Bayesian Methods

Generally, the idea here is to incorporate past knowledge by updating the prior information using the available data.

$\Theta \sim \Lambda$ prior distribution [before observing data]

$\underset{\sim}{X} \mid \Theta = \theta \sim P_\theta$

$\Longrightarrow R(\theta, \delta) = E[L(\theta, \delta(X)) \mid \Theta = \theta] = \int L(\theta, \delta) \, dP_\theta(x)$

and

$$E_\theta L(\theta, \delta) = E_\theta E_{X|\theta}[L(\theta, \delta) \mid \theta] = E_\theta R(\theta, \delta) = \int R(\theta, \delta) \, d\Lambda(\theta)$$

Expected risk (1)

- The estimator that minimizes (1) is called Bayes estimate
- " " " " the conditional expected loss given data. That is,

$$E(L(\theta, \delta) \mid X = x) \quad \text{called posterior risk}$$

The average is taken against the posterior distribution of $\theta \mid \underset{\sim}{X}$.

Theorem: Let $\theta \sim \Lambda$, $X \mid \theta \sim P_\theta$ and $L(\theta, \delta) \geq 0$ for all θ and all δ.

If (a) $E L(\theta, \delta_0) < \infty$ for some δ_0

(b) $\exists \, \delta_\Lambda(x)$ minimizing $E[L(\theta, \delta) \mid X]$ with respect to δ

then δ_Λ is a Bayes estimator.

Proof: Let δ be an arbitrary estimator. Then

$$E[L(\theta, \delta) \mid X] \geq E[L(\theta, \delta_\Lambda) \mid X]$$

$\Rightarrow \delta_\Lambda$ is the Bayes estimate.

Examples

① Weighted square error loss (SEL)

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

The Bayes estimator $\delta_\Lambda(x)$ is the one minimizes

$$E\left[w(\theta)(d-g(\theta))^2 \mid X\right] = d^2 E[W(\theta) \mid X] - 2d\, E[W(\theta) g(\theta) \mid X] + E[W(\theta) g^2(\theta) \mid X]$$

Diff. $\Rightarrow$

$$2d\, E[W(\theta) \mid X] - 2E[W(\theta) g(\theta) \mid X] = 0$$

$\Rightarrow$

$$\delta_\Lambda(x) = \frac{E[W(\theta) g(\theta) \mid X]}{E[W(\theta) \mid X]}$$

If $W(\theta) = 1 \Rightarrow \delta_\Lambda(x) = E(g(\theta) \mid X)$ Posterior mean of $g(\theta)$

If $\mathcal{P}$ is a dominated family with p_θ, the density of P_θ and if Λ is absolutely continuous with Lebesgue density, then the joint density of X and θ is

$$p(x \mid \theta)\, \lambda(\theta) = p_\theta(x)\, \lambda(\theta)$$

The marginal density of X is

$$q(x) = \int p(x \mid \theta)\, \lambda(\theta)\, d\theta$$

Then the posterior density of θ given X is

$$\lambda(\theta \mid x) = \frac{p(x \mid \theta)\, \lambda(\theta)}{q(x)}$$

Therefore, the Bayes estimator under the Weighted SEL is

$$\delta_\Lambda(x) = \frac{\int w(\theta) g(\theta) p(x \mid \theta) \lambda(\theta)\, d\theta}{\int w(\theta) p(x \mid \theta) \lambda(\theta)\, d\theta}$$

② Let $P_\theta \sim$ Binomial with n and θ. Suppose the prior dist. of θ is Beta(α, β) with

$$\lambda(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0<\theta<1$$

$$q(x) = \int p(x|\theta)\lambda(\theta)\,d\theta = \int \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha+x-1}(1-\theta)^{n-x+\beta-1}\,d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$\Rightarrow \lambda(\theta|x) = \frac{p(x|\theta)\lambda(\theta)}{q(x)} = \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}}$$

$$= \frac{\cancel{\binom{n}{x}} \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} \theta^{\alpha+x-1}(1-\theta)^{n-x+\beta-1}}{\cancel{\binom{n}{x}} \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} \frac{\Gamma(\alpha+x)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}}$$

