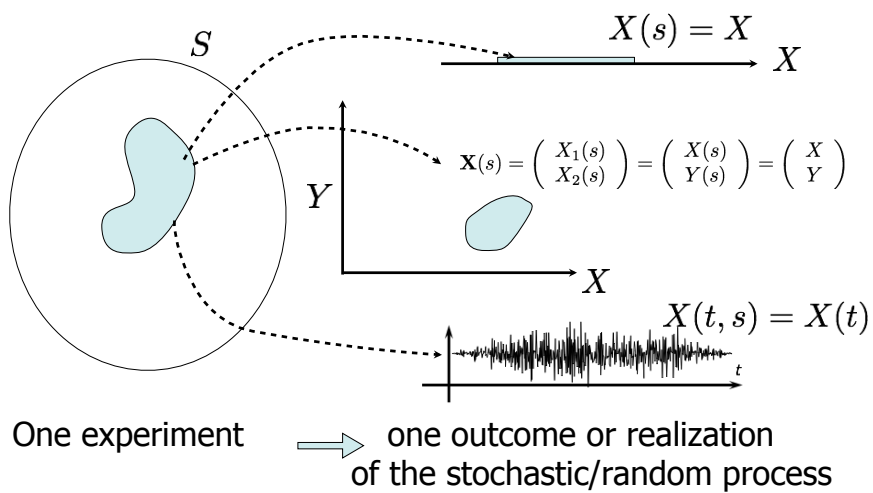


Prediction

Stochastic Processes Lecture 7

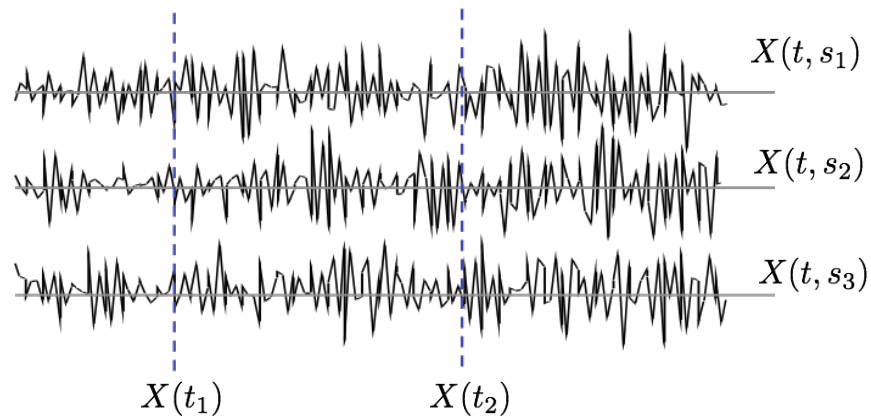
1

Summarizing



2

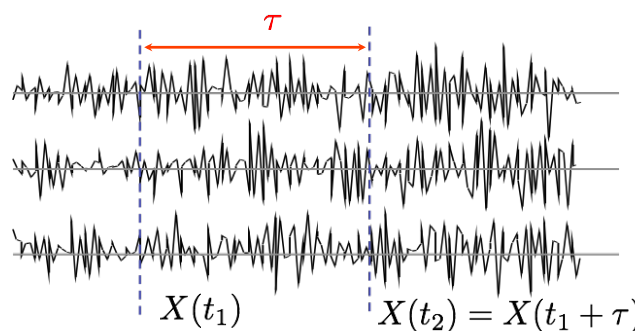
Description random process



3

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



4

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

5

What can we do with Stoch. Proc.?

