

22 January 21, 2015 \Rightarrow STAT 846

\Rightarrow Admissibility, Bayes Rules

\Rightarrow Theorem 2.3 : Assume that $\Theta = \{\theta_1, \dots, \theta_t\}$ is finite, and that the prior $\pi(\cdot)$ gives positive probability to each θ_i . Then a Bayes rule with respect to $\pi(\cdot)$ is admissible.

\Rightarrow Theorem 2.4 : if a Bayes rule is unique, it is admissible.

\Rightarrow theorem 2.5

(1) Let Θ be a subset of the real line;

(2) Assume that the risk functions $R(\theta, d)$

are continuous in θ for all decision rules d ;

(3) Suppose that for any $\varepsilon > 0$ and any θ the interval $(\theta - \varepsilon, \theta + \varepsilon)$ has positive probability under the prior $\pi(\cdot)$.

Then, a Bayes rule with respect to $\pi(\cdot)$ is admissible.

Chapter 3 : Bayesian methods

3.1 Fundamental elements of Bayesian inference

Bayes theorem (Law)

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

then the posterior density of θ , given x

$$\pi(\theta|x) = \pi(\theta) \cdot f(x; \theta) / \int_0^1 \pi(\theta) \cdot f(x; \theta) d\theta$$

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

\Rightarrow examples :

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

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

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

\Rightarrow A note:

$$\pi(\theta|x) = \frac{\pi(\theta) f(x; \theta)}{\int_0^1 \pi(\theta) f(x; \theta) d\theta} = \frac{C \pi(\theta) \cdot f(x; \theta)}{\int C \pi(\theta) \cdot f(x; \theta) d\theta}$$

C is free of θ

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

$$\begin{aligned} \text{example 1: } \pi(\theta) \cdot f(x; \theta) &\propto \theta^{a-1} \cdot (1-\theta)^{b-1} \cdot \theta^x \cdot (1-\theta)^{n-x} \\ &= \theta^{a+x-1} \cdot (1-\theta)^{n+b-x-1} \end{aligned}$$

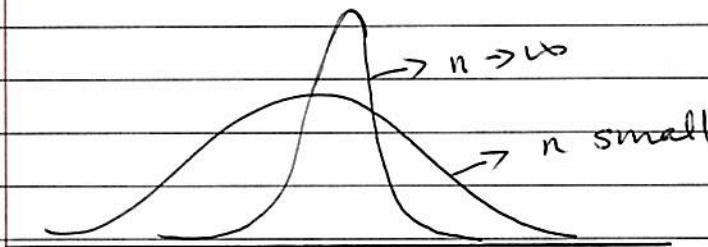
So, we can conclude that :

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

$$E(\theta | x) = \frac{a+x}{a+b+n} \quad \text{as } n \rightarrow \infty \quad E(\theta | x) \approx \frac{x}{n}$$

$$\text{Var}(\theta | x) = \frac{(a+x)(n+b-x)}{(a+b+n)^2(a+b+n+1)} \approx \frac{x(n-x)}{n^3}$$

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



example 2 :

$$x_1, x_2, \dots, x_n | \mu \quad x_i \sim N(\mu, \sigma^2)$$

assume σ^2 is known, $\mu \sim N(\mu_0, \sigma_0^2)$

To find $\pi(\mu | x_1, \dots, x_n)$ $\underline{x} = (x_1, \dots, x_n)$

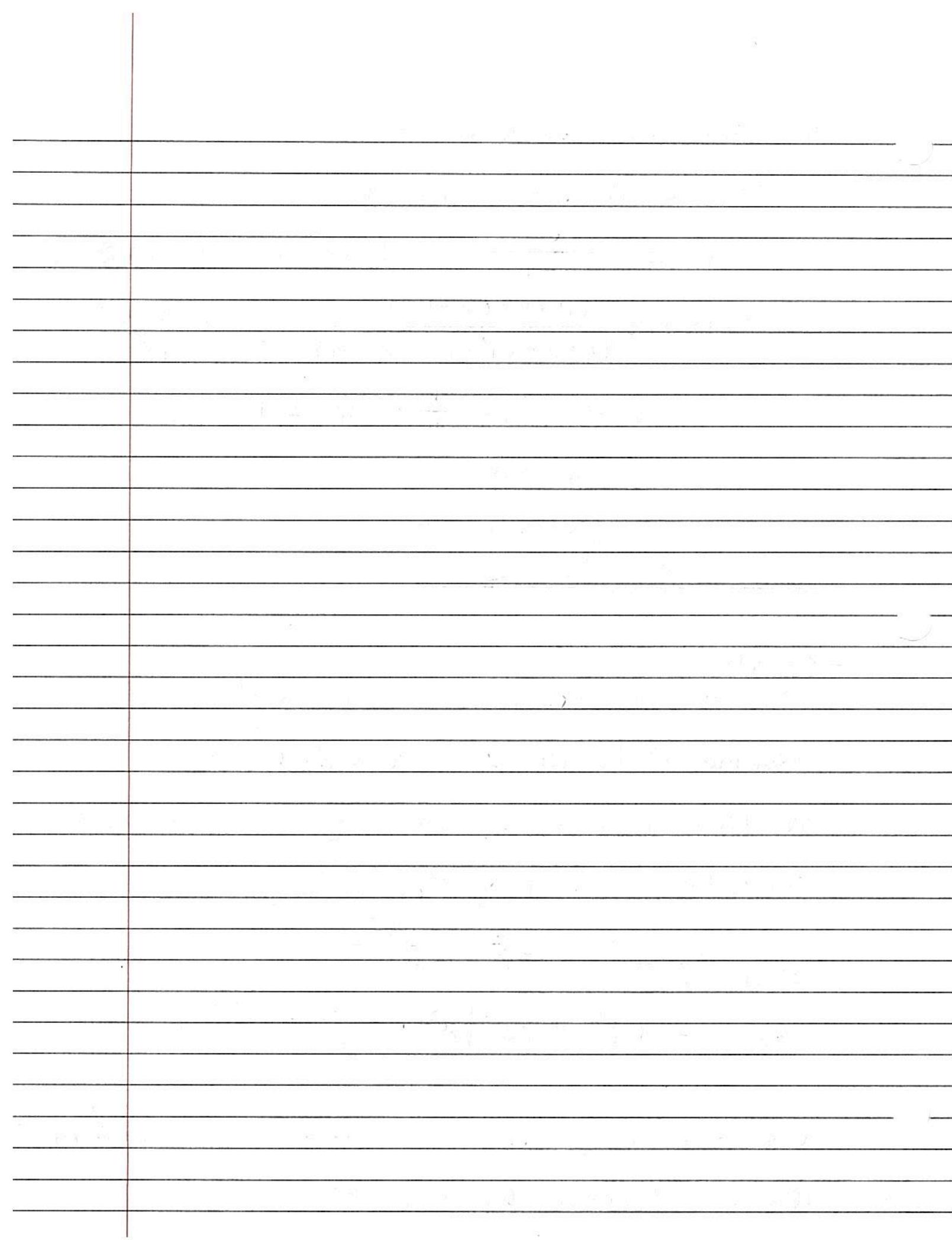
$$\pi(\mu | x_1, \dots, x_n) \propto \pi(\mu) \cdot f(\underline{x}; \theta)$$

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

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

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

We will find μ_1 and σ_1^2



$$\pi(\mu | x) \propto e^{-\frac{(\mu - \mu_1)^2}{2\sigma^2}} \propto e^{-\left(\frac{1}{2\sigma^2}\mu^2 - \frac{\mu_1}{\sigma^2}\mu\right)}$$

$$\text{So, } \frac{1}{2\sigma^2} = \frac{1}{2\sigma_0^2} + \frac{n}{2\sigma^2} \iff$$

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

$$\text{Let } \sigma_0 = \frac{1}{\sigma_0^2}; \sigma_1 = \frac{1}{\sigma_1^2}; \sigma = \frac{1}{\sigma^2}$$

$$\Rightarrow \sigma_1 = \sigma_0 + n \cdot \sigma$$

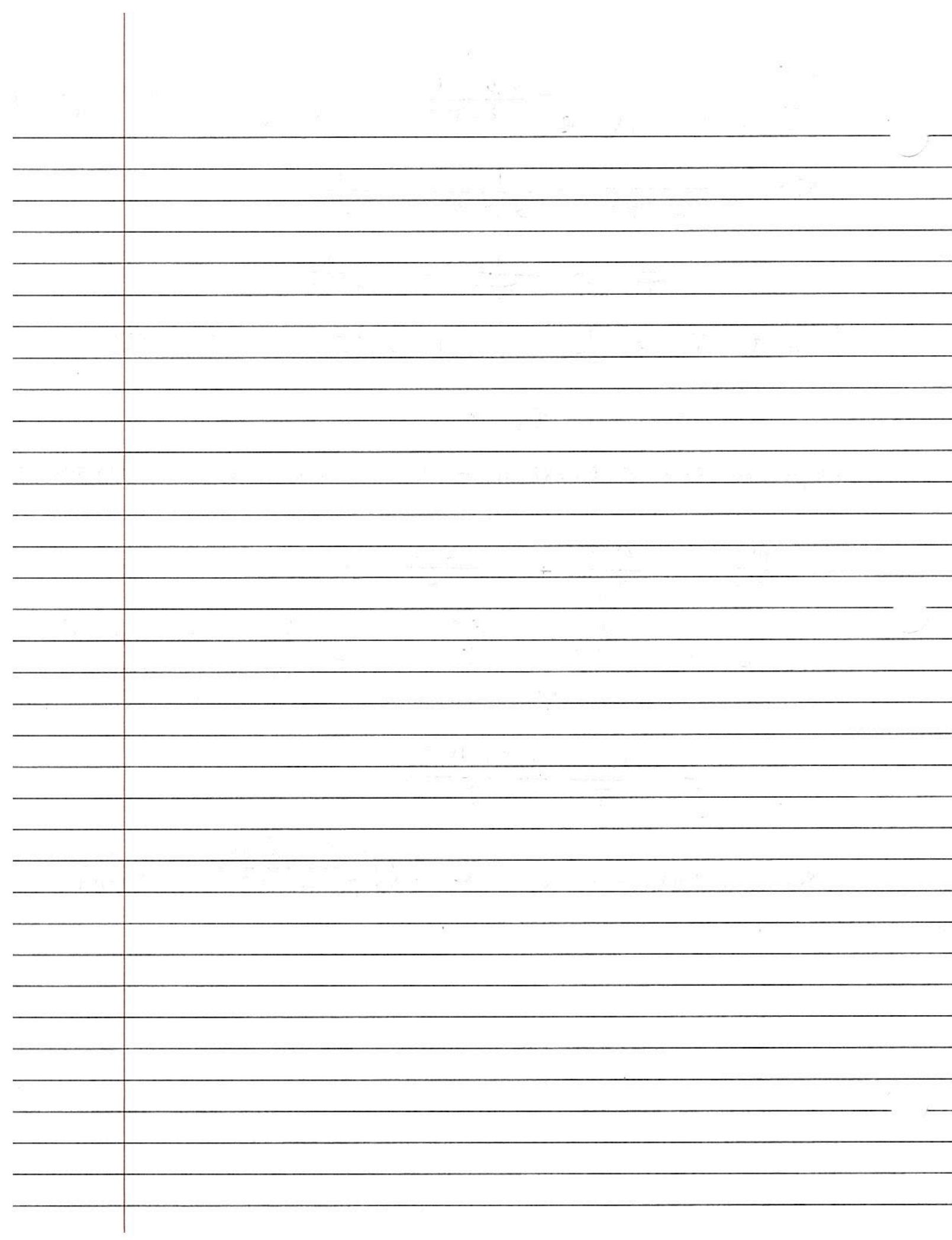
\Leftrightarrow Posterior precision = prior precision + $\sum_{i=1}^n$ Precision of x_i

