Математическое и алгоритмическое обеспечение статистического анализа данных типа времени жизни

Чимитова Екатерина Владимировна

05.13.17 – Теоретические основы информатики
Диссертация на соискание ученой степени доктора технических наук

Научный консультант: Лемешко Борис Юрьевич, д.т.н., профессор

Новосибирск – 2016

Цель диссертации: расширение возможностей математического аппарата и развитие компьютерных технологий исследования статистических закономерностей для решения задач анализа данных типа времени жизни.

Задачи:

1. Исследование статистических свойств оценок максимального правдоподобия (ОМП) параметров наблюдаемых законов по цензурированным справа выборкам, выборкам усеченных слева наблюдений и выборкам текущих состояний.

2. Исследование распределений статистик и мощности критериев согласия типа хи-квадрат по полным, группированным и цензурированным выборкам.

3. Исследование распределений статистик и мощности модифицированных критериев типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для цензурированных данных.

4. Разработка статистических критериев согласия для проверки простых и сложных гипотез по выборкам усеченных слева наблюдений и выборкам текущих состояний.

5. Исследование распределений статистик и мощности критериев согласия, применяемых к выборкам остатков, для проверки гипотез о виде параметрических регрессионных моделей.

6. Разработка методики проверки гипотезы о виде деградационной гамма-модели с использованием непараметрических критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

7. Разработка программного обеспечения для построения вероятностных моделей типа времени жизни по выборкам, содержащим полные, цензурированные, усеченные слева или интервальные наблюдения.

Положения, выносимые на защиту

На защиту выносятся:

1. Результаты исследования статистических свойств ОМП параметров законов распределения по цензурированным справа выборкам и выборкам усеченных слева наблюдений, свидетельствующие о существенном отличии свойств оценок в реальных ситуациях от асимптотических.
2. Метод оптимального разбиения цензурированной выборки на интервалы, позволяющий существенно повысить мощность применяемых критериев согласия типа хи-квадрат относительно заданной пары конкурирующих гипотез.
3. Результаты сравнительного анализа мощности критериев согласия для цензурированных справа выборок при проверке сложных гипотез о виде распределения. Рекомендации по выбору критериев.
4. Критерии согласия, предложенные для анализа выборок текущих состояний. Применение критериев базируется на статистическом моделировании требуемых распределений статистик, осуществляемом в интерактивном режиме проводимого анализа.

Положения, выносимые на защиту

На защиту выносятся:

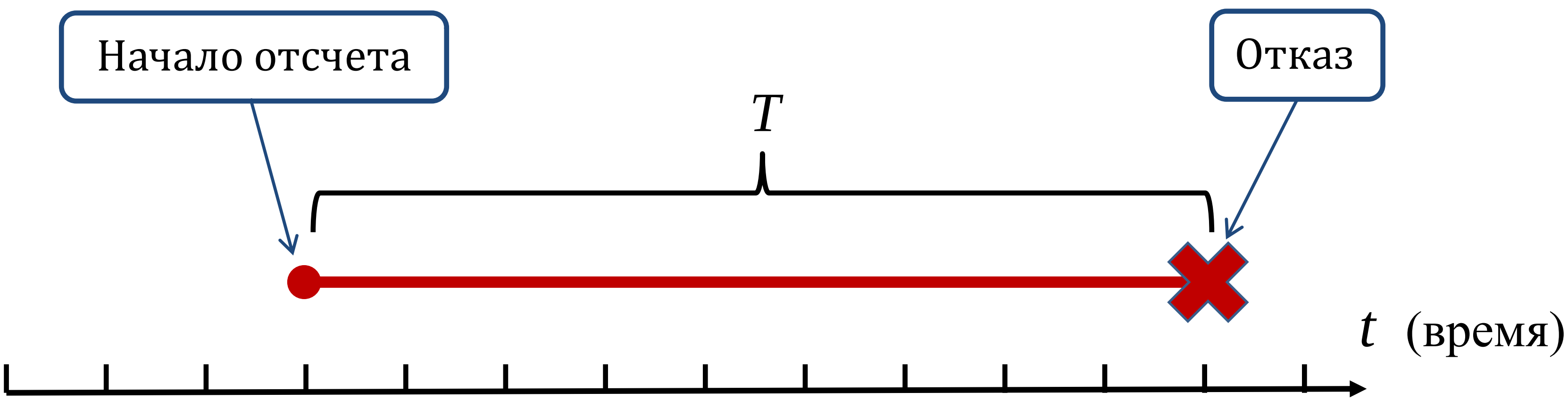
5. Методика проверки гипотез о виде параметрических регрессионных моделей по выборкам отказов, содержащим полные, цензурированные, усеченные слева или интервальные наблюдения, на основе критериев согласия, применяемых к выборкам остатков.
6. Алгоритмы моделирования распределений статистик критериев согласия при справедливости проверяемой гипотезы, существенно расширяющие сферу применения аппарата математической статистики для проверки гипотез о виде распределения по выборкам, содержащим полные, цензурированные, усеченные слева или интервальные наблюдения.
7. Методика проверки гипотезы о виде деградационной гамма-модели надежности с использованием критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

Данные типа времени жизни

В задачах анализа данных типа времени жизни особый интерес представляют группа или группы объектов (индивидуумов), для каждого из которых определено точечное событие, часто называемое **отказом**.

Для точного определения наработки до отказа необходимо выполнение следующих условий:

1. Четко установлено начало отсчета времени.
2. Выбран масштаб для измерения отсчета времени.
3. Точно определено само понятие отказа.



T – положительная случайная величина, характеризующая продолжительность жизни объекта (индивидуума)

Зависимость от объясняющих переменных

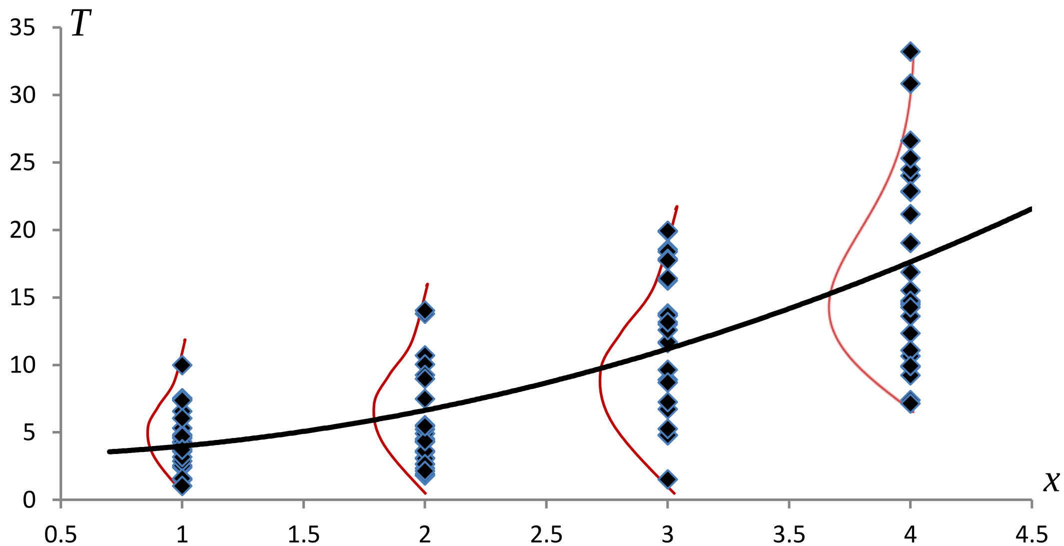
T – положительная случайная величина, характеризующая продолжительность жизни объекта (индивидуума)

$x = (x_1, x_2, \ldots, x_m)^T$ – вектор объясняющих переменных (ковариат)

План эксперимента: $\xi_n^q = \begin{Bmatrix} x^1 & \ldots & x^q \\ n_1 & \ldots & n_q \end{Bmatrix}.$

$$\boxed{T = H(x; \varepsilon), \qquad (1)}$$

где $H(\cdot)$ – неотрицательная функция, $\varepsilon \sim F_0(t),\ \ t > 0$
$F_0(t)$ – **базовая** функция распределения.



? Вероятность безотказной работы

? Средняя наработка до отказа

? Интенсивность отказов

...

Зависимость от объясняющих переменных

$$S_x(t) = g\left(S_0(t); x\right), \qquad (2)$$

где $S_x(t) = P\{T \geq t \mid x\} = 1 - F_x(t)$ - функция надежности, $g(\cdot):[0,1] \to [0,1]$

$S_0(t)$ - базовая функция надежности.

$$\boxed{S_x(t;\beta,\theta) = g\left(S_0(t;\theta); x; \beta\right)}$$

<u>Модель ускоренных испытаний (AFT-модель):</u>

$$S_x(t;\beta,\theta) = S_0\left(\frac{t}{r(x;\beta)};\theta\right)$$

<u>Модель пропорциональных интенсивностей:</u>

$$S_x(t;\beta,\theta) = \left(S_0(t;\theta)\right)^{r(x;\beta)}$$

$$r(x;\beta) = \exp\left(\beta_1\varphi_1(x) + \ldots + \beta_m\varphi_m(x)\right), \qquad (3)$$

где $\varphi_i(x)$ - элементарная функция от элементов вектора ковариат, $i = \overline{1,m}$

$S_0(t;\theta)$ соответствует некоторому параметрическому семейству распределений (экспоненциальное, гамма, Вейбулла и др.)

Проверка гипотезы о виде параметрической регрессионной модели

$$H_0: \ S_x(t) \in \left\{ g\left(S_0(t;\theta); x; \beta\right), \ \theta, \beta \in \Omega \right\}$$

Остатки для параметрической регрессионной модели по полной выборке с ковариатами $\left\{\left(T_1, x^{(1)}\right), \left(T_2, x^{(2)}\right), \ldots, \left(T_n, x^{(n)}\right)\right\}$ имеют вид:

$$e_i = S_0^{-1}\left(S_{x^{(i)}}\left(T_i; \hat{\beta}, \hat{\theta}\right)\right),$$

где $S_0^{-1}(\cdot)$ – функция, обратная к базовой функции надежности,

$\hat{\beta}$ и $\hat{\theta}$ – оценки максимального правдоподобия.

$$\boxed{H_0: \ F_e(t) \in \left\{ F_0\left(t; \theta\right), \theta \in \Omega \right\}}$$

Зависимость распределений статистик критериев согласия от свойств оценок параметров проверяемой модели.

Выбор интервалов группирования для критериев согласия типа хи-квадрат.

Необходимость разработки и исследования критериев согласия для выборок, содержащих цензурированные, усеченные или интервальные наблюдения.

Выборка

- Полная
- Цензурированная справа
- Усеченных слева наблюдений
- Процессов деградации
- Текущих состояний
- Группированная
- Интервальная

Цензурированная справа выборка

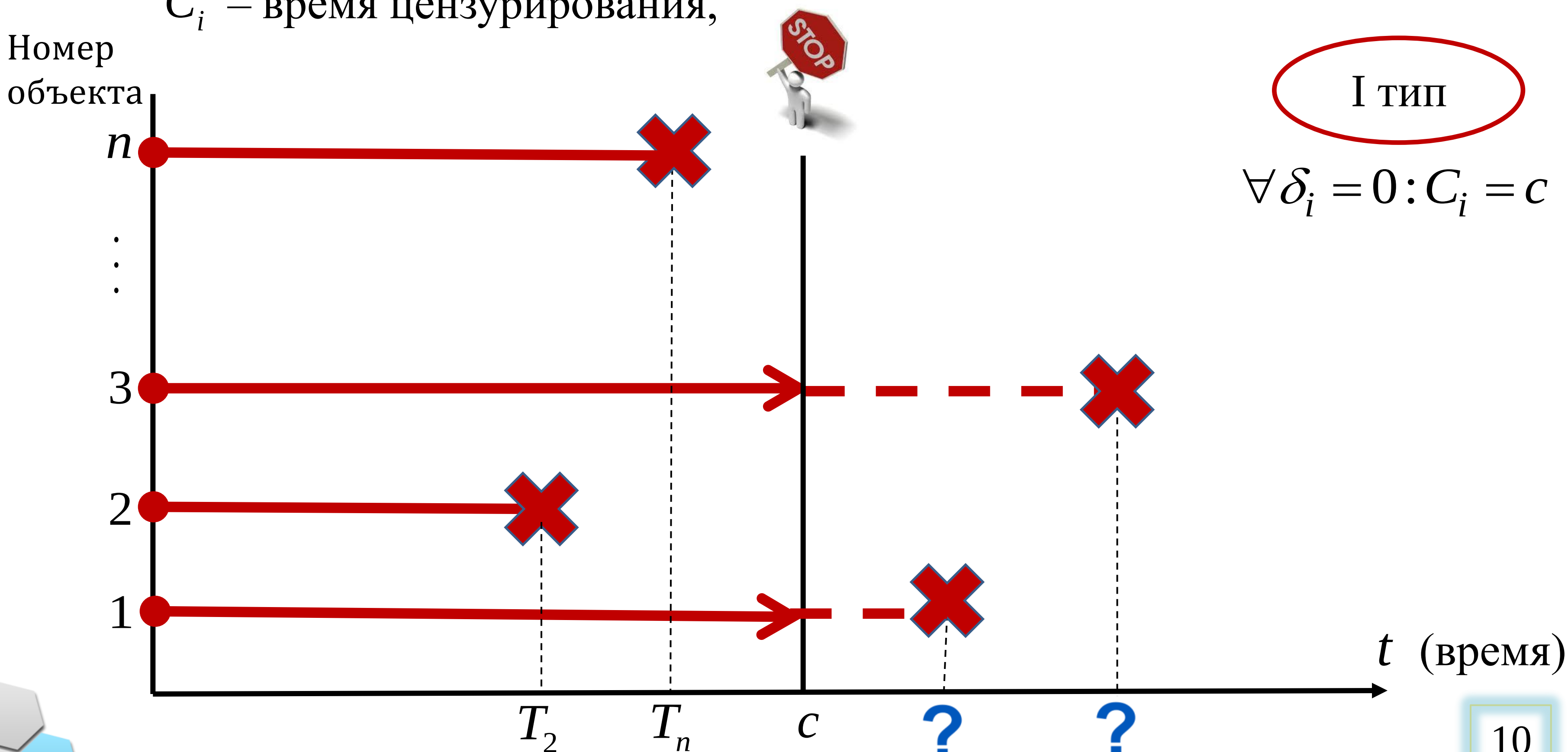
$$\mathbf{X}_n = \left\{ \left(X_1, \delta_1 \right), \left(X_2, \delta_2 \right), \ldots, \left(X_n, \delta_n \right) \right\},$$

$X_i = \min\left(T_i, C_i \right),$

T_i – время отказа,

C_i – время цензурирования,

$$\delta_i = \begin{cases} 1, \text{ если } X_i = T_i, \\ 0, \text{ если } X_i = C_i \end{cases} \qquad i = \overline{1, n}$$



I тип

$\forall \delta_i = 0 : C_i = c$

Цензурированная справа выборка

$$\mathbf{X}_n = \{(X_1,\delta_1),(X_2,\delta_2),\dots,(X_n,\delta_n)\},$$

$X_i = \min(T_i, C_i)$,

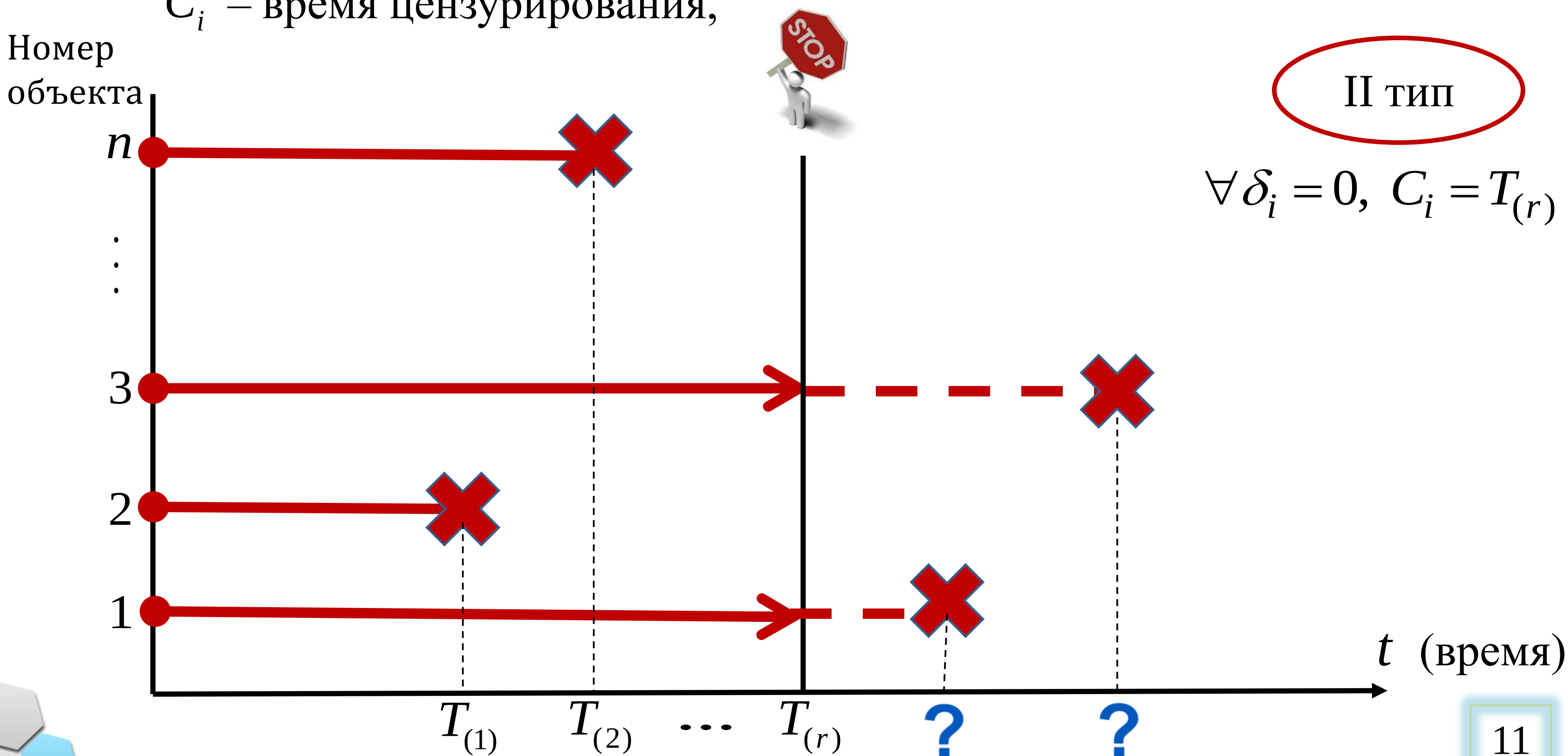
T_i – время отказа,

C_i – время цензурирования,

$$\delta_i = \begin{cases} 1, \text{ если } X_i = T_i, \\ 0, \text{ если } X_i = C_i \end{cases} \qquad i = \overline{1,n}$$

