In Danish: 2005-09-19 / SLB
Translated: 2008-05-28 / SLB

Bayesian Statistics, Simulation and Software

The beta-binomial distribution

I have translated this document, written for another course in Danish, almost as is. I have kept the references to Lee, the textbook used for that course.

Introduction

In Lee: *Bayesian Statistics*, the beta-binomial distribution is very shortly mentioned as the predictive distribution for the binomial distribution, given the conjugate prior distribution, the beta distribution. (In Lee, see pp. 78, 214, 156.) Here we shall treat it slightly more in depth, partly because it emerges in the WinBUGS example in Lee § 9.7, and partly because it possibly can be useful for your project work.

Bayesian Derivation

We make n independent Bernoulli trials (0-1 trials) with probability parameter π. It is well known that the number of successes x has the binomial distribution. Considering the probability parameter π unknown (but of course the sample-size parameter n is known), we have

$$x|\pi \sim \text{bin}(n, \pi),$$

where in *generic*[1] notation

$$p(x|\pi) = \binom{n}{x} \pi^x (1-\pi)^{1-x}, \quad x = 0, 1, \ldots, n.$$

We assume as prior distribution for π a beta distribution, i.e.

$$\pi \sim \text{beta}(\alpha, \beta),$$

[1] By this we mean that p is not a fixed function, but denotes the density function (in the discrete case also called the probability function) for the random variable the value of which is argument for p.

with density function

$$p(\pi) = \frac{1}{\mathrm{B}(\alpha, \beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1}, \quad 0 < \pi < 1.$$

I remind you that the beta *function* can be expressed by the gamma function:

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{1}$$

In Lee, § 3.1 is shown that the posterior distribution is a beta distribution as well,

$$\pi | x \sim \text{beta}(\alpha + x, \beta + n - x).$$

(Because of this result we say that the beta distribution is *conjugate* distribution to the binomial distribution.) We shall now derive the *predictive* distribution, that is finding $p(x)$. At first we find the simultaneous distribution

$$p(x, \pi) = p(\pi) p(x|\pi) = \frac{\binom{n}{x}}{\mathrm{B}(\alpha, \beta)} \pi^{\alpha+x-1}(1-\pi)^{\beta+n-x-1}.$$

Then we integrate π away, and we get the predictive distribution

$$p(x) = \binom{n}{x} \frac{\mathrm{B}(\alpha + x, \beta + n - x)}{\mathrm{B}(\alpha, \beta)}, \quad x = 0, 1, \ldots, n.$$

The emerged distribution is called the *beta-binomial distribution*, and we write

$$x \sim \text{betabin}(n, \alpha, \beta).$$

We use (1) and the property of the gamma function that

$$\Gamma(\alpha + x) = \Gamma(\alpha)\alpha(\alpha+1)\ldots(\alpha+x-1).$$

Hereby we get

$$p(x) = \binom{n}{x} \frac{\alpha(\alpha+1)\ldots(\alpha+x-1)\beta(\beta+1)\ldots(\beta+n-x-1)}{(\alpha+\beta)(\alpha+\beta+1)\ldots(\alpha+\beta+n-1)}. \tag{2}$$

Expectation and Variance

We apply two useful formulae which are derived in Lee, p. 26

$$\mathrm{E}(x) = \mathrm{E}\{\mathrm{E}(x|\pi)\}, \quad \mathrm{Var}(x) = \mathrm{E}\{\mathrm{Var}(x|\pi)\} + \mathrm{Var}\{\mathrm{E}(x|\pi)\}. \tag{3}$$

(There are other ways to derive expectation and variance, but if you understand the formulae (3) then you have the simplest derivation!) Expectation and variance for the binomial distribution are assumed to be so well known that I don't need to mention them here, and for the beta distribution we have (see Lee, p. 293)

$$\mathrm{E}(\pi) = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}(\pi) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

By means of (3) we get

$$\mathrm{E}(x) = E(n\pi) = \frac{n\alpha}{\alpha+\beta},$$

and

$$\begin{aligned}
\mathrm{Var}(x) &= \mathrm{E}\{n\pi(1-\pi)\} + \mathrm{Var}(n\pi) \\
&= n\mathrm{E}(\pi) - n\mathrm{E}(\pi^2) + n^2\mathrm{Var}(\pi) \\
&= n\{E(\pi) - \mathrm{Var}(\pi) - \mathrm{E}(\pi)^2 + n\mathrm{Var}(\pi)\} \\
&= \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

Introducing

$$\pi = \frac{\alpha}{\alpha+\beta},$$

we get

$$\mathrm{E}(x) = n\pi, \quad \mathrm{Var}(x) = n\pi(1-\pi)\frac{\alpha+\beta+n}{\alpha+\beta+1}$$

We see that the variance is higher than for the corresponding binomial distribution $\mathrm{bin}(n,\pi)$ and say that there is *overdispersion*.

An Urn Model

Finally I'll mention that when α and β are integers, the formula (2) gives rise to interpreting the distribution by an *urn model*,[2] i.e. a model where balls are randomly drawn from an urn. We start with an urn in which there are α red and β white balls. Now we randomly[3] draw n balls from the urn and count the number of red balls x in the sample. If the draws are with replacement, as you will know we get the binomial distribution $\mathrm{bin}(n,\pi)$, where $\pi = \alpha/(\alpha+\beta)$. If the draws are without replacement we get the *hypergeometric* distribution, not to be elaborated further here.[4]

[2] Here *urn* just means a big jar.

[3] By this we mean that all the balls in the urn have the same probability to be drawn. In practice the urn must be shaken well after each draw.

[4] I can't help mentioning that the hypergeometric distribution has *underdispersion*, by which we mean that the variance is lower than for the corresponding binomial distribution.

Now we assume that after each draw we not only put the ball back into the urn but further put in another ball of the same colour. For instance, if we first draw a red ball, then a white one, and then a red one, the probability is

$$\begin{aligned} P(R_1 W_2 R_3) &= P(R_1)P(W_2|R_1)P(R_3|R_1, W_2) \\ &= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta+1}\right)\left(\frac{\alpha+1}{\alpha+\beta+2}\right) \end{aligned}$$

By reordering the numerators we see that the probability is the same for two red and one white ball, irrespectively of the order in which they are drawn. This can be generalised to an arbitrary number of red and white balls, which is why we just need to count how many orders there are. If n balls are drawn and x are red, there are $\binom{n}{x}$ possible orders. We can therefore reason in the same way as by the derivation of the binomial distribution, and we get the formula (2).

It was the Hungarian mathematician Pólya who derived the beta-binomial distribution in this way, for which reason it is also called the *Pólya distribution*. Pólya used this urn model to describe the spread of contagious diseases. If you observe a sick person, this suggests that the disease is increasing, such that the number of “sick” balls in the urn is increased. If you observe a healthy person, this suggests that the disease is decreasing, such that the number of “healthy” balls in the urn is increased. He also considered the model where more than one extra ball of the drawn colour are added.