What to do this week (2022/03/8 - 03/10)?

Part I. Preliminaries

Part II. Parametric Inference

Part III. Nonparametric/Semi-parametric Inference

Part IV. Advanced Topics

Part IV.1 Counting Process Formulation

Part IV.1.1 Theoretical Preparation

Planning ... ...

- **Data Analysis Project**
  - Phase I. Analysis Plan (due by Mar 14)
  - Phase II. Presentation (the in-class presentation: Mar 31, Apr 5 and 7)
  - Phase III. Analysis Report (the final project: Apr 22)
- **Part IV. Advanced Topics** (*Homework 4*)
  - Part IV.1 Counting Process Formulation
    (Revisits to KM estm, logrank test, and Cox PH model)
  - Part IV.2 Selected Recent Topics in LIDA
    - Alternative models to Cox PH model: accelerated failure time model (AFT), linear transformation model, ...
    - Multivariate survival analysis
    - Other incomplete data structures: competing risk, interval censoring, current status, truncation, missing covariates in LIDA ...
    - *Alternative approach to the martingale-based one*
  - Part IV.3 Beyond Lifetime Data Analysis

Part IV.1 Counting Process Formulation

Revisits to the nonparametric/semi-parametric approaches of Part III ... ...

Part IV.1.1 Theoretical Preparation

- IV.1.1A Basic concepts
- IV.1.1B An introduction to stochastic process
- IV.1.1C Counting process and martingale: the key results

Part IV.1.1A Theoretical Preparation: Basic concepts

Probability Space: a triplet $(\Omega, \mathcal{F}, \mathcal{P})$

- the sample space Ω: a non-empty set

- $\mathcal{F}$ is $\subseteq 2^{\Omega}$ (the collection of all subsets of Ω), and a σ-algebra.

  (i) $\emptyset \in \mathcal{F}$;
  (ii) If $A \in \mathcal{F}$, $\bar{A} \in \mathcal{F}$ (closed under complements);
  (iii) If $A_j \in \mathcal{F}$, $j = 1, \ldots$, $\bigcup_j A_j \in \mathcal{F}$ (closed under countable unions, and thus under countable intersections)

- $\mathcal{P}$ is a probability measure: $\mathcal{F} \to [0, 1]$

  (i) $\mathcal{P}(\Omega) = 1$
  (ii) measure: non-negative, countable additive, $\mathcal{P}(\emptyset) = 0$

Part IV.1.1A Theoretical Preparation: Basic concepts

Random Variable: Given $(\Omega, \mathcal{F}, \mathcal{P})$, r.v. $X : \Omega \to \mathcal{R}$ (real-valued) and is measurable.
That is, $\forall x \in \mathcal{R}$, $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$.

- $F_X(x) = \mathcal{P}(\omega : X(\omega) \leq x)$ is the cumulative distribution of X.
- The r.v. X then induces another probability space, $(\mathcal{X}, \mathcal{B}, P_X)$:

  (i) $\mathcal{X}$ is the collection of all possible values of X, a subset of $\mathcal{R}$
  (ii) $\mathcal{B}$ is the Borel σ-algebra, the σ-algebra generated by all $(-\infty, x]$ sets
  (iii) P is the probability measure: $\mathcal{B} \to [0, 1]$ with $P_X((-\infty, x]) = F_X(x)$.

Part IV.1.1A Theoretical Preparation: Basic concepts

Example 4.1: Consider (i) “tossing an even coin”, (ii) “a student's mark at the final exam of STAT-475 taught by JHu”

Part IV.1.1A Theoretical Preparation: Basic concepts

Integration – Riemann Integral:

$$\int_a^b f(x)dx = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} f(u_i)(x_i - x_{i-1}),$$

provided the limit exists.

