Lecture 5

Optimal Estimation Theory and Its Applications

Prof. Yan LIANG

liangyan@nwpu.edu.cn

School of Automation,

Northwestern Polytechnical University

Autumn,2019

Outline

- A Posteriori estimation
- Maximum A Posteriori estimation(MAP)
- Comparison MAP with MLE
- Homework
- Appendix

A Posteriori estimation

Given the matter of "Z=z" and the conditional PDF $p(x|z)$,determine the best estimate.

—A Posteriori estimation (APE)

There are multiple performance indexes for designing APE:

1. Maximum A Posteriori estimation
2. Minimum Variance estimation
3. Minimum Error estimation

A Posteriori estimation

1. Maximum A Posteriori estimation

$$\hat{X}_{MAP} = \arg\max_{X} p\{X|Z\}$$

2. Minimum Variance estimation

$$\hat{X}_{MV} = E\left\{X|Z\right\} = \int xp(x|z)dx$$

3. Minimum Error estimation

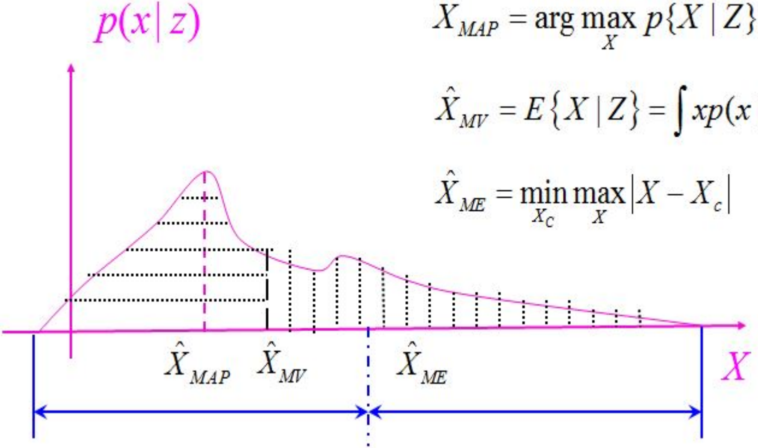
$$\hat{X}_{ME} = \min_{X_C} \max_{X} |X - X_c|$$

A Posteriori estimation

$$\hat{X}_{MAP} = \arg\max_{X} p\{X \mid Z\}$$

$$\hat{X}_{MV} = E\{X \mid Z\} = \int x p(x \mid z) dx$$

$$\hat{X}_{ME} = \min_{X_C} \max_{X} |X - X_c|$$



Maximum A Posteriori estimation

Maximum A Posteriori estimation (MAP)

$$\hat{X}_{MAP} = \arg\max_X p\{X|Z\}$$

The probability that a random variable X appears within the small neighborhood of $\hat{X}_{MAP}$ is largest.

Idea: Given Z=z, $X = \hat{X}_{MAP}$ is the most likely to occur.

In statistics,$\hat{X}_{MAP}$ is called as mode

Maximum A Posteriori estimation

Maximum A Posteriori estimation

$$\frac{\partial p(X|Z)}{\partial X}\bigg|_{X=\hat{X}_{MAP}} = 0$$ (if it is derivable)

$$\frac{\partial \ln p(X|Z)}{\partial X}\bigg|_{X=\hat{X}_{MAP}} = 0$$ A Posteriori Equation

Maximum A Posteriori estimation

Example 3.1: Consider a n-dimensional vector X,z_i is the m-dimensional vector of the i-th measurement, measurement error is v_i.Through k independent observations,we have

$$z_i = h_i X + v_i \qquad i = 1, 2, \cdots, k$$

$$Z = HX + V \qquad Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}_{km} \qquad H = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_k \end{bmatrix}_{km \times n} \qquad V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}_{km}$$

Maximum A Posteriori estimation

$\mathrm{X} \sim \mathrm{N}\{\mathrm{M}, \mathrm{P}\} \mathrm{and} \mathrm{V} \sim \mathrm{N}\{0, \mathrm{R}\}$ are independent each other. Please determine the MAP estimate of X.

