Теория вероятностей и математическая статистика

Кракашова Ольга Анатольевна

доцент, канд. экон. наук,
доцент кафедры «Статистики, эконометрики и оценки рисков»
РГЭУ (РИНХ)

Лекция № 6

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН, СВЯЗАННЫЕ С ПОВТОРНЫМИ ИСПЫТАНИЯМИ

Биномиальное распределение (схема Бернулли)

- **Пример 1.** Монета подбрасывается 4 раза, пусть X - означает число появившихся гербов.
- **Пример 2.** Известно, что в определенном городе 30% горожан предпочитают добираться на работу личным автомобилем. Случайно выбраны 8 человек. Пусть Y - число людей в выборке, предпочитающих личный транспорт.
- **Пример 3.** Известно, что 15% деталей, произведенных автоматом, - бракованные. В порядке случайного отбора взято 12 деталей. Пусть Z - число дефектных деталей.
- Что характерно для случайных величин X, Y и Z?
- **Это – примеры ДСВ, подчиняющихся вероятностному закону распределения, известному как биномиальное распределение.**
- **Биномиальное распределение базируется на эксперименте, состоящем в последовательности испытаний Бернулли (схеме повторных испытаний).**

- **Испытания Бернулли** - это последовательность *n* идентичных испытаний, удовлетворяющих следующим условиям:
- 1. Каждое испытание имеет два исхода, называемые успех и неуспех. Эти два исхода - взаимно несовместные и противоположные события.
- 2. Вероятность успеха, обозначаемая *p*, остается постоянной от испытания к испытанию. Вероятность неуспеха обозначается *q*, где $q=1 - p$.
- 3. Все *n* испытаний - независимы. То есть вероятность наступления события в любом из испытаний не зависит от результатов других испытаний.
- Случайная величина, для которой вычисляется число успехов в *n* повторных испытаниях, где *p* - вероятность успеха в любом из заданных испытаний, а $q = (1 - p)$ -соответствующая вероятность неуспеха, подчиняется закону биномиального распределения с параметрами *n* и *p*.

- **Пример 1.** Монета подбрасывается 4 раза, пусть ДСВ X - означает число появившихся гербов. При четырех подбрасываниях монеты случайная величина X - число выпадений герба, принимает возможные значения X_i: *1,2,3,4*.
- Рассмотрим определенное событие, когда $X = 2$. Событие состоит в том, что при 4-х подбрасываниях монеты 2 раза выпадет герб.
- Определим вероятность этого события, то есть $P(X=2)$. Подсчитаем, сколькими способами может осуществиться данное событие.
- При 4-х бросаниях монеты герб появится два раза в одной из следующих 6-ти последовательностей:

ГГЦЦ, ГЦГЦ, ГЦЦГ, ЦГГЦ, ЦГЦГ, ЦЦГГ.

- Исходя из независимости 4-х испытаний, вероятность любой последовательности, скажем (Г, Г, Ц, Ц) есть ppqq. Очевидно, что порядок появления цифры или герба не влияет на вероятность. Вероятность p^2q^2 есть вероятность для любой из 6-ти перечисленных комбинаций.
- Поскольку все шесть возможных комбинаций ведут к событию X=2, то умножим результат на 6, что даст нам $6p^2q^2$. Для идеальной монеты $p = q = 0,5$; отсюда $P(X=2) = 6(0,5)^4 = 0,375$.
- Точно также можно вычислить другие вероятности P(X=0), P(X=1), P(X=3), P(X=4).

- Обобщим процедуру вычисления вероятности появления некоторого события точно *m* раз в *n* последовательных испытаниях, удовлетворяющих условиям схемы повторных испытаний:
- 1. Вероятность любой заданной последовательности, в которой событие появляется *m* раз в *n* испытаниях с вероятностью успеха в каждом отдельном испытании *p* и *q* - неуспеха, равна $p^m q^{n-m}$.
- Для опыта с подбрасыванием монеты при p=q=0,5, n= 4 и m= 2 получим
- $P(X=2) = (0{,}5)^2(0{,}5)^2 = (0{,}5)^4$.
- 2. Число различных комбинаций в испытаниях, в результате которых наступит точно *m* успехов, равно числу сочетаний из *n* элементов по *m* элементов в каждом:

$$C_n^m = \frac{A_n^m}{P_m} \quad \text{или} \quad \frac{n!}{m!(n-m)!}$$

Для примера с подбрасыванием монеты

$$C_4^2 = \frac{4\cdot 3}{1\cdot 2} = 6 \quad \text{или} \quad \frac{4!}{2!2!}$$

- 3. Поскольку существует C_n^m комбинаций и каждая комбинация имеет вероятность $p^m q^{n-m}$, то вероятность m успехов в n испытаниях есть результат двух описанных выше действий.
- Будем использовать символ $P_{n,m}$ для обозначения вероятности $P(X=m)$ в n испытаниях с вероятностью успеха в каждом отдельном испытании – p:

$$P(X=m) = P_{n,m} = C_n^m p^m q^{n-m} =$$

$$= \left(\frac{n!}{m!(n-m)!}\right) p^m q^{n-m} \qquad (1)$$

где p - вероятность успеха в отдельном испытании, $q=1-p$, n - число испытаний, m - число успешных испытаний.

