СЕМ - Статистика и емпирични методи

Лекция 1. (03.10.2019 г.)

СЕМ

↓

статистика ⟶ Теория на вероятностите (фундамент за статистиката)

Случаен експеримент. Събития. Действия със събития

1. Случаен експеримент

Експеримент ↗ Условия на провеждане
↘ Резултат

Видове експерименти ↗ Детерминирани
↘ Случайни

При детерминираните експерименти резултатът се определя еднозначно. При случайните експерименти резултатът <u>не се</u> определя еднозначно. Случайните експерименти се срещат по-често в практиката в сравнение с детерминираните.

Основно пространство - Ω - множеството на всички възможни изходи на случайния експеримент (СЕ).

Пример: Хвърляне на зар: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Елементарна теория на вероятностите (ЕТВ)

Обща теория на вероятностите (ОТВ)

ЕТВ - Ω е краен или изброимо безкраен

ОТВ - Ω е неизброимо безкраен

Изброима безкрайност: $\mathbb{N}$: $1, 2, \dots, n, n+1, \dots$; възможно е някакво индексиране с естествени числа на последователните елементи

Неизброима безкрайност: $\mathbb{R}$: $(0, 1)$ има са твърде много (неизброим брой) елементи

$n \to \frac{1}{n}$

Събитие – част от изходите на СЕ

$A, B, C, \ldots \subset \Omega$

k_A – # изходи, при които настъпва A

Приемаме, че $k_\Omega < \infty$ (ЕТВ).

$$0 \le k_A \le k_\Omega$$

Ако $k_A = 0$ – невъзможно събитие – $\varnothing$

Ако $k_A = 1$ – елементарно събитие

Ако $k_A = k_\Omega$ – сигурно събитие

в общия случай: $A = \Omega$

Нека $A = \{1, 3, 5\}$, $B = \{5, 6\}$.

Действия със събития

1. Обединение – $A \cup B$ – събитие, съдържащо ел. събития или на A, или на B; $A \cup B = \{1, 3, 5, 6\}$

2. Сечение – $A \cap B$ – събитие, съдържащо ел. събития и на A, и на B; $A \cap B = \{5\}$

Ако $A = \{1, 3, 5\}$, $B = \{4, 6\}$, то $A \cap B = \varnothing$

Настъпва, ако A и B настъпват едновременно.

$A \cap B = \varnothing$ – несъвместими събития (две събития са несъвместими, ако не могат да настъпят едновременно)

3. Допълнение – $\overline{A}$ – събитие, съдържащо ел. събития, <u>не</u> принадлежащи на A

Настъпва, ако A не настъпва.

$$\boxed{\begin{aligned} A \cup \overline{A} &= \Omega \\ A \cap \overline{A} &= \varnothing \end{aligned}}$$

$n = 2$

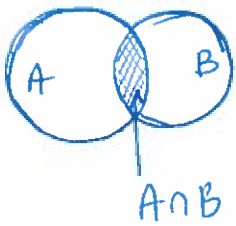
Пълна група несъвместими събития

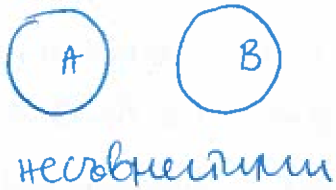
$A_1, A_2, \dots, A_n$ – ПГНС, ако едновременно са изпълнени условията:

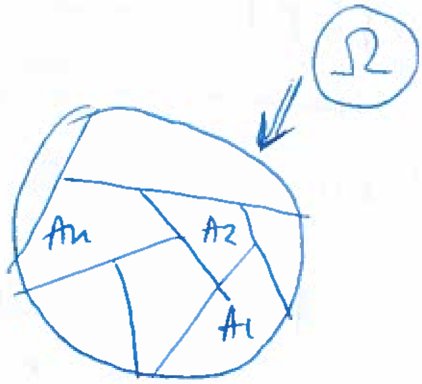
1) $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$

2) $A_i \cap A_j = \emptyset$, $i \neq j$

Диаграми на Вен







Вероятност

Хвърляне на монета $n = 100$

2^{-100} е вероятността и 100-те пъти да се падне ези

$45 < \#\text{ЕЗИ} < 55$ е много вероятен случай

1. Статистическа вероятност

n – #повторения на СЕ

A – събитие

k – абсолютна честота на настъпване на A

$\frac{k}{n}$ – относителна честота на настъпване на A

$$\boxed{P(A) = \lim_{n \to \infty} \frac{k}{n}}$$

$n = 20000 \quad \frac{k}{n} \approx 0{,}503$

10.10.2019г.

Свойства на вероятността.

1) $0 \leq P(A) \leq 1$

2) $P(\emptyset) = 0$; $P(\Omega) = 1$;

2. Класическа вероятност - при СЕ с краен брой равновероятни изходи

$$k_\Omega < \infty \ ; \quad P(\omega) = \frac{1}{k_\Omega} = const$$

$$\boxed{P(A) = \frac{k_A}{k_\Omega}}$$

3. Общо определение на вероятност при ЕТВ

$$P : \begin{cases} \Omega \longrightarrow (0,1) \\ \omega \longrightarrow P(\omega) \end{cases}$$

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Независими събития

1. Условна вероятност

Пример: СЕ - "Хвърляне на 2 монети"

$$\Omega = \{EE, ET, TE, TT\}$$

$$A = \{EE\}, \quad B = \{EE, ET\}$$

$$P(A) = \frac{1}{4}$$

$\underbrace{P(A|B)}_{\text{условна}}$ - вероятност на A, ако е настъпило B

условна вероятност на A, ако е настъпило B

$\cup$ $\cap$ $-$ $|$

или и не ако

$$P(A|B) = \frac{1}{2}$$

$$P(A|B) = \frac{k_{A\cap B}}{k_B} = \frac{k_{A\cap B}/k_\Omega}{k_B/k_\Omega} = \frac{P(A\cap B)}{P(B)}$$

2. Независими събития

B – като в т. 1 $\quad A = \{EE, TE\}$

$$P(A) = \frac{1}{2}\ ; \quad P(A|B) = \frac{1}{2}$$

Деф. Независимо събитие е такова събитие, при което $P(A) = P(A|B)$. (настъпването на събитието B не влияе върху настъпването на събитието A).

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

умножаваме на кръст и получаваме:

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

оттук следва, че независимите събития са винаги съвместими!!!

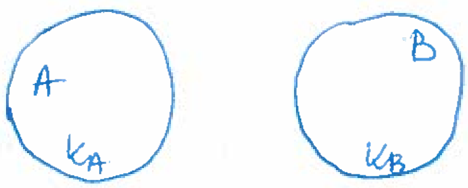
3. Формули за събиране и умножаване на вероятности

а) събиране

A, B – събития $\quad P(A \cup B) = ?$

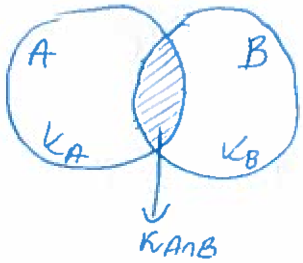
Случай $A \cap B = \emptyset$

$$P(A \cup B) = \frac{k_{A \cup B}}{k_\Omega} = \frac{k_A + k_B}{k_\Omega}$$



$$P(A \cup B) = \frac{k_{A \cup B}}{k_\Omega} = \frac{k_A + k_B}{k_\Omega} = P(A) + P(B)$$

Общ случай: $A \cap B \neq \emptyset$



$$P(A \cup B) = \frac{K_{A\cup B}}{K_\Omega} = \frac{K_A + K_B - K_{A\cap B}}{K_\Omega}$$

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Ако A и B са независими, тогава имаме

$$P(A \cup B) = P(A) + P(B) - P(A).P(B)$$

Пример: A и B независими $\quad P(A) = 0{,}7 \;;\; P(B) = 0{,}5$

$$P(A \cup B) = 0{,}7 + 0{,}5 - 0{,}7 . 0{,}5 = 0{,}85$$

б) умножаване

$$P(A \cap B) = ?$$

$$P(A|B) = P(A \cap B) / P(B), \text{ т.е. } P(A|B) = \frac{P(A\cap B)}{P(B)}$$

$$P(A \cap B) = P(A|B) . P(B)$$

Когато A и B са независими, то $P(A|B) = P(A) \Rightarrow$

$$\Rightarrow P(A \cap B) = P(A) . P(B)$$

4. Формула за пълната вероятност

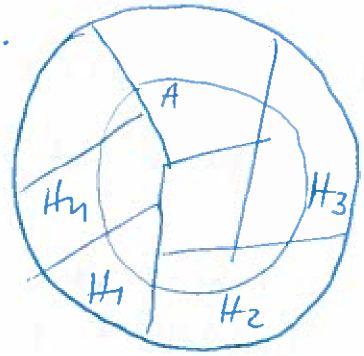
$H_1, H_2, \ldots, H_n$ – ПГНС

A – събитие

Дадени са $P(H_1), P(H_2), \dots, P(H_n)$

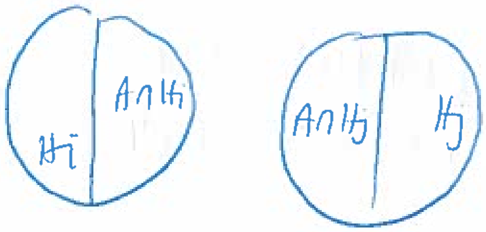
$P(A|H_1), P(A|H_2), \dots, P(A|H_n)$.

Търси се $P(A) = ?$



$$A = (A\cap H_1)\cup(A\cap H_2)\cup\dots\cup(A\cap H_n)$$

Тогава:



$$\boxed{P(A) = \sum_{i=1}^{n} P(A\cap H_i) = \sum_{i=1}^{n} P(A|H_i)P(H_i)}$$

Пример:

ФМИ → момичета – 55%
ФМИ → момчета – 45%

Среден успех момичета по МА → 4^{80}

Среден успех момчета по МА → 4^{20}

Колко е средният успех на студентите по МА?

⚠ примерът некоректен

$H_1 = \{\text{момиче}\}$ $\quad P(A|H_1) = 4^{.80}$

$H_2 = \{\text{момче}\}$ $\quad P(A|H_2) = 4^{.20}$

$A = \{\text{оценка по МА}\}$ $\quad 4{,}8 \cdot 0{,}55 + 4{,}2 \cdot 0{,}45$.

