Аннотация

Объем работы – 96 страниц, состоит из введения, 5 глав, 21 рисунка и 27 таблиц, заключения, списка литературы и приложения.

Объект исследования – вероятностная модель, основанная на усеченных слева и цензурированных справа данных.

Цель работы – исследование методов математической статистики для анализа цензурированных справа выборок усеченных слева наблюдений.

В результате работы были проведены исследования оценок максимального правдоподобия по цензурированным справа выборкам усеченных слева наблюдений в зависимости от степени усечения и цензурирования, процента усеченных наблюдений в выборке и её объема. В том числе было проведено исследование потери информации Фишера от усечения. Исследованы распределения статистик и мощности критериев согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга. Показано влияние усеченности данных на свойства непараметрической оценки Каплана-Мейера.

Практическая ценность и реализация результатов заключается в разработке программного модуля для системы статистического анализа данных типа времени жизни LiTiS, позволяющему моделировать цензурированную справа выборку усеченных слева наблюдений, вычисления оценок максимального правдоподобия, построения оценки Каплана-Мейера и осуществлять проверку простых и сложных гипотез о согласии с помощью непараметрических критериев для усеченных слева и цензурированных справа данных.

Содержание

Введение

Современное состояние и актуальность темы исследования. (2 стр)

Статистические методы широко используются во многих сферах деятельности, в которых возникает необходимость в обработке, анализе и прогнозировании данных типа времени жизни или работы, таких как медицина, финансовая сфера, демография, промышленность, контроль качества, маркетинг и другие.

Так как область применения математической статистики довольно широка, то и формы представления и регистрации данных, с которыми приходится сталкиваться разнообразны. Например, в рамках некоторого эксперимента, направленного на исследование безотказности работы, некоторая часть объектов продолжает функционировать после окончания наблюдения, в таком случае, приходится прибегать к цензурированию данных, храня лишь неполную информацию о времени работы. Такого рода данные часто используются при оценке и контроле надежности технических устройств.

При работе с данными приходится сталкиваться и с ситуацией, когда объект попал под наблюдение с какой-то наработкой. Такие случаи широко распространены, потому что часто моменты начала работы объектов и момент начала эксперимента – это разные моменты времени. Время между началом эксплуатации и началом наблюдения будем называть временем усечения, а данные – усеченными слева.

В последние годы за рубежом появилось множество публикаций, относящихс
частности, Балакришнан Н. и Митра Д. разработали ЕМ-алгоритм и исследовали свойства ОМП для цензурированной справа выборки, содержащей усеченные слева наблюдения, из распределения Вейбулла [6] или логнормального распределения [7].

В [8] разработан новый подход к спецификации модели пропорциональных
Построению вероятностных моделей на основе данных типа времени жизни в зависимости от объясняющих переменных посвящена работа [4]. При выборе

подхода к исследованию надежности объектов необходимо исходить из имеющейся информации о проводимых экспериментах.

Зачастую может возникнуть ложное мнение, что между усечением и цензурированием нет разницы. В своей работе […] Манделл ставит цель показать читателю различие в вероятностых моделях, построенных на основе усеченных слева данных и цензурированных справа данных.

Если априорные данные отсутствуют и нельзя сделать предположение о виде вероятностной модели, то для оценивания надежности используют, например, непараметрическую оценку Каплана-Мейера. В [10] описано построение оценки Каплана-Мейера функции надежности цензурированных данных, а в [11] рассматривается непараметрическая оценка для случая, когда выборка содержит усеченные данные. В [9] была рассмотрена непараметрическая оценка Нельсона-Аалена, которая была применена для анализа выживаемости по усеченным слева и цензурированным справа данным.

Основной задачей анализа надежности объектов является предсказание времени безотказной работы устройств. Например, в статье … авторы проводят статистический анализ на реальных данных о сроке службы трансформаторов высокого напряжения, расположенных на территории США в период с 1980 года. Для прогнозирования остаточного ресурса силовых трансформаторов используется параметрическая модель Кокса (?). Полученные результаты могут быть использованы для планирования расходов на техническое обслуживание и капитальный ремонт оборудования.

Вместе с тем подробные исследования влияния усеченности данных на статистические свойства оценок максимального правдоподобия параметров распределений не проводились.

Цель и задачи исследований. Целью данной магистерской работы является разработка математического и алгоритмического обеспечения для вычисления оценок максимального правдоподобия, статистик критериев согласия и построения оценки Каплана-Мейера для усеченных слева и

цензурированных справа данных. Для достижения цели были поставлены и решены следующие задачи:

- исследование статистических свойств оценок максимального правдоподобия параметров распределении …;
- исследование потери информации Фишера от усечения;
- исследование оценок Каплана-Мейера для цензурированных справа выборок усеченных слева наблюдений;
- исследование распределений статистик и мощности критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для усеченных слева и цензурированных справа данных;
- исследование мощности критериев согласия по цензурированным справа выборок усеченных слева наблюдений;
- разработка программного обеспечения для анализа данных о выживаемости по усеченным слева и цензурированным справа данным.

Методы исследования. Для решения поставленных задач использовались методы статистического анализа, теории вероятности, математической статистики и компьютерного моделирования.

Научная новизна диссертационной работы заключается в следующем:

- в результатах исследований ОМП параметров распределений по цензурированным справа выборкам усеченных слева наблюдений. Показано, как влияет на точность ОМП количество оцениваемых параметров распределения, вид распределения, степень усечения, процент усеченных наблюдений;
- разработана методика вычисления информационного количества Фишера в выборке усеченных наблюдений;
- показано, как влияет на изменение потери информации Фишера по выборкам усеченных наблюдений, полученных из полных выборок;

– в результатах исследования влияния степени усечения и процента наблюдений в выборке из усечённого закона распределения для оценки Каплана-Мейера;
– разработана методика проверки простых и сложных гипотез о виде распределения по выборкам, содержащим усеченные слева наблюдения;
– в результатах исследования критериев Колмогорова, Андерсона-Дарлинга, Крамера-Мизеса-Смирнова.

Основные положения, выносимые на защиту. На защиту выносятся следующие результаты:

– результаты исследования статистических свойств оценок максимального правдоподобия параметров распределений по цензурированных справа выборкам усеченных слева наблюдений;
– результаты исследования статистических свойств оценки Каплана-Мейера для усеченных слева и цензурированных справа данных;
– результаты исследования распределений статистик и мощностей статистик критериев согласия;
– алгоритмы моделирования цензурированных справа выборок усеченных слева наблюдений.
– ,,,

Достоверность и обоснованность научных положений, рекомендаций и выводов подтверждается:

– корректной работой разработанных методов, достоверностью описанных выводов и адекватностью полученных результатов;
– равенство полученных результатов моделирования и теоретических результатов.

Личный творческий вклад автора заключается:

– в описании алгоритма моделирования цензурированных выборок усеченных слева наблюдений при разных значения степени усечения и цензурирования, процента усеченных наблюдений;
–
– в вычислении потери информации Фишера от усечения слева;
– в формулировании этапов исследования распределений статистик непараметрических критериев согласия по цензурированным справа выборкам усеченным слева наблюдений;

Практическая ценность и реализация результатов работы. Полученные в работе результаты могут быть использованы в прикладных задачах статистического анализа усеченных слева и цензурированных данных. Сформулированы рекомендации по корректному применению критериев Колмогорова, Крамера-Мизеса-Смирнова и Андресона-Дарлинга для цензурированных справа выборок усеченных слева наблюдений. Разработанное программное обеспечение может применяться в теории надёжности, выживаемости, медицине, экономике для прикладных задач, связанных с усеченными слева и цензурированными справа данными.

Апробация работы. Результаты исследования докладывались на всероссийской научной конференции молодых ученых “Наука. Технология. Инновации”, Новосибирск, 2014г.; Российской научно-технической конференции "Обработка информационных сигналов и математическое моделирование", Новосибирск, 2015г.

Публикации. Основные результаты исследований по теме диссертации опубликованы в 2 печатных работах. Принята к печати статья в рецензируемый журнал «Вестник СибГУТИ». ССЫЛКИ

Структура работы. Диссертация состоит из введения, 5 глав основного содержания, заключения, списка литературы и 3 приложений. Основная честь

содержания изложена на … страницах, включая … рисунков, … таблиц и списка литературы из … источников.

Краткое содержание работы. В первой главе представлены основные понятия и определения, используемые в работе. Формулируются задачи в соответствии с поставленными целями.

Во второй главе исследуются свойства оценок, полученных методом максимального правдоподобия, в зависимости от количества усеченных или цензурированных наблюдений в выборке и её полного объема. Проводится изучение влияния усечения на информационное количество в выборках ограниченного объема. Проведен анализ свойств эффективности и состоятельности оценок максимального правдоподобия для цензурированных данных.

В третьей главе проведено исследование точности и скорости сходимости непараметрических оценок Каплана-Мейера по цензурированным справа выборкам усеченных слева наблюдений.

В четвертой главе были исследованы мощности и распределения статистик критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для усеченных слева и цензурированных справа данных в зависимости от процента усеченных и цензурированных наблюдений в выборке и глубины усечения.

В пятой главе описаны и протестированы разработанные программы для вычисления функции правдоподобия, оценки Каплана-Мейера, вычисления статистики критериев согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для цензурированных справа выборок усеченных слева наблюдений. Приведены примеры анализа выживаемости на реальных данных.

В заключении приводятся основные результаты и выводы.

1. Методы статистического анализа цензурированных справа выборок усеченных слева наблюдений

1.1. Основные понятия

В ситуации, когда момент начала эксперимента наступил позже начала времени жизни объекта, необходимо учитывать, что данные являются усеченными слева.

Когда объект остается рабочим после окончания эксперимента, то есть отказ произошел после некоторого определенного момента времени, в этом случае, данные будут называться цензурированными справа. Заметим, что такая выборка будет содержать меньше информации, чем полная.

Выделяют три вида цензурирования. Первым типом цензурирования будем называть вероятность попадания в интервал цензурирования, то есть фиксируется момент цензурирования. Под вторым типом цензурирования будем понимать отношение количества цензурированных наблюдений к полному объему выборки. Существует и еще один тип цензурирования – случайный или третий тип, в этом случае моменты отказов и моменты цензурирования C_i являются независимыми случайными величинами, а значение времени жизни выбирается по принципу $X_i = \min(T_i, C_i)$.

Таким образом, данные можно представить в виде выборки:

$$(X_1, \tau_1, \delta_1), (X_2, \tau_2, \delta_2), ..., (X_n, \tau_n, \delta_n),$$

где n – объем выборки, X_i – время жизни или момент цензурирования i-го объекта, τ_i – время усечения i-го объекта, δ_i – индикатор цензурирования i-го объекта, который принимает значения 0, если наблюдение цензурированное, и 1, если наблюдение полное, $i = \overline{1,n}$.

Введем основные понятия, которые будем использовать для описания данных. Под временем жизни понимается непрерывная неотрицательная случайная величина ξ с функцией распределения $F(t) = P\{t > \xi\}$, $t \geq 0$.

Усечением слева называется операция, когда время жизни преобразуется в зависимости от параметра усечения, то есть функция распределения для усеченных слева данных определяется через условную функцию распределения и принимает вид:

$$F^{LT}(t) = F(t \mid t > \tau) = \frac{F(t) - F(\tau)}{1 - F(\tau)}, t > \tau, \quad (1)$$

где τ – момент усечения, $F(\tau)$ – функция распределения времен отказов.

Значение неусеченной функции распределения в момент усечения будем называть степенью усечения, а степенью цензурирования будем называть процент цензурированных наблюдений в выборке.

По аналогии с функцией распределения, плотность распределения определяется соотношением:

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

Функция надежности – функция, описывающая вероятность получить отказ после определенного момента времени t:

$$S(t) = P\{\xi > t\},\ t \geq 0. \quad (3)$$

Также функция надежности может быть записана через функцию распределения:

$$S(t) = 1 - F(t). \quad (4)$$

Для усеченных слева данных функция надежности выглядит:

$$S^{LT}(t) = \frac{S(t)}{1 - F(\tau)}, t > \tau. \quad (5)$$

Функция интенсивности – функция, которая описывает вероятность отказа в течение малого промежутка времени при условии, что до этого момента отказа не произошло. Функция может быть записана в виде:

$$\lambda^{LT}(t) = \lambda(t) = \frac{f(t)}{(1 - F(t))} = \frac{f(t)}{S(t)}, t > \tau. \quad (6)$$

Функция риска – в общем виде определяется соотношением:

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau = -\ln\big(S(t)\big). \tag{7}$$

В случае усеченных слева данных функция риска определяется следующим образом:

$$\Lambda^{LT}(t) = \Lambda(t) - \Lambda(\tau),\ t > \tau\,. \tag{8}$$

1.2. Метод максимального правдоподобия

Оценивание неизвестных параметров будем проводить, используя метод максимального правдоподобия. Для этого введем функцию максимального правдоподобия:

$$L(\mathbb{X}_n, \theta) = \prod_{i=1}^{n} f(X_i, \theta), \tag{9}$$

где $\mathbb{X}_n = (X_1, ..., X_n)$ – выборка; θ – параметр, оценку которого необходимо найти; $f(x_i, \theta)$ – в случае непрерывного распределения – это плотность вероятности случайной величины x_i, а в дискретном случае — как вероятность.

Для данных, усеченных слева и цензурированных справа, функция максимального правдоподобия выражается соотношением:

$$\begin{aligned} L(\mathbb{X}_n, \theta) = &\prod_{i \in S_1} \big\{ f(X_i, \theta) \big\}^{\delta_i} \big\{ 1 - F(X_i, \theta) \big\}^{1-\delta_i} \times \\ &\times \prod_{i \in S_2} \left\{ \frac{f(X_i, \theta)}{1 - F(\tau_i, \theta)} \right\}^{\delta_i} \left\{ \frac{1 - F(X_i, \theta)}{1 - F(\tau_i, \theta)} \right\}^{1-\delta_i}, \end{aligned} \tag{10}$$

где $F(X_i, \theta)$ – функция распределения; $f(X_i, \theta)$ – плотность; δ_i – индикатор цензурированных данных; S_1 – множество индексов i, для которых случайная величина X_i является неусеченной; S_2 – множество индексов i, для которых случайная величина X_i является усеченной.

Оценкой максимального правдоподобия (ОМП) параметра называется точка параметрического множества Θ, в которой функция максимального правдоподобия $L(\mathbb{X}_n, \theta)$ достигает максимума:

$$L(\mathbb{X}_n, \hat{\theta}) = \sup_{\theta \in \Theta} L(\mathbb{X}_n, \theta).$$

Прологарифмируем функцию правдоподобия:

$$\ln L(\mathbb{X}_n, \theta) = \sum_{i \in S_1} \left[\delta_i \ln f(X_i, \theta) + (1 - \delta_i) \ln\left(1 - F(X_i, \theta)\right) \right] + \\ + \sum_{i \in S_1} \left[\delta_i \ln\left(\frac{f(X_i, \theta)}{1 - F(\tau_i, \theta)} \right) + (1 - \delta_i) \ln\left(\frac{1 - F(X_i, \theta)}{1 - F(\tau_i, \theta)} \right) \right]. \quad (11)$$

Если для любой выборки $\mathbb{X}_n$ из выборочного пространства X максимум $L(\mathbb{X}_n, \theta)$ достигается во внутренней точке θ и $L(\mathbb{X}_n, \theta)$ дифференцируема θ, то ОМП θ удовлетворяет уравнению $\frac{\partial \ln L(\mathbb{X}_n, \theta)}{\partial \theta_i} = 0, i = 1, ..., r,$ которое называется уравнением правдоподобия. Решая систему, составленную из уравнений максимального правдоподобия, находят оценку неизвестных параметров.

1.2.1. Свойства оценок максимального правдоподобия

В исследовании будем проверять следующие свойства оценки максимального правдоподобия:

- Асимптотическая несмещенность.

ОМП является асимптотически несмещенной, то есть $M\theta \to \theta, n \to \infty$.

- Состоятельность.

Оценка $T_n(\mathbb{X}_n)$ некоторой функции $\tau(\theta)$ называется состоятельной, если при $n \to \infty, T_n \xrightarrow{P} \tau(\theta), \forall \theta \in \Theta$.

- Эффективность.

Для любой несмещенной оценки $T(\mathbb{X}_n)$ параметрической функции $\tau(\theta)$ справедливо неравенство Рао-Крамера:

$$D\left[T(\mathbb{X}_n)\right] \geq \frac{\left[\tau'(\theta)\right]}{ni(\theta)} \quad , \quad (12)$$

где $i(\theta)$ – количество информации Фишера, содержащейся в одном наблюдении [11].

Оценка, при которой достигается нижняя граница неравенства Рао-Крамера, называется эффективной.

- Асимптотическая нормальность.

Оценка θ называется асимптотически нормальной, если $F_{\theta}(t) \to \Phi(t)$, где $\Phi(t)$ – функция распределения нормального закона [12].

Оценки максимального правдоподобия входят в класс наилучших асимптотически нормальных оценок. Однако при ограниченных объемах выборок и значительной степени цензурирования законы распределения ОМП весьма далеки от асимптотически нормального и, более того, оказываются асимметричными, а сами оценки смещенными. В следующей главе будут проведены исследования свойств оценок на различных объемах выборки. [13, 14].

1.2.2. Функция правдоподобия в случае распределения Вейбулла для усеченных слева и цензурированных справа данных

Предположим, что величина t распределена по закону Вейбулла. Рассмотрим подробнее случай нахождения оценки методом максимального правдоподобия, когда значения времени жизни подчиняются закону распределения Вейбулла, который выглядит следующим образом:

$$F(t,\theta) = 1 - \exp\left(-\left(\frac{t}{\theta_0}\right)^{\theta_1}\right), \tag{13}$$

где θ_0 – параметр масштаба, θ_1 – параметр формы.

Плотность распределения Вейбулла и функция надежности, соответственно, выражаются как:

$$f(t,\theta) = \frac{\theta_1}{\theta_0} \cdot \left(\frac{t}{\theta_0}\right)^{\theta_1 - 1} \exp\left(-\left(\frac{t}{\theta_0}\right)^{\theta_1}\right), \tag{14}$$

$$S(t,\theta) = \exp\left(-\left(\frac{t}{\theta_0}\right)^{\theta_1}\right). \tag{15}$$

Функция правдоподобия:

$$L(\mathbb{X}_n,\theta)=\prod_{i\in S_1}\{f(X_i,\theta)\}^{\delta_i}\{S(X_i,\theta)\}^{1-\delta_i}\times\prod_{i\in S_2}\left\{\frac{f(X_i,\theta)}{S(\tau_i,\theta)}\right\}^{\delta_i}\left\{\frac{S(X_i,\theta)}{S(\tau_i,\theta)}\right\}^{1-\delta_i}=$$

$$=\prod_{i\in S_1}\left[\left\{\frac{\theta_1}{\theta_0}\cdot\left(\frac{X_i}{\theta_0}\right)^{\theta_1-1}\right\}^{\delta_i}\cdot\exp\left(-\left(\frac{X_i}{\theta_0}\right)^{\theta_1}\right)\right]\times \qquad (16)$$

$$\times\prod_{i\in S_2}\left[\left\{\frac{\theta_1}{\theta_0}\cdot\left(\frac{X_i}{\theta_0}\right)^{\theta_1-1}\right\}^{\delta_i}\cdot\exp\left(\left(\frac{1}{\theta_0}\right)^{\theta_1}\left(\tau_i^{\theta_1}-X_i^{\theta_1}\right)\right)\right].$$

Логарифм функции правдоподобия:

$$\ln L(\mathbb{X}_n,\theta)=\sum_{i\in S_1}\left[\delta_i\cdot\ln\left(\frac{\theta_1}{\theta_0}\cdot\left(\frac{X_i}{\theta_0}\right)^{\theta_1-1}\right)-\left(\frac{X_i}{\theta_0}\right)^{\theta_1}\right]+$$

$$+\sum_{i\in S_2}\left[\delta_i\cdot\ln\left(\frac{\theta_1}{\theta_0}\cdot\left(\frac{X_i}{\theta_0}\right)^{\theta_1-1}\right)+\left(\frac{1}{\theta_0}\right)^{\theta_1}\left(\tau_i^{\theta_1}-X_i^{\theta_1}\right)\right]. \qquad (17)$$

Таким образом, получен логарифм функции правдоподобия, соответствующий распределению Вейбулла.

1.3. Непараметрическая оценка Каплана-Мейера

В случае если нет априорной информации или нельзя сделать предположение о виде распределения случайной величины, возникает необходимость использовать непараметрический подход.

Одним из примеров такой оценки является оценка Каплана-Мейера, которая также называется множительной оценкой. Далее рассмотрим, как она может быть представлена для усеченных слева и цензурированных справа данных.

Обозначим, через $a_1<a_2<\ldots<a_k=T$, $k\le n$, моменты времени, в которые были зафиксированные полные события, где T – время последнего полного наблюдения. Тогда оценку Каплана-Мейера для усеченных данных можно вычислить по формуле [5]:

$$\hat{F}_n(t) = \begin{cases} 0, t < a_1 \\ 1 - \prod_{j:a_j \le x} \left(1 - \frac{d_j}{r_j}\right), t \ge a_1 \end{cases}, \quad (18)$$

где r_j – число объектов, наблюдаемых в момент t, таких что время начала их эксплуатации меньше t, d_j – число объектов, отказавших в момент t.

В качестве расстояния между множительной оценкой и теоретическим законом будем использовать статистику Колмогорова [15]:

$$D_n = \sup_{t<\infty} \left| \hat{F}_n(t) - F(t) \right|, \quad (19)$$

где $\hat{F}_n(t)$ – оценка Каплана-Мейера, $F(t)$ – теоретическая функция распределения, n – объём выборки.

1.4. Моделирование псевдослучайных величин

Моделировать данные, необходимые для анализа, будем с помощью метода Монте-Карло.

Метод Монте-Карло – это общее название группы методов для решения различных задач с помощью случайных последовательностей.

Идея метода заключается в следующем. Вместо того чтобы описывать исследуемый случайный процесс аналитически, составляется алгоритм, имитирующий этот процесс. В алгоритм включаются специальные процедуры для моделирования случайности. Конкретные вычисления в соответствии с алгоритмом складываются каждый раз по-иному, со своими результатами. Множество реализаций алгоритма используется как некий искусственно полученный статистический материал, обработав который методами математической статистики, можно получить любые характеристики: вероятности событий, математические ожидания, дисперсии случайных величин и т.п.

Как правило, программа составляется для осуществления одного случайного испытания. Затем это испытание повторяется N раз, причем каждый опыт не зависит от остальных, и результаты всех опытов усредняются [16].

Для моделирования выборок случайных величин воспользуемся *методом обратной функции*.

Стандартный метод моделирования случайной непрерывной величины (метод обратной функции) – преобразование вида $\xi = \varphi(\alpha)$, где $\varphi(t)$ – строго непрерывная и монотонная на отрезке функция, α – случайная величина; для ξ задана плотность распределения $f_\xi(t), a \le t \le b$ (границы a и b могут быть и бесконечными). Предположим, что $\varphi(t)$ монотонно возрастает, и найдем функцию распределения для $\eta = \varphi(\alpha)$:

$$F_\eta(t) = \int_a^t f_\eta(x)dx = P(\varphi(\alpha) < t) = P(\alpha < \varphi^{-1}(t)) = \varphi^{-1}(t).$$

Отсюда $\varphi(\alpha) = F^{-1}(\alpha)$. С другой стороны $P(F^{-1}(\alpha) < t) = P(\alpha < F(t)) = F(t)$.

