COMMUN. STATIST.-THEOR. METH., A9(14), 1515-1525 (1980)

IDENTIFIABILITY OF DISTRIBUTIONS UNDER COMPETING RISKS AND COMPLEMENTARY RISKS MODEL

A. P. Basu and J. K. Ghosh

University of Missouri-Columbia and Indian Statistical Institute

Key Words and Phrases: competing risks; complementary risks; Identifiability; reliability; biometry; distribution of minimum and maximum; series and parallel system.

ABSTRACT

Let $X_1, X_2, \ldots, X_p$ be p random variables with cdf's $F_1(x), F_2(x), \ldots, F_p(x)$ respectively. Let $U = \min(X_1, X_2, \ldots, X_p)$ and $V = \max(X_1, X_2, \ldots, X_p)$. In this paper we study the problem of uniquely determining and estimating the marginal distributions $F_1, F_2, \ldots, F_p$ given the distribution of U or of V.

First the problem of competing and complementary risks are introduced with examples and the corresponding identification problems are considered when the X_i's are independently distributed and U(V) is identified, as well as the case when U(V) is not identified. The case when the X_i's are dependent is considered next. Finally the problem of estimation is considered.

1. INTRODUCTION

Let $X_1, \ldots, X_p$ be p random variables with cdf's $F_1(x), \ldots, F_p(x)$ respectively. Let $U = \min(X_1, \ldots, X_p)$ and $V = \max(X_1, \ldots, X_p)$. We shall use the following definitions. Let I be an integer valued random variable ($I = 1,2,\ldots,p$). (U,I) is called an *identified minimum* if $I = k$ when $U = \min(X_1,X_2,\ldots,X_p) = X_k$. That is, we observe the minimum U and know which coordiante of $X = (X_1,X_2,\ldots,X_p)$ is the minimum. Similarly, (V,I) is called an identified maximum if $I = k$ when $V = X_k$. In many physical situations we are interested in uniquely determining the marginal distributions $F_1,\ldots,F_p$ given the distribution of U, V, (U,I), or of (V,I).

The first problem, that is the problem of determining the cdf of X_i's from that of U is called the problem of competing risks and arises in a number of different contexts, as given below:

(a) An individual may be subject to p causes of death $C_1,C_2,\ldots,C_p$. Let X_i represent the lifetime of an individual exposed to cause C_i alone ($i = 1,2,\ldots,p$). However, X_i's are not observable and we need to infer about the F_i's based on observations on U.

(b) Consider a p component system where the components with lifetimes $X_1,\ldots,X_p$ are connected in series. Then only the system lifetime $U = \min(X_1,\ldots,X_p)$ will be observable.

(c) Let X_1 be the amount demanded and X_2 be the amount available for supply for an elastic good at a given price p. Then the amount actually transacted in the market will be $U = \min(X_1,X_2)$.

For a bibliography of the literature in this area, see David and Moeschberger (1978). Associated with the problems of competing risks is a dual problem of *complementary risks* where we would like to identify the cdf of X_i from that of $V = \max(X_1,\ldots X_p)$.

The problem of complementary risks also arises naturally in a number of physical situations as can be seen from the following examples.

(d) Consider a p-component system where the components are connected in parallel. Here system lifetime is given by $V = \max(X_1,X_2,\ldots,X_p)$ which is observable. Although in many situations each of the X_i will be known when V is known, there are situations where only V will be observable. For example, during the flight of a twin engine plane or a space satellite, individual components (engines) are not easy to monitor. However, system lifetime is readily available.

(c) Consider the failure of internal body organs like kidneys or similar organs which are duplicated. Here exact time of failure of a single organ may not be known, but that of the second organ to fail (a fatal incident) will be known.

It will be seen that theories of competing and complementary risks are closely related and results in one area lead to or suggest similar results in the other area. In section 2 we consider the case when the components are independently distributed and U and/or V are identified. The case when X_i's are independent and U(V) is not identified is also considered in this section. The case of dependent X_i's is discussed in section 3. Finally, in section 4 the problem of estimation is considered.

2. INDEPENDENT RANDOM VARIABLES

Let $X_1,\dots,X_p$ be independently distributed. If the X_i's are iid with common cdf $F_X(x)$ then it can be obtained trivially from the cdf $F_U(x)$ of U, or $F_V(x)$ of V.

Next assume X_i's are independent but not identically distributed. Berman (1963) has shown that the distribution of the identified minimum (U,I) uniquely determines that of the X_i's.

Since $\mathrm{Max}(X_1,\dots,X_p) = -\mathrm{Min}(-X_1,\dots,-X_p)$, it follows that the distribution of the identified maximum (V,I) also uniquely determines that of the X_i's as the identification problem can be restated in terms of the minimum.

