Домашнее задание 1

Задание 1

a) По формуле ковариации имеем: $cov(\hat{F}_n(x), \hat{F}_n(y)) = \mathrm{E}[\hat{F}_n(x)\hat{F}_n(y)] - \mathrm{E}\hat{F}_n(x)\mathrm{E}\hat{F}_n(y)$

Поскольку $\hat{F}_n$ − эмпирическая функция распределения, то по ее свойству: $\mathrm{E}\hat{F}_n(x) = F(x)$, $\mathrm{E}\hat{F}_n(y) = F(y)$.

$$\mathrm{E}[\hat{F}_n(x)\hat{F}_n(y)] = \text{[по определению эмпирической ф.р.]} = \mathrm{E}\left[\frac{1}{n^2}\sum_{i,j}^{n} I\{X_i \le x, X_j \le y\}\right] = \text{[линейность МО]} =$$
$$= \frac{1}{n^2}\sum_{i,j}^{n}\mathrm{E}I\{X_i \le x, X_j \le y\} = \frac{1}{n^2}\sum_{i\ne j}^{n}P(X_i \le x, X_j \le y) + \frac{1}{n^2}\sum_{i=j}^{n}P(X_i \le x, X_j \le y) = \text{[т.к. } X_i \text{ независимы]} =$$
$$= \frac{1}{n^2}\sum_{i\ne j}^{n}P(X_i \le x)P(X_j \le y) + \frac{1}{n^2}\sum_{i=1}^{n}P(X_i \le \min(x,y)) = \frac{n(n-1)}{n^2}F(x)F(y) + \frac{n}{n^2}F(\min(x,y)) = \frac{(n-1)F(x)F(y) + F(\min$$

Таким образом, искомая $cov(\hat{F}_n(x), \hat{F}_n(y)) = \dfrac{(n-1)F(x)F(y) + F(\min(x,y))}{n} - F(x)F(y) = \dfrac{F(\min(x,y)) - F(x)F(y)}{n}$

b) По свойствам билинейности и симметричности ковариации для любых двух случайных величин X и Y:

$\mathrm{V}[X-Y] = cov(X-Y, X-Y) = cov(X,X) + cov(X,-Y) + cov(-Y,X) + cov(Y,Y) = \mathrm{V}X + \mathrm{V}Y - 2cov(X,Y)$

Поскольку $\hat{F}_n$ − эмпирическая функция распределения, то по ее свойству: $\mathrm{V}\hat{F}_n(x) = \dfrac{F(x)(1-F(x))}{n}$ (★)

Тогда

$$\mathrm{V}\hat{\theta} = \mathrm{V}[\hat{F}_n(a) - \hat{F}_n(b)] = \mathrm{V}\hat{F}_n(a) + \mathrm{V}\hat{F}_n(b) - 2cov(\hat{F}_n(a), \hat{F}_n(b)) = \text{[по пункту а) (здесь } a < b\text{) и (★)]} =$$
$$= \frac{F(a)(1-F(a))}{n} + \frac{F(b)(1-F(b))}{n} - \frac{2F(a)(1-F(b))}{n} =$$
$$= \frac{F(b) - F(a) - (F(b)-F(a))^2}{n} = \frac{(F(b)-F(a))(1-(F(b)-F(a)))}{n} = \frac{\theta(1-\theta)}{n}$$

Значит, $se(\hat{\theta}) = \sqrt{\mathrm{V}\hat{\theta}} = \sqrt{\dfrac{\theta(1-\theta)}{n}}$.

Заметим, что $T(F)$ - линейный статистический функционал, т.к. $T(kF) = kF(a) - kF(b) = kT(F)$ и $T(F+G) = F(a) + G(a) - (F(b) + G(b)) = T(F) + T(G)$. Поэтому $T(\hat{F}_n) \approx \mathcal{N}(T(F), \hat{se}^2)$, где $\hat{se}$ − оценка для стандартного отклонения $T(\hat{F}_n)$, и доверительный интервал для $T(F)$ с доберительной вероятностью $1-\alpha$ будет иметь вид:

$$T(\hat{F}_n) \pm z_{\alpha/2}\hat{se}$$

В данном случае по посчитанному ранее получим, что приближенный доверительный интервал размера $1-\alpha$ для θ имеет вид:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

c)

