

ЛЕКЦИЯ 6

ЦЕНЗУРИРОВАНИЕ

Лекция 6

Пример 1

На испытания поставлены *7 экземпляров* однотипного оборудования.

Поставлена задача найти точечную оценку параметра распределения, считая, что времена отказов подчиняются экспоненциальному распределению.

В ходе испытаний были зафиксированы отказы трех экземпляров оборудования с временами **1, 2 и 5** единиц времени.

По прошествии **6** единиц времени было принято решение прекратить испытания.

Пример 1 (продолжение)

Предположим, было принято решение проводить процедуру оценивания, основываясь только на наблюдаемых значениях времени отказа, считая, что количество объектов равно 3.

Используем МНК.

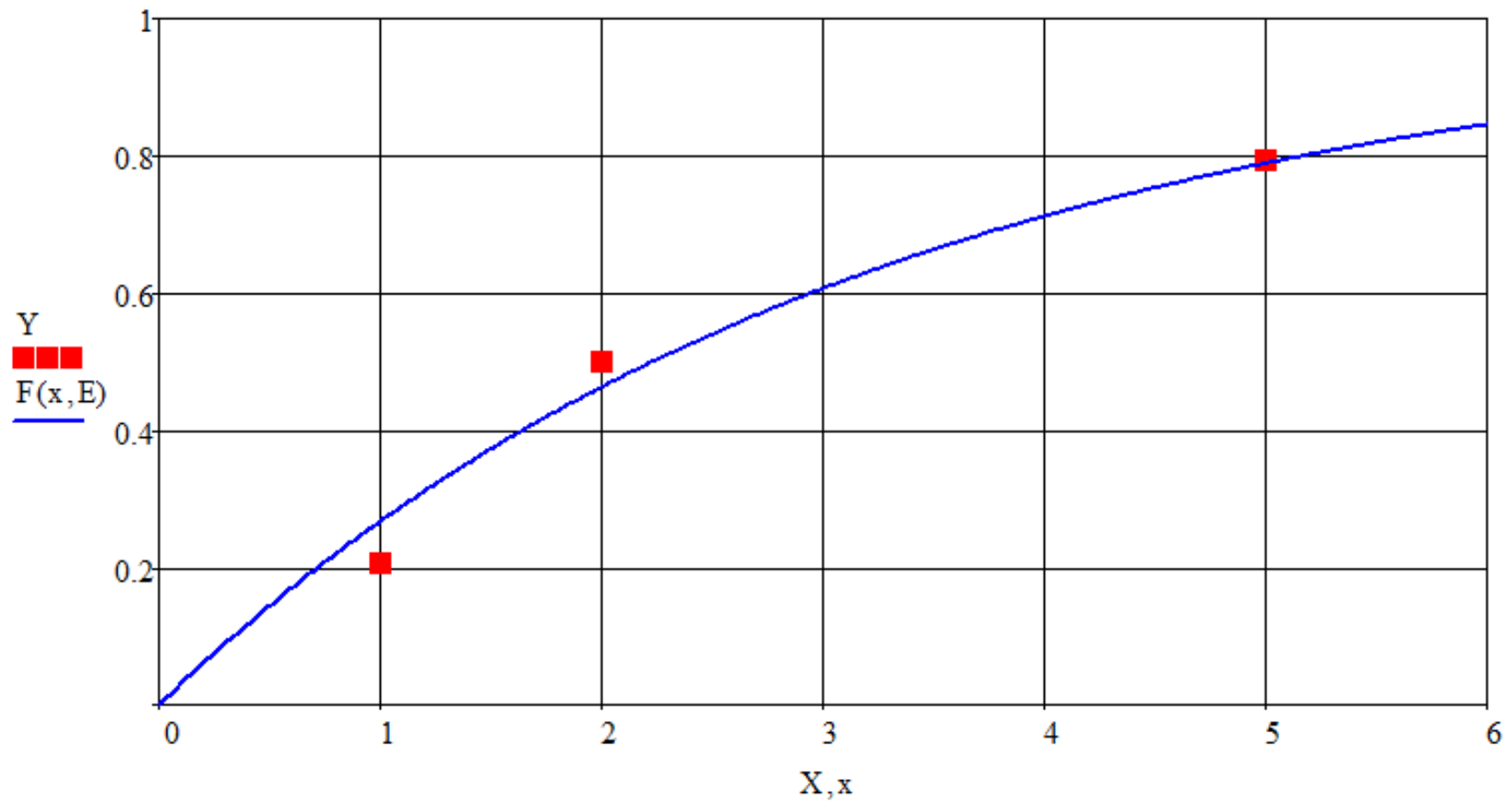
$$X = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \quad Y = \begin{bmatrix} 0,206 \\ 0,5 \\ 0,794 \end{bmatrix}$$

$$RSS(\lambda) = \sum_{i=1}^3 [Y_i - F(X_i, \lambda)]^2$$

$$E_{MНК} = 0,311$$

$$T_{MНК} = 3,215$$

Пример 1 (продолжение)



Пример 1 (продолжение)

