СЕМ, Л-03
(2020-10-15)

Равномерна вероятност

Равномерната вероятност е вероятностно разпределение, при което всички елементарни събития (изходи) в дадено пространство от събития имат еднаква вероятност да се случат.

За крайно вероятностно пространство с n на брой еднакво възможни изхода, вероятността за всяко отделно събитие е:

$$\mathbb{P}(A) = \frac{\text{брой на благоприятни изходи за } A}{\text{общ брой възможни изходи}}$$

Основни характеристики:
- Симетричност: Всички елементарни събития са еднакво вероятни
- Нормираност: Сумата от вероятностите на всички елементарни събития е 1
- Простота: Изчислението на вероятности се свежда до броене на благоприятни случаи

Примери:
⊕ Хвърляне на зар. При хвърляне на честен шестстранен (честен) зар:
- Пространство от събития: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Вероятност за всяко число: $p_{\omega_i} = \frac{1}{6}$.
- Вероятност за четно число: $p_{\omega_i \equiv 0 \mod 2} = \frac{3}{6} = \frac{1}{2}$.

⊕ Теглене на карта от тесте. От тесте от 52 карти (без жокери), ако теглим на случаен принцип:
- Вероятност за асо: $p_{ace} = \frac{4}{52} = \frac{1}{13}$
- Вероятност за пика: $p_{spade} = \frac{13}{52} = \frac{1}{4}$

Формално представяне:
За пространство от събития $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$. Вероятностната мярка се дефинира като: $\mathbb{P}(s_i) = \frac{1}{n}$ за всички $i = \overline{1, n} = 1, 2, \dots, n$. За всяко събитие $A \subseteq \mathcal{S}$: $\mathbb{P}(A) = \frac{|A|}{n}$, където $|A|$ е броят на елементите в A (благоприятните изходи).

$\Omega = \{\omega_1, \omega_2, \dots, \omega_k, \dots\}$
Когато имаме изброимо много елементи (събития) също имаме дискретна вероятност и множество от числа (вероятности), които вървят с всяко едно събитие $\{p_1, p_2, \dots, p_k, \dots\}$, $p_i \geq 0$, $\Sigma_{i=1}^{\infty} p_i = 1$. Но в този случай обаче, не е възможно да зададем равномерна вероятност.

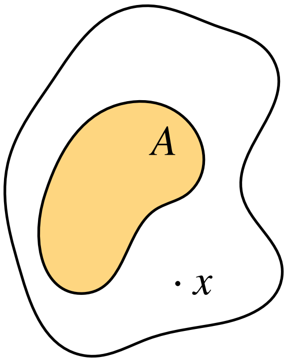
Може да я зададем по следния по-абстрактен начин: $\mathbb{P}(\{\omega_i\}) = p_i,\ i \geq 1$. $A \subseteq \Omega,\ \mathbb{P}(A) = \Sigma_{w_i \in A} p_i$.

Геометрична вероятност

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

За събитие A в геометрично пространство Ω:

$$\mathbb{P}(A) = \frac{\text{мярка на благоприятната област } A}{\text{мярка на цялата област}}$$



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

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}; \text{всяко } A \text{ е отворено}$$

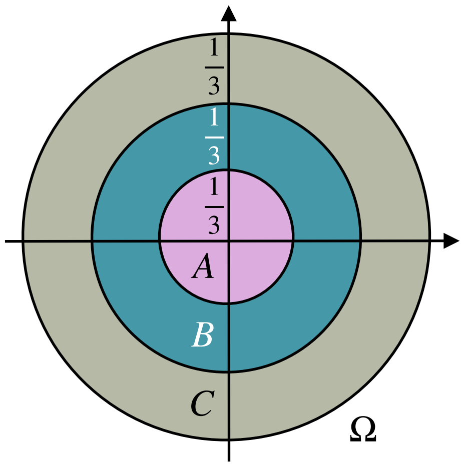
Основни характеристики:
- Безкрайни изходи: Пространството от възможни изходи е безкрайно (континуум)
- Непрекъснатост: Работим с непрекъснати величини
- Геометрични мерки: Използваме дължина за отсечки, площ за фигури, обем за тела

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: $\dfrac{\mu(A)}{\mu(\Omega)}$. Това е пример за равномерна вероятност, но равномерна само върху Ω. Това е така, защото самата вероятност зависи само от площта (мярката) на (събитието) A – не зависи от нейната локация или форма, а само от нейната площ.

$$\mathbb{P}(\{x\}) = \frac{|\{x\}|}{|\Omega|} = 0.$$

Площта на една точка е 0 и следователно вероятността на една точка е 0 (има безброй много други точки от каквато и да е площ).

