On the sum of t and Gaussian random variables.

Guy P. Nason*

August 8, 2005

Abstract

This article derives the probability density function (pdf) of the sum of a normal random variable and a (sphered) Student's-t distribution on odd degrees of freedom greater than or equal to three. Apart from its intrinsic interest applications of this result include Bayesian wavelet shrinkage, Bayesian posterior density derivations, calculations in the theoretical analysis of projection indices and computation of certain moments.

Some key words: sum of Gaussian and Student's t, characteristic function, wavelet shrinkage, error function.

1 Introduction

Let X be a Gaussian random variable with mean μ, variance σ^2 and density $\phi_{\mu,\sigma}(x)$. Let T_ν be a random variable distributed according to Student's t-distribution on ν degrees of freedom. Both Gaussian and Student's t-distributions are amongst the most important distributions in statistics. In many situations it would be very useful to know the density function of their sum $Y = X + T_\nu$ which can be represented as the convolution of the density functions of X and T_ν as follows:

$$f_Y(y) = \int_{-\infty}^{\infty} \phi_{\mu,\sigma}(y-x)\, t_\nu(x)\, dx = \int_{-\infty}^{\infty} \phi_{\mu+y,\sigma}(x) t_\nu(x)\, dx =< \phi_{\mu+y,\sigma}, t_\nu >, \quad (1)$$

where $<,>$ is the usual inner product $< f, g >= \int_{-\infty}^{\infty} f(x)g(x)dx$. When $\nu = 1$ the density of Y is known from the result of Kendall (1938):

$$< \phi_{\mu,\sigma}(x), t_1(x) >= \sqrt{2}\, \mathrm{Re} \left\{ e^{z^2} \operatorname{erfc}(z) \right\} /(\pi\sigma), \quad (2)$$

where $z = d - ip/2$, $d = 1/(\sqrt{2}\sigma)$ and $p = \sqrt{2}\mu/\sigma$. Here erfc is the complementary error function (see Lebedev, 1965; Abramowitz and Stegun, 1972) and Re means 'take the real part'.

*Department of Mathematics, University Walk, University of Bristol, Bristol, BS8 1TW, England E-mail: G.P.Nason@bristol.ac.uk

Like Kendall (1938) we make use of Parseval's relation to initiate our derivation of the density of $X + T_\nu$ for $\nu = 3$ in Theorem 1 and for general odd ν in Theorem 2. In all that follows T_ν is (slightly) redefined to be the following sphered Student's t distribution: this is just the usual Student's t but scaled to have unit variance.

Definition 1 (Sphered Student's t) *The* sphered Student's t-density *on $\nu \geq 3$ degrees of freedom is defined by $t_\nu : \mathbb{R} \to (0, \infty)$ such that*

$$t_\nu(x) = \pi^{-1/2}(\nu-2)^{-1/2}\frac{\Gamma\left\{\frac{1}{2}(\nu+1)\right\}}{\Gamma\left\{\frac{1}{2}\nu\right\}}\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2}. \tag{3}$$

A multivariate version of this sphered distribution appears in Nason (2001).

Before presenting our Theorems the following list sets out some reasons why the inner product (1) is important.

1. It provides an explicit formula for the density function of the sum of two of the most important random variables in statistics!

2. In Bayesian wavelet shrinkage Johnstone and Silverman (2004, 2005) show that excellent performance is obtained by using heavy-tailed distributions as part of a wavelet coefficient mixture prior instead of the standard normal. A quantity of interest is the convolution of the heavy-tailed distribution with the standard normal. It has been mooted that Student's t-distribution might also be an interesting distribution to use in this context and again the convolution in (1) is useful for deriving the posterior mean based on a Student's t mixture prior.

