Математика и механика

УДК 519.2

К ВОПРОСУ О ТОЧНОСТИ АСИМПТОТИЧЕСКИХ ПРИБЛИЖЕНИЙ ФОРМУЛЫ БЕРНУЛЛИ

И.Г. Карпов, А.Н. Грибков

Тамбовский государственный технический университет
E-mail: GribkovAlexey@yandex.ru

Предложены выражения для аппроксимации формулы Бернулли, позволяющие более точно определять вероятности появления события А ровно k раз при большом количестве независимых испытаний, если в каждом из них событие А наступает с вероятностью p.

Ключевые слова:
Формула Бернулли, независимые испытания, аппроксимация.
Key words:
Bernoulli's formula, independent trials, approximation.

В теории вероятностей при решении целого ряда практических задач приходится сталкиваться со следующей схемой проведения испытаний: производится N независимых испытаний, в результате каждого из них происходит либо событие A с вероятностью p, либо противоположное ему событие C с вероятностью q=1–p. Два исхода каждого испытания (наступление события A либо события C) обычно обозначают символами «1» и «0» или называют «успехом» и «неудачей». Такая схема испытаний впервые была рассмотрена Я. Бернулли и носит его имя [1–3].

Вероятность $P_N(k)$ того, что событие A при N испытаниях наступит ровно k раз (k=1,2,...,N) определяется по формуле Бернулли [1–3]

$$P_N(k) = \frac{N!}{(N-k)!k!} p^k (1-p)^{N-k}, \qquad (1)$$

представляющей собой биномиальное распределение. Основными числовыми характеристиками распределения (1) являются математическое ожидание m, дисперсия D, коэффициент асимметрии K_a и коэффициент эксцесса $K_э$, определяемые соотношениями [4]:

$$m = Np; \; D = Npq; \; K_a = \frac{q-p}{\sqrt{Npq}}; \; K_э = \frac{1-6pq}{Npq}.$$

С помощью формулы Бернулли можно вычислять вероятности появления событий при произвольном, в том числе большом, числе испытаний N. Однако ее практическое использование сопряжено с трудностями уже при N>10, что вызвано необходимостью проведения операций над очень большими числами. Эти трудности можно преодолеть, например, путем использования специальных рекуррентных алгоритмов вычисления факториалов и степеней больших чисел, но это сопряжено с увеличением объемов вычислений и не всегда практически удобно. Значительного упрощения удается достичь также применением программных систем компьютерной математики. Однако и этот путь не может в полной мере решить проблему размерности задач и операций над весьма большими числами, возникающую, например, при анализе нескольких десятков и сотен испытаний. В то же время исследования асимптотического поведения вероятностей $P_N(k)$ появления ровно k раз события A при N испытаниях при стремлении N к бесконечности дают возможность получать приближенные, но достаточно точные с практической точки зрения, значения этих вероятностей по значительно более простым выражениям, чем формула Бернулли (1).

В настоящее время в качестве асимптотических (приближенных) формул для вычисления вероятностей $P_N(k)$ используют формулу

$$P_N(k) \approx \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(k-Np)^2}{2Npq}\right], \qquad (2)$$

которая является следствием локальной теоремы Муавра–Лапласа, а также формулу Пуассона при $p<0{,}05$ [4]

$$P_N(k) \approx \frac{(Np)^k}{k!}\exp[-Np]. \qquad (3)$$

В течение последних 30 лет пристальное внимание многих исследователей привлекает семейство обратно гауссовских распределений [4–7]. В качестве аппроксимирующего распределения из этого семейства распределений широко используется распределение Вальда

$$p(x) = \frac{0{,}5c}{\sqrt{\pi\,\lambda\,(x-\mu)}}\times$$
$$\times\exp\left[c-\lambda\,(x-\mu)-\frac{0{,}25c^2}{\lambda\,(x-\mu)}\right];\ x>\mu, \qquad (4)$$

где $c>0$, $\lambda>0$, $-\infty<\mu<\infty$ – параметры распределения (4). Оно имеет асимметричную форму с $K_a>0$ и также, как гауссовское и пуассоновское распределения, является безгранично делимым распределением.