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

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

$$= \frac{\mu_0 \sigma_0^2 + \bar{x} \cdot n \sigma^2}{\sigma_0^2 + n \sigma^2}$$

$$\text{So, } \mu | x_1, \dots, x_n \sim N\left(\frac{\mu_0 \sigma_0^2 + \bar{x} \cdot n \sigma^2}{\sigma_0^2 + n \sigma^2}, \frac{1}{\sigma_0^2 + n \sigma^2}\right)$$



<< January 23, 2015 >> STAT 846

Examples:

Parameter	$\theta \sim$	1	2	3
(prior Dis. $\pi(\theta)$)	$\pi(\theta)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$X \theta \sim$	$\theta = 1$	$\theta = 0$	$\theta = 1$
Data Distribution $\pi(x \theta)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

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

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

$\theta \backslash X$	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	$\frac{1}{3} \times 0$	$\frac{1}{3} \times 0$	$\frac{1}{3} \times \frac{1}{4}$	$\frac{1}{3} \times \frac{1}{2}$	$\frac{1}{3} \times \frac{1}{4}$

$\pi(\theta x=0) \sim$	θ	1	2	3
$\pi(\theta) \cdot \pi(x \theta)$	$\pi(\theta)$	1	0	0

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

$\pi(\theta x=1) \sim$	θ	1	2	3
$\pi(\theta)$		$\frac{2}{3}$	$\frac{1}{3}$	0

$\pi(\theta x=2) \sim$	θ	1	2	3
$\pi(\theta)$		$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

$$2) X_1, \dots, X_n | (\mu, \sigma^2) \sim N(\mu, \frac{1}{\sigma^2})$$

$\sigma^2 \sim \text{Gamma}(\alpha, \beta)$

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{\alpha-1} e^{-\beta/\sigma^2}$$

$$\mu | \sigma^2 \sim N(\nu, \frac{1}{K\sigma^2})$$

$$\pi(\mu | \sigma^2) = \frac{1}{\sqrt{2\pi}} (K\sigma^2)^{\frac{1}{2}} e^{-\frac{K\sigma^2}{2}(\mu - \nu)^2}$$

$$\pi(\sigma^2, \mu) \propto \sigma^{\alpha-1} e^{-\sigma^2(\beta + \frac{K}{2}(\mu - \nu)^2)}$$

$$\pi(\sigma^2, \mu | X_1, X_2, \dots, X_n) \propto \pi(\sigma^2, \mu) \cdot f(X_1, \dots, X_n; \sigma^2, \mu)$$

$$\propto \sigma^{\alpha-1} e^{-\sigma^2(\beta + \frac{K}{2}(\mu - \nu)^2)} \prod_{i=1}^n \sigma^{\frac{1}{2}} e^{-\frac{\sigma^2}{2}(X_i - \mu)^2}$$

$$\propto \sigma^{\alpha+1-\frac{1}{2}} \cdot e^{-\sigma^2(\beta' + \frac{K'}{2}(\mu - \nu')^2)}$$

$$\text{where, } \alpha' = \alpha + \frac{n}{2}$$

$$\beta' = \beta + \frac{1}{2} \cdot \frac{nK}{n+K} (\bar{X} - \nu)^2 + \frac{1}{2} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$K' = K + n$$

$$\nu' = \frac{K\nu + n\bar{X}}{K+n}$$

\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.

\Rightarrow General form of Bayes rule, give a fixed prior.

\Rightarrow Risk function:

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

Bayes Risk:

$$r(\pi, d(x)) = \int_{\Theta} \bar{\pi}(\theta) \cdot R(\theta, d(x)) d\theta$$

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

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

where, $f(x) = \int_{\Theta} \bar{\pi}(\theta) \cdot f(x, \theta) d\theta$

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

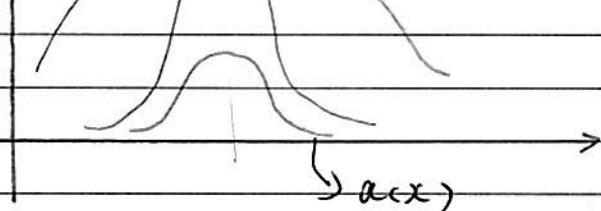
To find Bayes Rule $d(x)$ to minimize $r(\bar{\pi}, d(x))$ is equivalent to that for each x , we find $d(x)$

to minimize $\underbrace{\int_{\Theta} L(\theta, d(x)) \bar{\pi}(\theta | x) d\theta}_{a(x)} = E_{\theta|x} \{ L(\theta, d(x)) \}$

Called expected posterior loss

$$\rightarrow a_2(x) = d_2(x) f(x)$$

$$\rightarrow a_1(x) = d_1(x) f(x)$$



1. *Chlorophytum comosum* L. (Liliaceae)

2. *Clivia miniata* (L.) Ker-Gawler (Amaryllidaceae)

3. *Crinum asiaticum* L. (Amaryllidaceae)

4. *Crinum asiaticum* L.

5. *Crinum asiaticum* L.

6. *Crinum asiaticum* L.

7. *Crinum asiaticum* L.

8. *Crinum asiaticum* L.

9. *Crinum asiaticum* L.

10. *Crinum asiaticum* L.

11. *Crinum asiaticum* L.

12. *Crinum asiaticum* L.

\Rightarrow Now to find the Bayes Rule:

examples:

1) Point estimate:

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

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

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

$$\text{Solving } \frac{\partial}{\partial d}(A(d(x))) = E_{\theta|x}\left(\frac{\partial}{\partial d}(\theta - d)^2\right) = E_{\theta|x}(2(\theta - d)) \\ = 0$$

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

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

2) Point estimate:

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

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

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

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

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

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

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

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

the old school with a lot of fun

and a great time

at the beach

swimming

and a lot of fun at the beach

and a great time at the beach

and a great time at the beach

and a

great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

and a great time at the beach

« January 26, 2015 » STAT 846

⇒ Examples of Bayes Rules

1) Point estimates:

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

$$\begin{aligned} E_{\theta|x} (|\theta - d(x)|) &= \int_{-\infty}^{\infty} |\theta - d(x)| \pi(\theta|x) d\theta \\ &= r(d(x)|x) \end{aligned}$$

$$\begin{aligned} \frac{\partial r(d(x)|x)}{\partial d(x)} &= \int_{-\infty}^{\infty} [1x I(\theta < d) + (-1) I(\theta > d)] \pi(\theta|x) d\theta \\ &= \int_{-\infty}^d \pi(\theta|x) d\theta - \int_d^{\infty} \pi(\theta|x) d\theta = 0 \end{aligned}$$

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

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

2) In hypothesis testing

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

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

100% of the time

comes from the bottom of

the sand

the sand is very soft

it is soft

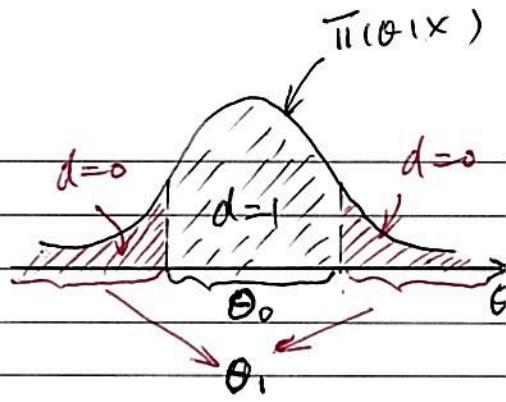
and it is very soft

so it is very soft

so it is very soft

the sand is very soft

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



Bayes Rule: $d(x)$:

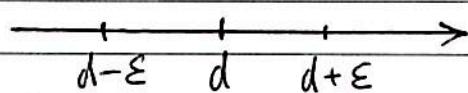
$$d(x) = \begin{cases} 1 & \text{if } \Pr(\theta \in \Theta_1 | x) \geq \Pr(\theta \in \Theta_0 | x) \\ 0 & \text{if } \Pr(\theta \in \Theta_0 | x) \geq \Pr(\theta \in \Theta_1 | x) \end{cases}$$

We will come back in chapter 4.

\Rightarrow Lindley paradox

3) Interval estimate

$$\mathcal{A}_0 = \{d\} = \theta$$



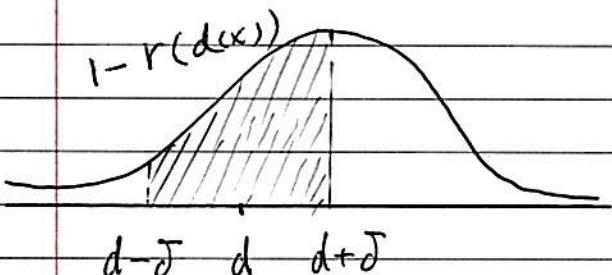
$$\mathcal{A}_1 = \{(d-\delta, d+\delta) \mid d \in \theta\}$$

Prescribe the length of interval δ

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

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

the Bayes Rule $d(x)$ is



$$d(x) = \operatorname{argmin}_d \Pr(d-\delta < \theta < d+\delta | x)$$

= highest Posterior Density

= HPD

1. What is the difference between a primary and secondary market?

2. What is the difference between a primary and secondary market?

3. What is the difference between a primary and secondary market?

4. What is the difference between a primary and secondary market?

5. What is the difference between a primary and secondary market?

6. What is the difference between a primary and secondary market?

7. What is the difference between a primary and secondary market?

8. What is the difference between a primary and secondary market?

9. What is the difference between a primary and secondary market?

10. What is the difference between a primary and secondary market?

11. What is the difference between a primary and secondary market?

12. What is the difference between a primary and secondary market?

13. What is the difference between a primary and secondary market?

14. What is the difference between a primary and secondary market?

15. What is the difference between a primary and secondary market?

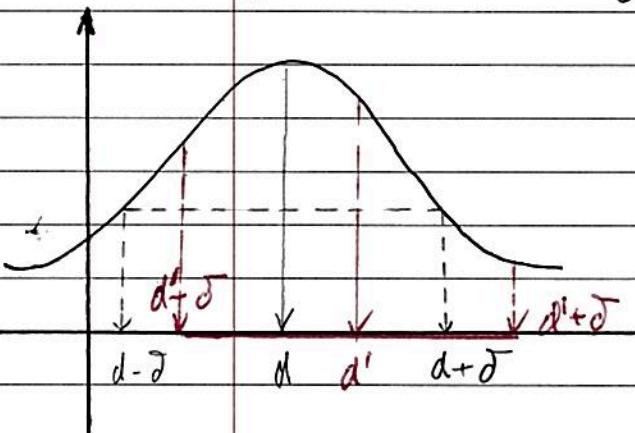
16. What is the difference between a primary and secondary market?