- Modeling of real signals
- System identification
- Prediction

6

This week: prediction

- Consider random variables
 - Estimating X : blind estimate
 - Minimum mean squared error
 - Linear estimation
- Characteristics of the sample mean
- Markov and Chebychev inequality
- Weak law of large numbers
- Central limit theorem

7

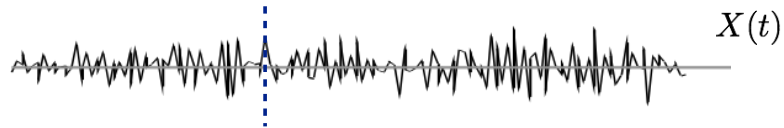
Estimation of a random variable



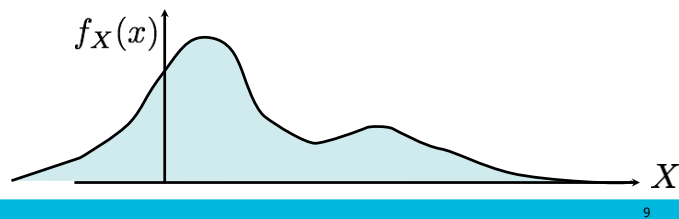
- Most simple case:
- Assume X is WSS, but we know $f_X(x)$.
- Without doing any measurements, can we estimate what X is? What would be our best gamble? (blind estimate)

8

Estimation of a random variable



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

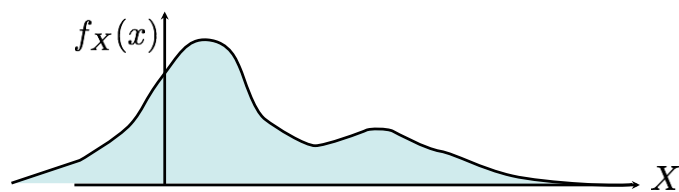


9

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):

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



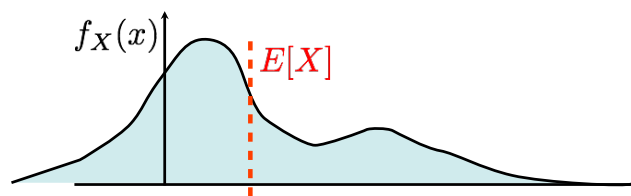
10

Blind estimate

- Minimize the MSE $e = E[(X - \hat{x}_B)^2]$

$$= E[X^2] - 2\hat{x}_B E[X] + \hat{x}_B^2$$
- Set derivative to zero, and solve:

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

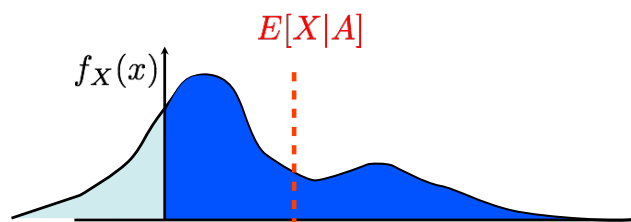


11

Estimation given an event

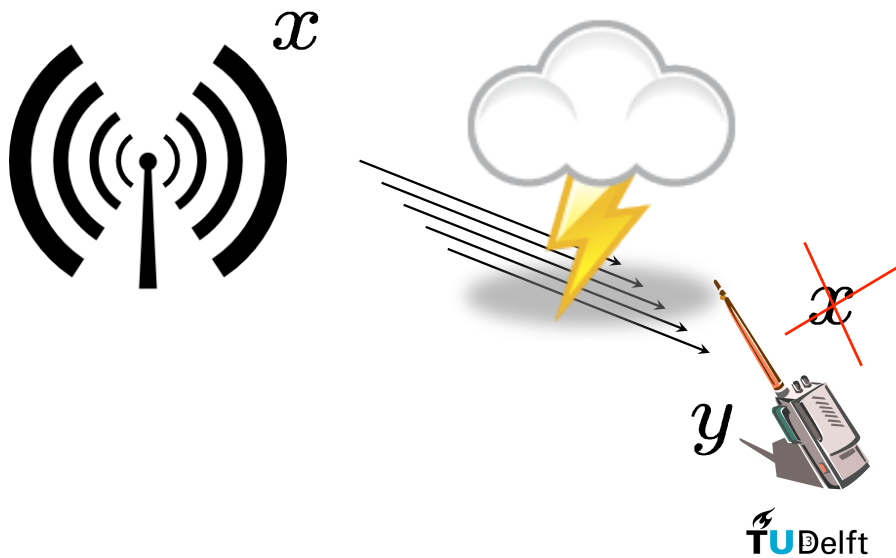
- Now assume we know that X comes from an event A

