

Obsah

0	Než začnete číst	3
1	Přístupy k pravděpodobnosti	3
1.1	Bayesovský přístup	3
1.2	Individualistický přístup	3
1.3	Přístup frekvenční	3
2	Podmíněná pravděpodobnost	4
2.1	Úloha o zruinování hráče	4
3	Cauchy-Schwartzova nerovnost	10
4	Diskrétní náhodná veličina	11
4.1	$Ge(p)$	12
4.2	Negativně binomické rozdělení	12
4.3	Hypergeometrické rozdělení	13
4.4	Poissonovo rozdělení	14
4.5	Rozdělení rovnoměrné (diskrétní)	14
4.6	Náhodná procházka	15
5	Čebyševova nerovnost a Čebyševova věta	15
6	Spojité náhodné veličiny	17
6.1	Příklady rozdělení	18
6.1.1	Rovnoměrné	18
6.1.2	Exponenciální rozdělení	19
6.1.3	Normální (Gaussovo) rozdělení	20
6.2	Centrální limitní věta	21
7	Simulace náhody na počítačích	23
7.1	Inverzní metoda	23
7.2	Zamítací metoda	24
8	Rozhodování mezi dvěma tvrzeními	24
8.1	Sekvenční přístup	24
8.2	Přístup přes CLV	25
8.2.1	Oboustranná varianta	25
8.3	Kvalita postupu	25

9	Třídění	26
9.1	Generování náhodných permutací	26
9.2	Kotouč	26
9.3	Párový test	28

0 Než začnete číst

Tento text vznikl přepisem poznámek z přednášek Prof. RNDr. Jaromíra Antocha, CSc. do trochu čitelné podoby. Původně tento dokument byl zamýšlen jako moje pomůcka.

Neručím za naprostou správnost všech písmenek na následujících stranách.

Jaromír Kryš

Př. 1.
5. 10. 2009

1 Přístupy k pravděpodobnosti

Půjde nám o formalizování výsledků náhodných pokusů (například mince, kostka, tahání čísel z klobouku).

Pokusy, které mají statistickou stabilitu: Při velkém opakování pokusů se výsledek stabilizuje.

Stabilita \equiv matematický model, který se snaží tuto stabilitu (četnost výskytů) popsat.

1.1 Bayesovský přístup

Počítá se pomocí podílu šancí následujícím způsobem: Mějme 2 jevy A_1 a A_2 tak, že šance na nastání těchto jevů je $a_1 : a_2$. Potom $P(A_1) = \frac{a_1}{a_1 + a_2}$ a $P(A_2) = \frac{a_2}{a_1 + a_2}$. Pro n jevů můžeme psát obecně $P(A_i) = \frac{a_i}{\sum_{j=1}^n a_j}$ pro $1 \leq i \leq n$.

1.2 Individualistický přístup

(Klasická pravděpodobnost, Princip symetrie)

Model: Konečný počet možných základních (elementárních) výsledků a všechny výsledky jsou nepopsatelné jednoduššími a mají tu samou šanci nastat.

Pravděpodobnost jakéhokoliv jevu J složeného z jevů základních se spočítá vztahem $P(J) = \frac{a}{b}$, kde a značí počet příznivých výsledků a b značí počet všech možných výsledků.

1.3 Přístup frekvenční

(četnostní, statistický)

$$P(zdar) = \frac{n_z}{n} \approx p$$

Matematicky se to neprosadilo.

Za většinou pravděpodobností bych si nějaký takovýto pokus měl umět představit = PŘÁNÍ.

Důvod k PŘÁNÍ je, že současná pravděpodobnost se snaží četnosti modelovat.

Z důvodů matematické přesnosti, algoritmizace atd. se prosadil model přes teorii množin jako základní, byť ne jediný.

Definice: Diskrétní pravděpodobnostní prostor je trojice (Ω, \mathcal{A}, P) kde:

Ω je množina spočetná nebo konečná,

\mathcal{A} je soubor množin Ω uzavřen na spočetné průniky a spočetná sjednocení; prvky \mathcal{A} jsou náhodnými jevy a prvky Ω jsou elementární náhodné jevy,

P je definována následujícím způsobem: $P(\Omega) = 1$ (říkáme, že Ω je jistý jev)

$\forall A \in \mathcal{A}: 0 \leq P(A) \leq 1$

$\forall A_1, A_2$ navzájem disjunktní platí $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$ \square

Z definice vyplývá $P(\emptyset) = 0$. \emptyset nazýváme jev nemožný.

Př. 2.
12. 10. 2009

Rozdělení pravděpodobnosti - jednotkovou pravděpodobnost rozdělím mezi elementární jevy.

Speciální případ (klasická pravděpodobnost): Nechť $\text{card } \Omega = n$ a $P(\omega_i) = \frac{1}{n}$. Potom hovoříme o úloze tzv. klasické pravděpodobnosti, kde $P(A) = \frac{\aleph}{n}$, kde \aleph je počet případů příznivých pro A a n je počet všech případů.

Definice: Nechť $\Omega = \bigcup_i H_i$ a $\forall i \neq j, H_i \cap H_j = \emptyset$. Pak $\{H_i\}_i$ nazveme úplný systém jevů.

Nechť $A = A \cap \Omega = A \cap \left(\bigcup_i H_i\right) = \bigcup_i (A \cap H_i)$ a nechť H_i jsou disjunktní. Pak $A \cap H_i$ jsou také disjunktní.

Věta (O úplné pravděpodobnosti): Jestliže $\{H_i\}$ je úplný systém jevů na (Ω, \mathcal{A}, P) , pak $\forall A \in \mathcal{A}: P(A) = \sum_i P(A \cap H_i)$.

2 Podmíněná pravděpodobnost

Příklad šťastných a nešťastných studentů: Podstatné je zjistit, co se stalo v prvním kroku. 1. student byl buď šťastný nebo nešťastný. Následující studenty tedy podmíním výsledkem prvního tahu. 1. tah je jistý jev (nějak to dopadnout muselo).

Definice: Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní model a A, B jsou jevy. Podmíněnou pravděpodobnostní nazveme číslo $P(A|B) = \frac{P(A \cap B)}{P(B)}$, pokud $P(B) > 0$.

Věta: $P(A \cap B) = P(A|B) \cdot P(B)$

Poznámka: Četnostní interpretace naší definice

Mám n pokusů. Vyberu jen ty, kdy nastalo B ; zbytek mě nezajímá. Mezi těmito jevy se podívám, kdy také nastalo A .

Pokud $P(\omega_i) = \frac{1}{\text{card } \Omega}$, pak by $P(A|B) = \frac{n_{A \cap B}}{n_B} = \frac{\frac{n_{A \cap B}}{n}}{\frac{n_B}{n}} = \frac{P(A \cap B)}{P(B)}$, tedy četnostní interpretace naší definice v případě klasické pravděpodobnosti.

Poznámka: Hlavním cílem je zúžit prostor výsledků, přes který pravděpodobnost hledám.

2.1 Úloha o zruinování hráče

Hráč A má i Kč. Hráč B má $n - i$ Kč. Buď 1 Kč vyhraju ($P = \frac{1}{2}$), nebo 1 Kč prohraju. $\Omega = \{1_V, 1_P\}$.

$$p_i = P(A \text{ vyhraje všechny peníze}) \equiv P(VV_i)$$

$$P(VV_i) = P(VV_i|1_V) \cdot P(1_V) + P(VV_i) \cdot P(1_P)$$

Dostáváme tedy $p_i = p_{i+1} \cdot \frac{1}{2} + p_{i-1} \cdot \frac{1}{2}$. Navíc víme, že $p_0 = 0$ a $p_n = 1$. Dostaneme tedy $\forall i \in \{1, \dots, n-1\} 2 \cdot p_i = p_{i+1} + p_{i-1}$. Jednoduchým odvozením dostaneme $\forall i \in \{2, \dots, n\} p_i = i \cdot p_1$. A tedy $p_i = \frac{1}{n}$ a $p_i = \frac{i}{n}$.

p_i netvoří rozdělení pravděpodobnosti

$$\sum_{i=1}^n p_i = \frac{1}{n} \sum_{i=1}^n i = \frac{n \cdot (n+1)}{2n} = \frac{n+1}{2}.$$

Rozumný pravděpodobnostní prostor je $\Omega = \{\text{vyhrají vše, prohrají vše}\}$.

Příklad: Na rozmyšlení

Polyovo urnové schéma:

Mějme koule různých barev v osudí. Zamícháme je a vytáhneme jednu kouli. Kouli poté vrátíme a sní přidáme k koulí té samé barvy pro $k \in \mathbb{Z}$. Zajímá nás pravděpodobnost, že v j -tém tahu vytáhnou bílou kouli.

$k = -1$ odpovídá tahům bez vracení

$k = 0$ odpovídá tahům s vracením

Př. 3.
19. 10. 2009

Příklad: Mějme 2 dokonalé kostky. Chceme zjistit s jakou pravděpodobností padne součet 5 dříve, než součet 7. Můžu to zjistit následujícími způsoby:

- Podíl šancí: rychle dostanu výsledek $\frac{2}{5}$.

- Podmínit výsledkem prvního pokusu: $H_1 = S_5, H_2 = S_7, H_3 = \text{zbytek}$

$$P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + P(A|H_3) \cdot P(H_3)$$

$$P(A|H_1) = 1$$

$$P(H_1) = \frac{1}{9}$$

$$P(A|H_2) = 0$$

$$P(A|H_3) = P(A)$$

$$P(H_3) = \frac{13}{18}$$

$$\text{Z předchozího dostáváme } P(A) = \frac{2}{5}.$$

- Čekáním: Zajímá nás posloupnost $\{H_3 \dots H_3 H_1\}$. Definujme jev E_i jako jev, kdy S_5 nastal v

čase i a S_7 nenastal v žádném z časů $1, 2, \dots, i$. Potom $P(A) = \sum_{i=1}^{\infty} P(E_i) = \frac{1}{9} \cdot \sum_{i=1}^{\infty} \left(\frac{13}{18}\right)^i =$

$$\frac{1}{9} \cdot \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}.$$

Definice: Řekneme, že jevy A a B jsou nezávislé, jestliže platí $P(A \cap B) = P(A) \cdot P(B)$.

