Title

Semiparametric estimation for the additive hazards model with left-truncated and right-censored data.

Journal

Biometrika, 100(4)

ISSN

0006-3444

Authors

Huang, Chiung-Yu
Qin, Jing

Publication Date

2013

DOI

10.1093/biomet/ast039

Peer reviewed

HHS Public Access

Author manuscript

Biometrika. Author manuscript; available in PMC 2015 August 03.

Published in final edited form as:

Biometrika. 2013 ; 100(4): 877–888. doi:10.1093/biomet/ast039.

Semiparametric estimation for the additive hazards model with left-truncated and right-censored data

CHIUNG-YU HUANG and

Division of Oncology Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A

JING QIN

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A

CHIUNG-YU HUANG: cyhuang@jhmi.edu; JING QIN: jingqin@niaid.nih.gov

Summary

Survival data from prevalent cases collected under a cross-sectional sampling scheme are subject to left-truncation. When fitting an additive hazards model to left-truncated data, the conditional estimating equation method (Lin & Ying, 1994), obtained by modifying the risk sets to account for left-truncation, can be very inefficient, as the marginal likelihood of the truncation times is not used in the estimation procedure. In this paper, we use a pairwise pseudolikelihood to eliminate nuisance parameters from the marginal likelihood and, by combining the marginal pairwise pseudo-score function and the conditional estimating function, propose an efficient estimator for the additive hazards model. The proposed estimator is shown to be consistent and asymptotically normally distributed with a sandwich-type covariance matrix that can be consistently estimated. Simulation studies show that the proposed estimator is more efficient than its competitors. A data analysis illustrates application of the method.

Keywords

Canadian Study of Health and Aging; Composite likelihood; Estimating equation; Martingale; Prevalent sampling

1. Introduction

For left-truncated and right-censored data, conventional methods of statistical inference are usually based on the conditional likelihood, conditioning on the truncation times. Suppose that we are interested in estimating the distribution of the survival time T^* in a target population. Let A^* be an independent truncation time, so that (T^*, A^*) is observed if and only if $T^* \geq A^*$. Let (T, A) denote the observed survival and truncation times; then (T, A) has the same joint distribution as (T^*, A^*) given that $T^* \geq A^*$. Let f and S denote the density and survival functions of the survival time T^*, and let h denote the density function of the truncation time A^*. Then the joint density function of (T, A) evaluated at (t, a) can be expressed as

$$\frac{f(t)\,h\,(a)}{\int_0^\infty S(u)\,h\,(u)\,\mathrm{d}u}=\frac{f(t)}{S\,(a)}\times\frac{S(a)\,h\,(a)}{\int_0^\infty S(u)\,h\,(u)\,\mathrm{d}u}\quad(0\le a\le t),$$

where $f(t)/S(a)$ is the conditional density of T given A and $S(a)\,h\,(a)/\int_0^\infty S(u)\,h\,(u)\,\mathrm{d}u$ is the marginal density of A. Denote by C the potential censoring time after study enrolment; that is, the observation of the residual survival time $T - A$ is subject to censoring time C. Let $Y = \min(T, A + C)$ be the observed survival time and let $\Delta = I\,(T \le A + C)$ be the indicator function of a failure event. Under an assumption of independent censoring, the full likelihood function of independent and identically distributed data (Y_i, A_i, Δ_i) $(i = 1, \ldots, n)$ can be decomposed as $L_F = L_C \times L_M$, where

$$L_{\mathrm{C}}=\prod_{i=1}^{n}\left\{\frac{f(Y_i)}{S(A_i)}\right\}^{\Delta_i}\left\{\frac{S(Y_i)}{S(A_i)}\right\}^{1-\Delta_i}$$

is the conditional likelihood of (Y, Δ) given the truncation time A and

$$L_{\mathrm{M}}=\prod_{i=1}^{n}\frac{S(A_i)\,h\,(A_i)}{\int_0^\infty S(u)\,h\,(u)\,\mathrm{d}u}$$

is the marginal likelihood of A. In cases where the truncation time density function h is left unspecified, Wang (1991, § 3) showed that the conditional nonparametric likelihood L_C is fully efficient with respect to the full nonparametric likelihood L_F.

When the effects of a $p \times 1$ vector of covariates Z^* are modelled through the proportional hazards model (Cox, 1972)

$$\lambda(t|z)=\lambda_0\,(t)\,exp(\beta_0' z),$$

where $\lambda\,(t \mid z)$ is the conditional hazard function of T^* given $Z^* = z$, the profile likelihood after profiling out the baseline hazard function $\lambda_0\,(t)$ from the conditional likelihood L_C is equivalent to the partial likelihood for the truncated data. Wang et al. (1993, Property 3.1.1) and Kalbfleisch & Lawless (1991) further showed that the maximum partial likelihood estimator is fully efficient with respect to the conditional likelihood L_C. However, it is known that the information loss due to ignoring the information about β_0 in the marginal likelihood L_M can be substantial (Huang et al., 2012), especially when the truncation time has a known distribution.

In many applications, the appropriateness of the Cox model may be questionable, as the assumption of multiplicative covariate effects can be violated, especially when continuous covariates are involved. The additive hazards model (Aalen, 1980; Cox & Oakes, 1984; Thomas, 1986; Breslow & Day, 1987; Lin & Ying, 1994; Martinussen & Scheike, 2002a), which focuses on modelling the difference in the risk, has been regarded as an appealing alternative because researchers are often more interested in the risk difference attributed to the risk factors. Specifically, the additive hazards model assumes that the conditional hazard function of T^* given $Z^* = z$ takes the form

$$\lambda(t|z)=\lambda_0(t)+\beta_0' z, \quad (1)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β_0 is a $p \times 1$ vector of parameters. Define $\Lambda_0(t)=\int_0^t \lambda_0(u)\,\mathrm{d}u$, so that $\Lambda_0(t)$ is the baseline cumulative hazard function. Model (1) is equivalent to assuming that the conditional survival function of T^* given $Z^* = z$ is

$$S(t|z)=S_0(t)\,exp(-\beta_0' zt),$$

where $S_0(t) = \exp\{-\Lambda_0(t)\}$ is the baseline survival function.