Основная цель данной работы – предложить выражения, аналогичные (2) и (3), для аппроксимации формулы Бернулли (1), которые позволят более точно определять вероятности $P_N(k)$ появления события A ровно k раз при большом количестве независимых испытаний, если в каждом из них событие A наступает с вероятностью p, а также рассмотреть условия, при которых возможен точный расчет вероятности $P_N(k)$ без особых затруднений.

Отметим сразу, что если количество независимых испытаний $N\le 10$, то расчет вероятности $P_N(k)$ появления события A ровно k раз можно производить точно по формуле Бернулли (1) либо с помощью рекуррентных формул

$$P_N(k) = \frac{p\,(N+1-k)}{(1-p)k}P_N(k-1),\ \ P_N(0)=(1-p)^N; \quad (5)$$

$$P_N(k) = \frac{(1-p)(k+1)}{p\,(N-k)}P_N(k+1),\ \ P_N(N)=p^N. \quad (6)$$

Если количество независимых испытаний $N>10$, то удобнее прямые вычисления вероятности $P_N(k)$ появления события A ровно k раз производить по рекуррентным формулам (5) и (6). При этом если выполняются условия

$$P_N(k)\ge 0{,}0001;\ 0\le k\le 10, \qquad (7)$$

то используется рекуррентная формула (5).

При выполнении условий

$$P_N(k)\ge 0{,}0001;\ N-10\le k\le N \qquad (8)$$

используется рекуррентная формула (6). Верхние граничные значения параметров N и p, при которых еще выполняются условия (7) или (8), приведены в таблице.

Рассмотрим теперь приближенные формулы для расчета вероятности $P_N(k)$ появления события A ровно k раз при различных значениях вероятности p, если количество независимых испытаний $N>10$:

1. Пусть $0<p\le 0{,}02$. В этом случае можно использовать формулу Пуассона (3).

2. Если $0<p\le 0{,}45$; $N'\ge 6/pq$, то для аппроксимации формулы (1) можно использовать распределение Вальда (4)

$$P_N(k) \approx \frac{0{,}5c}{\sqrt{\pi\,\lambda\,(k-\mu)}}\times$$
$$\times\exp\left[c-\lambda\,()k-\mu-\frac{0{,}25c^2}{\lambda\,(k-\mu)}\right], \qquad (9)$$

$$\lambda=\frac{1{,}5}{1-2p};\ \ c=\frac{9\,(N+0{,}5)\,pq}{(1-2\,p)^2};$$

$$\mu = Np-\frac{9\,(N+0{,}5)\,pq}{1-2p}.$$

3. Если $0{,}45<p<0{,}55$; $N'\ge 6/pq$, то для аппроксимации формулы (1) используется гауссовское распределение

$$P_N(k)\approx\frac{1}{\sqrt{2\pi\,(N+0{,}5)\,pq}}\exp\left[-\frac{(k-Np)^2}{2\,(N+0{,}5)\,pq}\right].(10)$$

Оно отличается от распределения (2) величиной дисперсии, что позволяет обеспечить более высокую точность аппроксимации распределения (1).

Таблица. *Значения параметров N и p, при которых выполняются условия (7) или (8)*

Условия (7)	p	0,001	0,002	0,005	0,01	0,02	0,05	0,1	0,15	0,2
	N	1500	900	300	190	100	44	24	17	14
Условия (8)	p	0,999	0,998	0,995	0,99	0,98	0,95	0,9	0,85	0,8
	N	1500	900	300	190	100	44	24	17	14

4. Если $0{,}55<p<1$; $N'\ge 6/pq$, то для аппроксимации формулы (1) используется распределение Вальда (4) с $K_a<0$

$$P_N(k)\approx\frac{0{,}5c}{\sqrt{\pi\,\lambda\,(\mu-k)}}\times$$
$$\times\exp\left[c-\lambda\,(\mu-k)-\frac{0{,}25c^2}{\lambda\,(\mu-k)}\right],$$

$$\lambda=\frac{1{,}5}{2p-1};\ \ c=\frac{9\,(N+0{,}5)\,pq}{(1-2p)^2};$$

$$\mu=Np+\frac{9\,(N+0{,}5)\,pq}{2p-1}.$$

5. Пусть $0{,}98\le p<1$; $N>10$. В этом случае для аппроксимации формулы Бернулли (1) можно использовать распределение Пуассона с $K_a<0$

$$P_N(k)\approx\frac{(Nq)^{N-k}}{(N-k)!}\exp[-Nq].$$

Рассмотрим два примера.

