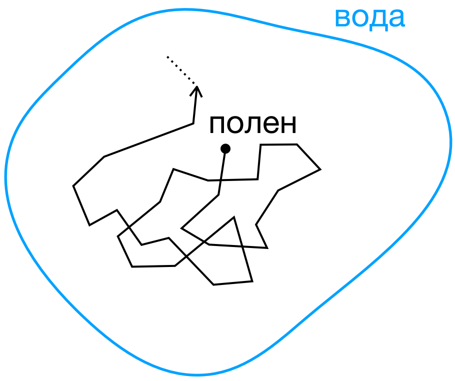
СЕМ, лекция 1
(2020-10-01)

През 1827 г., Робърт Браун, поставя частица полен върху вода и забелязва непрекъснато и хаотично движение. Той търси причината за това движение, което по-късно е наречено „брауново движение“.



Допускането на това, че частицата има вътрешна енергия, която да поражда движението, е довело до такъв избор на частицата (полен), който да осигури липсата на такава енергия.

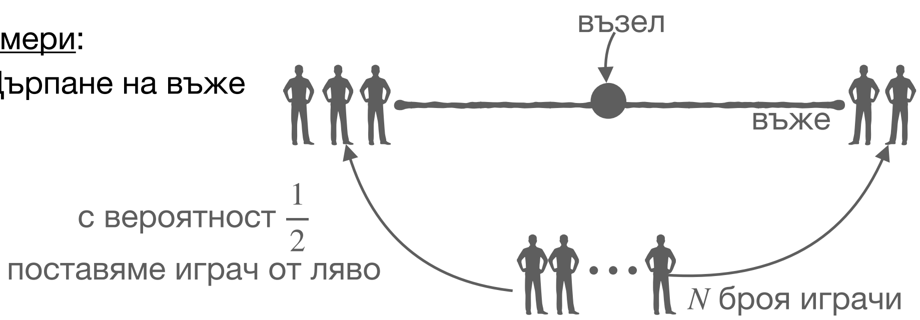
През 1905 г. Алберт Айнщайн обяснява истинската причина за това движение в семинарна статия. Благодарение на неговата кинетична теория на молекулите, той показва, че частица, поставена върху стояща вода бива удряна от молекулите на водата. Във всеки един момент от време, частицата ще я удрят множество молекули във всевъзможни посоки, което ще предизвиква рязка смяна на посоката и движението ѝ, в случай че сме взели достатъчно малка частица. Всичко това се случва, тъй като огромно количество молекули удрят полена едновременно във всеки един момент и в този момент, резултатната сила в очакване е 0 (което означава, че очакваното движение е нулево), но реализираната резултатна сила е в някаква посока и частицата се движи в нея. Това движение е универсално, тъй като то е резултат от всевъзможна колекция от движения.

В следващите години след Айнщайн, Жан Перан и колектив, успяват с помощта на това брауново движение да приближат броя на молекулите в изотопа C_{12} на въглерода. За времето си това е било постижение, което им е донесло нобелова награда.

Чисто физически е ясно, че движението не е съвсем случайно, тъй като и молекулите имат своите скорости и посоки на движение. Оказва се, че ако третираме молекулите като някакви случайни частици и приближим този процес, ние може да получим едно много добро приближение на истинското движение. То разбира се няма да е реалното брауново движение в истинския смисъл на думата, но то ще даде толкова добро приближение, че ние ще може да приближим други константни величини и куп други неща, на база на това приближение.

<u>Примери</u>:

⊕ Дърпане на въже



Средният брой хора, които ще се разполагат отляво и отдясно ще е $\frac{N}{2}$. Ако допуснем, че всеки човек дърпа въжето с еднаква сила, то в очакване възела ще е неподвижен, но реално той ще се движи във всеки един момент.

⨁ Системата „Бонус-Малус"

Това е известната система за глоби, при която шофьор, който нарушава правилника за движение по-често се глобява с по-големи суми, а такъв, който го нарушава по-рядко – с по-малки суми на глобите. Примерна система:

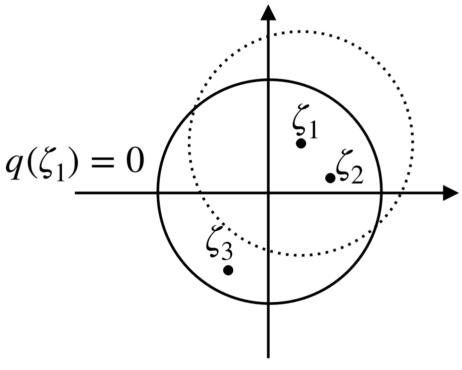
Брой нарушения									
1	2	3	4	5	6	7	8	9	10

Коефициент									
0.7	0.8	0.9	1.0	1.1	1.2	1.4	1.4	1.5	1.6

Идеята е, ако например глобата за дадено нарушение е 200 лв., шофьор, който е правел до 3 нарушения да заплаща 0.9 x 200 лв., такъв който е правел до 9 нарушения да заплаща 1.5 x 200 лв. и т.н. По-този начин ще се стимулират водачите да правят по-малко нарушения или по-точно да ги ограничават колкото се може повече.

Казуса, който възниква касае държавния апарат за събиране на данъци и застрахователи (ако например имаме подобен застрахователен проблем). Това е именно казуса: как да се изчислят тези коефициенти така, че сумарно платените глоби да са в очакване колкото ще са платените глоби ако системата няма плаващи коефициенти. Т.е. $P = \mathbb{E}\left[\frac{f_{\text{статично}}(X_n)}{f_{\text{плаващо}}(X_n)}\right] \approx 1$, където X_n цената, която заплаща даден водач с n нарушения. В случай, че $P > 1$ – ще се появи недоволство в данъкоплатците в името на държавата / застрахователите, тъй като сумарно ще се събират по-малко данъци от нарушителите. Ако пък $P < 1$, данъкоплатците / водачите отново ще са недоволни, тъй като ще заплащат сумарно повече за глоби, отколкото при фиксираната/статината система.

Преди около 6 години (~2015 г.), акад. Сендов формулира следната хипотеза: Ако вземем един полином $q(\zeta)$ и знаем, че нулите на полинома са в единичния кръг,



тогава ако вземем окръжност с радиус 1 и център която и да е нула на полинома, то в този единичен кръг ще има поне една нула от производната $\frac{\partial}{\partial \zeta}\left[q(\zeta)\right] = 0.$

Професор Теранс Тау доказва тази хипотеза за всички полиноми от достатъчно голяма степен $\forall n \geq n_0$ (т.е. останалите са краен брой). Неговото доказателство включва много вероятностни аргументи.

Дефиниция: (**Случаен експеримент**) Опит или експеримент, при който не може предварително да определим кой от възможните изходи ще се сбъдне.

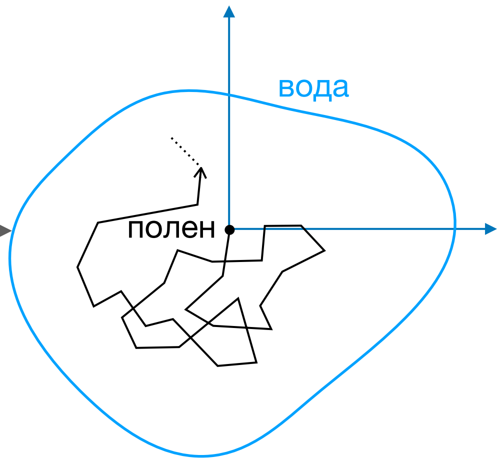
Всеки възможен елементарен изход ще означаваме с ω и ще наричаме елементарно събитие.

Дефиниция: (**Множество от всички елементарни събития**) С Ω ще означаваме съвкупността от всички елементарни събития на даден случаен експеримент.

⊕ $\Omega = \{$'ези', 'тура'$\}$; $\Omega = \{0, 1\}$

⊕ $\Omega = \{1, 2, 3, 4, 5, 6\}$; $\omega_5 = \{5\}$

⊕ $\Omega = \{$всички криви от $f(0) = (0{,}0)\}$



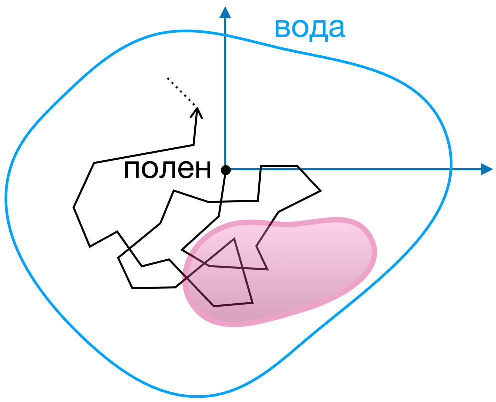
⊕ Тото „6 от 49“. Броя на всички елементарни изходи е равен на $\binom{49}{6}$.

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_{13\,983\,816}\}$

⊕ $\Omega = \mathbb{R}^+ = (0, \infty) \rightarrow$ времена на живот на CPU. Например $\omega = 23.5$ (23 месеца и половина)

Дефиниция: (**Събитие**) Всяко подмножество $A \subseteq \Omega$ наричаме събитие.

⊕ $\Omega = \{$клиенти$\}$, $A = \{$клиентите, които са жени$\}$
⊕ $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 3, 5\}$, признака по групиране е нечетност
⊕ $A = \{$всички криви, чиито траектории достигат розовия регион$\}$



Операции с множества

- $A \subseteq B \Leftrightarrow \omega \in A \Rightarrow \omega \in B$
- $A = B \Leftrightarrow \omega \in A \Rightarrow \omega \in B$ и $\omega \in B \Rightarrow \omega \in A$

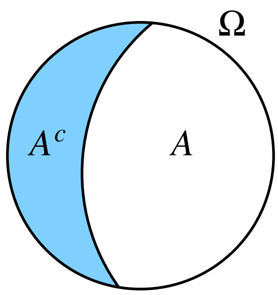
Дефиниция: (**„или“**), $A, B \subseteq \Omega$, то под обединението на събитията $A \cup B$ разбираме всички $\omega \in A$ или $\omega \in B$.

$\oplus$
$A = \{$хора между 20 и 30 г.$\}$
$B = \{$гласували за партия X$\}$
$A \cup B = \{$хора или между 20 и 30 г. или гласували за партия X$\}$

Дефиниция: (**„и“**), $A, B \subseteq \Omega$, то под сечението на събитията $A \cap B$ разбираме всички $\omega \in A$ и $\omega \in B$.

$\oplus$
$A = \{$хора между 20 и 30 г.$\}$
$B = \{$гласували за партия X$\}$
$A \cap B = \{$хора между 20 и 30 г. и гласували за партия X$\}$

Дефиниция: (**„отрицание“**), $A \subseteq \Omega$, то под допълнението на събитието A разбираме всички $\omega \notin A$ и бележим с A^c (понякога ще се случва да го бележим и с $\overline{A}$).



$\oplus$
$A^c = \{$всички хора, които са по-млади от 20 г. и по-възрастни от 30 г.$\}$

Свойства

а) (комутативност)	$A \cap B = B \cap A$ и $A \cup B = B \cup A$
б) (асоциативност)	$(A \cap B) \cap C = A \cap (B \cap C)$
б) (дистрибутивност)	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
г) (Закони на **де Морган**)	$(A \cup B)^c = A^c \cap B^c$
	$(A \cap B)^c = A^c \cup B^c$

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ за някое (поне едно) } i\}, \text{ където } A_i \in \Omega,\ \forall i \geq 1$$

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ за всяко } i\}, \text{ където } A_i \in \Omega,\ \forall i \geq 1$$

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c, \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c.$$

<u>Дефиниция</u>: (**σ-алгебра**). Нека Ω е съвкупност от елементарни събития. $\mathscr{A}$ е колекция от събития/подмножества на Ω. Наричаме $\mathscr{A}$ σ-алгебра, ако:

а) $\emptyset \in \mathscr{A}$;
б) $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$;
в) $A_i \in \mathscr{A},\ \forall i \geq 1 : \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$.

(Принадлежност на празното множество и затвореност относно допълнение и обединение на крайно или безкрайно обединения на множества от $\mathscr{A}$. Оказва се, че това е достатъчно (виж следствието по-долу).)

Ако премахнем „безкрайно“ обединение и оставим само „крайно“, то ще останем само с алгебра без σ.

<u>Следствие</u>: Ако $\mathscr{A}$ е σ-алгрбра, то:
а) $\Omega \in \mathscr{A}$
б) $A_i \in \mathscr{A},\ \forall i \geq 1$, то $\bigcap_{i=1}^{\infty} A_i \in \mathscr{A}$ (т.е. имаме и затвореност относно крайно/безкрайно сечение)

<u>Доказателство</u>:
а) $\emptyset^c = \Omega$, но $\emptyset \in \mathscr{A} \Rightarrow \Omega \in \mathscr{A}$;
б) $A_i \in \mathscr{A} \Rightarrow A_i^c \in \mathscr{A},\ \forall i \geq 1 \Rightarrow$

$$\bigcup_{i=1}^{\infty} A_i^c \in \mathscr{A} \Rightarrow \left(\bigcup_{i=1}^{\infty} A^c\right)^c \in \mathscr{A} \overset{\text{де Морган}}{\Rightarrow} \bigcap_{i=1}^{\infty} A_i \in \mathscr{A}.$$

$\oplus\ \Omega = \{0,1\}$
$\mathscr{A}_1 = \{\emptyset, \Omega\}$
$\mathscr{A}_2 = 2^\Omega = \left\{\emptyset, \Omega, \{0\}, \{1\}\right\}$

$\oplus\ \Omega = \{1, 2, \ldots, n\},\ |\Omega| = n$
$\mathscr{A} = 2^\Omega \Rightarrow \mathscr{A}$ има 2^n елемента.

⊕ „6 от 49“. $\Omega = \left\{ \omega_1, \ldots, \omega_{\binom{49}{6}} \right\}$. $\mathscr{A} = 2^n = 2^{13\,983\,816}$.

<u>Дефиниция</u>: (**Борелова сигма алгебра**) Ако $\mathscr{B}$ е произволна колекция от събития от Ω, то $\sigma(\mathscr{B})$ е най-малката/най-грануларната σ-алгебра, такава, че $\mathscr{B} \subseteq \sigma(\mathscr{B})$, т.е. $\forall B \in \mathscr{B}$, то $B \in \sigma(\mathscr{B})$.

$$\sigma(\mathscr{B}) = \bigcap_{\substack{\sigma_T \text{ е } \sigma\text{-алгебра} \\ \mathscr{B} \subseteq \sigma_T}} \sigma_T$$

върху $\sigma(\mathscr{B})$ може добре да се дефинира понятието дължина (мярка на Лебег).

$\Omega = \mathbb{R} = (-\infty, \infty)$; $\mathscr{B} = \{$всички отворени интервали$\}$

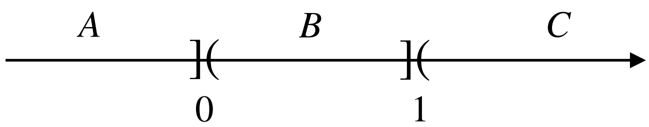
(a, b), (a, ∞), $(-\infty, b)$
$\mathscr{B}(\mathbb{R}) = \sigma(\mathscr{B})$ се нарича борелова σ-алгебра.

$x \in \mathbb{R}$, $\{x\} \overset{?}{\in} \mathscr{B}(\mathbb{R})$. отговорът е „ДА“, тъй като може да представим точката $\{x\}$ по следния начин:

$\{x\} = \bigcap_{i=1}^{\infty} \left(x - \frac{1}{i}, x + \frac{1}{i} \right) \in \mathscr{B}(\mathbb{R})$. Също така $[a, b) = \{a\} \cup (a, b) \in \mathscr{B}(\mathbb{R})$

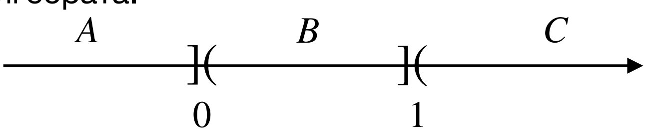
$[a, b] = \{a\} \cup (a, b) \cup \{b\} \in \mathscr{B}(\mathbb{R})$ и т.н.

Естествено, ако ни интересува само трихотомията,



то може да се ограничим само до σ-алгебрата
$\mathscr{A} = \{\emptyset, \mathbb{R}, A, B, C, A \cup B, B \cup C, C \cup A\}$, която има кардиланост $|\mathscr{A}| = 8$.

<u>Дефиниция</u>: (**Атом**) Ако $\mathscr{A}$ е σ-алгебра, то $A \in \mathscr{A}$ се нарича атом, ако от $B \subseteq A$ и $B \in \mathscr{A} \Rightarrow B = \emptyset$, т.е. не съществува нетривиално подсъбитие на $\mathscr{A}$, което е част от σ-алгебрата.

⊕ 

$|\mathscr{A}| = 8$. Атомите на $\mathscr{A}$ са A, B, C.

⊕ За $\mathscr{B}(\mathbb{R})$ атомите са $\{x\}$, т.е. всяка една точка от реалната права.

$\oplus$ $T = 10\ 001$ - брой тиражи на „6 от 49“. За всяко едно теглене имаме $\Omega_i = \left\{ \omega_1^{(i)}, \omega_2^{(i)}, \ldots, \omega_{\binom{49}{6}}^{(i)} \right\}$, $1 \le i \le T$ шесторки, които се падат в i-тия тираж.

Формална конструкция: всяка една от всевъзможна шесторка от $\omega^{(1)}$.

$$\Omega = \bigotimes_{i=1}^{T} \Omega_i = \{(\omega_{\cdot}^{(1)}, \omega_{\cdot}^{(2)}, \ldots, \omega_{\cdot}^{(T)})\}$$

$A = \{$паднали са се две еднакви наредени шесторки в T тиража$\}$

$A = \bigcup_{i=1}^{10\ 000} A_i$, където $A_i = \{\omega \in \Omega : \omega_{\cdot}^{(i)} = \omega_{\cdot}^{i+1}\}$.

$\oplus$ Имаме два пощенски плика A и B, в които има съответно сумите a и b. Нямаме никаква априорна информация за сумите, а човека, който ги е поставил в пликовете знае, че $a < b$. Избираме случайно с вероятност $\frac{1}{2}$ и отваряме съответния плик. Виждаме сумата x в плика, който сме избрали ($x = a$ или $x = b$), но нямаме никаква информация за това дали останалата сума е по-голяма или по-малка. Човека, който е сложил сумите в пликовете знае, но ние не. Дава ни се шанс, ако искаме, да си сменим плика. При пожелана смяна, ние със сигурност ще вземем сумата в новия плик, а ако откажем смяната ще останем със сумата от първоначално избрания плик. Да означим събитието $C = \{$печелим по-голямата сума $b\}$. Съществува ли такава стратегия, за която $\mathbb{P}(C) \ge \frac{1}{2}$? (т.е. има ли стратегия, която ви позволява да спечелите по-голямата от двете суми, с вероятност по-голяма от $\frac{1}{2}$)

Интуицията подвежда и се оказва, че има такава стратегия.

СЕМ, лекция 2

(2020-10-08)

<u>Дефиниция</u>: (**Вероятност**) Нека $\mathscr{A}$ е σ-алгебра върху множество от елементарни събития Ω. Тогава изображението $\mathbb{P} : \mathscr{A} \to [0,1]$ се нарича вероятност, ако са изпълнени следните три условия:

1) $\mathbb{P}(\Omega) = 1$

2) Ако $A \in \mathscr{A}$ и $A^c = \Omega \backslash A$, то $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

3) Ако $A_i \in \mathscr{A}$, $\forall i \geq 1$ и $A_i \cap A_j = \emptyset$ за $i \neq j$, то $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(A_k\right)$

(Ако имаме редица от непресичащи се събития, то вероятността поне едно от тях да се случи (операцията „или“) е сумата на индивидуалните вероятности. В този смисъл вероятността е мярка, тъй като това е най-класическото свойство на мярката)

<u>Следствие</u>: Нека имаме $\mathbb{P} : \mathscr{A} \to [0,1]$ е вероятност. Тогава са изпълнени следните свойства $(A, B \in \mathscr{A})$:

a) $\mathbb{P}(\emptyset) = 0$

b) Ако $B \subseteq A$, то $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$, $\forall A \in \mathscr{A}$

c) Ако $A \subseteq B$, то $\mathbb{P}(A) \leq \mathbb{P}(B)$ (Монотонност)

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

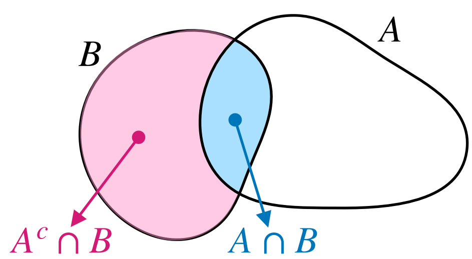
e) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, то $\mathbb{P}\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{n \to \infty} \mathbb{P}(A_n)$ (Непрекъснатост)

f) Ако имаме A_i, $i \geq 1$, то $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right)$. Това свойство е изпълнено и за всяко крайно обединение от събития: $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}\left(A_i\right)$, за някое $n \in \mathbb{N}$.

<u>Доказателство</u>:

a) $\emptyset = \Omega^c \overset{2)}{\Rightarrow} \mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 0$

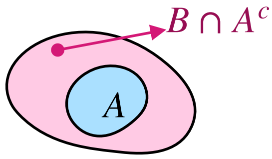
b) $A, B \in \mathscr{A}$



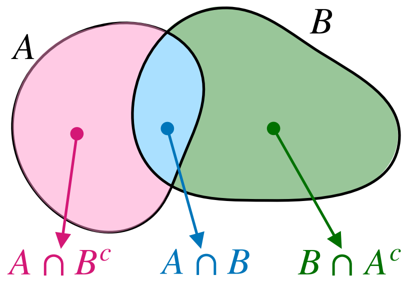
$$\Rightarrow B = (A \cap B) \cup (A^c \cap B) \Rightarrow \mathbb{P}(B) \overset{3)}{=} \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$

c) $B = A \cup (B \cap A^c)$

$\mathbb{P}(B) \overset{3)}{=} \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$



d) $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (B \cap A^c)$



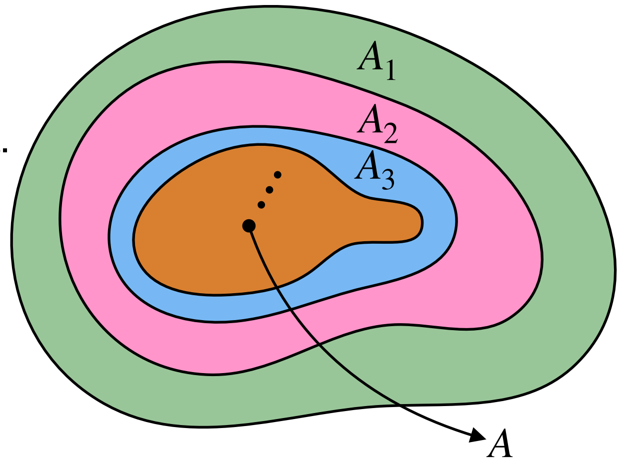
$\mathbb{P}(A \cup B) \overset{3)}{=} \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c) =$

$\overset{b)}{=} \mathbb{P}(A) + \mathbb{P}(B \cap A^c) + \underbrace{\mathbb{P}(A \cap B) - \mathbb{P}(A \cap B)}_{\text{добавяме и изваждаме}} =$

$\overset{b)}{=} \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

e) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$

$A = \bigcap_{j=1}^{\infty} A_j$ е множеството, което принадлежи на всяко едно от събитията.



Под $A \backslash B$ разбираме множеството A без множеството B, т.е. $A \backslash B = A \cap B^c$.

Цел: Да докажем, че $\mathbb{P}(A) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$.

$A_1 = \bigcup_{j=1}^{\infty} A_j \backslash A_{j+1} \cup A \overset{3)}{\Rightarrow}$

$1 \geq \mathbb{P}(A_1) = \sum_{j=1}^{\infty} \mathbb{P}(A_j \backslash A_{j+1}) + \mathbb{P}(A) \Rightarrow \sum_{j=1}^{\infty} \mathbb{P}(A_j \backslash A_{j+1}) < \infty$ е сходящ ред.

От друга страна, $A_n = \bigcup_{j=n}^{\infty} A_j \backslash A_{j+1} \cup A \Rightarrow \mathbb{P}(A_n) = \sum_{j=n}^{\infty} \mathbb{P}(A_j \backslash A_{j+1}) + \mathbb{P}(A)$.

Граничен преход: $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A) + \underbrace{\lim\limits_{n\to\infty} \sum_{j=n}^{\infty} \mathbb{P}(A_j \backslash A_{j+1})}_{=0} = \mathbb{P}(A)$.

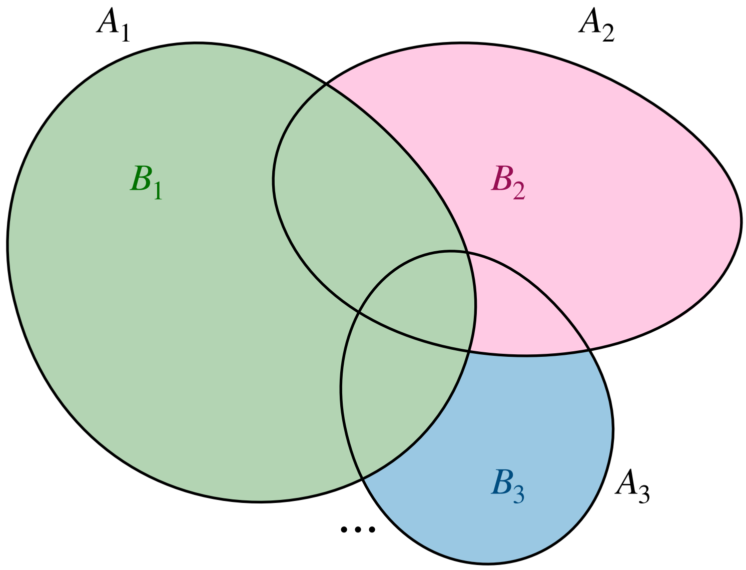
f) $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mathbb{P}(A_j)$

$\left(\sum_{j=1}^{\infty} \mathbb{P}(A_j) \geq \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right)\right.$, това може да се докаже по индукция, използвайки $\mathbb{P}(A \cup B) \overset{д)}{\leq} \mathbb{P}(A) + \mathbb{P}(B)$, но ние ще докажем директно по-общия случай за безкраен брой множества.$\Big)$

$A_1 = B_1$
$B_2 = A_2 \backslash A_1 = A_2 \cap A_1^c$
$B_3 = A_3 \backslash (A_1 \cup A_2)$
$\dots$

$B_n = A_n \backslash \left(\bigcup_{j=1}^{n-1} A_j \right) \subseteq A_n$

$\dots$



$B_i \cap B_j = \emptyset, \; i \neq j$

От друга страна имаме, че:

$\mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j \right) \overset{3)}{=} \sum_{j=1}^{\infty} \mathbb{P}(B_j)$, но $B_j \subseteq A_j \overset{c)}{\Rightarrow} \mathbb{P}(B_j) \leq \mathbb{P}(A_j)$. Т.е. е в сила

$\mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j \right) \overset{3)}{=} \sum_{j=1}^{\infty} \mathbb{P}(B_j) \leq \sum_{j=1}^{\infty} \mathbb{P}(A_j)$. Остана да докажем, че $\bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j$ (тъй като, ако това е изпълнено, то ще може да го заместим в предходното равенство и да получим желания резултат).

$\bigcup_{j=1}^{\infty} B_j \subseteq \bigcup_{j=1}^{\infty} A_j$ е очевидно, тъй като $B_j \subseteq A_j$,$\forall j$. Защо, обаче $\bigcup_{j=1}^{\infty} B_j \supseteq \bigcup_{j=1}^{\infty} A_j$?

Нека вземем елемент $\omega \in \bigcup_{j=1}^{\infty} A_j \Rightarrow \omega \in A_k$ за някое k. Да вземем $k = \min\{j \geq 1 : \omega \in A_j\}$ (най-малкия номер k на множество, в което елемента ω принадлежи – знаем, че със сигурност ще има поне едно такова множество)

$B_k = A_k \backslash \bigcup_{j=1}^{k-1} A_j$, но $\omega \notin \bigcup_{j=1}^{k-1} A_j \Rightarrow \omega \in B_k \Rightarrow \omega \in \bigcup_{j=1}^{\infty} B_j \Rightarrow$

$\Rightarrow \mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j \right) = \mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j \right) \overset{3)}{=} \sum_{j=1}^{\infty} \mathbb{P}\left(B_j \right) \leq \sum_{j=1}^{\infty} \mathbb{P}(A_j)$, което искахме да докажем.

Примери:

$\oplus_1$: $\Omega = \{0,1\}$; $\mathscr{A} = \big\{\emptyset, \Omega, \{0\}, \{1\}\big\} = 2^{\Omega}$

$\mathbb{P}\big(\{0\}\big) := p$, $\mathbb{P}\big(\{0\}\big) := 1 - p$, $p \in [0,1]$, то $\mathbb{P}$ е вероятност.

$\oplus_2$: Дискретна вероятност:

$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\} \simeq \{1, 2, \dots, N\}$
$\mathscr{A} = 2^{\Omega}$, $\left(p_i\right)_{i=1}^{N}$: $p_i \geq 0$, $\forall i \geq 1$ и $\sum_{i=1}^{N} p_i = 1$.

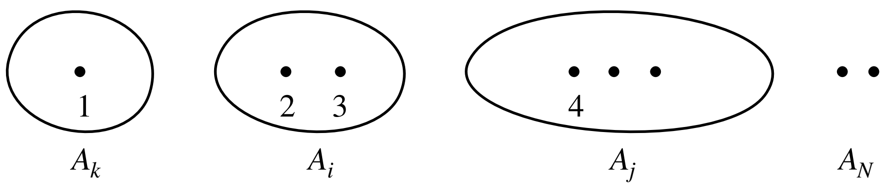
$\forall A \subseteq \Omega$, $\mathbb{P}(A) := \sum_{i \in A} p_i$

$\mathbb{P} : \mathscr{A} \to [0,1]$ е вероятност
$A = \{1, 3, 5\}$, $\mathbb{P}(A) = p_1 + p_3 + p_5$.

Проверка: По дефиниция $\mathbb{P}(\Omega) = \sum_{i \in \Omega} p_i = \sum_{i=1}^{N} p_i = 1$

Ако $A \subseteq \Omega$, то $\mathbb{P}(A^c) = \sum_{i \in A^c} p_i = \sum_{i \notin A} p_i = \sum_{i=1}^{N} p_i - \sum_{i \in A} p_i = 1 - \mathbb{P}(A)$.

Нека $A_1, \dots, A_k$ са непресичащи се събития:



$$\mathbb{P}\left(\bigcup_{j=1}^{k} A_j\right) = \sum_{i \in \bigcup_{j=1}^{k} A_j} p_i = \sum_{j=1}^{k} \sum_{i \in A_j} p_i = \left(\begin{matrix} i \text{ принадлежи на точно} \\ \text{едно от събитията} \end{matrix}\right) = \sum_{j=1}^{k} \mathbb{P}(A_j).$$

$\oplus_3$: $\Omega = \{1,2,\dots,N\}$
Ако дефинираме $p_i = \dfrac{1}{N}$, $1 \leq i \leq N$, то $\mathbb{P}(A) = \sum_{i \in A} \dfrac{1}{N} = \dfrac{|A|}{N}$ се нарича равномерна вероятност! Това означава, че всяко едно от елементарните събития има равен шанс да се сбъдне.

$\oplus_4$: $\Omega = \{1, 2, \dots, N\}$
$A = \{i \leq N : i$ е четно$\}$

$A^c = \{i \leq N : i \text{ е нечетно}\}$
$\mathbb{P}(A) = \dfrac{|A|}{N} = p$
$\mathbb{P}(A^c) = \dfrac{N - |A|}{N} = 1 - p$

$\Omega = \{0,\ 1\},\ \mathscr{A} = (\Omega, \emptyset, A, A^c) \subseteq 2^\Omega$

$\oplus_5$: Дискретни вероятности

$\Omega = \{\omega_1, \omega_2, \dots\} \simeq \{1, 2, \dots\}$
$\mathscr{A} = 2^\Omega$
Дадена е редица $(p_i)_{i=1}^\infty : p_i \geq 0, \forall i \geq 1$ и $\sum\limits_{i=1}^\infty p_i = 1$.
Ако $A \subseteq \mathscr{A}$, $\mathbb{P}(A) := \sum\limits_{i \in A} p_i$, то $\mathbb{P} : \mathscr{A} \to [0,1]$ задава вероятност.

$\oplus_6 : \Omega = \{1, 2, \dots\}$
$p_i = \dfrac{c}{i^2},\ i \geq 1$
$\mathbb{P}(A) = \sum\limits_{i \in A} p_i = c \sum\limits_{i \in A} \dfrac{1}{i^2}; \qquad \sum\limits_{i=1}^\infty = 1 \Leftrightarrow c = \dfrac{6}{\pi^2}.$

Следователно само за $c = \dfrac{6}{\pi^2}$ ще може да дефинираме вероятност.

$\oplus_7 : \Omega = \{0, 1, \dots\}$
$p_i = qp^i,\ i \geq 0, p + q = 1, p \in (0,1).$
$\sum\limits_{i=0}^\infty qp^i = \dfrac{q}{1-p} = 1$ (геометрична прогресия и дефиниране на геометрично рзпределение)

$\oplus_8 : \Omega = \{1, 2, \dots\}$ тук не може да дефинираме равномерно разпределение, т.е. $p_i = p_j$ за всяко i, j.

СЕМ, лекция 3

(2020-10-15)

$\Omega = \{1, 2, \ldots, N\}$, $p_i = \dfrac{1}{N}$.

Равномерна вероятност е такава вероятност, при която всяко едно от събитията е с равна вероятност да се сбъдне. Тя е прототип на случая, в който нямаме никаква априорна информация за модела и не искаме на нито един от възможните изходи да придаваме по-голяма или по-малка възможност за сбъдване от тези на останалите.

$\Omega = \{w_1, w_2, \ldots, w_m, \ldots\}$

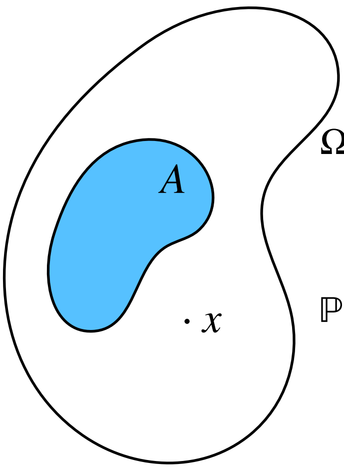
Когато имаме изброимо много елементи (събития) също имаме дискретна вероятност и множество от числа (вероятности), които вървят с всяко едно събитие $\{p_1, p_2, \ldots, p_n, \ldots\}$, $p_i \geq 0$, $\displaystyle\sum_{i=1}^{\infty} p_i = 1$. В този случай, обаче, не е възможно да зададем равномерна вероятност.

$\mathbb{P}(\{w_i\}) = p_i$, $i \geq 1$;

$A \subseteq \Omega$, $\mathbb{P}(A) = \displaystyle\sum_{w_i \in A} p_i$.

Геометрична вероятност

Тази вероятност е пример за вероятностно разпределение върху неизброимо множество. Най-простия прототип, който ще вземем, отново отговаря на равномерна вероятност.



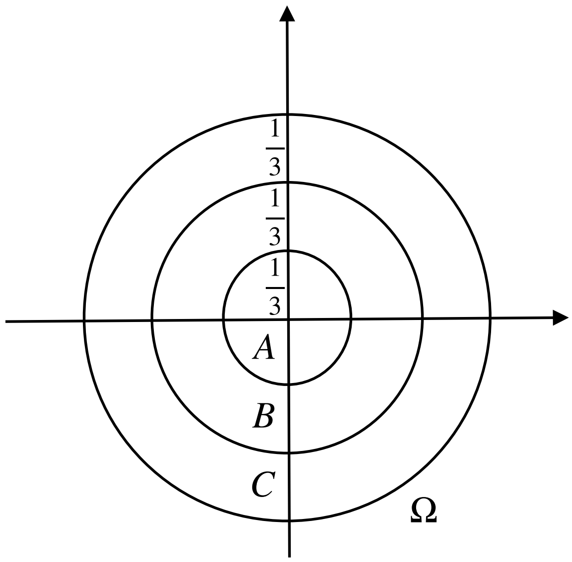
$\Omega \subseteq \mathbb{R}^2 : |\Omega| = \displaystyle\iint_{\Omega} \mathrm{d}\,x < \infty$

$\mathbb{P}(A) = \dfrac{|A|}{|\Omega|}$; всяко A отворено

Вероятността нещо да се случи в A, като подмножество на Ω ($A \subseteq \Omega$) е равна на площта (мярката) на A върху площта на Ω $\left(\dfrac{\mu(A)}{\mu(\Omega)}\right)$. Това е пример за равномерна вероятност, но равномерна само върху Ω. Това е така, защото самата вероятност зависи само от площта на (събитието) A – не зависи от нейната локация или форма, а само от нейната площ.

$\mathbb{P}(\{x\}) = \dfrac{|\{x\}|}{|\Omega|} = 0$. Площта на една точка е равна на 0. Вероятността на една точка е равна на 0 (има безбройно много други точки от каквато и да е площ). Вероятността да оценим конкретна точка е равна на 0 (площта и е нулева).

⊕



$$\Omega = \{x^2 + y^2 \leq 1\}$$

$$A = \left\{x^2 + y^2 \leq \frac{1}{9}\right\}$$

$$B = \left\{\frac{1}{9} \leq x^2 + y^2 \leq \frac{4}{9}\right\}$$

$$C = \left\{\frac{4}{9} \leq x^2 + y^2 \leq 1\right\}$$

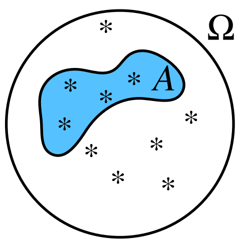
Ако се стреля хаотично (напълно аматьорски), то вероятностите да се улучат съответно множествата A, B и C са съответно:

$$A: \quad \mathbb{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\pi \times \left(\frac{1}{3}\right)^2}{\pi \times 1^2} = \frac{1}{9};$$

$$B: \quad \mathbb{P}(B) = \frac{\mu(A \cup B) - \mu(A)}{\mu(\Omega)} = \frac{\pi \times \left(\frac{2}{3}\right)^2 - \pi\left(\frac{1}{3}\right)^2}{\pi \times 1^2} = \frac{4}{9} - \frac{1}{9} = \frac{3}{9} = \frac{1}{3};$$

$$C: \quad \mathbb{P}(C) = \frac{\mu(A \cup B \cup C) - \mu(A \cup B)}{\mu(\Omega)} = \frac{\pi \times 1^2 - \pi \times \left(\frac{2}{3}\right)^2}{\pi \times 1^2} = 1 - \frac{4}{9} = \frac{5}{9}.$$

<u>Идея на Монте Карло алгоритмите</u>: Имаме, например, лицето на Ω, което в нашия случай е кръг и искаме да пресметнем лицето на $A \subseteq \Omega$:



$\dfrac{A(N)}{N}$

брой на точките попаднали в A
брой на всички хвърлени точки

Колкото повече точки включваме в изследването, толкова по-добро ще е приближението към шлощта на A. По този начин реално може да пресметнем интеграл без числени методи.

<u>Дефиниция</u>: (**Вероятностно пространство**) Наредена тройка $(\Omega, \mathscr{A}, \mathbb{P})$, където Ω е пространство от елементарни събития; $\underbrace{\mathscr{A}}_{\text{колекция от подмножества на } \Omega} \subseteq 2^{\Omega}$ е σ-алгебра и

$\mathbb{P} : \mathscr{A} \to [0,1]$ е вероятностна мярка (вероятното пространство).

$\oplus$ $\quad \Omega = \{1, 2, \ldots, N\}; \; \mathscr{A} = 2^{\Omega}$

$$\underbrace{\mathbb{P}(\{i\}) = \frac{1}{N}, \forall i \geq 1}$$

В случая когато сме взели вероятностите на всяко елементарно събитие да са равни – ще имаме равномерно вероятностно пространство. Т.е. всяко $i \geq 0$ има един и същ шанс да се сбъдне/падне.