Direct maximization of the full likelihood under the additive hazard model with respect to (λ_0, β_0) is computationally cumbersome because it involves the nonparametric component λ_0 in a complicated way; moreover, to the best of our knowledge, the asymptotic properties of the maximum likelihood estimator have not been formally studied. A natural idea would be to extend existing methods for right-censored data to accommodate left-truncation. For example, one could estimate the additive hazards model by further conditioning the estimating function proposed by Lin & Ying (1994) on the truncation time A. This estimating function is an analogue of the partial likelihood score function under the Cox model. It is expected that the extended conditional estimating equation approach is not efficient, as it is not based on maximizing the conditional likelihood L_C or the full likelihood L_F. Moreover, the information about β_0 in the marginal likelihood is not used in the estimation procedure. In this paper, we apply the pairwise likelihood method to eliminate the nuisance parameter λ_0 from the marginal likelihood, and we combine the conditional estimating function and the marginal pairwise pseudo-score function to improve efficiency in estimating β_0. Our method provides a computationally tractable parameter estimator and, as demonstrated by the simulation studies, offers substantial efficiency gains over the conditional estimating equation approach. We believe that the proposed method provides a useful tool for studying left-truncated and right-censored survival data.

2. Model estimation

2.1. Conditional estimating equation method

In the absence of truncation, or equivalently when $A \equiv 0$, Lin & Ying (1994) obtained closed-form estimators for the regression parameters β_0 and the cumulative baseline hazard function $\Lambda_0(t)$ for the additive hazards model (1) via the estimating equation approach. By further conditioning on the truncation time A, one can modify the estimating equations to accommodate left-truncation. Specifically, define a counting process for the observed failure events, $N_i(t) = \Delta_i I(Y_i \le t)$, and for the at-risk process, $R_i(t) = I(A_i \le t \le Y_i)$. Let Z_i be the covariate of subject i. Assume that the censoring time C is noninformative in the sense that $\mathrm{pr}\{T^* \in [t, t + \Delta t) \mid T^* \ge t, T^* \ge A^*, C \ge t, Z^*\} = \mathrm{pr}\{T^* \in [t, t + \Delta t) \mid T^* \ge t, T^* \ge A^*, Z^*\}$. It can be verified that $M_i(t)=N_i(t) - \int_0^t R_i(u)\{d\Lambda(u)+\beta' Z_i \,\mathrm{d}u\}$ is a local square-integrable martingale when $\beta = \beta_0$ and $\Lambda(t) = \Lambda_0(t)$, suggesting that one may estimate β_0 and Λ_0 by solving the two estimating equations $\sum_{i=1}^{n}\int_0^t \mathrm{d}\,M_i\,(u)=0$ and $\sum_{i=1}^{n}\int_0^\tau Z_i\,\mathrm{d}\,M_i\,(u)=0$, where $t \in [0, \tau]$ and τ is a prespecified time-point. Solving the first estimating equation $\sum_{i=1}^{n}\int_0^t \mathrm{d}\,M_i\,(u)=0$ yields

$$\hat{\Lambda}\,(t,\beta)=\int_0^t \frac{\sum_{i=1}^{n}\{\mathrm{d}\,N_i\,(u)-R_i\,(u)\,\beta^{'}\,Z_i\,\mathrm{d}u\}}{\sum_{i=1}^{n}R_i\,(u)}. \quad (2)$$

Substituting $\hat{\Lambda}(t, \beta)$ into the second estimating equation yields a closed-form estimating equation $\sum_{i=1}^{n}\phi_i\,(\beta)=0$, where

$$\phi_i\,(\beta)=\int_0^\tau\{Z_i-\overline{Z}(t)\}\,\mathrm{d}\,M_i\,(t), \quad (3)$$

with $\overline{Z}(t)=\{\sum_{i=1}^{n}Z_i\,R_i\,(t)\}/\{\sum_{i=1}^{n}R_i\,(t)\}$. The estimating function (3) does not depend on the nonparametric component Λ. In fact, solving $\sum_{i=1}^{n}\phi_i\,(\beta)=0$ for β yields an estimate

$$\hat{\beta}_\phi=\left[\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}(t)\}^{\otimes 2}R_i\,(t)\,\mathrm{d}t\right]^{-1}\left[\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}\,(t)\}\,\mathrm{d}\,N_i\,(t)\right]$$

for β_0, where $a^{\otimes 2} = aa'$. Define the estimating function $\phi\,(\beta)=n^{-1}\sum_{i=1}^{n}\phi_i\,(\beta)$. It follows from standard counting process theorems that $n^{1/2}\phi(\beta_0)$ converges in distribution to a zero-mean multivariate normal distribution with variance-covariance matrix B_1, where B_1 can be consistently estimated by $\hat{B}_1=n^{-1}\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}(t)\}^{\otimes 2}\,\mathrm{d}\,N_i\,(t)$, provided that the Z_i (i = 1, …, n) are bounded. We can show by Taylor series expansion and the central limit theorem that $n^{1/2}(\hat{\beta}_\phi - \beta_0)$ converges weakly to a zero-mean multivariate normal distribution with variance-covariance matrix $B_2^{-1}\,B_1\,B_2^{-1}$, which can be consistently estimated by $\hat{B}_2^{-1}\,\hat{B}_1\,\hat{B}_2^{-1}$ where $\hat{B}_2=n^{-1}\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}\,(t)\}^{\otimes 2}R_i\,(t)\,\mathrm{d}t$.

Although the maximum partial likelihood estimator under the Cox model is fully efficient with respect to the truncation likelihood of (Y, Δ) given A (Wang et al., 1993), the efficiency can be improved greatly if the information about β_0 in the marginal likelihood can be used in the estimation procedure. However, under the Cox model, it is unclear how this can be achieved because the marginal likelihood involves $\Lambda_0\,(t)$ in a complicated way. On the other hand, for the additive hazards model, we shall show in the next section that we are able to incorporate the marginal likelihood in the estimation procedure.

