Lecture 3

The binomial model

After suspicious performance in the weekly soccer match, 37 mathematical sciences students, staff, and faculty were tested for performance enhancing drugs. Let $Y_i = 1$ if athlete i tests positive and $Y_i = 0$ otherwise. A total of 13 mathletes tested positive.

What is the sampling model $p(y_1, ..., y_{37}|\theta)$?

Assume a uniform prior distribution on $p(\theta)$. Write the pdf for this distribution.

In what larger class of distributions does this distribution reside? What are the parameters?

Note that $E[\theta] = \frac{\alpha}{\alpha+\beta}$ and $Var[\theta] = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ if

Now compute the posterior distribution, $p(\theta|\boldsymbol{y})$.

$$
\begin{aligned}
p(\theta|\boldsymbol{y}) &= \frac{\mathcal{L}(\theta|\boldsymbol{y})p(\theta)}{\int_\theta \mathcal{L}(\theta|\boldsymbol{y})p(\theta)d\theta} \\
&= \frac{\mathcal{L}(\theta|\boldsymbol{y})p(\theta)}{p(\boldsymbol{y})} \\
&\propto \mathcal{L}(\theta|\boldsymbol{y})p(\theta) \\
&\propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^{13+1-1}(1-\theta)^{37-13+1-1}
\end{aligned}
$$

Conjugate Priors

We have shown that a beta prior distribution and a binomial sampling model lead to a beta posterior distribution. This class of beta priors is **conjugate** for the binomial sampling model.

Def: Conjugate *A class $\mathcal{P}$ of prior distributions for θ is called conjugate for a sampling model $p(y|\theta)$ if $p(\theta) \in \mathcal{P} \rightarrow p(\theta|y) \in \mathcal{P}$.*

Conjugate priors make posterior calculations simple,but

Intuition about Prior Parameters

Note the posterior expectation can be written as:

$$E[\theta|y] = \frac{\alpha + y}{\alpha + \beta + n}$$

Now what do we make of:

- α:

- β:

- $\alpha + \beta$:

If $n >> \alpha + \beta$

If $n << \alpha + \beta$

Predictive Distributions

An important element in Bayesian statistics is the (posterior) predictive distribution, in this case let Y^* be the outcome of a future experiment. We are interested in computing:

$$Pr(Y^* = 1|y_1, ..., y_n) = \int Pr(Y^* = 1|\theta, y_1, ..., y_n)p(\theta|y_1, ..., y_n)d\theta$$

Note that the predictive distribution does not depend on any unknown quantities, but rather only the observed data. Furthermore, Y^* is not independent of $Y_1, ..., Y_n$ but depends on them through θ.

Posterior Intervals

With a Bayesian framework we can compute **credible intervals**.

Credible Interval:

Recall in a frequentist setting

$$Pr(l(y) < \theta < u(y)|\theta) = \tag{1}$$

Note that in some settings Bayesian intervals can also have frequentist coverage probabilities, at least asymptotically.

Quantile based intervals With quantile based intervals, the posterior quantiles are used with $\theta_{\alpha/2}, \theta_{1-\alpha/2}$ such that:

Quantile based intervals are typically easy to compute.

Highest posterior density (HPD) region: A 100 $\times$ (1-α)% HPD region consists of a subset of the parameter space, $s(y) \subset \Theta$ such that

The Poisson Model

Now assume the National Park Services records daily totals of tourists caught breaking the rules. This data can be modeled with a Poisson model.

