On simulating tail proportion ratios

A tail occupies some fraction of a distribution, and F is the size of that fraction. The location of the cut-point is C. For a left tail, the quantile of the cut-point is equal to F and C is the upper bound of the tail. For a right tail, the quantile is $1 - F$ and C is the lower bound of the tail. The difference between C and the mean is x. For example, if we select the right tail of the standard normal distribution and a cut-point at the 90th percentile, $F = 0.1$ and $x = C \approx 1.28155$.

The tail proportion ratio T is a relational measure of density beyond a given cut-point, which compares the proportions of two distributions above or below C in the form of a ratio.

Imagine two normal distributions with general properties: $N(M_1, {s_1}^2)$ and $N(M_2, {s_2}^2)$, with means M_1 and M_2 and standard deviations s_1 and s_2, respectively. Let $M = M_1 - M_2$ and $s = s_1/s_2$. For our purposes, the given variables are T, F, M and we want to solve for the appropriate s. We do not wish to select a value for M itself, but a value of Cohen's d which can then be converted to M. So we would like a function with inputs T, F, d and an output s. But s and x are interdependent, and the conversion from d to M depends on s, so M, x, and s are all interdependent. A system of equations is needed. The equations below will use the error function, erf, to provide tractable expressions. The following properties will be exploited with the aim of solving for the standard deviation ratio s.

Let $t = \dfrac{x-\mu}{\sigma\sqrt{2}}$. A normal distribution $N(\mu, \sigma^2)$ has mean μ and standard deviation σ.

$$\int_y^z N(\mu,\sigma^2)\,dt = \int_y^z \frac{e^{-t^2/2}}{\sigma\sqrt{2\pi}}\,dt = \frac{\operatorname{erf}\left(\dfrac{z-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\dfrac{y-\mu}{\sigma\sqrt{2}}\right)}{2}$$

$$\lim_{z\to\infty} \operatorname{erf}(z) = 1 \quad \text{and} \quad \lim_{z\to-\infty} \operatorname{erf}(z) = -1$$

We can thus express the cumulative distribution function, $\Phi(z)$:

$$\int_{-\infty}^z N(\mu,\sigma^2)\,dt = \frac{\operatorname{erf}\left(\dfrac{z-\mu}{\sigma\sqrt{2}}\right) + 1}{2}$$

And its complement, $1 - \Phi(z)$:

$$\int_z^\infty N(\mu,\sigma^2)\,dt = \frac{1 - \operatorname{erf}\left(\dfrac{z-\mu}{\sigma\sqrt{2}}\right)}{2}$$

The error function has an inverse function, erf^{-1}, as well as an inverse complementary function, erfc^{-1}. During algebraic manipulation we will use two further properties:

$$\operatorname{erf}^{-1}(1-z) = \operatorname{erfc}^{-1}(z) \quad \text{and} \quad \operatorname{erf}^{-1}(z-1) = -\operatorname{erfc}^{-1}(z)$$

The following pages will sometimes alternate between results for the left tail and right tail, with clear marking. The results are highly similar; some terms differ in sign.

Variable formulations in the left tail

The first given variable is F. Let there be a mixed normal distribution comprised in equal parts by $N(M_1, {s_1}^2)$ and $N(M_2, {s_2}^2)$. Thus $C = x + .5(M_1 + M_2)$. In the left tail, F is the ratio of $\Phi(C)$ to the total area under the curve. The properties of the normal distribution dictate that the total area under the curve is 1, so $F = \Phi(C)$.

$$F = \int_{-\infty}^{C} \frac{N(M_1, {s_1}^2) + N(M_2, {s_2}^2)}{2}\, dt = \frac{\mathrm{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right) + \mathrm{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right) + 2}{4}$$

The second given variable is T. Here we will consider the two halves of our mixed distribution separately, and formulate an expression of T that refers to the left tail. In general, $\Phi(C)$ of $N(M_1, {s_1}^2)$ is not equal to that of $N(M_2, {s_2}^2)$. T is the ratio of the former to the latter.

$$T = \frac{\int_{-\infty}^{C} N(M_1, {s_1}^2)\, dt}{\int_{-\infty}^{C} N(M_2, {s_2}^2)\, dt} = \frac{\mathrm{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right) + 1}{\mathrm{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right) + 1}$$

Variable formulations in the right tail

F in the right tail is equal to the complement of $\Phi(C)$.

$$F = \int_{C}^{\infty} \frac{N(M_1, {s_1}^2) + N(M_2, {s_2}^2)}{2}\, dt = \frac{2 - \mathrm{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right) - \mathrm{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right)}{4}$$

