Likelihood inferences with interval–censored data

Ramon Oller[1], Guadalupe Gómez[2] and M. Luz Calle[3]

[1] Dept. de Matemàtica i Informàtica.
Universitat de Vic.
Sagrada Família 7, 08500–Vic (Spain).
E-mail: ramon.oller@uvic.es

[2] Dept. d'Estadística i Investigació Operativa.
Universitat Politècnica de Catalunya.
Pau Gargallo 5, 08028–Barcelona (Spain).
E-mail: lupe.gomez@upc.es

[3] Dept. d'Informàtica i Matemàtica.
Universitat de Vic.
Sagrada Família 7, 08500–Vic (Spain).
E-mail: callem@uvic.es

LIKELIHOOD INFERENCES WITH INTERVAL–CENSORED DATA

RAMON OLLER[1] AND GUADALUPE GÓMEZ[2] AND M. LUZ CALLE[3]

[1]Grup de Recerca en Modelització de Sistemes Biològics, Dept. de Matemàtica i Informàtica, Facultat d'Empresa i Comunicació, Universitat de Vic, Carrer de la Sagrada Família 7, 08500–Vic, e-mail: ramon.oller@uvic.es

[2]Dept. d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Carrer de Pau Gargallo 5, 08028–Barcelona, e-mail: lupe.gomez@upc.es

[3]Grup de Recerca en Modelització de Sistemes Biològics, Dept. d'Informàtica i Matemàtica, Escola Politècnica Superior, Universitat de Vic, Carrer de la Sagrada Família 7, 08500–Vic, e-mail: callem@uvic.es

Data de recepció: 12/02/03
Data de publicació: 28/04/03

ABSTRACT
In survival data analysis the interval censoring problem has been usually treated via maximum likelihood inferences. In order to make use of a simpler expression of the likelihood function, standard methods suppose that conditions producing censoring do not affect the survival process. This paper is about formal conditions that ensure the validity of such a simplified likelihood. We state different notions of noninformative censoring appeared in the literature and we define the analogous constant–sum condition derived in the context of right censoring. We prove that the simplified likelihood produces correct inferences when these conditions hold. We discuss the identifiability of the distribution function of the failure time based on interval–censored data and we study the testability of the constant–sum condition.

RESUM
En l'anàlisi de la supervivència el problema de les dades censurades en un interval es tracta, usualment, via l'estimació per màxima versemblança. Amb l'objectiu d'utilitzar una expressió simplificada de la funció de versemblança, els mètodes estàndards suposen que les condicions que produeixen la censura no afecten el temps de fallada. En aquest article formalitzem les condicions que asseguren la validesa d'aquesta versemblança simplificada. Així, precisem diferents condicions de censura no informativa i definim una condició de suma constant anàloga a la derivada en el context de censura per la dreta. També demostrem que les inferències obtingudes amb la versemblança simplificada són correctes quan aquestes condicions són certes. Finalment, tractem la identificabilitat de la funció distribució del temps de fallada a partir de la informació observada i estudiem la possibilitat de contrastar el compliment de la condició de suma constant.

RESUMEN
En análisis de la supervivencia el problema de los datos censurados en un intervalo se trata, habitualmente, mediante la estimación por máxima verosimilitud. Con el objetivo de utilizar una expresión simplificada de la función de verosimilitud, los métodos estándar suponen que las condiciones que producen la censura no afectan el tiempo de fallo. En este artículo formalizamos las condiciones que aseguran la validez de esta verosimilitud simplificada. Así, precisamos diferentes condiciones de censura no informativa i definimos una condición de suma constante análoga a la derivada en el contexto de censura por la derecha. También demostramos que las inferencias obtenidas con la verosimilitud simplificada són correctas cuando estas condiciones son ciertas. Finalmente, tratamos la identificabilidad de la función de distribución del tiempo de fallo a partir de la información observada y estudiamos la posibilidad de contrastar el complimiento de la condición de suma constante.

---

1 Introduction

Interval censoring mechanisms arise when the event of interest cannot be directly observed and it is only known to have occurred during a random interval of time. This type of censored data has been extensively analyzed during the last years. Inference methods are mainly based on what we will refer to as the simplified likelihood, that is, the likelihood we would obtain if the censoring intervals were fixed in advance and we ignore the randomness of the intervals. Turnbull (1976), Groneboom and Wellner (1992) and Shick and Yu (2000) among other authors approach the estimation of the distribution function via this simplified likelihood. In this paper we discuss different conditions under which such likelihood–based inferences are correct. Williams and Lagakos (1977) in the context of right censoring and Betensky (2000) in the context of current status data addressed the same problem. Sufficient conditions for the appropriateness of the simplified likelihood with interval–censored data are introduced in the papers of Self and Grossman (1986) and Gómez *et al.* (2003). In a more general censoring framework, Heitjan and Rubin (1991), Heitjan (1993) and Gill *et al.* (1997) develop and characterize a closely related concept, the so called coarsening at random conditions. This paper surveys different definitions of the noninformative condition for interval–censored data, introduces the notion of the constant–sum model and justifies the validity of the simplified likelihood that has been widely used.

The remain of the paper is organized as follows. Section 2 introduces the notation, different noninformative censoring conditions and states their equivalences. In Section 3 we generalize the constant–sum condition, as introduced by Williams and Lagakos (1977), in the context of right–censoring. We distinguish between this condition that ensure that the inference process can omit the randomness of the intervals and noninformative conditions that ensure that the censoring mechanism cannot affect the event time. We state the relationship between these two concepts. Section 4 revises known censoring models and reduces the general concepts of the previous section to the underlying model. In Section 5 we consider whether it is possible to test either the constant–sum condition or the noninformative condition based on observable data.

