

М. Г. Лапина

преподаватель департамента экономики Санкт-Петербургской школы экономики и менеджмента Национального исследовательского университета «Высшая школа экономики»

Г. М. Фридман

докт. техн. наук, зав. кафедрой экономической кибернетики и математических методов в экономике Санкт-Петербургского государственного экономического университета

ОТНОСИТЕЛЬНАЯ АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ ОЦЕНОК, ПОЛУЧЕННЫХ МЕТОДОМ КАПЛАНА — МЕЙЕРА И МЕТОДОМ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ (НА ПРИМЕРЕ ПРОДАЖИ АВИАБИЛЕТОВ)

Введение

Задачи анализа выживаемости, в которых изучаются закономерности появления ожидаемых событий у представителей наблюдаемой выборки с течением времени, находят применение во многих областях, таких как теория надежности, медицина, социология, экономические и финансовые исследования, страховая сфера. Например, в медико-биологических задачах ожидаемым событием может являться выздоровление пациента, рецидив заболевания либо летальный исход, а в демографических — вступление в брак или прибытие в новый город. Характерной особенностью при этом является то, что момент наступления ожидаемого события известен не для всех элементов наблюдаемой выборки. Данные, которые содержат неполную информацию, называют цензурированными. Причины цензурирования в зависимости от объекта исследования могут быть различны: выбывание из исследования, включение в исследование в его середине или в конце, установление каких-либо ограничений по времени или количеству и т. п.

В задачах управления доходами (*Revenue Management*, далее — *RM*) также применимы методы анализа выживаемости, при этом ожидаемым событием служит удовлетворение спроса на какой-либо продукт, а причиной цензурирования является ограничение (верхняя граница), наложенное на общий объем продаж этого продукта. Восстановление данных общего спроса (в том числе оценка параметров его распределения) по наблюдаемой выборке является целью исследования в этом классе задач *RM*.

Для анализа цензурированных данных существуют различные параметрические и непараметрические методы, обладающие по сравнению друг с другом рядом преимуществ и недостатков. Как правило, среди преимуществ непараметрических методов выделяют следующие:

- 1) менее жесткие по сравнению с параметрическими методами начальные допущения;
- 2) большая робастность;
- 3) относительная несложность вычислений (в большинстве случаев).

К недостаткам же относятся:

1) меньшая эффективность, чем у параметрических методов; эффективность несмещенной оценки $\hat{\theta}$ параметра θ :

$$eff(\hat{\theta}) = \frac{1/I(\theta)}{\text{var } \hat{\theta}}, \quad (1)$$

где $\frac{1}{I(\theta)}$ — нижняя граница неравенства Рао—Крамера, $\text{var } \hat{\theta}$ — дисперсия оценки $\hat{\theta}$, $0 \leq eff(\hat{\theta}) \leq 1$;

2) трудоемкость при применении к большим массивам данных.

Выделяются несколько разновидностей цензурирования: 1-го, 2-го типов, случайное цензурирование (справа, слева), а также интервальное цензурирование. В статье рассмотрено случайное правое цензурирование. При этом типе цензурирования данных для оценки параметров или восстановления функции надежности исходной случайной величины широко распространен метод Каплана—Мейера (Kaplan, Meier, 1958) (непараметрический метод), а также метод максимального правдоподобия (стандартный параметрический метод оценки параметров).

Метод Каплана—Мейера обладает рядом преимуществ: оценки, полученные с его помощью, имеют асимптотически нормальное распределение с дисперсией, рассчитываемой по формуле Гринвуда (Greenwood, 1926), в рамках этого метода не требуется никаких предположений о типе распределения данных, единственное требование — независимость исходной и цензурирующей случайных величин. Однако метод Каплана—Мейера (как представитель непараметрических методов) может обладать меньшей эффективностью, чем параметрические методы.

Первое и широко цитируемое исследование относительной асимптотической эффективности оценок (т. е. эффективности при неограниченном росте размера выборки), полученных методом Каплана—Мейера и методом максимального правдоподобия, проводилось в статье (Miller, 1983). В ней рассмотрены два варианта исходных распределений, подвергающихся цензурированию, а именно экспоненциальное распределение и распределение Вейбулла. В качестве цензурирующих распределений использовано экспоненциальное и равномерное распределение. В статье продемонстрировано, что если исходное распределение выбрано корректно, асимптотическая эффективность оценок, полученных методом Каплана—Мейера, ниже, особенно для высокого процента цензурированных элементов (рассмотренный максимум — 50%) или значений функции надежности, близких к 0 или 1.

