Случайни величини с дискретни разпределения. Задачи, в които възникват. Пораждащи моментите функции, математическо очакване и дисперсия.

П.Матеев

София, 2012

Дефиниция на (дискретно и) целочислено разпределение на случайна величина. Свойства на вероятностите (неотрицателност и нормираност). За всяко от разпределенията — равномерно (дискретно), биномно, геометрично, Поасоново и хипергеометрично – да се посочи пример (задача), при който то възниква. Пресмятане на математическото очакване и дисперсията на всяко от тези разпределения. При пресмятанията може да се използва пораждаща моментите функция, но тя трябва да се определи за всяко целочислено разпределение и да се изведат основните й свойства.

Съдържание

1 Случайна величина **2**

1.1 Случаен експеримент 2

1.2 Вероятностно пространство 2

1.3 Дискретно разпределение 4

2 Математическо очакване **7**

2.1 Моменти 7

2.2 Пораждаща моментите функция 9

3 Някои дискретни разпределения **10**

3.1 Равномерно дискретно разпределение 10

3.2 Биномно разпределение 11

3.3 Геометрично разпределение 12

3.4 Поасоново разпределение 13

3.5 Хипергеометрично разпределение 14

1 Случайна величина

1.1 Случаен експеримент

Случайна величина е основното понятие в теорията на вероятностите и статистическия анализ на данни. В математиката (и в програмирането), най-просто казано, различаваме два вида величини:

- константи, които са фиксирани и не променят стойността си;
- променливи величини, които могат да приемат произволна стойност от определено множество.

Област на изменение на променливата може да са целите числа, реални числа, вектори от дадена размерност или произволно, но добре описано множество. *Случайната величина* заема място между променливите и константите в смисъл, че към всяка от стойностите, които може да приема променливата сме прикачили очакването ни или надеждата величината да приеме точно тази конкретна стойност. Очакването ни измерваме или по точно оценяваме в скала от 0 до 1 (или от 0 до 100%) и го наричаме *вероятност*.

Например, дадена логическа променлива може да приема две стойности - "$True$" и "$False$" или "1" и "0". Измерваме очакването ни величината да има стойността "1" с число между 0 и 1. Колкото по-вероятна е тази стойност, толкова по-малко вероятна е другата възможна стойност "0". Очакването ни се разпределя между тези две стойности и това разпределение се определя от конкретния случай.

Обикновено стойността на дадена величина се определя еднозначно като някоя от възможните стойности след провеждане на експеримент (или наблюдение). Почти винаги резултатът не може да бъде точно известен предварително и има, в известна степен, игра на случая. Казваме, че е проведен *случаен експеримент*.

1.2 Вероятностно пространство

Математическият модел на случаен експеримент е *вероятностно пространство* $(\Omega, \mathcal{A}, P)$. Първите два символа в тройката означават

- множеството от елементарните събития $\Omega = \{\omega_1, \omega_2, \dots\}$ - множеството от всички възможни резултати от експеримента;
- алгебрата $\mathcal{A} = \{A | A \subset \Omega\}$ от подмножества A на Ω, наричаме ги *събития*, за които може да бъде определена (оценена или измерена) вероятността.

(Булова) алгебра е множество от подмножества на дадено множество, които изпълняват условията:
(i) $\Omega \in \mathcal{A}$;
(ii) $A \in \mathcal{A} \quad \Rightarrow \quad A^c = \{\omega \in \Omega | \omega \notin A\} \in \mathcal{A}$;
(iii) $A \in \mathcal{A} \quad \& \quad B \in \mathcal{A} \quad \Rightarrow \quad A \cup B \in \mathcal{A}$.

Алгебра е $\sigma-$алгебра, когато (iii) е в сила и за обединение на изброимо много подмножества на Ω:

за всяка редица $A_k \in \mathcal{A}, \quad k = 1, \dots, n, \dots$ е в сила $\bigcup_k A_k \in \mathcal{A}$.

Двойката $(\Omega, \mathcal{A})$ наричаме *измеримо ($\sigma-$измеримо) пространство* - пространството на наблюденията/измерванията.

Вероятността на събитието $A \in \mathcal{A}$, означаваме с $P(A)$, е число в интервала $[0,1]$ (понякога се представя в проценти). Вероятността е мярка на очакването да се наблюдава конкретно събитие A. Вероятността $P : \Omega \to \mathrm{R}$ е ограничена, положителна и адитивна (σ-адитивна) функция, дефинирана върху множеството от събития $A \in \mathcal{A}$:

1. $P(\Omega) = 1 \qquad$ (ограничена)
2. $P(A) \geq 0, \quad \forall A \in \mathcal{A} \qquad$ (положителна)
3. $P(A \cup B) = P(A) + P(B) \qquad$ (адитивна)

Последното равенство е в сила за всяка двойка непресичащи се събития $A \in \mathcal{A}$, $B \in \mathcal{A}$, $A \cap B = \emptyset$. При безкраен брой на елементарните изходи е необходимо да се постави ограничение за непрекъснатост, което е еквивалентно на безкрайна изброима адитивност ($\sigma-$адитивност):

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n),$$

където $A_1, A_2, \ldots, A_n, \ldots$ е редица от непресичащи се събития:

$A_i \in \mathcal{A}, i = 1, \ldots, n, \ldots,\ A_j \cap A_k = \emptyset$ за всяка двойка $j \neq k$.

Вероятностното пространство се състои от измеримо пространство и вероятностна функция (вероятностна мярка или вероятностно разпределение). Върху едно измеримо пространство могат да бъдат определени различни вероятностни разпределения или различни модели на експеримента.

> *Случайна величина* наричаме (реална) еднозначна функция $X(\omega)$, дефинирана върху множеството от елементарни събития Ω и за която е изпълнено условието:
>
> (*) Множествата от типа $\{\omega \in \Omega | X(\omega) \leq x\}$ принадлежат на алгебрата (σ-алгебрата) от събития $\mathcal{A}$ за всяка стойност на $x \in \mathrm{R}$.

