Approximating random inequalities with Edgeworth expansions

John D. Cook*

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

Approximating random inequalities with Edgeworth expansions

John D. Cook

Abstract

Random inequalities of the form Prob (x $>$ y + ?) Prob often appear as part of Bayesian clinical trial methods. Simulating trial designs could require calculating millions of random inequalities. When these inequalities require numerical integration, or worse random sampling, the inequality calculations account for the large majority of the simulation time. In this paper we show how to approximate random inequalities using Edgeworth expansions. The calculations required to use these expansions can be done in closed form, as we will see below. Although the calculations are elementary, they are also somewhat tedious, and so we include Python code to illustrate how to use the approximations in practice. We make no distributional assumptions on the random variables X and Y other than requiring that the necessary moments exist. The accuracy of the approximation will depend on how well the densities of these random variables are approximated by the Edgeworth expansions.

Approximating random inequalities with Edgeworth expansions

John D. Cook

November 3, 2012

Abstract

Random inequalities of the form

$$\text{Prob}(X > Y + \delta)$$

often appear as part of Bayesian clinical trial methods. Simulating trial designs could require calculating millions of random inequalities. When these inequalities require numerical integration, or worse random sampling, the inequality calculations account for the large majority of the simulation time.

In this paper we show how to approximate random inequalities using Edgeworth expansions. The calculations required to use these expansions can be done in closed form, as we will see below. Although the calculations are elementary, they are also somewhat tedious, and so we include Python code to illustrate how to use the approximations in practice.

We make no distributional assumptions on the random variables X and Y other than requiring that the necessary moments exist. The accuracy of the approximation will depend on how well the densities of these random variables are approximated by the Edgeworth expansions.

1 Approximations and integrals

Random inequalities between normal variates can be computed in closed form [1]. Using a normal approximation corresponds to a 0th order Edgeworth expansion, and works surprisingly well for beta random variables [2]. However, normal approximations have symmetric densities and can be a poor fit for random variables that are significantly non-symmetric. Adding one more term to the Edgeworth expansion creates an asymmetric approximation (if the density being approximated is asymmetric) and hence improves accuracy.

Let X and Y be independent random variables whose third moments exist. Let $\mu_X = E(X)$, $\sigma_X^2 = Var(X)$, and $\gamma_X = E(X^3)$ and similar for Y. Without loss of generality, we can assume $\mu_x = 0$ and $\sigma_X = 1$. This is because

$$\text{Prob}(X > Y + \delta) = \text{Prob}\left(\frac{X - \mu_X}{\sigma_X} > \frac{Y + \delta - \mu_X}{\sigma_X}\right)$$

and so if necessary we replace Y with $(Y + \delta - \mu_X)/\sigma_X$. We can also assume $\delta = 0$ by absorbing the δ into the new definition of Y.

Let f_X denote the PDF of X and F_Y the CDF of Y. Then

$$\text{Prob}(X > Y) = \int_{-\infty}^{\infty} f_X(x)\, F_Y(x)\, dx.$$

We will approximate this integral by using Edgeworth approximations.

Denote the PDF of a Normal(μ, σ) random variable by

$$\phi(x; m, s) = \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(x-m)^2}{2s^2}\right).$$

When $\mu = 0$ and $\sigma = 1$, denote $\phi(x; \mu, \sigma)$ simply as $\phi(x)$. Similarly, Denote the CDF of a Normal(μ, σ) random variable by $\Phi(x; \mu, \sigma)$.

The first-order Edgeworth approximations of f_X and F_Y are

$$f_X(x) \approx g_X(x) = \phi(x)\left(1 + \frac{\gamma_X}{6} H_3(x)\right)$$

and

$$F_Y(x) \approx G_Y(x) = \Phi(x;\mu_Y,\sigma_Y) - \frac{\gamma_Y\,\sigma_Y}{6} H_2\left(\frac{x-\mu_Y}{\sigma_Y}\right)\phi(x;\mu_Y,\sigma_Y)$$

where $H_2(x) = x^2 - 1$ and $H_3(x) = x^3 - 3x$ are the second and third Hermite polynomials.

