



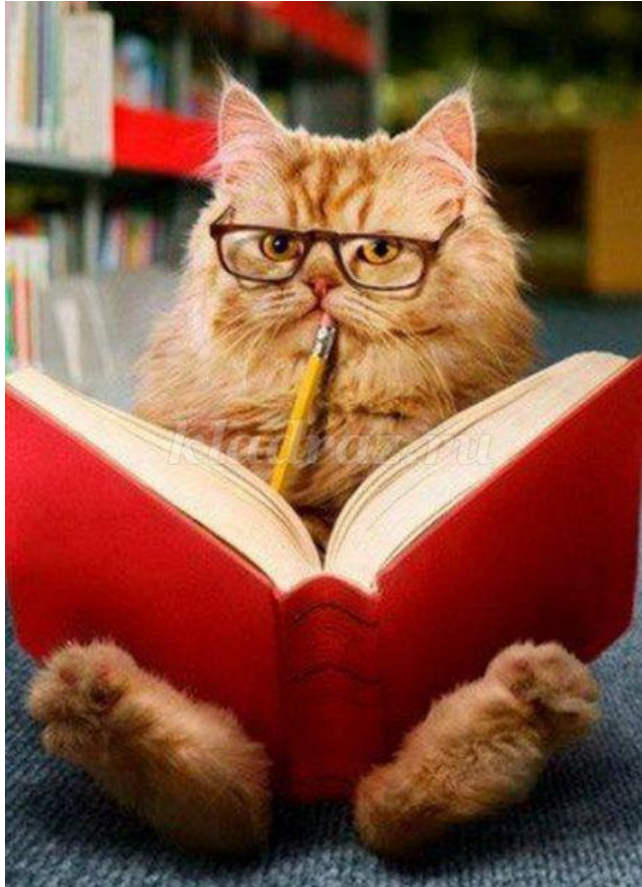
УНИВЕРСИТЕТ ИТМО

ЗАНЯТИЕ 3

ТЕОРИЯ И ПРАКТИКА СТАТИСТИЧЕСКИХ ВЫВОДОВ

Раздел 1. Генеративная статистика и метод Монте-Карло

Александр Валерьевич Бухановский, Анна Владимировна Калюжная



Статистический вывод:

использование выборочных данных для получения и формализации знаний о свойствах генеральной совокупности



Оценить количественные характеристики генеральной совокупности (вопрос «насколько



точно?»)

Выбрать и настроить модель описания генеральной совокупности (вопрос «подходит или нет?»)




Валидировать модель, т.е. оценить ее качество (вопрос «насколько

Метод Монте-Карло



Генеративный (регенеративный) подход: исследование статистических свойств на основе «размножения» заданной выборки

Механизмы генерации:

 На основе априорных знаний о свойствах генеральной

 совокупности

На основе действий с выборкой
Метод Монте-Карло (Метод статистических испытаний): синтез процессов и явлений с использованием случайных чисел

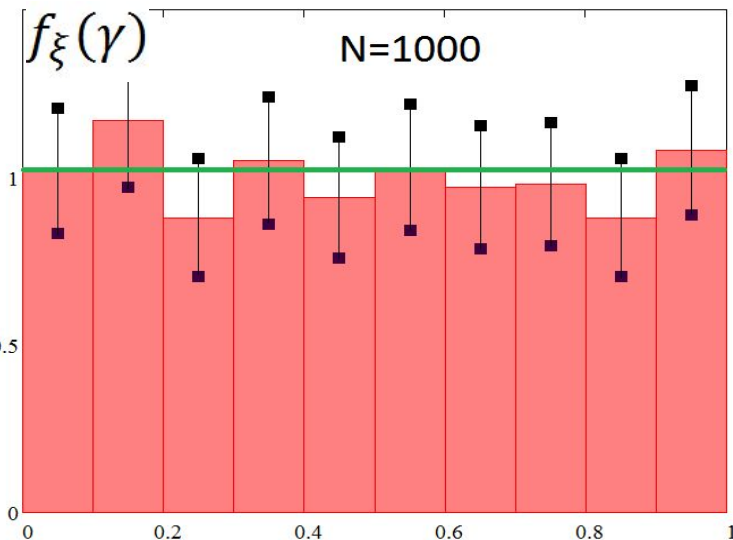


Случайные числа



Значения γ равномерно распределенной случайной величины

$$f_{\xi}(\gamma) = \begin{cases} 1, & \gamma \in [0,1] \\ 0, & \gamma \notin [0,1] \end{cases}$$



Псевдослучайные числа (ПСЧ): на основе детерминированного алгоритма (заданный период повторяемости)

Случайные числа (СЧ): на основе физических механизмов или



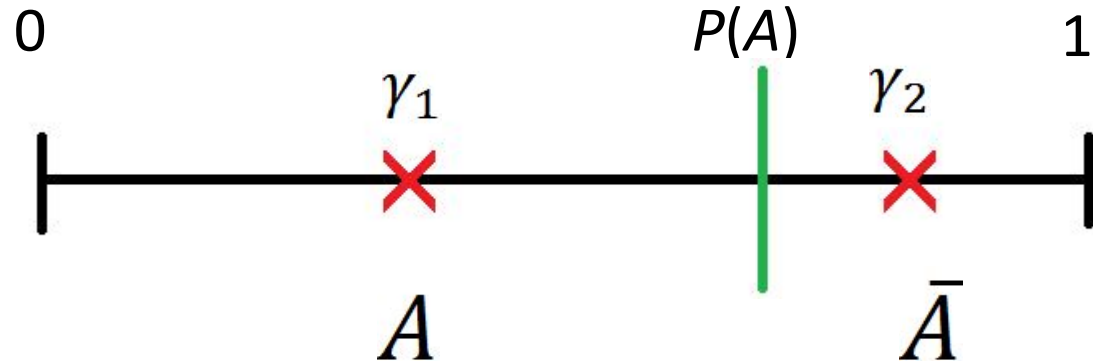
Моделирование случайных событий

Геометрическая интерпретация случайного события заданной вероятности $P(A)$ - **имитация попадания** точки в интервал $[0, P(A)]$



Выбор варианта реализации события по значению γ :

$$\begin{cases} A, \gamma \leq P(A) \\ \bar{A}, \gamma > P(A) \end{cases}$$



Моделирование дискретных величин

Закон распределения
(вероятности набора событий)

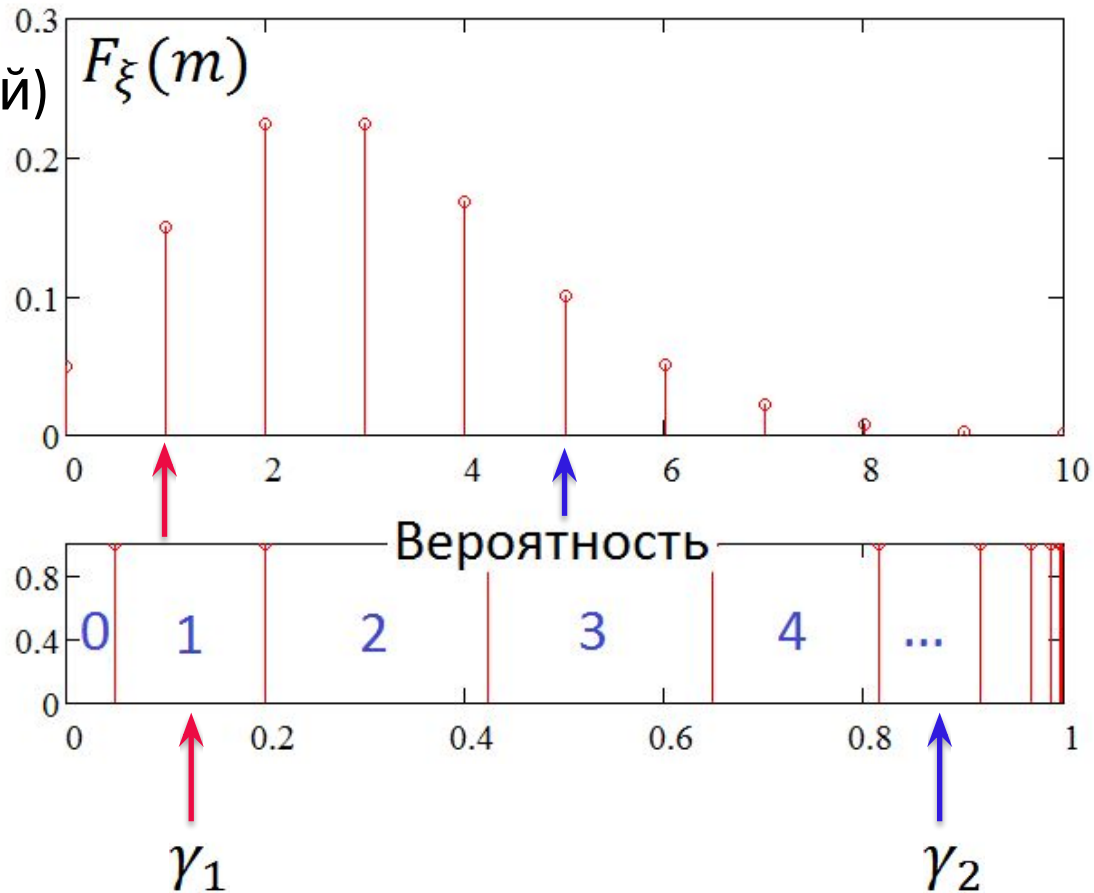
$$F_{\xi}(m) = P(\xi = x_m), \quad m \in \mathbb{N}$$

Выбор реализации события

$$A_m: \xi = x_m$$

при попадании точки в
интервал

$$\gamma \in \left[\sum_{k=0}^m F_{\xi}(k), \sum_{k=0}^{m+1} F_{\xi}(k) \right]$$



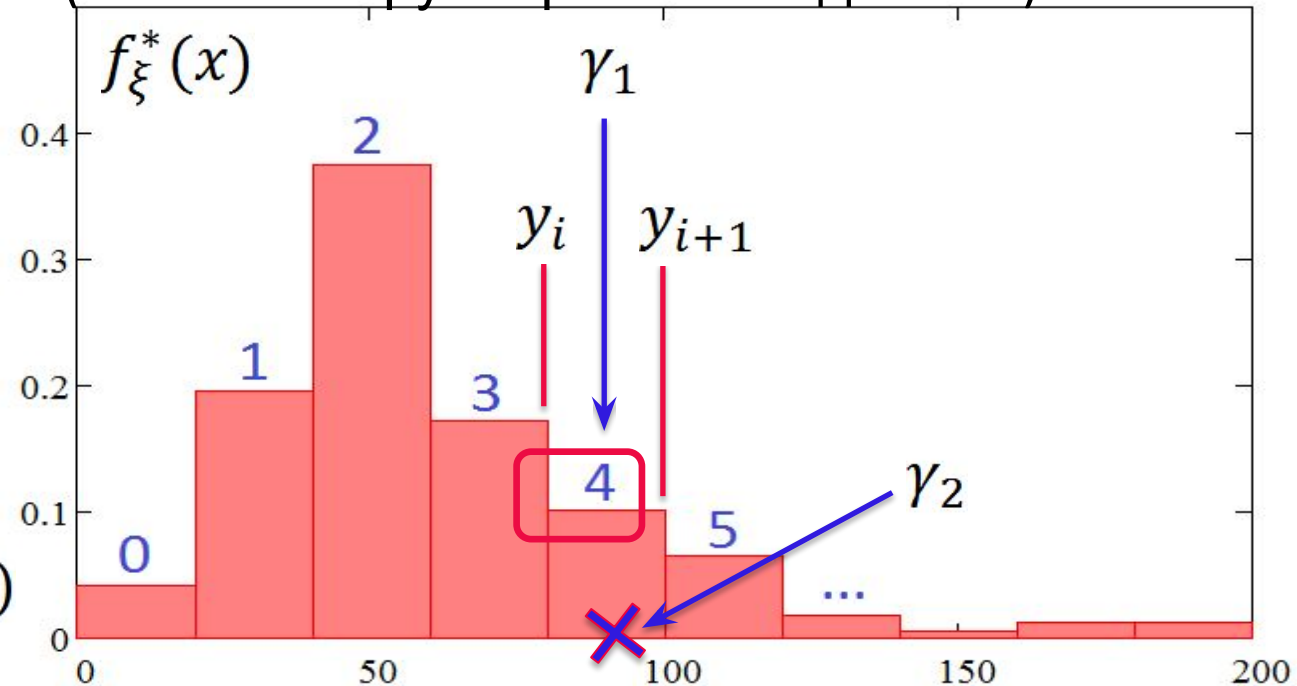
Моделирование непрерывных величин (1/3)



Дискретно-равномерное приближение

(на основе группированных данных)

1. Выбор номера столбца i
2. Выбор значения x внутри столбца



$$x = y_i + \gamma(y_{i+1} - y_i)$$



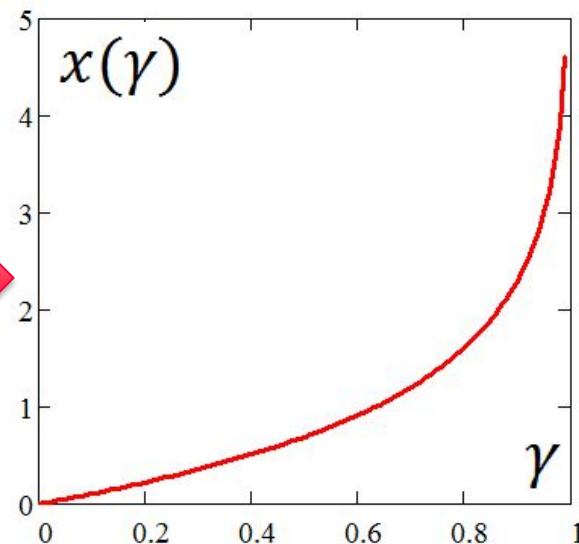
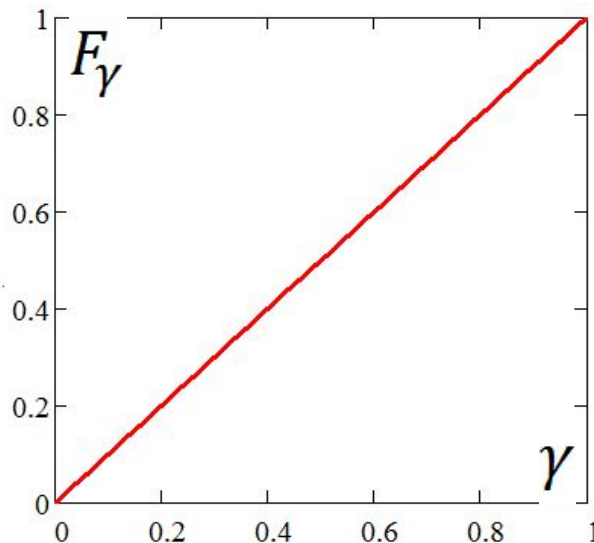
Метод обратной функции