Следовательно, в предположении монотонного возрастания $\varphi(t)$ мы получаем единственную моделирующую формулу $\xi = F^{-1}(\alpha)$, которая представляет собой стандартный метод моделирования случайной непрерывной величины [17].

1.5. Критерии согласия

При проверке согласия различают простые и сложные гипотезы. Гипотеза вида $H_0 : F(x) = F(x,\theta)$, где $F(x,\theta)$ – функция распределения вероятностей, с которой проверяется согласие наблюдаемой выборки независимых одинаково распределенных величин, называется *простой*, если θ – известное значение параметра (скалярного или векторного).

Гипотеза вида H_0: $F(x) \in \{F(x,\theta), \theta \in \Omega\}$ называется *сложной*, если в качестве неизвестного параметра θ используется его оценка $\hat{\theta}$, вычисленная по той же выборке, по которой проверяется гипотеза о согласии.

Мощностью критерия называется величина $1-\beta = P\{H_1 \mid H_1\}$, т.е. вероятность принять справедливую альтернативную гипотезу H_1.

В данной работе будут проведены исследования непараметрических критериев Колмогорова, ω^2 Крамера-Мизеса-Смирнова и Ω^2 Андерсона-Дарлинга. [12]

В критерии Колмогорова используют статистику

$$S_m = \frac{\left(6nD_n^+ + 1\right)^2}{9n}, \qquad (20)$$

где $D_n = \max\left(D_n^+, D_n^-\right)$, $D_n^+ = \max\limits_{1\le i\le n}\left\{\frac{i}{n} - F(X_{(i)},\theta)\right\}$, $D_n^- = \max\limits_{1\le i\le n}\left\{F(X_{(i)},\theta) - \frac{i-1}{n}\right\}$, $X_{(i)}$ – i-ый элемент упорядоченной выборки $X_1, X_2, ..., X_n$. При проверке простой гипотезы эта статистика в пределе подчиняется распределению χ^2 с числом степеней свободы, равным 2.

В критерии ω^2 Крамера-Мизеса-Смирнова применяется статистика вида

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F(X_{(i)},\theta) - \frac{2i-1}{2n}\right\}^2, \qquad (21)$$

которая при проверке простой гипотезе в пределе подчиняется закону с функцией распределения $a1(s)$.

Для критерия Ω^2 Андресона-Дарлинга статистика имеет следующий вид

$$S_\Omega = n\Omega_n^2 = -n - 2\sum_{i=1}^{n}\left\{\frac{2i-1}{2n}\ln F(X_{(i)},\theta) + \left(1 - \frac{2i-1}{2n}\right)\ln(1 - F(X_{(i)},\theta))\right\}. \qquad (22)$$

в пределе эта статистика подчиняется закону с функцией распределения $a2(s)$.

Все вышеперечисленные предельные распределения критериев описаны для случая рассмотрения простой гипотезы. В случае сложной гипотезы на распределения статистик критериев согласия значительное влияние оказывают различные изменения свойств оценок параметров распределения.

Для проверки гипотез о согласии по цензурированным выборкам вместо эмпирической функции распределения $F_n(t)$ предлагается использовать оценку функции распределения Каплана-Мейера при вычислении значений статистик [10], [14], [18], [22].

В модифицированном критерии Колмогорова в качестве меры расстояния между эмпирическим и теоретическим законом распределения используется величина (19), а в качестве статистики с поправкой Большева

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{23}$$

где $D_n = \max\left(D_n^+, D_n^-\right)$, $D_n^+ = \max\limits_{1\le i\le n}\left\{\hat{F}_n(t_i) - F(t_i,\theta)\right\}$, $D_n^- = \max\limits_{1\le i\le n}\left\{F(t_i,\theta) - \hat{F}_n\left(t_{i-1}\right)\right\}$.

В модифицированном критерии Крамера-Мизеса-Смирнова применяется статистика вида:

$$S_\omega = \frac{n}{3}\cdot F(t_i,\theta) + n\cdot\sum_{j=1}^{n-1}\left[\hat{F}_n^2\left(t_j\right)\left(F\left(t_{j+1},\theta\right) - F\left(t_j,\theta\right)\right) - \right.$$

$$\left. -\hat{F}_n\left(t_j\right)\left(F^2\left(t_{j+1},\theta\right) - F^2\left(t_j,\theta\right)\right) + \frac{1}{3}\left(F^3\left(t_{j+1},\theta\right) - F^3\left(t_j,\theta\right)\right)\right]. \tag{24}$$

Статистика модифицированного критерия Андерсона-Дарлинга вычисляется следующим образом:

$$S_\Omega = n\cdot\left\{-F\left(t_1,\theta\right) + \sum_{j=1}^{n-1}\left[F\left(t_j,\theta\right) - F\left(t_{j+1},\theta\right) + \hat{F}_n^2\left(t_j\right)\cdot\ln\frac{F\left(t_{j+1},\theta\right)}{F\left(t_j,\theta\right)} - \right.\right. \tag{25}$$

$$\left.\left. -\left(1-\hat{F}_n\left(t_j\right)\right)^2\ln\left\{\frac{1-F\left(t_{j+1},\theta\right)}{1-F\left(t_j,\theta\right)}\right\}\right]\right\}.$$

Проверяемая гипотеза о согласии отвергается при больших значениях статистики. В случае модифицированных критериев аналитических предельных распределений статистики не существуют, поэтому при проверке гипотез о согласии можно основываться на статистики, полученные в результате статистического моделирования [5, 19].

Пусть имеется выборка усеченных слева наблюдений $(X_1,\tau_1),(X_2,\tau_2),...,(X_n,\tau_n)$. Если случайная величина является усеченной, то ее закон распределения будет иметь вид (1). Преобразуем эту выборку следующим образом:

$$U_i = \frac{F(X_i) - F(\tau_i)}{1 - F(\tau_i)},\ i = \overline{1,n}$$

Таким образом, в случае справедливости H_0 полученная выборка $U_1, U_2, ..., U_n$ будет распределена по равномерному закону $U(0,1)$ и задача проверки гипотезы о согласии с законом распределения для выборки усеченных слева наблюдений $(X_1, \tau_1), (X_2, \tau_2), ..., (X_n, \tau_n)$ сводится к проверке гипотезы о согласии с равномерным законом для выборки $U_1, U_2, ..., U_n$.

1.6. Выводы

В данной главе проведен обзор понятий и определений, используемых в работе. Описан метод максимального правдоподобия. Получена функция правдоподобия и её логарифм для случая, когда время жизни подчиняются закону распределения Вейбулла. Дано описание непараметрической оценки Каплана-Мейера для усеченных данных.

Для достижения поставленной цели предусмотрено решение следующих задач:

1) реализация и тестирование программных модулей для моделирования данных, оценивания параметров с помощью метода максимального правдоподобия и построения непараметрической оценки Каплана-Мейера;
2) анализ влияния на статистические свойства полученных оценок максимального правдоподобия:
   - объема выборки;
   - степени усечения;
   - степени цензурирования;
   - процента усеченных наблюдений в выборке;
3) исследование статистических свойств оценки Каплана-Мейера;
4) на примере реальных данных построить параметрическую и непараметрическую оценку функции распределения отказов;

5) исследование распределений статистик и мощности непараметрических критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для проверки простых и сложных гипотеза о согласии по усеченным слева и цензурированным справа данным;
6) разработать программное обеспечение для моделирования цензурированных справа выборок усеченных слева наблюдений, вычисления ОМП и оценки Каплана-Мейера, проверки гипотез о согласии.

2. Исследование свойств оценок максимального правдоподобия

Проведем исследование свойств оценок, полученных методом максимального правдоподобия, в зависимости от количества усеченных или цензурированных наблюдений в выборке и её полного объема.

2.1. Исследование свойств оценок максимального правдоподобия по усеченным слева данным

Выберем значения параметров, которые будем использовать для данного исследования. В качестве распределения продолжительностей жизни возьмем распределение Вейбулла с истинными значениями параметров формы и масштаба $\theta_0 = 2$, $\theta_1 = 2$. Количество неполных наблюдений будем изменять с помощью цензурирования первого типа, то есть степень цензурирования будет варьироваться в зависимости от момента окончания эксперимента. Аналогично, в зависимости от выбора точки усечения процент усеченных данных также будет меняться.

На рисунке 1 представлены графики плотности распределения оценок параметра масштаба распределения Вейбулла при разных процентах усеченных наблюдений в полной выборке.

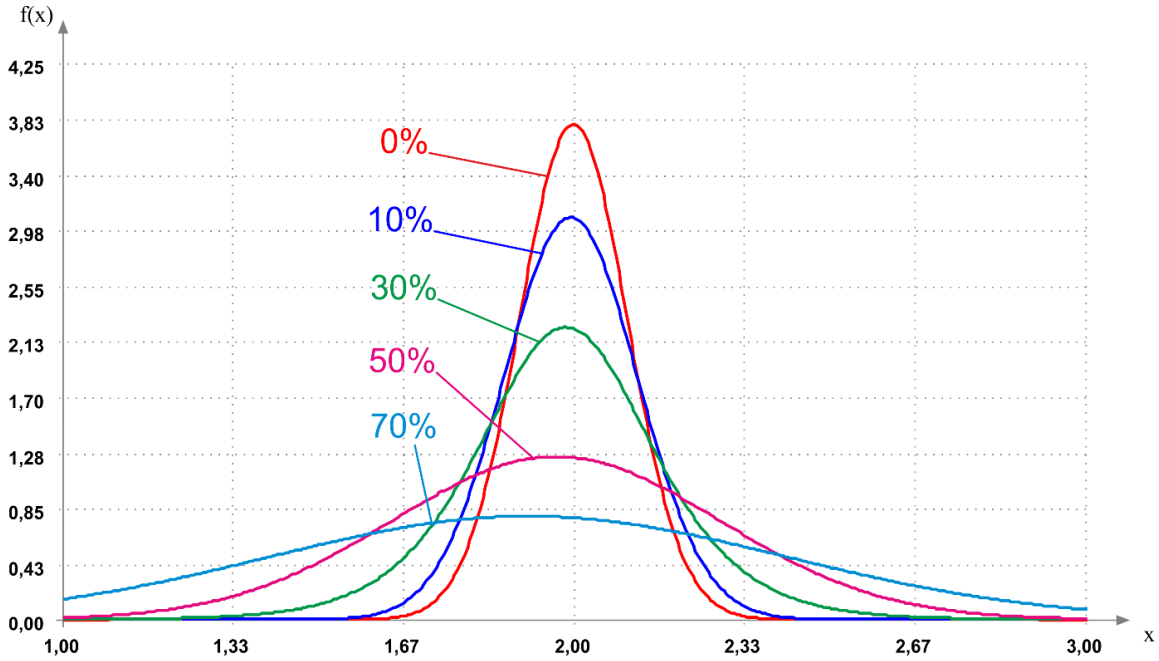


Рисунок 1 – Плотности ОМП параметра масштаба θ_0 при разных процентах усеченных наблюдений в полной выборке объемом $n=100$

Из рисунка 1 видно, что при увеличении количества цензурированных наблюдений смещение относительно истинного значения параметра масштаба и выборочная дисперсия оценки увеличиваются. Таким образом, можно сказать, что полученные ОМП являются смещенными при значительном проценте усеченных данных.

Проведем исследование оценок максимального правдоподобия для выборок с разным процентом усеченных наблюдений. Для этого зафиксируем степень усечения $d=0.3$ и с помощью методов статистического моделирования рассчитаем значения основных статистических показателей для выборок объемом 200, приведенные в таблице 1.

Таблица 1 – Статистические свойства ОМП параметром распределения Вейбулла при степени усечения $d=0.3$. $n=200$

Процент усеч. наблюдений	$M\hat{\theta}-\theta_{ист.}$		$D\hat{\theta}$	
	масштаб	форма	масштаб	Форма
0%	0.00037	0.01371	0.00556	0.01241
10%	-0.00080	0.01408	0.00579	0.01306
30%	0.00027	0.01597	0.00636	0.01491
50%	-0.00170	0.01577	0.00747	0.01811

Продолжение таблицы 1

70%	-0.00335	0.01777	0.00968	0.02355
100%	-0.01564	0.01158	0.02054	0.04964

Как видно из таблицы 1, смещение относительно истинного значения не зависит от количества усеченных наблюдений в выборке, однако, выборочная дисперсия увеличивается с увеличением процента усеченных наблюдений в выборке.

2.2. Исследование свойств оценок максимального правдоподобия по цензурированным данных

С целью исследования свойств оценки максимального правдоподобия на цензурированных справа данных проведем эксперимент при тех же условиях. Рассмотрим случай, когда выборка содержит только неусеченные наблюдения, а степень цензурирования имеет различные значения. На рисунке 2 представлены плотности оценки максимального правдоподобия параметра масштаба при варьировании количества цензурированных данных.

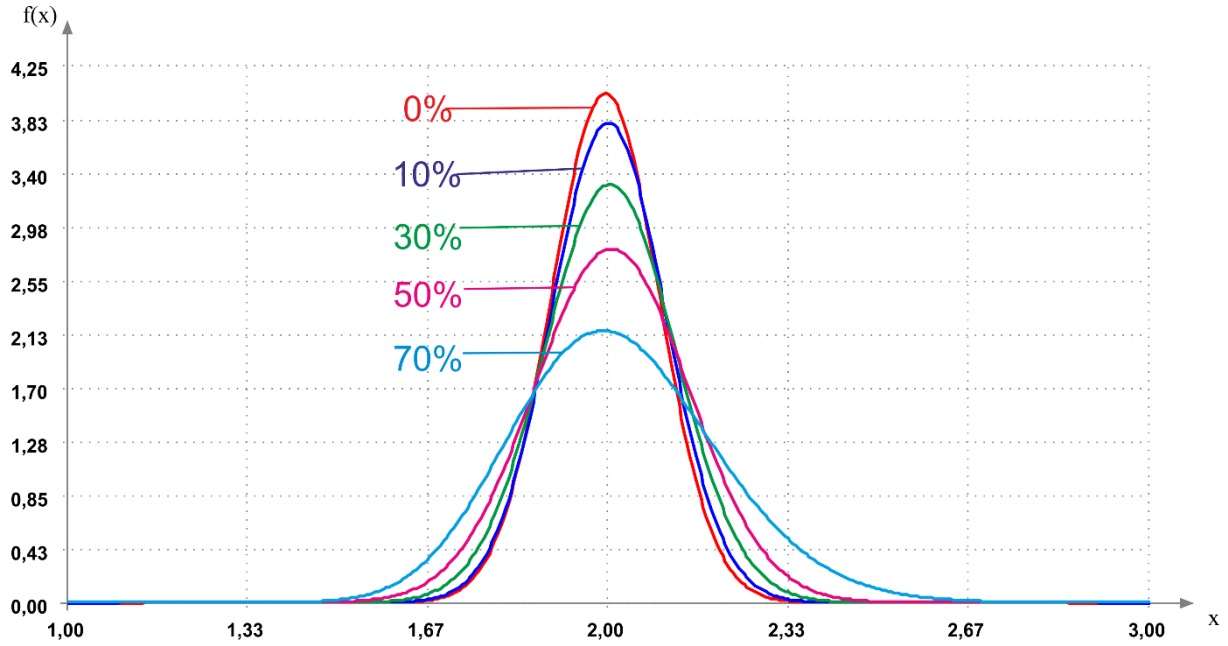


Рисунок 2 – Плотности ОМП параметра масштаба θ_0 для выборки без усеченных данных при разных значениях процента цензурирования. Объем выборки $n=100$

Исходя из данных представленных на рисунке 2, можно сделать вывод о том, что процедура усечения в меньшей степени влияет на точность оценивания, по сравнению с процедурой цензурирования.

Для обобщения исследований смещения цензурированных и усеченных данных в таблицах 2 и 3 приведены значения сдвига, дисперсии, полученных с помощью метода максимального правдоподобия, для параметров сдвига и масштаба на выборках объема 100 и 200 при разной степени усечения и процентах цензурирования.

Таблица 2 – Статистические свойства ОМП параметров распределения Вейбулла. $n=100$

Степень усеч.	Средний процент ценз.	$M\hat{\theta}-\theta_{ист.}$		$D\hat{\theta}$	
		масштаб	форма	масштаб	Форма
0	0%		0.0272	0.0111	0.0255
	10%	0.0013	0.0238	0.0117	0.0314
	30%	0.0034	0.0229	0.0150	0.0460
	50%	0.0119	0.0386	0.0278	0.0734
	70%	0.0383	0.0553	0.0946	0.1408
0.1	0%	-0.0056	0.0360	0.0169	0.0527
	10%	-0.0101	0.0228	0.0177	0.0732
	30%	-0.0153	0.0220	0.0199	0.1298
	50%	-0.0237	0.0220	0.0294	0.2663
	70%	-0.0694	0.0469	0.1246	0.7988
0.3	0%	-0.0164	0.0400	0.0419	0.1017
	10%	-0.0540	0.0015	0.0656	0.1703
	30%	-0.1071	-0.0008	0.1127	0.3583
	50%	-0.1610	0.0451	0.1886	0.7904
	70%	-0.3496	0.2710	0.3052	2.3817
0.5	0%	-0.0426	0.0374	0.1037	0.1843
	10%	-0.1209	-0.0177	0.1869	0.3021
	30%	-0.2433	-0.0350	0.3334	0.6803
	50%	-0.3616	0.0569	0.4482	1.5199
	70%	-0.5253	0.5265	0.4690	5.4089
0.7	0%	-0.0835	0.0417	0.2634	0.3277
	10%	-0.2466	-0.0591	0.4544	0.5546
	30%	-0.4027	-0.0496	0.6701	1.1857
	50%	-0.5899	0.0983	0.7828	2.9789
	70%	-0.8433	0.4436	0.5060	11.3840

Таблица 3 – Статистические свойства ОМП параметром распределения Вейбулла. $n = 200$

Степень усеч.	Средний процент ценз.	$M\hat{\theta} - \theta_{ист.}$		$D\hat{\theta}$	
		масштаб	форма	масштаб	Форма
0	0%	0.0003	0.0136	0.0054	0.0125
	10%	-0.0001	0.0132	0.0058	0.0158
	30%	0.0020	0.0147	0.0074	0.0231
	50%	0.0068	0.0197	0.0131	0.0363
	70%	0.0220	0.0254	0.0408	0.0654
0.1	0%	-0.0033	0.0178	0.0081	0.0251
	10%	-0.0056	0.0098	0.0087	0.0363
	30%	-0.0054	0.0132	0.0091	0.0651
	50%	-0.0106	0.0168	0.0115	0.1390
	70%	-0.0133	0.0108	0.0383	0.3928
0.3	0%	-0.0126	0.0168	0.0209	0.0523
	10%	-0.0315	-0.0061	0.0288	0.0830
	30%	-0.0666	-0.0244	0.0534	0.1823
	50%	-0.1398	-0.0402	0.1037	0.4439
	70%	-0.2579	0.1015	0.1748	1.3558
0.5	0%	-0.0356	0.0006	0.0502	0.0864
	10%	-0.0982	-0.0524	01039	0.1655
	30%	-0.2186	-0.1179	0.2207	0.3880
	50%	-0.3603	-0.1081	0.3371	0.9072
	70%	-0.5461	0.0711	0.3474	2.9340
0.7	0%	-0.0937	-0.0305	0.1442	0.1663
	10%	-0.2381	-0.1270	0.3057	0.3224
	30%	-0.4569	-0.2276	0.5301	0.7036
	50%	-0.7153	-0.2820	0.6033	1.6111
	70%	-0.9955	-0.4094	0.2572	4.2810

2.3. Потери информации Фишера от усечения

Пусть имеется полная выборка времен жизни $X_1, X_2, ..., X_n$ из распределения $F(t;\theta)$, где $\theta \in \Theta$ – это вектор параметров размерности. Если условия проведения эксперимента таковы, что наблюдению доступны только те отказы, для которых время жизни больше некоторой наперед заданной величины τ, называемой временем усечения, то в результате получаем выборку усеченных усеченных слева величин $(X_{(1)},\tau),(X_{(2)},\tau),...,(X_{(M)},\tau)$, объем

которой представляет собой случайную величину $M \le n$, принадлежащую биномиальному распределению $Bi\left(n,1-F(\tau)\right)$.

Информационное количество Фишера в выборке, полученной путем усечения из полной выборки $X_1, X_2, ..., X_m$, имеет вид:

$$I^{LT}(\theta) = \sum_{m=0}^{n}\left(P\{M=m\}\cdot m \cdot i^{LT}\left(\theta \mid M=1\right)\right) = \\ = \sum_{m=0}^{n}\left(C_n^m (1-d)^m d^{n-m} \cdot m \cdot i^{LT}\left(\theta \mid M=1\right)\right) = n(1-d) i^{LT}\left(\theta \mid M=1\right), \quad (26)$$

где $i^{LT}\left(\theta \mid M=1\right)$ – это информационное количество Фишера о параметре θ в усеченном наблюдении:

$$i^{LT}\left(\theta \mid M=m\right) = \int_{\tau}^{\infty}\left(\frac{\partial \ln f^{LT}(t;\theta)}{\partial \theta}\right)^T \frac{\partial \ln f^{LT}(t;\theta)}{\partial \theta} f^{LT}(t;\theta)dt. \quad (27)$$

Поскольку для оценивания s неизвестных параметров требуется как минимум s наблюдений, то выражение (26) примет вид:

$$I^{LT}(\theta) = \frac{n(1-d) i^{LT}\left(\theta \mid M=1\right)}{1-\sum_{m=0}^{s-1} C_n^m (1-d)^m d^{n-m}}. \quad (28)$$

Понятно, что при больших объемах выборок величиной $\sum_{m=0}^{s-1} C_n^m (1-d)^m d^{n-m}$ в выражении (28) 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 $X_1, X_2, ..., X_n$. В таблице 4 для закона распределения Вейбулла найдены значения $\det I^{LT}(\theta)/\det I(\theta)$ для параметров распределения Вейбулла в зависимости от степени усечения при $n \ge 100$.

Таблица 4 – Отношение информационного количества Фишера в выборке усеченных слева наблюдений к информационному количеству в исходной полной выборке

d	О параметре θ_0 распределения Вейбулла	О параметре θ_1 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 таблице 4, 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 $\det D\left[\theta\right] / \det D\left[\theta^{LT}\right]$, где θ – ОМП неизвестного параметра распределения по полной выборке, θ^{LT} – ОМП по выборке усеченных слева наблюдений. Результаты моделирования представлены в таблице 5. Количество N моделируемых выборок, по которым исследовались законы распределения оценок по выборкам объема n, было взято равным

100000. При исследовании распределений оценок выборки моделировались по закону Вейбулла с параметром масштаба $\theta_0 = 2$ и формы $\theta_1 = 2$.

Таблица 5 – Относительная эффективность оценивания параметров распределения Вейбулла в зависимости от объема выборки

О параметре θ_0 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_1 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 таблице 5, видно, что с ростом объема выборки величина относительной эффективности оценивания параметров распределения Вейбулла $D\left[\theta\right] / D\left[\theta^{LT}\right]$ 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_0, только параметра θ_1 и одновременном оценивании двух параметров. Наблюдаемые отклонения на уровне погрешности моделирования.