Then

$$\mathrm{Prob}(X > Y) \approx \int_{-\infty}^{\infty} g_X(x)\,G_Y(x)\,dx.$$

This integral equals

$$A - \frac{\sigma_Y\gamma_Y}{6}B + aC - \frac{\gamma_X\gamma_Y\sigma_Y}{6}D$$

where

$$\begin{aligned}
A &= \int_{-\infty}^{\infty}\phi(x)\,\Phi(x;\mu_Y,\sigma_Y)\,dx\\
B &= \int_{-\infty}^{\infty}H_2\left(\frac{x-\mu_Y}{\sigma_Y}\right)\phi(x)\,\Phi(x;\mu_Y,\sigma_Y)\,dx\\
C &= \int_{-\infty}^{\infty}H_3(x)\,\phi(x)\,\Phi(x;\mu_Y,\sigma_Y)\,dx\\
D &= \int_{-\infty}^{\infty}H_3(x)\,H_2\left(\frac{x-\mu_y}{\sigma_Y}\right)\phi(x)\,\Phi(x;\mu_Y,\sigma_Y)\,dx.
\end{aligned}$$

These integrals can be evaluated in closed form as we will show bellow. Note that adding more terms to the Edgeworth expansion would add more integrals, but it would not add a new *type* of integral, *i.e.* all additional integrals could be computing using the techniques that follow.

In [1] we show that

$$A = \Phi\left(\frac{-\mu_Y}{\sqrt{1+\sigma_Y^2}}\right).$$

For the rest of the integrals we need the following result: The product of two normal PDFs is given by the equation

$$\phi(x;\mu_1,\sigma_1)\,\phi(x;\mu_2,\sigma_2) = \phi(\mu_1;\mu_2,\sqrt{\sigma_1^2+\sigma_2^2})\,\phi(x,\mu,\sigma)$$

where

$$\mu = \frac{\sigma_1^{-2}\mu_1 + \sigma_2^{-2}\mu_2}{\sigma_1^{-2} + \sigma_2^{-2}}$$

and

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

For our purposes $\mu_1 = 0$ and $\sigma_1 = 0$.

It follows that

$$\begin{aligned}
B &= \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2}) \int_{-\infty}^{\infty} H_2\left(\frac{x-\mu_Y}{\sigma_Y}\right) \phi(x; \mu, \sigma) \\
&= \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2}) \int_{-\infty}^{\infty} \left(\frac{x^2}{\sigma_Y^2} - \frac{2\mu_Y}{\sigma_Y^2} + \frac{\mu_Y^2}{\sigma_Y^2} - 1\right) \phi(x; \mu, \sigma) \\
&= \frac{1}{\sigma_Y^2}\, \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2}) \left(\sigma^2 + \mu^2 - 2\mu_Y\mu + \mu^2 - \sigma_Y^2\right).
\end{aligned}$$

To evaluate C, we use the relation

$$\frac{d}{dx} - H_2(x)\, \phi(x) = H_3(x)\, \phi(x)$$

to integrate by parts.

We have

$$\begin{aligned}
C &= \int_{-\infty}^{\infty} H_3(x)\, \phi(x)\, \Phi(x; \mu_Y, \sigma_Y)\, dx \\
&= \int_{-\infty}^{\infty} H_2(x)\, \phi(x)\, \phi(x; \mu_Y, \sigma_Y)\, dx \\
&= \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2}) \int_{-\infty}^{\infty} (x^2 - 1)\, \phi(x; \mu, \sigma)\, dx \\
&= \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2})\, (\sigma^2 + \mu^2 - 1)
\end{aligned}$$

