Guided Study (Week 4 & 5): Bayesian Updating, Predictive Distributions, Transformations

Problem Sheet + Worked Solutions

Guided Study Session

Q1: Binomial model + Beta prior

Model: $Y_1, \ldots, Y_N \mid p \overset{i.i.d.}{\sim} \mathrm{Binom}(n, p)$.

Tasks

(a) With prior $p \sim \mathrm{Beta}(\alpha, \beta)$ derive the posterior.

(b) After observing y_{N+1}, derive $\pi(p \mid y_1, \ldots, y_N, y_{N+1})$ by *updating*.

(c) Show you get the same result if you observe all $N + 1$ points at the start.

Q1(a)

Goal: derive $\pi(p \mid y) \propto \pi(y \mid p)\pi(p)$, with $y = \{y_1, \ldots, y_N\}$.

Work in pairs:

- What is the likelihood $\pi(y \mid p)$ up to proportionality (ignore constants not depending on p)?
- Multiply by the Beta prior kernel $p^{\alpha-1}(1-p)^{\beta-1}$.
- Identify the resulting distribution family.

Q1(a)

Goal: derive $\pi(p \mid y) \propto \pi(y \mid p)\pi(p)$, with $y = \{y_1, \ldots, y_N\}$.

Work in pairs:

- What is the likelihood $\pi(y \mid p)$ up to proportionality (ignore constants not depending on p)?
- Multiply by the Beta prior kernel $p^{\alpha-1}(1-p)^{\beta-1}$.
- Identify the resulting distribution family.

Hint: $\prod_{i=1}^{N} p^{y_i}(1-p)^{n-y_i} = p^{\sum y_i}(1-p)^{nN-\sum y_i}$.

Q1(a) Solution (posterior)

Likelihood:

$$\pi(y \mid p) = \prod_{i=1}^{N} \binom{n}{y_i} p^{y_i} (1-p)^{n-y_i} \;\propto\; p^{\sum_{i=1}^{N} y_i} (1-p)^{nN - \sum_{i=1}^{N} y_i}.$$

Q1(a) Solution (posterior)

Likelihood:

$$\pi(y \mid p) = \prod_{i=1}^{N} \binom{n}{y_i} p^{y_i}(1-p)^{n-y_i} \propto p^{\sum_{i=1}^{N} y_i}(1-p)^{nN-\sum_{i=1}^{N} y_i}.$$

Prior kernel: $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$.

Q1(a) Solution (posterior)

Likelihood:

$$\pi(y \mid p) = \prod_{i=1}^{N} \binom{n}{y_i} p^{y_i}(1-p)^{n-y_i} \;\propto\; p^{\sum_{i=1}^{N} y_i}(1-p)^{nN-\sum_{i=1}^{N} y_i}.$$

Prior kernel: $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$.

Posterior kernel:

$$\pi(p \mid y) \propto p^{\sum y_i+\alpha-1}(1-p)^{nN-\sum y_i+\beta-1}.$$

Q1(a) Solution (posterior)

Likelihood:

$$\pi(y \mid p) = \prod_{i=1}^{N} \binom{n}{y_i} p^{y_i} (1-p)^{n-y_i} \;\propto\; p^{\sum_{i=1}^{N} y_i} (1-p)^{nN - \sum_{i=1}^{N} y_i}.$$

Prior kernel: $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$.

Posterior kernel:

$$\pi(p \mid y) \propto p^{\sum y_i + \alpha - 1}(1-p)^{nN - \sum y_i + \beta - 1}.$$

Therefore

$$p \mid y \sim \mathrm{Beta}\Big(\alpha + \sum_{i=1}^{N} y_i,\ \beta + nN - \sum_{i=1}^{N} y_i\Big).$$

Q1(b) Sequential update: one more observation

We observe a new $y_{N+1} \sim \mathrm{Binom}(n, p)$ independently.

Guided step:

$$\pi(p \mid y, y_{N+1}) \propto \pi(y_{N+1} \mid p)\, \pi(p \mid y).$$

Q1(b) Sequential update: one more observation

We observe a new $y_{N+1} \sim \mathrm{Binom}(n, p)$ independently.

Guided step:

$$\pi(p \mid y, y_{N+1}) \propto \pi(y_{N+1} \mid p)\, \pi(p \mid y).$$

Write kernels:

$$\pi(y_{N+1} \mid p) \propto p^{y_{N+1}}(1-p)^{n-y_{N+1}}.$$

Q1(b) Sequential update: one more observation

We observe a new $y_{N+1} \sim \mathrm{Binom}(n, p)$ independently.