Казваме, че случайната величина е $(\Omega, \mathcal{A})-$измерима функция. От условието (*) следва, че за всяко Борелово множество B (изброимо много сечения и обединения на отворени/затворени крайни/безкрайние интервали в R, също и допълнения, т.е. елемент на Бореловата $\sigma-$алгебра $\mathcal{B}$) е в сила $\{\omega | X(\omega) \in B\} \in \mathcal{A}$.

Функция на разпределение на случайната величина X определяме като вероятността на събитието $\{\omega \in \Omega | X(\omega) \leq x\}$ (съкратено записваме като $\{X \leq x\}$ или само $X \leq x$) и означаваме с $F_X(x)$:

$$F_X(x) = P(\{X \leq x\}).$$

Индексът X на F при подразбране се изпуска.

Функцията на разпределение на всяка случайна величина е дефинирана за всяко $x \in \mathrm{R}$, като $F(x) \in [0,1]$ и притежава следните очевидни свойства

1. $F(x') \leq F(x'')$ за $x' < x''$

2. $\lim_{x\to-\infty} F(x) = 0$

3. $\lim_{x\to+\infty} F(x) = 1$

4. $\lim_{\varepsilon\to 0, \varepsilon>0} F(x+\varepsilon) = F(x)$

Последното условие е изпълнено за всяко $x \in \mathrm{R}$. Казваме, че функцията е непрекъсната отдясно и записваме съкратено $F(x+) = F(x)$.

1.3 Дискретно разпределение

> Случайна величина с дискретно разпределение или *дискретна случайна величина* наричаме случайна величина, която приема краен брой или изброимо много различни стойности с вероятност по-голяма от нула.

Ако означим тези стойности с $x_0, x_1, x_2, \ldots, x_n$ (съответно $x_0, x_1, x_2, \ldots, x_n, \ldots$ при изброимо много стойности), то събитията $H_k = \{\omega | X(\omega) = x_k\}$ образуват пълна група от събития или разбиване на множеството на елементарните събития Ω:

$$\bigcup_k H_k = \Omega, \qquad \& \quad \forall (k,j) | k \neq j : \quad H_j \cap H_k = \emptyset.$$

Означаваме съответните вероятности $P(H_k) = P(X = x_k)$ с p_k, $k = 0, 1, 2, \ldots, n(, \ldots)$, които са крайна (изброима) редица от положителни числа със сума равна на 1.

В резюме, *разпределението* на всяка дискретна случайна величина се задава с краен брой (или изброимо много) двойки числа (x_k, p_k), като $p_k > 0$ и $\sum_k p_k = 1$. Когато няма опасност от многозначност, ще означваме стойностите на случайната величина X с x и множеството им (крайно или изброимо) с $\mathbb{X}$, съответно вероятностите $P(X = x) = p_x$, $\quad \sum_x p_x = 1$, като сумирането е по всички $x \in \mathbb{X}$. Аналогично събитията H_x, за $x \in \mathbb{X}$ образуват крайно или изброимо разбиване на Ω. Множеството $\{p_x\}_{\mathbb{X}}$ определя разпределение на вероятностите върху разбиването $\{H_x\}_{\mathbb{X}}$ и върху самото $\mathbb{X}$.

Случайната величина с краен брой стойности (с положителни вероятности) се нарича *проста случайна величина.*

Целочислена случайна величина указва, че числата x_k са цели и можем да приемем $x_k = k$ и и съответните им вероятности p_k.

Константата е частн случай на случайна величина - една стойност с вероятност 1.

На всяко събитие $A \in \mathcal{A}$ съответства случайната величина *индикатор* на събитието A:

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

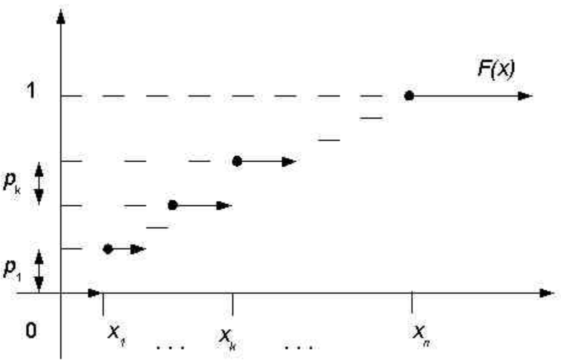
Функцията на разпределение на дискретната случайна величина X с разпределение $p_x = P(X = x) = P(H_x)$ е сума от вероятности

$$F_X(x) = P(X \leq x) = P(\bigcup_{u \leq x} H_u) = \sum u \leq x P(H_u) = \sum u \leq x p_u.$$

Функцията на разпределение на всяка дискретна случайна величина е стъпаловидна, непрекъсната отдясно функция, с краен брой или изброимо много прекъсвания. Без ограничение, можем да предположим, че стойностите са наредени в нарастващ ред $x_1 < x_2 < \cdots < x_n$ и нека са краен брой:

$$F(x) = \begin{cases} 0 & x < x_1 \\ \sum_{i=1}^{k} p_i & x \in [x_k, x_{k+1}), \quad k = 1, \ldots, n-1 \\ 1 & x \in [x_n, +\infty) \end{cases}$$

Схематична графика е представена на Фигура 1



Фигура 1: Функция на разпределение на проста случайна величина

Функцията на разпределение на константа c (изродена случайна величина, която приема единствена стойност с вероятност 1) има само едно стъпало в точката c с височина 1.

Индикаторът на събитие A с вероятност $P(A) = p$ има функция на разпределение с две стъпала: в нулата с височина $1-p$ и с височина p при $x = 1$.