3. O'Hagan and Forster (2004, 11.2) describe the Bayesian linear model

   $$\mathbf{y} = \mathbf{X}\beta + \epsilon,$$

   where $\epsilon \sim N(0, \sigma^2 I)$ and hence the likelihood is $f(y|\beta, \sigma^2) \sim N(\beta X, \sigma^2 I)$. In Section (11.67–8) they describe Student's t distribution as a useful family of heavy-tailed *prior* distributions for the β parameter. For example, $\beta \sim t_\nu(\mathbf{m}, \mathbf{W}) = t(\beta)$ and with an independent prior, f, for σ^2 their joint prior for (β, σ^2) is $f(\beta, \sigma^2) \propto t(\beta) f(\sigma^2)$. O'Hagan and Forster (2004) describe the posterior distribution $f(\beta, \sigma^2|\mathbf{y})$ as intractable since β cannot be integrated out analytically. However, our work shows that the posterior distribution *can* now be computed exactly (in one dimension, for odd degrees of freedom) since the marginal likelihood is given by $f(y) = \int_{-\infty}^{\infty} \phi_{y,\sigma^2}(\beta) t_3(\beta) d\beta$, which is just the convolution in (1). It should be pointed out that σ^2 is generally unknown in practice and further calculations and consideration would be required, especially for the more realistic multivariate case.

4. In exploratory projection pursuit the L^2 distance is sometimes used as a projection index. For example, Hall (1989), defined a projection index that measured the dissimilarity between a (projected) density $f(x)$ and a reference standard normal density $\phi(x)$ by $L(f, \phi) = \int_{-\infty}^{\infty} \{f(x) - \phi(x)\}^2 \, dx$. One of the design goals for projection indices is to ensure that they are not unduly influenced by "pseudo-outliers" and produce projection solutions which reflect true clustering in a data set rather than one large cluster and an outlier. Some suggestions have been made about how to robustify indices, e.g. Friedman (1987) and Cook *et al.* (1993). An alternative suggestion in Nason (2001) was to replace the reference normal distribution in $L(f, \phi)$ by Student's t or the Laplace distribution. The experimental setup in Nason (2001) required analytical evaluation of $L(\phi_{\mu,\sigma}, t_3)$ which itself requires evaluation of our desired convolution (1).

5. The inner product of Student's t on ν d.f. and the $\phi_{\mu,\sigma}$ density can been viewed as $\mathbb{E}(1 + X^2)^{-(\nu+1)/2}$ where $X \sim N(\mu, \sigma^2)$. (Likewise $\mathbb{E}\exp(-X^2/2)$ for $X \sim t_3$). Although maybe not of major current interest these expectations are similar to the negative moments of a distribution, see Cressie *et al.* (1981).

6. There is a great deal of current interest in convolutions of random variables, especially those convolutions between variables from different families. For example, see Nadarajah and Kotz (2005) and references therein. For a more comprehensive treatment of the Student's t-distribution in general see Kotz and Nadarajah (2004).

Note: A small R software package, called `NORMT3` has been written to compute the convolution density for $\nu = 3$. See the R CRAN archive for this package and associated documentation.

The next section establishes the distribution of $X + T_3$ and the more general result for $X + T_\nu$ for odd ν. Section 3 concludes.

2 Inner Product of Student's t and Gaussian densities

We first derive our inner product formula for $t_\nu(x)$ with $\phi_{\mu,\sigma}$ for $\nu = 3$. The $t_3(x)$ is the most heavy tailed Student's t that possesses a mean and variance and is thus of most use for designing robust projection indices out of all the $t_\nu(x)$ for $\nu \geq 3$, see Nason (2001). We later discuss the situation for more general ν which is more complicated and relies on the characteristic function (c.f.) of t_ν for odd $\nu > 3$.

Theorem 1 *The inner product of $t_3(x)$ and $\phi_{\mu,\sigma}(x)$ is given by*

$$\begin{aligned} < \phi_{\mu,\sigma}(x), t_3(x) > &= \frac{b}{\pi}\exp(a/2)\left[\left\{(1-a)\cos(\mu a) + \frac{pb}{2}\sin(\mu a)\right\} I_{C1}(p,d)\right. \\ &+ \left\{(1-a)\sin(\mu a) - \frac{pb}{2}\cos(\mu a)\right\} I_{S1}(p,d) \\ &+ \left. be^{-d^2}/2\right], \end{aligned} \tag{4}$$

where $a = \sigma^{-2}$, $b = \sqrt{2}/\sigma$, $d = a/b$, $p = \mu b$,

$$I_{C1}(p,d) = \int_d^\infty \cos(px)e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}e^{-p^2/4}\left[1 - \operatorname{Re}\left\{\operatorname{erf}(c) - \operatorname{erf}(c-d)\right\}\right], \tag{5}$$

and

$$I_{S1}(p,d) = \int_d^\infty \sin(px)e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}e^{-p^2/4}\operatorname{Im}\left\{\operatorname{erf}(c-d)\right\}, \tag{6}$$

where $c = ip/2$, $i = \sqrt{-1}$, Re *and* Im *extract the real and imaginary parts and* $\operatorname{erf}(z) = \int_0^z \exp(-t^2)\,dt$ *is the error function which coincides with $\sqrt{2\pi}\{\Phi(x) - 1/2\}$ for real $x > 0$.*

The proof of this theorem appears in the appendix.

