A quantile alternative for kurtosis

Moors, J.J.A.

Publication date:
1987

POSTBUS 90153
5000 LE TILBURG

[redacted]

FACULTEIT DER ECONOMISCHE WETENSCHAPPEN

REEKS "TER DISCUSSIE"

[redacted]

A QUANTILE ALTERNATIVE FOR KURTOSIS

J.J.A. Moors

No. 87.08

A QUANTILE ALTERNATIVE FOR KURTOSIS

J.J.A. Moors

Summary. Recently, MOORS (1986) showed that kurtosis is easily interpreted as a measure of dispersion around the two values $\mu \pm \sigma$. For this dispersion an alternative measure, based on quantiles, is proposed here. It is shown to have several desirable properties: (i) the measure exists even for distributions for which no moments exist, (ii) it is not influenced by the (extreme) tails of the distribution, and (iii) the calculation is simple (and is even possible by graphical means).

1. Introduction

For any distribution the kurtosis k will be defined here as the normalized fourth central moment (provided it exists). So, any random variable X with expectation $\mu := E(X)$, variance $\sigma^2 := V(X)$ and $E(X^4) < \infty$ has

$$k = E(X-\mu)^4/\sigma^4$$

Until recently, the interpretation of kurtosis used to be rather controversial. Most statistical textbooks describe kurtosis in terms of peakedness (versus flatness), while some seek the explanation in the presence of heavy tails or in a combination of the two.

Inspired by DARLINGTON (1970), MOORS (1986b) gave a new and useful interpretation of k. Introduction of the standardized variable $Z := (X-\mu)/\sigma$ gives $k = E(Z^4) = V(Z^2) + [E(Z^2)]^2$, implying

$$(1.1) \quad k = V(Z^2) + 1$$

Therefore, k can be seen as a measure of the dispersion of Z^2 around its expectation 1 or, equivalently, the dispersion of Z around the values -1 and +1. So, k measures the dispersion of X around the two values $\mu \pm \sigma$; it is an inverse measure for the concentration of the distribution of X in these two points. Indeed, (1) attains the minimum value 1 for a symmetric two-point distribution. High kurtosis therefore may arise in two situations:

(i) concentration of probability mass near μ (corresponding to a peaked unimodal distribution), or

(ii) concentration of probability mass in the tails of the distribution.

The existence of these two possibilities explains the past confusion about the meaning of kurtosis.

The first three moments of a given distribution measure location, dispersion and skewness, respectively. For these characteristics of the distribution well-known alternative measures exist, based on quantiles: the median Q, the half interquantile range R and the quantile coefficient of skewness S. Defining an i-th quartile Q_i by

$$(1.2) \quad P(X < Q_i) \le i/4, \quad P(X > Q_i) \le 1 - i/4, \; i = 1, 2, 3$$

leads to the expressions

$$(1.3) \quad \begin{cases} Q = Q_2 \\ R = (Q_3-Q_1)/2 \\ S = (Q_3-2Q_2+Q_1)/(Q_3-Q_1) \end{cases}$$

Up to now, an analogous quantile coefficient of kurtosis did not exist, most probably because a good understanding of the meaning of kurtosis was missing. The closest to such a measure came GROENEVELD & MEEDEN (1984); for symmetric distribution, they took as quantile alternative to kurtosis the skewness coefficient S of the 'upper half' of the distribution. See Section 4 for a further discussion.

In the next section a new quantile measure for the dispersion of a distribution around the values $\mu \pm \sigma$ will be presented. In view of the interpretation of kurtosis, discussed above, this measure can be seen as an alternative to k. Some theoretical properties of the new measure are derived in Section 2 as well, while the behaviour for different types of (symmetrical) distributions is considered in Section 3. The final Section 4 gives a short discussion of the results and surveys related literature.

2. Definition and properties

For any random variable X the octiles E_i are defined by

$$(2.1) \quad P(X < E_i) \le i/8, \quad P(X > E_i) \le 1 - i/8, \; i = 1, 2, \ldots, 7$$

Note the similarity with (1.2); $E_{2i} = Q_i$ (i=1,2,3). For continuous X with distribution function F, the octiles are unique and (2.1) can be simplified to

$$(2.2) \quad F(E_i) = i/8 \; , \; i = 1,2,\ldots,7$$

If Y is defined by Y = aX + b, it is easily checked that $aE_i + b$ is an octile of Y.