⊕



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

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

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

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

Имаме мишената дартс показана на фигурата по-горе. Ако се стреля хаотично (напълно аматьорски) по нея, то вероятностите да се улучат съответно множествата A, B и C са съответно:

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

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

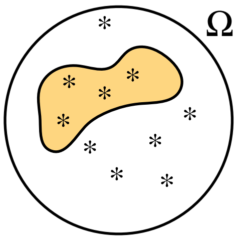
$$C : \mathbb{P}(C) = \frac{\mu(A \cup B \cup C) - \mu(A \cup B)}{\mu(\Omega)} = \frac{\pi \times 1^2 - \pi \times \left(\frac{2}{3}\right)^2}{\pi \times 1^2} = 1 - \frac{4}{9} = \frac{5}{9}.$$

Вероятностното пространство от примера се дефинира по следния наичн:

- $\Omega = \{x^2 + y^2 \le 1\}$.
- $\mathscr{A} = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, C \cup A\}$ е σ-алгебра
- $\mathbb{P}$: $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(A) = 1/9$, $\mathbb{P}(B) = 3/9$, $\mathbb{P}(C) = 5/9$, $\mathbb{P}(A \cup B) = 4/9$, $\mathbb{P}(B \cup C) = 8/9$, $\mathbb{P}(C \cup A) = 6/9$.

Монте Карло алгоритми

Нека използваме примера по-горе за да обясним идеята на Монте Карло алгоритмите. Знаем лицето на Ω, което в нашия случай е кръг и искаме да пресметнем лицето на $A \subseteq \Omega$. Започваме да стреляме аматьорски (без да се прицелваме) по мишената. След като натрупаме значително количество точки, може да направим оценка на площта на A като преброим каква част от точките попаднали в A са като процент от всички точки в мишената Ω. Това което твърдят Монте Карло алгоритмите е, че и лицата на тези сектори ще има същото отношение.
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.



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

$A(N)$ → брой на точките попаднали в A

N → брой на всички точки

Основна идея в още една аналогиа

Представете си, че искате да преброите всички бели камъчета на плаж, но имате време да проверите само малка, случайно избрана част от тях. Ако преброите камъчетата в няколко различни, случайни проби от плажа и осредните резултата, ще получите много добра оценка за общия брой. В този смисъл методът „Монте Карло“ замества тотално (и невъзможно) преброяване чрез „умна“ статистическа извадка.

Как работят на практика?

Основната структура на алгоритъм от тип Монте Карло обикновено включва следните стъпки:

1. **Дефиниране на вероятностно пространство:** Проблемът се формулира така, че отговорът да бъде свързан с вероятност от някакво събитие или средна стойност.

2. **Генериране на случайни проби:** Компютърът генерира огромен брой произволни сценарии или входни данни (напр. случайни хвърляния на зар, случайни пътища, случайни точки в пространство).

3. **Изчисляване и събиране на резултати:** За всяка „случайна проба“ се изчислява конкретният резултат или се проверява дали е изпълнено дадено условие.

4. **Статистически анализ:** Накрая се усредняват резултатите от всички симулации, за да се получи окончателната оценка.

Кога се използват?

Тези алгоритми са особено ценни, когато класическите математически методи са:

- Твърде сложни или невъзможни за аналитично решаване.
- Много бавни при реалистични размери на проблема.
- Въз основа на симулация и случаен шанс по своята същност.

Класически пример: Изчисляване на числото π.