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline

X = pd.read_csv('fijiquakes.dat', delim_whitespace=True)
X.head()
```

Out[1]:

	Obs.	lat	long	depth	mag	stations
0	1	-20.42	181.62	562	4.8	41
1	2	-20.62	181.03	650	4.2	15
2	3	-26.00	184.10	42	5.4	43
3	4	-17.97	181.66	626	4.1	19
4	5	-20.42	181.96	649	4.0	11

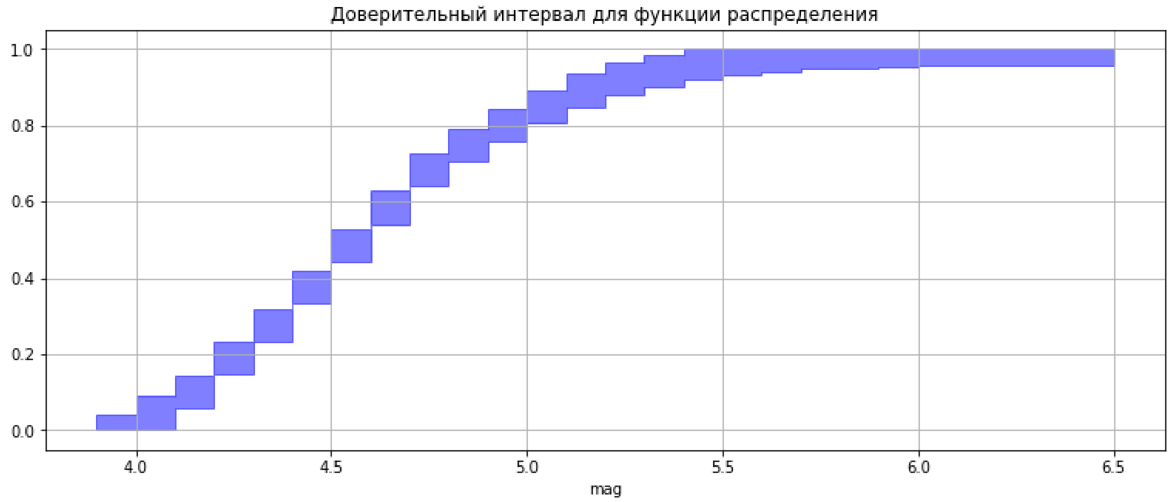
In [2]:

```
y_un = np.unique(np.array(X["mag"]), return_counts=True)
mag = y_un[0]  # отсортированные значения амплитуд
num = y_un[1].cumsum()/y_un[1].sum()  # значения эмпирической ф.р. в точках из mag
# добавим крайние точки
mag = np.insert(mag.tolist(), [0, y_un[0].size], [mag.min() - 0.1, mag.max() + 0.1]
num = np.insert(num.tolist(), [0, y_un[0].size], [0, 1])
# продублируем точки, чтобы график стал кусочно-гладким
mag = np.repeat(mag, repeats = 2)
num = np.repeat(num, repeats = 2)
mag = np.delete(mag, 0)
num = np.delete(num, num.size - 1)

f, (t) = plt.subplots(1, 1, figsize=(13, 5))
t.plot(mag, num)
t.set_title("Эмпирическая функция распределения амплитуды")
t.set_xlabel(u"mag")
t.grid()
```

Построим приближенный 95% доверительный интервал для F по следствию из неравенства Дворецкого-Кифера-Вольфовица.

In [3]:

```
from math import log

epsn = (0.5/X["mag"].size * log(2/0.05)) ** 0.5
Lx = np.where(num - epsn > 0, num - epsn, np.zeros(mag.size))
Ux = np.where(num + epsn < 1, num + epsn, np.ones(mag.size))
f, (t) = plt.subplots(1, 1, figsize=(13, 5))
t.fill_between(mag, Lx, Ux, color='blue', alpha=0.5)
t.set_title("Доверительный интервал для функции распределения")
t.set_xlabel(u"mag")
t.grid()
```



Подсчитаем и построим приближенный 95% доверительный интервал для значения $F(4.9) - F(4.3)$, используя формулу из пункта b.

In [21]:

```
from scipy import stats as st

b = 4.9
a = 4.3
th = num[mag == b][1] - num[mag == a][1]
se = (th * (1 - th)/X["mag"].size)**.5
z = st.norm.ppf(.025, 0, 1)
print("Приближенный доверительный интервал для F(4.9) – F(4.3): [" \
      + str(round(th + z * se, 4)) + ", " + str(round(th - z * se, 4)) + "]")
```

```
Приближенный доверительный интервал для F(4.9) – F(4.3): [0.4951, 0.55
69]
```

Задание 2

Рассмотрим выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$, где $n = 100$, соответсвующие двум группам людей. Будем считать, что X_j (или Y_j) принимает значение 1, если j-ый пациент в данной группе выздоровел, иначе $-\ 0$. Тогда $X_i\ -$ независимые одинаково распределенные случайные величины из распределения Бернулли с параметром p_1, а $Y_i\ -\ i.i.d.$ из Бернулли с параметром p_2.

Получим оценки $\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{90}{100} = 0.9$ и $\hat{p}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{85}{100} = 0.85$

Поэтому искомая оценка для $\theta = p_1 - p_2$ будет $\hat{\theta} = \hat{p}_1 - \hat{p}_2 = 0.9 - 0.85 = 0.05$

Найдем для нее стандартную ошибку.

$\mathrm{V}\hat{p}_1 =$ [из независимости X_i] = $\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}X_i = \frac{1}{n^2}\sum_{i=1}^{n}p_1(1-p_1) = \frac{1}{n^2}np_1(1-p_1) = \frac{p_1(1-p_1)}{n}$

Аналогично $\mathrm{V}\hat{p}_2 =$ [из независимости Y_i] $= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}Y_i = \frac{1}{n^2}\sum_{i=1}^{n}p_2(1-p_2) = \frac{1}{n^2}np_2(1-p_2) = \frac{p_2(1-p_2)}{n}$

Откуда $\mathrm{V}\hat{\theta} = \mathrm{V}[\hat{p}_1 - \hat{p}_2] =$ [из независимости p_1 и p_2] =
$\mathrm{V}\hat{p}_1 + \mathrm{V}\hat{p}_1 = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{n} = \frac{p_1(1-p_1) + p_2(1-p_2)}{n}$

Искомая $se = \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{n}}$, ее оценка

$\hat{se} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)}{n}} = \sqrt{\frac{0.9 * 0.1 + 0.85 * 0.15}{100}} \approx 0.047$

Согласно центральной предельной теореме $\hat{\theta} \approx \mathcal{N}(\theta, \hat{se}^2)$. Тогда приближенный доверительный интервал для θ с доверительной вероятностью $1 - \alpha$ имеет вид:

$$\hat{\theta} \pm z_{\alpha/2}\hat{se}$$

Для $\alpha = 0.2$ (вероятности 80%) получим приближенный доверительный интервал $[-0.0098, 0.1098]$

Для $\alpha = 0.05$ (вероятности 95%) получим приближенный доверительный интервал $[-0.0414, 0.1414]$

In [24]:

```
#Ниже приведены вычисления, использовавшиеся в задании

se = ((0.9*0.1 + 0.85*0.15)/100)**.5
round(se, 3)
```

Out[24]:

```
0.047
```

In [25]:

```
from scipy import stats as st
z1 = st.norm.ppf(.1, 0, 1)
z2 = st.norm.ppf(.025, 0, 1)
th = 0.05
print ("80% доверительный интервал: [" + str(round(th + z1 * se, 4)) + ", " + str(r
print ("95% доверительный интервал: [" + str(round(th + z2 * se, 4)) + ", " + str(r
```

```
80% доверительный интервал: [-0.0098, 0.1098]
95% доверительный интервал: [-0.0414, 0.1414]
```

Задание 3

In [57]:

```
import numpy as np

N = 50
standard_data = np.random.normal(0, 1, N)
log_data = np.exp(standard_data)
log_data
```

Out[57]:

```
array([ 0.43457146,  0.05311461,  5.39610666,  0.42602509,  0.4043529
4,
        5.38791756,  1.66196525,  0.2640057 ,  0.86252361,  1.1677115
5,
        0.13337221,  1.01827117,  0.29508236,  1.28660986,  1.2613008
5,
        0.94629758,  1.21110126,  0.59527912,  0.16426217,  0.5811718
2,
        3.37638508,  0.14159539,  0.21260484,  0.65010659,  3.8342538
6,
        0.44928088,  0.39852588,  0.20408565,  3.06435576,  1.7005716
,
        0.79885349,  1.09512004,  0.32625981,  1.09175708,  1.7787966
8,
        0.51617873,  0.48962882,  0.78679329,  0.83781214,  1.2615710
3,
        1.17470884,  2.57579925,  2.66966812,  2.84936148,  3.5261733
5,
        0.42964436,  1.40215343,  1.12320859,  4.42340928,  0.5094381
3])
```

Заметим, что поскольку $\theta = T(F) = \int r(x)dF(x)$ - линейный статистический функционал, то $\hat{\theta} = T(\hat{F}_n) = \int r(x)d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} r(X_i)$, где $X_1, \ldots, X_n \ - \ i.i.d$ выборка из распределения F с эмпирической функцией распределения F_n. Здесь $r(x) = \frac{(x-\mu)^3}{\sigma^3}$, $F \ - \$ логнормальное распредление. Используем этот факт для построения оценки на каждом шаге бутстрепа.

Для построения доверительных интервалов применим три подхода на остнове бустрепа:

$1)$ Эфронов доверительный интервал

Для бутстрепа используем статистику $\hat{\theta}$. В качестве квантилей q_1^* и q_2^* возьмем полученные порядковые статистики $\hat{\theta}^*_{[B\alpha/2]}$ и $\hat{\theta}^*_{[B(1-\alpha/2)+1]}$ соответсвенно. Получим интервал $[q_1^*, q_2^*]$.

$2)$ Квантильный (рецентрированный) доверительный интервал

Для бутстрепа используем статистику $\hat{\theta} - \theta$, у которой бутстреповским аналогом является $\hat{\theta}^* - \hat{\theta}$. В качестве квантилей $q_1^{*\%}$ и $q_2^{*\%}$ возьмем полученные порядковые статистики с номерами $[B\alpha/2]$ и $[B(1-\alpha/2)+1]$ соответственно. Получим интервал $[\hat{\theta} - q_2^{*\%}, \hat{\theta} - q_1^{*\%}]$.

$3)$ Т-квантильный доверительный интервал

Для бутстрепа используем статистику $\dfrac{\hat{\theta} - \theta}{se(\hat{\theta})}$, у которой бутстреповским аналогом является $\dfrac{\hat{\theta}^* - \hat{\theta}}{se^*(\hat{\theta})}$. В качестве квантилей $q_1^{*\%t}$ и $q_2^{*\%t}$ возьмем полученные порядковые статистики с номерами $[B\alpha/2]$ и $[B(1-\alpha/2)+1]$ соответственно. Получим интервал $[\hat{\theta} - se(\hat{\theta})q_2^{*\%t}, \hat{\theta} - se(\hat{\theta})q_1^{*\%t}]$.

В данном методе также используем, что в рассматриваем случае

$$se(\hat{\theta}) = \sqrt{\mathrm{V}\hat{\theta}} = \sqrt{\mathrm{V}\sum_{i=1}^{n}\frac{(X_i-\mu)^3}{\sigma^3}} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}\frac{(X_i-\mu)^3}{\sigma^3}} = \sqrt{\frac{1}{n^2}n\mathrm{V}\frac{(X_1-\mu)^3}{\sigma^3}} = \frac{1}{\sqrt{n}}se(\frac{(X_1-\mu)^3}{\sigma^3})$$

где в качестве оценки последней стандартной ошибки можно использовать выборочную.

In [59]:

```
B = 50000  # фиксируем количество бутстреп-выборок
alpha = 0.05

values_e = []
values_q = []
values_tq = []

theta_est =  1/N * np.sum((log_data - log_data.mean())**3/(log_data.var()**1.5))  #
se_est = (1/N)**.5 * np.std((log_data - log_data.mean())**3/(log_data.var()**1.5))

for i in range (B):
    data = np.random.choice(log_data, N)  # генерируем бутстреп-выборку
    values_e += [1/N * np.sum((data - data.mean())**3/(data.var()**1.5))]
    values_q += [1/N * np.sum((data - data.mean())**3/(data.var()**1.5)) - theta_es
    values_tq += [(1/N * np.sum((data - data.mean())**3/(data.var()**1.5)) - theta_
                  ((1/N)**.5 * np.std((data - data.mean())**3/(data.var()**1.5)))]

