1 Diskrétní náhodné veličiny

Definice 1.1 (Diskrétní náhodná veličina)**.** *Náhodnou veličinu X nazýváme diskrétní, pokud obor hodnot R_X je nejvýše spočetná množina, tzn. pokud existuje taková posloupnost $x_1, \ldots, x_n, \ldots$ že $X^{-1}\{x_1, \ldots, x_n, \ldots\} = \Omega$.*

Distribuční funkci diskrétní náhodné veličiny můžeme zapsat například jako

$$\mathrm{F}_X(x) = \mathrm{P}(X \leq x) = \sum_{m|x_m \leq x} \mathrm{P}(X = x_m) = \sum_{m=1}^{N,\infty} \mathrm{P}(X = x_m)\mathbb{I}_{[x_m, +\infty]}(x) \tag{1}$$

Diskrétní hustota pravděpodobnosti (frekvenční funkce) je funkce definována jako

$$f_X(x) = \begin{cases} \mathrm{P}(X = x_k) & X = x_k \\ 0 & jinak \end{cases} \tag{2}$$

Poznámka 1.2. $\mathrm{P}\left(X = x_k\right)$ *můžeme označit jako p_k. Přitom platí*

$$\sum_k p_k = 1$$

Definice 1.3 (Diracovo rozdělení)**.** *Buď X náhodná veličina, a nechť existuje $c \in \mathbb{R}$ takové, že*

$$\begin{gathered} \mathrm{P}\left(X = c\right) = 1 \\ \mathrm{P}\left(X \neq c\right) = 0 \end{gathered} \tag{3}$$

Značíme $X \sim \delta_c$.

Definice 1.4 (Alternativní/Bernoulliho rozdělení s parametrem p)**.** *Nechť $p \in [0, 1]$. Uvažujme náhodnou veličinu X, která může nabývat pouze dvou hodnot, například $0, 1$, a dále nechť platí*

$$\begin{gathered} \mathrm{P}(X = 1) = p \\ \mathrm{P}(X = 0) = 1 - p \end{gathered} \tag{4}$$

Značíme $X \sim A(p)$.

Definice 1.5 (Binomické rozdělení)**.** *Opakujme n-krát experiment s náhodnou veličinou X, která má alternativní rozdělení, přičemž uvažujeme $\mathrm{P}(A) = p$ (pravděpodobnost úspěchu) a tedy $\mathrm{P}\left(A^{\mathbb{C}}\right) = 1-p$ (pravděpodobnost neúspěchu). Počet příznivých jevů při n opakováních je potom*

$$X = \sum_{j=1}^{n} X_j \tag{5}$$

Značíme $X \sim Bi(n, p)$.

Pro pravděpodobnost poté platí

$$\mathrm{P}_n(X = k) = \mathrm{P}_n\left(\sum_{\pi(\hat{n})} \{X_{i_1} = 1, \ldots, X_{i_k} = 1, X_{i_{k+1}} = 0, \ldots, X_{i_n} = 0\}\right) =$$

$$= \sum_{\pi(\hat{n})} \mathrm{P}\left(X_{i_1} = 1, \ldots, X_{i_k} = 1, X_{i_{k+1}} = 0, \ldots, X_{i_n} = 0\right) =$$

$$= \binom{n}{k} \prod_{i=1}^{k} \mathrm{P}(x_i = 1) \prod_{i=k+1}^{n} \mathrm{P}(x_i = 0) = \binom{n}{k} p^k (1-p)^{n-k}$$

Z binomické věty také vyplývá, že

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

Příklad 1.6. *Uvažujme šachovou partii, ve které jsou dva stejně silní soupeři (pravděpodobnost výhry i prohry je $\frac{1}{2}$). Rozhodněte, zda je pravděpodobnější*

1. *vyhrát 3 partie ze 4, nebo 5 partií z 8.*
2. *vyhrát alespoň 3 partie ze 4, nebo alespoň 5 partií z 8.*

1.

$$\mathrm{P}\left(X = 3\right) = \mathrm{P}_4(3) = \binom{4}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = \frac{1}{4}$$

$$\mathrm{P}\left(X = 5\right) = \mathrm{P}_8(5) = \binom{8}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right) = \frac{7}{32}$$

2.

$$\mathrm{P}\left(X = 3 \vee X = 4\right) = \mathrm{P}\left(X = 3\right) + \mathrm{P}\left(X = 4\right) \mathrm{P}_4(3) + \mathrm{P}_4(4) =$$

$$\frac{1}{4} + \binom{4}{4} \left(\frac{1}{2}\right)^4 = \frac{5}{16}$$

$$\mathrm{P}\left(X = 5 \vee X = 7 \vee X = 8\right) = \sum_{k=5}^{8} \mathrm{P}_8\left(k\right) =$$

$$= \sum_{k=5}^{8} \binom{8}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{8-k} = \frac{93}{256}$$

