An Introduction to M-Distributions

D. C. Vaughan
Department of Mathematics, Wilfrid Laurier University
Waterloo, Ontario Canada N2L 3C5

1. Introduction

Blattberg and Gonedes (1974), McLeay (1986) and Praetz (1972) illustrate that symmetric non-normal distributions in share price returns are justified through empirical studies. Corrado and Su (1996) and Harris, Küçüközmen, and Yilmaz (2004) find empirical evidence that conditional distributions of asset returns exhibit both non-normal kurtosis and skewness. Non-normality arises in other scientific investigations, such as gene expression (see Hardin and Wilson (2009) and chemical and nuclear measurements (see Currie (2001)). Thus, there is a wide range of applications for models that permit both non-normal kurtosis and skewness.

Several efforts have been made to create flexible models that allow wide ranges of combinations of skewness and kurtosis. Ferreira and Steel (2006) give a comprehensive overview of such methods, showing that they can be viewed as special cases of a unified structure. Such methods typically start with a symmetric distribution and then introduce a skewing mechanism through multiplication by an appropriate function.

The goal is to develop a very flexible family of distributions that have, at least asymptotically, virtually any combination of tail behaviours, such as normal-Cauchy, or normal-exponential. We want to retain explicit construction, along with good analytic properties. In particular, in a location-scale formulation, we want to be able to apply maximum likelihood.

2: Mixed Tail Distributions

First, we define a class of continuous **weighting** functions as follows:

Definition 1: We say the continuous functions w_1 and w_2 are a **symmetric weighting pair** (swp) on $\mathbb{R}$ if:

1. w_1 and w_2 are analytic, non-negative and monotone on $\mathbb{R}$;
2. $\lim_{x\to-\infty} w_1(x) = 0$; $\lim_{x\to\infty} w_1(x) = 1$; $\lim_{x\to-\infty} w_2(x) = 1$; $\lim_{x\to\infty} w_2(x) = 0$;
3. $w_1(x) = w_2(-x)$;

and

4. $w_1(x) + w_2(x) \equiv 1$ for all $x \in \mathbb{R}$.

The weights can in fact be more general than this (with the introduction of an appropriate normalizing constant), but swps provide a wide range of distributions with good properties.

Definition 2: M-Distributions: We say the random variable X with values in $\mathbb{R}$ has an M-distribution (or **mixed tail distribution**) if its density f is given by

$$f(x) \;=\; w_1(x)g_1(x) + w_2(x)g_2(x)$$

where g_1 and g_2 are densities symmetric about $x = 0$ on $\mathbb{R}$.

As with the weights, the densities do not have to be symmetric, but often families of distributions that incorporate skewness start from a specific symmetric density.

To illustrate the usefulness of choosing a swp, we have:

Theorem: The function f given in Definition 2 is a density function on $\mathbb{R}$.

Proof: This is the result of a change of variable:

$$\int_{-\infty}^{\infty}(w_1(x)g_1(x) + w_2(x)g_2(x))dx = (\int_0^{\infty} + \int_{-\infty}^{0})(w_1(x)g_1(x) + w_2(x)g_2(x))dx$$
$$= \int_0^{\infty}(w_1(x)g_1(x) + w_2(x)g_2(x))dx + \int_0^{\infty}(w_2(x)g_1(x) + w_1(x)g_2(x))dx$$
$$= \int_0^{\infty}(w_1(x)g_1(x) + w_2(x)g_1(x) + w_2(x)g_2(x) + w_1(x)g_2(x))dx$$
$$= \int_0^{\infty}(g_1(x) + g_2(x))dx = 1$$

since the densities g_1 and g_2 are assumed symmetric at 0.

