In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import maxwell
import scipy.stats as sts
from scipy.stats import geom
from random import random
from collections import Counter
import copy
import math
from math import *
from random import *
import pandas as pd
import calendar
import statsmodels.api as sm

plt.style.use('ggplot')  # Красивые графики
plt.rcParams['figure.figsize'] = (15, 5)  # Размер картинок
```

Содержание

Теоретическое введение

Выборочное среднее и выборочная дисперсия

Наиболее важными характеристиками случайной величины ξ являются ее моменты $\alpha_k = M\xi_k$, а также цетральные моменты $\mu^k = M(\xi - \alpha 1)^k$ (когда они существуют). Их статистическими аналогами, вычисляемыми по соответсвующей выборке $X = (X_1, \ldots, X_n)$, являются выборочные моменты соотаетсвенно обычные:

$$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

и центральные:

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\alpha}_1)^k$$

$\hat{\alpha}_1$ (принято обозначать, как $\overline{X}$) называют выборочным средним, μ^2 - выборочной дисперсией. Таким образом, выборочное среднее и выборочная дисперсия являются статистическими аналогами теоритических среднего (математического ожидания) $M\xi$ и дисперсии $D\xi$, когда они существуют.

Найдем математическое ожидание и дисперсию выборочного среднего и выборочной дисперсии:

$$M\overline{X} = \frac{1}{n}\sum_{i=1}^{n} MX_i = M\xi = \alpha_1$$

$$D\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n}D\xi = \frac{\mu_2}{n}$$

Для выборочной дисперсии введем обозначение: $Y_i = X_i - \alpha_1$:

$$S^2 = \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2$$

Поскольку $MY_i = 0, MY_i^2 = \mu_2$ и $MY_iY_j = MY_jY_i = 0, (i \neq j)$, то:

$$M\overline{Y}^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} MY_iY_j = \frac{1}{n^2}\sum_{i=1}^{n} MY_i^2 = \frac{\mu_2}{n}$$

Отсюда следует, что

$$MS^2 = \frac{n-1}{n}\mu_2$$

Перейдём к вычислению DS^2

$$(S^2)^2 = \frac{1}{n^2}(\sum_{i=1}^{n} Y_i^2)^2 - \frac{2}{n}\overline{Y}^2\sum_{i=1}^{n} Y_i^2 + \overline{Y}^4$$

Так как случайные величины $Y_1, \ldots, Y_n$ независимы и $MY_i = 0$, то в правой части равенства

$$M\overline{Y}^4 = \frac{1}{n^4}(n\mu_4 + 3n(n-1)\mu_2^2) = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3}$$

Аналогично находим

$$\frac{1}{n^2}M(\sum_{i=1}^{n} Y_i^2) = \frac{\mu_4 + (n-1)\mu_2^2}{n} = M(\overline{Y}^2\sum_{i=1}^{n} Y_i^2)$$

С учётом этих соотношений по формуле

$$DS^2 = M(S^2)^2 - (MS^2)^2$$

Теперь рассмотрим свойства выборочных среднего и дисперсии при неограниченном возрастании объема выборки n, которые дадут нам ответ на вопрос, оценками каких параметров рапределений они являются. Чтобы подчеркнуть зависимость моментов $\hat{\alpha}_k, \hat{\mu}_k$ от объема выборки, будем в дальнейшем приписывать дополнительный индекс n: $\hat{\alpha}_{nk}, \hat{\mu}_{nk}$

$$M\hat{\alpha}_{nk} = \frac{1}{n}\sum_{i=1}^{n} MX_i^k = M\xi^k = \hat{\alpha}_k$$

$$D\hat{\alpha}_{nk} = \frac{1}{n^2}\sum_{i=1}^{n} DX_i^k = \frac{1}{n}D\xi^k = \frac{1}{n}(M\xi^{2k} - (M\xi^k)^2) = \frac{\alpha_{2k} - \alpha_k^2}{n}$$

На основании неравенства Чебышева, отсюда следует, что для любого $\epsilon > 0$ при $n \to \infty$

$$P|\hat{\alpha}_{nk} - \alpha_k| < \epsilon \to 1$$

т.е. выборочный момент $\hat{\alpha}_{nk}$ сходится по вероятности при $n \to \infty$ к соответствующему теоретическому моменту α_k . Таким образом, $\hat{\alpha}_{nk}$ можно использовать в качестве оценки α_k, когда объем выборки достаточно велик. Аналогичное утверждение справедливо и для центральных моментов:

$$P|\mu_{nk} - \mu_k| < \epsilon \to 1$$

т.е. μ_{nk} можно использовать в качестве оценки μ_k , когда объем выборки достаточно велик.

