Statistical Theory II Practical Labs

Lab 4: Bayesian Inference — Conjugate Models

Paulius Kazlauskas

Outline

Conjugate Models: The Big Idea

Beta–Binomial Model

Gamma–Poisson Model

Writing R Functions

R Reference

Conjugate Models: The Big Idea

What is a Conjugate Prior?

Definition

A prior is **conjugate** to a likelihood if the posterior belongs to the same distributional family as the prior.

Why it matters

Updating reduces to simple arithmetic on the parameters — no integration needed.

Likelihood	**Parameter**	**Conjugate prior**	**Posterior**
Binomial(n, π)	π	Beta(a_0, b_0)	Beta$(a_0 + y,\ b_0 + n - y)$
Poisson(μ)	μ	Gamma(r, v)	Gamma$(r + \sum y_i,\ v + n)$

Shared Architecture of Both Models

Both conjugate models follow **exactly the same logic**:

1. **Choose a prior** whose parameters match your prior belief (method of moments).
2. **Observe data** and update the parameters using the conjugate formula.
3. **Summarise the posterior:** mean, variance, credible interval.
4. **Decompose the posterior mean** into a weighted average of prior and data mean.
5. **Approximate with a normal** when posterior parameters are large (> 10).

The weights reveal how much the data dominate over prior beliefs.

Beta–Binomial Model

Beta–Binomial: Setup

Model

$\pi \sim \mathrm{Beta}(a_0, b_0)$ (prior)

$Y \mid \pi \sim \mathrm{Binomial}(n, \pi)$ (likelihood)

$\pi \mid y \sim \mathrm{Beta}(a_1, b_1)$ (posterior)

Update Rule

$$a_1 = a_0 + y \qquad b_1 = b_0 + n - y$$

Each success adds 1 to a; each failure adds 1 to b.

Posterior Summary

$$E(\pi \mid y) = \frac{a_1}{a_1 + b_1}$$

$$\mathrm{Var}(\pi \mid y) = \frac{a_1 b_1}{(a_1 + b_1)^2(a_1 + b_1 + 1)}$$

$$\mathrm{SD}(\pi \mid y) = \sqrt{\mathrm{Var}(\pi \mid y)}$$

Beta–Binomial: Posterior Mean Decomposition

Posterior mean as a weighted average

$$E(\pi \mid y) = \underbrace{\frac{a_0 + b_0}{a_0 + b_0 + n}}_{w_{\text{prior}}} \cdot \underbrace{\frac{a_0}{a_0 + b_0}}_{\text{prior mean}} + \underbrace{\frac{n}{a_0 + b_0 + n}}_{w_{\text{data}}} \cdot \underbrace{\frac{y}{n}}_{\text{data mean}}$$

Prior weight

$$w_{\text{prior}} = \frac{a_0 + b_0}{a_0 + b_0 + n}$$

Large when $a_0 + b_0 \gg n$: prior dominates.

Data weight

$$w_{\text{data}} = \frac{n}{a_0 + b_0 + n}$$

Large when $n \gg a_0 + b_0$: data dominate.

$a_0 + b_0$ is a *prior sample size*: how many pseudo-observations your beliefs are worth.

Eliciting a Beta Prior from Beliefs

Problem

Prior belief: mean $= \mu_0$, SD $= \sigma_0$. Find a_0, b_0 so $\mathrm{Beta}(a_0, b_0)$ matches.

Method of Moments

$$\mu_0 = \frac{a_0}{a_0 + b_0} \qquad \sigma_0^2 = \frac{a_0 b_0}{(a_0 + b_0)^2 (a_0 + b_0 + 1)}$$

$$\Rightarrow \quad a_0 + b_0 = \frac{\mu_0(1 - \mu_0)}{\sigma_0^2} - 1 \qquad a_0 = \mu_0(a_0 + b_0)$$

Women in economics (today's Q2)

$\mu_0 = 0.5,\ \sigma_0 = 0.1 \ \Rightarrow\ a_0 + b_0 = 24,\ a_0 = b_0 = 12$. Prior: $\pi \sim \mathrm{Beta}(12, 12)$.

Credible Intervals

What is a credible interval?

A $(1-\alpha)\times 100\%$ CI $[\mathrm{Low}, \mathrm{Up}]$ satisfies:

$$P(\mathrm{Low} \le \pi \le \mathrm{Up} \mid y) = 1-\alpha$$

A direct probability statement about the parameter.

Exact — Beta quantiles

$\mathrm{Low} = F^{-1}_{\mathrm{Beta}}(\alpha/2), \quad \mathrm{Up} = F^{-1}_{\mathrm{Beta}}(1-\alpha/2)$

`qbeta(alpha/2, a1, b1)`

Normal approximation

When $a_1 > 10$ and $b_1 > 10$:

$$E(\pi|y) \pm 1.96 \cdot \mathrm{SD}(\pi|y)$$

Credible $\neq$ Confidence interval

A credible interval is a probability statement about π. A frequentist CI is not.

Flat prior

Beta(1, 1) = Uniform(0, 1).

Posterior: Beta($1+y$, $1+n-y$).