II тип

$\forall \delta_i = 0,\ C_i = T_{(r)}$



Цензурированная справа выборка

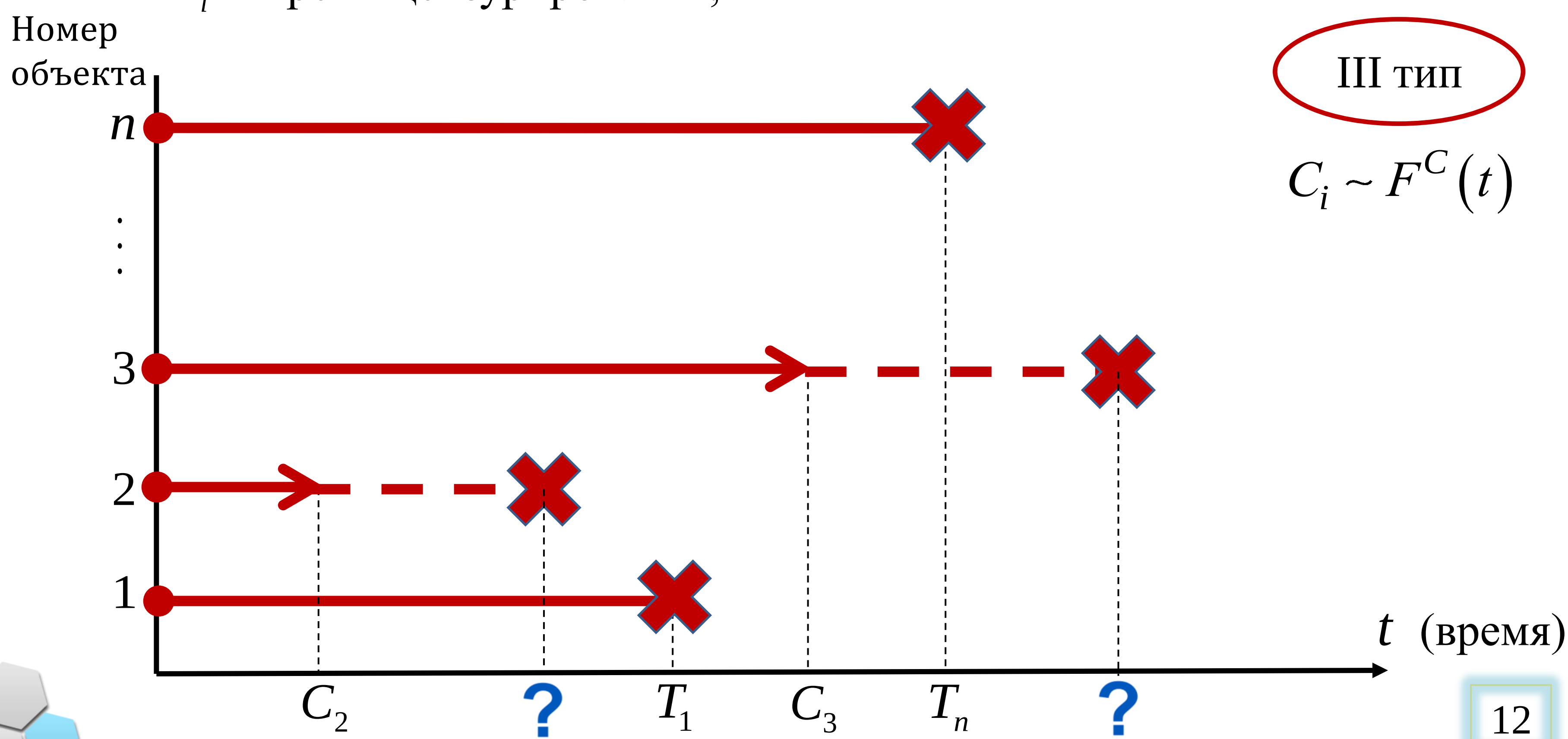
$$\mathbf{X}_n = \{(X_1, \delta_1), (X_2, \delta_2), \ldots, (X_n, \delta_n)\},$$

$X_i = \min(T_i, C_i)$,

T_i – время отказа,

C_i – время цензурирования,

$$\delta_i = \begin{cases} 1, \text{ если } X_i = T_i, \\ 0, \text{ если } X_i = C_i \end{cases} \qquad i = \overline{1, n}$$



III тип

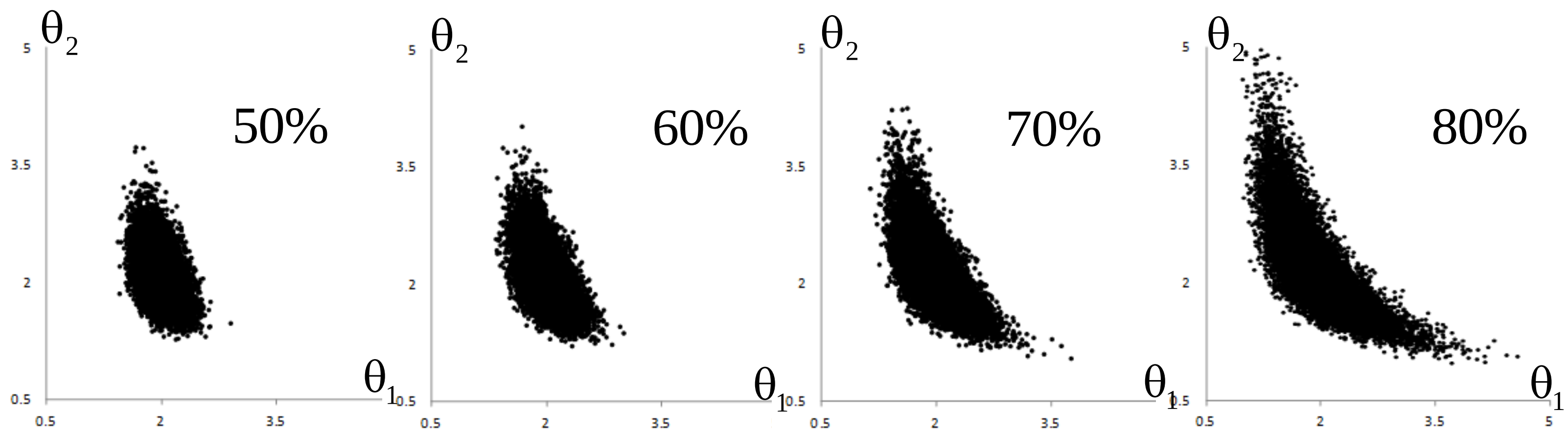
$C_i \sim F^C(t)$

Исследование свойств ОМП параметров распределений по цензурированным справа выборкам

- Показано, что при ограниченных объемах выборок и высоких степенях цензурирования распределение ОМП оказывается асимметричным и далеким от нормального.
- При различных типах и степенях цензурирования найдены оценки смещения ОМП и оценки относительной эффективности оценивания

$\det D[\hat{\theta}] / \det D[\hat{\theta}^c]$, где $\hat{\theta}$ – ОМП по полной выборке,

$\hat{\theta}^c$ – ОМП по цензурированной выборке.

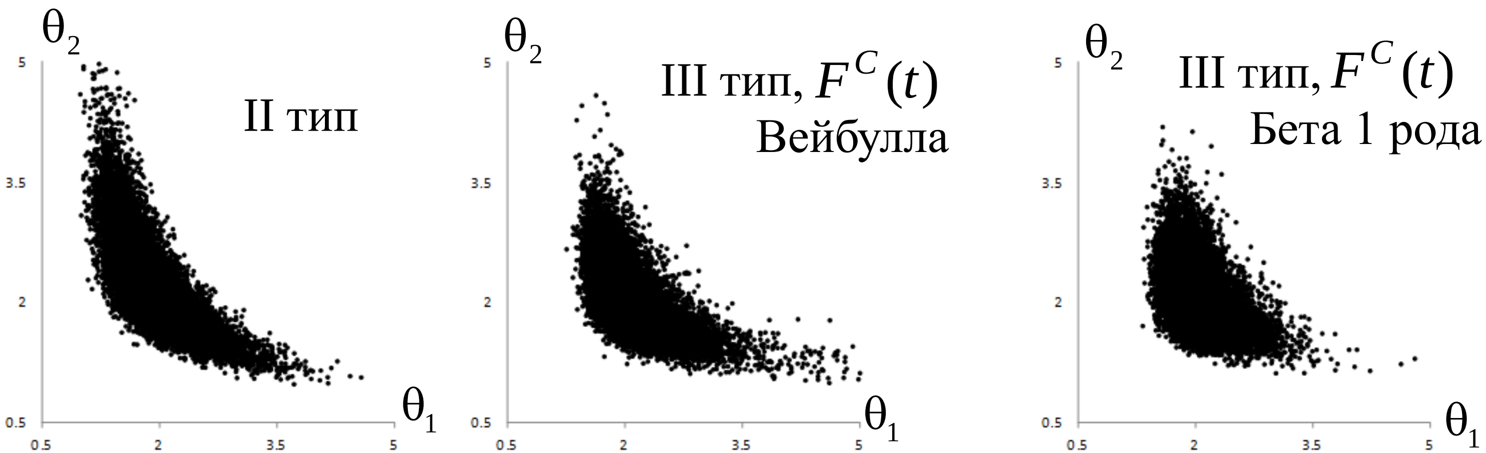


ОМП параметров распределения Вейбулла

$n = 100$, II тип цензурирования

Исследование свойств ОМП параметров распределений по цензурированным справа выборкам

Показано, что при случайном цензурировании и таком распределении, что моменты цензурирования в вариационном ряду расположены примерно равномерно, дисперсия и смещение ОМП оказываются наименьшими по сравнению с другими типами цензурирования.



ОМП параметров распределения Вейбулла
$n=100$, степень цензурирования 80%

Критерии согласия типа χ^2 для цензурированных выборок

$$H_0:\ F(t)\in\{F_0(t;\theta),\ \theta\in\Omega\}$$

Наблюдаемая область $[0,\tau]$ разбивается на k непересекающихся интервалов $I_1,\ldots,I_k$ граничными точками:

$$0=a_{(0)}<a_{(1)}<\ldots<a_{(k-1)}<a_{(k)}=\tau,$$

где τ – последнее полное наблюдение в вариационном ряду выборки.

Обобщенный критерий Пирсона-Фишера

$$\boxed{\tilde{X}_n^2=\xi_n^T(\tilde{\theta}_n)\hat{\Sigma}^{+}\xi_n(\tilde{\theta}_n), \qquad (4)}$$

$$\xi_n(\theta)=\left(\frac{n\hat{p}_1-np_1(\theta)}{\sqrt{np_1(\theta)}},\ldots,\frac{n\hat{p}_k-np_k(\theta)}{\sqrt{np_k(\theta)}}\right)^T,$$

$\hat{p}_j=\hat{F}_n(a_{(j)})-\hat{F}_n(a_{(j-1)})$ – эмпирические вероятности попадания в интервал,

$p_j(\theta)=F_0(a_{(j)};\theta)-F_0(a_{(j-1)};\theta)$ – теоретические вероятности, $j=1,\ldots,k$,

$\hat{F}_n(t)$ – непараметрическая оценка Каплана-Мейера,

$\tilde{\theta}_n$ – оценка параметра распределения $F_0(t;\theta)$ по методу минимума χ^2 (по группированной выборке),

$$G_{\tilde{X}_n^2}(y\,|\,H_0)\to\chi^2_{k-s-1},\quad n\to\infty$$

Критерии согласия типа χ^2 для цензурированных выборок

$$H_0:\ F(t)\in\{F_0(t;\theta),\ \theta\in\Omega\}$$

Наблюдаемая область $[0,\tau]$ разбивается на k непересекающихся интервалов $I_1,\ldots,I_k$ граничными точками:

$$0=a_{(0)}<a_{(1)}<\ldots<a_{(k-1)}<a_{(k)}=\tau,$$

где τ – последнее полное наблюдение в вариационном ряду выборки.

Критерий Никулина-Рао-Робсона

$$\boxed{\tilde{Y}_n^2=Z^T\hat{V}^-Z \qquad (5)}$$

$$Z=\frac{1}{\sqrt{n}}\left(U_1-E_1,\ldots,U_k-E_k\right)^T,$$

$U_j=N(a_{(j)})-N(a_{(j-1)})$ – наблюдаемое число отказов, $N(t)=\sum_{i=1}^{n}\mathbf{1}\{X_i\le t,\delta_i=1\}$,

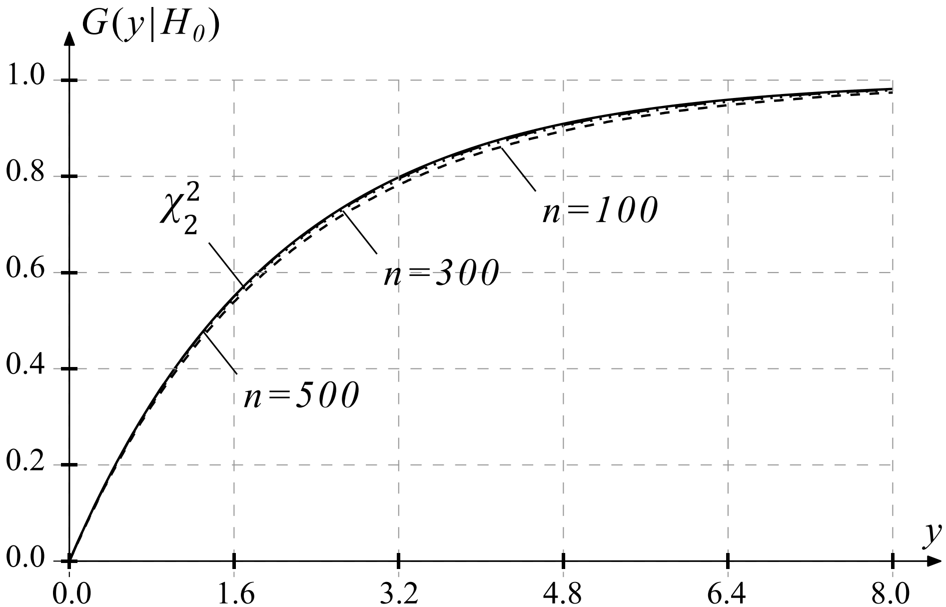
$E_j=\int\limits_{a_{(j-1)}}^{a_{(j)}}\lambda_0\left(u;\hat{\theta}\right)Y(u)du$ – ожидаемое число отказов, $Y(t)=\sum_{i=1}^{n}\mathbf{1}\{X_i\ge t\}$, $j=1,\ldots,k$,

$\hat{\theta}$ – ОМП параметра распределения $F_0(t;\theta)$ по исходной цензурированной выборке.

