University of Texas, MD Anderson Cancer Center
UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series

Year 2012 *Paper* 76

Fast approximation of Beta inequalities

John D. Cook*

*The University of Texas MD Anderson Cancer Center, jdcook@mdanderson.org

Fast approximation of Beta inequalities

John D. Cook

Abstract

Many Bayesian clinical trial methods are based on random inequalities. For some distribution families, these inequalities can be computed in closed form. For example, [1] gives closed-form solutions to computing

P(X $>$ Y)

when X and Y are either independent normal or independent gamma random variables. However the case of beta random variables is very important and no closed form solution for this case is known. Such inequalities must be evaluated numerically. Simulation programs using these inequalities spend nearly all their time computing the inequalities. This report presents a close-form approximation for beta inequalities that is two orders of magnitude faster to evaluate.

Fast approximation of Beta inequalities

John D. Cook
Department of Biostatistics
P. O. Box 301402, Unit 1409
The University of Texas, M. D. Anderson Cancer Center
Houston, Texas 77230-1402, USA
cook@mdanderson.org

October 13, 2012

Abstract

Many Bayesian clinical trial methods are based on random inequalities. For some distribution families, these inequalities can be computed in closed form. For example, [1] gives closed-form solutions to computing

$$P(X > Y)$$

when X and Y are either independent normal or independent gamma random variables. However the case of beta random variables is very important and no closed form solution for this case is known. Such inequalities must be evaluated numerically. Simulation programs using these inequalities spend nearly all their time computing the inequalities. This report presents a close-form approximation for beta inequalities that is two orders of magnitude faster to evaluate.

1 Normal approximation

It is well known that a beta distribution may sometimes be approximated by a normal distribution. This approximation becomes exact asymptotically as the beta distribution parameters increase. For small parameters, the approximation is best when the parameters are nearly equal.

The key idea in this report is

$$P(X_B > Y_B) \approx P(X_N > Y_N)$$

where X_B and Y_B are independent beta random variables and X_N and Y_N are their normal approximations formed by moment matching. This approximation can be surprisingly accurate, even for small parameter values.

Let μ_X be the common mean of X_B and X_N. If the beta parameters of X_B are a and b, then

$$\mu_X = \frac{a}{a+b}.$$

Also, let σ_X^2 be the common variance of X_B and X_N. In terms of the beta parameters, this is

$$\sigma_X^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

Then

$$P(X_N > Y_N) = \Phi\left(\frac{\mu_X - \mu_Y}{(\sigma_X^2 + \sigma_Y^2)^{1/2}}\right).$$

If the distribution parameters of X_B and Y_B take on integer values between 1 and 10 inclusive, the maximum absolute error occurs when $X_B \sim$ Beta(1, 3) and $Y_B \sim$ Beta(3, 10). In this case $P(X_B > Y_B) = 0.4835$ and $P(X_N > Y_N) = 0.5342$ for an absolute error of 0.05069. The average absolute error over the parameter values is 0.006676.

If we search instead over integer parameters between 10 and 100 inclusive, the maximum absolute error occurs when $X_B \sim$ Beta(10, 31) and $Y_B \sim$ Beta(32, 100). In this case $P(X_B > Y_B) = 0.4927$ and $P(X_N > Y_N) = 0.5078$ for an absolute error of 0.0151. The average absolute error over the parameter values is 0.0006416.

The time required to compute the beta inequality directly through numerical integration and the time required to compute the inequality via the normal approximation depend on how particular software is implemented. However, in one benchmark, the normal approximation was 510 times faster than the numerical integration.

2 Shifted inequalities

The approximation given here can be extended to

$$P(X_B > Y_B + \delta) \approx P(X_N > Y_N + \delta)$$

.

In this case one simply replaces μ_Y with $\mu_Y + \delta$ in the calculations above.

3 Numerical tactics

Much of the work that goes into designing Bayesian clinical trials is trial-and-error simulation to find parameters that give the Bayesian design the desired frequentist operating characteristics. (Perhaps one day Bayesian trials will be designed on Bayesian principles, but for now this is rarely done.) One could use the approximation presented here during the search for desired parameters and then use slower but more exact computational methods for confirmation.

Another approach would be to always use the approximation given here, but switch to a more accurate method only when the former method suggests that the difference in accuracy may matter. For example, if a trial is designed to stop when $P(X > Y)$ is 0.95 and the approximate value of the inequality is in the range (0.94, 0.96), software could recalculate the inequality using a more accurate method.

4 References

John D. Cook. Numerical Computation of Stochastic Inequality Probabilities. MDACC technical report UTMDABTR-008-03, (2003), revised 2005.

Available at http://biostats.bepress.com/mdandersonbiostat/paper46/