In case the extremum (U or V) is not identified, one can still uniquely determine $F_k(x)$ under certain conditions. To this end we consider the following theorem.

Theorem 1. Let F be a family of pdf on R_1 with support (a,b) which are continuous and are positive to the left of some point A and *such that* if f and g are any two distinct members of F then

$$\lim_{x \to a} (f(x)/g(x))$$

exists and equals either 0 or ∞. Let $X_1,\ldots,X_p$ be independent random variables with respective pdf's $f_1,f_2,\ldots,f_p$ in F and $Y_1,Y_2,\ldots,Y_q$ be independent random variables with respective pdf's belonging to F. If $\min(X_1,\ldots,X_p)$ and $\min(Y_1,\ldots,Y_q)$ have identical distributions, then $p = q$ and there exists a permutation $(k_1,k_2,\ldots,k_p)$ of $(1,2,\ldots,p)$ such that the pdf of Y_i is f_{k_i} $(i = 1,2,\ldots,p)$.

The proof is similar to the proof of a similar theorem for maximum given by Anderson and Ghurye (1977).

As an application of the above theorem, one can prove the identifiability of (univariate) normal distributions.

On the other hand if F is the family of negative exponential distributions

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad x \geq 0 ,$$

then the conditions of the above theorem are not met and, in fact, the distributions are not identifiable.

Note, however, if the maximum is observed both normal and exponential distributions are identifiable.

However, there are situations of interest in reliability theory, where the conditions of the above theorem are not met, yet the underlying family of distributions are identifiable. We consider below the family of gamma and Weibull distributions, each of which contains the exponential distribution as a special case.

Consider, first, the case of the gamma distributions.

<u>Theorem 2.</u> Let the pdf of X_i be given by

$$f_i(x) \equiv f(x;\alpha_i,\beta_i) = \frac{e^{-x/\beta_i}\, x^{\alpha_i - 1}}{\beta^{\alpha_i}\,\Gamma(\alpha_i)}, \quad \alpha_i > 0,\ \beta_i > 0 \quad (i = 1,2,3,4) \tag{1}$$

where α_1 and α_2 are not *both equal to one*. Let X_1 and X_2 be independent random variables. Similarly, let X_3 and X_4 be independent. If the distribution of $\min(X_1,X_2)$ is identical with that of $\min(X_3,X_4)$, then either

$$(\alpha_1,\alpha_2) = (\alpha_3,\alpha_4) \quad \text{and} \quad (\beta_1,\beta_2) = (\beta_3,\beta_4)$$

or, (2)

$$(\alpha_1,\alpha_2) = (\alpha_4,\alpha_3) \quad \text{and} \quad (\beta_1,\beta_2) = (\beta_4,\beta_3)$$

Proof. To prove the theorem we need the following.

Lemma 1.

$$J = \int_x^\infty e^{-y}\, y^{\alpha-1}\, dy \sim e^{-x}\, x^{\alpha-1} \quad \text{as} \quad x\to\infty, \alpha > 0 . \tag{3}$$

Proof of Lemma 1. Suppose first $0 < \alpha \le 1$. Then integrating by parts

$$J = e^{-x}\, x^{\alpha-1} + (\alpha-1)\int_x^\infty e^{-y}\, y^{\alpha-2}\, dy \le e^{-x}\, x^{\alpha-1}. \tag{4}$$

Similarly,

$$\begin{aligned} J &= e^{-x}\, x^{\alpha-1} + (\alpha-1)e^{-x}\, x^{\alpha-2} + (\alpha-1)(\alpha-2)\int_x^\infty e^{-y}\, y^{\alpha-3}\, dy \\ &\ge e^{-x}\, x^{\alpha-1} + (\alpha-1)e^{-x}\, x^{\alpha-2} \\ &= e^{-x}\, x^{\alpha-1}\left[1 + \frac{\alpha-1}{x}\right]. \end{aligned} \tag{5}$$

The lemma follows from (4) and (5). Now let $m < \alpha \le m+1$, where m is a positive integer. Then by repeated integration by parts,

$$J = e^{-x}\, x^{\alpha-1} + (\alpha-1)e^{-x}\, x^{\alpha-2} + \ldots + (\alpha-1)\ldots(\alpha-m)\int_x^\infty e^{-y}\, y^{(\alpha-m)}dy.$$

Since $0 < (\alpha-m) \le 1$, by applying the first part of the proof to the integral on the right hand side.

