

Lecture 8: Structure of the Bayes rule: posterior expected 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

This lecture

Previous lecture

Lecture 7: Prior distribution, Bayes risk and Bayes rule.

Today

- The Bayes Principle
 - Structure of the Bayes Rule

Structure of the Bayes rule

The following result is the single most useful computational tool in Bayesian decision theory. It converts a problem of minimising an integral over the whole sample space into a *pointwise* minimisation.

Theorem 8.1 (Structure of the Bayes rule)

Let $L(\theta, a) \geq 0$, let $\mathbf{X}$ have density $f(\mathbf{x} \mid \theta)$ and let θ have prior density $g(\theta)$. Define the marginal (or prior predictive) density of $\mathbf{X}$ and the posterior density of θ given $\mathbf{X} = \mathbf{x}$,

$$p(\mathbf{x}) = \int_{\Theta} f(\mathbf{x} \mid \theta) g(\theta) d\theta, \quad \pi(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta) g(\theta)}{p(\mathbf{x})}, \quad p(\mathbf{x}) > 0.$$

Then a rule δ_g is Bayes with respect to g if, for (almost) every $\mathbf{x}$ with $p(\mathbf{x}) > 0$, the action $\delta_g(\mathbf{x})$ minimises the posterior expected loss

$$\rho(\pi, a \mid \mathbf{x}) = \int_{\Theta} L(\theta, a) \pi(\theta \mid \mathbf{x}) d\theta.$$

Proof.

By the definition of Bayes risk,

$$r(g, \delta) = \int_{\Theta} R(\theta, \delta) g(\theta) d\theta = \int_{\Theta} \left[\int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x} | \theta) d\mathbf{x} \right] g(\theta) d\theta.$$

The integrand is non-negative, so by Tonelli's theorem the order of integration may be interchanged:

$$r(g, \delta) = \int_{\mathcal{X}} \left[\int_{\Theta} L(\theta, \delta(\mathbf{x})) f(\mathbf{x} | \theta) g(\theta) d\theta \right] d\mathbf{x}.$$

Now write $f(\mathbf{x} | \theta)g(\theta) = \pi(\theta | \mathbf{x})p(\mathbf{x})$, which is the statement of Bayes' theorem.

Proof (cont.)

Since $p(\mathbf{x})$ does not involve θ , it may be taken outside the inner integral:

$$r(g, \delta) = \int_{\mathcal{X}} \underbrace{\left[\int_{\Theta} L(\theta, \delta(\mathbf{x})) \pi(\theta | \mathbf{x}) d\theta \right]}_{= \rho(\pi, \delta(\mathbf{x}) | \mathbf{x})} p(\mathbf{x}) d\mathbf{x}.$$

The outer integrand is a product of the non-negative weight $p(\mathbf{x})$ and the posterior expected loss at $\mathbf{x}$. Since the value $\delta(\mathbf{x})$ may be chosen *separately for each $\mathbf{x}$* , and since $p(\mathbf{x}) \geq 0$, the whole integral is minimised by choosing, for each $\mathbf{x}$, that action which minimises $\rho(\pi, a | \mathbf{x})$. Hence

$$\delta_g(\mathbf{x}) = \arg \min_{a \in \mathcal{A}} \int_{\Theta} L(\theta, a) \pi(\theta | \mathbf{x}) d\theta,$$

whenever the minimum exists. □

Remark 8.2

The proof requires only that the interchange of integration be permissible, which Tonelli's theorem guarantees for non-negative integrands. If $r(g, \delta) = \infty$ for every δ , the notion of a Bayes rule is vacuous; we tacitly assume $r(g) < \infty$.

Practice (Lecture Notes, Chapter 1)

- Short-answer questions (Section 1.12): Q9.
- Exercises (Section 1.13): 1.15.

Next lecture

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