There are several symmetric weighting pairs available, including:

$$\frac{\exp(-ax)}{\exp(ax)+\exp(-ax)} \quad \text{and} \quad \frac{\exp(ax)}{\exp(ax)+\exp(-ax)} \tag{1}$$
$$\frac{\exp(-ax^3)}{\exp(ax^3)+\exp(-ax^3)} \quad \text{and} \quad \frac{\exp(ax^3)}{\exp(ax^3)+\exp(-ax^3)} \tag{2}$$

(and other odd powers of x) and

$$\frac{\tan^{-1}(x)}{\pi} + 1/2 \quad \text{and} \quad \frac{\cot^{-1}(x)}{\pi}. \tag{3}$$

$$\frac{\exp(ax)}{\exp(ax)+\exp(-bx^3)} \quad \text{and} \quad \frac{\exp(-bx^3)}{\exp(ax)+\exp(-bx^3)} \tag{4}$$
$$\frac{\exp(ax)}{\exp(ax)+\exp(-bx)} \quad \text{and} \quad \frac{\exp(-bx)}{\exp(ax)+\exp(-bx)} \tag{5}$$

In fact, let $\alpha_1(x)$ and $\alpha_2(x)$ be odd functions on $\mathbb{R}$ with appropriate limiting behaviours. Then

$$\frac{e^{\alpha_1(x)}}{e^{\alpha_1(x)}+e^{\alpha_2(x)}} \quad \text{and} \quad \frac{e^{\alpha_2(x)}}{e^{\alpha_1(x)}+e^{\alpha_2(x)}} \tag{6}$$

is a swp.

Non-symmetric pair: The weights can be extended to non-symmetric pairs, such as:

$$\frac{1}{\exp(2x+2)+1} \quad \text{and} \quad \frac{1}{1+\exp(-2x-2)}. \tag{7}$$

Weighted distributions using such pairs will need a normalizing constant, but many of the calculations, including moments and percentage points, will be of the same level of complication as for the symmetric weighting pairs.

Non-symmetric weighting pairs will give greater control of convergence to tail distributions. However, for the purposes of introducing M-distributions, we will consider only symmetric weighting pairs.

Mixture distributions arise frequently in the literature and take the form:

$$\pi_1 g_1(x) + \pi_2 g_2(x)$$

where the π_is are fixed constants representing the proportion of each subpopulation in the whole. An experimental unit is from one of the subpopulations. For example, in Macdonald and Pitcher (1979) and Macdonald (1987), the lengths of 523 pike sampled in 1965 from Heming Lake, Manitoba, Canada are modeled using five age categories, and mixtures of gammas are used to represent the whole population. Other distributions such as normals and lognormals are also appropriate; see http://www.math.mcmaster.ca/peter/mix/demex/expike.html.

M-distributions can be viewed in two ways: First, they satisfy the conditions of being density functions defined on $\mathbb{R}$ with left and right tails having different asymptotic behaviours, including: normal-Student's t; normal-exponential; and exponential-Student's t. The distributions retain excellent analytic properties. They are explicit functions, and therefore provide simple models for data exploring.

Second, they can be viewed as the result of two underlying distributions that are continuously mixed according to continuous weighting functions w_1 and w_2. An observation then is a value that results from that mixture.

In general, we want one tail in

$$f(x) \;=\; w_1(x)g_1(x) + w_2(x)g_2(x)$$

to be dominated (and usually quickly) by g_1 and the other by g_2. Thus, we choose the weights w_1 and w_2 so that w_1 increases from 0 to 1, while w_2 decreases from 1 to 0. For a given pair of densities, we also want $w_1 g_1$ to go to zero much faster than $w_2 g_2$ as $x \to -\infty$, and $w_2 g_2$ to go to 0 much faster than $w_1 g_1$ as $x \to \infty$ (or vice-versa). These considerations, and the requirement that the weights do not change the analytic properties of the densities, led to the choices listed above.

