The binomial model

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

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

$$\propto$$

$$\propto$$

$$\propto$$

The posterior expectation, $E[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n}$, is a function of prior information and the data.

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

Conjugate priors make posterior calculations simple, but might not always be the best representation of prior beliefs.

Intuition about Prior Parameters

Note the posterior expectation can be written as:

$$\begin{aligned} E[\theta|y] &= \frac{\alpha + y}{\alpha + \beta + n} \\ &= \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta}\right) + \frac{n}{\alpha+\beta+n}\left(\frac{y}{n}\right) \end{aligned}$$

Now what do we make of:

- α

- β

- $\alpha + \beta$

Predictive Distributions

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

$$Pr(Y^* = 1|y_1, ..., y_n) \quad =$$

Note that the predictive distribution does not depend on any unknown quantities, but rather only the observed data.

Now let's compute the posterior predictive distribution in this specific example with $n = 37$ and $\sum y_i = 13$.

Gelman and Shalizi talked about posterior predictive checks, how do we simulate from this predictive distribution?

Posterior Intervals

With a Bayesian framework we can compute **credible intervals**.
Credible Interval:

Recall in a frequentist setting

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:

1. $Pr(\theta < \theta_{\alpha/2}|Y = y) =$ and
2. $Pr(\theta > \theta_{1-\alpha/2}|Y = y) =$

Quantile based intervals are typically easy to compute.

Highest posterior density (HPD) region

A 100 $\times$ (1-α)% HPD region consists of

All points in the HPD region have higher posterior density than those not in region. Additionally the HPD *region* need not be a continuous interval. HPD regions are typically more computationally intensive to compute than quantile based intervals.

Sketch a posterior distribution where the HPD and quantile based intervals would be different.

The Poisson Model

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

Properties of the Poisson distribution:

- $E[Y] =$
- $Var(Y) =$
- Iif $Y_i \sim Poisson(\theta_i)$, then $\sum_i^n Y_i \sim$

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

$$\begin{aligned} p(\theta|y_1, ..., y_n) &\propto p(\theta)\mathcal{L}(\theta|y_1, ..., y_n) & (1) \\ &\propto p(\theta) \times \theta^{\sum y_i} \exp(-n\theta) & (2) \end{aligned}$$

What is the conjugate prior for a Poisson sampling model? Why?

Thus the conjugate prior class will have the form $\theta^{c_1} \exp(c_2\theta)$. This is the kernel of a

A positive quantity θ has a

Properties :

- $E[\theta] =$
- $Var(\theta) =$

Posterior Distribution

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

$$\begin{aligned} p(\theta|y_1, ..., y_n) &= \frac{p(\theta)p(y_1, ..., y_n|\theta)}{p(y_1, ..., y_n)} \\ &= \{\theta^{a-1} \exp(-b\theta)\} \times \{\theta^{\sum y_i} \exp(-n\theta)\} \times \{c(y_1, ..., y_n, a, b)\} \\ &\propto \end{aligned}$$

So $\theta|y_1,...,y_n \sim Gamma($
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 in interpreted as

- a is interpreted as

When $n >> b$

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:

$$p(y|\theta) = h(y)c(\phi)\exp(\phi t(y)), \tag{3}$$

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."

More on Priors

Noninformative Priors

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

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

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

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

$$\begin{aligned} I(\theta) &= -E\left[\frac{\partial^2}{\partial \mu^2}\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)\right] \\ &= \frac{1}{\sigma^2} \end{aligned}$$

Hence in this case the Jeffreys prior for μ is a constant. 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.