PRAVDĚPODOBNOST A STATISTIKA, PŘEDNÁŠKA 16. 3. 2020

Doporučená literatura: Likeš J., Machek J.: Počet pravděpodobosti. SNTL, Praha 1982.

Diskrétní rozdělení

Alternativní rozdělení

$X \sim A(p)$ ... náhodná veličina X má alternativní rozdělení s parametrem $p \in (0,1)$.

Pravděpodobnostní funkce: $P(X = 1) = p, P(X = 0) = 1 - p$

Číselné charakteristiky: $E(X) = p, var(X) = p(1 - p)$

Binomické rozdělení

$X \sim Bi(n, p)$ ... náhodná veličina X má binomické rozdělení řádu n s parametrem $p \in (0,1)$.

Pravděpodobnostní funkce: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k \in \{0,1, \dots, n\}$

Číselné charakteristiky: $E(X) = np, var(X) = np(1 - p)$

Poznámky

- Označme A možný výsledek náhodného pokusu. Potom náhodná veličina vyjadřující četnost jevu A v n nezávislých opakováních náhodného pokusu má binomické rozdělení s parametry n a $p = P(A)$. (např. počet 6, které padnou v n hodech hrací kostrkou)
- Jsou-li $X_1, \dots, X_n$ nezávislé náhodné veličiny s alternativním rozdělením s parametrem p, potom náhodná veličina $X = X_1 + \cdots + X_n$ má binomické rozdělení $Bi(n, p)$.

Hypergeometrické rozdělení

$X \sim HGe(N, M, n)$ ... náhodná veličina X má hypergeometrické rozdělení s parametry N, M, n, kde $0 < M \leq N$, $0 < n \leq N$.

Pravděpodobnostní funkce: $P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \ \max\{0, n - (N - M)\} \leq k \leq \min\{M, n\}$

Číselné charakteristiky: $E(X) = n\frac{M}{N}, var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$

Poznámka

- Celkem bylo vyrobeno N výrobků, mezi nimi je M $(\leq N)$ výrobků vadných. Jestliže náhodně vybereme n $(\leq N)$ výrobků z celkového počtu, potom náhodná veličina vyjadující počet vadných výrobků mezi n vybranými má hypergeometrické rozdělení.
  (Sportka - je losováno $M = 6$ čísel z $N = 49$, $n = 6$ čísel označeno na tiketu. Počet uhádnutých čísel má hypergeometrické rozdělení.)

Geometrické rozdělení

$X \sim Ge(p)$ ... náhodná veličina X má geometrické rozdělení s parametrem $p \in (0,1)$.

Pravděpodobnostní funkce: $P(X = k) = p(1-p)^k, k \in \{0,1,\dots\}$

Číselné charakteristiky: $E(X) = \frac{1-p}{p}, var(X) = \frac{1-p}{p^2}$

Poznámka

- Označme A možný výsledek náhodného pokusu („úspěch“). Jestliže opakujeme daný náhodný pokus, potom náhodná veličina vyjadřující počet „neúspěchů“ do prvníbo „úspěchu“ má geometrické rozdělení s parametrem $p = P(A)$.
  (např. X ... počet hodů hrací kostkou, než nám prvně padne číslo 1)

Negativně binomické rozdělení

$X \sim NBi(n, p)$ ... náhodná veličina X má negativně binomické rozdělení s parametry n a $p \in (0,1)$.

Pravděpodobnostní funkce: $P(X = k) = \binom{n+k-1}{k} p^n (1-p)^k, k \in \{0,1,\dots\}$

Číselné charakteristiky: $E(X) = n\frac{1-p}{p}, var(X) = n\frac{1-p}{p^2}$

Poznámka

- Označme A možný výsledek náhodného pokusu („úspěch“). Jestliže opakujeme daný náhodný pokus, potom náhodná veličina vyjadřující počet „neúspěchů“ do n-tého „úspěchu“ má geometrické rozdělení s parametrem $p = P(A)$.
  (např. hod hrací kostkou, úspěch = padne číslo 1, neúspěch = nepadne číslo 1; X ... počet neúspěchů do n-tého „úspěchu“)

Poissonovo rozdělení

$X \sim Po(\lambda)$ ... náhodná veličina X má Poissonovo rozdělení s parametrem $\lambda > 0$.

