STAT 513 fa 2020 Lec 09 slides

Bayesian Inference

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Consider a parameter $\theta \in \Theta$.

- *Frequentist paradigm*: θ is a fixed constant.
- *Bayesian paradigm*: θ is a random variable.

The hierarchical model

Let $X_1, \dots, X_n$ be rvs (our data) and let θ be a rv in $\Theta \subset \mathbb{R}$. Assume

$$X_1, \dots, X_n | \theta \sim f(x_1, \dots, x_n | \theta)$$
$$\theta \sim \pi(\theta),$$

where

- $f(\cdot|\theta)$ is the joint pmf/pdf of $X_1, \dots, X_n$, conditional on θ (*data distribution*).
- $\pi(\cdot)$ is the marginal pmf/pdf of the parameter θ (*prior distribution*).

Posterior distribution

The *posterior distribution* of θ is the distribution of θ conditional on $X_1, \dots, X_n$.

The conditional pdf/pmf of $\theta | X_1, \dots, X_n$ is given by

$$\pi(\theta | X_1, \dots, X_n) = \begin{cases} \dfrac{f(X_1, \dots, X_n | \theta)\pi(\theta)}{\int_\Theta f(X_1, \dots, X_n | \tilde{\theta})\pi(\tilde{\theta}) d\tilde{\theta}} & \text{if } \theta \text{ is continuous} \\ \dfrac{f(X_1, \dots, X_n | \theta)\pi(\theta)}{\sum_{\tilde{\theta} \in \Theta} f(X_1, \dots, X_n | \tilde{\theta})\pi(\tilde{\theta})} & \text{if } \theta \text{ is discrete.} \end{cases}$$

Exercise: Derive the above.

Bayesian estimation with posterior mean

A typical Bayesian estimator of θ is the *posterior mean* of θ, which is

$$\hat{\theta}_{\text{Bayes}} = \mathbb{E}[\theta|X_1,\dots,X_n] = \begin{cases} \int_\Theta \theta\pi(\theta|X_1,\dots,X_n)d\theta & \text{if } \theta \text{ is continuous} \\ \sum_{\theta\in\Theta} \theta\pi(\theta|X_1,\dots,X_n) & \text{if } \theta \text{ is discrete.} \end{cases}$$

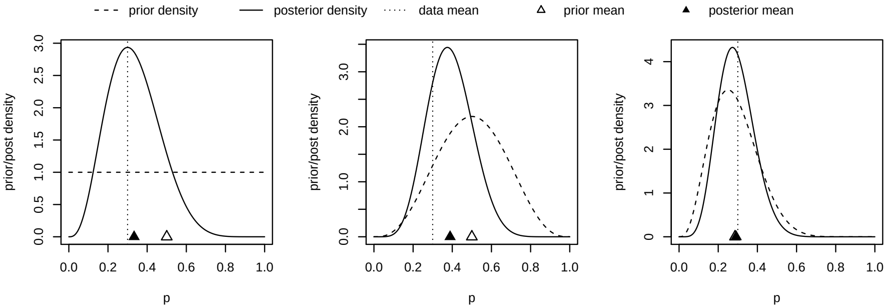
Exercise: Show that the above estimator is the solution to

$$\hat{\theta}_{\text{Bayes}} = \underset{a}{\text{argmin}} \quad \mathbb{E}[(\theta - a)^2|X_1,\dots,X_n].$$

Exercise: Suppose we have

$$Y|p \sim \text{Binomial}(n, p)$$
$$p \sim \text{Beta}(\alpha, \beta).$$

1. Find the posterior distribution of $p|Y$.
2. Find an expression for $\hat{p}_{\text{Bayes}} = \mathbb{E}[p|X_1, \dots, X_n]$.
3. Suppose $n = 10$ and $Y = 3$ is observed. Find the posterior mean of p under
   - $\alpha = 1$, $\beta = 1$
   - $\alpha = 4$, $\beta = 4$
   - $\alpha = 4$, $\beta = 10$



prior density
posterior density
data mean
prior mean
posterior mean
prior/post density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
p
prior/post density
0.0
1.0
2.0
3.0
0.0
0.2
0.4
0.6
0.8
1.0
p
prior/post density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
p

Shortcut to finding the posterior distribution

1. Find $g(\theta)$ such that $f(X_1, \ldots, X_n|\theta)\pi(\theta) \propto g(\theta)$.
2. Then $\pi(\theta|X_1, \ldots, X_n) = Cg(\theta)$, where $C = [\int_\Theta g(\theta)d\theta]^{-1}$.