Finally,

$$\begin{aligned}
D &= \int_{-\infty}^{\infty} H_3(x)\, H_2\left(\frac{x-\mu_y}{\sigma_Y}\right) \phi(x)\, \Phi(x; \mu_Y, \sigma_Y)\, dx \\
&= \phi(0; \mu_Y, \sqrt{1+\sigma_Y^2}) \int_{-\infty}^{\infty} H_3(x)\, H_2\left(\frac{x-\mu_y}{\sigma_Y}\right) \phi(x; \mu, \sigma)\, dx
\end{aligned}$$

$$= \frac{1}{\sigma_Y^2}\phi(0;\mu_Y,\sqrt{1+\sigma_Y^2})\int_{-\infty}^{\infty}(x^3-3)(x^2-2\mu_Y x+\mu_Y^2-\sigma_Y^2)\,\phi(x;\mu,\sigma)\,dx$$
$$= \frac{1}{\sigma_Y^2}\phi(0;\mu_Y,\sqrt{1+\sigma_Y^2})\,(m_5-2\mu_Y m_4+(\mu_Y^2-\sigma_Y^2-3)m_3+6\mu_Y m_2-3(\mu_Y^2-\sigma_Y^2)m_1)$$

where

$$m_r = E(W^r)$$

and $W \sim \mathrm{Normal}(\mu,\sigma)$.

2 Sample code

```
from scipy.stats import norm

from scipy import sqrt


def phi(x, mu = 0, sigma = 1):

    "Normal pdf"

    return norm.pdf(x, mu, sigma)


def Phi(x, mu = 0, sigma = 1):

    "Normal cdf"

    return norm.cdf(x, mu, sigma)


def H2(x):

    "Second Hermite polynomial, probabilist version"

    return x*x - 1.


def H3(x):

    "Third Hermite polynomial, probabilist version"
```

```
    return x**3 - 3*x


def edge_ineq(mu_x, sigma_x, gamma_x, mu_y, sigma_y, gamma_y, delta):
    "Calculate P(X > Y + delta) given mean, stdev, and 3rd moment of X and Y"
    "gamma_x = E(X^3), gamma_y = E(Y^3)"

    mu_y = (mu_y + delta - mu_x)/sigma_x
    sigma_y /= sigma_x

    mu = mu_y/(sigma_y**2 + 1.)
    sigma = sigma_y**2/(sigma_y**2 + 1.)
    product_normalization = phi(0, mu_y, sqrt(sigma_y**2 + 1.))

    # nth moments of normal distribution
    m1 = mu
    m2 = sigma**2 + mu*2
    m3 = mu*(mu**2 + 3*sigma**2)
    m4 = mu**4 + 6*mu**2*sigma**2 + 3*sigma**4
    m5 = mu*(mu**4 + 10*mu**2*sigma**2 + 15*sigma**4)

    integral1 = Phi(-mu_y/sqrt(sigma_y**2 + 1.))

    integral2 = product_normalization
    integral2 *= -gamma_y/(6.*sigma_y)
    integral2 *= (m2 - 2.*mu_y*mu + mu_y**2 - sigma_y**2)
```

```
integral3 = product_normalization
integral3 *= gamma_x/6.
integral3 *= m2 - 1

integral4  = product_normalization
integral4 *= -gamma_x*gamma_y/(36. * sigma_y)
s =  m5 - 2.*mu_y*m4 + (mu_y**2 - sigma_y**2 - 3.)*m3
s += 6*mu_y*m2 - 3*(mu_y**2 - sigma_y**2)*m1
integral4 *= s

return integral1 + integral2 + integral3 + integral4
```

3 References

[1] John D. Cook. Numerical computation of stochastic inequality probabilities (2003). UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series. Working Paper 46.
http://www.bepress.com/mdandersonbiostat/paper46/

[2] John D. Cook Fast approximation of beta inequalities (2012). UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series. Working Paper 76.
http://www.bepress.com/mdandersonbiostat/paper76/