A letter to Dr. Renee Moore, Statistics professor at NCSU

Hello Dr. Moore,

I hope all is well with mother and newborn. I have thought of you both and wanted to check in, but also wanted to share what follows, and with travel (yes, it continues - I'm away more than home - do you know of anyone who needs a local data analyst?) it's taken me a while to type it. In addition, the NCSU mail system habitually stripped the Latex images from my drafts, so I had to restart several times. That's why you are receiving this as an attachment.

Some years ago I began using the Johnson S-Family of probability distributions to characterize observed non-normal data (from "Statistical Modeling Techniques" by S. Shapiro and A. Gross). There are three members of the family (kind of like the Moores, huh?): S_L, S_U, and S_B, each a lognormal transformation with three or four parameters (for scale, position, shape, etc). To "fit" a distribution, of course, means to express observed moments or percentiles in multiple equations in terms of the fitted distribution parameters (one equation for each parameter) and then solve for the parameters. Using the equations from Shapiro and Gross I would get an occasional "bad" fit - first percentile after the last observation, that sort of thing - and having observed numerous typographic and algebraic errors in the book I decided to derive the equations myself (the ones for moments, which I prefer, being more objective than those for percentiles, which have to be chosen). And that is where I was when thinking of you. But the question is, "Where did I end up?"

The S_U distribution is defined as

$$z = \gamma + \eta \sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right)$$

where

$$z \sim N(0,1)$$

$$\Rightarrow x = \lambda \sinh\left(\frac{z-\gamma}{\eta}\right) + \epsilon = \lambda \frac{\mathrm{e}^{\frac{z-\gamma}{\eta}} - \mathrm{e}^{-\frac{z-\gamma}{\eta}}}{2} + \epsilon$$

$$\Rightarrow x^k = \sum_{i=0}^{k} \binom{k}{i} \left[\lambda \frac{\mathrm{e}^{\frac{z-\gamma}{\eta}} - \mathrm{e}^{-\frac{z-\gamma}{\eta}}}{2}\right]^i \epsilon^{k-i}$$

(from a binomial expansion)

$$\Rightarrow x^k = \sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^i \binom{k}{i} \sum_{j=0}^{i} \binom{i}{j} \left(\mathrm{e}^{\frac{z-\gamma}{\eta}}\right)^j \left(-\mathrm{e}^{-\frac{z-\gamma}{\eta}}\right)^{i-j} \epsilon^{k-i}$$

(from another binomial expansion of the bracketed factor)

$$= \sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^{i} \binom{k}{i} \sum_{j=0}^{i} \binom{i}{j} \left(\mathrm{e}^{\frac{z-\gamma}{\eta}}\right)^{j} \left(-\mathrm{e}^{-\frac{z-\gamma}{\eta}}\right)^{i-j} \epsilon^{k-i}$$

$$= \sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^{i} \epsilon^{k-i} \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \mathrm{e}^{(2j-i)\frac{z-\gamma}{\eta}}$$

And now for the moments [remember that $z \sim N(0,1)$]:

$$E\left[f(z)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(z) \mathrm{e}^{-z^2/2} dz$$

$$\Rightarrow E\left[x^k\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^{i} \epsilon^{k-i} \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \mathrm{e}^{(2j-i)\frac{z-\gamma}{\eta}}\right] \mathrm{e}^{-z^2/2} dz$$

$$= \sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^{i} \epsilon^{k-i} \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{(2j-i)\frac{z-\gamma}{\eta}} \mathrm{e}^{-z^2/2} dz\right]$$

where the contents of the brackets becomes

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\left[\left(z-\frac{2j-i}{\eta}\right)^2 - \left(\frac{2j-i}{\eta}\right)^2 + 2\gamma\left(\frac{2j-i}{\eta}\right)\right]/2} dz$$

$$= \mathrm{e}^{-\left[\left(\frac{2j-i}{\eta}\right)^2 + 2\gamma\left(\frac{2j-i}{\eta}\right)\right]/2} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\left(z-\frac{2j-i}{\eta}\right)^2/2} dz\right]$$

$$= \mathrm{e}^{-\left[\left(\frac{2j-i}{\eta}\right)^2 + 2\gamma\left(\frac{2j-i}{\eta}\right)\right]/2}$$

$$\Rightarrow E\left[x^k\right] = \sum_{i=0}^{k} \left(\frac{\lambda}{2}\right)^{i} \epsilon^{k-i} \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \mathrm{e}^{\frac{4j^2-4ij+i^2-(4j-2i)\gamma\eta}{2\eta^2}}$$