17.10.2019

Пример (преинтерпретиран) ⊛:

Поток → 55% момичета
Поток → 45% момчета

60% от момичетата имат успех $> 5^{00}$

40% от момчетата имат успех $> 5^{00}$

? % от всички студенти имат успех > 5⁰⁰

$H_1 = \{$момичета$\}$ $\quad P(H_1) = 0,55$

$H_2 = \{$момчета$\}$ $\quad P(H_2) = 0,45$

$A = \{$успехът на студент от потока е $> 5^{00}\}$

$P(A|H_1) = 0,6$

$P(A|H_2) = 0,4$

Тогава $P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) =$

$= 0,6 \cdot 0,55 + 0,4 \cdot 0,45 =$

$= 0,33 + 0,18 = 0,51.$

5. Формула на Бейс

$H_1, H_2, \dots, H_n$ - ПГНС

A - събитие

Дадено:

$P(H_i), P(A|H_i), i = \overline{1,n}$

Търси се: $P(H_i|A)$

$$\boxed{P(H_i|A) = \frac{P(A \cap H_i)}{P(A)} = \frac{P(A|H_i) \cdot P(H_i)}{\sum\limits_{i=1}^{m} P(A|H_i) \cdot P(H_i)}}$$

$P(H_i)$ - априорни вероятности на H_i (вероятн. до опита)

$P(H_i|A)$ - апостериорни вероятности (вероятн. след опита)

Към примера (*):

$$P(H_1|A) = \frac{P(A|H_1) \cdot P(H_1)}{P(A|H_1)P(H_1) + P(A|H_2) \cdot P(H_2)} = \frac{0,55 \cdot 0,6}{0,51} \approx 0,65$$

Основни вероятностни схеми

1. Схема на Я. Бернули - провеждат се $\underline{n}$ независими прости опити с една и съща вероятност за успех $\underline{p}$

прости опити → $\underline{\text{успех}}$ (p), $\underline{\text{неуспех}}$ → $1-p=q$

$k = \#$ успехи $\quad k = 0, 1, 2, \dots, n$

$A_k = \{$настъпват $\underline{k}$ успеха$\} \quad P(A_k) = ?$

$k=0 \quad A_0 = \underbrace{HH\dots H}_{n} \quad P(A_0) = \underbrace{q\cdot q \cdots q}_{n} = q^n = 1.1.q^{n-0}$

$k=1 \quad A_1 = yHH\dots H \cup HyH\dots H \cup HHy\dots H \cup \dots \cup HHH\dots Hy$

$P(A_1) = pq^{n-1} + pq^{n-1} + \dots + pq^{n-1} = npq^{n-1} = C_n^1 p^1 q^{n-1}$

$k=2 \quad A_2 = yyHH\dots H \cup yHyH\dots H \cup \dots \quad P(A_2) = C_n^2 p^2 q^{n-2}$

$\vdots$

Обща формула: $\boxed{P(A_k) = C_n^k \cdot p^k \cdot q^{n-k}}$

$C_n^k = \dfrac{n(n-1)\dots(n-k+1)}{k!}; \quad C_n^2 = \dfrac{n(n-1)}{2}$

Свойство: $C_n^k = C_n^{n-k}$ - симетрия

2. „Схема с топки"

урна $\longrightarrow$ N-топки: M - бели, N-M - черни

Изваждат се без връщане $\underline{n}$ топки. Нека $\underline{m}$ на брой от топките в извадката са бели. Тогава $\underline{n-m}$ - # черни топки в извадката.

$A_m = \{$в извадката има m ~~т~~ бели топки$\}$

$$P(A_m) = \frac{k_{A_m}}{k_\Omega}$$

Пример: Тото 2 (6 от 49)

$N = 49$ $\quad n = 6$

$M = 6$ $\quad m = 6$

(в най-добрия сл

$$k_\Omega = C_N^n$$

<u>Правило за броене:</u> Сложен експеримент $E \begin{matrix} \nearrow E_1 \\ \searrow E_2 \end{matrix} \Big\}$ етапи на E

E_1 има k_1 възможни изходи

Независимо от E_1, E_2 има k_2 възможни изхода.

Тогава E има $k_1 k_2$ - изхода.

С помощта на това правило можем да определим, че $k_{A_m} = C_M^m \cdot C_{N-M}^{n-m}$.

Следователно $\boxed{P(A_m) = \frac{k_{A_m}}{k_\Omega} = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}}$

Решение на примера „Тото 2":

$$P(A_6) = \frac{C_6^6 \cdot C_{43}^0}{C_{49}^6} = \frac{1 \cdot 1}{C_{49}^6} = \frac{1}{13983816}$$

Ако схемата е с връщане, то тя е схема на Бернули с брой опити n, вероятност за успех $p = \frac{M}{N}$.

Сх. на Бернули - <u>n опита</u> (1), <u>независими</u> (2) с една и съща вероятност за успех (3)

p_1 - вероятност първата топка да е бяла

p_2 - вероятност втората топка да е бяла

$$p_1 = \frac{M}{N} \qquad p_2 = \begin{cases} \frac{M}{N-1}, & \text{ако първата топка е черна} \\ \frac{M-1}{N-1}, & \text{ако първата топка е бяла} \end{cases}$$

$p_1 \neq p_2 \Rightarrow$ усл. (3) е нарушено, както и усл. (2)

Пример: Нека $n=10$, $M=400$, $N=1000$.

Тогава $p_1 = \frac{400}{1000} = 0,4$, $p_2 = \begin{cases} \frac{400}{999}, \text{ ако} \dots \\ \frac{399}{999}, \text{ ако} \dots \end{cases} \Rightarrow p_1 \approx p_2$

Ако $n \ll M, N-M$, схемата с топки може да се замени със схемата на Бернули (апроксимация).

3. Схема „до първи успех" – правят се независими прости опити с една и съща вероятност за успех p до настъпване на успех, след което опитите се преустановяват (прекратяват).

$A_k = \{$правят се k опита$\}$, $k = 1, 2, 3, 4, 5, \dots$

$P(A_k) = ?$

$k=1 \quad A_1 = У \quad P(A_1) = p$

$k=2 \quad A_2 = НУ \quad P(A_2) = pq$

$k=3 \quad A_3 = ННУ \quad P(A_3) = pq^2$

$\vdots$

Следователно в общия случай имаме $\boxed{P(A_k) = pq^{k-1}}$

4. Схема на Поасон – схема на Бернули при $n \to \infty$

Без допълнителни условия $P(A_k) \to 0 \quad \forall k$

$$0 < p = \lim_{n\to\infty} \frac{k}{n} \Rightarrow k \to \infty$$

Условия на Поасон: $n \to \infty$, $p \to 0$, така че $np = \lambda = const$

При тези условия $\boxed{P(A_k) = \frac{\lambda^k}{k!} e^{-\lambda}}$

От схемата на Бернули е известно, че $P(A_k) = C_n^k p^k q^{n-k}$

$$P(A_k) = C_n^k p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \cdot p^k \cdot q^{n-k} =$$

$$= \frac{np(np-p)\dots(np-kp+p)}{k!} \cdot q^{n-k} =$$

$$= \frac{\lambda(\lambda-\overset{\overset{0}{\uparrow}}{p})\dots(\lambda-kp+\overset{\overset{0}{\uparrow}}{p})}{k!} \cdot q^{n-k} =$$

$$= \frac{\lambda^k}{k!} \cdot q^{n-k}$$

$$q^{n-k} = (1-p)^{n-k}$$

Знаем, че $(1+x)^{\frac{1}{x}} \xrightarrow[x \to 0]{} e$.

Тогава $(1-p)^{n-k} \underset{x=-p}{=} (1+x)^{-\frac{\lambda}{x}-k} \to e^{-\lambda}$

$np = \lambda$

$-nx = \lambda \Rightarrow n = -\frac{\lambda}{x}$

С това формулата на Поасон е изведена.

<u>Пример:</u> Vivacom $\to$ 1% грешки

Реализират се 300 разговора месечно от даден клиент.

$P = ?$ за 3 грешки

Схема на Бернули: $n = 300$, $k = 3$

$p = 0{,}01$

$$P(A_3) = C_{300}^3 \cdot 0{,}01^3 \cdot 0{,}99^{297}$$

Схема на Поасон: $\lambda = np = 3$. Приемаме, че $e = 2{,}71$.

$$P(A_3) = \frac{3^3}{6} \cdot e^{-3} = \frac{9}{2 \cdot e^3} = \frac{9}{39{,}8} = \frac{90}{398} \approx 0{,}23$$

Дискретни случайни величини

1. Случайна величина

$$\xi : \begin{cases} \Omega \to \mathbb{R} \\ \omega \to \underset{\text{число}}{\xi(\omega)} \end{cases}$$

2. Видове случайни величини (сл.в.) $\nearrow$ дискретни, $\searrow$ непрекъснати

- дискретни - изолирани точки от числовата права

- непрекъснати - плътно запълване на интервал от числовата права

- По своето естество дискретните сл. в. са по-прости от непрекъснатите.

3. Закон на разпределение - съответствие между стойности на сл.в. и вероятностите на тези стойности. Това съответствие лесно може да се зададе в табличен вид:

ξ	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

$$\boxed{p_n = P(\xi = x_n)}$$

$$0 < p_n < 1, \quad n = 1, 2, \dots$$

$$\sum_n p_n = 1$$

4. Числови характеристики, моменти

а) математическо очакване

$$E\xi = \sum_n x_n p_n$$

* грубо казано, това е средната стойност на сл.в.

* пример: алкохолното съдържание в дадена алк. напитка (примерно в 100 бутилки)

Свойства на E_ξ:

1) Ако ξ – сл. в. и c – число, то $E(c\xi) = cE_\xi$

$$E(c\xi) = \sum_n c x_n p_n = c\sum_n x_n p_n = cE_\xi$$

2) Ако $\xi = c = const$, то $E_\xi = c$

ξ	c
p	1

3) Ако ξ, η – сл. в., то $E(\xi+\eta) = E_\xi + E_\eta$

б) дисперсия (на бвм. разсейване)

$$D_\xi = \sum_n (x_n - E_\xi)^2 p_n > 0$$

31.10.2019г.

