УДК 519.24

ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК И ВИЗУАЛИЗАЦИЯ БИНОМИАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ (С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМИЧЕСКОГО ЯЗЫКА R)

Борис Игнатьевич Смагин

доктор экономических наук, профессор

bismagin@mail.ru

Мичуринский государственный аграрный университет

г. Мичуринск, Россия

Аннотация. Биномиальный закон распределения имеет широкое применение в анализе случайных процессов. В статье дан анализ испытаний Бернулли, лежащих в основе формирования биномиального закона. Дана оценка математического ожидания и дисперсии, приведена локальная теорема Лапласа, применяемая при больших независимых испытаниях. Показано также применение алгоритмического языка R для вычисления вероятности и визуализации биномиального закона распределения.

Ключевые слова: испытания Бернулли, биномиальный закон, математическое ожидание, дисперсия, алгоритмический язык R, локальная и интегральная теоремы Лапласа.

При практическом применении теории вероятностей часто приходится встречаться с задачами, в которых один и тот же опыт или аналогичные опыты повторяются неоднократно. В результате каждого опыта может появиться или не появиться некоторое событие А, причем нас интересует не результат каждого отдельного опыта, а общее число появлений события А в результате серии опытов. Например, если производится группа выстрелов по одной и той же цели, нас, как правило, интересует не результат каждого выстрела, а общее число попаданий. В подобных задачах требуется уметь определять вероятность любого заданного числа появлений события в результате серии опытов.

Несколько опытов называются независимыми, если вероятность того или иного исхода каждого из опытов не зависит от того, какие исходы имели другие опыты. Например, несколько последовательных бросаний монеты представляют собой независимые опыты. Несколько последовательных извлечений карты из колоды представляют собой независимые опыты при условии, что вынутая карта каждый раз возвращается в колоду и карты перемешиваются; в противном случае это – зависимые опыты.

Рассмотрим независимые опыты, которые проводятся в одинаковых условиях. В этом случае вероятность события А во всех опытах одна и та же. Совокупность данных опытов принято называть испытаниями Бернулли.

Производится n независимых опытов, в каждом из которых может появиться или не появиться некоторое событие А; вероятность появления события А в каждом опыте равна p, а вероятность непоявления $q = 1 - p$. Требуется найти вероятность $P_{m,n}$ того, что событие А в этих n опытах появится ровно m раз.

Рассмотрим событие B_m, состоящее в том, что событие А появится в n опытах ровно m раз. Это событие может осуществиться различными способами. Разложим событие B_m на сумму произведений событий, состоящих в появлении или не появлении события А в отдельном опыте. Будем обозначать A_i появление события А в i-м опыте; $\overline{A}_i$ – непоявление события А в i-м опыте. Очевидно, каждый вариант появления события B_m (каждый член суммы) должен состоять из m появлений события A и n – m непоявлений, т. е. из m событий A и n – m событий $\overline{A}$ с различными индексами, причем в каждое произведение событие A должно входить m раз а $\overline{A}$ должно входить n – m раз. Число всех комбинаций такого рода равно C_n^m, т. е. числу способов, какими можно из n опытов выбрать m, в которых произошло событие A. Вероятность каждой такой комбинации, по теореме умножения для независимых событий, равна $p^m q^{n-m}$. Так как комбинации между собой несовместны, то, по теореме сложения, вероятность события B_m равна

$$P_{m,n} = \underbrace{p^m q^{n-m} + \ldots + p^m q^{n-m}}_{C_n^m \ раз} = C_n^m p^m q^{n-m}$$

Таким образом, если производится n независимых опытов, в каждом из которых событие A появляется с вероятностью p, то вероятность того, что событие A появится ровно m раз, выражается формулой

$$P_{m,n} = C_n^m p^m q^{n-m}$$

где q = 1 – p. Данное распределение вероятностей называется биномиальным распределением [1].

Данная формула описывает, как распределяются вероятности между возможными значениями некоторой случайной величины – числа появлений события A при т опытах. В связи с тем, что вероятности $P_{m,n}$ по форме представляют собой члены разложения бинома $(p+q)^n$, распределение вероятностей вида называется биномиальным распределением.