Věta: Nechť platí $P(A|B) = \frac{P(A \cap B)}{P(B)}$ a nechť A a B jsou nezávislé jevy. Pak $P(A|B) = P(A)$.

Věta: Jestliže A a B jsou nezávislé jevy, pak také A a \overline{B} , \overline{A} a B , \overline{A} a \overline{B} jsou nezávislé.

Definice: Řekneme, že $\{A_i\}_{i=1}^n$ jsou sdruženě nezávislé, jestliže $\forall k \in \{2, \dots, n\}$ platí $P(A_{i_1} \cap \dots \cap$

$$A_{i_k}) = \prod_{j=1}^k P(A_{i_j}).$$

Definice: Řekneme, že A_1, \dots, A_n jsou:

- po dvou nezávislé, jestliže $\forall i, j \quad i \neq j \quad P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$
- po třech nezávislé, jestliže $\forall i, j, k \quad i \neq j \neq k, \quad P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cdot P(A_k)$

Věta: Jsou-li jevy sdruženě nezávislé, pak jsou nezávislé po dvou a po třech.

Poznámka: předchozí věta naopak neplatí

Příklad: Mějme pravidelný čtyřstěn s obarvenými stranami na následující barvy: červená, modrá, bílá a ČMB.

$$P(\text{cervena}) = P(\text{modra}) = P(\text{bila}) = \frac{1}{2}$$

$$P(\text{CMB}) = \frac{1}{4}$$

$$P(\text{CM}) = P(\text{CB}) = P(\text{MB}) = \frac{1}{4}$$

Příklad: Nejjednodušší čekání je na 1_Z v posloupností hodů mincí. Pokud $P(Z) = p$ a $P(N) = 1 - p$, pak $P(NN \dots NZ) = (1 - p)^{i-1} \cdot p$.

Příklad: Dáváme krev, za 14 dní dostaneme dopis, že test dopadl pozitivně. Chceme spočítat $P(\text{mám AIDS} \mid \text{test dopadl pozitivně})$.

$$P(A_+ \mid T_+) = \frac{P(A_+ \cap T_+)}{P(T_+)} = \frac{P(T_+ \mid A_+) \cdot P(A_+)}{P(T_+)} = \frac{P(T_+ \mid A_+) \cdot P(A_+)}{P(T_+ \mid A_+) \cdot P(A_+) + P(T_+ \mid A_-) \cdot P(A_-)}$$

$$P(A_+) = \frac{3}{10000}, \quad P(A_-) = 1 - P(A_+)$$

Věta (Bayesova věta, Věta o inverzní pravděpodobnosti): Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní model, kde $\Omega = \bigcup_i H_i$ a $\forall i \neq j \quad H_i \cap H_j = \emptyset$. Dále nechť A je jev. Potom

$$P(H_i \mid A) = \frac{P(H_i \cap A)}{P(A)} = \frac{P(A \mid H_i) \cdot P(H_i)}{\sum_i P(A \mid H_i) \cdot P(H_i)}$$

Př. 4.
26. 10. 2009

Definice: Mějme pravděpodobnostní prostor (Ω, \mathcal{A}, P) . Potom náhodná veličina je funkce $X : (\Omega, \mathcal{A}, P) \rightarrow (R_1, \mathcal{B})$.

Náhodnou veličinu definujeme proto, abychom si zjednodušili život. Zajímá nás jak náhodnou veličinu charakterizovat (popsat).

Příklad: Hod mincí n -krát nezávisle dokonalou mincí. V tomto případě $P(Z) = P(N) = \frac{1}{2}$.

Zajímá nás $P(\#Z = k)$, kde $k \in \{0, 1, \dots, n\}$.

Ω obsahuje všechny posloupnosti nul a jedniček délky n . Tedy $|\Omega| = 2^n$. Označme $\omega = \{0, 1\}_n$. Pak

$$P(\omega) = \frac{1}{2^n} \text{ (díky nezávislosti jevů a dokonalosti mince). Zajímá mne } P_i = P(X = i) = \binom{n}{i} \cdot \frac{1}{2^n}$$

$$\sum_{i=0}^n \binom{n}{i} \cdot \frac{1}{2^i} \cdot \frac{1}{2^{n-i}}$$

Definice: Základní charakterizací bude tzv. rozdělení pravděpodobnosti: $P(X = x_i)$, kde x_i jsou možné hodnoty náhodné veličiny X .

Je-li $\text{card } \Omega$ konečná (spočetná), pak i $\text{card } X$ je konečná (spočetná).

v našem případě $x_i = 0, 1, \dots, n$

$$P_i = P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$

$P(\Omega) = 1$, $\Omega = \bigcup_i H_i$, $H_i \rightarrow x_i$, zde H_i je třída ekvivalence těch ω , které se mi zobrazí na ten samý bod ω_i .

$$P\left(\bigcup_i H_i\right) = \sum_i P(H_i) = 1$$

Rozdělení pravděpodobností náhodné veličiny X je dáno:

- hodnotami x_i , jichž nabývá
- pravděpodobnostmi p_i s nimiž jich nabývá

Pro tyto p_i platí $\sum_i p_i = 1$.

Poznámka: Rozdělení pravděpodobnosti se užívá proto, že jsem jednotkovou pravděpodobnost Ω rozdělil na menší kousky.

Poznámka: Pravděpodobnosti odpovídající náhodné veličině X měřím v (Ω, \mathcal{A}, P) :

$$P(X = i) = \sum P(\omega) \quad \forall \omega : X(\omega) = x_i$$

Poznámka: Často umím Ω zkonstruovat. Podívejme se na celou situaci obráceně: dostaneme tabulku $\begin{bmatrix} x_i \\ p_i \end{bmatrix}$, kde $\sum_i p_i = 1$ a $p_i \geq 0$. Každá takováto tabulka je rozdělení nějaké náhodné veličiny. Stačí vzít $\Omega = \{\omega_1, \omega_2, \dots\}$ a X identitu, tedy $(\omega_i \equiv x_i) \Rightarrow p(\omega) = p(x_i) = p_i$.

Poznámka: Nechtě f a g jsou funkce. Pak i $g(f)$ je funkce. Nechtě f a X jsou náhodné veličiny, pak $f(X)$ je opět náhodná veličina.

Definice: Nechtě X je náhodná veličina. Potom distribuční funkce je definována vztahem $F(x) = P(X \leq x)$ pro $x \in \mathbb{R}$.

Dohoda: Distribuční funkce se zpravidla značí F, G, H, \dots .
Náhodné veličiny se zpravidla značí X, Y, Z, \dots .

$$(i) \quad \lim_{x \rightarrow \infty} F(x) = 1, \quad \lim_{x \rightarrow -\infty} F(x) = 0$$

(ii) jedná se o funkci po částech konstantní se skoky velikosti p_i v bodech x_i

Graf rozdělení pravděpodobnosti se mnohem snáze interpretuje, než odpovídající graf distribuční funkce.

Definice: Nechtě X je dáno $\begin{bmatrix} i \\ p_i \end{bmatrix}$ pro $i \in \mathcal{N}_0$. Takovou X nazveme celočíselnou pravděpodobnostní veličinou. Potom její vytvářející funkce je definována vztahem $A(x) = \sum_i p_i \cdot x^i$.

Tedy $A(1) = \sum_i p_i \cdot 1^i = \sum_i p_i = 1$. Platí $p_i = \frac{A^{(i)}(0)}{i!}$.

Poznámka: Vytvářející funkce $A(x)$ jednoznačně určuje celočíselnou náhodnou veličinu.

$$\textbf{Důkaz:} \quad A'(x) = \sum_i i \cdot p_i \cdot x^{i-1}$$

po dosazení $x = 0$ dostáváme $0^0 = 1$ a $0^n = 0$

$$A''(x) = \sum_i i \cdot (i-1) \cdot p_i \cdot x^{i-2} \quad \square$$

Poznámka: Cesta jde podobně pro libovolné náhodné veličiny s tím, že se zavede pojem charakteristické funkce (zmíníme později).

$$\varphi(t) = Ee^{itX} = \sum_i p_i^{itx_i}$$

Toto je nástroj!

Realizace náhodné veličiny:

$$\begin{bmatrix} -15 & 1/8 \\ -5 & 1/4 \\ 0 & 1/2 \\ 20 & 1/8 \end{bmatrix}$$

tedy po n hrách dostáváme následující:

$$\left. \begin{array}{lll} \frac{n}{8} & -15 & -15 \cdot \frac{n}{8} \\ \frac{n}{4} & -5 & -5 \cdot \frac{n}{4} \\ \frac{n}{2} & 0 & 0 \cdot \frac{n}{2} \\ \frac{n}{8} & 20 & 20 \cdot \frac{n}{8} \end{array} \right\} \frac{\text{celkova vyhra}}{n} = \text{průměrná výhra } \sum p_i \cdot x_i$$

Definice: Střední hodnota náhodné veličiny X je $EX = \sum X(\omega) \cdot P(\omega)$ pro $\omega \in \Omega$

Věta: Platí $EX = \sum_i p_i x_i$.

$$EX = \sum_{\omega} X(\omega)P(\omega) = \sum_{i|X(\omega)=x_i} X(\omega)P(\omega)$$

$$\sum_{\omega|X(\omega)=x_i} P(\omega) = p_i$$

Toto platí jen při velkém opakování naší náhodné veličiny, tedy při velkém počtu her. Není to vůbec rozumná charakteristika jedné hry!

Nechť $\{x_i, p_i\}$ jsou všechny dvojice, kde x_i je plat zaokrouhlený na 1000 a p_i je procentuální část obyvatel, kteří mají plat x_i . Pak $\sum x_i p_i$ je průměrný plat.