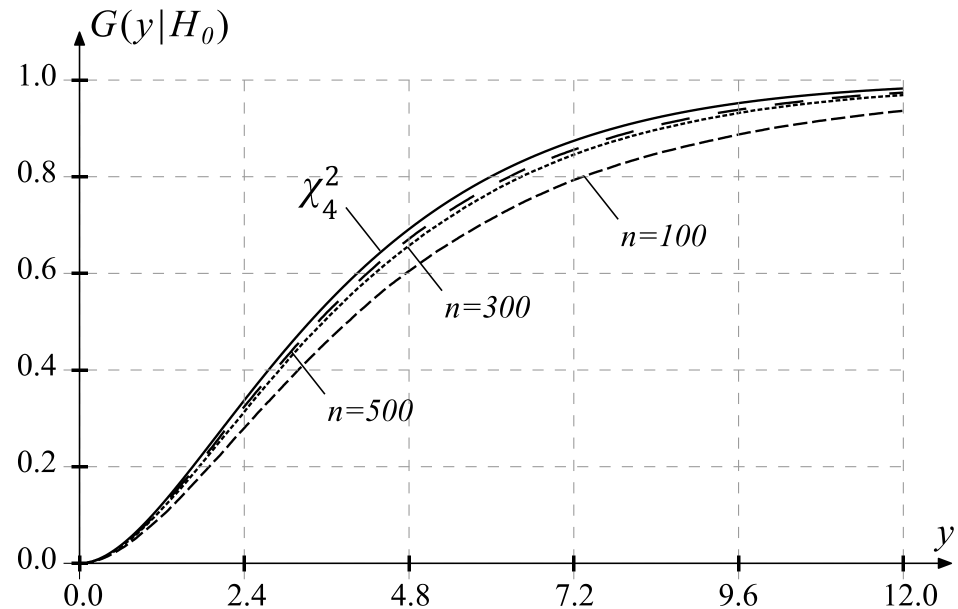
$$G_{\tilde{Y}_n^2}\left(y\,|\,H_0\right)\to\chi_r^2,\quad n\to\infty,\quad r=rank\left(\hat{V}^-\right)$$

Исследование распределений статистик критериев согласия типа χ^2

Показано, что в случае цензурированных выборок распределения статистики обобщенного критерия Пирсона-Фишера сходятся к соответствующему χ^2_{k-s-1} -распределению существенно быстрее, чем распределения статистики Никулина-Рао-Робсона – к своему предельному закону распределения.



Распределения статистики Пирсона–Фишера

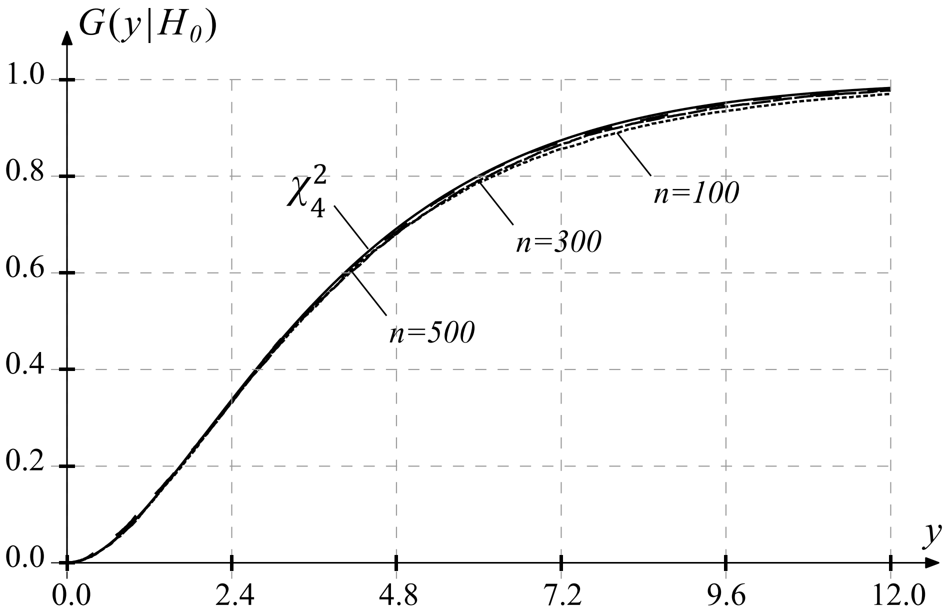


Распределения статистики Никулина-Рао-Робсона

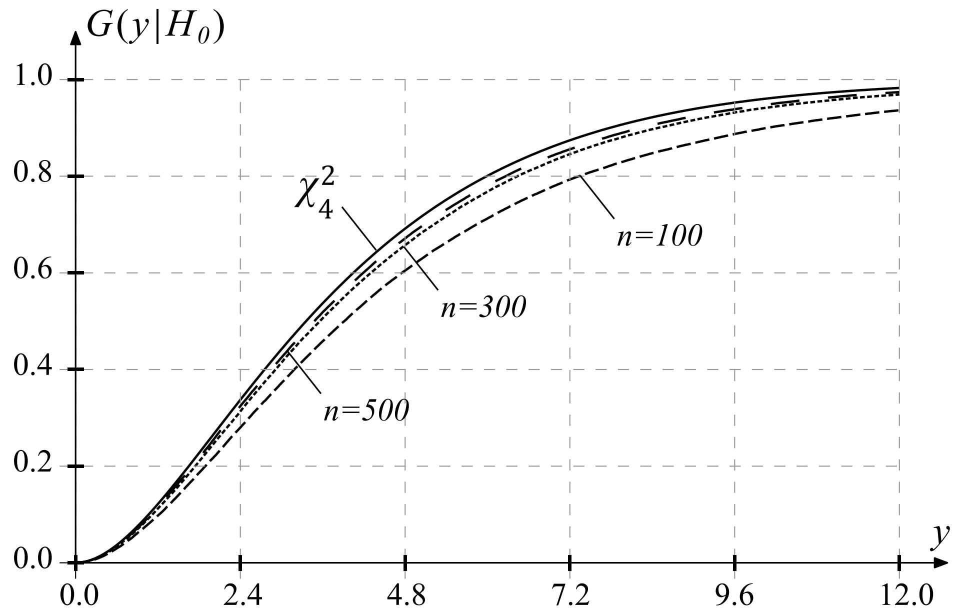
II тип цензурирования, 50%, k =5, равночастотное группирование

Исследование распределений статистик критериев согласия типа χ^2

Показано, что в случае III типа цензурирования распределения статистик критериев ближе к предельным законам, чем при I и II типах цензурирования (при тех же объемах выборок и степенях цензурирования).



Случайное цензурирование (III тип), $F^C(t)$ – бета-распределение 1 рода



II тип цензурирования

Распределения статистики Никулина-Рао-Робсона, степень цензурирования 50%, k =5, равночастотное группирование

Алгоритм оптимального группирования цензурированной выборки

$$H_0: F(t) = F_0\left(t;\theta^0\right) \qquad H_1: F(t) = F_1\left(t;\theta^1\right)$$

Предложен алгоритм оптимального группирования, который позволяет существенно увеличить мощность критериев согласия типа χ^2

Оптимальное разбиение на интервалы $I_1,\ldots,I_k$ может быть получено в результате решения следующей задачи:

$$n\sum_{j=1}^{k}\frac{\left(p_j^1\left(\theta^1\right)-p_j^0\left(\theta^0\right)\right)^2}{p_j^0\left(\theta^0\right)} \to \max_{0<a_{(1)}<a_{(2)}<\cdots<a_{(k-1)}<\tau} \qquad (6)$$

$$p_j^0\left(\theta^0\right)=\int_{a_{(j-1)}}^{a_{(j)}}\left(1-F^C(u)\right) f_0\left(u;\theta^0\right)du, \qquad p_j^1\left(\theta^1\right)=\int_{a_{(j-1)}}^{a_{(j)}}\left(1-F^C(u)\right) f_1\left(u;\theta^1\right)du$$

В качестве $F^C(t)$ предлагается взять оценку Каплана-Мейера*, которая строится по инвертированной выборке:

$$\mathbf{X}_n' = \left\{\left(X_1,\delta_1'\right),\left(X_2,\delta_2'\right),\ldots,\left(X_n,\delta_n'\right)\right\}, \qquad \delta_i' = 1-\delta_i, \qquad i=\overline{1,n}$$

* Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations // Journal of Am. Stat. Assoc. – 1958. – Vol. 53. – P. 457-481.

Алгоритм оптимального группирования цензурированной выборки

1. С использованием метода простого случайного поиска на интервале $(0,\tau)$ найти приближенное решение задачи максимизации (6).

2. Принять полученное решение за начальное приближение

$$\bar{a}^0 = \left(a_{(1)}^0, a_{(2)}^0, \ldots, a_{(k-1)}^0\right)^T .$$

3. Методом прямого направленного поиска минимизировать функцию

$$Q(\bar{a}) = -n\sum_{j=1}^{k} \frac{\left(p_j^1\left(\theta^1\right) - p_j^0\left(\theta^0\right)\right)^2}{p_j^0\left(\theta^0\right)} +$$

$$+\gamma_1\left(-a_{(1)} + \left|a_{(1)}\right|\right)^2 + \gamma_2\left(a_{(1)} - a_{(2)} + \left|a_{(1)} - a_{(2)}\right|\right)^2 + \ldots +$$

$$+\gamma_{k-1}\left(a_{(k-2)} - a_{(k-1)} + \left|a_{(k-2)} - a_{(k-1)}\right|\right)^2 + \gamma_k\left(a_{(k-1)} - \tau + \left|a_{(k-1)} - \tau\right|\right)^2$$

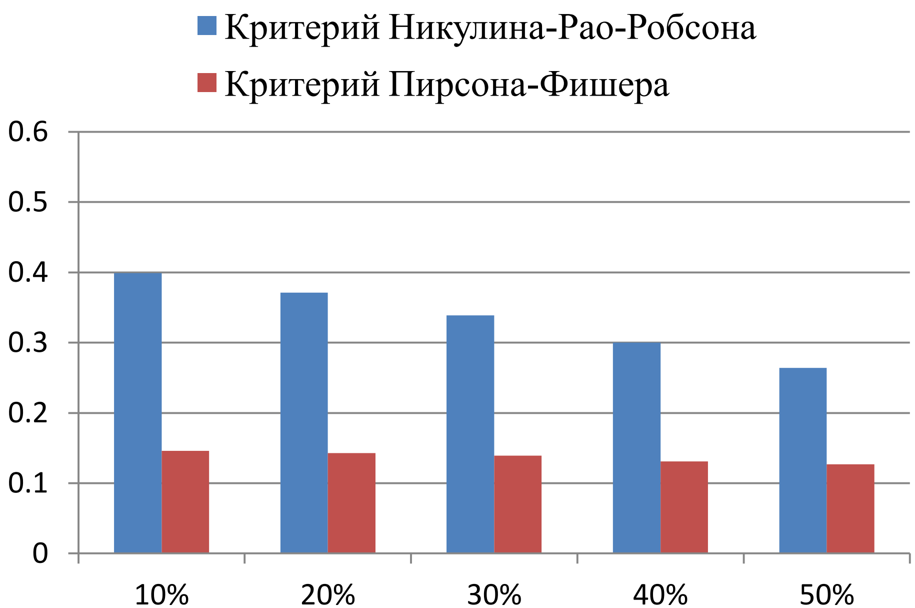
по переменным $a_{(1)}, a_{(2)}, \ldots, a_{(k-1)}$, где $\gamma_j > 0,\ j = \overline{1,k}$ − коэффициенты штрафа, величины которых выбираются достаточно большими.

Исследование мощности критериев согласия типа χ^2

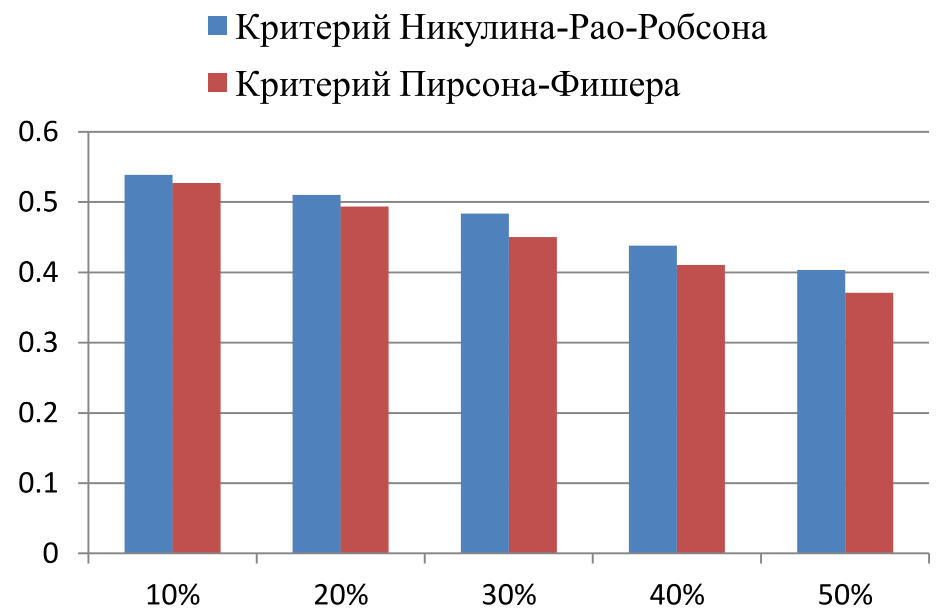
H_0 : распределение Вейбулла, H_1 : гамма-распределение $\alpha = 0.1$

Показано, что мощность критерия Никулина-Рао-Робсона выше мощности обобщенного критерия Пирсона-Фишера.

Мощность критериев согласия типа χ^2 выше при использовании оптимального группирования.



Равночастотное группирование



Оптимальное группирование

Случайно цензурированные выборки, n =200, k = 4

Непараметрические критерии согласия для цензурированных выборок

Модификации непараметрических критериев:

Колмогорова

$$D_n = \sup_{|t|<\tau} \left| \hat{F}_n(t) - F_0\left(t;\hat{\theta}\right) \right|$$

Крамера-Мизеса-Смирнова

$$\omega^2 = \int_0^{\tau} \left(\hat{F}_n(t) - F_0\left(t;\hat{\theta}\right) \right)^2 dF_0\left(t;\hat{\theta}\right)$$

Андерсона-Дарлинга

$$\Omega^2 = \int_0^{\tau} \left(\hat{F}_n(t) - F_0\left(t;\hat{\theta}\right) \right)^2 \frac{dF_0\left(t;\hat{\theta}\right)}{F_0\left(t;\hat{\theta}\right)\left(1 - F_0\left(t;\hat{\theta}\right)\right)}$$

Оценка Каплана-Мейера

Таблицы верхних процентных точек критериев согласия для цензурированных I и II типа выборок

Построены таблицы верхних процентных точек модифицированных критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для проверки **простых** гипотез по цензурированным I и II типа выборкам.

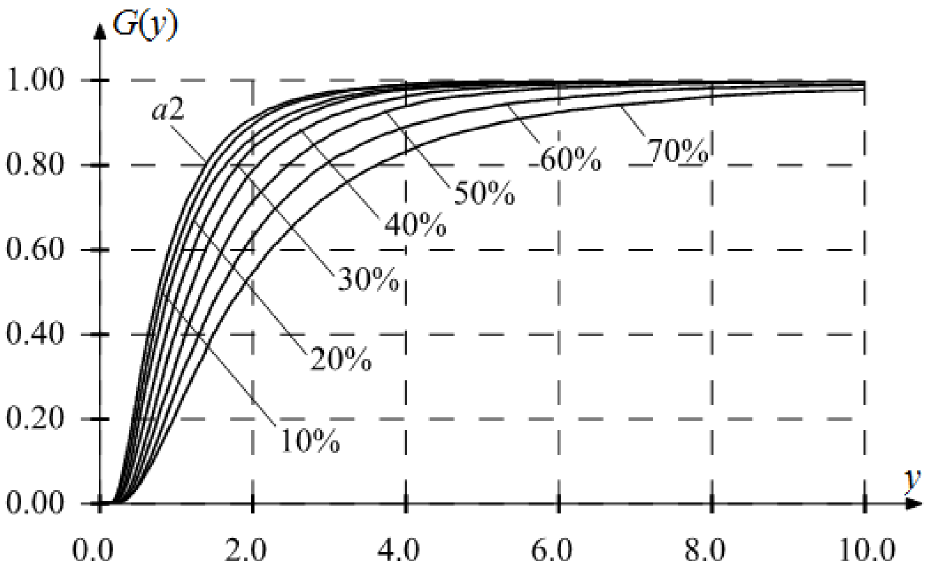
Степень цензурирования	Уровень значимости α		
	0.01	0.05	0.1
0%	1.628	1.358	1.224
10%	1.628	1.358	1.224
20%	1.621	1.347	1.209
30%	1.621	1.347	1.209
40%	1.600	1.321	1.181
50%	1.551	1.273	1.133
60%	1.467	1.198	1.062
70%	1.342	1.087	0.960
80%	1.151	0.927	0.815
90%	0.851	0.682	0.599

Верхние процентные точки модифицированного критерия Колмогорова для цензурированных I и II типа выборок

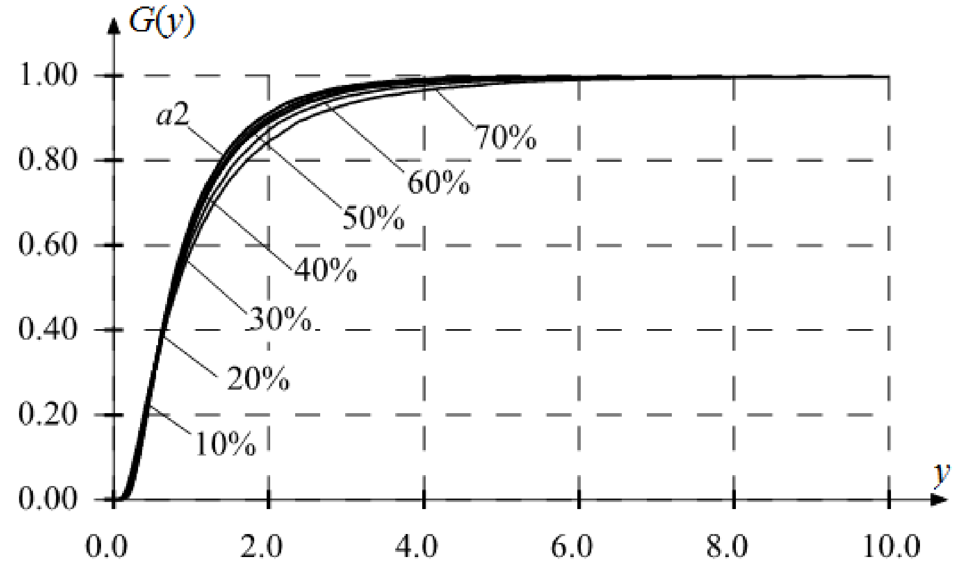
Аналогичные таблицы построены для проверки **сложных** гипотез относительно законов Вейбулла, экспоненциального и логнормального.

