§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

ПАВЕЛ Е. РЯБОВ[1]

[1]Финансовый университет при Правительстве Российской Федерации

PERyabov@fa.ru

2 апреля 2024 г.

§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

План лекции

- Биномиальное распределение. Геометрическое расперделение. Распределение Пуассона.
- Производящая функция и ее свойства.
- Начальные и центральные моменты дискретных случайных величин.
- Асимметрия распределения (The moment coefficient of *skewness*)
- Эксцесс распределения (Kurtosis Excess)

§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

План лекции

- Биномиальное распределение. Геометрическое расперделение. Распределение Пуассона.
- Производящая функция и ее свойства.
- Начальные и центральные моменты дискретных случайных величин.
- Асимметрия распределения (The moment coefficient of *skewness*)
- Эксцесс распределения (Kurtosis Excess)

§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

План лекции

- Биномиальное распределение. Геометрическое расперделение. Распределение Пуассона.
- Производящая функция и ее свойства.
- Начальные и центральные моменты дискретных случайных величин.
- Асимметрия распределения (The moment coefficient of *skewness*)
- Эксцесс распределения (Kurtosis Excess)

§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

План лекции

- Биномиальное распределение. Геометрическое раcперделение. Распределение Пуассона.
- Производящая функция и ее свойства.
- Начальные и центральные моменты дискретных случайных величин.
- Асимметрия распределения (The moment coefficient of ***skewness***)
- Эксцесс распределения (Kurtosis Excess)

§5. Основные дискретные распределения и их числовые характеристики. Начальные и центральные моменты дискретных случайных величин.

План лекции

- Биномиальное распределение. Геометрическое расперделение. Распределение Пуассона.
- Производящая функция и ее свойства.
- Начальные и центральные моменты дискретных случайных величин.
- Асимметрия распределения (The moment coefficient of ***skewness***)
- Эксцесс распределения (Kurtosis Excess)

Биномиальное распределение

Определение

Биномиальным распределением с параметрами n ***и*** p *называется распределение числа успехов в последовательности из* n *независимых испытаний с вероятностью успеха "*p*" в каждом испытании.*

Обозначение: $X \sim Bin(n;p)$ *и как класс из библиотеки* ***scipy.stats:*** $X = binom(n,p)$.

Пусть X – число успехов в последовательности из n независимых испытаний. Тогда

$$X \sim Bin(n;p) \Leftrightarrow p_k = \mathbb{P}(X = k) \overset{\text{формула Бернулли}}{=} \mathbb{P}_n(k) = C_n^k p^k q^{n-k}, \qquad (1)$$

$$k = 0, 1 \ldots, n; q = 1 - p.$$

Биномиальное распределение, $X \sim Bin(n;p)$

Теорема

Пусть с.в. $X \sim Bin(n;p)$ *(т.е. случайная величина* X *распределена по биномиальному закону распределения с параметрами* n *и* p*). Тогда*

$$\mathbb{E}(X) = np; \qquad \mathbb{V}ar(X) = npq, \quad \text{где} \quad q = 1 - p. \tag{2}$$

Пример

Производятся 5120 *независимых испытаний, состоящих в том, что одновременно подбрасываются* 9 *(симметричных) монет. Пусть* X *– число испытаний, в которых выпало* 2 *герба. Найдите* $\mathbb{E}(X)$ *и* $\mathbb{V}ar(X)$.

Решение. *Заметим, что с.в.* $X \sim Bin(n;p)$*, где* $n = 5120$*. Найдем для этого закона параметр "p" – вероятность успеха (события* A*) в одном испытании, что при подбрасывании* 9 *монет выпадает* 2 *герба. Тогда* $p = \mathbb{P}(A) = \mathbb{P}_9(2) = C_9^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^7 = \frac{36}{512}$ *и* $q = 1 - p = \frac{476}{512}$*. Применяя формулы теоремы, находим* $\mathbb{E}(X) = np = 5120 \cdot \frac{36}{512} = 360$,
$\mathbb{V}ar(X) = npq = 5120 \cdot \frac{36}{512} \cdot \frac{476}{512} = 334{,}6875$.

Геометрическое распределение

Определение

Геометрическим распределением с параметром *“p” называется распределение числа испытаний* ***до*** *первого успеха в серии независимых испытаний с вероятностью успеха “p” в каждом испытании.*
Обозначение: с.в. $X \sim Geom(p)$ *и как класс из библиотеки* ***scipy.stats****:*
$X = geom(p)$.

