Stat 710: Mathematical Statistics Lecture 2

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Definition 4.1 (Bayes action)

Let $\mathscr{A}$ be an action space in a decision problem and $L(\theta, a) \geq 0$ be a loss function

For any $x \in \mathscr{X}$, a *Bayes action* w.r.t. Π is any $\delta(x) \in \mathscr{A}$ such that

$$E[L(\vec{\theta}, \delta(x))|X = x] = \min_{a \in \mathscr{A}} E[L(\vec{\theta}, a)|X = x]$$

where the expectation is w.r.t. the posterior distribution $P_{\theta|x}$

Remarks

- The Bayes action minimizes the posterior expected loss
- x is fixed, although $\delta(x)$ depends on x
- The Bayes action depends on the prior
- The Bayes action depends on the loss function

Definition 4.1 (Bayes action)

Let $\mathscr{A}$ be an action space in a decision problem and $L(\theta, a) \geq 0$ be a loss function

For any $x \in \mathscr{X}$, a *Bayes action* w.r.t. Π is any $\delta(x) \in \mathscr{A}$ such that

$$E[L(\vec{\theta}, \delta(x))|X = x] = \min_{a \in \mathscr{A}} E[L(\vec{\theta}, a)|X = x]$$

where the expectation is w.r.t. the posterior distribution $P_{\theta|x}$

Remarks

- The Bayes action minimizes the posterior expected loss
- x is fixed, although $\delta(x)$ depends on x
- The Bayes action depends on the prior
- The Bayes action depends on the loss function

Proposition 4.1 (Existence and uniqueness of Bayes actions)

Assume the conditions in Theorem 4.1; $L(\theta,a)$ is convex in a for each fixed θ; for each $x \in \mathscr{X}$, $E[L(\vec{\theta},a)|X=x] < \infty$ for some a.
(i) If $\mathscr{A} \subset \mathscr{R}^p$ is compact, then a Bayes action exists for each $x \in \mathscr{X}$.
(ii) If $\mathscr{A} = \mathscr{R}^p$ and $L(\theta,a)$ tends to ∞ as $\|a\| \to \infty$ uniformly in $\theta \in \Theta_0 \subset \Theta$ with $\Pi(\Theta_0) > 0$, then a Bayes action exists for each $x \in \mathscr{X}$.
(iii) In (i) or (ii), if $L(\theta,a)$ is strictly convex in a for each fixed θ, then the Bayes action is unique.

Proof

The convexity of L implies that $E[L(\vec{\theta},a)|X=x]$ as a function of a with any fixed x is convex and continuous. Result (i) follows from the fact that any continuous function on a compact set attains its minimum. Result (ii) follows from the fact that

$$\lim_{\|a\|\to\infty} E[L(\vec{\theta},a)|X=x] \geq \lim_{\|a\|\to\infty} \int_{\Theta_0} L(\theta,a)dP_{\theta|x} = \infty$$

Result (iii) holds because $E[L(\vec{\theta},a)|X=x]$ is strictly convex in a

Proposition 4.1 (Existence and uniqueness of Bayes actions)

Assume the conditions in Theorem 4.1; $L(\theta,a)$ is convex in a for each fixed θ; for each $x \in \mathscr{X}$, $E[L(\vec{\theta},a)|X=x] < \infty$ for some a.
(i) If $\mathscr{A} \subset \mathscr{R}^p$ is compact, then a Bayes action exists for each $x \in \mathscr{X}$.
(ii) If $\mathscr{A} = \mathscr{R}^p$ and $L(\theta,a)$ tends to ∞ as $\|a\| \to \infty$ uniformly in $\theta \in \Theta_0 \subset \Theta$ with $\Pi(\Theta_0) > 0$, then a Bayes action exists for each $x \in \mathscr{X}$.
(iii) In (i) or (ii), if $L(\theta,a)$ is strictly convex in a for each fixed θ, then the Bayes action is unique.

Proof

The convexity of L implies that $E[L(\vec{\theta},a)|X=x]$ as a function of a with any fixed x is convex and continuous. Result (i) follows from the fact that any continuous function on a compact set attains its minimum. Result (ii) follows from the fact that

$$\lim_{\|a\|\to\infty} E[L(\vec{\theta},a)|X=x] \geq \lim_{\|a\|\to\infty} \int_{\Theta_0} L(\theta,a)dP_{\theta|x} = \infty$$

Result (iii) holds because $E[L(\vec{\theta},a)|X=x]$ is strictly convex in a

Example 4.1: the estimation of $\vartheta = g(\theta)$

Assume $\int_\Theta [g(\theta)]^2 d\Pi < \infty$, $\mathscr{A} =$ the range of $g(\theta)$, and
$L(\theta, a) = [g(\theta) - a]^2$ (squared error loss).
Using the argument in Example 1.22, we obtain the Bayes action

$$\delta(x) = \frac{\int_\Theta g(\theta) f_\theta(x) d\Pi}{m(x)} = \frac{\int_\Theta g(\theta) f_\theta(x) d\Pi}{\int_\Theta f_\theta(x) d\Pi},$$

which is the posterior expectation of $g(\vec{\theta})$, given $X = x$.