$\oplus$ $\quad \Omega = \{x^2 + y^2 \leq 1\}; \; A, B, C; \; \Omega = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, C \cup A\}$

$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1, \mathbb{P}(A) = \frac{1}{9}, \mathbb{P}(B) = \frac{3}{9}, \mathbb{P}(C) = \frac{5}{9}, \mathbb{P}(A \cup B) = \frac{4}{9},$$

$$\mathbb{P}(B \cup C) = \frac{8}{9}, \mathbb{P}(C \cup A) = \frac{6}{9}.$$

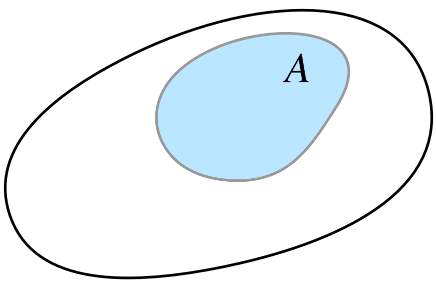
$\oplus$ $\quad \Omega = \{x^2 + y^2 \leq 1\}; \quad \underbrace{\mathbf{B}(\Omega) = \mathscr{A}}_{\text{бореловата сигма алгебра}} \; : \; \mathbb{P}(A) = \dfrac{|A|}{|\Omega|}; \qquad \forall A \in \Omega.$

Условна вероятност

$(\Omega, \mathscr{A}, \mathbb{P})$. Нашият модел трябва да е такъв, че да може да промени вероятностното пространство максимално добре спрямо постъпваща информация.

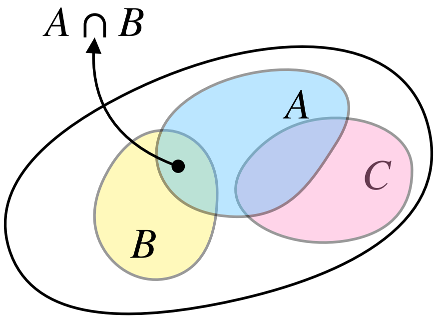
Изкуственият интелект, бейсовата статистика и много други имат зад себе си условни вероятности освен всичко останало, което включват.

$A \in \mathscr{A}$ настъпва. Първоначално тръгваме с Ω, но в даден етап настъпва събитието A. Това вече е в състояние да промени цялото вероятностно пространство/разпределение. Въпроса е как това се случва (как се променя)?



Дефиниция: (**Условна вероятност**) Нека $(\Omega, \mathscr{A}, \mathbb{P})$ е вероятностно пространство и е такова, че $A \in \mathscr{A} : \mathbb{P}(A) > 0$. Тогава условна вероятност при условие A наричаме $\mathbb{P}_A(B) = \mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \ \forall B \in \mathscr{A}$.

По този начин се изчисляват вероятностите на всяко едно B.



Т.е. ние знаем, че е настъпило събитието A и тази част от B, която не е в A – не ни интересува! Трябва да пресмятаме вероятността да се сбъдне частта на B, която попада в A $(A \cap B)$, като новото вероятностно състояние вече е A $(\Omega \mapsto A)$. Т.е. ние вече „живеем“ в $(A, \mathscr{A} \cap A, \mathbb{P}_A)$.

При настъпването на A се променя вероятностното пространство.
$A \cap \mathscr{A} = \{B \cap A \,|\, B \in \mathscr{A}\}$.

Най-доброто число, което бихме задали за вероятност на събитието B, при положение, че знаем (че се е случило) A е $\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$, както интуитивно, така и чисто в математически смисъл.

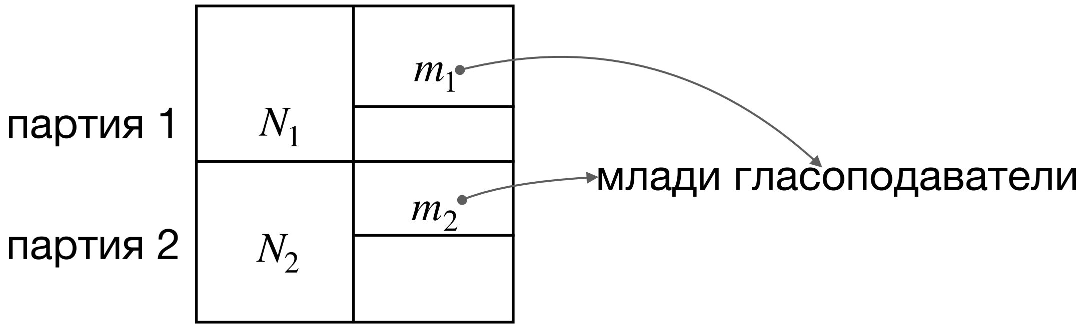
⊕ „6 от 49“

Пуснали сме фиш: $B = \{1, 2, 3, 4, 5, 6\}$. По една или друга причина ние научаваме, че са се паднали числата 1 и 2 в настоящия тираж (някога когато не е имало интернет). Т.е. $A = \{\underbrace{w}_{\text{шесторки}} \in \Omega \,|\, 1 \text{ и } 2 \in w\}$.

$\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(A)}$, тъй като $B \subseteq A$. Следователно

$\mathbb{P}(B|A) = \dfrac{\mathbb{P}(B)}{\mathbb{P}(A)} = \dfrac{\frac{1}{\binom{49}{6}}}{\frac{\binom{47}{4}}{\binom{47}{6}}} = \dfrac{\frac{1}{13\,983\,816}}{\frac{1}{128\,365}} \dfrac{1}{\binom{47}{4}} = \dfrac{1}{178\,365}$. Т.е. вероятността за печалба нараства значително (от порядъка на $70 - 80$ пъти).

⊕ Имаме две партии на някакви избори - Π_1 и Π_2.



Пита се някакъв човек – за коя партия е гласувал и той се окачва, че е млад. Нека $A = \{\text{млад}\}$ и $B = \{\text{гласувал за } \Pi_1\}$. Сега питаме – каква е вероятността да е гласувал за Π_1, ако се знае, че е млад?

$\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\frac{m_1}{N_1 + N_2}}{\frac{m_1 + m_2}{N_1 + N_2}} = \dfrac{m_1}{m_1 + m_2}$, т.е. числото така се променя, че не зависи от общия брой гласоподаватели, а само от младите такива (m_1 и m_2).

Независимост

<u>Дефиниция</u>: (**Независимост**) Две събития A и B се наричат независими, ако $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ $\big($Ако $\mathbb{P}(A) > 0 \Rightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$, т.е. независимостта означава, че случването на събитието A не ни носи никаква информация за $B\big)$.

<u>Дефиниция</u>: (**Взаимна независимост**) Дадени са събития $A_1, A_2, \ldots, A_n$. Казваме, че те са независими взаимно (в съвкупност), ако

$\forall M \in \{1, \ldots, n\}$ $(M \neq \emptyset$, M не е празното$)$ $\mathbb{P}\left(\bigcap_{i \in M} A_i\right) = \prod_{i \in M} \mathbb{P}(A_i)$.

Вероятността да се случи кое да е подмножество от M се разпада на произведението на вероятностите да се случат съставните (елементарните) множества от M.

<u>Теорема</u>: Нека $A_1, A_2, \ldots, A_n$ са n събития, така че $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) > 0$. Тогава

$$(*)\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right) \times \mathbb{P}\left(A_{n-1} \middle| \bigcap_{i=1}^{n-2} A_i\right) \times \ldots \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1)$$

Доказателство: По индукция. За $n = 1 : \mathbb{P}(A_1) = \mathbb{P}(A_1)$. Нека допуснем, че (*) е вярно за $n = k$, т.е. : $\mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right) = \mathbb{P}\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \times \ldots \times \mathbb{P}(A_2 | A_1)\mathbb{P}(A_1)$

(Индукционна хипотеза). Ще докажем верността на твърдението и за $n = k + 1$ (Индукционна стъпка):

$$\mathbb{P}\left(\bigcap_{i=1}^{k+1} A_i\right) = \mathbb{P}\left(A_{k+1} \cap \bigcap_{i=1}^{k} A_i\right) = \mathbb{P}\left(A_{k+1} \middle| \bigcap_{i=1}^{k} A_i\right) \mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right)$$

Следствие: Ако $A_1, A_2, \ldots, A_n$ са независими (ще разбираме, че са взаимно независими / зависими в съвкупност), то $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i)$.

$\oplus$ „6 от 49"

$$\underbrace{\Omega = \left\{w = (w_1^{(1)}, w_2^{(2)}, \ldots, w^{10\,001})\right\}}_{\text{всички паднали се 10 001 наредени шесторки до сега}}$$

$$A = \bigcup_{i=1}^{10\,000} A_i,\ A_i = \left\{w \in \Omega \,|\, w^{(i)} = w^{(i+1)}\right\}$$

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{10\,000}\right) = 1 - \mathbb{P}\left(\overline{\bigcup_{i=1}^{10\,000} A_i}\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{10\,000} \overline{A_i}\right) = 1 - \prod_{i=1}^{10\,000} \mathbb{P}(\overline{A_i}).$$

$\overline{A_i}$ са независими, тъй като ако $\overline{A_i} = \left\{w \in \Omega \,|\, w^{(i)} \neq w^{(i+1)}\right\}$ и $\overline{A_j} = \left\{w \in \Omega \,|\, w^{(j)} \neq w^{(j+1)}\right\}$, то за $|i - j| \geq 2$, $\overline{A_i}$ и $\overline{A_j}$ са независими. Освен това може да се покаже, че A_i и A_{i+1} също са независими.

$$\mathbb{P}(\overline{A_i} \cap \overline{A_{i+1}}) = \mathbb{P}(\overline{A_i}) \times \mathbb{P}(\overline{A_{i+1}}) = 1 = \prod_{i=1}^{10\,000} \mathbb{P}(\overline{A_i}) = 1 - \left(\mathbb{P}(\overline{A_1})\right)^{10\,000} =$$

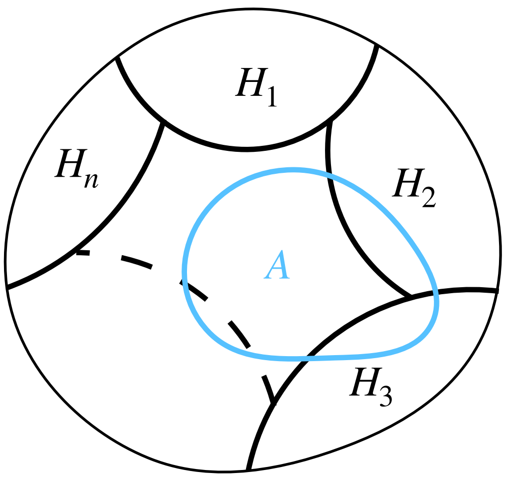
$$= 1 - \left(\frac{\binom{49}{6} - 1}{\binom{49}{6}}\right)^{10\,000} = 1 - \left(1 - \frac{1}{\binom{49}{6}}\right)^{10\,000} \approx 1 - \left(1 - 10\,000 \times \frac{1}{\binom{49}{6}}\right) \approx \frac{1}{1400}$$

Формула за пълната вероятност

<u>Дефиниция</u>: (**Пълна група от събития**) $H_1, H_2, \ldots, H_n$ се нарича пълна група от събития, ако $H_i \cap H_j = \emptyset, \forall i \neq j, i \leq n, j \leq n$ и $\dot{\bigcup}_{i=1}^{n} H_i = \Omega$. $\left(\dot{\bigcup} \right.$ е символ за обединение на непресичащи се множества$\left. \right)$.

<u>Теорема</u>: (**Формула за пълната вероятност**) Нека $H_1, H_2, \ldots, H_n$ е пълна група от събития в Ω и $A \in \mathscr{A}$. Тогава $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \,|\, H_i)\mathbb{P}(H_i)$ с конвенцията, че ако $\mathbb{P}(H_i) = 0$, то $\mathbb{P}(A \,|\, H_i)\mathbb{P}(H_i) = 0$.

<u>Доказателство</u>: $A = A \cap \Omega = A \cap \dot{\bigcup}_{i=1}^{n} H_i = \dot{\bigcup}_{i=1}^{n} A \cap H_i$.



Тогава

$$\mathbb{P}(A) = \mathbb{P}\left(\dot{\bigcup}_{i=1}^{n} A \cap H_i \right) \underbrace{=}_{\text{непресичащи се}} \sum_{i=1}^{n} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{n} \mathbb{P}(A \,|\, H_i)\mathbb{P}(H_i),$$

като накрая използвахме, че $\mathbb{P}(A \cap B) = \mathbb{P}(B \,|\, A)\mathbb{P}(A)$, което е формулата за условна вероятност.

<u>Теорема</u>: (**Формула на Бейс**) Нека $H_1, H_2, \ldots, H_n$ е пълна група от събития в Ω и $A \in \Omega$. Тогава $\mathbb{P}(H_k \,|\, A) = \dfrac{\mathbb{P}(A \,|\, H_k)\mathbb{P}(H_k)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A \,|\, H_k)\mathbb{P}(H_k)}{\sum_{i=1}^{n} \mathbb{P}(A \,|\, H_i)\mathbb{P}(H_i)}, \forall\, 1 \leq k \leq n.$

<u>Доказателство</u>: От една страна имаме, че $\mathbb{P}(A \cap H_k) = \mathbb{P}(A \,|\, H_k)\mathbb{P}(H_k)$, но от друга страна $\mathbb{P}(H_k \,|\, A)\mathbb{P}(A) = \mathbb{P}(H_k \cap A)$

$$\Rightarrow \mathbb{P}(H_k \,|\, A) = \frac{\mathbb{P}(H_k)\mathbb{P}(A \,|\, H_k)}{\underbrace{\mathbb{P}(A)}_{\text{формула за пълната вероятност}}}.$$ Формулата на Бейс показва как актуализираме нашата априорна вероятност за хипотеза да се случи, при положение че е настъпила някаква информация A.

$\oplus$ Допускаме, че на някакво летище има бърз COVID-19 тест, който е способен да индикира дали даден човек е заразен или не. Инфектираните са $1\,\%$ от посетителите на летището.

$I\ (infected) \longrightarrow 99\,\%$. Ако човек **е** носител на висруса, теста с $99\,\%$ засича вярно и индикира, за наличието на вирус.

$H\ (healthy) \longrightarrow 80\,\%$. Ако човек **не е** носител на заразата, тогава теста правилно връща индикатор (т.е. не реагира) с $80\,\%$ вероятност, че е здрав.

Да се пресметне вероятността, даден човек да е заразен, при положение, че теста е реагирал положително за наличие на вирус?

<u>Решение</u>:

Нека $A = \{$теста е реагирал положително за вирус (аларма)$\}$.

Търси се $\mathbb{P}(I\,|\,A)$.

$$\mathbb{P}(I\,|\,A) = \frac{\mathbb{P}(I \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A\,|\,I)\mathbb{P}(I)}{\mathbb{P}(A\,|\,I)\mathbb{P}(I) + \mathbb{P}(A\,|\,H)\mathbb{P}(H)} = \frac{99\% \times 1\,\%}{99\% \times 1\% + 20\% \times 99\,\%} =$$

$$= \frac{99}{99 + 20 \times 99} = \frac{1}{21}$$. Тук излолзвахме, че I и H са пълна група от събития, тъй като $I = \overline{H}$.

Оказва се, че теста не е много полезен. Получава се така, защото заразените са много малък процент от популацията (посетителите на летището), а грешката от $20\,\%$ е твърде голяма.

$\oplus$ $p \ll 10\,\%$ заразени.

Взимат се няколко проби и се смесват и се тества смеската. По този начин с един тест се правят n проби (където n е броя на извадката).

$$n - \text{проби накуп} = \begin{cases} \text{нито един заразен - жертваме 1 тест, но правим извод за цялата извадка;} \\ \text{има заразен, тогава правим } n \text{ индивидуални теста .} \end{cases}$$

Как да подберем размера на извадката n, така че да минимизираме използваните тестове.

Например при $p = 5\,\% = 0{,}05$, $p = 2\,\% = 0{,}02$. Да се помисли за домашно (случайни величини – предстои да се вземат).

$\oplus$ Имаме две числа a и b (две суми пари в два плика), за които ние не знаем нищо, освен че са положителни $a, b > 0$, но някой друг (водещия на играта, например) знае, че $a < b$.

Може ли да измислим стратегия, при която $\mathbb{P}(b) > \dfrac{1}{2}$ (избираме по-голямата сума с вероятнос по голяма от $50\,\%$)?

<u>Решение</u>:

Нека $A = \{$вижда a в $I^{\text{-ви}}$ плик$\}$ и $B = \{$вижда b във $II^{\text{-ри}}$ плик$\}$.

Знаем, че $\mathbb{P}(A) = \mathbb{P}(B) = \dfrac{1}{2}$. Нека още $C = \{$прави се смяна на пликовете$\}$.

$$\mathbb{P}(b) = \mathbb{P}(A \cap C) + \mathbb{P}(B \cap \overline{C}) = \mathbb{P}(C\,|\,A)\mathbb{P}(A) + \mathbb{P}(\overline{C}\,|\,B)\mathbb{P}(B) =$$

$$= \frac{1}{2}\left(\mathbb{P}(C\,|\,A) + \mathbb{P}(\overline{C}\,|\,B)\right) = \frac{1}{2}\left(\mathbb{P}(C\,|\,A) + 1 - \mathbb{P}(C\,|\,B)\right) = \frac{1}{2} + \frac{1}{2}\left(\mathbb{P}(C\,|\,A) - \mathbb{P}(C\,|\,B)\right)$$

Тук използвахме, че $\mathbb{P}(C\,|\,B) + \mathbb{P}(\overline{C}\,|\,B) = \dfrac{\mathbb{P}(C \cap B) + \mathbb{P}(\overline{C} \cap B)}{\mathbb{P}(B)} = 1.$

Т.е. задачата се свежда до въпроса: може ли да определим такава стратегия, при която $\mathbb{P}(C\,|\,A) > \mathbb{P}(C\,|\,B)$.

Нека разгледаме няколко случая, за да добием представа за сложността на въпроса:

Нека
$J = \mathbb{P}(C\,|\,A) - \mathbb{P}(C\,|\,B) \Rightarrow J = \mathbb{P}(C) \times \left(\mathbb{P}(A\,|\,C) - \mathbb{P}(B\,|\,C)\right) = \mathbb{P}(C) \times \left(\mathbb{P}(B) - \mathbb{P}(A)\right)$

1 сл. Ако никога не сменяме, т.е. $\mathbb{P}(C) = 0 \to J = 0$;

2 сл. Ако сменяме на всяко второ теглене, т.е. $\mathbb{P}(C) = \dfrac{1}{2} \to J = 0$;

3 сл. Ако винаги сменяме, т.е. $\mathbb{P}(C) = 1 \to J = 0$.

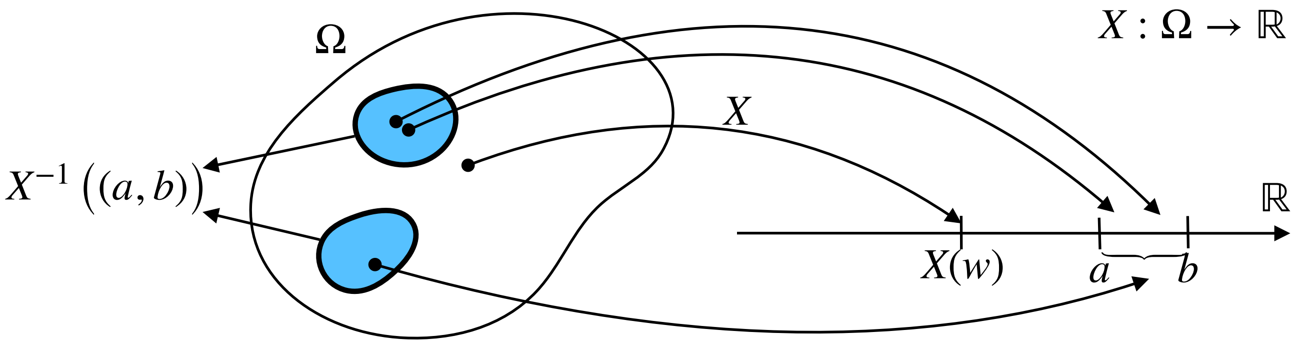
Обаче, ако си дефинираме стратегията по следния начин: ако виждаме числото x сменяме с вероятност e^{-x}. Т.е. $\mathbb{P}(C\,|\,\text{виждаме } x) = e^{-x} \Rightarrow J = e^{-a} - e^{-b} > 0.$

СЕМ, лекция 4

(2020-10-22)

Случайни величини

$V = (\Omega, \mathscr{A}, \mathbb{P})$ - вероятностно пространство.



Случайната величина X не е нито случайна, нито величина. Тя е грубо казано функция/изображение, което съпоставя на всяко елементарно събитие w от Ω - някакво реално число.

За да бъде X случайна величина, тя трябва да удовлетворява някакви критерии.

<u>Дефиниция</u>: (**Случайна величина**) Нека V е вероятностно пространство. Тогава $X : \Omega \to \mathbb{R}$ е случайна величина, тогава когато $\forall a < b,\ a, b \in \mathbb{R}$ е в сила $X^{-1}\left((a,b)\right) \in \mathscr{A}$, където $X^{-1}(\underset{B \subseteq \mathbb{R}}{B}) = \{w \in \Omega \,|\, X(w) \in B\}$. Т.е. трябва да имаме възможността да кажем каква е вероятността x да е между a и b.

Всички елементарни събития, към които, като приложим изображението X отиват в интервала (a,b) са множеството B.

<u>Факт</u>: Вярно е, че $X^{-1}(I) \in \mathscr{A}$, ако $I = (a,b];\ I = [a,b];\ I = \{x\},\ x \in \mathbb{R}$. Всеки интервал (a,b) от $\mathbb{R}$ има прообраз $B \subseteq \Omega$ и се изпраща в него с $X^{-1}\left((a,b)\right)$. Някой интервали може да се изпращат в празното множество $\varnothing$.

Това изображение $X : \Omega \to \mathbb{R}$ се нарича случайна величина, ако може да придаваме вероятност на прообразите му - на всички множества, които изпращаме във всеки един интервал.

<u>Теорема</u> (**Свойства на случайни величини**): Нека V е вероятностно пространство и X и Y са случайни величини ($X, Y : \Omega \to \mathbb{R}$). Тогава е в сила:

а) $aX \pm bY$ е случайна величина, $\forall a, b \in \mathbb{R}$;

б) cX е случайна величина, $\forall c \in \mathbb{R}$ (частен случай на а) : $a = c$ и $b = 0$);

в) XY е случайна величина;

г) ако $\mathbb{P}(Y = 0) = 0$, то $\dfrac{X}{Y}$ е случайна величина.

Случайните величини са функции, за които не може да знаем как действат навсякъде, тъй като това би било или твърде сложно или твърде скъпо, а в някои случаи може дори да не е ясно как да въведем пространството от елементарни събития Ω.

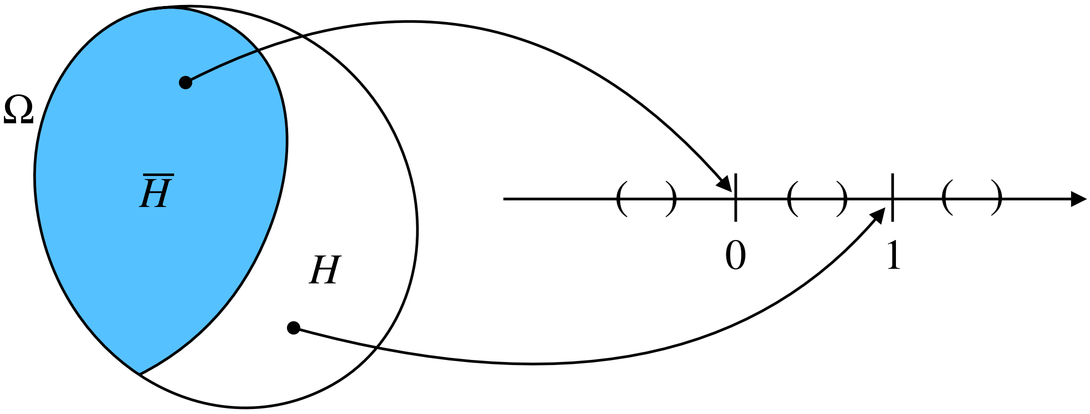
Дискретни случайни величини

Дефиниция: (**Индикаторна функция**) Нека Ω е множество от елементарни събития и $H \subseteq \Omega$. Тогава 1_H $(1_{\{H\}})$ се нарича индикаторна функция, ако
$1_H = \begin{cases} 1, & \text{ако } w \in H \\ 0, & \text{ако } w \in \overline{H} \end{cases}$. Грубо казано: $1_H : \Omega \to \mathbb{R}$.

Лема: Нека V е вероятностно пространство и $H \in \mathscr{A}$. Тогава $1_H : \Omega \to \mathbb{R}$ е случайна величина.

Доказателство:

Ако $X(w) := 1_H(w)$, то $X^{-1}\left(\{0\}\right) = \overline{H}$, а $X^{-1}\left(\{1\}\right) = H$.



$\forall a < b$ е вярно, че $X^{-1}(a,b) = \begin{cases} \emptyset, & \text{ако } a \geq 1 \text{ или } b \leq 0 \text{ или } a > 0 \text{ и } b < 1 \\ \Omega, & \text{ако } 0 \in (a,b) \text{ и } 1 \in (a,b) \\ H, & \text{ако } 1 \in (a,b) \text{ и } 0 \notin (a,b) \\ \overline{H}, & \text{ако } 1 \notin (a,b) \text{ и } 0 \in (a,b) \end{cases}$

Който и интервал (a,b) да вземем - изходите ще са един от 4-те възможни: $\emptyset$, Ω, H и $\overline{H}$, които са σ алгебра. Т.е. дефиницията е изпълнена (за всеки случай $X^{-1}\left((a,b)\right) \in \mathscr{A} \Rightarrow X = 1_H$ е сучайна величина). С това лемата е доказана.

$\mathscr{H} \subseteq \mathscr{A} : X(w) = 1_H(w) \in \{0,1\}$

$H = \{w \in \Omega \,|\, X(w) = 1\} = \{X = 1\}$
$\overline{H} = \{w \in \Omega \,|\, X(w) = 0\} = \{X = 0\}$

Имаме две възможности: $\mathbb{P}(X = 1) = p$ и $\mathbb{P}(X = 0) = q = 1 - p$, $p \in [0,1]$

Нека $V^* = (\Omega^*, \mathscr{A}^*, \mathbb{P}^*)$ е друго вероятностно пространство и $\mathscr{H}^* \subseteq \mathscr{A} : X^* = 1_{H^*}$ е вероятността $\mathbb{P}(X^* = 1) = p \Rightarrow$ вероятностно тези две случайни величини X и X^* не са различими.

Означения (за удобство):

$$\overline{x} = \begin{cases} (x_1, x_2, \ldots, x_n) - n \text{ различни числа} \\ (x_1, x_2, \ldots, x_k, \ldots) - \text{ изброимо много различни числа} \end{cases}$$

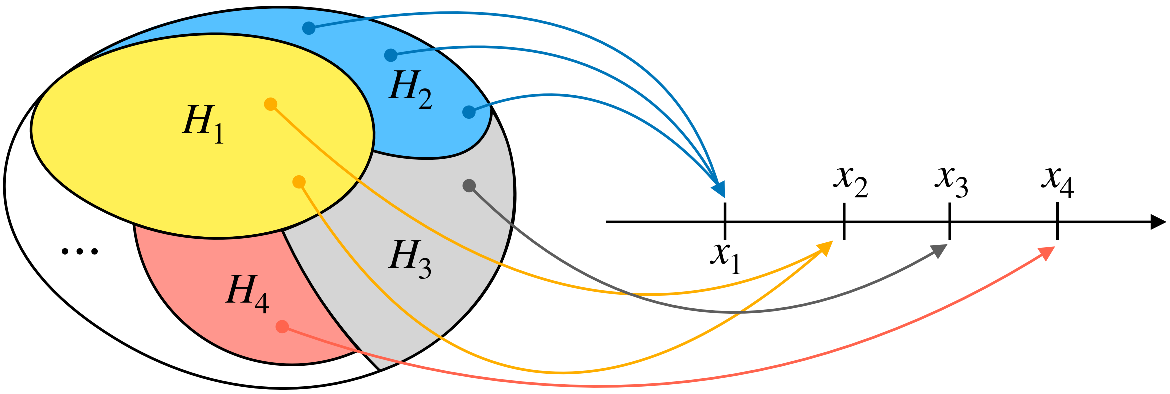
V е вероятностно пространство.

$$\mathscr{H} = \begin{cases} H_1, \ldots, H_n \text{ пълна група от събития във } V \\ (H_i)_{i \geq 1}, \text{ където } H_i \in A;\ H_i \cap H_j = \emptyset,\ i \neq j,\ \bigcup_{i=1}^{\infty} H_i = \Omega \end{cases}$$

<u>Дефиниция</u>: (**Дискретна случайна величина**) Нека V е вероятностно пространство. Дадени са $\overline{x}$ и $\mathscr{H}$. Тогава

$$X(w) = \sum_{j=1}^{n} x_j 1_{H_j}(w) \left(\text{или когато имаме изброим брой: } X(w) = \sum_{i=1}^{\infty} x_j 1_{H_j}(w) \right) \text{ се}$$

нарича дискретна случайна величина (взима или n различни или най-много изброимо много на брой различни стойности умножени по индикаторната функция). Кратък запис: $X = \sum_j x_j 1_{H_j}$.



$$H_j = \{X = x_j\} = \{w \in \Omega \,|\, X(w) = x_j\}$$

<u>Дефиниция</u>: (**Разпределение на дискретна случайна величина**)

Нека $X = \sum_j x_j 1_{H_j}$ е дискретна случайна величина. Тогава таблицата

X	x_1	x_2	$\cdots$	x_k	$\cdots$
$\mathbb{P}(X = x_j)$	p_1	p_2	$\cdots$	p_k	$\cdots$

където $\mathbb{P}(X = x_j) = p_j = \mathbb{P}(H_j)$ и $\sum_{j=1} p_j = 1$, се нарича разпределение на X.

$\oplus$ Измерваме дните, в които дадено CPU работи (функционира) и X случайната величина, която измерва броя на тези дни. Може да моделираме X по два начина:

- $X \in \mathbb{N}_0^+ = \{0, 1, 2, \dots\}$
- $X \in \{0, 1, 2, \dots, 1000\}$

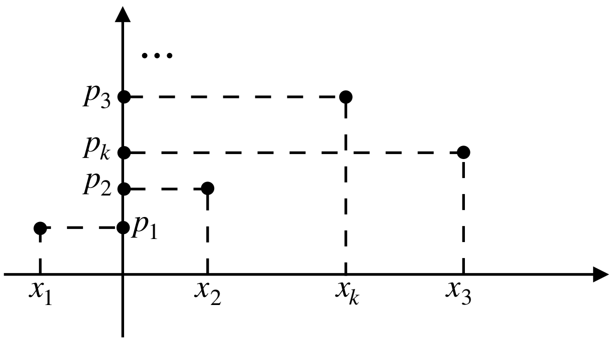
X	0	1	2	$\cdots$
$\mathbb{P}$	p_0	p_1	p_2	$\cdots$

$\sum_{j=0}^{\infty} p_j = 1$

X	0	1	$\cdots$	1000
$\mathbb{P}$	p_0	p_1	$\cdots$	p_{1000}

$\sum_{j=0}^{1000} p_j = 1$

Дефиниция: (**Хистограма**) Графиката по-долу се нарича хистограма:



$\oplus$ Хвърляме два зара. X и Y са случайните величини - точките от 1 до 6, съответно паднали се на $1^{\text{-вия}}$ зар. $Z = X + Y$.

Z	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Смяна на променливите на дискретни случайни величини

X - сл. вел., $g : \mathbb{R} \to \mathbb{R}$; $Y = g(X)$ - искаме да знаем дали Y е случайна величина.

Ако $X = \sum_j x_j 1_{H_j}$, то $Y = \sum_j g(x_j) 1_{H_j}$ е случайна величина и ако положим $y_i = g(x_j)$, то $Y = \sum_j y_j 1_{H_j}$.

За $g(x_m) = g(x_k)$, $m \neq k$ ще получим повтаряемост на някои стойности, но това няма да е грешка, просто за удобство и икономичност може да ги обединим като $\mathscr{H} = H_m \cup H_k$.

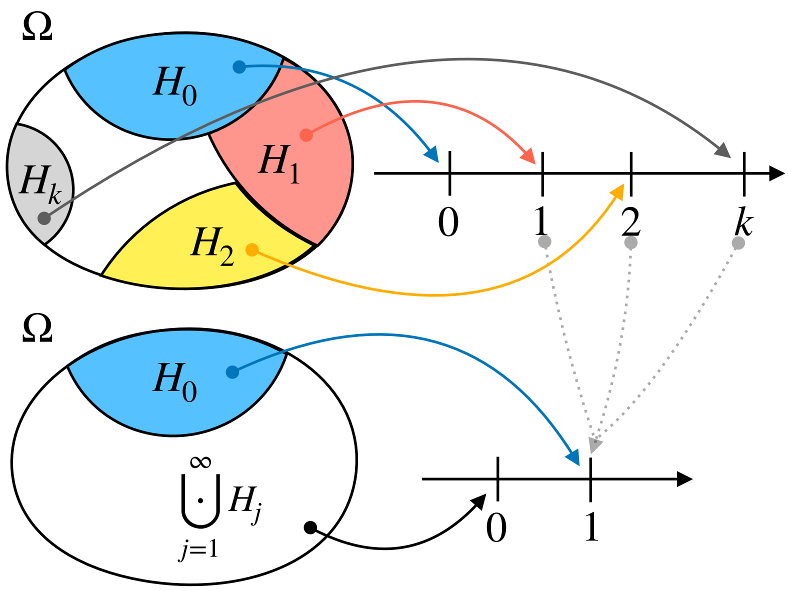
X	y_1	y_2	$\cdots$	y_m	$\cdots$	$\cancel{y_k}$	$\cdots$
$\mathbb{P}$	p_1	p_2	$\cdots$	$p_m + p_k$	$\cdots$	$\cancel{p_k}$	$\cdots$

$\oplus$ Имаме някакво CPU. $X = \sum_{j=0}^{\infty} j 1_{\{H_j\}}$, където $H_j = \{CPU$ работи точно j дни$\}$

$Y = g(X)$, където $g(n) = \begin{cases} 0, & \text{ако } n = 0 \\ 1, & \text{ако } n \geq 1 \end{cases}$.

X	0	1	2	$\cdots$
$\mathbb{P}$	p_0	p_1	p_2	$\cdots$

$Y = g(X)$	0	1
$\mathbb{P}$	p_0	$p_1 + p_2 + \ldots$



X, Y - дискретни случайни величини, $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Тогава $Z = g(Z, Y)$ е дискретна случайна величина.

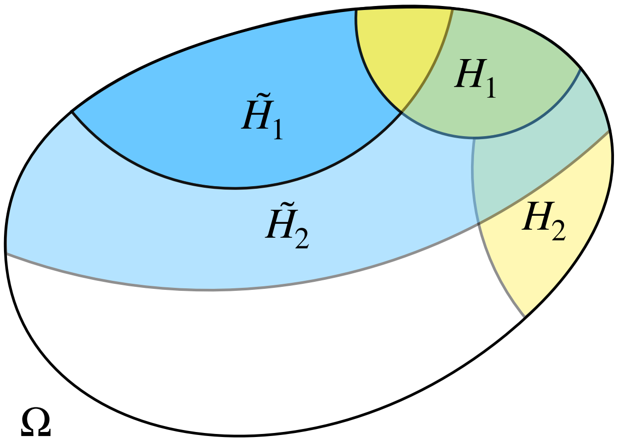
X	x_1	x_2	$\cdots$
$\mathbb{P}$	p_1	p_2	$\cdots$

$$\sum_i p_i = 1$$

$$Z = \sum_i \sum_j g(x_i, y_j) 1_{\{X=x_i, Y=y_j\}}$$

Y	y_1	y_2	$\cdots$
$\mathbb{P}$	q_1	q_2	$\cdots$

$$\sum_i q_i = 1$$



$$H_1 = \{X = x_1\};\ H_2 = \{X = x_2\};\ \dots$$

$$\tilde{H}_1 = \{Y = y_1\};\ \tilde{H}_2 = \{Y = y_2\};\ \dots$$

$$T_{ij} = H_i \cap \tilde{H}_j$$

Независимост на дискретни случайни величини

Дефиниция: (**Независимост на дискретни сл. вел.**) Нека X, Y са дискретни случайни величини във вероятностното пространство V. Тогава

$$\underbrace{X \perp\!\!\!\perp Y}_{X \text{ и } Y \text{ са независими}} \Leftrightarrow \mathbb{P}(X = x_j;\ Y = y_k) = \mathbb{P}(X = x_j \cap Y = y_k) \stackrel{def.}{=}$$

$$= \mathbb{P}(X = x_j)\mathbb{P}(Y = y_k), \forall j, k.$$

$\oplus$ $\quad \Omega = \{(0, 0);\ (0, 1);\ (1, 0);\ (1, 1)\};\ \mathscr{A} = 2^{\Omega};$

$\mathbb{P}\left(\{0, 0\}\right) = \mathbb{P}\left(\{0, 1\}\right) = \mathbb{P}\left(\{1, 0\}\right) = \mathbb{P}\left(\{1, 1\}\right) = \dfrac{1}{4}$ (имаме равномерна вероятност върху четирите елемента). Това е математическа конструкция на простия пример с хвърлянето на две монети.

$X : \Omega \to \mathbb{R} \qquad \underbrace{X(w)}_{\text{1-ва кордината}} = w(1)$. Например $X\left(\{0, 1\}\right) = 0$

$Y : \Omega \to \mathbb{R} \qquad \underbrace{Y(w)}_{\text{2-ра кордината}} = w(2)$. Например $Y(\{0, 1\}) = 1$

Тоест, първата монета е „тура“, а втората монета - „ези“ (ако сме дефинирали събитието „ези“ с 1-ца (за успех)).

X и Y са независими ($X \perp\!\!\!\perp Y$), т.к. $\mathbb{P}(X = i,\ Y = j) = \mathbb{P}(X = i)\mathbb{P}(Y = j)$, $\forall i, j \in \{0,1\}$.

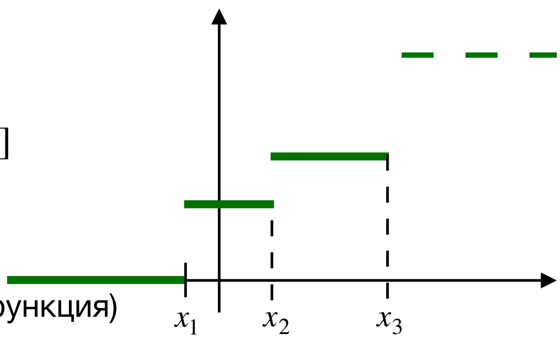
<u>Дефиниция</u>: (**Функция на разпределение на случайна величина**) Нека X е сл. вел. във вероятностно пространство V. Тогава $F_X(x) = \mathbb{P}(X < x)$, $\forall x \in (-\infty, \infty)$, се нарича функция на разпределение на X.