$$J \sim e^{-x}\, x^{\alpha-1} \quad \text{for large } x. \tag{6}$$

We now prove Theorem 2. Let F_i be the cdf of X_i and $\overline{F}_i = 1 - F_i$, $i = 1,2,3,4$. By the conditions given

$$\overline{F}_1(x)\, \overline{F}_2(x) = \overline{F}_3(x)\, \overline{F}_4(x) \quad \text{for all } x. \tag{7}$$

Differentiating the logarithm of both sides with respect to x,

$$\frac{f_1(x)}{\overline{F}_1(x)} + \frac{f_2(x)}{\overline{F}_2(x)} = \frac{f_3(x)}{\overline{F}_3(x)} + \frac{f_4(x)}{\overline{F}_4(x)} \quad \text{for all } x. \tag{8}$$

Making $x \to \infty$ in (7) and using Lemma 1, we have

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = \frac{1}{\beta_3} + \frac{1}{\beta_4}, \tag{9}$$

$$\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4, \tag{10}$$

and

$$\beta_1^{\alpha_1 - 1} \beta_2^{\alpha_2 - 1} \Gamma(\alpha_1)\Gamma(\alpha_2) = \beta_3^{\alpha_3 - 1} \beta_4^{\alpha_3 - 1} \Gamma(\alpha_3)\Gamma(\alpha_4). \tag{11}$$

Also note,

$$\lim_{x \to 0} \overline{F}_j(x) = 1, \tag{12}$$

and

$$\lim_{x \to 0} \frac{f_i(x)}{f_j(x)} = \begin{matrix} 0, \text{ if } \alpha_i > \alpha_j \\ \infty, \text{ if } \alpha_i < \alpha_j \\ cij, \text{ if } \alpha_i = \alpha_j \end{matrix} \tag{13}$$

where

$$c_{ij} = (\beta_j/\beta_i)^{\alpha i}.$$

Now consider the case when $\alpha_1 < \alpha_2$. Dividing (8) throughout by $f_1(x)$ and making $x \to 0$ we note, in view of (12) and (13), that the left hand side of (8) tends to 1. So the right hand side of (8) must also tend to one. By (13) this implies at least one of α_3, α_4 is equal to α_1. Without any loss of generality, let $\alpha_3 = \alpha_1$. Then, by (10) $\alpha_2 = \alpha_4 > \alpha_1$. Also, by (13), we have $c_{31} = 1$. Since $\alpha_1 = \alpha_3$ we must have $\beta_1 = \beta_3$. Hence, $\overline{F}_1(x) = \overline{F}_3(x)$ and hence, from (7), $\overline{F}_2(x) = \overline{F}_4(x)$. This proves the theorem when $\alpha_1 < \alpha_2$. The case $\alpha_1 > \alpha_2$ can be treated similarly.

Now consider the case when $\alpha_1 = \alpha_2$, but the common value is not equal to 1. As before if we divide (8) throughout by $f_1(x)$ and take limits as $x \to 0$, the left hand side tends to a constant; by (13) this implies at least one of α_3, α_4, say α_3 equals α_1. Then α_4 equals α_2 by (10). Thus

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha, \text{ say.}$$

Hence, by (11)

$$(\beta_1\beta_2)^{\alpha-1} = (\beta_3\beta_4)^{\alpha-1}, \text{ where } \alpha \neq 1.$$

That is, $\beta_1\beta_2 = \beta_3\beta_4$. This and (9) imply one of β_3, β_4 equals β_1 and the other equals β_2. This completes the proof.

Let $X_i \sim W(p_i, \theta_i)$. That is let X_i follow the Weibull distribution with density function

$$f_i(x) = \frac{p_i}{\theta_i} x^{p_i - 1} e^{-x^{p_i}/\theta_i} \qquad x > 0,\ (\theta_i, p_i > 0) .$$

Here

$$\bar{F}_i(x) = 1 - F_i(x) = e^{-x^{p_i}/\theta_i},\quad f_i(x)/\bar{F}_i(x) = \frac{p_i}{\theta_i} x^{p_i - 1} ,$$

and as $x \to 0$,

$$\frac{f_i(x)}{f_j(x)} = \begin{cases} \theta_j/\theta_i, & p_i = p_j \\ 0, & p_i > p_j \\ \infty, & p_i < p_j \end{cases} \tag{14}$$

<u>Theorem 3.</u> Let $X_i \sim W(p_i, \theta_i)$, $(i = 1,2,3,4)$ be independent Weibull random variables. If the distribution of $\min(X_1, X_2)$ is the same as that of $\min(X_3, X_4)$, then either

$$(p_1, \theta_1) = (p_3, \theta_3), \quad (p_2, \theta_2) = (p_4, \theta_4) ,$$

or

$$(p_1, \theta_1) = (p_4, \theta_4) \text{ and } (p_2, \theta_2) = (p_3, \theta_3)$$

provided $p_1 \neq p_2$.