1. Начертайте квадрат, вписан в кръг.
2. Генерирайте много случайни точки в квадрата (генериране на случайни координати).
3. Пребройте колко точки попадат вътре в кръга.
4. Съотношението Точки в кръга / Общо точки е пропорционално на съотношението на площите. Оттам може да се оцени $\pi \approx 4 \times$ (Точки в кръга / Общо точки).

Колкото повече случайни точки се генерират, толкова по-точно е приближението. Това демонстрира идеята: сложен проблем (площ на кръг) се решава чрез проста, повторяема случайна проба.

Основни типове Монте Карло симулации

- Класически Монте Карло: За оценка на интеграли и очаквани стойности (като примера с π).
- Марковски вериги Монте Карло (MCMC: Markov Chain Monte Carlo): Използва се, когато е трудно директно да се вземат проби от желано разпределение. Алгоритъмът „се скита“ случайно през възможни състояния, така че накрая посещенията да отговарят на търсеното разпределение (много често използвани в машинно обучение и байесови статистики).

Условна вероятност

Моделът, който искаме да изградим, трябва да е такъв, че да може да промени вероятностното пространство максимално добре спрямо постъпваща информация.

Изкуственият интелект, Бейсовата статистика и редица други съвременни области разчитат в своята основа на концепцията за условна вероятност. Тя представлява не само технически инструмент, а фундаментален начин на мислене за причинно-следствените връзки и несигурността в данните.

1. **Начално състояние (преди информация)**
   Работим в пълното вероятностно пространство $(\Omega, \mathscr{A}, \mathbb{P})$. Това са всички възможни изходи, които могат да се случат.
2. **Настъпване на събитието A**
   Когато събитието A (където $A \in \mathscr{A}$) настъпи, това не просто „се случва“ — то предоставя нова, ключова информация, която фундаментално променя нашия поглед върху възможностите.
3. **Механизъм на промяната**
   Настъпването на A действа като филтър или лупа, която налага:

- **Рестрикция на пространството:** Ограничава се до изходите, съвместими с A. Вече не разглеждаме цялото Ω, а само неговото подмножество A.
- **Пренасяне на вероятностната маса:** Цялата вероятностна маса (общо 1) се пренасочва към A. Събития извън A получават вероятност 0.
- **Ренормализация:** За да запазим общата вероятност равна на 1 в новото ограничено пространство, всички вероятности се преизчисляват пропорционално.

Дефиниция (Условна вероятност). Нека $(\Omega, \mathscr{A}, \mathbb{P})$ е вероятностно пространство и е такова, че $B \in \mathscr{A} : \mathbb{P}(B) > 0$. Условна вероятност на всяко събитие A при условие B се дефинира като $\mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Философски смисъл:
Тоест ние знаем, че е настъпило събитието B и тази част от A, която не е в B – вече не ни интересува. Трябва да пресметнем вероятността да се сбъдне частта на A, която попада в B $(A \cap B)$, като новото вероятностно пространство вече е B $(\Omega \mapsto B)$. Тоест ние вече се намираме в новото вероятностно пространство $(B, \mathscr{A} \cap B, \mathbb{P}_B)$.

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

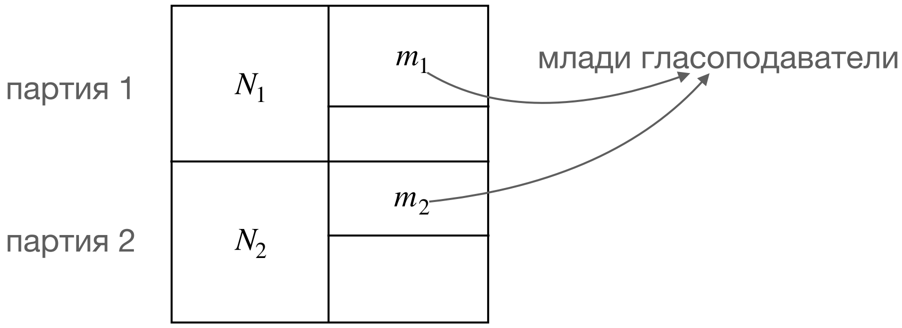
⊕ Тото „6 от 49“ (условна вероятност)

Пуснали сме фиш: $B = \{4, 8, 9, 24, 38, 49\}$. По една или друга причина ние научаваме, че са се паднали числата 4 и 8 в настоящия тираж (някога когато не е имало интернет). Тоест $A = \{\underbrace{\omega}_{\text{шесторки}} \in \Omega \mid 4 \text{ и } 8 \in \omega\}$.

