

Исследование свойств оценок максимального правдоподобия параметров распределения Вейбулла по усеченным слева данным¹

Е. В. Чимитова, Е. О. Ермилова

С усеченными слева данными часто приходится иметь дело в задачах анализа надежности, когда отбор в выборку производится при условии, что продолжительность безотказной работы исследуемых устройств не меньше некоторой величины. В данной работе методами статистического моделирования исследуются свойства оценок максимального правдоподобия (ОМП) параметров распределения Вейбулла в зависимости от объема выборки, степени усечения и процента наблюдений из усеченного распределения. Получено выражение для информационного количества Фишера о параметрах распределения Вейбулла, содержащегося в выборке усеченных слева наблюдений. Вычислены потери информации Фишера от усечения слева, на основе которых можно судить о точности ОМП. Полученные в работе результаты могут быть полезны при построении вероятностных моделей надежности по усеченным слева данным.

Ключевые слова: усеченные слева данные, распределение Вейбулла, оценка максимального правдоподобия, информационное количество Фишера.

1. Введение

В задачах статистического анализа данных типа времени жизни, как правило, выборки оказываются цензурированными справа и содержат усеченные слева наблюдения. Исследованию методов оценивания параметров распределений по цензурированным данным посвящено огромное количество публикаций, в частности [1–3]. Однако упоминаний об исследованиях статистических свойств оценок максимального правдоподобия (ОМП) по выборкам, содержащим усеченные слева наблюдения, судя по публикациям, в отечественных изданиях практически не встречается.

Для объяснения понятия усеченных слева случайных величин рассмотрим следующий пример. Пусть имеется генеральная совокупность технических устройств, принятых в эксплуатацию в разные моменты времени. Обозначим через $F(t)$ функцию распределения случайной величины T – продолжительности работы до отказа. Предположим, что регистрация отказов с целью исследования надежности началась в некоторый момент времени t_0 после начала эксплуатации устройств. При этом в выборку попадают только те объекты из этой же генеральной совокупности, которые не отказали до момента времени t_0 .

¹ Работа выполнена при поддержке Министерства образования и науки РФ в рамках проектной части государственного задания (№ 2.541.2014К)

Это означает, что полученные независимые наблюдения о продолжительности наработки до отказа T_1, T_2, \dots, T_n являются усеченными слева случайными величинами. Необходимо отметить, что если рассматривать данную выборку как обычную, значения которой принадлежат распределению $F(t)$, то полученная картина будет отражать слишком оптимистичный результат, поскольку чем больше продолжительность работы устройства, тем больше у него шансов попасть в эту выборку. В то время как объекты с ранними отказами в выборку не попадают.

Обозначим через D_i время с начала эксплуатации i -го объекта до начала эксперимента t_0 . Тогда распределение усеченной случайной величины T_i имеет вид:

$$F_{LT}(t) = \frac{F(t) - F(D_i)}{1 - F(D_i)}, \quad t > D_i. \quad (1)$$

Аналогично, плотность распределения определяется соотношением:

$$f_{LT}(t) = \frac{f(t)}{1 - F(D_i)}, \quad t > D_i. \quad (2)$$

В выборку могут быть включены также устройства, которые начали эксплуатироваться после начала эксперимента t_0 . В этом случае соответствующий элемент выборки представляет собой полное наблюдение, для которого момент усечения $D_i = 0$.

Выборку усеченных слева наблюдений можно представить в следующем виде:

$$X_n = T_1, D_1, T_2, D_2, \dots, T_n, D_n, \quad (3)$$

где n – объём выборки, T_i – время отказа i -го объекта, D_i – время усечения, $i = \overline{1, n}$. В такой выборке содержатся наблюдения с функциями распределения $F(t)$ и $F_{LT}(t)$. По существу мы имеем выборку, принадлежащую смеси распределений $F(t)$ и $F_{LT}(t)$.

На практике выборки отказов вида (3) встречаются довольно редко, так как при испытаниях на надёжность или при анализе выживания результаты наблюдений чаще всего оказываются ещё и цензурированными справа:

$$X_n = T_1, D_1, \delta_1, X_2, D_2, \delta_2, \dots, X_n, D_n, \delta_n, \quad (4)$$

где n – объём выборки, X_i – время отказа или момента цензурирования i -го объекта, D_i – момент усечения, δ_i – индикатор цензурирования.

В связи с широким использованием закона распределения Вейбулла в анализе данных типа времени жизни, будем проводить исследование, ориентируясь на этот закон. Функция распределения Вейбулла имеет вид:

$$F(t; \theta) = 1 - \exp \left[- \left(\frac{t}{\theta_1} \right)^{\theta_2} \right], \quad (5)$$

где $\theta_1 > 0$ – параметр масштаба, $\theta_2 > 0$ – параметр формы.