Пример 1. Вероятность изделия некоторого производства оказаться бракованным равна 0,005. Чему равна вероятность того, что из 10000 наудачу отобранных изделий окажется ровно 55 бракованных изделий?

Решение. В соответствии с условием примера имеем p=0,005, N=10000, $P(A)=P_N(55)$. Прямые вычисления по формуле Бернулли (1) возможны только с использованием компьютерных технологий. Так, например, в среде Mathcad имеется встроенная функция dbinom, с помощью которой можно вычислить вероятность $P(A)=P_N(55)$. При этом получим $P_N(55)$=dbinom(55,N,p)=0,042. Так как p=0,005, $N>6/pq$=1206, то в качестве приближенных формул используем формулы (3) и (9). Расчет по ним приводит к тому же результату. С помощью формулы (2) получим $P_N(55)$=0,044. Здесь появляется погрешность в третьем знаке после запятой.

Пример 2. Вероятность p появления события A при каждом испытании равна 0,45. Производится 150 независимых испытаний. Определить вероятность $P_N(k)$ того, что событие A наступит ровно 80 раз.

Решение. В соответствии с условием примера имеем p=0,45, N=150, $P(A)=P_N(80)$. Прямые вычисления по формуле Бернулли (1) в среде Mathcad позволяют найти вероятность $P(A)=P_N(80)$. При этом получим $P_N(80)=8{,}057\cdot10^{-3}$. Расчет по формуле (2) дает $P_N(80)=7{,}983\cdot10^{-3}$, по формуле (9) – $P_N(80)=8{,}078\cdot10^{-3}$ и по формуле (10) – $P_N(80)=8{,}025\cdot10^{-3}$. Следовательно, формулы (9) и (10) позволяют точнее производить расчет вероятности $P_N(k)$, чем формула (2).

Выводы

Предложены выражения, аналогичные (2) и (3), для аппроксимации формулы Бернулли (1). Они позволяют более точно определять вероятности $P_N(k)$ появления события A ровно k раз при большом количестве независимых испытаний, если в каждом из них событие A наступает с вероятностью p. Рассмотрены также условия, при которых возможен точный расчет вероятности $P_N(k)$ по формуле Бернулли без особых затруднений.

СПИСОК ЛИТЕРАТУРЫ

1. Гнеденко Б.В. Курс теории вероятностей. – М.: Наука, 1988. – 448 с.
2. Левин Б.Р. Теоретические основы статистической радиотехники. – М.: Радио и связь, 1989. – 656 с.
3. Вероятность и математическая статистика / гл. ред. Ю.В. Прохоров. – М.: Большая Российская энциклопедия, 1999. – 910 с.
4. Вадзинский Р.Н. Справочник по вероятностным распределениям. – СПб.: Наука, 2001. – 296 с.
5. Джонсон Н.Л., Коц С., Балакришнан Н. Одномерные непрерывные распределения. Ч. 1. – М.: БИНОМ. Лаборатория знаний, 2010. – 703 с.
6. Истигечева Е.В. Оценивание параметров гиперболического и обратного гауссовского распределений // Известия Томского политехнического университета. – 2006. – Т. 309. – № 6. – С. 11–13.
7. Бостанджиян В.А. Распределение Пирсона, Джонсона, Вейбулла и обратное нормальное. Оценивание их параметров. – Черноголовка: Редакционно-издательский отдел ИПХФ РАН, 2009. – 240 с.

Поступила 14.05.2012 г.