2 Terminology and Noninformative models

Let T be the random variable of interest. In our setting T is a positive random variable representing the time until the occurrence of a certain event $\mathcal{E}$ with unknown right-continuous distribution function $W(t) = \mathrm{Prob}\{T \leq t\}$. Data is said to be interval–censored when the time to $\mathcal{E}$ is unknown and instead we observe a time interval $\lfloor L, R \rfloor$ where L is the last observed time before the event $\mathcal{E}$ has occurred and R indicates the first time the event $\mathcal{E}$ has been observed. We use the $\lfloor L, R \rfloor$ notation to indicate an interval that can be closed, open or half open depending on the interval censoring model. For example, Peto (1973) and Turnbull (1976) consider closed intervals $[L, R]$, while Groeneboom and Wellner (1992) suppose half open intervals $(L, R]$ and Yu *et al.* (2000) define a mixed interval–censored scheme that involves open, closed and half open intervals. In each of these situations, we are in fact formally observing a random censoring vector (L, R), such that $T \in \lfloor L, R \rfloor$ with probability 1. A model for interval–censored data is determined by the joint distribution, $F_{L,R,T}$, between the random variable T and the observables (L, R), under the constraint that

$$P(T \in \lfloor L, R \rfloor) = \iiint_{\{(l,r,t): t \in \lfloor l,r \rfloor\}} dF_{L,R,T}(l, r, t) = 1.$$

The marginal laws of the survival time and the observables are characterized, respectively, by

$$dW(t) = \iint_{\{(l,r): t \in \lfloor l,r \rfloor\}} dF_{L,R,T}(l, r, t)$$

and

$$dF_{L,R}(l, r) = \int_{\{t: t \in \lfloor l,r \rfloor\}} dF_{L,R,T}(l, r, t) = P(L \in dl, R \in dr, T \in \lfloor l, r \rfloor). \tag{1}$$

Expression (1) is the contribution to the likelihood of an individual with observed interval $\lfloor l, r \rfloor$. The goal of this paper is to define conditions under which this contribution can be reduced to $P(T \in \lfloor l, r \rfloor)$. This probability is what we refer as simplified likelihood.

We introduce three possible definitions for noninformativeness of the interval censoring mechanism in theorem 2.1. The first characterization has been proposed in Self and Grossman (1986). Gómez *et al.* (2003) uses the second definition to derive the simplified likelihood while the third definition follows from the coarsening at random notion used in Heitjan and Rubin (1991), Heitjan (1993) and Gill *et al.* (1997).

Theorem 2.1 *The following properties define the noninformative condition and are equivalent:*

(a) The conditional distribution of T given L and R satisfies

$$dF_{T|L,R}(t|l,r) = \frac{dW(t)}{P(T \in \lfloor l,r \rfloor)} \; \mathbf{1}_{\{t:t\in\lfloor l,r\rfloor\}}(t)$$

that is, censoring in $\lfloor l,r \rfloor$ provides the same information as T being in $\lfloor l,r \rfloor$.

(b) The conditional distribution of L and R given T satisfies that

$$dF_{L,R|T}(l,r|t) = \frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)} \; \mathbf{1}_{\{(l,r):t\in\lfloor l,r\rfloor\}}(l,r) \tag{2}$$

that is, the observables (l,r) are not influenced by the specific value of T in $\lfloor l,r \rfloor$.

(c) The conditional distribution of L and R given T satisfies that

$$dF_{L,R|T}(l,r|t) = dF_{L,R|T}(l,r|t') \qquad \textit{on } \{(l,r) : t \in \lfloor l,r \rfloor \textit{ and } t' \in \lfloor l,r \rfloor\}$$

that is, two specific values of T that are consistent with the observables always provide the same information.

Proof:

(a) implies *(b)*:

If $dF_{T|L,R}(t|l,r) = \dfrac{dW(t)}{P(T \in \lfloor l,r \rfloor)} \; \mathbf{1}_{\{t:t\in\lfloor l,r\rfloor\}}(t)$, then for any (l,r,t) such that $t \in \lfloor l,r \rfloor$, following the usual rules for conditional distributions, we have

$$dF_{L,R|T}(l,r|t) = \frac{dF_{L,R,T}(l,r,t)}{dW(t)} = \frac{dF_{T|L,R}(t|l,r)dF_{L,R}(l,r)}{dW(t)}$$

$$= \frac{dW(t)dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)dW(t)} = \frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)}$$

(b) implies *(c)*:

If $dF_{L,R|T}(l,r|t) = \dfrac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)}\ \mathbf{1}_{\{(l,r):t\in\lfloor l,r\rfloor\}}(l,r)$, then clearly

$$\frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)}\ \mathbf{1}_{\{(l,r):t\in\lfloor l,r\rfloor\}}(l,r) = \frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)}\ \mathbf{1}_{\{(l,r):t'\in\lfloor l,r\rfloor\}}(l,r)$$

on $\{(l,r) : t \in \lfloor l,r \rfloor$ and $t' \in \lfloor l,r \rfloor\}$

(c) implies *(a)*:

If $dF_{L,R|T}(l,r|t) = dF_{L,R|T}(l,r|t')$ on $\{(l,r) : t \in \lfloor l,r \rfloor$ and $t' \in \lfloor l,r \rfloor\}$, then for any fixed (l,r,t) such that $t \in \lfloor l,r \rfloor$

$$dF_{L,R}(l,r) = \int_{\{s:s\in\lfloor l,r\rfloor\}} dF_{L,R,T}(l,r,s) = \int_{\{s\in\lfloor l,r\rfloor\}} dF_{L,R|T}(l,r|s)dW(s)$$

$$= \int_{\{s\in\lfloor l,r\rfloor\}} dF_{L,R|T}(l,r|t)dW(s) = dF_{L,R|T}(l,r|t)P(T \in \lfloor l,r \rfloor)$$

Then, if we use this last equality and we follow the usual rules for conditional distributions, we have

$$dF_{T|L,R}(t|l,r) = \frac{dF_{L,R,T}(l,r,t)}{dF_{L,R}(l,r)} = \frac{dF_{L,R|T}(l,r|t)\ dW(t)}{dF_{L,R|T}(l,r|t)\ P(T \in \lfloor l,r \rfloor)} = \frac{dW(t)}{P(T \in \lfloor l,r \rfloor)}$$

$\diamond$

3 Constant–sum models

The definition of the constant–sum condition extends Williams and Lagakos's one in the context of right censoring (1977). The analogous condition proposed here is based on the marginal laws of the censoring model, W and $F_{L,R}$.

