Empirical Processes & Survival Analysis

Lecture 2
The Empirical Processes

Lu Mao lmao@biostat.wisc.edu

Objectives

By the end of this lecture, you will

1. understand the theoretical characterizations of weak convergence in functional spaces such as $l^\infty(\mathcal{F})$, and how they specialize in the case of EP;
2. learn Glivenko-Cantelli and Donsker Theorems for classes of functions by bounding their bracketing numbers;
3. see how these techniques are applied to elementary survival analysis problems.

Contents

Weak Convergence in Metric Spaces

For empirical processes, we are dealing with $l^\infty(\mathcal{F})$ equipped with the uniform norm. Since there is lack of an obvious counterpart of "distribution functions" on $l^\infty(\mathcal{F})$, we must first address the question:

What is meant by weak convergence in $(l^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ *anyway?*

Definition 2.1 (Weak Convergence in General Metric Spaces)

Let $(\mathcal{D}, d)$ be a metric space. Then, $\mathbb{X}_n \rightsquigarrow \mathbb{X}$ in $\mathcal{D}$ if

$$Ef(\mathbb{X}_n) \to Ef(\mathbb{X})$$

for every bounded and continuous functional $f : \mathcal{D} \to \mathbb{R}$.

Weak Convergence in Metric Spaces

Remark 2.1

For empirical processes, $\mathcal{D} = l^\infty(\mathcal{F})$, $\mathbb{X}_n = \mathbb{G}_n$, $\mathbb{X} = \mathbb{G}_P$, and $d(\mathbb{G}_n, \mathbb{G}_P) = ||\mathbb{G}_n - \mathbb{G}_P||_\mathcal{F} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n f - \mathbb{G}_P f|$

Remark 2.2

Under Definition 2.1, one can show that the Continuous Mapping Theorem, Portmanteau Lemmas, relationships between weak, in-probability, and almost-sure convergences, Slutsky's Theorem, and etc., all in their usual forms, hold with this new definition of weak convergence.

Weak Convergence in Metric Spaces

Exercise 2.1 (Continuous Mapping Theorem)

Show that if $\mathbb{X}_n \rightsquigarrow \mathbb{X}$ in $\mathcal{D}$, and $g : \mathcal{D} \to \mathcal{E}$ is continuous, then

$$g(\mathbb{X}_n) \rightsquigarrow g(\mathbb{X}) \text{ in } \mathcal{E}.$$

In particular, if $\mathbb{G}_n \rightsquigarrow \mathbb{G}_P$ in $(l^\infty(\mathcal{F}), ||\cdot||_\mathcal{F})$, then

$$||\mathbb{G}_n||_{\mathcal{F}_0} \rightsquigarrow ||\mathbb{G}||_{\mathcal{F}_0}$$

for any $\mathcal{F}_0 \subset \mathcal{F}$.

The result of Exercise 2.1 justifies the construction of confidence bands suggested in §1.1.

Weak Convergence in Metric Spaces

Obviously, one cannot rely on Definition 2.1 for proving weak convergence, as it is infeasible to test exhaustively all bounded, continuous functionals.

To find a better approach, invoke the concept of *tightness*, which can be viewed as a probabilistic version of compactness.

Recall that a random element in a Euclidean space is tight if for every $\epsilon > 0$, there exists a compact set K such that $\mathrm{P}(X \notin K) < \epsilon$.

Weak Convergence in Metric Spaces

A sequence of random vectors $\{X_n\}$ is asymptotically tight if for every $\epsilon > 0$, there exists a compact set K such that

$$\limsup_{n\to\infty} P(X_n \notin K) < \epsilon.$$

That is, X_n is eventually contained in a compact set with large probability.

Clearly, in the Euclidean case, asymptotic tightness is synonymous to asymptotic boundedness, denoted as $O_P(1)$.

Weak Convergence in Metric Spaces

As one would expect from the property of compact set, the Prohorov Theorem states that if X_n is asymptotically tight, then there exists a sub-sequence of X_n that is weakly convergent.

How would Prohorov Theorem aid us in proving $X_n \rightsquigarrow X$. First, note that $X_n \rightsquigarrow X$ if *every sub-sequence of X_n contains a further sub-sequence that converges weakly to X*.

Weak Convergence in Metric Spaces

But this statement can be further decomposed into two conditions:

(a) Every sub-sequence of X_n, if convergent at all, must converge to the same limit

(b) Every sub-sequence of X_n contains a further convergent sub-sequence

We have seen that (b) is implied by asymptotic tightness of X_n and the Prohorov Theorem.