Припомняме формулата за дисперсия:

$$D_\xi = \sum_n (x_n - E_\xi)^2 p_n$$

Ще формализираме следната формула:

$$D_\xi = E(\xi^2) - (E_\xi)^2$$

Знаем, че $D_\xi = \sum_n (x_n - E_\xi)^2 p_n =$

$$= \sum_n \left(x_n^2 - 2E_\xi \cdot x_n + (E_\xi)^2\right) p_n =$$

$$= \sum_n x_n^2 p_n - 2E_\xi \sum_n x_n p_n + (E_\xi)^2 \sum_n p_n =$$

$$= E(\xi^2) - 2(E_\xi)^2 + (E_\xi)^2 \cdot 1 =$$

$$= E(\xi^2) - (E_\xi)^2$$

Свойства на $D\xi$:

1) А ξ - сл. в. и c - число, то $D(c\xi) = c^2 D\xi$

$$D(c\xi) = E(c^2\xi^2) - (E(c\xi))^2 =$$

$$= c^2 E(\xi^2) - c^2 (E\xi)^2 =$$

$$= c^2 \left[E(\xi^2) - (E\xi)^2\right] = \boxed{c^2 D\xi}$$

2) Ако $\xi = c = const$, то $D\xi = 0$

3) Ако ξ, η - сл. в., то $D(\xi+\eta) = D\xi + D\eta + 2(E(\xi\eta) - E\xi E\eta)$

$$D(\xi+\eta) = E(\xi+\eta)^2 - (E(\xi+\eta))^2 =$$

$$= E(\xi^2+\eta^2+2\xi\eta) - (E\xi + E\eta)^2 =$$

$$= \underbrace{E(\xi^2)}_{1} + \underbrace{E(\eta^2)}_{2} + \underbrace{2E(\xi\eta)}_{3} \underbrace{- (E\xi)^2}_{1} \underbrace{- (E\eta)^2}_{2} \underbrace{- 2E\xi . E\eta}_{3} =$$

$$= D\xi + D\eta + 2(E(\xi\eta) - E\xi E\eta)$$

Ковариация на ξ и η: $cov(\xi, \eta) = E(\xi\eta) - E\xi . E\eta$

Следователно $D(\xi+\eta) = D\xi + D\eta + 2cov(\xi, \eta)$

Деф. Случайните величини ξ и η се наричат некорелирани точно тогава, когато $cov(\xi, \eta) = 0$.

$cov(\xi, \xi) = D\xi > 0$ - свойство 1 на ковариацията

$cov(c\xi, \eta) = c\, cov(\xi, \eta)$ - свойство 2 на ковариацията

Деф. Сл. величини ξ и η се наричат __независими__, [1] ако събитията, заключаващи се в това те да приемат кои да е две стойности, са независими.

Илюстрация:

ξ	x_1	x_2	....	x_n	....
p	p_1	p_2	....	p_n	....

η	y_1	y_2	...	y_m	...
p	q_1	q_2	...	q_m	....

$\{\xi = x_n\}, \{\eta = y_m\}$ - незав. за $\forall n, m$, т.е. $P\{\xi = x_n, \eta = y_m\} = p_n q_m$.

Теорема Ако случайните величини ξ и η са независими, то те са некорелирани, т.е. $\operatorname{cov}(\xi, \eta) = 0$.

__Доказателство:__ $E(\xi\eta) = \sum_n \sum_m x_n y_m P(\xi = x_n, \eta = y_m)$

$$\Rightarrow E(\xi\eta) = \sum_n \sum_m x_n y_m p_n q_m =$$

$$= \sum_n x_n p_n \sum_m y_m q_m = E\xi \cdot E\eta$$

Тогава $\operatorname{cov}(\xi, \eta) = E(\xi\eta) - E\xi E\eta = E\xi E\eta - E\xi E\eta = 0$

Моменти от ред k: $m_k = E(\xi^k)$

Централни моменти от ред k: $\mathring{m}_k = E(\xi - E\xi)^k$

$$\boxed{\begin{aligned} m_1 &= E\xi \\ \mathring{m}_2 &= D\xi \end{aligned}}$$

$$m_k = E(\xi^k) = \sum_n x_n^k p_n$$

$$\mathring{m}_k = E(\xi - E\xi)^k = \sum_n (x_n - m_1)^k p_n$$

к-четни – число или $+\infty$

к-нечетни – не съществува, число, $+\infty$ или $-\infty$

5. Пораждащи функции

сл. в. $\to \xi$ $\qquad \varphi_\xi(t) = E(t^\xi),\ t > 0$

$$\varphi_\xi(t) = E(t^\xi) = \sum_n t^{x_n} . p_n$$

$$\varphi_\xi'(t) = \sum_n x_n . t^{x_n - 1} . p_n \ ; \ \varphi_\xi'(1) = E\xi = m_1$$

$$\varphi_\xi''(t) = \sum_n x_n(x_n - 1) . t^{x_n - 2} . p_n \ ; \ \varphi_\xi''(1) = \sum_n x_n^2 p_n - \sum_n x_n p_n =$$

$$= m_2 - m_1$$

Основни дискретни разпределения $\qquad$ 7.11.19

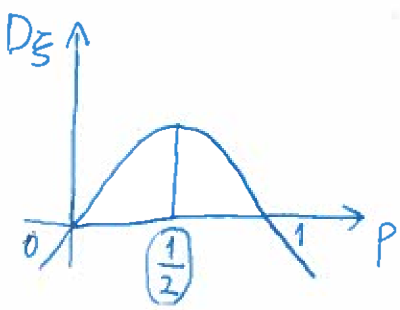
1. Разпределение на Бернули

ξ	0	1
p	$1-p=q$	p

$$E\xi = 0.q + 1.p = p$$

$$D\xi = E(\xi^2) - (E\xi)^2 = p - p^2 =$$

$$= p(1-p) = pq$$



$$\varphi_\xi(t) = E(t^\xi) = 1.q + t.p = tp + q$$

2. Биномно разпределение – $\xi_1, \xi_2, \dots, \xi_n$ – независими величини с разпр. на Бернули с вероятност за успех p.

$$S_n = \xi_1 + \xi_2 + \dots + \xi_n \sim Bi(n,p)$$

S_n	0	1	$\dots$	k	$\dots$	n
P	q^n	$n.p.q^{n-1}$	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	p^n

$$ES_n = E\xi_1 + E\xi_2 + \dots + E\xi_n = np$$

$$DS_n = D\xi_1 + D\xi_2 + \dots + D\xi_n = npq$$

(тъй като са независими величини)

$$\varphi_{S_n}(t) = E(t^{S_n}) = E\left(\prod_{i=1}^{n} t^{\xi_i}\right) = \prod_{i=1}^{n} E(t^{\xi_i}) = (q+tp)^n$$

(некорелираност)

3. Геометрично разпределение – $\xi_1, \xi_2, \dots, \xi_n, \dots$ – независими величини с разпр. на Бернули с вероятност за успех p.

$$\tau = \min(n : \xi_n = 1)$$

τ	1	2	$\dots$	n	$\dots$
P	p	pq	$\dots$	pq^{n-1}	$\dots$

$$E\tau = 1.p + 2pq + 3pq^2 + \dots npq^{n-1} + \dots =$$

$$= 1-q + 2(1-q)q + 3(1-q)q^2 + \dots =$$

$$= 1\underbrace{-q + 2q}\underbrace{-2q^2 + 3q^2}\underbrace{-3q^3 + \dots}$$

$$\Rightarrow E\tau = 1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q} = \frac{1}{p}$$

$D\tau$ - дисперсия на геом. разпр. (самостоятелно)

$$\varphi_\tau(t) = E(t^\tau) = \sum_{n=1}^{\infty} t^n p q^{n-1} = pt \sum_{n=0}^{\infty} (tq)^n = \frac{pt}{1-qt}$$

$$D\tau = E(\tau^2) - (E\tau)^2 = \ldots$$

4. Разпределение на Поасон

π	0	1	2	$\ldots$	n	$\ldots$
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\ldots$	$\frac{\lambda^n}{n!} e^{-\lambda}$	$\ldots$

$E\pi = ?$, $D\pi = ?$, $\varphi_\pi(t) = ?$

$$\begin{array}{ccc} \text{Схема на Бернули} & \xrightarrow[np \to \lambda]{n\to\infty,\ p \to 0} & \text{Схема на Поасон} \\ \downarrow & & \downarrow \\ Bi(n,p) & \xrightarrow[np = \lambda]{n\to\infty,\ p\to 0} & \text{Разпр. на Поасон} \end{array}$$

$ES_n \to E\pi$

$DS_n \to D\pi$

$\varphi_{S_n}(t) \to \varphi_\pi(t)$

$$E\pi = \lim_{n\to\infty} ES_n = \lim_{n\to\infty} np = \lambda$$

$$D\pi = \lim_{n\to\infty} DS_n = \lim_{n\to\infty} npq = \lambda \lim q = \lambda \lim (1-p) = \lambda$$

$$\varphi_\pi(t) = \lim_{n\to\infty} \varphi_{S_n}(t) = \lim_{n\to\infty} (q+pt)^n =$$

$$= \lim_{n\to\infty} (1 + p(t-1))^n = \lim_{\substack{n\to\infty \\ p\to 0}} (1+p(t-1))^{\frac{\lambda}{p}}$$

Известно е, че $(1+x)^{\frac{1}{x}} \xrightarrow[x\to 0]{} e$.