(на основе преобразования квантилей)

$$x = F_{\xi}^{-1}(\gamma)$$

Для плотности
распределения
(напр.,
ядерной оценки)

$$x: \int_{-\infty}^x f_{\xi}(y) dy = \gamma$$



Геометрический метод

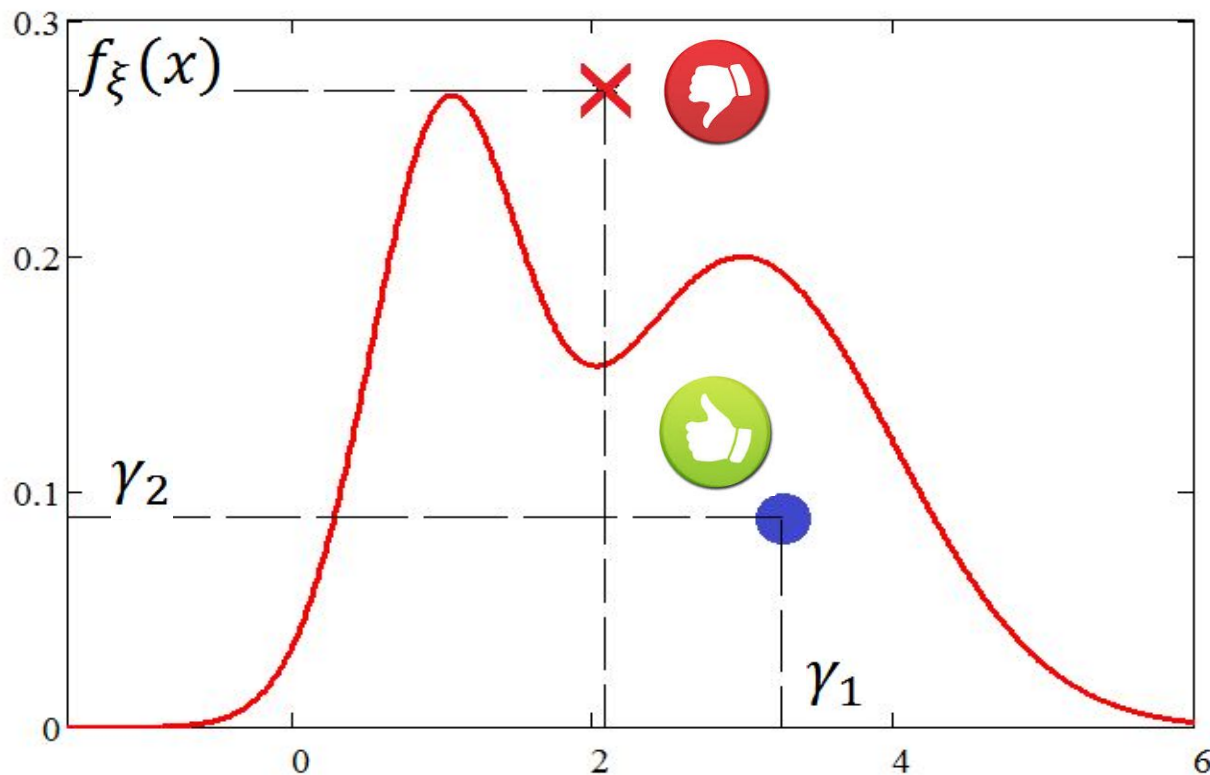
(на основе плотности распределения)



$$x = \gamma_1$$

при условии,

$$\text{что } f_{\xi}(\gamma_1) \geq \gamma_2$$



Моделирование гауссовых случайных

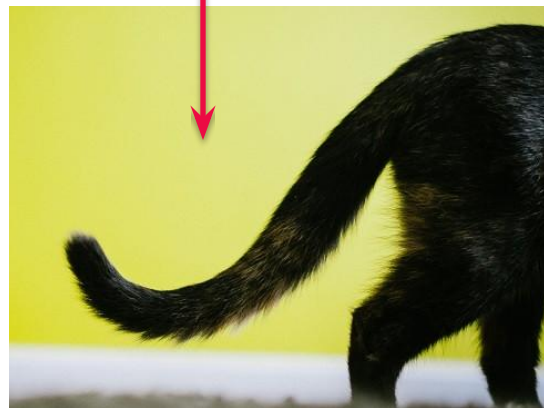
ВАРИАЦИЙ



Нормированное распределение Гаусса
 $N(0,1)$

1) На основе **центральной предельной теоремы**:

$$x = \sum_{k=1}^{12} \gamma_i - 6$$



2) На основе **преобразования Бокса-Мюллера**:

$$x_1 = \sin(2\pi\gamma_1) \sqrt{-2 \ln \gamma_2}$$

$$x_2 = \cos(2\pi\gamma_1) \sqrt{-2 \ln \gamma_2}$$

Интервальное оценивание на основе Монте-

Карло

Для избранной статистики $\Xi^* = T_P(x_1, \dots, x_N)$



Задать аналитическую модель распределения $F_\xi(x, \Theta)$ или ее оценку $F_\xi^*(x) = F_\xi(x, \Theta^*)$



Выбрать метод моделирования случайной величины с этим распределением



Сгенерировать M выборок $(x_1^{(j)}, \dots, x_N^{(j)})$, $j = 1 \dots M$



Рассчитать по каждой выборке точечную оценку ξ_j^* статистики Ξ^* и



По выборке $(\xi_1^*, \dots, \xi_M^*)$ определить характерные параметры, в том числе границы $\beta\%$ -доверительного интервала:

$$I_1^{(\beta)} = \xi_{(\lfloor M(1-\beta)/2 \rfloor + 1)}^*$$

$$I_2^{(\beta)} = \xi_{(\lfloor M(1+\beta)/2 \rfloor + 1)}^*$$

Задача об отборе на выставку кошек (1/2)



Найти интервальные оценки математического ожидания и дисперсии для: (а) выборки из всех 109 кошек, (б) для случайно отобранных 10

```
Чтение исходных данных - чистая выборка (кошки)
X := READPRN("109_Cats.dat")      N := rows(X)      i := 0..N-1      N = 109
uβ := 1.96      Квантиль распределения Гаусса для 95% интервала      β := 0.95
```

Точечные оценки по выборке

```
Mx := mean(X)      Mx = 5.806      Среднее значение
Sx := stdev(X)      Sx = 0.425      Выборочное СКО
Dx := var(X)      Dx = 0.181      Выборочная дисперсия
```

Доверительные интервалы для выборочных средних и дисперсий - расчет по асимптотическим формулам

95%-доверительные интервалы для выборочного среднего

```
IMniz := Mx - uβ * Sx / sqrt(N)      IMver := Mx + uβ * Sx / sqrt(N)
IMniz = 5.726      IMver = 5.886
```

95%-доверительные интервалы для выборочной дисперсии

```
IDniz := Dx * (N - 1) / qchisq(0.975, N)      IDver := Dx * (N - 1) / qchisq(0.025, N)
IDniz = 0.14      IDver = 0.238
```

Доверительные интервалы для выборочных средних и дисперсий - расчет методом Монте-Карло

```
M := 10000      Число испытаний метода Монте-Карло      j := 0..M-1
```

```
k := 0..11
Yi,j := Mx + Sx * (sum(rnd(1) - 6, k))
```

Генератор M выборок гауссовых случайных чисел со средним Mx и СКО Sx, каждая из N значений

Расчет средних и дисперсий для каждой из выборок

```
Mx_MCj := mean(Y^j)      Dx_MCj := var(Y^j)
```

Расчет границ доверительных (толерантных) интервалов по выборкам Mx_MC и Dx_MC - как соответствующих квантилей

```
no(p) := floor(M*p)      Mx_MC_sort := sort(Mx_MC)      Dx_MC_sort := sort(Dx_MC)
```

95%-доверительные интервалы для выборочного среднего

```
Mx_MC_sort_no(0.025) = 5.727      Mx_MC_sort_no(0.97) = 5.882
```

95%-доверительные интервалы для выборочной дисперсии

```
Dx_MC_sort_no(0.025) = 0.136      Dx_MC_sort_no(0.975) = 0.23
```

Задача об отборе на выставку кошек (2/2)

Чтение исходных данных - чистая выборка (кошки)

```
X := READPRN("10_SmallSampleCats.dat")
```

```
N := rows(X)
```

```
i := 0..N-1
```

```
N = 10
```

```
uβ := 1.96    Квантиль распределения Гаусса для 95% интервала
```

Точечные оценки по выборке

```
Mx := mean(X)    Mx = 5.968    Среднее значение
```

```
Sx := stdev(X)   Sx = 0.39     Выборочное СКО
```

```
Dx := var(X)     Dx = 0.152    Выборочная дисперсия
```

Доверительные интервалы для выборочных средних и дисперсий - расчет по асимптотическим формулам

95%-доверительные интервалы для выборочного среднего

$$IMniz := Mx - u\beta \cdot \frac{Sx}{\sqrt{N}} \quad IMver := Mx + u\beta \cdot \frac{Sx}{\sqrt{N}}$$

$$IMniz = 5.726 \quad IMver = 6.21$$

95%-доверительные интервалы для выборочной дисперсии

$$IDniz := \frac{Dx \cdot (N-1)}{qchisq(0.975, N)} \quad IDver := \frac{Dx \cdot (N-1)}{qchisq(0.025, N)}$$

$$IDniz = 0.067 \quad IDver = 0.421$$

Доверительные интервалы для выборочных средних и дисперсий - расчет методом Монте-Карло

```
M := 10000    Число испытаний метода Монте-Карло    j := 0..M-1
```

```
k := 0..11
```

```
Yi,j := Mx + Sx ·  $\left( \sum_k \text{rnd}(1) - 6 \right)$  Генератор M выборок гауссовых случайных чисел со средним Mx и СКО Sx, каждая из N значений
```

Расчет средних и дисперсий для каждой из выборок

```
Mx_MCj := mean(Y(j))    Dx_MCj := var(Y(j))
```

Расчет границ доверительных (толерантных) интервалов по выборкам Mx_MC и Dx_MC - как соответствующих квантилей

```
no(p) := floor(M·p)    Mx_MC_sort := sort(Mx_MC)    Dx_MC_sort := sort(Dx_MC)
```

95%-доверительные интервалы для выборочного среднего

$$Mx_MC_sort_{no(0.025)} = 5.73 \quad Mx_MC_sort_{no(0.975)} = 6.213$$

95%-доверительные интервалы для выборочной дисперсии

$$Dx_MC_sort_{no(0.025)} = 0.04 \quad Dx_MC_sort_{no(0.975)} = 0.287$$

Формирование M *псевдовыборок*

$$\left(x_1^{(j)}, \dots, x_N^{(j)}\right), j = 1 \dots M$$

на основе случайного *выбора с возвращением* из исходной выборки:



Непараметрические методы: джекнайф



Формирование N **подвыборок**





$$\left(x_1^{(j)}, \dots, x_{N-1}^{(j)} \right), j = 1 \dots N$$

путем **выкалывания** одного из членов исходной выборки:



Методическая
проблема джекнайфа
и бутстрепа:
смещенность оценок

Обобщение: генеративная статистика

-  – **Метод Монте-Карло:** универсальный инструмент статистического оценивания, когда аналитика не справляется
-  – **Методы моделирования случайных величин:** могут использовать как параметрические, так и непараметрические оценки распределений
-  – **Интервальное оценивание** на основе метода Монте-Карло: просто, надежно, но чувствительно к модели распределения
-  – **Методы бутстреп и джекнайф:** не зависят от распределения, но оценки могут быть смещены

Автор признателен всем котикам
(и хозяевам) за возможность
некоммерческого использования
их изображений, размещенных
в публичном Интернете



Александр Валерьевич Бухановский, Анна Владимировна Калюжная
boukhanovsky@mail.ifmo.ru

IT'sMOre than a
UNIVERSITY



УНИВЕРСИТЕТ ИТМО

ЗАНЯТИЕ 3

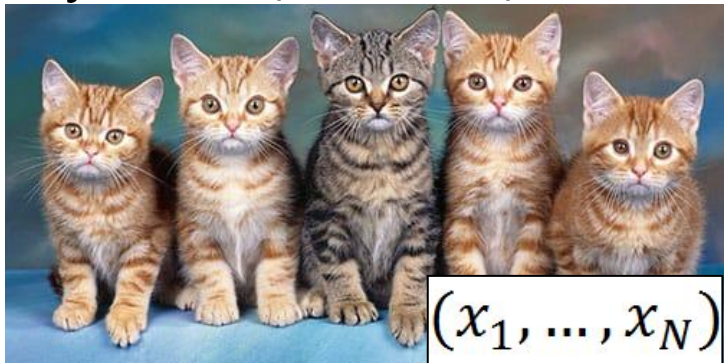
ТЕОРИЯ И ПРАКТИКА СТАТИСТИЧЕСКИХ ВЫВОДОВ

Раздел 2. Проверка гипотез.
Непараметрические статистические критерии

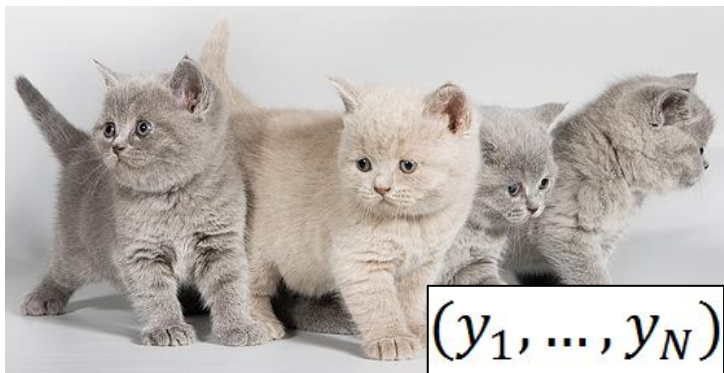
Александр Валерьевич Бухановский, Анна Владимировна Калюжная

Проверка статистических гипотез

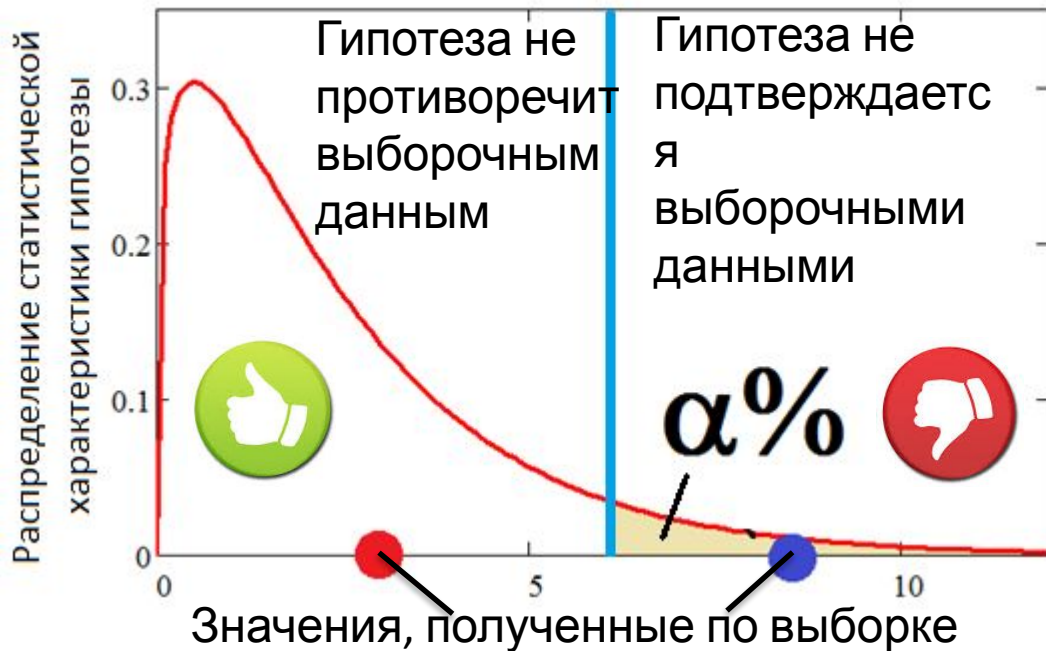
Нулевая (базовая) гипотеза *Статистический критерий*: механизм



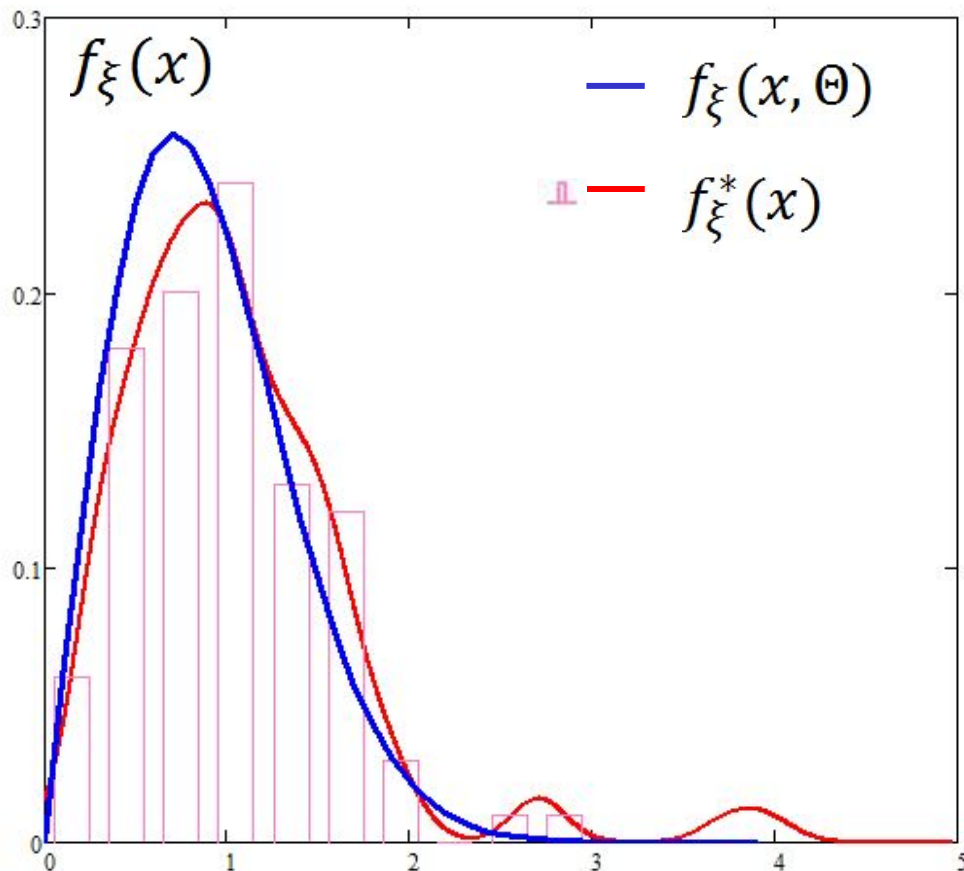
? || ?