- With the definition, the integrand needs to be almost continuous.
- e.g. $f(x) = 1, 0$ if x is rational or not

Part IV.1.1A Theoretical Preparation: Basic concepts

Integration – Lebesgue Integral:
Given a measure space $(\mathcal{E}, \mathcal{S}, \mu)$ with $\mathcal{E}$ an Euclidean space and the Lebesgue measure μ, $\int_E f d\mu = \int_{E_x} f(x)\mu(dx)$

(i) for a set indicator, $\int 1_S d\mu = \mu(S)$ for $S \in \mathcal{S}$.

(ii) for a simple function, $\int \sum a_k 1_{S_k} d\mu = \sum_k a_k \mu(S_k)$ for disjoint $S_k \in \mathcal{S}$.

(iii) for a non-negative function,
$\int_E f d\mu = \sup \left\{ \int_E g d\mu : 0 \leq g \leq f,\ g\ simple \right\}$, provided exist
(iv) for a general function, $f = f^+ - f^-$ and
$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$, provided $\int |f| d\mu < \infty$.

Part IV.1.1A Theoretical Preparation: Basic concepts

- If $f(\cdot)$ is Riemann-integrable, it is Lebesgue-integrable. e.g. the example of the indicator of rational numbers
- the commonly used properties
  - linearity
  - monotonicity
  - monotone convergence theorem

Part IV.1.1A Theoretical Preparation: Basic concepts

Riemann-Stieltjes Integral

$$\int_a^b f(x)dg(x) = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} f(u_i)[g(x_i) - g(x_{i-1})], \quad \textit{provided exist}$$

Lebesgue-Stieltjes Integral

$$\int_a^b f(x)dg(x) = \int_a^b f(x)\mu_g(dx)$$

Provided $g : [a, b] \to \mathcal{R}$ with bounded variation, there exists the unique Boreal measure μ_g on $[a, b]$ such that $\mu_g((s, t]) = g(t) - g(s)$.

Example 4.2. the expectation of a r.v. X

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Stochastic Process A collection of r.v.s defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and indexed by t in a set I:

$$\{X(\omega; t) : \omega \in \Omega, t \in I\},$$

donoted by $X(\cdot)$

- $I = \{1, 2, 3, \ldots\} \Longrightarrow$ a sequence of r.v.s
- $I = [0, \infty) \Longrightarrow$ a time continuous process
- $I = [0, 1] \times [0, 1] \Longrightarrow$ a random field

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Examples

- survival process of a subject
- counts of asthma attacks over time of a child
- air pollution level over time at Vancouver airport
- air pollution level over time across Canada

- *Realization/Sample Path*
  With $\omega_0 \in \Omega$,
  $$\{X(\omega_0; t) : t \in I\},$$
  a function of $t \in I$
  - continuous sample path
  - cadlag path: right-continuous and left-limit-exist

- *Modification vs Indistinguishable*
  - If $\forall t \in I$, $P(\omega : X(\omega, t) = Y(\omega, t)) = 1$, $X(\cdot)$ is a modification of $Y(\cdot)$.
  - If $P(\omega : X(\omega; t) = Y(\omega; t), \forall t) = 1$, $X(\cdot)$ and $Y(\cdot)$ are indistinguishable.

- *Filtration (History)* Given a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, a sequence of non-decreasing σ-algebra $\mathcal{F}_t \subseteq \mathcal{F}$ for $t \in I$: $\{\mathcal{F}_t : t \in I\}$
  - "history" $\mathcal{H}_t = \sigma(X_s : 0 \leq s \leq t)$

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Gaussian Process $X(\cdot) = \{X_t : t \in I\}$ is a Gaussian process, if any its finite dimensional distribution are multivariate normal, characterized by mean $\mu(t) = E(X_t)$ and covariance $c(s,t) = Cov(X_t, X_s)$.

Special Cases:

- Wiener Process (Brownian Motion)
- Brownian Bridge on $[0,1]$
- Gaussian Random Walk

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Poisson Process A stochastic process $\{N(\omega; t) : \omega \in \Omega, t \geq 0\}$ is the Poisson process of rate ρ, if, as $\delta \to 0+$,

- (i) $P(N(t+\delta) - N(t) = 1|\mathcal{H}_t) = \rho\delta + o(\delta)$, and
- (ii) $P(N(t+\delta) - N(t) > 1|\mathcal{H}_t) = o(\delta)$, so that
- (iii) $P(N(t+\delta) - N(t) = 0|\mathcal{H}_t) = 1 - \rho\delta + o(\delta)$.

