Lab 3.8: Posterior Distributions
Conjugate and Non-Conjugate Priors

Grid approximation, log-likelihoods, and numerical normalisation

(Computer Lab)

Lab aim

What you will practise

- Working with posterior distributions when:
  - the prior *is* conjugate (closed-form posterior exists),
  - the prior *is not* conjugate (no simple closed form).
- Implementing likelihoods and posteriors in R using built-in density functions.
- Numerical normalisation of an unnormalised posterior on a grid (Trapezoidal rule).
- Extracting summaries: posterior mode (MAP), posterior mean, visual comparison.

Key idea

Even if we cannot write down a posterior in closed form, we can still compute with it numerically.

Conjugate priors (definition)

Definition (conjugacy)

A prior distribution $\pi(\theta)$ is a **conjugate prior** for a likelihood model $\pi(y \mid \theta)$ if the posterior $\pi(\theta \mid y)$ belongs to the same distributional family as the prior:

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

Conjugate priors (definition)

Definition (conjugacy)

A prior distribution $\pi(\theta)$ is a **conjugate prior** for a likelihood model $\pi(y \mid \theta)$ if the posterior $\pi(\theta \mid y)$ belongs to the same distributional family as the prior:

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

- Conjugacy is convenient: the posterior has a known form (e.g. Beta–Binomial, Gamma–Poisson).
- Sometimes conjugate priors are not appropriate (expert constraints, interpretability).
- Then we rely on computation: grids, numerical integration, simulation.

Example 3.12: Data (goals in 50 games)

Data

We observe total number of goals in each of 50 games:

```
y <- c(2, 6, 2, 3, 4, 3, 4, 3, 1, 2, 3, 2, 6, 6, 2, 3, 5, 1, 2, 2,
       4, 2, 5, 3, 6, 4, 1, 2, 7, 8, 4, 3, 7, 3, 3, 5, 2, 6, 1, 3,
       7, 4, 2, 6, 8, 8, 4, 5, 7, 4)

hist(y, main = "", xlab = "Number of goals scored")
mean(y)
```

Example 3.12: Data (goals in 50 games)

Data

We observe total number of goals in each of 50 games:

```
y <- c(2, 6, 2, 3, 4, 3, 4, 3, 1, 2, 3, 2, 6, 6, 2, 3, 5, 1, 2, 2,
       4, 2, 5, 3, 6, 4, 1, 2, 7, 8, 4, 3, 7, 3, 3, 5, 2, 6, 1, 3,
       7, 4, 2, 6, 8, 8, 4, 5, 7, 4)

hist(y, main = "", xlab = "Number of goals scored")
mean(y)
```

- The sample mean is about 3.92.
- A Poisson model is a natural first model for counts.

Poisson likelihood model

Model

$$Y_1, \ldots, Y_n \mid \lambda \overset{i.i.d.}{\sim} \text{Poisson}(\lambda), \qquad n = 50.$$

Poisson pmf:

$$\Pr(Y = y \mid \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y \in \{0, 1, 2, \ldots\}, \ \lambda > 0.$$

Poisson likelihood model

Model

$$Y_1, \ldots, Y_n \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Poisson}(\lambda), \qquad n = 50.$$

Poisson pmf:

$$\Pr(Y = y \mid \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y \in \{0, 1, 2, \ldots\}, \ \lambda > 0.$$

Likelihood

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

Computing likelihood and log-likelihood in R

Why log-likelihood?

Likelihood values can be extremely small (underflow), so we often compute on the log scale.

```
lambda <- seq(0, 10, 0.01) # grid

likelihood.function <- function(lambda, y) prod(dpois(y, lambda))
loglik.function <- function(lambda, y) sum(dpois(y, lambda, log = TRUE))

likelihood <- sapply(lambda, likelihood.function, y = y)
loglik <- sapply(lambda, loglik.function, y = y)

plot(lambda, likelihood, xlab = expression(lambda),
     ylab = "likelihood", type = "l")
plot(lambda, loglik, xlab = expression(lambda),
     ylab = "log-likelihood", type = "l")
```

Trapezoidal rule (normalising a grid posterior)

Trapezoidal rule

On a grid $a = x_0 < x_1 < \cdots < x_n = b$ with spacing h,

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\left(f(x_0) + f(x_n) + 2\sum_{i=1}^{n-1} f(x_i)\right).$$

