FREUND'S BIVARIATE EXPONENTIAL DISTRIBUTION AND CENSORING

Sue Leurgans and Wei-Yann Tsai

University of Wisconsin-Madison

and

John Crowley

University of Washington and Fred Hutchinson Cancer Research Ctr.

0. SUMMARY

In some problems, a bivariate random vector (T_1,T_2) with bivariate cumulative distribution function F is observed for each of n independent subjects, but the coordinates may be subject to censoring. In the first section, we describe several mechanisms which can generate the censorship. The usual nonparametric approaches to estimation of F are then shown to be unsatisfactory. Therefore, in the third section, we describe a parametric model due to Freund (1961). This model is studied not because all data can be forced to fit this specific parametric form, but because this model suggests some approaches to the nonparametric problem. These ideas, together with some relationships to the work of other authors, are outlined in the fourth section.

1. Bivariate Times and Censoring Patterns

In this section, we outline several mechanisms which can generate bivariate censored times. A distinction between univariate and bivariate censoring is developed.

Univariate censoring arises naturally in two similar contexts. Firstly, the experimental units may contain two similar components (such as ears, elbows, knees, kidneys or engines) whose survival is being studied. Alternatively, the experimental units may contain two dissimilar components whose survival is being studied. Unlike competing risk problems, neither component is essential for the survival of the experimental unit. In both cases, censoring occurs when the experimental unit is removed from observation before both components have been observed to fail. Examples include lifetimes of pumps and hoses on 15 tractors given in Barlow and Proschan (1977) and the times of first responses to treatment (as observed at one site, perhaps a head or a tumor) and times of first sign of toxicity or death, as discussed by Lagakos (1976).

In the examples above, all times for any experimental unit are measured on a single clock from a common origin. However, double clocks are natural when studying the times required from initiation of treatment until the first sign of response in two successive courses of treatment in the same patient (see Gross and Lam (1981)) or the lifetimes of two paired subjects (siblings or other kin). The time until response to treatment and the length of the subsequent disease-free interval also requires two separate clocks. Indeed, the random variables need not be times in the usual sense. Variables could be cumulative dose or cumulative cost. Censoring would occur when an experimental unit (or component) is removed from observation for reasons independent of both responses.

To model the censoring, independent censoring vectors (C_1,C_2) are postulated to exist for each bivariate vector (T_1,T_2). We suppose that the vectors (C_1,C_2) form a sample from a bivariate distribution G. While such an assumption will not always be valid, it permits censoring times to differ. The observed quantities are then $X_i = \min(T_i,C_i)$ and $D_i = [T_i \leq C_i]$ $(i=1,2)$. (The symbol $[A]$ denotes the indicator function of the event A.)

When a single clock governs both times, censoring occurs when an experimental unit is removed from observation. Since C_1 will always equal C_2 in this case, this censoring structure will be referred to as univariate censoring. When the censoring times can differ, the censoring will be called bivariate.

The distinction between univariate and bivariate censoring is clear when the observations (X_1,X_2) are plotted in R^2. If $D_i=0$, the i^{th} coordinate is censored and an arrow parallel to the i^{th} axis is drawn from (X_1,X_2). If $D_i=1$, the i^{th} coordinate is observed exactly, and the arrow is omitted. If the censoring is bivariate, the diagram can resemble Figure 1.