2.4. Исследование точности ОМП по выборкам усеченных слева наблюдений

Пусть имеется выборка усеченных слева наблюдений вида:

$$\mathbb{X}_n = \{(X_1, \tau_1), (X_2, \tau_2), ..., (X_n, \tau_n)\},$$

где n – объем выборки, X_i – время отказа i-го объекта, τ_i – время усечения, $i = \overline{1,n}$. Если $\tau_i = 0$, 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 τ_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 выживаемости, 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 $0 < \gamma \le 1$ 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

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

где $i^{LT}(\theta \mid M=1)$ – 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. 6 представлены значения информационного количества Фишера $\det i^{LT}(\theta \mid M=1)$ о параметрах θ_0 или θ_1 в одном наблюдении из усеченного распределения Вейбулла при $\theta_0 = 2$ и $\theta_1 = 2$ 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. 6, 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 6 – Информационное количество Фишера в одном наблюдении из усеченного распределения Вейбулла с параметрами $\theta_0 = 2$ и $\theta_1 = 2$

d	$i^{LT}(\theta_0 \mid M=1)$	$i^{LT}(\theta_1 \mid M=1)$	$i^{LT}(\theta \mid M=1),\ \theta = (\theta_0, \theta_1)$
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 таблице 6 величины соответствует 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_0 остается неизменной, о параметре θ_1 увеличивается, а о векторном параметре – уменьшается.

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

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

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

В таблице 7 представлены отношения минимальной дисперсии $I_{\gamma}^{-1}(\theta_1)$ к оценке реальной дисперсии $\hat{D}\left[\hat{\theta}_1\right]$ 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 7 – Оценки величины $I_{\gamma}^{-1}(\theta_1) \cdot D^{-1}\left[\theta_1\right]$ 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. 7, с ростом объема выборки дисперсия ОМП стремится к соответствующей асимптотической дисперсии $I_{\gamma}^{-1}(\theta_1)$. Отметим, что в случае степени усечения $d=0.1$ и $d=0.3$, когда информация Фишера $i^{LT}(\theta_1 | M=1)$ в наблюдении по усеченному закону меньше, чем в полном наблюдении (см. табл. 6), 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 $I_{\gamma}^{-1}(\theta_0)\cdot\hat{D}^{-1}\left[\hat{\theta}_0\right]$ 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. 8 представлены отношения $\det I_{\gamma}^{-1}(\theta)$ к $\det\hat{D}\left[\hat{\theta}\right]$ 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 $\det \hat{D}\left[\hat{\theta}\right]$ к $\det I_{\gamma}^{-1}(\theta)$ не столь существенно как при увеличении от 75% до 100%.

Таблица 8 – Оценки величины $\det I_{\gamma}^{-1}(\theta) \cdot \det \hat{D}^{-1}\left[\hat{\theta}\right]$ 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

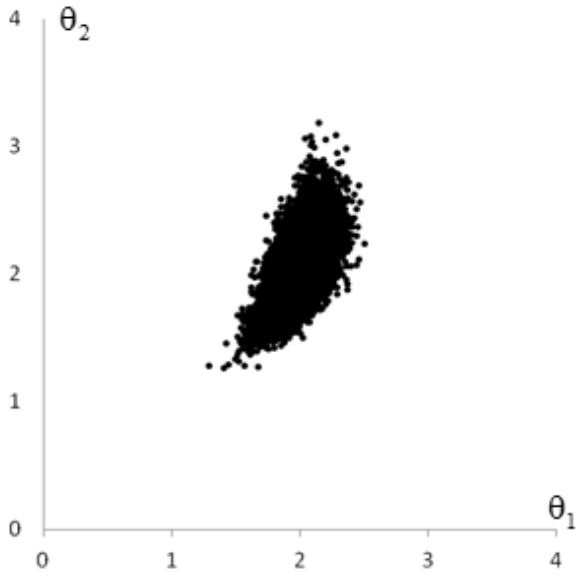
Продолжение таблицы 8

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

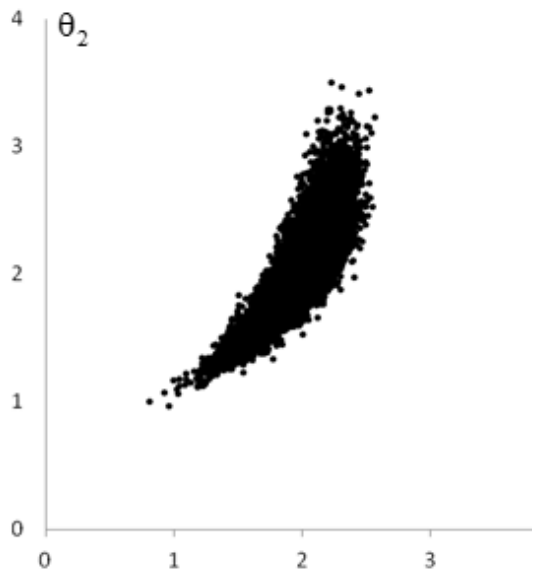
75%	0.868	0.943	0.961	0.965	0.976	1.004
100%	0.305	0.426	0.519	0.627	0.692	0.774

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. 3 – 4. 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 .

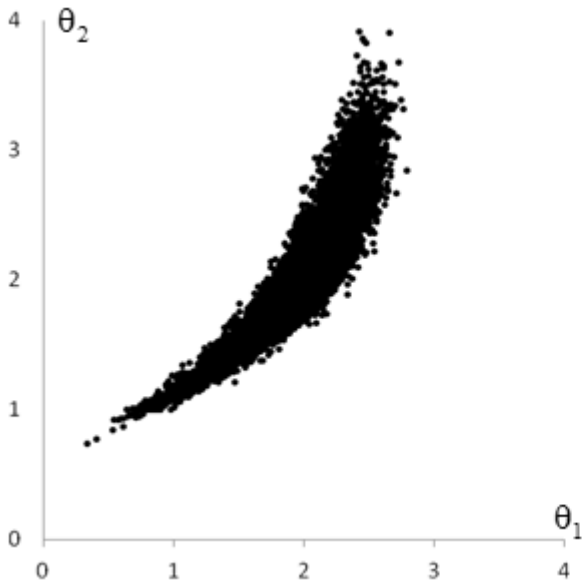
Как можно видеть, при увеличении степени и величины $\gamma \cdot 100\%$ эллипсоиды рассеяния ОМП деформируются и становятся все больше ассиметричными относительно осей. Это свидетельствует о том, что закон распределения ОМП параметров масштаба и формы распределения Вейбулла (по выборкам из смеси усеченных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



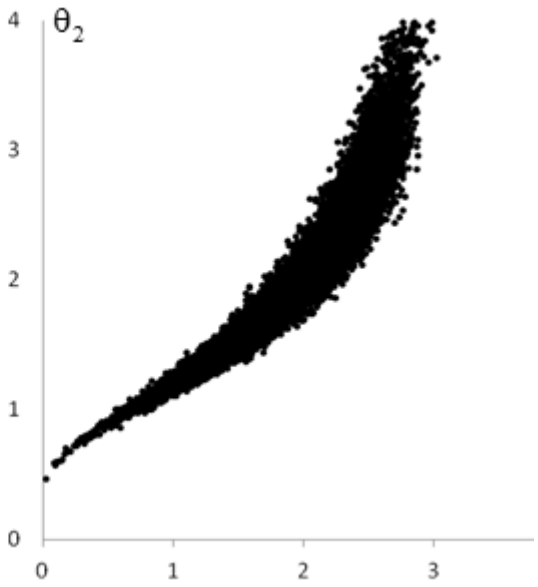
(а) $d=0.1$



(б) $d=0.3$

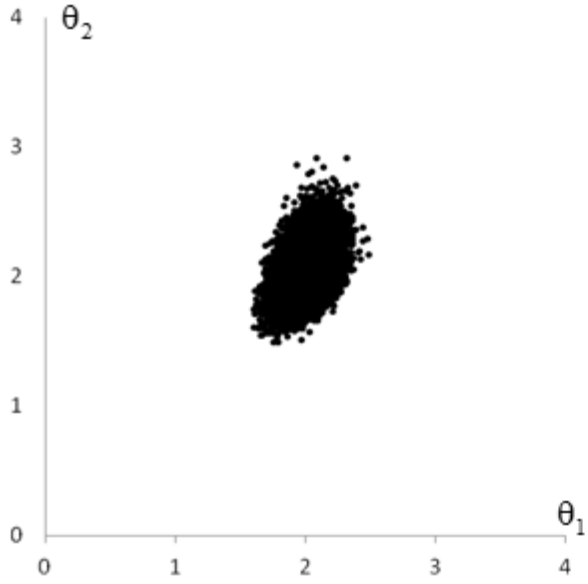


(в) $d = 0.5$

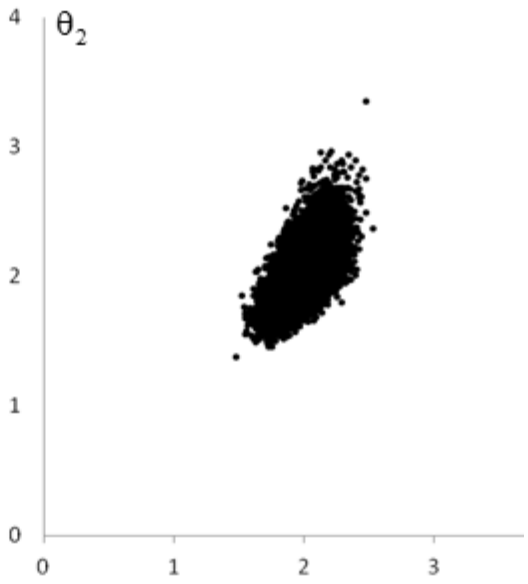


(в) $d = 0.7$

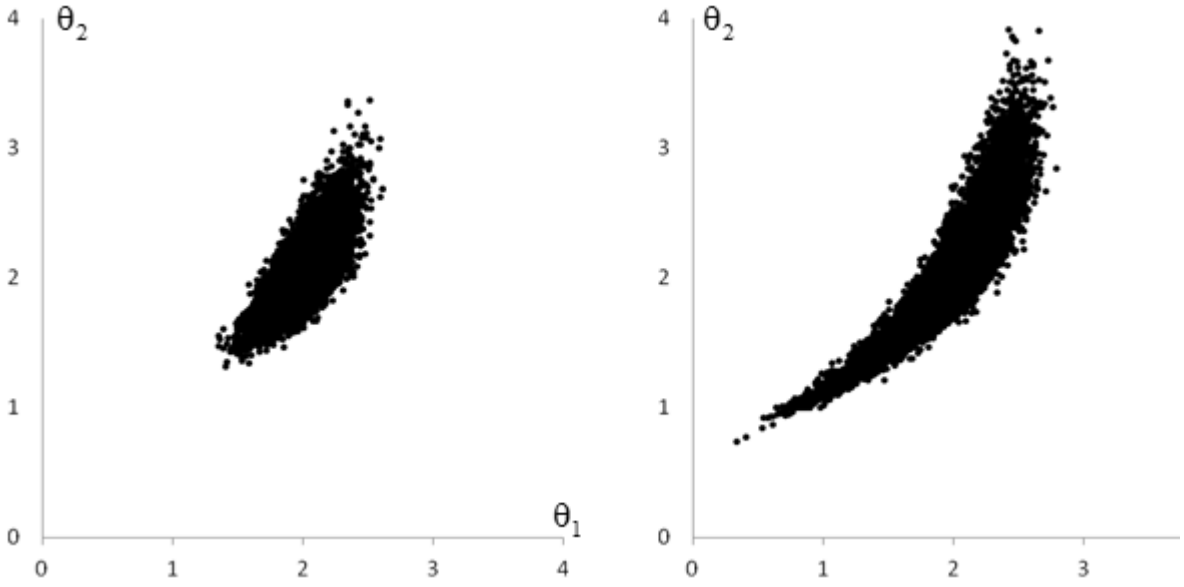
Рисунок 3 – Диаграммы рассеяния ОМП параметров масштаба и формы распределения Вейбулла по выборкам усеченных наблюдений объема $n = 100$, 100% наблюдений из усеченного распределения Вейбулла



(а) 25%



(б) 50%



(в) 75% (г) 100%

Рисунок 4 – Диаграммы рассеяния ОМП параметров масштаба и формы распределения Вейбулла по выборкам усеченных слева наблюдений объема $n=100$ при степени усечения $d=0.5$

На рис. 3 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. 3 – 4 наглядно объясняют результаты, полученные в табл. 8, 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.

2.4. Исследование свойства эффективности оценок максимального правдоподобия по цензурированным справа данным

С целью исследования эффективности оценки максимального правдоподобия на цензурированных справа данных проведем эксперимент при аналогичных условиях. Рассмотрим случай, когда выборка содержит только неусеченные наблюдения, а степень цензурирования имеет различные значения и рассчитаем отношения количества информации Фишера в наблюдении цензурированной выборки к количеству информации в полной, представленные в таблице 9.

Таблица 9 – Отношение количества информации Фишера в цензурированной справа выборке к количеству информации в исходной полной выборке

Степень цензури-рования	О параметре θ_0 распределения Вейбулла	О параметре θ_1 распределения Вейбулла	О двух параметрах распределения Вейбулла
5%	0.9500	0.8461	0.8346
10%	0.9000	0.7410	0.7133
20%	0.8000	0.5920	0.5240
30%	0.7000	0.4949	0.3791
40%	0.6000	0.4343	0.2657
50%	0.5000	0.4010	0.1771
60%	0.4000	0.3878	0.1093
70%	0.3000	0.3859	0.0595
80%	0.2000	0.3814	0.0257
Степень цензури-рования	О параметре θ_0 логнормального распределения	О параметре θ_1 логнормального распределения	О двух параметрах логнормального распр.
5%	0.9931	0.9235	0.9166
10%	0.9831	0.8558	0.8389
20%	0.9563	0.7375	0.6943
30%	0.9206	0.6392	0.5615
40%	0.8753	0.5599	0.4399
50%	0.8183	0.5000	0.3296
60%	0.7467	0.4601	0.2311
70%	0.6550	0.4400	0.1457
80%	0.5336	0.4360	0.0754

Как и ожидалось, при увеличении значения степени цензурирования значение отношения количества информации Фишера по цензурированной

выборке к количеству информации Фишера по полной выборке уменьшается, что говорит нам о том, что точность оценки параметров уменьшается.

Проверим результаты расчета теоретической величины асимптотической дисперсии, сравнив с практическими расчетами значения относительной эффективности для выборок с различными объемами. В случае цензурированной выборки это величина $I^C(\theta)/I(\theta)$ в скалярном случае и величина $\det I^C(\theta)/\det I(\theta)$ в векторном случае, возникающим при оценивании двух параметров одновременно. При построении распределений оценок выборки моделировались по закону Вейбулла с параметром масштаба $\theta_0 = 2$ и формы $\theta_1 = 2$. Затем вычислялась оценка по смоделированной полной выборке, после чего выборка цензурировалась, и находилась оценка $\hat{\theta}^c$. В результате по полученным выборкам оценок объема $N = 100000$ вычислялось отношение $\det D[\hat{\theta}]/\det D[\hat{\theta}^c]$. Эти величины представлены в таблице 10.

Таблица 10 – Относительная эффективность оценивания параметра распределения Вейбулла по цензурированным I типа выборкам по сравнению с оцениванием по полной выборке в зависимости от объема выборки n

Степень цензури-рования	*n* = 100	*n* = 200	*n* = 300	*n* = 500	*n* = 700	*n* = 1000
5%	0.8447	0.8364	0.8364	0.8367	0.8357	0.8344
10%	0.7262	0.7179	0.7190	0.7158	0.7144	0.7130
20%	0.5276	0.5239	0.5245	0.5272	0.5240	0.5243
30%	0.3718	0.3733	0.3765	0.3806	0.3783	0.3763
40%	0.2521	0.2553	0.2605	0.2639	0.2638	0.2608
50%	0.1575	0.1647	0.1694	0.1740	0.1744	0.1736
60%	0.0840	0.0948	0.0998	0.1050	0.1058	0.1068
70%	0.0345	0.0456	0.0502	0.0540	0.0555	0.0571
80%	0.0073	0.0138	0.0173	0.0205	0.0216	0.0232

При увеличении объема выборки точность получаемых оценок несколько увеличивается, а значение относительной эффективности приближается к значению асимптотической дисперсии. Значения относительной эффективности

полученные методом компьютерного моделирования в целом подтверждают теоретические результаты, представленные в таблице 9.

2.5. Выводы

Обобщим результаты, приведенные в таблицах 2 и 3 и главы в целом. Проанализируем свойства оценок максимального правдоподобия при варьировании объема выборки, количества цензурированных и усеченных наблюдений.

В исследовании были подтверждены свойства оценки максимального правдоподобия для данных усеченных слева цензурированных справа:

1. при возрастании степени цензурирования или степени усечения смещение относительно истинного значения и выборочная дисперсия увеличивается, что подтверждает такое свойство как несмещенность оценки;
2. при увеличении степени усечения дисперсия возрастает, а смещение близко к нулю и почти не меняется;
3. дисперсия оценки максимального правдоподобия при увеличении объема уменьшается, значит, выполняется свойство состоятельности;
4. в случае оценивания двух параметров при увеличении степени усечения или степени цензурирования относительная эффективность оценок уменьшается;
5. 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;
6. 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.

7. при увеличении степени цензурирования точность ОМП падает.

8. ДОБАВИТЬ ПРО ЦЕНЗУРИРОВАНИЕ

Таким образом, было показано, что на оценку максимального правдоподобия изменение степени усечения в большую сторону влияет больше, чем увеличение процента цензурированных наблюдений.

3. Исследование свойств оценок Каплана-Мейера

Как говорилось в первой главе, в случае, когда априорные данные не содержат в себе информацию о виде распределения и значения параметров, для оценки функции распределения или функции надежности используют оценку Каплана-Мейера.

Проведем исследование и посмотрим, как ведет себя множительная оценка в зависимости от типа цензурирования, процентного содержания полных и усеченных наблюдений, объема выборки и распределений времен жизни.

3.1. Исследование влияния степени цензурирования на оценку Каплана-Мейера для цензурованных III типа данных

В этом исследовании будем использовать случайный способ цензурирования. В качестве закона распределения моментов окончания экспериментов возьмем семейство распределений Вейбулла. Путем варьирования значений параметров масштаба и формы были получены желаемые количества неполных наблюдений. Полученные параметры распределений цензурирования приведены в таблице 11.

Таблица 11 – Параметры масштаба и формы для распределений Вейбулла, определяющих моменты цензурирования

Процент цензурирования	Параметр масштаба	Параметр формы
0	20	10
25	3	1
50	1	1
75	1	1.6

Точка усечения для всех наблюдений будет задаваться одинаковой, таким образом, выборка будет содержать только усеченные данные. В таком случае, теоретическая функция распределения описывается формулой (1).

В качестве распределений времен жизни выбраны:

1) распределение Вейбулла с параметрами $\theta_0 = 1$, $\theta_1 = 1$,

2) Гамма-распределение с плотностью:

$$f(t)=\frac{1}{\theta_1^{\theta_0}\Gamma(\theta_0)}t-\theta_2^{\theta_0-1}e^{-t-\theta_2/\theta_1}$$ с параметрами $\theta_0=0$, $\theta_1=1$, $\theta_2=1$.

С помощью реализованной компьютерной программы была получена оценка Каплана-Мейера для усеченной функции распределения закона Вейбулла в точке усечения $\tau=3$, представленная на рисунке 5.

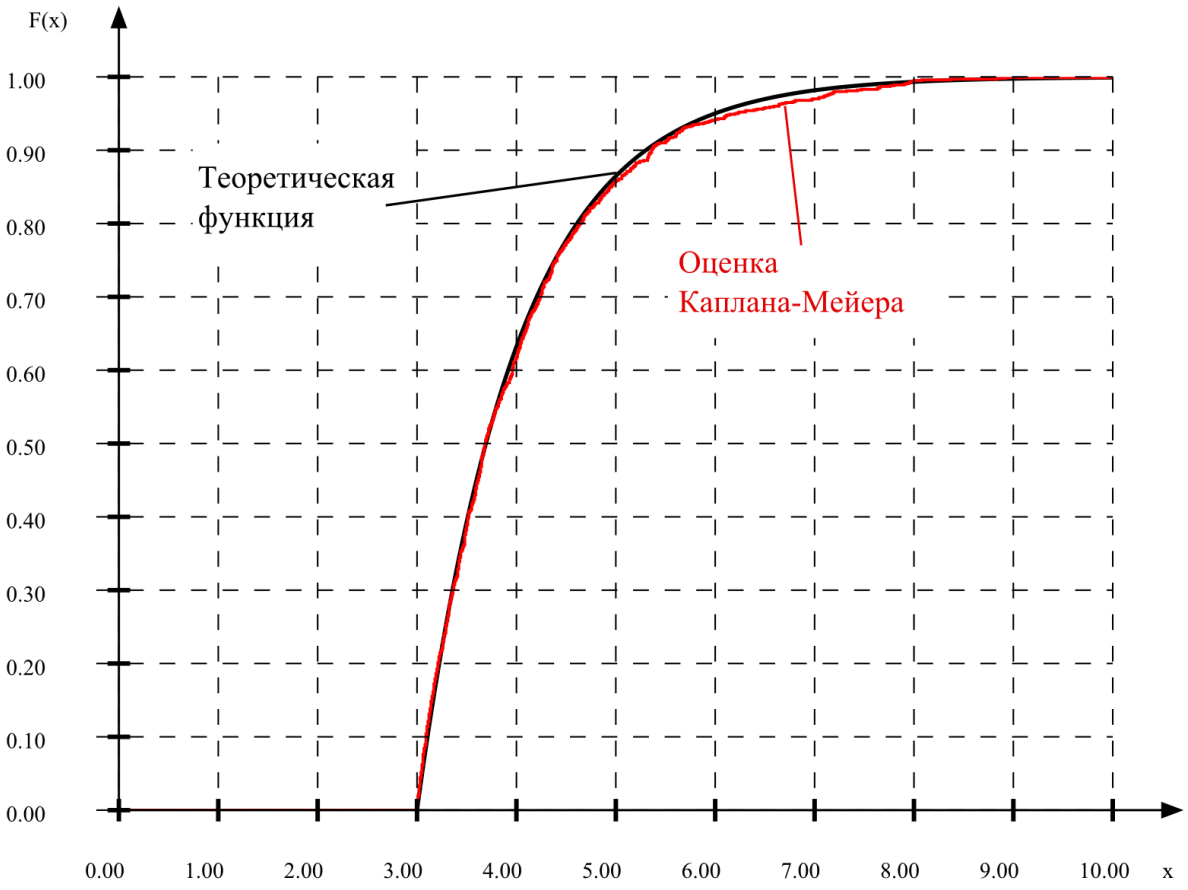


Рисунок 5 – Теоретическая функция и оценка Каплана-Мейера для распределения Вейбулла ($\theta_0=1$, $\theta_1=1$) для полной выборки. Точка усечения $\tau=3$. Объем выборки $n=1000$

По графику видно, что полученная функция распределения почти что совпадает с теоретической, из этого можно сделать вывод о том, что программный модуль работает правильно.

Ниже приведены расчеты среднего расстояния D_n между истинной функцией распределение и полученной в результате моделирования

множительной оценкой при проведении 100 экспериментов. В таблицах 12 и 13 приведены результаты для Вейбулла и гамма-распределений соответственно.

