

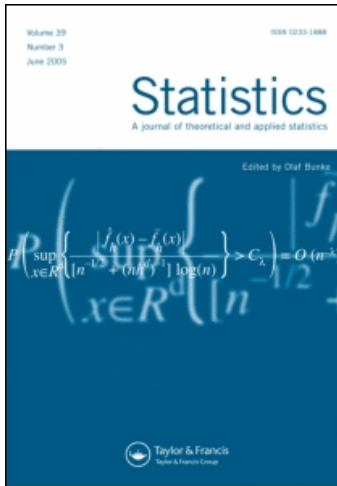
This article was downloaded by: [Yanev, George]

On: 12 February 2011

Access details: Access Details: [subscription number 933399554]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Statistics

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713682269>

Characterizations of Student's t -distribution via regressions of order statistics

George P. Yanev^a; M. Ahsanullah^b

^a Mathematics Department, The University of Texas - Pan American, Edinburg, TX, USA ^b

Management Sciences Department, Rider University, Lawrenceville, NJ, USA

First published on: 12 February 2011

To cite this Article Yanev, George P. and Ahsanullah, M.(2011) 'Characterizations of Student's t -distribution via regressions of order statistics', *Statistics*, First published on: 12 February 2011 (iFirst)

To link to this Article: DOI: 10.1080/02331888.2010.535904

URL: <http://dx.doi.org/10.1080/02331888.2010.535904>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Characterizations of Student’s t -distribution via regressions of order statistics

George P. Yanev^{a*} and M. Ahsanullah^b

^aMathematics Department, The University of Texas – Pan American, Edinburg, TX, USA;

^bManagement Sciences Department, Rider University, Lawrenceville, NJ, USA

(Received 12 May 2010; final version received 19 October 2010)

Utilizing regression properties of order statistics, we characterize a family of distributions introduced by Akhundov *et al.* [New characterizations by properties of midrange and related statistics, *Commun. Stat. Theory Methods* 33(12) (2004), pp. 3133–3143], which includes the t -distribution with two degrees of freedom as one of its members. Then we extend this characterization result to t -distribution with more than two degrees of freedom.

Keywords: order statistics; characterizations; t -distribution; regression

1. Discussion of the results

The Student’s t -distribution is widely used in statistical inferences when the population standard deviation is unknown and is substituted by its estimate from the sample. Recently, Student’s t -distribution was also considered in financial modelling by Ferguson and Platen [1] and as a pedagogical tool by Jones [2]. The probability density function (pdf) of the t -distribution with ν degrees of freedom (t_ν -distribution) is given for $-\infty < x < \infty$ and $\nu = 1, 2, \dots$ by

$$f_\nu(x) = c_\nu \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \quad \text{where} \quad c_\nu = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \quad (1)$$

and $\Gamma(x)$ is the gamma function.

The vast majority of characterization results for univariate continuous distributions based on ordered random variables is concentrated to exponential and uniform families. It was not until recently, within the last seven to eight years, that some characterizations were obtained for t_ν -distribution with $\nu = 2$ and $\nu = 3$. In this note, we communicate generalizations of these recent results for t_ν -distribution when $\nu \geq 2$. Let X, X_1, X_2, \dots, X_n for $n \geq 3$ be independent random variables with common cumulative distribution function (cdf) $F(x)$. Assume that $F(x)$ is absolute continuous with respect to the Lebesgue measure. Let $X_{1:3} \leq X_{2:3} \leq \dots \leq X_{n:n}$ be the

*Corresponding author. Email: yanevgp@ytpa.edu

corresponding order statistics. Nevzorov *et al.* [3] (see also [4]) and Akhundov and Nevzorov [5] proved characterizations for the t_ν -distribution when $\nu = 2$ and $\nu = 3$, respectively, assuming, in addition, $n = 3$. Here we extend these results to the general case of any $\nu \geq 2$ and any $n \geq 3$.

Let $Q(x)$ be the quantile function of a random variable with cdf $F(x)$, i.e. $F(Q(x)) = x$ for $0 < x < 1$. Akhundov *et al.* [6] proved that for $0 < \lambda < 1$, the relation

$$E[\lambda X_{1:3} + (1 - \lambda)X_{3:3} \mid X_{2:3} = x] = x \quad (2)$$

characterizes a family of probability distributions with quantile function

$$Q_\lambda(x) = \frac{c(x - \lambda)}{\lambda(1 - \lambda)(1 - x)^\lambda x^{1-\lambda}} + d, \quad 0 < x < 1, \quad (3)$$

where $0 < c < \infty$ and $-\infty < d < \infty$. Let us call this family of distributions – Q -family.

