EE2S31 Signal Processing – Stochastic Processes

Lecture 4: Estimation of a RV – Ch. 12

Alle-Jan van der Veen

–

4 May 2022

Today: Ch. 12 Estimation of a RV

We want to estimate the realization of a random variable, while the random variable itself cannot be observed, e.g.

- Due to noise
- No available sensor
- Not available in time (e.g., prediction)

What is the best estimate?

Today: Ch. 12 Estimation of a RV

We want to estimate the realization of a random variable, while the random variable itself cannot be observed, e.g.

- Due to noise
- No available sensor
- Not available in time (e.g., prediction)

What is the best estimate?

- Define "best": need cost function (e.g., mean square error)
- What knowledge is available? (data model, observations)
- Feasibility/max complexity of calculations

There are many answers!

This chapter shows 4 estimators: MMSE, MAP, ML, and LMMSE.

Estimation starts with a data model

RV of interest: X; observations $y_1, \ldots, y_N$ of RV Y.

Often we have a **forward** data model, e.g.

- Linear model: $Y = AX + N$
- Autoregressive model: $X_{n+1} = AX_n + V_n$
- State-space model:

$$\begin{cases} X_{n+1} & = & AX_n + Bu_n + V_n \\ Y_n & = & CX_n + Du_n + W_n \end{cases}$$

This naturally connects to the conditional PDF $f_{Y|X}(y|x)$.

But, given an observation $Y = y$, it is also natural to work with $f_{X|Y}(x|y)$. This relates to an **inverse** data model.

Estimation of RVs that can't be observed

Imagine we cannot observe X itself, but we want to estimate it using some related observations (and knowledge on the statistics).

What can we do if we have

- Only knowledge of statistics of X? (Blind estimate)
- Some information about X, e.g., $X \in A$? (e.g., $X \geq 5$)
- Knowledge of a related variable? (e.g., observe $Y = X + N$)

Notation

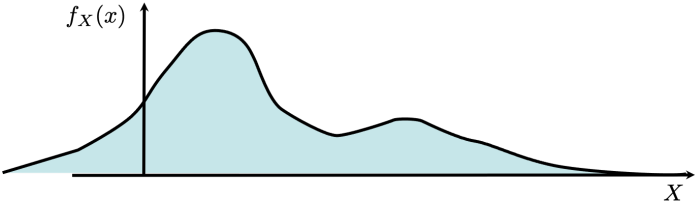
X is the RV of interest, Y is an RV that we can observe.

- If the observation is $Y = y$, then our estimate of the (unobserved) realization of X is a function of y, denoted by $\hat{x}(y)$.
- The same function, but now leaving Y random, is denoted as $\hat{X}(Y)$. This is a random variable
- We can use the PDF $f_Y(y)$ to evaluate expressions for $\hat{X}(Y)$, such as $\mathrm{E}[\hat{X}(Y)]$ or $\mathrm{var}[\hat{X}(Y)]$

Minimum Mean-Squared Error

Without *any* measurements, what can we do?

- Let's assume we know the prior pdf, $f_X(x)$:

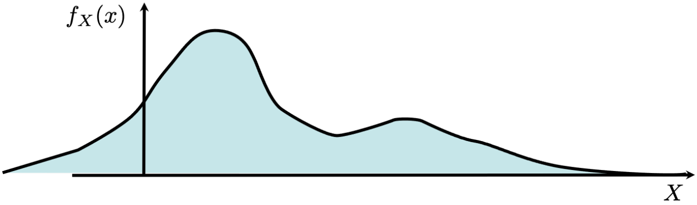


How do we know which value is the best estimate of X?

Minimum Mean-Squared Error

Without *any* measurements, what can we do?

- Let's assume we know the prior pdf, $f_X(x)$:



How do we know which value is the best estimate of X?