17. What is the difference between a primary and secondary market?

\Rightarrow A special case

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

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



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

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

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

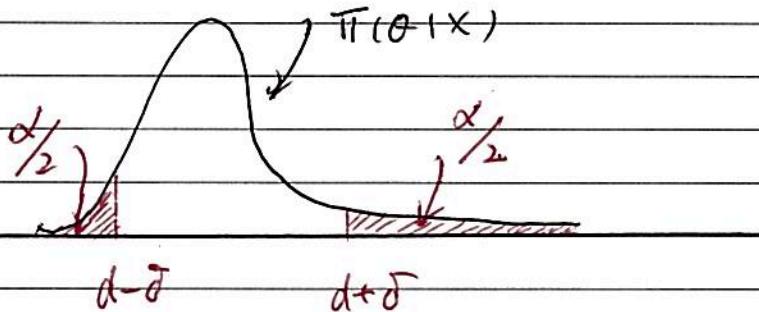
then we search I such that $\text{length}(I)$ is minimized.

\Rightarrow In practice, two alternatives :

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

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

Equal tailed interval



and writing to it

is a good way to learn grammar

and writing to it

is a good way to learn grammar

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

where, $\mu = \bar{E}(\theta|x)$ $\sigma^2 = \text{Var}(\theta|x)$

Based on Bayesian Asymptotic

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

\Rightarrow point estimate: how to find the minimax Rule

$$x| \theta \sim \text{Bin.}(n, \theta)$$

$$\theta \sim \text{Beta}(a, b)$$

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

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

the Bayes Rule is

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

If $R(\theta, d(x)) = \text{constant}$
 for all $\theta \in (0, 1)$
 By theorem 2.2, this
 calculated the risk function $d(x)$ is minimax.

$$\text{Let } c = a+b+n$$

$$R(\theta, d) = \bar{E}_x \left\{ (\theta - d(x))^2 \right\} = \bar{E}_x \left\{ \left(\theta - \frac{a+x}{a+b+n} \right)^2 \right\}$$

$$= \bar{E}_x \left\{ \left(\theta - \frac{a+x}{c} \right)^2 \right\} = \frac{1}{c^2} \bar{E}_x \left\{ (c\theta - a - x)^2 \right\}$$

$$= \frac{1}{c^2} \bar{E}_x \left\{ (c\theta - a)^2 - 2(c\theta - a)x + x^2 \right\}$$

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

1. *S. sordidus* (Gmelin) - Schlegel's Kingbird

2. *S. albocristatus* (Gmelin) - Yellow-bellied Kingbird

3. *S. superbus* (Gmelin) - Great-tailed Kingbird

4. *S. m. albiventris* (Gmelin) - Western Kingbird

5. *S. m. mitchelli* (Gmelin) - Mountain Kingbird

6. *S. m. albocristatus* (Gmelin)

7. *S. m. albocristatus* (Gmelin)

8. *S. m. albocristatus* (Gmelin)

9. *S. m. albocristatus* (Gmelin)

10. *S. m. albocristatus* (Gmelin)

11. *S. m. albocristatus* (Gmelin)

12. *S. m. albocristatus* (Gmelin)

13. *S. m. albocristatus* (Gmelin)

14. *S. m. albocristatus* (Gmelin)

15. *S. m. albocristatus* (Gmelin)

16. *S. m. albocristatus* (Gmelin)

17. *S. m. albocristatus* (Gmelin)

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

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

We write that

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

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

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

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

1. In the first place, we must consider the nature of the
2. We must also consider the nature of the
3. We must also consider the nature of the
4. We must also consider the nature of the
5. We must also consider the nature of the
6. We must also consider the nature of the
7. We must also consider the nature of the
8. We must also consider the nature of the
9. We must also consider the nature of the
10. We must also consider the nature of the

<< January 28, 2015 >> STAT 846

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

$$P \sim \text{Beta}(a, b)$$

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

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

\Rightarrow Shrinkage and James-Stein estimation.

$$X_i | \mu_i \sim N(\mu_i, 1) \text{ for } i=1, 2, \dots, p$$

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

A decision rule $d = (d_1, d_2, \dots, d_p)$

Loss function:

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

$$\text{Let } x = (x_1, x_2, \dots, x_p)^T, d = (d_1, d_2, \dots, d_p)^T$$

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

$$\Rightarrow \text{A special case } a=0 \Rightarrow d^0(x) = x$$

$$\Rightarrow \text{When } a>0, d^a(x) \text{ closer to } (0, 0, \dots, 0)^T$$

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

d^0 .

$$\text{the risk function: } R(\mu, d^0(x)) = \bar{E}_x \{ \|x - \mu\|^2 \}$$

1. *Chlorophytum comosum* (L.) Willd.
2. *Chlorophytum comosum* (L.) Willd.
3. *Chlorophytum comosum* (L.) Willd.
4. *Chlorophytum comosum* (L.) Willd.
5. *Chlorophytum comosum* (L.) Willd.
6. *Chlorophytum comosum* (L.) Willd.
7. *Chlorophytum comosum* (L.) Willd.
8. *Chlorophytum comosum* (L.) Willd.
9. *Chlorophytum comosum* (L.) Willd.
10. *Chlorophytum comosum* (L.) Willd.
11. *Chlorophytum comosum* (L.) Willd.
12. *Chlorophytum comosum* (L.) Willd.
13. *Chlorophytum comosum* (L.) Willd.
14. *Chlorophytum comosum* (L.) Willd.
15. *Chlorophytum comosum* (L.) Willd.
16. *Chlorophytum comosum* (L.) Willd.
17. *Chlorophytum comosum* (L.) Willd.
18. *Chlorophytum comosum* (L.) Willd.
19. *Chlorophytum comosum* (L.) Willd.
20. *Chlorophytum comosum* (L.) Willd.

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

\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 $N(0, 1)$

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

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

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

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

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

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

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

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

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

$$\Rightarrow \text{Risk function } R(\mu, d^a) = E_X (||\mu - d^a(x)||^2)$$

$$= E_X \left\{ \left(||\mu - \left(1 - \frac{\alpha}{||x||^2} \right) \cdot X|| \right)^2 \right\}$$

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

1. *Leucosia* *leucostoma* (L.)

2. *Leucosia* *leucostoma* (L.)

3. *Leucosia* *leucostoma* (L.)

4. *Leucosia* *leucostoma* (L.)

5. *Leucosia* *leucostoma* (L.)

6. *Leucosia* *leucostoma* (L.)

7. *Leucosia* *leucostoma* (L.)

8. *Leucosia* *leucostoma* (L.)

9. *Leucosia* *leucostoma* (L.)

10. *Leucosia* *leucostoma* (L.)

11. *Leucosia* *leucostoma* (L.)

12. *Leucosia* *leucostoma* (L.)

$$= \bar{E}_X \left\{ \| \mu - x \|^2 \right\} - 2a \cdot \bar{E}_X \left\{ \frac{x^T(\mu - \mu)}{\| x \|} \right\} + a^2 \cdot \bar{E}_X \left\{ \left(\frac{1}{\| x \|^2} \right) \right\}$$

where $\bar{E}_X \left(\frac{x^T(\mu - \mu)}{\| x \|^2} \right)$

$$= \sum_{i=1}^P \bar{E}_X \left(\frac{x_i (\bar{x}_i - \mu_i)}{\sum_{j=1}^P x_j^2} \right) . h(x_i) = \frac{x_i}{\sum_{j=1}^P x_j^2}$$

$$= \sum_{i=1}^P \bar{E}_X \left\{ \frac{1}{\partial x_i} \cdot \left(\frac{\bar{x}_i}{\sum_{j=1}^P x_j^2} \right) \cdot \frac{\sum_{j=1}^P x_j^2 - 2x_i^2}{\left(\sum_{j=1}^P x_j^2 \right)^2} \right\}$$

$$= \left. \begin{aligned} & (P-2) \bar{E}_X \left(\frac{1}{\| x \|^2} \right) \end{aligned} \right\} \text{Note(1)}$$

$$= P - 2a \cdot (P-2) \bar{E}_X \left(\frac{1}{\| x \|^2} \right) + a^2 \bar{E}_X \left(\frac{1}{\| x \|^2} \right)$$

$$= P - (2a(P-2) + a^2) \bar{E}_X \left(\frac{1}{\| x \|^2} \right)$$

$$\text{so, } R(\mu, d^a) < R(\mu, d^o) = P$$

$$\Leftrightarrow 2a(P-2) - a^2 > 0$$

$$\Leftrightarrow 2(P-2) > a$$

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

$$a < 2(P-2)$$

\Leftrightarrow therefore d^o is inadmissible

Note 1:

$$\sum_{i=1}^P \mathbb{E}_x \left\{ \frac{\sum_{j=1}^P x_j^2 - 2x_i^2}{\left(\sum_{j=1}^P x_j^2 \right)^2} \right\}$$

$$= \mathbb{E}_x \left\{ \sum_{i=1}^P \frac{\|x\|^2 - 2x_i^2}{\|x\|^4} \right\}$$

$$= \mathbb{E}_x \left\{ \frac{1}{\|x\|^2} \cdot \sum_{i=1}^P 1 - \frac{2}{\|x\|^4} \cdot \sum_{i=1}^P x_i^2 \right\}$$

$$= \mathbb{E}_x \left(\frac{P}{\|x\|^2} - \frac{2\|x\|^2}{\|x\|^4} \right)$$

$$= \mathbb{E}_x \left(\frac{P-2}{\|x\|^2} \right)$$

$$= (P-2) \mathbb{E}_x \left(\frac{1}{\|x\|^2} \right)$$

<<January 30, 2015 >> STAT 846

=> Home Run Examples

pre-season Regular Season

	y_i	n_i	x_i	$N_i(AB)$	$\hat{\mu}_i^{JS}$	\bar{n}_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: $\hat{HR}_i = \left(\frac{y_i}{n_i}\right) \times N_i$

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

$y_i | n_i, p_i \sim \text{Bin}(n_i, p_i)$

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

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

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

We will then apply JS to x_1, \dots, x_p to estimate μ_1, \dots, μ_p , then we can estimate