Исследование распределений статистик непараметрических критериев согласия при случайном цензурировании

Показано, что в случае III типа цензурирования распределения статистик модифицированных критериев согласия зависят от объема выборки, типа и степени цензурирования, а также от распределения моментов цензурирования.



$F^C(t)$ – бета 1-го рода



$F^C(t)$ – Вейбулла

Распределения статистики модифицированного критерия Андерсона-Дарлинга при проверке простой гипотезы при различных степенях цензурирования

Алгоритм непараметрического моделирования случайно цензурированных выборок

Чтобы смоделировать случайно цензурированную выборку в соответствии с механизмом цензурирования исходной (эталонной) выборки необходимо:

1) смоделировать полную выборку объемом n в соответствии с распределением, соответствующим проверяемой нулевой гипотезе: $T_i = F_0^{-1}(\zeta_i;\theta)$, где $\zeta_i \sim \mathrm{Uniform}(0,1)$, $i = 1,2,\dots,n$;
2) инвертировать исходную цензурированную выборку;
3) построить оценку Каплана-Мейера $\hat{F}_n^C(t)$ по инвертированной выборке;
4) смоделировать выборку ξ_i, равномерно распределенную на интервале (0,1), и вычислить значения C_i:
   a. если $\xi_i < \hat{F}_n^C(c_1)$, то $C_i = \dfrac{\xi_i \cdot c_1}{\hat{F}_n^C(c_1)}$,
   b. если $\xi_i \in \left(\hat{F}_n^C(c_j), \hat{F}_n^C(c_{j+1})\right]$, то $C_i = c_j + \dfrac{\left(\xi_i - \hat{F}_n^C(c_j)\right)\cdot\left(c_{j+1} - c_j\right)}{\left(\hat{F}_n^C(c_{j+1}) - \hat{F}_n^C(c_j)\right)}$, $j = 1,2,\dots,r$,
   c. если $\xi_i > \hat{F}_n^C(c_r)$, то $C_i = c_r + c_r\left(\xi_i - \hat{F}_n^C(c_r)\right)$,

   где $c_1,\dots,c_r$ – упорядоченные по возрастанию различные моменты цензурирования в исходной выборке, r – количество различных моментов цензурирования в исходной выборке, $i = 1,2,\dots,n$;
5) $X_i = \min(T_i, C_i), \delta_i = 1\{T_i \le C_i\}, i = 1,2,\dots,n$.

Преобразование цензурированной выборки в псевдополную

Для проверки *сложной гипотезы* о виде распределения на основе преобразования цензурированной выборки в псевдополную необходимо:

1) Оценить методом максимального правдоподобия значения неизвестных параметров распределения $F_0(t;\theta)$ по исходной цензурированной выборке $(X_1,\delta_1),(X_2,\delta_2),...,(X_n,\delta_n)$;
2) Преобразовать выборку $(X_1,\delta_1),(X_2,\delta_2),...,(X_n,\delta_n)$ в псевдополную $(\hat{X}_1,\hat{\delta}_1),(\hat{X}_2,\hat{\delta}_2),...,(\hat{X}_n,\hat{\delta}_n)$:
$$\hat{X}_i = \begin{cases} X_i = T_i, \text{ если } \delta_i = 1; \\ \hat{T}_i = F_0^{-1}(\xi_i;\hat{\theta}),\ \xi_i \sim \text{Uniform}\left[F_0(C_i;\hat{\theta}), 1\right], \text{ если } \delta_i = 0; \end{cases}$$
3) Вычислить оценки максимального правдоподобия $\tilde{\theta}$ неизвестных параметров по псевдополной выборке $\hat{X}_1,\hat{X}_2,...,\hat{X}_n$;
4) Вычислить значение статистики критерия согласия S^* по полученной выборке и теоретическому распределению $F_0(t;\tilde{\theta})$;
5) Оценить достигнутый уровень значимости $\alpha^* = 1 - G\left(S^* \mid H_0\right)$;

Отклонить проверяемую гипотезу $H_0 : F(t) \in \{F(t;\theta), \theta \in \Omega\}$, если $\alpha^* < \alpha$, где α – заданный уровень значимости.

Исследование мощности непараметрических критериев согласия для цензурированных выборок

H_0 : распределение Вейбулла,

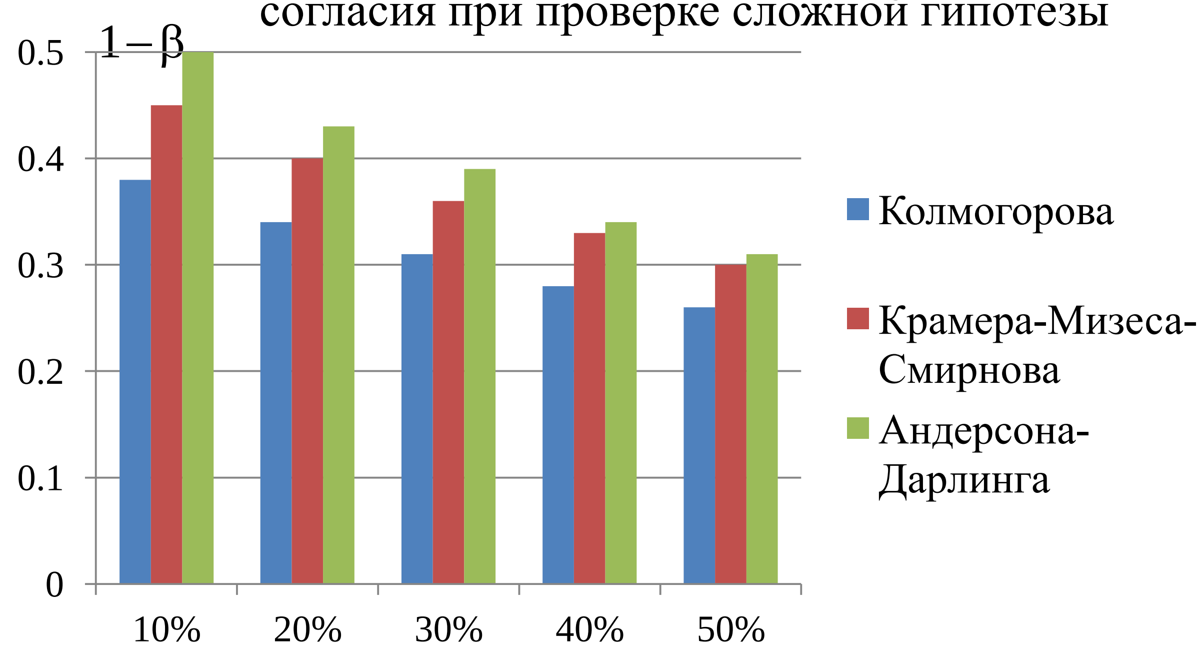
H_1 : гамма-распределение

Цензурирование II типа

$n = 200$

$\alpha = 0.1$

Оценки мощности модифицированных критериев согласия при проверке сложной гипотезы



Показано, что классические критерии согласия по псевдополным выборкам незначительно уступают по мощности модифицированным критериям.

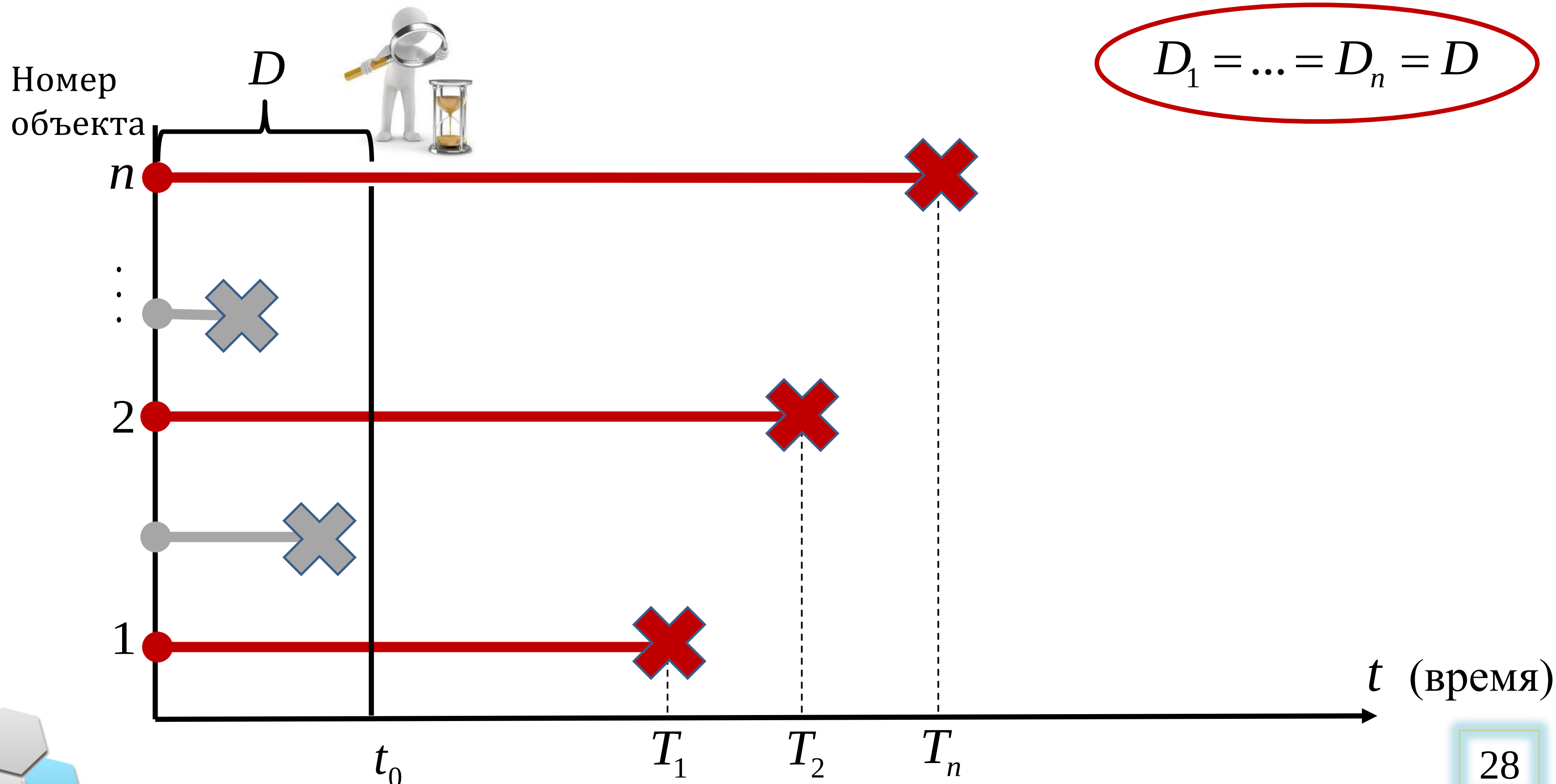
Для большинства рассмотренных случаев наиболее предпочтительным по мощности оказался критерий типа Андерсона-Дарлинга.

Выборка усеченных слева наблюдений

$$\mathbf{X}_n = \{(T_1, D_1), (T_2, D_2), \dots, (T_n, D_n)\}$$

T_i – время отказа, D_i – время усечения, $i = \overline{1, n}$

$D_1 = \dots = D_n = D$

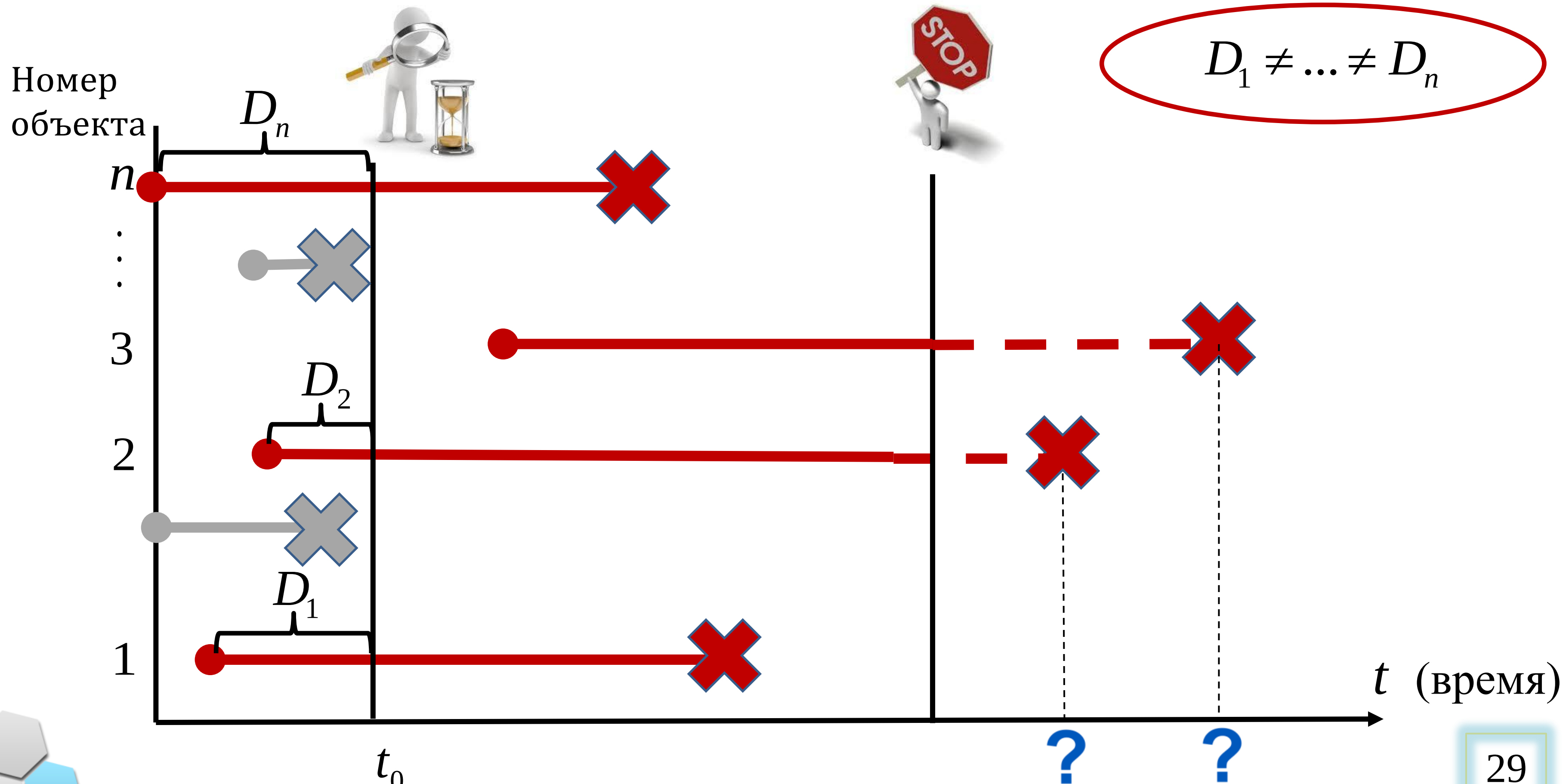


Выборка усеченных слева наблюдений

$$\mathbf{X}_n = \{(T_1, D_1, \delta_1), (T_2, D_2, \delta_2), \ldots, (T_n, D_n, \delta_n)\}$$

T_i – время отказа, D_i – время усечения, δ_i – индикатор цензурирования, $i = \overline{1, n}$

$D_1 \neq \ldots \neq D_n$

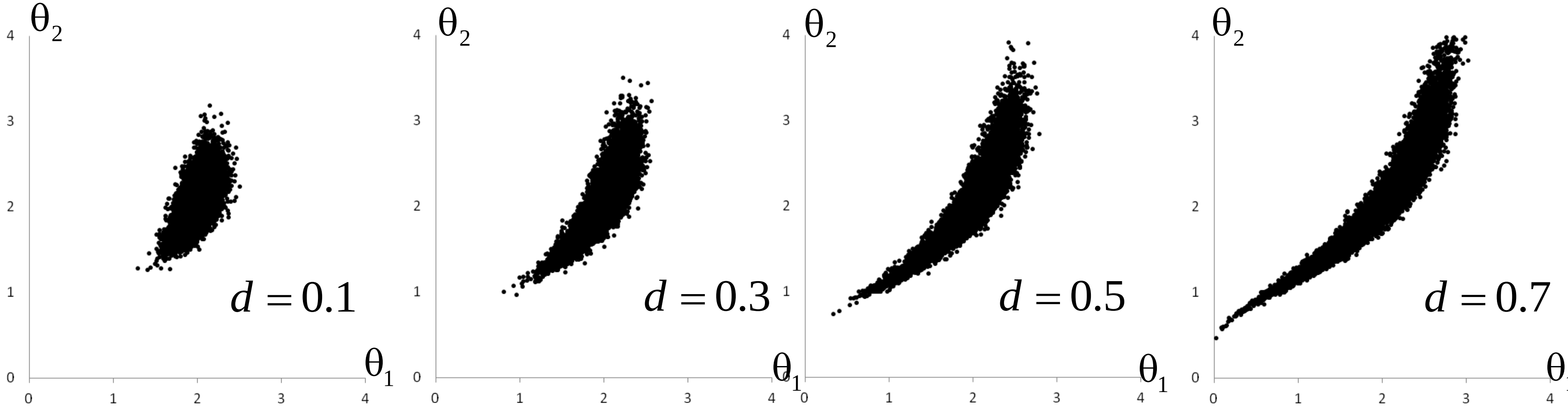


Исследование свойств ОМП параметров распределений по выборкам усеченных слева наблюдений

Показано, что на свойства ОМП при ограниченных объемах выборок существенное влияние оказывают степень усечения и процент наблюдений в выборке, принадлежащих усеченному распределению.