$\oplus$

X	x_1	x_2	x_3	$\cdots$
$\mathbb{P}$	p_1	p_2	p_3	$\cdots$

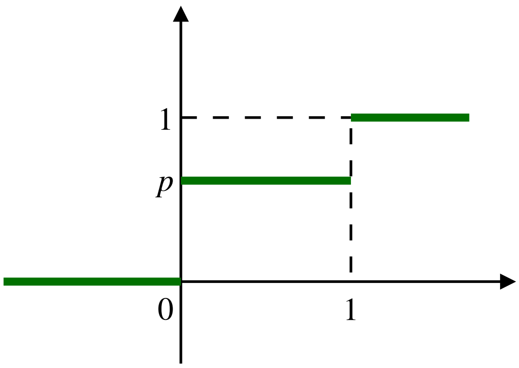
и $x_1 < x_2 < x_3 < \ldots$

$$\Rightarrow F_X(x) = \begin{cases} 0, & \text{ако } x \leq x_1 \\ p_1, & \text{ако } x \in (x_1, x_2] \\ p_1 + p_2, & \text{ако } x \in (x_2, x_3] \\ \cdots & \\ p_1 + \ldots + p_k, & \text{ако } x \in (x_k, x_{k+1}] \end{cases}$$

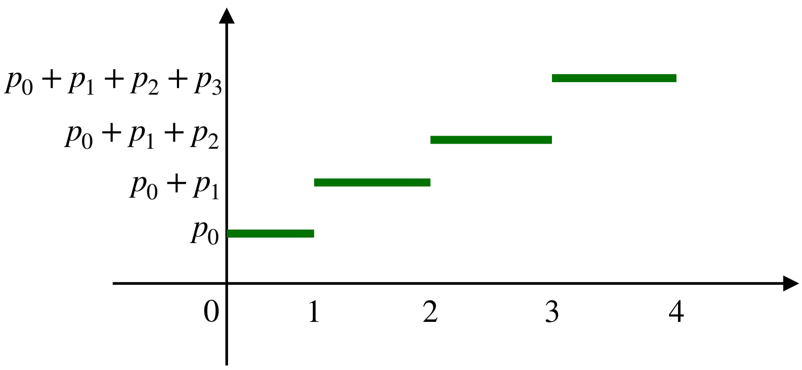


стъпаловидна (нарастваща функция)

$\oplus$ $X = 1_H$



$\oplus$ CPU



<u>Свойства</u>: $\lim\limits_{x \to \infty} F_X = 1$, $\lim\limits_{x \to -\infty} F_X = 0$.

Математическо очакване (за дискретни случайни величини)

Дефиниция: Нека X е дискретна сл. вел. Ако $\sum_j x_j p_j$ е добре деф. (т.е. е крайна), то

$$\mathbb{E}X = \sum_j x_j p_j = \sum_j \underbrace{x_j}_{\substack{\text{възможна} \\ \text{стойност на } X}} \times \underbrace{\mathbb{P}(X = x_j)}_{\substack{\text{вероятност за} \\ \text{тази стойност } x_j}} \quad \text{е очакването на } X.$$

Когато имаме краен брой стойности - тяхната сума ще е винаги добре дефинирана. Обаче, когато имаме изброимо много стойности, то тогава може сумата да не е крайна.

⊕ Ако X взема краен брой стоиности: $x_1, \ldots, x_n$, то $\sum_{j=1}^{n} x_j p_j < \infty \Rightarrow \mathbb{E}X = \sum_{j=1}^{n} x_j p_j$ винаги съществува.

⊕ Ако X е такава случайна величина, че $\mathbb{P}(X = x_j) = \dfrac{6}{\pi^2} \times \dfrac{1}{j^2}$, $j \geq 1$. Тогава

$\sum_{j=1}^{\infty} \mathbb{P}(X = j) = \dfrac{6}{\pi^2} \sum_{j=1}^{\infty} \dfrac{1}{j^2} = 1$, тъй като $\sum_{j=1}^{\infty} \dfrac{1}{j^2} = \dfrac{\pi^2}{6}$. Но тази случайна величина

няма очакаване, тъй като

$$\mathbb{E}X = \sum_{j=1}^{\infty} j \times \frac{6}{\pi^2} \times \frac{1}{j^2} = \frac{6}{\pi^2} \sum_{j=1}^{\infty} \frac{1}{j} = \frac{6}{\pi^2} \times \underbrace{(\text{хармоничния ред})}_{\text{не схожда}} = \infty.$$

Коментар: $f(a) = \sum_j (x_j - a^2) p_j$ е функция на a и тя се минимизира, когато $a = \mathbb{E}X$.

$\min_{a \in \mathbb{R}} f(a) = f(\mathbb{E}X)$. Т.е. $\mathbb{E}X$ минимизира квадратичната грешка.

⊕ Ако имаме равномерно разпределение върху $\{x_1, x_2, \ldots, x_n\}$, то тогава $\mathbb{E}X = \sum_{j=1}^{n} x_j \times \dfrac{1}{n} = \dfrac{1}{n} \sum_{j=1}^{n} x_j = \overline{X}$ (което е средно аритметичното на X).

Средно аритметичното е математическото очакване на равномерното разпределение върху дадени точки.

СЕМ, лекция 5

(2020-10-29)

Да се върнем на примера от лекция 3 за правене на COVID-19 тест на извадка от n души едновременно.

Нека X_n е случайната величина, която отговаря на n човека, които се събират в една проба. Тогава:

$$X_n = \begin{cases} 1, & \text{ако всичките } n \text{ не са заразени;} \\ n+1, & \text{ако има поне един заразен от всички } n \text{ души в извадката} \\ & \text{(+1, т.к. вече сме изхабили един тест за цялата извадка)} \end{cases}$$

Тъй като вероятността един човек да е заразен е равна на p, то вероятността да не е заразещ ще е равна на $1 - p$.

$$X_n = \begin{cases} 1, & (1-p)^n \\ n+1, & 1-(1-p)^n \end{cases}$$

X_n	1	$n+1$
$\mathbb{P}$	$(1-p)^n$	$1-(1-p)^n$

$\mathbb{E}X_n = 1 \times (1-p)^n + (n+1) \times \left(1-(1-p)^n\right) = \cancel{(1-p)^n} + n - n(1-p)^n + 1 - \cancel{(1-p)^n} =$ $= n + 1 - n(1-p)^n$ (трансцедентна функция). Остана само да минимизираме тази функция $\dfrac{\mathbb{E}X_n}{n}$, за дадено p. Например при

$p = 0{,}05 \Rightarrow \min\limits_{n \geq 1} \dfrac{\mathbb{E}X_n}{n} = \dfrac{\mathbb{E}X_5}{5} = 4.626216$. Т.е. за $n = 5$ ще използваме най-малко тестове в очакване, за да определим всеки един от посетителите на летището дали е носител на вирус.

Wolfram Alpha: $\min \left\{1 + \dfrac{1}{x} - 0.95^x\right\} \approx 0.426216$ в $x \approx 5.02239$

<u>Коментар</u>: Имаме $5k$ човека, които да тестваме.

$X_5(1), X_5(2), \ldots, X_5(k)$.

$$\underbrace{X_5(1) + X_5(2) + \ldots + X_5(k)}_{k} = 2{,}1310 = \mathbb{E}X_5.$$

Свойства на математическото очакване

а) Ако $g : \mathbb{R} \to \mathbb{R}$ и $Y = g(X)$, то $\mathbb{E}Y = \mathbb{E}g(X) = \sum\limits_j g(x_j)p_j$, ако тази сума е добре дефинирана, което следва директно от дефиницията на Y;

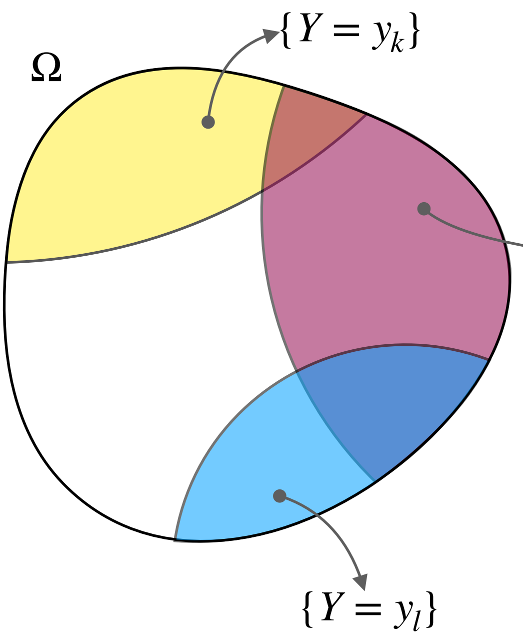
б) Ако $X \geq 0 \Rightarrow \mathbb{E}X \geq 0$ (позитивност) $\Leftarrow \mathbb{E}X = \sum_j \underbrace{x_j}_{\geq 0} \underbrace{p_j}_{\geq 0} \geq 0;$

в) Ако $X \equiv c\ (const.)$:

X	c
$\mathbb{P}$	1

$\Rightarrow \mathbb{E}X = c \Leftarrow \mathbb{E}X = c \times 1 = c.$

г) $\mathbb{E}cX = c\mathbb{E}X$, където $c \in \mathbb{R}$ е константа. $\Leftarrow$ ако вземем функцията $g(x) = cx$, то $Y = cX$. Тогава имаме, че $\mathbb{E}Y = \sum_j cx_jp_j = c\sum_j x_jp_j = c\mathbb{E}X.$



X, Y- дискретни случайни величини върху Ω

$$\{X = x_j\} = \{X = x_j\} \cap \dot{\bigcup_i} \{Y = y_i\} =$$
$$= \bigcup_i \{X = x_j\} \cap \{Y = y_i\}$$

$$\underbrace{\mathbb{P}(X = x_j) = p_j = \sum_i \mathbb{P}(X = x_j \cap Y = y_i)}_{\text{формула за пълната вероятност}}$$

д) Нека X, Y са две дискретни случайни величини. Тогава $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$

Доказателство:

$$\mathbb{E}(X+Y) = \sum_i \sum_j (x_j + y_i)\mathbb{P}(X = x_j \cap Y = y_i) =$$
$$= \sum_i \sum_j x_j\mathbb{P}(X = x_j \cap Y = y_i) + \sum_i \sum_j y_i\mathbb{P}(X = x_j \cap Y = y_i) =$$
$$= \sum_j x_j \underbrace{\sum_i \mathbb{P}(X = x_j \cap Y = y_i)}_{\mathbb{P}(X = x_j)\text{, т.к. изчерпваме}} + \sum_i y_i \underbrace{\sum_j \mathbb{P}(X = x_j \cap Y = y_i)}_{\mathbb{P}(Y=y_i)} =$$

$\forall$ възм. ст-ти на y

сечени с x_j

ф-ла за пълната вер.

$$= \sum_j x_j\mathbb{P}(X = x_j) + \sum_i y_i\mathbb{P}(Y = y_i) = \mathbb{E}X + \mathbb{E}Y,$$ което искахме да докажем.

Твърдение: Нека X и Y са дискретни случайни величини. Нека в допълнение имаме, че $X \perp\!\!\!\perp Y$. Тогава $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.

Доказателство:

Въвеждаме функцията $g(x, y) = xy$.

$$\mathbb{E}XY = \mathbb{E}g(X, Y) =$$

$$= \sum_j \sum_i x_j y_i \mathbb{P}(X = x_j \cap Y = y_i) \overset{X \perp\!\!\!\perp Y}{=} \sum_j \sum_i x_j x_i \mathbb{P}(X = x_j)\mathbb{P}(Y = y_i) =$$

$$= \sum_j x_j \mathbb{P}(X = x_j) \sum_i y_i \mathbb{P}(Y = y_i) = \mathbb{E}X\mathbb{E}Y,$$ което искахме да докажем.

⊕ Игра на рулетка (18 червени, 18 черни, 1 зелено).

Разглеждаме два вида игри:
Играч 1: играе само на черно с по 1лв.
Играч 2: играе само на числото 5 с по 1лв.

Очевидно двете игри са коренно различни. Нека се опитаме да ги оценим по някакъв начин и посочим и дори притеглим разликите.

Първо нека пресметнем математическото очакване на тяхната възвращаемост (печалба/загуба) за играта на двата играча.

Нека X е играча, който играе само на едно чисо (в случая това е числото 5, но реално може да бъде което и да е число от рулетката). Тогава:

X	-1	35
$\mathbb{P}$	$\frac{36}{37}$	$\frac{1}{37}$

$$\mathbb{E}X = -1 \times \frac{36}{37} + 35 \times \frac{1}{37} = -\frac{1}{37}$$

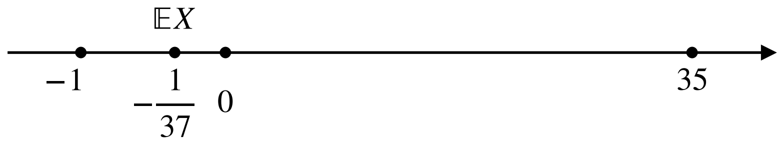
За по-консервативния играч Y, залагащ само на червено, ще имаме:

Y	-1	1
$\mathbb{P}$	$\frac{19}{37}$	$\frac{18}{37}$

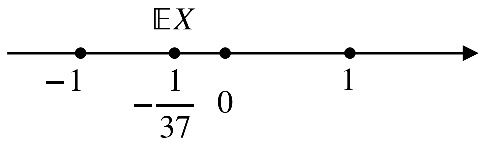
$$\mathbb{E}Y = -1 \times \frac{19}{37} + 1 \times \frac{18}{37} = -\frac{1}{37}$$

Оказва се, че $\mathbb{E}X = \mathbb{E}Y = -\frac{1}{37}$, т.е. и двамата играча губят средно по 1лв. на всеки 37 изиграни от тях игри или средно по ≈ 3ст. на всяка игра.

Дългосрочно, и двамата играча ще имат средно една и съща печалба или загуба, но игрите се различават доста по своя характер. Играча X ще печели много повече (с много по-голямо отклонение от средното) при печалба, но много по-рядко. Докато играча Y ще печели много по-често, но с много по-малко отклонение от средното.



X (само на едно число) :



Y (на 18 от 37 числа) :

При играча X волатилността е много по-голяма от тази при играча Y.

Извода е, че очакването не може да различи двете игри. Необходим е друг инструмент, който да може да направи тази разлика, да я посочи и да я измери количествено.

Дисперсия

<u>Дефиниция</u>: (**Дисперсия**) Ако сумата $\sum_j (X_j - \mathbb{E}X)^2 p_j$ е добре дефинирана, то $DX = \sum_j (x_j - \mathbb{E}X)^2 p_j$ е дисперсията на X.

За рисковия играч (този, който залага само на едно единствено число):

$$DX = \left(-1 + \frac{1}{37}\right)^2 \frac{36}{37} + \left(35 - \frac{1}{37}\right)^2 \frac{1}{37} \approx 33.97;$$

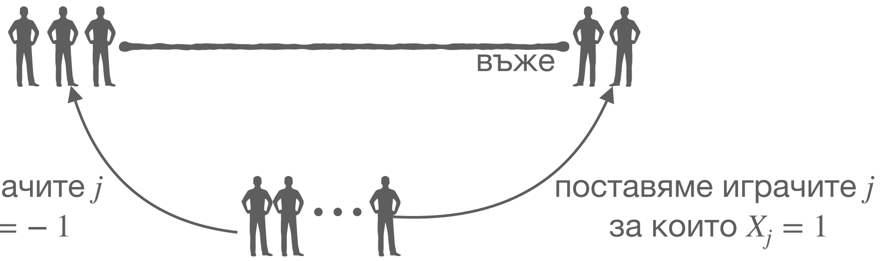
За консервативния играч (този който залага на 18 числа едновременно):

$$DX = \left(-1 + \frac{1}{37}\right)^2 \frac{19}{37} + \left(1 + \frac{1}{37}\right)^2 \frac{18}{37} \approx 0.999.$$

<u>Дефиниция</u>: (**Стандартно отклонение**) Ако X е случайна величина и DX е крайно, то $\sqrt{DX}$ се нарича стандартно отклонение на X.

$\oplus$ Имаме $2n$ играча и въже, което се дърпа от двата края. Ще нареждаме играчите от ляво и от дясно с вероятност $\frac{1}{2}$ и тази страна, от към която има повече играчи ще се измества от към техния край и ще спечели (допускаме, че всеки играч дърпа с еднаква сила).

Нека дефинираме случайна величина $X_i = \begin{cases} 1, & \text{с вероятност } p = \frac{1}{2} \\ -1, & \text{с вероятност } 1 - p = \frac{1}{2} \end{cases}$, където ако се падне 1, i-тия играч се праща от дясно, иначе от ляво (ако се падне -1).



$$Y = \sum_{j=1}^{2n} X_i = s \begin{cases} > 0 \Rightarrow \text{печели дясната страна} \\ < 0 \Rightarrow \text{печели лявата страна} \\ = 0 \Rightarrow \text{равенство между двете страни} \end{cases}$$

$$\mathbb{E}Y = \mathbb{E}\sum_{i=1}^{2n} X_i + \sum_{i=1}^{2n} \mathbb{E}X_i = \sum_{i=1}^{2n} \left(\frac{1}{2} \times 1 - \frac{1}{2} \times 1 \right) = \sum_{i=1}^{2n} 0 = 0.$$

$$\mathbb{P}(Y) = \frac{1}{2^n}\binom{2n}{n} \overset{\text{асимптотично}}{=} O\left(\frac{1}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} 0.$$

Последния ред показва, че въпреки очакването да настъпи равенство, ние почти никога няма да го имаме. Тази игра може да се разглежда като опростен аналог на брауновото движение на частиците, на които деистват равни сили от всякъде и се очаква да са статични, но всъщност непрекъснато се движат.

<u>Твърдение</u>: $DX \overset{(1)}{=} \mathbb{E}(X - \mathbb{E}X)^2 \overset{(2)}{=} \mathbb{E}X^2 - (\mathbb{E}X)^2$.

<u>Доказателство</u>:

$g(x) = (x - \mathbb{E}X)^2$, то

$\mathbb{E}g(X) = \mathbb{E}(X - \mathbb{E}X)^2 \overset{def.}{=} \sum_j (x_j - \mathbb{E}X)p_j \overset{def.}{=} DX \Rightarrow (1)$ е доказано;

$$DX = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}\left(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2\right) =$$

$$= \mathbb{E}X^2 + E\left(\underbrace{2}_{\text{число}} \times \underbrace{X}_{\text{сл. вел.}} \times \underbrace{\mathbb{E}X}_{\text{число}} \right) + \mathbb{E}(\underbrace{\mathbb{E}X}_{\text{число}})^2 =$$

$= \mathbb{E}X^2 - 2\mathbb{E}X\mathbb{E}X + (\mathbb{E}X)^2\mathbb{E}1 = \mathbb{E}X^2 - (\mathbb{E}X)^2 \Rightarrow$ и (2) е доказано.

Свойства на дисперсията:

а) $DX \geq 0 \Leftrightarrow DX = \mathbb{E}(X - \mathbb{E}X)^2 \geq 0$, понеже $(X - \mathbb{E}X)^2 \geq 0$ и очакването на тази случайна величина е неотрицателно;

б) $\mathbb{E}X^2 \geq (\mathbb{E}X)^2 \Leftarrow 0 \leq DX = \mathbb{E}X^2 - (\mathbb{E}X)^2$;

в) Ако $X = c$ е константа, тогава $DX = 0 \Leftarrow DX = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(c - c) = 0$;

г) $D(cX) = c^2 DX \Leftarrow D(cX) = \mathbb{E}(cX)^2 - (\mathbb{E}cX)^2 = c^2\mathbb{E}X^2 - c^2(\mathbb{E}X)^2 = c^2 DX$;

д) Ако X и Y са независими ($X \perp\!\!\!\perp Y$), дискретни, случайни величини, то $D(X+Y) = DX + DY$.

Доказателство:

Нека $Z = X + Y$ е нова случайна величина. Тогава

$$DZ \overset{def.}{=} \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathbb{E}\left(X + Y - \mathbb{E}(X+Y)\right)^2 = \mathbb{E}\left(X - \mathbb{E}X + Y - \mathbb{E}Y\right)^2 =$$

$$= \mathbb{E}\left((X - \mathbb{E}X) + (Y - \mathbb{E}Y)^2 + 2(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) \overset{\text{линейност на } \mathbb{E}}{=}$$

$$= \mathbb{E}(X - \mathbb{E}X)^2 + \mathbb{E}(Y - \mathbb{E}Y)^2 + 2\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) =$$

$$= DX + DY + 2\mathbb{E}\left(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y\right) =$$

$$= DX + DY + 2\mathbb{E}XY - 2\mathbb{E}X\mathbb{E}Y - \cancel{2\mathbb{E}Y\mathbb{E}X} + \cancel{2\mathbb{E}X\mathbb{E}Y} =$$

$$= DX + DY - 2\mathbb{E}Y\mathbb{E}X + 2\mathbb{E}XY \overset{\text{независимост}}{=} DX + DY - 2\mathbb{E}Y\mathbb{E}X + \cancel{2\mathbb{E}X\mathbb{E}Y} = DX + DY$$

Пораждаща функция

В тази тема X винаги ще е целочислена и неотрицателна, дискретна случайна величина, т.е. $X \in \mathbb{N}_0^+$.

Пораждащата функция прилича на торбичка, в която сме събрали всички индивидуални вероятности $X = k$, както и друга информация, като сме я компресирали по начин, по който да може да изваждаме само това което ни трябва - на цената на някакви операции.

Дефиниция: (**Пораждаща функция**) Нека $X \in \mathbb{N}_0^+$ е случайна величина. Тогава функцията $g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$, за $|s| < 1$, се нарича пораждаща функция на X.

⊕

X	0	1
$\mathbb{P}$	$1-p$	p

$$g_X(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X=k) = s^0(1-p) + s^1(p) =$$
$$= \underbrace{1-p+ps}_{\substack{\text{полином от първа} \\ \text{степен на } s}}$$

⊕

X	1	2	$\cdots$	n
$\mathbb{P}$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$

$$g_X(s) = \sum_{k=1}^{\infty} s^k \mathbb{P}(X=k) =$$
$$= s^1 \times \frac{1}{n} + s^2 \times \frac{1}{n} + \ldots + s^n \times \frac{1}{n} =$$
$$= \frac{1}{n}\left(s^1 + \ldots + s^n\right) \overset{|s|<1}{=} \frac{s(1-s)^n}{n(1-s)}$$

Свойства на пораждащата функция:

а) $g_X^{(n)}(0) = n!\mathbb{P}(X=n) \Leftarrow g_X^{(n)}(s) = \sum_{k=0}^{\infty} (s^k)^{(n)} \mathbb{P}(X=k) =$

$= \sum_{k=n}^{\infty} (s^k)^{(n)} \mathbb{P}(X=k) \overset{s=0}{=} \underbrace{(s^n)^{(n)} \mathbb{P}(X=n)}_{n!\mathbb{P}(X=n)}$. Следователно $\mathbb{P}(X=n) = \dfrac{g_X^{(n)}(0)}{n!}$;

б) $g'_X(1) = \mathbb{E}X \Leftarrow \dfrac{\partial}{\partial s} g_X(s) = \dfrac{\partial}{\partial s}\mathbb{E}s^X = \mathbb{E}\dfrac{\partial}{\partial s}s^X = \mathbb{E}Xs^{X-1}\Big|_{s=1} = \mathbb{E}X$;

в) $DX = g''_X(1) + g'_X(1) - \left(g'_X(1)\right)^2$

<u>Доказателство</u>:

$$g'_X(1) = \frac{\partial}{\partial s}\sum_k s^k \mathbb{P}(X=k) = \sum_k k s^{k-1}\mathbb{P}(X=k)\Big|_{s=1} = \underbrace{\sum_k k\mathbb{P}(X=k)}_{=\mathbb{E}X};$$

$$g''_X(1) = \frac{\partial}{\partial s}\sum_k k s^{k-1} \mathbb{P}(X=k) = \sum_k k(k-1) s^{k-2}\mathbb{P}(X=k)\Big|_{s=1} = \sum_k k(k-1)\mathbb{P}(X=k)$$

$g''_X(1) + g'_X(1) = \sum_k k^2 \mathbb{P}(X=k) \overset{def.}{=} \mathbb{E}X^2$. От друга страна: $\left(g'_X(1)\right)^2 = (\mathbb{E}X)^2$.

Следователно: $g''_X(1) + g'_X(1) - \left(g'_X(1)\right)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 \overset{def.}{=} DX$, което искахме да докажем.

г) <u>Твърдение</u>: Нека $X \perp\!\!\!\perp Y$ и $X \in \mathbb{N}_0^+$ и $Y \in \mathbb{N}_0^+$. Тогава $g_{X+Y}(s) = g_X(s)g_X(s)$.

<u>Доказателство</u>:

$$g_{X+Y}(s) = \mathbb{E}s^{X+Y} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s^{j+i}\mathbb{P}(X = j \cap Y = i) =$$

$$= \sum_{i=0}^{\infty} s^i\mathbb{P}(Y = i)\sum_{j=0}^{\infty} s^j\mathbb{P}(X = j) = \mathbb{E}s^Y\mathbb{E}s^X = g_X(s)g_Y(s).$$

CEM, лекция 6

(2020-11-05)

Дефиниция: (**Пораждаща функция / преговор**) Ако $X \in \mathbb{N}_0^+$, то
$g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$, $|s| < 1$ се нарича пораждаща функция.

$\oplus$ $\quad \mathbb{P}(X = 1) = p$ и $\mathbb{P}(X = 0) = 1 - p \Rightarrow g_X(s) = 1 - p + ps$.

Пораждащата функция е като една торбичка, която носи много обекти и може да ги вадим един по един чрез различни операции, като например:

$$\begin{cases} \mathbb{E}X = g_X'(1) \\ DX = g_X''(1) + g_X'(1) - \left(g_X'(1)\right)^2 \\ \mathbb{P}(X = k) = \frac{g_X^{(k)}(0)}{k!} \end{cases}$$

Дефиниция: (**Независимост на (дискретни) случайни величини**) Случайните величини $X_1, X_2, \ldots, X_N$ са независими в съвкупност, ако $\forall\ 1 \leq m \leq N$ и $\{j_1, j_2, \ldots, j_m\} \subseteq \{1, 2, \ldots, N\}$ и възможни стойности $(x_1, x_2, \ldots, x_m)$ е вярно, че $\mathbb{P}(X_{j_1} = x_1 \cap X_{j_2} = x_2 \cap \ldots \cap X_{j_m} = x_m) = \prod_{i=1}^{m} \mathbb{P}(X_{j_i} = x_i)$.
Вероятността да се сбъдне кой да е подбор от случайни величини се разпада на произведението на индивидуалните вероятности.

Твърдение: Ако $X_1, \ldots, X_n$ са независими в съвкупност, целочислени случайни величини и $Y = \sum_{j=1}^{n} X_j$, то $g_Y(s) = \prod_{j=1}^{n} g_{X_j}(s)$.

Логика на доказателството: $g_Y(s) = \mathbb{E}s^{\sum_{j=1}^n X_j} = \mathbb{E}\prod_{j=1}^{n} s^{X_j}$.

Коментари:

$\mathbb{E}\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} \mathbb{E}X_j$, винаги (когато очакванията са добре дефинирани)

$D\sum_{j=1}^{n} X_j = \sum_{j=1}^{n} DX$, ако $X_1, X_2, \ldots, X_n$ са независими в съвкупност.

Някои целочислени случайни величини

$X_1, X_2, \ldots, X_m, \ldots$ - взаимно независими (дискретни) сл. вел.

X_i	0	1
$\mathbb{P}$	q	p

$i \geq 1$
$p + q = 1$

p и q са фиксирани $\forall i$ (не зависят от i). Все едно имаме n експеримента, които са еднородни по своя характер (хвърляне на монета).

Тази постановка се нарича схема на Бернули.

А. Разпределение на Бернули

$X \in Ber(p)$, ако имаме разпределението

X_i	0	1
$\mathbb{P}$	q	p

$p + q = 1$

$$\begin{cases} \mathbb{E}X = 0.q + p.1 = p \\ \mathbb{E}X^2 = 0^2\,.\,q + 1^2\,.\,p = p \end{cases} \Rightarrow DX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1-p) = pq$$

$$g_X(s) = \mathbb{E}s^X = s^0 q + s^1 p = q + ps = 1 - p + ps$$

Б. Биномно разпределение

Взимаме само първите n случайни величини от схемата на Бернули: $X_1, X_2, \ldots, X_n$ и образуваме $X = \sum_{i=1}^{n} X_i$, което ни отразява броя успехи измежду n експеримента.

$X \in Bin(n, p)$ се нарича биномно разпределена (дискретна) случайна величина с параметри n и p.

<u>Твърдение</u>: Нека X е биномно разпределена сл. вел. с параметри n и p. Тогава:

а) $g_X(s) = (q + ps)^n$

б) $\mathbb{E}X = np$, $DX = npq$

в)

X	0	1	$\ldots$	k	$\ldots$	n
$\mathbb{P}$	q^n	npq^{n-1}	$\ldots$	$\binom{n}{k} p^k q^{n-k}$	$\ldots$	p^n

k успеха измежду n експеримента
всеки един измежду които се случва
независимо от останалите с вероятност $\underbrace{p}_{\text{за успех}}$

<u>Доказателство</u>:

а) $X = \sum_{j}^{n} X_j$ (сума на независими бернулиеви случайни величини) $\Rightarrow$

$$g_X(s) \overset{\text{твърдение}}{=} \prod_{j=1}^{n} g_{X_j}(s) = \prod_{j=1}^{n} (q + ps) = (q + ps)^n \,;$$

б) $$\mathbb{E}X = \mathbb{E}\sum_{j=1}^{n} X_j \overset{\text{лин. функц.}}{=} \sum_{j=1}^{n} \mathbb{E}X_j = \sum_{j=1}^{n} p = np \,;$$

$$DX = D\sum_{j=1}^{n} X_j \overset{\text{незав.}}{=} \sum_{j=1}^{n} DX_j = \sum_{j=1}^{n} pq = npq \,;$$

в) Комбинаторно избираме k експеримента от общо n, които да са успешни по $\binom{n}{k}$ начина и умножаваме по вероятността за успех p - k пъти и k пъти по вероятността за неуспех $\Rightarrow \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$.

Но може да сведем комбинаторния проблем до чисто аналитичен и да го получим по следния начин:

$k!\mathbb{P}(X = k) = g_X^{(k)}(0),\ 0 \le k \le n.$

Но, $g_X^{(k)}(0) = \dfrac{\partial^k}{\partial s^k}(q + ps)^n \Big|_{s=0} = n(n-1)\ldots(n-k+1)q^{n-k}p^k \Rightarrow$

$$\Rightarrow \mathbb{P}(X = k) = \frac{n(n-1)\ldots(n-k+1)q^{n-k}p^k}{k!} = \frac{n!}{(n-k)!k!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}$$

$\oplus$ Колко заразени биха влезли в България от чужбина, ако вероятността човек да е заразен е $p = 1\,\%$ и са се върнали $100\,000$ души. Тогава, не лош модел би бил случайната величина $X \sim Bin(n = 100\,000,\ p = 0.01)$, която брой колко от върналите се в България са заразени.

B. Геометрично разпределение

Тук може да си мислим за горния пример, но по следния начин: Колко незаразени хора ще влязат в България, преди да влезе първия заразен, върнал се от чужбина?

<u>Дефиниция</u>: $X \in Ge(p)$ и $X = \min\left\{j \ge 1 \,\Big|\, \sum_{i=1}^{j} X_i = 1\right\} - 1$. Т.е. най-малкото j, за което сумата $\sum_{i=1}^{j} X_i$ става единица, като от нея вадим 1-ца, за да извадим последната стъпка (успех/заразен посетител). Това е броя неуспехи до първи успех (в нашия случай с „успех“ контраинтуитивно сме маркирали заразен човек).

$\underbrace{0\ 0\ 0}_{\text{неуспехи}}\ \underbrace{1}_{\text{успех}} \Rightarrow X = 3$

$\underbrace{0}_{\text{неуспех}}\ \underbrace{1}_{\text{успех}} \Rightarrow X = 1$

$\underbrace{1}_{\text{успех}} \Rightarrow X = 0$

Твърдение: $X \in Ge(p)$. Тогава:

а)

X	0	1	2	$\dots$	k	$\dots$
$\mathbb{P}$	p	pq	pq^2	$\dots$	pq^k	$\dots$

$\Rightarrow \mathbb{P}(X = k) = pq^k$

б) $g_X(s) = \dfrac{p}{1 - qs}, \ |s| < 1$

Доказателство:

а)

$\mathbb{P}(X = 0) = \mathbb{P}(X_1 = 1) = p$ $\qquad 1$

$\mathbb{P}(X = 1) = \mathbb{P}(X_1 = 0;\ X_2 = 1) = pq$ $\qquad 01$

$\dots$ $\qquad \dots$

$\mathbb{P}(X = k) = \mathbb{P}(X_1 = 0;\ X_2 = 0;\ \dots X_k = 0;\ X_{k+1}) = pq^k$ $\qquad \underbrace{00...0}_{k}1$

$\overset{\text{твърдение}}{\Rightarrow}$ а) е доказано.

б) $$g_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k pq^k = p \sum_{k=0}^{\infty} s^k q^k \overset{|qs|<1}{=} \frac{p}{1 - qs}.$$

Следствие: $X \sim Ge(p) \Rightarrow \mathbb{E}X = \dfrac{q}{p}$ и $DX = \dfrac{q}{p^2}$

Доказателство: $g_X(s) = \dfrac{p}{1 - qs}$

$$\mathbb{E}X = g_X'(1) = p\frac{\partial}{\partial s}\left(\frac{1}{1 - qs}\right) = p(-1)\frac{\frac{\partial}{\partial s}(1 - qs)}{(1 - qs)^2}\Bigg|_{s=1} = \frac{pq}{(1 - q)^2} = \frac{pq}{p^2} = \frac{q}{p};$$

$$g_X''(1) = \frac{\partial}{\partial s}\left(\frac{pq}{(1 - qs)^2}\right) = pq\frac{\partial}{\partial s}\left(\frac{1}{(1 - qs)^2}\right) = pq(-2)\frac{\frac{\partial}{\partial s}(1 - qs)}{(1 - qs)^3} =$$

$$= \frac{2pq^2}{(1-qs)^3}\bigg|_{s=1} = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2};$$

$$DX = g''_X(1) + g'_X(1) - \left(g'_X(1)\right)^2 = \frac{\cancel{2}q^2}{p^2} + \frac{q}{p} - \cancel{\left(\frac{q}{p}\right)^2} = \frac{q^2}{p^2} + \frac{q}{p} =$$

$$= \frac{q^2+qp}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2}$$

⊕ Брой „тури“ при хвърляне на монета до първо „ези“ $\sim Ge\left(p = \frac{1}{2}\right)$;

⊕ Брой неуспешни атаки (в баскетбол) до първи кош. Ако грубо казано не се променят играчите и имат вероятност за успех, например $80\,\%$ (за кош), то броя на неуспешните атаки до първия успех може да се моделира добре като се допусне, че е разпределен геометрично с параметър $p = 0.8$.

Твърдение: (**Безпаметност на геометричното разпределение**) Нека $X \in Ge(p)$. Тогава $\forall n \geq 0$ и $k \geq 0$: $\mathbb{P}(X \geq m+k \,|\, X \geq m) = \mathbb{P}(X \geq k) = q^k$.

Тоест, ако сме наблюдавали нашия процесор да работи 20 месеца, то вероятността да работи следващия месец е грубо казано същата, като вероятността да работи в момента, в който сме го стартирали в първия месец до края на първия месец.

Доказателство:

$\mathbb{P}(X \geq l) = \sum_{j=l}^{\infty} pq^j = pq^l \sum_{j=0}^{\infty} q^j = pq^l \frac{1}{1-q} = q^l$. Тогава,

$$\mathbb{P}(X \geq m+k \,|\, X \geq m) = \frac{\mathbb{P}(X \geq m+k \cap X \geq m)}{\mathbb{P}(X \geq m)} = \frac{\mathbb{P}(X \geq m+k)}{\mathbb{P}(X \geq m)} =$$

$$= \frac{q^{m+k}}{q^m} = q^k = \mathbb{P}(X \geq k).$$

Г. Отрицателно биномно разпределение

Това разпределение се конструира от няколко геометрично разпределени случайни величини. Т.е. тук ще моделираме броя на неуспехите, но не до първия успех, а до r-тия успех.

Дефиниция: $X \in NB(r, p)$ и $X = \min\left\{ j \geq 1 \,\Big|\, \sum_{i=1}^{j} X_j = r \right\} - r$.

вероятност за успех
в схемата на бернули

Това е броя неуспехи до r-тия успех.

$r=2$ $\qquad \underbrace{0\,0\,0}\,1\,\underbrace{0\,0\,0\,0\,0}\,1 \qquad X=8$

общо 8 неуспеха

$r=4$ $\qquad 1\,1\,\underbrace{0}\,1\,1 \qquad X=1$

$r=1$ $\qquad \underbrace{0\,0\,0}\,1 \qquad X=3$

$\oplus$ За $r=1 \to NB(1,p)=Ge(p)$

Твърдение: Ако $X \in NB(r,p)$, то $X=\sum_{j=1}^{r} Y_j$, където Y_j са геометрични с вероятност p $\left(Y_j \in Ge(p)\right)$, за $1 \le j \le r$ и Y_j са независими в съвкупност.

За $r=2 : X \in NB(2,p)$ $\qquad \underbrace{0\,0\,0}\,1\,\underbrace{0\,0}\,1$

$X \stackrel{?}{=} \sum_{j=1}^{2} Y_j$, където $Y_j \in Ge(p)$ и Y_j са независими в съвкупност.

Ще проверим, че $Y_1 \perp\!\!\!\perp Y_2$ $\big($Ясно е, че $X=Y_1+Y_2$ и $Y_1 \in Ge(p)$ и $Y_2 \in Ge(p)\big)$

$\mathbb{P}(Y_1=l \cap Y_2=m) \stackrel{?}{=} \mathbb{P}(Y_1=l)\mathbb{P}(Y_2=m), \forall\, l,m \ge 0.$

$\mathbb{P}(Y_1=l \cap Y_2=m)=$

$$=\mathbb{P}(\underbrace{X_1,0,\ldots,X_l=0,\,X_{l+1}=1,\,X_{l+2}=0,\ldots,X_{l+m}=0,\,X_{l+m+1}=1}_{\text{независими експерименти от схемата на Бернули}})=$$

$$=\underbrace{\mathbb{P}(X_1=0)\ldots\mathbb{P}(X_2=0)\mathbb{P}(X_{l+1}=1)}\,\underbrace{\mathbb{P}(X_{l+2}=0)\ldots\mathbb{P}(X_{l+m}=0)\mathbb{P}(X_{l+m+1}=1)}=$$

$=\mathbb{P}(Y_1=l)\mathbb{P}(Y_2=m) \Rightarrow$ независими.

Твърдение: Ако $X \sim NB(r,p)$, то $g_X(s)=\left(\dfrac{p}{1-qs}\right)^r$, $\mathbb{E}X=\dfrac{rq}{p}$ и $DX=\dfrac{rq}{p^2}$.

Доказателство:

$$\mathbb{E}X = \mathbb{E}\sum_{j=1}^{r} \underbrace{Y_j}_{\substack{\text{геом.}\\ \text{незав.}}} = \sum_{j=1}^{r}\mathbb{E}Y_j = r\frac{q}{p};$$

$$DX = D\sum_{j=1}^{r} Y_j \overset{\text{незав.}}{=} \sum_{j=1}^{r} DY_j = r\frac{q}{p^2};$$

$$g_X(s) \overset{\text{твърдение}}{=} \prod_{j=1}^{r} g_{Y_j}(s) = \prod_{j=1}^{r}\frac{p}{1-qs} = \left(\frac{p}{1-qs}\right)^r.$$

Твърдение: $X \sim NB(r,p)$. Тогава:

X	0	1	$\dots$	k	$\dots$
$\mathbb{P}$				$\binom{r+k-1}{k}p^rq^k$	

Аналитичен подход:

От пораждащата функция знаем, че $g_X(s) = \left(\dfrac{p}{1-qs}\right)^r$, но освен това знаем, че

$$k!\mathbb{P}(X=k) = g_X^{(k)}(s)\Big|_{s=0}$$

$$g_X(s) = \frac{p^r}{(1-qs)^r}, \text{ но } \frac{1}{(1-x)^r} \underset{|x|<1}{\overset{\substack{\text{ред на}\\ \text{Тейлър}}}{=}} \sum_{k=0}^{\infty}\binom{r+k-1}{k}x^k$$

$$\frac{\partial^n}{\partial x^n}\frac{1}{(1-x)^r}\Bigg|_{x=0} = r(r+1)\dots(r+1-n)\frac{1}{(1-x)^{r+n}}\Bigg|_{x=0} = r(r+1)\dots(r-n+1)$$

$$\Rightarrow g_X(s) = p^r\sum_{k=0}^{\infty}\binom{r+k-1}{k}\underbrace{q^ks^k}_{x^k}$$

$$g_X^{(k)}(0)\Big|_{s=0} = p^r\binom{r+k-1}{k}q^kk! = k!\mathbb{P}(X=k)$$

$$\Rightarrow \mathbb{P}(X=k) = \binom{r+k-1}{k}p^rq^k.$$

Комбинаторен подход:

k нули и $r-1$ единици трябва да се поставят на $r+k-1$ позиции след което да се последват от 1-ца:

$$\underbrace{\binom{r+k-1}{k}}_{\text{нули}} \underbrace{q^k}_{\text{вер.}} \; p^{r-1}p$$

Д. Поасоново разпределение

Дефиниция: Нека $\lambda > 0$. Казваме, че X е поасоново разпределена случайна величина с параметър λ и бележим с $X \sim Pois(\lambda)$, ако $\mathbb{P}(X=k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$, $k \geq 0$

$$1 \stackrel{?}{=} \sum_{k=0}^{\infty} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda} \times \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\substack{\text{развитие в ред} \\ \text{на Тейлър} \\ \text{за exp}}} = e^{-\lambda}e^{\lambda} = 1$$

Какво може да се моделира добре с поасоновото разпределение на случайна величина:

- ⊕ Брой постъпили заявки към сървър за единица време ;
- ⊕ Брой катастрофи за 1-ца време на дадена територия/кръстовище ;
- ⊕ Брой насекоми за единица площ ;
- ⊕ Брой голове за 1-ца време ;
- ⊕ Брой звезди в някакъв регион.

