Topics in Survival Analysis

Notes on large sample theory for counting processes

Lecture: Professor Mai Zhou
Notes (Spring 2009)
Updated (Fall 2011)

Department of Statistics

The University of Kentucky

1 Introduction

Suppose $N_i(t), i = 1, 2, \dots, n$, are independent Poisson processes with parameter $\lambda > 0$. The kind of limit we are interested can be seen in this simple example: as $n \to \infty$

$$U_n(t) = \frac{\sum_{i=1}^n [N_i(t) - \lambda t]}{\sqrt{n\lambda}} \quad \text{converge to ??}$$

For any fixed t, the above converge to a normal distribution by ordinary CLT. As $n \to \infty$ we should get a "normal process" in the limit, it still has independent increments (because for every n $U_n(t)$ is, please verify), with continuous sample path (because $U_n(t)$ has jumps of size $1/\sqrt{n}$ order), – that kind of normal process is called a Brownian Motion.

(remark: if we divide by $\sqrt{n\lambda t}$ in the above then we will lose independent increments property).

Definition A Brownian Motion $B(t)$ is a stochastic process with $B(0) \equiv 0$, has independent increments, with continuous sample path (i.e. $B(t, \omega)$ is continuous in t) and the distribution of $B(t)$ is Normal $N(0, Var = t)$.

A major result we will be discussing in this note is a Central Limit Theorem for the processes. The above is just a special case: as $n \to \infty$

$$U_n(t) \xrightarrow{\mathcal{D}} B(t) \ .$$

We shall cover more general type of processes then independent increments. We shall cover martingale processes.

[Excersice: verify that the processes $U_n(t)$ and $B(t)$ are martinagles in t with self-exciting filtration.]

But first, what do we mean by converge in distribution of stochastic processes?

2 Space C[0, 1] and D[0, 1]; Convergence in distribution

The space of all continuous function on $[0, 1]$ is denoted by $C[0, 1]$. We notice that the Brownian motions live on this space.

The space of all Cadleg functions on $[0, 1]$ is denoted by $D[0, 1]$. We notice that the Poisson processes U_n above lives on this space.

We define the convergence in distribution as follows.

Definition: If for every bounded continuous function $f(\cdot)$, we have $Ef(X_n) \to Ef(X)$ then we say the sequence of random elements X_n converge to X in distribution as $n \to \infty$.

Remark: this definition works for random variables, k dimensional random vectors, as well as random processes. In those case the function f will be a real function, an R^k to R^1 function or a $C[0,1]$ ($D[0,1]$) to R^1 map.

In order to define continuous functions, we need to define a metric. The distance on R^1 or R^k will always be the Euclidian distance for us. The distance on $C[0,1]$ and $D[0,1]$ can either be the uniform distance:

$$d(x,y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$$

or a so called Skorohod distance.

In this note, we shall always use the uniform distance and avoid the Skorohod distance.

Example: The Kaplan-Meier estimator is a random element lives on $D[0,1]$ space.

Therefore, a sequence of stochastic processes, $x_n(t)$, live on $D[0,1]$ is said to converge in distribution to $x(t)$, if $Ef(x_n(t)) \to Ef(x(t))$ for every bounded, continuous function f.

We assume the limiting process $x(t)$ always lives on a separable subspace of $D[0,1]$, like $C[0,1]$.

3 Some Properties of Brownian Motion

Some calculation formula involve Brownian motion. We know from the normal distribution definition of BM that for every t, $EB(t) = 0$ and $E[B(t)^2] = t$.

Lemma 1 The covariance computation. $E(B(t_1)B(t_2)) = E(B(t_1)B(t_1)) = t_1$, for $0 < t_1 < t_2$.

Proof:

$$\begin{aligned}
E(B(t_1)B(t_2)) &= E(B(t_1)[B(t_2) - B(t_1) + B(t_1)]) & (1)\\
&= E(B(t_1)[B(t_2) - B(t_1)]) + E(B(t_1)B(t_1)) & (2)\\
&= E(B(t_1))E(B(t_2) - B(t1)) + t_1 \quad \text{(independent increments)} & (3)\\
&= t_1 & (4)
\end{aligned}$$

Lemma 2 $V(B(t_2) - B(t_1)) = t_2 - t_1$, for $t_1 < t_2$.

