Chapter 3: Bayesian methods

3.1 Fundamental elements of Bayesian inferrence

Bayes theorem (Law)

suppose $\theta \sim \pi(\theta)$; $x|\theta \sim f(x;\theta)$

then the posterior density of θ, given x

$$\pi(\theta|x) = \pi(\theta) \cdot f(x;\theta) \Big/ \int_\Theta \pi(\theta) \cdot f(x;\theta)\, d\theta$$

$$\pi(\theta|x) \propto \pi(\theta) \cdot \text{likelihood}$$

$\Rightarrow$ examples:

$$X|\theta \sim Bin(n, \theta) \Rightarrow \pi(x|\theta) = \binom{n}{x} \cdot \theta^x (1-\theta)^{n-x}$$

$$\theta \sim Beta(a, b) \Rightarrow \pi(\theta) = \frac{\theta^{a-1} \cdot (1-\theta)^{b-1}}{B(a, b)}$$

$$B(a, b) = \int_0^1 \theta^{a-1} \cdot (1-\theta)^{b-1}\, d\theta$$

$\Rightarrow$ A note:

$$\pi(\theta|x) = \frac{\pi(\theta) f(x;\theta)}{\int_\Theta \pi(\theta) f(x;\theta) d\theta} = \frac{c\,\pi(\theta) \cdot f(x;\theta)}{\int c\,\pi(\theta) \cdot f(x;\theta)\, d\theta}$$

c is free of θ

$\Rightarrow$ how to find the posterior $\pi(\theta|x)$

example 1: $\pi(\theta) \cdot f(x;\theta) \propto \theta^{a-1} \cdot (1-\theta)^{b-1} \cdot \theta^x \cdot (1-\theta)^{n-x}$

$$= \theta^{a+x-1} \cdot (1-\theta)^{n+b-x-1}$$

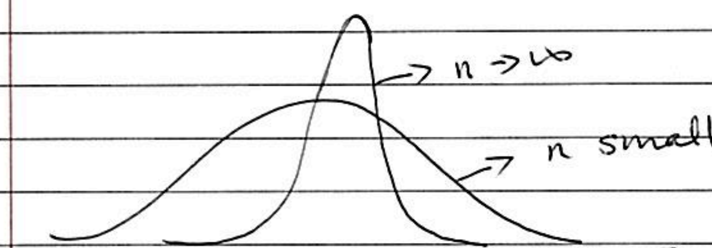
So, we can conclude that:

$$\theta \mid x \sim B(a+x,\ b+n-x)$$

$$E(\theta \mid x) = \frac{a+x}{a+b+n} \qquad \text{as } n \to \infty \quad E(\theta \mid x) \simeq \frac{x}{n}$$

$$\mathrm{var}(\theta \mid x) = \frac{(a+x)(n+b-x)}{(a+b+n)^2(a+b+n+1)} \approx \frac{x(n-x)}{n^3}$$

$$= \frac{x}{n} \cdot \left(1 - \frac{x}{n}\right) \cdot \frac{1}{n} = \hat{p}(1-\hat{p}) \cdot \frac{1}{n}$$



example 2:

$$x_1, x_2, \cdots, x_n \mid \mu \qquad x_i \sim N(\mu, \alpha^2)$$

assume α^2 is know, $\mu \sim N(\mu_0, \sigma_0^2)$

To find $\pi(\mu \mid x_1, \cdots, x_n)$ $\quad \underset{\sim}{x} = (x_1, \cdots, x_n)$

$$\pi(\mu \mid x_1, \cdots, x_n) \propto \pi(\mu) \cdot f(\underset{\sim}{x}; \theta)$$

$$= e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\alpha^2}}$$

$$\propto e^{-\left\{\left(\frac{1}{2\sigma_0^2} + \frac{n}{2\alpha^2}\right)\mu^2 + \left(-\frac{\mu_0}{\sigma_0^2} - \frac{\Sigma x_i}{\sigma^2}\right)\mu\right\}}$$

We see $\mu \mid \underset{\sim}{x}$ is a normal, suppose $N(\mu_1, \sigma_1^2)$

We will find μ_1 and σ_1^2

$$\pi(\mu \mid \underset{\sim}{x}) \propto e^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}} \propto e^{-\left(\frac{1}{2\sigma_1^2}\mu^2 - \frac{\mu_1}{\sigma_1^2}\mu\right)}$$

So, $\dfrac{1}{2\sigma_1^2} = \dfrac{1}{2\sigma_0^2} + \dfrac{n}{2\sigma^2} \iff$

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

Let $\tau_0 = \dfrac{1}{\sigma_0^2}$; $\tau_1 = \dfrac{1}{\sigma_1^2}$; $\tau = \dfrac{1}{\sigma^2}$

$$\Rightarrow \tau_1 = \tau_0 + n \cdot \tau$$

$\iff$ posterior precision = prior precision + $\sum_{i=1}^{n}$ Precision of x

So, $\dfrac{\mu_1}{\sigma_1^2} = \dfrac{\mu_0}{\sigma_0^2} + \dfrac{\sum x_i}{\sigma^2} \iff$

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_1^2}} = \frac{\mu_0\tau_0 + \bar{x} \cdot n \cdot \tau}{\tau_1}$$

$$= \frac{\mu_0\tau_0 + \bar{x} \cdot n\tau}{\tau_0 + n\tau}$$

So, $\mu \mid x_1, \cdots, x_n \sim N\left(\dfrac{\mu_0\tau_0 + \bar{x} \cdot n\tau}{\tau_0 + n\tau}, \dfrac{1}{\tau_0 + n\tau}\right)$

<< January 23, 2015 >> STAT 846

Examples :

1) parameter $\theta \sim$ (prior Dis. $\pi(\theta)$)

θ	1	2	3
$\pi(\theta)$	$1/3$	$1/3$	$1/3$

Data Distribution $\pi(x|\theta)$

$x \mid \theta \sim$	$\theta-1$	θ	$\theta+1$
$\pi(x\mid\theta)$	$1/4$	$1/2$	$1/4$

$$\pi(\theta \mid x) \propto \pi(\theta) \cdot \pi(x \mid \theta)$$

$\pi(\theta) \cdot \pi(x \mid \theta)$

θ \ x	0	1	2	3	4
1	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$	0	0
2	$\frac{1}{3}\times 0$	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$	0
3	$\frac{1}{3}\times 0$	$\frac{1}{3}\times 0$	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$

$\pi(\theta \mid x=0) \sim$

θ	1	2	3
$\pi(\cdot)$	1	0	0

$$\pi(\theta \mid x) = \frac{\pi(\theta) \cdot \pi(x \mid \theta)}{\sum_{\theta=1}^{3} \pi(\theta) \cdot \pi(x \mid \theta)}$$

$\pi(\theta \mid x=1) \sim$

θ	1	2	3
$\pi(\cdot)$	$\frac{2}{3}$	$\frac{1}{3}$	0

$\pi(\theta \mid x=2) \sim$

θ			
$\pi(\cdot)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

2) $X_1, \cdots, X_n \mid (\mu, \tau) \sim N(\mu, \frac{1}{\tau})$

$\tau \sim Gamma(\alpha, \beta)$

$$\pi(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \cdot e^{-\beta\tau}$$

$\mu \mid \tau \sim N(\upsilon, \frac{1}{k\tau})$

$$\pi(\mu \mid \tau) = \frac{1}{\sqrt{2\pi}} (k\tau)^{\frac{1}{2}} e^{-\frac{k\tau}{2}(\mu-\upsilon)^2}$$

$$\pi(\tau, \mu) \propto \tau^{\alpha-\frac{1}{2}} e^{-\tau(\beta + \frac{k}{2}(\mu-\upsilon)^2)}$$

$$\pi(\tau, \mu \mid x_1, x_2, \cdots, x_n) \propto \pi(\tau, \mu) \cdot f(x_1, \cdots x_n; \tau, \mu)$$

$$\propto \tau^{\alpha-\frac{1}{2}} \cdot e^{-\tau(\beta + \frac{k}{2}(\mu-\upsilon)^2)} \prod_{i=1}^{n} \tau^{\frac{1}{2}} \cdot e^{-\frac{\tau}{2}(x_i-\mu)^2}$$

$$\propto \tau^{\alpha'-\frac{1}{2}} \cdot e^{-\tau(\beta' + \frac{k'}{2}(\mu-\upsilon')^2)}$$

where, $\alpha' = \alpha + n/2$

$$\beta' = \beta + \frac{1}{2} \cdot \frac{nk}{n+k}(\bar{x}-\upsilon)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$k' = k + n$$

$$\upsilon' = \frac{k\upsilon + n\bar{x}}{k+n}$$

$\Rightarrow$ Conjugate prior : if the parametric form of the prior and the posterior is the same, then we say this prior is conjugate to this problem.