В последние годы за рубежом появилось множество публикаций, относящихся к проблеме построения вероятностных моделей по усеченным слева и цензурированным справа данным. Обширные исследования по данной тематике приведены в [3–12]. В частности, Балакришнан Н. и Митра Д. разработали EM-алгоритм и исследовали свойства ОМП для цензурированной справа выборки, содержащей усеченные слева наблюдения, из распределения Вейбулла [4] или логнормального распределения [5]. В [6] разработан новый подход к спецификации модели пропорциональных интенсивностей Кокса для выборки усеченных слева наблюдений, который предполагает существование зависимости между временем усечения и временем отказа объекта. Построению вероятностных моделей на

основе данных типа времени жизни в зависимости от объясняющих переменных посвящена работа [7]. При выборе подхода к исследованию надежности объектов необходимо исходить из имеющейся информации о проводимых экспериментах.

Если априорные данные отсутствуют и нельзя сделать предположение о виде вероятностной модели, то для оценивания надежности используют, например, непараметрическую оценку Каплана–Мейера. В [8] описано построение оценки Каплана–Мейера функции надежности цензурированных данных, а в [9] рассматривается непараметрическая оценка для случая, когда выборка содержит усеченные данные. В [10] была рассмотрена непараметрическая оценка Нельсона–Аалена, которая была применена для анализа выживаемости по усеченным слева и цензурированным справа данным.

В вышеперечисленных работах рассматриваются усеченные слева и цензурированные справа данные. Главная особенность настоящей работы заключается в том, что свойства ОМП параметров распределения Вейбулла рассматриваются в зависимости от параметров усечения без учета цензурирования данных намного более подробно. Важно иметь представление о том, как усечение данных влияет на потери информации Фишера, на относительную эффективность и диаграмму рассеивания ОМП параметров распределения Вейбулла.

2. Потери информации Фишера от усечения

Пусть имеется полная выборка времен жизни T_1, T_2, \dots, T_n из распределения $F(t; \theta)$, где $\theta \in \Theta$ – вектор параметров размерности s . Если условия проведения эксперимента таковы, что наблюдению доступны только те отказы, для которых время жизни больше некоторой наперед заданной величины D , называемой временем усечения, то в результате получаем выборку усеченных слева величин $T^{(1)}, D, T^{(2)}, D, \dots, T^{(M)}, D$, объем которой представляет собой случайную величину $M \leq n$, принадлежащую биномиальному распределению $Bi_{n, 1-F(D)}$.

Обозначим через $d = F(D)$ вероятность попадания в усеченную область и будем называть данную величину степенью усечения.

Функция плотности усеченной случайной величины $T^{(1)}, D$ определяется соотношением:

$$f_{LT}(t; \theta) = \frac{f(t; \theta)}{1 - F(D; \theta)}, \quad t > D, \quad (6)$$

где D – время усечения, $F(t; \theta)$ – функция распределения времен отказов.

Тогда информационное количество Фишера в усеченной выборке, полученной из полной выборки T_1, T_2, \dots, T_n , имеет вид:

$$\begin{aligned} I_{LT}(\theta) &= \sum_{m=0}^n P_{M=m} \cdot m \cdot i_{LT}(\theta | M=1) = \\ &= \sum_{m=0}^n C_n^m (1-d)^m d^{n-m} \cdot m \cdot i_{LT}(\theta | M=1) = n(1-d) \cdot i_{LT}(\theta | M=1), \end{aligned} \quad (7)$$

где $i_{LT}(\theta | M=1)$ – информационное количество Фишера о параметре θ в усеченном наблюдении:

$$i_{LT} \theta | M = 1 = \int_D \left(\frac{\partial \ln f_{LT}(t; \theta)}{\partial \theta} \right)^T \frac{\partial \ln f_{LT}(t; \theta)}{\partial \theta} f_{LT}(t; \theta) dt. \quad (8)$$

Поскольку для оценивания s неизвестных параметров требуется как минимум s наблюдений, то выражение (7) примет вид:

$$I_{LT}(\theta) = \frac{n(1-d) \cdot i_{LT} \theta | M = 1}{1 - \sum_{m=0}^{s-1} C_n^m (1-d)^m d^{n-m}}. \quad (9)$$

Понятно, что при больших объемах выборок величиной $\sum_{m=0}^{s-1} C_n^m (1-d)^m d^{n-m}$ в выражении (9) можно пренебречь.

О потерях информации Фишера от усечения слева будем судить по величине $I_{LT}(\theta) / I(\theta)$, где $I(\theta)$ – информационное количество Фишера в полной выборке T_1, T_2, \dots, T_n . В табл. 1 представлены найденные значения $\det I_{LT}(\theta) / \det I(\theta)$ для параметров распределения Вейбулла в зависимости от степени усечения при $n \geq 100$.