Solution

(1) determine p(x)

$$p\{X = x\} = \frac{1}{(2\pi)^{\frac{n}{2}}|P|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x - M)^T P^{-1}(x - M)\}$$

Maximum A Posteriori estimation

(2) determine p(Z)

$$E(Z) = HE(X) + E(V) = HM$$
$$\mathrm{Var(Z)} = HPH^{\mathrm{T}} + \mathrm{R}$$

By the fact that Z=HX+V, and both X and V are Gaussian,Z is also Gaussian, i.e.,

$$p\{Z = z\} = \frac{\exp\{-\frac{1}{2}(z - HM)^T (HPH^T + R)^{-1}(z - HM)\}}{\left|2\pi\left(HPH^T + R\right)\right|^{\frac{1}{2}}}$$

Maximum A Posteriori estimation

(3) determine p(V)

By the fact that V is Gaussian, we have

$$p\{V = v\} = \frac{\exp\{-\frac{1}{2}v^T R^{-1} v\}}{|2\pi R|^{\frac{1}{2}}}$$

Maximum A Posteriori estimation

(4) determine p(X,Z)

$$P\{X,V\}=P(X)P(V)=\boxed{P(X)P(Z\text{-}HX)}$$

where

Joint PDF of X and Z

$$p\{V = Z - HX = z - Hx\} = \frac{\exp\left\{-\frac{1}{2}(z-Hx)^T R^{-1}(z-Hx)\right\}}{|2\pi R|^{\frac{1}{2}}}$$

Maximum A Posteriori estimation

(4) determine p(X|Z)

According to the Bayesian formula, we have

$$p\{X=x|Z=z\} = \frac{p\{X=x,Z=z\}}{p\{Z=z\}}$$
$$= \frac{\left|2\pi\left(HPH^T+R\right)\right|^{\frac{1}{2}}}{|2\pi R|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}(x-a)^T K^{-1}(x-a)\}$$

where

$$a = M + KH^T R^{-1}(z - HM)$$
$$K = (P^{-1} + H^T R^{-1} H)^{-1} \quad \text{or} \quad K = P - PH^T(HPH^T + R)^{-1}HP$$

Maximum A Posteriori estimation

(4) determine the MAP estimate of X

$$\begin{aligned}\max_X p\{X|Z\} &\Leftrightarrow \min_X (X-a)^T K^{-1}(X-a)\\ &\Leftrightarrow \hat{X}_{MAP} = a\end{aligned}$$

Hence we have

$$\hat{X}_{MAP} = M + KH^T R^{-1}(Z - HM)$$

with

$$K = (P^{-1} + H^T R^{-1} H)^{-1} \quad \text{or} \quad K = P - PH^T(HPH^T + R)^{-1}HP$$

Comparison MAP with MLE

(1) Performance deficiency of ML in the small sample number case

Example 3.2: Consider the experiment of throwing coins, estimate the probability of the coin is in face.

Define the probability of random matter

$$\theta = p\left\{\text{The coin is face.}\right\}$$

Through n independent experiments, it is found that the coin is in face h times. We have

$$p\left(Z|\theta\right) = \frac{n!}{h!n-h!}\theta^{h}(1-\theta)^{n-h}$$

Comparison MAP with MLE

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0 \Rightarrow \frac{h}{\theta} - \frac{n-h}{1-\theta} = 0 \Rightarrow \hat{\theta}_{ML} = \frac{h}{n}$$

In the case that n=1, $\hat{\theta}_{ML} = 0 \quad \text{or} \quad 1$

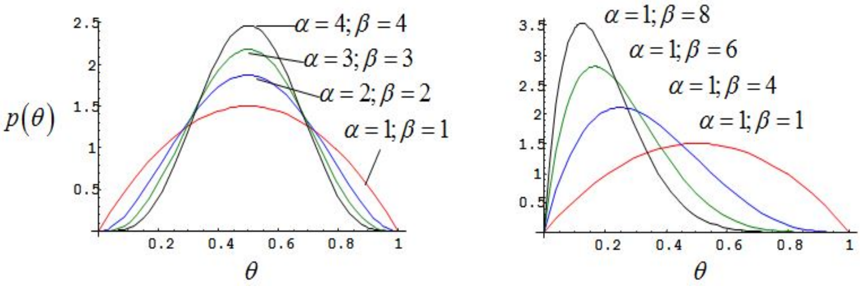
In such situation, the ML estimate may significantly deviate from the truth.

Comparison MAP with MLE

In MAP estimate,we have a priori PDF

$$p\left(\theta\right)=\frac{1}{c}\theta^{\alpha}(1-\theta)^{\beta} \quad \text{with} \quad c=\int_{0}^{1}\theta^{\alpha}(1-\theta)^{\beta}d\theta$$

Beta distribution



In $\alpha+\beta$ independent experiments, the coin is in face α times.