Př. 5.
2. 11. 2009

Příklad:

a) hod kostkou: $P(X = i) = \frac{1}{6}$ pro $i \in \{1, 2, 3, 4, 5, 6\}$

$$EX = \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{21}{6} = 3,5$$

b) rovnoměrné rozdělení: $f(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{jinak} \end{cases}$

$$EX = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

c) Cauchyho rozdělení: $f(x) = \frac{1}{\pi \cdot (1+x^2)}$

$$EX = \int_{-\infty}^{\infty} \frac{x}{\pi \cdot (1+x^2)} dx = \text{není definováno}$$

$$EX < \infty \Rightarrow X \in \mathcal{L}_1$$

Vlastnosti střední hodnoty:

$$1) X, Y \in \mathcal{L}_1 \Rightarrow E(aX + bY + c) = aEX + bEY + c$$

$$2) P(X \geq 0) = 1 \Rightarrow EX \geq 0$$

$$3) X \in \mathcal{L}_1 \Rightarrow |X| \in \mathcal{L}_1$$

$$\text{Diskrétní případ: } E\Phi(X) = \sum_{i \in \mathbb{N}} \Phi(x_i) p_i$$

$$\text{Spojitý případ: } E\Phi(X) = \int_{-\infty}^{\infty} \Phi(x) f(x) dx$$

Definice: Ostatní momenty:

- 1) N -tý obecný moment je definován jako EX^N .
- 2) N -tý centrální moment je definován jako $E(X - EX)^N$.
- 3) N -tý absolutní moment je definován jako $E|X|^N$.
- 4) N -tý absolutní centrální moment je definován jako $E|X - EX|^N$.
- 5) Momentová vytvořující funkce je definována jako $\psi(t) = Ee^{tx}$.

Poznámka: Rozptyl je druhý centrální moment:

$$\begin{aligned} \text{var}X &= E(X - EX)^2 = E(X^2 - 2X \cdot EX + (EX)^2) = EX^2 - 2 \cdot EX \cdot EX + (EX)^2 = \\ &= EX^2 - (EX)^2 \end{aligned}$$

Příklad: Rovnoměrné rozdělení: Víme, že $EX = \frac{1}{2}$. Snadno spočítáme $EX^2 = \int_0^1 x^2 dx = \frac{1}{3}$.

Dostáváme tedy $\text{var}X = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$.

Věta: Nechť $\psi(t)$ je konečná pro $t < t_0$, kde $t_0 > 0$. Potom platí $EX^r = \left(\frac{\partial^r \psi(t)}{\partial t^r}\right)(0)$.

Příklad: Nechť X je exponenciální rozdělení:

$$f(x) = \lambda \cdot e^{-\lambda x} \text{ pro } x > 0 \text{ a}$$

$$f(x) = 0 \text{ jinak}$$

$$\psi(t) = Ee^{tx} = \int_0^\infty e^{tx} \cdot \lambda \cdot e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx = \lambda \cdot \left[\frac{e^{(t-\lambda)x}}{t-\lambda} \right]_0^\infty = \frac{\lambda}{\lambda-t}$$

$$\psi'(t) = \frac{\lambda}{(\lambda-t)^2}$$

$$\psi''(t) = \frac{2\lambda}{(\lambda-t)^3}$$

$$\begin{aligned} EX &= \frac{1}{\lambda} \quad EX^2 = \frac{2}{\lambda^2} \\ \text{var}X &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \square \end{aligned}$$

Definice: Mějme nezávislé veličiny $X, Y \in \mathcal{L}_2$.

Kovariaci definujeme jako $\text{cov}(X, Y) = E((X - EX) \cdot (Y - EY))$.

Korelaci definujeme jako $\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}X \cdot \text{var}Y}}$, jestliže $\text{var}X \cdot \text{var}Y > 0$. \square

Nechť $X, Y \in \mathcal{L}_2$ a $a, b, c \in \mathbb{R}$. Potom

$$\begin{aligned} \text{var}(aX + bY + c) &= E(aX + bY + c - aEX - bEY - c)^2 = E(a \cdot (X - EX) + b \cdot (Y - EY))^2 = \\ &= E(a \cdot (X - EX))^2 + 2 \cdot a \cdot b \cdot E((X - EX)(Y - EY)) + E(b^2 \cdot (Y - EY)^2) = \\ &= a^2 \cdot \text{var}X + 2ab \cdot \text{cov}(X, Y) + b^2 \cdot \text{var}Y \square \end{aligned}$$

3 Cauchy-Schwartzova nerovnost

$$\left(\int f \cdot g d\mu \right)^2 \leq \int f^2 d\mu \cdot \int g^2 d\mu$$

$$\left. \begin{array}{l} f = X - EX \\ g = Y - EY \\ \mu = p \end{array} \right\} \text{ pak } (\text{cov}(X, Y))^2 \leq \text{var}X \cdot \text{var}Y$$

pak tedy $\text{cor}(X, Y) \in [-1, 1]$

$$\text{cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

$$\text{var}X = EX^2 - (EX)^2$$

$$\text{var}X = \text{cov}(X, X)$$

Definice: Necht' X_1, \dots, X_n jsou náhodné veličiny a necht' $B_1, \dots, B_n \subset \mathbb{R}$. X_1, \dots, X_n jsou nezávislé právě tehdy, když $P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$. \square

Diskrétní náhodná veličina: Jestliže X_1, \dots, X_n jsou nezávislé veličiny, pak $\forall x_1, \dots, x_n \in \mathbb{R}$:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$

Spojitá náhodná veličina: Necht' X_1, \dots, X_n jsou nezávislé. Pak $\forall x_1, \dots, x_n \in \mathbb{R}$: $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$

Necht' X_1, \dots, X_n jsou nezávislé a necht' $x_1, \dots, x_n \in \mathbb{R}$. Potom $F(x_1 \dots x_n) = \prod_{i=1}^n F_i(x_i)$.

X a Y jsou sdružené distribuční funkce, pokud $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.

$$F_X(x) = P(X \leq x) \quad \square$$

$$X \sim U([-1, 1])$$

$$f(x) = \begin{cases} \frac{1}{2} & x \in [-1, 1] \\ 0 & \text{jinak} \end{cases}$$

$$Y = x^2$$

$$EX = \int_{-1}^1 x \cdot \frac{1}{2} dx = 0$$

$$EX^3 = \int_{-1}^1 x^3 \cdot \frac{1}{2} dx = 0$$

$$\text{cov}(X, Y) = E((X - EX) \cdot (Y - EY)) = EX^3 - EX \cdot EX^2 = 0$$

Př. 6.
9. 11. 2009

Hrajeme hru s následujícími možnými výsledky:

x_i	p_i	při n hrách
-20	$\frac{1}{8}$	$n \cdot \frac{1}{8}$ -20
-5	$\frac{1}{4}$	$n \cdot \frac{1}{4}$ -5
0	$\frac{1}{2}$	$n \cdot \frac{1}{2}$ 0
30	$\frac{1}{8}$	$n \cdot \frac{1}{8}$ 30

V tomto případě $EX = 0$. Nyní změňme výsledek 30 na výsledek 29, zbytek tabulky zůstane stejný. Po této drobné úpravě dostáváme $EX = -\frac{1}{8}$.

Celkový zisk po n hrách bude přibližně $n \cdot \frac{1}{8}(-20) + n \cdot \frac{1}{4}(-5) + n \cdot \frac{1}{2}(0) + n \cdot \frac{1}{8}(29)$.

$$\frac{\text{celkový zisk}}{n} = \text{průměrný zisk} = \sum_i x_i \cdot p_i = EX$$

Střední hodnota má smysl pouze pokud máme hodně realizací náhodné veličiny. Nemá smysl, jdeme-li si zahrát jednou nebo dvakrát.

Věta (o konvoluci): Mějme nezávislé náhodné veličiny X a Y , kde $X = \begin{pmatrix} i \\ p_i \end{pmatrix}_i$ a $Y = \begin{pmatrix} i \\ q_i \end{pmatrix}_i$.

Dále necht' $Z = X + Y$. Pak $Z = \begin{pmatrix} j \\ r_j \end{pmatrix}_j$, kde r_j je konvolucí z p_i a q_i .

$$\begin{aligned} \textbf{Důkaz: } P(Z = i) &= P\left(\bigcup_{j=0}^i (X = j \wedge Y = i - j)\right) = \sum_{j=0}^i P(X = j \wedge Y = j) = \\ &= \sum_{j=0}^i P(X = j) \cdot P(Y = i - j) = \sum_{j=0}^i p_j \cdot q_{i-j} \quad \square \end{aligned}$$

Poznámka: $X = \begin{pmatrix} x_i \\ p_i \end{pmatrix}, Y = \begin{pmatrix} y_i \\ q_i \end{pmatrix}$

$$Z = X + Y \quad \begin{pmatrix} x_i + y_i \\ p_{ij} \end{pmatrix} \equiv \begin{pmatrix} z_k \\ r_k \end{pmatrix}$$

$$P(Z = z_k) = P\left(\bigcup_{(i,j): x_i + y_j = z_k} (X = x_i \wedge Y = y_j)\right) = \sum \dots \quad \square$$

$$X = \begin{pmatrix} i \\ p_i \end{pmatrix} \quad A(x) = \sum_i p_i x^i$$

$$Y = \begin{pmatrix} i \\ q_i \end{pmatrix} \quad B(x) = \sum_i q_i x^i$$

$$Z = X + Y \quad C(x) = ?$$

$$X \text{ a } Y \text{ jsou nezávislé, tedy } P(Z = i) = \sum_{j=0}^i p_j q_{i-j}. \text{ A pak tedy } C(x) = \sum_i \left(\sum_{j=0}^i p_j q_{i-j} \right) x^i.$$

Věta: Jsou-li X a Y nezávislé celočíselné náhodné veličiny, potom $C(x) = A(x) \cdot B(x)$.