Таблица 1. Отношение информационного количества Фишера в выборке усеченных слева наблюдений к информационному количеству в исходной полной выборке

d	О параметре θ_1 распределений экспоненциального, Релея и Вейбулла	О параметре θ_2 распределения Вейбулла	О двух параметрах распределения Вейбулла
0.1	0.9000	0.6595	0.4008
0.2	0.8000	0.6186	0.2211
0.3	0.7000	0.6141	0.1261
0.4	0.6000	0.6122	0.0709
0.5	0.5000	0.5990	0.0380
0.6	0.4000	0.5657	0.0187
0.7	0.3000	0.5051	0.0079
0.8	0.2000	0.4080	0.0025
0.9	0.1000	0.2590	0.0004

Анализируя результаты, представленные в табл. 1, можно отметить, что наиболее существенные потери в информации Фишера от усечения слева наблюдаются в случае оценивания одновременно двух параметров рассматриваемых распределений. Например, в случае распределения Вейбулла при степени усечения $d = 0.5$ выборка содержит 50 % от полной информации Фишера при оценивании только параметра масштаба, примерно 60 % при оценивании только параметра формы и при этом не более 4 % от полной информации при оценивании двух параметров данного распределения.

Проведем исследование методом Монте-Карло точности ОМП параметров распределения отказов по выборке усеченных слева наблюдений в зависимости от объема выборки n и степени усечения. Для этого рассмотрим изменение величины $\det D[\hat{\theta}] / \det D[\hat{\theta}_{LT}]$, где $\hat{\theta}$ – ОМП неизвестного параметра распределения по полной выборке, $\hat{\theta}_{LT}$ – ОМП по выборке усеченных слева наблюдений. Результаты моделирования представлены в табл. 2. Количество N моделируемых выборок, по которым исследовались законы распределения оценок по выборкам объема n , было принято равным 100000. При

исследовании распределений оценок выборки моделировались по закону Вейбулла с параметром масштаба $\theta_1 = 2$ и формы $\theta_2 = 2$.

Таблица 2. Относительная эффективность оценивания параметров распределения Вейбулла в зависимости от объема выборки

О параметре θ_1 распределения Вейбулла						
d	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
0.1	0.9009	0.9004	0.9010	0.8984	0.8985	0.8995
0.3	0.6982	0.6978	0.7019	0.6994	0.6967	0.7044
0.5	0.4986	0.4961	0.4969	0.4975	0.4957	0.5032
0.7	0.2950	0.2964	0.2971	0.2987	0.2971	0.3014
О параметре θ_2 распределения Вейбулла						
d	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
0.1	0.6363	0.6531	0.6496	0.6573	0.6584	0.6546
0.3	0.5519	0.5888	0.5952	0.6053	0.6063	0.6054
0.5	0.5385	0.5747	0.5799	0.5912	0.5919	0.5899
0.7	0.4491	0.4810	0.4882	0.4951	0.4975	0.4981
О двух параметрах распределения Вейбулла						
d	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
0.1	0.3745	0.3964	0.4035	0.4010	0.4009	0.3993
0.3	0.0884	0.1107	0.1137	0.1201	0.1248	0.1225
0.5	0.0152	0.0232	0.0250	0.0292	0.0329	0.0311
0.7	0.0002	0.0021	0.0025	0.0031	0.0039	0.0042

По данным, представленным в табл. 2, видно, что с ростом объема выборок величина относительной эффективности оценивания параметров распределения Вейбулла $\det D[\hat{\theta}] / \det D[\hat{\theta}_{LT}]$ повышается и стремится к асимптотической величине $\det I_{LT}(\theta) / \det I(\theta)$. Как можно заметить, это подтверждается для любой степени усечения при рассмотрении всех случаев: при оценивании только параметра θ_1 , только параметра θ_2 и одновременном оценивании двух параметров. Наблюдаемые отклонения – на уровне погрешности моделирования.

3. Исследование точности ОМП по выборкам усеченных слева наблюдений

Пусть имеется выборка усеченных слева наблюдений вида:

$$X_n = \{(T_1, D_1), (T_2, D_2), \dots, (T_n, D_n)\},$$

где n – объем выборки, T_i – время отказа i -го объекта, D_i – время усечения, $i = \overline{1, n}$. Если $D_i = 0$, то i -е наблюдение является полным.

В выборке могут содержаться как полные наблюдения, так и наблюдения усеченных случайных величин, причем времена усечения могут быть различными. Такого рода выборки обычно являются результатом наблюдения за объектами, начиная с некоторого момента времени t_0 . При этом начало отсчета времени для некоторых объектов (момент начала

эксплуатации) оказывается раньше момента начала наблюдения t_0 (начала исследования). В этом случае наблюдаемые случайные величины являются усеченными слева и время усечения D_i равно разности между t_0 и началом отсчета времени для i -го объекта. Если же отсчет времени начался позже начала наблюдения t_0 , то соответствующее наблюдение является полным.

