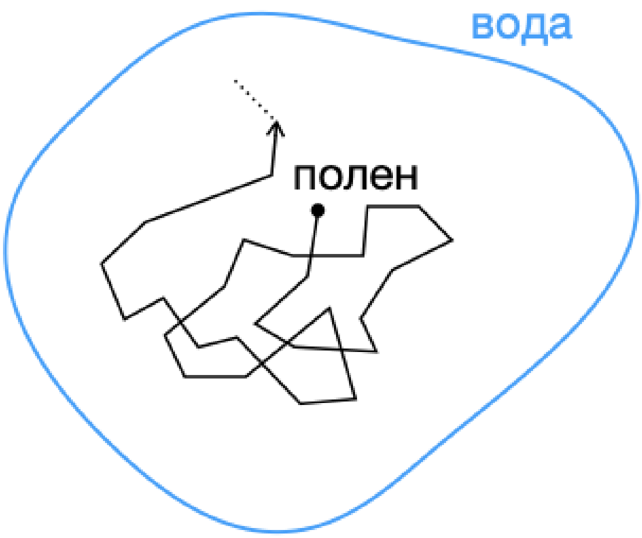
1 Лекция 1

01.10.2020

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.

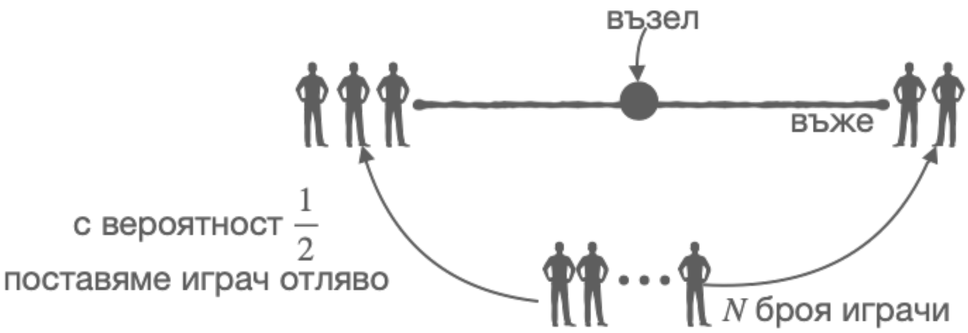
През 1905 г. Алберт Айнщайн обяснява истинската причина за това движение в семинарна статия. Благодарение на неговата кинетична теория на молекулите, той показва, че частица, поставена върху стояща вода бива удряна от молекулите на водата. Във всеки един момент от време, частицата ще бива удряна от множество 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 е с очакване 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.

След Айнщайн, Жан Перан и колектив успяват с помощта на това брауново движение да приближат броя на молекулите в изотопа C_{12} на въглерода. Това тяхно постижение им носи Нобелова награда през 1926 г.

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 това приближение.

Примери:

★ Дърпане на въже: Разполагаме случайно N души от двете страни.



Средният брой хора, които ще се разполагат отляво и отдясно ще е $N/2$. 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.

★ 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	1.1	1.2	1.3	1.4	1.5	1.6

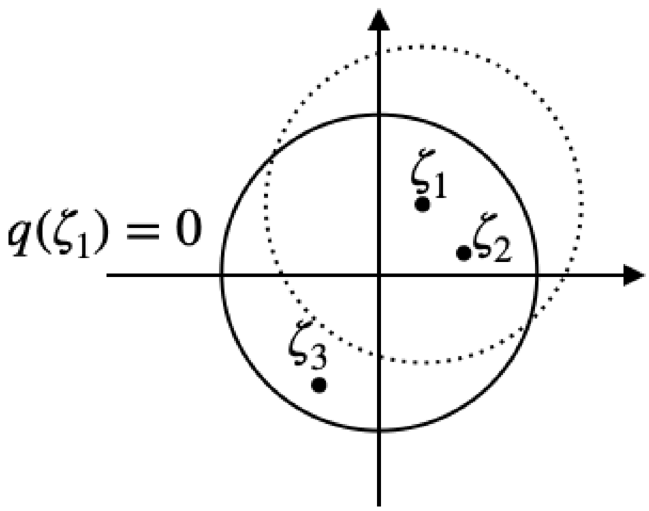
Идеята е, ако например глобата за дадено нарушение е 200 лв., шофьор, който е правел до 3 нарушения да заплаща 0.9×200 лв., такъв който е правел до 9 нарушения да заплаща 1.5×200 лв. и т.н. По-този начин ще се стимулират водачите да правят по-малко нарушения или по-точно да ги ограничават колкото се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

Казусът, който възниква касае държавния апарат за събиране на данъци и застрахователи, а именно: как да се изчислят тези коефициенти, така че сумарно платените глоби да са в очакване колкото биха били 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 е цената, която заплаща даден водач с n нарушения. В случай, че $P > 1$, ще се появи недоволство в данъкоплатците в името на държавата/застрахователите, т.к. 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 статичната система.

★ Преди около 60 години (1959 г.), акад. Сендов формулира следната хипотеза: Ако $q(\zeta)$ е полином и знаем, че нулите на полинома са в единичния кръг, то ако вземем окръжност с радиус 1 и център, която и да е нула на полинома (например ζ_1 на картинката), то в този единичен кръг (окръжността с пунктир) ще има поне една нула ζ на производната, т.е. $f'(\zeta) = 0$.



Професор Терънс Тау доказва през 2020 г. тази хипотеза за всички полиноми от достатъчно голяма степен $n \geq n_0$ (т.е. останалите са краен брой). Неговото доказателство включва много вероятностни аргументи.

Доказателство за всички n все още не е открито.

Дефиниция 1. ***Случаен експеримент*** *ще наричаме 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****.*

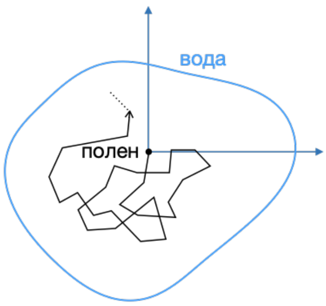
Дефиниция 2. *С Ω ще означаваме* ***множеството от всички елементарни събития*** *на даден случаен експеримент.*

<u>Примери:</u>

★ $\Omega = \{$ези, тура$\}$; $\Omega = \{0, 1\}$.

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

★ В примера с полена, $\Omega = \{$всички криви в равнината $f : f(0) = (0, 0)\}$



★ Тото „6 от 49“.Броят на всички елементарни изходи е равен на $\binom{49}{6} = 13983816$, т.е.

$$\Omega = \{\omega_1, \omega_2, ..., \omega_{13983816}\}.$$

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

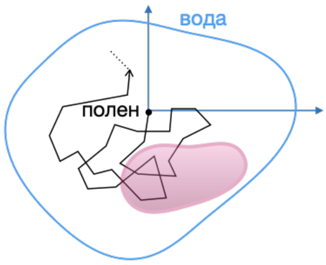
Дефиниция 3. *Всяко подмножество $A \subseteq \Omega$ наричаме* ***събитие****.*

Примери:

★ $\Omega = \{$клиенти$\}$, $A = \{$клиенти, които са жени$\}$.

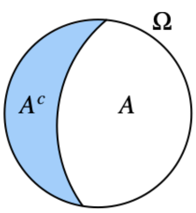
★ $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{$всички нечетни числа$\} = \{1, 3, 5\}$.

★ $A = \{$всички криви от $(0, 0)$, които достигат до розовия регион$\}$



1.1 Операции с множества

- $A \subseteq B \Leftrightarrow (\omega \in A \Rightarrow \omega \in B)$.
- $A = B \Leftrightarrow (A \subseteq B$ и $A \subseteq B)$.
- За $A, B \subseteq \Omega$, под обединението на събитията $A \cup B$ разбираме всички $\omega \in A$ **или** $\omega \in B$. Например ако $A = \{$хора между 20 и 30 години$\}$ и $B = \{$гласували за партия X$\}$, то $A \cup B =\{$хора или между 20 и 30 години, или гласували за партия X$\}$. Разбира се, последното или не е изключващо.
- За $A, B \subseteq \Omega$, под сечение на събитията $A \cap B$ разбираме всички $\omega \in A$ **и** $\omega \in B$. Например ако $A = \{$хора между 20 и 30 години$\}$ и $B = \{$гласували за партия X$\}$, то $A \cup B = \{$хора между 20 и 30 години **и** гласували за партия X$\}$.
- За $A \subseteq \Omega$, под допълнението на събитието A разбираме всички $\omega \notin A$ и бележим с A^c (или с $\overline{A}$). За горния пример, $A^c = \{$всички хора, които са по-млади от 20 г. или по-възрастни от 30 г.$\}$.



