УДК 519.214.4

Оценки вероятностей уклонений сумм для случайных величин Бернулли

А. Н. Архангельский*, П. В. Кириченко, Г. М. Пиголкин

Показаны оценки вероятностей отклонения среднеарифметического из независимых одинаково распределенных бернуллиевских случайных величин от вероятности «успеха».

Ключевые слова: бернуллиевские, биномиальные и сопряженные случайные величины, моменты случайных величин.

Пусть на вероятностном пространстве (Ω, A, P) заданы независимые одинаково распределенные бернуллиевские случайные величины ξ_n, $n = 1, 2, \ldots$, принимающие значения единица, когда появляется «успех» и ноль — при «неудаче» с вероятностями $p\in(0, 1)$ и $q = 1 - p$ соответственно. Напомним, что обычно значение p называют вероятностью «успеха», а q — вероятностью «неудачи». Положим, что $S_n = \xi_1 + \ldots + \xi_n$.

Ниже показаны результаты, которые можно использовать в задачах математической статистики по малым выборкам, где недостаточно простого неравенства Чебышева, а требуются интервальные оценки повышенной точности. Таким образом, целью работы является получение двухсторонних оценок для вероятностей отклонений сумм одинаково распределенных случайных величин от среднего (т.е. вероятности «успеха») в однородной схеме Бернулли.

Полученные оценки, в частности, уточняют известное неравенство Гефдинга [1] и оценку С. В. Нагаева [2, теорема 2] для одинаково распределенных величин.

Начнем с получения нижних оценок.

Теорема 1. Пусть для p, $q = 1 - p$, ε и n выполняются следующие условия:

(i) $0 < \varepsilon < q < 1$;

$$(ii)\ n \ge n_0 = \left[3,662\left(\frac{(p+\varepsilon)^2+(q-\varepsilon)^2}{(p+\varepsilon)(q-\varepsilon)}\right)^2+1\right],$$

где $[\cdot]$ — целая часть числа.

(iii)

$$\left(\frac{(p+\varepsilon)^2+(q-\varepsilon)^2}{(p+\varepsilon)(q-\varepsilon)}\right)\times$$
$$\times\left(\sqrt{\frac{2\pi}{n}+\frac{\varepsilon^2(1-\varepsilon)(p+\varepsilon)}{p^2(q-\varepsilon)}}+(\pi-1)\sqrt{\frac{\varepsilon^2(1-\varepsilon)(p+\varepsilon)}{p^2(q-\varepsilon)}}\right)<1.3.$$

Тогда справедливы оценки:

$$P\left(\frac{S_n}{n}-p\ge\varepsilon\right)\ge Q(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)} \quad (1)$$

и

$$P\left(\frac{S_n}{n}-p\le-\varepsilon\right)\ge Q(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}, \quad (2)$$

где

$$Q(\varepsilon,n)=\sqrt{\frac{\pi}{2}}\left(\sqrt{2\pi+\frac{n\varepsilon^2(1-\varepsilon)^2(p+\varepsilon)}{p^2(q-\varepsilon)}}+\right.$$
$$\left.+(\pi-1)\sqrt{\frac{n\varepsilon^2(1-\varepsilon)^2(p+\varepsilon)}{p^2(q-\varepsilon)}}\right)^{-1}- \quad (3)$$
$$-0,9568\cdot\frac{(p+\varepsilon)^2+(q-\varepsilon)^2}{\sqrt{n(p+\varepsilon)(q-\varepsilon)}}.$$

* arkhang@gmail.com

Следствие. Пусть выполняются условия (i) — (iii). Тогда справедлива оценка:

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\geq 2Q(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}, \quad (4)$$

где $Q(\varepsilon, n)$ определено в (3).

Доказательство. Воспользуемся методом сопряженных распределений. Пусть $F(x)$ обозначает функцию распределения ξ_i и $R(t) = E\exp(t\xi_i)$. Определим функцию распределения сопряженных случайных величин $\tilde{\xi}_i$ равенством:

$$\tilde{F}(x)=R^{-1}(t)\int_{-\infty}^{x}e^{tu}dF(u)=$$
$$=\begin{cases}0, & \text{если} \quad x\leq 0;\\ q/(pe^t+q), & \text{если} \quad 0<x\leq 1;\\ 1, & \text{если} \quad x>1.\end{cases}$$