⇒ General form of Bayes rule, give a fixed prior.

⇒ Risk function:

$$R(\theta, d) = \int_X L(\theta, d(x))\, f(x,\theta)\, dx$$

Bayes Risk:

$$r(\pi, d(x)) = \int_\theta \pi(\theta) \cdot R(\theta, d(x))\, d\theta$$

$$= \int_\theta \pi(\theta) \left[\int_X L(\theta, d(x)) f(x,\theta)\, dx \right] d\theta$$

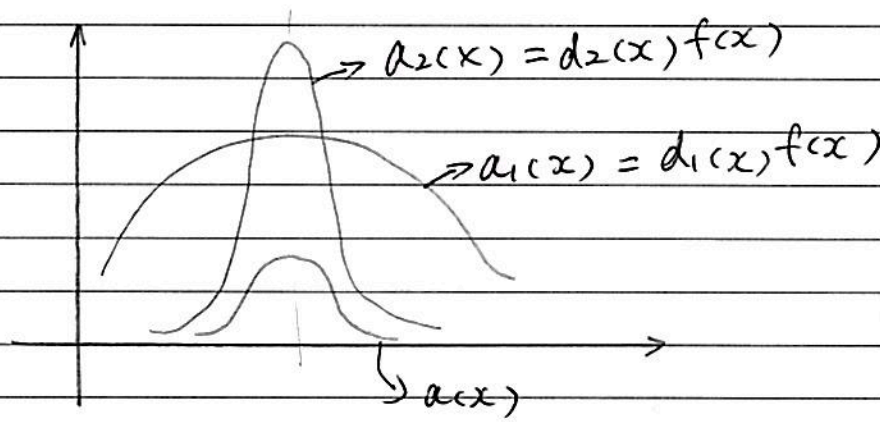
$$= \int_X f(x) \left\{ \int_\theta L(\theta, d(x)) \cdot \frac{\pi(\theta) \cdot f(x,\theta)}{f(x)}\, d\theta \right\} dx$$

where, $f(x) = \int_\theta \pi(\theta) \cdot f(x,\theta)\, d\theta$

$$= \int_X f(x) \cdot \left[\int_\theta L(\theta, d(x))\, \pi(\theta | x)\, d\theta \right] dx$$

To find Bayes Rule $d(x)$ to minimize $r(\pi, d(x))$ is equivalent to that for each x, we find $d(x)$ to minimize $\int_\theta \underbrace{L(\theta, d(x))\, \pi(\theta|x)}_{a(x)}\, d\theta = E_{\theta|x}\{L(\theta, d(x))\}$

Called expected posterior loss



$\Rightarrow$ How to find the Bayes Rule :

examples :

1) Point estimate :

$$L(\theta, a) = (\theta - a)^2$$

$$E_{\theta|x}(L(\theta, d(x))) = E_{\theta|x}((\theta - d(x))^2) = A(d(x))$$

To find a value $d(x)$ to minimize $A(d(x))$,

Solving $\frac{\partial}{\partial d}(A(d(x))) = E_{\theta|x}\left(\frac{\partial}{\partial d}(\theta - d)^2\right) = E_{\theta|x}(2(\theta - d)(-1))$

$$= 0$$

$\Rightarrow E_{\theta|x}(\theta) - d(x) = 0 \quad \Rightarrow d(x) = E_{\theta|x}(\theta)$

So, the Bayes Rule is the mean of posterior θ.

2) Point estimate :

Loss function : $L(\theta, d) = |\theta - d|$

Solution: $E_{\theta|x}(L(\theta, d)) = E_{\theta|x}(|\theta - d|)$

$$= \int_{-\infty}^{\infty} |\theta - d| \, \pi(\theta|x) \, d\theta$$

$$= \int_{-\infty}^{d} (d - \theta)\pi(\theta|x) \, d\theta + \int_{d}^{\infty} (\theta - d)\pi(\theta|x) \, d\theta$$

Differential : $\frac{\partial}{\partial d}\{E_{\theta|x}(L(\theta, d))\} = 0$

$$\Rightarrow \int_{-\infty}^{d} \pi(\theta|x) \, d\theta - \int_{d}^{\infty} \pi(\theta|x) \, d\theta = 0$$

$\because \int_{-\infty}^{\infty} \pi(\theta|x) \, d\theta = 1$

$$\Rightarrow \int_{-\infty}^{d} \pi(\theta|x) \, d\theta = \int_{d}^{\infty} \pi(\theta|x) \, d\theta = \frac{1}{2}$$

《 January 26, 2015 》 STAT 846

⇒ Exampres of Bayes Rules

1) point estimates:

$$L(\theta, a) = |\theta - a|$$

$$E_{\theta|x}(|\theta - d(x)|) = \int_{-\infty}^{\infty} |\theta - d(x)| \, \pi(\theta|x) \, d\theta$$

$$= r(d(x)|x)$$

$$\frac{\partial r(d(x)|x)}{\partial d(x)} = \int_{-\infty}^{\infty} [1 \times I(\theta < d) + (-1) I(\theta > d)] \pi(\theta|x) \, d\theta$$

$$= \int_{-\infty}^{d} \pi(\theta|x) \, d\theta - \int_{d}^{\infty} \pi(\theta|x) \, d\theta = 0$$

$$\Rightarrow Pr(\theta < d|x) = Pr(\theta > d|x)$$

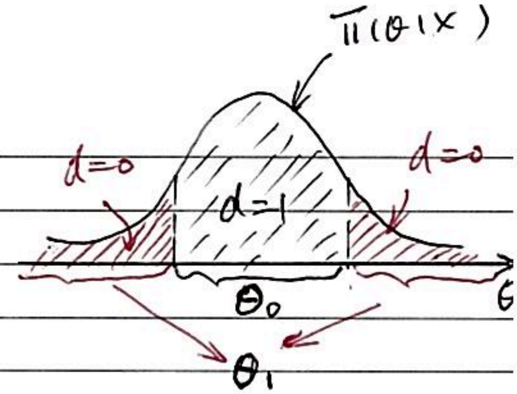
$d(x)$ = median of $\pi(\theta|x)$

2) In hypothesis testing

$$L(\theta, a) = \begin{cases} 1 & \text{if } \theta \in \Theta_0, \ a = 1 \\ 1 & \text{if } \theta \in \Theta_1, \ a = 0 \\ 0 & \text{o/w} \end{cases}$$

$$r(d|x) = \begin{cases} E_{\theta|x}\{L(\theta, d=1)\}, & \text{if } d = 1 \\ E_{\theta|x}\{L(\theta, d=0)\} & \text{if } d = 0 \end{cases}$$

$$= \begin{cases} Pr(\theta \in \Theta_0 | x) & \text{if } d=1 \\ Pr(\theta \in \Theta_1 | x) & \text{if } d=0 \end{cases}$$



Bayes Rule: $d(x)$:

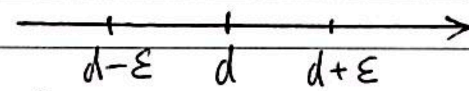
$$d(x) = \begin{cases} 1 & \text{if } Pr(\theta \in \Theta_1 | x) \geq Pr(\theta \in \Theta_0 | x) \\ 0 & \text{if } Pr(\theta \in \Theta_0 | x) \geq Pr(\theta \in \Theta_1 | x) \end{cases}$$

We will come back in chapter 4.