Definition 3.1 *A censoring model is constant–sum if and only if the following equation holds*

$$\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)} = 1 \quad \textit{for any } t > 0 \textit{ such that } dW(t) \neq 0. \tag{3}$$

Theorem 3.2 *If a censoring model is constant–sum, then the simplified likelihood given by* $P(T \in \lfloor l,r \rfloor)$ *is a proper basis for inferences.*

Proof:

We first note that the full contribution to the likelihood given in equation (1) can be written as

$$P(L \in dl, R \in dr, T \in \lfloor l,r \rfloor) = P(T \in \lfloor l,r \rfloor) \cdot dK(l,r)$$

where

$$dK(l,r) = P(L \in dl, R \in dr|\ T \in \lfloor l,r \rfloor)$$

$$= \frac{P(L \in dl, R \in dr,\ T \in \lfloor l,r \rfloor)}{P(T \in \lfloor l,r \rfloor)} = \frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)}.$$

Thus, any likelihood–based inference will only be based on the distribution function W and the conditional measure K subject to the constraints: (i) W is a distribution function, (ii) $dK \geq 0$ and (iii)

$$\iint_{\{0 \leq l \leq r\}} P(T \in \lfloor l,r \rfloor) dK(l,r) = 1$$

which can be equivalently written as

$$\int_0^{+\infty} \left(\iint_{\{(l,r):t \in \lfloor l,r \rfloor\}} dK(l,r) \right) dW(t) = 1.$$

Therefore, if condition (iii) does not impose additional parametric relations between W and dK, then inferences based on the simplified likelihood will be valid, at least concerning the survival time distribution. If we assume a constant–sum model then the following equality

$$\iint_{\{(l,r):t \in \lfloor l,r \rfloor\}} dK(l,r) = 1 \qquad \text{for any } t > 0 \text{ such that } dW(t) \neq 0$$

holds, which in turn implies that there is no constrain between W and dK in condition (iii) and the theorem follows.

◇

The following two propositions prove that the noninformative condition is a sufficient but not necessary condition for a model to be constant–sum. In this sense, the second proposition shows that the constant–sum notion expresses in terms of marginal laws the restriction on the conditional laws expressed by the noninformative notion.

Proposition 3.3 *If a censoring model is noninformative then the model is constant-sum.*

Proof:

Indeed, for any $t > 0$ such that $dW(t) \neq 0$, it follows from equation (2) that

$$\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P(T\in\lfloor l,r\rfloor)} = \iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} dF_{L,R|T}(l,r|t) = 1$$

and, consequently, the constant–sum condition holds.

$\diamond$

Proposition 3.4 *If a censoring model, F_{L_1,R_1,T_1}, satisfies the constant– sum condition, then it always exists a noninformative model, F_{L_2,R_2,T_2}, such that $W_2 = W_1$ and $F_{L_2,R_2} = F_{L_1,R_1}$.*

Proof:

Define F_{L_2,R_2,T_2} by

$$dF_{L_2,R_2,T_2}(l,r,t) = \frac{dW_1(t)dF_{L_1,R_1}(l,r)}{P(T_1\in\lfloor l,r\rfloor)},$$

which defines a probability measure such that $T \in \lfloor L, R\rfloor$ with probability one,

$$\iiint_{\{(l,r,t):t\in\lfloor l,r\rfloor\}} dF_{L_2,R_2,T_2}(l,r,t)$$

$$= \int_0^{+\infty} dW_1(t)\left(\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L_1,R_1}(l,r)}{P(T_1\in\lfloor l,r\rfloor)}\right) = \int_0^{+\infty} dW_1(t) = 1.$$

Furthermore, for any $t > 0$

$$dW_2(t) = \iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} dF_{L_2,R_2,T_2}(l,r,t)$$

$$= dW_1(t) \left(\iint_{\{(l,r): t \in \lfloor l,r \rfloor\}} \frac{dF_{L_1,R_1}(l,r)}{P(T_1 \in \lfloor l,r \rfloor)} \right) = dW_1(t)$$

and for any (l,r) such that $0 \leq l \leq r$

$$dF_{L_2,R_2}(l,r) = \int_{\{t: t \in \lfloor l,r \rfloor\}} dF_{L_2,R_2,T_2}(l,r,t) =$$

$$= dF_{L_1,R_1}(l,r) \left(\int_{\{t: t \in \lfloor l,r \rfloor\}} \frac{dW_1(t)}{P(T_1 \in \lfloor l,r \rfloor)} \right) = dF_{L_1,R_1}(l,r).$$

Finally, it follows that $F_{L_2,R_2|T_2}$ satisfies equation (2) for any (l,r,t) such that $t \in \lfloor l,r \rfloor$ and $dW_2(t) \neq 0$,

$$dF_{L_2,R_2|T_2}(l,r|t) = \frac{dF_{L_2,R_2,T_2}(l,r,t)}{dW_2(t)} = \frac{dW_1(t) dF_{L_1,R_1}(l,r)}{P(T_1 \in \lfloor l,r \rfloor) dW_2(t)} = \frac{dF_{L_2,R_2}(l,r)}{P(T_2 \in \lfloor l,r \rfloor)}.$$

$\diamond$

For the sake of completeness, it is interesting to remark that for any $t > 0$ the constant–sum condition can be expressed as

$$\iint_{\{(l,r): t \in \lfloor l,r \rfloor\}} \frac{dW(t)}{P(T \in \lfloor l,r \rfloor)} \, dF_{L,R}(l,r) = dW(t). \tag{4}$$

Equation (4) is the well-known self–consistent equation which is the basis of the nonparametric maximum likelihood estimation of W, see Turnbull (1976) or Gómez et al. (2003).

4 Examples

We discuss the meaning of the noninformative and constant–sum conditions for the particular cases of right–censored data, double–censored data and interval–censored data case k. The results for right–censored data and interval–censored data case 1 are similar to those in Williams and Lagakos (1977) and Betensky (2000), respectively.

Example 4.1 (Right censoring) *Right censored-data arise when the event of interest can only be observed if the survival time does not exceed the value of a positive random censoring variable, C. The observed data for an individual is traditionally expressed by the pair (X, δ) where $X = \min(T, C)$ and $\delta = \mathbf{1}_{\{T \leq C\}}$. Using interval censoring notation, the vector of observables is,*

$$(L, R) = (T, T) \cdot \delta + (C, +\infty) \cdot (1 - \delta)$$

and the observed intervals are defined as

$$\lfloor l, r \rfloor = \begin{cases} [l, r] & \text{if } l = r \\ (l, r) & \text{if } r = +\infty. \end{cases}$$

Thus, the joint distribution function for L, R, T is given by:

$$dF_{L,R,T}(l, r, t) = \begin{cases} P(C \geq t, T \in dt) & \text{if} \quad l = t = r \\ P(C \in dl, T \in dt) & \text{if} \quad l < t \quad \textit{and} \quad r = +\infty \\ 0 & \text{otherwise.} \end{cases}$$

In this setting, the noninformative and the constant–sum conditions become, respectively,

$$P(C \in dl | T = t) = P(C \in dl | T > l) \quad \text{ for any } t > l > 0$$

and

$$P(C \geq t | T = t) + \int_0^{t^-} P(C \in dl | T > l) = 1 \quad \text{ for any } t > 0.$$

The proofs are postponed to appendix A. If we assume the usual independence between the variables T and C, then both conditions are clearly satisfied.