Здесь $\tilde{\xi}_i$ — независимые, одинаково распределенные случайные величины, имеющие распределение Бернулли с параметром, равным $\tilde{p} = pe^t/(pe^t+ q)$. Положим $\sum_{i=1}^{n}\tilde{\xi}_i$. Поскольку $\tilde{S}_n$ имеет биномиальное распределение $Bi(n, \tilde{p})$, то $E\tilde{S}_n = n\tilde{p}$ и $D\tilde{S}_n = n\tilde{\sigma}^2$, где $\tilde{\sigma}^2 = \tilde{p}\tilde{q}$, $\tilde{q} = q/(pe^t + q)$. Легко проверить, что

$$P(S_n<x)=R^n(t)\int_{-\infty}^{x}\mathrm{e}^{-tu}dP(\tilde{S}_n<u).$$

Поэтому получаем

$$P\left(\frac{S_n}{n}-p\geq\varepsilon\right)=\mathrm{e}^{-tn\tilde{p}}R^n(t)\times$$
$$\times\int_{\sqrt{n}(\varepsilon+p-\tilde{p})/\tilde{\sigma}}^{+\infty}\mathrm{e}^{-t\tilde{\sigma}\sqrt{n}y}d\tilde{F}_n(y), \quad (5)$$

где $\tilde{F}_n(y)=P\left((\tilde{S}_n-n\tilde{p})/(\sqrt{n}\tilde{\sigma})<y\right)$.

Найдем t из условия равенства нулю нижней границы интеграла в (5), т.е. из равенства $\varepsilon+p-\tilde{p}=0$ или, что эквивалентно, из уравнения $(\varepsilon + p)(pe^t + q) = pe^t$. Если выполняется условие (i) теоремы 1, то t существует и равно

$$t=\ln\left(\frac{1+\varepsilon/p}{1-\varepsilon/q}\right)>0. \quad (6)$$

Ясно, что для такого t получаем $\tilde{p} = p + \varepsilon$, $\tilde{q} = q - \varepsilon$, $R_n(t) = 1/(1 - \varepsilon/q)$. Обозначим

$$r_n(y) = \tilde{F}n(y) - \Phi(y) \text{ и } r_n=\sup_{y\in R}\left|r_n(y)\right|,$$

где $\Phi(y)=(2\pi)^{-1/2}\int_{-\infty}^{y}\mathrm{e}^{-u^2/2}du$

Из классического неравенства Берри–Эссеена следует, что

$$r_n\leq An^{-1/2}E\,|\,\tilde{\xi}_i-E\tilde{\xi}_1|^3\left(D\tilde{\xi}_1\right)^{-3/2}=A\left(\tilde{p}^2+\tilde{q}^2\right)\left(n\tilde{p}\tilde{q}\right)^{-1/2},$$

где A = 0,4784 (см. [3]). Из (5), с учетом описанного, получим

$$P\left(\frac{S_n}{n}-p\geq\varepsilon\right)\geq\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}\times$$
$$\times\left(\frac{1}{\sqrt{2\pi}}\int_0^{+\infty}\mathrm{e}^{-t\tilde{\sigma}\sqrt{n}y-y^2/2}dy-2r_n\right), \quad (7)$$

где t задается выражением (6). Заметим, что интеграл

$$\int_0^{+\infty}\exp(-t\tilde{\sigma}\sqrt{n}y-y^2/2)dy=$$
$$=\exp(t^2\tilde{\sigma}^2n)\int_{t\tilde{\sigma}\sqrt{n}}^{+\infty}e^{-y^2/2}dy=I(t\tilde{\sigma}\sqrt{n})$$

представляет собой известное отношение Миллса. Воспользовавшись нижней оценкой этого отношения (см., например [4]), получим

$$I(t\tilde{\sigma}\sqrt{n})\geq\pi\left(\sqrt{t^2\tilde{\sigma}^2n+2\pi}+(\pi-1)t\tilde{\sigma}\sqrt{n}\right)^{-1}.$$

Используя последнюю оценку и неравенство Берри–Эссеена, из (7) получим оценку (1). Для того чтобы она была нетривиальной, требуется, чтобы $Q(\varepsilon, n)$ из (3) было больше нуля. Это несложно, если

$$\frac{\tilde{p}\tilde{q}}{\tilde{p}^2+\tilde{p}^2}>A\sqrt{\frac{8}{\pi}}\left(\sqrt{t^2\tilde{\sigma}^2+2\pi/n}+(\pi-1)t\tilde{\sigma}\right). \quad (8)$$

Условие (ii) теоремы 1 следует из (8), если положить $t = 0$ и воспользоваться неравенством $16A^2 < 3{,}662$. Легко проверить, что из (6) следует

$$t \leq \varepsilon/p + \varepsilon/(q(1 - \varepsilon/q)) = \varepsilon(1-\varepsilon)/(p(q - \varepsilon)).$$

Поэтому неравенство (8) будет справедливо, если выполняются условия (ii) и (iii) теоремы 1, поскольку $A^{-1}\sqrt{\pi/8}>1{,}3$.

