Lecture 5B
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture

- Review
- Compute posterior distribution for transformed parameters and multiple parameters
- Compute posterior estimates and credible intervals for transformed parameters and multiple parameters.
- Choose a prior distribution.

Review: point estimates

- Suppose we know the posterior distribution $p(\theta \mid y)$ for a one-dimensional parameter θ computed from

Posterior distribution $\propto$ prior distribution $\times$ likelihood

- We can obtain point estimates of θ by summarising the center of the posterior $p(\theta \mid y)$ using e.g.,
  - mean
  - median
  - mode

- We can also obtain a $1 - \alpha$-probability or credible interval for θ.

Review: Credible or probability intervals

- A $1 - \alpha$-probability or credible interval for θ is an interval (θ_L, θ_U) such that

$$P(\theta_L < \theta < \theta_U) = 1 - \alpha.$$

- The probabilities are calculated from the posterior distribution pmf or pdf

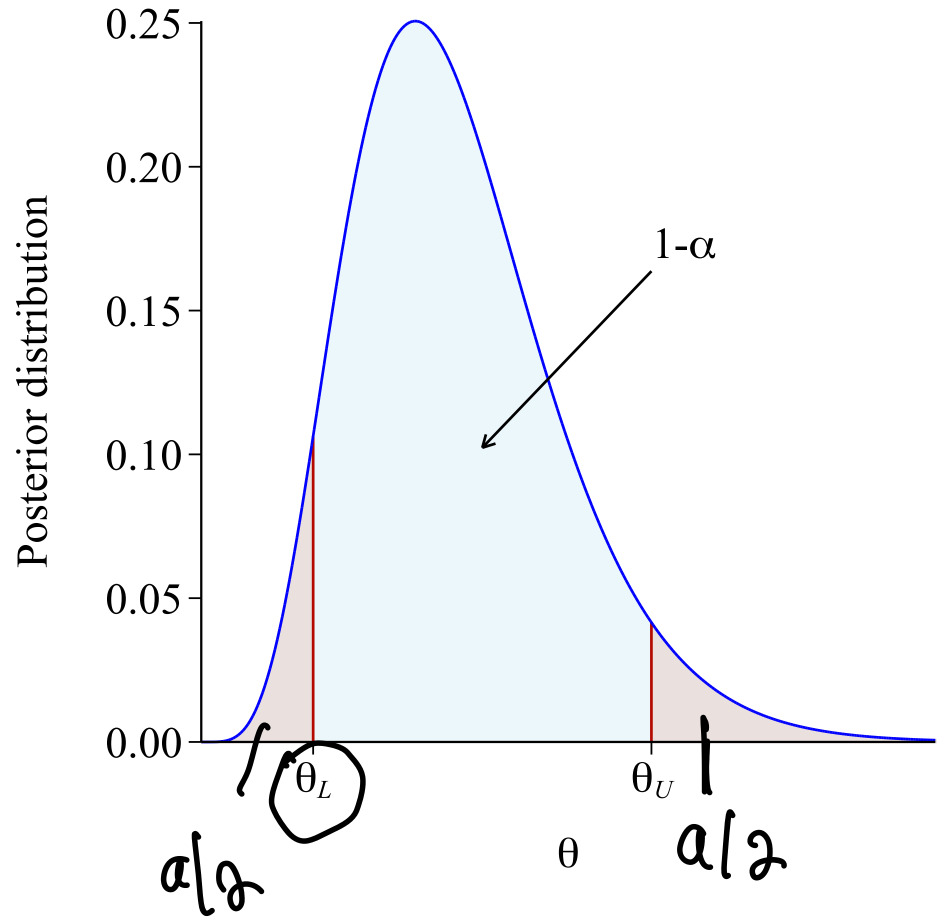
$$p(\theta \mid y)$$

Review: Equal tail intervals or symmetric probability intervals

- A (1-α)%equal-tail or symmetric probability interval is an interval (θ_L, θ_U) such that

$$P(\theta < \theta_L) = \alpha/2$$
$$P(\theta > \theta_U) = \alpha/2$$

- It's symmetric because the amount of probability remaining on either side of the interval is the same, $\alpha/2$.



