

Санкт-Петербургский государственный университет

Направление (математика)

Профиль (теория вероятностей)

Ильин Алексей Геннадьевич

Категориальные методы анализа данных типа времени жизни

Выпускная квалификационная работа

Научный руководитель:

Кандидат физико-математических наук, доцент С.В.Малов

Рецензент:

Доктор технических наук, профессор Уткин Л. В.

Санкт-Петербург

2016

SAINT-PETERSBURG STATE UNIVERSITY

Main Field of Study (Mathematics)

Area of Specialization (probability theory)

Aleksei Ilin

Categorical methods in survival analysis

Final qualifying work

Scientific supervisor:

Candidate of Physics and Mathematics, Associate Professor S.V.Malov

Reviewer:

Doctor of Technical Sciences, Professor Utkin L.V.

Saint-Petersburg

2016

2

Введение

Целью данной выпускной квалификационной работы было описать основные модели и методы обработки данных типа времени жизни, основанных на точечных процессах и применение их для построения категориальных тестов. Данное направление в современной математике активно развивается. Полученные на сегодняшний день результаты позволяют применять их в различных отраслях (медицина, машиностроение, психология и др.). В качестве основы для написания, были использованы следующие материалы:

1. Основные понятия и определения

*Основан на материале: Fleming [1], Ширяев[2], Лебедев[3], Кокс, Оукс[5]
Даны вводные определения, необходимы для других разделов*

2. Оценки Нельсона-Аалена и Каплана-Мейера, представление с помощью случайного процесса.

*Основанной на материале: Fleming[1]
Даны базовые понятия оценок Нельсона-Аалена и Каплана-Мейера*

3. Состоятельность оценки Нельсона-Аалена и Каплана-Мейера.

*Основан на материале Fleming
Описаны состоятельность оценок Нельсона-Аалена и Каплана-Мейера*

4. Мартингальная Центральная Предельная Теорема.

Основан на материале: Fleming[1]

5. Категориальный данные.

*Основан на материале: Malov&O`Brien [4]
Описывается метод построения классических категориальных тестов и категориальных тестов типа времени жизни.*

Оглавление

1. Основные понятия и определения
 - 1.1 Введение в данные типа времени жизни.
 - 1.2 Точечные процессы.
 - 1.3. Представление данных с помощью точечных процессов
 - 1.4 Описание цензурированных данных с помощью точечных процессов
 2. Оценки Нельсона-Аалена и Каплана-Мейера, представление с помощью случайного процесса.
 - 2.1 Введение
 - 2.2 Непараметрическое оценивание функции отказов.
 - 2.3 Оценка Каплана-Мейера
 3. Состоятельность оценки Нельсона-Аалена и Каплана-Мейера.
 4. Мартингальная Центральная Предельная Теорема.
 5. Категориальный данные.
 - 5.1 Классические категориальные и их применение в анализе данных типа времени жизни.
 - 5.2 Категориальные тесты типа времени жизни.
- Список литературы

1. Основные понятия и определения.

1.1 Введение в данные типа времени жизни.

В анализе *данных типа времени жизни* рассматриваются группы объектов исследования (индивидов), для которых определено точечное событие, называемое отказом. Отказ происходит после некоторого интервала времени (наработки до отказа) для каждого объекта только один раз. Примерами наработок до отказа могут служить: продолжительность работы механизмов машин в технике, жизни больных при клинических исследованиях и др.

Для точного определения наработки до отказа необходимо три условия:

1. четко установить начало отсчета времени;
2. выбрать масштаб для измерения отсчета времени
3. точным должно быть само понятие отказа

При анализе данных типа времени жизни, у некоторых объектов может не наблюдаться отказ в течение полного времени исследования. Некоторые больные (многие, как хотелось бы надеяться) будут жить к концу клинического исследования. Объект, который наблюдался и не имел отказов в течение нескольких дней, а затем был выключен из числа наблюдавшихся, имеет наработку до отказа. Такое неполное наблюдение наработки до отказа называется *цензурированием*. Заметим, что как и отказ, цензурирование является *точечным событием*, а период наблюдения цензурированных объектов должен регистрироваться.

Предположим, что при отсутствии цензурирования i -й объект в выборке объема n имеет наработку до отказа T_i , (где T_i — случайная величина). Предположим также, что имеется такой интервал времени U_i при котором наблюдение над этим объектом прекращается в момент U_i , если отказ не произойдет раньше. Тогда в действительности наблюдаются величины $X_i = \min(T_i, U_i)$ совместно с индикаторной переменной $\delta_i = 1$, если $T_i \leq U_i$ (объект нецензурирован), $\delta_i = 0$, если $T_i > U_i$ (объект цензурирован). Величины U_i для объектов, отказ которых наблюдался, считаются *нереализованными моментами цензурирования* в противоположность реализованным моментам цензурирования. Термин «потенциальный момент цензурирования» обычно употребляется в том случае, когда U_i рассматриваются независимо от того, произошло ли цензурирование или отказ.

Цензурирование типа I, при котором все U_i равны, $U_i = U$ (где U не случайная величина). Например, в некоторых приложениях все моменты U_i известны, например, если единственная причина цензурирования состоит в запланированном окончании испытаний через заранее определенный срок. При цензурировании типа II наблюдение прекращается после достижения заранее определенного числа отказов d , так что U становится случайной величиной. В общем случае предполагается, что T, U — случайные величины.

1.2 Точечные процессы.

Определение 1.1: Случайным процессом мы будем называть семейство случайных величин $\{\xi_t\}$, определенных на вероятностном пространстве (Ω, \mathcal{F}, P) , где t – принадлежит некоторому параметрическому множеству T .

Параметр t чаще всего понимают как время. Также случайный процесс можно представить в виде отображения $\xi: \Omega \times T \rightarrow R$. Если мы зафиксируем t , то получим случайную величину ξ_t , если зафиксируем ω , то получим функцию $\xi(\omega, \cdot)$, заданную на множестве T , которую называют траекторией случайного процесса или реализацией процесса.

Множество исходов Ω удобно трактовать как набор реализаций.

Реализации могут обладать свойствами: непрерывность, дифференцируемость.

Определение 1.2: Случайный процесс $X(t, \omega)$ измерим, если отображение $X: (\Gamma \times \Omega, \sigma(\mathcal{D} \times \mathcal{F})) \rightarrow (R, \mathcal{B})$ – измеримо, где $\sigma(\mathcal{D} \times \mathcal{F})$ – минимальная σ -алгебра, содержащая все множества вида $I \times A$, $I \in \mathcal{D}$, $A \in \mathcal{F}$; \mathcal{B} – Борелевская σ -алгебра на прямой.

Определение 1.3: Процесс $\xi(t)$ называется процессом с независимыми приращениями, если для любых $0 = t_0 < t_1 < t_2 < \dots < t_n$ возрастающих моментов времени случайные величины $\xi(t_0), \xi(t_1) - \xi(t_0), \xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$ являются независимыми.

Разность $\xi(t_n) - \xi(t_{n-1})$ называется приращением.

Определение 1.4: Процесс $\{X_t\}_{t \in R_+}$

- Интегрируем, если $\sup_{0 \leq t < \infty} E|X(t)| < \infty$
- Квадратично интегрируем, если $\sup_{0 \leq t < \infty} E|X(t)|^2 < \infty$
- Ограничен, если $P(\sup_{0 \leq t < \infty} |X(t)| < M) = 1, M < \infty$
- Равномерно интегрируем, если $\lim_{n \rightarrow \infty} \sup_{0 \leq t < \infty} E(|X(t)| \cdot I_{\{X(t) > n\}}) = 0$

Определение 1.5: Семейство под σ -алгебр $\{F_t, t \geq 0\}$ σ -алгебры F такое, что $F_s \subseteq F_t, s \leq t$ называется фильтрацией.

Пусть $\{F_t, t \geq 0\}$ – фильтрация. Определим σ -алгебры $F_{t+} = \bigcap_{h>0} F_{t+h}$ и $F_{t-} = \bigcap_{h>0} F_{t-h}$. Фильтрация непрерывна справа, если $F_{t+} = F_t, \forall t \in R_+$

Определение 1.6: *Стохастическим базисом* будем называть вероятностное пространство (Ω, \mathcal{F}, P) с непрерывной справа фильтрацией $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$

Определение 1.7: Случайный процесс $\{X_t\}_{t \in \mathbb{R}_+}$ согласован с фильтрацией $\{\mathcal{F}_t, t \geq 0\}$, если для любого $t \geq 0$ случайная величина X_t является \mathcal{F}_t измеримой.

Определение 1.8: *Интегралом (стохастическим)* измеримого процесса $\{X_t\}_{t \geq 0}$ по процессу ограниченной вариации $\{Y_t\}_{t \geq 0}$, заданному на том же вероятностном пространстве, будем называть случайную величину

$$X \cdot Y_t(\omega) = \int_0^t X(u, \omega) Y(du, \omega), \omega \in \Omega$$

где потраекторный интеграл понимается в смысле Лебега-Стилтьеса.

Определение 1.9: Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, X – случайная величина, $\mathcal{A} \subseteq \mathcal{F}$. *Условным математическим ожиданием* $E(X|\mathcal{A})$ назовем случайную величину, удовлетворяющую следующим условиям:

- $E\{X|\mathcal{A}\}$ - случайная величина измеримая относительно σ -алгебры \mathcal{A}
- Для любых $B \in \mathcal{A}$ $\int_B E\{X|\mathcal{A}\} dP = \int_B X dP$

Условное математическое ожидание существует и единственно с точностью до множества нулевой меры.

Свойства математического ожидания:

1. Линейность - $E\{\alpha X + \beta Y|\mathcal{A}\} = \alpha E\{X|\mathcal{A}\} + \beta E\{Y|\mathcal{A}\}$, где X, Y - случайные величины
2. Пусть X – не зависит от σ -алгебры \mathcal{A} , то $E\{X|\mathcal{A}\} = EX$
3. Пусть X, Y - случайные величины и X измерим относительно \mathcal{A} тогда $E\{XY|\mathcal{A}\} = X E\{Y|\mathcal{A}\}$
4. Пусть есть две σ -алгебры $M \subseteq \mathcal{A} \subseteq \mathcal{F}$, M – более бедная $E\{X|M\} = E\{E\{X|\mathcal{A}\}|M\}$

Определение 1.10: Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, \mathcal{F}_t - возрастающий поток σ -алгебр (фильтрация), $\forall s < t, s, t \in T; \mathcal{F}_s \subseteq \mathcal{F}_t$

$X(t), t \in T$ называется *мартингалом*, если

1. $\forall t, E|X(t)| < \infty$,
2. $X(t)$ согласован с \mathcal{F}_t
3. $\forall s < t, s, t \in T; E(X(t)|\mathcal{F}_s) = X(s)$ с точностью до множеств нулевой меры.

Определение 1.11: Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, \mathcal{F}_t - возрастающий поток σ -алгебр (фильтрация), $\forall s < t; s, t \in T; \mathcal{F}_s \subseteq \mathcal{F}_t$. Случайный процесс $X(t), t \in T$ называется *субмартингалом*, если

1. $\forall t, E|X(t)| < \infty$
2. $X(t)$ согласован с \mathcal{F}_t
3. $\forall s < t; s, t \in T; E(X(t) | \mathcal{F}_s) \geq X(s)$ с точностью до множеств нулевой меры.