Extension to location-scale model: We can introduce location ($\mu \in \mathbb{R}$) and scale ($\sigma > 0$) parameters in the usual way. The form of the density is:

$$f(x) = \frac{1}{\sigma}[w_1\left(\frac{x-\mu}{\sigma}\right) g_1\left(\frac{x-\mu}{\sigma}\right) + w_2\left(\frac{x-\mu}{\sigma}\right) g_2\left(\frac{x-\mu}{\sigma}\right)].$$

Maximum likelihood can be applied to finding estimates of μ and σ as long as it applies to g_1 and g_2 individually. For the purposes of this introduction to M-distributions, we will consider properties under the assumption $\mu = 0$ and $\sigma = 1$.

The mode is usually unique, and relatively easy to compute in many cases. For example, if we use swp (2), and the densities g_1 and g_2 each have a unique mode at 0, then so will f. If we arrange it so that g_1 and g_2 to have the same value at 0 and they have unique modes at 0, then the density f will have a unique mode at 0. Other swps may not have this property, as illustrated in the diagrams that follow.

In choosing an appropriate M-distribution to model a data set, we often make use of skewness and kurtosis of the data as a guide. We will use the usual Pearson kurtosis measure, namely $\beta_2 = \mu_4/\mu_2$. We use the Arnold-Groeneveld (1995) skewness measure: $(1 - 2F(mode)$ for illustration purposes.

Consider the following density functions:

$$g_1(x) = \exp(-x^2/2)/\sqrt{2\pi};$$

$$g_2(x) = \frac{\exp(-x^2/8)}{2\sqrt{2\pi}}, \quad g_3(x) = \frac{\text{sech}(\pi x/2)}{2}, \quad g_4(x) = \frac{1}{\pi(1+x^2)}$$

with swp:

$$w_1(x) = \frac{e^{x^3}}{e^{x^3} + e^{-x^3}}, \quad w_2(x) = \frac{e^{-x^3}}{e^{x^3} + e^{-x^3}}.$$

Examples of M distributions include:

$$f_{NN}(x) = \frac{\exp(-x^3)}{\exp(x^3) + \exp(-x^3)} g_1(x) + \frac{\exp(x^3)}{\exp(x^3) + \exp(-x^3)} g_2(x)$$

$$f_{NSH}(x) = \frac{\exp(-x^3)}{\exp(x^3) + \exp(-x^3)} g_1(x) + \frac{\exp(x^3)}{\exp(x^3) + \exp(-x^3)} g_3(x)$$

$$f_{NC}(x) = \frac{\exp(-x^3)}{\exp(x^3) + \exp(-x^3)} g_1(x) + \frac{\exp(x^3)}{\exp(x^3) + \exp(-x^3)} g_4(x)$$