T in the right tail is the same ratio, but instead uses the complement of $\Phi(C)$.

$$T = \frac{\int_{C}^{\infty} N(M_1, {s_1}^2)\, dt}{\int_{C}^{\infty} N(M_2, {s_2}^2)\, dt} = \frac{1 - \mathrm{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right)}{1 - \mathrm{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right)}$$

We can now solve for s in terms of $\{T, F, M\}$. Later we will convert M to d and solve in terms of $\{T, F, d\}$. Some of the substitutions made in the following pages are:

$$C = x + .5(M_1 + M_2) \qquad M = M_1 - M_2 \qquad s = \frac{s_1}{s_2}$$

Simplify the F equation (left tail)

$$F = \frac{\operatorname{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right) + \operatorname{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right) + 2}{4}$$

$$4F = \operatorname{erf}\left(\frac{x - .5(M_1 - M_2)}{s_1\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x + .5(M_1 - M_2)}{s_2\sqrt{2}}\right) + 2$$

$$4F - 2 - \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

Simplify the F equation (right tail)

$$F = \frac{2 - \operatorname{erf}\left(\frac{C - M_1}{s_1\sqrt{2}}\right) - \operatorname{erf}\left(\frac{C - M_2}{s_2\sqrt{2}}\right)}{4}$$

$$4F = 2 - \operatorname{erf}\left(\frac{x - .5(M_1 - M_2)}{s_1\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x + .5(M_1 - M_2)}{s_2\sqrt{2}}\right)$$

$$4F - 2 + \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

Simplify the T equation (left tail)

$$T = \frac{\operatorname{erf}\left(\dfrac{C - M_1}{s_1\sqrt{2}}\right) + 1}{\operatorname{erf}\left(\dfrac{C - M_2}{s_2\sqrt{2}}\right) + 1}$$

$$T \operatorname{erf}\left(\frac{x + .5(M_1 - M_2)}{s_2\sqrt{2}}\right) + T = \operatorname{erf}\left(\frac{x - .5(M_1 - M_2)}{s_1\sqrt{2}}\right) + 1$$

$$T - 1 - \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) + T \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

Simplify the T equation (right tail)

$$T = \frac{1 - \operatorname{erf}\left(\dfrac{C - M_1}{s_1\sqrt{2}}\right)}{1 - \operatorname{erf}\left(\dfrac{C - M_2}{s_2\sqrt{2}}\right)}$$

$$T - T \operatorname{erf}\left(\frac{x + .5(M_1 - M_2)}{s_2\sqrt{2}}\right) = 1 - \operatorname{erf}\left(\frac{x - .5(M_1 - M_2)}{s_1\sqrt{2}}\right)$$

$$T - 1 + \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) - T \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

Solve the system for x (left tail)

$$4F - 2 - \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) - \left[T - 1 - \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) + T\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right)\right] = 0$$

$$4F - T - 1 - (T+1)\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

$$\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = \frac{4F - T - 1}{T+1} = \frac{4F}{T+1} - 1$$

$$\frac{x + .5M}{s_2\sqrt{2}} = \operatorname{erf}^{-1}\left(\frac{4F}{T+1} - 1\right) = -\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)$$

$$x = -\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)\ s_2\sqrt{2} - .5M$$

Solve the system for x (right tail)

$$4F - 2 + \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) - \left[T - 1 + \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) - T\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right)\right] = 0$$

$$4F - T - 1 + (T+1)\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = 0$$

$$\operatorname{erf}\left(\frac{x + .5M}{s_2\sqrt{2}}\right) = \frac{T + 1 - 4F}{T+1} = 1 - \frac{4F}{T+1}$$

$$\frac{x + .5M}{s_2\sqrt{2}} = \operatorname{erf}^{-1}\left(1 - \frac{4F}{T+1}\right) = \operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)$$

$$x = \operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)\ s_2\sqrt{2} - .5M$$

Solve the system for s (left tail)

$$4F - 2 - \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) - \left(\frac{4F}{T+1} - 1\right) = 0$$

$$\operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) = 4F - 2 - \frac{4F}{T+1} + 1 = \frac{4FT - T - 1}{T+1} = \frac{4FT}{T+1} - 1$$

$$\frac{x - .5M}{s_1\sqrt{2}} = \operatorname{erf}^{-1}\left(\frac{4FT}{T+1} - 1\right) = -\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)$$

$$s_1 = \frac{x - .5M}{-\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)\sqrt{2}} = \frac{-\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)\, s_2\sqrt{2} - M}{-\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)\sqrt{2}}$$

$$s = \frac{\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right) + \frac{M}{s_2\sqrt{2}}}{\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)}$$