1. Оценка $\hat{\theta}(X)$ параметра θ называется несмещенной, если:

$$E(\hat{\theta}(X)) = \theta$$

2. Оценка $\hat{\theta}(X) = \hat{\theta}_n(X_1, \ldots, X_n)$ параметра θ называется состоятельной, если при $n \to \infty$ соблюдается:

$$\hat{\theta}_n(X_1, \ldots, X_n) \xrightarrow{p} \theta$$

При этом для проверки состоятельности достаточно убедиться, что соблюдены следующие два условия:

$$\lim_{n\to\infty} E(\hat{\theta}_n(X_1, \ldots, X_n)) = \theta$$

$$\lim_{n\to\infty} Var(\hat{\theta}_n(X_1, \ldots, X_n)) = 0$$

Выборочное среднее является несмещенной оценкой для теоретического математического ожидания.

$$\hat{\alpha}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$M\hat{\alpha}_1 = M(\frac{1}{n} \sum_{i=1}^{n} x_i) = \frac{1}{n} \cdot n \cdot MX = MX$$

$$M\hat{\alpha}_1 = MX$$

Выборочное среднее является состоятельной оценкой для теоретического математического ожидания.

$$\hat{\alpha}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\lim_{n\to\infty} \hat{\alpha}_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \lim_{n\to\infty} \frac{1}{n} \cdot n \cdot MX = MX$$

Выборочная дисперсия S^2 является состоятельной и несмещенной оценкой для теоретической дисперсии.

Продолжим исследование свойств выборочных моментов для больших выборок и рассмотрим теперь асимптотическое поведение их выборочных распределений.
Если распределение случайной величины ν_n сходится при $n \to \infty$ к распределению случайной величины ν и при этом $\zeta(\nu) = N(\mu, \sigma^2)$, то будем писать $\zeta(\nu_n) \to N(\mu, \sigma^2)$. Далее иногда будем говорить, что случайная величина ν_n асимптотически нормальна $N(\mu_n, \sigma_n^2)$, и записывать это следующим образом:
$\zeta(\nu_n) \approx N(\mu_n, \sigma_n^2)$, если $\zeta(\frac{\nu_n - \mu_n}{\sigma_n}) \to N(0, 1)$.
Найдем сначала асимптотические распределения выборочных моментов $\hat{\alpha}_{nk}$. Величина $n\hat{\alpha}_{nk} = \sum_{i=1}^{n} X_i^k$ является суммой независимых одинаково распределенных случайных величин. Если конечен момент $\alpha_{2k} = M\xi^{2k}$, то к этой сумме можно применить центральную предельную теорему теории вероятностей. Так как $MX_i^k = \alpha^k$, $DX_i^k = \alpha_{2k} - \alpha_k^2$, то величина

$$\frac{n\hat{\alpha}_{nk} - n\alpha_k}{\sqrt{n(\alpha_{2k} - \alpha_k^2)}} = \frac{\hat{\alpha}_{nk} - \alpha_k}{\sqrt{\frac{\alpha_{2k} - \alpha_k^2}{n}}}$$

асимтотически нормальна $N(0, 1)$. Таким образом справедлива следующая теорема:

Если конечен теоритический момент α_{2k} , то при $n \to \infty$ выборочный момент $\hat{\alpha}_{nk}$ асимптотически нормален $N(\alpha_k, \frac{\alpha_{2k} - \alpha_k^2}{n})$
Из теоремы следует, что если существует теоретическая дисперсия, то выборочное среднее $\hat{\alpha}_{n1}$ асимптотически нормально $N(\alpha_1, \frac{\mu_2}{n})$
Из теоремы об асимптотической нормальности функций от выборочных моментов следует, что асимптотически нормальными являются и центральные выборочные моменты $\hat{\mu}_{nk}$, поскольку они являются непрерывными функциями (многочленами) от обычных выборочных моментов.

Доверительный интервал

Определение: γ - доверительным интервалом для g называется такой случайный интервал $(T_1(X), T_2(X)), T_1(X) < T_2(X)$, который содержит внутри себя (накрывает) неизвестное значение g с вероятностью, не меньшей γ:

$$P\{T_1(X) < g < T_2(X)\} \geq \gamma$$

Здесь $T_1(X)$ и $T_2(X)$ - некоторые статистики (функции от выборки), называемые соответственно нижней и верхней доверительными границами, а γ - задаваемый заранее доверительный уровень, который обычно выбирается близким к 1. Длина доверительного интервала характеризует точность локализации оцениваемой характеристики g, а величина γ является показателем надежности доверительного интервала.

