Conjugate prior of a binomial distribution

Jacob L. Fine

July 9th, 2024

Suppose we have some random variable X that follows a binomial distribition, $p(x|\theta) \sim \text{Binomial}(n, \theta)$ and we wish to derive a closed-form expression for the prior density of the mean parameter θ. It is often useful to choose a prior density function $\pi(\theta)$ such that it is a member of the same family of distributions as its posterior density $\pi(\theta|x)$. If this is the case, we say that $\pi(\theta)$ is the conjugate prior of the likelihood function. It is useful to choose a conjugate prior as the prior distribution, since it is easier to interpret, i.e., as simply containing updated parameters of the prior from the same distribution. It has a nice closed-form expression that can be easier to work with as well.

We will show that a Beta distribution is the conjugate prior of the Binomial likelihood function, which is why many often choose it to model the mean parameter's distribution. The expression for the distribution of $\theta \sim \text{Beta}(\alpha, \beta)$ is

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

And if $X \sim \text{Binomial}(n, \theta)$, we have that

$$f(x|\theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

By Bayes's rule, the expression for the posterior density θ conditioned on X with respect to the likelihood function $L(\theta)$ and the prior $\pi(\theta)$ is

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)}$$

We often write $f(x|\theta)$ as $L(\theta)$ to imply that it is a likelihood function. We can also express $f(x)$ using the likelihood function as the integral

$$f(x) = \int_{\theta} L(\theta)\pi(\theta)d\theta$$

Since the integral is just some positive constant, we often ignore it in the context of deriving posteriors/priors and just write the proportionality relation

$$\pi(\theta|x) \propto L(\theta)\pi(\theta)$$

We can ignore the denominator since we just want to show that the kernel of the prior and posterior are of the same family of distributions. We therefore substitute $L(\theta) = f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ and $\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ to obain that

$$\implies \pi(\theta|x) \propto \left[\binom{n}{x}\theta^x(1-\theta)^{n-x}\right]\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]$$

Simplifying, we obtain

$$\implies \pi(\theta|x) \propto \binom{n}{x}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

In the above, both $\binom{n}{x}$ and $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ are positive values, which implies that

$$\implies \pi(\theta|x) \propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$$

Since the RHS is evidently the kernel of a Beta distribution, we have shown that the conjugate prior of a binomial likelihood function is in fact a Beta distribution.