PRAVDĚPODOBNOST A STATISTIKA

PETR HÁJEK

Cílem textu je dát porozumění matematickému základu teorie pravděpodobnosti a matematické statistiky, přičemž u studenta se nepředpokládá znalost teorie míry. Tato okolnost silně limituje možnost výkladu. Můj přístup je tedy založen na vybudování intuitivní představy o pravděpodobnosti využitím kombinatorických modelů, které mají konečný charakter, nebo jsou na základě konečného přístupu pochopitelné. Jde zejména o nekonečný model házení mincí, resp. náhodné procházky. V této části je můj text motivován monografií W. Fellera.

Ve zbylé části je proveden standartní výklad pravděpodobnosti a statistiky, s důrazem na matematický obsah, který víceméně sleduje skripta Dupač, Hušková.

K přípravě přednášky byla použita následující (doporučená) literatura:

V. Rogalewicz, Pravděpodobnost a statistika pro inženýry (FEL ČVUT 1998),
V. Dupač, M. Hušková, Pravděpodobnost a matematická statistika (MFF 2005)
F. Jaroš, Pravděpodobnost a statistika (skripta VŠCHT 1998)
J. Novovičová, Pravděpodobnost a matematická statistika (skripta ČVUT 2006)
M. Navara, Pravděpodobnost a matematická statistika (skripta ČVUT 2007)
J. Anděl, Matematika náhody (Matfyzpress 2007)
W. Feller, An Introduction to Probability Theory and its Applications (Wiley, 1970)

1. Úvod

Historicky byly motivací teorie pravděpodobnosti hazardní hry. Motivační úloha. Dva hráči hrají sérii her o celkovou sázku, kterou získá ten hráč, který jako první získá 6 vítězství. Předpokládáme, že šance obou hráčů v každé hře jsou vyrovnané. Za stavu, kdy hráč A vyhrál již 5 her a hráč B vyhrál 3 hry musela být série přerušena. Jakým způsobem si mají hráči rozdělit celkovou sázku, aby to odpovídalo jejich šanci na celkové vítězství při pokračování hry?

Odpověď získáme s použitím představy že tyto situace mají jistou pravděpodobnost vyjádřitelnou číslem z $[0, 1]$ (v procentech vynásob 100). Od tohoto čísla přitom očekáváme následující vlastnosti:

1. Předpokládáme, že pokud lze v popisu situace najít symetrii, potom pravděpodobnosti symetrických jevů by měly být stejné. Tento předpoklad umožňuje vytváření jednoduchých pravděpodobnostních modelů.
2. Pokud se dva jevy navzájem vylučují, potom je pravděpodobnost že nastane alespoň jeden z nich rovna součtu jejich pravděpodobností. Speciálně, pravděpodobnost určitého jevu je rovna jedna mínus pravděpodobnost negace tohoto jevu.
3. Při opakování pokusu za stejných podmínek má poměr úspěšných případů ke všem jistou limitu (pravděpodobnost), která se shoduje s přiřazenou pravděpodobností.

K řešení našeho problému vytvoříme pravděpodobnostní model situace. Model obsahuje množinu Ω která obsahuje popis všech možných průběhů událostí, nebo situací, které považujeme za stejně pravděpodobné. Námi zkoumaný jev bude v tomto kontextu jistou podmnožinou Ω. Je třeba zdůraznit, že model podstatně závisí na správnosti našeho porozumění situaci, resp. na použití veškeré relevantní informace kterou máme k dispozici, a jeho praktickým testem je splnění podmínky 3.

Nechť Ω je neprázdná množina. Pokud $\mathcal{A}$ je systém nějakých podmnožin Ω, $\emptyset \in \mathcal{A}$, který je uzavřený na doplňky a konečná sjednocení a průniky, říkáme, že $\mathcal{A}$ je algebra podmnožin Ω. Nejjednodušším příkladem takové algebry je algebra všech podmnožin Ω.

Nejjednodušší pravděpodobnostní model je popsán následovně.

Definice 1.1. *Nechť* $(\Omega, \mathcal{F}, P)$ *je uspořádaná trojice, kde*
1. Ω*-je neprázdná konečná množina, kterou budeme nazývat pravděpodobnostní prostor, nebo prostor elementárních jevů, a jejíž prvky budeme nazývat elementárními jevy.*
2. $\mathcal{F}$ *je množina všech podmnožin* Ω *(včetně* $\emptyset$*), zvaných jevy.*
3. $P : \mathcal{F} \to [0,1]$ *je funkce která splňuje následující vlastnost. Pro všechny podmnožiny* $A \subset \Omega$*, platí* $P(A) = \frac{|A|}{m}$

Je zřejmé, že pro každý disjunktní konečný systém $A_j \in \mathcal{F}$ platí $P(\cup A_j) = \sum P(A_j)$. Máme následující interpretaci množinových operací ve vztahu k jevům:
$A^c = \Omega \setminus A$ je jev opačný k A, tedy negace.
$A \cap B$ je množina, která představuje jev kdy A i B nastaly současně
$A \cup B$ je množina která představuje jev kdy nastal jev A nebo B.

Příklady modelů:

Házení férovou mincí.

Házení kostkou.

♣ Opakované házení férovou mincí nebo kostkou $\Omega = \{(a_1, \ldots, a_k) : a_i \in \{0,1\}\}$, $|\Omega| = 2^k$, a aplikace tohoto modelu na následující problémy:

Jaká je pravděpodobnost, že rodina se čtyřmi dětmi má dva chlapce a dvě dívky?

Jaká je pravděpodobnost, že padne právě m krát Hlava v n-hodech, nebo že padne alespoň m krát? $p = \frac{1}{2}$, $\binom{n}{m} p^m (1-p)^{n-m}$, $\sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k}$

Jaké je řešení motivační úlohy?

Jaká je pravděpodobnost, že při 2 hodech kostkou součet hodnot je m?

♣ Zobecnění předchozí situace, k-násobně opakované náhodné tahání prvků z množiny n rozlišitelných prvků s vracením. $\Omega = \{(a_1, \ldots, a_k) : a_i \in \{1, \ldots, n\}\}$. $|\Omega| = n^k$.

Jaká je pravděpodobnost, že při k hodech kostkou vyjdou navzájem různá čísla? $\frac{6 \times \ldots (6-k+1)}{6^k}$.

Jaká je pravděpodobnost, že při náhodném umístění k koulí do k různých krabic bude v každé krabici jedna koule? $\frac{k!}{k^k}$.

♣ k-násobně opakované náhodné tahání prvků z množiny n rozlišitelných prvků bez vracení. $\Omega = \{(a_1, \ldots, a_k) : a_i \in \{1, \ldots, n\}, a_i$ navzájem různá$\}$. $|\Omega| = n(n-1)\ldots(n-k+1)$.

Táhneme postupně čísla z množiny $\{1, \ldots 2n\}$. Jaká je pravděpodobnost, že prvních n čísel bude sudých? $\frac{(n!)^2}{(2n)!}$.

Jaká je pravděpodobnost, že při k tazích s vracením z množiny $\{1, \dots, n\}$ prvcích byl vytažen 1? $1 - \frac{(n-1)\dots(n-k)}{n\dots(n-k+1)} = \frac{k}{n}$

Jaká je pravděpodobnost, že při postupném tažení papírku s písmeny MMAA z pytlíku vyjde slovo MAMA? $\Omega = \{(a_1, a_2, a_3, a_4) : a_i \in \{1, 2, 3, 4\}\}$ permutace, přičemž $1, 2$ je přiřazeno Ω a zbytku A.

Jaká je pravděpodobnost, že skupina 2 žen a 3 mužů usazená náhodně kolem kulatého stolu sestává z obou skupin sedících pohromadě? Ω je množina všech 2 prvkových podmnožin cyklu o 5 prvcích (10 prvku), příznivých případů je 5.

V pytlíku je 50 černých a 50 bílých koulí. Vytáhneme náhodně 5 koulí. Jaká je pravděpodobnost, že vytáhneme 3 černé a 2 bílé koule? Ω bude množina všech 5-ti prvkových podmnožin 100 prvkové množiny. $|A|$ bude součin všech tříprvkových z jedné poloviny a dvouprvkových z druhé.

Vytáhneme náhodný počet koulí z pytlíku o n koulích. Jaká je pravděpodobnost, že počet je sudý? Ω budou všechny podmnožiny $\{1, \dots, n\}$. Dokažte, že počet podmnožin s lichým počtem prvků je stejný jako počet podmnožin se sudým počtem prvků. Užij $(1-1)^n = \sum (-1)^k \binom{n}{k} a^k b^{n-k}$

Vytáhneme náhodně dvě čísla z množiny $\{1, \dots, n\}$. Jaká je pravděpodobnost, že první číslo je menší než druhé? $\Omega = \{(a, b) : a \neq b \in \{1, ..., n\}\}$, užití symetrie.

Kolik je posloupností $a_1 < a_2 < a_3$ sestávajících z čísel z množiny $\{1, \dots, 7\}$? Jaká je pravděpodobnost, že tři náhodně vybraná navzájem různá čísla z této množiny splňují tyto nerovnosti? Ω bude množina všech uspořádaných trojic z dané množiny, je jich $7 \times 6 \times 5$. množinu A budeme popisovat pomoci relace ekvivalence na Ω, která da dohromady trojice obsahující stejná čísla (v různých pořadích). Výsledek $\frac{1}{6}$.

Pokud jevy nejsou disjunktní, vlastnost součtu pravděpodobností neplatí, a místo toho platí následující formule.

Věta 1.2. *(Věta o inkluzi-exkluzi)*

$$P(A_1 \cup \dots \cup A_n) = \sum_{j=1}^{n} P(A_j) - \sum_{1 \leq i < j \leq n}^{n} P(A_i \cap A_j)$$

$$+ \sum_{1 \leq i < j < k \leq n}^{n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n)$$

Důkaz. Z formule $(1-1)^n = \sum (-1)^k \binom{n}{k}$. Použijeme na každý prvek ze sjednocení, který je kódován tím, ve kterých z A_i leží a ve kterých neleží (vytvoříme rozklad Ω pomoci kódu $(a_1, \dots, a_n), a_i \in \{0, 1\}$). □

Jaká je pravděpodobnost, že náhodné celé číslo z intervalu $[1, 1000]$ není dělitelné ani 7 ani 12 ani 15?

Ve skupině 4 manželských párů byly náhodně vytvořeny taneční páry. Jaká je pravděpodobnost, že aspoň jeden muž tančí se svojí ženou? Že nikdo netančí se svojí ženou? Ω budou všechna možná spárování.

2. Obecné pravděpodobnostní prostory

Představa o konečné množině elementárních jevů a rovnoměrně rozdělené pravděpodobnosti je příliš limitující, a je třeba ji zobecnit.

♣ Motivační příklady:

Hod neférovou kostkou, pravděpodobnost H je p, pro O je $1-p$. Po 2, resp. n hodech je pravděpodobnostní model $\Omega = \{(a_1, \ldots, a_n) : a_i \in \{H, O\}\}$, avšak jednotlivé elementární jevy nemají stejnou pravděpodobnost.

Nekonečná diskrétní pravděpodobnost. Házíme mincí, dokud nepadne H. Jaká je pravděpodobnost, že hra skončí před 6 tahem? Jaká je pravděpodobnost, že hra skončí po konečně mnoha hodech? Přirozený model je nekonečná $\Omega = \{(a_i) : a_1, \ldots, a_n = O, a_{n+1} = H\}$. Ani zde není pravděpodobnost elementárních jevů stejná.

Nediskrétní pravděpodobnosti-geometrie. Jaká je pravděpodobnost, že náhodně zvoleny bod z $[-1, 1]$ je nezáporný, resp. kladný, resp. roven 0?

Definice 2.1. *Pravděpodobnostní prostor*
Říkáme, že uspořádaná trojice $(\Omega, \mathcal{F}, P)$ tvoří pravděpodobnostní prostor, jestliže:
1. Ω-je neprázdná množina, kterou budeme nazývat pravděpodobnostní prostor, nebo prostor elementárních jevů, a jejíž prvky budeme nazývat elementárními jevy.
2. $\mathcal{F}$ je jistá množina podmnožin Ω (která obsahuje $\emptyset$), zvaných jevy. Tato množina splňuje axiomy tzv. σ-algebry, to znamená, že pokud $A \in \mathcal{F}$ pak $A^c = \Omega \backslash A \in \mathcal{F}$. Pokud $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, potom

$$\cup_{i=1}^{\infty} A_i \in \mathcal{F}, \;\; \cap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

3. $P : \mathcal{F} \to [0,1]$ je funkce (v matematice obvykle nazývaná míra) která splňuje následující axiomy:
1.$P(\Omega) = 1$.
2. Pro disjunktní spočetný systém $A_j \in \mathcal{F}$ platí $P(\cup A_j) = \sum P(A_j)$.

Vlastnosti.
$P(A^c) + P(A) = 1$, Pokud máme rostoucí posloupnost jevů $A_1 \subset A_2 \subset \ldots$, (resp. klesající posloupnost $A_1 \supset A_2 \supset \ldots$), tak platí

$$P(\cup_{n=1}^{\infty} A_n) = \lim P(A_n), resp.$$

$$P(\cap_{n=1}^{\infty} A_n) = \lim P(A_n)$$

Speciálně, pokud máme klesající posloupnost $A_1 \supset A_2 \supset \ldots$, pro kterou platí $\cap_{n=1}^{\infty} A_n = \emptyset$, potom $\lim P(A_n) = 0$.

Diskuse.
♣ Pro zjednodušení budeme někdy používat konvenci nahrazení uspořádané trojice $(\Omega, \mathcal{F}, P)$ dvojicí (Ω, P) s tím, že mlčky předpokládáme, že systém $\mathcal{F}$ obsahuje všechny množiny se kterými právě pracujeme. Tento předpoklad lze rozumně matematicky zdůvodnit.

♣ Nejjednodušší příklad Dirakova míra δ_x v bodě $x \in \Omega$, nebo obecněji diskrétní míra $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$, kde $a_i > 0, \sum a_i = 1$. Zde lze volit $\mathcal{F}$ množinu všech podmnožin Ω. $\mu(A) = \sum_{x_i \in A} a_i$.

Pokud $\mathcal{A}$ je je algebra podmnožin Ω, potom lze vytvořit její sigma uzávěr, což je nejmenší σ-algebra $\mathcal{F}$ podmnožin Ω, která obsahuje $\mathcal{A}$.

Věta 2.2. *(Hahn-Kolmogorov)*
Nechť $\mathcal{A}$ je algebra podmnožin množiny Ω, $P : \mathcal{A} \to [0,1]$ je aditivní funkce, která splňuje dodatečnou podmínku: Pro každou klesající posloupnost $A_1 \supset A_2 \supset \dots$, pro kterou $\cap_{n=1}^{\infty} A_n = \emptyset$, platí $\lim P(A_n) = 0$.
Potom P lze rozšířit na sigma uzávěr $\mathcal{F}$ tak, že uspořádaná trojice $(\Omega, \mathcal{F}, P)$ tvoří pravděpodobnostní prostor.

Bez důkazu. Podmínka monotonie je též nutná, jak jsme již viděli.

♣ Nyní vytvoříme model nekonečného házení obecnou mincí (nebo náhodné procházky).
Zvolme pevné $p, q > 0, p + q = 1$,...H padne s pravděpodobností p, O padne s pravděpodobností q.

$$\Omega = \{(a_i)_{i=1}^{\infty}, a_i \in \{H, O\}\},$$

Symboly H, O lze nahradit libovolnou dvojicí symbolů, např. $\{-1, 1\}$ pro model náhodné procházky, nebo $\{0, 1\}$ pro Bernoulliho model. Algebra podmnožin $\mathcal{A}$ generována množinami

$$A(b_1, \dots, b_n) = \{(a_i)_{i=1}^{\infty} : a_j = b_j,\ j = 1, \dots n\}$$

s odpovídající pravděpodobností

$$P(A(b_1, \dots, b_n)) = p^k q^{n-k},\ \ k \text{ je počet } H \text{ v posloupnosti } b_j.$$

Použitím kompaktnosti $\{0,1\}^I N$, a toho že $A(b_1, \dots, b_n)$ jsou clopen, vyjde:

$$\cap_{n=1}^{\infty} A_n = \emptyset \Rightarrow \cap_{n=1}^{N} A_n = \emptyset$$

pro dostatečně velké $N \in \mathbb{N}$. Podle Věty 2.2 lze tedy P rozšířit na sigma obal $\mathcal{A}$, označený $\mathcal{F}$.
♣ Příklady jevů z $\mathcal{F}$, a jejich pravděpodobností:
Jednobodová podmnožina Ω (tedy jeden konkrétní průběh hry).
Padla alespoň jednou H.
Padlo nekonečně mnoho H.
Padlo nekonečně mnoho H i O.
V průběhu hry padla nějaká konkrétní konečná posloupnost.

♣ Varianta předchozí konstrukce pro házení kostkou.

♣ Lebesgueova míra na $[0,1]$. Aplikace předchozí konstrukce na jednotkový interval, požitím dvojkového rozvoje desetinného čísla. Vznik translačně invariantní míry, geometrická pravděpodobnost.

Pravděpodobnost, že náhodně vybraný bod z $[0,1]$ je větší než r. Model $\Omega = [0,1]$.

Geometrická pravděpodobnost-házení šipkou do terče. Model $\Omega = \{(x,y) : x^2 + y^2 \leq 1\}$. Diskuse rozdělení pravděpodobnosti.

♣ Nechť $f : \mathbb{R}^n,\ f \geq 0$ je spojitá funkce, $\int_{\mathbb{R}^n} f = 1$. Položme $\Omega = \mathbb{R}^n$, $\mathcal{A}$ algebru generovanou všemi polouzavřenými kvádry, $P(A) = \int_A f$. Potom P je aditivní a splňuje podmínky Věty 2.2, tedy P lze rozšířit na sigma uzávěr $\mathcal{F}$ tak, že uspořádaná trojice $(\Omega, \mathcal{F}, P)$ tvoří pravděpodobnostní prostor.

3. Nezávislost a konstrukce součinových prostorů

Definice 3.1. *Říkáme, že jevy $A, B \in \mathcal{F}$ jsou nezávislé, pokud platí $P(A \cap B) = P(A)P(B)$.*

Tvrzení 3.2. *Pokud A, B jsou nezávislé potom A, B^c jsou též nezávislé.*

Příklady:

Náhodný výběr karty z balíčku, A červená karta, B eso apod.

Na Vánoční besídku přišlo n dětí. Každé dítě přineslo stejně zabalený dárek a dalo ho do koše. Na konci besídky si děti náhodně vyberou dárek z koše. Jev A Pepíček si odnáší svůj vlastní dárek, B Anička si odnáší svůj vlastní dárek.

Hod dvěma kostkami, A-součet hodů je k, B-na první kostce padlo n.