2.2. A combined estimating equation approach

Assume that the density function h of the underlying truncation time does not depend on Z^* and is not degenerate. Under the additive hazards model, the marginal density function of A given $Z = z$ is

$$\frac{S(a|z)\,h\,(a)}{\mu\,(z)}=\frac{S_0\,(a)\,exp\,(-\beta_0^{'}\,za)\,h\,(a)}{\mu\,(z)},$$

where $\mu(z)=\int_0^\infty S(u|z)h\,(u)\,\mathrm{d}u=\int_0^\infty S_0\,(u)\,exp\,(-\beta_0^{'}\,zu)\,h\,(u)\,\mathrm{d}u$. As a result, the conditional density function of A given $A \le \tau$ and $Z = z$ is

$$\frac{S(a|z)\,h\,(a)}{\int_0^\tau S(u|z)\,h\,(u)\,\mathrm{d}u}=\frac{S_0(a)\,exp\,(-\beta_0^{'}\,za)\,h\,(a)}{\int_0^\tau S_0(u)\,exp\,(-\beta_0^{'}\,zu)\,h\,(u)\,\mathrm{d}u}\quad (a \le \tau).$$

Because τ is usually set to be the maximum of the observed survival times, all observed truncation times satisfy $A \le \tau$ in practice. Obtaining an estimate for β_0 by maximizing the marginal likelihood of A is a challenge because the integral in the denominator of the conditional likelihood does not have a closed form with λ_0 and h unspecified. In the spirit of the pairwise pseudo-likelihood method (Kalbfleisch, 1978; Liang & Qin, 2000), we propose an alternative estimation procedure that does not involve the nonparametric components λ_0 and h and thus has the advantage of computational convenience.

As in Liang & Qin (2000), we apply the conditional argument of Kalbfleisch (1978) in a pairwise fashion to eliminate nuisance parameters in the marginal distribution of A. By further conditioning on having observed the values $\{A_i, A_j\}$ for a given pair but without knowing the order, the pairwise pseudolikelihood of (A_i, A_j) conditional on $A_i \le \tau$, $A_j \le \tau$ and (Z_i, Z_j), for $i < j$, is

$$\frac{\frac{S(A_i|Z_i)\,h\,(A_i)}{\int_0^\tau S(u|Z_i)\,h\,(u)\,\mathrm{d}u}\times\frac{S(A_j|Z_j)\,h\,(A_j)}{\int_0^\tau S(u|Z_j)\,h\,(u)\,\mathrm{d}u}}{\frac{S(A_i|Z_i)\,h\,(A_i)}{\int_0^\tau S(u|Z_i)\,h\,(u)\,\mathrm{d}u}\times\frac{S(A_j|Z_j)\,h\,(A_j)}{\int_0^\tau S(u|Z_j)\,h\,(u)\,\mathrm{d}u}+\frac{S(A_i|Z_j)\,h\,(A_i)}{\int_0^\tau S(u|Z_j)\,h\,(u)\,\mathrm{d}u}\times\frac{S(A_j|Z_i)\,h\,(A_j)}{\int_0^\tau S(u|Z_i)\,h\,(u)\,\mathrm{d}u}},$$

and this equals

$$\frac{exp\,(-\beta_0^{'}\,Z_i\,A_i-\beta_0^{'}\,Z_j\,A_j)}{exp\,(-\beta_0^{'}\,Z_i\,A_i-\beta_0^{'}\,Z_j\,A_j)+exp\,(-\beta_0^{'}\,Z_j\,A_i-\beta_0^{'}\,Z_i\,A_j)}=\frac{1}{1+exp\,\{\beta_0^{'}\,(Z_i-Z_j)\,(A_i-A_j)\}}$$

for $A_i, A_j \le \tau$. Interestingly, the pairwise pseudolikelihood depends on the regression parameter β but not on the baseline hazard function $\lambda\,(t)$ nor on the truncation time density function h. Define the function $\rho_{ij} = \rho(A_i, Z_i, A_j, Z_j) = (Z_i - Z_j)(A_i - A_j)$. We estimate β_0 by maximizing the log pairwise pseudolikelihood

$$\sum_{1\le i<j\le n} -\;log\left\{1+exp(\beta^{'}\,\rho_{ij})\right\}. \quad (4)$$

To derive the maximum pairwise pseudolikelihood estimator $\hat{\beta}_\psi$ for β_0, one can solve the normalized pseudo-score equation $\psi(\beta) = 2\{n(n-1)\}^{-1}\sum_{1\le i<j\le n}\psi_{ij}(\beta) = 0$ where

$$\psi_{ij}(\beta)=\psi_{ij}(\beta;A_i,Z_i,A_j,Z_j)=\frac{-\rho_{ij}}{1+exp(-\beta'\rho_{ij})}. \quad (5)$$

The proposed method can be applied even when there is no additional follow-up after enrolment. In this case, the conditional distribution of Y given A is degenerate, as $Y = A$ is observed; hence the conditional estimating equation method does not work, and the inference can be based only on the marginal likelihood of A. Moreover, the method works when A^* has a discrete distribution.

Because $-\log\{1 + \exp(\beta'\rho_{ij})\}$ is the loglikelihood of A_i and A_j conditional on $A_i \le \tau$, $A_j \le \tau$, $\{Z_i, Z_j\}$ and the order statistics of $\{A_i, A_j\}$, the pairwise pseudolikelihood (4) achieves its maximum at the true parameter value as $n \to \infty$. By the conditional Kullback–Leibler information inequality (Andersen, 1970), the maximum pairwise pseudolikelihood estimator $\hat{\beta}_\psi$ is consistent for β_0. Next, it is easy to see that $\psi_{ij}(\beta)$ is permutation-symmetric in its arguments (A_i, Z_i) and (A_j, Z_j), and hence $\psi(\beta)$ is a U-statistic of order 2. It can further be shown that $E\{\psi_{ij}(\beta_0)\} = 0$ and $E\{\psi_{ij}(\beta_0)^{\otimes 2}\} < \infty$, provided that the covariate is bounded. Applying the projection method developed by Hoeffding (1948), we can show that $n^{1/2}\psi(\beta_0)$ converges to a normal distribution with mean zero and variance-covariance matrix $V_1 = 4E\{\psi_{12}(\beta_0)'\psi_{13}(\beta_0)\}$. The asymptotic property of $\hat{\beta}_\psi$ can be established using the delta method. In fact, $n^{1/2}(\hat{\beta}_\psi - \beta_0)$ converges to a zero-mean multivariate normal distribution with variance-covariance matrix $\sum_\phi=V_2^{-1}V_1V_2^{-1}$, where $V_1 = 4E\{\psi_{12}(\beta_0)\psi_{13}(\beta_0)\}$ and $V_2=-E\{\partial\psi_{12}(\beta_0)/\partial\beta\}=E[\rho_{12}^{\otimes 2}exp(-\beta_0'\rho_{12})/\{1+exp(-\beta_0'\rho_{12})\}^2]$ can be consistently estimated (Sen, 1960) by

