Statistics & Probability Letters 17 (1993) 225–230
North-Holland

18 June 1993

Characterizations of identically distributed independent random variables using order statistics

R.B. Bapat and Subhash C. Kochar

Indian Statistical Institute, New Delhi, India

Received December 1991

Abstract: Suppose $X_1,\ldots,X_n$ are independent random variables of continuous type with proportional hazard (or failure) rates. Let $X_{(r)}$ denote the rth order statistic and let $I_{(r)}=i$ if $X_i=X_{(r)}$. It is proved that X_i's are identically distributed if and only if $X_{(r)}$ and $I_{(r)}$ are independent for some $r\in\{2,\ldots,n\}$. The second characterization is in terms of order statistics from subsamples.

Keywords: Order statistics; hazard rate; competing risks.

1. Introduction and the main result

Suppose $X_1,\ldots,X_n$ are independent absolutely continuous random variables with f_i, F_i and $\bar{F}_i$ $(=1-F_i)$ denoting, respectively, the probability density function, the distribution function and the survival function of X_i, $i=1,\ldots,n$. We shall use the notation $r_i(\cdot)$ for the hazard (or failure) rate of X_i.

For $1\leqslant r\leqslant n$, let $X_{(r)}$ denote the rth order statistic and let $I_{(r)}=i$ if $X_{(r)}=X_i$. Since the observations are assumed to be of continuous type, the functions $I_{(r)}$'s are uniquely defined with probability one.

X_i's are said to have proportional hazard rates if there exist positive constants γ_i, $i=1,\ldots,n$, such that

$$r_i(x)=\gamma_i r_1(x)\quad \text{for all } x \text{ and } i=2,\ldots,n, \tag{1.1}$$

or equivalently, if the survival functions satisfy the relations

$$\bar{F}_i(\cdot)=\bar{F}_1^{\gamma_i}(\cdot)\quad \text{for all } x \text{ and } i=2,\ldots,n. \tag{1.2}$$

X_i's are said to belong to the proportional hazard family if they satisfy (1.1). This family of distributions has many interesting properties. The following result due to Armitage (1959), Allen (1963) and Sethuraman (1965) is very useful in the theory of competing risks.

Theorem 1.1. X_i's *belong to the proportional hazard family* (1.1) *if and only if* $X_{(1)}$ *and* $I_{(1)}$ *are independent*. □

In the context of competing risks theory, Theorem 1.1 has the following interpretation. If the n independent risks are acting simultaneously on a subject or a system in an effort to fail it then the time to

Correspondence to: Subhash C. Kochar, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA.

failure is independent of the cause of failure if and only if their lifetimes belong to the proportional hazard family.

In many applications, it is of interest to know whether the n different risks are equally fatal, that is, whether the n random variables X_i's are identically distributed. The following result gives a necessary and sufficient condition for the homogeneity of X_i's in the proportional hazard family.

Theorem 1.2. *Let $X_1, \ldots, X_n$ be independent random variables with proportional hazard rates. Then $X_1, \ldots, X_n$ are identically distributed if and only if $X_{(r)}$ and $I_{(r)}$ are independent for some $r \in \{2, \ldots, n\}$.*

To prove the main theorem we shall need the following lemma.

Lemma 1.3. *Let $c_1, \ldots, c_n$ be real numbers and let $0 < d_1 < \cdots < d_n$. Suppose*

$$\sum_{i=1}^{n} c_i u^{d_i} = 0, \quad 0 \leqslant u \leqslant 1.$$

Then $c_i = 0$, $i = 1, 2, \ldots, n$.

Proof. Let $0 < u_1 < \cdots < u_n < 1$. Then

$$\sum_{i=1}^{n} c_i u_j^{d_i} = 0, \quad j = 1, 2, \ldots, n.$$

Thus

$$\begin{bmatrix} u_1^{d_1} & \cdots & u_1^{d_n} \\ \vdots & \ddots & \vdots \\ u_n^{d_1} & \cdots & u_n^{d_n} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{1.3}$$

The $n \times n$ matrix appearing on the left hand side of (1.3) is nonsingular (see, for example, Polya and Szego, 1976, p. 46). Therefore $c_i = 0$, $i = 1, 2, \ldots, n$, and the proof is complete. □

Proof of Theorem 1.2. It is obvious that if $X_1, \ldots, X_n$ are identically distributed then $X_{(r)}$ and $I_{(r)}$ are independent for any $r \in \{1, \ldots, n\}$.