Example 4.2 (Double censoring) *Data is said to be double–censored when the event of interest can only be observed inside the window $[C_1, C_2]$, where C_1 and C_2 are positive random variables and $C_1 < C_2$ (Chang and Yang, 1987). The observed data for an individual is of the form (X, δ, γ) where $\delta = \mathbf{1}_{\{T < C_1\}}$, $\gamma = \mathbf{1}_{\{T \leq C_2\}}$ and $X = C_1 \cdot \delta + T \cdot (1 - \delta) \cdot \gamma + C_2 \cdot (1 - \delta) \cdot (1 - \gamma)$. In the interval censoring framework, the vector of observables can be expressed as*

$$(L, R) = (0, C_1) \cdot \delta + (T, T) \cdot (1 - \delta) \cdot \gamma + (C_2, +\infty) \cdot (1 - \delta) \cdot (1 - \gamma)$$

and intervals are defined as

$$\lfloor l, r \rfloor = \begin{cases} [l, r) & \text{if } l = 0 \\ [l, r] & \text{if } l = r \\ (l, r) & \text{if } r = +\infty. \end{cases}$$

In this model the joint probability law of the survival and the observables is given by,

$$dF_{L,R,T}(l, r, t) = \begin{cases} P(C_1 \in dr, T \in dt) & \text{if} \quad l = 0 \quad \textit{and} \quad t < r \\ P(C_1 \le t, C_2 \ge t, T \in dt) & \text{if} \quad l = t = r \\ P(C_2 \in dl, T \in dt) & \text{if} \quad l < t \quad \textit{and} \quad r = +\infty \\ 0 & \text{otherwise.} \end{cases}$$

Under a double censoring setup the noninformative condition is expressed through the following two equalities:

- $P(C_1 \in dr | T = t) = P(C_1 \in dr | T < r)$ *for any* $0 < t < r$
- $P(C_2 \in dl | T = t) = P(C_2 \in dl | T > l)$ *for any* $t > l > 0$.

Furthermore, the constant–sum condition reduces to

$$\int_t^{+\infty} P(C_1 \in dr | T < r) + P(C_1 \le t, C_2 \ge t | T = t)$$

$$+ \int_0^{t^-} P(C_2 \in dl | T > l) = 1.$$

We observe again that independence between T and (C_1, C_2) implies both conditions. Details on the computations are given in appendix B.

Example 4.3 (Interval–censored data, case k) *This interval censoring scheme has been largely studied, specially the case 1 and case 2 (Groeneboom and Wellner, 1992; Schick and Yu, 2000). In the interval–censored model, case 1 or current status data, the event is only known to be larger or smaller than an observed monitoring time. The interval–censored model, case 2, consider two monitoring times, X_1 and X_2 with $X_1 < X_2$, where it is only possible to determine whether the event of interest occurs before the first monitoring time ($T \le X_1$), between the two monitoring times ($X_1 < T \le X_2$), or after the last monitoring* *time ($T > X_2$). Although interval censoring case 2 looks like the double censoring model, it is fundamentally different because the value of T is unknown inside the window $(X_1, X_2]$. The general case k model consider k positive random monitoring times, $X_1 \leq \cdots \leq X_k$, such that the event of interest can only be determined to have occurred before, between or after that times. The vector of observables is*

$$(L, R) = (0, X_1)\mathbf{1}_{\{T \leq X_1\}} + \sum_{j=2}^{k} \Big\{(X_{j-1}, X_j)\mathbf{1}_{\{X_{j-1} < T \leq X_j\}}\Big\} + (X_k, +\infty)\mathbf{1}_{\{T > X_k\}}.$$

Thus, the intervals are defined as,

$$\lfloor l, r \rfloor = \begin{cases} (l, r) & \text{if } r = +\infty. \\ (l, r] & \text{otherwise} \end{cases}$$

The joint distribution function for L, R, T is expressed as

$$dF_{L,R,T}(l,r,t) = \begin{cases} P(X_1 \in dr, T \in dt) & \text{if } l = 0 \textit{ and } t \leq r \\ \displaystyle\sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr, T \in dt) & \text{if } 0 < l < t \leq r < +\infty \\ P(X_k \in dl, T \in dt) & \text{if } l < t \textit{ and } r = +\infty \\ 0 & \text{otherwise.} \end{cases}$$

In this model, as it is shown in appendix C, the noninformative condition can be written as,

- $P(X_1 \in dr | T = t) = P(X_1 \in dr | T \leq r) \quad \text{for any } 0 < t \leq r$
- $\displaystyle\sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr | T = t) = \sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr | l < T \leq r)$
  for any $0 < l < t \leq r$
- $P(X_k \in dl | T = t) = P(X_k \in dl | T > l) \quad \text{for any } t > l > 0.$

We can also see that, for any $t > 0$ such that $dW(t) \neq 0$, the constant–sum equation is

$$\int_{t-}^{+\infty} P(X_1 \in dr | T \leq r) + \sum_{j=2}^{k} \int_0^{t^-} \int_{t^-}^{+\infty} P(X_{j-1} \in dl, X_j \in dr | l < T \leq r)$$

$$+ \int_0^{t^-} P(X_k \in dl | T > l) = 1$$

and furthermore (see proposition C.1), when T is continuous, it can be simplified to

$$P(X_1 \in dt|T \leq t) + \sum_{j=2}^{k} \int_{\{l:l\in[0,t)\}} P(X_{j-1} \in dl, X_j \in dt|l < T \leq t)$$

$$= \sum_{j=2}^{k} \int_{\{r:r\in(t,+\infty]\}} P(X_{j-1} \in dt, X_j \in dr|t < T \leq r) + P(X_k \in dt|T > t)$$

Again, when the model satisfies the usual assumption of independence between the survival time, T, and the monitoring times, $(X_1, \ldots, X_k)$, all the above equations hold.

5 Identifiability and testability problems

In this section we discuss the identifiability of the distribution of the failure time, W, and the testability of the constant–sum condition on the basis of the observables, $F_{L,R}$.