For $E_6 > E_2$, the quantity T is defined as

$$(2.3) \quad T = \frac{(E_7-E_5) + (E_3-E_1)}{E_6 - E_2}$$

T is proposed here as alternative to k with the following argument as basic justification. The two terms in the numerator are large (small) if relatively little (much) probability mass is concentrated in the neighbourhood of E_6 and E_2, corresponding with large (small) dispersion around (roughly) $\mu \pm \sigma$. Compare Figure 2.1.

Since both distributions in Figure 2.1 have equal E_2 and E_6, the broken line shows the distribution with the larger T-value.

The denumerator in (2.3) is a normalizing constant, which guarantees the invariance of T under linear transformations. Hence, T is constant over any class of distributions determined by a location-scale parameter, e.g. the class $\{N(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ of normal distributions. As is well-known, k has this property too.

Of course, T takes values in $\mathbb{R}^+ := [0,\infty)$. By way of illustration consider the symmetrical three-point distribution, defined by $P(X=0) = p$,

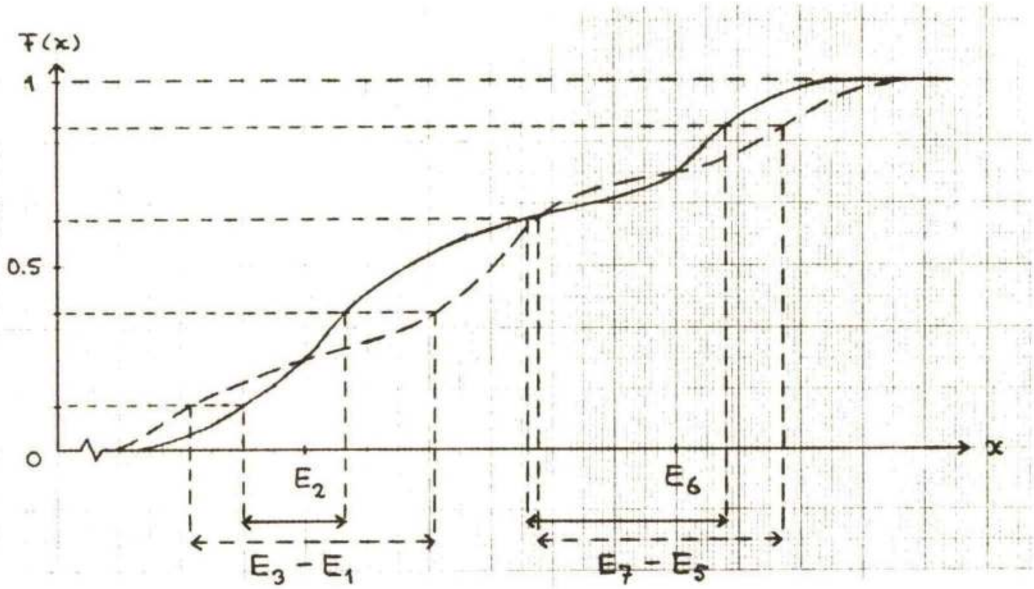

Figure 2.1. Octiles for two continuous distributions

$P(X=-1) = P(X=+1) = (1-p)/2$. It is easily checked that for this distribution

$$T = \begin{cases} 0, & 0 \le p < \frac{1}{4} \\ 1, & \frac{1}{4} < p < \frac{1}{2} \end{cases}$$

holds. For $p = \frac{1}{4}$ or $p = \frac{1}{2}$, T is not uniquely determined, while T does not exist for $p > \frac{1}{2}$. On the other hand, $k = 1/(1-p)$.

Note the analogy between (2.3) and (1.3). For distributions that are symmetrical around 0, (2.3) can be simplified to

(2.4) $\quad T = (E_7 - E_5)/E_6$

3. Behaviour of T

The new measure T will be calculated now for a number of (classes of) symmetrical distributions. Guidelines for the selection of these distributions were (i) the wish to cover a wide range of k-values and (ii) ease of calculation. Because of the invariance property discussed before, it suffices to find T (and k) for one representative in any class determined by a location-scale parameter. For simplicity, a distribution with mean zero will be chosen throughout, so that (2.4) is applicable.