4 Diskrétní náhodná veličina

$$X \begin{cases} 0 & 1-p=q \\ 1 & p \end{cases}$$

$$EX = p, \text{ var } X = EX^2 - (EX)^2 = p - p^2 = p \cdot (1-p) = p \cdot q$$

Je to náhodná procházka po mříži v 1. kvadrantu s jednou pohlcující bariérou $y = 1 - x$. \square

Hodíme nezávisle n -krát a ptáme se, jaká je pravděpodobnost, že padlo právě k výher.

$$\Omega = \{0, 1\}_n$$

$$\text{Binomické rozdělení: } Y_n \dots p_k = P(Y_n = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

$$Y_n = \sum_{i=1}^n X_i$$

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

X_i jsou nezávislé $\rightarrow X_i \sim \text{Alt}(p)$

$$EY_n = E\left(\sum_{i=1}^n X_i\right) = n \cdot p$$

$$\text{var}Y_n = \sum_i \text{var}X_i = npq$$

Lze chápat jako náhodnou procházku po mříži v 1. kvadrantu s pohlcující bariérou $y = n - x$.

$$X_i^* \begin{cases} 1 & p \\ -1 & 1-p \end{cases}$$

$$EX_i = p + (-1) \cdot (1-p)$$

4.1 Ge(p)

Čekání na první zdar. Konám nezávislé náhodné pokusy X_i , z nichž každý $X_i \sim \text{Alt}(q)$.

$$p_k = P(\text{první zdar nastal právě v } k\text{-tém pokusu}) = q^{k-1}p$$

$$p_k^* = P(\text{kolik nezdarů bylo před prvním zdarem}) = q^k p$$

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} p \cdot q^{k-1} = p \cdot \frac{1}{1-q} = 1 \text{ (ověřili jsme, že jde o rozdělení pravděpodobnosti)}$$

$$EX = \sum_{k=0}^{\infty} k \cdot p \cdot q^{k-1} \text{ (} p \text{ lze vytknout před sumu)}$$

$$A(x) = \sum_{k=1}^{\infty} p \cdot q^{k-1} \cdot x^k = \frac{p}{q} \cdot \sum_{k=1}^{\infty} (qx)^k = \frac{p}{q} \cdot \frac{1}{1-qx}$$

$$EX = A'(1)$$

$$A^*(x) = \sum_k pq^k x^k = \frac{p}{1-qx}$$

Poznámka: Geometrické rozdělení nemá paměť (\Rightarrow pokusy jsou nezávislé).

$$P(X > x + y | X > y) = P(X > x)$$

Je to jediné takové diskrétní rozdělení!!!

4.2 Negativně binomické rozdělení

Dělám alternativní nezávislé pokusy.

$p_k = P(r\text{-tý zdar nastal právě v } k\text{-tém pokusu}) = \binom{k-1}{r-1} \cdot p^r \cdot q^{k-r}$ (v posloupnosti délky $k-1$ musí být libovolně rozmístěno $r-1$ úspěchů, navíc je jeden úspěch na konci, zbytek jsou neúspěchy)

$$p_k^* = P(\text{před } r\text{-tým zdarem je právě } k \text{ nezdarů})$$

$$\text{NBi}(r, p): W_r = \sum_{i=1}^r Z_i, \text{ kde } Z_i \text{ jsou nezávislé a } Z_i \sim \text{Ge}(p). \text{ Pak } EX_r = r \cdot EZ_1 \text{ a } \text{var}W_r = r \cdot \text{var}Z_1.$$

Na rozmyšlení: Čekání na jev, že zdar padne třikrát za sebou.

4.3 Hypergeometrické rozdělení

Mějme 49 čísel. Z toho vytáhneme 6. Jaká je pravděpodobnost, že uhádnou právě 3 čísla?

$$P(A) = \frac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}}$$

Provádím tahy bez vracení.

Koule $\begin{cases} \text{bílé} & A \\ \text{černé} & B \end{cases}$

Ke kontrole vytáhnou n koulí. Zajímá mně pravděpodobnost jevu α = mezi vytaženými je právě i černých:

$$P(\alpha) = \frac{\binom{A}{n-i} \cdot \binom{B}{i}}{\binom{A+B}{n}}$$

Koule $\begin{cases} \text{bílé} & A \\ \text{černé} & B \\ \text{červené} & C \end{cases}$

Ke kontrole vytáhnou n koulí. Zajímá mě pravděpodobnost jevu β = mezi vytaženými koulemi je i bílých, j černých a $n-i-j$ červených koulí:

$$P(\alpha) = \frac{\binom{A}{i} \cdot \binom{B}{j} \cdot \binom{C}{n-i-j}}{\binom{A+B+C}{n}}$$

Toto pravidlo lze snadno zobecnit pro k typů koulí.

Jak si představit $\text{HPG}(A, B, n)$ jakou součet náhodných veličin?

- (I) Vezmu vytažené koule a každé se zeptám, zda je bílá. Výsledek pro i -tou kouli označíme X_i . Toto X_i je s pravděpodobností p rovné 1 a s pravděpodobností $1-p$ rovné 0. **DŮLEŽITÉ:** X_i nejsou nezávislé!

$$E(X_1 + \dots + X_n) = n \cdot EX_1 = n \cdot \frac{A}{A+B}$$

- (II) Vezmu bílé koule a zeptám se každé z nich, jestli byla vytažena. Pro i -tou kouli dostávám výsledek X_i^* , nabývající hodnot 0 nebo 1.

náhodná veličina je $\sum_{i=1}^A X_i^*$

$p^* = P(\alpha)$, kde α = i -tá koule byla vytažena

$$\sum_i \frac{\binom{A}{i} \cdot \binom{B}{n-i}}{\binom{A+B}{n}} = E\text{HPG}(A, B, n)$$

$$E\left(\sum_{i=1}^A X_i^*\right) = A \cdot EX_1^*$$

$$\begin{aligned}\text{var} \left(\sum_{i=1}^n x_i \right) &= \sum_{i=1}^n \text{var} X_i + \sum_{i \neq j} \text{cov} (X_i, X_j) \\ \text{cov} (X, Y) &= E(X - EX)(Y - EY) = EXY - EXEY \\ X, Y \text{ nezávislé} &\Rightarrow \text{cov} (X, Y) = 0 \text{ (neplatí naopak)} \\ \text{var} \left(\sum_{i=1}^n X_i \right) &= n \cdot \text{var} X_1 + 2 \cdot \binom{n}{2} \cdot \text{cov} (X_1, X_2) \\ \text{cov} (X_1, X_2) &= EX_1 X_2 - EX_1 EX_2, \text{ kde } EX_1 EX_2 \text{ znám (je to malé záporné číslo, tedy} \\ &\text{rozptyl je menší, než u binomického rozdělení) a } X_1 \cdot X_2 \text{ nabývá hodnot 0 nebo 1, tedy} \\ &EX_1 X_2 = 0 \cdot P(X_1 = 0 \vee X_2 = 0) + 1 \cdot P(X_1 = 1 \wedge X_2 = 1)\end{aligned}$$

Výhoda: Vedle definic stačí umět spočítat dvě pravděpodobnosti.

$$\frac{\binom{A}{i} \cdot \binom{B}{n-i}}{\binom{A+B}{n}} = \frac{\frac{A!}{i!(A-i)!} \cdot \frac{B!}{(n-i)!(B-n+i)!}}{\frac{(A+B)!}{n!(A+B-n)!}} = \binom{n}{i} \cdot \prod_{j=0}^{i-1} \frac{A-j}{A+B-j} \cdot \prod_{j=0}^{n-i-1} \frac{B-j}{A+B-j-i}, \text{ kde}$$

$$\frac{A-j}{A+B-j} \approx p \text{ a } \frac{B-j}{A+B-j-i} \approx 1-p, \text{ tedy můžeme upravit na } \binom{n}{i} p^i (1-p)^i.$$

DCV: Zkuste si představit (modelovat) HPG (A, B, n) jako náhodnou procházku v 1. kvadrantu.

4.4 Poissonovo rozdělení

$X \dots 0, 1, 2, \dots$

$$p_i = P(X = i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

$$\sum_i p_i = \sum_i x^i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \cdot \sum_i \frac{(x\lambda)^i}{i!} = e^{-\lambda} \cdot e^{x\lambda} = e^{\lambda(x-1)}$$

$$X_1 \sim P_0(\lambda_1) \quad e^{\lambda_1(x-1)}$$

$$X_2 \sim P_0(\lambda_2) \quad e^{\lambda_2(x-1)}$$

$$X_1 + X_2 \sim e^{(\lambda_1 + \lambda_2)(x-1)}$$

Věta: Součet dvou náhodných veličin s Poissonovým rozdělením má zase Poissonovo rozdělení s parametrem rovným součtu parametrů jednotlivých členů.

Příklad: Do rozlišitelných přihrádek dáváme náhodně rozlišitelné kuličky. Jestliže pošleme počet kuliček i počet přihrádek do nekonečna, dostaneme Poissonovo rozdělení. Budou-li kuličky nerozlišitelné, pak je limitní rozdělení geometrické.

4.5 Rozdělení rovnoměrné (diskrétní)

X nabývá hodnot $0, 1, \dots, n$ a platí $P(X = i) = \frac{1}{n+1}$. Další variantou je X nabývá hodnot

$1, 2, \dots, n$, potom ale $P(X = i) = \frac{1}{n}$.

Myšlenkou tohoto rozdělení je: Nevím-li nic, dám všem tu stejnou šanci nastat.

$$EX = \sum_{i=0}^n i \cdot \frac{1}{n+1} = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = \frac{n}{2}$$

$$\text{var} X = \sum_{i=0}^n \left(i^2 \cdot \frac{1}{n+1} \right) - \left(\frac{n}{2} \right)^2 = \frac{n(n+1)(2n+1)}{6(n+1)} \cdot \left(\frac{n}{2} \right)^2$$

4.6 Náhodná procházka

Definice: Necht' X_1, X_2, \dots jsou náhodné veličiny (obecně libovolné). Dále necht' $S_0 = z \in \mathbb{R}_1$ a $S_k = \sum_{i=0}^k X_i$. Potom $\{S_k : k \in \mathbb{N}_0\}$ nazveme náhodnou procházkou.

Poznámka: My budeme zkoumat ty nejjednodušší, kdy $X_i = \begin{cases} 1 & p \\ -1 & 1-p \end{cases}$. Pro úplnou jednoduchost $p = \frac{1}{2}$.

