Схема Бернулли

Определение.

Схемой Бернулли называется последовательность независимых испытаний, в каждом из которых возможны лишь два исхода —«успех» и «неудача», при этом «успех» в одном испытании происходит с вероятностью p, «неудача» — с вероятностью $q = 1 - p$.

Теорема. (Формула Бернулли).

Обозначим через ν_n число успехов в n испытаниях схемы Бернулли. Тогда для любого $k = 0, 1, \ldots, n$

$$\mathsf{P}(\nu_n = k) = C_n^k\, p^k\, (1-p)^{n-k} = C_n^k\, p^k\, q^{n-k}.$$

Доказательство. Событие

$$A = \{\nu_n = k\}$$

Означает, что в серии из n испытаний произошло ровно k успехов.

Рассмотрим один из плагоприятствующих событию A исходов.

$$(\underbrace{у, у, \ldots, у}_{k},\ \underbrace{н, \ldots, н}_{n-k})$$

(у – «успех», н – «неудача»)

Т.к. испытания независимы, то вероятность такого элементарного исхода равна

$$p^k \ (1-p)^{n-k}$$

Первые k испытаний завершились успехом, а остальные $(n - k)$ - неудачей.

Другие, благоприятствующие нашему событию исходы, отличаются от данного лишь иным расположением $\boldsymbol{k}$ успехов по $\boldsymbol{n}$ местам.

Число благоприятствующих исходов равно числу сочетаний из $\boldsymbol{n}$ элементов по $\boldsymbol{k}$, а вероятность события A равна сумме вероятностей всех элементарных событий, составляющих данное.

- Определение. Набор чисел

$$\left\{C_n^k\, p^k\, (1-p)^{n-k}, k = 0, 1, \ldots, n\right\}$$

Называется биномиальным распределением вероятностей и обозначается

$$\mathbf{B}_{n,p} \text{ или } \mathbf{B}(n,p)$$

$$(p+q)^n = C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + \ldots + C_n^n p^n q^0$$

Наиболее вероятное число успехов

$$\frac{\mathsf{P}(\nu_n = k)}{\mathsf{P}(\nu_n = k-1)} = \frac{n!}{k!(n-k)!}\frac{(k-1)!(n-k+1)!}{n!}\frac{p^k q^{n-k}}{p^{k-1}q^{n-k+1}} =$$