Pravděpodobnostní funkce: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k \in \{0,1,\dots\}$

Číselné charakteristiky: $E(X) = var(X) = \lambda$

Poznámka

- Počet nezávislých událostí (poruch, dotazů na server, žádostí o obsluhu apod.) během časového intervalu pevné délky má obvykle Poissonovo rozdělení.

Multinomické rozdělení

$\boldsymbol{X} = (X_1, X_2, \dots, X_k) \sim Multi(p_1, p_2, \dots, p_k)$ ... náhodný vektor $\boldsymbol{X}$ má multinomické rozdělení řádu $n \in N^+$ s parametry $p_i \in (0,1)$, kde $\sum_{i=1}^{k} p_i = 1$.

Pravděpodobnostní funkce: $P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} p_1^{n_1} \cdot p_2^{n_2} \cdot \dots \cdot p_k^{n_k}$,

kde $n = \sum_{i=1}^{k} n_i, n_i \in N$

Poznámky

- Nechť $B_1, B_2, \dots, B_k$ je úplná soustava jevů náhodného pokusu, $P(B_i) = p_i$ (t.j. $\sum_{i=1}^{k} p_i = 1$). Označíme-li X_i $(i = 1, \dots, k)$ četnost jevu B_i v n nezávislých opakováních uvažovaného náhodného pokusu, potom $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ má multinomickérozdělení.
- Pro jednotlivé složky X_i vektoru $\boldsymbol{X}$ platí $X_i \sim Bi(n, p_i)$, tedy $E(X_i) = np_i$, $var(X_i) = np_i(1 - p_i)$.

Spojitá rozdělení

Rovnoměrné rozdělení

$X \sim R(a, b)$ ... náhodná veličina X má rovnoměrné rozdělení na intervalu (a, b), kde $a < b$.

Hustota: $$f(x) = \begin{cases} 1/(b-a), & \text{pro } x \in (a,b) \\ 0, & \text{jinde} \end{cases}$$

Distribuční funkce: $$F(x) = \begin{cases} 0 & \text{pro } x \leq a \\ (x-a)/(b-a) & \text{pro } a < x < b \\ 1 & \text{pro } x \geq b \end{cases}$$

Kvantil: $x_\alpha = a + \alpha(b - a)$

Číselné charakteristiky: $E(X) = (a+b)/2$, $var(X) = (b-a)^2/12$

Exponenciální rozdělení

$X \sim Exp(\lambda)$ ... náhodná veličina X má exponenciální rozdělení s parametrem $\lambda > 0$.

Hustota: $$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{pro } t \geq 0 \\ 0, & \text{jinde} \end{cases}$$

Distribuční funkce: $$F(t) = \begin{cases} 1 - e^{-\lambda t}, & \text{pro } t \geq 0 \\ 0, & \text{jinde} \end{cases}$$

Kvantil: $t_\alpha = -\ln(1-\alpha)/\lambda$

Číselné charakteristiky: $E(X) = 1/\lambda$, $var(X) = 1/\lambda^2$

Poznámky

- Exponenciální rozdělení je také označováno jako rozdělení bez paměti. Důvodem je následující fakt - jestliže X má exponenciální rozdělení, potom zřejmě platí
$$P(X > t + s | X > s) = P(X > t)$$
Možná „spolehlivostní“ interpretace - pravděpodobnost, že zařízení bude pracovat bez poruchy po následující dobu délky t nezáleží na tom, jak dlouho dosud zařízení pracovalo (odpovídá provozu zařízení v ustálenem stavu).
- Pokud doby mezi dvěma po sobě jdoucími událostmi jsou nezávislé náhodné veličiny s exponenciálním rozdělením s paramatrem λ, potom počet událostí v časovém intervalu délky t, má Poissonovo rozdělení s parametrem λt. Označíme-li tedy $N(t)$ počet událostí v časovém intervalu délky t, potom
$$P(N(t) = k) = {(\lambda t)^k}/{k!} \cdot e^{-\lambda t}, k = 0,1, \dots$$
- Teorie spolehlivosti:
Označíme-li $X \sim Exp(\lambda)$ dobu do poruchy, potom
    - $R(t) = 1 - F(t) = e^{-\lambda t}$ ... je tzv. funkce přežití (pravděpodobnost, že zařízení přežije dobu t).
    - $\lambda(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{R(t)}$ ... je tzv. intenzita poruchy. V případě exponenciálního rozdělení platí $\lambda(t) = \lambda$, tedy je konstantní v čase.
    - $\lambda(t)dt$ lze interpretovat jako (podmíněnou) pravděpodobnost toho, že se zařízení porouchá v intervalu $(t, t + dt\rangle$, jestliže se do doby t neporouchalo.
    - $f(t)dt$ lze interpretovat jako (nepodmíněnou) pravděpodobnost toho, že se zařízení porouchá v intervalu $(t, t + dt\rangle$).

