Aust. N. Z. J. Stat. **47**(4), 2005, 543–547

A CHARACTERIZATION OF CONTINUOUS DISTRIBUTIONS VIA REGRESSION ON PAIRS OF RECORD VALUES

I. Bairamov[1], M. Ahsanullah[2] and Anthony G. Pakes[3*]

Izmir University of Economics, University of South Florida and The University of Western Australia

Summary

The exponential type is characterized in terms of the regression of a (possibly non-linear) function of a record value with its adjacent record values as covariates. Monotone transformations extend this result to more general settings, and these are illustrated with some specific examples.

Key words: characterization; record values; regression.

1. Introduction

Let $X_1, X_2, \ldots$ be independent copies of a random variable X whose distribution function is denoted by F. There have been many studies on characterizations of F via linear regression relations of one order statistic on one or two other order statistics. See Wesolowski & Ahsanullah (1997) for references and the most complete results. Similarly, the regression of one record value on another has received attention. Denote record times by $L(1) = 1$ and, for $n > 1$,

$$L(n) = \min\{j\colon\ j > L(n-1) \text{ and } X_j > X_{L(n-1)}\},$$

and corresponding record values by $X(n) = X_{L(n)}$; see Nevzorov (2001).

Fix positive integers n and s, and real constants a and b. The one-variable linear regression problem is to determine all F for which

$$\mathrm{E}\big(X(n+s) \mid X(n) = x\big) = ax + b\,. \tag{1}$$

Nagaraja (1977) first addressed this in the adjacent case $s = 1$, showing that solutions exist if $a > 0$ and that solutions comprise a unique type whose form depends on the value of a relative to unity. Ahsanullah & Wesolowski (1998) proved that the same solutions are determined in the case $s = 2$. Finally Dembinska & Wesolowski (2000) proved that the general case can be reduced to an application of the Lau–Rao theorem for solving the extended Cauchy equation. See Arnold *et al.* (1998 Section 4.4.2,3) for a compact account of the Lau–Rao theorem.

A formal extension of these results replaces (1) with $\mathrm{E}(h(X(n+s)) \mid X(n) = x) = ax + b$, where h is a continuous and strictly monotonic function. In the case $s = 1$ Franco & Ruiz

Received February 2004; revised June 2004; accepted November 2004.
* Author to whom correspondence should be addressed.
1 Dept of Mathematics, Izmir University of Economics, 35330 Balcova, Izmir, Turkey.
2 Dept of Mathematics, University of South Florida, Tampa, Florida 33629, USA.
3 School of Mathematics & Statistics, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009. e-mail: pakes@maths.uwa.edu.au

(1996) investigate the deeper problem of determining F from a specification of this function, which need not be linear. In the adjacent case, these problems, as well as order-statistics versions, can be transformed to characterization by the form of the conditional expectation of X given truncation events such as $\{X \le x\}$. These connections are reviewed by Pakes (2004) who shows how product integration yields simple solutions requiring very weak conditions, if any, on F.

In this paper we investigate the characterization of F in terms of the bivariate regression function $\mathrm{E}(H(X(n)) \mid X(n-1)=u, X(n+1)=v)$ where $\ell_F \le u < v \le r_F$ and $\ell_F = \inf\{x\colon\ F(x)>0\}$ and $r_F = \sup\{x\colon\ F(x)<1\}$, respectively, are the left and right extremities of F. Pakes (2004 Section 5) alludes briefly to this problem, but we aim to give an elementary treatment for the continuous case.