- Use a proper distortion measure (cost function).
  A measure that is often minimized is the mean-squared error:

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

MMSE – Blind Estimate

- Define the mean squared error (MSE):

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

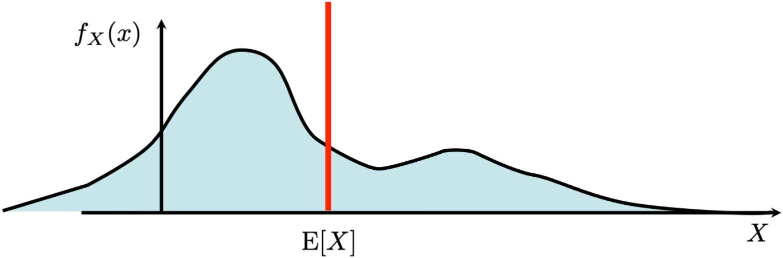
- Minimize MSE by taking derivative to $\hat{x}$ and setting it to zero:

$$\frac{\mathrm{d}e}{\mathrm{d}\hat{x}} = -2\mathrm{E}[X] + 2\hat{x} = 0 \quad \Rightarrow \quad \hat{x} = \mathrm{E}[X]$$

- The minimum MSE is $e = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mathrm{var}[X]$

Is this what we expect?

MMSE — Blind Estimate



- $\mathrm{E}[X]$ is the "most typical" value for X.

  Knowing only $f_X(x)$, taking $\hat{x} = \mathrm{E}[X]$ is the best we can do under MSE.

Expected value of sums of random variables

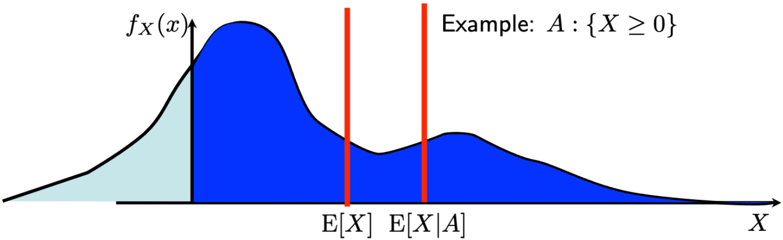
Imagine that we have additional information:

$$X \in A$$

Can we use this information to improve our estimate?

$$e = \mathrm{E}[(X - \hat{x})^2 | A] = \mathrm{E}[X^2 | A] - 2\hat{x}\,\mathrm{E}[X|A] + \hat{x}^2$$

$$\frac{\mathrm{d}e}{\mathrm{d}\hat{x}} = -2\,\mathrm{E}[X|A] + 2\hat{x} = 0 \quad \Rightarrow \quad \hat{x} = \mathrm{E}[X|A]$$



MMSE – Estimation Given a Random Variable

We cannot observe X directly, but we can observe a related random variable Y, e.g.

$$Y = X + N.$$

- In that case, we can make $\hat{X}$ a function of Y, say $\hat{X}(Y)$. Suppose the observation is $Y = y$:

  $$e(y) = \mathrm{E}[(X - \hat{x}(y))^2 | Y = y] = \mathrm{E}[X^2 | Y = y] - 2\hat{x}(y)\,\mathrm{E}[X | Y = y] + \hat{x}^2(y)$$

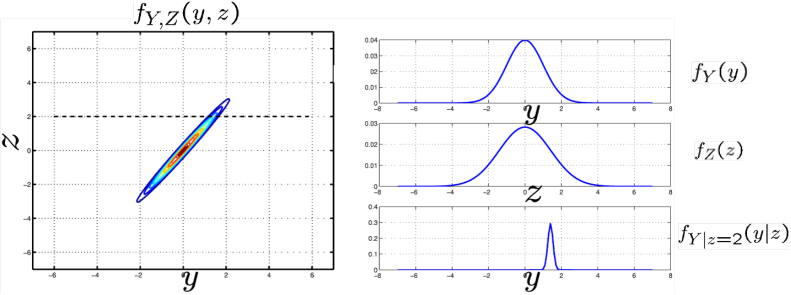
  $$\frac{\partial e(y)}{\partial \hat{x}(y)} = -2\,\mathrm{E}[X|Y = y] + 2\hat{x}(y) = 0 \quad \Rightarrow \quad \hat{x}(y) = \mathrm{E}[X|Y = y]$$

- We can also say $\hat{X}(Y) = \mathrm{E}[X|Y]$: a random variable
- The minimum MSE is
  $e(y) = \mathrm{E}[X^2|Y = y] - \mathrm{E}[X|Y = y]^2 = \mathrm{var}[X|Y = y]$

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

Problem 12.1.6

A signal X and noise Z are independent Gaussian(0,1) random variables, and $Y = X + Z$ is a noisy observation of the signal X.