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

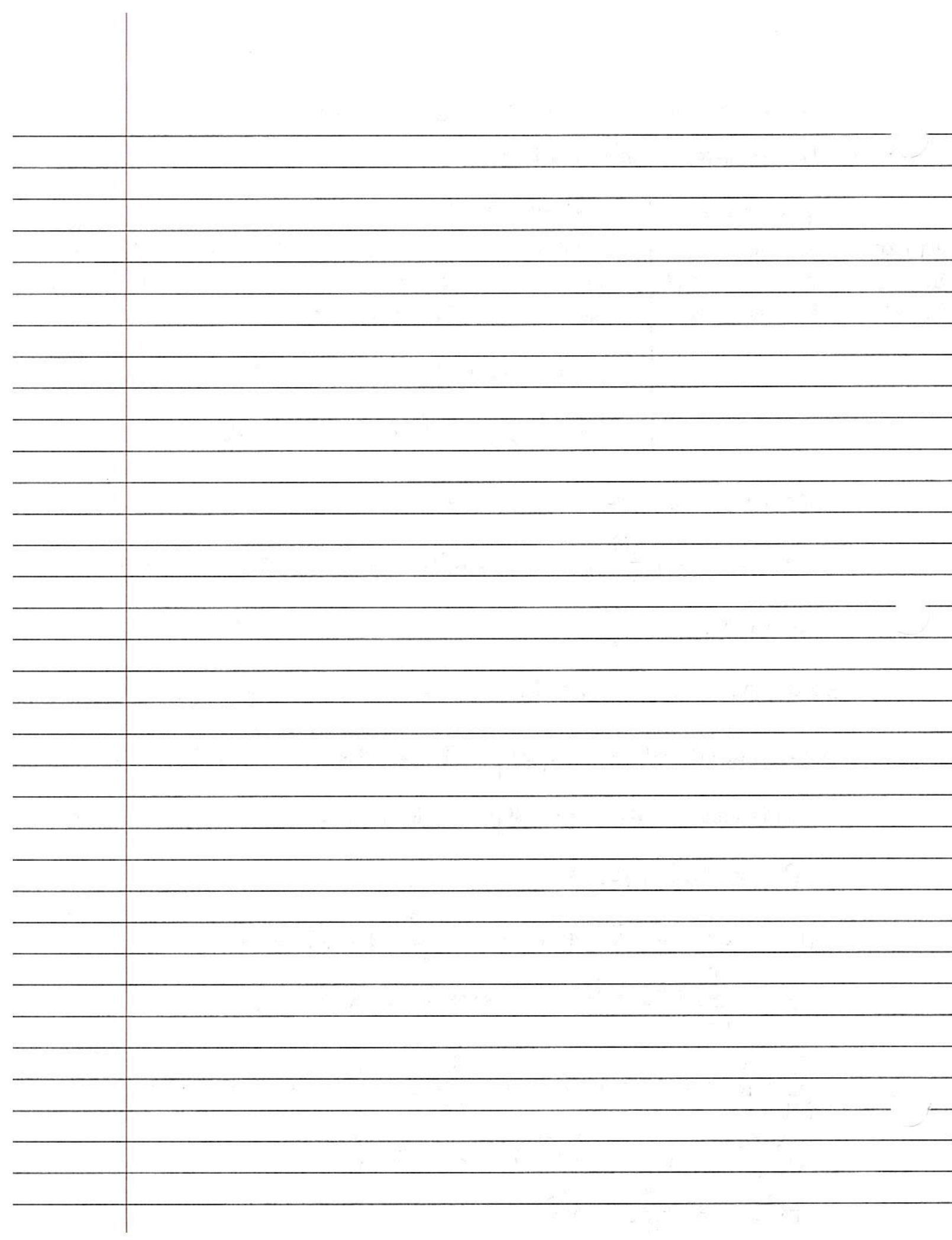
$$d^{p-2}(x) = \bar{x} + \left(1 - \frac{p-2}{V}\right)(X - \bar{x})$$

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

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

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

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



\Rightarrow Remarks:

when $p=1$, and z , \bar{x} is admissible under square loss.

(*) \Rightarrow Review: $x|\mu \sim N(\mu, \sigma^2)$

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

What is the marginal distribution of X

Method 1 $\Rightarrow f(x) = \int_{-\infty}^{\infty} f(x|\mu) \cdot \pi(\mu) d\mu$

Method 2 $\Rightarrow x = \mu + z$ where $\mu \sim N(\mu_0, \sigma_0^2)$
 $z \sim N(0, \sigma^2)$

(*) \Rightarrow A general result: $x_i \stackrel{iid}{\sim} N(\mu_i, \sigma_i^2)$,

then $\sum_{i=1}^n a_i x_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$, the

distribution of X is $N(\mu_0, \sigma_0^2 + \sigma^2)$

\Rightarrow predictive distribution

Given x_1, \dots, x_n we want to estimate x^* by

finding $f(x^* | x_1, \dots, x_n) = \int f(x^* | \theta, x_1, \dots, x_n) f(\theta | x_1, \dots, x_n) d\theta$

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

\Rightarrow A special case $x_1, \dots, x_n, x^* | \theta \stackrel{iid}{\sim} f(x|\theta)$

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

Example: $x_1, \dots, x_n, x^* | \mu \stackrel{iid}{\sim} N(\mu, \sigma^2)$ σ^2 known

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

* Question is how to find the predictive distribution

$$f(x^* | x_1, \dots, x_n)$$

Sol: $\mu | x_1, \dots, x_n \sim N(\mu_1, \sigma_1^2)$

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

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

$$\begin{aligned} x^* | \mu &\sim N(\mu, \sigma^2) \\ \mu | x_1, \dots, x_n &\sim N(\mu_1, \sigma_1^2) \end{aligned} \quad \Rightarrow x^* | x_1, \dots, x_n \sim N(\mu_1, \sigma_1^2 + \sigma^2)$$

\Rightarrow Empirical Bayes:

$$\Rightarrow \text{Review: 1)} x_i | \mu_i \sim N(\mu_i, 1)$$

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

the Marginal distribution of x_i

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

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

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, $z \sim \text{Gamma}(\alpha, 1)$

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

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

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

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

« 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 $Z_i \sim N(0, 1)$

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

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

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

\Rightarrow Empirical Bayes:

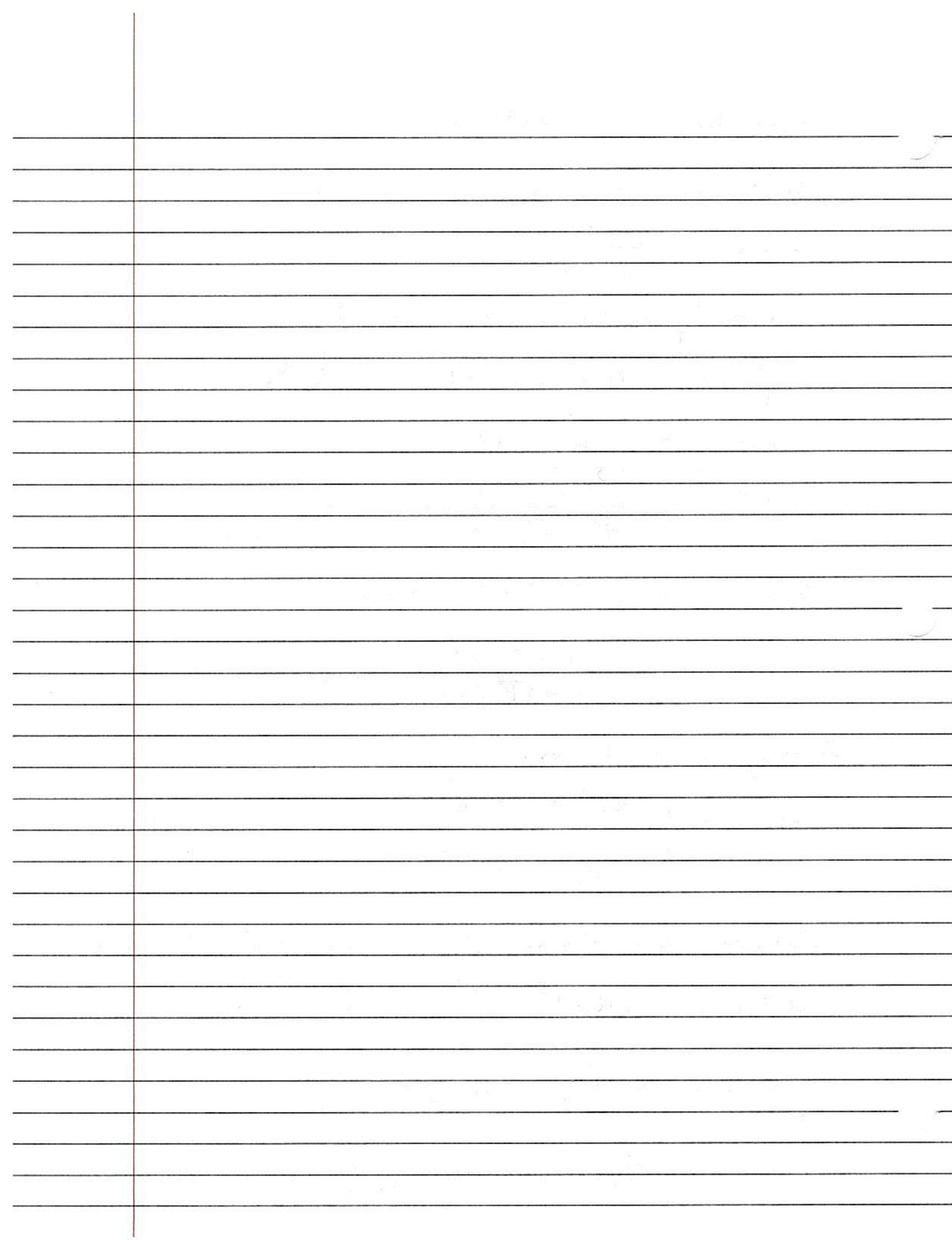
$$\text{Model: } \begin{cases} X_i | \mu_i \sim N(\mu_i, 1) \\ \mu_i \sim N(0, \sigma^2) \end{cases} \quad \text{for } i = 1, 2, \dots, p$$

Suppose we know σ^2 , Bayes estimator for

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

$$\mu_i | X_i \sim N\left(\frac{\sigma^2}{1+\sigma^2} X_i, \frac{\sigma^2}{1+\sigma^2}\right)$$

$$\hat{\mu}^T(x) = E_{\mu|X}(\mu|x) = \frac{\sigma^2}{1+\sigma^2} \cdot x$$



$$\begin{aligned}
 \Rightarrow \text{Bayes Risk: } R(\pi^T, \delta^T) &= \mathbb{E}_{\mu \sim \pi^T} \{ R(\mu, \delta^T) \} \\
 &= \mathbb{E}_{\mu} \mathbb{E}_{x | \mu} \{ \| \mu - \delta^T(x) \|^2 | \mu \} \\
 &= \mathbb{E}_x \mathbb{E}_{\mu | x} \{ \| \mu - \delta^T(x) \|^2 | x \} \\
 &= \mathbb{E}_x \mathbb{E}_{\mu | x} \left\{ \sum_{i=1}^P (\mu_i - \delta^T(x)_i)^2 | x \right\} \\
 &= \mathbb{E}_x \left\{ \sum_{i=1}^P \mathbb{E}_{\mu | x} (\mu_i - \delta^T(x)_i)^2 | x \right\} \\
 &= \mathbb{E}_x \left\{ \sum_{i=1}^P \text{Var} (\mu_i | x_i) \right\} \quad \therefore = \mathbb{E}_x \left\{ \sum_{i=1}^P \frac{\sigma^2}{1+\sigma^2} \right\} = P \cdot \frac{\sigma^2}{1+\sigma^2}
 \end{aligned}$$

\Rightarrow Empirical Bayes estimate:

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

$$\begin{aligned}
 x_i | \mu_i &\sim N(\mu_i, 1) \Rightarrow x_i \sim N(0, 1 + \sigma^2) \\
 \mu_i &\sim N(0, \sigma^2)
 \end{aligned}$$