Posterior on a grid + trapezoidal normalisation

```
lambda <- seq(0, 10, 0.01)
h <- 0.01

likelihood <- sapply(lambda, function(lam) prod(dpois(y, lam)))
prior <- dnorm(lambda, mean = 5, sd = 1)

post_unnorm <- likelihood * prior

Zhat <- (h/2) * (post_unnorm[1] + post_unnorm[length(post_unnorm)] +
               2 * sum(post_unnorm[2:(length(post_unnorm)-1)]))

posterior <- post_unnorm / Zhat

plot(lambda, posterior, type = "l",
     xlab = expression(lambda), ylab = "posterior density")

lambda[which.max(posterior)] # MAP on the grid
```

Exercise 3.1: show Gamma posterior

Prior

$$\lambda \sim \mathrm{Exp}(0.1) \quad \Rightarrow \quad \pi(\lambda) \propto e^{-0.1\lambda}.$$

Exercise 3.1: show Gamma posterior

Prior

$$\lambda \sim \mathrm{Exp}(0.1) \quad \Rightarrow \quad \pi(\lambda) \propto e^{-0.1\lambda}.$$

Posterior kernel

Poisson likelihood kernel: $e^{-n\lambda}\lambda^{\sum y_i}$. Then

$$\pi(\lambda \mid y) \propto \lambda^{\sum y_i} e^{-(n+0.1)\lambda},$$

so

$$\lambda \mid y \sim \mathrm{Gamma}\Big(\sum_{i=1}^{n} y_i + 1,\ n + 0.1\Big)$$

(shape, rate), with $n = 50$.

Exercise 3.2: posterior kernel (no closed form)

Model

$$Y_1, \ldots, Y_N \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Exp}(\lambda), \quad f(y \mid \lambda) = \lambda e^{-\lambda y}, \; y \geq 0.$$

Exercise 3.2: posterior kernel (no closed form)

Model

$$Y_1, \ldots, Y_N \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Exp}(\lambda), \quad f(y \mid \lambda) = \lambda e^{-\lambda y}, \; y \geq 0.$$

Prior

$$\lambda \sim \mathrm{Beta}(\alpha, \beta), \quad \lambda \in (0, 1).$$

Exercise 3.2: posterior kernel (no closed form)

Model

$$Y_1, \dots, Y_N \mid \lambda \overset{i.i.d.}{\sim} \mathrm{Exp}(\lambda), \quad f(y \mid \lambda) = \lambda e^{-\lambda y}, \; y \geq 0.$$

Prior

$$\lambda \sim \mathrm{Beta}(\alpha, \beta), \quad \lambda \in (0, 1).$$

Posterior (up to proportionality)

$$\pi(\lambda \mid y) \propto \lambda^{N+\alpha-1}(1-\lambda)^{\beta-1} \exp\Big(-\lambda \sum_{i=1}^{N} y_i\Big), \quad \lambda \in (0, 1).$$

Exercise 3.3: Beta–Binomial conjugacy

Model

$$X_1, \ldots, X_N \mid p \overset{i.i.d.}{\sim} \mathrm{Binomial}(100, p), \qquad p \sim \mathrm{Beta}(\alpha, \beta).$$

Let $T = \sum_{i=1}^{N} X_i$ and $M = 100N$.

Exercise 3.3: Beta–Binomial conjugacy

Model

$$X_1, \ldots, X_N \mid p \overset{i.i.d.}{\sim} \mathrm{Binomial}(100, p), \qquad p \sim \mathrm{Beta}(\alpha, \beta).$$

Let $T = \sum_{i=1}^{N} X_i$ and $M = 100N$.

Posterior

$$p \mid x \sim \mathrm{Beta}(\alpha + T,\ \beta + M - T).$$

Exercise 3.3: Beta–Binomial conjugacy

Model

$$X_1, \ldots, X_N \mid p \overset{i.i.d.}{\sim} \mathrm{Binomial}(100, p), \qquad p \sim \mathrm{Beta}(\alpha, \beta).$$

Let $T = \sum_{i=1}^{N} X_i$ and $M = 100N$.

Posterior

$$p \mid x \sim \mathrm{Beta}(\alpha + T,\ \beta + M - T).$$

Posterior mean

$$\mathbb{E}[p \mid x] = \frac{\alpha + T}{\alpha + \beta + M}.$$