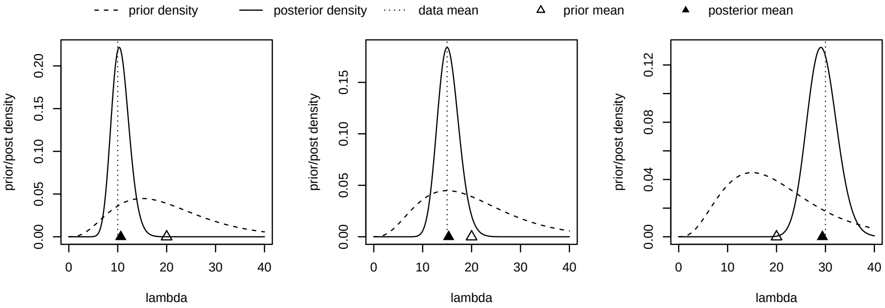
Can sometimes find C in Step 2 by identifying the distribution.

Exercise: Find the posterior distribution in the previous example with this strategy.

Exercise: Suppose

$$X_1, \dots, X_n | \lambda \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$$
$$\lambda \sim \text{Gamma}(\alpha, \beta).$$

1. Find the posterior distribution of $\lambda | X_1, \dots, X_n$.
2. Find an expression for $\hat{\lambda}_{\text{Bayes}} = \mathbb{E}[\lambda | X_1, \dots, X_n]$.
3. Under $\alpha = 4$ and $\beta = 5$ and a sample size of $n = 3$, compute the posterior mean of $\lambda | X_1, \dots, X_n$ when $\bar{X}_n = 10, 15, 30$.



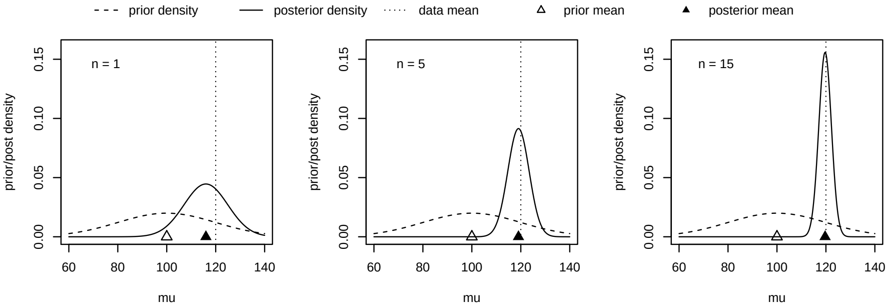
prior density
posterior density
data mean
prior mean
posterior mean
prior/post density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
lambda
prior/post density
0.00
0.05
0.10
0.15
0
10
20
30
40
lambda
prior/post density
0.00
0.04
0.08
0.12
0
10
20
30
40
lambda

Exercise: Let σ^2 be a known constant and suppose

$$\begin{aligned} Y_1, \dots, Y_n | \mu &\overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2) \\ \mu &\sim \text{Normal}(\mu_0, \tau^2), \end{aligned}$$

where $Y_1, \dots, Y_n$ are conditionally independent given μ.

1. Find the posterior distribution of $\mu | Y_1, \dots, Y_n$.
2. Find an expression for $\hat{\mu}_{\text{Bayes}} = \mathbb{E}[\mu | Y_1, \dots, Y_n]$.
3. Let $\sigma = 10$, $\tau = 20$, and $\mu_0 = 100$. Suppose that under sample sizes of $n = 1, 5, 15$, the sample mean $\bar{Y}_n = 120$ is observed. Compute the posterior mean of $\mu | Y_1, \dots, Y_n$ in each case.



prior density
posterior density
data mean
prior mean
posterior mean
n = 1
n = 5
n = 15
prior/post density
0.00
0.05
0.10
0.15
60
80
100
120
140
mu

Conjugacy

In the Bayesian setup, a *conjugate prior* is a prior distribution under which the posterior distribution belongs to the same family of distributions.

Examples we have seen:

- In Beta-Binomial model, posterior was Beta.
- In Gamma-Poisson model, posterior was Gamma.
- In Normal-Normal model, posterior was Normal.

We do not *need* to choose a conjugate prior; it just makes hand calculations easier.

Conjugacy mostly for textbook examples. In real life we use  power.

A $(1-\alpha)100\%$ frequentist conf. int. is any interval (L_n, U_n) such that

$$P(L_n < \theta < U_n) = 1 - \alpha,$$

where L_n and U_n are based on data $X_1, \dots, X_n$ and θ is fixed.

Bayesian credible interval

A $(1-\alpha)100\%$ *Bayesian credible interval* is any interval (L_n, U_n) such that

$$P(L_n < \theta < U_n | X_1, \dots, X_n) = 1 - \alpha,$$

where L_n and U_n are based data $X_1, \dots, X_n$ and θ is a random variable.

