Numerical Computation of Stochastic Inequality Probabilities

Numerical Computation of Stochastic Inequality Probabilities

Abstract

This paper addresses the problem of numerically evaluating

$$P(X > Y) \quad (1)$$

for independent continuous random variables X and Y. This calculation arises in the design of clinical trials and as such appears in the inner loop of simulations of these trials. An early example of this is given by Thompson (1933), with more recent examples by Giles et al (2003), and Berry (2003, 2004). It is worthwhile to optimize the calculation of these probabilities as they may be computed millions of times in the course of simulating a single trial. Techniques such as memoization (Orwant 2002) can eliminate redundant calculations of such probabilites throughout a simulation, but the need for a large number of evaluations remains.

After considering how to compute (1) in general, we present optimizations for important special cases in which X and Y both belong to one of the following families of classical distributions: exponential, gamma, inverse gamma, normal, Cauchy, beta, and Weibull.

Numerical Computation of Stochastic Inequality Probabilities

John D. Cook
Division of Quantitative Sciences, Unit 1409
The University of Texas, M. D. Anderson Cancer Center
P. O. Box 301402, Houston, Texas 77030, USA
cook@mdanderson.org

First published December 10, 2003. This revision August 8, 2008.

Abstract

This paper addresses the problem of numerically evaluating

$$P(X > Y) \tag{1}$$

for independent continuous random variables X and Y. This calculation arises in the design of clinical trials and as such appears in the inner loop of simulations of these trials. An early example of this is given by Thompson (1933), with more recent examples by Giles et al (2003), and Berry (2003, 2004). It is worthwhile to optimize the calculation of these probabilities as they may be computed millions of times in the course of simulating a single trial. Techniques such as memoization (Orwant 2002) can eliminate redundant calculations of such probabilites throughout a simulation, but the need for a large number of evaluations remains.

After considering how to compute (1) in general, we present optimizations for important special cases in which X and Y both belong to one of the following families of classical distributions: exponential, gamma, inverse gamma, normal, Cauchy, beta, and Weibull.

Keywords: adaptive clinical trials, beta distribution, gamma distribution

1 General Distributions

For a continuous random variable X, let f_X and F_X be the PDF and CDF respectively of X and let $G_X = 1 - F_X$.

We will consider all density functions to be defined over $(-\infty, \infty)$ though they may be identically zero over parts of this domain.

The probability that $X > Y$ is given by

$$P(X > Y) = \int_{-\infty}^{\infty} f_X(t) F_Y(t)\, dt.$$

Let $\eta = \sqrt{\varepsilon/2}$. Define

$$\begin{aligned} a^* &= \min(F_X^{-1}(\eta), F_Y^{-1}(\eta)) \\ b^* &= \max(G_X^{-1}(\eta), G_Y^{-1}(\eta)) \end{aligned}$$

Then

$$0 < \int_{-\infty}^{a^*} f_X(t) F_Y(t)\, dt < \int_{-\infty}^{a^*} f_X(t) F_Y(a^*)\, dt = F_X(a^*) F_Y(a^*) < \varepsilon/2$$

since F_X is an increasing function and we freeze it at its largest value over the region of integration.

Also,

$$\begin{aligned} \int_{b^*}^{\infty} f_X(t) F_Y(t)\, dt &= \int_{b^*}^{\infty} f_X(t) F_Y(b^*)\, dt + \int_{b^*}^{\infty} f_X(t)(F_Y(t) - F_Y(b^*))\, dt \\ &= G_X(b^*) F_Y(b^*) + \int_{b^*}^{\infty} f_X(t)(F_Y(t) - F_Y(b^*))\, dt. \end{aligned}$$

Now

$$\begin{aligned} \int_{b^*}^{\infty} f_X(t)(F_Y(t) - F_Y(b^*))\, dt &< \int_{b^*}^{\infty} f_X(t)(1 - F_Y(b^*))\, dt \\ &= G_X(b^*) G_Y(b^*). \end{aligned}$$

It follows that

$$\int_{-\infty}^{\infty} f_X(t) F_Y(t)\, dt \approx \int_{a^*}^{b^*} f_X(t) F_Y(t)\, dt + G_X(b^*) F_Y(b^*). \qquad (2)$$

with an error less than ε. Also, the right side of (2) is a lower bound as well as an approximation.

