Mid-term revision by examples: Bayesian Updating, Predictive Distributions, Transformations

Remaining questions in problem sheet + worked solutions

Warm-up problem-sheet 2 Q3: True / False (set-up)

For each statement, decide if it is **True** or **False**.

(a) The likelihood function is proportional to the posterior distribution.

(b) A 99% credible interval captures 99% of the posterior probability.

(c) If a set of random variables are exchangeable, then we can reorder them without changing their joint distribution.

(d) Bayesian and frequentist methods always lead to significantly different estimates.

Q3(a): Question

Statement (a)

(a) The likelihood function is proportional to the posterior distribution.

Decide: **True** or **False**?

Q3(a): Answer

Answer

False.

Explanation

Posterior $\propto$ likelihood $\times$ prior:

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\,\pi(\theta).$$

So the likelihood alone is not proportional to the posterior unless the prior is constant (with care about support/properness).

Q3(b): Question

Statement (b)

(b) A 99% credible interval captures 99% of the posterior probability.

Decide: **True** or **False**?

Q3(b): Answer

Answer

True.

Explanation

By definition, a 99% credible interval C is constructed so that

$$\Pr(\theta \in C \mid y) = 0.99.$$

Q3(c): Question

Statement (c)

(c) If a set of random variables are exchangeable, then we can reorder them without changing their joint distribution.

Decide: **True** or **False**?

Q3(c): Answer

Answer

True.

Explanation

Exchangeability means the joint distribution is invariant under permutations of indices:

$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)}) \quad \text{for any permutation } \sigma.$$

Q3(d): Question

Statement (d)

(d) Bayesian and frequentist methods always lead to significantly different estimates.

Decide: **True** or **False**?

Q3(d): Answer

Answer

False.

Explanation

With large samples or vague priors (e.g. $\pi(\theta) \propto 1$ over the relevant region), Bayesian and frequentist conclusions often coincide (Bernstein–von Mises intuition).

Q3: Summary

Statement	True/False
(a) Likelihood $\propto$ posterior	**False**
(b) 99% credible interval has 99% posterior mass	**True**
(c) Exchangeable $\Rightarrow$ permutation invariance	**True**
(d) Bayes vs frequentist always very different	**False**

Q4: Statement of the problem

Given

Pareto with scale $\alpha = 1$ and shape β:

$$\pi(x \mid \alpha = 1, \beta) = \frac{\beta}{x^{\beta+1}}, \qquad x > 1,\ \beta > 0.$$

Data $y = \{y_1, \ldots, y_N\}$ are i.i.d. from this model.
Prior:

$$\pi(\beta) \propto \beta^{a-1} e^{-b\beta}.$$

Task

Derive the posterior $\pi(\beta \mid y)$ and identify its distribution.

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1)$$

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1) = \prod_{i=1}^{N} \frac{\beta}{y_i^{\beta+1}}$$

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1) = \prod_{i=1}^{N} \frac{\beta}{y_i^{\beta+1}} = \frac{\beta^N}{\prod_{i=1}^{N} y_i^{\beta+1}}.$$

Separate the β-dependent and constant parts

Write

$$\prod_{i=1}^{N} y_i^{\beta+1}$$

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1) = \prod_{i=1}^{N} \frac{\beta}{y_i^{\beta+1}} = \frac{\beta^N}{\prod_{i=1}^{N} y_i^{\beta+1}}.$$

Separate the β-dependent and constant parts

Write

$$\prod_{i=1}^{N} y_i^{\beta+1} = \left(\prod_{i=1}^{N} y_i\right)\left(\prod_{i=1}^{N} y_i^{\beta}\right).$$

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1) = \prod_{i=1}^{N} \frac{\beta}{y_i^{\beta+1}} = \frac{\beta^N}{\prod_{i=1}^{N} y_i^{\beta+1}}.$$

Separate the β-dependent and constant parts

Write

$$\prod_{i=1}^{N} y_i^{\beta+1} = \left(\prod_{i=1}^{N} y_i\right)\left(\prod_{i=1}^{N} y_i^{\beta}\right).$$

So

$$\pi(y \mid \beta, \alpha = 1) = \frac{\beta^N}{\left(\prod_{i=1}^{N} y_i\right)\left(\prod_{i=1}^{N} y_i^{\beta}\right)}$$

Q4: Likelihood (show the exact simplification)

Start from the i.i.d. product

$$\pi(y \mid \beta, \alpha = 1) = \prod_{i=1}^{N} \frac{\beta}{y_i^{\beta+1}} = \frac{\beta^N}{\prod_{i=1}^{N} y_i^{\beta+1}}.$$

Separate the β-dependent and constant parts

Write

$$\prod_{i=1}^{N} y_i^{\beta+1} = \left(\prod_{i=1}^{N} y_i\right)\left(\prod_{i=1}^{N} y_i^{\beta}\right).$$

So

$$\pi(y \mid \beta, \alpha = 1) = \frac{\beta^N}{\left(\prod_{i=1}^{N} y_i\right)\left(\prod_{i=1}^{N} y_i^{\beta}\right)} \propto \frac{\beta^N}{\prod_{i=1}^{N} y_i^{\beta}}$$

because $\prod_{i=1}^{N} y_i$ does not depend on β.