Всяка случайна величина може да се представи като линйна комбинация на индикатори:

$$X = \sum_k x_k . I_{H_k}$$

където събитията $H_k = \{\omega | X(\omega) = x_k\}$ са определеното по-горе крайно (изброимо) разбиване на Ω.

Нека X и Y са две дискретни случайни величини. Разпределението на X е определено от елементите $x \in \mathbb{X}$ и вероятностите p_x и нека $\{H_x\}_{\mathbb{X}}$ е съответното му разбиване на Ω. Аналогично за разпределението на Y имаме стойности $y \in \mathbb{Y}$ с вероятности $P(Y = y) = p_y$ и съответното разбиване е $\{H_y\}_{\mathbb{Y}}$. Множеството от непразните сечения $H_x \cap H_y$ също образува разбиване.

Вероятността за съвместното събитите $\{\omega | X(\omega) = x \,\&\, Y(\omega) = y\}$ означаваме с двоен индекс:

$$p_{xy} = P(\{X = x\} \cap \{Y = y\}) = P(H_x \cap H_y).$$

Множеството $\{p_{xy}\}_{\mathbb{X}\times\mathbb{Y}}$ определят съвместно разпределение върху $\mathbb{X}\times\mathbb{Y}$. За всяко $x \in \mathbb{X}$ и за всяко $y \in \mathbb{Y}$, от адитивното свойство на вероятностите, следват равенствата:

$$\sum_y p_{xy} = p_x, \qquad \sum_x p_{xy} = p_y,$$

сумирането в първата сума е по всички $y \in \mathbb{Y}$, във втората по $x \in \mathbb{X}$. Разпределенията $\{p_x\}_{\mathbb{X}}$ и $\{p_y\}_{\mathbb{Y}}$ наричаме *маргинални* разпределения (на съвместното разпределение). Едни и същи маргинални рзпределения могат да съответстват на различни съвместни разпределения.

Условно разпределение на случайната величина Y при условие събитието $X = x$, определяме като условните вероятности

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{P(H_x \cap H_y}{P(H_x)} = \frac{p_{xy}}{p_x}.$$

Естествено е изискването събитието $X = x$ да е с положителна вероятност и за такива x е определено означението $p_{y|x} = P(Y = y|X = x)$.

Формулата за пълната вероятност ни дава връзка между съвместно, маргинално и условно разпределения:

$$p_y = \sum_x p_{y|x}.p_x,$$

като сумирането е по стойностите на x, за които $p_x > 0$.

Независимост на случайни величини. Случайните величини X и Y наричаме независими, ако за всяка двойка (x, y) събитията H_x и H_y са независими. В този случай условното разпределение съвпада с безусловното (маргинално) разпределение $p_y = p_{y|x}$ $(p_x = p_{x|y})$, съвместното е произведение на маргиналните $p_{xy} = p_x.p_y$

Означаваме $X \perp\!\!\!\perp Y$.

В общия случай независимоста на две случайни величини се изразява със съответните функции на разпределение:

$$F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\})P(\{Y \leq y\}) = F_X(x).F_Y(y).$$

Накрая да отбележим, че сума на две случайни величини е случайна величина. Наистина, нека означим $Z = X+Y$. Величината Z приема стойност $z = x + y$, когато $\omega \in \bigcup_{x+y=z}(H_x \cap H_y)$ - обединението е по онези сечения на H_x и H_y, чиято сума е равна на z.

Множеството $\bigcup_{x+y=z}(H_x \cap H_y)$ е подмножеството на Ω, определя се от величината z и естествено означаваме с H_z. За всяка стойност на $z \in R$ множеството H_z принадлежи на алгебрата $\mathcal{A}$, тъй като е обединение от сечения на елементи от $\mathcal{A}$:

$$\forall z \in R \quad H_z \in \mathcal{A}$$

или величината Z е случайна величина ($Z : \Omega \to R$ е измерима функция).

2 Математическо очакване

2.1 Моменти

Математичско очакване на дискретна случайна величина X е претеглената сума от стойностите ѝ с тегла съответните им вероятности (E като *expectation*):

$$\mathbb{E}\,X = \sum_k x_k.p_k \qquad \text{или по-просто} \qquad \mathbb{E}\,X = \sum_x x.p_x.$$

Когато сумата е безкраен ред казваме, че очакването съществува, ако редът е абсолютно сходящ: $\mathbb{E}\,|X| < +\infty$.

Очевидно, ако случайна величина е константа, т.е. $X = c$ с вероятност 1, то математическото ѝ очакване е $\mathbb{E}\,X = 1.c = c$.

Математичското очакване определя параметър, характеризиращ положението (локацията) на разпределението и затова се нарича още *средна стойност*. Наистина, нека изместим всички стойности на случайната величина X на едно и също растояние c или с други думи определим случайна величина $Y = X + c$ със стойности $y_k = x_k + c$. Математическото очакване на Y или $X + c$ е

$$\mathbb{E}\,(X + c) = \sum_k (x_k + c).p_k = \sum_k x_k.p_k + c.\sum_k p_k = \mathbb{E}\,X + c.$$

Аналогично се получава $\mathbb{E}\,(c.X) = c.\mathbb{E}\,X$.