Usually, we want to use Y to estimate X; however, in this problem we will use Y to estimate the noise Z.

- Find $\hat{Z}(Y)$, the MMSE estimator of Z given Y

Problem 12.1.6

A signal X and noise Z are independent Gaussian(0,1) random variables, and $Y = X + Z$ is a noisy observation of the signal X.

Usually, we want to use Y to estimate X; however, in this problem we will use Y to estimate the noise Z.

- Find $\hat{Z}(Y)$, the MMSE estimator of Z given Y

The MMSE estimator of Z given Y is always the conditional expectation. Random variables Y and Z are jointly (bivariate) Gaussian. Then (see Thm. 7.15)

$$\hat{Z}(Y) = \mathrm{E}[Z|Y] = \rho_{ZY} \frac{\sigma_Z}{\sigma_Y}(Y - \mu_Y) + \mu_Z$$

Problem 12.1.6 (cont'd)

In this case, it is given that $\mu_Z = 0$, $\sigma_Z = 1$, and

$$\mu_Y = \mathrm{E}[Y] = \mathrm{E}[X] + \mathrm{E}[Z] = 0$$

Since X and Z are independent,

$$\begin{aligned} \sigma_Y^2 &= \sigma_X^2 + \sigma_Z^2 = 2 \\ \rho_{Z,Y} &= \frac{\mathrm{cov}[Z, Y]}{\sigma_Y \sigma_Z} \\ &= \frac{\mathrm{E}[Z(X+Z)] - 0}{\sigma_Y \sigma_Z} \\ &= \frac{\mathrm{E}[Z]\mathrm{E}[X] + \mathrm{E}[Z^2]}{\sigma_Y \sigma_Z} = \frac{\sigma_Z}{\sigma_Y} = \frac{1}{\sqrt{2}} \end{aligned}$$

It follows that

$$\hat{Z}(Y) = \mathrm{E}[Z|Y] = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} (Y - 0) + 0 = \frac{Y}{2}$$

(Not intuitive: why not Y? (See also Pr. 12.2.9))

Problem 12.1.6 (cont'd; cf. Pr. 12.1.7)

Now derive this from first principles:

To compute $\mathrm{E}[Z|Y]$, first compute $f_{Z|Y}(z|y)$.

- Forward model: given $Z = z$, then $Y = X + z$, hence (Thm. 6.4)

$$f_{Y|Z}(y|z) = f_X(y - z)$$

- Use Bayes:

$$f_{Z|Y}(z|y) = \frac{f_{Y|Z}(y|z)\, f_Z(z)}{f_Y(y)} = \frac{f_X(y-z) f_Z(z)}{\int f_X(y-z)\, f_Z(z) \mathrm{d}z}$$

$$\mathrm{E}[Z|Y = y] = \int z\, f_{Z|Y}(z|y) \mathrm{d}z = \frac{\int z\, f_X(y-z) f_Z(z) \mathrm{d}z}{\int f_X(y-z) f_Z(z) \mathrm{d}z}$$

Problem 12.1.6 (cont'd)

- Insert Gaussian models for X and Z: (for normalization constant c)

$$\hat{z}(y) = \mathrm{E}[Z|Y=y] = \frac{\int z\, \cancel{c}\, e^{-(y-z)^2/2} e^{-z^2/2} \mathrm{d}z}{\int \cancel{c}\, e^{-(y-z)^2/2} e^{-z^2/2} \mathrm{d}z}$$

$$= \frac{\int z\, e^{-(z-\frac{1}{2}y)^2} \cancel{e^{-y^2/4}} \mathrm{d}z}{\int e^{-(z-\frac{1}{2}y)^2} \cancel{e^{-y^2/4}} \mathrm{d}z}$$

$$= \frac{\int z\, c'\, e^{-(z-\frac{1}{2}y)^2} \mathrm{d}z}{1}$$

This evaluates to the expected value of a Gaussian$(\frac{1}{2}y, \frac{1}{2})$.
Thus, $\hat{z}(y) = \mathrm{E}[Z|Y=y] = \frac{1}{2}y$. The mean square error is $e^* = \mathrm{var}[Z|Y] = \frac{1}{2}$.