Weak Convergence in Metric Spaces

Now we argue that (a) is automatic for the EP $\mathbb{G}_n$.

To see this, we use the fact that the law of a random element $\mathbb{G}$ in $l^\infty(\mathcal{F})$ is completely determined by its finite-dimensional distributions, i.e., distributions of

$$\big(\mathbb{G}(f_1), \cdots, \mathbb{G}(f_k)\big), \quad , f_1, \cdots, f_k \in \mathcal{F}, k = 1, 2, \cdots .$$

But if some sub-sequence of $\mathbb{G}_n$ converges weakly to some limit process $\mathbb{G}^*$, by the continuous mapping theorem, its finite-dimensional evaluations also converge weakly to the finite-dimensional evaluations of $\mathbb{G}^*$.

By the ordinary central limit theorem, the finite-dimensional distributions of $\mathbb{G}^*$ must be the same as those of $\mathbb{G}_P$ regardless of the sub-sequence. Since the law of $\mathbb{G}^*$ is completely determined by its finite-dimensional distributions, it is equal to the law of $\mathbb{G}_P$.

Weak Convergence in $l^\infty(\mathcal{F})$

Hence, weak convergence of the EP is *equivalent* to it asymptotic tightness. It remains to characterize asymptotic tightness in $l^\infty(\mathcal{F})$.

To that end, think about what kind of sets are compact in a functional space.

Caution: the intuition inherited from the Euclidean space about (relative) compactness as being bounded may fail in functional spaces.

Weak Convergence in $l^\infty(\mathcal{F})$

Consider the sequence of functions $\{x_n(t), n \geq 2\}$ on $[0,1]$ defined by

$$x_n(t) = nt \cdot \mathbf{1}\big(t \in [0, n^{-1}]\big) + (2 - nt) \cdot \mathbf{1}\big(t \in (n^{-1}, 2n^{-1}]\big)$$

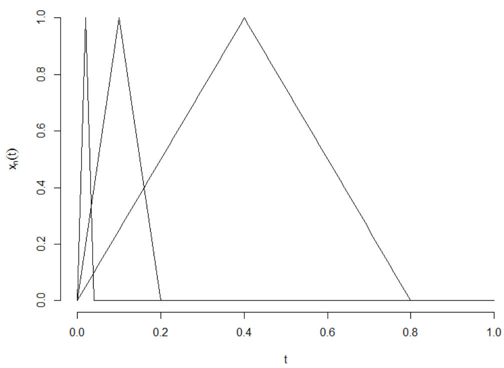

Figure 1: x_n contains no uniformly convergent sub-sequence despite being uniformly bounded.

Weak Convergence in $l^\infty(\mathcal{F})$

The key idea missing is *equi-continuity*. For $\{x_n\}$, the *modulus of continuity* is not controlled as $n \to \infty$. Specifically, denote $\phi_n(\delta) = \sup_{|t-s|<\delta} |x_n(t) - x_n(s)|$, then

$$\limsup_{n\to\infty} \phi_n(\delta) = 1,$$

no matter how small $\delta > 0$ is.

A sequence $\{x_n\}$ is equi-continuous if $\limsup_{n\to\infty} \phi_n(\delta) \downarrow 0$ as $\delta \downarrow 0$.

Weak Convergence in $l^\infty(\mathcal{F})$

Indeed, a version of the Arzelà-Ascoli Theorem states that a sequence $\{x_n\} \subset l^\infty(\mathcal{F})$ is relatively compact iff there exists a metric ρ such that $(\mathcal{F}, \rho)$ is totally bounded * and the x_ns are equi-continuous.

Hence, it is expected that asymptotic tightness in $l^\infty(\mathcal{F})$ is equivalent to a form of *asymptotic equi-continuity (in probability)*. Furthermore, because of the special structure of Gaussian random variables, the metric ρ can always be taken to be the standard deviation metric.

---

*$(\mathcal{F}, \rho)$ being totally bounded means that for every $\epsilon > 0$, there exist a finite number of ϵ balls under ρ that cover $\mathcal{F}$.