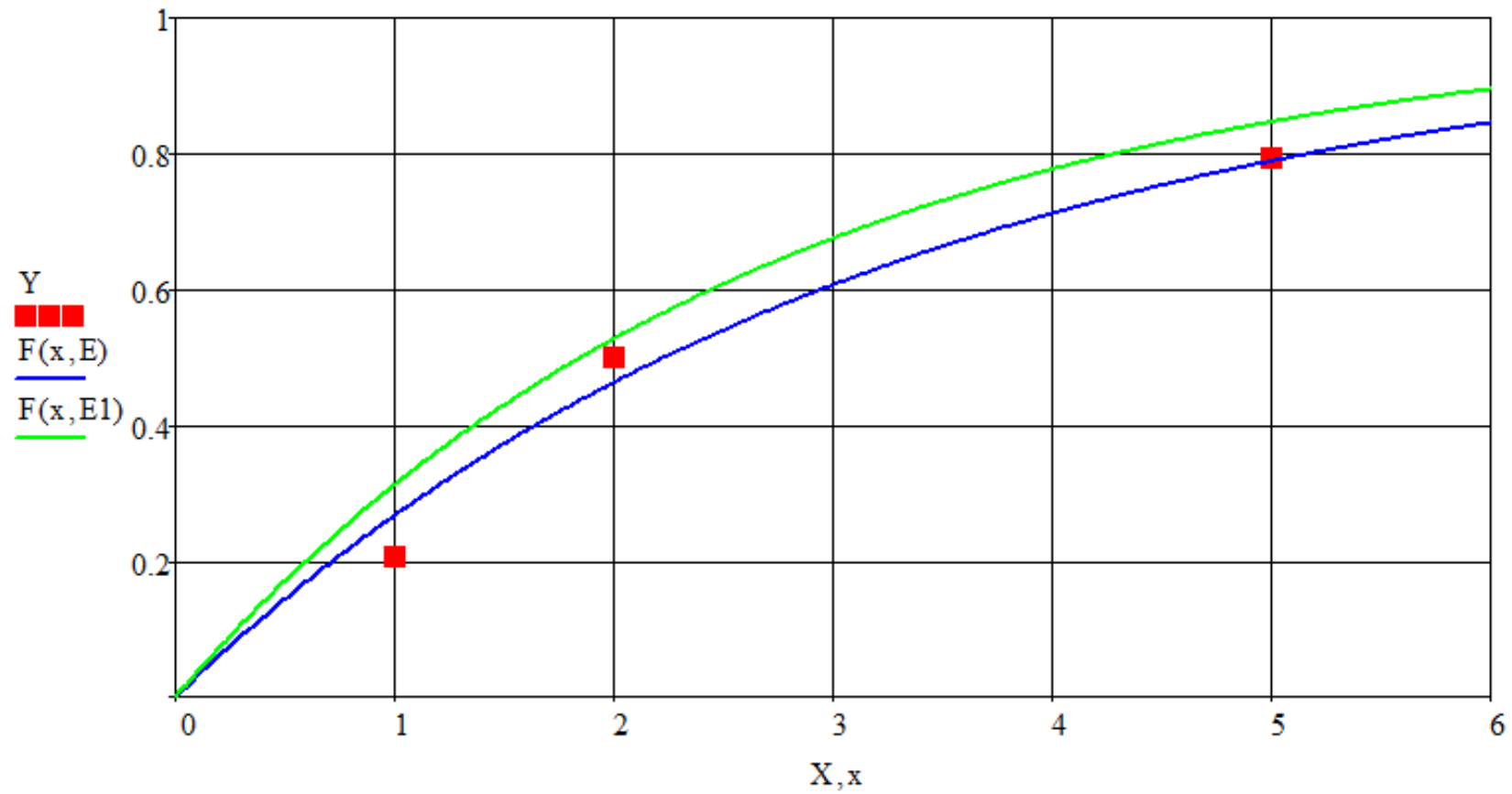
Используем ММП.

$$X = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \quad \Lambda(\lambda) = \sum_{i=1}^3 \ln f(X_i, \lambda)$$

$$E_{\text{ММП}} = 0,375$$

$$T_{\text{ММП}} = 2,667$$

Пример 1 (продолжение)



Пример 1 (продолжение)

Однако, тот факт, что некоторые экземпляры оборудования не отказали к моменту прекращения наблюдения, представляет собой информацию.

Отказ от использования этой информации означает искажение результатов оценивания.

Предположим теперь, что для МНК используются значения Y , рассчитанные, исходя из того, что общее количество экземпляров оборудования равно 7.

Пример 1 (продолжение)