проверки нулевой гипотезы путем сравнения выборочных данных с теоретическим «эталоном»




Гипотеза о виде распределения (пример)




Критерий согласия


Базовая гипотеза H_0 :

 $f_{\xi}(x, \Theta) = f_{\xi}^*(x)$


Альтернатива H_1 :

 $f_{\xi}(x, \Theta) \neq f_{\xi}^*(x)$

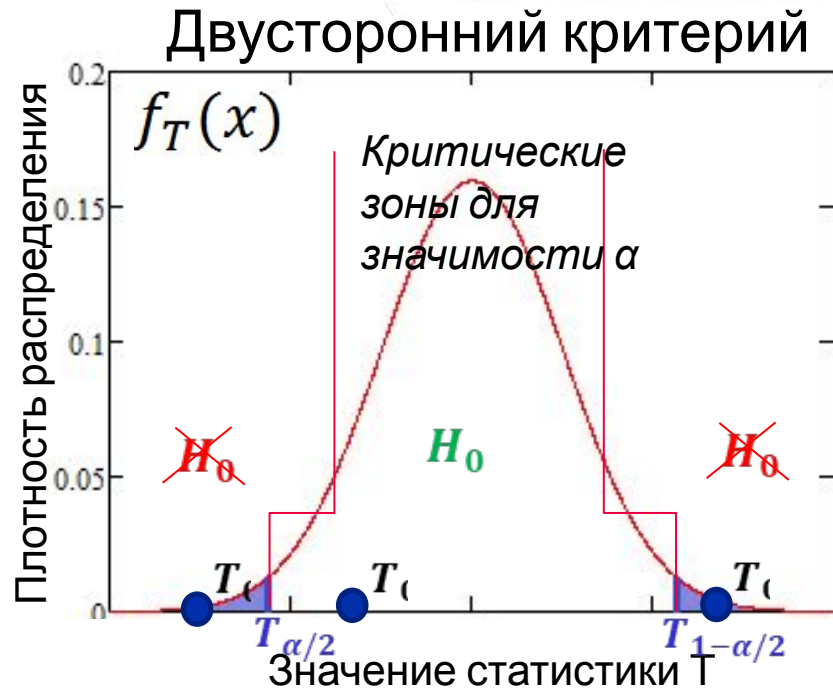
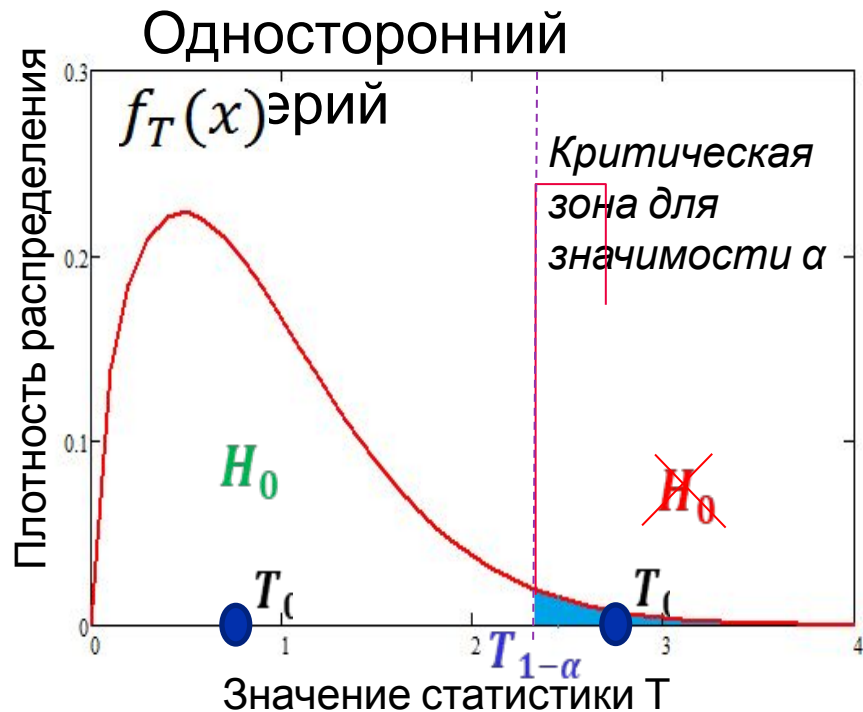
Теоретическая характеристика

 $T = \mathfrak{G}[f_{\xi}]$

Выборочная

 $T_0 = \mathfrak{F}[f_{\xi}, f_{\xi}^*(x_i)]$

Критерии: сравнение характеристик

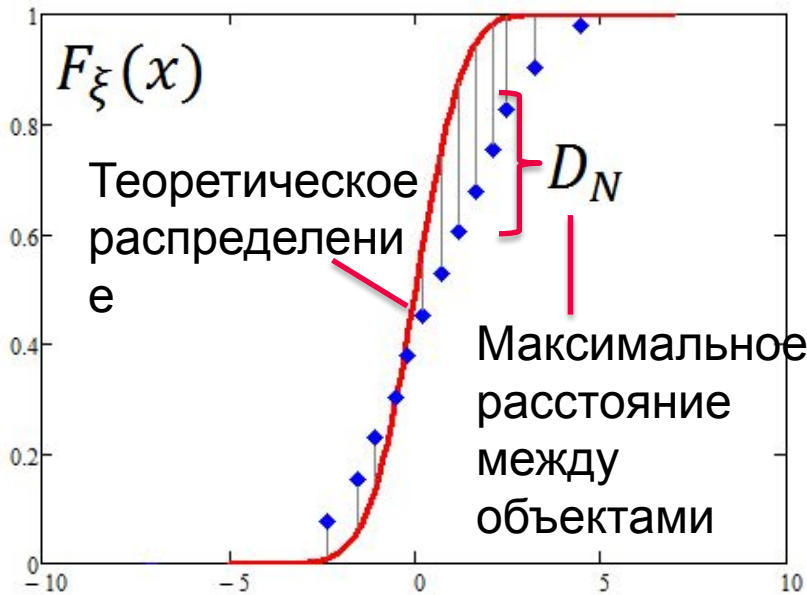


Ошибка проверки непараметрических гипотез

(*ошибка*

1 рода) – с вероятностью α отвергнуть верную

Критерий Колмогорова



Проверка гипотезы **о виде (модели)** распределения

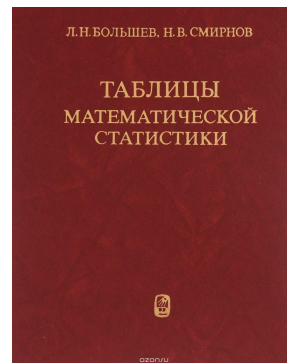
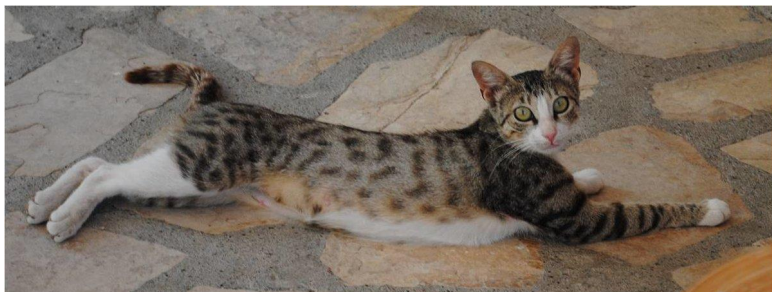
Статистическая характеристика:

$$D_N = \max |F^*(x) - F(x)|$$

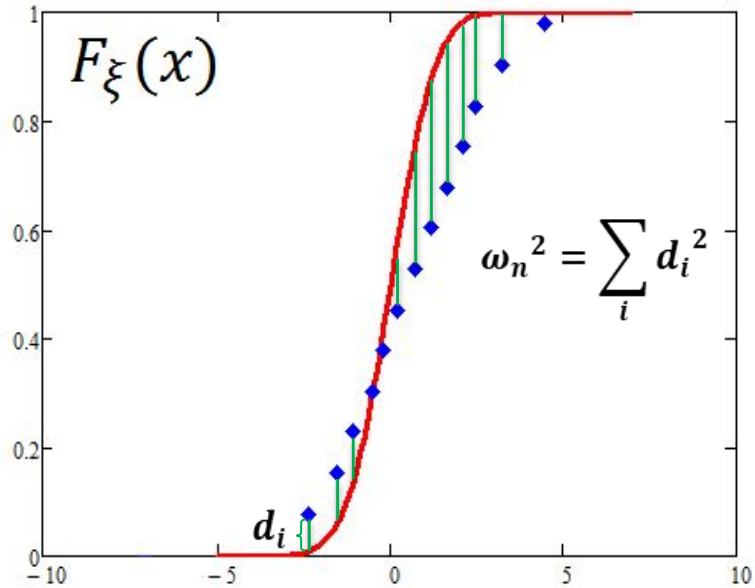
Критерий проверки (**отрицания** H_0):

$$\sqrt{N}D_N \geq T_{1-\alpha}$$

$T_{1-\alpha}$ - квантиль
распределения
Колмогорова



Критерии ω^2 (Крамера-вон Мизеса-Смирнов)



Проверка гипотезы **о виде (модели)** распределения

Статистическая характеристика:

$$N\omega_N^2 = \frac{1}{12N} + \sum_{i=1}^N \left[F(x_{(i)}) - \frac{2i-1}{2N} \right]^2$$

Критерий проверки (**отрицан** H_0)

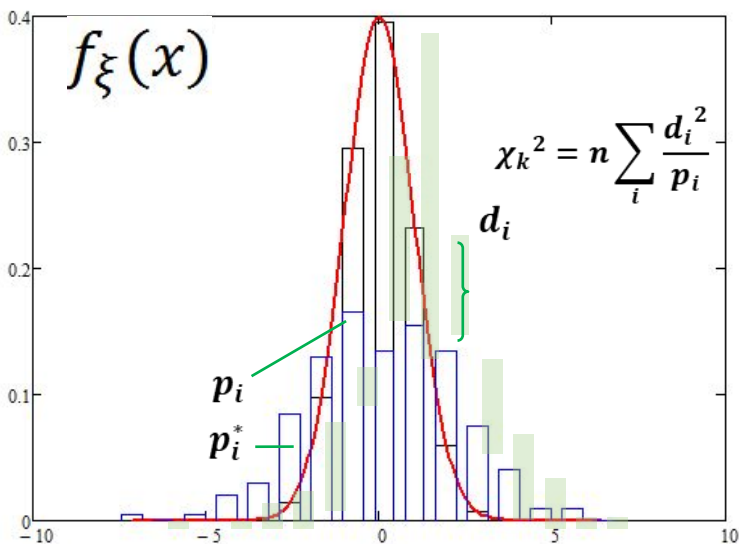
):

$$N\omega_N^2 \geq T_{1-\alpha}$$

$T_{1-\alpha}$ - квантиль
табулированного



Критерий χ^2 (Пирсона)



Проверка гипотезы **о виде (модели)** распределения *по группированным данным*

Статистическая характеристика :

$$\chi_N^2 = N \sum_{j=1}^M \frac{(p_j^* - p_j)^2}{p_j}$$

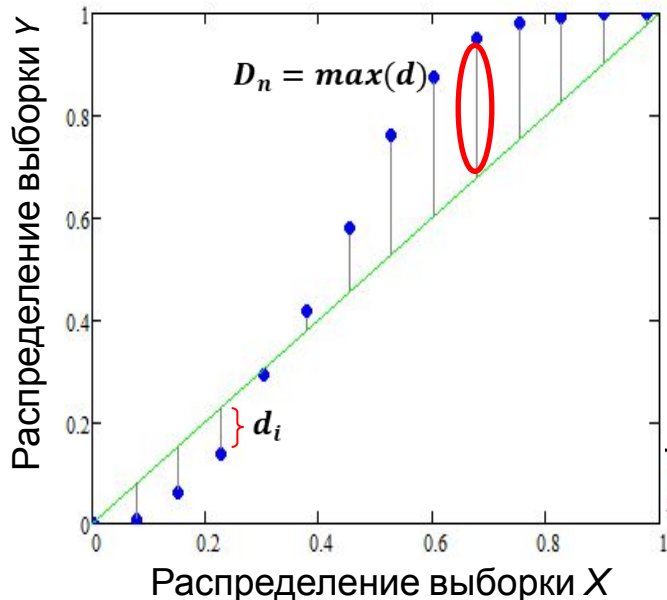
Критерий проверки (**отрицания** H_0):

$$\chi_N^2 \geq T_{1-\alpha}$$

$T_{1-\alpha}$ - квантиль
распределения χ_{N-1}^2

Критерии для проверки однородности

Вероятностный



Базовая гипотеза $H_0 : f_{\xi}^*(x) = g_{\xi}^*(x)$

Критерий Колмогорова →

См $D_{NM} = \max |F^*(x) - G^*(x)|,$

Критерий ω^2 → Розенблатта

$$\frac{NM}{N+M} \omega_{NM}^2 = \frac{1}{N+M} \left(\frac{1}{6} + \frac{1}{M} \sum_{i=1}^N (R_i - i)^2 + \frac{1}{N} \sum_{j=1}^M (S_j - j)^2 \right) - \frac{2}{3}$$

Правила проверки
—
как для критериев

1.4

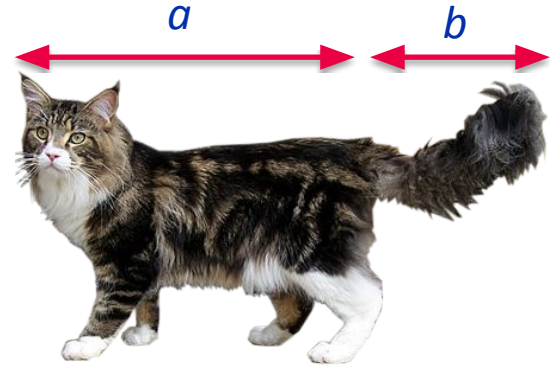


Задача о поддельных мейн-кунах (1/3)

В одном питомнике разводят крупных котов и продают как мейн-кунов. Однако ассоциация заводчиков мейн-кунов засомневалась в том, что эти коты действительно принадлежат породе, а не являются смесью с норвежской лесной кошкой. Для того, чтобы принять или отклонить гипотезу о принадлежности котов из питомника породе мейн-кун, заводчики решили сравнить распределения характерного параметра кошек между собой. В качестве параметров для сравнения могут быть выбраны отношения длин тел животных, хвостов или хвоста к длине тела.

Известно, что в среднем длина тел котов норвежской породы и мейн-кунов совпадает и составляет 47 - 50 см. Известно также, что отношение длины хвоста к телу у породистых норвежских кошек должно быть больше 0,9, а у породистых мейн-кунов больше $3/5$. Для чистоты эксперимента сравнение решили проводить только на взрослых котах, которых в питомнике оказалось 30. Из базы ассоциации мейн-кунов для сопоставления были выбраны 43 самых породистых кота.

Однако, стоит учесть, что если кошки из питомника являются смесью пород, тогда ими могут быть унаследованы как признаки



$$x = a/b$$

Задача о поддельных мейн-кунах (2/3)

1) Выборки длин тел мейн-кунов и котов из питомника:

$$BLM := M^{(1)}$$

$$BLF := F^{(1)}$$

2) Выборки длин хвостов мейн-кунов и котов из питомника:

$$TLM := M^{(1)}$$

$$TLF := F^{(1)}$$

3) Выборки отношений длин хвостов к телам мейн-кунов и котов из питомника:

$$RM := M^{(2)}$$

$$RF := F^{(2)}$$

File Edit View Insert Format Tools Symbolics Window Help

Normal Arial 10 B I U

```
M := READPRN("Meinkun_data.dat")      F_ := READPRN("Fam_cats_data.dat")      m := rows(M)      n := rows(F)
M - мейн-куны                          F - коты из питомника                  m = 43            n = 30
                                          j = 0..m-1        i = 0..n-1
```

Получаем эмпирические функции распределения на основе ядерных оценок с гауссовым ядром:

Для отношения длин тел к хвостам:

$$K_gauss(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$$

Для длин тел:

$$hBLM := 1.1$$

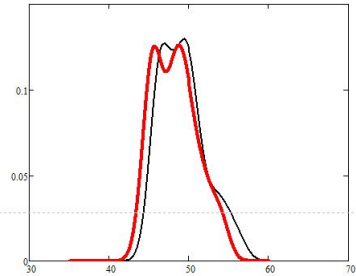
$$hBLF := 1.1$$

$$fBLM(y) := \frac{1}{m \cdot hBLM} \sum_j K_gauss\left(\frac{y - BLM_j}{hBLM}\right)$$

$$fBLF(y) := \frac{1}{n \cdot hBLF} \sum_i K_gauss\left(\frac{y - BLF_i}{hBLF}\right)$$

$$d := 0.1$$

$$y := 35.35 + d..60$$



$$z := 35.35 + d..60$$

Для длин хвостов:

$$hTLM := 1.1$$

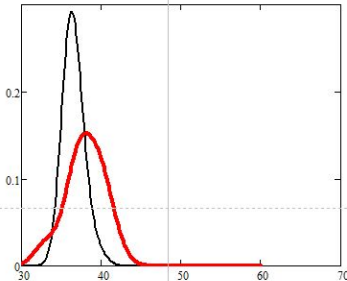
$$hTLF := 1.3$$

$$fTLM(y) := \frac{1}{m \cdot hTLM} \sum_j K_gauss\left(\frac{y - TLM_j}{hTLM}\right)$$

$$fTLF(y) := \frac{1}{n \cdot hTLF} \sum_i K_gauss\left(\frac{y - TLF_i}{hTLF}\right)$$

$$d := 0.1$$

$$yt := 30.30 + d..60$$



$$zt := 30.30 + d..60$$

$$hRM := 0.05$$

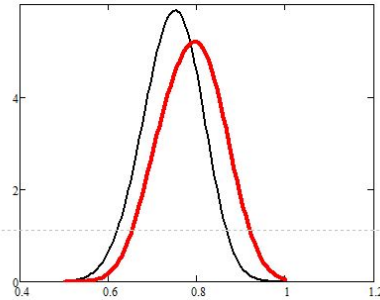
$$hRF := 0.05$$

$$fRM(y) := \frac{1}{m \cdot hRM} \sum_j K_gauss\left(\frac{y - RM_j}{hRM}\right)$$

$$fRF(y) := \frac{1}{n \cdot hRF} \sum_i K_gauss\left(\frac{y - RF_i}{hRF}\right)$$

$$dr := 0.005$$

$$xr := 0.5, 0.5 + dr..1$$



$$zr := 0.5, 0.5 + dr..1$$

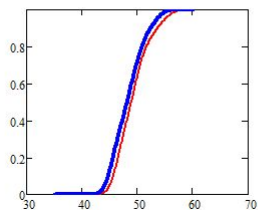
Задача о поддельных мейн-кунах (3/3)