Theorem 1 only deals with the case $\nu = 3$. The proof relies on the characteristic function (c.f.) of the Student's t distribution shown by Stuart and Ord (1994, Ex. 3.13) for odd d.f., $\nu = 2m - 1$, to be given by

$$\chi_{T_\nu}(t) = \frac{k_m \pi}{2^{2m-2}(m-1)!}e^{-|t|}\sum_{j=0}^{m-1}(2|t|)^{m-1-j}(m-1+j)^{[2j]}/j! \tag{7}$$

where $k_m = \Gamma(m)/\{\Gamma(\frac{1}{2})\Gamma(m-\frac{1}{2})\}$ and $x^{[r]} = x!/(x-r)!$ The c.f. of the *regular* Student's t-distribution can be obtained from (7) and is equal to $\chi_{T_\nu}\left(\{(\nu-2)/\nu\}^{1/2}\,t\right)$. Note that χ_{T_ν} can be written as $K_\nu P_m(|t|)e^{-|t|}$ where K_ν is a constant and $P_m(z)$ is a polynomial of order m.

For more general odd d.f. ν, Theorem 2 makes use of the c.f. in (7). The higher order terms of $P_m(z)$ in (7) require the evaluation of quantities such as

$$I_{Cr}(p,d) = \int_d^\infty \cos(pz)z^{r-1}\exp(-z^2)\,dz, \tag{8}$$

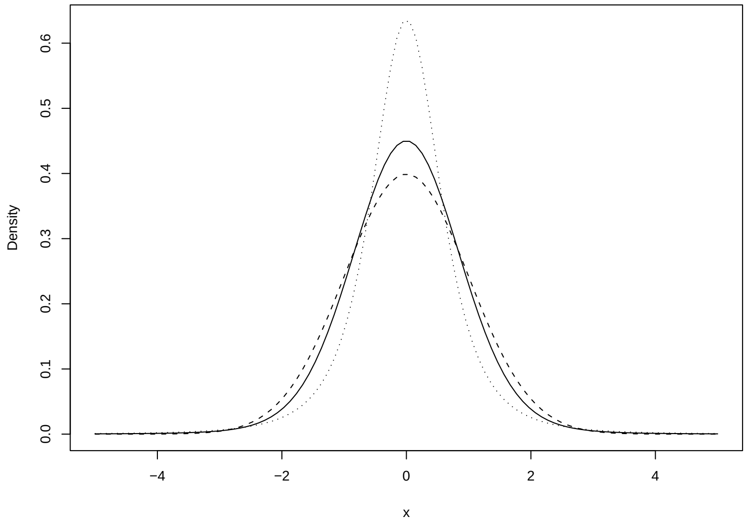

Figure 1: Central part of p.d.f. of $Y = X + T_3$ (solid line); standard normal density for X (dashed line); sphered Student's t distribution on 3 degrees of freedom for T_3 (dotted line).

for $p, d \in \mathbb{R}$, integer $r \geq 1$ and similarly $I_{Sr}(p, d)$ for sin. These integrals can be evaluated recursively although the combination of the recursion and their linear combination specified by the c.f. formula (14) does not result in a particularly neat formula.

Theorem 2 *The inner product of $t_\nu(x)$ and $\phi_{\mu,\sigma}(x)$ is given by*

$$< \phi_{\mu,\sigma}(x), t_\nu(x) >= \frac{bK_\nu}{\pi} \exp(a/2) \sum_{r=0}^{m} q_r \left\{ \cos(\mu a) I_{C_{r+1}}(p, d) + \sin(\mu a) I_{S_{r+1}}(p, d) \right\}, \quad (9)$$

where $m = (\nu + 1)/2$, a, b, p, d are as defined in Theorem 1. The coefficients $\{q_r\}_{r=0}^m$ are coefficients of the polynomial $Q_m(z) = P_m(bz - a)$.