Definice 1.7 (Geometrické/Pascalovo rozdělení)**.** *Uvažujme nekonečnou posloupnost pokusů s veličinou s alternativním rozdělením*

$$\mathrm{P}(A) = p$$

$$\mathrm{P}\left(A^{\mathbb{C}}\right) = 1 - p$$

a buď X počet pokusů před prvním výskytem jevu A. Platí, že

$$\mathrm{P}\left(X = k\right) = p(1-p)^k \tag{6}$$

Platí

$$\sum_{k=0}^{\infty} p(1-p)^k = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1 - (1-p)} = 1$$

Definice 1.8 (Negativně binomické rozdělení)**.** *Opakujme jev nekonečně krát, a nechť náhodná veličina Y značí počet neúspěchů před m-tým úspěchem, přičemž $\mathrm{P}(A) = p$. Potom*

$$\mathrm{P}\left(Y = k\right) = \binom{k+m-1}{k} p^m (1-p)^k$$

Definice 1.9 (Hypergeometrické rozdělení)**.** *Uvažujme zásobníkový model, ve kterém je r červených a $N-r$ bílých kuliček. Opakujme n-krát tah bez vracení, a jako náhodnou veličinu X uvažujme počet červených kuliček v n-tici.*

$$\mathrm{P}\left(X = x\right) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \tag{7}$$

Značíme $X \sim Hyp(N, r, n)$.

Příklad 1.10. *Uvažujme rybník, ve kterém je N ryb, a nechť r z nich je označeno. Chytněme n ryb, a jako náhodnou veličinu X uvažujme počet označených ryb mezi n chycenými. Jaká je statistika N?*

Chytíme n ryb, spočítáme označené a odhadneme $N = \widehat{N}$, protože víme kolik je r.

$$\frac{r}{\widehat{N}} = \frac{x}{n}$$

$$\widehat{N} = \frac{r}{x} n$$

Věta 1.11. *Pro N velká, $\frac{n}{N}$ malá platí*

$$\mathrm{H}(N, r, n) \doteq \mathrm{B}_i\left(n, \frac{r}{N}\right)$$

Definice 1.12 (Poissonovské rozdělení)**.** *Říkáme, že náhodná veličina $X : \Omega \to \mathbb{N}_0$ má poissonovské rozdělení s parametrem $\lambda > 0$, pokud*

$$\mathrm{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, \ldots \tag{8}$$

Značíme $X \sim Po(\lambda)$

Věta 1.13 (Poissonova)**.** *Uvažujme posloupnost náhodných veličin X_n s binomickým rozdělením, $X_n \sim B_i(n, p_n)$, a nechť $np_n \to \lambda$ (nebo $np_n = \lambda$), $\lambda > 0$. Potom*

$$\lim_{n\to\infty} \mathrm{P}_n(x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{9}$$

Důkaz.

$$\lim_{n\to+\infty} \mathrm{P}_n(x) = \lim_{n\to+\infty} \binom{n}{x} p_n^x (1-p_n)^{n-x} = \left\{ p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right\} =$$

$$= \lim_{n\to+\infty} \binom{n}{x} \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^x \left(1 - \frac{\lambda}{n} - o\left(\frac{1}{n}\right)\right)^{n-x} =$$

$$= \lim_{n\to+\infty} \frac{\lambda^x}{x!} \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \underbrace{\left(1 + no\left(\frac{1}{n}\right)\right)^x}_{\to 1}$$

$$\underbrace{\left(1-\frac{\lambda}{n}-o\left(\frac{1}{n}\right)\right)^{n-x}}_{\to e^{-\lambda}}=\frac{\lambda^x}{x!}e^{-\lambda}$$

□

Poznámka 1.14. *V předchozím důkazu se používají značky $o()$ a $O()$ definované:*

$$a_n = o(b_n)(n\to\infty) \iff \frac{a_n}{b_n}\to 0$$

$$a_n = O(b_n) \iff \exists k : \frac{a_n}{b_n}\le k$$

Věta 1.15 (Zákon řídkých jevů)**.** *Uvažujme jev A v čase t (se začátkem v $t_0=0$). Počet výskytu jevu do času t označme jako X_t. Nechť dále platí:*

1. *$X_{t+h}-X_t$ nezávisí na t*
2. *$\mathrm{P}\left((X_{t+h}-X_t)=1\right)=\lambda h+o(h), \lambda>0$, přičemž při $h\to 0+$ platí*
$$\frac{o(h)}{h}\to 0$$
3. *$\mathrm{P}\left(X_{t+h}-X_t>1\right)=o(h)$ při $h\to 0+$*
4. *Funkce $p_k(t)=\mathrm{P}(X_t=k)$ je diferencovatelná v t pro všechna $k\in\mathbb{N}_0$.*

Potom

$$\mathrm{P}\left(X_t=k\right)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \tag{10}$$

(λ nazýváme intenzitou řídkého jevu).

