Вероятностные модели.
Практика 15.
Предельные теоремы теории вероятностей

Предельные теоремы ТВ устанавливают связь между теоретическими и экспериментальными характеристиками случайных величин при большом числе испытаний над ними.

Предельные теоремы позволяют выяснять свойства сумм и средних арифметических нескольких случайных величин, при неограниченном возрастании числа слагаемых. Рассматриваемые теоремы лежат в основе асимптотических методов в теории вероятностей и математической статистике.

Обычно рассматривают два крупных блока предельных теорем: закон больших чисел и центральную предельную теорему.

Закон больших чисел утверждает, что среднее арифметическое большого числа независимых случайных величин мало отличается от среднего арифметического из их математических ожиданий.

Согласно центральной предельной теореме достаточно большая сумма примерно одинаковых случайных величин ведет себя приближенно как нормально распределенная случайная величина.

Сначала рассмотрим неравенства Маркова и Чебышева, которые можно использовать для грубой оценки вероятности событий, связанных со случайными величинами, распределение которых неизвестно.

Теорема (Неравенство Маркова). *Если неотрицательная случайная величина $X(\omega)$ имеет математическое ожидание m_x, то для любого положительного числа ε справедливо неравенство Маркова*

$$P(X(\omega) \geq \varepsilon) \leq \frac{m_x}{\varepsilon}.$$

Пример 1. Оценить вероятность того, что при 3600 независимых бросаниях игральной кости число появлений 6 очков будет не меньше 900.

Решение. Пусть X - число появлений 6 очков при 3600 бросаниях кости. Тогда $m_x = 3600 \cdot \frac{1}{6} = 600$.

Воспользуемся теперь неравенством Маркова при $\varepsilon = 900$:

$P(X \geq 900) \leq \frac{600}{900} = \frac{2}{3}$.

Теорема (Первое неравенство Чебышева). *Если случайная величина X имеет математическое ожидание m_x и дисперсию $D_x = \sigma_x^2$, то для любого положительного числа ε справедливо неравенство Чебышева*

$$P(|X - m_x| \geq \varepsilon) \leq \frac{\sigma_x^2}{\varepsilon^2}.$$

Отсюда следует, **второе неравенство Чебышева**:

$$P(|X - m_x| < \varepsilon) = P(m_x - \varepsilon < X < m_x + \varepsilon) \geq 1 - \frac{\sigma_x^2}{\varepsilon^2}.$$

Пример 2. Вероятность наступления события A в каждом из 1000 независимых опытов равна 0,8. Найдите вероятность того, что число наступлений события A в этих 1000 опытах отклонится от своего математического ожидания по абсолютной величине меньше чем на 50.

Решение. Пусть X - число наступлений события A в указанных 1000 опытах. Имеем дело со схемой Бернулли и соответствующим ей биномиальным распределением. Тогда $m_x = np = 1000 \cdot 0.8 = 800$ и $D = \sigma_x^2 = npq = 1000 \cdot 0.8 \cdot 0.2 = 160$. Используем второе неравенство Чебышева:

$P(|X - 800| < 50) \geq 1 - \frac{160}{50^2} = 0.936$.

Пример 3. Дисперсия каждой из 1000 независимых случайных величин x_k (k = 1,2,..., 1000) равна 4. Оценить вероятность того, что отклонение средней арифметической этих величин от средней арифметической их математических ожиданий по абсолютной величине не превзойдет 0.1.

Решение. Среднее арифметическое случайных величин x_k: $\bar{x} = \frac{1}{1000}\sum_{k=1}^{1000} x_k$;

среднее арифметическое математических ожиданий

$$\overline{m} = \frac{1}{1000}\sum_{k=1}^{1000} m_k = \frac{1}{1000} M\left(\sum_{k=1}^{1000} x_k\right) = \frac{1}{1000} M\left(\sum_{k=1}^{1000} x_k\right) = M(\overline{x}).$$

Здесь учтено свойство линейности математического ожидания для любых случайных величин X_k: $M\left(\sum_{k=1}^{n} a_k X_k + b\right) = \sum_{k=1}^{n} a_k M(X_k) + b$. В нашем случае $a_k = 1, b = 0$.

