# Chapter 4. Bayesian Analysis (C2)

## ♠ Bayesian Decision Theory

### ♢ Posterior Decision Analysis

**• Posterior Expected Loss and Bayes Action**

The *posterior expected loss* of an action $a$, when the posterior distribution is $\pi(\theta|\boldsymbol{x})$ is

$$\rho(\pi(\theta|\boldsymbol{x}), a) = \int_{\Theta} L(\theta, a) dF^{\pi(\theta|\boldsymbol{x})}(\theta).$$

A (posterior) *Bayes action*, to be denoted by $\delta^{\pi}(\boldsymbol{x})$ is any action $a \in \mathcal{A}$ which minimizes $\rho(\pi(\theta|\boldsymbol{x}), a)$, or equivalently which minimizes

$$\int_{\Theta} L(\theta, a) f(\boldsymbol{x}|\theta) dF^{\pi}(\theta).$$

• **Bayes Rule**: A Bayes rule $\delta^{\pi}$ minimizes the Bayes risk $r(\pi, \delta)$.

• **Result 1**: *A Bayes rule $\delta^\pi$ can be found by choosing an action which minimizes the posterior expected loss.*

• **Note**: The Bayes rule $\delta^\pi$ need not be unique. When $m(\boldsymbol{x}) = 0$, $\delta^\pi$ can be defined arbitrarily. Furthermore, if $r(\pi, \delta) = \infty$ for all $\delta$, then any decision rule is a Bayes rule.

- **Result 2**: *If $\delta$ is a nonrandomized estimator, then*

$$r(\pi, \delta) = \int_{\{\boldsymbol{x}:\, m(\boldsymbol{x})>0\}} \pi(\pi(\theta|\boldsymbol{x}), \delta(\boldsymbol{x})) dF^m(\boldsymbol{x}).$$

**Proof**: By definition,

$$\begin{aligned} r(\pi, \rho) &= \int_\Theta R(\theta, \delta) dF^\pi(\theta) \\ &= \int_\Theta \int_\mathcal{X} L(\theta, \delta(\boldsymbol{x})) dF^{\boldsymbol{X}|\theta}(\boldsymbol{x}) dF^\pi(\theta). \end{aligned}$$

Since $L(\theta, \delta) \geq -K > -\infty$ and all measures above are finite, Fubini's theorem can be employed to interchange orders of integration and obtain

$$r(\pi, \delta) = \begin{cases} \int_\mathcal{X} \left[\int_\Theta L(\theta, \delta(\boldsymbol{x})) f(\boldsymbol{x}|\theta) dF^\pi(\theta)\right] d\boldsymbol{x}, \\ \sum_\mathcal{X} \left[\int_\Theta L(\theta, \delta(\boldsymbol{x})) f(\boldsymbol{x}|\theta) dF^\pi(\theta)\right], \end{cases}$$

in the cases of continuous and discrete $\mathcal{X}$, respectively. Finally, noting that, if $m(\boldsymbol{x}) = 0$, then $f(\boldsymbol{x}|\theta) = 0$ almost everywhere with respect to $\pi(\theta)$, the definition of $\pi(\theta|\boldsymbol{x})$ and $\rho(\pi(\theta|\boldsymbol{x}), \delta)$ yield the result.

• **Generalized Bayes Rule**:

If $\pi$ is an improper prior, but $\delta^\pi$ is an action which minimizes

$$\rho(\pi(\theta|\boldsymbol{x}), a) = \int_\Theta L(\theta, a) dF^{\pi(\theta|\boldsymbol{x})}(\theta)$$

for each $\boldsymbol{x}$ with $m(\boldsymbol{x}) > 0$, then $\delta^\pi$ is called a generalized Bayes rule.

◇ **Estimation**

• **Result 3**: *If $L(\theta, a) = (\theta - a)^2$, the Bayes rule is*

$$\delta^\pi(\boldsymbol{x}) = E^{\pi(\theta|\boldsymbol{x})}[\theta],$$

*which is the posterior mean of $\theta$.*

• **Result 4**: *If* $L(\theta, a) = w(\theta)(\theta - a)^2$, *the Bayes rule is*

$$\delta^\pi(\boldsymbol{x}) = \frac{E^{\pi(\theta|\boldsymbol{x})}[\theta w(\theta)]}{E^{\pi(\theta|\boldsymbol{x})}[w(\theta)]}.$$

**Proof**: Since

$$\rho(\pi(\theta|\boldsymbol{x}), a) = \int_\Theta w(\theta)(\theta - a)^2 dF^{\pi(\theta|\boldsymbol{x})}(\theta),$$

we set

$$\frac{d\rho(\pi(\theta|\boldsymbol{x}), a)}{da} = -2\int_\Theta w(\theta)(\theta - a) dF^{\pi(\theta|\boldsymbol{x})}(\theta) = 0$$

and the minimum of $\rho(\pi(\theta|\boldsymbol{x}), a)$ is obtained at

$$\begin{aligned}\delta^\pi(\boldsymbol{x}) &= \frac{\int_\Theta w(\theta)\theta dF^{\pi(\theta|\boldsymbol{x})}(\theta)}{\int_\Theta w(\theta) dF^{\pi(\theta|\boldsymbol{x})}(\theta)} \\ &= \frac{E^{\pi(\theta|\boldsymbol{x})}[\theta w(\theta)]}{E^{\pi(\theta|\boldsymbol{x})}[w(\theta)]}.\end{aligned}$$