Proof:

$$\begin{aligned}
V(B(t_2) - B(t_1)) &= V(B(t_2)) + V(B(t_1)) - 2Cov(B(t_1), B(t_2)) & (5)\\
&= t_2 + t_1 - 2(E(B(t_1)B(t_2)) & (6)\\
&= t_2 + t_1 - 2t_1 \qquad \text{(By Lemma 1)} & (7)\\
&= t_2 - t_1 & (8)
\end{aligned}$$

We verified in the previous section as exercise that $B(t)$ is a martingale.

Lemma 3 The process $B^2 - t$ is also a martingale.

What about the process $a(t)B(t)$ where $a(t)$ is a non-random function? This process is no longer a BM but we can easily see that it has the following properties.

a. $a(0)B(0) = 0$

b. $a(t)B(t) \sim N(0, a^2(t)t)$, and similarly for the joint distributions $(a(t_1)B(t_1), a(t_2)B(t_2))$

c. $a(t)B(t)$ does not have independent incrementS unless $a(t)$ is constant

d. $a(t)B(t)$ has continuous sample path, if $a(t)$ is a continuous function.

Is it a martingale? [exercise]

Gaussian process is just a name we gave to those process similar to the Brownian motion except do not have independent increment, stationary property.

4 Transformations/Integrations of processes

4.1 Two Generalizations of Poisson Processes

We consider two types of generalizations/transformations for Poisson processes $N(t)$: time accelaration/decelaration and jump size change.

1. time change: $N(t)$ becomes $N(g(t))$ where $g(t)$ is an increasing function (clock $g(t)$), with $g(0) = 0$. After the time change, the size of jumps are still $+1$. But the waiting time between jumps are no longer iid exponential. We shall use the term: the crazy clock function $g(t)$.

2. jump size change: $P(t) = \int_0^t f(s)dN(s)$. This process $P(t)$ has jump size $f(T_1), f(T_2), \ldots$ where T_i are the location/time of the jumps of $N(t)$. The time of the jumps are not changed.

If we apply both changes simultaneously, then we have: $P(t) = \int_0^t f(s)dN(g(s))$. The resulting process $P(t)$ is a pure jump process. (jump size not always one, and waiting time between jumps are not always exponential).

Web pages with applet that you can interactively do time change and integration (jump size change) of a Poisson process is available at:

http://www.ms.uky.edu/∼mai/java/stat/countpro.html

http://www.ms.uky.edu/∼mai/java/stat/mart.html

Question: what is the mean of the pure jump process $P(t)$?, What is the conditional mean of the increment of $P(t)$? Does $P(t)$ have independent increments?

Remark: The change $a(t)N(t)$ not only change the jump sizes, but also change the Poisson process between jumps.

Notice: the jump size function (and clock g) can depend on 'history' of the process, and still make the game 'fair'. In the applets, the red line minus the black have mean zero.

Example: Write down a process similar to the Poisson process $N(t)$, except the jump sizes are $1, 1/2, 1/3, 1/4, \cdots$,

$$P(t) = \int_0^t \frac{1}{1 + N(s-)} dN(s)$$

here the jump size function f is 'predictable'.

5 Martingales Related to $N(t)$

Definition A stochastic process $M(t)$, $t \in [0, T)$ and an increasing σ fields $\mathcal{F}_t$ are called a martingale, if (1) for every t $M(t)$ is $\mathcal{F}_t$ measurable; (2) $E(M(t)|\mathcal{F}_s) = M(s)$ for any $0 \leq s < t < T$.

The increasing *sigma* fields will be called a filtration: $\mathcal{F}_t$

The following facts are not hard to verify. If $N(t)$ is a Poisson process with intensity λ, then

$$M(t) = N(t) - \lambda t = N(t) - \int_0^t \lambda dt$$

is a martingale, with respect to the filtration $\mathcal{F}_t = \sigma\{\}$.