Определение 1.12: Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, \mathcal{F}_t - возрастающий поток σ -алгебр (фильтрация), $\forall s < t; s, t \in T; \mathcal{F}_s \subseteq \mathcal{F}_t$. Случайный процесс $X(t), t \in T$ называется *супермартингалом*, если

1. $\forall t, E|X(t)| < \infty$
2. $X(t)$ согласован с \mathcal{F}_t
3. $\forall s < t; s, t \in T; E(X(t) | \mathcal{F}_s) \leq X(s)$ с точностью до множеств нулевой меры.

Определение 1.13: Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, \mathcal{F}_t - возрастающий поток σ -алгебр, $t \in T$. Случайная величина τ – называется *моментом остановки* относительно \mathcal{F}_t , если $\forall t, \text{событие} \{\tau \leq t\} \in \mathcal{F}_t$

Свойства моментов остановки:

1. для любого $t_0 = \text{const}$ – есть момент остановки, так как $\{t_0 \leq t\} = \begin{cases} \Omega \\ \emptyset \end{cases}$
2. τ – момент остановки, при постоянной $t_0 > 0$, тогда $t+t_0$ момент остановки. $\{\tau+t_0 \leq t\} = \{\tau \leq t - t_0\} \in \mathcal{F}_{t-t_0} \in \mathcal{F}_t$
3. τ, σ – моменты остановки, тогда $\max\{\tau, \sigma\} \equiv \tau \vee \sigma$ - момент остановки
4. τ, σ – моменты остановки, тогда $\min\{\tau, \sigma\} \equiv \tau \wedge \sigma$ - момент остановки
5. $\tau \sim \mathcal{F}_t$ $\{\tau \leq s\} \in \mathcal{F}_t$, то есть для любого t событие $\{\tau \leq s\} \cap \{\tau \leq t\} \in \mathcal{F}_t$
6. $X(k), k=1, 2, \dots$ - последовательность случайных величин (случайный процесс с дискретным временем) $\mathcal{F}_n = \sigma(X(k), 1 \leq k \leq n)$ - поток σ – алгебр, порожденных последовательностью $X(k)$, τ - момент остановки относительно потока \mathcal{F}_n . Тогда последовательность $X(\tau)$ - \mathcal{F}_t измерима.

Определение 1.14: Пусть (Ω, \mathcal{F}, P) , $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$ – стохастический базис. *Предсказуемой*, называется σ -алгебра, содержащая все множества вида $(\Gamma \cap (t, \infty)) \times A, A \in \mathcal{F}_t, t \geq 0$.

Система подмножеств вида

$$(\Gamma \cap (a, b)) \times A, A \in \mathcal{F}_a, a, b \geq 0$$

является порождающей для предсказуемой σ -алгебры.

Определение 1.15: Процесс $\{X_t\}_{t \geq 0}$ называется *предсказуемым* в стохастическом базисе $(\Omega, \mathcal{F}, \mathbb{P})$, если он измерим относительно предсказуемой σ -алгебры.

Пусть $T \geq 0$ – случайная величина (время безотказной работы)

$F(t) = P(T \leq t)$ – функция распределения случайной величины T

$S(t) = 1 - F(t)$ – функция отказа случайной величины T

Предположим, что T абсолютно непрерывна с плотностью f

Определение 1.16:

1) *Интенсивностью отказа* случайной величины T будем называть следующую функцию:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = -\frac{dS(t)}{dt} / S(t) = -\frac{d \ln S(t)}{dt} = \frac{f(t)}{S(t)}, t \in R$$

2) *Накопленной интенсивностью* будем называть функцию:

$$\Lambda(t) = \int_0^t \lambda(u) du; t \in R$$

Замечание:

1) В рассматриваемом случае абсолютно непрерывной S справедливо равенство

$$S(t) = \exp\{-\Lambda(t)\}, t \in R$$

Определение 1.17 (Класс D)

Процесс X – процесс класса (D) (класса Дирихле), если множество случайных величин $\{X_T: T \text{ — конечный момент остановки}\}$ равномерно интегрируемо.

Определение 1.18:

Пусть τ — случайный момент, для которого существует неубывающая последовательность $(\tau_n)_{n \geq 1}$ моментов остановки таких, что:

a) $\tau(\omega) = \lim_{n \rightarrow \infty} \tau_n(\omega), \omega \in \Omega;$

b) $\tau_n(\omega) < \tau(\omega), \omega \in \{\tau(\omega) > 0\}.$

Последовательность $(\tau_n)_{n \geq 1}$ будем называть *предвещающей последовательностью* для τ .

Теорема (неравенство Ленгляра-Реболledo) 1.1:

Пусть X, Y — неотрицательные процессы из класса $D \cap F$ при этом $X_0 = Y_0 = 0, Y$ – возрастающий процесс, $F = (F_t)_{t \geq 0}$. Пусть Y доминирует процесс X , в том смысле, что для любого момента остановки τ

$$EX_\tau \leq EY_\tau \tag{1.1}$$

Тогда для любого момента остановки T и чисел $a > 0, b > 0$

$$P(\sup_{t \leq T} X_t \geq a) \leq \frac{1}{a} E \left[Y_T \wedge (b + \sup_{t \leq T} \Delta Y_t) \right] + P(Y_T \geq b) \quad (1.2)$$

Если к тому же процесс Y предсказуем, то

$$P(\sup_{t \leq T} X_t \geq a) \leq \frac{1}{a} E[Y_T \wedge b] + P(Y_T \geq b) \quad (1.3)$$

Доказательство. Будем сначала предполагать, что $T < \infty$.

Обозначим

$$\alpha = \inf(t : X_t > a), \beta = \inf(t : T_t \geq b)$$

и заметим, что

$$P(\sup_{t \leq T} X_t > a) = P(\sup_{t \leq T \wedge \beta} X_t > a, \beta > T) + P(\sup_{t \leq T} X_t > a, \beta \leq T) \leq P(\sup_{t \leq T \wedge \beta} X_t > a) + P(\beta \leq T) \quad (1.4)$$

Оценим члены, входящие в правую часть (1.4). Имеем

$$P(\beta \leq T) = P(Y_T \geq b) \quad (1.5)$$

Далее,

$$\left\{ \sup_{t \leq T \wedge \beta} X_t > a \right\} \subseteq \left\{ X_{\alpha \wedge T \wedge \beta} \geq a \right\}$$

Отсюда в силу неравенства Чебышева и (1.11) получаем

$$P(\sup_{t \leq T \wedge \beta} X_t \geq a) \leq P(X_{\alpha \wedge T \wedge \beta} \geq a) \leq \frac{1}{a} EX_{\alpha \wedge T \wedge \beta} \leq \frac{1}{a} EY_{\alpha \wedge T \wedge \beta} \quad (1.6)$$

Заметим, что

$$\begin{aligned} Y_{\alpha \wedge T \wedge \beta} &\leq T_{T \wedge \beta} = Y_T \wedge Y_{T \wedge \beta} = Y_T \wedge T_{T \wedge \beta} = Y_T \wedge [T_{(T \wedge \beta)^-} + \Delta Y_{T \wedge \beta}] \leq \\ &\leq Y_T \wedge \left[Y_{\beta^-} + \sup_{t \leq T} \Delta T_t \right] \leq T_T \wedge \left[b + \sup_{t \leq T} \Delta T_t \right] \end{aligned}$$

Отсюда и из (1.6)

$$P(\sup_{t \leq T \wedge \beta} X_t > a) \leq \frac{1}{a} E \left[Y_T \wedge (b + \sup_{t \leq T} \Delta Y_t) \right]$$

Пусть $a_k, k \geq 1$ такая числовая последовательность, что $0 < a_k < a, k \geq 1$ и $a_k \uparrow a$. Тогда, согласно доказанному

$$P(\sup_{t \leq T} X_t \geq a_k) \leq \frac{1}{a_k} E \left[Y_T \wedge (b + \sup_{t \leq T} \Delta Y_t) \right] + P(Y_T \geq b), k \geq 1 \quad (1.7)$$

а предельный переход при $k \rightarrow \infty$ в этом неравенстве приводит к неравенству (1.2).

Для доказательства (1.3) заметим, что поскольку $Y \in V^+ \cap P$, то момент β является предсказуемым. Пусть $(\beta_n)_{n \geq 1}$ — предвещающая последовательность моментов останова для β . Тогда из (1.4)

$$P(\sup_{t \leq T} X_t > a) \leq P(\sup_{t \leq T \wedge \beta_n} X_t > a) + P(\beta_n \leq T) \quad (1.8)$$

Аналогом неравенства (1.6) будет неравенство

$$P(\sup_{t \leq T \wedge \beta_n} X_t > a) \leq \frac{1}{a} EY_{\alpha \wedge T \wedge \beta_n}$$

Но $Y_{\alpha \wedge T \wedge \beta_n} \leq Y_{T \wedge \beta_n} = Y_T \wedge Y_{\beta_n} \leq Y_T \wedge b$, и значит

$$P(\sup_{t \leq T \wedge \beta_n} X_t > a) \leq \frac{1}{a} E[Y_T \wedge b] \quad (1.9)$$

Оценим $P(\beta_n \leq T)$. Имеем

$$\begin{aligned} P(\beta_n \leq T) &= P(\beta_n \leq T, \beta - \beta_n > \varepsilon) + P(\beta_n \leq T, \beta - \beta_n \leq \varepsilon) \leq \\ &\leq P(\beta - \beta_n > \varepsilon) + P(\beta \leq T + \varepsilon) = P(\beta - \beta_n > \varepsilon) + P(Y_{T+\varepsilon} \geq b) \end{aligned}$$

Пусть $0 < a_k < a$, $k \geq 1$ и $a_k \uparrow a$ и $b_{ij} \geq 1$ – числовая последовательность со свойствами: $0 < b_j < b$, $j \geq 1$, $b_j \uparrow b$, где b_j – точки непрерывности функции распределения случайной величины Y_t

Тогда

$$P(\beta_n \leq T) \leq P(\beta - \beta_n > \varepsilon) + P(Y_{T+\varepsilon} > b_j), j \geq 1 \quad (1.10)$$

Из (1.8)-(1.10) находим

$$P(\sup_{t \leq T} X_t > a_k) \leq \frac{1}{a_k} E[Y_T \wedge b] + P(\beta - \beta_n > \varepsilon) + P(Y_{T+\varepsilon} > b_j)$$

Предельный переход $\lim_k \lim_{\varepsilon \rightarrow 0} \lim_n$ в этом неравенстве дает требуемое неравенство (1.3)

Установим теперь справедливость неравенства (1.2) и (1.3) для произвольного момента остановки T (не обязательно конечного).