В данной статье рассматривается случай, когда ожидаемым событием является удовлетворение спроса на какой-либо продукт, например на авиабилеты различных ценовых классов или на гостиничные номера разного уровня комфортности. Если спрос на продукт превышает фактически установленный (ввиду ограничений по вместимости или накладываемых пределов бронирования) верхний предел продаж, то значение наблюдаемой случайной величины, соответствующей фактическому спросу, будет цензурировано (Лапина, Носова, Фридман, 2011; Lapina, Fridman, 2013; Talluri, Ryzin, 2004) (см. табл. 1). Таким образом, наблюдаемая цензурированная выборка будет содержать данные по количеству продаж данного продукта, а не по фактическому спросу на него.

Таблица 1

Пример появления цензурированной наблюдаемой выборки для данных по спросу*

Общий спрос (единицы продукта)	18	20	21	19	20	17	21	23	18	22
Предел бронирования	20	19	17	22	21	18	19	21	22	20
Наблюдаемый спрос	18	19	17	19	20	17	19	21	18	20

* Цензурированные элементы наблюдаемой выборки помещены в серые ячейки таблицы.

При моделировании использовано предположение о том, что спрос на некоторый продукт является случайной величиной, подчиняющейся усеченному нормальному распределению и подвергшейся цензурированию. В качестве цензурирующих взяты случайные величины с тремя типами распределений: усеченного нормального, экспоненциального и равномерного. Анализу подлежал вопрос о том, какой метод, параметрический или непараметрический, обладает большей эффективностью при восстановлении исходной случайной величины (общего спроса на продукт) по цензурированной наблюдаемой выборке (числу продаж продукта).

Постановка задачи

Пусть X и Y — независимые абсолютно непрерывные случайные величины с функциями распределений и $F(x; \theta)$ и $Y(y, \theta)$, где $\theta \in \Theta \subset R^1$ — параметры распределений. Распределение случайной величины X назовем исходным распределением, распределение случайной величины Y — цензурирующим распределением. В случае случайного правого цензурирования будем наблюдать (Z, δ) , где $Z = \min(X, Y)$ и $\delta = I(X \leq Y)$. $I(A)$ — функция-индикатор, принимающая значения 0, если $X \leq Y$, и 1, если $X > Y$.

Обозначим через $\bar{z} = (z_1, z_2, \dots, z_n)$ упорядоченную выборку из n наблюдений случайной величины Z , через $\bar{\delta} = (\delta_1, \delta_2, \dots, \delta_n)$ — соответствующую выборку значений индикаторов цензурированности. Процентом цензурирования будем называть относительную долю элементов с $\delta = 1$, т. е. элементов, подвергшихся цензурированию.

Оценка функции надежности (функции выживаемости) $S(x) = 1 - F(x)$ по методу Каплана—Мейера рассчитывается следующим образом:

$$S(x) = \prod_{y_i \leq x} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \quad (2)$$

$S(x)$ является состоятельной асимптотически нормальной оценкой $S(x)$ с дисперсией:

$$\text{var}_{asy} S(x) = \frac{1}{n} (1 - F(x))^2 \int_0^x \frac{f(u) du}{(1 - F(u))^2 (1 - G(u))} \quad (3)$$

где $f(x) = dF(x) / dx$.

Для применения параметрического метода максимального правдоподобия необходимо задать плотность распределения цензурированной случайной величины,

а затем сформировать функцию максимального правдоподобия, максимизация которой даст необходимые оценки параметров, которые позволят перейти к функции надежности исходной случайной величины.

Плотность совместного распределения будет выглядеть следующим образом:

$$h(z, \delta; \theta) = [\bar{G}(z, \theta) f(z, \theta)]^\delta [\bar{F}(z, \theta) g(z, \theta)]^{1-\delta} \quad (4)$$

Здесь f и g — плотности распределений, $\bar{F} = 1 - F$ и $\bar{G} = 1 - G$ — функции надежности. Тогда функция максимального правдоподобия для выборки (z_i, δ_i) , $i = 1, \dots, n$ имеет следующий вид (Miller, 1981; Prakasa Rao, 1995):

$$L(z; \theta) = \prod_{i=1}^n [\bar{G}(z_i, \theta) f(z_i, \theta)]^\delta [\bar{F}(z_i, \theta) g(z_i, \theta)]^{1-\delta}. \quad (5)$$

Остановимся на частном случае, когда цензурирующее распределение G не зависит от параметров θ . В этом случае функция максимального правдоподобия запишется:

$$L(z; \theta) = \prod_{i=1}^n [f(z_i, \theta)]^\delta [\bar{F}(z_i, \theta)]^{1-\delta}, \quad (6)$$

так как при максимизации функции $L(z; \theta)$ компоненты \bar{G} и g ведут себя как константы.