Применим второе неравенство Чебышева к заданной средней арифметической:

$$P(\overline{x} - M(\overline{x}) = P\left(\left|\frac{1}{1000}\sum_{k=1}^{1000} x_k - \frac{1}{1000}\sum_{k=1}^{1000} m_k\right| < 0.1\right) \ge 1 - \frac{D(\overline{x})}{0.1^2}.$$

Найдем дисперсию средней арифметической и учтем, что дисперсия суммы независимых случайных величин равно сумме дисперсий:

$$D(\overline{x}) = D\left(\frac{1}{1000}\sum_{k=1}^{1000} x_k\right) = \frac{1}{1000^2} D\left(\sum_{k=1}^{1000} x_k\right) = \frac{1}{(1000)^2}\sum_{k=1}^{1000} D(x_k) = \frac{1000D}{(1000)^2} = \frac{D}{1000}.$$

Здесь учтено свойство дисперсии для независимых случайных величин X_k:

$$D\left(\sum_{k=1}^{n} a_k X_k + b\right) = a_k^2 \sum_{k=1}^{n} D(X_k).$$

Теперь неравенство Чебышева примет вид:

$$P\left(\left|\frac{1}{1000}\sum_{k=1}^{1000} x_k - \frac{1}{1000}\sum_{k=1}^{1000} m_k\right| < 0.1\right) \ge 1 - \frac{D/1000}{0.01^2} = 1 - \frac{4}{1000 \cdot 0.1^2} = 0.6$$

Закон больших чисел в форме Чебышева:

Теорема (Чебышева). *Если случайные величины в последовательности $X_1, X_2, \ldots, X_n, \ldots$ попарно независимы, одинаково распределены, имеют одинаковые математические ожидания m и одинаковые дисперсии D, то последовательность средних арифметических*

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad n \in \mathbb{N}$$

сходится по вероятности к математическому ожиданию m:

$$\bar{X}_n \xrightarrow{P} m.$$

или $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right) = 1$.

Закон больших чисел в форме Бернулли:

Теорема (Бернулли). *Относительная частота $\frac{k}{n}$ события A при n независимых испытаниях по схеме Бернулли сходится по вероятности к вероятности $p = P(A)$ события A при одном испытании:*

$$\frac{k}{n} \xrightarrow{P} P(A) \text{ при } n \to \infty.$$

или $\lim\limits_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1$.

Отметим, что если произведено конечное число n независимых испытаний по схеме Бернулли, то для относительной частоты числа «Успехов» $\frac{k}{n}$ второе неравенство Чебышева имеет вид

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \ge 1 - \frac{pq}{n\varepsilon^2}.$$

Пример 4. Оценить вероятность того, что частота появления шестерки в 10000 независимых бросаниях игральной кости отклонится от вероятности появления шестерки по абсолютной величине меньше чем на 0,01.

Решение. Имеем дело со схемой Бернулли. Применительно к данной задаче второе неравенство Чебышева запишется в виде:

$$P\left(\left|\frac{k}{n} - p\right| < 0.01\right) = 1 - \frac{\frac{1}{6}\cdot\frac{5}{6}}{10000 \cdot 0.01^2} \approx 0.861$$

Центральная предельная теорема в форме Ляпунова:

Теорема (Ляпунова). *Пусть случайные величины в последовательности $X_1, X_2, \ldots, X_n, \ldots$ независимы, одинаково распределены, имеют равные математические ожидания m и равные дисперсии $D = \sigma^2$.*

Тогда функция распределения $F_{Z_n}(z)$, $n \in \mathbb{N}$ центрированной и нормированной суммы данных случайных величин

$$Z_n = \frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}}, \quad n \in \mathbb{N}$$

сходится для всех $z \in \mathbb{R}$ к функции распределения Лапласа

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z).$$

Следствием центральной предельной теоремы является теорема Муавра–Лапласа, относящаяся к схеме Бернулли.

Теорема (Муавра–Лапласа). *Пусть случайная величина B_n имеет биномиальное распределение и определена как число появлений события A при n независимых испытаниях по схеме Бернулли с вероятностью $p = P(A)$ успеха события A в каждом испытании.*

Тогда функции распределения $F_{\bar{B}_n}$, $n \in \mathbb{N}$ центрированной и нормированной последовательности случайных величин $\bar{B}_n = \frac{B_n - np}{\sqrt{npq}}$, $n \in \mathbb{N}$ сходятся для всех $z \in \mathbb{R}$ к функции распределения Лапласа $\lim\limits_{n\to\infty} F_{\bar{B}_n}(z) = \Phi(z)$.

Теорема Муавра–Лапласа теоретически обосновывает возможность приближения дискретного биномиального распределения непрерывным нормальным распределением с характеристиками $m = pn$ и $\sigma = \sqrt{npq}$.

В частности, для вычисления вероятности попадания биномиальной случайной величины на дискретный интервал $[k_1, k_2)$ при достаточно больших значениях npq применяется интегральная формула Муавра – Лапласа

$$P(k_1 \le B_n < k_2) = P\left(\frac{k_2 - np}{\sqrt{npq}} \le \frac{B_n - np}{\sqrt{npq}} < \frac{k_1 - np}{\sqrt{npq}}\right) =$$
$$= \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) + O\left(\frac{1}{\sqrt{npq}}\right),$$

а для вычисления вероятности любого из возможных значений $k = 0, 1, \ldots, n$ биномиальной случайной величины – локальная формула Муавра–Лапласа

$$P(B_n = k) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} + O\left(\frac{1}{\sqrt{npq}}\right).$$

Высокая точность приближения достигается в тех случаях, когда выполняются условия $npq > 5$ и $0,1 < p < 0,9$. При условии $npq > 25$ приближение можно применять при любых вероятностях $p = P(A)$ события A.