Recall that Poisson random variable have mean λ *and* variance also λ. We have, that the following is also a martingale:

$$[M(t)]^2 - \lambda t$$

Now assume $N(t)$ is a standard Poisson process (i.e. with $\lambda = 1$). It should be obvious that for a (non-random) crazy clock function $g(t)$, the time changed poisson process,

$$N(g(t)) - g(t)$$

is still a martingale (with respect to $\mathcal{F}_{g(t)}$).

Next, we work on a particular choice of $g(t)$. For any given positive random variable T with distribution $F_T(t)$, define $g(t) = \Lambda(t) = H(t) =$ cumulative hazard function of T. Thus $g'(t) = \lambda(t) = h(t) =$ hazard function of T.

Theorem Suppose $N(t)$ is a standard Poisson process (with $\lambda = 1$). Then the waiting time for the first jump of the new process $N(g(t))$ is equal to T in distribution, if we take $g(t) = H_T(t)$.

Proof: Denote the first jump time of $N(g(t))$ as X_1. We compute

$$P(X_1 > t) = P(N(g(t)) = 0) = \exp(-g(t))$$

by the Poisson distribution of $N(g(t))$. Since we took $g(t) = H_T(t) = -\log[1 - F_T(t)]$ (assume continuous $F_T(t)$), we may conclude

$$P(X_1 > t) = 1 - F_T(t) \ .$$

Or, $P(X_1 \le t) = F_T(t)$. i.e. $X_1 = T$ in distribution. QED

This tells us what $g(t)$ to use if we want the waiting time of first jump to have certain distribution $F_T(\cdot)$.

Next we make the subsequent jumps all of size zero. This way there will be one and only one jump. Leave the clock as $g(t) = H(t)$, we shall make all the subsequent jumps size $= 0$ after the first jump.

Using the jump size zero argument, when $g(t) =$ cumulative hazard

$$\int_0^t I[s \le T] dN(g(s))$$

is a one jump process and is identical to $I[t \ge T]$.

And we have

$$\int_0^t I[s \le T] dN(g(s)) - \int_0^t I[s \le T] dg(s)$$

is a martingale.

5.1 Censoring – as Thinning/Spliting of a Counting Process

Recall, we may thinning a Poisson process by further classify the type of its jumps. Similar thing also works here for the jump process (?). The type of jump we use here will be censoring.

Let $X > 0$ be the failure time and $C > 0$ be the censoring time. Assume they are independent. Let $Z = \min(X, C)$.

The counting process based on the Z is $N(t) = I[t \ge Z]$. The intensity for this (one jump) counting process is $h_z(t)I[Z \ge t]$. (i.e. $I[t \ge Z] - \int_0^t h_z(s)I[Z \ge s]ds$ is a martingale).

Censoring is to split the event of jump into two types: either a death type or a censoring type, indicated by the value $\delta = I[X \le C] = I[Z = X]$.

After the thinning, we have two counting processes. $N^x(t) = I[Z \le t, \delta = 1]$ and $N^c(t) = I[Z \le t, \delta = 0]$. They have intensity $h_x(t)I[Z \ge t]$ and $h_c(t)I[Z \ge t]$ respectively.

Notice we have $N^x(t) + N^c(t) = N(t)$, and $h_x(t)I[Z \ge t] + h_c(t)I[Z \ge t] = h_z(t)I[Z \ge t]$.

We shall continue the discussion of those one-jump counting process and martingales in section ?. But here we digress to discuss the Brownian motion, which is the limit process of those martingales.

6 Transformations/Integrations of Brownian Motion

Now let us look at the same two transformations for the Brownian motion.

1. Time change: $B(t)$ changes to $B(g(t))$. A Brownian motion with clock $g(t)$, not a standard Brownian motion. Here $g(t)$ is also the variance function.