Comparison MAP with MLE

$$p\left(Z|\theta\right) = \frac{n!}{h!n-h!}\theta^h(1-\theta)^{n-h}$$

$$p\left(\theta|Z\right) = \frac{\frac{1}{c}\theta^\alpha(1-\theta)^\beta \frac{n!}{h!n-h!}\theta^h(1-\theta)^{n-h}}{\int\limits_0^1 \frac{1}{c}\theta^\alpha(1-\theta)^\beta \frac{n!}{h!n-h!}\theta^h(1-\theta)^{n-h} d\theta}$$

$$= \frac{\theta^{\alpha+h}(1-\theta)^{\beta+n-h}}{\int\limits_0^1 \theta^{\alpha+h}(1-\theta)^{\beta+n-h} d\theta}$$

Comparison MAP with MLE

In the case that $\alpha = 1$,$\beta = 1$,we have

$$p\left(\theta|Z\right) == \frac{\theta^{1+h}(1-\theta)^{1+n-h}}{\int\limits_0^1 \theta^{1+h}(1-\theta)^{1+n-h}d\theta} \Rightarrow \hat{\theta}_{MAP} = \frac{h+1}{n+2}$$

In the case that $n = 1$, $\hat{\theta}_{MAP} = 1/3 \quad \text{or} \quad 2/3$ $\qquad \hat{\theta}_{ML} = 0 \quad \text{or} \quad 1$

Conclusion: the MAP estimate is closer to the actual value $1/2$ than the ML estimate if the proper apriori information is introduced.

Comparison MAP with MLE

(2) The equivalence between MAP and ML

Based on Bayesian formula, we have

$$
\begin{aligned}
& \boxed{p\{X \mid Z\}} \\
&= \frac{\boxed{p\{X, Z\}}}{\boxed{p\{Z\}}} \\
&= \frac{\boxed{p\{Z \mid X\}}\,\boxed{p\{X\}}}{p\{Z\}}
\end{aligned}
$$

A Posteriori PDF of X

Joint PDF of X and Z

PDF of Z

A Priori PDF of X

Likelihood Function of X

Comparison MAP with MLE

(2) The equivalence between MAP and ML

$$\frac{\partial \ln p(x|z)}{\partial x} = \frac{\partial \ln p(z|x)}{\partial x} + \frac{\partial \ln p(x)}{\partial x} - \frac{\partial \ln p(z)}{\partial x}$$

$$= \frac{\partial \ln p(z|x)}{\partial x} + \frac{\partial \ln p(x)}{\partial x}$$

In general,

$$\frac{\partial \ln p(x)}{\partial x} \neq 0 \Rightarrow \hat{X}_{MAP} \neq \hat{X}_{ML}$$

Comparison MAP with MLE

(2) The equivalence between MAP and ML

In the case that X is uniformly distributed, we have

$$\frac{\partial \ln p(x)}{\partial x} = 0$$

In the case that there is no a priori information about $\mathrm{X}, \mathrm{p}\{\mathrm{X}\}$ can be looked as the Gaussian distribution with infinite variance. Thus we have

$$\tfrac{\partial \ln p(x)}{\partial x} = 0 \Leftrightarrow P^{-1}(X - M) = 0$$

In these cases, the equivalence between MAP and ML holds.

Comparison MAP with MLE

Example3.3:

Conditions

A fishpond has carps (Fish 1) and crucians (Fish 2).

Fish 1 and Fish 2 are different in the color.

Requirement

Please design a fish sorting system to automatically discern Fish 1 and Fish 2.

