Chapter 8.

4. $P(I \mid X=k) = ?$

(a)

$$P(X=k \mid I) = 10 \cdot P(X_i = k \mid I)$$

$$\cancel{= 10 \cdot (P(X_i = k/10;\ \text{know} \mid \theta) + P(X_i = k/10;\ \text{guess} \mid \theta))}$$

since $X_i \sim \cancel{\text{Poisso}}\ \text{Bernoulli}(\theta_i + \frac{1}{3}(1-\theta_i))$

$$X \sim \text{~~~}(\tfrac{1}{3} + \tfrac{2}{3}\theta_i, 10)$$

$$P(X=k \mid I=i) = \binom{10}{k}(\tfrac{1}{3}+\tfrac{2}{3}\theta_i)^k \cdot (\tfrac{2}{3}-\tfrac{1}{3}\theta_i)^{10-k}.$$

$$P(I=i) = \tfrac{1}{3},$$

$$I_{MAP} = \arg\max_I P(\cancel{X=k}\ I \mid X=k) = \arg\max_I P(X=k \mid I=i)$$

$$= \arg\max_i \binom{10}{k}(\tfrac{1}{3}+\tfrac{2}{3}\theta_i)^k \cdot (\tfrac{2}{3}-\tfrac{1}{3}\theta_i)^{10-k} \cdot \tfrac{1}{3}$$

$$= \text{~~~}$$

(b) $\cancel{P(M=m \mid \theta=\theta_i) = \binom{5}{m} \cdot \left(\frac{\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^m \left(\frac{\frac{1}{3}-\frac{1}{3}\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^{5-m}}$

$$\cancel{\theta_{MAP} = \arg\max_\theta \binom{5}{m} \cdot \left(\frac{\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^m \cdot \left(\frac{\frac{1}{3}-\frac{1}{3}\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^{5-m}}.$$

$$\cancel{\theta_{LMS} = E[\theta \mid M=m]}$$

= estimate M.

$$P(M=m \mid X=k, \theta=\theta_i) = P(M=m \mid X=k, \theta=\theta_i) = \binom{k}{m}\left(\frac{\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^m \left(\frac{\frac{1}{3}-\frac{1}{3}\theta_i}{\frac{1}{3}+\frac{2}{3}\theta_i}\right)^{k-m}$$

$$M_{MAP} = \arg\max_m P(M=m \mid X=5) = \arg\max_m \sum_{i=1}^{3} P(M=m \mid X=5, \theta=\theta_i) \cdot P(\theta=\theta_i \mid X=5)$$

$$= \arg\max_m \sum_{i=1}^{3} P(M=m \mid X=5, \theta=\theta_i) \cdot P(X=5 \mid \theta=\theta_i) \cdot P(\theta=\theta_i)$$

$$m_{LMS} = E(M \mid X=5)$$

$$= \sum_{m=0}^{5} \cancel{m} \, P_{M|X}(M=m \mid X=5)$$

$$= \sum_{m=0}^{5} P_{M|X,\theta}(M=m \mid X=5, \theta=\theta_i) \cdot \sum_{i=1}^{3} P_{\theta|X}(\theta=\theta_i \mid X=5)$$

$$= \sum_{m=0}^{5} P_{M|X,\theta}(M=m \mid X=5, \theta=\theta_i) \cdot \sum_{i=1}^{3} P_{\theta|X}(\theta=\theta_i \mid X=5).$$

9. $P(\theta \neq g_n(x_1, \cdots, x_n) \mid X_1 = x_1, \cdots, X_n = x_n)$

$\geq P(\theta = g_{n-1}(x_1, \cdots, x_n) \mid X = x_1, \cdots, X_n = x_n)$

$= P(\theta = g_{n-1}(x_1, \cdots x_{n-1}) \mid X = x_1, \cdots X_{n-1}, x_{n-1})$

$\Rightarrow e_n(x_1, \cdots, x_n) \leq e_{n-1}(x_1, \cdots, x_{n-1})$. $\square$

13. (a) $E(Y_1 \mid Y_1 + \cdots + Y_n) + E(Y_2 \mid Y_1 + \cdots + Y_n) + \cdots + E(Y_n \mid Y_1 + \cdots + Y_n)$

$= E(Y \mid Y)$

$= E(Y \mid Y=y) = \sum_{Y=y'} y' \cdot P_{Y|Y}(y' \mid y)$

$= y \cdot P_{Y|Y}(y \mid y)$

$= Y.$

$\Rightarrow E(Y_1 \mid Y_1 + \cdots + Y_n) = \frac{Y}{n}.$

(b) $\theta \sim \text{Normal}(0, k)$, $W \sim \text{Normal}(0, m)$.

assume that $X_i \sim \text{Normal}(0, 1)$.