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Now simplify the product term:

$$\prod_{i=1}^{N} y_i^{-\beta}$$

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Now simplify the product term:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(\log \prod_{i=1}^{N} y_i^{-\beta} \right)$$

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Now simplify the product term:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(\log \prod_{i=1}^{N} y_i^{-\beta} \right) = \exp\left(\sum_{i=1}^{N} \log\left(y_i^{-\beta} \right) \right).$$

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Now simplify the product term:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(\log \prod_{i=1}^{N} y_i^{-\beta} \right) = \exp\left(\sum_{i=1}^{N} \log\left(y_i^{-\beta} \right) \right).$$

Use $\log\left(y_i^{-\beta} \right) = -\beta \log y_i$:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(-\beta \sum_{i=1}^{N} \log y_i \right).$$

Q4: Convert the product into an exponential using logs

From the previous slide:

$$\pi(y \mid \beta) \propto \beta^N \left(\prod_{i=1}^{N} y_i^{-\beta} \right).$$

Now simplify the product term:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(\log \prod_{i=1}^{N} y_i^{-\beta} \right) = \exp\left(\sum_{i=1}^{N} \log\left(y_i^{-\beta} \right) \right).$$

Use $\log\left(y_i^{-\beta} \right) = -\beta \log y_i$:

$$\prod_{i=1}^{N} y_i^{-\beta} = \exp\left(-\beta \sum_{i=1}^{N} \log y_i \right).$$

Likelihood kernel in β

$$\pi(y \mid \beta) \propto \beta^N \exp\left(-\beta \sum_{i=1}^{N} \log y_i \right).$$

Q4: Posterior and identification as Gamma

Multiply likelihood and prior kernels:

$$\pi(\beta \mid y) \propto \pi(y \mid \beta)\pi(\beta) \propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \left[\beta^{a-1} e^{-b\beta}\right].$$

Q4: Posterior and identification as Gamma

Multiply likelihood and prior kernels:

$$\pi(\beta \mid y) \propto \pi(y \mid \beta)\pi(\beta) \propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \left[\beta^{a-1} e^{-b\beta}\right].$$

Combine terms:

$$\pi(\beta \mid y) \propto \beta^{N+a-1} \exp\left(-\beta \left(b + \sum_{i=1}^{N} \log y_i\right)\right).$$

Q4: Posterior and identification as Gamma

Multiply likelihood and prior kernels:

$$\pi(\beta \mid y) \propto \pi(y \mid \beta)\pi(\beta) \propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \left[\beta^{a-1} e^{-b\beta}\right].$$

Combine terms:

$$\pi(\beta \mid y) \propto \beta^{N+a-1} \exp\left(-\beta \left(b + \sum_{i=1}^{N} \log y_i\right)\right).$$

Posterior

This is Gamma-shaped, so

$$\beta \mid y \sim \left(N + a,\ b + \sum_{i=1}^{N} \log y_i\right).$$

Q4: Posterior and identification as Gamma

Multiply likelihood and prior kernels:

$$\pi(\beta \mid y) \propto \pi(y \mid \beta)\pi(\beta) \propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \left[\beta^{a-1} e^{-b\beta}\right].$$

Combine terms:

$$\pi(\beta \mid y) \propto \beta^{N+a-1} \exp\left(-\beta \left(b + \sum_{i=1}^{N} \log y_i\right)\right).$$

Posterior

This is Gamma-shaped, so

$$\beta \mid y \sim \left(N + a,\ b + \sum_{i=1}^{N} \log y_i\right).$$

Key statistic

The data influence the posterior through the sufficient statistic $\sum_{i=1}^{N} \log y_i$ (given $\alpha = 1$).

Problem sheet 3 Q2: Poisson model + Gamma prior

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

(a) Prior $\lambda \sim \text{Gamma}(\alpha, \beta)$. 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:

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

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 $\alpha = 1$,

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

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 $\alpha = 1$,

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

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:

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

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.

Negative Binomial distribution $\mathrm{NegBin}(r, p)$: quick intro

Interpretation (one common parametrisation)

Run i.i.d. Bernoulli trials with success probability p. Stop when you have observed r successes. Let K be the **number of failures before the r-th success**. Then $K \sim \mathrm{NegBin}(r, p)$, with support $K \in \{0, 1, 2, \dots\}$.

PMF (matches the hint used in our question)

$$\Pr(K = k) = \binom{k + r - 1}{k}(1 - p)^k p^r = \frac{\Gamma(k + r)}{\Gamma(r)\, k!}(1 - p)^k p^r, \qquad k = 0, 1, 2, \dots$$

Mean / variance (for this parametrisation)

$$\mathbb{E}[K] = \frac{r(1 - p)}{p}, \qquad \mathrm{Var}(K) = \frac{r(1 - p)}{p^2}.$$

(Variance is larger than the mean unless $p = 1$, hence "overdispersed Poisson" behaviour.)

Why it appeared in our sheet

If $\tilde{Y} \mid \lambda \sim \mathrm{Pois}(\lambda)$ and $\lambda \mid y \sim \mathrm{Gamma}(a, b)$, then integrating out λ gives $\tilde{Y} \mid y \sim \mathrm{NegBin}\left(r = a,\ p = \frac{b}{b+1}\right)$ under the same $\mathrm{NegBin}(r, p)$ convention. :

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.