The proof of this theorem appears in the appendix.

Figures 1 and 2 show the central and tail parts of the density functions respectively and compares them to the standard normal distribution and $t_3(x)$. Note in the tail that f_Y has heavier tails than ϕ but lighter tails than t_3 as one would expect.

3 Conclusions

This article provides an explicit analytical formula for the pdf of the sum, Y, of a standard normal random variable and a (sphered) Student's t variable on three degrees of freedom. We also provide a general formula for higher degrees of freedom. The result also has

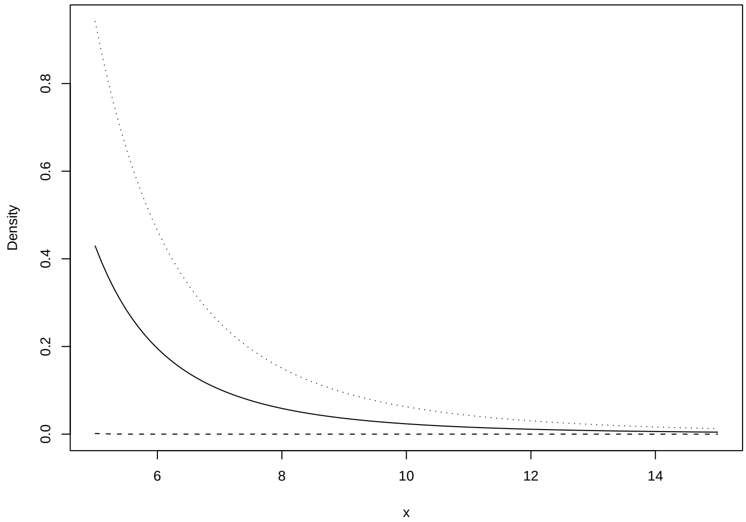

Figure 2: Tail part of p.d.f. of $Y = X + T_3$ (solid line); standard normal density for X (dashed line); sphered Student's t distribution on 3 degrees of freedom for T_3 (dotted line)

direct bearing on computations in theoretical studies of exploratory projection indices, calculations involving convolutions of ϕ with t_3 in Bayesian wavelet shrinkage, posterior density computations in Bayesian statistics and possibly certain moment calculations.

Acknowledgments

The author would like to express thanks to Nicolas Chopin, Idris Eckley, Natalya Schofield and Bernard Silverman and a referee for several helpful comments. The author gratefully acknowledges support from EPSRC Grant GR/M10229 and an EPSRC Advanced Research Fellowship.

A Proofs

Lemma 1 *Define $I_{Cr}(p, d)$ and $I_{Sr}(p, d)$ as in (8) for $p, d \in \mathbb{R}$.*

Then

$$I_{C_{r+1}} = -pI_{S_r}(p, d)/2 + (r-1)I_{C_{r-1}}(p, d)/2 + d^{r-1}e^{-d^2}\cos(pd)/2, \tag{10}$$

and

$$I_{S_{r+1}} = pI_{C_r}(p, d)/2 - (r-1)I_{S_{r-1}}(p, d)/2 - d^{r-1}e^{-d^2}\sin(pd)/2, \tag{11}$$

for $r \geq 1$. *In particular,*

$$I_{S2}(p,d) = \frac{p}{2}\left\{I_{C1}(p,d) + \sin(pd)e^{-d^2}/p\right\} \text{ and } I_{C2}(p,d) = \frac{p}{2}\left\{\cos(pd)e^{-d^2}/p - I_{S1}(p,d)\right\}. \quad (12)$$

Proof of Lemma 1 We prove the formula for $I_{C_{r+1}}$, that for $I_{S_{r+1}}$ is similar. The proof is a simple integration by parts: for $p \neq 0$

$$\begin{aligned} I_{S_r}(p,d) &= \int_d^\infty \sin(pz) z^{r-1} e^{-z^2}\, dz \\ &= \left[z^{r-1}e^{-z^2}\cos(pz)/p\right]_\infty^d + \int_d^\infty \left\{(r-1)z^{r-2}e^{-z^2} - 2z^r e^{-z^2}\right\}\cos(pz)/p\, dz \\ &= p^{-1}d^{r-1}e^{-d^2}\cos(pd) + p^{-1}(r-1)I_{C_{r-1}} - 2p^{-1}I_{C_{r+1}} \end{aligned}$$

Rearrangement gives the formula in (10). It is easy to check that formulae (10)–(12) are valid for $p = 0$.