THEOREM 1 (Q -family) *Assume that $E|X| < \infty$ and $n \geq 3$ is a positive integer. The random variable X belongs to the Q -family if and only if for some $2 \leq k \leq n - 1$ and some $0 < \lambda < 1$*

$$\begin{aligned} & \lambda E \left[\frac{1}{k-1} \sum_{i=1}^{k-1} (X_{k:n} - X_{i:n}) \mid X_{k:n} = x \right] \\ & = (1 - \lambda) E \left[\frac{1}{n-k} \sum_{j=k+1}^n (X_{j:n} - X_{k:n}) \mid X_{k:n} = x \right]. \end{aligned} \quad (4)$$

Note that Equation (4) can be written as

$$\lambda E \left[\frac{1}{k-1} \sum_{j=1}^{k-1} X_{j:n} \mid X_{k:n} = x \right] + (1 - \lambda) E \left[\frac{1}{n-k} \sum_{j=k+1}^n X_{j:n} \mid X_{k:n} = x \right] = x. \quad (5)$$

Clearly for $n = 3$ and $k = 2$ Equation (5) reduces to Equation (2). It is also worth mentioning here that, as Balakrishnan and Akhundov [7] reported, the Q -family, for different values of λ , approximates well a number of common distributions including Tukey lambda, Cauchy, and Gumbel (for maxima).

Notice that t_2 -distribution belongs to the Q -family, having quantile function [2]

$$Q_{1/2}(x) = \frac{2^{1/2}(x - 1/2)}{x^{1/2}(1 - x)^{1/2}}, \quad 0 < x < 1.$$

Nevzorov *et al.* [3] proved that if $E|X| < \infty$ then X follows t_2 -distribution if and only if

$$E[X_{2:3} - X_{1:3} \mid X_{2:3} = x] = E[X_{3:3} - X_{2:3} \mid X_{2:3} = x]. \quad (6)$$

This also follows directly from Equation (2) with $\lambda = 1/2$. Recall that the cdf of t_2 -distribution [2] is

$$F_2(x) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{1+x^2}} \right).$$

Setting $\lambda = 1/2$ in Equation (4), we obtain the following corollary of Theorem 1.

COROLLARY 1 (t_2 -distribution) Assume that $E|X| < \infty$ and $n \geq 3$ is a positive integer. Then

$$F(x) = F_2\left(\frac{x - \mu}{\sigma}\right) \text{ for } -\infty < \mu < \infty, \sigma > 0, \tag{7}$$

if and only if for some $2 \leq k \leq n - 1$

$$E\left[\frac{1}{k-1} \sum_{i=1}^{k-1} (X_{k:n} - X_{i:n}) \mid X_{k:n} = x\right] = E\left[\frac{1}{n-k} \sum_{j=k+1}^n (X_{j:n} - X_{k:n}) \mid X_{k:n} = x\right]. \tag{8}$$

Relation (8) can be interpreted as follows. Given the value of $X_{k:n}$, the average deviation from $X_{k:n}$ to the observations less than it equals the average deviation from the observations greater than $X_{k:n}$ to it.

Remark 1 (i) Notice that Equation (8) reduces to Equation (6) when $n = 3$ and $k = 2$. (ii) Let us set $n = 2r + 1$ and $k = r + 1$ for an integer $r \geq 1$. Let $M_{2r+1} = X_{r+1:2r+1}$ be the median of the sample $X_1, X_2, \dots, X_{2r+1}$. Then Equation (8) implies

$$E\left[\sum_{i=1}^r (M_{2r+1} - X_{i:2r+1}) \mid M_{2r+1} = x\right] = E\left[\sum_{j=r+2}^{2r+1} (X_{j:2r+1} - M_{2r+1}) \mid M_{2r+1} = x\right].$$

If, in addition, $\bar{X}_{2r+1} = \sum_{i=1}^{2r+1} X_i / (2r + 1)$ is the sample mean, then Equation (8) reduces to Nevzorov *et al.* [3] t_2 -distribution characterization relation

$$E[\bar{X}_{2r+1} \mid M_{2r+1} = x] = x.$$

Let us now turn to the case of t_ν -distribution with $\nu \geq 3$. Akhundov and Nevzorov [5] extended Equation (6) to a characterization of t_3 -distribution as follows. If $EX^2 < \infty$ then X follows t_3 -distribution if and only if

$$E[(X_{2:3} - X_{1:3})^2 \mid X_{2:3} = x] = E[(X_{3:3} - X_{2:3})^2 \mid X_{2:3} = x]. \tag{9}$$

We generalize this in two directions: (i) characterizing t_ν -distribution with $\nu \geq 3$ and (ii) considering a sample of size $n \geq 3$. The following result holds.