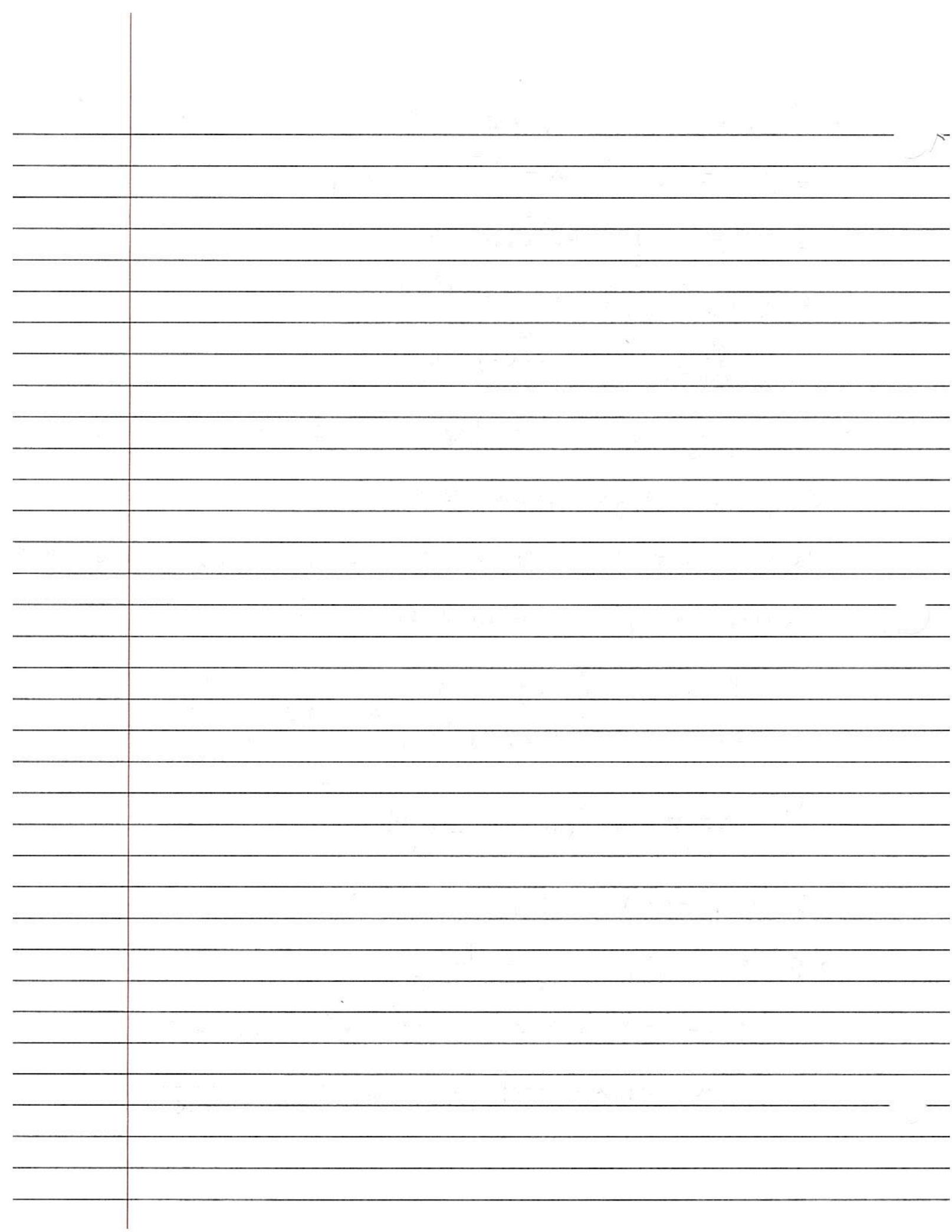
$$\text{If } \frac{x_i}{\sqrt{1+\sigma^2}} \sim N(0, 1) \Rightarrow \sum_{i=1}^P \frac{x_i^2}{1+\sigma^2} \sim \chi_p^2$$

$$\mathbb{E}_x \left(\frac{1}{\|x\|^2} \right) = \frac{1}{P-2}$$

$$\Rightarrow \mathbb{E}_x \left\{ \frac{1}{\|x\|^2} \right\} = \frac{1}{(P-2)(1+\sigma^2)}$$

$$\mathbb{E}_x \left(1 - \frac{P-2}{\|x\|^2} \right) = 1 - \frac{1}{1+\sigma^2} = \frac{\sigma^2}{1+\sigma^2}$$

We can replace $\frac{\sigma^2}{1+\sigma^2}$ by $(1 - \frac{P-2}{\|x\|^2})$



{ A note :

Another approach to estimate σ^2

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

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

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

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

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

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

to find the Bayes risk of d

$$r(\pi^T, d^{n-2}(x)) = E_{\mu} E_{x|\mu} (\|u - d(x)\|^{p-2} | x)$$

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

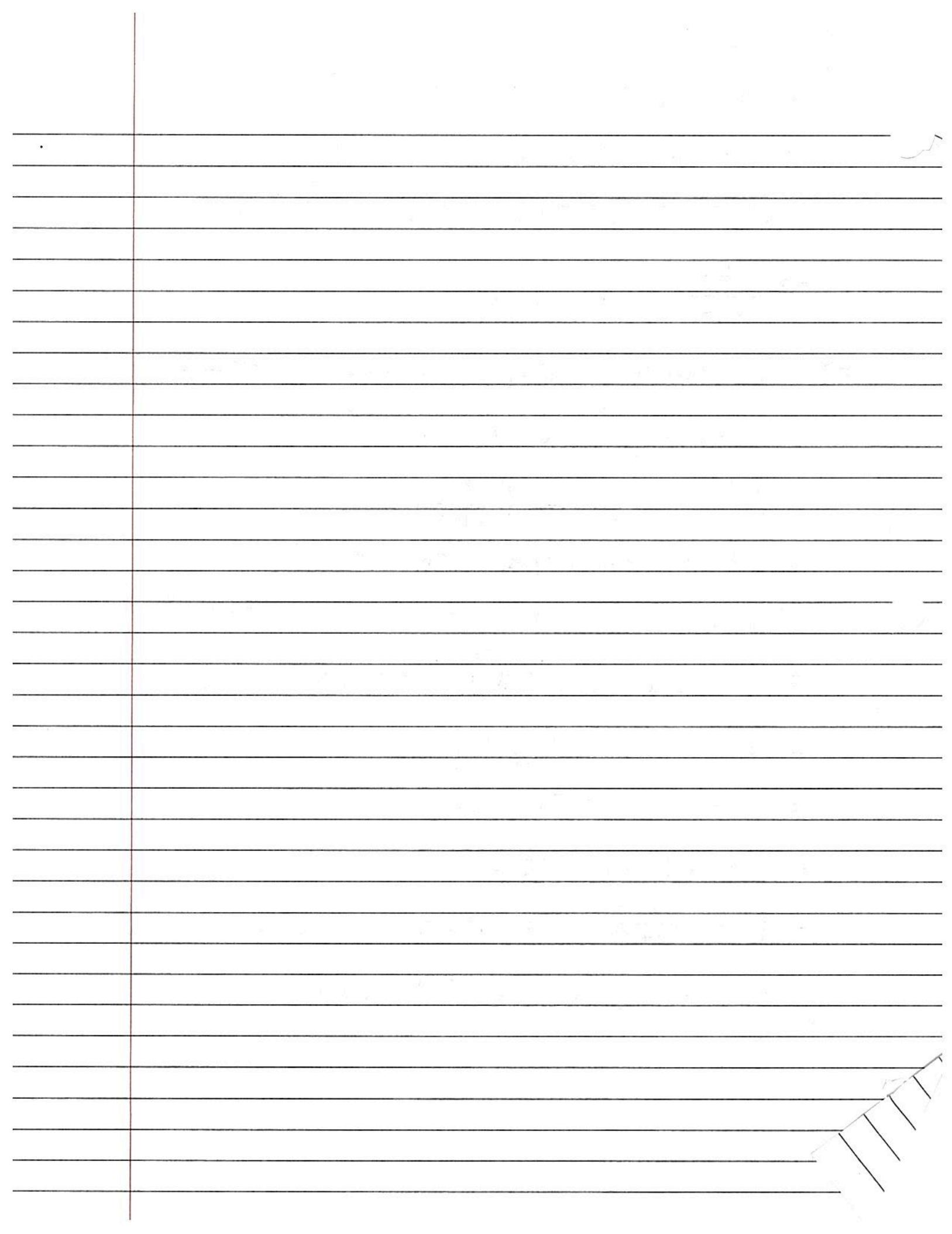
$$= E_x \left\{ \left(p - \frac{p-2}{\|x\|^2} \right) \right\}$$

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

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

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

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



<< Feb. 04, 2015 >> STAT 846

⇒ choice of prior { how to choose the prior }

Methods: 1) Empirical Bayes:

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

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

$$f(x|T) = \int f(x|\theta) f(\theta|T) d\theta$$

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

2) physical method by Bayes:

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

An example : θ : recombination rate

3) Non-informative prior by Jeffrey and Laplace

Roughly, Bayes inference = MLE

$$\text{Example: } x_1, \dots, x_n | \mu \sim N(\mu, \sigma^2) \quad \mu \sim N(\mu_0, \sigma_0^2)$$

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

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

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 V of S students

θ is unknown, can not be replicated.

θ is a R.V only because θ is unknown.

$$\theta \sim N(1, 20^{-1})$$

- $\Pi(\theta)$ 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 conjugate prior.

6) hierarchical Modelling

$$x|\theta \sim f(x|\theta), \theta_1, \dots, \theta_n | \sigma \sim f(\theta|\sigma)$$

$$\sigma \sim \Pi(\sigma)$$

Hierarchical modelling example:

$$x_i | \mu_i \sim N(\mu_i, 1) \quad \left\{ \begin{array}{l} Y_{i1}, \dots, Y_{in} | \mu_i \sim N(\mu_i, \sigma_i^2) \\ x_i = \bar{Y}_i \end{array} \right\}$$