Очевидно, что определение биномиального закона корректно, т.к. основное свойство ряда распределения $\sum_{i=0}^{n} p_i = 1$ выполнено, ибо $\sum_{i=0}^{n} p_i$ есть не что иное, как сумма всех членов разложения бинома Ньютона:

$$\sum_{i=0}^{n} C_n^m \cdot p^m \cdot q^{n-m} = \left(p+q\right)^n = 1^n = 1.$$

Если известна вероятность $P_{m,n}$, то $P_{m+1,n}$ можно вычислить, пользуясь рекуррентной формулой. Так как $P_{m,n} = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} \cdot p^m \cdot q^{n-m}$ и

$P_{m+1,n} = C_n^{m+1} p^{m+1} q^{n-m-1} = \frac{n!}{(m+1)!(n-m-1)!} \cdot p^{m+1} \cdot q^{n-m-1}$, то

$$\frac{P_{m+1,n}}{P_{m,n}} = \frac{n! \cdot p^{m+1} \cdot q^{n-m-1}}{(m+1)!(n-m-1)!} \cdot \frac{m!(n-m)!}{p^m \cdot q^{n-m} \cdot n!} = \frac{n-m}{m+1} \cdot \frac{p}{q}.$$

Таким образом,

$$P_{m+1,n} = \frac{n-m}{m+1} \cdot \frac{p}{q} \cdot P_{m,n}$$

Поскольку p/q – константа, то можно очень быстро вычислить последующий член при известном предыдущем члене распределения.

Оценку основных числовых характеристик биномиального закона распределения дает следующая

Теорема. Математическое ожидание случайной величины X, распределенной по биномиальному закону, M{X} = np а ее дисперсия D{X} = npq.

Доказательство. Случайную величину X – число наступлений события A в n независимых испытаниях можно представить в виде суммы n независимых случайных величин $X = X_1 + X_2 + \ldots + X_n$, каждая из которых имеет один и тот же закон распределения, т.е. $X = \sum_{k=1}^{n} X_k$. Для произвольного $k = \overline{1,n}$ имеем

X_k	1	0
p_k	p	q

Таким образом, $M\{X_k\} = 1 \cdot p + 0 \cdot q = p$.

Учитывая тот факт, что математическое ожидание суммы случайных величин равно сумме их математических ожиданий, получим

$$M\left\{X\right\}=M\left\{\sum_{k=1}^{n}X_k\right\}=\sum_{k=1}^{n}M\left\{X_k\right\}=np.$$

Аналогично, для произвольного k оценим величину дисперсии:

$$D\left\{X_k\right\}=\left(1-p\right)^2\cdot p+\left(0-p\right)^2\cdot q=q^2p+p^2q=pq\left(p+q\right)=pq.$$

Так как дисперсия суммы независимых случайных величин равна сумме их дисперсий, будем иметь:

$$D\left\{X\right\}=D\left\{\sum_{k=1}^{n}X_k\right\}=\sum_{k=1}^{n}D\left\{X_k\right\}=npq.$$

Легко видеть, что использовать формулу Бернулли при больших значениях n достаточно трудно, т.к. формула требует выполнения действий над огромными числами. Возникает вопрос: нельзя ли вычислить интересующую нас вероятность, не прибегая к формуле Бернулли? Ответ на этот вопрос дает локальная теорема Лапласа, опирающаяся на тот факт, что распределение стандартизованной биномиальной случайной величины $X'=\dfrac{X-np}{\sqrt{npq}}$ при $n\to\infty$ стремится к нормальному распределению с параметрами (0, 1), т. е. для любых действительных чисел α и β (где α < β)

$$\lim_{n\to\infty}P\left(\alpha<X'<\beta\right)=\frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-u^2/2}du$$

Это утверждение называется теоремой Муавра – Лапласа. На ней основывается целый ряд формул для приближенных вычислений значений вероятностей биномиального распределения. В частности, локальная теорема Лапласа позволяет приближенно найти вероятность появления события ровно m раз в n испытаниях, если число испытаний достаточно велико.

Локальная теорема Лапласа. Если вероятность p появления события A в каждом испытании постоянна и отлична от нуля и единицы, то вероятность $P_{m,n}$ того, что событие A появится в n испытаниях ровно m раз приближенно равна значению функции

$$y = \frac{1}{\sqrt{npq}} \cdot \varphi(x),\ \ \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2},\ \ x = \frac{m - np}{\sqrt{npq}}.$$

Имеются таблицы, в которых помещены значения четной функции φ(x).

Пример. Найти вероятность того, что событие А появится ровно 80 раз в 400 испытаниях, если вероятность появления этого события в каждом испытании равна 0,2.

Решение. $P_{80,400} = \frac{1}{\sqrt{400 \cdot 0,2 \cdot 0,8}} \cdot \varphi(x);\ x = \frac{m - np}{\sqrt{npq}} = \frac{80 - 400 \cdot 0,2}{8} = 0.$

По таблице φ(0) = 0,3989. Тогда искомая вероятность $P_{80,400} = \frac{0,3989}{8} = 0,04986.$

При использовании алгоритмического языка R отсутствует проблема больших чисел, т.е. расчеты проводятся легко при любом количестве испытаний (n) и любом количестве появлений события А [2]. В частности, рассмотрим предыдущий пример, используя функцию dbinom.

dbinom(m, n, p) – вероятность того, что случайная величина примет заданное значение. Если x – нецелое или отрицательное число, то будет выдано предупреждение и результатом будет ноль.

x – целочисленный неотрицательный вектор – вектор значений случайной величины.

n – один из параметров биномиального распределения – число испытаний Бернулли

p – второй параметр биномиального распределения – вероятность «успеха» в одном испытании Бернулли.

