Об оценке параметров специальной модели кривых дожития

Коробейников Антон Иванович, гр. 522

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., доц. Барт А.Г.
к.ф.-м.н., доц. Алексеева Н.П.
Рецензент: к.ф.-м.н., доц. Некруткин В.В.

Санкт-Петербург
2007г.

Кривые дожития

- τ — время до наступления события, называемого *отказом*.
- *Кривой дожития* называется функция $S(t) = \mathbf{P}(\tau > t)$.
- Приложения в теории надежности, медицине, биологии, страховой математике и т.п.
- Генеральная модель кривой дожития — функция распределения: $F(t) = 1 - S(t)$.

Постановка задачи

- Рассматривается специальная модель кривых дожития (Барт А.Г., Бондаренко Б.Б., Бойко В.И., 1980).
- Решается задача построения оценок параметров модели.
- Исследуются статистические свойства полученных оценок.
- Свойства оценок проверяются на модельных выборках.
- Оценки используются для обработки реальных данных из стоматологии и кардиологии.

Специальная модель кривых дожития

Модельная функция распределения задается уравнением

$$F(x;\,\eta,\tau,\mu) = 1 - \exp\left(-\eta\left(\frac{x-\mu}{\tau}\right)\right)\cos\left(\frac{\pi}{2}\left(\frac{x-\mu}{\tau}\right)\right),$$

$$\eta > 0,\ \mu < x < \mu + \tau.$$

Задача оценивания параметров данного распределения относится к так называемому «нерегулярному типу»:

1. Носитель распределения зависит от значений неизвестных параметров μ и τ.
2. Плотность распределения имеет ненулевые пределы на границе носителя.

Обозначения

- $X_1, \ldots, X_n$ — набор из n независимых одинаково распределенных случайных величин с модельной функцией распределения.
- η_0, τ_0, μ_0 — истинные значения параметров η, τ, μ.
- L_n — логарифм функции правдоподобия:

$$L_n(\eta, \tau, \mu) = \frac{1}{n} \sum_{i=1}^{n} \log f(X_i; \eta, \tau, \mu).$$

- Порядковые статистики: $X_{[1;n]} < \cdots < X_{[n;n]}$.
- $\tilde{L}_n$ — логарифм модифицированной функции правдоподобия:

$$\tilde{L}_n(\eta, \tau, \mu) = \frac{1}{n} \sum_{i=2}^{n-1} \log f\left(X_{[i;n]}; \eta, \tau, \mu\right).$$

Известные факты

- Пусть $\bar{\eta}_n$ — оценка максимального правдоподобия для η при известных значениях μ_0, τ_0.
Оценка $\bar{\eta}_n$ необходимо удовлетворяет уравнению:

$$\frac{\partial L_n}{\partial \eta}(\bar{\eta}_n, \tau_0, \mu_0) = 0.$$

- К $\bar{\eta}_n$ применимы результаты классической теории оценок макимального правдоподобия в регулярном случае.
- Известно (Cheng, Traylor, 1995), что оценки параметров, полученные максимизацией непосредственно логарифма функции правдоподобия L_n по всем трем аргументам (в нерегулярном случае), могут быть несостоятельными.

Процедура оценивания

Двухстадийная процедура:

1. Эффективное оценивание параметров τ и μ при помощи порядковых статистик:

$$\hat{\mu}_n = X_{[1;n]}, \quad \hat{\tau}_n = X_{[n;n]} - X_{[1;n]}.$$

2. Получение оценки $\tilde{\eta}_n$ как локального максимума $\tilde{L}_n$:

$$\tilde{\eta}_n = \arg\max_{\eta>0} \tilde{L}_n\left(\eta, \hat{\tau}_n, \hat{\mu}_n\right).$$

Оценка $\tilde{\eta}_n$ должна удовлетворять *уравнению правдоподобия*

$$\frac{\partial \tilde{L}_n}{\partial \eta}\left(\tilde{\eta}_n, \hat{\tau}_n, \hat{\mu}_n\right) = 0.$$

Процедура оценивания

Двухстадийная процедура:

1. Эффективное оценивание параметров τ и μ при помощи порядковых статистик:

$$\hat{\mu}_n = X_{[1;n]}, \quad \hat{\tau}_n = X_{[n;n]} - X_{[1;n]}.$$

2. Получение оценки $\tilde{\eta}_n$ как локального максимума $\tilde{L}_n$:

$$\tilde{\eta}_n = \arg\max_{\eta > 0} \tilde{L}_n\,(\eta, \hat{\tau}_n, \hat{\mu}_n).$$

Оценка $\tilde{\eta}_n$ должна удовлетворять *уравнению правдоподобия*

$$\frac{\partial \tilde{L}_n}{\partial \eta}\,(\tilde{\eta}_n, \hat{\tau}_n, \hat{\mu}_n) = 0.$$

