Week 3 (part 1): Sept 12 - Sept 14

The binomial model

Example. 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 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?

Beta distribution. Recall, $\theta \sim Beta(\alpha, \beta)$ if:

$E[\theta] =$

$Var[\theta] =$

Now compute the posterior distribution, $p(\theta|\boldsymbol{y})$, where $\boldsymbol{y} = \sum_i^n y_i$.

$$p(\theta|\boldsymbol{y}) \quad =$$

Compute the posterior expectation, $E[\theta|\boldsymbol{y}] =$
Note that this 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.

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:

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

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

$$Pr(l(y) < \theta < u(y)|\theta) =$$

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.

2.

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

1.

2.

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.

Exercise. Sketch a distribution where the HPD and "equal" quantile based intervals do not agree.

The Poisson Model

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

Properties of the Poisson distribution:

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

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

$$\begin{aligned} p(\theta|y_1, ..., y_n) &\propto p(\theta)\mathcal{L}(\theta|y_1, ..., y_n) \\ &\propto \end{aligned}$$

Thus the conjugate prior class will have the form

A positive quantity θ has a distribution if:

$$p(\theta) =$$

Properties of

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

Posterior Distribution

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

$$\begin{aligned} p(\theta|y_1, ..., y_n) &= \frac{p(\theta)p(y_1, ..., y_n|\theta)}{p(y_1, ..., y_n)} \\ &= \\ &\propto \end{aligned}$$

So $\theta|y_1, ..., y_n \sim$

Note that

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

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 $n >> b$ the information from the data dominates the prior distribution.

Predictive distribution

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

$$\begin{aligned} p(y^*|y_1, ..., y_n) &= \int p(y^*|\theta, y_1, ..., y_n)p(\theta|y_1, ..., y_n)d\theta \\ &= \int \\ &= \int \\ &= ... \\ &= ... \end{aligned}$$

You can (and likely will) show that $p(y^*|y_1, ..., y_n) \sim$

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:

$$\tag{1}$$

where ϕ is the unknown parameter and $t(y)$ is the sufficient statistic.

The using the class of priors, where $p(\phi) \propto$, is a conjugate prior. There are similar considerations to the Poisson case where can be thought of as a "prior sample size" and is a "prior guess."