Weibullovo rozdělení

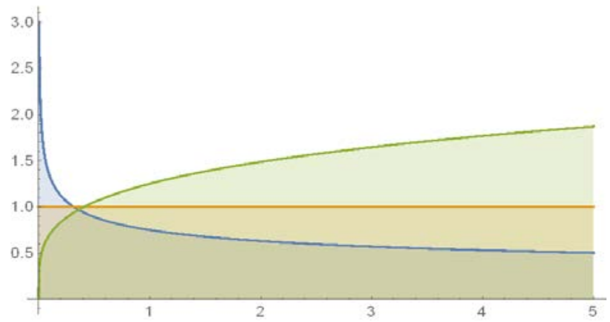
$X \sim W(\alpha, \lambda)$ ... náhodná veličina X má Weibullovo rozdělení s parametry $\alpha, \lambda > 0$.

Hustota: $$f(t) = \begin{cases} \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha}, & \text{pro } t \geq 0 \\ 0, & \text{jinde} \end{cases}$$

Distribuční funkce: $$F(t) = \begin{cases} 1 - e^{-\lambda t^\alpha}, & \text{pro } t \geq 0 \\ 0, & \text{jinde} \end{cases}$$

Poznámky

- Platí $W(1, \lambda) = Exp(\lambda)$.
- V teorii spolehlivosti se Weibullovo rozdělení používá jako obecnější (v porovnání s exponenciálním) rozdělení doby do poruchy, které umožňuje popsat i nekonstantní intenzity poruch. Pro intenzitu poruchy totiž platí $\lambda(t) = \lambda \alpha t^{\alpha-1}$. Následující grafy zobrazují průběh intenzity $\lambda(t)$ pro $\lambda = 1$ a $\alpha \in \{0{,}75; 1; 1{,}25\}$.



Normální rozdělení

$X \sim N(\mu, \sigma^2)$ ... náhodná veličina X má normální rozdělení s parametry $\mu \in R, \sigma^2 > 0$.

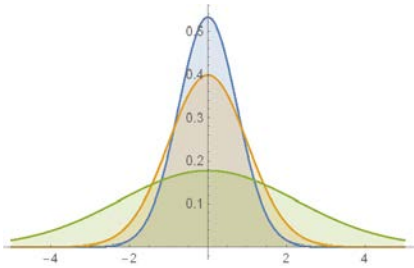
Distribuční funkce: $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du$

Hustota: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mu \in R, \sigma^2 > 0$

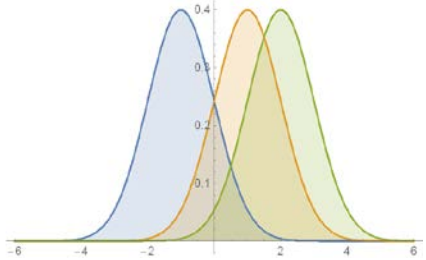
Číselné charakteristiky: $E(X) = \mu, var(X) = \sigma^2$ ($\sigma > 0$ ... je tzv. směrodatná odchylka)

Poznámky

- Nejdůležitější rozdělení v teorii pravděpodobnosti i matematické statistice. Lze ho použít tam, kde je zdrojem „měnlivosti“ součet velkého množství nezávislých vlivů. Limitní rozdělení celé řady rozdělení.
- Následující grafy zobrazují průběhy hustoty $N(\mu, \sigma^2)$ pro různé střední hodnoty a sm. odchylky:



stejná střední hodnota $\mu = 0$
sm. odchylky $\sigma = 0{,}75; 1; 2{,}25$



různé střední hodnoty $\mu = -1; 1; 2$,
stejná sm. odchylka σ =1

- Z důvodu symetrie normálního rozdělení $N(\mu, \sigma^2)$ platí:

$$F(x) = 1 - F(2\mu - x), \quad \varphi(x) = \varphi(2\mu - x), \quad x_{1-\alpha} = 2\mu - x_\alpha.$$

