Lecture 5A
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture will

- Review
- Compute Bayes point estimates given a pmf or pdf posterior distribution.
- Construct credible intervals given a pmf or pdf posterior distribution.

Review: Bayesian updating

Bayesian updating: Using Bayes' theorem to update a prior distribution to a posterior distribution given data and the likelihood.

- Observed data y come from $p(y \mid \theta)$, where θ is unknown.
- Prior distribution, $p(\theta)$ of θ (pmf or pdf).
- Likelihood: $p(y \mid \theta)$ (discrete or continuous)

Bayes' theorem

$$p(\theta \mid y) = \frac{p(\theta)\, p(y \mid \theta)}{p(y)}$$

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

Review: Conjugate priors

- A prior is conjugate to a likelihood, $p(y \mid \theta)$, if the posterior is the same type of distribution as the prior.

- Advantage: Bayesian updating reduces to modifying the parameters of the prior distribution.

Review: Examples of likelihood/conjugate prior pairs

	hypothesis	data	prior	likelihood	posterior
Bernoulli/Beta	$\theta \in [0,1]$	$x = 0$ or $x = 1$	Beta(α, β)	Bernoulli(θ)	Beta$(\alpha+1, \beta)$ or Beta$(\alpha, \beta+1)$
Binomial/Beta (fixed n)	$\theta \in [0,1]$	$x = k$	Beta(α, β)	binomial(n, θ)	Beta$(\alpha+k, \beta+n-k)$
Geometric/Beta	$\theta \in [0,1]$	$x = k$	Beta(α, β)	geometric(θ)	Beta$(\alpha+k, \beta+1)$
Normal/Normal (fixed σ^2)	$\theta \in \mathbb{R}$	x	$N(\mu_0, \sigma_0^2)$	$N(\theta, \sigma^2)$	$N(\mu_1, \sigma_1^2)$
Normal/gamma (fixed θ)	$\tau = 1/\sigma^2 > 0$	$x \in \mathbb{R}$	gamma(α, β)	$N(\theta, \sigma^2)$	gamma$(\alpha+0.5, \beta+0.5(x-\theta)^2)$
Exponential/Gamma	$\lambda > 0$	$x > 0$	gamma(α, β)	exponential(λ)	gamma$(1+\alpha, x+\beta)$

Board question

Which are conjugate priors for the following pairs likelihood/prior?

1. Exponential/Normal
2. Exponential/Gamma
3. Binomial/Normal

Normal example, both parameters unknown

- If μ and $\tau = 1/\sigma^2$ are unknown then there is a bivariate distribution which is conjugate.
- Marginal distribution

$$\tau \sim \text{Gamma}$$

and conditional distribution

$$\mu \mid \tau \sim \text{Normal}.$$

- The joint prior distribution is the product of these two.
- The posterior is of the same form.
- We're not going into details in this module.

Bayesian inference

- Data y come from $p(y \mid \theta)$, where θ is unknown.
- We have seen how to calculate the posterior distribution for parameter θ by

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

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

- In the Bayesian framework, all our inferences about θ are based on the posterior distribution $p(\theta \mid y)$.
- This includes point estimates.
- For a single parameter, we can summarize the posterior distribution just as we would normally summarize a distribution.

Point estimates

- Suppose we know the posterior distribution $p(\theta \mid y)$ for a one-dimensional parameter θ.
- We could summarise the center of the posterior $p(\theta \mid y)$ using e.g.,
  - mean
  - median
  - mode
- Mean or median are most common.
- Mode may be used if it's difficult to calculate mean or median.

Point estimates

Summaries of $p(\theta \mid y)$ as point estimates for θ.

- Posterior mean, for a pdf posterior density

$$\hat{\theta}_{\text{B}} = \int_{\theta} \theta p(\theta \mid y)\, d\theta$$

- Median, $\hat{\theta}_m$

$$P(\theta \leq \hat{\theta}_m | y) = 0.5.$$

- Mode or maximum a posteriori (MAP)

$$\hat{\theta}_{\text{MAP}} = \text{argmax}_{\theta} p(\theta \mid y).$$

Point estimates for Beta posterior pdf

Beta$(k+\alpha, n-k+\beta)$ posterior distribution.