В силу того, как формируется выборка в задаче анализа надежности, она может представлять собой смесь элементов, принадлежащих усеченным законам вида $F_{LT}(t; \theta)$ с различной степенью усечения, и элементов, принадлежащих $F(t; \theta)$. В частном случае может наблюдаться смесь двух законов вида $F(t; \theta) + \gamma \cdot F_{LT}(t; \theta)$, где $0 < \gamma \leq 1$ задает долю присутствия наблюдений усеченной случайной величины. В решаемых задачах анализа выживаемости и надежности известно, какое наблюдение принадлежит (соответствующему) усеченному, а какое полному закону. Поэтому нет принципиальных проблем с записью функции правдоподобия, а следовательно, и с поиском оценок. Информационное количество Фишера о параметре θ в этом случае представляет собой линейную комбинацию

$$I_\gamma(\theta) = n\gamma \cdot i_{LT}(\theta | M = 1) + n(1 - \gamma) \cdot i(\theta),$$

где $i_{LT}(\theta | M = 1)$ – информационное количество Фишера, содержащееся в одном наблюдении из усеченного распределения, $i(\theta)$ – информационное количество Фишера о параметре θ , содержащееся в одном полном наблюдении.

В табл. 3 представлены значения информационного количества Фишера $\det i_{LT}(\theta | M = 1)$ о параметрах θ_1 или θ_2 в одном наблюдении из усеченного распределения Вейбулла при $\theta_1 = 2$ и $\theta_2 = 2$ для различных степеней усечения. Значения в первой строке таблицы при $d = 0.0$ соответствуют информационному количеству $i(\theta)$ в полном наблюдении.

На основе значений, представленных в табл. 3, можно рассчитать значения информации Фишера $I_\gamma(\theta)$ в выборке из смеси двух законов вида $F(t; \theta) + \gamma \cdot F_{LT}(t; \theta)$ при различных процентах наблюдений из усеченного распределения $\gamma \cdot 100\%$.

Таблица 3. Информационное количество Фишера в одном наблюдении из усеченного распределения Вейбулла с параметрами $\theta_1 = 2$ и $\theta_2 = 2$

d	$i_{LT}(\theta_1 M = 1)$	$i_{LT}(\theta_2 M = 1)$	$\det i_{LT}(\theta M = 1),$ $\theta = \theta_1, \theta_2$
0.0	1.0000	0.4559	0.4112
0.1	1.0000	0.3341	0.2035
0.2	1.0000	0.3526	0.1421
0.3	1.0000	0.4000	0.1059
0.4	1.0000	0.4652	0.0810
0.5	1.0000	0.5462	0.0626
0.6	1.0000	0.6448	0.0481
0.7	1.0000	0.7676	0.0363
0.8	1.0000	0.9301	0.0261
0.9	1.0000	1.1809	0.0168

При $d = 0.0$ представленные в табл. 3 величины соответствуют информационному количеству Фишера в одном наблюдении, принадлежащему исходному распределению

Вейбулла $F(t; \theta)$. Важно отметить, что при увеличении степени усечения информация Фишера о параметре θ_1 остается неизменной, о параметре θ_2 увеличивается, а о векторном параметре – уменьшается.

Информационное количество Фишера определяет нижнюю границу дисперсии несмещенных оценок. ОМП являются асимптотически эффективными, т.е. при $n \rightarrow \infty$ достигается нижняя граница неравенства Рао–Крамера:

$$D[\hat{\theta}] \geq I_{\gamma}^{-1}(\theta).$$

Реальную же картину точности ОМП при ограниченных объемах выборок можно увидеть, оценив величину $I_{\gamma}^{-1}(\theta) \cdot D^{-1}[\hat{\theta}]$, которая при $n \rightarrow \infty$ должна стремиться к 1.

В табл. 4 представлены отношения минимальной дисперсии $I_{\gamma}^{-1}(\theta_2)$ к оценке реальной дисперсии $\hat{D}[\hat{\theta}_2]$ ОМП параметра формы (при известном параметре масштаба) распределения Вейбулла по выборкам из смеси усеченного и неусеченного распределения Вейбулла при различных процентах усеченных наблюдений $\gamma \cdot 100\%$ и при различной степени усечения d .