X_i jsou nezávislé, $S_0 = 0$, S_k je stav naší kapsy po k hrách

$$S_k \in [-k, k] \cap \mathbb{Z}$$

$$EX_i = 0$$

$$\text{var} X_i = EX_i^2 - (EX_i)^2 = 1 \rightarrow (\forall k) ES_k = 0$$

$$S_k^* = \frac{1}{\sqrt{k}} \cdot S_k$$

$$\text{var} S_k^* = \frac{1}{k} \cdot k = 1$$

Zdá se

- (1) Zvětšuje se amplituda S_k .
- (2) Doby mezi návraty do nuly se prodlužují.

$$\limsup_{k \rightarrow \infty} S_k^* = +\infty$$

$$\liminf_{k \rightarrow \infty} S_k^* = -\infty$$

$$\textbf{ZIL:} \text{ Platí } \limsup_{k \rightarrow \infty} \frac{1}{\sqrt{2k \cdot \log \log k}} \cdot S_k^* = 1 \text{ a } \liminf_{k \rightarrow \infty} \frac{1}{\sqrt{2k \cdot \log \log k}} \cdot S_k^* = -1.$$

5 Čebyševova nerovnost a Čebyševova věta

Věta (Čebyševova nerovnost): Necht' náhodná veličina X má $EX = \mu$ a $\text{var} X = \sigma^2 < \infty$. Potom $P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$.

Věta (Čebyševova věta): Necht' náhodné veličiny X_1, X_2, \dots jsou nezávislé stejně rozdělené a necht' $EX_1 = \mu$, $\text{var} X_1 = \sigma^2 < \infty$. Potom $P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$. Pro $n \rightarrow \infty$ dostáváme $\frac{\sigma^2}{n \cdot \varepsilon^2} \rightarrow 0$.

Význam: Průměr nezávislých stejně rozdělených náhodných veličin je z hlediska vzdálenosti měření pravděpodobností libovolně blízko své střední hodnotě.

$$\textbf{Důkaz (věta):} \quad E \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \cdot n \cdot EX_1 = EX_1 = \mu$$

$$\text{var} \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n^2} \sum_{i=1}^n \text{var} X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \quad \square$$

$$\text{Důkaz (nerovnost): } \text{var}X = E(X - EX)^2 = \sum_k p_k \cdot (x_k - EX)^2 \geq \sum_{k||x_k - EX| \geq \varepsilon} p_k \cdot \varepsilon^2 =$$

$$\varepsilon^2 \cdot \sum_{k||x_k - EX| \geq \varepsilon} p_k = P(|X - EX| \geq \varepsilon)$$

$$g: X \rightarrow (X - EX)^2$$

$$Eg(X) = \sum_i g(x_i)p_i \quad \square$$

Poznámka: Důkaz ukazuje, že jsme zvolili co možná nejjednodušší ohraničení (useknutí) \Rightarrow nepříliš užitečné v praxi.

$$X_i \sim \text{Alt}(p = \frac{1}{2}); X_1, \dots, X_{1000}$$

$$P\left(\left|\sum_{i=1}^{1000} X_i - 500\right| \geq 10\right) \leq \frac{250}{100} = 2.5$$

$$P\left(\left|\sum_{i=1}^{1000} X_i - 500\right| \geq 100\right) \leq \frac{250}{10000}$$

Poznámka: Existují četná zobecnění.

Věta: Nechť X_1, X_2, \dots jsou nezávislé náhodné veličiny takové, že $EX_i = \mu_i$, $\text{var}X_i = \sigma_i^2$

a $\exists c \in \mathbb{R} \sigma_i^2 \leq c$. Potom $P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \sum_{i=1}^n EX_i\right| \geq \varepsilon\right) \leq \frac{\text{var} \frac{1}{n} \sum_{i=1}^n X_i}{\varepsilon^2} \leq \frac{c}{n\varepsilon^2} \rightarrow 0$ pro n jdoucí k ∞ .

$$\text{var} \frac{1}{n} \sum X_i = \frac{1}{n^2} \sum \text{var}X_i \leq \frac{n \cdot c}{n^2}$$

$$\text{var} \sum_{i=1}^n X_i = \sum \text{var}X_i + \sum \sum_{i \neq j} \text{cov}(x_i, x_j)$$

Aplikace: Simulace a oprávnění k nim.

$$\text{Pokus je } X_i \sim \text{Alt}(p = \frac{\pi}{4}). X_i \begin{cases} 0 & 1 - \frac{\pi}{4} \\ 1 & \frac{\pi}{4} \end{cases}$$

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n I[\text{zasáhl cíl}]$$

$$\frac{\#Z}{n}; pokus_i \leftrightarrow \text{Alt}(\cdot)_i$$

Otázkou je, kolik udělat pokusů abych byl dostatečně blízko s dosti velkou pravděpodobností.

$$P(|X_n - \frac{\pi}{4}| \leq 10^{-xx}) \geq 0,9 \dots 9, \text{ kde } 10^{-xx} \text{ je přesnost a } 0,9 \dots 9 \text{ je spolehlivost}$$

Platí: $p(|X - EX| \leq \varepsilon) \geq 1 - \frac{\text{var}X}{\varepsilon^2}$ (Čebyševova nerovnost v doplňkovém tvaru)

$$X \dots \mu$$

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \left(\sum_{i=1}^n X_i + X_n \right) = \frac{n-1}{n} \overline{X_{n-1}} + \frac{1}{n} X_n$$

Rostoucí n znamená rozšiřování délky simulace. Pro k platných číslic potřebuju $c \cdot 10^{2k}$ pokusů, kde c je nějaká konstanta.

6 Spojitá náhodná veličina

Do teď: $\left\{ \begin{matrix} x_i \\ p_i \end{matrix} \right\}_i$

$\text{card } \{x_i\}$ je spočetná

$X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}_1, \mathcal{B}_1)$ (zde $p_i \geq 0$ a $\sum_i p_i = 1$)

$F(x) = P(X \leq x) \quad \forall x \in \mathbb{R}_1$

$EX = \sum_{\omega} X(\omega)P(\omega) \equiv \sum_i x_i p_i$

$\text{var} X = E(E - EX)^2 = EX^2 - (EX)^2$

Jestliže $g : \mathbb{R}_1 \rightarrow \mathbb{R}_1$ je funkce, pak $g(X)$ je také náhodná veličina a $Eg(X) = \sum_i g(x_i)p_i$.

Vytvořující funkce je $A(x) = \sum_i p_i x^i$.

Nyní: $\Omega \subseteq \mathbb{R}_1$, \mathcal{A} jsou intervaly a nad nimi je cosi, čemu se říká σ -algebra uzavřená na spočetné průniky a sjednocení, tedy jsou-li $\{A_i\}_i$ disjunkttní, pak $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$.

Bude nás zajímat $P(a < X \leq b)$.

Definice: Náhodná veličina je definována jako $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}_1, \mathcal{B}_1)$.

Základní charakteristikou bude distribuční funkce.

Definice: Distribuční funkci náhodné veličiny X definujeme jako $F(x) = P(X \leq x)$.

Vlastnosti: $F(x)$ je neklesající a zprava spojitá, $P(X < -\infty) = 0$, $P(X < +\infty) = 1$.

Definice: Hustotou náhodné veličiny X rozumíme nezápornou funkci $f(x)$, která splňuje $\int_{\mathbb{R}_1} f(x)dx = 1$.

Poznámka: Hned vidíme, že $F(x) = \int_{-\infty}^x f(t)dt$.

Poznámka: Hustota popisuje přírůstek: $\int_x^{x+\Delta} f(t)dt = F(x+\Delta) - F(x) \approx \Delta \cdot f(x)$. Tedy $f(x) \approx \frac{F(x+\Delta) - F(x)}{\Delta}$. Tedy pro $\Delta \rightarrow 0$ dostáváme $f(x) \approx F'(x)$.

Úprava definice distribuční funkce:

Definice: Mějme nezápornou funkci $f(x)$ takovou, že $\int_{\mathbb{R}_1} f(x)dx = 1$. Potom se jedná o hustotu některé náhodné veličiny X .

Poznámka: Podstatná je distribuce pravděpodobnosti (zde vyjádřené přes distribuční funkci) a k ní si potom odvodím jinou charakteristiku, například hustotu.

Definice: Charakteristická funkce je definována jako $\varphi(t) = e^{itX}$.

Pozor: Charakteristická funkce má všechny hezké vlastnosti jako funkce vytvořující, ale matematicky se dostanu do komplexního světa.

Poznámka: V praxi častěji pracujeme s hustotami než s distribučními funkcemi. Lépe vidíme co se děje.

Poznámka: Většina důležitých rozdělení je (podobně jako v diskrétním světě) parametrizována. Typicky 1 - 3 parametry, někdy i 4. Význam parametrů je vztažen (často, ne vždy), ke střední hodnotě, rozptylu, posunutí ...

Definice: Nechť X je náhodná veličina, $F(X)$ je její distribuční funkce a $f(x)$ její hustota. Potom:

- $EX = \int_{\mathbb{R}_1} x \cdot f(x) dx$
- $Eg(X) = \int_{\mathbb{R}_1} g(x) f(x) dx$
- $\text{var} X = E(X - EX)^2 = EX^2 - (EX)^2 = \int_{\mathbb{R}_1} x^2 f(x) dx - \left(\int_{\mathbb{R}_1} x f(x) dx \right)^2$
- $\text{cov}(X, Y) = E(X - EX)(Y - EY) = EXY - EXEY = \int_{\mathbb{R}_1} \int_{\mathbb{R}_1} x \cdot y \cdot f(x, y) dx dy$

Poznámka: Všechny vlastnosti střední hodnoty a rozptylu zůstanou zachovány.

Definice: Mějme náhodné veličiny X_1, \dots, X_n . $P(X_1 \leq x_1, \dots, X_n \leq x_n)$ nazýváme sdruženou distribuční funkcí vektoru (x_1, \dots, x_n) .

6.1 Příklady rozdělení

6.1.1 Rovnoměrné

Značíme $R(0, 1)$.

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \in (0, 1) \\ 0 & x \geq 1 \end{cases}$$

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & x \in (0, 1) \\ 1 & x \geq 1 \end{cases}$$

$$EX = \frac{1}{2}, \text{var} X = \frac{1}{12}$$

Věta: Nechť náhodná veličina má distribuční funkci $F(x)$ a inverzní distribuční funkci $F^{-1}(t)$. Dále nechť $U \sim R(0, 1)$. Potom $Z = F^{-1}(U) \sim F(x)$.

Důkaz: $P(z \leq x) = P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x) \square$

Význam: Libovolnou distribuční funkci a tím i libovolnou odpovídající náhodnou veličinu dostanu touto transformací. Je to univerzální transformace.