$\theta = \sum_{i=1}^{k} X_i$, $\quad W = \sum_{j=1}^{m} X_j$.

$$\Rightarrow E(\theta \mid \theta + W) = E\left(\sum_{i=1}^{k} X_i \,\middle|\, \sum_{j=1}^{m} X_j + \sum_{i=1}^{k} X_i\right) = \frac{k}{k+m} \cdot \text{Normal}(0, k+m)$$

$$= \frac{k}{k+m}(\theta + W)$$

(c) $\theta \sim \text{Poisson}(k)$, $W \sim \text{Poisson}(m)$

$\theta + W \sim \text{Poisson}(k+W)$.

$$E(\theta \mid \theta+W) = \frac{k}{k+m} \cdot (\theta + W).$$

20. (Estimation with spherically invariant PDF's.)

(spherical 球形的).

(quadratic 二次的) (scalar 标量) 标量函数

(a) $E(\theta \mid X=x) = \int_{-\infty}^{+\infty} \theta \cdot f_{\theta|X}(\theta|x)\, d\theta$.

$$f_{\theta,x}(\theta,x) = h\big(a(\theta-\bar{\theta})^2 + b(x-\bar{x})^2 - 2c(\theta-\bar{\theta})(x-\bar{x})\big)$$

scalar function! 返回一个值 return single value.

$$f_X(x) = \int_{-\infty}^{+\infty} \theta \cdot f_{\theta,x}(\theta,x)\, d\theta,$$

$$f_{\theta|x}(\theta|x) = \frac{h(q(\theta,x))}{\int_{-\infty}^{+\infty} \cancel{\theta}\cdot f_{\theta,x}(\theta,x)\, d\theta}$$

$$E(\theta \mid X=x) = \frac{\int_{-\infty}^{+\infty} \theta \cdot f_{\theta,x}(\theta,x)\, d\theta}{\int_{-\infty}^{+\infty} f_{\theta,x}(\theta,x)\, d\theta}. \quad \text{(LMS)}.$$

linear LMS: $\theta_{LLMS} = E(\theta) + \dfrac{\text{cov}(\theta,X)}{\text{var}(X)} \cdot (X - E(X))$.

$$= \bar{\theta} + \frac{E((\theta-\bar{\theta})\cdot(X-\bar{X}))}{\text{var}(X)} \cdot (X - E(X))$$

(b) $q(\theta,x) \geq 0$, h is ~~monototically~~ monotonically decreasing,

$$\theta_{MAP} = \arg\max_{\theta} \frac{h(q(\theta,x))}{\int_{-\infty}^{+\infty} f_{\theta,X}(\theta,x)\, d\theta}, \quad \frac{d f_{\theta|X}(\theta|x)}{d\theta} = \frac{h'(q(\theta,x)) \frac{dq}{d\theta}}{f_X(x)}$$

if $h'(q(\theta,x)) = 0$, $\cancel{\text{且}}$ $\frac{dq(\theta,x)}{d\theta} \neq 0$,

θ is the inflection point, not maximization point.

if $\frac{dq(\theta,x)}{d\theta} = 0$, $\Leftrightarrow$ $2a(\hat{\theta}-\bar{\theta}) - 2c(x-\bar{x}) = 0$

$\theta_{MAP} = \frac{c}{a}\cdot(x-\bar{x}) + \bar{\theta}$.

q is quadratic function, so q has only one maximization point for each fixed point $\cancel{\theta}$, x.

now we proove $\theta_{MAP} = \theta_{LMS}$,

$$\theta_{LMS} = E(\theta \mid X = x)$$

$$= \frac{\int_{-\infty}^{+\infty} \theta f_{\theta,X}(\theta,x)\,d\theta}{f_X(x)},$$

$$f_{\theta,X}(\theta,x) = h(q(\theta,x)) = h(a(\theta-\bar{\theta})^2 + b(x-\bar{x})^2 - 2c(\theta-\bar{\theta})\cdot(x-\bar{x}$$

$$= \qquad ?\ \times$$

2). (Linear LMS)

find a, b, c to minimize $E((\theta - aX - bY - c)^2)$

firstly, the requirement ~~a~~constraint is

$$E(\theta) = E(aX + bY + c)$$

since $E(\theta \mid X = x) = \arg\min_{\hat{\theta}} E((\theta-\hat{\theta})^2 \mid X = x)$ $\hat{\theta}$ is function of x.