2. A characterization of the exponential type

Let $\bar F(x) = 1 - F(x)$ and, for $\ell_F \le x < r_F$, let $B(x) = -\log \bar F(x)$. The Markov dependence of record values (Arnold, 1998 p.28 or Nevzorov, 2001 p.68) can be used to show that

$$\Pr\big(X(n) \in dx \mid X(n-1)=u, X(n+1)=v\big) = \frac{dB(x)}{B(v)-B(u)},$$

whence

$$\mathrm{E}\big(H\big(X(n)\big) \mid X(n-1)=u, X(n+1)=v\big) = \frac{1}{B(v)-B(u)} \int_u^v H(x)\, dB(x). \qquad (2)$$

We draw two conclusions from this expression. First, if F is the standard exponential distribution function then $B(x)=x$ $(x \ge 0)$ and hence the right-hand side of (2) equals

$$\frac{1}{v-u}\int_u^v H(x)\,dx = \frac{h(v)-h(u)}{v-u},$$

where $h'(x) = H(x)$. Second, if $h(x) = cx+b$, where c and b are real constants, then the right-hand side of (2) equals c for any choice of F. Consequently, without loss of generality we can restrict attention to the regression relation

$$\mathrm{E}\big(h'\big(X(n)\big) \mid X(n-1)=u, X(n+1)=v\big) = \frac{h(v)-h(u)}{v-u} \qquad (\ell_F < u < v < r_F), \quad (3)$$

subject to a restriction on h which excludes affine functions. Our first result does this under an additional restriction to absolutely continuous F. In the sequel, in any reference to (2) we understand that $H(x) = h'(x)$.

Theorem 1. *Suppose F is absolutely continuous with density function f, that h is continuous in $[\ell_F, r_H]$ and continuously differentiable in (ℓ_F, r_F), and that almost everywhere in this open interval,*

$$h'(x) \ne \frac{h(x)-h_1}{x-\ell_F}, \qquad (4)$$

where $h_1 = h(\ell_F+)$. Then (3) holds if and only if $\ell_F > -\infty$, $r_F = \infty$ and

$$F(x) = 1 - e^{-c(x-\ell_F)} \qquad (x \ge \ell_F), \qquad (5)$$

where $c>0$ is an arbitrary constant.

Proof. Equation (5) implies (3). For the converse, the continuity of F entails $B(\ell_F) = 0$, and hence letting $u \to \ell_F+$ in (3), it follows from (2) that

$$\int_{\ell_F}^{v} h'(x)B'(x)\,dx = B(v)\frac{h(v)-h_1}{v-\ell_F} \qquad (\ell_F < v < r_F)\,.$$

Differentiating and re-arranging terms yields the identity

$$\Big(h'(v) - \frac{h(v)-h_1}{v-\ell_F}\Big)B'(v) = \frac{B(v)}{v-\ell_F}\Big(h'(v) - \frac{h(v)-h_1}{v-\ell_F}\Big)\,.$$

The condition (4) implies that $B'(v)/B(v) = 1/(v-\ell_F)$ and hence that $\bar{F}(v) = e^{-c(v-\ell_F)}$. It follows that $\ell_F > -\infty$ and $c > 0$, and then the continuity of F implies that $r_F = \infty$.

We conjecture that the absolute continuity assumption is unnecessary. Lemma 1 shows that we can transfer this assumption to a further smoothness assumption about h.

Lemma 1. *If $h''(x)$ exists in (ℓ_F, r_F), and (3) and (4) both hold, then F is absolutely continuous.*

Proof. Again with $u = \ell_F$, integration by parts of the integral in (2) and re-arranging yields

$$\int_{\ell_F}^{v} B(x)h''(x)\,dx = B(v)\Big(h'(v) - \frac{h(v)-h_1}{v-\ell_F}\Big)\,.$$

The integral is absolutely continuous and the coefficient of $B(v)$ is differentiable, and so it follows from the assumptions that B is absolutely continuous.

Differing choices of h yield many characterizations of the exponential type. One choice satisfying Lemma 1 is $h(x) = \frac{1}{2}x^2$.