First of all, the following simple representative distributions will be taken into consideration:

(i) the (standard)normal distribution N(0,1),

(ii) the triangular distribution Tr(-1,1) with density

$$f(x) = 1 - |x|,\ 0 \le |x| \le 1$$

(iii) the uniform distribution U(-1,1),

(iv) the double triangular distribution DT(-1,1) with density

$$f(x) = \begin{cases} 2|x|, & 0 \le |x| \le \frac{1}{2} \\ 2-2|x|, & \frac{1}{2} \le |x| \le 1 \end{cases}$$

(v) the inverse triangular distribution IT(-1,1) with density $f(x) = |x|,\ -1 \le x \le 1$.

The densities of the less familiar distributions are drawn in Figure 3.1.

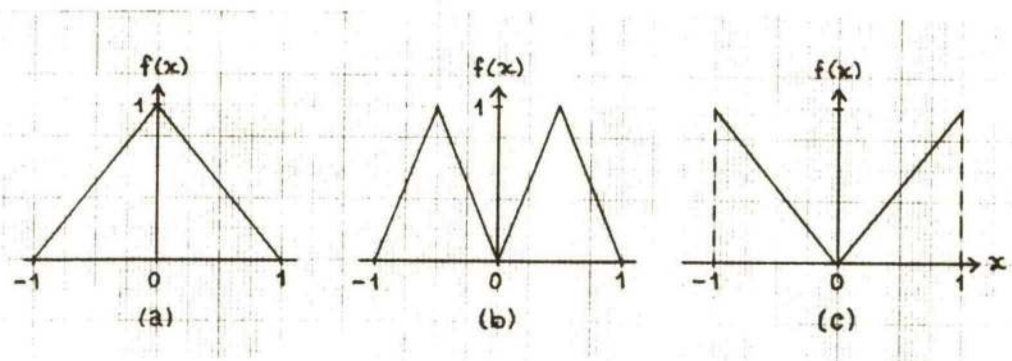

Figure 3.1. Triangular (a), double triangular (b) and inverse triangular (c) distribution

Table 3.1 gives the value of the measure T for these (classes of) distributions and the auxiliary quantities E_5, E_6 and E_7; besides, k is given as well as $E(X^2)$ and $E(X^4)$. Table 3.1 is arranged according tot decreasing k-values.

Table 3.1. Values of T and k for selected distributions

Distribution	E_5	E_6	E_7	T	$E(X^2)$	$E(X^4)$	k
N	0.3186	0.6745	1.1503	1.233	1	3	3
Tr	0.1340	0.2929	0.5	1.250	1/6	1/15	2.4
U	0.25	0.5	0.75	1	1/3	1/5	1.8
DT	0.3536	0.5	0.6464	0.586	7/24	31/240	1.518
IT	0.5	0.7071	0.8660	0.518	1/2	1/3	1.333

Secondly, the class $\{Q(a), -\frac{1}{2} \le a \le 1\}$ is considered, where Q(a) is the probability distribution with the (quadratic) density

(3.1.) $f(x) = (3ax^2+1-a)/2, \ -1 \le x \le 1$

(cf. MOORS 1986a). The property

$$E(X^{2n}) = \frac{4an + 2n + 3}{(2n+1)(2n+3)}, \ n = 0,1,2,\ldots$$

is easily checked and implies

(3.2) $$k = \frac{45(8a+7)}{7(4a+5)^2}$$

so that k is an decreasing function of a. Table 3.2 shows k and T for selected values of a; the calculations show that T is increasing in k.