A more specific case

$g(\theta) = \theta^j$ for some integer $j \geq 1$
$f_\theta(x) = e^{-\theta}\theta^x I_{\{0,1,2,...\}}(x)/x!$ (the Poisson distribution) with $\theta > 0$
Π has a Lebesgue p.d.f. $\pi(\theta) = \theta^{\alpha-1}e^{-\theta/\gamma}I_{(0,\infty)}(\theta)/[\Gamma(\alpha)\gamma^\alpha]$
(the gamma distribution $\Gamma(\alpha, \gamma)$ with known $\alpha > 0$ and $\gamma > 0$)
Then, for $x = 0, 1, 2, ...$, and some function $c(x)$,

$$f_\theta(x)\pi(\theta)/m(x) = c(x)\theta^{x+\alpha-1}e^{-\theta(\gamma+1)/\gamma}I_{(0,\infty)}(\theta),$$

This is the gamma distribution $\Gamma(x+\alpha, \gamma/(\gamma+1))$.

Example 4.1: the estimation of $\vartheta = g(\theta)$

Assume $\int_\Theta [g(\theta)]^2 d\Pi < \infty$, $\mathscr{A}$ = the range of $g(\theta)$, and $L(\theta, a) = [g(\theta) - a]^2$ (squared error loss).
Using the argument in Example 1.22, we obtain the Bayes action

$$\delta(x) = \frac{\int_\Theta g(\theta) f_\theta(x) d\Pi}{m(x)} = \frac{\int_\Theta g(\theta) f_\theta(x) d\Pi}{\int_\Theta f_\theta(x) d\Pi},$$

which is the posterior expectation of $g(\vec{\theta})$, given $X = x$.

A more specific case

$g(\theta) = \theta^j$ for some integer $j \geq 1$
$f_\theta(x) = e^{-\theta}\theta^x I_{\{0,1,2,\ldots\}}(x)/x!$ (the Poisson distribution) with $\theta > 0$
Π has a Lebesgue p.d.f. $\pi(\theta) = \theta^{\alpha-1} e^{-\theta/\gamma} I_{(0,\infty)}(\theta)/[\Gamma(\alpha)\gamma^\alpha]$
(the gamma distribution $\Gamma(\alpha, \gamma)$ with known $\alpha > 0$ and $\gamma > 0$)
Then, for $x = 0, 1, 2, \ldots$, and some function $c(x)$,

$$f_\theta(x)\pi(\theta)/m(x) = c(x)\theta^{x+\alpha-1} e^{-\theta(\gamma+1)/\gamma} I_{(0,\infty)}(\theta),$$

This is the gamma distribution $\Gamma(x+\alpha, \gamma/(\gamma+1))$.

Without actually working out the integral $m(x)$, we know that

$$c(x) = (1+\gamma^{-1})^{x+\alpha}/\Gamma(x+\alpha),$$

$$\delta(x) = c(x)\int_0^\infty \theta^{j+x+\alpha-1}e^{-\theta(\gamma+1)/\gamma}d\theta.$$

The integrand is proportional to the p.d.f. of the gamma distribution $\Gamma(j+x+\alpha, \gamma/(\gamma+1))$.
Hence

$$\begin{aligned}\delta(x) &= c(x)\Gamma(j+x+\alpha)/(1+\gamma^{-1})^{j+x+\alpha} \\ &= (j+x+\alpha-1)\cdots(x+\alpha)/(1+\gamma^{-1})^j.\end{aligned}$$

In particular, $\delta(x) = (x+\alpha)\gamma/(\gamma+1)$ when $j=1$.

Conjugate prior

An interesting phenomenon is that the prior and the posterior are in the same parametric family of distributions.
Such a prior is called a *conjugate* prior.

Without actually working out the integral $m(x)$, we know that

$$c(x) = (1+\gamma^{-1})^{x+\alpha}/\Gamma(x+\alpha),$$

$$\delta(x) = c(x) \int_0^\infty \theta^{j+x+\alpha-1} e^{-\theta(\gamma+1)/\gamma} d\theta.$$

The integrand is proportional to the p.d.f. of the gamma distribution $\Gamma(j+x+\alpha, \gamma/(\gamma+1))$.
Hence

$$\begin{aligned} \delta(x) &= c(x)\Gamma(j+x+\alpha)/(1+\gamma^{-1})^{j+x+\alpha} \\ &= (j+x+\alpha-1)\cdots(x+\alpha)/(1+\gamma^{-1})^j. \end{aligned}$$

In particular, $\delta(x) = (x+\alpha)\gamma/(\gamma+1)$ when $j=1$.

Conjugate prior

An interesting phenomenon is that the prior and the posterior are in the same parametric family of distributions.
Such a prior is called a *conjugate* prior.