=> Lindley paradox

3) Interval estimate

$$\mathcal{A}_1 = \{d\} = \Theta$$



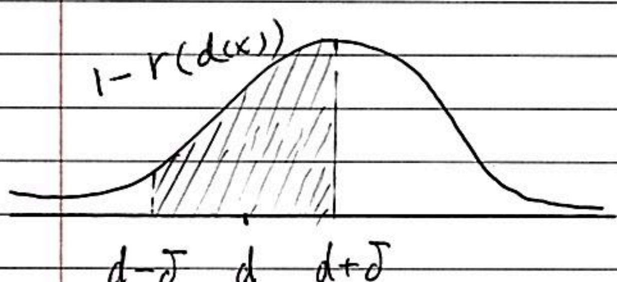
$$\mathcal{A}_1 = \{(d-\delta, d+\delta) \mid d \in \Theta\}$$

Prescribe the length of in interval δ

$$L(\theta, d) = \begin{cases} 0 & \theta \in (d-\delta, d+\varepsilon) \\ 1 & \theta \notin (d-\delta \quad d+\varepsilon) \end{cases}$$

$$r(d|x) = E_{\theta|x}\{L(\theta, d)\} = Pr\{\theta \notin (d-\delta, d+\delta) | x\}$$

the Bayes Rule $d(x)$ is



$$d(x) = \underset{d}{\arg\min}\, Pr(d-\delta < \theta < d+\delta \mid x)$$

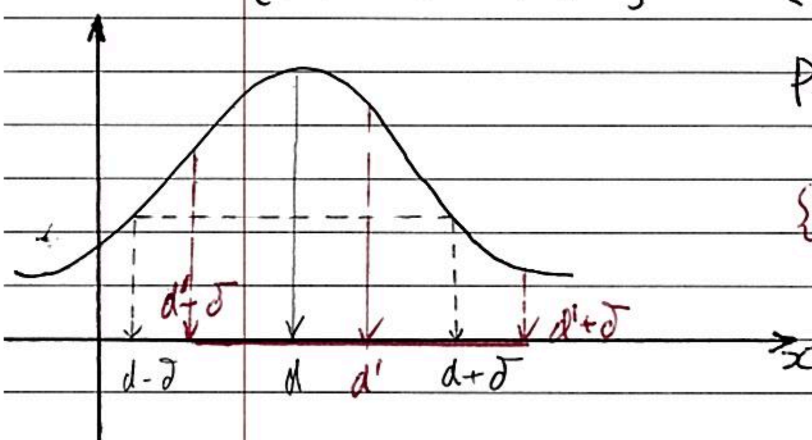
= highest Posterior Density

= HPD

⇒ A special case

If $\pi(\theta|x)$ is unimodel, the Bayes Rule interval has the form

$\{\theta \mid \pi(\theta|x) \geq c\}$ OR $\{d \mid \pi(d+\delta|x) = \pi(d-\delta|x)\}$



$P(\theta \in d' \pm \delta|x) > Pr(\theta \in d \pm \delta|x)$

{But it is higher d to evaluate $\pi(\theta|x)$, when $\pi(\theta)$ OR $P(x|\theta)$ doesn't have close form}

⇒ in practice, we prescribe $Pr(\theta \in I|x) = 1-\alpha$

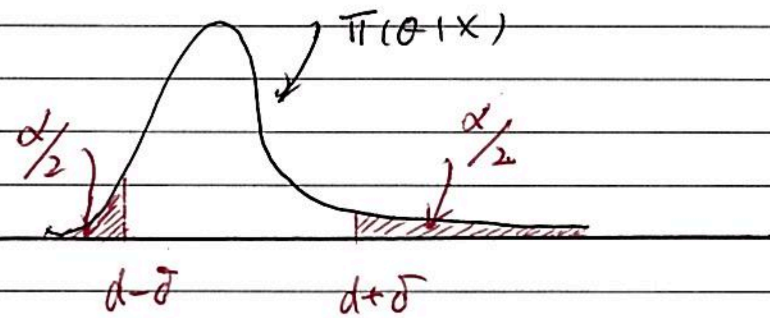
then we search I such that length(I) is minimized.

⇒ In practice, two alternatives:

a) $d(x)$ is set such that

$$Pr(\theta > d(x)|x) = Pr(\theta < d(x)|x)$$

Equil tailed interval



2) Normal interval $(\mu - Z_{\frac{\alpha}{2}} \cdot \sigma \quad \mu + Z_{\frac{\alpha}{2}} \cdot \sigma)$

where, $\mu = E(\theta | x) \qquad \sigma^2 = Var(\theta | x)$

Based on Bayesian Asymptotic

$$\theta | x \sim N(\mu, \sigma^2)$$

$\Rightarrow$ point estimate : how to Find the minimax Rule

$X | \theta \sim Bin.(n, \theta)$

$\theta \sim Beta(a, b)$

$L(\theta, a) = (\theta - a)^2$

$\theta | x \sim Beta(a+x, b+n-x)$

the Bayes Rule is

$$d(x) = E(\theta | x) = \frac{a+x}{a+b+n}$$

calculated the Risk function

If $R(\theta, d(x)) =$ constant for all $\theta \in (0, 1)$

By theorem 2.2, this $d(x)$ is minimax.

Let $c = a+b+n$

$$R(\theta, d) = E_X\{(\theta - d(x))^2\} = E_X\{(\theta - \frac{a+x}{a+b+n})^2\}$$

$$= E_X\{(\theta - \frac{a+x}{c})^2\} = \frac{1}{c^2} E_X\{(c\theta - a - x)^2\}$$

$$= \frac{1}{c^2} E_X\{(c\theta - a)^2 - 2(c\theta - a) \cdot x + x^2\}$$

$$= \frac{1}{c^2} \cdot (c\theta - a)^2 - \frac{2}{c^2}(c\theta - a) E_X(x) + \frac{1}{c^2} E_X(x^2)$$

Note that : $E(X) = n\theta$, $E(X^2) = (n\theta)^2 + n\theta(1-\theta)$

$= \frac{1}{c^2}\{(c\theta - a)^2 - 2(c\theta - a) \cdot n\theta + n^2\theta^2 + n\theta(1-\theta)\}$

To search a, b such that

$R(\theta, d) = \text{constant}$ for all θ

We write that

$$\begin{cases} n + 2na - 2ac = 0 & \text{coefficent of } \theta \\ -n + n^2 - 2nc + c^2 = 0 & \text{coefficent of } \theta^2 \end{cases}$$

$$\Rightarrow \begin{cases} a = \sqrt{n}/2 \\ b = \sqrt{n}/2 \end{cases}$$

the minimax rule is $d(x) = \dfrac{\sqrt{n}/2 + x}{n + \sqrt{n}}$

Note that : $d(x) \xrightarrow{\text{as } n \to \infty} \dfrac{\sqrt{n}/2 + n\theta}{n +} \to \theta$

« January 28, 2015 » STAT 846

$X \sim Bin(n, p) \Rightarrow f(x, p) = \binom{n}{x} p^x (1-p)^{n-x}$

$p \sim Beta(a, b)$

Bayes estimator: $\hat{p}_{BS} = \frac{x+a}{a+b+n}$ $\quad \hat{p} = \frac{X}{n}$.

So $\hat{p}_{BS}$ shrinks $\hat{p}$ to $\frac{a}{a+b+n}$

$\Rightarrow$ Shrinkage and Jame-stein estimation.