- $U \sim N(0,1)$ ... tzv. normované (standardizované) normální rozdělení
  - $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{v^2}{2}} dv$ ... distribuční funkce normovaného normálního rozdělení

    z důvodu symetrie platí $\Phi(-u) = 1 - \Phi(u)$, (hodnoty $\Phi(u)$ tabelovány, obvykle pro $u \geq 0$)
  - $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ ... hustota normovaného normálního rozdělení

    z důvodu symetrie platí $\varphi(-u) = \varphi(u)$, (hodnoty $\varphi(u)$ tabelovány, obvykle pro $u \geq 0$)
  - u_α ... $\alpha\%$ kvantil normovaného normálního rozdělení

    z důvodu symetrie platí $u_{1-\alpha} = -u_\alpha$ (hodnoty u_α tabelovány, obvykle pro $\alpha \geq 0{,}5$).
- Má-li $X \sim N(\mu, \sigma^2)$ distribuční funkcí $F(x)$, hustotou $f(x)$ a kvantil x_α,

  $U \sim N(0,1)$ distribuční funkcí $\Phi(u)$, hustotou $\varphi(u)$ a kvantil u_α, potom platí:

$$X = \mu + \sigma U, \quad F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x_\alpha = \mu + \sigma u_\alpha.$$

V matematické statistice se kromě normálního rozdělení velmi často setkáme s následujícími rozděleními (v těchto případech budeme pracovat především s jejich kvantily, které jsou tabelovány).

Rozdělení χ^2

$X \sim \chi^2(n), n \in N^+$ ... náhodná veličina X má rozdělení „chí-kvadrát“ s n stupni volnosti.

Kvantil: $\chi^2_\alpha(n)$ ... α% kvantil rozdělení „chí-kvadrát“ s n stupni volnosti (hodnoty tabelovány)

Číselné charakteristiky: $E(X) = n, var(X) = 2n$

Poznámka

- Jsou-li $U_1, \dots, U_n$ nezávislé náhodné veličiny mající každá rozdělení $N(0,1)$, potom $X = U_1^2 + \dots + U_n^2$ má rozdělení $\chi^2(n)$.
- Pro $n \geq 30$ lze použít aproximaci $\chi^2_\alpha(n) = \frac{1}{2}\left(\sqrt{2n-1} + u_\alpha\right)^2$, kde u_α je α% kvantil $N(0,1)$

Rozdělení t (Studentovo)

$X \sim t(n), n \in N^+$ ... náhodná veličina X má rozdělení t s n stupni volnosti.

Kvantil: $t_\alpha(n)$ ... α% kvantil rozdělení t s n stupni volnosti (hodnoty tabelovány)

Číselné charakteristiky: $E(X) = 0 \text{ (pro } n > 1), var(X) = \frac{n}{n-2} \text{ (pro } n > 2)$

Poznámky

- Rozdělení $t(n)$ je symetrické kolem 0 a proto platí $t_{1-\alpha}(n) = -t_\alpha(n)$.
- Jsou-li $U \sim N(0,1)$ a $Y \sim \chi^2(n)$ nezávislé náhodné veličiny, potom $X = \frac{U}{\sqrt{Y/n}}$ má rozdělení $t(n)$.

Rozdělení F (Fischerovo-Snedecorovo)

$X \sim F(m,n)$, $m, n \in N^+$ ... náhodná veličina X má rozdělení F s m a n stupni volnosti.

Kvantil: $F_\alpha(m,n)$ ... α% kvantil rozdělení F s m a n stupni volnosti (hodnoty tabelovány)

Číselné charakteristiky: $E(X) = \frac{n}{n-2} \text{ (pro } n > 2), var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ (pro } n > 4)$

Poznámky

- Pro kvantily rozdělení $F(m,n)$ platí $F_{1-\alpha}(m,n) = {}^1/_{F_\alpha(n,m)}$.
- Jsou-li $Y_1 \sim \chi^2(m)$ a $Y_2 \sim \chi^2(n)$ nezávislé náhodné veličiny, potom $X = \frac{Y_1/m}{Y_2/n}$ má rozdělení $F(m,n)$.