2. There is no jump for $B(t)$, so that the 'jump size change' name does not make sense, but the integration can still be defined. (as stochastic integrals).

The integration is actually equivalent (in distribution) to a (variance change) time change:

$$\int_0^t f(s)dB(s) \stackrel{\mathcal{D}}{=} B^*(g(t))$$

where B^* is a Brownian motion; the clock $g(t) = \int_0^t f^2(s)ds$.

Notice the analogy: sum of independent, zero mean normal r.v. is equivalent in distribution to a constant multiple of a single normal r.v. and the constant multiple of a zero mean normal random variable is again a zero mean normal random variable with variance multiplied by c^2.

If $|f(s)| \equiv 1$, then $\int_0^t f(s)dB(s) = B^*(t)$, another Browning motion.

Heuristically, $\int_0^t f(s)dB(s) \simeq \sum_i f(t_i)[B(t_{i+1}) - B(t_i)]$. Using the independent increment property, the right hand side is normally distributed with mean 0 and $V = \sum_i f^2(t_i)V(B(t_{i+1}) - B(t_i)) = \sum_i f^2(t_i)(t_{i+1} - t_i) = \int_0^t f^2(s)ds$.

A Word of Caution: since $B(t, \omega)$ is not differentiable, and not Bounded Variation, the integration cannot be defined as path-wise Stieljes integral.

Some more types of change/transformations involving $B(t)$:

3. $\int_0^T B(s)dg(s)$ assume $g(s)$ is bounded variation. So this is the Steljes integral. Obviously it is $\sim N(0, V)$. Think about how to calculate V.

This is useful for calculating confidence interval for mean survival time estimator. Assume we know the distribution of Kaplan Meier, then $\hat{E}(X) = \int_0^\infty [1 - \hat{F}(X)]dx$, what is $V(\hat{E}(X))$?

For the Brownian motion. Same kind of transformations.

7 Variance of L^2 martingales

For a process with independent increments, the variance is easier to compute. (just add the variance of each increments up to time t).

If a process do not have independent increments, but is only a martingale (with finite variance) then the following is valid: Assume $M(t)$ has mean zero

$$Var\{M(t)\} = E\left\{\sum_i E[(M(t_{i+1}) - M(t_i))^2|\mathcal{F}_{t_i}]\right\}$$

where $0 = t_0 < t_1 < \cdots < t_n = t$.

Notice this is similar to 'variance of a sum is sum of the variances'. The cross product term becomes zero after conditional expectation. When the interval t_i, t_{i+1} becomes smaller the above sum (those inside the $\{\}$ on right hand side) becomes

$$V(t) \triangleq \langle M(t) \rangle$$

where

$$dV(s) = V(s+ds) - V(s) = E[(M(s+ds) - M(s))^2|\mathcal{F}_s] \ .$$

This integral $V(t)$ is called "predictable variation process" of the original martingale $M(t)$. Notice $V(t)$ is predictable and $Var(M(t)) = EV(t)$. It is also true that $M^2(t) - V(t)$ is again a martingale.

For Poisson martingale, this $V(t)$ is particularly easy. By the independent increment and Poisson distribution property we easily see that $dV(s) = \lambda ds$ for the Poisson martingale.

For the Brownian motion $BM(t)$ we have $\langle BM(t) \rangle = t$.

For an integration of a martingale, $\int_0^t f(s)dM(s)$ if $f(t)$ is predictable then it is easy to see that the predictable variation process of this integral is $\int_0^t f^2(s)dV(s)$.

8 Central Limit Theorem for Counting Process Martingales

8.1 More on counting process martingales

Consider n independent pairs of positive random variables X_i, C_i, $i = 1, 2, \cdots, n$; where X_i and C_i are independent. Define $Z_i = \min(X_i, C_i)$ and $\delta_i = I[Z_i = X_i]$. Denote the hazard function of X_i by $h_i^x(s)$, the hazard function of C_i by $h_i^c(s)$. Notice $h_i^z(s) = h_i^x(s) + h_i^c(s)$.

