Journal of Biometrics & Biostatistics

Hobbs, J Biom Biostat 2015, 6:2
DOI: 10.4172/2155-6180.1000232

Research Article Open Access

On Nonparametric Hazard Estimation

Hobbs BP*

Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

Abstract

The Nelson-Aalen estimator provides the basis for the ubiquitous Kaplan-Meier estimator, and therefore is an essential tool for nonparametric survival analysis. This article reviews martingale theory and its role in demonstrating that the Nelson-Aalen estimator is uniformly consistent for estimating the cumulative hazard function for right-censored continuous time-to-failure data.

Keywords: Asymptotic theory; Counting process; Hazard estimation; Martingale theory; Nelson-Aalen estimator; Time-to-failure analysis

Introduction

The Nelson-Aalen estimator Nelson [1,2] provides the foundation for the ubiquitous Kaplan-Meier survival estimator which consists of its product-integral. This article reviews martingale theory and its role in demonstrating that the Nelson-Aalen estimator is asymptotically uniformly consistent for the cumulative hazard function for right-censored continuous time-to-failure data and demonstrates its application using simulation.

Fundamental principles

Nonparametric statistical models for censored data were developed using counting processes which decompose into a martingales and integrated intensity processes. Martingale methods facilitate direct evaluation of small and large sample properties of hazard estimators for right censored failure time data. This section provides formal definition of a martingale process as well as reviews several intrinsic properties that are used to study small and large sample properties of hazard estimators for right censored failure time data.

Counting processes

A family of sub-σ-algebras $\{F_t: t \geq 0\}$ of a σ-algebra F is called increasing if $s \leq t$ implies $F_s \subset F_t$. An increasing family of sub-σ-algebras is called a filtration. When $\{F_t: t \geq 0\}$ is a filtration, the σ-algebra $\bigcap_{h>0} \mathcal{F}_{t+h}$ is usually denoted by F_{t+}. The corresponding limit from the left, F_{t-}, is the smallest σ-algebra containing all the sets in $\bigcup_{h>0} \mathcal{F}_{t-h}$ and is written $\sigma\{\bigcup_{h>0} \mathcal{F}_{t-h}\}$. A filtration $\{F_t: t \geq 0\}$ is right-continuous if, for any t, $F_{t+}=F_t$. A stochastic basis is a probability space (Ω, F, P) with a right-continuous filtration $\{F_t: t \geq 0\}$, and is denoted by $(\Omega, F, \{F_t: t \geq 0\}, P)$. A stochastic process $\{X(t): t \geq 0\}$ is adapted to a filtration if, for every $t \geq 0$, X (t) is F_t-measurable. A counting process is a stochastic process $\{N (t): t \geq 0\}$ adapted to a filtration $\{F_t: t \geq 0\}$ with N (0)=0 and N (t) < ∞ almost surely (a.s.), and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities with jumps of size +1. If N is a counting process, f is some function of time, and $0 \leq s < t \leq \infty$, then $\int_s^t f(u)dN(u)$, is the Stieltjes integral representation of the sum of the values of f at the jump times of N in the interval (s, t].

Martingales

Let $X=\{X(t): t \geq 0\}$ denote a right-continuous stochastic process with left-hand limits and $\{F_t: t \geq 0\}$ a filtration, defined on a common probability space. X is called a martingale with respect to $\{F_t: t \geq 0\}$ if, X is adapted to $\{F_t: t \geq 0\}$, $E|X (t)| < \infty$ for all $t < \infty$, and $E[X (t + s)|F_t]=X (t)$ a.s. for all $s \geq 0$, $t \geq 0$. Thus, a martingale is essentially a process that has no drift and whose increments are uncorrelated. If $E[X (t + s)|F_t] \geq X (t)$ a.s. X is a *submartingale.* Two fundamental properties of martingales are, for any h > 0,

$$E[X (t)|F_{t-h}]=X (t - h), \quad \text{and}$$

$$E[X (t) - X (t - h)|F_{t-h}]=X (t - h) - X (t - h)=0.$$

Predictable processes

The stochastic process X is said to be predictable with respect to filtration F_t if for each t, the value of X(t) is specified by F_{t-} and therefore is F_{t-}-measurable [4]. Theorem A.1 in Appendix A is the version of the Doob-Meyer Decomposition Theorem provided by Fleming and Harrington [3], which states that for any right-continuous nonnegative submartingale X there is a unique increasing right-continuous predictable process, A, such that A(0)=0 and X−A is a martingale. Also, there is a unique process A so that for any counting process, N, with finite expectation, N−A is a martingale. This is shown in the Corollary A.2 [3]. The process A in Corollary A.2 is referred to as the compensator for the submartingale X.