Theorem 2. *The continuous random variable X has the exponential type (5) if and only if*

$$\mathrm{E}\big(X(n) \mid X(n-1) = u, X(n+1) = v\big) = \tfrac{1}{2}(u+v) \qquad (\ell_F < u < v < \infty)\,.$$

3. Monotone transformations

In this section we give a formal generalization of Theorem 1 which arises from monotone transformation of X. Specifically, we give a regression condition which specifies the form of the distribution function G of a random variable Y. The corresponding record values are denoted by $Y(n)$.

Theorem 3. *Suppose that:*

(i) *a random variable Y has a continuous distribution function G supported on $[\ell_G, r_G]$;*

(ii) *the function R is continuous and strictly increasing in (ℓ_G, r_G) and*

$$\tau = R(\ell_G+) > -\infty \quad \textit{and} \quad R(r_G) = \infty\,; \tag{6}$$

and

(iii) *h is twice differentiable in (ℓ, ∞).*

Then

$$\mathrm{E}\big(h'\big(R\big(Y(n)\big)\big) \mid Y(n-1)=s, Y(n+1)=t\big) = \frac{h(R(t))-h(R(s))}{R(t)-R(s)} \qquad (\ell_G < s < t < r_G) \tag{7}$$

if and only if

$$G(y) = 1 - e^{-c(R(y)-\tau)} \qquad (\ell_G < y < r_G)\,, \tag{8}$$

where $c>0$ is an arbitrary constant. If (ii) is replaced by

(ii$'$) *the function R is continuous and strictly decreasing in (ℓ_G, r_G) and*

$$R(\ell_G+) = \infty \quad \textit{and} \quad \tau = R(r_G)\,;$$

then (7) holds if and only if

$$G(y) = e^{-c(R(y)-\tau)} \qquad (\ell_G < y < r_G)\,, \tag{9}$$

where $c>0$ is an arbitrary constant.

Proof. If $X = R(Y)$ then the random variables $R(Y(j))$ $(j = n-1, n, n+1)$ have the same joint distribution as $X(n-1), X(n), X(n+1)$. It follows that (7) is equivalent to (3) with $u = R(s)$ and $v = R(t)$. Theorem 1 and Lemma 1 imply that $R(Y)$ has an exponential type distribution (5) (with τ replacing ℓ_F) and the assumptions about R yield

$$G(y) = P\big(R(Y) \le R(y)\big) = P\big(X \le R(y)\big)\,,$$

whence (8). The argument is reversible. Equation (6) is necessary for the continuity of G. Altering some details shows that when (ii$'$) replaces (ii) then (7) is equivalent to (9).

Combining Theorems 2 and 3 yields the following result which gives many special cases.

Theorem 4. *If (i) and (ii) (resp. (ii$'$)) of Theorem 3 hold then (8) (resp. (9)) holds if and only if*

$$\mathrm{E}\big(R\big(Y(n)\big) \mid Y(n-1)=s, Y(n+1)=t\big) = \tfrac{1}{2}\big(R(s)+R(t)\big) \qquad (\ell_G < s < t < r_G)\,.$$

Examples 1–3 follow from Theorem 4 with Condition (ii).

Example 1. If $\ell_G = 0$, $r_G = \infty$, and $R(y) = y^\alpha$ for some constant $\alpha > 0$, then $\tau = 0$ and Y has the Weibull distribution with $G(y) = 1 - \exp(-cy^\alpha)$ if and only if

$$\mathrm{E}\big(Y^\alpha(n) \mid Y(n-1)=s, Y(n+1)=t\big) = \tfrac{1}{2}(s^\alpha + t^\alpha) \qquad (0 < s < t < \infty)\,.$$

Example 2. If $\ell_G = 0$, $r_G = 1$ and $R(y) = -\log(1-y)$ then $\tau = 0$ and $G(y) = 1-(1-y)^c$ if and only if

$$\mathrm{E}\big(\log\big(1-Y(n)\big) \mid Y(n-1)=s, Y(n+1)=t\big) = \tfrac{1}{2}\log\big((1-s)(1-t)\big) \qquad (0 \le s < t < 1)\,.$$