$$= \frac{(n-k+1)p}{kq} = 1 + \frac{(n-k+1)p}{kq} - 1 = 1 + \frac{np+p-k}{kq}.$$

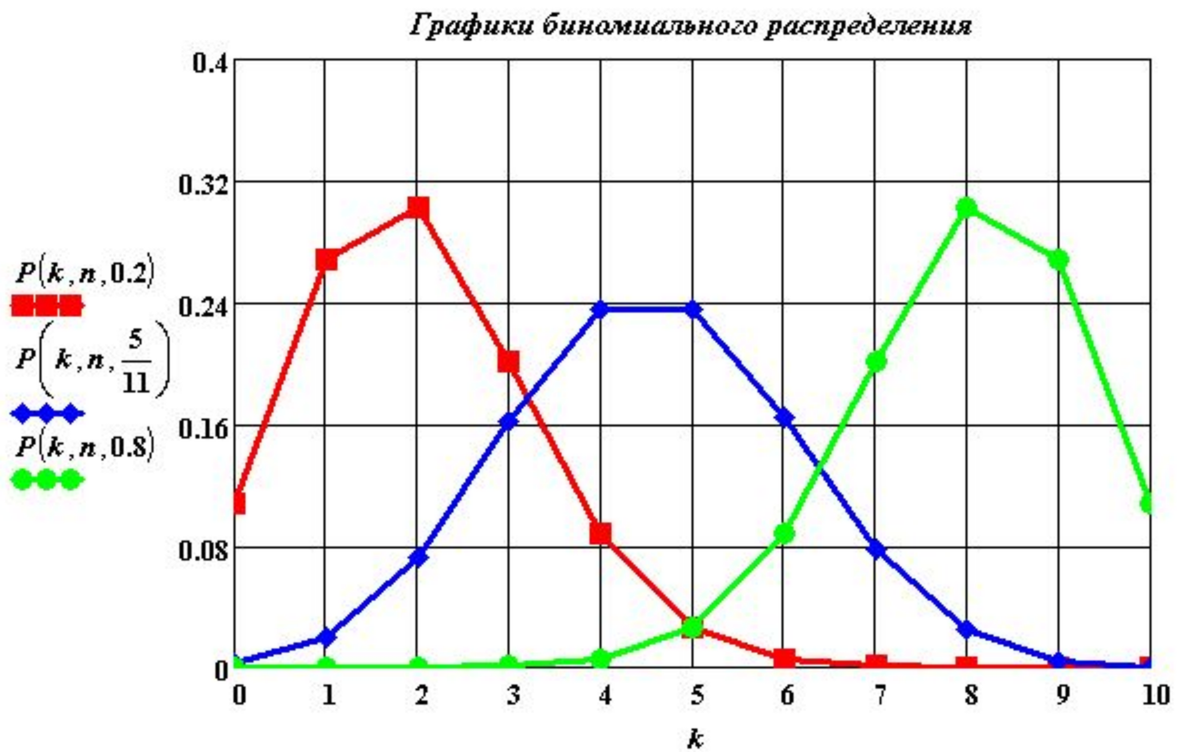
Видим, что

(a) $\mathsf{P}(\nu_n = k) > \mathsf{P}(\nu_n = k-1)$ при $np+p-k > 0$, то есть при $k < np+p$;

(b) $\mathsf{P}(\nu_n = k) < \mathsf{P}(\nu_n = k-1)$ при $np+p-k < 0$, то есть при $k > np+p$;

(c) $\mathsf{P}(\nu_n = k) = \mathsf{P}(\nu_n = k-1)$ при $np+p-k = 0$, что возможно лишь если $np+p$ — целое число.