Table 3.2. Values of T and k for selected distributions Q(a)

a	E_5	E_6	E_7	T	k
-0.5	0.1683	0.3473	0.5579	1.122	2.143
-0.25	0.2016	0.4142	0.6566	1.098	2.009
0	0.25	0.5	0.75	1	1.8
0.25	0.3222	0.5961	0.8177	0.831	1.607
0.5	0.4239	0.6823	0.8612	0.641	1.443
0.75	0.5366	0.7474	0.8894	0.472	1.306
1	0.6300	0.7937	0.9086	0.351	1.190

Next, the class of double gamma distributions $\{DG(\rho): \rho > 0\}$ will be considered where $DG(\rho)$ has the (symmetric) density

$$(3.3)\qquad f(x) = \frac{1}{2\Gamma(\rho)} |x|^{\rho-1}e^{-|x|} , \; x \in \mathbb{R}$$

Note that these distributions are bimodal for $\rho > 1$; they can be obtained from a more general class by putting the scale parameter equal to 1. From

$$(3.4)\qquad E(X^{2n}) = \Gamma(\rho+2n)/\Gamma(\rho), \; n = 0,1,2,\ldots$$

it follows

$$(3.5)\qquad k = \frac{(\rho+3)(\rho+2)}{(\rho+1)\rho}$$

so that k is decreasing in ρ. To find the octiles, first note that these are the quartiles of the well-known gamma distribution $\Gamma(1,\rho)$. Further, $\chi_n^2 = \Gamma(\frac{1}{2},n/2)$ leads to the simple expression

$$E_{i+4} = \tfrac{1}{2}\, \chi^2_{2\rho;i/4}, \; i = 1,2,3$$

for the octiles of DG(ρ). Table 3.3. shows the values of T and k for selected values of ρ.

Table 3.3. Values of T and k for selected distributions DG(ρ)

ρ	T	k	ρ	T	k
0.5	2.686	11.667	10	0.433	1.418
1	1.585	6	15	0.352	1.275
1.5	1.224	4.2	20	0.304	1.205
2	1.032	3.333	30	0.248	1.135
3	0.820	2.5	40	0.214	1.101
5	0.622	1.867	50	0.191	1.081

Again, T is increasing in k according to this table.

Finally, the class of exponential power distributions $\{EP(\alpha): 0 < \alpha \le 1\}$ will be discussed, extensively used by BOX & TIAO (1973). The density of EP(α) can be written as

$$(3.6) \qquad f(x) = \omega(\alpha)\exp[-c(\alpha)|y|^{1/\alpha}]$$

where

$$(3.7) \qquad \omega(\alpha) = \frac{1}{2\alpha}\left[\frac{\Gamma(3\alpha)}{\Gamma^3(\alpha)}\right]^{\frac{1}{2}} \quad c(\alpha) = \left[\frac{\Gamma(3\alpha)}{\Gamma(\alpha)}\right]^{1/2\alpha}$$

This class contains as special cases the double exponential distribution ($\alpha = 1$), the standard normal ($\alpha = \frac{1}{2}$) and - as limiting case - the uniform distribution $U(-\sqrt{3},\sqrt{3})$ ($\alpha \to 0$).

Some algebra give

$$(3.8) \qquad E(X^{2n}) = \Gamma((2n+1)\alpha)\Gamma^{n-1}(\alpha)/\Gamma^{n}(3\alpha)$$

so that

(3.9) $$k = \frac{\Gamma(5\alpha)\Gamma(\alpha)}{\Gamma^2(3\alpha)}$$

Table 3.4 shows the octiles as well as T and k for selected values of α. The octiles were obtained by numerical integration; the results agree with Table 3.2.2 of BOX & TIAO (1973), except the value of E_6 for $\alpha = 0.25$, which is 0.80 according to Box and Tiao.

Table 3.4. Values of T and k for selected distributions $EP(\alpha)$

α	E_5	E_6	E_7	T	k
0	0.4330	0.8660	1.2990	1	1.8
0.125	0.4197	0.8397	1.2683	1.011	1.923
0.25	0.3900	0.7863	1.2290	1.067	2.188
0.375	0.3543	0.7293	1.1903	1.146	2.548
0.5	0.3186	0.6745	1.1503	1.233	3
0.625	0.2853	0.6231	1.1089	1.322	3.553
0.75	0.2549	0.5754	1.0665	1.410	4.222
0.875	0.2277	0.5311	1.0234	1.498	5.029
1	0.2034	0.4901	0.9803	1.585	6

Figure 3.1 shows the values of T and k for all distributions considered. A simple relation does not exist, as was to be expected: after all, T and k are quite different measures. Roughly speaking, however, T is an increasing function of k. For comparison Figure 3.2 shows the analogous picture for the two familiar measures of dispersion, R and σ.

