More on Bayesian Methods

Rebecca C. Steorts
Bayesian Methods and Modern Statistics: STA 360/601

Lecture 4

Today's menu

- Review of notation
- When are Bayesian and frequentist methods the same?
- Example: Normal-Normal
- Posterior predictive inference
- Example
- Credible Intervals
- Example

Notation

$p(x\|\theta)$	likelihood
$\pi(\theta)$	prior
$p(x) = \int p(x\|\theta)\pi(\theta)\ d\theta$	marginal likelihood
$p(\theta\|x) = \frac{p(x\|\theta)\pi(\theta)}{p(x)}$	posterior probability
$p(x_{new}\|x) = \int p(x_{new}\|\theta)\pi(\theta\|x)\ d\theta$	predictive probability

Another conjugate example

Suppose

$$X_1 \dots X_n \mid \lambda \overset{iid}{\sim} \mathsf{Poisson}(\lambda)$$

$$\lambda \sim \mathsf{Gamma}(\alpha, \beta).$$

Find $p(\lambda \mid X)$.

$$p(\lambda \mid X) = \prod_{i=1}^{n} \left[\lambda^{x_i} e^{-\lambda}/x_i!\right] \times \frac{\beta^\alpha}{\gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \times \lambda} \quad (1)$$

$$\propto \lambda^{n\bar{x}} e^{-n\lambda} \times \lambda^{\alpha-1} e^{-\beta \times \lambda} \quad (2)$$

$$\propto \lambda^{n\bar{x}+\alpha-1} e^{-\lambda(n+\beta)} \quad (3)$$

$$\lambda \mid X \sim \mathsf{Gamma}(n\bar{x} + \alpha, n + \beta)$$

Normal-Normal

$$\begin{aligned} X_1, \ldots, X_n | \theta &\overset{iid}{\sim} \mathsf{N}(\theta, \sigma^2) \\ \theta &\sim \mathsf{N}(\mu, \tau^2), \end{aligned}$$

where σ^2 is known. Calculate the distribution of $\theta | x_1, \ldots, x_n$. Using a ton of math and algebra, you can show that

$$\begin{aligned} \theta | x_1, \ldots, x_n &\sim N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \right) \\ &= N\left(\frac{n\bar{x}\tau^2 + \mu\sigma^2}{n\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2} \right). \end{aligned}$$

Two Useful Things to Know

Definition

The reciprocal of the variance is referred to as the *precision*. Then

$$\text{Precision} = \frac{1}{\text{Variance}}.$$

Suppose the loss we assume is squared error. Let $\delta(x)$ be an estimator of true parameter θ. Then

$$\begin{aligned} MSE(\delta(x)) &= Bias^2 + Variance && (4) \\ &= \left\{\theta - E_\theta[\delta(x)]\right\}^2 + E_\theta\left[\left\{\delta(x) - E_\theta[\delta(x)]\right\}^2\right] && (5) \end{aligned}$$

Theorem

Let δ_n be a sequence of estimators of $g(\theta)$ with mean squared error $E(\delta_n - g(\theta))^2$. Let $b_n(\theta)$ be the bias.

(i) *If $E[\delta_n - g(\theta)]^2 \to 0$ then δ_n is consistent for $g(\theta)$.*

(ii) *Equivalent to the above, δ_n is consistent if $b_n(\theta) \to 0$ and $Var(\delta_n) \to 0$ for all θ.*

(iii) *In particular (and most useful), δ_n is consistent if it is unbiased for each n and if $Var(\delta_n) \to 0$ for all θ.*

We omit the proof since it requires Chebychev's Inequality along with a bit of probability theory. See Problem 1.8.1 in TPE for the exercise of proving this.

Normal-Normal Revisited

We write the posterior mean and posterior variance out.

$$
\begin{aligned}
E(\theta|x) &= \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}. \\
&= \frac{\frac{n\bar{x}}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} + \frac{\frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}.
\end{aligned}
$$

$$
V(\theta|x) = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}.
$$

Can someone given an explanation of what's happening here? How does this contrast frequentist inference?

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent.

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent.
First consider the posterior mean as $n \to \infty$.