Comparison MAP with MLE

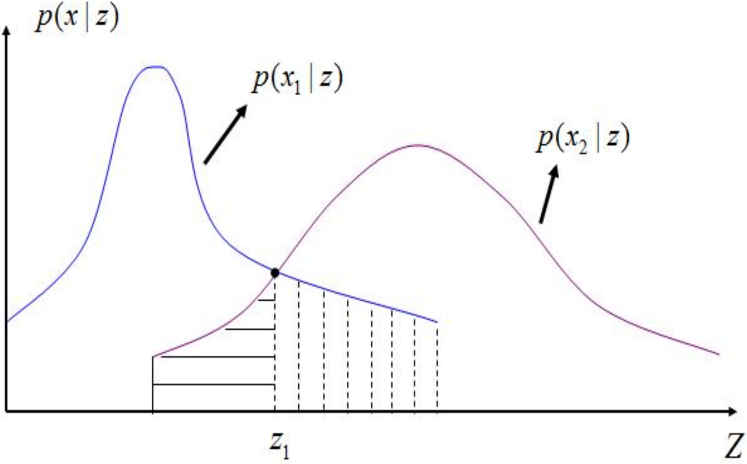
Solution

Dual-mode pattern recognition

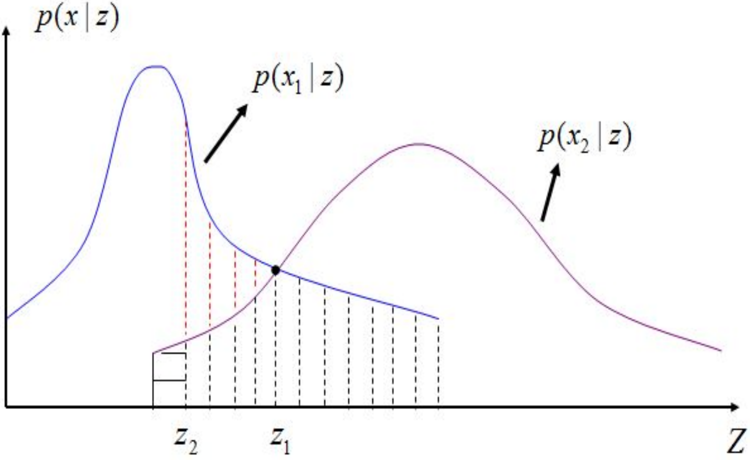
X=1 and X=2 represent "Fish 1" and "Fish 2", respectively.

The color of each sample is normalized within (0, 255). It is the measurement.

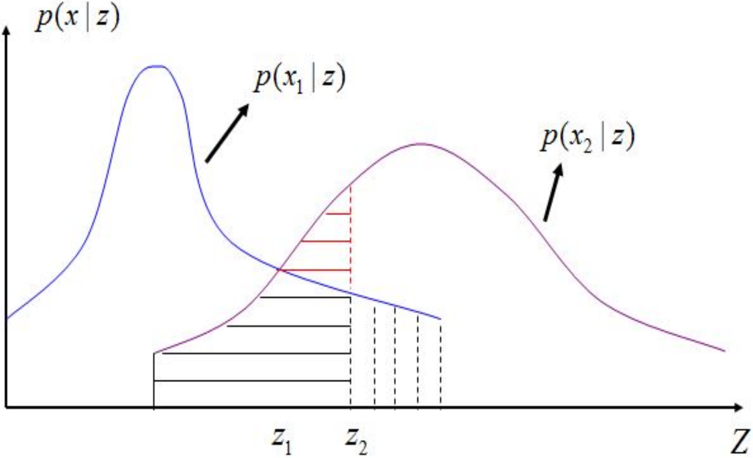
Comparison MAP with MLE



Comparison MAP with MLE



Comparison MAP with MLE



Comparison MAP with MLE

The MAP estimate always has the least mis-sort probability!

Comparison MAP with MLE

Step 1: determine a priori PDF

Net the fishes within the suitable zone of the fish pond, get N_1 piece of Fish1 and N_2 piece of Fish2.

Hence we have the a priori PDF

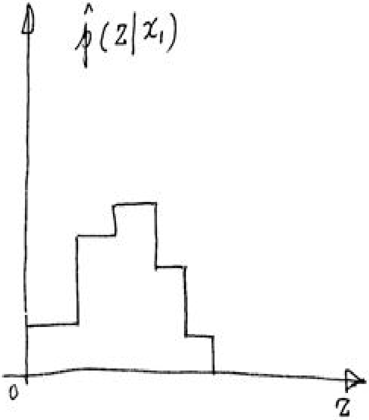
$$P\{x_1\} = N_1/N$$
$$P\{x_2\} = N_2/N$$

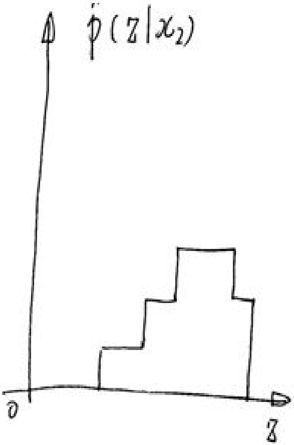
where the total sample number $N = N_1 + N_2$.

Comparison MAP with MLE

Step 2: determine the color histogram

To Fish1 and Fish2, respectively get the color histogram.