$$\hat{V}_1=\frac{4}{n-1}\sum_{i=1}^{n}\left\{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\psi_{ij}(\hat{\beta}_\psi)\right\}^{\otimes 2}$$

and

$$\hat{V}_2=\frac{1}{n(n-1)}\sum_{i\neq j}\frac{\rho_{ij}^{\otimes 2}exp(-\hat{\beta}_\psi'\rho_{ij})}{\{1+exp(-\hat{\beta}_\psi'\rho_{ij})\}^2}.$$

Both the estimating function φ and the pseudo-score function ψ yield consistent estimates of β_0. To combine the information about β_0 in the marginal likelihood and that in the conditional estimating function (3), we propose to estimate β_0 by solving

$$\xi(\beta)=\phi(\beta)+\psi(\beta)=0$$

for β. Let $\hat{\beta}$ be the proposed estimator satisfying $\xi(\hat{\beta}) = 0$. Because $\varphi(\beta)$ is linear in β, the consistency of $\hat{\beta}$ follows directly from the consistency of $\hat{\beta}_\psi$. With the asymptotic normality properties of $\hat{\beta}_\varphi$ and $\hat{\beta}_\psi$ established above, the asymptotic normality of $\hat{\beta}$ follows from the asymptotic independence (van der Vaart & Wellner, 1996, Example 1.4.6) of $\varphi(\beta)$ and $\psi(\beta)$. The large-sample properties can also be established by using the results that $\xi(\beta) = 2\{n(n-1)\}^{-1}\sum_{1\le i<j\le n}[\{\varphi_i(\beta)+\varphi_j(\beta)\}/2+\psi_{ij}(\beta)]$ is a U-statistic of order 2 and that $\varphi_i(\beta_0)$ and $\psi_{ij}(\beta_0)$ are orthogonal. Specifically, by using the facts that $E\{\mathrm{d}M_i(t) \mid A_i, Z_i\} = 0$ and that $\psi_{ij}(\beta_0)$ only involves (A_i, Z_i) and (A_j, Z_j), by double expectation we have that $E\{\varphi(\beta_0)'\psi(\beta_0)\} = 0$ and $\mathrm{var}\{\xi(\beta_0)\} = \mathrm{var}\{\varphi(\beta_0)\} + \mathrm{var}\{\psi(\beta_0)\}$. By the central limit theorem for U-statistics, $n^{1/2}\xi(\beta_0)$ converges in distribution to a normal distribution with mean zero and variance-covariance matrix $\mathrm{var}\{n^{1/2}\varphi(\beta_0)\} + \mathrm{var}\{n^{1/2}\psi(\beta_0)\} = B_1 + V_1$. Then the asymptotic normality of $\hat{\beta}$ follows upon applying a Taylor expansion to $\xi(\beta)$. The large-sample properties of the proposed estimator $\hat{\beta}$ are summarized in the following theorem.

Theorem 1—*Assume the following regularity conditions:* (a) β_0 *lies in a compact set* B*;* (b) Z *is bounded; and* (c) *the matrix* $B_2 + V_2$ *is positive definite. Then, as* $n \to \infty$, $n^{1/2}(\hat{\beta} - \beta_0)$ *converges in distribution to a zero-mean multivariate normal distribution with variance matrix* $(B_2+V_2)^{-1}(B_1+V_1)(B_2+V_2)^{-1}$.

The regression parameter β_0 may not be identifiable based on the marginal likelihood when the underlying truncation time distribution A^* depends on the covariates Z^*. In contrast, β_0 can be estimated consistently by applying the conditional estimating equation method even if A^* is correlated with Z^*, as long as the conditional independence of T^* and A^* given Z^* holds.

An estimator of the baseline cumulative hazard function can be obtained by replacing β in (2) with $\hat{\beta}$, that is, $\hat{\Lambda}(t, \hat{\beta})$. Applying counting process theory, we can show that $n^{1/2}\{\hat{\Lambda}(t, \hat{\beta}) - \Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process on $[0, \tau]$. The proof of the asymptotic properties of $\hat{\Lambda}(t, \hat{\beta})$ follows closely that for the estimated baseline cumulative hazard function in Lin & Ying (1994), and thus we omit it. Finally, as suggested by a referee, the cumulative distribution function of A^*, $H(a) = \int_0^a h(u)\,\mathrm{d}u$, can be estimated by

$$\hat{H}(a) = \frac{\sum \hat{S}_0^{-1}(A_i)\, exp(\hat{\beta}' Z_i A_i)\, I(A_i \le a)}{\sum \hat{S}_0^{-1}(A_i)\, exp(\hat{\beta}' Z_i A_i)}$$

where $\hat{S}_0(t) = \exp\{-\hat{\Lambda}(t, \hat{\beta})\}$.

Naturally, we can consider applying the conditional argument to three or more truncation times by conditioning on their order statistics and expect a potential gain in efficiency. However, doing so increases the computational burden, while the small gain in efficiency would not be sufficient to justify the use of a more complicated procedure. Readers are referred to Diao et al. (2012) for additional discussion.