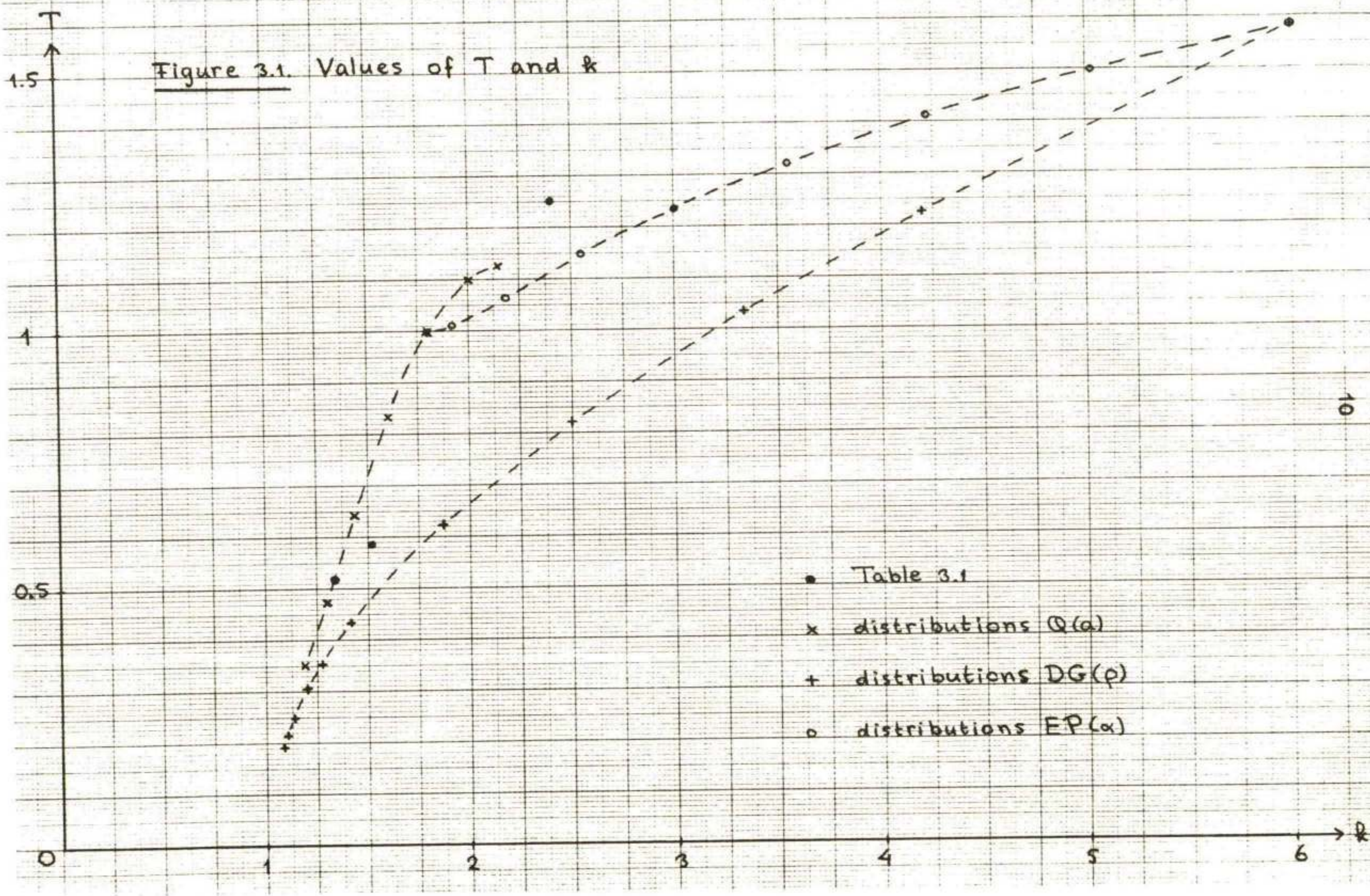
Figure 3.1. Values of T and k
T
1.5
1
0.5
0
1
2
3
4
5
6
k
10
Table 3.1
distributions Q(a)
distributions DG(ρ)
distributions EP(α)