Следните свойства са класически:

- (комутативност) $A \cap B = B \cap A$ и $A \cup B = B \cup A$.
- (асоциативност) $(A \cap B) \cap C = A \cap (B \cap C)$.
- (дистрибутивност) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ и $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- (Закони на **де Морган**) $(A \cup B)^c = A^c \cap B^c$ и $(A \cap B)^c = A^c \cup B^c$. Аналогично, за (изброим) безкраен брой множества:
$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c \quad \text{и} \quad \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c\,.$$
Можем да опишем и с думи последните количества: за $A_1, A_2, \dots \in \Omega$
$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ за поне едно } i\} \quad \text{и} \quad \bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ за всички } i\}\,.$$

Дефиниция 4. *Нека Ω е съвкупност от елементарни събития и $\mathcal{A}$ е колекция от събития (т.е. от подмножества на Ω). Наричаме $\mathcal{A}$ **σ-алгебра**, ако:*

i) $\emptyset \in \mathcal{A}$;

ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;

iii) $(\forall i \geq 1 : A_i \in \mathcal{A}) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Ако последното свойство е изпълнено само за краен брой множества A_i, наричаме $\mathcal{A}$ просто алгебра.

Твърдение 5. *Ако $\mathcal{A}$ е σ-алгебра, то:*

i) $\Omega \in \mathcal{A}$;

ii) $(\forall i \geq 1 : A_i \in \mathcal{A}) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ *(т.е. имаме затвореност относно крайно/безкрайно сечение).*

Доказателство. i) Тъй като $\emptyset^c = \Omega$ и $\emptyset \in \mathcal{A}$, то $\Omega \in \mathcal{A}$.

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

□

Примери за σ-алгебри:

★ Ако $\Omega = \{0,1\}$, двете възможности са $\mathcal{A}_1 = \{\emptyset, \Omega\}$ и $\mathcal{A}_2 = 2^{\Omega} = \{\emptyset, \Omega, \{0\}, \{1\}\}$. Всъщност каквото и да е Ω, $\{\emptyset, \Omega\}$ и множеството от всички подмножества на Ω - 2^{Ω} са σ-алгебри. Те се наричат тривиални.

★ За $\Omega = \{1, 2, ..., n\}$, $|\Omega| = n$, $\mathcal{A} = 2^{\Omega}$ има 2^n елемента.

★ „6 от 49“. $\Omega = \left\{\omega_1, ..., \omega_{\binom{49}{6}}\right\}$. $|2^{\Omega}| = |2^n| = 2^{13983816}$.

Дефиниция 6. *Ако $\mathcal{B}$ е произволна колекция от събития от Ω, то $\sigma(\mathcal{B})$ е най-малката σ-алгебра, такава, че $\mathcal{B} \subseteq \sigma(\mathcal{B})$, т.е. за всяко $B \in \mathcal{B}$, $B \in \sigma(\mathcal{B})$.*

Важен пример!

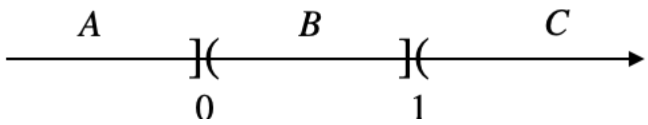
★ Да разгледаме случая $\Omega = \mathbb{R} = (-\infty, \infty)$ и $\mathcal{B} = \{$всички отворени интервали, т.е. от типа $(a,b), (a,\infty), (-\infty, b)\}$. $\sigma(\mathcal{B}) := \mathcal{B}(\mathbb{R})$ се нарича борелова σ-алгебра.

Да вземем $x \in \mathbb{R}$. Изпълнено ли е, че $\{x\}{\in}\mathcal{B}(\mathbb{R})$? Отговорът е „ДА“, тъй като може да представим точката $\{x\}$ по следния начин (виж ii) от Твърдение 5):

$$\{x\} = \bigcap_{i=1}^{\infty} \left(x - \frac{1}{i}, x + \frac{1}{i} \right) \in \mathcal{B}(\mathbb{R}).$$

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

Естествено, ако ни интересува само трихотомията



т.е. единствено в кой/кои от горните интервали попадаме, може да се ограничим само до σ-алгебрата $\mathcal{A} = \{\emptyset, \mathbb{R}, A, B, C, A \cup B, B \cup C, C \cup A\}$, която има кардиналност $|\mathcal{A}| = 8$, а не е безкрайна като бореловата.

Дефиниция 7. *Ако $\mathcal{A}$ е σ-алгебра, то $A \in \mathcal{A}$ се нарича* ***атом****, ако от $B \subseteq A$ и $B \in \mathcal{A}$, следва, че $B = \emptyset$, т.е. не съществува нетривиално подсъбитие на $\mathcal{A}$, което е част от σ-алгебрата.*

Примери:

★ За трихотомията отгоре, $|\mathcal{A}| = 8$ и атомите на $\mathcal{A}$ са A, B и C.

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

★ $T = 10001$ - брой тиражи на „6 от 49“. За всяко едно теглене имаме $\Omega_i = \left\{ \omega_1^{(i)}, \omega_2^{(i)}, ..., \omega_{\binom{49}{6}}^{(i)} \right\}$.

Разглеждайки тиражите накуп, можем да конструираме формално Ω като всички възможни комбинация от изтеглени шесторки:

$$\Omega = \bigotimes_{i=1}^{T} \Omega_i = \{(\omega_{\cdot}^{(1)}, \omega_{\cdot}^{(2)}, ..., \omega_{\cdot}^{(T)}) : \omega^{(i)} \in \Omega_i\},$$

където интерпретираме $\omega^{(i)}$ като падналата се шесторка в i-тия тираж

В такъв случай, можем да опишем събитието $A = \{$паднали са се две еднакви наредени шесторки в два последователни тиража$\}$ чрез

$$A = \bigcup_{i=1}^{10000} \{\omega \in \Omega : \omega_{\cdot}^{(i)} = \omega_{\cdot}^{(i+1)}\}.$$

Задача за мислене:

★ Разполагаме с два пощенски плика A и B, в които има съответно сумите a и b. Нямаме никаква априорна информация за сумите в тях, а човека, който ги е поставил в пликовете знае, че $a < b$. Избираме случайно с вероятност $1/2$ 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. Човекът, 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 $\mathbb{P}(C) \geq 1/2$, т.е. има ли стратегия, която ни позволява да спечелим по-голямата от двете суми, с вероятност по-голяма от $1/2$?

Интуицията подвежда и се оказва, че има такава стратегия.

2 Лекция 2

08.10.2020

Дефиниция 8. *Нека $\mathcal{A}$ е $\sigma-$алгебра върху множество от елементарни събития Ω. Тогава изображението $\mathbb{P} : \mathcal{A} \to [0,1]$ се нарича вероятност, ако са изпълнени следните три условия:*

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

ii) *Ако $A \in \mathcal{A}$ и $A^c = \Omega \setminus A$, то $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;*

iii) *Ако за всяко $i \geq 1$, $A_i \in \mathcal{A}$ и $A_i \cap A_j = \emptyset$ за $i \neq j$, то*

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(A_k\right),$$

т.е. 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.

Забележка 1. Знаците $\sqcup$ и $\uplus$ се използват за обозначаване на обединение на непресичащи се множества. В такъв случай, горното свойство може да се запише като

$$\mathbb{P}\left(\bigsqcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(A_k\right),$$

Твърдение 9. *Нека $\mathbb{P} : \mathcal{A} \to [0,1]$ е вероятност. Тогава за $A, B \in \mathcal{A}$ са изпълнени следните свойства:*

1. $\mathbb{P}(\emptyset) = 0$;
2. *Ако $A \subseteq B$, то $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$;*
3. *(Монотонност) Ако $A \subseteq B$, то $\mathbb{P}(A) \leq \mathbb{P}(B)$;*
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
5. *(Непрекъснатост) Ако $A_i \in \mathcal{A}$, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, то*

$$\mathbb{P}\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{n\to\infty} \mathbb{P}(A_n);$$

6. *Ако $A_i \in \mathcal{A}$, то*

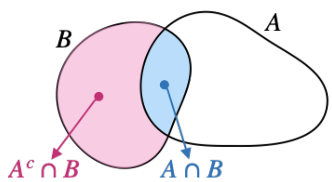
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right).$$

Това свойство е изпълнено и за всяко крайно обединение от събития, т.е., ако $n \in \mathbb{N}$

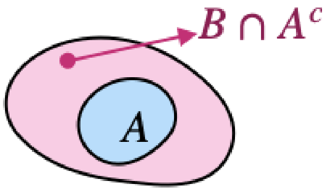
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}\left(A_i\right).$$

Доказателство. 1. Тъй като $\emptyset = \Omega^c$, то от *ii)* в дефиниция 1 следва, че $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 0$.