Таблица 12 – Отклонение оценок Каплана-Мейера от теоретической функции распределения Вейбулла для третьего типа цензурирования

Точка усечения	Процент цензурирования	Объем выборки				
		100	200	500	1000	2000
1.5	0	0.08586	0.05952	0.03875	0.02838	0.02014
	25	0.09653	0.06809	0.04400	0.03132	0.02160
	50	0.13122	0.09300	0.05895	0.04415	0.03239
	75	0.14598	0.12278	0.09160	0.08696	0.07380
3	0	0.08187	0.06011	0.03769	0.02843	0.01857
	25	0.09618	0.06879	0.04249	0.02890	0.02122
	50	0.12749	0.08846	0.06152	0.04568	0.03233
	75	0.14964	0.13897	0.09937	0.09214	0.07285

Таблица 13 – Отклонение оценок Каплана-Мейера от теоретической функции гамма-распределения для третьего типа цензурирования

Точка усечения	Процент цензурирования	Объем выборки				
		100	200	500	1000	2000
1.5	0	0.08485	0.06096	0.03850	0.02827	0.01972
	25	0.09767	0.06757	0.04290	0.03032	0.02313
	50	0.13001	0.09222	0.05835	0.04600	0.03274
	75	0.13853	0.12582	0.09974	0.09077	0.07233
3	0	0.08538	0.05646	0.03026	0.02416	0.01736
	25	0.09251	0.07285	0.04293	0.02494	0.02200
	50	0.12514	0.08665	0.05889	0.04194	0.03194
	75	0.14714	0.14222	0.10956	0.08614	0.07984

Из таблиц 12 и 13 можно сделать вывод, что чем больше объем выборки, тем ближе к теоретической функции будет находиться оценка Каплана-Мейера. Также же необходимо отметить, что при увеличении процента цензурирования точность оценки снижается.

Как и следовало ожидать, полученные результаты не зависят от выбора вида распределений времен жизни. Экспериментально был получен и тот результат, что точность множительной оценки не зависит от выбора точки усечения.

Рисунок 6 иллюстрирует зависимость от объема выборки средних расстояний между теоретической функцией распределения Вейбулла и её

оценкой Каплана-Мейера, также стоит заметить, что представленные графики соответствуют зависимости типа $\overline{D}_n(n) = \alpha_0 n^{\alpha_1}$, где α_0, α_1 – действительные коэффициенты, их значения приведены в таблице 14.

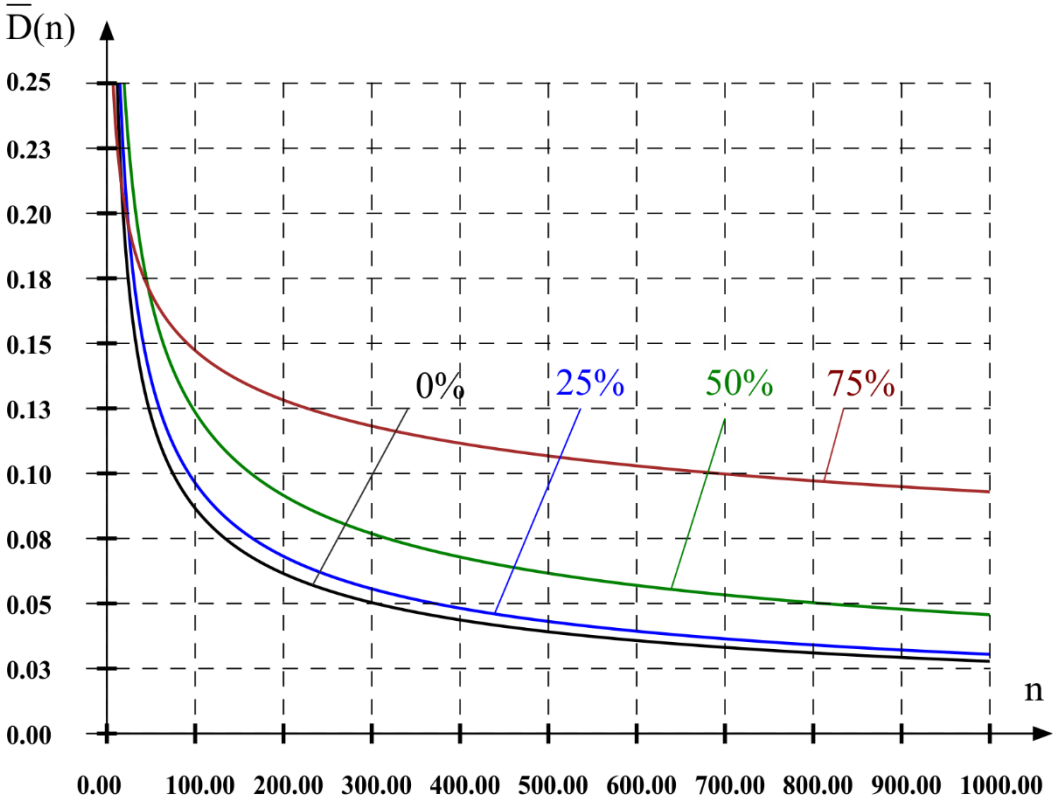


Рисунок 6 – Зависимость средних отклонений $\overline{D}_n(n)$ от объема выборки при разных степенях цензурирования для выборки усеченных наблюдений

Из рисунка 6, очевидно, что точность оценки повышается при росте полного объема выборки, однако, при возрастании количества неполных наблюдений величина $\overline{D}_n(n)$ увеличивается.

Таблица 14 – Коэффициенты средних отклонений $\overline{D}_n(n) = \alpha_0 n^{\alpha_1}$ и коэффициент детерминации R^2

Процент цензурирования	α_0	α_1	R^2
0	0.8416	-0.5	0.9975
25	0.9693	-0.5	0.9981

Продолжение таблицы 14

50	0.9031	-0.4	0.9978

75	0.3699	-0.2	0.9642

В таблице 14 помимо коэффициентов α_0, α_1 приведены значения коэффициента детерминации R^2. Так как его значение близко к 1, то можно сделать вывод о том, что аппроксимирующая функция достаточно хорошо описывает экспериментальные данные.

На рисунке 7 показано, как полученные результаты расчетов средних отклонений $\overline{D}_n(n)$ в некоторых точках n аппроксимируются функцией $\overline{D}_n(n) = 0.9693 \cdot n^{-0.5}$.

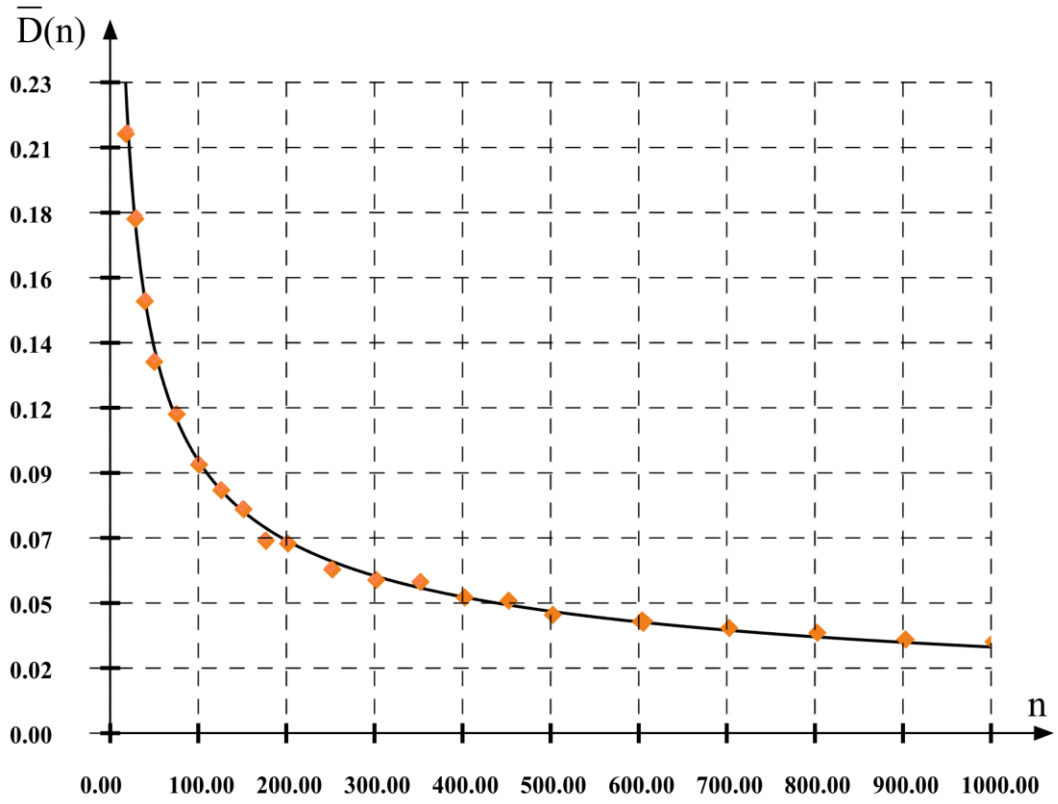


Рисунок 7 – Зависимость средних отклонений $\overline{D}_n(n)$ от объема выборки при степени цензурирования 25% для выборки, содержащей усеченные наблюдения

Рисунок 7 иллюстрирует, что полученная функция $\overline{D}_n(n) = 0.9693 \cdot n^{-0.5}$ хорошо аппроксимирует средние отклонения $\overline{D}_n(n)$.

3.2. Исследование влияния усечения на оценку Каплана-Мейера для цензурованных III типа данных

Проведем исследование зависимости оценки Каплана-Мейера от процента цензурирования. Для этого воспроизведем эксперимент при условиях аналогичных в пункте 3.1. для распределения Вейбулла.

В качестве теоретической функции в данном случае будем рассматривать функцию вида (13).

Таблица 15 – Отклонение оценок Каплана-Мейера от теоретической функции распределения Вейбулла в зависимости от степени цензурирования и процента усеченных наблюдений в неполных выборках с разным объемом. Количество экспериментов – 100

Процент усечения	Процент цензурирования	Объем выборки				
		100	200	500	1000	2000
0	0	0.08875	0.06145	0.03759	0.02557	0.02038
	25	0.09394	0.06900	0.04332	0.03011	0.02115
	50	0.12341	0.09237	0.06218	0.04468	0.03526
25	0	0.10168	0.07039	0.04555	0.03182	0.02149
	25	0.11727	0.07759	0.04790	0.03423	0.02513
	50	0.18404	0.13375	0.09781	0.07769	0.05778
50	0	0.12220	0.08551	0.05482	0.03918	0.02729
	25	0.13650	0.09396	0.05966	0.04619	0.03094
	50	0.18731	0.12398	0.08778	0.06551	0.04522
75	0	0.16385	0.11539	0.07884	0.05411	0.03790
	25	0.18701	0.13839	0.08442	0.05996	0.04356
	50	0.25905	0.19417	0.12455	0.08995	0.06283
100	0	0.95070	0.95050	0.95031	0.95027	0.95024
	25	0.95074	0.95046	0.95032	0.95027	0.95024
	50	0.95067	0.95041	0.95031	0.95027	0.95024

Результаты компьютерного моделирования для разных процентах усечения и цензурирования при изменении объема выборки приведены в таблице 15. Из нее видно, что резкое уменьшение точности наблюдается для процента усечения, величина которой превышает 75%. Рассмотрим подробнее случай, когда процент усечения принимает значения от 90% до 100%.

Таблица 16 – Отклонение оценок Каплана-Мейера от теоретической функции распределения Вейбулла в зависимости от процента усеченных наблюдений в полных выборках с разным объемом. Количество экспериментов – 300

Процент усечения	Объем выборки				
	100	200	500	1000	2000
90	0.26675	0.18871	0.11675	0.08306	0.06065
91	0.26326	0.19704	0.12753	0.08990	0.06333
92	0.28700	0.20663	0.12992	0.09673	0.06843
93	0.30669	0.22054	0.13884	0.10479	0.07116
94	0.33224	0.24204	0.15307	0.10588	0.07906
95	0.35849	0.25827	0.16724	0.11707	0.08513
96	0.40304	0.28284	0.18465	0.12982	0.09832
97	0.44961	0.32346	0.21449	0.15512	0.10772
98	0.54594	0.39647	0.25511	0.18485	0.13556
99	0.72410	0.53462	0.37222	0.25554	0.18077
100	0.95073	0.95048	0.95031	0.95026	0.95024

Результаты расчетов в таблице 16 показывают, что если в выборке присутствует хотя бы один процент неусеченных наблюдений, то оценка Каплана-Мейера приближается к функции распределения, не учитывающей усеченные наблюдения, однако, при выборке, в которой процент усечения равен 100%, множительная оценка лежит ближе к функции распределения описывающейся формулой (1).

Графическая интерпретация представлена на рисунке 8.

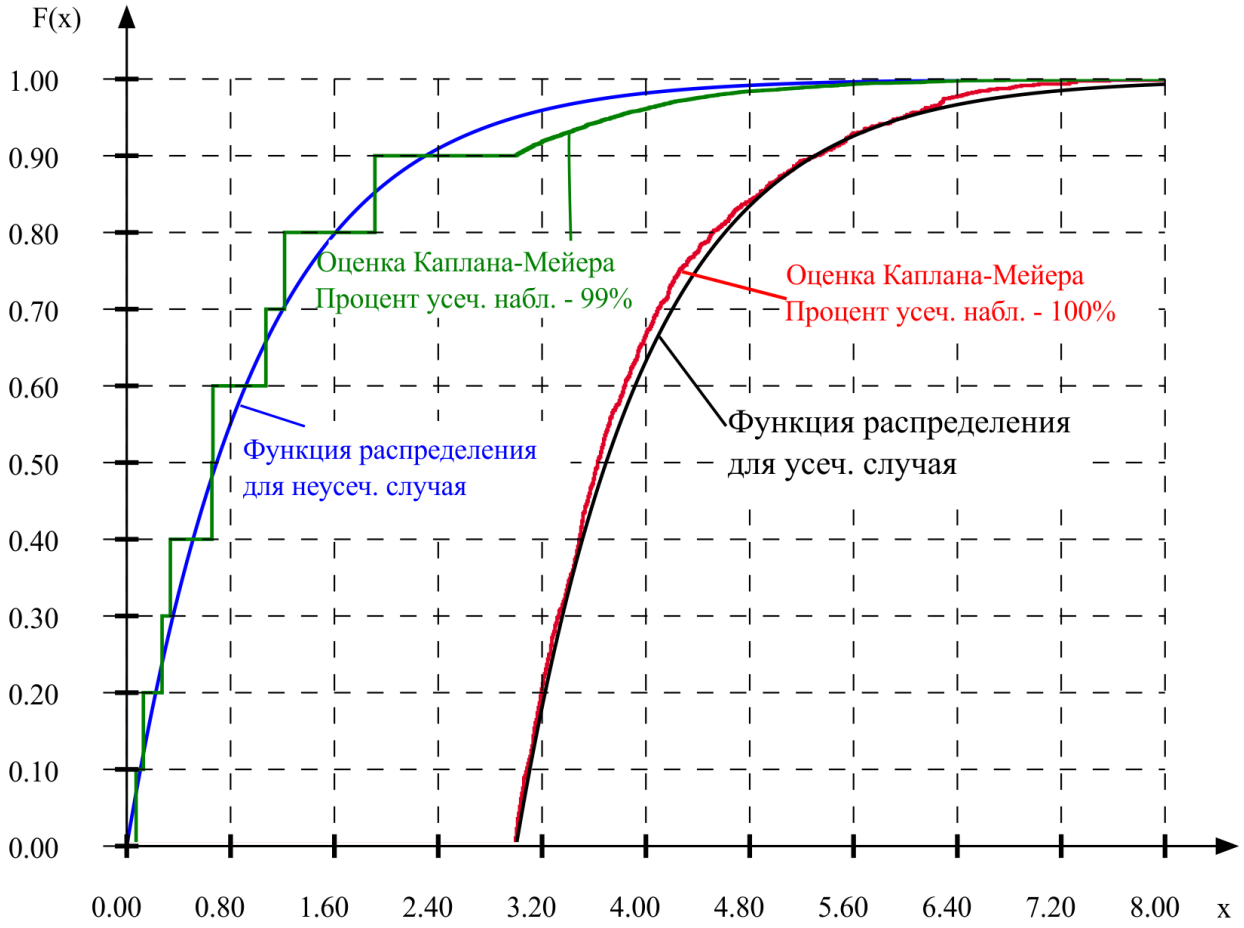


Рисунок 8 – Теоретические функции распределения для усеченного и неусеченного случаев и оценки Каплана-Мейера для 99%, 100% процентов усеченных наблюдений в выборках

Как видно из рисунка 8, оценка Каплана-Мейера для выборки, в которой все объекты являются усеченными, лежит ближе к функции распределения для усеченного случай, в то время как график множительной оценки для выборки, содержащей 99% усеченных наблюдений, расположен около функции распределения для неусеченного случая.

3.3. Исследование влияния цензурирования на оценку Каплана-Мейера для цензурованных I типа данных

Рассмотрим, как повлияет на свойства оценка Каплана-Мейера, если использование первого типа цензурирования при этом момент цензурирования будем рассчитывать с помощью метода обратной функции.

Эксперимент проводился в тех же условиях, что и для третьего типа цензурирования, когда времена жизни были подчинены закону распределения

Вейбулла. Исследуем, как зависит точность оценки Каплана-Мейера от точки усечения, процента цензурирования и объема выборки при первом типе цензурирования.

Ниже на рисунке 9 представлены графики теоретической функции и оценки Каплана Мейера для закона распределения Вейбулла.

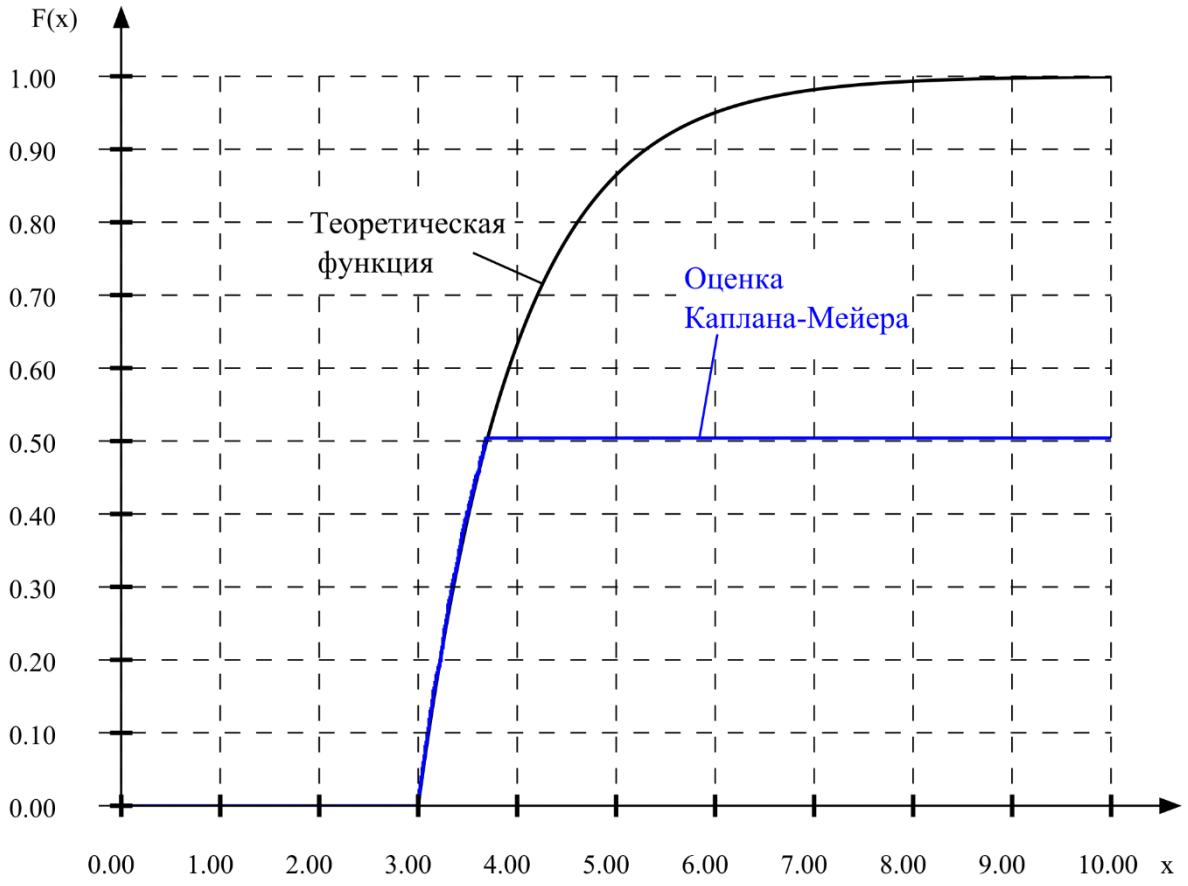


Рисунок 9 – Теоретическая функция и оценка Каплана-Мейера для распределения Вейбулла ($\theta_0 = 1$, $\theta_1 = 1$) для выборки с показателем цензурирования 50% (1 тип). Точка усечения $\tau = 3$. Объем выборки $n = 1000$.

Из рисунка 9 видно, что в случае цензурированных данных оценка Каплана-Мейера строится только на наблюдаемой области, то есть до точки последнего отказа, поэтому в данном случае при степени цензурирования 50% оценка Каплана-Мейера принимает значении от 0 до 0.5.

Таблица 17 – Отклонение оценок Каплана-Мейера от теоретической функции распределения Вейбулла для I типа цензурирования

Точка усечения	Процент цензурирования	Объем выборки				
		100	200	500	1000	2000
1.5	0	0.08487	0.06093	0.03888	0.02666	0.01921
	25	0.08117	0.06310	0.03815	0.02679	0.01925
	50	0.07870	0.05464	0.03420	0.02350	0.01695
	75	0.05285	0.04398	0.02658	0.01900	0.01180
3	0	0.08644	0.06133	0.03944	0.02767	0.01861
	25	0.08438	0.06014	0.03590	0.02821	0.01877
	50	0.07251	0.05469	0.03324	0.02288	0.01687
	75	0.05383	0.03892	0.02521	0.01801	0.01304

В таблице 17 приведены отклонения среднего значения статистики Колмогорова для разных объемов выборки при варьировании точки усечения и процента цензурирования. Видно, что точность повышается с увеличением количества неполных наблюдений, и это противоположно третьему типу цензурирования, такой эффект возникает, потому что при увеличении степени цензурирования наблюдаемая область уменьшается.

3.4. Выводы

В данной главе проведено исследование вопросов построения оценки Каплана-Мейера для разных типов цензурирования выборки. В результате были сделаны следующие выводы:

1) точность оценки Каплана-Мейера не зависит от выбора точки усечения;
2) точность оценки Каплана-Мейера не зависит от вида распределения времен жизни;
3) получена оценка скорости сходимости множительной оценки к истинному распределению отказов при различных степенях цензурирования. В частности, для степени цензурирования от 0% до 25% скорость сходимости равна примерно $O\left(\frac{1}{\sqrt{n}}\right)$;
4) при увеличении степени цензурирования III типа точность множительной оценки ухудшается;

5) при увеличении степени цензурирования I типа точность множительной оценки улучшается;
6) при проценте усеченных данных в выборке меньше 100% оценка Каплана-Мейера расположена вблизи к функции распределения, не учитывающей усечение.