Таблица 4. Оценки величины $I_{\gamma}^{-1}(\theta_2) \cdot \hat{D}^{-1}[\hat{\theta}_2]$ для ОМП параметра формы распределения Вейбулла по выборкам усеченных слева наблюдений в зависимости от объема выборки n

Степень усечения $d = 0.1$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.949	0.974	0.984	0.987	0.984	0.990
50 %	0.944	0.971	0.984	0.981	0.998	0.991
75 %	0.942	0.977	0.971	0.985	0.994	0.995
100 %	0.931	0.963	0.975	0.983	0.996	0.995
Степень усечения $d = 0.3$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.945	0.968	0.979	0.984	0.989	0.990
50 %	0.925	0.966	0.985	0.980	0.987	0.985
75 %	0.918	0.957	0.972	0.985	0.989	0.990
100 %	0.900	0.949	0.968	0.978	0.987	0.992
Степень усечения $d = 0.5$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.952	0.969	0.980	0.991	0.991	0.999
50 %	0.941	0.970	0.984	0.982	0.996	0.992
75 %	0.942	0.972	0.985	0.983	0.997	0.995
100 %	0.935	0.960	0.974	0.983	0.988	0.995
Степень усечения $d = 0.7$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.952	0.981	0.984	1.001	0.988	0.992
50 %	0.948	0.976	0.985	0.997	1.001	0.997
75 %	0.959	0.983	0.985	0.991	0.992	0.993
100 %	0.961	0.980	0.995	0.987	1.000	1.000

Как видно из табл. 4, с ростом объема выборки оценка дисперсии ОМП стремится к своей асимптотической величине $I_{\gamma}^{-1} \theta_2$. Отметим, что в случае степени усечения $d = 0.1$

и $d = 0.3$, когда информация Фишера $i_{LT}(\theta_2 | M = 1)$ в одном наблюдении по усеченному закону меньше, чем по полному (см. табл. 3), точность ОМП параметра формы падает с ростом количества усеченных наблюдений в выборке. Этого не наблюдается при $d = 0.5$ и $d = 0.7$, когда информационное количество Фишера по усеченному закону больше, чем по полному.

Необходимо отметить, что в случае оценивания параметра масштаба распределения Вейбулла (при известном параметре формы) оценки отношения $I_\gamma^{-1}(\theta_1) \cdot \hat{D}^{-1}[\hat{\theta}_1]$ по выборкам усеченных слева наблюдений оказываются близкими к 1 независимо от степени усечения и процента наблюдений из усеченного распределения, поскольку информационное количество Фишера о параметре масштаба в наблюдении усеченной случайной величины $i_{LT}(\theta | M = 1)$ совпадает с информацией в полном наблюдении.

В табл. 5 представлены отношения $\det I_\gamma^{-1}(\theta)$ к $\det \hat{D}[\hat{\theta}]$ для ОМП векторного параметра распределения Вейбулла. В данном случае информационное количество Фишера в одном наблюдении усеченного распределения значительно меньше, чем в полном наблюдении, и уменьшается с ростом степени усечения. Поэтому для всех рассмотренных значений степени усечения точность ОМП параметров по выборке усеченных наблюдений падает с увеличением процента наблюдений из усеченного распределения. Отметим также, что при увеличении величины $\gamma \cdot 100\%$ от 25 % до 75 % уменьшение скорости сходимости $\det \hat{D}[\hat{\theta}]$ к $\det I_\gamma^{-1}(\theta)$ не столь существенно, как при увеличении от 75 % до 100 %.

Таблица 5. Оценки величины $\det I_\gamma^{-1}(\theta) \cdot \det \hat{D}^{-1}[\hat{\theta}]$ для ОМП параметров масштаба и формы распределения Вейбулла по усеченным слева выборкам в зависимости от объема выборки n

Степень усечения $d = 0.1$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.932	0.972	0.985	0.989	0.992	0.996
50 %	0.942	0.966	0.975	0.973	0.997	0.998
75 %	0.917	0.963	0.980	0.980	0.981	0.987
100 %	0.909	0.955	0.962	0.970	0.974	0.986
Степень усечения $d = 0.3$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.920	0.970	0.978	0.988	0.987	0.990
50 %	0.908	0.947	0.979	0.971	0.984	0.986
75 %	0.885	0.938	0.960	0.965	0.989	1.004
100 %	0.749	0.868	0.891	0.937	0.957	0.971
Степень усечения $d = 0.5$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.925	0.968	0.971	0.983	0.987	1.030
50 %	0.912	0.946	0.972	0.988	0.986	1.007
75 %	0.873	0.937	0.944	0.983	0.974	0.972
100 %	0.529	0.675	0.745	0.831	0.870	0.893

Степень усечения $d = 0.7$						
$\gamma \cdot 100\%$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
25 %	0.946	0.972	0.978	1.001	1.006	0.981
50 %	0.923	0.968	0.974	0.975	0.976	1.001
75 %	0.868	0.943	0.961	0.965	0.976	1.004
100 %	0.305	0.426	0.519	0.627	0.692	0.774

В значительной мере закономерности в изменении статистических свойств ОМП параметров распределения Вейбулла (в зависимости от степени и процента усеченных слева наблюдений) объясняет картина, представленная на рис. 1, 2. На этих рисунках приведены диаграммы рассеяния ОМП векторного параметра θ , полученные для выборок объемом $n = 100$ при различных значениях степени усечения и процента усеченных наблюдений $\gamma \cdot 100\%$.

