Proposition: $E[(X-b)^2]$ is minimized when $b = E[X]$

Proof. $E[(X-b)^2] = E[X^2] - 2b\,E[X] + b^2 = g(b)$

$$g'(b) = -2E[X] + 2b \overset{\text{set}}{=} 0$$

$$b = E[X]$$

Implication: If $\mathcal{L}_\theta(\hat\theta) = (\hat\theta-\theta)^2$,

then $E[\mathcal{L}_\theta(\hat\theta) \mid \vec{x}]$ is minimized when $\hat\theta = E[\theta \mid \vec{x}]$ = posterior mean

2

From Tuesday's example,

$$\hat\theta_{\substack{Bayes\\L2}} = \frac{(\alpha+\sum X_i)\beta}{n\beta+1}$$

$$= \frac{1}{n\beta+1}\underbrace{(\alpha\beta)}_{\text{prior mean}} + \frac{n\beta}{n\beta+1}\underbrace{(\bar{X})}_{\text{MLE of }\theta}$$

Proposition: $E[|X-b|]$ is minimized when $b = \text{median}(X)$

Proof: $E[|X-b|] = \int_{-\infty}^{\infty} |x-b| f(x)dx = g(b)$ ③

$$g'(b) = \int_{-\infty}^{\infty} \frac{\partial}{\partial b} |x-b| f(x)dx$$

$$= \int_{-\infty}^{\infty} \begin{Bmatrix} -1 & x>b \\ 0 & x=b \\ 1 & x<b \end{Bmatrix} f(x)dx$$

$$= -\int_{b}^{\infty} f(x)dx + \int_{-\infty}^{b} f(x)dx \overset{set}{=} 0$$

$$= -P(X>b) + P(X<b) = 0$$

④

$$P(X>b) = P(X<b)$$

$\therefore$ b is median(X)

Implication: If $\mathcal{L}_\theta(\hat{\theta}) = |\hat{\theta} - \theta|$,

then the Bayes estimator is the posterior median.

Back to our example,

$\hat{\theta}_{\substack{Bayes \\ L1}}$ would not have a closed form.

⑤

There are other possible loss functions

Example. $\mathcal{L}_\theta(\hat\theta) = \begin{cases} 0 & \hat\theta = \theta \\ 1 & \text{elsewhere} \end{cases}$

$$E[\mathcal{L}_\theta(\hat\theta) \mid \vec{x}]$$

$$= 0 \cdot P[\theta = \hat\theta \mid \vec{x}] + 1 \cdot P[\theta \neq \hat\theta \mid \vec{x}]$$

$$= P[\theta \neq \hat\theta \mid \vec{x}]$$

$$= 1 - P[\theta = \hat\theta \mid \vec{x}]$$

To minimize this, maximize $P[\theta = \hat\theta \mid \vec{x}]$ ⑥

Implication: Choose $\hat\theta$ to be the posterior mode

This gives you an MAP estimator

↑

maximum a posteriori

Note: $\pi(\theta \mid \vec{x}) = \dfrac{\pi(\theta) L(\theta)}{m(\vec{x})}$ is proportional to $\pi(\theta) L(\theta)$

<u>Binomial / Beta example</u>

$$X \sim \text{Bino}(n, \theta), \quad \theta \sim \text{Beta}(\alpha, \beta)$$

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$0 < \theta < 1$$

$$\pi(\theta|x) \propto \pi(\theta) L(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

8

$$\propto \theta^{\alpha-1+x} (1-\theta)^{\beta-1+n-x}$$

$$\sim \text{Beta}(\alpha+x, \beta+n-x)$$

For L_2 loss, $\hat{\theta}$ = posterior mean

$$= \frac{\alpha+x}{\alpha+x+\beta+n-x} = \frac{\alpha+x}{\alpha+\beta+n}$$

$$= \frac{\alpha+\beta}{n+\alpha+\beta} \underbrace{\left(\frac{\alpha}{\alpha+\beta}\right)}_{\text{prior mean}} + \frac{n}{n+\alpha+\beta} \underbrace{\left(\frac{X}{n}\right)}_{\text{MLE}}$$

9

<u>Simulation</u>:

Assume that you have access to a $U(0,1)$ random number generator

Let $F(x)$ be a cdf (continuous)

$$\text{Let } Y = F^{-1}(U) \qquad u = F(y)$$

$$\frac{du}{dy} = f(y)$$

$$g(y) = h(u)\left|\frac{du}{dy}\right|$$
$$= 1\,|f(y)| = f(y)$$

10

That is: start with a sample from $U(0,1)$

apply F^{-1} to each number generated.

The resulting sample comes from a distribution with cdf F.

Simulate some values from an exponential distribution.

$$f(x) = \lambda e^{-\lambda x}$$
$$F(x) = 1 - e^{-\lambda x} = u$$
$$e^{-\lambda x} = 1 - u$$

$$-\lambda x = \ln(1-u)$$

$$x = -\frac{1}{\lambda}\ln(1-u)$$

HW #7 due 3/7

$\left.\begin{matrix} 5.59 \\ 60 \end{matrix}\right\}$ simulation

$\left.\begin{matrix} 7.22 \\ .23 \end{matrix}\right\}$ Bayes

Stat 562 HW#7

5.59 Prove that the algorithm of Example 5.6.7 generates a beta(a, b) random variable.

5.60 Generalize the algorithm of Example 5.6.7 to apply to any bounded pdf; that is, for an arbitrary bounded pdf $f(x)$ on $[a, b]$, define $c = \max_{a \le x \le b} f(x)$. Let X and Y be independent, with $X \sim$ uniform(a, b) and $Y \sim$ uniform$(0, c)$. Let d be a number greater than b, and define a new random variable

$$W = \begin{cases} X & \text{if } Y < f(X) \\ d & \text{if } Y \ge f(X). \end{cases}$$

(a) Show that $P(W \le w) = \int_a^w f(t)dt/[c(b-a)]$ for $a \le w \le b$.

(b) Using part (a), explain how a random variable with pdf $f(x)$ can be generated. (*Hint:* Use a geometric argument; a picture will help.)

7.22 This exercise will prove the assertions in Example 7.2.16, and more. Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\theta, \sigma^2)$ population, and suppose that the prior distribution on θ is $\mathrm{n}(\mu, \tau^2)$. Here we assume that σ^2, μ, and τ^2 are all known.

(a) Find the joint pdf of $\bar{X}$ and θ.

(b) Show that $m(\bar{x}|\sigma^2, \mu, \tau^2)$, the marginal distribution of $\bar{X}$, is $\mathrm{n}(\mu, (\sigma^2/n) + \tau^2)$.

(c) Show that $\pi(\theta|\bar{x}, \sigma^2, \mu, \tau^2)$, the posterior distribution of θ, is normal with mean and variance given by (7.2.10).

7.23 If S^2 is the sample variance based on a sample of size n from a normal population, we know that $(n-1)S^2/\sigma^2$ has a χ^2_{n-1} distribution. The conjugate prior for σ^2 is the *inverted gamma* pdf, IG(α, β), given by

$$\pi(\sigma^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{(\sigma^2)^{\alpha+1}} e^{-1/(\beta\sigma^2)}, \quad 0 < \sigma^2 < \infty,$$

where α and β are positive constants. Show that the posterior distribution of σ^2 is IG$(\alpha + \frac{n-1}{2}, [\frac{(n-1)S^2}{2} + \frac{1}{\beta}]^{-1})$. Find the mean of this distribution, the Bayes estimator of σ^2.