Intensity Specification: The intensity of $N(\cdot)$ is

$$\lim_{\delta \to 0+} \frac{1}{\delta} P(N(t+\delta) - N(t) \geq 1|\mathcal{H}_t) = \rho$$

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

- *Interval Specification:* $N(\cdot)$ is a Poisson process with rate ρ if the subsequent points where $N(\cdot)$ have jumps are at times $X_1, X_1 + X_2, \ldots$ and r.v.s $X_1, X_2, \ldots$ (the gap times) are iid $\sim \rho e^{-\rho x}$.
  - $T_r = X_1 + \ldots + X_r$, the time of the rth point (event), $\sim \frac{1}{\Gamma(r)} \rho (\rho t)^{r-1} e^{-\rho t}$, a Gamma distn.
  - a convenient way to simulate a Poisson process
- *Counting Specification* $N(\cdot)$ is a Poisson process with rate ρ if $\forall A_1, \ldots, A_k$ disjoint sets of $\mathcal{B}(0, \infty)$, $N(A_1), \ldots, N(A_k)$ are $\perp\!\!\!\perp$ and with the Poisson distn of mean $\rho |A_j|$, where $N(A)$, a non-negative integer r.v., is the count of events over time period A.
  - $|A_j|$ is the Lebesgue measure of A_j: the length of A_j if it's an interval.

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Extensions of Poisson Process:

The intensity of a Poisson process $N(\cdot)$ with rate of ρ is

$$\lim_{\delta \to 0+} \frac{1}{\delta} P(N(t+\delta) - N(t) \geq 1 | \mathcal{H}_t) = \rho.$$

time-homogeneous

- *time-inhomogeneous Poisson process* $N(\cdot)$ with its intensity function of $\rho(t)$:

$$\lim_{\delta \to 0+} \frac{1}{\delta} P(N(t+\delta) - N(t) \geq 1 | \mathcal{H}_t) = \rho(t).$$

- *mixed Poisson process* Conditional on $\xi \sim G(\cdot)$, $N(\cdot)$ is a Poisson process with rate of $\xi\rho$
  - $N(\cdot)$'s increments are not indepdent.
  - *overdispersion*

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Concepts of Convergence with Stochastic Process

Recall ... ...

- With $\{x_1, x_2, \ldots\}$, a sequence of constants,

$$\lim_{n\to\infty} x_n = x^*$$

- With $\{X_1, X_2, \ldots\}$, a sequence of r.v.s,
  - $\lim_{n\to\infty} X_n = X^*$ a.s. (almost surely)
  - $\lim_{n\to\infty} X_n = X^*$ in prob
  - $\lim_{n\to\infty} X_n = X^*$ in distn (weak convergence)

With $\{X_1(\cdot), X_2(\cdot), \ldots\}$, a sequence of stochastic processes, $||X_n(\cdot) - X^*(\cdot)|| = \sup_{t \in I} |X_n(t) - X^*(t)|$,

- $\lim_{n\to\infty} X_n(\cdot) = X^*(\cdot)$ in prob
  $\forall \epsilon > 0, P(\omega : ||X_n(\cdot) - X^*(\cdot)|| > \epsilon) \to 0$ as $n \to \infty$

- $\lim_{n\to\infty} X_n(\cdot) = X^*(\cdot)$ a.s. (almost surely)
  $||X_n(\cdot) - X^*(\cdot)|| \to 0$ a.s.

- $\lim_{n\to\infty} X_n(\cdot) = X^*(\cdot)$ in distn (weak convergence)

  if $\forall f$, real valued, bounded, measurable on $(\mathcal{M}, \delta)$,
  $\int_{\mathcal{M}} f d\mathcal{P}_n \to \int_{\mathcal{M}} f d\mathcal{P}^*$ as $n \to \infty$. That is,
  $E\big[f(X_n)\big] \to E\big[f(X^*)\big]$.