$\Rightarrow$ ~~且~~ $E(\theta \mid X = x)$~~其~~ $= \arg\min_{\hat{\theta}} E((\theta-\hat{\theta})^2)$

~~$=E(\theta)$~~. $\Rightarrow E(\theta \mid X = x) = \hat{\theta} \Rightarrow E(\theta) = E(\hat{\theta})$

~~when~~

~~so~~ ~~$E(\theta)=$~~ $\Rightarrow E(\theta - \hat{\theta}) = 0$.

$$E(\theta) = E(aX + bY + c)$$

we have $c = E(\theta) - aE(X) - bE(T)$,

substitute into

$$E((\theta - aX - bT - c)^2) = Var(\theta - aX - bT)$$

$$= Var(\theta) - 2cov(\theta, aX + bT) + Var(aX + bT)$$

$$= Var(\theta) - 2a\,cov(\theta, X) - 2b\,cov(\theta, T) + a^2 Var(X) + b^2 Var($$

$$= Var(X)\cdot\left(a - \frac{cov(\theta, X)}{Var(X)}\right)^2 + Var(T)\cdot\left(b - \frac{cov(\theta, T)}{Var(T)}\right) + C$$

when $a = \frac{cov(\theta, X)}{Var(X)}$, $b = \frac{cov(\theta, T)}{Var(T)}$, it can be minimized.

$c = E(\theta) - aE(X) - bE(T)$.

22. $E(\theta - a_1X_1 - a_2X_2 - \cdots - a_nX_n - b) = 0$

$\Rightarrow b = E(\theta) - \sum_{i=1}^{n} a_i E(X_i)$

$$E\left(\left(\theta - \sum_{i=1}^{n} a_i X_i - E(\theta) + \sum_{i=1}^{n} a_i E(X_i)\right)^2\right) = Var\left(\theta - \sum_{i=1}^{n} a_i X_i\right)$$

$$= Var\left(\left(1 - \sum_{i=1}^{n} a_i\right)\theta - \sum_{i=1}^{n} a_i W_i\right)$$

$$= Var(\theta)\cdot\left(1 - \sum_{i=1}^{n} a_i\right)^2 + \sum_{i=1}^{n} a_i^2 Var(W_i) \triangleq RHS$$

we need $\frac{\partial RHS}{\partial a_1} = \cdots = \frac{\partial RHS}{\partial a_n} = 0$, it is the ~~require~~ necessary condition to ~~imply~~ minimization. (necessary and suffici condition)

$\Rightarrow 2a_i Var(W_i) - 2Var(\theta)\cdot\left(1 - \sum_{i=1}^{n} a_i\right) = 0$

$\Rightarrow a_i = \frac{\sigma_0^2}{\sigma_i^2}\left(1 - \sum_{i=1}^{n} a_i\right)$, assume that $a_i = \frac{1}{\sigma_i^2}\cdot k$ $(k > 0)$.

$\Rightarrow k = \sigma_0^2\left(1 - \sum_{i=1}^{n} \frac{1}{\sigma_i^2} k\right) \Rightarrow k = \frac{1}{\sum_{i=0}^{n} \sigma_i^2}$, $\Rightarrow a_i = \frac{1}{\sigma_i^2}\cdot\frac{1}{\sum_{i=0}^{n} \sigma_i^2}$.

23. $\tilde{\Theta}_L = \hat{\Theta}_L - \Theta$, associated error. (linear LMS)

$\tilde{\Theta} = \hat{\Theta} - \Theta$, LMS.

(a) $E(\tilde{\Theta}_L) = E(\hat{\Theta}_L) - E(\Theta)$, $\hat{\Theta}_L = E(\Theta) + \frac{cov(\Theta, X)}{Var(X)}(X - E(X))$

$= 0$

(b) $cov(\tilde{\Theta}_L, X) = E((\hat{\Theta}_L - \Theta)(X - E(X)))$

$= E((\hat{\Theta}_L - \Theta)X) - E(\hat{\Theta}_L - \Theta) \cdot E(X)$.