Square integrable martingales

A martingale, X(t) is called square integrable if $E[X^2(t)]=var[X(t)] < \infty$ for all $t \leq \tau$, or equivalently, if $E[X^2(\tau)] < \infty$ [4]. The variance of a square integrable martingale X(t) is estimated using the predictable variation process,

$$\langle X\rangle(t) = \int_0^t var[dX(u) \mid \mathcal{F}_{u^-}].$$

For martingales X_i, X_j,

$$\langle X_i, X_j\rangle(t) = \int_0^t cov[dM_i(u), dM_j(u) \mid \mathcal{F}_{u^-}]$$

is the predictable covariation process of X_i and X_j. If $(X_i, X_j)(t)=0$ for all t, then X_i and X_j are orthogonal martingales [3]. Suppose X_i, X_j are orthogonal martingales, for all $i \neq j$. Then $\langle \sum_{i=1}^m X_i\rangle(t) = \sum_{i=1}^m \langle X_i\rangle(t)$.

***Corresponding author:** Hobbs BP, Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA, Tel: 713-794-4633; E-mail: bphobbs@mdanderson.org

Received May 29, 2015; **Accepted** June 10, 2015; **Published** June 18, 2015

Citation: Hobbs BP (2015) On Nonparametric Hazard Estimation. J Biom Biostat 6: 232. doi:10.4172/2155-6180.1000232

Citation: Hobbs BP (2015) On Nonparametric Hazard Estimation. J Biom Biostat 6: 232. doi:10.4172/2155-6180.1000232

Localization

A nonnegative random variable τ is a stopping time with respect to filtration $\{F_t\}$ if $\{\tau \le t\} \in F_t$ for all $t \ge 0$. An increasing sequence of random times τ_m, m=1, 2, ..., is a localizing sequence with respect to a filtration if each τ_m is a stopping time and $\lim_{m\to\infty} \tau_m = \infty$ [3]. A stochastic process $X=\{X(t): t \ge 0\}$ is a local martingale (submartingale) with respect to a filtration $\{F_t: t \ge 0\}$ if there exists a localizing sequence $\{\tau_n\}$ such that, for each n, $X_n=\{X(t \wedge \tau_n): 0 \le t < \infty\}$ is an F_t-martingale (submartingale). If X_n is a martingale and a square integrable process, X_n is a square integrable martingale and X is called a local square integrable martingale. It can be shown that any martingale is a local martingale by simply taking $\tau_n=n$. An adapted process $X=\{X(t): t \ge 0\}$ is locally bounded if, for a suitable localizing sequence $\{\tau_n\}$, $X_n=\{(t \wedge \tau_n): t \ge 0\}$ is a bounded process for each n [3].

Martingale approach to censored failure time data

Suppose T and U are nonnegative, independent random variables, and assume that the distribution of T has a density. Define variable $X=(T \wedge U)$ to be the censored observation of the failure time variable T and $\delta=I(T \le U)$, the indicator for the event of an uncensored observation. The martingale approach to censored data uses the counting process $\{N(t): t \ge 0\}$ given at time t by $N(t)=I(X \le t, \delta=1)=\delta I(T \le t)$. We are interested in estimating the conditional rate at which N jumps in small intervals. Define the distribution and survival functions as $F(t)=P\{T \le t\}$ and $S(t)=1-F(t)$ and let $C(u)=P\{U > u\}$.