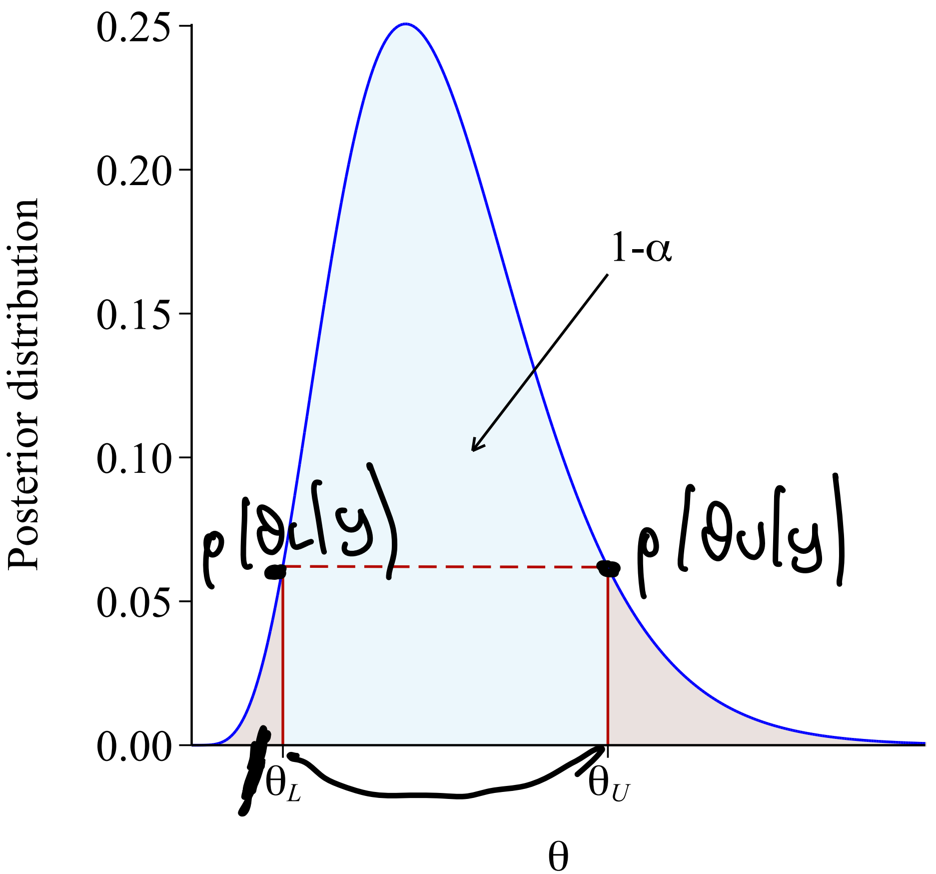
Review: Highest posterior density (HPD) intervals

- Let $p(\theta|y)$ be a unimodal posterior density for θ.
- A (1-α)% highest posterior density (HPD) interval is an interval (θ_L, θ_U) such that

$$P(\theta_L < \theta < \theta_U) = 1 - \alpha$$
$$p(\theta_L \mid y) = p(\theta_U \mid y)$$

- The interval captures the "most likely" values of the unknown, θ.

unimodal



Review: Highest posterior density (HPD) intervals

- Of all possible $(1-\alpha)\%$ credible intervals, the HPD interval is the **shortest.**
- If the density posterior density is unimodal and symmetric then the symmetric interval and the HPD interval coincide. Otherwise they do not.
- Finding a HPD interval in a non-symmetric distribution is not straightforward.

Transformed parameters

- Suppose we have arrived at a posterior distribution $p(\theta \mid y)$ for a parameter θ.

ψ is random

- Let $\psi = g(\theta)$, where g is a monotonic transformation of θ (increasing or decreasing), 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 monotonic transformations of the random variable.
- Let g be a monotonic function, and differentiable
- Suppose we have random variables X and Y with $Y = g(X)$.
- Their pdfs are related by

$$f_X(x) = |g'(x)|\, f_Y(g(x)) \quad \text{or}$$

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X(g^{-1}(y))$$

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 $\psi = \theta^3$.

The probability of observing 3 heads in the next 3 coin tosses

ψ is random

Solution

Last time we show that the posterior density, $p(\theta|x)$, of θ is beta$(6,2)$. Thus,

$$p(\theta|x) = \frac{\theta^5(1-\theta)}{B(6,2)} = 42\,\theta^5(1-\theta), \quad \theta \in [0,1]$$

We want to find the posterior density, $p(\psi|x)$, of $\psi = \theta^3 = g(\theta)$.

So $g(\theta) = \theta^3$ is monotone increasing with inverse $\theta = g^{-1}(\psi) = \psi^{1/3}$. The range of g is $[0,1]$, because

$\left[\psi = \theta^3 \Leftrightarrow \psi^{1/3} = \theta\right]$ $0 \le \theta^3 \le 1$ for all $0 \le \theta \le 1$. So $\psi \in [0,1]$.

