Билеты по курсу «Математическая статистика»

-1. Чем занимается математическая статистика.

Есть некий набор данных и по этому набору данных нужно сделать выводы о характере вероятностного распределения и то, откуда эти данные пришли.

0. Сведения из теории вероятностей.

Случайная величина.

Есть исходное вероятностное пространство $(\Omega, \mathcal{A}, P)$, где Ω — пространство элементарных исходов (множество, содержащее всевозможные взаимоисключающие результаты данного эксперимента), $\mathcal{A}$ — сигма-алгебра пространства элементарных исходов Ω (система подмножеств Ω, содержащая пустое множество, каждое дополнение к элементу множества $\mathcal{A}$, объединение любого числа (в т. ч. и бесконечного) элементов множества $\mathcal{A}$), P — вероятностная мера, заданная на сигма-алгебре $\mathcal{A}$, имеющая вид функции $P: \mathcal{A} \to [0,1]$.

Есть некоторое измеримое отображение $X: (\Omega, \mathcal{A}, P) \overset{X}{\to} (\mathbb{R}, \mathfrak{B}, P_X)$, где $\mathbb{R}$ — множество действительных чисел, $\mathfrak{B}$ – борелевская сигма-алгебра (минимальная сигма-алгебра, которая содержит все открытые интервалы), P_X — распределение случайной величины.

Измеримое отображение. Есть два пространства с выделенными сигма-алгебрами на них: $(X, \mathcal{A})$ и $(Y, \mathcal{B})$. Отображение $\phi: X \to Y$ называется измеримым, если $\forall\, B \in \mathcal{B}: \phi^{-1}(B) \in \mathcal{A}$. То есть, прообраз любого подмножества $\mathcal{B}$ содержится в $\mathcal{A}$.

Распределение случайной величины $P_X(\mathfrak{B}) = P(\{\omega: X(\omega) \in \mathfrak{B}\}) = P(X \in \mathfrak{B})$.

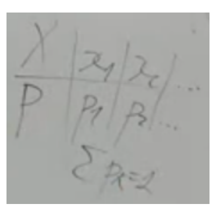
Функция распределения случайной величины X: $F_X(t) = P(X \leq t)$. Её свойства: монотонно возрастает, $F_X(-\infty) = 0, F_X(+\infty) = 1$, непрерывна справа. Функция распределения однозначно определяем распределение случайной величины.

Тогда X – случайная величина.

Типы распределений.

Дискретные случайные величины и их распределения.

X — дискретная случайная величина $\Leftrightarrow \exists$ множество E не более чем счетное, такое, что $P(X \in E) = 1$. Иначе говоря, можно задать таблицу:



X	x_1	x_2	...
P	p_1	p_2	...

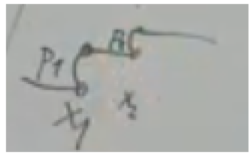
$\sum p_k = 1$

Вырожденное распределение: $P(X = const) = 1$.

Распределение Бернулли: $Bern(p)$. $X \sim Bern(p)$ (такая запись означает, что случайная величина X имеет распределение Бернулли с параметром p). Распределение Бернулли характеризуется тем, что случайная величина принимает значение 1 с вероятностью p и значение 0 с вероятностью $(1-p)$.

Биномиальное распределение: $Bin(n,p)\colon P(X=k)=C_n^k p^k q^{n-k},\ q=1-p, k=\overline{0,n}.$

В дискретном случае функция распределения имеет кусочно-постоянный вид:



Геометрическое распределение: $Geom(p)$. Есть две различные трактовки: число неудач до первого успеха, либо номер первого успеха. Рассмотрим первую трактовку:

0	1	2	3	n
p	p(1-p)	...		p(1-p)^n

Дискретное равномерное распределение.

Распределение Пуассона: $Pois(\lambda>0)\colon P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}, k\in Z_+$

Абсолютно непрерывные распределения.

X – непрерывна $\Longrightarrow \exists\, p(x)\geq 0\colon\ P(X\in\mathfrak{B})=\int_{\mathfrak{B}} p(x)dx$. ($p(x)$ – плотность распределения непрерывной случайной величины).

Свойство абсолютно непрерывных распределений: $if\ X-\text{непр}, \text{то}\ P(X=const)=0.$

Тогда функцию распределения можем расписать через плотность: $F_X(t)=\int_{-\infty}^{t}p(x)dx$ и $p(x)=F_X'(x)$.