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

$$= \frac{\frac{1}{\binom{49}{6}}}{\frac{\binom{47}{4}}{\binom{47}{6}}} = \frac{\binom{47}{6}}{\binom{49}{6}\binom{47}{4}} \approx \frac{1}{232{,}290}.$$ Тоест вероятността за печалба нараства значително (около 60 пъти).

⊕ Имаме две партии на някакви избори — П$_1$ и П$_2$.



Каква е вероятността гласоподавател да е гласувал за П$_1$, ако се знае, че е млад? Дефинираме две събития: $A = \{$млад$\}$ и $B = \{$гласувал за П$_1\}$.

$\mathbb{P}(B\,|\,A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\frac{m_1}{N_1+N_2}}{\frac{m_1+m_2}{N_1+N_2}} = \dfrac{m_1}{m_1+m_2}$, тоест 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 $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$. Следствие: ако $\mathbb{P}(A) > 0 \Rightarrow \mathbb{P}(B\,|\,A) = \mathbb{P}(B)$, тоест независимостта означава, че случването на събитието A не ни носи никаква информация за B.

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

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

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.

- База: За $n = 1 : \mathbb{P}(A_1) = \mathbb{P}(A_1)$ е тривиално.
- Индукционна хипотеза: Нека допуснем, че равенството от условието е изпълнено за някое $n = k$, тоест:

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right) = \mathbb{P}\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \times \ldots \times \mathbb{P}(A_2 \,|\, A_1)\mathbb{P}(A_1)$$

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

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

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

⊕ „6 от 49“ (в два поредни тиража да се паднат едни и същи числа (ненаредени)). Това е примера от Л-01. При него тиражите са общо 10,000.

$$\underbrace{\Omega = \left\{T_1, T_2, \ldots, T_{10{,}000}\right\}}_{\substack{\text{всички паднали се 10,000} \\ \text{ненаредени шесторки до сега}}}$$

Дефинираме събитията: $A_i = \{T_i = T_{i+1}\}$ за $i = 1, 2, \ldots, 9999$. Търсим:

$$\mathbb{P}\left(\bigcup_{i=1}^{9999} A_i\right).$$

Основната вероятност за всяко фиксирано i: $p = \mathbb{P}(A_i) = \dfrac{1}{\binom{49}{6}} = \dfrac{1}{13{,}983{,}816}$.

$$p \approx 7.1511 \times 10^{-8}$$

Проблеми със зависимостите: Събитията A_i не са независими, защото:

- A_i и A_{i+1} са зависими (споделят T_{i+1})
- Например: $A_1 \cap A_2$ означава $T_1 = T_2 = T_3$
- Но A_i и A_j са независими при $|i - j| > 1$

Това е поредица от събития със застъпване.

Може да използваме приближение:

$$\mathbb{P}\left(\bigcup_{i=1}^{9999} \{T_i = T_{i+1}\}\right) = 1 - (1-p)^{9999} + O(p^2) \approx 0.000714785 \approx \frac{1}{1400}.$$

Формула за пълната вероятност

Дефиниция (Пълна група от събития). Нека $(\Omega, \mathscr{A}, \mathbb{P})$ е вероятностно пространство. Наборът от събития $\{A_1, A_2, \ldots, A_n\}$ (или безкрайна изброима поредица $\{A_i\}_{i=1}^{\infty}$) се нарича пълна група от събития, ако са изпълнени следните две условия:

- Несъвместимост (дизюнктност):
  $$A_i \cap A_j = \emptyset \text{ за всички } i \neq j$$
  Тоест, никакви две събития не могат да се случат едновременно.Ю

- Изчерпателност:
  $\bigcup_{i=1}^{n} A_i = \Omega$ (за краен случай) или $\bigcup_{i=1}^{\infty} A_i = \Omega$ (за безкраен изброим случай)
  Тоест, всички заедно покриват цялото вероятностно пространство.