Твърдение:

$X \in Pois(\lambda)$. Тогава:

а) $g_X(s) = e^{-\lambda+\lambda s}$, за $|s| \leq 1$

б) $\mathbb{E}X = DX = \lambda$

Доказателство:

а) $$g_X(s) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda}e^{\lambda s};$$

б) $\mathbb{E}X = g'_X(1) = \lambda e^{-\lambda} e^{\lambda s}\Big|_{s=1} = \lambda$

$$g''_X(1) = \frac{\partial}{\partial s}\left(\lambda e^{-\lambda} e^{\lambda s}\right) = \lambda^2 e^{-\lambda} e^{\lambda s}\Big|_{s=1} = \lambda^2$$

$$DX = g'_X(1) + g''_X(1) - \left(g'_X(1)\right)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = \mathbb{E}X.$$

<u>Теорема</u>: (**Поасон**) Нека $\forall\, n \geq 1 : X_n \sim Bin(n, p_n)$, където $p_n = \dfrac{\lambda}{n} + \dfrac{V_n}{n}$, където $\lambda > 0$ и $\lim\limits_{n\to\infty} V_n = 0$.

Т.е. $P_N = \dfrac{\lambda}{n} + O\left(\dfrac{1}{n}\right)$. Тогава $\forall\, k \geq 0 : \lim\limits_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$ или казано по друг начин, $X \sim Pois(\lambda)$.

$\oplus$ $p = 0{,}01$ - вероятност на зараза и $n = 1000$.

$$\mathbb{P}(X = 50) = \binom{1000}{50} \times 0{,}01^{50} \times 0{,}99^{950} \underset{\lambda = np = 10}{\overset{Pois}{\approx}} \frac{10^{50}}{50!}e^{-10}$$

$$\mathbb{P}(X = 0) \approx \frac{10^0}{0!}e^{-10} = e^{-10}.$$

СЕМ, лекция 7

(2020-11-12)

Теорема (**Поасон / преговор**) Нека $X_n \in Bin(n, p)$, където $p_n = \frac{\lambda}{n} + \frac{V_n}{n}$, $\lim\limits_{n\to\infty} V_n = 0$ $\left(\text{т.е. } \frac{V_n}{n} \text{ клони по-бързо от } \frac{\lambda}{n} \text{ към } 0\right)$. Тогава за $\forall k \geq 0$ е вярно, че $\lim\limits_{n\to\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \mathbb{P}(X = k)$, където $X \in Pois(\lambda)$.

$\oplus$ При $n \geq 100$ и $np \leq 20$ може да считаме, че $\mathbb{P}(X_n = k) \approx \mathbb{P}(X = k)$, където $X \in Pois(\lambda)$ и $\lambda = np$.

Нека например $n = 1000$ души се завръщат в България за някакъв период от време и вероятността за всеки един от тях да внесе някакъв вирус е $p = 0.001$. Искаме да приближим относително добре вероятността да влязат точно k на брой заразени.

$$\tilde{X} \in Bin\left(n = 1000, p = \frac{1}{1000}\right)$$

$$\mathbb{P}(\tilde{X} = 3) = \binom{1000}{3}\left(\frac{1}{1000}\right)^3\left(\frac{999}{1000}\right)^{997}$$

$$\lambda = np = 1 \leq 20, n = 1000 \geq 100 \Rightarrow \mathbb{P}(\tilde{X} = 3) \approx \frac{1}{3!}.e^{-1} = \frac{1}{6e} \approx \frac{1}{6}.\frac{1}{2.7182...}$$

Доказателство: $X_n \in Bin(n, p_n)$, тогава $g_{X_n} = (1 - p_n + p_n s)^n$. Ако $g_{X_n}(s) \underset{n\to\infty}{\longrightarrow} g_X(s)$, където $g_X(s) = e^{-\lambda}.e^{\lambda s}$, то може да заключим искания резултат, тъй като $e^{-\lambda}.e^{\lambda s}$ е пораждаща функция на $X \in Pois(\lambda)$.

От $g_{X_n}(s) \underset{x\to\infty}{\longrightarrow} g_X(s) \Rightarrow \lim\limits_{n\to\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $\forall k \geq 0$

$$\lim_{n\to\infty} g_{X_n}(s) = \lim_{n\to\infty}\left(1 - \frac{\lambda}{n} - \underbrace{\frac{V_n}{n} + \frac{\lambda}{n}s + \frac{V_n}{n}s}_{<o\left(\frac{1}{n}\right)}\right)^n = \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}(1-s)\right)^n = e^{-\lambda(1-s)} = e^{\lambda s}e^{-\lambda}$$

Е. Хипергеометрично разпределение

Имаме N обекта, от които M са маркирани $(0 \leq M \leq N)$. Избират се n обекта и случайната величина X е броя маркирани измежду тези N $(n \leq N)$. Тогава казваме, че X е разпределено хипергеометрично с параметри N, M и n и бележим $X \in HG(N, M, n)$.

<u>Твърдение</u>: Нека $X \in HG(N, M, n)$. Тогава:

а) $\mathbb{P}(X = k) = \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, като $\max\left(0, n - (N - M)\right) \leq k \leq \min(n, M)$

б) $\mathbb{E}X = n \, . \, \dfrac{M}{N}$, $DX = n \, . \, \dfrac{M}{N} \, . \, \dfrac{N-M}{N} \, . \, \dfrac{N-n}{N-1}$.

<u>Доказателство</u>:

а) $\mathbb{P}(X = k) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$.

б)

$$\underbrace{\underbrace{\_\ \_\ \_\ \_\ \overset{i}{\_}\ \_\ \_}_{n}\ \_\ \_\ \_\ \_\ \_\ \_}_{N}$$

$X = \sum_{i=1}^{n} X_i$, където $X_i = \begin{cases} 1, & \text{ако на } i\text{-тата позиция } \textbf{има} \text{ маркиран обект} \\ 0, & \text{ако на } i\text{-тата позиция } \textbf{няма} \text{ маркиран обект} \end{cases}$

X_i е бернулиево разпределено: $X_i \in Ber(p_i)$. $p_i = \mathbb{P}(X_i = 1) = \dfrac{\binom{N-1}{M-1}}{\binom{N}{M}} = \dfrac{M}{N}$.

$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}X_i = n \, . \, \dfrac{M}{N}$.

$\oplus$ $X_i \in Ber(p)$ - поведението на i-тия клиент в даден магазин (купува или не)

$X = \sum_{i=1}^{n} X_i$, $X \in Bin(n, p)$ - броя на извършилите покупка клиенти от първите n клиента.

$Y = \min\left\{ \sum_{j=1}^{n} X_j = 1 \right\} - 1 \in Ge(p)$ - броя на не извършили покупка клиенти, до идването на първия извършил покупка клиент.

$Z = \min\left\{\sum_{j=1}^{n} X_j = r\right\} - r \in NB(r,p)$ - броя на не извършили покупка клиенти, до идването на r-тия извършил покупка клиент.

Съвместно разпределение на дискретни случайни величини

Нека X, Y са дискретни случайни величини. Интересуваме се от (X, Y).
Дефиниция. Нека (X, Y) са две дискретни случайни величини. Тогава таблицата по-долу се нарича съвместно разпределение на X и Y:

$Y \backslash X$	x_1	x_2	$\cdots$	x_n	$\cdots$	
y_1	p_{11}	p_{21}	$\cdots$	p_{n1}	$\cdots$	$\sum_i p_{i1}$
y_2	p_{12}	p_{22}	$\cdots$	p_{n2}	$\cdots$	$\sum_i p_{i2}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
y_k	p_{1k}	p_{2k}	$\cdots$	p_{nk}	$\cdots$	$\sum_i p_{ik}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
	$\sum_j p_{1j}$	$\sum_j p_{2j}$		$\sum_j p_{nj}$		

Където $0 \leq p_{ij} = \mathbb{P}(X = x_i;\ Y = y_j)$, за $\forall i, j \in$ Table Indexes. $\sum_{i,j} p_{ij} = 1$.

Маргинално разпределение на X:

X	x_1	x_2	$\cdots$	x_k	$\cdots$
	$\sum_j p_{1j}$	$\sum_j p_{2j}$	$\ldots$	$\sum_j p_{nj}$	$\ldots$

Разпределенията само на X или само на Y се наричат маргинални разпределения.

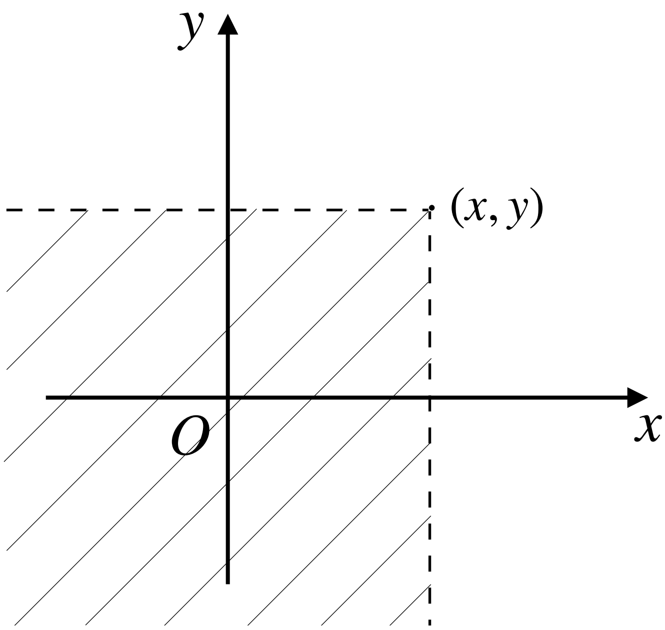
$\oplus$ Хвърляме два зара $(1,...,6)$. Нека X е броят шестици, а Y е броят единици. Търси се (X, Y). За едно хвърляне може да имаме $0,1,2$ шестици или единици.

$Y \backslash X$	0	1	2	
0	$\left(\frac{4}{6}\right)^2$	$\binom{2}{1} \cdot \frac{1}{6} \cdot \frac{4}{6}$	$\left(\frac{1}{6}\right)^2$	$\frac{25}{36}$
1	$\binom{2}{1} \cdot \frac{1}{6} \cdot \frac{4}{6}$	$\binom{2}{1} \cdot \frac{1}{6} \cdot \frac{1}{6}$	0	$\frac{10}{36}$
2	$\left(\frac{1}{6}\right)^2$	0	0	$\frac{1}{36}$
	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	$1 \backslash 1$

Дефиниция: (**Функция на разпределение на случайни величини**) Нека (X, Y) се състои от *произволни* случайни величини. Тогава
$F_{XY}(x, y) = \mathbb{P}(X < x \cap Y < y) = \mathbb{P}(X < x;\ Y < y)$ дефинирана за $\forall (x, y) \in \mathbb{R}^2$ се нарича функция на разпределение.

$F_X(x) = F_{XY}(x, \infty)$

$F_Y = F_{XY}(\infty, y)$



Дефиниция: (**Независимост** $\perp\!\!\!\perp$) Произволни случайни величини X и Y са независими ($X \perp\!\!\!\perp Y$) $\Leftrightarrow$
$F_{XY}(x, y) = F_X(x)F_Y(y) = F_{XY}(x, \infty)F_{XY}(\infty, y), \forall (x, y) \in \mathbb{R}^2$.

Ковариация на X и Y

Линейна зависимост $Y = aX + b$.
Показател за това до колко имаме линейна зависимост между произволни случайни величини X и Y е ковариацията.

<u>Дефиниция</u>: (**Ковариация**) Нека X и Y са случайни величини с $DX < \infty$ и $DY < \infty$. Тогава $cov(X,Y) = \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right]$ се нарича ковариация.

<u>Твърдение</u>:
а) $cov(X,Y) = \mathbb{E}\tilde{X}\tilde{Y}$, където $\tilde{X} = X - \mathbb{E}X,\ \tilde{Y} = Y - \mathbb{E}Y$
б) $cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$

Доказателство:

а) $cov(X,Y) = \mathbb{E}\left[\underbrace{(X-\mathbb{E}X)}_{\tilde{X}}\underbrace{(Y-\mathbb{E}Y)}_{\tilde{Y}}\right] = \mathbb{E}\tilde{X}\tilde{Y}$.

б)

$$cov(X,Y) = \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right] = \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}XY) \overset{\substack{\text{лин.}\\ \text{функц.}\\ \mathbb{E}}}{=} \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$$

<u>Следствие</u>: Ако $X \perp\!\!\!\perp Y$, то $cov(X,Y) = 0$
<u>Доказателство</u>: Ако $X \perp\!\!\!\perp Y$, то $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y \Rightarrow \underbrace{\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y}_{cov(X,Y)} = 0$.

$X \longmapsto 10X,\ Y \longmapsto 10Y \ \Rightarrow \tilde{X} \longmapsto 10\tilde{X},\ \tilde{Y} \longmapsto 10\tilde{Y}$
$\Rightarrow cov(10X, 10Y) = 100 \times cov(X,Y)$.

<u>Дефиниция</u>: (**Корелация**) Нека X, Y са случайни величини и $DX < \infty$ и $DY < \infty$. Тогава $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$ се нарича коефициент на корелация между X и Y.

Целта на корелацията е да мери някаква степен на линейност между X и Y.

<u>Твърдение</u>: Нека X и Y са случайни величини с крайни дисперсии и нека $\overline{X} = \dfrac{X-\mathbb{E}X}{\sqrt{DX}}$ и $\overline{Y} = \dfrac{Y-\mathbb{E}Y}{\sqrt{DY}}$. Тогава $\rho(X,Y) = \mathbb{E}\overline{X}\,\overline{Y}$ и
$\mathbb{E}\overline{X} = \mathbb{E}\overline{Y} = 0,\ D\overline{X} = \mathbb{E}\overline{X}^2 = D\overline{Y} = \mathbb{E}\overline{Y}^2 = 1$.

<u>Доказателство</u>:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{DX}\sqrt{DY}} = \frac{\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right]}{\sqrt{DX}\sqrt{DY}} = \mathbb{E}\left[\underbrace{\left(\frac{X-\mathbb{E}X}{\sqrt{DX}}\right)}_{=\overline{X}}\underbrace{\left(\frac{Y-\mathbb{E}Y}{\sqrt{DY}}\right)}_{=\overline{Y}}\right] = \mathbb{E}\overline{X}\,\overline{Y}$$

$$\mathbb{E}\overline{X} = \mathbb{E}\frac{X-\mathbb{E}X}{\sqrt{DX}} = \frac{1}{\sqrt{DX}}\mathbb{E}(X-\mathbb{E}X) = 0$$

$D\overline{X} = D\dfrac{X - \mathbb{E}X}{\sqrt{DX}} = \dfrac{1}{(\sqrt{DX})^2}D(X - \underbrace{\mathbb{E}X}_{const.}) = \dfrac{DX}{DX} = 1$, тъй като $D(X - c) = DX$,

защото $\mathbb{E}\big(X - c - \mathbb{E}(X - c)\big)^2 = \mathbb{E}(X - \mathbb{E}X)^2$.

$D\overline{X}^2 = D\overline{X} + (\underbrace{\mathbb{E}\overline{X}}_{=0})^2 = D\overline{X} = 1.$

Твърдение: Нека X и Y са случайни величини с крайни дисперсии. Тогава:

а) $|\rho(X, Y)| \leq 1$

б) $|\rho(X, Y)| = 1 \Leftrightarrow \exists a, b : a, b \in \mathbb{R} \ \& \ Y = aX + b$

(Колкото е по-голям коефициента на корелация, толкова по-добра е линейната зависимост между X и Y)

Доказателство:

а) $(\overline{X} - \overline{Y})^2$ е неотрицателно случайна величина

$\Rightarrow 0 \leq \mathbb{E}(\overline{X} - \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 - 2\mathbb{E}\overline{X}\,\overline{Y} = 2\left(1 - \rho(X, Y)\right) \Rightarrow \rho(X, Y) \leq 1$

Аналогично,

$0 \leq \mathbb{E}(\overline{X} + \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 + 2\mathbb{E}\overline{X}\,\overline{Y} = 2\big(1 + \rho(X, Y)\big) \Rightarrow -1 \leq \rho(X, Y)$

б) $Y = aX + b \Leftrightarrow Y = \mathbb{E}X = a(X - \mathbb{E}X) + a\mathbb{E}X + b - \mathbb{E}Y \Leftrightarrow$

$\Leftrightarrow \dfrac{Y - \mathbb{E}Y}{\sqrt{DY}} = \underbrace{\dfrac{a}{\sqrt{DY}}\sqrt{DX}}_{v}\,\dfrac{X - \mathbb{E}X}{\sqrt{DX}} + \underbrace{\dfrac{a\mathbb{E}X + b - \mathbb{E}Y}{\sqrt{DY}}}_{w} \Leftrightarrow \overline{Y} = v\overline{X} + w$

$0 = \mathbb{E}\overline{Y} = \mathbb{E}(v\overline{X} + w) = v\underbrace{\mathbb{E}\overline{X}}_{=0} + w \Rightarrow w = 0 \Rightarrow \overline{Y} = v\overline{X}$

$D\overline{Y} = v^2 D\overline{X} \Rightarrow v^2 = 1$ или $v = \pm 1$.

$Y = aX + b \Leftrightarrow \overline{Y} = v\overline{X}$ и $v^2 = 1$

б) Нека $Y = aX + b$ или $\overline{Y} = v\overline{X}$. Тогава

$\rho(X, Y) \overset{\text{твърдение}}{=} \mathbb{E}\overline{X}\,\overline{Y} = v\mathbb{E}\overline{X}^2 = v \Rightarrow \rho(X, Y) = \pm 1.$

Нека $|\rho(X, Y)| = 1$. Тогава $|\mathbb{E}(\overline{X}\,\overline{Y})| = 1$. Да допуснем, че $\mathbb{E}\overline{X}\,\overline{Y} = 1$ (аналогично и за другия случай: $\mathbb{E}\overline{X}\,\overline{Y} = -1$). Тогава

$\mathbb{E}(\overline{X} - \overline{Y})^2 = \mathbb{E}\overline{X}^2 + \mathbb{E}\overline{Y}^2 - 2\mathbb{E}\overline{X}\,\overline{Y} = 0 \Rightarrow \mathbb{E}(\overline{X} - \overline{Y}^2) = 0 \Rightarrow \overline{X} = \overline{Y}$, защото и $D(\overline{X} - \overline{Y}) = 0$.

СЕМ, лекция 8

(2020-11-19)

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{DX}\sqrt{DY}},\ |\rho(X,Y)| \leq 1 \text{ и } |\rho(X,Y)| = 1 \Leftrightarrow Y = aX + b.$$

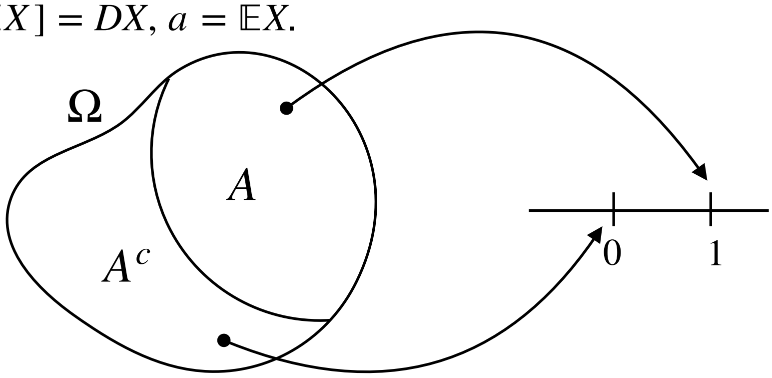
Когато X и Y са дискретни: $cov(X,Y) = \sum_{i,j}(x_i - \mathbb{E}X)(y_i - \mathbb{E}Y)p_{ij}$

$$DX = \sum_i (x_i - \mathbb{E}X)^2 \mathbb{P}(X = x_i)$$

Условно математическо очакване (УМО)

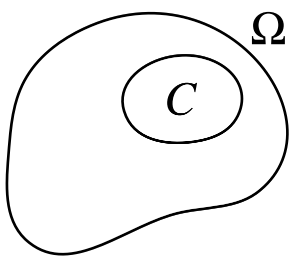
Знаем, че $\min_{a\in\mathbb{R}}(X-a)^2 = \mathbb{E}[X - \mathbb{E}X] = DX,\ a = \mathbb{E}X.$

Ако $Y = \begin{cases} 1, p \\ 0, 1-p \end{cases}$

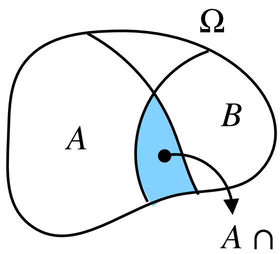


$A = \{Y = 1\}$
$A^c = \{Y = 0\}$

$Y = 1_A = 1_{\{Y=1\}}$
$p = \mathbb{E}Y = \mathbb{E}1_A = \mathbb{E}1_{\{Y=1\}} = \mathbb{P}(A) = \mathbb{P}(Y=1)$

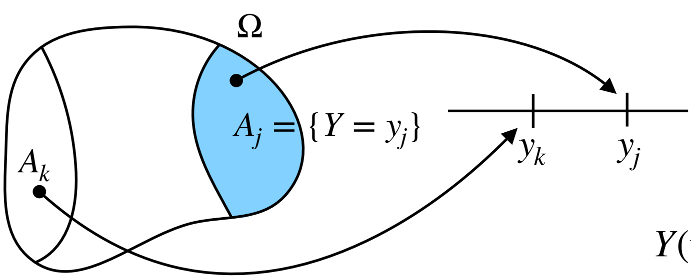


По-общо $C \subseteq \Omega$
$\mathbb{E}1_C = \mathbb{P}(C)$



A, B - множества. Тогава $1_A 1_B = 1_{A\cap B}$

Ако Y е дискретна случайна величина, то $Y = \sum_j y_j.1_{A_j}$, където A_j е пълна група от събития и $\mathbb{P}(A_j) = \mathbb{P}(Y = y_j),\ \ A_j = \{Y = y_j\}$



$$Y(w) = 0 + \ldots + 0 + y_k.1 + 0 + \ldots + 0$$

$\oplus$ Случайна величина X и наблюдаваме $Y = \begin{cases} 1, \, p \\ 0, \, 1-p \end{cases}$, $Y = 1_A$, където $A = \{Y = 1\}$.

(Пример: X е клиент влязъл в магазин, а Y е дали клиента е мъж или жена)

$G: \{0,1\} \to \mathbb{R}$

$\min\limits_G \mathbb{E}[X - G(Y)]^2 = ?$ От всички функции G искаме да вземем тази, която минимизира квадратичната грешка.

$\mathbb{E}[X - f(Y)]^2$, $f(x) = ?$

$G(Y) = a.1_A + b.1_B = aY + b(1 - Y)$, тъй като $1 - Y = 1_{A^c}$

$$\min_{a,b} \mathbb{E}[X - a1_A - b1_{A^c}]^2 = \min_{a,b}(\mathbb{E}X^2 - a^2\mathbb{E}1_A + b^2\mathbb{E}1_B - 2a\mathbb{E}X1_A - 2b\mathbb{E}X1_{A^c} - 0) = f(a,b)$$

Интересувме се от $\min\limits_{a,b} f(a,b) \Rightarrow \begin{cases} 0 = \frac{\partial f}{\partial a} = 2a\mathbb{E}1_A - 2\mathbb{E}1_A X \\ 0 = \frac{\partial f}{\partial b} = 2b\mathbb{E}1_{A^c} - 2\mathbb{E}X1_{A^c} \end{cases} \Rightarrow$

$$\boxed{a = \frac{\mathbb{E}X1_A}{\mathbb{E}1_A} \quad \text{и} \quad b = \frac{\mathbb{E}X1_{A^c}}{\mathbb{E}1_{A^c}}}$$

$$G(Y) = \underbrace{\frac{\mathbb{E}X1_A}{\mathbb{E}1_A}}_{a} \times 1_A + \underbrace{\frac{\mathbb{E}X1_{A^c}}{\mathbb{E}1_{A^c}}}_{b} \times 1_{A^c} = \frac{\mathbb{E}XY}{\mathbb{P}(Y=1)} \times 1_{\{Y=1\}} + \frac{\mathbb{E}(1-Y)}{\mathbb{P}(Y=0)} \times 1_{\{Y=0\}}$$

<u>Дефиниция</u>: (**Условно очакване**) Нека X и Y са две случайни величини. Тогава

$$\mathbb{E}[X \,|\, Y] = f(y): \ \min_G \mathbb{E}[X - G(Y)]^2 = \mathbb{E}[X - f(Y)^2]$$

$\oplus$ $X = 1_B$

Ω

$$\mathbb{E}[X \,|\, Y] = \mathbb{E}[1_B \,|\, Y] =$$

$$= \frac{\mathbb{E}1_B1_A}{\mathbb{P}(A)}1_A + \frac{\mathbb{E}1_B1_A}{\mathbb{P}(A^c)}1_{A^c} =$$

$$= \underbrace{\mathbb{P}(B \,|\, A)} \times 1_A + \underbrace{\mathbb{P}(B \,|\, A^c)} \times 1_{A^c}$$

<u>Твърдение</u>: Нека X и Y са случайни величини, като Y е дискретна.

Y	Y_1	$\dots$	y_j	$\dots$
$\mathbb{P}$	p_1	$\dots$	p_j	$\dots$

$$Y = \sum_j y_j 1_{A_j}, \quad A_j = \{Y = y_j\}, \quad \mathbb{P}(A_j) = p_j.$$

Тогава $\mathbb{E}[X \,|\, Y] = \sum_j \dfrac{\mathbb{E}X1_{A_j}}{\mathbb{P}(A_j)} 1_{A_j}$. Количеството $\mathbb{E}[X \,|\, Y = y_j] = \dfrac{\mathbb{E}X1_{A_j}}{\mathbb{P}(A_j)}$ се нарича условно очакване на X при положение (условие) $Y = y_j$.

$\oplus$ $X = 1_B,\ B = \{X = 1\}$

$$\mathbb{E}[1_B \,|\, Y = y_j] = \frac{\mathbb{E}1_B 1_{A_j}}{\mathbb{P}(A_j)} = \frac{\mathbb{P}(B \cap A_j)}{\mathbb{P}(A_j)} = \mathbb{P}[B \,|\, A_j] = \mathbb{P}[X = 1 \,|\, Y = y_j]$$

$\oplus$ $X = \sum_i x_i 1_{B_i}$, X е дискретна случайна величина.

$Y = \sum_j y_j 1_A$ и $p_{ij} = \mathbb{P}(X = x_i \cap Y = y_j)$

$$\mathbb{E}[X \,|\, Y] = \sum_j \frac{\mathbb{E}X1_{A_j}}{\mathbb{P}(A_j)} 1_{A_j}; \qquad \underbrace{\mathbb{E}X1_{A_j}} = \mathbb{E}\left(\sum_i x_i 1_{B_i}\right) = \sum_i x_i \mathbb{E}1_{B_i} 1_{A_j} =$$

$$= \sum_i x_i \mathbb{P}(B_i \cap A_i) = \underbrace{\sum_i x_i p_{ij}}$$

$$\mathbb{E}[X \,|\, Y = y_j] = \frac{\sum_i x_i p_{ij}}{\mathbb{P}(A_j)} = \sum_i x_i \frac{\mathbb{P}(A_j \cap B_i)}{\mathbb{P}(A_j)} = \sum_i x_i \mathbb{P}(B_i \,|\, A_j) = \sum_i \underline{x_i}\, \underbrace{\mathbb{P}(X = x_i \,|\, Y = y_i)}$$

<u>Твърдение</u>: X и Y са случайни величини, като Y е дискретна. Тогава $\mathbb{E}[X \,|\, Y]$ е дискретна случайна величина и $\mathbb{E}[X \,|\, Y] = \sum_j \mathbb{E}[X \,|\, Y = y_j] 1_{A_j}$.

$\mathbb{E}[X \,\vert\, Y]$	$\dots$	$\mathbb{E}[X \,\vert\, Y = y_j]$
$\mathbb{P}$	$\dots$	$\mathbb{P}(A_j)$

Свойства на условните математически очаквания

<u>Теорема</u>: Нека X, Z са случайни величини и Y е дискретна случайна величина - $Y = \sum_j y_j 1_{A_j}$. Тогава:

а) $\mathbb{E}[aX + bZ \,|\, Y] = a\mathbb{E}[X \,|\, Y] + b\mathbb{E}[Z \,|\, Y]$
б) Ако $X \perp\!\!\!\perp Y$, то $\mathbb{E}[X \,|\, Y] = \mathbb{E}X$
в) Ако $X = g(Y)$, то $\mathbb{E}[X \,|\, Y] = g(Y)$
г) $\mathbb{E}\Big[\mathbb{E}[X \,|\, Y]\Big] = \mathbb{E}X$
д) $\mathbb{E}\Big[f(U, Y) \,|\, Y = y_j\Big] = \mathbb{E}f(U, y_j)$, където

U е конкретна случайна величина $U = X$;
U е вектор от случайни величини $U = (X_i)_{i \geq 1}$
U е редица от случайни величини $U = (X_1, \ldots, X_n)$, ако $U \perp\!\!\!\perp Y$.

<u>Доказателство</u>:

а) $$\mathbb{E}[aX + bZ \,|\, Y] = \sum_j \frac{\mathbb{E}\Big[(aX + bZ)1_{A_j}\Big]}{\mathbb{P}(A_j)} 1_{A_j} = \sum_j \frac{a\mathbb{E}X1_{A_j} + b\mathbb{E}Z1_{A_j}}{\mathbb{P}(A_j)} 1_{A_j} =$$

$$= a \underbrace{\sum_j \frac{\mathbb{E}X1_{A_j}}{\mathbb{P}(A_j)} 1_{A_j}} + b \underbrace{\sum_j \frac{\mathbb{E}Z1_{A_j}}{\mathbb{P}(A_j)} 1_{A_j}} = a\mathbb{E}[X \,|\, Y] + b\mathbb{E}[Z \,|\, Y].$$

б) Нека X е дискретна (ще го докажем само в този случай, макар и да е вярно и ако не е дискретна)

$$\Rightarrow \mathbb{E}[X \,|\, Y] = \sum_j \underbrace{\mathbb{E}[X \,|\, Y = y_j]}\, 1_{A_j} = \mathbb{E}[X \,|\, Y = y_j] = \sum_j x_i \mathbb{P}(X = x_i \,|\, Y = y_j) =$$

$$\overset{X \perp\!\!\!\perp Y}{=} \sum_j \frac{\mathbb{P}(X = x_i)\cancel{\mathbb{P}(Y = y_j)}}{\cancel{\mathbb{P}(Y = y_i)}} = \underbrace{\sum_j \mathbb{E}1_{A_j}}_{\text{пълна група от събития}} = \mathbb{E}X.$$

в) $$\mathbb{E}[X \,|\, Y] = \sum_j \frac{\mathbb{E}\Big[g(Y)1_{A_j}\Big]}{\mathbb{P}(A_j)} 1_{A_j} = S,$$

имаме, че $g(Y)1_{A_j} = g(Y).1_{\{Y = y_j\}} = g(y_j)1_{\{Y = y_j\}}$

$$\Rightarrow S = \sum_j \frac{\mathbb{E}\Big[g(y_j)1_{A_j}\Big]}{\mathbb{P}(A_j)} 1_{A_j} = \sum_j g(y_j) \frac{\cancel{\mathbb{P}(A_j)}}{\cancel{\mathbb{P}(A_j)}} 1_{A_j} = g(Y) = X.$$

г) Нека X е дискретна.

$$\mathbb{E}[X \mid Y] = \sum_j \mathbb{E}[X \mid Y = y_j] 1_{A_j}$$

$$\mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}\left[\sum_j \mathbb{E}[X \mid Y = y_j] 1_{A_j}\right] = \sum_j \mathbb{E}[X \mid Y = y_j] \mathbb{E} 1 A_j =$$

$$= \sum_j \mathbb{E}[X \mid Y = y_j] \mathbb{P}(Y = y_j) = \sum_j \frac{\mathbb{E} X 1_{A_j}}{\cancel{\mathbb{P}(A_j)}} \cancel{\mathbb{P}(A_j)} = \mathbb{E} \sum_j X 1_{A_j} = \mathbb{E} X \underbrace{\sum_j 1_{A_j}}_{=1} = \mathbb{E} X.$$

д) $\oplus$ Имаме следния модел:
В даден магазин влизат клиенти, като знаем, че мъжете закупуват нещо с вероятност $1/3$, а жените с вероятност $2/3$.

$$X = \begin{cases} 1, & \mathbb{P}(Y) \\ 0, & \mathbb{P}(Y) \end{cases}, \quad Y = \begin{cases} \frac{1}{3}, & \text{мъж (вероятност 1/2)} \\ \frac{2}{3}, & \text{жена (вероятност 1/2)} \end{cases}$$

закупува или не

$X = Z_1.1_{\{Y=\frac{1}{3}\}} + Z_2.1_{\{Y=\frac{2}{3}\}}$, където $Z_1 \in Ber\left(\frac{1}{3}\right)$, а $Z_2 \in Ber\left(\frac{2}{3}\right)$.

$$X = f(Z_1, Z_2, Y) = \begin{cases} Z_1, & Y = \frac{1}{3} \\ Z_2, & Y = \frac{2}{3} \end{cases}$$

$$\mathbb{E}X = \mathbb{P}(X=1) = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}\left[X \mid Y = \frac{1}{3}\right] \mathbb{P}\left(Y = \frac{1}{3}\right) + \mathbb{E}\left[X \mid Y = \frac{2}{3}\right] \mathbb{P}\left(Y = \frac{2}{3}\right) =$$

$$= \mathbb{E}\left[Z_1 \mid Y = \frac{1}{3}\right] \times \underbrace{\frac{1}{2} + \frac{1}{2}}_{\text{мъж или жена}} \times \mathbb{E}\left[Z_2 \mid Y = \frac{2}{3}\right] = \frac{1}{2}\left(\mathbb{E}Z_1 + \mathbb{E}Z_2\right) = \frac{1}{2}\left(\frac{1}{3} + \frac{2}{3}\right) = \frac{1}{2}$$

$\oplus$ $U = (X_j)_{j \geq 1}$, X_j са независими една от друга случайни величини и $X_j \in Ber(p)$ (хвърляне на нечестна монета с вероятност p за ези и q за тура)

Нека $N \in Ge(r)$ и N не зависи от U. Търсим $\mathbb{E} \sum_{j=1}^{N+1} X_j$.

$f(U,N) = \sum_{j=1}^{n+1} X_j$, ако $N = n$. Тогава

$$\mathbb{E}\sum_{j=1}^{N+1} X_j = \mathbb{E}\left[f(U,N)\right] = \mathbb{E}\left[\mathbb{E}f(U,N)\,|\,N\right] = \mathbb{E}\sum_{n=0}^{\infty}\mathbb{E}\left[f(U,N)\,|\,N=n\right]1_{N=n} =$$

$$= \mathbb{E}\left[\sum_{n=0}^{\infty}\mathbb{E}\left[f(U,n)\right]1_{N=n}\right] = \mathbb{E}\sum_{n=0}^{\infty}\left[\mathbb{E}\sum_{j=1}^{n+1}X_j\right]1_{\{N=n\}} = \sum_{n=0}^{\infty}\mathbb{E}\sum_{j=1}^{n+1}X_j\,.\,\mathbb{E}1_{N=n} = \sum_{n=0}^{\infty}\mathbb{E}\sum_{j=1}^{n+1}X_j\,.\,\mathbb{P}(N=n) =$$

$$= \sum_{n=0}^{\infty}(n+1)p(1-r)^n r = pr\sum_{n=0}^{\infty}(n+1)(1-r)^n = T$$

$$\left[\quad \frac{1}{(1-x)^2} = \frac{\partial}{\partial x}\left(\frac{1}{1+x}\right) \overset{|x|<1}{=} \frac{\partial}{\partial x}\left(\sum_{n=0}^{\infty}x^n\right) = \sum_{n=1}^{\infty}n\,x^{n-1} = \sum_{n=0}^{\infty}(n+1)x^n \quad\right]$$

$$\Rightarrow T \overset{x=1-r}{=} pr\frac{1}{\big(1-(1-r)\big)^2} = pr\,.\,\frac{1}{r^2} = \frac{p}{r}.$$

Условни разпределения

Нека X и Y са дискретни случайни величини. Тогава под разпределение на X при условие $Y = y_j$ се разбира следната таблица:

$X\,\|\,Y = y_j$	$\dots$	x_i	$\dots$
$\mathbb{P}$	$\dots$	$\mathbb{P}(X = x_i\,\|\,Y = y_j)$	$\dots$

$$\sum_j \mathbb{P}(X = x_i\,|\,Y = y_j) = 1,\ \forall j$$

$\oplus$ Хвърляме два зара $(1,...,6)$. Нека X е броят шестици, а Y е броят единици. Търси се (X, Y). За едно хвърляне може да имаме $0,1,2$ шестици или единици.

$Y\backslash X$	$X = 0$	1	2	
$Y = 0$	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\mathbb{P}(Y=0) = \frac{25}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\mathbb{P}(Y=1) = \frac{10}{36}$
2	$\frac{1}{36}$	0	0	$\mathbb{P}(Y=2) = \frac{1}{36}$
	$\mathbb{P}(X=0) = \frac{25}{36}$	$\mathbb{P}(X=1) = \frac{10}{36}$	$\mathbb{P}(X=2) = \frac{1}{36}$	

Условно разпределение:

$X \mid Y$	$X=0$	1	2	
$Y=0$	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$	$\sum_{i=0}^{2} \mathbb{P}(X \mid Y=i)=1$
1	$\frac{8}{10}$	$\frac{2}{10}$	0	$\dots$
2	1	0	0	$\dots$

Ж. Полиномно разпределение

Имаме n - независими експеримента. Всеки експеримент има r възможни стойности с вероятност $p_0, p_1, \dots, p_{r-1}$ и $p_0 + p_1 + \dots + p_{r-1} = 1$. Тогава $(X_0, X_1, \dots, X_{r-1})$ са случайнит величини X_i - брой експерименти измежду n, които са върнали i за $0 \le i \le r-1$.

Забележка: $(X_0, \dots, X_{r-1})$ вече не са независими!

$J = \mathbb{P}(X_0 = k_0, \dots, X_{r-1} = k_{r-1})$, където $k_0 + \dots + k_{r-1} = n$ и $k_i \in \mathbb{N}_0$

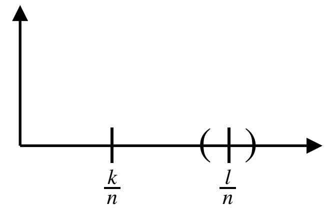
$$J = \binom{n}{k_0} p_0^{k_0} \binom{n-k_0}{k_1} p_1^{k_1} \dots \binom{n-k_0-k_1-\dots-k_{r-2}}{k_{r-1}} p_{r-1}^{k_{r-1}}.$$

СЕМ, лекция 9
(2020-11-26)

Непрекъснати случайни величини (НСВ)

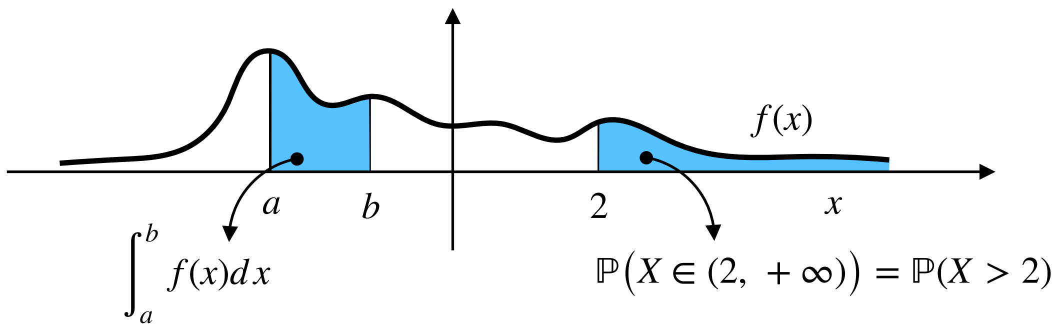
$X: \ \Omega \to \mathbb{R}$ и X приема неизброимо много стойности. $(\Omega, \mathscr{A}, \mathbb{P})$ - вероятностно пространство.