Пусть с.в. X обозначает число испытаний ***до первого успеха***, тогда ряд распределения (распределение вероятностей) имеет вид

Значения X	1	2	$\dots$	k	$\dots$
Вероятность $p_k = \mathbb{P}(X = k)$	p	qp	$\dots$	$q^{k-1}p$	$\dots$

Пример

Симметричная монета бросается до тех пор, пока не выпадет герб. Пусть X *– число бросков. Тогда с.в.* $X \sim Geom\left(p = \frac{1}{2}\right)$.

Геометрическое распределение

Теорема

Пусть с.в. $X \sim Geom(p)$. *Тогда* $\mathbb{E}(X) = \frac{1}{p}$, $\mathbb{V}ar(X) = \frac{q}{p^2}$.

Пример

Симметричная игральная кость подбрасывается до тех пор, пока не выпадут все грани. Найдите математическое ожидание и дисперсию числа бросаний.

Решение. Пусть X обозначает число бросаний до выпадения всех граней. Представим X в виде ***суммы шести независимых случайных величин***

$$X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6.$$

Здесь X_1 – число бросаний до появления ***какой-либо*** из граней (какой— все равно). Поэтому $X_1 = 1$. Далее, X_2 – число последующих бросаний до появления какой-либо ***другой*** грани, X_3 – число бросаний до появления ***третьей*** (еще не выпадавшей) грани и т.д. Например, если в результате серии бросаний появились грани: 6; 1; 1; 1; 2; 3; 6; 5; 2; 3; 1; 2; 4, то в этом случае $X = 13$, а $X_1 = 1, X_2 = 1, X_3 = 3, X_4 = 1, X_5 = 2, X_6 = 5$. Каждая с.в. X_k распределена по геометрическому закону (а их сумма (X) – ***нет***) с параметром $p_k = \frac{6-(k-1)}{6}$, $X_k \sim Geom\left(p_k = \frac{7-k}{6}\right)$, $k = 1, 2, \ldots, 6$. Используя свойства для математических ожиданий и дисперсий, а также формулы теоремы, находим $\mathbb{E}(X_k) = \frac{1}{p_k} = \frac{6}{6-(k-1)}$. И тогда получаем, что

$$\begin{aligned}
&\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \ldots + X_6) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_6) = \\
&= 6\left(\frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1\right) = 14{,}7; \\
&\mathbb{V}ar(X) = \mathbb{V}ar(X_1 + X_2 + \ldots + X_6) = \mathbb{V}ar(X_1) + \ldots + \mathbb{V}ar(X_6) = \\
&= \frac{\frac{1}{6}}{\left(\frac{5}{6}\right)^2} + \frac{\frac{2}{6}}{\left(\frac{4}{6}\right)^2} + \frac{\frac{3}{6}}{\left(\frac{3}{6}\right)^2} + \frac{\frac{4}{6}}{\left(\frac{2}{6}\right)^2} + \frac{\frac{5}{6}}{\left(\frac{1}{6}\right)^2} = 38{,}99.
\end{aligned} \tag{3}$$

Упражнение

Предположим, что игральная кость ***не является симметричной*** и "вероятности" в ней распределены так, что "вероятность" каждой грани пропорциональна её номеру. Такая (несимметричная) игральная кость подбрасывается до тех пор, пока не выпадут все грани. Найдите математическое ожидание и дисперсию числа бросаний. (Задача сложная, но все равно имеет аналитическое решение!!!)

Ответ: $\mathbb{E}(X) = \frac{1315957}{50160} \approx 26{,}23518740031898;$
$\mathbb{V}ar(X) \approx 340{,}0258005534602.$

Распределение Пуассона

Определение

С.в. X имеет ***распределение Пуассона с параметром*** *$\lambda > 0$, если*

$$p_k = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \ldots$$

Обозначение: $X \sim \Pi(\lambda)$ и как класс из библиотеки ***scipy.stats****:*
$X = poisson(\mu)$ ($\mu = \lambda$).