Thus, the posterior density of ψ is

$$p(\psi|x) = \frac{d}{d\psi} g^{-1}(\psi) \cdot p_\theta\left(g^{-1}(\psi)\,|\,x\right)$$

$$= \frac{1}{3}\psi^{-2/3} \cdot p_\theta\left(\psi^{1/3}|x\right)$$

$$= \frac{1}{3}\psi^{-2/3} \cdot 42\,\psi^{5/3}(1-\psi)^{1/3}$$

$$= 14\,\psi\left(1-\psi^{1/3}\right), \quad 0 \le \psi \le 1$$

Mean of transformed parameters

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

$$\boxed{E(g(X)) \neq g(E(X)).}$$

- So, if $\hat{\theta}_B$ is the posterior mean of θ, $g(\hat{\theta}_B)$ is NOT the posterior mean of ψ, if g is nonlinear!

- 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.

When g is linear function, $g(x) = a + bx$, then

$$\mathbb{E}(Y) = \underline{\mathbb{E}(g(X))} = \mathbb{E}(a+bX)$$
$$= a + b\,\mathbb{E}(X)$$
$$= \underline{g(\mathbb{E}(X))}$$

- $X \sim U[0,1]$, $f_X(x) = 1 \ \forall x \in [0,1]$ $\quad \mathbb{E}(X) = \frac{1}{2}$

Let $Y = e^X = g(X)$.

$$\mathbb{E}(Y) = \mathbb{E}(e^X) \neq e^{\mathbb{E}(X)} = e^{1/2}$$

How do we compute $\mathbb{E}(Y) = \mathbb{E}(g(X))$

We use

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_0^1 g(x)\, f_X(x)\, dx$$
$$= \int_0^1 e^x \cdot 1\, dx = e^1 - 1$$

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 under increasing, one-to-one transformations
- If $(q_{0.025}, q_{0.975})$ is a 0.95 equal-tail credible intervals for θ, then $(g(q_{0.025}), g(q_{0.975}))$ is a 0.95 equal-tail credible intervals for $\psi = g(\psi)$ for any g monotonic increasing transformations.

. The median is preserved under non-linear monotonic transformations g.

Let θ_m be the posterior median of θ, that is

$$P(\theta \leq \theta_m | x) = 0.5$$

Let $\psi = g(\theta)$ where $g \uparrow$. Then,

$$0.5 = P(\theta \leq \theta_m | x) = P(g(\theta) \leq g(\theta_m) | x)$$
$$= P(\psi \leq g(\theta_m) | x)$$

$$\Rightarrow \quad P(\psi \leq g(\theta_m) | x) = 0.5$$

Thus $\psi_m = g(\theta_m)$ is the median of ψ.

Similarly, if g is $\downarrow$ you can show that

$$P(\psi \leq g(\theta_m) | x) = 0.5$$

. What about other quantiles?

Suppose that $q_{0.25}$ is the first quantile of θ. Then

$$0.25 = P(\theta \le q_{0.25} | x)$$

Let $\psi = g(\theta)$, where $g \uparrow$. Then

$$\begin{aligned} 0.25 &= P(\theta \le q_{0.25} | x) \\ &= P(g(\theta) \le g(q_{0.25}) | x) \\ &= P(\psi \le g(q_{0.25}) | x) \end{aligned}$$

So if g is increasing, $\psi_{0.25} = g(q_{0.25})$ is the 0.25-quantile of ψ.

$\Rightarrow$ The quantiles are preserved under monotone increasing transformations

Thus, if g is $\uparrow$ and

$\rightarrow (q_{0.025}, q_{0.975})$

is a 95% equal-tail credible interval for θ, then

$$(g(q_{0.025}), g(q_{0.975}))$$

is a 95% equal-tail credible interval for $\psi = g(\theta)$.