2. Можем да представим $B = (A \cap B) \sqcup (A^c \cap B)$ и следователно, от *iii)*, $P(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$.

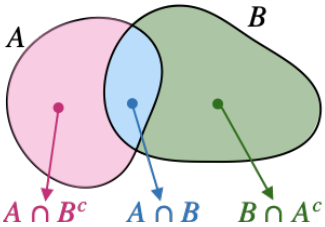


3. Подобно на 2., тъй като $B = A \sqcup (B \cap A^c)$, то $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$.



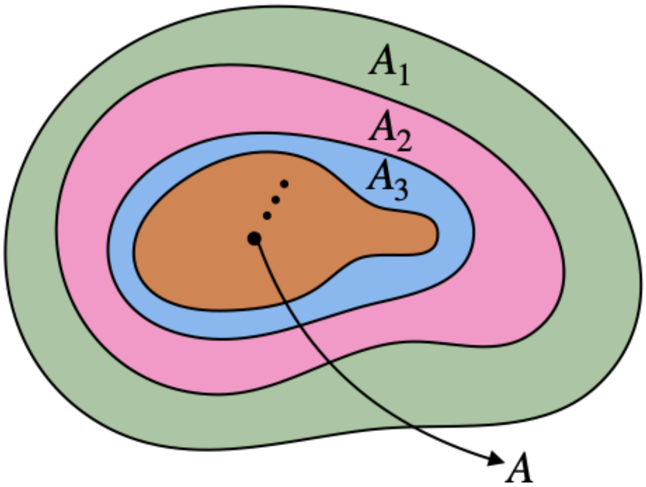
4. От $A \cup B = A \sqcup (A^c \cap B)$ и 2., получаваме, че

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\,.$$



5. Нека $A_i \in \mathcal{A}$ и $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$

   Да припомним, че $A = \bigcap_{j=1}^{\infty} A_j$ е множеството, което принадлежи на всяко едно от събитията A_i, а под $A \setminus B$ разбираме множеството A без множеството B, т.е. $A \setminus B = A \cap B^c$.



   Целта ни е да докажем, че $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

   Да конструираме множествата $B_1 = A_1 \setminus A_2$, $B_2 = A_2 \setminus A_3, B_3 = A_3 \setminus A_4$ и т.н. $B_j = A_j \setminus A_{j+1}$. Тъй като $A_1 = \bigsqcup_{j=1}^{\infty} A_j \setminus A_{j+1} \sqcup A$, то

$$1 \geq \mathbb{P}(A_1) = \sum_{j=1}^{\infty} \mathbb{P}(B_j) + \mathbb{P}(A)\,,$$

   което води до заключението, че

$$\sum_{j=1}^{\infty} \mathbb{P}(B_j) < \infty$$

   е сходящ ред. В частност, $\lim_{n\to\infty} \sum_{j=n}^{\infty} \mathbb{P}(B_j) = 0$.

   От друга страна, $A_n = \bigsqcup_{j=n}^{\infty} B_j \sqcup A$. Следователно $\mathbb{P}(A_n) = \sum_{j=n}^{\infty} \mathbb{P}(B_j) + \mathbb{P}(A)$, което след граничен преход дава желания резултат:

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A) + \lim_{n\to\infty} \sum_{j=n}^{\infty} \mathbb{P}(B_j) = \mathbb{P}(A)\,.$$

6. Целта ни ще е да докажем, че $\mathbb{P}(\bigcup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ [1].

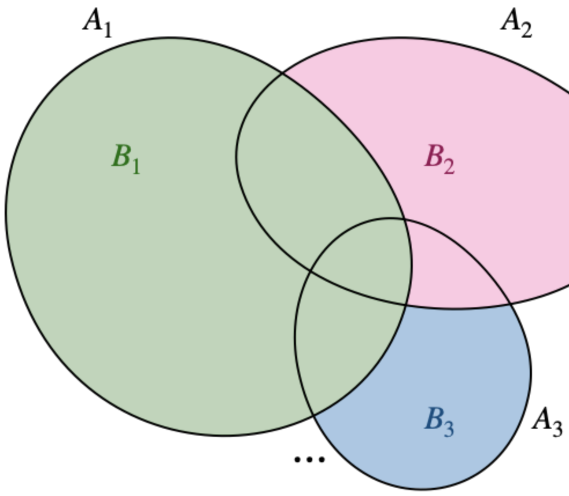
[1] Това свойство се нарича сигма субадитивност

Да отбележим първо, че от 4. по индукция следва, че за всяко $n \in \mathbb{N}$

$$\sum_{j=1}^{n} \mathbb{P}(A_j) \geq \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) .$$

Ние ще се опитаме да усилим този резултат.

Да дефинираме, подобно на миналата точка, множествата $B_1 := A_1$, $B_2 := A_2 \setminus A_1 = A_2 \cap A_1^c$, $B_3 := A_3 \setminus (A_1 \cup A_2)$, $\cdots$, $B_n := A_n \setminus (\bigcup_{j=1}^{n-1} A_j)$, $\cdots$.



Първо, да отбележим, че за всяко $n \in \mathbb{N}, B_n \subset A_n$, което води до $\mathbb{P}(B_n) \leq P(A_n)$ и, съчетавайки с дефиницията на B_i, до $B_i \cap B_j = \emptyset$ за $i \neq j$.

Следователно

$$\mathbb{P}\left(\bigsqcup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(B_j) \leq \sum_{j=1}^{\infty} \mathbb{P}(A_j) .$$

Ако докажем, че $B := \bigsqcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty} A_j := A$, то, от последното неравенство, това директно ще доведе до началното твърдение.

Тъй като $B_n \subset A_n$, то $B \subset A$ е очевидно. Да проверим, че и $A \subset B$. Нека $\omega \in A$. Тогава ω принадлежи най-малко на едно A_i, тъй като е в тяхното обединение. Нека вземем най-малкия индекс k, за който $\omega \in A_k$. Ако $k = 1$, то $\omega \in A_1 = B_1$ и значи $\omega \in B$. Ако $k \geq 1$, то, от минималността на k, $\omega \in A_k$, но $\omega \notin A_{k_1}$. В такъв случай $\omega \in A_k \setminus A_{k-1} = B_k$ и следователно във всички случаи $\omega \in A$, води до $\omega \in B$, което искахме да докажем.

$\square$

Примери за вероятности $\mathbb{P}$:

★ Ако $\Omega = \{0, 1\}$ и $\mathcal{A} = \{\emptyset, \Omega, \{0\}, \{1\}\} = 2^{\Omega}$, можем да дефинираме $\mathbb{P}(\{0\}) := p$, $\mathbb{P}(\{0\}) := 1 - p$ за някакво $p \in [0, 1]$. Това е модел на евентуално нечестна монета с две страни.

★ Дискретна вероятност: ако имаме краен брой елементарни изходи

$$\Omega = \{\omega_1, \omega_2, ..., \omega_N\} \simeq \{1, 2, ..., N\} \tag{1}$$

и можем да считаме, че $\{1, 2, ..., N\}$. Стандартният избор в дискретна ситуация е $\mathcal{A} = 2^{\Omega}$ и нека изберем числата $p_1, p_2, \ldots$, така че $\sum_{i=1}^{N} p_i = 1$. В такъв случай, можем да дефинираме за всяко $A \in \mathcal{A}$

$$\mathbb{P}(A) := \sum_{i \in A} p_i . \tag{2}$$

Например, ако $A = \{1, 3, 5\}$, то $\mathbb{P}(A) = p_1 + p_3 + p_5$.

Проверка за коректност: По дефиниция $\mathbb{P}(\Omega) = \sum_{i=1}^{N} p_i = 1$ и ако $A_1, ..., A_k$ са непресичащи се събития, то

$$\mathbb{P}\left(\bigcup_{j=1}^{k} A_j\right) = \sum_{i\in\bigcup_{j=1}^{k} A_j} p_i = \sum_{j=1}^{k}\sum_{i\in A_j} p_i = \sum_{j=1}^{k}\mathbb{P}(A_j)\,,$$

като използвахме критично, че всеки елемент i се съдържа най-много в едно A_j.