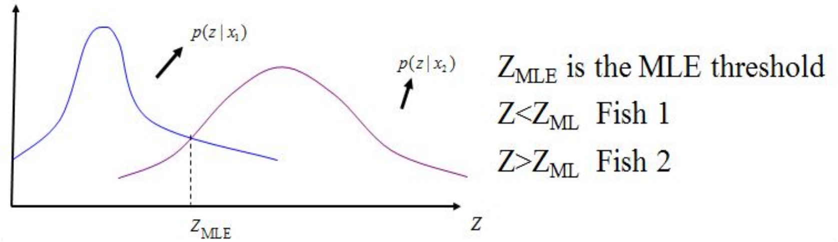




Comparison MAP with MLE

Step 3: determine the likelihood function

Normalize the color histogram as the sampled PDF. If N is not enough large, the past similar data can be utilized.



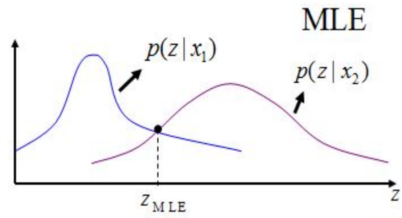
Comparison MAP with MLE

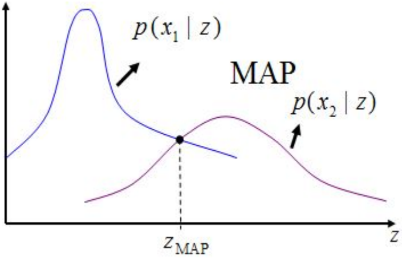
Step 4: determine the conditional PDF

Z_2 is the MAP threshold

If $\mathrm{Z} < \mathrm{Z}_{\mathrm{MAP}}$ then Fish 1; Fish 2 otherwise MAP

If $\mathrm{Z} < \mathrm{Z}_{\mathrm{MLE}}$ then Fish 1; Fish 2 otherwise MLE





Homework

Homework 1:

Given $\mathrm{E}\{\mathrm{X}\}, \mathrm{E}\{\mathrm{XZ}\} \quad \text{and} \quad \mathrm{E}\{\mathrm{XZ}^2\}$, and the MMSE estimate of has the following expression

$$\hat{X} = cZ^2 + aZ + b$$

Please estimate a, b and c, respectively.

Homework

Homework 2: Suppose X and Z are Gaussian, and their joint PDF is

$$p\left(x,z\right) = \frac{\exp\left\{\frac{-\left[\frac{(x-m_x)^2}{{\delta_x}^2}-\frac{2r(x-m_x)(z-m_z)}{\delta_x\delta_z}+\frac{(z-m_z)^2}{{\delta_z}^2}\right]}{2(1-r^2)}\right\}}{2\pi\delta_x\delta_z\sqrt{1-r^2}}$$

where

$$X \sim N\{m_x, \sigma_x^2\}, Z \sim N\{m_z, \sigma_z^2\}, E(X-m_x)(Z-m_z) = r\sigma_x\sigma_z$$

Please determine the MMSE estimate of X.

Homework

Hint

First determine the conditional PDF

$$p\left(x|z\right)=\frac{\exp\left\{\left[\frac{-1}{2\delta_x^z\left(1-r^2\right)}\right]\left[x-m_x-\frac{r\delta_x}{\delta_z}\left(z-m_z\right)\right]^2\right\}}{\delta_x\sqrt{2\pi\left(1-r^2\right)}}$$

And then utilized the integration for estimation

$$\hat{X}_{MV}=E[X|Z]=m_x+\frac{r\sigma_x}{\sigma_z}(z-m_z)$$

Appendix

The derivative of f(scalar) with respect to A(matrix)

$$\frac{\partial f}{\partial A} \triangleq \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1m}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2m}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f}{\partial a_{n1}} & \frac{\partial f}{\partial a_{n2}} & \cdots & \frac{\partial f}{\partial a_{nm}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial A_1} \\ \frac{\partial f}{\partial A_2} \\ \vdots \\ \frac{\partial f}{\partial A_n} \end{bmatrix} = \left[\frac{\partial f}{\partial a_{ij}}\right]_{n\times m},$$

$$\text{with} \quad A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} = [a_{ij}]_{n\times m}$$

Appendix

The derivative of f(scalar) with respect to A(matrix)

$$\frac{\partial \mathrm{trace}(Q)}{\partial A^T} = \left(\frac{\partial \mathrm{trace}(Q)}{\partial A}\right)^T$$

$$\frac{\partial \mathrm{trace}(AB)}{\partial A} = B^T \qquad \frac{\partial \mathrm{trace}(BA^T)}{\partial A} = B$$

$$\frac{\partial \mathrm{trace}(ABA^T)}{\partial A} = A(B + B^T) \qquad \frac{\partial \mathrm{trace}(A^T BA)}{\partial A} = (B + B^T)A$$