Definition 5.1 *Two interval censoring models, $\{F_{L_1,R_1,T_1}\}$ and $\{F_{L_2,R_2,T_2}\}$, are said to be indistinguishable if the marginal distribution of the observables is the same, that is, $F_{L_1,R_1} = F_{L_2,R_2}$.*

It is clear, see proposition D.1, that we might find two indistinguishable censoring models with two different failure time distributions. Thus, the distribution of the failure time cannot be identified on the basis of the observables unless we assume some kind of restriction on the model.

We will show now that if we restrict to the class of indistinguishable constant–sum models, then we can identify the probability of the failure time in the observable intervals. More precisely, we can ensure, from the following theorem, that any constant–sum model in a class of indistinguishable models has the same simplified likelihood.

Theorem 5.2 *Let $\{F_{L,R,T_1}\}$ and $\{F_{L,R,T_2}\}$ be two indistinguishable constant–sum models such that $dW_1 \neq 0$ if and only if $dW_2 \neq 0$, then $P(T_1 \in \lfloor l, r \rfloor) = P(T_2 \in \lfloor l, r \rfloor)$ $dF_{L,R}$ almost surely .*

Proof:

If model $\{F_{L,R,T_1}\}$ is constant–sum then

$$\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P(T_1 \in \lfloor l,r\rfloor)} = 1,$$

which implies that

$$\int_0^{+\infty} dW_2(t) \iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P(T_1 \in \lfloor l,r\rfloor)} = 1$$

which in turn implies that

$$\iint \frac{P(T_2 \in \lfloor l,r\rfloor)}{P(T_1 \in \lfloor l,r\rfloor)} dF_{L,R}(l,r) = 1.$$

Analogously, it is clear that starting with model $\{F_{L,R,T_2}\}$ it follows that

$$\iint \frac{P(T_1 \in \lfloor l,r\rfloor)}{P(T_2 \in \lfloor l,r\rfloor)} dF_{L,R}(l,r) = 1.$$

Thus, the two equations and lemma D.2 prove the statement of this theorem.

$\diamond$

Summarizing, the simplified likelihood produces wrong inferences and it is not identifiable without the assumption of the constant–sum condition. Thus, it remains to look again at equation (3) and study the testability of this property on the basis of the observables, $F_{L,R}$. Then, it follows that the only way to test the constant–sum property is to search for conditions on the observables, $F_{L,R}$, which ensure the existence of a distribution function, $\widetilde{W}$, that solves the following equation

$$\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)} = 1 \quad \text{for any } t>0 \text{ such that } d\widetilde{W}(t) \neq 0. \tag{5}$$

If such a solution exists, it is not possible to know from the observables, $F_{L,R}$, whether the model is constant–sum or not because we can always construct a constant–sum model that is indistinguishable from the underlying model, see proposition D.3. On the other hand, if no solution exists, we clearly know from the observables, $F_{L,R}$, that the underlying model is neither constant–sum nor noninformative.

At this point, the question to answer is when equation (5) has solution. This is an open problem though it is solved in two special cases, discrete data and current status data, see examples below.

Example 5.3 (Discrete data) *Let $\{F_{L,R,T}\}$ be a model such that L, R and T have the same finite supports, then equation (5) has always solution and therefore in the discrete case it is not possible to test the constant–sum condition. See proposition D.4 for the justification.*

Example 5.4 (Current status data) *In this example we study the problem of the testability of the constant–sum condition in the interval censoring model case 1 and extent the results in Betensky (2000). If we note by X the random monitoring time, then the vector of observables in this model is,*

$$(L,R) = (0,X)\mathbf{1}_{\{T\leq X\}} + (X,+\infty)\mathbf{1}_{\{T>X\}}$$

with probability law

$$dF_{L,R}(l,r) = \mathbf{1}_{\{0\}}(l)P(X\in dr, T\leq r) + \mathbf{1}_{\{+\infty\}}(r)P(X\in dl, T>l)$$

If we suppose that T is continuous, see proposition C.1, the constant–sum condition reduces to

$$\frac{dF_{L,R}(0,t)}{W(t)} = \frac{dF_{L,R}(t,+\infty)}{1-W(t)}$$

or, equivalently,

$$W(t) = \frac{dF_{L,R}(0,t)}{dF_{L,R}(0,t)+dF_{L,R}(t,+\infty)} = P(T\leq t|X=t)$$

Following Example 4.3, the noninformative condition can be written as,

$$P(X\in dx|T=t) = \begin{cases} \dfrac{P(X\in dx, T\leq x)}{W(x)} & \text{for any } 0<t\leq x \\[2ex] \dfrac{P(X\in dx, T>x)}{1-W(x)} & \text{for any } t>x>0 \end{cases}$$

This last expression together with the above constant–sum characterization, $W(x) = P(T \leq x|X = x)$, implies that

$$P(X \in dx|T = t) = P(X \in dx) \quad \text{for any } x, t > 0$$

That is, in the case of current status data the noninformative condition is equivalent to the independence of T and X.

In this framework, equation (5) reduces to

$$\widetilde{W}(t) = P(T \leq t|X = t)$$

and the existence of a solution is equivalent to the condition that $P(T \leq t|X = t)$ is a distribution function. Rabinowitz (2000) proposes to test the constant–sum condition with a rank statistic which detects decreasing trends on this probability.

Finally, as an illustration of a class of current status models which covers all different conditions studied in this paper, we consider (T, X) to be in the family of two dimensional log–normal distributions. In this class the joint density function of T and X is of the form:

$$f(t,x) = \frac{\exp(-\frac{1}{2(1-\rho^2)}\left[\frac{(\ln(t)-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(\ln(t)-\mu_1)(\ln(x)-\mu_2)}{\sigma_1\sigma_2} + \frac{(\ln(x)-\mu_2)^2}{\sigma_2^2}\right])}{(2\pi)\sigma_1\sigma_2 tx\sqrt{1-\rho^2}}$$

It can be shown that this class of models have the following characterizations:

- *the model is noninformative if and only if $\rho = 0$,*
- *the model is constant–sum if and only if $\rho = 0$ or $\left\{\rho = \frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\right.$ and $\mu_1 = \mu_2$ and $\left.\sigma_1 < \sigma_2\right\}$,*
- *equation (5) has solution if and only if $\rho < \frac{\sigma_2}{\sigma_1}$.*

ACKNOWLEDGEMENTS

The authors are grateful to the GRASS group for their suggestions and fruitful discussions. This research was partially supported by the Dirección General de Enseñanza Superior e Investigación Científica Grant PB98–0919 and by the National Institute of Health Grant AI24643.