$$A : \{X \geq 0\}$$
- In an identical manner: $\hat{x}_A = E[X|A]$

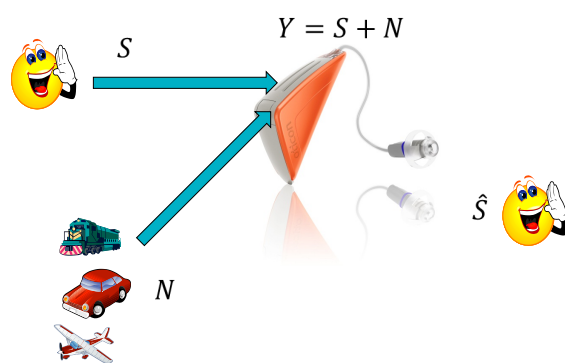


12

Estimate Given a Realization



Estimate Given a Realization



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!



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: $\mathcal{R} = E[\mathcal{C}(X, \hat{X}(Y))]$

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

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



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

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

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 .

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

Bayesian Estimation: MMSE

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

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

Bayesian Estimation: MMSE

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

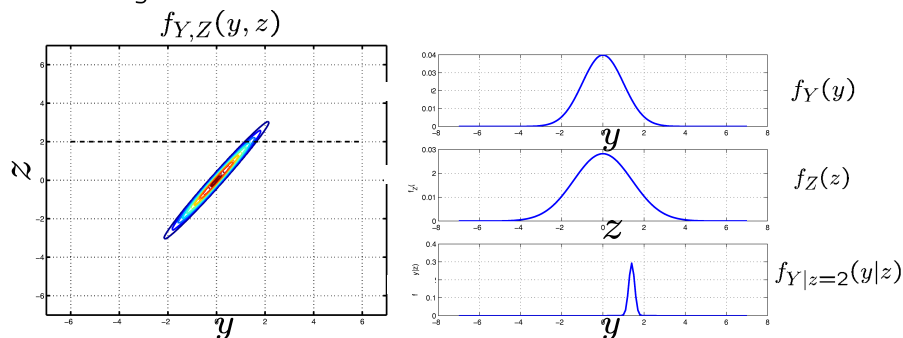
$$\hat{x}_{\text{MMSE}}(y) = E[X|y]$$

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

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 $\hat{y} = E\{Y|z = 2\} \approx 1.5$.

Minimum mean square estimation

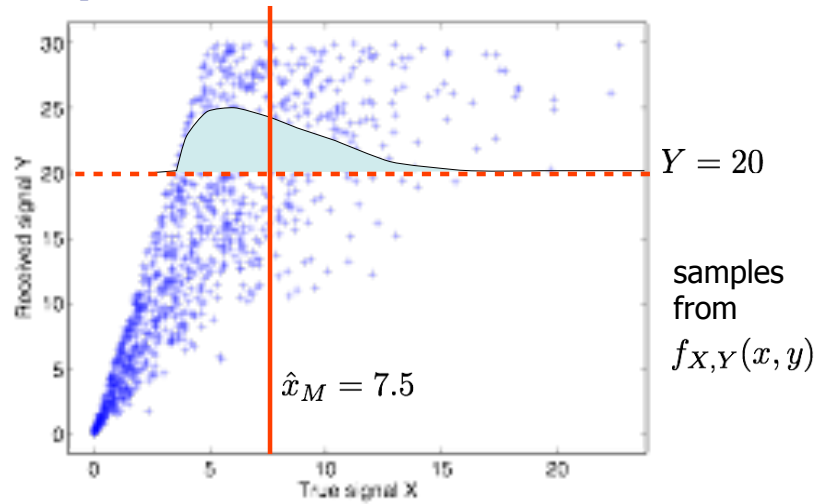
- When you know the joint pdf

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

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

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

Example



23

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

$$E[X], E[Y], Var[X], Var[Y], Cov[X, Y]$$

??