Generalization: Bayesian estimation of a RV

Let's do this for more general cost functions.

- Define a non-negative cost function $C(X, \hat{X}(Y))$, e.g.

$$C(X, \hat{X}(Y)) = \left(X - \hat{X}(Y)\right)^2$$

- Minimize expected costs: $e = \mathrm{E}[C(X, \hat{X}(Y))]$

Since both X and $\hat{X}(Y)$ are random variables, we can express e as

$$e = \int \int C(x, \hat{x}(y)) \, f_{X,Y}(x, y) \, \mathrm{d}x\mathrm{d}y$$

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

$$e = \int \underbrace{\int C(x,\hat{x}(y)) f_{X|Y}(x|y)\mathrm{d}x}_{e(y)}\, f_Y(y)\mathrm{d}y$$

Notice that

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

Bayesian estimation of a RV

We can simplify our problem:

- To minimize e, it is sufficient to minimize

$$e(y) = \int C(x, \hat{x}(y))\, f_{X|Y}(x|y)\, \mathrm{d}x$$

for each realization (observation) y:

$$\hat{x}(y) = \arg\min_{\hat{x}(y)} \int C(x, \hat{x}(y))\, f_{X|Y}(x|y)\, \mathrm{d}x$$

- For the MMSE cost function we find

$$\begin{aligned}\frac{\mathrm{d}e(\hat{x}(y))}{\mathrm{d}\hat{x}(y)} &= \frac{\mathrm{d}}{\mathrm{d}\hat{x}(y)} \int (x - \hat{x}(y))^2 f_{X|Y}(x|y)\, \mathrm{d}x \\ &= -2 \int (x - \hat{x}(y))\, f_{X|Y}(x|y)\, \mathrm{d}x \\ &= 0\end{aligned}$$

Bayesian estimation of a RV

$$\Leftrightarrow \quad \int \hat{x}(y) f_{X|Y}(x|y) \mathrm{d}x = \int x \, f_{X|Y}(x|y) \, \mathrm{d}x$$

$$\hat{x}(y) \underbrace{\int f_{X|Y}(x|y) \mathrm{d}x}_{1} = \mathrm{E}[X|Y = y]$$

The result is (again)

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

- The optimal estimator under the squared error condition (the MMSE) is the conditional expectation
- What about other cost functions?

Bayesian Estimation: MAP

Estimators that are derived using the **uniform cost function** are often referred to as **maximum a posteriori (MAP) estimators**:

$$C(x, \hat{x}(y)) = \begin{cases} 0 & |x - \hat{x}(y)| < \epsilon, \\ 1 & \text{otherwise} \end{cases}$$

- Finding the MAP estimator:

$$\begin{aligned} \min_{\hat{x}(y)} e(y) &= \min_{\hat{x}(y)} \int C(x, \hat{x}(y))\, f_{X|Y}(x|y)\, \mathrm{d}x \\ &= \min_{\hat{x}(y)} \int_{|x-\hat{x}(y)|\geq\epsilon} f_{X|Y}(x|y)\, \mathrm{d}x \\ &= \min_{\hat{x}(y)} 1 - \int_{|x-\hat{x}(y)|<\epsilon} f_{X|Y}(x|y)\, \mathrm{d}x \end{aligned}$$

Maximum A Posteriori (MAP) Estimator

Because the integral is over an arbitrarily small region around $\hat{x}(y)$, $e(y)$ is minimized when the PDF $f_{X|Y}(x|y)$ is maximized;

$$\hat{x}_{\text{MAP}}(y) = \arg\max_{x} \ f_{X|Y}(x|y)$$

- The MAP estimate $\hat{x}(y)$ is the value of x that maximizes the conditional density $f_{X|Y}(x|y)$.

