Сравнение кривых дожития с разными формами цензурирования

Караваева Валерия Андреевна, 422 группа

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — к.ф.-м.н., доц. Н.П. Алексеева
Рецензент: к.ф.-м.н., доц. П.В. Шпилёв

Санкт-Петербург
2015г.

Основные определения

Рассматриваем данные типа времени жизни.

- **Отказ** — точечное событие, которое является индикатором конца исследования для объекта.
- **Дожитие** τ_j — время до наступления *отказа* j-ого объекта.
- **Кривая дожития** — это $S(t) = \mathbb{P}(\tau > t)$, т. е. вероятность того, что отказ наступит после момента t.

Основные определения

Рассматриваем данные типа времени жизни.

- **Отказ** — точечное событие, которое является индикатором конца исследования для объекта.
- **Дожитие** τ_j — время до наступления *отказа* j-ого объекта.
- **Кривая дожития** — это $S(t) = \mathbb{P}(\tau > t)$, т. е. вероятность того, что отказ наступит после момента t.
- **Цензурирование** — это неполное наблюдение за временем ожидания отказа.
- **Момент цензурирования** для j-ого объекта — c_j.
- **Задача** анализа дожития: оценка кривых дожития и их сравнение для разных групп индивидов.

Формы цензурирования и оценки кривой дожития

- **Правостороннее цензурирование**
  $Y_j = (\tilde{\tau}_j, \delta_j)$, где $\tilde{\tau}_j \;=\; \min(\tau_j,\ c_j)$, $\delta_j = \mathbb{I}(\tau_j > c_j)$,
  $j = 1, \ldots, m$.
  Оценка: Каплана-Мейера (1958).

- **Интервальное цензурирование**
  $Y_j = \big(L_j, R_j\big)$, где $\tau_j \in (L_j, R_j)$, $0 \leq L_j < R_j$, $j = 1, \ldots, m$.
  Оценка: непараметрическая оценка Тёрнбулла (1976).

- **Медианное цензурирование**
  $Y_i = \big(\mu_i, n_i\big)$, где $\mu_i,\ n_i$ — выборочная медиана и объём подгруппы i соответственно, $i = 1, \ldots, k$.

Мотивация

Проверка эффективности лечения онкологических больных в России и за рубежом на основании данных с ***правым (RR)*** *и* ***медианным (FM)*** *типами цензурирования.*

Цели и задачи

Задачи

- *Построение оценки кривой дожития в условиях медианного цензурирования параметрическим и непараметрическим методом.*
- *Сравнение кривых дожития в условиях медианного и правого цензурирования.*

Непараметрическая оценка кривой дожития

Утверждение (DasGupta Anirban, 2008)

Пусть $\mu(n)$ *выборочная медиана группы объёма* n,

$$a_n = \left\lceil \frac{n+1}{2} \right\rceil, b_n = \left\lfloor \frac{n+1}{2} \right\rfloor, \ \mathbb{B}_x(a_n, b_n) = \frac{\int\limits_0^x t^{a_n - 1}(1-t)^{b_n - 1} dt}{\int\limits_0^1 t^{a_n - 1}(1-t)^{b_n - 1} dt},$$

— *регуляризованная неполная бета-функция,* $x \in [0, 1]$.

Тогда

$$P\{\mu(n) \leq x\} = \Psi_n(x) = \mathbb{B}_x(a_n, b_n).$$

Оценка кривой дожития (Ю.А. Матвеева, 2011)

Пусть (μ_i, n_i), $i = 1, \ldots, k$, $\mu_1 \le \mu_2 \le \cdots \le \mu_k$,
ν_i *— позиция наблюдения* μ_i *в упорядоченном ряду,*
p_i *— вероятностная нагрузка в* i*-ой точке,* $p_i \ge 0$, $\sum\limits_{i=0}^{k} p_i = 1$.
Тогда задача оценивания кривой дожития сводится к максимизации функции правдоподобия в точках $p_1, \ldots, p_k$.

$$\sum_{i=1}^{k} \log \left[\Psi_{n_i} \left(\sum_{j \le \nu_i} p_j \right) - \Psi_{n_i} \left(\sum_{j < \nu_i} p_j \right) \right] \to \max .$$

Результаты. Параметрическая оценка кривой дожития

- Функция дожития $S_{n_i}(t;\theta_i)$ в i-ой группе объёма n_i,
- $\hat{\theta}_i$ оценка параметра распределения с помощью (μ_i, n_i), где $i = 0, \ldots, k$, $\mu_0 = 0$, $n_0 = 0$.

Результаты. Параметрическая оценка кривой дожития

- Функция дожития $S_{n_i}(t;\theta_i)$ в i-ой группе объёма n_i,
- $\hat{\theta}_i$ оценка параметра распределения с помощью (μ_i, n_i), где $i = 0, \ldots, k$, $\mu_0 = 0$, $n_0 = 0$.