To prove the converse, suppose that $X_{(r)}$ is independent of $I_{(r)}$ for some $r \in \{2, \ldots, \mathrm{n}\}$. Then there exist positive constants $\beta_1 = 1, \beta_2, \ldots, \beta_n$ such that

$$P\left[I_{(r)} = i,\ X_{(r)} \leqslant t\right] = \beta_i P\left[I_{(r)} = 1,\ X_{(r)} \leqslant t\right] \quad \text{for all } t,\ i = 1, 2, \ldots, n. \tag{1.4}$$

Let $N_i = \{1, 2, \ldots, n\} \setminus \{i\}$. Differentiating (1.4) with respect to t and after some simple calculations it follows that for any $t \in (-\infty, \infty)$,

$$\frac{1}{\beta_i} f_i(t) \sum_{S \subset N_i, |S| = r-1} \prod_{j \in S} F_j(t) \prod_{j \in N_i \setminus S} \bar{F}_j(t) \tag{1.5}$$

is constant in i for $i = 1, 2, \ldots, n$.

From (1.2) and (1.5) we see that for any $t \in (-\infty, \infty)$,

$$\frac{\gamma_i}{\beta_i} \bar{F}_1^{\gamma_i}(t) \sum_{S \subset N_i, |S| = r-1} \prod_{j \in S} \left(1 - \bar{F}_1^{\gamma_j}(t)\right) \prod_{j \in N_i \setminus S} \bar{F}_1^{\gamma_j}(t) \tag{1.6}$$

is constant in i for $i = 1, 2, \ldots, n$.

Let $u=\bar{F}_1(t)$. Then the coefficient of $u^{\gamma_1+\cdots+\gamma_n}$ in (1.6) is

$$\frac{\gamma_i}{\beta_i}\binom{n-1}{n-r}(-1)^{r-1}. \tag{1.7}$$

By Lemma 1.3, (1.7) must be constant in i for $i=1,2,\ldots,n$ and hence $\gamma_i/\beta_i=\gamma_j/\beta_j$, $i\neq j$.

Now divide the expression in (1.6) by $(\gamma_1/\beta_1)\prod_{j=1}^{n}\bar{F}_1^{\gamma_j}(t)$ to conclude that for any $t\in(-\infty,\infty)$ which satisfies $0<F_1(t)<1$,

$$\sum_{S\subset N_i,|S|=r-1}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right) \tag{1.8}$$

is constant in i for $i=1,2,\ldots,n$.

Equating (1.8) for $i=1,2$,

$$\sum_{S\subset N_1,|S|=r-1}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right)=\sum_{S\subset N_2,|S|=r-1}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right). \tag{1.9}$$

If $S\subset N_1\cap N_2$, then the corresponding term occurs on either side of (1.9) and hence can be cancelled. Thus

$$\sum_{2\in S\subset N_1,|S|=r-1}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right)=\sum_{1\in S\subset N_2,|S|=r-1}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right).$$

This leads to

$$\begin{aligned}&\left(\bar{F}_1^{-\gamma_2}(t)-1\right)\sum_{S\subset N_1\cap N_2,|S|=r-2}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right)\\&\quad=\left(\bar{F}_1^{-\gamma_1}(t)-1\right)\sum_{S\subset N_1\cap N_2,|S|=r-2}\prod_{j\in S}\left(\bar{F}_1^{-\gamma_j}(t)-1\right)\end{aligned}$$

and therefore $(\bar{F}_1^{-\gamma_2}(t)-1)=(\bar{F}^{-\gamma_1}(t)-1)$. It follows that X_1, X_2 are identically distributed. Similarly it can be shown that X_1 and X_i are identically distributed and the proof is complete. □

A similar result can be obtained for the dual family of distributions satisfying

$$F_i(x)=F_1^{\alpha_i}(\alpha),\quad i=2,\ldots,n, \tag{1.10}$$

where α_i's are positive constants. It can be seen that (1.10) holds if and only if

$$\tilde{r}_i(x)=\alpha_i\tilde{r}_1(x)\quad\text{for all } x,\ i=2,\ldots,n, \tag{1.11}$$

where $\tilde{r}_i=f_i/F_i$ denotes the 'survival rate' of X_i. It has the following interpretation. Given that a unit with survival rate $\tilde{r}_i$ has failed at time t, the probability that it survived upto time $t-\delta$ is approximately equal to $\delta\tilde{r}_i(t)$. We call the family of distributions satisfying (1.10) as the 'proportional survival rate' family. It is well known that $X_{(n)}$ and $I_{(n)}$ are independent if and only if $X_1,\ldots,X_n$ belong to the proportional survival rate family. We have the following result whose proof is similar to that of Theorem 1.2.