THEOREM 2 (t_ν -distribution) Assume $EX^2 < \infty$. Let $n \geq 3$ and $\nu \geq 3$ be positive integers. Then

$$F(x) = F_\nu\left(\frac{x - \mu}{\sigma}\right) \text{ for } -\infty < \mu < \infty, \sigma > 0, \tag{10}$$

where $F_\nu(x)$ is the t_ν -distribution cdf if and only if for some $2 \leq k \leq n - 1$

$$\begin{aligned} E\left[\frac{1}{k-1} \sum_{i=1}^{k-1} \left(\frac{\nu-1}{2} X_{k:n} - (\nu-2) X_{i:n}\right)^2 \mid X_{k:n} = x\right] \\ = E\left[\frac{1}{n-k} \sum_{j=k+1}^n \left((\nu-2) X_{j:n} - \frac{\nu-1}{2} X_{k:n}\right)^2 \mid X_{k:n} = x\right]. \end{aligned} \tag{11}$$

Remark 2 (i) Notice that if $n = 3$, $k = 2$, and $\nu = 3$, then Equation (11) reduces to Equation (9).
(ii) Let us set $\nu = 3$, $n = 2r + 1$ and $k = r + 1$ for an integer $r \geq 1$. If, as before, $M_{2r+1} = X_{r+1:2r+1}$ is the median of the sample $X_1, X_2, \dots, X_{2r+1}$, then Equation (11) implies the following equality between the sum of squares of the deviations from the sample median

$$E \left[\sum_{i=1}^r (M_{2r+1} - X_{i:2r+1})^2 \mid M_{2r+1} = x \right] = E \left[\sum_{j=r+2}^{2r+1} (X_{j:2r+1} - M_{2r+1})^2 \mid M_{2r+1} = x \right].$$

2. Proofs

To prove our results, we need the following two lemmas.

LEMMA 1 [7] *The cdf $F(x)$ of a random variable X with quintile function (3) is the only continuous cdf solution of the equation*

$$[F(x)]^{2-\lambda}[1-F(x)]^{1+\lambda} = cF'(x), \quad c > 0. \quad (12)$$

LEMMA 2 *Let $r \geq 1$ and $n \geq 2$ be integers. Then*

$$\begin{aligned} \frac{1}{k-1} \sum_{i=1}^{k-1} E[X_{i:n}^r \mid X_{k:n} = x] &= \frac{1}{F(x)} \int_{-\infty}^x t^r dF(t), \quad 2 \leq k \leq n; \\ \frac{1}{n-k} \sum_{j=k+1}^n E[X_{j:n}^r \mid X_{k:n} = x] &= \frac{1}{1-F(x)} \int_x^{\infty} t^r dF(t), \quad 1 \leq k \leq n-1. \end{aligned} \quad (13)$$

Proof Using the standard formulas for the conditional density of $X_{j:n}$ given $X_{k:n} = x$ ($j < k$) [8, Theorem 1.1.1], we obtain for $r \geq 1$

$$\begin{aligned} &\frac{1}{k-1} \sum_{j=1}^{k-1} E[X_{j:n}^r \mid X_{k:n} = x] \\ &= \frac{1}{(k-1)[F(x)]^{k-1}} \sum_{j=1}^{k-1} \binom{k-2}{j-1} \int_{-\infty}^x [F(t)]^{j-1} [F(x) - F(t)]^{k-1-j} t^r dF(t) \\ &= \frac{1}{[F(x)]^{k-1}} \sum_{i=0}^{k-2} \binom{k-2}{i} \int_{-\infty}^x [F(t)]^i [F(x) - F(t)]^{k-2-i} t^r dF(t) \\ &= \frac{1}{F(x)} \int_{-\infty}^x t^r dF(t). \end{aligned}$$

This verifies (13). The second relation in the lemma's statement can be proved similarly. ■