Математическото очакване на сума от две случайне величини X и Y, на случайната величина $Z = X+Y$ е сумата от очакванията $\mathbb{E}\,Z = \mathbb{E}\,(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y$. Наистина

$$\begin{aligned}
\mathbb{E}\,Z &= \mathbb{E}\,(X + Y) = \sum_z z.P(\bigcup_{x+y=z} (H_x \cap H_y) \\
&= \sum_z z. \sum_{x,y|x+y=z} p_{xy} = \sum_z \sum_{x+y=z} (x + y)p_{xy} = \\
&= \sum_{x,y} (x + y)p_{xy} = \sum_x x.p_x + \sum_y y.p_y = \\
&= \mathbb{E}\,X + \mathbb{E}\,Y.
\end{aligned}$$

Ако двете случайни величини X и Y са независими, т.е. $p_{xy} = p_x.p_y$, то очакването на произведение на две случайни величини е произведение от очакванията им:

$$\mathbb{E}\,(X.Y) = \sum_{x,y} x.y.p_{xy} = \sum_{x,y} x.y.p_x.p_y = \sum_x x.p_x.\sum_y y.p_y = (\mathbb{E}\,X).(\mathbb{E}\,Y).$$

Тъй като различни съвместни разпределения могат да имат едни и същи маргинални разпределения, то разликата $\mathbb{E}\,(X.Y) - (\mathbb{E}\,X).(\mathbb{E}\,Y)$ може да приема различни стойности, включително и нула.

Наистина, дори разликата да е равна на нула, не следва независимост. Например, нека случайната величина X приема две различни стойности -1, 0 и $+1$ и съответно с вероятности $1/4$, $1/2$ и $1/4$. Очевидно $\mathbb{E}\,X = 0$. Случайната величина $Y = X^2$ има разпределение върху 0 и 1 и двете стойности са с вероятност $1/2$. Очевидно X и Y са зависими, тъй като при условие $X = x$ при всяка от трите стойности на x, условното разпределение е различно. Въпреки това очакването на произведението им $\mathbb{E}\,(X.Y) = \mathbb{E}\,(X^3) = \mathbb{E}\,X = 0$ съвпада с произведението на очакванията $\mathbb{E}\,X.\mathbb{E}\,Y = 0 \times 1 = 0$.

Вариация (или дисперсия) на дискретна случайна величина X наричаме математическото очаквне на квадратите от отклоненията на случайната величина от математическото ѝ очакване (означаваме $\mathbb{V}ar$ като *variance*):

$$\mathbb{V}ar\,(X) = \mathbb{E}\,(X - \mathbb{E}\,X)^2.$$

Нарича се още *втори централен момент.* Понякога е удобно представянето

$$\begin{aligned}\mathbb{V}ar\,(X) &= \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \mathbb{E}\,(X^2 - 2.X.\mathbb{E}\,X + (\mathbb{E}\,X)^2) = \\ &= \mathbb{E}\,(X^2) - 2.\mathbb{E}\,X.\mathbb{E}\,X + (\mathbb{E}\,X)^2 = \\ &= \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2,\end{aligned}$$

вариацията се представя като разлика на втори (нецентрален) момент минус квадрата на първия момент.

Очевидно, ако случайна величина е константа, т.е. $X = c$ с вероятност 1, вариацията ѝ е нула: $\mathbb{V}ar\,X = \mathbb{E}\,(X - \mathbb{E}\,X)^2 = 1.(c - c)^2 = 0$. Обратното също е в сила: ако вариацията на случайна величина X е равна на нула, следва че $x = \mathbb{E}\,X$ е единствената стойност с положителна вероятност и вероятността ѝ е равна на 1, т.е. е константа.

Вариацията зависи само от мащаба или скалата на измерване и не зависи от транслация. Наистина

$$\begin{aligned}\mathbb{V}ar\,(X + c) &= \mathbb{E}\,((X + c) - \mathbb{E}\,(X + c))^2 = \mathbb{E}\,(X + c - \mathbb{E}\,X - c))^2 = \\ &= \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \\ &= \mathbb{V}ar\,X.\end{aligned}$$

При умножение с константа c е в сила

$$\begin{aligned}\mathbb{V}ar\,(c.X) &= \mathbb{E}\,[(c.X) - \mathbb{E}\,(c.X)]^2 = \mathbb{E}\,[c.(X - \mathbb{E}\,X)]^2 = \mathbb{E}\,[c^2.(X - \mathbb{E}\,X)^2] = \\ &= c^2.\mathbb{E}\,[c^2.(X - \mathbb{E}\,X)^2] = \\ &= c^2.\mathbb{V}ar\,X\end{aligned}$$

или вариацията зависи линейно от квадрата на множителя.

Мярката за мащаба или разсейването около средната стойност е квадратният корен от вариацията на случайната величина. Нарича се *стандартно отклонение* и е прието е да се означава със σ_X, т.е. $\sigma_X^2 = \mathbb{V}ar\,X$.

Ковариацията, съвместната вариация, на двете случайни величини X и Y е математическото очакване на произведението от отклоненията от очакванията им $\mathrm{cov}(X, Y) = \mathbb{E}\,(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y)$.

Ковариацията е равна на нула, когато двете случайни величини са независими. Наистина, след разкриване на скобите

$$\begin{aligned}\mathrm{cov}(X, Y) &= \mathbb{E}\,(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y) = \mathbb{E}\,(X.Y - X.\mathbb{E}\,Y - Y.\mathbb{E}\,X + \mathbb{E}\,X.\mathbb{E}\,Y) = \\ &= \mathbb{E}\,(X.Y) - \mathbb{E}\,X.\mathbb{E}\,Y.\end{aligned}$$

Обратното не е верно. Както видяхме по-горе е възможно $\mathbb{E}\,(X.Y) = \mathbb{E}\,X.\mathbb{E}\,Y$ и двете случайни величини са даже функционално зависими.