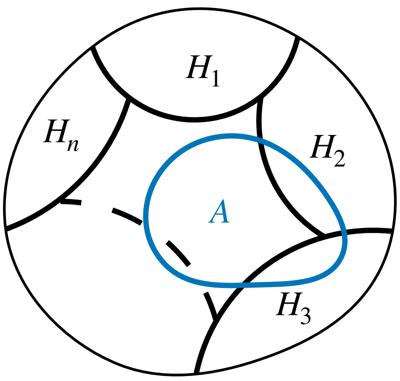
Теорема (Формула за пълната вероятност). Нека $\{H_1, H_2, \ldots, H_n\}$ е пълна група от събития с $\mathbb{P}(H_i) > 0$ за всяко i. Тогава за всяко събитие A е в сила:

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

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

Доказателство:

$A = A \cap \Omega = A \cap \bigcup_{i=1}^{n} H_i = \bigcup_{i=1}^{n} A \cap H_i$. Следователно:



$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{n} A \cap H_i\right) \underbrace{=}_{\text{непресичащи се събития}} \sum_{i=1}^{n} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid H_i)\mathbb{P}(H_i)$$

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

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

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

Формулата на Бейс (Томас Бейс (1701-1761), Кралство Великобритания) показва как актуализираме нашата априорна вероятност за хипотеза да се случи, при положение че е настъпила някаква информация A.

⊕ 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 (тоест 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 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.

⊕ Оптимизационен пример. За $p \ll 10\,\%$ 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 $p = 5\,\%$, $p = 2\,\%$.

⊕ *Решение* на задачата/парадокса с двата плика от L01.

Контраинтуитивно, такава стратегия съществува.

- Именуваме следните събития: $A = \{\text{вижда } a \text{ в 1}^{\text{-ви}} \text{ плик}\}$,
- $B = \{\text{вижда } b \text{ във 2}^{-\text{ри}} \text{ плик}\}$ и
- $C = \{\text{прави се смяна на пликовете}\}$.

Знаем, че $\mathbb{P}(A) = \mathbb{P}(B) = \dfrac{1}{2}$ (равно вероятно е да изберем който и да е плик).

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

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

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

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 $J > 0$.

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

1 сл. Ако никога не сменяме, тоест $\mathbb{P}(C) = 0$: $J = 0 \Rightarrow \mathbb{P}(C|A) = \mathbb{P}(C|B) = 0$.

2 сл. Ако винаги сменяме, тоест $\mathbb{P}(C) = 1$: $\mathbb{P}(A|C) = \dfrac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} =$
$= \mathbb{P}(A \cap C) = \mathbb{P}(A) = \dfrac{1}{2}$ и аналогично $\mathbb{P}(B|C) = \mathbb{P}(B) = \dfrac{1}{2}$. Следователно отново имаме $J = 0$.

3 сл. Ако сменяме на всяко второ теглене: $J = 0$, тъй като $\mathbb{P}(A|C) = \dfrac{1}{2} \times \mathbb{P}(A) = \dfrac{1}{4}$ и аналогично и за $\mathbb{P}(B|C) = \dfrac{1}{4}$.

3 сл. Ако сменяме на всяко трето теглене: $J = 0$, тъй като $\mathbb{P}(A|C) = \dfrac{1}{3} \times \mathbb{P}(A) = \dfrac{1}{6}$ и аналогично и за $\mathbb{P}(B|C) = \dfrac{1}{6}$.

…

За да разработим такава стратегия, трябва да мислим по-нестандартно. Тя неизбежно ще бъде по-съфистицирана от подходите, които следвахме в предишните си опити за моделиране.

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 всяка k-та игра".

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,

$J = \mathbb{P}(C \mid A) - \mathbb{P}(C \mid B) = e^{-a} - e^{-b} > 0$ 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 2020.

$\square$