Процедура оценивания

Двухстадийная процедура:

1. Эффективное оценивание параметров τ и μ при помощи порядковых статистик:

$$\hat{\mu}_n = X_{[1;n]}, \quad \hat{\tau}_n = X_{[n;n]} - X_{[1;n]}.$$

2. Получение оценки $\tilde{\eta}_n$ как локального максимума $\tilde{L}_n$:

$$\tilde{\eta}_n = \arg\max_{\eta>0} \tilde{L}_n\left(\eta, \hat{\tau}_n, \hat{\mu}_n\right).$$

Оценка $\tilde{\eta}_n$ должна удовлетворять *уравнению правдоподобия*

$$\frac{\partial \tilde{L}_n}{\partial \eta}\left(\tilde{\eta}_n, \hat{\tau}_n, \hat{\mu}_n\right) = 0.$$

Оценки $\hat{\tau}_n$ и $\hat{\mu}_n$

Утверждение

Для оценок $\hat{\tau}_n$, $\hat{\mu}_n$ имеет место состоятельность:

1. $\hat{\mu}_n - \mu_0 = O_p\left(\frac{1}{n}\right),\ n \to \infty.$
2. $\tau_0 - \hat{\tau}_n = O_p\left(\frac{1}{n}\right),\ n \to \infty.$

Следствие

Асимптотически:

- $n\left(\hat{\mu}_n - \mu_0\right) \sim Exp\left(\eta_0\right).$

Замечание

Оценки $\hat{\tau}_n$ и $\hat{\mu}_n$ демонстрируют сверхсходимость: дисперсия оценок имеет порядок $\frac{1}{n^2}$, против обычного $\frac{1}{n}$.

Оценка $\tilde{\eta}_n$

Теорема

1. Оценка $\tilde{\eta}_n$, удовлетворяющая уравнению правдоподобия: $\dfrac{\partial \tilde{L}_n}{\partial \eta}(\tilde{\eta}_n, \hat{\tau}_n, \hat{\mu}_n) = 0$ существует и притом единственна.
2. С вероятностью, стремящейся к 1 при $n \to \infty$ имеет место $\tilde{\eta}_n - \bar{\eta}_n = o_p\left(\dfrac{1}{n^{1-\varepsilon}}\right)$, $\forall \varepsilon : 0 < \varepsilon < 1$.

Следствие

1. $\tilde{\eta}_n$ — состоятельная оценка для параметра η.
2. $\tilde{\eta}_n$ — асимптотически эффективна.
3. Случайная величина $\sqrt{n}\,(\tilde{\eta}_n - \eta_0)$ асимптотически нормальна.

Изучение свойств оценок на модельных выборках

1. Исследование различных способов моделирования кривой дожития:
   - Метод обратных функций для малых значений η.
   - Метод отбора (из сдвиговой модификации экспоненциального распределения) для не очень малых значений η.
2. Проверка теоретических результатов на модельных выборках:
   - Изучение скорости сходимости оценок параметров.
   - Изучение асимптотического распределения оценок параметров.

Изучение свойств оценок на модельных выборках

1. Исследование различных способов моделирования кривой дожития:
   - Метод обратных функций для малых значений η.
   - Метод отбора (из сдвиговой модификации экспоненциального распределения) для не очень малых значений η.
2. Проверка теоретических результатов на модельных выборках:
   - Изучение скорости сходимости оценок параметров.
   - Изучение асимптотического распределения оценок параметров.

Цензурирование

При проведении процедуры сбора реальных данных типично *цензурирование* — наблюдение над индивидом прекращается до наступления момента отказа.

Модель цензурирования:

- $\tau_j \in \{0, 1\}$ — показатель цензурирования.
- c_j — момент цензурирования.
- Наблюдается пара $\left(Y_j, \tau_j\right)$, где $Y_j = \tau_j c_j + \left(1 - \tau_j\right) X_j$.
- X_j, τ_j, c_j — все зависимы.

Типичный случай —«цензурирование справа»:

$$\tau_j = \mathbb{I}\left(X_j > c\right), \quad Y_j = \min\left\{c, X_j\right\}, \quad c_j = c,$$

где $\mathbb{I}$ — индикатор множества, c — известная константа.