Weak Convergence in $l^\infty(\mathcal{F})$

Theorem 2.1 (Weak Convergence of $\mathbb{G}_n$)

The empirical process $\mathbb{G}_n$ converges weakly to a tight P-Brownian bridge in $(l^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ for some $\mathcal{F} \subset L_2(P)$ iff $\mathcal{F}$ is totally bounded with respect to ρ, and for every $\epsilon > 0$

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} P\left(\sup_{\rho(f,g)<\delta} |\mathbb{G}_n(f-g)| > \epsilon \right) = 0. \qquad (2.1)$$

Remark 2.3

Clearly, for the sufficiency part, the standard deviation norm can be replaced by the $L_2(P)$ norm, which hereafter we adopt as the meaning of ρ.

Weak Convergence in $l^\infty(\mathcal{F})$

Remark 2.4

Define the modulus of continuity (MOC) of EP by

$$\phi_n(\delta) = E||\mathbb{G}_n||_{\mathcal{F}_\delta},$$

where $\mathcal{F}_\delta = \{f - g : \rho(f,h) < \delta, f, g \in \mathcal{F}\}$. Then, by Markov inequality, (2.1) is implied by

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \phi_n(\delta) = 0.$$

To place a bound on the MOC as well as showing total boundedness of $\mathcal{F}$, we need the idea of finite approximation.

Contents

Bracketing Numbers

Definition 2.2 (Bracketing Number)

An ϵ-bracket $[l, u]$ in $L_2(P)$ denotes all the functions f with $l \leq f \leq u$ for two functions l and u with $P(u-l)^r < \epsilon^r$. The *bracketing number*

$$N_{[]}(\epsilon, \mathcal{F}, L_r(P))$$

is the minimum number of ϵ-brackets needed to cover $\mathcal{F}$.

The utility of finite approximation is immediately evidenced in the following Glivenko-Cantelli theorem of uniform law of large numbers.

Glivenko-Cantelli Theorem

Theorem 2.2 (Glivenko-Cantelli)

If for every $\epsilon > 0$

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty,$$

then $\mathcal{F}$ is Glivenko-Cantelli in the sense that

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_n - P| f \to_P 0.$$

Glivenko-Cantelli Theorem

Proof of Theorem 2.2:

Given $\epsilon > 0$, let $[l_i, u_i]$ $(i = 1, \cdots, k)$ be an enumeration of $L_1(P)$ brackets covering $\mathcal{F}$. For any $f \in \mathcal{F}$, suppose $f \in [l_i, u_i]$, then

$$\mathbb{P}_n f - Pf \leq \mathbb{P}_n u_i - P u_i + \epsilon,$$

and

$$\mathbb{P}_n f - Pf \geq \mathbb{P}_n l_i - P l_i - \epsilon.$$

The result follows by the ordinary law of large numbers. □

Glivenko-Cantelli Theorem

Example 2.1 (Empirical Distribution Function)

For the empirical distribution function $1(X \leq t), t \in \mathbb{R}$, $\mathcal{F}$ is the class of all indicators $\{1(X \leq t), t \int \mathbb{R}\}$. Choose appropriate knots $-\infty = \xi_0 < \xi_1 < \cdots < \xi_k = \infty$ such that $F(\xi_i-) - F(\xi_{i-1}) < \epsilon$, $i = 1, \cdots, k$. Then, the brackets $[\mathbf{1}(X \leq \xi_{i-1}), \mathbf{1}(X < \xi_i)]$, $i = 1, \cdots, k$, cover $\mathcal{F}$.

It is easy to show that the number of brackets k is to the order of ϵ^{-1}, which is finite. By Theorem 2.2, the classical Glivenko-Cantelli Theorem holds.

Bracketing Entropy

Definition 2.3 (Bracketing Entropy)

The bracketing entropy of class $\mathcal{F}$ is defined as

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon.$$

Lemma 2.1 (Modulus of Continuity)

Suppose $\mathcal{F}$ has a square-integrable envelope function[a] F, then

$$E||\mathbb{G}_n||_{\mathcal{F}_\delta} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(P)),$$

where $A \lesssim B$ means $A \leq cB$ for some positive constant c.

[a] F is an envelope of $\mathcal{F}$ means $\sup_{f \in \mathcal{F}} |f| \leq F$.

Bracketing Entropy

Theorem 2.3 (Donsker Theorem)

If for some $\delta > 0$

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) < \infty,$$

then $\mathcal{F}$ is P-Donsker in the sense that $\mathbb{G}_n$ converges weakly to $\mathbb{G}_P$ in $l^\infty(\mathcal{F})$.