Пусть L_T обозначает правую часть любого из неравенств (1.2) или (1.3). Положим $T_n = T \wedge n$. Тогда

$$P(\sup_{t \leq T_n} X_t \geq a) \leq L_{T_n} \leq L_T \quad (1.11)$$

Заметим, что $\sup_{t \leq T_n} X_t \uparrow \sup_{t \leq T} X_t$. Выберем последовательность a_k , $k \geq 1$, такую, что $a_k \uparrow a$ и a_k, a – точки непрерывности функции распределения случайной величины $\sup_{t \leq T} X_t$. Очевидно, что

$$P(\sup_{t \leq T_n} X_t > a_k) \leq P(\sup_{t \leq T_n} X_t \geq a) \leq L_T$$

Отсюда

$$P(\sup_{t \leq T} X_t > a_k) = \lim_n P(\sup_{t \leq T_n} X_t > a_k) \leq L_T$$

и

$$P(\sup_{t \leq T} X_t \geq a) = \lim_k P(\sup_{t \leq T} X_t > a_k) \leq L_T$$

что и доказывает неравенства (1.2) и (1.3) для произвольного момента остановки $T \leq \infty$.

Определение 1.19: Случайный процесс \tilde{X}_t называется *компенсатором* процесса X_t , если

1. \tilde{X}_t -предсказуем
2. \tilde{X}_t -имеет ограниченную вариацию на любом конечном отрезке, содержащемся в Γ .
3. Процесс $X_t - \tilde{X}_t$ -мартингал

Теорема 1.3 (разложение Дуба-Мейера): Пусть $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$, – стохастический базис, $\{X_t\}_{t \geq 0}$ - непрерывный справа неотрицательный субмартингал класса D . Тогда существует единственный с точностью до неразличимости непрерывный справа равномерно интегрируемый мартингал $\{M_t\}_{t \geq 0}$ и непрерывный справа возрастающий предсказуемый процесс $\{A_t\}_{t \geq 0}$ такой, что $E|A_t| < \infty$ и

$$X_t = A_t + M_t, \quad \text{п.н.}$$

Теорема 1.4 (существования): Пусть $\{M_t\}_{t \geq 0}$ и $\{M'_t\}_{t \geq 0}$ - непрерывные справа мартингалы в стохастическом базисе $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$, такие, что $E(M_t)^2 < \infty$ и $E(M'_t)^2 < \infty, t \geq 0$. Тогда существует непрерывный справа предсказуемый процесс $\langle M, M' \rangle$ такой, что

1. $\langle M, M' \rangle(0) = 0, E\langle M, M' \rangle(t) < \infty$
2. $\langle M, M' \rangle$ - разность двух непрерывных справа возрастающих процессов ($\langle M, M' \rangle$ - процесс ограниченной вариации)
3. Процесс $M_t M'_t - \langle M, M' \rangle(t)$ – мартингал

Лемма 1: Пусть $a, b \in \mathbb{R}$. Тогда, как в условиях предыдущей теоремы.

- a) $E(a M_t + b M'_t | \mathcal{F}_s) = a M_s + b M'_s, s \leq t$
- b) $E(a M_t + b M'_t)^2 < \infty$
- c) Процесс $E(a M_t + b M'_t)^2$ – субмартингал.

Доказательство (Леммы): а) следует из свойства линейности условного математического ожидания. Для доказательства б) оценим

$$\begin{aligned} E(a M_t + b M'_t)^2 &= a^2 E(M_t)^2 + b^2 E(M'_t)^2 + 2ab E(M_t M'_t) \leq \\ &\leq a^2 E(M_t)^2 + b^2 E(M'_t)^2 + ab(E(M_t)^2 + E(M'_t)^2) < \infty. \end{aligned}$$

Используя неравенство Йенсена и мартингальное свойство получаем неравенство

$E((a M_t + b M'_t)^2 | \mathcal{F}_s) \geq (E(a M_t + b M'_t | \mathcal{F}_s))^2 = (a M_s + b M'_s)^2, s \leq t$
 которое доказывает с). Лемма доказана.

Доказательство (теоремы существования): Согласно Лемме условия теоремы (Дуба-Мейера) выполнены для субмартигалов $(M_t + M'_t)^2$ и $(M_t - M'_t)^2$ Тогда существуют разложения

$$(M_t + M'_t)^2 = M_t^+ + \langle M + M', M + M' \rangle (t)$$

и

$$(M_t - M'_t)^2 = M_t^- + \langle M - M', M - M' \rangle (t)$$

где M_t^+ и M_t^- - соответствующие мартигалы . Пусть

$$\langle M, M' \rangle (t) = \frac{1}{4} (\langle M + M', M + M' \rangle (t) - \langle M - M', M - M' \rangle (t))$$

Тогда $M_t M'_t - \langle M, M' \rangle (t)$ - мартигал. Остается отметить, что

$\frac{1}{4} \langle M + M', M + M' \rangle (t)$ и $\frac{1}{4} \langle M - M', M - M' \rangle (t)$ - возрастающие процессы.

Теорема доказана.

Определение 1.20. Пусть $(\Omega, \mathcal{F}, \mathbb{F}, P)$, $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$, - стохастический базис. Непрерывный справа случайный процесс $\{N_t\}_{t \geq 0}$ будем называть *точечным*, если

1. $\{N_t\}_{t \geq 0}$ согласован с \mathbb{F}
2. Траектории $N(\cdot, \omega)$ - кусочно-постоянные функции,
3. Скачки $N(\cdot, \omega)$ имеют величину 1 почти наверное.

1.3 Представление данных с помощью точечных процессов

Пусть T, U - неотрицательные случайные величины. Введем величины $X = \min(T, U) = T \wedge U$ и $\delta = I_{\{T \leq U\}}$, а также случайные процессы $\{N(t)\}_{t \geq 0}$ и $\{N^U(t)\}_{t \geq 0}$, определяемые равенствами

$$N(t) = I_{\{X \leq t, \delta = 1\}} = \delta \cdot I_{\{X \leq t\}}, N^U(t) = I_{\{X \leq t, \delta = 0\}} = (1 - \delta) I_{\{X \leq t\}}, t \in R_+$$

Предположим, что $F_t = \sigma\{(N(u), N^U(u)), 0 \leq u \leq t\}, t \geq 0$.

Теорема 1.5 (о независимом цензурировании). Пусть случайная величина T имеет абсолютно непрерывное распределение. Тогда случайный процесс

$$M(t) = N(t) - \int_0^t I_{\{X \geq u\}} \lambda(u) du - F\text{-мартигал если, и только если}$$

$$\lambda(t) = \frac{-\frac{d}{du} P(T \geq u, U \geq t)_{u=t}}{P(T \geq t, U \geq t)} \quad (*)$$

для любого $t \geq 0: P(X > t) > 0$

Комментарий.

1. Если T и U независимы, то условие (*) очевидно выполнено.
2. Правая часть (*) может быть переписана в виде

$$\lim_{h \rightarrow 0} \frac{P(t \leq T \leq t+h | T \geq t, U \geq t)}{h}$$

Теорема 1.6 (о независимом цензурировании): Следующие утверждения равносильны:

1. Случайный процесс

$$M(t) = N(t) - \int_0^t I_{\{X \geq u\}} d\Lambda(u)$$

2. Для любого $t \geq 0: P(X > t) > 0$ выполнено равенство

$$\frac{dF(t)}{1 - F(t-)} = \frac{-dP(T \geq t, U \geq T)}{P(T \geq t, U \geq T)}$$

3. Для любого $t \geq 0: P(X > t) > 0$ выполнено равенство

$$P(t \leq T \leq t+dt | T \geq t) = P(t \leq T \leq t+dt | T \geq t, U \geq t)$$

Определение 1.21: Будем говорить, что пара (T, U) удовлетворяет условию независимости цензурирования, если выполнено каждое из условий 1) - 3)

Пусть $(\Omega, \mathcal{F}, \mathbb{F}, P)$, $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$, – стохастический базис. Рассмотрим стохастический интеграл вида $L(t) = \int_0^t H(u) dM(u)$, где $N(t) = A(t) + M(t)$ – разложение Дуба-Мейера процесса $N(t)$, $H(t)$ - предсказуемый процесс

Теорема 1.7 (О мартингальном свойстве стохастического интеграла): Пусть $\{N(t)\}_{t \geq 0}$ - точечный процесс, $EN(t) < \infty, t \geq 0; N(t) = A(t) + M(t)$ - разложение Дуба-Мейера в стохастическом базисе $(\Omega, \mathcal{F}, \mathbb{F}, P)$, $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$; $\{N(t)\}_{t \geq 0}$ - ограниченный на любом ограниченном множестве процесс

Тогда $L(t) = \int_0^t H(u) dM(u)$ - мартингал.

1.4 Описание цензурированных данных с помощью точечных процессов

Пусть $(T_{ij}, U_{ij}), j = 1, 2, \dots, n_{ij}, i = 1, 2, \dots, m$ - набор независимых серий независимых в каждой серии неотрицательных и конечных с вероятностью 1 моментов отказов и цензурирования соответственно. Считаем, что $F_i(t) = P(T_{ij} < t) = P(T_{i1} < t)$ — функция распределения моментов отказов в i -й серии, $S_i(t) = 1 - F_i(t)$ — соответствующая функция отказа, а функция отказа величины U_{ij} имеет вид $C_{i,j}(t) = P(U_{ij} > t), t \in \mathbb{R}$.

Данные, полученные в результате эксперимента могут быть записаны в виде набора пар $(X_{ij}, \delta_{ij}), j = 1, 2, \dots, n_{ij}, i = 1, 2, \dots, m$ где

$$X_{ij} = \min(T_{ij}, U_{ij}) = T_{ij} \wedge U_{ij}; \quad \delta_{ij} = I_{\{T_{ij} \leq U_{ij}\}}$$

Обозначим $\pi_{i,j} = P(X_{i,j} \geq t)$. Введем случайные процессы

$$N_{i,j}(t) = \delta_{i,j} I_{\{X_{i,j} \leq t\}}; \quad N_{i,j}^U(t) = (1 - \delta_{i,j}) I_{\{X_{i,j} \leq t\}}; \quad Y_{i,j}(t) = I_{\{X_{i,j} \leq t\}}, \quad t \geq 0$$

а также

$$N_{i,\bullet}(t) = \sum_{j=1}^{n_i} N_{i,j}(t); \quad N_{i,\bullet}^U(t) = \sum_{j=1}^{n_i} N_{i,j}^U(t); \quad Y_{i,\bullet}(t) = \sum_{j=1}^{n_i} Y_{i,j}(t); \quad Y_{i,\bullet}^U(t) = \sum_{j=1}^{n_i} Y_{i,j}^U(t); \quad , t \geq 0$$

(точка обозначает суммирование по соответствующему индексу)

Определение 1.22. Будем говорить, что цензурирование в i -й серии независимо, если компенсатор $N_{i,\bullet}(t)$ относительно фильтрации $\{\mathcal{F}_t\}_{t \geq 0}$ имеет вид

$$\Lambda_{i,\bullet}^*(t) = \int_0^t Y_{i,\bullet}(u) \Lambda_i(du)$$

где $\Lambda_i(dt) = -\frac{S_i(dt)}{S_i(t_-)}$

Покажем, что если каждая пара (T_{ij}, U_{ij}) удовлетворяет условию независимости цензурирования, т.е.

$$\frac{dF_i(u)}{1 - F_i(u_-)} = \frac{-dP(T_{i,j} \geq u, U_{i,j} \geq T_{i,j})}{P(T_{i,j} \geq u, U_{i,j} \geq T_{i,j})}$$

или

$$P(u \leq T_{i,j} < u + du | T \leq u) = P(u \leq T_{i,j} < u + du | T \geq u, U \geq u)$$

$i=1, \dots, k, j=1, \dots, n_i$, то условие независимости цензурирования выполнено.