Квадратът на ковариацията е ограничен от произведението на вариациите $[\mathrm{cov}(X,Y)]^2 \le \sigma_X^2.\sigma_Y^2$. Наистина, неравенството

$$\mathbb{E}\ [(X - \mathbb{E}\,X) + t.(Y - \mathbb{E}\,Y)]^2 \ge 0$$

е в сила за всяко реално число t. Полиномът спрямо t отляво е от втора степен. От неравенството следва, че полиномът не може да има два различни корени, може да има най-много един (двоен) реален корен. Условието за това е дискриминантата на квадратното уравнение

$$\mathbb{E}\,[(X - \mathbb{E}\,X) + t.(Y - \mathbb{E}\,Y)]^2 = \sigma_X^2 + 2t.\mathrm{cov}(X,Y) + t^2.\sigma_Y^2 = 0,$$

да е отрицателна или най-много нула: $D = [2.\mathrm{cov}(X,Y)]^2 - 4.\sigma_X^2.\sigma_Y^2 \le 0$, което е еквивалентно на $[\mathrm{cov}(X,Y)]^2 \le \sigma_X^2.\sigma_Y^2$. Равенство се достига, когато X е линейна функция на Y, т.е. когато $(X - \mathbb{E}\,X) + t.(Y - \mathbb{E}\,Y = 0)$.

Коефициент на корелация на две случайни величини X и Y наричаме нормираната им с произведението на стандартните им отклонения ковариация. Означаваме с $\rho(X,Y)$:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\sigma_X^2.\sigma_Y^2}}.$$

Сумата на две случайни величини не винаги е равна на сумата от дисперсиите на събиремите:

$$\begin{aligned}
\mathbb{V}ar\,(X+Y) &= \mathbb{E}\,[(X+Y) - \mathbb{E}\,(X+Y)]^2 = \mathbb{E}\,[(X - \mathbb{E}\,X) + (Y - \mathbb{E}\,Y)]^2 = \\
&= \mathbb{E}\,[(X - \mathbb{E}\,X)^2 + 2(X - \mathbb{E}\,X).(Y - \mathbb{E}\,Y) + (Y - \mathbb{E}\,Y)^2] = \\
&= \mathbb{V}ar\,X + 2.\mathbb{E}\,(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y) + \mathbb{V}ar\,Y = \\
&= \mathbb{V}ar\,X + \mathbb{V}ar\,Y + 2.\overbrace{\mathrm{cov}(X,Y)}^{=}.
\end{aligned}$$

С други думи $\mathbb{V}ar\,(X+Y) = \mathbb{V}ar\,X + \mathbb{V}ar\,Y$ само когато $\mathrm{cov}(X,Y) = 0$ или ковариацията, съответно също и корелацията, е равна на нула. Наричаме ги *некорелирани*. Две независими случайни величини са винаги некорелирани. Както видяхме в примера по-горе, некорелираните са случайни величини X и $Y = X^2$ са зависими.

2.2 Пораждаща моментите функция

Пораждаща моментите функция или функция $M_X(t)$, пораждаща моментите на случайната величина X е математическото очакване на случайната величина $\mathbf{e}^{t.X}$:

$$M_X(t) = \mathbb{E}\,\mathbf{e}^{t.X}.$$

Забележителното на тази функция е, че винаги при $t = 0$ е равна на 1: $M_X(0) = 1$. Ако освен това съществува производна от ред k в околност на нулата - за $|t| < h$, при някакво $h > 0$, съществуват и всички производни от ред $m < k$. Производната от ред k на функцията $M_X(t)$ имат следния вид:

$$\frac{d^k}{dt^k} M_X(t) = \frac{d^k}{dt^k}\mathbb{E}\,\mathbf{e}^{t.X} = \mathbb{E}\,\frac{d^k}{dt^k}\mathbf{e}^{t.X} = \mathbb{E}\,X^k\mathbf{e}^{t.X}.$$

Стойността ѝ при $t = 0$ съвпада с математическото очакване на X^k:

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}\,(X^k \mathbf{e}^{0.X}) = \mathbb{E}\,(X^k).$$

Когато разпределението на X е такова, че функцията пораждаща моментите му има първи две производни в околност на нулата, математическо очакване и вариация се определят от стойностите на производните при $t = 0$:

$$\begin{aligned}
\mathbb{E}\,X &= M'(0), \\
\mathbb{V}ar\,X &= \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \mathbb{E}\,(X^2 - 2.X.\mathbb{E}\,X + (\mathbb{E}\,X)^2) = \\
&= \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = M''(0) - (M'(0))^2.
\end{aligned}$$

3 Някои дискретни разпределения

3.1 Равномерно дискретно разпределение

Случайна величина X има дискретно равномерно разпределение, когато е дискретна, крайна (проста) $\mathbb{X} = \{x_1, \ldots, x_n\}$ и с равни вероятности, константа, за всяка от стойностите ѝ: $P(X = x_k) = c$. От условието $\sum_x p_x = 1$ се определя стойността на константата $c = \frac{1}{n}$.

Математическото очакване на X е средната аритметична от стойностите в $\mathbb{X}$

$$\mathbb{E}\,X = \sum_{k=1}^{n} x_k.\frac{1}{n} = \frac{x_1 + \cdots + x_n}{n} = \bar{x}$$

.

Вариацията на X е средното квадратично отклонение (средната аритметична на квадратите на отклоненията) от средната аритметична:

$$\mathbb{V}ar\,X = \sum_{k=1}^{n} (x_k - \bar{x})^2.\frac{1}{n} = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}.$$

Примери.

- Случайна величина X с равномерно разпределение върху целите числа $1, 2, \ldots, n$ има математическо очакване

  $$\mathbb{E}\,X = \sum_{k=1}^{n} \frac{1}{n}.k = \frac{1}{n}.\sum_{k=1}^{n} k = \frac{1}{n}.\frac{n.(n+1)}{2} = \frac{n+1}{2}$$

  и дисрелсия

  $$\begin{aligned}
  \mathbb{V}ar\,X &= \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = \frac{1}{n}.\sum_{k=1}^{n} k^2 - \left(\frac{n+1}{2}\right)^2 = \\
  &= \frac{1}{n}.\frac{(2n+1).n.(n+1)}{6} - \frac{(n+1)^2}{4} = \\
  &= \frac{n+1}{12}.\left[2.(2n+1) - 3.(n+1)\right] = \frac{(n+1).(n-1)}{12} = \frac{n^2-1}{12}.
  \end{aligned}$$

- Равномерното разпределение е модел при честни хазартни игри като зарове, тото, за които се предполага еднакви вероятности за всеки резултат.