  The name **maximum a posteriori** is derived from the fact that the density $f_{X|Y}(x|y)$ is often called the posterior density.

- Similarly, the **blind** estimator for this cost function is

$$\hat{x} = \arg\max_{x} \ f_X(x)$$

  which is the *mode* of the PDF of X.

Maximum Likelihood Estimator

Notice that

$$\hat{x}(y) = \arg\max_x \; f_{X|Y}(x|y) \;=\; \arg\max_x \; \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$$
$$= \arg\max_x \; f_{Y|X}(y|x)\, f_X(x)$$

If the prior $f_X(x)$ is non-informative (e.g., uniform over the whole region of interest), we obtain

$$\hat{x}(y) = \arg\max_x f_{Y|X}(y|x)$$

which is known as the **maximum likelihood** estimate.

- $f_{Y|X}(y|x)$ regarded as a function of x is called a *likelihood function*
- This estimator does not depend on the prior
- ML and MAP are identical if the prior is constant (= uniform).

Problem 12.3.4

Flip a coin n times. For each flip, the probability of heads is $Q = q$, independent of all other flips. Q is a Uniform(0,1) random variable. K is the number of heads in n flips.

- What is the ML estimator of Q given K?

Problem 12.3.4

Flip a coin n times. For each flip, the probability of heads is $Q = q$, independent of all other flips. Q is a Uniform(0,1) random variable. K is the number of heads in n flips.

- What is the ML estimator of Q given K?

Given $Q = q$, the conditional PMF of K is (binomial)

$$P_{K|Q}(k|q) = \begin{cases} \binom{n}{k} q^k (1-q)^{n-k} & k = 0, 1, \cdots, n, \\ 0 & \text{otherwise} \end{cases}$$

The ML estimator of Q given $K = k$ is

$$\hat{q}_{\mathrm{ML}}(k) = \arg\max_{0 \le q \le 1} \; P_{K|Q}(k|q)$$

Problem 12.3.4 (cont'd)

Differentiating $P_{K|Q}(k|q)$ with respect to q and setting equal to zero yields

$$\frac{dP_{K|Q}(k|q)}{dq} = \binom{n}{k}\left(kq^{k-1}(1-q)^{n-k} - (n-k)q^k(1-q)^{n-k-1}\right) = 0$$

$$\Rightarrow \quad k(1-q) = (n-k)q$$

The maximizing value is $q = k/n$ so that

$$\hat{Q}_{\mathrm{ML}}(K) = \frac{K}{n}$$

- This is intuitive: to estimate Q, simply count the relative frequency of Heads.
- Finding the MMSE is much harder!

Linear MMSE estimation of a RV

Bayesian estimators:

- MMSE: $\hat{x}(y) = \mathrm{E}[X|Y = y]$
- MAP: $\hat{x}(y) = \arg\max_x \; f_{X|Y}(x|y)$

These estimators

- are generally non-linear functions of the observations;
- involve the posterior density $f_{X|Y}(x|y)$.

The non-linearity makes it sometimes difficult to derive and/or implement these estimators.

Moreover, what if the density $f_{X|Y}(x|y)$ is unknown and cannot be estimated from the data?

Linear MMSE estimation of a RV

The Linear MMSE estimator constrains the estimator $\hat{X}_{\text{lin}}$ to have a linear relationship with the observable RV:

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

- To find the constants a and b, we again minimize the MSE:

$$\arg\min_{a,b} \mathrm{E}[(X - \hat{X}_{\text{lin}}(Y))^2]$$

$$e = \mathrm{E}[(X - \hat{X}_{\text{lin}}(Y))^2] = \mathrm{E}[(X - aY - b)^2]$$

Expanding gives:

$$e = \mathrm{E}[X^2] - 2a\,\mathrm{E}[XY] - 2b\,\mathrm{E}[X] + a^2\mathrm{E}[Y^2] + 2ab\,\mathrm{E}[Y] + b^2$$

Linear MMSE estimation of a RV (cont'd)

