Повесьте ваши уши на гвоздь внимания !!!!!!

Неравенство Чебышева.

Неравенство Маркова (или лемма Чебышева)

Если случайная величина X принимает только неотрицательные значения и имеет математическое ожидание EX, то для любого положительного числа α справедливо неравенство: $P(X \geq \alpha) \leq \frac{MX}{\alpha}.$

Теорема (*неравенство Чебышева*):

Если случайная величина X имеет математическое ожидание EX и дисперсию DX, то для любого ε > 0 справедливо неравенство:

$$\boxed{P(|X - MX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.}$$

Теорема Чебышева. Закон больших чисел (ЗБЧ).

<u>Введем понятие сходимости по вероятности:</u>

$$\lim_{n\to\infty} P\left(\left|X_n - a\right| < \varepsilon\right) = 1 \quad \textit{для сколь угодно}$$

малого положительного ε;

или другая более компактная форма записи :

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} a.$$

Формулировка ЗБЧ в форме Чебышева П.Л. ***(теорема Чебышева):***

Если дисперсии n независимых случайных величин $X_1, X_2, \ldots, X_n$ ограничены сверху одной и той же константой: $DX_i \le C,\ i=1, 2, \ldots, n,$

то для любого сколь угодно малого положительного числа ε

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{m_1+m_2+\ldots+m_n}{n}\right|<\varepsilon\right)=1, \text{где} \quad m_i = MX_i.$$

$$\text{Или} \quad \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{P} \frac{\sum_{i=1}^{n} m_i}{n}.$$

Следствия ***из теоремы Чебышева***:

Первое следствие: **Теорема Хинчина**

Если независимые случайные величины X_1, $X_2, \ldots, X_n$ имеют одинаковые математические ожидания, равные m, то

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n \to \infty]{P} m.$$

Это соотношение является основой <u>выборочного метода</u> (статистических исследований). Если мы хотим узнать истинное значение какого-то параметра m, нам нужно несколько раз экспериментально получить значения X_i этого параметра и затем на основе этих значений вычислить их среднее арифметическое. Вычисленная величина будет достаточно хорошим приближением истинного значения параметра, причем чем больше включено в расчет экспериментальных значений, тем более точное приближение истинного значения параметра будет получено.

Второе следствие: **Теорема Бернулли**

Пусть проводится *n* независимых испытаний, в каждом из которых событие А может произойти с одной и той же вероятностью *p* (схема Бернулли). При неограниченном возрастании числа опытов *n* частота события А сходится по вероятности к вероятности *p* этого события в отдельном испытании:

$$\frac{k}{n} \xrightarrow[n\to\infty]{P} p$$

Здесь k - количество случаев, когда событие А наблюдалось.

Третье следствие:

ЗБЧ может быть распространен и на зависимые случайные величины (это обобщение принадлежит Маркову А.А.):

Если имеются зависимые случайные величины $X_1, X_2, \ldots, X_n$ и если при

$$\text{при} \quad n \to \infty \quad \frac{D\left(\sum_{i=1}^{n} X_i\right)}{n^2} \xrightarrow[n\to\infty]{} 0, \quad \text{что было выполнено в предыдущей}$$

$$\text{теореме,} \quad \text{то} \quad \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{P} \frac{\sum_{i=1}^{n} m_i}{n}.$$

Смысл и формулировка центральной предельной теоремы (ЦПТ). Интегральная теорема Муавра-Лапласа как следствие ЦПТ.

Эта теорема утверждает, что распределение суммы большого числа независимых и сравнимых по вкладам в сумму случайных величин близко к нормальному закону распределения.

Иначе:

если $Y_n = X_1 + X_2 + \ldots + X_n$, причем

1) Слагаемых много;
2) Слагаемые независимые;
3) Слагаемые сравнимы по вкладам в сумму, т.е. нет слагаемого, которое было бы по вкладу существенно больше остальных,

то ЦПТ утверждает, что СВ Y_n подчиняется нормальному закону распределения.

Именно поэтому нормальный закон распределения так широко применяется в практических задачах, ибо в реальных задачах исследуемые случайные величины часто есть результат сложения многих других случайных величин.

Упрощенная математическая формулировка ЦПТ:

Если X_1 , X_2 ,…, X_n – независимые случайные величины, для каждой из которых существует математическое ожидание $EX_i = m_i$ и дисперсия $DX_i=\sigma_i^{\ 2}$, а также выполняется некоторое дополнительное условие , то закон распределения $Y_n = X_1 + X_2 + \ldots + X_n$ при $n \to \infty$ асимптотически приближается к нормальному закону распределения с параметрами

$$m_{Y_n} = \sum_{i=1}^{n} m_i \qquad u \qquad DY_n = \sum_{i=1}^{n} \sigma_i^{\ 2}.$$

Что касается упомянутого в формулировке теоремы дополнительного условия, то оно сложно записывается математически, но означает, что вклад каждого слагаемого в сумму ничтожно мал, т.е. слагаемые соразмерны по своим вкладам в сумму.

Из ЦПТ для схемы испытаний Бернулли вытекает как следствие интегральная теорема Муавра – Лапласа.

Многомерная случайная величина и закон ее распределения.

Пусть имеется система случайных величин (СВ), причем эта система может состоять как из дискретных, так и из непрерывных СВ. Будем рассматривать их как координаты случайного вектора.

Определение. n-мерной случайной величиной или случайным вектором называется упорядоченный набор n случайных величин

$$\overrightarrow{X} = (X_1, X_2, ..., X_n)$$

Для описания поведения многомерной СВ должен быть введен закон ее распределения:

$$F(x_1, x_2, ..., x_n) = (опред.) =$$
$$= P(X_1 < x_1, X_2 < x_2, ..., X_n < x_n)$$

Эта функция выражает вероятность совместного выполнения неравенств в правой части этого соотношения.

