

Lecture 9: Bayes estimators under squared error, weighted squared error and absolute error loss

Statistical Inference–II ■ Unit I: Decision Theory

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

Previous lecture

Lecture 8: Structure of the Bayes rule: posterior expected loss.

Today

- The Bayes Principle
 - Bayes Rules under Standard Loss Functions

(i) Squared error loss

The Structure Theorem reduces the determination of a Bayes rule to a calculus problem for each fixed $\mathbf{x}$. We now carry this out for the most important loss functions in estimation: squared error (and its weighted form), absolute error, small-neighbourhood 0–1 loss and LINEX loss.

Corollary 9.1 (Bayes estimator under SELF is the posterior mean)

Under $L(\theta, a) = (a - \theta)^2$, the Bayes estimator is the posterior mean

$$\delta_g(\mathbf{x}) = \mathbb{E}[\theta \mid \mathbf{X} = \mathbf{x}] = \int_{\Theta} \theta \pi(\theta \mid \mathbf{x}) d\theta,$$

provided $\mathbb{E}(\theta^2 \mid \mathbf{x}) < \infty$ (so that the posterior mean and the posterior variance exist). The minimum posterior expected loss at $\mathbf{x}$ is the posterior variance $\text{Var}(\theta \mid \mathbf{x})$, and the Bayes risk of the problem is

$$r(g) = \mathbb{E}[\text{Var}(\theta \mid \mathbf{X})] = \int_{\mathcal{X}} \text{Var}(\theta \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}.$$

Proof.

For fixed $\mathbf{x}$, put $h(a) = \mathbb{E}[(a - \theta)^2 \mid \mathbf{x}]$. Expanding,

$$h(a) = a^2 - 2a \mathbb{E}(\theta \mid \mathbf{x}) + \mathbb{E}(\theta^2 \mid \mathbf{x}).$$

Then $h'(a) = 2a - 2\mathbb{E}(\theta \mid \mathbf{x})$, which vanishes at $a = \mathbb{E}(\theta \mid \mathbf{x})$, and $h''(a) = 2 > 0$, confirming a minimum. At this point

$$h(\mathbb{E}(\theta \mid \mathbf{x})) = \mathbb{E}(\theta^2 \mid \mathbf{x}) - \{\mathbb{E}(\theta \mid \mathbf{x})\}^2 = \text{Var}(\theta \mid \mathbf{x}).$$

The expression for $r(g)$ follows by substituting into

$$r(g, \delta_g) = \int_{\mathcal{X}} \rho(\pi, \delta_g(\mathbf{x}) \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x},$$

which was established in the proof of the Structure Theorem. □

Remark 9.2 (A common error)

It is *not* correct to say that “the posterior variance is the Bayes risk”.

- The posterior variance $\text{Var}(\theta \mid \mathbf{x})$ is a function of $\mathbf{x}$; it is the *posterior expected loss* at $\mathbf{x}$, that is, the conditional Bayes risk.
- The Bayes risk of the problem is its *expectation over the marginal distribution of $\mathbf{X}$* .

The two coincide only in the degenerate case where $\text{Var}(\theta \mid \mathbf{x})$ does not depend on $\mathbf{x}$ — which does happen, for instance, in the normal–normal model (example below), but not in the beta–binomial model (example below).

Corollary 9.3 (Bayes estimator under weighted squared error loss)

Let $L(\theta, a) = w(\theta)(a - \theta)^2$ with $w(\theta) > 0$, and suppose that $\mathbb{E}\{w(\theta) \mid \mathbf{x}\}$ and $\mathbb{E}\{w(\theta)\theta^2 \mid \mathbf{x}\}$ are finite. Then the Bayes estimator is

$$\delta_g(\mathbf{x}) = \frac{\mathbb{E}\{w(\theta) \theta \mid \mathbf{x}\}}{\mathbb{E}\{w(\theta) \mid \mathbf{x}\}} = \frac{\int_{\Theta} w(\theta) \theta \pi(\theta \mid \mathbf{x}) d\theta}{\int_{\Theta} w(\theta) \pi(\theta \mid \mathbf{x}) d\theta}.$$

Proof.

