Предельные теоремы теории вероятностей. Основные понятия математической статистики.

Лекция 16

Числовые характеристики суммы независимых случайных величин

Пусть $x_1, x_2 \ldots x_n$ взаимно независимые, одинаково распределенные случайные величины с параметрами: $M[x_k] = m; \quad \mathcal{D}[x_k] = \sigma^2$.

Случайная величина $\bar{x} = \frac{\sum_{k=1}^{n} x_k}{n}$, которую называют **средним арифметическим**, имеет характеристики: $M[\bar{x}] = M\left[\frac{\sum_{k=1}^{n} x_k}{n}\right] = \frac{1}{n} M[\sum_{k=1}^{n} x_k] = \frac{1}{n} nm = m;$

$$\mathcal{D}[\bar{x}] = \mathcal{D}\left[\frac{\sum_{k=1}^{n} x_k}{n}\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad ; \quad \sigma[\bar{x}] = \frac{\sigma}{\sqrt{n}}$$

Каждое слагаемое н*ормированной и центрированой случайной величины*

$$Y_n = \frac{\bar{x} - M[\bar{x}]}{\sigma[\bar{x}]} = \frac{\sum_{k=1}^{n} \frac{x_k}{n} - \sum_{k=1}^{n} \frac{m}{n}}{\frac{\sigma}{\sqrt{n}}} = \sum_{k=1}^{n} \frac{x_k - m}{\sigma\sqrt{n}}$$ имеет характеристики:

$$M[Y_k] = M\left[\frac{x_k - m}{\sigma\sqrt{n}}\right] = \frac{1}{\sigma\sqrt{n}}(M[x] - M[m]) = \frac{1}{\sigma\sqrt{n}}(m - m) = 0$$

$$\mathcal{D}[Y_k] = \mathcal{D}\left[\frac{x_k - m}{\sigma\sqrt{n}}\right] = \frac{1}{\sigma^2 n}(\mathcal{D}[x_k] - \mathcal{D}[m]) = \frac{\sigma^2}{\sigma^2 n} = \frac{1}{n}$$

Поскольку $M[Y_k] = 0$, $\mathcal{D}[Y_k] = M[Y_k^2] - M[Y_k] = M[Y_k^2] = \frac{1}{n}$

Характеристическая функция суммы независимых случайных величин

Характеристическая функция каждого слагаемого:

$$G_{Y_k}(t) = M\left[e^{ity_k}\right] = M\left[1 + iy_k t + y_k^2 \frac{(it)^2}{2!} \dots\right] = 1 + M[Y_k]it + M\left[Y_k^2\right]\frac{(it)^2}{2!}$$

$$= 1 + \frac{1}{n}\frac{(it)^2}{2!} + \cdots$$

Характеристическая функция суммы $\sum_{k=1}^{n} Y_k$:

$$G_{Y_n}(t) = \left(G_{Y_k}(t)\right)^n = \left(1 + \frac{1}{n}\frac{(it)^2}{2!} \dots\right)^n$$

При $n \to \infty$ получаем неопределенность $[1^\infty]$, которую раскрываем, используя основное логарифмическое тождество и разложение в ряд логарифмической функции $\left(\ell n\,(1+x)_{x\to 0} = x + 0(x)\right)$:

$$\lim_{n\to\infty} G_{Y_n}(t) = e^{\lim\limits_{n\to\infty} n\ell n\left(1+\frac{1}{n}\frac{(it)^2}{2!}\dots\right)} = e^{\lim\limits_{n\to\infty} n\left(\frac{1}{n}\left(\frac{-t^2}{2}\right)\right)} = e^{-\frac{t^2}{2}}.$$

В результате сформулируем **центральную предельную теорему.**

Центральная предельная теорема

Если случайные величины $x_1,\ x_2\ \ldots .\ x_n$ независимы и одинаково распределены, а также имеют конечные математическое ожидание и дисперсию: $\boldsymbol{M}\ [\boldsymbol{x_k}] = \boldsymbol{m};\quad \boldsymbol{\mathcal{D}}[\boldsymbol{x_k}] = \boldsymbol{\sigma^2}$,

то для любого действительного x закон распределения нормированного и центрированного среднего арифметического $\boldsymbol{n}$ случайных величин при $\boldsymbol{n} \to \infty$ стремится к нормальному закону распределения с параметрами $\boldsymbol{M}\ [\boldsymbol{x}] = \boldsymbol{m} = \boldsymbol{o}$ **и** $\boldsymbol{\sigma} = \boldsymbol{1}$ (подробнее смотри Теоремы. docx) **:**

$$\boldsymbol{F_{Y_n}(x) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz\ ;\quad z = \frac{x-m}{\sigma}\ ;\quad G_{Y_n}(t) \to e^{-\frac{t^2}{2}},\qquad n \to \infty}$$

Таким образом**, нормальное распределение** является предельной формой распределения суммы большого числа случайных величин, из которых ни одна не доминирует над другой.