С ростом степени усечения распределение ОМП существенно отклоняется от многомерного нормального закона.

Степень усечения – вероятность попадания в усеченную область: $d = F(D)$



ОМП параметров распределения Вейбулла
$n = 100$

Методика проверки гипотез о виде распределения по выборкам усеченных наблюдений

$$H_0: \; F_T(t) \in \left\{F_0\left(t;\theta\right), \theta \in \Omega\right\}$$

$$P\left\{T_i < t \,|\, T_i \ge D_i\right\} = \frac{F(t) - F(D_i)}{1 - F(D_i)},$$

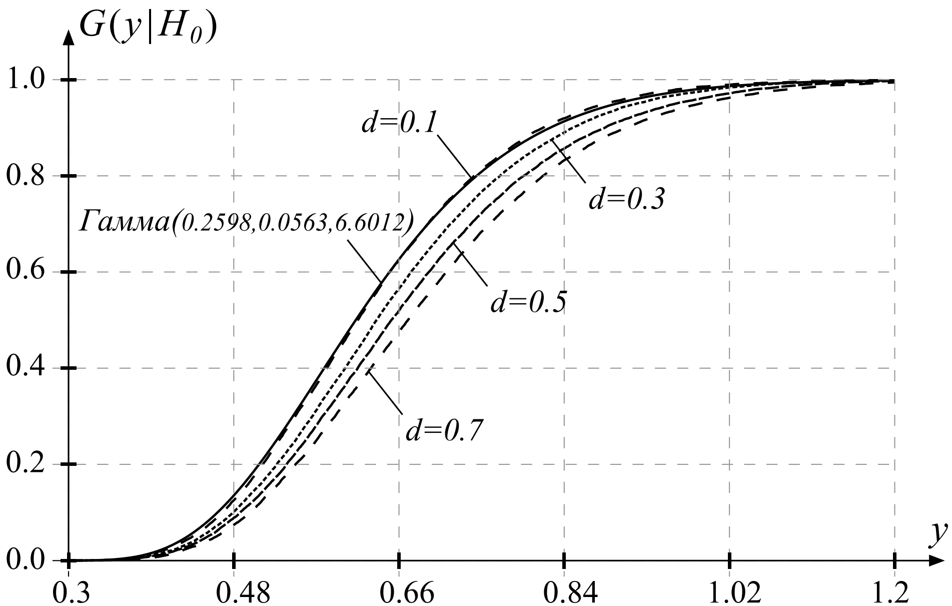
$$U_i = \frac{F_0\left(T_i;\hat{\theta}\right) - F_0\left(D_i;\hat{\theta}\right)}{1 - F_0\left(D_i;\hat{\theta}\right)}, \qquad i = \overline{1,n}$$

$$H_0: \; F_U(t) = t, \; 0 < t < 1 \qquad \left(\text{Uniform}\left(0,1\right)\right)$$

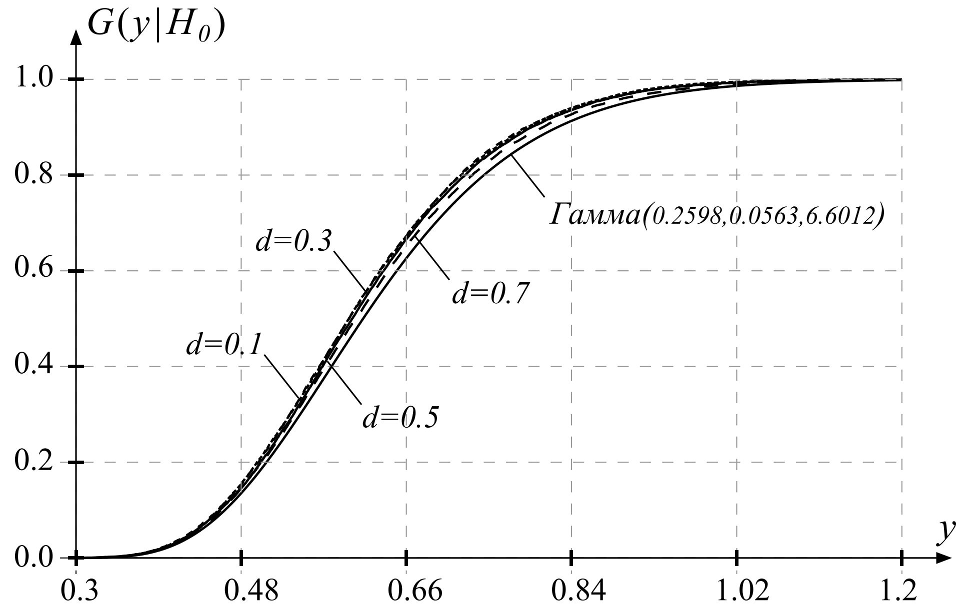
Для проверки гипотезы о равномерном распределении можно использовать критерии согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга

Исследование распределений статистик критериев согласия по выборкам усеченных наблюдений

Показано, что при проверке сложных гипотез по выборкам усеченных слева наблюдений распределения статистик критериев согласия зависят от степени усечения d и процента наблюдений из усеченного распределения в выборке.



50% наблюдений из усеченного распределения



100% наблюдений из усеченного распределения

Распределения статистики Колмогорова при проверке сложной гипотезы о распределении Вейбулла, n = 50

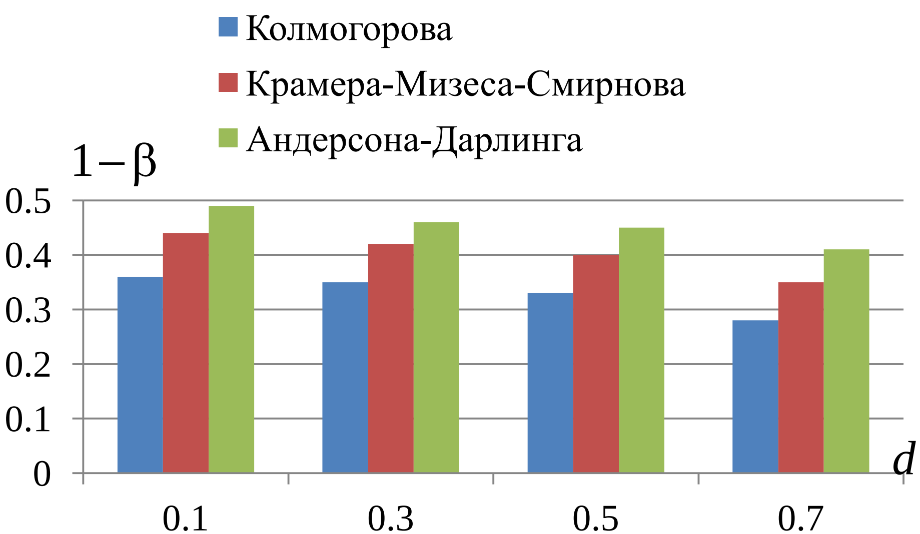
Сравнительный анализ мощности критериев согласия по выборкам усеченных слева наблюдений

Показано, что для рассмотренных в работе пар близких конкурирующих гипотез наиболее предпочтительным по мощности является критерий типа Андерсона-Дарлинга.

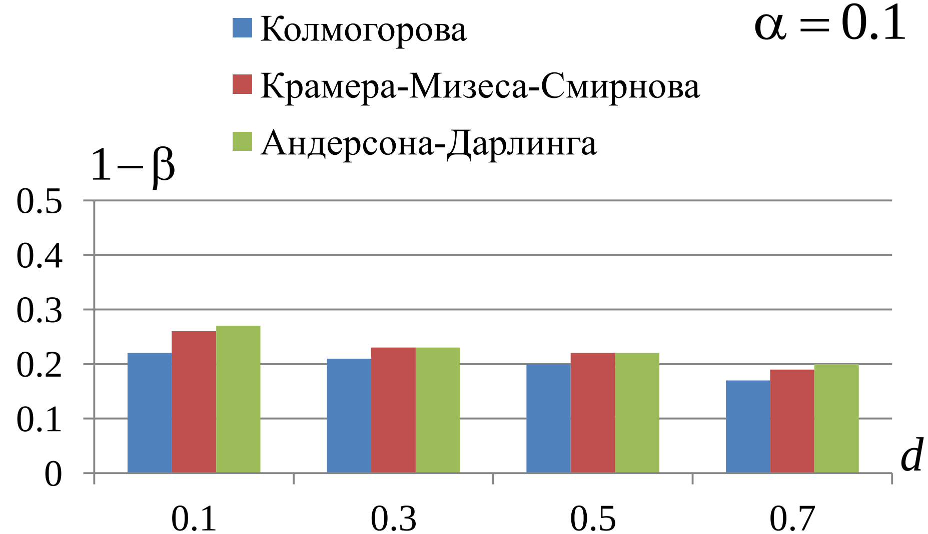
H_0 : распределение Вейбулла, H_1 : гамма-распределение

$n = 200$

$\alpha = 0.1$



25% наблюдений из усеченного распределения



75% наблюдений из усеченного распределения

Проверка гипотезы о виде параметрической регрессионной модели по цензурированным справа выборкам усеченных слева наблюдений

$$H_0:\ S_x(t)\in\left\{g\left(S_0(t;\theta);x;\beta\right),\ \theta,\beta\in\Omega\right\}$$

$$\mathbf{Z}_n=\left\{\left(X_1,D_1,\delta_1,x^{(1)}\right),\left(X_2,D_2,\delta_2,x^{(2)}\right),\ldots,\left(X_n,D_n,\delta_n,x^{(n)}\right)\right\}$$

$$\Downarrow$$

$$\left(X_1^0,D_1^0,\delta_1\right),\left(X_2^0,D_2^0,\delta_2\right),\ldots,\left(X_n^0,D_n^0,\delta_n\right)$$

$$X_i^0=S_0^{-1}\left(S_{x^{(i)}}\left(X_i;\hat\beta,\hat\theta\right)\right),\quad D_i^0=S_0^{-1}\left(S_{x^{(i)}}\left(D_i;\hat\beta,\hat\theta\right)\right),\ i=\overline{1,n}$$

$$\Downarrow$$

$$(U_1,\delta_1),(U_2,\delta_2),\ldots,(U_n,\delta_n),\ U_i=\frac{F_0\left(X_i^0;\hat\theta\right)-F_0\left(D_i^0;\hat\theta\right)}{1-F_0\left(D_i^0;\hat\theta\right)},\ i=\overline{1,n}$$

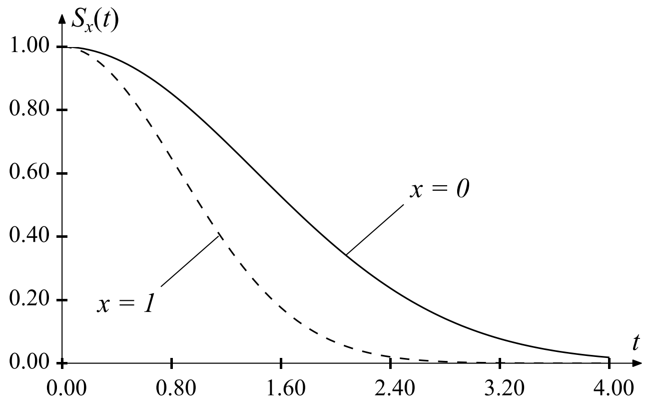
$$H_0:\ F_U(t)=t,\ 0<t<1\qquad\left(\mathrm{Uniform}(0,1)\right)$$

Алгоритм моделирования распределений статистик критериев согласия

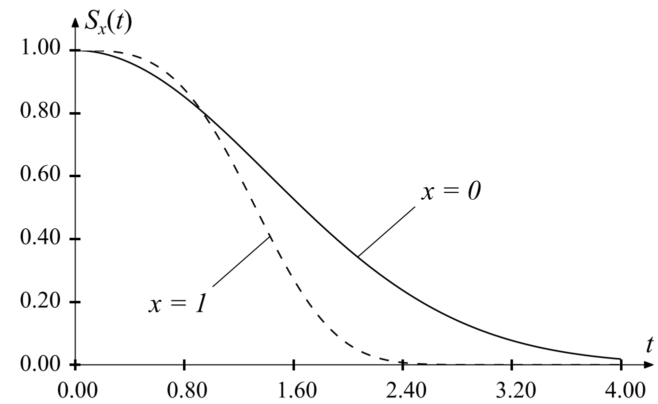
1. На основе построенной параметрической регрессионной модели, сгенерировать выборку $\mathbf{Y}_n$, аналогичную исходной выборке $\mathbf{Z}_n$, по заданному плану эксперимента, выполнив следующую последовательность действий:
   1.1. Сгенерировать усеченную слева выборку $\mathbf{Y}_n$ без цензурирования. Для формирования i-го элемента выборки $\mathbf{Y}_n$, $i=\overline{1,n}$, генерируется последовательность величин $\xi_j \sim \mathrm{Uniform}(0,1)$, $j=1,2,\ldots$, пока не выполнится условие $T_i = F_{x^i}^{-1}(\xi_j) \geq D_i$, где D_i – время усечения i-го наблюдения выборки $\mathbf{Z}_n$.
   1.2. В соответствии с типом цензурирования выборки $\mathbf{Z}_n$ преобразовать полученную выборку $\mathbf{Y}_n$ в цензурированную.
2. По полученной выборке $\mathbf{Y}_n$ оценить параметры проверяемой модели методом максимального правдоподобия.
3. Сформировать выборку остатков и вычислить значение статистики критерия согласия.
4. Повторить пункты 1 – 4 N раз. По полученной выборке статистик объёма N построить эмпирическую функцию распределения статистики критерия $G_N(S \mid H_0)$.

Исследование мощности критериев согласия, применяемых к выборкам остатков

$H_0: \ \Lambda_x(t;\beta) = e^{\beta^T x} \cdot \Lambda_0(t;\theta)$



$H_1: \ \Lambda_x(t;\beta,\gamma,\theta) = \exp\left(\beta^T x\right)\left\{\Lambda_0(t;\theta)\right\}^{\exp\left(\gamma^T x\right)}$



Показано, что в случае цензурированных справа выборок усеченных слева наблюдений критерий типа Андерсона-Дарлинга уступает по мощности критериям типа Колмогорова и Крамера-Мизеса-Смирнова.