Remark 2.5

The conditions of Theorem 2.3 require that the bracketing entropy be finite, i.e., $\sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))}$ be integrable in ϵ. Once this is satisfied, by Lemma 2.1, the modulus of continuity of the EP goes to 0 as $\delta \downarrow 0$.

Bracketing Entropy

Remark 2.5 (cont'd)

So, the EP is asymptotically equi-continuous and, by Theorem 2.1 and its ensuing Remark, converges weakly in $l^\infty(\mathcal{F})$.

Remark 2.6

If

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim \epsilon^{-r},$$

for some $r \in (0, 2)$. Then the condition of Theorem 2.3 is satisfied with

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) \lesssim \delta^{1-r/2}.$$

Bracketing Entropy

Example 2.1 (Empirical Distribution Function, cont'd)

Following the calculation of the bracketing number of $\{\mathbf{1}(X \leq t) : t \in \mathbb{R}\}$ under $L_1(P)$, the bracketing number under $L_2(P)$ needs the "knots" $-\infty = \xi_0 < \xi_1 < \cdots < \xi_{k'} = \infty$ to satisfy $\sqrt{F(\xi_i-) - F(\xi_{i-1})} < \epsilon$, $i = 1, \cdots, k'$. So that k' is to the order of ϵ^{-2}.

Since the function $\sqrt{\log(1/\epsilon)}$ is clearly integrable, by Theorem 2.3, we have that the classical EP is a Donsker class.

Take-Home Messages

Remark 2.7 (§2.1 and §2.2 in a nutshell)

1. It is well-known that a weakly convergent sequence must be tight;
2. For the EP, tightness turns out to be sufficient for weak convergence given the ordinary central limit theorem;
3. In functional spaces, tightness is more complicated than (uniform) boundedness and requires equi-continuity, which means the modulus of continuity (MOC; i.e., the greatest variation over any small ball of diameter δ) must vanish as $\delta \downarrow 0$;
4. For the EP, the MOC is quantified by the bracketing entropy, i.e., the integral of the square root of the log bracketing number.

Contents

Examples of Donsker Classes

Example 2.2 (Parametric Classes)

Let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$, where Θ is a bounded subset of $\mathbb{R}^k$. Suppose f_θ is Lipschitz in θ in the sense that

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq m(x)||\theta_1 - \theta_2||, \tag{2.2}$$

for every $\theta_1, \theta_2 \in \Theta$ and some $m(x)$ with $Pm^2 < \infty$. Then, after some thought, one finds that the ϵ-bracketing number of $\mathcal{F}$ is of the same order as the number of ϵ-balls that cover Θ. That is

$$N_{[]}(, \epsilon, \mathcal{F}, L_2(P)) \lesssim \epsilon^{-k}.$$

Hence, $\mathcal{F}$ is a Donsker class.

Examples of Donsker Classes

Example 2.3 (Z-Estimation with a Smooth Parameter)

If the estimating function ψ_θ belongs to the smooth parametric class $\mathcal{F}$ defined in 2.2, which is shown to be Donsker therein. Then, we can show that the AC condition (1.2a) holds, i.e.,

$$\mathbb{G}_n \psi_{\widehat{\theta}_n} = \mathbb{G}_n \psi_0 + o_P(1)$$

if $\rho(\psi_{\widehat{\theta}_n}, \psi_{\theta_0}) \to_P 0$.

Suppose $\widehat{\theta}_n \to_P \theta_0$. By the Lipschitz condition 2.2,

$$\rho(\psi_{\widehat{\theta}_n}, \psi_{\theta_0}) \leq \sqrt{Pm^2} ||\widehat{\theta}_n - \theta_0|| \to_P 0.$$

The AC result follows.

Examples of Donsker Classes

Example 2.4 (Mean Absolute Deviation)

Consider estimation of the absolute deviation from the mean, i.e., $\theta = P|X - \mu|$, where $\mu = PX$. A natural nonparametric estimator is

$$\widehat{\theta}_n = \mathbb{P}_n |X - \mathbb{P}_n X|.$$

But, what is the asymptotic distribution of $\widehat{\theta}_n$?

The complication arises from the estimated μ (otherwise $\widehat{\theta}_n$ would have been in i.i.d. averaged form).

