EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

BOSTON UNIVERSITY · CISE

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 17 - Outline

1. **Estimation theory**
   a. **ML and MAP estimation**
   b. **MMSE estimation**
   c. **LLSE estimation**
   d. **Extensions to random vectors**

Chapter 9 **Estimation Theory**

An estimation problem is an *N*-ary test of hypotheses problem (*N* may be ∞ or the hypothesis may be a continuous variable).

Review of the *N*-ary test of hypothesis problem:

Given observation *X*, which of hypotheses $H_1, H_2, .. H_N$ is true? The decision $U \in \{U_1, U_2, .. \text{ or } U_N\}$ depends on *X*.

ML: Select hypothesis for which $P[X|H_j]$ is maximum

MAP: Select hypothesis for which $P[H_j|X]$ is maximum (or $P[X|H_j]\, P[H_j]$ is maximum)

Bays Risk: Select hypothesis for which expected cost $E[C(U(X),H)]$ is minimum

In estimation theory we regard the set of hypotheses as a random variable *H* (discrete or continuous)

The estimation problem:

Given observation *X*, estimate the RV *H*. The estimator *U* is a random variable that depends on *X*

ML: Select H for which $P[X|H]$ is maximum

MAP: Select H for which $P[H|X]$ is maximum (or $P[X|H]\, P[H]$ is maximum)

Bays Risk: Select *H* for which $E[C(U(X),H)]$ is minimum

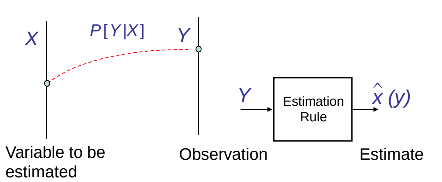
Change of Notations:

Detection	Estimation
Observation *X*	Observation *Y*
Hypothesis *H*	Variable to be estimated *X*
Decision *U*	Estimator $\hat{x}(y)$

Estimation problem:

Given observation *Y*, find an estimator $\hat{x}(y)$ for the value of RV *X*

(Think of *X* and *Y*, and also $\hat{x}(y)$ as RVs)



Properties

1) Mean error (bias)

$$B = E[\hat{x}(y) - X] = E[\hat{x}(y)] - E[X]$$

If *B* = 0, i.e.,

$$E[\hat{x}(y)] = E[X]$$

the estimator is said to be unbiased

2) Mean square error (MSE)

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

9.1 ML and MAP Estimation

Maximum Likelihood (ML) Estimate

ML estimate of X given Y is the value of X that best explains Y, i.e., for which the likelihood function $p_{Y|X}(y|x)$ is maximum.

$$\hat{x}_{ML}(y) = \arg\max_{x\in S_x} p_{Y|X}(y|x) \quad (Y \text{ discrete})$$
$$= \arg\max_{x\in S_x} f_{Y|X}(y|x) \quad (Y \text{ continuous})$$

Maximum A Posteriori (MAP) Estimate

MAP estimate of X given Y = value of X for which $P_{X|Y}(x|y)$, or $P_{Y|X}(y|x)\, P_X(x)$, is maximum.

$$\hat{x}_{MAP}(y)$$
$$= \arg\max_{x\in S_x} p_{X|Y}(x|y) = \arg\max_{x\in S_x} p_{X,Y}(x,y) \quad (Y \text{ discrete})$$
$$= \arg\max_{x\in S_x} f_{X|Y}(x|y) = \arg\max_{x\in S_x} f_{X,Y}(x,y) \quad (Y \text{ continuous})$$

MAP vs ML

As in detection, ML estimation is same as MAP estimation when prior information on X is uniform

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x) = Cp_{Y|X}(y|x)$$

Example

- Number of customers arriving at a service station X is a binomial RV with parameters (4, 0.5)
- Service time Y is exponentially distributed with parameter $\lambda(X) = 1/(1+X)$
- Service time Y was observed and found to be $y = 2$
- Find an ML and a MAP estimate of X

$$p_X(x) = \binom{4}{x} 0.5^x (1-0.5)^{4-x} = \binom{4}{x} 0.5^4$$
$$f_{Y|X}(y|x) = \lambda(x)e^{-\lambda(x)y} = \frac{1}{1+x}e^{-\frac{y}{1+x}}, \quad y \ge 0$$

