Bayesian Inference and Computation

Problem sheet 2

What we will practice today

- Turning i.i.d. models into a **likelihood function**.
- Finding **MLEs** by differentiating the **log-likelihood**.
- Recognising **conjugacy** and writing down the **posterior**.
- Computing **posterior expectations** and comparing to frequentist estimators.
- Building **credible intervals** from posterior quantiles (via R).

Big idea

Bayes: $\pi(\theta \mid y) \propto \pi(y \mid \theta)\, \pi(\theta)$

Problem 1: Geometric model setup

Assume

$$Y_1, \ldots, Y_N \mid p \overset{\text{i.i.d.}}{\sim} \mathrm{Geom}(p), \qquad p \in (0, 1],$$

with pmf

$$\pi(x \mid p) = (1-p)^{x-1} p, \qquad x \in \{1, 2, 3, \ldots\}.$$

We observe data $y = (y_1, \ldots, y_N)$.

Tasks

1. Derive the likelihood and the MLE $\hat{p}(y)$.
2. With $p \sim \mathrm{Beta}(\alpha, \beta)$, derive $\pi(p \mid y)$.
3. Compare MLE vs posterior mean; when do they match and what happens to the prior?

1(a) Likelihood for the geometric model

Start from independence:

$$\pi(y \mid p) = \prod_{i=1}^{N} \pi(y_i \mid p) = \prod_{i=1}^{N} (1-p)^{y_i - 1} p.$$

1(a) Likelihood for the geometric model

Start from independence:

$$\pi(y \mid p) = \prod_{i=1}^{N} \pi(y_i \mid p) = \prod_{i=1}^{N} (1-p)^{y_i - 1} p.$$

Collect terms: let $S = \sum_{i=1}^{N} y_i$.

$$\pi(y \mid p) = (1-p)^{\sum (y_i - 1)} \, p^N = (1-p)^{S-N} \, p^N.$$

1(a) Likelihood for the geometric model

Start from independence:

$$\pi(y \mid p) = \prod_{i=1}^{N} \pi(y_i \mid p) = \prod_{i=1}^{N} (1-p)^{y_i - 1} p.$$

Collect terms: let $S = \sum_{i=1}^{N} y_i$.

$$\pi(y \mid p) = (1-p)^{\sum (y_i - 1)} p^N = (1-p)^{S-N} p^N.$$

Likelihood function

$$L(p; y) \equiv \pi(y \mid p) \propto (1-p)^{S-N} p^N \qquad \text{(as a function of } p\text{)}.$$

1(a) MLE via log-likelihood

Take logs (monotone transform, so maximiser unchanged):

$$\ell(p) \equiv \log L(p; y) = (S - N)\log(1 - p) + N \log p.$$

1(a) MLE via log-likelihood

Take logs (monotone transform, so maximiser unchanged):

$$\ell(p) \equiv \log L(p; y) = (S - N) \log(1 - p) + N \log p.$$

Differentiate w.r.t. p:

$$\frac{d\ell}{dp} = (S - N) \cdot \frac{-1}{1 - p} + \frac{N}{p} = -\frac{S - N}{1 - p} + \frac{N}{p}.$$

1(a) MLE via log-likelihood

Take logs (monotone transform, so maximiser unchanged):

$$\ell(p) \equiv \log L(p; y) = (S - N)\log(1 - p) + N \log p.$$

Differentiate w.r.t. p:

$$\frac{d\ell}{dp} = (S - N) \cdot \frac{-1}{1 - p} + \frac{N}{p} = -\frac{S - N}{1 - p} + \frac{N}{p}.$$

Set to zero:

$$-\frac{S - N}{1 - p} + \frac{N}{p} = 0 \quad \Longrightarrow \quad \frac{N}{p} = \frac{S - N}{1 - p}.$$

1(a) MLE via log-likelihood

Take logs (monotone transform, so maximiser unchanged):

$$\ell(p) \equiv \log L(p; y) = (S - N)\log(1-p) + N \log p.$$

Differentiate w.r.t. p:

$$\frac{d\ell}{dp} = (S-N)\cdot\frac{-1}{1-p} + \frac{N}{p} = -\frac{S-N}{1-p} + \frac{N}{p}.$$

Set to zero:

$$-\frac{S-N}{1-p} + \frac{N}{p} = 0 \quad\Longrightarrow\quad \frac{N}{p} = \frac{S-N}{1-p}.$$

Solve for p:

$$N(1-p) = p(S-N) \implies N = pS \implies \boxed{\hat{p}_{\text{MLE}}(y) = \frac{N}{\sum_{i=1}^{N} y_i}}.$$

(You were told you do *not* need to verify it is a maximum.)