При выполнении условий регулярности (Rao, 1973) оценки параметров θ , полученные методом максимального правдоподобия, являются состоятельными асимптотически нормальными с дисперсией $I^{-1}(\theta)/n$, где $I(\theta)$ — информационная матрица Фишера. Для случая одного параметра:

$$I(\theta) = \int_0^\infty \left(\frac{\partial \log f_\theta(u)}{\partial \theta} \right)^2 f_\theta(u) (1 - G(u)) du + \int_0^\infty \left(\frac{\partial \log(1 - F_\theta(u))}{\partial \theta} \right)^2 (1 - F_\theta(u)) g(u) du. \quad (7)$$

Для случая двух параметров $\theta = (\theta_1, \theta_2)$:

$$I_{11}(\theta_1, \theta_2) = \int_0^\infty \left(\frac{\partial \log f_{\theta_1, \theta_2}(u)}{\partial \theta_1} \right)^2 f_{\theta_1, \theta_2}(u) (1 - G(u)) du + \int_0^\infty \left(\frac{\partial \log(1 - F_{\theta_1, \theta_2}(u))}{\partial \theta_1} \right)^2 (1 - F_{\theta_1, \theta_2}(u)) g(u) du, \quad (8)$$

$$I_{12}(\theta_1, \theta_2) = \int_0^\infty \left(\frac{\partial \log f_{\theta_1, \theta_2}(u)}{\partial \theta_1} \right) \left(\frac{\partial \log f_{\theta_1, \theta_2}(u)}{\partial \theta_2} \right) f_{\theta_1, \theta_2}(u) (1 - G(u)) du + \int_0^\infty \left(\frac{\partial \log(1 - F_{\theta_1, \theta_2}(u))}{\partial \theta_1} \right) \left(\frac{\partial \log(1 - F_{\theta_1, \theta_2}(u))}{\partial \theta_2} \right) (1 - F_{\theta_1, \theta_2}(u)) g(u) du. \quad (9)$$

$$I_{22}(\theta_1, \theta_2) = \int_0^\infty \left(\frac{\partial \log f_{\theta_1, \theta_2}(u)}{\partial \theta_2} \right)^2 f_{\theta_1, \theta_2}(u) (1 - G(u)) du + \int_0^\infty \left(\frac{\partial \log(1 - F_{\theta_1, \theta_2}(u))}{\partial \theta_2} \right)^2 (1 - F_{\theta_1, \theta_2}(u)) g(u) du. \quad (10)$$

Тогда соответствующая оценка функции $S_{\hat{\theta}(x)}$ для случая $\theta \in \Theta \subset R^1$ будет иметь асимптотически нормальное распределение с дисперсией:

$$\text{var}_{asy} S_{\theta}(x) = \frac{1}{n} I^{-1}(\theta) \left(\frac{\partial S_{\theta}(x)}{\partial \theta} \right)^2, \quad (11)$$

выражение которой получено с помощью дельта-метода (Cramer, 1946; Oehlert, 1992). Для случая нескольких параметров $\theta \in \Theta \subset R^k$:

$$\text{var}_{asy} S_{\theta}(x) = \frac{1}{n} \nabla S^T I^{-1}(\theta) \nabla S, \quad (12)$$

где ∇S — градиент функции $S_{\theta(x)}$.

Так как относительная эффективность двух несмещенных оценок рассчитывается как отношение их обратных дисперсий, отношение обратных асимптотических дисперсий для метода Каплана—Мейера и метода максимального правдоподобия говорит об относительной асимптотической эффективности (обозначим ее через $R(x)$) оценок функции надежности, полученных этими двумя методами. В результате расчетов для конкретного исходного распределения и различных цензурирующих распределений выявлен характер зависимостей относительных асимптотических эффективностей оценок от процента цензурирования, значения функции надежности и типа цензурирующего распределения.

Сравнительное тестирование методов

При моделировании спроса одним из самых часто используемых распределений является нормальное. При работе со средними и большими значениями спроса хвост, приходящийся на отрицательные значения, весьма мал и вероятность получить отрицательное значение спроса практически нулевая. При работе же с малыми значениями спроса влияние хвоста уже существенно, поэтому необходимо произвести усечение нормального распределения по нулю или воспользоваться, например, гамма-распределением.

Рассмотрим усеченное по нулю нормальное распределение в качестве исходного распределения. В качестве цензурирующих распределений выбраны усеченное нормальное, экспоненциальное и равномерное распределения.