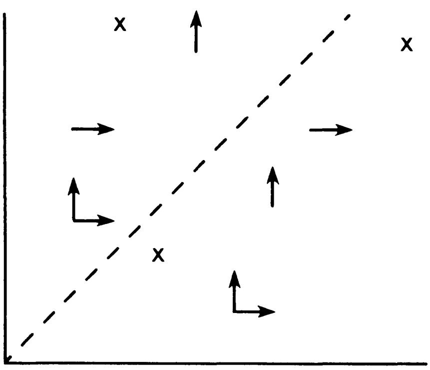

Figure 1: Schematic Diagram for Bivariate Censoring

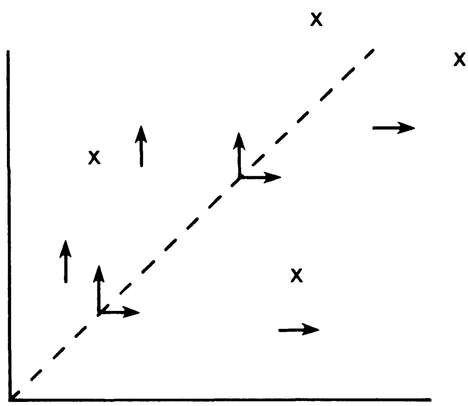

Figure 2: Schematic Diagram for Univariate Censoring

However, if the censoring is univariate, Figure 1 is impossible. In univariate censoring, if exactly one coordinate is censored, it must be the coordinate with the larger value. Consequently, if a point has one arrow attached, that arrow must point away from the diagonal lime $X_1 = X_2$. Furthermore, if both coordinates are censored, since the censoring variable must be the same for each component, the two coordinates must have the same value and all points with two arrows must be based on the diagonal. Figure 2 indicates a possible diagram when univariate censoring is present. Both of these situations are clearly different from competing risks problems, when at most one lifetime can be observed exactly on each experimental unit and observation of the surviving component is censored at the end of the first lifetime, resulting in a diagram similar to Figure 3:

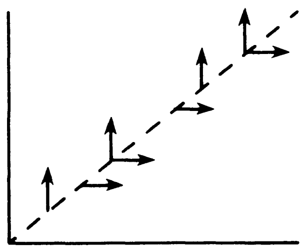

Figure 3: Schematic Diagram for Competing Risks with (Univariate) Censoring

2. Nonparametric Approaches

The principles of generalized maximum likelihood estimation and of self-consistency are often used to derive the product limit estimator of the cumulative distribution function of a single survival time in the presence of random censoring. In this section, these principles are shown to be inadequate for estimation of bivariate cumulative distribution functions.

In regular parametric settings, maximum likelihood estimators are well-known to have optimal asymptotic properties. The likelihood can be viewed as the Radon-Nikodym derivative of a parametrized probability measure with respect to a carrier measure. Since Radon-Nikodym derivatives can often be computed even when the "parameter" is not finite dimensional and a likelihood is not defined, Kiefer and Wolfowitz (1956) suggested a Generalized Maximum Likelihood Estimator (GMLE) for nonparametric problems. In parametric settings, the GMLE reduces to the usual maximum likelihood estimator. However, the generalized maximum likelihood principle is not known to guarantee any optimal properties, as occurs in the finite dimensional case. Johansen (1978) showed that the product limit estimator of Kaplan-Meier (1957) is the GMLE of F in the class of all CDF's.

Another property of the product limit estimator was established by Efron (1967), who named the property self-consistency. In the univariate problem, an estimator $\hat{F}$ is said to be self-consistent if

$$1 - \hat{F}(t) = \sum_{i=1}^{n} \frac{d_i[t_i \geq t]}{n} + \sum_{i=1}^{n} \frac{(1-d_i)}{n} \frac{1-\hat{F}(t)}{1-\hat{F}(t_i)} ,$$

or, the proportion estimated to survive past t is equal to the proportion of the subjects observed to survive past t plus the sum for all individuals censored before t of the estimated conditional probability of surviving past t, given survival to the censoring time. Efron showed that, up to possible indeterminacy for $t \geq t_{(n)}$, the only self-consistent estimator of the cumulative distribution function is the product limit estimator. Thus the GMLE is self-consistent.

