Beta Distributions
Class 14, 18.05
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be familiar with the 2-parameter family of beta distributions and its normalization.

2. Be able to update a beta prior to a beta posterior in the case of a binomial likelihood.

2 Beta distribution

The beta distribution beta(a, b) is a two-parameter distribution with range $[0, 1]$ and pdf

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$$

We have made an applet so you can explore the shape of the Beta distribution as you vary the parameters:

`http://mathlets.org/mathlets/beta-distribution/`.

As you can see in the applet, the beta distribution may be defined for any real numbers $a > 0$ and $b > 0$. In 18.05 we will stick to integers a and b, but you can get the full story here: `http://en.wikipedia.org/wiki/Beta_distribution`

In the context of Bayesian updating, a and b are often called hyperparameters to distinguish them from the unknown parameter θ representing our hypotheses. In a sense, a and b are 'one level up' from θ since they parameterize its pdf.

2.1 A simple but important observation!

If a pdf $f(\theta)$ has the form $c\theta^{a-1}(1-\theta)^{b-1}$ then $f(\theta)$ is a beta(a, b) distribution and the normalizing constant must be

$$c = \frac{(a+b-1)!}{(a-1)!\,(b-1)!}.$$

This follows because the constant c must normalize the pdf to have total probability 1. There is only one such constant and it is given in the formula for the beta distribution.

A similar observation holds for normal distributions, exponential distributions, and so on.

2.2 Beta priors and posteriors for binomial random variables

Example 1. Suppose we have a bent coin with unknown probability θ of heads. We toss it 12 times and get 8 heads and 4 tails. Starting with a flat prior, show that the posterior pdf is a beta$(9, 5)$ distribution.

answer: This is nearly identical to examples from the previous class. We'll call the data from all 12 tosses x_1. In the following table we call the leading constant factor in the posterior column c_2. Our simple observation will tell us that it has to be the constant factor from the beta pdf.

The data is 8 heads and 4 tails. Since this comes from a binomial(12, θ) distribution, the likelihood $p(x_1|\theta) = \binom{12}{8}\theta^8(1-\theta)^4$. Thus the Bayesian update table is

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$1 \cdot d\theta$	$\binom{12}{8}\,\theta^8(1-\theta)^4$	$\binom{12}{8}\,\theta^8(1-\theta)^4\,d\theta$	$c_2\,\theta^8(1-\theta)^4\,d\theta$
total	1		$T = \binom{12}{8}\int_0^1 \theta^8(1-\theta)^4\,d\theta$	1

Our simple observation above holds with $a = 9$ and $b = 5$. Therefore the posterior pdf

$$f(\theta|x_1) = c_2\theta^8(1-\theta)^4$$

follows a beta$(9, 5)$ distribution and the normalizing constant c_2 must be

$$c_2 = \frac{13!}{8!\,4!}.$$

Note: We explicitly included the binomial coefficient $\binom{12}{8}$ in the likelihood. We could just as easily have given it a name, say c_1 and not bothered making its value explicit.

Example 2. Now suppose we toss the same coin again, getting n heads and m tails. Using the posterior pdf of the previous example as our new prior pdf, show that the new posterior pdf is that of a beta$(9+n, 5+m)$ distribution.

answer: It's all in the table. We'll call the data of these $n+m$ additional tosses x_2. This time we won't make the binomial coefficient explicit. Instead we'll just call it c_3. Whenever we need a new label we will simply use c with a new subscript.

hyp.	prior	likelihood	Bayes posterior	numerator
θ	$c_2\theta^8(1-\theta)^4\,d\theta$	$c_3\,\theta^n(1-\theta)^m$	$c_2c_3\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$	$c_4\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$
total	1		$T = \int_0^1 c_2c_3\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$	1

Again our simple observation holds and therefore the posterior pdf

$$f(\theta|x_1, x_2) = c_4\theta^{n+8}(1-\theta)^{m+4}$$

follows a beta$(n+9, m+5)$ distribution.

Note: Flat beta. The beta$(1, 1)$ distribution is the same as the uniform distribution on $[0, 1]$, which we have also called the flat prior on θ. This follows by plugging $a = 1$ and $b = 1$ into the definition of the beta distribution, giving $f(\theta) = 1$.

Summary: If the probability of heads is θ, the number of heads in $n+m$ tosses follows a binomial$(n+m,\theta)$ distribution. We have seen that if the prior on θ is a beta distribution then so is the posterior; only the parameters a, b of the beta distribution change! We summarize precisely how they change in a table. We assume the data is n heads in $n+m$ tosses.

hypothesis	data	prior	likelihood	posterior
θ	$x = n$	beta(a,b)	binomial$(n+m,\theta)$	beta$(a+n, b+m)$
θ	$x = n$	$c_1\theta^{a-1}(1-\theta)^{b-1}\,d\theta$	$c_2\theta^n(1-\theta)^m$	$c_3\theta^{a+n-1}(1-\theta)^{b+m-1}\,d\theta$

2.3 Conjugate priors

In the literature you'll see that the beta distribution is called a conjugate prior for the binomial distribution. This means that if the likelihood function is binomial, then a beta prior gives a beta posterior. In fact, the beta distribution is a conjugate prior for the Bernoulli and geometric distributions as well.

We will soon see another important example: the normal distribution is its own conjugate prior. In particular, if the likelihood function is normal with known variance, then a normal prior gives a normal posterior.

Conjugate priors are useful because they reduce Bayesian updating to modifying the parameters of the prior distribution (so-called hyperparameters) rather than computing integrals. We saw this for the beta distribution in the last table. For many more examples see:
`http://en.wikipedia.org/wiki/Conjugate_prior_distribution`