The integral above can be integrated using an adaptive integration routine. (`DQAG` from `QUADPACK` (Piessens et al 1983) is a popular example of such a routine.) However, such routines may need help finding dramatic changes in the integrand. For example, if f_X is a spike centered in the tail of F_Y, the numerical routine may miss the mass under f_X entirely. For this reason we recommend an optimization routine to first find the maximum of the integrand and split the domain of integration into three intervals, the middle interval being centered around the maximum.

2 Special Distributions

2.1 Exponential

If X has an exponential with mean μ_X and Y has an exponential with mean μ_Y then it is well known that

$$P(X > Y) = \frac{\mu_X}{\mu_X + \mu_Y}. \tag{3}$$

See, for example, the work of Casella and Berger (2001).

2.2 Gamma

Let X have a gamma (α_X, β_X) distribution and Y a gamma (α_Y, β_Y) distribution. One can show via the transformation theorem that the random variable B defined by

$$B = \frac{\beta_X Y}{\beta_X Y + \beta_Y X}$$

has a beta (α_Y, α_X) distribution. It follows that

$$P(X > Y) = P\left(B < \frac{\beta_X}{\beta_X + \beta_Y}\right) \tag{4}$$

and so $P(X > Y)$ can be computed by evaluating an incomplete beta function. See the work of Abramowitz and Stegun (1970) and DiDonato and Morris (1992) for details on computing the incomplete beta function.

2.3 Inverse Gamma

If Z has an inverse gamma distribution with parameters (α, β) then $1/Z$ has a gamma distribution with parameters $(\alpha, 1/\beta)$. Thus for random variables

X and Y with inverse gamma distributions, we use

$$P(X > Y) = P(1/Y > 1/X) \tag{5}$$

to take advantage of the method above for comparing gamma random variables.

2.4 Normal

Let X, Y, and Z be normal random variables with parameters (μ_X, σ_X), (μ_Y, σ_Y), and (0,1) respectively.

$$\begin{aligned}
P(X > Y) &= P(0 > Y - X) \\
&= P(0 > \mu_Y - \mu_X + (\sigma_X^2 + \sigma_Y^2)^{1/2} Z) \\
&= P\left(Z < \frac{\mu_X - \mu_Y}{(\sigma_X^2 + \sigma_Y^2)^{1/2}}\right) \\
&= \Phi\left(\frac{\mu_X - \mu_Y}{(\sigma_X^2 + \sigma_Y^2)^{1/2}}\right)
\end{aligned}$$

where Φ is the distribution function of a standard normal. Abramowitz and Stegun (1970) provide methods of evaluating Φ.

2.5 Cauchy

Let X, Y, and C be Cauchy random variables with parameters (μ_X, σ_X), (μ_Y, σ_Y), and (0,1) respectively.

$$\begin{aligned}
P(X > Y) &= P(0 > Y - X) \\
&= P(0 > \mu_Y - \mu_X + (\sigma_X + \sigma_Y) C) \\
&= P\left(C < \frac{\mu_X - \mu_Y}{\sigma_X + \sigma_Y}\right) \\
&= \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{\mu_X - \mu_Y}{\sigma_X + \sigma_Y}\right)
\end{aligned}$$

2.6 Beta

2.6.1 Symmetries and Recurrences

Define

$$g(a, b, c, d) = P(X > Y)$$

where X and Y are distributed as beta(a, b) and beta(c, d) respectively. The function has three fundamental symmetries. First of all, g obviously satisfies

$$g(a, b, c, d) = 1 - g(c, d, a, b). \tag{6}$$

The change of variables $u = 1 - x$ reveals that

$$g(a, b, c, d) = g(d, c, b, a) \tag{7}$$

as well. Thompson (1933) was aware of these two symmetries but was apparently not aware that

$$g(a, b, c, d) = g(d, b, c, a) \tag{8}$$

as well.