4. Исследование критериев согласия для усеченных слева и цензурированных справа данных

В данной главе проведем исследование непараметрических критериев согласия при проверке простых и сложных гипотез для усеченных слева и цензурированных справа данных в зависимости от процента усеченных наблюдений в выборке, степени усечения и цензурирования. Покажем результаты исследований на примере проверки гипотез о согласии с распределением Вейбулла с параметрами $\theta_0 = 2$, $\theta_1 = 2$. Объем моделируемых выборок статистик $N = 16600$.

4.1. Исследование распределений статистик критериев согласия для усеченных слева данных

Исследование проверки простой гипотезы для усеченных слева данных показало, что распределения статистик критериев не зависит от процента усеченных наблюдений в выборке и степени усечения, поэтому далее будет рассмотрено исследование на примере проверки сложной гипотезы.

На рисунках 10 и 11 приведены распределения статистики Колмогорова при разных значениях степени усечения для выборок объема $n = 200$, когда процент усеченных наблюдений в выборках равен 25% и 100%, соответственно.

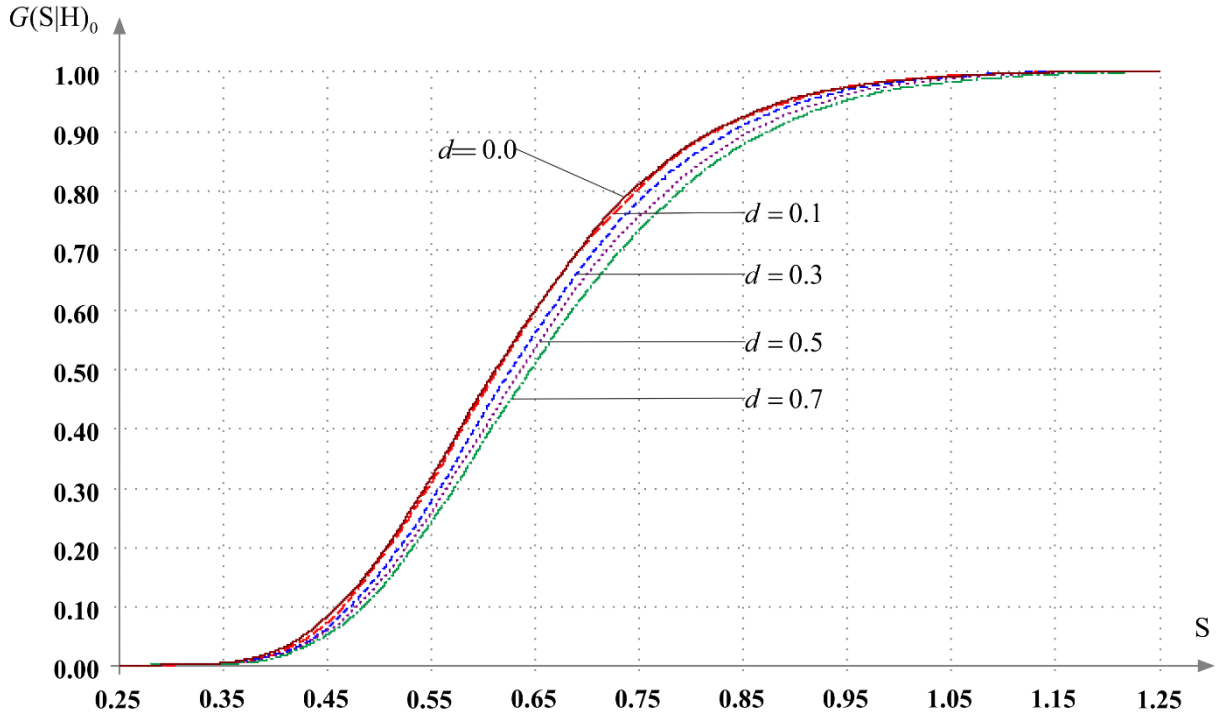


Рисунок 10 – Распределения статистики Колмогорова при разной степени усечения. Процент усеченных наблюдений в выборке равен 25%

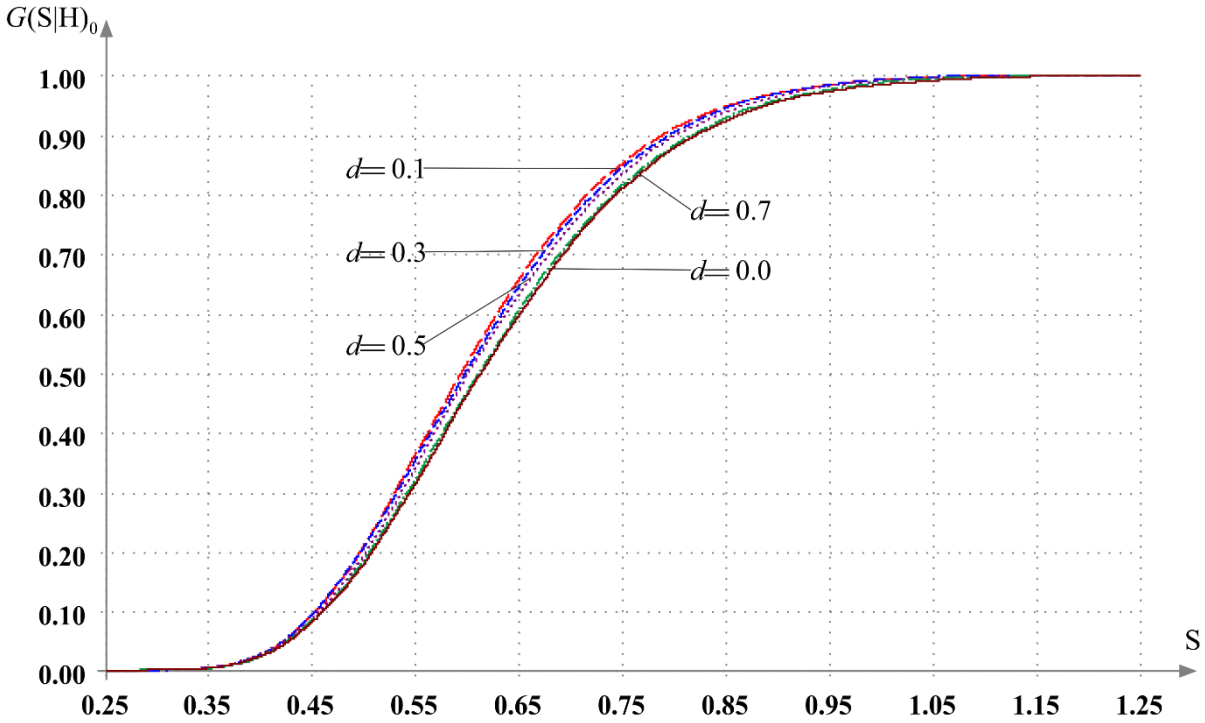


Рисунок 11 – Распределения статистики Колмогорова при разной степени усечения. Процент усеченных наблюдений в выборке равен 100%

Как видно по рисунку 10, для выборки, которая содержит 25% усеченных слева наблюдений, с увеличением степени усечения распределение статистики

Колмогорова заметно отклоняется от предельного закона распределения, однако в случае, когда в выборке все наблюдении являются усеченными (рис. 11), наблюдается обратная закономерность, то есть при увеличении степени усечения отклонение уменьшается. Аналогичные результаты получены и для остальных критериев.

Проведем исследование с

ТУТ НАДО ЧТО ТО ЕЩЕ ДОБАВИТЬ

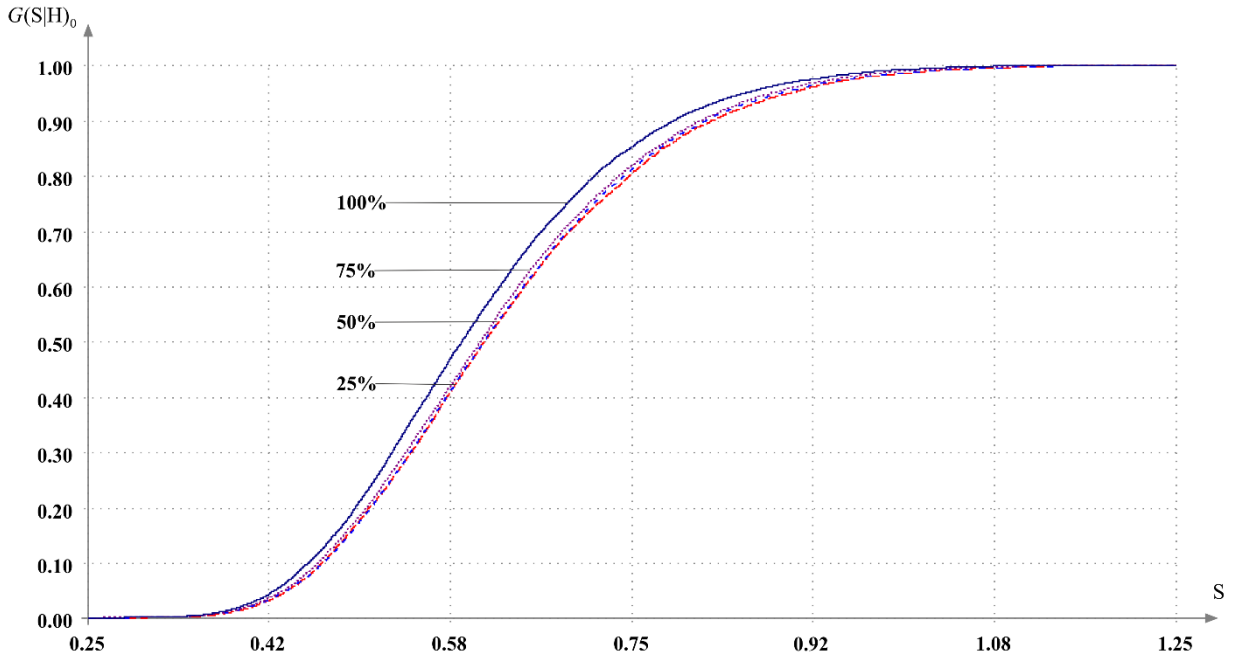


Рисунок 12 – Распределения статистики Колмогорова при разных процентах усеченных наблюдений в выборке. Степень усечения $d = 0.1$

На рисунок 12 иллюстрирует, что при степени усечения равной 0.1, с ростом процента усеченных наблюдений в выборке отклонение распределения статистики Колмогорова от предельного распределения увеличивается. Это справедливо и для других значениях степени усечения.

4.2. Исследование распределений статистик критериев согласия для цензурированных справа данных

Распределения статистик критериев согласия для цензурированнных данных исследовались при проверке простой и сложной гипотез для данных при цензурированнии II типа в зависимости от количества цензурированных

наблюдений в выборке. На рисунках 13 и 14 представлены соответствующие функции распределения статистик для простой и сложной гипотез.

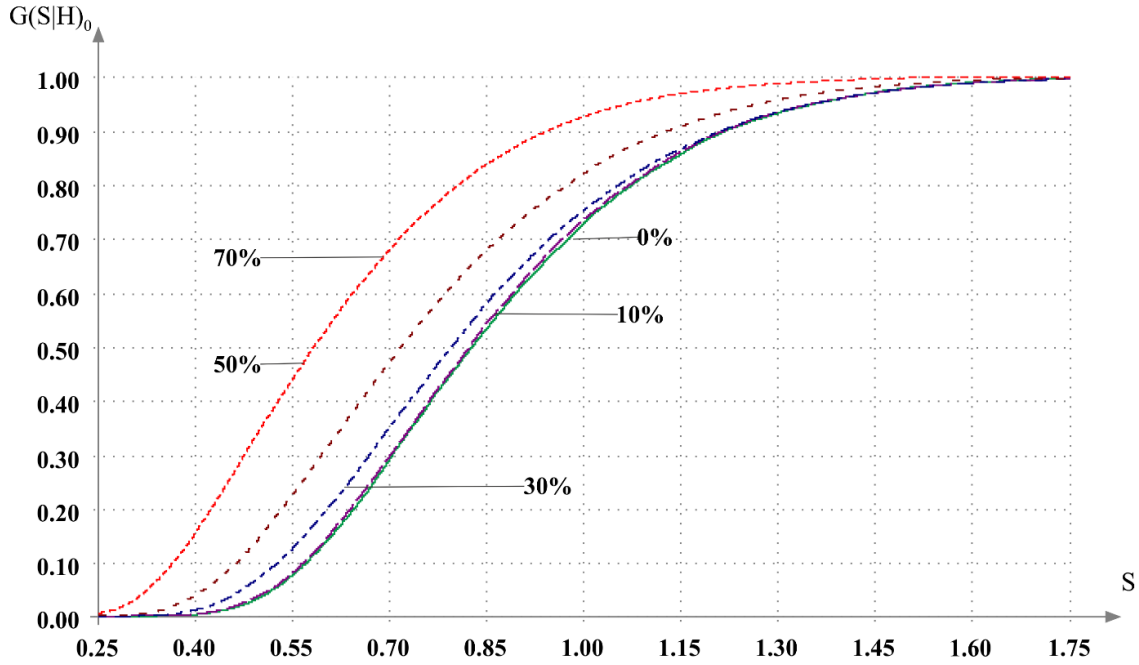


Рисунок 13 – Распределения статистики Колмогорова при проверке простой гипотезы по цензурированным справа данным для разных степеней цензурирования (II тип цензурирования). $n = 200$

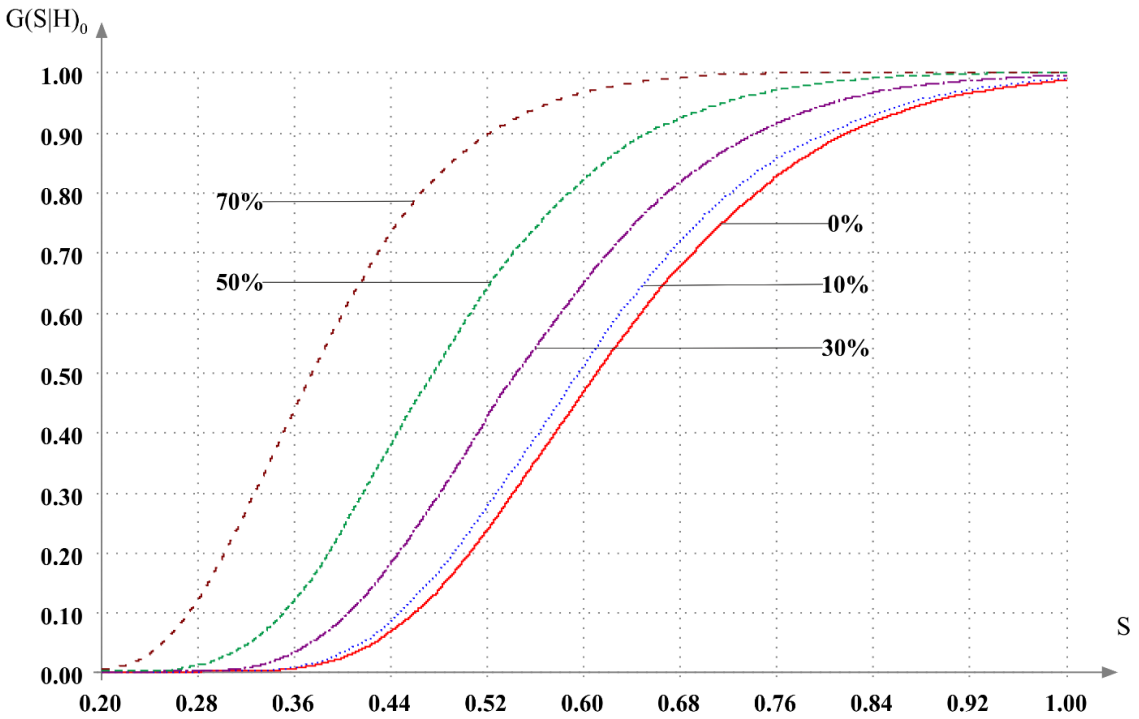


Рисунок 14 – Распределения статистики Колмогорова при проверке сложной гипотезы по цензурированным справа данным для разных степеней цензурирования, (II тип цензурирования). $n = 200$

Из рисунков 13 и 14 следует, что при увеличении степени цензурирования функции распределения в обоих случаях увеличивается смещение влево от функции распределения статистики для полных данных. Аналогичные результаты были получены для распределения статистик критериев Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

Проведем аналогичное исследование для III типа цензурирования. Отказы моделировались из распределения Вейбулла с параметрами $\theta_0 = 2$, $\theta_1 = 2$. Значения параметров распределения моментов цензурирования были подобраны таким образом, что были получены желаемые степени цензурирования. Значения параметров распределений приведены в таблице 18.

Таблица 18 – Параметры масштаба и формы для распределений, определяющих моменты цензурирования

Процент цензурирования	Распределение Вейбулла	
	Параметр масштаба	Параметр формы
10	16	1
30	4.8	1
50	2	1.8
70	1.2	1

На рисунках 15 и 16 представлены распределения статистики Колмогорова при проверке простых и сложных гипотез, соотвественно.

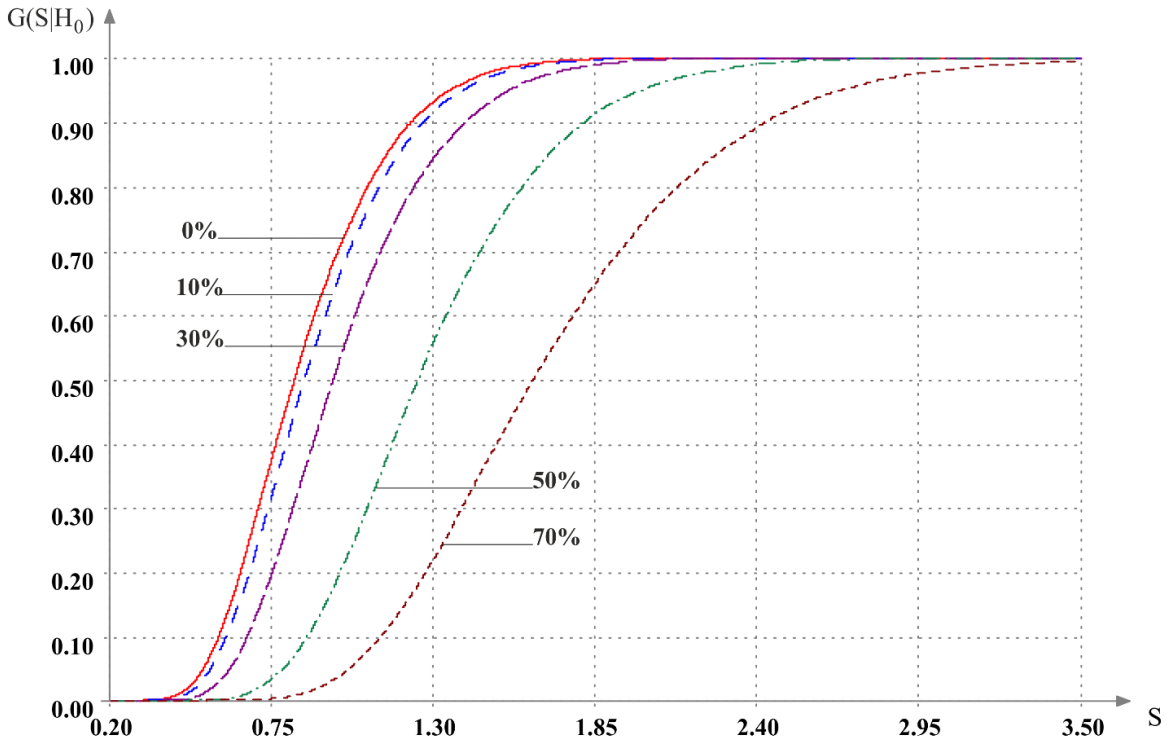


Рисунок 15 – Распределения статистики Колмогорова при проверке простой гипотезы по цензурированным справа данным для разных степеней цензурирования (III тип цензурирования). $n = 200$

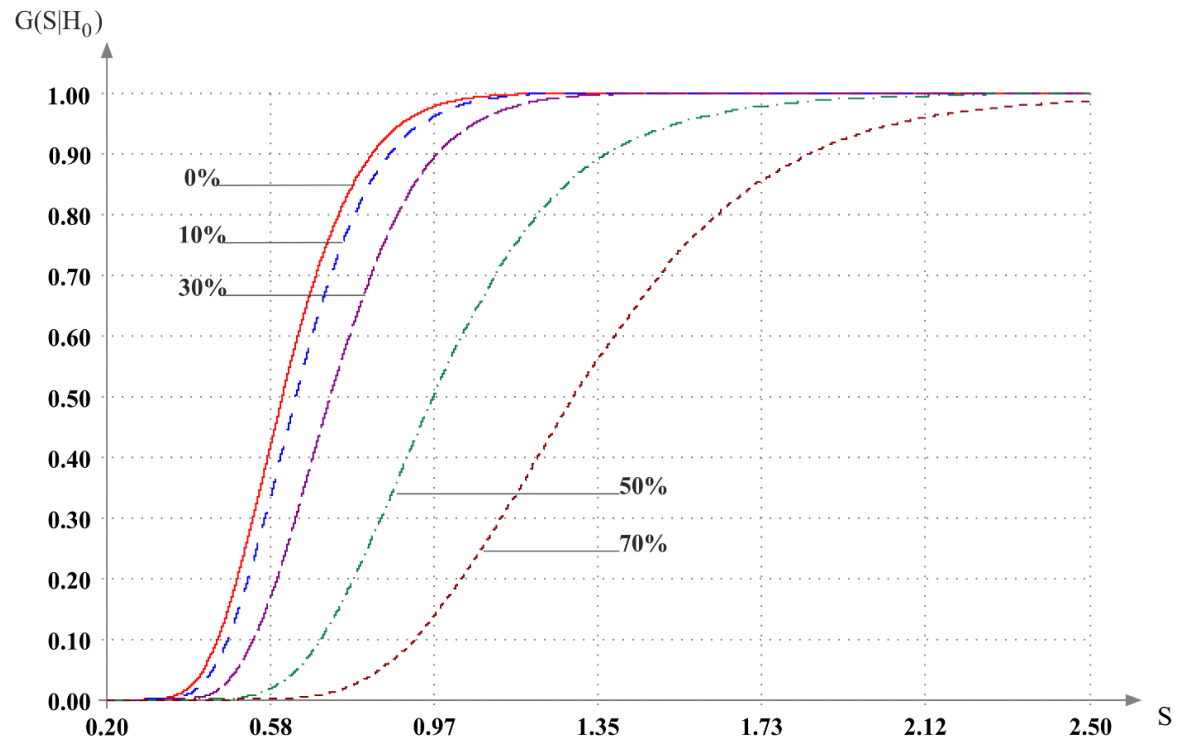


Рисунок 16 – Распределения статистики Колмогорова при проверке сложной гипотезы по цензурированным справа данным для разных степеней цензурирования (III тип цензурирования). $n = 200$

Как можно заметить, и в случае проверки простых гипотез, и в случае проверки сложных гипотез, при увеличении процента неполных наблюдений в выборках, для данной пары гипотез статистики критериев согласия смещаются вправо.

4.3. Исследование мощности критериев согласия по усеченным слева и цензурированным справа данным

Мощность критериев согласия может зависеть от многих факторов, к примеру, от степени усечения и цензурирования, процента наблюдений из усеченного распределения. В качестве конкурирующих гипотез рассматривается пара H_0: Вейбулл(2, 2, 0) и H_1: Гамма(0.5577, 3.1215, 0). Для исследования было смоделировано $N = 16600$ выборок объемом $n = 200$. При проверке сложных гипотез для оценивания параметров распределения используется метод максимального правдоподобия.