★ Ако $\Omega = \{1, 2, ..., N\}$, вероятността

$$\mathbb{P}(A) = \sum_{i\in A}\frac{1}{N} = \frac{|A|}{N}$$

наричаме равномерна. В частност $\mathbb{P}(\{i\}) = 1/N$ за всички $1 \le i \le N$, т.е. всяко едно от елементарните събития има равен шанс да се сбъдне.

★ Нека отново $\Omega = \{1, 2, ..., N\}$ и се интересуваме от събитията $A = \{i \le N : i \text{ е четно}\}$ и $A^c = \{i \le N : i \text{ е нечетно}\}$. Ако вероятността е равномерна, то

$$\mathbb{P}(A) = \frac{|A|}{N} =: p \quad \text{и} \quad \mathbb{P}(A^c) = \frac{N-|A|}{N} = 1-p\,.$$

Ако се интересуваме само от тези две събития обаче, можем да работим не върху цялата сигма алгебра 2^Ω, а само върху $\mathcal{A} = (\Omega, \emptyset, A, A^c) \subseteq 2^\Omega$.

★ По-общо, под дискретна вероятност разбираме не само такива върху крайни множества, но и такива с изброим безкраен брой елементи, т.е.

$$\Omega = \{\omega_1, \omega_2, ...\} \simeq \{1, 2, ...\}$$

и можем да считаме, че $\Omega = \mathbb{N}, \mathcal{A} = 2^{\mathbb{N}}$. В такъв случай, ако $p_i \ge 0$ са числа, за които $\sum_{i=1}^{\infty} p_i = 1$, то можем да дефинираме вероятност чрез

$$\mathbb{P}(A) := \sum_{i\in A} p_i$$

за $A \in \mathcal{A}$.

★ Ако $\Omega = \{1, 2, ...\}$, тъй като

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}\,,$$

то можем да дефинираме вероятност чрез $\mathbb{P}(\{n\}) := p_n = 6/(\pi n)^2$.

★ За $\Omega = \{0, 1, ...\} := \mathbb{N}_0$, то възможен избор е $p_i = qp^i, p + q = 1$ за $i \in \mathbb{N}_0$. Защо тази дефиниция е коректна?

★ Ако $\Omega = \{1, 2, ...\}$, не можем да дефинираме равномерно разпределение. Виждате ли защо?

3 Лекция 3

15.10.2020

Да припомним, че равномерна вероятност е такава, при която всички елементарни събития имат една и съща вероятност да се сбъднат. Тя е прототип на случая, в който нямаме никаква априорна информация за модела и не искаме да придаваме на нито един от възможните изходи по-голяма или по-малка възможност за сбъдване от тези на останалите.

Математически, разглеждаме $\Omega = \{1, 2, \dots, N\}$ и вероятност $p_i := \mathbb{P}(\{i\}) = 1/N$ за $1 \le i \le N$.

Ако преминем към изброима безкрайност от елементарни събития, т.е.

$$\Omega = \{w_1, w_2, ..., w_n, ...\} \simeq \mathbb{N}\,,$$

и вероятност върху Ω

$$p_i := \mathbb{P}(\{\omega_i\}) \ge 0, \quad \sum_{i=1}^{\infty} p_i = 1\,,$$

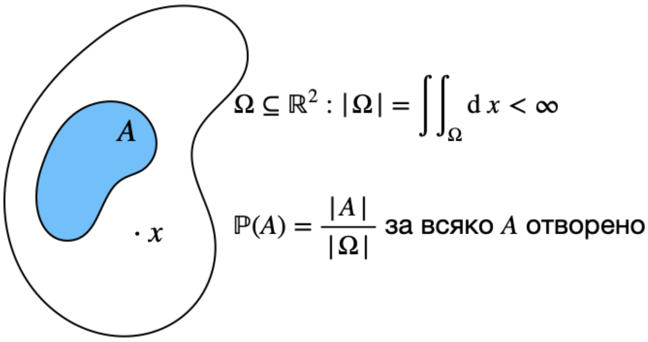
то не е възможно да зададем равномерна вероятност. Наистина, ако всички p_i са равни, тъй като са безкраен брой, няма как да се сумират до 1.

Все пак върху безкрайни Ω съществуват много други интуитивни вероятности (вероятностни разпределения), с които постепенно ще се запознаем.

Геометрична вероятност.

Тази вероятност е пример за вероятностно разпределение върху неизброимо множество. Най-простият прототип, който ще разгледаме, отново разпределя "равномерно" вероятностите.

Да дефинираме вероятността нещо да се случи в $A \subseteq \Omega$), като площта (мярката) на A върху площта на цялото Ω.

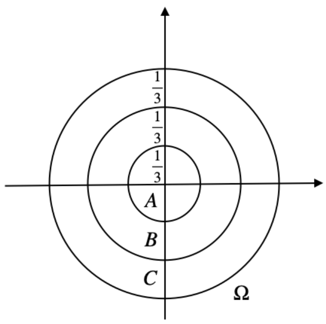


Това е пример за равномерна вероятност върху Ω - тя зависи само от площта на (събитието) A и не зависи например от нейното разположение или форма.

Площта на една точка $x \in \Omega$ е равна на 0 и значи $\mathbb{P}(\{x\}) = \mu(\{x\})/\mu(\Omega) = 0$.

Примери:

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



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

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

★ *Идея на Монте Карло алгоритмите*: Разполагаме например

и искаме да пресметнем лицето на синята област $A \subseteq \Omega$. Тъй като вероятността да попаднем в A е $\mu(A)/\mu(\Omega)$, идеята е да генерираме голям брой случайни точки от Ω и да оценим тази вероятност като

$$\frac{\mu(A)}{\mu(\Omega)} \approx \frac{\text{брой точки, попаднали в } A}{\text{общ брой генерирани точки}}.$$

Ако знаем $\mu(\Omega)$, получаваме оценка за $\mu(A)$. Колкото повече точки включваме в изследването, толкова по-добро ще е приближението към площта на A. По този начин може да пресметнем интеграл без числени методи.

Един задълбочен анализ на такива алгоритми би включвал колко "много" точки ще са ни нужни $(10, 100, 1000, 10000, \dots)$, за да получим приближение, което има грешка от например 0.01.

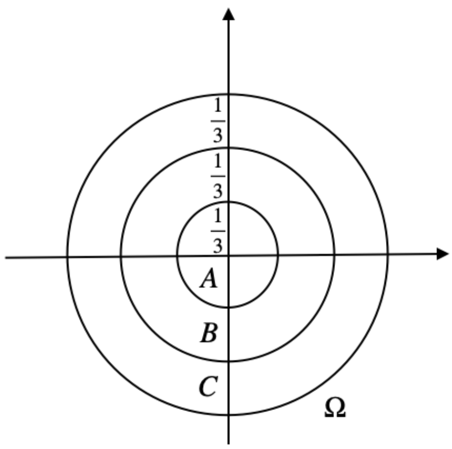
Вероятностно пространство.

Дефиниция 10. *Наредена тройка $(\Omega, \mathcal{A}, \mathbb{P})$, където Ω е пространство от елементарни събития, $\mathcal{A} \subseteq 2^\Omega$ е σ-алгебра и $\mathbb{P} : \mathcal{A} \to [0,1]$ е вероятностна мярка се нарича* ***вероятностно пространство****.*

Примери за вероятностни пространства:

★ Едно вероятностно пространство е например тройката $\Omega = \{1, 2, ..., N\}$, $\mathcal{A} = 2^\Omega$ и вероятността, дефинирана чрез $\mathbb{P}(\{i\}) = 1/N$ за $i \in \Omega$. Това е математическата формулировка на равномерно вероятностно пространство, т.е. всяко елементарно събитие има един и съща вероятност да се сбъдне.

★ Да се върнем на примера с мишената.



В него $\Omega = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, $\mathcal{A} = 2^\Omega = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, C \cup A\}$ и вероятностите на всички възможни събития са съответно

$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1, \mathbb{P}(A) = \frac{1}{9}, \mathbb{P}(B) = \frac{3}{9}, \mathbb{P}(C) = \frac{5}{9}, \mathbb{P}(A \cup B) = \frac{4}{9}, \mathbb{P}(B \cup C) = \frac{8}{9}, \mathbb{P}(C \cup A) = \frac{6}{9}.$$

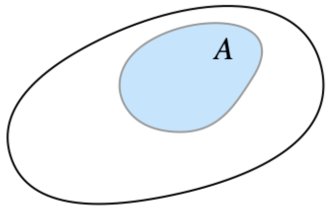
★ Ако ни интересуват не само трите региона по-горе, а всички "достатъчно хубави" множества в кръга (например отворените), можем да разгледаме тройката $\Omega = \{x^2 + y^2 \leq 1\}$, Бореловата сигма алгебра $\mathcal{A} = \mathcal{B}(\mathbb{R})$ и вероятността за $\mathbb{P}(A) = S_A / S_\Omega$ $A \in \mathcal{A}$.

