Advanced Methods in Biostatistics II

Lecture 14

December 12, 2017

Bayesian Inference

- Most statistical methods covered in introductory statistics courses are frequentist (or classical) methods.
- Bayesian inference is an alternative approach that provides a somewhat different perspective.
- Bayesian inference is the process of inductive learning using Bayes' rule.

Classical vs Bayesian Approach

- The frequentist (or classical) point of view:
  - Probabilities describe long run relative frequencies.
  - Parameters are fixed unknown constants. Because they do not fluctuate, no useful probability statements can be made about them.
  - Statistical procedures should be designed to have well-defined long run frequency properties (e.g., hypothesis tests and confidence intervals).

Classical vs Bayesian Approach

- The Bayesian point of view:
  - Probabilities describe a degree of belief.
  - Probability statements can be made about parameters, even though they are fixed constants.
  - Inferences are made about a parameter θ by producing a probability distribution for it.

Bayesian statistics

- In Bayesian analysis, one chooses a density $p(\theta)$, the prior, that expresses our beliefs about θ before we see any data.
- Then one chooses a statistical model $p(\mathbf{y}|\theta)$, the likelihood, that reflects our belief about $\mathbf{y}$ given θ.
- After observing $\mathbf{y}$, we update our beliefs and calculate the posterior distribution $p(\theta|\mathbf{y})$.

Bayes calculations

- Updating is is done as follows:

$$
\begin{aligned}
p(\text{Param}|\text{Data}) &= \frac{p(\text{Param}, \text{Data})}{p(\text{Data})} \\
&\propto p(\text{Data}|\text{Param}) \times p(\text{Param}) \\
&= \text{Likelihood} \times \text{Prior}.
\end{aligned}
$$

- The posterior distribution is used for subsequent inference.

The Prior distribution

- The prior distribution is a subjective distribution, based on the experimenter's belief and is formulated prior to viewing the data.
- In the Bayesian framework the choice of prior is crucial.
- If no prior information about the parameters are available, non-informative priors can be used. These types of priors let the data 'speak for itself'.
- One can also choose the priors in such a way that the posterior lies in the same family of distributions as the prior (conjugate priors).

The Posterior distribution

- The posterior distribution contains all current information about the parameter θ.
- Numerical summarizes (e.g., mean, median, mode) of the distribution are used to obtain point estimates of the parameter.
- We can also make probability statements about the parameter of interest and create posterior intervals.

Credible intervals

- For example, a credible interval is the Bayesian analogue of a confidence interval.
- Given a posterior distribution on a parameter θ, a $1 - \alpha$ credible interval $[L, U]$ is an interval such that:

$$P(L \leq \theta \leq U \mid \mathbf{y}) = 1 - \alpha.$$

The Binomial model

- Consider a series of coin flips, $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$.
- The likelihood associated with this experiment is

$$\begin{aligned} p(x_1, \ldots x_n | \theta) &\propto \theta^{\sum_i x_i} (1-\theta)^{n - \sum_i x_i} \\ &= \theta^x (1-\theta)^{n-x} \end{aligned}$$

  where $x = \sum_i x_i$.
- Notice the likelihood depends only on the total number of successes.

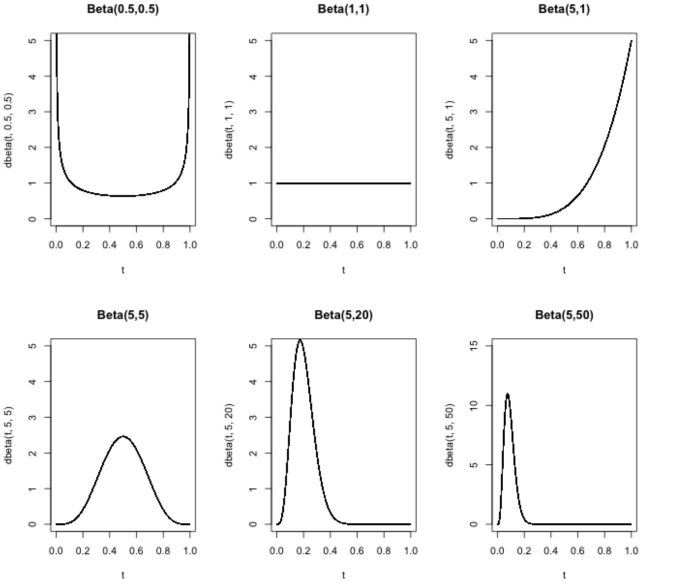
The Binomial model

- Consider putting a Beta(α, β) prior on θ.
- This can be written as follows:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

  for $0 \leq \theta \leq 1$, and shape parameters $\alpha, \beta > 0$.
- Here the terms α and β are refered to as hyperparameters.