We can apply the generalized method of moments (Hansen, 1982) to further improve estimation efficiency. Define $\eta(\beta) = \{\varphi(\beta)', \psi(\beta)'\}'$ and let W be a positive-definite weight function matrix. A consistent estimator of β_0 can be obtained by minimizing $\hat{\beta}_W = \arg\min_\beta \eta(\beta)'W^{-1}\eta(\beta)$. The optimal matrix which yields an efficient estimator is $W = \mathrm{var}\{\eta(\beta)\}$. Using the optimal weight function matrix and the fact that $\varphi(\beta_0)$ and $\psi(\beta_0)$ are orthogonal,

$n^{1/2}(\hat{\beta}_W - \beta_0)$ converges in distribution to a zero-mean multivariate normal distribution with variance matrix

$$[E\{\partial_\eta(\beta_0)/\partial\beta\}' W(\beta_0)^{-1} E\{\partial_\eta(\beta_0)/\partial\beta\}]^{-1} = (B_2 B_1^{-1} B_2 + V_2 V_1^{-1} V_2)^{-1} \text{ as } n \to \infty.$$

3. Simulations and data analysis

3.1. Simulations

We conducted two sets of Monte Carlo simulations to examine the finite-sample performance of our method. In the first set of simulations, the time-independent covariate Z^* was generated from a Un(0, 1) random variable. The survival time T^* was generated from the additive hazards models $\lambda(t \mid Z) = 1 + \beta_0 Z$ and $\lambda(t \mid Z) = 0.5\, t^{1/2} + \beta_0 Z$, with $\beta_0 = 1$ in both scenarios. The two models correspond to constant and increasing hazards. The underlying truncation time A^* was independently generated from a Un(0, 100) random variable and an exponential random variable with a mean of 10. To form a prevalent cohort of sample size n, realizations of (A^*, T^*, Z^*) were generated until n subjects satisfied the sampling constraint $A^* \leq T^*$. The censoring time for the residual survival time $T^* - A^*$ was generated from a uniform distribution, Un(0, τ_c), where τ_c was selected so that the censoring rate was approximately 0%, 30% or 50%. In each simulation, we generated 1000 datasets, each with $n = 200$.

Four different methods were applied to estimate the regression parameter: (a) $\hat{\beta}_\varphi$, the conditional estimating equation estimator (Lin & Ying, 1994) obtained by solving $\varphi(\beta) = 0$; (b) $\hat{\beta}_\psi$, the marginal pairwise pseudolikelihood estimator obtained by solving $\psi(\beta) = 0$; (c) $\hat{\beta}$, the proposed estimator obtained by solving $\varphi(\beta) + \psi(\beta) = 0$; and (d) $\hat{\beta}_W$, the generalized method of moments estimator. Table 1 summarizes the empirical bias, standard error and relative efficiency of the four estimators. All are close to their estimands under different truncation time distributions. Compared with the conditional estimating equation estimator $\hat{\beta}_\varphi$, the relative efficiency of the marginal pairwise pseudolikelihood estimator $\hat{\beta}_\psi$ and the proposed estimator $\hat{\beta}$ increases with the censoring rate. This is not surprising, as the truncation time A is observed for all subjects, while the uncertainty in the conditional estimating equation increases with the censoring rate. The efficiency gain in the generalized method of moments estimator $\hat{\beta}_W$ was not as high as that of the combined estimating equation estimator $\hat{\beta}$. Intuitively, this is because estimation of the optimal weight function involves estimation of the second moments of $\varphi(\beta_0)$ and $\psi(\beta_0)$, which requires a larger sample size to obtain the benefits of an efficient generalized method of moments estimator.

The second set of simulations investigates the efficiency gain for truncation time distributions with different skewness coefficients. Table 2 summarizes the simulation results for the conditional estimating equation approach, the marginal pairwise pseudolikelihood approach, and the proposed estimator under the additive hazards model $\lambda(t \mid Z) = 1 + Z$, where Z was generated from a Un(0, 1) random variable. In this set of simulations, the censoring percentage was set to zero. For the continuous case, we simulated $A^*/2$ from beta distributions with parameter values (1, 1), (4, 1) and (1, 4), which illustrate uniform, negatively skewed and positively skewed distributions. Similarly, for the discrete case, we simulated A^* from a discrete uniform distribution taking values on $\{0, 1, 2\}$, as well as from binomial distributions with size 2 and success probabilities 0.25 and 0.75; these also illustrate uniform, negatively skewed and positively skewed distributions. Together with Table 1, these simulation results suggest that the efficiency of the estimation procedures increases with the upper limit of the support of the underlying truncation time distribution. Moreover, negatively skewed truncation time distributions lead to higher efficiency than positively skewed ones.

3.2. The Canadian Study of Health and Aging

We illustrate the proposed estimation procedures by analysing data from the Canadian Study of Health and Aging, one of the largest epidemiological studies of dementia. Alzheimer's disease and vascular dementia are the top two leading causes of dementia affecting the elderly. Alzheimer's disease, which accounts for approximately 50–70% of all dementia diagnoses, destroys nerve cells and thus causes the brain to degenerate. It is a progressive disease, eventually leading to a loss of ability to perform daily living tasks. Vascular dementia, accounting for 20–30% of the cases, is caused by stroke or small-vessel disease that interrupts the supply of oxygen to the brain and damages the cortex, which is associated with learning, memory and language. One frequently raised question is the impact of dementia on life expectancy. People with dementia have reduced survival compared with those without dementia. Moreover, studies suggest that older people with vascular dementia have worse survival than those with Alzheimer's disease (Wolfson et al., 2001). Compared to Alzheimer's disease, however, vascular dementia has been understudied.

In the first phase of the Canadian Study of Health and Aging, a total of 1132 persons aged 65 and older with dementia were identified by surveying an age-stratified random sample of 9008 community residents and 1255 residents of institutions in Canada. For each dementia case, a diagnosis of possible Alzheimer's disease, probable Alzheimer's disease, or vascular dementia was assigned, and the date of dementia onset was determined by interviewing care-givers. Information on mortality was obtained at the follow-up data collection in 1996. As pointed out in Wolfson et al. (2001), the Canadian Study of Health and Aging had a prevalent cohort study design, because survival data were collected from a prevalent cohort of dementia patients who had not experienced the failure event, death, at the time of recruitment. Hence the survival time is subject to left-truncation, where the truncation time is the duration from the onset of dementia to enrolment. The observation of the residual survival time after enrolment is censored by the end of the follow-up.