Таблица 2

Функции и плотности используемых распределений

Распределение	Функция распределения	Плотность распределения
Усеченное нормальное	$F(x) = \frac{-\frac{1}{2} \text{Erfc}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \frac{1}{2} \text{Erfc}\left(\frac{-x+\mu}{\sqrt{2}\sigma}\right)}{1 - \frac{1}{2} \text{Erfc}\left(\frac{\mu}{\sqrt{2}\sigma}\right)},$ $x \geq 0$	$f(x) = \frac{\frac{(x-\mu)^2}{e^{-2\sigma^2}}}{\sqrt{2\pi}\sigma \left(1 - \frac{1}{2} \text{Erfc}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)},$ $x \geq 0$
Экспоненциальное	$F(x) = \lambda e^{-\lambda x}, x \geq 0$	$f(x) = 1 - e^{-\lambda x}, x \geq 0$
Равномерное на интервале $[0; \max]$	$F(x) = \frac{x}{\max}, 0 \leq x \leq \max$	$f(x) = \frac{x}{\max}, 0 \leq x \leq \max$

Для сравнительного анализа асимптотических эффективностей, присущих двум методам в зависимости от процента цензурирования и, соответственно, обеспечения необходимого процента цензурирования, использовалась следующая формула:

$$P(X > Y) = p,$$

где X и Y — случайные величины с заданными распределениями, p — необходимый процент цензурирования. С учетом заданных параметров распределения случайной величины X и заданных процентов цензурирования p неизвестный параметр распределения случайной величины Y , обеспечивающий необходимый процент цензурирования, рассчитывался численно.

В качестве параметров исходной случайной величины приняты $\mu_1 = 18$, $\sigma_1 = 4$, параметр масштаба цензурирующего распределения $\sigma_2 = 4$. Варианты процентов цензурирования: 25, 50, 75, 90, 98%. В табл. 3 приведены рассчитанные параметры цензурирующих распределений.

Таблица 3

Значения параметров цензурирующего распределения, необходимые для обеспечения следующих процентов цензурирования: 25%, 50%, 75%, 90%, 98%

Процент цензурирования p , %	Усеченное нормальное распределение, μ^*	Экспоненциальное распределение, λ	Равномерное распределение, \max
25	21,82	0,016	72,00
50	18,00	0,039	36,00
75	14,18	0,080	23,83
90	10,74	0,136	18,48
98	7,27	0,244	13,58

Табл. 4 и рис. 1 иллюстрируют результаты расчетов для усеченного нормального цензурирующего распределения. В табл. 4 представлены значения относительной асимптотической эффективности $R(x)$ метода Каплана—Мейера по сравнению с методом максимального правдоподобия для различных значений функции надежности $S(x) = P(X > x)$ и различных процентов цензурирования $p = P(Y < X)$. Напомним, что значения функции $R(x)$ меньше 1 свидетельствуют о превосходстве асимптотической эффективности оценок, полученных методом максимального правдоподобия по сравнению с методом Каплана—Мейера.

На рис. 1 изображены функции относительной асимптотической эффективности $R(x)$. Числовые расчеты ясно демонстрируют, что асимптотическая эффективность непараметрического метода снижается по сравнению с асимптотической эффективностью параметрического метода в следующих случаях: при увеличении процента цензурирования и приближении x к нулю или его устремлении на бесконечность (что соответствует нулевому и единичному значениям функции надежности).

Таблица 4

Относительная асимптотическая эффективность метода Каплана—Мейера по сравнению с методом максимального правдоподобия*

$S(x) p, \%$	0,9	0,75	0,5	0,25	0,1
25	0,65	0,66	0,66	0,67	0,47
50	0,66	0,65	0,66	0,57	0,28
75	0,64	0,61	0,62	0,34	0,10
90	0,57	0,59	0,47	0,16	0,03
98	0,53	0,48	0,20	0,03	0,002

* Исходное распределение: усеченное нормальное. Цензурирующее распределение: усеченное нормальное.

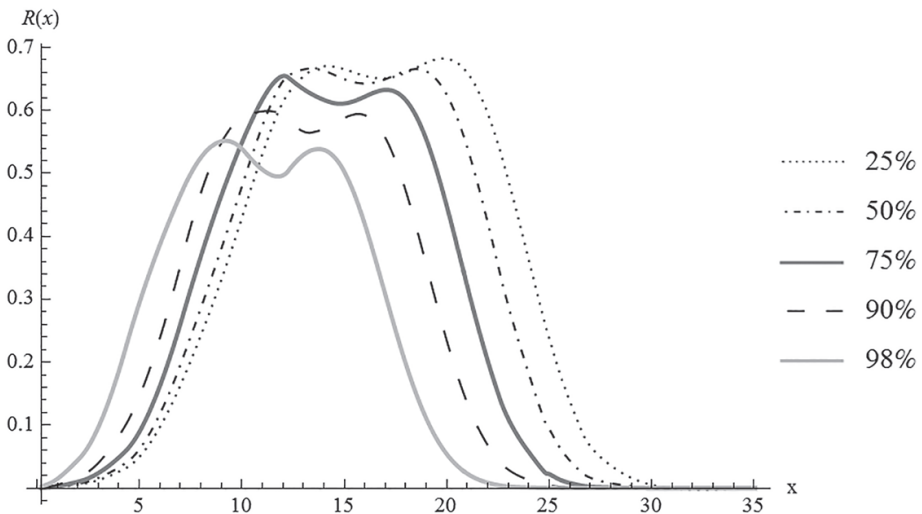


Рис. 1. Относительная асимптотическая эффективность метода Каплана—Мейера по сравнению с методом максимального правдоподобия как функция от возможных значений случайной величины X для различных процентов цензурирования

Исходное распределение: усеченное нормальное. Цензурирующее распределение: усеченное нормальное.