$X_i \mid \mu_i \sim N(\mu_i, 1)$ for $i = 1, 2, \cdots, p$

We want to estimate $\mu_1, \mu_2, \cdots, \mu_p$

A decision Rule $d = (d_1, d_2, \cdots d_p)$

Loss function:

$$L(\mu, d) = \sum_{i=1}^{p} (\mu_i - d_i)^2 = \| \mu - d \|^2$$

Let $x = (X_1, X_2, \cdots, X_p)^T$, $d = (d_1, d_2, \cdots, d_p)^T$

A class of d^a: $d^a(x) = \left(1 - \frac{a}{\|x\|^2}\right) \cdot X$

$\Rightarrow$ A special case $a = 0 \Rightarrow d^0(x) = X$

$\Rightarrow$ When $a > 0$, $d^0(x)$ closer to $(0, 0, \cdots, 0)^T$

$\Rightarrow$ We will show that, when $p \geq 3$, d^a dominates d^0.

the Risk function: $R(\mu, d^0(x)) = E_X\{\|X - \mu\|^2\}$

$= \sum_{i=1}^{p} E_{x_i}\{(x_i-\mu_i)^2\} = \sum_{i=1}^{p} 1 = p$

$\Rightarrow$ Stein Lemma: if $x_i \sim N(\mu_i, 1)$, then

$$E_x\{(x_i-\mu_i)h(x)\} = E_x\left(\frac{\partial h(x)}{\partial x_i}\right)$$

proof: suppose that $p=1$, Let φ be pdf of $N(0,1)$

$$\frac{\partial \varphi(x-\mu)}{\partial x} = -\varphi(x-\mu)\cdot(x-\mu)$$

$\left\{\text{Note that: since } \varphi(x-\mu) = \frac{1}{\sqrt{2\pi}}\cdot \exp^{-\frac{(x-\mu)^2}{2}}\right\}$

$$E_x\{(x-\mu)h(x)\} = \int (x-\mu)h(x)\cdot\varphi(x-\mu)\,dx$$

$$= -\int_{-\infty}^{\infty} h(x)\, d(\varphi(x-\mu))$$

$$= -\left\{h(x)\varphi(x-\mu)\Big|_{-\infty}^{\infty} - \int \varphi(x-\mu)\,dh(x)\right\}$$

$$= -h(x)\varphi(x-\mu)\Big|_{-\infty}^{\infty} + \int \varphi(x-\mu)\,dh(x)$$

$$= \int \varphi(x-\mu)\cdot\frac{\partial h(x)}{\partial x}\cdot dx$$

$$= E_x\left\{\frac{\partial h(x)}{\partial x}\right\}$$

therefore, $E_x\{(x_i-\mu_i)h(x)\} = E_x\left(\frac{\partial h(x)}{\partial x}\right)$

$\Rightarrow$ Risk function $R(\mu, d^a) = E_x(\|\mu - d^a(x)\|^2)$

$$= E_x\left\{\left(\left\|\mu - \left(1-\frac{a}{\|x\|^2}\right)\cdot x\right\|\right)^2\right\}$$

$$= E_x\left\{\left(\left\|(\mu-x) + \frac{a}{\|x\|^2}\right\|\right)^2\right\}$$

$$= E_X\{\|\mu - X\|^2) - 2a \cdot E_X\left\{\frac{X^T(X-\mu)}{\|X\|}\right\} + a^2 \cdot E_X\left\{\left(\frac{1}{\|X\|^2}\right)\right\}$$

$$\left\{\text{where } E_X\left(\frac{X^T(X-\mu)}{\|X\|^2}\right)\right.$$

$$= \sum_{i=1}^{p} E_X\left(\frac{X_i(X_i-\mu_i)}{\sum_{j=1}^{p} X_j^2}\right) \quad . \quad h(X_i) = \frac{X_i}{\sum_{j=1}^{p} X_j^2}$$

$$= \sum_{i=1}^{p} E_X\left\{\frac{\partial}{\partial X_i}\cdot\left(\frac{X_i}{\sum_{j=1}^{p} X_j^2}\right)\right\} = \sum_{i=1}^{p} E_X\left\{\frac{\sum_{j=1}^{p} X_j^2 - 2X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right\}$$

Note(1)

$$\left.= (p-2) E_X\left(\frac{1}{\|X\|^2}\right)\right\}$$

$$= p - 2a\cdot(p-2) E_X\left(\frac{1}{\|X\|^2}\right) + a^2 E_X\left(\frac{1}{\|X\|^2}\right)$$

$$= p - (2a(p-2) - a^2) E_X\left(\frac{1}{\|X\|^2}\right)$$

so, $R(\mu, d^a) < R(\mu, d^0) = p$

$\Leftrightarrow \quad 2a(p-2) - a^2 > 0$

$\Leftrightarrow \quad 2(p-2) > a$

when $p \geq 3$ there exist $a > 0$ such that

$a < 2(p-2)$

$\Leftrightarrow$ therefore d^0 is inadmissible

Note 1:

$$\sum_{i=1}^{p} E_x \left\{ \frac{\sum_{j=1}^{p} x_j^2 - 2x_i^2}{\left(\sum_{j=1}^{p} x_j^2\right)^2} \right\}$$

$$= E_x \left\{ \sum_{i=1}^{p} \frac{\|x\|^2 - 2x_i^2}{\|x\|^4} \right\}$$

$$= E_x \left\{ \frac{1}{\|x\|^2} \cdot \sum_{i=1}^{p} 1 - \frac{2}{\|x^4\|} \cdot \sum_{i=1}^{p} x_i^2 \right\}$$

$$= E_x \left(\frac{p}{\|x\|^2} - \frac{2\|x\|^2}{\|x\|^4} \right)$$

$$= E_x \left(\frac{p-2}{\|x\|^2} \right)$$

$$= (p-2) E_x \left(\frac{1}{\|x\|^2} \right)$$

<< January 30, 2015 >> STAT 846

=> Home Run Examples

	pre-season			Regular season						
	y_i	n_i	x_i	N_i (AB)	$\hat{\mu}_i^{JS}$	μ_i	$\hat{\mu}_i = x_i$	HR_i	$\hat{HR}_i^{JS}$	$\hat{HR}_i$
Sosa	7	58	-6.56	509	-7.12	-6.18	-6.56	70	50	61
M. G	9	59	-5.90	643	-6.71	-7.06	-5.90	66	75	98

Note that : $\hat{HR}_i = (y_i / n_i) \times N_i$

$61 = \frac{7}{58} \times 509 \quad 98 = \frac{9}{59} \times 643$

$y_i \mid n_i, p_i \sim Bin(n_i, p_i)$

$x_i = f_{n_i}(\frac{y_i}{n_i})$, when $n_i = n$ $f_n(\frac{y}{n}) = \sqrt{n} \cdot \sin^{-1}(2\frac{y}{n} - 1)$

$SD(y/n) = \sqrt{p(1-p)/n}$, $SD(x_i) = \sigma$

New model : $x_i \mid \mu_i \sim N(\mu_i, \sigma^2)$; $\mu = f_{n_i}(p_i)$

We will then apply J-S to $x_1, \dots, x_p$ to estimate $\mu_1, \dots, \mu_p$, then we can estimate

$p_i = f_{n_i}^{-1}(\mu_i)$

$d^{p-2}(x) = \bar{X} + (1 - \frac{p-2}{V})(X - \bar{X})$

$V = \sum_{i=1}^{p}(x_i - \bar{x})^2$, where, $X = (x_1, \cdots x_p)^T$

$\sum_{i=1}^{p}(\hat{\mu}_i - \mu_i)^2 = 19.68 \qquad \sum_{i=1}^{p}(\hat{\mu}^{JS} - \mu_i)^2 = 8.07$

$\hat{HR}_i^{JS} = f_{N_i}^{-1}(\hat{\mu}_i^{JS}) \cdot N_i \quad \approx \hat{p}^{JS}$

$\hat{HR}_i = \frac{y_i}{n_i} \times N_i$

⇒ Remarks:

when $p=1$, and 2, $\bar{X}$ is admissible under squre Loss.

(*) ⇒ Review: $X \mid \mu \sim N(\mu, \sigma^2)$

$\mu \sim N(\mu_0\ \sigma_0^2)$

What is the marginal distribution of X

Method 1 ⇒ $f(x) = \int_{-\infty}^{\infty} f(x \mid \mu) \cdot \pi(\bar{u})\, du$

Method 2 ⇒ $X = \mu + Z$ where $\mu \sim N(\mu_0, \sigma_0^2)$

$Z \sim N(0, \sigma^2)$

(*) ⇒ A general result: $X_i \overset{iid}{\sim} N(\mu_i\ \sigma_i^2)$,

then $\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$, the distribution of X is $N(\mu_0, \sigma_0^2 + \sigma^2)$

⇒ predictive distribution

Given $X_1, \cdots X_n$ we want to estimate x^* by finding $f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \theta, x_1 \cdots x_n) f(\theta \mid x_1 \cdots x_n) d\theta$

This is a generalization of $f(x) = \int f(x \mid \theta) f(\theta) d\theta$

⇒ A special case $X_1, \cdots, X_n, X^* \mid \theta \overset{iid}{\sim} f(x \mid \theta)$

$f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \theta) \pi(\theta \mid x_1, \cdots, x_n)\, d\theta$

