Re: Overlapping probability distributions

Source: http://sci.tech−archive.net/Archive/sci.math/2009−02/msg00152.html

---

- *From*: Ray Vickson <RGVickson@xxxxxxx>
- *Date*: Sun, 1 Feb 2009 17:15:33 −0800 (PST)

On Feb 1, 4:42 pm, sert <je...@xxxxxxxxxxx> wrote:

> We have two probability distributions, A and B. We need to determine the probability that a value randomly picked from A will be higher than another value randomly picked from B. Is it possible to do this in a deterministic way or do we need to do use a Monte−Carlo simulation?
>
> For example, suppose we know that the heights of men and women conform to the normal distribution, with known mean and deviation. What is the probability that a random man is taller than a random woman?

If X is from A and Y is from B, then (given the assumed *independence* of the two draws) the pair (X,Y) has the bivariate distribution A(x)*B (y). Now you want the probability P{X > Y}, computed from the bivariate distribution. I will assume continuous distributions with _densities_ a(x) and b(y) in the following−−−−if the distributions are discrete, instead, just replace integrals by sums. Method: P{X > Y} = integral of a(x)*b(y) dx*dy over the 2D region R = {x > y}. This can be expressed as integral{y = −infinity ..infinity} integral{x = y .. infinity} a(x)*b(y) dx dy = integral{y = −inf..inf} b(y)*AA(y) dy, where AA(y) = P{X > y} = integral{x=y..inf} a(x) dx. So, the problem now reduces to a 1−dimensional integration. In some cases the functions b(y) and AA(y) are simple enough to allow an explicit integration in closed form; if not, you can easily do /numerical integration/, which, in 1 dimension is well−studied and has effective methods widely available. Of course, this assumes that AA(y) is easily computed. If necessary, do a numerical integration also to get AA(y) for all the y values needed for the main numerical integration. Alternatively, you can write P{X > Y} = integral{x=−inf..inf} a(x)*B (x) dx, where B(x) = P{Y < x} = integral{y=−inf..x} b(y) dy; this may be easier than the other way round.

For the specific case of normal A and B, say with means m (men) and w (women) and variances vm and vw, resp., the random variable D = X−Y is normal again, with mean d = m−w and variance vd = vm + vw. Now you just want the probability P{D > 0} = P{N(0,1) > −d/sqrt(vd)}, which can be found from normal tables, or by pushing a button on a scientific calculator, or by using a spreadsheet, or whatever. [Note: the fact that D has a normal distribution again is a standard property about linear combinations of independent normal random variables; see, eg., http://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables or http://mathworld.wolfram.com/NormalSumDistribution.html or http://www.xycoon.com/nor_properties6.htm . This last reference has an incorrect formula for the variance; it should be sigma^2 = sum (c_i)^2 * (sigma_i)^2 [not sum c_i * (sigma_i)^2].

R.G. Vickson
.