СЕМ, Л-05
(2020-10-29)

Да се върнем на оптимизационния пример от Л-03 за правене на COVID-19 тест на извадка от n души едновременно.

Нека X_n е случайната величина, която отговаря на сумата заплатена за тестване на n човека, които се събират в една проба (цената на един тест е 1 лв.). Тогава:

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

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

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

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

$$\mathbb{E}[X_n] = 1 \times (1-p)^n + (n+1) \times \left(1-(1-p)^n\right) = \cancel{(1-p)^n} + n - n(1-p)^n + 1 - \cancel{(1-p)^n} =$$
$= n + 1 - n(1-p)^n$, което е трансцендентна функция

Остана само да минимизираме тази функция $\dfrac{\mathbb{E}[X_n]}{n}$, за дадено p. Например при

$$p = 0{,}05 \Rightarrow \min_{n \geq 1} \frac{\mathbb{E}[X_n]}{n} = \frac{\mathbb{E}[X_5]}{5} = 4.626216.$$

Тоест за $n = 5$ ще използваме най-малко тестове в очакване, за да определим всеки един от посетителите на летището дали е носител на вирус.

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

Свойства на математическото очакване

1. **Линейност**
   За всякакви случайни величини X, Y (дефинирани върху едно и също вероятностно пространство) и константи $a, b \in \mathbb{R}$:

   $$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

   Това важи и за произволен краен брой случайни величини:

   $$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i \mathbb{E}[X_i]$$

2. **Очакване на константа**
   Ако c е константа (разглеждана като изродена случайна величина), то:
   $$\mathbb{E}[c] = c$$

3. **Монотонност**
   Ако $X \le Y$ (с вероятност 1), то:
   $$\mathbb{E}[X] \le \mathbb{E}[Y]$$

4. **Очакване на функция на случайна величина**
   Ако $g : \mathbb{R} \to \mathbb{R}$ е измерима функция, то за дискретна X:

   $$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) \cdot p_X(x),$$

   При условие, че редът е абсолютно сходящ.

5. **Независимост**
   Ако X и Y са независими случайни величини, то:

   $$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

   Обратното не е вярно: равенството $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ не гарантира независимост.

6. **Очакване на индикаторна случайна величина**
   За събитието A с индикатор $\mathbf{1}_A$ (който е 1, ако A се случи, и 0 иначе):
   $$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A).$$

7. **Неравенство на Йенсен**
   Ако $\varphi : \mathbb{R} \to \mathbb{R}$ е изпъкнала функция и X е случайна величина с крайно очакване, то:
   $$\varphi\big(\mathbb{E}[X]\big) \le \mathbb{E}\left[\varphi(X)\right]$$
   Пример: Вземайки $\varphi(x) = x^2$, получаваме $\big(\mathbb{E}[X]\big)^2 \le \mathbb{E}[X^2]$.

Доказателство:

1. (Линейност) Нека X, Y са дискретни случайни величини с носители $\mathscr{X}$, $\mathscr{Y}$ и съвместна функция на вероятностната маса (PMF) $p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y)$.

   Нека $a, b \in \mathbb{R}$.

$$\begin{aligned}\mathbb{E}[aX+bY] &= \sum_{x\in\mathscr{X}}\sum_{y\in\mathscr{Y}}(ax+by)p_{X,Y}(x,y) \\ &= a\sum_{x\in\mathscr{X}}\sum_{y\in\mathscr{Y}}xp_{X,Y}(x,y) + b\sum_{x\in\mathscr{X}}\sum_{y\in\mathscr{Y}}yp_{X,Y}(x,y)\,.\end{aligned}$$

   Но

$$\sum_{y\in\mathscr{Y}}p_{X,Y}(x,y) = p_X(x),\ \sum_{x\in\mathscr{X}}p_{X,Y}(x,y) = p_Y(y).$$

   Следователно:

$$\begin{aligned}\mathbb{E}[aX+bY] &= a\sum_{x\in\mathscr{X}}xp_X(x) + b\sum_{y\in\mathscr{Y}}yp_Y(y) \\ &= a\mathbb{E}[X] + b\mathbb{E}[Y]\,.\end{aligned}$$

   $\square$

2. (Очакване на константа) Нека $c \in \mathbb{R}$. Разглеждаме c като случайна величина, която приема стойност c с вероятност 1. Тогава:

$$\mathbb{E}[X] = \sum_{\text{единствена стойност}} c\cdot\mathbb{P}(c=c) = c\cdot 1 = c$$

   $\square$

3. (Монотонност) Нека $X \le Y$ с вероятност 1. Това означава, че за всяка двойка (x,y) с $x > y$ имаме $p_{X,Y}(x,y) = 0$.
   За простота, ако са дискретни и дефинирани върху едно и също вероятностно пространство, можем да разгледаме $Z = Y - X \ge 0$ с вероятност 1.

