ch. 4.

Bayesian Estimation

$$\underset{\sim}{X} \sim f(x|\theta) = P_\theta(x)$$

Suppose θ is a random variable with prior probability density $\pi(\theta)$, $\theta \in \Omega$

$X|\theta \sim f(x|\theta)$, { likelihood based on x / sampling distribution of data $\underset{\sim}{x}$.

The joint p.d.f of $(\underset{\sim}{X}, \theta)$ is

$$f(\underset{\sim}{x}, \theta) = f(x|\theta)\pi(\theta), \quad x \in \mathcal{X}, \theta \in \Omega$$

The marginal p.d.f of $\underset{\sim}{X}$ is $m(x) = \int_\Omega f(x|\theta)\pi(\theta)\,d\theta$.

The posterior density of θ given $\underset{\sim}{x}$ (after observing $\underset{\sim}{x}$)

(statistician's opinion about θ after observing $\underset{\sim}{x}$):

$$\pi(\theta|\underset{\sim}{x}) = f(x|\theta)\pi(\theta)/m(x).$$

Example 1. $X_1, \dots, X_n \overset{iid}{\sim} \text{Bin}(1, \theta)$.

Take $\theta \sim \text{Beta}(a, b)$, $a, b > 0$ are known

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}, \quad 0 < \theta < 1,$$

$$f(\underset{\sim}{x}|\theta) = \theta^{\sum^n x_i}(1-\theta)^{n-\sum^n x_i}$$

$Y = \sum_{i=1}^n X_i$ is complete and sufficient.

> Lemma. If $T(\underset{\sim}{X})$ is sufficient,
>
> $$\pi(\theta|\underset{\sim}{x}) = \pi(\theta|T(x)).$$
>
> Proof. From Factorization Th.

clearly $\pi(\theta|\underset{\sim}{x}) \propto \theta^{y+a-1}(1-\theta)^{n-y+b-1}$,

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

$$\Rightarrow E(\theta \mid \underset{\sim}{X}) = \frac{a+y}{a+y+b+(n-y)} = \frac{a+y}{a+b+n}$$

$$= \underbrace{\frac{a}{a+b}}_{E(\theta)} \cdot \underbrace{\frac{a+b}{a+b+n}}_{\xrightarrow{n\to\infty} 0} + \underbrace{\bar{X}}_{\text{sample mean}} \underbrace{\frac{n}{a+b+n}}_{\xrightarrow{n\to\infty} 1}.$$

$$Var(\theta \mid y) = \frac{(a+y)(b+n-y)}{(a+b+n)^2(a+b+n+1)}.$$

Example 2. $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 is known.

$\theta \sim N(\mu, \tau^2)$

$$\Rightarrow f(X, \theta) \propto \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \theta)^2\right\} \frac{\exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}}{\sqrt{2\pi}\,\tau}.$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n\sqrt{2\pi}\,\tau} \exp\left\{-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2 - \frac{(\theta-\mu)^2}{2\tau^2} - \frac{1}{2\sigma^2}\sum(X_i - \bar{X})^2\right\}$$

$$\Rightarrow \pi(\theta \mid \bar{X}) \propto \exp\left\{-\frac{1}{2}\left[\frac{n}{2\sigma^2}(\bar{X}-\theta)^2 + \frac{1}{\tau^2}(\theta-\mu)^2\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2 - 2\theta\left(\frac{n}{\sigma^2}\bar{X} + \frac{\mu}{\tau^2}\right) + \frac{n}{\sigma^2}\bar{X}^2 + \frac{\mu^2}{\tau^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\underbrace{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}_{=\frac{1}{\sigma_B^2}}\left(\theta^2 - 2\theta \underbrace{\frac{\frac{n}{\sigma^2}\bar{X} + \frac{1}{\tau^2}\mu}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}}_{\mu_B} + (\quad)^2\right] + \cdots\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_B^2}(\theta-\mu_B)^2\right\}$$

$\therefore\ \theta \mid \bar{X} \sim N(\mu_B, \sigma_B^2)$

μ_B is the weighted average between the prior mean and $\bar{X}$,

Def. $\pi \in \Pi$ is a family of conjugate priors for $f \in \mathcal{F}$.
if $\pi(\theta \mid \underset{\sim}{x}) \in \Pi$, $\forall f$, $\forall x \in \mathcal{X}$.

Thus, Beta is conjugate for Binomial
Normal is conjugate for normal.

