

QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Practical 6

2023-2024

Transformed parameters

Consider the `iris` dataset in R. Let y_1, \dots, y_n be the column `iris$Sepal.Length`, and assume that they are independent. Suppose that each $y_i \sim N(\mu, \sigma^2)$, where we assume μ as known, equal to 5, and let $\tau = 1/\sigma^2$. For τ , we assign a prior distribution that is `Gamma(2, 2)`. The posterior distribution is `Gamma(α, β)` for some α, β . Once you have found these parameters, define variables in R for the two parameters. Then generate a random sample of size 10,000 from this posterior distribution, and store it in the vector `tau_post`:

```
tau_post = ...
```

Calculate the mean and median of this random sample. Also calculate a 95% equal tail credible interval for τ .

τ is the precision of the distribution that is assumed to have generated the data. Suppose that we want to make inferences about the standard deviation

$$\sigma = \frac{1}{\sqrt{\tau}}$$

Using the posterior sample for τ , generate a sample from the posterior distribution for σ as follows

```
sigma_post = 1/sqrt(tau_post)
```

Once we have run this command, then `sigma_post` is a vector of the same length as `tau_post`.

Summarize the sample `sigma_post` to obtain an estimate of the posterior mean and median for σ , and a 95% equal tail credible interval.

How do each of the two point estimates and the interval endpoints compare to the corresponding estimates for τ ?

Simulation study: Comparing two binomials

We looked at this problem in lecture 5B as an example of simulating from a joint posterior for the Binomial case.

Suppose that the data are the outcome of a clinical trial of two treatments for a serious illness, the number of deaths after each treatment. Let the data be k_i deaths out of n_i patients, $i = 1, 2$ for the two treatments, and the two unknown parameters are q_1 and q_2 , the probability of death with each treatment.

Assuming a binomial model for each outcome, and independent $\text{Beta}(\alpha_i, \beta_i)$ priors for each parameter, the posterior distribution is

$$q_i \mid k_i \sim \text{Beta}(k_i + \alpha_i, n_i - k_i + \beta_i), \quad i = 1, 2.$$

We have independent prior distributions and likelihood, so the posterior distributions are also independent.

$$p(q_1, q_2 \mid k_1, k_2) = p(q_1 \mid k_1) p(q_2 \mid k_2) \propto p(k_1 \mid q_1) p(q_1) p(k_2 \mid q_2) p(q_2)$$

However, it is useful to think in terms of the joint posterior density for q_1 and q_2 , as then we can make probability statements involving both parameters. In this case, one quantity of interest is the probability $P(q_2 < q_1)$, i.e. does the second treatment have a lower death rate than the first.

We can approximate the probability by generating a sample of (q_1, q_2) pairs from the joint density. To generate the sample, we just need to generate each parameter from its beta posterior distribution, then once we have the sample, we just count what proportion of pairs has $q_2 < q_1$ to estimate $P(q_2 < q_1)$.

Take the observed data to be $n_1 = 240, k_1 = 15; n_2 = 280, k_2 = 12$. Assuming a uniform prior distribution for each of q_1 and q_2 , simulate two vectors in R, each of length 10,000, containing random samples from the posterior distributions of q_1 and q_2

```
Nsim = 10000
q_sim1 = ...
q_sim2 = ...
```

The following commands are two different ways of counting the proportion of times that each element of the vector `q_sim2` is less than the corresponding element of the vector `q_sim1`.

```
sum(q_sim2 < q_sim1)/length(q_sim1)
mean(q_sim2 < q_sim1)
```

This proportion is a Monte Carlo estimate of the posterior probability $P(q_2 < q_1)$.

Generate a posterior sample for the difference $d = q_1 - q_2$, using the two vectors already generated.

```
d = q_sim1 - q_sim2
```

Use this sample to calculate the posterior median for d , and an equal tail 95% credible interval.

The 95% credible interval can be calculated using either of the two ways shown in practical 5 to summarize the vector `post_lambda`. If we have a posterior sample `theta`, we can output the posterior median and credible interval together using this command

```
quantile(theta, probs=c(0.5, 0.025, 0.975))
```

We may also be interested in the risk ratio, that is $r = q_2/q_1$. Generate a posterior sample for r , and again find the posterior median for r , and a 95% credible interval. Draw a histogram of the posterior density for r . Calculate an estimate of the posterior probability $P(r < 0.5)$, i.e. the probability that the death rate on treatment 2 is less than half the death rate on treatment 1.