МНК

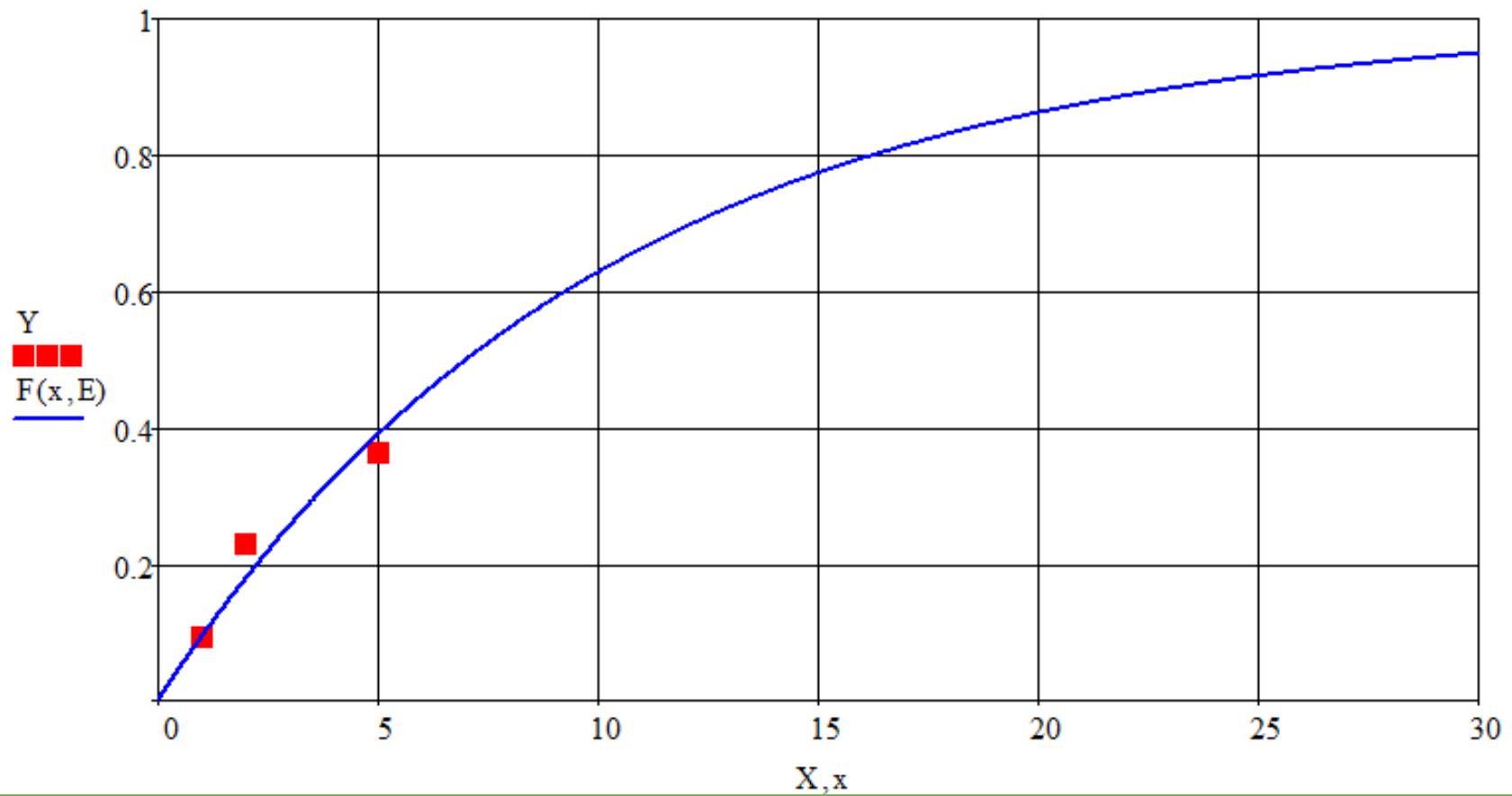
$$X = \begin{bmatrix} 1 \\ 2 \\ 5 \\ ? \\ ? \\ ? \\ ? \end{bmatrix} \quad Y = \begin{bmatrix} 0,094 \\ 0,228 \\ 0,364 \\ 0,5 \\ 0,636 \\ 0,772 \\ 0,906 \end{bmatrix}$$

$$RSS(\lambda) = \sum_{i=1}^3 [Y_i - F(X_i, \lambda)]^2$$

$$E_{МНК} = 0,099$$

$$T_{МНК} = 10,099$$

Пример 1 (продолжение)



Пример 1 (продолжение)

Видно, что учет даже части из имеющейся информации существенно изменяет результат оценки параметра.

ММП позволяет учесть больше информации.

Обозначим количество наблюдаемых отказов через $N_F = 3$, а число экземпляров оборудования, которые не отказали к моменту прекращения наблюдения через $N_R = 4$.

Также, обозначим через $T_R = 6$ момент окончания наблюдения.

Пример 1 (продолжение)

Имеющиеся данные можно представить в следующем виде:

1	<i>F</i>	
2	<i>F</i>	отказы
5	<i>F</i>	
6	<i>R</i>	
6	<i>R</i>	прекращение наблюдений
6	<i>R</i>	
6	<i>R</i>	

Пример 1 (продолжение)

Информацию о том, что некоторый экземпляр оборудования **не отказал** к моменту времени T_{sus} можно учесть через вероятность этого события:

$$\Pr[X \geq T_R] = 1 - F(T_R, \lambda)$$

Для использования этой информации необходимо модифицировать логарифмическую функцию правдоподобия следующим образом:

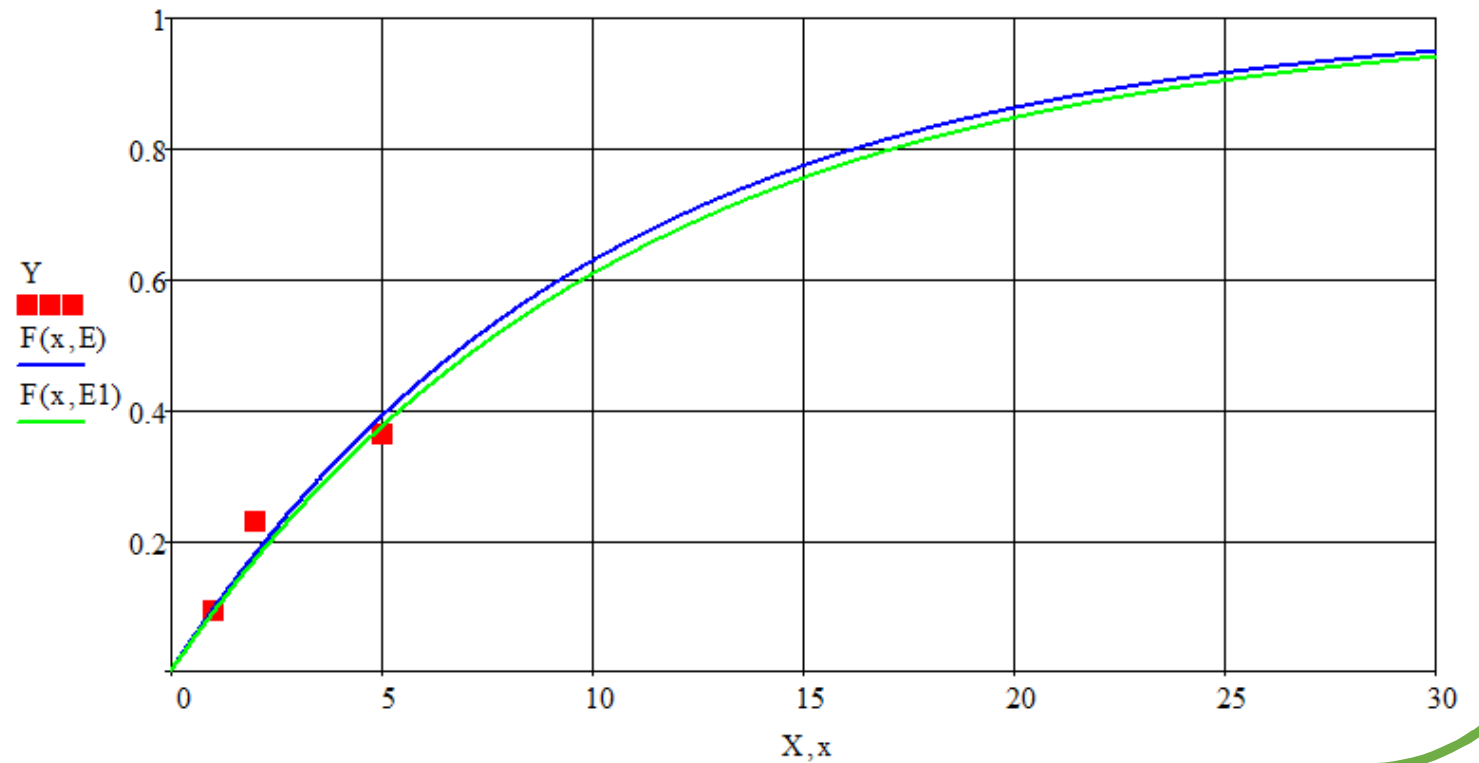
$$\Lambda(\lambda) = \sum_{i=1}^{N_F=3} \ln f(X_i, \lambda) + N_R \cdot \ln[1 - F(T_R, \lambda)]$$