1(b) Prior and posterior: Beta conjugacy

Assume a Beta prior:

$$p \sim \mathrm{Beta}(\alpha, \beta), \qquad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}.$$

1(b) Prior and posterior: Beta conjugacy

Assume a Beta prior:

$$p \sim \mathrm{Beta}(\alpha, \beta), \qquad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}.$$

Bayes' rule:

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

1(b) Prior and posterior: Beta conjugacy

Assume a Beta prior:

$$p \sim \mathrm{Beta}(\alpha, \beta), \qquad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}.$$

Bayes' rule:

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

Substitute likelihood and prior:

$$\begin{aligned} \pi(p \mid y) &\propto (1-p)^{S-N} p^{N} \cdot p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{N+\alpha-1}(1-p)^{(S-N)+\beta-1}. \end{aligned}$$

1(b) Prior and posterior: Beta conjugacy

Assume a Beta prior:

$$p \sim \mathrm{Beta}(\alpha, \beta), \qquad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}.$$

Bayes' rule:

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

Substitute likelihood and prior:

$$\begin{aligned}\pi(p \mid y) &\propto (1-p)^{S-N} p^{N} \cdot p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{N+\alpha-1}(1-p)^{(S-N)+\beta-1}.\end{aligned}$$

Recognise Beta form:

$$\boxed{p \mid (Y = y) \;\sim\; \mathrm{Beta}\big(\alpha + N,\; \beta + S - N\big).}$$

Interpretation

α, β behave like **pseudo-counts**: they add to the data evidence.

1(c) Compare MLE to posterior mean

MLE: $\hat{p}_{\text{MLE}} = \dfrac{N}{S}.$

1(c) Compare MLE to posterior mean

MLE: $\hat{p}_{\text{MLE}} = \dfrac{N}{S}$.

Posterior mean (Beta): If $p \mid y \sim \mathrm{Beta}(\alpha + N, \beta + S - N)$, then

$$\mathbb{E}[p \mid y] = \frac{\alpha + N}{(\alpha + N) + (\beta + S - N)} = \boxed{\frac{\alpha + N}{\alpha + \beta + S}}.$$

1(c) Compare MLE to posterior mean

MLE: $\hat{p}_{\text{MLE}} = \dfrac{N}{S}.$

Posterior mean (Beta): If $p \mid y \sim \text{Beta}(\alpha + N, \beta + S - N)$, then

$$\mathbb{E}[p \mid y] = \frac{\alpha + N}{(\alpha + N) + (\beta + S - N)} = \boxed{\frac{\alpha + N}{\alpha + \beta + S}}.$$

When does $\mathbb{E}[p \mid y] = \hat{p}_{\text{MLE}}$?

Set

$$\frac{\alpha + N}{\alpha + \beta + S} = \frac{N}{S}.$$

Cross-multiply:

$$(\alpha + N)S = N(\alpha + \beta + S) \implies \alpha(S - N) = N\beta.$$

So (for a *fixed dataset*) one can match the MLE by choosing

$$\boxed{\beta = \alpha \frac{S - N}{}}$$

1(c) Compare MLE to posterior mean

MLE: $\hat{p}_{\text{MLE}} = \dfrac{N}{S}$.

Posterior mean (Beta): If $p \mid y \sim \text{Beta}(\alpha + N, \beta + S - N)$, then

$$\mathbb{E}[p \mid y] = \frac{\alpha + N}{(\alpha + N) + (\beta + S - N)} = \boxed{\frac{\alpha + N}{\alpha + \beta + S}}.$$

When does $\mathbb{E}[p \mid y] = \hat{p}_{\text{MLE}}$?

Set

$$\frac{\alpha + N}{\alpha + \beta + S} = \frac{N}{S}.$$

Cross-multiply:

$$(\alpha + N)S = N(\alpha + \beta + S) \implies \alpha(S - N) = N\beta.$$

So (for a *fixed dataset*) one can match the MLE by choosing

$$\boxed{\beta = \alpha \frac{S - N}{}}$$

1(c) Compare MLE to posterior mean

When does $\mathbb{E}[p \mid y] = \hat{p}_{\text{MLE}}$?