In a 1980 Stanford Ph.D. dissertation, Muñoz studied nonparametric estimation of a bivariate distribution function in the presence of univariate censoring. He showed that the GMLE is self-consistent. He also showed that the GMLE is supported on three kinds of sets: points, rays and regions. The points of support are those (X_{1i},X_{2i}) with $(D_{1i},D_{2i}) = (1,1)$. The rays of support are sets $\{(x,y): x = X_{1i},\ y \geq X_{2i}\}$ with $(D_{1i},D_{2i}) = (1,0)$ or $\{(x,y): x \geq X_{1i},\ y = X_{2i}\}$ with $(D_{1i},D_{2i}) = (0,1)$. One region of support may exist: $\{(x,y): x \geq X_{1i},\ y \geq X_{2i}\}$ will be a region of support if $(D_{1i},D_{2i}) = (0,0)$ and the region contains no other points, rays or regions. Thus the support of the GMLE is the minimal set in which the true times corresponding to observed times must lie.

Muñoz showed that the mass of each set is determined, but the distribution of the mass within the set is completely arbitrary. Since, under random censorship, a non-negligible proportion of the observations will be censored in a single component, a non-negligible proportion of the mass is not located by the GMLE. Therefore there are self-consistent estimators of bivariate distribution functions which do not converge to the correct limit. The fact that self-consistency alone is inadequate is recognized in the calculations of Muñoz's example, despite a theorem which states that bivariate self-consistent estimators are asymptotically consistent. However, Campbell (1981) establishes that self-consistent estimators of discrete distributions are asymptotically consistent.

3. Freund's Model

Since the completely nonparametric approaches outlined above are unsatisfactory, a simple parametric model introduced by Freund (1961) will be described in this section. We will show that the resulting joint density for (T_1,T_2) is a tractable curved exponential family when univariate censoring is present. Subfamilies described by Block and Basu (1974) and by Lagakos (1976) are seen to be much less tractable under censoring.

3.1 Freund's Distribution

We suppose that the pair of times being studied can be recorded from a single clock. The experimental unit can be thought of as being under continuous observation, changing state whenever clocktime passes T_1 or T_2. If T_1 and T_2 are jointly absolutely continuous, the states and transitions possible at time t are indicated in Figure 4.

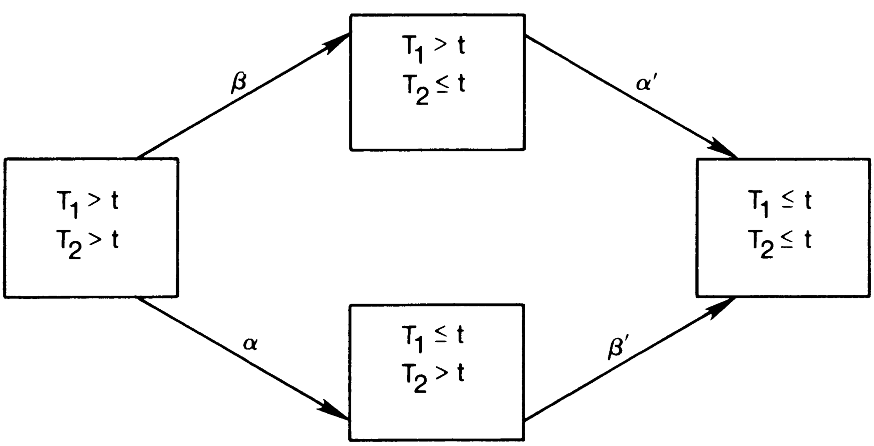

Figure 4: States and Transitions for Freund Model

If the Markov property is assumed, the times between transitions will have exponential distributions with the four positive parameters indicated in the diagram. Since the Markov property implies that the difference between the two times must be independent of the exact value of the smaller random variable, this model gives the following joint density for T_1 and T_2:

$$(1) \qquad f_{T_1,T_2}(t_1,t_2) = \begin{cases} \alpha e^{-(\alpha+\beta)t_1} \beta' e^{-\beta'(t_2-t_1)} & 0 < t_1 < t_2 \\ \beta e^{-(\alpha+\beta)t_2} \alpha' e^{-\alpha'(t_1-t_2)} & 0 < t_2 < t_1 \end{cases} .$$

This density was introduced by Freund (1961), who showed that the marginal distributions are not exponential. Freund also calculated the expectations, variances and covariance of T_1 and T_2. He showed that the correlation coefficient need not be non-negative, but can range from -1/3 to 1.