$$\mu_1, \dots, \mu_n | \sigma^2 \sim N(\theta, \sigma^2)$$

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

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

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

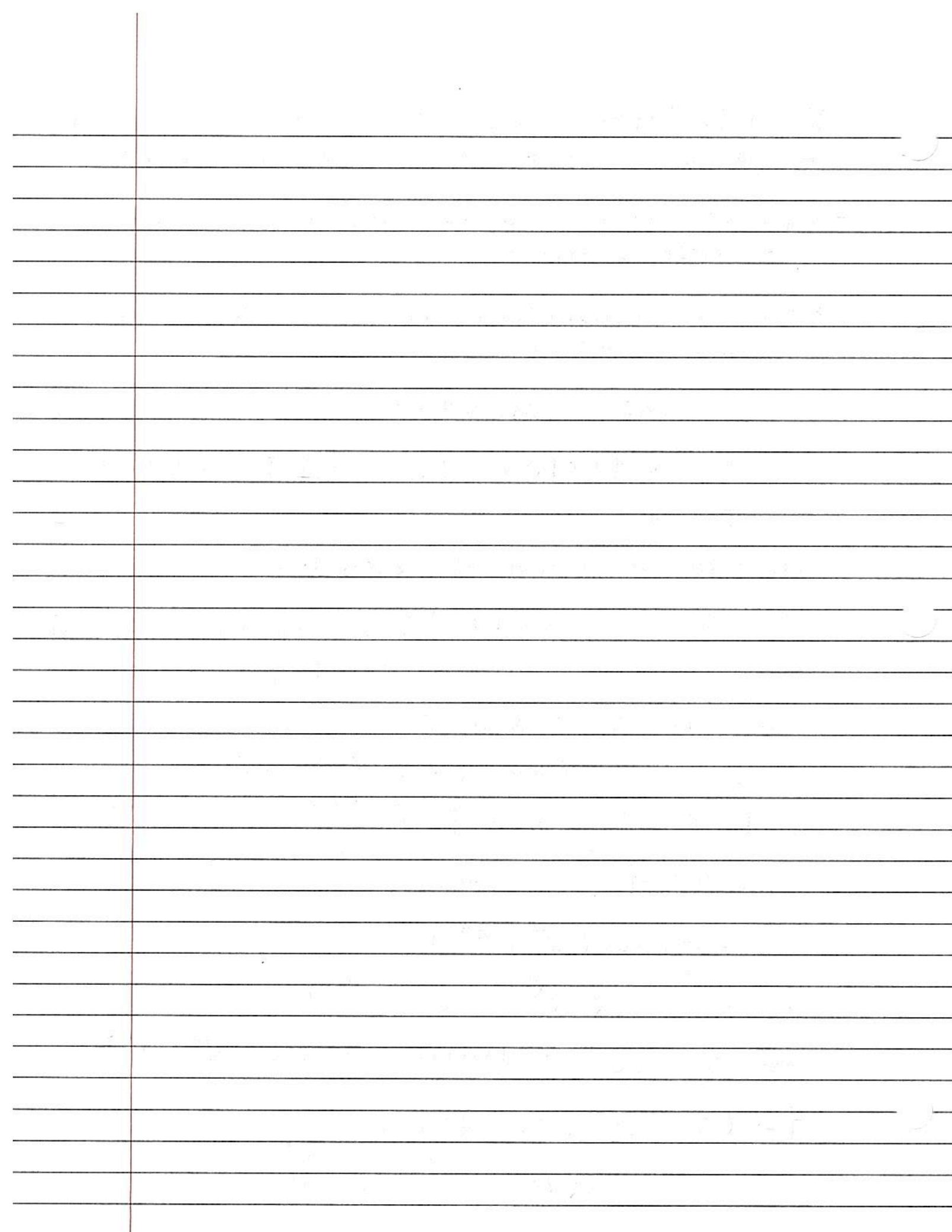
$$x \sim \text{Gamma}(\alpha^*, \beta^*)$$

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

$$\text{Let } \sigma^2 = \frac{1}{x} \sim \text{inverse-Gamma}(\alpha^*, \beta^*)$$

$$f_{\sigma^2}(\sigma^2) d\sigma^2 \propto f_x(\frac{1}{\sigma^2}) d\frac{1}{\sigma^2}$$

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



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

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

$$f(\mu_1, \dots, \mu_n, \theta, \tau^2 | x_1, x_2, \dots, x_n)$$

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

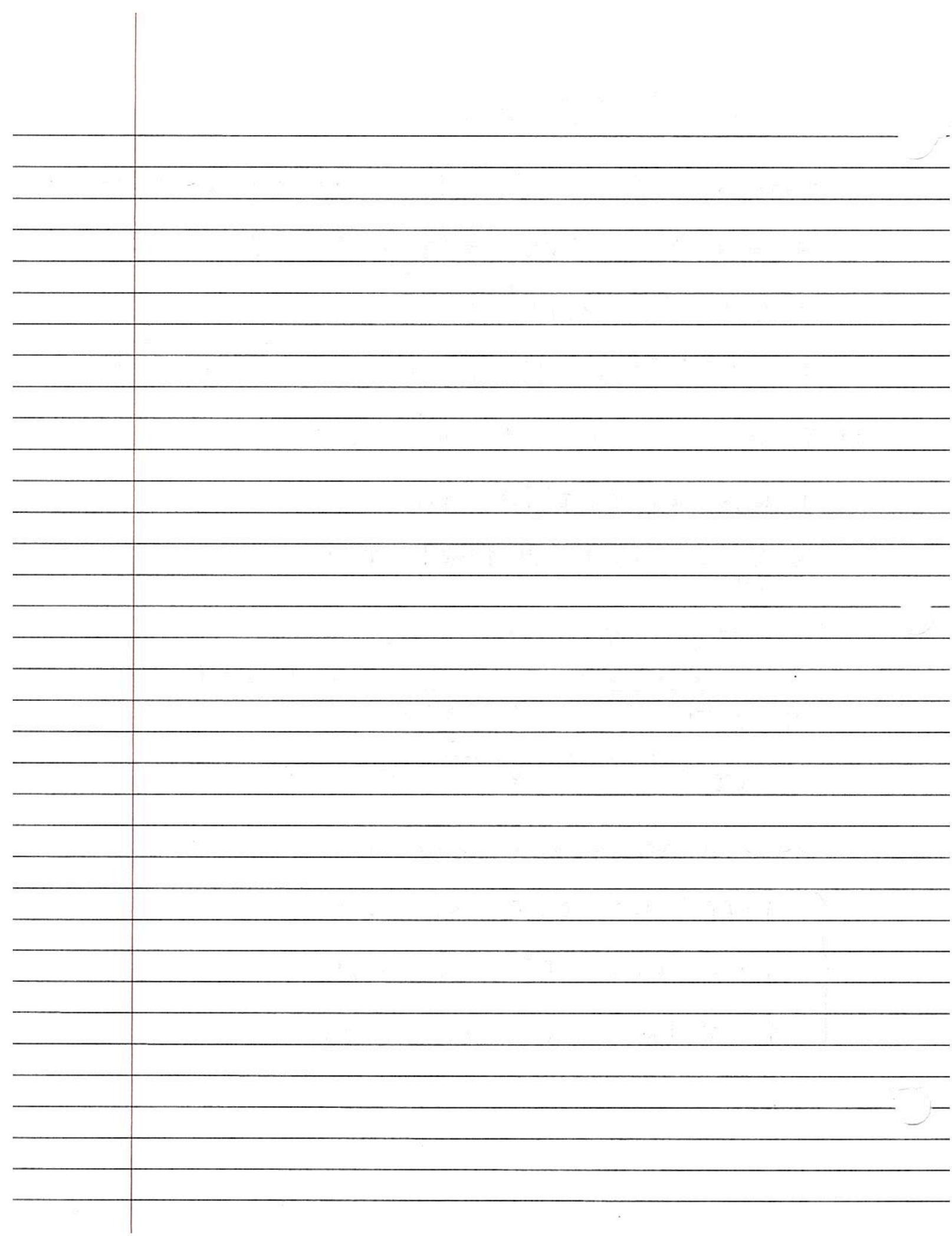
$$\propto \prod_{i=1}^n \frac{(x_i - \mu_i)^2}{2} \times (\tau^2)^{-\frac{n}{2}} \cdot e^{-\sum_{i=1}^n \frac{(\mu_i - \theta)^2}{2\tau^2}}$$

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

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

We will derive full conditionals :

$$\left\{ \begin{array}{l} f(\mu_i | \mu_{-i}, \theta, \tau^2, x_1, \dots, x_n) \\ f(\theta | \mu_{1:n}, \tau^2, x_1, \dots, x_n) \\ f(\tau^2 | \mu_{1:n}, \theta, x_1, \dots, x_n) \end{array} \right.$$



<< Feb. 06, 2015 >> STAT 846

\Rightarrow Example:

$$x_i | \mu_i \sim N(\mu_i, 1) \quad \text{for } i=1, \dots, p$$

$$\mu_i | \theta \sim N(\theta, \tau^2)$$

$$\theta \sim N(\theta_0, \sigma_0^2)$$

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

$$f(\mu_1, \dots, \mu_p, \theta, \tau^2) \propto \prod_{i=1}^p f(x_i | \mu_i) \cdot \prod_{i=1}^p f(\mu_i | \theta, \tau^2) \cdot \Pi_\theta(\theta) \cdot \Pi_{\tau^2}(\tau^2)$$

$$\text{Data: } D : \rightarrow x_1, \dots, x_p$$



$$\text{Parameter: } P : \rightarrow \mu_1, \dots, \mu_p$$



$$\text{Hyper-parameter: HP: } \rightarrow \theta, \tau^2$$



$$\Pi_\theta(\theta) \quad \Pi_{\tau^2}(\tau^2)$$

\Rightarrow Full Conditionals

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

$$\left. \begin{array}{l} x_i | \mu_i \sim N(\mu_i, 1) \\ \mu_i | \theta, \tau^2 \sim N(\theta, \tau^2) \end{array} \right\} \Rightarrow$$

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

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

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

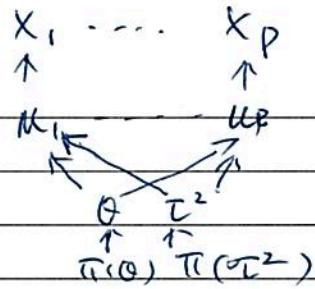
$$\mu_1, \dots, \mu_p | \theta, \tau^2 \sim N(\theta, \tau^2)$$

$$\theta \sim N(\theta_0, \sigma_0^2)$$

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

15

<< Feb. 09, 2015 >> STAT 846



⇒ Baseball example:

$$\text{Let } X = (x_1, \dots, x_p)$$

$$f(\mu_1, \dots, \mu_p, \theta, \tau^2 | X) \propto \prod_{i=1}^p f(x_i | \mu_i) \cdot \prod_{i=1}^p f(\mu_i | \theta, \tau^2) \cdot \pi(\theta) \cdot \pi(\tau^2)$$

After we have samples:

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

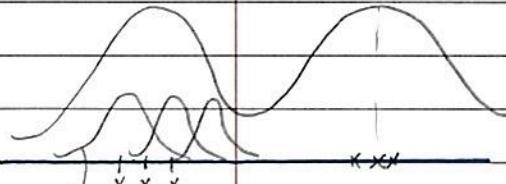
To look at $f(\mu_j | X)$ marginal posterior of μ_j

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

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

(where $f(\mu_j | X, \theta, \tau^2) \sim N(\theta + \tau^2(x_j - \theta), \tau^2)$)