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

$\Rightarrow$ $\theta | X=x \sim$ Beta$(\alpha+x, n-x+\beta)$ conjugate since the posterior family is coming from the same class for prior.

$$\Rightarrow \delta_\Lambda(x) = E(\theta|X) = \frac{\alpha+X}{n+\alpha+\beta}$$

$$= \frac{n}{n+\alpha+\beta} \underset{\substack{\uparrow \\ \text{UMVUE}}}{\frac{X}{n}} + \left(1-\frac{n}{n+\alpha+\beta}\right) \underset{\substack{\uparrow \\ \text{Prior mean}}}{\frac{\alpha}{\alpha+\beta}}$$

Theorem: Let $\theta \sim \Lambda$ and $X|\theta \sim P_\theta$ and $L(\theta, d) = (g(\theta) - d)^2$. Then no unbiased estimator $\delta(X)$ can be a Bayes estimator unless $E(\delta(X) - g(\theta))^2 = 0$.

Proof: Suppose $\delta(X)$ is a Bayes estimator and it is unbiased for estimating $g(\theta)$. Then

$$\delta(X) = E(g(\theta)|X)$$

Since $\delta(X)$ is unbiased, $E(\delta(X)|\theta) = g(\theta)$, $\forall \theta$

Conditioning on X,

$$E(g(\theta)\delta(X)) = EE(g(\theta)\delta(X)|X) = E[\delta(X)E(g(\theta)|X)]$$
$$= E\delta^2(X)$$

Conditioning on θ,

$$E(g(\theta)\delta(X)) = EE(g(\theta)\delta(X)|\theta)$$
$$= E[g(\theta)E(\delta(X)|\theta)]$$
$$= E g^2(\theta)$$

It follows that

$$E(\delta(X) - g(\theta))^2 = E\delta^2(X) + E g^2(\theta) - 2E[\delta(X) \cdot g(\theta)]$$
$$= E\delta^2(X) + E g^2(\theta) - E\delta^2(X) - E g^2(\theta)$$
$$= 0$$

Minimax Estimation

Def: An estimator δ is minimax if for any estimator δ'

$$r^*_\delta = \sup_{\theta \in \Theta} R(\theta, \delta) \le \sup_{\theta\in\Theta} R(\theta, \delta') = r^*_{\delta'}$$

That is, δ is minimizing the maximal risk.

The average risk (Bayes risk) of δ_Λ is

$$r_\Lambda = \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta)$$

A prior dist. Λ is said to be least favorable if $r_\Lambda \ge r_{\Lambda'}$, for all prior Λ'.

Example (Minimax Estimation)

Let $X \sim B(1, P)$, $P \in \Theta = \{1/4, \frac{1}{2}\}$ and $\mathcal{D} = \{a_1, a_2\}$ = all decisions open to statistician. Let the loss function be defined as

	a_1	a_2
$P_1 = \frac{1}{4}$	1	4
$P_2 = \frac{1}{2}$	3	2

$P(X=1) = P = 1 - P(X=0)$

The set of decision rules include four functions: d_1, d_2, d_3, d_4 defined by $d_1(0) = d_1(1) = a_1$; $d_2(0) = a_1$, $d_2(1) = a_2$; $d_3(0) = a_2$, $d_3(1) = 1$; $d_4(0) = d_4(1) = a_2$. The risk function takes the following values:

		d_1	d_2	d_3	d_4
X	0	a_1	a_1	a_2	a_2
	1	a_1	a_2	a_1	a_2
$R(P_1, d)$		1	7/4	13/4	4
$R(P_2, d)$		3	5/2	5/2	2
$\max_{P_1,P_2} R(P,d)$		3	5/2	13/4	4
$\min_i \max_{P_1,P_2} R(P,d)$			5/2		

$\Rightarrow$ The minimax sol. is $d_2(x) = a_1$ if $X=0$ & a_2 if $X=1$.