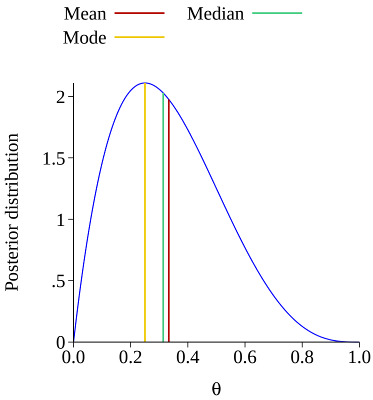
- Mean:

$$\frac{k+\alpha}{n+\alpha+\beta}$$

- Mode:

$$\frac{k+\alpha-1}{n+\alpha+\beta-2}$$

- No simple formula for median but we can use computer.



Quantile function

- For a RV Θ, let $F(\theta)$ be the cdf

$$P(\Theta \leq \theta) = F(\theta)$$

- If F is strictly increasing and continuous, then $F^{-1}(q)$, $q \in (0, 1)$ is the unique real number θ_q such that

$$F(\theta_q) = q$$

- We call θ_q the q-quantile of Θ.
- The quantile function is the inverse function of the cdf

$$Q = F^{-1}$$

- If $q = F(\theta_q)$ for some $q \in (0, 1)$, then $Q(q) = \theta_q$.

Quantile function

- E.g. if $q = 0.5$ and $m = \theta_{0.5} = F^{-1}(1/2)$ is the median,

$$F(\theta_{0.5}) = 0.5$$

$$Q(0.5) = \theta_{0.5}$$

- We call $F^{-1}(1/4)$ the first quantile and $F^{-1}(3/4)$ the third quantile.

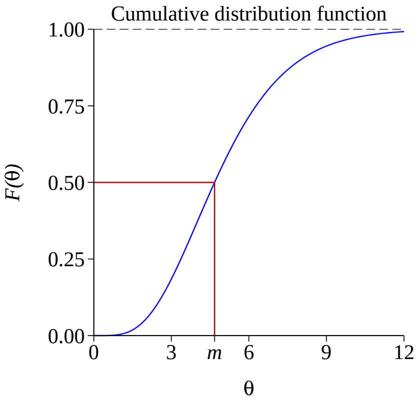
Finding the median

- Let $F(\theta)$ be the cdf

$$P(\Theta \leq \theta) = F(\theta)$$

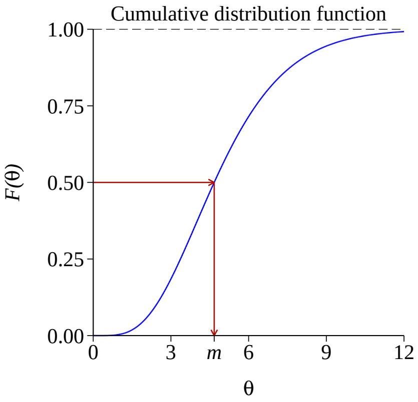
- If m is the median, then $F(m) = 0.5$.
- Half the probability mass is below, and half is above

$$P(\Theta \leq m) = 0.5$$



Finding the median

- So if we can find the inverse function of the cdf, we can find the median.
- The inverse of the cdf is called the quantile function.



Finding the median

- We have seen examples where the posterior distribution is in a well-known family of distributions.
- E.g. beta, gamma, or normal.
- Each one has a simple formula for the mean.
- For beta or gamma, there is no direct formula for the median (or the cdf).
- But we can use functions in R.
- E.g. for the gamma distribution `pgamma` returns the cdf and `qgamma` returns the quantile function (inverse of cdf).

Board question

Bent coin with unknown probability θ.
Flat prior: $p(\theta) = 1$ on $[0, 1]$
Data: toss 27 times and get 15 heads.

1. Find the posterior mean
2. Find the posterior median.
3. Find the MAP.

Uncertainty in parameters

- In Bayesian inference, any statements about uncertainty are based on the posterior distribution $p(\theta \mid y)$.
- For a single summary of uncertainty, we can calculate the posterior standard deviation.
- This is just the square root of the variance of the distribution.
- For example, for the beta($\alpha + k$, $\beta + n - k$) pdf, the posterior variance of θ is

$$\mathrm{var}(\theta \mid k) = \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}.$$

Confidence intervals

In frequentist inference (i.e. non-Bayesian inference), confidence intervals are used to express a range of uncertainty around a parameter estimate.