В таблице 19 приведены мощности критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга по выборкам усеченных наблюдений в зависимости от степени усечения и процента наблюдений из усеченного распределения.

Таблица 19 – Оценки мощности непараметрических критериев согласия по выборкам усеченных наблюдений

Степень усечения	Процент усеченных наблюдений			
	25%	50%	75%	100%
Критерий Колмогорова				
0.1	0.365	0.306	0.222	0.174
0.3	0.357	0.287	0.227	0.128
0.5	0.328	0.260	0.199	0.120
0.7	0.285	0.207	0.177	0.106
Критерий Крамера-Мизеса-Смирнова				
0.1	0.440	0.366	0.255	0.199
0.3	0.426	0.340	0.243	0.132
0.5	0.403	0.303	0.221	0.115
0.7	0.356	0.257	0.192	0.109
Критерий Андерсона-Дарлинга				
0.1	0.489	0.389	0.267	0.213

0.3	0.462	0.436	0.235	0.131
0.5	0.445	0.307	0.217	0.115
0.7	0.418	0.275	0.188	0.105

Мощность критериев согласия для выборок, содержащих усеченные наблюдения падает при увеличении степени усечения и процента наблюдений из усеченного распределения. Из рассмотренных критериев наименьшая мощность оказалась у критерия Колмогорова. Критерии Крамера-Мизеса-Смирнова и Андерсона-Дарлинга показали примерно одинаковую мощность для процента усеченных наблюдений в выборке больше 50%, однако при проценте усечения равным 25% мощность критерия Андерсона-Дарлинга оказалась выше.

Распределение статистик критериев согласия по неполным выборкам зависит непосредственно от степени и типа цензурирования. В таблице 20 приведены значения мощности критериев согласия для простой и сложной гипотез по выборкам с II типом цензурирования в зависимости от степени цензурирования.

Таблица 20 – Оценки мощности непараметрических критериев согласия по цензурированным выборкам (II тип)

Средняя степень цензурирования	Критерий Колмогорова	Критерий Крамера-Мизеса-Смирнова	Критерий Андерсона-Дарлинга
Простая гипотеза			
0%	0.331	0.332	0.617
10%	0.345	0.342	0.319
30%	0.347	0.356	0.332
50%	0.351	0.324	0.291
70%	0.239	0.184	0.158
Сложная гипотеза			
0%	0.422	0.520	0.597
10%	0.339	0.397	0.432
30%	0.275	0.325	0.329
50%	0.230	0.275	0.257
70%	0.191	0.224	0.192

Как видно из таблицы 20 с ростом степени усечения мощность критериев согласия уменьшается как при проверке простой гипотезы, так и сложной.

4.4. Выводы

В данной главе проведено исследование непараметрических критериев согласия при проверке простых и сложных гипотез для усеченных слева и цензурированных справа данных в зависимости от процента усеченных наблюдений выборке, степени усечения и цензурирования. В результате были сделаны следующие выводы:

1) при проценте усеченных наблюдений в выборке равным 25% с увеличением степени усечения распределение статистики Колмогорова заметно отклоняется от предельного закона распределения, однако в случае, при проценте усеченных наблюдений в выборке равным 100% наблюдается обратная закономерность.
2) с увеличением степени усечения возрастает отклонение распределения статистик от предельного распределения.
3) с увеличением степени цензурирования функции распределения статистик критериев согласия сдвигаются влево для II типа цензурирования и вправо для III типа цензурирования.
4) было показано, что мощность критериев согласия падает при увеличении процента усеченных наблюдений в выборке и степени усечения при увеличении степени цензурирования для цензурированной справа выборки.

5. Описание разработанных программ и примеры статистического анализа усеченных слева и цензурированных справа данных

В рамках исследовательской работы была выполнена реализация программных модулей, на основе программной системы LiTiS (*Life Time Statistics*), используемой для решения задач в области статистического анализа, обработки данных о продолжительности жизни, построения вероятностных моделей надежности и выживаемости.

Программные модули были разработаны с целью решить такие задачи:

- моделирование цензурированной I или III типом выборки наблюдений из усеченного слева распределения;
- получение и максимизация функции правдоподобия при наличии усеченных наблюдений;
- построение оценки Каплана-Мейера при наличии усеченных наблюдений;
- расчет расстояния Колмогорова между теоретической функцией распределения и оценкой Каплана-Мейера;
- вычисление статистик непараметрических критериев согласия для выборок усеченных слева наблюдений.

5.1. Алгоритмы моделирования цензурированных справа выборок усеченных слева наблюдений

Рассмотрим подробнее, как происходит процесс моделирования выборки цензурированной I или III типом.

В качестве входных параметров при использовании I типа необходимо указать:

- объем выборки n;
- распределение моментов отказа $F(x)$;
- значение момента времени начала эксперимента (момент усечения) T_0;
- значение момента времени конца эксперимента (момент цензурирования) T;
- массив моментов начала эксплуатации объектов D_i.

В качестве выходного результата программного модуля получаем цензурированную справа выборку усеченных слева наблюдений.

Алгоритм моделирования выборки, цензурированной I типом, заключается в следующих шагах:

1. Используя метод обратной функции, генерируем случайную величину, распределенную по закону $F(x)$ – продолжительность жизни (работы) i-го объекта и обозначим его как X_i;

2. Если конец эксплуатации после начала эксперимента, то переходим на пункт 3, иначе на пункт 1.

3. Если эксперимент начался после начала эксплуатации, то наблюдение является усеченным, а значение времени усечения равно разнице между моментом усечения и началом эксплуатации.

$$\tau_i = T_0 - D_i.$$

4. Если отказ произошел до момента цензурирования, то наблюдение полное $\delta_i = 1$, иначе $X_i = T - D_i$, $\delta_i = 0$.

5. Если не достигнут необходимый размер выборки $i < n$, то переходим на пункт 1, иначе выполнение алгоритма закончено, на выходе получена смоделируемая выборка.

В качестве входных параметров при использовании III типа необходимо указать:

- объем выборки n;
- распределение моментов отказа $F(x)$;
- распределение моментов цензурирования $F_c(x)$;
- массив моментов усечения τ_i.

Алгоритм моделирования выборки, цензурированной III типом, заключается в следующих шагах:

1. Используя метод обратной функции, генерируем случайную величину, распределенную по закону $F(x)$ – момент отказа i-го объекта и обозначим его как T_i;

2. Используя метод обратной функции, генерируем случайную величину, распределенную по закону $F_c(x)$ – момент цензурирования i-го объекта и обозначим его как C_i;

3. Присваиваем значение $X_i = \min(T_i, C_i)$.

Если наблюдение полное $X_i = T_i$, то $\delta_i = 1$, иначе $\delta_i = 0$.

4. Если не достигнут необходимый размер выборки $i < n$, то переходим на пункт 1, иначе выполнение алгоритма закончено, на выходе получена смоделируемая выборка.

5.2. Тестирование программных модулей

Проведем тестирование программных модулей для вычисления функции правдоподобия, оценки Каплана-Мейера,

Для проверки на правильную реализацию функции правдоподобия был выполнен тест, входные данные приведены таблице 21.

Таблица 21 – Входные данные для теста функции правдоподобия

X_i	δ_i	τ_i
10	1	0
11	1	3
12	0	5
13	0	6
14	0	6
15	1	0

На рассматриваемых данных значение функции правдоподобия равно -32.5794. Тест показал, что результат, полученный программным модулем, совпадает с фактическим значением, поэтому можно сделать вывод, что программный модуль реализован правильно.

Тестирование программного модуля, считающего оценку Каплана-Мейера, проводилось на тесте, представленном в таблице 22. Моменты времени, в которые проводились замеры приведены в таблице 23.

Таблица 22 – Входные данные для теста оценки Каплана-Мейера

X_i	δ_i	τ_i
6	0	3
6	1	2
6	1	0
6	1	1
7	1	2
9	0	8
10	0	5
10	1	9
11	0	0
13	1	10
16	1	0
17	0	0
19	0	1
20	0	3
22	1	15
23	1	0
25	0	20
32	0	3
32	0	0
34	0	0
35	0	33

Таблица 23 – Результаты теста оценки Каплана-Мейера

a_i	Фактическое значение	Расчетное значение
6	0.8	0.8
7	0.727272727	0.727272727
10	0.661157025	0.661157025
13	0.587695133	0.587695133
16	0.522395674	0.522395674
22	0.435329728	0.435329728
23	0.348263783	0.348263783

По результатам теста Каплана-Мейера, которые приведены в таблице 23, видно, что программный модуль, рассчитывающий множительную оценку в некоторые моменты времени, работает верно, так как результат совпадает со значениями, полученные не программным путем.

Для тестирования программного модуля расчета значения статистики непараметрических критериев согласия при проверке сложной гипотезы для усеченных слева данных, представленных в таблице 24.

Таблица 24 – Входные данные для теста критериев согласия

X_i	δ_i	τ_i
1.264160	1	0.649186
3.490003	1	0.649186
2.061245	1	0.649186
1.917127	1	0.649186
1.151289	1	0.649186
2.778297	1	0.649186
1.812017	1	0.649186
1.543187	1	0.000000
2.869661	1	0.000000
5.787537	1	0.000000
1.399435	1	0.000000
2.410444	1	0.000000
1.772045	1	0.000000
1.408563	1	0.000000
2.024336	1	0.000000

Таблица 25 – Результаты теста критериев согласия

Название критерия	Фактическое значение	Расчетное значение
Колмогоров	0.73645	0.73645
Крамера-Мизеса-Смирнова	0.10140	0.10140
Андерсона-Дарлинга	0.67457	0.67457

При проверке гипотезы H_0: Вейбулл(2, 2, 0) были получены значения расчета статистик непараметрических критериев согласия Колмогорова, ω^2 Крамера-Мизеса-Смирнова и Ω^2 Андерсона-Дарлинга, которые представлены в таблице 25. Из этих результатов видно, что полученные значения совпадают с фактическими, значит, программный модуль работает правильно.

5.3. Графический интерфейс

В рамках данной работы был улучшен графический интерфейс программной системы LiTiS для взаимодействия пользователя с выборками усеченных слева случайных величин.

Для работы с усеченными данными были реализованы функции открытия, редактирования и сохранения выборки, содержащей усеченные наблюдения и построение графика непараметрической оценки Каплана-Мейера. На рисунке 17 представлен скриншот приложения после открытия выборки.

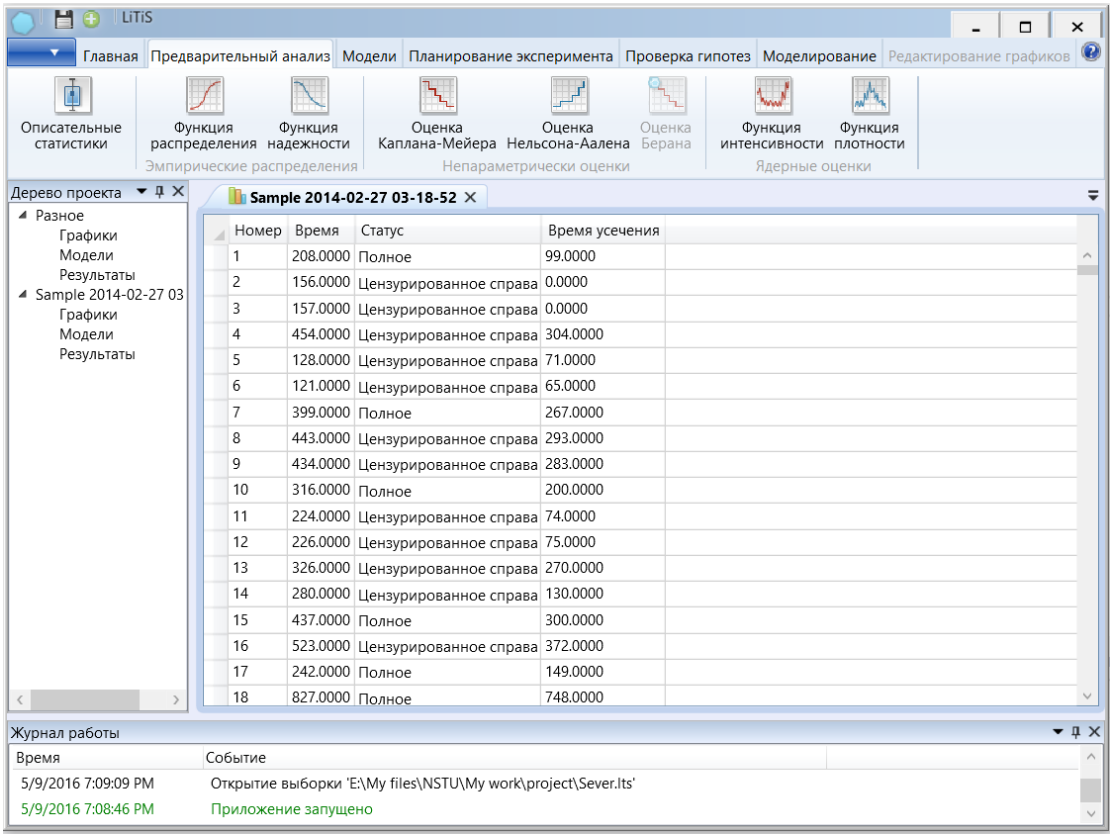


Рисунок 17 – Снимок пользовательского интерфейса

Окно с выборкой представляет собой таблицу из трех или четырех столбцов: номер наблюдения, время, статус (полное, цензурированное справа или цензурированное слева) и время усечения. В данном окне пользователь может просмотреть и изменить выборку. После изменений новую выборку можно сохранить в файл. Также необходимо отметить, что при открытии выборки, не содержащей усеченные наблюдения, столбец «Время усечения» не будет отображен и работа с ней будет осуществляться, как с данными без учета усечения.

Используя «Мастер анализа», пользователь может посчитать оценку максимального правдоподобия по выборке в соответствии с выбранным распределением или с помощью идентификации распределения. Пример пользовательского интерфейса представлен на рисунке 18.

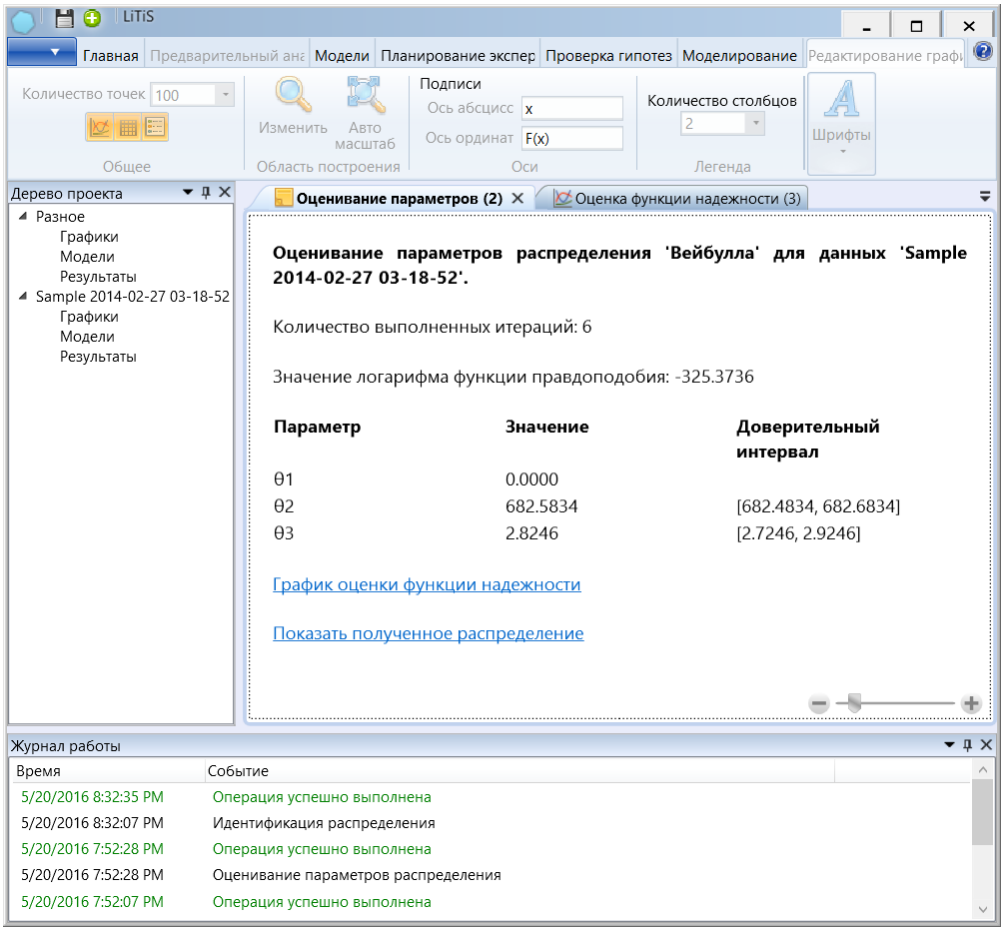


Рисунок 18 – Оценивание параметров распределения

В новой вкладке «Оценивание параметров» приведены результаты оценивания, а также присутствует возможность построения графика оценки функции надежности. Воспользовавшись ей, можно получить график оценки Каплана-Мейера и оценки функции надежности. Пример подобного изображения представлен на рисунке 19.

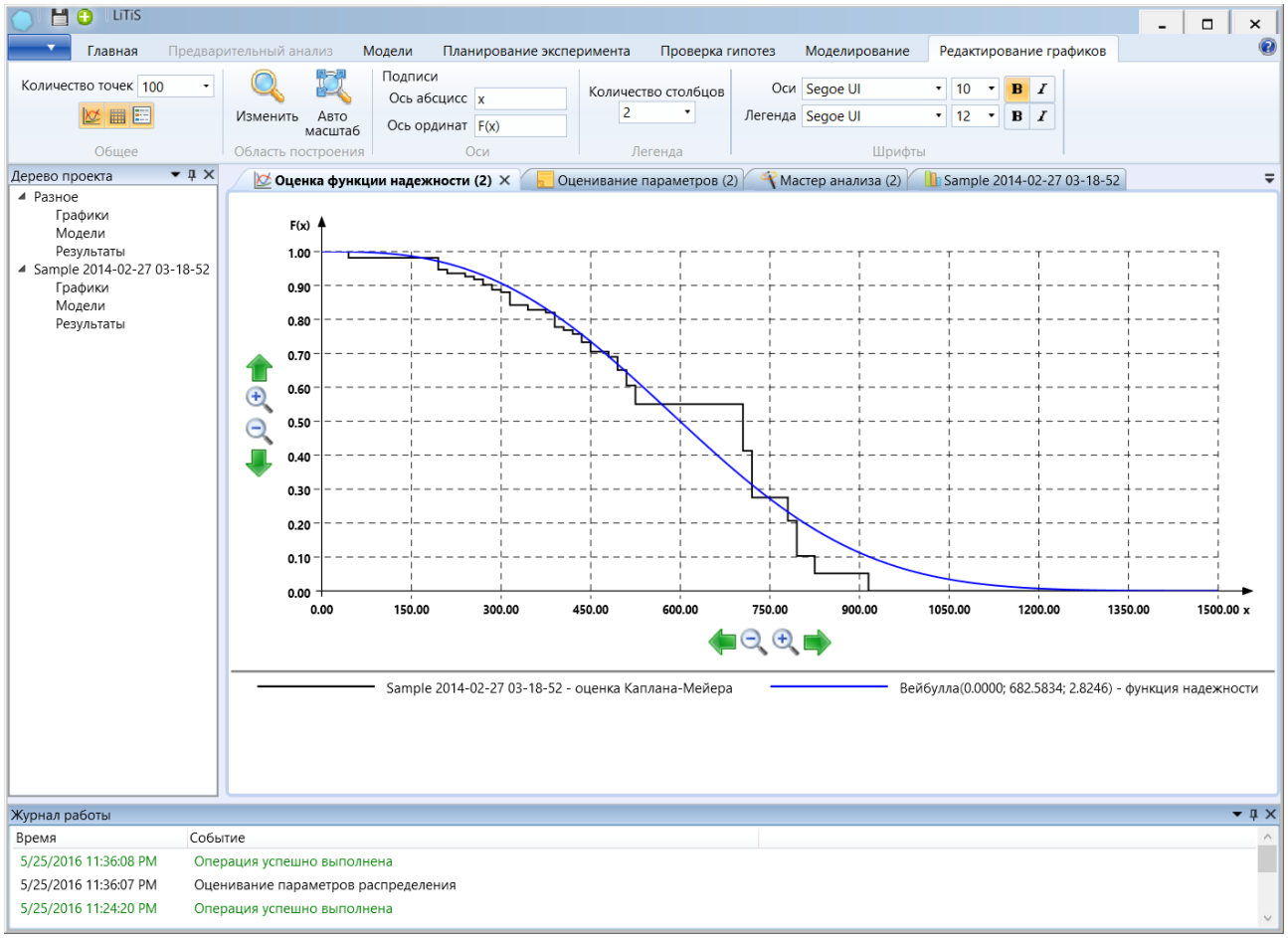


Рисунок 19 – Построение оценки Каплана-Мейера

При построении графика оценки Каплана-Мейера используется (18), а для построения функции надежности используются полученные ранее оценки максимального правдоподобия параметров распределения.

5.4. Анализ данных о выживаемости трудящихся промышленных предприятий Севера

С 1991 года по 2001 год сотрудниками научного центра клинической и экспериментальной медицины СО РАМН был проведен скрининг среди работников акционерной компании «АЛРОСА», проживающих на Севере [18]. На основе полученных данных было проведено исследование, направленное на выявление основных действующих факторов риска хронических неинфекционных заболеваний трудящихся. Данные приведены в приложении Б.

Проведем анализ продолжительности жизни северян, основываясь на оценке Каплана-Мейера. Началом эксперимента будем считать дату приезда человека на Север, а окончанием – дату последнего наблюдения или смерти. В связи с тем, что большинство людей приехало на Север до начала скрининга выборка будет содержать большой процент усеченных наблюдений.

Всего выборка содержит 711 наблюдений. Её процент усеченных наблюдений соответствует величине 97%, а степень цензурирования – 93%. Все расчеты времен жизни и усечения будут проводится в пересчете на месяцы.

Было выдвинуто предположение, что времена жизни описаны с помощью закона Вейбулла. С помощью метода максимального правдоподобия были получены оценки параметров распределения $\theta_0 = 682.5834$, $\theta_1 = 2.8246$.

При помощи полученной функции распределения можно рассчитать прогноз вероятности выживания. Результаты приведены в таблице 26.

Таблица 26 – Вероятность выживания для разных замеров по времени

Замер по времени, лет	Вероятность выживания
10	0.99265
20	0.94912
30	0.84863
40	0.69081
50	0.49921
60	0.31264
70	0.16578

Посмотрим, как влияет на теоретическую функцию распределения и оценку Каплана-Мейера учет наличия усеченных наблюдений. Для этого, используя те же данные, составим выборку в которой не будет хранится информация, содержащая время усечения. На рисунке 20 представлены два вида графиков, учитывающих и не учитывающих усечения.

