SOME CONSEQUENCES OF A CHARACTERIZATION THEOREM BASED ON TRUNCATED MOMENTS

Wolfgang Glänzel

Library of the Hungarian Academy of Sciences, Information Science and Scientometrices Research Unit, Budapest, Hungary

Summary. Let X be an arbitrary real-valued random variable and let g and h be two real functions such that $E(g(X)|X \geq x)$ and $E(h(X)|X \geq x)$ are defined. It was shown that under relatively mild conditions the distribution of X is uniquely determined by the functions g, h and $\lambda = E(g(X)|X \geq x)/E(h(X)|X \geq x)$ (Glänzel, 1987). In the present paper we show that this characterization is stable in the sense of weak convergence. Some consequences of this stability theorem and the above characterization are discussed.

Keywords. Characterization; truncated moments; representation; stability; weak convergence; Irwin's system; normal distribution

1 INTRODUCTION

In the last fifteen years, several interesting results concerning characterizations based on truncated moments have been published. Among the first results, those of Kotlarski (1972) and Gupta (1975) deserve to be mentioned here. The book by Galambos and Kotz (1978) on characterizations of probability distributions has already devoted one chapter to this topic. The papers by Kotz and Shanbhag (1980) and Shanbhag and Kotz (1987) presented very general results and showed a broad spectrum of applications of characterizations by conditional expectations in the case of univariate and multivariate distributions, respectively. In order to improve the applicability of this method, Glänzel et al. (1984) substituted the truncated moment by the ratio of of different truncated moments of the same random variable, and showed that also under this condition the distribution of the random variable is characterized. This result was generalized in the papers by Glänzel (1987, 1991) where the author applied the obtained results to the characterization of relatively wide classes of continuous as well as discrete distributions (Irwin's system, Pearson system, discrete Pearson system, Cauchy distribution). In the present paper we show that the characterization based on the ratio of two truncated moments is stable in the sense of weak convergence. Thereby the spectrum of applications can be extended. This will be illustrated by three examples.

2 A STABILITY THEOREM

Let $(\Omega, \mathcal{A}, P)$ be a given probability space and let $X : \Omega \to H$ be a random variable, where $H = [a, b]$ for some real $a < b$; $a \geq -\infty$ and $b \leq +\infty$. The distribution function of X is denoted

The final publication is available at Springer via http://dx.doi.org/10.1080/02331889008802273

by F. First of all we give a result of an earlier paper by the author (Glänzel, 1987) since it will be the base for the further discussion:

Theorem 1 (A representation theorem)**.** *Let g and h be two real functions defined on H such as*

$$E(g(X)|X \geq x) = \lambda(x) \cdot E((h(x)|X \geq x)\,; \quad x \in H$$

is defined. Assume that $g, h \in C^1(H)$ and λ is a left-continuous function with an at most countable set of discontinuity points on H. Further assume that either $\lambda(x)h(x) > g(x)$ or $\lambda(x)h(x) < g(x)$ on int H. *Finally assume that*

$$\int_H \frac{\lambda h' - g'}{\lambda h - g} = -\infty\,.$$

Then F is uniquely determined by the functions g, h and λ.

The following result, a generalisation of a stability theorem by Kotz and Shanbhag (1980), shows that this characterization is stable in the sense of weak convergence.

Let $\{X_n\}$ be a sequence of random variables with the distributions $\{F_n\}$ and let X be a random variable with the distribution F such that $X_n : \Omega \to H_n = [a, b_n]$ and $X : \Omega \to H = [a, b]$. Assume that $g, h \in C^1(H), g_n, h_n \in C^1(H_n), g_n \to g, h_n \to h$ and

$$\lambda_n(x) = \frac{E\big(g_n(X_n)|X_n \geq x\big)}{E\big(h_n(X_n)|X_n \geq x\big)}\,; \quad x \in H_n$$

and the conditions of the characterization theorem are satisfied.