Theorem (3 basic martingales) Using the notation above, we have

$$M_i(t) = I[Z_i \le t] - \int_0^t h_i^z(s) I[Z_i \ge s]ds = I[Z_i \le t] - H_i(t \wedge Z_i)$$

(where $a \wedge b = \min(a, b)$) is a martingale. Similarly,

$$M_{1i}(t) = I[Z_i \le t, \delta_i = 1] - \int_0^t h_i^x(s) I[Z_i \ge s]ds$$

and

$$M_{2i}(t) = I[Z_i \le t, \delta_i = 0] - \int_0^t h_i^c(s) I[Z_i \ge s] ds$$

are also martingales. The filtration $\mathcal{F}_t$ needed to define the 3 martingales can taken to be the 'history' based on the Z_i, δ_i: $\mathcal{F}_t = \sigma\{Z_i I[Z_i \le t], \delta_i I[Z_i \le t] \quad i = 1, \cdots, n;\}$

This fundamental fact is proved the reference books.

Here we can just view $I[Z_i \le t]$ as the result of transformations from a Poisson process (as discussed in section 5).

And the other two processes as the 'thinning' or splitting of this Poisson process. The basic fact is that a (time changed) Poisson process minus the clock is a martingale.

Notice $M_i(t) = M_{1i}(t) + M_{2i}(t)$. We mainly use $M_{1i}(t)$ in survival analysis.

8.2 The CLT

Consider $\int_0^t g_{ni}(s) dM_i(s) = \delta_i I[Z_i \le t] g_{ni}(Z_i) - \int_0^t g_{ni}(s) h_i^x(s) I[Z_i \ge s] ds$, and the summation/average. If the function g_{ni} are 'predictable', then this integration is also a martingale.

Let

$$U_n(t) = \sum_{i=1}^n \int_0^t g_{ni}(s) dM_i(s) \ . \tag{9}$$

If for all i, $g_{ni}(t)$ are integrable functions that can only depend on the past wrt $\mathcal{F}_t$ (at any time t), then $U_n(t)$ is also a martingale (by above theorem) and $U_n(t) \xrightarrow{\mathcal{D}} BM(V(t))$ with $V(t)$ being increasing and non-random **if** two conditions below are satisfied. ($BM(\cdot)$ is a Brownian motion).

The first condition is ("convergence of (conditional) variance")

$$\sum_{i=1}^n \int_0^t g_{ni}^2(s) h_i^x(s) I[Z_i \ge s] ds \xrightarrow{\mathcal{P}} V(t) \tag{10}$$

where $V(t)$ must be a non-random function.

The second condition is a Lindeberg condition:

$$\forall \epsilon > 0, \quad \sum_{i=1}^n \int_0^t g_{ni}^2(s) I[|g_{ni}(s)| > \epsilon] h_i^x(s) I[Z_i \ge s] ds \xrightarrow{\mathcal{P}} 0 \ . \tag{11}$$

Remark: the Lindeberg condition controls the jump size of the $U_n(t)$. It says the variance of the jump part of U_n (with jump size $> \epsilon$) is going to zero in probability.

8.3 Use the CLT in Survival Analysis

For a concrete example we now suppose that X_i are identically distributed (in addition of independent), so the hazard function of X_i are the same: $h_i^x(s) = h^x(s)$. Then,

$$M_{1i}(t) = I[Z_i \le t] - \int_0^t I[Z_i \ge s] h^x(s) ds$$

If we define

$$f_{ni}(s) = \frac{\sqrt{n}}{R(s)}$$

where $R(s) = \sum_i I[Z_i \ge s]$ (please verify it is predictable) and

$$U_n(t) = \sum_{i=1}^n \frac{\sqrt{n} I[Z_i \le t]\delta_i}{R(Z_i)} - \int_0^t \sqrt{n} h^x(s) ds = \sqrt{n}\left[\sum_{i=1}^n \frac{\delta_i I[Z_i \le t]}{R(Z_i)} - H_X(t)\right] \xrightarrow{\mathcal{D}} B(V(t))$$

We can easily check the two conditions and find the expression of $V(t)$. (left as exercise).