В сформулированной ранее теореме [Если конечен теоретический момент α_{2k} , то при $n \to \infty$ выборочный момент $\hat{\alpha}_{nk}$ асимптотически нормален $N(\alpha_k, \frac{\alpha_{2k}-\alpha_k^2}{n})$] можно заменить асимптотическую дисперсию $\frac{\alpha_{2k}-\alpha_k^2}{n}$ ее оценкой $\frac{\hat{\alpha}_{n,2k}-\hat{\alpha}_{nk}^2}{n}$. Это дает искомый асимптотический γ-доверительный интервал для момента α_k вида:

$$(\hat{\alpha}_{nk} \mp c_\gamma \sqrt{\frac{\hat{\alpha}_{n,2k} - \hat{\alpha}_{nk}^2}{n}})$$

Полагая здесь k = 1, получим соответствующий интервал для теоретического среднего $\alpha_1 = M\xi$:

$$(\overline{X} \mp \frac{c_\gamma S}{\sqrt{n}})$$

Чтобы построить асимптотический γ-доверительный интервал для теоретической дисперсии $\mu_2 = D\xi$, надо просто воспользоваться результатом теоремы об асимптотической нормальности выборочной дисперсии [$\zeta(\frac{\sqrt{n}(S^2-\mu_2)}{\sqrt{\hat{\mu}_{n4}-S^4}}) \to N(0,1)$]: искомый интервал есть

$$(S^2 \mp c_\gamma \sqrt{\frac{\hat{\mu}_{n4} - S^4}{n}})$$

Оптимальность оценок

Для построения теории оптимального оценивания прежде всего надо договориться о мере точности оценок, т.е. уточнить смысл приближенного равенства $T(X) \approx g$. Если статистика T(x) используется для оценивания g, то одной из разумных мер расхождения между ними является $(T(X) - g)^2$, или квадратичная ошибка. Но так как это величина случайная используется среднеквадратичная ошибка (с. к. о.) $\Delta(T) = M(T(X) - g)^2$.

Определение: Оценка минимизирующая с. к. о. в данном классе оценок T_g называется оптимальной в среднеквадратичном смысле и обозначается T^* :

$$T^* = argmin_{T \in T_g} \Delta(T)$$

Пусть требуется оценить заданную параметрическую функцию $\tau(\theta)$ в модели $F = F(x; \theta), \theta \in \Theta$ по соответствующей выборке $X = (X_1, \ldots, X_n)$. Обозначим τ_τ класс всех несмещенных оценок $T = T(X)$ для $\tau(\theta)$ и предположим, что он не пуст. Дополнительно предположим, что дисперсии всех оценок из класса τ_τ конечны: $D\theta T = M_\theta(T - \tau(\theta))2 < \infty$, в этом случае мерой точности оценок является их дисперсия.

Утверждение: Для несмещенных оценок среднеквадратичное отклонение совпадает с ее дисперсией, а для смещенной оценки больше ее дисперсии.

Доказательство:

$$M\theta(T - \tau)^2 = M(T - MT + MT - \tau)^2 = M(T - MT)^2 + M(MT - \tau)^2 + 2M((T - MT)(MT - \tau)) = DT + b^2 = 0 \iff MT = \tau$$

Теорема Рао-Блэкуэлла-Колмогорова: Оптимальная оценка, если она существует, является функцией от достаточной статистики.

По определению достаточная статистика T = T(X) называется полной, если для всякой функции $\varphi(T)$ из того, что

$$M_\theta \varphi(T) = 0, \forall \theta$$

следует $\varphi(t) \equiv 0$ на всем множестве значений статистики T.

Теорема: Если существует полная достаточная статистика, то всякая функция от нее является оптимальной оценкой своего математического ожидания.

Итак, пусть существует полная достаточная статистика T = T(X) и требуется оценить заданную параметрическую функию $\tau(\theta)$.

Тогда:

1)Если существует какая-то несмещенная оценка $\tau(\theta)$, то существует и несмещенная оценка, являющаяся функцией от T; можно так же сказать, что если нет несмещенныхоценок вида H(T), то класс несмещенных оценок τ_τ для $\tau(\theta)$ пуст;

2)оптимальная (н.о.р.м.д.) оценка когда она существует, всегда является функцией от T и она однозначно определяется уравнением $M_\theta H(T) = \tau(\theta)$

3)оптимальную оценку τ^* можно искать по формуле:

3.1. Геометрическое распределение

In [2]:

```
# Создание случайной величины с геометрическим распределением, зависящим
# от параметра p
p = 0.5
geom_rv = sts.geom(p)
```

In [3]:

```
#Генерация выборки объема n = 5 с выводом
for n in [5]:
    means_5 = []
    for i in range(5):
        sample = geom_rv.rvs(n)
        means_5.append(sample)
        print(sample)
```

```
[2 1 1 1 2]
[1 1 2 3 1]
[3 1 1 1 2]
[1 3 1 3 3]
[2 7 1 3 3]
```

In [4]:

```
#Генерация выборки объема n = 10 с выводом
for n in [10]:
    means_10 = []
    for i in range(5):
        sample = geom_rv.rvs(n)
        means_10.append(sample)
        print(sample)
```

```
[3 3 2 2 1 1 1 1 1 1]
[1 1 1 1 2 1 3 3 1 2]
[1 5 1 1 4 1 1 1 1 2]
[1 3 1 2 1 4 3 2 1 2]
[2 1 4 1 1 3 7 3 1 2]
```

In [5]:

```python
#Генерация выборки объема n = 100 ,без вывода
for n in [100]:
    means_100 = []
    for i in range(5):
        sample = geom_rv.rvs(n)
        means_100.append(sample)
```

In [6]:

```python
#Генерация выборки объема n = 1000 ,без вывода
for n in [1000]:
    means_1000 = []
    for i in range(5):
        sample = geom_rv.rvs(n)
        means_1000.append(sample)
```

In [7]:

```python
#Генерация выборки объема n = 100000 ,без вывода
for n in [100000]:
    means_100000 = []
    for i in range(5):
        sample = geom_rv.rvs(n)
        means_100000.append(sample)
```

3.1.1 Нахождение выборочного среднего и выборочной дисперсии геометрического распределения

Выборочное среднее

In [8]:

```
#n=5
vs_5 = []
for i in range(5):
    vs_5.append(np.mean(means_5[i]))
print(vs_5)
#n=10
vs_10 = []
for i in range(5):
    vs_10.append(np.mean(means_10[i]))
print(vs_10)
#n=100
vs_100 = []
for i in range(5):
    vs_100.append(np.mean(means_100[i]))
print(vs_100)
#n=1000
vs_1000 = []
for i in range(5):
    vs_1000.append(np.mean(means_1000[i]))
print(vs_1000)
#n=100000
vs_100000 = []
for i in range(5):
    vs_100000.append(np.mean(means_100000[i]))
print(vs_100000)
```

```
[1.4, 1.6, 1.6, 2.2, 3.2]
[1.6, 1.6, 1.8, 2.0, 2.5]
[2.23, 2.19, 1.97, 2.09, 1.89]
[1.949, 2.038, 1.993, 1.942, 1.977]
[1.99304, 2.00094, 1.99217, 1.99835, 2.00392]
```

$M\xi = \frac{1}{p}$, при p = 0.5 $M\xi = 2$

In [9]:
```
#Сравнение
p = 0.5
geom.stats(p, moments = 'm')
```
Out[9]:
```
array(2.)
```

Выборочная дисперсия

In [10]:

```
#n=5
vd_5 = []
for i in range(5):
    vd_5.append(round(np.var(means_5[i]),6))
print(vd_5)
#n=10
vd_10 = []
for i in range(5):
    vd_10.append(round(np.var(means_10[i]),6))
print(vd_10)
#n=100
vd_100 = []
for i in range(5):
    vd_100.append(round(np.var(means_100[i]),6))
print(vd_100)
#n=1000
vd_1000 = []
for i in range(5):
    vd_1000.append(round(np.var(means_1000[i]),6))
print(vd_1000)
#n=100000
vd_100000 = []
for i in range(5):
    vd_100000.append(round(np.var(means_100000[i]),6))
print(vd_100000)
```