   Тогава $\mathbb{E}[Z] = \sum_{z\ge 0} zp_Z(z) \ge 0$, защото всички $z \ge 0$ и вероятностите са неотрицателни.

   От линейността:

$$\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Y-X] \ge 0 \Rightarrow \mathbb{E}[X] \le \mathbb{E}[Y]$$

   $\square$

4. (Очакване на функция на случайна величина) Нека X има PMF $p_X(x)$ с носител $\mathscr{X}$ и $g : \mathbb{R} \to \mathbb{R}$.

Дефинираме $Y = g(X)$. Това е нова случайна величина с носител $\mathcal{Y} = \{g(x) : x \in \mathcal{X}\}$,
PMF на Y:

$$p_Y(y) = \sum_{x \in \mathcal{X} : g(x) = y} p_X(x)$$

Тогава:

$$\mathbb{E}[g(X)] = \mathbb{E}[Y] = \sum_{y \in \mathcal{Y}} y p_Y(y) = \sum_{y \in \mathcal{Y}} u \sum_{x : g(x) = y} p_X(x),$$

За фиксирано y вътрешната сума се взема по всички x с $g(x) = y$. Можем да сменим реда на сумирането, като сумираме директно по $x \in \mathcal{X}$:

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} \sum_{y = g(x)} y p_X(x) = \sum_{x \in \mathcal{X}} g(x) p_X(x),$$

Защото за фиксирано x имаме точно едно $y = g(x)$.

$\square$

5. (Независимост) Нека X, Y са независими дискретни случайни величини. Тогава съвместната им PMF е:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y) \text{ за всички } x, y$$

Тогава:

$$\mathbb{E}[XY] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (xy) p_{X,Y}(x, y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (xy) p_X(x) p_Y(y) =$$
$$= \left(\sum_{x \in \mathcal{X}} x p_X(x) \right) \left(\sum_{y \in \mathcal{Y}} y p_Y(y) \right) = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

$\square$

6. (Очакване на индикаторна случайна величина) Нека $A \in \mathcal{A}$ е събитие. Индикаторът $\mathbb{I}_A$ е случайна величина:

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

PMF:

$$p_{\mathbf{1}_A}(1) = \mathbb{P}(A), p_{\mathbf{1}_A}(0) = 1 - \mathbb{P}(A)$$

Тогава:

$$\mathbb{E}[\mathbf{1}_A] = 1 \cdot \mathbb{P}(A) + 0 \cdot (1 - \mathbb{P}(A)) = \mathbb{P}(A).$$

$\square$

7. (Неравенство на Йенсен) Нека φ е изпъкнала функция. Това означава, че за всяко x_0 съществува поддържаща права $l(x) = ax + b$ такава,ч е:

$$l(x_0) = \varphi(x_0) \text{ и } l(x) \leq \varphi(x) \text{ за всички } x.$$

Избираме $x_0 = \mathbb{E}[X]$. Тогава $l(x) = ax + b$ с $l(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$ и $l(x) \leq \varphi(x)$ за всички x.

От свойствата на очакването:

$$\begin{aligned}
\varphi(\mathbb{E}[X]) &= l(\mathbb{E}[X]) \\
&= \mathbb{E}[l(X)] \qquad \text{(от линейността, тъй като } l(X) = ax + b\text{)} \\
&\leq \mathbb{E}[\varphi(X)] \qquad \text{(от монотонността, тъй като } l(X) \leq \varphi(X)\text{)}
\end{aligned}$$

□

⊕ **Игра на европейска рулетка**

За да сравним стратегиите на двама играчи на рулетка, първо ще дефинираме условията на играта:

Описание на рулетката:

Играта се играе с европейска рулетка, която има 37 числа:

- 18 черни
- 18 черни
- 1 зелено (числото 0)

Стратегии на играчите:

- Играч 1: Залага винаги на черно по 1 евро на завъртане.
- Играч 2: Залага винаги на конкретно число (например 8) по 1 евро на завъртане.

Въпреки че и двамата рискуват еднаква сума на завъртане, вероятностите и възможните изходи са съществено различни. За да оценим разликата, първо изчисляваме математическото очакване (очакваната печалба или загуба) за всеки от тях.