Definice 3.3. *Říkáme, že jevy $A_1, \dots, A_n \in \mathcal{F}$ jsou nezávislé (nebo též stochasticky nezávislé) pokud pro libovolný výběr navzájem různých $j_1, \dots, j_k$ z množiny $\{1, \dots, n\}$ platí $P(A_{j_1} \cap \dots \cap A_{j_k}) = P(A_{j_1}) \dots P(A_{j_k})$.*

Hoďme dvakrát mincí. A-první hod H, B-druhý hod H, C-oba hody stejné. Potom ABC jsou po dvou nezávislé ale nikoli nezávislé.

♣ Pokud jsou A_i nezávislé, potom je možné libovolnou část z nich zaměnit za A_i^c a jevy zůstávají nezávislé.

Základní metoda vytváření pravděpodobnostního modelu nezávislých jevů je následující.

Věta 3.4. *(Kartézský součin pravděpodobnostních prostorů)*
Nechť $(\Omega_1, \mathcal{F}_1, P_1)$, $(\Omega_2, \mathcal{F}_2, P_2)$ jsou pravděpodobnostní prostory. Položme $\Omega = \Omega_1 \times \Omega_2$, a σ-algebra $\mathcal{F}$ budiž sigma obal algebry generované množinami $A \times B$, kde $A \in \mathcal{F}_1, B \in \mathcal{F}_2$. Potom existuje pravděpodobnost P na $\mathcal{F}$ která splňuje pro každé A, B z $\mathcal{F}_1, \mathcal{F}_2$,

$$P(A \times B) = P(A \times \Omega_2 \cap \Omega_1 \times B) = P_1(A)P_2(B) = P(A \times \Omega_2)P(\Omega_1 \times B).$$

Takto vytvořený pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$ nazýváme kartézským součinem původních prostorů.

Konstrukci lze provést pro libovolný konečný počet pravděpodobnostních prostorů.

Tato konstrukce již byla námi použita pro modely opakovaného tahání prvků s vracením, resp. házení mincí.

♣ Lebesgueova míra na $[0, 1] \times [0, 1]$. Měření plošného obsahu.

Jaká je pravděpodobnost, že dvojice náhodně vybraných bodů z $[-1, 1]$ má vzdálenost nejvýše 1?

Na kružnici poloměru 1 jsou náhodně vybrané dva body. Jaká je pravděpodobnost, že jejich vzdálenost je nejvýše d?

Rozřežeme úsečku náhodně na tři části. Jaká je pravděpodobnost, že z těchto úseček lze sestavit trojúhelník?

4. Podmíněná pravděpodobnost

Z hlediska praktických aplikací teorie pravděpodobnosti, je naše hodnocení pravděpodobnosti určitých jevů je do značné míry subjektivní, a záleží na množství informací, které máme o daném jevu k dispozici. Jinými slovy, dodatečné informace mohou ovlivnit náš pravděpodobnostní model. To samozřejmě neznamená, že matematická pravděpodobnost jevů není jednoznačně určena, naopak tato je určena jednoznačně, jakmile máme k dispozici náš model. Přechod od jednoho modelu jevu ke druhému lze často uskutečnit použitím konceptu podmíněné pravděpodobnosti.

Motivační příklady.

Experiment dvou hodů kostkou. Jaká je pravděpodobnost, že první hod byl 6? Jak budeme tuto pravděpodobnost hodnotit, pokud víme, že celkový součet hodů je 9, resp. 8?

Hod dvěma mincemi. Jaká je pravděpodobnost, že padly dvě hlavy, pokud víme, že padla alespoň jedna hlava? Výsledek $\frac{1}{3}$ je překvapující (čekali bychom $\frac{1}{2}$).

Definice 4.1. *Podmíněná pravděpodobnost*
Nechť $A, B \in \mathcal{F}$ jsou dva jevy. Podmíněná pravděpodobnost jevu A za podmínky B je definována následovně

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

10 hodů mincí. Víme-li, že padla hlava 7krát, jaká je pravděpodobnost ze prvním hodem padla hlava? Ω budou sedmiprvkové podmnožiny z $\{1, \ldots, 10\}$, A budou šestiprvkové z 9.

Roztržitý profesor matematiky zapomíná v obchodě deštník s pravděpodobností $\frac{1}{4}$. Dnes cestou domu nakupoval ve třech obchodech a domu přišel bez deštníku. Jaká je pravděpodobnost, že deštník zapomněl v jednotlivých obchodech? Označme A_i jev deštník byl zapomenut v i-tém obchodě. Potom $A = \cup A_i$ je jev ze deštník byl zapomenut. Platí $P(A_1) = \frac{1}{4}, P(A_2) = \frac{3}{4}\frac{1}{4}, P(A_3) = \frac{3}{4}\frac{3}{4}\frac{1}{4}$. Vyjde $P(A_1|A) = 0.43, P(A_2|A) = 0.32, P(A_3|A) = 0.24$.

Tvrzení 4.2. *Nechť $A, B \in \mathcal{F}$ jsou dva jevy s nenulovou pravděpodobností. Tyto jevy jsou nezávislé právě když platí*

$$P(A|B) = P(A).$$

Definice 4.3. *Předpokládejme, že jevy $F_1, \ldots, F_n \in \mathcal{F}$ jsou disjunktní s nenulovou pravděpodobností a jejich sjednocením je celý prostor Ω. Takový systém se nazývá úplný systém jevů s nenulovou pravděpodobností na Ω.*

Tvrzení 4.4. *Věta o úplné pravděpodobnosti*
Předpokládejme, že jevy $F_1, \ldots, F_n$ jsou úplný systém jevů s nenulovou pravděpodobností na Ω. Potom platí

$$P(A) = \sum_{j=1}^{n} P(F_j)P(A|F_j)$$

♣ Hodíme mincí. Pokud je výsledek hlava, hodíme jednu kostku. Pokud je výsledek orel, hodíme dvě kostky. Jaká je pravděpodobnost , že na kostce padla právě jednou šestka? $\Omega = \{(H, i), (O, i, j)$.

A..padla právě jednou šestka, B na minci padla hlava.

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c) = \frac{1}{2}\frac{1}{6} + \frac{1}{2}\frac{10}{6^2} \doteq 0.22$$

♣ Hodíme kostkou a vybereme náhodně tolik karet z balíku 32 karet, kolik padlo na kostce. Jaká je pravděpodobnost, že vytáhneme aspoň jedno eso?

A..alespon 1 eso, F_j na kostce padlo j. $P(F_j) = \frac{1}{6}$,

$$P(A|F_j) = 1 - \frac{\binom{28}{j}}{\binom{32}{j}}$$

$$P(A) = \sum_{j=1}^{6} P(F_j)P(A|F_j)$$

♣ Model pro pojišťovnu automobilů. Předpokládáme, že řidiči jsou buď dobří, nebo špatní. Dobrý řidič má nehodu za rok s pravděpodobností $p = 0.06$, špatný řidič má nehodu s pravděpodobností $s = 0.6$. Předpokládáme, že poměr dobrých řidičů ke špatným je v populaci $5 : 1$. Jaká je pravděpodobnost, že řidič bude mít v prvním roce nehodu? Jestliže měl řidič v prvním roce nehodu, jaká je pravděpodobnost, že bude mít nehodu i příští rok?

A jev nehoda v prvním roce, B jev nehoda dva roky po sobě.

$$P(A) = \frac{1}{6}s + \frac{5}{6}p = 0.15$$

$$P(B) = \frac{1}{6}s^2 + \frac{5}{6}p^2 = 0.063$$

$$P(B|A) = 0.42$$

♣ Urnový model. V koši je n bílých a m černých koulí. Vytáhneme jednu z nich náhodně, vrátíme ji zpět a přidáme k koulí téže barvy. Proces opakujeme. jaká je pravděpodobnost vytažení bíle koule j-tým tahem? $\Omega_j = \{(a_1, \dots, a_j),\ a_i \in \{0,1\}\}$, $P((0)) = \frac{n}{n+m}$, $P((1)) = \frac{m}{n+m}$.

Pro $j = 1$ je výsledek $\frac{n}{n+m}$. Pro $j = 2$ zvolím $B = F_1$ první tah byla bílá koule, $B^c = F_2$ první tah byla černá koule, A druhý tah je bílá koule.

$$P(B) = \frac{n}{n+m}, P(B^c) = \frac{m}{n+m}, P(A|B) = \frac{n+k}{n+m+k}, P(A|B^c) = \frac{n}{n+m+k}$$

$$P(A) = P(B)P(A|B)+P(B^c)P(A|B^c) = \frac{n}{n+m}\frac{n+k}{n+m+k}+\frac{m}{n+m}\frac{n}{n+m+k} = \frac{n}{n+m} \blacksquare$$

Obecný případ dokážeme indukcí podle j, za předpokladu, že výsledek platí pro všechna menší j a všechny možnosti m, n, k.

Indukční krok pro $j + 1$.

Zvolím $B = F_1$ první tah byla bílá koule, $B^c = F_2$ první tah byla černá koule, A $j + 1$ tah je bílá koule. Výpočet je nyní stejný jako pro $j = 2$, kde využijeme indukčního předpokladu výsledku pro menší hodnoty j.

$$P(B) = \frac{n}{n+m}, P(B^c) = \frac{m}{n+m}, P(A|B) = \frac{n+k}{n+m+k}, P(A|B^c) = \frac{n}{n+m+k}$$

$$P(A) = P(B)P(A|B)+P(B^c)P(A|B^c) = \frac{n}{n+m}\frac{n+k}{n+m+k}+\frac{m}{n+m}\frac{n}{n+m+k} = \frac{n}{n+m} \blacksquare$$

♣ Máme $N+1$ krabic $K_i, i = 0, \dots N$, každá K_i obsahuje i bílých a $N-i$ černých míčků. Vybereme náhodně jednu krabici, a a provedeme následující experiment. Budeme vytahovat náhodný míček n krát po sobě s vracením do krabice. Pokud všech n výsledků dalo bílý míček, jaká je pravděpodobnost, že $n+1$ výběr bude opět bílý míček?

Označme F_i jev výběru krabice K_i, $P(F_i) = \frac{1}{N+1}$. Jev A znamená, že vyšlo n bílých míčků po sobě, B znamená že $n+1$ tah je opět bílý míček. $P(A|F_i) = \frac{i^n}{N^n}$, tedy pro $n \ll N$ dostaneme odhad

$$P(A) = \sum_{i=0}^{N} \frac{1}{N+1} (\frac{i}{N})^n \doteq \int_0^1 x^n dx = \frac{1}{n+1}$$

Tedy

$$P(A \cap B) = P(B) \doteq \frac{1}{n+2}$$

$$P(B|A) = \frac{P(B)}{P(A)} \doteq \frac{n+1}{n+2}$$

5. Bayesův vzorec

Věta 5.1. *Bayesův vzorec*

Předpokládejme, že jevy $F_1, \dots, F_n \in \mathcal{F}$ tvoří úplný systém jevů s nenulovou pravděpodobností na Ω. Potom pro každý jev A splňující $P(A) > 0$ platí

$$P(F_j|A) = \frac{P(F_j \cap A)}{P(A)} = \frac{P(A|F_j)P(F_j)}{\sum_{j=1}^{n} P(F_j)P(A|F_j)}$$

Pravděpodobnosti $P(F_j|A)$ se říká aposteriorní pravděpodobnost, $P(F_j)$ apriorní pravděpodobnost. Typická situace spočívá v tom, že po uskutečnění náhodného výběru jednoho z jevů F_j se snažíme pomocí dodatečného "experimentu"ve formě jevu A zlepšit náš odhad toho, který z jevů byl ve skutečnosti vybrán.

♣ Test nemoci je u 1 procenta zdravých falešně pozitivní, a u 10 procent nemocných falešně negativní. Podíl nemocných v populaci je 0,1 procenta. Kolik procent lidi ma pozitivní test? Jaká je pravděpodobnost, že pozitivní pacient je skutečně nemocný?

A..pozitivní test, B nemocný, P(B)=0.001.

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c) = 0.01089$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = 0.0826$$

♣ Máme férovou minci a neférovou minci která dává jenom H. Vybereme jednu z nich náhodně, a hodíme jí dvakrát s výsledkem HH. Jaká je pravděpodobnost, že mince je férová?

B...mince je férová, B^c mince je falešná, A padlo HH

$$P(B) = P(B^c) = \frac{1}{2}, P(A|B) = \frac{1}{4}, P(A|B^c) = 1$$
$$P(B|A) = \frac{\frac{1}{2}\frac{1}{4}}{\frac{1}{2}\frac{5}{4}} = \frac{1}{5}$$

♣ Máme tři krabice K_1, K_2, K_3, ve kterých je n_j bílých a m_j černých míčků. Z náhodně vybrané krabice vytáhneme bez vracení dva míčky, jsou bílý a černý. Jaká je pravděpodobnost, že byly taženy z krabice K_j?

A je výběr bílého a černého míčku, F_j je výběr krabice K_j v prvním kroku, tedy $P(F_j) = \frac{1}{3}$. $P(A|F_j) = \frac{n_j m_j}{\binom{n_j+m_j}{2}}$.

$$P(F_j|A) = \frac{P(A|F_j)P(F_j)}{\sum_{j=1}^{3} P(F_j)P(A|F_j)}$$

♣ Tři lovci se snaží ulovit jelena. První trefí jelena s pravděpodobností 0.3, druhý 0.4 a třetí 0.5. Jelen je zabit při jednom zásahu s pravděpodobností 0.4, při dvou zásazích 0.7 a při třech 0.9. Střelci vystřelili naráz. Jaká je pravděpodobnost, že jelen bude zabit. Byl-li zabit, jaká je pravděpodobnost, že se tak stalo jedním zásahem třetího lovce? Dvěma zásahy lovců jedna a dva?

Označíme $F_{(0,0,0)}, F_{(0,0,1)} \dots, F_{(1,1,1)}$ úplný systém jevů podle všech možností zásahu, tj. $F_{(0,0,0)}$ netrefil nikdo, $F_{(0,0,1)}$ trefil jenom třetí lovec atd. Platí $P(F_{(0,0,1)}) = 0.7 \cdot 0.4 \cdot 0.5$, $P(F_{(1,0,1)}) = 0.3 \cdot 0.4 \cdot 0.5$ atd.

Označme A jev jelen byl zabit. Podle věty o celkové pravděpodobnosti,

$$P(A) = 0P(F_{(0,0,0)}) + 0.2(P(F_{(0,0,1)}) + P(F_{(0,1,0)}) + P(F_{(1,0,0)}))+$$
$$+0.5(P(F_{(0,1,1)}) + +P(F_{(1,1,0)}) + P(F_{(1,0,1)})) + 0.9P(F_{(1,1,1)}) = 0.287$$

$$P(F_{(0,0,1)}|A) = \frac{P(A|F_{(0,0,1)})P(F_{(0,0,1)})}{\sum P(F_j)P(A|F_j)} = \frac{0.2 \cdot 0.7 \cdot 0.6 \cdot 0.5}{0.287} = 0.146...$$

$$P(F_{(1,1,0)}|A) = \frac{P(A|F_{(1,1,0)})P(F_{(1,1,0)})}{\sum P(F_j)P(A|F_j)} = \frac{0.5 \cdot 0.3 \cdot 0.4 \cdot 0.5}{0.287} = 0.104...$$

6. Aplikace v teorii her

♣ Ruinování hráče. Hráči A,B hrají hru skládající se z partií, kde A vítězí s pravděpodobností p, B vítězí s pravděpodobností $q = 1 - p$. Na začátku hry ma hráč A z korun, hráč B $a - z$ korun, po každé partii poražený předá vítězi 1 korunu. Jaká je pravděpodobnost zruinování hráče A,B?

Uvažme model nekonečného průběhu hry. Hru budeme modelovat jako nekonečné házení mincí, tedy binární strom s pravděpodobností větvení p, q. Symbol -1 znamená že vyhrál B, symbol 1 že vyhrál A. $\Omega_j = \{(a_1, \dots),\ a_i \in \{-1, 1\}\}$, $P(A(b_1, \dots, b_k)) = p^{\text{počet } 1} q^{\text{počet } -1}$. Jev zruinování hráče A, označíme $\tilde{A} \in \mathcal{F}$, splňuje $(a_1, a_2, \dots) \in \tilde{A}$ právě když

$\sum_{i=1}^{N} a_i$ poprvé vyleze z intervalu $(-z, a - z)$ pro nějaké N, a jeho hodnota je $-z$.

Tato množina je skutečně jev, a tedy má definovanou pravděpodobnost. Označme q_z pravděpodobnost, ze A bude zruinován. Podle věty o úplné pravděpodobnosti,

$$q_z = pq_{z+1} + qq_{z-1}$$

Řešení této diferenční rovnice (viz poznámka níže) má tvar (pro $p \neq q$)

$$q_z = A + B(\frac{q}{p})^z$$

resp. (pro $p = q$)

$$q_z = A + Bz$$

Dopočítáním konstant A, B, podle okrajových podmínek $q_0 = 1, q_a = 0$, dostaneme

$$q_z = \frac{(\frac{q}{p})^a - (\frac{q}{p})^z}{(\frac{q}{p})^a - 1}$$

Zruinování hráče B splňuje podobně

$$p_z = \frac{(\frac{p}{q})^a - (\frac{p}{q})^{a-z}}{(\frac{p}{q})^a - 1}$$

a lze ověřit $p_z + q_z = 1$, tedy hra skončí s pravděpodobností jedna.

Pro $p = q$ vyjde $q_z = 1 - \frac{z}{a}, p_z = \frac{z}{a}$.

♣ Poznámka. Prostor posloupností $\{x_n\}_{n=0}^{\infty}$ splňujících diferenční rovnici

$$x_{n+2} = Px_{n+1} + Qx_n,\ P, Q \in \mathbb{R},$$

je lineární prostor dimenze 2. Pokud má charakteristická rovnice $r^2 = Cr + D$ dva různé reálné kořeny r_1, r_2, potom každé řešení diferenční rovnice lze zapsat ve tvaru

$$x_n = Cr_1^n + Dr_2^n.$$

Pokud má charakteristická rovnice $r^2 = Cr + D$ dvojnásobný reálný kořen r, potom každé řešení diferenční rovnice lze zapsat ve tvaru

$$x_n = Cr^n + nDr^{n-1}.$$

♣ Dva hráči hází opakovaně jednou mincí, která dává $P(H) = p, P(O) = q$. První hráč zvítězí jakmile padne α H po sobě, druhý hráč jakmile padne β O po sobě. Jaká je pravděpodobnost vítězství prvního hráče?

Označme A jev α H po sobě padlo dříve než padlo β O po sobě.

Zavedeme úplný systém jevů $F_j,\ j = 0, \ldots, \alpha$:

$$F_j = A(b_1, \ldots, b_{j+1}),\ \ b_1 = \cdots = b_j = H, b_{j+1} = O,\ \ j = 0, \ldots, \alpha - 1$$

$$F_\alpha = A(b_1, \ldots, b_\alpha),\ \ b_1 = \ldots b_\alpha = H.$$

Označme U první padla H, V první padla O, tedy $V = F_0$ a $U = \cup_{j=1}^{\alpha} F_j$.

