(Y - E[Y]) + E[X]$$

TU Delft 38

Example: the linear MMSE estimator (1)

Remember the previous example:

$$f_{X,Y}(x,y) = \begin{cases} 10x & 0 \le x \le y^2 \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What would $\hat{X}_{\text{lin}(Y)}$ in this case be?

We already know that $f_Y(y) = \int_0^{y^2} 10x\,dx = 5y^4$

$f_X(x)$ is given by: $f_X(x) = \int_{\sqrt{x}}^{1} 10x\,dy = 10(x - x^{3/2})$

TU Delft 19

Example: the linear MMSE estimator (2)

compute $E[X]$, $E[Y]$, $E[Y^2]$ and $E[XY]$

- $E[X] = \int_0^1 x f_X(x)dx = 10/21$
- $E[Y] = \int_0^1 y f_Y(y)dy = 5/6$
- $E[Y^2] = \int_0^1 y^2 f_Y(y)dy = 5/7$
- $E[XY] = \int_0^1 xy f_{X,Y}(x,y)dxdy = 10/24$

Substitution in

$$\hat{X}_{\text{lin}(Y)} = \frac{E[XY] - E[Y]E[X]}{E\left[Y^2\right] - E[Y]^2}(Y - E[Y]) + E[X]$$

TU Delft 20

Example: the linear MMSE estimator (3)

Substitution in

$$\hat{X}_{\text{lin}}(Y) = \frac{E[XY] - E[Y]E[X]}{E\left[Y^2\right] - E[Y]^2}(Y - E[Y]) + E[X]$$

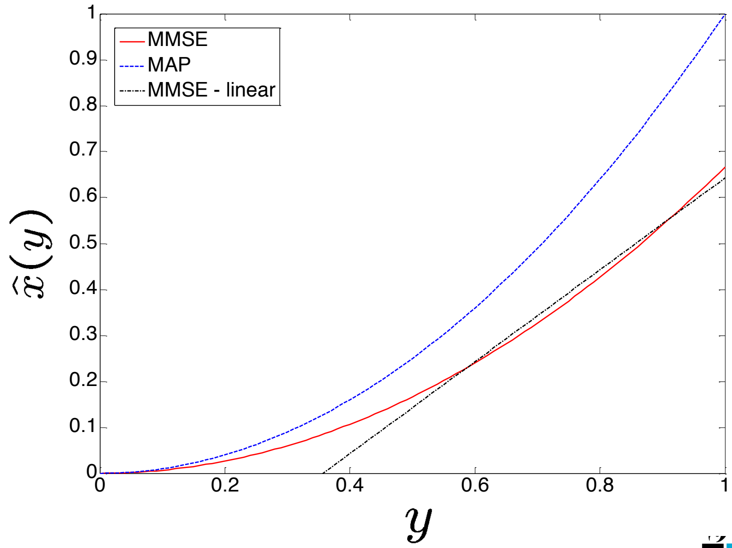
leads to:

$$\hat{X}_{\text{lin}}(Y) = Y - 5/14$$

How does this linear estimator compare with the MAP non-linear MMSE estimator?

TU Delft

Bayesian Estimation: Example



TU Delft

Linear estimation of X given Y

- And when we don't have a single Y, but a vector **Y**?

$$(y_{11}, .., y_{1d}) \to x_1$$
$$(y_{21}, .., y_{2d}) \to x_2$$
$$\vdots$$
$$(y_{N1}, .., y_{Nd}) \to x_N$$

- Linear estimation:

$$\hat{x}_L = a_0 y_0 + a_1 y_1 + ... + a_{n-1} y_{n-1}$$

- In vector notation:

$$\hat{X}_L(y) = \mathbf{a}^T \mathbf{Y}$$

43

TUDelft
Delft University of Technology

Linear estimation from vector Y

- Fill in the predictor $\hat{X}_L(y) = \mathbf{a}^T \mathbf{Y}$
  in the mean square error, set derivative to 0:

$$e_L = E\left[(X - a_0 Y_0 - a_1 Y_1 - ... - a_{n-1} Y_{n-1})^2\right]$$
$$\frac{\partial e_L}{\partial a_i} = 2E\left[Y_i (X - a_0 Y_0 - a_1 Y_1 - ... - a_{n-1} Y_{n-1})\right] = 0$$

$$E[XY_i] = a_0 E[Y_i Y_0] + a_1 E[Y_i Y_1] + ... + a_{n-1} E[Y_i Y_{n-1}]$$

$$r_{Y_i X} = a_0 r_{Y_i Y_0} + a_1 r_{Y_i Y_1} + ... + a_{n-1} r_{Y_i Y_{n-1}}$$

$$\mathbf{R}_{\mathbf{Y}X} = \mathbf{R}_{\mathbf{Y}} \mathbf{a} \qquad \to \mathbf{a} = \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{R}_{\mathbf{Y}X}$$