3.2 Inference

In this sub-section, we show that Freund's distribution is a curved exponential family under univariate censoring and derive the closed form maximum-likelihood estimators. Bivariate censoring causes the dimension of the statistic to be random (and stochastically increasing with n). At the end of this sub-section a simpler alternative to the maximum-likelihood estimator is suggested for bivariate censoring.

In the presence of univariate censoring with density g and distribution function G, the likelihood for Freund's model is

$$(2) \qquad e^{\eta^T(\alpha,\beta;\alpha',\beta')\, z_+(\underset{\sim}{x},\underset{\sim}{d})} \prod_{i=1}^{n} \left[(1-G(x_i)) \left(\frac{g(\underset{\sim}{x}_i)}{1-G(\underset{\sim}{x}_i)} \right)^{1-d_{1i}d_{2i}} \right]$$

where

$$\eta(\alpha,\beta;\alpha',\beta')=\begin{pmatrix} -(\alpha+\beta) \\ -\alpha' \\ -\beta' \\ \ell n(\frac{\alpha\beta'}{\alpha'\beta}) \\ \ell n(\alpha'/\alpha) \\ \ell n\ \alpha \\ \ell n\ \beta \end{pmatrix}, \quad z(\underset{\sim}{x},\underset{\sim}{d}) = \begin{pmatrix} \min(x_1,x_2) \\ (x_1-x_2)_+ \\ (x_2-x_1)_+ \\ [x_1 < x_2]d_1d_2 \\ d_1d_2 \\ d_1 \\ d_2 \end{pmatrix}$$

$z_+(\underset{\sim}{x},\underset{\sim}{d}) = \sum_{i=1}^{n} z(\underset{\sim}{x}_i,\underset{\sim}{d}_i)$, and $(x)_+ = x[x > 0]$. Because four of the seven coordinates of η are non-linear functions of the parameters, this is a four-parameter curved exponential family with a seven-dimensional sufficient statistic (see Efron (1978)). If no censoring is present, $G \equiv 0$ and the last three components of z_+ are all equal to n. Therefore $\eta^T(\alpha,\beta;\alpha',\beta')\ z_+(\underset{\sim}{x},\underset{\sim}{d})$ is an affine function of four sufficient statistics, and Freund's distribution is a regular exponential family in the absence of censoring. In either case, the theory of exponential families implies that the maximum-likelihood estimator is given by the solution of

$$\begin{pmatrix} -1 & 0 & 0 & 1/\alpha & -1/\alpha & 1/\alpha & 0 \\ -1 & 0 & 0 & -1/\beta & 0 & 0 & 1/\beta \\ 0 & -1 & 0 & -1/\alpha' & 1/\alpha' & 0 & 0 \\ 0 & 0 & -1 & 1/\beta' & 0 & 0 & 0 \end{pmatrix} \frac{\underset{\sim}{z}_+}{n} = \underset{\sim}{0} \quad .$$

If all components z are positive, the maximum-likelihood estimator is obtained by solving four one-parameter equations. Each of the resulting estimators is the ratio of a number of occurrences to a total exposure time:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} D_{1i}(1-D_{2i}[T_{1i} \geq T_{2i}])}{\sum_{i=1}^{n} \min(T_{1i},T_{2i})}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} D_{2i}(1-D_{1i}[T_{1i} < T_{2i}])}{\sum_{i=1}^{n} \min(T_{1i},T_{2i})}$$

(3)
$$\hat{\alpha}' = \frac{\sum_{i=1}^{n} D_{1i}D_{2i}[T_{1i} > T_{2i}]}{\sum_{i=1}^{n} (T_{1i} - T_{2i})_+}$$

$$\hat{\beta}' = \frac{\sum_{i=1}^{n} D_{1i}D_{2i}[T_{1i} < T_{2i}]}{\sum_{i=1}^{n} (T_{2i} - T_{1i})_+} \quad .$$