Теоремы Муавра - Лапласа

Рассматриваем биномиальное распределение (схема Бернулли): вероятность того, что при n испытаниях событие A появится k раз:

$$P(n,k,p) = c_n^k p^k q^{n-k}; \quad p(A) = p; \quad p + q = 1; \quad p(\bar{A}) = q.$$

При достаточно больших значениях npq биномиальное распределение приближенно заменяют нормальным распределением :

$$M[x] = np; \quad D[x] = npq; \quad \sigma[x] = \sqrt{npq} \quad \Longrightarrow \quad z = \frac{x-m}{\sigma} = \frac{k-np}{\sqrt{npq}}$$

Вероятность того, что при n испытаниях событие A появится k раз:

$$P(n,k,p) \approx \frac{1}{\sqrt{2\pi}\ \sqrt{npq}} e^{-\frac{z^2}{2}}; \quad z = \frac{k-np}{\sqrt{npq}}$$ **(Локальная теорема Муавра –Лапласа)**

Вероятность того, что при n истытаниях число событий A удовлетворяет **условию $k_1 \le k < k_2$**:

$$P(k_1 \le k < k_2) = \phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_1 - np}{\sqrt{npq}}\right),$$

(Интегральная теорема Муавра-Лапласа)

Закон больших чисел в форме Бернулли

Найдем вероятность того, что относительная частота события отличается от его вероятности не более, чем на $\mathcal{E}$: $\boldsymbol{P\left(\left|\frac{k}{n}-p\right| \leq \mathcal{E}\right)} = ?$

С учётом того, что $\left|\frac{k-np}{n}\right| \leq \mathcal{E} \to |k-np| \leq \mathcal{E}\cdot n \to -\mathcal{E}n \leq k-np \leq \mathcal{E}n \to$

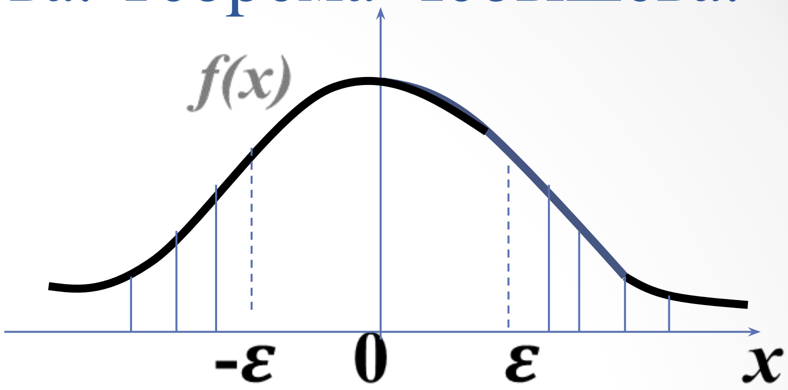
$np-\mathcal{E}n \leq k \leq np+\mathcal{E}n$, используем интегральную теорему Муавра-Лапласа получаем: $P\left(\left|\frac{k}{n}-p\right| \leq \mathcal{E}\right) = P(np-\mathcal{E}n \leq k \leq np+\mathcal{E}n) == \phi\left(\frac{np+\mathcal{E}n-np}{\sqrt{npq}}\right) - \phi\left(\frac{np-\mathcal{E}n-np}{\sqrt{npq}}\right) == \phi\left(\frac{\mathcal{E}n}{\sqrt{npq}}\right) - \phi\left(\frac{-\mathcal{E}n}{\sqrt{npq}}\right) = \phi\left(\mathcal{E}\sqrt{\frac{n}{pq}}\right) - 1 + \phi\left(\mathcal{E}\sqrt{\frac{n}{pq}}\right) = 2\phi\left(\mathcal{E}\sqrt{\frac{n}{pq}}\right) - 1.$ $\lim\limits_{n\to\infty} P\left(\left|\frac{k}{n}-p\right| \leq \mathcal{E}\right) = \lim\limits_{n\to\infty}\left(2\phi\left(\mathcal{E}\sqrt{\frac{n}{pq}}\right) - 1\right) = 2\phi(\infty) - 1 = 1$

$$\boldsymbol{\lim_{n\to\infty} P\left(\left|\frac{k}{n}-p\right| \leq \mathcal{E}\right) = 1}$$

Относительная частота события в $\boldsymbol{n}$ независимых испытаниях при $\boldsymbol{n \to \infty}$ стремится к вероятности одного испытания

Неравенство Чебышева. Теорема Чебышева.