Likelihood function $f_{Y|X}(2|x) = \frac{1}{1+x}e^{-\frac{2}{1+x}}$

ML Estimate: $\hat{x}_{ML}(2) = \arg\max_x \frac{1}{1+x}e^{-\frac{2}{1+x}}$

0	1	2	3	4
0.14	0.18	0.17	0.15	0.13

$\hat{x}_{ML}(2) = 1$

MAP Estimate: $\hat{x}_{MAP}(2) = \arg\max_x f_{Y|X}(2|x)p_X(x)$
$$= \arg\max_x \frac{1}{1+x}e^{-\frac{2}{1+x}}\binom{4}{x}0.5^4$$

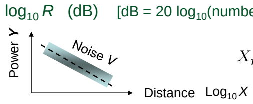
0	1	2	3	4
0.01	0.05	0.06	0.04	0.01

$$\hat{x}_{MAP}(2) = 2$$

Quiz 9.3

Receiver at distance X from a radio beacon measures the beacon power to be

$X = Y - 40 - 40 \log_{10} R$ (dB) [dB = 20 $\log_{10}$(number)]



$$X_{number} = \frac{10^{\frac{Y}{20}-2}}{R^2}$$

Y, called the shadow fading factor, is noise, modeled as N(0,8) independent of X

Receiver location is unknown, with PDF

$$f_R(r) = \begin{cases} \frac{2r}{10^6} & 0 \le r \le 10^3 \\ 0 & else \end{cases}$$

–Find ML and MAP estimates of R given $X = x$

- ML $\quad f_{X|R}(x|r) = N(-40 - 40\log_{10}(r), 8)$

$$\hat{r}_{ML}(x) = \arg\max_r f_{X|R}(x|r)$$
$$= \arg\min_r (x + 40 + 40\log_{10}(r))^2$$
$$\Rightarrow \hat{r}_{ML}(x) = 10^{-\frac{x}{40}-1}$$

- MAP

$$f_{X|R}(x|r)f_R(r) \sim r\exp\left\{-\frac{(x+40+40\log_{10} r)^2}{2\sigma^2}\right\} \quad 0 \le r \le 10^3$$
$$\hat{r}_{MAP}(x) = \arg\max_r f_{X|R}(x|r)f_R(r)$$
$$\Rightarrow \hat{r}_{MAP}(x) = \begin{cases} 10^{\frac{\ln 10}{200}-\frac{x}{40}-1} & \text{if } 10^{\frac{\ln 10}{200}-\frac{x}{40}-1} \le 10^3 \\ 10^3 & \text{if } 10^{\frac{\ln 10}{200}-\frac{x}{40}-1} > 10^3 \end{cases}$$

9.2 Minimum Mean Square Error (MMSE) Estimation

The MMSE Estimator of X given the observation Y is the estimator $\hat{x}(y)$ with the smallest mean square error (MSE):

$$e = E[(X - \hat{x}(y))^2]$$ (Joint expectation over X,Y)

The estimator is denoted $\hat{x}_{MMSE}$ or simply $\hat{x}_M$

1) Simple Case: Blind Estimation (No Observation)

In this case, the estimator is a constant (does not depend on the random observation y) $\hat{x}(y) = C$

$$\begin{aligned} e &= E[(X - \hat{x}(y))^2] \\ &= E[(X - C)^2] = E[X^2 - 2XC + C^2] \\ &= E[X^2] - 2CE[X] + C^2 \\ &= E[X^2] - E[X]^2 + (C - E[X])^2 \end{aligned}$$

The value of C for which e minimum is: $C = E[X]$

Therefore, the blind MMSE estimator is $\boxed{\hat{x}_M(y) = E[X]}$

Given no measurement, the MMSE estimator is the prior mean of X

2) More Complex Case: a property of X is observed

- Observe $X \in A \subset S_X$
  - Learn about X
  - Conditional PDF $f_{X|A}(x)$ or PMF $p_{X|A}(x)$
- MSE: Iterated expectation

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

  - Minimize the inner expectation for which A *is given*
  - Can be done because A is observed, so estimate is a constant given knowledge of A!!!
    - Use calculus to minimize…

- Consider any estimator that is a function of observed set A
  - Then,

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

$$\begin{aligned} E[(X - \hat{x})^2|A] &= E[X^2 - 2\hat{x}X + \hat{x}^2|A] \\ &= E[X^2|A] - 2\hat{x}E[X|A] + \hat{x}^2 \\ &\quad \text{because } \hat{x} \text{ function of } A \\ &= E[X^2|A] - E[X|A]^2 + (\hat{x} - E[X|A])^2 \end{aligned}$$