Рассмотрим фильтрацию $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$, где

$$\mathcal{F}_t = \sigma(N_{i,j}(u), N_{i,j}^U(u), 0 \leq u \leq t, j=1, \dots, n_i, i=1, \dots, m)$$

Как было отмечено ранее, при выполнении условия независимости цензурирования, случайный процесс

$$M_{i,j}(t) = N_{i,j}(t) - \int_0^t Y_{i,j}(u) d\Lambda_i(u)$$

где $\Lambda_i(t) = \int_0^t (S_i(u_-))^{-1} dF_i(u)$, является мартингалом относительно фильтрации $F^{ij} = \{F_t^{ij}, t \geq 0\}$, $F_t^{ij} = \sigma(N_{i,j}(u), N_{i,j}^u(u), 0 \leq u \leq t)$. Принимая во внимание независимость пар (T_{ij}, U_{ij}) заключаем, что $M_{i,j}(t)$ – F - мартингал при всех $j=1, \dots, n_i, i=1, \dots, m$. Следовательно случайный процесс

$$M_i(t) = \sum_{j=1}^{n_i} M_{i,j}(t) = N_{i,\bullet}(t) - \int_0^t Y_{i,\bullet}(u) d\Lambda_i(u)$$

является F - мартингалом при любом $i \in \{1, \dots, m\}$

Рассмотрим некоторые примеры цензурирования справа.

Цензурирование 1-го типа и его обобщения.

Пусть $U_{ij} = u_0 = const, j = 1, \dots, n_i, i=1, \dots, m$, т.е. каждая тестируемая система наблюдается строго определенное время. Такой тип цензурирования носит название цензурирования 1-го типа. Обобщением цензурирования данного типа является цензурирование с независимыми временами отказов и цензурирований, т.е. T_{ij} и U_{ij} - являются независимыми при любых i, j . Например, если наблюдения начинаются в случайные моменты времени, а завершаются в строго фиксированный момент и продолжительность интервала времени с момента начала наблюдений до момента отказа не зависит от времени начала наблюдений, то T_{ij} и U_{ij} являются независимыми при любых i, j . Очевидно, что условие независимости цензурирования в рассматриваемой модели выполнено.

2. Оценки Нельсона-Аалена и Каплана-Мейера, представление с помощью случайного процесса.

2.1 Введение

Пусть

T_1, T_2, \dots, T_n – выборка из распределения с функцией отказа S ;

$L \leq n$ – число различных наблюдаемых отказов;

$T_1^* < T_2^* < \dots < T_L^*$ – последовательные времена отказов;

D_k – число отказов в момент времени $T_k^*, k = 1, 2, \dots, L$;

$0 = t_0 < t_1 < \dots < t_m = t$ – разбиение интервала $[0, t]$;

d_l – число наблюдавшихся отказов интервале $[t_{l-1}, t_l)$;

y_l – число элементов, находящихся под наблюдением, т.е. не отказавших и не цензурированных, к моменту времени t_{l-1}

При $\Delta t \rightarrow 0$

$$\Lambda(t + \Delta t) - \Lambda(t) \approx \lambda(t) \Delta t$$

$$\Lambda(t + \Delta t) - \Lambda(t) \approx P\{t \leq T < t + \Delta t | T \geq t\}$$

Таким образом, при $y_l > 0$ естественной оценкой для $\Lambda(t_l) - \Lambda(t_{l-1})$ выглядит

отношение $\frac{d_l}{y_l}$, т.е. $\hat{\Lambda} = \sum_{l: t_l \leq t} \frac{d_l}{y_l}$

При $m \rightarrow \infty$ и $\max_{1 \leq l \leq m} |t_l - t_{l-1}| \rightarrow 0, d_l = 0$, в пределе

$$\hat{\Lambda}(t) = \sum_{k: T_k^* \leq t} \frac{D_1}{Y_l^*} \quad (\text{оценка Нельсона-Аалена})$$

где Y_l^* – число элементов не отказавших и не цензурированных на момент времени $T_k^*, k=1, \dots, L$, $\hat{\Lambda}$ – оценка Нельсона-Аалена накопленной интенсивности Λ .

Рассмотрим оценку функции отказа S , при $\frac{D_1}{Y_l^*} \rightarrow 0$

$$\hat{S}(t) = \prod_{k: T_k^* \leq t} \left(1 - \frac{D_1}{Y_l^*} \right) \quad (\text{оценка Каплана-Мейера функции отказа})$$

Пусть T_j определена при $\forall T_j \leq U_j, X_j = \min(T_j, U_j), j=1, \dots, n$,

$X_1^* < \dots < X_m^*, k=1, \dots, L$ – последовательные моменты отказов и цензурирований;

$$D_i^* = \sum_{j=1}^n \mathbf{I}_{\{X_j = T_i\}} \quad (\text{число наблюдений})$$

$$D_i^f = \sum_{j=1}^n \mathbf{I}_{\{X_j \leq U_j, X_j = T_i\}} \quad (\text{число отказов})$$

$$D_i^c = D_i^* - D_i^f \quad (\text{число цензурирований})$$

В момент времени $X_i, i=1, \dots, m$, получаем

$$\hat{S}(X_k^*) = \prod_{i=1}^k \left(1 - \frac{D_i^f}{Y_i^*} \right)$$

Соответствующее дискретное распределение имеет атомы в точках $T_{(k)}^*$ величины

$$\delta(T_{(k)}^*) = \hat{S}(T_{(k)}^*) - \hat{S}(T_{(k-1)}^*) = \prod_{i=1}^{k-1} \left(1 - \frac{D_i^f}{Y_i^*} \right) \frac{D_k^f}{Y_k^*} = S(T_{(k-1)}^*) \frac{D_k^f}{Y_k^*}$$

С учетом соотношений

$$1 - \frac{D_i^f}{Y_i^*} = \frac{Y_i^* - D_i^f}{Y_i^*} = \frac{Y_i^* - D_i^*}{Y_i^*} \frac{Y_i - D_i^f}{Y_i^* - D_i^*} = \frac{Y_{i+1}^*}{Y_i^*} \left(1 + \frac{D_i^* - D_i^f}{Y_{i+1}^*} \right)$$

получаем следующее выражение

$$\hat{S}(X_{(k)}^*) = \frac{Y_{k+1}^*}{n} \prod_{i=1}^k \left(1 + \frac{D_i^c}{Y_{i+1}^*} \right)$$

$$\delta(X_{(k)}^*) = \frac{D_k^f}{n} \prod_{i=1}^{k-1} \left(1 + \frac{D_i^c}{Y_{i+1}^*} \right)$$

2.2 Непараметрическое оценивание функции отказов.

В данном разделе рассматривается случай однородных отказов, т.е. $m = 1$.

Пусть

$$\Lambda_1 = \Lambda(t), t \geq 0, N_{1,j}(t) = N_j(t), N_{1,j}^U(t) = N_j^U(t), Y_{1,j}(t) = Y_j(t)$$

Пусть $(\Omega, \mathcal{F}, \mathcal{F}, P)$, $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$; - стохастический базис, где

$$\mathcal{F}_t = \sigma\{(N_j(u), N_j^U(u)), 0 \leq u \leq t\}, j = 1, \dots, n\}$$

Случайный процесс

$$M(t) = N_{\bullet}(t) - \int_0^t Y_{\bullet}(u) d\Lambda(u)$$

Где $N_{\bullet}(t) = \sum_{j=1}^n N_j(t)$; $Y_{\bullet}(t) = \sum_{j=1}^n Y_j(t)$ является \mathcal{F} мартингалом

Введем случайные процессы

$$\frac{I_{\{Y_{\bullet}(t) > 0\}}}{Y_{\bullet}(t)} = \begin{cases} \frac{1}{Y_{\bullet}(t)}, & Y_{\bullet}(t) > 0 \\ 0, & Y_{\bullet}(t) = 0 \end{cases}$$

и

$$\mathcal{M}(t) = \int_0^t \frac{I_{\{Y_{\bullet}(u) > 0\}}}{Y_{\bullet}(u)} dM(u) = \int_0^t \frac{dN_{\bullet}(u)}{Y_{\bullet}(u)} - \int_0^t I_{\{Y_{\bullet}(u) > 0\}} d\Lambda(u)$$

Поскольку $\frac{I_{\{Y_{\bullet}(u) > 0\}}}{Y_{\bullet}(u)}$ - предсказуем, то случайный процесс $\{\mathcal{M}(t)\}_{t \geq 0}$ - мартингал.

В частности отсюда следует, что

$$E \int_0^t \frac{dN_{\bullet}(u)}{Y_{\bullet}(u)} = E \int_0^t I_{\{Y_{\bullet}(u) > 0\}} d\Lambda(u)$$

Определение 2.1. Статистику

$$\hat{\Lambda}(t) = \int_0^t \frac{dN_{\bullet}(u)}{Y_{\bullet}(u)}$$

Будем называть *оценкой Нельсона-Аалена* накопленной интенсивности Λ . Рассмотрим, также, статистику

$$\Lambda^*(t) = \int_0^t I_{\{Y_{\bullet}(u) > 0\}} d\Lambda(u)$$

Следующая теорема дает некоторые свойства оценки Нельсона-Аалена.

Теорема 2.1 (об оценке Нельсона-Аалена): Пусть $\Lambda(t) < \infty, t \geq 0$ Тогда

$$1. E(\hat{\Lambda}(t) - \Lambda^*(t)) = 0, t \geq 0$$

$$2. \quad E(\hat{\Lambda}(t) - \Lambda(t)) = - \int_0^t \prod_{j=1}^n (1 - \pi_j(u)) d\Lambda(u), \text{ и если } \pi_j(u) = \pi(u), u \leq t, \text{ то}$$

$$E(\hat{\Lambda}(t) - \Lambda(t)) = - \int_0^t (1 - \pi(u))^n d\Lambda(u) \geq -(1 - \pi(t))^n \Lambda(t)$$

$$3. \quad \sigma_*^2(t) = E(\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)))^2 = E \int_0^t \frac{I_{\{Y(u) > 0\}}}{\bar{Y}(t)} (1 - \Delta\Lambda(u)) d\Lambda(u), \text{ где } \bar{Y} = \frac{Y_*(t)}{n}$$

Замечание.

1. При $\pi_j(s) = \pi(s), s \geq 0$ утверждение 2) теоремы доказывает асимптотическую несмещенность оценки Нельсона-Аалена $\hat{\Lambda}(t)$ накопленной интенсивности $\Lambda(t)$ при любом $t: \pi(t) > 0$.
2. Принимая во внимание усиленный закон больших чисел заключаем, что $\frac{n}{Y_*(t)} \xrightarrow{n \rightarrow \infty} \frac{1}{\pi(t)}$ почти наверное при $t: \pi(t) > 0$.