We can define the hazard function to be $\lambda(t)=\frac{-d[\log S(t)]}{dt}$ and cumulative hazard function $\Lambda(t)=\int_0^t \frac{dF(u)}{1-F(u^-)}$. Since we assumed that T and U are independent,

$$\lambda(t)=\lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\frac{P\{t \le T < t+\Delta t\}}{P\{T \ge t\}}=\lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}P\{t \le T < t+\Delta t \mid T \ge t, U \ge t\}. \quad (1)$$

Therefore, if we let $N(t^-)=\lim_{s\uparrow t} N(s)$ then

$\lambda(t)\Delta t \approx E[N((t+\Delta t)^-) - N(t^-) \mid T \ge t, U \ge t]$.

Thus, the hazard function $\lambda(t)$ represents the conditional average rate of change in N over $[t, t+\Delta t)$, given that both the censoring and failure time exceed or equal t. For each fixed t, $A(t)=\int_0^t I(X \ge u)\lambda(u)du$ is a random variable commonly referred to as the integrated intensity process which approximates the number of jumps by N over (0, t]. Let

$$M(t)=I(X \le t, \delta=1)-\int_0^t I(X \ge u)\lambda(u)du = N(t)-A(t).$$

Notice that $E[N(t)]=E[A(t)]$ (see (B.1) in Appendix B). Furthermore, suppose that $F_s=\sigma\{N(u), I(X \le u, \delta=0): 0 \le u \le s\}$ is the filtration for the process M (t) and F_{s-} is the information in N(u) and $I(X \le u, \delta=0)$ up to, but not including time s. Both $\{X < s\}$ and $\{X \ge s\}$ are F_{s-} -measurable since $\{X < s\}=\bigcup_{n=1}^{\infty}\{X \le s-\frac{1}{n}\}$. Furthermore, dN(s) is a Bernoulli random variable such that

$$E[dN(s) \mid F_{s-}]=I(X \ge s)\lambda(s)ds=dA(s). \quad (2)$$

When we have independent censoring it follows that $P\{s \le T<s+ds \mid T \ge s\}=P\{s \le T < s+ds \mid T \ge s, U \ge s\}$ and $E[dA(s) \mid F_{s-}]=dA(s)$ (see (B.2) in Appendix B). Therefore, the change in $M(t)=N(t)-A(t)$ over an infinitesimal interval (s−ds, s], $dM(s)=dN(s)-dA(s)$ has expectation 0 given F_{s-}. Therefore, M is a martingale with respect to $\{F_s\}$.

Nelson-Aalen estimator

Suppose time-to-failure data are collected over a finite interval [0, τ] on n subjects with independent failure times. The Nelson-Aalen estimator is a nonparametric estimator of their common cumulative hazard function, $\Lambda(t)$, in the presence of right censoring. Suppose T_i and U_i are the failure and censoring times and $N_i=\{N_i(t), t \ge 0\}$ the observed counting process for the ith subject. Let $\{Y_i(t), t \ge 0\}$ denote a process such that $Y_i(t)=1$ if and only if the ith subject is uncensored and has survived to time t^-. It is referred to the at risk process and assumed left-continuous. For each $t > 0$, let $F_{t-}=\sigma\{N_i(u),Y_i(u), i=1, ..., n; 0 \le u \le t\}$ denote the filtration up to, but not including t. Let $N.(t)=\sum_{i=1}^n N_i(t)$ and $Y.(t)=\sum_{i=1}^n Y_i(t)$ denote the aggregate processes that count the numbers of total failures and at risk in the interval (0, t], respectively. Furthermore, let $J(t)=I_{\{Y.(t)>0\}}$ indicate whether at least one subject is at risk at time t and suppose that $\frac{0}{0}=0$. The Nelson-Aalen estimator follows as

$$\hat{\Lambda}(t)=\int_0^t \frac{J(u)}{Y.(u)}dN.(u), \qquad 0 \le t \le \tau. \quad (3)$$