Полагаме $x = p(t-1) \Rightarrow p = \frac{x}{t-1}$

Тогава $\varphi_\pi(t) = \lim_{p \to 0} (1 + p(t-1))^{\frac{\lambda}{p}} = \lim_{px \to 0} (1+x)^{\frac{\lambda(t-1)}{x}} =$

$$= \boxed{e^{\lambda(t-1)}}$$

Гранични свойства на схемата на Бернули.
Закон за големите числа (ЗГЧ) на Я.
Бернули, теореми на Моавър-Лаплас

$n \to \infty$
p - фиксирано

1. Неравенство на Чебишев
$\xi \geq 0$ - случ. величина $\quad \varepsilon > 0$

$$\boxed{P(\xi > \varepsilon) < \frac{E\xi}{\varepsilon}}$$

$$E\xi = \sum_n x_n p_n = \sum_{n: x_n < \varepsilon} x_n p_n + \sum_{n: x_n \geq \varepsilon} x_n p_n \geq \sum_{n: x_n \geq \varepsilon} x_n p_n \geq$$

$$\geq \varepsilon \underbrace{\sum_{n: x_n \geq \varepsilon} p_n}_{= P(\xi \geq \varepsilon)}$$

Следствие: ξ - произв., $\varepsilon > 0$

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$$

Прилагаме нер. на Чебишев към $(\xi - E\xi)^2$ и числото ε^2:

$$P((\xi - E\xi)^2 > \varepsilon^2) \leq \frac{E(\xi - E\xi)^2}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2}$$

$$P(|\xi - E\xi| > \varepsilon) < P((\xi - E\xi)^2 > \varepsilon^2) \leq \frac{E(\xi - E\xi)^2}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2}$$

Използвахме, че $a > b \Leftrightarrow a^2 > b^2 \; (\forall a, b > 0)$

2. Закон за големите числа на Я. Бернули (ЗГЧ)

Схема на Бернули: $n = \#$опити

p – вер. за успех в 1 опит

k – $\#$успехи в n опита

ЗГЧ гласи: $\boxed{P\left(\left|\frac{k}{n} - p\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0}$

$k = S_n \sim Bi(n, p)$

$Ek = np$

$Dk = npq$

$$E\left(\frac{k}{n}\right) = \frac{1}{n} Ek = \frac{1}{n} \cdot np = p$$

$$D\left(\frac{k}{n}\right) = \frac{1}{n^2} Dk = \frac{1}{n^2} \cdot npq = \frac{pq}{n}$$

$$P(|\xi - E\xi| > \varepsilon) < \frac{D\xi}{\varepsilon^2}$$

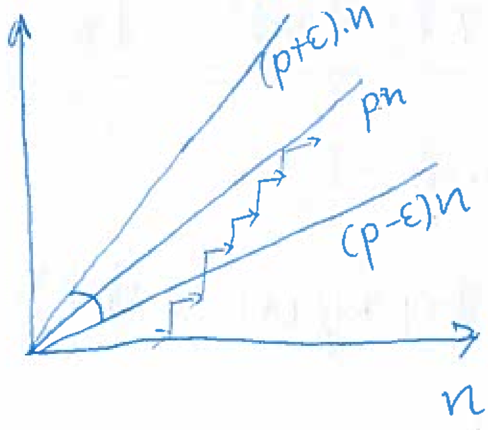
$$\xi = \frac{k}{n}$$

$$P\left(\left|\frac{k}{n} - p\right| > \varepsilon\right) < \frac{pq}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

Геометрична интерпретация

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \xrightarrow[n\to\infty]{} 1$$

$$P((p-\varepsilon)n < k < (p+\varepsilon)n) \xrightarrow[n\to\infty]{} 1$$



Вероятността да е вътре в ъгъла се увеличава. Обратно, вероятността за попадане извън ъгъла намалява неограничено.

Функцията е случайна стъпаловидна.

3. Теореми на Моавър-Лаплас

$$P(pn-\varepsilon n<k<pn+\varepsilon n)\xrightarrow[n\to\infty]{}1$$

Нека вместо $\varepsilon.n$ вземем $\varepsilon.\sqrt{n}$

$$P(pn-\varepsilon\sqrt{n}<k<p.n+\varepsilon\sqrt{n})\to ?$$

Нормиран брой успехи

ξ - сл. в. $\quad \underset{\substack{\downarrow\\ \text{число}}}{E\xi}, D\xi<\infty$

Центрирана случ. величина $\to \xi-E\xi\ (=0)$

Нормирана случ. величина: $\dfrac{\xi-E\xi}{\sqrt{D\xi}}$

$\xi=k$

$k-np$ е ц. сл. в.

$\dfrac{k-np}{\sqrt{npq}}$ е норм. сл. в.

$$\boxed{P\left(a < \frac{k - np}{\sqrt{npq}} < b\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx}$$

$$P(a\sqrt{pqn} + np < k < b\sqrt{npq} + np) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Избираме $a\sqrt{pq} = -\varepsilon$ и $b\sqrt{pq} = \varepsilon$.

Тогава:

$$P(pn - \varepsilon\sqrt{n} < k < pn + \varepsilon\sqrt{n}) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon}{\sqrt{pq}}}^{\frac{\varepsilon}{\sqrt{pq}}} e^{-\frac{x^2}{2}} dx$$

Пример: Хвърляне на монета: $n = 100$

$$P = ? \quad 45 < \#E < 55 \quad (P(45 < k < 55))$$

$$p = \frac{1}{2}$$

$$P = \sum_{k=46}^{54} C_{100}^k \cdot 2^{-100}$$

$$P\left(\frac{45-50}{5} < \frac{k-50}{5} < \frac{55-50}{5}\right)$$

$$P\left(-1 < \frac{k-50}{5} < 1\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{x^2}{2}} dx \approx 0,68$$

4. Сходимости по вероятност и разпределение. Централна гранична теорема

Деф. Редицата случайни величини $\xi_1, \xi_2, \dots, \xi_n, \dots$ е сходяща по вероятност към сл. в. ξ, ако $P(|\xi_n - \xi| > \varepsilon) \xrightarrow[n\to\infty]{} 0$, $\varepsilon > 0$. Обозначение: $\xi_n \xrightarrow{P} \xi$

$\xi_n = \frac{k}{n}$; $\xi = p$ $\quad\quad \frac{k}{n} \xrightarrow{P} p$

$\boxed{\text{Деф.}}$ Редицата сл. в. $\xi_1, \xi_2, \dots, \xi_n, \dots$ е сходяща по разпределение към сл. в. ξ, ако е изпълнено $P(\xi_n \leq x) \longrightarrow P(\xi \leq x)$ за всяко x, за което функцията $P(\xi \leq x)$ е непрекъсната.

Означение: $\xi_n \xrightarrow{d} \xi$

- p - probability, d - distribution

$$P\left(a < \frac{k - np}{\sqrt{npq}} < b\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

$\xi_n = \frac{k - np}{\sqrt{npq}}$; $b = x$; $a = -\infty$ ← полагания в Теор. на М.-Л.

$\xi = Z$ - стандартно нормално разпределена сл. в., а функцията $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ е нейна вероятностна плътност.

$$\boxed{\frac{k - np}{\sqrt{npq}} \xrightarrow{d} Z}$$

$\boxed{\text{Т}}$ Централна гранична теорема (ЦГТ)

(*) Нека $\xi_1, \xi_2, \dots, \xi_n, \dots$ - редица независими еднакво разпределени сл. величини, като $D\xi_1 < \infty$ и $S_n = \xi_1 + \xi_2 + \dots + \xi_n$.

Тогава $\frac{S_n - ES_n}{\sqrt{DS_n}} \xrightarrow{d} Z$.

Ако $\xi_1, \xi_2, \dots, \xi_n, \dots$ имат разпределение на Бернули с вероятност за успех p, то $S_n = k$, $ES_n = np$, $DS_n = npq$. С други думи, ЦГТ е обобщение на теоремата на Моавър-Лаплас. (от голяма важност !!! на теста ще има въпроси по това)

Случаен експеримент с неизброимо много изходи. Аксиоми на Колмогоров

Пример: Безбройно много (∞) хвърляния на монета

$$\Omega = \{\omega : \omega = \omega_1 \omega_2 \dots \omega_n \dots \quad \omega_n = \begin{Bmatrix} E \\ T \end{Bmatrix}\}.$$

i - изоморфизъм (взаимно еднозначно съответствие)

$$i : \begin{cases} \Omega \longrightarrow [0, 1] \\ \omega = \omega_1 \omega_2 \dots \longrightarrow 0, \alpha_1 \alpha_2 \dots \underset{\text{двоично}}{(2)} \end{cases} \qquad \alpha_n = \begin{cases} 0, & \omega_n = E \\ 1, & \omega_n = T \end{cases}$$

$A_1 = \{\omega : \omega_1 = E\}$ - събитие "първо хвърляне е E"

$P(A_1) = \frac{1}{2} \qquad i(A_1) = [0, \frac{1}{2})$, т.е. $P(A_1) = \frac{1}{2} = |i(A_1)|$

$A_{11} = \{\omega : \omega_1 = \omega_2 = E\}$ - "първите две хвърляния са E"

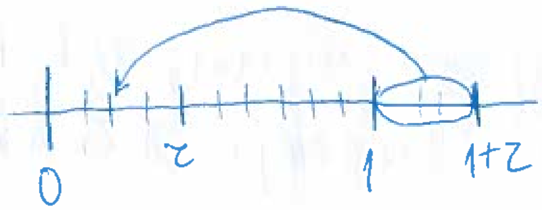
$P(A_{11}) = \frac{1}{4} \qquad i(A_{11}) = [0, \frac{1}{4})$, т.е. $P(A_{11}) = \frac{1}{4} = |i(A_{11})|$

21.11.2019г. $[0, 1]$, разликата е ~~х~~рационално число

$\varepsilon \begin{cases} > 0 \\ = 0 \end{cases}$ множеството от всички такива

M_0, $|M_0| = \varepsilon$

$M_r = r + M_0 \pmod 1$, r - рационално число



$\Rightarrow \varepsilon = |M_r| = r + M_0 \pmod 1$

Следователно не е възможно нито $\varepsilon = 0$, нито $\varepsilon > 0$.
Тогава получаваме, че множеството <u>няма дължина</u>.

<u>Ст. Банах (1924)</u>

$n \geq 3 \quad S_1^{(n)} \quad S_2^{(n)}$

$k = k(n)$

• Нека имаме две събития A и B.

A, B - събития
$A \cup B$ - събитие
$A \cap B$ - събитие
$\bar{A}$ - събитие

$A_1, A_2, \ldots, A_n, \ldots$ - събития

$\bigcup_n A_n$, $\bigcap_n A_n$ - събития

$\boxed{\sigma\text{-алгебра}}$
↓
всички събития са интервали или точки.

• Борелова σ-алгебра в $\mathbb{R}^1$ - минимална σ-алгебра, която съдържа интервали от вида $(a, b]$, $a < b$
$\{a\}, \{b\}$ - изолирани точки, (a, b) - отворени интервали
$[a, b]$ - затворени интервали, интервали от вида $[a, b)$
Възможно е $a = -\infty$, $b = \infty$.

Пример: $\{b\} = \bigcap_n (b - \frac{1}{n}, b]$

сечение на изброимо много интервали

Случайна величина: $\xi : \begin{cases} \Omega \to \mathbb{R}^1 \\ (\Omega, \mathcal{F}) \to (\mathbb{R}^1, \beta) \end{cases}$ $\xi(\omega)$

$\mathcal{F}$ – σ-алгебра от събитията в Ω

β – Борелова σ-алгебра в $\mathbb{R}^1$

$\boxed{\xi^{-1}(B) \in \mathcal{F}}$, $B \in \beta$

↓

Условие за измеримост

• Случайните величини трябва да удовлетворяват условието за измеримост.

• Разпределение на сл. в.

При дискр. сл. в.