• **Example 1**: Assume $X \sim N(\theta, \sigma^2)$, where $\theta$ is unknown and $\sigma^2$ is known. Let $\pi(\theta)$ be a $N(\mu, \tau^2)$ density. Then

$$\theta | x \sim N(\mu(x), 1/\rho),$$

where

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}x = x - \frac{\sigma^2}{\sigma^2 + \tau^2}(x - \mu),$$

and $\rho = \text{precision} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$.

Thus, the Bayes rule under the squared-error loss is

$$\delta^\pi(x) = \frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}x = x - \frac{\sigma^2}{\sigma^2 + \tau^2}(x - \mu).$$

If we let $\tau^2 \to \infty$, i.e., we use $\pi(\theta) = 1$, then the generalized Bayes rule under the squared-error loss is

$$\delta^\pi(x) = x.$$

• **Example 2**: Assume $X \sim \text{Bin}(n, \theta)$, $\theta \sim \mathcal{B}e(\alpha, \beta)$ and $w(\theta) = [\theta(1-\theta)]^{-1}$. Then, we have

$$\theta | x \sim \mathcal{B}e(\alpha + x, \beta + n - x)$$

and the Bayes rule under the weighted squared-error loss is

$$\begin{aligned}
\delta^\pi(x) = & \frac{E^{\pi(\theta|x)}[(1-\theta)^{-1}]}{E^{\pi(\theta|x)}[\theta^{-1}(1-\theta)^{-1}]} \\
= & \frac{\int_0^1 \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-2} d\theta}{\int_0^1 \theta^{\alpha+x-2}(1-\theta)^{\beta+n-x-2} d\theta} \\
= & \frac{B(\alpha + x, \beta + n - x - 1)}{B(\alpha + x - 1, \beta + n - x - 1)} \\
= & \frac{\alpha + x - 1}{\alpha + \beta + n - 2}.
\end{aligned}$$

What is the Bayes rule under the squared-error loss?

- **Bayes rule under the quadratic loss**:

For the quadratic loss $L(\boldsymbol{\theta}, \boldsymbol{a}) = (\boldsymbol{\theta} - \boldsymbol{a})'Q(\boldsymbol{\theta} - \boldsymbol{a})$ ($\boldsymbol{\theta}$ and $\boldsymbol{a}$ are now vectors and $Q$ is a positive definite matrix), then the Bayes rule (or Bayes estimator) is the posterior mean

$$\boldsymbol{\delta}^{\pi}(\boldsymbol{x}) = E^{\pi(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}].$$

- **Result 5**: *If $L(\theta, a) = |\theta - a|$, any median of $\pi(\theta|\boldsymbol{x})$ is a Bayes estimator of $\theta$.*

**Proof**: Let $m$ denote a median of $\pi(\theta|\boldsymbol{x})$, and let $a > m$ be another action. Note that

$$\begin{aligned} & L(\theta, m) - L(\theta, a) \\ &= \begin{cases} m - a & \text{if } \theta \le m, \\ 2\theta - (m + a) & \text{if } m < \theta < a, \\ a - m & \text{if } \theta \ge a. \end{cases} \end{aligned}$$

It follows that

$$L(\theta, m) - L(\theta, a) \le (m - a)I_{(-\infty, m]}(\theta) + (a - m)I_{(m, \infty)}(\theta).$$

Since $P(\theta \le m|\boldsymbol{x}) \ge \frac{1}{2}$, so that $P(\theta > m|\boldsymbol{x}) \le \frac{1}{2}$, it can be concluded that

$$
\begin{aligned}
&E^{\pi(\theta|\boldsymbol{x})}[L(\theta, m) - L(\theta, a)] \\
\leq&(m-a)P(\theta \leq m|\boldsymbol{x}) + (a-m)P(\theta > m|\boldsymbol{x}) \\
\leq&(m-a)\frac{1}{2} + (a-m)\frac{1}{2} = 0,
\end{aligned}
$$

establishing that $m$ has posterior expected loss at least as small as $a$. A similar argument holds for $a < m$.

- **Result 6**: *If*

$$
L(\theta, a) = \begin{cases} K_0(\theta - a) & \text{if } \theta - a \geq 0, \\ K_1(a - \theta) & \text{if } \theta - a < 0, \end{cases}
$$

*any* $(K_0/(K_0 + K_1))$*-fractile of* $\pi(\theta|\boldsymbol{x})$ *is a Bayes estimate of* $\theta$.

## $\diamondsuit$ **Hypothesis Testing**

Consider $H_0$: $\theta \in \Theta_0$ versus $H_1$: $\theta \in \Theta_1$.