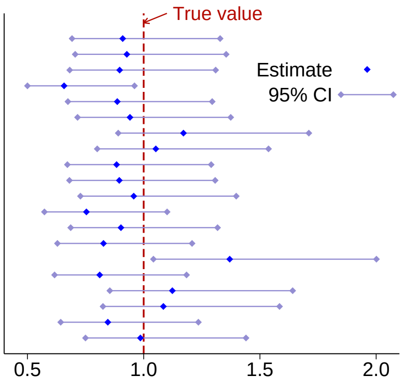
- Suppose random samples $Y = (Y_1, \ldots, Y_n)$ are repeatedly generated.
- For each sample we can estimate the true parameter θ by $\hat{\theta}(Y)$, and also construct an interval estimator $(\theta_L(Y), \theta_U(Y))$ based on the random sample $Y = (Y_1, \ldots, Y_n)$.
- A 95% confidence interval is an interval $(\theta_L(Y), \theta_U(Y))$ that covers θ with probability 0.95

$$P\left(\theta_L(Y) \leq \theta \leq \theta_U(Y)\right) = 0.95$$

- The probability 0.95 refers to the random interval $(\theta_L(Y), \theta_U(Y))$, and not the parameter and is called the coverage probability.

Confidence intervals illustrated

- Generate repeated samples from some distribution.
- Estimate $\hat{\theta}$ and a 95% confidence interval for $\hat{\theta}$ each time.
- 95% of the random intervals should contain the true value.



Interpretation of confidence intervals

- In classical statistics, it is NOT correct to say θ lies in the interval $(\theta_L(y), \theta_U(y))$ with probability 0.95 since θ is assumed to be fixed.

- The interval $(\theta_L(y), \theta_U(y))$ is one of the possible realised values of the random interval $(\theta_L(Y), \theta_U(Y))$ when $Y = y$, and since θ is fixed, θ is in $(\theta_L(y), \theta_U(y))$ with probability 0 or 1.

- **Long-run frequency interpretation.** With frequentist confidence intervals, when we say that the interval $(\theta_L(y), \theta_U(y))$ has 0.95 chance of coverage we only mean that, in the long run, with repeated sampling, the intervals trap the parameter θ 95% of the time.

Credible intervals

- In the Bayesian framework, we can say that θ lies inside the interval with some probability, not 0 or 1.
- θ is a random variable with a probability distribution.
- After seeing the data y, this is the posterior distribution

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

- As well as summarizing the posterior with a point estimate, we can directly calculate an interval for θ using the posterior distribution.
- They are called credible intervals or probability intervals.

Credible intervals

- For some $\alpha \in [0,1]$, a $100(1-\alpha)\%$ credible or probability interval for θ is an interval (θ_L, θ_U) such that

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

  E.g. $\alpha = 0.05$ for a 95% credible interval.
- More generally, (θ_L, θ_U) is a p-probability or credible interval for θ such that

$$P(\theta_L < \theta < \theta_U) = p$$

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

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

Credible intervals

- There are many ways to compute a p-credible interval.
- In particular, notice that the p-credible interval for θ is not unique.

- **Example:** Between the 0.05 and 0.55 quantiles is a 0.5 probability interval. Another 0.5-probability interval goes from 0.25 to the 0.75 quantiles.
- Thus we have 0.5 probability intervals $[\theta_{0.05}, \theta_{0.55}]$ and $[\theta_{0.25}, \theta_{0.75}]$.

Equal tail intervals or symmetric probability intervals

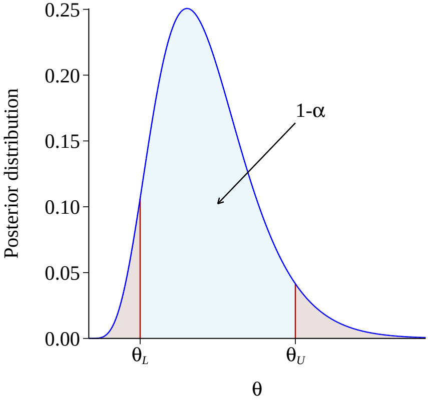
- Posterior pdf shown.
- $100(1-\alpha)\%$ interval.

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

- Equal probability outside each end.

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

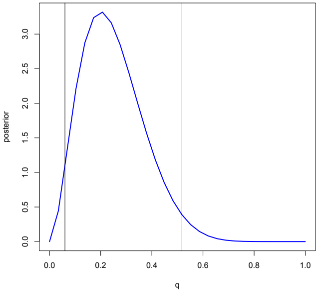
- **Example:** If $\alpha = 0.5$, the interval $[\theta_{0.25}, \theta_{0.75}]$ is symmetric because the amount of probability remaining on either side of the interval is the same, namely 0.25.



Board question: beta credible interval

Bent coin with unknown probability θ.
Flat prior: $p(\theta) = 1$ on $[0, 1]$
Data: toss 10 times and get 2 heads.