To prove this symmetry in the first and last arguments, we note that

$$g(a, b, c, d) = \int_0^1 \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)} I_x(c, d)\, dx \tag{9}$$

where $I_x(c, d)$ is the incomplete beta function, the CDF of a beta(c, d) random variable. Equation 26.5.16 from the work of Abramowitz and Stegun (1970) says that

$$I_x(c, d) = \frac{1}{cB(c, d)} x^c (1-x)^d + I_x(c+1, d). \tag{10}$$

Substituting (10) in (9) shows that

$$g(a, b, c, d) = \frac{B(a+c, b+d)}{cB(a, b)B(c, d)} + g(a, b, c+1, d). \tag{11}$$

Define

$$\begin{aligned} h(a, b, c, d) &= \frac{B(a+c, b+d)}{B(a, b)B(c, d)} && (12) \\ &= \frac{\Gamma(a+c)\Gamma(b+d)\Gamma(a+b)\Gamma(c+d)}{\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(d)\Gamma(a+b+c+d)} && (13) \end{aligned}$$

Since

$$\lim_{c\to\infty} g(a,b,c,d) = 0$$

it follows that

$$g(a,b,c,d) = \sum_{n=0}^{\infty} \frac{h(a,b,c+n,d)}{c+n}.$$

Each term is symmetric in a and d and therefore g is as well.

Combining the basic symmetries reveals further symmetries. For example, (8) and (7) shows that g is also symmetric in its second and third arguments.

The symmetry relations and equation (11) combine to show that

$$\begin{aligned}
g(a+1,b,c,d) &= g(a,b,c,d) + h(a,b,c,d)/a \\
g(a,b+1,c,d) &= g(a,b,c,d) - h(a,b,c,d)/b \\
g(a,b,c+1,d) &= g(a,b,c,d) - h(a,b,c,d)/c \\
g(a,b,c,d+1) &= g(a,b,c,d) + h(a,b,c,d)/d
\end{aligned}$$

In clinical trial simulations, one often starts by evaluating $g(a,b,c,d)$ for small arguments and increments different arguments by 1 as outcomes are recorded. In this setting we can evaluate g from scratch one time and repeatedly apply the recurrence relationships above.

If we need to directly evaluate g for moderate-sized parameters, we can use the recurrence relationships to reduce the problem to evaluating g with parameters in the interval $[1,2)$. (If the parameters are integers, the problem can be reduced to $g(1,1,1,1)$ which clearly equals 1/2.) This approach is inefficient for large values of the parameters because the number of calculations required is proportional to the size of the parameters. However, asymptotic methods are available for large parameters. Therefore we present two methods for evaluating g, one for small parameters and one for large parameters, and rely on the recurrence relationships in the middle.

2.6.2 Small arguments

We begin by expanding the integrand of (9) in series:

$$g(a,b,c,d) = \frac{1}{B(a,b)B(c,d)} \int_0^1 \left(x^{a+c-1} \sum_{j=0}^{\infty} \alpha_j x^j \right) \left(\sum_{k=0}^{\infty} \beta_k x^k \right) dx$$

where

$$\alpha_j = \binom{b-1}{j}(-1)^j$$

and

$$\beta_k = \binom{d-1}{k}\frac{(-1)^k}{c+k}.$$

We combine the series in the integrand to

$$g(a,b,c,d) = \frac{1}{B(a,b)B(c,d)}\int_0^1 x^{a+c-1}\sum_{n=0}^{\infty}\gamma_n x^n\,dx$$

where γ_n is the convolution

$$\gamma_n = \sum_{j+k=n}\alpha_j\beta_k.$$

We could exchange the order of integration and summation above. However, the rate of convergence would be unacceptably slow since the radius of convergence of the series is 1. By splitting the domain of integration into two halves, we obtain more rapid convergence.