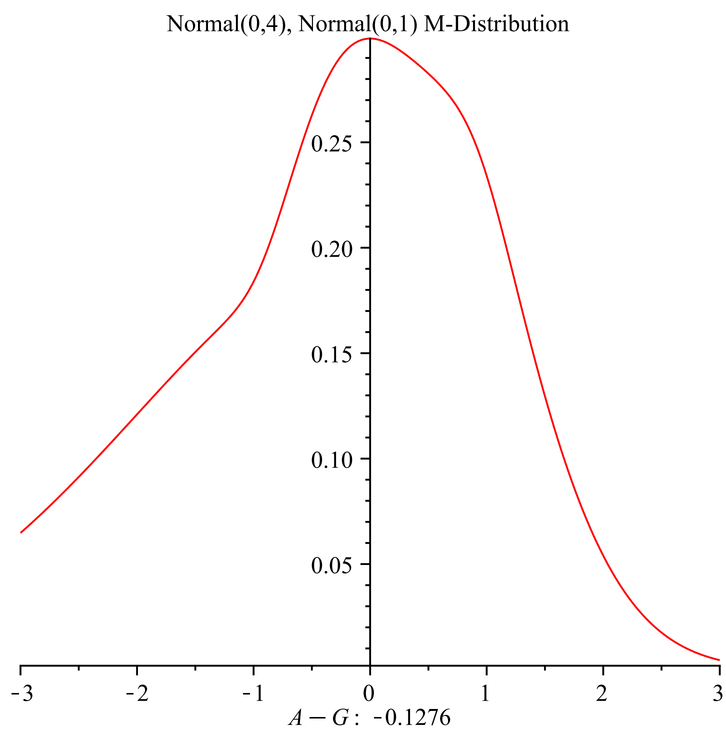
and

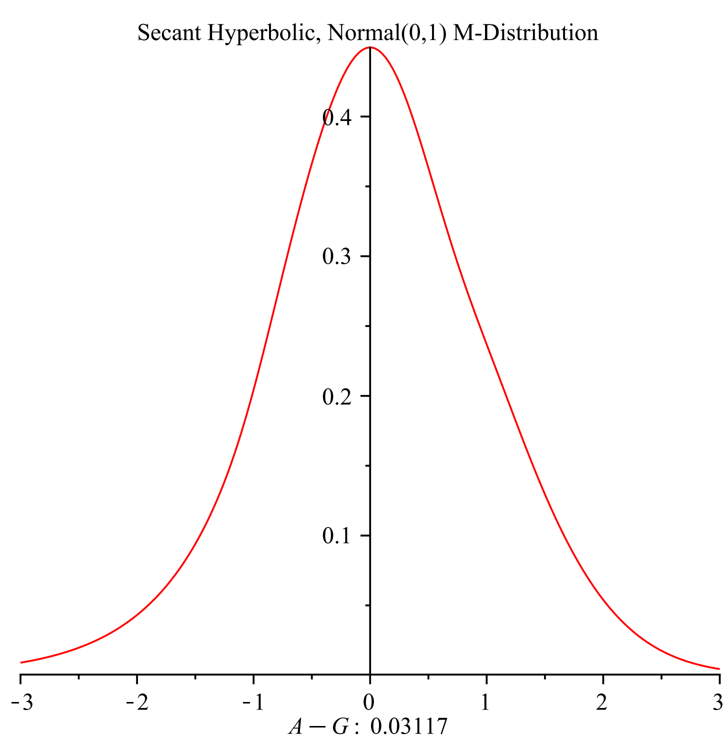
$$f_{SHC}(x) = \frac{\exp(-x)}{\exp(x) + \exp(-x)} g_3(x) + \frac{\exp(x)}{\exp(x) + \exp(-x)} g_4(x).$$

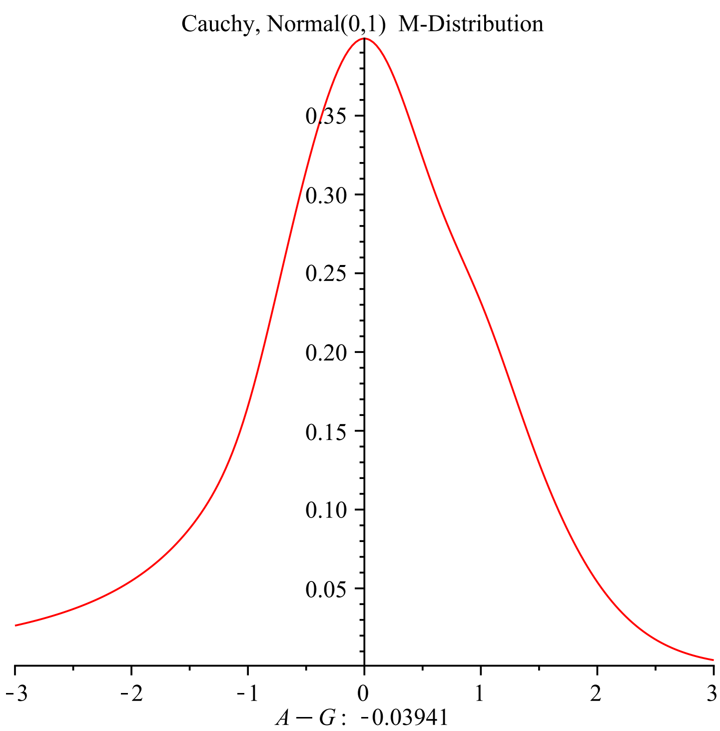
Here, NN stands for Normal-Normal, NSH for Normal-Secant Hyperbolic, NC for Normal-Cauchy, and SHC for Secant Hyperbolic - Cauchy. Thus we have (asympotically) a normal distribution in one tail, and a Student's t distribution in the other, etc. Note that if we want the distributions to reach the asymptotic tails faster, constants can be used in the weight functions to speed up the process.