Set

$$\frac{\alpha + N}{\alpha + \beta + S} = \frac{N}{S}.$$

Cross-multiply:

$$(\alpha + N)S = N(\alpha + \beta + S) \implies \alpha(S - N) = N\beta.$$

So (for a *fixed dataset*) one can match the MLE by choosing

$$\boxed{\beta = \alpha \frac{S - N}{N}}.$$

1(c) Compare MLE to posterior mean

When does $\mathbb{E}[p \mid y] = \hat{p}_{\text{MLE}}$?

Set

$$\frac{\alpha + N}{\alpha + \beta + S} = \frac{N}{S}.$$

Cross-multiply:

$$(\alpha + N)S = N(\alpha + \beta + S) \implies \alpha(S - N) = N\beta.$$

So (for a *fixed dataset*) one can match the MLE by choosing

$$\boxed{\beta = \alpha \frac{S - N}{N}}.$$

Important: If you want the equality to hold *for all datasets*, the only way is the limiting/improper choice $\alpha = \beta = 0$ (see next slide).

1(c) What happens with $\alpha = \beta = 0$?

If $\alpha = \beta = 0$, then formally

$$\pi(p) \propto p^{-1}(1-p)^{-1},$$

which is $\mathrm{Beta}(0,0)$.

1(c) What happens with $\alpha = \beta = 0$?

If $\alpha = \beta = 0$, then formally

$$\pi(p) \propto p^{-1}(1-p)^{-1},$$

which is $\mathrm{Beta}(0, 0)$.

Key point

$\mathrm{Beta}(0, 0)$ is **not a proper prior**:

$$\int_0^1 \frac{1}{p(1-p)}\, dp = \infty.$$

So it is an **improper prior** (it does not integrate to 1).

1(c) What happens with $\alpha = \beta = 0$?

If $\alpha = \beta = 0$, then formally

$$\pi(p) \propto p^{-1}(1-p)^{-1},$$

which is $\mathrm{Beta}(0,0)$.

Key point

$\mathrm{Beta}(0,0)$ is **not a proper prior**:

$$\int_0^1 \frac{1}{p(1-p)}\, dp = \infty.$$

So it is an **improper prior** (it does not integrate to 1).

With this improper prior,

$$\mathbb{E}[p \mid y] = \frac{N}{S} = \hat{p}_{\mathrm{MLE}},$$

but you must be cautious: you are relying on an improper prior.

Practical note: Proper weakly-informative Betas (e.g. $\alpha = \beta = 1$ uniform, or $\alpha = \beta = \frac{1}{2}$ Jeffreys for Bernoulli-type problems) are often preferred.

Problem 2: Infectious period model

You observe $n = 100$ infected individuals.

$$\sum_{i=1}^{100} t_i = 870 \text{ days}, \qquad t_i > 0.$$

Model (clinician advice):

$$T_i \mid \theta \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(5, \theta) \quad (\text{shape} = 5, \text{ rate} = \theta).$$

Prior:

$$\theta \sim \mathrm{Exp}(0.01) \quad (\text{rate } 0.01).$$

Tasks

1. Derive $\pi(\theta \mid t)$.
2. Obtain a 95% credible interval using R.

2) Likelihood in θ (Gamma with known shape)

Gamma density (shape $k = 5$, rate θ):

$$f(t \mid \theta) = \frac{\theta^5}{\Gamma(5)} t^{5-1} \exp(-\theta t) = \frac{\theta^5}{\Gamma(5)} t^4 \exp(-\theta t).$$

2) Likelihood in θ (Gamma with known shape)

Gamma density (shape $k = 5$, rate θ):

$$f(t \mid \theta) = \frac{\theta^5}{\Gamma(5)} t^{5-1} \exp(-\theta t) = \frac{\theta^5}{\Gamma(5)} t^4 \exp(-\theta t).$$

Independence gives the likelihood:

$$\begin{aligned} \pi(\boldsymbol{t} \mid \theta) &= \prod_{i=1}^{n} \frac{\theta^5}{\Gamma(5)} t_i^4 e^{-\theta t_i} \\ &\propto \theta^{5n} \exp\left(-\theta \sum_{i=1}^{n} t_i\right), \end{aligned}$$

where we dropped terms not involving θ.