Табл. 5 и соответствующий ей рис. 2 иллюстрируют результаты расчета для случая экспоненциального распределения.

Как и для предыдущего случая, расчеты показали, что асимптотическая эффективность метода Каплана—Мейера снижается по сравнению с асимптотической эффективностью метода максимального правдоподобия в случае увеличения процента цензурирования и приближения значений функции надежности к нулю или единице. Но следует отметить, что снижение эффективности наблюдается уже в меньших размерах, нежели в случае усеченного нормального распределения.

Таблица 5

Относительная асимптотическая эффективность метода Каплана—Мейера по сравнению с методом максимального правдоподобия*

$S(x) p, \%$	0,9	0,75	0,5	0,25	0,1
25	0,62	0,64	0,62	0,66	0,60
50	0,62	0,62	0,61	0,65	0,57
75	0,59	0,58	0,60	0,62	0,51
90	0,54	0,53	0,59	0,57	0,40
98	0,46	0,51	0,55	0,41	0,21

* Исходное распределение: усеченное нормальное. Цензурирующее распределение: экспоненциальное.

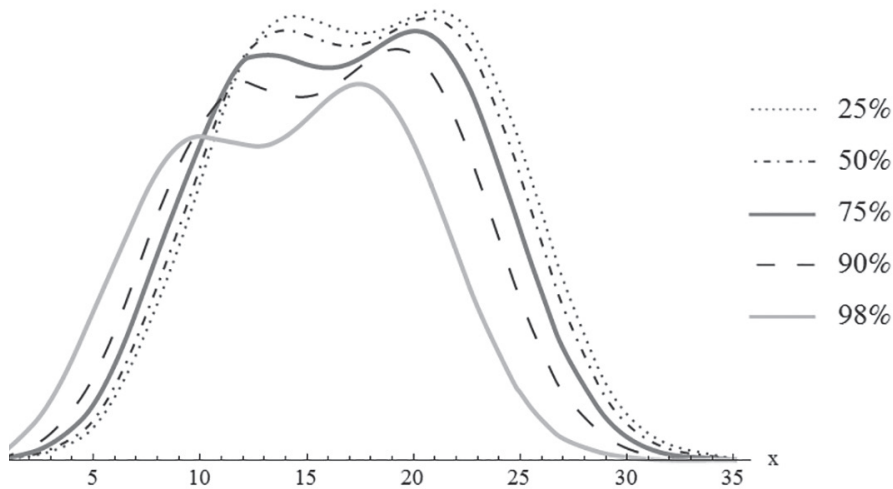


Рис. 2. Относительная асимптотическая эффективность метода Каплана—Мейера по сравнению с методом максимального правдоподобия как функция от возможных значений случайной величины X для различных процентов цензурирования

Исходное распределение: усеченное нормальное. Цензурирующее распределение: экспоненциальное.

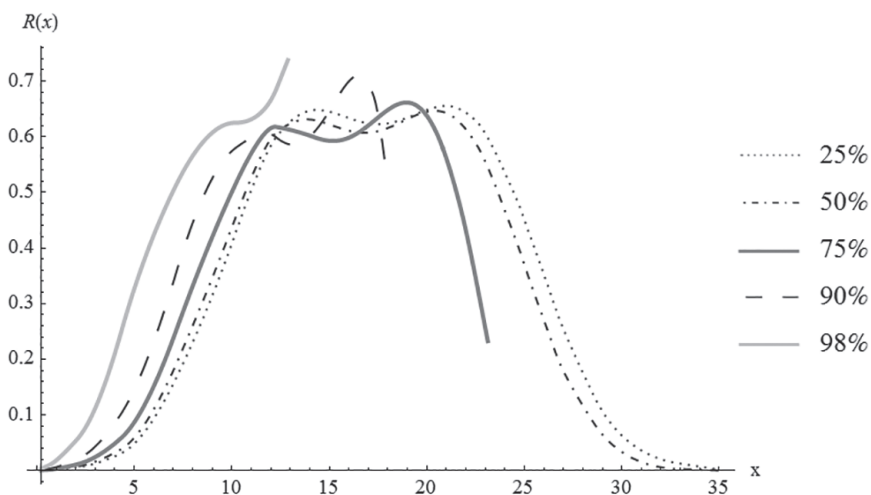
Табл. 6 и рис. 3 иллюстрируют результаты для равномерного $[0; \max]$ цензурирующего распределения. Значения относительной эффективности могут быть получены только для тех значений функции надежности, которым соответствуют $x < \max$. Выявленные закономерности наблюдаются и для данного случая, единственным исключением является скачок относительной эффективности для 98% цензурирования и значения функции надежности 0,9. Это свидетельство того, что цензурирование может как повысить, так и понизить эффективность методов — в зависимости от влияния цензурирующего распределения $G(y; \theta)$ на исходное распределение $F(x; \theta)$.