References

Betensky, R.A. (2000). On nonidentifiability and noninformative censoring for current status data. *Biometrika* 87, 218–221.

Chang, M.N. and Yang, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* 16, 1536–1547.

Gill, R.D., van der Laan, M.J. and Robins, J.M. (1997). Coarsening at random: characterizations, conjectures, counter–examples. In *Proceedings First Seattle Symposium on Biostatistics: Survival Analysis*, Springer–Verlag, 255–294.

Gómez, G., Calle, M.L. and Oller, R. (2003). Frequentist and bayesian approaches for interval–censored data. *Statistical Papers*, in press.

Groeneboom, P. and Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag, Basel.

Heitjan, D.F. (1993). Ignorability and coarse data: Some biomedical examples. *Biometrics* 49, 1099–1109.

Heitjan, D.F. and Rubin, D.B. (1991). Ignorability and coarse data. *Ann. of Statist.* 19, 2244–2253.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Appl. Statist.* 22, 86–91.

Rabinowitz, D. (2000). Testing current status data for dependent censoring. *Statist. Probab. Lett.* 48, 213–216.

Self, S.G. and Grossman, E.A. (1986). Linear rank tests for interval–censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* 42, 521–530.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval–censored data. *Scand. J. Statist.* 27, 45–55.

Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* 38, 290–295.

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: constant–sum and variable–sum models. *Biometrika* 64, 215–224.

Yu, Q., Li, L. and Wong, G.Y.C. (2000). On consistency of the self–consistent estimator of survival functions with interval–censored data. *Scand. J. Statist.* 27, 35–44.

A Appendix

Proof of results in Example 4.1

The definition of the right censoring model, $F_{L,R,T}$, implies that

$$dF_{L,R|T}(l,r|t) = \begin{cases} P(C \geq t|T=t) & \text{if} \quad l=t=r \\ P(C \in dl|T=t) & \text{if} \quad l<t \quad \text{and} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

and

$$dF_{L,R}(l,r) = \begin{cases} P(C \geq l, T \in dl) & \text{if} \quad l=r \\ P(C \in dl, T > l) & \text{if} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)} = \begin{cases} P(C \geq l|T=l) & \text{if} \quad l=r \\ P(C \in dl|T > l) & \text{if} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

If we impose the second characterization of the noninformative condition, then the following equations should be satisfied:

- If $l=r=t$, $P(C \geq t|T=t) = P(C \geq l|T=l)$, but in this case the equality always holds.
- If $l<t$ and $r=+\infty$, $P(C \in dl|T=t) = P(C \in dl|T>l)$.

On the other hand, if we impose the constant–sum condition to the model, then for any $t>0$ such that $dW(t) \neq 0$:

$$P(C \geq t|T=t) + \int \mathbf{1}_{(l,+\infty)}(t) P(C \in dl|T>l) = 1$$

and it follows that,

$$P(C \geq t|T=t) + \int_0^{t^-} P(C \in dl|T>l) = 1$$

B Appendix

Proof of results in Example 4.2

The definition of the double censoring model, $F_{L,R,T}$, implies that

$$dF_{L,R|T}(l,r|t) = \begin{cases} P(C_1 \in dr|T=t) & \text{if} \quad l=0 \quad \text{and} \quad t<r \\ P(C_1 \leq t, C_2 \geq t|T=t) & \text{if} \quad l=t=r \\ P(C_2 \in dl|T=t) & \text{if} \quad l<t \quad \text{and} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

and

$$dF_{L,R}(l,r) = \begin{cases} P(C_1 \in dr, T<r) & \text{if} \quad l=0 \\ P(C_1 \leq l, C_2 \geq l, T \in dl) & \text{if} \quad l=r \\ P(C_2 \in dl, T>l) & \text{if} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)} = \begin{cases} P(C_1 \in dr|T<r) & \text{if} \quad l=0 \\ P(C_1 \leq l, C_2 \geq l|T=l) & \text{if} \quad l=r \\ P(C_2 \in dl|T>l) & \text{if} \quad r=+\infty \\ 0 & \text{otherwise} \end{cases}$$

If we impose the second characterization of the noninformative condition, then the following equations should be satisfied:

- If $l=0$ and $t<r$, $P(C_1 \in dl|T=t) = P(C \in dr|T<r)$.
- If $l=t=r$, $P(C_1 \leq t, C_2 \geq t|T=t) = P(C_1 \leq l, C_2 \geq l|T=l)$, but in this case the equality always holds.
- If $t>l$ and $r=+\infty$, $P(C_2 \in dl|T=t) = P(C_2 \in dl|T>l)$.

Furthermore, for any $t>0$ such that $dW(t) \neq 0$, the constant–sum condition simplifies to:

$$\int \mathbf{1}_{[0,r)}(t) P(C_1 \in dr|T<r)$$

$$+P(C_1 \leq t, C_2 \geq t|T=t) + \int \mathbf{1}_{(l,+\infty)}(t) P(C_2 \in dl|T>l) = 1$$

and, consequently,

$$\int_t^{+\infty} P(C_1 \in dr|T < r) + P(C_1 \leq t, C_2 \geq t|T = t)$$

$$+ \int_0^{t^-} P(C_2 \in dl|T > l) = 1$$

C Appendix

Proof of results in Example 4.3

The definition of the case k interval censoring model, $F_{L,R,T}$, implies that

$$dF_{L,R|T}(l,r|t) = \begin{cases} P(X_1 \in dr|T = t) & \text{if } l = 0 \text{ and } t \leq r \\ \sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr|T = t) & \text{if } 0 < l < t \leq r < +\infty \\ P(X_k \in dl|T = t) & \text{if } l < t \text{ and } r = +\infty \\ 0 & \text{otherwise} \end{cases}$$

and

$$dF_{L,R}(l,r) = \begin{cases} P(X_1 \in dr, T \leq r) & \text{if } l = 0 \\ \sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr, l < T \leq r) & \text{if } 0 < l < r < +\infty \\ P(X_k \in dl, T > l) & \text{if } r = +\infty \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\frac{dF_{L,R}(l,r)}{P(T \in \lfloor l,r \rfloor)} = \begin{cases} P(X_1 \in dr|T \leq r) & \text{if } l = 0 \\ \sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr|l < T \leq r) & \text{if } 0 < l < r < +\infty \\ P(X_k \in dl|T > l) & \text{if } r = +\infty \\ 0 & \text{otherwise} \end{cases}$$

If we impose the second characterization of the noninformative condition, then the following equations should be satisfied:

- If $l = 0$ and $t \leq r$, $P(X_1 \in dr|T = t) = P(X_1 \in dr|T \leq r)$.