	За играч 1 (залагащ на черно):	**За играч 2 (залагащ на конкретно число):**
Ω	○ Печалба: ако излече черно, печели 1 евро (нетна печалба). ○ Загуба: ако не излезе черно губи 1 евро.	○ Печалба: ако излезе залаганото число, печели 36 евро (нетна печалба). ○ Загуба: ако не излезе залаганото число, губи 1 евро.
$\mathbb{P}$	○ Печалба (черно): 18/37 ○ Загужа (червено или зелено): 19/37	○ Вероятност за печалба: 1/37 ○ Вероятност за загуба: 36/37
$\mathbb{E}$	$\mathbb{E}_{\text{черно}} = 1 \times \frac{18}{37} + (-1) \times \frac{19}{37} = -\frac{1}{37} \approx -0.027$	$\mathbb{E}_{\text{число}} = 35 \times \frac{1}{37} + (-1) \times \frac{36}{37} = -\frac{1}{37} \approx 0.027$

И при двете стратегии излиза, че при достатъчно продължителна игра, играчът очаква да губи средно около 2.7 цента на завъртане.

Очевидно е обаче, че двете игри са коренно различни. Тази разлика е в риск-профила. Играч 1 ще печели много по-често но много по-малко от играч 2. Изводът е, че очакването не може да различи двете игри. Необходим е друг инструмент, който да може да направи тази разлика, да я посочи и да я измери количествено.

Дисперсия

Дисперсия (в статистиката и теорията на вероятностите) е мярка за разпространението (разсейването) на стойностите на случайна величина спрямо нейната средна стойност (математическо очакване). Тя количествено показва колко стойностите се отклоняват от средната.

Дефиниция (Дисперсия). Ако сумата $\sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 \cdot p_X(x)$ е добре дефинирана,

то $\mathrm{Var}[X] = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right] = \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 \cdot p_X(x)$ е дисперсията на X.

Това е средноаритметичното на квадратите на отклоненията на всяка стойност от средната.

Да се върнем отново на примера с рулетката и двата играчи.

За консервативния играч 1 (този който залага на 18 числа едновременно):

$$\mathrm{Var}[X] = \left(-1 + \frac{1}{37}\right)^2 \frac{19}{37} + \left(1 + \frac{1}{37}\right)^2 \frac{18}{37} = \frac{1368}{1369} \approx 0.9992.$$

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

$$\mathrm{Var}[X] = \left(-1 + \frac{1}{37}\right)^2 \frac{36}{37} + \left(35 + \frac{1}{37}\right)^2 \frac{1}{37} = \frac{46,656}{1369} \approx 34.080.$$

Дефиниция (Стандартно отклонение). Ако X е случайна величина и $\mathrm{Var}[X]$ е крайно, то $\sqrt{\mathrm{Var}[X]}$ се нарича стандартно отклонение на X.

Твърдение: $\mathrm{Var}[X] \overset{(1)}{=} \mathbb{E}[(X - \mathbb{E}[X])^2] \overset{(2)}{=} \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$

Доказателство:
Нека означим $\mu = \mathbb{E}[X]$. Тогава

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[(X-\mu)^2] = \\
&= \mathbb{E}[X^2 - 2\mu X + \mu^2] = \\
&= \mathbb{E}[X^2] - 2\mu \underbrace{\mathbb{E}[X]}_{\mu} + \mu^2 = \\
&= \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Свойства на дисперсията

1. **Неотрицателност:** $\mathrm{Var}[X] \geq 0$.
   Равенство $\mathrm{Var}[X] = 0$ е тогава и само тогва, когато X е константа с вероятност 1.
   Това следва директно от дефиницията. Освен това

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2 \Leftarrow 0 \leq \mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

2. **Влияние на изместване и мащабиране:** За всякакви константни $a, b \in \mathbb{R}$:

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$$

   Доказателство:
   Нека $Y = aX + b$. Тогава $\mathbb{E}[Y] = a\mu + b$.

$$\begin{aligned} \mathrm{Var}[Y] &= \mathbb{E}\left[\left(aX + b - (a\mu + b)\right)^2\right] = \\ &= \mathbb{E}\left[(aX - a\mu)^2\right] = \\ &= a^2 \mathbb{E}\left[(X - \mu)^2\right] = a^2 \cdot \mathrm{Var}[X] \end{aligned}$$

   Важно: Константата b не влияе на дисперсията.

3. **Дисперсия на сума на независими случайни величини:**
   Ако $X_1, X_2, \ldots, X_n$ са независими, то:

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

   Доказателство (за две):

   Нека X, Y са независими.

$$\begin{aligned} \mathrm{Var}[X + Y] &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mu_X + \mu_Y)^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2 . \end{aligned}$$

   От независимостта $\mathbb{E}[XY] = \mu_X\mu_Y$. Тогава:

$$\mathrm{Var}[X + Y] = (\mathbb{E}[X^2] - \mu_X^2) + (\mathbb{E}[Y^2] - \mu_Y^2) = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Пораждаща функция

Пораждаща (моментна) функция на вероятностите (Probability Generating Function, PGF) е мощен аналитичен инструмент за работа с целочислени неотрицателни дискретни случайни величини.