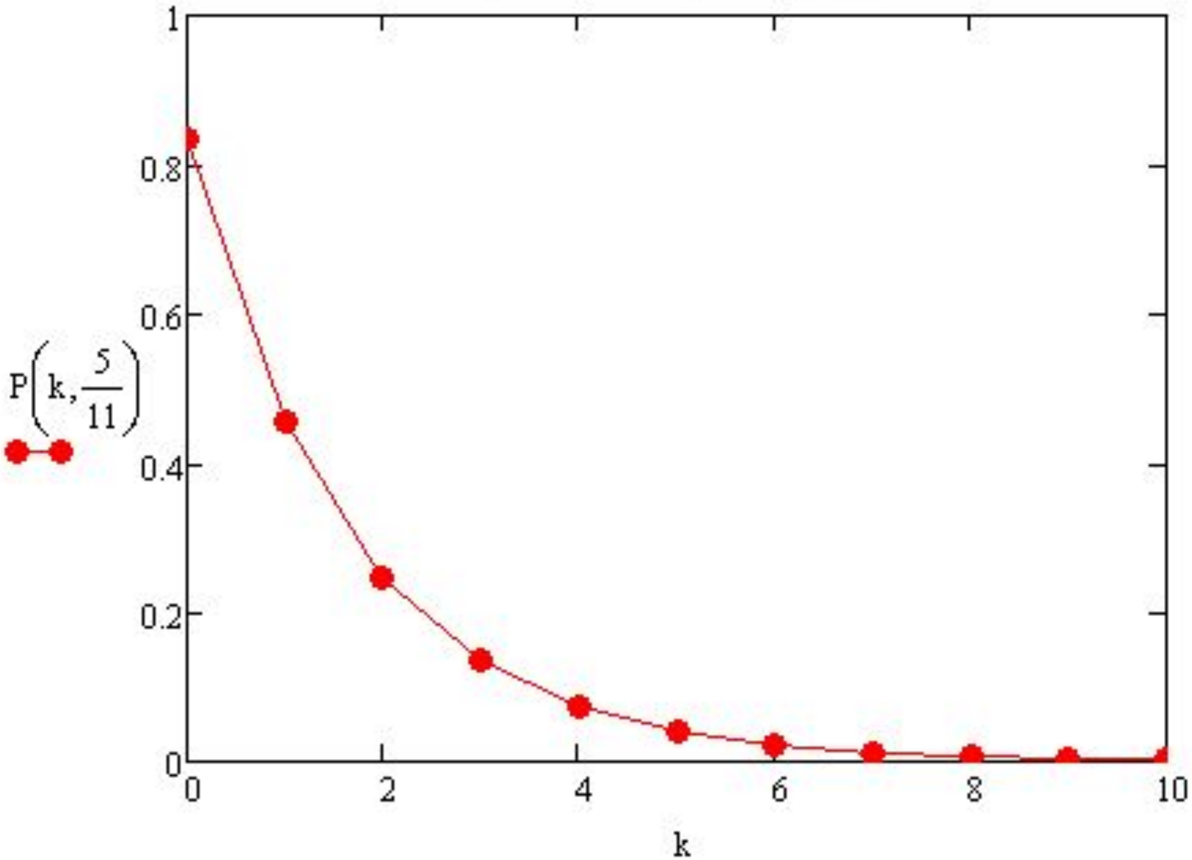
$$np - q < k^* < np + p$$



Графики биномиального распределения
0.4
0.32
0.24
0.16
0.08
0
0
1
2
3
4
5
6
7
8
9
10
k
P(k, n, 0.2)
P(k, n, 5/11)
P(k, n, 0.8)

$n := 10 \quad k := 0..n \qquad P(k,p) := p\cdot(1-p)^{k-1}$

Геометрическое распределение



Номер первого успешного испытания в схеме Бернулли

Введем величину τ, принимающую

значения из $\{1, 2, 3, \dots\}$, равную *номеру первого успешного испытания*.

Теорема *Вероятность того, что первый успех произойдет в испытании с номером* $k \in \mathbb{N} = \{1, 2, 3, \dots\}$, *равна* $\mathsf{P}(\tau = k) = p\, q^{k-1}$.

Доказательство. Действительно, $\mathsf{P}(\tau = k) = \mathsf{P}(\underbrace{н,\ н,\ \dots,\ н}_{k-1}, у) = p\, q^{k-1}$.

Определение Набор чисел $\{p\, q^{k-1},\ k = 1, 2, 3, \dots\}$ называется *геометрическим распределением вероятностей* и обозначается $\mathbf{G}_p$ или $\mathbf{G}(p)$.

Выбор без возвращения

Рассмотрим урну, содержащую N шаров, из которых K шаров — белые, $N - K$ шаров — черные.

Из урны наудачу выбирают n шаров

Вероятность $P_{N,K}(n,k)$ того, что будет выбрано ровно k белых и $n-k$ черных находится по формуле

