

Биномиальное распределение

В настоящей и нескольких следующих заметках мы рассмотрим математические модели случайных событий. *Математическая модель* — это математическое выражение, представляющее случайную величину. Для дискретных случайных величин это математическое выражение известно под названием функция распределения.¹

Если задача позволяет явно записать математическое выражение, представляющее случайную величину, можно вычислить точную вероятность любого ее значения. В этом случае можно вычислить и перечислить все значения функции распределения. В деловых, социологических и медицинских приложениях встречаются разнообразные распределения случайных величин. Одним из наиболее полезных распределений является биномиальное.

Биномиальное распределение используется для моделирования ситуаций, характеризующихся следующими особенностями.

- Выборка состоит из фиксированного числа элементов n , представляющих собой исходы некоего испытания.
- Каждый элемент выборки принадлежит одной из двух взаимоисключающих категорий, исчерпывающих все выборочное пространство. Как правило, эти две категории называют успех и неудача.
- Вероятность успеха p является постоянной. Следовательно, вероятность неудачи равна $1 - p$.
- Исход (т.е. удача или неудача) любого испытания не зависит от результата другого испытания. Чтобы гарантировать независимость исходов, элементы выборки, как правило, получают с помощью двух разных методов. Каждый элемент выборки случайным образом извлекается из бесконечной генеральной совокупности без возвращения или из конечной генеральной совокупности с возвращением.

Биномиальное распределение используется для оценки количества успехов в выборке, состоящей из n наблюдений. Рассмотрим в качестве примера оформление заказов. Чтобы сделать заказ клиенты компании *Saxon Company* могут воспользоваться интерактивной электронной формой и послать ее в компанию. Затем информационная система проверяет, нет ли в заказах ошибок, а также неполной или недостоверной информации. Любой заказ, вызывающий сомнения, помечается и включается в ежедневный отчет об исключительных ситуациях. Данные, собранные компанией, свидетельствуют, что вероятность ошибок в заказах равна 0,1. Компания хотела бы знать, какова вероятность обнаружить определенное количество ошибочных заказов в заданной выборке. Например, предположим, что клиенты заполнили четыре электронных формы. Какова вероятность, что все заказы окажутся безошибочными? Как вычислить эту вероятность? Под успехом будем понимать ошибку при заполнении формы, а все остальные исходы будем считать неудачей. Напомним, что нас интересует количество ошибочных заказов в заданной выборке.

Какие исходы мы можем наблюдать? Если выборка состоит из четырех заказов, ошибочными могут оказаться один, два, три или все четыре, кроме того, все они могут оказаться правильно заполненными. Может ли случайная величина, описывающая количество неправильно заполненных форм, принимать какое-либо иное значение? Это невозможно, поскольку количество неправильно заполненных форм не может превышать объем выборки n или быть отрицательным. Таким образом, случайная величина, подчиняющаяся биномиальному закону распределения, принимает значения от 0 до n .

Допустим, что в выборке из четырех заказов наблюдаются следующие исходы:

Первый заказ **Второй заказ** **Третий заказ** **Четвертый заказ**
Ошибочный Ошибочный Правильный Ошибочный

Какова вероятность обнаружить три ошибочных заказа в выборке, состоящей из четырех заказов, причем в указанной последовательности? Поскольку предварительные исследования показали, что вероятность ошибки при заполнении формы равна 0,10, вероятности указанных выше исходов вычисляются следующим образом:

¹ Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 307–313

Первый заказ	Второй заказ	Третий заказ	Четвертый заказ
$p = 0,10$	$p = 0,10$	$1-p = 0,90$	$p = 0,10$

Поскольку исходы не зависят друг от друга, вероятность указанной последовательности исходов равна: $p * p * (1-p) * p = 0,1 * 0,1 * 0,9 * 0,1 = 0,0009$. Если же необходимо вычислить количество вариантов выбора X объектов из выборки, содержащей n элементов, следует воспользоваться формулой сочетаний (1):

$$(1) \frac{n!}{X!(n-X)!}$$

где $n! = n * (n-1) * (n-2) * \dots * 2 * 1$ — факториал числа n , причем $0! = 1$ и $1! = 1$ по определению.

