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.

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

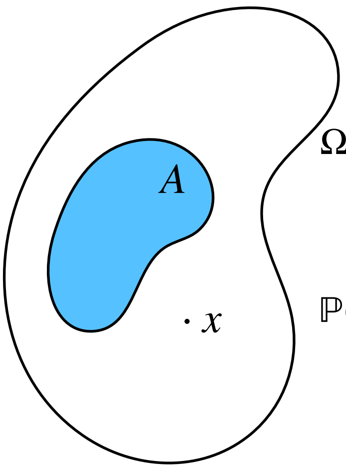
Когато имаме изброимо много елементи (събития) също имаме дискретна вероятност и множество от числа (вероятности), които вървят с всяко едно събитие $\{p_1, p_2, \ldots, p_n, \ldots\}$, $p_i \geq 0$, $\displaystyle\sum_{i=1}^{\infty} p_i = 1$. В този случай, обаче, не е възможно да зададем равномерна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



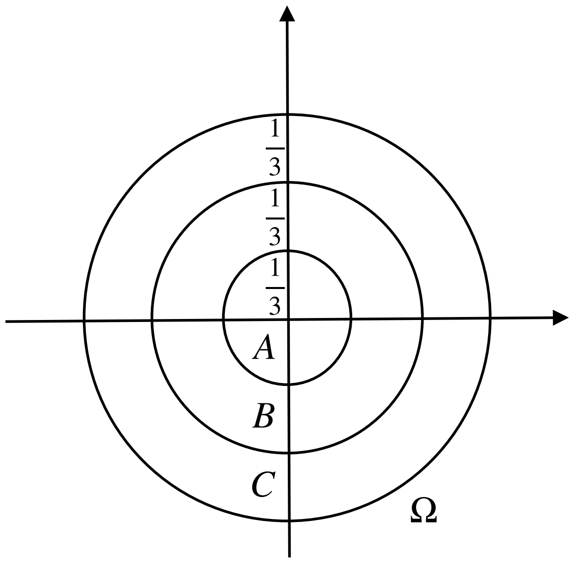
$\Omega \subseteq \mathbb{R}^2 : |\Omega| = \displaystyle\iint_{\Omega} \mathrm{d}\,x < \infty$

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

⊕



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

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

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

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

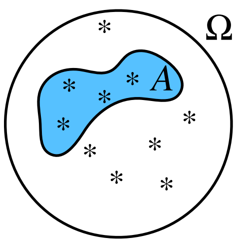
Ако се стреля хаотично (напълно аматьорски), то вероятностите да се улучат съответно множествата A, B и C са съответно:

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

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

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

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



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

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.

Дефиниция (Вероятностно пространство). Наредена тройка $(\Omega, \mathscr{A}, \mathbb{P})$, където Ω е пространство от елементарни събития; $\underbrace{\mathscr{A}}_{\substack{\text{колекция от} \\ \text{подмножества на } \Omega}} \subseteq 2^\Omega$ е σ-алгебра и $\mathbb{P} : \mathscr{A} \to [0,1]$ е вероятностна мярка (вероятното пространство).

⊕ $$\underbrace{\begin{array}{l} \Omega = \{1, 2, \ldots, N\};\ \mathscr{A} = 2^\Omega \\ \mathbb{P}(\{i\}) = \dfrac{1}{N}, \forall i \geq 1 \end{array}}$$

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.

⊕ $$\Omega = \{x^2 + y^2 \leq 1\};\ A, B, C;\ \Omega = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, C \cup A\}$$
$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1, \mathbb{P}(A) = \frac{1}{9}, \mathbb{P}(B) = \frac{3}{9}, \mathbb{P}(C) = \frac{5}{9}, \mathbb{P}(A \cup B) = \frac{4}{9},$$
$$\mathbb{P}(B \cup C) = \frac{8}{9}, \mathbb{P}(C \cup A) = \frac{6}{9}.$$

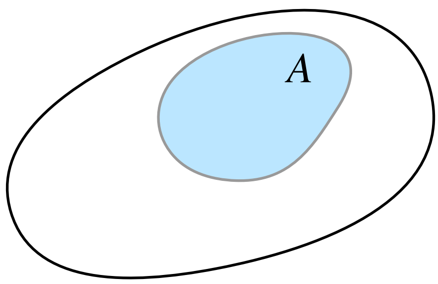
⊕ $$\Omega = \{x^2 + y^2 \leq 1\};\quad \underbrace{\mathbf{B}(\Omega) = \mathscr{A}}_{\substack{\text{бореловата} \\ \text{сигма алгебра}}} \ :\ \mathbb{P}(A) = \frac{|A|}{|\Omega|};\qquad \forall A \in \Omega.$$

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.