<u>Proof.</u> Since $\bar{F}_1(x)\,\bar{F}_2(x) = \bar{F}_3(x)\,\bar{F}_4(x)$ for all x, we have, taking logarithm of both sides,

$$x^{p_1}/\theta_1 + x^{p_2}/\theta_2 = x^{p_3}/\theta_3 + x^{p_4}/\theta_4 \quad \text{for all } x \tag{15}$$

Since $p_1 \neq p_2$, suppose without loss of generality $p_1 < p_2$. Considerin the behavior of (15) as $x \to 0$ and $x \to \infty$, we get $p_1 = \min(p_3, p_4) = p_3$, say, and $p_2 = \max(p_3, p_4) = p_4$, say. From the linear independence of powers of x, we now get from (15), $\theta_1 = \theta_3$, $\theta_2 = \theta_4$. This completes the proof.

Remark. Let $p_1 = p_2 = p$, say. Then by transforming to $Y_i = X_i^p$, one sees from the case of exponential distributions that the scale parameters θ_1, θ_2 are not identifiable. In fact this is also clear from (15).

3. CASE OF DEPENDENT RANDOM VARIABLES

In this section we consider the case when the X_i's are dependent with joint cdf F. Basu and Ghosh (1978) have shown the difficulty in identification based on U or (U,I) unless the class of distributions is restricted to a specific parametric family. Results for some special cases are described below.

Basu and Ghosh (1978) have considered identifiability of a number of bivariate families of distributions useful as models in life testing based on the distribution of U or that of (U,I). These include the bivariate exponential distributions of Marshall and Olkin (1967), Block and Basu (1974), and Gumbel (1960). In particular, they showed that parameters of Marshall-Olkin distribution are identifiable given the distribution of the identified minimum (U,I). However, the parameters are not identifiable if only U is observable. For the Block-Basu (1974) and the Freund distribution (1961) the parameters are not at all identifiable, even when the identified minimum is available. Similar results for the identifiability of parameters based on (U,I) have been obtained for the two bivariate exponential models proposed by Gumbel. However, if only U is observable, none of the above models, except the following is identifiable.

Consider the following bivariate distribution of Gumbel:

$$F(x,y) = (1 - e^{-\lambda_1 x})\,(1 - e^{-\lambda_2 y})\,(1 + \lambda_{12}\, e^{-\lambda_1 x-\lambda_2 y})\,. \quad (16)$$

<u>Theorem 4</u>. If U is observable, λ_{12} is identifiable and (λ_1,λ_2) is identifiable up to a permutation. That is, if (X_1,X_2) and (X_1',X_2') follow the bivariate exponential (16) with parameters $(\lambda_1,\lambda_2,\lambda_{12})$ and $(\lambda_1',\lambda_2',\lambda_{12}')$ respectively, and if $\min(X_1,X_2)$ and $\min(X_1',X_2')$

have the same distribution then either $(\lambda_1, \lambda_2, \lambda_{12}) = (\lambda_1', \lambda_2', \lambda_{12}')$ or $(\lambda_1, \lambda_2, \lambda_{12}) = (\lambda_2', \lambda_1', \lambda_{12}')$.

We omit the proof.

Next assume (X_1, X_2) follow the bivariate Weibull distribution defined by

$$\overline{F}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2) = \exp -\{\lambda_1 x_1^{p_1} + \lambda_2 x_2^{p_2} + \lambda_{12} \max(x_1^{p_1}, x_2^{p_2})\}.$$

By similar arguments one can show that if U is observable p_1 and p_2 are identifiable up to permutation. However, λ_1 and $\lambda_2 + \lambda_{12}$ or λ_2 and $\lambda_1 + \lambda_{12}$ are identifiable.

Because of the inherent physical interpretation based on the minimum, the Marshall-Olkin and the Block-Basu model are not suitable models for the complementary risks problem. However Gumbel's bivariate exponential distributions are still suitable as models.

Consider the first model with cdf

$$F(x,y) = 1 - e^{-\lambda_1 x} - e^{-\lambda_2 y} + e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} xy} \tag{17}$$

Unlike the case of minimum, we can prove the following general theorem.

Theorem 5. Let $V = \max(X_1, X_2)$ be observable, and (X_1, X_2) follow the bivariate exponential distribution with cdf (17). λ_{12} is identifiable and $(\lambda_1, \lambda_{12})$ is identifiable up to a permutation.