This last case is actually the (process) CLT for the Nelson-Aalen estimator, in the sense that, $\sqrt{n}(\hat{\Lambda}(t) - \Lambda(t)) = U_n(t)$.

There are parallel results concern the martingales M_{2i} and M_i.

A special case, which we will use in the empirical likelihood.

Theorem 3 Given a non-random function $f(t)$. Suppose

$$\sigma^2 = \int_0^\infty \frac{f^2(t)}{P(Z \ge t)} d\Lambda_x(t) < \infty$$

then we have

$$\sqrt{n}\left(\int_0^\infty f(t) d\hat{\Lambda}(t) - \int_0^\infty f(t) d\Lambda_x(t)\right) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

where $\sigma^2 =$ above integral, and $\hat{\Lambda}(t)$ is the Nelson-Aalen estimator.

For the central limit theorem about the Kaplan-Meier estimate, there are two approaches.

(1) for a given integral with respect to the Kaplan-Meier

$$\sqrt{n}\int_0^\infty g(t) d[\hat{F}(t) - F(t)]$$

we try to find a difference integral with respect to the Nelson-Aalen

$$\sqrt{n}\int_0^\infty g^*(t) d[\hat{\Lambda}(t) - \Lambda(t)]$$

such that the two integrals are very close to each other. Indeed, Akritas (2000) explicitly found the $g^*(t)$ and show the difference is going to zero in probability. See Akritas (2000) for this approach. In general, this is called i.i.d. representation of the Kaplan-Meier estimator.

(2) Use the (Duhamel's equation) identity

$$\frac{\hat{F}(t) - F(t)}{1 - F(t)} = \int_{-\infty}^{t} \frac{1 - \hat{F}(s-)}{1 - F(s)} d[\hat{\Lambda}(s) - \Lambda(s)]$$

and the already known results about the Nelson-Aalen estimator. We can show the left hand side is a martingale in t.

References:

Akritas, M. (2000) The central limit theorem under censoring. Bernoulli 6(6), 1109-1120.

Homework:

For a continuous, positive random variable Z, define

$$M(t) = I[Z \le t] - \int_0^t I[Z \ge s] h^z(s) ds$$

where $h^z(\cdot)$ is the hazard function of Z. Formally define the 'self-exciting' history filtration

$$\mathcal{F}_t = \sigma\{M(s), 0 \le s \le t\} \ .$$

(1). Verify that $M(t)$, is an $\mathcal{F}_t$ martingale in t. (for $0 < t$)

(2). Verify that

$$M^2(t) - \int_0^t I[Z \ge s] h^z(s) ds$$

is also an $\mathcal{F}_t$ martingale in t.

(3). Verify that, for predictable function $f(t)$, wrt $\mathcal{F}_t$,

$$\int_0^t f(s) dM(s)$$

is also a $\mathcal{F}_t$ martingale.

(4). Verify that,

$$\left(\int_0^t f(s) dM(s) \right)^2 - \int_0^t f^2(s) I[Z \ge s] h^z(s) ds$$

is also a $\mathcal{F}_t$ martingale.

Finally, we have parallel result for two more related martingales. We just write one of them down here. let $T = \min(Z, C)$, also let $\delta = I[Z \le C]$, where C is a positive random variable independent of Z. Define

$$M(t) = I[T \le t, \delta = 1] - \int_0^t I[T \ge s] h^z(s) ds$$

then $M(t)$ is a martingale.

Similarly $[M(t)]^2 - \int_0^t I[T \ge s] h^z(s) ds$ is also a martingale.

For predictable function $f(t)$, the following two are martingale

$$\int_0^t f(s) dM(s)$$

$$\left(\int_0^t f(s)dM(s)\right)^2 - \int_0^t f^2(s)I[T \geq s]h^z(s)ds$$

May be we can call one type the mean martingale, the other the variance martingale for counting process.