ξ	x_1	x_2	...	x_n	...
P	p_1	p_2	...	p_n	...

При непрек. сл. в. → функция на разпределение

$$F(x) = P(\xi \le x) = P(\xi^{-1}((-\infty, x]))$$

Свойства на $F(x)$:

1) $0 \le F(x) \le 1$, $F(x)$ е монотонно ненамаляваща

2) $F(x)$ е непрекъсната отгоре и има граница отдолу за всяко x;

3) $\lim_{x \to \infty} F(x) = 1$;

4) $\lim_{x \to -\infty} F(x) = 0$.

Да разгледаме свойствата:

$0 \le F(x) \le 1$ - очевидно (събитие - вероятност)

$$P(\xi \le x) = F(x) \le F(y) = P(\xi \le y)$$

$$x < y$$

$$\{\xi \le x\} \subset \{\xi \le y\}$$

монотонно ненамаляваща

$$F(x) = P(\xi \le x) = P(\xi^{-1}((-\infty, x]))$$

$$\bigcap_{n=1}^{\infty} (-\infty, x_n] = \underbrace{(-\infty, x]}_{\text{лъч}} \qquad F(x_n) \to F(x),\ x_n \downarrow x$$

$$\bigcup_{n=1}^{\infty} (-\infty, x_n] = \underbrace{(-\infty, x)}_{\text{отворен лъч}} \qquad F(x_n) \to P(\xi^{-1}((-\infty, x))),\ x_n \uparrow x$$

В общия случай $P(\xi^{-1}((-\infty, x))) \ne F(x)$.

съвпадение само при 0

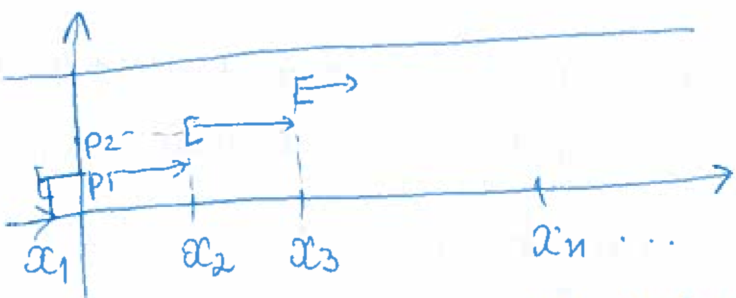
$x_n \uparrow \infty$

$\bigcup (-\infty, x_n] = \mathbb{R}^1$ цялата числова права $\to 1$

$x_n \downarrow -\infty \quad \to \emptyset$ невъзможно събитие $\to 0$.

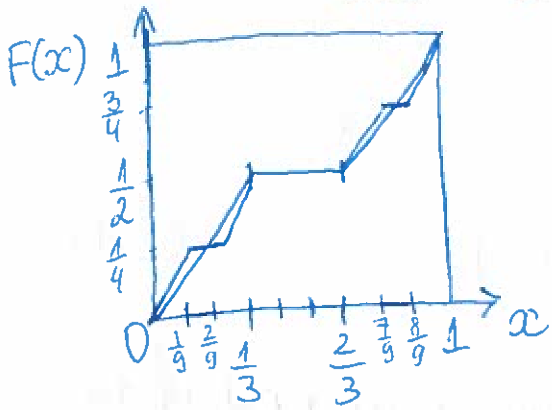
Видове функции на разпределение:

1) Дискретни ф.р.



2) Абсолютно непрекъснати ф.р. - ако $\exists f(x)$ (ф-я, която се нарича вероятностна плътност) такава, че $\boxed{F(x) = \int_{-\infty}^{x} f(x)dx}$

3) Сингулярни функции на разпределение стълба на Г. Кантор



$F_1(x)$

$F_2(x)$

$F_3(x)$

$$F(x) = \lim_n F_n(x)$$

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \ldots = \frac{1}{3}\left(1 + \frac{2}{3} + \frac{4}{9} + \ldots\right)$$

$$= \frac{1}{3} \cdot \frac{1}{1-\frac{2}{3}} = \frac{1}{3} \cdot 3 = \boxed{1}$$

$$0 < F(x) \neq \int_0^x F'(x)dx = 0$$

$0, \alpha_1 \alpha_2 \ldots (3) \longrightarrow 0, \frac{\alpha_1}{2} \frac{\alpha_2}{2} \ldots (2)$

тричен вид — двоичен вид

28.11.2019г.

функция на разпределение на сл. в. ξ: $F(x) = P(\xi \leq x)$

$$F(x) = P(\xi \leq x) = \int_{-\infty}^{x} f(x)dx$$ – абсолютно непрекъсната ф-я на разпределение

$$P(\xi \in (a, b]) = F(b) - F(a)$$

Случай на абс. непр. $F(x)$

$$0 = P(\xi = x) \leq P(x - \varepsilon < \xi \leq x + \varepsilon) = \int_{x-\varepsilon}^{x+\varepsilon} f(x)dx \longrightarrow 0$$

интервалът става все по-малък и по-малък

Вероятностен смисъл на $f(x)$ (вер. плътност):

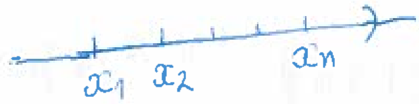
$$dF(x) = F'(x)dx = f(x)dx$$

$$F(x+dx) - F(x) = dF(x) = F'(x)dx = f(x)dx$$

$$P(\xi \in (x, x+dx)) = F(x+dx) - F(x) = dF(x) = F'(x)dx = f(x)dx$$

• Математическо очакване на абсолютно непрекъсната случайна величина:

ξ - дискр. сл. в. $\quad E\xi = \sum_n x_n p_n$



ξ - абс. непр. сл. в. $\quad E\xi = \int_{-\infty}^{+\infty} x f(x)dx$

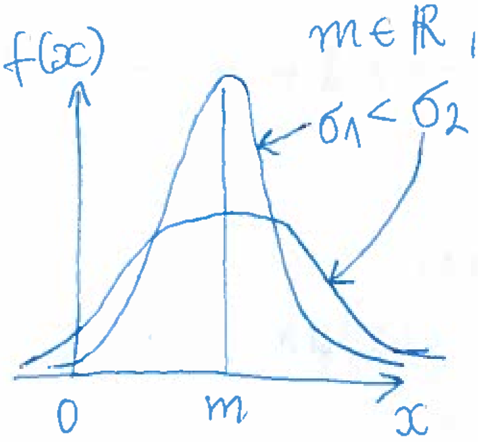
$$D\xi = E(\xi - E\xi)^2 = \int_{-\infty}^{+\infty} (x - E\xi)^2 f(x)dx$$

Основни абсолютно непрекъснати разпределения

1). Нормално разпределение $\quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$

$m, \sigma > 0$ - параметри

$m \in \mathbb{R},\ \sigma > 0,\ -\infty < x < \infty$



Ако $m=0$, $\sigma=1$, говорим за стандартно нормално разпределение

Z означен на ст. н.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$P(a < Z < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

$$P\left(a < \frac{x - np}{\sqrt{npq}} < b\right) \to P(a < Z < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Нормално разпред. означаване с $\mathcal{N}(m, \sigma)$.

Свойства на $\mathcal{N}(m, \sigma)$:

1) Ако $\xi \sim \mathcal{N}(m, \sigma)$, то $m = E\xi$, $\sigma^2 = D\xi$.

$$E\xi = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

$$= \int_{-\infty}^{+\infty} (x - m + m) \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

$$= \int_{-\infty}^{\infty} (x-m) \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} dx + m \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-m)^2}{2\sigma^2}} dx$$

По форм. на Нютон-Лайбниц $\int_a^b f(x)dx = F(b) - F(a)$

Тогава $\int_{-\infty}^{\infty} f(x)dx = \lim\limits_{b \to \infty} F(b) - \lim\limits_{a \to -\infty} F(a) = 1 - 0 = 1$

Полагаме горе $\boxed{x - m = t}$. Тогава получаваме

$$E\xi = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2\sigma^2}} dt + m \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-m)^2}{2\sigma^2}} d$$

$$\underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2\sigma^2}} dt}_{\text{нечетна функция в симетричен интервал}} = 0.$$

$$= 0 + m \int_{-\infty}^{\infty} f(x)dx = m.1 = m$$

$$\boxed{E\xi = m}$$

$$D\xi = \int_{-\infty}^{\infty} (x-m)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

$$= \sigma^2 \int_{-\infty}^{\infty} \left(\frac{x-m}{\sigma}\right)^2 \cdot \frac{1}{\sqrt{2\pi}.\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

$$= \sigma^2 \int_{-\infty}^{\infty} \left(\frac{x-m}{\sigma}\right)^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} d\frac{x-m}{\sigma} \qquad \underline{\underline{t = \frac{x-m}{\sigma}}}$$

$$= \sigma^2 \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt =$$

$$= \sigma^2 \underbrace{\int_{-\infty}^{\infty} (t-0)^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}_{=1} = \sigma^2 DZ$$

t^2 се представя като $t \cdot t$ → внасяме под диференциала. Тогава можем да внесем и експонентата. Използваме интегриране по части, за да докажем, че инт. е 1.

2) Ако $\xi \sim N(m, \sigma)$, a – число, то $\xi + a \sim N(m+a, \sigma)$ → 2a)

и $a\xi \sim N(ma, \sigma|a|)$ → 2б)

$$F_{\xi+a}(x) = P(\xi + a \le x) = P(\xi \le x-a) = F_\xi(x-a)$$

$$f_{\xi+a}(x) = F'_{\xi+a}(x) = F'_\xi(x-a) = f_\xi(x-a)$$

$$f_\xi(x-a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a-m)}{2\sigma^2}}$$ – това е вероятностна плътност на нормално разпр. с параметри $m+a, \sigma$, т.е. $N(m+a, \sigma$

За $a\xi$. Нека $a>0$.

$$F_{\xi a}(x) = P(\xi a \le x) = P\left(\xi \le \frac{x}{a}\right) = F_\xi\left(\frac{x}{a}\right)$$

$$f_{\xi a}(x) = F'_{\xi a}(x) = F'_\xi\left(\frac{x}{a}\right)\cdot\frac{1}{a} = f_\xi\left(\frac{x}{a}\right)\cdot\frac{1}{a}$$

$$f_{\xi a}(x) = \frac{1}{a} f_\xi\left(\frac{x}{a}\right) = \frac{1}{\sqrt{2\pi}\,\sigma . a}\, e^{\frac{-\left(\frac{x}{a}-m\right)^2}{2\sigma^2}} =$$

$$= \frac{1}{\sqrt{2\pi}\,a\sigma}\, e^{-\frac{(x-ma)^2}{2\sigma^2 a^2}} \rightarrow \mathcal{N}(ma, \sigma|a|)$$

При $a<0$, неравенството си сменя посоката.
Оставено за самостоятелно обмисляне.