1. Use R to construct a symmetric 95% credible interval
2. `qbeta(c(0.025,0.975),shape1=3,shape2=9)`



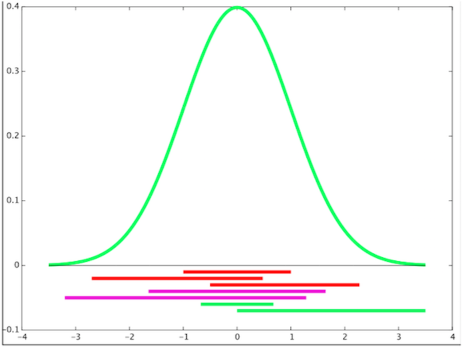
3. A beta(3,9) posterior distribution with vertical bars indicating a 95% probability interval.

Board question: Normal credible set

Let $x_1, \ldots, x_n$ an i.i.d from $N(\theta, \sigma^2)$ where σ^2 is known. Let θ have prior $N(\mu, \tau^2)$, where μ and τ are known.

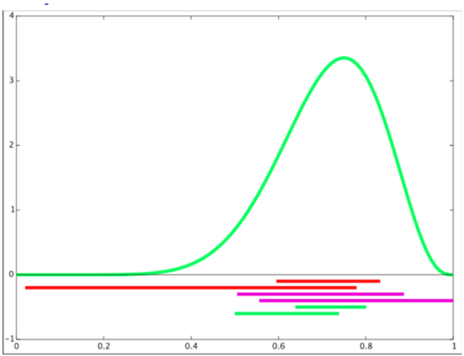
- Find a $1 - \alpha$ credible interval for θ.

Probability intervals for beta distributions



Red = 0.68, magenta = 0.9, green = 0.5

Probability intervals for normal distributions



Red = 0.68, magenta = 0.9, green = 0.5

Remarks

- For a fixed, p, different p-credible intervals for θ may have different $widths$.

- Since the width can vary for fixed p, a larger p does now always mean a larger width. But if a p_1-credible interval is fully contained in a p_2-credible interval, then p_1 is smaller than p_2.

- As in classical statistics, we can obtain a smallest credible interval by centering the interval under the highest part of the pdf posterior. Such an interval is called highest posterior density interval and is usually a good choice since it contains the most likely values.

Board question

To convert an 80% probability interval to a 90% interval should you shrink it or stretch it?

1. Shrink.
2. Stretch.

Highest posterior density (HPD) intervals

- If the posterior density $p(\theta|y)$ is unimodal, then for a given values of α, the 1- α- shortest credible interval for θ is given by

$$\{\theta : p(\theta|y) \geq k\},$$

where k is chosen so that

$$\int_{\{\theta : p(\theta|y) \geq k\}} p(\theta|y)\, d\theta = 1 - \alpha.$$

- The set $\{\theta : p(\theta|y) \geq k\}$ is called the highest posterior density (HPD) interval, as it consists of the values of the parameter θ for which the posterior density is highest.

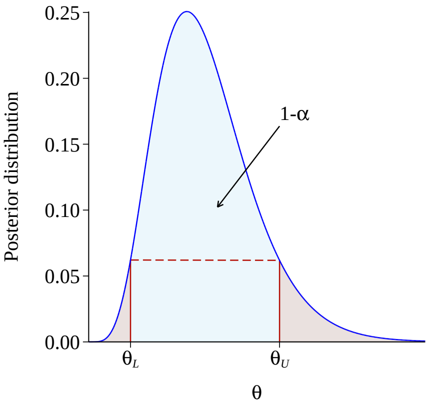
Highest posterior density (HPD) intervals

- Posterior pdf shown. We need to find θ_L and θ_U
- $100(1-\alpha)\%$ interval.

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

- Equal height to posterior density at θ_L and θ_U.

$$p(\theta_L \mid y) = p(\theta_U \mid y)$$



Calculating credible intervals

- Some textbooks emphasise the highest posterior density interval.
- However, it is usually difficult to calculate.
- The equal tail interval is easier to find computationally.
- For named distributions, just like for the median, we can use the quantile functions in R, `qgamma`, `qnorm` etc.

Suppose our posterior distribution for θ is Gamma(a, b).

Posterior median:

```
qgamma(0.5, shape=a, rate=b)
```

Equal tail 95% credible interval limits:

```
qgamma(0.025, shape=a, rate=b)
qgamma(0.975, shape=a, rate=b)
```