Beta distribution



The Binomial model

- The posterior distribution is given by

$$\begin{aligned} p(\theta \mid x_1, \ldots x_n) &\propto p(x_1, \ldots x_n|\theta) \times p(\theta) \\ &\propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \end{aligned}$$

- Hence, the posterior distribution is Beta$(x+\alpha, n-x+\beta)$.

The Binomial model

- The posterior mean is

$$\begin{aligned} E[\theta \mid x_1, \ldots x_n] &= \frac{x+\alpha}{n+\alpha+\beta} \\ &= \delta\frac{x}{n} + (1-\delta)\frac{\alpha}{\alpha+\beta} \end{aligned}$$

- Therefore, the posterior mean is a weighted average of the MLE ($\hat{p} = x/n$) and the prior mean $\tilde{p} = \alpha(\alpha+\beta)^{-1}$.
- The weight is given by the term:

$$\delta = \frac{n}{n+\alpha+\beta}.$$

The Binomial model

- Note, as $n \to \infty$ for fixed α and β, $\delta \to 1$ and the MLE dominates.
- That is, as we collect more data, the prior becomes less relevant and the data, in the form of the likelihood, dominates.
- In contrast, for fixed n, as either α or β go to infinity (or both), the prior dominates ($\delta \to 0$).
- For the Beta distribution α or β going to infinity makes the distribution much more peaked.
- Thus, if we are more certain of our prior distribution, the data matters less.

Conjugate prior

- In Bayesian analysis, if the posterior is in the same family as the distribution, the prior is called a conjugate prior for the likelihood function.
- The Beta distribution is a conjugate prior for the Binomial distribution, as it gives rise to a posterior that follows a Beta distribution.

The Poisson model

- Now, let $X_1, \ldots, X_n \sim$ Poisson(θ).
- The likelihood associated with this experiment is

$$\begin{aligned} p(x_1, \ldots x_n | \theta) &\propto \theta^{\sum_i x_i} e^{-n\theta} \\ &\propto \theta^{x} e^{-n\theta} \end{aligned}$$

where $x = \sum_i x_i$.

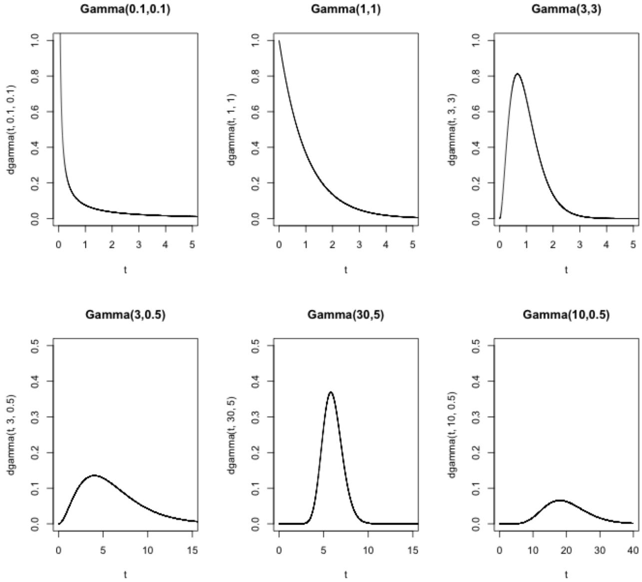
The Poisson model

- Consider putting a Gamma(α, β) prior on θ.
- This can be written as follows:

$$p(\theta) = \frac{\beta^{\alpha}\theta^{\alpha-1}e^{-\beta\theta}}{\Gamma(\alpha)}$$

for $\theta > 0$, and shape parameters α, $\beta > 0$.

Gamma distribution



The Poisson model

- The posterior is the given by

$$
\begin{aligned}
p(\theta \mid x_1, \ldots x_n) &\propto \theta^x e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \theta^{x+\alpha-1} e^{-(\beta+n)\theta}.
\end{aligned}
$$

- Thus the posterior is Gamma$(x + \alpha, \beta + n)$.
- The Gamma distribution is a conjugate prior for the Poisson distribution.

The Poisson model

- The posterior mean is:

$$
\begin{aligned}
E[\lambda \mid x] &= \frac{x+\alpha}{\beta+n} \\
&= \delta\frac{x}{n} + (1-\delta)\frac{\alpha}{\beta}
\end{aligned}
$$

  where x/n is the MLE (the observed rate) and a/b is the prior estimate.