Asymptotic uniform consistency of the Nelson-Aalen estimator

In this section, we demonstrate how to prove that (3) is an uniformly consistent estimator of $\Lambda(t)=\int_0^t \frac{dF(u)}{1-F(u)}$, for continuous distribution function, $F(t)=P\{T \le t\}$. Note that by assuming continuous time we have assumed that no two of the counting processes $N_1(t)$, ..., $N_n(t)$ jump at the same moment. Theorem A.4, known as Lenglart's Inequality, and related Corollary A.5 are used to produce the result [3]. Two regularity conditions are necessary. First, $\inf_{t\in(0,\tau]} Y.(t) \to \infty$ in probability as $n \to \infty$. This implies that the number of subjects at risk at each time point becomes large for large n. Second, $\Lambda(\tau) < \infty$, thus, $F(t) < 1, \forall t \in [0, \tau]$.

First notice,

$$\left|\hat{\Lambda}(t)-\Lambda(t)\right|$$

$$\le \left|\int_0^s \frac{J(u)}{Y.(u)}dN.(u)-\int_0^s J(u)\frac{dF(u)}{1-F(u)}\right| + \left|\int_0^s [1-J(u)]\frac{dF(u)}{1-F(u)}\right|$$

$$= \left|\int_0^s \frac{J(u)}{Y.(u)}[dN.(u)-Y.(u)d\Lambda(u)]\right| + \left|\int_0^s [1-J(u)]\frac{dF(u)}{1-F(u)}\right|$$

(see e.g. (B.3) in Appendix B). We demonstrated in Section 2 that $N_i(t)$ has compensator $A_i(t)=\int_0^t Y_i(u)d\Lambda(u)$. A simple alteration of (B.1) in Appendix B can be used to show that for $A.(t)=\sum_{i=1}^n \int_0^t Y_i(u)d\Lambda(u)$, $E[N.(t) \mid \mathcal{F}_{t}]=nE[N_i(t) \mid \mathcal{F}_t]=nE[A_i(t) \mid \mathcal{F}_t]=E[A.(t) \mid \mathcal{F}_t]$.

Also, since $A_i(0)=\Lambda(0)-\Lambda(0)=0$, for all i=1, ..., n then $A.(0)=0$. Now we must show that $A.(t) < \infty$ and $A.(t)$ is locally bounded $\forall t \in [0, \tau]$. For all fixed n=1, 2... and all $t \in [0, \tau]$, it follows that

$$A_i(t) \le \int_0^t d\Lambda(u) = -\left(\log(S(t))-\log(S(0))\right)$$
$$=\log(1-F(0))-\log(1-F(t))=-\log(1-F(t)).$$

The second regularity condition imposes that $F(t) < 1, \forall t \in [0, \tau]$. Thus, we can conclude that $A.(t) < \infty, \forall t \in [0, \tau]$. If we consider the localizing sequence $\tau_m=\sup\{t: \sup_{0\le s\le t}|A.(s)|< m\} \wedge m$, for m=1, 2, ..., and stopping process $Q_m=A.(t \wedge \tau_m)$, then A. satisfies the conditions for locally bounded. Thus, Theorem A.3 in Appendix A, which is given by [3], implies that $M.(t)=N.(t)-\sum_{i=1}^n A_i(t)$ is a local square integrable martingale. In inequality (B.3) in Appendix B we noted that $\int_0^t \frac{J(u)}{Y.(u)}[dN.(u) - Y.(u)d\Lambda(u)] = \int_0^t \frac{J(u)}{Y.(u)} dM.(u)$. Equation (2) as well as Corollary A.2 in Appendix A suggest that A.(t) is a predictable process. Therefore, so is $\frac{J(t)}{Y.(t)}$. In addition, since $0 \le \frac{J(t)}{Y.(t)} \le 1$ for all t ≥ 0 and n>0, it is clearly locally bounded. Corollary A.5 in Appendix A can be used to show that $\sup_{0 \le s \le t}\{\int_0^t \frac{J(u)}{Y.(u)} dM.(u)\}^2 \to 0$, in probability where $H(s) = \frac{J(s)}{Y.(s)}$. In particular,

$$P\left\{\sup_{0 \le s \le t}\left\{\int_0^t \frac{J(u)}{Y.(u)} dM.(u)\right\}^2 \ge \epsilon\right\} \le \frac{\eta}{\epsilon} + P\left\{\int_0^t \frac{J(u)}{Y.^2(u)} d\langle M.\rangle(u) \ge \eta\right\}. \quad (4)$$