Označ $u = P(A|U), v = P(A|V)$. Platí

$$P(A) = P(A|V)P(V) + P(A|U)P(U) = qv + pu = P(A|V)P(V) + P(A \cap U).$$

$$P(A \cap U) = \sum_{j=0}^{\alpha} P(A \cap U|F_j)P(F_j) = P(A \cap U|V)P(V) + \sum_{j=1}^{\alpha} P(A \cap U|F_j)P(F_j)$$

Přitom $P(A \cap U|V) = 0$, $P(F_j) = p^j q,\ 1 \leq j < \alpha$, $P(F_\alpha) = p^\alpha$.

Dále $P(A \cap U|F_j) = P(A|V) = v$, $j = 1, \ldots, \alpha - 1$, $P(A \cap U|F_\alpha) = 1$.

$$P(A \cap U) = \sum_{j=1}^{\alpha-1} p^j qv + p^\alpha = p^\alpha + vqp\frac{1 - p^{\alpha-1}}{1 - p} = p^\alpha + pv(1 - p^{\alpha-1})$$

Tedy

$$u = P(A|U) = \frac{P(A \cap U)}{P(U)} = p^{\alpha-1} + v(1 - p^{\alpha-1}).$$

Podobným postupem, kdy úplný systém jevů závisí na počtu počátečních O, dojdeme v vyjádření

$$v = P(A|V) = u(1 - q^{\beta-1}).$$

Dostali jsme systém rovnic v neznámých u, v, který vyřešíme a vyjádříme

$$P(A) = pu + qv = p^{\alpha-1}\frac{1 - q^\beta}{p^{\alpha-1} + q^{\beta-1} - p^{\alpha-1}q^{\beta-1}}$$

Opačnou situaci výhry, jev B kdy padlo β O po sobě dřív než padlo α H po sobě dává

$$P(B) = q^{\beta-1}\frac{1 - p^\alpha}{p^{\alpha-1} + q^{\beta-1} - p^{\alpha-1}q^{\beta-1}}$$

Protože $x + y = 1$, pravděpodobnost, že hra skončí v konečném čase vítězstvím jednoho hráče je 1.

7. Náhodná procházka

Model náhodné procházky na množině $\mathbb{Z}$. V čase 0 se nacházíme v počátku (nebo obecněji v jiném bodě), a každou další časovou jednotku, měřenou přirozenými čísly, se s pravděpodobností $\frac{1}{2}$ pohneme buď doleva nebo doprava. Budeme analyzovat dlouhodobé chování typické trajektorie, zejména návraty do počátku.

Grafické vyjádření pozice v case n je (n, x), $n \in \mathbb{N}_0, x \in \mathbb{Z}$.

Definice 7.1. *Procházka z počátku do bodu (n, x) je n-tice $(s_0, \dots, s_n)$, $s_0 = 0, |s_{k+1} - s_k| = 1$, $s_n = x$. Analogicky, nekonečná procházka je $(s_0, \dots)$ splňující totéž.*

Označíme-li $\Omega = \{(a_i)_{i=1}^{\infty}, a_i \in \{-1, 1\}\}$, máme bijekci mezi posloupností kódující jednotlivé kroky a nekonečnou procházkou $(s_0, \dots)$ $(a_i)_{i=1}^{\infty}$ z Ω, $s_k = \sum_{i=1}^{k} a_i$.

Počet procházek z počátku $(0, 0)$ do (n, x) je roven

$$N_{n,x} = \binom{n}{\frac{n-x}{2}}.$$

Pravděpodobnost, že procházka z počátku projde přes (n, x) je rovna

$$p_{n,x} = \frac{N_{n,x}}{2^n} = \binom{n}{\frac{n-x}{2}} 2^{-n}.$$

Lemma 7.2. *Princip zrcadlení*

Nechť $A = (n, x), B = (m, y)$, $m > n \geq 0, x, y > 0$, $A' = (n, -x)$ je reflexe A. Potom počet navzájem různých procházek z A do B, které se dotknou (nebo protnou) horizontální osu je roven počtu všech procházek z A' do B. Jejich počet je roven $N_{m-n,x+y} = \binom{m-n}{\frac{m-n+x-y}{2}}$.

Věta 7.3. *Nechť $n, x \in \mathbb{N}^+$. Potom počet procházek $(s_0, \dots, s_n)$ z počátku do (n, x) pro které $s_1, \dots, s_n > 0$ je roven*

$$N_{n-1,x-1} - N_{n-1,x+1} = \frac{x}{n}\binom{n}{\frac{n-x}{2}} = \frac{x}{n} N_{n,x}.$$

Důkaz. Je zřejmé, že tento počet odpovídá počtu procházek z bodu $(1, 1)$ do (n, x) které se nedotknou horizontální osy, a tedy s použitím principu zrcadlení dostaneme náš výraz. Podrobně:

Polož $n = p + q, x = p - q$. Potom $N_{n,x} = \binom{p+q}{p}$, a naše formule je ekvivalentní

$$\binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \frac{p-q}{p+q}\binom{p+q}{p}$$

□

Budeme studovat pravděpodobnost návratu do počátku po $2v$ krocích.

Označme f_{2v} pravděpodobnost, že první návrat do počátku se uskutečnil po $2v$ krocích, tedy $s_j \neq 0, j = 1, \dots 2v - 1$, $s_{2j} = 0$, u_{2v} pravděpodobnost, že po $2v$ krocích došlo k návratu ($u_0 = 1$).

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \cdots + f_{2n} u_0$$

Použitím Stirlingovy formule $n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$,

$$u_{2v} = \binom{2v}{v} 2^{-2v} \sim \frac{1}{\sqrt{\pi v}}$$

Lemma 7.4. *Pravděpodobnost, že nedojde k návratu až do času $2n$ je stejná jako že dojde k návratu v čase $2n$.*

$$P\{s_j \neq 0, j = 1, \dots 2n\} = P(s_{2n} = 0) = u_{2n}$$

Důkaz. Chceme dokázat, že

$$P\{s_j > 0, j = 1, \dots 2n\} = P\{s_j < 0, j = 1, \dots 2n\} = \frac{1}{2} u_{2n}.$$

Je zřejmé, že z Věty 7.3,

$$P\{s_j > 0, j = 1, \dots 2n\} = \sum_{r=1}^{\infty} P\{s_j > 0, j = 1, \dots 2n-1, s_{2n} = 2r\} =$$

$$= \sum_{r=1}^{\infty} \frac{N_{2n-1,2r-1} - N_{2n-1,2r+1}}{2^n} =$$

$$\sum_{r=1}^{\infty} \frac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1}) = \frac{1}{2} p_{2n-1,1} = \frac{1}{2} u_{2n}$$

□

Věta 7.5. *Pravděpodobnost prvního návratu náhodné procházky do počátku v čase $2n$ je rovna $f_{2n} = \frac{1}{2n-1} u_{2n} \sim \frac{1}{2n-1} \frac{1}{\sqrt{\pi n}}$.*

Důkaz. Jev, že první návrat do počátku se odehrál v čase $2n$ je stejný jako rozdíl jevů kdy k návratu nedošlo až do času $2(n-1)$ mínus jev k návratu nedošlo až do času $2n$, tedy

$$f_{2n} = u_{2n-2} - u_{2n} = \binom{2(n-1)}{n-1} 2^{-2(n-1)} - \binom{2n}{n} 2^{-2n} =$$

$$= \frac{2^2 n \cdot 2(n-1) \dots n - 2n \dots (n+1)}{(n-1)! n 2^{2n}} = \frac{1}{2n-1} u_{2n}$$

□

Věta 7.6. *Pravděpodobnost návratu náhodné procházky do počátku je rovna jedné.*

Důkaz. Platí $f_2 + f_4 + \cdots + f_{2n} = 1 - u_{2n}$, takže pravděpodobnost návratu do počátku (pro nekonečnou procházku) po konečně mnoha krocích je rovna jedné. □

Věta 7.7. *Zákon arkussinu pro poslední návštěvu*

Pravděpodobnost, že poslední návrat do počátku před krokem $2n$ je v kroku $2k$ je dána

$$\alpha_{2k,2n} = u_{2k} u_{2n-2k}$$

Důkaz. Naše situace je $s_{2k} = 0, s_j \neq 0, j > 2k$. počet takových procházek je roven počtu procházek z $(0,0)$ do $(2k,0)$ krát počet procházek z $(0,0)$ délky $2n-2k$, které neprotínají počátek, tedy podle lematu máme výsledek. □

Položme

$$f(x) = \frac{1}{\pi \sqrt{x(x-1)}},\ 0 < x < 1$$

Podle naší aproximace, dostaneme

$$\alpha_{2k,2n} \sim \frac{1}{n} f(\frac{k}{n})$$

Integrací

$$\sum_{k < xn} \alpha_{2k,2n} \simeq \frac{2}{\pi} \arcsin \sqrt{x}$$

Pravá strana nezávisí na n.

Bez důkazu:

Věta 7.8. *Zákon arkussinu pro délku pobytu*

Pravděpodobnost, že množina $\{j, 0 \leq j \leq 2n, s_j > 0\}$ *má* $2k$ *prvků je rovna* $\alpha_{2k,2n}$.

Důsledek

Věta 7.9. *Zákon arkussinu pro délku pobytu*

Pro $0 < x < 1$ *pravděpodobnost, že* $\leq xn$ *časových jednotek je stráveno na pozitivní straně a* $\geq (1-x)n$ *na negativní straně má pro* $n \to \infty$ *limitu* $\frac{2}{\pi} \arcsin \sqrt{x}$.

Tyto výsledky jsou paradoxní a ve zdánlivém rozporu se zákonem velkých čísel. Numericky, pro velká n s pravděpodobností 0.2 částice stráví 0.976 času na jedné straně. S pravděpodobností 0.1 stráví 0.994 času na jedné straně.

8. Náhodná veličina

Definice 8.1. *Náhodná veličina je reálná funkce $X : \Omega \to I\!R$, která splňuje následující vlastnost: Pro každý interval $B \subset I\!R$ množina $\{t : t \in \Omega, X(t) \in B\}$ patří do $\mathcal{F}$.*

Poznámka. Podmínka z definice je technického charakteru, a je splněna pro všechny funkce, které můžeme potkat v aplikacích. Pokud je Ω spočetná množina, je splněna vždy.

Příklady:

Definice 8.2. *Diskrétní veličina $X : \Omega \to I\!R$ je taková, jejíž obor hodnot je konečný (nebo spočetný) $\{x_1, x_2, \dots\}$. Označíme $p_X(x_i) = P(x = x_i)$, kde $p_X(x_i) > 0$, a platí $\sum p_X(x_i) = 1$. Přiřazení $x_i \to p_X(x_i)$ se říká pravděpodobnostní rozdělení diskrétní veličiny X.*

Při práci s jednou diskrétní veličinou X budeme často vynechávat dolní index X, t.j. $p(x_i) = p_X(x_i)$.

Příklady:

Bernoulliho veličina, pro $0 < p < 1$. $\Omega = \{0, 1\}$, $P(1) = p, P(0) = q = 1 - p$, $X(0) = 0, X(1) = 1$.

Hodím minci 10krát, X je počet hlav. Zvolíme $\Omega = \{(a_1, .., a_{10}) : a_i \in \{0, 1\}\}$.

Házení kostkou dokud nepadne H, X je počet hodů. $\Omega = \{(a_1, ..., a_{n+1}); a_i = O, a_{n+1} = H\}$.

Spojitá funkce na $[0, 1]$ s Lebesgueovou mírou.

Náhodný bod ve čtverci $[0, 1] \times [0, 1]$, X je vzdálenost od počátku.

Definice 8.3. *Distribuční funkce $F_X : I\!R \to [0, 1]$ náhodné veličiny X je definovaná následovně*

$$F_X(t) = P(X \leq t)$$

Distribuční funkce předchozích příkladů.

Tvrzení 8.4. *Distribuční funkce ma následující vlastnosti:*
1. $0 \leq F_X(t) \leq 1$
2 je neklesající

$$3. \lim_{t \to -\infty} F_X(t) = 0, \lim_{t \to \infty} F_X(t) = 1.$$

4. F_X je zprava spojitá (tj. $F_X(t) = \lim_{s \to t^+} F_X(s)$) a má nejvýše spočetně mnoho bodů nespojitosti.

Tvrzení 8.5. *Mezi pravděpodobnostním rozdělením a distribuční funkcí diskrétní veličiny je vzájemně jednoznačný vztah*

$$F_X(t) = \sum_{x_i \leq t} p_X(x_i)$$

Věta 8.6. *Náhodná veličina $X : (\Omega, P) \to I\!R$ přenáší míru P z prostoru (Ω, P) na pravděpodobnostní míru μ na prostoru $(I\!R, \mu)$ podle formule*

$$\mu(A) = P(\{\omega \in \Omega, X(\omega) \in A\}) = P(X^{-1}(A)), \quad A \subset I\!R$$

♣ Pokud X je diskrétní veličina s oborem hodnot $\{x_1, \dots x_n\}$, potom μ bude diskrétní míra

$$\sum_{i=1}^{n} p(x_i)\delta_{x_i}$$

♣ Vztah distribuční funkce F_X a obrazu míry μ je následující:

$$F_X(t) = \mu((-\infty, t])$$

Lze ukázat, že míra μ je již funkcí F_X určena jednoznačně. Pro případ množiny $A = (a, b]$ je to vidět hned,

$$\mu(A) = F_X(b) - F_X(a)$$

a odtud lze dále odvodit, že μ je již určena jednoznačně.

Věta 8.7. *Nechť $X : (\Omega, P) \to \mathbb{R}$ je náhodná veličina, μ je obrazem míry P. Potom veličina X má stejnou distribuční funkci jako náhodná veličina $y(x) = x$ na prostoru $(\mathbb{R}, \mu)$.*

Tato jednoduchá věta ma zásadní význam, neboť nám umožňuje při studiu náhodných veličin kdykoliv nahradit Ω množinou $\mathbb{R}$ a pracovat s odpovídající distribuční funkcí.

9. Střední hodnota náhodné veličiny-diskrétní případ

Hlavní charakteristikou náhodné veličiny je její střední hodnota, která je definována

$$E(X) = \int_{\Omega} X dP.$$

Tato hodnota je fakticky integrálem veličiny podle pravděpodobnostní míry P, ale takto obecná definice by vyžadovala vybudování celé teorie integrace podle obecné míry P. Proto se omezíme na dva hlavní případy, tedy diskrétní a později spojitou pravděpodobnostní veličinu, kdy lze střední hodnotu vyjádřit s použitím konečných součtů, resp. obvyklých integrálů v $\mathbb{R}$.

♣ Střední hodnota obecně nemusí existovat, ani pro spojité nebo diskrétní (ale nekonečně hodnotové) proměnné.

Věta 9.1. *Střední hodnota diskrétní veličiny X s hodnotami v množině čísel $\{x_1, \dots\}$ splňuje*

$$E(X) = \sum x_i P(X = x_i) = \sum x_i p(x_i).$$

Příklady na střední hodnotu:

Hod nesymetrickou minci, 1 padne s pravděpodobností p, 0 s pravděpodobností $(1-p)$. $E(X) = p$.

Hodím dvěma kostkami, X je součet obou hodu, resp. součin, resp. rozdíl.

♣ Dva hráči se střídají v hodu kostkou. Pokud hodí stejné číslo, hra pokračuje dalším hodem. Vyhrává ten, kdo hodí vyšší číslo. S jakou pravděpodobností hra skončí po konečně mnoha cyklech? Najděte pravděpodobnostní rozdělení délky hry a jeho střední hodnotu.

Jeden cyklus hry je reprezentován dvojicí $(k, l) \in \{1, \dots, 6\}^2 = M$. Označíme $R = \{(k, k), k \in \{1, \dots, 6\}\}, Q = M \setminus P$. Náš model průběhu hry bude následující.

$$\Omega = \{(a_1, ..., a_{n-1}, b_n),\ a_j \in R, b_n \in Q,\ n \in \mathbb{N}\} \cup \{(a_1, \dots),\ a_j \in P\}.$$

Označme $A_n = \{(a_1, ..., a_{n-1}, b_n),\ a_j \in R, b_n \in Q\}$, $A_\infty = \{(a_1, \dots),\ a_j \in P\}$. Je zřejmé, že $A_n, n \in \mathbb{N} \cup \{\infty\}$ je úplný systém jevů na Ω (v tom smyslu, že jsou disjunktní a pokrývají Ω). Platí

$P(A_1) = \frac{6^2-6}{6^2}$, $P(A_2) = \frac{6}{6^2}\frac{6^2-6}{6^2}$ etc. $P(A_n) = \frac{1}{6^{n-1}}\frac{5}{6}$. Tedy, (jak víme $\sum_{n=0}^{\infty} q^n = \frac{1}{(1-q)}$)

$$P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \frac{5}{6^n} = 1.$$

Odtud plyne, že $P(A_\infty = 0)$ a hra skončí v konečném čase s pravděpodobností 1. Délku hry představuje náhodná veličina $X : \Omega \to \mathbb{N}$,

$$X(\{(a_1, ..., a_{n-1}, b_n)\}) = n$$

Platí $X(A_n) = n$, tedy $p_X(n) = P(A_n)$,

$$E(X) = \sum_{n=1}^{\infty} nP(A_n) = \sum_{n=1}^{\infty} n \frac{1}{6^{n-1}} \frac{5}{6} = \frac{6}{5}$$

(Pomocí známé formule $\sum_{n=0}^{\infty} nq^{n-1} = \frac{1}{(1-q)^2}$ dostaneme výsledek $\frac{6}{5}$).

Tvrzení 9.2. *Nechť $\{X_i\}_{i=1}^n$ jsou diskrétní náhodné veličiny na Ω, $a_i \in I\!R$. Potom platí*

$$E(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i E(X_i).$$

Důkaz. Pro dvě konečné veličiny, kde $A_i = \{\omega : X = a_i\}, B_j = \{\omega : Y = b_j\}$, zavést zjemnění $C_{i,j} = \{\omega : X = a_i, Y = b_j\}$.

♣ Bernoulliho (nebo též binomické) rozdělení $X = Binom(n,p)$ jako rozdělení součtu stejně rozdělených Y_j, s hodnotou 1 o pravděpodobnosti p a 0 o pravd. $(1-p)$, $X = \sum_{i=1}^n Y_i$. Tedy $\Omega = \{(a_1,\dots,a_n), a_i = 0,1, P(1) = p\}$, $Y_i(\omega) = a_i$. Binomické rozdělení $X = Binom(n,p)$ počítá počet úspěchů n nezávislých hodů mincí s pravděpodobností p. (tedy součet hodu nesymetrickou mincí, 1 padne s pravděpodobnosti p, 0 s pravděpodobnosti $(1-p)$). Rozdělení splňuje

$$b(j;n,p) = P(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$$

Protože Bernoulliho rozdělení $X = Binom(n,p)$ je možno modelovat jako součet stejně rozdělených Y_j, s $E(Y_j) = p$, dostaneme

$$E(X) = E(Y_1) + \cdots + E(Y_n) = np$$

V domě, který má n poschodí je výtah. V přízemí nastoupí p osob, každá z nich nezávisle na ostatních vystoupí se stejnou prvděpodobností v kterémkoli patře. Jaká je střední hodnota počtu zastávek výtahu?