- Well, we can do linear estimation:

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

24

Linear estimation of X given Y

- We want to minimize the mean square error:

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

- Let's be brave, and expand:

$$= E[(X - aY - b)^2]$$

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

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

25

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:

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

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

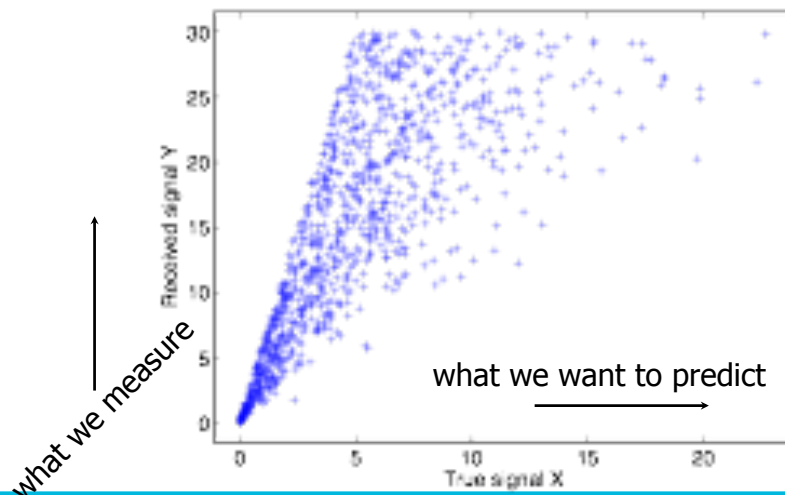
- Solve:

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

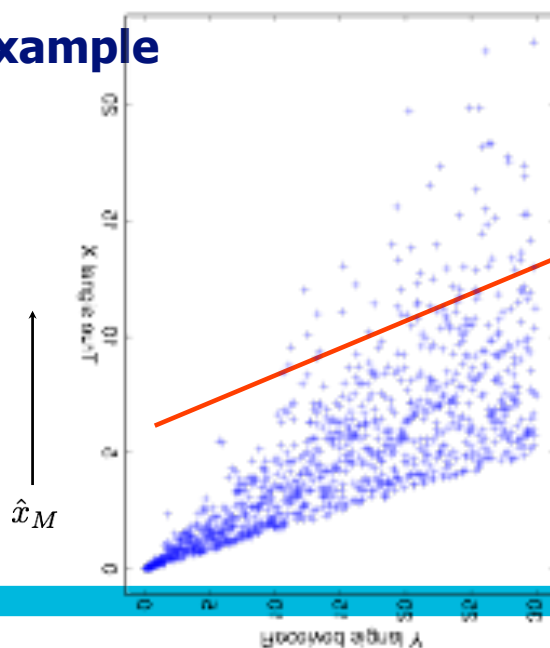
26

Example

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



27

Example

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

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

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

28

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

- From the given examples, I estimated:

$$\text{Cov}[X, Y] = 25.2$$

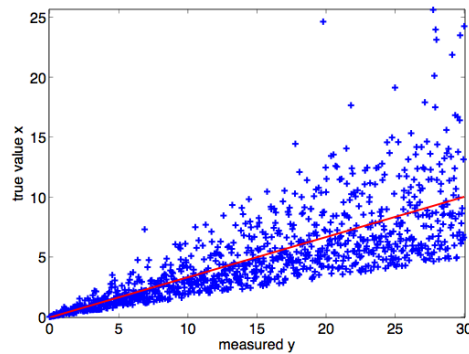
$$\text{Var}[Y] = 75.2$$

$$E[X] = 5.0$$

$$E[Y] = 15.0$$

$$a^* = \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} = 0.33$$

$$b^* = E[X] - a^*E[Y] = -0.05$$



29

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:

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

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

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

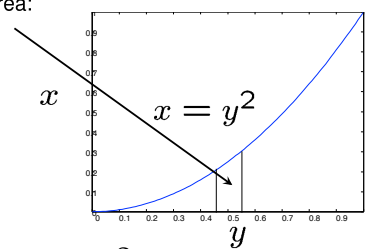
Bayesian Estimation: Example

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

computing the marginal $f_Y(y)$:

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

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



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

Bayesian Estimation: Example

MMSE estimator:

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

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

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

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:

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



Linear MMSE estimation of a RV

Setting

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

and

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

and solving for a and b then leads to

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



Example: the linear MMSE estimator (1)

Remember the previous example:

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

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



Example: the linear MMSE estimator (2)

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

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

Substitution in

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



Example: the linear MMSE estimator (3)

Substitution in

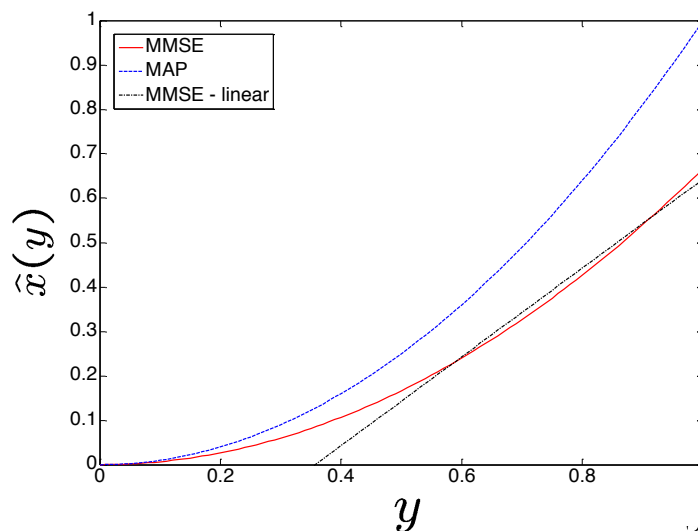
$$\hat{X}_{\text{lin}}(Y) = \frac{E[XY] - E[Y]E[X]}{E[Y^2] - E[Y]^2} (Y - E[Y]) + E[X]$$

leads to:

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

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

Bayesian Estimation: Example



Linear estimation of X given Y

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

$$(y_{11}, \dots, y_{1d}) \rightarrow x_1$$

$$(y_{21}, \dots, y_{2d}) \rightarrow x_2$$

$$\vdots$$

- Linear estimation: $(y_{N1}, \dots, y_{Nd}) \rightarrow x_N$

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

- In vector notation:

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

39

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:

$$e_L = E[(X - a_0 Y_0 - a_1 Y_1 - \dots - a_{n-1} Y_{n-1})^2]$$

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

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

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

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

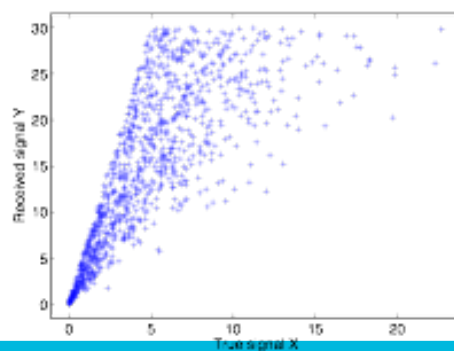
40

Prediction from samples

41

Estimating Expected Values

- And, how do I get $E(X)$? And $\text{Cov}[X, Y]$, 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)

42

Sample mean, discrete case

$$\begin{aligned}
 E[X] &= \sum_{k=1}^M x_k P[X = x_k] \\
 &\simeq \sum_{k=1}^M x_k \frac{n_k}{n} \quad (\text{Sample } n \text{ realizations from } P_X) \\
 &= \frac{1}{n} \left(\underbrace{(x_1 + \dots + x_1)}_{n_1} + \underbrace{(x_2 + \dots + x_2)}_{n_2} + \dots \right) \\
 &= \frac{1}{n} \sum_{i=1}^n x_i \quad \text{result of the } i\text{-th repeat, consider as random var.}
 \end{aligned}$$

43

Expected value of sample mean

- Is the sample mean a good estimator?

44

Expected value of sample mean

- Sums in general

$$W_n = X_1 + X_2 + \dots + X_n$$

- Expected value

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

- 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!