Solve the system for s (right tail)

$$4F - 2 + \operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) + \left(1 - \frac{4F}{T+1}\right) = 0$$

$$\operatorname{erf}\left(\frac{x - .5M}{s_1\sqrt{2}}\right) = 1 - 4F + \frac{4F}{T+1} = \frac{T + 1 - 4FT}{T+1} = 1 - \frac{4FT}{T+1}$$

$$\frac{x - .5M}{s_1\sqrt{2}} = \operatorname{erf}^{-1}\left(1 - \frac{4FT}{T+1}\right) = \operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)$$

$$s_1 = \frac{x - .5M}{\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)\sqrt{2}} = \frac{\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right)\, s_2\sqrt{2} - M}{\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)\sqrt{2}}$$

$$s = \frac{\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right) - \frac{M}{s_2\sqrt{2}}}{\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right)}$$

So the standard deviation ratio s cannot be solved for purely in terms of $\{T, C, M\}$. Either s_1 or s_2 must also be in the expression. The solution in terms of s_1 is easily derived from the solution in terms of s_2 above. Here are both:

$$s \;=\; \frac{\operatorname{erfc}^{-1}\left(\dfrac{4F}{T+1}\right) \pm \dfrac{M}{s_2\sqrt{2}}}{\operatorname{erfc}^{-1}\left(\dfrac{4FT}{T+1}\right)} \;=\; \frac{\operatorname{erfc}^{-1}\left(\dfrac{4F}{T+1}\right)}{\operatorname{erfc}^{-1}\left(\dfrac{4FT}{T+1}\right) \mp \dfrac{M}{s_1\sqrt{2}}}$$

Quantile Function

A quantile function inputs a quantile and outputs an absolute location. For ease of implementation in a computer program, we will now express the solutions by using the normal distribution's quantile function, Φ^{-1}. For $0 < z < 2$:

$$\operatorname{erfc}^{-1}(z) = \frac{-\Phi^{-1}(z/2)}{\sqrt{2}}$$

Therefore:

$$\operatorname{erfc}^{-1}\left(\frac{4F}{T+1}\right) = -\Phi^{-1}\left(\frac{2F}{T+1}\right)\Big/\sqrt{2}$$

$$\operatorname{erfc}^{-1}\left(\frac{4FT}{T+1}\right) = -\Phi^{-1}\left(\frac{2FT}{T+1}\right)\Big/\sqrt{2}$$

And:

$$s \;=\; \frac{\operatorname{erfc}^{-1}\left(\dfrac{4F}{T+1}\right)}{\operatorname{erfc}^{-1}\left(\dfrac{4FT}{T+1}\right) \mp \dfrac{M}{s_1\sqrt{2}}} \;=\; \frac{\Phi^{-1}\left(\dfrac{2F}{T+1}\right)}{\Phi^{-1}\left(\dfrac{2FT}{T+1}\right) \pm \dfrac{M}{s_1}}$$

$$s \;=\; \frac{\operatorname{erfc}^{-1}\left(\dfrac{4F}{T+1}\right) \pm \dfrac{M}{s_2\sqrt{2}}}{\operatorname{erfc}^{-1}\left(\dfrac{4FT}{T+1}\right)} \;=\; \frac{\Phi^{-1}\left(\dfrac{2F}{T+1}\right) \mp \dfrac{M}{s_2}}{\Phi^{-1}\left(\dfrac{2FT}{T+1}\right)}$$

Note that the $\pm$ and $\mp$ switched when converting from erfc^{-1} to Φ^{-1}.

Convert raw mean difference M to Cohen's d (left tail)

$$d = \frac{M}{\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}} \quad \therefore \quad M = d\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}$$

$$\text{Let } a = \Phi^{-1}\left(\frac{2F}{T+1}\right) \text{ and } b = \Phi^{-1}\left(\frac{2FT}{T+1}\right)$$

$$s = \frac{a - \frac{M}{s_2}}{b} \quad \therefore \quad s_1 = \frac{as_2 - M}{b} = \frac{as_2 - d\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}}{b}$$

$$bs_1 = as_2 - d\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}$$

$$\left(\sqrt{2}\,(as_2 - bs_1)\right)^2 = (d\sqrt{{s_1}^2 + {s_2}^2})^2$$

$$2a^2{s_2}^2 - 4abs_1s_2 + 2b^2{s_1}^2 = d^2{s_1}^2 + d^2{s_2}^2$$

We can arrange this equation in two highly similar ways, to solve for either s_1 or s_2 using the quadratic formula.