Two standard constructions of Bayesian credible intvls are choosing L_n and U_n st

1. *Equal tails interval*:

$$P(\theta < L_n | X_1, \dots, X_n) = P(\theta > U_n | X_1, \dots, X_n) = \alpha/2.$$

2. *Highest posterior density interval*:

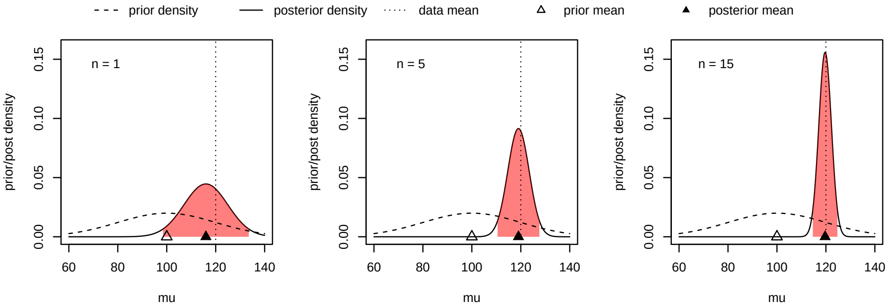
$$(L_n, U_n) \text{ is smallest interval st } P(L_n < \theta < U_n | X_1, \dots, X_n) = 1 - \alpha.$$

*If posterior is symmetric, the two constructions yield the same interval.

Exercise: Compute 95% Bayesian credible intervals for μ in the model

$$\begin{aligned} Y_1, \ldots, Y_n | \mu &\overset{\text{ind}}{\sim} \text{Normal}(\mu, 10^2) \\ \mu &\sim \text{Normal}(100, 20^2), \end{aligned}$$

when $\bar{Y}_n = 120$ under the sample sizes $n = 1, 5, 15$.


prior density
posterior density
data mean
prior mean
posterior mean
n = 1
n = 5
n = 15
prior/post density
0.00
0.05
0.10
0.15
60
80
100
120
140
mu

For $\theta \in \Theta$, consider testing H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$ based on $X_1, \ldots, X_n$.

Prior and posterior probabilities

The *prior probabilities* of H_0 and H_1 are

$$\pi_0 = P(\theta \in \Theta_0) \quad \text{and} \quad \pi_1 = P(\theta \in \Theta_1).$$

The *posterior probabilities* of H_0 and H_1 conditional on $X_1, \ldots, X_n$ are

$$p_0 = P(\theta \in \Theta_0 | X_1, \ldots, X_n) \quad \text{and} \quad p_1 = P(\theta \in \Theta_1 | X_1, \ldots, X_n).$$

For $\theta \in \Theta$, consider testing H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$ based on $X_1, \dots, X_n$.

Prior and posterior odds and the Bayes factor

- The *prior odds* in favor of H_0 over H_1 are π_0/π_1.
- The *posterior odds* in favor of H_0 over H_1 given $X_1, \dots, X_n$ are p_0/p_1.

The *Bayes factor* in favor of H_0 over H_1 is the odds ratio

$$B = \frac{p_0/p_1}{\pi_0/\pi_1}.$$

The Bayes factor represents how much the data have tipped the odds towards H_0:

$$\textit{posterior odds} = B \times \textit{prior odds}$$

$$B = \frac{p_0/p_1}{\pi_0/\pi_1}.$$

- A ________________(large/small) Bayes factor indicates that the data carry ________________(much/little) evidence in favor of H_0 over H_1.
- If the Bayes factor is ________________(less than/greater than) 1, the data have changed our prior beliefs in favor of ________________(H_0/H_1).

Exercise: Suppose

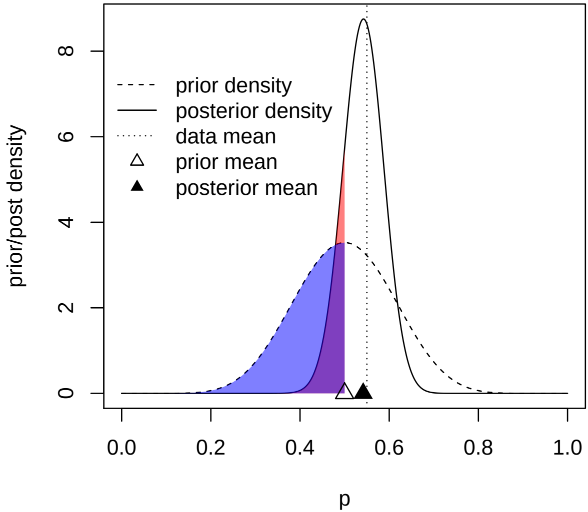
$$Y|p \sim \text{Binomial}(n, p)$$
$$p \sim \text{Beta}(\alpha, \beta)$$

and consider testing

$$H_0\colon p \leq 1/2 \text{ versus } H_1\colon p > 1/2.$$

Let $n = 100$ and let $Y = 55$ be observed, and set $\alpha = 10$ and $\beta = 10$.

1. Find the posterior probability p_0 of H_0, that is of the event $p \leq 1/2$.
2. Find $P(Y \geq 55|p = 1/2)$ and interpret it from the frequentist perspective.
3. Find the prior odds in favor of H_0 over H_1.
4. Find the posterior odds in favor of H_0 over H_1.
5. Compute the Bayes factor of the data in favor of H_0 over H_1.



prior density
posterior density
data mean
prior mean
posterior mean
prior/post density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
p