Утверждение

Абсолютная частота выбывания на промежутке $[\mu_{j-1}, \mu_j)$

$$\Phi([\mu_{j-1}, \mu_j)) = \sum_{i=1}^{k} n_i \left[S_{n_i}\left(\mu_{j-1};\ \hat{\theta}_i\right) - S_{n_i}\left(\mu_j;\ \hat{\theta}_i\right) \right].$$

Выбор модели на основании функции риска

Функция интенсивности (риск)

$$h(t) = \lim_{\Delta \to 0} \frac{P(t < \tau < t + \Delta | \tau > t)}{\Delta} = [-\ln(S(t))]'.$$

Выбор модели на основании функции риска

Функция интенсивности (риск)

$$h(t) = \lim_{\Delta \to 0} \frac{P(t<\tau<t+\Delta|\tau>t)}{\Delta} = [-\ln(S(t))]'.$$

Утверждение

Пусть $S_{\mathrm{U}}(t;0,\theta) = 1 - t/\theta,$ *при* $t \in [0,\theta],$

$$S_{\Gamma}(t;\lambda,\theta) = 1 - \frac{\gamma(\lambda, t/\theta)}{\Gamma(\lambda)},$$ *где* $t \in [0,\infty),$

$$\gamma(\lambda, t/\theta) = \int\limits_0^{t/\theta} e^{-x}x^{\lambda-1}dx,\ \Gamma(\lambda) = \int\limits_0^{\infty} e^{-x}x^{\lambda-1}dx.$$

Тогда

- $h_{\Gamma}(t) = 1/\theta,$ *при* $\lambda = 1.$
- $h'_{\mathrm{U}}(t) = 1/(\theta - t)^2 > 0.$
- $h'_{\Gamma}(t) < 0,$ *при* $\lambda < 1.$
- $h'_{\Gamma}(t) > 0,$ *при* $\lambda > 1.$

Результаты. Параметрическая оценка кривой дожития

Модель	***Оценка*** $\hat{\theta}$	***Абсолютная частота выбывания*** $\tau \in [\mu_{j-1}, \mu_j)$
$\mathrm{Exp}[\theta]$	$\frac{\ln 2}{\mu}$	$\sum\limits_{i=1}^{k} n_i \left(e^{-\mu_j \frac{\ln 2}{\mu_i}} - e^{-\mu_{j-1} \frac{\ln 2}{\mu_i}} \right)$
$\mathrm{U}[0, \theta]$	2μ	$(\mu_j - \mu_{j-1}) \sum\limits_{i=j}^{k} \frac{n_i}{2(\mu_i - \mu_0)}$
$\Gamma[\lambda, \theta]$	$\frac{\gamma\left(\lambda, \frac{\mu}{\hat{\theta}}\right)}{\Gamma(\lambda)} = \frac{1}{2}$	$\sum\limits_{i=1}^{k} \frac{n_i}{\Gamma(\lambda)} \left[\gamma\left(\lambda, \frac{\mu_j}{\hat{\theta}_i}\right) - \gamma\left(\lambda, \frac{\mu_{j-1}}{\hat{\theta}_i}\right) \right]$

Таблица: Оценка кривых дожития при медианном цензурировании.

Результаты. Оценка кривой дожития

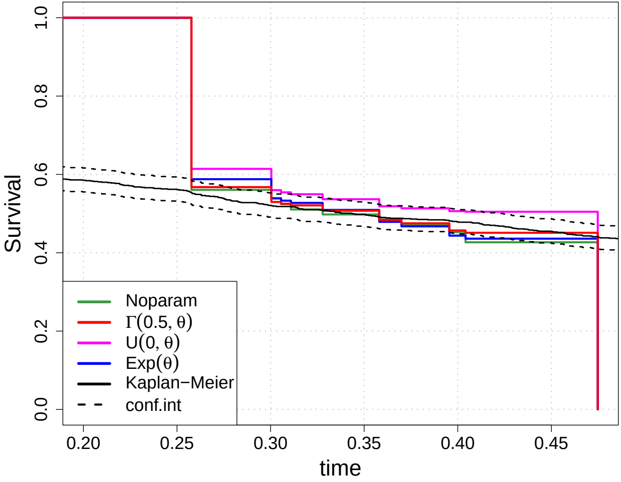
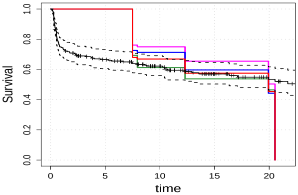
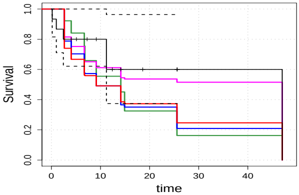


Рис.: Модельные данные. $\Gamma\left(\frac{1}{3}, 4\right)$ — убывающий риск.