```
[0.24, 0.64, 0.64, 0.96, 4.16]
[0.64, 0.64, 1.96, 1.0, 3.25]
[2.1371, 3.7939, 2.1491, 2.4019, 1.9979]
[1.984399, 2.132556, 1.824951, 1.734636, 2.008471]
[1.973252, 2.001179, 1.984249, 2.009987, 2.025725]
```

$$D\xi = \frac{q}{p^2} = \frac{1-p}{p^2}$$

При p = 0.5, $D\xi = 2$

In [11]:
```
#Сравнение
geom_stats(p_moments = 'v')
```
Out[11]:
```
array(2.)
```

Как видно из полученных значений, чем больше объём выборки, тем менее отличаются выборочное среднее от теоретического математического ожидания и выборочная дисперсия от теоретической дисперсии

3.1.2 Построение доверительного интервала для выборочного среднего

Положим $\gamma = 0.95$ и найдем доверительный интервал для выборочного среднего.

$$\Phi(c_\gamma) = \frac{\gamma}{2} = 0.475$$

Из таблицы значений функции Лапласа $c_\gamma \approx 1.96$

In [12]:

```
#n=5
print('n = 5')
for i in range(5):
    print('(', vs_5[i], '-+ 1.96 *', np.sqrt(vd_5[i]/5), ') = (', vs_5[i], '-+', round(1.96*np.s
#n=10
print('n = 10')
for i in range(5):
    print('(', vs_10[i], '-+ 1.96 *', np.sqrt(vd_10[i]/10), ')= (', vs_10[i], '-+', round(1.96*n
#n=100
print('n = 100')
for i in range(5):
    print('(', vs_100[i], '-+ 1.96 *', np.sqrt(vd_100[i]/100), ') = (', vs_100[i], '-+', round(1
#n=1000
print('n = 1000')
for i in range(5):
    print('(', vs_1000[i], '-+ 1.96 *', np.sqrt(vd_1000[i]/1000), ') = (', vs_1000[i], '-+', rou
#n=100000
print('n = 100000')
for i in range(5):
    print('(', vs_100000[i], '-+ 1.96 *', np.sqrt(vd_100000[i]/100000), ') = (', vs_100000[i], '
```