Следствие от св-во 2): Ако $\xi \sim \mathcal{N}(m, \sigma)$, то

$$Z = \frac{\xi - m}{\sigma} \sim \mathcal{N}(0,1).$$

Z-трансформация

Ето как се получава:

Прилагане 2а) към ξ и $a=-m \Rightarrow \xi - m \sim \mathcal{N}(0,\sigma)$

Прилагане 2б) към $\xi-m$ и $a=\frac{1}{\sigma} \Rightarrow Z=\frac{\xi-m}{\sigma} \sim \mathcal{N}(0,1)$

2). Гама разпределение

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\, dx \underset{\substack{\text{смяна:}\\ x\to\frac{x}{\beta}}}{=} \beta^{1-p}\int_0^\infty x^{p-1}\cdot e^{-\frac{x}{\beta}}\, d\left(\frac{x}{\beta}\right) =$$

$$= \beta^{-p}\int_0^\infty x^{p-1}\cdot e^{-\frac{x}{\beta}}\, dx$$

$$1 = \int_0^{\infty} \frac{x^{p-1} \cdot e^{-\frac{x}{\beta}}}{\beta^p \cdot \Gamma(p)} dx$$

Това равенство показва, че функцията $f(x)$

$$f(x) = \begin{cases} \frac{x^{p-1} \cdot e^{-\frac{x}{\beta}}}{\beta^p \cdot \Gamma(p)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

е вероятностна плътност на разпр., $\Gamma(p, \beta)$

p, β - параметри на разпр.

$\beta > 0$

Частни случаи: $p = 1 \rightarrow$ експоненциално разпр. $\mathcal{E}(\beta)$

05.12.2019г.

<u>Частни случаи</u>: 1) $\overset{\text{експон.}}{\mathcal{E}(\beta)}$ $\quad f(x) = \frac{1}{\beta} \cdot e^{-\frac{x}{\beta}}$

2) χ^2-разпр. $\quad \beta = 2, \; p = \frac{n}{2}, \; n = 1, 2, \ldots$

„хи-квадрат" $\quad$ n - степени на свобода

$$f(x) = \frac{x^{p-1} \cdot e^{-\frac{x}{\beta}}}{\Gamma(p) \cdot \beta^p}$$

$\uparrow$

плътност на $\Gamma(p, \beta)$

$\chi^2(n)$ - разпр.

$$f(x) = \frac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{\Gamma\left(\frac{n}{2}\right) \cdot 2^{\frac{n}{2}}} \rightarrow \text{плътност на } \chi^2(n)\text{-разпр.}$$

$\boxed{\Gamma(p+1) = p\,\Gamma(p)}$ - основно функционално уравнение на $\Gamma(p)$ - функцията

$p = n \qquad \Gamma(n) = (n-1)!$

Ако $\xi \sim \Gamma(p,\beta)$, то:

$$E\xi = \int_0^\infty x f(x)\,dx = \int_0^\infty \frac{x^p e^{-\frac{x}{\beta}}}{\Gamma(p)\beta^p}\,dx = \frac{\beta}{\Gamma(p)}\int_0^\infty \left(\frac{x}{\beta}\right)^p e^{-\frac{x}{\beta}}\,d\left(\frac{x}{\beta}\right) =$$

$$= \frac{\beta}{\Gamma(p)}\int_0^\infty x^{p+1-1}\cdot e^{-x}\,dx = \frac{\beta\,\Gamma(p+1)}{\Gamma(p)} = \beta\cdot p$$

$E(\xi^2) = \dots$ (самостоятелно)

3). Бета-разпр.

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx \quad , a,b>0$$

Функцията $f(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & 0<x<1 \\ 0, & x \overline{\in} (0,1) \end{cases}$ е

вероятностна плътност.

<u>Частни случаи</u>: 1) $a=b=1$, $f(x) = \begin{cases} 1, & x\in(0,1) \\ 0, & x\overline{\in}(0,1)\end{cases}$

стандартно равномерно разпределение

2) $a=b=\frac{1}{2}$, $f(x) = \begin{cases} \dfrac{1}{B(\frac{1}{2},\frac{1}{2})}\cdot\dfrac{1}{\sqrt{x(1-x)}}, & 0<x<1 \\ 0, & x\overline{\in}(0,1)\end{cases}$

разпределение на арксинуса

$$F(x) = \int_0^x f(x)\,dx = \frac{1}{B(\frac{1}{2},\frac{1}{2})}\int_0^x \frac{dx}{\sqrt{x(1-x)}} = \frac{2}{B(\frac{1}{2},\frac{1}{2})}\int_0^x \frac{d\sqrt{x}}{\sqrt{1-x}} \overset{\sqrt{x}=t}{=}$$

$$= \frac{2}{B(\frac{1}{2},\frac{1}{2})} \int_0^{x^2} \frac{dt}{\sqrt{1-t^2}} = \frac{2}{B(\frac{1}{2},\frac{1}{2})} \cdot \arcsin(x^2)$$

$$\Rightarrow F(x) = \frac{2}{B(\frac{1}{2},\frac{1}{2})} \arcsin(x^2) = \frac{2}{\pi} \arcsin(x^2)$$

Ако $\xi \sim B(a, b)$, то:

$$E\xi = \frac{1}{B(a,b)} \int_0^1 x^a (1-x)^{b-1} dx = \frac{B(a+1,b)}{B(a,b)}$$

$$E(\xi^2) = \frac{B(a+2,b)}{B(a,b)}$$

Зависимост: $\boxed{B(a,b) = \frac{\Gamma(a+b+1}{\Gamma(a)\Gamma(b}}$

Случайни вектори

? неслучаен вектор ?

неслучаен вектор – наредена n-торка от числа/обекти

$\underbrace{X_1, X_2, \dots, X_n}_{\text{случайни величини}}$

$(X_1, X_2, \dots, X_n)$ - случаен вектор

случайна точка в n-мерното пространство

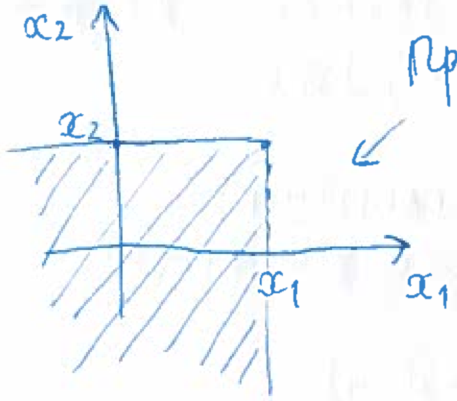
$\underbrace{\xi: \Omega \to \mathbb{R}^1}_{\text{сл. величина}}$

$(X_1, X_2, \dots, X_n): \Omega \to \mathbb{R}^n$

• Разпределение на $(X_1, X_2, \dots, X_n)$

съвместна функция на разпределение:

$$F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$



Пример в двумерното пространство

Кога функция на $\underline{\underline{n}}$ променливи е съвместна функция на разпределение на случаен вектор?

$$F(x_1, \infty) = F_{X_1}(x_1)$$

↳ функц. на разпр. на X_1

$$F(\infty, x_1, x_3, \dots, x_n) = F_{X \setminus X_1}(x_2, \dots, x_n)$$

условия за съгласуваност

- Съвместна вероятностна плътност:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

- Вероятностен смисъл на $f(x_1, \dots, x_n)$

$$\mathbb{P}\big(X_1 \in (x_1, x_1 + dx_1), X_2 \in (x_2, x_2 + dx_2), \dots, X_n \in (x_n, x_n + dx_n)\big) =$$

$$= f(x_1, x_2, \dots, x_n)\, dx_1\, dx_2 \dots dx_n$$

$$(*)\quad \mathbb{P}\big((x_1, x_2, \dots, x_n) \in D\big) = \int_D f(x_1, x_2, \dots, x_n)\, dx_1\, dx_2\, dx_3 \dots dx_n$$

↑ област в $\mathbb{R}^n$

$\boxed{\text{Деф.}}$ Случайни вектори с независими координати

$(X_1, X_2, \dots, X_n)$ е вектор с независими координати, ак. случайните вел. $X_1, X_2, \dots X_n$ са независими в съвкупност т.е. ако $\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \dots \mathbb{P}(X_n \in I_n)$ за всеки набор интервали $I_1, I_2, \dots, I_n$.

$\boxed{\text{T}}$ Теорема. Следните условия са еквивалентни:

1) $(X_1, X_2, \dots, X_n)$ е с независими координати;

2) $F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n)$ (съвм. ф-я на разпр.)

3) $f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n)$ (съвм. вер. плътност)

Достатъчно е да докажем $1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 1)$

$1) \Rightarrow 2)$) $I_1 = (-\infty, x_1]$

$I_2 = (-\infty, x_2]$

$\dots\dots$

$I_n = (-\infty, x_n]$

$2) \Rightarrow 3)$) Очевидно, след пресм. на $\dfrac{\partial^n F(\dots)}{\partial x_1 \dots \partial x_n}$ от двете страни на 2).

$3) \Rightarrow 1)$) от (*) за $D = I_1 \times I_2 \times \dots \times I_n$

$$\dots = \int_{I_1 \times \dots \times I_n} f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n)\, dx_1\, dx_2 \dots dx_n =$$

$$= \int_{I_1} f_{X_1}(x_1)\, dx_1 \int_{I_2} f_{X_2}(x_2)\, dx_2 \dots \int_{I_n} f_{X_n}(x_n)\, dx_n =$$

$$= \mathbb{P}(X_1 \in I_1)\, \mathbb{P}(X_2 \in I_2) \dots \mathbb{P}(X_n \in I_n)$$

$\Rightarrow 1)$.

Теоремата е доказана.

12.12.2019г. миналия път - сл. вектори с независ. коорд.
този път - сл. вектори със завис. коорд.

