Lecture 5B
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture will

- Compute posterior distribution for transformed parameters and multiple parameters
- Compute posterior estimates and intervals for transformed parameters and multiple parameters.

Transformed parameters

- Suppose we have arrived at a posterior distribution $p(\theta \mid y)$ for a parameter θ.
- Let $\psi = g(\theta)$ be a monotonic, increasing transformation of θ, e.g., $\psi = \log(\theta)$, $\sqrt{\theta}$ or θ^3.
- **Questions:**
  - How do we make inferences about ψ?
  - Which posterior summary statements about θ carry over to ψ?
    -E.g. if $\tilde{\theta}$ is the posterior mean for θ, is $g(\tilde{\theta})$ the posterior mean for ψ?

Transforming random variables

- The shape of a probability density changes under nonlinear transformations of the random variable.
- Let g be a monotonic function.
- Suppose we have random variables X and Y with $Y = g(X)$.
- Their pdfs are related by

$$f_X(x) = |g'(x)| \, f_Y(g(x))$$

Example: Posterior of transformed parameters

- Bent coin with probability of success θ
- Flat prior on θ: $p(\theta) = 1$ for all $\theta \in [0, 1]$.
- $k = 5$ heads in $n = 6$ tosses.
- Find the posterior distribution of θ
- Find the posterior distribution of $\psi = \theta^3$.

Mean of transformed parameters

- Mean is not preserved by the transformation since for a nonlinear g

$$E(g(X)) \neq g(E(X)).$$

- The posterior density changes shape, so the mode is not preserved by the transformation.

- Also the endpoints of the highest posterior density credible intervals are not preserved.

Quantiles of transformed parameters

- Quantiles are preserved under nonlinear monotone transformations, so median is preserved.

- If θ_m be the posterior median for θ, then $g(\theta_m)$ is the posterior median for ψ.

- Similarly, equal tail credible intervals are preserved
- If $(q_{0.025}, q_{0.975})$ is a 0.95-credible intervals for θ, then $(g(q_{0.025}, g(q_{0.975}))$ is a 0.95-credible intervals for ψ.

More than one parameter

- We have covered one-parameter examples so far.
- We have considered conjugate priors - the simplest examples have one unknown parameter.
- Computational methods allow models with many parameters.
- And priors don't need to be conjugate.

Multiple parameters

- Let $\theta = (\theta_1, \ldots, \theta_K)$ be a vector of parameters.
- Then we can still use Bayes' theorem to compute the joint posterior

$$p(\theta_1, \ldots, \theta_K \mid y) \propto p(\theta_1, \ldots, \theta_K)\, p(y \mid \theta_1, \ldots, \theta_K)$$

- We still base our estimates on the joint posterior $p(\theta \mid y)$.
- For predictions of future data, we use the entire joint distribution.

Marginal distribution

- For point estimates of individual parameters, we typically use the marginal distribution.
- For example, if $\theta = (\theta_1, \theta_2, \theta_3)$, the marginal posterior distribution for θ_1 is
$$p(\theta_1 \mid y) = \int \int p(\theta_1, \theta_2, \theta_3 \mid y) \, d\theta_2 \, d\theta_3$$
- The computational methods used for Bayesian inference make going from joint to marginal distribution easy.
- No need to explicitly evaluate the integral.

Example: comparing two Binomials

- In a clinical trial, suppose we have n_1 control patients and n_2 treatment patients.
- x_1 control patients survive and x_2 treatment patients survive.
- Then

$$X_1 \sim \text{binomial}(n_1, p_1) \quad X_2 \sim \text{binomial}(n_2, p_2)$$

- We want to estimate $\tau = P(p_2 < p_1)$.

Example: comparing two Binomials

- The prior is $f(p_1, p_2) = 1, \quad 0 < p_1 < 1$, $0 < p_2 < 1$
- The posterior is

$$f(p_1, p_2 | x_1, x_2) \propto p_1^{x1}(1 - p_1)^{n1 - x1} p_2^{x2}(1 - p_2)^{n2 - x2}$$

- Notice that p_1, p_2 live on a square, and that

$$f(p_1, p_2 | x_1, x_2) \propto f(p_1 | x_1) f(p_2 | x_2),$$

  where $f(p_1 | x_1) = p_1^{x1}(1 - p_1)^{n1 - x1}$, $f(p_2 | x_2) = p_2^{x2}(1 - p_2)^{n2 - x2}$.
- Thus, p_1 and p_2 are independent under the posterior

Simulation

- Notice that $f(p_1|x_1) \sim \text{beta}(x_1 + 1, n_1 - x_1 + 1)$, $f(p_2|x_2) \sim \text{beta}(x_2 + 1, n_2 - x_2 + 1)$
- Let $P_{11}, \dots, P_{1B}$ a random sample from $\text{beta}(x_1 + 1, n_1 - x_1 + 1)$
- Let $P_{21}, \dots, P_{2B}$ a random sample from $\text{beta}(x_2 + 1, n_2 - x_2 + 1)$
- Then (P_{1i}, P_{2i}), $i = 1, \dots, B$ is a sample from $f(p_1, p_2|x_1, x_2)$.
- We estimate τ by counting the proportion of pairs (P_{1i}, P_{2i}) such that $P_{2i} < P_{1i}$