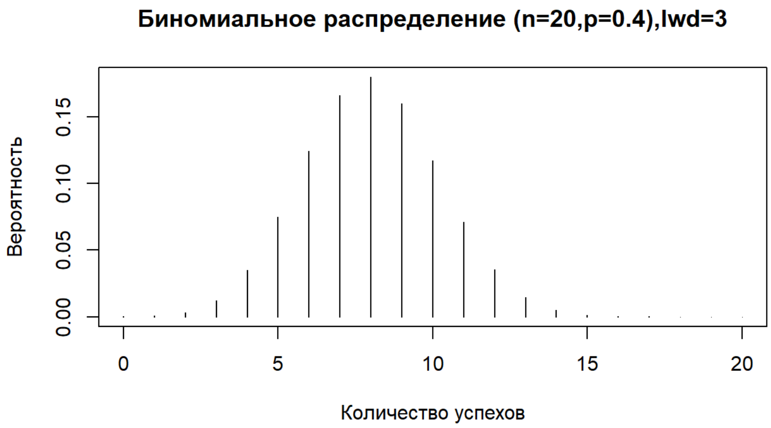
Используя данную функцию (расчеты ведем на платформе RStudio), получим. Отметим, что здесь мы имеем результат более точный, чем при использовании локальной теоремы Лапласа.

```
> dbinom(80,400,0.2)
[1] 0.04981327
```

Чтобы построить гистограмму биномиального распределения в R, можно использовать следующие функции: dbinom(m, n, p) и plot(m,y,type=’h’).

Например, следующий код иллюстрирует построение функции массы вероятности для биномиального распределения с количеством испытаний n = 20 и вероятностью «успеха» = 0,4.

```
> usp <- 0:20
> plot(usp,dbinom(usp,20,0.4),type = 'h',xlab = "Количество успехов",ylab="Веро
ятность",main = 'Биномиальное распределение (n=20,p=0.4),lwd=3')
```



В практических приложениях довольно часто возникает задача вычисления вероятности $P_n(k_1,k_2)$ того, что событие А появится в n испытаниях не менее k_1 раз и не более k_2 раз. Ответ на этот вопрос дает интегральная теорема Лапласа.

Интегральная теорема Лапласа. Если вероятность p появления события А в каждом испытании постоянна и отлична от нуля и единицы, то вероятность $P_n(k_1,k_2)$ того, что событие А появится в n испытаниях не менее k_1 раз и не более k_2 раз приближенно равна определенному интегралу

$$P_n(k_1,k_2)\approx\frac{1}{\sqrt{2\pi}}\int_{x'}^{x''}e^{-z^2/2}dz,\ \ x'=\frac{k_1-np}{\sqrt{npq}},\ \ x''=\frac{k_2-np}{\sqrt{npq}}.$$

При применении этой теоремы пользуются таблицами функции Лапласа

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-z^2/2}dz,$$

т.к. доказано, что $P_n(k_1,k_2)=\Phi(x'')-\Phi(x')$.

Список литературы:

1. Джонсон Н.Л, Коц С., Кемп А.У. Одномерные дискретные распределения // М.: Лаборатория знаний, 2017. 563с.
2. Зарядов И.С. Статистический пакет R: Теория вероятностей и математическая статистика // М.: РУДН. 2010. 141с.

UDC 519.24

ESTIMATION OF NUMERICAL CHARACTERISTICS AND VISUALIZATION OF THE BINOMIAL DISTRIBUTION LAW (USING THE ALGORITHMIC LANGUAGE R)

Boris I. Smagin

doctor of economics, professor

bismagin@mail.ru

Michurinsk State Agrarian University

Michurinsk, Russia

Abstract. The binomial distribution law has wide application in the analysis of random processes. The article analyzes the Bernoulli tests underlying the formation of the binomial law. The estimation of mathematical expectation and variance is given, and the local Laplace theorem is given, which is used for large independent tests. The application of the algorithmic language R to calculate probability and visualize the binomial distribution law is also shown.

Keywords: Bernoulli tests, binomial law, mathematical expectation, variance, algorithmic language R, local and integral Laplace theorems.

Статья поступила в редакцию 01.02.2024; одобрена после рецензирования 20.03.2024; принята к публикации 22.03.2024.

The article was submitted 01.02.2024; approved after reviewing 20.03.2024; accepted for publication 22.03.2024.