- In this case

$$\delta = \frac{n}{\beta+n}.$$

- As $n \to \infty$ the MLE dominates.

Posterior inference

- The posterior distribution is the basis for all Bayesian inference.
- In the previous examples our use of conjugate priors lead to simple expressions for the posterior means and variances.
- However, even if the posterior is known, it is sometimes difficult to obtain exact values of certain posterior quantities.

Monte Carlo methods

- For example, we may want to calculate the probability that $\theta \in A$ for some arbitrary set A, or alternatively the posterior distribution of some function of the parameters (e.g., $\max\{\theta_1, \ldots \theta_n\}$).
- Obtaining exact values for these posterior quantities can be difficult.
- By generating random samples from the posterior, all quantities of interest can be approximated using Monte Carlo methods.

Monte Carlo methods

- Let $g(\theta)$ be some function of θ.
- Suppose we want to estimate $E(g(\theta) \mid \mathbf{x})$.
- Generate an i.i.d sequence $\theta_1, \ldots \theta_N$ from the posterior distribution of θ.
- Estimate $E(g(\theta) \mid \mathbf{x})$ using

$$\bar{g} = \frac{1}{N} \sum_{i=1}^{N} g(\theta_i)$$

- As $N \to \infty$, $\bar{g} \to E(g(\theta) \mid \mathbf{x})$.

Example - Monte Carlo methods

- In a sample of $n = 860$ individuals, $x = 441$ said that they agreed with a Supreme Court ruling that prohibited state or local goverments from requiring the reading of religious texts in public schoool.
- Let θ be the population proportion agreeing with the ruling.

Example - Monte Carlo methods

- In this example, the likelihood follows a Binomial model.
- Recall if we put a Beta(α, β) prior on θ, then the posterior distribution is Beta($x + \alpha, n - x + \beta$).
- Here we choose $\alpha = 1$ and $\beta = 1$.
- Thus, the posterior distribution of θ is Beta($442, 420$).

Example - Monte Carlo methods

- Suppose we are interested in performing inference on the log odds:
$$\gamma = \log\left(\frac{\theta}{1-\theta}\right).$$
- Let us use Monte Carlo methods to describe the posterior and perform inference.

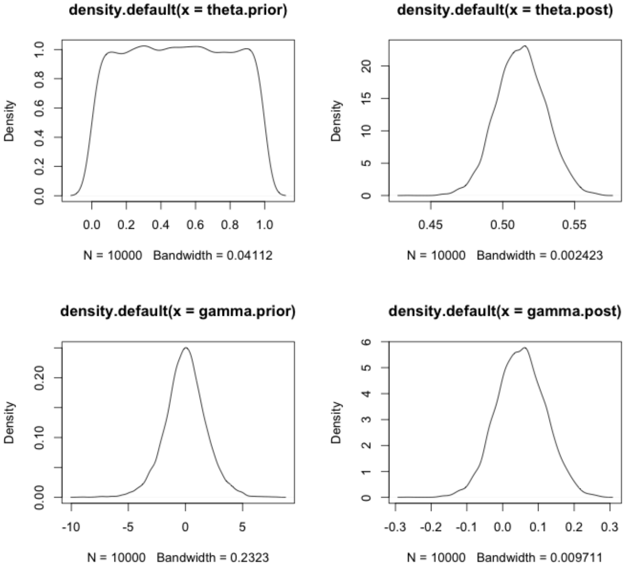
Example - Monte Carlo methods

```
> a=1
> b=1
> nsamp=10000
> theta.prior = rbeta(nsamp,a,b)
> gamma.prior = log(theta.prior/(1-theta.prior))

> theta.post = rbeta(nsamp, a+441, b+860-441)
> gamma.post = log(theta.post/(1-theta.post))

> par(mfrow=c(2,2))
> plot(density(theta.prior))
> plot(density(theta.post))
> plot(density(gamma.prior))
> plot(density(gamma.post))
```

Example - Monte Carlo methods



Posterior inference

- Sampling from the posterior is effective when it can be implemented.
- However, it is often difficult in practice.
- For most probability distributions there is no simple way to simulate random variables of that particular distribution.

MCMC

- Markov-chain Monte-Carlo (MCMC) is a method for sampling from a posterior distribution.
- A Markov chain is generated that has the desired distribution as its stationary distribution.
- The state of the chain after a large number of steps is used as a sample from the desired distribution.
- Can be extremely computationally expensive.

Variational Bayes

- Variational Bayes (VB) is an approach towards approximating the posterior density which is less computationally intensive than MCMC.
- It allows one to approximate the posterior density with another density that has a more analytically tractable form.