

Lecture 12: Normal–normal and Poisson–gamma models; different losses, different estimates

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 11: Worked examples: linear rules; the beta–binomial model.

Today

- The Bayes Principle
 - Worked Examples on Bayes Estimation

Example 12.1 (Normal-normal: Bayes estimator of a mean)

Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, 1)$ and let the prior for μ be $N(0, \tau^2)$. Under squared error loss, obtain the posterior distribution, the Bayes estimator, and the Bayes risk.

Solution

Write $S = \sum_{i=1}^n x_i$ and $\bar{x} = S/n$. The likelihood is

$$f(\mathbf{x} \mid \mu) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right\},$$

and the prior is

$$g(\mu) = \frac{1}{\sqrt{2\pi} \tau} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}.$$

Solution (cont.)

Step 1: posterior kernel. Retaining only factors involving μ ,

$$\pi(\mu \mid \mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \left[\sum_i (x_i - \mu)^2 + \frac{\mu^2}{\tau^2} \right] \right\}.$$

Expanding $\sum_i (x_i - \mu)^2 = \sum_i x_i^2 - 2\mu S + n\mu^2$ and discarding $\sum_i x_i^2$ (free of μ),

$$\pi(\mu \mid \mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \left[\left(n + \frac{1}{\tau^2} \right) \mu^2 - 2\mu S \right] \right\}.$$

Solution (cont.)

Step 2: complete the square. Let $A = n + \tau^{-2}$. Then

$$A\mu^2 - 2\mu S = A \left(\mu^2 - \frac{2\mu S}{A} \right) = A \left(\mu - \frac{S}{A} \right)^2 - \frac{S^2}{A},$$

and since S^2/A is free of μ ,

$$\pi(\mu \mid \mathbf{x}) \propto \exp \left\{ -\frac{A}{2} \left(\mu - \frac{S}{A} \right)^2 \right\}.$$

This is recognised as a normal kernel, so

$$\mu \mid \mathbf{x} \sim N \left(\frac{S}{n + \tau^{-2}}, \frac{1}{n + \tau^{-2}} \right)$$

Solution (cont.)

Step 3: Bayes estimator. By the corollary “Bayes estimator under SELF is the posterior mean”,

$$\hat{\mu}_{\text{Bayes}} = \mathbb{E}(\mu \mid \mathbf{x}) = \frac{S}{n + \tau^{-2}} = \frac{n\tau^2}{n\tau^2 + 1} \bar{x},$$

a *shrinkage* estimator: the sample mean pulled towards the prior mean 0 by the factor $n\tau^2/(n\tau^2 + 1) \in (0, 1)$. As $n \rightarrow \infty$ or $\tau^2 \rightarrow \infty$ the factor tends to 1 and the Bayes estimator approaches $\bar{x}$.

Solution (cont.)

Step 4: Bayes risk. The posterior variance is $(n + \tau^{-2})^{-1}$, which does *not* depend on $\mathbf{x}$. Hence by the same corollary

$$r(g) = \mathbb{E}[\text{Var}(\mu \mid \mathbf{X})] = \frac{1}{n + \tau^{-2}} = \frac{\tau^2}{n\tau^2 + 1}.$$

Remark 12.2 (Two important special cases)

- (i) Taking $n = 1$ and writing X for X_1 , the Bayes estimator is $\tau^2 X / (1 + \tau^2)$ with Bayes risk $\tau^2 / (1 + \tau^2)$, exactly matching the example “Bayes risk of a linear rule, normal model”. This confirms that the best linear rule is in fact the overall Bayes rule in this model.
- (ii) Taking $\tau^2 = 1$ (standard normal prior), the posterior is $N(S/(n+1), 1/(n+1))$, the Bayes estimator is $S/(n+1) = n\bar{x}/(n+1)$, and the Bayes risk is $1/(n+1)$.

Note in particular the denominator $n+1$: precisions add. The posterior precision is the data precision $n \cdot 1$ (n observations, each of precision 1) plus the prior precision $\tau^{-2} = 1$, that is, $n+1 = n \cdot 1 + \tau^{-2}$; the prior contributes exactly the information of one extra observation.