Define

$$\begin{aligned} g_\ell(a,b,c,d) &= \int_0^{1/2}\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}I_x(c,d)\,dx \\ g_r(a,b,c,d) &= \int_{1/2}^{1}\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}I_x(c,d)\,dx \end{aligned}$$

so that

$$g(a,b,c,d) = g_\ell(a,b,c,d) + g_r(a,b,c,d).$$

By symmetry we can show that

$$g_r(a,b,c,d) = I_{1/2}(b,a) - g_\ell(b,a,d,c)$$

and so it suffices to be able to compute g_ℓ.

We have

$$g_\ell(a,b,c,d) = \frac{1}{B(a,b)B(c,d)}\sum_{n=0}^{\infty}\frac{\gamma_n}{(a+c+n)2^{a+c+n}}$$

and must now decide where to truncate the infinite sum.

Note that

$$\begin{aligned} |\gamma_n| &= \left| \sum_{i=0}^{n} \binom{b-1}{n-1} \binom{d-1}{i} \frac{1}{c+i} \right| \\ &\leq \left| \sum_{i=0}^{n} \binom{b-1}{n-i} \binom{d-1}{i} \frac{1}{i} \right| \\ &= \left| (d-1) \binom{b+d-3}{n-1} \right| \end{aligned}$$

(The absolute value signs are necessary since b and d may not be integers.)

It follows that for $N > b + d - 2$, the error in the series truncated after N terms is bounded by

$$\frac{|d-1|}{B(a,b)B(c,d)} \sum_{n=N+1}^{\infty} \frac{1}{(a+c+i)2^{a+c+i}}$$

and the sum is bounded by

$$\frac{1}{(a+c+N)2^{a+c+N}}.$$

¿From this it is simple to determine the size of N necessary for the desired accuracy.

Because of the alternating signs in the series, there can be a numerical loss of precision when using this series to evaluate $g(a, b, c, d)$ for large parameters. However, because of the recurrence relations and the asymptotitc approximation below, it is unnecessary to use the series approximation for large parameters.

If we were numerically integrating (9) to evaluate g, arguments less than 1 would be a special concern since they cause the integrand to be singular at one or both ends. However, in the development of this section, arguments less than 1 do not require special treatment.

2.6.3 Large arguments

Wise (1960) showed that if X is a beta random variable then $(-\log X)^{1/3}$ is approximately normal, and the approximation improves as the parameters become larger. The approximation also improves if the parameters are approximately equal. For example, if all parameters are between 4 and 5, the error in computing g is bounded by 0.0002. Therefore, for large arguments, we reduce the problem of comparing beta random variables to that of comparing normal random variables as described above. Wise gives equations for approximating the mean and variance of the transformed variables. However, we find the the approximations given by the delta method are more accurate. If X is distributed as beta(a, b) then the mean and variance of $Y = (-\log X)^{1/3}$ are approximately

$$\left(-\log \frac{a}{a+b}\right)^{1/3}$$

and

$$\frac{1}{9}\left(-\log \frac{a}{a+b}\right)^{-4/3} \frac{ab}{(a+b)^2(a+b+1)}$$

respectively.

Depending on one's accuracy requirements, it may be more efficient to use the recurrence relations to transform the problem of small arguments to one with moderately large arguments.

2.7 Weibull

Let X have a Weibull distribution with shape a and scale b and let Y have a Weibull distribution with shape c and scale d.

$$P(X < Y) = \int_0^\infty \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a - \left(\frac{x}{d}\right)^c\right) dx.$$

The difficulty in evaluating this integral comes if $a < 1$ in which case the integrand is singular at 0.

