Quiz #2

Dustin Leatherman

October 5, 2020

1 1

Let n be the unknown customers that visit a store on the day of a sale. The number of customers that make a purchase is distributed by $Y|n \sim Bin(n, 0.2)$ where 0.2 is the known probability of making a purchase given the customer visited the store. The prior $n \sim Poisson(5)$. Give an expression for the posterior of n. What is the support (i.e., set of values with positive probability) of the posterior distribution?

Prior: $n \sim Poisson(5)$
Likelihood: $Y|n \sim Bin(n, 0.2)$
Posterior:

$$\begin{aligned} f(n|Y) =& \frac{f(Y|n)f(n)}{f(Y) = \int f(Y|n)f(n)dn} \\ =& \frac{\binom{n}{Y}0.2^Y 0.8^{n-Y} \cdot \frac{5^Y e^{-5}}{Y!}}{\int \binom{n}{Y}0.2^Y 0.8^{n-Y} \cdot \frac{5^Y e^{-5}}{Y!} dn} \\ =& \frac{\binom{n}{Y}0.8^{n-Y}}{\int \binom{n}{Y}0.8^{n-Y} dn} \end{aligned} \tag{1}$$

This looks close to a binomial but doesn't quite match since $0.2Y$ cancels out. The support of the posterior is $(0, \infty)$ since there cannot be a negative number of customers.

2 2

It is known that 25% of untreated patches of forest will be infected by bark beetles. A company has developed a new intervention and applied it to 50 (independent) patches of forest. They will record the number of patches that become infected. Describe a Bayesian Analysis plan to analyze these data and test whether the pesticide is more effective than no treatment (provide all details including likelihood, prior, and how you will summarize the posterior).

The response variable is the number of infected patches of forest. This can be interpreted as a number of "successes" which is modeled best by a Binomial distribution.

$$Y|\theta \sim Bin(n = 50, \theta)$$

It is known that 25% of untreated patches of forest will be infected by bark beetles. Since the analysis is trying to model the percentage infected and we have prior knowledge of the infection rate which falls between 0 and 1, the Beta-Binomial conjugate Prior can be used.

Let θ be Beta Random Variable which represents the infection rate. Then, $E(\theta) = \frac{a}{a+b}$. a and b can be chosen so that the expected value matches the infection rate of 0.25. For $\theta \sim Beta(a = 1, b = 3)$, the expected value is $\frac{1}{1+3} = \frac{1}{4}$.

Thus a reasonable choice for a Prior is

$$\theta \sim Beta(1, 3)$$

This means that the posterior is

$$\begin{aligned} \theta|Y \sim & Beta(Y_i + a, n - Y_i + b) \\ \sim & Beta(Y_i + 1, 53 - Y_i) \end{aligned} \tag{2}$$

Testing Pesticides

Let $g(\theta_1, \theta_2, Y_1, Y_2)$ be an indicator function which returns 1 when $P(\theta_1|Y_1) > P(\theta_2|Y_2)$ and 0 otherwise. The average value of the outputs of this indicator function yield the probability that $P(\theta_1|Y_1) > P(\theta_2|Y_2)$.

$$\theta^* = \frac{1}{\frac{m+n}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m} g(\theta_i, \theta_j, Y_i, Y_j)$$

If θ^* exceeds a predefined threshold (perhaps 0.5 to start), then there can be some confidence that the pesticide is successful. The converse result where θ^* falls below a threshold would indicate that the pesticide is less effective than the control.