Theorem: Suppose $\theta \sim \Lambda$ such that $\underbrace{r_\Lambda}_{\text{Bayes risk}} = \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta)$

$$= r^*_{\delta_\Lambda} = \underbrace{\sup_{\theta \in \Theta} R(\theta, \delta_\Lambda)}_{\text{maximal risk}}$$

Then

(i) δ_Λ is minimax

(ii) If δ_Λ is the unique Bayes Sol. with respect to Λ, it is unique minimax

(iii) Λ is a least favorable.

Proof:

(i)
$$\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta)\, d\Lambda(\theta)$$
$$\geq \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta)$$
$$= \sup_\theta R(\theta, \delta_\Lambda)$$

(ii) If δ_Λ is unique $\Rightarrow$
$$\int R(\theta, \delta)\, d\Lambda(\theta) > \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta)$$
$$\Rightarrow \quad r_\Lambda > \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta) = r^*_{\delta_\Lambda} = \sup_\theta R(\theta, \delta_\Lambda)$$

$\Rightarrow$ strict inequality assures the uniqueness.

(iii) Let Λ' be some other dist. of θ. Then
$$r_{\Lambda'} = \int R(\theta, \delta_{\Lambda'})\, d\Lambda'(\theta) \leq \int R(\theta, \delta_\Lambda)\, d\Lambda'(\theta)$$
$$= \sup_\theta R(\theta, \delta_\Lambda)$$
$$= r^*_{\delta_\Lambda} = r_\Lambda \quad \text{(assumption)}$$

$\Rightarrow$ Λ is least favorable dist.

<u>COR 1:</u> If δ_Λ has constant risk, then it is minimax ($r_\Lambda = r^*_{\delta_\Lambda}$)

<u>COR 2:</u> Let $\omega_\Lambda = \{\theta : R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda)\}$
Then δ_Λ is minimax if $\Lambda(\omega_\Lambda) = 1$.

<u>Example:</u> $X \sim B(n, p)$, loss is SEL function
Consider the estimate of $p = \hat{p} = \frac{X}{n} \Longrightarrow$ The risk is

$$R(p, \delta) = \frac{pq}{n}, \text{ which has a max. at } p = \tfrac{1}{2}.$$

Define

$$\delta^*(X) = \begin{cases} \frac{X}{n} & \text{with prob. } 1-\varepsilon \\ \frac{1}{2} & \text{" " } \varepsilon \end{cases}, \quad \varepsilon = \frac{1}{n+1}$$

$$E(\delta^* - p)^2 = E\left(\frac{X}{n} - p\right)^2 (1-\varepsilon) + \left(\frac{1}{2} - p\right)^2 \varepsilon$$
$$= \frac{pq}{n} \cdot \frac{n}{n+1} + \left(\frac{1}{2} - p\right)^2 \frac{1}{n+1}$$
$$= \frac{1}{4(n+1)} < \frac{1}{4n} = \sup_{p=\frac{1}{2}} \frac{pq}{n}$$

$\Rightarrow \delta = \frac{X}{n}$ is not minimax

<u>Def.:</u> An estimator δ_Λ is Bayes WRT prior Λ if it minimizes
(*) $\int_{H} R(\theta, \delta_\Lambda) d\Lambda(\theta)$. That is, if δ_0 minimizes (*) $\Longrightarrow$

$\delta_0 = \delta_\Lambda$ a.e. Then δ_Λ is unique Bayes.

<u>Theorem:</u> Any unique Bayes estimator is admissible

Proof: Suppose δ' dominates δ_Λ which is unique Bayes
$\Longrightarrow R(\theta, \delta') \le R(\theta, \delta_\Lambda) \quad \forall \theta$
$\int R(\theta, \delta') d\Lambda(\theta) \le \int R(\theta, \delta_\Lambda) d\Lambda(\theta)$
By uniqueness, $R(\theta, \delta') = R(\theta, \delta_\Lambda)$, $\forall \theta \Longrightarrow \delta'$ does not dominate δ_Λ.