Тогда, в силу утверждения 3) теоремы

$$\sigma_*^2(t) = \int_0^t E \left(\frac{I_{\{Y(u) > 0\}}}{\bar{Y}(t)} (1 - \Delta\Lambda(u)) d\Lambda(u) \right) \xrightarrow{n \rightarrow \infty} \sigma^2(t)$$

3. Поскольку

$$\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^t \frac{n}{Y_*(u)} I_{\{Y(u) > 0\}} dM_j(u),$$

где правая часть данного равенства - нормированная сумма независимых одинаково распределенных случайных величин с нулевым средним и учитывая условие

$$\sqrt{n}(\Lambda^*(t) - \Lambda(t)) \xrightarrow{n \rightarrow \infty} 0 \quad \text{п.н.}$$

естественно предположить асимптотическую нормальность оценки Нельсона-Аалена $\hat{\Lambda}$. Исследуем дисперсию величины $\hat{\Lambda}$

$$D\hat{\Lambda}(t) = \frac{1}{n} \sigma_*^2(t) + 2E \left((\hat{\Lambda}(t) - \Lambda^*(t)) (\Lambda^*(t) - E\Lambda^*(t)) \right) - D\Lambda^*(t)$$

Отметим, что

$$nP(Y_*(s) = 0) = n \prod_{j=1}^n (1 - \pi_j(t)) \xrightarrow{n \rightarrow \infty} 0$$

Если $\pi_j(t) > \varepsilon > 0, j = 1, \dots$ Тогда

$$\begin{aligned}
D(\sqrt{n}\Lambda^*(t)) &= E\left(\int_0^t \sqrt{n}\left(\mathbf{I}_{\{Y(u)>0\}} - P(Y(u) > 0)\right)d\Lambda(u)\right)^2 = \\
&= \left(\int_0^t \sqrt{n}(1 - P(Y(u) > 0))d\Lambda(u)\right)^2 P(Y(u) > 0) + \\
&+ \left(\int_0^t P(Y(u) > 0)d\Lambda(u)\right)^2 nP(Y(u) = 0) \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

Следовательно,

$$D\hat{\Lambda}(t) \sim \frac{1}{n} \sigma_*^2(t)$$

Определим

$$\hat{\sigma}^2(t) = n \int_0^t \frac{\mathbf{I}_{\{Y(u)>0\}}}{Y^2(u)} \left(1 - \frac{\Delta N(u) - 1}{Y(u) - 1}\right) dN(u) = \int_0^t \frac{\mathbf{I}_{\{Y(u)>0\}}}{\bar{Y}^2(u)} \left(1 - \frac{\Delta \bar{N}(u) - 1}{Y(u) - 1}\right) d\bar{N}(u)$$

где $\bar{Y} = \frac{Y}{n}$, $\bar{N} = \frac{N}{n}$. В непрерывном случае данный объект выглядит следующим образом:

$$\hat{\sigma}^2(t) = n \int_0^t \frac{\mathbf{I}_{\{Y(u)>0\}}}{Y^2(u)} dN(u) = \int_0^t \frac{\mathbf{I}_{\{Y(u)>0\}}}{\bar{Y}^2(u)} d\bar{N}(u)$$

Теорема 2.2 (об оценке дисперсии): Пусть условие независимости цензурирования выполнено, $\{N(t)\}_{t \geq 0}$ точечный процесс с накопленной интенсивностью $\Lambda(t) < \infty, t \geq 0$. Тогда

$$\frac{E(\hat{\sigma}^2(t) - \sigma_*^2(t))}{n} = \int_0^t P(Y(u) = 1) \Delta \Lambda(u) d\Lambda(u)$$

Замечание. Если $\Lambda(t), t \geq 0$, непрерывна, то $E(\hat{\sigma}^2(t) - \sigma_*^2(t)) = 0, t \geq 0$, т.е. при некоторых дополнительных ограничениях, рассмотренных выше, $\hat{\sigma}^2(t)$ асимптотически несмещенная оценка $\sigma^2(t)$.

Доказательство. проведем лишь для случая с непрерывной Λ . Отметим, что в этом случае

$$\frac{E(\hat{\sigma}^2(t) - \sigma_*^2(t))}{n} = \int_0^t \frac{\mathbf{I}_{\{Y(u)=0\}}}{Y^2(u)} (Y(u)d\Lambda(u) - dN(u))$$

Следовательно $\hat{\sigma}^2(t) - \sigma_*^2(t)$ - мартингал. Доказательство завершено.

2.3 Оценка Каплана-Мейера

Принимая во внимание соотношение $S(t) = \exp(-\Lambda(t))$ введем оценку Каплана-Мейера по правилу $S(t) = \exp(-\Lambda(t)), t \geq 0$. Тогда $d\Lambda = -\frac{dS(t)}{S(t)}$. Пусть $d\hat{\Lambda} = -\frac{d\hat{S}(t)}{\hat{S}(t)}$.

Переносим $\hat{S}(t-)$ в правую часть. После интегрирования получаем

$$d\hat{S}(t) = 1 - \int_0^t \hat{S}(u_-) d\hat{\Lambda}(u)$$

Тогда

$$\Delta\hat{S}(t) = \hat{S}(t) - \hat{S}(t_-) = -\hat{S}(t_-) \frac{\Delta N_{\cdot}(t)}{Y_{\cdot}(t)}$$

Принимая во внимание ступенчатый характер функции $\hat{\Lambda}$ получаем следующее представление оценки Каплана-Мейера:

$$\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N_{\cdot}(s)}{Y_{\cdot}(s)} \right)$$

Теорема 2.3 (О мартингальном представлении оценки Каплана-Мейера)

Пусть $S(t) > 0$. Тогда

$$\frac{\hat{S}(t) - S(t)}{S(t)} = - \int_0^t \frac{\hat{S}(u_-)}{S(u)} \left(\frac{dN_{\cdot}(t)}{Y_{\cdot}(t)} - d\Lambda(u) \right)$$

Теорема 2.4 (о состоятельности) Пусть S - абсолютно непрерывна. Тогда для любого $t \in [0, \infty)$ при условии сходимости по вероятности

$$Y_{\cdot}(t) \xrightarrow[n \rightarrow \infty]{} 0$$

Имеет место сходимость по вероятности

$$\sup_{0 \leq s \leq t} |\hat{\Lambda}(s) - \Lambda(s)| \xrightarrow[n \rightarrow \infty]{} 0$$

и

$$\sup_{0 \leq s \leq t} |\hat{S}(s) - S(s)| = \sup_{0 \leq s \leq t} |\hat{F}(s) - F(s)| \xrightarrow[n \rightarrow \infty]{} 0$$

где $\hat{F}(t) = 1 - \hat{S}(t), t \geq 0$

Теорема 2.5 Пусть $\{X(t): t \geq 0\}$ непрерывный справа процесс согласованный с фильтрацией $\{F_t: t \geq 0\}$ – является мартингалом (субмартингалом, супермартингалом), и пусть τ согласованные с F_t момент остановки, тогда $\{X(t \wedge \tau), t \geq 0\}$ - тоже является мартингалом (субмартингалом, супермартингалом)

3. Состоятельность оценок Нельсона-Аалена и Каплана-Мейера.

Определение 3.1: Пусть f - функция, определенная на интервале $[a,b]$ где $-\infty \leq a < b \leq \infty$

$\sup_{a \leq s \leq b} |f(s)|$ - супремум нормы f

$\sup_{a \leq s \leq b} |f(s) - g(s)|$ - расстояние, индуцируемое этой нормой между двумя функциями f и g

Следствие 3.1 (из Теоремы 1.1 неравенства Ленгляра) Пусть N будет точечный процесс, а $M = N - A$ – соответствующий равномерно интегрируемый, квадратично-интегрируемый мартингал. Пусть H согласованный непрерывный слева процесс и ограниченный справа или, в более общем плане, предсказуемый и локально ограниченный процесс. Тогда для любого момента остановки T такого, что $P\{T < \infty\} = 1$, и любых $\varepsilon, \eta > 0$,

$$P\left\{\sup_{t \leq T} \left\{ \int_0^t H(s) dM(s) \right\}^2 \geq \varepsilon\right\} \leq \frac{\eta}{\varepsilon} + P\left\{\int_0^t H^2(s) d\langle M, M \rangle(s) \geq \eta\right\}$$

Доказательство. Пусть $\{\tau_k; k = 1, 2, \dots\}$ локализирующая последовательность такая, что для любого k , $N(\bullet \wedge \tau_k)$, $A(\bullet \wedge \tau_k)$ и $H(\bullet \wedge \tau_k)$ являются процессами, ограниченные по k (напомним, что A всегда локально ограничена), и $M(\bullet \wedge \tau_k)$ является квадратично интегрируемым мартингалом.

По теореме о моментах остановки (теорема 2.5)

$$E\{X_k(t \wedge T) - Y_k(t \wedge T)\} = 0, \forall t \geq 0 \quad (4.1)$$

где

$$X_k(t) \equiv \left\{ \int_0^{t \wedge \tau_k} H(s) dM(s) \right\}^2$$

и

$$Y_k(t) \equiv \int_0^{t \wedge \tau_k} H^2(s) d\langle M, M \rangle(s)$$

При $t \rightarrow \infty$, $X_k(t \wedge T) \rightarrow X_k(T)$, по теореме о мажорируемой сходимости $EX_k(t \wedge T) \rightarrow EX_k(T)$. Следовательно, по теореме монотонной сходимости, $EY_k(t \wedge T) \uparrow EY_k(T)$ и $EY_k(T) < \infty$. Таким образом, для любого k

$$EX_k(T) = EY_k(T)$$

По неравенству Ленгляра,

$$P_{1k} \equiv P \left\{ \sup_{t \leq T} \left\{ \int_0^{t \wedge \tau_k} H(s) dM(s) \right\}^2 \geq \varepsilon \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \int_0^{T \wedge \tau_k} H^2(s) d\langle M, M \rangle(s) \geq \eta \right\} \equiv \frac{\eta}{\varepsilon} + P_{2k}$$

по теореме о монотонной сходимости при $k \rightarrow \infty$

$$P_{2k} \uparrow P \left\{ \int_0^T H^2(s) d\langle M, M \rangle(s) \geq \eta \right\} \equiv P_2$$

Поэтому для любого k , $P_{1k} \leq \eta/\varepsilon + P_2$. По теореме о мажорируемой сходимости для $k \rightarrow \infty$

$$P_{1k} \rightarrow P \left\{ \sup_{t \leq T} \left\{ \int_0^t H(s) dM(s) \right\}^2 \geq \varepsilon \right\}$$

Следствие доказано. \square

Теорема 3.2:

1. Пусть T время отказа- случайная величина с непрерывной функцией распределения $F(s) = P\{T \leq s\}$ и накопленной интенсивностью

$$\Lambda(s) = \int_0^s \frac{dF(v)}{1 - F(v)}$$

$$\text{Если } t \in (0; \infty), \text{ то } \bar{Y}(t) \xrightarrow{P} \infty, n \rightarrow \infty \quad (4.2)$$

Тогда

$$\sup_{0 \leq s \leq t} \left| \int_0^s \frac{d\bar{N}(v)}{\bar{Y}(v)} - \Lambda(s) \right| \xrightarrow{P} 0, n \rightarrow \infty \quad (4.3)$$

и

$$\sup_{0 \leq s \leq t} \left| \hat{F}(s) - F(s) \right| \xrightarrow{P} 0, n \rightarrow \infty \quad (4.4)$$

где $1 - \hat{F}$ - оценка Каплана-Мейера.