Оценки по выборке с цензурированием

Модифицированная двухстадийная процедура:

1. Оценивание параметра μ при помощи порядковых статистик:
$$\hat{\mu}_n = X_{[1;n]}.$$
2. Получение оценок $\tilde{\eta}_n^c$ и $\tilde{\tau}_n^c$ как локального максимума $\tilde{L}_n^c$:

$$\tilde{L}_n^c(\eta, \tau, \mu) = \frac{1}{n}\left[\sum_{i:\,\tau_i=0} \log f(Y_i; \eta, \tau, \mu) + \sum_{i:\,\tau_i=1} \log\left(1 - F(Y_i; \eta, \tau, \mu)\right)\right],$$

$$(\tilde{\eta}_n^c,\ \tilde{\tau}_n^c) = \arg\max_{\eta>0,\tau>0} \tilde{L}_n^c(\eta, \tau, \hat{\mu}_n).$$

Оценки по выборке с цензурированием

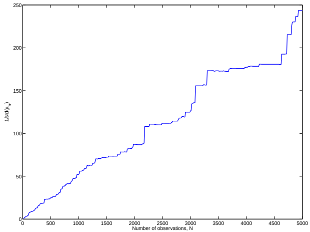
Модифицированная двухстадийная процедура:

1. Оценивание параметра μ при помощи порядковых статистик:
$$\hat{\mu}_n = X_{[1;n]}.$$
2. Получение оценок $\tilde{\eta}_n^c$ и $\tilde{\tau}_n^c$ как локального максимума $\tilde{L}_n^c$:

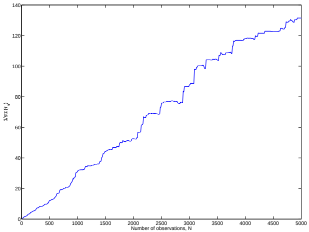
$$\tilde{L}_n^c(\eta, \tau, \mu) = \frac{1}{n}\left[\sum_{i:\,\tau_i=0} \log f(Y_i; \eta, \tau, \mu) + \sum_{i:\,\tau_i=1} \log\left(1 - F(Y_i; \eta, \tau, \mu)\right)\right],$$

$$(\tilde{\eta}_n^c,\ \tilde{\tau}_n^c) = \operatorname*{arg\,max}_{\eta>0,\tau>0} \tilde{L}_n^c(\eta, \tau, \hat{\mu}_n).$$

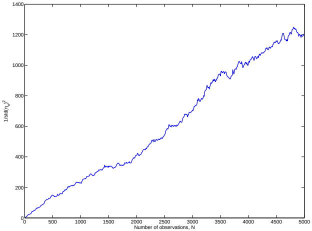
Результаты моделирования (без цензурирования)



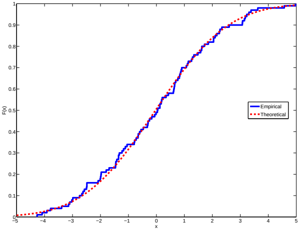
(a) $(\mathrm{std}\,\hat{\mu}_n)^{-1} = O(n)$



(b) $(\mathrm{std}\,\hat{\tau}_n)^{-1} = O(n)$



(c) $(\mathrm{std}\,\tilde{\eta}_n)^{-2} = O(n)$



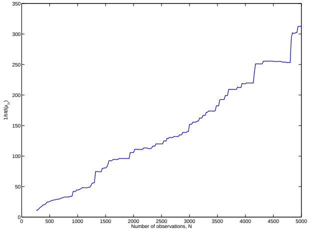
(d) $\tilde{\eta}_n \sim N(\eta_0, \sigma^2)$

Результаты моделирования (с цензурированием)

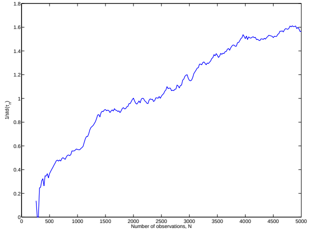
Оценки по выборке с цензурированием:

1. Оценка $\hat{\mu}_n$ демонстрирует сверхсходимость, оценки $\tilde{\tau}_n^c$, $\tilde{\eta}_n^c$ — обычную скорость сходимости.
2. $\hat{\mu}_n$ асимптотически распределена экспоненциально, распределение оценок $\tilde{\tau}_n^c$, $\tilde{\eta}_n^c$ хорошо аппроксимируется нормальным.
3. Оценка $\tilde{\eta}_n^c$ смещена. Смещение вызвано цензурированием и одновременным оцениванием параметра τ.

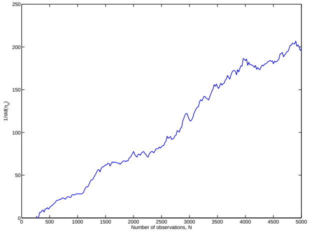
Стандартное отклонение оценок (с цензурированием)



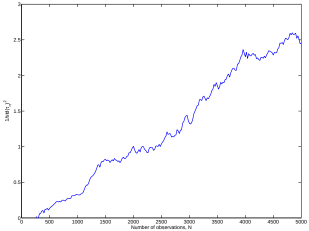
(e) $(\mathrm{std}\,\hat{\mu}_n)^{-1}$



(f) $(\mathrm{std}\,\tilde{\tau}_n^c)^{-1}$



(g) $(\mathrm{std}\,\tilde{\eta}_n^c)^{-2}$



(h) $(\mathrm{std}\,\tilde{\tau}_n^c)^{-2}$

Пример из кардиологии