Доказательство оценки (2) проводится совершенно аналогично. Оценка (4) очевидным образом следует из неравенств (1) и (2).

Для того чтобы оценить качество нижних оценок, желательно иметь верхние той же вероятности. Можно воспользоваться результатами работы [5]. Однако, оценки из [5] точны лишь при больших значениях n. Более наглядным будет сравнение полученной оценки (1) с верхней оценкой из работы [1]. В следующей теореме будет получена верхняя оценка, являющаяся одним из многочисленных уточнений неравенства Хефдинга.

Теорема 2. Пусть выполняется условие

(j) $0 < \varepsilon < q < 1$.

Тогда справедливы оценки

$$P\left(\frac{S_n}{n}-p\geq\varepsilon\right)\leq U(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)} \qquad (9)$$

и

$$P\left(\frac{S_n}{n}-p\leq-\varepsilon\right)\leq U(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}, \qquad (10)$$

где

$$U(\varepsilon,n)=\sqrt{\frac{\pi}{2}}\left(\sqrt{2\pi+(\pi-2)\frac{n\varepsilon^2(1+\varepsilon)^2(q-\varepsilon)}{q^2(p+\varepsilon)}}+\right.$$
$$\left.+2\sqrt{\frac{n\varepsilon^2(1+\varepsilon)^2(q-\varepsilon)}{q^2(p+\varepsilon)}}\right)^{-1}+ \qquad (11)$$
$$+0,9568\frac{(p+\varepsilon)^2+(q-\varepsilon)^2}{\sqrt{n(p+\varepsilon)(q-\varepsilon)}}.$$

Следствие. Пусть выполняется условие (j). Тогда справедлива оценка

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\leq 2U(\varepsilon,n)\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}, \qquad (12)$$

где $U(\varepsilon, n)$ определено в (11).

Замечание. Отмеченное выше неравенство Гефдинга имеет вид (12), где $U(\varepsilon, n)=1$, а в нашем неравенстве $U(\varepsilon, n) < 1$.

Доказательство теоремы 2 практически полностью повторяет доказательство предыдущей теоремы. Изменение коснется лишь нескольких моментов. Вместо нижнего неравенства (7) будет справедливо неравенство:

$$P\left(\frac{S_n}{n}-p\geq\varepsilon\right)\leq\left(1-\frac{\varepsilon}{q}\right)^{-n(q-\varepsilon)}\left(1+\frac{\varepsilon}{p}\right)^{-n(p+\varepsilon)}\times$$
$$\times\left(\frac{1}{\sqrt{2\pi}}\int_0^{+\infty}e^{-t\tilde{\sigma}\sqrt{n}y-y^2/2}dy+2r_n\right), \qquad (13)$$

где t снова задается выражением (6). Ясно, что

$$t\geq\varepsilon/(p(1+\varepsilon/p))+\varepsilon/q=\varepsilon(1+\varepsilon)/(q(p+\varepsilon)).$$

Воспользовавшись верхней оценкой отношения Миллса (см. [4]), получим

$$\int_0^{+\infty}e^{-t\tilde{\sigma}\sqrt{n}y-y^2/2}dy\leq$$
$$\leq\pi\left(\sqrt{(\pi-2)^2t^2\tilde{\sigma}^2n+2\pi}+2t\tilde{\sigma}\sqrt{n}\right)^{-1}.$$

Используя теперь оценку Берри–Эссеена для r_n, из (13) получим требуемое неравенство (9).

Неравенство (10) доказывается точно также. Неравенство (12) — очевидное следствие неравенств (9) и (10).

Литература

1. **Hoeffding W.** Probability inequalities for sums of bounded random variables // J. Am. Stat. Assoc. 1963. N 58. P. 13 — 30.
2. **Нагаев С.В.** Нижние границы для вероятностей больших уклонений сумм независимых случайных величин // Теория вероятности и ее применение. 2001. № 46. Вып. 4. С. 785 — 792.
3. **Королев В.Ю., Шевцова И.Г.** О верхней оценке абсолютной постоянной в неравенстве Берри–Эссеена // Теория вероятности и ее применение. 2009. № 54. Вып. 4. С. 671 — 695.
4. **Люк Ю.** Специальные математические функции и их аппроксимации. М.: Мир, 1980.
5. **Antonov S.N., Kruglov V.M.** Sharpened versions of a Kolmogorov's inequality // Stat. & Prob. Lett. 2010. N 80. P. 155 — 160.

Статья поступила в редакцию 23.10.2015