$\boxed{\text{Decision Theory point of view.}}$

$f(\underset{\sim}{x} \mid \theta) \in \mathcal{F}$, family of distributions (densities) for $\underset{\sim}{x}$

Ω = parameter space.

$\pi(\theta)$: distribution (may be a density) on Ω.

$\mathcal{D} = \{\delta(x) = \text{decision rule}\}$.

$L(\theta, \delta(x))$ loss when θ is truth, and use $\delta(x)$ as estimator for $g(\theta)$.

$R(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta(x)) f(x \mid \theta)\, dx$: risk function of θ for δ.

Def. The Bayes risk is

$$r(\pi, \delta) = E_\pi R(\theta, \delta) = \int_\Omega R(\theta, \delta)$$

Remark.
$$r(\pi, \delta) = E_\pi E_\theta L(\theta, \delta(\underset{\sim}{x}))$$
$$= \int_\Omega \int_{\mathcal{X}} L(\theta, \delta(x)) f(x \mid \theta) \pi(\theta)\, dx\, d\theta$$

Def. An estimator δ_π is Bayes wrt prior π if

$$r(\pi, \delta_\pi) = \inf_\delta r(\pi, \delta).$$

Note: minimizer (existence) must be shown

Construction Theorem.

Suppose there is δ_π with $r(\pi, \delta_\pi) < \infty$,

If $d = \delta_\pi(x)$ minimizes

$$E(L(\theta, d) \mid X = x).$$

for a.e. all x, then δ_π is Bayes.

Proof. For any δ,

$$E(L(\theta, \delta_\pi(x)) \mid X = x) \leq E(L(\theta, \delta(x)) \mid X = x) \quad , \text{a.e. } x.$$

Take expectation on both sides wrt marginal dist of X, we get

$$r(\pi, \delta_\pi) \leq r(\pi, \delta), \qquad \forall \delta.$$

$\therefore$ δ_π is Bayes.

Corollary. If $d = \delta(x)$ minimizes $\int_\Omega L(\theta, d) f(x|\theta) \pi(\theta) d\theta$, $\forall x$,

then δ is Bayes.

Proof. Let $g(x, \theta) = f(x|\theta)\pi(\theta)$,

$$m(x) = \int_{\Theta} g(x, \theta)\, d\theta$$

$$\pi(\theta | \underset{\sim}{x}) = g(x, \theta) / m(x).$$

$$\int_\Omega L(\theta, d) f(x|\theta)\pi(\theta) d\theta = \left(\int_\Omega L(\theta, d) \pi(\theta|x) d\theta\right) \underbrace{m(x)}_{\text{does not depend on } \theta}.$$

//

Remark $E(L(\theta, d) \mid X = x) =$ posterior loss

$$\min_d E(L(\theta, d) \mid X = x) = E(L(\theta, \delta_\pi(x)) \mid X = x) \equiv \text{posterior risk}.$$

Example 1. under squared error loss. $L(\theta, d) = (\theta - d)^2$.

$E((\theta-d)^2 \mid X = x)$ is minimized by $d = E(\theta \mid X)$.

$\therefore\ \delta_\pi(x) = E(\theta \mid X = x)$ is Bayes.

Example 2. under absolute error loss. $L(\theta, d) = |\theta - d|$

$\delta(x) = \text{med}(\theta \mid X)$ is Bayes.

Normal Example (revisit): $X_1, \ldots, X_n$ iid $N(\theta, \sigma^2)$ σ^2 known

prior $\theta \sim N(\mu, \tau^2)$,

i) $(\theta \mid \underset{\sim}{X}) \sim N\left(\dfrac{\tau^2}{\tau^2 + \sigma^2/n}\bar{X} + \dfrac{(\sigma^2/n)\mu}{\tau^2 + \sigma^2/n},\ \dfrac{\sigma^2}{n}\left(\dfrac{\tau^2}{\tau^2 + \sigma^2/n}\right)\right)$.

ii) Under $L(\theta, d) = (\theta - d)^2$; $\delta_\pi(x) = E(\theta \mid \underset{\sim}{X})$.

iii) The posterior risk under $L(\theta, d) = (d - \theta)^2$

$$E((\theta - \delta_\pi(x))^2 \mid X = x) = \frac{\sigma^2}{n}\left(\frac{\tau^2}{\tau^2 + \sigma^2/n}\right)$$

The Bayes risk. $r(\pi, \delta_\pi) = E_\pi\left(E_\theta(\theta - \delta_\pi)^2\right)$

$$= E_m E((\theta - \delta_\pi)^2 \mid X) = E_m\left(\frac{\sigma^2}{n}\left(\frac{\tau^2}{\tau^2 + \sigma^2/n}\right)\right)$$

$$= \frac{\sigma^2}{n}\left(\frac{\tau^2}{\tau^2 + \sigma^2/n}\right) < \frac{\sigma^2}{n}.$$

Theorem, Under sq. error loss, if δ is Bayes wrt π, and an unbiased estimator of θ, then $r(\pi, \delta) = 0$

proof. unbiased. $E(\delta \mid \theta) = \theta$ $\quad E(\delta\theta \mid \theta) = \theta^2$

$\Rightarrow E(\delta\theta) = E(\theta^2)$.