Example 12.3 (Poisson–gamma)

Let $X_1, \dots, X_n$ be i.i.d. $\text{Poisson}(\theta)$ and let the prior be $\theta \sim \text{Gamma}(\alpha, \beta)$ with density

$$g(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0.$$

Obtain the Bayes estimator of θ under SELF.

Solution

With $S = \sum_i x_i$,

$$f(\mathbf{x} \mid \theta) = \frac{e^{-n\theta} \theta^S}{\prod_i x_i!} \propto e^{-n\theta} \theta^S.$$

Solution (cont.)

Hence

$$\pi(\theta \mid \mathbf{x}) \propto \theta^S e^{-n\theta} \cdot \theta^{\alpha-1} e^{-\beta\theta} = \theta^{S+\alpha-1} e^{-(n+\beta)\theta},$$

which is the kernel of $\text{Gamma}(S + \alpha, n + \beta)$. Therefore

$$\hat{\theta}_{\text{Bayes}} = \mathbb{E}(\theta \mid \mathbf{x}) = \frac{S + \alpha}{n + \beta} = \frac{n}{n + \beta} \bar{x} + \frac{\beta}{n + \beta} \cdot \frac{\alpha}{\beta},$$

again a weighted average of the sample mean $\bar{x}$ and the prior mean α/β .

Solution (cont.)

The posterior expected loss is

$$\text{Var}(\theta \mid \mathbf{x}) = \frac{S + \alpha}{(n + \beta)^2},$$

and the Bayes risk is

$$r(g) = \mathbb{E} \left[\frac{S + \alpha}{(n + \beta)^2} \right] = \frac{n \mathbb{E}_g(\theta) + \alpha}{(n + \beta)^2} = \frac{n\alpha/\beta + \alpha}{(n + \beta)^2} = \frac{\alpha}{\beta(n + \beta)},$$

using $\mathbb{E}(S) = \mathbb{E}\{\mathbb{E}(S \mid \theta)\} = n\mathbb{E}_g(\theta) = n\alpha/\beta$.

Note 12.4 (Conjugate families)

The beta–binomial, normal–normal and Poisson–gamma examples illustrate *conjugacy*:

- the beta family is conjugate to the binomial likelihood,
- the normal to the normal (with known variance), and
- the gamma to the Poisson.

A family $\mathcal{G}$ of priors is *conjugate* for a likelihood if the posterior remains in $\mathcal{G}$ for every prior in $\mathcal{G}$ and every possible sample. Conjugacy is a computational convenience, not a statistical principle.

Example 12.5 (Different losses, different Bayes estimates)

The waiting times (minutes) of a commuter at a bus stop are modelled as i.i.d. $U(0, \theta)$, where θ is the unknown interval between buses. Five waiting times were

$$4.0, \quad 10.5, \quad 7.5, \quad 16.0, \quad 13.0.$$

The prior for θ is the Pareto distribution

$$g(\theta) = \frac{\alpha \beta^\alpha}{\theta^{\alpha+1}}, \quad \theta > \beta,$$

with $\alpha = 2$ and $\beta = 5$. Find the Bayes estimates of θ under squared error, absolute error and small-neighbourhood 0–1 loss.

Solution

Likelihood. With $x_{(n)} = \max_i x_i$,

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbf{1}(0 < x_i < \theta) = \theta^{-n} \mathbf{1}(\theta > x_{(n)}).$$

Posterior. Put $m = \max(\beta, x_{(n)})$. For $\theta > m$,

$$\pi(\theta \mid \mathbf{x}) \propto \theta^{-n} \cdot \theta^{-(\alpha+1)} = \theta^{-(\alpha+n+1)},$$

and $\pi(\theta \mid \mathbf{x}) = 0$ for $\theta \leq m$.

Solution (cont.)

This is the Pareto density with shape $\alpha + n$ and scale m :

$$\pi(\theta \mid \mathbf{x}) = \frac{(\alpha + n) m^{\alpha+n}}{\theta^{\alpha+n+1}}, \quad \theta > m,$$

so that

$$\mathbb{P}(\theta > t \mid \mathbf{x}) = \left(\frac{m}{t}\right)^{\alpha+n}, \quad t \geq m.$$

(The Pareto family is therefore conjugate for the uniform model.)