$\Omega = \{1,\dots,n\}^p$. Zavedeme indikátorové náhodné veličiny, $Y_i = 1$ pokud výtah zastaví v i-tém patře, $Y_i = 0$ pokud nezastaví. Celkový počet zastavení výtahu je $Y = \sum_1^n Y_i$. Pravděpodobnost že nikdo nevystoupí v i-tém patře je $(1-\frac{1}{n})^p$, tedy

$$P(Y_i = 1) = 1 - (1-\frac{1}{n})^p$$

Jelikož $EY_i = P(Y_i = 1)$,

$$EY = n(1-(1-\frac{1}{n})^p)$$

Standartní postup spočívá ve výpočtu $EY = \sum kP(Y = k)$, ale tento výpočet by byl složitější. (platí $P(Y = k) = \frac{1}{n^p}(\binom{n}{k}\sum_{j=0}^{k-1}(-1)^j\binom{k}{j}(k-j)^p)$)

Definice 9.3. *Rozptyl (variance) diskrétní náhodné veličiny X je definován*

$$\sigma_X^2 = D(X) = E(X - (EX))^2.$$

Výraz σ_X se nazývá střední kvadratická odchylka (nebo směrodatná odchylka) od střední hodnoty.

Tvrzení 9.4. *platí $D(X) = E(X^2) - (EX)^2$.*

Najděte distribuční funkci, střední hodnotu a rozptyl veličiny:

1. Hod nesymetrickou minci, 1 padne s pravděpodobnosti p, 0 s pravděpodobnosti $(1-p)$. $D(X) = p(1-p)$.

$$2. P(X = -2) = \frac{1}{3}, P(X = 1) = \frac{4}{9}, P(X = 2) = \frac{2}{9}$$

Definice 9.5. *Kovariance dvou diskrétních náhodných veličin na Ω je definována jako*

$$Cov(X,Y) = E((X-EX)(Y-EY)) = E(XY) - E(X)E(Y).$$

Veličiny pro které je $Cov(X,Y) = 0$ se nazývají nekorelované.

Tvrzení 9.6. *Dva jevy A, B jsou nezávislé právě když jim odpovídající indikátorové funkce jsou nekorelované.*

Tvrzení 9.7. *Platí*

$$D(X_1 + \cdots + X_n) = \sum_{j=1}^{n} D(X_j) + 2\sum_{j<k} Cov(X_j, X_k).$$

Speciálně, pokud X_k jsou po dvou nekorelované, potom platí

$$D(X_1 + \cdots + X_n) = \sum_{j=1}^{n} D(X_j).$$

♣ Příklad. Bernoulliho rozdělení $X = Binom(n,p)$ jako součet stejně rozdělených Y_j, s hodnotou 1 o pravděpodobnosti p a 0 o pravd. $(1-p)$. Tedy $\Omega = \{(a_1,\dots,a_n), a_i = 0,1, P(1) = p\}$, $Y_i(\omega) = a_i$.

Dokažte, že Y_j jsou po dvou nekorelované, tedy $Cov(Y_j, Y_k) = 0$. Použitím této vlastnosti dostaneme

$$D(X) = D(Y_1) + \cdots + D(Y_n) = np(1-p)$$

♣ Příklad. Na Vánoční besídku přišlo n dětí. Každé dítě přineslo stejně zabalený dárek a dalo ho do koše. Na konci besídky si děti náhodně vyberou dárky z koše. Jaká je střední hodnota počtu všech dětí, které si domu odnášejí svůj vlastní dárek? Jsou jevy že i a j dítě si odnášejí své vlastní dárky nezávislé?

Ω jsou všechny permutace $\{1,\dots,n\}$, tedy rozdělení dárků. X_i je náhodná veličina rovna 1 pokud si dítě odnáší svůj dárek, jinak 0. Protože počet permutaci fixujících j je roven $(n-1)!$, je $P(X_i = 1) = \frac{1}{n}$. Střední hodnota je tedy rovna střední hodnotě

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

Poznámka; jevy že i a j dítě si odnášejí své vlastní dárky nejsou nezávislé. Tedy ani veličiny X_i nejsou nezávislé.

$$D(X_i) - EX_i^2 - (EX_i)^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$$

$$P(X_iX_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

$$Cov(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

$$D(S) = n\frac{n-1}{n^2} + n(n-1)\frac{1}{n^2(n-1)} = 1$$

10. Náhodné veličiny-spojitý případ

♣ Náhodné veličiny lze studovat jako obecné funkce na Ω, avšak z technických důvodů se omezujeme na dva hlavní případy, tedy diskrétní a spojité.

Definice 10.1. *Budeme říkat, že veličina $X : \Omega \to \mathbb{R}$ je spojitá, pokud její distribuční funkce F_X je spojitá a po částech spojitě diferencovatelná, tedy existuje po částech spojitá nezáporná funkce $f_X : \mathbb{R} \to \mathbb{R}$ (nazývaná hustota pravděpodobnostního rozdělení X) taková, že až na konečně mnoho bodů $t \in \mathbb{R}$ platí*

$$F_X'(t) = f_X(t).$$

Tvrzení 10.2. *Distribuční funkce spojité náhodné veličiny X splňuje*

$$F_X(t) = P(X \le t) = \int_{-\infty}^{t} f_X(s)ds.$$

Obraz $(\mathbb{R}, \mu)$ míry (Ω, P) splňuje

$$\mu(B) = P(X \in B) = \int_B f_X(x)dx,\ B \subset \mathbb{R}.$$

Střední hodnota splňuje

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx.$$

Rozptyl (variance) splňuje

$$\sigma_X^2 = D(X) = E(X-(EX))^2 = E(X^2)-E(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx - (\int_{-\infty}^{\infty} x f_X(x)dx)^2.$$

Tvrzení 10.3. *Nechť $\{X_i\}_{i=1}^n$ jsou náhodné veličiny na Ω, $a_i \in \mathbb{R}$. Potom platí*

$$E(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i E(X_i).$$

♣ Stejnoměrně rozdělení na intervalu $[a, b]$.
Hustota: $f(x) = \frac{1}{b-a}$ uvnitř intervalu, jinak 0.
$E(X) = \frac{a+b}{2}$, $D(X) = \frac{(b-a)^2}{12}$.

Rozřežeme náhodně úsečku $[0, 1]$ na dvě části. Najděte distribuční funkci a střední délku kratší části. $\Omega = [0, 1]$.

Najděte distribuční funkci a střední vzdálenost náhodného bodu v kruhu od počátku.

Na úsečce jsou vybrány náhodně dva body, X je jejich vzdálenost. Najděte distribuční funkci a střední hodnotu X. $\Omega = [0, 1]^2$.

Najděte distribuční funkci a střední výšku náhodného bodu na obvodu kola nad zemí.

Na kružnici jsou vybrány náhodně dva body, X je jejich vzdálenost. Najděte distribuční funkci a střední hodnotu X. $\Omega = [0, 2\pi]^2$, nebo $\Omega = [0, 2\pi]$.

11. Některé důležité typy rozdělení

♣ Binomické rozdělení $X = Binom(n, p)$ počítá počet úspěchů n nezávislých hodů mincí s pravděpodobností p, tedy

$$b(j; n, p) = P(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$$

Pro n velké, p malé a $\lambda = np$ rozumné velikosti chceme odhadnout tyto hodnoty. Takže pro pevné k a dostatečně velké n platí

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{(n-k+1)p}{kq} \approx \frac{np}{kq} \approx \frac{\lambda}{k}.$$

Pro výpočet těchto hodnot pro $\lambda = np$: použitím logaritmu a Taylorovy aproximace,

$$b(0, n, p) = (1-p)^n = (1 - \frac{\lambda}{n})^n \approx e^{-\lambda}.$$

Indukcí spojením předchozích odhadů vyjde pro pevné p-malé, n-velké, a k pevné

$$b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

Binomické rozdělení a Poissonovo rozdělení (následuje) jsou blízká (pro $\lambda = np$) ve smyslu distribuční funkce pokud platí $n \geq 20$ a $p \leq 0.05$ nebo $n \geq 100$ a $np \leq 10$.

Platí $E(X) = np$ (suma pravděpodobnostních veličin, i.e . E je suma E), $D(X) = np(1-p)$. (Užitím nezávislosti sčítanců a nulové kovariance)

♣ Poissonovo rozdělení $Poiss(\lambda), \lambda > 0$ má rozdělení

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

Označím jako $p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$. použitím Taylorova rozvoje $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$, dostaneme $\sum_{k=0}^{\infty} p(k; \lambda) = 1$.

$$E(X) = \sum_{k=1}^{\infty} kp(k; \lambda) = \lambda \sum_{k=1}^{\infty} p(k-1; \lambda) = \lambda$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p(k; \lambda) = \lambda \sum_{k=1}^{\infty} kp(k-1; \lambda) =$$

$$\lambda \sum_{k=1}^{\infty} (k-1)p(k-1; \lambda) + \lambda \sum_{k=1}^{\infty} p(k-1; \lambda) = \lambda^2 + \lambda$$

Tedy $D(X) = \lambda$.

Poissonovo rozdělení je model pro rozdělení počtu náhodných událostí, které jsou stejnoměrně rozdělené v čase, a navzájem nezávislé. Přitom každá z nich má malou pravděpodobnost, ale jejich počet je velký. Typické příklady: počet telefonních hovorů v centrále (mnoho volajících, ale každý jen málo kdy), internetový provoz, původní příklad byl počet zabitých v armádě v důsledku kopnutí koněm.

Odvození Poissonova rozdělení. Představme si jednotkový časový interval rozdělený na n podintervalů stejné délky. Množinu konečně mnoha bodů na intervalu budeme považovat za výsledek náhodného procesu. Každý podinterval má stejnou pravděpodobnost p_n, že bude obsahovat některý z těchto bodů, a jejich nezávislost znamená, že budeme používat Bernoulliho proměnné.

Předpokládáme tedy, že pravděpodobnost toho že přesně k podintervalů je neprázdných je $b(k, n, p_n)$. Nyní nechme $n \to \infty$. Pravděpodobnost, že celý interval neobsahuje žádný bod musí mít konečnou limitu, tedy posloupnost $b(0, n, p_n) = (1 - p_n)^n$ musí mít limitu. Tedy $\log(1 - p_n)^n = -np_n + o(np_n) \to -\lambda \in \mathbb{R}$. Tedy

$$b(0, n, p_n) = \binom{n}{0} {p_n}^0 (1 - p_n)^n \to e^{n \log(1-p_n)} \to e^{-np_n} \to e^{-\lambda}$$

$$b(k, n, p_n) = \binom{n}{k} {p_n}^k (1 - p_n)^{n-k} \to \frac{n^k}{k!} {p_n}^k (1 - p_n)^{n-k} =$$
$$= \frac{\lambda^k}{k!} (1 - p_n)^n (1 - p_n)^{-k} \to \frac{\lambda^k}{k!} e^{-\lambda}$$

Pro aplikace, pravděpodobnost nalezení k bodu v intervalu délky t bude rovna

$$p(k; \lambda t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

♣ Předpokládejme, že na ústřednu přijde průměrně 5 hovorů za hodinu. Jaká je pravděpodobnost, že na ústřednu přijde aspoň jeden hovor? $\lambda = 5$. Tedy $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-5}$.

♣ Předpokládejme, že ulicí projede za hodinu průměrně λ automobilů. S pravděpodobností p auto překračuje rychlostní limit. Najděte pravděpodobnostní rozdělení počtu pokut udělených za hodinu, pokud bude prováděno měření rychlosti.

Rozdělení počtu automobilů za hodinu je Poissonovým rozdělením $Poiss(\lambda)$ (podle informace o střední hodnotě). Označme A_n jev projelo n automobilů, a B_j jev bylo uděleno j pokut (tedy bylo j překročení rychlosti).

Víme, že

$$P(A_n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, \ldots$$

$$P(B_j | A_n) = \binom{n}{j} p^j (1 - p)^{n-j}, \quad j \geq n,$$

$$P(B_j) = \sum_{n=0}^{\infty} P(B_j | A_n) P(A_n) =$$

$$= \sum_{n=0}^{\infty} \binom{n}{j} p^j (1 - p)^{n-j} \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=0}^{\infty} \frac{n!}{j!(n-j)!} p^j (1 - p)^{n-j} \frac{\lambda^n}{n!} e^{-\lambda} =$$
$$= \frac{e^{-\lambda} p^j \lambda^j}{j!} \sum_{n=j}^{\infty} \frac{(1 - p)^{n-j} \lambda^{n-j}}{(n-j)!} = \frac{e^{-\lambda} p^j \lambda^j}{j!} e^{(1-p)\lambda} = \frac{e^{-p\lambda} (p\lambda)^j}{j!}$$

Je vidět, že výsledné rozdělení počtu pokut je $Poiss(p\lambda)$.

♣ Exponenciální rozdělení: parametr $\lambda > 0$, hustota je $f(x) = \lambda e^{-\lambda x}, x \geq 0$ jinak 0.

platí:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}, x \geq 0$$

$$E(X) = \int_0^{\infty} \lambda t e^{-\lambda t} dt = [-te^{-\lambda t}]_0^{\infty} - \int_0^{\infty} -e^{-\lambda t} dt = -[\frac{1}{\lambda} e^{-\lambda t}]_0^{\infty} = \frac{1}{\lambda},$$

$$E(X^2) = \int_0^\infty \lambda t^2 e^{-\lambda t} dt = [-t^2 e^{-\lambda t}]_0^\infty + \int_0^\infty 2te^{-\lambda t} dt = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2},$$

Tedy $D(X) = \frac{1}{\lambda^2}$.
$P(X \geq x) = e^{-\lambda x}$, odtud máme
$P(X \geq x + y | X \geq y) = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = P(X \geq x)$

X je vhodným modelem doby čekání na výskyt nějakého jevu (doba čekání na poruchu), např. může představovat životnost (délku života) žárovky, nebo částice (radioaktivní rozpad částic). Tedy zbývající čas čekání na událost má stejnou distribuční funkci jako na začátku.

♣ Pokud má veličina spojité rozdělení, a splňuje $P(X \geq x + y | X \geq y) = P(X \geq x)$, potom je nutně exponenciální.
Označme $u(x) = P(X \geq x)$, platí tedy

$$u(x + y) = u(x)u(y), \; u(0) = 1$$

$$u'(x) = \lim_{y \to 0} u(x) \frac{u(y) - 1}{y} = u(x)u'(0) = \lambda u(x)$$

Řešení je $u(x) = \lambda e^{-\lambda x}$.

Příklad: Žárovka ma střední dobu života 1000 hodin. Jaká je pravděpodobnost, že praskne mezi 1000 a 1500 hodinami provozu. $\lambda = \frac{1}{1000}$

$P(1000 < X < 1500) = P(X > 1000) - P(X > 1500) = e^{-\lambda 1000} - e^{-\lambda 1500} = e^{-1} - e^{-1.5}$ $\blacksquare$

♣ Normální rozdělení, parametry $\mu \in I\!\!R, \sigma > 0$, $N(\mu, \sigma^2)$ má hustotu
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Je třeba ověřit

$$\int_{I\!\!R} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

$E(X) = \mu$, $D(X) = \sigma^2$. Odvození pro případ $\sigma = 1, \mu = 0$. Je třeba užít trik s dvojným integrálem.

$$\left(\int_{I\!\!R} e^{\frac{-x^2}{2}} dx\right)^2 = \left(\int_{I\!\!R} e^{\frac{-x^2}{2}} dx\right)\left(\int_{I\!\!R} e^{\frac{-y^2}{2}} dy\right) = \int\int_{I\!\!R^2} e^{\frac{-x^2-y^2}{2}} dxdy.$$

V polárních souřadnicích $x = \rho \cos\phi, y = \rho \sin\phi$ dostaneme

$$\int\int_{I\!\!R^2} e^{\frac{-x^2-y^2}{2}} dxdy = \int_0^{2\pi} \int_0^\infty \rho e^{\frac{-\rho^2}{2}} d\rho d\phi.$$

Substitucí $\frac{\rho^2}{2} = u$ vyjde uvnitř $-\int_0^\infty e^{-u} du$, takže celkový integrál bude 2π, tedy

$$\int_{I\!\!R} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1.$$

Obecně, substituce $u = \frac{x-\mu}{\sigma}$:

$$\int_{I\!\!R} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{I\!\!R} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1.$$

Střední hodnota nula pro $\mu = 0$ plyne ze symetrie, pro varianci per partes $u = x, v = xe^{-\frac{x^2}{2}}$

$$\int_{I\!R} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = [-\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx]_{-\infty}^{\infty} + \int_{I\!R} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1.$$

Normální rozdělení je též aproximací binomického, pro velká n a rozumná p, ale jeho význam je větší-centrální limitní věta.

12. Transformace náhodných veličin

Následující tvrzení plyne okamžitě z měřitelnosti náhodných veličin.

Tvrzení 12.1. *Nechť X je náhodná veličina na Ω a $g : \mathbb{R} \to \mathbb{R}$ je spojitá funkce. Potom $Y = g \circ X$ je též náhodná veličina na Ω.*

Pro diskrétní veličiny X snadné. Příklad: X budiž veličina výsledku hodu kostkou, $g : \{1, \dots, 6\} \to \{0, 1\}$ podle parity.

Věta 12.2. *Nechť X je spojitá náhodná veličina s hustotou f_X, $g : \mathbb{R} \to \mathbb{R}$ spojitá funkce. Potom platí*

$$E(g \circ X) = \int\limits_{-\infty}^{\infty} g(t) f_X(t) dt$$

Důkaz. Pro rostoucí g, $Y = g \circ X$:

$$F_Y(t) = P(Y \le t) = P(g \circ X \le t) = P(X \le g^{-1}(t)) = F_X(g^{-1}(t)).$$

$$f_Y(t) = F_Y(g^{-1}(t))' = (g^{-1}(t))' f_X(g^{-1}(t)) = \frac{1}{g'(g^{-1}(t))} f_X(g^{-1}(t)).$$

$$E(g \circ X) = \int\limits_{-\infty}^{\infty} t(g^{-1}(t))' f_X(g^{-1}(t)) dt = [subst. u = g^{-1}(t)] = \int\limits_{-\infty}^{\infty} g(u) f_X(u) du$$

$\square$

Příklad.
X obecné, $g(x) = rx + s$ lineární transformace pro $r > 0$. Vyjde

$$F_Y(y) = F_X(\frac{y-s}{r}), f_Y(y) = \frac{1}{r} f_X(\frac{y-s}{r}).$$

X normální rozdělení $N(\mu, \sigma^2)$, $Y = rX + s$ lineární transformace, potom Y je $N(r\mu + s, r^2\sigma^2)$.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$$f_Y(x) = \frac{1}{r\sigma\sqrt{2\pi}} e^{-\frac{(\frac{x-s}{r}-\mu)^2}{2\sigma^2}} = \frac{1}{r\sigma\sqrt{2\pi}} e^{-\frac{(x-s-r\mu)^2}{2r^2\sigma^2}}.$$

13. Náhodné vektory

Naším cílem bude nyní přejít od studia jedné náhodné veličiny ke studiu konečných množin (posloupností) náhodných veličin a jejich vzájemných vztahů.