2.1. Proof of Theorem 1

First, we show that Equation (4) implies Equation (3). Applying Lemma 2, for the left-hand side of Equation (5), we obtain

$$\begin{aligned} & \frac{\lambda}{k-1} \sum_{j=1}^{k-1} E[X_{j:n} | X_{k:n} = x] + \frac{1-\lambda}{n-k} \sum_{j=k+1}^n E[X_{j:n} | X_{k:n} = x] \\ &= \frac{\lambda}{F(x)} \int_{-\infty}^x t \, dF(t) + \frac{1-\lambda}{1-F(x)} \int_x^{\infty} t \, dF(t). \end{aligned} \tag{14}$$

Further, since $E|X| < \infty$, we have

$$\lim_{x \rightarrow -\infty} xF(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} x(1-F(x)) = 0. \tag{15}$$

Therefore, integrating by parts, we obtain

$$\begin{aligned} \frac{\lambda}{F(x)} \int_{-\infty}^x t \, dF(t) + \frac{1-\lambda}{1-F(x)} \int_x^{\infty} t \, dF(t) &= x - \frac{\lambda}{F(x)} \int_{-\infty}^x F(t) \, dt \\ &\quad + \frac{1-\lambda}{1-F(x)} \int_x^{\infty} (1-F(t)) \, dt. \end{aligned} \tag{16}$$

Thus, from Equations (14) and (16), it follows that Equation (4) is equivalent to

$$\lambda(1-F(x)) \int_{-\infty}^x F(t) \, dt = (1-\lambda)F(x) \int_x^{\infty} (1-F(t)) \, dt.$$

The last equation can be written as

$$-\frac{\lambda}{1-\lambda} \int_{-\infty}^x F(t) \, dt \frac{d}{dx} \left[\int_x^{\infty} (1-F(t)) \, dt \right] = \int_x^{\infty} (1-F(t)) \, dt \frac{d}{dx} \left[\int_{-\infty}^x F(t) \, dt \right],$$

which leads to

$$\int_{-\infty}^x F(t) \, dt = c \left(\int_x^{\infty} (1-F(t)) \, dt \right)^{-\lambda/(1-\lambda)} \quad c > 0.$$

Differentiating both sides with respect to x , we obtain

$$\int_x^{\infty} (1-F(t)) \, dt = c_1 \left(\frac{1}{F(x)} - 1 \right)^{1-\lambda}, \quad c_1 > 0.$$

Differentiating one more time, we have

$$[F(x)]^{2-\lambda} [1-F(x)]^{1+\lambda} = c_2 F'(x), \quad c_2 > 0, \tag{17}$$

which is Equation (12). Referring to Lemma 1 we see that Equation (4) implies Equation (3).

To complete the proof of the theorem, it remains to verify that $F(x)$ with quantile function (3) satisfies Equation (4). Differentiating Equation (3) with respect to x we obtain

$$Q'_\lambda(x) = c(1-x)^{-(1+\lambda)} x^{-(2-\lambda)} \quad c > 0.$$

On the other hand, since $F(Q_\lambda(x)) = x$, we have $Q'_\lambda(x) = [F'(Q_\lambda(x))]^{-1}$. (Note that the right-hand side is clearly differentiable, and thus so is the left-hand side.) Therefore,

$$(1-x)^{1+\lambda}x^{2-\lambda} = cF'(Q_\lambda(x)),$$

which is equivalent to Equation (17) and thus to Equation (4). This completes the proof.

2.2. Proof of Theorem 2

Notice that Equation (11) can be written as

$$\begin{aligned} & (\nu-1)x \left[\frac{1}{n-k} \sum_{j=k+1}^n E[X_{j:n} | X_{k:n} = x] - \frac{1}{k-1} \sum_{j=1}^{k-1} E[X_{j:n} | X_{k:n} = x] \right] \\ &= (\nu-2) \left[\frac{1}{n-k} \sum_{j=k+1}^n E[X_{j:n}^2 | X_{k:n} = x] - \frac{1}{k-1} \sum_{j=1}^{k-1} E[X_{j:n}^2 | X_{k:n} = x] \right]. \end{aligned}$$