Пример 1 (продолжение)

Далее определяем максимум модифицированной функции.

$$E_{\text{ММП}} = 0,094$$

$$T_{\text{ММП}} = 10,667$$



Лекция 6

В общем случае, когда за различными экземплярами оборудования прекращают наблюдать в разное время T_{Rj} , логарифмическая функция правдоподобия принимает вид:

$$\Lambda(\Theta) = \sum_{i=1}^{N_F} \ln f(X_i, \Theta) + \sum_{j=1}^{N_R} \ln [1 - F(T_{Rj}, \Theta)]$$

где Θ – вектор оцениваемых параметров, X_i – времена наблюдаемых отказов, T_{Rj} – времена прекращения наблюдения.

Лекция 6

В процессе анализа надежности приходится сталкиваться с ситуациями, когда определенная часть объектов или систем не отказывает за период наблюдения,

а другая часть отказывает, но моменты отказов точно неизвестны.

Лекция 6

В таких ситуациях возникает необходимость проведения статистического анализа надежности на основе специфических выборок, основной особенностью которых является отсутствие сведений о моментах отказов контролируемой части изделий.

Это явление носит название *цензурирование данных*, а получаемые в результате выборки — *цензурированными выборками*.

Лекция 6

Под данными, применительно к задачам надежности, понимают фиксированные значения наработок изделий, полученные по результатам испытаний или эксплуатационных наблюдений.

Данными цензурированной выборки являются как наработки отказавших объектов, так и наработки неотказавших объектов, а также могут быть интервалы времени, в течение которых объект отказал, но момент отказа точно неизвестен.

Прежде чем перейти к описанию структуры представления данных цензурированных выборок, приведем некоторые определения.

Лекция 6

Событие, приводящее к прекращению испытаний или эксплуатационных наблюдений объекта до наступления отказа (предельного состояния) изучаемого характера, либо к обнаружению отказа изучаемого характера в пределах известного интервала наработки, называется *цензурированием*.

Цензурированной выборкой называется выборка, элементами которой являются значения наработки до отказа и наработки до цензурирования, либо только значения наработки до цензурирования.

Лекция 6

Как было отмечено ранее, цензурирование — это процесс возникновения неопределенности момента отказа объекта, причем интервал неопределенности известен аналитику.

Интервал неопределенности — интервал наработки, внутри которого произошел либо произойдет отказ объекта, причем точное значение наработки до отказа неизвестно.

Этот интервал может быть неограниченным справа, тогда говорят о *цензурировании справа*, либо ограниченным справа, тогда говорят о *цензурировании слева*. Если интервал неопределенности момента отказа ограничен слева и справа, то говорят о *цензурировании интервалом*.

Пример 2

На испытания поставлены *7 экземпляров* однотипного оборудования.

Поставлена задача найти точечную оценку параметра распределения, считая, что времена отказов подчиняются экспоненциальному распределению.

Непосредственно, наблюдения начались только через 6 единиц времени, при этом 3 экземпляра оборудования уже отказали. В ходе испытаний были зафиксированы отказы четырех экземпляров оборудования с временами **7, 8, 18** и **29** единиц времени.