- $\psi = e^{\theta} = g(\theta)$ where $g(x) = e^{x} \uparrow$

$$\left(e^{q_{0.025}}, e^{q_{0.975}}\right)$$ is a 95% credible interval 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

prior × likelihood

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

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

(more about this next lectures

$$\theta = (\theta_1, \theta_2)$$

$$\mathbb{E}(\theta) = \iint \underbrace{\theta_1 \theta_2 \, p(\theta_1, \theta_2 | y)}_{\text{posterior mean}} \, d\theta_1 \, d\theta_2 .$$

$$\psi = g(\theta_1, \theta_2)$$

$$\mathbb{E}(\psi) = \iint \underbrace{g(\theta_1, \theta_2) \, p(\theta_1, \theta_2 | y)}_{\text{posterior mean of } \psi} \, d\theta_1 \, d\theta_2 .$$

- If we want to compute the marginal posterior density of θ_1, $p(\theta_1 | y)$, then

$$p(\theta_1 | y) = \int_{\theta_2} p(\theta_1, \theta_2 | y) \, d\underline{\theta_2} .$$

$$\boxed{f_Y(y) = \int_x f_{XY}(x, y) \, dx}$$

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 and x_2 are independent,

$$x_1 \sim \text{binomial}(n_1, p_1) \quad x_2 \sim \text{binomial}(n_2, p_2)$$

→ independent

- We want to estimate $\tau = P(p_2 < p_1)$, the survival success for treatment group is lower than the survival success for control group.
- We might want to estimate the difference in proportions $d = p_2 - p_1$ or the log-odds $\log(\frac{p_1}{1-p_1})$, p_1/p_2, the ratio of risks.

$\theta = (p_1, p_2) \rightarrow$ two-dimensional

Example: comparing two Binomials

- The prior is $f(p_1, p_2) = 1$, $\quad 0 < p_1 < 1$, $0 < p_2 < 1$
- By independence of the data, the posterior is

Binomial likelihood

$$\underbrace{f(p_1, p_2|x_1, x_2)}_{} \propto \underbrace{p_1^{x_1}(1-p_1)^{n_1-x_1}p_2^{x_2}(1-p_2)^{n_2-x_2}}_{} \quad f(p_1,p_2)$$

- 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 $f(p_1|x_1) = p_1^{x_1}(1-p_1)^{n_1-x_1}$, $f(p_2|x_2) = p_2^{x_2}(1-p_2)^{n_2-x_2}$.

- Thus, p_1 and p_2 are independent under the posterior
- Also $f(p_1|x_1) \sim \text{beta}(x_1 + 1, n_1 - x_1 + 1)$,
  $f(p_2|x_2) \sim \text{beta}(x_2 + 1, n_2 - x_2 + 1)$

My joint posterior density, $p(p_1, p_2 | x_1, x_2)$, is

$$p(p_1, p_2 | x_1, x_2) \propto p(p_1, p_2) \times p(x_1, x_2 | p_1, p_2)$$

$$= 1 \; p(x_1 | p_1) \, p(x_2 | p_2) \quad \text{[by independence of } x_1 \text{ and } x_2\text{]}$$

$$\propto p_1^{x_1} (1-p_1)^{n_1 - x_1} \; p_2^{x_2} (1-p_2)^{n_2 - x_2}$$

beta(x_1+1, n_1-x_1+1)

$$= p_1^{(x_1+1)-1} (1-p_1)^{(n_1-x_1+1)-1} \; p_2^{(x_2+1)-1} (1-p_2)^{(n_2-x_2+1)-1}$$

$$= p(p_1 | x_1) \times p(p_2 | x_2)$$

Thus p_1 and p_2 are independent under the posterior.

$$p_1 \sim \text{beta}(x_1+1, n_1 - x_1 + 1)$$

$$p_2 \sim \text{beta}(x_2+1, n_1 - x_2 + 1)$$

Simulation

- Let $P_{11}, \ldots, P_{1B}$ a random sample from beta$(x_1 + 1, n_1 - x_1 + 1)$.
- Let $P_{21}, \ldots, P_{2B}$ a random sample from beta$(x_2 + 1, n_2 - x_2 + 1)$.
- Then (P_{1i}, P_{2i}), $i = 1, \ldots, B$ 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 $P_{2i} < P_{1i}$

$$\hat{\tau} = \frac{1}{M} \sum_{i=1}^{M} \underbrace{I}_{\text{indicator function}}(P_{2i} < P_{1i})$$

$d = p_2 - p_1 = g(p_1, p_2)$ → This is random

We can estimate d by its posterior mean

$$\hat{d}_B = \underbrace{\iint g(p_1,p_2)\, p(p_1,p_2 \mid x_1,x_2)\, dp_1\, dp_2}_{\text{Difficult to do.}}$$

⟹ Use simulations

Using R,

$d_i = p_{2i} - p_{1i} \qquad i = 1, \ldots, B$

$d_i,\ i = 1, \ldots, B$ is a sample from d

I can estimate d by taking the sample mean of $\{d_1, \ldots, d_B\}$