- MMSE Estimator: $\boxed{\hat{x}_M = E[X|A]}$

3) Harder Case: Another RV Y is observed

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

$$\begin{aligned} E[(X - \hat{x})^2|Y = y] &= E[X^2 - 2\hat{x}X + \hat{x}^2|Y = y] \\ &= E[X^2|Y = y] - 2\hat{x}E[X|Y = y] + \hat{x}^2 \\ &\quad \text{because } \hat{x} \text{ function of } observed\ y \\ &= E[X^2|Y = y] - E[X|Y = y]^2 + (\hat{x} - E[X|Y = y])^2 \end{aligned}$$

The error is minimum when $\hat{x}_M(y) = E[X|Y = y]$

So that $\boxed{\hat{x}_M(y) = E[X|Y = y]}$

This estimator may be computed from the conditional PDF $f_{X|Y}(x|y)$ or PMF $p_{X|Y}(x|y)$

MMSE estimator is *unbiased*

$$\begin{aligned} E_Y[\hat{x}_M(y)] &= E_Y[E[X|Y = y]] \\ &= E[X] \text{ (iterated expectation)} \end{aligned}$$

expectation with respect to y

unconditional expectation of X

MMSE Orthogonality Property

MMSE estimator $E[X|Y=y]$ is orthogonal (and uncorrelated) with error

$Z = X – E[X|Y=y]$

$$E[Z] = E[X] - E\{E[X|Y = y]\} = E[X] - E[X] = 0$$

$$E\big[ZE[X|Y = y]\big] = E\Big\{E\big[(X - E[X|Y = y])E[X|Y = y]\big|Y = y\big]\Big\}$$

$$E\big[(X - E[X|Y = y])E[X|Y = y]\big|Y = y\big] = E[X|Y = y]^2 - E[X|Y = y]^2 = 0$$

Example: Quiz 9.1

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^y 2(x+y)dx = 3y^2,\ 0 \le y \le 1$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{2(x+y)}{3y^2} & 0 \le x \le y \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{x}_M(y) = E[X|Y=y] = \int_0^y x\frac{2x+2y}{3y^2}dx = \int_0^y \frac{2x^2+2xy}{3y^2}dx = \frac{2y}{9} + \frac{y}{3} = \frac{5}{9}y$$

9.3 Linear Least Squares Estimator (LLSE)

Sometimes, expectations are hard to compute. Idea: Restrict estimator to be a linear function of the observation y.

$$\hat{x}_{LLSE}(y) \equiv \hat{x}_L(y) = ay + b \text{ for some } a, b$$

To determine *a, b* substitute form into expected mean square error, and choose *a, b* to minimize!

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

$$= \frac{Cov(X,Y)}{\sigma_Y^2}y + E[X] - \frac{Cov(X,Y)}{\sigma_Y^2}E[Y]$$

$$= E[X] + \frac{Cov(X,Y)}{\sigma_Y^2}(y - E[Y])$$

$$= E[X] + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y - E[Y])$$

LLSE estimator is easy to compute. It depends only on means, variances, covariance. Not necessary to know conditional PDFs or PMFs

Proof

$$e = E[(X - \hat{x}_L)^2]$$
$$= E[(X - aY - b)^2]$$
$$= E[X^2 - 2aXY + a^2Y^2 - 2Xb + 2abY + b^2]$$
$$= E[X^2] - 2aE[XY] + a^2E[Y^2] - 2bE[X] + 2abE[Y] + b^2$$

Differentiate with respect to *a, b* and set to 0

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

Solve simultaneous equations:

$$\frac{\partial}{\partial a}e - E[Y]\frac{\partial}{\partial b}e = -2E[XY] + 2aE[Y^2] + 2bE[Y] + 2E[X]E[Y] - 2aE[Y]^2 - 2bE[Y] = 0$$
$$\Rightarrow a(E[Y^2] - E[Y]^2) = (E[XY] - E[X]E[Y])$$
$$\Rightarrow a\sigma_Y^2 = Cov(X,Y) \Rightarrow a = \frac{Cov(X,Y)}{\sigma_Y^2}$$
$$b = E[X] - aE[Y] = E[X] - \frac{Cov(X,Y)}{\sigma_Y^2}E[Y]$$