Example 3. If $\ell_G = a > 0$, $r_G = \infty$ and $R(y) = \log(y/a)$ then $\tau = 0$ and Y has the Pareto distribution with

$$G(y) = 1 - (a/y)^c \qquad (y \ge a)$$

if and only if

$$\mathrm{E}\big(\log Y(n) \mid Y(n-1)=s, Y(n+1)=t\big) = \log\sqrt{st} \qquad (a \le s < t < \infty)\,.$$

This is unexpected insofar as the form of the regression relation is independent of a.

Examples 4 and 5 arise from Theorem 4 with Condition (ii$'$).

Example 4. If $\ell_G = -\infty$, $r_G = \infty$, and $R(y) = e^{-y}$ then $\tau = 0$ and Y has the Gumbel extremal process marginal distribution function $G(y) = \exp(-ce^{-y})$ if and only if

$$\mathrm{E}\big(e^{-Y(n)} \mid Y(n-1) = s, Y(n+1) = t\big) = \tfrac{1}{2}(e^{-s} + e^{-t}) \qquad (-\infty < s < t < \infty)\,.$$

The case $c = 1$ is the standard Gumbel distribution function $\Lambda(y)$.

Example 5. If $\ell_G = -\infty$, $r_G = \infty$, and $R(y) = \log(1 + e^{-y})$ then $\tau = 0$ and Y has a generalized logistic distribution, $G(y) = (1 + e^{-y})^{-c}$ if and only if

$$\begin{aligned}\mathrm{E}\big(\log(1 + e^{-Y(n)}) \mid Y(n-1) = s, Y(n+1) = t\big)& \\ = \tfrac{1}{2}\log\big((1 + e^{-s})(1 + e^{-t})\big)& \qquad (-\infty < s < t < \infty)\,.\end{aligned}$$

The constructions based on Condition (ii) are particular cases of the following. Suppose $A(y)$ is a continuous distribution function with support $[\ell_G, r_G]$, and strictly increasing therein, and if $R(y) = -\log \bar{A}(y)$ then

$$G(y) = 1 - \big(1 - A(y)\big)^c$$

for some $c > 0$ if and only if (7) holds. Thus $A(y) = 1 - \exp(-y^\alpha)$ for Example 1, $yI_{[0,1]}(y)$ for Example 2, and $1 - a/y$ for Example 3. Similarly, taking $R(y) = -\log A(y)$ yields $G(y) = (A(y))^c$ if and only if (7) holds. Thus $A(y) = \Lambda(y)$ for Example 4, and $A(y) = (1 + e^{-y})^{-1}$ for Example 5.

References

AHSANULLAH, M. & WESOLOWSKI, J. (1998). Linearity of best predictors for non-adjacent record values. *Sankhyā Ser. B* **60**, 221–227.

ARNOLD, B.C., BALAKRISHNAN, N. & NAGARAJA, H.N. (1998). *Records.* New York: Wiley.

DEMBINSKA, A. & WESOLOWSKI, J. (2000). Linearity of regression for non-adjacent record values. *J. Statist. Plann. Inference* **90**, 195–205.

FRANCO, M. & RUIZ, J.M. (1996). On characterization of continuous distributions by conditional expectation of record values. *Sankhyā Ser. A* **58**, 135–141.

NAGARAJA, H.N. (1977). On a characterization based on record values. *Austral. J. Statist.* **19**, 70–73.

NEVZOROV, V.B. (2001). *Records: Mathematical Theory.* Providence RI: American Mathematical Society.

PAKES, A.G. (2004). Product integration and characterization of probability laws. *J. Appl. Statist. Sci.* **13**, 11–31.

WESOLOWSKI, J. & AHSANULLAH, M. (1997). On characterizing distributions via linearity of regression for order statistics. *Austral. J. Statist.* **39**, 69–78.