Predictive Probability Interim Analysis

John Cook

February 26, 2006
Revised: March 19, 2014

1 Introduction

It is natural to ask in the middle of a trial how likely it is that the trial will reach one conclusion or another, or even to reach no conclusion at all. Predictive probabilities provide a mechanism for answering that question.

Predictive probabilities are most easily understood in the context of binary outcomes. This note focuses on binary outcomes, though the same principles apply more broadly to other outcomes, such as time-to-event.

2 Preliminaries for beta random variables

Let

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

denote the beta function. A beta(a,b) random variable is one with probability density function

$$\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}.$$

Let Y be a binomial random variable with probability of success θ where θ has a beta(a,b) prior. The prior predictive probability of s successes and f failures in Y is

$$\int_0^1 \binom{s+f}{s}\theta^s(1-\theta)^f\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\,d\theta$$

which simplifies to

$$\binom{s+f}{s}\frac{B(s+a,f+b)}{B(a,b)}.$$

Note that the predictive probability of *one* success is $E[\theta]$. However, the predictive probability of two successes is *not* simply $E[\theta]^2$.

Suppose now that we have observed s_0 successes on Y and f_0 failures and want to find the *posterior* predictive probability of s more successes and f more failures on Y. This is the same as the *prior* predictive probability of s successes and f failures with the prior on Y updated to beta(s_0+a, f_0+b), which amounts to

$$\binom{s+f}{s} \frac{B(s+s_0+a,\, f+f_0+b)}{B(s_0+a,\, f_0+b)}.$$

2.1 Examples

Suppose you just started watching a high school basketball game. You've noticed that one of the players, Jones, has made two out of his last five free throws. You want to estimate the probability of Jones making his next free throw. Not having anything else to go on, you use your recent observations to form a beta(2, 3) prior on the probability of Jones making a free throw. You calculate the predictive probability of Jones making his next free throw at 2/5, exactly his average so far, and start to wonder about the benefit is of all this predictive probability machinery.

But next you decide to look a little further into the future and wonder about the outcomes of Jones's next two free throws. Now things get a little more interesting. Again using a beta(2, 3) prior, you find that the predictive probability of Jones making his next two free throws is

$$\binom{2+0}{2} \frac{B(2+2,\, 3+0)}{B(2,\, 3)} = \frac{1}{5}.$$

The predictive probability for one out of two free throws is

$$\binom{1+1}{1} \frac{B(2+1,\, 3+1)}{B(2,\, 3)} = \frac{2}{5}$$

and the predictive probability for his missing two free throws is

$$\binom{0+2}{2} \frac{B(2+0,\, 3+2)}{B(2,\, 3)} = \frac{2}{5}.$$

Note that these are not quite what you would have gotten if you had simply assumed that Jones makes 2 out of 5 free throws. For in that case, you'd estimate his probabilities of missing 0, 1, or 2 of his next free throws as $(\frac{2}{5})^2$, $2(\frac{2}{5})(\frac{3}{5})$, and $(\frac{3}{5})^2$ respectively. These frequentist estimates work out to 0.16, 0.48, and 0.36, whereas the corresponding predictive probabilities are 0.2, 0.4, and 0.4. What accounts for the difference? *The Bayesian approach takes into account the uncertainty in the estimate of Jones's free throw ability.* After all, you'd only observed five free throw attempts when you started to wonder about probabilities. If you had observed that Jones made 20 out of his last 50 attempts and chosen a beta(20, 30) prior, the predictive probabilities would have been (0.164706, 0.470588, 0.364706), closer to the empirical estimates.

3 Application to interim analysis

Let θ_A and θ_B be the probabilities of response on arms A and B. Given the data collected so far, we want to calculate the probability that at the end of the trial we will conclude $\theta_A > \theta_B$ or $\theta_B > \theta_A$. (In general these probabilities will not add up to 1 because it is possible we will not have enough evidence to reject the null hypothesis $\theta_A = \theta_B$.)

There are two commonly used ways determine whether $\theta_A > \theta_B$ or vice versa, one frequentist and one Bayesian. The idea behind both calculations is simple: we look at all possible outcomes, and add up the probabilities of the outcomes that can lead to each conclusion.

3.1 Frequentist approach

The term "frequentist" may seem a bit odd, because we're doing Bayesian statistics here. What we're calling the frequentist approach is calculating the *Bayesian* predictive probability of a *frequentist* test result.