Properties of the LLSE Estimator

- LLSE estimate is unbiased: $b = E[X] - aE[Y]$

$$E[aY + b] = E[E[X] + a(Y - E[Y])] = E[X] + aE[Y - E[Y]] = E[X]$$

- Mean Square Estimation Error $Z = X - aY - b$ is uncorrelated (orthogonal) with LLSE estimate!

$$E[(X - aY - b)(aY + b)]$$
$$= E\left\{\left[(X - E[X]) - a(Y - E[Y])\right]\left[a(Y - E[Y]) + E[X]\right]\right\}$$
$$= E\left\{(X - E[X])E[X]\right\} + aE\left\{(X - E[X])(Y - E[Y])\right\} - aE\left\{(Y - E[Y])E[X]\right\} - a^2E\left\{(Y - E[Y])^2\right\}$$
$$= aCov(X,Y) - a^2\sigma_Y^2 = a\left[Cov(X,Y) - \frac{Cov(X,Y)}{\sigma_Y^2}\sigma_Y^2\right] = 0$$

- Mean Square Error of LLSE:

$$E[(X - aY - b)^2] = E[ZX] - E[Z(aY + b)]$$
$$= E[ZX] \text{ uncorrelated LLSE, } Z$$
$$\text{(plug in } b\text{)}\ = E[X(X - E[X]) - aX(Y - E[Y])]$$
$$= E[(X - E[X])^2 - a(X - E[X])(Y - E[Y])]$$
$$\text{(plug in } a\text{)}\ e_L^* = \sigma_X^2 - \frac{Cov(X,Y)^2}{\sigma_Y^2}$$
$$= \sigma_X^2 - \rho_{X,Y}^2\sigma_X^2$$

- Estimation Error Z is zero-mean, orthogonal to Y
  - Geometry of best approximation

$$E[Z] = E[(X - E[X]) - a(Y - E[Y])] = 0$$
$$E[ZY] = E[(X - E[X])Y - a(Y - E[Y])Y]$$
$$= E[(X - E[X])(Y - E[Y]) - a(Y - E[Y])^2]$$
$$= Cov(X,Y) - a\sigma_Y^2 = 0$$

- e_{MMSE} is no larger than e_{LLSE}
  - MMSE is best estimator, linear or nonlinear
  - LLSE is best linear estimator
  - Best of larger class is at least as good as best in subclass
- e_{MMSE} is average of conditional covariance over possible values of Y
  - Hard to compute, depends on form of conditional density or PMF
- e_{LLSE} is easy to compute

9.4 MMSE and LLSE estimators for Gaussians

For the special case of *X,Y* jointly Gaussian, the conditional density $f_{X|Y}(x|y)$ is also Gaussian with mean and variance

$$E[X|Y] = E[X] + \frac{Cov(X,Y)}{\sigma_Y^2}(Y - E[Y])$$
$$Var[X|Y] = \sigma_X^2 - \frac{Cov(X,Y)^2}{\sigma_Y^2}$$

Therefore, MMSE is ***Linear***! and LLSE and MMSE estimators are the same

–This is rare! (but Gaussians are special…)

LLSE estimator for general variables can be interpreted as using Gaussian approximation

Example: Quiz 9.2

Temperature sensor measures $T \sim N(0,9)$ and transmits.
Receiver measures voltage $R = T+V$.
Quantization noise V is uniform in [-3,3] and independent of T.
Determine LLSE estimate of T

$E[R] = E[T] + E[V] = 0$
$Var[R] = Var[T] + Var[V]$ (independent) = 9 + 3 = 12
$Cov[T,R] = Var[T] + Cov[T,V] = 9 + 0 = 9$

$\rho_{T,R} = 9/3\sqrt{12}$

LLSE estimate:
$$\hat{T}_L(r) = E[T] + \rho_{T,R}\frac{\sigma_T}{\sigma_R}(r - E[R]) = \hat{T}_L(r) = 0 + \frac{9}{12}(r-0) = \frac{3}{4}r$$

Optimal error :
$$e_L^* = \sigma_T^2 - \frac{Cov(T,R)^2}{\sigma_R^2} = 9 - \frac{81}{12} = \frac{27}{12} = 2.25$$