Пример

Распределение Пуассона используется в тех случаях, когда мы имеем дело с ***редкими событиями****, такими как, например, клиенты, звонящие в справочный центр; посетители веб-сайта; радиоактивный распад атомов; фотоны, поступающие в космический телескоп; изменения цены акций; количество аварий на дорогах.*

Распределение Пуассона

Теорема

Пусть $X \sim \Pi(\lambda)$. *Тогда* $\mathbb{E}(X) = \mathbb{V}ar(X) = \lambda$.

Производящая функция

Пусть X – дискретная случайная величина с законом распределения $p_k = \mathbb{P}(X = x_k), k = 0, 1, 2, \ldots$ (В наших примерах возможные значения $x_k = k$).

Определение

Производящей функцией *дискретной случайной величины X называется функция*

$$\psi_X(s) = \mathbb{E}\left(s^X\right) = \sum_{k=0}^{+\infty} p_k s^k = p_0 + p_1 s + p_2 s^2 + \ldots$$

Область сходимости: $|s| < 1$ *и* $s = 1$.

Производящая функция

Теорема

Производящая функция суммы независимых случайных величин X и Y равна произведению производящих функций слагаемых, т.е

$$\psi_{X+Y}(s) = \psi_X(s) \cdot \psi_Y(s).$$

- Пусь $X \sim Bin(n; p)$. Тогда $\psi_X(s) = (q + ps)^n$, где $q = 1 - p$;
- Пусть $X \sim Geom(p)$. Тогда $\psi_X(s) = \frac{ps}{1-qs}$, где $q = 1 - p$;
- Пусть $X \sim \Pi(\lambda)$. Тогда $\psi_X(s) = e^{\lambda(s-1)}$.

Теорема

Если $X \sim \Pi(\lambda)$, $Y \sim \Pi(\mu)$ и X, Y – независимые случайные величины, то с.в. $X + Y \sim \Pi(\lambda + \mu)$ (сумма независимых случайных величин, распределенных по закону Пуассону, распределена по тому же закону Пуассона).

Производящая функция

Полезные соотношения

Теорема

Пусть $\psi_X(s)$ *– производящая функция д.с.в.* X. *Тогда*

$$\mathbb{E}(X) = \psi'_X(1);$$
$$\mathbb{V}ar(X) = \psi''_X(1) + \psi'_X(1) - [\psi'_X(1)]^2 .$$

Начальные и центральные моменты (дискретных) случайных величин

Определение

***Начальным моментом** k**-го порядка** $\nu_k(X)$, где $k \in \mathbb{N}$, дискретной случайной величины X называют математическое ожидание k-ой степени X, т.е.*

$$\nu_k(X) = \mathbb{E}\left[X^k\right] = \sum_{i=0}^{\infty} x_i^k p_i = \sum_{i=0}^{\infty} x_i^k \mathbb{P}(X = x_i),$$

если ряд, стоящий в правой части, сходится абсолютно, т.е. $\sum_{i=0}^{+\infty} |x_i|^k p_i < +\infty$

В частности, $\nu_1(X) = \mathbb{E}(X)$.

Начальные и центральные моменты (дискретных) случайных величин

Определение

***Центральным моментом** k-го **порядка** $\mu_k(X)$, где $k \in \mathbb{N}$, дискретной случайной величины X называют математическое ожидание k-ой степени **отклонения** $X - \mathbb{E}[X]$, т.е.*

$$\mu_k(X) = \mathbb{E}\left[(X - E[X])^k\right] = \sum_{i=0}^{\infty}(x_i - E[X])^k p_i = \sum_{i=0}^{\infty}(x_i - \mathbb{E}[X])^k \mathbb{P}(X = x_i),$$

если ряд, стоящий в правой части, сходится абсолютно.

В частности, $\mu_1(X) = 0$, $\mu_2(X) = \mathbb{V}ar(X)$.