Условна вероятност.

Следващата ситуация, която ще опитаме да формализираме, е постъпването на нова информация като например дали някакво събитие се е случило, постъпването/реализирането на нови данни и др.

Възможен сценарий е да разполагаме с изграден алгоритъм за изкуствен интелект, но да искаме да променяме параметрите на база нови данни, които идват в реално време.

Малко по-формално, да предположим, че $A \in \mathcal{A}$ настъпва.



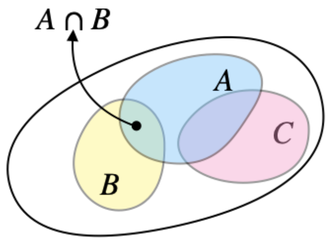
Първоначално възможните елементарни събития са елементите на Ω, но ако знаем, че събитието A е настъпило, то много от тях всъщност са излишни, тъй като сме сигурни, че няма да се реализират. Това променя цялото вероятностно пространство и съответно разпределение. Въпросът е как това се случва?

Дефиниция 11. *Нека $(\Omega, \mathcal{A}, \mathbb{P})$ е вероятностно пространство и $A \in \mathcal{A}$ е такова, че $\mathbb{P}(A) > 0$. В такъв случай, можем да дефинираме нова вероятност върху $(\Omega, \mathcal{A})$ чрез*

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

*за всяко $B \in \mathcal{A}$. Тази вероятност, която ще бележим с $P_A(B)$ или $P(A|B)$, се нарича **условна вероятност при условие** A.*

Интерпретацията е, че знаем, че е настъпило събитието A и тази част от B, която не е в A не ни интересува! Вече, ако искаме да пресметнем новата вероятност за сбъдване на B, е нужна вероятността само на тази част, която попада в $\mathcal{A}$, т.е. $A \cap B$.



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

Примери:

★ Тото 6 от 49

Пуснали сме фиш $B = \{1, 2, 3, 4, 5, 6\}$. При тегленето чуваме само, че последните изтеглени числа са 2 и 4, т.е. се е случило събитието $A = \{\text{шесторки } \omega \in \Omega : 2, 4 \in \omega\}$ Тъй като $B \subset A$, то

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

и следователно

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{1/\binom{49}{6}}{\binom{47}{4}/\binom{47}{6}} = \frac{1}{13983816} \cdot \frac{\binom{47}{6}}{\binom{47}{4}} = \frac{1}{178365}\,.$$

Оказва се, че вероятността за печалба нараства значително (от порядъка на $70 - 80$ пъти).

★ На избори се явяват две партии - Π_1 и Π_2 със съответен брой избиратели N_1, N_2, от които m_1 и m_2 са определени като млади.

Пита се случаен млад човек за коя партия е гласувал. Нека $A = \{\text{млади}\}$ и $B = \{\text{гласували за } \Pi_1\}$. Сега питаме каква е вероятността да е гласувал за Π_1? Имайки информацията A, това е точно

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{m_1}{N_1+N_2}}{\frac{m_1+m_2}{N_1+N_2}} = \frac{m_1}{m_1+m_2}\,.$$

Забележете, че тази вероятност не зависи от общия брой гласоподаватели, а само от броя на младите такива - m_1 и m_2!

Независимост

Дефиниция 12. *Две събития A и B се наричат независими, ако $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*
Ако $\mathbb{P}(A) > 0$, то от горното би следвало, че

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

т.е. независимостта би означавала, че случването на събитието A не ни носи никаква информация за B.

Дефиниция 13. *Събитията $A_1, A_2, ..., A_n$ се наричат независими в съвкупност (или взаимно независими), ако за всяко подмножество I на $\{1, 2, \ldots, n\}$*

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i)\,.$$

Теорема 14. *Нека $A_1, A_2, ..., A_n$ са n събития, за които е изпълнено*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) > 0\,.$$

Тогава

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right) \times \mathbb{P}\left(A_{n-1} \middle| \bigcap_{i=1}^{n-2} A_i\right) \times \cdots \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1)\,. \tag{3}$$

Доказателство. По индукция. За $n = 1 : \mathbb{P}(A_1) = \mathbb{P}(A_1)$. Нека допуснем, че (3) е вярно за n. Тогава, тъй като

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

след използване на индукционното предположение, получаваме точно искания резултат за $n+1$ и индукцията е завършена. □

Пример:

★ 6 от 49

Да разгледаме първите 10001 тиража на "6 от 49". Тогава можем да считаме, че

$$\Omega = \{w = (w_1, w_2, \ldots, w_{10001})\}\,,$$

където $w_i = \left(w_i^{(1)}, \ldots, w_i^{(6)}\right)$ е i-тата паднала се шесторка. Искаме да оценим вероятността на събитието A - да се паднат еднакви числа в два поредни тиража, т.е.

$$A = \bigcup_{i=1}^{10000} A_i, \quad A_i = \{w \in \Omega | w_i = w_{i+1}\}\,.$$

Оказва се, че събитията A_i, а значи и $\overline{A_i}$ (опитайте да докажете като упражнение), са независими в съвкупност (формалното доказателство е доста тромаво, а интуитивно можем да мислим, че, ако $w_i = w_{i+1}$ това не ни носи никаква информация за наличието на други еднакви поредни шесторки). Следователно

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{i=1}^{10000} A_i\right) = 1 - \mathbb{P}\left(\overline{\bigcup_{i=1}^{10000} A_i}\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{10000} \overline{A_i}\right) = 1 - \prod_{i=1}^{10000} \mathbb{P}(\overline{A_i}) \\
&= 1 - \mathbb{P}(\overline{A_1})^{10000} = 1 - \left(\frac{\binom{49}{6} - 1}{\binom{49}{6}}\right)^{10000} = 1 - \left(1 - \frac{1}{\binom{49}{6}}\right)^{10000} \\
&\approx 1 - \left(1 - 10000 \cdot \frac{1}{\binom{49}{6}}\right) \approx \frac{1}{1400}\,.
\end{aligned}$$

Формула за пълната вероятност

Дефиниция 15. *Групата от множества $H_1, H_2, ..., H_n$ се нарича* ***пълна група от събития****, ако за всеки различни $1 \le i, j \le n$, $H_i \cap H_j = \emptyset$ и $\bigcup_{i=1}^{n} H_i = \Omega$.*

В друг контекст, казваме, че $H_1, \ldots, H_n$ е разбиване на Ω. Както отбелязахме и по-рано, можем да използваме символа $\sqcup$ или $\uplus$ за обединение на непресичащи се множества, т.е. $\Omega = \sqcup_{i=1}^{n} H_i$.

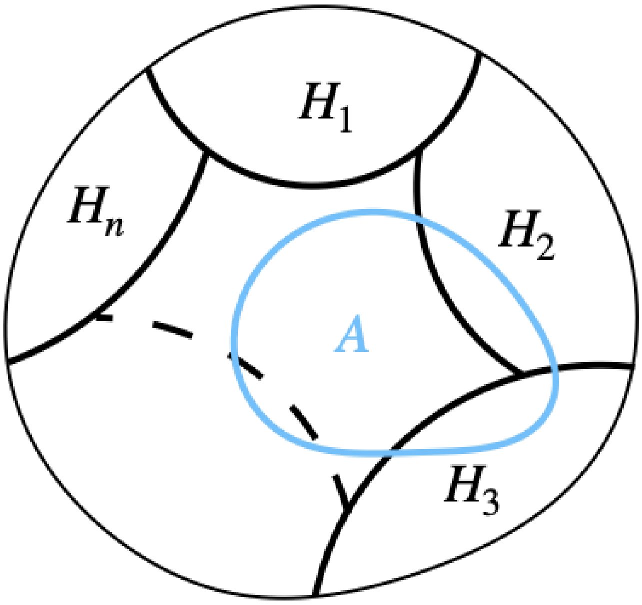
Теорема 16 (**Формула за пълната вероятност**). *Нека* $H_1, H_2, ..., H_n$ *е пълна група от събития в* Ω *и* $A \in \mathcal{A}$. *Тогава*

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

с конвенцията, че ако $\mathbb{P}(H_i) = 0$, *то* $\mathbb{P}(A|H_i)\mathbb{P}(H_i) = 0$.