Равномерное распределение: $U[a,b], p(x)=\mathbb{I}(x\in[a,b])\cdot\frac{1}{b-a}$,

где $\mathbb{I}(condition)\Leftrightarrow if condition\ true\ than\ 1, else\ 0.$

Свойство равномерного распределения (линейность): пусть $X\sim U[0,1], Y=bX+c, b>0\ \ \Longrightarrow Y\sim U[c,b+c]$.

Нормальное распределение: $\mathcal{N}(\mu,\sigma^2)\colon p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{2\sigma^2}\right), \mu$ –математическое ожидание, σ^2 – дисперсия, σ – стандартное отклонение.

Стандартное нормальное распределение: $\mathcal{N}(0,1)$.

Свойство стандартного нормального распределения: $X\sim\mathcal{N}(0,1), Y=\sigma X+\mu\ \ \Longrightarrow\ \ Y\sim\mathcal{N}(\mu,\sigma^2)$.

Также обратное свойство:

$$Y\sim\mathcal{N}(\mu,\sigma^2)\Rightarrow\frac{X-\mu}{\sigma}\sim\mathcal{N}(0,1)$$

Экспоненциальное распределение: $Exp(\lambda): p(x) = \lambda e^{-\lambda x}\mathbb{I}(x \geq 0)$.

Независимость.

Случайные величины независимы в совокупности, если $P(X_1 \in \mathfrak{B}_1, \dots, X_n \in \mathfrak{B}_n) = \prod_{i=1}^{n} P(X_i \in \mathfrak{B}_i)$, $\mathfrak{B}_i \in \mathcal{B}$(борелевская сигма-алгебра).

Для независимых случайных величин также можно расписать их функции распределения следующим образом: $P(X_1 \leq t_1, \dots, X_n \leq t_n) = \prod_{k=1}^{n} F_{X_k}(X_k \leq t_k)$.

Тогда можно рассматривать многомерную функцию совместного распределения $F(t_1, \dots, t_k)$.

Также, если случайные величины независимы, то их совместная плотность $p(x_1, \dots, x_n) = \prod_{k=1}^{n} p(x_k)$ (для непрерывных величин), а для дискретного случая: $P(X_1 = k_1, \dots, X_n = k_n) = \prod_{j=1}^{n} P(X_j = k_j)$.

Сумма независимых случайных величин.

$S_n \sim Bin(n, p)$, тогда S_n можно рассматривать как сумму случайных величин, то есть $S_n = \sum_{i=1}^{n} X_i$, $X_i - i.i.d.(independent, identical\ disturb), Bern(p)$.

Свойство нормального распределения: есть набор случайных величин $Y_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$,

тогда величина $Y_1 + \dots + Y_n$ будет иметь следующее распределение $\mathcal{N}\left(\sum \mu_k, \sum \sigma_k^2\right)$.

Числовые характеристики случайных величин.

Рассмотрим интеграл $\int_{\mathbb{R}} g(x)dF(x)$, где $F(x)$ — некоторая функция распределения, $g(x)$ – непрерывная функция. Он равен $= \lim\limits_{m \to \infty} \sum_{k=1}^{m} g(x_k)(F(x_k) - F(x_{k-1})) = \begin{cases} \sum p_i g(x_i) - \text{дискретный случай} \\ \int_{\mathbb{R}} g(x)p(x)dx - \text{непрерывный случай} \end{cases}$.

Математическое ожидание случайной величины X: $E(X) = \int_{\mathbb{R}} x dF(x)$ (данный интеграл имеет смысл, если существует интеграл $\int_{\mathbb{R}} |x| dF(x)$). Итого, по аналогии с рассмотренным выше интегралом, получим знакомые формулы для математических ожиданий дискретной и непрерывной случайной величины.

Свойства матожидания: линейно, $E(XY) = E(X) \cdot E(Y)$ (если X и Y независимы).

Дисперсия (variance) случайной величины X: $E(X - EX)^2 = E(X^2) - \left(E(X)\right)^2$. Корень из дисперсии – стандартное отклонение.

Свойства дисперсии: $D(aX) = a^2 D(X), D(X) \geq 0, D(X) = 0 \Leftrightarrow X \text{ имеет вырожденное распределение}, D(X_1 + X_2) = D(X_1) + D(X_2)$ (если X_1 и X_2 – независимы).