Guided step:

$$\pi(p \mid y, y_{N+1}) \propto \pi(y_{N+1} \mid p)\,\pi(p \mid y).$$

Write kernels:

$$\pi(y_{N+1} \mid p) \propto p^{y_{N+1}}(1-p)^{n-y_{N+1}}.$$

Multiply:

$$\pi(p \mid y, y_{N+1}) \propto p^{(\sum_{i=1}^{N} y_i)+y_{N+1}+\alpha-1}(1-p)^{nN-\sum_{i=1}^{N} y_i+(n-y_{N+1})+\beta-1}.$$

Q1(b) Sequential update: one more observation

We observe a new $y_{N+1} \sim \mathrm{Binom}(n, p)$ independently.

Guided step:

$$\pi(p \mid y, y_{N+1}) \propto \pi(y_{N+1} \mid p)\, \pi(p \mid y).$$

Write kernels:

$$\pi(y_{N+1} \mid p) \propto p^{y_{N+1}} (1-p)^{n-y_{N+1}}.$$

Multiply:

$$\pi(p \mid y, y_{N+1}) \propto p^{(\sum_{i=1}^{N} y_i) + y_{N+1} + \alpha - 1} (1-p)^{nN - \sum_{i=1}^{N} y_i + (n - y_{N+1}) + \beta - 1}.$$

So

$$p \mid y, y_{N+1} \sim \mathrm{Beta}\Big(\alpha + \sum_{i=1}^{N+1} y_i,\ \beta + n(N+1) - \sum_{i=1}^{N+1} y_i\Big).$$

Q1(c) All-at-once vs sequential: why the same?

If we observe all $N+1$ points at once, the likelihood is

$$\pi(y_1, \ldots, y_{N+1} \mid p) \propto p^{\sum_{i=1}^{N+1} y_i}(1-p)^{n(N+1)-\sum_{i=1}^{N+1} y_i}.$$

Q1(c) All-at-once vs sequential: why the same?

If we observe all $N+1$ points at once, the likelihood is

$$\pi(y_1, \ldots, y_{N+1} \mid p) \propto p^{\sum_{i=1}^{N+1} y_i} (1-p)^{n(N+1) - \sum_{i=1}^{N+1} y_i}.$$

Multiplying by the same Beta prior yields the same Beta posterior:

$$p \mid y_1, \ldots, y_{N+1} \sim \mathrm{Beta}\Big(\alpha + \sum_{i=1}^{N+1} y_i,\ \beta + n(N+1) - \sum_{i=1}^{N+1} y_i\Big).$$

Q1(c) All-at-once vs sequential: why the same?

If we observe all $N+1$ points at once, the likelihood is

$$\pi(y_1, \ldots, y_{N+1} \mid p) \propto p^{\sum_{i=1}^{N+1} y_i}(1-p)^{n(N+1)-\sum_{i=1}^{N+1} y_i}.$$

Multiplying by the same Beta prior yields the same Beta posterior:

$$p \mid y_1, \ldots, y_{N+1} \sim \mathrm{Beta}\Big(\alpha + \sum_{i=1}^{N+1} y_i,\ \beta + n(N+1) - \sum_{i=1}^{N+1} y_i\Big).$$

Takeaway: Conjugacy makes Bayesian updating *associative*:

$$\text{Prior} \xrightarrow{\text{data}} \text{Posterior} \xrightarrow{\text{new data}} \text{Updated posterior}$$

equals

$$\text{Prior} \xrightarrow{\text{all data at once}} \text{Posterior}.$$

Q2: Poisson model + Gamma prior

Model: $Y_1, \dots, Y_N \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Pois}(\lambda)$.

(a) Prior $\lambda \sim \mathrm{Gamma}(\alpha, \beta)$. Derive posterior.

(b) Fix $\alpha = 1$. Discuss effect of β on posterior.

(c) Derive posterior predictive for a new observation $\tilde{y}$.