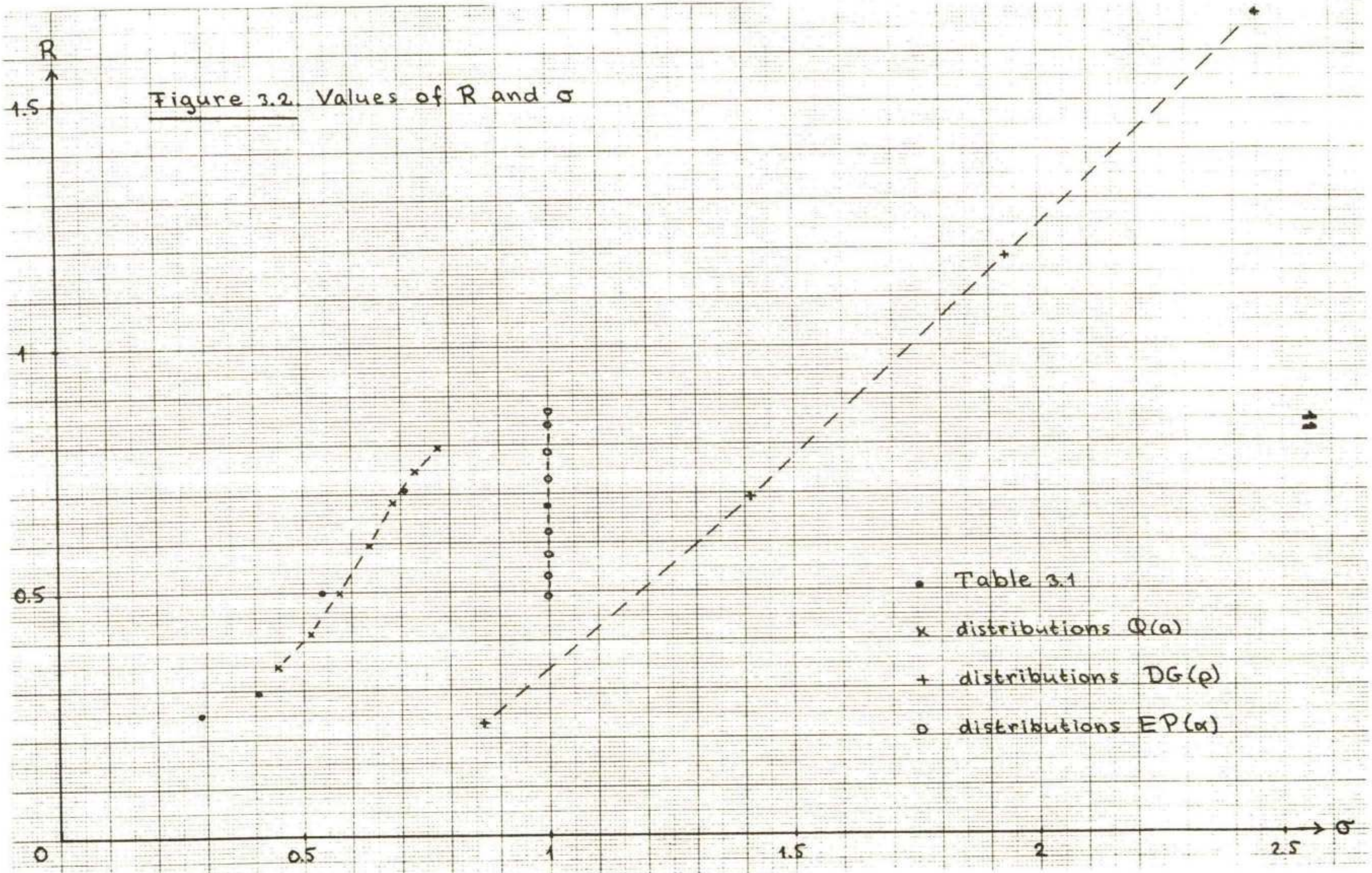

Figure 3.2 Values of R and σ

4. Discussion

The dispersion of a distribution around the two values $\mu \pm \sigma$ can be measured by means of the kurtosis k. An alternative measure T based on quantiles was introduced here. Just as k, T is dimensionless and invariant under location-scale transformations.