Užití: V simulacích. Umím-li generovat náhodnou veličinu z $R(0, 1)$, pak umím generovat jakoukoliv jednorozměrnou náhodnou veličinu, diskrétní i spojitou.

Problém: Jak najít $F^{-1}(t)$?

Řešení:

- (1) Zkusím najít z rovnosti $F(x) = t$:

$$F(x) = 1 - e^{-\mu x} = t$$

$$1 - t = e^{-\mu x}$$

$$\log(1 - t) = -\mu x$$

$$-\frac{1}{\mu} \log(1 - t) = x$$

- (2) Numericky vyřeším rovnici $F(x) = t$.

- (3) Místo s $F^{-1}(t)$, kterou neumím vyjádřit explicitně, pracuji s $F_*^{-1} = \begin{cases} \sup \{x \mid F(x) \leq t\} \\ \inf \{x \mid F(x) \geq t\} \end{cases}$

Pokud $F^{-1}(t)$ existuje, pak $F^{-1}(t) \equiv F_*^{-1}(t)$.

Definice: Obecně rovnoměrné rozdělení je dáno následovně:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$$

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & x \in (a, b) \\ 1 & x \geq b \end{cases}$$

6.1.2 Exponenciální rozdělení

Může být zadáno následujícími způsoby:

$$(1) f(x, \lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0, \lambda > 0 \\ 0 & \text{jinak} \end{cases}$$

$$(2) f(x, \mu) = \begin{cases} \frac{1}{\mu} \cdot e^{-\frac{x}{\mu}} & x \geq 0, \mu > 0 \\ 0 & \text{jinak} \end{cases}$$

Distribuční funkce:

$$(1) F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

$$(2) F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\frac{x}{\mu}} & x > 0 \end{cases}$$

Střední hodnota:

$$(1) EX = \frac{1}{\lambda}$$

$$(2) EX = \mu$$

Rozptyl:

$$(1) \text{var}X = \frac{1}{\lambda^2}$$

$$(2) \text{var}X = \mu^2$$

Toto rozdělení nemá paměť a je to jediné takové rozdělení.

To znamená, že $P(X > x + y | X > x) = P(X > y)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(X > x + y | X > x) = \frac{P(X > x + y)}{P(X > x)} = \frac{1 + (1 + e^{-\lambda(x+y)})}{1 + (1 + e^{-\lambda x})} = e^{-\lambda y} = P(X > y)$$

Poznámka: Interpretujme X jako délku života. Potom $P(X > x) = P(\text{pacient přežije dobu větší jak } X)$.

Poznámka: Analogií v diskrétním světě je rozdělení $Ge(p)$.

Připomenutí: exponenciální rozdělení

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0, \lambda > 0 \\ 0 & \text{jinak} \end{cases}$$

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

$$EX = \frac{1}{\lambda}; \text{var}X = \frac{1}{\lambda^2}$$

(1) nemá paměť: X je nejčastěji interpretována ve spolehlivosti, přežívání atd. jako doba života výrobku, tedy $P(X > x + y | X > y) = P(X > y)$.

(2) Pravděpodobnost okamžitého selhání:

$$P(x < X \leq x + \Delta | X > x) = \frac{P(x < X \leq x + \Delta \cap X > x)}{P(X > x)} = \frac{P(x < X \leq x + \Delta)}{1 - P(X \leq x)} =$$

$$\frac{F(x + \Delta) - F(x)}{1 - F(x)} = \Delta \cdot \frac{\frac{F(x + \Delta) - F(x)}{\Delta}}{1 - F(x)} \approx \Delta \cdot \frac{f(x)}{1 - F(x)}$$

Definice: Mějme náhodnou veličinu X s distribuční funkcí $F(x)$ a hustotou $f(x)$. Potom intenzitou (poruch) rozumíme funkci $\Lambda(x) = \frac{f(x)}{1 - F(x)}$.

Poznámka: Pro exponenciální rozdělení je $\Lambda(x) = \frac{\lambda e^{-\lambda x}}{1 - (1 - e^{-\lambda x})} = \lambda$, neboli je konstatní.

Poznámka: Na intenzitu (poruch) se často můžeme dívat jako na převrácenou střední hodnotu (při interpretaci).

Poznámka: Konstatní intenzita je pouze pro exponenciální rozdělení. Klesající intenzita je sen každého zákazníka. Rostoucí intenzita je běžnou realitou. Vanovitá intenzita (v nízkém věku vysoká pravděpodobnost, ve středním nízká a ve vysokém zase vysoká).

6.1.3 Normální (Gaussovo) rozdělení

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x - \mu)^2}{2\sigma^2}} \text{ pro } x \in \mathbb{R}_1, \mu \in \mathbb{R}_1 \text{ a } \sigma^2 > 0$$

$$EX = \mu; \text{var}X = \sigma^2$$

$$\text{Pro standartní } N(0, 1) \text{ platí } f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Jestliže $X \sim N(0, 1)$, pak $\mu + \sigma \cdot X \sim N(\mu, \sigma^2)$.

Naopak, jestliže $X \sim N(\mu, \sigma^2)$, pak $\frac{X - \mu}{\sigma} \sim N(0, 1)$.

Platí: Nechť $X \sim N(\mu_X, \sigma_X^2)$ a $Y \sim N(\mu_Y, \sigma_Y^2)$ jsou nezávislé, pak $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Tedy sčítání zachovává typ distribuce.

Poznámka: Mezi podstatnými rozděleními (jež známe) má tuto vlastnost $Bi(n, p)$, $Po(\lambda)$ a normální rozdělení.

6.2 Centrální limitní věta

Věta (Moivre-Laplace): Nechť X_1, X_2, \dots jsou nezávislé náhodné veličiny s $\text{Alt}(p)$, $p > 0$.

Potom $P\left(\frac{\sum_{i=1}^n X_i - np}{\sqrt{n \cdot p \cdot (1-p)}} \leq x\right) \approx \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt \equiv P(N(0,1) \leq x) \equiv$ distribuční funkce

standartního normálního rozdělení, pro $n \gg 0$.

Poznámka: $E \sum_{i=1}^n X_i = n \cdot EX_i = n \cdot p$

$\text{var} \sum_{i=1}^n X_i = n \cdot p \cdot (1-p)$

Věta (Centrální limitní věta): Nechť X_1, X_2, \dots jsou nezávislé náhodné veličiny takové, že $EX_i =$

μ a $\text{var} X_i = \sigma^2 < \infty$, stejně rozdělené. Potom $P\left(\frac{\sum_{i=1}^n X_i - E \sum_{i=1}^n X_i}{\sqrt{\text{var} \sum_{i=1}^n X_i}} \leq x\right) \approx \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt,$

pro $n \gg 0$.

Otázka: Co to znamená $n \gg 0$? Jak velké n ?

Odpověď: Aproximace je velmi dobrá již pro desítky porovnání!!! \square

Lidé jsou zvyklí průměrovat. Zde je hlavní reálná aplikace.

Hlavní předpoklad je nezávislost.

Rozdělení součtu náhodné veličiny vždy mohou dostat přes konvoluci $\sum_{j=0}^i p_j \cdot q_{i-j}$. Centrální limitní věta je jednodušší.

Možná aplikace (dříve dost používaná): Pro simulaci náhodných čísel z normálního rozdělení:

- Nagenervuj 12 čísel z $R(0,1)$ a označ U_i .

- Vrať $\frac{\sum_{i=1}^{12} 2U_i - 6}{1}$. Platí $EU_i = \frac{1}{2}$ a $\text{var} U_i = \frac{1}{12}$.

Aplikace: V simulacích určit počet simulací potřebných pro výpočet s danou přesností a spolehlivostí. Mějme jev A , který nastává s pravděpodobností $P(A) = p$. Dále mějme náhodnou veličinu $I_A \sim \text{Alt}(p)$. Chceme (náhodný) pokus zopakovat vícekrát (\equiv simulovat) a neznámou hodnotu $P(A)$ odhadnout.

$$T_n = \frac{\sum_{i=1}^n I_A^i}{n}$$

$$ET_n = \frac{1}{n} \sum_{i=1}^n EI_A^i = \frac{1}{n} \cdot n \cdot P(A) = P(A)$$

Z Čebyševovy věty (zákona velkých čísel) dostáváme $P(|T_n - P(A)| \leq \varepsilon) \geq 1 - \delta$. Zde ε značí

přesnost a $1 - \delta$ značí spolehlivost. Oboje volíme MY!

Klasická aplikace: Výpočet integrálů. Obecně dostaneme omezenou funkci $g(x)$ a chceme spočítat $\int_a^b g(x)dx$ pro $-\infty < a, b < \infty$. Tento integrál převedeme na $K \int_0^1 h(x)dx$, pro který platí: $\forall x \in (0,1) : 0 \leq h(x) \leq 1$. Nyní stačí nahradit výpočet integrálu BMC (Brute Monte Carlo) algoritmem:

```
begin
  S:=0;
  for i:=1 to OPAK do begin
    nagenereuj X a Y nahodne velicity z R(0,1);
    if Y <= h(X) then inc(S);
  end;
  return S/OPAK;
end.
```

Př. 11.
14. 12. 2009

Příklad: Mějme funkci $h(x)$, která na intervalu $[0,1]$ nabývá hodnot z intervalu $[0,1]$. Chceme spočítat $\int_a^b h(x)dx$, tedy plochu pod grafem funkce $h(x)$. Označme M množinu všech bodů pod grafem $h(x)$ a označme Ω dvourozměrný interval $[0,1]^2$. Pak platí $P(M) = \int_a^b h(x)dx$. $P(M)$ lze počítat algoritmem Brute Monte Carlo, viz výše.

Ale kolik náhodných bodů musím vzít, aby byla aproximace dostatečně přesná?