Theorem 2. *Assume that $H_n = H$ for each $n \in \mathbb{N}$, $h_n(X)$, $g_n(X)$ are uniformly integrable and the family $\{F_n\}$ is relative compact. Then under the above conditions $X_n \stackrel{\mathcal{D}}{\to} X$ if and only if λ_n converges weakly to λ. ($X_n \stackrel{\mathcal{D}}{\to} X$ and $\lambda_n \Rightarrow \lambda$ denote the convergence in distribution and the weak convergence of functions and measures, respectively.)*

Proof. Since the proof of this theorem partially parallels with the proofs of Proposition 4 of Kotz and Shanbhag (1980) and Theorem 4 of Shanbhag and Kotz (1987), we give only a sketch which is restricted to the essential line of thoughts of the proof.

1. Assume that $X_n \stackrel{\mathcal{D}}{\to} X$. From Rubin's theorem $h_n(X_n) \stackrel{\mathcal{D}}{\to} h(X)$ and $g_n(X_n) \stackrel{\mathcal{D}}{\to} g(X)$ follow. Hence and from the uniform integrability the statement is obtained.
2. To establish the 'only if' part, we assume that λ_n converges weakly to λ but $X_n \stackrel{\mathcal{D}}{\to} X$ does not hold.

    2.1. Helly's selection theorem ensures that there is a subsequence $\{F_{n_\alpha}\} \subset \{F_n\}$ converging weakly to some left-continuous, monotone non-descending, non-negative and bounded function F^*. Because of the relative compactness of $\{F_n\}$ the function F^* must be a distribution function on H. Thus $F^* \neq F$ must be assumed.

2.2. From the 'if' part and the assumption $\lambda_n \Rightarrow \lambda$ the following result is obtained:

$$\lambda_{n_\alpha}(x) \Rightarrow \lambda(x) = \lambda^*(x) = \int_x^b g\, dF^* \cdot \left(\int_x^b h\, df^* \right)^{-1}$$

for every $x \in H$. Hence and from Theorem 1 $F \equiv F^*$ follows which contradicts the indirect assumption. Thus the proof is complete

□

Theorem 3. *Assume that the definition of λ is extended in the following way:*

$$\lambda(x) = \begin{cases} E(g(X))/E((h(X))\,, & \text{if} \quad x < a \\ E(g(X)|X \geq x)/E((h(X)|X \geq x)\,, & \text{if} \quad a \leq x \leq b \\ g(b)/h(b)\,, & \text{if} \quad x > b\,. \end{cases}$$

Put $b^* = \sup \bigcup_n H_n$ and $H^* = [-\infty, b^*]$. Then the validity of Theorem 2 can be extended to the sets $H_n = H^*$.

Proof. Only two things must be proved: (1) The above extension of the definition of λ does not cause the loss of validity of the statement of Theorem 1 and (2) $F(b^*) = 1$. Either statement can easily be proved by indirect assumption. □

Remark. If $h_n \equiv 1$ for every $n \in \mathbb{N}$, a version of Proposition 4 of Kotz and Shanbhag (1980) is obtained.

The following corollary shows that we need no sequence of random variables. For given functions g and h a suitable sequence of left-continuous real functions define a corresponding distribution if certain conditions are met.

Corollary. *Let $\{\lambda_n\}$ be a sequence of left-continuous real functions and g_n, $h_n \in C^1(H_n)$ real functions defined on a set H such that the conditions of the characterization theorem are satisfied (g_n/h_n is of bounded variation, $\lambda_n h_n - g_n$ does not vanish nor change its sign on* int H *and*

$$\int_H \frac{\lambda_n h_n' - g_n'}{\lambda_n h_n - g_n} = -\infty\,).$$

Assume that $g_n \to g$, $h_n \to h$. If $\{\lambda_n\}$ converges weakly to some left-continuous function λ and the functions λ, g and h satisfy the conditions of the characterization theorem then the according distribution functions F_n defined by λ_n, g_n and h_n converge weakly to a distribution function F which is uniquely determined by λ, g and h.

3 APPLICATIONS

In the followings we give examples for some possible applications of the above Theorems 1 and 3. In some earlier papers we showed that relatively large classes of probability distributions can be represented by a joined characterization based on the ratio of two truncated moments (cp. Glänzel et al., 1984, Glänzel, 1987, 1991). Theorem 3 ensures that these classes are closed with respect to the weak convergence.