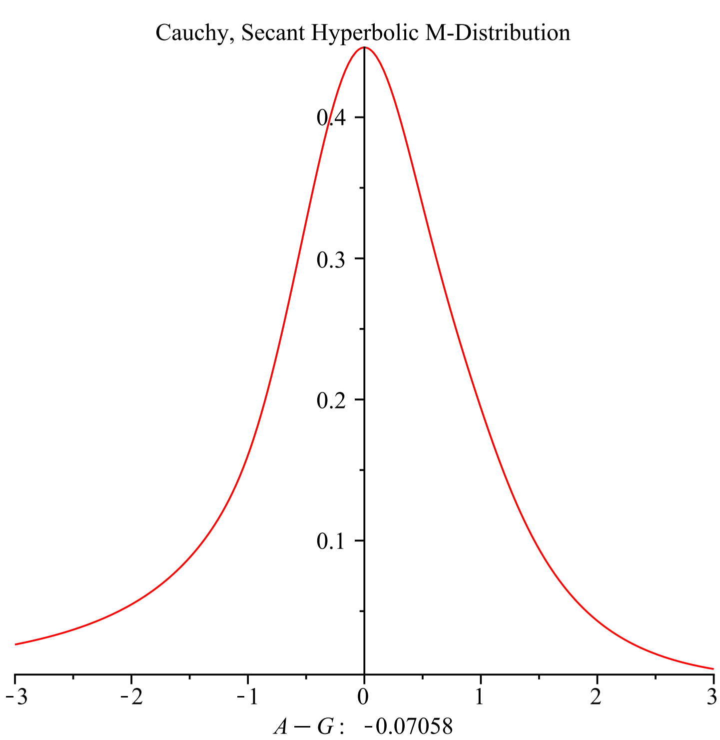
Plots of the densities resulting from the above M-distributions are given as follows. The A-G skewness is included for comparison purposes.

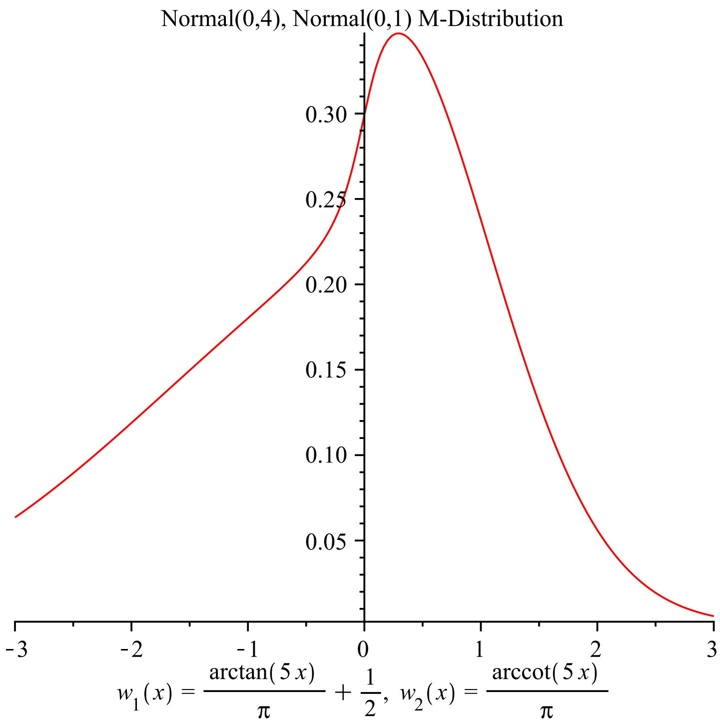
If we consider a location-scale model of the form $f((x-\mu)/\sigma)/\sigma$, then maximum likelihood methods apply. An alternative approach is to use Tiku's (Tiku (1967, 1968), Tiku and Suresh (1992)) modified maximum likelihood, which expresses the