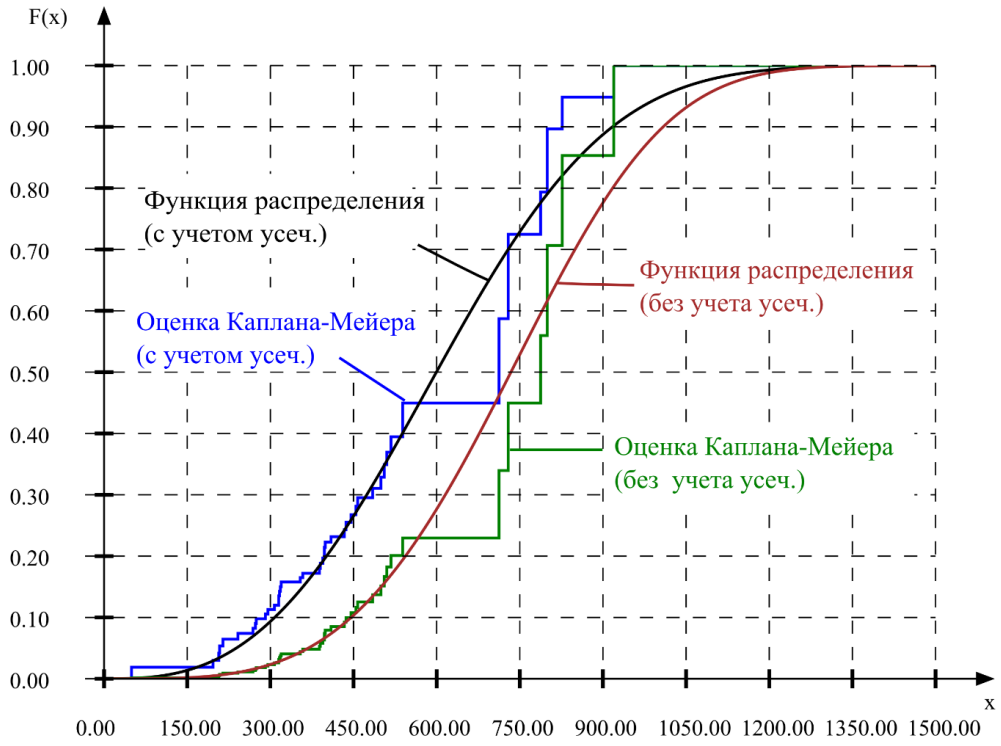


Рисунок 20 – Теоретическая функция распределения Вейбулла и оценка Каплана-Мейера для примера с данными о времени жизни северян с учетом и без учета усечения

Как видно из рисунка 20, при учете процесса усечения данных графики распределения и оценки Каплана-Мейера смещаются влево относительно аналогичных графиков, не учитывающих усечения. Таким образом, если факт усечения не будет учтен при построении вероятностной модели, вероятность выживаемости будет завышенной. Так, например, с вероятностью 0.95 для случая, когда учитывается усечение данных, продолжительность жизни составляет 208 месяцев (17 лет 4 месяца), а для варианта без учета усечения эта величина больше и составляет 359 месяцев (29 лет 11 месяцев).

5.5. Анализ данных о времени безотказной работы

В работе [6] был представлен пример на основе данных о времени работы машин. Таблица с данными приведена в приложении А. Этот эксперимент начался в 1980 году, а закончился в 2008. Необходимо учесть, что некоторые объекты начали эксплуатироваться до начала наблюдения, таким образом примем за время усечения разницу между началом их установки и 1980 годом.

В результате имеем, что из 100 объектов под наблюдением 40% являются усеченными, а 50% цензурированными.

Выдвинем предположение, что времена жизни объектов распределены по закону Вейбулла с оцененными методом максимального правдоподобия параметром масштаба 34.3948 и параметром формы 2.9309. Рассчитаем оценку Каплана-Мейера для исходных данных и сравним её с теоретической.

При помощи полученной функции распределения можно рассчитать прогноз вероятности безотказной работы. Результаты приведены в таблице 27.

Таблица 27 – Вероятность безотказной работы для разных замеров по времени

Замер по времени	Вероятность безотказной работы
10	0.97359
20	0.81537
30	0.51178
40	0.21085
50	0.05001

На рисунке 21 представлены графики оценки Каплана-Мейера и функции распределения для двух случаев, учитывающих и не учитывающих усечение.

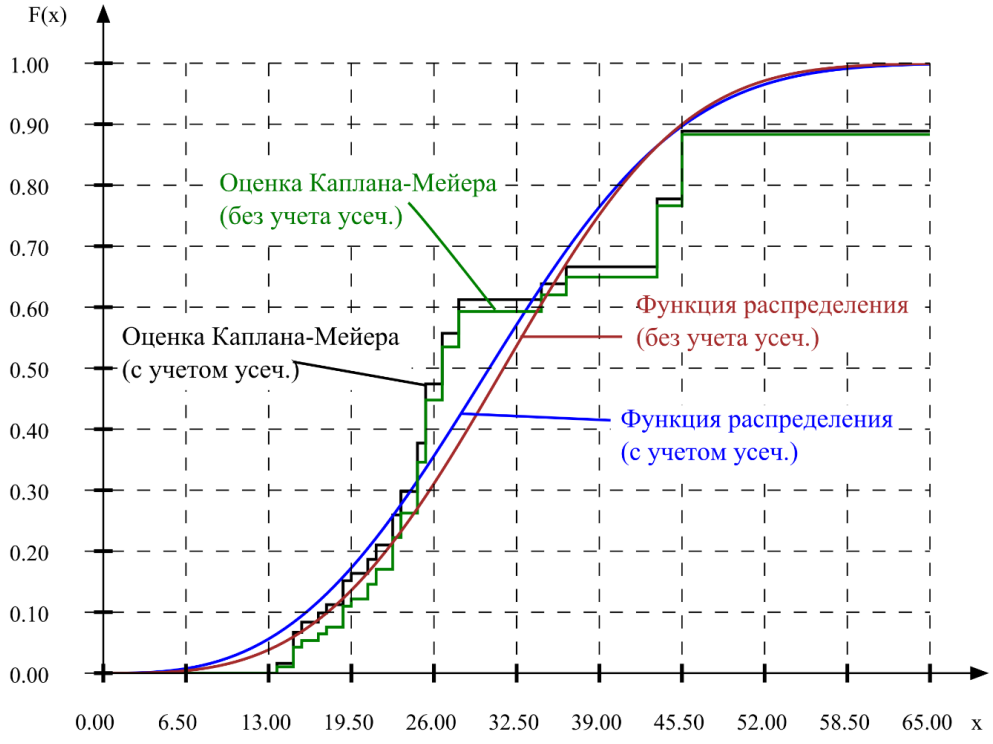


Рисунок 21 – Теоретическая функция распределения Вейбулла и оценка Каплана-Мейера для примера с данными о времени работы машин с учетом и без учета усечения

По графикам на рисунке 21 видно, что на этих экспериментальных данных графики оценки Каплана-Мейера и функция распределения с учетом и без учета усечения различаются не так сильно как в рассмотренном примере в пункте 4.2. Также из графика множительной оценки можно сделать вывод, что с вероятностью 0.95 продолжительность работы объекта составляет 14 лет.

5.6. Выводы

В этой главе было проведено тестирование разработанных программных модулей, в результате проведенных тестов можно сделать вывод о том, что функции моделирования выборки, цензурированной I или III типом, вычисления статистики непараметрических критериев согласия, построения оценки Каплана-Мейера и функции правдоподобия реализованы верно.

В результате проведенных в этой главе исследований на примере экспериментальных данных можно сделать следующий вывод, что в случае, когда не учитывается факт усечения времени работы объекта значение вероятности выживаемости будет завышенным по сравнению со случаем, когда учитывается усечение. Однако, на разных экспериментальных данных величина разницы между функциями распределения с учетом и без учета усечения разная.

Заключение

В соответствии с целью данной работы получены следующие основные результаты:

1) разработаны программные модули, позволяющие моделировать выборки, содержащие усеченные слева цензурированные справа наблюдения при использовании I или III типа цензурирования, вычислять оценки Каплана-Мейера и оценки максимального правдоподобия;
2) процент усеченных наблюдений в выборке влияет на смещение и выборочную дисперсию наименьшим образом, однако, степень усечения

оказывает большее влияние в сравнении со степенью цензурирования при рассмотрении свойств оценок максимального правдоподобия;

3) при изменении степени усечения или цензурирования относительная эффективность полученных оценок уменьшается.
4) проведенное исследование зависимости множительной оценки от процента усеченных наблюдений показало, что при проценте усечения меньше 100% полученная оценка Каплана-Мейера расположена вблизи к функции распределения для неусеченного случая;
5) было получено, что расстояние Колмогорова между теоретической функцией распределения и оценкой Каплана-Мейера не зависит от выбора точки усечения. При сравнении разных способов цензурирования было выявлено, что при использовании I типа цензурирования точность оценки Каплана-Мейера выше, чем при аналогичном значении процента неполных наблюдения при использовании случайного способа цензурирования;
6) при проценте усеченных наблюдений равным 100%, полученные распределения статистик критерия проверки сложной гипотезы вне зависимости от значения степени усечения оказываются смещенными влево от предельного распределения, при остальных процентах усеченных наблюдений в выборке – вправо;
7) при увеличении степени цензурирования функции распределения статистик непараметрических критериев согласия увеличивают свое смещение относительно предельного закона распределения;
8) мощность критериев согласия падает при увеличении степени или процента усеченных наблюдений в выборке;
9) мощность критериев согласия падает при увеличении степени цензурирования для цензурированной справа выборки.

Список литературы

1) Cohen, A. C. Truncated and censored samples: theory and applications / A. C. Cohen. – New York : Marcel Dekker, 1991. – p. 328.
2) Balakrishnan, N. Order Statistics and Inference: Estimation / N. Balakrishnan, A.C. Cohen. – Boston : Academic Press, 1991.
3) Meeker, W.Q. Statistical Methods for Reliability Data / W.Q. Meeker, L.A. Escobar. – New York : John Wiley & Sons, 1998. – p. 712.
4) Su, Y.-R. Modeling left-truncated and right-censored survival data with longitudinal covariates / Y.-R. Su, J.-L. Wang ; Institute of Mathematical Statistics // The Annals of Statistics. – 2012. – Vol. 40, No. 3. – p. 1465-1488.
5) Bagdonavicius, V. Nonparametric Tests for Censored Data / V. Bagdonavicius, J. Kruopis, M.S. Nikulin. – London : Wiley-ISTE, 2010. – 233 c.
6) Balakrishnan, N. Left truncated and right censored Weibull data and likelihood infence with an illusrtation / N. Balakrishnan, D. Mitra // Computational Statistics and Data Analysis. – 2012. – 56. – p. 4011-4012.
7) Balakrishnan, N. Likelihood inference for lognormal data with left truncation and right censoring with an illustration / N. Balakrishnan, D. Mitra // Journal of Statistical Planning and Inference. – 2011. – 141. – p. 3536-3553.
8) Li, J. Cox Model Analysis with the Dependently Left / J. Li // Mathematics Theses. – Georgia State University, Atlanta, GA,, 2010. – Paper 88. – C. 0-0.
9) Pan, W. A Nonparametric Estimator of Survival Functions / W. Pan, R. Chappell // Lifetime Data Analysis. – Boston, 1998. – 4. – p. 187-202.
10) Кокс, Д.Р. Анализ данных типа времени жизни / Д.Р. Кокс, Д. Оукс, пер. с англ. О.В. Селезнева – М. : Финансы и статистика, 1988. – 191 с.
11) Постовалов С. Н. Математическая статистика [Электронный ресурс] : учебное пособие / С. Н. Постовалов, Е. В. Чимитова, В. С. Карманов ; Новосиб. гос. техн. ун-т. – Новосибирск. – 2012.
12) Лемешко Б. Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход :

монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. // – Новосибирск : Изд-во НГТУ. – 2011. – 888 с.

13) ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения.

14) Лемешко Б.Ю., Постовалов С.Н., Чимитова Е.В. К оцениванию параметров законов распределений и проверке гипотез по цензурированным выборкам // Тр. V международной конференции "Актуальные проблемы электронного приборостроения" АПЭП-2000. Новосибирск, 2000. - Т. 7. - С. 188-191.

15) Лемешко, Б.Ю. Модифицированные критерии согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для случайно цензурированных выборок. Ч. 2 / Б.Ю. Лемешко, Е.В. Чимитова, М.А. Ведерникова // Научный вестник НГТУ. – 2013. – № 1(50). – С. 3-16.

16) Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие. – Новосибирск: Изд-во НГТУ, 2004. – 119 с.

17) Цой, Е.Б. Моделирование и управление в экономике (часть I) : курс лекций / Е.Б. Цой, И.В. Самочернов. – Новосибирск : Изд-во НГТУ, 2003. – 104 с.

18) Кейль, В.Р. Здоровье трудящихся промышленных предприятий Севера: Стратегия разработки оздоровительных программ / В.Р. Кейль, И.Ю. Кузнецова, И.М. Митрофанов и др. – Новосибирск : Наука, 2005. – 231 с.

19) Hong, Y., Meeker, W. Q., and McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. The Annals of Applied Statistics, 3:857 – 879

20) Mandel, M. (2007). Censoring and truncation - highlighting the differences. *The American Statistician.* 61 (4):321-324.Article at Am. Stat. website. Full Text.

21) Hjort N. L. On Inference in Parametric Survival Data / N. L. Hjort // International Statistical Review. – 1992. – Vol. 60. – № 3. – P. 355-387

22) Koziol J. A. A Cramer-von Mises statistic for randomly censored data / J. A. Koziol, S. B. Green // Biometrika. – 1976. – Vol.63. – № 3. – P. 465-474.

23) Nair V. Plots and tests for goodness of fit with randomly censored data / V. Nair // Biometrika. – 1981. – Vol. 68. – P. 99-103.

24) Reineke D. Estimation of hazard, density and survival functions for randomly censored data / D. Reineke, J. Crown // Journal of Applied Statistics. –

2004. – Vol. 31. – № 10. – P. 1211-1225

25) Chimitova E. Application of classical Kolmogorov, Cramer-von MisesSmirnov and Anderson-Darling tests for censored samples / E. Chimitova, H. Liero, M. Vedernikova // Proceedings of the International Workshop AMSA. – Novosibirsk: Publ. of NSTU, 2011. – P. 176-185

26) Nikulin M. Nonparametric goodness-of-fit tests for censored data / M. Nikulin, B. Lemeshko, E. Chimitova, A. Tsivinskaya // Proceedings of the 7th international conference on “Mathematical methods in reliability”: Theory. Methods. Applications, Beijing, China. – 2011. – P. 817-823

27) НТИ

28) ПОПОВСКАЯ

29)

Приложение А. Данные об отказах машин

Таблица А.1 – Выборка отказов машин, приведенная в работе [6]

Номер	Время установки	Индикатор усечения	Время усечения	Время отказа	Индикатор цензурирования	Время жизни
1	1984	1	*	*	0	24
2	1990	1	*	2001	1	11
3	1983	1	*	2002	1	19
4	1981	1	*	2000	1	19
5	1985	1	*	*	0	23
6	1991	1	*	*	0	17
7	1982	1	*	*	0	26
8	1990	1	*	*	0	18
9	1983	1	*	1999	1	16
10	1992	1	*	*	0	16
11	1983	1	*	*	0	25
12	1989	1	*	*	0	19
13	1985	1	*	*	0	23
14	1982	1	*	*	0	26
15	1983	1	*	*	0	25
16	1981	1	*	*	0	27
17	1985	1	*	*	0	23
18	1981	1	*	*	0	27
19	1988	1	*	2002	1	14
20	1983	1	*	*	0	25
21	1984	1	*	*	0	24
22	1989	1	*	*	0	19
23	1988	1	*	*	0	20
24	1982	1	*	*	0	26
25	1981	1	*	*	0	27
26	1986	1	*	*	0	22
27	1987	1	*	*	0	21
28	1990	1	*	1997	1	7
29	1980	1	*	1996	1	16
30	1980	1	*	*	0	28
31	1981	1	*	*	0	27
32	1983	1	*	1997	1	14
33	1980	1	*	*	0	28
34	1984	1	*	*	0	24
35	1982	1	*	*	0	26
36	1980	1	*	*	0	28
37	1985	1	*	2007	1	22
38	1993	1	*	*	0	15
39	1983	1	*	*	0	25
40	1980	1	*	*	0	28
41	1981	1	*	2001	1	20
42	1989	1	*	*	0	19
43	1993	1	*	*	0	15
44	1983	1	*	*	0	25
45	1993	1	*	*	0	15
46	1987	1	*	*	0	21
47	1994	1	*	*	0	14
48	1985	1	*	2007	1	22
49	1981	1	*	*	0	27
50	1983	1	*	2004	1	21
51	1982	1	*	*	0	26
52	1981	1	*	*	0	27

Продолжение таблицы А.1

Номер	Время установки	Индикатор усечения	Время усечения	Время отказа	Индикатор цензурирования	Время жизни
53	1986	1	*	*	0	22
54	1980	1	*	1990	1	10
55	1980	1	*	1994	1	14
56	1982	1	*	*	0	26
57	1990	1	*	2008	1	18
58	1985	1	*	*	0	23
59	1983	1	*	*	0	25
60	1982	1	*	*	0	26
61	1963	0	17	1996	1	33
62	1963	0	17	2001	1	38
63	1961	0	19	1998	1	37
64	1961	0	19	1992	1	31
65	1960	0	20	1984	1	24
66	1964	0	16	2004	1	40
67	1961	0	19	1994	1	33
68	1977	0	3	1998	1	21
69	1963	0	17	1987	1	24
70	1960	0	20	1991	1	31
71	1961	0	19	1983	1	22
72	1964	0	16	1995	1	31
73	1963	0	17	1998	1	35
74	1961	0	19	2001	1	40
75	1960	0	20	1988	1	28
76	1974	0	6	2006	1	32
77	1978	0	2	1995	1	17
78	1962	0	18	*	0	46
79	1963	0	17	1993	1	30
80	1960	0	20	1998	1	38
81	1962	0	18	2007	1	45
82	1960	0	20	1990	1	30
83	1962	0	18	1980	1	18
84	1961	0	19	1981	1	20
85	1964	0	16	1989	1	25
86	1964	0	16	1987	1	23
87	1960	0	20	2006	1	46
88	1961	0	19	1992	1	31
89	1964	0	16	*	0	44
90	1963	0	17	1991	1	28
91	1973	0	7	*	0	35
92	1964	0	16	*	0	44
93	1972	0	8	1984	1	12
94	1962	0	18	2007	1	45
95	1963	0	17	1997	1	34
96	1964	0	16	1987	1	23
97	1964	0	16	2002	1	38
98	1971	0	9	*	0	37
99	1965	0	15	1990	1	25
100	1962	0	18	1994	1	32

Приложение Б. Данные о времени жизни работников на Севере

Таблица Б.1 – Выборка данных о времени жизни работников на Севере акционерной компании «АЛРОСА» [13]

i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i
1	208	1	99	47	352	0	203	93	314	0	164
2	156	0	0	48	316	1	241	94	360	0	210
3	157	0	0	49	319	1	181	95	181	0	32
4	454	0	304	50	158	0	0	96	206	0	56
5	128	0	71	51	213	0	156	97	518	1	386
6	121	0	65	52	261	0	111	98	503	0	353
7	399	1	267	53	136	0	80	99	281	2	272
8	443	0	293	54	433	0	282	100	308	1	235
9	434	0	283	55	386	0	236	101	299	0	149
10	316	1	200	56	115	0	59	102	492	0	342
11	224	0	74	57	319	0	305	103	382	0	241
12	226	0	75	58	251	0	101	104	411	1	326
13	326	0	270	59	400	2	393	105	262	2	223
14	280	0	130	60	798	2	757	106	252	0	103
15	437	1	300	61	69	2	50	107	446	0	296
16	523	0	372	62	391	1	345	108	477	0	327
17	242	1	149	63	183	0	127	109	329	0	179
18	827	1	748	64	67	0	0	110	281	0	132
19	265	0	209	65	439	0	289	111	519	0	370
20	147	0	0	66	166	0	110	112	113	0	65
21	503	0	352	67	353	0	203	113	481	0	330
22	459	1	375	68	703	0	620	114	137	0	81
23	374	0	264	69	537	0	387	115	306	0	285
24	218	0	162	70	303	0	247	116	479	0	328
25	258	0	202	71	467	0	318	117	592	2	587
26	70	0	14	72	428	0	372	118	317	1	170
27	434	0	283	73	117	0	61	119	436	0	380
28	333	0	183	74	483	0	355	120	427	0	277
29	277	0	221	75	730	1	605	121	395	0	246
30	181	0	124	76	254	0	104	122	318	0	168
31	100	0	44	77	421	0	270	123	513	0	363
32	474	0	323	78	385	0	329	124	396	0	247
33	456	1	326	79	287	0	138	125	458	0	308
34	274	1	189	80	395	1	281	126	220	0	71
35	471	0	320	81	540	1	488	127	253	0	104
36	171	0	20	82	649	0	500	128	568	0	418
37	501	1	375	83	191	0	42	129	557	0	408
38	222	0	165	84	390	0	241	130	280	0	131
39	390	2	352	85	525	0	376	131	298	0	148
40	188	0	132	86	457	0	307	132	372	2	340
41	477	0	421	87	464	0	326	133	399	0	249
42	234	0	83	88	427	0	277	134	921	1	809
43	381	0	325	89	329	0	273	135	379	0	323
44	258	0	108	90	567	0	417	136	303	0	247
45	239	0	89	91	516	0	366	137	386	0	236

46	355	1	232	92	443	0	293	138	388	0	332

Продолжение таблицы Б.1

i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i
139	413	0	357	185	216	1	137	231	357	0	253
140	96	0	40	186	414	0	264	232	162	0	106
141	197	0	141	187	371	0	222	233	168	0	18
142	298	0	147	188	197	0	47	234	237	0	180
143	402	0	346	189	156	0	0	235	75	0	19
144	280	0	142	190	320	1	256	236	295	0	144
145	430	0	374	191	303	0	154	237	477	0	421
146	396	0	340	192	481	0	331	238	206	0	56
147	302	0	246	193	336	0	186	239	313	0	257
148	298	0	242	194	209	1	96	240	349	0	269
149	404	0	348	195	298	0	149	241	199	0	157
150	433	0	377	196	266	0	116	242	351	0	310
151	263	0	244	197	388	0	332	243	98	0	56
152	165	0	109	198	511	1	360	244	519	0	478
153	458	0	307	199	359	1	303	245	225	0	201
154	126	0	69	200	176	0	120	246	285	0	244
155	229	0	79	201	424	0	302	247	296	0	255
156	310	2	291	202	356	0	206	248	335	0	294
157	389	1	239	203	422	0	273	249	369	0	328
158	338	0	187	204	486	1	396	250	442	0	400
159	120	0	64	205	435	1	375	251	338	0	297
160	210	0	153	206	156	0	100	252	297	0	256
161	227	0	77	207	488	0	338	253	411	0	370
162	330	0	273	208	200	0	85	254	333	0	292
163	322	0	266	209	387	0	331	255	159	0	118
164	236	0	180	210	660	0	604	256	131	0	90
165	393	0	242	211	364	0	215	257	531	0	490
166	456	0	400	212	464	0	314	258	188	0	147
167	371	0	314	213	387	0	237	259	258	0	217
168	331	0	275	214	287	0	137	260	65	0	24
169	713	1	615	215	233	0	177	261	104	0	63
170	446	1	365	216	289	0	233	262	374	0	333
171	789	1	666	217	170	0	113	263	127	0	86
172	507	1	370	218	312	0	172	264	252	0	227
173	305	0	156	219	138	0	81	265	46	0	0
174	314	0	165	220	144	0	87	266	241	0	200
175	276	0	126	221	495	0	345	267	198	0	157
176	198	1	122	222	90	0	34	268	393	0	352
177	395	0	245	223	127	0	0	269	391	0	359
178	400	1	344	224	487	0	339	270	241	0	200
179	398	2	369	225	459	0	309	271	329	0	288
180	233	0	83	226	201	0	144	272	648	0	607
181	235	0	85	227	97	0	41	273	335	0	294
182	131	0	75	228	208	0	58	274	291	0	249
183	110	0	53	229	146	0	90	275	264	0	223
184	800	1	676	230	354	0	203	276	127	0	86