$= E\left(\left(E(\Theta) - \Theta + \frac{cov(\Theta, X)}{Var(X)} \cdot (X - E(X))\right)X\right) - E(\sim) \cdot E(X)$

$= 0$

(d) $\hat{\Theta} = E(\Theta | X)$, $\tilde{\Theta} = \hat{\Theta} - \Theta$.

$E((\tilde{\Theta} - E(\tilde{\Theta}))(h(X) - E h(X)))$

$= E((\tilde{\Theta}) \cdot h(X))$

$= E((E(\Theta|X) - \Theta) \cdot h(X))$

$= E(E(\Theta|X) \cdot h(X) - \Theta h(X))$

$= \cancel{E(E(\Theta|X) \cdot h(X)|X)}\ E(E(\Theta \cdot h(X)|X)) - E(\Theta \cdot h(X))$

$= E(\Theta \cdot h(X)) - E(\Theta \cdot h(X)) = 0$.

(c) $var(\tilde{\Theta}_L) = \cancel{Var(\hat{\Theta}_L) + var(\Theta)}\ Var(\hat{\Theta}_L - \Theta)$

$= var(\cancel{E(\Theta)} - \Theta + \frac{cov(\Theta, X)}{Var(X)}(X - E(X)))$

$= var(\frac{cov(\Theta, X)}{Var(X)} X - \Theta)$

$= \frac{cov(\Theta, X)^2}{Var(X)} + Var(\Theta)$

(c) $Var(\theta) = var(\hat{\theta}_L - \tilde{\theta}_L) = var(\hat{\theta}_L) + var(\tilde{\theta}_L)$.

(e) $cov(\tilde{\theta}, x) = E(\hat{\theta}(X - E(X)))$

$$= E(\hat{\theta} X - \hat{\theta} \cdot E(X))$$

$$= E(E(\theta|X)X) - E(\theta|X)E(X)$$

$$= E(\theta X) - E(\theta|X) \cdot E(X).$$

not necessarily equal to 0.

(f)

24. $E(\tilde{\theta}_L) = 0$

$cov(\tilde{\theta}_L, X_i) = E(\tilde{\theta}_L \cdot (X_i - E(X_i)))$

$\cancel{= E((\hat{\theta} E(\theta|X_1 - X_n) - \theta)(X_i - E(X_i)))}$

set $\hat{\theta}_L = a_1 x_1 + \cdots + a_n x_n + b$.

$b = E(\theta) - E(a_1 x_1 + \cdots + a_n x_n)$.

$E(\tilde{\theta}_L \cdot x_i) = E((a_1 x_1 + \cdots + a_n X_n + b - \theta) \cdot x_i)$

$E(\tilde{\theta}_L) = 0$,

$E((a_1 x_1 + \cdots + a_n x_n + b - \theta)^2)$, is the minimization,

$$\frac{\partial \sim}{\partial a_1} = E(2(a_1 x_1 + \cdots + a_n x_n + b - \theta) \cdot a_1) = 0.$$

$\cancel{\Rightarrow a_1 E(x_1) + \cdots + a_n E(x_n)}$

Since $E((\hat{\theta}_L - \theta)^2)$ reach the minimum value for linear functi $g(x_1, \ldots x_n)$, that is $E((g(x_1 \ldots x_n) - \theta)^2) \geq E((\hat{\theta}_L - \theta)^2)$

set $E((\hat{\theta}_L + a_1 X_1 - \theta)^2)$, reach the minimum when $a_1 = 0$.

when $a_1 = 0$, and we have $\left.\frac{\partial E((\hat{\theta}_L + a_1 x_1 - \theta)^2)}{\partial a_1}\right|_{a_1=0} = 0$

~~(Roye's Proposition)~~

(Rolle's Theorem)

$\Rightarrow E(2x_1 \cdot (\hat{\theta}_L + a_1 x_1 - \theta)) = 0$, when $a_1 = 0$ (Rolle's Theorem).

$\Rightarrow E(2x_1 \cdot \hat{\theta}_L - 2x_1 \theta) = 0 \Rightarrow E(x_1 \cdot (\hat{\theta}_L - \theta)) = 0$

$\Leftrightarrow E(x_1 \cdot \tilde{\theta}_L) = 0$. $\square$ ~~we get prooved.~~

~~yes as we can.~~