Мода случайной величины X: точка максимума плотности (непрерывный случай); значение, соответствующее максимальной вероятности (дискретный случай).

Квантиль уровня α: $\begin{cases} P(X \leq q_\alpha) \geq \alpha \\ P(X \geq q_\alpha) \geq 1 - \alpha \end{cases}$

Медиана – квантиль уровня $\frac{1}{2}$.

Определение квантиля в непрерывном случае: $F(q_\alpha) = \alpha$.

Свойство: $|E(X) - median(X)| \leq \sqrt{D(X)}$.

Ковариация случайных величин X, Y: $cov(X,Y) = E\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right) = E(XY) - E(X)E(Y)$.

Свойства ковариации: билинейность, $cov(X,X) = D(X)$, X и $Y -$ независимы $\Longrightarrow cov(X,Y) = 0$.

Коэффициент корреляции случайных величин X, Y: $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{D(X)D(Y)}}$

Свойства: $|\rho(X,Y)| \leq 1, if\ |\rho(X,Y)| = 1 \Leftrightarrow X = aY + b$.

Ковариационная матрица случайного вектора X: $\sum = E((X - E(X))(X - E(X))^T) = matrix\ \left(cov(X_i, X_j)\right)_{i,j=\{1,..,n\}}$

Свойства ковариационной матрицы: симметричность матрицы ($\Sigma = \Sigma^T$), $\Sigma \succcurlyeq 0$ (неотрицательно определена), $cov(AX + b) = AXA^T$.

Важные вероятностные неравенства.

Неравенство Маркова: $(X \geq 0)\quad P(X \geq \delta) \leq \frac{E(X)}{\delta}$

Неравенство Чебышева: $P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$

Минимизация матожиданием и медианой.

Что минимизирует математическое ожидание? (При условии, что оно существует.)

$$E(X) = arg\min_{a \in \mathbb{R}} E(X - a)^2$$

Что минимизирует медиана? Она минимизирует абсолютное отклонение

$$medX = arg\min_{a \in \mathbb{R}} E|X - a|$$

Многомерные распределения.

Полиномиальное (мультиномиальное) распределение. Это обобщение биномиального распределения. У него два параметра: $n -$ независимых испытаний, $m -$ исходов (в биномиальном распределении 2 исхода в каждом испытании). Для каждого из исходов у нас есть вероятность: вектор $p = [p_1 \quad \dots \quad p_m]^T$, $\sum p_i = 1$.

Есть вектор случайных величин $X = \begin{bmatrix} X^{(1)} \\ \dots \\ X^{(m)} \end{bmatrix}$, $X^{(i)} -$ число результатов типа i в n независимых испытаниях.

Тогда можем расписать вероятность $P\left(X^{(1)} = n_1, \dots, X^{(m)} = n_m\right) = \frac{n!}{n_1! \cdot \ldots \cdot n_m!} p_1^{n_1} \cdot \ldots \cdot p_m^{n_m}$, где $\sum_{k=1}^{m} n_k = n$.

Стандартное $n -$ мерное нормальное распределение.

Вектор $[X_1 \quad \dots \quad X_n]^T$, плотность: $p_0(x) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}x^T x\right)$.

Каждое из $X_i \sim \mathcal{N}(0,1), X_i$ – независимы.

Многомерное нормальное распределение. Зададим как линейное преобразование над стандартным $n -$ мерным нормальным распределением.

$$Y = AX + b, \qquad X \sim \mathcal{N}(0, I_n = E_n = diagonal\_matrix(n)) \text{ (стандартное многомерное нормальное распределение)}$$

$$\Longrightarrow Y \sim \mathcal{N}(b, AA^T)$$

если ковариационная матрица $\sum \succ 0$, то тогда можем написать плотность $p(x) = \frac{1}{(\sqrt{2\pi})^n \sqrt{det\sum}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \cdot \sum^{-1} \cdot (x-\mu)\right)$, где $\mu -$ вектор мат ожиданий.

Предельные теоремы.

Центральная предельная теорема: есть последовательность независимых, одинаково распределенных случайных величин $X_1, \dots, X_n, \dots - i.i.d.; E(X_1) = \mu, D(X_1) = \sigma^2$, тогда $\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma^2}} \underset{n\to\infty}{\overset{d}{\underset{\lim}{\to}}}$ (сходится по распределению) к случайной величине $Y \sim \mathcal{N}(0,1)$.