$H_0: \ \Lambda_x(t;\beta) = e^{\beta^T x} \cdot \Lambda_0(t;\theta)$, базовое экспоненциальное распределение;

$H_1: \ \Lambda_x(t;\beta) = e^{\beta^T x} \cdot \Lambda_0(t;\theta)$, базовое распределение Вейбулла

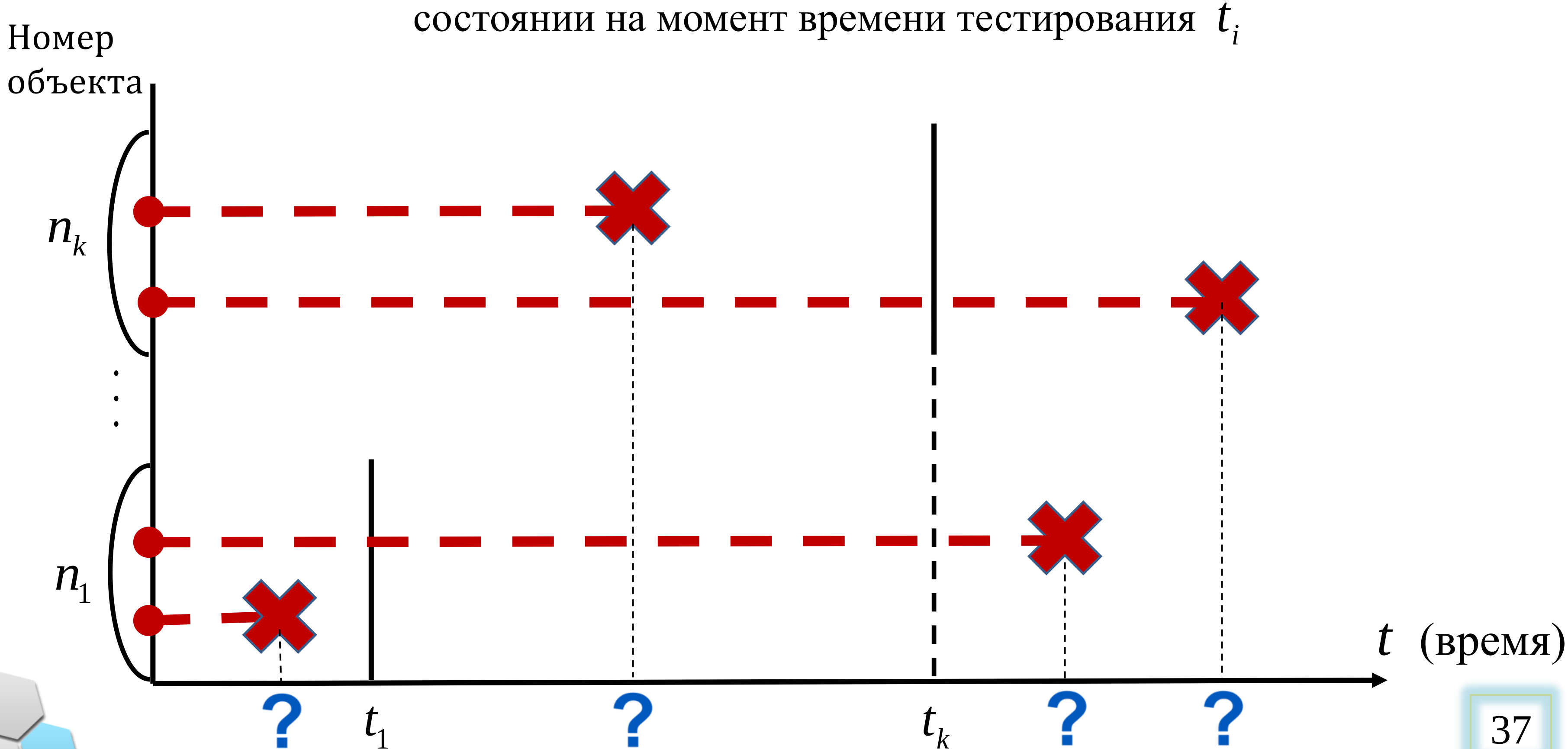
Выборка текущих состояний

$$\mathbf{X}_n = \{(t_i, K_i, n_i),\ i = 1,\ldots,k\}$$

$t_1, t_2, \ldots, t_k$ – моменты времени тестирования,

n_i – количество объектов, тестируемых в момент времени t_i,

K_i – количество устройств, оказавшихся в неработоспособном состоянии на момент времени тестирования t_i



Оптимальное планирование испытаний на надежность устройств одноразового срабатывания

Оптимальные моменты времени тестирования устройств одноразового срабатывания будем искать в виде дискретного нормированного плана:

$$\varepsilon_k = \begin{Bmatrix} t_1, & t_2, & \ldots, & t_k \\ p_1, & p_2, & \ldots, & p_k \end{Bmatrix}, \qquad p_i = n_i / n, \qquad \sum_{i=1}^{k} p_i = 1$$

$$\det I_{CS}(\theta) \to \max_{t_1, t_2, \ldots, t_k;\; n_1, \ldots, n_k} \qquad (7)$$

при условиях: $0 < t_1 < t_2 < \ldots < t_k, \; \sum_{i=1}^{k} n_i = n, \;\; n_i \in \mathbf{N}, \; i = \overline{1,k}$

$$I_{CS}(\theta) = \sum_{i=1}^{k} \frac{n_i}{F(t_i;\theta)\big(1 - F(t_i;\theta)\big)} \frac{\partial F(t_i;\theta)}{\partial \theta} \frac{\partial F(t_i;\theta)}{\partial \theta^T}$$

- ✓ Экспоненциальное
- ✓ Вейбулла
- ✓ Логнормальное
- ✓ Гамма

Непараметрическая ОМП функции распределения по выборке текущих состояний

$$\ln L(\mathbf{X}_n) = \sum_{i=1}^{k} \left(K_i \ln F_i + (n_i - K_i) \ln (1 - F_i) \right) \to \max_{0 \le F_1 \le F_2 \le \ldots \le F_k \le 1} \qquad (8)$$

где $F_1 = F(t_1),\ F_2 = F(t_2), \ldots,\ F_k = F(t_k)$

Предложен алгоритм вычисления непараметрической ОМП:

1. Вычислить начальные значения $\hat{F}_i = \dfrac{K_i}{n_i}$.
2. Если найденные значения $\hat{F}_1, \ldots, \hat{F}_k$ удовлетворяют ограничениям $\hat{F}_1 \le \hat{F}_2 \le \ldots \le \hat{F}_k$, то решение найдено, иначе перейти на шаг 3.
3. Найти наименьший номер $i < k$, для которого $\hat{F}_i > \hat{F}_{i+1}$.
4. Пересчитать значения, удовлетворяющие неравенству $\hat{F}_i > \hat{F}_{i+1} \ge \ldots \ge \hat{F}_{i+m}$:

$$\hat{F}_i = \ldots = \hat{F}_{i+m} = \frac{\sum_{j=i}^{i+m} K_j}{\sum_{j=i}^{i+m} n_j}.$$

5. Повторять шаги 3-4 до тех пор, пока не выполнится ограничение $\hat{F}_1 \le \hat{F}_2 \le \ldots \le \hat{F}_k$.

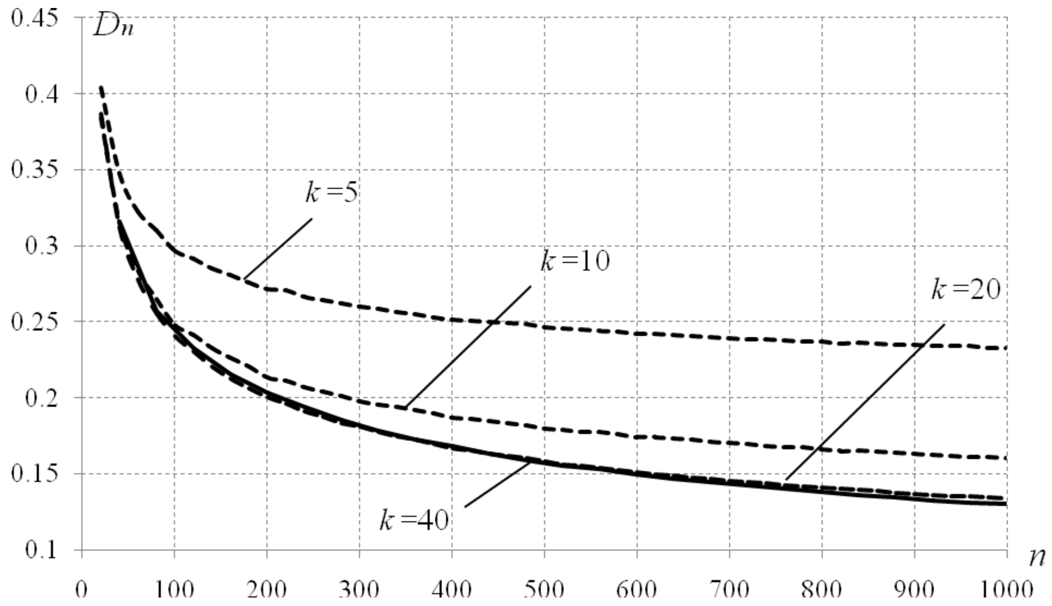
Непараметрическая ОМП функции распределения по выборке текущих состояний

$$\ln L(\mathbf{X}_n) = \sum_{i=1}^{k}\Big(K_i \ln F_i + (n_i - K_i)\ln(1 - F_i)\Big) \to \max_{0\le F_1 \le F_2 \le \ldots \le F_k \le 1} \qquad (8)$$

где $F_1 = F(t_1),\ F_2 = F(t_2), \ldots,\ F_k = F(t_k)$

Предложен алгоритм вычисления непараметрической ОМП.

Проведено исследование скорости сходимости непараметрической ОМП к истинной функции распределения отказов.



Зависимость расстояния

$$D_n = \sup_{0<t<t_k}\left|\hat{F}(t) - F(t)\right|$$

от объема выборки n, при различном числе моментов тестирования k

Разработка критериев согласия для выборок текущих состояний

Модификации непараметрических критериев:

- Колмогорова

$$D_n = \sup_{|t|<\tau} \left| \hat{F}(t) - F_0\left(t;\hat{\theta}\right) \right|$$

- Крамера-Мизеса-Смирнова

$$\omega^2 = \int_0^{\tau} \left(\hat{F}(t) - F_0\left(t;\hat{\theta}\right) \right)^2 dF_0\left(t;\hat{\theta}\right)$$

- Критерий типа хи-квадрат

$$Z_n^2 = \sum_{i=1}^{k} \frac{(\nu_i - u_i)^2}{u_i},$$

$\nu_i = n_i \hat{F}(t_i)$ – эмпирическое число отказов к моменту времени t_i,

$u_i = n_i F_0\left(t_i;\hat{\theta}\right)$ – ожидаемое число отказов к моменту времени t_i в соответсвии с проверяемой гипотезой, $i = \overline{1,k}$.

Непараметрическая ОМП

Разработка критериев согласия для выборок текущих состояний

- Критерий типа Уайта*

$$I(\theta) = -M\left[\frac{\partial^2 \ln L(\mathbf{X}_n;\theta)}{\partial\theta^2}\right] = M\left[\left(\frac{\partial \ln L(\mathbf{X}_n;\theta)}{\partial\theta}\right)^2\right]$$

$$A_n\left(\hat{\theta}\right) = -\frac{1}{n}\sum_{i=1}^{k}\left[K_i\frac{\partial^2 \ln F_0\left(t_i;\hat{\theta}\right)}{\partial\theta^2} + (n_i - K_i)\frac{\partial^2 \ln\left(1 - F_0\left(t_i;\hat{\theta}\right)\right)}{\partial\theta^2}\right] \quad \text{или}$$

$$B_n\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{k}\left[K_i\frac{\partial \ln F_0\left(t_i;\hat{\theta}\right)}{\partial\theta}\left(\frac{\partial \ln F_0\left(t_i;\hat{\theta}\right)}{\partial\theta}\right)^T + (n_i - K_i)\frac{\partial \ln\left(1 - F_0\left(t_i;\hat{\theta}\right)\right)}{\partial\theta}\left(\frac{\partial \ln\left(1 - F_0\left(t_i;\hat{\theta}\right)\right)}{\partial\theta}\right)^T\right]$$

$$S_W = \sqrt{n}\,\frac{\left|\det A_n\left(\hat{\theta}\right) - \det B_n\left(\hat{\theta}\right)\right|}{\det B_n\left(\hat{\theta}\right)},$$

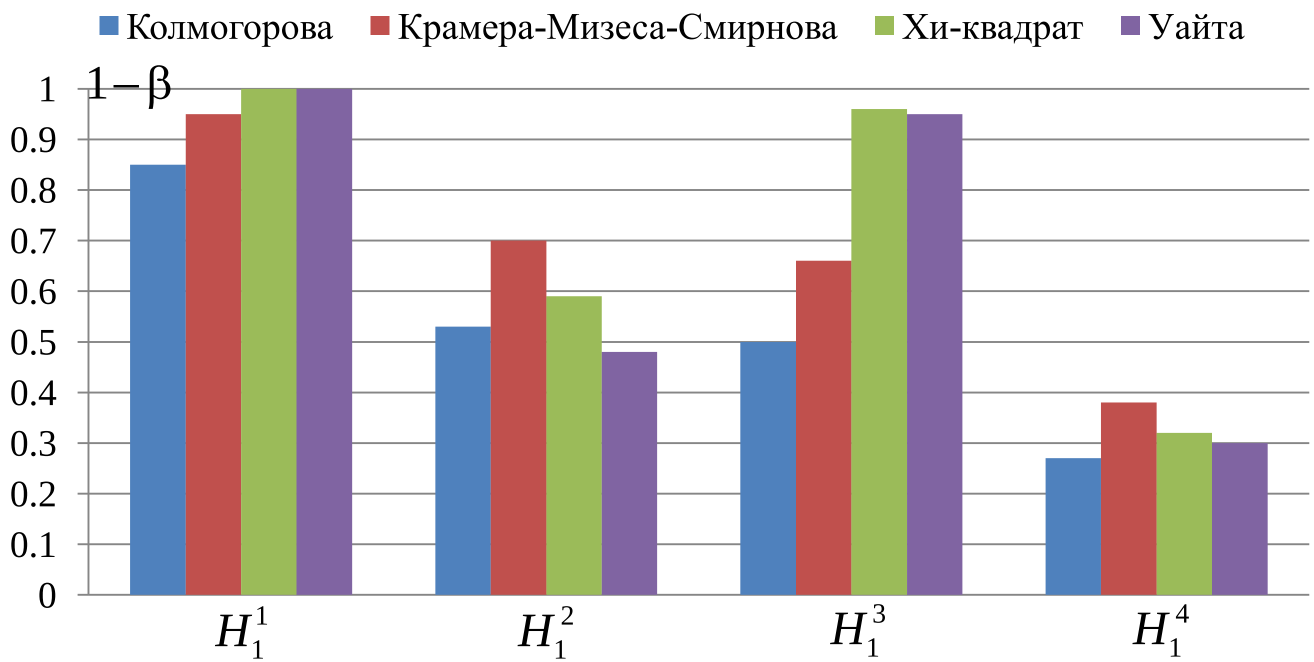
$\hat{\theta}$ – ОМП параметра распределения $F_0(t;\theta)$

* White H. Maximum likelihood estimation of misspecified models // Econometrica. – 1982. – Vol. 50. – P. 1-25.

Сравнительный анализ мощности критериев согласия для выборок текущих состояний

Показано, что для рассмотренных пар конкурирующих гипотез более предпочтительными из предложенных являются критерии согласия типа хи-квадрат и типа Крамера-Мизеса-Смирнова.

H_0 : экспоненциальное распределение



H_1^1 : Вейбулла $(\theta_2 = 0.5)$

H_1^2 : Вейбулла $(\theta_2 = 1.5)$

H_1^3 : гамма $(\theta_2 = 0.5)$

H_1^4 : гамма $(\theta_2 = 1.5)$

$n = 200$

$k = 20$

$\alpha = 0.1$

Оценки мощности критериев согласия при проверке сложной гипотезы

Проверка гипотезы о виде параметрической регрессионной модели по выборке текущих состояний

$$H_0: \ S_x(t) \in \left\{ g\left(S_0(t;\theta); x; \beta\right), \ \theta, \beta \in \Omega \right\}$$

$$\mathbf{Z}_n = \left\{ (t_i, x^{(i,j)}, K_{ij}, n_{ij}), \quad i = \overline{1,k}, \quad j = \overline{1,q_i} \right\}$$

Момент времени тестирования	План эксперимента	Количество отказавших устройств
t_1	$\xi_{n_1}^{q_1} = \left\{ \begin{matrix} x^1 & \dots & x^{q_1} \\ n_{11} & \dots & n_{1q_1} \end{matrix} \right\}$	$K_{11}, \dots, K_{1q_1}$
$\dots$	$\dots$	$\dots$
t_k	$\xi_{n_k}^{q_k} = \left\{ \begin{matrix} x^1 & \dots & x^{q_k} \\ n_{k1} & \dots & n_{kq_k} \end{matrix} \right\}$	$K_{k1}, \dots, K_{kq_k}$

Выборка остатков:

$$\left\{ (t_{ij}^0, K_{ij}, n_{ij}), \ i = \overline{1,k}, \ j = \overline{1,q_i} \right\},$$

$$t_{ij}^0 = S_0^{-1}\left(S_{x^j}\left(t_i; \hat{\beta}, \hat{\theta} \right) \right)$$

$$H_0: \ F_e(t) \in \left\{ F_0(t;\theta), \ \theta \in \Omega \right\}$$

Показано, что для ряда пар конкурирующих гипотез наибольшую мощность демонстрирует критерий типа Уайта, в то время как для других пар гипотез более мощными оказываются критерии типа Колмогорова и Крамера-Мизеса-Смирнова. Критерий типа хи-квадрат во всех рассмотренных ситуациях уступил по мощности упомянутым критериям.