Hint (given): Negative Binomial pmf with parameters r, p:

$$\pi(k \mid r, p) = \frac{\Gamma(k + r)}{\Gamma(r)\, k!} (1 - p)^k p^r, \quad k \in \{0, 1, 2, \dots\}.$$

Q2(a) Posterior derivation

Likelihood kernel:

$$\pi(y \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \;\propto\; \lambda^{\sum_{i=1}^{N} y_i} \, e^{-N\lambda}.$$

Q2(a) Posterior derivation

Likelihood kernel:

$$\pi(y \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \propto \lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda}.$$

Gamma prior (rate parameterization):

$$\pi(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Q2(a) Posterior derivation

Likelihood kernel:

$$\pi(y \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \propto \lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda}.$$

Gamma prior (rate parameterization):

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Posterior kernel:

$$\pi(\lambda \mid y) \propto \lambda^{\sum y_i + \alpha - 1} e^{-(N+\beta)\lambda}.$$

Q2(a) Posterior derivation

Likelihood kernel:

$$\pi(y \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \;\propto\; \lambda^{\sum_{i=1}^{N} y_i} e^{-N\lambda}.$$

Gamma prior (rate parameterization):

$$\pi(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \;\propto\; \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Posterior kernel:

$$\pi(\lambda \mid y) \propto \lambda^{\sum y_i + \alpha - 1} e^{-(N+\beta)\lambda}.$$

Therefore

$$\lambda \mid y \sim \mathrm{Gamma}\Big(\alpha + \sum_{i=1}^{N} y_i,\; \beta + N\Big).$$

Q2(b) Effect of β when $\alpha = 1$

With $\alpha = 1$,

$$\lambda \mid y \sim \mathrm{Gamma}\Big(1 + \sum_{i=1}^{N} y_i,\ \beta + N\Big).$$

Q2(b) Effect of β when $\alpha = 1$

With $\alpha = 1$,

$$\lambda \mid y \sim \mathrm{Gamma}\left(1 + \sum_{i=1}^{N} y_i,\ \beta + N\right).$$

For $\mathrm{Gamma}(a, b)$ (shape a, rate b):

$$\mathbb{E}[\lambda \mid y] = \frac{1 + \sum y_i}{\beta + N}, \qquad \mathrm{Var}(\lambda \mid y) = \frac{1 + \sum y_i}{(\beta + N)^2}.$$

Q2(b) Effect of β when $\alpha = 1$

With $\alpha = 1$,

$$\lambda \mid y \sim \mathrm{Gamma}\left(1 + \sum_{i=1}^{N} y_i,\ \beta + N\right).$$

For $\mathrm{Gamma}(a, b)$ (shape a, rate b):

$$\mathbb{E}[\lambda \mid y] = \frac{1 + \sum y_i}{\beta + N}, \qquad \mathrm{Var}(\lambda \mid y) = \frac{1 + \sum y_i}{(\beta + N)^2}.$$

Interpretation:

- Larger β (stronger prior pull toward smaller λ) $\Rightarrow$ smaller posterior mean and variance.
- Smaller β $\Rightarrow$ weaker prior (data dominates more), larger mean/variance.

Q2(c) Posterior predictive $\pi(\tilde{y} \mid y)$

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y} \mid \lambda)\, \pi(\lambda \mid y)\, d\lambda, \qquad \pi(\tilde{y} \mid \lambda) = \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!}.$$

Q2(c) Posterior predictive $\pi(\tilde{y} \mid y)$

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y} \mid \lambda)\, \pi(\lambda \mid y)\, d\lambda, \qquad \pi(\tilde{y} \mid \lambda) = \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!}.$$

Plug in posterior $\lambda \mid y \sim \mathrm{Gamma}(\alpha + \sum y_i, \beta + N)$ and integrate:

$$\pi(\tilde{y} \mid y) = \frac{\Gamma\left(\tilde{y} + \alpha + \sum y_i\right)}{\tilde{y}!\, \Gamma\left(\alpha + \sum y_i\right)} \left(\frac{\beta + N}{\beta + N + 1}\right)^{\alpha + \sum y_i} \left(\frac{1}{\beta + N + 1}\right)^{\tilde{y}}.$$

Q2(c) Posterior predictive $\pi(\tilde{y} \mid y)$

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y} \mid \lambda)\, \pi(\lambda \mid y)\, d\lambda, \qquad \pi(\tilde{y} \mid \lambda) = \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!}.$$

Plug in posterior $\lambda \mid y \sim \mathrm{Gamma}(\alpha + \sum y_i, \beta + N)$ and integrate:

$$\pi(\tilde{y} \mid y) = \frac{\Gamma\big(\tilde{y} + \alpha + \sum y_i\big)}{\tilde{y}!\, \Gamma\big(\alpha + \sum y_i\big)} \left(\frac{\beta + N}{\beta + N + 1}\right)^{\alpha + \sum y_i} \left(\frac{1}{\beta + N + 1}\right)^{\tilde{y}}.$$