Доказателство. Имаме, че

$$A = A \cap \Omega = A \cap \bigsqcup_{i=1}^{n} H_i = \bigsqcup_{i=1}^{n} (A \cap H_i) \ .$$



Следователно, от дефиницията на $\mathbb{P}$ и $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$, получаваме

$$\mathbb{P}(A) = \mathbb{P}\left(\bigsqcup_{i=1}^{n} A \cap H_i\right) = \sum_{i=1}^{n} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{n} \mathbb{P}(A|H_i)\mathbb{P}(H_i) \, .$$

□

Теорема 17 (**Формула на Бейс**). *Нека* $H_1, H_2, ..., H_n$ *е пълна група от събития в* Ω, $A \in \Omega$ *и* $\mathbb{P}(A) > 0$. *Тогава за всяко* $1 \leq k \leq n$

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\sum_{i=1}^{n} \mathbb{P}(A|H_i)\mathbb{P}(H_i)} \, .$$

Доказателство. Имаме, че $\mathbb{P}(A \cap H_k) = \mathbb{P}(A|H_k)\mathbb{P}(H_k)$, както и $\mathbb{P}(H_k|A)\mathbb{P}(A) = \mathbb{P}(H_k \cap A)$. Следователно

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\mathbb{P}(A)} \, .$$

Второто равенство от условието получаваме като приложим формулата за пълната вероятност за $\mathbb{P}(A)$.

□

Формулата на Бейс показва как актуализираме нашата априорна вероятност за хипотеза да се случи, при положение, че е настъпила някаква информация A.

Примери:

★ На летище се предлага бърз COVID тест, който индикира дали даден човек е заразен или не. Знае се, че инфектираните са 1% от посетителите на летището.

Ако човек **е** носител на вируса, тестът засича вярно с вероятност с 99% и индикира за наличието на вирус.

Ако човек **не е** носител на заразата, тогава тестът правилно връща с вероятност 80%, че е здрав.

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

Решение:

Нека $Pos = \{$тестът е положителен$\}$ и $Sick = \{$човекът е болен$\}$. Търсим $\mathbb{P}(Sick|Pos)$. От формулата на Бейс

$$\mathbb{P}(Sick|Pos) = \frac{\mathbb{P}(Sick \cap Pos)}{\mathbb{P}(Pos)} = \frac{\mathbb{P}(Pos|Sick)\mathbb{P}(Sick)}{\mathbb{P}(Pos|Sick)\mathbb{P}(Sick) + \mathbb{P}(Pos|\overline{Sick})\mathbb{P}(\overline{Sick})} = \frac{99\% \times 1\%}{99\% \times 1\% + 20\% \times 99\%} = \frac{1}{21}\,.$$

Оказва се, че теста не е много полезен. Получава се така, защото заразените са много малък процент от популацията, а грешката от 20% е твърде голяма в сравнение с него.

★ Да разгледаме метода, когато се взимат n проби, смесват се и се тества получената смес. Тестът ще е положителен, ако поне една от пробите е положителна. В такъв случай ще трябва да направим n индивидуални теста, за да открием кой/кои са болни.

От друга страна, ако всички са здрави, ще сме го открили само с един тест, а не с n.

Да допуснем, че вероятността човек да е болен е 2%. Как да подберем n, така че да минимизираме използваните тестове?

★ В два плика са сложени съответно a и b лева, като знаем единствено, че $a, b > 0$. Някой друг (водещия на играта например) знае, че $a < b$.

Възможно ли е да измислим стратегия, при която $\mathbb{P}(\text{да изберем } b) > 1/2$, т.е. да избирам по-голямата сума с вероятност по-голяма от 50%)?

Решение:

Нека $A = \{$виждаме a в първия плик$\}$ и $B = \{$виждаме b във втория плик$\}$.

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

$$\begin{aligned}
\mathbb{P}(\text{избираме } b) &= \mathbb{P}(A \cap C) + \mathbb{P}(B \cap \overline{C}) = \mathbb{P}(C|A)\mathbb{P}(A) + \mathbb{P}(\overline{C}|B)\mathbb{P}(B) \\
&= \frac{1}{2}\left(\mathbb{P}(C|A) + \mathbb{P}(\overline{C}|B)\right) = \frac{1}{2}\left(\mathbb{P}(C|A) + 1 - \mathbb{P}(C|B)\right) \\
&= \frac{1}{2} + \frac{1}{2}\left(\mathbb{P}(C|A) - \mathbb{P}(C|B)\right)\,,
\end{aligned}$$

като използвахме, че

$$\mathbb{P}(C|B) + \mathbb{P}(\overline{C}|B) = \frac{\mathbb{P}(C \cap B) + \mathbb{P}(\overline{C} \cap B)}{\mathbb{P}(B)} = 1\,.$$

Следователно 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

$$J := \mathbb{P}(C|A) - \mathbb{P}(C|B) = \mathbb{P}(C) \cdot (\mathbb{P}(A|C) - \mathbb{P}(B|C)) = \mathbb{P}(C) \cdot (\mathbb{P}(B) - \mathbb{P}(A))\,.$$

(а) Ако никога не сменяме, т.е. $\mathbb{P}(C) = 0$, то $J = 0$.

(б) Ако сменяме на всяко второ теглене, т.е. $\mathbb{P}(C) = \frac{1}{2}$, то отново $J = 0$.

(в) Ако винаги сменяме, т.е. $\mathbb{P}(C) = 1$, също получаваме $J = 0$.

Да дефинираме обаче стратегията по следния начин: ако виждаме числото x, сменяме с вероятност e^{-x}. Тогава $\mathbb{P}(C|\text{виждаме } x) = e^{-x}$ и $J = e^{-a} - e^{-b} > 0$.

4 Лекция 4

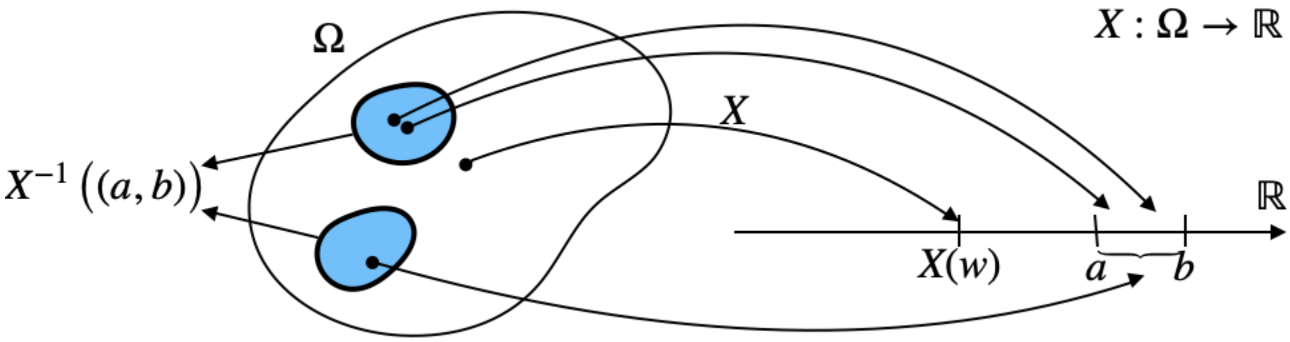
22.10.2020

Целта ни в тази лекция ще е да се запознаем с понятието случайна величина. Често при дефинирането на това понятие се отбелязва, че една случайна величина X не е нито случайна, нито величина. Това е грубо казано функция/изображение, което съпоставя на всяко елементарно събитие $\omega \in \Omega$ някакво реално число.

За да бъде функцията $X : \Omega \to \mathbb{R}$ случайна величина, тя трябва да има определени свойства.

Дефиниция 18. *Нека $V = (\Omega, \mathcal{A}, \mathbb{P})$ е вероятностно пространство. Функцията $X : \Omega \to \mathbb{R}$ се нарича* ***случайна величина****, когато за всеки две реални числа a и b, $a < b$ е в сила $X^{-1}((a, b)) \in \mathcal{A}$, където за $B \subseteq \mathbb{R}$, $X^{-1}(B) := \{\omega \in \Omega | X(\omega) \in B\}$.*

Ако горното свойство е изпълнено, това ни позволява да кажем каква е вероятността X да бъде между a и b, тъй като тогава множеството $\{\omega \in \Omega | X(\omega) \in (a, b)\}$ би било в $\mathcal{A}$, откъдето е дефинирана и вероятността $\mathbb{P}$.