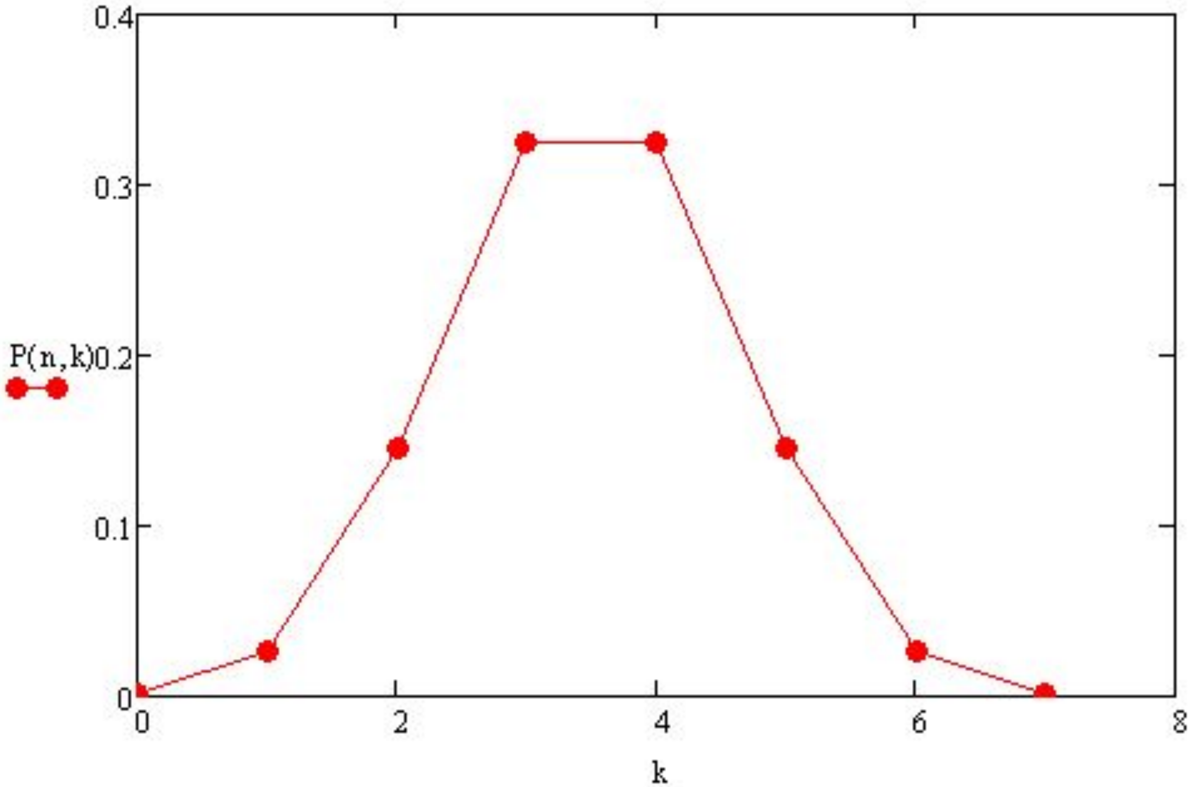
$$P_{N,K}(n,k) = \frac{C_K^k \, C_{N-K}^{n-k}}{C_N^n}.$$

Такое распределение вероятностей называется *гипергеометрическим*

$N := 20 \qquad K := 7 \qquad P(n,k) := C(K,k) \cdot \frac{C(N-K, n-k)}{C(N,n)}$

$k := 0..n$

Гипергеометрическое распределение вероятностей



Предельное поведение гипергеометрического распределения

Теорема *Если* $N \to \infty$ *и* $K \to \infty$ *так, что* $K/N \to p \in (0,1)$, *то для любых фиксированных* n, $0 \leqslant k \leqslant n$

$$P_{N,K}(n,k) = \frac{C_K^k \, C_{N-K}^{n-k}}{C_N^n} \to C_n^k \, p^k \, (1-p)^{n-k}.$$

Независимые испытания с несколькими исходами

Пусть в одном испытании возможны m исходов. Обозначим их цифрами $1, 2, \ldots, m$. Пусть исход i в одном испытании случается с вероятностью p_i, $1 \leqslant i \leqslant m$, и $\sum\limits_{1}^{m} p_i = 1$. Обозначим через $P(n_1, \ldots, n_m)$ вероятность того, что в $n = n_1 + \ldots + n_m$ независимых испытаниях исход 1 появился n_1 раз, исход 2 — n_2 раз, ... , исход m — n_m раз.