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



This is an application of Rao-Blackwell

$$\text{Kernel Density} \quad \varphi\left(\frac{x_i - \bar{x}}{n}\right) \cdot \frac{1}{h}$$

formulas

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

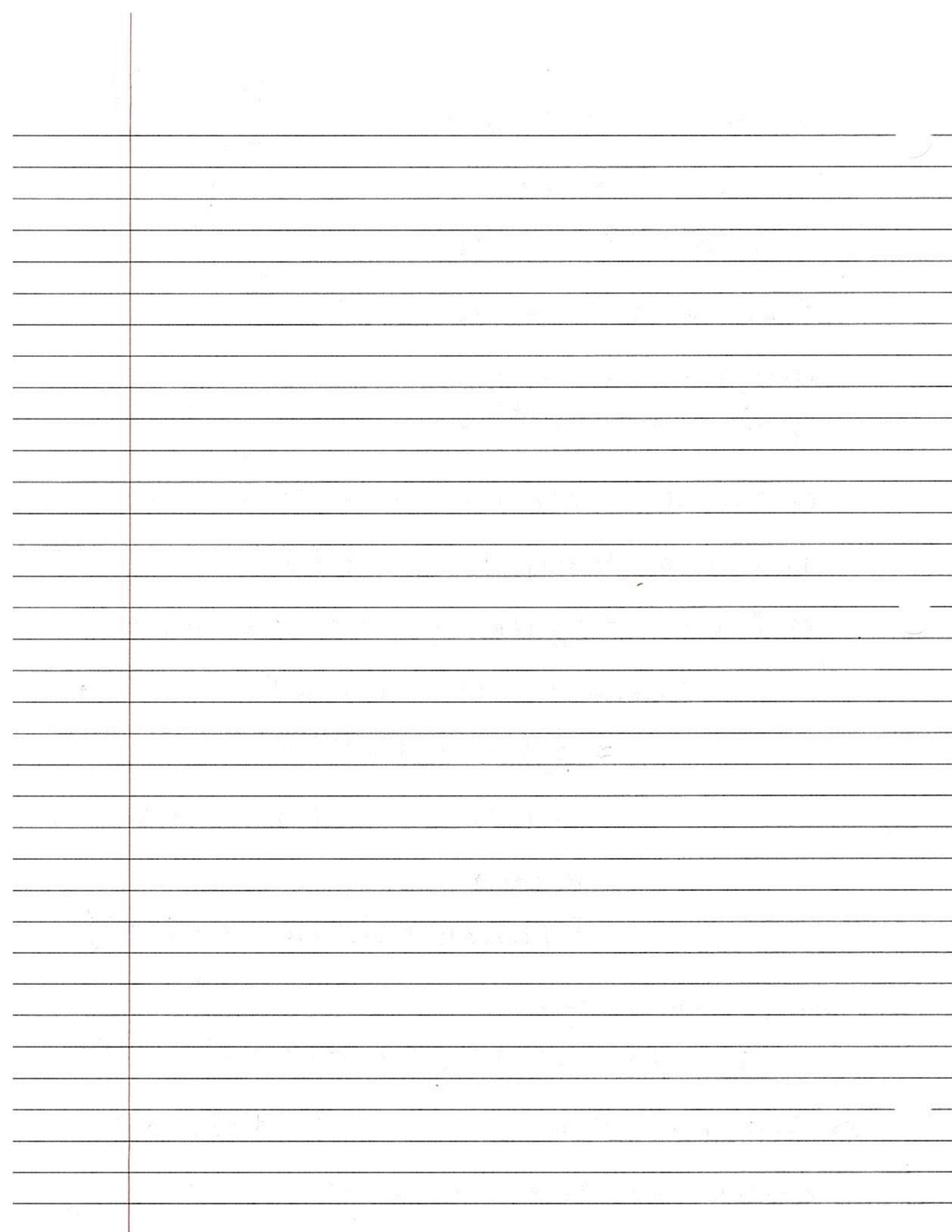
$$\mathbb{E}\{t(\mu, \theta)\} = \mathbb{E}_\theta\{\mathbb{E}_\mu(t(\mu, \theta) | \theta)\}$$

make a note in here:

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

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

Suppose, we have an estimator of θ, τ^2 by



looking $f(x_i | \theta, \tau^2)$, denoted by $\hat{\theta}$, $\frac{1}{\hat{\tau}^2} (= \frac{\hat{\tau}^2}{1 + \hat{\tau}^2})$

$$f(\mu_i | x) \approx f(\mu_i | x, \hat{\theta}, \hat{\tau}^2) = N(\hat{\theta} + \frac{1}{\hat{\tau}^2} (\hat{\theta} - x_i), \frac{1}{\hat{\tau}^2})$$

$$\text{in particular: J-S: } \hat{\theta} = \bar{x} \quad \frac{1}{\hat{\tau}^2} = 1 - \frac{P-3}{V}$$

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

\Rightarrow predictive distribution (Another Method.)

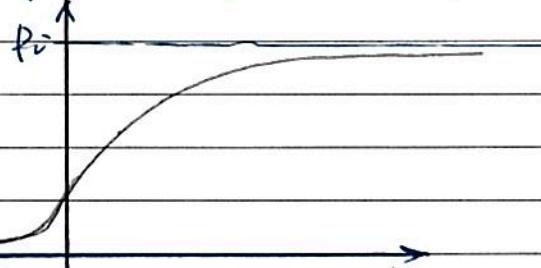
$$\text{Example: } y_i | n_i, p_i \sim \text{Bin}(n_i, p_i)$$

↑
pre-season

$$z_i | N_i, p_i \sim \text{Bin}(N_i, p_i)$$

↑
full season.

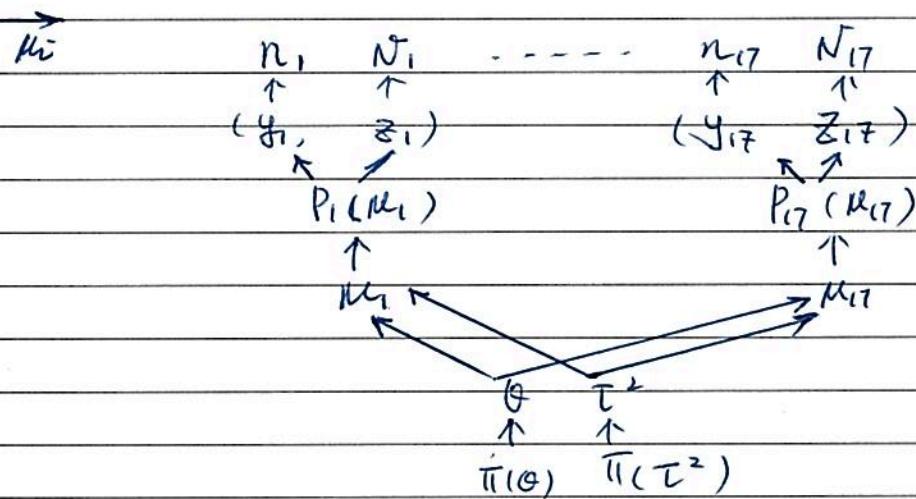
$$p_i = e^{\mu_i} / (1 + e^{\mu_i}) \quad (\text{inv-logistic transformation})$$



$$\mu_1, \dots, \mu_P | \theta, \tau^2 \sim N(\theta, \tau^2)$$

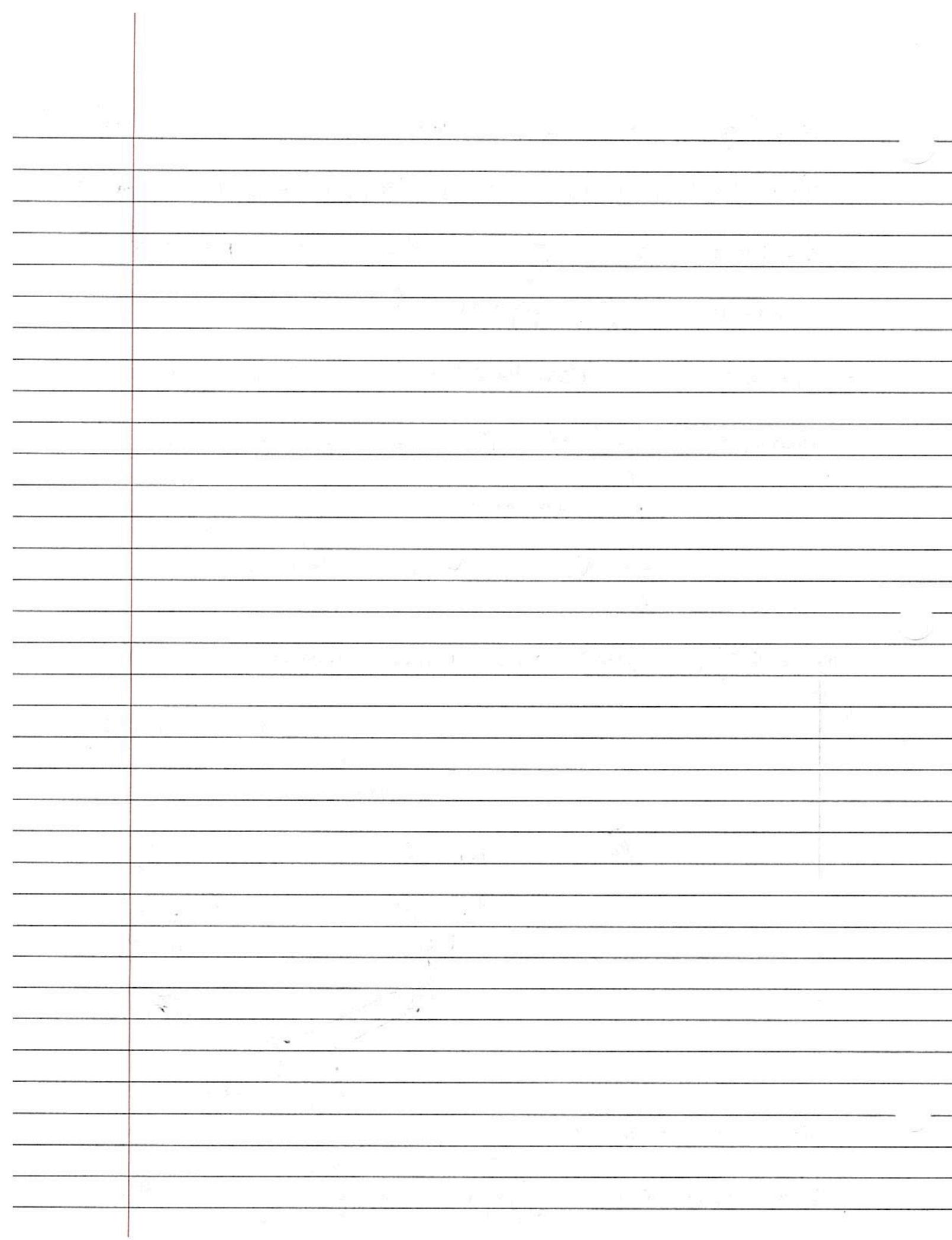
$$\tau^2 \sim \text{inv-Gamma}(\alpha^*, \beta^*)$$

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



Posterior of μ_i, θ, τ^2

$$f(\mu_1, \dots, \mu_7, \theta, \tau^2) \propto \prod_{i=1}^{17} f(y_i | \mu_i) \cdot \prod_{i=1}^{17} f(\mu_i | \theta, \tau^2) \cdot \pi(\theta) \pi(\tau^2)$$



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

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

\Rightarrow Gibbs sampling

$$\begin{cases} 1) \mu_i | y_i, \theta, \tau^2 \\ 2) \theta | \mu_1, \dots, \mu_7, \tau^2, y_1, \dots, y_7 \\ 3) \tau^2 | \theta, \mu_1, \dots, \mu_7, y_1, \dots, y_7 \end{cases}$$

\Rightarrow Metropolis - Hastings

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

Starting from $\mu^{(0)}$

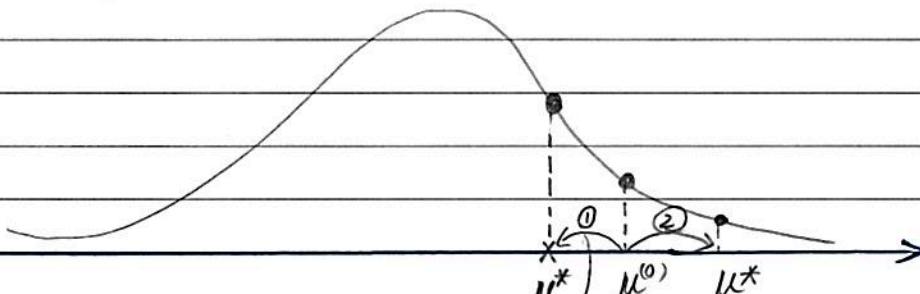
1) propose μ^* from P

2) Draw $u \sim \text{unif}(0, 1)$

3) if $u < \min \{ 1, \frac{\pi(\mu^*) P(\mu^* | \mu^{(0)})}{\pi(\mu^{(0)}) P(\mu^{(0)} | \mu^*)} \}$