Горната картинка изобразява, че множеството $X^{-1}((a, b))$ е съставено точно от тези елементарни събития $\omega \in \Omega$, за които $X(\omega) \in (a, b)$.

Забележка 2. Ако X е случайна величина, е вярно, че $X^{-1}(I) \in \mathcal{A}$ не само за $I = (a, b)$, но и за $I = (a, b]$, $I = [a, b]$ и $I = \{a\}$. Всеки интервал (a, b) от $\mathbb{R}$ има праобраз $X^{-1}((a, b)) \subseteq \Omega$, като този праобраз може да бъде и празното множество $\emptyset \subseteq \Omega$.

За да не отбелязваме във всяко твърдение, оттук нататък ще работим винаги върху вероятностното пространство $V = (\Omega, \mathcal{A}, \mathbb{P})$.

Теорема 19 (**Свойства на случайни величини**). *Нека $X, Y : \Omega \to \mathbb{R}$ са случайни величини и $a, b \in \mathbb{R}$. Тогава е в сила:*

1. *За всеки $a, b \in \mathbb{R}$, $aX \pm bY$ е случайна величина;*
2. *aX е случайна величина (частен случай на 1. за $b = 0$);*
3. *XY е случайна величина;*
4. *ако $\mathbb{P}(Y = 0) = 0$, то X/Y е случайна величина.*

Случайните величини са функции, за които е възможно да не 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. В такъв случай бихме могли да работим и само с информация, например от типа $\mathbb{P}(X \in (a, b))$ и др. без да прецизираме Ω или $X(\omega)$ за всички $\omega \in \Omega$.

Дискретни случайни величини

Дефиниция 20. *Нека $H \subseteq \Omega$. Тогава функцията $1_H : \Omega \to \mathbb{R}$, дефинирана чрез*

$$1_H(\omega) = \begin{cases} 1, & \omega \in H; \\ 0, & \omega \notin H. \end{cases}$$

се нарича ***индикаторна функция****.*

Твърдение 21. *Нека $H \in \mathcal{A}$. Тогава функцията $1_H : \Omega \to \mathbb{R}$ е случайна величина.*

Доказателство. Тъй като $X(w) := 1_H(w)$, то

$$X^{-1}(\{0\}) = \overline{H} \quad \text{и} \quad X^{-1}(\{1\}) = H.$$

Следователно за всички $a < b$ е вярно, че

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

В такъв случай, независимо кой интервал (a,b) изберем, $X^{-1}(a,b)$ ще бъде една от четирите възможности: $\emptyset$, Ω, H, или $\overline{H}$. За да е изпълнена дефиницията на случайна величина, трябва $X^{-1}(a,b)$ да е елемент на σ-алгебрата $\mathcal{A}$: $\emptyset$ и Ω са нейни елементи по дефиниция, а по условие H също е неин елемент и отново от дефиницията на σ-алгебра, то и $\overline{H}$ също принадлежи на $\mathcal{A}$. Следователно, $X = 1_H$ е случайна величина и лемата е доказана. $\square$

Ще отбележим, че ако X е индикаторна случайна величина, т.е. $X \equiv 1_H$ за някакво $H \in \mathcal{A}$, то X се характеризира чрез $\mathbb{P}(X = 1)$ и $\mathbb{P}(X = 0) = 1 - \mathbb{P}(X = 1)$. Това е удобно при моделиране на всякакви събития с два изхода - успех/неуспех, ези/тура и др.

Да въведем за удобство следните обекти:

- $\overline{x}$ за (евентуално изброимо безкраен) вектор от различни числа:

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

- вероятностно пространство V.
- пълна група от събития $\mathcal{H} = \{H_1, H_2, \dots\}$ във V.

Дефиниция: (*Дискретна случайна величина*) Функцията $X : \Omega \to \mathbb{R}$, дефинирана чрез $X(\omega) = \sum_{j=1}^{n} x_j 1_{H_j}(\omega)$, т.е. приемаща n различни стойности $x_1, \dots, x_n$ (или $X(\omega) = \sum_{j=1}^{\infty} x_j 1_{H_j}(\omega)$, когато приема изборимо безкраен брой стойности $x_1, x_2, \dots$) се нарича дискретна случайна величина. За краткост, ще използваме и записа

$$X = \sum_j x_j 1_{H_j},$$

където индексното множество се подразбира от контекста и $H_j = \{X = x_j\} = \{w \in \Omega : X(w) = x_j\}$.

Дефиниция: (*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	x_3	$\dots$	x_k	$\dots$
$\mathbb{P}$	p_1	p_2	p_3	$\dots$	$p_k = \mathbb{P}(X = x_k) = P(H_k)$	$\dots$

се нарича разпределение на X.

★ Нека X е случайната величина, която измерва броя дни които работи дадено CPU без повреда. Можем да изберем да моделираме X чрез

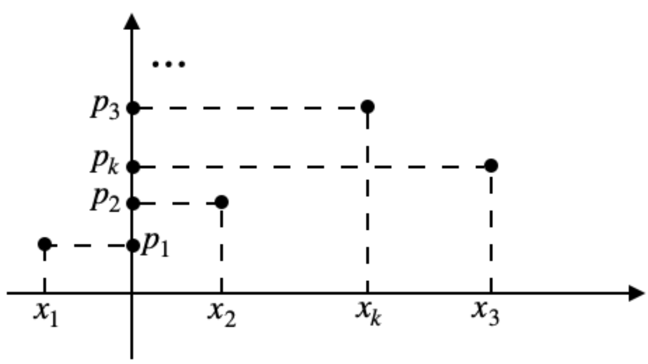
$$X \in \mathbb{N}_0^+ = \{0, 1, 2, ...\}$$

или например чрез

$$X \in \{0, 1, 2, ..., 1000\},$$

в случай, че се интересуваме само до ден 1000.

Дефиниция: (*Хистограма*) Графиката, която изобразява точките $(x_j, \mathbb{P}(X = x_j)$ се нарича хистограма (виж по-долу за илюстрация).



Хистограма на дискретна случайна величина със стойности x_i и съответни вероятности за тях $p_i = \mathbb{P}(X = x_i)$.

★ Хвърляме два стандартни зара. Нека X и Y са случайните величини, съответстващи на падналите се точки на всеки от тях. Нека Z е падналият се сбор, т.е. $Z = X + Y$. По-долу може да видите разпределението и хистограмата на Z.

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

Разпределение на $Z = X + Y$.

Хистограма на $Z = X + Y$.

Смяна на променливите за дискретни случайни величини

Нека X е дискретна случайна величина, $g : \mathbb{R} \to \mathbb{R}$ и $Y = g(X)$. Искаме да се уверим, че Y също е добре дефинирана случайна величина, т.е. Y може да се запише като

$$Y = \sum_i y_i \mathbb{1}_{\widetilde{H}_i}$$

за различни y_i и пълна група от събития $(\widetilde{H}_i)$.

Да запишем X в каноничния ѝ вид:

$$X = \sum_j x_j 1_{H_j}.$$

В такъв случай,

$$Y = g(X) = \sum_j g(x_j) 1_{H_j}.$$

Все пак, трябва да имаме предвид, че е възможно, $g(x_m) = g(x_k)$ за някои $m \neq k$ и следователно не можем просто да положим $y_i = g(x_i)$, а трябва да дефинираме $y_1, y_2, \dots$ като различните стойности на $g(x_1), g(x_2), \dots$. След това $\widetilde{H}_i$ се формират като обединението на тези H_j, за които $y_i = g(x_j)$.