Example 1. Let X be a non-negative integer-valued random variable. Assume that $\{X_k\}$ is a sequence of random variables having binomial distributions with parameters

$$n_k = k \quad \text{and} \quad p_k = b/k \quad (0 < b \leq 1)\,.$$

Then the characterization equation has the following form (cp. Glänzel et al., 1984):

$$E(X_k|X_k \geq n) = n + b - b \cdot n \cdot (1 + 1/k) \cdot E(1/(X_k + 1|X_k \geq n)\,; \quad n \in H_k\,,$$

where $H_k = \{i \in \mathbb{Z} : 0 \leq i \leq k\}$. Assume that $X_n \xrightarrow{\mathcal{D}} X$ where X has a Poisson distribution with parameter b. According to the notations of Theorem 3 we have:

$$\begin{aligned} g_k(X_k) &= X_k - b \\ h_k(X_k) &= 1 - b \cdot (1 + 1/k)/(X_k + 1) \\ \lambda_k(n) &= n \quad \text{for every } k \in \mathbb{N} \quad \text{and every } n \in H_K\,. \end{aligned}$$

Since the conditions of Theorem 3 are met, the following characterization is obtained:

$$\frac{E(X - n|X \geq n)}{1 - E(n/(X + 1)|X \geq n)} = b\,; \qquad n \in H = \mathbb{N} \cup \{0\}\,,$$

which is indeed the characterization of the Poisson distribution (cp. Glänzel et al., 1984).

The following example illustrates the applicability of the above results to limit-distribution problems.

Example 2.

Proposition. *Let $\{X_i\}$ be a sequence of totally independent random variables. Put $Y_n = \sum_{i=1}^{n} X_i$, $M_n = \sum_{i=1}^{n} E(X_i)$ and $S_n = \left\{\sum_{i=1}^{n} D^2(X_i)\right\}^{1/2}$. Then*

$$(Y_n - M_n)/S_n \xrightarrow{\mathcal{D}} N(0,1) \quad \textit{as} \quad n \to \infty$$

if and only if

$$\begin{aligned} &E\big(\{(Y_n - M_n)/S_n\}^2|(Y_n - M_n)/S_n \geq x\big) \\ &\Rightarrow E\big((Y_n - M_n)/S_n|(Y_n - M_n)/S_n \geq x\big) \cdot x + 1 \quad \textit{as} \quad n \to \infty\,. \end{aligned}$$

Proof. The proposition is a consequence of Theorem 2 and the characterization of the normal distribution: The random variable $X : \Omega \to \mathbb{R}$ has a normal distribution $N(m, \sigma)$ if and only if

$$E\big(X(X-x)|(X-x) \geq 0\big) = m \cdot E\big((X-x)|(X-x) \geq 0\big) + \sigma^2 ; \quad x \in \mathbb{R}$$

(cp. Glänzel, 1987) □

The concluding example shows how statistical tests and parameter estimations can be developed from Theorems 1 and 3, if empirical distributions are given.

Example 3. Let $\boldsymbol{X} = \{X_i\}_{i=1}^n$ be a given sample where $X_i : \Omega \to H$ are independent identically distributed random variables and let F be their common distribution function. For given real functions and $g,\ h \in C^1(H)$ we define the empirical truncated moments as follows:

$$\bar{e}_g(x) = \frac{\sum_{i=1}^{n} g(X_i)\, \chi(X_i \geq x)}{\sum_{i=1}^{n} \chi(X_i \geq x)} ; \qquad x \in H ,$$

provided the ratio has a sense. $\bar{e}_h(x)$ is defined in analogous manner. According to Glivenko's theorem $\bar{F}_n \to F$ with probability 1, where $\bar{F}_n$ is the empirical distribution functions for the sample size n. As a consequence of Theorems 1 and 2, $\overline{\lambda}_n = \bar{e}_g(x)/\bar{e}_h(x)$ converges (weakly) to $\lambda = E(g(X)|X \geq x)/E(h(X)|X \geq x)$. Reversely, if $\overline{\lambda}_n \Rightarrow \lambda$, then $\bar{F}_n \Rightarrow \bar{F}$ where $\bar{F}$ is a distribution function determined by g, h and λ according to Theorem 1, and due to Glivenko's theorem $\bar{F} \equiv F$ with probability 1. Thus F is 'approximately' characterized by the given functions g and h and the ratio of empirical truncated moments $\overline{\lambda}_n$.