$\oplus \ \mathbb{P}\left(T = \frac{k}{n}\right), 0 \le k \le n$



$\oplus$ Интересуваме се от случайната величина $X \in (a, b)$, $\quad \big(T \in (1^\circ C, 3^\circ C)\big)$

$\oplus$ Фундаменталните теореми включват непрекъснати случайни величини.



<u>Дефиниция</u>: X е (абсолютно) непрекъсната случайна величина, ако $\exists f_X : \mathbb{R} \to \mathbb{R}$, такава, че: а) $f_X(x) \ge 0, \ \forall x \in \mathbb{R}$,

б) $\displaystyle\int_{-\infty}^{+\infty} f_X(x)dx = 1$,

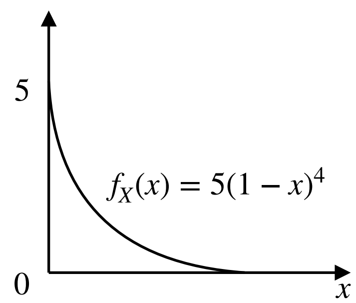
в) $\mathbb{P}\big(X \in (a, b)\big) = \displaystyle\int_a^b f_X(x)dx, \ \forall a, b \in (-\infty, +\infty), \ a < b$,

f_X се нарича плътност на X.

$\oplus$ Дадена застрахователна компания обслужва здравните полици на някаква малка фирма. Разходите за обслужване на годишна полица е $M = 100\,000X$, където застрахователната компания е оценила, че X приема стойности в интервала $(0,1)$ и

$$f_X(x) = \begin{cases} 5(1-x)^4, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

Търси се $\mathbb{P}(M > 10\,000)$ - вероятността цената за обслужване на 1 година да е по-голяма от $10\,000$.



Първо нека проверим, дали функцията в действителност може да бъде плътност на X. Проверка:

$$\int_0^1 5(1-x)^4 dx \overset{y=1-x}{=} 5\int_0^1 y^4 dy = 5.\frac{y^5}{5}\Bigg|_0^1 = 1,$$ т.е. плътността е добре определена.

$$\mathbb{P}(100\,000X > 10\,000) = \mathbb{P}(X > \frac{1}{10}) = \int_{\frac{1}{10}}^{\infty} f_X(x)dx =$$

$$= 5\int_{\frac{1}{10}}^{1} (1-x)^4 dx \overset{y=1-x}{=} 5\int_0^{\frac{9}{10}} y^4 dy == 5\frac{y^5}{5}\Bigg|_0^{\frac{9}{10}} = \left(\frac{9}{10}\right)^5 \approx 0.59.$$

<u>Твърдение</u>: Нека X е непрекъсната случайна величина (НСВ). Тогава $\mathbb{P}(X=c)=0$, $\forall c \in \mathbb{R}$. Следователно
$\mathbb{P}\big(X\in[a,b]\big) = \mathbb{P}\big(X\in[a,b)\big) = \mathbb{P}\big(X\in(a,b]\big) = \mathbb{P}\big(X\in(a,b)\big)$,
$\forall a,b \in (-\infty,+\infty)$, $a<b$. Случайната величина няма маса, т.е. има нулева вероятност в конкретна точка.

<u>Доказателство</u>: $\{X=c\} = \bigcap_{n\geq 1}\left\{X\in\left((c-\frac{1}{n}, c+\frac{1}{n}\right)\right\}$, $\forall n\geq 1$, но

$$Q = \int_{c-\frac{1}{n}}^{c+\frac{1}{n}} f_X(x)dx \Rightarrow$$

$$\mathbb{P}(X=c) \leq \lim_{n\to\infty}\int_{c-\frac{1}{n}}^{c+\frac{1}{n}} f_X(x)dx = \int_{c-\frac{1}{n}}^{c+\frac{1}{n}} f_X(x)dx = \int_c^c f_X(x)dx = 0 \Rightarrow$$

$\mathbb{P}(X=c)\leq 0$, но $\mathbb{P}(X=c)\geq 0 \Rightarrow \mathbb{P}(X=c=0)$. Сега,

$\{X\in[a,b]\} = \{X\in(a,b)\}\cup\{X=a\}\cup\{X=b\}$, което е обединение на непресичащи се събития. Следователно,

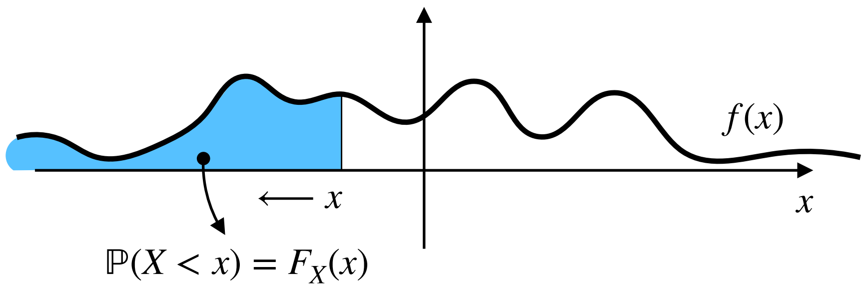
$$\mathbb{P}\left(X\in[a,b]\right) = \mathbb{P}\left(X\in(a,b)\right) + \underbrace{\mathbb{P}(X=a)}_{=0} + \underbrace{\mathbb{P}(X=b)}_{=0} = \mathbb{P}\big(X\in(a,b)\big).$$

<u>Дефиниция</u>: (**Функция на разпределение на НСВ**) Нека X е НСВ с плътност f_X. Тогава функцията $F_X(x) = \mathbb{P}(X<x) = \int_{-\infty}^{\infty} f_X(y)dy$, $\forall x\in(-\infty,+\infty)$ се нарича функция на разпределението на X.

<u>Свойства</u>:

- ако f_X е непрекъсната в точка x_0, то $\frac{\partial}{\partial x}F_X\Big|_{x=x_0} \quad f_X(x_0)$

- $F_X(-\infty) = \lim_{x\to-\infty} \mathbb{P}(X < x) = \lim_{x\to-\infty} \int_{-\infty}^{+\infty} f_X(y)dy = 0$



Аналогично

- $F_X(+\infty) = 1$
- $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X < x) + \mathbb{P}(X = x) = \mathbb{P}(X < x)$

Смяна на променливите на НСВ

Дадена е НСВ X с плътност f_X.
$Y = g(X)$ е някаква детерминираща функция. Питаме се дали може да пресметнем плътността на Y.
Първото нещо, в което трябва да се обедим е, че не за всяка функция g, Y има плътност.

Нека например $g : \mathbb{R} \to \{0,1\}$, като $g(x) = \begin{cases} 1, \ x \geq 0 \\ 0, \ x < 0 \end{cases}$. Тогава
$Y = g(X) = \begin{cases} 1, & \mathbb{P}(X \geq 0) \\ 0, & \mathbb{P}(X < 0) \end{cases} \Rightarrow Y \in Ber\left(\mathbb{P}(X \geq 0)\right)$, т.е. за плътност на Y не може и да става дума, тъй като Y не приема неизброимо много стойности, а само краен брой такива - само две $\{0,1\}$.

Ще изследваме конкретен клас от функции, за които Y има плътност и тя се пресмята сравнително лесно, чрез f_X и свойствата на g.

<u>Теорема</u>: (**Смяна на променливите на НСВ**) Нека X е НСВ с плътност f. Нека g е строго монотонно растяща или намаляваща (е монотонна) функция. Тогава
$Y = g(X)$ е НСВ с плътност $\varphi(y) = f\left(g^{-1}(y)\right) . \left| \dfrac{\partial}{\partial y}\left(g^{-1}(y)\right) \right|$ или
$\varphi = f\left(h(y)\right) . \left| h'(y) \right|$, където $h = g^{-1}$.

<u>Доказателство</u>: Нека $g \uparrow$ (g е строго растяща).

$$\underbrace{\mathbb{P}\left(Y \in (a,b)\right)}_{=\int_a^b ?} = \mathbb{P}\left(g(X) \in (a,b)\right) \overset{g\uparrow}{=} \mathbb{P}\left(X \in \left(g^{-1}(a), g^{-1}(b)\right)\right) \overset{h=g^{-1}}{=}$$

$$= \mathbb{P}\Big(X \in \big(h(a), h(b)\big)\Big) = \int_{h(a)}^{h(b)} f(x)dx \underset{\begin{cases} h(a) = h(v) \Rightarrow v = a \\ h(b) = h(v) \Rightarrow v = b \end{cases}}{\overset{x=h(v)}{=}} \int_a^b g\,\big(h(v)\big)\,dh(v) =$$

$$= \int_a^b \underbrace{f\,\big(h(v)h'(v)\big)}\,dv = \int_a^b \varphi(v)dv \Rightarrow \varphi(y) = f\,\big(h(y)\big)\,h'(y).$$

Аналогично и за $g \downarrow$ (с дребни промени)

$\varphi(y) = f\,\big(h(y)\big) \quad \underbrace{\big(-h'(u)\big)}_{>0 \Leftarrow g\downarrow,\ h=g^{-1}}$. Във всеки случай $\varphi(y) = f\,\big(h(y)\big)\,.\,\big|h'(y)\big|$.

Математическо очакване на НСВ

Нека X е НСВ. Под очакване на X се разбира $\mathbb{E}X = \displaystyle\int_{-\infty}^{\infty} xf_X(x)dx$, ако $\displaystyle\int_{-\infty}^{\infty} |x|f_X(x)dx < \infty$ (е крайно).

$\Big[$ Припомняне: За X дискретно $\mathbb{E}X = \displaystyle\sum_i x_i p_i$ $\Big]$

Свойства:

- $\mathbb{E}cX = c\mathbb{E}X$

  $\mathbb{E}Y = \displaystyle\int_{-\infty}^{\infty} y\,\underbrace{f_Y(y)}\,dy$; (може да се покаже с функцията $g(x) = cx$)
- $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$
- $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$, ако $X \perp\!\!\!\perp Y$
- $Y = g(X)$, то $\mathbb{E}Y = \mathbb{E}g(X) = \displaystyle\int_{-\infty}^{\infty} g(x)f_X(x)dx$, където $g \uparrow$ или $g \downarrow$.

$g \uparrow$, то $f_Y(y) = f_X\,\big(h(y)\big)\,.\,h'(y)$, където $h(y) = g^{-1}(y)$.

$$\mathbb{E}Y = \int_{-\infty}^{\infty} yf_Y(y)dy = \int_{-\infty}^{\infty} yf_X\,\big(h(y)\big)\,.\,h'(y)dy \underset{y=g(x)\Rightarrow dy=g'(x)dx}{\overset{x=h(y)}{=}}$$

$$= \int_{-\infty}^{\infty} g(x)f_X(x)\underbrace{h'\,\big(g(x)\big)\,g'(x)}_{\underset{(*)}{=}1}\,dx = \int_{-\infty}^{\infty} g(x)f_X(x)dx.$$

$$(*) \quad x = h\left(g(x)\right) \Big| \frac{\partial}{\partial x} \Rightarrow 1 = h'\left(g(x)\right) . g'(x).$$

<u>Дефиниция</u>: (**Дисперсия**) Нека X е НСВ с плътност f_X. Тогава, ако $\int_{-\infty}^{\infty} x^2 f_X(x) \mathrm{d}\, x < \infty$ (е крайно), то под дисперсия на X разбираме

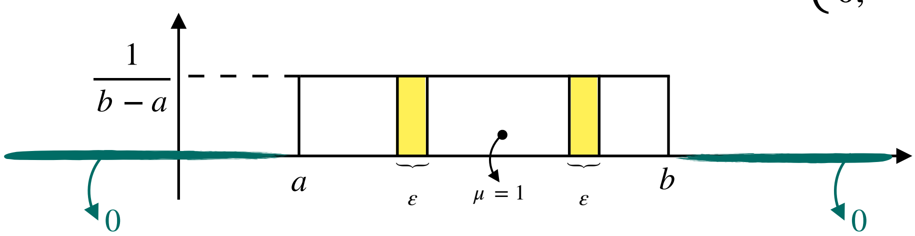
$$DX = \mathbb{E}\underbrace{[X - \mathbb{E}X]^2}_{g(x)} = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f_X(x) \mathrm{d}\, x$$

<u>Свойства</u>:
- $DcX = c^2 DX$
- $D(X + c) = DX$
- $D(X + Y) = DX + DY \Leftrightarrow X \perp\!\!\!\perp Y$

Видове НСВ

А. Равномерно разпределена НСВ

<u>Дефиниция</u>: За $a < b$ казваме, че $X \in Unif(a, b)$, ако $f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}$.



Ако $Y = \dfrac{X - a}{b - a}$, то $Y \in Unif(0,1)$. Нека $g(x) = \dfrac{x - a}{b - a}$, тогава $g(x) \uparrow$.

$$h(y) = g^{-1}(y) = (b - a)y + a$$

$$f_Y(y) = f_X\Bigg(\underbrace{\overbrace{(b - a)y + a}^{=x}}_{y \in (0,1) \Leftrightarrow x \in (a,b)} \Bigg)(b - a) \Rightarrow f_Y(y) = \begin{cases} \frac{1}{b-a}(b - a) = 1, & y \in (0,1) \\ 0, & y \notin (0,1) \end{cases}$$

Нека $X \in Unif(a, b)$. Тогава

$$Y = \frac{X - a}{b - a} \in Unif(0,1) \Rightarrow \mathbb{E}Y = \frac{1}{b - a}\mathbb{E}(X - a) = \frac{1}{b - a}(\mathbb{E}X - a)$$ и

$$DY = \left(\frac{1}{b - a}\right)^2 D(X - a) = \frac{DX}{(b - a)^2},$$ но

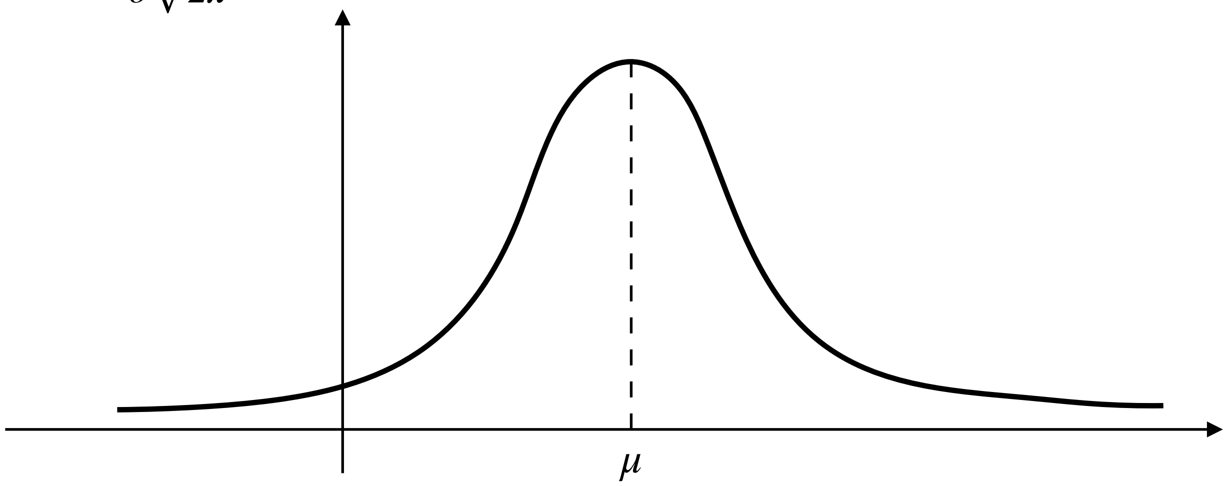
$$\mathbb{E}Y = \int_0^1 y.1\,\mathrm{d}\,y = \frac{1}{2} \Rightarrow \mathbb{E}X = \frac{b-a}{2} + a = \frac{a+b}{2};$$

$$DY = \int_0^1 \left(y - \frac{1}{2}\right)^2 \times 1\,\mathrm{d}\,y = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12} \Rightarrow DX = \frac{1}{12}(b-a)^2.$$

- $DX \geq 0,\ DX = \mathbb{E}X - (\mathbb{E}X)^2$

Б. Нормално разпределена НСВ

Казваме, че $X \in \mathcal{N}orm(\mu, \sigma^2)$ или $X \in \mathcal{N}(\mu, \sigma^2)$, където $\mu \in \mathbb{R}$, $\sigma^2 > 0$, ако $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$.



$$\frac{1}{\sigma 2\pi} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1,\ \forall \mu, \sigma.$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

F_X комулативната функция на нормалното разпределение. За пресмятането и се ползват трансформации до стандартното нормално разпределение (ползват таблици). Това е така, тъй като не може да интегрираме ($F_X(x)$ няма явен вид). Интеграла се приближава числено.

(За домашно може да пресметнем комулативната функцията на равномерното разпределение, което пропуснахме)

Ако $X \sim \mathcal{N}(\mu, \sigma^2)$, то $Y = \frac{1}{\sigma}(X - \mu) \sim \mathcal{N}(0,1)$ (линейно транслиране) и се нарича стандартно нормално разпределение.

$g(x) = \dfrac{x-\mu}{\sigma}$ е $\uparrow$. $h(y) = \sigma y + \mu$, $h'(y) = \sigma$. Тогава
$f_Y(y) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sigma y + \mu - \mu)^2}{2\sigma^2}.\sigma} = \dfrac{1}{\sqrt{2\pi}} \,.\, e^{-\frac{y^2}{2}}$, $\forall y \in (-\infty, \infty)$.

Нека $X \sim \mathcal{N}(\mu, \sigma^2)$, тогава $Y = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ и

$$\mathbb{E}Y = \frac{1}{\sigma}\mathbb{E}X - \frac{\mu}{\sigma} \Rightarrow \mathbb{E}Y = \int_{-\infty}^{\infty} y f_X(y)\mathrm{d}\,y = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \underbrace{ye^{-\frac{y^2}{2}}}_{\text{нечетна}}\ \mathrm{d}\,y = 0.$$

Следователно $0 = \mathbb{E}Y = \dfrac{1}{\sigma}\mathbb{E}X - \dfrac{\mu}{\sigma}$ или $\mathbb{E}X = \mu$.

$DY = \mathbb{E}Y^2 - (\underbrace{\mathbb{E}Y}_{=0})^2 = \mathbb{E}Y^2 = \dfrac{1}{\sigma^2}DX$ $\left(\text{т.к. } Y = \dfrac{X-\mu}{\sigma} \text{ и свойства на } D\right)$

$\mathbb{E}Y^2 = \displaystyle\int_{-\infty}^{\infty} y^2 f_Y(y)\mathrm{d}\,y = \underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}}\,\mathrm{d}\,y}_{I} = 1$. Следователно

$DY = 1 = \dfrac{DX}{\sigma^2} \Rightarrow DX = \sigma^2$.

Сега, за интеграла I: знаем, че $1 = \dfrac{1}{\sqrt{2\pi}\sigma}\displaystyle\int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}\,y$ за $\forall \sigma > 0$.

$$\Rightarrow \sqrt{2\pi}\sigma = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}\,y \Big| \frac{\partial}{\partial\sigma}.$$

$$\sqrt{2\pi} = \int_{-\infty}^{\infty} \frac{\partial}{\partial\sigma} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}\,y = \int_{-\infty}^{\infty} \frac{y^2}{\cancel{2}} \,.\, \frac{\cancel{2}}{\sigma^3} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}\,y = \int_{-\infty}^{\infty} \frac{y^2}{\sigma^3} e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}\,y \overset{\sigma=1}{=} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}}\,\mathrm{d}\,y$$

Този метод е известен като метода на физика Ричард Файнман. Друга алтернатива за пресмятане на интеграла I е чрез интегриране по части.

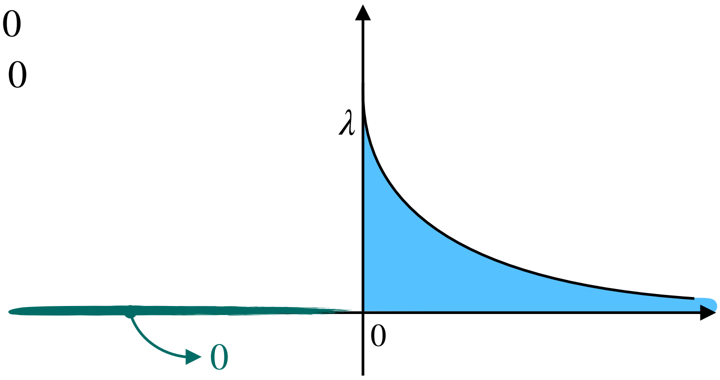
СЕМ, лекция 10

(2020-12-03)

B. Експоненциално разпределена НСВ

<u>Дефиниция</u>: Казваме, че случайната величина X е експоненциално разпределена с параметър $\lambda > 0$ и бележим $X \in Exp(\lambda)$, ако X има плътност от вида

$$f_X = \begin{cases} \lambda e^{-\lambda x}, \ x \geq 0 \\ 0, \qquad x < 0 \end{cases}$$



Проверка: $\displaystyle\int_{-\infty}^{\infty} f_X(x)\mathrm{d}\,x = \int_0^{\infty} \lambda e^{-\lambda x}\,\mathrm{d}\,x \overset{\begin{cases} y = -\lambda x \\ \mathrm{d}\,y = -\lambda\,\mathrm{d}\,x \\ \ \Rightarrow \mathrm{d}\,x = \frac{\mathrm{d}\,y}{-\lambda} \\ \lambda > 0 \Rightarrow y \in (0, -\infty) \end{cases}}{=} \int_0^{-\infty} \lambda e^y \frac{\mathrm{d}\,y}{-\lambda} =$

$= -\displaystyle\int_{-\infty}^{0} -1e^y\,\mathrm{d}\,y = \int_{-\infty}^{0} e^y\,\mathrm{d}\,y = e^y\Big|_{-\infty}^{0} = 1 - 0 = 1$. Следователно плътността е

добре дефинирана и X съществува.

Функция на разпределение $F_X(x) = \mathbb{P}(X < x)$:

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \int_0^x \lambda e^{-\lambda y}\,\mathrm{d}\,y = 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

$\overline{F}_X = \mathbb{P}(X \geq x) = \begin{cases} e^{-\lambda x}, x > 0 \\ 1, \quad\ x \leq 0 \end{cases}$, където $\overline{F}_X$ се нарича „опашка“ на X.

$$\mathbb{E}X = \int_0^{\infty} x\,.\,\lambda e^{-\lambda x}\,\mathrm{d}\,x \overset{(*)}{=} \int_0^{\infty} x\,\mathrm{d} - e^{-\lambda x} \underbrace{=}_{\substack{\text{интегриране} \\ \text{по части}}} -xe^{-\lambda x}\Big|_0^{\infty} - \int_0^{\infty} -e^{-\lambda x}\,\mathrm{d}\,x =$$

$$= 0 - 0 - \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}\, \mathrm{d}(-\lambda x) = -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty = 0 - \left(-\frac{1}{\lambda}\right) = \frac{1}{\lambda}.$$

$$\left[\quad (*) \quad \frac{\partial}{\partial x}\left[-e^{-\lambda x}\right] = \frac{\partial}{\partial x}[-\lambda x] - e^{-\lambda x} = \lambda e^{-\lambda x} \quad \right]$$

Втори подход (Файнман): Ние знаем, че

$$1 = \lambda\int_0^\infty e^{-\lambda x}\, \mathrm{d}\, x, \forall \lambda > 0 \Rightarrow \frac{1}{\lambda} = \int_0^\infty e^{-\lambda x}\, \mathrm{d}\, x \Big| \frac{\partial}{\partial \lambda} \Rightarrow -\frac{1}{\lambda^2} = \int_0^\infty - xe^{-\lambda x}\, \mathrm{d}\, x$$

$\Rightarrow \frac{1}{\lambda} = \lambda \int_0^\infty xe^{-\lambda x}\, \mathrm{d}\, x = \mathbb{E}X$, което искахме да докажем.

$$\mathbb{E}X^2 = \int_0^\infty x^2 \lambda e^{-\lambda x}\, \mathrm{d}\, x = \frac{1}{\lambda^2}\,.\,\lambda\,.\,\frac{1}{-\lambda}\int_0^\infty (-\lambda x)^2\,.\,e^{-\lambda x}\, \mathrm{d}(-\lambda x) =$$

$$= -\frac{1}{\lambda^2}\int_0^\infty e^{-\lambda x}\, \mathrm{d}\, 2(-\lambda x) = -\frac{2}{\lambda^2}\left(e^{-\lambda x}\Big|_0^\infty\right) = -\frac{2}{\lambda^2}(0-1) = \frac{2}{\lambda^2}.$$

Окончателно, ако $X \sim Exp(\lambda)$, то $\mathbb{E}X = \frac{1}{\lambda}$ и $DX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

Експоненциалното разпределение е непрекъснат аналог на геометричното разпределение в смисъла на свойството безпаметност (memoryless). Ако вземем $t > 0,\, s > 0$, то $\mathbb{P}(X > t + s\,|\,X > t) = \mathbb{P}(X > s)$.

<u>Доказателство</u>:

$$\frac{\mathbb{P}(X > t + s \cap X > t)}{\mathbb{P}(X > t)} \overset{\{t\}\subseteq\{t+s\}}{=} \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s)$$

Преди да продължим, нека въведем следните нотации, които ще използваме по-долу:

$X = (X_1, X_2)$, f_X - плътносъ на X, $x = (x_1, x_2)$
$Y = (Y_1, Y_2)$, f_Y - плътносъ на Y, $y = (y_1, y_2)$

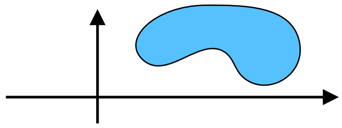
$g : \mathbb{R}^2 \to \mathbb{R}^2$. Под $y = g(x)$ се разбира $(y_1, y_2) = \big(g_1(x_1, x_2), g_2(x_1, x_2)\big) = g(x)$.

Двумерна непрекъсната случайна величина

<u>Дефиниция</u>: $X = (X_1, X_2)$ е вектор от НСВ с плътност f_X, ако е изпълнено:

- $f_X(x) \geq 0, \forall x \in \mathbb{R}^2$

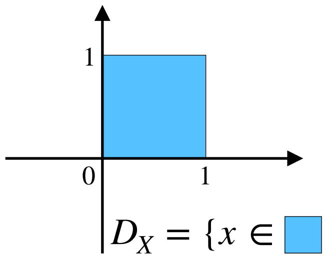
- $\int_0^{\infty}\int_0^{\infty} f_X(x_1, x_2)\underbrace{\mathrm{d}\, x_1 \,\mathrm{d}\, x_2}_{dx} = \int_{\mathbb{R}^2} f_X(x)\mathrm{d}\, x = 1$



- $\mathbb{P}(X \in D) = \int_D f_X(x)dx,\ \forall$ отворени/затворени множества $D \subseteq \mathbb{R}^2$

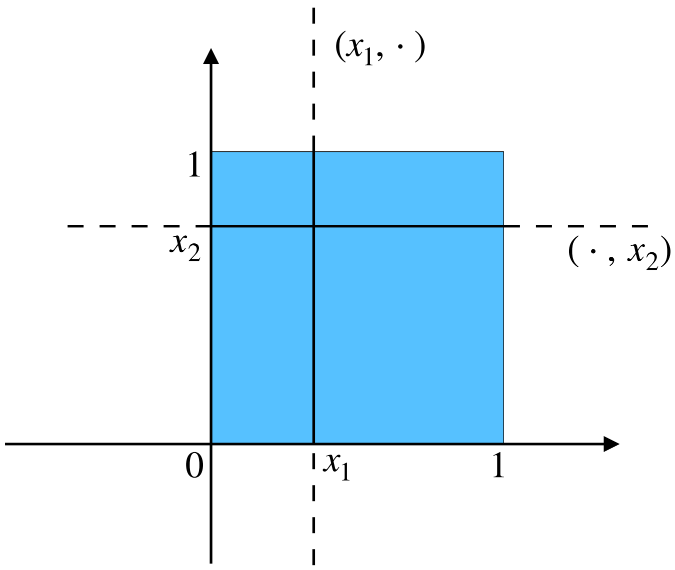
<u>Дефиниция</u>: (**Носител на случайна величина**) Нека f_X е плътността на X. Тогава $D_X = \{x \in \mathbb{R}^2 : f_X(x) > 0\}$ и D_X се нарича носител на X.

Смисъла на D_X е да показва какви са възможните стойности на случайния вектор.

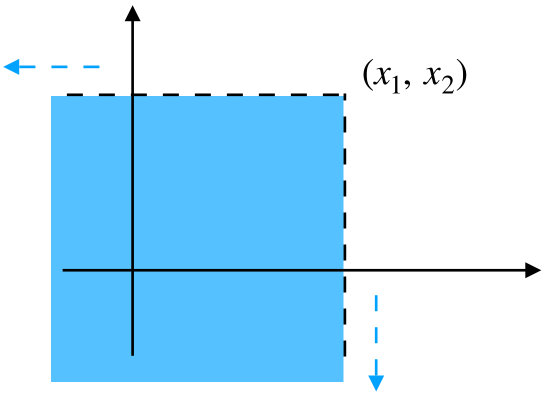


$$f_X(x) = \begin{cases} 1, & x_1, x_2 \in (0,1) \\ 0, & x_1, x_2 \notin (0,1) \end{cases}$$

<u>Дефиниция</u>: (**Маргинални разпределения**) Нека X е вектор от НСВ с плътност f_X. Тогава $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2)\mathrm{d}\, x_2$ е маргиналното разпределение на X_1 и $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_X(x_1, x_2)\mathrm{d}\, x_1$ е маргиналното разпределение на X_2



<u>Дефиниция</u>: (**Функция на разпределение**) X е вектор от случайни величини. Тогава $F_X(x) = \mathbb{P}(X_1 < x_1, X_2 < x_2),\ \forall x \in \mathbb{R}^2$ се нарича функция на разпределение на X.



Ако X е вектор от НСВ, то X има следната функция на разпределение:

$$F_X(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_X(v_1, v_2)\mathrm{d}\, v_1\,\mathrm{d}\, v_2$$

Плътността на X в точката x е равна на $f_X(x) = \dfrac{\partial^2}{\partial x_1 \partial x_2} F_X \Big|_{x=x_1+x_2}$

<u>Дефиниция</u>: (**Независимост на две непрекъснати случайни величини**)
Нека $X = (X_1, X_2)$. Тогава $X_1 \perp\!\!\!\perp X_2$, когато $F_X(x) = F_{X_1}(x_1)F_{X_2}(x_2)$ или $\mathbb{P}(X_1 < x_1, X_2 < x_2) = \mathbb{P}(X_1 < x_1)\mathbb{P}(X_2 < x_2)$ за всяко $x = (x_1, x_2) \in \mathbb{R}^2$. Ако X е вектор от НСВ, то независимостта е еквивалентна на
$f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2),\ \forall x = (x_1, x_2) \in \mathbb{R}^2$.

Т.е. двумерната плътност X се разпада на произведението на двете маргинални плътности на X_1 и $X_2 \Leftrightarrow X_1 \perp\!\!\!\perp X_2$.

Обобщение за n мерен случай:

<u>Дефиниция</u>: (**Съвкупна независимост**) Нека $X_1, X_2, \ldots, X_n$ са случайни величини, такива, че $X = (X_1, X_2, \ldots, X_n)$ има плътност f_X. Т.е. са изпълнени условията:

а) $f_X(x) \geq 0, \forall x \in \mathbb{R}^n$

б) $\displaystyle\int_{\mathbb{R}^n} f_X(x)\mathrm{d}\, x = 1$

в) $\forall x \in \mathbb{R}^n:\ \mathbb{P}(X \in D) = \displaystyle\int_D f_X(x) \underbrace{\mathrm{d}\, x}_{\mathrm{d}\, x_1 \ldots \mathrm{d}\, x_n}$

Тогава $X_1, X_2, \ldots, X_n$ са независими в съвкупност

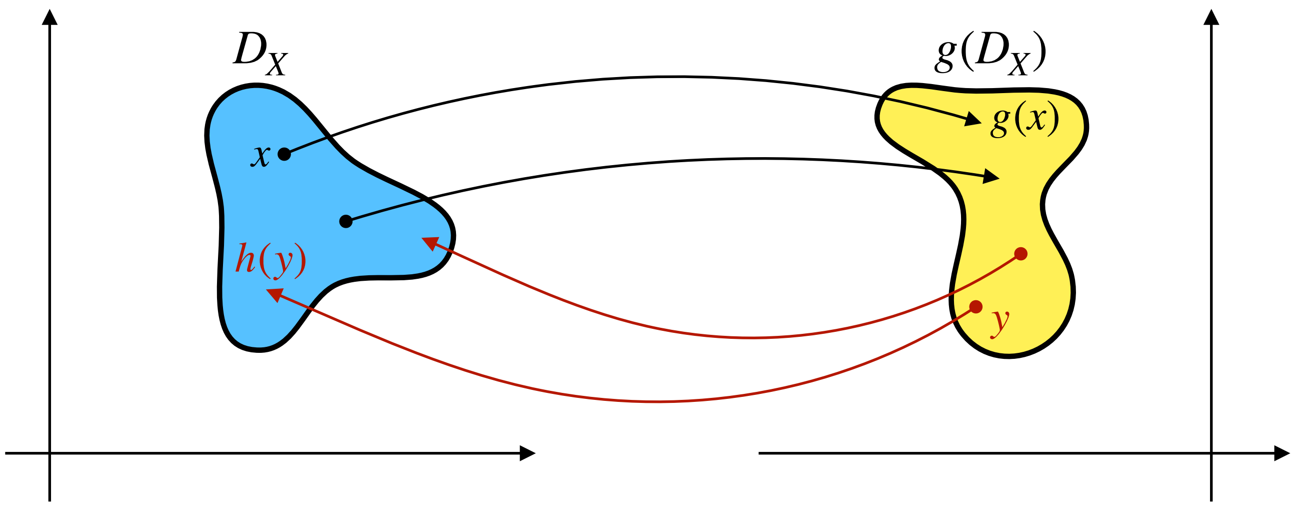
$$\Leftrightarrow f_{X_{i_1}X_{i_2}\ldots X_{i_k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \prod_{j=1}^{k} f_{X_{i_j}}(x_{i_j})$$

Смяна на променливите

Имаме $g:\mathbb{R}^2 \to \mathbb{R}^2$, $Y = g(X)$ и ще знаем плътността f_X. Въпроса е: кога и как ще може да изчислим f_Y?

$D_X = \{x \in \mathbb{R}^2 : f_X(x) > 0\}$,

$g(D_X) = \{y \in \mathbb{R}^2, \exists x \in DX : y = g(x)\}$



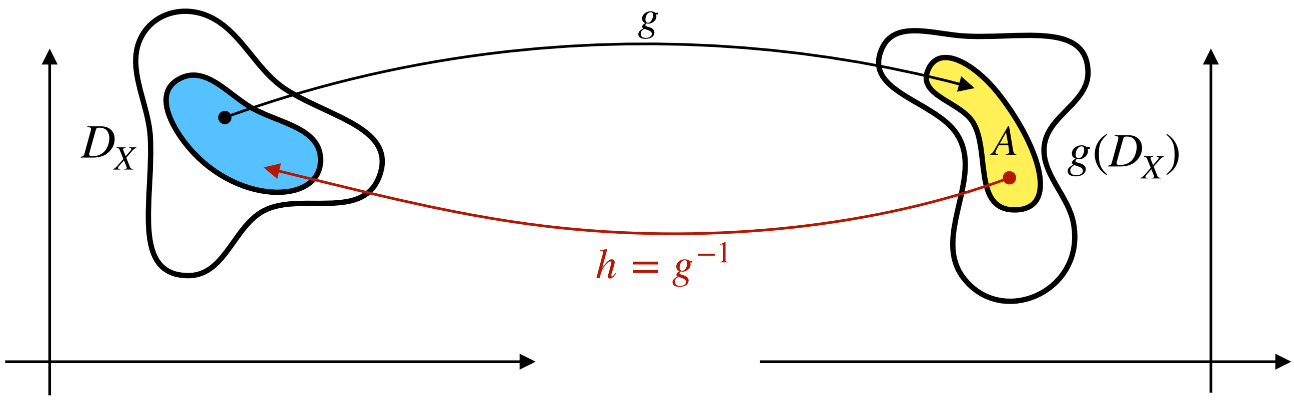
Ако g е взаимно еднозначно вътху D_X, то може да дефинираме и
$h(y) = g^{-1}(y)$, $y \in g(D_X)$.

<u>Теорема</u>: (**Смяна на променливите**) Нека X е вектор от НСВ(2) (две непрекъснати случайни величини) и $g:\mathbb{R}^2 \to \mathbb{R}^2$ е функция. Нека $Y = g(X)$.
Ако $g: D_X \to g(D_X)$ е взаимно еднозначно с обратна функция $h = g^{-1}$, h, g са непрекъснати, h има непрекъснати производни и $\forall y \in g(D_X)$ е изпълнено:

$$0 \neq \left| det \begin{vmatrix} \frac{\partial h_1}{\partial y_1}(y_1, y_2) & \frac{\partial h_1}{\partial y_2}(y_1, y_2) \\ \frac{\partial h_2}{\partial y_1}(y_1, y_2) & \frac{\partial h_2}{\partial y_2}(y_1, y_2) \end{vmatrix} \right| =: |J(y)|$$, то Y е вектор от НСВ с плътност

$$f_Y(y) = \begin{cases} f_X\left(h(y)\right) | \underbrace{J(y)}_{\text{Якобиан на смяната}} |, & y \in g(D_X) \\ 0, & y \notin g(D_X) \end{cases}$$ и $D_Y = g(D_X)$.

Доказателство: Ще покажем, че за $\forall A \subseteq g(D_X)$: $\mathbb{P}(Y \in A) = \int_A f_Y(y) \mathrm{d}\, y$.



$\mathbb{P}(Y \in A) = \mathbb{P}\left(g(x) \in A\right) = \mathbb{P}\left(X \in h(A)\right) =$

$$= \int_{x \in h(A)} f_X(x) \mathrm{d}\, x \overset{x=(x_1,x_2)=h(y)=\left(h_1(y),\, h_2(y)\right)}{=} \int_{y \in A} \underbrace{f_X\left(h(y)\right) |J(y)|} \mathrm{d}\, y \Rightarrow$$

$f_Y(y) = f_X\left(h(y)\right) |J(y)|$ е плътността на Y.

$\oplus$ Нека $V_1, V_2, \ldots, V_n$ са независими в съвкупност НСВ, т.е. $V_i \in \mathcal{N}(\mu_i, \sigma_i^2)$. Тогава $\sum_{i=1}^{n} V_i \in N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$. Ще покажем, че е изпълнено за $n = 2$. От принципа на математическата индукция ще следва за всяко $n \in \mathbb{N}$.

$n = 2$: $V_1 + V_2 = \mu_1 + \sigma_1 Z_1 + \mu_2 + \sigma_2 Z_2$, където $Z_1, Z_2 \in \mathcal{N}(0, 1)$ и
$Z_1 \perp\!\!\!\perp Z_2 \Rightarrow V_1 + V_2 = \left(\mu_1 + \mu_2\right) + \sigma_1 Z_1 + \sigma_2 Z_2$.

Поставяме $X_1 = \sigma_1 Z_1 \in \mathcal{N}(0, \sigma_1^2)$ и $X_2 = \sigma_2 Z_2 \in N(0, \sigma_2^2) \Rightarrow$

$V_1 + V_2 = \mu_1 + \mu_2 + X_1 + X_2$. Ако $X_1 + X_2 \in \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, то
$V_1 + V_2 \in \left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$.