Zvolme $(X_i, Y_i) \in [0,1]^2$. Označme $Z_i = I_{[(X_i, Y_i) \in M]} = \begin{cases} 0 & 1 - P(M) \\ 1 & P(M) = p \end{cases}$. Potom tedy dostáváme $p = \frac{1}{n} \sum_{i=1}^n Z_i = \overline{Z_n}$. Navíc $\text{var} \overline{Z_n} = \frac{p(1-p)}{n}$. Nyní chceme spočítat $P(|\overline{Z_n} - p| \leq \varepsilon) \geq 1 - \delta$. To lze udělat následujícími způsoby:

(I) Čebyševova nerovnost: $P(|X - EX| \geq \varepsilon) \leq \frac{\text{var} X}{\varepsilon^2}$

$$P(|X - EX| \leq \varepsilon) \geq 1 - \frac{\text{var} X}{\varepsilon^2}$$

$$X \dots \overline{Z_n}; EX = p$$

$$P(|\overline{Z_n} - p| \leq \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2} = 1 - \delta$$

Pak tedy dostáváme $n \approx \frac{p(1-p)}{\delta\varepsilon^2}$. Zde nastává problém, že n závisí na p , kterou ovšem neznáme. Pokud nevím vůbec nic, zvolím nejhorší variantu, v tomto případě $p = \frac{1}{2}$. Potom dostaneme $n \approx \frac{1}{4\delta\varepsilon^2}$. Pro $\delta = 0.05$ dostávám $n \approx \frac{5}{\varepsilon^2}$. Pro $\delta = 0.01$ dostávám $n \approx \frac{25}{\varepsilon^2}$.

(II) Centrální limitní věta: $\left| \frac{1}{n} \sum Z_i - p \right| \leq \varepsilon$

$$\begin{aligned} \left| \frac{\sum Z_i - p}{np} \right| &\leq \varepsilon \\ \left| \sum Z_i - np \right| &\leq n \cdot \varepsilon \\ \left| \frac{\sum Z_i - np}{\sqrt{n \cdot p \cdot (1-p)}} \right| &\leq \frac{\varepsilon \cdot n}{\sqrt{n \cdot p \cdot (1-p)}} \\ -\frac{\varepsilon n}{\sqrt{npq}} &\leq \frac{\sum Z_i - np}{\sqrt{n \cdot p \cdot (1-p)}} \leq \frac{\varepsilon n}{\sqrt{npq}} \end{aligned}$$

$A = \frac{\varepsilon n}{\sqrt{npq}} \Rightarrow n = \frac{A^2 p(1-p)}{\varepsilon^2}$, kde A najdu v tabulkách nebo spočtu.

Pro $\delta = 0.05$ a $p = \frac{1}{2}$ dostáváme $n \approx \frac{1}{\varepsilon^2}$.

Pro $\delta = 0.01$ a $p = \frac{1}{2}$ dostáváme $n \approx \frac{1.65}{\varepsilon^2}$.

7 Simulace náhody na počítačích

Základní stavební kámen je $R(0, 1)$.

Nejdříve se náhodná čísla brala z telefonního seznamu: Vezmu třetí až sedmou pozici telefonního čísla a to bude desetinná část náhodného čísla.

Později von Neuman navrhl, že stačí vzít deterministický algoritmus a předpis, co od náhody čekáme. Říkejme tomu pseudonáhoda.

Možným algoritmem je: $X_{n+1} = (a \cdot X_n + b) \bmod M$ + lineární transformace na interval $[0, 1)$.

Co od algoritmu chceme:

- rychlý
- přenositelný mezi platformami
- čísla nezávislá
- měl frekvenčně danou distribuci
- nedekódovatelný
- případné další požadavky

Náhoda se smrskne na baterii testů.

Generovat z nějaké distribuce: $X \dots f(x), F(x), F^{-1}(t)$, exponenciální, tedy $f(x) = e^{-\lambda x}$

Výsledkem budou čísla $x_i \geq 0$, která mají následující strukturu: krátká doba obsluhy bude mnohem pravděpodobnější, než dlouhá doba obsluhy.

Nechť $X \sim \text{Alt}(p)$.

0, 1, 0, 0, 1, 0, 0, 0, 1, 1, ...

0, 1, 0, 1, 0, 1, 0, 1, 0, 1, ...

Druhý řádek jsme asi nechtěli, i když je přesný pro $p = \frac{1}{2}$.

7.1 Inverzní metoda

$$X \sim F(x), F^{-1}(t)$$

$$R \sim R(0, 1)$$

$$Z = F^{-1}(R) \sim F$$

$$F(x) = 1 - e^{-\lambda x}$$

$$F^{-1}(t) = -\frac{1}{\lambda} \log(1 - t)$$

$$1 - e^{-\lambda x} = t$$

$$1 - t = e^{-\lambda x}$$

$$\log(1 - t) = -\lambda x$$

$$R \sim R(0, 1): P(1 - R \leq x) = P(R > 1 - x) = 1 - P(R \leq 1 - x)$$

$$1 - R \sim R(0, 1): P(1 - R \leq x) = 1 - (1 - x) = x$$

7.2 Zamítací metoda

$$f(x) \dots x \in [a, b]; 0 \leq f(x) \leq c$$

Zvolme $X, Y \in R(0, 1)$. Označme $X^* = a + (b - a) \cdot X$ a $Y^* = c \cdot Y$.

$$R(a, b) \dots g(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{jinak} \end{cases} \Rightarrow \exists M f(x) \leq M g(x)$$

Vezmu dvojici (X^*, Y^*) , kde X^* je potenciální kandidát. Potom Y^* je náhodná veličina, podle které se rozhodnu zda kandidáta přijmu, či ne.

Tahám $X \sim g(x)$ a $R \sim R(0, 1)$.

Př. 12.
4. 1. 2010

8 Rozhodování mezi dvěma tvrzeními

Kamarád přinesl minci. Potom

Hypotéza H mince je dokonalá $\equiv p = \frac{1}{2}$

Alternativa A mince je falešná $\equiv \{p > \frac{1}{2}, p < \frac{1}{2}, p \neq \frac{1}{2}, p = 0.4\}$

8.1 Sekvenční přístup

Realita: Uděláme n hodů a na jejich základě se rozhodneme. Cca v $\frac{n}{2}$ případech padne zdar a v $\frac{n}{2}$ případech padne nezdar. Pravděpodobnost, že po $2n$ hodech padne přesně n zdarů, je asi $\frac{1}{\sqrt{\pi n}}$.

Kdy kolegu nařknu, z podvodu? Když bude zdarů příliš moc nebo příliš málo.

$$X_i \sim \text{Alt} \left(p = \frac{1}{2} \right)$$

$$Y = \sum_{i=1}^{1000} X_i \sim \text{Bi} \left(1000, \frac{1}{2} \right)$$

$$EY = 500, \text{var}Y = 250$$

V intervalu $(500 - 3\sqrt{250}, 500 + 2\sqrt{250})$ by měla ležet většina výsledků.

	H	A
H	OK	KO ₁
A	KO ₂	OK

(sloupec značí realitu, řádek rozhodnutí)

Naším cílem je minimalizovat pravděpodobnost chybných rozhodnutí.

Dohoda: Stanovíme rozhodovací pravidlo na základě chyby KO₁, tzv. chyby prvního druhu (budeme značit α). Pravděpodobností chyby KO₂ měříme kvalitu naší procedury.

$$\alpha = P(H \text{ zamítneme} | H \text{ platí}) = P \left(\sum_{i=1}^n X_i > C \middle| p = \frac{1}{2} \right) = P \left(\text{Bi} \left(n, \frac{1}{2} \right) \right)$$

$$Y_n = \sum_{i=1}^n X_i \sim \text{Bi}\left(n, \frac{1}{2}\right)$$

$$\Rightarrow P(Y_n > C) = \alpha \sim P(\text{Bi}(n, \frac{1}{2}) > C) = \alpha$$

$P(Y_n \leq C_\alpha) = 1 - \alpha$... největší C_α takové, aby rovnost platila

$P(Y_n > C)$ je distribuční funkce Y_n , ta má skoky velikosti $\binom{n}{i} p^i (1-p)^{n-i}$, $n = 0, 1, \dots, n$

Interpretace α : Ze sta dokonalých mincí **neoprávněně** 5 kusů zamítnu a 95 přijmu.

$$\alpha = 0.05 \quad 526$$

$$\alpha = 0.01 \quad 538$$

8.2 Přístup přes CLV

$$\begin{aligned} \alpha &= P\left(\sum_{i=1}^n X_i > C \mid p = \frac{1}{2}\right) = P(Y_n > C) = 1 - P\left(\frac{Y_n - EY_n}{\sqrt{\text{var}Y_n}} \leq \frac{C - EY_n}{\sqrt{\text{var}Y_n}}\right) = \\ &= 1 - P\left(N(0, 1) \leq \frac{C - \frac{n}{2}}{\sqrt{\frac{n}{4}}}\right) \equiv P(N(0, 1) \leq K_\alpha) = 1 - \alpha \end{aligned}$$

$Y_n \sim \text{Bi}(n, \frac{1}{2})$, $EY_n = \frac{n}{2}$, $\text{var}Y_n = \frac{n}{4}$, navíc z CLV víme $\frac{Y_n - EY_n}{\sqrt{\text{var}Y_n}} \sim N(0, 1)$

$$K_\alpha = \frac{C - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \Rightarrow C = \frac{n}{2} + K_\alpha \frac{\sqrt{n}}{2}$$

K_α najdu v tabulkách nebo spočítám programem případně kalkulačkou.

$$\alpha = 0.05 \quad K_\alpha = 1.64 \quad 525.93$$

$$\alpha = 0.01 \quad K_\alpha = 2.33 \quad 536.84$$

Dohoda: Většinou se snažíme hypotézu zamítnout.