Проверка на однородность выборок с помощью критерия Смирнова:

Квантиль распределения Колмогорова для уровня значимости 5%:

$D_{cr} = 1.36$

$$FBLM(z) := \int_{35}^z fBLM(y) dy \quad FBLF(z) := \int_{35}^z fBLF(y) dy$$



Для длин тел:

$$L_x := \text{floor} \left[\frac{(\max(BLM, BLF) - \min(BLM, BLF))}{d} \right]$$

$$L_x = 119$$

$$p := 0..L_x$$

$$X(p) := \min(BLM, BLF) + 0.1 \cdot p$$

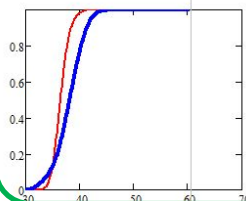
$$D_p := |FBLM(X(p)) - FBLF(X(p))|$$

$$DD := \sqrt{\frac{(n-m)}{n+m}} \cdot \max(D)$$

$$DD = 0.5$$

$DD < D_{cr}$, поэтому H_0 принята

$$FTLM(z) := \int_{30}^{zt} fTLM(y) dy \quad FTLF(z) := \int_{30}^{zt} fTLF(y) dy$$



Для длин хвостов:

$$L_{xt} := \text{floor} \left[\frac{(\max(TLM, TLF) - \min(TLM, TLF))}{d} \right]$$

$$L_{xt} = 100$$

$$p_{xt} := 0..L_{xt}$$

$$X_t(p_t) := \min(TLM, TLF) + d \cdot p_t$$

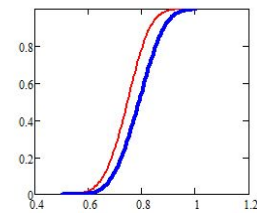
$$D_{p_t} := |FTLM(X_t(p_t)) - FTLF(X_t(p_t))|$$

$$DD_t := \sqrt{\frac{(n-m)}{n+m}} \cdot \max(D_t)$$

$$DD_t = 1.539$$

$DD_t > D_{cr}$, поэтому H_0 отвергнута

$$FRM(zr) := \int_{0.5}^{zr} fRM(yr) dyr \quad FRF(zr) := \int_{0.5}^{zr} fRF(yr) dyr$$



Для отношения длин тел к хвостам:

$$L_{xr} := \text{floor} \left[\frac{(\max(RM, RF) - \min(RM, RF))}{dr} \right]$$

$$L_{xr} = 51$$

$$p_r := 0..L_{xr}$$

$$X_r(p_r) := \min(RM, RF) + dr \cdot p_r$$

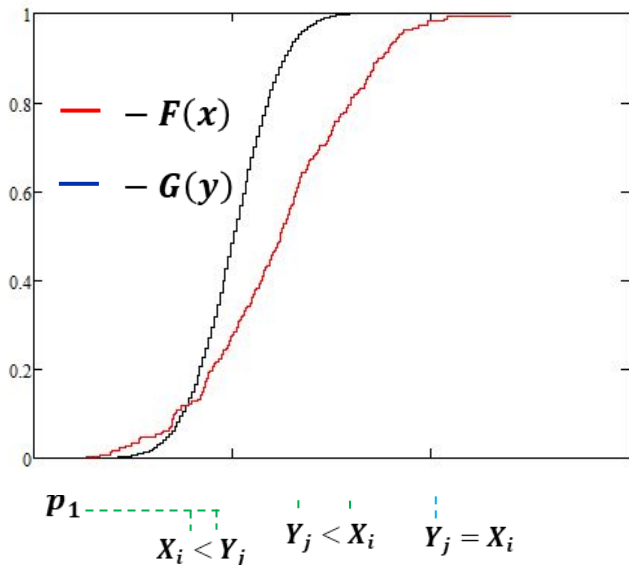
$$D_{p_r} := |FRM(X_r(p_r)) - FRF(X_r(p_r))|$$

$$DD_r := \sqrt{\frac{(n-m)}{n+m}} \cdot \max(D_r)$$

$$DD_r = 1.003$$

$DD_r < D_{cr}$, поэтому H_0 принята

Критерий ранговых сумм Вилкоксона



Неоднородность как превалирование

Базовая гипотеза H_0 $F(x) = G(x)$

Альтернативная гипотеза H_2 $F(x) < G(x)$

Выборки (x_1, \dots, x_N) и (y_1, \dots, y_M)

Статистическая характеристика:

$$U_{st} = \frac{\min\{U_X, U_Y\} - \frac{1}{2}MN}{\sqrt{\frac{1}{12}MN(M+N+1)}}$$







Критерий проверки
(**отрицание** H_0):

$$U_{st} \notin [Z_{\alpha/2}, Z_{1-\alpha/2}]$$

$$Z \sim \mathcal{N}(0,1)$$

$$U_Y = \sum_{i=1}^N \sum_{j=1}^M [x_i > y_j] \quad U_X = \sum_{i=1}^N \sum_{j=1}^M [x_i < y_j]$$

Обобщение: непараметрические критерии

-  – **Решаемые задачи:** проверка соответствия распределения определенному закону, проверка однородности двух выборок.
-  – **Все критерии несовершенны:** критерий Колмогорова прост, но чувствителен к выбросам, критерий согласия Пирсона – для группированных данных (все проблемы гистограмм), критерий Крамера-вон Мизеса-Смирнова табулирован не для всех распределений
-  – **Проверку можно строить по-разному:** просто проверять **неоднородность**, модифицируя критерии для теоретических распределений (Смирнова, Розенблатта), или оценивать превалирование (критерии Вилкоксона и Манна-Уитни).
-  – **Проверка гипотез – не панацея:** она может только опровергать,

Автор признателен всем котикам
(и хозяевам) за возможность
некоммерческого использования
их изображений, размещенных
в публичном Интернете



Александр Валерьевич Бухановский, Анна Владимировна Калюжная
boukhanovsky@mail.ifmo.ru

IT'sMOre than a
UNIVERSITY



УНИВЕРСИТЕТ ИТМО

ЗАНЯТИЕ 3

ТЕОРИЯ И ПРАКТИКА СТАТИСТИЧЕСКИХ ВЫВОДОВ

Раздел 3. Проверка гипотез. Параметрические
статистические критерии

Александр Валерьевич Бухановский, Анна Владимировна Калюжная

Логика параметрических критериев



Объект исследования

– Θ

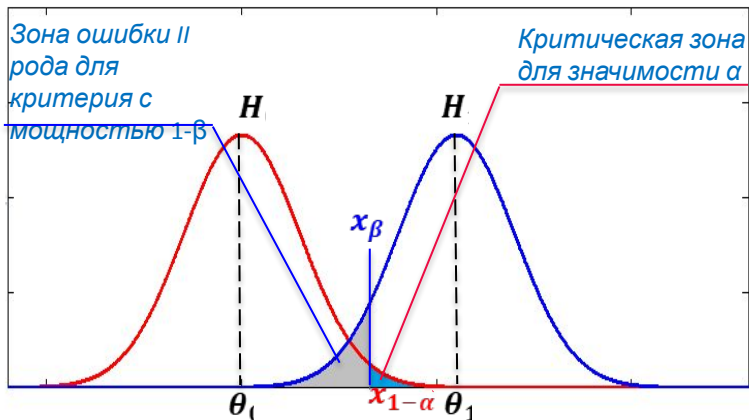
набор параметров,
характеризующих
модель

распределения

генеральной
совокупности $F(x) \equiv F(x, \Theta)$

А правда ли, что параметр $\Theta^* = \Theta_0$ (базовая гипотеза);
или все-таки параметр $\Theta^* = \Theta_1$ (альтернативная гипотеза)?

Ошибки первого и второго рода

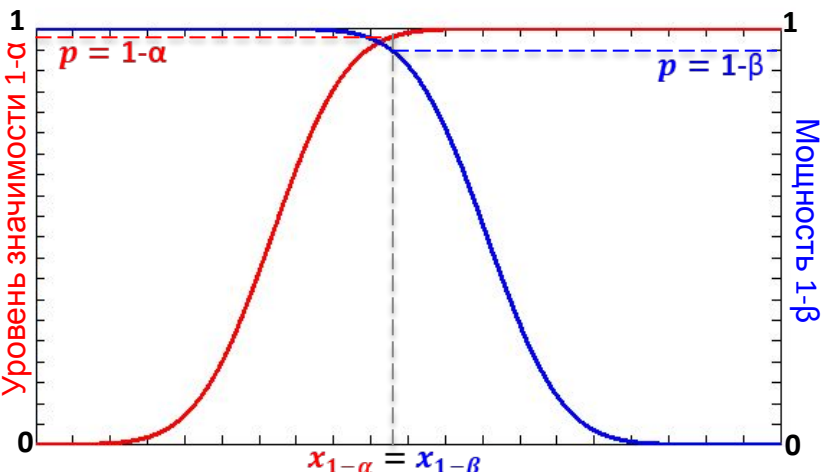


Ошибка статистического оценивания (вероятность $1-\beta$):
упустить истинное значение из $\beta\%$ -доверительного интервала

Ошибки проверки гипотез:

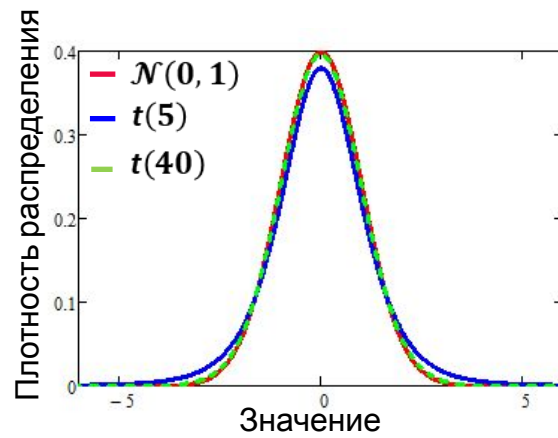
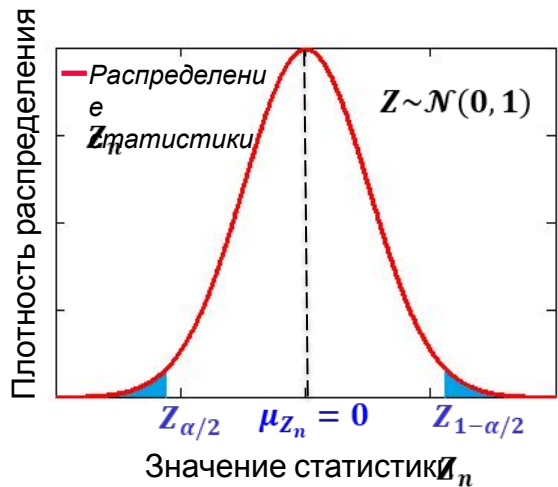
А) **Ошибка I рода** (вероятность α):
отвергнуть верную гипотезу

Б) **Ошибка II рода** (вероятность γ):
принять неверную гипотезу



Построение наиболее **МОЩНЫХ** (в своем классе) критериев: $1 - \gamma \rightarrow \max$
при заданном уровне значимости α

Критерий для среднего значения



Базовая гипотеза: $\bar{x} = \mu$

Для больших выборок:

Статистическая характеристика $Z_N = \left| \frac{\bar{x} - \mu}{s/\sqrt{N}} \right|$

Критерий проверки: $Z_N > |Z_{1-\alpha}|$ $Z \sim \mathcal{N}(0, 1)$

Для малых выборок ($N < 30$):

$$t_N = \sqrt{N-1} \frac{|\bar{x} - \mu|}{s}$$

Критерий

проверки:

$$t_N > |t_{1-\alpha, N-1}|$$



χ^2 критерий для выборочной дисперсии



Базовая гипотеза: $s^2 = \sigma^2$

Статистическая характеристика:

$$\chi_N^2 = \frac{(N - 1)s^2}{\sigma^2}$$

Критерий проверки (**отрицание** H_0)

):

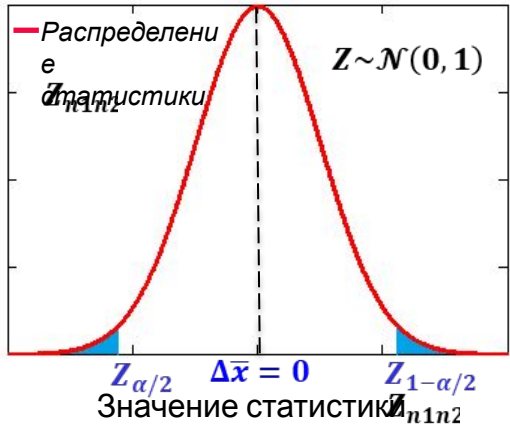
$$\chi_N^2 \in \left[0, \chi^2_{\frac{\alpha}{2}, N-1} \right] \cup \left[\chi^2_{1-\frac{\alpha}{2}, N-1}, \infty \right]$$

$\chi^2_{\frac{\alpha}{2}, N-1}$ - табличная квантиль
распределения χ^2 с $N-1$
степенью свободы

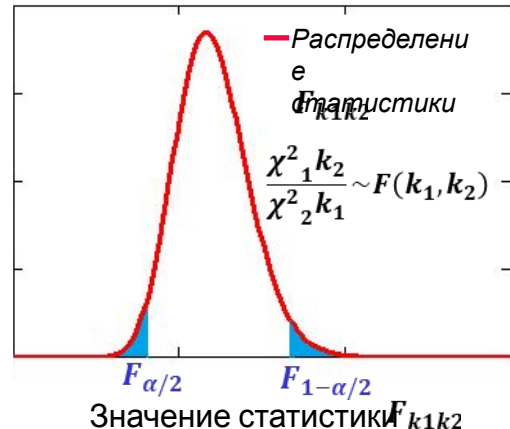


Критерии для равенства средних и дисперсий

Плотность распределения



Плотность распределения



Равенство двух средних значений:

Статистическая характеристика
Критерий

$$Z_{NM} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{N} + \frac{s_2^2}{M}}}$$

$$Z_{NM} > |Z|_{1-\alpha}$$

Равенство двух выборочных дисперсий:

Статистическая характеристика:
Критерий

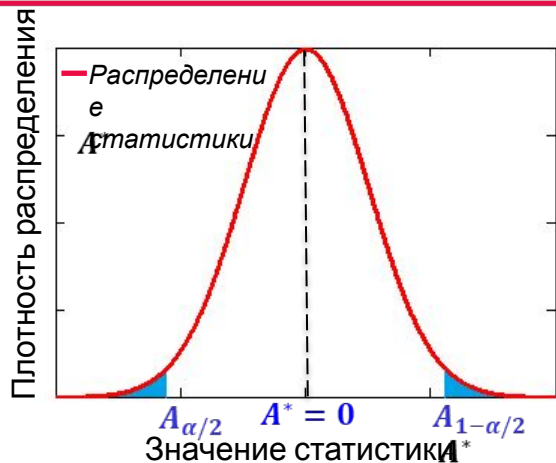
$$F_{NM} = \frac{s_1^2}{s_2^2}$$

проверки:

$$F_{NM} \in \left[0, F_{\frac{\alpha}{2}, N-1, M-1} \right] \cup \left[F_{1-\frac{\alpha}{2}, N-1, M-1}, \infty \right)$$

F - табулированное распределение

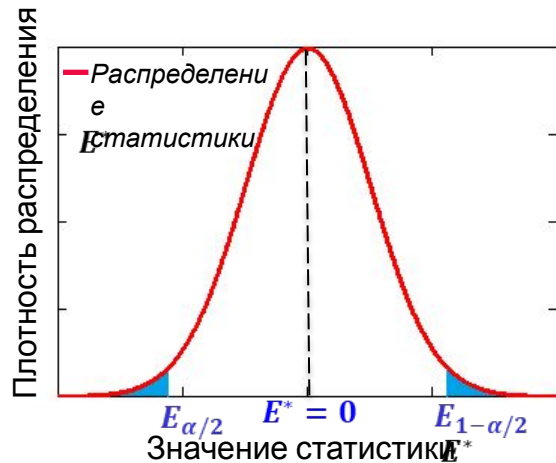
Критерии для асимметрии и эксцесса



**Суждение об асимметрии
распределения $A^* = 0$**

Базовая гипотеза:
Критерий проверки $A^* \in \left(-\infty, A_{\frac{\alpha}{2}}\right) \cup \left(-\infty, A_{1-\frac{\alpha}{2}}\right)$

$$A \sim \mathcal{N}\left(0, \sqrt{\frac{6N(N-1)}{(N+3)(N+1)(N-2)}}\right)$$








**Суждение о выраженности пика
распределения $E^* = 0$**

Базовая гипотеза:
Критерий проверки $E^* \in \left(-\infty, E_{\frac{\alpha}{2}}\right) \cup \left(-\infty, E_{1-\frac{\alpha}{2}}\right)$

$$E \sim \mathcal{N}\left(0, \sqrt{\frac{24N(N-1)^2}{(N+5)(N+3)(N-2)(N-3)}}\right)$$

Метод Монте-Карло для проверки гипотез

Для базовой гипотезы $H_0: \Theta^* = \Theta_0$ и альтернативы $H_1: \Theta^* = \Theta_1$

-  Задать аналитическую модель распределения $F_\xi(x, \Theta)$
-  Выбрать метод моделирования случайной величины с этим распределением
-  Сгенерировать M выборок по закону $F_\xi(x, \Theta_0)$, и M выборок по закону $F_\xi(x, \Theta_1)$
-  Рассчитать по каждой выборке точечную оценку параметра Θ_1^* в Θ_0^* и
-  Оценить распределения $F_{\Theta_1}(x)$ $F_{\Theta_0}(x)$ оценок параметров

Для заданного значения уровня значимости α найти ошибку второго рода γ .