- The optimal parameters are found by setting partial derivatives to zero:

$$\begin{aligned}\frac{\partial e}{\partial a} &= -2\mathrm{E}[XY] + 2a\mathrm{E}[Y^2] + 2b\mathrm{E}[Y] = 0 \\ \frac{\partial e}{\partial b} &= -2\mathrm{E}[X] + 2a\mathrm{E}[Y] + 2b = 0\end{aligned}$$

- Solving for a and b then leads to

$$a^* = \frac{\mathrm{cov}[X, Y]}{\mathrm{var}[Y]} = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y} \quad \text{and} \quad b^* = \mathrm{E}[X] - a^*\mathrm{E}[Y]$$

Linear MMSE

- The Linear MMSE estimator is

$$\hat{X}_{\text{lin}}(Y) = a^* Y + b^* \; = \; \frac{\text{cov}[X,Y]}{\text{var}[Y]} Y + \text{E}[X] - \frac{\text{cov}[X,Y]}{\text{var}[Y]} \text{E}[Y]$$

$$= \; \frac{\text{cov}[X,Y]}{\text{var}[Y]} (Y - \text{E}[Y]) + \text{E}[X]$$

  We have seen a similar result for bivariate Gaussian variables.

- This can be generalized to the case where we estimate a RV X from a vector random process $\boldsymbol{Y}$:

$$\hat{X}_{\text{lin}}(\boldsymbol{Y}) = \boldsymbol{C}_{X\boldsymbol{Y}} \boldsymbol{C}_{\boldsymbol{Y}}^{-1} (\boldsymbol{Y} - \text{E}[\boldsymbol{Y}]) + \text{E}[X]$$

  and to the case where we estimate a vector random process $\boldsymbol{X}$ from another vector random process $\boldsymbol{Y}$:

$$\hat{\boldsymbol{X}}_{\text{lin}}(\boldsymbol{Y}) = \boldsymbol{C}_{\boldsymbol{X}\boldsymbol{Y}} \boldsymbol{C}_{\boldsymbol{Y}}^{-1} (\boldsymbol{Y} - \text{E}[\boldsymbol{Y}]) + \text{E}[\boldsymbol{X}]$$

Problem 12.4.2

$\boldsymbol{X}$ is a three-dimensional random vector with $\mathrm{E}[\boldsymbol{X}] = 0$ and autocorrelation matrix $\boldsymbol{R_X}$ with elements $r_{ij} = (-0.80)^{|i-j|}$.

Use X_1 and X_2 to form a linear estimate of X_3: $\hat{X}_3 = a_1 X_1 + a_2 X_2$.

Problem 12.4.2

$\boldsymbol{X}$ is a three-dimensional random vector with $E[\boldsymbol{X}] = 0$ and autocorrelation matrix $\boldsymbol{R_X}$ with elements $r_{ij} = (-0.80)^{|i-j|}$.

Use X_1 and X_2 to form a linear estimate of X_3: $\hat{X}_3 = a_1 X_1 + a_2 X_2$.

In this problem, we view $\boldsymbol{Y} = [X_1 \;\; X_2]^T$ as the observation and $X = X_3$ as the variable we wish to estimate. Then $\hat{X}_{\text{lin}}(\boldsymbol{Y}) = \boldsymbol{R}_{X\boldsymbol{Y}} \boldsymbol{R}_{\boldsymbol{Y}}^{-1} \boldsymbol{Y}$ where

$$\boldsymbol{R_Y} = E[\boldsymbol{Y}\boldsymbol{Y}^T] = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] \\ E[X_2 X_1] & E[X_2^2] \end{bmatrix} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

$$\boldsymbol{R}_{X\boldsymbol{Y}} = E[X\boldsymbol{Y}^T] = E[X_3[X_1 \;\; X_2]] = [E[X_1 X_3],\; E[X_2 X_3]] = [0.64,\; -0.8]$$

$$\boldsymbol{R}_{X\boldsymbol{Y}} \boldsymbol{R}_{\boldsymbol{Y}}^{-1} = [0.64,\; -0.8] \begin{bmatrix} 25/9 & 20/9 \\ 20/9 & 25/9 \end{bmatrix} = [0,\; -0.8]$$