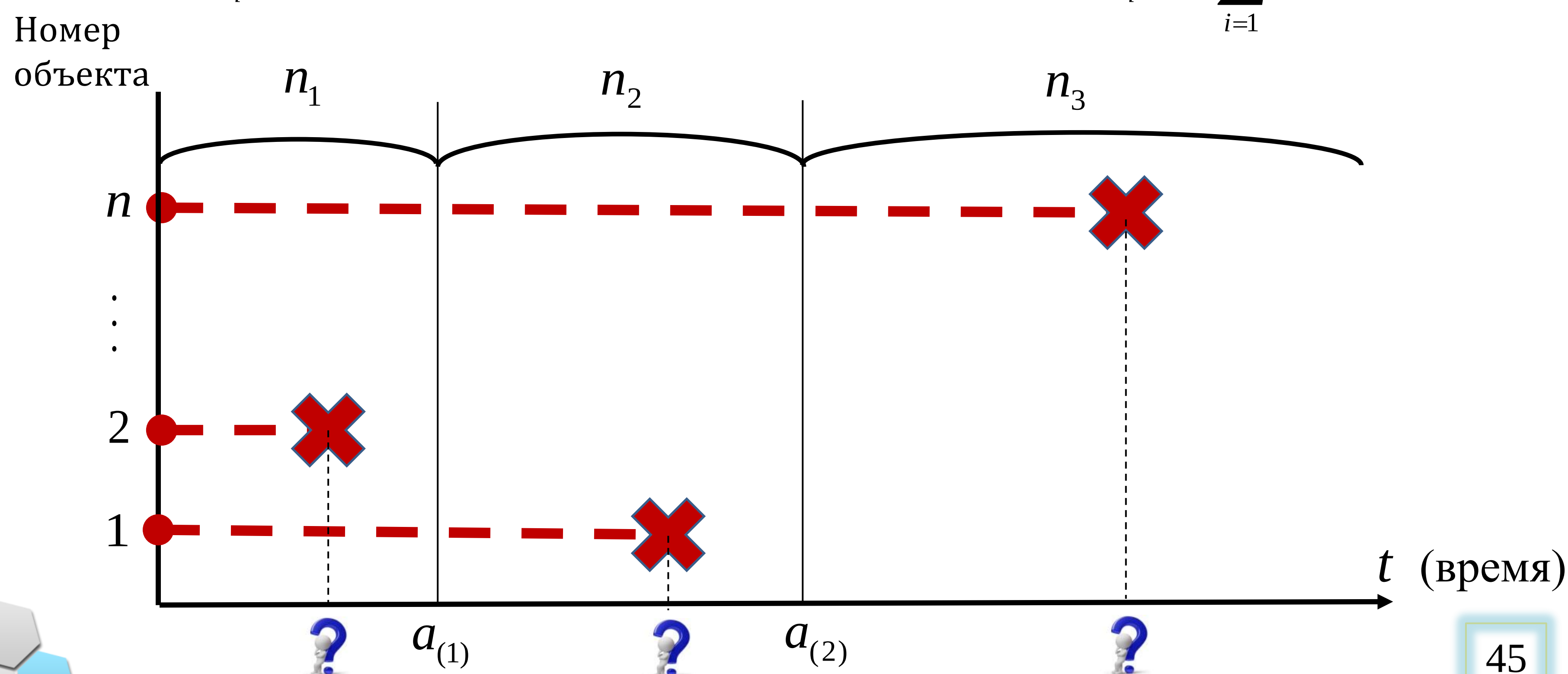
Группированная выборка

$$\mathbf{X}_n = \{(I_1, n_1), (I_2, n_2), \ldots, (I_k, n_k)\},$$

$I_1, I_2, \ldots, I_k$ – непересекающиеся интервалы с граничными точками:

$$0 = a_{(0)} < a_{(1)} < \ldots < a_{(k-1)} < a_{(k)} = +\infty,$$

n_i – количество отказов, попавших в интервал I_i, $\quad \sum_{i=1}^{k} n_i = n$



Проверка гипотезы о виде параметрической регрессионной модели по группированной выборке

$$H_0:\ S_x(t) \in \left\{ g\left(S_0(t;\theta); x; \beta\right),\ \theta, \beta \in \Omega \right\}$$

Статистика критерия хи-квадрат Пирсона:

$$X_n^2 = n \sum_{i=1}^{q} \sum_{j=1}^{k} \frac{\left(n_{ij} / n_i - P_{ij}\left(\hat{\eta}, \hat{\theta}\right) \right)^2}{P_{ij}\left(\hat{\eta}, \hat{\theta}\right)}$$

n_{ij} – число объектов, исследуемых при значении вектора ковариат x^i, $i = \overline{1,q}$;

$$n = \sum_{i=1}^{q} \sum_{j=1}^{k} n_{ij}$$

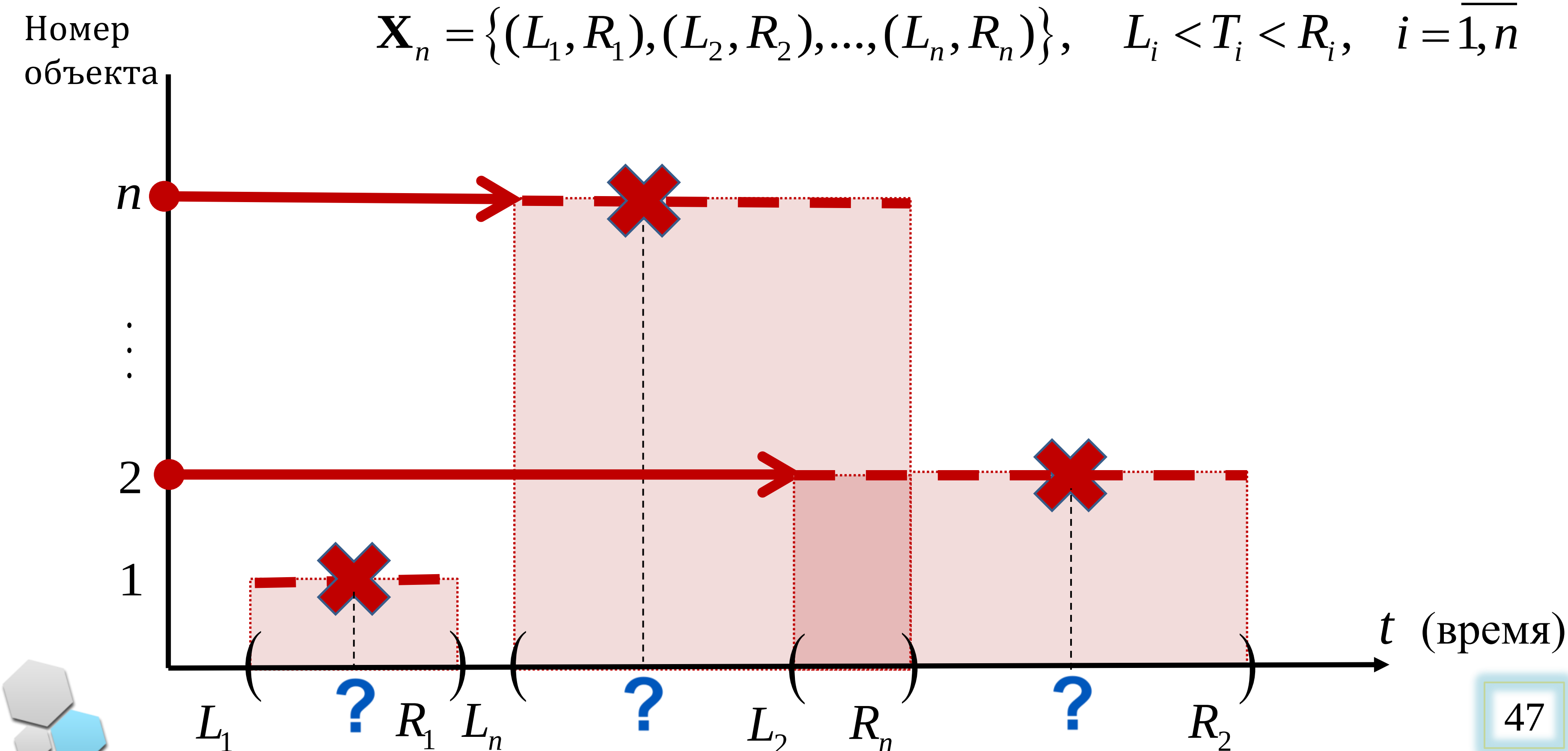
В пределе при $n \to \infty$ статистика X_n^2 имеет χ^2-распределение с числом степеней свободы $kq - (m+s) - 1$, где k – число интервалов группирования, q – число опорных точек плана эксперимента (количество различных значений вектора ковариат), m – размерность вектора регрессионных параметров, s – размерность вектора параметров базового распределения.

Проверка гипотезы о виде параметрической регрессионной модели по группированной выборке

$$H_0: \; S_x(t) \in \left\{ g\left(S_0(t;\theta); x; \beta\right), \; \theta, \beta \in \Omega \right\}$$

Группированная выборка отказов → Интервальная выборка остатков

$$\mathbf{X}_n = \left\{ (L_1, R_1), (L_2, R_2), \ldots, (L_n, R_n) \right\}, \quad L_i < T_i < R_i, \quad i = \overline{1, n}$$



Непараметрические критерии согласия для интервальных выборок

Модификации непараметрических критериев:

Колмогорова

$$D_n = \sup_{|t|<\tau} \left| \hat{F}(t) - F_0\left(t;\hat{\theta}\right) \right|$$

Крамера-Мизеса-Смирнова

$$\omega^2 = \int_0^\tau \left(\hat{F}(t) - F_0\left(t;\hat{\theta}\right) \right)^2 dF_0\left(t;\hat{\theta}\right)$$

Андерсона-Дарлинга

$$\Omega^2 = \int_0^\tau \left(\hat{F}(t) - F_0\left(t;\hat{\theta}\right) \right)^2 \frac{dF_0\left(t;\hat{\theta}\right)}{F_0\left(t;\hat{\theta}\right)\left(1 - F_0\left(t;\hat{\theta}\right)\right)}$$

Оценка Тернбулла*

Методами компьютерного моделирования проведено исследование скорости сходимости оценки Тернбулла к истинной функции распределения в зависимости от длины интервалов наблюдений

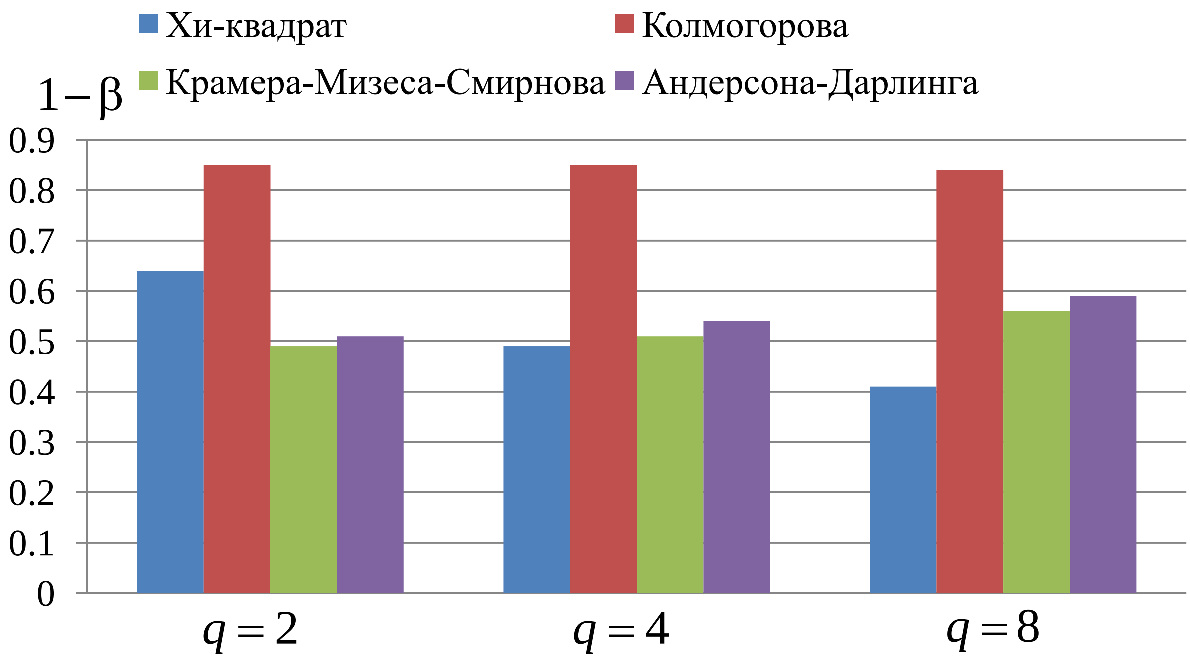
* Turnbull B.W. The empirical distribution function with arbitrarily grouped, censored and truncated data // J. R. Statist. Soc. Ser. B. – 1976. – Vol. 38. – P. 290-295.

Сравнительный анализ мощности критериев проверки гипотезы о виде модели по группированной выборке

Показано, что для рассмотренных пар конкурирующих гипотез критерий согласия типа Колмогорова на основе интервальной выборки остатков является наиболее предпочтительным по мощности.

H_0 : модель ускоренных испытаний, $F_0(t;\theta)$ – экспоненциальное, $r(x;\beta)=e^{\beta x}$

H_1 : модель ускоренных испытаний, $F_0(t;\theta)$ – гамма-распределение, $r(x;\beta)=e^{\beta x}$



$n=160$

$k=4$

$\alpha=0.1$

Оценки мощности критериев согласия при проверке сложной гипотезы

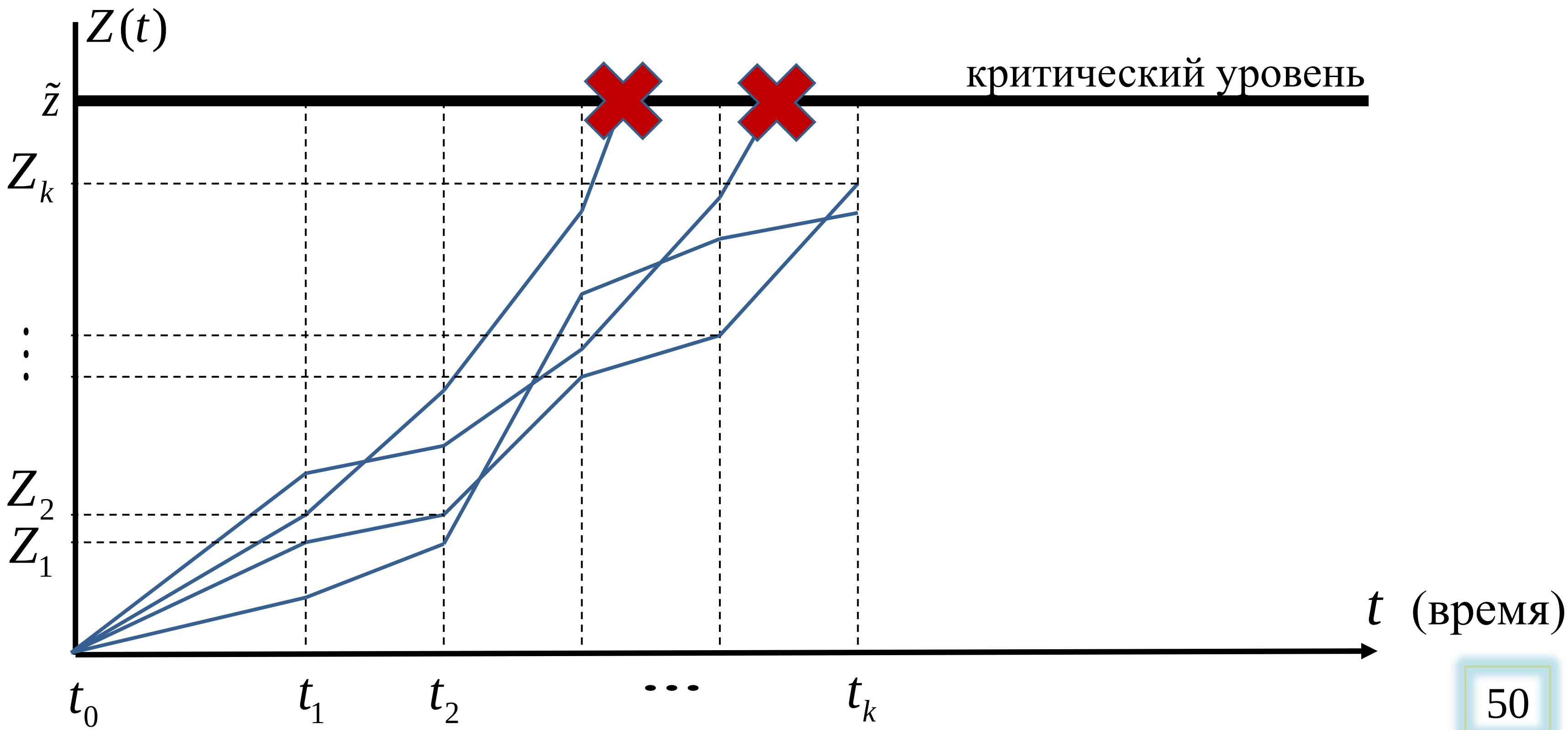
Данные о процессах деградации

Измерения показателя деградации: $Z^i = \left\{(0, Z_0^i), (t_1^i, Z_1^i), \ldots, (t_{k_i}^i, Z_{k_i}^i)\right\}, \quad i = \overline{1, n}$

Выборка приращений:

$$\mathbf{Z}_n = \left\{ \left(\left\{ X_j^1 = Z_j^1 - Z_{j-1}^1, \; j = \overline{1, k_1} \right\}, x^{(1)} \right), \ldots, \left(\left\{ X_j^n = Z_j^n - Z_{j-1}^n, \; j = \overline{1, k_n} \right\}, x^{(n)} \right) \right\}$$

$x^{(i)}$ – значение вектора ковариат для i-го объекта, $i = \overline{1, n}$



Деградационная гамма-модель

Пусть деградационный процесс $Z(t)$ удовлетворяет следующим условиям:

1) $Z(0)=0$;
2) $Z(t)$ является случайным процессом с независимыми приращениями;
3) $\Delta Z(t)=Z(t+\Delta t)-Z(t)\ \sim\ f_{Gamma}\left(t;\sigma,\Delta\nu(t)\right)=\left(\dfrac{t}{\sigma}\right)^{\Delta\nu(t)-1}\dfrac{e^{-t/\sigma}}{\sigma\cdot\Gamma\left(\Delta\nu(t)\right)}$,

где $\Delta\nu(t)=\nu(t+\Delta t)-\nu(t)$ – параметр формы и $\sigma>0$ – параметр масштаба.