С целью экономии времени изложение выполним для двумерного случая; при этом будем понимать, что все утверждения справедливы и для n>2:

$$F(x,y) = P(X < x, Y < y)$$

Рассмотрены свойства функции F(x,y).

Могут быть получены **частные (маргинальные)** функции распределения на основе функции совместного распределения двух случайных величин:

$$F(x,+\infty) = F_X(x) \qquad F(+\infty,y) = F_Y(y)$$

Для двумерной непрерывной случайной величины (X,Y) функция совместного распределения может быть представлена в виде:

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\cdot du\cdot dv,$$ *причем если продифференцировать это равенство по x и y, то найдем другую формулу связи между этими функциями:*

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\cdot\partial y}$$

Для функции f(x,y), которая называется плотностью совместного распределения, справедливы те же **свойства**, которые были получены для функции f(x) в одномерном случае.

Зная плотность совместного распределения двух случайных величин, можно найти плотность **частного (маргинального**) распределения одной случайной величины:

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\cdot dy;\quad f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\cdot dx$$

Для независимых случайных величин X и Y независимы события {X<x} и {Y<y}, откуда следует:

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

Для непрерывных СВ из данного соотношения, дифференцируя его по x и y, получим:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

Для зависимых СВ эти равенства не выполняются:

$$F(x, y) \neq F_X(x) \cdot F_Y(y); \qquad f(x, y) \neq f_X(x) \cdot f_Y(y)$$

Стохастическая зависимость двух случайных величин. Ковариация и коэффициент корреляции.

Если случайные величины зависимы, влияют на поведение друг друга, то следует количественно описать степень их влияния друг на друга.

Определение.

Ковариацией двух СВ X и Y называется математическое ожидание произведения соответствующих центрированных СВ:

$$\boxed{\mathrm{cov}\,(X, Y) = E((X - EX) \cdot (Y - EY)) =}$$

$$\begin{cases} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(x_i - EX)\cdot(y_j - EY)\cdot p_{ij} - ДСВ;\ p_{ij} \to (x_i; y_j) \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x - EX)\cdot(y - EY)\cdot f(x, y)\cdot dxdy - НСВ \end{cases}$$

Рассмотрены свойства ковариации.

<u>Вывод</u>:
ковариация не улавливает сложные виды связей между X и Y. Ковариация отслеживает наличие только **линейной связи** между СВ. При наличии такой линейной связи (стохастической) ковариация отлична от 0.

Определение:

<u>Коэффициентом корреляции</u> двух СВ X и Y называется отношение их ковариации к произведению стандартных отклонений этих величин:

$$\boxed{\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_X\cdot\sigma_Y}}$$

Рассмотрены свойства коэффициента корреляции.
Значения, принимаемые коэффициентом корреляции:

$$\boxed{-1\le\rho\le 1}$$

Определение.
Случайные величины называются <u>некоррелированными</u>, если их коэффициент корреляции равен нулю. Случайные величины называются <u>коррелированными,</u> если их коэффициент корреляции отличен от нуля.

Было показано, что если случайные величины независимые, то они некоррелированные, а из некоррелированности случайных величин еще не следует их независимость. <u>Из некоррелированности нормальных СВ следует их независимость</u> (в общем случае это не так.)

Коэффициент корреляции характеризует степень линейной зависимости между случайными величинами X и Y в стохастическом смысле и не может отражать более сложных видов зависимостей между случайными величинами.

Графически показана стохастическая линейная связь между случайными величинами при различных значениях коэффициента корреляции.

Введено уравнение линейной регрессии, наилучшим образом описывающим связь между случайными величинами:

$$y = EY + \rho \cdot \frac{\sigma_y}{\sigma_x} \cdot \left(x - EX\right)$$

Для вычисления коэффициента корреляции между двумя количественными признаками на практике используется **<u>линейный коэффициент корреляции Пирсона:</u>**

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Введем коэффициент корреляции для изучения тесноты связи между порядковыми случайными величинами.

Если n объектов совокупности пронумеровать в соответствии с возрастанием или убыванием изучаемого признака, то говорят, что объекты **ранжированы** по этому признаку. Присвоенный номер называется **<u>рангом.</u>**

<u>Коэффициент ранговой корреляции Спирмена</u> вычисляется по формуле:

$$\rho_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)}, \quad \textit{где}$$

r_i – *ранги по первому признаку*; s_i – *ранги по второму признаку*;

$d_i = (r_i - s_i)$ – *разность рангов* i*–ого объекта*; $|\rho_s| \le 1$

В случае совпадения рангов при вычислении коэффициента ранговой корреляции следует брать среднее арифметическое рангов, приходящихся на данные объекты, причем каждому объекту присваивается это среднее арифметическое значение. В формулу вводятся поправки на совпадающие ранги T_a и T_b . Формула приобретает такой вид:

$$\rho_s = \frac{\frac{1}{6}\cdot\left(n^3 - n\right) - \sum_{i=1}^{n} d_i^2 - T_a - T_b}{\sqrt{\frac{1}{6}\cdot\left(n^3 - n\right) - 2T_a}\cdot\sqrt{\frac{1}{6}\cdot\left(n^3 - n\right) - 2T_b}}$$

$$T_a = \frac{1}{12}\cdot\sum_{k=1}^{m_a}\left(a_k^3 - a_k\right), \qquad T_b = \frac{1}{12}\cdot\sum_{k=1}^{m_b}\left(b_k^3 - b_k\right), \quad где$$

a_k, b_k – *объемы каждой группы с совпадающими рангами по первому и по второму признаку;*

m_a, m_b – *количество групп с совпадающими рангами по первому и по второму признаку.*

Благодарю за внимание!