Таблица 6

**Относительная асимптотическая эффективность метода Каплана—Мейера
по сравнению с методом максимального правдоподобия***

$S(x) p, \%$	0,9	0,75	0,5	0,25	0,1
25	0,62	0,64	0,62	0,65	0,60
50	0,62	0,62	0,62	0,65	0,55
75	0,61	0,59	0,65	0,59	0,23
90	0,59	0,66	0,51	—	—
98	0,74	—	—	—	—

* Исходное распределение: усеченное нормальное. Цензурирующее распределение: равномерное на интервале $[0; \max]$.



**Рис. 3. Относительная асимптотическая эффективность метода Каплана—Мейера
по сравнению с методом максимального правдоподобия как функция от возможных значений
случайной величины X для различных процентов цензурирования**

Исходное распределение: усеченное нормальное. Цензурирующее распределение: равномерное на интервале $[0; \max]$.

Рис. 4 и 5 демонстрируют поведение функций, соответствующих относительной асимптотической эффективности для случаев трех различных цензурирующих распределений: усеченного нормального, экспоненциального и равномерного. Так, например, для 25-процентного цензурирования функции с экспоненциальным и равномерным цензурирующими распределениями довольно близки, снижение же относительной асимптотической эффективности для случая усеченного нормального распределения наблюдается для меньших по сравнению с двумя другими случаями x . Для 75-процентного цензурирования снижение относительной асимптотической эффективности для случая экспоненциального распределения наблюдается уже для меньших по сравнению с равномерным случаем значений x .

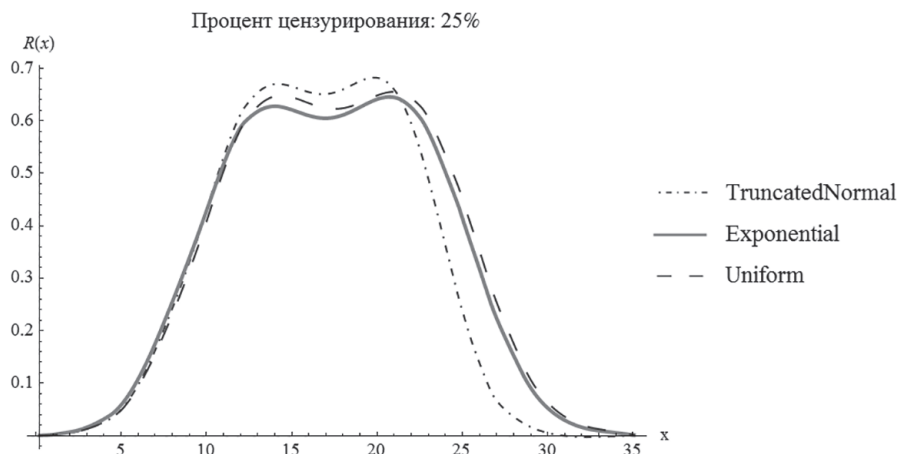


Рис. 4. Относительная асимптотическая эффективность метода Каплана—Мейера по сравнению с методом максимального правдоподобия как функция от возможных значений случайной величины X для различных процентов цензурирования.

Исходное распределение: усеченное нормальное. Цензурирующие распределения: усеченное нормальное, экспоненциальное, равномерное на интервале $[0; \max]$. Процент цензурирования: 25%.

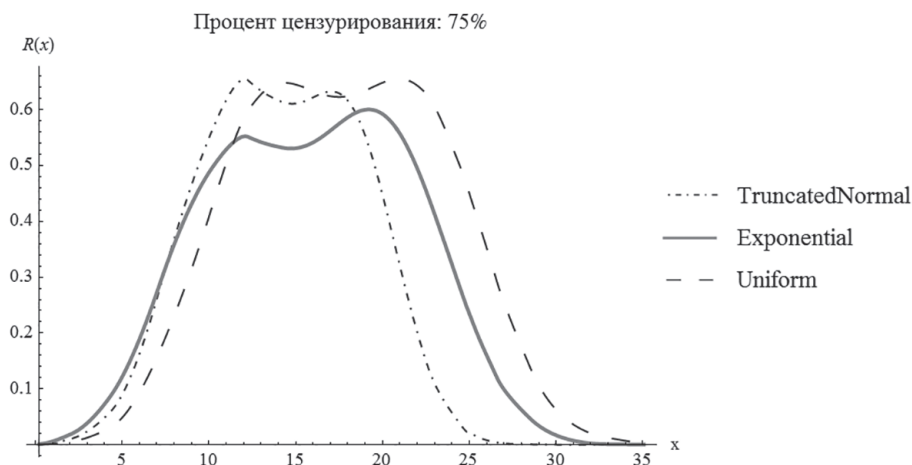


Рис. 5. Относительная асимптотическая эффективность метода Каплана — Мейера по сравнению с методом максимального правдоподобия как функция от возможных значений случайной величины X для различных процентов цензурирования.