Как можно видеть, при увеличении степени и величины $\gamma \cdot 100\%$ эллипсоиды рассеяния ОМП деформируются и становятся все больше ассиметричными относительно осей. Это свидетельствует о том, что закон распределения ОМП параметров масштаба и формы распределения Вейбулла (по выборкам из смеси усеченных и полного закона) существенно отклоняется от многомерного нормального закона и ассиметричен. Причём чем больше степень усечения и процент усеченных наблюдений в выборке, тем сильнее такое отклонение.

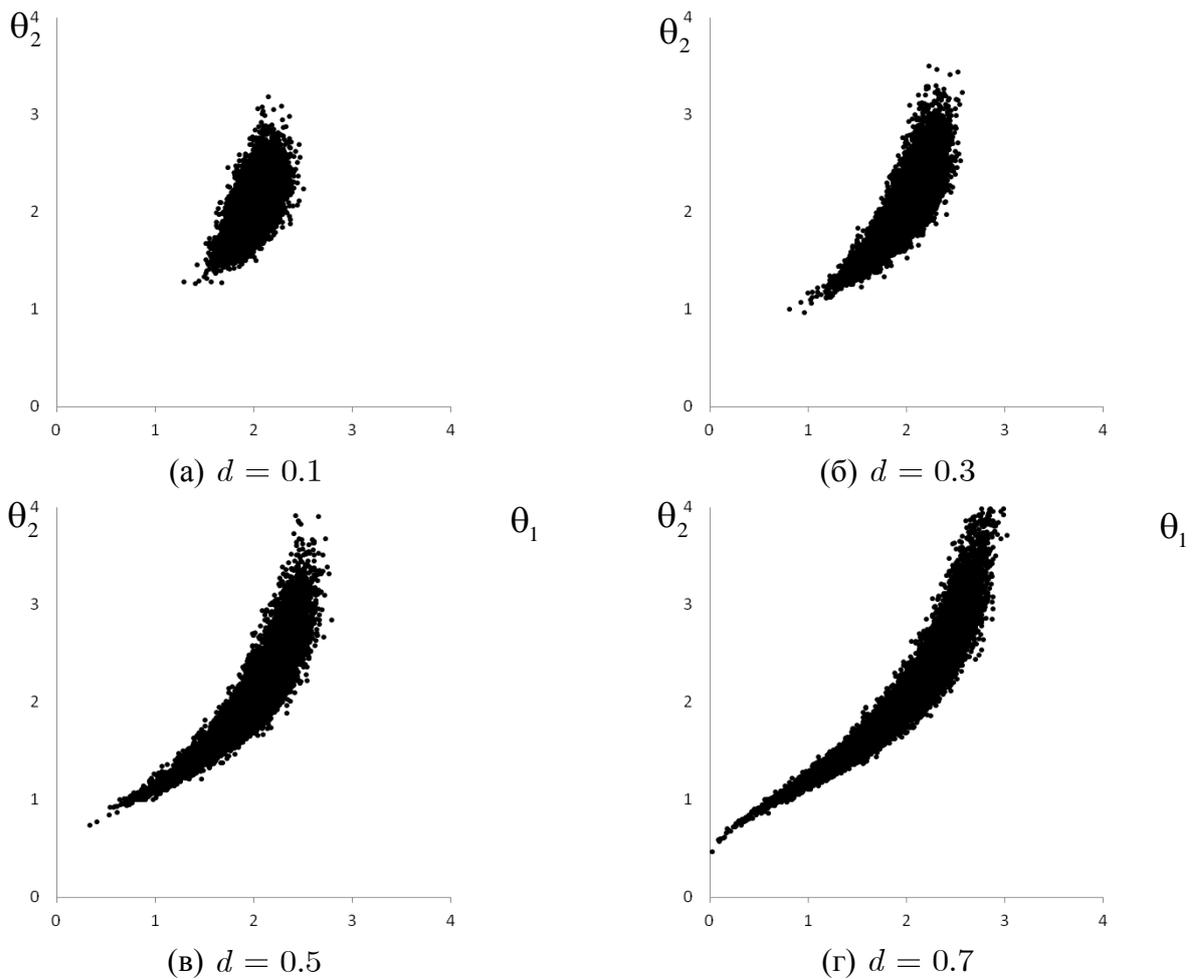


Рис. 1. Диаграммы рассеяния ОМП параметров масштаба и формы распределения Вейбулла θ усеченным выборкам объема $n = 100$, 100 % наблюдений из усеченного распределения Вейбулла