2. Предположим, что распределение времени отказа может меняться с увеличением объема выборки n , т.е. $P\{T_j^n \leq s\} = F^n(s), j=1, \dots, n$.

а) Если условие (4.2) выполняется, и $\limsup_{n \rightarrow \infty} F^n(t) < 1$ (4.5)

следовательно при $n \rightarrow \infty$

$$\sup_{0 \leq s \leq t} \left| \int_0^s \frac{d\bar{N}(v)}{\bar{Y}(v)} - \Lambda^n(s) \right| \xrightarrow{P} 0 \quad (4.6)$$

и

$$\sup_{0 \leq s \leq t} \left| \hat{F}(s) - F^n(s) \right| \xrightarrow{P} 0 \quad (4.7)$$

б) Если $u \in (0; \infty)$ такое, что (4.2) и (4.5) справедливы для всех $t < u$ и F является непрерывной функцией распределения такой, что для всех $t \leq u$,

$$\lim_{n \rightarrow \infty} F^n(t) = F(t), \quad (4.8)$$

то (4.7) справедливо при замене t на u .

Доказательство. Достаточно доказать 2). Начнем с 2) а). Пусть $s \in [0, t]$. Тогда

$$\begin{aligned} \left| \int_0^s \frac{d\bar{N}(v)}{\bar{Y}(v)} - \Lambda^n(s) \right| &\leq \left| \int_0^s \frac{d\bar{N}(v)}{\bar{Y}(v)} - \int_0^s I_{\{\bar{Y}(v) > 0\}} d\Lambda^n(v) \right| + \\ &+ \left| \int_0^s I_{\{\bar{Y}(v) = 0\}} d\Lambda^n(v) \right| \leq \left| \int_0^s \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} dM(v) \right| + I_{\{\bar{Y}(t) = 0\}} \Lambda^n(t) \end{aligned}$$

При условии (4.2) и (4.5), $I_{\{\bar{Y}(t) = 0\}} \Lambda^n(t) \xrightarrow{P} 0$, для получения (4.6) достаточно, чтобы

$$\sup_{0 \leq s \leq t} \left\{ \int_0^s \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} dM(v) \right\}^2 \xrightarrow{P} 0$$

по следствию 3.1 в неравенстве Ленгляра имеем

$$P \left\{ \sup_{0 \leq s \leq t} \left\{ \int_0^s \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} dM(v) \right\}^2 \geq \varepsilon \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \int_0^s \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} d\Lambda^n(v) \geq \eta \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \frac{\Lambda^n(t)}{\bar{Y}(t)} > \eta \right\}$$

Из условия (4.2) и (4.5) следует, что второе слагаемое в правой части стремится к нулю при $n \rightarrow \infty$ и $\eta > 0$. Тогда условие сходимости оценки Нельсона-Аалена выполняется.

При любом фиксированном t таком, что (4.2) и (4.5) справедливы, имеем,

$$P \left\{ \frac{\hat{F}(x) - F^n(x)}{1 - F^n(x)} = Z(x), \text{ при } x \in [0, t] \right\} \xrightarrow{n \rightarrow \infty} 1$$

где

$$Z(x) = \int_0^t \frac{\{1 - \hat{F}(v-)\}}{\{1 - F^n(v)\}} \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} dM(v)$$

Условие (4.5) подразумевает

$$\liminf_{n \rightarrow \infty} \inf_{s \in [0, t]} \{1 - F^n(s)\} = \liminf_{n \rightarrow \infty} [1 - F^n(t)]$$

Следовательно, чтобы получить $\sup_{0 \leq s \leq t} |\hat{F}(s) - F^n(s)| \xrightarrow{P} 0$

достаточно показать, что $\sup_{0 \leq s \leq t} \{Z(s)\}^2 \xrightarrow{P} 0$

По следствию 3.1

$$P \left\{ \sup_{0 \leq s \leq t} \{Z(s)\}^2 \geq \varepsilon \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \int_0^t \frac{\{1 - \hat{F}(v-)\}^2}{\{1 - F^n(v)\}^2} \frac{I_{\{\bar{Y}(v) > 0\}}}{\bar{Y}(v)} d\Lambda^n(v) \geq \eta \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \frac{\Lambda^n(t)}{\{1 - F^n(t)\}^2 \bar{Y}(t)} \geq \eta \right\}$$

Из (4.2) и (4.5) следует, что второе слагаемое в правой части предыдущего уравнения, сходится к нулю при $n \rightarrow \infty$ и любом $\eta > 0$. Так как ε и η случайные величины, то (4.7) выполняется.

Пусть теперь u такая, что (4.2) и (4.5) справедливы для всех $t < u$, и, кроме того, в доп. слагаемом уравнение (4.8) выполняется. Для выполнения (2) (б) мы должны показать, что

$$\sup_{0 \leq s \leq t} |\hat{F}(s) - F^n(s)| \xrightarrow{P} 0, n \rightarrow \infty \quad (4.9)$$

или, что то же самое, для любого $\varepsilon > 0$

$$P \left\{ \sup_{0 \leq s \leq u} |\hat{F}(s) - F^n(s)| > \varepsilon \right\} < \varepsilon \quad (4.10)$$

при достаточно больших n

Для любой точки $t_0 < u$, часть 2) а) устанавливает однородность свойств на $[0, t_0]$, поэтому доказательство (2) (б) зависит от нахождения $t_0 < u$, для которых мы можем показать равномерную сходимость на $[t_0, u]$. Поскольку

$$|F^n(u) - F^n(t_0)| \leq |F^n(u) - F(u)| + |F(u) - F(t_0)| + |F(t_0) - F^n(t_0)|$$

из непрерывности F и уравнения (4.8) следует существование точки t_0 , $0 < t_0 < u$ и целого числа $n' \geq 1$ такого, что

$$|F^n(u) - F^n(t_0)| < \frac{\varepsilon}{4}$$

для любого $n \geq n'$

В формуле. (4.7), имеется второе число n'' такое, что для всех $n \geq n''$

$$P\left\{\left|\hat{F}^n(t_0) - F^n(t_0)\right| > \frac{\varepsilon}{4}\right\} < \frac{\varepsilon}{2} \quad (4.11)$$

Для $t_0 \leq s \leq u$, неравенство

$$\hat{F}(t_0) \leq \hat{F}(s) \leq \hat{F}(u)$$

и

$$F^n(t_0) \leq F^n(s) \leq F^n(u)$$

означает, что

$$\left|\hat{F}(s) - F^n(s)\right| \leq \left|\hat{F}(u) - F^n(t_0)\right| + \left|\hat{F}(t_0) - F^n(u)\right|$$

вследствие этого

$$\begin{aligned} \sup_{t_0 \leq s \leq u} \left|\hat{F}(s) - F^n(s)\right| &\leq \left|\hat{F}(u) - F^n(t_0)\right| + \left|\hat{F}(t_0) - F^n(u)\right| \leq \\ &\leq \left|\hat{F}(u) - F^n(u)\right| + 2\left|F^n(u) - F^n(t_0)\right| + \left|\hat{F}(t_0) - F^n(t_0)\right| \end{aligned}$$

Если $\left|\hat{F}(u) - F^n(u)\right| \rightarrow 0$ по вероятности, то будет существовать третье число n''' , такое что для всех $n \geq n'''$

$$P\left\{\left|\hat{F}^n(u) - F^n(u)\right| > \frac{\varepsilon}{4}\right\} < \frac{\varepsilon}{2}$$

тогда для $n \geq \max(n', n'', n''')$

$$P\left\{\sup_{t_0 \leq s \leq u} \left|\hat{F}(s) - F^n(s)\right| > \varepsilon\right\} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Для завершения доказательства мы должны показать, что $\left|\hat{F}(u) - F^n(u)\right| \xrightarrow{P} 0$. Рассмотрим отдельно два случая $F(u) = 1$ и $F(u) < 1$. Если $F(u) = 1$, тогда существуют t_0 и n' такие, что $\left|F^n(t_0 - 1)\right| < \frac{\varepsilon}{3}, \forall n \geq n'$ Далее,

$$\begin{aligned} |\hat{F}(u) - F^n(u)| &\leq |1 - \hat{F}(u)| + |1 - F^n(u)| \leq \\ &\leq |1 - \hat{F}(t_0)| + |1 - F^n(t_0)| \leq 2|1 - F^n(t_0)| + |F^n(t_0) - \hat{F}(t_0)| \end{aligned}$$

Так как $P\left\{|F^n(t_0) - \hat{F}(t_0)| > \frac{\varepsilon}{3}\right\} \xrightarrow{P} 0$ то существует $n^* > n'$, такое что $\forall n \geq n^*, P\left\{|F^n(t_0) - \hat{F}(t_0)| > \frac{\varepsilon}{3}\right\} < \varepsilon$ и $|\hat{F}(u) - F^n(u)| \xrightarrow{P} 0$

Допустим $F(u) = 1 - \delta^* < 1$

Из неравенства

$$\begin{aligned} |\hat{F}(u) - F^n(u)| &\leq |\hat{F}(u) - \hat{F}(t_0)| + |\hat{F}(t_0) - F^n(t_0)| + \\ &+ |F^n(t_0) - F^n(u)| \end{aligned} \quad (4.12)$$

и уравнения (4.7) и (4.8), и непрерывности F следует, что последними двумя правыми слагаемыми, можно пренебречь. Нам нужно только показать, что первое слагаемое стремится к нулю по вероятности, при $n \rightarrow \infty$. Допустим существует n' такая, что $\forall n \geq n'$

$$1 - F^n(t) > 1 - \delta^* / 2, \forall t \in [0, u]$$

Тогда $\forall n \geq n'$ и для $\forall t_0 \leq u$

$$\begin{aligned} \int_{t_0}^u d\Lambda^n(t) &\leq \int_{t_0}^u \frac{dF^n(t)}{1 - F^n(t)} < \frac{2}{\delta^*} \int_{t_0}^u dF^n(t) = \frac{2}{\delta^*} \{F^n(u) - F^n(t_0)\} = \\ &= \frac{2}{\delta^*} \{F^n(u) - F(u) + F(u) - F(t_0) + F(t_0) - F^n(t_0)\} \end{aligned}$$