$\underline{n=2}$ $f(x,y) = f_1(x).f_2(y)$ сл. вектор (X, Y)

съвместна вер. плътност на X и Y; вер. плътност на X; вер. плътност на Y; независими

Условна вероятностна плътност на Y при условие, че $X=x$

$$P(Y\in(y,y+dy) \mid X=x) = ?$$

$$P(Y\in(y,y+dy) \mid X=x) = f_{Y|X}(y|x)\,dy$$

↳ условна вероятн. плътност

A, B - събития $\quad P(A|B) = \dfrac{P(A\cap B)}{P(B)}$

$$A = \{Y\in(y,y+dy)\}, \quad B=\{X=x\}$$

Заменяйки, получаваме:

$$P(Y\in(y,y+dy) \mid X=x) = \frac{P(Y\in(y,y+dy), X=x)}{P(X=x)}$$

$\cancel{B=\{X=x\}}$ - не върши работа, понеже води до $\boxed{\text{неопределеност от вида } \frac{0}{0}}$

$B=\{X\in(x,x+dx)\}$ - това вече върши работа

$$P(Y\in(y,y+dy) \mid X\in(x,x+dx)) = \frac{P(Y\in(y,y+dy), X\in(x,x+dx))}{P(X\in(x,x+dx))} =$$

$$= \frac{f(x,y)\,\cancel{dx}\,dy}{f_1(x)\,\cancel{dx}} = \frac{f(x,y)}{f_1(x)}\,dy$$

Получихме следната формула:

$$\boxed{f_{Y|X}(y|x) = \frac{f(x,y)}{f_1(x)}}$$

$$\underbrace{(X,Y)}_{\text{независ.}} \to f(x,y) = f_1(x).f_2(y)$$

$$\underbrace{(X,Y)}_{\text{зависими}} \to f(x,y) = f_{Y|X}(y|x).f_1(x)$$

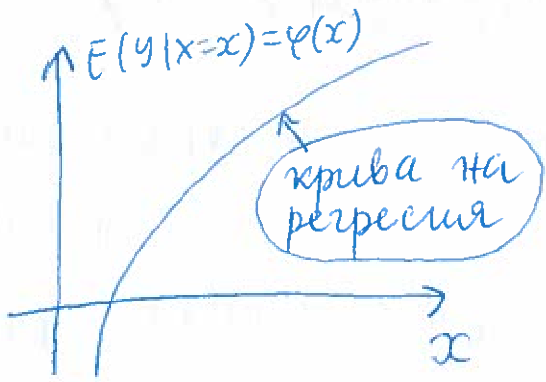
Условно математическо очакване на Y при условие, че $X=x \to E(Y|X=x)$

Условна дисперсия на Y при условие, че $X=x \to D(Y|X=$

$$E(Y|X=x) = \int_{-\infty}^{+\infty} x\, f_{Y|X}(y|x)\,dy$$

$$D(Y|X=x) = \int_{-\infty}^{+\infty} (x - E(Y|X=x))^2 . f_{Y|X}(y|x)\,dy$$

Графика на $E(Y|X=x)$



Това са твърде сложни формули.
Тяхното приложение най-добре би се илюстрирало с конкретен пример.

Пример: Двумерен нормален вектор (Двумерно нормално разпределение)

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho)^2}\left(\frac{(x-m_1)^2}{\sigma_1^2}-2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right)}$$

$\underbrace{m_1,\sigma_1}_{\text{параметри на } X},\underbrace{m_2,\sigma_2}_{\text{параметри на } Y},\rho$ - параметри

ρ – коеф. на корелация на X и Y

$$\rho=\frac{cov(X,Y)}{\sqrt{DX.DY}}$$

$$|\rho|\le 1$$

Ако $\rho=0$, $f(x,y)=f_1(x).f_2(y)$ (т.е. X и Y - независ.

При двумерните вектори понятията независимост и некорелираност са равносилни (от едното следва другото и обратно).

$f_{Y|X}(y|x)$, $E(Y|X=x)$, $D(Y|X=x)=?$

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_1(x)}$$

$$f_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}$$. Тогава $f_{Y|X}(y|x)=$

$$=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\cdot e^{-I}$$

$$I=e^{-\frac{1}{2(1-\rho)^2}\left(\frac{(x-m_1)^2}{\sigma_1^2}-2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right)}-\frac{(x-m_1)^2}{2\sigma_1^2}\Rightarrow$$

Отметка: $\frac{1}{1-\rho^2}-1=\frac{1-(1-\rho^2)}{1-\rho^2}=\frac{\rho^2}{1-\rho^2}$

$$\Rightarrow I = \frac{1}{2(1-\rho^2)}\left(\rho^2\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right) \Rightarrow$$

точен квадрат

$$\Rightarrow I = \frac{1}{2(1-\rho^2)}\left(\frac{y-m_2}{\sigma_2} - \rho\cdot\frac{x-m_1}{\sigma_1}\right)^2 =$$

$$= \frac{1}{2\sigma_2^2(1-\rho^2)}\left(y - m_2 - \rho\frac{\sigma_2}{\sigma_1}(x-m_1)\right)^2$$

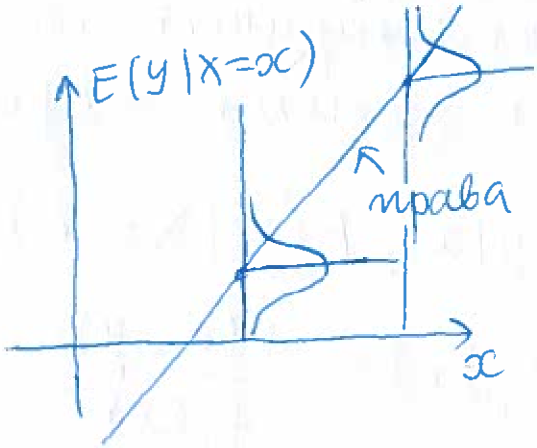
Но това е вероятностна плътност на норм. разпр.

$$N\left(m_2 + \rho\frac{\sigma_2}{\sigma_1}(x-m_1),\ \sigma_2\sqrt{1-\rho^2}\right) \ !$$

$E(Y|X=x) = m_2 + \rho\frac{\sigma_2}{\sigma_1}(x-m_1)$ – закон за линейната регресия

$D(Y|X=x) = \sigma_2^2(1-\rho^2)$

Графика на $E(Y|X=x)$



• При двумерен вектор кривата на регресия представлява права.

$E(Y|X=x) = m_2 + \rho\frac{\sigma_2}{\sigma_1}(x-m_1)$ – закон за линейната регресия

$D(Y|X=x) = \sigma_2^2(1-\rho^2)$ – закон за хомоскедастичност

19.12.2019г. Основни понятия и задачи на Статистиката

Основни понятия → Генерална съвкупност (популация); → Извадка

Понятие, асоциирано с генерална съвкупност, е „случайна величина".

$$X: \begin{cases} \Omega \to \mathbb{R} \\ \omega \to X(\omega) \end{cases}$$

С други думи, случайната величина е подходящ математически модел на генералната съвкупност.

Сега, какво представлява извадката? Ако случайната величина е мат. модел на генералната съвкупност, то мат. модел на извадката е <u>случаен вектор</u>.

Извадка $\rightarrow$ случаен вектор $(X_1, X_2, \ldots, X_n)$

обем на извадката

<u>Деф.</u> Извадката представлява случаен вектор, който притежава следните свойства:

1) $X_1, X_2, \ldots, X_n$ са дефинирани върху същото Ω като генералната съвкупност X;

2) $X_1, X_2, \ldots, X_n$ имат същото разпределение като генералната съвкупност X, формалният запис е $X_i \overset{d}{=} X$, $i = 1, 2, \ldots, n$;

3) $X_1, X_2, \ldots, X_n$ (още се наричат елементи на извадката) са <u>независими</u>.

<u>Забележка:</u> Извадката включва няколко наблюдения $((X_1, X_2, \ldots, X_n))$ и това е с цел да се обективизира истината за генералната съвкупност. Изключително важна роля играе <u>независимостта</u> (в противен случай е възможно да има изкривявания ~~в~~ резултатите, ~~х~~ свързани с генералната съвкупност.

Основни задачи на Статистиката → параметрични
→ Непараметрични

Характеризация на параметричните задачи:

• Видът на разпределението на генералната съвкупност X е известен, но един или повече от параметрите са неизвестни.

Характеризация на непараметричните задачи:

• Видът на разпределението на генералната съвкупност X е неизвестен. В този смисъл непараметричните задачи са по-комплицирани.

Ние ще се занимаваме с параметричните задачи и по-конкретно със задача за оценка на неизвестните параметри на X.

θ – неизвестен параметър на разпр. на X

$X_1, X_2, \dots, X_n$ – извадка

$\theta = ?$

Тази задача се решава по 3 начина:

* Метод на точковите оценки
* Метод на интервалните оценки (или още се нарича метод на доверителните интервали)
* Метод на статистическите хипотези (тестове)

Забележка. Методите са подредени по сложност. С увеличаване на сложността обаче и резултатите стават по-задоволителни и акуратни.

Метод на точковите оценки

1. Идея на метода

$\boxed{\text{Деф. 1}}$ Статистика – всяка функция, която зависи само и единствено от извадката.

$$\varphi(X_1, X_2, \dots, X_n)$$

С други думи, статистиката е функция от случайни величини.

$\boxed{\text{Деф. 2}}$ Точкова оценка – статистика, приета за стойност на неизвестния параметър.