T is very analogous to the familiar quantile measures for dispersion and skewness and therefore has similar advantages. The calculation is easy and can even be done by graphical means. T is not very sensible to the extreme tails of the distribution and hence more robust than k. Distributions for which k is not defined still can have finite T. E.g. no moments exist for the Cauchy distribution; however, from the distribution function in the standard case

(4.1) $$F(x) = \frac{1}{2} + \frac{1}{\pi} \text{ arc tg } x$$

easily follows

(4.2) $$E_{i+4} = \text{tg } (i\pi/8), \; i = 1, 2, 3$$

so that T = 2.

Based on the convexity considerations of VAN ZWET (1964) and OJA (1981), GROENEVELD & MEEDEN (1984) discussed several so-called tailness measures for symmetric distributions only. The measures considered are in fact skewness measures, applied to the 'upper half' of a distribution; starting from S in (1.3) this leads to the tailness measure

(4.3) $$b_1^* = \frac{E_7 - 2E_6 + E_5}{E_7 - E_5}$$

As an alternative to the kurtosis k, b_1^* has definite drawbacks. If the two halves of the symmetric distribution are symmetric as well, b_1^* equals zero, of course, while k (and T) may vary. E.g., consider the class of double triangular distributions $\{DT(c): c \geq 1\}$ with densities f(x) sketched in Figure 4.1.

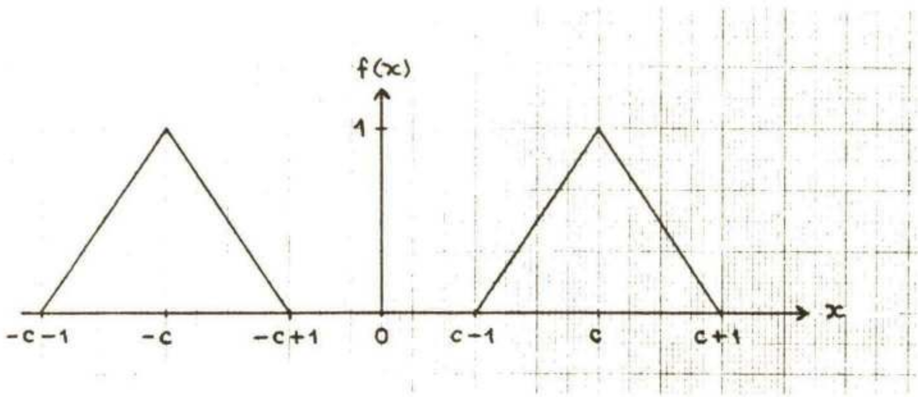

Figure 4.1. Double triangular distribution DT(c)

It is easy to check that k ranges from 1 to 1.518 and T from 0 to 0.586, while b_1^* is identically zero. This reflects the fact that b_1^* essentially is a measure of skewness. On the other hand, T is a (relative) measure of dispersion: (2.4) can be read as 2R/Q, where R and Q from (1.3) apply to the 'upper half' of the distribution. As a consequence, b_1^* is not monotone in k for some classes of distributions, e.g. for $\{Q(a) : -\frac{1}{2} \leq a \leq \frac{1}{2}\}$.

Values of T were calculated for some (classes of) probability distributions, all of them symmetrical (around 0). Calculations for skew distributions are in progress. For choosing an appropriate probability model to describe empirical data, characterizations of large classes of distributions in the so-called (β_1,β_2)-plane are a useful tool (cf. PEARSON 1954); here β_1 and β_2 (= k) are the third and fourth standardized moments. The aim is to construct similar characterizations in the (S,T)-plane.