$p \neq \frac{1}{2} \Rightarrow p < \frac{1}{2} \vee p > \frac{1}{2}$, pro oba případy jsme schopní najít rozhodovací pravidlo

8.2.1 Oboustranná varianta

$$\alpha = \alpha_1 + \alpha_2: \begin{array}{ll} \alpha_1 > 0 & \alpha_1 \dots C_{\alpha_1} \\ \alpha_2 > 0 & \alpha_2 \dots C_{\alpha_2} \end{array}$$

a zamítneme tehdy, když buď $\sum X_i < C_{\alpha_i}$ nebo $\sum X_i > C_{\alpha_i}$. Nemám-li explicitní důvod se bát jednotlivých (pod)alternativ, volím $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

8.3 Kvalita postupu

Alternativu si mohu představit jako všechna $p > \frac{1}{2}$. K hypotéze $p = \frac{1}{2}$ jsem našel C_α .

silofunkce \equiv doplněk chyby druhého druhu bráno jako funkce parametru (zde p)

Problém: Jak obecně zvolit funkci pozorování, na jejímž základě chceme testovat. Většinou chceme dělat testy o parametrech nějakého náhodného modelu, často o parametrech určitého rozdělení:

I) parametr je nějak svázán se $\begin{cases} \text{střední hodnotou} \\ \text{rozptylem} \end{cases}$ daného rozdělení

II) příslušná hustota se dá šikovně přepsat

Příklad: Máme distribuci, kdy se parametr váže ke střední hodnotě:

$$\begin{array}{ll} X \sim \text{Alt}(p) & EX = p \\ X \sim N(\mu, \sigma^2) & EX = \mu \\ X \sim \text{Po}(\lambda) & EX = \lambda \end{array}$$

Podle ZVČ (Čebyšev) víme, že průměr je dobrým odhadem střední hodnoty.

$P(|\bar{X}_n - E\bar{X}_n| > \varepsilon) \rightarrow 0$ (pro $n \rightarrow \infty$) $\Rightarrow \bar{X}_n$ se nabízí jako dobrá testovací statistika

(i) Intuitivně cítíme, že velký rozdíl \bar{X}_n od očekávané hodnoty mluví proti H .

(ii) Nabízí se užít CLV: $\left| \frac{\sum X_i}{n} - p \right| > C_1 \Rightarrow |\sum X_i - np| > C_2$

Definice: Nechť X_1, \dots, X_n jsou nezávislé náhodné veličiny s hustotou $f(x, \theta)$. Nechť sdružená hustota $\prod_{i=1}^n f(x_i, \theta)$ se dá napsat ve tvaru $h(x) \cdot c(\theta) \exp\{T(x) \cdot Q(\theta)\}$. Pak řekneme, že se jedná o rozdělení z exponenciální třídy hustot.

Př. 13. 11. 1. 2010

9 Trídění

Zajímá nás závislost počtu operací na rozsahu úlohy. Nageneruji 50 náhodných permutací délky 10^7 , seřídím je a zapíšu potřebné charakteristiky.

9.1 Generování náhodných permutací

Hledám zobrazení $\{1, \dots, n\} \rightarrow n!$ takové, že každá permutace má stejnou šanci nastat, tedy $P(\text{perm} \in \mathcal{P}) = \frac{1}{n!}$.

Naivní algoritmus: Náhodně vyberu prvek množiny $\{1, \dots, n\}$ a podle indexu vygeneruji příslušnou permutaci. Díky částečnému uspořádání je tato permutace jednoznačně určena. Zjevně mají všechny permutace stejnou šanci nastat.

Algoritmus: Prohazování:

```
for(int i = n; i >= 2; --i) {
    nagenervuj nahodne cislo z R{1, ..., n};
    swap(i, n - i + 1);
}
```

Podstatné je, že tento algoritmus teoreticky skutečně zaručuje rovnoměrné rozdělení na \mathcal{P} .

9.2 Kotouč

Věta (Kotouč I): Umím seřadit posloupnost délky 10^7 v průměru za ν_0 se směrodatnou odchylkou Δ .

Položme toto tvrzení za standard.

Vytvořím jinou proceduru. Nagenervuji X_1, \dots, X_{50} a ptám se na poměr čas/složitost.

Otázka: Jsem lepší, než standard \equiv kotouč?

- (i) Předpokládám o všech svých časech nějakou distribuci, například normální rozdělení $N(\nu, \sigma^2)$.
- (ii) Stanovím hypotézu $H: \nu = \nu_0$ a k ní alternativu $A: \nu < \nu_0$ (jsem lepší). Alternativu A lze také zvolit jako $\nu > \nu_0$ (jsem horší) nebo $\nu \neq \nu_0$ (nejsem jako Kotouč).

Intuitivně zamítnu H když můj průměr je menší, než Kotoučův.

$$P(|X - \nu| \geq \varepsilon) \leq \frac{\text{var} X}{\varepsilon^2}$$

$$\left(E\bar{X}_n = \nu; \quad \text{var}\bar{X}_n = \frac{\sigma^2}{n} \right)$$

$$P(|\bar{X}_n - \nu| \geq \varepsilon) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \rightarrow 0 \text{ pro } n \rightarrow \infty, \text{ test tedy založíme na průměru}$$

X_1, \dots, X_n nechť jsou nezávislé

$$\hat{\nu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n \sim N\left(\nu, \frac{\sigma^2}{n}\right)$$

Budu zamítat tehdy, když $\bar{X}_n < \nu_0$. Ale o kolik menší má smysl?

$$E\bar{X}_n = \frac{\nu \cdot n}{n} = \nu$$

$$\bar{X}_n - \nu_0 < C_\alpha$$

$$P(H \text{ zamítneme} | H \text{ platí}) = \alpha$$

$$P(\bar{X}_n - \nu_0 < C_\alpha | \nu = \nu_0)$$

$$\bar{X}_n | \nu = \nu_0 \Rightarrow \bar{X}_n \sim N\left(\nu_0, \frac{\sigma^2}{n}\right)$$

$$X \sim N(\nu, \sigma^2) \Rightarrow \frac{X - \nu}{\sigma} \sim N(0, 1)$$

$$Y \sim N(0, 1) \Rightarrow (\nu + \sigma \cdot Y) \sim N(\nu, \sigma^2)$$

$$P\left(\frac{N(0, \sigma^2)}{\sigma} < \frac{C_\alpha}{\sigma}\right), \frac{N(0, \sigma^2)}{\sigma} \text{ je náhodná veličina s } N(0, 1) \text{ a } \frac{C_\alpha}{\sigma} \text{ označím } K$$

K je α kvantil $N(0, 1)$, pro $\alpha = 0.05$ je $K = -1.64$

Tedy $C_\alpha = \sigma \cdot K$.

Závěr: Na hladině významnosti α zamítneme H : *jsme stejně dobří jako standard* ve prospěch alternativy A : *jsme lepší*, pokud $\bar{X}_n \leq \nu_0 + \sigma \cdot K$, kde K je α -kvantil $N(0, 1)$.

$$\sigma^2 \text{ většinou neznáme a musíme odhadovat: } \hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$$

Co se stane, když σ^2 nahradím odhadem? Distribuce $\frac{X-\nu}{\hat{\sigma}}$ **již není nadále normální!!!**

Dobré zprávy:

- (i) Existuje přesné rozdělení a nazývá se t (Studentovo).
- (ii) Když $n \rightarrow \infty$, pak se Studentovo rozdělení blíží k normálnímu.
- (iii) Místo do tabulek pro $N(0, 1)$ se dívám do tabulek rozdělení t .
- (iv) Je-li $n \geq 30$, pak neudělám velkou chybu, použiji-li místo t -rozdělení rozdělení normální.

Věta (Kotouč II):

$$\begin{array}{lll} \text{Alg}_K & K_1, \dots, K_n \sim N(\nu_1, \sigma_1^2) & H: \nu_1 = \nu_2 \quad K \text{ i } JA \text{ třídí stejně rychle} \\ \text{Alg}_{JA} & JA_1, \dots, JA_m \sim N(\nu_2, \sigma_2^2) & A: \nu_1 > \nu_2 \quad \text{Kotouč třídí pomaleji} \end{array}$$

Každý si nageněroval 50 náhodných permutací a použil se ten samý počítač.

H zamítneme pokud $\hat{\nu}_1 - \hat{\nu}_2 \gg 0$, tedy pokud $\hat{\nu}_1 - \hat{\nu}_2 > C_\alpha$.

$$\begin{aligned}\hat{\nu}_1 &\sim N\left(\nu_1, \frac{\sigma_1^2}{n}\right) & \hat{\nu}_1 &= \frac{1}{n} \sum_{i=1}^n K_i \\ \hat{\nu}_2 &\sim N\left(\nu_2, \frac{\sigma_2^2}{m}\right) & \hat{\nu}_2 &= \frac{1}{m} \sum_{i=1}^m JA_i\end{aligned}$$

$$\alpha = P(\hat{\nu}_1 - \hat{\nu}_2 > C_\alpha | \nu_1 = \nu_2) = P\left(\frac{\hat{\nu}_1 - \hat{\nu}_2 - (\nu_1 - \nu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} > \frac{C_\alpha - (\nu_1 - \nu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} | \nu_1 = \nu_2\right)$$

Dostáváme tedy $P\left(N(0, 1) > \frac{C_\alpha}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right)$. Označme $\frac{C_\alpha}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = K$. Dostáváme se do té samé situace jako předtím. Neznámé σ_i^2 odhadneme:

$$\begin{aligned}\hat{\sigma}_1^2 &= \frac{1}{n} \sum_{i=1}^n (K_i - \bar{K})^2 \\ \hat{\sigma}_2^2 &= \frac{1}{m} \sum_{i=1}^m (JA_i - \overline{JA})^2\end{aligned}$$

Věta (Kotouč III): Nagenarovat $n = 50$ permutací a na každou z nich aplikovat oba algoritmy.

Na každém mám 2 párová řešení $K_1^*, \dots, K_n^* \sim N(\nu_1, \sigma_1^2)$
 $JA_1^*, \dots, JA_n^* \sim N(\nu_2, \sigma_2^2)$. Nyní položíme $Y_i^* = K_i^* - JA_i^* \sim N(\nu, \sigma^2)$ a mám místo dvou výběrů jeden. A to je situace, kterou jsme začali.

9.3 Párový test

Lze použít jen na přirozených párech (tedy levá ruka \times pravá ruka ano, Sparta \times Slavie ne).

Kdy to přináší zisk? Užití párového místo dvouvýběrového testu? Je to tehdy, když je velká kladná korelace mezi K_i^* a JA_i^* !

$$\text{var}Y_i^* = \text{var}K_i^* + \text{var}JA_i^* - \text{cov}(K_i^*, JA_i^*)$$

Je-li $\text{cov}(K_i^*, JA_i^*) > 0$, pak jsem snížil rozptyl Y_i^* , tedy jsem zvýšil šanci, že přelezu C_α .