Оттук нататък се интересуваме от $X_1 \in \mathcal{N}(0, \sigma_1^2)$, $X_2 \in \mathcal{N}(0, \sigma_2^2)$ и тяхната сума, където $X_1 \perp\!\!\!\perp X_2$.

$$f_X(x) = f_{X_1 X_2}(x_1, x_2) = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2\sigma_1^2}}}_{f_{X_1}(x_1)} \times \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2\sigma_2^2}}}_{f_{X_2}(x_2)} = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x_1^2}{2\sigma_1^2} - \frac{x_2^2}{2\sigma_2^2}} > 0$$ - новата съвместна плътност.

$D_X = \mathbb{R}^2$ (диапазона на X)

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x_1^2}{\sigma_1^2} - \frac{x_2^2}{\sigma_2^2}}, \; D_X = \mathbb{R}^2$$

$\begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = X_2 \end{cases} \Rightarrow Y = g(X)$ (линейно неизродено изображение)

$$\Rightarrow g(D_X) = g(\mathbb{R}^2) = \mathbb{R}^2$$

$$X_1 + X_2 \in N(0, \sigma_1^2 + \sigma_2^2)$$

$$y = g(x) \Leftrightarrow x = h(y) = \begin{pmatrix} y_1 & -y_2 \\ y_2 & \end{pmatrix}$$

$$|J(y)| = \left| \det \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} \right| = 1,$$

$$f_Y(y) = f_X\left(h(y)\right).1 = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(y_1-y_2)^2}{2\sigma_1^2} - \frac{y_2^2}{2\sigma_2^2}}, \; \forall y = (y_1, y_2) \in \mathbb{R}^2.$$

$$Y_1 = X_1 + X_2, \; Y_2 = X_2$$

$$f_{Y_1}(y_1) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{(y_1-y_2)^2}{2\sigma_1^2} - \frac{y_2^2}{2\sigma_2^2}} \, \mathrm{d}y^2 \overset{(*)}{=} \frac{e^{-\frac{y_1^2 c}{2}}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(by_2 - ay_1)^2} \, \mathrm{d}y_2 \overset{by_2 = w}{=}$$

$$= \frac{e^{-\frac{y_1^2 c}{2}}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(w - ay_1)^2} \, \mathrm{d}w \overset{v = w - ay_1}{=} \frac{e^{-\frac{y_1^2 c}{2}}}{2\pi\sigma_1\sigma_2 b} \underbrace{\int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2} \, \mathrm{d}v}_{\sqrt{2\pi}} = \frac{e^{-\frac{y_1^2 c}{2}}}{\sqrt{2\pi}\sigma_1\sigma_2 b}.$$

Като $(*)$ е следното допускане: $\dfrac{(y_1 - y_2)^2}{\sigma_1^2} + \dfrac{y_2^2}{\sigma_2^2} \overset{(*)}{=} c y_1^2 + (b y_2 - a y_1)^2$, т.е.

допускаме, че левия израз може да се представи като десния за някой $a, b, c \in \mathbb{R}$. Остава да намерим тези параметри a, b и c.

$$f_{Y_1}(y_1) = \frac{e^{-\frac{y_1^2 c}{2}}}{\sqrt{2\pi}\sigma_1\sigma_2 b}.$$

Връщаме се в $(*)$, за да намерим b и c.

$$\frac{(y_1-y_2)^2}{\sigma_1^2}+\frac{y_2^2}{\sigma_2^2}=cy_1^2+(by_2-ay_1)^2=cy_1^2+b^2y_2^2-2aby_1y_2+a^2y_1^2$$

$$\frac{y_1^2}{\sigma_1^2}-\frac{2y_1y_2}{\sigma_1^2}+\frac{y_2^2}{\sigma_1^2}+\frac{y_2^2}{\sigma_2^2}=y_2^2\underbrace{\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}_{b^2}-\frac{2y_1y_2}{\sigma_1^2}+y_1^2+\frac{y_1^2}{\sigma_1^2}=$$

$$=\left(y_2^2b^2-\frac{2y_1y_2}{\sigma_1^2}\cdot\frac{b}{b}+\frac{y_1^2}{\sigma_1^4b^2}\right)-\frac{y_1^2}{\sigma_1^4b^2}+\frac{y_1^2}{\sigma_1^2}$$

$$b^2=\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}$$

$$a^2=\frac{1}{\sigma_1^4}\cdot\frac{1}{b^2}=\frac{\sigma_2^2}{(\sigma_1^2+\sigma_2^2)\sigma_1^2}$$

$$c=\frac{1}{\sigma_1^2}-\frac{1}{a^2}=\frac{1}{\sigma_1^2}\left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)=\frac{1}{\sigma_1^2+\sigma_2^2}$$

$$f_Y(y_1)=\frac{e^{-\frac{y_1^2c}{2}}}{\sqrt{2\pi}\sigma_1\sigma_2b}=\frac{e^{-\frac{y_1^2}{2}\cdot\frac{1}{\sigma_1^2+\sigma_2^2}}}{\sqrt{2\pi}\cancel{\sigma_1\sigma_2}\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\cancel{\sigma_1\sigma_2}}}=\frac{e^{-\frac{y_1^2}{2(\sigma_1^2+\sigma_2^2)}}}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}$$ - плътност на

$\mathcal{N}(0,\,\sigma_1^2+\sigma_2^2)\Rightarrow Y_1=X_1+X_2\in\mathcal{N}(0,\,\sigma_1^2+\sigma_2^2)$.

Г. Гама разпределение

<u>Дефиниция</u>: (Гама разпределени случайни величини) Казваме, че случайната непрекъснатa величина X е гама разпределена с параметри $\alpha,\,\beta>0$ и бележим $X\in\Gamma(\alpha,\beta)$, ако има плътност $f_X(x)=\begin{cases}\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}, & x>0\\ 0, & x\le 0\end{cases}$, където

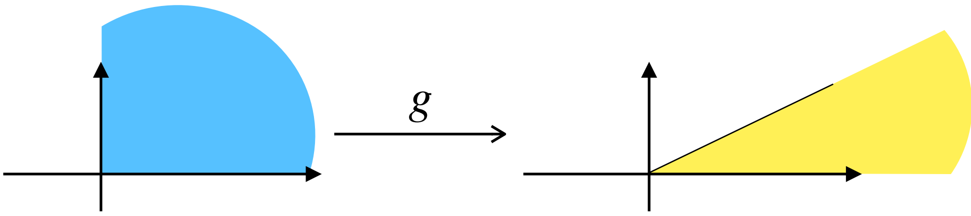
$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx.$$

$\oplus$ $\alpha=1, X\in\Gamma(1,\beta)=Exp(\beta)$. $\quad f_X(x)=\dfrac{\beta e^{-\beta x}}{\Gamma(1)}=\beta e^{-\beta x},\ x>0$

Твърдение: Ако $X_1 \in \Gamma(\alpha_1, \beta)$ и $X_2 \in \Gamma(\alpha_2, \beta)$ и $X_1 \perp\!\!\!\perp X_2$, то $X_1 + X_2 \in \Gamma(\alpha_1 + \alpha_2, \beta)$

Следствие: $X_i \in \Gamma(\alpha_i, \beta),\ i = 1, \dots, n$ и $X_1, \dots, X_n$ са независими в съвкупност, то $X_1 + \ldots + X_2 \in \Gamma(\alpha_1 + \ \ldots \ + \alpha_n, \beta)$

Упътване: $\begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = X_2 \end{cases}$, $f_X(x_1, x_2) = \dfrac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\beta(x_1+x_2)},\ x_1, x_2 \geq 0.$

$D_X =$ 

Свойства: Нека $X_1, X_2, \dots, X_n$ са независими в съвкупност експоненциално разпределени случайни величини $\in Exp(\beta) \sim \Gamma(1, \beta)$. Тогава

$$H = \sum_{i=1}^{n} X_i \sim \Gamma(n, \beta)$$

$$\mathbb{E}H = n\mathbb{E}X_1 = \frac{n}{\beta},\quad DH = \frac{n}{\beta^2}$$

Най-общо: $X \in \Gamma(\alpha, \beta) \Rightarrow \mathbb{E}X = \dfrac{\alpha}{\beta},\quad DX = \dfrac{\alpha}{\beta^2}$

Д. Хи квадрат разпределение

Дефиниция: Казваме че една случайна непрекъсната велиина X е Хи квадрат разпределена с параметър n и бележим $X \in \mathcal{X}^2(n)$, ако има плътност

$$f_X(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Ако $X \in \mathcal{X}^2(n)$, то $X \in \Gamma\left(\dfrac{n}{2}, \dfrac{1}{2}\right)$.

СЕМ, лекция 11

(2020-12-10)

Дефиниция: Две случайни величини са равни тогава и само тогава, когато имат еднакви функции на разпределение $X = Y \Leftrightarrow F_X = F_Y$. Ако X, Y са непрекъснати случайни величини, то от дясната страна на $\Leftrightarrow$ може да сложим и равенство на плътностите $f_X = f_Y$.

Твърдение: Нека $Z_1, Z_2, \ldots, Z_n$ са независими в съвкупност, стандартно нормално разпределени, случайни величини. $Z_i \in \mathcal{N}(0,1), \forall 1 \leq i \leq n$. Тогава

$X = Z_1^2 + Z_2^2 + \ldots + Z_n^2 = \sum_{i=1}^{n} Z_i^2 \in \mathcal{X}^2(n)$.

Доказателство: Ще докажем, че $Z_1^2 \in \mathcal{X}^2(1) = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Понеже Z_i^2 са независими и еднакво разпределени, то от предишната лекция

$\Rightarrow \sum_{i=1}^{n} Z_i^2 \in \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, което по дефиниция е $\mathcal{X}^2(n)$.

Т.е. трябва да докажем само, че $Z_1^2 = \mathcal{X}^2(1)$. Този факт не може да се докаже директно със смяна, защото функцията, която имаме $Y_1 = Z_1^2 = g(Z_1), g(x) = x^2$, която функция g не е монотонна и еднозначна. Ще трябва да използваме директен подход.

Ще се опитаме да пресметнем плътността на Y_1 и от нейната производна да намерим функцията на разпределението.

$$x \geq 0, \mathbb{P}(Y_1 < x) = \mathbb{P}(Z_1^2 < x) = \mathbb{P}(-\sqrt{x} < Z_1 < \sqrt{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{y^2}{2}} dy =$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{y^2}{2}} dy.$$

$$f_{Y_1}(x) = \frac{\partial}{\partial x} \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2\sqrt{x}} e^{-\frac{x}{2}} = \frac{x^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{x}{2}} = \frac{x^{-\frac{1}{2}} e^{-\frac{x}{2}}}{\Gamma\left(\frac{1}{2}\right)},$$ което е

плътността на $\mathcal{X}^2(1)$.

E. t-разпределение

Случайна величина $Y = \frac{Z}{\sqrt{\frac{S}{n}}}$, където $Z \in \mathcal{N}(0,1)$, $Z \perp\!\!\!\perp S$ и $S \in \mathcal{X}^2(n)$, се нарича t-разпределена случайна величина с n степени на свобода.

$\oplus$ $X_1, \ldots, X_n \in N(0,1)$ независими. Означаваме $\overline{X} = \frac{1}{n}, \ \sum_{i=1}^{n} X_i \sim N\left(0, \frac{1}{n}\right)$

$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2, \ \underbrace{S \perp\!\!\!\perp \overline{X}}, nS \in \mathscr{X}^2(n).$

Видове сходимост на случайни величини

$X_n : \Omega \to \mathbb{R}, n \geq 1$ и $X : \Omega \to \mathbb{R}$ са случайни величини във вероятностното пространство $V = (\Omega, \mathscr{A}, \mathbb{P})$ (т.е. имаме едни единствено вероятностно пространство и X_i, $i = \overline{1,n}$, X са функции на елементарни събития в числата)

<u>Дефиниция</u>: (**Сходимост почти сигурно (п.с.)**) Казваме, че $X_n \xrightarrow[n\to\infty]{\text{п.с.}} X \Leftrightarrow \mathbb{P}(L) = 1$, където събитието $\underset{\in \mathscr{A}}{L} = \{ \lim\limits_{n\to\infty} X_n = X\} = \{w \in \Omega : \lim\limits_{n\to\infty} X_n(w) = X(w)\}$.

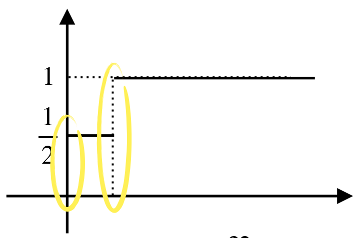
<u>Дефиниция</u>: (**Сходимост по вероятност**) Казваме, че $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \Leftrightarrow \forall \varepsilon > 0 :$ $\lim\limits_{n\to\infty} \mathbb{P}(A_n, \varepsilon) = 0$, където $A_{n,\varepsilon} = \{ |X_n - X| > \varepsilon\} = \{w \in \Omega : |X_n(w) - X(w)| > \varepsilon\}$

<u>Дефиниция</u>: Ако F_X е функция на разпределение, то с C_{F_X} означаваме всички точки x, за които F е непрекъсната в x. $C_{F_X} = \{x \in \mathbb{R} : \ F_X$ е непрекъсната в $x\}$ и $x \in C_{F_X} \Leftrightarrow \mathbb{P}(X = x) = 0$.

$\oplus$ Ако X е непрекъсната случайна величина, то $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R} \Rightarrow C_{F_X} = \mathbb{R}$

<u>Дефиниция</u>: (**Сходимост по разпределение**) Казваме, че $X_n \xrightarrow{d} X \Leftrightarrow \lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x), \forall x \in C_{F_X}$ (тук никъде не споменаваме вероятностно пространство).

$$X = \begin{cases} 1, p = \frac{1}{2} \\ 0, p = \frac{1}{2} \end{cases}$$



<u>Твърдение</u>: Нека $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ и $A = \bigcup\limits_{i=1}^{\infty} A_i$. Тогава $\mathbb{P}(A) = \lim\limits_{n\to\infty} \mathbb{P}(A_i)$.

Нека $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ и $A = \bigcap\limits_{i=1}^{\infty} A_i$. Тогава $\mathbb{P}(A) = \lim\limits_{n\to\infty} \mathbb{P}(A_i)$.

Теорема: Нека $(X_n)_{n=1}^{\infty}$ е редица от случайни величини и X е случайна величина.

а) Ако $X_n \xrightarrow[n\to\infty]{\text{п.с.}} X \Rightarrow X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$

б) Ако $X_n \xrightarrow[n\to\infty]{\mathbb{P}} \Rightarrow X_n \xrightarrow[n\to\infty]{d} X$

в) Обратните индикации на а) и б) не са верни.

Доказателство:

а) Знаем, че $1 = \mathbb{P}(L)$, където $L = \{\lim\limits_{n\to\infty} X_n = X\} \stackrel{?}{=} \bigcap\limits_{r=1}^{\infty}\bigcup\limits_{n=1}^{\infty}\bigcap\limits_{k\geq n} A^c_{k,\frac{1}{r}}$, където

$\Pi_{k,r} = A^c_{k,\frac{1}{r}} = \{\, |X_k - X| \leq \dfrac{1}{r}\}$

За фиксирано k: $\quad \ldots \supseteq \Pi_{k,r-1} \supseteq \Pi_{k,r} \supseteq \Pi_{k,r+1} \supseteq \ldots$

Въвеждаме $B_{n,r} = \bigcap\limits_{k\geq n} \Pi_{k,r}:\quad \ldots \supseteq B_{k,r-1} \supseteq B_{k,r} \supseteq B_{k,r+1} \supseteq \ldots$ за фиксирано k.

Но при фиксирано r имаме следното: $\ldots \subseteq B_{k-1,r} \subseteq B_{k,r} \subseteq B_{k+1,r} \subseteq \ldots$

$L \stackrel{?}{=} \bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=1}^{\infty} B_{n,r}$.

Въвеждаме още един запис $C_r = \bigcup\limits_{n=1}^{\infty} B_{n,r} \Rightarrow L \stackrel{?}{=} \bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=1}^{\infty} B_{n,r} = \bigcap\limits_{r=1}^{\infty} C_r$

За фиксирано r: $\quad \ldots \supseteq C_{r-1} \supseteq C_r \supseteq C_{r+1} \supseteq \ldots$

Следователно $L \stackrel{?}{=} C = \bigcap\limits_{r=1}^{\infty} C_r$

Нека $\overline{w} \in L$ ще докажем, че то принадлежи и на C.

От допускането $\Rightarrow \forall r \geq 1,\ \exists n_r L n > n_r,\ |X_n(\overline{w}) - X(\overline{w})| \leq \dfrac{1}{r} \Rightarrow$

$\overline{w} \in \Pi_{n,r},\ \forall n \geq n_r \Rightarrow \overline{w} \in B_{n_r,r} \subseteq \bigcup\limits_{n=1}^{\infty} B_{n,r} = C_r,\ \forall r \Rightarrow w \in C = \bigcap\limits_{r=1}^{\infty} C_r$

Нека $\overline{w} \in C \Rightarrow \overline{w} \in C_r,\ \forall r \geq 1 \Rightarrow \exists n_r:\ \overline{w} \in B_{n_r,\,r},\ \forall n \geq n_r \Rightarrow \overline{w} \in B_{n,r},\ \forall n \geq n_r$

$\Rightarrow \overline{w} \in \Pi_{k,r},\ \forall k \geq n_r$

$|X_k(\overline{w}) - X(\overline{w})| \le \frac{1}{r}, \forall k \ge n_r.$

Успяхме да покажем, че $L = C$.

$1 = \mathbb{P}(L) = \mathbb{P}(C) \Rightarrow \mathbb{P}(C_r) = 1, \forall r; \quad C \subseteq C_r, \forall r \ge 1$

$1 = \mathbb{P}(C_r) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_{n,r}\right) =$ $\qquad \ldots \subseteq B_{n,r} \subseteq B_{n+1,r} \subseteq \ldots$

$= \lim_{n\to\infty} \mathbb{P}(B_{n,r}) \le$ $\qquad \ldots B_{n,r} \subseteq \Pi_{n,r}$

$\le \lim_{n\to\infty} \mathbb{P}(\Pi_{n,r}) \le 1$

$\Rightarrow \lim_{n\to\infty} \mathbb{P}(\Pi_{n,r}) = \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \le \frac{1}{r}\right) = 1 \Rightarrow \lim_{n\to\infty} \mathbb{P}(A_{n,r}) = 0$ $\square$

б) $X_n \xrightarrow[n\to\infty]{\mathbb{P}} \overset{?}{\Rightarrow} X_n \xrightarrow[n\to\infty]{d} X.$

От първата сходимост $\Rightarrow \forall \varepsilon > 0: \lim_{n\to\infty} \mathbb{P}(A_{n,\varepsilon}) = 0, A_{n,\varepsilon} = \{|X_n - X| > \varepsilon\}$

$\varepsilon = \frac{1}{r}$, достатъчно е да разгледаме само тези ε, тъй като ако $\varepsilon \in \left(\frac{1}{r}, \frac{1}{r-1}\right)$, то $A_{n,\frac{1}{r-1}} \subseteq A_{n,\varepsilon} \subseteq A_{n,\frac{1}{r}}$.

Втората сходимост (по разпределение) означава, че $F_{X_n} \to F_X$, за $\forall x \in C_{F_X}$. Т.е. избираме $x \in C_{F_X}$ и целим да докажем, че $\lim_{n\to\infty} \mathbb{P}(X_n < x) = \mathbb{P}(X < x)$

Фиксираме $\varepsilon > 0$.

$\{X < x - \varepsilon\} \cap A_{n,\varepsilon} \subseteq \{X_n < x\} \cap A_{n,\varepsilon}^c \subseteq \{X_n < x\} = \{X_n < x\} \cap \overbrace{(A_{n,\varepsilon} \cup A_{n,\varepsilon}^c)}^{=\Omega} =$

$= \{X_n < x\} \cap A_{n,\varepsilon}^c \cup \{X_n < x\} \cap A_{n,\varepsilon} \subseteq \{X < x + \varepsilon\} \cup A_{n,\varepsilon}$

$\underbrace{\underbrace{\mathbb{P}(X < x - \varepsilon \cap A_{n,\varepsilon}^c)}_{=\mathbb{P}(X<x-\varepsilon)-\mathbb{P}(X<x-\varepsilon\cap A_{n,\varepsilon})}}_{\ge \mathbb{P}(X<x-\varepsilon)-\mathbb{P}(A_n,e)} \le \mathbb{P}(X_n < x) \le \mathbb{P}(X < x + \varepsilon) + \mathbb{P}(A_{n,\varepsilon})$

$$\Rightarrow \underbrace{\mathbb{P}(X < x - \varepsilon)}_{F_X(x-\varepsilon)} - \underbrace{\mathbb{P}(A_{n,\varepsilon})}_{\underset{n\to\infty}{\longrightarrow} 0} \leq \underbrace{\mathbb{P}(X_n < x)}_{F_{X_n}} \leq \underbrace{\mathbb{P}(X < x + \varepsilon)}_{F_X(x+\varepsilon)} + \underbrace{\mathbb{P}(A_{n,\varepsilon})}_{\underset{n\to\infty}{\longrightarrow} 0}$$

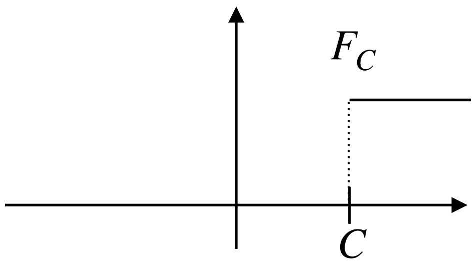
$$\Rightarrow F_X(x-\varepsilon) = \mathbb{P}(X > x - \varepsilon) \leq \lim_{n\to\infty} \mathbb{P}(X_n < x) \leq \mathbb{P}(X < x + \varepsilon) = F_X(X+\varepsilon)$$

$\Rightarrow$ При $\varepsilon \to 0$: $F_X(x) \leq \lim\limits_{n\to\infty} \mathbb{P}(X_n < x) \leq F_X(x) \Rightarrow \lim\limits_{n\to\infty} (X_n < x) = F_X(x)$

$F_{X_n}(x) \to F_X(x)$ $\square$

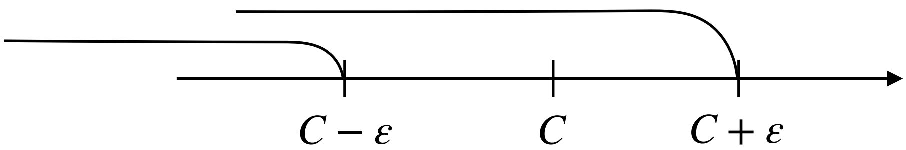
<u>Твърдение</u>: Ако $X_n \overset{d}{\to} \underset{const.}{C}$, то и $X_n \overset{\mathbb{P}}{\to} C$.

<u>Доказателство</u>: $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_C(x), \forall x \in \mathbb{R}\backslash\{C\}$



Цел: $\forall \varepsilon > 0 : \ \mathbb{P}(|X_n - C| \leq \varepsilon) \underset{n\to\infty}{\longrightarrow} 1$

$$\mathbb{P}(|X_n - C| \leq \varepsilon) = \mathbb{P}(X_n \leq C + \varepsilon) - \mathbb{P}(X_n \leq C - \varepsilon)$$



Но $\mathbb{P}(X_n \leq C - \varepsilon) = F_C(x - \varepsilon) = 0 \Rightarrow \lim\limits_{n\to\infty} \mathbb{P}(|X_n - C| \leq \varepsilon) = \lim\limits_{n\to\infty} \mathbb{P}(X_n \leq C + \varepsilon)$

$$1 \geq \mathbb{P}(X_n \leq C + \varepsilon) \geq \mathbb{P}(X_n < C + \varepsilon) \underset{n\to\infty}{\longrightarrow} 1 = F_C(C + \varepsilon)$$

$$\Rightarrow 1 = \lim_{n\to\infty} \mathbb{P}(|X_n - C| \leq \varepsilon).$$

Неравенство на Чебишев

<u>Твърдение</u>: (**Чебишев**) Нека X е случайна величина с очакванр $\mathbb{E}X$ и дисперсия DX. Нека $a > 0$. Тогава $\mathbb{P}(|X - \mathbb{E}X| > a) \leq \dfrac{DX}{a^2}$

Доказателство: $A = \{ |X - \mathbb{E}X| > a\} = \{(X - \mathbb{E}X)^2 > a^2\}$
$\mathbb{P}(A) \leq ?$

$$DX = \mathbb{E}(X - \mathbb{E}X)^2.1 = \mathbb{E}(X - \mathbb{E}X)^2.1_A + \mathbb{E}(X - \mathbb{E}X)^2.1_{A^c} \geq \mathbb{E}(X - \mathbb{E}X)^2.1_A \geq$$
$$\geq a^2\mathbb{E}1_A = a^2\mathbb{P}(A) \Rightarrow \mathbb{P}(A) \leq \frac{DX}{a^2} \qquad \square$$

$\oplus$ $a = b\sqrt{DX} \Rightarrow \mathbb{P}(|X - \mathbb{E}X| > b\sqrt{DX}) \leq \dfrac{1}{b^2}$

Закон за големите числа (ЗГЧ)

Дефиниция: Нека имаме редица от еднакво разпределени и независими (i.i.d.) случайни величини $(X_i)_{i=1}^{\infty}$ с очаквания съответно $\mathbb{E}X_i$. Казваме, че за X е изпълнено (слаб) ЗГЧ, ако

$$\frac{\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Казваме, че за X е изпълнено (усилен) ЗГЧ, ако $\dfrac{\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)}{n} \xrightarrow[n\to\infty]{\text{п.с.}} 0.$

$\oplus$ Ако $\mathbb{E}X_i = \mathbb{E}X_1 = c, \forall i \geq 1$, тогава $\dfrac{\sum_{i=1}^{n}X_i}{n} \xrightarrow[n\to\infty]{\mathbb{P}(\text{п.с.})} \mathbb{E}X_1 = c$

Дефиниция: Наричаме $X = (X_i)_{i=1}^{\infty}$ редица от независими в съвкупност и еднакво разпределени (НЕР) случайни величини, ако $X_i \overset{d}{=} X_1, \forall i$ и всички случайни величини са независими. $(F_{X_i} = F_{X_1}, \mathbb{E}X_i = \mathbb{E}X_1, \ldots)$

Теорема: Нека $X = (X_i)_{i=1}^{\infty}$ от НЕР случайни величини. Нека в допълнение

$\mathbb{E}|X_1| < \infty$ и $\mathbb{E}X_1 = \mu \in (-\infty, \infty)$.

Тогава $\dfrac{\sum_{i=1}^{n}X_i}{n} \xrightarrow{\mathbb{P}(\text{п.с.})} \mu = \mathbb{E}X_1$.

СЕМ, лекция 12

(2020-12-17)

Централна Гранична Теорема (ЦГТ)

ЗГЧ: $\left(X_i\right)_{i=1}^{\infty}$ е редица от независими и еднакво разпределени (i.i.d.) случайни величини с $\mathbb{E}|X_1| < \infty$, $\mathbb{E}X_1 = \mu$ и $\sigma = \sqrt{DX_1}$.

$$S_n = \sum_{j=1}^{n} X_j \Rightarrow \frac{S_n}{n} \xrightarrow[n\to\infty]{\text{п.с.}(\mathbb{P})} \mu; \qquad \frac{S_n}{n} - \mu \approx \frac{\sigma}{\sqrt{n}} H_n$$

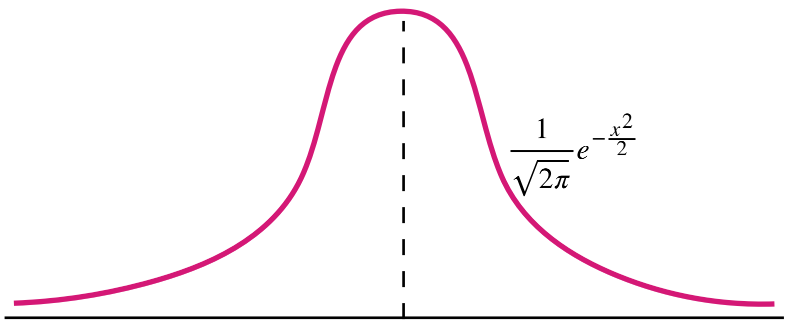
<u>Теорема</u>: (**ЦГТ**). Нека $X = \left(X_i\right)_{i=1}^{\infty}$ е редица от независими, еднакво разпределени случайни величини със $\sigma^2 = DX_1 < \infty$ и $\mu = \mathbb{E}X_1$. Тогава

$$Z_n \stackrel{def.}{:=} \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} Z \in \mathcal{N}(0,1).$$

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \mathbb{P}(Z < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \,\mathrm{d}\, x.$$

$$\mathbb{P}(Z \geq x) = 1 - \Phi(x) = \overline{\Phi}(x).$$

ЦГТ измерва каква е грешката в ЗГЧ. Т.е. ние знаем, че $\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu$, почти сигурно (траекторно) и скалираме грешката по $\frac{\sigma}{\sqrt{n}}$ и виждаме, че се държи като нормално разпределение.



Често се дава за пример пресмятането на средно IQ за група хора или тяхна височина и т.н., но за да бъде IQ-то нормална разпределена случайна величина е необходимо да бъде средно аритметичен ефект от множество независими характеристики. А дали това е така?

Във физиката ЦГТ може да ни даде каква е грешката между акумулиран ефект и среден ефект, който сме измерили в някаква система.

Следствия:

$\oplus$ $\quad Z_n := \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, където $\mu = \mathbb{E}X_1$, $\sigma = \sqrt{DX_1}$. ЦГТ гласи следното:

Ако се интересуваме от

$$\mathbb{P}\left(Z_n \in (a,b)\right) = \mathbb{P}\left(S_n \in (n\mu + a\sigma\sqrt{n}, n\mu + b\sigma\sqrt{n})\right) \sim \mathbb{P}\left(Z \in (a,b)\right) =$$

$$= \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \times \underbrace{\int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y}_{\substack{\text{този интеграл}\\ \text{не се интересува}\\ \text{от това с какви}\\ \text{случайни величини}\\ \text{сме стартирали.}\\ \text{Той зависи само от } a \text{ и } b}} .$$

$$\mathbb{P}(Z_n < a) \xrightarrow[n\to\infty]{} \Phi(a) = \mathbb{P}(Z < a); \qquad \mathbb{P}(Z_n \geq a) \xrightarrow[n\to\infty]{} \overline{\Phi}(a) = 1 - \Phi(a)$$

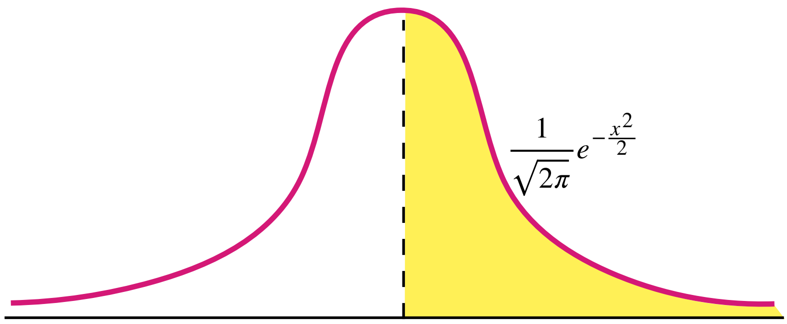
$$\mathbb{P}(Z_n < \mu) \xrightarrow[n\to\infty]{} \Phi(\mu).$$

Т.е. ЦГТ дава инструмент, с който може да приближаваме вероятности.

$\oplus$ $\quad \mu = 0$, $\sigma^2 = 1$, тогава $\dfrac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} Z \in \mathcal{N}(0,1)$, където $S_n = \sum_{j=1}^{n} X_j$.

$$\mathbb{P}\left(S_n \in (a\sqrt{n}, b\sqrt{n})\right) \sim \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y.$$

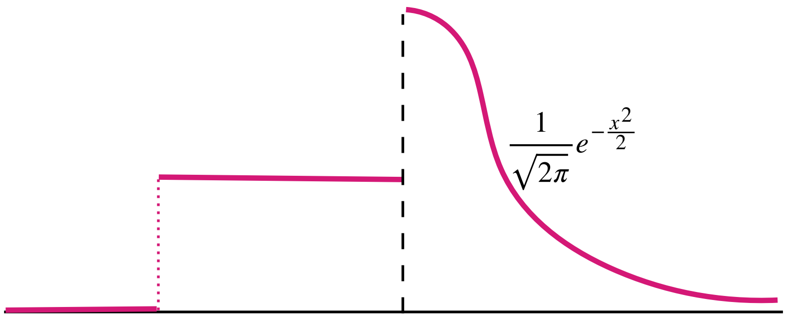
$$\mathbb{P}(S_n > 0) = \mathbb{P}\left(\frac{S_n}{\sqrt{n}} > 0\right) \sim \mathbb{P}(Z > 0) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{y^2}{2}}\,\mathrm{d}y = \frac{1}{2}.$$



⊕ Хвърляме зарче $6\,000\,000$ пъти. Каква е вероятността измежду тези $6\,000\,000$ пъти да сме хвърлили повече от $1\,000\,000$ пъти 6-ца? Решение: (*виж последния пример от зад. 10 от домашното*).

$$\mu = 0, \sigma^2 = 1, S_n = \sum_{j=1}^{n} X_j$$

$$\mathbb{P}(S_n > 0) \sim \frac{1}{2}, f_{X_1}(x) = \begin{cases} \frac{1}{2}e^{-x}, & x > 0 \\ \frac{1}{4}, & x \in (-2, 0) \\ 0, & \text{иначе} \end{cases}$$



$$\mathbb{E}X_1 = \int_{-2}^{0} x\frac{1}{4}\,\mathrm{d}\,x + \int_{0}^{\infty} x\frac{1}{2}e^{-x}\,\mathrm{d}\,x = \frac{1}{4}\frac{x^2}{2}\Big|_{-2}^{0} - \frac{1}{2}\int_{0}^{\infty} x\,\mathrm{d}\,e^{-x} =$$

$$= -\frac{1}{4}\times\frac{4}{2} - \underbrace{\frac{1}{2}xe^{-x}\Big|_{0}^{\infty}}_{=0} + \frac{1}{2}\int_{0}^{\infty} e^{-x}\,\mathrm{d}\,x = -\frac{1}{2} + \frac{1}{2}\left(-e^{-x}\right)\Big|_{0}^{\infty} = -\frac{1}{2} + \frac{1}{2} = 0.$$

Т.е. дори и за случайни величини, които са много далеч от симетрия, ако ги сумираме всички от тях, то вероятността да видим нещо положително е $\sim \frac{1}{2}$.

$$X_i = \begin{cases} 1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}, \quad S_n = \mathscr{D}_n - \Lambda_n = \sum_{j=1}^{n} X_j.$$

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} > a\right) \sim \overline{\Phi}(a),$$

$$\mathbb{P}\left(\mathscr{D}_n > \Lambda_n + a\sqrt{n}\right) = \mathbb{P}\left(\mathscr{D}_n > \frac{n}{2} + \frac{a}{2}\sqrt{n}\right) \sim \overline{\Phi}(a).$$

Функция на моментите (Ф.М.)

Дефиниция: Нека X е случайна величина. Ако $\mathbb{E}e^{tX}$ съществува за $t \in (-\varepsilon, \varepsilon)$ и някое $\varepsilon > 0$, то $M_X(t) = \mathbb{E}e^{tX}$ за $t \in (-\varepsilon, \varepsilon)$ се нарича функция на моментите.

$\oplus$ $\mathbb{E}e^{tX} = \sum_i e^{tx_i}\mathbb{P}(X = x_i)$ - винаги съществува за $\forall t$, ако стойностите са краен брой. Но ако не са, тази сума можеда не е сумируема и да отива към ∞.

Но ако вчемем $x_i = j$ и $\mathbb{P}(X = j) = \dfrac{1}{j^2}$, то няма да може да направим сумировката за $t > 0$.

Ако имаме непрекъсната случайна величина, функцията на моментите се изчислява чрез интеграла $\mathbb{E}e^{tX} = \int_{-\infty}^{\infty} e^{tx}f_X(x)\mathrm{d}\,x$, който може да съществува само за някаква част от t, но е важно да съществува за $t \in (-\varepsilon, \varepsilon)$, за да може да го наречем функция на моментите.

$X \sim Unif(0, 1)$.

$$M_X(t) = \mathbb{E}e^{tX} = \int_0^1 e^{tx} \times f_X(x)\mathrm{d}\,x = \int_0^1 e^{tx}\frac{1}{1-0}\,\mathrm{d}\,x = \frac{1}{t}\int_0^1 e^{tx}\,\mathrm{d}\,tx = \frac{1}{t}e^{tx}\Big|_0^1 = \frac{e^t - 1}{t}$$

е добре дефинирано за всяко $t \in \mathbb{R}$.

Винаги когато нашата случайна величина може да приема краен брой стойности или възможните стойности са в някакъв компакт или някакво крайно множество (в случая са от 0 до 1), то винаги функцията на моментите е добре дефиниране за всяко t.

Дефиниция: X е случайна величина. Тогава:

а) $\mathbb{E}X^k$ се нарича момент от ред $k \geq 1$;
б) $\mathbb{E}|X|^k$ Се нарича абсолютен момент от ред k;
в) $\mathbb{E}(X - \mathbb{E}X)^k$ се нарича централен момент от ред k; $\quad DX = \mathbb{E}(X - \mathbb{E}X)^2$
г) $\mathbb{E}|X - \mathbb{E}X|^k$ се нарича абсолютен централен момент от ред k.

Свойства на M_X. Ще допускаме, че $M_X(t)$ е добре дефинирана за $t \in (-\varepsilon, \varepsilon)$

а) $M_X(0) = \mathbb{E}e^{0.X} = \mathbb{E}1 = 1$;
б) $\dfrac{\partial}{\partial t^k}M_X(t)\Big|_{t=0} = \mathbb{E}X^k$, за $\forall k \geq 1$;

в) $M_X(t) = \mathbb{E}e^{tX} \overset{\text{ред на Тейлър}}{=} \mathbb{E}\sum_{k=0}^{\infty}\frac{t^k}{k!}X^k = \sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbb{E}X^k;$

г) Ако $M_{X_n}(t) \underset{n\to\infty}{\longrightarrow} M_X(t)$, за $t \in (-\varepsilon, \varepsilon)$, където $(X_n)_n$ е редица от случайни величини, то $X_n \xrightarrow[n\to\infty]{d} X$;

д) $X \overset{d}{=} Y \Leftrightarrow M_X = M_Y$;

е) Ако $X \perp\!\!\!\perp Y$ и M_X, M_Y са добре дефинирани за $t \in (-\varepsilon, \varepsilon)$, то
$M_{X+Y}(t) = \mathbb{E}e^{t(X+Y)} = \mathbb{E}e^{tX}e^{tY} = M_X(t)M_Y(t) \overset{X\perp\!\!\!\perp Y}{=} \mathbb{E}e^{tX}\mathbb{E}e^{tY}$.

Нека например X, Y са непрекъснати с плътности f_X, f_Y. Как да докажем, че $M_X(t) = M_Y(t)$.

1. $f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x-y)f_Y(y)\mathrm{d}\,y;$

2. $\int_{-\infty}^{\infty} e^{tX}f_{X+Y}(x)\mathrm{d}\,x = \int_{-\infty}^{\infty} e^{tX}f_X(x)\mathrm{d}\,x \times \int_{-\infty}^{\infty} e^{tX}f_Y(x)\mathrm{d}\,x;$

ж) Ако $Y = aX + b$, то $M_Y(t) = e^{bt}M_X(at)$, за всяко t, такова че $M_X(at)$ е добре дефинирано.

Ако M_X е добре дефинирано за $t \in (-\varepsilon, \varepsilon)$, то $M_X(at)$ е добре дефинирано за $-\varepsilon < at < \varepsilon$ и следователно $M_Y(t)$ е добре дефинирано за $-\frac{\varepsilon}{|a|} < t < \frac{\varepsilon}{|a|}$.

$M_Y(t) = \mathbb{E}e^{t(aX+b)} = \mathbb{E}e^{bt}e^{taX} = e^{bt}\mathbb{E}e^{atX} = e^{bt}M_X(at)$.

<u>Твърдение</u>: $X \in \mathcal{N}(\mu, \sigma^2)$, то за $\forall t \in \mathbb{R}$, то $M_X(t) = e^{\mu t}e^{\frac{\sigma^2 t^2}{2}}$.