Пример 5. Вероятность наступления события А в каждом из 900 независимых испытаний равна р = 0,8. Найдите вероятность того, что событие А произойдет: а) 750 раз; б) от 710 до 740 раз.

Решение: Так как npq = 900 · 0.8 · 0.2 = 144 >25 , то можно воспользоваться локальной и интегральной формулами Муавра-Лапласа.

а) $P_{900}(750) = \frac{1}{\sqrt{2\pi \cdot 900 \cdot 0.8 \cdot 0.2}} e^{-\frac{(750 - 900\cdot 0.8)^2}{2\cdot 900\cdot 0.8\cdot 0.2}} = \frac{1}{\sqrt{288\pi}} e^{-\frac{1}{0.32}} \approx 0.00146$

При расчетах можно воспользоваться табулированной функцией плотности нормального распределения $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Тогда

$$P_{900}(750) = \frac{1}{\sqrt{900 \cdot 0.8 \cdot 0.2}} \varphi(x)$$

В нашем случае

$x = \frac{750 - 900 \cdot 0.8}{\sqrt{900 \cdot 0.8 \cdot 0.2}} = \frac{1}{\sqrt{0.16}} = 2.5$, $\varphi(2.5) = 0.0175$ и

$$P_{900}(750) = \frac{1}{12} \varphi(2.5) \approx 0.00146$$

б) Воспользуемся аппроксимацией биномиального распределения нормальным распределением $N(pn, \sqrt{pnq})$ и применим интегральную формулу Муавра-Лапласа:

$$P(k_1 \le B_n < k_2) \cong \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right);$$

по условию n=900, k_1=710, k_2=740, p=0.8, q=0.2

$$P(710 \le k < 740) = \Phi\left(\frac{740 - 720}{\sqrt{144}}\right) - \Phi\left(\frac{710 - 720}{\sqrt{144}}\right) = \Phi(1.67) - \Phi(-0.83) =$$

$$= \Phi(1.67) - 1 + \Phi(0.83) = 0.9525 - 1 + 0.7967 = 0.7492$$

Пример 6. Вероятность того, что электролампочка, изготовленная данным заводом, является бракованной, равна 0,02. Для контроля отобрано наудачу 1000 лампочек. Оценить вероятность того, что частота бракованных лампочек в выборке отличается от вероятности 0,02 менее чем на 0,01.

Решение. Пусть k – число бракованных лампочек в выборке. Нам нужно оценить вероятность выполнения неравенства

$\left|\frac{k}{1000} - 0{,}02\right| < 0{,}01$. Оно равносильно неравенству $11 \le k < 30$. Следовательно

$P\left(\left|\frac{k}{1000} - 0{,}02\right| < 0{,}01\right) = P_{1000}(11 \le k < 30)$. Тогда n=1000, k_1=11, k_2=30, p=0.02, q=0.98

$npq = 1000 \cdot 0{,}02 \cdot 0{,}98 = 19{,}6$, что вполне приемлемо.

$$P(11 \le k < 30) = \Phi\left(\frac{30 - 20}{\sqrt{19.6}}\right) - \Phi\left(\frac{11 - 20}{\sqrt{19.6}}\right) = \Phi(2.26) - \Phi(-2.03) =$$

$$= \Phi(2.26) - 1 + \Phi(2.03) = 0.9881 - 1 + 0.9788 = 0.9669$$

Распределение Пуассона (см. Памятку)

Пример. Пусть в среднем за 30 суток в городе Энске возникает 60 ситуаций, требующих оперативно вмешательства МЧС. На сколько вызовов в сутки должна быть рассчитана Энская служба МЧС, чтобы с вероятностью не меньшей 0.96 выехать на все поступившие вызовы.

Решение. Пусть служба рассчитана на k заявок в сутки, которое пока неизвестно. Предположим, что m – действительное число поступивших за сутки вызовов. Тогда k найдем из условия $P(m \le k) \ge 0.96$. Среднее число заявок за сутки $\lambda=60/30=2$. Воспользуемся формулой Пуассона

$P(m \le k) = e^{-\lambda} \sum_{m=0}^{k} \frac{\lambda^m}{m!}$, чтобы подобрать k.

$$e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} + \frac{2^5}{5!}\right) \approx 0.983 \Rightarrow k = 5$$

ИДЗ. Методичка Попов-Яблонская. Задание 3, стр. 64-66, задание 4, стр. 67-68.

Срок сдачи 23:59 17.12.2021.