Таким образом, существует t_0 и $n'' > n'$, такая что для $\forall n \geq n''$

$$\int_{t_0}^u d\Lambda^n(t) < \frac{\varepsilon^3}{108} \quad (4.13)$$

И так, для $\forall n \geq n''$

$$\begin{aligned} P\left\{|\hat{F}(u) - \hat{F}(t_0)| > \frac{\varepsilon}{3}\right\} &= P\left\{\int_{t_0}^u \mathbf{I}_{\{\bar{Y}(s) > 0\}} \left\{1 - \hat{F}(s-)\right\} \frac{d\bar{N}(s)}{\bar{Y}(s)} > \frac{\varepsilon}{3}\right\} \leq \\ &\leq P\left\{\int_{t_0}^u \mathbf{I}_{\{\bar{Y}(s) > 0\}} \left[\frac{d\bar{N}(s)}{\bar{Y}(s)} - d\Lambda^n(s)\right] > \frac{\varepsilon}{6}\right\} + P\left\{\int_{t_0}^u d\Lambda^n(s) > \frac{\varepsilon}{6}\right\} = \\ &= P\left\{\left[\int_{t_0}^u \frac{\mathbf{I}_{\{\bar{Y}(s) > 0\}}}{\bar{Y}(s)} dM(s)\right]^2 > \frac{\varepsilon^2}{36}\right\} \leq \eta \frac{36}{\varepsilon^2} + P\left\{\int_{t_0}^u \frac{\mathbf{I}_{\{\bar{Y}(s) > 0\}}}{\bar{Y}(s)} d\Lambda^n(s) \geq \eta\right\} = \frac{\varepsilon}{3} \end{aligned}$$

Где последнее неравенство следует из Следствия 3.1, а последнее уравнение-из подстановки $\eta = \frac{\varepsilon^3}{108}$

Мы можем предположить, что выбор t_0 в (4.12) удовлетворяет

$$|F^n(t_0) - F^n(u)| < \frac{\varepsilon}{3} \text{ для } \forall n \geq n'' . \text{ Поэтому}$$

$$|\hat{F}^n(t_0) - F^n(t_0)| \xrightarrow{P} 0, \Rightarrow |\hat{F}^n(u) - F^n(u)| \xrightarrow{P} 0, \text{ при } F(u) < 1$$

Если $F^n \equiv F$, тогда если $\bar{Y}(t) \xrightarrow{P} 0, n \rightarrow \infty$, для $\forall t \in [0, u]$, тогда оценка Каплана-Мейера $\hat{S}(t)$ - равномерно состоятельная оценка.

$$S(t) = \exp\left\{-\int_0^t \lambda(s) ds\right\} \text{ на } [0, u]$$

4. Мартингальная Центральная Предельная Теорема.

Теперь мы используем мартингальную структуру $U^{(n)} \equiv \sum_{i=1}^n \int H_i^{(n)} dM_i^{(n)}$, чтобы установить асимптотические результаты для последовательностей статистик. Пусть n -размер выборки.

Будем считать набором конечномерных распределений $(U_1^{(n)}, \dots, U_r^{(n)})$. Для любых n ,

$\{N_{i,l}^{(n)} : i=1, \dots, n, l=1, \dots, r\}$ - многомерный точечный процесс со стохастическим базисом $(\Omega, F, \{F_t : t \geq 0\}, P)$ (4.1)

с непрерывным компенсатором $A_{i,l}^{(n)}$ для $N_{i,l}^{(n)}$ (4.2)

$M_{i,l}^{(n)} \equiv N_{i,l}^{(n)} - A_{i,l}^{(n)}$ являются квадратично интегрируемым мартингалом, $\{M_{i,l}^{(n)}\}^2$; $A_{i,l}^{(n)}$ является квадратично интегрируемым мартингалом, и $M_{i,l}^{(n)} M_{i',l'}^{(n)}$ является квадратично интегрируемым мартингалом, если $i \neq i'$ или $l \neq l'$

Для любых i, l мы также предполагаем, что

$H_{i,l}^{(n)}$ является ограниченным F_t - предсказуемым процессом. (4.3)

$$U_{i,l}^{(n)}(t) = \int_0^t H_{i,l}^{(n)}(s) dM_{i,l}^{(n)}(s)$$

$$U_1^{(n)}(t) = \sum_{i=1}^n \int_0^t H_{i,l}^{(n)}(s) dM_{i,l}^{(n)}(s)$$

для любого $\varepsilon > 0$

$$U_{i,l,\varepsilon}^{(n)}(t) = \int_0^t H_{i,l}^{(n)}(s) I_{\{|H_{i,l}^{(n)}(s)| \geq \varepsilon\}} dM_{i,l}^{(n)}(s)$$

и

$$U_{1,\varepsilon}^{(n)}(t) = \sum_{i=1}^n U_{i,1,\varepsilon}^{(n)}(t)$$

$U_{i,1}^{(n)}, U_1^{(n)}, U_{i,1,\varepsilon}^{(n)}$ и $U_{1,\varepsilon}^{(n)}$ квадратично интегрируемые мартингалы. Процесс $U_{1,\varepsilon}^{(n)}$ содержит все скачки процесса $U_1^{(n)}$, размера не более чем ε .

$$\langle U_1^{(n)}, U_1^{(n)} \rangle(t) = \sum_{i=1}^n \int_0^t \{H_{i,1}^{(n)}(s)\}^2 dA_{i,1}^{(n)}(s)$$

и

$$\langle U_{1,\varepsilon}^{(n)}, U_{1,\varepsilon}^{(n)} \rangle(t) = \sum_{i=1}^n \int_0^t \{H_{i,1}^{(n)}(s)\}^2 I_{\{|H_{i,1}^{(n)}(s)| \geq \varepsilon\}} dA_{i,1}^{(n)}(s)$$

Основным результатом является асимптотическое распределение $\{U_1^{(n)} : 1 = 1, \dots, r\}$, при $n \rightarrow \infty$, и опирается на понятие слабой сходимости или сходимости по распределению, случайных процессов. Когда последовательность процессов $\{X_n\}$ слабо сходится к пределу процесса X , мы пишем $X_n \Rightarrow X$.

Сходимость по распределению для последовательности случайных величин, записывается как $X_n \xrightarrow{D} X$. Некоторые результаты используют сходимость в L^p норме или по вероятности. Мы пишем, что $X_n \xrightarrow{L^p} X$, когда последовательность случайных величин $\{X_n\}$ удовлетворяет

$$\lim_{n \rightarrow \infty} E(|X_n - X|^2) = 0$$

и $X_n \xrightarrow{P} X$, когда

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

для любого $\varepsilon > 0$

Переходя к пределу $U_1^{(n)}$ мы, чтобы упростить, сначала избавимся от аргумента 1, а затем опустим его в обозначениях. Изначально, мы будем считать, что для каждого n существует константа $C^{(n)} < \infty$ (которую можно увеличить с n до ∞) такую, что с почти наверное,

$$N_i^{(n)}, A_i^{(n)}, H_i^{(n)}, i = 1, \dots, n \text{ ограничены } C^{(n)} \quad (4.4)$$

Из этого следует, что, $U_i^{(n)}, U^{(n)}, U_{i,\varepsilon}^{(n)}, U_\varepsilon^{(n)}$ являются квадратично-интегрируемыми мартингалами. Понятно, что процессы $N_i^{(n)}(\cdot \wedge \tau^{(n)}), A_i^{(n)}(\cdot \wedge \tau^{(n)}), H_i^{(n)}(\cdot \wedge \tau^{(n)})$ удовлетворяет условию (4.4), и, что $U_i^{(n)}(\cdot \wedge \tau^{(n)}), U^{(n)}$ слабо сходятся к тому же пределу при $\tau^{(n)} \rightarrow \infty$.

Мы установим, в особых условиях, что $U^{(n)}$ слабо сходится к следующему процессу.

Определение 4.1. Стандартным Винеровским или процессом Броуновского движения называется процесс $W(t), t \geq 0$:

1. $W(0) = 0$ и $EW(t) = 0$ для любых t ;

2. $W(t)$ - процесс с независимыми приращениями;
3. $W(t) - W(s) \sim N(0, t-s), t < s$
4. $W(t)$ имеет непрерывные траектории.

Если f измеримая неотрицательная функция и $\alpha(t) = \int_0^t f^2(s) ds$, тогда $\int f dW$ - процесс, удовлетворяющий (1), (2) и (4), и

$$\text{var} \left\{ \int_0^t f(s) dW(s) \right\} = \alpha(t)$$

Процесс $U^{(n)}$ удовлетворяет условию (1) для всех n . По мартингалльному свойству, $U^{(n)}$ имеет некоррелированные приращения, так что (2) справедливо при больших n . По (теореме 2.4.4 Fleming),

$$\text{var} \{ U^{(n)}(t) \} = E \langle U^{(n)}, U^{(n)} \rangle(t)$$

если

$$\langle U^{(n)}, U^{(n)} \rangle(t) \xrightarrow{P} \alpha(t)$$

для некоторой подынтегральной функции f^2 , у $U^{(n)}$ существует дисперсия при достаточно больших n , тогда t заменяется на $\alpha(t)$. И, наконец, $U_\varepsilon^{(n)}$ содержит все скачки в $U^{(n)}$ размера ε или больше, $E U_\varepsilon^{(n)} = 0$ и

$$\text{var} \{ U_\varepsilon^{(n)}(t) \} = E \langle U_\varepsilon^{(n)}, U_\varepsilon^{(n)} \rangle(t)$$

если $\langle U_\varepsilon^{(n)}, U_\varepsilon^{(n)} \rangle(t) \xrightarrow{P} 0, n \rightarrow \infty$ для любых $\varepsilon > 0$, выборочные траектории должны стать гладкими при $n \rightarrow \infty$, и 4) также должно выполняться.

Мейер показал, что если U_1, U_2, \dots, U_r - являются квадратично интегрируемыми мартингалами, принимают значение 0 в нулевой момент времени, с непрерывными траекториями такими, что $\langle U_i, U_j \rangle(t) = \delta_{ij} \alpha_i(t)$ для некоторых возрастающих непрерывных справа функций $\alpha_i(t)$ и $\alpha_i(0) = 0$, тогда $\{ U_1, U_2, \dots, U_r \}$ являются независимыми гауссовскими процессами, равными нулю в нулевой момент времени, с независимыми приращениями и $\text{var} \{ U_i(t) \} = \alpha_i(t)$.

Ниже приводится формальная постановка результата (в частном случае при $r = 1$).

Теорема 4.1. Пусть f измеримая неотрицательная функция и $\alpha(t) = \int_0^t f^2(s) ds < \infty$ для всех $t > 0$. Для всех $t > 0$, и $n \rightarrow \infty$,

$$\langle U^{(n)}, U^{(n)} \rangle(t) \xrightarrow{P} \alpha(t) \tag{4.5}$$

и

$$\langle U_\varepsilon^{(n)}, U_\varepsilon^{(n)} \rangle(t) \xrightarrow{P} 0 \text{ для любых } \varepsilon > 0 \tag{4.6}$$

Тогда

$$U^{(n)} \Rightarrow U^\infty \equiv \int f dW \text{ на } D[0, \infty) \text{ и } n \rightarrow \infty,$$

где W - броуновское движение, и $D[0, \infty)$ - пространство функций на $[0, \infty)$, которые являются непрерывными справа, с конечными левосторонними пределами, наделенное топологией Скорохода.

5. Категориальный данные.