Пример 2 (продолжение)

МНК

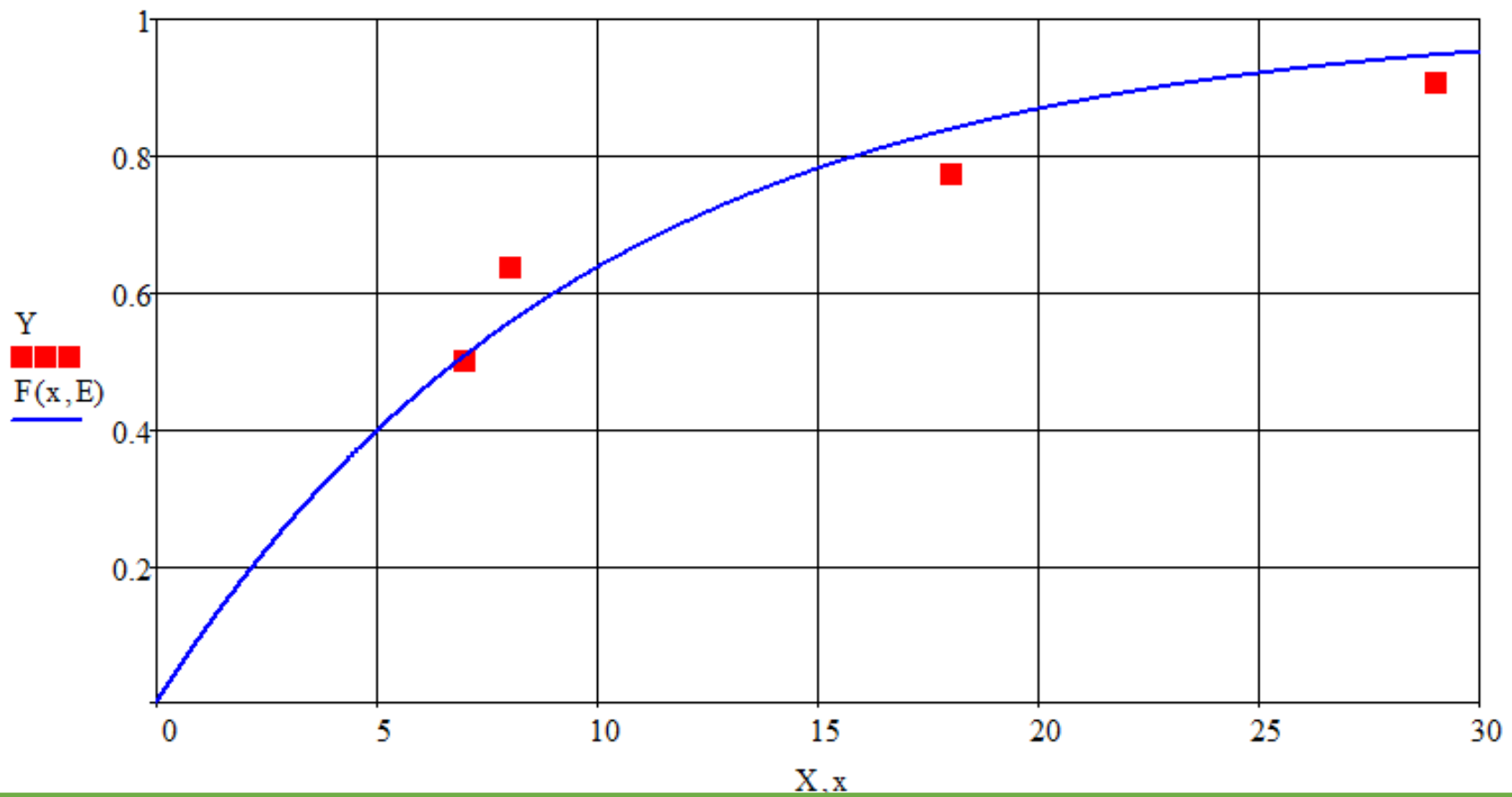
$$X = \begin{bmatrix} ? \\ ? \\ ? \\ 7 \\ 8 \\ 18 \\ 29 \end{bmatrix} \quad Y = \begin{bmatrix} 0,094 \\ 0,228 \\ 0,364 \\ 0,5 \\ 0,636 \\ 0,772 \\ 0,906 \end{bmatrix}$$

$$RSS(\lambda) = \sum_{i=4}^7 [Y_i - F(X_i, \lambda)]^2$$

$$E_{МНК} = 0,101$$

$$T_{МНК} = 9,869$$

Пример 2 (продолжение)



Пример 2 (продолжение)

ММП

Обозначим количество наблюдаемых отказов через $N_F = 4$, а число экземпляров оборудования, которые отказали к моменту начала наблюдения через $N_L = 3$.

Также, обозначим через $T_L = 6$ момент начала наблюдения.

Пример 2 (продолжение)

Имеющиеся данные можно представить в следующем виде:

6	<i>L</i>	
6	<i>L</i>	отказы до начала наблюдений
6	<i>L</i>	
7	<i>F</i>	
8	<i>F</i>	
18	<i>F</i>	наблюдаемые отказы
29	<i>F</i>	

Пример 2 (продолжение)

Информацию о том, что некоторый экземпляр оборудования отказал к моменту времени T_L можно учесть через вероятность этого события:

$$\Pr[X < T_L] = F(T_L, \lambda)$$

Для использования этой информации необходимо модифицировать логарифмическую функцию правдоподобия следующим образом:

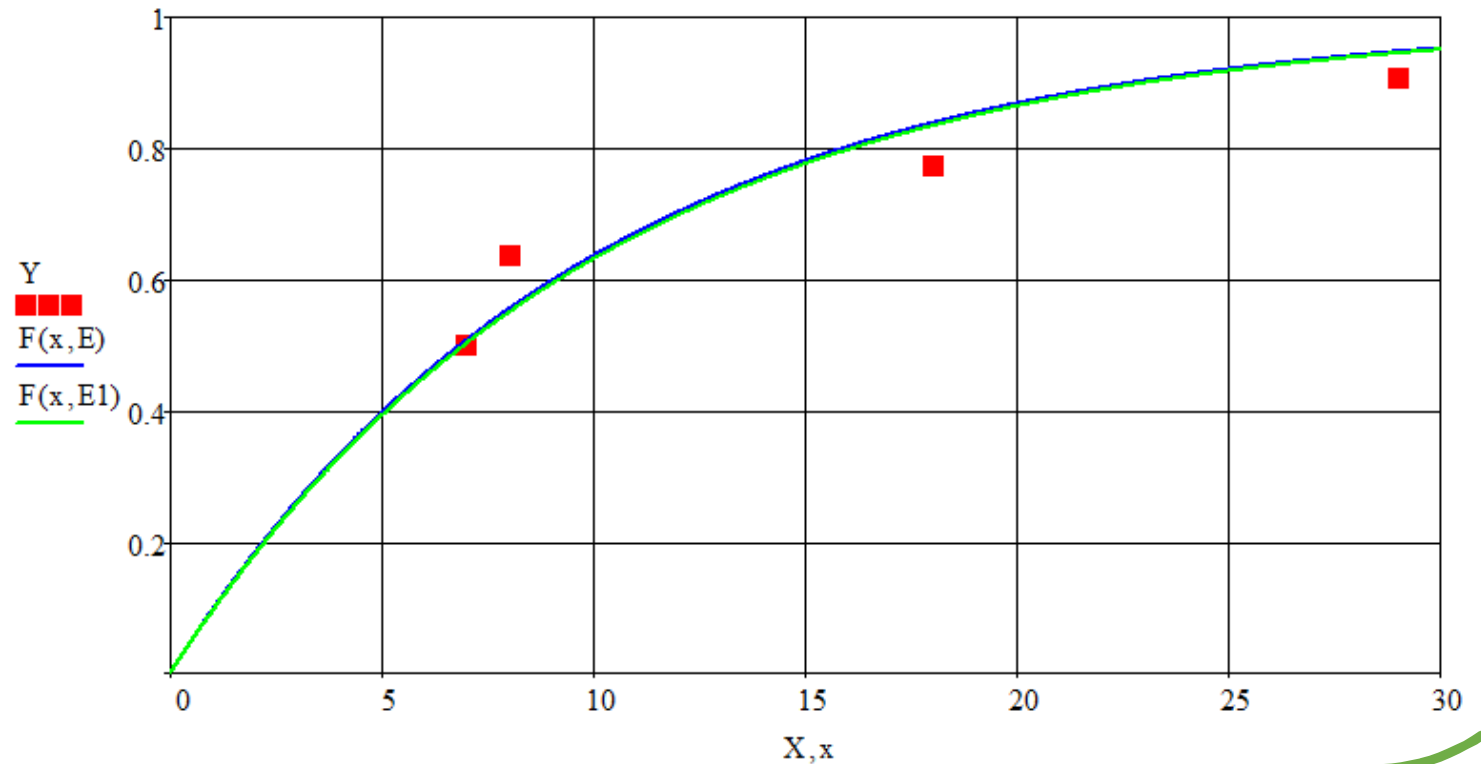
$$\Lambda(\lambda) = \sum_{i=1}^{N_F=4} \ln f(X_i, \lambda) + N_L \cdot \ln[F(T_L, \lambda)]$$

Пример 2 (продолжение)

Далее определяем максимум модифицированной функции.

$$E_{\text{ММП}} = 0,0998$$

$$T_{\text{ММП}} = 10,015$$



Лекция 6

В общем случае, когда за различными экземплярами оборудования начинают наблюдать в разное время T_{Lj} , логарифмическая функция правдоподобия принимает вид:

$$\Lambda(\Theta) = \sum_{i=1}^{N_F} \ln f(X_i, \Theta) + \sum_{j=1}^{N_L} \ln [F(T_{Lj}, \Theta)]$$

где Θ – вектор оцениваемых параметров, X_i – времена наблюдаемых отказов, T_{Lj} – времена начала наблюдения за j -ым объектом.

Пример 3

7 экземпляров однотипного оборудования поставлены на испытания на двух промышленных площадках (3 и 4).

В ходе испытаний на первой площадке зафиксированы времена отказов: 2, 5 и 18.

На второй площадке наблюдения проводились через интервалы в 10 единиц времени. Определено, что в интервале 0-10 произошло 3 отказа, 10-20 – 0 отказов, 20-30 – 1 отказ.

Поставлена задача найти точечную оценку параметра распределения, считая, что времена отказов подчиняются экспоненциальному распределению.

Пример 3 (продолжение)

Имеющиеся данные можно представить в следующем виде:

2		F	
5		F	наблюдаемые отказы
18		F	
0-10	3		
10-20	0		интервальные данные
20-30	1		

Пример 3 (продолжение)

Информацию о том, что некоторый экземпляр оборудования отказал на интервале $(a; b]$ можно учесть через вероятность этого события:

$$\Pr[a < X \leq b] = F(b, \lambda) - F(a, \lambda)$$

Для использования этой информации необходимо модифицировать логарифмическую функцию правдоподобия следующим образом:

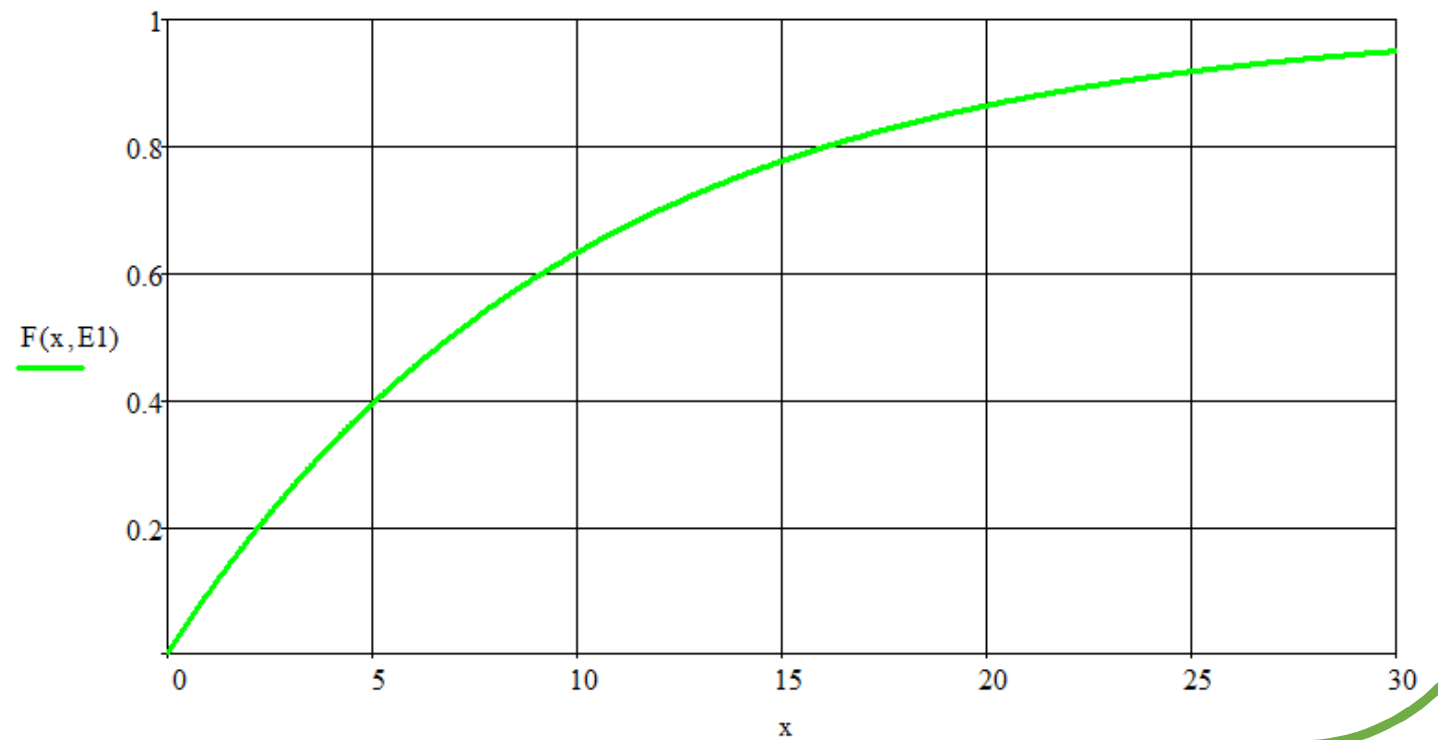
$$\Lambda(\lambda) = \sum_{i=1}^{N_F=3} \ln f(X_i, \lambda) + \sum_{j=1}^{N_I=4} \ln [F(b_j, \lambda) - F(a_j, \lambda)]$$

Пример 3 (продолжение)

Далее определяем максимум модифицированной функции.

$$E_{\text{ММП}} = 0,114$$

$$T_{\text{ММП}} = 8,753$$



Лекция 6

Общий вид модифицированной логарифмической функции правдоподобия, учитывающей правое, левое и интервальное цензурирование имеет вид:

$$\Lambda(\lambda) = \sum_{i=1}^{N_F} \ln f(X_i, \lambda) + \sum_{j=1}^{N_R} \ln [1 - F(T_{Rj}, \Theta)] + \sum_{k=1}^{N_L} \ln [F(T_{Lk}, \Theta)] + \sum_{n=1}^{N_I} \ln [F(b_n, \lambda) - F(a_n, \lambda)]$$

наблюдаемые
отказы

правое
цензурирование

левое
цензурирование

интервальное
цензурирование

Лекция 6

Необходимо обратить внимание на то, что интервальное цензурирование является наиболее общим видом цензурирования, так как при устремлении правой границы интервала к бесконечности этот вид цензурирования превращается в цензурирование **справа**, а при устремлении левой границы интервала к нулю — в цензурирование **слева**.

$$\Lambda(\lambda) = \sum_{i=1}^{N_F} \ln f(X_i, \lambda) + \sum_{j=1}^{N_I} \ln [F(b_j, \lambda) - F(a_j, \lambda)]$$

Правое: $F(\infty, \lambda) - F(a_j, \lambda) = 1 - F(a_j, \lambda)$

Левое: $F(b_j, \lambda) - F(0, \lambda) = F(b_j, \lambda) - 0$

Лекция 6

Так называемое *цензурирование типа I* применяется в ситуациях, когда тестирование заканчивается в определенный заранее момент времени (например, мы начинаем проверять 100 ламп и заканчиваем эксперимент после некоторого фиксированного заранее момента времени).

В этом случае время цензурирования фиксировано, и число отказавших единиц (в примере, ламп) является случайной величиной.

Лекция 6

При *цензурировании типа II* эксперимент продолжается до того момента, когда откажет (выйдет из строя) фиксированное заранее число проверяемых единиц (например, эксперимент заканчивается только после выхода из строя 50 ламп).

В этом случае число отказавших элементов фиксировано, а время является случайной величиной.

Лекция 6

При оценке параметров статистических моделей надежности с использованием ММП по цензурированным (I и II тип) выборкам логарифмическая функция правдоподобия должна быть изменена следующим образом:

$$\Lambda(\Theta) = \sum_{i=1}^N \ln f(T_i, \Theta) + M \cdot \ln[1 - F(T_{\text{ценз}}, \Theta)],$$

где N – число наблюдаемых отказов; $N + M$ – объем выборки; T_i – времена наблюдаемых отказов; $T_{\text{ценз}}$ – время цензурирования; Θ – оцениваемый параметр (вектор параметров).

Для цензурирования типа I $T_{\text{ценз}}$ задается заранее.