The second summation generates the following terms:

i	j	term
0	0	1
1	0	-$e^{(1+2\gamma\eta)/2\eta^2}$
1	1	$e^{(1-2\gamma\eta)/2\eta^2}$
2	0	$e^{(4+4\gamma\eta)/2\eta^2}$
2	1	-2$e^0 = -2$
2	2	$e^{(4-4\gamma\eta)/2\eta^2}$
$\vdots$	$\vdots$	$\vdots$

giving

$$E(x) = \epsilon + \frac{\lambda}{2}\left[e^{(1-2\gamma\eta)/2\eta^2} - e^{(1+2\gamma\eta)/2\eta^2}\right]$$

$$= \epsilon - \lambda\ e^{1/2\eta^2}\sinh(\gamma/\eta)$$

and

$$E(x^2) = \epsilon^2 - 2\lambda\epsilon e^{\frac{1}{\eta^2}}\sinh(\gamma/\eta) + \lambda^2\left[e^{\frac{2}{\eta^2}}\left(\cosh^2(\gamma/\eta) - \frac{1}{2}\right) - \frac{1}{2}\right]$$

giving

$$Var(x) = \lambda^2\left[e^{\frac{1}{\eta^2}}\cosh^2\left(\frac{\gamma}{\eta}\right)\left(e^{\frac{1}{\eta^2}-1}\right) - e^{\frac{1}{\eta^2}}\left(\frac{e^{\frac{1}{\eta^2}}}{2} - 1\right) - \frac{1}{2}\right]$$

I still have some simplification to do with the variance, but these are nonetheless pretty compact, aren't they? Higher order moments follow similar derivation, but with more and more(!) terms to simplify. (What would Moore + Moore equal? One Moore? Or do you prefer Moore! Or maybe, no Moore?)

And that's where I found myself - with fairly simple equations to derive parameters for a very elastic distribution (although they do increase in complexity as order increases). In a world that assumes normality (at times insists on it - F tests and such), it's nice to have methods to compensate for deviations.

But wait, there's more! I (like everyone) have used binomial probabilities for a long time. They work so well in pass/fail quality control tests. But, in working out the above problem, I recognized that

$$x^k = \sum_{i=0}^{k}\binom{k}{i}\left[\lambda\frac{e^{\frac{z-\gamma}{\eta}} - e^{-\frac{z-\gamma}{\eta}}}{2}\right]^i \epsilon^{k-i}$$

results in a trinomial expansion through one binomial expansion nested within another. And so I began to consider trinomial probabilities (I imagined baby names like "Trini," but thought it was going too far).

Trinomial Probabilities

Let p, q, r be the probabilities of mutually exclusive events A, B, and C such that $P(A)=p$, $P(B)=q$, and $P(C)=r$ and $p+q+r=1$. Let $n, x, y, z \in \mathbb{Z}$ such that $0<n, x, y, z$ and $n=x+y+z$, where n is a total number of independent trials and x is the number of trials resulting in outcome A, y is the number of trials resulting in outcome B, and z is the number resulting in outcome C.

Then $\exists \binom{n}{x}$ combinations of trials resulting in A, $\binom{y+z}{y}=\binom{n-x}{y}$ combinations of the remaining n-x trials resulting in B, and the remaining $\binom{z}{z}=1$ combination of remaining n-x-y=z trials resulting in C.

Since the trials are independent, the probability of a succession of events equals the product of the probabilities of the individual events. Therefore, P(a given sequence of x As, y Bs, and z Cs) = $p^x q^y r^z$ = $p^x q^y (1-p-q)^{n-x-y} \Longrightarrow$ P(any sequence having x As, y Bs, and z Cs) = $P(x, y, z) = \binom{n}{x}\binom{n-x}{y} p^x q^y (1-p-q)^{n-x-y}$

Finally, all combinations of x, y, z, where $x+y+z=n$ form a pdf since $(p+q+r)^n = [p+(q+r)]^n$

$= \sum_{x=0}^{n}\binom{n}{x} p^x (q+r)^{n-x}$ (binomial expansion) $= \sum_{x=0}^{n}\binom{n}{x} p^x \sum_{y=0}^{n-x}\binom{n-x}{y} q^y r^{n-x-y}$

$= \sum_{x=0}^{n}\binom{n}{x}\binom{n-x}{y} p^x q^y (1-p-q)^{n-x-y} = \Sigma P(x,y,z) = (p+q+r)^n = 1^n = 1$

Hope all remains well.

Tom