Our primary interest in this analysis is to examine whether people with vascular dementia have a higher risk of death than those with Alzheimer's disease. We considered a subset of the study data by excluding those with missing date of onset or classification of dementia subtype. Moreover, as in Wolfson et al. (2001), those with observed survival time of 20 or more years were excluded because these subjects are considered unlikely to have Alzheimer's disease or vascular dementia. The stationarity assumption that the incidence of dementia is constant over time was found to be reasonably met for these data using the method described in Wang (1991). Thus a total of 807 dementia patients were included in our analysis. Among them, 637 had a diagnosis of probable/possible Alzheimer's disease and 170 had a diagnosis of vascular dementia. In the second phase of the study, a total of

627 deaths were recorded, of which 491 subjects had a diagnosis of Alzheimer's disease and 136 of vascular dementia.

We fitted the additive hazards model to the subset of the dementia study and applied the methods presented in § 2 to estimate the regression parameter. To obtain a 95% confidence interval for the estimate, we adopted a nonparametric bootstrap method by sampling 807 subjects with replacement from the dataset. The resampling procedure was repeated 2000 times, and the 95% confidence interval was constructed by using the 2.5th and 97.5th percentiles of the 2000 estimates. We first applied the conditional estimation equation method (Lin & Ying, 1994) to compare vascular dementia and Alzheimer's disease in terms of the risk of death. The result suggests that vascular dementia is associated with worse survival, with an estimated risk difference of 0.025. However, the difference is not significant, as the corresponding 95% bootstrap confidence interval (−0.021, 0.078) covers zero. On the other hand, the proposed combined estimating equation method estimates a significant risk difference of 0.051, with 95% bootstrap confidence interval (0.006, 0.106). For comparison, we also fitted the Cox model. The estimates from the additive hazards and Cox models have the same signs, indicating the same directions of the covariate effects.

Graphical checking of the additivity assumption can be performed as follows. First, estimate the survival functions for vascular dementia and probable/possible Alzheimer's disease using the truncation product-limit estimator. Then obtain the estimated regression coefficient $\hat{\beta}_{\varphi}$ using the conditional estimating equation method. Let $\hat{S}_0(t)$ and $\hat{S}_1(t)$ be the estimated survival curves for vascular dementia and probable/possible Alzheimer's disease. Plot $\hat{S}_1(t)$ and $\hat{S}_0(t)\exp(-\hat{\beta}_{\varphi} t)$ against time t. Figure 1 shows that, except for the first two years after onset, the two curves almost overlap, thus suggesting that the additivity assumption is reasonably met for the Canadian Study of Health and Aging data. Model-checking methods will be investigated elsewhere.

4. Remarks

For right-censored survival data, it is well known that the censoring time distribution can be factored out from the full likelihood, so an attempt to model the censoring time distribution does not affect the derivation of the maximum likelihood estimator for the survival time distribution. For left-truncated data, when the underlying truncation time random variable A^* is allowed to depend on the covariate Z^* in an arbitrary way, the marginal likelihood L_{M} can be shown to be ancillary with respect to the full likelihood L_{F} by using the weak ancillarity argument of Wang et al. (1993). In other words, maximizing the conditional likelihood L_{C} would yield fully efficient estimation when h is allowed to depend on the covariate. On the other hand, when A^* is assumed to be independent of Z^* or when the conditional distribution of A^* given Z^* is parameterized, potential efficiency gains can be achieved by incorporating the information about β_0 in L_{M}. An important example is the special case where the underlying truncation time is known to have a uniform distribution. In this case, the survival times T can be viewed as a biased sample of the T^*s, where the sampling weight is proportional to the length of the survival time. Various authors, including Vardi (1989), Asgharian et al. (2002) and Qin et al. (2011), have considered nonparametric and semiparametric methods that exploit knowledge of the truncation time distribution to improve efficiency in estimating the survival time model.

Under the additive hazards model, the estimating equation method (Lin & Ying, 1994) does not maximize L_C, so there is room for improvement. When A^* is independent of the covariate, the weak ancillary argument of Wang et al. (1993) fails to hold, and the marginal likelihood is informative about β. In this paper we show that under the additive hazards model, efficiency can be increased even when the truncation time distribution h is unspecified, provided h does not depend on the covariate. The proposed estimator combines the conditional estimating function, constructed based on the distribution of (Y, Δ) conditional on A, and the pairwise pseudo-score function, constructed based on the marginal distribution of A, and has been shown via simulations to enjoy substantial gains in efficiency over the conditional estimating equation approach. The proposed method does not work when the covariate distribution is degenerate or when the distribution of the underlying truncation time is degenerate. As in conventional survival analysis, the efficiency of the proposed estimation procedures increases with the variability in the covariate distribution, although the relative efficiency compared to the conditional estimating equation approach does not necessarily change in the same direction. Similarly, the efficiency gain obtained through employing the pairwise pseudolikelihood depends on the underlying truncation time distribution. In general, truncation time distributions with larger support have a higher efficiency gain, and negatively skewed distributions have a higher relative efficiency than positively skewed ones.

Although we present only the results for time-independent covariates $Z_i(t)$, the proposed estimation procedure can be applied to handle time-dependent covariates by replacing ρ_{ij} in the estimating function (5) with $\int_0^\tau \{Z_i(u) - Z_j(u)\}\{I(A_i \geq u) - I(A_j \geq u)\}\,\mathrm{d}u$. We believe that the proposed method provides a useful tool for studying left-truncated and right-censored survival data. It would be interesting to extend our approach to other semiparametric models, including the additive-multiplicative hazard models considered by Lin & Ying (1995) and Martinussen & Scheike (2002b), among others.

As suggested by a referee, a model-checking method that examines the assumption of independence of A^* and Z^* can be formulated as follows. Suppose that A^* given $Z^* = z$ follows a semiparametric proportional likelihood ratio model (Luo & Tsai, 2012; Diao et al., 2012)

$$h(a|z) = \frac{exp(-\gamma_0 az)\,h_0(a)}{\int_0^\infty exp(-\gamma_0 uz)\,h_0(u)\,\mathrm{d}u},$$

where h_0 is an arbitrary density function. Then the marginal density of A given $Z = z$ is

$$\frac{S(a|z)h(a|z)}{\int_0^\infty S(u|z)\,h(u|z)\,\mathrm{d}u} = \frac{S_0(a|z)\,h_0(a)\,exp\{-(\beta_0'+\gamma_0')za\}}{\int_0^\infty S_0(u)\,h_0(u)\,exp\{-(\beta_0'+\gamma_0')zu\}\,\mathrm{d}u}.$$

The maximum pairwise likelihood estimator $\hat{\beta_\psi}$ converges to $\beta_0 + \gamma_0$ as $n \to \infty$. We can test the null hypothesis $\gamma_0 = 0$ by considering $\hat{\beta_\psi} - \hat{\beta_\varphi}$.