□

Proof of Theorem 1 First we use the definition of the Fourier transform of a function $f(x)$ that reads $\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\, dx$. Then the inner product of $t_3(x)$ with $\phi_{\mu,\sigma}(x)$ can be expressed in the Fourier domain by

$$\begin{aligned} < \phi_{\mu,\sigma}, t_3 > &= (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-i\omega\mu}\hat{\phi}(\sigma\omega)\hat{t}_3(-\omega)\, d\omega \\ &= \pi^{-1}\int_0^\infty e^{-\omega^2\sigma^2/2-\omega}(1+\omega)\cos(\mu\omega)\, d\omega, \end{aligned} \quad (13)$$

by Parseval's relation and using that the Fourier transform of $t_3(x)$ can be obtained from the formula for the c.f. of Student's t for odd d.f. in Example 3.13 from Stuart and Ord (1994):

$$\hat{t}_3(\omega) = (1 + |\omega|)e^{-|\omega|} \quad (14)$$

for $\omega \in \mathbb{R}$. Now complete the square in the exponential in (13) by letting $a = \sigma^{-2}$ and $b = \sqrt{2}/\sigma$ to obtain

$$< \phi_{\mu,\sigma}, t_3 >= \pi^{-1}\int_0^\infty \exp\left\{-\left(\frac{\omega + a}{b}\right)^2 + a/2\right\}(1+\omega)\cos(\mu\omega)\, d\omega.$$

Now substitute $z = (w + a)/b$ and let $d = a/b$ to obtain

$$
\begin{aligned}
< \phi_{\mu,\sigma}, t_3 > &= b\pi^{-1}\exp(a/2)\int_d^\infty \cos\{\mu(bz-a)\}e^{-z^2}(1-a+bz)\,dz \\
&= K\left[(1-a)\int_d^\infty \cos\{\mu(bz-a)\}e^{-z^2}\,dz\right. \\
&+ \left. b\int_d^\infty \cos\{\mu(bz-a)\}ze^{-z^2}\,dz\right], \qquad (15)
\end{aligned}
$$

where $K = b\pi^{-1}\exp(a/2)$. Now letting $p = \mu b$, (15) becomes

$$
\begin{aligned}
< \phi_{\mu,\sigma}, t_3 > &= K\left[(1-a)\left\{\cos(\mu a)I_{C1}(p,d) + \sin(\mu a)I_{S1}(p,d)\right\}\right. \\
&+ \left. b\left\{\cos(\mu a)I_{C2}(p,d) + \sin(\mu a)I_{S2}(p,d)\right\}\right], \qquad (16)
\end{aligned}
$$

where I_{C2} is defined as in (8) and similarly for I_{S2}. Substituting (12) from Lemma 1 into (16) yields the main formula given in (4).

For the integral formulae in (5) and (6) note first that formulae (7.4.6) and (7.4.7) in Abramowitz and Stegun (1972) show that

$$
\int_0^\infty \cos(pz)e^{-z^2}\,dz = \frac{\sqrt{\pi}}{2}e^{-p^2/4}, \text{ and } \int_0^\infty \sin(pz)e^{-z^2}\,dz = e^{-p^2/4}\int_0^{p/2} e^{t^2}\,dt. \qquad (17)
$$

The formulae we need in (5) and (6) are $\int_d^\infty$ whereas those in (17) are $\int_0^\infty$ so we need to find

$$
I_C(p,d) = \int_0^d \cos(pz)e^{-z^2}\,dz \text{ and } I_S(p,d) = \int_0^d \sin(pz)e^{-z^2}\,dz.
$$