<u>Доказателство</u>: $X = \mu + \sigma Z$, където $Z \in \mathcal{N}(0, 1)$.

$$M_Z(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tX}e^{-\frac{x^2}{2}}\,\mathrm{d}\,x = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2}\,\mathrm{d}\,x \overset{x-t=y}{=} e^{\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,\mathrm{d}\,y = e^{\frac{t^2}{2}}$$

$M_X(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t}e^{\frac{\mu^2 t^2}{2}}, \forall t \in \mathbb{R}$.

<u>Доказателство</u> (**ЦГТ**): $\left(X_i\right)_{i=1}^{\infty}$ е редица от независими, еднакворазпределени случайни величини с $\mathbb{E}X_1 = \mu, DX_1 = \sigma^2$.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Z.$$

X_1 има функция на моментите. $M_{X_1}(t)$ е добре дефинирана за $t \in (-\varepsilon, \varepsilon)$

$\dfrac{S_n - n\mu}{\sigma\sqrt{n}} = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{j=1}^{n}\dfrac{X_j - \mu}{\sigma} = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{j=1}^{n} Y_j = \dfrac{V_n}{\sqrt{n}} =: W_n$, където сме положили $Y_j = \dfrac{X_j - \mu}{\sigma}, \forall j \geq 1$ и $(Y_j)_{j=1}^{\infty}$ са независими с еднакво разпределение сл. вел.

$$\mathbb{E}Y_1 = \frac{\mathbb{E}X_1 - \mu}{\sigma} = 0;\ DY_1 = \frac{DX_1}{\sigma^2} = 1.$$

$$M_{Y_1}(t) = e^{-\frac{\mu}{\sigma}t}M_{X_1}\left(\frac{t}{\sigma}\right), \qquad \varepsilon < \frac{t}{\sigma} < t\varepsilon \Rightarrow -\sigma\varepsilon < t < \sigma\varepsilon$$

M_{Y_1} е добре дефинирана за $|t| < \sigma\varepsilon$.

Нека фиксираме t. Ще докажем, че $M_{W_t} \xrightarrow[n\to\infty]{} M_Z(t) = e^{\frac{t^2}{2}}$.

$$M_{W_n}(t) = \mathbb{E}e^{\frac{t}{\sqrt{n}}V_n} \overset{\text{деф.}}{=} \mathbb{E}e^{\frac{t}{\sqrt{n}}\sum_{j=1}^{n} Y_j} \overset{\text{незав.}}{=} \prod_{j=1}^{n} M_{Y_j}\left(\frac{t}{\sqrt{n}}\right) \overset{Y_j \overset{d}{=} Y_1,\ \forall j}{=} \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Ако $\left|\dfrac{t}{\sqrt{n}}\right| < \sigma\varepsilon$, то $M_{W_n}(t)$ е добре дефинирано.

$$M_{W_n}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right], M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}e^{\frac{t}{\sqrt{n}}Y_1}.$$

$$e^{\frac{t}{\sqrt{n}}Y_1} = 1 + \frac{t}{\sqrt{n}}Y_1 + \frac{t^2}{2n}Y_1^2 + \frac{\theta(Y_1)t^3Y_1^3}{3!n^{\frac{3}{2}}}; \quad |\theta(Y_1)| \leq 1;$$

$$M_{Y_1}(t) = \mathbb{E}e^{\frac{t}{\sqrt{n}}Y_1} = 1 + \frac{t^2}{2n} + \frac{\mathbb{E}\theta(Y_1)t^3Y_1^3}{3!n^{\frac{3}{2}}};$$

$$M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}\mathbb{E}\theta(Y_1)Y_1^3;$$

$$|\theta(Y_1)Y_1^3| \le |Y_1|^3 \Rightarrow |\mathbb{E}\left(\theta(Y_1)Y_1^3\right)| \le \mathbb{E}|Y_1|^3 = \rho_3.$$

$$M_{W_a}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{t^2}{2n} + \frac{t^2}{2n}\left(\frac{\rho_3}{\sqrt{n}}\right)\right]^n =$$

$$= \left[1 + \frac{t^2}{2n}\left(1 + \underbrace{\frac{\rho_3}{\sqrt{n}}}_{\to 0}\right)\right]^n \sim \left(1 + \frac{t^2}{2n}\right)^n \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}$$

$$\Rightarrow \lim_{n\to\infty} M_{W_n}(t) = e^{\frac{t^2}{2}} = M_Z(t)$$

Свойство $\Rightarrow W_n \xrightarrow[n\to\infty]{d} Z;\qquad Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Z.$

$$\mathbb{P}\left(Z_n \in (a,b)\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y = \Phi(b) - \Phi(a).$$

$\oplus\qquad X_i = \begin{cases} 1, & p \\ 0, & 1-p=q \end{cases},\qquad X_i \in Ber(p)$

$\dfrac{S_n}{n} \xrightarrow[n\to\infty]{\text{п.с.}} p,\quad \mathbb{E}n = \left|\dfrac{S_n}{n} - p\right|$

$$\mathbb{P}\left(|\mathbb{E}n > \varepsilon|\right) = \mathbb{P}\left(\underbrace{\left|\frac{S_n - np}{\sqrt{n}\sqrt{pq}}\right|}_{n\to\infty} > \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) = \mathbb{P}\left(|Z| > \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) = 2\overline{\Phi}\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) \le$$

$$\le 2\overline{\Phi}\left(\frac{\sqrt{n}\varepsilon}{\frac{1}{2}}\right) = 2\overline{\Phi}(2\sqrt{n}\varepsilon) = 2\int_{2\sqrt{n}\varepsilon}^{+\infty} e^{-\frac{y^2}{2}}\,\mathrm{d}y.$$

Тоерема на Берн-Есеен:

Нека $(X_i)_{i=1}^{\infty}$ са независими и еднакво разпределени случайни величини с $\mathbb{E}X_1 = \mu$, $DX_1 = \sigma^2$ и $\mathbb{E}|X_1 - \mathbb{E}X_1|^3 = \rho_3$.

Тогава $\operatorname{Sup}_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) - \underbrace{\Phi(x)}_{=\mathbb{P}(Z<x)} \right| \le 0{,}4748 \frac{\rho^3}{\sigma^{\frac{3}{2}}\sqrt{n}}$.

Следствие: $X \in Bin(n, p)$, то $\lim_{n\to\infty} \mathbb{P}\left(\frac{X - np}{\sqrt{npq}} \le x\right) = \Phi(x)$

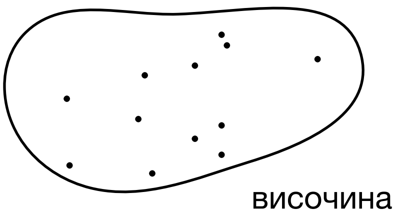
Доказателство: $X = \sum_{j=1}^{n} X_j$, $X_j \in Ber(p)$. Тогава прилагаме ЦГТ с $\mu = p$ и $\sigma^2 = pq = p(1-p)$.

СЕМ, лекция 13

(2021-01-07)

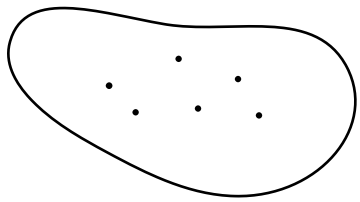
Статистика - точкови оценки

⊕ Всички българи

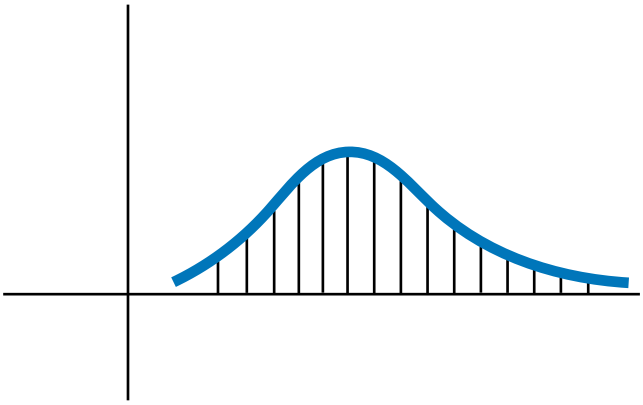


фиг. 1

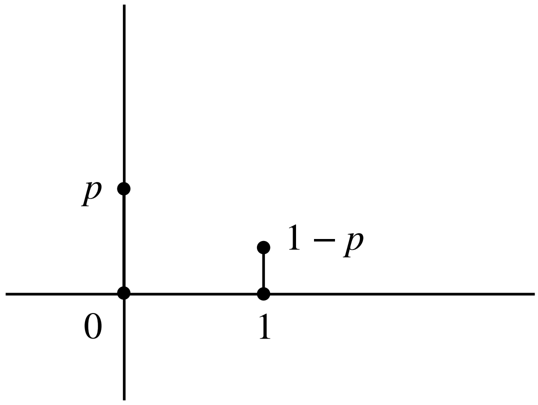
⊕ Гласоподаватели в САЩ



фиг. 2



фиг. 3



фиг. 4

Ако знаехме предварително височината на всеки от българите, щяхме да можем да си построим такава крива на разпределението, като от фиг. 1 и аналогично за гласоподавателите от фиг. 2.
Проблема е, че или е невъзможно или е твърде скъпо (по някой път обектите са цели функции или някакви сложни конфигурации). Целта е, наблюдавайки някаква подизвадка от обекти от цялата популация, да разберем нещо за синята крива на разпределението, за хистограмата, за средното или за пропорцията $\dfrac{p}{1-p}$ от фиг. 2 и т.н.

Най-ефективния начин за намирането на такава информация е като си направим една подизвадка от n човека, на които ще направим желаната характеристика и на база на тази информация ще се опитаме да извлечем нещо (дисперсия, средно и т.н.) за цялата популация.

Проблемът, който възниква е свързан с това да се направи избора напълно случайно. Т.е. всеки един обект от популацията да има равни шансове за избор спрямо останалите.

Ако работим само върху извадка от обекти с определен признак, то има риск да имаме пристрастие към нашите резултати и те няма да са показателни.

Постановка: Имаме случайна величина X с функция на разпределение F_X или плътност f_X. Искаме да разберем възможно най-много за някоя от характеристиките на X (μ, σ, F_X, f_X).

$\overrightarrow{X} = (X_1, \ldots, X_n)$, $(X_i)_{i=1}^n$ са независими случайни величини, които се реализрат чрез избор от генералната съвкупност (от всички изучавани обекти).

Цел: Някаква информация за X и най-вече F_X и f_X.

Допускания: Когато например изследваме гласоподавателите в САЩ (избира се между двама) ние имаме естествена рестрикция и знаем, че гласоподавател гласува за „партия 1“ или „партия 2“, което е конкретен клас разпределение (бернулиево - или за едната или за другата, ако е гласувал).

✤ X е някакъв клас разпределение, който се характеризира с някакъв вектор от параметри θ. Т.е. $F_X(x, \theta)$ и $f_X(x, \theta)$ (функцията на разпределение и плътност на X зависят от θ)

⊕ Ако допуснем, че $X \in \mathcal{N}(\mu, \sigma^2)$, където $\theta = (\mu, \sigma^2)$,
$f_X(x, \theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$; В някой ситуации може да допускаме и, че $\theta = \mu$ при $\sigma^2 = 1$ $\big(\mathcal{N}(\mu, 1)\big)$

⊕ (при гласоподавателите, както споменахме, нещата са още по прости)
$X \in \mathrm{Ber}(p)$, т.е. $\theta = p = \mathbb{P}(X = 1)$. Трябва да приближим възможно най-добре този параметър p;

⊕ Класа на Гама разпределение: $\theta = (\alpha, \beta)$;
⊕ Класа на експоненциално разпределение: $\theta = (\lambda)$ и т.н.

Ще се опитваме да търсим информация при някакви допускания за разпределение на неизвестната случайна величина (било то гласоподавател, клетки, функции или каквото и да е)

Цел: $\overrightarrow{X} = \underbrace{(X_1, \ldots, X_m)}_{\text{априори}}$ и искаме да оценим θ на базата на априорните слчайни величини. Означаваме с $\hat{\theta} = \hat{\theta}(\overrightarrow{X})$, което е случайна величина.
Ако имаме конкретна реализация $X_i = x_i$, то $\hat{\theta} = \hat{\theta}(x_1 \ldots, x_n)$ е число.

Дефиниция: Оценката $\hat{\theta} = \hat{\theta}(\overrightarrow{X})$ на неизвестното θ се нарича ТОЧКОВА ОЦЕНКА на θ или статистика на θ.

Въпроса сега е: Как може да направим добри точкови оценки?

A. МЕТОД НА МАКСИМАЛНОТО ПРАВДОПОДОБИЕ (ММП)

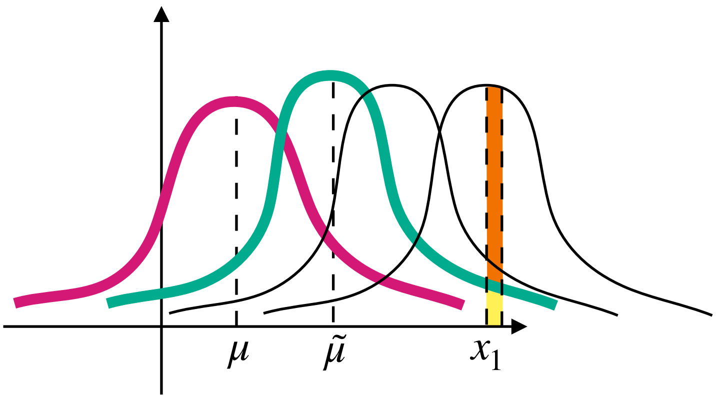
Този метод изхожда от следната схема:
Търсим информация за X с плътност $f_X(x, \theta)$ (допускаме, че я има), която зависи от някакви параметри θ. Наблюдаваме $\overrightarrow{X} = (X_1, \ldots, X_n)$ - n случайни обекта, за които сме извадили някакв бстойности, съответно $(x_1, \ldots, x_n)$. На базата на тези стойности трябва да конструираме по някакъв оптимизационен (смислен/облягащ се на някакви закономерности) начин - оценка за параметъра θ.

Първо ще погледнем какво е съвместното разпределение на $X_1, \ldots, X_n$ и това ще

бъде $f_{\overrightarrow{X}}(x_1, \ldots, x_n, \theta) \overset{\text{независими, еднакво разпр.}}{=} \underbrace{\prod_{j=1}^{n} f_X(x_j, \theta)}_{\text{съвместното разпределение}} \xrightarrow{\text{ще се опитаме максимизираме по } \theta}$

$\longrightarrow$ функция на максималното правдоподобие.

<u>Пример</u>: Нека например имаме $X \in \mathcal{N}(\mu, 1), X_1 = x_1$; оценка: $\hat{\theta} = x_1$ $(\theta = \mu)$



$$\mathbb{P}\left(X_1 \in (x_1 - \varepsilon, x_1 + \varepsilon)\right) = \frac{1}{\sqrt{2\pi}} \int_{x_1-\varepsilon}^{x_1+\varepsilon} e^{-\frac{(x-\theta)^2}{2}} \, \mathrm{d}\,x \approx \frac{2\varepsilon}{\sqrt{2\pi}} e^{-\frac{-(x_1-\theta)^2}{2}} = 2\varepsilon f_X(x_1, \theta)$$

$\operatorname*{Sup}_{\theta} \mathbb{P}\left(X_1 \in (x_1 - \varepsilon, x_1 + \varepsilon)\right) = 2\varepsilon f_X(x_1, x_1)$. Т.е. взимаме там (околността) където плътността достига своя максимум.

<u>Дефиниция</u>: $\overrightarrow{X}$ е вектор от n независими, еднакво разпределени (копия) на X. Нека X има плътност $f_X(x, \theta)$ (допускаме, че се параметризира от някакъв параметър θ), където $\theta \in \Theta$ (допустимо множество). Тогава МПП (максимално правдоподобие/максимално правдоподобно приближение) $\hat{\theta}$ на θ е този вектор/стойност, който/която удовлетворява:

$$\mathrm{L}(X, \hat{\theta}) = \operatorname*{Sup}_{\theta\in\Theta} \underbrace{\mathrm{L}(\overrightarrow{X}, \theta)}_{\text{функция на макс. правдоподобие}} \stackrel{def.}{=} \operatorname*{Sup}_{\theta\in\Theta}\prod_{j=1}^{n} f_X(x_j, \theta)$$

$$\mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(\underbrace{x_j}_{\text{или цялото } X_j}, \theta) = f_{\overrightarrow{X}}(\overrightarrow{X}, \theta)$$

$\oplus$ $\overrightarrow{X} = (x_1, \dots, x_n) = \mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(x_j, \theta)$

$$\operatorname*{Sup}_{\theta\in\Theta}\prod_{j=1}^{n} f_X(x_j, \theta) \qquad \frac{\partial \mathrm{L}}{\partial\theta} = 0 \longmapsto \frac{\partial \ln \mathrm{L}}{\partial\theta} = 0$$

$\oplus$ Ако $\underbrace{X \in \mathcal{N}(\mu, \sigma^2)}_{X \text{ е от класа}}$ и $n \to$ наблюдения, то трябва да решим системата:

$\begin{cases} \frac{\partial \mathrm{L}}{\partial\mu} = 0 \\ \frac{\partial \mathrm{L}}{\partial\sigma^2} = 0 \end{cases} \Rightarrow (\mu, \sigma^2)$ - намираме тези μ и σ^2, които максимизират функцията на максималното правдоподобие, но е много по удобно да го правим за $\begin{cases} \frac{\partial \ln \mathrm{L}}{\partial\mu} = 0 \\ \frac{\partial \ln \mathrm{L}}{\partial\sigma^2} = 0 \end{cases}$, като $\ln$ също е нарастваща функция.

$\Theta = \mathbb{R} \times (0, \infty) \leftarrow$ максимизираме системата върху тази област.

Но не винаги ще може да имаме функция на правдоподобие, която да е диференцируема.

$\oplus$ $X \in \mathcal{U}nif(0, \theta)$, $\overrightarrow{X} = (X-1, \dots, X_n) = (x_1, \dots, x_n)$.

Означаваме $X^* = \max_{1\le j\le n} x_j$ $\quad\Big($ДС (допустими стойности): $\Theta = (0, \infty)\Big)$

$$\mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(x_j, \theta) = \prod_{j=1}^{n} \frac{1}{\theta} 1_{[0,\theta]}(x_j) = \frac{1}{\theta^n} 1_{[0,\theta]}(x^*) = \begin{cases} 0, & \theta < x^* \\ \theta^{-n}, & x^* \le \theta \end{cases}, \text{ но}$$

последната функция не е диференцируема.

Но пък от друга страна, много лесно се максимизира:

$$\operatorname*{Sup}_{\theta>0} \mathrm{L}(\overrightarrow{X}, \theta) = \mathrm{L}(\overrightarrow{X}, x^*) = \frac{1}{(x^*)^n} \Rightarrow \hat{\theta} = \max(X_1, \ldots, X_n)$$

$\Big(\hat{\theta}$ се нарича точкова оценка по метода на максималното правдоподобие, ако $\mathrm{L}(\overrightarrow{X}, \hat{\theta}) = \operatorname*{Sup}_{\theta\in\Theta} \mathrm{L}(\overrightarrow{X}, \theta)\Big)$

Ще изведем оценките по максималното правдоподобие, в случая, в който $X \in \mathcal{N}(\mu, \sigma^2)$.

<u>Твърдение</u>: X е случайна величина от класа на нормално разпределените случайни величини. $\overrightarrow{X} = (X_1, \ldots, X_n)$ са независими еднакво разпределени наблюдения. Тогава $\hat{\mu} = \overline{X} = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} X_j$, МПО (макс. правд. оценка) на неизвестния параметър μ.

МПО за σ^2 е изразът (макс. правдоподобна оценка)

$a)$ $\quad \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (x_j - \mu)^2$, ако μ се допусне, че е известно;

б) $\quad \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (x_j - \hat{x}_n)^2$, ако μ не е известно.

Разликата е чувствителна, тъй като в $a)$ използваме истинското μ, докато в б) използваме оценка.

<u>Доказателство</u>: $\mathrm{L}(\overrightarrow{X}, \theta) = \displaystyle\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}}, \quad \theta = (\mu, \sigma^2)$

$$\overrightarrow{\mathrm{L}}(\overrightarrow{X}, \theta) = \ln \mathrm{L}(\overrightarrow{X}, \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \sum_{j=1}^{n} \frac{(x_j-\mu)^2}{2\sigma^2}$$

$$0 = \frac{\partial \tilde{\mathrm{L}}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu) \Rightarrow \hat{\mu} = \underbrace{\frac{1}{n}\sum_{j=1}^{n} X_j =: \overline{X_n}}_{\text{не зависи от } \sigma} \to \text{МПО}$$

$$0 = \frac{\partial \tilde{\mathrm{L}}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{j=1}^{n}(X_j - \mu)^2 \xrightarrow{\mu} \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2, \ \mu \text{ - известно}$$

$$\xrightarrow{\hat{\mu}} \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2, \ \mu \text{ - неизвестно}$$

Има две възможни оценки за $\hat{\sigma}^2$ (максимална правдоподобна оценка) според това дали знаем средното или не.

Б. МЕТОД НА МОМЕНТИТЕ

Тук вече няма нужда да допускаме съществуването на нищо друго освен съществуването на моментите.

$\oplus$ X е случайна величина с параметър $\theta \in \mathbb{R}$. Знаем, че средното $\mathbb{E}X = \mu(\theta)$. За $\overrightarrow{X}$ от ЗГЧ имаме $\overline{X}_n^{(1)} = \dfrac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\mathbb{P}(\text{п.с.})} \mathbb{E}X = \mu(\theta)$

Ако може да решим $\theta = \mu^{-1}\left(\overline{X}_n^{(1)}\right)$, където взимаме (като приближение) $\overline{X}_n = \mu(\theta)$, то θ ще е оценена по метода на моментите. Т.е. тук използваме ЗГЧ, за да оценим θ.

<u>Дефиниция</u>: Нека X е случайна величина, $F_X(x, \theta)$, $\theta = (\theta_1, \ldots, \theta_s)$. Нека имаме $\overrightarrow{X} = (X_1, \ldots, X_n)$ случайни величини, независими и разпределени като X (прототипи н случайна величина X).

Означаваме: $\overline{X}_n^{(k)} = \dfrac{1}{n}\sum_{j=1}^{n} X_j^k$ и $\mu^k(\theta) = \mathbb{E}X^k$. Тогава решението на системата

$\overline{X}^{(k)} = \mu^k(\theta)$, $1 \leq k \leq s$ за θ, се нарича оценка по метода на моментите.

$\oplus$ $X \in \mathcal{U}nif(0, \theta)$ и $\overrightarrow{X}$, $\overline{X}_n^{(1)} = \dfrac{1}{n}\sum_{j=1}^{n} X_j$

$$\mathbb{E}X = \frac{\theta}{2} \Rightarrow \overline{X}_n^{(1)} = \frac{\theta}{2} \Rightarrow \hat{\theta} = 2\overline{X}_n^{(1)} \qquad \text{ММО}$$

$$\theta = \max(X_1, \ldots, X_n) \qquad \text{ММП}$$

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, $\overline{X}_n^{(1)} = \mu = \mathbb{E}X$, ММО=ММП

$$\frac{1}{n}\sum_{j=1}^{n} X_j^2 = \overline{X}_n^{(2)} = \mathbb{E}X^2 = \sigma^2 + \mu^2 \Rightarrow \sigma^2 = \overline{X}_n^{(1)} = \left(X_m^{(1)}\right)^2 \qquad \text{ММО=ММП}$$

Трябва да проверим, че $X_n^{(2)} - \left(X_n^{(1)}\right) = \dfrac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}_n^{(1)}\right)^2$. Трябва да намерим начин, който да ни показва колко добра е всяка една от оценките.

Свойства на точковите оценки/статистики

$a)$ Неизместеност.

<u>Дефиниция</u>: Казваме, че $\hat{\theta}$ е неизместена оценка на θ, ако $\mathbb{E}\hat{\theta}(\overrightarrow{X}) = \theta$. Когато $\theta = (\theta_1, \ldots, \theta_s)$, равенството се разбира като $\mathbb{E}\hat{\theta}_j(\overrightarrow{X}) = \theta_j$, $\forall\ 1 \le j \le s$.

Иначе, ако това не е изпълнено, тогава $\theta - \mathbb{E}\hat{\theta}(\overrightarrow{X})$ Се нарича систематична грешка.

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$ $\hat{\mu} = X_m^{(1)}$, $\hat{\sigma}^2 = \overline{X}^{(2)} - \left(X_n^{(1)}\right)^2$, $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$

$\mathbb{E}\hat{\mu} = \mathbb{E}\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}X_j = \frac{n\mu}{n} = \mu \Rightarrow \hat{\mu}$ е неизвестно.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2$$

Нека μ е известно, тогава оценката е $\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_j - \mu)^2 = \frac{1}{n}n\sigma^2 = \sigma^2$.

Ако μ е известно, то $\mathbb{E}\hat{\sigma}^2 = \mathbb{E}\frac{1}{n}\sum_{j=1}^{n}X_j^2 = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 =$

$$= \mathbb{E}X_1^2 - \mathbb{E}\left(\frac{1}{n^2}\sum_{j=1}^{n}X_j^2 + \frac{1}{n^2}\sum_{i\neq j}X_iX_j\right) = \mathbb{E}X_1^2 - \frac{1}{n}\mathbb{E}X_1^2 + \frac{1}{n^2}\sum_{i\neq j}\underbrace{\mathbb{E}X_iX_j}_{\mu^2 \text{ произведение на две независими очаквания}} =$$

$$= \left(1 - \frac{1}{n}\right)\mathbb{E}X_1^2 + \frac{\mu^2}{n^{\cancel{2}}}\cancel{n}(n-1) = \left(1 - \frac{1}{n}\right)\left(\sigma^2 + \mu^2\right) - \frac{(n-1)\mu^2}{n} = \frac{n-1}{n}\sigma^2,$$

$\hat{\sigma}^2 \neq \sigma^2$.

$$\underbrace{S^2}_{\text{оценка}} = \frac{n}{n-1}\hat{\sigma}^2 \Rightarrow \mathbb{E}S^2 = \sigma^2, \text{ т.е. } \hat{\mu} = \overline{X}_n^{(1)},\ S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X_n})^2.$$

$\oplus$ $X \in \mathcal{U}nif(0, \theta)$, $\theta = Y = \max\limits_{1\le j\le n}(X_j)$ е ММП

$\mathbb{E}Y = ?$

$$F_Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}\left(\max_{j\le n}(X_j)\le y\right)=\mathbb{P}\left(\bigcap_{j=1}^{n}\{X_j\le y\}\right)=\prod_{j=1}^{n}\mathbb{P}(X_j\le y)=\frac{y^n}{\theta^n}$$

$$F_Y(y)=\begin{cases}\frac{y^n}{\theta^n}, & y\in[0,\theta]\\ \text{за останалите стойности}\\ \text{не се интересуваме}\end{cases}\Rightarrow\begin{cases}n\frac{y^{n-1}}{\theta^n}, & y\in(0,\theta)\\ 0, & \text{иначе}\end{cases}$$

$$\Rightarrow \mathbb{E}Y=\frac{n}{\theta^n}\int_0^\theta y^{n-1}y\,\mathrm{d}y=\frac{n}{n+1}\theta.$$

Оценката на максималното правдоподобие е изместена оценка на θ.

$\mathbb{E}\hat{\theta}=\dfrac{n}{n+1}\theta;\ \tilde{\theta}=\dfrac{n+1}{n}\hat{\theta}.\quad \hat{\theta}=2\overline{X}_n^{(1)}$ е неизместена за θ.

б) Състоятелност на статистика (точкова оценка).

<u>Дефиниция</u>: Казваме, че $\hat{\theta}$ е състоятелна оценка на $\theta=(\theta_1,\ldots,\theta_s)$, ако $\lim\limits_{n\to\infty}\hat{\theta}_j(\overrightarrow{X})\overset{\mathbb{P}}{=}\theta_j(\overline{\theta}_j\xrightarrow{\mathbb{P}}\theta_j)$, $1\le j\le s$.

$\oplus$ $\quad X\in\mathcal{N}(\mu,\sigma^2)$, $\hat{\mu}=\dfrac{1}{n}\sum\limits_{j=1}^{n}X_j\xrightarrow[n\to\infty]{\mathbb{P}}\mu$ е състоятелна.

μ - известно; $\quad\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{j=1}^{n}(X_j-\mu)^2=\dfrac{1}{n}\sum\limits_{j=1}^{n}Y_j\xrightarrow[n\to\infty]{\mathbb{P}}\mathbb{E}Y_1=\sigma^2$ е състоятелна.

$\hat{\mu}$ и $\hat{\sigma}^2$ Са състоятелни оценки, когато μ е известно ($\hat{\mu}$ по принцип винаги е състоятелна)

$\overline{X}_n^{(k)}=\dfrac{1}{n}\sum\limits_{j=1}^{n}X_j^k$, то с полагането $Y_j=X_j^k\Rightarrow\overline{X}_n^{(k)}=\dfrac{1}{n}\sum\limits_{j=1}^{n}\underbrace{Y_j}_{\substack{\text{еднакво разпр.}\\ \text{независими, т.к.}\\ X_j \text{ са такива и само}\\ \text{сме ги вдигнали}\\ \text{на степен } k}}$.

Тоест $\overline{X}_n^{(k)}$ е състоятелна оценка за k-тия момент. Така, че тези оценки, които се намират по метода на моментите по принцип са състоятелни. Това е така, защото имаме ЗГЧ и той ни казва, че средното аритметично на наблюденията на степен k се схожда по вероятност (по траекторно) до k-тия момент на случайната величина, която изучаваме.

$\overline{X}_n^{(k)} \underbrace{\approx}_{\text{е мн. близо}} \mu^{(k)}(\theta) = \mathbb{E}X^k$, решаваме $\underbrace{\overline{X}_n^{(k)} = \mu^k(\theta)}_{\text{това в граница е тавтология}}$ и решаването му ни гарантира състоятелна оценка по принцип.

СЕМ, лекция 14

(2021-01-14)

Доверителни интервали

Постановка:

- Случайна величина X;
- $F_X(x, \theta)$ е разпределение, което искаме да разберем, като знаем че то зависи параметрично от някакъв параметър θ;
- $\theta = (\theta_1, \ldots, \theta_s)$ - имаме s на брой параметри (например при нормалното разпределение са $s = 2$: средно μ и стандартно отклонение σ)

$\overrightarrow{X}(X_1, \ldots, X_n)$ - вектор от n независимо еднакво разпределени наблюдения над X (прототипи на X). На база на тези наблюдения, които в крайна сметка ще бъдат сведени до някакви числа (за модела/за експеримента) трябва да намерим някаква оценка $\underbrace{\hat{\theta} = \hat{\theta}(\overrightarrow{X})}_{\text{точкова оценка}}$, която да я вземем близо до θ, така че да определи това разпределение $F_X(x, \theta)$.

Проблема на точковата оценка е, че сама по себе си тя е доста динамична. Това е логично, тъй като извадките могат да бъдат различни. Хубаво ще е освен тази $\hat{\theta}$, да имаме и някаква вероятност, с която истинския параметър θ да попада в интервал, който може да нарачем *доверителен*.

Ще разглеждаме само едномерни параметри θ $(\theta \in \mathbb{R})$.

Цел: Ще търсим две числа $I_1 = I_1(\overrightarrow{X}) < I_2 = I_2(\overrightarrow{X})$ такаива, за които $\mathbb{P}(I_1 < \theta < I_2) = \gamma$ (като γ обикновено е число по-голямо от 0.9 и по-малко от 0.999 и има следния смисъл: колкото по-малко е γ, толкова по-широки интервали ще се получават, за да може с по-голяма вероятност да хванем истинския параметър θ. Стандартно $\gamma = 0.95$, а $\gamma = 0.90$ е за не чак толкова важни изследвания. За медицински цели се използва $\gamma \geq 0.999$.)

Дефиниция: (**Централна статистика ЦС**) Казваме, че $T = T(\overrightarrow{X}, \theta)$ е централна статистика, ако:

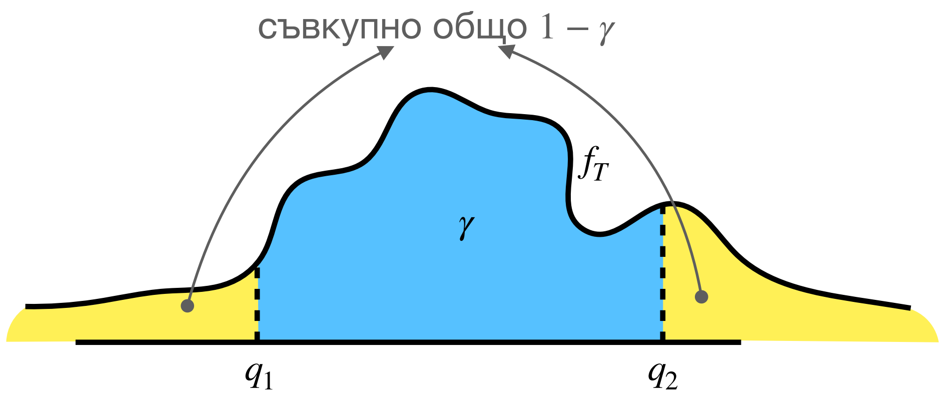
1) T е монотонна по θ
2) $\mathbb{P}(T < x) = F_T(x)$ не зависи от θ (T е функция на θ, но разпределението и не зависи от θ)

$\oplus$ Имаме $\overrightarrow{X}$ (вектор от наблюдения) и искаме да намерим някакъв доверителен интервал: (I_1, I_2) за θ, т.е. $\theta \in (I_1, I_2)$. Целта ни е да имаме някакво ниво на доверие $\gamma = \mathbb{P}(q_1 < T < q_2)$. За улеснение ще допуснем, че T расте по θ (но тя може и да намалява по θ).

Тъй като T е монотонна по θ, то

$$\gamma = \mathbb{P}(q_1 < T < q_2) \overset{\text{при фиксиран вектор на наблюдения } \overrightarrow{X}}{=} \mathbb{P}\left(T^{-1}(q_1) < \theta < T^{-1}(q_2)\right),$$ защото сме допускали, че T расте по θ. Ако T намаляваше по θ, щяхме да имаме $\gamma = \mathbb{P}\left(T^{-1}(q_2) < \theta < T^{-1}(q_1)\right)$. Следователно интервала ще е $I_1 = T^{-1}(q_1)$, $I_2 = T^{-1}(q_2)$.

f_T - плътността на централната статистика (ЦС)

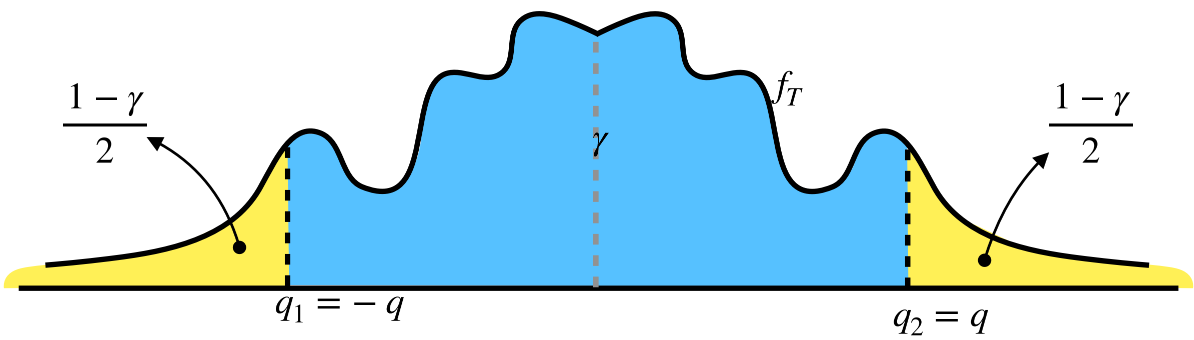


Има много начини, по които може да изберем q_1 и q_2, така че вероятността между тях да е γ.

Имаме параметър $\theta \in \mathbb{R}$, който искаме да оценяваме. Търсим I_1 и I_2, които да зависят от наблюденията $\overrightarrow{X}$ и $I_1(\overrightarrow{X}) < I_1(\overrightarrow{X})$. Те образуват т.нар. *доверителен интервал* за θ с ниво на доверие $\gamma = \mathbb{P}\left(\theta \in (I_1, I_2)\right)$.

T е централна статистика за θ, ако удовлетворява дефиницията за ЦС.

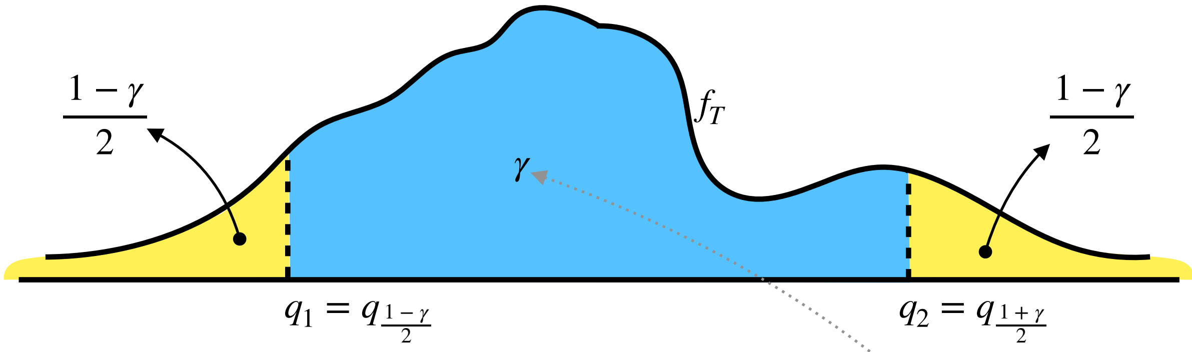
⊕ Ако T е симетрична случайна величина,



тогава се търсят такова q_1 и q_2, че $q_1 = -q = -q_2$, за които

$\mathbb{P}(T < -q) = 1 - \mathbb{P}(T < 1) = \dfrac{1-\gamma}{2}$ (за да имаме по средата вероятност γ)

<u>По-общо</u>: Ако имаме някакво несиметрично разпределение на T:



проверка : $\mathbb{P}(T < q_2) - \mathbb{P}(T < q_1) = \dfrac{1+\gamma}{2} - \dfrac{1-\gamma}{2} = \dfrac{1+\gamma-1+\gamma}{2} = \gamma$

Имаме много начини, по които може да изберем q_1 и q_2, но тези които демонстрираме по-горе са изпитани от практиката рецепти за избиране и имат конкретен смисъл за симетричните разпределения.

$$\underbrace{\gamma}_{\text{фиксирана}} = \mathbb{P}(q_1 < T < q_2) = \mathbb{P}\Big(\underbrace{T^{-1}(q_1) < \theta < T^{-1}(q_2)}_{T \text{ нараства}}\Big).$$

Тази вероятност γ е фиксирана и се задава предварително от изследователите. Какво може да оптимизираме ние като математици? - може да търсим q_1 и q_2 такива, за които е изпълнено:

$$\min_{\substack{q_1 < q_2 \\ \gamma = \mathbb{P}(q_1 < T < q_2)}} = \Big\{ \underbrace{|T^{-1}(q_2) - T^{-1}(q_1)|}_{\substack{\text{искаме най-малък} \\ \text{интервал, за да може} \\ \text{да свием опциите за} \\ \theta \text{ -максимално}}} \Big\}.$$

Т.е. минимизираме доверителния интервал при фиксирано ниво на доверие!

$\oplus$ $\underbrace{X \in \mathcal{N}(\mu, \sigma^2)}_{\substack{\text{случайна величина, която} \\ \text{искаме да изучаваме}}}$, σ^2 е известно, т.е. интересуваме се само от параметъра $\mu = \theta$ (едномерен). Искаме да видим как може да оценим μ и да намерим за него доверителен интервал.