Důkaz.

$$p_o(t+h)=\mathrm{P}\left(X_t=0\right)\mathrm{P}\left(X_{t+h}-X_t=0\right)=p_0(t)\left(1-\lambda h+o(h)\right)$$

$$\frac{\mathrm{d}p_0}{\mathrm{d}t}(t)=\lim_{h\to 0+}\frac{p_0(t+h)-p_0(t)}{h}=\lim_{h\to 0+}\left(-\lambda p_0(t)+p_0(t)\frac{o(h)}{h}\right)=-\lambda p_0(t)$$

$$p_k(t+h)=\sum_{j=0}^{k}\mathrm{P}\left(X_t=j\right)\mathrm{P}\left(X_{t+h}-X_t=k-j\right)=$$

$$=\sum_{j=0}^{k-2}\mathrm{P}\left(X_t=j\right)\underbrace{\mathrm{P}\left(X_{t+h}-X_t=k-j\right)}_{o(h)}+\underbrace{\mathrm{P}(X_{t+h}-X_t=1)}_{\lambda h+o(h)}p_{k-1}(t)+$$

$$+\underbrace{\mathrm{P}(X_{t+h}-X_t=0)}_{1-\lambda h+o(h)}p_k(t)=\sum_{j=0}^{k-2}\mathrm{P}\left(X_t=j\right)o(h)+$$

$$+\mathrm{P}(X_t=k-1)(\lambda h+o(h))+\mathrm{P}(X_t=k)(1-\lambda h+o(h))$$

$$\frac{\mathrm{d}p_k}{\mathrm{d}t}(t) = \lim_{h\to 0+} \frac{p_k(t+h) - p_k(t)}{h} = \lambda p_{k-1}(t) - \lambda p_k(t)$$

Máme tedy soustavu lineárních diferenciálních rovnic

$$p_0'(t) = -\lambda p_0(t)$$

$$p_k'(t) = \lambda(p_{k-1}(t) - p_k(t))$$

která je za počátečních podmínek

$$p_0(0) = 1$$

$$p_k(0) = 0$$

a po zavedení funkce $q_k(t) = p_k(t)e^{\lambda t}$ řešitelná rekurentně. □

Poznámka 1.16. *Důležitými předpoklady v zákonu řídkých jevů jsou*

1. *Nezávislost na minulosti.*
2. *Pravděpodobnost, že v $X_{t+h} - X_t$ nastane právě jedna událost je zhruba lineární funkcí délky intervalu.*

Příklad 1.17 (Příklad pro kuchaře)**.** *Na zadělání* 1000 *koláčů dáme* 10000 *rozinek. Najděte rozdělení počtu rozinek v náhodně vybraném koláči.*

1. *Jaká je pravděpodobnost, že na koláči bude více než 5 rozinek?*
2. *Jaká je pravděpodobnost, že na koláči nebude žádná rozinka?*

Vezměme libovolný koláč, ptejme se, zda je na něm n-tá rozinka, a tento pokus opakujme $10000x$ (protože právě tolik je rozinek). Každá rozinka může být na jednom z tisíce koláčů, a to se stejnou pravděpodobností, proto

$$p_k = \frac{1}{1000}$$

Jako X označme počet rozinek na koláči, přičemž $\lambda = n \cdot p_k = 10$. Potom dle zákona velkých čísel platí

$$\mathrm{P}_n(k) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{10^k}{k!}e^{-10} \qquad \forall k \in \widehat{10000}$$

takže

1. $$\mathrm{P}(X > 5) = 1 - \mathrm{P}(X = 0) - \ldots - \mathrm{P}(X = 5) \doteq 0,934$$
2. $$\mathrm{P}(X = 0) = \ldots$$

Poznámka 1.18. *V předchozím případě jsme ale předpokládali, že se na libovolný koláč vejde libovolný počet rozinek. Jinak by totiž nebyl splněn předpoklad o nezávislosti na minulosti.*

Definice 1.19 (Vícerozměrná diskrétní hustota)**.** *Vícerozměrnou diskrétní hustotu definijeme jako*

$$f_{\mathbb{X}} = \mathrm{P}\left(\cap_{j=1}^{n} \{X_j = x_j\}\right) \tag{11}$$

Definice 1.20 (Multinomické rozdělení)**.** *Uvažujme měřitelný prostor* $(\Omega, \mathcal{A})$*, a nechť* $A_1, \ldots, A_n \in \mathcal{A}$*. Nechť dále platí*

1. A_i *jsou disjunktní*
2. $\sum_i A_i = \Omega$
3. $\mathrm{P}(A_i) = p_i$

Buď X_j *počet opakování jevu jevu* A_j *v* n *pokusech. Přitom zřejmě platí*

$$\sum_{j=1^k} \mathrm{P}(A_j) = n$$

Potom

$$f_{\mathbb{X}}(x) = \mathrm{P}\left(X_1 = x_1, \ldots, X_k = x_k\right) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \tag{12}$$

Poznámka 1.21. *Značíme*

$$\frac{n!}{x_1! x_2! \cdots x_k!} = \binom{n}{x_1! x_2! \cdots x_k!}$$

Omezení tedy nutně je

$$0 \leq x_1 \leq n$$

$$0 \leq x_2 \leq n - x_1$$

$$0 \leq x_3 \leq n - x_1 - x_2$$

$$\vdots$$

$$x_k = n - x_1 - x_2 - \cdots - x_{k-1}$$

Potom

$$\sum_{j=1}^{k} X_j = n$$