Let $a_i$ =acceptance of $H_i$ for $i = 0, 1$ and

$$L(\theta, a_i) = \begin{cases} 0, & \text{if } \theta \in \Theta_i, \\ 1, & \text{if } \theta \in \Theta_j,\ j \neq i. \end{cases}$$

Then,

$$\begin{aligned} E^{\pi(\theta|\boldsymbol{x})}[L(\theta, a_1)] &= \int_{\Theta} L(\theta, a_1) dF^{\pi(\theta|\boldsymbol{x})}(\theta) \\ &= \int_{\Theta_0} dF^{\pi(\theta|\boldsymbol{x})}(\theta) = P(\Theta_0|\boldsymbol{x}). \end{aligned}$$

Similarly,

$$E^{\pi(\theta|\boldsymbol{x})}[L(\theta, a_0)] = P(\Theta_1|\boldsymbol{x}).$$

Hence the Bayes decision is simply the hypothesis with the larger posterior probability.

- **“$0 - K_i$” Loss**

Let

$$L(\theta, a_i) = \begin{cases} 0, & \text{if } \theta \in \Theta_i, \\ K_i, & \text{if } \theta \in \Theta_j,\ j \neq i. \end{cases}$$

Then, the posterior expected loss is

$$E^{\pi(\theta|\boldsymbol{x})}[L(\theta, a_1)] = K_1 P(\Theta_0|\boldsymbol{x}),$$

and

$$E^{\pi(\theta|\boldsymbol{x})}[L(\theta, a_0)] = K_0 P(\Theta_1|\boldsymbol{x}).$$

Therefore, in the Bayesian test, the null hypothesis $H_0$ is rejected (i.e., $a_1$ is taken) when

$$\frac{K_0}{K_1} > \frac{P(\Theta_0|\boldsymbol{x})}{P(\Theta_1|\boldsymbol{x})}.$$

Usually $\Theta_0 \cup \Theta_1 = \Theta$, in which case

$$P(\Theta_0|\boldsymbol{x}) = 1 - P(\Theta_1|\boldsymbol{x}).$$

It follows that

$$\frac{K_0}{K_1} > \frac{P(\Theta_0|\boldsymbol{x})}{P(\Theta_1|\boldsymbol{x})} = \frac{1 - P(\Theta_1|\boldsymbol{x})}{P(\Theta_1|\boldsymbol{x})} = \frac{1}{P(\Theta_1|\boldsymbol{x})} - 1,$$

or

$$P(\Theta_1|\boldsymbol{x}) > \frac{K_1}{K_0 + K_1}.$$

Thus in classical terminology, the rejection region of the Bayesian test is

$$C = \left\{ \boldsymbol{x} : \; P(\Theta_1|\boldsymbol{x}) > \frac{K_1}{K_0 + K_1} \right\}.$$

Typically, $C$ is of exactly the same form as the rejection region of a classical (say likelihood ratio) test.

• **Example 3**: Assume $X \sim N(\theta, \sigma^2)$, where $\theta$ is unknown and $\sigma^2$ is known. Let $\pi(\theta)$ be a $N(\mu, \tau^2)$ density. Then

$$\theta|x \sim N(\mu(x), 1/\rho),$$

where

$$\mu(x) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x = x - \frac{\sigma^2}{\sigma^2+\tau^2}(x-\mu),$$

and $\rho = \text{precision} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$.

Assume that it is desired to test $H_0$: $\theta \geq \theta_0$ versus $H_1$: $\theta < \theta_0$ under "$0 - K_i$" loss.

Then, the Bayes test rejects $H_0$ if

$$\begin{aligned}\frac{K_1}{K_0+K_1} &< P(\Theta_1|\boldsymbol{x}) \\ &= \left(\frac{\rho}{2\pi}\right)^{1/2}\int_{-\infty}^{\theta_0}\exp\left\{-\frac{\rho(\theta-\mu(x))^2}{2}\right\}d\theta \\ &=(2\pi)^{-1/2}\int_{-\infty}^{\rho^{1/2}(\theta_0-\mu(x))}\exp\left\{-\frac{\eta^2}{2}\right\}d\eta.\end{aligned}$$

Letting $z(\alpha)$ denote the $\alpha$-fractile of a $N(0,1)$ distribution, it follows that the Bayes test rejects $H_0$ if

$$\rho^{1/2}(\theta_0-\mu(x)) > z\left(\frac{K_1}{K_0+K_1}\right),$$

i.e.,

$$x < \theta_0 + \frac{\sigma^2}{\tau^2}(\theta_0-\mu) - \sigma^2\rho^{1/2}z\left(\frac{K_1}{K_0+K_1}\right).$$

The classical uniformly most powerful size $\alpha$ tests are of the same form, rejecting $H_0$ when

$$x < \theta_0 + \sigma z(\alpha).$$

Thus, the Bayes test with $\alpha = \frac{K_1}{K_0+K_1}$ and $\pi(\theta)=1$ (i.e., $\tau^2\to\infty$) is identical to the classical UMP test.

♠ **A Practice Problem**:

*Show that if* $P_\theta(\delta^\pi \neq \theta) = 1$*, then no unbiased estimator can be Bayes under the squared-error loss.*