Gamma–Poisson Model

Gamma–Poisson: Setup

Model

$\mu \sim \mathrm{Gamma}(r, v)$ (prior)

$Y_i \mid \mu \overset{\text{iid}}{\sim} \mathrm{Poisson}(\mu)$ (likelihood)

$\mu \mid \mathbf{y} \sim \mathrm{Gamma}(r_1, v_1)$ (posterior)

Update Rule

$$r_1 = r + \sum y_i \qquad v_1 = v + n$$

Posterior Summary

$$E(\mu \mid \mathbf{y}) = \frac{r_1}{v_1}$$

$$\mathrm{Var}(\mu \mid \mathbf{y}) = \frac{r_1}{v_1^2}$$

$$\mathrm{SD}(\mu \mid \mathbf{y}) = \sqrt{\frac{r_1}{v_1^2}}$$

Gamma–Poisson: Posterior Mean Decomposition

Posterior mean as a weighted average

$$E(\mu \mid \mathbf{y}) = \underbrace{\frac{v}{v+n}}_{w_{\text{prior}}} \cdot \underbrace{\frac{r}{v}}_{\text{prior mean}} + \underbrace{\frac{n}{v+n}}_{w_{\text{data}}} \cdot \underbrace{\frac{\sum y_i}{n}}_{\text{data mean}}$$

Prior weight

$$w_{\text{prior}} = \frac{v}{v+n}$$

v acts as a prior sample size.

Data weight

$$w_{\text{data}} = \frac{n}{v+n}$$

Large $n \Rightarrow$ data dominate.

Parallel with Beta–Binomial

Identical structure. Only the prior family and pseudo-sample-size differ.

Eliciting a Gamma Prior + Normal Approximation

Method of Moments

$$\mu_0 = \frac{r}{v} \qquad \sigma_0^2 = \frac{r}{v^2}$$

$$\Rightarrow \quad v = \frac{\mu_0}{\sigma_0^2} \qquad r = \frac{\mu_0^2}{\sigma_0^2}$$

Žalgiris (today)

$\mu_0 = 80, \ \sigma_0 = 10$

$v = 0.8, \quad r = 64$

Prior: $\mu \sim \mathrm{Gamma}(64, 0.8)$

Normal approximation

Approximate when r_1 is large:

$$\mu \mid \mathbf{y} \approx N(m_1, s_1^2)$$

$$\mathrm{CI} \approx m_1 \pm 1.96 \cdot s_1$$

where $m_1 = r_1 / v_1, \ s_1 = \sqrt{r_1} / v_1$.

Beta rule of thumb

Approximate when $a_1 > 10$ *and* $b_1 > 10$.

Writing R Functions

Defining a Function in R

```r
# Structure only -- you fill in the body
binbetaf <- function(a, b, n, y, alpha) {
  # compute: a1, b1, m_pos, var_pos, sd_pos,
  #          w_prior, w_data, low, up
  # print all results
  return(list(apos = a1, bpos = b1))
}
```

R Reference

R: Beta–Binomial Workflow

```r
a <- 12; b <- 12; n <- 1000; y <- 250; alpha <- 0.05
# Posterior parameters
a1 <- a + y;  b1 <- b + n - y          # 262, 762
# Posterior summaries
m_pos   <- a1 / (a1 + b1)
var_pos <- a1 * b1 / ((a1 + b1)^2 * (a1 + b1 + 1))
sd_pos  <- sqrt(var_pos)
# Weights
w_prior <- (a + b) / (a + b + n)       # 0.023
w_data  <- n / (a + b + n)             # 0.977
# Exact credible interval
qbeta(alpha / 2, a1, b1)
qbeta(1 - alpha / 2, a1, b1)
```

R: Gamma–Poisson Workflow

```r
y <- c(64, 72, 84, 73, 98, 85, 85, 94, 72, 93)
m <- 80;  s <- 10
# Prior elicitation
r <- m^2 / s^2;  v <- m / s^2          # 64, 0.8
# Posterior parameters
sum_y <- sum(y);  n <- length(y)        # 820, 10
r1 <- r + sum_y;  v1 <- v + n          # 884, 10.8
# Posterior summaries
pos_mean <- r1 / v1                     # 81.85
pos_var  <- r1 / v1^2
pos_sd   <- sqrt(pos_var)              # 2.75
# Exact 95% credible interval
qgamma(0.025, r1, v1)                   # 76.54
qgamma(0.975, r1, v1)                   # 87.33
# P(mu < 75)
pgamma(75, r1, v1)
```

Summary: Two Models, Same Logic

	Beta–Binomial	**Gamma–Poisson**
Prior	$\pi \sim \mathrm{Beta}(a_0, b_0)$	$\mu \sim \mathrm{Gamma}(r, v)$
Update	$a_1 = a_0 + y,\ b_1 = b_0 + n - y$	$r_1 = r + \sum y_i,\ v_1 = v + n$
Post. mean	$a_1/(a_1 + b_1)$	r_1/v_1
Post. var	$a_1 b_1/(a_1+b_1)^2(a_1+b_1+1)$	r_1/v_1^2
Prior weight	$(a_0+b_0)/(a_0+b_0+n)$	$v/(v+n)$
Data weight	$n/(a_0+b_0+n)$	$n/(v+n)$
Exact CI	`qbeta(·, a1, b1)`	`qgamma(·, r1, v1)`
Elicitation	$a_0+b_0 = \mu_0(1-\mu_0)/\sigma_0^2 - 1$	$v = \mu_0/\sigma_0^2,\ r = \mu_0^2/\sigma_0^2$