Example 2.4 (Mean Absolute Deviation, cont'd)

To linearize $\widehat{\theta}_n$, denote $\widehat{\mu}_n = \mathbb{P}_n X$ and note that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \mathbb{G}_n|X - \widehat{\mu}_n| + \sqrt{n}\Big(P|X - \widehat{\mu}_n| - P|X - \mu_0|\Big). \quad (2.3)$$

To treat the second term on the right, use Taylor expansion on the function

$$g(\mu) = P|X - \mu| = \Big(2F(\mu) - 1\Big)\mu - 2\int_{-\infty}^{\mu} x dF(x) + \mu_0,$$

where F is the distribution of X. By standard calculus,

$$g'(\mu_0) = 2F(\mu_0) - 1.$$

Example 2.4 (Mean Absolute Deviation, cont'd)

By the Delta method,

$$\sqrt{n}\Big(P|X - \widehat{\mu}_n| - P|X - \mu_0|\Big) = \Big(2F(\mu_0) - 1\Big)\mathbb{G}_n(X - \mu_0) + o_P(1).$$

Now, the first term on the right of (2.3) is $\mathbb{G}_n|X - \mu_0| + o_P(1)$ because the class of function $\{|X - \mu| : \mu \in [a, b]\}$ is Lipschitz. So, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \mathbb{G}_n\Big\{|X - \mu_0| - \theta_0 + \big(2F(\mu_0) - 1\big)(X - \mu_0)\Big\} + o_P(1).$$

Examples of Donsker Classes

Exercise 2.2 (Mean Absolute Deviation)

Propose an estimator for variance of $\widehat{\theta}_n$ in Example 2.4 and implement an inference procedure for it with computer code.

Examples of Donsker Classes

We continue to consider parametric families of functions.

Example 2.5 (Monotone Processes)

Denote $N(t)$ as a non-decreasing *cadlag* random process on $[0, \tau]$. Assume (without loss of generality) that $N(0) = 0$ a.s. and $PN^2(\tau) < \infty$.

Now we calculate the bracketing number of $\mathcal{F} = \{N(t) : t \in [0, \tau]\}$. An obvious way to find a partition $0 = \xi_0 < \xi_1 < \cdots < \xi_k = \tau$, and use the brackets $[N(\xi_{i-1}), N(\xi_i-)]$ such that $P(N(\xi_i-) - N(\xi_{i-1}))^2 < \epsilon^2$. But how to relate k, the number of brackets, to ϵ?

Examples of Donsker Classes

Example 2.5 (Monotone Processes, cont'd)

Define $F(t) = PN(t)N(\tau)/c$, where $c = PN(\tau)^2$. Then $F(t)$ is a distribution function on $[0, \tau]$ (Note the parallel with Example 2.1). Since

$$P(N(\xi_i-)-N(\xi_{i-1}))^2 \leq P(N(\xi_i-)-N(\xi_{i-1}))N(\tau) = c\Big(F(\xi_i-)-F(\xi_{i-1})\Big).$$

Thus, the partitions can be chose such that $F(\xi_i-) - F(\xi_{i-1}) < c^{-1}\epsilon^2$. Since F is a distribution function, the cardinality of the partition is to the order of ϵ^{-2}, i.e.,

$$N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim \epsilon^{-2}.$$

In particular, $\mathcal{F} = \{N(t) : t \in [0, \tau]\}$ is a Donsker class.

Examples of Donsker Classes

Finally, some "truly" infinite-dimensional classes.

Example 2.6 (Bounded Monotone/Bounded-Variation Functions)

Let $\mathcal{F}$ be the class of all non-decreasing functions $f : \mathbb{R} \to [-1, 1]$, or all functions with variations bounded by 1. The latter class contains the pair-wise differences between non-decreasing functions taking values in $[0, 1]$. Then, the bracketing entropy

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim \epsilon^{-1}.$$

In particular, by Remark 2.6, $\mathcal{F}$ is a Donsker class.

Examples of Donsker Classes

Remark 2.8 (Monotone Processes *vs* Functions)

The difference between Examples 2.5 and 2.6 must be observed. The former treats functions that are monotone in the (real-valued) index parameter (for each realized data point):

$$f_0(x;t) \text{ monotone in } t \text{ for each } x;$$

The latter deals with functions that are monotone in the (real-valued) data (and these functions may or may not be further indexed by a parameter):

$$f(x) \text{ monotone in } x.$$

Stability of Donsker Classes

Theorem 2.4 (Preservation of Donsker Classes)

Given any fixed Lipschitz function[a] $\phi : \mathbb{R}^2 \to \mathbb{R}$. If $\mathcal{F}$ and $\mathcal{G}$ are Donsker, then

$$\{\phi(f,g) : f \in \mathcal{F}, g \in \mathcal{G}\}$$

is Donsker.