Proof. Let $\max(X_1, X_2)$ and $\max(X_1', X_2')$ have cdf of the form (17) with parameters $(\lambda_1, \lambda_2, \lambda_{12})$ and $(\lambda_1', \lambda_2', \lambda_{12}')$ respectively. Without any loss of generality, assume $\min(\lambda_1, \lambda_2) = \lambda_1$, and $\min(\lambda_1', \lambda_2') = \lambda_1'$. Since $P(X_1 < x; X_2 < x) = P(X_1' < x, X_2' < x)$, we have, from (17),

$$e^{-\lambda_1 x} + e^{-\lambda_2 x} - e^{-(\lambda_1 + \lambda_2 + \lambda_{12} x)x} = e^{-\lambda_1' x} + e^{-\lambda_2' x} - e^{-(\lambda_1' + \lambda_2' + \lambda_{12}' x)x} \quad \text{for all } x . \tag{18}$$

Dividing both sides by $e^{-\lambda_1 x}$ and taking the limit as $x \to \infty$, the left side tends to 1. The right side must tend to 1 also. Since

$0 < \lambda_1' < \lambda_2' < \lambda_1' + \lambda_2' + \lambda_{12}'x$, we must have $\lambda_1 = \lambda_1'$. Eliminating equal terms from both sides of (18), and dividing by $e^{-\lambda_2 x}$, we obtain, using similar arguments

$$\lambda_2 = \lambda_2' .$$

And hence,

$$\lambda_{12} = \lambda_{12}' .$$

Identifiability of the bivariate normal distribution, using the observed minimum, has been considered by Basu and Ghosh (1978) and Nadàs (1971). If X follows the normal distribution, so does $-\underline{X}$ and hence the identification problem for the maximum can be restated in terms of the identification problem for the minimum. Also, any bivariate distribution obtained through strict monotone transformation of normal variables will be identifiable. The bivariate lognormal distribution is thus identifiable.

4. ESTIMATION OF PARAMETERS

Estimation of parameters based on (U,I), the identified minimum has been considered extensively (see David and Moeschberger, 1978). Basu and Ghosh (1978) have considered the problem of estimation of parameters based on U alone for the bivariate normal distribution.

As remarked before, the case of the estimation problem for the normal distribution based on V reduces to that of U. We shall therefore consider some of the other models.

The pdf of V, assuming independence, is given by $f_1(t)F_2(t) + f_2(t)F_1(t)$, so that the likelihood function can be readily obtained. Hence the likelihood equations can be written down easily and solved numerically.

In the case of Weibull distributions if the shape parameters are assumed to be equal, that is if $p_1 = p_2 = p$, the parameters can be estimated more readily using the methods of moments. The r^{th} raw moment of V is given by

$$E(V^r) = \mu_r' = \Gamma(1 + \frac{r}{p})(\theta_1^{\frac{r}{p}} + \theta_1^{\frac{r}{p}} - (\theta_1 + \theta_2)^{\frac{r}{p}}) . \qquad (19)$$

Method of maximum likelihood and method of moments can be used for the exponential and other distributions including the cases when the random variables are dependent. In particular, method of moments can be readily used for the two models proposed by Grumbel (1960).

ACKNOWLEDGMENT

This research was supported by the National Science Foundation Grant MCS78-04010 and by the Research Council of the Graduate School, University of Missouri-Columbia.

BIBLIOGRAPHY

Anderson, T. W. & Ghurye, S. G. (1977). Identification of parameters by the distribution of a maximum random variable. J. R. Statist. Soc. B. 39, 337-342.

Basu, A. P. & Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model. J. of Multivariate Analysis 8, 413-429.

Block, H. W. & Basu, A. P. (1974). A continuous bivariate exponential extension. J. Amer. Statist. Assoc. 69, 1031-1037.

Berman, S. M. (1963). Notes on extreme values, Competing risks, and semi-Markov processes. Ann. Math. Statist. 34, 1104-1106.

David, H. A. & Moeschberger, M. L. (1978). The Theory of Competing Risks. London: Charles Griffin and Co.

Freund, J. E. (1961). A bivariate extension of the exponential distribution. J. Amer. Statist. Assoc. 56, 971-977.

Gumbel, E. J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55, 698-707.

Marshall, A. W. & Olkin, I. (1967). A multivariate exponential distribution. J. Amer. Statist. Assoc. 62, 30-44.

Nadàs, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. Technometrics, 13, 201-202.

Received January, 1979; Revised November, 1979; Retyped March, 1980.

Refereed by N. L. Johnson, University of North Carolina, Chapel Hill, NC.