2) Likelihood in θ (Gamma with known shape)

Gamma density (shape $k = 5$, rate θ):

$$f(t \mid \theta) = \frac{\theta^5}{\Gamma(5)} t^{5-1} \exp(-\theta t) = \frac{\theta^5}{\Gamma(5)} t^4 \exp(-\theta t).$$

Independence gives the likelihood:

$$\begin{aligned}\pi(\boldsymbol{t} \mid \theta) &= \prod_{i=1}^{n} \frac{\theta^5}{\Gamma(5)} t_i^4 e^{-\theta t_i} \\ &\propto \theta^{5n} \exp\left(-\theta \sum_{i=1}^{n} t_i\right),\end{aligned}$$

where we dropped terms not involving θ.

Sufficient statistic

Only $\sum_{i=1}^{n} t_i$ matters for θ here.

2) Prior and posterior: Exp prior as Gamma

Exponential prior (rate 0.01):

$$\pi(\theta) = 0.01\, e^{-0.01\theta} \;\propto\; e^{-0.01\theta}.$$

2) Prior and posterior: Exp prior as Gamma

Exponential prior (rate 0.01):

$$\pi(\theta) = 0.01\, e^{-0.01\theta} \;\propto\; e^{-0.01\theta}.$$

This is a special case of Gamma:

$$\mathrm{Exp}(0.01) = \mathrm{Gamma}(1, 0.01) \quad \text{(shape 1, rate 0.01).}$$

2) Prior and posterior: Exp prior as Gamma

Exponential prior (rate 0.01):

$$\pi(\theta) = 0.01\, e^{-0.01\theta} \;\propto\; e^{-0.01\theta}.$$

This is a special case of Gamma:

$$\mathrm{Exp}(0.01) = \mathrm{Gamma}(1, 0.01) \quad \text{(shape 1, rate 0.01)}.$$

Posterior:

$$\begin{aligned}\pi(\theta \mid \boldsymbol{t}) &\propto \pi(\boldsymbol{t} \mid \theta)\pi(\theta) \\ &\propto \theta^{5n} \exp\left(-\theta \sum_{i=1}^{n} t_i\right) \cdot \exp(-0.01\theta) \\ &= \theta^{5n} \exp\left(-\theta \left(\sum_{i=1}^{n} t_i + 0.01\right)\right).\end{aligned}$$

Recognise Gamma kernel $\theta^{k-1}e^{-r\theta}$:

$$k - 1 = 5n \Rightarrow k = 5n + 1, \qquad r = \sum t_i + 0.01.$$

$$\boxed{\theta \mid \boldsymbol{t} \sim \mathrm{Gamma}\left(5n + 1, \ \sum_{i=1}^{n} t_i + 0.01\right).}$$

For $n = 100$ and $\sum t_i = 870$:

$$\boxed{\theta \mid \boldsymbol{t} \sim \mathrm{Gamma}(501, \ 870.01).}$$

2) 95% credible interval in R

A 95% equal-tailed credible interval uses posterior quantiles:

$$[q_{0.025},\ q_{0.975}], \quad \text{where } q_u = F^{-1}(u).$$

R code (rate parameterisation)

```
n <- 100
S <- 870
shape_post <- 5*n + 1        # 501
rate_post  <- S + 0.01       # 870.01

qgamma(c(0.025, 0.975), shape=shape_post, rate=rate_post)
```

2) 95% credible interval in R

A 95% equal-tailed credible interval uses posterior quantiles:

$$[q_{0.025},\ q_{0.975}], \quad \text{where } q_u = F^{-1}(u).$$

R code (rate parameterisation)

```
n <- 100
S <- 870
shape_post <- 5*n + 1        # 501
rate_post  <- S + 0.01       # 870.01

qgamma(c(0.025, 0.975), shape=shape_post, rate=rate_post)
```

For the given numbers, this returns approximately:

$$\boxed{(0.527,\ 0.627)}.$$

Interpretation: given the model and prior, there is 95% posterior probability that θ lies in this

Problem 3: True or False (concept check)

Decide if each statement is true or false:

1. The likelihood function is proportional to the posterior distribution.
2. A 99% credible interval captures 99% of the posterior probability.
3. If random variables are exchangeable, we can reorder them without changing their joint distribution.
4. Bayesian and frequentist methods always lead to significantly different estimates.