Theorem 1.4. *Let $X_1,\ldots,X_n$ be independent random variables with proportional survival rates. Then $X_1,\ldots,X_n$ are identically distributed if and only if $X_{(r)}$ and $I_{(r)}$ are independent for some $r\in\{1,\ldots,n-1\}$.* □

If X_i's are independent and identically distributed then it is immediate that $X_{(r)}$ is independent of $I_{(r)}$, for any $r\in\{1,\ldots,n\}$. Theorem 1.3 and Theorem 1.4 prove that the independence of $X_{(r)}$ and $I_{(r)}$,

$r \in \{1, n\}$ and that of $X_{(s)}$ and $I_{(s)}$, $s \in \{1, \ldots, n\} \backslash r$ imply the independence of all other pairs $X_{(j)}$ and $I_{(j)}$.

The case of dependent variables

Now we consider the case of dependent variables. Let $F(t_1, \ldots, t_n)$ and $\bar{F}(t_1, \ldots, t_n)$ denote the joint c.d.f. and the survival function of $(X_1, \ldots, X_n)$. Kochar and Proschan (1991) have generalized Theorem 1.1 as follows:

Theorem 1.5. (a) *$X_{(1)}$ and $I_{(1)}$ are independent if and only if there exist positive constants $\gamma_2, \ldots, \gamma_n$ such that*

$$g_i(t) = \gamma_i g_1(t) \quad \text{for all } t,\ i = 2, \ldots, n, \tag{1.12}$$

where

$$g_i(t) = -\left(\frac{(\partial/\partial t_i)\bar{F}(t_1, \ldots, t_n)}{\bar{F}(t_1, \ldots, t_n)} \right)_{t_j = t}, \quad j = 1, 2, \ldots, n, \tag{1.13}$$

is the cause specific hazard rate corresponding to the ith cause.

(b) *$X_{(n)}$ and $I_{(n)}$ are independent if and only if there exist positive constants $\gamma_2^*, \ldots, \gamma_n^*$ such that*

$$g_i^*(t) = \gamma_i^* g_1^*(t) \quad \text{for all } t,\ i = 1, 2, \ldots, n, \tag{1.14}$$

where

$$g_i^*(t) = \left(\frac{(\partial/\partial t_i)F(t_1, \ldots, t_n)}{F(t_1, \ldots, t_n)} \right)_{t_j = t}, \quad j = 1, 2, \ldots, n. \quad \square \tag{1.15}$$

It can be seen that if the components of $X = (X_1, \ldots, X_n)$ are exchangeable, that is, $(X_1, \ldots, X_n) =^{\text{dist}} (X_{i1}, \ldots, X_{in})$ for all permutations $(i_1, \ldots, i_n)$ of $\{1, 2, \ldots, n\}$, then (1.12) and (1.14) obviously hold implying thereby the independence of $X_{(1)}$ and $I_{(1)}$ and that of $X_{(n)}$ and $I_{(n)}$. The following counter example shows that the converse analogous to Theorem 1.2 does not hold for the dependent case.

Example 1.1. For $a > 0$, let the joint density of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } 0 < x_1 < x_2 < a, \\ 1/\alpha & \text{if } 0 < x_2 < x_1 < a, \\ \text{arbitrary} & \text{if } u = v, \end{cases}$$

where a and α are related by

$$a = \sqrt{2\alpha/(1+\alpha)}$$

It is easy to see that (1.12) and (1.14) hold implying, respectively, the independence of $X_{(1)}$ and $I_{(1)}$ and that of $X_{(n)}$ and $I_{(n)}$. However, X_1 and X_2 are not exchangeable.

2. A characterization in terms of order statistics from sub-samples of size $n - 1$

We assume throughout this section that $X_1, \ldots, X_n$ are independent random variables of continuous type with common support T. We also keep r fixed, $1 \leqslant r \leqslant n - 1$, throughout the subsequent discussion.

Denote by $X_{(r)}^{(i)}$ the rth order statistic corresponding to the random variabels $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ and let $F_{(r)}^{(i)}$ denote the distribution function of $X_{(r)}^{(i)}$. The main result of this section is the following.

Theorem 2.1. *The random variables $X_1, \ldots, X_n$ are identically distributed if and only if the n random variables $X_{(r)}^{(1)}, \ldots, X_{(r)}^{(n)}$ have the same distribution.*

We will need a preliminary result. The following notation will be used. Recall that $N_i = \{1, \ldots, n\} \setminus \{i\}$. Let N_{ij} denote $\{1, \ldots, n\} \setminus \{i, j\}$ and let $\eta_k(x)$ denote the probability that exactly k of the $n-2$ random variables $X_{i_1}, \ldots, X_{i_{n-2}}$ are less than x where $\{i_1, \ldots, i_{n-2}\} = N_{ij}$. We set $\eta_0(x) \equiv 1$.