- Емпирично разпределение. При n наблюдения - наблюдения на количествена характеристика на n обекта, когато няма други съображения за предпочитание на някое от наблюденията, всички са равностойни или с други думи са представителна (и хомогенна) извадка от генерална съвкупност (популация) на всяко наблюдение приписваме равни тегла по $\frac{1}{n}$. Емпиричното разпределение приписва вероятност $\frac{1}{n}$ на всяко наблюдение. То е равномерно, ако няма повторения. Ако няколко наблюдения, да кажем k, имат една и съща стойност, то естествено приписваме на такава стойност вероятност $\frac{k}{n}$ и равномерността се нарушава.

3.2 Биномно разпределение

Случайната величина X има биномно разпределение, когато стойностите ѝ са целите числа от 0 до n, $\mathbb{X} = \{0, 1, 2, \dots, n\}$, вероятностите им са

$$P(X = k) = \binom{n}{k} p^k.q^{n-k} = \frac{n!}{k!(n-k)!}, \qquad k = 0, 1, 2, \dots, n,$$

където параметърът $p \in [0, 1]$ и за удобство е положено $q = 1 - p$.

Знакът $\binom{n}{k}$ (а също и C_n^k - комбинации от n елемента k-ти клас) е известния биномен коефициент от развитието на Нютоновия бином

$$(a+b)^n = \binom{n}{0} a^0.b^n + \binom{n}{1} a^1.b^{n-1} + \dots + \binom{n}{k} a^k.b^{n-k} + \dots + \binom{n}{n} a^n.b^0.$$

Понякога означаваме $\binom{n}{k}$ с C_n^k - комбинации от n елемента k-ти клас, а с $Bi(n, p)$ класа на случайните величини с биномно разпределение $X \in Bi(n, p)$ или $X \sim Bi(n, p)$.

Математическо очакване и вариация. Използваме Нютоновия бином и получаваме за пораждащата моментите функция на X:

$$\begin{aligned} M(t) &= \mathbb{E}\left(\mathbf{e}^{t.X}\right) = \sum_{k=0}^{n} \binom{n}{k} p^k.q^{n-k}.\mathbf{e}^{t.k} = \\ &= \sum_{k=0}^{n} \binom{n}{k} (p.\mathbf{e}^t)^k.q^{n-k} = (p.\mathbf{e}^t + q)^n. \end{aligned}$$

Диференцираме два пъти

$$\begin{aligned} M'(t) &= \frac{d}{dt}(p.\mathbf{e}^t + q)^n = n.(p.\mathbf{e}^t + q)^{n-1}.p.\mathbf{e}^t, \\ M''(t) &= \frac{d}{dt}\left[n.p.(p.\mathbf{e}^t + q)^{n-1}.\mathbf{e}^t\right] = \\ &= n.p.(p.\mathbf{e}^t + q)^{n-1} + n.p.(n-1).(p.\mathbf{e}^t + q)^{n-2}.p.\mathbf{e}^t \end{aligned}$$

Заместваме с $t = 0$ и получаваме:

$$\begin{aligned}
\mathbb{E}\,X &= M'(0) = n.p.(p.\mathbf{e}^0 + q)^{n-1} = n.p.(p+q)^{n-1} = n.p, \\
\mathbb{E}\,X^2 &= M''(0) = n.p + n.p.(n-1).p = n.p + n.(n-1).p^2, \\
\mathbb{V}ar\,X &= \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = n.p + n^2.p^2 - n.p^2 - (n.p)^2 = \\
&= n.p - n.p^2 = n.p.q.
\end{aligned}$$

Пример. (Схема на Бернули) Провеждат се n опита (експеримента) независимо и при едни и същи условия. Наблюдава се дали събитието A се сбъдва или не. Алгебрата от събития във вероятностното пространство $(\Omega, \mathcal{A}, P)$ при един опит се състои от четирите множества $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$. Нека $I_A(\omega)$ е индикаторът на събитието A и нека $P(I_A = 1) = p$. Вероятностното пространство за един опит е изоморфно на породеното от индикатора, а именно $\Omega = \{1, 0\}$, алгебрата от събития е $\mathcal{A}$ е породена от разбиването $H_0 = \{0\}$, $H_1 = \{1\}$. Може да запишем, че $P(I_A = x) = p^x.(1-p)^{1-x}$ за $x \in \Omega = \{1, 0\}$.

Вероятностното пространство за n-те опита може да се конструира като декартово произведение на n-те пространства за всеки един от n-те опита $(\Omega^{\times n}, \mathcal{A}^{\times n}, P^n)$. Множеството от елементарните събития $\Omega^{\times n}$ се състои от всички вектори $(x_1, x_2, \ldots, x_n)$ като за x_i е стойността на индикатора на A в i-тия опит с възможните стойности 0 или 1. Вероятността на едно елементарно събитие е

$$P(\omega) = P(x_1, x_2, \ldots, x_n) = p^k.(1-p)^{n-k},$$

където k е броят на единиците в ω и $n-k$ - броят на нулите: $k = \sum_{i=1}^{n} x_i$.

Сумата на индикаторите или броя на успешните опити в серия от n независими еднотипни експеримента има биномно разпределение.