In the absence of censoring, these estimators reduce to those obtained by Freund (1961). It is clear that these estimators always take values inside the parameter space. Furthermore, the exponential family form implies that UMP tests are possible for all one-sided alternatives that can be specified in terms of a single linear transformation of the natural parameters. Thus the best test based on complete observations of the null hypothesis $\alpha' = \beta' = \alpha+\beta$ (stress-passing) against $\alpha' = \beta' > \alpha+\beta$ (increased stress) will not depend on $\alpha+\beta$. Clearly, fewer UMP tests exist in the presence of univariate censoring. The strong consistency and joint asymptotic normality of the estimators follow routinely from the strong law of large numbers and the central limit theorem applied to iid vectors Z.

If bivariate censoring is present, then the log-likelihood is not always linear in the parameters. To see this, note that the likelihood factor for terms with $D = (1,0)$ and $X_1 > X_2$ is

$$\int_{X_2} f_{X_1X_2}(x_1,x_2)dx_2 = \frac{\beta\alpha'}{\alpha+\beta-\alpha'} e^{-\alpha'(X_1-X_2)} e^{-(\alpha+\beta)X_2} + e^{-(\alpha+\beta)X_1} \frac{(\alpha-\alpha')(\alpha+\beta)}{\alpha+\beta-\alpha'} \quad .$$

This factor is not of exponential family form, but is a mixture of two exponential family densities, reflecting the fact that it may be unclear which parameters were acting on which experimental units. The product of terms of this type does not generate a sufficient statistic of fixed dimension. Since the sufficient statistic is more complex than that for univariate censoring or complete data, the solution of the maximum likelihood equations will generally be more difficult.

One way to simplify the estimation procedure can be thought of as modifying the observations to reflect the observations that would have been made if the censoring had been univariate. Formally, define

$$(X_1^*,X_2^*,D_1^*,D_2^*)^T = \begin{cases} \begin{pmatrix} \min(X_1,X_2) \\ \min(X_1,X_2) \\ 0 \\ 0 \end{pmatrix} & \text{if } \begin{cases} X_1 > X_2 \text{ and } D_1 = 0 \\ \text{or} \\ X_1 < X_2 \text{ and } D_2 = 0 \end{cases} \\ (X_1,X_2;D_1,D_2)^T & \text{otherwise} \end{cases} .$$

The estimators $\tilde{\alpha},\tilde{\beta},\tilde{\alpha}'$ and $\hat{\beta}'$ are obtained by applying the estimators (3) to the univariately censored $\{(X_{1i}^*,X_{2i}^*;D_{1i}^*,D_{2i}^*),\ 1 \leq i \leq n\}$. While it is clear that this approach throws away information and cannot always be efficient, this approach does provide closed form consistent estimators which are approximately normal and independent in large samples. The precise efficiency properties remain to be determined.

3.3 Subfamilies

Several sub-models of Freund's distribution have been proposed. Block and Basu (1974) point out that a three-parameter subfamily of Freund's distribution corresponds to the absolutely continuous component of the bivariate exponential distribution derived by Marshall and Olkin (1967). The three parameters are a linear function of the first three coordinates of η in (2) and correspond to the constraint that $\beta(\beta' - \beta) = \alpha(\alpha' - \alpha)$. The non-linearity of this function and the resulting curvature of the exponential family are reflected in the fact that the maximum-likelihood estimators for λ_1, λ_2 and λ_{12} do not have closed-form expressions, even for complete data. The sub-family characterization does imply that the maximum-likelihood equations based on complete data (univariate censoring) are determined by a four (seven) dimensional sufficient statistic. (See Section 7 of Block and Basu (1974) for complete data equations.)

Lagakos (1976) presents a three-parameter family for joint analysis of response time and survival time in cancer treatment studies. With a convention that response times are not observed after death, this family corresponds to Freund's family with the restriction that $\beta=\beta'$, a non-linear constraint on the natural parameters.

