

# On simulating tail proportion ratios

A tail occupies some fraction of a distribution, and  $C$  is the size of that fraction. The location of the cut-point is  $L$ . For a left tail, the quantile of the cut-point is equal to  $C$  and  $L$  is the upper bound of the tail. For a right tail, the quantile is  $1 - C$  and  $L$  is the lower bound of the tail. The difference between  $L$  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,  $C = 0.1$  and  $x = L \approx 1.28155$ .

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  $L$  in the form of a ratio.

Imagine two distributions: the standard normal distribution,  $N(0, 1)$ , and a normal distribution with mean  $M$  and standard deviation  $s$ ,  $N(M, s^2)$ . For our purposes,  $T$ ,  $C$ , and  $M$  are given variables 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  $T$ ,  $C$ ,  $d$  and an output  $s$ . But  $s$  and  $x$  are interdependent, and the conversion from  $d$  to  $M$  depends on  $s$ , so  $s$  and  $M$  are interdependent as well. A system of equations is needed. The equations below will use the error function,  $\text{erf}$ , to provide tractable expressions. The following properties will be exploited with the aim of solving for the standard deviation ratio  $s$ .

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

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

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

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

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

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

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

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

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

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  $C$ . Let there be a mixed normal distribution comprised in equal parts by  $N(M, s^2)$  and  $N(0, 1)$ . Thus  $L = x + .5M$ . In the left tail,  $C$  is the ratio of  $\Phi(L)$  to the total area under the curve. The properties of the normal distribution dictate that the total area under the curve is 1, so  $C = \Phi(L)$ .

$$C = \int_{-\infty}^L \frac{N(M, s^2) + N(0, 1)}{2} dt = \frac{\text{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) + \text{erf}\left(\frac{x + .5M}{\sqrt{2}}\right) + 2}{4}$$

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,  $\Phi(L)$  of  $N(M, s^2)$  is not equal to that of  $N(0, 1)$ .  $T$  is the ratio of the former to the latter.

$$T = \frac{\int_{-\infty}^L N(M, s^2) dt}{\int_{-\infty}^L N(0, 1) dt} = \frac{\text{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) + 1}{\text{erf}\left(\frac{x + .5M}{\sqrt{2}}\right) + 1}$$

### Variable formulations in the right tail

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

$$C = \int_L^{\infty} \frac{N(M, s^2) + N(0, 1)}{2} dt = \frac{2 - \text{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) - \text{erf}\left(\frac{x + .5M}{\sqrt{2}}\right)}{4}$$

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

$$T = \frac{\int_L^{\infty} N(M, s^2) dt}{\int_L^{\infty} N(0, 1) dt} = \frac{1 - \text{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right)}{1 - \text{erf}\left(\frac{x + .5M}{\sqrt{2}}\right)}$$

We can now solve for  $s$  in terms of  $\{T, C, M\}$ . Later we will convert  $M$  to  $d$  and solve in terms of  $\{T, C, d\}$ .

Simplify the  $C$  equation (left tail)

$$C = \frac{\operatorname{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x + .5M}{\sqrt{2}}\right) + 2}{4}$$

$$4C = \operatorname{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x + .5M}{\sqrt{2}}\right) + 2$$

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

Simplify the  $C$  equation (right tail)

$$C = \frac{2 - \operatorname{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x + .5M}{\sqrt{2}}\right)}{4}$$

$$4C = 2 - \operatorname{erf}\left(\frac{x - .5M}{s\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x + .5M}{\sqrt{2}}\right)$$

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

Simplify the  $T$  equation (left tail)

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

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

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

Simplify the  $T$  equation (right tail)

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

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

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

Solve the system for  $x$  (left tail)

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

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

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

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

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

Solve the system for  $x$  (right tail)

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

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

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

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

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

Solve the system for  $s$  (left tail)

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

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

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

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

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

Solve the system for  $s$  (right tail)

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

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

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

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

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

## Quantile Function

For ease of implementation in computer code, we will now express the solutions in terms of the normal distribution's quantile function,  $\Phi^{-1}$ . For  $0 < z < 2$ :

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

Therefore:

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

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

And:

$$s = \frac{\operatorname{erfc}^{-1}\left(\frac{4C}{T+1}\right) \pm \frac{M}{\sqrt{2}}}{\operatorname{erfc}^{-1}\left(\frac{4CT}{T+1}\right)} = \frac{\Phi^{-1}\left(\frac{2C}{T+1}\right) \pm M}{\Phi^{-1}\left(\frac{2CT}{T+1}\right)}$$

Note that if  $M = 0$ , the next step is not necessary.

Further, if  $M = 0$  and  $T = 1$ , the distributions within the mixed normal distribution are identical and the standard deviation ratio  $s$  always simplifies to 1.

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

$$d = \frac{M}{\sqrt{s^2 + 1}} \quad \therefore \quad M = d\sqrt{s^2 + 1}$$

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

$$s = \frac{a + M}{b} = \frac{a + d\sqrt{s^2 + 1}}{b}$$

$$sb = a + d\sqrt{s^2 + 1}$$

$$(sb - a)^2 = (d\sqrt{s^2 + 1})^2$$

$$b^2 s^2 - 2abs + a^2 = d^2 s^2 + d^2$$

$$(b^2 - d^2)s^2 - 2abs + a^2 - d^2 = 0$$

$$s = \frac{2ab \pm \sqrt{4a^2b^2 - 4(b^2 - d^2)(a^2 - d^2)}}{2(b^2 - d^2)}$$

$$s = \frac{ab \pm \sqrt{a^2d^2 + b^2d^2 - d^4}}{b^2 - d^2}$$

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



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

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

$$s = \frac{a - M}{b} = \frac{a - d\sqrt{s^2 + 1}}{b}$$

$$sb = a - d\sqrt{s^2 + 1}$$

$$(d\sqrt{s^2 + 1})^2 = (a - sb)^2$$

Because  $(a - sb)^2 = (sb - a)^2$ , it is clear that the solution for  $s$  will be the same as in the left tail.

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

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

If  $d = 0$ , the equation simplifies to  $s = a/b$ , which is identical to the previous equation if  $M = 0$ .

The results derived above generalize broadly. Although  $M$  here refers to the center of  $N(M, s^2)$ , it can refer to any mean difference between two distributions and converts in the same way to  $d$ . And although  $s$  here refers to the standard deviation of  $N(M, s^2)$ , it can refer to the standard deviation ratio of any pair of distributions. Note that for the current pair, the mean difference is  $M - 0 = M$  and the standard deviation ratio is  $s/1 = s$ .

## Summary

No mean difference:

$$s = \frac{\Phi^{-1}\left(\frac{2C}{T+1}\right)}{\Phi^{-1}\left(\frac{2CT}{T+1}\right)}$$

Mean difference expressed in terms of  $M$ :

$$s = \frac{\Phi^{-1}\left(\frac{2C}{T+1}\right) \pm M}{\Phi^{-1}\left(\frac{2CT}{T+1}\right)}$$

Mean difference expressed in terms of  $d$ :

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

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

In both cases, plus refers to the left tail and minus refers to the right tail.