This matches $\mathrm{NegBin}(r, p)$ with

$$r = \alpha + \sum_{i=1}^{N} y_i, \qquad p = \frac{\beta + N}{\beta + N + 1}.$$

Q2(c) Posterior predictive $\pi(\tilde{y} \mid y)$

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y} \mid \lambda)\, \pi(\lambda \mid y)\, d\lambda, \qquad \pi(\tilde{y} \mid \lambda) = \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!}.$$

Plug in posterior $\lambda \mid y \sim \mathrm{Gamma}(\alpha + \sum y_i, \beta + N)$ and integrate:

$$\pi(\tilde{y} \mid y) = \frac{\Gamma\big(\tilde{y} + \alpha + \sum y_i\big)}{\tilde{y}!\, \Gamma\big(\alpha + \sum y_i\big)} \left(\frac{\beta + N}{\beta + N + 1}\right)^{\alpha + \sum y_i} \left(\frac{1}{\beta + N + 1}\right)^{\tilde{y}}.$$

This matches $\mathrm{NegBin}(r, p)$ with

$$r = \alpha + \sum_{i=1}^{N} y_i, \qquad p = \frac{\beta + N}{\beta + N + 1}.$$

Takeaway: Poisson likelihood + Gamma posterior $\Rightarrow$ Negative Binomial posterior predictive.

Q3: Density of a transformed variable

Let $X \sim \pi(x)$ be continuous and $Y = h(X)$ where h is strictly monotonic and smooth.

Show:

$$\pi_Y(y) = \pi_X(x) \left| \frac{\partial x}{\partial y} \right| .$$

Then: if $X \sim \mathrm{Exp}(1)$, find the density of $Y = \sqrt{X}$.

Q3 Solution: general change of variables

If h is strictly increasing:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(h(X) \leq y) = \Pr\big(X \leq h^{-1}(y)\big) = F_X(h^{-1}(y)).$$

Q3 Solution: general change of variables

If h is strictly increasing:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(h(X) \leq y) = \Pr\big(X \leq h^{-1}(y)\big) = F_X(h^{-1}(y)).$$

Differentiate w.r.t. y:

$$\pi_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(h^{-1}(y)) = \pi_X(h^{-1}(y))h^{-1}(y)y.$$

Q3 Solution: general change of variables

If h is strictly increasing:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(h(X) \leq y) = \Pr\big(X \leq h^{-1}(y)\big) = F_X(h^{-1}(y)).$$

Differentiate w.r.t. y:

$$\pi_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(h^{-1}(y)) = \pi_X(h^{-1}(y))h^{-1}(y)y.$$

If h is decreasing, a minus sign appears; both cases combine to

$$\pi_Y(y) = \pi_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|.$$

Q3 Solution: $X \sim \mathrm{Exp}(1)$, $Y = \sqrt{X}$

Here $\pi_X(x) = e^{-x}$ for $x > 0$. Let $Y = \sqrt{X} \Rightarrow X = Y^2$ with $y > 0$.

Q3 Solution: $X \sim \mathrm{Exp}(1)$, $Y = \sqrt{X}$

Here $\pi_X(x) = e^{-x}$ for $x > 0$. Let $Y = \sqrt{X} \Rightarrow X = Y^2$ with $y > 0$.
Compute Jacobian:

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y} = 2y.$$

Q3 Solution: $X \sim \mathrm{Exp}(1)$, $Y = \sqrt{X}$

Here $\pi_X(x) = e^{-x}$ for $x > 0$. Let $Y = \sqrt{X} \Rightarrow X = Y^2$ with $y > 0$.
Compute Jacobian:

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y} = 2y.$$

Therefore

$$\pi_Y(y) = \pi_X(y^2)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = e^{-y^2} \cdot 2y, \qquad y > 0.$$

Q3 Solution: $X \sim \mathrm{Exp}(1)$, $Y = \sqrt{X}$

Here $\pi_X(x) = e^{-x}$ for $x > 0$. Let $Y = \sqrt{X} \Rightarrow X = Y^2$ with $y > 0$.
Compute Jacobian:

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y} = 2y.$$

Therefore

$$\pi_Y(y) = \pi_X(y^2)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = e^{-y^2} \cdot 2y, \qquad y > 0.$$

Check: looks like a Rayleigh-type shape; integrates to 1 on $(0, \infty)$.

Q4: Exponential model + invariant prior

Let $X_1, \ldots, X_n \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Exp}(\lambda)$ with density

$$\pi(x \mid \lambda) = \lambda e^{-\lambda x}, \quad x > 0.$$

(a) Construct an invariant (Jeffreys) prior for λ.