values_e = np.sort(np.array(values_e))
values_q = np.sort(np.array(values_q))
values_tq = np.sort(np.array(values_tq))
Xq1 = int(B * alpha/2 - 1)
Xq2 = int(B * (1 - alpha/2))

print("Эфронов 95% доверительный интервал: [" + str(values_e[Xq1]) + ", " + str(val
print("Квантильный 95% доверительный интервал: [" + str(theta_est - values_q[Xq2])
print("Т-квантильный 95% доверительный интервал: [" + str(theta_est - se_est * valu
      ", " + str(theta_est - se_est * values_tq[Xq1]) + "]")
```

```
Эфронов 95% доверительный интервал: [0.942997741164, 2.33181313301]
Квантильный 95% доверительный интервал: [0.834310897867, 2.2231262897
2]
Т-квантильный 95% доверительный интервал: [1.14962641704, 2.7371460359
1]
```

Задание 4

а) Найдем распределение $\hat{\theta}$: $F_{\hat{\theta}}(x) = F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \ldots, X_n \le x) =$ [из независимости X_i] $= \prod_{i=1}^{n} P(X_i \le x) = F^n(x)$, где $F(x) = \dfrac{x}{\theta} I\{x \in [0, \theta]\}$. Значит, $F_{\hat{\theta}}(x) = \dfrac{x^n}{\theta^n} I\{x \in [0, \theta]\}$. Его плотность $f_{\hat{\theta}}(x) = \dfrac{nx^{n-1}}{\theta^n} I\{x \in [0, \theta]\}$

In [13]:

```
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline

N = 50
uni_data = np.random.uniform(0, 1, N)  # генерируем выборку из равномерного распред
uni_data
```

Out[13]:

```
array([ 0.04684271,  0.06994152,  0.79317903,  0.15668857,  0.8023125
6,
        0.6082226 ,  0.85319143,  0.44823225,  0.96551698,  0.9340824
1,
        0.26719604,  0.15125082,  0.87105256,  0.45652741,  0.4614093
7,
        0.99212911,  0.4097798 ,  0.1416267 ,  0.84282902,  0.0597054
1,
        0.56687538,  0.61286604,  0.18136865,  0.26374206,  0.8070036
,
        0.65970093,  0.43024124,  0.33771941,  0.38965242,  0.3102999
3,
        0.22694643,  0.61578784,  0.27993732,  0.62933888,  0.1966787
8,
        0.8610192 ,  0.96595863,  0.02456348,  0.41202564,  0.4690654
,
        0.43563249,  0.21406134,  0.91585658,  0.29462573,  0.7749411
2,
        0.29739609,  0.18105059,  0.09844718,  0.66288807,  0.8794209
3])
```

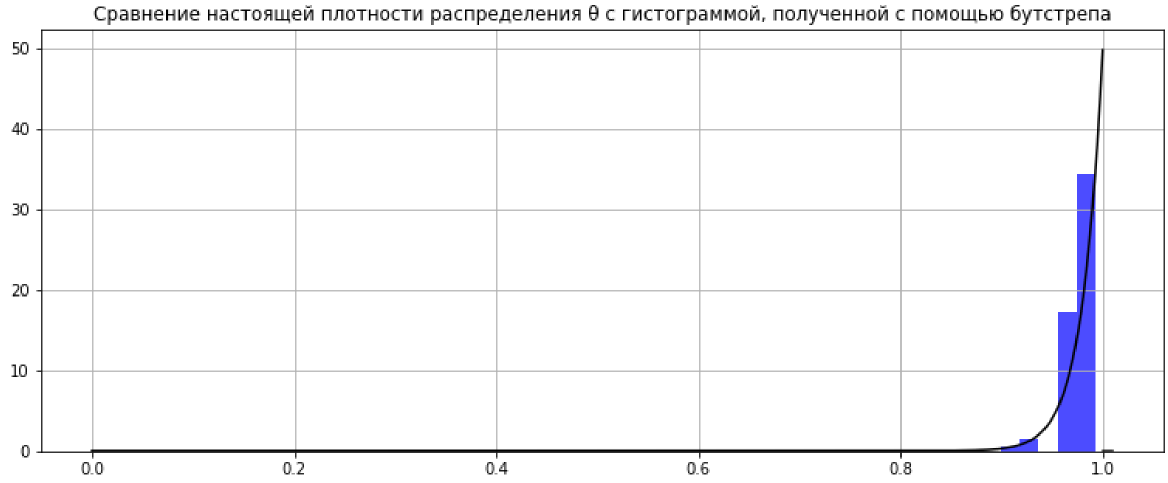
In [14]:

```
def f(x):  # определим функцию плотности
    res = N * x**(N-1)
    res[x <= 0] = 0
    res[x > 1] = 0
    return res
```

Сравним плотность настоящего распределения θ с гистограммой, полученной с помощью бутстрепа с числом выборок $B = 50000$.

In [15]:

```
B = 50000
values = []
for i in range (B):
    data = np.random.choice(uni_data, N)  # генерируем бутстреп-выборку
    values += [np.max(data)]
values = np.array(values)

xn = np.arange(0, 1, 0.0001)
(g), t = plt.subplots(1, 1, figsize=(13, 5))
t.hist(values,  color="blue", normed=True, alpha=0.7)
t.plot(xn, f(xn), color='black')
t.plot(np.array([1, 1.01]), np.array([0, 0]), color='black')
t.set_title("Сравнение настоящей плотности распределения θ с гистограммой, полученн
t.grid()
```



Выполнив несколько сравнений видим, что в большинстве случаев бутстреп работает достаточно плохо − гистограмма не заполняет график плотности, а иногда крайний правый столбец далек от 1.

b) Заметим, что поскольку распределение является непрерывным, то

$$0 \le P(\hat{\theta} = \theta) = P(X_{(n)} = \theta) \le \sum_{i=1}^{n} P(X_i = \theta) = \sum_{i=1}^{n} 0 = 0$$

и значит

$$P(\hat{\theta} = \theta) = 0$$

В то же время, если $\hat{\theta}^*$ − оценка, полученная с помощью бутстрепа, т.е. $\hat{\theta}^* = X^*_{(n)}$, то бутстреповским аналогом для $P(\hat{\theta} = \theta)$ является вероятность

$$P(\hat{\theta}^* = \hat{\theta}) = 1 - P(\hat{\theta}^* \ne \hat{\theta}) = 1 - P(X^*_{(n)} \ne \hat{\theta}) =$$

$$= 1 - P(X^*_1 \ne \hat{\theta}, X^*_2 \ne \hat{\theta}, \dots, X^*_n \ne \hat{\theta}) = \text{ [так как } X^*_i \text{ выбираются независимо] } =$$

$$= 1 - \prod_{i=1}^{n} P(X^*_i \ne \hat{\theta}) = 1 - (P(X^*_1 \ne \hat{\theta}))^n = 1 - (1 - P(X^*_1 = X_{(n)}))^n =$$

$$= \text{ [так как } X^*_i \text{ выбираются из начальной выборки равномерно] } = 1 - (1 - \frac{1}{n})^n \xrightarrow{n \to \infty} 1 - \frac{1}{e} \approx 0.632$$

(по второму замечательному пределу; при $n = 50$, как в примере, получится примерно 0.636)

Как следствие, в данном случае бутстреп работает плохо (значения вероятностей должны были быть примерно равны).

Задание 5

Заметим, что $\mathrm{V}T_n = \mathrm{V}\overline{X}_n^{-2} = \mathrm{E}\overline{X}_n^{-4} - (\mathrm{E}\overline{X}_n^{-2})^2 = \frac{1}{n^4}\mathrm{E}(\sum_{i=1}^{n} X_i)^4 - \left(\frac{1}{n^2}\mathrm{E}(\sum_{i=1}^{n} X_i)^2\right)^2 = \frac{1}{n^4}\left(\mathrm{E}(\sum_{i=1}^{n} X_i)^4 - \left(\mathrm{E}(\sum_{i=1}^{n} X_i)^2\right)^2\right)$