We give two examples:

(1) A non-negative integer-valued random variable X has a Waring distribution with parameters $\alpha > 1$ and N, i.e.,

$$P(X=k) = \frac{\alpha}{N+\alpha} \cdot \frac{N}{N+\alpha+1} \cdot \ldots \cdot \frac{N+k-1}{N+\alpha+k}; \quad k \geq 0 ,$$

if and only if

$$E(X|X \geq k) = a \cdot k + b ; \qquad k \geq 0 ,$$

where $a = \alpha/(\alpha-1)$ and $b = N/(\alpha-1)$. A sample $\{X_i\}_{i=1}^n$ can be considered to be taken from a Waring distributed populations if

$$\bar{e}_g(k) = \frac{\sum_{i=1}^{n} X_i\, \chi(X_i \geq k)}{\sum_{i=1}^{n} \chi(X_i \geq k)} \approx ak + b ; \quad 0 \leq k \leq \max\{X_i : X_i \in \boldsymbol{X}\} .$$

Hence a statistical test as well as parameter estimation can easily be developed (cp. Telcs et al., 1985).

(2) A similar test for the normal distribution is based on the characterization of the normal distribution given in Example 2. Accordingly a sample can be considered to be taken from a normal distributed population if

$$D(x) = m \cdot d(x) + \sigma^2 ; \qquad x \in \mathbb{R} ,$$

where $D(x)$ and $d(x)$ are estimators of the truncated moments $E\big(X(X-x)|X-x \geq 0\big)$ and $E\big((X-x)|X-x \geq 0\big)$, respectively. $D(x)$ and $d(x)$ can be calculated as described in the preceding case. This test can also be used as a parameter estimation. Interesting is that here the distribution function of a normal distribution has been transformed to a linear function of truncated moments.

Acknowledgement. *Contributed paper at the 18th European meeting of statisticians, August 1988, Berlin*

REFERENCES

Galambos, J., Kotz, S. (1978), Characterizations of probability distributions. A unified approach with an emphasis on exponential and related models. *Lecture Notes in Mathematics*, 675, Springer, Berlin–Heidelberg–New York.

Glänzel, W., Telcs, A., Schubert, A. (1984), Characterization by truncated moments and its application to Pearson-type distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66, 1984, 173–183. (Correction: *Probab. Th. Rel. Fields*, 74, 1987, 317.)

Glänzel, W. (1987), *A characterization theorem based on truncated moments and its application to some distribution families*. In: P. Bauer, F. Konecny, W. Wertz (Eds.), Mathematical Statistics and Probability Vol. B, D. Reidel Publishing Company Dordrecht-Holland, 75–84.

Glänzel, W. (1991),A Characterization Theorem for Pearson-Type distributions and discrete analogous. *Sankhya Ser. B*, 53, 17–24.

Gupta, R.C. (1975), On characterization of distributions by conditional expectations. *Commun. Statist.*, 4, 99–103.

Kotlarski, I.I. (1972), On a characterization of some probability distributions by conditional expectations. *Sankhya Ser. A* 34, 461–466.

Kotz, S., Shanbhag, D.N. (1980), Some new approaches to probability distributions. *Advanves in Applied Probability*, 12, 903–921.

Shanbhag, D.N., Kotz, S. (1987), Some new approaches to multivariate probability distributions. *J. Multivariate Anal.*, 22, 189–211.

Telcs, A., Glänzel, W., Schubert, A. (1985), Characterization and statistical test using truncated expectations for a class of skew distributions. *Math. Soc. Sci.*, 10 (1985) 169–178.