45

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:

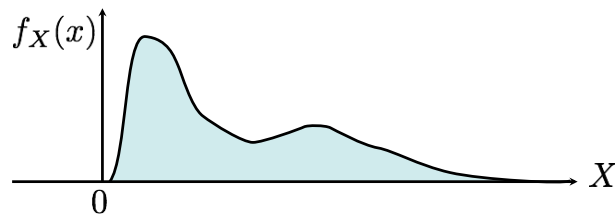
$$|Y - \mu_Y|$$

- How does a random variable differ from its expectation?
- We introduce the Chebyshev inequality, but for that we need the Markov inequality...

46

Markov Inequality

- Assume we have a random variable X :



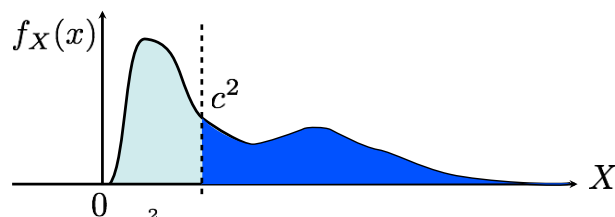
- Then you can show:

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

47

Markov Inequality

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

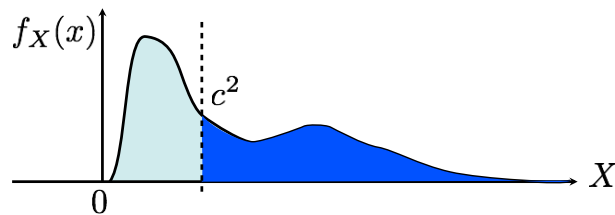


- Well:
$$E[X] = \int_0^{c^2} x f_X(x) dx + \int_{c^2}^{\infty} x f_X(x) dx$$
$$\geq \int_{c^2}^{\infty} x f_X(x) dx$$

48

Markov Inequality

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

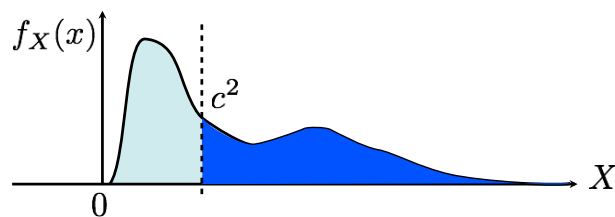


- Because $x \geq c^2$:
$$E[X] \geq \int_{c^2}^{\infty} x f_X(x) dx$$
$$\geq c^2 \int_{c^2}^{\infty} f_X(x) dx$$

49

Markov Inequality

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



- 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!

50

Example Age distribution of Holland

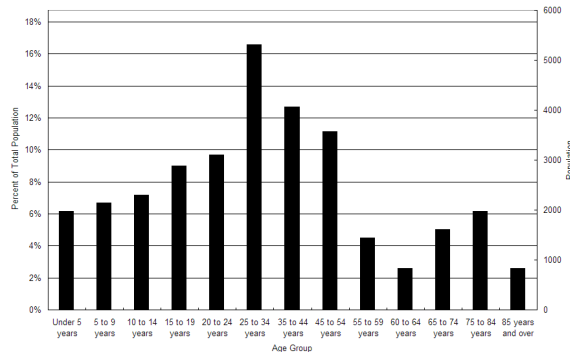
Age Distribution of Holland, MI

$$\mu = 36.1$$

$$\sigma^2 = 541$$

From the graph:

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



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

$$P[X \geq 65] \leq \frac{E[X]}{65} = \frac{36.1}{75} = 0.55$$

51

Chebyshev inequality

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

$$|Y - \mu_Y|$$

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

- Proof: for $c > 0$

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

52

What does it say?

- Looking at the Chebyshev inequality

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

- 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
- This can be used in Crime Scene Investigations.....

