5.2 A Description of Beliefs: The Beta Distribution

Zongyi Liu

2023-06-10

5.2 A Description of Beliefs: The Beta Distribution

Beta distribution can be expressed with two parameters, a and b, and be expressed as

$$p(\theta|a,b) = beta(\theta;a,b) = \theta^{(a-1)}(1-\theta)^{(b-1)}/B(a,b)$$

where $B(a,b)$ is simply a normalizin constant that ensures that the area under the beta density integrates to 1. In other words, the normalizer for the beta distribution is

$$B(a,b) = \int_0^1 d\theta\theta^{(a-1)}(1-\theta)^{(b-1)}$$

5.2.1 Specifying a Beta Prior

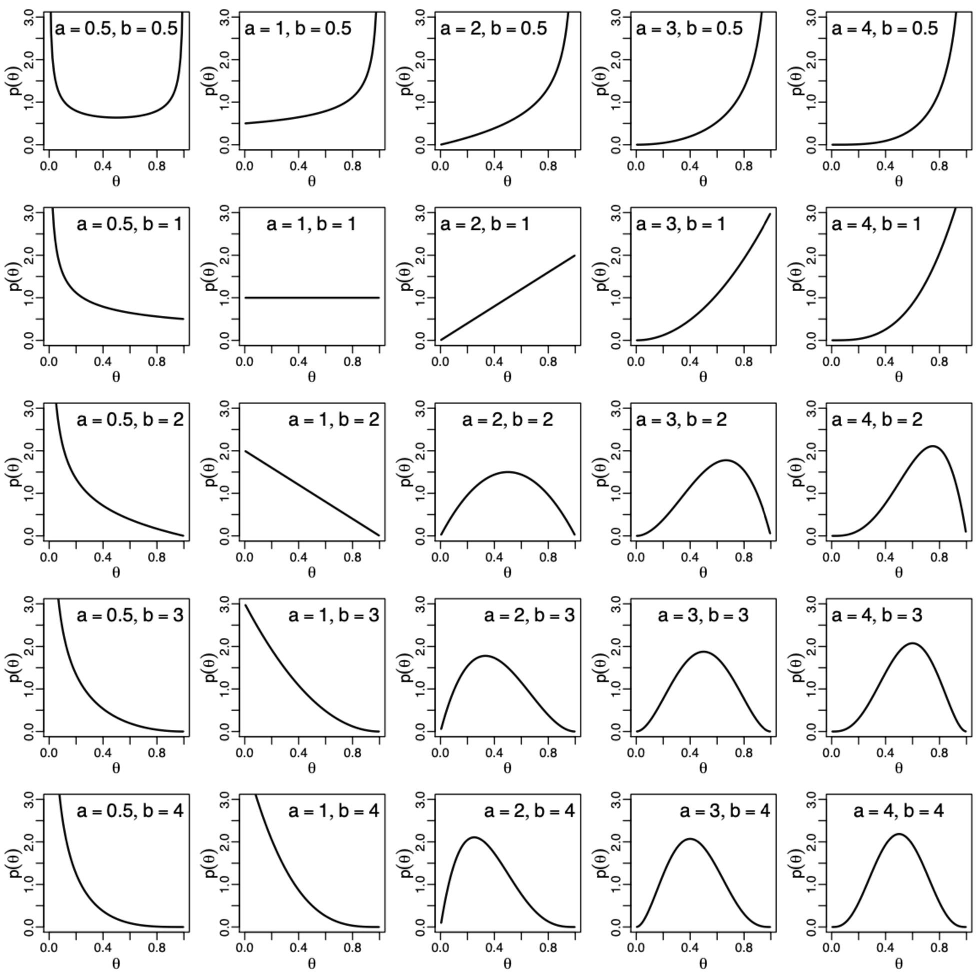
We would like to specify a beta distribution that describes our prior beliefs, thus we need to knoe the mean and variance of the beta distribution.

We can get than for beta$(\theta;a,b)$

- mean is $\bar{\theta} = a/(a+b)$
- standard deviation is $\sqrt{\bar{\theta}(1-\bar{\theta})/(a+b+1)}$

You can think of a and b in the prior as if they were previously observed data, in which there were a heads and b tails in a total of a + b flips.

We can get the beta distribution with different parameters plotted as below:



Instead of thinking in terms of a or b in the prior data, it's easier to think in terms of the mean proportion of heads in the prior data and its sample size, the mean proportion of heads, $m = a/(a+b)$, and the sample size, $n = a + b$. Solving these gives

$$a = mn, and b = (1 - m)n$$

Another way of establishing the shape parameters is by starting with the mean and standard deviation of the desired beta distribution, we should notice that the standard deviation must make sense in this context.

For a beta density with mean m and standard deviation s, the slope parameters are

$$a = m(\frac{m(1-m)}{s^2} - 1) and b = (1-m)(\frac{m(1-m)}{s^2} - 1)$$

In most applications, we would deal with beta distribution for which $a \geq 1$ and $b \geq 1$.

5.2.2 The Posterior Beta

Here we can combine the Bernoulli and Beta distribution together:

$$
\begin{aligned}
p(\theta|z,N) &= p(z,N|\theta)p(\theta)/p(z,N) \\
&= \theta^{z}\ (1-\theta)^{(N-z)}\ \ \theta^{(a-1)}\,(1-\theta)^{(b-1)} \,/\, [B(a,b)p(z,N)] \\
&= \theta^{((z+a)-1)}\ (1-\theta)^{((N-z+b)-1)} \Bigg/ \underbrace{[B(a,b)\,p(z,N)]}_{B(z+a,N-z+b)} \ .
\end{aligned}
$$

In this sequence of equations, we would follow the collection of powers of θ and of $(1-\theta)$, but we may balked here.

The transition was made by simply thinking about what the normalizing factor for the numerator must be. The numerator is $\theta^{((z-a)-1)}(1-\theta)^{((N-z+b)-1)}$, which is the numerator of a beta$(\theta; z+a, N-z+b)$ distribution.

We can also think about the prior and posterior means; the prior mean of θ is $a/(a+b)$, and the posterior mean is $(z+a)/[(z+a)+(N-z+b)] = (z+a)/(N+a+b)$, we can have

It indicates that the posterior mean is always somewhere between the prior mean and the proportion in the data.

$$
\underbrace{\frac{z+a}{N+a+b}}_{\text{posterior}} = \underbrace{\frac{z}{N}}_{\text{data}}\underbrace{\frac{N}{N+a+b}}_{\text{weight}} + \underbrace{\frac{a}{a+b}}_{\text{prior}}\underbrace{\frac{a+b}{N+a+b}}_{\text{weight}} .
$$

The mixing weight on the prior mean has N in its denominator, and so it decreases as N increases. The mixing weight on the data proportion increases as N increases. So the more data we have, the less is the influence of the prior, and the posterior mean gets closer to the proportion in the data.

The choice of prior n (which equals $a+b-2$) should represent the size of the new data set that would sway us away from our prior toward the data proportion.