Example: $X_1, \cdots, X_n, X^* \mid \mu \overset{iid}{\sim} N(\mu, \sigma^2)$ σ^2 known

$\mu \sim N(\mu_0, \sigma_0^2)$

Question is how to find the predictive distributio

$f(X^* \mid X_1, \cdots, X_n)$

Sol: $\mu \mid X_1, \cdots, X_n \sim N(\mu_1, \sigma_1^2)$

where, $\sigma_1^2 = \left[\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right]^{-1}$ $\mu_1 = \left[\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2}\cdot\bar{X}\right] \times \sigma_1^2$

$f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \mu)\, \pi(\mu \mid x_1, \cdots, x_n)\, d\mu$

$$\left.\begin{array}{l} x^* \mid \mu \sim N(\mu, \sigma^2) \\ \mu \mid X_1, \cdots X_n \sim N(\mu_1\ \sigma_1^2) \end{array}\right\} \Rightarrow x^* \mid x_1, \cdots X_n \sim N(\mu_1, \sigma_1^2 + \sigma^2)$$

$\Rightarrow$ Emprical Bayes:

$\Rightarrow$ Review: 1) $x_i \mid \mu_i \sim N(\mu_i, 1)$

$\mu_i \sim N(0\ \sigma^2)$

the marginal distribution of X_i

$x_i \sim N(\mu_i\ \ 1 + \sigma^2)$

2) $X \sim \text{Gamma}(\alpha, \lambda) \Rightarrow E(X^k) = ?$

proof: $f(x) = \frac{1}{\Gamma(\alpha)} \cdot \lambda^{\alpha} \cdot x^{\alpha - 1} e^{-\lambda x}$ for $x > 0$

Let $X = Z/\lambda$ where, $Z \sim \text{Gamma}(\alpha, 1)$

$Z \sim f(z) = \frac{1}{\Gamma(\alpha)} \lambda\, z^{\alpha-1} \cdot e^{-z}$ for $z > 0$

$$E(Z^k) = \int_0^\infty Z^k \cdot \frac{1}{\Gamma(\alpha)} Z^{\alpha-1} \cdot e^{-z}\, dz$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty Z^{\alpha+k-1} \cdot e^{-z}\, dz$$

$$= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$$

« Feb, 02, 2015 » STAT 846

$x \sim \text{Gamma}(\alpha, \lambda)$; λ is rate.

$$E(x^k) = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)}$$

$$x \sim \chi_p^2 = \text{Gamma}\left(\frac{p}{2}, \lambda = \frac{1}{2}\right)$$

χ_p^2 is distribution of $x = \sum_{i=1}^{p} z_i^2$

where $z_i \sim N(0, 1)$

$$E(x^k) = \frac{\Gamma(\frac{p}{2}+k)}{(\frac{1}{2})^k \Gamma(\frac{p}{2})} = \frac{\Gamma(\frac{p}{2}+k)}{\Gamma(\frac{p}{2})} \cdot 2^k$$

$$\Rightarrow E(x^{-1}) = \left\{\Gamma\left(\frac{p}{2}-1\right) / \Gamma\left(\frac{p}{2}\right)\right\} \cdot \frac{1}{2}$$

$$= \frac{\Gamma(\frac{p}{2}-1)}{(\frac{p}{2}-1)\Gamma(\frac{p}{2}-1)} \cdot \frac{1}{2} = \frac{1}{p-2} \quad (\text{for } p > 2)$$

$\Rightarrow$ Emparical Bayes:

$$\text{Model: } \begin{cases} x_i \mid \mu_i \sim N(\mu_i, 1) \\ \mu_i \sim N(0, \tau^2) \end{cases} \quad \text{for } i = 1, 2, \dots, p$$

Suppose we know τ^2, Bayes estimator for

$\mu = (\mu_1, \dots, \mu_p)^T$, $\delta^\tau(x) = E_{\mu|x}(\mu|x)$

$$\mu_i \mid x_i \sim N\left(\frac{\tau^2}{1+\tau^2} x_i, \frac{\tau^2}{1+\tau^2}\right)$$

$$\delta^\tau(x) = E_{\mu|x}(\mu|x) = \frac{\tau^2}{1+\tau^2} \cdot x$$

$\Rightarrow$ Bayes Risk: $r(\pi^\tau, \delta^\tau) = E_{\mu\sim\pi^\tau}\{R(\mu, \delta^\tau)\}$

$= E_\mu E_{x|\mu}\{\|\mu - \delta^\tau(x)\|^2 \mid \mu\}$

$= E_x E_{\mu|x}\{\|\mu - \delta^\tau(x)\|^2 \mid x\}$

$= E_x E_{\mu|x}\{\sum_{i=1}^{p}(\mu_i - \delta^\tau(x)_i)^2 \mid x\}$

$= E_x\{\sum_{i=1}^{p} E_{\mu|x}(\mu_i - \delta^\tau(x_i))^2 \mid x\}$

$= E_x\{\sum_{i=1}^{p} Var(\mu_i \mid x_i)\} = E_x\{\sum_{i=1}^{p}\frac{\tau^2}{1+\tau^2}\} = p\cdot\frac{\tau^2}{1+\tau^2}$

$\Rightarrow$ Emprical Bayes estimate:

We want to replace τ^2 with a statistic (of X) using marginal distribution of X

$$\left.\begin{array}{ll} x_i \mid \mu_i & \sim N(\mu_i, 1) \\ \mu_i & \sim N(0, \tau^2) \end{array}\right\} \Rightarrow x_i \sim N(0, 1+\tau^2)$$

If $\frac{x_i}{\sqrt{1+\tau^2}} \sim N(0,1) \Rightarrow \sum_{i=1}^{p}\frac{x_i^2}{1+\tau^2} \sim \chi_p^2$

$E_x\left(\frac{1+\tau^2}{\|x\|^2}\right) = \frac{1}{p-2}$

$\Rightarrow E_x\left\{\frac{1}{\|x\|^2}\right\} = \frac{1}{(p-2)(1+\tau^2)}$

$E_x\left(1 - \frac{p-2}{\|x\|^2}\right) = 1 - \frac{1}{1+\tau^2} = \frac{\tau^2}{1+\tau^2}$

We can replace $\frac{\tau^2}{1+\tau^2}$ by $\left(1 - \frac{p-2}{\|x\|^2}\right)$

{ A note :

Another approach to estimate σ^2

$$x_1, \cdots, x_p \overset{iid}{\sim} N(0, 1+\sigma^2)$$

$$\hat{\sigma}^2 = \frac{\|x\|^2}{p-2} - 1, \quad \widehat{1+\sigma^2} = \sum_{i=1}^{p} x_i^2 / p-2$$

$$\Rightarrow \frac{\hat{\sigma}^2}{1+\hat{\sigma}^2} = 1 - \frac{p-2}{\|x\|}$$

$\Rightarrow$ An emprical Bayes estimate replace $\frac{\sigma^2}{1+\sigma^2}$ in

$$\delta^{\sigma}(x) = \frac{\sigma^2}{1+\sigma^2} X \text{ by } 1 - \frac{p-2}{\|X\|^2}$$

$$d^{p-2}(x) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$$

to find the Bayes Risk of d^{p-2}

$$r(\pi^{\sigma}, d^{p-2}(x)) = E_{\mu} E_{x|\mu} \left(\|\mu - d^{p-2}(x)\|^2 | x \right)$$

$$= E_x E_{\mu|x} \left(\|\mu - \left(1 - \frac{p-2}{\|x\|^2}\right) x\|^2 | x \right)$$

$$= E_x \left\{ \left(p - \frac{p-2}{\|x\|^2} \right) \right\}$$

$$= p - (p-2)^2 E_x \left(\frac{1}{\|x\|^2} \right)$$

$$= p - (p-2) \times \frac{1}{1+\sigma^2}$$

$$= p - \frac{p-2}{1+\sigma^2} = r(\pi^{\sigma}, \delta^{\sigma}(x)) + \frac{2}{1+\sigma^2}$$

$$r(\pi^{\sigma}, \delta^{\sigma}(x)) = p\sigma^2 / 1+\sigma^2$$

<< Feb. 04. 2015 >> STAT 846

⇒ choice of prior {how to choice the prior}

Methods: 1) Empiracal Bayes:

$X|\theta \sim f(X|\theta)$

$\theta|\tau \sim f(\theta|\tau)$

$$f(x|\tau) = \int f(x|\theta) f(\theta|\tau) \, d\theta$$

we want to find a $\hat{\tau}$ from $f(x|\tau)$

2) physical method by Bayes:

$x_1, \dots, x_n | \theta \sim Bin(n; \theta)$ $\quad \theta \sim unif([0, 1])$

An example: θ: recombination Rate

3) Non-informative prior by Jeffrey and Laplace

Roughly, Bayes inference = MLE

Example: $x_1, \cdots, x_n | \mu \sim N(\mu, \sigma^2)$ $\quad \mu \sim N(\mu_0, \sigma_0^2)$

$$\mu | x_i \sim N\left\{\left(\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2} \cdot \bar{x}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right), \ * \right\}$$

if we set $\sigma_0^2 = +\infty$ $\quad E(\mu | x) = \bar{x}$

Jeffrey prior: $\pi(\theta) \propto \sqrt{I(\theta)}$

$$I(\theta) = E_x\left\{\frac{\partial^2}{\partial \theta^2}(\log f(x; \theta))\right\}$$

4) personal probability (Subjective)

- $\pi(\theta)$ reflects a person's judgement on θ

θ = Average heights of all U of S students

θ is unknow, can not be replicated.

θ is a R.V only because θ is unknow.

$\theta \sim N(170, 20^2)$

- $\pi(\theta)$ is information, external to data subjective to persons (not all we the same smart)

- inference results can be still judged with frequentist criterion

5) choose convenient prior, such as conditional conjuge prior.

6) hierarchical Modelling

$$X|\theta \sim f(x|\theta), \quad \theta_1, \cdots, \theta_n | \tau \sim f(\theta|\tau)$$

$$\tau \sim \pi(\tau)$$

Hierachical Modelling example:

$$x_i | \mu_i \sim N(\mu_i, 1) \quad \{ Y_{i1}, \cdots Y_{in} | \mu_i \sim N(\mu_i, \sigma_i^2)$$

$$x_i = \bar{Y}_i \}$$

$$\mu_1, \cdots \mu_n | \tau^2 \sim N(\theta, \tau^2)$$

$$\tau^2 \sim \text{inv-Gamma}(\alpha^*, \beta^*) \text{ OR}$$

$$f(\tau^2)\, d\tau^2 \propto (\tau^2)^{-(\alpha^*+1)} e^{-\frac{\beta^*}{\tau^2}} d\tau^2$$

$$\theta \sim \text{Unif}(-\infty, +\infty) \longrightarrow \text{Non-informative}$$

$$x \sim \text{Gamma}(\alpha^*, \beta^*)$$

$$f_x(x)\, dx \propto x^{\alpha^*-1} \cdot e^{-\beta^* \cdot x}\, dx$$

Let $\tau^2 = 1/x \sim$ inverse-Gamma (α^*, β^*)

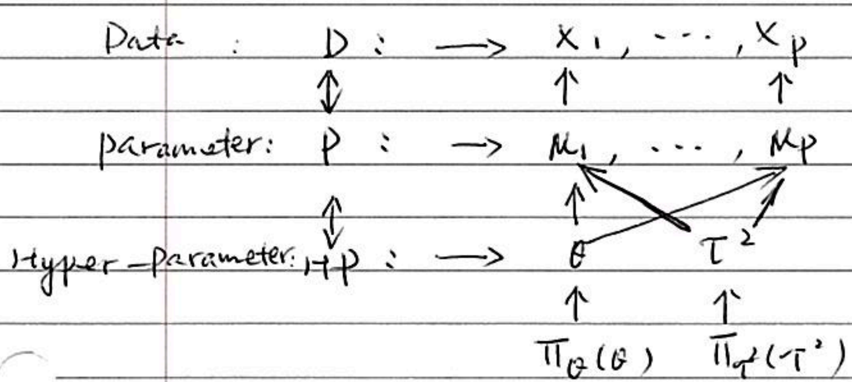
$$f_{\tau^2}(\tau^2)\, d\tau^2 \propto f_x\left(\frac{1}{\tau^2}\right) d\frac{1}{\tau^2}$$

$$\propto (\tau^2)^{-(\alpha^*-1)} e^{-\frac{\beta^*}{\tau^2}} (\tau^2)^{-2} d\tau^2$$

⇒ Example :

$$x_i \mid \mu_i \sim N(\mu_i, 1) \quad \text{for } i = 1, \cdots, p$$
$$\mu_i \mid \theta \sim N(\theta, \tau^2)$$
$$\theta \sim N(\theta_0, \sigma_0^2)$$
$$\tau^2 \sim \text{inv-Gamma}(\alpha^*, \beta^*)$$

$$f(\mu_1, \cdots, \mu_p, \theta, \tau^2) \propto \prod_{i=1}^{p} f(x_i \mid \mu_i) \cdot \prod_{i=1}^{p} f(\mu_i \mid \theta, \tau^2) \cdot \pi_\theta(\theta) \cdot \pi_{\tau^2}(\tau^2)$$



⇒ Full Conditionals

1) $\mu_i \mid x_i, \theta, \tau^2$

$$\begin{cases} x_i \mid \mu_i \sim N(\mu_i, 1) \\ \mu_i \mid \theta, \tau^2 \sim N(\theta, \tau^2) \end{cases} \Rightarrow$$

$$\mu_i \mid x_i \sim N\left(\frac{\frac{\theta}{\tau^2} + \frac{x_i}{1}}{\frac{1}{\tau^2} + \frac{1}{1}}, \frac{1}{\frac{1}{\tau^2} + 1} \right) = N\left(\theta + \tau_2^2(x_i - \theta) + \tau_2^2\right)$$

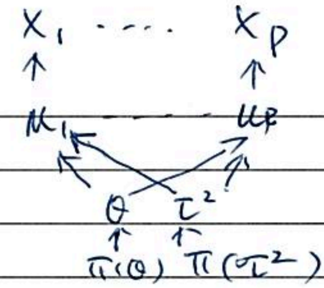
where, $\tau_2^2 = \frac{\tau^2}{1+\tau^2}$

2) $\theta \mid \tau^2, \mu_i, x_i$

$$\mu_1, \cdots, \mu_p \mid \theta, \tau^2 \sim N(\theta, \tau^2)$$
$$\theta \sim N(\theta_0, \sigma_0^2)$$

$$\theta \mid \tau^2, \mu_i, x_i \sim N\left(\frac{\frac{\theta_0}{\sigma_0^2} + \frac{p}{\tau^2} \cdot \bar{\mu}}{\frac{1}{\sigma_0^2} + \frac{p}{\tau^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{p}{\tau^2}} \right) = N\left(\bar{\mu}, \frac{\tau^2}{p}\right)$$

<< Feb. 09. 2015 >> STAT 846



⇒ Baseball example:

Let $X = (x_1, \cdots, x_p)$

$$f(\mu_1, \cdots \mu_p, \theta, \tau^2 \mid x) \propto \prod_{i=1}^{p} f(x_i \mid \mu_i) \cdot \prod_{i=1}^{p} f(\mu_i \mid \theta, \tau^2) \cdot \pi_0(\theta) \cdot \pi_{\tau^2}(\tau^2)$$

After we have samples:

$$\{ \mu_1^{(i)}, \cdots \mu_p^{(i)}, (\tau^2)^{(i)}, \theta^{(i)} \mid i = 1, 2, \cdots, N \}$$

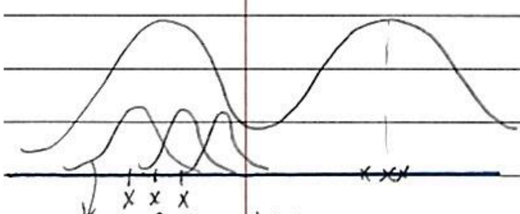
To look at $f(\mu_i \mid X)$ marginal posterior of μ_i

1) look at density $\hat{f}(\mu_j)$ of $\{ \mu_j^{(i)} \mid i = 1, 2, \cdots, N \}$

2) $f(\mu_j \mid X) = \int_N f(\mu_j \mid X, \theta\ \tau^2) f(\theta, \tau^2)\, d\theta\, d\tau^2$