[a] ϕ is Lipschitz if $|\phi(x_1,y_1) - \phi(x_2,y_2)| \lesssim |x_1 - x_2| + |y_1 - y_2|$.

Remark 2.9

If f and g range over Donsker classes, then $f+g$, minima $f \wedge g$, maxima $f \vee g$, fg if f and g are uniformly bounded, and $\phi(f)$ for any fixed Lipschitz function ϕ, etc., range over Donsker classes.

Infinite-Dimensional Estimating Functions

Proposition 2.1 (Asymptotic Continuity for Infinite-Dimensional Estimating Functions)

Let $\psi_\eta(t)$, $t \in \mathcal{T}$, be an infinite-dimensional estimating function for parameter η, and let $\widehat{\eta}_n$ be the Z-estimator such that $d(\widehat{\eta}_n, \eta_0) \to_P 0$ for some metric d. Then the asymptotic continuity holds, i.e.,

$$\mathbb{G}_n \psi_{\widehat{\eta}_n}(t) = \mathbb{G}_n \psi_{\eta_0}(t) + o_P(1)$$

uniformly in $t \in \mathcal{T}$, if for some $\delta > 0$ the class

$$\{\psi_\eta(t) : d(\eta, \eta_0) < \delta, t \in \mathcal{T}\}$$

is Donsker and $\sup_{t\in\mathcal{T}} \rho(\psi_{\widehat{\eta}_n}(t), \psi_{\eta_0}(t)) \to_P 0$.

Nelsen-Aalen Estimator for Recurrent Event

Example 1.3 (Nelsen-Aalen Estimator, cont'd)

In Lecture 1 we have heuristically derived the asymptotic distributions of the Nelsen-Aalen type estimator for the mean function $\mu(t)$ of recurrent event. For the arguments to go through, we need to show that

$$\mathbb{G}_n M_{\widehat{\mu}_n}(t) = \mathbb{G}_n M_{\mu_0}(t) + o_P(1),$$

uniformly in $t \in [0, \tau]$.

Assume that $PN(\tau)^2 < \infty$ and that $P(C \geq \tau) > 0$. First note that $\widehat{\mu}_n(t) = \int_0^t \frac{d\mathbb{P}_n N(s)}{\mathbb{P}_n Y(s)}$ converges in probability to $\mu_0(t)$ uniformly in t using the Donsker (Glivenko-Cantelli) results about monotone processes.

Nelsen-Aalen Estimator for Recurrent Event

Example 1.3 (Nelsen-Aalen Estimator, cont'd)

Hence, $\widehat{\mu}_n$ eventually ranges over uniformly bounded monotone functions on $[0, \tau]$; denote this class of functions as $\mathcal{U}$.

Then consider the class

$$\mathcal{F} = \{M_\mu(t) : \mu \in \mathcal{U}, t \in [0, \tau]\}.$$

We are dealing with functions of the form

$$N(t) - \mu(C \wedge t), \qquad \mu \in \mathcal{U}, t \in [0, \tau]$$

Nelsen-Aalen Estimator for Recurrent Event

Example 1.3 (Nelsen-Aalen Estimator, cont'd)

By Example 2.5, $\{N(t), t \in [0, \tau]\}$ is Donsker. Since for each fixed μ and t, the function $\mu(\cdot \wedge t)$ is a non-decreasing function, by Example 2.6, $\{\mu(C \wedge t) : \mu \in \mathcal{U}, t \in [0, \tau]\}$ is Donsker. Then, by the preservation property of Donsker classes given in Theorem 2.4, $\mathcal{F}$ is Donsker.

Finally, $L_2(P)$ convergence of $M_{\widehat{\mu}_n}$ to M_{μ_0} follows easily from the uniform convergence of $\widehat{\mu}_n$.

Concluding Remarks

This lecture is a crash course for empirical processes, from definition to application. To gain some intuition of why the function $\sqrt{\log x}$ comes up in the calculation of entropy, see the review paper by Pollard (1989).

For readers wishing to know more about EP, Jon A. Wellner's lecture notes (`https://www.stat.washington.edu/jaw/COURSES/EPWG/w13.html`) may be a good starting point.

References

- Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science*, 341-354.