The change of variables

$$u = \left(\frac{x}{b}\right)^a$$

transforms the above integral to

$$\int_0^\infty \exp\left(-u - (b/d)^c u^{c/a}\right) dx. \tag{14}$$

The latter integral is much better behaved: for all values of the paramters it is bounded at 0 and is monotone decreasing.

By swapping the role of X and Y if necessary, we may assume that $a < c$, which makes the integrand decay to zero faster and shortens the domain of integration.

3 Timings

For the following timing results, we calculated $P(X > Y)$ for beta, gamma, and Weibull distributions. In each case the calculation was carried out 100,000 times with parameters chosen randomly. For Weibull and gamma distributions, the shape and scale parameters were chosen uniformly from [0.5, 10.5]. For beta distributions, the parameters were chosen uniformly from (0, 100). Probabilities were calculated by simulation and by integration where the target accuracy was 0.01. Times are given in seconds.

Distribution	Simulation	Integration
beta	808.0	0.281
gamma	247.1	0.391
Weibull	447.9	7.468

4 Conclusions

We have given a method for determining finite limits of integration for computing

$$P(X > Y)$$

for general distributions as well as practical suggestions on how to carry out the resulting numerical integration. We have also discussed how to handle singularities which sometimes arise when comparing beta or Weibull distributions. In the case of exponential, normal, and Cauchy distributions, the probability can be calculated in terms of elementary functions. For gamma and inverse gamma distributions, the probability can be computed in terms of an incomplete beta distribution. For beta and Weibull distributions, we have not discovered closed forms in general but have presented efficient numerical approximations.

5 Addendum

The present version of this paper corrects an error in the results for the Cauchy distribution. Also, additional references have been added to results published since the first version of this paper appeared, namely Cook (2005, 2005) and Cook and Nadarajah (2006).

6 References

Abramowitz, M. and Stegun, I. (1970) *Handbook of Mathematical Functions*, Dover

Berry, D. A. (2003) Statistical Innovations in Cancer Research. In *Cancer Medicine e.6.* Chapter 33. London: BC Decker. (Ed: Holland J, Frei T et al.)

Berry, D. A. (2004) Bayesian statistics and the ethics of clinical trials. *Statistical Science* Volume 19(1):175-187. (To appear.)

Casella, G. and Berger, R. L. (2001) *Statistical Inference*, Brooks Cole, 2nd edition

Cook, J. D. Exact calculation of beta inequalities (2005). Technical Report UTMDABTR-006-05.

Cook, J. D. Numerical evaluation of gamma inequalities (2006). Technical Report UTMDABTR-001-06.

Cook, J. D. and Nadarajah, S. Stochastic Inequality Probabilities for Adaptively Randomized Clinical Trials. *Biometrical Journal.* 48 (2006) pp 256-365.

DiDonato, A. R. and Morris, A. H. (1992) Algorithm 708: Significant Digit Computation of the Incomplete Beta Function Ratios. *ACM Transactions on Mathematical Software (TOMS) archive* Volume 18, Issue 3. pp. 360 - 373

Giles, F. et al. (2003) Adaptive Randomization Study of Idarubicin and Cytarabine Versus Troxacitabine and Cytarabine Versus Troxacitibine and Idarubicin in Untreated Patients 50 Year or Older With Adverse Karyotype Acute Myeloid Leukemia *Journal of Clinical Oncology*, Vol 21, No 9, pp 1722-1727.

Orwant, J. (2002), *Computer Science & Perl Programming: Best of The Perl Journal*, O'Reilly

Piessens, R., Kapenga, E., Berhuber, C. W., and Kahaner, D. K. (1983) *QUADPACK: A Subroutine Package for Automatic Integration.* Springer-Verlag.

Thompson, W. R. (1933) On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, Volume 25, Issue 4. pp. 285 - 294.

Wise, M. E. (1960) On Normalizing the Incomplete Beta-Function for Fitting to Dose-Response Curves. *Biometrika*, Volume 47, Issue 1. pp. 173 - 175.