Normal(0,4), Normal(0,1) M-Distribution
0.25
0.20
0.15
0.10
0.05
-3
-2
-1
0
1
2
3
A − G: -0.1276


Secant Hyperbolic, Normal(0,1) M-Distribution
0.4
0.3
0.2
0.1
-3
-2
-1
0
1
2
3
A − G: 0.03117


Cauchy, Normal(0,1) M-Distribution
0.35
0.30
0.25
0.20
0.15
0.10
0.05
-3
-2
-1
0
1
2
3
A − G : -0.03941


Cauchy, Secant Hyperbolic M-Distribution
0.4
0.3
0.2
0.1
-3
-2
-1
0
1
2
3
A − G : -0.07058



likelihood equations in terms of order statistics, and then essentially linearizes non-linear terms in the likelihood equations using Taylor expansions around the expected values of the order statistics. It has been shown that using percentage points in place of the expected values of order statistics produces highly accurate and asymptotically fully efficient estimates. Percentage points can be evaluated by Newton-Raphson or other approximation methods.

In general, it should be noted that the size of the data set for which good results can be expected must be reasonably large. In the case of financial data, that is very often the case, as it is in experiments involving gene expression.

Another useful extension is to start with symmetric density g with variance 1 and then create the M-distribution through mixing two versions of scale-weighted forms of g as follows:

$$\frac{2\sigma}{\sigma^2+1}(w_1(x)g(x/\sigma)+w_2(x)g(\sigma x))$$

This is a similar construction to that of Fernandez and Steel (1998) model, but smooth throughout so maximum likelihood can be applied. By choosing different mixing constants (the value a in (1) and (2) or a and b in (4) and (5)) in the swps. This allows us to incorporate Arnold-Groeneveld values closer to 1 or -1.

4. Moments and Percentage Points

Some facts:

M-distributions have the same number of moments as the minimum for g_1, g_2.

If each has a moment generating function, then so will the M-distribution.

Half moments $M_{1/2}(n) = \int_0^\infty g(x)x^n \, dx$ and similar are useful in characterizing M-distributions.

Weighted "moments" are defined to be $M_w(n) = \int_{-\infty}^\infty w(x)g(x)x^n \, dx$, and are often relatively easy to compute.

Weighted half moments are defined by $M_{1/2,w}(n) = \int_0^\infty w(x)g(x)x^n \, dx$ and similar and often are straightforward to compute.

In the case of swps in M-distributions, $M_w(2k) = M_{1/2}(2k)$, and $M_w(2k+1) = 2M_{1/2,w}(2k+1) - M_{1/2}(2k+1)$.

Determining t_α such that $\alpha = g(t_\alpha)$, $0 < \alpha < 1$, by Newton-Raphson, requires only a few iterations of the approximation formula.

5. Alternatives

A number of alternative methods for introducing skewness exist in the literature. For example, Marshall and Olkin (1997) introduce a new parameter to generate skew distributions from symmetric ones. Fernandez and Steel (1998) develop a skewing mechanism that is based on joining two distributions of the same form but with different scale parameters by using the indicator function. Jones (2004) uses the beta distribution and transformations to create skew distributions from symmetric ones. Ferreira and Steel (2006) give an overview of multiplicative skewing mechanisms and bring results under one general framework.

In general, none of these methods allow for tails that have different limiting behaviour. The tails may have different power laws, such as $\mid x \mid^{-k}$ and $\mid x \mid^{-j}$, but not exponential decay in one tail, and a power law in the other. The simplicity of the M-distribution formulation is not equaled in these other structures, and the level of complexity makes them difficult to use and explore. Further analysis of the properties of M-distributions should generate a family that appeals to practitioners.