44

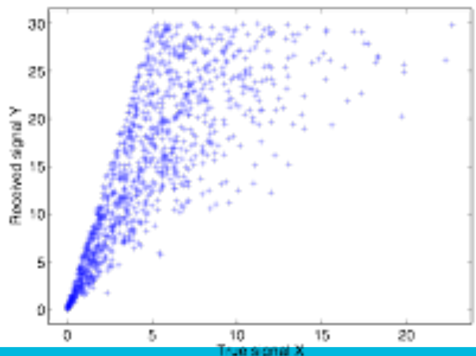
TUDelft
Delft University of Technology

Prediction from samples

TUDelft
Delft University of Technology

Estimating Expected Values

- And, how do I get E(X)? And Cov[X,Y], and ... when I do not know $f_{X,Y}(x,y)$?
- We need some realizations, some **examples**



(y_1, x_1)
(y_2, x_2)
(y_3, x_3)
$\dots$
(y_N, x_N)

TUDelft
Delft University of Technology

Sample mean, discrete case

$$E[X] = \sum_{k=1}^{M} x_k P[X = x_k]$$

(Sample n realizations from P_X)

$$\simeq \sum_{k=1}^{M} x_k \frac{n_k}{n}$$

$$= \frac{1}{n}\left(\underbrace{(x_1 + \ldots + x_1)}_{n_1} + \underbrace{(x_2 + \ldots + x_2)}_{n_2} + \ldots\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

result of the i-th repeat, consider as random var.

47

TUDelft Delft University of Technology

Expected value of sample mean

- Is the sample mean a good estimator?

48

TUDelft Delft University of Technology

Expected value of sample mean

- Sums in general

$$W_n = X_1 + X_2 + ... + X_n$$

- Expected value

$$E[W_n] = E[X_1] + E[X_2] + ... + E[X_n] = nE[X]$$

- Sample mean $M_n = \frac{1}{n} W_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

- Expected value of sample mean=expected value of X

$$E[M_n] = \frac{1}{n}(nE[X]) = \mu_X$$

Hurray!

49

TUDelft
Delft University of Technology

Analysis of sample mean

- That is good news, we can use the sample mean to estimate the expected value!
- To analyze M_n further we look more general to:

$$|Y - \mu_Y|$$

- How does a random variable differ from its expectation?

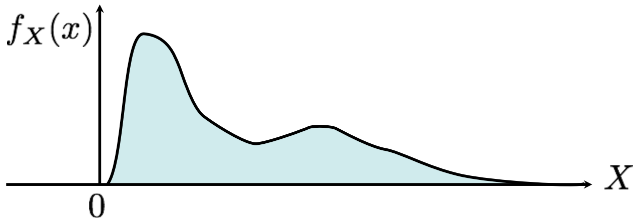
- We introduce the Chebyshev inequality, but for that we need the Markov inequality...

50

TUDelft
Delft University of Technology

Markov Inequality

- Assume we have a random variable X:

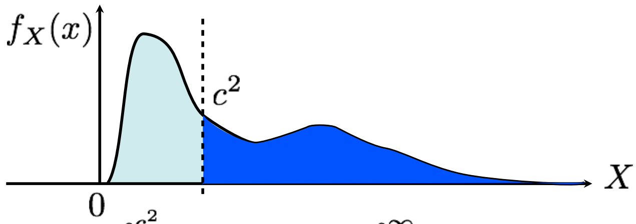


- Then you can show:

$$P[X \geq c^2] \leq \frac{E[X]}{c^2}$$

Markov Inequality

- How to show that $P[X \geq c^2] \leq \dfrac{E[X]}{c^2}$?

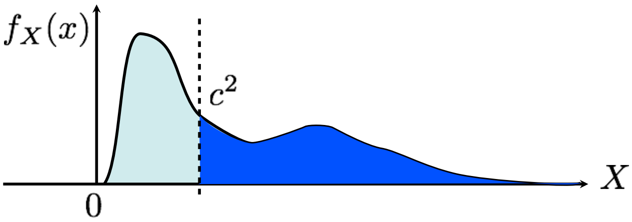


- Well: $E[X] = \int_0^{c^2} x f_X(x) dx + \int_{c^2}^{\infty} x f_X(x) dx$

$$\geq \int_{c^2}^{\infty} x f_X(x) dx$$

Markov Inequality

- How to show that $P[X \geq c^2] \leq \dfrac{E[X]}{c^2}$?