- **Это - формула Бернулли**.
- В формуле (1) m может принимать значения от 0 до n. Подставим $m=0,1,2,\ldots,n$ в (1).
- Построим биномиальное распределение числа выпадений герба при 4-х подбрасываниях монеты.

X =m			$P(x) = P_{4,m}$
0	$C_n^0\, p^0 q^n$	$\dfrac{4!}{0!(4-0)!} \cdot 0{,}5^0 \cdot 0{,}5^4$	0,0625
1	$C_n^1\, p^1 q^{n-1}$	$\dfrac{4!}{1!(4-1)!} \cdot 0{,}5^1 \cdot 0{,}5^{4-1}$	0,2500
2	$C_n^2\, p^2 q^{n-2}$	$\dfrac{4!}{2!(4-2)!} \cdot 0{,}5^2 \cdot 0{,}5^{4-2}$	0,3750
3	$C_n^3\, p^3 q^{n-3}$	$\dfrac{4!}{3!(4-3)!} \cdot 0{,}5^3 \cdot 0{,}5^{4-3}$	0,2500
4	$C_n^4\, p^4 q^{n-4}$	$\dfrac{4!}{4!(4-4)!} \cdot 0{,}5^4 \cdot 0{,}5^{4-4}$	0,0625

- Рассмотрим в качестве СВ X число m наступления некоторого события в n независимых испытаниях. Очевидно, общее число появлений этого события в испытаниях состоит из суммы чисел появлений события в отдельных испытаниях, то есть $X=m=X_1+X_2+...+ +X_n$, где X_i - число появлений события в i-том испытании $(i=1,2,...,n)$. Так как вероятность наступления события в каждом испытании постоянна и равна p (не наступления - q), то для каждой случайной величины X_i имеем распределение вероятностей:

x_i	**0**	**1**
p_i	**q**	**p**

Следовательно,

$$M(X_1) = M(X_2) = ... = M(X_n)$$

$$M(X_i) = 0 \cdot q + 1 \cdot p = p$$

Исходя из формулы математического ожидания СВ, получим:

$$M(X) = M(X_1) + M(X_2) + ... + M(X_n) =$$

$$= \sum_{i=1}^{n} M(X_i) = np$$

Математическое ожидание частоты биномиального распределения

$$M(X) = np \qquad (2)$$

Аналогично рассуждая и применяя формулу дисперсии СВ получим:

$$D(X_i) = M(X^2{}_i) - M^2(X_i) = 0^2\, q + 1^2\, p - p^2 = p(1\text{-}p) = pq$$

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n) = \sum_{i=1}^{n} D(X_i) = npq$$

Дисперсия частоты биномиального распределения:

$$\sigma^2 = D(X) = npq \qquad (3)$$

Стандартное отклонение биномиального распределения:

$$\sigma = \sqrt{npq}$$

- Используя формулы (2) и (3), найдем математическое ожидание и дисперсию случайной величины X - числа появления гербов при 4-х подбрасываниях монеты: $M(X) = np = 4 \cdot 0,5 = 2$.
- Полученное значение интуитивно понятно и без вычислений. При достаточно большой серии испытаний по четыре подбрасывания монеты, мы ожидаем, что в среднем при четырех подбрасываниях монеты приходится два герба.
- Дисперсия X есть $npq = 4 \cdot 0,5 \cdot 0,5 = 1,00$.

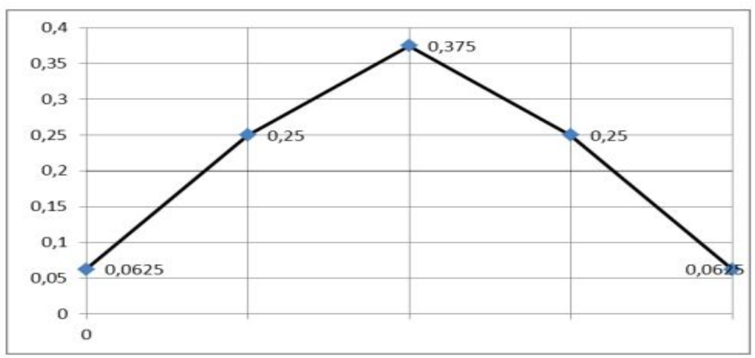


Рис.1. График биномиального распределения для примера 1

- Как видно из графика на рисунке 1 при $m=2$ вероятность достигает максимального значения.
- Частота m, равная 2, называется **вероятнейшим числом или вероятнейшей частотой (наивероятнейшей)**.
- **Вероятнейшей частотой наступления события** называется та частота, при которой вероятность достигает своего наибольшего значения и обозначается m_0:

$$np - q \le m_0 \le np + p \qquad (5)$$

- В этом неравенстве m0 может быть только целым числом.
- **Замечание:** Если np - целое число, то $m_0 = np$.
- **Пример 2.** Вероятность того, что выписанный продавцом чек будет оплачен, равна $0,9$. Какое наивероятнейшее число чеков будет оплачено, если выписано 40 чеков?
- Находим произведение $np=40\cdot 0,9=36$ (целое число), значит, $m_0=36$.
- Найдем m0 по формуле (5): $40\cdot 0,9 - 0,1 \le m_0 \le 40\cdot 0,9 + 0,9$

$$35,9 \le m_0 \le 36,9$$

- Какое целое число удовлетворяет этому двойному неравенству? $m_0=36$.

Распределение Пуассона

Если вероятность появления события **А** в n отдельных независимых испытаниях очень мала ($p<q$), то для вычисления вероятности применяется формула Пуассона:

$$P_{n,m} \approx \frac{\lambda^m}{m!} e^{-\lambda} \qquad (6)$$

где $\lambda = np$, n - число независимых испытаний с постоянной малой вероятностью p, e - основание натуральных логарифмов (e=2,71828), m - число появлений события ($m_i=0,1,2,3,\ldots$).

Для представления закона распределения Пуассона в виде ряда распределения придадим m целые неотрицательные значения $m= 0,1,2,\ldots,n$ и вычислим соответствующие им вероятности $P_{n,m}$:

Таблица 1.
Закон распределения Пуассона

m	0	1	2	3	...	k	...	n
Pn,m	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	$\frac{\lambda^3 e^{-\lambda}}{3!}$	...	$\frac{\lambda^k e^{-\lambda}}{k!}$	...	$\frac{\lambda^n e^{-\lambda}}{n!}$

Сумма вероятностей ряда равна 1.

Закон распределения Пуассона можно записать также в виде функции распределения

$$F(x) = P(m < x) = \sum_{m<x} P_{n,m} = \sum_{m<x} \frac{\lambda^m}{m!} e^{-\lambda} \qquad (7)$$

где знак $\sum\limits_{m<x}$ означает сумму вероятностей $\boldsymbol{P_{n,m}}$ для всех $\boldsymbol{m}$, меньших $\boldsymbol{n}$.

Поскольку вероятности $P_{n,m\geq 1}$ и $P_{n,0}$ есть вероятности противоположных событий, то

$$P_{n,m\geq 1} = 1 - P_{n,0} = 1 - \frac{\lambda^0}{0!}e^{-\lambda} = 1 - e^{-\lambda},$$

или

$$P_{n,m\geq 1} = 1 - e^{-\lambda} \quad (8)$$

Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, совпадают и равны параметру λ, который определяет этот закон, т.е.

$$M(X) = D(X) = \lambda \qquad (9)$$

Пример 3. Нас интересует число инкассаторов, прибывающих утром на автомобиле в банк в течение 15 минут. Если предположить, что вероятность прибытия автомобиля одинакова в любые два периода времени равной длины, и что прибытие или неприбытие автомобиля в любой период времени не зависит от прибытия или неприбытия в любой другой период времени, то последовательность прибытия инкассаторов в банк может быть описана распределением Пуассона.

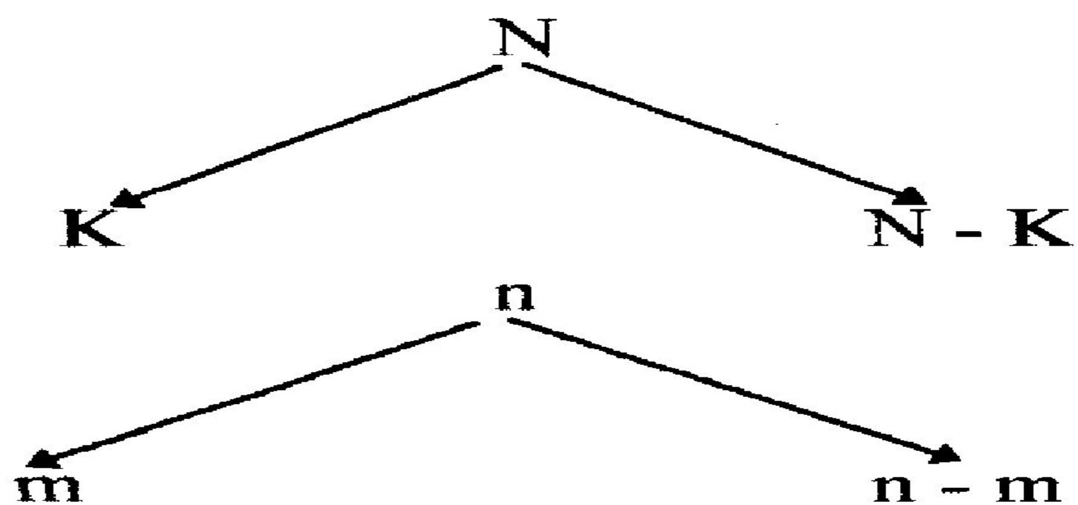
РЕШЕНИЕ. Анализ предыдущих данных показал, что среднее число инкассаторов, прибывающих в 15-ти минутный период, равно 10, тогда при $\lambda=10$ получаем: $\boldsymbol{P(m)=\lambda^m e^{-\lambda}/ m!} =$
$= \boldsymbol{10^m e^{-10}/m!}$, при $\boldsymbol{m=0,1,2,...}$