For fixed $\mathbf{x}$ put

$$\begin{aligned} h(a) &= \mathbb{E}\{w(\theta)(a - \theta)^2 \mid \mathbf{x}\} \\ &= a^2 \mathbb{E}\{w(\theta) \mid \mathbf{x}\} - 2a \mathbb{E}\{w(\theta)\theta \mid \mathbf{x}\} + \mathbb{E}\{w(\theta)\theta^2 \mid \mathbf{x}\}. \end{aligned}$$

Then

$$h'(a) = 2a \mathbb{E}\{w(\theta) \mid \mathbf{x}\} - 2 \mathbb{E}\{w(\theta)\theta \mid \mathbf{x}\},$$

which vanishes at the stated value, and $h''(a) = 2 \mathbb{E}\{w(\theta) \mid \mathbf{x}\} > 0$ confirms a minimum. □

Remark 9.4

Equivalently, $\delta_g(\mathbf{x})$ is the mean of the *weighted posterior* $\pi_w(\theta \mid \mathbf{x}) \propto w(\theta) \pi(\theta \mid \mathbf{x})$.

For example, let $X \sim \text{Binomial}(n, p)$ with a $\text{Beta}(\alpha, \beta)$ prior and $w(p) = \{p(1-p)\}^{-1}$.

- The posterior is $\text{Beta}(x + \alpha, n - x + \beta)$, so the weighted posterior is $\text{Beta}(x + \alpha - 1, n - x + \beta - 1)$ (when $x + \alpha > 1$ and $n - x + \beta > 1$), and

$$\delta_g(x) = \frac{x + \alpha - 1}{n + \alpha + \beta - 2}.$$

- With the uniform prior ($\alpha = \beta = 1$) this is x/n for $0 < x < n$; for the end-points $x = 0$ and $x = n$ see Exercise 1.3.

(ii) Absolute error loss

Corollary 9.5 (Bayes estimator under ALF is the posterior median)

Under $L(\theta, a) = |a - \theta|$, and assuming $\mathbb{E}(|\theta| \mid \mathbf{x}) < \infty$, any posterior median is a Bayes estimator, i.e. any value $m(\mathbf{x})$ satisfying

$$\mathbb{P}(\theta \leq m(\mathbf{x}) \mid \mathbf{x}) \geq \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\theta \geq m(\mathbf{x}) \mid \mathbf{x}) \geq \frac{1}{2}.$$

Proof.

We assume that the posterior has a density $\pi(\theta \mid \mathbf{x})$. For fixed $\mathbf{x}$ write

$$h(a) = \mathbb{E}[|a - \theta| \mid \mathbf{x}] = \int_{-\infty}^a (a - \theta)\pi(\theta \mid \mathbf{x})d\theta + \int_a^{\infty} (\theta - a)\pi(\theta \mid \mathbf{x})d\theta.$$

(continued)

Proof: Bayes estimator under ALF is the posterior median

Proof (cont.)

Differentiating under the integral sign (Leibniz's rule; the boundary terms cancel because the integrand vanishes at $\theta = a$),

$$\begin{aligned} h'(a) &= \int_{-\infty}^a \pi(\theta | \mathbf{x}) d\theta - \int_a^{\infty} \pi(\theta | \mathbf{x}) d\theta \\ &= \Pi(a | \mathbf{x}) - \{1 - \Pi(a | \mathbf{x})\} = 2\Pi(a | \mathbf{x}) - 1, \end{aligned}$$

where Π is the posterior distribution function.

Setting $h'(a) = 0$ gives $\Pi(a | \mathbf{x}) = \frac{1}{2}$, i.e. a is a posterior median. Since h' is non-decreasing, h is convex and the stationary point is a global minimum.

(For a posterior without a density the result still holds; the argument is given in Lecture Notes, Appendix A.) □

Practice (Lecture Notes, Chapter 1)

- Exercises (Section 1.13): 1.3, 1.12.

Next lecture

Lecture 10: Small-neighbourhood (MAP) and LINEX losses.