Это выражение часто обозначают как $\binom{n}{X}$. Таким образом, если $n = 4$ и $X = 3$, количество последовательностей, состоящих из трех элементов, извлеченных из выборки, объем которой равен 4, определяется по следующей формуле:

$$\binom{n}{X} = \frac{n!}{X!(n-X)!} = \frac{4!}{3!(4-3)!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} = 4$$

Следовательно, вероятность обнаружить три ошибочных заказа вычисляется следующим образом:

$$\begin{aligned} & (\text{Количество возможных последовательностей}) * \\ & (\text{вероятность конкретной последовательности}) = 4 * 0,0009 = 0,0036 \end{aligned}$$

Аналогично можно вычислить вероятность того, что среди четырех заказов окажутся один или два ошибочных, а также вероятность того, что все заказы ошибочны или все верны. Однако при увеличении объема выборки n определить вероятность конкретной последовательности исходов становится труднее. В этом случае следует применить соответствующую математическую модель, описывающую биномиальное распределение количества вариантов выбора X объектов из выборки, содержащей n элементов.

Биномиальное распределение

$$(2) P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

где $P(X)$ — вероятность X успехов при заданных объеме выборки n и вероятности успеха p , $X = 0, 1, \dots, n$.

Обратите внимание на то, что формула (2) представляет собой формализацию интуитивных выводов. Случайная величина X , подчиняющаяся биномиальному распределению, может принимать любое целое значение в диапазоне от 0 до n . Произведение $p^X (1-p)^{n-X}$ представляет собой вероятность конкретной последовательности, состоящей из X успехов в выборке, объем которой равен n .

Величина $\frac{n!}{X!(n-X)!}$ определяет количество возможных комбинаций, состоящих из X успехов в n испытаниях. Следовательно, при заданном количестве испытаний n и вероятности успеха p вероятность последовательности, состоящей из X успехов, равна

$$\begin{aligned} P(X) &= (\text{количество возможных последовательностей}) * (\text{вероятность конкретной} \\ & \text{последовательности}) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \end{aligned}$$

Рассмотрим примеры, иллюстрирующие применение формулы (2).

1. Допустим, что вероятность неверно заполнить форму равна 0,1. Какова вероятность того, что среди четырех заполненных форм три окажутся ошибочными? Используя формулу (2), получаем, что вероятность обнаружить три ошибочных заказа в выборке, состоящей из четырех заказов, равна

$$P(X=3) = \frac{4!}{3!(4-3)!} (0,1)^3 (1-0,1)^{4-3} = \frac{4!}{3!(4-3)!} (0,1)^3 (0,9)^1 = 4 \times 0,1 \times 0,1 \times 0,1 \times 0,9 = 0,0036$$

2. Допустим, что вероятность неверно заполнить форму равна 0,1. Какова вероятность того, что среди четырех заполненных форм не менее трех окажутся ошибочными? Как показано в предыдущем примере, вероятность того, что среди четырех заполненных форм три окажутся ошибочными, равна 0,0036. Чтобы вычислить вероятность того, что среди четырех заполненных форм не менее трех будут неправильно заполнены, необходимо сложить вероятность того, что среди четырех заполненных форм три окажутся ошибочными, и вероятность того, что среди четырех заполненных форм все окажутся ошибочными. Вероятность второго события равна

$$P(X = 4) = \frac{4!}{4!(4-4)!} (0,1)^4 (1-0,1)^{4-4} = \frac{4!}{4!0!} (0,1)^4 (0,9)^0 = 1 \times 0,1 \times 0,1 \times 0,1 \times 0,1 = 0,0001$$

Таким образом, вероятность того, что среди четырех заполненных форм не менее трех окажутся ошибочными, равна

$$P(X > 3) = P(X = 3) + P(X = 4) = 0,0036 + 0,0001 = 0,0037$$

3. Допустим, что вероятность неверно заполнить форму равна 0,1. Какова вероятность того, что среди четырех заполненных форм менее трех окажутся ошибочными? Вероятность этого события

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

Используя формулу (2), вычислим каждую из этих вероятностей:

$$P(X = 0) = \frac{4!}{0!(4-0)!} (0,1)^0 (1-0,1)^{4-0} = 0,6561$$

$$P(X = 1) = \frac{4!}{1!(4-1)!} (0,1)^1 (1-0,1)^{4-1} = 0,2916$$

$$P(X = 2) = \frac{4!}{2!(4-2)!} (0,1)^2 (1-0,1)^{4-2} = 0,0486$$

Следовательно, $P(X < 3) = 0,6561 + 0,2916 + 0,0486 = 0,9963$.

Вероятность $P(X < 3)$ можно вычислить иначе. Для этого воспользуемся тем, что событие $X < 3$ является дополнительным по отношению к событию $X > 3$. Тогда $P(X < 3) = 1 - P(X > 3) = 1 - 0,0037 = 0,9963$.

По мере увеличения объема выборки n вычисления, аналогичные проведенным в примере 3, становятся затруднительными. Чтобы избежать этих сложностей, многие биномиальные вероятности табулируют заранее. Некоторые из этих вероятностей приведены рис. 1. Например, чтобы получить вероятность, что $X = 2$ при $n = 4$ и $p = 0,1$, следует извлечь из таблицы число, стоящее на пересечении строки $X = 2$ и столбца $p = 0,1$.