Задача об отборе на выставку кошек

$$Mx0 := 5.97 \quad N := 30$$

$$Sx := 0.39$$

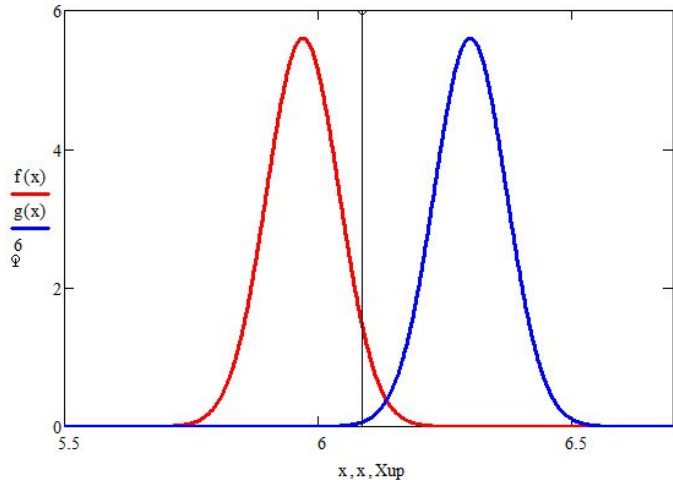
$$Mx1 := 6.3$$

$$x := 5.5, 5.501 \dots 7.0$$

$$f(x) := \text{dnorm}\left(x, Mx0, \frac{Sx}{\sqrt{N}}\right) \quad g(x) := \text{dnorm}\left(x, Mx1, \frac{Sx}{\sqrt{N}}\right)$$

$$Xup := \text{qnorm}\left(0.95, Mx0, \frac{Sx}{\sqrt{N}}\right)$$

$$Xup = 6.087$$



$$G := \int_{5.5}^{Xup} g(x) dx$$

$$G = 0.14\%$$

Средний вес по выборке из 30 кошек составил 5.97 кг, а СКО – 0.39 кг.

Можно ли с 10% уровнем значимости считать, что математическое ожидание веса кошек по генеральной совокупности составляет 6.0 кг?

Какова при этом вероятность реализации гипотезы, что математичес



3.1



2.3

Развитие: статистические игры и решения

Решение $d_i \in D$ и его **функция потерь** $L(x, d)$

Цель: выбор оптимальной последовательности решений

$$d_1 \rightarrow d_2 \rightarrow \dots \rightarrow d_m$$

на основе ограниченного числа наблюдений (x_1, \dots, x_N)
за случайной величиной Ξ



d_k



Риск: средние потери при решении ∞

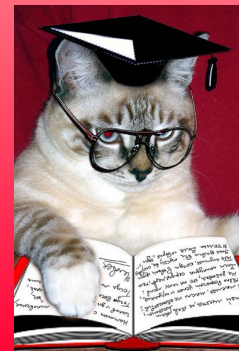
$$R_{d_k} = \int_{-\infty}^{\infty} L(x, d_k) f_{\Xi}(x) dx$$

Смысл игры: предугадать реакцию «противника» (значение Ξ) так, чтобы

минимизировать свои риски



Автор признателен всем котикам
(и хозяевам) за возможность
некоммерческого использования
их изображений, размещенных
в публичном Интернете



Александр Валерьевич Бухановский, Анна Владимировна Калюжная
boukhanovsky@mail.ifmo.ru

IT'sMO *re than a*
UNIVERSITY



УНИВЕРСИТЕТ ИТМО

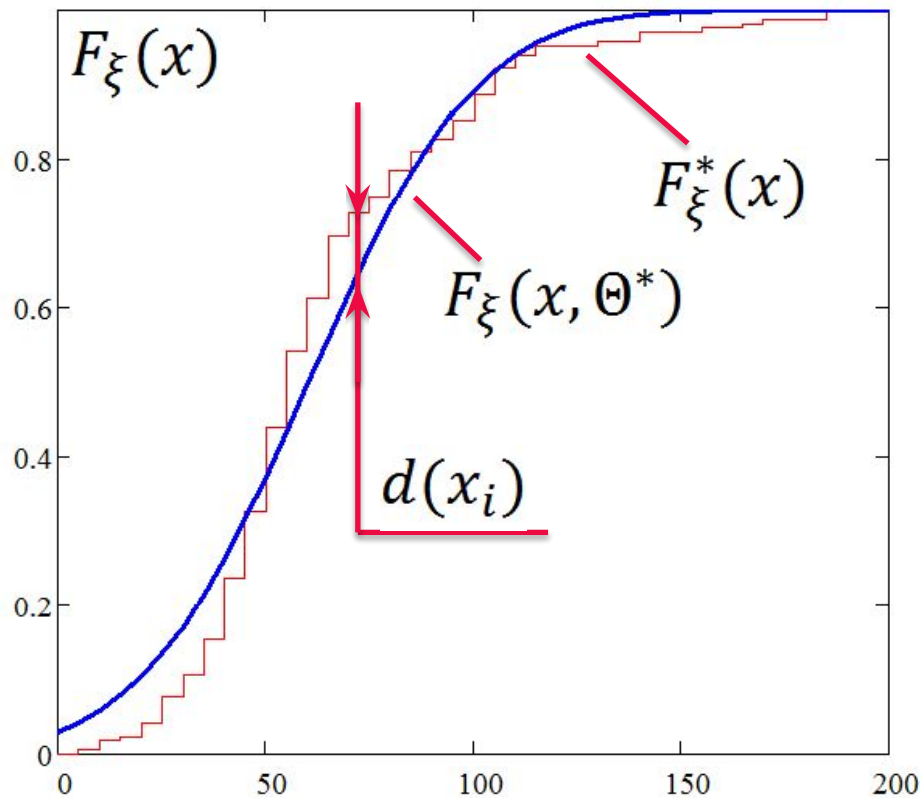
ЗАНЯТИЕ 3

ТЕОРИЯ И ПРАКТИКА СТАТИСТИЧЕСКИХ ВЫВОДОВ

Раздел 4. Конструктивные модели распределений

Александр Валерьевич Бухановский, Анна Владимировна Калюжная

Подбор распределения как задача приближения




Минимизация нормы невязки: $\|d(x)\|_{x \in [a, b]} \xrightarrow{\Theta} \min$

Идея метода наименьших квадратов (МНК)


Невязка как **евклидово расстояние**:

$$d(x_p) = (x_p^* - x_p(\Theta))^2$$

Идея метода (минимизация суммы квадратов невязок по L характерным квантилям):


$$\sum_{k=1}^L (x_{p_k}^* - x_{p_k}(\Theta))^2 \rightarrow \min_{\Theta}$$

Реализация метода (система уравнений):


$$\sum_{k=1}^L (x_{p_k}^* - x_{p_k}(\Theta)) \frac{\partial x_{p_k}(\Theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, l$$



МНК для линейно-масштабируемых величин

Квантиль линейно-масштабируемой величины:

$U_{p_k} = F_{\xi^{(0)}}^{-1}(p_k)$ - квантиль нормированного модельного распределения

$$x_{p_k}(\Theta) = \theta_1 + \theta_2 U_{p_k}$$



1.1

Реализация МНК:
система из двух
линейных
алгебраических
уравнений

$$\left\{ \begin{array}{l} L\theta_1 + \theta_2 \sum_{k=1}^L U_{p_k} = \sum_{k=1}^L x_{p_k}^* \\ \theta_1 \sum_{k=1}^L U_{p_k} + \theta_2 \sum_{k=1}^L U_{p_k}^2 = \sum_{k=1}^L U_{p_k} x_{p_k}^* \end{array} \right.$$

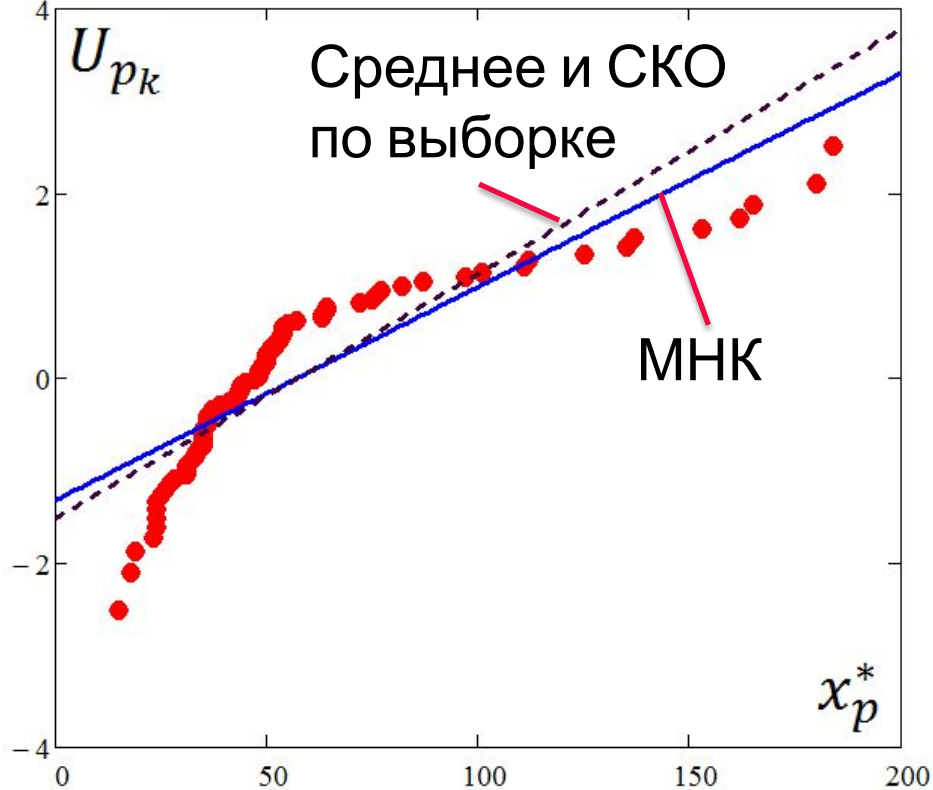
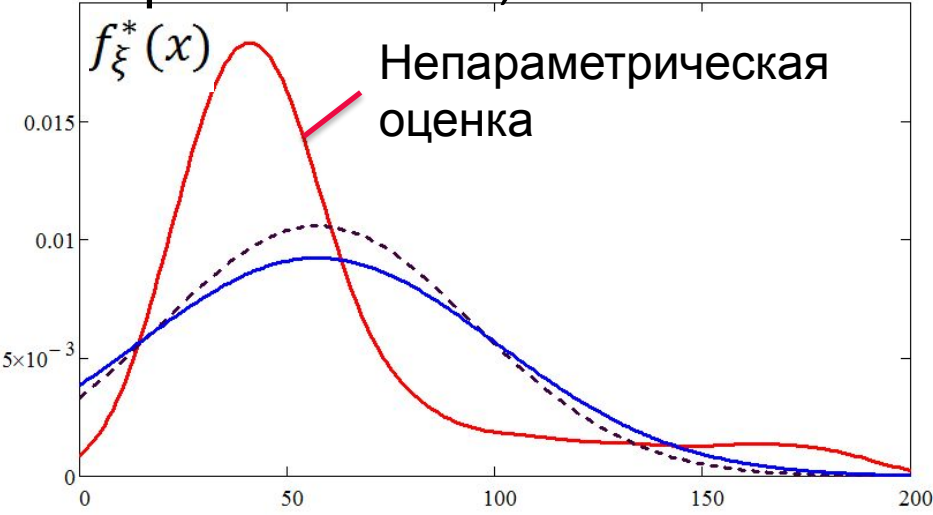
Простейшее приближение: МНК по всей выборке

Задача: приблизить модельное распределение к выборке, которая **не вполне** ему соответствует



Все выборочные квантили равнозначны,

$$L = N$$



Взвешенный МНК: учитываем особенности

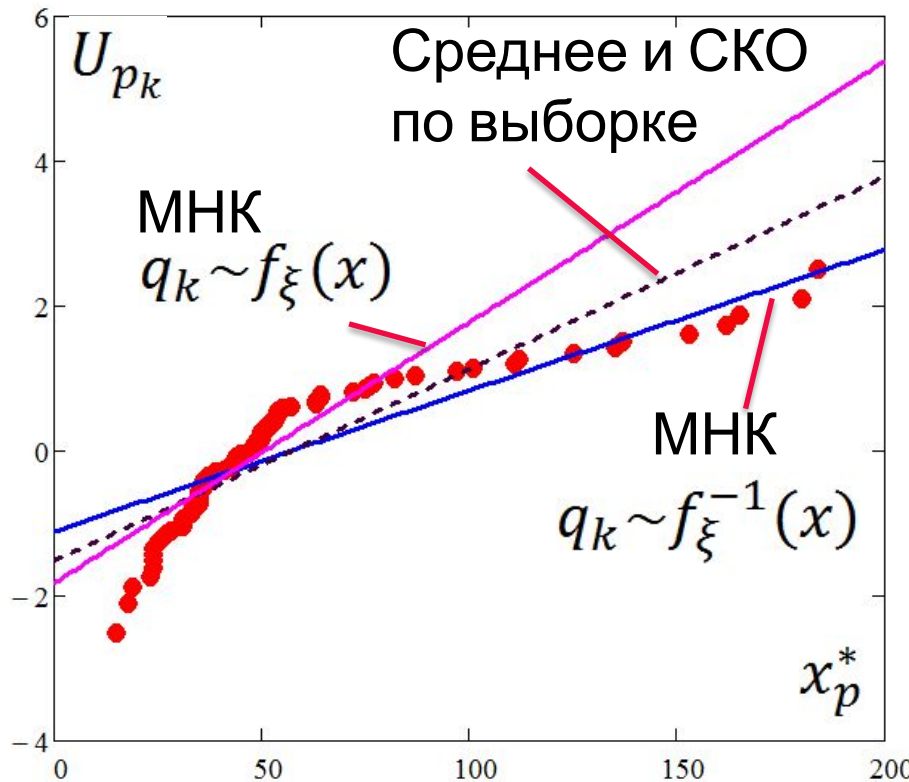
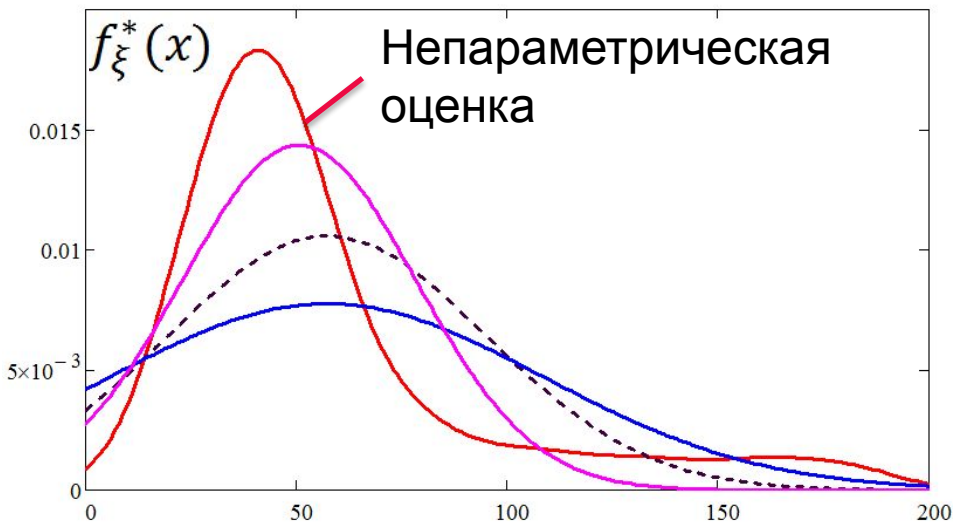
$$\sum_{k=1}^L q_k (x_{p_k}^* - x_{p_k}(\Theta))^2 \rightarrow \min_{\Theta}$$