Продолжение таблицы Б.1

i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i
277	101	0	60	323	389	0	348	369	438	0	396
278	319	0	278	324	337	0	296	370	399	0	358
279	451	0	410	325	427	0	386	371	202	0	161
280	246	0	205	326	253	0	212	372	226	0	185
281	363	0	322	327	305	0	264	373	370	0	329
282	228	0	217	328	311	0	290	374	276	0	235
283	304	0	263	329	451	0	410	375	328	0	295
284	51	1	36	330	128	0	87	376	100	0	59
285	382	0	341	331	278	0	236	377	422	0	381
286	158	0	116	332	339	0	298	378	311	0	292
287	315	0	273	333	245	0	204	379	135	0	93
288	369	0	328	334	277	0	249	380	543	0	502
289	329	0	288	335	94	0	53	381	107	0	65
290	342	0	301	336	614	0	573	382	556	0	515
291	149	0	108	337	285	0	244	383	589	0	548
292	394	0	352	338	282	0	240	384	435	0	394
293	345	0	304	339	225	0	184	385	139	0	98
294	329	0	288	340	271	0	261	386	467	0	433
295	91	0	50	341	360	0	331	387	164	0	123
296	100	0	58	342	75	0	34	388	306	0	265
297	367	0	326	343	110	0	69	389	104	0	63
298	315	0	274	344	295	0	282	390	181	0	140
299	332	0	291	345	309	0	268	391	226	0	184
300	229	0	188	346	447	0	406	392	367	0	326
301	292	0	260	347	62	0	21	393	237	0	196
302	386	0	345	348	279	0	237	394	448	0	407
303	64	0	22	349	96	0	65	395	413	0	372
304	102	0	61	350	376	0	335	396	94	0	53
305	90	0	48	351	207	0	198	397	439	0	398
306	260	0	219	352	303	0	290	398	106	0	65
307	100	0	59	353	186	0	144	399	309	0	268
308	56	0	15	354	385	0	344	400	403	0	362
309	337	0	295	355	48	0	0	401	187	0	146
310	193	0	176	356	66	0	25	402	556	0	515
311	343	0	317	357	413	0	372	403	330	0	320
312	484	0	443	358	522	0	484	404	346	0	336
313	104	0	62	359	315	0	274	405	118	0	77
314	133	0	91	360	375	0	334	406	342	0	301
315	424	0	383	361	335	0	294	407	99	0	58
316	301	0	283	362	482	0	441	408	339	0	298
317	607	0	566	363	227	0	186	409	129	0	110
318	358	0	316	364	263	0	222	410	270	0	229
319	519	0	478	365	273	0	261	411	99	0	58
320	465	0	438	366	231	0	189	412	361	0	341
321	362	0	321	367	455	0	414	413	568	0	527
322	124	0	83	368	369	0	328	414	424	0	383

Продолжение таблицы Б.1

i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i
415	48	0	0	461	255	0	214	507	175	0	134
416	92	0	78	462	49	0	0	508	88	0	47
417	354	0	313	463	124	0	83	509	204	0	169
418	35	0	0	464	440	0	399	510	64	0	23
419	372	0	331	465	518	0	477	511	146	0	132
420	57	0	15	466	204	0	163	512	513	0	472
421	58	0	17	467	325	0	284	513	175	0	134
422	332	0	291	468	231	0	190	514	333	0	291
423	41	0	0	469	317	0	276	515	73	0	32
424	142	0	131	470	57	0	16	516	339	0	298
425	272	0	231	471	51	0	0	517	254	0	213
426	316	0	286	472	530	0	489	518	534	0	493
427	301	0	259	473	260	0	219	519	390	0	348
428	367	0	326	474	423	0	382	520	380	0	370
429	249	0	232	475	353	0	312	521	94	0	53
430	498	0	457	476	243	0	201	522	63	0	22
431	53	0	0	477	635	0	594	523	245	0	204
432	297	0	282	478	485	0	444	524	209	0	167
433	110	0	68	479	308	0	267	525	377	0	336
434	207	0	166	480	361	0	333	526	338	0	297
435	201	0	191	481	383	0	342	527	341	0	300
436	328	0	287	482	283	0	242	528	282	0	241
437	103	0	62	483	46	0	0	529	344	0	303
438	53	0	0	484	325	0	284	530	180	0	139
439	440	0	399	485	297	0	256	531	496	0	455
440	56	0	15	486	342	0	301	532	204	0	163
441	487	0	457	487	250	0	237	533	269	1	243
442	422	0	381	488	316	0	275	534	521	0	480
443	48	0	0	489	381	0	340	535	322	0	280
444	52	0	0	490	403	0	362	536	306	0	265
445	262	0	221	491	198	0	189	537	369	0	348
446	321	0	280	492	147	0	106	538	251	0	232
447	55	0	14	493	427	0	398	539	492	0	451
448	296	1	273	494	100	0	59	540	334	0	293
449	282	0	241	495	505	0	464	541	47	0	0
450	287	0	246	496	69	0	28	542	397	0	382
451	256	0	214	497	363	0	322	543	287	0	246
452	176	0	135	498	207	0	165	544	287	0	270
453	340	0	299	499	506	0	465	545	500	0	459
454	417	0	376	500	191	0	150	546	367	0	326
455	227	0	186	501	269	0	228	547	507	0	467
456	70	0	54	502	349	0	308	548	160	0	118
457	357	0	316	503	176	0	135	549	298	0	257
458	503	0	462	504	239	0	226	550	246	0	232
459	164	0	123	505	551	0	510	551	341	0	300
460	239	0	222	506	341	0	300	552	348	0	307

Продолжение таблицы Б.1

i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i	i	X_i	δ_i	τ_i
553	432	0	391	599	246	0	205	645	102	0	62
554	40	0	18	600	443	0	402	646	281	0	241
555	383	0	342	601	274	0	233	647	300	0	260
556	340	0	299	602	111	0	69	648	71	0	31
557	223	0	181	603	456	0	415	649	429	0	389
558	311	0	270	604	214	0	173	650	473	0	433
559	125	0	84	605	270	0	229	651	220	0	180
560	353	0	312	606	234	0	210	652	524	0	484
561	254	0	213	607	265	0	224	653	484	0	444
562	575	0	534	608	252	0	211	654	333	0	293
563	283	0	242	609	174	0	132	655	227	0	186
564	53	0	0	610	193	0	151	656	94	0	54
565	334	0	293	611	274	0	233	657	276	0	236
566	103	0	62	612	404	0	363	658	414	0	395
567	59	0	18	613	403	0	362	659	423	0	383
568	376	0	335	614	516	0	475	660	359	0	319
569	391	0	350	615	352	0	311	661	275	0	265
570	126	0	85	616	205	0	164	662	82	0	41
571	176	0	135	617	365	0	324	663	297	0	257
572	118	0	93	618	181	0	140	664	278	0	237
573	410	0	369	619	403	0	362	665	402	0	362
574	237	0	196	620	513	0	472	666	513	0	473
575	342	0	301	621	101	0	59	667	372	0	333
576	343	0	302	622	229	0	188	668	363	0	323
577	115	0	91	623	66	0	25	669	317	0	277
578	406	0	364	624	285	0	244	670	354	0	326
579	375	0	334	625	317	0	276	671	313	0	273
580	343	0	322	626	344	0	303	672	121	0	81
581	168	0	126	627	59	0	17	673	175	0	135
582	287	0	246	628	396	0	355	674	363	0	323
583	271	0	230	629	284	0	243	675	173	0	133
584	197	0	156	630	157	0	116	676	113	0	72
585	175	0	134	631	187	0	146	677	149	0	132
586	114	0	73	632	168	0	127	678	342	0	302
587	76	0	35	633	275	1	267	679	293	0	253
588	414	0	373	634	155	0	114	680	154	0	114
589	111	0	70	635	496	0	455	681	393	0	353
590	199	0	180	636	497	0	456	682	170	0	130
591	487	0	446	637	86	0	45	683	366	0	326
592	202	0	161	638	60	0	27	684	337	0	297
593	375	0	334	639	206	0	166	685	312	0	271
594	451	0	410	640	477	0	437	686	276	0	236
595	165	0	124	641	467	0	427	687	336	0	296
596	104	0	62	642	151	0	141	688	358	0	318
597	438	0	397	643	519	0	478	689	280	0	240
598	373	0	331	644	365	0	354	690	104	0	64

Продолжение таблицы Б.1

i	X_i	δ_i	τ_i
691	131	0	101
692	85	0	44
693	292	1	273
694	304	0	264
695	424	0	384
696	257	0	216
697	525	0	485
698	311	0	271
699	654	0	614
700	389	0	349
701	381	0	341
702	94	0	53
703	649	0	609
704	463	0	423
705	511	0	470
706	516	0	475
707	77	0	36
708	261	0	241
709	267	0	226
710	315	0	274
711	398	1	362

Приложение В. Текст программы

```
/// <summary>
/// Моделирование усеченной слева и цензурированной справа (первый тип) выборки наблюдений.
/// </summary>
/// <param name="rv"> Распределение, которому подчиняются наблюдения выборки.</param>
/// <param name="size"> Объем выборки.</param>
/// <param name="censorPoint"> Значение момента времени конца эксперимента (Точка правого
цензурирования).</param>
/// <param name="installTimes"> Значения моментов времени начала эксплуатации объектов.
</param>
/// <param name="truncPoint"> Значение момента времени начала эксперимента (Точка усечения).
</param>
public static Sample<Observation> CreateSample(RVariate rv, int size, double censorPoint,
double[] installTimes, double truncPoint)
{
    // Итоговая выборка
    var resSample = new Sample<Observation>();
    double _value, _truncTime;
    ObservationType _observationType;
    int n = 0;

    while (n < size)
    {
        // Продлжительность жизни
        double t = rv.FInv(Generator.Rnd());

        // Если конец эксплуатации после начала эксперимента - включаем в выборку,
        // иначе не влючаем
        double failtureTime = t + installTimes[n];
        if (failtureTime >= truncPoint)
        {
            // Если начало эксперимента после начала экспуатации
            if (installTimes[n] < truncPoint)
            {
                // Наблюдение усеченное
                _truncTime = truncPoint - installTimes[n];
            }
            else
            {
                // Наблюдение неусеченное
                _truncTime = 0;
            }

            // Если отказ до момента цензурирования, то наблюдение полное
            if (failtureTime <= censorPoint)
            {
                // Наблюдение нецензурированное (полное)
                _observationType = ObservationType.Complete;
                _value = t;
            }
            else
            {
                // Наблюдение цензурированное
                _observationType = ObservationType.RightCensored;
                _value = censorPoint - installTimes[n];
            }
            resSample.AddObservation(new Observation(_value, _observationType, _truncTime));
            n++;
        }
    }
    return resSample;
}
```

```
/// <summary>
/// Моделирование усеченной слева и цензурированной справа (третий тип) выборки наблюдений.
/// </summary>
/// <param name="rv"> Распределение, которому подчиняются наблюдения выборки.</param>
/// <param name="rvc"> Распределение, которому подчиняются моменты цензурирования.</param>
/// <param name="size"> Объем выборки.</param>
/// <param name="truncPoint"> Значение моментов времени начала эксперимента (Точка
усечения). </param>
public static Sample<Observation> CreateSample(RVariate rv, RVariate rvc, int size, double[]
truncPoint)
{
    int n = 0;
    // Итоговая выборка
    var resSample = new Sample<Observation>();

       ObservationType observationType;
       while (n < size)
    {
        // Продолжительность жизни
        double t = rv.FInv(Generator.Rnd());
        // Точка цензурирования
        double c = rvc.FInv(Generator.Rnd());

        if (truncPoint[n] < t && truncPoint[n] < c)
        {
            double value;
            double truncTime = truncPoint[n];
            if (c >= t)
            {
                value = t;
                observationType = ObservationType.Complete;
            }
            else
            {
                value = c;
                observationType = ObservationType.RightCensored;
            }
            resSample.AddObservation(new Observation(value, observationType, truncTime));
            n++;
        }

    }
       return resSample;
}


/// ОМП.
public void EstimationOfEstimationVersionWithDepth()
{
    List<Graph> graphs = new List<Graph>();

    // Распределение, по которому моделируем
    var rv = new RVWeibull(0, 2, 2);

    // Истинные значения параметров формы и масштаба для распределения
    double trueForm = 2.0;
    double trueScale = 2.0;

    int size = 200;

    // Точка(год) усечения
    double dTR = 0.3;
    double truncYear = rv.FInv(dTR);
    int nTr = size*(100-10)/100; // степень усечения задается как количество в выборке
    // Год цензурирвования. Выбирается из массива censorYearArray
```

```
    double censorYear;

    // Года цензурирования
    double[] censorYearArray =
    {
        //0.0, 0.15, 0.3, 0.45
        0.0//, 0.1, 0.3, 0.5, 0.7
    };

    // Года установки машин
    double[] installationYear = new double[size];
    int N = 10000; // количество испытаний

    for (int i = 0; i < nTr; i++)
        installationYear[i] = truncYear;
    for (int i = nTr; i < (size - nTr); i++)
        installationYear[i] = 0.0;

    var nameFile = "Out_" + rv.Name + "_" + size.ToString() + "_" + ".txt";
    TextWriter writer = new StreamWriter(nameFile);
    writer.WriteLine("d     nCens    сдвиг    дисперсия");
    for (int ii = 0; ii < censorYearArray.Length; ii++)
    {
        var ff = ((double)((ii + 1.0)) * 100 / (double)(censorYearArray.Length)).ToString()
+ "%";
        Console.WriteLine(ff);
        // Берем текущий год цензурирования
        if (censorYearArray[ii] == 0.0)
        {
            censorYear = 3000;
        }
        else
        {
            censorYear = rv.FInv((1-censorYearArray[ii])*(1-dTR)+dTR);
        }
        if (censorYear < truncYear)
            Console.WriteLine("WOW! IDIOT!");
        else
        {
            // Названия файлов
            string str;
            // Выборки с оценками параметров масштаба и формы
            var SampleForm = new Sample<Observation>();
            var SampleScale = new Sample<Observation>();

            // Количество усеченных наблюдений
            double nCensMean = 0;
            // Среднее количество цензурированных
            double nTruncMean = 0;
            double nnCens = 0, nnTrunc = 0;


            for (int i = 0; i < N; i++)
            {
                var sample = SampleOperations.CreateSample(rv, size, censorYear,
installationYear, truncYear);
                int nCens = 0;
                int nTrunc = 0;

                for (int j = 0; j < size; j++)
                {
                    if (sample[j].Type == ObservationType.RightCensored)
                    {
                        nCens++;
```

```
                    nnCens++;
                }
                if (sample[j].TruncValue > 0.0)
                {
                    nTrunc++;
                    nnTrunc++;
                }
            }

            nCensMean += nCens;
            nTruncMean += nTrunc;

            // По этому распределению будет проихводиться оценка параметров
            var rv1 = new RVWeibull(0, 2, 2);
            // оценивать параметр сдвига
            rv1.Parameters[ParameterType.Shift].IsEstimated = false;
            // оценивать параметр масштаба
            rv1.Parameters[ParameterType.Scale].IsEstimated = true;
            // оценивать параметр формы
            rv1.Parameters[ParameterType.Form].IsEstimated = true;

            // Оценка максимального правдоподобия
            var function = new TSMaximumLikelihood(sample, rv1);
            // функционал для оценивания - логарифм функции правдоподобия

            Estimation estimation = new Estimation(function, MultivariateMethod.Gauss)
            {
                OptimumType = OptimumType.Maximum
            };

            estimation.Estimate(rv1, sample);

            // результат оценивания - измененные параметры распределения
            SampleForm.AddObservation(new
Observation(rv1.Parameters[ParameterType.Form].Value));
            SampleScale.AddObservation(new
Observation(rv1.Parameters[ParameterType.Scale].Value));
        }

        str = Math.Round(truncYear, 3).ToString() + "_" + Math.Round(censorYear,
3).ToString() + "_" + size.ToString() + "_SampleForm_isw.txt";
        SampleForm.SaveAsIsw(str);
        str = Math.Round(truncYear, 3).ToString() + "_" + Math.Round(censorYear,
3).ToString() + "_" + size.ToString() + "_SambleScale_isw.txt";
        SampleScale.SaveAsIsw(str);
        string str1 = truncYear.ToString() + "_" + censorYear.ToString();
        writer.WriteLine("{0}\t{1}\t{2}\t{3}\t{4}\t{5}\t{6}\t{7}\t", truncYear,
censorYearArray[ii], nnTrunc / (N * size) * 100, nnCens / (N * size) * 100,
SampleScale.Mean() - trueScale, SampleForm.Mean() - trueForm,
SampleScale.Variance(SampleScale.Mean()), SampleForm.Variance(SampleForm.Mean()));
        Console.WriteLine("{0}\t{1}\t{2}\t{3}\t{4}\t{5}\t{6}\t{7}\t", truncYear,
censorYearArray[ii], nnTrunc / (N * size) * 100, nnCens / (N * size) * 100,
SampleScale.Mean() - trueScale, SampleForm.Mean() - trueForm,
SampleScale.Variance(SampleScale.Mean()), SampleForm.Variance(SampleForm.Mean()));
        }
    }
        writer.Close();
}
```

```
namespace ConsoleApplication
{

    internal class TestsTruncated : SimulationObject
    {
        /// <summary>
        /// Конструтор.
        /// </summary>
        public TestsTruncated()
        {
            Console.Write("n=");
            n = ReadInt();
            Console.Write("V=");
            V = ReadInt();
            N = 16600;
            Dimension = 3; // размерность функции для моделирования
            FileNames = new string[Dimension]; // имена файлов для сохранения результатов
            var truncDegree = new[] { 0.5 };
            //FileNames[0] = string.Format("Kolm_n{0}_V{1}_{2}%.dat", n, V, 0);
            //FileNames[1] = string.Format("CMS_n{0}_V{1}_{2}%.dat", n, V, 0);
            //FileNames[2] = string.Format("AD_n{0}_V{1}_{2}%.dat", n, V, 0);
            for (int i = 0; i < 1; i++)
            {
                FileNames[i * 3] = string.Format("Tr_Simple_Kolm_n{0}_V{1}_{2}%.dat", n, V,
truncDegree[i] * 100);
                FileNames[i * 3 + 1] = string.Format("Tr_Simple_CMS_n{0}_V{1}_{2}%.dat", n,
V, truncDegree[i] * 100);
                FileNames[i * 3 + 2] = string.Format("Tr_Simple_AD_n{0}_V{1}_{2}%.dat", n,
V, truncDegree[i] * 100);
            }
        }

        /// <summary>
        /// Функция для многократного вызова в simulation.
        /// </summary>
        /// <returns>
        /// Статистики непар критериев
        /// </returns>
        public override double[] Function()
        {
            double[] res = new double[Dimension];
            //double censDegree = 0.3;

            var rvTrue = new RVWeibull(0, 2, 2); //Распределение, из которого моделируем
выборку Поменять на гамму для G(S|H0)
            //var rvTrue = new RVGamma(0, 0.5577, 3.1215); //Распределение, из которого
моделируем выборку Поменять на гамму для G(S|H0)
            var rv = new RVWeibull(0, 2, 2); //Нулевая гипотеза

            var truncDegree = new[] { 0.5 };

            for (int i = 0; i < 1; i++)
            {
                var installTimes = new double[n];
                double trDeep = rvTrue.FInv((double)V / 100);
                for (int j = 0; j < (int)(n * truncDegree[i]); j++) installTimes[j] = 0.0;
                for (int j = (int)(n * truncDegree[i]); j < n; j++) installTimes[j] =
trDeep;
                var sample = SampleOperations.CreateSample(rvTrue, n, 10000, installTimes,
trDeep);

                var estimation = new Estimation(new TSMaximumLikelihood(sample, rv))
                {
                    OptimumType = OptimumType.Maximum,
```

```
                    Eps = 1e-6,
                    MaxIterationNumber = 100000,
                };

                rv.Parameters[ParameterType.Shift].IsEstimated = false;
                rv.Parameters[ParameterType.Scale].IsEstimated = true;
                rv.Parameters[ParameterType.Form].IsEstimated = true;

                estimation.Estimate(rv, sample);
                var ravSample = sample.TruncatedToUniform(rv);
                var rav = new RVUniform(0, 1);

                var kolm = new TSKolmogorov(ravSample, rav);
                var CMS = new TSOmegaSmall(ravSample, rav);
                var AD = new TSOmegaBig(ravSample, rav);


                res[i * 3] = kolm.Function(rav.ParametersToVector());
                res[i*3 + 1] = CMS.Function(rav.ParametersToVector());
                res[i * 3 + 2] = AD.Function(rav.ParametersToVector());

            }
            return res;
        }

        /// <summary>
        /// Функция для многократного вызова в simulation.
        /// </summary>
        /// <returns>
        /// Статистики непар критериев
        /// </returns>
        public override double[] Function()
        {
            double[] res = new double[Dimension];
            //double censDegree = 0.3;

            var rvTrue = new RVWeibull(0, 2, 2); //Распределение, из которого моделируем
выборку Поменять на гамму для G(S|H0)
            //var rvTrue = new RVGamma(0, 0.5577, 3.1215); //Распределение, из которого
моделируем выборку Поменять на гамму для G(S|H0)
            var rv = new RVWeibull(0, 2, 2); //Нулевая гипотеза
            var censDegree = new[] { 0, 10, 30, 50, 70 };
            var censPoints = new double[censDegree.Length];
            for (int i = 0; i < censDegree.Length; i++)
            {
                censPoints[i] = rv.FInv((100.0 - censDegree[i])/100.0);
            }
            for (int i = 0; i < censDegree.Length; i++)
            {
                var installTimes = new double[n];
                for (int j = 0; j < (int)(n); j++)
                    installTimes[j] = 0.0;
                var sample = SampleOperations.CreateSample(rvTrue, n, censPoints[i],
installTimes, 0);
                var estimation = new Estimation(new TSMaximumLikelihood(sample, rv))
                {
                    OptimumType = OptimumType.Maximum,
                    Eps = 1e-6,
                    MaxIterationNumber = 100000,
                };

                rv.Parameters[ParameterType.Shift].IsEstimated = false;
                rv.Parameters[ParameterType.Scale].IsEstimated = true;
                rv.Parameters[ParameterType.Form].IsEstimated = true;
```

```
                estimation.Estimate(rv, sample);
                var kolm = new TSKolmogorov(sample, rv);
                var CMS = new TSOmegaSmall(sample, rv);
                var AD = new TSOmegaBig(sample, rv);

                res[i * 3] = kolm.Function(rv.ParametersToVector());
                res[i*3 + 1] = CMS.Function(rv.ParametersToVector());
                res[i*3 + 2] = AD.Function(rv.ParametersToVector());

            }
            return res;
        }
    }
}
```