To determine I_C and I_S define

$$
I(p,d) = \int_0^d e^{ipz}e^{-z^2}\,dz = I_C(p,d) + iI_S(p,d). \qquad (18)
$$

Let $c = ip/2$, completing the square in (18) and substituting $z \to -z$ we obtain

$$
I(p,d) = e^{-p^2/4}\int_{-d}^0 e^{-(z+c)^2}\,dz = e^{-p^2/4}\int_{\gamma_1} e^{-w^2}\,dw, \qquad (19)
$$

where γ_1 is the line in the complex plane $\gamma_1 : [0,-d] \to \mathbb{C}$ defined by $\gamma_1(t) = c + t$ for $t \in [0,-d]$. Let γ_2 and γ_3 be the lines defined by the vectors $(0,0) \to (-d,c)$ and $(0,0) \to (0,c)$ respectively. Then since e^{-w^2} is holomorphic on $\mathbb{C}$ we can use Cauchy's theorem to show

$$I(p,d) = e^{-p^2/4}\left(-\int_{\gamma_2} e^{-w^2}\,dw + \int_{\gamma_3} e^{-w^2}\,dw\right).$$

By definition of erf this means that

$$I(p,d) = \frac{\sqrt{\pi}}{2} e^{-p^2/4}\left\{\mathrm{erf}(c) - \mathrm{erf}(c-d)\right\}.$$

Taking real and imaginary parts yields I_C and I_S and hence the formulae in (5) and (6).

□

Proof of Theorem 2 This proof follows closely the proof of Theorem 1. Inserting the $K_\nu P_m(|\omega|)e^{-|w|}$ formula for the c.f. of t_ν yields the following general form of equation (13):

$$< \phi_{\mu,\sigma}, t_\nu >= \pi^{-1} K_\nu \int_0^\infty e^{-\omega^2\sigma^2/2-\omega} \cos(\mu\omega) P_m(\omega)\,d\omega. \tag{20}$$

Then making the same change of variable $\omega = bz - a$ as in Theorem 1 yields the generalization of (15)

$$< \phi_{\mu,\sigma}, t_\nu >= b\pi^{-1} K_\nu \exp(a/2) \int_d^\infty \cos\left\{\mu(bz-a)\right\} e^{-z^2} P_m(bz-a)\,dz. \tag{21}$$

Then using the trig formula for $\cos(A-B)$ on $\cos\{\mu(bz-a)\}$, writing $Q_m(z) = \sum_{r=0}^m q_r z^r = P_m(bz-a)$ and using the definition of I_{S_r} and I_{C_r} yields the result (9).

□

References

Abramowitz, M. and Stegun, I. (1972) *Handbook of mathematical functions.* New York: Dover.

Cook, D., Buja, A. and Cabrera, J. (1993) Projection pursuit indices based on expansions with orthonormal functions. *J. Comput. Graph. Statist.*, **2**, 225–250.

Cressie, N., Davis, A., Folks, D. and Policello, G. (1981) The moment-generating function and negative integer moments. *The American Statistican*, **35**, 148–150.

Friedman, J. (1987) Exploratory projection pursuit. *J. Am. Statist. Ass.*, **82**, 249–266.

Hall, P. (1989) On polynomial-based projection indices for exploratory projection pursuit. *Ann. Statist.*, **17**.

Johnstone, I. and Silverman, B. (2004) Needles and hay in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, **32**, 1594–1649.

— (2005) Empirical Bayes selection of wavelet thresholds. *Ann. Statist.*, **33**, 1700–1752.

Kendall, D. (1938) Effect of radiation damping and doppler broadening on the atomic absorption coefficient. *Zeitschrift für Astrophysik*, **16**, 308–317.

Kotz, S. and Nadarajah, S. (2004) *Multivariate t Distributions and their Applications.* Cambridge: Cambridge University Press.

Lebedev, N. (1965) *Special functions and their applications.* Englewood Cliffs, N.J.: Prentice-Hall.

Nadarajah, S. and Kotz, S. (2005) On the linear combination of exponential and gamma random variables. *Entropy*, **7**, 161–171.

Nason, G. (2001) Robust projection indices. *J. R. Statist. Soc. B*, **63**, 551–567.

O'Hagan, A. and Forster, J. (2004) *Bayesian Inference.* London: Arnold.

Stuart, A. and Ord, J. (1994) *Kendall's Advanced Theory of Statistics: Distribution Theory*, vol. 1. London: Arnold.