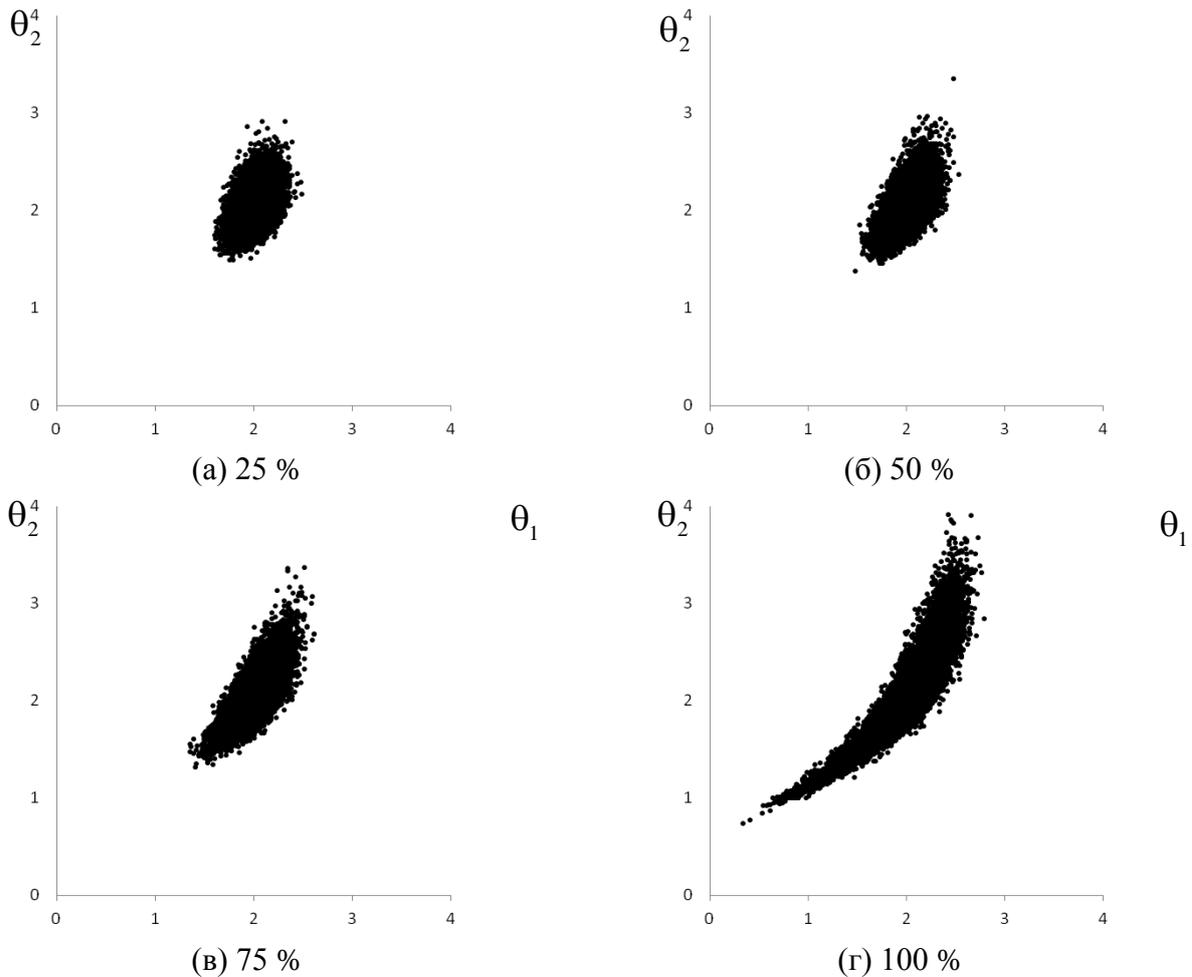


Рис. 2. Диаграммы рассеяния ОМП параметров масштаба и формы распределения Вейбулла θ_1 и θ_2^4 усеченным выборкам объема $n = 100$ при степени усечения $d = 0.5$

На рис. 2 следует обратить внимание на то, как при изменении процента усеченных наблюдений в выборке с 75 % до 100 % резко изменяется форма диаграммы рассеяния ОМП в сторону асимметричности, чего не наблюдается при изменении с 25 % до 75 %.

Рис. 1 и 2 наглядно объясняют результаты, полученные в табл. 5, из которой видно, что при ограниченных объемах выборок и высоком проценте усеченных наблюдений в выборке (более 75 %) значение определителя ковариационной матрицы ОМП векторного параметра распределения Вейбулла $\det \hat{D} \left[\hat{\theta} \right]$ существенно выше, чем нижняя граница неравенства Рао–Крамера $\det I_{\gamma}^{-1} \theta$, что говорит о потере свойства эффективности ОМП в условиях высокой степени усечения.