Целевая функция с весами q_k



Выборочные квантили

неравнозначны, $L = N$



МНК по избранным квантилям

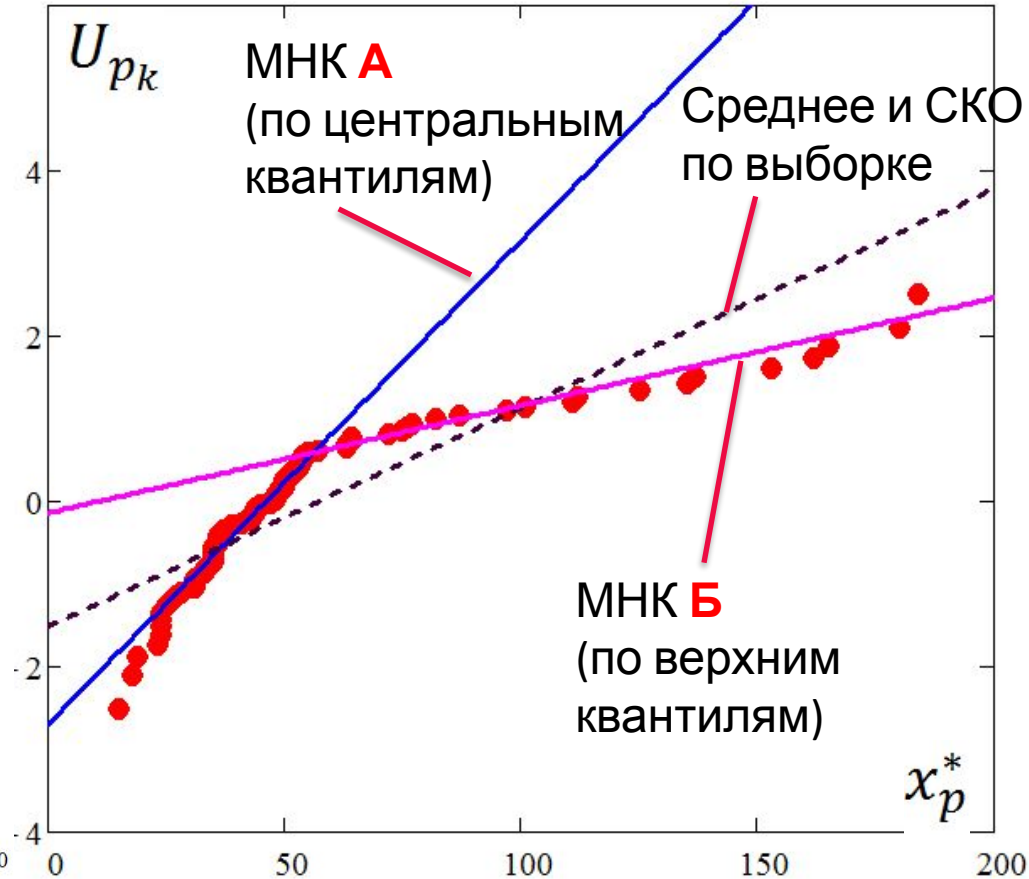
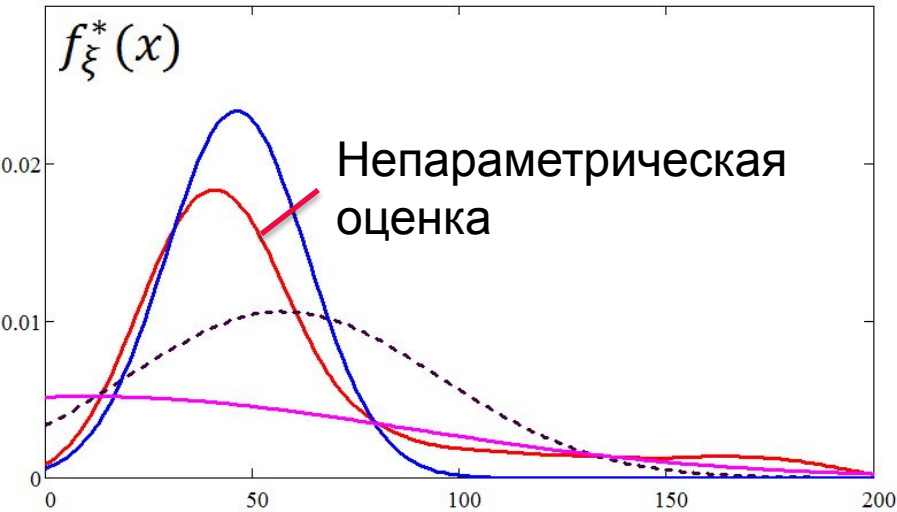
Предположение: избранные квантили равнозначны, $L < N$

А: квантили

(10%, 25%, 35%, 50%, 65%, 75%)

Б: квантили

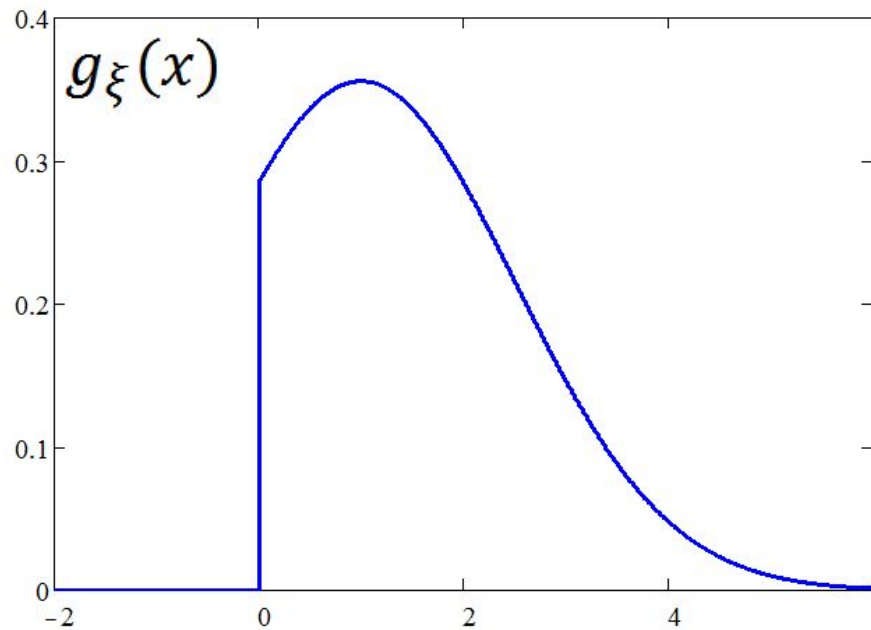
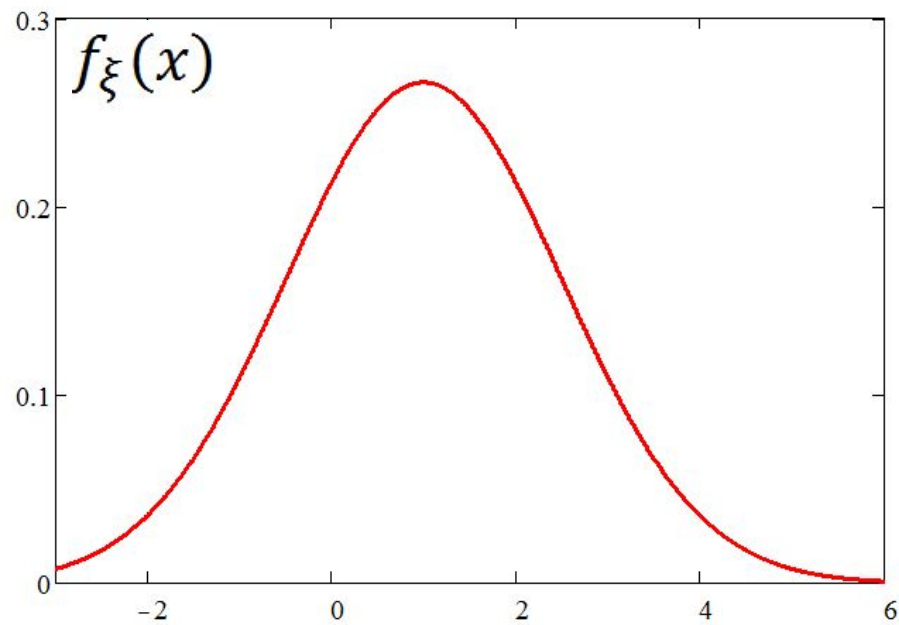
(75%, 85%, 90%, 95%, 97%, 99%)



Модель усеченного распределения



$$g_{\xi}(x) = \begin{cases} f_{\xi}(x)/C, & x \geq a \\ 0, & x < a \end{cases} \quad C = \int_a^{+\infty} f_{\xi}(x) dx$$

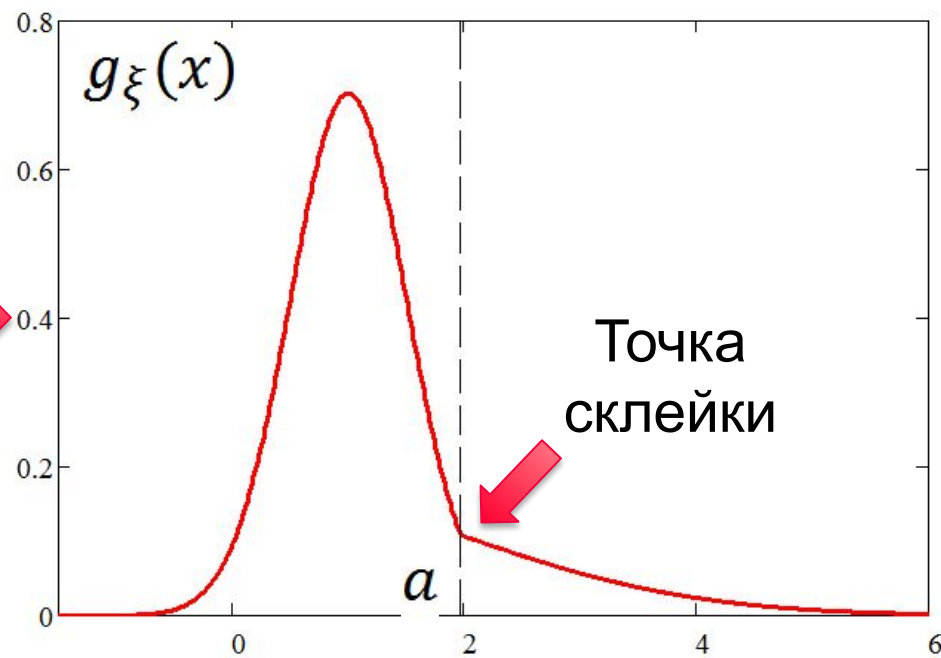
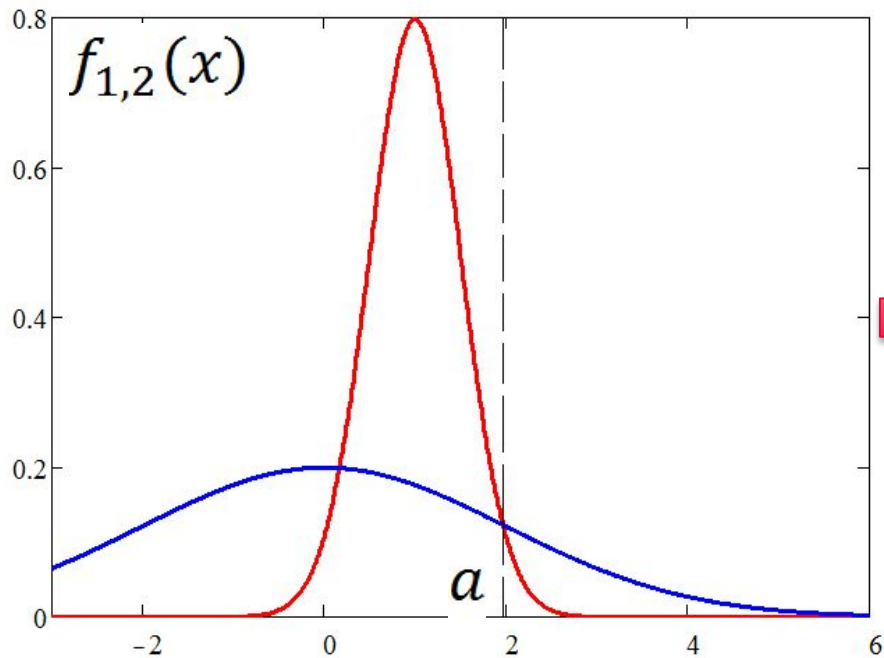


Модель склейки распределений



$$g_{\xi}(x) = \begin{cases} f_2(x)/C, & x \geq a \\ f_1(x)/C, & x < a \end{cases}$$

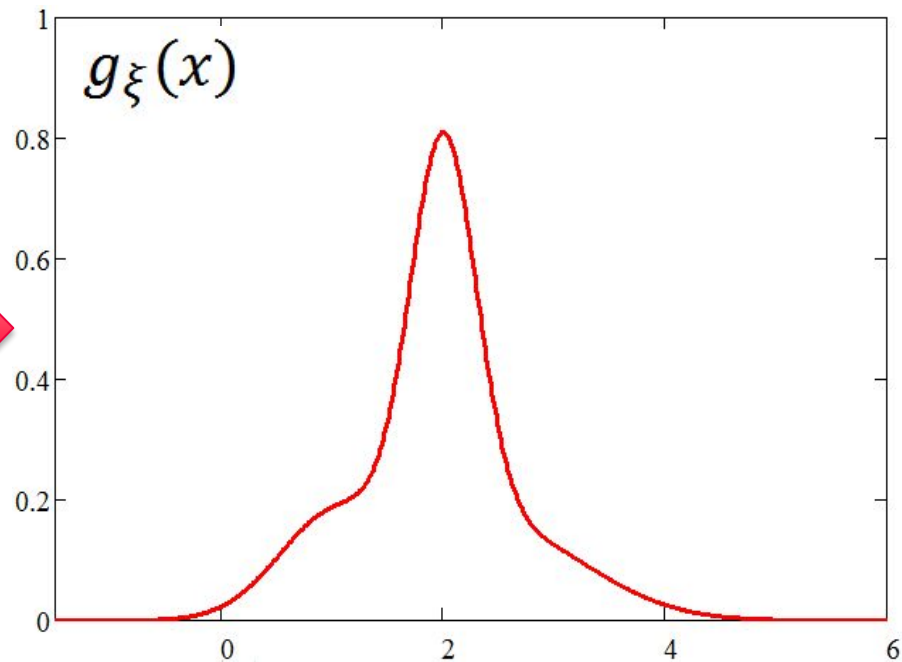
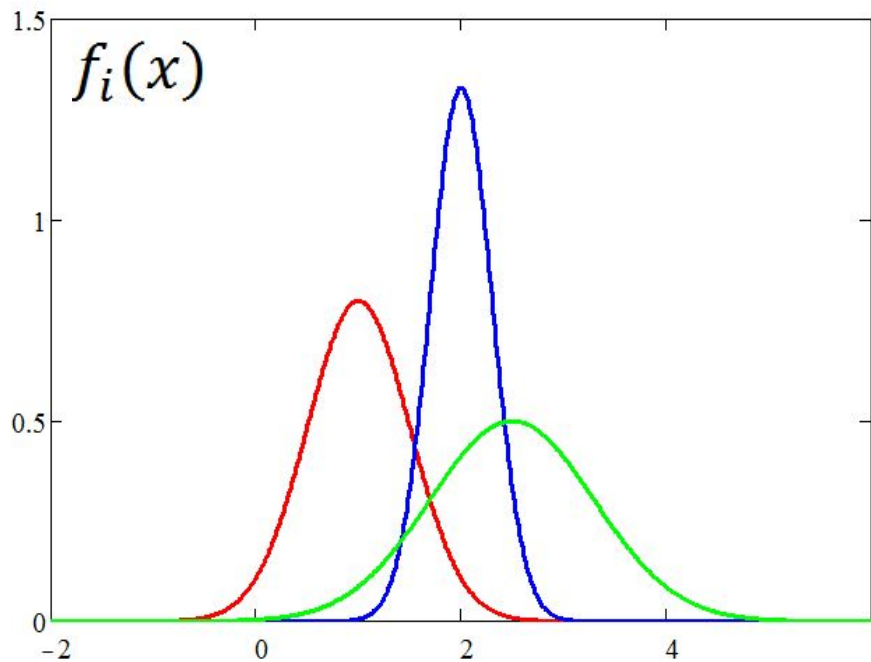
C –
нормировочная
константа



