Bayesian Conjugates

***This document is a lightning summary of Donovan and Mickey Chapters 10, 11, and 12**, which offer an introduction to three types of Bayesian conjugates. Perhaps we can see the forest for the trees more clearly working through these three chapters in one fell swoop.*

As you may have started to realize, parametrizing priors, likelihoods, and posteriors, and computing products of functions and their integrals is the bane of getting quickly to conclusions with Bayesian statistics.

Furthermore, once you have a posterior, you may get some more data, and the posterior after the first round of data collection becomes the prior going into the second round of data collection, and so you have yet more products of functions and integrals to compute. This seems never ending!

Wouldn't life be great if the posterior and the prior had the same functional form? If that were so, the posterior after a round of data collection which becomes the prior for the next round — if it is of the same form — wouldn't introduce a slew of new products and integrals to compute. Well, life is great. There are pairings of priors and likelihoods for which this is true, and these pairings are called "Bayesian conjugates."

Chapter 10 — Beta Function Priors are Conjugate to Binomial Likelihoods

Our canonical example will be field goal attempts.

Beta Function Priors

In my introduction to Chapter 10, I wrote down the beta function prior in the following form:

$$P(p) = \frac{1}{B(\alpha,\beta)}\, p^{\alpha-1}(1-p)^{\beta-1}$$

The parameter p is the probability of a successful field goal attempt (and it might be something small like 0.1 for field goals being attempted from 80 yards). $1/B(\alpha,\ \beta)$ is a normalization factor that you would have something like Mathematica compute for you. We think of it as a constant because in this context we are thinking of p as the variable, and p doesn't show up in $B(\alpha,\ \beta)$. It is there so that the area under the curve of $P(p)$ is 1. In Donovan and Mickey's jargon, α and β are called "hyperparameters." This is so you won't get confused between them and the parameter p which shows up in the binomial distribution. All three of these are parameters, and I will not often be saying "hyperparameters" because parameters appear in many contexts and we don't need to have a different name for each context.

Two facts about the beta function prior: (1) It has mean $\mu = \frac{\alpha}{\alpha+\beta}$, and (2) it has variance $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$. These facts can be important when you are picking a prior.

Binomial Likelihoods

Binomial likelihoods are old hat for you now, but it is handy to reproduce the formula. The likelihood of having n successes (field goals made) in N trials (field goals attempted) is $P(n \mid p) = \binom{N}{n} p^{N-n}(1-p)^n$.

Chapter 11 — Gamma Function Priors are Conjugate to Poisson Likelihoods

Our canonical example will be mug breakage rates.

Gamma Function Priors

In Chapter 11, Donovan and Mickey wrote down the beta function prior in the following form:

$$P(a) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, a^{\alpha-1}\, e^{-\beta a}$$

The parameter a is the mug breakage rate. $\beta^\alpha / \Gamma(\alpha)$ is also a normalization factor that you would have something like Mathematica compute for you. We think of this factor as a constant because it doesn't depend on a. The hyperparameters in the gamma function have absolutely nothing to do with the hyperparameters in the beta function.

Two useful facts about the gamma function prior: (1) It has mean $\mu = \frac{\alpha}{\beta}$, and (2) it has variance $\sigma^2 = \frac{\alpha}{\beta^2}$.

Poisson Likelihoods

Poisson likelihoods are old hat for you now, but it is handy to reproduce the formula. The likelihood of having n events (mug breakages) is $P(n \mid a) = \frac{a^n}{n!} e^{-a}$.

Chapter 12 — Gaussian Function Priors are Conjugate to Gaussian Likelihoods

Our canonical example will be bacteria survival time.

Gaussian Function Priors

In Chapter 12, Donovan and Mickey wrote down the gaussian function prior, but I am going to write it down with something more like Young's variable names:

$$P(m) = \frac{1}{\sqrt{2\pi}\ \sigma_m} e^{-(m-\mu)^2/2\,\sigma_m{}^2}$$

Something I don't want to deal with and that we really should deal with if we want to be completely honest is that σ_m is probably not going to be given. However, I am going to treat it as a given for the foreseeable future.

Of course, this Gaussian function prior (1) has mean μ, and (2) variance $\sigma_m{}^2$.

Gaussian Likelihoods

Gaussian likelihoods are old hat. The likelihood of getting the data point x (a bacterium survival time) is

$$P(x \mid m) = \frac{1}{\sqrt{2\pi}\ \sigma_x} e^{-(x-m)^2/2\,\sigma_x{}^2}.$$

Notice that I have allowed for the possibility that the variance, $\sigma_x{}^2$, in bacterium survival times has some different value than the variance, $\sigma_m{}^2$, in the prior. There is no reason for these variances to be the same.

The Posteriors

Having summarized all these formula in one handy-dandy place, we now have the fun of computing the posteriors:

$$P(p \mid n) = \frac{P(n \mid p)\, P(p)}{\int_0^1 P(n \mid p)\, P(p)\, dp}$$

$$P(a \mid n) = \frac{P(n \mid a)\, P(a)}{\int_0^\infty P(n \mid a)\, P(a)\, da}$$

$$P(m \mid x) = \frac{P(x \mid m)\, P(m)}{\int_{-\infty}^\infty P(x \mid m)\, P(m)\, dm}$$

We’ll do that algebra together in class. All we really need to focus on to understand why Bayesian conjugates are so great is the numerators. What happens in the integrals in the denominators is much less interesting (why!?), and we will usually just stuff those into Mathematica.