```
n = 5
( 1.4 -+ 1.96 * 0.21908902300206645 ) = ( 1.4 -+ 0.429414 )
( 1.6 -+ 1.96 * 0.35777087639996635 ) = ( 1.6 -+ 0.701231 )
( 1.6 -+ 1.96 * 0.35777087639996635 ) = ( 1.6 -+ 0.701231 )
( 2.2 -+ 1.96 * 0.4381780460041329 ) = ( 2.2 -+ 0.858829 )
( 3.2 -+ 1.96 * 0.9121403400793104 ) = ( 3.2 -+ 1.787795 )
n = 10
( 1.6 -+ 1.96 * 0.25298221281347033 )= ( 1.6 -+ 0.495845 )
( 1.6 -+ 1.96 * 0.25298221281347033 )= ( 1.6 -+ 0.495845 )
( 1.8 -+ 1.96 * 0.4427188724235731 )= ( 1.8 -+ 0.867729 )
( 2.0 -+ 1.96 * 0.31622776601683794 )= ( 2.0 -+ 0.619806 )
( 2.5 -+ 1.96 * 0.570087712549569 )= ( 2.5 -+ 1.117372 )
n = 100
( 2.23 -+ 1.96 * 0.14618823482072693 ) = ( 2.23 -+ 0.286529 )
( 2.19 -+ 1.96 * 0.19477936235648785 ) = ( 2.19 -+ 0.381768 )
```

*Округлено до 6 знаков после запятой.

3.1.3 Нахождение оптимальности рассматриваемых оценок

Условия регулярности:

1) Так как $0 < \theta < 1$, то $L(\overline{x}, \theta) > 0$. Функция дифференцируема по θ.

2) Воспользуемся равенством:

$$i_n(\theta) = M_\theta V^2(X; \theta) = n * i(\theta)$$

Найдём $i(\theta)$:

$$i(\theta) = -M_\theta \frac{\partial^2(lnf_\theta(x_1))}{\partial\theta}$$

$$\frac{\partial^2(ln(f(x1, \theta)))}{\partial\theta} = \frac{\partial^2(xln((1-\theta)\theta))}{\partial\theta}$$

Первая производная:

$$\frac{x\theta + \theta - 1}{(\theta - 1)\theta}$$

Вторая производная:

$$-\frac{\theta^2(x+1) + 2\theta - 1}{(\theta-1)^2\theta^2}$$

$$M(x) = \frac{1-\theta}{\theta}$$

$$M_\theta(\frac{\theta^2(x+1+2\theta-1)}{(\theta-1)^2\theta^2}) = M_\theta(\frac{\theta^2 x}{(\theta-1)^2\theta^2}) + \frac{\theta + 2\theta - 1}{(\theta-1)^2\theta^2} = (\frac{\theta^2 M(x)}{(\theta-1)^2\theta^2}) + \frac{3\theta - 1}{(\theta-1)^2\theta^2} = (-\frac{\theta^2(1-\theta)}{(\theta-1)^2\theta^3}) + \frac{3}{(\theta}$$

$$-\frac{1}{(\theta-1)\theta} + \frac{3\theta - 1}{(\theta-1)^2\theta^2} = -\frac{\theta^2 - 4\theta + 1}{(\theta-1)^2\theta^2}$$

В силу того, что $0 < \theta < 1$, то информация Фишера конечна, следовательно, свойство выполняется.

3) Выборочное среднее не зависит от параметра θ, то свойство выполняется.

$$f(x) = (1-p)^x p$$

Запишем логарифмическую функцию правдоподобия:

$$L(\overline{x};\theta) = \prod_{i=1}^{n} (1-\theta)^{x_i}\theta$$

Прологарифмируем обе части:

$$lnL(\overline{x};\theta) = \sum_{i=1}^{n} ln(1-\theta)^{x_i}\theta = \sum_{i=1}^{n} ln(\theta) + \sum_{i=1}^{n} ln((1-\theta)x_i)$$

$$\frac{\partial(ln(L(x,\theta)))}{\partial\theta} = \sum_{i=1}^{n}(\frac{1}{\theta}) + \sum_{i=1}^{n}(-\frac{x_i}{1-\theta})$$

Так как

$$\sum_{i=1}^{n}(\frac{x_i}{1-\theta}) = \sum_{i=1}^{n}(\frac{1}{\theta}), \theta \neq 0; 1,$$

то

$$\frac{1}{1-\theta}\overline{X} = \frac{1}{\theta}$$

$$\overline{X} = \frac{1-\theta}{\theta}$$

$$M(\frac{1-\theta}{\theta}) = M\overline{X} = M\xi$$

А значит, оценка несмещенная. Также оценка является состоятельной, в силу того, что

$$M\overline{X} \overset{n\to\infty}{\overset{P}{\to}} \overline{X}$$

Теперь приступим к оцениванию эффективности данной оценки:

Вспомним, что

$$-M_\theta \frac{\partial^2(ln(f_\theta(X_1)))}{\partial\theta} = i_1(\theta)$$

f(x) геометрического распределения можно дважды продифференцировать.

3.2 Распределения Максвелла

In [13]:
```
# Создание случайной величины с распределением Максвелла, зависящим
# от параметра lambda
lambd=1.0
maxwell_rv=sts.maxwell(scale=lambd)
```

In [14]:
```
#Генерация выборки объема n = 5 с выводом
for n in[5]:
    means__5=[]
    for i in range(5):
        sample=maxwell_rv.rvs(n)
        means__5.append(sample)
        print(sample)
```

```
[2.41609712 1.76922215 1.5930344  1.83891646 2.32443766]
[1.91594045 1.51313849 2.17701851 2.98251721 1.74690197]
[0.54612996 1.82129461 2.12808232 1.87961106 0.53487278]
[1.58352172 1.74733789 2.13514589 2.59575027 2.23143757]
[0.86938855 0.91584446 2.23479136 0.67373006 1.37449349]
```

In [15]:

```
#Генерация выборки объема n = 10 с выводом
for n in[10]:
    means__10=[]
    for i in range(5):
        sample=maxwell_rv.rvs(n)
        means__10.append(sample)
        print(sample)
```

```
[1.29333112 2.49794698 0.39525199 2.28555043 0.44500091 2.38964386
 1.30222352 2.28941905 1.25920271 0.90757726]
[0.2591424  0.3525709  1.56350926 2.61469714 1.67857905 1.5425198
 1.1630406  0.3912105  1.628316   1.65754265]
[1.73446419 1.28225808 1.62562587 1.88573595 1.72514081 1.35559837
 1.32751044 0.87079062 3.27650123 2.19524686]
[0.91181159 1.58967585 0.6389368  2.57554735 1.87469683 1.5959847
 0.94075626 1.0759609  2.43485971 2.19496471]
[2.34641351 2.46242359 3.04918384 2.24460863 1.58035144 2.55553114
 0.61721371 1.7907363  0.3763505  1.00948433]
```

In [16]:

```
#Генерация выборки объема n = 100 без вывода
for n in[100]:
    means__100=[]
    for i in range(5):
        sample=maxwell_rv.rvs(n)
        means__100.append(sample)
```

In [17]:

```
#Генерация выборки объема n = 1000 без вывода
for n in[1000]:
    means__1000=[]
    for i in range(5):
        sample=maxwell_rv.rvs(n)
        means__1000.append(sample)
```

In [18]:

```python
#Генерация выборки объема n = 100000 без вывода
for n in[100000]:
    means__100000=[]
    for i in range(5):
        sample=maxwell_rv.rvs(n)
        means__100000.append(sample)
```

3.2.1 Нахождение выборочного среднего и выборочной дисперсии геометрического распределения

Выборочное среднее

In [19]:

```
#n=5
vs__5 = []
for i in range(5):
    vs__5.append(round(np.mean(means__5[i]),6))
print(vs__5)
#n=10
vs__10 = []
for i in range(5):
    vs__10.append(round(np.mean(means__10[i]),6))
print(vs__10)
#n=100
vs__100 = []
for i in range(5):
    vs__100.append(round(np.mean(means__100[i]),6))
print(vs__100)
#n=1000
vs__1000 = []
for i in range(5):
    vs__1000.append(round(np.mean(means__1000[i]),6))
print(vs__1000)
#n=100000
vs__100000 = []
for i in range(5):
    vs__100000.append(round(np.mean(means__100000[i]),6))
print(vs__100000)
```

```
[1.988342, 2.067103, 1.381998, 2.058639, 1.21365]
[1.506515, 1.285113, 1.727887, 1.583319, 1.80323]
[1.588683, 1.601758, 1.624953, 1.511346, 1.698355]
[1.566829, 1.647546, 1.574126, 1.629635, 1.579183]
[1.593033, 1.593984, 1.593523, 1.596336, 1.59679]
```

In [20]:

```
#Сравнение
2*np.sqrt(2/np.pi)
```

Out[20]: 1.5957691216057308

$$M\xi = 2\lambda\sqrt{\frac{2}{\pi}}$$

При $\lambda = 1.0$ $M\xi = 2 \cdot \sqrt{\frac{2}{\pi}} \approx 1.5957691216057308$

Выборочная дисперсия

In [21]:

```
#n=5
vd__5 = []
for i in range(5):
    vd__5.append(round(np.var(means__5[i]),6))
print(vd__5)
#n=10
vd__10 = []
for i in range(5):
    vd__10.append(round(np.var(means__10[i]),6))
print(vd__10)
#n=100
vd__100 = []
for i in range(5):
    vd__100.append(round(np.var(means__100[i]),6))
print(vd__100)
#n=1000
vd__1000 = []
for i in range(5):
    vd__1000.append(round(np.var(means__1000[i]),6))
print(vd__1000)
#n=100000
vd__100000 = []
for i in range(5):
    vd__100000.append(round(np.var(means__100000[i]),6))
print(vd__100000)
```

```
[0.104509, 0.256464, 0.482708, 0.129369, 0.313464]
[0.585148, 0.505908, 0.388416, 0.418193, 0.716524]
[0.461868, 0.452764, 0.410254, 0.445777, 0.275561]
[0.450162, 0.455311, 0.444754, 0.427367, 0.445206]
[0.456508, 0.451218, 0.451222, 0.45127, 0.455822]
```

In [22]:

```
#Сравнение
(3*np.pi-8)/np.pi
```

Out[22]:

```
0.4535209105296745
```

$$D\xi = \frac{3\pi - 8}{\pi} \cdot \lambda$$

При $\lambda = 1.0$ $D\xi = \frac{3\pi-8}{\pi} \approx 0.4535209105296745$

Как видно из полученных значений, чем больше объём выборки, тем менее отличаются выборочное среднее от теоретического математического ожидания и выборочная дисперсия от теоретической дисперсии.

3.2.2 Построение доверительного интервала для выборочного среднего

Положим $\gamma = 0.95$ и найдем доверительный интервал для выборочного среднего.

$$\Phi(c_\gamma) = \frac{\gamma}{2} = 0.475$$

Из таблицы значений функции Лапласа $c_\gamma \approx 1.96$

In [23]:

```python
#n=5
print('n = 5')
for i in range(5):
    print('(', vs__5[i], '-+ 1.96 *', np.sqrt(vd__5[i]/5), ') = (', vs__5[i], '-+', round(1.96*n
#n=10
print('n = 10')
for i in range(5):
    print('(', vs__10[i], '-+ 1.96 *', np.sqrt(vd__10[i]/10), ')= (', vs__10[i], '-+', round(1.9
#n=100
print('n = 100')
for i in range(5):
    print('(', vs__100[i], '-+ 1.96 *', np.sqrt(vd__100[i]/100), ') = (', vs__100[i], '-+', roun
#n=1000
print('n = 1000')
for i in range(5):
    print('(', vs__1000[i], '-+ 1.96 *', np.sqrt(vd__1000[i]/1000), ') = (', vs__1000[i], '-+',
#n=100000
print('n = 100000')
for i in range(5):
    print('(', vs__100000[i], '-+ 1.96 *', np.sqrt(vd__100000[i]/100000), ') = (', vs__100000[i]
```

```
n = 5
( 1.988342 -+ 1.96 * 0.1445745482441498 ) = ( 1.988342 -+ 0.283366 )
( 2.067103 -+ 1.96 * 0.22647913811210074 ) = ( 2.067103 -+ 0.443899 )
( 1.381998 -+ 1.96 * 0.31071144169470166 ) = ( 1.381998 -+ 0.608994 )
( 2.058639 -+ 1.96 * 0.16085334935897358 ) = ( 2.058639 -+ 0.315273 )
( 1.21365 -+ 1.96 * 0.2503853030830684 ) = ( 1.21365 -+ 0.490755 )
n = 10
( 1.506515 -+ 1.96 * 0.24189832574865003 )= ( 1.506515 -+ 0.474121 )
( 1.285113 -+ 1.96 * 0.22492398716010706 )= ( 1.285113 -+ 0.440851 )
( 1.727887 -+ 1.96 * 0.19708272374817637 )= ( 1.727887 -+ 0.386282 )
( 1.583319 -+ 1.96 * 0.20449767724842255 )= ( 1.583319 -+ 0.400815 )
( 1.80323 -+ 1.96 * 0.2676796592944634 )= ( 1.80323 -+ 0.524652 )
n = 100
( 1.588683 -+ 1.96 * 0.0679608710950647 ) = ( 1.588683 -+ 0.133203 )
( 1.601758 -+ 1.96 * 0.0672877403395299 ) = ( 1.601758 -+ 0.131884 )
```

*Округлено до 6 знаков после запятой.

3.2.3 Нахождение оптимальности рассматриваемых оценок

Для простоты предположим, что все частоты $p_i, i = \overline{1, n}$ равны единице.
Запишем функцию максимального правдоподобия для закона Максвелла.

$$L(\lambda) = (\frac{2}{\pi})^{\frac{n}{2}} \frac{\prod_{i=1}^{n} x_i^2}{\lambda^{3n}} e^{-\frac{1}{2\lambda^2} \sum_{i=1}^{n} x_i^2}$$

$$ln(L(\lambda)) = 2\sum_{i=1}^{n} lnx_j - 3nln\lambda - \frac{n}{2}ln\frac{2}{\pi} - \frac{1}{2\lambda^2}\sum_{i=1}^{n} x_i^2$$

$$\frac{\partial(lnL(\lambda))}{\partial\lambda} = -\frac{3n}{\lambda} + \frac{\sum_{i=1}^{n} x_i^2}{\lambda^3} = 0$$

Переходя к статистическому ряду (не все p_i равны 1, $i = \overline{1, n}$), получим уравнение для нахождения λ:

$$\lambda = \sqrt{\frac{\sum_{i=1}^{m} p_i x_i^2}{3n}}$$

Оценка методом моментов:
Поскольку по выборке оценивается лишь один параметр, то для нахождения λ используемся оценку математического ожидания.

$$M[X] = \overline{X} = 2\lambda\sqrt{\frac{2}{\pi}},$$

где $\overline{X} = \frac{1}{n}\sum_{i=1}^{m} x_i p_i$, $\sum_{i=1}^{m} p_i = n$
Отсюда $\lambda = \sqrt{\frac{\pi}{8}}\overline{X}$