Using the variances of increments of M.(t), one can show that, $\langle M.\rangle(t) = \sum_{i=1}^{n} \langle M_i\rangle(t) + 2\sum_{i<j}\langle M_i, M_j\rangle(t)$ (see e.g. (B.4) in Appendix B). The assumption of continuous time implies that for all $i \ne j$, M_i, and M_j are orthogonal martingales. In other words, $N_i(t)$ and $N_j(t)$ cannot jump at the exact same moment in time, $\forall i \ne j$ and ∀t ≥ 0. Therefore, we have $\langle M.\rangle(t) = \sum_{i=1}^{n} \langle M_i\rangle(t)$. Now for any i=1, ..., n and any t>0, $\langle M_i\rangle(t) = E[M_i^2(t) \mid \mathcal{F}_{t^-}]$, because $E[M_i \mid \mathcal{F}_{t^-}] = 0$. Since $dN_i(t)$ is a Bernoulli random variable ∀t>0 and ∀i, and $dA_i(t)$ is approximately 0 over an infinitesimally small interval,

$$E[dM_i^2(t) \mid \mathcal{F}_{t^-}] = E[(dN_i(t) - dA_i(t))^2 \mid \mathcal{F}_{t^-}] = E[dN_i^2(t) \mid \mathcal{F}_{t^-}] - dA_i^2(t)$$

$$= dA_i(t)\left(1 - dA_i(t)\right) \approx dA_i(t).$$

Thus, taking the predictable variation process equal to its compensator: $\langle M_i\rangle(t) = A_i(t) = \int_0^t Y_i(u) d\Lambda(u)$, it follows from (4) that

$$P\left\{\int_0^t \frac{J(u)}{Y.^2(u)} d\langle M.\rangle(u) \ge \eta\right\} = P\left\{\int_0^t \frac{J(u)}{Y.(u)} d\Lambda(u) \ge \eta\right\} \le P\{\frac{\Lambda(t)}{Y.(t)} > \eta\}$$

From our first regularity condition we have Y.(t) → ∞ in probability, as n → ∞. Therefore, $P\{\frac{\Lambda(t)}{Y.(t)} > \eta\} \to 0$ in probability for any η>0. Thus, all that is left to show is that $\sup_{s \in [0,\tau]} \left| \int_0^s [1\text{-J(u)}] \frac{dF(u)}{1\text{-F(u)}} \right| \to 0$ probability as n → ∞. Noting that $\left| \int_0^s [1\text{-J(u)}] \frac{dF(u)}{1\text{-F(u)}} \right|$ is bounded by |Λ(s) – I (Y.(s) > 0)Λ(s)| e. g.

$$\left| \int_0^s [1\text{-J(u)}] \frac{dF(u)}{1\text{-F(u)}} \right| = \left| \Lambda(s) - \int_0^s I(Y.(u) > 0) d\Lambda(u) \right| \le \left| \Lambda(s) - I(Y.(s) > 0)\Lambda(s) \right|,$$

from our first regularity condition we can conclude that

$$\lim_{n \to \infty} P\{\left| I(\inf_{s \in [0,\tau]} Y.(s) > 0) - 1 \right| \le \nu\} = 1, \quad \forall \nu > 0.$$

Therefore we have,

$$\sup_{0 \le s \le \tau} \left| \Lambda(s) - I(Y.(s) > 0)\Lambda(s) \right| \le \Lambda(\tau) \left| 1 - I(\inf_{0 \le s \le \tau} Y.(s) > 0) \right|.$$

Thus, we can use Slutsky's Theorem to conclude that $\sup_{s \in [0,\tau]} \left| \int_0^s [1\text{-J(u)}] \frac{dF(u)}{1\text{-F(u)}} \right| \to 0$ in probability as n → ∞. Therefore, the Nelson-Aalen estimator is asymptotically uniformly consistent for the cumulative hazard function under regularity conditions: $inf_{t \in (0,\tau]} Y.(t) \to \infty$ in probability as n → ∞ and Λ(τ) < ∞.