Теорема *Для любого n и любых целых $n_1 \geqslant 0, \ldots, n_m \geqslant 0$ таких, что $n_1 + \ldots + n_m = n$, верна формула:*

$$P(n_1, \ldots, n_m) = \frac{n!}{n_1! \ldots n_m!} \, p_1^{n_1} \cdot \ldots \cdot p_m^{n_m}.$$

Полиномиальное распределение

Предельные теоремы с схеме Бернулли

Теорема Пуассона

Предположим, нам нужна вероятность получить не менее десяти успехов в 1000 испытаний схемы Бернулли с вероятностью успеха 0.003. Вероятность этого события равна любому из следующих выражений:

$$\sum_{k=10}^{1000} C_{1000}^k \ (0.003)^k \ (0.997)^{1000-k} \ = \ 1 - \sum_{k=0}^{9} C_{1000}^k \ (0.003)^k \ (0.997)^{1000-k},$$

и вычисление даже одного слагаемого в каждом из этих выражений весьма проблематично.

Теорема (Теорема Пуассона).
Пусть $n \to \infty$, $p_n \to 0$ *так, что* $np_n \to \lambda > 0$. *Тогда для любого* $k \geqslant 0$ *вероятность получить* k *успехов в* n *испытаниях схемы Бернулли с вероятностью успеха* p_n *стремится к величине* $\dfrac{\lambda^k}{k!}\, e^{-\lambda}$:

$$\mathsf{P}(\nu_n = k) = C_n^k \, p_n^k \, (1-p_n)^{n-k} \ \to \ \frac{\lambda^k}{k!}\, e^{-\lambda} \ \text{ при } \ n \to \infty, p_n \to 0 \ \text{ так, что } np_n \to \lambda > 0.$$

Доказательство. Положим $\lambda_n = n \cdot p_n \to \lambda > 0$. $C_n^k \sim \dfrac{n^k}{k!}$ при фиксированном k и при $n \to \infty$. Тогда

$$C_n^k \, p_n^k \, (1-p_n)^{n-k} = C_n^k \, \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \sim \frac{\cancel{n}^k}{k!} \frac{\lambda_n^k}{\cancel{n}^k} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^n}_{\downarrow \atop e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-k}}_{\downarrow \atop 1} \to \frac{\lambda^k}{k!} \, e^{-\lambda}.$$

Определение 23. Пусть $\lambda > 0$ — некоторая постоянная. Набор чисел

$$\left\{ \frac{\lambda^k}{k!} \, e^{-\lambda}, k = 0, 1, 2, \ldots \right\}$$

называется *распределением Пуассона с параметром* λ.

Теорема 17 (Теорема Пуассона с оценкой погрешности).

Пусть $A \subseteq \{0, 1, 2, \ldots, n\}$ *— произвольное множество целых неотрицательных чисел,* ν_n *— число успехов в* n *испытаниях схемы Бернулли с вероятностью успеха* p, $\lambda = n \cdot p$. *Тогда*

$$\left| \mathsf{P}(\nu_n \in A) - \sum_{k \in A} \frac{\lambda^k}{k!}\, e^{-\lambda} \right| = \left| \sum_{k \in A} C_n^k\, p^k\, (1-p)^{n-k} - \sum_{k \in A} \frac{\lambda^k}{k!}\, e^{-\lambda} \right| \leqslant np^2 = \frac{\lambda^2}{n}.$$

Предельная теорема Муавра-Лапласа

Пусть A — *событие, которое может произойти в любом из* n *независимых испытаний с одной и той же вероятностью* $p = \mathsf{P}(A)$. *Пусть* $\nu_n(A)$ — *число осуществлений события* A *в* n *испытаниях. Тогда* $\dfrac{\nu_n(A) - np}{\sqrt{np(1-p)}} \Rightarrow \mathbf{N}_{0,1}$. *Иначе говоря, для любых вещественных* $x < y$ *при* $n \to \infty$ *имеет место сходимость*

$$\mathsf{P}\left(x \leqslant \frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \leqslant y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt;$$