Дискретные случайные величины

Используя свойства для математических ожиданий, можно получить формулы, которые связывают начальные и центральные моменты. Например,

$$\mu_2 = \nu_2 - \nu_1^2,$$
$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3,$$
$$\mu_4 = \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4.$$

Теорема

Если X и Y – независимые случайные величины, то

$$\mu_3(X+Y) = \mu_3(X) + \mu_3(Y).$$

Формула обобщается на сумму n независимых случайных величин:

$$\mu_3(X_1 + \ldots + X_n) = \mu_3(X_1) + \mu_3(X_2) + \ldots + \mu_3(X_n).$$

Теорема

*Для **независимых** случайных величин X и Y справедлива формула*

$$\mu_4(X+Y) = \mu_4(X) + \mu_4(Y) + 6\mathbb{V}ar(X)\cdot\mathbb{V}ar(Y).$$

В общем случае, для n ***независимых*** случайных величин $X_1, X_2, \ldots, X_n$ справедлива формула:

$$\mu_4\Big(\sum_{k=1}^{n} X_k\Big) = \sum_{k=1}^{n} \mu_4(X_k) + 6\sum_{1\leqslant i<j\leqslant n}\sum \mathbb{V}ar(X_i)\mathbb{V}ar(X_j).$$

Упражнение

- Пусть $X \sim Bin(n; p)$. Тогда

$$\mu_3(X) = pq(1 - 3pq),$$
$$\mu_4(X) = npq(1 + 3npq - 6pq),$$

  где $q = 1 - p$;
- Пусть $X \sim \Pi(\lambda)$. Тогда

$$\nu_3(X) = \lambda^3 + 3\lambda^2 + \lambda,$$
$$\nu_4(X) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.$$

Асимметрия распределения (The moment coefficient of *skewness*)

Определение

Асимметрией распределения (The moment coefficient of skewness) *случайной величины X называется отношение третьего центрального момента $\mu_3(X)$ к кубу стандартного отклонения $\sigma(X)^3$, т.е.*

$$\mathrm{As}(X) = \frac{\mu_3(X)}{(\sigma(X))^3},$$

где $\sigma(X) = \sqrt{\mathbb{V}ar(X)}$.

Свойства асимметрии распределения

Утверждение

- $\mathrm{As}(X) = \nu_3\left(\frac{X-\mathbb{E}(X)}{\sigma(X)}\right) = \mu_3\left(\frac{X-\mathbb{E}(X)}{\sigma(X)}\right)$;
- *Если распределение* ***симметрично*** *относительно математического ожидания* $\mathbb{E}(X)$, *то* $\mathrm{As}(X) = 0$ *(С.в.* X *имеет* ***симметричное*** *распределение относительно* $\mathbb{E}(X)$, *если* $\mathbb{P}(X \leqslant E(X) - x) = \mathbb{P}(X \geqslant \mathbb{E}(X) + x)$ *для любого* x *).*
- *Пусть* X *–* ***непрерывная*** *случайная величина. Если* $\mathrm{As}(X) > 0$, *то плотность распределения вероятностей (метод pdf)* $f_X(x)$ ***"скошена влево"*** *относительно* $\mathbb{E}(X)$; *Если же* $\mathrm{As}(X) < 0$, *то плотность распределения вероятностей (метод pdf)* $f_X(x)$ ***"скошена вправо"***.

Неформально говоря, коэффициент асимметрии положителен, если ***"правый хвост"*** распределения ***"длиннее левого"***, и отрицателен в противном случае.

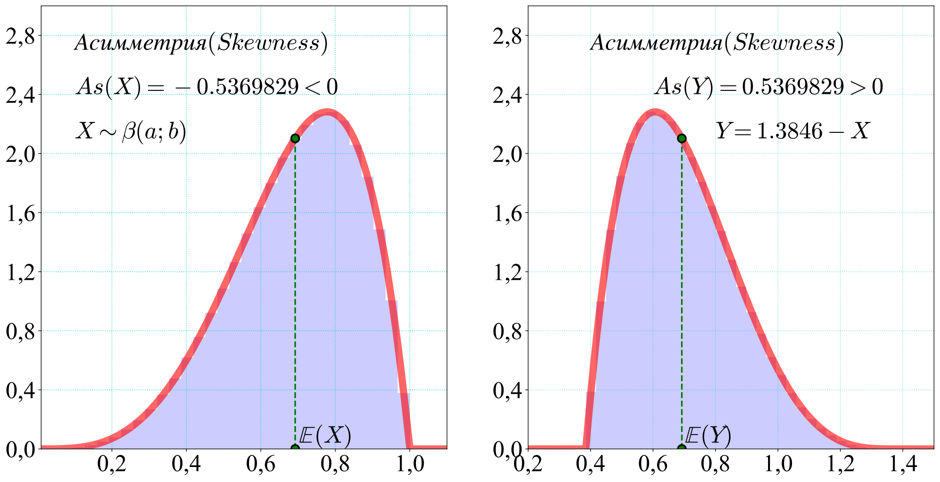

Рис.1. *Асимметрия непрерывного распределения (The moment coefficient of* ***skewness*** *).*

Упражнение

- Пусть $X \sim Bin(n;p)$. Покажите, что $\mathrm{As}(X) = \frac{1-2p}{\sqrt{npq}}$, $q = 1 - p$;
- Пусть $X \sim \Pi(\lambda)$. Покажите, что $\mathrm{As}(X) = \frac{1}{\sqrt{\lambda}}$.