3) Solutions with brief justification

1. **False.**

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

   Posterior is proportional to *likelihood times prior*, not likelihood alone.

3) Solutions with brief justification

1. **False.**

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

   Posterior is proportional to *likelihood times prior*, not likelihood alone.

2. **True.** A 99% credible interval is constructed to contain 0.99 of posterior mass (e.g. equal-tailed, HPD, etc., depending on definition).

3) Solutions with brief justification

1. **False.**
$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\,\pi(\theta).$$
Posterior is proportional to *likelihood times prior*, not likelihood alone.
2. **True.** A 99% credible interval is constructed to contain 0.99 of posterior mass (e.g. equal-tailed, HPD, etc., depending on definition).
3. **True.** Exchangeability means the joint distribution is invariant under permutations:
$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)}) \quad \forall \text{ permutations } \sigma.$$

3) Solutions with brief justification

1. **False.**

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

   Posterior is proportional to *likelihood times prior*, not likelihood alone.

2. **True.** A 99% credible interval is constructed to contain 0.99 of posterior mass (e.g. equal-tailed, HPD, etc., depending on definition).

3. **True.** Exchangeability means the joint distribution is invariant under permutations:

$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)}) \quad \forall \text{ permutations } \sigma.$$

4. **False.** When n is large and the prior is weak (or regular), Bayesian and frequentist conclusions often *nearly coincide* (heuristically: posterior dominated by likelihood; formally: Bernstein–von Mises type results).

Problem 4: Pareto model with unknown shape

Pareto with scale $\alpha = 1$ and shape $\beta > 0$:

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

Data $y_1, \ldots, y_N$ i.i.d. from this model.
Prior:

$$\beta \sim \mathrm{Gamma}(a, b) \quad \text{(shape } a\text{, rate } b\text{)}.$$

Task

Derive the posterior $\pi(\beta \mid \boldsymbol{y})$.

4) Likelihood for Pareto shape β

Likelihood:

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

4) Likelihood for Pareto shape β

Likelihood:

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

Separate the β-dependent part:

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

The first bracket does not involve β (constant for the posterior kernel).

4) Likelihood for Pareto shape β

Likelihood:

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

Separate the β-dependent part:

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

The first bracket does not involve β (constant for the posterior kernel).
Rewrite the second bracket using logs:

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

So (up to constants):

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

4) Posterior: Gamma conjugacy

Prior:

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

4) Posterior: Gamma conjugacy

Prior:

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

Posterior kernel:

$$\begin{aligned}
\pi(\beta \mid \boldsymbol{y}) &\propto \pi(\boldsymbol{y} \mid \beta)\pi(\beta) \\
&\propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \cdot \left[\beta^{a-1} e^{-b\beta}\right] \\
&= \beta^{(N+a)-1} \exp\left(-\beta \left(b + \sum_{i=1}^{N} \log y_i\right)\right).
\end{aligned}$$

4) Posterior: Gamma conjugacy

Prior:

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

Posterior kernel:

$$\begin{aligned}
\pi(\beta \mid \boldsymbol{y}) &\propto \pi(\boldsymbol{y} \mid \beta)\pi(\beta) \\
&\propto \left[\beta^N e^{-\beta \sum \log y_i}\right] \cdot \left[\beta^{a-1} e^{-b\beta}\right] \\
&= \beta^{(N+a)-1} \exp\left(-\beta\left(b + \sum_{i=1}^{N} \log y_i\right)\right).
\end{aligned}$$

Recognise Gamma:

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

Sanity check

Summary: the patterns you should recognise

- **Likelihood from i.i.d.**: $\pi(\boldsymbol{y} \mid \theta) = \prod_i \pi(y_i \mid \theta)$.
- **MLE**: maximise $\log L(\theta)$ by differentiation.
- **Conjugacy**: prior $\times$ likelihood keeps the same family:
  - Geom/Bernoulli-type p with Beta prior $\Rightarrow$ Beta posterior.
  - Gamma rate parameter with Gamma/Exp prior $\Rightarrow$ Gamma posterior.
  - Pareto shape with Gamma prior $\Rightarrow$ Gamma posterior.
- **Credible intervals**: posterior quantiles (e.g. `qgamma`).

One-line Bayes

Posterior $\propto$ Likelihood $\times$ Prior.