4. Выводы

В настоящей работе получено выражение для информационного количества Фишера о параметрах распределения Вейбулла, содержащегося в выборке усеченных слева наблюдений. Вычислены потери в информации Фишера от усечения слева, на основе которых можно судить о точности оценок максимального правдоподобия. Так, например,

при оценивании только параметра масштаба при известном значении параметра формы распределения Вейбулла в случае усечения слева не происходит потери точности оценивания с ростом степени усечения. В то время как при оценивании обоих параметров одновременно точность ОМП стремительно падает с ростом степени усечения и процента усеченных наблюдений в выборке.

В результате проведенных исследований методами статистического моделирования показано, что степень усечения и процент усеченных наблюдений в выборке оказывают существенное влияние на распределение ОМП при ограниченных объемах выборок при оценивании двух параметров распределения Вейбулла. В условиях высокого процента усеченных наблюдений в выборке и с ростом степени усечения распределение ОМП существенно отклоняется от многомерного нормального закона, что говорит о потере свойства эффективности ОМП.

Литература

1. *Баталова З. Г.* Анализ точности метода максимального правдоподобия для случайно цензурированных выборок. В кн.: Статистические методы обработки результатов наблюдений при контроле качества и надежности машин и приборов. Л., ЛДНТП, 1979. С. 14–17.
2. *Баскаков В. Н.* Обобщенный метод минимального расстояния для цензурированных данных // Надежность и контроль качества. 1995. № 7. С. 3–7.
3. *Лемешко Б. Ю., Гильдебрант С. Я., Постовалов С. Н.* К оцениванию параметров надежности по цензурированным выборкам // Заводская лаборатория. Диагностика материалов. 2001. Т. 67, № 1. С. 52–64.
4. *Balakrishnan N., Mitra D.* Left truncated and right censored Weibull data and likelihood inference with an illustration // Computational Statistics and Data Analysis. 2012. V. 56, № 12. P. 4011–4012.
5. *Balakrishnan N., Mitra D.* Likelihood inference for lognormal data with left truncation and right censoring with an illustration // Journal of Statistical Planning and Inference. 2011. V. 141, № 11. P. 3536–3553.
6. *Li J.* Cox Model Analysis with the Dependently Left // Mathematics Theses. Georgia State University, Atlanta, GA, 2010. Paper 88.
7. *Su Y.-R., Wang J.-L.* Modeling left-truncated and right-censored survival data with longitudinal covariates // The Annals of Statistics. 2012. V. 40, № 3. P. 1465–1488.
8. *Кокс Д. Р., Оукс Д.* Анализ данных типа времени жизни / пер. с англ. О.В. Селезнева. М: Финансы и статистика, 1988. 191 с.
9. *Bagdonavicius V., Kruopis J., Nikulin M.S.* Nonparametric Tests for Censored Data. London, Wiley (ISTE – 2010). P. 233.
10. *Pan W. A., Chappell R.* Nonparametric Estimator of Survival Functions // Lifetime Data Analysis. Boston. 1998. № 4. P. 187–202.
11. *Cohen A. C.* Truncated and censored samples: theory and applications. New York: Marcel Dekker, 1991. 328 p.
12. *Balakrishnan N., Cohen A.C.* Order Statistics and Inference: Estimation. Boston: Academic Press, 1991.

Статья поступила в редакцию 27.11.2015;
переработанный вариант – 10.03.2016.

Чимитова Екатерина Владимировна

к.т.н., доцент кафедры теоретической и прикладной информатики НГТУ,
тел. (383) 346-06-00, e-mail: chimitova@corp.nstu.ru.

Ермилова Елена Олеговна

студент 2-го курса магистратуры факультета прикладной математики и информатики
НГТУ, e-mail: ermilovaeo@gmail.com.

The research of properties of maximum likelihood estimates of Weibull distribution parameters by left truncated data

E. Chimitova, E. Ermilova

Left truncated data often arise in problems of reliability analysis, when sampling is carried out under condition that the lifetime is not less than a certain value. This paper is devoted to the research of statistical properties of maximum likelihood estimates (MLE) of Weibull distribution parameters depending on the sample size, truncation level, and the number of truncated observations in a sample. The research is carried out by means of computer simulation technique. The statement of the Fisher information contained in a sample of left truncated observations for distribution parameters has been obtained. The Fisher information losses from truncation and relative efficiency of MLE have been estimated. The results obtained in this paper can be useful for the problems of probabilistic reliability models construction by left truncated data.

Keywords: left truncated data, Weibull distribution, maximum likelihood estimators, Fisher information.