5.1 Классические категориальные и их применение в анализе данных типа времени жизни.

Категориальный план эксперимента основан на классификации объектов исследования по группам. Распределение по группам определяется по одной или нескольким характеристикам. В методе анализа категориальных данных используется асимптотическая нормальность выборочных оценок вероятностей.

Рассмотрим группу наблюдений с временем отказа T , как переменную отклика, и z в качестве ковариаты. Предположим, что заданы две категории времен отказа : $T \leq t_1$ и $T > t_1$. В случае цензурирования справа данные не могут быть правильно классифицированы, потому что мы не можем определенно сказать, сколько объектов наблюдения откажут к моменту времени t_1 . Различные виды стратегий классификации цензурированных к моменту t_1 в любом случае приводят к смещенности наблюдаемых частот как оценок искомых вероятностей. Рассмотрим три возможные стратегии:

- 1) все объекты наблюдений, цензурированные до момента остановки t_0 удаляются из анализа;
- 2) все цензурированные объекты наблюдений, включая объекты без симптомов заболевания в конечной точке, удаляются из анализа
- 3) все объекты наблюдений, классифицируются по времени события (отказа или цензурирования).

Обозначим ожидаемые вероятности $q_{i|z}^{(k)}, i=1,2, k=1,2,3$.

Наблюдаемые вероятности в этом случае определяются как:

$$q_{1|z}^{(1)} = P(T \leq t_0 | \{T \leq U\} \cup \{U > t_0\}; z) = \frac{P(T \leq U \wedge t_0 | z)}{1 - P(U < T \wedge t_0 | z)}$$

$$q_{1|z}^{(2)} = P(T \leq t_0 | T \leq U; z) = \frac{P(T \leq U \wedge t_0 | z)}{P(T \leq U | z)}$$

$$q_{1|z}^{(3)} = P(T \wedge U \leq t_0; z); \quad q_{2|z}^{(k)} = 1 - q_{1|z}^{(k)}$$

Предположим, что распределение времени цензурирования U не зависит от z и имеет функцию отказа $G(x) = P(U > x), G(0) = 0$. Тогда при независимости T и U ,

$$q_{1|z}^{(1)} = \frac{\int_0^t G(x) dF_z(x)}{S_z(t_0)G(t_0) + \int_0^t G(x) dF_z(x)}; \quad q_{1|z}^{(2)} = \frac{\int_0^{t_0} G(x) dF_z(x)}{\int_0^\infty G(x) dF_z(x)}; \quad q_{1|z}^{(3)} = 1 - \bar{G}(t_0)S_x(t_0)$$

Соответствующие смещения:

$$q_{1|z}^{(1)} - p_{1|z} = S_z(t_0)(G(t_0)(S_z(t_0) - 1)) + \frac{\int_0^t G(x) dF_z(x)}{S_z(t_0)G(t_0) + \int_0^{t_0} G(x) dF_z(x)} =$$

$$= \frac{S_z(t_0) \int_0^{t_0} (1 - S_z(x)) d\bar{G}(x)}{S_z(t_0)G(t_0) + \int_0^{t_0} G(x) dF_z(x)} = S_z(t_0) \frac{P(T \leq U < t_0)}{1 - P(U \leq T \wedge t_0)}$$

где $p_{1|z} = F_z(t_0)$ и $\bar{G}(x) = 1 - G(x)$;

$$q_{1|z}^{(2)} = \frac{\int_0^{t_0} (G(x) - C) dF_z(x)}{C}, \quad \text{где } C = \int_0^\infty G(x) dF_z(x)$$

и $q_{1|z}^{(3)} - p_{1|z} = S_z(t_0)G(t_0)$

Заметим, что классические категориальные тесты фактически применяют для $\tilde{H}_0 : q_{1|1} = \dots = q_{1|d}$

5.2 Категориальные тесты типа времени жизни.

Пусть $0 < t_1 < \dots < t_s < \gamma_T$, тогда $S_i(t_{k_1}) - S_i(t_{k_2}) > 0$ для всех $1 < k_2 < k_1 < \gamma_T$

$$H_0 : S_1(t_1) = \dots = S_d(t_1) \text{ для всех } l=1, \dots, s$$

Обозначим $\theta_{ij} = S_i(t_j)$. Введем параметры $\theta_i = (\theta_{i1}, \dots, \theta_{is})'$. Пусть $\psi_i = A\theta_i$, где $A = \|a_{ij}\|$ - $(d-1) \times d$ -матрица линейно независимых сравнений, то есть $\sum_{j=1}^d a_{ij} = 0$, для всех i и $\text{rk}(A) = d-1$. Далее H_0 может быть переписана в терминах сравнений:

$$H_0 : \psi_1 = \dots = \psi_{d-1} = 0$$

Пусть $\hat{\theta}_{ij} = \hat{S}_i(t_j)$ оценка для θ_{ij} , где \hat{S}_i - оценка Каплана-Мейера для $S_i, i=1, \dots, d$.

Используя сходимость:

$$\sqrt{n} \left(\frac{\hat{S}_i(t)}{S_i(t)} - 1 \right) \Rightarrow W_{\sigma_i^2(t)}$$

Где W_t - стандартный Винеровский процесс и $\sigma_i^2(t)$ - некоторая положительная неубывающая функция, тогда получаем сходимость

$$\sqrt{n} (\hat{\theta}_{i1} - \theta_{i1}, \dots, \hat{\theta}_{is} - \theta_{is})' \Rightarrow N(0, \Sigma_i)$$

где ковариационная матрица $\Sigma_i = \|\sigma_{i:qr}\|$ такая, что $\sigma_{i:qr} = \theta_{iq} \theta_{ir} \sigma_i^2(t_q \wedge t_r)$, $q, r=1, \dots, s$

Мы используем $\hat{\Sigma}_i, \hat{\sigma}_{i:qr} = \hat{\theta}_{iq} \hat{\theta}_{ir} \hat{\sigma}_i^2(t_q \wedge t_r)$, где $\hat{\sigma}_i^2(t)$ - состоятельная оценка $\sigma_i^2(t)$

Введем случайный вектор $\theta = (\theta_{11}, \dots, \theta_{1s}, \dots, \theta_{d1}, \dots, \theta_{ds})'$ и соответствующую ему оценку $\hat{\theta} = (\hat{\theta}_{11}, \dots, \hat{\theta}_{1s}, \dots, \hat{\theta}_{d1}, \dots, \hat{\theta}_{ds})'$, тогда

$$\sqrt{n} (\hat{\theta} - \theta) \Rightarrow N(0, \Sigma) \quad (5.1)$$

где $\Sigma = \text{diag}(1_1 \Sigma_1, \dots, 1_n \Sigma_d)$ - блочно-диагональная матрица, $1_i = \frac{n}{n_i}, i=1, \dots, d$

Свяжем все a_{ij} диагональной матрицей $A_{ij} = a_{ij} I_s$, где I_s - единичная матрица размера s и построим матрицу B размера $(d-1)s \times ds$ из блоков A_{ij} в соответствующем порядке. Очевидно, что B является матрицей линейно независимых сравнений, и нулевую гипотезу можно переписать в векторной форме

$$H_0 : B\theta = 0$$

Принимая во внимание (5.1), для нулевой гипотезы получим

$$n \hat{\theta}' \hat{Q}^{-1} \hat{\theta} \Rightarrow \chi_{(d-1)s}^2$$

где $\hat{Q} = B'(B\hat{\Sigma}B')^{-1}B$

Аналогичный тест может быть основан на оценке Нельсона-Аалена $\Lambda_i(t)$ и $\Lambda_i(t) = -\log S_i(t)$ при непрерывном $S_i(t), i = 1, \dots, d$
 Введем параметр $k_i = (k_{i1}, \dots, k_{is})'$, где $k_{ij} = \Lambda_i(t_j)$.

Так как $\varphi_i = Ak_i, i = 1, \dots, d$, где A матрица сравнений, тогда нулевая гипотеза может быть переписана в виде

$$H_0 : \varphi_1 = \dots = \varphi_{d-1} = 0$$

Используя функциональную предельную теорему для оценки Нельсона-Аалена, получаем, что

$$\sqrt{n_i} (\hat{\Lambda}_i(x) - \Lambda_i(x)) \Rightarrow W_{\tau_i^2(x)}$$

где $\tau_i^2(x)$ является соответствующей функции дисперсии при $i = 1, \dots, d$.

Тогда

$$\sqrt{n_i} (\hat{k}_{i1} - k_{i1}, \dots, \hat{k}_{is} - k_{is})' \Rightarrow N(0, \gamma_i)$$

где ковариационная матрица $\gamma_i = \|\tau_{i:qr}\|$, такая что $\tau_{i:qr} = \tau_i^2(t_q \wedge t_r), q, r = 1, \dots, s$. Мы используем $\hat{\Sigma}_i, \hat{\tau}_{i:qr}^2(t_q \wedge t_r)$ где $\hat{\tau}_{i:qr}^2(x)$ - состоятельная оценка $\tau_{i:qr}^2(x), i = 1, \dots, d$

Введем случайный вектор $\hat{k} = (\hat{k}_{11}, \dots, \hat{k}_{1s}, \dots, \hat{k}_{d1}, \dots, \hat{k}_{ds})'$, тогда

$$\sqrt{n} (\hat{k} - k) \Rightarrow N(0, \gamma) \quad (5.2)$$

где $\gamma = \text{diag}(l_1 \gamma_1, \dots, l_n \gamma_d)$ - блок-диагональная матрица $l_i = \frac{n}{n_i}, i = 1, \dots, d$. Нулевая

гипотеза может быть записана в следующем виде:

$$H_0: B_k = 0$$

В силу (5.2), получаем, что при нулевой гипотезе

$$n \hat{k}' \hat{R}^{-1} \hat{k} \Rightarrow \chi_{(d-1)s}^2$$

где $\hat{R} = B'(B \hat{\gamma} B')^{-1} B$

Список литературы

- 1) Fleming, T.R. & Harrington, D.P. (1991). Counting Processes and Survival Analysis. Wiley, New York
- 2) Липцер Р.Ш., Ширяев А.Н., Теория мартингалов, Наука, 1986, 512 с.
- 3) Лебедев В.А. Мартингалы, сходимость вероятностных мер и стохастические уравнения — М изд-во МАИ, 1996, 348 с.
- 4) Malov S.V. & O'Brien S.J (2013) On Survival Categorical Methods with Applications in Epidemiology and AIDS Research. In Applied Methods of Statistical Analysis. Applications in Survival Analysis, Reliability and Quality Control. Proceedings of the International Workshop AMSA'13 (Novosibirsk, September 25-27, 2013), 173–180.
- 5) Кокс Д.Р., Оукс Д., Анализ данных типа времени жизни, "Финансы и статистика" 1988, 192 с.
- 6) Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. ,Statistical Models Based on Counting Processes. Springer, New York. ,1993
- 7) Бородин А.Н., Лекции по «Случайным процессам», 2016
- 8) Ананьевский С.М., Невзоров В.Б. Теория вероятностей с примерами и задачами: учеб. пособие СПбГУ, Издательство 2013г. 244 с.