В тази тема, ако не е указано друго, под X винаги ще разбираме целочислена и неотрицателна, дискретна случайна величина, тоест $X : \mathscr{A} \to \mathbb{N}^{+}_{\{0\}}$.

Пораждащата моментна функция прилича на торбичка или цифров пакет с данни, в който цялата вероятностна информация за X е запазена и компресирана в една удобна за работа математическа форма. Разгръщайки я (чрез диференциране и други операции), можем да извличаме конкретни характеристики на разпределението — като моменти или вероятности — точно когато ни потрябват.

Дефиниция (Пораждаща функция). Нека $X \in \mathbb{N}_0$ е дискретна случайна величина.

Пораждащата функция на вероятностите (PGF) на X се дефинира като:

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \cdot \mathbb{P}(X = k), \ |s| < 1.$$

$\oplus$ $\Omega = \{x_1, x_2\}$. $\mathbb{P}(X = x_1) = p$, $\mathbb{P}(X = x_2) = 1 - p$:

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

$\oplus$ $\Omega = \{x_1, x_2, \ldots, x_n\}$. $\mathbb{P}(X = x_1) = \ldots = \mathbb{P}(X = x_n) = \dfrac{1}{n}$

$$G_X(s) = \sum_{k=0}^{n} s^k \mathbb{P}(X = k) = s^0 \times \frac{1}{n} + s^1 \times \frac{1}{n} + \ldots + s^n \times \frac{1}{n} =$$
$$= \frac{1}{n}\left(s^1 + \ldots + s^n\right) \underset{n \to \infty}{\overset{|s|<1}{=}} \frac{s(1-s)^n}{n(1-s)} .$$

Основни свойства на пораждащата функция

1. **Връзка с вероятностите. Функция на масата на вероятностите**
   (PMF: Probability Mass Function)
   PGF поражда вероятностите чрез диференциране:

$$P_k = \mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}, \text{ където } G_X^{(k)}(0) = \frac{\mathrm{d}^k G_X(s)}{\mathrm{d}s^k}\Big|_{s=0}.$$

   Доказателство:

$$G_X^{(k)}(0) = \left(s^0 \cdot p_0 + s^1 \cdot p_1 + s^2 \cdot p_2 + \cdots\right)^{(k)} = k! \cdot p_k \Rightarrow \frac{G_X^{(k)}(0)}{k!} = \mathbb{P}(X = k).$$

2. **Връзка с вероятностите. Функция на разпределение**
   (CDF: Cumulative Distribution Function)

$$F_X(k) = \mathbb{P}(X \le k) = [s^k]\frac{G_X(s)}{1 - s},$$

   където $[s^k]f(s)$ означава „коефициентът пред s^k в реда на $f(s)$“.

   Доказателство:

   Тъй като $\dfrac{1}{1-s} = 1 + s + s^2 + s^3 + \ldots$ за $|s| < 1$, то и следователно

$$\frac{G_X(s)}{1-s} = (p_0 + p_1 s + p_2 s^2 + \ldots)(1 + s + s^2 + \ldots).$$

   Коефициентът пред s^k в това произведение е:

$$p_0 + p_1 + \cdots + p_k = F_X(k).$$

   Следователно:

$$F_X(k) = [s^k]\frac{G_X(s)}{1-s}.$$

3. **Математическо очакване**

$$\mathbb{E}[X] = G'_X(1)$$

   Където $G'_X(1) = \dfrac{\mathrm{d}G_X(s)}{\mathrm{d}s}\Big|_{s=1}$ (ако производната съществува).

   Доказателство:

$$G'_X(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \Rightarrow G'_X(1) = \sum_{k=1}^{\infty} k p_k = \mathbb{E}[X].$$

4. **Дисперсия**

$$\mathrm{Var}[X] = G_X''(1) + G_X'(1) - \left[G_X(1)'\right]^2.$$

Доказателство:

Знаем, че $\mathrm{Var}[X] = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2$. От $G_X''(s) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2}$, имаме:

$$G_X''(1) = \sum_{k=2}^{\infty} k(k-1)p_k = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X].$$

Следователно:

$$\mathbb{E}[X^2] = G_X''(1) + G_X'(1)$$

Тогава:

$$\mathrm{Var}[X] = [G_X''(1) + G_X'(1)] - [G_X'(1)]^2.$$

5. **Пораждаща функция на сума от независими случайни величини**

Ако $X_1, X_2, \ldots, X_n$ са независими целочислени неотрицателни случайни величини, то за $S_n = \Sigma_{i=1}^n X_i$:

$$G_{S_n}(s) = \prod_{i=1}^{n} G_{X_i}(s).$$

Доказателство: От независимостта:

$$\mathbb{E}[s^{S_n}] = \mathbb{E}[s^{X_1} \cdots s^{X_n}] = \prod_{i=1}^{n} \mathbb{E}[s^{X_i}].$$