Acknowledgement

I am much indebted to Ruud M.J. Heuts and Rob W.M. Suykerbuyk for their careful reading of an earlier draft. Besides, the former drew my attention to the exponential power distributions, the latter did many of the calculations.

References

BOX, G.E.P. & TIAO, G.C. (1973), Bayesian inference in statistical analysis, Addison Wesley.

DARLINGTON, Richard B. (1970), Is kurtosis really "peakedness"?, The American Statistician, 24, No. 2, 19-22.

GROENEVELD, R.A. & MEEDEN, G. (1984), Measuring skewness and kurtosis, The Statistician 33, 391-399.

MOORS, J.J.A. (1986a), Het gebruik van 'order statistics' in locatieschatters: een eenvoudig voorbeeld, in: Liber Amicorum Jaap Muilwijk, 58-66 (in Dutch).

MOORS, J.J.A. (1986b), The meaning of kurtosis: Darlington reexamined, The American Statistician, 40, No. 4, 283-284.

OJA, H. (1981), On location,scale, skewness and kurtosis of univariate distributions, Scandinavian Journal of Statistics 8, 154-168.

VAN ZWET, W.R. (1964), Convex transformations of random variables, Math. Centrum, Amsterdam.

IN 1986 REEDS VERSCHENEN

01 F. van der Ploeg
Monopoly Unions, Investment and Employment: Benefits of Contingent Wage Contracts

02 J. van Mier
Gewone differentievergelijkingen met niet-constante coëfficiënten en partiële differentievergelijkingen (vervolg R.T.D. no. 84.32)

03 J.J.A. Moors
Het Bayesiaanse Cox-Snell-model by accountantscontroles

04 G.J. van den Berg
Nonstationarity in job search theory

05 G.J. van den Berg
Small-sample properties of estimators of the autocorrelation coefficient

06 P. Kooreman
Huishoudproduktie en de analyse van tijdsbesteding

07 R.J. Casimir
DSS, Information systems and Management Games

08 A.J. van Reeken
De ontwikkeling van de informatiesysteemontwikkeling

09 E. Berns
Filosofie, economie en macht

10 Anna Harańczyk
The Comparative Analysis of the Social Development of Cracow, Bratislava, and Leipzig, in the period 1960-1985

11 A.J. van Reeken
Over de relatie tussen de begrippen: offer, resultaat, efficiëntie, effectiviteit, produktiviteit, rendement en kwaliteit

12 A.J. van Reeken
Groeiende Index van Informatiesysteemontwikkelmethoden

13 A.J. van Reeken
A note on Types of Information Systems

14 A.J. van Reeken
Het probleem van de Componentenanalyse in ISAC

15 A. Kapteyn, P. Kooreman, R.J.M. Willemse
Some methodological issues in the implementation of subjective poverty definitions

16 I. Woittiez
Preference Interdependence and Habit Formation in Family Labor Supply

17 A.J. van Reeken
A new concept for allocation of joint costs: Stepwise reduction of costs proportional to joint savings

18 A.J. van Reeken
Naar een andere aanpak in de systemering

19 J.G. de Boer, N.J.W. Greveling
Informatieplanning met behulp van referentie-informatiemodellen 1. Totstandkoming bedrijfsinformatiemodellen

20 J.G. de Boer, N.J.W. Greveling
Informatieplanning met behulp van referentie-informatiemodellen 2. Een methode voor informatieplanning

21 W. Reijnders
Direct Marketing: "Van tactiek naar strategie"

22 H. Gremmen
A four economy computer simulation game

IN 1987 REEDS VERSCHENEN

01 J.J.A. Moors
Analytical Properties of Bayesian Cox-Snell Bounds in Auditing

02 H.P.A. Mulders, A.J. van Reeken
DATAAL - een hulpmiddel voor onderhoud van gegevensverzamelingen

03 Drs. A.J. van Reeken
Informatisering en de beloning van arbeid

04 J. Kriens, P.C. van Batenburg
Bayesian Discovery Sampling: a simple model of Bayesian Inference in Auditing.

05 Prof.Dr. J.P.C. Kleijnen
Simulatie

06 Rommert J. Casimir
Characteristics and implementation of decision support systems

07 Rommert J. Casimir
Infogame, the model