Definice 13.1. *Pro zadaný pravděpodobnostní prostor* (Ω, Σ, S) *a náhodné veličiny* $X_j : \Omega \to I\!R$ *definujeme náhodný vektor jako zobrazeni* $X = (X_1, \dots, X_n) : \Omega \to I\!R^n$.

Věta 13.2. *Náhodný vektor*$X : (\Omega, P) \to I\!R^n$ *přenáší míru* P *z prostoru* (Ω, P) *na pravděpodobnostní míru* μ *na prostoru* $(I\!R^n, \mu)$ *podle formule*

$$\mu(A) = P(\{\omega \in \Omega, X(\omega) \in A\}) = P(X^{-1}(A)), \quad A \subset I\!R^n$$

Definice 13.3. *Distribuční funkce* $F_X : I\!R^n \to I\!R$ *náhodného vektoru* $X = (X_1, \dots, X_n)$ *je definována následovně*

$$F_X(t_1, \dots, t_n) = P(\cap_{j=1}^n (X_j \le t_j))$$

♣ Vztah distribuční funkce F_X a obrazu míry μ je následující. Nechť $F_n : I\!R^n \to I\!R$ je pevně zvoleno. Označme operátor $\Delta_{a_i,b_i} : I\!R^n \to I\!R$, pro $a_i \le b_i$, definovaný formulí

$$\Delta_{a_i,b_i} F_n(x_1, \dots, x_n) = F_n(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots) - F_n(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots).$$

Jelikož

$$F_n(x_1, \dots, x_n) = \mu((-\infty, x_1] \times \dots \times (-\infty, x_n]),$$

potom platí

$$\Delta_{a_1,b_1} \dots \Delta_{a_n,b_n} F_n(x_1, \dots, x_n) = \mu((a, b]) \ge 0$$

kde $(a, b] = (a_1, b_1] \times \dots \times (a_n, b_n]$.

Věta 13.4. *Funkce* $F_n : I\!R^n \to I\!R$ *je distribuční funkcí náhodného vektoru právě když splňuje podmínky*

$$F_n : I\!R^n \to [0, 1]$$

$$F_n(x^k) \searrow F_n(x), \text{ pokud } x_k \searrow x$$

$$F_n(+\infty, \dots, +\infty) = 1, \ F_n(x_1, \dots, -\infty, \dots, x_n) = 0$$

$$\Delta_{a_1,b_1} \dots \Delta_{a_n,b_n} F_n(x_1, \dots, x_n) \ge 0$$

Distribuční funkce obsahuje samozřejmě celou informaci o každé ze složek vektoru. Marginálním rozdělením vektoru F_X se rozumí rozdělení F_{X_j} jednotlivých složek vektoru.

Tvrzení 13.5. *Nechť* $F_X : I\!R^n \to I\!R$ *je distribuční funkce náhodného vektoru* $X = (X_1, \dots, X_n)$. *Potom marginální rozdělení jednotlivých proměnných lze získat*

$$F_{X_j}(t_j) = \lim_{t_i \to \infty, i \ne j} F_X(t_1, \dots, t_n) = \lim_{t_i \to \infty, i \ne j} P(\cap_{j=1}^n (X_j \le t_j))$$

Definice 13.6. *Rozdělení pravděpodobnosti náhodného vektoru* X *pro diskrétní veličiny je nezáporná funkce*

$$p_X(a_1, \dots, a_n) = P(X_1 = a_1, \dots, X_n = a_n), a_j \in I\!R$$

Pro spojitě rozdělený vektor je to spojitá funkce (pravděpodobnostní hustota) $f_X : I\!R^n \to I\!R^+$ *s vlastností*

$$P((X_1,\dots,X_n)\in S)=\int\dots\int_S f_X(x_1,\dots,x_n)dx_1\dots dx_n,$$

♣ Poznámka. Spojitost hustoty f_X je poněkud příliš restriktivní podmínka. V obecnější formulaci by stačila integrovatelnost, avšak za účelem technického zjednodušení výkladu budeme používat spojité funkce, resp. funkce které jsou restrikcemi spojitých funkcí na jednoduché podmnožiny R^n, jako např. otevřené množiny apod.

Věta 13.7. *Pro spojitě rozdělený vektor platí*

$$f_X(a_1,\dots,a_n)=\frac{\partial^n}{\partial x_1\dots\partial x_n}F_X(a_1,\dots,a_n)$$

Důkaz.

$$F_X(a_1,\dots,a_n)=\int_{-\infty}^{a_n}\dots\int_{-\infty}^{a_1} f_X(x_1,\dots,x_n)dx_1\dots dx_n,$$

$$\frac{\partial F_X(a_1,\dots,a_n)}{\partial x_1}=\int_{-\infty}^{a_n}\dots\int_{-\infty}^{x_2} f_X(a_1,x_2,\dots,x_n)dx_2\dots dx_n,\quad etc$$

$$\frac{\partial^n}{\partial x_1\dots\partial x_n}F_X(a_1,\dots,a_n)=f_X(a_1,\dots,a_n)$$

□

Hustota marginálního rozdělení splňuje

$$f_{X_j}(a_j)=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f_X(x_1,x_2,\dots,a_j,\dots,x_n)dx_1\dots dx_n,$$

♣ Příklady

Nechť $\Omega=[0,1], X(t)=t, Y(t)=t$, resp. $X(t)=t, Y(t)=1-t$. Najděte distribuční funkci vektoru (X,Y). Nechť $\Omega=[0,1]^2, X(x,y)=x, Y(x,y)=y$, najděte distribuční funkci vektoru (X,Y). Porovnejte tyto distribuční funkce a take distribuční funkce jednotlivých veličin X,Y ve všech případech.

Věta 13.8. *Nechť $X=(X_1,\dots,X_n):(\Omega,P)\to\mathbb{R}^n$ je spojitě rozdělený náhodný vektor s hustotou f_X. Potom náhodný vektor $(x_1,\dots,x_n):\mathbb{R}^n\to\mathbb{R}^n$ definovaný na pravděpodobnostním prostoru $(\mathbb{R}^n,\mu)$, kde μ je obrazem míry P z Věty 13.2, má stejnou distribuční funkci jako X.*

Důkaz této věty je jednoduchý výpočet. Její význam spočívá v tom, že nám umožňuje nahradit vektor X konkrétnějším vektorem souřadnicových funkcí z $\mathbb{R}^n$. Z hlediska pravděpodobnosti jsou tyto vektory nerozlišitelné. Větu lze tedy použít pro jakékoliv výpočty. Například:

Věta 13.9. *Nechť $X=(X_1,\dots,X_n)$ je spojitě rozdělený náhodný vektor veličin definovaných na Ω s hustotou f_X, $g:\mathbb{R}^n\to\mathbb{R}$ je spojitá funkce. Potom $Y=g\circ X=g(X_1,\dots,X_n):\Omega\to\mathbb{R}$ je náhodná veličina, pro kterou platí: Distribuční funkce F_Y splňuje*

$$F_Y(t)=\int\dots\int_{\{(x,1,\dots,x_n):\,g(x_1,\dots,x_n)\le t\}} f_X(x_1,\dots,x_n)dx_1\dots dx_n,$$

$$E(Y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_1, \dots, t_n) f_X(t_1, \dots, t_n) dt_1 \dots dt_n$$

Speciálně, pro původní složky vektoru platí

$$E(X_j) = E(x_j \circ X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} t_j f_X(t_1, \dots, t_n) dt_1 \dots dt_n$$

14. Nezávislost náhodných veličin

Náš hlavní zájem se týká nezávislých veličin, nejdůležitějšího pojmu v teorii pravděpodobnosti.

Definice 14.1. *Náhodné veličiny obsažené v náhodném vektoru (X, Y) (resp. $X = (X_1, \dots, X_n)$) se nazývají nezávislé pokud platí*

$$F_{(X,Y)}(a, b) = F_X(a)F_Y(b), a, b \in \mathbb{R}.$$

$$F_X(a_1, \dots, a_n) = F_{X_1}(a_1) \dots F_{X_n}(a_n), a_j \in \mathbb{R}$$

Ekvivalentně,

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b), a, b \in \mathbb{R}.$$

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) = P(X_1 \leq a_1) \dots P(X_n \leq a_n), a_j \in \mathbb{R}$$

♣ Ekvivalentně, všechny jevy $X_j^{-1}(A_j)$ jsou nezávislé, pro libovolné množiny A_j. Toto lze dokázat postupem po složkách, s využitím faktu, že míra je určena hodnotami na intervalech.

♣ Důsledek je, že pokud $X_1, \dots, X_n$ jsou nezávislé, potom $g_1(X_1), \dots, g_n(X_n)$ jsou také nezávislé, pro libovolné transformace g_j veličin X_j. Podobně platí, že $g_1(X_1)$ a $g(X_2, \dots, X_n)$ jsou nezávislé.

Věta 14.2. *Nechť X_j jsou spojitě rozdělené náhodné veličiny s hustotami f_j, Potom*

$$f(a_1, \dots, a_n) = f_1(a_1) \dots f_n(a_n) : \mathbb{R}^n \to \mathbb{R}$$

je funkce hustoty na $\mathbb{R}^n$ taková, že proměnné $x_j : \mathbb{R}^n \to \mathbb{R}$ na tomto $\Omega = \mathbb{R}^n$ s mírou danou hustotou f, jsou nezávislé veličiny se stejným rozdělením jako X_j. Lze tedy pro každou konečnou množinu veličin (resp. rozdělení) zkonstruovat náhodný vektor, jehož složky jsou nezávislé s předepsanými rozděleními.

Takto sestrojeny náhodný vektor je přitom určen jednoznačně.

Tvrzení 14.3. *Pokud $X = (X_1, \dots, X_n)$ je spojitě rozdělený náhodný vektor s hustotou $f_X : \mathbb{R}^n \to \mathbb{R}$, jehož složky jsou nezávislé veličiny s hustotami f_j, potom platí*

$$f_X(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$$

Důkaz.

$$f_X(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X(x_1, \dots, x_n) = \prod_{j=1}^n \frac{\partial}{\partial x_j} F_j(x_j) = \prod_{j=1}^n f_j(x_j)$$

□

Tvrzení 14.4. *Pokud X, Y jsou nezávislé veličiny (pro důkaz předpokládejme spojitě rozdělené) potom*

$$E(XY) = E(X)E(Y)$$

Důkaz:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E(X)E(Y)$$

Pokud X, Y jsou nezávislé veličiny (pro důkaz předpokládejme spojitě rozdělené s hustotami f, g), potom jsou také nekorelované. Korelace je tedy rychlým testem na nezávislost veličin (nutná podmínka). Opačné tvrzení neplatí.

$$Cov(X,Y) = \int\int xyf_X(x)f_Y(y) - \int xf_X(x)\int yf_Y(y) = 0.$$

Tedy pro nezávislé veličiny tedy platí

$$D(X_1 + \cdots + X_n) = \sum_{j=1}^{n} D(X_j)$$

♣ $X_1, \ldots, X_n$ jsou stejně rozdělené nezávislé veličiny, $Z = \max_j\{X_j\}$ (resp. min). Najděte rozdělení Z. Distribuční funkce $Z = maxX_j$ je $F_Z(x) = F_X(x)^n$.

♣ Pro X, Y nezávislé veličiny a $Z = X + Y$ dostaneme vyjádření pro distribuční funkci:

Diskrétní případ

$$p_Z(c) = \sum_j p_X(c - b_j)p_Y(b_j).$$

kde b_j běží přes všechny nenulové hodnoty $p_Y(b_j)$.

Příklad dva hody kostkou.

♣ Ve spojitém případě pro distribuční funkci platí

$$F_Z(a) = \int\int_{x+y\le a} f_Z(x,y)dxdy.$$

Užitím nezávislosti

$$F_Z(a) = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_Z(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy.$$

♣ Diferencováním podle a, hustoty splňují:

$$f_Z(a) = F_Z'(a) = \int_{-\infty}^{\infty} F_X'(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy.$$

Příklad X, Y nezávislé veličiny s rozdělením $N(0,1)$, $Z = X + Y$ bude mit rozdělení $N(0,2)$.

$$f_Z(z) = \int_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-y)^2})(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2})dy = \int_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}})^2 e^{-\frac{1}{2}(2y^2-2yz+z^2)}dy$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{4}z^2}\int_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}})e^{-\frac{1}{2}(2y^2-2yz+\frac{1}{2}z^2)}dy$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}z^2} \int\limits_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}}) e^{-\frac{1}{2}(\sqrt{2}(y-z/2))^2)} dy$$

$$= \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{4}z^2} \int\limits_{-\infty}^{\infty} (\frac{1}{\sqrt{\pi}}) e^{-\frac{1}{2}(\sqrt{2}(y-z/2))^2)} dy = \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-\frac{z^2}{2\cdot 2}}$$

♣ Indukcí, pokud $X_1, \ldots, X_n$ jsou nezávislé veličiny s normálním rozdělením $N(\mu_j, \sigma_j^2)$, a c_j konstanty, potom $X = \sum_{j=1}^n c_j X_j$ má normální rozdělení $N(\mu, \sigma^2)$ kde $\mu = \sum_{j=1}^n c_j \mu_j$, $\sigma^2 = \sum_{j=1}^n c_j^2 \sigma_j^2$. Tento výsledek dokážeme s použitím charakteristických funkcí dále.

Věta 14.5. *Nechť $X_1, \ldots, X_n$ jsou nezávislé veličiny s normálním rozdělením $N(0, \sigma^2)$. Nechť $U = (u_{ij})_{i,j=1}^n$ je ortonormální matice,*

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

kde $Y = (Y_1, \ldots, Y_n)$ je nový náhodný vektor. Potom Y_j jsou nezávislé veličiny s normálním rozdělením $N(0, \sigma^2)$.

Důkaz. Pro pravděpodobnostní hustotu f vektoru $X = (X_1, \ldots, X_n)$ platí

$$f(x_1, \ldots, x_n) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{(x_1)^2}{2\sigma^2}} \ldots e^{-\frac{(x_n)^2}{2\sigma^2}} = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{\|x\|^2}{2\sigma^2}}$$

Funkce hustoty je rotačně symetrická, tedy $f(x) = f(Ux)$, $x \in I\!R^n$. Tedy, označíme-li μ míru na $I\!R^n$ s hustotou f, $A \subset I\!R^n$,

$$P(X \in A) = \mu(x \in A) = \int \ldots \int\limits_A f(x) d\mu =$$

$$= \int \ldots \int\limits_{U^T(A)} f(x) d\mu = P(X \in U^T(A)) = P(UX \in A)$$

Volbou $A_0 = \{(x_1, \ldots, x_n\} : \ x_1 \leq t_1, \ldots, x_n \leq t_n\}$ získáme:

$$F_X(t_1, \ldots, t_n) = P(X \in A_0) = P(UX \in A_0) = P(Y \in A_0) = F_Y(t_1, \ldots, t_n).$$

Tim jsme dokázali, že distribuční funkce vektoru Y je rovna distribuční funkci vektoru X, speciálně tedy Y_j jsou nezávislé veličiny s požadovaným rozdělením. □

Věta 14.6. *Nechť (X, Y) jsou nezávislé spojité veličiny s hustotami f, g. Potom $Z = XY$, resp. $Z = \frac{X}{Y}$ má hustotu*

$$h(y) = \int\limits_{-\infty}^{\infty} f(\frac{y}{x}) g(x) \frac{1}{|x|} dx$$

resp. za předpokladu $g(x) = 0, \ x < 0$,

$$h(y) = \int\limits_0^{\infty} x f(yx) g(x) dx$$

Důkaz. Z Věty 13.9, pro součin, použitím záměny pořadí integrace a diferencování,

$$F_Z(t) = \int_{-\infty}^{0} \int_{\frac{t}{x}}^{\infty} f(x)g(y)dydx + \int_{0}^{\infty} \int_{-\infty}^{\frac{t}{x}} f(x)g(y)dydx$$

$$f_Z(t) = \frac{dF_Z(t)}{dt} = -\int_{-\infty}^{0} f(x)g(\frac{t}{x})\frac{1}{x}dx + \int_{0}^{\infty} f(x)g(\frac{t}{x})\frac{1}{x}dx = h(t)$$

Podobně pro podíl.

□

15. Zákon velkých čísel

Věta 15.1. *Čebyševova nerovnost. Nechť X je náhodná veličina, $a > 0$. Potom platí*

$$P(|X - E(X)| \geq a) \leq \frac{1}{a^2} D(X).$$

Důkaz: označ $\mu = E(X)$, spojitý případ.

$$D(X) = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq a} (x-\mu)^2 f_X(x)dx$$
$$\geq \int_{|x-\mu|\geq a} a^2 f_X(x)dx = a^2 P(|X - E(X)| \geq a).$$

Nechť $X_1, \ldots, X_n$ jsou stejně rozdělené nezávislé veličiny, $\overline{X}_n = \frac{\sum_{j=1}^n X_j}{n}$ budeme značit náhodnou veličinu, která reprezentuje průměr.

$$D(\frac{X_1 + \cdots + X_n}{n}) = \sum D(\frac{1}{n}X_j) = \sum \frac{1}{n^2} D(X_j) = \frac{1}{n^2}\sum D(X_j) = \frac{D(X_j)}{n}.$$

Je vidět, že D klesá k nule pro rostoucí n.

Věta 15.2. *Zákon velkých čísel (slabý)*

Nechť $X_1, \ldots, X_n$ jsou stejně rozdělené nezávislé (stačí nekorelované) veličiny, jejichž střední hodnota je μ a variance σ^2. Potom platí pro každé $\varepsilon > 0$:

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Důkaz. platí $E(\overline{X}_n) = \mu$, $D(\overline{X}_n) = \frac{\sigma^2}{n}$. Podle Čebyševovy nerovnosti

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{1}{\varepsilon^2}\frac{\sigma^2}{n}.$$

Příklad. Hodíme férovou mincí 100 krát. Odhadněte pravděpodobnost, že počet hlav je v rozmezí $40 - 60$.

Model: X_j je Benoulliho veličina s hodnotami $0, 1$ s pravděpodobností $\frac{1}{2}$, výsledek 1 modeluje hlavu. Víme, že $\mu = \frac{1}{2}$, $\sigma^2 = \frac{1}{4}$. Potom $\overline{X}_{100}$ je veličina představující průměrnou hodnotu. Zajímá nás pravděpodobnost, že

$$P(\frac{4}{10} \leq \overline{X}_{100} \leq \frac{6}{10})$$

Opačný jev znamená. že $|\overline{X}_{100} - \frac{1}{2}| > \frac{1}{10}$. Podle odhadu Čebyševa máme

$$P(|\overline{X}_{100} - \frac{1}{2}| > \frac{1}{10}) \leq 10^2 \frac{\frac{1}{4}}{100} = \frac{1}{4}.$$

Pravděpodobnost, že výsledek bude v rozmezí 40-60 je tedy nejméně $1 - \frac{1}{4}$.