Part IV.1.1B Theoretical Preparation: An introduction to Stochastic Process

Often-Used Results (an analogue in stochastic processes to its version in r.v.s)

- Slutsky's Theorem. If $X_n(\cdot) \to X(\cdot)$ in distn and $Y_n(\cdot) \to m(\cdot)$ in prob, $m(\cdot)$ a constant function, then
  $X_n(\cdot) + Y_n(\cdot) \to X(\cdot) + m(\cdot)$ in distn and
  $X_n(\cdot)Y_n(\cdot) \to m(\cdot)X(\cdot)$ in distn

  e.g. when $m(\cdot) = a$

- Continuous Mapping Theorem. If $X_n(\cdot) \to X(\cdot)$ in distn, $\forall f$, continuous, $f \circ X_n(\cdot) = f\left[X_n(\cdot)\right] \to f \circ X(\cdot)$ in distn.

  e.g. $Y_n = \sup_t |X_n(t)|$

Example. $T_1, \ldots, T_n \sim F(\cdot)$ iid. The empirical distn based on the data:

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i \leq t)$$

- With $t = t_0$
- Over $t \in [0, \infty)$

Empirical Process Theory

Part IV.1.1C Martingale: An Introduction

Martingale Definition
Consider a stochastic process $X(\cdot) = \{X(\omega; t) : \omega \in \Omega, t \geq 0\}$ on $(\Omega, \mathcal{F}, \mathcal{P})$, adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$.

Suppose $X(\cdot)$ is right-continuous left-hand limits. It is a martingale wrt the filtration $\{\mathcal{F}_t\}$ if

- (a) $\forall t$, $E|X(t)| < \infty$ (integrable)
- (b) $\forall t$, $E\big[X(t+s)|\mathcal{F}_t\big] = X(t)$ a.s. for $s \geq 0$.

- *"fair game" in gambling*: $E(X(t)) = E(X(0))$
- *Sub (Super) Martingale:* "$=$" in (b) of a martingale definition is replaced by $\geq$ ($\leq$)

Part IV.1.1C Martingale: An Introduction

Example. *Random Walk* $X_t = \sum_{j=1}^{t} Y_j$, $t = 1, 2, \ldots$ and iid

$$Y_j = \begin{cases} 1, & \text{with } Pr\ 1/2 \\ -1, & \text{with } Pr\ 1/2 \end{cases}$$

Part IV.1.1C Martingale: An Introduction

- *Predictable.* stochastic process $X(\cdot)$ is predictable wrt $\{\mathcal{H}_t\}$ if $X(t)$ is determined by $\mathcal{H}_{t-} = \sigma(X(u) : 0 \le u < t)$.

  Processes with left-continuous sample path are predictable.

- *Stopping Time.* r.v. τ is a stopping time wrt a filtration $\{\mathcal{F}_t\}$ if $\{\tau \le t\} \in \mathcal{F}_t, \forall t$.

- *Stopped Process.* Suppose $X(\cdot)$ is a stochastic process adapted to $\{\mathcal{F}_t\}$, and τ a stopping time wrt $\{\mathcal{F}_t\}$. We call $X^\tau(\cdot)$ a stopped process by τ: $X^\tau(t) = X(t \wedge \tau)$.

- *Notion of "Localization" in Theory.*
  - A sequence of stopping times, non-decreasing $\{\tau_n\}$ is called "localizing sequence" if $P(\tau_n \geq t) \to 1$ as $n \to \infty$ for $t \in \mathcal{T}$.
  - We say $X(\cdot)$ has a certain property locally if there is a localizing sequence $\{\tau_n\}$ such that, $\forall n$, $I(\tau_n > 0)X^{\tau_n}(\cdot)$ has the property.

e.g. $X(\cdot)$ is locally bounded.

e.g. $X(\cdot)$ is a local martingale.