Let X_A be the number of successes observed to date on arm A, Y_A a hypothetical number of future successes, and N_A the total number of patients who will be treated on arm A. Denote the analogous quantities for arm B similarly. Define

$$\hat{\theta}_A = \frac{X_A + Y_A}{N_A},$$

$$\hat{\theta}_B = \frac{X_B + Y_B}{N_B},$$

and

$$\hat{\theta} = \frac{X_A + Y_A + X_B + Y_B}{N_A + N_B}.$$

The usual frequentist two-sample statistic for the equality of θ_A and θ_B is

$$Z = \frac{\hat{\theta}_A - \hat{\theta}_B}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{N_A} + \frac{1}{N_B}\right)}}.$$

This test concludes $\theta_A > \theta_B$ if $Z > z_{1-\alpha^*/2}$ and concludes $\theta_B > \theta_A$ if $Z < z_{\alpha^*/2}$ where α^* is the significance level of the test.

The predictive probability of concluding, for example, $\theta_A > \theta_B$ is simply the sum of the predictive probabilities of all (Y_A, Y_B) outcomes for which the test would make this conclusion.

3.2 Bayesian approach

The usual Bayesian approach would be to conclude treatment A is superior if the posterior probability

$$P(\theta_A > \theta_B \,|\, \text{data})$$

is larger than some cutoff c. The predictive probability of concluding treatment A is superior is the sum of the predictive probabilities of all (Y_A, Y_B) outcomes for which

$$P(\theta_A > \theta_B \,|\, \text{data}) > c.$$

3.3 Example

Suppose 50 patients have been treated so far, 25 on arm A and 25 on arm B, in a trial that was design to run to 100 patients. There have been 10 successes on arm A and 16 on arm B. How likely is it that if this trial were run to completion, the trial would reject the null hypothesis that the response rates on the two arms, θ_A and θ_B, are equal at a 5% significance level? How likely is the trial to decide in favor of A? In favor of B? Use the frequentist method with a beta(0.6, 0.4) prior for the probability of response on each arm.

In the future, we will treat 25 more patients on each arm. We examine all possible pairs (s_A, s_B) of success counts on arms A and B with $s_A \geq 0$ and $s_B \geq 0$. We determine which of these pairs lead to the conclusion that $\theta_A > \theta_B$ at the 2.5% significance level, and add up their predictive probabilities. We find that the set of such pairs is the set (s_A, s_B) such that $s_A - s_B \geq 16$. (In general, the description of the points for which the test passes will not be this simple, but here is it especially tidy.) This set of outcomes has total predictive probability of about 3×10^{-6}. The data collected so far imply that it is very unlikely the trial will select A as the superior treatment. If A were an investigational treatment and B were a placebo, the trial should stop at this point.

Next we find the probability of the trial concluding arm B is superior. Again we determine which (s_A, s_B) pairs lead to this conclusion. The description of this set of points is a little messier than the corresponding set above. If there are no more responses on A and at least 3 more on B, the trial will conclude in favor of A. Or if there is only one more response on A and at least 5 more on B. We continue, for each number of successes on A, determining the minimum number of successes on B. If there are 23 or more future successes on A, there cannot be enough responses on B to make the trial conclude B is superior. When we add up the predictive probabilities of each of these pairs, we get 0.6886.

Note that the probability of declaring A the winner is essentially zero, and the probability of declaring B the winner is 0.6886. These numbers do not sum to 1, because the predictive probability of not rejecting the null is 0.3114.

To illustrate the predictive probability calculations, we show how one of the terms in the sum leading to declaring B superior wold be calculated. We example the possibility of 11 future successes on A and 18 future successes on B. In this case we find $z = -2.613$, less than the cutoff of -1.96 for the 2.5%

significance level. The predictive probability of 11 successes on A is

$$\binom{11+14}{11}\frac{B(10+11+0.6,\ 15+14+0.4)}{B(10+0.6,\ 15+0.4)}$$

and the predictive probability of 18 successes on B is

$$\binom{18+7}{18}\frac{B(16+18+0.6,\ 9+7+0.4)}{B(16+0.6,\ 9+0.4)}$$

and so their product, 0.01154, is the predictive probability of the pair.

4 Resources

J. Kyle Wathen has written software to calculate the predictive probabilities discussed in this note. His software also supports time-to-event outcomes. This software is available on the MD Anderson Biostatistics software download site:

`https://biostatistics.mdanderson.org/SoftwareDownload`

For more information on posterior predictive probability, see *Bayesian Data Analysis* by Gelman, Carlin, Stern, and Rubin.