9.4 LLSE for Random Vectors

- Random vectors **X**, **Y**
  - Means μ_X, μ_Y
  - Covariance and cross covariance Σ_X, Σ_Y, Σ_{XY}
- MMSE estimate of **X** given observed **Y** = y:
$$\hat{\underline{x}}_M(\underline{y}) = E[\underline{X}|\underline{Y} = \underline{y}]$$
- LLSE estimate:
$$\hat{\underline{x}}_L(\underline{y}) = \underline{\mu}_X + \Sigma_{\underline{XY}}\Sigma_{\underline{Y}}^{-1}(\underline{y} - \underline{\mu}_Y)$$
- For joint Gaussians, MMSE and LLSE are same

- MMSE Error $\underline{Z}_M = \underline{X} - \hat{\underline{x}}_M(\underline{y})$
  - Uncorrelated (orthogonal) to $\underline{Y}$
- LLSE Error $\underline{Z}_L = \underline{X} - \hat{\underline{x}}_L(\underline{y})$
  - Uncorrelated (orthogonal) to **Y**
  - Simple expression for covariance of $\mathbf{Z}_L$

$$\Sigma_{\underline{Z}_L} = \Sigma_{\underline{X}} - \Sigma_{\underline{XY}}\Sigma_{\underline{Y}}^{-1}\Sigma_{\underline{YX}}$$

Example: Quiz 9.4

X random vector, **W** random vector independent of **X**

– Observe **Y** = **X** + **W**

$$E_{\underline{W}} = 0 \quad \Sigma_{\underline{W}} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$$
$$E_{\underline{X}} = 0 \quad \Sigma_{\underline{X}} = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$

– Find the LLSE of X_2 given Y_2 and the mean square error of the estimate

– Find the LLSE of X_2 given observations of both Y_1 and Y_2 and the mean square error

- LLSE of X_2 given Y_2

$$Cov(X_2, Y_2) = Var(X_2) = 1 \quad Var(Y_2) = Var(X_2) + Var(W_2) = 1.1$$
$$\hat{x}_2(y_2) = 0 + \frac{1}{1.1}(y_2 - 0) = \frac{10}{11}y_2$$
$$e_L^* = 1 - \frac{1}{1.1} = \frac{1}{11}$$

- LLSE of X_2 given Y_1, Y_2:

$$\Sigma_{X_2\underline{Y}} = \begin{pmatrix} E[X_2Y_1] & E[X_2Y_2] \end{pmatrix} = \begin{pmatrix} E[X_2X_1] & E[X_2^2] \end{pmatrix} = \begin{pmatrix} -0.9 & 1 \end{pmatrix}$$
$$\Sigma_{\underline{Y}} = \Sigma_{\underline{X}} + \Sigma_{\underline{W}} = \begin{pmatrix} 1.1 & -0.9 \\ -0.9 & 1.1 \end{pmatrix}$$
$$\hat{x}_2(\underline{y}) = 0 + \Sigma_{X_2\underline{Y}}\Sigma_{\underline{Y}}^{-1}(\underline{y} - 0)$$

- Trivial to do in MATLAB…

$$\Sigma_{\underline{Y}}^{-1} = \begin{pmatrix} \frac{11}{4} & \frac{9}{4} \\ \frac{9}{4} & \frac{11}{4} \end{pmatrix}$$

$$\hat{x_2}(\underline{y}) = \begin{pmatrix} -0.9 & 1 \end{pmatrix} \begin{pmatrix} \frac{11}{4} & \frac{9}{4} \\ \frac{9}{4} & \frac{11}{4} \end{pmatrix} \underline{y}$$

$$= \begin{pmatrix} -0.225 & 0.725 \end{pmatrix} \underline{y}$$

$$e^* = 1 - \begin{pmatrix} -0.9 & 1 \end{pmatrix} \begin{pmatrix} \frac{11}{4} & \frac{9}{4} \\ \frac{9}{4} & \frac{11}{4} \end{pmatrix} \begin{pmatrix} -0.9 \\ 1 \end{pmatrix}$$

$$= 0.0725$$

More information ➔ Less error uncertainty

Extensions to Random Vectors

Extensions are straightforward:

- LLSE computed in same manner as Gaussians were before
- MMSE is again conditional expectation
- Have nice orthogonality properties as before
- MAP and ML estimators defined as optimizations over vectors, not scalars
- Joint Gaussian vectors