Silný zákon velkých čísel je tvrzení, že $|\overline{X}_n(\omega) - \mu| \to 0$ platí pro skoro všechna $\omega \in \Omega$.

Věta 15.3. *(refomulace pro distribuční funkci) Platí, že $F_{\overline{X}_n} \to F$ kde F je skoková distribuční funkce se skokem 1 v bode μ.*

Víme, že $E(\overline{X}_n) = \mu$, $D(\overline{X}_n) \to 0$. Při jiné normalizaci součtů $\sum D(X_j) = \sum (X_j - \mu)^2$ lze dosáhnout toho, aby $\frac{1}{\sqrt{n}} D(\sum X_j) = \sigma^2$. Tento přístup vede na centrální limitní větu.

16. Centrální limitní věta

Věta 16.1. *Centrální limitní věta.*

Nechť $X_1, \dots, X_n$ jsou nezávislé stejně rozdělené veličiny se střední hodnotou μ a variancí σ^2, $E|X_1|^3 < \infty$. Nechť distribuční funkce proměnné $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ má tvar

$$G_n(x) = P(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Potom $G_n(x) \to \Phi(x)$ konverguje v každém bodě pro $n \to \infty$.

Důkaz bude proveden níže s použitím charakteristických funkcí (Fourrierovy transformace).

Definice 16.2. *Charakteristická funkce (Fourierova transformace) náhodné veličiny X je definována*

$$\phi_X(t) = E(e^{itX}).$$

Ekvivalentně, pro X s hustotou f_X je vidět, že jde o Fourrierovu transformaci.

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx.$$

Některé důležité vlastnosti: ϕ je uniformně spojitá na $\mathbb{R}$.

$$\phi_X(0) = 1, |\phi_X(t)| \leq 1,$$
$$\phi_{cX+d}(t) = e^{itd}\phi_X(ct), c, d \in \mathbb{R}$$
$$\frac{d^n}{dt^n}\phi_X(t)|_{t=0} = i^n E(X^n), n \in \mathbb{N}.$$

Pokud $X_1, \dots, X_n$ jsou nezávislé veličiny,

$$\phi_{X_1+\dots+X_n}(t) = E(e^{it\sum_j X_j}) = E(\prod_j e^{itX_j}) = \prod_j E(e^{itX_j}) = \prod_{j=1}^{n} \phi_{X_j}(t).$$

Věta 16.3. *(věta o spojitosti) Nechť F je distribuční funkce veličiny s charakteristickou funkcí ϕ. Potom $F_n \to F$ konverguje slabě (tj. bodově v každém bodě spojitosti F) právě když $\phi_n \to \phi$ bodově v každém bodě.*

Tuto větu lze přeformulovat tak, že Fourrierova transformace je zobrazení, které má inverzi a které je v jistém smyslu spojité v obou směrech.

♣ Pokud $Z = N(\mu, \sigma^2)$, potom $\phi_Z(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$. Speciálně, pokud Z ma rozdělení $N(0,1)$ pak $\phi_Z(t) = e^{-\frac{t^2}{2}}$.

Důkaz. pro $N(0,1)$.

$$g(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{itx} dx$$

$$\frac{d}{dt} g(t) = \frac{1}{\sqrt{2\pi}} \frac{d}{dt} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{itx} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} (ix) e^{itx} dx = \frac{-i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d}{dx}(e^{-\frac{x^2}{2}}) e^{itx} dx$$

per partes

$$= [e^{-\frac{x^2}{2}+itx}]_{-\infty}^{\infty} - \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (e^{-\frac{x^2}{2}}) e^{itx} dx = -tg(t)$$

Tato diferenciální rovnice

$$g'(t) = f(t)g(t)$$

$$\frac{g'(t)}{g(t)} = (\ln g(t))' = f(t) = -t$$

má obecné řešení $ce^{\int f(t)dt}$, tedy v našem případě má řešení $g(t) = ce^{-\frac{t^2}{2}}$. Počáteční podmínka $c = g(0) = 1$.

Nyní zbývá použít jednoznačnost charakteristické funkce spolu s předchozím výpočtem.

Důkaz centrální limitní věty.

Podle věty o spojitosti stačí dokázat konvergenci odpovídajících charakteristických funkci. Máme

$$Y = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{j=1}^n X_j - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j,$$

kde $Z_j = \frac{X_j - \mu}{\sigma}$. Normalizace je provedena tak, že platí $E(Z_j) = 0, D(Z_j) = 1$, $E(Y) = 0, D(Y) = 1$

$$\phi_{\sum Z_j}(t) = \phi_{Z_1}(t)^n.$$

$$\phi_Y(t) = \phi_{\sum Z_j/\sqrt{n}}(t) = \phi_{\sum Z_j}(t/\sqrt{n}) = \phi_{Z_1}(t/\sqrt{n})^n.$$

Taylor okolo 0:

$$\phi_{Z_1}(x) = \phi_{Z_1}(0) + x\phi'_{Z_1}(0) + \frac{x^2}{2}\phi''_{Z_1}(0) + o(x^2).$$

Máme $x = \frac{t}{\sqrt{n}}$ a

$$\phi_{Z_1}(0) = 1, \phi'_{Z_1}(0) = iE(Z_1) = i(0) = 0, \phi''_{Z_1}(0) = i^2 E(Z_1^2) = -1.$$

$$\phi_{Z_1}(t/\sqrt{n}) = 1 - \frac{t^2}{2n} + o(\frac{t^2}{n}).$$

$$\phi_Y(t) = \phi_{Z_1}(t/\sqrt{n})^n = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-\frac{t^2}{2}}.$$

Rychlost konvergence je odhadnuta v následující větě.

Věta 16.4. *(Berry Essén)*
Nechť X_j jsou nezávislé, stejně rozdělené veličiny s nulovou střední hodnotou, $D(X_1) = \sigma^2$, $\rho = E(|X_1|^3) < \infty$, G_n distribuční funkce $\frac{\frac{1}{n}\sum_{j=1}^n X_j - \mu}{\sigma/\sqrt{n}}$, $\Phi = F_{N(0,1)}$. Potom platí

$$\sup_{x \in I\!R} |G_n(x) - \Phi(x)| \leq \frac{\rho}{2\sigma^3\sqrt{n}}$$

Poznámka. Binomické rozdělení lze aproximovat je-li $np(1-p) \geq 9$.

V praxi se používá $n \geq 30$ (e.g. Novovicova p. 71).

Věta 16.5. *Pokud $X_1, \ldots, X_n$ jsou nezávislé veličiny s normálním rozdělením $N(\mu_j, \sigma_j^2)$, a c_j konstanty, potom $X = \sum_{j=1}^n c_j X_j$ má normální rozdělení $N(\mu, \sigma^2)$ kde $\mu = \sum_{j=1}^n c_j \mu_j$, $\sigma^2 = \sum_{j=1}^n c_j^2 \sigma_j^2$.*

Důkaz.

$$\phi_X(t) = \prod_{j=1}^{n} \phi_{c_j X_j}(t) = \prod_{j=1}^{n} \phi_{X_j}(c_j t) = \prod_{j=1}^{n} e^{ic_j t\mu_j - \frac{\sigma_j^2 c_j^2 t^2}{2}} = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

$\square$

17. Statistika základní pojmy

Definice 17.1. *Nechť F je pravděpodobnostní rozdělení. Toto rozdělení (veličinu) nazveme* **základním souborem**. *Posloupnost $X = (X_1, \dots, X_n)$ nezávislých veličin se stejným rozdělením F se nazývá* **náhodný výběr** *o rozsahu n ze základního souboru. Množinu všech hodnot X nazveme* **výběrovým prostorem** *a každý bod $(x_1, \dots, x_n)$ této množiny nazveme* **realizací náhodného výběru**.

Tato posloupnost reprezentuje provedení n nezávislých experimentů se stejným rozdělením F.

♣ Formulace problému.
Ve statistice vycházíme z představy, že existují pravděpodobnostní rozdělení F_θ určitého typu, závisející na jednom nebo několika parametrech $\theta \in \Theta$, které jsou nám neznámé. Každé hodnotě θ přísluší odpovídající náhodný výběr X o rozsahu n. Naším cílem je pomocí souboru naměřených dat, tedy realizace náhodného výběru, odhadnout hodnoty těchto neznámých parametrů θ.

Příklady. Házíme mincí, tedy zkoumáme nula-jedničkové rozdělení s neznámým parametrem $\theta = p \in [0, 1]$
Zákon chyb- předpokládáme normální rozdělení, neznáme parametry $\mu \in I\!R, \sigma > 0$, $N(\mu, \sigma^2)$ má hustotu $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Definice 17.2. *Náhodná veličina $T(X) = T(X_1, \dots, X_n)$ definovaná na výběrovém prostoru, se nazývá* **výběrový odhad** *nebo* **statistika**.

Definice 17.3. **Výběrový průměr** *je definován jako statistika*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Výběrový rozptyl:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Hodnota S_n se nazývá **směrodatná odchylka** *náhodného výběru.*

Tyto dvě hlavní statistiky budeme dále studovat, jakožto odhady střední hodnoty, resp. variance apriori neznámého rozdělení F.

Definice 17.4. *Statistika T se nazývá* **nestranným odhadem** *parametru θ , jehož hodnota nemusí být apriori známa, pokud platí*

$$E(T) = \theta.$$

Statistika T se nazývá **konzistentním odhadem** *parametru θ, pokud platí*

$$\lim_{n\to\infty} P(|T - \theta| < \varepsilon) = 1.$$

Přesněji,

$$\forall \varepsilon > 0\ \forall \delta > 0\ \exists n_0 \in I\!N,\ (n > n_0) \Rightarrow\ P(|T_n - \theta| < \varepsilon) > 1 - \delta$$

Věta 17.5. *Nechť X je veličina se střední hodnotou μ a rozptylem σ^2. Potom výběrový průměr a výběrový rozptyl jsou nestrannou a konzistentní statistikou pro střední hodnotu $E(X)$, resp. varianci $D(X)$.*

Důkaz. Výběrový průměr.

$$E(\overline{X}_n) = \mu$$

Nestrannost je zřejmá, konzistence je ekvivalentní (slabému) zákonu velkých čísel.

Rozptyl. Potřebuji jednoduchou identitu:

$$\sum_{i=1}^{n}(X_i-\mu)(\overline{X}_n-\mu)=n(\overline{X}_n-\mu)^2$$

Takže

$$\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=\sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X}_n)^2=$$

$$\sum_{i=1}^{n}(X_i-\mu)^2-2\sum_{i=1}^{n}(X_i-\mu)(\overline{X}_n-\mu)+n(\overline{X}_n-\mu)^2=$$

$$\sum_{i=1}^{n}(X_i-\mu)^2-2\sum_{i=1}^{n}(X_i-\mu)(\overline{X}_n-\mu)+\sum_{i=1}^{n}(X_i-\mu)(\overline{X}_n-\mu)$$

Potom

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-2\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)(\overline{X}_n-\mu)+\frac{n}{n-1}(\overline{X}_n-\mu)^2=$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{n}{n-1}(\overline{X}_n-\mu)^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{n}{n-1}(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu))^2$$

Protože $(X_i-\mu)^2$ jsou nezávislé, dostaneme nestrannost pro rozptyl:

$$E(S_n^2)=\frac{n}{n-1}\sigma^2-\frac{n}{n-1}\frac{1}{n}\sigma^2=\sigma^2$$

Ze zákona velkých čísel aplikovaného na $\overline{X}_n$ a též na $\sum(X_i-\mu)^2$ ve vyjádření

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{n}{n-1}(\overline{X}-\mu)^2$$

plyne konzistence pro rozptyl.

Pro stejný parametr lze obecně řečeno zavést mnoho nestranných statistik (odhadu). Např. pro střední hodnotu lze zvolit X_j apod. V praxi je nejlepší volit odhady které mají co nejmenší rozptyl.

Definice 17.6. *Odhad $\phi(X_1,\dots,X_n)$ pro θ se nazývá nejlepší nestranný odhad, pokud je nestranný, a pokud pro každý jiný odhad $\psi(X_1,\dots,X_n)$ platí*

$$\forall\theta\in\Theta\quad D_\theta(\psi(X_1,\dots,X_n))\geq D_\theta(\phi(X_1,\dots,X_n)).$$

Lze ukázat, že pro náhodný výběr odvozený od normálního rozdělení $N(\mu,\sigma^2)$ jsou výběrový průměr a rozptyl nejlepším odhadem.

18. Bodové odhady, Metoda maximální věrohodnosti

Tato metoda se hodně používá, její odhady jsou konzistentní a pro velké n též nejlepší (tj. mají minimální rozptyl podle odhadu Cramér-Rao-[Dupač], nebo Dekking). Nemusí byt nestranné, ale jsou asymptoticky nestranné.

Motivační příklad: Nechť $X_1, \dots, X_4$ je náhodný výběr z nula-jedničkového rozdělení s parametrem $p = 0.2$ nebo $p = 0.8$. Realizací náhodného výběru máme $0, 0, 1, 0$. pravděpodobnost tohoto výsledku je $P = p(1-p)^3$. Pro $p = 0.2$ je to 0.1024, pro $p = 0.8$ je to 0.0064. Je tedy zřejmé, že za odhad p vezmeme hodnotu s pravděpodobnějším výsledkem.

Uvažujeme nyní pravděpodobnostní rozdělení F_θ, které závisí na jednom parametru $\theta \in \mathbb{R}$. Nechť $f(t, \theta)$ je spojitě diferencovatelná funkce dvou proměnných, která představuje pro každou hodnotu parametru θ pravděpodobnostní hustotu f_θ. Hustota pravděpodobnosti náhodného výběru $(X_1, \dots, X_n)$ (veličiny X_j jsou nezávislé, tedy hustota rozdělení výběru je součin hustot) je

$$L(\theta)(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i, \theta)$$

Této funkci se v tomto kontextu říká funkce věrohodnosti.

Definice 18.1. *Maximálně věrohodný odhad θ je hodnota $h(x_1, \dots, x_n)$ která maximalizuje funkci věrohodnosti $L(\theta)$ pro všechna $(x_1, \dots, x_n)$ z výběrového prostoru. Odpovídající statistika $T(X_1, \dots, X_n) = h(X_1, \dots, X_n)$ se tedy nazývá maximálně věrohodný odhad θ.*

Poznámka. Maximálně věrohodný odhad nemusí byt vždy nestranný, ale lze dokázat, že za určitých obecných předpokladů je asymptoticky nestranný (pro $n \to \infty$), a má řadu dalších pozitivních vlastností.

Poněvadž $L(\theta) \geq 0$, použitím logaritmu to je ekvivalentní maximalizaci výrazu

$$\log L(\theta) = \sum_{j=1}^{n} \log(f(x_i, \theta))$$

tedy

$$\sum_{j=1}^{n} \frac{d}{d\theta} \log(f(x_i, \theta)) = 0$$

♣ Předpokládáme, že můžeme rozdělení popsat hustotou

$$f(x) = (k+1)x^k, \quad x \in [0, 1].$$

kde $k > -1$ je neznámý parametr.

Naměřené hodnoty:

0.2 0.4 0.5 0.7 0.8 0.8 0.9 0.9

Odhadněte hodnotu k metodou maximální věrohodnosti.

$$L(k) = \prod_{i=1}^{8} (k+1) x_i^k$$

$$\log L(k) = 8 \log(k+1) + k \sum_{i=1}^{8} \log x_i$$

$$\frac{d}{dk}\log L(k) = \frac{8}{k+1} + \sum_{i=1}^{8}\log x_i = 0$$

Numerickým řešením dostaneme $k = 0.89$.

♣ Diskrétní příklad. Jev který nastává s neznámou pravděpodobností p nastal k-krát v průběhu n nezávislých pokusů. Odhadněte p.

$$L(p) = p^k(1-p)^{n-k}$$

$$\log L(p) = l(p) = k\log p + (n-k)\log(1-p)$$

$$l'(p) = \frac{k-np}{p(1-p)}$$

$l'(p) = 0$ pro $p = \frac{k}{n}$. Toto je rovno odhadu pro střední hodnotu Bernoulliho veličiny, podle výsledků výše.

♣ Diskrétní případ- Poissonovo rozdělení. Nechť $x_1, \ldots, x_n$ je realizace náhodného výběru. $Poiss(\lambda), \lambda > 0$ má rozdělení

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

věrohodnostní funkce je

$$L(\lambda, x_1, \ldots, x_n) = \lambda^{\sum x_j} e^{-n\lambda}\frac{1}{x_1! \ldots x_n!}$$

$$\log L(\lambda, x_1, \ldots, x_n) = \sum x_j \log\lambda - n\lambda - \sum_{j=1}^{n}(\sum_{i=1}^{x_j}\log i)$$

Řešení věrohodnostní rovnice splňuje

$$\sum x_j \frac{1}{\lambda} - n = 0$$

$$\lambda = \overline{x}_n = E(X)$$

♣ Rozdělení s hustotou závislou na parametru σ^2, náhodný výběr $x_1, \ldots, x_n$.

$$f(x, \sigma^2) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}, \ \ x \geq 0.$$

Funkce věrohodnosti má tvar

$$L(\sigma^2) = \prod_{i=1}^{n}[\frac{x_i}{\sigma^2}e^{-\frac{x_i^2}{2\sigma^2}}]$$

$$\ln L(\sigma^2) = -n\ln\sigma^2 + \ln(\prod_{i=1}^{n}x_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2$$

$$\frac{d}{d\sigma^2}\ln L(\sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}x_i^2 = 0$$

Řešení je

$$\sigma^2 = \frac{1}{2n}\sum_{i=1}^{n}x_i^2.$$

♣ Metodu maximální věrohodnosti lze použít i v případě více parametrů.