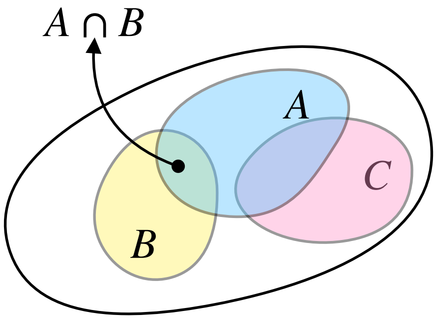
Изкуственият интелект, бейсовата статистика и много други имат зад себе си условни вероятности освен всичко останало, което включват.

$A \in \mathscr{A}$ 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" във новото вероятностно пространство $(A, \mathscr{A} \cap A, \mathbb{P}_A)$.

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

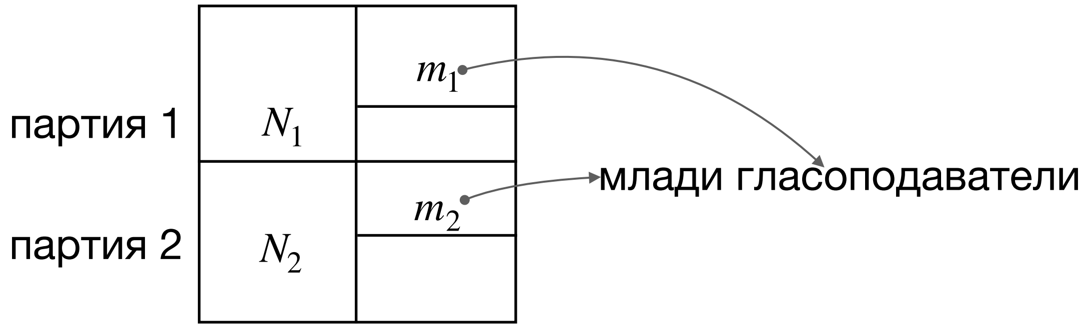
Най-доброто число, което бихме задали за вероятност на събитието B, при положение, че знаем (че се е случило) A е $\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$, както интуитивно, така и чисто в математически смисъл.

$\oplus$ „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).

$\oplus$ 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

Дефиниция (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 $B\big)$.

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

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

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

Теорема. 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“

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

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

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

$\overline{A_i}$ 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.

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

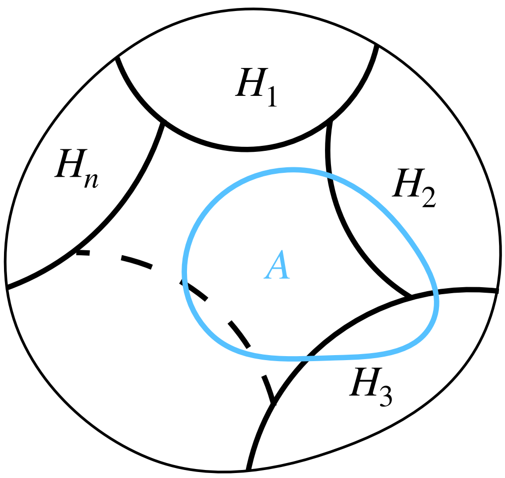
$$= 1 - \left(\frac{\binom{49}{6} - 1}{\binom{49}{6}}\right)^{10\,000} = 1 - \left(1 - \frac{1}{\binom{49}{6}}\right)^{10\,000} \approx 1 - \left(1 - 10\,000 \times \frac{1}{\binom{49}{6}}\right) \approx \frac{1}{1400}$$

Формула за пълната вероятност

Дефиниция (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$\left. \right)$.

Теорема (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 $\mathbb{P}(A \,|\, H_i)\mathbb{P}(H_i) = 0$.

Доказателство. $A = A \cap \Omega = A \cap \dot{\bigcup}_{i=1}^{n} H_i = \dot{\bigcup}_{i=1}^{n} A \cap H_i$.



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.

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

Доказателство. От една страна имаме, че $\mathbb{P}(A \cap H_k) = \mathbb{P}(A \,|\, H_k)\mathbb{P}(H_k)$, но от друга страна $\mathbb{P}(H_k \,|\, A)\mathbb{P}(A) = \mathbb{P}(H_k \cap A)$

$$\Rightarrow \mathbb{P}(H_k \,|\, A) = \frac{\mathbb{P}(H_k)\mathbb{P}(A \,|\, H_k)}{\underbrace{\mathbb{P}(A)}_{\substack{\text{формула за} \\ \text{пълната вероятност}}}}.$$ 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?

Решение.

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 изполвахме, 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).

$\oplus$ Разполагате с два пощенски плика A и B, в които има съответно положителните суми a и b. Нямате никаква априорна информация за сумите, освен това, че са положителни, а човекът, който ги е поставил в пликовете знае, че $a < b$. Избирате случайно (равновероятно) един от двата плика. Виждате сумата x в плика, който сте избрали ($x = a$ или $x = b$), но нямате никаква информация за това дали останалата сума е по-голяма или по-малка. Човекът, който е сложил сумите в пликовете знае, но Вие – не. Дава ви се шанс, ако искате, да си смените плика. При пожелана смяна, Вие със сигурност ще вземем сумата в новоизбрания плик, а ако откажете смяна ще останете със сумата от първоначално избрания плик. Да означим събитието $C = \{$печелите по-голямата сума $b\}$.