Part IV.1.1C Martingale: An Introduction

Important Martingale Results

- Provided $\{M(t) : t \in \mathcal{T}\}$ is a martingale wrt $\{\mathcal{F}_t\}$ and $M(0) = 0$,
  - $E(M(t)) = 0, \forall t \in \mathcal{T}$
  - $Cov(M(t+s) - M(t), M(t)) = 0$ and $Cov(M(t), M(t+s)) = Var(M(t))$, $\forall t \in \mathcal{T}, s > 0$

- Provided $\{M_1(t) : t \in \mathcal{T}\}$ and $\{M_2(t) : t \in \mathcal{T}\}$ are martingales wrt $\{\mathcal{F}_t\}$, $M(\cdot) = aM_1(\cdot) + bM_2(\cdot)$ is a martingale.

- Provided $\{M(t) : t \in \mathcal{T}\}$ is a martingale wrt $\{\mathcal{F}_t\}$, $M^2(\cdot)$ is a submartingale wrt $\{\mathcal{F}_t\}$.

Part IV.1.1C Martingale: An Introduction

Doob-Meyer Decomposition. If $X(\cdot)$ is a submartingale adapted to $\{\mathcal{F}_t : t \geq 0\}$, there exists a unique, right-continuous, non-decreasing, predictable process $A(\cdot)$ with $A(0) = 0$ and $M(\cdot) = X(\cdot) - A(\cdot)$ is a martingale wrt $\{\mathcal{F}_t : t \geq 0\}$. The process $A(\cdot)$ is called the *compensator* for $X(\cdot)$.

Example. $N(\cdot)$ is a Poisson process with intensity function ρ. (i) $N(\cdot)$ is a submartingale. (ii) Its compensator $A(\cdot)$ is $A(t) = \int_0^t \rho(u)du$, $t \geq 0$.

This result holds for any counting process $N(\cdot)$.

Part IV.1.1C Martingale: An Introduction

- If $M(.)$ is martingale and (locally) square integrable, M^2 is a (local) sub-martingale and has a compensator, denoted by $< M, M > (.)$ such that $M^2(.) - < M, M > (.)$ is a martingale.

  The process $< M, M > (.)$ is called the predictable quadratic variation of $M(.)$.

  - If $E[M(t)] = 0$, then $var[M(t)] = E[M^2(t)] = E[< M, M > (t)]$.
  - If the zero mean martingale $M = N - A$ satisfies $E[M^2(t)] < \infty$ and $A(.)$ is continuous, then $< M, M > (.) = A(.)$, a.s.

Part IV.1.1C Martingale: An Introduction

- $M_1, M_2, \ldots$ are zero mean martingales defined on the same filtration.
  - If both M_1 and M_2 are (locally) square integrable, then there is a right continuous predictable process $< M_1, M_2 >$ such that $< M_1, M_2 > (0) = 0$, $E[\big| < M_1, M_2 > \big|] < \infty$, and $M_1(.)M_2(.) - < M_1, M_2 > (.)$ is a martingale.
  - If $< M_1, M_2 >= 0$, a.s., M_1 and M_2 are called orthogonal.
  - If $M_j = N_j - A_j$, where N_j is a counting process and A_j its (continuous) compensator, then if N_i and N_j don't jump at the same ime, $< M_i, M_j > (.) = 0$, a.s.

Suppose that $N(.)$ is a counting process with $E[N(t)] < \infty$, that $A(.)$ is the compensator for $N(.)$ and $H(.)$ a bunded, predictable process. Define $Q(.)$ by

$$Q(t) = \int_0^t H(s)dM(s), \text{ where } M(.) = N(.) - A(.).$$

- $Q(.)$ is a zero-mean martingale.
- $var[Q(t)] = E[Q^2(t)]$
- $Q^2(.)$ is a sub-martingale, and thus there is $< Q, Q > (.)$ such that $Q^2(.) - < Q, Q > (.)$ is a martingale. Hence, $var[Q(t)] = E[< Q, Q > (t)]$
- If $A(.)$ is continuous, $< Q, Q > (t) = \int_0^t H^2(s)dA(s)$, a.s.
- If M_1, M_2 are martingales, H_1, H_2 bounded predictable processes, and Q_1, Q_2 are defined in the same way as $Q(.)$, the predictable quadratic covariance process $< Q_1, Q_2 >$ satisfies $< Q_1, Q_2 > (t) = \int_0^t H_1(s)H_2(s)d < M_1, M_2 > (s)$.