То есть $P\left(t_1 \le \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma^2}} \le t_2\right) \to \Phi(t_2) - \Phi(t_1) = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}} dt$ ($\Phi(t) -$ функция распределения стандартного нормального распределения).

Закон больших чисел: есть набор независимых, одинаково распределенных случайных величин $X_1, \dots, X_n, \dots - i.i.d.; E(X_1) = \mu$ тогда $P\left(\left|\frac{\sum x_k}{n} - \mu\right| > \varepsilon\right) \to_{n\to\infty} 0$.

Что то же самое, что $P\left(\left|\frac{\sum x_k}{n}\right| \le \varepsilon\right) \to_{n\to\infty} 1$.

Многомерная центральная предельная теорема: есть последовательность независимых, одинаково распределенных векторов случайных величин $X_1, \dots, X_n, \dots - i.i.d.; E(X_1) = \mu$ (вектор мат ожиданий), $D(X_1) = \sum \succcurlyeq 0$ (ковариационная матрица, неотрицательно определенная), тогда $\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma^2}} \underset{n\to\infty}{\overset{d}{\underset{\lim}{\to}}}$ (сходится по распределению) к многомерному стандартному нормальному распределению $\mathcal{N}(0, \sum)$.

1. Выборка. Эмпирическая функция распределения и её свойства. Способы визуализации выборки.

Каким свойством должна обладать выборка для того, чтобы мы могли сделать значимые выводы? Выборка должна происходить из некоторое вероятностного распределения, выборка должна быть достаточного размера (репрезентативность). Так как при анализе выборки мы будем часто применять предельные теоремы, а их применять можно только в случае достаточно большой выборки.

Дана выборка объема $n: X_1, \dots, X_n$. Данная выборка – это реализация n $i.i.d.$ случайных величин с теоритической функцией распределения $F(x)$.

Кратко, можно написать так: $X_1, \dots, X_n \sim F(x)$ или, если мы знаем распределение: $X_1, \dots, X_n \sim Exp(1)$.

$\mu_n(x) = \sum_{K=1}^{n} \mathbb{I}(X_K \le x)$ –функция, считающая число элементов выборки, которые не превосходят числа x.

Тогда эмпирическая функция распределения – это $F_n(x) = \frac{\mu_n(x)}{n}$ (монотонна по x, принимает значения $[0,1]$). Она задает эмпирическое распределение.

Вариационный ряд – отсортированная по возрастанию выборка: $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$.

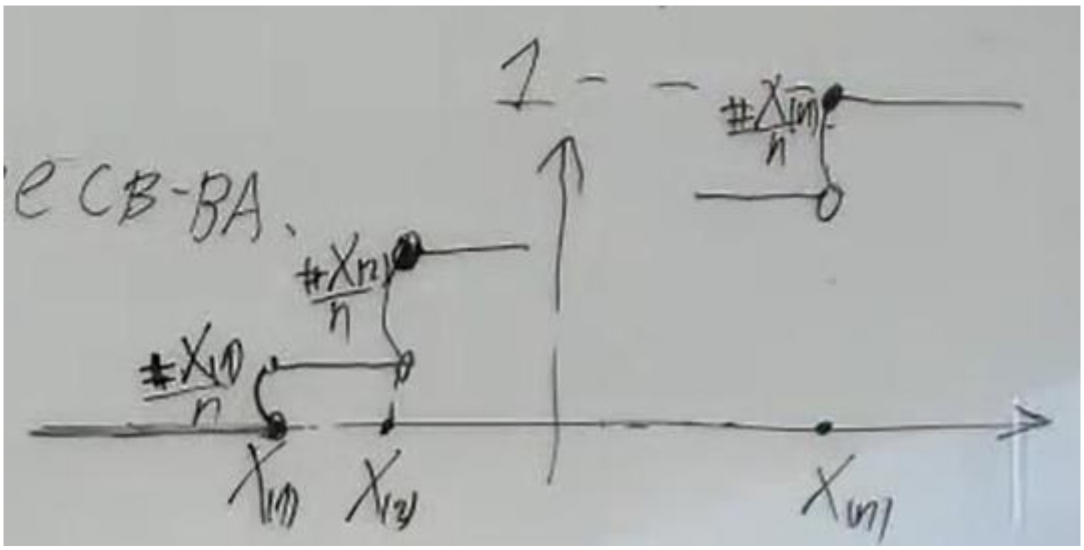


Рисунок 1. Эмпирическая функция распределения в общем виде.