Najděte maximálně věrohodný odhad parametru normálního rozdělení $N(\mu, \sigma^2)$ pro náhodný výběr $x_1, \dots, x_n$. Hustota splňuje
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Hledáme kritický bod (σ^2 je pro nás formální symbol neznámého parametru, nikoliv druhá mocnina), tj.

$$\frac{\partial}{\partial\mu}\log L(\mu, \sigma^2) = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial}{\partial\sigma^2}\log L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

Řešení pro konkrétní realizaci náhodného výběru je

$$\mu = \overline{x}_n, \sigma^2 = \frac{n-1}{n}s_n^2,$$

takže řešení pro náhodný výběr je (jak je vidět pro rozptyl nevyjde výběrový rozptyl, takže tento odhad není nestranný, nicméně je vidět, že je konzistentní)

$$\mu = \overline{X}, \sigma^2 = \frac{n-1}{n}S_n^2$$

19. Intervalové odhady

Definice 19.1. *Nechť* $X_1, \ldots, X_n$ *je náhodný výběr z rozdělení které patří do množiny* $\{F_\theta : \theta \in \Theta\}$. *Dvojice statistik*

$$\{(\phi_L(X_1, \ldots, X_n), \phi_U(X_1, \ldots, X_n))\}$$

se nazývá **intervalový odhad** *(někdy též* **interval spolehlivosti**, *resp.* **konfidenční interval***) parametru* θ *o* **koeficientu spolehlivosti** $(1-\alpha)$ *jestliže platí*

$$P_\theta(\phi_L(X_1, \ldots, X_n) < \theta < \phi_U(X_1, \ldots, X_n)) = 1 - \alpha, \quad \theta \in \Theta$$

Dvojice statistik

$$\{(\phi_L(X_1, \ldots, X_n), \phi_U(X_1, \ldots, X_n))\}$$

se nazývá dolní, resp. horní odhad parametru θ *o koeficientu spolehlivosti* $1-\alpha$, *jestliže platí*

$$P_\theta(\phi_L(X_1, \ldots, X_n) < \theta) = 1 - \alpha, \quad \theta \in \Theta$$

resp.

$$P_\theta(\theta < \phi_U(X_1, \ldots, X_n)) = 1 - \alpha, \quad \theta \in \Theta$$

Konfidenční interval neznamená, že pravděpodobnost skutečné hodnoty ležící v tomto intervalu se rovná kvantilu, protože na prostoru Θ nemáme zadefinovanou pravděpodobnostní míru, a protože hodnota odhadu je funkcí realizace náhodného výběru. Místo toho znamená, že při mnoha opakováních experimentu bude poměrná četnost toho, že intervalový odhad obsahuje skutečnou hodnotu parametru, rovna kvantilu.

Definice 19.2. *Nechť distribuční funkce* F *je spojitá a rostoucí,* $0 < \beta < 1$. *Potom číslo* $x_\beta = F^{-1}(\beta)$ *se nazývá* β*-kvantil pro* F.

Je zřejmé, že platí

$$P(x_{\frac{\alpha}{2}} < X < x_{1-\frac{\alpha}{2}}) = F(x_{1-\frac{\alpha}{2}}) - F(x_{\frac{\alpha}{2}}) = 1 - \alpha, \quad \alpha < \frac{1}{2}.$$

♣ Kvantily pro normální rozdělení $N(0,1)$ mají značení $u_{1-\alpha} = F^{-1}(1-\alpha)$ (nebo v jiném značení $z_\alpha = u_{1-\alpha}$) a lze je najít v tabulkách. Platí

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Věta 19.3. *Je-li* X *náhodný výběr o velikosti* n *náhodné veličiny* $N(\mu, \sigma^2)$, *se známou hodnotou* σ, *potom konfidenční interval s koeficientem spolehlivosti* $(1-\alpha)$ *pro* μ *je dán*

$$\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$\overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$ *je dolní odhad* μ *s koeficientem spolehlivosti* $(1-\alpha)$.
$\overline{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}$ *je horní odhad* μ *s koeficientem spolehlivosti* $(1-\alpha)$.

Důkaz. Použijeme toho, že součet nezávislých normálně rozdělených veličin je opět normálně rozdělená veličina (Věta 16.5). Veličina $Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ má rozdělení $N(0,1)$, tedy platí

$$P_\mu(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \quad \mu \in I\!R$$

Úpravou získáme

$$P_\mu(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}_n < \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha, \quad \mu \in I\!R$$

$$P_\mu(\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha, \quad \mu \in \mathbb{R}$$

Postup pro jednostranné odhady je stejný. □

♣ Určete 90-procentní interval spolehlivosti pro střední hodnotu μ za předpokladu, že máte náhodný výběr o $n = 50$ z normálního rozdělení se známou směrodatnou odchylkou $\sigma = 12.1$ a aritmetický průměr je $\overline{x}_{50} = 36.38$.

Koeficient spolehlivosti je 0.9, tedy $\alpha = 0,10$, z tabulek určíme $z_{\alpha/2} = z_{0.05} = 1.645$. Dosazením

$$(36.38 - 1.645\frac{12.1}{\sqrt{50}}, 36.38 + 1.645\frac{12.1}{\sqrt{50}}) = (33.6, 39.2)$$

20. Základní věta statistiky

K nalezení intervalových odhadů pro normální rozdělení s neznámou hodnotou σ je třeba zavést nové náhodná veličiny, rozdělení Studentovo a χ^2.

♣ Funkce $\Gamma : I\!R^+ \to I\!R$ je definována následovně

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

platí $\Gamma(n+1) = n!$, asymptorická Stirlingova formule $\Gamma(z+1) \approx \sqrt{2\pi z}(\frac{z}{e})^z$, $z \to \infty$.

Věta 20.1. *χ^2 o n-stupních volnosti:*
Nechť $X_1, \dots, X_n$ jsou nezávislé veličiny s rozdělením $N(0,1)$. Potom $Y_n = \sum_{j=1}^n X_j^2$ ma hustotu (pro $t \leq 0$ je to 0)

$$g_n(t) = \frac{t^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{t}{2}}, \quad t > 0.$$

a platí $E(Y_n) = n, D(Y_n) = 2n$. Toto rozdělení se též značí χ_n^2. Distribuční funkce se obvykle značí G_n.

Dukaz viz skripta Dupac.

Věta 20.2. *(Základní věta statistiky) Nechť $X_j, j = 1, \dots, n$ jsou nezávislé veličiny stejně rozdělené $N(\mu, \sigma^2)$. Potom výběrový průměr $\overline{X}_n$ a výběrový rozptyl S_n^2 jsou nezávislé veličiny.*

Náhodná veličina $(n-1)\frac{S_n^2}{\sigma^2}$ má χ_{n-1}^2 rozdělení o $(n-1)$ stupních volnosti.

Důkaz. Budeme předpokládat, že $\mu = 0$, v obecném případě lze přejít k veličinám $X_j - \mu$. Označme $X = (X_1, \dots, X_n)$ náhodný vektor.

Nechť $U = (u_{ij})_{i,j=1}^n$ je ortonormální matice pro kterou $u_{1j} = \frac{1}{\sqrt{N}}$. Nechť

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{N}} & \cdots & \frac{1}{\sqrt{N}} \\ u_{21} & \cdots & u_{2n} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

kde $Y = (Y_1, \dots, Y_n)$ je nový náhodný vektor. Platí

$$Y_1 = \overline{X}_n \sqrt{n}$$

$$\sum_{j=1}^n Y_j^2 = Y^T Y = X^T X = \sum_{j=1}^n X_j^2$$

použitím vztahu který jsme již odvodili dříve ($\mu = 0$)

$$\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \sum_{i=1}^n X_i^2 - 2\sum_{i=1}^n X_i \overline{X}_n + n\overline{X}_n^2 = \sum_{j=1}^n X_j^2 - n\overline{X}_n^2,$$

tedy

$$\sum_{j=2}^n Y_j^2 = \sum_{j=1}^n (X_j - \overline{X}_n)^2.$$

Jak jsme dokázali výše ve Větě 14.5, jsou $Y_1, \dots, Y_n$ navzájem nezávislé veličiny, a tedy jsou Y_1 a $\sum_{j=1}^n (X_j - \overline{X}_n)^2$ též nezávislé veličiny.

Poslední tvrzení naší věty plyne okamžitě ze vztahu

$$\sum_{j=2}^{n}\frac{Y_j^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j-\overline{X}_n)^2=(n-1)\frac{S_n^2}{\sigma^2}.$$

Věta 20.3. *Studentovo (nebo také t_n-rozdělení) o n stupních volnosti:*
Nechť U,V jsou nezávislé veličiny s rozděleními $N(0,1)$ pro U a χ_n^2 o n-stupních volnosti pro V . Potom $T=\frac{U}{\sqrt{V}}\sqrt{n}$ má hustotu

$$h_n(t)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}},\ \ t\in I\!R.$$

Distribuční funkce se obvykle značí H_n.

Důkaz. Podle Věty 20.1,

$$f_V(t)=\frac{t^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}e^{-\frac{t}{2}},\ \ t>0.$$

Máme

$$F_{\sqrt{V}}(T)=F_V(T^2)$$

$$f_{\sqrt{V}}(T)=\frac{d}{dT}F_{\sqrt{V}}(T)=\frac{d}{dT}F_V(T^2)=2Tf_V(T^2)=2T\frac{T^{n-2}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}e^{-\frac{T^2}{2}}$$

Hustota veličiny $\sqrt{V}$ tedy splňuje

$$q_n(x)=\frac{x^{n-1}e^{-\frac{x^2}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})},\ x>0$$

Veličina $\sqrt{n}U$ má normální rozdělení $N(0,n)$
Podle Věty 14.6, kde $X=\sqrt{n}U$, $Y=\sqrt{V}$, $Z=\frac{X}{Y}$ má hustotu

$$f_Z(t)=\int_0^\infty xf_X(tx)f_Y(x)dx=\int_0^\infty x\frac{1}{\sqrt{2\pi n}}\,e^{-\frac{x^2t^2}{2n}}q_n(x)dx=$$

$$=\int_0^\infty\frac{x^ne^{-\frac{x^2t^2}{2n}-\frac{x^2}{2}}}{\sqrt{2\pi n}\,2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}dx=\int_0^\infty\frac{x^ne^{-x^2\frac{1}{2}(1+\frac{t^2}{n})}}{\sqrt{2\pi n}\,2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}dx$$

substitucí $s=\frac{1}{2}x^2(1+\frac{t^2}{n})$, $(dx=\frac{ds}{x(1+\frac{t^2}{n})},\ x^{n-1}=\frac{(2s)^{\frac{n-1}{2}}}{(1+\frac{t^2}{n})^{\frac{n-1}{2}}})$

$$=\frac{1}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}\int_0^\infty s^{\frac{n+1}{2}-1}e^{-s}ds=$$

$$=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}\to\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}},\ \ n\to\infty.$$

□

Toto rozdělení je symetrické a platí $E(T)=0, D(T)=\frac{n}{n-2}$. Pro $n>30$ platí $T\doteq N(0,1)$.

Důsledek předchozích vět a definic je následující.

Věta 20.4. *Náhodná veličina $T=\frac{\overline{X}_n-\mu}{S_n}\sqrt{n}$ má t_{n-1}-rozdělení o $(n-1)$ stupních volnosti.*

Důkaz. Položme $U = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$, $V = \frac{(n-1)S_n^2}{\sigma^2}$. Potom U, V jsou nezávislá a mají rozdělení $N(0,1)$, resp. χ^2_{n-1}. Přitom platí

$$T = \frac{U}{\sqrt{V}}\sqrt{n-1}.$$

Kvantily obou těchto rozdělení jsou tabelovány, a nesou značení (pro n-stupňů volnosti) $\chi^2_{1-\alpha,n}$, resp. $t_{1-\alpha,n}$. Studentovo rozdělení o n-stupních volnosti je symetrické, a platí že $t_{1-\alpha,n} \to z_\alpha$ pro $n \to \infty$. Pro naše účely při práci s tabulkami kvantilu tedy budeme nahrazovat kvantily Studentova rozdělení kvantily normálního rozdělení.

Věta 20.5. *Je-li X náhodný výběr o velikosti n náhodné veličiny $N(\mu, \sigma^2)$, s neznámou hodnotou σ, potom je konfidenční interval spolehlivosti $(1-\alpha)$ pro μ dán*

$$\overline{X}_n - t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}}$$

Konfidenční interval o spolehlivosti $(1-\alpha)$ pro σ^2 je dán

$$\frac{(n-1)S_n^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)S_n^2}{\chi^2_{\frac{\alpha}{2},n-1}}$$

Důkaz. Použijeme toho, že veličina $Z = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ má t_{n-1}-rozdělení o $(n-1)$ stupních volnosti, tedy platí

$$P_{\mu,\sigma^2}(-t_{1-\alpha/2,n-1} < \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < t_{1-\alpha/2,n-1}) = 1-\alpha, \quad \mu \in I\!R$$

Úpravou získáme

$$P(\mu - t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}} < \overline{X}_n < \mu + t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}}) = 1-\alpha, \quad \mu \in I\!R$$

$$P(\overline{X}_n - t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}}) = 1-\alpha, \quad \mu \in I\!R$$

Úpravou podobně jako v případě známé σ získáme výsledek.

Použijeme toho, že veličina $Z = \frac{(n-1)S_n^2}{\sigma^2}$ má χ^2_{n-1}-rozdělení o $(n-1)$ stupních volnosti, tedy platí

$$P_{\mu,\sigma^2}(\chi^2_{\alpha/2,n-1} < \frac{(n-1)S_n^2}{\sigma^2} < \chi^2_{1-\alpha/2,n-1}) = 1-\alpha, \quad \mu \in I\!R$$

odtud plyne odhad ve větě. □

Věta 20.6. *Je-li $X = (X_1, \ldots, X_n)$ náhodný výběr o velikosti n náhodné veličiny s konečným rozptylem $D(X_1) > 0$, potom je konfidenční interval spolehlivosti $(1-\alpha)$ pro střední hodnotu $\mu = E(X_1)$ dán asymptoticky, pro $n \to \infty$, vztahem*

$$\overline{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

$\overline{X}_n - z_\alpha \frac{S_n}{\sqrt{n}}$ *je asymptotický dolní odhad μ o spolehlivosti $(1-\alpha)$.*

$\overline{X}_n + z_\alpha \frac{S_n}{\sqrt{n}}$ *je asymptotický horní odhad μ o spolehlivosti $(1-\alpha)$.*

Důkaz: Označme $\sigma^2 = D(X_1)$. Podle centrální limitní věty má náhodná veličina $Z_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ asymptoticky rozdělení $N(0,1)$. Dále platí, že asymptoticky $\frac{S_n}{\sigma} \to 1$, tedy platí, že $Y_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ má opět asymptoticky rozdělení $N(0,1)$. Tedy asymptoticky pro $n \to \infty$ platí

$$P_{\mu,\sigma^2}(-z_{\alpha/2} < Y_n < z_{\alpha/2}) = 1 - \alpha, \quad \mu, \sigma^2 \in I\!R.$$

Úpravou podobně jako v případě známé σ získáme výsledek.

V praxi se ze zkušenosti považuje za dostatečně velký soubor kde $n > 30$.

♣ Pr. Má být zřízeno nové vlakové spojení mezi dvěma městy. V průběhu roku bylo provedeno 30 náhodných měření počtu cestujících. Jejich aritmetický průměr je $\overline{x} = 450$ a výběrová odchylka $s_{30} = 30$. Určete 99 procentní interval spolehlivosti pro střední počet cestujících.

Řešení: Koeficient spolehlivosti je 0.99, tedy $\alpha = 0.01$. Z tabulky určíme $z_{\alpha/2} = 2.567$, tedy

Tudíž,

$$(450 - 2.567\frac{30}{\sqrt{30}}, 450 + 2.567\frac{30}{\sqrt{30}})$$

21. Přípustná chyba odhadu

Definice 21.1. *Přípustná chyba odhadu je definována jako*
$$\Delta = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Tedy při zadaném koeficientu spolehlivosti a přípustné chybě odhadneme minimální velikost souboru, která bude tyto podmínky splňovat,
$$n = (\frac{z_{\alpha/2}\sigma}{\Delta})^2$$

♣ U náhodně vybraných domácností byly sledovány výdaje za pohonné látky auta. Byl udělán závěr, že směrodatná odchylka σ je 413 Kč. Určete rozsah výběru nutný k tomu, abychom měli 95 procentní spolehlivost, že přípustná chyba odhadu je rovna 15.

Řešení. $\Delta = 15$, $\alpha = 0.05$, tedy $z_{\alpha/2} = 1.96$.
$$n = (\frac{z_{\alpha/2}\sigma}{\Delta})^2 \doteq 2912$$

♣ Hodíme 100krat kostkou. Jaká je pravděpodobnost, že součet hodů bude z intervalu $[320, 380]$?

Známe $\mu = 3.5$, $\sigma^2 = \frac{2(2.5^2+1.5^2+0.5^2)}{6} = 2.91$, $\sigma = 1.7$. veličina $Z = \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$ má přibližně rozdělení $N(0,1)$, tedy platí

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha,$$

$$P(-z_{\alpha/2} < \frac{\overline{X}_{100}-\mu}{\sigma/\sqrt{100}} < z_{\alpha/2}) = 1-\alpha,$$
$$P(-z_{\alpha/2} < \frac{\overline{X}_{100}-3.5}{0.17} < z_{\alpha/2}) = 1-\alpha.$$

Chceme aby $\overline{X}_{100} \in [3.2, 3.8]$, tedy

$$\frac{\overline{X}_{100}-3.5}{0.17} \in [-1.764, 1.764].$$

Zde $z_{\alpha/2} = 1.764$, tedy $\frac{\alpha}{2} \doteq 0.04$, $\alpha \doteq 0.08$. pravděpodobnost výsledku je tedy přibližně 92 procent.

♣ Máme minci, která dává pannu s pravděpodobností 70 procent. Kolik pokusných hodů je třeba udělat, abychom mohli s jistotou 90 procent očekávat, že padne víc pannen než orlů?

Položíme X_1 nula jedničkové rozdělení s pravděpodobnostmi 0.7, 0.3. Známe $\mu = 0.7$, $\sigma^2 = 0.7\cdot 0.3^2 + 0.3\cdot 0.7^2 = 0.21$, $\sigma = 0.458$, $\alpha = 0.1$ tedy $z_\alpha = 1.29$. Padne víc pannen než orlů znamená, že $\overline{X}_n > 0.5$. Náš cíl je tedy

$$P(\overline{X}_n > 0.5) = 1-\alpha.$$

veličina $Z = \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$ má přibližně rozdělení $N(0,1)$, tedy platí

$$P(Z > -z_\alpha) = 1-\alpha,$$

$$-1.29 < \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$$
$$\frac{-1.29\times 0.458}{\sqrt{n}} + 0.7 < \overline{X}_n$$

Tedy požadavek bude splněn, pokud

$$0.5 \leq \frac{-1.29 \times 0.458}{\sqrt{n}} + 0.7 < \overline{X}_n$$

$$\sqrt{n} > \frac{1.29 \cdot 0.458}{0.2} = 2.95$$

a tedy $n > 8$.

♣ Bod se pohybuje z počátku číselné osy každou vteřinu o jedna doleva nebo zůstane na místě, obojí s pravděpodobnosti jedna polovina. Určete, za jak dlouhou dobu n budeme moci s jistotou 95 procent očekávat. že vzdálenost bodu od počátku je aspoň 40 procent n v jednotkách délky?