Bayes, $E(\theta \mid X) = \delta(X) \Rightarrow E(\theta\delta \mid X) = \delta^2 \Rightarrow E(\theta\delta) = E(\delta^2)$

$\therefore\ r(\pi, \delta) = E(\theta - \delta)^2 = E(\theta^2) - 2E(\theta\delta) + E(\delta^2) = 0$.

Remark. Bayes estimator is not unbiased, but for the normal example, $\delta_\pi = \delta_{\mu,\tau}(X)$.

$$\lim_{\tau\to\infty} \delta_{\mu,\tau}(X) = \bar{X},$$

Different prior?

Properties of Bayes Estimators

a. Admissibility

Th. If $\left.\begin{cases} \pi(\theta) > 0 & \text{on } \Omega, \\ R(\theta, \delta_\pi) & \text{is continuous in } \theta \\ r(\pi, \delta_\pi) < \infty \end{cases}\right\}$ then δ_π is admissible.

proof. If not, there is δ, $\ni$

$$R(\theta, \delta) \leq R(\theta, \delta_\pi), \quad \forall \theta \in \Omega$$

$$R(\theta, \delta) < R(\theta, \delta_\pi) \quad \text{some } \theta_0 \in \Omega$$

then $$\underbrace{\int_\Omega R(\theta, \delta)\pi(\theta)\,d\theta}_{= r(\pi, \delta)} < \underbrace{\int_\Omega R(\theta, \delta_\pi)\pi(\theta)\,d\theta}_{= r(\pi, \delta_\pi)}.$$

this is contridiction!

b. Minimaxity

Recall, δ_0 is minimax if

$$\sup_\theta R(\theta, \delta_0) \leq \sup_\theta R(\theta, \delta), \quad \forall \theta \in \Omega.$$

- different idea, many results are known.
- Basic idea: minimax δ generally has $R(\theta, \delta) = \text{constant}$ (equalizer rule)

Theorem. A Bayes rule δ_π that is also an equalized rule is minimax.

Proof. If not, there is δ_0 $\exists$.

$$K = \sup_\theta R(\theta, \delta_0) < \sup_\theta R(\theta, \delta_\pi) = R(\theta, \delta_\pi), \ \forall \theta$$

$$\Rightarrow \quad r(\pi, \delta_0) = \int R(\theta, \delta_0)\pi(\theta)d\theta \leq K < \underbrace{\int R(\theta, \delta_\pi)\pi(\theta)d\theta}_{= r(\pi, \delta_\pi)}$$

This is contridiction! since δ_π is Bayes.

Definition. π is called least favarable prior

- The simple idea extends to certain limits of Bayes rule as well.

- Result, $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\bar{X}$ is minimax.

- Example 1.7. P311, minimax estimator of θ in $Bin(n, \theta)$, under sq. error loss.,
  Bayes estimator $E(\theta|X)$ under prior $Beta\left(\sqrt{\frac{n}{4}}, \sqrt{\frac{n}{4}}\right)$ is minimax.

Example, $X_1, \cdots, X_n \overset{iid}{\sim} N_p(\theta, \sigma^2 I_p)$, $\theta \in R^p$, σ^2 is known.

$\bar{X}$ is sufficient & complete statistic for θ.

$\bar{X}$ is minimax.

$\bar{X}$ is admissible if $p=1,2$, but not if $p \geq 3$!

Since $\bar{X} \sim N_p(\theta, \frac{\sigma^2}{n} I_p)$, without loss of generity, assume $\sigma^2/n = 1$, so $X \sim N_p(\theta, I_p)$!

loss: $L(\underset{\sim}{\theta}, \underset{\sim}{d}) = \|\underset{\sim}{\theta} - \underset{\sim}{d}\|^2 = \sum_{i=1}^{p} (\theta_i - d_i)^2$,

$\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_p)'$, $\underset{\sim}{d} = (d_1, \cdots, d_p)' \in R^p$.