Для цензурирования типа II $T_{\text{ценз}} = \max T_i$.

Лекция 6

Цензурирование в совокупности с ММП предоставляет еще одну возможность анализа.

Пример 4

Пусть имеется система, которая может отказать по двум причинам: А и В. Необходимо по эксплуатационным данным оценить параметры распределения Вейбулла для отказов по каждой причине.

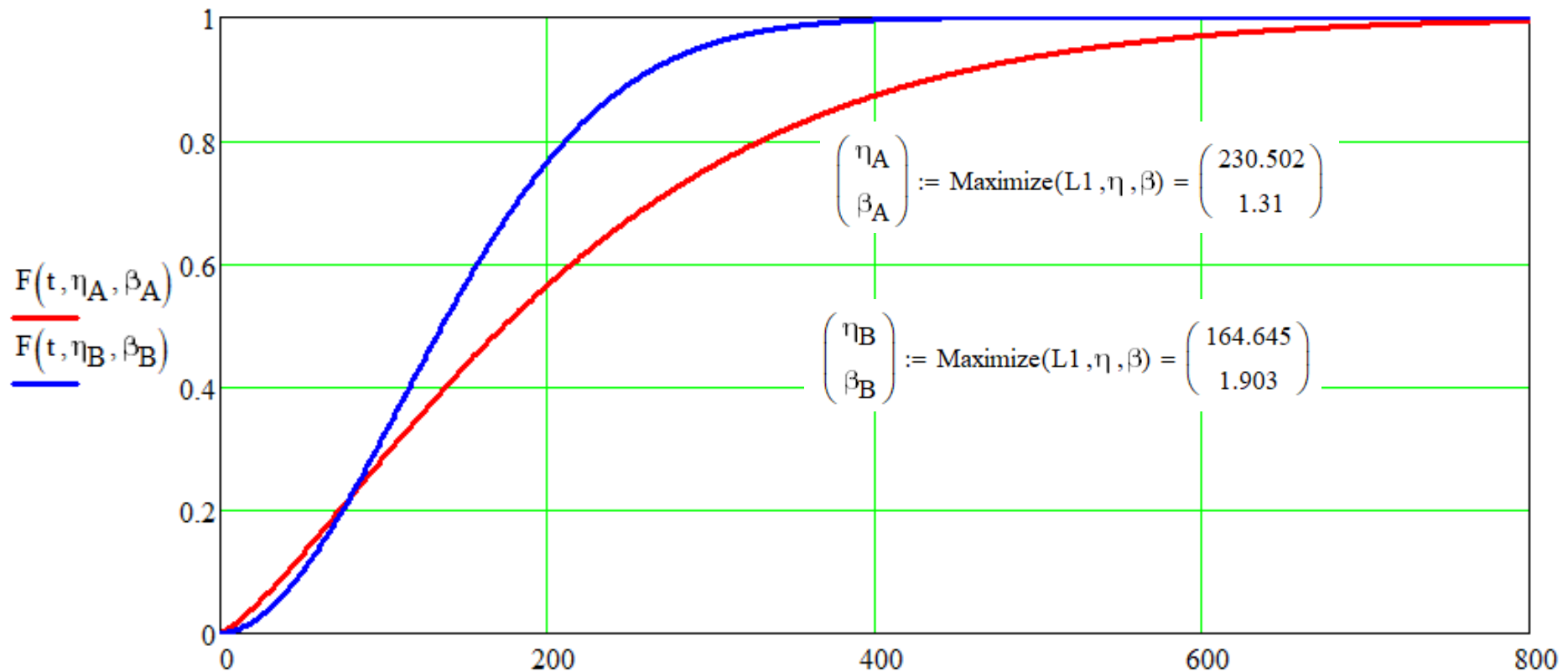
Предположим, имеется выборка из $N = 20$ времен отказов оборудования, причем каждому отказу соответствует одна из причин.

Пример 4 (продолжение)

Время	Причина	Время	Причина	Время	Причина	Время	Причина
8	A	55	B	93	B	167	B
40	B	57	B	114	B	169	B
41	A	65	B	130	B	198	B
42	A	71	A	135	B	226	A
49	A	90	B	148	A	263	A

При определении оценок параметров **для причины A** будем считать времена отказов по причине B временами правого цензурирования, и наоборот.

Пример 4 (продолжение)



Индивидуальное домашнее задание

Исходные данные:

- Выборка времен отказа системы объема N из ЛР №4;
- Вероятностная модель надежности, признанная лучшей для этой выборки по результатам ЛР №5.

Задание:

1. Модифицировать выборку с полными данными так, чтобы
 - около 10% значений были цензурированы справа;
 - около 30% значений были цензурированы справа;
 - около 60% значений были цензурированы справа;
 - около 10% значений были цензурированы слева;

Индивидуальное домашнее задание

- около 30% значений были цензурированы слева;
- около 60% значений были цензурированы слева.

Все выборки с цензурированными данными должны быть сохранены в файлы Excel. Каждому значению времени в файле должно соответствовать одно из обозначений: F – для наблюдаемого отказа, R – для времени правого цензурирования, L – для времени левого цензурирования.

2. Разделить выборку с полными данными на 8 непересекающихся интервалов;
на 12 непересекающихся интервалов.

Индивидуальное домашнее задание

3. Сохранить сгруппированные данные в файлы Excel. Каждому объекту из выборки должны соответствовать два числа: значения левой и правой границы интервала, на котором произошел отказ.
4. Для каждой цензурированной выборки определить точечные оценки значений параметров модели методом максимального правдоподобия.
5. Сравнить полученные значения оценок со значениями, полученными для полной выборки.

Отчет и полученные выборки необходимо направить по адресу alexeyfremov@tpu.ru