Ние знаем, че $\hat{\mu} = \overline{X_n} = \dfrac{1}{n}\sum_{j=1}^{n} X_j$. Тогава $T(\overrightarrow{X}, \mu) = \dfrac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \in \mathcal{N}(0,1)$, σ е известно число. Това е така, защото

$$\sum_{j=1}^{n} X_j \overset{\substack{\text{линейност} \\ \text{на } \mathcal{N} \\ \text{и незав.}}}{\in} \mathcal{N}(n\mu, n\sigma^2) \Rightarrow \frac{1}{n}\sum_{j=1}^{n} X_j \in \mathcal{N}\left(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

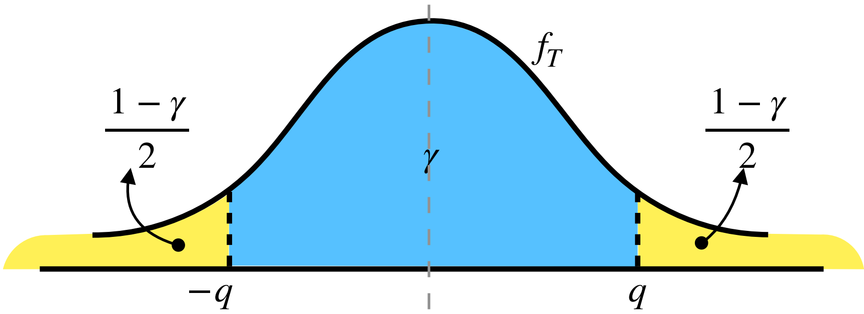
[По-подробно обяснение на свойството линейност на нормалното разпределение:

$$\overline{X_n} = \frac{\sum_{j=1}^{n} X_j}{n} \in \mathcal{N}(x, y^2)\,.\ \mathbb{E}\overline{X_n} = \frac{n\mu}{n} = x;\ \operatorname{Var}(\overline{X_n}) = \frac{1}{n^2}\sum_{j=1}^{n}\sigma^2 = \frac{\sigma^2}{n} = y^2\Big]$$

T е намаляваща функция по μ и $\mathbb{P}(T \le x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,\mathrm{d}\,y$ - не зависи от $\mu \overset{def.}{\Rightarrow}$

T е централна статистика за μ (тя е монотонна и намаляваща по μ и нейното разпределение съвпада с $\mathcal{N}(0,1)$, т.е. не зависи от θ)

Тогава, $\gamma = \mathbb{P}(-q < T < q)$, тъй като $\mathcal{N}(0,1)$ е симетрично:



Това ни гарантира, че $(-q, q)$ ще е най-тесния интервал, тъй като в $(-\infty, -q)$ и (q, ∞) е сбита най-малко маса (навсякъде извън тези интервали е сбита повече маса)

$q = q_{\frac{1}{2}+\frac{\gamma}{2}}$, който квантил го има в таблицата за нормалното стандартно разпределение.

$$\gamma = \mathbb{P}(-q < T < q) = \mathbb{P}\left(-q_{\frac{1}{2}+\frac{\gamma}{2}} < \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} < q_{\frac{1}{2}+\frac{\gamma}{2}}\right) \Rightarrow$$

$$\Rightarrow \mathbb{P}\left(\mu \in (\underbrace{\overline{X_n} - \frac{\sigma}{\sqrt{n}} \times q_{\frac{1}{2}+\frac{\gamma}{2}}}_{I_1}, \underbrace{\overline{X_n} + \frac{\sigma}{\sqrt{n}} \times q_{\frac{1}{2}+\frac{\gamma}{2}}}_{I_2})\right) = \gamma \Rightarrow$$

$$\Rightarrow I_1 = \overline{X_n} - \frac{\sigma}{\sqrt{n}} \times q_{\frac{1}{2}+\frac{\gamma}{2}};\ I_2 = \overline{X_n} + \frac{\sigma}{\sqrt{n}} \times q_{\frac{1}{2}+\frac{\gamma}{2}}.$$

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, но този път не знаем σ.

Как да конструираме доверителен интервал само за μ? Припомняме, че

$$\underbrace{\hat{\mu} = \overline{X_n}}_{\text{независимо дали знаем или не } \sigma} \quad \text{и оценката за дисперсията е } s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X_n})^2 \text{ и е неизместена.}$$

Фактора $\dfrac{1}{n-1}$ го има, тъй като тя е неизместена оценка за дисперсията.

<u>Твърдение</u>: Имаме, че $X \in \mathcal{N}(\mu, \sigma^2)$ и $\overrightarrow{X} = (X_1, \ldots, X_n)$ са наблюдения над X. Тогава е вярно, че:

а) $\hat{\mu}$ е независимо от s^2 : $\hat{\mu} \perp\!\!\!\perp s^2$;

б) $(n-1)\dfrac{s^2}{\sigma^2} \in \mathcal{X}^2(n-1)$.

$T = \dfrac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}$. Ако знаехме σ, последното щеше да е разпределено като $\mathcal{N}(0,1)$

неизвестно

(както направихме в предходния пример), но ние не знаем σ. Но друго, което знаем е, че:

$\in \mathcal{N}(0,1)$

$$T = \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{n-1}{n-1} \cdot \frac{s^2}{\sigma^2}}} = \frac{\overline{X_n} - \mu}{\frac{s}{\sqrt{n}}} \text{ е централна статистика!}$$

$\in \mathcal{X}^2(n+1)$

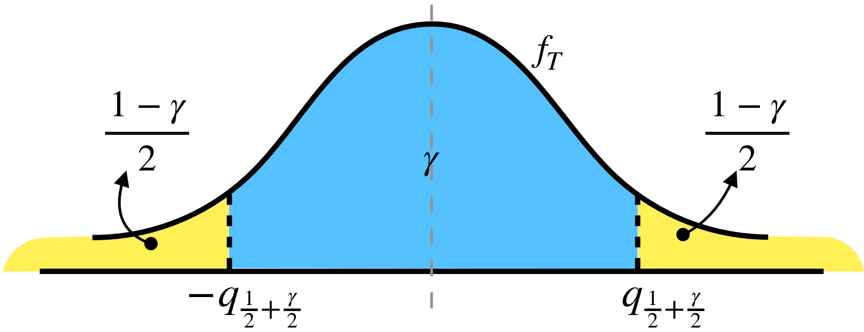
Тъй като $\dfrac{\overline{X_n} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{Z}{\sqrt{\frac{Y}{n-1}}}$, където $Z \in \mathcal{N}(0,1)$, $Y \in \mathcal{X}^2(n+1)$ и $Z \perp\!\!\!\perp Y$.

<u>Заключения</u>:

- T е намаляваща по μ
- $T \in t(n-1)$ и не зависи от $\mu \Rightarrow T$ е централна статистика за μ.

$$T = \frac{Z}{\sqrt{\frac{Y}{n-1}}};\ Z = \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}};\ Y = \frac{(n-1)s^2}{\sigma^2}.$$

T е симетрична, тъй като в числителя имаме симетрична случайна величина (Z) $\Rightarrow$



$$\Rightarrow \mathbb{P}(-q_{\frac{1}{2}+\frac{\gamma}{2}} < T < q_{\frac{1}{2}+\frac{\gamma}{2}}) = \mathbb{P}\left(\mu \in \left(\overline{X}_n - q_{\frac{1}{2}+\frac{\gamma}{2}} \times \frac{s}{\sqrt{n}}, \overline{X}_n + q_{\frac{1}{2}+\frac{\gamma}{2}} \times \frac{s}{\sqrt{n}}\right)\right),$$

т.е. имаме, че I_1 и I_2 зависят от s и квантилите не са от нормалното разпределение а от $t - \text{Student's}$ разпределението.

<u>Бележки</u>: $T - \dfrac{\overline{X}_n - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{Z}{\sqrt{\frac{Y}{n-1}}}$, $Z \perp\!\!\!\perp Y$, $Z \in \mathcal{N}(0,1)$, $Y \in \mathcal{X}^2(n-1)$

- $Y = \sum_{j=1}^{n-1} V_j$, където $(V_j)_{j=1}^{n-1}$ са независими и еднакво разпределени (i.i.d.) с $\mathcal{X}^2(1)$.

  От ЗГЧ: $\dfrac{Y}{n-1} \xrightarrow[n\to\infty]{\text{п.с.}} \mathbb{E}V_1 = 1$. Следователно за големи n : $T \approx \dfrac{Z}{1} \approx \mathcal{N}(0,1)$;

- Ако знаем дисперсията: $\dfrac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$ за произволни $X_1, \ldots, X_n, \ldots$

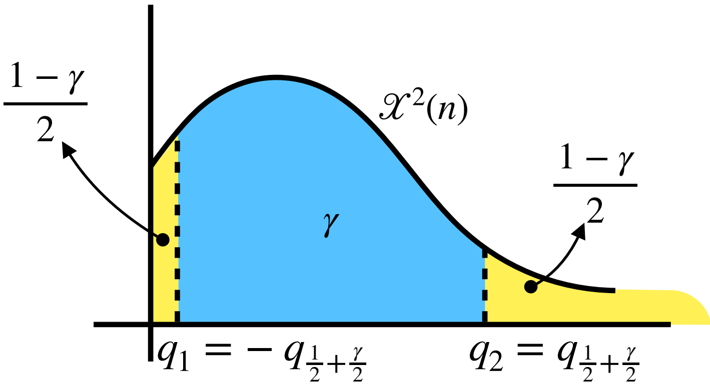
  $\mathbb{E}X_1 = \mu;\ \operatorname{Var}(X_1) = \sigma^2$.

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, знаем μ, $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2$. Искаме да оценим дисперсията.

Трябва да си конструираме централна статистика. ЦС не трябва да зависи от σ, но в $\hat{\sigma}^2$, X_j зависи от σ и трябва да отстраним тази зависимост.

Нагаждаме: $\dfrac{n\hat{\sigma}^2}{\sigma^2} = \sum_{j=1}^{n}\underbrace{\left(\dfrac{X_j - \mu}{\sigma}\right)^2}_{Z_j^2} = \sum_{j=1}^{n} Z_j^2$, но $Z_j \in \mathcal{N}(0,1)$, тъй като центрираме и нормираме с неизвестна дисперсия. Следователно $\sum_{j=1}^{n} Z_j^2 \in \mathcal{X}^2(n)$, защото сумира n квадрати на независими нормални стандартно разпределени $Z \Rightarrow$ статистиката

$T = \dfrac{n\hat{\sigma}^2}{\sigma^2}$ е централна по дефиниция (разпределението и е $\mathcal{X}^2(n)$ - не зависи от σ^2 и тя е монотонно намаляваща по σ).
За квантилите:

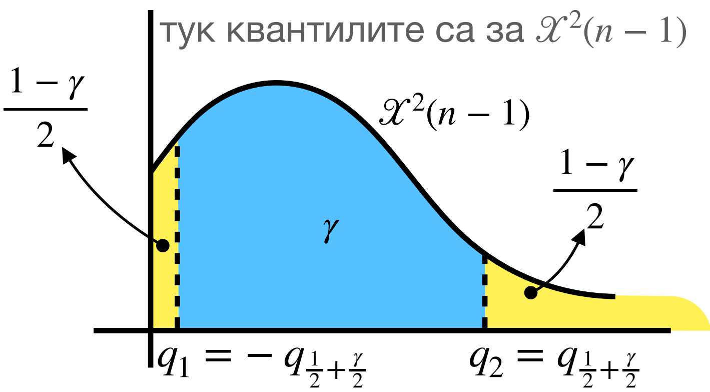


$$\gamma = \mathbb{P}(q_1 < T < q_2) = \mathbb{P}\left(\frac{n\hat{\sigma}^2}{q_2} < \frac{n\hat{\sigma}^2}{q_1}\right) \Rightarrow \begin{cases} I_1 = \frac{n\hat{\sigma}^2}{q_2} \\ I_2 = \frac{n\hat{\sigma}^2}{q_1} \end{cases}$$

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, не знаем μ и се интересуваме от σ^2. Знаем, че s^2 е неизместена оценка за σ^2:

$(n-1)s^2 = \sum_{j=1}^{n} (X_j - \overline{X_n})^2$ и нашата цел е да конструираме статистика за σ^2, т.е. да намерим ЦС, отговаряща на дефиницията. Ако разделим на σ^2, от дясно не можем да направим никакво заключение: $\dfrac{(n-1)s^2}{\sigma^2} = \dfrac{\sum_{j=1}^{n}(X_j - \overline{X_n})^2}{\sigma^2}$, но знаем от твърдението, че $T = \dfrac{(n-1)s^2}{\sigma^2} \in \mathcal{X}^2(n-1)$ (с една степен по-малко, защото сме изхабили една степен за оценката на μ). От n наблюдения все едно имаме $n-1$ наблюдения.

$\Rightarrow$ T е централна статистика (монотонна намаляваща по σ^2 и независима от σ^2)



$$\gamma = \mathbb{P}\left(q_1 < \frac{(n-1)s^2}{\sigma^2} < q_2\right) = \mathbb{P}\left(\frac{(n-1)s^2}{q_2} < \sigma^2 < \frac{(n-1)s^2}{q_1}\right)$$

$$I_1 = \frac{1}{q_2}(n-1)s^2;\ I_2 = \frac{1}{q_1}(n-1)s^2$$

При голямо n $(n > 30$ например) може да ползваме както от предходния пример - все едно знаем σ.

Проверка на хипотези

Теорията за тестване на хипотези е въведена и развита за първи път от Нейман и Пиърсън през 40-те години на миналия век. Тя най-често се задава със следната математическа постановка:

X е случайна величина с функция на разпределение $F_X(x, \theta)$, нулева хипотеза H_0 и алтернативна такава H_1, където θ е оценявания параметър.

$H_0 : \theta = \theta_0$
$H_1 : \theta = \theta_1$

Искаме да конструираме някакво множество $W \in \mathbb{R}^n$ такова, че ако $\overrightarrow{X}$ попадне в W $(\overrightarrow{X} \in W)$, тогава отхвърляме H_0 (в полза на H_1), ако вектора от наблюдения $\overrightarrow{X}$ не попадне в W, то тогава не отхвърляме H_1.

H_0 и H_1 се наричат прости хипотези (нулева и алтернативна). Прости хипотези са $\theta = \theta_1$ (число). Сложни хипотези са $\theta > \theta_i$, $\theta \neq \theta_i$, $\theta \in I$ и т.н.

Цел: При наблюденията $\overrightarrow{X} \in \mathbb{R}^n$, търсим да конструираме $W \subseteq \mathbb{R}^n$: ако $\overrightarrow{X} \in W$, то отхвърляме H_0 и приемаме H_1, а ако $\overrightarrow{X} \in \overline{W}$, то приемаме H_0.

$$W \subseteq \mathbb{R}^n : \begin{cases} \overrightarrow{X} \in W \Rightarrow \text{отхвърляме } H_0 \\ \overrightarrow{X} \in \overline{W} \Rightarrow \text{приемаме } H_0 \end{cases}$$

Грешки, които може да допуснем:

- Грешка от $I^{\text{-ви}}$ род: Да отхвърлим H_0, когато H_0 е вярна, т.е. $\alpha = \mathbb{P}(\overrightarrow{X} \in W \,|\, H_0)$
- Грешка от $II^{\text{-ри}}$ род: При положение, че е вярна хипотезата H_1, ние сме приели H_0, т.е. $\beta = \mathbb{P}(\overrightarrow{X} \in \overline{W} \,|\, H_1)$

$\pi = 1 - \beta$ се нарича *мощност* на W.

$\oplus$ H_0 : дадена ваксина е вредна ($\theta = \theta_0$)
H_1 : ваксината не е вредна ($\theta = \theta_1$)

Има по-голям резон за H_0 да вземем по-опасната/рисковата хипотеза. Това е тази хипотеза, която е по-вероятно да я отхвърлим. Това е така, защото грешката от I-ви род ще бъде контролирана/задавана от изследователя (той ще казва дали иска/допуска да е 0.01, 0.05, 0.1 и т.н.)

Дефиниция: (**Оптимална критична област**) При фиксирана грешка от I-ви род α, $W^* \subseteq \mathbb{R}^n$ се нарича оптимална критична област (ОКО), ако

$$\mathbb{P}(\overrightarrow{X} \notin W^* \mid H_1) = \min_{\substack{\underbrace{W \subseteq \mathbb{R}^2}_{\text{всички крит. области}} \\ \alpha = \mathbb{P}(X \in W \mid H_0)}} \mathbb{P}(\overrightarrow{X} \notin W \mid H_1).$$

Постановка: X е случайна величина; $F_X(x, \theta)$ е разпределението на X, което зависи от накакъв параметър θ, но допускаме, че $f_X(x, \theta)$ е плътността на X (т.е. допускаме, че $\frac{\partial x}{\partial} F_X(x, \theta)$ съществува). Въвеждаме

$$f_{\overrightarrow{X}}(x, \theta) = \underbrace{L(X, \theta)}_{\text{ф-я на правдоподобие}} = \prod_{j=1}^{n} f_X(x_j, \theta), \text{ където } x \in \mathbb{R}^n \text{ и } x = (x_1, \ldots, x_n).$$

Тогава е верен следния резултат:

Лема: (**Нейман-Пиърсън**) Нека X удовлетворява горните условия от постановката и тестваме следната хипотеза:

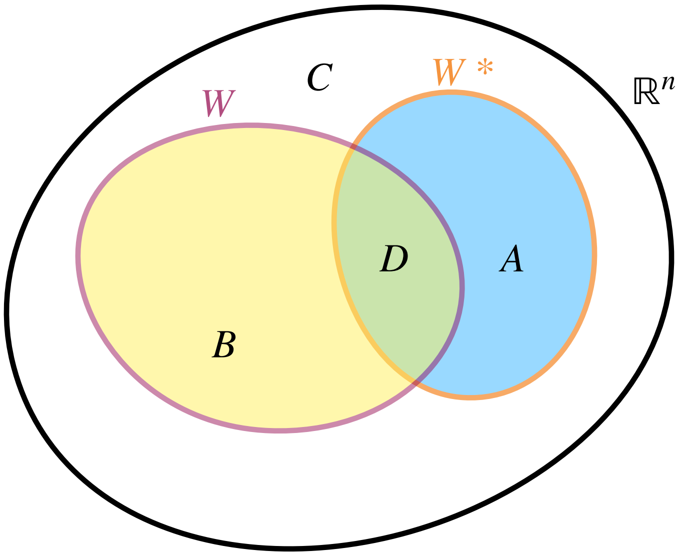
$H_0 : \theta = \theta_0$ срещу $H_1 : \theta = \theta_1$.

Ако $L_0(x) = L(x, \theta_0)$ и $L_1(x) = L(x, \theta_1)$ и

$\exists k \geq 0 : W^* \subseteq \{x \in \mathbb{R}^n : L_1(x) \geq kL_0(x)\}, \overline{W^*} \subseteq \{X \in \mathbb{R}^2 : L_1(x) \leq kL_0(x)\}$ и $\alpha = \mathbb{P}(X \in W^* \mid H_0)$ е зададена, то W^* е ОКО (оптимална критична област).

Доказателство:

$$\underbrace{\alpha = \mathbb{P}(\overrightarrow{X} \in W^* \mid H_0) = \mathbb{P}(\overrightarrow{X} \in W \mid H_0)}_{\text{имаме}} \Rightarrow \underbrace{\mathbb{P}(\overrightarrow{X} \notin W^* \mid H_1) \leq \mathbb{P}(\overrightarrow{X} \notin W \mid H_1)}_{\text{искаме да докажем}}$$



$$\mathbb{P}(\overrightarrow{X} \notin W \,|\, H_1) \overset{\theta=\theta_1}{=} \int_{\overline{W}} L_1(x)\mathrm{d}\,x = \int_A L_1(x)\mathrm{d}\,x + \int_C L_1(x)\mathrm{d}\,x + \underbrace{\int_B L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x}_{} =$$

$$= \int_{\overline{W*}} L_1(x)\mathrm{d}\,x + \int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x = \mathbb{P}(\overrightarrow{X} \in W* \,|\, H_1) + \underbrace{\int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{dx}}_{\overset{?}{\geq} 0} \geq$$

$\geq \mathbb{P}(\overrightarrow{X} \notin W* \,|\, H_1)$, което искахме да докажем.

Т.е. всичко се свежда до това да проверим, че $\int_A L_1(x)\mathrm{d}\,x - \int_B L_1(x)\mathrm{d}\,x \geq 0$, но

$$\int_A L_1(d)\mathrm{d}\,x - \int_B L_1()x\,\mathrm{d}\,x \geq \underbrace{k\int_A L_0(x)\mathrm{d}\,x - k\int_B L_0(x)\mathrm{d}\,x}_{\overset{?}{=}0}$$

$$\alpha = \int_{W*} L_0(x)\mathrm{d}\,x = \int_A L_0(x)\mathrm{d}\,x + \cancel{\int_D l_0(x)\mathrm{d}\,x} = \int_W L_0(x)\mathrm{d}\,x = \int_B L_0(x)\mathrm{d}\,x + \cancel{\int_D L_0(x)\mathrm{d}\,x}$$

$$\Rightarrow \int_A L_0(x)\mathrm{d}\,x = \int_B L_0(x)\mathrm{d}\,x.$$

СЕМ, лекция 15

(2021-01-21)

Припомняне:

- Имаме случайна величина X, която искаме „да разберем“, т.е. да извлечем някаква информация за нея, и чиято функция на разпределение зависи от някакъв параметър θ (едномерен);
- $\overrightarrow{X}$ са някакви наблюдения дадени като n-мерен вектор;
- α (алфа) е предварително зададена грешка от I-ви род;
- Тестваме две прости хипотези:
  $H_0 : \theta = \theta_0$ (базова)
  $H_1 : \theta = \theta_1$ (алтернативна)

Означаваме: $L_0(x) = L(x; \theta_0)$ (функция на правдоподобие за $\theta = \theta_0$) и $L_1(x) = L(x; \theta_1)$ (функция на правдоподобие за $\theta = \theta_1$).

Търсим $W^* \subseteq \mathbb{R}^n$ (област на $\mathbb{R}^n$), така, че когато нашия n-мерен вектор от наблюдения попадне в него, ние да отхвърляме нулевата хипотеза и да приемаме алтернативната. $\alpha = \mathbb{P}(\overrightarrow{X} \in W^* \,|\, H_0)$ е грешка от първи род, т.е. да попаднем в W^* и да отхвърлим нулевата хипотеза, но тя да е била вярна. Тази грешка е презададена (предефинирана) и се контролира от изследователя. Целта на оптималната критична област W^* е да намери тази област, която минимизира грешката от първи род.

$$\beta = \min_{\alpha = \mathbb{P}(\overrightarrow{X} \in W | H_0)} \mathbb{P}(\overrightarrow{X} \notin W \,|\, H_1).$$

Лемата на Нейман-Пиърсън е „добра“, защото ни характеризира даден критетии, по който да определим дали една област е оптимална критична област и по тази лема знаем, че W^* е о.к.о. (оптимална критична област), ако съществува някаква константа K (K може да зависи от n и от θ, но не може да зависи от x), за която

$$W^* \in \{x \in \mathbb{R}^n : L_1(x) > K \times L_0(x)\}$$
$$W^{*c} \in \{x \in \mathbb{R}^n : L_1(x) \leq K \times L_0(x)\}$$

и ако знаем, че е изпълнено равенството: $\alpha = \mathbb{P}(\overrightarrow{X} \in W^* \,|\, H_0)$, то W^* е о.к.о.

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, където σ^2 е известно и искаме да построим оптимална критична област за тестване на хипотезата на μ (за намиране на средното, знаейки каква е дисперсията)

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu = \mu_1$$

Допускаме за улеснение, че $\mu_1 > \mu_0$. При зададено α.

$$L_0(x) = (\sqrt{2\pi}\sigma)^{-n} e^{-\frac{\sum_{j=1}^{n}(x_j-\mu_0)^2}{2\sigma^2}},\ L_1(x) = (\sqrt{2\pi}\sigma)^{-n} e^{-\frac{\sum_{j=1}^{n}(x_j-\mu_1)^2}{2\sigma^2}}.$$

От лемата на Нейман-Пиърсън знаем, че оптималните критични области се намират лесно с неравенства от вида:

$$\{X \in \mathbb{R}^n : L_1(x) \geq K \times L_0(x)\} = \left\{ X \in \mathbb{R}^n : \frac{-\sum_{j=1}^{n}(x_j-\mu_1)^2}{2\sigma^2} \geq \ln K - \frac{\sum_{j=1}^{n}(x_j-\mu_0)^2}{2\sigma^2} \right\} =$$

$$= \left\{ X \in \mathbb{R}^n : -\cancel{\frac{\sum_{j=1}^{n} x_j^2}{2\sigma^2}} + \frac{1}{\sigma^2}\mu_1 \sum_{j=1}^{n} x_j - \underbrace{\frac{n\mu_1^2}{2\sigma^2}}_{\text{не зависи от наблюденията } \vec{X}} \geq \right.$$

$$\left. \geq -\cancel{\frac{\sum_{j=1}^{n} x_j^2}{2\sigma^2}} + \frac{1}{\sigma^2}\mu_0 \sum_{j=1}^{n} x_j - \underbrace{\frac{n\mu_0^2}{2\sigma^2}}_{\text{не зависи от наблюденията } \vec{X}} \right\} =$$

$$= \left\{ X \in \mathbb{R}^n : \frac{1}{\sigma^2} \underbrace{(\mu_1 - \mu_0)}_{\substack{\mu_1 > \mu_0 \\ \text{по допускане}}} \sum_{j=1}^{n} x_j \geq K_1 \right\} = \text{, където } K_1 = \ln K - \frac{n\mu_0^2}{2\sigma^2} + \frac{n\mu_1^2}{2\sigma^2}.$$

$$= \left\{ X \in \mathbb{R}^n : \sum_{j=1}^{n} x_j \geq K_2 \right\} = \text{, където } K_2 = \frac{K_1\sigma^2}{\mu_1 - \mu_0}$$

$= \overline{X}$

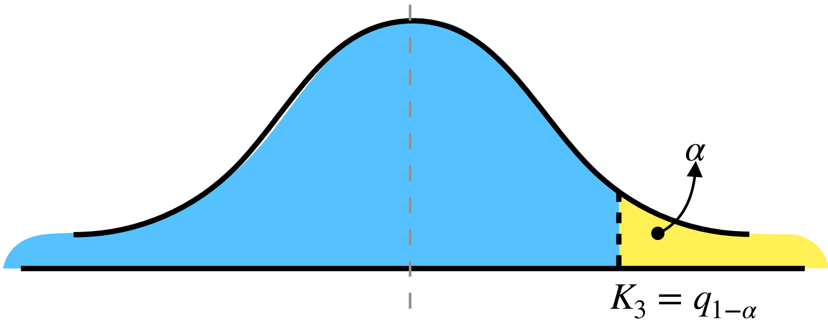
$$= \left\{ X \in \mathbb{R}^n : \frac{\sum_{j=1}^{m} x_j}{n} \geq \frac{K_2}{n} \right\} = \left\{ X \in \mathbb{R}^n : \overline{X} \geq \frac{K_2}{n} \right\} =$$

$$= \left\{ X \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{K_2}{\sigma\sqrt{n}} = K_3 \right\}.$$

$$\alpha = \mathbb{P}(\vec{X} \in W \,|\, H_0) = \mathbb{P}\big(\underbrace{L_1(\vec{X}) \geq K_0 L_0(\vec{X})}_{} \,|\, H_0\big) = \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq K_3 \,|\, H_0 \right),$$

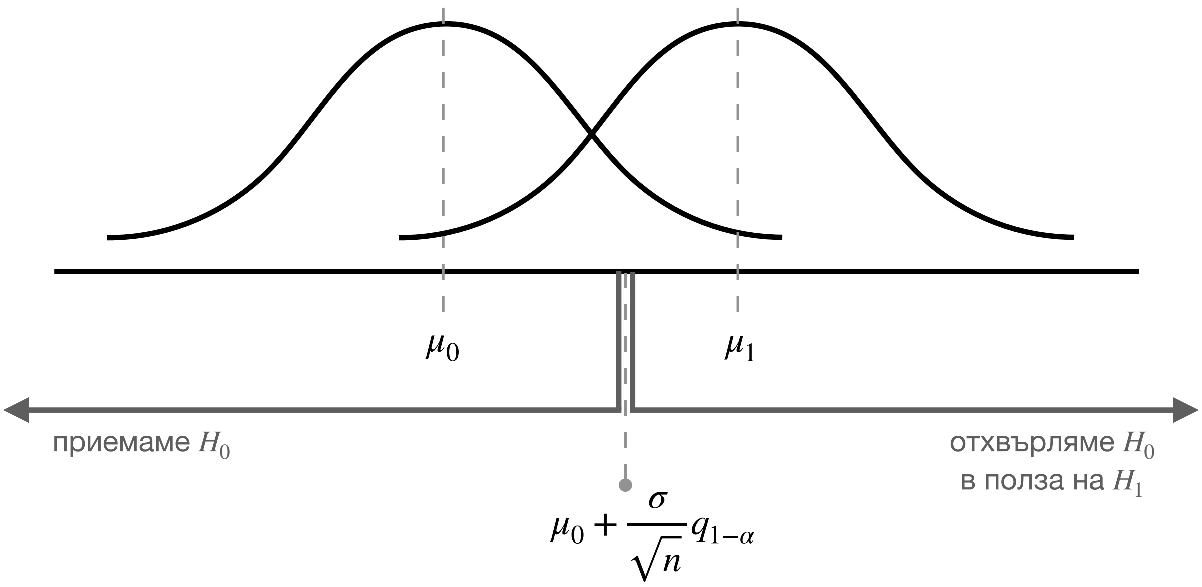
където $\overline{X} = \dfrac{\sum_{j=1}^{n} x_j}{n}$.

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \in \mathcal{N}(0,1) = \mathbb{P}(Z \geq K_3)$$



$$\Rightarrow K_3 = q_{1-\alpha}$$

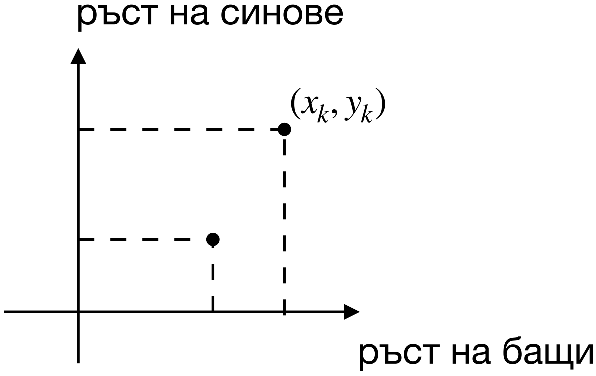
о.к.о.: $\left\{ \overline{X} \geq \mu_0 + q_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right\}$



Линейна регресия

(Галтон)

Нека a е средния ръст на мъжете.

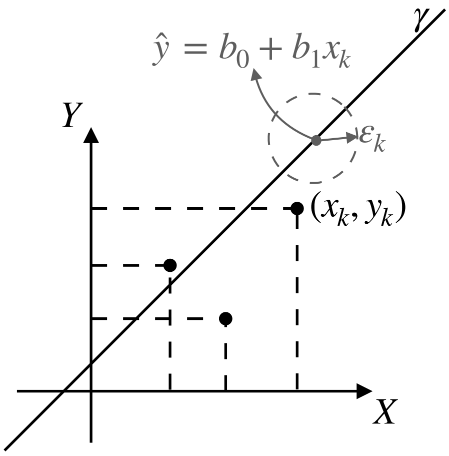


$y_{\text{син}} = a + \beta(x_{\text{баща}} - a)$.

Галтон е забелязал, че по неговите данни, коефициента бета е $\beta = 0.6$. Тоест, ако бащата е 10 см. над средния ръст, то сина му ще е с 6 см. над средния ръст. Тази по-слаба зависимост е влязла в теорията като регрес (завръщане) към средното. Синовете на високите бащи не са чак толкова високи в средно както бащите им, а са на около половината от отклонението на бащата над средния ръст за мъжете.

Модел: $Y = \beta_0 + \beta_1 X + \varepsilon$ грешка

отклик предиктор



Допускаме че в множеството от точки $(x_1, x_2, \ldots, x_n)$ и $(y_1, y_2, \ldots, y_n)$ има някакъв линеен модел. Т.е. предполагаме, че има линеен модел $y_k = b_0 + b_1 x_k + \varepsilon_k$. Тоест имаме някаква права γ (от чертежа). Искаме да си построим линеен модел, а не някакъв друг, за да не рискуваме да интерполираме, тъй като интерполацията няма добра статистическа стойност. Т.е. не е добре да обхванем всички данни с много сложна крива и в момента, в който добавим данни – кривата ни да е твърде динамична и да няма никаква прогнозна сила. Интерполацията може да мине през n точки, но при добавянето на $n+1$-вата точка – точността на кривата да рухне.

$\overline{X} = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} x_k$ и $\overline{Y} = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} y_k$.

Търсим: $\displaystyle\min_{b_0,\, b_1} \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 = \min_{b_0,\, b_1} \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k)^2$.

(С квадратични грешки се смята по-лесно, а освен това имат и статистическо значение. Въпреки това тук може да имаме най-разнообразни метрики, които искаме да оптимизираме (например абсолютната стойност или максималното отклонение измежду всички възможни отклонения и т.н.))

Искаме да минимизираме функцията по-горе по две променливи. За целта ще си вземем производната по b_0 и тя трябва да бъде нула и аналогично за производната по b_1:

$$0 = \frac{\partial}{\partial b_0} \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k)^2 = -2 \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k) \Rightarrow$$

$$\Rightarrow \sum_{k=1}^{n} y_k - n b_0 - b_1 \sum_{k=1}^{n} x_k = 0 \Rightarrow n\overline{Y} - n b_0 - n b_1 \overline{X} = 0 \Rightarrow \overline{Y} = b_0 + b_1 \overline{X}$$

$$0 = \frac{\partial}{\partial b_1} \sum_{k=1}^{n} (y_k - b_0 - b_1 x_k)^2 = -2 \sum_{k=1}^{n} (b_0 + b_1 x_k - y_k) x_k \Rightarrow$$

$$\Rightarrow n\overline{X}(\overline{Y} - b_1 \overline{X}) + b_1 \sum_{k=1}^{n} x_k^2 - \sum_{k=1}^{n} x_k y_k$$

$$\begin{cases} b_0 = \overline{Y} - b_1 \overline{X} \\ b_1 = \dfrac{\sum_{k=1}^{n} x_k y_k - n\overline{Y}\overline{X}}{\sum_{k=1}^{n} x_k^2 - n(\overline{X})^2} = \dfrac{\sum_{k=1}^{n} (y_k - \overline{Y})(x_k - \overline{X})}{\sum_{k=1}^{n} (x_k - \overline{X})^2} = \dfrac{\sum_{k=1}^{n} (x_k - \overline{X}) y_k}{\sum_{k=1}^{n} (x_k - \overline{X})^2} \end{cases}$$

$$A = \sum_{k=1}^{n} (X_k - \overline{X})^2$$

Допускаме, че Y_k като отговор на X_k е случайна величина, в смисъл, че съществуват неизвестни коефициенти β_0, β_1, които при зададено X_k дават следната линейна зависимост, където ε_k е случайна грешка.
Допускаме, че $Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k$, където $k = \overline{1,n}$.

Правим и следните допускания за епсилон грешките:
$(\varepsilon_i)_{i=1}^{n}$ са независими еднакво разпределени случайни величини, като $\varepsilon_i \in \mathcal{N}(0, \sigma^2)$

хомоскедастичност

Т.е. грешките са нормално разпределени и независими една от друга. Т.е. нямаме системна грешка. Хомоскедастичността е малко по-тежко допускане, но тя придава простота на модела.

$$Y_k \in \mathcal{N}(\beta_0 + \beta_1 X_k, \sigma^2)$$

$$\begin{cases} \hat{\beta}_0 = b_0 = \overline{Y} - b_1 \overline{X} \\ \hat{\beta}_1 = b_1 = \underbrace{\dfrac{\sum_{k=1}^{n} X_k Y_k - n\overline{X}\overline{Y}}{\sum_{k=1}^{n} X_k^2 - n(\overline{X})^2}}_{(1)} = \underbrace{\dfrac{\sum_{k=1}^{n} (Y_k - Y)(X_k - X)}{\sum_{k=1}^{n} (X_k - \overline{X})^2}}_{(2)} = \underbrace{\dfrac{\sum_{k=1}^{n} (X_k - \overline{X}) Y_k}{\sum_{k=1}^{n} (X_k - \overline{X})^2}}_{(3)} \end{cases}$$

$A = \sum_{k=1}^{n} (X_k - \overline{X})^2$ си остава същото, тъй като е фиксирано число в знаменателя.

$$\mathbb{E}b_1 \overset{(3)}{=} \frac{1}{A}\sum_{k=1}^{n}(X_k-\overline{X})\mathbb{E}Y_k = \frac{1}{A}\sum_{k=1}^{n}(X_k-\overline{X})(\beta_0-\beta_1 X_k) =$$

$$= \frac{\beta_0}{A}\underbrace{\cancel{\sum_{k=1}^{n}(X_k-\overline{X})}}_{=0} + \frac{\beta_1}{A}\underbrace{\sum_{k=1}^{n}(X_k-\overline{X})X_k}_{=A} = \beta_1\,.$$

Оказва се, че очакването на b_1 е равно на β_1, което ни казва, че b_1 е неизместена оценка на β_1.

$$\mathbb{E}b_0 = \mathbb{E}\overline{Y} - \overline{X}\mathbb{E}b_1 = \frac{1}{n}\sum_{k=1}^{n}Y_k - \beta_1\overline{X} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}(\beta_0 + \cancel{\beta_1 X_k}) - \cancel{\beta_1\overline{X}} = \beta_0\,.$$

И b_0 и b_1 са неизместени оценки на неизвестните парамвтри β_0 и β_1. По този начин знаем, че нямаме систематична грешка, когато правим тези оценки.

$$Db_1 \overset{(3)}{=} D\frac{\sum_{k=1}^{n}(X_1-\overline{X})Y_k}{A} = \frac{1}{A^2}\sum_{k=1}^{n}(X_n-\overline{X})^2\underbrace{DY_k}_{\sigma^2} =$$

$$= \sigma^2\frac{\cancel{\sum_{k=1}^{n}(X_n-\overline{X})^2}}{A^{\cancel{2}}} = \frac{\sigma^2}{A}\,.$$

$$\Rightarrow b_1 = \underset{\text{оценка}}{\hat{\beta}_1} \in \mathcal{N}\left(\beta_1, \frac{\sigma^2}{A}\right)$$

Това означава, че вече може да тестваме хипотези за b_1.

За дисперсията на b_0 по същата логика може да докажем, че:

$$Db_0 = \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{A}\right) \Rightarrow b_0 = \underset{\text{оценка}}{\hat{\beta}} \in \mathcal{N}\left(\beta_0, \sigma(\frac{1}{n}+\frac{\overline{X}^2}{A})\right).$$

Двете дисперсии клонят към нула.

<u>Оценка на σ^2 (ако не го знаем априорно)</u>: $Y_k \in \mathcal{N}(\beta_0+\beta_1 X_k, \sigma^2)$. Проблема е, че не знаем β_0 и β_1, тъй като, ако допуснем, че ги знаем, щяхме да имаме $\dfrac{Y_k-\beta_0-\beta_1 X_k}{\sigma} \in \mathcal{N}(0,1)$ и тогава

$$\frac{\sum_{k=1}^{n}(Y_k-\beta_0-\beta_1X_k)^2}{\sigma^2}\in\mathcal{X}^2(n).$$

Но, ако са ни верни допусканията за модела, тогава:

$\dfrac{\sum_{k=1}^{n}(Y_k-b_0-b_1X_k)^2}{\sigma^2}\in\mathcal{X}^2(n-2)$ („изхабили“ (използвали) сме две степени на свобода (две данни), за да оценим b_0 и b_1)

$$\mathbb{E}\frac{\sum_{k=1}^{n}(Y_k-b_0-b_1X_k)^2}{\sigma^2}=n-2$$

$\hat{\sigma}^2=\dfrac{\sum_{k=1}^{n}(Y_k-b_0-b_1X_k)^2}{n-2}$, т.е. $\sigma^2=\mathbb{E}\hat{\sigma}^2$.

Оттук нататък ние може да тестваме хипотези. Може да си конструираме множество хипотези от следния вид:

$$H_0:\beta_1=\tilde{\beta}$$
$$H_1:\beta_1=\approx{\beta}$$

$$Y=\beta_0+\beta_1X+\varepsilon$$

$\dfrac{b_1-\tilde{\beta}}{\sqrt{\sigma^2/A}}\in\mathcal{N}(0,1)$ при H_0.