Исходное распределение: усеченное нормальное. Цензурирующие распределения: усеченное нормальное, экспоненциальное, равномерное на интервале $[0; \max]$. Процент цензурирования: 75%.

Важно отметить, что приведенные в статье результаты получены в предположении, что закон распределения, заложенный в алгоритм метода максимального правдоподобия, совпадает с законом распределения случайной величины X описывающей неограниченный спрос. Если эти распределения не совпадают, относительная асимптотическая эффективность методов может принимать другие значения (Meier, Karrison, Chappell, Xie, 2004; Aranda-Ordaz, 1987; Klein, Moeschberger, 1989).

Пример использования сравнимых методов восстановления для прогнозирования пассажирского спроса на авиабилеты

Одним из основных приложений методов, которые обсуждены в статье, является построение прогноза общего спроса на некоторый продукт по историческим данным о количестве продаж этого продукта. В частности, такая задача возникает при определении стратегии продаж авиабилетов, причем продуктом в этом случае является так называемый «ценовой класс», т. е. сочетание полетного маршрута (последовательности стыкующихся между собой рейсов) и цены, за которую продается билет на маршрут. Основная цель при этом — вычисление оптимальных пределов бронирования, ограничивающих число продаж сверху и обеспечивающих наибольшую ожидаемую прибыль. Естественно, что более точный прогноз спроса может привести к более эффективной стратегии продаж. Единственной базой, на которой строится прогноз, является историческая информация о количестве проданных билетов по каждому ценовому классу для ранее отправленных рейсов одного и того же направления, а продажи, в свою очередь, определяются в том числе наложенными пределами бронирования. Так возникает цензурированная выборка («наблюдаемый спрос»), по которой необходимо восстановить спрос неограниченный. Эффективность выбранного авиакомпанией математического метода восстановления спроса играет большую роль в повышении точности прогноза и, соответственно, в увеличении общей прибыли от продаж авиабилетов.

В терминах, принятых в статье, случайная величина X соответствует общему спросу на конкретный ценовой класс в определенный временной срез, а случайная величина Y описывает ограничения (пределы бронирования), наложенные на количество возможных продаж в этом ценовом классе для данного рейса. Тогда случайная величина Z соответствует наблюдаемому спросу, т. е. цензурированной случайной величине.

Для моделирования исторических данных воспользуемся первой (из рассмотренных ранее) комбинацией распределений: в качестве исходного и цензурирующего распределений примем усеченное нормальное.

Таблица 7

Оценки параметров исходного спроса и соответствующие величины средней абсолютной ошибки в процентах (Mean Absolute Percentage Error, MAPE), полученные методом максимального правдоподобия (ML) и методом Каплана—Мейера (KM)

p (%)	$\hat{\mu}$ и $\hat{\sigma}$, ML	MAPE (%), ML	$\hat{\mu}$ и $\hat{\sigma}$, KM	MAPE (%), KM
25	{17,99, 3,96}	{1,69, 6,9}	{17,98, 3,91}	{1,69, 7,50}
50	{18,00, 3,98}	{1,46, 7,86}	{17,92, 3,80}	{1,83, 10,44}
75	{17,96, 3,92}	{2,26, 10,12}	{17,62, 3,39}	{3,03, 17,09}
90	{18,00, 3,96}	{4,47, 18,14}	{17,08, 2,82}	{6,00, 30,00}
98	{18,15, 4,02}	{10,00, 29,37}	{15,30, 1,65}	{15,14, 58,64}

В табл. 7 представлены оценки $\hat{\mu}$ и $\hat{\sigma}$, параметров распределения случайной величины X (соответствующей общему спросу), а также соответствующие величины средней абсолютной ошибки в процентах (Mean Absolute Percentage Error, MAPE), полученные методом максимального правдоподобия (maximum likelihood, ML) и методом Каплана—Мейера (Kaplan—Meier, KM) для различных процентов

цензурирования p наблюдаемой выборки. Истинные значения этих параметров таковы: $\hat{\mu} = 18$, $\hat{\sigma} = 4$. Наблюдаемый спрос был представлен сгенерированной случайной выборкой из 100 элементов, что ориентировочно соответствует двухгодичной статистике (104 недели), собираемой авиакомпанией. Для исключения возможных выбросов числовые расчеты повторялись по 100 раз (каждый раз наблюдаемая выборка генерировалась заново), в качестве окончательного результата использовалось среднее значение.