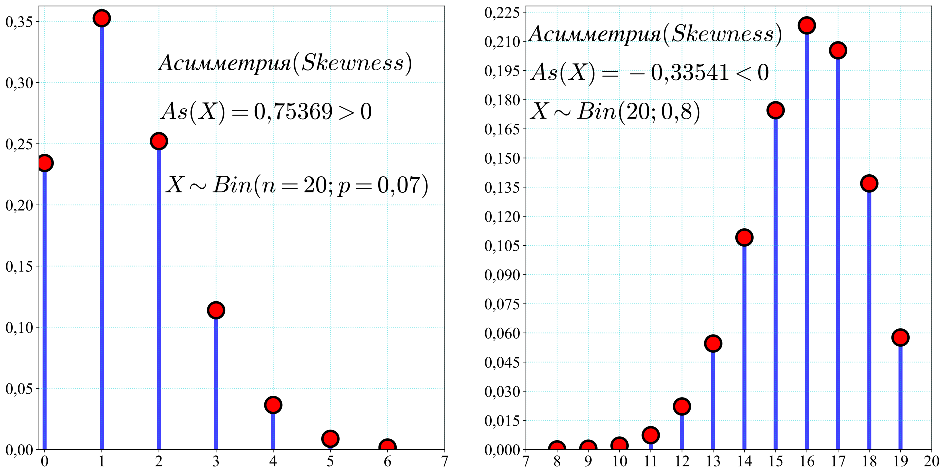

Рис.2. *Асимметрия дискретного распределения распределения (The moment coefficient of* ***skewness****).*

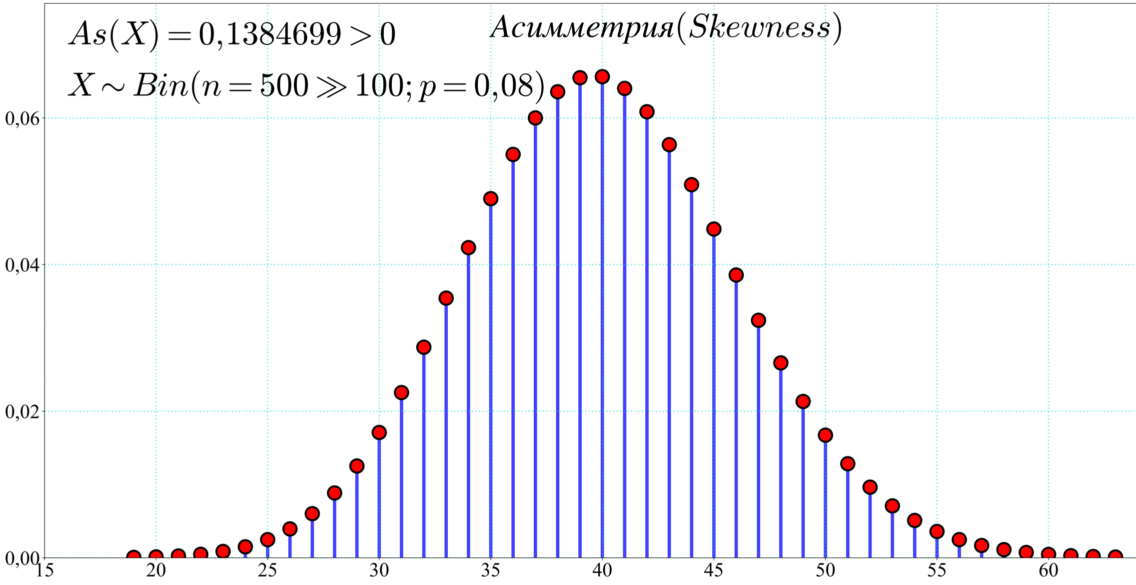

Рис.3. *Асимметрия дискретного распределения распределения (The moment coefficient of* ***skewness****).*

Эксцесс распределения (Kurtosis Excess)

Определение

Эксцессом (Kurtosis Excess) *(избыточный куртозис) распределения случайной величины X называется величина*

$$\mathrm{Ex}(X) = \frac{\mu_4(X)}{(\sigma(X))^4} - 3.$$

Kurtosis *(от греческого "куртос" выгнутый) распределения определяется по формуле*

$$Kurt(X) = \frac{\mu_4(X)}{(\sigma(X))^4}$$

Куртозис характеризует "тяжесть хвостов распределения" (tailedness).