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent.
First consider the posterior mean as $n \to \infty$.

$$E(\theta|x) = \frac{\frac{1}{n}\frac{n\bar{x}}{\sigma^2} + \frac{1}{n}\frac{\mu}{\tau^2}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \to \frac{\frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma^2}} = \bar{x} \quad \text{as } n \to \infty.$$

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent.
First consider the posterior mean as $n \to \infty$.

$$E(\theta|x) = \frac{\frac{1}{n}\frac{n\bar{x}}{\sigma^2} + \frac{1}{n}\frac{\mu}{\tau^2}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \to \frac{\frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma^2}} = \bar{x} \quad \text{as } n \to \infty.$$

Now consider

$$E[\bar{x}] = E[E[\bar{x} \mid \theta]] = \theta. \text{ (unbiased)}$$

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent. First consider the posterior mean as $n \to \infty$.

$$E(\theta|x) = \frac{\frac{1}{n}\frac{n\bar{x}}{\sigma^2} + \frac{1}{n}\frac{\mu}{\tau^2}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \to \frac{\frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma^2}} = \bar{x} \quad \text{as } n \to \infty.$$

Now consider

$$E[\bar{x}] = E[E[\bar{x} \mid \theta]] = \theta. \;\; \text{(unbiased)}$$

In the case of the posterior variance, divide the denominator and numerator by n. Then

$$V(\theta|x) = \frac{\frac{1}{n}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \approx \frac{\sigma^2}{n} \to 0 \quad \text{as } n \to \infty.$$

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent. First consider the posterior mean as $n \to \infty$.

$$E(\theta|x) = \frac{\frac{1}{n}\frac{n\bar{x}}{\sigma^2} + \frac{1}{n}\frac{\mu}{\tau^2}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \to \frac{\frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma^2}} = \bar{x} \quad \text{as } n \to \infty.$$

Now consider

$$E[\bar{x}] = E[E[\bar{x} \mid \theta]] = \theta. \ \text{(unbiased)}$$

In the case of the posterior variance, divide the denominator and numerator by n. Then

$$V(\theta|x) = \frac{\frac{1}{n}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \approx \frac{\sigma^2}{n} \to 0 \quad \text{as } n \to \infty.$$

A consistent estimator

Let $\hat{\delta}(x) = E[\theta \mid \boldsymbol{X}]$. Show that the posterior mean is consistent. First consider the posterior mean as $n \to \infty$.

$$E(\theta|x) = \frac{\frac{1}{n}\frac{n\bar{x}}{\sigma^2} + \frac{1}{n}\frac{\mu}{\tau^2}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \to \frac{\frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma^2}} = \bar{x} \quad \text{as } n \to \infty.$$

Now consider

$$E[\bar{x}] = E[E[\bar{x} \mid \theta]] = \theta. \text{ (unbiased)}$$

In the case of the posterior variance, divide the denominator and numerator by n. Then

$$V(\theta|x) = \frac{\frac{1}{n}}{\frac{1}{n}\frac{n}{\sigma^2} + \frac{1}{n}\frac{1}{\tau^2}} \approx \frac{\sigma^2}{n} \to 0 \quad \text{as } n \to \infty.$$

Thus, the posterior mean is consistent by our Theorem, part (iii).

Posterior Predictive Distributions

- We have just seen how estimation can be done in Bayesian analysis.
- Another goal might be prediction.
- That is given some data y and a new observation $\tilde{y}$, we may wish to find the conditional distribution of $\tilde{y}$ given y.
- This distribution is referred to as the *posterior predictive distribution*.
- That is, our goal is to find $p(\tilde{y}|y)$.