Simulation Study

This section demonstrates the martingale approach for analysis of right-censored failure time data using simulation. Let A_1, A_2, ... denote independent uniform (0, τ) random variables where τ is a known constant, and let $x_i(t)$=I (A_i<t), i=1, ..., n. $N_i(t)$ counts observed failures for the ith subject observed over the interval (0, τ_i], where τ_i is a constant $0 < \tau_i \le \tau$, i=1, ..., n. We assume that the intensity function of $N_i(t)$ with respect to the filtration $F_t = \sigma\{N_i(u), Y_i(u^+), X_i(u^+), i=1, ..., n; 0 \le u \le t\}$ is

$$\lambda_i(t) = Y_i(t) \exp[x_i(t)\beta]\alpha, \quad (5)$$

for 0 < t < τ, i=1, ..., n, and α > 0. We simulated 30 realizations of the process [N.(t), 0 < t < 10] and the corresponding martingale when τ=10, n=10, τ_i=i, for i=1, ..., 10 and fixed "baseline hazard" α=1. For convenience, we fixed β=1. Let z_i=exp(1), if x_i=1 and=1 otherwise. We approximated continuous time by partitioning [0, 10) into disjoint intervals, t_j of length d_t=0.1. Now it follows that $E[dN_i(t_j)] = z_i \alpha dt$ at each $t_j \in [0, \tau)$. For each subject, the process was simulated by iterating through each time interval t_j within each subject. At each t_j we draw a single sample, $dN_i(t_j)$, from the Poisson density with rate parameter $E[dN_i(t_j)]$. If $dN_i(t_j)$>0, then we have observed the failure time for the ith subject to be t_j. Therefore, we set counting processes $N_i(k)$=1 and $Y_i(k)$=0, $\forall t_j \le k \le \tau$. Furthermore, the aggregated counting process N.(k)=N.(k) + 1, $\forall t_j \le k \le \tau$.

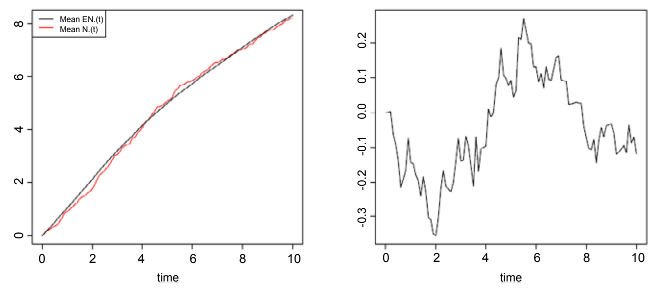

Figure 1: Results for the simulated mean aggregated counting process N.(t) and expectation (right) and the corresponding simulated mean martingale (left).

Figure 1 provides the resulting expected and simulated mean aggregated counting process N.(t) (right panel) as well as the simulated mean martingale (left panel). As expected, the simulated mean aggregated counting process N.(t) mirrors closely its expectation. Furthermore, the corresponding simulated mean aggregated martingale illustrates a process that has no random drift.

Discussion

Martingale characterizations play a critical role in the study of large sample properties of the Nelson-Aalen and thereby Kaplan-Meier estimators. This article reviewed martingale theory and its role in demonstrating that the Nelson-Aalen estimator is asymptotically uniformly consistent for the cumulative hazard function for right-censored continuous time-to-failure data and demonstrates its application using simulation.

Acknowledgement

The author was partially supported by NIH/NCI grants R01CA157458 and P30CA016672.

References

1. Nelson W (1969) Hazard plotting for incomplete failure data. Journal of Quality Technology 1: 27-52.
2. Nelson W (1972) Theory and applications of hazard plotting for censored failure data. Technometrics 14: 945-965.

3. Fleming TR, Harrington DP (1991) Counting Processes and Survival Analysis. John Wiley and Sons.
4. Kalbfleisch JD, Prentice RL (2002) The Statistical Analysis of Failure Time Data. (2nd edition), John Wiley and Sons.