The optimum linear estimator of X_3 given X_1 and X_2 is

$$\hat{X}_3 = -0.8\, X_2$$

Estimation of a RV: summary

- **Minimum mean-squared error (MMSE)** estimation:
  Minimize $\mathrm{E}[(X - \hat{X})^2]$
  - **Blind estimation** (no observation): we can only use $f_X(x)$
    $\hat{x} = \mathrm{E}[X]$, i.e., take the mean
  - Estimation of X given a related variable Y:

$$\hat{X}(Y) = \arg\min_{\hat{X}(Y)} \mathrm{E}[(X - \hat{X})^2 \mid Y] = \mathrm{E}[X|Y]$$

    Uses $f_{X|Y}(x|y)$ (inverse model)

Estimation of a RV: summary

- Uniform cost function:

  Minimize $C(X,\hat{X}) = \begin{cases} 0, & |X-\hat{X}| < \epsilon \\ 1, & \text{otherwise} \end{cases}$

  - **Blind**: $\hat{x} = \arg\max_x f_X(x)$, i.e., take the *mode*
  - **Maximum A Posteriori (MAP)**:

$$\hat{x}(y) = \arg\max_x f_{X|Y}(x|y) = \arg\max_x f_{Y|X}(y|x) \cdot f_X(x)$$

  - **Maximum Likelihood (ML)**: $\hat{x}(y) = \arg\max_x f_{Y|X}(y|x)$ ignores $f_X(x)$

In most estimation problems, ML is the standard choice.

Estimation of a RV: summary

Special case:

- **Linear MMSE estimation**: $\hat{x}(y) = ay + b$; uses only second-order statistics:

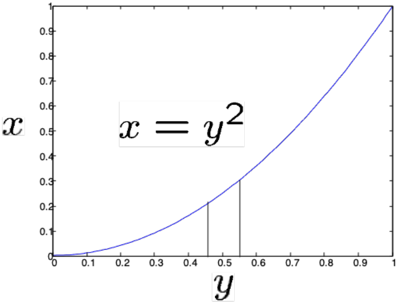
$$\hat{x}(y) = a\,y + b = \frac{\text{cov}[X, Y]}{\text{var}[Y]}\,(y - \text{E}[Y]) + \text{E}[X]$$

Example

Let the joint density of X and Y be given by

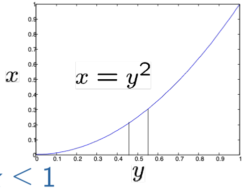
$$f_{X,Y}(x,y) = \begin{cases} 10x & 0 \leq x \leq y^2, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We observe realizations of Y and want to estimate X.



Example

- First compute the marginals:



$$
\begin{aligned}
f_Y(y) &= \int_0^{y^2} 10x\,\mathrm{d}x = 5y^4\,, \quad 0 \le y \le 1 \\
f_X(x) &= \int_{\sqrt{x}}^{1} 10x\,\mathrm{d}y = 10(x - x^{3/2})\,, \quad 0 \le x \le 1
\end{aligned}
$$

- Then the conditional densities are

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{10x}{5y^4} = \frac{2x}{y^4}\,, \quad 0 \le x \le y^2 \\
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{1-\sqrt{x}}\,, \quad \sqrt{x} \le y \le 1
\end{aligned}
$$

Example

- **MMSE**: compute the conditional expectation:

$$\hat{x}_{\text{MMSE}}(y) = E[X|Y=y] = \int x\, f_{X|Y}(x|y)\,\mathrm{d}x = \int_0^{y^2} \frac{2x^2}{y^4}\,\mathrm{d}x = \frac{2}{3}y^2$$

- **MAP**: maximize the a posteriori density $f_{X|Y}(x|y)$:

$$\hat{x}_{\text{MAP}}(y) = \arg\max_x \; f_{X|Y}(x|y) = \arg\max_x \; \frac{2x}{y^4}\,, \quad 0 \le x \le y^2$$

  The maximum over x is achieved for $x = y^2$, so $\hat{x}_{\text{MAP}}(y) = y^2$.