$$(2b^2 - d^2){s_1}^2 - 4abs_2s_1 + (2a^2 - d^2){s_2}^2 = 0$$

$$(2a^2 - d^2){s_2}^2 - 4abs_1s_2 + (2b^2 - d^2){s_1}^2 = 0$$

$$s_1 = \frac{4abs_2 \pm \sqrt{16a^2b^2{s_2}^2 - 4(2b^2 - d^2)(2a^2 - d^2){s_2}^2}}{2(2b^2 - d^2)} = s_2\,\frac{2ab \pm d\sqrt{2a^2 + 2b^2 - d^2}}{2b^2 - d^2}$$

$$s_2 = \frac{4abs_1 \mp \sqrt{16a^2b^2{s_1}^2 - 4(2a^2 - d^2)(2b^2 - d^2){s_1}^2}}{2(2a^2 - d^2)} = s_1\,\frac{2ab \mp d\sqrt{2a^2 + 2b^2 - d^2}}{2a^2 - d^2}$$

This provides us with two solutions for s in terms of $\{T, F, d\}$.

$$s = \frac{2ab \pm d\sqrt{2a^2 + 2b^2 - d^2}}{2b^2 - d^2} = \frac{2a^2 - d^2}{2ab \mp d\sqrt{2a^2 + 2b^2 - d^2}}$$

Convert raw mean difference M to Cohen's d (right tail)

We use the same equation except with M's altered sign.

$$s_1 = \frac{as_2 + M}{b} = \frac{as_2 + d\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}}{b}$$

$$bs_1 = as_2 + d\sqrt{\frac{{s_1}^2 + {s_2}^2}{2}}$$

$$\left(\sqrt{2}\,(bs_1 - as_2)\right)^2 = (d\sqrt{{s_1}^2 + {s_2}^2})^2$$

Because $(as_2 - bs_1)^2 = (bs_1 - as_2)^2$, it is clear that the solutions will be the same as in the left tail.

$$s \;=\; \frac{2ab \pm d\sqrt{2a^2 + 2b^2 - d^2}}{2b^2 - d^2} \;=\; \frac{2a^2 - d^2}{2ab \mp d\sqrt{2a^2 + 2b^2 - d^2}}$$

We can simplify the equations by redefining a and b. Increase them both by a factor of $\sqrt{2}$:

$$\text{Let } a = \Phi^{-1}\left(\frac{2F}{T+1}\right)\sqrt{2} \text{ and } b = \Phi^{-1}\left(\frac{2FT}{T+1}\right)\sqrt{2}$$

$$s \;=\; \frac{ab \pm d\sqrt{a^2 + b^2 - d^2}}{b^2 - d^2} \;=\; \frac{a^2 - d^2}{ab \mp d\sqrt{a^2 + b^2 - d^2}}$$

Every instance of 2 is now absorbed into a and b.

With regard to the $\pm$: plus applies to the left tail, minus applies to the right tail.

Naturally, the $\mp$ is reversed: minus applies to the left tail, plus applies to the right tail.

If $d = 0$, the equation simplifies to $s = {}^{a}\!/_{b}$, which is identical to the previous equation if $M = 0$.

Summary

No mean difference:

$$s = \frac{\Phi^{-1}\left(\dfrac{2F}{T+1}\right)}{\Phi^{-1}\left(\dfrac{2FT}{T+1}\right)}$$

Mean difference expressed in terms of M:

$$s = \frac{\Phi^{-1}\left(\dfrac{2F}{T+1}\right)}{\Phi^{-1}\left(\dfrac{2FT}{T+1}\right) \pm \dfrac{M}{s_1}} = \frac{\Phi^{-1}\left(\dfrac{2F}{T+1}\right) \mp \dfrac{M}{s_2}}{\Phi^{-1}\left(\dfrac{2FT}{T+1}\right)}$$

Mean difference expressed in terms of d:

$$s = \frac{ab \pm d\sqrt{a^2+b^2-d^2}}{b^2-d^2} = \frac{a^2-d^2}{ab \mp d\sqrt{a^2+b^2-d^2}}$$

$$a = \Phi^{-1}\left(\frac{2F}{T+1}\right)\sqrt{2} \quad \text{and} \quad b = \Phi^{-1}\left(\frac{2FT}{T+1}\right)\sqrt{2}$$

In both cases of the $\pm$: plus refers to the left tail and minus refers to the right tail.
In both cases of the $\mp$: minus refers to the left tail and plus refers to the right tail.