Acknowledgments

The authors thank Professors Christina Wolfson, Masoud Asgharian and Ian McDowell for kindly sharing the Canadian Study of Health and Aging data. The core study was funded by the National Health Research and Development Program of Health Canada Project. Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, Bayer Incorporated, and the British Columbia Health Research Foundation Projects. The authors also thank the referees, associate editor, editor, and Dr. Dean Follmann for their comments, which have helped to improve the presentation of this article. This work was conducted when the first author was a mathematical statistician at the National Institute of Allergy and Infectious Diseases.

References

Aalen, OO. A Model for Nonparametric Regression Analysis of Counting Processes. Vol. 2. New York: Springer; 1980. Lecture Notes in Statistics

Andersen EB. Asymptotic properties of conditional maximum-likelihood estimators. J R Statist Soc B. 1970; 32:283–301.

Asgharian M, M'Lan CE, Wolfson DB. Length-biased sampling with right censoring: An unconditional approach. J Am Statist Assoc. 2002; 97:201–9.

Breslow, NE.; Day, NE. Statistical Methods in Cancer Research : The Design and Analysis of Cohort Studies. Vol. II. Lyon: International Agency for Research on Cancer; 1987. IARC Scientific Publications No. 82

Cox DR. Regression models and life-tables. J R Statist Soc B. 1972; 34:187–220. with Discussion.

Cox, DR.; Oakes, D. Analysis of Survival Data. London: Chapman & Hall; 1984.

Diao G, Ning J, Qin J. Maximum likelihood estimation for semiparametric density ratio model. Int J Biostatist. 2012; 8 Article 16.

Hansen LP. Large sample properties of generalized method of moments estimators. Econometrica. 1982; 82:1029–54.

Hoeffding W. A class of statistics with asymptotically normal distribution. Ann Math Statist. 1948; 19:293–325.

Huang C-Y, Qin J, Follmann DA. A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. Biometrika. 2012; 99:199–210. [PubMed: 23843659]

Kalbfleisch JD. Likelihood methods and nonparametric tests. J Am Statist Assoc. 1978; 73:167–70.

Kalbfleisch JD, Lawless JF. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. Statist Sinica. 1991; 1:19–32.

Liang KY, Qin J. Regression analysis under non-standard situations: A pairwise pseudolikelihood approach. J R Statist Soc B. 2000; 62:773–86.

Lin DY, Ying Z. Semiparametric analysis of the additive risk model. Biometrika. 1994; 81:61–71.

Lin DY, Ying Z. Semiparametric analysis of general additive-multiplicative hazard models for counting processes. Ann Statist. 1995; 23:1712–34.

Luo X, Tsai W-Y. A proportional likelihood ratio model. Biometrika. 2012; 99:211–22.

Martinussen T, Scheike TH. Efficient estimation in additive hazards regression with current status data. Biometrika. 2002a; 89:649–58.

Martinussen T, Scheike TH. A flexible additive multiplicative hazard model. Biometrika. 2002b; 89:283–98.

Qin J, Ning J, Liu H, Shen Y. Maximum likelihood estimations and EM algorithms with length-biased data. J Am Statist Assoc. 2011; 106:1434–49.

Sen PK. On some convergence properties of U-statistics. Calcutta Statist Assoc Bull. 1960; 10:1–18.

Thomas, DC. Use of auxiliary information in fitting nonproportional hazards models. In: Moolgavkar, SH.; Prentice, RL., editors. Modern Statistical Methods in Chronic Disease Epidemiology. New York: Wiley; 1986. p. 197-210.

Van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. New York: Springer; 1996.

Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. Biometrika. 1989; 76:751–61.

Wang M-C. Nonparametric estimation from cross-sectional survival data. J Am Statist Assoc. 1991; 86:130–43.

Wang MC, Brookmeyer R, Jewell N. Statistical models for prevalent cohort data. Biometrics. 1993; 49:1–11. [PubMed: 8513095]

Wolfson C, Wolfson DB, Asgharian M, M'Lan CE. A reevaluation of the duration of survival after the onset of dementia. New Engl J Med. 2001; 344:1111–6. [PubMed: 11297701]