Remarks

- Whether a prior is conjugate involves a pair of families; one is the family $\mathscr{P} = \{f_\theta : \theta \in \Theta\}$ and the other is the family from which Π is chosen.
- Example 4.1 shows that the Poisson family and the gamma family produce conjugate priors.
- Many pairs of families in Table 1.1 (page 18) and Table 1.2 (pages 20-21) produce conjugate priors.
- Under a conjugate prior, Bayes actions often have explicit forms (in x) when the loss function is simple.
- Even under a conjugate prior, the integral in $\delta(x)$ in Example 4.1 involving a general g may not have an explicit form.
- In general, numerical methods have to be used in evaluating the integrals in $\delta(x)$ under general loss functions.
- More discussions on the computation of Bayes actions are given in §4.1.4.

Generalized Bayes action

The minimization in Definition 4.1 is the same as the minimizing

$$\int_{\Theta} L(\theta, \delta(x)) f_\theta(x) d\Pi = \min_{a \in \mathscr{A}} \int_{\Theta} L(\theta, a) f_\theta(x) d\Pi$$

$\delta(x)$ is called a *generalized Bayes action*.

This is still defined even if Π is not a probability measure but a σ-finite measure on Θ, in which case $m(x)$ may not be finite.

If $\Pi(\Theta) \neq 1$, Π is called an *improper prior*.

A prior with $\Pi(\Theta) = 1$ is then called a proper prior.

The following is a reason why we need to discuss improper priors and generalized Bayes actions.

In many cases, one has no past information and has to choose a prior subjectively.

In such cases, one would like to select a *noninformative* prior that tries to treat all parameter values in Θ equitably.

A noninformative prior is often improper.

Example 4.3

Suppose that $X = (X_1, ..., X_n)$ and X_i's are i.i.d. from $N(\mu, \sigma^2)$, where $\mu \in \Theta \subset \mathscr{R}$ is unknown and σ^2 is known.
Consider the estimation of $\vartheta = \mu$ under the squared error loss.
If $\Theta = [a, b]$ with $-\infty < a < b < \infty$, then a noninformative prior that treats all parameter values equitably is the uniform distribution on $[a, b]$.
If $\Theta = \mathscr{R}$, however, the corresponding "uniform distribution" is the Lebesgue measure on $\mathscr{R}$, which is an improper prior.
If Π is the Lebesgue measure on $\mathscr{R}$, then

$$(2\pi\sigma^2)^{-n/2} \int_{-\infty}^{\infty} \mu^2 \exp\left\{ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\} d\mu < \infty.$$

By differentiating a in

$$(2\pi\sigma^2)^{-n/2} \int_{-\infty}^{\infty} (\mu - a)^2 \exp\left\{ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\} d\mu$$

Example 4.3 (continued)

and using the fact that $\sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$, where $\bar{x}$ is the sample mean of the observations $x_1,\dots,x_n$, we obtain that

$$\delta(x) = \frac{\int_{-\infty}^{\infty} \mu \exp\left\{-n(\bar{x}-\mu)^2/(2\sigma^2)\right\} d\mu}{\int_{-\infty}^{\infty} \exp\left\{-n(\bar{x}-\mu)^2/(2\sigma^2)\right\} d\mu} = \bar{x}.$$

Thus, the sample mean is a generalized Bayes action under the squared error loss.

From Example 2.25 and Exercise 91 in §2.6, if Π is $N(\mu_0,\sigma_0^2)$, then the Bayes action is

$$\mu_*(x) = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} \qquad \text{and} \qquad c^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}$$

Note that in this case $\bar{x}$ is a limit of $\mu_*(x)$ as $\sigma_0^2 \to \infty$.

More detailed discussions of the use of improper priors, see Jeffreys (1939, 1948, 1961), Box and Tiao (1973), and Berger (1985).

Example 4.3 (continued)

and using the fact that $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$, where $\bar{x}$ is the sample mean of the observations $x_1, ..., x_n$, we obtain that

$$\delta(x) = \frac{\int_{-\infty}^{\infty} \mu \exp\left\{-n(\bar{x}-\mu)^2/(2\sigma^2)\right\} d\mu}{\int_{-\infty}^{\infty} \exp\left\{-n(\bar{x}-\mu)^2/(2\sigma^2)\right\} d\mu} = \bar{x}.$$

Thus, the sample mean is a generalized Bayes action under the squared error loss.

From Example 2.25 and Exercise 91 in §2.6, if Π is $N(\mu_0, \sigma_0^2)$, then the Bayes action is

$$\mu_*(x) = \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\bar{x} \qquad \text{and} \qquad c^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}$$

Note that in this case $\bar{x}$ is a limit of $\mu_*(x)$ as $\sigma_0^2 \to \infty$.

More detailed discussions of the use of improper priors, see Jeffreys (1939, 1948, 1961), Box and Tiao (1973), and Berger (1985).