- If $0 < l < t \leq r < +\infty$, $\sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr | T = t) = \sum_{j=2}^{k} P(X_{j-1} \in dl, X_j \in dr | l < T \leq r)$.
- If $t > l$ and $r = +\infty$, $P(X_k \in dl | T = t) = P(X_k \in dl | T > l)$.

Furthermore, for any $t > 0$ such that $dW(t) \neq 0$, the constant–sum condition can be written as:

$$\int \mathbf{1}_{(0,r]}(t) P(X_1 \in dr | T \leq r)$$

$$+ \sum_{j=2}^{k} \iint_{(0,+\infty)^2} \mathbf{1}_{(l,r]}(t) P(X_{j-1} \in dl, X_j \in dr | l < T \leq r)$$

$$+ \int \mathbf{1}_{(l,+\infty)}(t) P(X_k \in dl | T > l) = 1$$

and, consequently,

$$\int_{t-}^{+\infty} P(X_1 \in dr | T \leq r) + \sum_{j=2}^{k} \int_{0}^{t^-} \int_{t^-}^{+\infty} P(X_{j-1} \in dl, X_j \in dr | l < T \leq r)$$

$$+ \int_{0}^{t^-} P(X_k \in dl | T > l) = 1$$

The following proposition simplifies the constant–sum equation when T is continuous.

Proposition C.1 *If intervals cannot be singletons and T is a positive continuous random variable with $dW(t) \neq 0$ for any $t > 0$, then the constant–sum condition is equivalent to the following equality for any $t > 0$*

$$\int_{\{l: l \in [0,t)\}} \frac{dF_{L,R}(l,t)}{P(T \in (l,t])} = \int_{\{r: r \in (t,+\infty]\}} \frac{dF_{L,R}(t,r)}{P(T \in (t,r])}$$

Proof:

We suppose that $\lfloor l, r \rfloor = (l, r]$ without loss of generality because T is continuous and the intervals cannot be singletons. If constant–sum condition holds and we define $dK(l,r) = \dfrac{dF_{L,R}(l,r)}{P(T \in \lfloor l, r \rfloor)}$, then for any $t > 0$ it follows that

$$1 = \int_{[0,t)} \int_{[t,+\infty]} dK(l,r)$$

This property implies that for any $0 < a < b < +\infty$

$$\int_{[0,a)} \int_{[a,+\infty]} dK(l,r) = \int_{[0,b)} \int_{[b,+\infty]} dK(l,r)$$

$$\iff \int_{[0,a)} \int_{[a,b)} dK(l,r) + \int_{[0,a)} \int_{[b,+\infty]} dK(l,r)$$
$$= \int_{[0,a)} \int_{[b,+\infty]} dK(l,r) + \int_{[a,b)} \int_{[b,+\infty]} dK(l,r)$$

$$\iff \int_{[0,a)} \int_{[a,b)} dK(l,r) = \int_{[a,b)} \int_{[b,+\infty]} dK(l,r)$$

$$\iff \int_{[0,a)} \int_{[a,b)} dK(l,r) + \int_{[a,r)} \int_{[a,b)} dK(l,r)$$
$$= \int_{[a,b)} \int_{[b,+\infty]} dK(l,r) + \int_{[a,b)} \int_{(l,b)} dK(l,r)$$

$$\iff \int_{[0,r)} \int_{[a,b)} dK(l,r) = \int_{[a,b)} \int_{(l,+\infty]} dK(l,r)$$

$$\iff \int_{[0,t)} \int_{[a,b)} dK(l,t) = \int_{[a,b)} \int_{(t,+\infty]} dK(t,r)$$

Thus, we have proved that $\displaystyle\int_{\{l:l\in[0,t)\}} dK(l,t) = \int_{\{r:r\in(t,+\infty]\}} dK(t,r)$ for any interval $[a,b)$. Using a monotone class theorem this result extends to the σ-algebra on $(0,+\infty)$.

Let us prove the reciprocal, that is, we suppose that for any $t > 0$

$$\int_{\{l:l\in[0,t)\}} dK(l,t) = \int_{\{r:r\in(t,+\infty]\}} dK(t,r)$$

Then, it follows, from the above equivalences, that for any $0 < a < b < +\infty$

$$\int_{[0,a)} \int_{[a,+\infty]} dK(l,r) = \int_{[0,b)} \int_{[b,+\infty]} dK(l,r) = k \leq 1$$

Thus, it means that

$$\int_0^{+\infty} \left(\iint_{\{(l,r):t\in(l,r]\}} dK(l,r) \right) dW(t) = k$$

and this equality is only possible if $k = 1$.

$\diamond$

D Appendix

Proof of results in Section 5

Proposition D.1 *Let $\{F_{L_1,R_1,T_1}\}$ be any model for interval–censored data with T_1 being continuous and $dW_1 \neq 0$ for any $t > 0$, then there always exists another indistinguishable model with different failure time.*

Proof:

Consider a set A such that $0 < P((L_1, R_1) \in A) < 1$ and define the following censoring scheme,

$$F_{L_2,R_2} = F_{L_1,R_1}$$

and

$$T_2 = \begin{cases} T_1 & \text{if } (L_1, R_1) \notin A \\ \dfrac{L_1 + R_1}{2} & \text{if } (L_1, R_1) \in A \text{ and } R_1 \neq +\infty \\ 2L_1 & \text{if } (L_1, R_1) \in A \text{ and } R_1 = +\infty \end{cases}$$

We obtain a new model $\{F_{L_2,R_2,T_2}\}$ where $F_{L_2,R_2} = F_{L_1,R_1}$ and clearly $W_2 \neq W_1$.

$\diamond$

Lemma D.2 *Let μ be a probability measure and f a μ–measurable positive function such that*

$$\int f \; d\mu = 1 \qquad \textit{and} \qquad \int \frac{1}{f} \; d\mu = 1$$

then $f = 1$ μ–almost surely.