Local Properties:

- M is a local martingale (sub-martingale) if $\exists$ a localizing sequence $\{\tau_n\}$ such that for each n, $M_n(.) = \{M(t) : 0 \leq t \leq \tau_n\}$ is a martingale (sub-martingale).
- If as above but we also have $\sup_t E[M_n^2(t)] < \infty$ for all n, M is called a locally square-integrable martingale.
- **Extended Doob-Meyer Decomposition**: If $X(.)$ is a local sub-martingale, $\exists$ a right continuous, nondecreasing predictable process $A(.)$ such that $X(.) - A(.)$ is a local martingale.

By the extended Doob-Meyer Decomposition,

- If $N(.)$ is a counting process (not necessaryly $E[N(t)] < \infty$), then $N(.)$ is a local sub-martingale. Thus, $\exists$ a unique $A(.)$ (nondecreasing, right-continuous, predictable) such that $M(.) = N(.) - A(.)$ is a local martingale.

- If $A(.)$ is locally bounded, then $M(.)$ above is a local square integrable martingale.

- If $H(.)$ is locally-bounded and predictable, and $M(.)$ is a local martingale, then $Q = \int HdM$ is a locally square integrable martingale.

Part IV.1.1C Martingale: An Introduction

Martingale Central Limit Theorem (CLT)
For $i = 1, \ldots, n$, let $N_{in}(.)$ be counting process, $A_{in}(.)$ its continuous compensator, $H_{in}(.)$ locally bounded predictable process. Define

$$M_{in} = N_{in} - A_{in}, \quad U_{in} = \int H_{in} dM_{in}, \quad U_{in,\epsilon} = \int H^*_{in} dM_{in},$$

$$H^*_{in} = H_{in} \mathbf{1}[|H_{in}| \geq \epsilon], \quad \bar{U}_n = \sum_i U_{in}, \quad , \bar{U}_{n,\epsilon} = \sum_i U_{in,\epsilon}$$

Suppose that

(a) $< \bar{U}_n, \bar{U}_n > (t) \to \alpha(t)$ in prob, $\forall t \geq 0$ and some function $\alpha(.)$.

(b) $< \bar{U}_{n,\epsilon}, \bar{U}_{n,\epsilon} > (t) \to 0$ in prob, $\forall t \geq 0$ and $\forall \epsilon \geq 0$.

Then $\bar{U}_n \to U = \int f dW$ weakly, where $W(.)$ is a Wiener process and $\int_0^t f^2(s) ds = \alpha(t)$.

Part IV.1.1C Martingale: An Introduction

That is, the limiting process $U(.)$ is a zero-mean Gaussian process with independent increments and $var[U(t)] = \alpha(t)$.
By the definitions of N_{in}, A_{in}, and H_{in},

$$< \bar{U}_n, \bar{U}_n > (t) = \sum_{i=1}^{n} \int_0^t H_{in}^2(s) dA_{in}(s),$$

$$< \bar{U}_{n,\epsilon}, \bar{U}_{n,\epsilon} > (t) = \sum_{i=1}^{n} \int_0^t H_{in}^{*2}(s) dA_{in}(s).$$

What to study next?

Part IV. Advanced Topics

- **Part IV.1 Counting Process Formulation** (Revisits to KM estm, logrank test, and Cox PH model)
  - *Part IV.1.1 Theoretical Preparation*
  - **Part IV.1.2 Counting Process Formulation in LIDA and Applications: Revisits to KM, logrank, Cox PH**
- *Part IV.2 Selected Recent Topics in LIDA*
- *Part IV.3 Beyond Lifetime Data Analysis*