Lemma 2.2. *Let $i, j \in \{1, \ldots, n\}$, $i \neq j$. Then*

$$F_{(r)}^{(i)}(x) - F_{(r)}^{(j)}(x) = \eta_{r-1}(x)\left[F_j(x) - F_i(x)\right].$$

Proof. We have

$$\begin{aligned} F_{(r)}^{(i)}(x) &= \sum_{k=r}^{n-1} \sum_{S \subset N_i, |S|=k} \prod_{l \in S} F_l(x) \prod_{l \in N_i \setminus S} \bar{F}_l(x) \\ &= \sum_{k=r}^{n-2} \left\{ F_j(x) \sum_{S \subset N_{ij}, |S|=k-1} \prod_{t \in S} F_l(x) \prod_{t \in N_{ij} \setminus S} \bar{F}_l(x) \right. \\ &\qquad \left. + \bar{F}_j(x) \sum_{S \subset N_{ij}, |S|=k} \prod_{l \in S} F_l(x) \prod_{l \in N_{ij} \setminus S} \bar{F}_l(x) \right\} + \prod_{l \neq i} F_l(x) \\ &= \sum_{k=r}^{n-2} \left[F_j(x)\eta_{k-1}(x) + \bar{F}_j(x)\eta_k(x) \right] + F_j(x)\eta_{n-2}(x). \end{aligned} \tag{2.1}$$

Similarly

$$F_{(r)}^{(j)}(x) = \sum_{k=r}^{n-2} \left\{ F_i(x)\eta_{k-1}(x) + \bar{F}_i(x)\eta_k(x) \right\} + F_i(x)\eta_{n-2}(x). \tag{2.2}$$

Subtracting (2.2) from (2.1) we get

$$\begin{aligned} F_{(r)}^{(i)}(x) - F_{(r)}^{(j)}(x) &= \left(F_j(x) - F_i(x)\right) \left\{ \sum_{k=r}^{n-2} \left(\eta_{k-1}(x) - \eta_k(x)\right) + \eta_{n-2}(x) \right\} \\ &= \eta_{r-1}(x)\left(F_j(x) - F_i(x)\right), \end{aligned}$$

and the proof is complete. □

Proof of Theorem 2.1. The "only if" part is obvious. To prove the "if" part, let $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$. If $x \in T$, then $\eta_{r-1}(x) > 0$ and since $F_{(r)}^{(i)}(x) = F_{(r)}^{(j)}(x)$, we conclude from Lemma 2.2 that $F_i(x) = F_j(x)$. Similarly we can prove that $F_1(x) = \cdots = F_n(x)$ for all $x \in T$. □

Remarks. 1. We note the following consequence of Lemma 2.2. Let $i, j \in \{1, \ldots, n\}$, $i \neq j$. Then

$$F_{(r)}^{(i)}(x) - F_{(r)}^{(j)}(x) \gtreqless 0 \text{ according as } F_j(x) - F_i(x) \gtreqless 0.$$

In particular $X_i \geqslant^{\text{st}} X_j$, if and only if $X_{(r)}^{(i)} \leqslant^{\text{st}} X_{(r)}^{(j)}$ where the superscript 'st' denotes stochastic ordering.

2. As shown by Bapat and Beg (1989), for fixed x and n the η_k's form a log-concave sequence.

3. It follows clearly from Lemma 2.2 that $E[X_{(r)}^{(i)}] - E[X_{(r)}^{(j)}] \geqslant E(X_j) - E(X_i)$. A similar inequality holds for other moments.

Acknowledgement

Thanks are due to Prof. B. Ramachandran for helpful discussions on the problem and for providing Example 1.1.

References

Allen, W.R. (1963), A note on the conditional probability of failure when the hazards are proportional, *Oper. Res.* **11**, 658–659.

Armitage, P. (1959), The comparison of survivial curves, *J. Roy. Statist. Soc. Ser. A* **122**, 279–300.

Bapat, R.B. and M.I. Beg (1989), Order statistics for nonidentically distributed variables and permanents, *Sankhyā Ser. A* **51**, 78–93.

Kochar, S.C. and F. Proschan (1991), Independence of time and cause of failure in the multiple dependent competing risks model, *Statist. Sinica* **1**, 295–299.

Polya, G. and G. Szego (1976), *Problems and Theorems in Analysis, Vol. II* (Springer, Berlin).

Sethuraman, J. (1965), On a characterization of the three limiting types of the extreme, *Sankhyā Ser. A* **27**, 357–364.