(where $f(\mu_i \mid X, \theta, \tau^2) \sim N(\theta + \tau_2(X_i - \theta), \tau_2)$

$$\approx \sum_{i=1}^{N} f(\mu_j \mid X, \theta^{(i)}, (\tau^2)^{(i)})$$



Kernal Density

$\varphi\left(\frac{X_i - \bar{X}}{h}\right) \cdot \frac{1}{h}$

$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \varphi\left(\frac{X_i - \bar{X}}{h}\right) \cdot \frac{1}{h}$

This is an application of Rao-Blackwell formula

$$E\{t(\mu, \theta)\} = E_\theta\{ E_\mu(t(\mu, \theta) \mid \theta) \}$$

make a Note in here:

$$f(\mu_j \mid X, \theta, \tau^2) = \int f(\mu_j\, \mu_{-j} \mid X, \theta, \tau^2)\, d\mu$$

⇒ Empirical Bayes estimator of $f(\mu_j \mid X)$

Suppose, we have an estimator of θ, τ^2 by looking $f(x|\theta, \tau^2)$, denoted by $\hat{\theta}$, $\hat{\tau}_2 \left(=\frac{\hat{\tau}^2}{1+\hat{\tau}^2}\right)$

$$f(\mu_j|x) \approx f(\mu_j|x, \hat{\theta}, \hat{\tau}^2) = N(\hat{\theta} + \hat{\tau}_2(\hat{\theta} - x_i), \hat{\tau}_2)$$

in particular : J-S : $\hat{\theta} = \bar{x}$ $\quad \hat{\tau}_2 = 1 - \frac{p-3}{V}$

where, $V = \sum_{i=1}^{p} (x_i - \bar{x})^2$

$\Rightarrow$ predictive distribution (Another Method.)

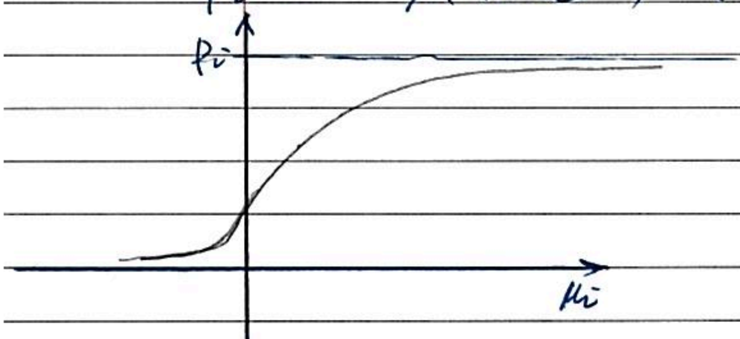
Example : $y_i | n_i \; p_i \sim Bin(n_i, p_i)$

pre-season

$z_i | N_i, p_i \sim Bin(N_i, p_i)$

full season

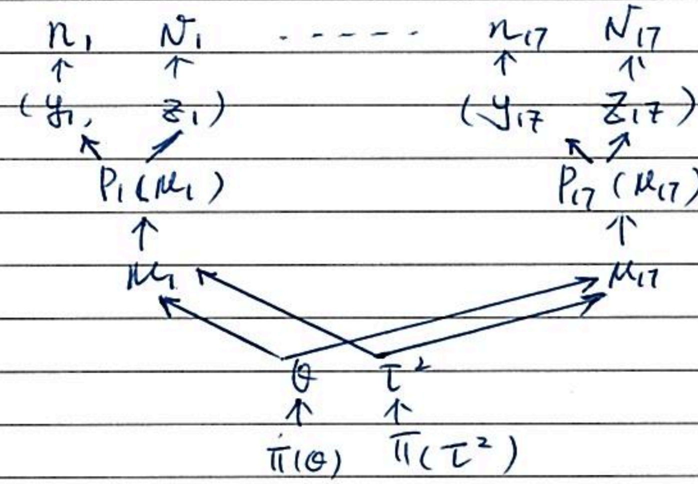
$p_i = e^{\mu_i}/(1+e^{\mu_i})$ (inv-logistic transformation)



$\mu_1, \cdots, \mu_p | \theta, \tau^2 \sim N(\theta, \tau^2)$

$\tau^2 \sim$ inv-Gamma (α^*, β^*)

$\theta \sim N(0, \sigma_0^2)$



Posterior of μ_i, θ, τ^2

$$f(\mu_1, \cdots \mu_{17}, \theta, \tau^2) \propto \prod_{i=1}^{17} f(y_i|\mu_i) \cdot \prod_{i=1}^{17} f(\mu_i|\theta, \tau^2) \cdot \pi(\theta), \pi(\tau^2)$$

$$f(y_i \mid \mu_i) = \binom{n_i}{y_i} \left(\frac{e^{\mu_i}}{1+e^{\mu_i}}\right)^{y_i} \cdot \left(1 - \frac{e^{\mu_i}}{1+e^{\mu_i}}\right)^{n_i - y_i}$$

We don't have close form for $f(\mu_i \mid y_i, \theta, \tau^2)$

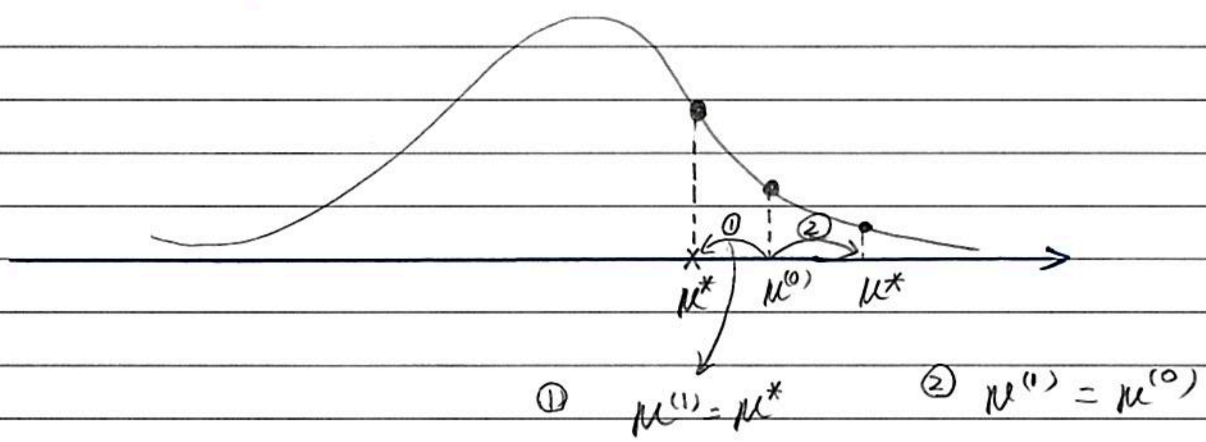
$\Rightarrow$ Gibbs sampling

$$\begin{cases} 1)\ \mu_i \mid y_i, \theta, \tau^2 \\ 2)\ \theta \mid \mu_1, \cdots \mu_{17}, \tau^2, y_1, \cdots y_{17} \\ 3)\ \tau^2 \mid \theta, \mu_1, \cdots \mu_{17}, y_1, \cdots y_{17} \end{cases}$$