Найдем сначала $\mathrm{E}(\sum_{i=1}^{n} X_i)^4 = \mathrm{E}(X_1 + X_2 + \ldots + X_n)^4$

По полиномиальной теореме $(X_1 + X_2 + \ldots + X_n)^4 = \sum_{\alpha_1+\alpha_2+\ldots+\alpha_n=4} \frac{4!}{\alpha_1!\alpha_2!\ldots\alpha_n!} X_1^{\alpha_1}X_2^{\alpha_2}\ldots X_n^{\alpha_n}$

Поэтому т.к. X_i одинаково распределенные с.в., то

$$\mathrm{E}(X_1 + X_2 + \ldots + X_n)^4 = n\frac{4!}{4!}\mathrm{E}X_i^4 + n(n-1)\frac{4!}{3!1!}\mathrm{E}X_i^3X_j + \frac{n(n-1)}{2}\frac{4!}{2!2!}\mathrm{E}X_i^2X_j^2 + n\frac{(n-1)(n-2)}{2}\frac{4!}{2!1!1!}\mathrm{E}X_i^2X_jX_k + C_n^4 4$$

Индексы в каждом слагаемом различны; каждое слагаемое соответсвует фиксированному разбиению числа 4 на сумму, отраженному в степени при X_i, и выборе номеров полученных слагаемых для α_i.

Например, разбиению $2 + 1 + 1$ соответствует четвертое слагаемое (см. степени), а способов зафисировать для такого разбиения конкретные индексы α_i всего: n(выбрать индекс элемента, который берется 2 раза) $\cdot \frac{(n-1)(n-2)}{2}$ (зафиксировать произвольные два различных индекса из оставшихся, т.е. C_{n-1}^2). Аналогично подсчитано для других слагаемых (там формулы уже более простые).

Таким образом, т.к. $X_i -$ независимы:

$$\circeq n\mathrm{E}X^4 + 4n(n-1)\mathrm{E}X^3\mathrm{E}X + 3n(n-1)\mathrm{E}X^2\mathrm{E}X^2 + 6n(n-1)(n-2)\mathrm{E}X^2(\mathrm{E}X)^2 + n(n-1)(n-2)(n-3)(\mathrm{E}X)^4 =$$

$$= n\mathrm{E}X^4 + (4n^2 - 4n)\mathrm{E}X^3\mathrm{E}X + (3n^2 - 3n)\mathrm{E}X^2\mathrm{E}X^2 + (6n^3 - 18n^2 + 12n)\mathrm{E}X^2(\mathrm{E}X)^2 + (n^4 - 6n^3 + 11n^2 - 6n)(\mathrm{E}X)^4$$

Найдем теперь

$$\left(\mathrm{E}(\sum_{i=1}^{n} X_i)^2\right)^2 = \left(\mathrm{E}(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_iX_j)\right)^2 = \left(n\mathrm{E}X^2 + n(n-1)(\mathrm{E}X)^2\right)^2 = n^2\mathrm{E}X^2\mathrm{E}X^2 + 2n^2(n-1)\mathrm{E}X^2(\mathrm{E}X)^2 + n^2(n-1)^2(\mathrm{E}X$$

$$= n^2\mathrm{E}X^2\mathrm{E}X^2 + (2n^3 - 2n^2)\mathrm{E}X^2(\mathrm{E}X)^2 + (n^4 - 2n^3 + n^2)(\mathrm{E}X)^4$$

Поэтому получим

$$\mathrm{V}T_n = \frac{1}{n^4}\left(n\mathrm{E}X^4 + (4n^2 - 4n)\mathrm{E}X^3\mathrm{E}X + (2n^2 - 3n)\mathrm{E}X^2\mathrm{E}X^2 + (4n^3 - 16n^2 + 12n)\mathrm{E}X^2(\mathrm{E}X)^2 + (-4n^3 + 10n^2 - 6n)(\mathrm{E}X)^4\right) =$$

$$= \frac{1}{n^3}\mathrm{E}X^4 + \frac{4n-4}{n^3}\mathrm{E}X^3\mathrm{E}X + \frac{2n-3}{n^3}\mathrm{E}X^2\mathrm{E}X^2 + \frac{4n^2 - 16n + 12}{n^3}\mathrm{E}X^2(\mathrm{E}X)^2 + \frac{-4n^2 + 10n - 6}{n^3}(\mathrm{E}X)^4$$