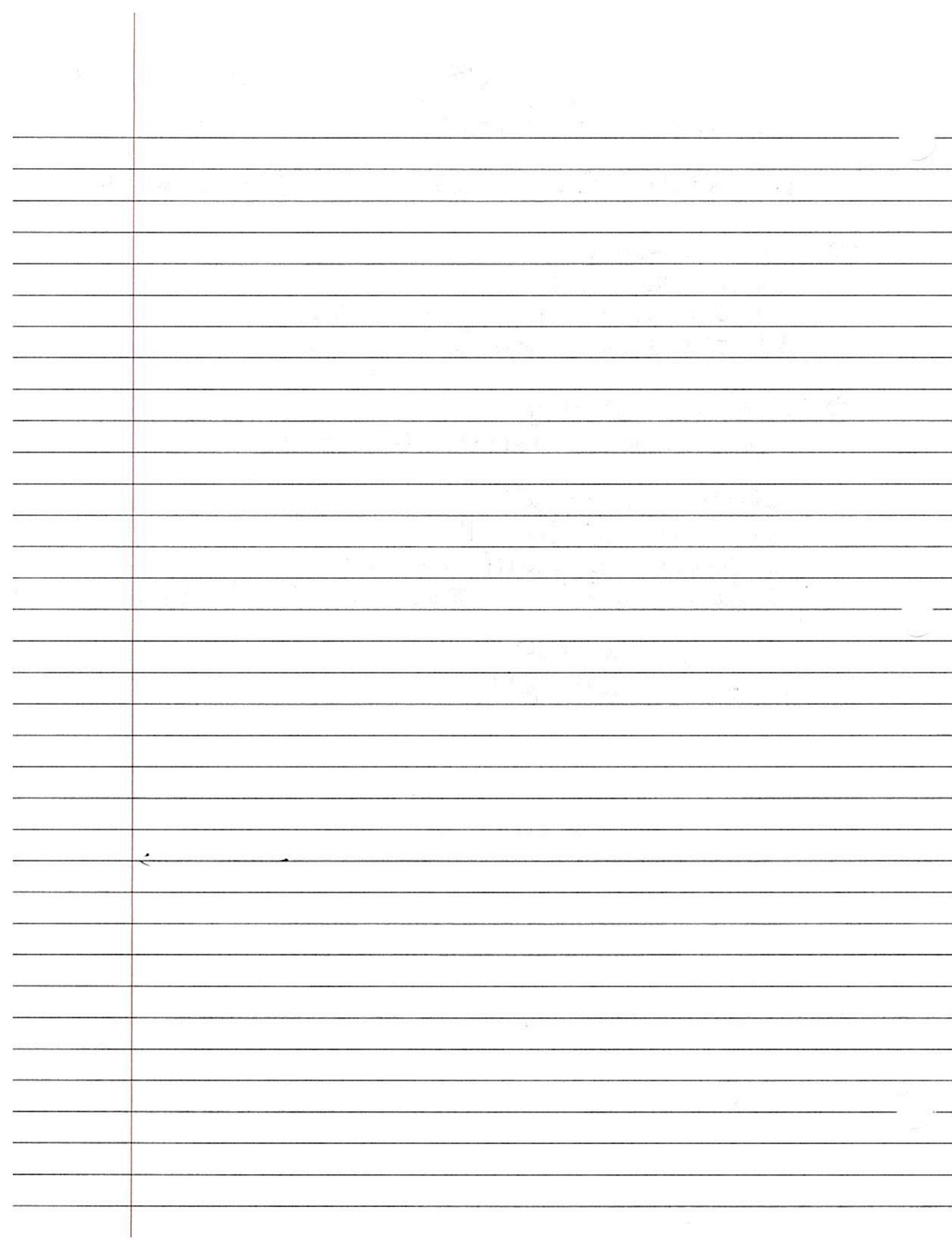
Accept. $\mu^{(1)} = \mu^*$

o/w $\mu^{(1)} = \mu^{(0)}$



$$\textcircled{1} \quad \mu^{(1)} = \mu^*$$

$$\textcircled{2} \quad \mu^{(1)} = \mu^{(0)}$$



< Feb. 11, 2015 > STAT 846

$\Pr(z_i=k | y_{1:17})$

Example:

$$\begin{array}{ccccccc} n_1 & N_1 & \cdots & n_{17} & N_{17} \\ \uparrow & \uparrow & & \uparrow & \uparrow \\ (y_1, z_1) & \cdots & (y_{17}, z_{17}) \\ \uparrow & \uparrow & & \uparrow & \uparrow \\ p_1 & \cdots & p_{17} \\ \uparrow & & \uparrow \\ \mu_1 & \cdots & \mu_{17} \end{array}$$

$$y_i | \mu_i \sim \text{Bin}(n_i, p_i)$$

$$z_i | \mu_i \sim \text{Bin}(N_i, p_i)$$

n_i, N_i are co-variates

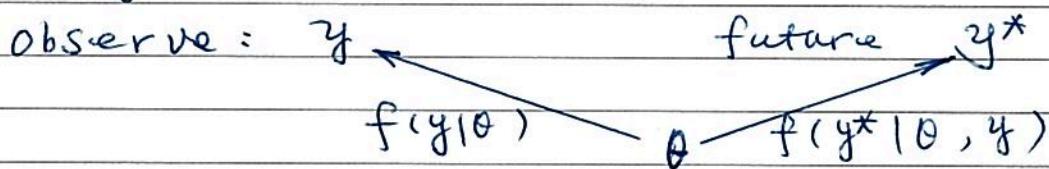
$$\mu_1, \dots, \mu_{17} | \theta, \tau^2 \sim N(\theta, \tau^2)$$

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

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

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

\Rightarrow in general term:



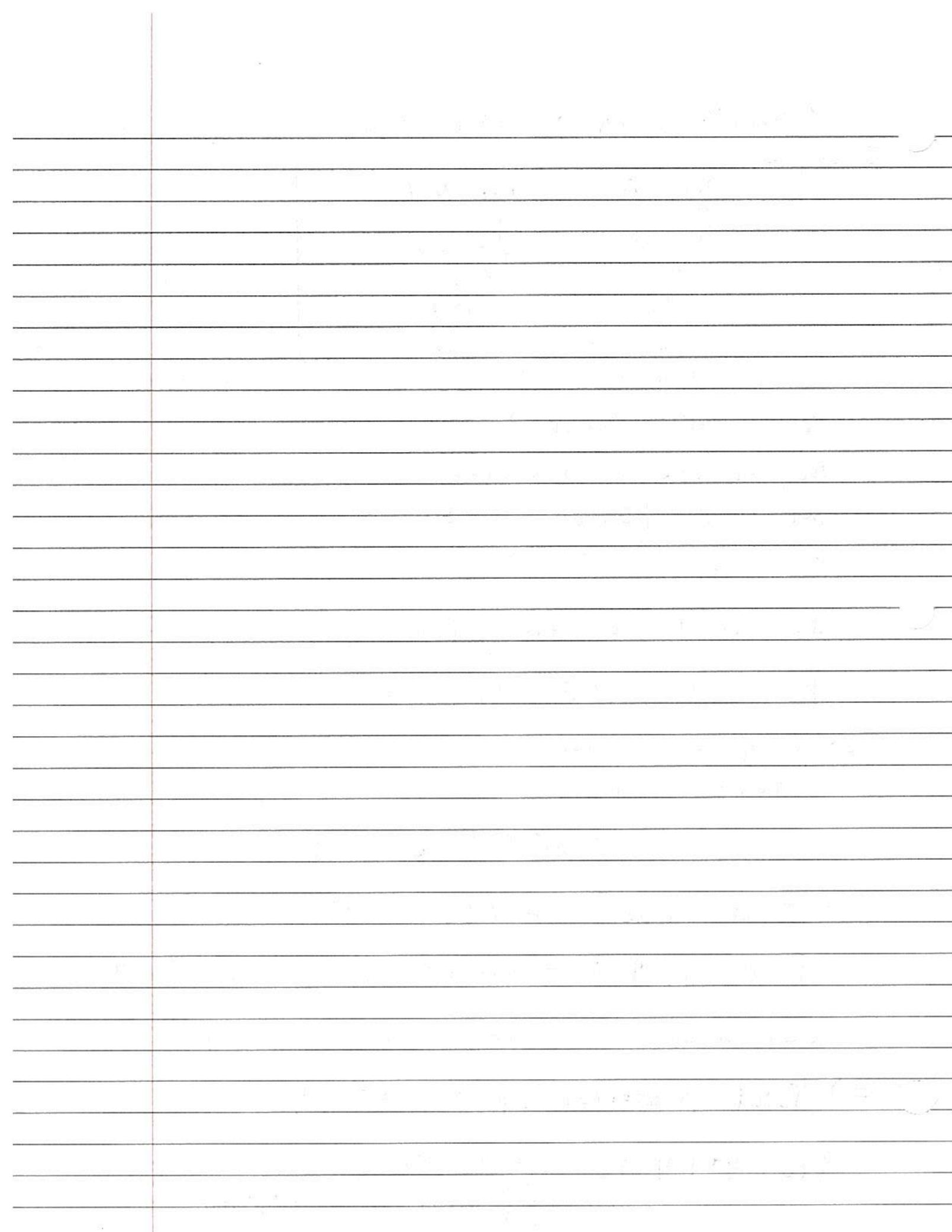
\Rightarrow Joint model for (θ, y, y^*) :

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

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

\Rightarrow Joint posterior of $\theta, y^* | y$

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



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

(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^* .

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

the first time I saw it I thought it was a bird
but then I realized it was a butterfly
it was a small brown butterfly with some
yellow spots on its wings
I think it might be a species of skipper butterfly
I saw it flying around in a field near my house
I tried to catch it but it was too fast
I just watched it for a few minutes before it flew away

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

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

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

\Rightarrow Re-Mark:

Method 1 : is Rao-Blackwellization of Method 2.

$$E_x(t(x)) = E_y(E_x(t(x) | Y)) = E_y(\tilde{t}(Y))$$

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

\Rightarrow Back to example:

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

OR. Draw $\theta^{(i)}, z_j^{(i)} \sim f(\theta, z_j | y_1, \dots, y_{17})$

then discard $z_j^{(i)}$

$$\hat{f}(z_j = k | y_1, \dots, y_{17}) \sim \frac{\sum_{i=1}^N \text{dbin}(k; N_j, p_j^{(i)})}{N}$$

Method 2: Draw $\theta^{(i)}, z_j^{(i)} \sim f(\theta, z_j | y_1, \dots, y_{17})$

using $z_j^{(1)}, \dots, z_j^{(N)}$ only $(z_j^{(i)} \sim f(z_j | y_1, \dots, y_{17}))$

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

$$\Rightarrow \text{Method 1: } \text{Var}(\bar{x}) = \frac{\text{Var}(x_i)}{N}$$

$$\text{Method 2: } \text{Var}(\bar{y}) = \frac{\text{Var}(y_i)}{N}$$