Recall, $Y \sim Poisson(\theta)$ if

Properties of the Poisson distribution:

- E[Y]=

- Var(Y) =

- $\sum_i^n Y_i \sim Poisson(\theta_1 + ... + \theta_n)$ if

Conjugate Priors for Poisson

Recall conjugate priors for a sampling model have a posterior model from the same class as the prior. Let $y_i \sim Poisson(\theta)$, then

$$p(\theta|y_1,...,y_n) \propto p(\theta)\mathcal{L}(\theta|y_1,...,y_n) \quad (2)$$

$$\propto p(\theta) \times \theta^{\sum y_i} \exp(-n\theta) \quad (3)$$

Thus the conjugate prior class will have the form

A positive quantity θ has a

Properties of

- $E[\theta] =$

- $Var(\theta) =$

Posterior Distribution

Let $Y_1, ..., Y_n \sim Poisson(\theta)$ and $p(\theta) \sim gamma(a, b)$, then

$$p(\theta|y_1, ..., y_n) = \frac{p(\theta)p(y_1, ..., y_n|\theta)}{p(y_1, ..., y_n)} \tag{4}$$

Note that

$$E[\theta|y_1, ..., y_n] = \frac{a + \sum y_i}{b + n}$$

So now a bit of intuition about the prior distribution. The posterior expectation of θ is a combination of the prior expectation and the sample average:

- b

- a

*W

Predictive distribution

The predictive distribution, $p(y^*|y_1, ..., y_n)$, can be computed as:

$$\begin{aligned}
p(y^*|y_1,...,y_n) &= \int p(y^*|\theta, y_1,...,y_n)p(\theta|y_1,...,y_n)d\theta \\
&= \int p(y^*|\theta)p(\theta|y_1,...,y_n)d\theta \\
&= \int \left\{ \frac{\theta^{y^*}\exp(-\theta)}{y^*!} \right\} \left\{ \frac{(b+n)^{a+\sum y_i}}{\Gamma(a+\sum y_i)} \theta^{a+\sum y_i - 1} \exp(-(b+n)\theta) \right\} \\
&= ... \\
&= ...
\end{aligned}$$

You can (and likely will) show that $p(y^*|y_1,...,y_n) \sim NegBinom(a + \sum y_i, b+n)$.

Exponential Families The binomial and Poisson models are examples of one-parameter exponential families. A distribution follows a one-parameter exponential family if it can be factorized as:

$$p(y|\theta) = h(y)c(\phi)\exp(\phi t(y)), \tag{5}$$

where ϕ is the unknown parameter and $t(y)$ is the sufficient statistic. The using the class of priors, where $p(\phi) \propto c(\phi)^{n_0} \exp(n_0 t_0 \phi)$, is a conjugate prior. There are similar considerations to the Poisson case where n_0 can be thought of as a "prior sample size" and t_0 is a "prior guess."

Prior Distribution Choice

A noninformative prior, $p(\theta)$,

Example 1. Suppose θ is the probability of success in a binomial distribution, then the uniform distribution on the interval $[0, 1]$ is a noninformative prior.

Example 2. Suppose θ is the mean parameter of a normal distribution. What is a noninformative prior distribution for the mean?

Invariant Priors

Recall ideas of variable transformation (from Casella and Berger): Let X have pdf $p_x(x)$ and let Y = g(X), where g is a monotone function. Suppose $p_X(x)$ is continuous and that $g^{-1}(y)$ has a continuous derivative. Then the pdf of Y is given by

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

Example. Let $p_x(x) = 1$, for $x \in [0, 1]$ and let $Y = g(x) = -\log(x)$, then

Now if $p_x(x)$ had been a prior on X, the transformation to y and $p_y(y)$

Jeffreys Priors

The idea of invariant priors was addressed by Jeffreys. Let $p_J(\theta)$ be a Jeffreys prior if:

$$p_J(\theta) = [I(\theta)]^{1/2},$$

where $I(\theta)$ is the expected Fisher information given by

$$I(\theta) = -E\left[\frac{\partial^2 \log p(X|\theta)}{\partial \theta^2}\right]$$

Example. Consider the Normal distribution and place a prior on μ when σ^2 is known. Then the Fisher information is

A similar derivation for the joint prior $p(\mu, \sigma) = \frac{1}{\sigma}$

Advantages and Disadvantages of Objective Priors

Advantages

Disadvantages

Advantages and Disadvantages of Subjective Priors

Advantages

Disadvantages

In many cases weakly-informative prior distributions are used that have some of the benefits of subjective priors without imparting strong information into the analysis.