James-Stein Theorem. If $p \geq 3$, (P273)

$\delta_{JS}(\underset{\sim}{X}) = (1 - \frac{p-2}{\|X\|^2}) \underset{\sim}{X}$ is better than X,

i.e. $E_\theta \|\underset{\sim}{\theta} - \delta_1(\underset{\sim}{X})\|^2 < E_\theta \|\underset{\sim}{\theta} - \underset{\sim}{X}\|^2$, $\forall \theta \in R^p$.

Proof Write $\delta^{JS}(X) = \underset{\sim}{X} - \underset{\sim}{g}(X)$, where $\underset{\sim}{g}(X) = \frac{p-2}{\|X\|^2} \underset{\sim}{X}$.

clearly $\underset{\sim}{g}(\underset{\sim}{X})$ is differentiable.

From the same argument as Example 5.16, on Page 31, we have

$$E_\theta [g(X)]' (\underset{\sim}{X} - \theta) = E_\theta \sum_{i=1}^{p} \frac{\partial}{\partial X_i} g_i(X) \quad (*)$$

$\therefore R(\theta, \delta_{JS}) = E_\theta \|\underset{\sim}{X} - \underset{\sim}{\theta} - g(X)\|^2$.

$= E_\theta (\underset{\sim}{X} - \underset{\sim}{\theta})'(\underset{\sim}{X} - \underset{\sim}{\theta}) + E_\theta |g(X)|^2 - 2 E_\theta [g(X)]' (X - \theta)$

$$\overset{(*)}{=} p + E_\theta \frac{(p-2)^2}{\|\underset{\sim}{X}\|^2} - 2 E_\theta \sum_{i=1}^{p} \frac{\partial}{\partial X_i}\left[\frac{(p-2)X_i}{\|X\|^2}\right].$$

$$= p + (p-2)^2 E \frac{1}{\|X\|^2} - 2(p-2) E_\theta \underbrace{\sum_{i=1}^{p} \frac{\|X\|^2 - 2X_i^2}{\|X\|^4}}_{p\frac{1}{\|X\|^2} - 2\frac{1}{\|X\|^2}}$$

$$= p + (p-2)^2 E_\theta\left(\frac{1}{\|X\|^2}\right) - 2(p-2)^2 E_\theta\left(\frac{1}{\|X\|^2}\right)$$

$$= p - (p-2)^2 E_\theta \frac{1}{\|X\|^2}.$$

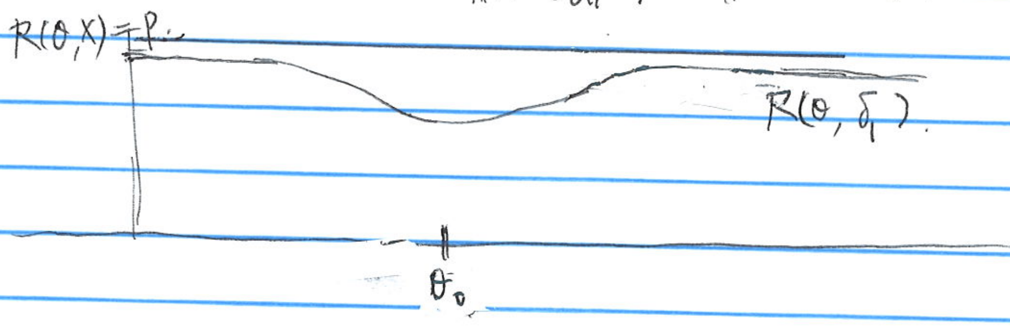
But $E_\theta \|\theta - X\|^2 = p = R(\theta, X)$, so

$$R(\theta, \delta_{JS}) - R(\theta, X) = -(p-2)^2 E_\theta\left(\frac{1}{\|X\|^2}\right) < 0, \quad \forall \theta \in R^p.$$

$\Rightarrow$ result.

Variations.

1. $\delta_1(X) = \theta_0 + (X - \theta_0)\left(1 - \frac{p-2}{\|X - \theta_0\|^2}\right)$ is better than X.



2. If $p > 3$, $\delta_2(X) = \bar{X} 1_p + (X - \bar{X} 1_p)\left(1 - \frac{p-3}{\Sigma(X_i - \bar{X})^2}\right)$ is better than X.

3. $\delta_{JS}^{+}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right)_+ \cdot X$ is better than δ_{JS}.

So δ_{JS} is not admissible.

Application: Efron & Morris, 75 JASA,
77 Scientific American.

Example: Improve an baseball players betting average for season.