3.3 Геометрично разпределение

Дискретно целочислено, но не крайно, разпределение определено с $\mathbb{X} = \{0, 1, 2, \ldots\}$ и вероятности $p_x = p.(1-p)^x, \quad x \in \mathbb{X}$ наричаме геометрично с параметър $p \in [0, 1]$ и означаваме с $Ge(p)$. Случайна величина $X \sim Ge(p)$ означава, че

$$P(X = k) = p_k = p.(1-p)^k = p.q^k, \quad (p + q = 1), \quad k = 0, 1, 2, \ldots.$$

Условието $\sum_x p_x = 1$ следва от сумата на безкрайната геометрична прогресия, откъдето е и иметона разпределението:

$$\sum_{k=0}^{\infty} p.q^k = \lim_{n \to \infty} p.\frac{1 - q^{n+1}}{1 - q} = p.\frac{1}{1-q} = 1.$$

Математическо очакване и вариация. За пораждащата моментите функция на $X \sim Ge(p)$ получаваме :

$$\begin{aligned}
M(t) &= \mathbb{E}\,(\mathbf{e}^{t.X}) = \sum_{k=0}^{\infty} p.q^k.\mathbf{e}^{t.k} = \\
&= = \sum_{k=0}^{\infty} p.\left[q.\mathbf{e}\right]^{t.k} = \\
&= p.\frac{1}{1 - q.\mathbf{e}^t}.
\end{aligned}$$

Диференцираме два пъти

$$\begin{aligned}
M'(t) &= \frac{d}{dt}\left[\frac{p}{1-q.\mathbf{e}^t}\right] = \frac{p.(-1).(-q).\mathbf{e}^t}{(1-q.\mathbf{e}^t)^2} = \frac{p.q.\mathbf{e}^t}{(1-q.\mathbf{e}^t)^2},\\
M''(t) &= \frac{d}{dt}\left[\frac{p.q.\mathbf{e}^t}{(1-q.\mathbf{e}^t)^2}\right] = \frac{p.q.\mathbf{e}^t.(1-q.\mathbf{e}^t)^2 - p.q.\mathbf{e}^t.2.(1-q.\mathbf{e}^t).(-q).\mathbf{e}^t}{(1-q.\mathbf{e}^t)^4} =\\
&= \frac{p.q.\mathbf{e}^t + p.q^2.\mathbf{e}^{2.t}}{(1-q.\mathbf{e}^t)^3}.
\end{aligned}$$

Заместваме с $t = 0$ и получаваме:

$$\begin{aligned}
\mathbb{E}\,X &= M'(0) = \frac{p.q.\mathbf{e}^0}{(1-q.\mathbf{e}^0)^2} = \frac{p.q}{(1-q)^2} = \frac{q}{p},\\
\mathbb{E}\,X^2 &= M''(0) = \frac{p.q+p.q^2}{(1-q)^3} = \frac{q+q^2}{p^2},\\
\mathbb{V}ar\,X &= \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = \frac{q+q^2}{p^2} - \frac{q^2}{p^2} = \frac{q}{p^2}.
\end{aligned}$$

Примери. Геометричното разпределение е подходящ модел за изследване надеждност на изделия при многократна употреба (веднъж стомна за вода, дваж . . . и накрая я няма ;); класически пример е стрелба в цел до поразяване.

3.4 Поасоново разпределение

Дискретно целочислено, но не крайно, разпределение определено с $\mathbb{X} = \{0, 1, 2, \ldots\}$ и вероятности

$$p_x = \mathbf{e}^{-\lambda}.\frac{\lambda^x}{x!}, \quad x \in \mathbb{X}$$

наричаме Поасаново с параметър $\lambda > 0$ и означаваме с $Po(\lambda)$. Случайна величина $X \sim Po(\lambda)$ означава, че

$$P(X = k) = \mathbf{e}^{-\lambda}.\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Изпълнението на условието $\sum_x p_x = 1$ следва от разлагането на експоненциалната функция в безкраен ред:

$$\mathbf{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Математическо очакване и вариация. За пораждащата моментите функция на $X \sim Po(\lambda)$ получаваме :

$$\begin{aligned}
M(t) &= \mathbb{E}\,\mathbf{e}^{t.X} = \sum_{k=0}^{\infty} \mathbf{e}^{-\lambda}.\frac{\lambda^k}{k!}.\mathbf{e}^{t.k} = \mathbf{e}^{-\lambda}.\sum_{k=0}^{\infty} \frac{\mathbf{e}^{k.\ln\lambda}}{k!}.\mathbf{e}^{t.k} =\\
&= \mathbf{e}^{-\lambda}.\exp\{\mathbf{e}^{t+\ln\lambda}\} = \exp\{\lambda\mathbf{e}^t - \lambda\} =\\
&= \exp\{\lambda(\mathbf{e}^t - 1)\}.
\end{aligned}$$

Диференцираме два пъти

$$\begin{aligned}
M'(t) &= \frac{d}{dt}\exp\{\lambda(\mathbf{e}^t-1)\} = \exp\{\lambda(\mathbf{e}^t-1)\}.\lambda\mathbf{e}^t, \\
M''(t) &= \frac{d}{dt}\exp\{\lambda(\mathbf{e}^t-1)\}.\lambda\mathbf{e}^t = \\
&= \exp\{\lambda(\mathbf{e}^t-1)\}.\lambda\mathbf{e}^t + \exp\{\lambda(\mathbf{e}^t-1)\}.(\lambda\mathbf{e}^t)^2.
\end{aligned}$$

Заместваме с $t = 0$ и получаваме:

$$\begin{aligned}
\mathbb{E}\,X &= M'(0) = \exp\{\lambda(\mathbf{e}^0-1)\}.\lambda\mathbf{e}^0 = \exp\{\lambda(1-1)\}.\lambda = \lambda, \\
\mathbb{E}\,X^2 &= M''(0) = \exp\{\lambda(\mathbf{e}^0-1)\}.\lambda\mathbf{e}^0 + \exp\{\lambda(\mathbf{e}^0-1)\}.(\lambda\mathbf{e}^0)^2 = \lambda+\lambda^2, \\
\mathbb{V}ar\,X &= \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = \lambda+\lambda^2-\lambda^2 = \lambda.
\end{aligned}$$

Примери. Разпределението на Поасон е приближение на биномното разпределение за големи (растящи) стойности n и постоянно произведението $n.p$ равно (или с граница) λ.