Referring to Lemma 2 with $r = 1$ and $r = 2$, we see that this is equivalent to

$$\begin{aligned} & (\nu-1)x \left[\frac{1}{1-F(x)} \int_x^\infty t \, dF(t) - \frac{1}{F(x)} \int_{-\infty}^x t \, dF(t) \right] \\ &= (\nu-2) \left[\frac{1}{1-F(x)} \int_x^\infty t^2 \, dF(t) - \frac{1}{F(x)} \int_{-\infty}^x t^2 \, dF(t) \right]. \end{aligned} \quad (18)$$

Let us assume that $EX = 0$ and $EX^2 = 1$. Hence,

$$\int_x^\infty t \, dF(t) = - \int_{-\infty}^x t \, dF(t) \quad \text{and} \quad \int_x^\infty t^2 \, dF(t) = 1 - \int_{-\infty}^x t^2 \, dF(t)$$

and thus Equation (18) is equivalent to

$$\begin{aligned} -(\nu-1)x \left(\frac{1}{1-F(x)} + \frac{1}{F(x)} \right) \int_{-\infty}^x t \, dF(t) &= \frac{\nu-2}{1-F(x)} - (\nu-2) \left(\frac{1}{1-F(x)} + \frac{1}{F(x)} \right) \\ &\quad \times \int_{-\infty}^x t^2 \, dF(t) \end{aligned}$$

Multiplying the above equation by $F(x)[1-F(x)]$, we find

$$-(\nu-1)x \int_{-\infty}^x t \, dF(t) = (\nu-2) \left[F(x) - \int_{-\infty}^x t^2 \, dF(t) \right]. \quad (19)$$

Differentiating both sides with respect to x , we obtain

$$-(\nu-1) \int_{-\infty}^x t \, dF(t) = f(x)(x^2 + \nu - 2).$$

Since the left-hand side of the above equation is differentiable, we have that $f'(x)$ exists. Differentiating both sides with respect to x , we find

$$\frac{f'(x)}{f(x)} = -\frac{\nu+1}{\nu-2} \frac{x}{1+x^2/(\nu-2)}.$$

Integrating both sides and making use of the fact that $f(x)$ is a pdf, we obtain

$$f(x) = c \left(1 + \frac{x^2}{v-2} \right)^{-(v+1)/2} \quad \text{where} \quad c = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}}. \quad (20)$$

It is not difficult to see that if a random variable Z has the pdf (20), then

$$X = Z \sqrt{\frac{v}{v-2}}$$

follows t_v -distribution, i.e. its pdf is given by Equation (1). Thus, we have proved that Equation (11) implies Equation (10) when $\mu = 0$ and $\sigma^2 = 1$. The result now follows in the general case by considering the linear transformation $Y = \sigma X + \mu$.

To complete the proof, we need to verify that Equation (11) holds when X has a cdf given by Equation (10). If X has pdf (1) (i.e. cdf (10)), then we define

$$Z = X \sqrt{\frac{v-2}{v}},$$

which has pdf (20). Now, it is not difficult to verify that Equation (20) satisfies Equation (19), which in turn is equivalent to Equation (11). The proof is complete.

References

- [1] K. Ferguson and E. Platen, *On the distributional characterization of daily log-returns of a world stock index*, Appl. Math. Finance 13(1) (2006), pp. 19–38.
- [2] M.C. Jones, *Student's simplest distribution*, Statistician 51(1) (2002), pp. 41–49.
- [3] V.B. Nevzorov, N. Balakrishnan, and M. Ahsanullah, *Simple characterizations of Student's t_2 -distribution*, Statistician 52(3) (2003), pp. 395–400.
- [4] V.B. Nevzorov, *On a property of Student's distribution with two degrees of freedom*, Zap. Nauchn. Sem. POMI, 294 (2002), pp. 148–157 (in Russian). (English Translation: J. Math. Sci. 127(1) (2005), pp. 1757–1762.)
- [5] I. Akhundov and V.B. Nevzorov, *A simple characterization of Student's t_3 distribution*, Stats. Probab. Lett. 80(5–6) (2010), pp. 293–295.
- [6] I.S. Akhundov, N. Balakrishnan, and V.B. Nevzorov, *New characterizations by properties of midrange and related statistics*, Commun. Stat. Theory Methods 33(12) (2004), pp. 3133–3143.
- [7] N. Balakrishnan and I. Akhundov, *A characterization by linearity of the regression function based on order statistics*, Stat. Probab. Lett. 63(4) (2003), pp. 435–440.
- [8] M. Ahsanullah and V.B. Nevzorov, *Ordered Random Variables*, NOVA Sci. Publ., Huntington, NY, 2001.