Закон больших чисел

Содержание

Закон больших чисел. Основные понятия

Экспериментальные и теоретические исследования показывают, что с ростом числа повторных испытаний суммарное поведение случайных событий почти совсем утрачивает случайный характер и становится всё более прогнозируемым. Для практических целей важно знать, при выполнении каких условий это происходит.

Эти условия указываются в теоремах, которые носят название закона больших чисел. Закон больших чисел позволяет установить общую закономерность, присущую случайным событиям.

В частности, проявлением закона больших чисел служит тот факт, что вероятность больших отклонений относительной частоты от «истинного» значения вероятности случайного события уменьшается с ростом числа испытаний. Кроме того, при большом числе наблюдений факторы, не связанные с существом самого процесса, а проявляющиеся только случайным образом, взаимно погашаются.

Рассмотрим теоремы, отражающие суть закона больших чисел.

Теорема Бернулли

Эта теорема изложена в сочинении Якоба Бернулли «Искусство предугадывания» (Ars conjectandi, 1713), изданном уже после его смерти, и явилась, по существу, первой попыткой изложения закона больших чисел.

Суть её формулировалась следующим образом.

Если вероятность события A постоянна ($P(A)=const$), *то при неограниченном возрастании числа независимых испытаний* ($n\to\infty$) *частость* $\frac{m}{n}$ *появления события A может принимать значения, сколь угодно близкие к его вероятности, то есть* $\left|\frac{m}{n}-p\right|<\varepsilon$, *где* ε *– сколь угодно малое положительное число.*

Однако, как показали дальнейшие исследования, с этим утверждением нельзя согласиться полностью, так как существует ненулевая вероятность, что эта разность может оказаться значительной, например, если в результате испытаний окажется, что m=0. Вместе с тем, несомненно, эта теорема сыграла большую роль при рассмотрении предельных законов теории вероятностей. Теорема в оригинале имела очень сложное доказательство, современная формулировка и более простое её доказательство были даны П. Л. Чебышевым в 1846 году (ниже мы представим новую формулировку теоремы Бернулли).

Теорема Чебышева

Вероятность того, что средняя арифметическая $\overline{X}$ *попарно независимых одинаково распределённых случайных величин* X_1, X_2, …, X_n, *имеющих одинаковые математические ожидания* $M(X_i)=\mu$ *и ограниченные дисперсии* $D(X_i)\leq C$ (i=1, 2, …, n), *отличается от средней арифметической их математических ожиданий меньшей, чем на наперёд заданную величину* $\varepsilon>0$, *с ростом числа испытаний стремится к единице*:

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|<\varepsilon\right)=1. \qquad (1)$$

Формулировка теоремы Чебышева может быть изложена и в более краткой форме: *при достаточно большом числе испытаний* ($n\to\infty$) *случайной величины* X *среднее арифметическое полученных результатов будет сколь угодно мало отличаться от её истинного значения* ($\overline{X}\to\mu$).

Следствие. *Если в результате конечного числа* n *независимых испытаний случайная величина* X, *имеющая математическое ожидание* μ *и дисперсию* σ^2, *приняла значения* x_1, x_2, …, x_n, *то с вероятностью, сколь угодно близкой к единице, можно утверждать, что отклонение средней арифметической* $\overline{X}$ *наблюдаемых значений величины от её математического ожидания не превысит некоторой малой положительной величины* $\varepsilon>0$, *где* $\varepsilon=\varepsilon(\sigma^2,n)$.

Взаимосвязь вероятности P, величины ошибки ε, числа испытаний n и дисперсии σ^2 может быть представлена формулой:

$$P\left(\left|\overline{X}-\mu\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{n\cdot\varepsilon^2} \qquad (2)$$

или

$$P\left(\left|\overline{X}-\mu\right|<\varepsilon\right)>1-\frac{\sigma^2}{n\cdot\varepsilon^2}. \qquad (3)$$

Замечание. При $n\to\infty$ правая часть неравенства (2) стремится к нулю, а неравенства (3) стремится к единице, так как σ^2 и ε – постоянные величины.

Пример. Для установления «истинной» массы (математического ожидания массы) новорождённой девочки на территории некоторого региона провели 70 измерений. Дисперсия массы составила 0,2 кг^2. Найти вероятность того, что отклонение (ошибка) среднего значения массы от математического ожидания (истинного значения) не превысит 0,15 кг.

Решение. Из условия: n=70; σ^2=0,2 кг^2; ε=0,15 кг. Воспользуемся формулой (3):

$$P\left(\left|\overline{X}-\mu\right|<\varepsilon\right)>1-\frac{\sigma^2}{n\cdot\varepsilon^2}=1-\frac{0,2}{70\cdot 0,15^2}=0,873.$$

Ответ: $P>0,873$ – то есть с уверенностью более 87% мы можем утверждать, что полученный нами результат не будет отличаться от «истинного» значения более чем на 0,15 кг.

Замечание. Формулой (3) можно также воспользоваться и если необходимо определить число испытаний n, при котором возможная ошибка не превысит наперёд заданную допустимую величину ε, то есть, начиная с некоторого n, будет выполняться условие $\left|\overline{X}-\mu\right|\le\varepsilon$.

Пример. (Рассмотрим предыдущий пример, но при другой постановке вопроса). Пусть «истинную» массу μ новорождённой девочки пытаются определить через среднее арифметическое. И пусть проведённая в результате исследований оценка дисперсии массы дала величину, равную 0,2 кг^2. Спрашивается, каково должно быть число испытаний n, чтобы отличие среднего от «истинного» значения массы с вероятностью более 0,95 было меньше 0,1 кг?