		P			
n	X	0,01	0,02	...	0,10
4	0	0,9606	0,9224		0,6561
	1	0,0388	0,0753		0,2916
	2	0,0006	0,0023		0,0486
	3	0,0000	0,0000		0,0036
	4	0,0000	0,0000		0,0001

Рис. 2. Биномиальная вероятность при $n = 4$, $X = 2$ и $p = 0,1$

Биномиальное распределение можно вычислить с помощью функции Excel =БИНОМ.РАСП() (рис. 2), имеющей 4 параметра: число успехов – X , число испытаний (или объем выборки) – n , вероятность успеха – p , параметр *интегральная*, принимающий значения ИСТИНА (в этом случае вычисляется вероятность **не менее** X событий) или ЛОЖЬ (в этом случае вычисляется вероятность **точно** X событий).

СУММ		X ✓ f_x		=БИНОМ.РАСП(
	A	B	C	D	E	F	G
1	Объем выборки	4					
2	Вероятность успеха	0,1					
3							
4	=БИНОМ.РАСП(
5	БИНОМ.РАСП(число_успехов; число_испытаний; вероятность_успеха; интегральная)						
6							

Рис. 2. Параметры функции =БИНОМ.РАСП()

Для вышеприведенных трех примеров расчеты приведены на рис. 3 (см. также Excel-файл). В каждом столбце приведено по одной формуле. Цифрами показаны ответы на примеры соответствующего номера).

	A	B	C	D	E	F
1	Объем выборки	4				
2	Вероятность успеха	0,1				
3						
4	Таблица биномиальных вероятностей					
5		X	P(X)	P(<=X)	P(<X)	P(>=X)
6		0	0,6561	0,6561	=D6-C6	1
7		1	0,2916	=БИНОМ.РАСП(\$B7;\$B\$1;\$B\$2;ИСТИНА)	0,6561	0,3439
8		2	=БИНОМ.РАСП(\$B8;\$B\$1;\$B\$2;ЛОЖЬ)	0,9963	0,9477	=1-E8
9		3	0,0036	0,9999	0,9963	0,0037
10		4	0,0001	1	0,9999	0,0001
11						

Рис. 3. Расчет биномиального распределения в Excel для $n = 4$ и $p = 0,1$

Свойства биномиального распределения

Биномиальное распределение зависит от параметров n и p . Биномиальное распределение может быть, как симметричным, так и асимметричным. Если $p = 0,05$, биномиальное распределение является симметричным независимо от величины параметра n . Однако, если $p \neq 0,05$, распределение становится асимметричным. Чем ближе значение параметра p к $0,05$ и чем больше объем выборки n , тем слабее выражена асимметрия распределения. Таким образом, распределение количества неправильно заполненных форм смещено вправо, поскольку $p = 0,1$ (рис. 4).

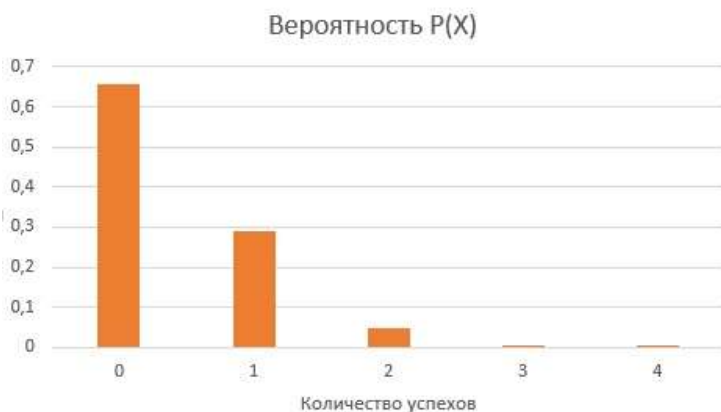


Рис. 4. Гистограмма биномиального распределения при $n = 4$ и $p = 0,1$

Математическое ожидание биномиального распределения равно произведению объема выборки n на вероятность успеха p :

$$(3) M = E(X) = np$$

В среднем, при достаточно долгой серии испытаний в выборке, состоящей из четырех заказов, может оказаться $p = E(X) = 4 \times 0,1 = 0,4$ неправильно заполненных форм.

Стандартное отклонение биномиального распределения

$$(4) \sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{np(1-p)}$$

Например, стандартное отклонение количества неверно заполненных форм в бухгалтерской информационной системе равно:

$$\sigma = \sqrt{4 * 0,1 * 0,9} = 0,60$$

Предыдущая заметка [Ковариация и ее применение в финансовом деле](#)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](#)