Съществува ли стратегия, за която $\mathbb{P}(C) > 50\,\%$? Тоест, има ли стратегия, която ви позволява да спечелите по-голямата от двете суми, с вероятност СТРОГО по-голяма от $50\,\%$? Обосновете отговора си.

Решение.

Контраинтуитивно, такава стратегия съществува!

Именуваме следните събития: $A = \{$вижда a в $I^{\text{-ви}}$ плик$\}$,
$B = \{$вижда b във $II^{-\text{ри}}$ плик$\}$ и $C = \{$прави се смяна на пликовете$\}$.

Знаем, че $\mathbb{P}(A) = \mathbb{P}(B) = \dfrac{1}{2}$ (равно вероятно е да изберем който и да е от двата плика). Нека още $C = \{$прави се смяна на пликовете$\}$.

$$\mathbb{P}(b) = \mathbb{P}\underbrace{(A \cap C)}_{\substack{\text{избира се}\\ \text{плик A и}\\ \text{се прави}\\ \text{смяна}}} + \mathbb{P}\underbrace{(B \cap \overline{C})}_{\substack{\text{избира се}\\ \text{плик B и}\\ \text{НЕ се прави}\\ \text{смяна}}} =$$

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

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

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

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

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

Тоест 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,
$J = \mathbb{P}(C|A) - \mathbb{P}(C|B)$. Търсим стратегия, при която $J > 0$.

$$
\begin{aligned}
J &= \mathbb{P}(C|A) - \mathbb{P}(C|B) \\
&= \frac{\mathbb{P}(C \cap A)}{\mathbb{P}(A)} - \frac{\mathbb{P}(C \cap B)}{\mathbb{P}(B)} = \\
&= \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(A)} \times \frac{\mathbb{P}(C)}{\mathbb{P}(C)} - \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)} \times \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = \\
&= \mathbb{P}(C) \times \left(\frac{1}{\mathbb{P}(A)} \times \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} - \frac{1}{\mathbb{P}(B)} \times \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} \right) = \\
&= \mathbb{P}(C) \times \left(\frac{1}{2} \times \mathbb{P}(A|C) - \frac{1}{2} \times \mathbb{P}(B|C) \right) = \\
&= \frac{\mathbb{P}(C)}{2} \times \left(\mathbb{P}(A|C) - \mathbb{P}(B|C) \right).
\end{aligned}
$$

1 сл. Ако никога не сменяме: $J = 0$, т.к. $\mathbb{P}(C) = 0$ и $\mathbb{P}(A|C) = 0 = \mathbb{P}(B|C)$.
2 сл. Ако винаги сменяме: $J = 0$, т.к. $\mathbb{P}(A|C) = \mathbb{P}(A) = 1$ и аналогично $\mathbb{P}(B|C) = \mathbb{P}(B) = 1$.
3 сл. Ако сменяме на всяко второ теглене: $J = 0$, тъй като $\mathbb{P}(A|C) = \frac{1}{2} \times \mathbb{P}(A) = \frac{1}{4}$ и аналогично и за B.
4 сл. Ако сменяме на всяко трето теглене: $J = 0$, тъй като $\mathbb{P}(A|C) = \frac{1}{3} \times \mathbb{P}(A) = \frac{1}{6}$ и аналогично и за B.
...

Изглежда, че каквото и да направим, винаги ще е равновероятно да останем с която и да е от двете суми.

Сега да се върнем на $J = \mathbb{P}(C|A) - \mathbb{P}(C|B)$. Това равенство в този вид е много по-мощно от равенството до което достигаме в синьото уравнение. Но благодарение на равенството от синьото уравнение може да пресметнем J при някакви стратегии от вида „сменяме плика на всяко k-то теглене“.

Какво е всъщност $\mathbb{P}(C|A)$? Това е „вероятността да сменим пликовете при положение, че сме избрали плик A“. Ние не знаем дали сме ибрали плик A, но знаем каква е сумата в него! Имаме наредба на събитията. Първо виждаме каква е сумата в избрания плик и после решаваме дали го сменим или не. Нека използваме тази налична информация, за да се опитаме да формулираме желаната стратегия.

Ако си дефинираме стратегията по следния начин:
Ако виждаме числото (сумата) x, то винаги сменяме с вероятност e^{-x}.
Тоест $\mathbb{P}(C|\text{виждаме } x) = e^{-x}$. Следователно,

$$J = \mathbb{P}(C|A) - \mathbb{P}(C|B) = e^{-a} - e^{-b} > 0 \text{ и това е валидна стратегия!}$$

В случая взехме неперовото число e, но твърдението е в сила за всяко положително реално число число, може да вземем например числото 2023.

Коментар: Нетривиална задача е да се намери такова число като функция на сумата в двата плика, което да максимизира вероятността за взимане на плика с по-голямата сума в очакване (в средния случай). □