Решение. Воспользуемся формулой (3):

$$P\left(\left|\overline{X}-\mu\right|<\varepsilon\right)>1-\frac{\sigma^2}{n\cdot\varepsilon^2}>0,95,$$

$$1-\frac{0,2}{n\cdot 0,1^2}>0,95\Leftrightarrow -\frac{1}{n}>\frac{(0,95-1)\cdot 0,01}{0,2}\Leftrightarrow n>\frac{0,2}{0,05\cdot 0,01}=400.$$

Ответ: при n > 400 будет выполняться заданное условие.

Теорема Бернулли («современная формулировка»)

С вероятностью, сколь угодно близкой к единице, можно предположить, что при неограниченном числе независимых испытаний относительная частота появления события А как угодно мало будет отличаться от вероятности появления его в отдельном испытании, то есть разность между относительной частотой события и его вероятностью может быть представлена в виде предела:

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p(A)\right|<\varepsilon\right)=1. \qquad (4)$$

Таким образом, при $n\to\infty$ относительная частота $\frac{m}{n}$ появлений случайного события A стремится к его вероятности $p(A)$.

Следствие 1. *При ограниченном числе измерений (испытаний) n для оценки вероятности того, что ошибка при определении вероятности события A через нахождение относительной частоты не превысит некоторое число ε>0, можно воспользоваться формулой, которая следует из теоремы Бернулли:*

$$P\left(\left|\frac{m}{n}-p(A)\right|<\varepsilon\right)>1-\delta. \tag{5}$$

Причём ε и δ при достаточно больших n связаны соотношениями:

$$\delta \geq \frac{p\cdot q}{n\cdot\varepsilon^2} \text{ и } n \geq \frac{p\cdot q}{\delta\cdot\varepsilon^2}\text{, где } p = p(A),\ q = 1 - p.$$

$$P\left(\left|\frac{m}{n}-p(A)\right|<\varepsilon\right)>1-\frac{p\cdot q}{n\cdot\varepsilon^2} \tag{6}$$

Пример. Вероятность спроса на некоторую лекарственную форму равна p=0,8. Оценить вероятность того, что при 1000 обращениях в аптеку отклонение частоты спроса на данную лекарственную форму от вероятности спроса по абсолютной величине будет меньше 0,05.

Решение. По условию: $p(A)=0{,}8$; $n=10^3$; $\varepsilon=0{,}05$; $q=1-p=1-0{,}8=0{,}2$. Воспользовавшись формулой (6), имеем:

$$P\left(\left|\frac{m}{n}-p(A)\right|<\varepsilon\right)>1-\frac{p\cdot q}{n\cdot\varepsilon^2}=1-\frac{0{,}8\cdot 0{,}2}{10^3\cdot 0{,}05^2}=1-\frac{0{,}16}{2{,}5}=0{,}936.$$

Ответ: искомая вероятность $P > 0{,}936$.

Следствие 2. *Выражение* (6) *может быть использовано и для определения числа испытаний n, при котором ошибка оценки вероятности случайного события A через относительную частоту его появления не будет превышать некоторое наперёд заданное положительное число ε.*

Пример. (Рассмотрим предыдущий пример, но при другой постановке вопроса). В условиях ограниченного числа наблюдений вероятность некоторого события исследуется через относительную частоту его появления. Предположим, что таким образом исследуется вероятность спроса на некоторую лекарственную форму. Пусть априори известна вероятность спроса в её единичном проявлении $p(A)$=0,8. Спрашивается: каково должно быть число n исследований обращения в аптеку, чтобы ошибка при определении вероятности через относительную частоту не превысила ε=0,15, а степень нашей уверенности (вероятность) была бы не менее $p\geq 0{,}95$?

Решение. По условию $p(A)=0{,}8$; $\varepsilon<0{,}15$; $P>0{,}9$; $q=1-p(A)=0{,}2$. Воспользуемся формулой (4):

$$P > 1 - \frac{p(A)\cdot q}{n\cdot\varepsilon^2} = 1 - \frac{0{,}8\cdot 0{,}2}{n\cdot 0{,}05^2} > 0{,}9 \Leftrightarrow \frac{0{,}8\cdot 0{,}2}{n\cdot 0{,}15^2} \le 0{,}1 \Leftrightarrow$$

$$n \ge \frac{0{,}16}{0{,}1\cdot 0{,}0225} = 71{,}1.$$

Таким образом, число исследований обращения в аптеку должно быть не менее 72.

Пример. При изготовлении одноразового медицинского инструмента, по данным ОТК, брак составляет 1,3%. Найти вероятность того, что при осмотре партии, состоящей из 1000 изделий, выявится отклонение от установленного процента брака меньше, чем на 1%.

Решение. Таким образом, нам требуется оценить:

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \text{ при } p=0{,}013,\ \varepsilon=0{,}01 \text{ и } n=1000.$$

По теореме Бернулли $P>1-\delta$, откуда:

$$\delta \ge \frac{p\cdot q}{n\cdot\varepsilon^2} = \frac{0{,}013\cdot 0{,}987}{1000\cdot 0{,}01^2} = 0{,}128.$$

Искомая вероятность равна $P>1-\delta=1-0{,}128=0{,}872$. Таким образом, вероятность того, что при осмотре партии, состоящей из 1000 изделий, выявится отклонение от установленного процента брака меньше, чем на 1%, будет равна 0,872.

Литература

1. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли. – М.: Прогресс, 1976. – 496 с.
2. Гланц С. Медико-биологическая статистика / С. Гланц. - М.: Практика, 1998. - 459 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – М.: Высш. шк., 2003. – 479 с.
4. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи) / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 67 с.
5. Новиков Д. А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д. А. Новиков, В. В. Новочадов. – Волгоград: Изд-во ВГМУ, 2005. – 84 с.
6. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Киров: Изд-во ВГГУ, 2015. – 158 с.