Posterior Predictive Distributions

Consider

$$
\begin{aligned}
p(\tilde{y}|y) &= \frac{p(\tilde{y}, y)}{p(y)} && (6)\\
&= \frac{\int_\theta p(\tilde{y}, y, \theta)\ d\theta}{p(y)} && (7)\\
&= \frac{\int_\theta p(\tilde{y}|y, \theta)p(y, \theta)\ d\theta}{p(y)} && (8)\\
&= \int_\theta p(\tilde{y}|y, \theta)p(\theta|y)\ d\theta. && (9)
\end{aligned}
$$

In most contexts, if θ is given, then $\tilde{y}|\theta$ is independent of y, i.e., the value of θ determines the distribution of $\tilde{y}$, without needing to also know y. When this is the case, we say that $\tilde{y}$ and y are *conditionally independent* given θ. Then the above becomes

$$p(\tilde{y}|y) = \int_\theta p(\tilde{y}|\theta)p(\theta|y)\ d\theta.$$

Theorem

If θ is discrete and $\tilde{y}$ and y are conditionally independent given θ, then the posterior predictive distribution is

$$p(\tilde{y}|y) = \sum_{\theta} p(\tilde{y}|\theta)p(\theta|y).$$

If θ is continuous and $\tilde{y}$ and y are conditionally independent given θ, then the posterior predictive distribution is

$$p(\tilde{y}|y) = \int_{\theta} p(\tilde{y}|\theta)p(\theta|y)\; d\theta.$$

Negative Binomial Distribution

- We reintroduce the Negative Binomial distribution.
- The binomial distribution counts the numbers of successes in a fixed number of iid Bernoulli trials.
- Recall, a Bernoulli trial has a fixed success probability p.
- Suppose instead that we count the number of Bernoulli trials required to get a fixed number of successes. This formulation leads to the *Negative Binomial distribution*.
- In a sequence of independent Bernoulli(p) trials, let X denote the trial at which the rth success occurs, where r is a fixed integer.
  Then

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x = r, r+1, \ldots$$

and we say $X \sim$ Negative Binom(r, p).

Negative Binomial Distribution

- There is another useful formulation of the Negative Binomial distribution.
- In many cases, it is defined as $Y =$ number of failures before the rth success. This formulation is statistically equivalent to the one given above in term of $X =$ trial at which the rth success occurs, since $Y = X - r$. Then

$$f(y) = \binom{r+y-1}{y} p^r (1-p)^y, \ y = 0, 1, 2, \ldots$$

  and we say $Y \sim$ Negative Binom(r, p).
- When we refer to the Negative Binomial distribution in this class, we will refer to the second one defined unless we indicate otherwise.

$$\begin{aligned} X|\lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{Gamma}(a, b) \end{aligned}$$

Assume that $\tilde{X}|\lambda \sim \text{Poisson}(\lambda)$ is independent of X. Assume we have a new observation $\tilde{x}$. Find the posterior predictive distribution, $p(\tilde{x}|x)$. Assume that a is an integer. First, we must find $p(\lambda|x)$.

Recall

$$\begin{aligned} p(\lambda|x) &\propto p(x|\lambda)(p(\lambda) \\ &\propto e^{-\lambda}\lambda^x\lambda^{a-1}e^{-\lambda/b} \\ &= \lambda^{x+a-1}e^{-\lambda(1+1/b)}. \end{aligned}$$

Thus, $\lambda|x \sim$ Gamma$(x+a, \frac{1}{1+1/b})$, i.e., $\lambda|x \sim$ Gamma$(x+a, \frac{b}{b+1})$. Finish the problem for homework.

- Suppose that X is the number of pregnant women arriving at a particular hospital to deliver their babies during a given month.
- The discrete count nature of the data plus its natural interpretation as an arrival rate suggest modeling it with a Poisson likelihood.
- To use a Bayesian analysis, we require a prior distribution for θ having support on the positive real line. A convenient choice is given by the Gamma distribution, since it's conjugate for the Poisson likelihood.

The model is given by

$$\begin{aligned} X|\lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{Gamma}(a, b). \end{aligned}$$

- We are also told 42 moms are observed arriving at the particular hospital during December 2007. Using prior study information given, we are told $a = 5$ and $b = 6$.
- (We found a, b by working backwards from a prior mean of 30 and prior variance of 180).

We would like to find several things in this example:

1. Plot the likelihood, prior, and posterior distributions as functions of λ in `R`.
2. Plot the posterior predictive distribution where the number of pregnant women arriving falls between [0,100], integer valued.
3. Find the posterior predictive probability that the number of pregnant women arrive is between 40 and 45 (inclusive). Do this for homework.
4. You are expected to have this done by early this week or next week since you have an exam on Thursday, Feb 11 (in class). (This material will not be turned in but could appear on the exam).