Означение: $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n) \neq \theta$

2. Видове точкови оценки

а) неизместени – ако $E\hat{\theta} = \theta$

грешка на оценката:

$$\hat{\theta} - \theta = \hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta = \underbrace{(\hat{\theta} - E\hat{\theta})}_{\text{случайна грешка}} + \underbrace{(E\hat{\theta} - \theta)}_{\text{систематична грешка}}$$

Примери:

1) Ако $\theta = EX$, тогава оценката $\overline{X} = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ средноаритметично е неизместена.

$$E\overline{X} = E\,\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} E(X_1 + X_2 + \dots + X_n) =$$

$$= \frac{1}{n}(EX_1 + EX_2 + \dots + EX_n) \overset{\text{от св-вото на извадката (св-во(2))}}{=}$$

$$= \frac{1}{n} \cdot n EX = EX = \theta.$$

2) Ако $\theta = DX$

метод на моментите:

	X_1	X_2	$\dots$	X_n
	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

$DX = E(X-EX)^2$ е известно

$$\cancel{\hat\theta = \sum} \quad \hat\theta = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$$

Методът на моментите препоръчва оценката

$$\boxed{\hat\theta = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2} \rightarrow \text{това е изместена оценка!!!}$$

$$E\sum_{i=1}^{n}(X_i-\bar X)^2 = E\sum_{i=1}^{n}(X_i^2-2\bar X X_i+\bar X^2) =$$

$$= E\left(\sum_{i=1}^{n}X_i^2-2n\bar X^2+n\bar X^2\right) =$$

$$= E\left(\sum_{i=1}^{n}X_i^2-n\bar X^2\right)$$

Приемаме, че $EX=0$. $(DX = E(X)^2-(EX)^2)$

Тогава $E\sum_{i=1}^{n}(X_i-\bar X)^2 = E\left(\sum_{i=1}^{n}X_i^2-n\bar X^2\right) =$

$$= nDX - nE\bar X^2$$

$$E(\bar X^2) = E\left(\frac{X_1+\dots+X_n}{n}\right)^2 = \frac{1}{n^2}E(X_1+X_2+\dots+X_n)^2 =$$

$$= \frac{1}{n^2}E\left(\sum_{i=1}^{n}X_i^2+2\sum_{i<j}X_iX_j\right) =$$

$$= \frac{1}{n^2}\left(nDX+2\sum_{i<j}E(X_iX_j)\right) \underset{\text{от св-вото на извадката (св-во(3))}}{=}$$

$\Rightarrow$

$$E(\bar{X}^2) = \frac{DX}{n} + \frac{2}{n^2} \underbrace{\sum_{i<j} EX_i \cdot EX_j}_{=0} = \frac{DX}{n}$$

Тогава $E\sum_{i=1}^{n}(X_i-\bar{X})^2 = nDX - nE\bar{X}^2 = nDX - \cancel{n}\cdot\frac{DX}{\cancel{n}} =$

$$= nDX - DX = (n-1)DX$$

! Неизместената оценка е $\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$

поправена дисперсия на извадката

$$X_i - \bar{X} = X_i - \underset{\substack{\| \\ EX}}{EX_i} - (\bar{X} - \underset{\substack{\| \\ EX}}{E\bar{X}})$$

(согл. св-во (2) на изв.)

$$\bar{X} - E\bar{X} = \frac{X_1 - EX_1 + X_2 - EX_2 + \dots}{n}$$

09.01.2020г.

б) състоятелни оценки – ако $\hat{\theta}(X_1,\dots,X_n) \xrightarrow[n\to\infty]{P} \theta$

$\frac{k}{n} \xrightarrow[n\to\infty]{P} p$, $P(|\frac{k}{n}-p|>\varepsilon) \xrightarrow[n\to\infty]{} 0$

$P(|\hat{\theta}-\theta|>\varepsilon) \xrightarrow[n\to\infty]{} 0$ – характеризира състоятелна оценка

в) неизместени оценки с минимална дисперсия

$\theta = EX$ $\quad \bar{X} = \frac{X_1+X_2+\dots+X_n}{n}$, X_1 (първото наблюдение е неизместена оценка)

неизм. о-ка

X_2; $\frac{X_1+X_2}{2}$ – също неизместени

$DX_1 = DX_2 = DX$

$$D\frac{X_1+X_2}{2} = \frac{1}{4}D(X_1+X_2) = \frac{1}{4}DX_1 + \frac{1}{4}DX_2 = \frac{1}{4}.2DX = \frac{DX}{2}.$$

$$D\frac{X_1+X_2+\ldots+X_n}{n} = \frac{1}{n^2}D(X_1+X_2+\ldots+X_n) = \frac{1}{n^2}nDX = \frac{DX}{n}.$$

Лема Ако неизм. оценка с минимална дисперсия съществува, то тя е единствена.
(При краен брой неизм. оценки тя винаги съществува, но при безкраен брой -не - пример $\{\frac{1}{n}\}$)

Доказателство: Нека $\hat{\theta}_1, \hat{\theta}_2$ са две неизместени оценки с минимална дисперсия за параметъра θ.

Тогава $E\hat{\theta}_1 = E\hat{\theta}_2 = \theta$;
$D\hat{\theta}_1 = D\hat{\theta}_2 = d$.

Разглеждаме оценката $\frac{\hat{\theta}_1+\hat{\theta}_2}{2}$. Тя е неизместена $\left(\frac{1}{2}\cdot 2\theta = \theta\right)$. Това е първо. Второ, за дисперсията имаме $D\frac{\hat{\theta}_1+\hat{\theta}_2}{2} = \frac{1}{4}D(\hat{\theta}_1+\hat{\theta}_2) = \frac{D\hat{\theta}_1 + D\hat{\theta}_2 + 2\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2)}{4}$

$$= \frac{2d+2\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2)}{4} = \frac{d+\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2)}{2} \geq d$$ (понеже d е минимална дисперсия, всяка друга дисперсия е $>d$)

$$\frac{d+\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2)}{2} \geq d \Rightarrow \boxed{\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2) \geq d} \quad (1)$$

$$\boxed{\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2) \leq \sqrt{D\hat{\theta}_1 . D\hat{\theta}_2} = d} \quad (2)$$

От това допускане (и (1) и (2)) получихме, че $\operatorname{cov}(\hat{\theta}_1,\hat{\theta}_2)=$

Продължаваме.

$$\rho_{\hat{\theta}_1, \hat{\theta}_2} = \frac{d}{d} = 1 \Rightarrow \hat{\theta}_1 = a \cdot \hat{\theta}_2 + b$$

Изчисляване дисперсията на лявата страна на равенството, както и на дясната страна.

$$D\hat{\theta}_1 = D(a \cdot \hat{\theta}_2 + b)$$

$cov(X,Y) = E(X.Y) - EX.EY$

$$d = a^2 . d \Rightarrow a = \pm 1$$

а не може да е -1

Ако $a = -1$, то $\hat{\theta}_1 = b - \hat{\theta}_2$.

Тогава $a = 1$. $\hat{\theta}_1 = \hat{\theta}_2 + b$, но неизвестните са две

$\Rightarrow b = 0 \Rightarrow \hat{\theta}_1 = \hat{\theta}_2 \Rightarrow$ единствеността е доказана.

Граница на Рао-Крамер. Неравенство на Рао-Крамер

$t = t(X_1, X_2, \dots, X_n)$ - неизм. оценка за $\tau(\theta)$.

$f(x_1, x_2, \dots, x_n | \underline{\theta})$ - съвместна вероятностна плътност на извадката $X_1, X_2, \dots, X_n$

$$Dt \geq \frac{(\tau'(\theta))^2}{D\left(\frac{\partial}{\partial \theta} \ln f(X_1, X_2, \dots, X_n|\theta)\right)}$$

$\rightarrow$ информационно количество на Фишър

Деф. Ако за неизм. оценка t в неравенството на Рао-Крамер се достига еквивалентност /равенство, то говорим за <u>ефективна</u> оценка.

<u>Забележка</u>. Ефективната оценка е с минимална дисперсия.

Пример: $X \sim N(\theta, 1)$ $X_1, X_2, \dots, X_n$ - извадка

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$ - неизм. оценка за θ

Ще докажем, че $\bar{X}$ ~~е~~, освен неизместена, е и ефективна оценка. $\tau(\theta) = \theta$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$$

$$f(x_1, x_2, \dots, x_n|\theta) = (2\pi)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2}((x_1-\theta)^2 + (x_2-\theta)^2 + \dots + (x_n-\theta)}$$

$$\ln f(x_1, x_2, \dots, x_n|\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$$

$$\frac{\partial}{\partial\theta}\ln f(x_1, x_2, \dots, x_n|\theta) = \sum_{i=1}^{n}(x_i - \theta)$$

$$D\bar{X} \stackrel{?}{=} \frac{1}{\underbrace{D\left(\sum_{i=1}^{n}(x_i-\theta)\right)}_{n}}$$

$$D\bar{X} = \frac{1}{n}$$

Очевидно, равенството е налице. По-точна от тази дисперсия няма.

16.01.2020г.

R $\longrightarrow$ 20%

Задачи $\longrightarrow$ 20% ориентировъчно

Тест (теор.) $\longrightarrow$ 60%

12 въпроса / 6 въпроса за 3 (среден) на теста

$\downarrow$

90 минути

Пример:

X \ Y	1	2	3	4	5
1	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{15}$
2	$\frac{1}{6}$	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{60}$	$\frac{1}{30}$
3	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{30}$	?

? Р-е: 1 − (всички дадени вероятности)

а) Запълнете празната клетка

б) $\operatorname{cov}(X, Y) = ?$

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

$E(XY) = \binom{\text{ст.}}{X} \cdot \binom{\text{ст.}}{Y} \cdot (\text{вероятност}) + \ldots +$ общото

$E(XY) = 1 \cdot 1 \cdot \frac{1}{15} + 1 \cdot 2 \cdot \frac{1}{10} + \ldots +$ (за всяка клетка i, j)

EX?

X	1	2	3
	$\frac{20}{60}$		

сумиране по редове

EY?

Y	1	2	3

сумиране по стълбове

в) Независими ли са X и Y?

Ако $\operatorname{cov} = 0$ → въпросът остава

Ако $\operatorname{cov} \neq 0$ → X и Y са зависими

$\operatorname{cov} = 0$ → допълнителна проверка се изисква

$\{X=2\}$; $\{Y=2\}$ - независ.?

Проверката е:

$$\underbrace{P(\{X=2;\ Y=2\})}_{\frac{1}{30}} = P(X=2)P(Y=2)$$

Ако се окажат зависими, спиране → зависими

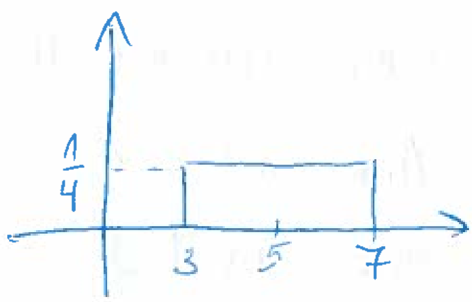
Ако са независими, продължаване със следващите клетки. Ако всички са независими → само тогава са независими

Какво е крива на регресия?
- зависимост на условното очакване на една величина от стойността, приета от др. величина

$\boxed{!}$ Основните дефиниции трябва да се знаят.
Напр.: Дайте определение на <u>извадка</u>.

Експоненциалното разпределение принадлежи към семейството на:
а) бета разпр.
б) гама разпр.
в) нормално разпр.
г) нито едно от посочените

Примерна задача:
дискр. $X \to 3, 4, 5, 6, 7$
непрек. $Y \to U(3, 7)$



а) $EX \overset{?}{=} EY$

б) $DX \overset{?}{=} DY$

Целта е да се знаят основните понятия, като мат. очакване, дисперсия, зависимост/независимост, корелираност/некорелираност, извадка и др.