Упражнение

- Пусть $X \sim Bin(n;p)$. Тогда

$$\mathrm{Ex}(X) = \frac{1-6pq}{npq}, q = 1 - p;$$

- Пусть $X \sim \Pi(\lambda)$. Тогда

$$\mathrm{Ex}(X) = \frac{1}{\lambda}.$$

- Для любого распределения с конечным центральным моментом четвертого порядка выполняются неравенства

$$a)\, \mathrm{Ex}(X) \geqslant -2;$$
$$b) Kurt(X) \geqslant [\mathrm{As}(X)]^2 + 1$$

Свойства асимметрии и эксцесса распределения

Теорема

- ***Асимметрия*** *и* ***эксцесс*** *распределения инвариантны относительно линейной замены случайной величины:*

$$\mathrm{As}(aX+b)=\mathrm{As}(X), a>0;$$
$$\mathrm{Ex}(aX+b)=\mathrm{Ex}(X).$$

- *Для независимых случайных величин* $X_1, X_2, \ldots, X_n$ *справедливы соотношения:*

$$a)\ \mathrm{As}(X_1+X_2+\ldots+X_n)=a_1\,\mathrm{As}(X_1)+\ldots+a_n\,\mathrm{As}(X_n);$$
$$b)\ \mathrm{Ex}(X_1+X_2+\ldots+X_n)=b_1\,\mathrm{Ex}(X_1)+\ldots+b_n\,\mathrm{Ex}(X_n),$$

где $a_k=\frac{\sigma^3(X_k)}{[\sigma^2(X_1)+\ldots+\sigma^2(X_n)]^{\frac{3}{2}}}$, $b_k=\frac{\sigma^4(X_k)}{[\sigma^2(X_1)+\ldots+\sigma^2(X_n)]^2}$.

Свойства асимметрии и эксцесса распределения

Следствие

Если все $X_1, X_2, \ldots, X_n$ – независимые случайные величины и имеют ***одинаковое*** *распределение, то*

$$a)\ \mathrm{As}(X_1 + \ldots X_n) = \frac{1}{\sqrt{n}}\,\mathrm{As}(X_1) \to 0;$$

$$b)\ \mathrm{Ex}(X_1 + \ldots X_n) = \frac{1}{n}\,\mathrm{Ex}(X_1) \to 0 \quad \text{при} \quad n \to +\infty.$$

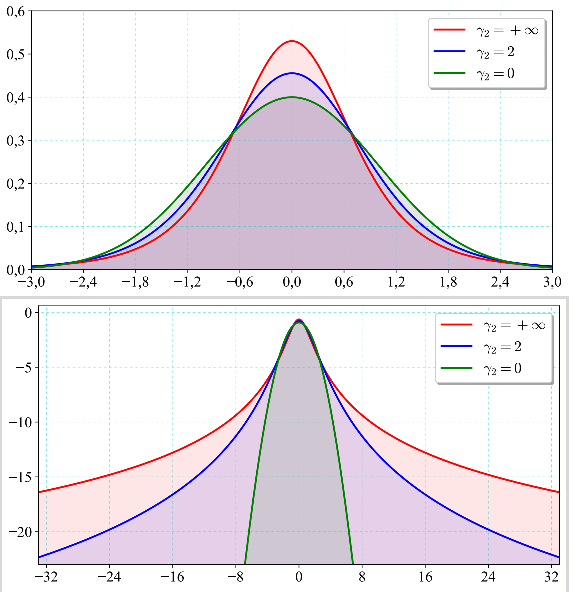

Рис.4. *Эксцесс непрерывного распределения* (**Kurtosis Excess**) ($W.pdf(x)$ и $W.logpdf(x)$.