Если необходимо узнать вероятность прибытия пяти инкассаторов в течение 15 минут, то при $\boldsymbol{m}$ = 5 получим:

$\boldsymbol{P(5) = 10^5 e^{-10}/5! = 0{,}0378.}$

Гипергеометрическое распределение

- Вероятность появления события ровно *m* раз в *n* независимых повторных испытаниях вычисляется по формулам Бернулли и Пуассона. Как вычисляются вероятности появления события ровно *m* раз в *n* зависимых повторных испытаниях?
- **Пример 2.** В урне *N* шаров, среди которых *K* белых и *(N-K)* черных. Без возвращения извлечены *n* шаров. Какова вероятность того, что в выборке из *n* шаров окажется *m* белых (и соответственно *(n-m)* черных) шаров. Изобразим ситуацию на схеме:



СВ $X = m$ - число белых шаров в выборке объемом в n шаров. Число всех возможных случаев отбора n шаров из N равно числу сочетаний из N по n (C_N^n), а число случаев отбора m белых шаров из имеющихся K белых шаров (*и значит,* $(n-m)$ черных шаров из $(N-K)$ *имеющихся черных*) равно произведению:

$$C_k^m \cdot C_{N-K}^{n-m}$$

Отбор каждого из m белых шаров может сочетаться с отбором любого из (**n-m**) черных.

Необходимо определить вероятность того, что в выборке из n шаров окажется ровно m белых шаров. По формуле для вероятности события в классической модели, вероятность получения в выборке m белых шаров (то есть вероятность того, что случайная величина X примет значение m) равна

$$P_{m,N} = \frac{C_k^m \cdot C_{N-k}^{n-m}}{C_N^n} \qquad (10)$$

Где C_N^n - общее число всех единственно возможных, равновозможных и несовместных исходов, $C_k^m C_{N-K}^{n-m}$ - число исходов благоприятствующих интересующему нас событию, $\boldsymbol{m \leq n}$, если $\boldsymbol{n \leq k}$ и $\boldsymbol{m \leq k}$, если $\boldsymbol{k < n}$.
Итак, вероятность появления интересующего нас события ровно $\boldsymbol{m}$ раз в $\boldsymbol{n}$ зависимых испытаниях вычисляется по формуле (10), которая задает значения **гипергеометрического распределения для** $\boldsymbol{m = 0,1,2,..,n}$**. -** распределения вероятностей значений случайной величины в $\boldsymbol{n}$ повторных зависимых испытаниях.

Если по формуле (10) вычислить вероятности для всех возможных значений m и поместить их в таблицу, то получим ряд распределения, называемый **гипергеометрическим законом распределения**.

Таблица 2. Гипергеометрический закон распределения

M	0	1	2	3	...	n
P(X=m)	$\frac{C_k^0 \cdot C_{N-k}^n}{C_N^n}$	$\frac{C_k^1 . C_{N-k}^1}{C_N^n}$	$\frac{C_k^2 . C_{N-k}^2}{C_N^n}$	$\frac{C_k^3 . C_{N-k}^3}{C_N^n}$	...	$\frac{C_k^m \cdot C_{N-k}^0}{C_N^n}$

Математическое ожидание и дисперсия случайной величины ***m***, распределенной по гипергеометрическому закону, определяются формулами:

$$\boldsymbol{M(m) = n\cdot\theta} \qquad (11)$$

$$\boldsymbol{D(m) = n\cdot\theta\,(1 - \theta)\cdot(1 - \frac{n-1}{N-1})}, \qquad (12)$$

где $\boldsymbol{\theta}$ - доля единиц с интересующим нас признаком в совокупности **N**, т.е. $\boldsymbol{\theta} = \frac{K}{N}$, а

$(1-\frac{n-1}{N-1})$ - называется поправкой для бесповторной выборки.