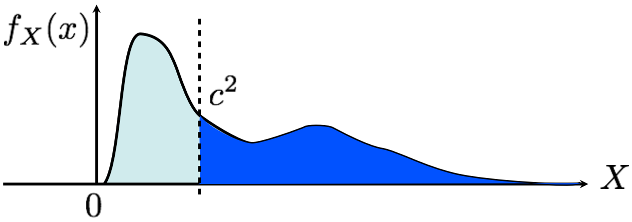


- Because $x \geq c^2$: $E[X] \geq \int_{c^2}^{\infty} x f_X(x) dx$

$$\geq c^2 \int_{c^2}^{\infty} f_X(x) dx$$

Markov Inequality

- How to show that $P[X \geq c^2] \leq \dfrac{E[X]}{c^2}$?



- Because $x \geq c^2$, therefore:

$$E[X] \geq c^2 \int_{c^2}^{\infty} f_X(x) dx = c^2 P[X \geq c^2]$$

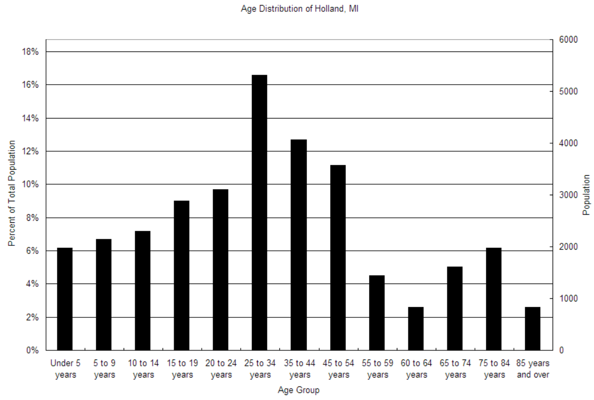
- Tadah!

Example Age distribution of Holland

$$\mu = 36.1$$
$$\sigma^2 = 541$$

From the graph:

$$P[X > 64] = 0.14$$



- What is the probability that somebody is older than 64?

$$P[X \geq 65] \leq \frac{E[X]}{65} = \frac{36.1}{65} \approx 0.65$$

55

TU Delft
Delft University of Technology

Chebyshev inequality

- Using the Markov inequality $P[X \geq c^2] \leq \dfrac{E[X]}{c^2}$ we can say something about

$$|Y - \mu_Y|$$

- Chebyshev ineq: $P[|Y - \mu_Y| \geq c] \leq \dfrac{Var[Y]}{c^2}$

- Proof: for c>0

$$P[|Y - \mu_Y| \geq c] = P[(Y - \mu_Y)^2 \geq c^2]$$
$$\leq \frac{E[(Y - \mu_Y)^2]}{c^2} = \frac{Var[Y]}{c^2}$$

56

TU Delft
Delft University of Technology

What does it say?

- Looking at the Chebyshev inequality

$$P[|Y - \mu_Y| \geq c] \leq \frac{Var[Y]}{c^2}$$

- The probability that a random variable Y deviates from its expected value directly depends on the variance of that variable
- It does **not** depend on the exact pdf of Y!
- Typically the bound is not tight (but tighter than the Markov inequality)

Variance of the sample mean

- Sums in general

$$Var[W_n] = Var[X_1] + Var[X_2] + ... + Var[X_n] + \\ + \sum_{\text{all } i,j,\, i \neq j} Cov[X_i, X_j]$$

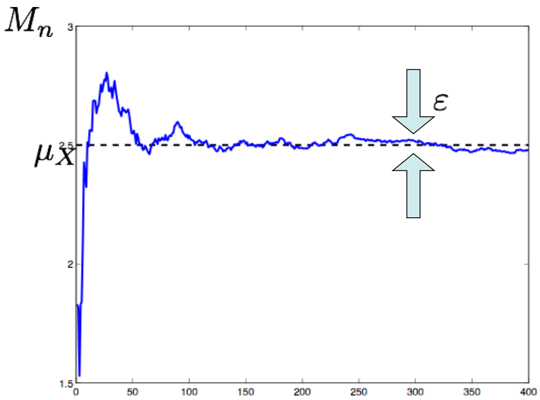
- Sample mean
  - Covariances are 0, because independent experiments:

$$Var[M_n] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ = \frac{1}{n^2}(nVar[X]) = \frac{1}{n}Var[X]$$

Weak law of large numbers

- Good news:

$$\lim_{n \to \infty} P[|M_n - \mu_X| > \varepsilon] = 0$$



TU Delft
Delft University of Technology

So…

- The sample mean $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is good, because:
- $E[M_n] = \mu_X$ (we converge to something useful)
- $P[|M_n - \mu_X| \geq c] \leq \frac{Var[M_n]}{c^2} = \frac{\frac{1}{n}Var[X]}{c^2} \underset{n \to \infty}{\to} 0$

  (we converge pretty fast)

TU Delft
Delft University of Technology

Covered Today

- Chapter 6, 7 and chapter 9.
- Key terms
  - Blind estimation
  - Linear estimation
  - Sample mean
  - Markov and Chebyshev inequality
  - Weak law of large numbers