Author Manuscript

Author Manuscript

Author Manuscript

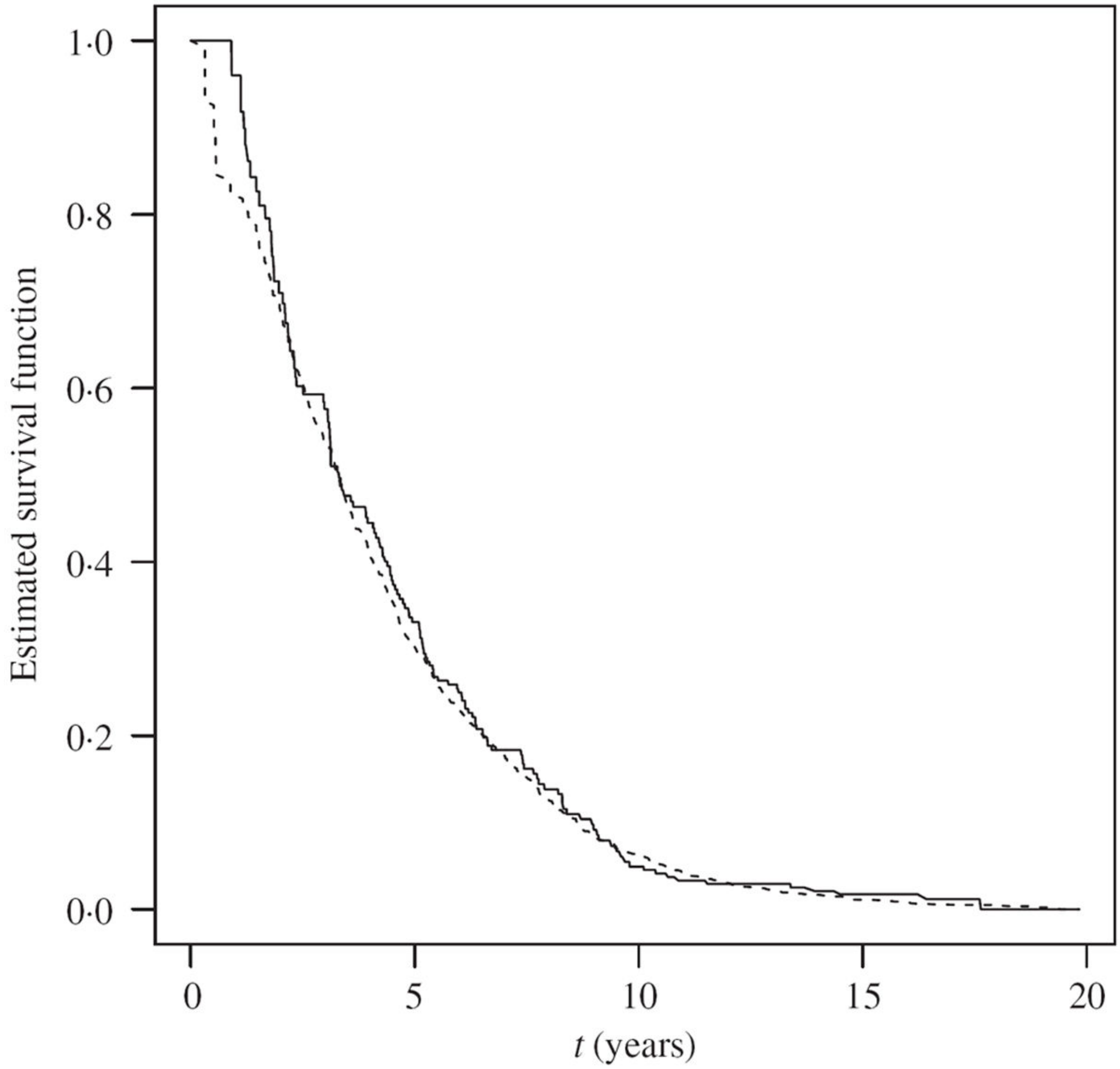

Fig. 1.

Estimated survival functions: the solid curve is $\hat{S}_1(t)$ and the dashed curve is $\hat{S}_0(t)\exp(-\hat{\beta_\varphi} t)$, where $\hat{S}_0(t)$ and $\hat{S}_1(t)$ are the truncation product-limit estimators for vascular dementia and probable/possibleAlzheimer's disease and $\hat{\beta_\varphi}$ is the conditional estimating equation estimator.

Table 1

Summary statistics for four estimators under different truncation time distributions and censoring proportions

	$\hat{\beta}_{\varphi}$		$\hat{\beta}_{\psi}$			$\hat{\beta}$			$\hat{\beta}_{W}$		
Cens	**Bias**	**SE**	**Bias**	**SE**	**RE**	**Bias**	**SE**	**RE**	**Bias**	**SE**	**RE**
Scenario I: $A^* \sim$ Un(0, 100), $\lambda(t \mid Z) = 1 + Z$											
0%	0	39	4	41	0.87	1	30	1.70	0	30	1.61
25%	2	46	4	40	1.31	2	31	2.17	0	31	2.11
50%	3	56	3	41	1.82	2	34	2.75	−2	35	2.60
Scenario II: $A^* \sim$ Un(0, 100), $\lambda(t \mid Z) = 0.5\, t^{1/2} + Z$											
0%	2	30	3	33	0.81	1	22	1.80	0	23	1.72
25%	2	32	6	34	0.87	3	23	1.87	2	24	1.76
50%	1	37	6	34	1.15	3	25	2.13	1	27	1.90
Scenario III: $A^* \sim$ exp(10), $\lambda(t \mid Z) = 1 + Z$											
0%	1	40	3	44	0.83	1	30	1.81	−1	31	1.70
25%	−1	44	5	42	1.13	0	30	2.11	−2	32	1.86
50%	0	54	7	42	1.65	2	33	2.63	−1	34	2.52
Scenario IV: $A^* \sim$ exp(10), $\lambda(t \mid Z) = 0.5\, t^{1/2} + Z$											
0%	2	29	4	36	0.63	1	23	1.61	1	23	1.56
25%	3	33	6	36	0.82	3	24	1.81	2	25	1.74
50%	3	37	6	36	1.05	3	26	2.03	2	27	1.86

Cens, censoring percentage; Bias and SE, empirical bias (×100) and empirical standard deviation (×100) of 1000 regression parameter estimates; RE, empirical variance of the competing method divided by that of $\hat{\beta}_{\varphi}$. The true value of the regression parameter is $\beta_0 = 1$.

Table 2

Summary statistics for three estimators with different truncation time distributions under the additive hazards model $\lambda(t \mid Z) = 1 + Z$. The censoring rate was set to *0*%

		$\hat{\beta}_{\varphi}$		$\hat{\beta}_{\psi}$			$\hat{\beta}$		
A*	V_{A*}	**Bias**	**SE**	**Bias**	**SE**	**RE**	**Bias**	**SE**	**RE**
2 × Be(1, 1)	0.33	2	41	3	55	0.54	2	32	1.57
2 × Be(4, 1)	0.11	1	40	0	65	0.39	1	34	1.37
2 × Be(1, 4)	0.11	0	38	5	106	0.13	0	36	1.13
DU{0, 1, 2}	1	0	38	5	51	0.55	1	31	1.52
Bi(2, 0.75)	0.38	3	39	1	37	1.11	2	28	1.96
Bi(2, 0.25)	0.38	1	40	3	76	0.27	1	35	1.30

V_{A*}, variance of *A**; Bias and SE, empirical bias (×100) and empirical standard deviation (×100) of 1000 regression parameter estimates; RE, empirical variance of the competing method divided by that of $\hat{\beta}_{\varphi}$; Be(*a*, *b*), beta distributionwith mean *a*/(*a* + *b*); DU{0, 1, 2}, discrete uniformdistribution taking values on {0, 1, 2}; Bi(2, *p*), binomial distribution with size 2 and success probability *p*.