(b) Derive posterior using this prior.

(c) What do you notice about this prior?

Q4(a) Jeffreys prior via Fisher information

Single-observation log-likelihood:

$$\log \pi(X \mid \lambda) = \log \lambda - \lambda X.$$

Q4(a) Jeffreys prior via Fisher information

Single-observation log-likelihood:

$$\log \pi(X \mid \lambda) = \log \lambda - \lambda X.$$

Derivatives:

$$\frac{\partial}{\partial \lambda} \log \pi(X \mid \lambda) = \frac{1}{\lambda} - X, \qquad \frac{\partial^2}{\partial \lambda^2} \log \pi(X \mid \lambda) = -\frac{1}{\lambda^2}.$$

Q4(a) Jeffreys prior via Fisher information

Single-observation log-likelihood:

$$\log \pi(X \mid \lambda) = \log \lambda - \lambda X.$$

Derivatives:

$$\frac{\partial}{\partial \lambda} \log \pi(X \mid \lambda) = \frac{1}{\lambda} - X, \qquad \frac{\partial^2}{\partial \lambda^2} \log \pi(X \mid \lambda) = -\frac{1}{\lambda^2}.$$

Fisher information:

$$I(\lambda) = \mathbb{E}\left[-\frac{\partial^2}{\partial \lambda^2} \log \pi(X \mid \lambda)\right] = \frac{1}{\lambda^2}.$$

Q4(a) Jeffreys prior via Fisher information

Single-observation log-likelihood:

$$\log \pi(X \mid \lambda) = \log \lambda - \lambda X.$$

Derivatives:

$$\frac{\partial}{\partial \lambda} \log \pi(X \mid \lambda) = \frac{1}{\lambda} - X, \qquad \frac{\partial^2}{\partial \lambda^2} \log \pi(X \mid \lambda) = -\frac{1}{\lambda^2}.$$

Fisher information:

$$I(\lambda) = \mathbb{E}\left[-\frac{\partial^2}{\partial \lambda^2} \log \pi(X \mid \lambda)\right] = \frac{1}{\lambda^2}.$$

Jeffreys prior:

$$\pi(\lambda) \propto \sqrt{I(\lambda)} = \frac{1}{\lambda}, \qquad \lambda > 0.$$

Q4(b) Posterior with Jeffreys prior

Likelihood for n i.i.d. observations:

$$\pi(x \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Q4(b) Posterior with Jeffreys prior

Likelihood for n i.i.d. observations:

$$\pi(x \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Multiply by prior $\pi(\lambda) \propto 1/\lambda$:

$$\pi(\lambda \mid x) \propto \lambda^{n-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Q4(b) Posterior with Jeffreys prior

Likelihood for n i.i.d. observations:

$$\pi(x \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Multiply by prior $\pi(\lambda) \propto 1/\lambda$:

$$\pi(\lambda \mid x) \propto \lambda^{n-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Identify Gamma kernel:

$$\lambda \mid x \sim \mathrm{Gamma}\left(n, \sum_{i=1}^{n} x_i\right),$$

(shape n, rate $\sum x_i$).

Q4(c) What do you notice about the prior?

Consider the integral over $(0, \infty)$:

$$\int_0^\infty \frac{1}{\lambda}\, d\lambda = \infty.$$

Q4(c) What do you notice about the prior?

Consider the integral over $(0, \infty)$:

$$\int_0^\infty \frac{1}{\lambda}\, d\lambda = \infty.$$

So the Jeffreys prior $\pi(\lambda) \propto 1/\lambda$ is improper.

Q4(c) What do you notice about the prior?

Consider the integral over $(0, \infty)$:

$$\int_0^\infty \frac{1}{\lambda}\, d\lambda = \infty.$$

So the Jeffreys prior $\pi(\lambda) \propto 1/\lambda$ is improper.

But: the posterior $\mathrm{Gamma}(n, \sum x_i)$ is a proper distribution for $n \geq 1$. So it is still usable (common in objective Bayes).

Wrap-up: patterns to remember

- **Conjugacy:** Beta–Binomial and Gamma–Poisson give posteriors in same family.
- **Sequential = batch:** updating doesn't depend on when data arrives.
- **Posterior predictive:** integrate out parameter; Gamma–Poisson $\Rightarrow$ Negative Binomial.
- **Transforms:** $\pi_Y(y) = \pi_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$.
- **Jeffreys prior:** $\pi(\lambda) \propto \sqrt{I(\lambda)}$ can be improper but yield proper posterior.