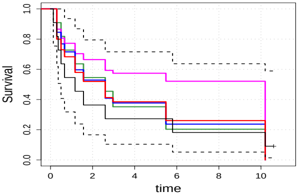
Диагноз мультиглиобластома.



(a) Тот. $\Gamma\left(\frac{1}{2}, \theta\right)$



(b) Криодест. $\mathrm{U}(0, \theta)$



(c) Биопсия. $\Gamma\left(\frac{1}{2}, \theta\right)$

Сравнение кривых дожития

Задача

Сравнение дожития в двух группах

$$\mathrm{I}: (\tilde{\tau}_1, \delta_1), (\tilde{\tau}_2, \delta_2), \ldots, (\tilde{\tau}_m, \delta_m).$$
$$\mathrm{II}: (\mu_1, n_1), (\mu_2, n_2), \ldots, (\mu_k, n_k).$$

Преобразование данных I типа к II.

- Разделение данных на k подгрупп объёма $m_i = Cn_i$ случайным образом, где $i = 1, \ldots, k$, $C = const$.

$$\overbrace{\tau_1, \ldots, \tau_{m_1}}, \ldots, \overbrace{\tau_{m_1+\cdots+m_{k-1}+1}, \ldots, \tau_{m_1+\cdots+m_k}}.$$

- $\mu_i^{(1)}$ — выборочная медиана i-ой подгруппы.
- Задача сводится к виду

$$\mathrm{I}_\xi : \left(\mu_1^{(1)}, m_1\right), \left(\mu_2^{(1)}, m_2\right), \ldots, \left(\mu_k^{(1)}, m_k\right).$$
$$\mathrm{II} : \left(\mu_1^{(2)}, n_1\right), \left(\mu_2^{(2)}, n_2\right), \ldots, \left(\mu_k^{(2)}, n_k\right).$$

Классические методы сравнения:

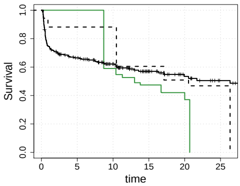
- Логранговый критерий.
- Критерий Гехана.

Проблема использования: медианное цензурирование.

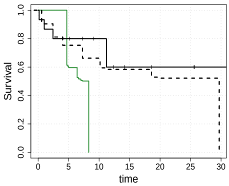
- **Преобразование** данных $\left(\mu_j^{(z)}, n_j^{(z)}\right)$, $j = 1, \dots, k$, $z = \{1, 2\}$, $\mu_1^{(z)} \leq \dots \leq \mu_{2k}^{(z)}$ к интервальному цензурированию $\left[\mu_{j-1}^{(z)}, \mu_j^{(z)}\right)$.
- **Оценка вероятностной нагрузки**.

$$V_j = \begin{cases} \Phi\left(\left[\mu_{j-1}^{(z)}, \mu_j^{(z)}\right)\right), & \text{в параметрическом,} \\ \hat{p}_j \sum\limits_{i=1}^{j} n_i^{(z)}, & \text{в непараметрическом случае.} \end{cases}$$

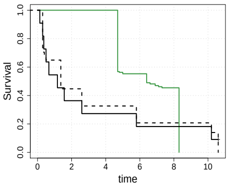
- **Метод сравнения.** Логранговый критерий для интервального цензурирования.



(d) Тот. $\Gamma\left(\frac{1}{2}, \theta\right)$



(e) Криодест. $\mathrm{U}(0, \theta)$



(f) Биопсия. $\Gamma\left(\frac{1}{2}, \theta\right)$

—— Правое ценз. (I) - - - Преобраз.(I_ξ) —— Мед. ценз. (II)

Сравнение	Значимость		Тотальное удаление	Криодест.	Биопсия
RR (I)/AM (II)	p		0.056	0.047	0.02
RR(I_ξ)/ AM(II)	p	min	0.005	0.009	0.003
		med	0.01	0.02	0.014
		max	0.031	0.045	0.06

Таблица: p-значения лограНгового критерия для интервального ценз.

Итоги и планы

Итоги

- Построены параметрическая и непараметрическая оценки для кривых дожития в условии медианного цензурирования.
- Произведено сравнение кривых дожития с разными формами цензурирования.

Итоги и планы

Итоги

- Построены параметрическая и непараметрическая оценки для кривых дожития в условии медианного цензурирования.
- Произведено сравнение кривых дожития с разными формами цензурирования.

Планы

- Оценка формы риска.
- Разработка специализированного метода сравнения кривых дожития для данных с медианным цензурированием.