Here $n = 5$, $x_{(5)} = 16$ and $\beta = 5$, so

$$m = 16 \quad \text{and the posterior shape is} \quad \alpha + n = 7.$$

Solution (cont.)

Squared error: posterior mean.

$$\begin{aligned}\mathbb{E}(\theta \mid \mathbf{x}) &= \int_m^\infty \theta \frac{(\alpha + n) m^{\alpha+n}}{\theta^{\alpha+n+1}} d\theta = (\alpha + n) m^{\alpha+n} \cdot \frac{m^{-(\alpha+n-1)}}{\alpha + n - 1} \\ &= \frac{(\alpha + n) m}{\alpha + n - 1} = \frac{7 \times 16}{6} = 18.67.\end{aligned}$$

Absolute error: posterior median. Solve $\mathbb{P}(\theta > t \mid \mathbf{x}) = (m/t)^{\alpha+n} = \frac{1}{2}$:

$$t = m 2^{1/(\alpha+n)} = 16 \times 2^{1/7} = 16 \times 1.1041 = 17.67.$$

Small-neighbourhood loss: posterior mode. The posterior density is decreasing on (m, ∞) , so the mode is the left end-point: $\tilde{\theta} = m = 16$.

Solution (cont.)

Comparison.

$$\text{mode} = 16.00 < \text{median} = 17.67 < \text{mean} = 18.67$$

- The three estimates differ because the posterior is skewed to the right.
- The choice of loss function therefore matters: the MAP estimate equals the largest waiting time, which is certainly below θ , whereas squared error, which penalises large errors heavily, pushes the estimate furthest out.

Example 12.6 (LINEX loss in a Poisson problem)

The numbers of failures of a machine in $n = 10$ successive weeks are i.i.d. $\text{Poisson}(\theta)$ with total $S = 25$. With the gamma prior of Example 12.3, taking $\alpha = 2$ and $\beta = 1$, find the Bayes estimate of the failure rate θ under squared error loss, and under LINEX loss with $c = 1$ and with $c = -1$.

Solution

By Example 12.3 the posterior is

$$\text{Gamma}(S + \alpha, n + \beta) = \text{Gamma}(27, 11) \quad (\text{shape } 27, \text{rate } 11).$$

Under squared error loss the Bayes estimate is the posterior mean $27/11 = 2.4545$.

Solution (cont.)

For LINEX loss we need $\mathbb{E}(e^{-c\theta} \mid \mathbf{x})$. For a $\text{Gamma}(A, B)$ posterior (rate B), and $B + c > 0$,

$$\begin{aligned}\mathbb{E}(e^{-c\theta} \mid \mathbf{x}) &= \int_0^\infty e^{-c\theta} \frac{B^A}{\Gamma(A)} \theta^{A-1} e^{-B\theta} d\theta \\ &= \frac{B^A}{\Gamma(A)} \cdot \frac{\Gamma(A)}{(B+c)^A} = \left(\frac{B}{B+c} \right)^A.\end{aligned}$$

By Corollary 10.3,

$$\delta_L = -\frac{1}{c} \log \left(\frac{B}{B+c} \right)^A = \frac{A}{c} \log \left(1 + \frac{c}{B} \right) = \frac{S + \alpha}{c} \log \left(1 + \frac{c}{n + \beta} \right).$$

Solution (cont.)

With $A = 27$ and $B = 11$:

$$c = 1 : \quad \delta_L = 27 \log \frac{12}{11} = 27 \times 0.08701 = 2.3493,$$

$$c = -1 : \quad \delta_L = -27 \log \frac{10}{11} = 27 \times 0.09531 = 2.5734.$$

- As Remark 10.4 predicts, $c = 1$ (over-estimation costly) gives an estimate below the posterior mean 2.4545, and $c = -1$ (under-estimation costly) gives one above it.
- In reliability work, where underestimating a failure rate can be dangerous, a negative c is often appropriate.

Practice (Lecture Notes, Chapter 1)

- Exercises (Section 1.13): 1.5, 1.6.

Next lecture

Lecture 13: Bayes rules and admissibility.