53

Variance of the sample mean

- Sums in general

$$\begin{aligned} \text{Var}[W_n] &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] + \\ &+ \sum_{\text{all } i, j, i \neq j} \text{Cov}[X_i, X_j] \end{aligned}$$

- Sample mean

- Covariances are 0, because independent experiments:

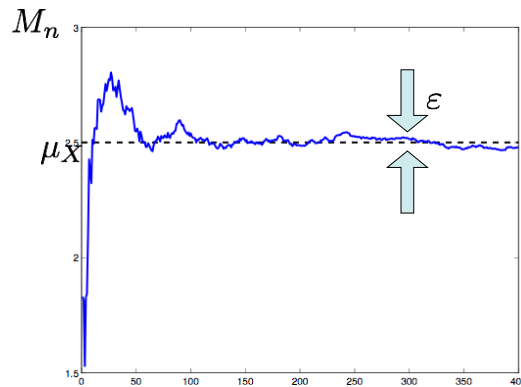
$$\begin{aligned} \text{Var}[M_n] &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] \\ &= \frac{1}{n^2} (n \text{Var}[X]) = \frac{1}{n} \text{Var}[X] \end{aligned}$$

54

Weak law of large numbers

- Good news:

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



55

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)
- $P[|M_n - \mu_X| \geq c] \leq \frac{\text{Var}[M_n]}{c^2} = \frac{\frac{1}{n} \text{Var}[X]}{c^2} \xrightarrow{n \rightarrow \infty} 0$
(we converge pretty fast)

56

Probability density function of M_n

- If X_1 and X_2 are independent random variables, then for $W_2 = X_1 + X_2$

$$f_{W_2}(w) = f_{X_1}(w) * f_{X_2}(w)$$

- For $n > 2$, this can be generalized to

$$f_{W_n}(w) = f_{X_1}(w) * f_{X_2}(w) * \dots * f_{X_n}(w)$$

57

Central Limit Theorem

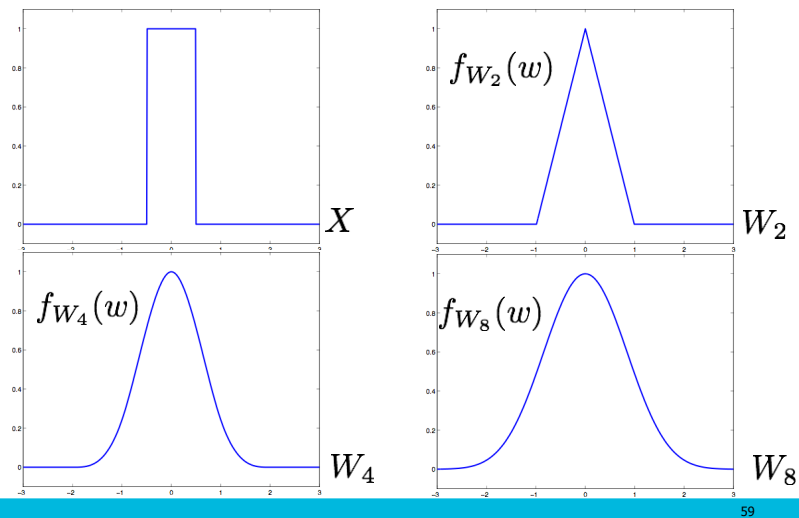
- When X_i are iid random variables
 - with expected value μ_X
 - and variance σ^2
- then for $W_n = X_1 + X_2 + \dots + X_n$

$$f_{W_n}(w) = f_{X_1}(w) * f_{X_2}(w) * \dots * f_{X_n}(w)$$

- can be approximated by a Gaussian pdf
 - with expected value $n\mu_X$
 - and variance $n\sigma_X^2$

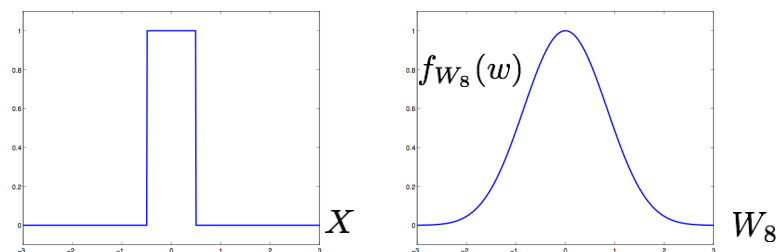
58

Illustration Central Limit Theorem



59

Illustration Central Limit Theorem



- PDF of the sample mean

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

can be approximated by a Gaussian with an expected value of μ_X and variance $\frac{\sigma_X^2}{n}$

60

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
 - Central Limit Theorem

61