6. Conclusions

The purpose of this paper is to introduce a family of skew distributions that can have different tail behaviours, and with various extensions including location-scale models. These are explicit functions that have reasonable analytic properties which allow for estimates of location and scale parameters through maximum likelihood. As such, they have the potential to be used as models for financial data (which often exhibit both skewness and high levels of kurtosis), as well as data from other scientific investigations.

This is a highly flexible family for which standard values, including moments and percentage points, are relatively easy to calculate. Because of the explicit nature of the density and the continuity and differentiability properties, maximum likelihood can be applied. The explicit form of the cdf makes calculating percentage points easy to evaluate, so that modified maximum likelihood (see, e.g., Tiku (1967, 1968) and Tiku and Suresh (1992)) can be applied. Because of these properties, it is a viable alternative to current skewing mechanisms applied to standard symmetric distributions.

References

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring Skewness with Respect to the Mode. *The American Statistician*, **49**, (1), pp. 34-38.

Blattberg, R. C. and Gonedes, N. J.(1974). A comparison of the stable and Student distributions as statistical models for stock prices. *J. Business* **47**, 244-280.

Corrado, C. and Su, T. (1996). Skewness and kurtosis in S&P 500 index returns implied by option prices. *J. Financial Research*, **19**, 175-192.

Currie,L. A. (2001). Some case studies of skewed (and other abnormal) data distributions arising in low-level environmental research. *Fresenius J. Anal. Chem.*, **370**, 705-718.

Ferreira, T. A. and Steel, M. F. j. (2006). A constructive representation of univariate skewed distributions. *JASA*, **101**, 823-828.

Fischer, M. and Vaughan, D. C. (2002): Classes of Skew Generalized Hyperbolic Secant Distributions. Discussion paper: Department of Statistics and Econometrics, University of Erlangen-Nurnberg

Fischer, M. and Vaughan, D. C. (2010). The beta-hyperbolic secant distribution. *Austrian J. Stats.*, **39**, 245-258.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series and Products, Seventh Edition.* Academic Press, New York (Alan Jeffrey and Daniel Zwillinger, Editors).

Hardin, J. and Wilson, J. (2009). A note on oligonucleotide expression values not being normally distributed. *Biostatistics*, **10**, 446-450.

Harris, R. D. F.; Küçüközmen, C. C. and Yilmaz, F. (2004). Skewness in the conditional distribution of daily equity returns. *Applied Financial Economics* **14**, 195-202.

Johnson, N. L.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2, Second Edition.* John Wiley and Sons: New York.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test* **13**, 1-43.

Macdonald, P.D.M. and T.J. Pitcher (1979). Age-groups from size-frequency data: a versatile and efficient method of analysing distribution mixtures. Journal of the Fisheries Research Board of Canada **36**, 987-1001.

Macdonald, P.D.M. (1987). Analysis of length-frequency distributions. In R.C. Summerfelt and G.E. Hall [editors], Age and Growth of Fish, Iowa State University Press, Ames, Iowa, 371-384.

Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika 84: 641652.

Malik, H.J. (1967). Exact distribution of the quotient of independent generalized gamma variables. *Canadian Mathematical Bulletin* **10**, 463-465.

McLeay, S. (1986). Student's t and the distribution of financial ratios. *J. Bus. Fin. Accounting* **13**, 209-222

Praetz, P. D. (1972). The distribution of share price changes. *J. of Business* **45**, 49-55.

Tiku, M. L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika* 54, 155-165.

Tiku, M. L. (1968). Estimating the parameters of normal and logistic distributions from censored samples. *Austral. J. Statist.* 10, 64-74.

Tiku, M. L. and Suresh, R.P. (1992). A new method of estimation for location and scale parameters. *J. Stat. Plan. Inf.* 30, 281-292.

Vaughan, D. C. (2002). The Generalized Secant Hyperbolic distribution and its properties. *Commun. Stat. - Theory Meth.* **31**, 219-238.