- $$\boldsymbol{P(|x| \geq \mathcal{E}) \leq \frac{\mathcal{D}[x]}{\mathcal{E}^2}}$$



Дисперсия совпадает со вторым начальным моментом:

$\sigma^2 = \mathcal{D}[x] = M[x^2] - M^2[x] = M[x^2] = \int_{-\infty}^{\infty} x^2 f(x)dx \geq \int x^2\, f(x)dx_{|x|\geq\mathcal{E}} \geq \mathcal{E}^2 \int f(x)dx_{|x|\geq\mathcal{E}}$: площадь под графиком $f(x)$ равна 1 и она больше, чем площадь под " хвостами " распределений. $\Longrightarrow$ $\int_{|x|\geq\mathcal{E}} f(x)dx = P(|x| \geq \mathcal{E})$

Для случайной **величины** $\boldsymbol{X = \bar{x} - m}$ – отклонение среднего от $M[x]$ $\Longrightarrow$

$M[X] = M[\bar{x}] - m = 0;\quad \mathcal{D}[x] = \mathcal{D}[\bar{x}] + \mathcal{D}[m] = \frac{\sigma^2}{n}$ $\Longrightarrow$ неравенство Чебышева:

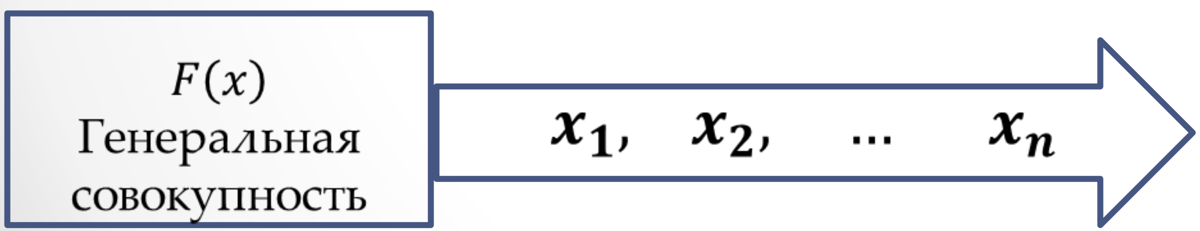
$$P(|X| \geq \mathcal{E}) \leq \frac{\sigma^2}{\mathcal{E}n} \to P(|x| \leq \mathcal{E}) = P(|\bar{x} - M[x]| \leq \mathcal{E}) \geq 1 - \frac{\sigma^2}{n\mathcal{E}^2}$$

$\boldsymbol{\lim_{n\to\infty} P(|\bar{x} - M[x]| \leq \mathcal{E}) = 1}$ (теорема Чебышева).

Математическая статистика

позволяет получать обоснованные выводы о видах распределения, параметрах и других свойствах случайных величин по совокупности наблюдений над ними – *выборке*.

Пусть случайная величина $\boldsymbol{X}$ распределена по закону $F(x)$ и наблюдается в эксперименте $\boldsymbol{D}$, а опыт повторяется $\boldsymbol{n}$ раз в одних и тех же условиях. В результате получаем последовательность наблюдений значений случайной величины или **$\boldsymbol{n}$ случайно отобранных объектов** $\{\boldsymbol{x_1}, \quad \boldsymbol{x_2}, \quad \ldots \quad \boldsymbol{x_n}\}$, которую **называют выборкой** из генеральной совокупности с законом распределения $F(x)$:



$\boldsymbol{n}$ - объем выборки

Далее все выводы делаются на основе выборки.

Основные задачи математической статистики

1. Сбор статистического материала (получение выборки)

2. Результаты наблюдений, записанные в порядке регистрации неудобны для анализа. Поэтому **вторая задача** статистического описания - **получение такого представления выборки, которое позволяет выявить характерные особенности распределения** (*группировка данных по интервалам, определение частот элементов выборки, построение полигона частот, гистограммы, эмпирической функции распределения*)

3. Получение числовых характеристик выборки и оценка параметров распределения.

4. На основе полученных оценок и характерных особенностей распределения выборки выдвигается гипотеза (предположение) о виде распределения генеральной совокупности или строится другая вероятностная модель описания данных

5. Выполняется проверка **статистической значимости** (оценка погрешности) и **адекватности** (соответствия модели экспериментальным данным) построенной вероятностной модели.