Proof:

If we sum the two integrals,

$$\int \left(f + \frac{1}{f} \right) \; d\mu = 2$$

Then we can rewrite this result as,

$$\int \left\{ \left(f + \frac{1}{f} \right) \mathbf{1}_{\{f \neq 1\}} + 2 \cdot \mathbf{1}_{\{f=1\}} \right\} \; d\mu = 2$$

$$\Longrightarrow \int \left(f + \frac{1}{f} - 2 \right) \mathbf{1}_{\{f \neq 1\}} \; d\mu = 0$$

$$\Longrightarrow \mu(\left(f + \frac{1}{f} - 2 \right) \mathbf{1}_{\{f \neq 1\}} = 0) = 1 \Longrightarrow \mu(\mathbf{1}_{\{f \neq 1\}} = 0) = 1$$

$\diamond$

Proposition D.3 *Let $\{F_{L,R,T}\}$ be a censoring model and $\widetilde{W}$ be a distribution function such that*

$$\iint_{\{(l,r):t \in \lfloor l,r \rfloor\}} \frac{dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r \rfloor)} = 1 \quad \textit{for any } t > 0 \textit{ such that } d\widetilde{W}(t) \neq 0$$

Then,

$$dF_{L_1,R_1,T_1}(l,r,t) = \frac{d\widetilde{W}(t) \; dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r \rfloor)}$$

defines a non–informative (constant–sum) censoring model which is indistinguishable from $\{F_{L,R,T}\}$.

Proof:

This proof is analogous to the one in Proposition 3.4. First we prove that dF_{L_1,R_1,T_1} defines a probability measure such that $T \in \lfloor L, R \rfloor$ with probability one,

$$\iiint_{\{(l,r,t):t\in\lfloor l,r\rfloor\}} dF_{L_1,R_1,T_1}(l,r,t)$$

$$= \int_0^{+\infty} d\widetilde{W}(t) \left(\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)} \right) = \int_0^{+\infty} d\widetilde{W}(t) = 1.$$

Then, we look at the marginal laws of the new censoring model,

$$dW_1(t) = \iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} dF_{L_1,R_1,T_1}(l,r,t)$$

$$= d\widetilde{W}(t) \left(\iint_{\{(l,r):t\in\lfloor l,r\rfloor\}} \frac{dF_{L_1,R_1}(l,r)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)} \right) = d\widetilde{W}(t)$$

and

$$dF_{L_1,R_1}(l,r) = \int_{\{t:t\in\lfloor l,r\rfloor\}} dF_{L_1,R_1,T_1}(l,r,t) =$$

$$= dF_{L,R}(l,r) \left(\int_{\{t:t\in\lfloor l,r\rfloor\}} \frac{d\widetilde{W}(t)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)} \right) = dF_{L,R}(l,r).$$

Thus, we have just showed that the censoring model F_{L_1,R_1,T_1} is indistinguishable from $F_{L,R,T}$. It just remains to see that $F_{L_1,R_1|T_1}$ satisfies equation (2) for any (l,r,t) such that $t \in \lfloor l,r \rfloor$ and $dW_1(t) \neq 0$,

$$dF_{L_1,R_1|T_1}(l,r|t) = \frac{dF_{L_1,R_1,T_1}(l,r,t)}{dW_1(t)} = \frac{d\widetilde{W}(t)\; dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)\; d\widetilde{W}(t)} = \frac{dF_{L,R}(l,r)}{P_{\widetilde{W}}(\lfloor l,r\rfloor)} = \frac{dF_{L_1,R_1}(l,r)}{P(T_1 \in \lfloor l,r\rfloor)}.$$

$\diamond$

Proposition D.4 *If $\{F_{L,R,T}\}$ is a censoring model such that L, R and T have the same finite supports, then equation (5) has always solution.*

Proof:

Let $I = \{x_1, \ldots, x_m\}$ be the finite support of L, R and T. Consider the problem of maximization of

$$l(\boldsymbol{w}) = l(w_1, \ldots, w_m) = \sum_{l,r \in I} \ln(P_W(\lfloor l, r\rceil)) dF_{L,R}(l,r)$$

over $w_i = P_W(x_i) \geq 0$ with $\displaystyle\sum_{i=1}^{m} w_i = 1$.

Notice that the directional derivative of this function $d_i(\boldsymbol{w})$ which considers the effect of increasing the i^{th} component by a small positive amount ϵ and divides all the components by $1+\epsilon$ in order to keep the sum equal to 1 corresponds to

$$d_i(\boldsymbol{w}) = \lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon}\, l\left(\frac{w_1}{1+\epsilon}, \ldots, \frac{w_i+\epsilon}{1+\epsilon}, \ldots, \frac{w_m}{1+\epsilon}\right) =$$

$$= \frac{\partial l(\boldsymbol{w})}{\partial w_i} - \sum_{j=1}^{m} w_j \frac{\partial l(\boldsymbol{w})}{\partial w_j} = \sum_{\substack{l,r \in I \\ x_i \in \lfloor l,r\rceil}} \frac{dF_{L,R}(l,r)}{P_W(\lfloor l,r\rceil)} - \sum_{j=1}^{m} w_j \sum_{\substack{l,r \in I \\ x_j \in \lfloor l,r\rceil}} \frac{dF_{L,R}(l,r)}{P_W(\lfloor l,r\rceil)} =$$

$$= \sum_{\substack{l,r \in I \\ x_i \in \lfloor l,r\rceil}} \frac{dF_{L,R}(l,r)}{P_W(\lfloor l,r\rceil)} - \sum_{l,r \in I} dF_{L,R}(l,r) \sum_{\substack{j=1,\ldots,m \\ x_j \in \lfloor l,r\rceil}} \frac{w_j}{P_W(\lfloor l,r\rceil)} =$$

$$= \sum_{\substack{l,r \in I \\ x_i \in \lfloor l,r\rceil}} \frac{dF_{L,R}(l,r)}{P_W(\lfloor l,r\rceil)} - \sum_{l,r \in I} dF_{L,R}(l,r) = \sum_{\substack{l,r \in I \\ x_i \in \lfloor l,r\rceil}} \frac{dF_{L,R}(l,r)}{P_W(\lfloor l,r\rceil)} - 1$$

Then the concavity of the function l($\boldsymbol{w}$) ensures a solution and the Kuhn–Tucker conditions are necessary and sufficient for optimality, that is, $\widetilde{\boldsymbol{w}}$ is a maximum if and only if, for every i, either $d_i(\widetilde{\boldsymbol{w}}) = 0$ or $d_i(\widetilde{\boldsymbol{w}}) \leq 0$ when $\widetilde{w}_i = 0$. Henceforth, it is obvious that the maximum $\widetilde{W}$ is a solution of equation (5).

$\diamond$