Тогда функция надёжности принимает вид:

$$S_x(t)=P\{T>t\}=P\{Z(t)<\tilde{z}\}=F_{Gamma}\left(\tilde{z};\sigma,\frac{m_x(t;\beta,\gamma)}{\sigma}\right), \qquad (9)$$

где $\tilde{z}$ – критическое значение показателя деградации,

$m_x(t)=\sigma\nu\left(\dfrac{t}{r(x;\beta)};\gamma\right)$ – положительная, возрастающая функция тренда,

$r(x;\beta)$ – функция от ковариат.

Проверка гипотезы о виде деградационной гамма-модели

$$H_0 : F_{X_j^i}(t) \in \left\{ F_{Gamma}\left(t; \sigma, p_j^i(\beta,\gamma)\right), \sigma, \beta, \gamma \in \Omega \right\}, \ i = \overline{1,n}, \ j = \overline{1,k},$$

где $p_j^i = \dfrac{m_{x^{(i)}}\left(t_j^i; \gamma, \beta\right) - m_{x^{(i)}}\left(t_{j-1}^i; \gamma, \beta\right)}{\sigma}$ – параметр формы гамма-распределения.

Введём следующее преобразование приращений:

$$U_j^i = F_{Gamma}\left(X_j^i; \hat{\sigma}, \hat{p}_j^i\right), \ i = \overline{1,n}, \ j = \overline{1,k_i} \qquad (10)$$

При справедливости гипотезы $H_0 : U_j^i \sim \text{Uniform}(0,1), \ i = \overline{1,n}, \ j = \overline{1,k_i}$

Для проверки гипотезы о равномерном распределении можно использовать критерии согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

Исследование распределений статистик и мощности критериев согласия для деградационной гамма-модели

Показано, что распределения статистик критериев типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга при проверке сложной гипотезы о виде деградационной гамма-модели зависят от

- ✓ вида функции тренда,
- ✓ вида функции от ковариат,
- ✓ плана эксперимента (опорных точек плана и их количества),
- ✓ моментов времени замера показателя деградации.

В результате исследований мощности рассматриваемых критериев показано, что предложенный метод позволяет проверять предположения как о виде распределения приращений, так и о виде функции тренда и функции от ковариат.

Программная система статистического анализа данных типа времени жизни «LiTiS»

Анализ данных

- Описательные статистики
- Оценивание параметров моделей надежности
- Идентификация моделей
- Графики функций и визуализация данных
- Вычисление показателей надежности
- Мастер анализа

Проверка гипотез

- Гипотезы о виде распределения / модели
- Гипотезы об однородности распределений
- Гипотезы о параметрах моделей

Моделирование

- Распределений оценок параметров моделей
- Распределений статистик критериев согласия, критериев однородности

Разработано совместно с Румянцевым А.В., Семёновой М.А., Галановой Н.С. и Деминым В.А.

Правообладатель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Новосибирский государственный технических университет».

Свидетельства о государственной регистрации программы для ЭВМ № 2012618138, № 2012618143 и № 2014661905.

Статистический анализ результатов эксперимента на стойкость режущего инструмента

T - суммарная длина отверстий, просверленных инструментом до затупления, мм

x_1 - скорость подачи на оборот, мм/об

x_2 - частота вращения, об/мин

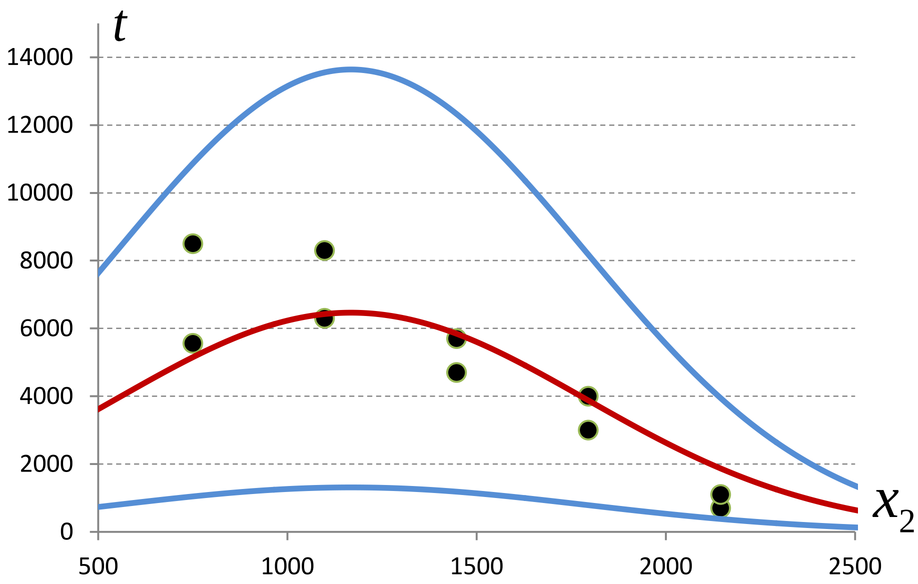
Стойкость T, мм

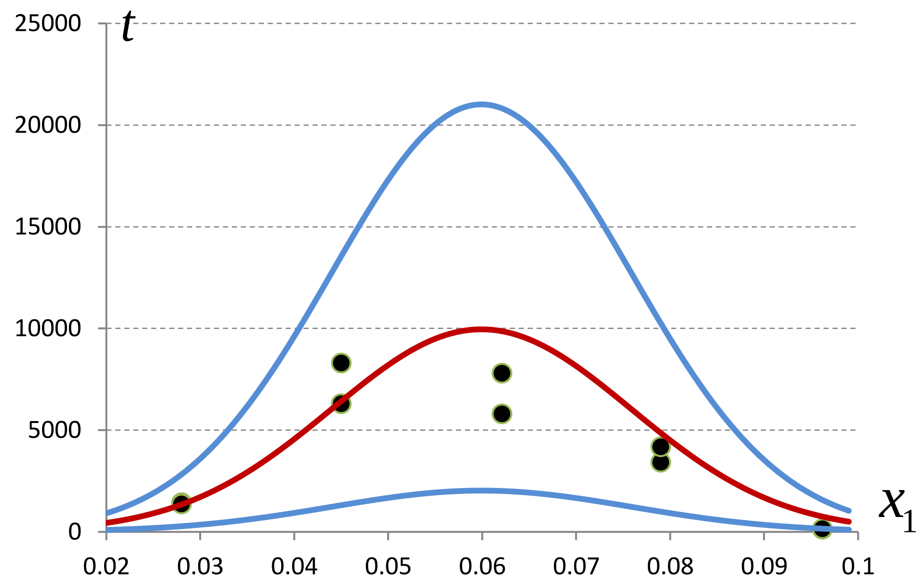
x_1, мм/об	x_2, об/мин				
	750	1098	1447	1795	2145
0.0280	570 390	1430 1370	**3600** **1800**	1400 1200	430 250
0.0450	5560 8500	**8300** **6300**	4700 5700	**4000** **3000**	700 1100
0.0621	**4040** **5640**	5800 7800	**6130** **4230**	3330 4070	**590** **810**
0.0790	3150 3850	**3420** **4180**	2760 3800	**1350** **1650**	470 690
0.0962	1910 3170	130 150	**100** **140**	30 50	9 11

Параметрическая регрессионная модель стойкости режущего инструмента

$$S_x(t;\beta,\theta) = P\{T \ge t\} = \exp\left\{-\left(\frac{t}{\theta_1 \cdot r(x;\beta)}\right)^{\theta_2}\right\}, \qquad \theta_1, \theta_2 > 0$$

$$r(x;\beta) = \exp\{\beta_1 x_1^2 + \beta_2 x_1 + \beta_3 x_2^2 + \beta_4 x_2 + \beta_5 x_1 x_2\} \qquad (11)$$





Гипотеза о виде модели **принимается** по критериям типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга, примененным к выборке остатков, $\alpha_n > \alpha = 0.05$

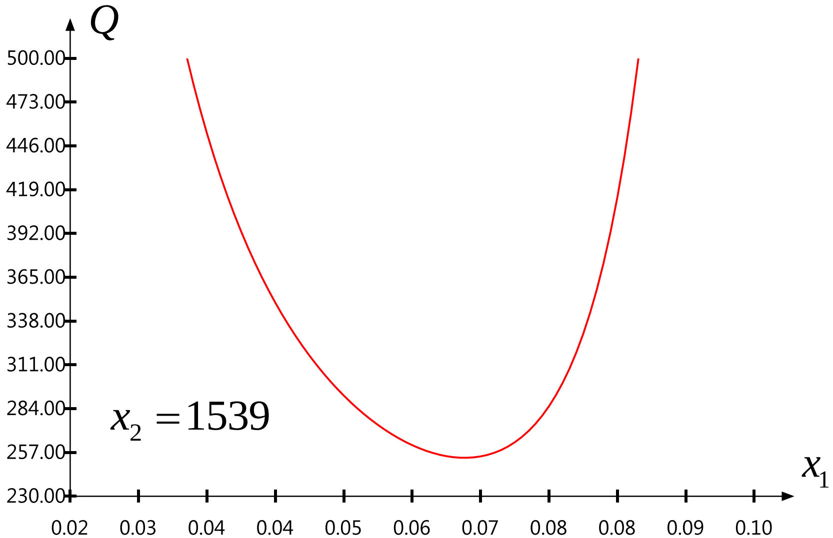
Оптимизация режимов обработки

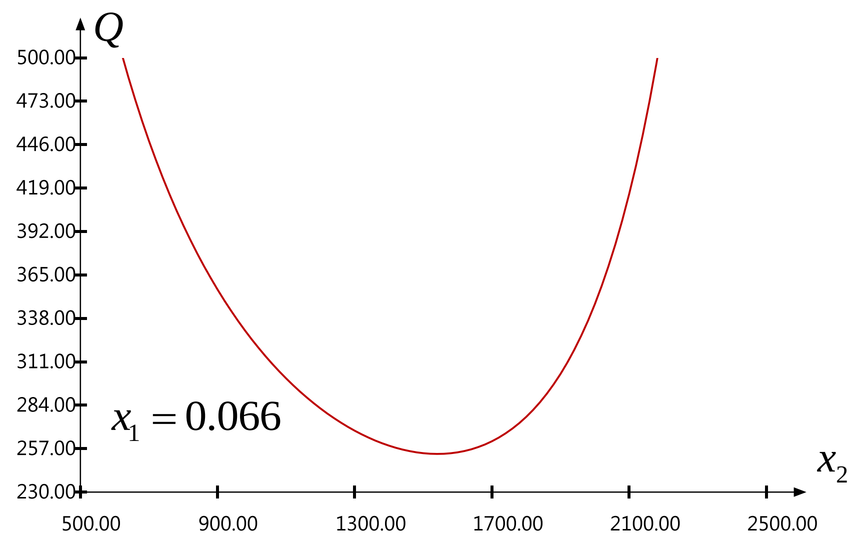
$$Q(x_1,x_2)=\frac{C}{x_1x_2}+\frac{D}{\bar{T}\left(x_1,x_2\right)}\to\min_{x_1,x_2}$$

$$\bar{T}\left(x_1,x_2\right)=\theta_1\exp\left\{\beta_1x_1^2+\beta_2x_1+\beta_3x_2^2+\beta_4x_2+\beta_5x_1x_2\right\}\cdot\Gamma\left(1+\frac{1}{\theta_2}\right)$$

$\Gamma(\cdot)$ – гамма-функция Эйлера, *C*, *D* – величины, зависящие от выбранного критерия и характера учитываемых затрат♠

$$x_1^*=0.066,\ x_2^*=1539,\ Q^*=253.715$$





♠ Карманов В.С. Исследование математических моделей стойкости режущего инструмента // Научный вестник НГТУ. – 2006. – № 2(23). – С. 55-64.

Статистический анализ выживаемости больных раком легкого после операции

В ФГБУ «НИИ онкологии им. Н.Н. Петрова» Минзрава РФ прослежены пациенты, оперированные по поводу немелкоклеточного рака легкого в возрасте от 70 до 84 лет.

T – продолжительность жизни пациента от операции до смерти от рака, в месяцах.

x_1 – категория опухоли (связана с размером и локализацией опухоли)

x_2 – состояние регионарных лимфатических узлов (связана с локализацией и количеством пораженных метастазами лимфатических узлов)

x_3 – наличие сахарного диабета

Получена случайно цензурированная выборка,
объем выборки $n = 613$, из них 169 цензурированных наблюдений.

Параметрическая регрессионная модель выживаемости больных раком легкого после операции

$$\Lambda_x(t;\beta,\theta)=\exp(\beta_1 x_1+\beta_2 x_2+\beta_3 x_3)\cdot\Lambda_0(t;\theta),$$

где $\Lambda_x(t)=-\ln(S_x(t))$ – кумулятивная функция риска.

В качестве базовой функции риска $\Lambda_0(t;\theta)$ выбрано логнормальное распределение

Гипотеза о виде модели **принимается** по критериям согласия типа Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга и типа χ^2 Никулина-Рао-Робсона, примененным к выборке остатков, $\alpha_n>\alpha=0.05$

Для проверки гипотезы использовались разработанные алгоритмы:

- ✓ моделирования случайно цензурированной выборки,
- ✓ оптимального группирования цензурированной выборки.

Заключение

В соответствии с поставленными целями получены следующие основные результаты:

1. Методами статистического моделирования исследованы свойства ОМП параметров распределений по цензурированным справа выборкам и выборкам усеченных слева наблюдений. Показано, что при ограниченных объемах выборок распределения ОМП с ростом степени цензурирования и степени усечения существенно отклоняются от асимптотического многомерного нормального закона.

2. Для ряда законов распределения, наиболее часто используемых в задачах теории надежности, получены оптимальные планы испытаний изделий одноразового срабатывания на надежность с позиции точности ОМП параметров распределений, получаемых по выборкам текущих состояний.

3. На основании результатов исследования распределений статистик и мощности критериев согласия типа хи-квадрат в ситуациях полных и цензурированных выборок сформулированы рекомендации по применению рассмотренных критериев при проверке простых и сложных гипотез.

4. Предложен алгоритм оптимального группирования цензурированной выборки, позволяющий существенно увеличить мощность критериев согласия типа χ^2 относительно конкретной пары конкурирующих гипотез.

Заключение

5. Построены таблицы верхних процентных точек критериев типа Колмогорова, Андерсона-Дарлинга, Крамера-Мизеса-Смирнова для цензурированных I и II типа выборок для проверки простых гипотез. Для случая проверки сложных гипотез аналогичные таблицы построены относительно законов Вейбулла, экспоненциального и логнормального.

6. Разработан алгоритм моделирования случайно цензурированных выборок, обеспечивающий возможность применения модифицированных критериев согласия типа Колмогорова, Андерсона-Дарлинга, Крамера-Мизеса-Смирнова в условиях неизвестного распределения моментов цензурирования.

7. Разработана методика проверки простых и сложных гипотез о виде распределения по цензурированным данным с использованием классических критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга, базирующаяся на преобразовании исходной цензурированной выборки в псевдополную.

8. Показано, что для большинства рассмотренных случаев проверки сложных гипотез по цензурированным выборкам наиболее предпочтительным по мощности является модифицированный критерий типа Андерсона-Дарлинга.

Заключение

9. Предложена методика проверки гипотез о виде распределения по выборкам усеченных слева наблюдений с использованием непараметрических критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

10. Предложены новые статистические критерии согласия для проверки гипотез по выборкам текущих состояний. Применение критериев предполагает использование распределений статистик, получаемых методами статистического моделирования, в том числе в ходе проводимого анализа (в интерактивном режиме).

11. Разработана методика проверки гипотезы о виде параметрической регрессионной модели надежности на основе выборок остатков. Предложены алгоритмы моделирования, обеспечивающие нахождение распределения статистики, соответствующего справедливости проверяемой гипотезы и необходимого для оценки достигнутого уровня значимости по применяемому критерию.

12. Для проверки гипотезы о виде параметрической регрессионной модели по группированным выборкам с ковариатами и анализа получающихся интервальных выборок остатков предложены модификации критериев Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

Заключение

13. В результате исследования мощности критериев согласия, применяемых к выборкам остатков, при проверке сложных гипотез о виде параметрической регрессионной модели показано следующее. В случае данных без усечения, как правило, наибольшей мощностью среди рассматриваемых обладает критерий типа Андерсона-Дарлинга. При наличии усеченных слева наблюдений он уже уступает в мощности критериям типа Колмогорова и Крамера-Мизеса-Смирнова.

14. В случае проверки гипотезы о виде параметрической регрессионной модели по выборкам текущих состояний показано, что для ряда пар конкурирующих гипотез наибольшую мощность демонстрирует критерий типа Уайта, в то время как для других пар гипотез более мощными оказываются критерии типа Колмогорова и Крамера-Мизеса-Смирнова. Критерий типа хи-квадрат во всех рассмотренных ситуациях уступил по мощности упомянутым критериям.

15. Предложена методика проверки сложной гипотезы о виде деградационной гамма-модели с использованием непараметрических критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга.

16. Разработана программная система статистического анализа данных типа времени жизни «LiTiS».