Admissibility of Linear Estimators of $N(\theta,1)$

$X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$

Let Λ be a normal prior on θ, $\theta \sim N(\mu, \sigma^2)$

$$\lambda(\theta \mid \underset{\sim}{x}) \propto f(\underset{\sim}{x} \mid \theta)\, \lambda(\theta)$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum(x_i-\theta)^2} \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2\sigma^2}(\theta-\mu)^2}$$

$$\propto e^{-\frac{1}{2}\sum x_i^2 + \theta \sum x_i - \frac{n}{2}\theta^2 - \frac{1}{2\sigma^2}\theta^2 + \frac{\mu}{\sigma^2}\theta - \frac{\mu^2}{2\sigma^2}}$$

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

$$= e^{-\frac{1}{2}\left(n+\frac{1}{\sigma^2}\right)\left[\theta - \left(\frac{\sum x_i + \frac{\mu}{\sigma^2}}{n + \frac{1}{\sigma^2}}\right)\right]}$$

$$\Rightarrow \theta \mid \underset{\sim}{x} \sim N\left(\frac{\sum x_i + \frac{\mu}{\sigma^2}}{n + \frac{1}{\sigma^2}}, \frac{1}{n+\frac{1}{\sigma^2}}\right)$$

$$\Rightarrow \delta_\Lambda(x) = \left(\frac{n}{n+\frac{1}{\sigma^2}}\right)\bar{X} + \left(\frac{\frac{1}{\sigma^2}}{n+\frac{1}{\sigma^2}}\right)\mu$$

$$= a\bar{X} + b$$

where $a = \dfrac{n}{n+\frac{1}{\sigma^2}}$, $b = \dfrac{\mu}{n\sigma^2+1}$

$\Rightarrow$ $\delta_\Lambda(x) = a\bar{X} + b$ is a unique Bayes estimator for $0 < a < 1$

$\Rightarrow$ It is admissible.

<u>Notes</u>: Let $X \sim P_\theta$ with $EX = \theta$, $Var(X) = \sigma^2$.

(1) If $a = 0$, $\delta_\Lambda = b$; which is trivially admissible

(2) $\delta_{a,b}$ is inadmissible if (i) $a < 0$ (ii) $a > 1$ (iii) $a = 1$ & $b \neq 0$

$$R(\theta, \delta_{a,b}) = E[(aX+b)-\theta]^2$$
$$= E\left[a(X-\theta)+(a-1)\theta+b\right]^2$$
$$= a^2\sigma^2 + \left[(a-1)\theta+b\right]^2$$

(i) If $a<0$, $(a-1)^2>1 \Rightarrow R(\theta, \delta_{a,b}) \geq \left[(a-1)\theta+b\right]^2$
$$= (a-1)^2\left[\theta+\frac{b}{a-1}\right]^2$$
$$> \left(\theta+\frac{b}{a-1}\right)^2 = R(\theta, \delta_{0,\frac{-b}{a-1}})$$

Thus, $aX+b$ is dominated by constant est.
$$\Rightarrow \delta = \frac{-b}{a-1}$$

(ii) If $a>1$, then $R(\theta, \delta_{a,b}) \geq a^2\sigma^2 > \sigma^2 = R(\theta, X)$
$\Rightarrow \delta_{a,b}$ is inadmissible

(iii) If $a=1$, $R(\theta, \delta_{a,b}) = \sigma^2 + b^2 > \sigma^2 = R(\theta, X)$
$\Rightarrow \delta_{a,b}$ is inadmissible.

Lemma: If δ_Λ is the Bayes estimator WRT Λ and $r_\Lambda = E(\delta_\Lambda(X)-g(\theta))^2$ is the Bayes risk. Then

(i) $r_\Lambda = \int_X Var(g(\theta)|X)\, dP(x)$

(ii) If $Var(g(\theta)|X)$ does not depend on X, then $r_\Lambda = Var(g(\theta)|X)$

Proof: (i) $r_\Lambda = \int_{\Theta}\int_X (\delta_\Lambda(x)-g(\theta))^2 dP(x)\, d\Lambda(\theta) = \int_X\int_{\Theta}(\delta_\Lambda(x)-g(\theta))^2 d\Lambda\, dP$
$$= \int_X Var(g(\theta)|X)\, dP(x)$$

(ii) If $Var(g(\theta)|X)$ does not depend on X,
$$r_\Lambda = Var(g(\theta)|X)\int_X dP(x) = Var(g(\theta)|X).$$