Выразим теоретические центральные моменты через начальные теоретические моменты:

$$\mathrm{E}(X - \mathrm{E}X)^2 = \mathrm{E}(X^2 - 2X\mathrm{E}X + (\mathrm{E}X)^2) = \mathrm{E}X^2 - (\mathrm{E}X)^2$$

$$\mathrm{E}(X - \mathrm{E}X)^3 = \mathrm{E}(X^3 - 3X^2\mathrm{E}X + 3X(\mathrm{E}X)^2 - (\mathrm{E}X)^3) = \mathrm{E}X^3 - 3\mathrm{E}X^2\mathrm{E}X + 2(\mathrm{E}X)^3$$

$$\mathrm{E}(X - \mathrm{E}X)^4 = \mathrm{E}(X^4 - 4X^3\mathrm{E}X + 6X^2(\mathrm{E}X)^2 - 4X(\mathrm{E}X)^3 + (\mathrm{E}X)^4) = \mathrm{E}X^4 - 4\mathrm{E}X^3\mathrm{E}X + 6\mathrm{E}X^2(\mathrm{E}X)^2 - 3(\mathrm{E}X)^4$$

$v_{boot} = \frac{4\bar{X}_n^2\hat{\alpha}_2}{n} + \frac{4\bar{X}_n\hat{\alpha}_3}{n^2} + \frac{\hat{\alpha}_4}{n^3} + \frac{\hat{\alpha}_2^2(2n-3)}{n^3}$ является бутстреповским аналогом (оценкой с помощью бутстрепа) для

$$\frac{4(\mathrm{E}X)^2\mathrm{E}(X - \mathrm{E}X)^2}{n} + \frac{4\mathrm{E}X\mathrm{E}(X - \mathrm{E}X)^3}{n^2} + \frac{\mathrm{E}(X - \mathrm{E}X)^4}{n^3} + \frac{\left(\mathrm{E}(X - \mathrm{E}X)^2\right)^2(2n-3)}{n^3} =$$

что по формулам выше для центральных моментов можно преобразовать как

$$= \frac{4(\mathrm{E}X)^2\left(\mathrm{E}X^2 - (\mathrm{E}X)^2\right)}{n} + \frac{4\mathrm{E}X\left(\mathrm{E}X^3 - 3\mathrm{E}X^2\mathrm{E}X + 2(\mathrm{E}X)^3\right)}{n^2} + \frac{\mathrm{E}X^4 - 4\mathrm{E}X^3\mathrm{E}X + 6\mathrm{E}X^2(\mathrm{E}X)^2 - 3(\mathrm{E}X)^4}{n^3} + \frac{\left(\mathrm{E}X^2 - (\mathrm{E}X)^2\right)^2(}{n^3}$$

$$= \frac{4\mathrm{E}X^2(\mathrm{E}X)^2 - 4(\mathrm{E}X)^4}{n} + \frac{4\mathrm{E}X^3\mathrm{E}X - 12\mathrm{E}X^2(\mathrm{E}X)^2 + 8(\mathrm{E}X)^4}{n^2} + \frac{\mathrm{E}X^4 - 4\mathrm{E}X^3\mathrm{E}X + 6\mathrm{E}X^2(\mathrm{E}X)^2 - 3(\mathrm{E}X)^4}{n^3} + \frac{\left(\mathrm{E}X^2\mathrm{E}X^2 - 2\mathrm{E}X^2\right.}{}$$

$$= \frac{1}{n^3}\mathrm{E}X^4 + \left(\frac{4}{n^2} - \frac{4}{n^3}\right)\mathrm{E}X^3\mathrm{E}X + \frac{2n-3}{n^3}\mathrm{E}X^2\mathrm{E}X^2 + \left(\frac{4}{n} - \frac{12}{n^2} + \frac{6}{n^3} - \frac{4n-6}{n^3}\right)\mathrm{E}X^2(\mathrm{E}X)^2 + \left(-\frac{4}{n} + \frac{8}{n^2} - \frac{3}{n^3} + \frac{2n-3}{n^3}\right)(\mathrm{E}$$

$$= \frac{1}{n^3}\mathrm{E}X^4 + \frac{4n-4}{n^3}\mathrm{E}X^3\mathrm{E}X + \frac{2n-3}{n^3}\mathrm{E}X^2\mathrm{E}X^2 + \frac{4n^2 - 12n + 6 - 4n + 6}{n^3}\mathrm{E}X^2(\mathrm{E}X)^2 + \frac{-4n^2 + 8n - 3 + 2n - 3}{n^3}(\mathrm{E}X)^4 = \mathrm{V}T_n$$

что и требовалось доказать.