Bounds for g-and-k cdf

Dennis Prangle

The g-and-k distribution is defined by its quantile function:

$$x = Q(z) = A + B\left(1 + c\frac{1-\exp(-gz)}{1+\exp(-gz)}\right)(1+z^2)^k z \tag{1}$$

where z is the corresponding standard normal quantile. To find the cdf we must numerically solve $z = Q^{-1}(x)$. This requires providing bounds $\underline{z}(x) \leq Q^{-1}(x) \leq \bar{z}(x)$. This document describes the derivation of some bounds.

We make the assumption that

$$0 \leq c < 1 \tag{2}$$

Following Rayner and MacGillivray (2002), typically $c = 0.8$ and values outside $[0, 1)$ seem of little practical interest.

First make the change of variable $y = (x - A)/B$ so that

$$y = R(z) = \left(1 + c\frac{1-\exp(-gz)}{1+\exp(-gz)}\right)(1+z^2)^k z \tag{3}$$

and we require $\underline{z} \leq R^{-1}(y) \leq \bar{z}$.

Consider $f(z) = (1-\exp(-gz)/(1+\exp(-gz))$. It is straightforward to check that $|f(z)| < 1$, $\text{sign}(f(z)) = \text{sign}(z)$, $\lim_{z\to-\infty} f(z) = -1$, $\lim_{z\to\infty} f(z) = 1$.

Using (2) it follows that $\text{sign}(y) = \text{sign}(z)$. So for $y = 0$ the cdf is known exactly. Also we can take we can take $\underline{z} = 0$ for $y > 0$ and $\bar{z} = 0$ for $y < 0$.

Case $k \geq 0$

In this case $(1+z^2)^k \geq 1$.

Suppose $y > 0$. Then $z > 0$ so $1 + cf(z) \geq 1$. Thus (3) gives $y \geq z$ so we can take $\bar{z} = y$.

Suppose instead $y < 0$. Then $z < 0$ and $1 + cf(z) \geq 1 - c$ (which is positive by (2)). Thus (3) gives $|y| \geq (1-c)|z|$, and $z \geq y/(1-c)$ allowing $\underline{z} = y/(1-c)$.

Case $k < 0$

From Rayner and MacGillivray (2002), $k > -1/2$ is required for a valid distribution. For clarity, let $m = -k$.

Consider the case $y > 0$. Then as above $1 + cf(z) \geq 1$ and (3) gives

$$y \geq (1+z^2)^{-m} z \tag{4}$$

Suppose $z \leq 1$, then $(1+z^2) \leq 2$, so $(1+z^2)^{-m}z \geq 2^{-m}z$ and (4) gives

$$y \geq 2^{-m}z > z/2$$

(n.b. $2^{-m} \geq 2^{-1/2}$, but we use 1/2 as a lower bound for convenience in the code) so we can take $\bar{z} = 2y$.

Suppose $z > 1$. Then $(1+z^2)^{-m} = z^{-2m}(1+z^{-2})^{-m} \geq z^{-2m}2^{-m} > z^{-2m}/2$. Hence (4) gives

$$y \geq z^{1-2m}/2$$

so that we can use $\bar{z} = (2y)^{1/(1-2m)}$.

For the case $y < 0$, adapting the argument of the $k \geq 0$ case gives

$$\underline{z} = \begin{cases} 2y/(1-c) & \text{for } |z| \leq 1 \\ -[2|y|/(1-c)]^{1/(1-2m)} & \text{for } |z| > 1 \end{cases}$$

Finally, note that it is easy to evaluate $|z| \leq 1$ by testing $|y| \leq R(1)$.

Summary

- For $y = 0$, $z = 0$ and it is not necessary to find bounds.
- For $k \geq 0$ and $y > 0$, $(\underline{z}, \bar{z}) = (0, y)$.
- For $k \geq 0$ and $y < 0$, $(\underline{z}, \bar{z}) = (y/(1-c), 0)$.
- For $k < 0$ and $R(1) \geq y > 0$, $(\underline{z}, \bar{z}) = (0, 2y)$.
- For $k < 0$ and $y > R(1)$, $(\underline{z}, \bar{z}) = (0, [2y]^{1/(1-2m)})$.
- For $k < 0$ and $R(-1) \leq y < 0$, $(\underline{z}, \bar{z}) = (2y/(1-c), 0)$.
- For $k < 0$ and $y < R(-1)$, $(\underline{z}, \bar{z}) = -([2|y|/\{1-c\}]^{1/(1-2m)}, 0)$.

References

Rayner, G. D. and MacGillivray, H. L. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing*, 12(1):57–75.