★ Отново да дефинираме X да бъдат броят дни на работа преди повреда на дадено CPU и да запишем $X = \sum_{j=0}^{\infty} j 1_{\{H_j\}}$, където $H_j = \{CPU$ работи точно j дни$\}$. Ако искаме да проверяваме дали това CPU е работило поне 100 дни без проблем, то подходяща случайна величина е $Y = g(X)$, където

$$g(n) = \begin{cases} 0, & n = 0, 1, 2, \dots 99; \\ 1, & n \geq 100. \end{cases}$$

Тогава стойностите на Y са 0 и 1 и се реализират съответно върху множествата $\widetilde{H}_0 = H_0 \cup H_1 \cup H_2 \cup \dots \cup H_{99}$ и $\widetilde{H}_1 = H_{100} \cup H_{101} \cup \dots$

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

Независимост на дискретни случайни величини

Дефиниция: (*Независимост на дискретни случайни величини*) Нека X, Y са дискретни случайни величини във вероятностното пространство V. Тогава казваме, че X и Y са независими и пишем $X \perp\!\!\!\perp Y$, ако за всеки j и k е изпълнено, че

$$\mathbb{P}(X = x_j, Y = y_k) = \mathbb{P}\left(\{X = x_j\} \cap \{Y = y_k\}\right) = \mathbb{P}(X = x_j)\mathbb{P}(Y = y_k).$$

★ Да разгледаме хвърлянето на две честни монети. Математически можем да опишем това чрез избора $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$, $\mathcal{A} = 2^{\Omega}$ и $\mathbb{P}(\{0,0\}) = \mathbb{P}(\{0,1\}) = \mathbb{P}(\{1,0\}) = \mathbb{P}(\{1,1\}) = 1/4$, т.е например кодираме ези/тура с $0/1$.

Да дефинираме $X_i : \Omega \to \mathbb{R}$ чрез $X_i(\omega) = i$-та компонента на ω (например $X(\{0,1\}) = 0$ и $Y(\{0,1\}) = 1$. В този случай може да проверите, че е изпълнена дефиницията за независимост, т.е. за всеки $i, j \in \{0,1\}$,

$$\mathbb{P}(X_1 = i, X_2 = j) = \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = j).$$

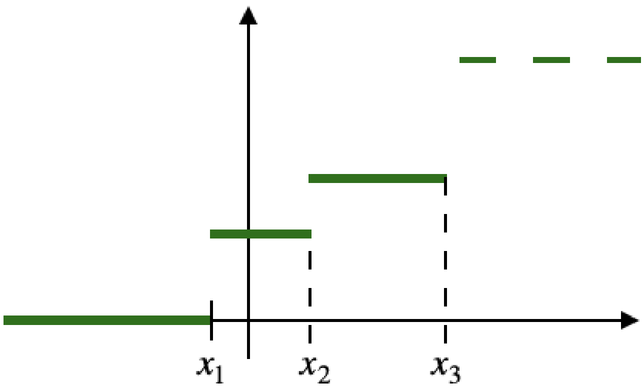
Това е математическия формализъм за интуитивно ясното, че двете монети не влияят една на друга.

Дефиниция 22. *(Функция на разпределение на случайна величина.[2]) Нека X е случайна величина във вероятностното пространство V. Тогава функцията $F_X : \mathbb{R} \to [0,1]$, дефинирана чрез $F_X(x) = \mathbb{P}(X \leq x)$, се нарича функция на разпределение на X.*

★ Ако X е дискретна случайна величина със стойности $x_1 < x_2 < \ldots$, то

$$F_X(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x \in [x_1, x_2) \\ p_1 + p_2, & x \in [x_2, x_3) \\ \ldots \\ p_1 + \ldots + p_k, & x \in [x_k, x_{k+1}) \\ \ldots \end{cases}$$

и F_X е стъпаловидна нарастваща функция (виж по-долу за илюстрация).

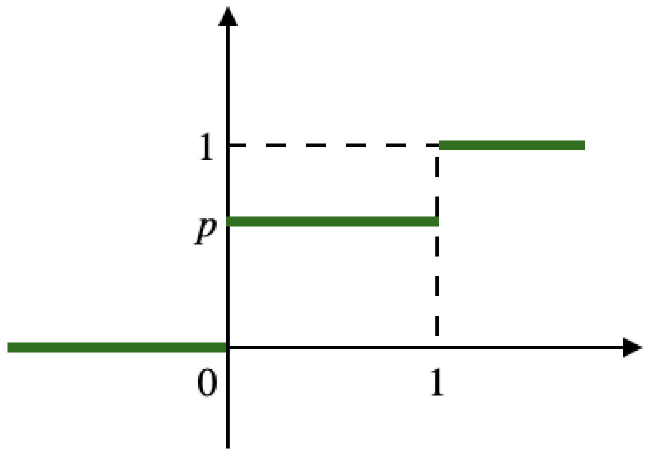


Скица на функция на разпределение на дискретна случайна величина.[3]

★ Ако $X = 1_H$ за някакво множество $H \subset \mathcal{A}$ с $\mathbb{P}(H) = p$, то

$$F_X(x) = \begin{cases} 0, & x < 0 \\ p, & x \in [0,1) \\ 1 & x \geq 1. \end{cases}$$

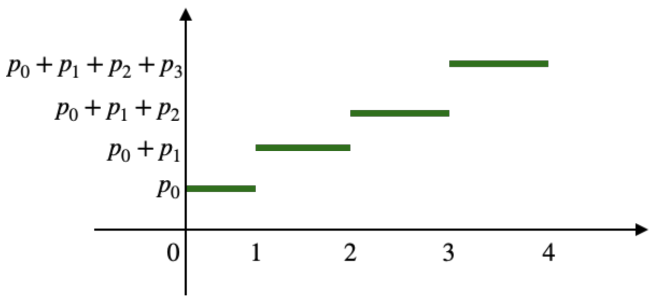
[2]Често срещано съкращание за функцията на разпределение е cdf от английското наименование cumulative distribution function



Скица на функцията на разпределение на X при $\mathbb{P}(X = n) = p_n$ за $n = 0, 1, 2, \ldots$

★ В примера от по-рано с броя дни на работа на CPU, т.е. $P(X = n) = p_n$ с $n = 0, 1, 2, \ldots$, то

$$F_X(x) = \begin{cases} 0, & x < 0 \\ p_0, & x \in [0, 1) \\ p_0 + p_1 & x \geq [1, 2) \\ \ldots \end{cases}$$



Скица на функция на разпределение на дискретна случайна величина.

В следващото твърдение отбелязваме някои свойства на F_X.

Твърдение 23. *(Свойства на фунцията на разпределение F_X.)*

1. *F_X е растяща фунцкия.*
2. $\lim_{x \to -\infty} F_X(x) = 0$ *и* $\lim_{x \to \infty} F_X(x) = 1$.
3. *F_X е непрекъсната отдясно, т.е. за всяко $x_0 \in \mathbb{R}$,* $\lim_{x \to x_0+} F_X(x) = F_X(x_0)$
4. *F_X има граници отляво, т.е. за всяко $x_0 \in \mathbb{R}$ съществува границата* $\lim_{x \to x_0-} F_X(x)$.

Математическо очакване на дискретни случайни величини

Дефиниция 24. *(Очакване на дискретна случайна величина.) Нека X е дискретна случайна величина. Ако сумата $\sum_j x_j \mathbb{P}(X = x_j)$ е крайна, то дефинираме математическо очакване $\mathbb{E}[X]$ на X като*

$$\mathbb{E}[X] := \sum_j x_j \mathbb{P}(X = x_j).$$

Ако X приема краен брой стойности, със сигурност $\sum_j x_j \mathbb{P}(X = x_j)$ е крайно (защо?) и значи $\mathbb{E}[X]$ е добре дефинирано. Ако обаче приема безкраен брой стойности, то това може и да не е изпълнено.

★ Да припомним, че
$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \pi^2/6 =: C,$$
като от значение в случая ще е само, че сумата е крайна, а не конкретната ѝ стойност. Следователно можем да дефинираме случайна величина X върху естествените числа чрез
$$\mathbb{P}(X = n) = \frac{1}{Cn^2}.$$
Следователно
$$\sum_{n=1}^{\infty} n\mathbb{P}(X = n) = \frac{1}{C} \sum_{n=1}^{\infty} \frac{1}{n},$$
което не е крайно и следвателно X няма математическо очакване според горната дефиниция.

★ Ако имаме равномерно разпределение върху $\{x_1, x_2, ..., x_n\}$, т.е. $\mathbb{P}(X = x_i) = 1/n$, то
$$\mathbb{E}[X] = \sum_j x_j \mathbb{P}(X = x_j) = \frac{x_1 + \cdots + x_n}{n} =: \overline{x},$$
което е средното аритметично на $x_1, \ldots, x_n$).

Забележка 3. Математическото очакване може да се въведе и по друг начин: като отговор на въпроса „Коя константа $a \in \mathbb{R}$ приближава най-добре X в квадратичен смисъл?“. По-формално, търсим a, което да минимизира функцията f, зададена чрез
$$f(a) := \mathbb{E}\left[(X - a)^2\right] = \sum_j (x_j - a)^2 \mathbb{P}(X = x_j).$$
Ако диференцирате по a, може да се убедите, че горната функция се минимизира точно когато $a = \mathbb{E}[X]$.