Модель смеси распределений

$$g_{\xi}(x) = \sum_{i=1}^Q p_i f_i(x)$$

$$\sum_{i=1}^Q p_i = 1$$

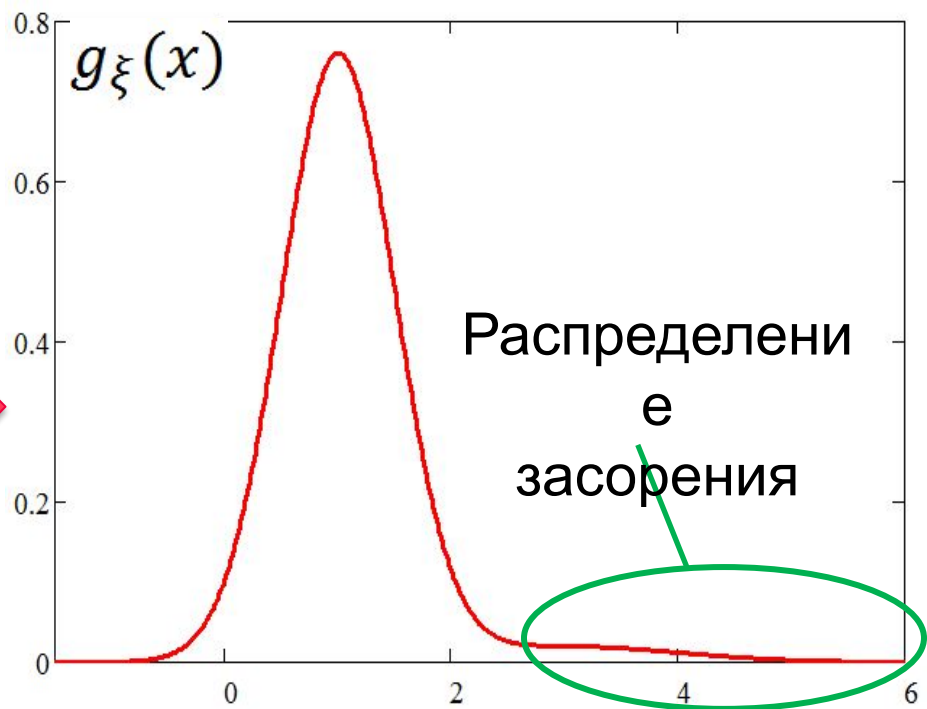
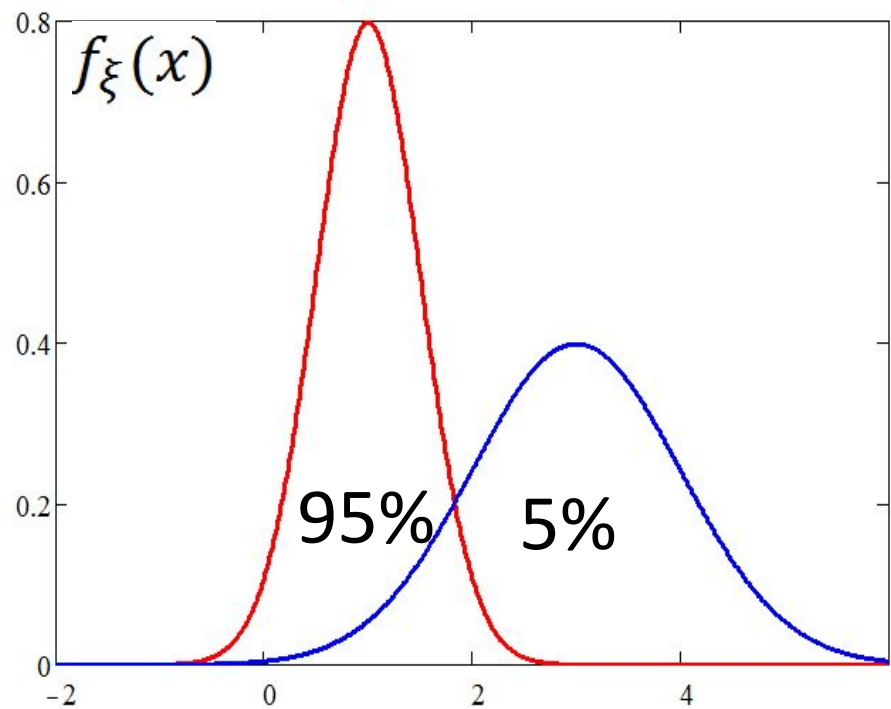


Модель распределения с засорением



$$g_{\xi}(x) = (1 - \varepsilon)f_{\xi}(x, \Theta) + \varepsilon f_{\xi}(x, m\Theta)$$

m - параметр масштаба засорения



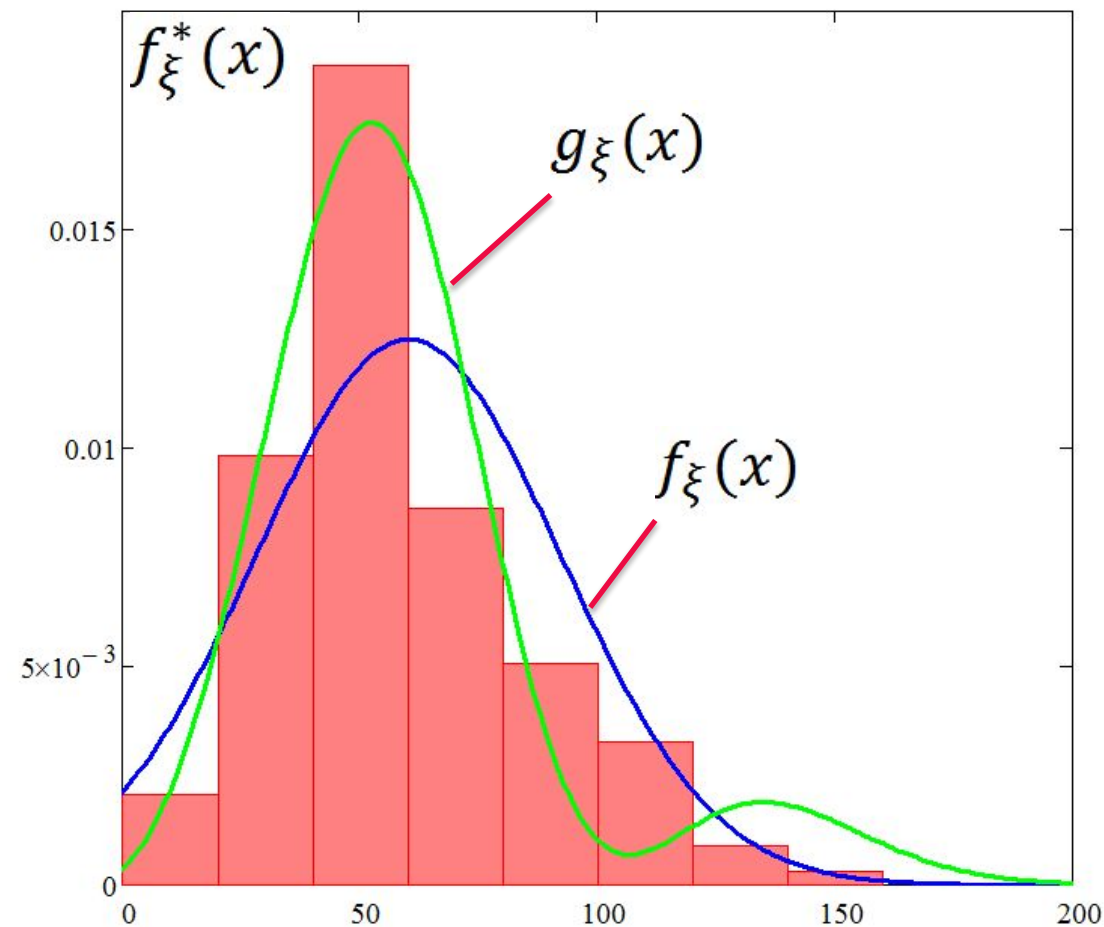
Модель в форме ряда Эджворта



Разложение в окрестности
модельного

$$g_{\xi}(x) = f_{\xi}(x) \sum_{j=1}^n a_j \varphi_j(x)$$

$$a_j = f_j(m_1, \dots, m_n)$$



Обобщение: конструктивные распределения

- **Метод МНК:** удачный механизм для «натягивания» моделей распределений на реальные данные
- **Механизмы управления детализацией модели распределения:** выбор весов и расположения квантилей для МНК.
- **Составные модели распределений** (усечение, склейка): применение МНК к разным частям вариационного ряда
- **Модели на основе смесей:** просто использовать, сложно оценивать
- **Разложения в ряд Эджворта:** эффективно, но лишь при малых отклонениях

Автор признателен всем котикам
(и хозяевам) за возможность
некоммерческого использования
их изображений, размещенных
в публичном Интернете



Александр Валерьевич Бухановский
boukhanovsky@mail.ifmo.ru

IT'sMOre than a
UNIVERSITY



УНИВЕРСИТЕТ ИТМО

ЗАНЯТИЕ 3

ТЕОРИЯ И ПРАКТИКА СТАТИСТИЧЕСКИХ ВЫВОДОВ

Раздел 5. Практическая реализация с
использованием инструментов компьютерной
математики

Александр Валерьевич Бухановский, Анна Владимировна Калюжная



1. MATLAB



2. R



3. Python



4. MS Excel



5. Statistica



Интервальное оценивание на основе Монте-Кар

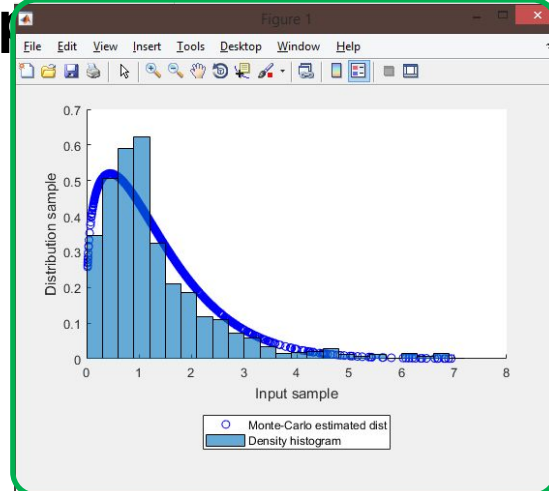
В MATLAB

```

1 clear,clc,close all
2 %% Open the file
3 waves = dlmread('dataset.txt');
4 p = mle('weib',waves);
5 pd = wblpdf(waves,p(1),p(2));
6 a=0;
7 b=max(waves);
8 %% Geometric method for generating samples
9 N = 100;
10 M = 100;
11 X = zeros(N, M);
12 for i = 1:N
13     for j = 1:M
14         y1=rand;
15         y2=rand;
16         y1m=a+(b-a)*y1;
17         y2m=max(pd)*y2;
18         while y2m>wblpdf(y1m,p(1),p(2))
19             y1=rand;
20             y2=rand;
21             y1m=a+(b-a)*y1;
22             y2m=max(pd)*y2;
23         end
24         X(i,j) = y1m;
25     end
26 end
    
```

```

27 %% Point estimation of lambda and k
28 point_est = zeros (M,2);
29 for i=1:M
30     phat = mle('weib',X(:,i));
31     point_est(i,1) = phat(1);
32     point_est(i,2) = phat(2);
33 end
34 %% Interval estimation by quantiles 2.5% and 97.5%
35 lambda_q1 = quantile(point_est(:,1)',0.025);
36 lambda_q2 = quantile(point_est(:,1)',0.975);
37 k_q1 = quantile(point_est(:,2)',0.025);
38 k_q2 = quantile(point_est(:,2)',0.975);
39 l = mean(point_est(:,1));
40 k = mean(point_est(:,2));
41 array_k = [k_q1, k, k_q2];
42 array_l = [lambda_q1,l,lambda_q2];
43 Y = wblpdf(waves,l,k);
44 figure(1)
45 hold on
46 plot(waves,Y,'bo')
47 h = histogram(waves,'Normalization','pdf');
48 ylabel('Distribution sample')
49 xlabel('Input sample')
50 legend({'Monte-Carlo estimated dist','Density histogram'},'Location','southoutside');
51 %% Tabel of results
52 T1 = array2table(array_l, 'VariableNames',{'lambda1','Lambda','Lambda2'});
53 display(T1);
54 T2 = array2table(array_k, 'VariableNames',{'k1','k','k2'});
55 display(T2);
56
    
```



Command Window

New to MATLAB? See resources for [Getting Started](#).

T1 =

lambda1	Lambda	Lambda2
1.2157	1.4161	1.6441

T2 =

k1	k	k2
1.1044	1.2924	1.5805



Тест Колмогорова-Смирнова в MATLAB

```
1
2 %% Чтение данных из файла
3 waves = dlmread('dataset.txt');
4 phat = mle('weib', waves);
5 y = zeros(1485, 2);
6 for i=1:1485
7     y(i, 1) = waves(i);
8     y(i, 2) = wblcdf(waves(i), 1, 1);
9 end
10 %% Проверка гипотезы о распределении
11 %% Вейбулла критерием Колмогорова-Смирнова
12 H = kstest(waves, y);
13 if H == 1
14     disp('Нулевая гипотеза отвергается')
15 else
16     disp('Нулевая гипотеза может быть принята')
17 end
```

Command Window

New to MATLAB? See resources for [Getting Started](#).

```
Нулевая гипотеза отвергается
fx >>
```

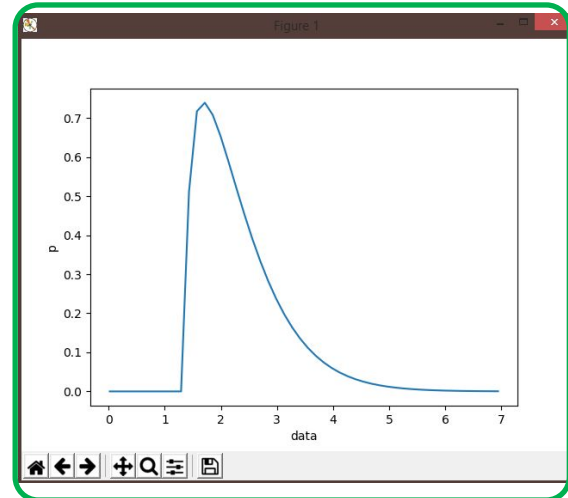
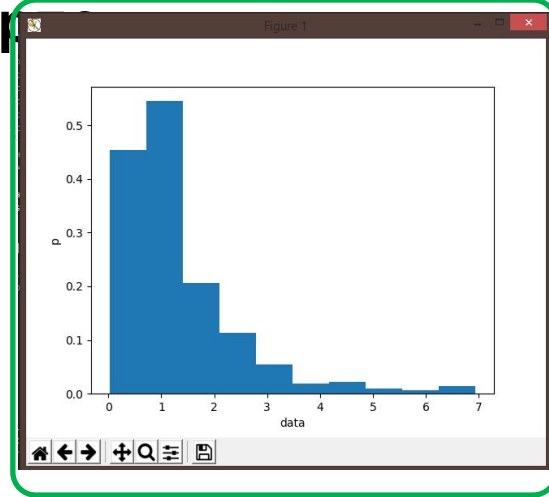
Интервальное оценивание на основе Монте-Кар

в Python

```
File Edit View Project Build Debug Team Tools Test Analyze Win
Debug Any CPU Attach
fit_distr.py Lecture_3.py* x
# Импорт пакетов
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats
import seaborn as sns
import random

# Чтение значений случайной величины из файла
data = np.loadtxt('dataset.txt')
a = 0
b = max(data)
params = scipy.stats.weibull_min.fit(data)
pdf = scipy.stats.weibull_min.pdf(data, *params)
cdf = scipy.stats.weibull_min.cdf(data, *params)
X = []
for i in range(100):
    X.append([])
    for j in range(100):
        y1 = random.random()
        y2 = random.random()
        y1m = a + (b-a)*y1
        y2m = max(pdf)*y2
        while y2m > scipy.stats.weibull_min.pdf(y1m,*params):
            y1 = random.random()
            y2 = random.random()
            y1m = a + (b-a)*y1
            y2m = max(pdf)*y2
        X[i].append(y1m)
```

```
ki = []
li = []
for i in range(100):
    params = scipy.stats.weibull_min.fit(X[i:])
    ki.append(params[0])
    li.append(params[2])
k1 = np.quantile(ki,0.025)
k2 = np.quantile(ki,0.975)
l1 = np.quantile(li,0.025)
l2 = np.quantile(li,0.975)
k = np.mean(ki)
l = np.mean(li)
print("k param")
print(k1, " ", k, " ", k2)
print("lambda param")
print(l1, " ", l, " ", l2)
plt.hist(sorted(data), 10, normed=1)
plt.xlabel('data')
plt.ylabel('p')
pdf = scipy.stats.weibull_min.pdf(x, *(k,l))
plt.plot(x, pdf)
plt.show()
```



```
CAProgram Files (x86)\Microsoft Visual Studio\Shared\Python36_64\python
k param
1.2438805598748481 1.2568268668234348 1.3141535399463202
lambda param
1.3547662100147817 1.4001340128978037 1.4230322081626214
```



Тест Колмогорова-Смирнова в Python

```
File Edit View Project Build Debug Team Tools Test Analyze
Debug Any CPU
fit_distr.py Lecture_3.py*
# Импорт пакетов
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats
import seaborn as sns
import random

# Чтение значений случайной величины из файла
data = np.loadtxt('dataset.txt')
```

```
k = np.mean(ki)
l = np.mean(li)
print("k param")
print(k1, " ", k, " ", k2)
print("lambda param")
print(l1, " ", l, " ", l2)

x = np.linspace(np.min(data), np.max(data))
print(scipy.stats.kstest(data, 'norm'))
```

```
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python36_64\python.exe
k param
1.2255553775709278 1.2450224347644534 1.2704436286733507
lambda param
1.3778728553729875 1.3933293279795977 1.4307822839498954
KstestResult(statistic=0.535224978668905, pvalue=0.0)
Press any key to continue . . .
```

Метод наименьших квадратов в Statistica(1/2)

The screenshot displays the Statistica software interface during a nonlinear regression analysis. The main window shows a data table with a 'Category' column and a 'Mean' column. The 'Basic Statistics and Tables' dialog is open, with 'Frequency tables' selected. The 'Advanced Models' menu is open, highlighting 'Nonlinear Estimation'. The 'Nonlinear Estimation' dialog is open, showing the function to be estimated: $y = v_2 + b_3 / (\sqrt{2\pi}) \cdot b_1 \cdot v_1 \cdot \exp(-1) \cdot (\log(v_1 - b_2))^2 / (2 \cdot b_1^2)$. The 'User-Specified Regression, Least Squares' dialog is also open, showing the same function. A graph displays the fitted function and observed data points. The 'Summary: Parameter estimates' dialog is open, showing the fitted 2D function and observed values.