Анализ результатов позволяет сделать вывод о том, что с увеличением процента цензурированных элементов оценки параметров общего спроса, полученные по функции надежности, восстановленной с помощью метода Каплана—Мейера, становятся существенно менее точными (недооцененными) по сравнению с оценками, полученными по методу максимального правдоподобия. Использование заниженных оценок параметров приведет к некорректному прогнозу пассажирского спроса, не соответствующим спросу пределам бронирования и, как следствие, к потерям прибыли.

Заключение

В статье проведен сравнительный анализ асимптотической эффективности оценок, полученных методами Каплана—Мейера и максимального правдоподобия. Такое сравнение имеет важное практическое применение в задачах управления доходами, когда в роли цензурируемой случайной величины выступает спрос на некоторый продукт. Выявление и дальнейшее использование наиболее эффективного метода восстановления спроса по его цензурированной выборке позволяет снизить вероятность недооценки общего спроса, неизбежно ведущей к уменьшению прибыли.

Рассмотрена случайная величина, подчиняющаяся усеченному нормальному распределению и цензурированная справа другой случайной величиной, распределенной по одному из законов: усеченному нормальному, экспоненциальному либо равномерному распределением. Установлено, что параметрический метод максимального правдоподобия в случае корректного предположения о распределении случайной величины является асимптотически более эффективным по сравнению с непараметрическим методом Каплана—Мейера. Это преимущество особенно ярко проявляется для высокой доли цензурированных элементов наблюдаемой выборки и для значений функции надежности, близких к 0 или 1. Выявлено, что увеличение процента цензурирования может как повысить, так и понизить эффективность методов. Относительная асимптотическая эффективность методов зависит и от вида цензурирующего распределения. Выбор усеченного нормального цензурирующего распределения ведет к меньшим, чем для цензурирующих экспоненциального и равномерного распределений, величинам относительной асимптотической эффективности метода Каплана—Мейера по сравнению с методом максимального правдоподобия при одних и тех же значениях случайной величины X . При этом с увеличением процента цензурирования различия в поведении относительной асимптотической эффективности для рассмотренных цензурирующих распределений становятся более явными.

Источники

Лапина М. Г., Носова Е. В., Фридман Г. М. Сравнение статистических и эвристических методов восстановления общего пассажирского спроса по данным о количестве продаж авиабилетов // Финансы и бизнес. 2011. № 3. С. 142–149.

- Aranda-Ordaz F. J.* Relative Efficiency of the Kaplan-Meier Estimator under Contamination // Communications in Statistics — Simulation and Computations. 1987. Vol. 16. P. 987–997.
- Cramer H.* Mathematical Methods of Statistics. Princeton. N. J., 1946.
- Greenwood M.* The Natural Duration of Cancer. Reports on Public Health and Medical Subjects 33. Her Majesty's Stationery Office, London. 1926. P. 1–26.
- Kaplan E. L., Meier P.* Nonparametric Estimation From Incomplete Observations // Journal of the American Statistical Association. 1958. Vol. 53. P. 457–481.
- Klein J. P., Moeschberger M. L.* The Robustness of Several Estimators of the Survivorship Function with Randomly Censored Data // Communications in Statistics — Simulation and Computations. 1989. Vol. 18. P. 1087–1112.
- Lapina M., Fridman G.* A New Single-class Demand Unconstraining Method // Proceedings of the 31th International Conference “Mathematical Methods in Economics 2013”. Brno, 2013. P. 183–189.
- Meier P., Karrison T., Chappell R., Xie H.* The price of Kaplan-Meier // Journal of the American Statistical Association. 2004. Vol. 99. N 467. P. 890–896.
- Miller R. G.* Survival Analysis. N. Y., 1981.
- Miller R. G.* What Price Kaplan – Meier? // Biometrics. 1983. Vol. 39. P. 1077–1081.
- Oehlert G. W.* A note on the Delta method // The American Statistician. 1992. Vol. 46. N 1. P. 27–29.
- Prakasa Rao B. L. S.* Remarks on Cramer-Rao Type Integral for Randomly Censored Data // Analysis of Censored Data, IMS Lecture Notes — Monograph Series / Ed. by H. L. Koul, J. V. Deshpande. Vol. 27. Hayward, (CA), 1995. P. 8–31.
- Rao C. R.* Linear Statistical Inference and its Applications, 2nd ed. N. Y., 1973.
- Talluri K. T., Ryzin van G. J.* The Theory and Practice of Revenue Management. Boston, 2004. P. 474–478, 485–486.