- **ML**: maximize the likelihood function $f_{Y|X}(y|x)$.
  As function of x, it is monotonically increasing, the maximum over x is achieved for $\sqrt{x} = y$, so

$$\hat{x}_{\text{ML}}(y) = \arg\max_x f_{Y|X}(y|x) = y^2$$

Example

- **Linear MMSE**:

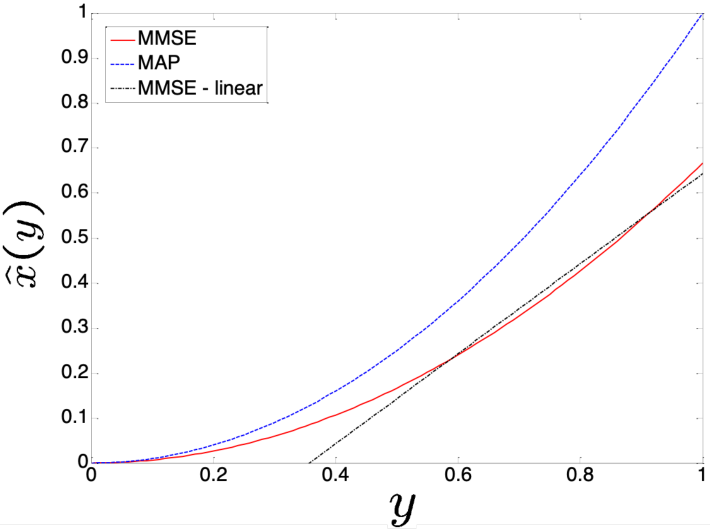
  Compute the moments:

$$
\begin{aligned}
\mathrm{E}[X] &= \int_0^1 x\, f_X(x)\, \mathrm{d}x = 10/21 \\
\mathrm{E}[Y] &= \int_0^1 y\, f_Y(y)\, \mathrm{d}y = 5/6 \\
\mathrm{E}[Y^2] &= \int_0^1 y^2\, f_Y(y)\, \mathrm{d}y = 5/7 \\
\mathrm{E}[XY] &= \int_0^1 \int_0^{y^2} xy\, f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y = 10/24
\end{aligned}
$$

  Then

$$
\begin{aligned}
\hat{X}_{\mathrm{lin}}(Y) &= \frac{\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]}{\mathrm{E}[Y^2] - \mathrm{E}[Y]^2}(Y - \mathrm{E}[Y]) + \mathrm{E}[X] \\
&= Y - 5/14
\end{aligned}
$$

Example



MMSE
MAP
MMSE - linear
$\hat{x}(y)$
y
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Example

Estimation errors (MSE):

- **MMSE**:

$$E[(X - \hat{X}(Y))^2] = \int_0^1 \int_0^{y^2} \left(x - \frac{2}{3}y^2\right)^2 10x \, dx dy = 0.0309$$

- **MAP, ML**:

$$E[(X - \hat{X}(Y))^2] = \int_0^1 \int_0^{y^2} (x - y^2)^2 \, 10x \, dx dy = 0.0926$$

- **LMMSE**:

$$E[(X - \hat{X}(Y))^2] = \int_0^1 \int_0^{y^2} \left(x - y + \frac{5}{14}\right)^2 10x \, dx dy = 0.0312$$

Some suggested exercises

Ch. 12: 12.1.3, 12.1.5, 12.2.1, 12.2.3, 12.2.5, 12.3.3, 12.4.3

Errata

- Eqn (12.8): x is missing in the integral; $\int_0^r x \frac{1}{r} \mathrm{d}x$
- Theorem 12.5: "Discrete" repeated the definition. The new result is:

$$\hat{x}_{\mathrm{MAP}}(y_j) = \arg\max_x P_{Y|X}(y_j|x) P_X(x)$$

  (Some typos also two lines above Theorem 12.5)
- Definition 12.2: "MAP" should be "ML"
- Solution of Problem 12.1.5: above and below eqn (1), $0 \leq y \leq 1$ (not "2"); eqn (4) gives the total MSE (averaging over Y), but it was asked to give the MSE for $Y = 0.5$.
- Problem 12.1.7, above eqn (3): $Z \geq x - y$ (not "$\leq$")
- Problem 12.2.7: above (6): $n = 2$, not 1. In (6) and (7) also replace 1 by 2.