Basic Statistics and Tables: P1 in ...

Quick | Descriptive statistics | Correlation matrices | t-test, independent, by groups | t-test, independent, by variables

Variables: Mean

Quick | Advanced | Robust | Normal

Summary: Statistics

Frequency tables

Advanced Models

General Linear

Generalized Linear/Nonlinear

Stepwise Model Builder

General Regression

General Partial Least Squares

NIPALS

Variance Components

Survival

Cox Proportional Hazards

Nonlinear Estimation

Fixed Nonlinear Regression

Log-Linear

Time Series/Forecasting

Structural Equation

Nonlinear Estimation: P1 in 2012-Obs-WW3... ?

Quick | User-specified regression, least squares | User-specified regression, custom loss function

User-Specified Regression, Least Squares: P1... ?

Quick | Function to be estimated

Function: $y = v_2 + b_3 / (\sqrt{2\pi}) \cdot b_1 \cdot v_1 \cdot \exp(-1) \cdot (\log(v_1 - b_2))^2 / (2 \cdot b_1^2)$

Model: $y = v_2 + b_3 / (\sqrt{2\pi}) \cdot b_1 \cdot v_1 \cdot \exp(-1) \cdot (\log(v_1 - b_2))^2 / (2 \cdot b_1^2)$

$y = (1319.35) / (\sqrt{2\pi}) \cdot (0.672724) \cdot x \cdot \exp(-1) \cdot (\log(x) - (0.131554))^2 / (2 \cdot (0.672724)^2)$

Frequency table: Hs

Category	Mean	Count
1.000000 < x <= 0.000000	0.5	738

Quick | Advanced | Residuals | Review

Summary: Parameter estimates

Predicted values, Residuals, etc.

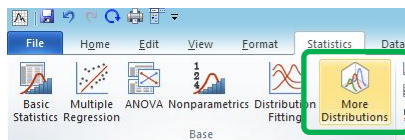
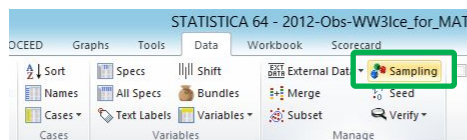
Iteration history

Analysis of Variance

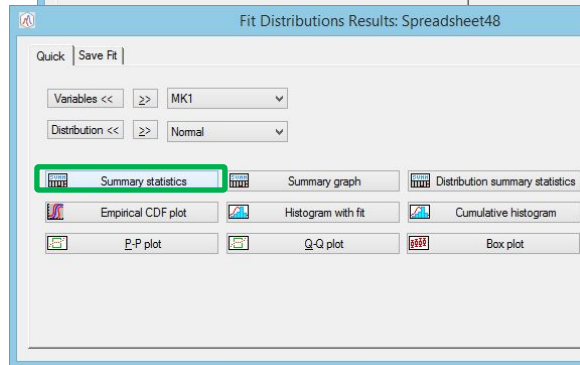
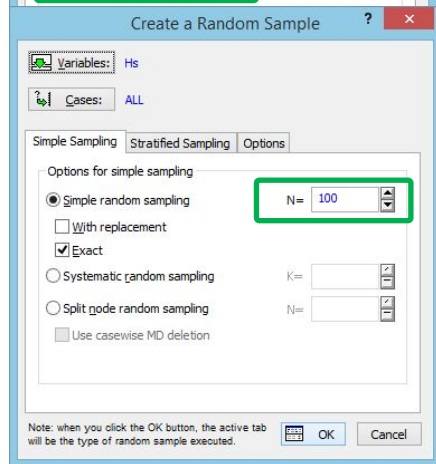
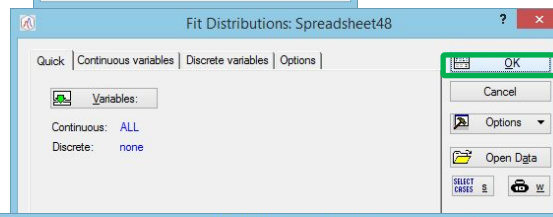
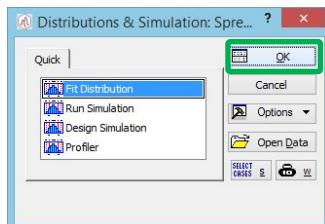
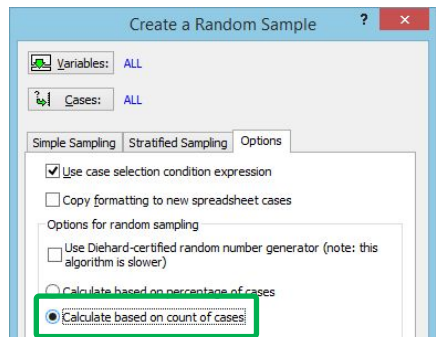
Fitted 2D function & observed vals

Fitted 3D function & observed vals

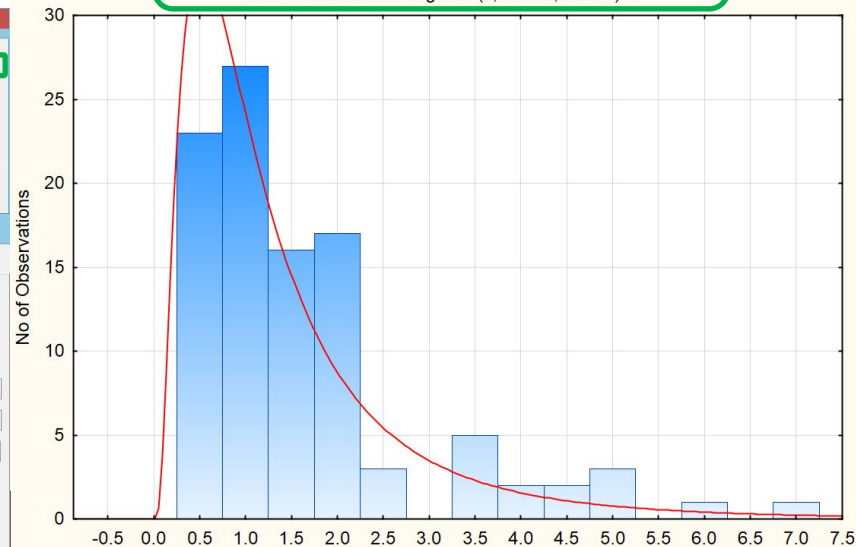
Метод Монте-Карло в Statistica(2/2)



Variable	Mean	Confidence -95.000%	Confidence 95.000%	Lower Quartile	Upper Quartile	Std.Dev.	Confidence SD -95.000%	Confidence SD +95.000%
MK1	1.324220	1.090281	1.558159	0.544500	1.731500	1.179000	1.035170	1.369619
MK2	1.149380	0.962073	1.336687	0.593000	1.377500	0.943984	0.828825	1.096603
MK3	1.324220	1.090281	1.558159	0.544500	1.731500	1.179000	1.035170	1.369619
MK4	1.554220	1.285686	1.822754	0.722500	1.911500	1.353349	1.188250	1.572152
MK5	1.439780	1.184912	1.694648	0.579000	1.779500	1.284479	1.127781	1.492147



Histogram of MK1
 $MK1 = 100 * 0.5000 * \text{Lognorm}(x, -0.0710, 0.8777)$
 Histogram of MK2
 $MK2 = 100 * 0.5000 * \text{Lognorm}(x, -0.1518, 0.8667)$
 Histogram of MK3
 $MK3 = 100 * 0.5000 * \text{Lognorm}(x, -0.0797, 0.9200)$
 Histogram of MK4
 $MK4 = 100 * 0.5000 * \text{Lognorm}(x, 0.0859, 0.9214)$
 Histogram of MK5
 $MK5 = 100 * 0.5000 * \text{Lognorm}(x, 0.0295, 0.8258)$



Проверка статистических гипотез в Excel (1/1)

excel-examples.xlsx - Excel (Сбой активации продукта)

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Надстройки Acrobat Team Что вы хотите сделать?

Буфер обмена Шрифт Выравнивание Число Стили

Гиперссылка 3 Обычный Обычный 3 Обычный

E20 : X ✓ fx =F.ТЕСТ(B2:B102;H3:H173)

Выборка	Характеристика	Значение	Логнорм. распределение
0,36	Выборочное среднее	0,63	Значение
0,32	Выборочная дисперсия	0,21	Вероятность
0,28	Медиана	0,54	0,05
0,24	Выборочное СКО	0,46	0,10
0,23			0,15
0,23			0,20
0,19			0,25
0,16			0,30
0,15			0,35
0,16			0,40
0,17			0,45
0,22			0,50
0,25			0,55
0,24			0,60
0,20			0,65
0,17			0,70
0,14			0,75
0,12			0,80
0,14			0,85
0,60			0,90
			0,95

Интервал для параметра	Нижняя граница	Верхняя граница
Доверительный интервал для среднего	0,53	0,72
Доверительный интервал для дисперсии	0,16	0,29

Характеристика	Значение
Мат. ожидание для распределения Гаусса	-0,86
СКО для распределения Гаусса	0,66

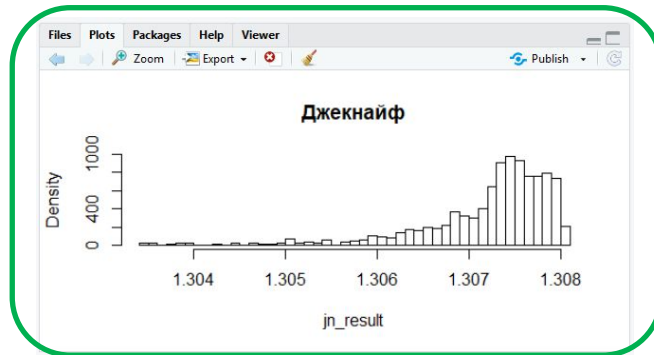
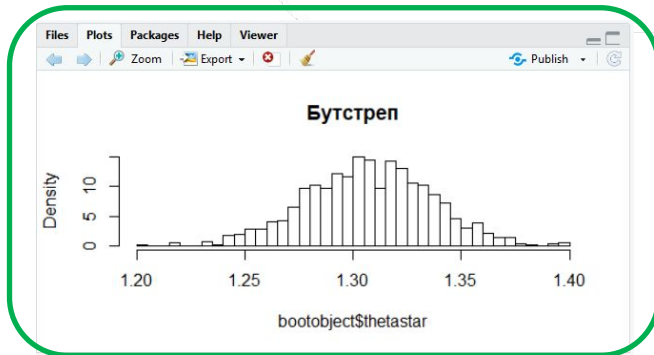
Название теста для проверки гипотезы	Значение	H0
Тест Стьюдента	1,94421E-32	ОТВЕРГНУТА
F-тест	1,87195E-39	ОТВЕРГНУТА

Логнорм. распределение	Значение	Вероятность
0,05	0,06	
0,10	0,53	
0,15	1,15	
0,20	1,57	
0,25	1,75	
0,30	1,76	
0,35	1,66	
0,40	1,51	
0,45	1,35	
0,50	1,18	
0,55	1,02	
0,60	0,88	
0,65	0,76	
0,70	0,65	
0,75	0,56	
0,80	0,48	
0,85	0,41	
0,90	0,35	
0,95	0,30	

Интервальное оценивание на основе Монте-Карло в R

(1/2)

```
stat3.R x
Source on Save
4
5 library(bootstrap)
6 library(fitdistrplus)
7 library(goftest)
8
9 #Чтение значений случайное величины
10 real_sample=read.table("X.csv",sep=";",header=TRUE)$x
11
12 #функция вычисления среднего по выборке
13 meanfun <- function(data, indices){
14   d <- data[indices]
15   return(mean(d))
16 }
17
18 #Моделирование распределения оценки мат. ожидания методом бутстреп
19 br_result = bootstrap(x=real_sample,nboot=1000,theta=meanfun)$thetastar
20 hist(br_result,breaks = 50,freq = FALSE,main="")
21 title("Бутстреп")
22 pint=quantile(br_result,c(0.025,0.975))
23 c(pint[1],mean(br_result),pint[2])
24
25
26 #Моделирование распределения оценки мат. ожидания методом джекнайф
27 jn_meanfun <- function(d){
28   return(mean(d))
29 }
30 jn_result=jackknife(real_sample,jn_meanfun)$jack.values
31 hist(jn_result,breaks = 50,freq = FALSE,main="")
32 title("Джекнайф")
33 pint=quantile(jn_result,c(0.025,0.975))
34 c(pint[1],mean(jn_result),pint[2])
```



```
> c(pint[1],mean(br_result),pint[2])
2.5% 97.5%
1.252100 1.307469 1.368992
> c(pint[1],mean(jn_result),pint[2])
2.5% 97.5%
1.252100 1.307207 1.368992
> |
```

Проверка статистических гипотез в R (2/2)

```
stat3.R x
32
33 #функция оценки параметров для логнормального распределения на основе метода максимального правдоподобия
34 theor_dist_fit_mle=fitdist(data=real_sample,distr="lnorm",method="mle")
35 #на основе метода квантилей
36 theor_dist_fit_qme=qmedist(data=real_sample,distr="lnorm",probs=c(0.25, 0.75))
37
38 #Проверка гипотезы методом Крамера-Мизеса
39 cvm_res=cvm.test(real_sample,"plnorm",mean=theor_dist_fit_mle$estimate[1],
40                 sd=theor_dist_fit_mle$estimate[2])
41 cvm_res
42
43 #Проверка гипотезы методом Андерсена
44 ad_res=ad.test(real_sample,"plnorm",mean=theor_dist_fit_mle$estimate[1],
45               sd=theor_dist_fit_mle$estimate[2])
46 ad_res
47
48 #Проверка гипотезы критерием Колмогорова-Смирнова
49 ks_res=ks.test(real_sample,"plnorm",mean=theor_dist_fit_mle$estimate[1],
50               sd=theor_dist_fit_mle$estimate[2])
51 ks_res$statistic
```

```
> cvm_res
```

```
Cramer-von Mises test of goodness-of-fit
Null hypothesis: log-normal distribution
```

```
data: real_sample
omega2 = 1.526, p-value = 0.0001498
```

```
> ad_res
```

```
Anderson-Darling test of goodness-of-fit
Null hypothesis: log-normal distribution
```

```
data: real_sample
An = 8.1689, p-value = 9.216e-05
```

```
> ks_res
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: real_sample
D = 0.068686, p-value = 1.644e-06
alternative hypothesis: two-sided
```

```
Console Terminal x
D:/Projects/Lectures/ЦГ/Статистика-3/
> if (cvm_res$statistic>=qCvm(c(0.95))) {
+   print("гипотеза H0 отклонена")
+ }
[1] "Гипотеза H0 отклонена"
```

```
> print(paste("Омега-квадрат =",round(ad_res$statistic,2),"95% квантиль распределения =",round(qAD(c(0.95)),2)))
[1] "Омега-квадрат = 8.17 95% квантиль распределения = 2.49"
```

```
> if (ad_res$statistic>=qAD(c(0.95))) {
+   print("Гипотеза H0 отклонена")
+ }
[1] "Гипотеза H0 отклонена"
```

```
> print(paste("p-value =",round(ks_res$p.value,6)))
[1] "p-value = 2e-06"
```

```
> if (ks_res$p.value<0.05) {
+   print("Гипотеза H0 отклонена")
+ }
[1] "Гипотеза H0 отклонена"
```