$\Rightarrow$ Metropolis – Hasting

$\mu \sim \pi(\mu)$, repeat N times

$$\begin{cases} \text{Starting from } \mu^{(0)} \\ 1)\ \text{propose } \mu^* \text{ from } P \\ 2)\ \text{Draw } u \sim \text{unif}([0,1]) \\ 3)\ \text{if } u < \min\left\{1, \ \pi(\mu^*) P(\mu^* \mid \mu^{(0)}) / \pi(\mu^{(0)}) \pi(\mu^{(0)} \mid \mu^*)\right. \\ \quad \text{Accept. } \mu^{(1)} = \mu^* \\ \quad \text{o/w } \ \mu^{(1)} = \mu^{(0)} \end{cases}$$



$$= (\tau^2)^{-(\alpha^*+1)} e^{-\frac{\beta^*}{\tau^2}} d\tau^2$$

Textbook : $\tau = \sqrt{\tau^2}$ $\quad f_\tau(\tau)\, d\tau = f_{\tau^2}(\tau^2)\, d\tau^2$

$$= f_{\tau^2}(\tau^2) \cdot 2\tau\, d\tau = (\tau^2)^{-(\alpha^*+1)} \tau \cdot e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$$= (\tau^2)^{-(\alpha^*+1)} \cdot (\tau^2)^{\frac{1}{2}} e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$$= (\tau^2)^{-\alpha^*-\frac{1}{2}} e^{-\frac{\beta^*}{\tau^2}} d\tau = (\tau^2)^{-(\alpha^*+\frac{1}{2})} e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$\Rightarrow$ Joint posterior of μ, θ, τ^2

$$f(\mu_1, \cdots, \mu_n, \theta, \tau^2 \mid x_1, x_2, \cdots, x_n)$$

$$\propto \prod_{i=1}^{n} f(x_i \mid \mu_i) \cdot \prod_{i=1} f(\mu_i \mid \theta, \tau^2)$$

$$\cancel{(\pi_\theta(\theta) \cdot \pi_{\tau^2}(\tau^2) \times d\theta\, d\mu_i\, d\tau^2)}$$

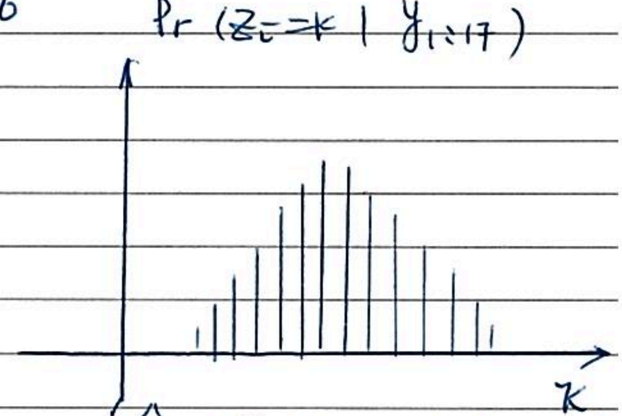
$$\propto e^{-\sum_{i=1}^{n} \frac{(x_i-\mu_i)^2}{2}} \times (\tau^2)^{-\frac{n}{2}} \cdot e^{-\sum_{i=1}^{n} \frac{(\mu_i-\theta)^2}{2\tau^2}}$$

$$(\tau^2)^{-(\alpha^*+1)} \cdot e^{-\frac{\beta^*}{\tau^2}} d\theta\, d\mu_i\, d\tau^2$$

We will derive full conditionals :

$$\begin{cases} f(\mu_i \mid \mu_{-i};\ \theta, \tau^2, x_1, \cdots, x_n) \\ f(\theta \mid \mu_{1:n}, \tau^2, x_1, \cdots, x_n) \\ f(\tau^2 \mid \mu_{1:n}, \theta, x_1, \cdots, x_n) \end{cases}$$

<< Feb. 11, 2015 >> STAT 846



$\hat{Z}_i = E(Z_i = k \mid y_{1:17})$

Example:

$$\begin{tikzcd}
n_1 & N_1 & \cdots & n_{17} & N_{17} \\
(y_1, \arrow[u] & z_1) \arrow[u] & \cdots & (y_{17} \arrow[u] & z_{17}) \arrow[u] \\
p_1 \arrow[u] \arrow[ur] & & \cdots & p_{17} \arrow[u] \arrow[ur] & \\
\mu_1 \arrow[u] & & \cdots & \mu_{17} \arrow[u] &
\end{tikzcd}$$

$y_i \mid \mu_i \sim Bin(n_i, p_i)$

$z_i \mid \mu_i \sim Bin(N_i, p_i)$

n_i, N_i are co-variates

$\mu_1, \cdots, \mu_{17} \mid \theta, \tau^2 \sim N(\theta, \tau^2)$

$\theta \sim \pi_\theta$, $\tau^2 \sim \pi_{\tau^2}$

We want to find $\hat{Pr}(z_i = k \mid y_1, \cdots y_{17})$ for $i = 1, \cdots 17$,

prediction PMF of $z_i \mid y_{1:17}$,

=> in general term:

$$\begin{tikzcd}
\text{observe: } y & & \text{future } y^* \\
& \theta \arrow[ul, "f(y \mid \theta)"] \arrow[ur, "f(y^* \mid \theta, y)"'] &
\end{tikzcd}$$

=> Joint Model for (θ, y, y^*):

$$f(\theta, y, y^*) = \pi(\theta) f(y \mid \theta) \cdot f(y^* \mid \theta, y)$$

"Our Goal to find the $f(y^* \mid y)$"

=> Joint posterior of $\theta, y^* \mid y$

$$f(\theta, y^* \mid y) = \frac{f(\theta, y, y^*)}{\iint f(\theta, y, y^*)\, dy^* d\theta}$$

$$= \frac{\pi(\theta)\, f(y|\theta)\, f(y^*|\theta, y)}{\left(\int f(y^*|y,\theta)\, dy^*\right)\left(\int \pi(\theta) f(y|\theta)\, d\theta\right)}$$

(Note: $\int f(y^*|y,\theta)\, dy^* = 1$)

$$= \frac{\pi(\theta) \cdot f(y|\theta)}{\int \pi(\theta) f(y|\theta)\, d\theta} \cdot f(y^*|\theta, y)$$

$$= f(\theta|y) \cdot f(y^*|\theta, y)$$

(Note that: if omitting $|y$, $f(\theta, y^*) = f(\theta) \cdot f(y^*|\theta)$)

$$\Rightarrow f(y^*|y) = \int f(\theta, y^*|y)\, d\theta$$

$$= \int f(\theta|y) \cdot f(y^*|\theta, y)\, d\theta$$

$$= E_{\theta|y}\{f(y^*|\theta, y)\}$$

$\Rightarrow$ Two approaches to finding $f(y^*|y)$

Method 1:

$$f(y^*=k|y) = E_{\theta|y}\{f(y^*=k|\theta, y)\}$$

Given samples of $\theta^{(i)} \sim f(\theta|y)$

$$\hat{f}(y^*=k|y) = \sum_{i=1}^{N} f(y^*=k|\theta^{(i)}, y) / N$$

Method 2: $f(y^*|y)$ is marginal of $f(\theta, y^*|y)$

for discrete y^*.

$$f(y^*=k|y) = E_{y^*,\theta|y}\{I(y^*=k)\}$$

Given samples of $(y^{*,(i)}, \theta^{(i)}) \sim f(\theta, y^* \mid y)$

$$\hat{f}(y^* = K \mid y) = \sum_{i=1}^{N} I(y^{*,i} = K) \,/\, N$$

If y^* is continuous density, apply kernal estimate to $y^{*,1}, \ldots, y^{*,N}$

$\Rightarrow$ Re-Mark:

Method 1 : is Rao-Blackwellization of Method 2.

$$E_X(t(X)) = E_Y(\underbrace{E_X(t(X) \mid Y)}_{\tilde{t}(Y)}) = E_Y(\tilde{t}(Y))$$

$\Rightarrow$ Note that $Var(t(X)) > Var(\tilde{t}(Y))$

$\Rightarrow$ Back to example:

Method 1 : Draw $\theta^{(i)} \sim f(\theta \mid y_1, \cdots, y_{17})$

OR. Draw $\theta^{(i)}, Z_j^{(i)} \sim f(\theta, Z_j \mid y_{1:17})$

then discard $Z_j^{(i)}$

$$\hat{f}(Z_j = K \mid y_{1:17}) \sim \frac{\sum_{i=1}^{N} dbino(K; N_j, P_j^{(i)})}{N}$$

Method 2 : Draw $\theta^{(i)}, Z_j^{(i)} \sim f(\theta, Z_j \mid y_{1:17})$

using $Z_j^{(1)} \cdots Z_j^{(N)}$ only $(Z_j^{(i)} \sim f(Z_j \mid y_{1:17}))$

$$\hat{f}(z_j = k \mid y_{1:17}) = \frac{\sum_{i=1}^{N} I(z_j^{(i)} = k)}{N}$$

$\Rightarrow$ Method 1: $Var(\bar{X}) = \dfrac{Var(x_i)}{N}$

Method 2: $Var(\bar{Y}) = \dfrac{Var(y_i)}{N}$