То е класически модел на "редки" събития. Нека T е моментът на настъпване на "рядкото" събитие и да означим настъпването му в единичния интервал $[0,1]$ с $A = \{T \in [0,1]\}$. Означаваме вероятността му с $P(A) = p_1 = \lambda$. Вероятността за повторно настъпване на събитието е "пренебрежимо малка". Следователно, математическото очакване "почти съвпада" с вероятността λ.

Сега нека разделим единичния интервал на n последователни интервала от време с дължина $\frac{1}{n}$ и нека $A_i = \{T \in [(i-1)/n, i/n]\}$ е настъпване на "редко" събитие в i-тия интервал.

Допускаме сега, че вероятността на всяко събитие A_i е една и съща $p_n = \frac{\lambda}{n}$ и събитията $A_1, A_2, \ldots, A_n$ са независими. Броят на интервалите, в които редките събития са настъпили е случайна величина с биномно разпределение $Bi(n, p_n)$ с математическо очаквне $n.p_n = \lambda$.

Вероятността p_n клони към 0 при $n \to \infty$ и произведението $n.p_n = n.\frac{\lambda}{n} = \lambda$ е константа. При тези условия (Теорема на Поасон), биномните вероятности клонят към вероятностите на Поасоновото разпределение с параметър λ. Това е основание да използваме приближението:

$$\mathbf{e}^{-\lambda}\frac{\lambda^k}{k!} \approx \frac{n!}{k!.(n-k)!}p^k.(1-p)^{n-k}, \quad k = 0, 1, 2, \ldots.$$

и в практически случаи на големи n и малко p. Апроксимираме вероятностите на биномното разпределение $Bi(n, p)$ с вероятностите на Поасоново разпределение с параметър $\lambda = n.p$.

3.5 Хипергеометрично разпределение

Хипергеометричното разпределение е целочислено разпределение с вероятности

$$p_k = \frac{\binom{M}{k}\cdot\binom{N}{n-k}}{\binom{M+N}{n}} = \frac{\frac{M!}{k!.(M-k)!}\cdot\frac{N!}{(n-k)!.(N-n+k)!}}{\frac{(M+N)!}{n!.(M+N-n)!}},$$

когато всички факториели имат неотрицателен аргумент и $p_k = 0$ в останалите случаи. Хипергеометричното разпределение е крайно. Нека X е с хипергеометрично разпределение:

$P(X=k)=p_k,\ k=0,1,2,\dots,n$, а ако $k>M$ или $k<n-N$ вероятността $p_k=0$.

Да проверим, че сумата от вероятностите p_k е равна на 1. От урна с $M+N$ топки изваждаме n от тях по $\binom{M+N}{n}$ начина. Нека M са бели и N сини. От извадките точно $\binom{M}{k}.\binom{N}{n-k}$ са с k бели и $n-k$ сини. Броят на белите k може да е $0,1,\dots,n$. Следователно

$$\binom{M+N}{n}=\sum_{k=0}^{n}\binom{M}{k}.\binom{N}{n-k},$$

от което следва, че $\sum_k p_k=1$.

Математическо очакване и вариация. Ще ги определим директно като използваме първо, че $k.\binom{M}{k}=M.\binom{M-1}{k-1}$ и $\binom{M+N}{n}=\frac{M+N}{n}\binom{M+N-1}{n-1}$:

$$\mathbb{E}\,X=\sum_k k.p_k=\sum_k\frac{k.\binom{M}{k}.\binom{N}{n-k}}{\binom{M+N}{n}}=\frac{M.n}{M+N}.\underbrace{\sum_k\frac{\binom{M-1}{k-1}.\binom{N}{n-k}}{\binom{M+N-1}{n-1}}}_{=1},$$

или $\mathbb{E}\,X=\frac{M.n}{M+N}$.

Аналогично за очакването на случайната величина $X.(X-1)$ (втори факторилен момент на X) получаваме

$$\begin{aligned}\mathbb{E}\,(X.(X-1)) &= \sum_k\frac{k.(k-1).\binom{M}{k}.\binom{N}{n-k}}{\binom{M+N}{n}}=\\ &= \frac{M.(M-1).n.(n-1}{(M+N).(M+N-1)}.\underbrace{\sum_k\frac{\binom{M-2}{k-2}.\binom{N}{n-k}}{\binom{M+N-2}{n-2}}}_{=1}.\end{aligned}$$

Вариацията ще получим от равенството:

$$\begin{aligned}\mathbb{V}ar\,X &= \mathbb{E}\,(X.(X-1))+\mathbb{E}\,X-(\mathbb{E}\,X)^2=\\ &= \frac{M.(M-1).n(n-1}{(M+N)(M+N-1)}+\frac{M.n}{M+N}-\frac{(M.n)^2}{(M+N)^2},\end{aligned}$$

което след привеждане под общ знаменател и разкриване скобите в числителя получаваме

$$\mathbb{V}ar\,X=\frac{n.M.N.(M+N-n)}{(M+N)^2.(M+N-1)}.$$

Пример. В партида от $M+N$ изделия M са дефектни, а N не са. При случайна извадка без връщане с обем n, вероятността броят на дефектните изделия в извадката X има хипергеометрично разпределение

$$P(X=k)=\frac{\binom{M}{k}.\binom{N}{n-k}}{\binom{M+N}{n}}.$$

Броят на познатите числа в един фиш при едно теглене от играта “6 от 49” има хипергеометрично разпределение

$$P(X=k)=\frac{\binom{6}{k}.\binom{43}{6-k}}{\binom{49}{6}}.$$