Since neither family exhibits a compelling superiority over the Freund family, we suggest considering the full family whenever either subfamily is fitted.

4. Extensions

One way to extend Freund's model to a nonparametric family is to allow the parameters to be functions of time, permitting α' and β' to depend on the first failure times. This yields the functions $\alpha(t)$, $\beta(t)$, $\alpha'(t|y)$ and $\beta'(t|y)$, corresponding to $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_{1|2}(t|y)$ and $\lambda_{2|1}(t|x)$ of Cox (1972). None of these functions correspond to hazard gradients. In his dissertation research, Mr. Tsai is investigating nonparametric estimators of these functions in the presence of censoring. When F is absolutely continuous, in order to obtain consistent estimators of $\alpha'(t|y)$ and $\beta'(t|x)$, some form of smoothing is required, since otherwise no more than one datum could be used to estimate each conditional function.

Other researchers have imposed additional structure. If the experimental units are assumed to have a non-stationary Markov property, $\alpha'(t|y) = \alpha'(t)$ and

$\beta'(t|x) = \beta'(t)$. Nonparametric tests for this model are described in Aalen, Borgan, Keiding and Thormann (1980) and in Voelkel (1980). In some cases, it is more reasonable to believe that when one component fails, the other component begins to age differently. The semi-Markov property requires instead that $\alpha'(t|y) = \alpha'(t-y)$ and $\beta'(t|x) = \beta'(t-x)$. Lagakos, Sommer and Zelen (1978) and Voelkel (1980) studied this model.

Freund's distribution has been extended to more than one time and to allow a positive probability that $T_1 = T_2$. For some such extensions and additional references, see Block (1975) and Proschan and Sullo (1975).

ACKNOWLEDGEMENT

This research was partially supported by NIH grant 2R01-CA-18332.

REFERENCES

Aalen, O., Borgan, O., Keiding, N., and Thormann, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. Scandinavian Journal of Statistics 7, 161-171.

Barlow, R.E. and Proschan, F. (1977). Techniques for analyzing multivariate failure data, in Tsockos, C.P. and Shimi, I.N. eds., Theory and Applications of Reliability 1, 373-396.

Block, H.W. (1975). Continuous multivariate exponential extensions, in Barlow, Fussell and Singpurwalla, eds., Reliability and Fault Tree Analysis, SIAM: 285-306.

Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension. Journal of the American Statistical Association 69, 1031-1037.

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. Biometrika 68, 417-422.

Cox, D.R. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society B 34, 187-220.

Efron, B. (1967). The two-sample problem with censored data. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV, University of California Press, Berkeley, California, 831-853.

Efron, B. (1978). The geometry of exponential families. Annals of Statistics 6, 363-376.

Freund, J.E. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association 56, 971-977.

Gross, A.J. and Lam, C.F. (1981). Paired observations from a survival distribution. Biometrics 37, 505-512.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. Scandinavian Journal of Statistics 5, 195-199.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics 27, 887-906.

Lagakos, S.W. (1976). A stochastic model for censored-survival data in the presence of an auxiliary variable. Biometrics 32, 551-559.

Lagakos, S.W., Sommer, C.J. and Zelen, M. (1978). Semi-Markov models for partially censored data. Biometrika 65, 311-317.

Muñoz, A. (1980a). Nonparametric observation from censored bivariate observations. Technical Report, Division of Biostatistics, Stanford University.

Muñoz, A. (1980b). Consistency of the self-consistent estimator of the distribution function from censored observations. Technical Report, Division of Biostatistics, Stanford University.

Proschan, F. and Sullo, R. (1975). Estimating the parameters of a bivariate exponential distribution in several sampling situations, in Proschan, F. and Serfling, R.J., eds. Reliability and Biometry, SIAM, 1974, 423-440.

Voelkel, J. (1980). Multivariate counting processes and the probability of being in response function. Ph.D. dissertation, Statistics Department, University of Wisconsin-Madison.