Položíme X_1 nula jedničkové rozdělení s pravděpodobnostmi $\frac{1}{2}$. Známe $\mu = 0.5$, $\sigma^2 = 0.5 \cdot 0.5^2 + 0.5 \cdot 0.5^2 = 0.25$, $\sigma = 0.5$, $\alpha = 0.05$ tedy $z_\alpha = 1.645$. Očekáváme, že $\overline{X}_n > 0.4$. Náš cíl je tedy

$$P(\overline{X}_n > 0.4) = 1 - \alpha.$$

Veličina $Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ má přibližně rozdělení $N(0,1)$, tedy platí

$$P(Z > -z_\alpha) = 1 - \alpha,$$

$$-1.645 < \frac{\overline{X}_n - 0.5}{0.5/\sqrt{n}}$$

$$\frac{-1.645 \times 0.5}{\sqrt{n}} + 0.5 < \overline{X}_n$$

Tedy požadavek bude splněn, pokud

$$0.4 \leq \frac{-1.645 \times 0.5}{\sqrt{n}} + 0.5 < \overline{X}_n$$

Tedy

$$\sqrt{n} > \frac{1.645 \cdot 0.5}{0.1} = 8.2$$

a tedy $n > 67$.

♣ Člun má nosnost 5000 kg. Hmotnost cestujících je náhodná veličina X se střední hodnotou $E(X) = 70$ kg a směrodatnou odchylkou 20 kg. Kolik cestujících může cestovat, aby pravděpodobnost přetížení byla menší než 1 procento?

Známe $\mu = 70$, $\sigma^2 = 400$, $\sigma = 20$. Pro $\alpha = 0.01$ máme $z_\alpha = 2.326$. Veličina $Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ má přibližně rozdělení $N(0,1)$. Chceme aby platilo

$$P(n\overline{X}_n < 5000) \geq 0.99,$$

Tedy úpravou

$$P(\overline{X}_n < \frac{5000}{n}) \geq 0.99,$$

Víme, že platí

$$P(\frac{\overline{X}_n - 70}{20/\sqrt{n}} < 2.326) \geq 0.99,$$

$$P(\overline{X}_n < 70 + \frac{20 \cdot 2.326}{\sqrt{n}}) \geq 0.99,$$

Takže pokud bude platit

$$70 + \frac{20 \cdot 2.326}{\sqrt{n}} < \frac{5000}{n}$$

potom bude též platit hledaný odhad. Řešením nerovnice bude $n \leq 70$.

♠ Pr. Opakovaná měření koncentrace látky vedla k násl. výsledkům:

$$0.2, 0.23, 0.21, 0.16, 0.18, 0.19, 0.14, 0.18, 0.21$$

Najděte oboustranné 90-procentní odhady střední hodnoty, rozptylu a směrodatné odchylky.

Platí $n = 9$, $\overline{x}_9 = 0.189$, $s^2 = 7.6 \cdot 10^{-4}, s = 2.76 \cdot 10^{-2}$.

Intervalový odhad pro $\alpha = 0.1$, kvantil $t_{1-\alpha/2,n-1} = q_{t(8)}(0.95) = 1.86$:

$$\overline{X}_n - t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}}$$

$$0.189 - \frac{2.76 \cdot 10^{-2}}{3}1.86 < \mu < 0.189 - \frac{2.76 \cdot 10^{-2}}{3}1.86$$

Konfidenční interval o spolehlivosti $(1-\alpha)$ pro σ^2 je dán, pro kvantily

$$\chi^2_{1-\frac{\alpha}{2},n-1} = 15.51, \;\; \chi^2_{\frac{\alpha}{2},n-1} = 2.73$$

$$\frac{(n-1)S_n^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)S_n^2}{\chi^2_{\frac{\alpha}{2},n-1}}$$

$$3.9 \cdot 10^{-4} < \sigma^2 < 2.2 \cdot 10^{-3}$$

Odtud

$$1.97 \cdot 10^{-2} < \sigma < 4.7 \cdot 10^{-2}$$

Věta 21.2. *Předpokládáme, že máme dva nezávislé výběry $X = (X_1, \ldots, X_{n_1})$ z rozdělení $N(\mu_1, \sigma^2)$ a $Y = (Y_1, \ldots, Y_{n_2})$ z rozdělení $N(\mu_2, \sigma^2)$. Označme $\overline{X}, \overline{Y}.S_1^2, S_2^2$ výběrové průměry a rozptyly. Potom statistiky $\overline{X}, \overline{Y}$ jsou nezávislé a statistika $\overline{X} - \overline{Y}$ ma rozdělení $N(\mu_1 - \mu_2, \frac{n_1+n_2}{n_1 n_2}\sigma^2)$.*

$$S^{*2} = \frac{1}{n_1 + n_2 - 2}((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2)$$

je nestranný odhad σ^2 a $(n_1 + n_2 - 2)S^{*2}/\sigma^2$ má $\chi^2(n_1 + n_2 - 2)$ rozdělení, a

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S^* \sqrt{\frac{n_1+n_2}{n_1 n_2}}}$$

má rozdělení $t_{n_1+n_2-2}$.

Jsou-li μ_1, μ_2, σ neznámé parametry je

$$((\overline{X} - \overline{Y}) - t_{1-\alpha/2,n_1+n_2-2}S^*\sqrt{\frac{n_1+n_2}{n_1 n_2}}, (\overline{X} - \overline{Y}) + t_{1-\alpha/2,n_1+n_2-2}S^*\sqrt{\frac{n_1+n_2}{n_1 n_2}})$$

intervalový odhad funkce $\mu_1 - \mu_2$ o spolehlivosti $1 - \alpha$.

♣ Je třeba stanovit intervalový odhad o spolehlivosti 0.95 pro rozdíl obsahu chlóru g/l ve dvou nádržích. Bylo odebráno $n_1 = 25$ vzorků z první nádrže a $n_2 = 10$ z druhé nádrže, s výsledky:

$$\overline{X} = 34.48, \;\; S_1^2 = 1.7482$$

$$\overline{Y} = 35.59, \;\; S_1^2 = 1.7121$$

Předpokládáme, že vzorky jsou normálně rozdělené, použijeme předchozí obecnou větu. Kvantil

$$t_{0.975,33} = 2.035$$

po dosazení dostaneme interval

$$(0.106, 2.114)$$

22. Testování hypotéz

Motivační příklad. Továrna vyrábí potravinu, která nesmí obsahovat v průměru více než jedno promile nežádoucí látky (toto je předpoklad, nebo hypotéza kterou chceme ověřit resp. otestovat). Na základě náhodného výběru měření této koncentrace chceme rozhodnout, jestli podmínka je splněna. Pokud náš závěr bude, že podmínka splněna není, bude nutno přistoupit k nákladné opravě, resp. reorganizaci výrobní linky. Je tedy třeba maximálně snížit riziko chybného hodnocení, ve smyslu falešného poplachu.

Formalizace problému.

Předpokládáme, že máme náhodný výběr $X = (X_1, \ldots, X_n)$, který pochází od základního souboru X_1, jenž má neznámý parametr $\theta \in \Theta$. Předpokládáme, že o parametru θ existují dvě navzájem si konkurující hypotézy, **nulová hypotéza** H_0: $\theta \in \Theta_0$ a **alternativní hypotéza** H_1: $\theta \in \Theta_1$ (často $\Theta_1 = \mathbb{R} \setminus \Theta_0$).

Testem nulové hypotézy H_0 proti H_1 rozumíme rozhodovací proces, založený na náhodném výběru (resp. jeho realizaci), X na jehož základě zamítneme nebo nezamítneme hypotézu H_0.

Máme následující možností.

1. platí H_0 a naše rozhodnutí je nezamítnout H_0.

2. **Chyba 1. druhu**: zamítneme H_0, ačkoliv je pravdivá. Závažnější než chyba 2 druhu, H_0-nevinen odsouzen

3. **Chyba 2. druhu**: přijmeme H_0, ačkoliv je nepravdivá. Méně závažná než chyba 1, H_1 vinen neodsouzen (platí H_1).

4. platí H_1 a naše rozhodnutí je zamítnout H_0.

Test je popsán **kritickým oborem** $W \subset \mathbb{R}^n$ realizace výběru X. Pokud $X \in W$ pak hypotézu H_0 zamítneme, v opačném případě ji nezamítneme.

Předpokládejme, že testovací proces dává chybu 1 druhu s pravděpodobností α, která se nazývá **hladina významnosti** testu, tedy platí

$$P_\theta(X \in W) \le \alpha, \;\; \forall \theta \in \Theta_0$$

Hodnotu α se budeme snažit minimalizovat. Chyba 2 druhu nastává s pravděpodobností β,

$$\beta(\theta) = P_\theta(X \notin W), \;\; \theta \in \Theta_1.$$

Číslo $1-\beta$ se nazývá **síla testu**, a tuto hodnotu se budeme snažit maximalizovat, při zachování α.

Při testování při pevné velikosti vzorku je zřejmé, že při snaze zmenšit α, budeme zároveň zvyšovat sílu testu (snižovat β).

Věta 22.1. *(Neymanova Pearsonova věta) Nechť p_1, p_2 jsou pravděpodobnostní hustoty na $\mathbb{R}^n$. Nechť pro $\alpha \in (0,1)$ existuje $c > 0$ takové, že pro množinu*

$$W^* = \{x \in \mathbb{R}^n : p_1(x) \ge c p_0(x)\}$$

platí

$$\int_{W^*} p_0(x)dx = \alpha$$

Pak pro libovolnou měřitelnou množinu $W \subset \mathbb{R}^n$ splňující

$$\int_W p_0(x)dx \leq \alpha$$

platí

$$\int_{W^*} p_1(x)dx \geq \int_W p_1(x)dx$$

Důkaz.

$$\int_{W^*} p_1(x)dx - \int_W p_1(x)dx = \int_{W^*\setminus W} p_1(x)dx - \int_{W\setminus W^*} p_1(x)dx$$

$$\geq \int_{W^*\setminus W} cp_0(x)dx - \int_{W\setminus W^*} cp_0(x)dx$$

$$= c(\int_{W^*} p_0(x)dx - \int_W p_0(x)dx) \geq 0$$

□

Důsledek 22.2. *Nechť X je náhodný výběr s parametrem θ_0 nebo $\theta_1 \neq \theta_0$, jejich odpovídající hustoty v $\mathbb{R}^n$ jsou p_0, p_1. Máme-li testovat H_0 proti H_1, potom test s kritickým oborem W^* má nejmenší pravděpodobnost chyby druhého druhu β (je tedy nejsilnější) mezi všemi testy na hladině α.*

Platí, pro libovolnou množinu W, kritický výběr na hladině α,

$$\beta(\theta_1) = P_{\theta_1}(X \notin W) = \int_{\mathbb{R}^n\setminus W} p_1(x)dx = 1 - \int_W p_1(x)dx \geq 1 - \int_{W^*} p_1(x)dx$$

Věta 22.3. *Nechť X je náhodný výběr z rozdělení $N(\mu, \sigma^2)$, kde μ je neznámý parametr a σ^2 známá konstanta. Nechť H_0 je hypotéza $\mu = \mu_0$, a H_1 je alternativní hypotéza $\mu = \mu_1 > \mu_0$, $\alpha > 0$. Potom nejsilnější test H_0 na hladině α je dán vztahem (který nezávisí na μ_1):*

$$\frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} \geq z_\alpha$$

Důkaz. Podle předchozí věty najdeme W^* tak, aby platilo

$$\prod_{j=1}^n\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_j-\mu_1)^2}{2\sigma^2}}\right) \geq c\prod_{j=1}^n\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_j-\mu_0)^2}{2\sigma^2}}\right)$$

tedy

$$e^{-\frac{\sum_{j=1}^n (x_j-\mu_1)^2}{2\sigma^2}} \geq ce^{-\frac{\sum_{j=1}^n (x_j-\mu_0)^2}{2\sigma^2}}$$

logaritmováním

$$\sum_{j=1}^n (x_j^2 - \mu_1 x_j + \mu_1^2) - \sum_{j=1}^n (x_j^2 - \mu_0 x_j + \mu_0^2) \leq c_1$$

$$\overline{x}_n \geq c_2$$

Konstantu c_2 určíme z podmínky

$$P_{\mu_0}(X \in W^*) = \alpha$$

tedy

$$P_{\mu_0}(\overline{X}_n \geq c_2) = 1 - \Phi(\frac{c_2 - \mu_0}{\sigma}\sqrt{n}) = \alpha$$

$$c_2 = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Nulovou hypotézu zamítneme, je-li

$$\frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} \geq z_\alpha$$

□

♣ Odhad chyby 2 druhu, jako funkce $\mu(= \mu_1 > \mu_0)$:

$$\beta(\mu) = P_\mu(X \notin W^*),\ \mu \in I\!R$$

Je-li μ skutečná hodnota parametru je $\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}$ rozdělení $N(0,1)$,

$$\beta(\mu) = P_\mu(\frac{\overline{X_n} - \mu_0}{\sigma/\sqrt{n}} \leq z_\alpha)$$

$$= P_\mu(\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} < z_\alpha + \frac{(\mu_0 - \mu)}{\sigma/\sqrt{n}})$$

$$= \Phi(z_\alpha + \frac{(\mu_0 - \mu)}{\sigma/\sqrt{n}})$$

Funkce β je tedy spojitá a klesající a platí

$$\beta(\mu) \geq 1 - \alpha,\ \mu < \mu_0,\ \ \beta(\mu_0) = 1 - \alpha$$

$$\beta(\mu) \leq 1 - \alpha,\ \mu > \mu_0,$$

$$\lim_{\mu \to \infty} \beta(\mu) = 0$$

Chceme-li aby pravděpodobnost chyby 2. druhu byla v pevném bodě $\mu_1 > \mu_0$ menší nebo rovna β, budeme požadovat

$$\Phi(z_\alpha + \frac{(\mu_0 - \mu_1)}{\sigma/\sqrt{n}}) \leq \beta$$

což vede k požadavku

$$n \geq (\frac{z_\alpha + z_\beta}{\mu_1 - \mu_0}\sigma)^2$$

Konkrétní příklad, $\mu_0 = 0, \sigma^2 = 1, n = 16, \alpha = 0.05$.
V tomto případě

$$\beta(\mu) = \Phi(1.645 - 4\mu)$$

hodnoty:

μ	0.1	0.2	0.3	0.4	0.5	0.8	1.0
$\beta(\mu)$	0.89	0.81	0.67	0.51	0.36	0.06	0.01

Obdobným způsobem, ale s použitím maximálně věrohodných odhadů v průběhu důkazu lze odvodit následující kritérium (viz Dupač p. 128)

Věta 22.4. *Nechť X je náhodný výběr z rozdělení $N(\mu, \sigma^2)$, kde μ i σ^2 jsou neznámé parametry. Nechť H_0 je hypotéza $\mu = \mu_0$ a H_1 je alternativní hypotéza $\mu \neq \mu_0$, $\alpha > 0$. Potom H_0 zamítneme na hladině α je-li:*

$$\frac{|\overline{X}_n - \mu_0|}{S_n}\sqrt{n} \geq t_{1-\frac{\alpha}{2},n-1}$$

♣ Test o střední hodnotě normálního rozdělení s neznámým rozptylem s hladinou významnosti α, $H_0 : \mu = \mu_0$. Zamítáme H_0 pokud platí:

$$|\frac{\overline{X} - \mu_0}{S_n/\sqrt{n}}| > t_{1-\alpha/2,n-1} \approx z_{\frac{\alpha}{2}}$$

Je vidět, že H_0 zamítáme na hladině testu α právě tehdy, když oboustranný $100(1-\alpha)$ procentní interval spolehlivosti pro μ neobsahuje hypotetickou hodnotu μ_0.

♣ Zařízení váží součástky s chybou, která má rozdělení $N(\mu, \sigma^2)$, $\sigma = 0.66$g, μ neznáme. Otestujte na hladině významnosti 5 procent, zda je možné, aby měření nebylo zatíženo systematickou chybou (tj. střední hodnota chyb μ_0 byla nulová), pokud při $n = 9$ kontrolních měřeních byly naměřeny tyto chyby (v gramech):

$$0.3, 0.4, -0.8, 0.1, -1.3, -1.1, -0.6, 0.2, -0.5$$

platí $\mu_0 = 0$, $\overline{x} = -3.3/9 \doteq -0.367$g. Směrodatná odchylka $\sigma/\sqrt{n} = 0.22$ g. Testovací statistiku

$$|\frac{\overline{x}}{\sigma}\sqrt{n}| = |\frac{-0.367}{0.22}| = 1.67$$

porovnáme s kvantilem $z_(0.025) = 1.96$ a nulovou hypotézu, že zařízení nemá systematickou chybu, nezamítáme.

♣ Testujte na hladině významnosti $\alpha = 0.1$, jestli následující normálně rozdělená data mají střední hodnotu $\mu = 5$.

$$6.1, 1.2, 3.4, 8.1, 5.1, 6.0, 4.7, 7.4$$

Předpokládáme rozdělení $N(\mu, \sigma^2)$, testujeme $H_0 : \mu = 5$ proti hypotéze $H_1 : \mu \neq 5$, používáme statistiku

$$T = \frac{\overline{X} - \mu}{S_n}\sqrt{n}$$

kde $n = 8, \overline{X} = 4.825, \mu = 5, S_8 = 2.059$, která má rozdělení t_7. Oboustranný test, tedy porovnáme $|T|$ s kvantilem $t_{1-\frac{\alpha}{2},7} = 1.89$. platí $|T| = 0.24 < 1.89$ proto nezamítáme H_0.

♣ Výrobce tvrdí, že spotřeba automobilu je 6 litru na 100 km. Průměrná spotřeba u 49 uživatelů ale byla 6.4 l na 100 km. Dále byla naměřena směrodatná odchylka $\sigma = 1.6$ l na 100 km. Testujte na hladině 5 procent, zda měl výrobce pravdu.

Volíme H_0: spotřeba 6 l, H_1 : spotřeba $\neq 6$. Za platnosti H_0 ma veličina

$$\frac{\overline{X} - 6}{1.6}\sqrt{49}$$

rozdělení $t(48)$ Vypočteme hodnotu

$$t = \frac{6.4 - 6}{1.6} \cdot 7 = 1.75$$

Jelikož $t_{0.975,48} = 2.011 > |t|$ hypotézu H_0 nezamítáme.

Pokud za alternativní hypotézu zvolíme H_1 : spotřeba > 6 l, pak hodnotu $|t| = 1.75$ porovnáváme s kvantilem $t_{0.95,48} = 1.68 < |t|$ a nulovou hypotézu zamítáme.

Alternativní přístup k testování hypotéz. spočívá v udání P-hodnoty hypotézy, tj. vypočteme hodnotu testové statistiky a k ní **nejmenší obor zamítnutí** při kterém bychom mohli na základě této hodnoty zamítnout nulovou hypotézu proti dané alternative. Hladina významnosti odpovídající tomuto kritickému oboru je P-hodnota.

Definice 22.5. *P-hodnota testu hypotézy je rovna nejmenší hladině významnosti, na které nulová hypotéza H_0 může být zamítnuta.*

Pridat obrazek eg novovicova p. 102

♣ Při paralelním zjišťování obsahu cukru v melase polarimetrickou metodou se naměřily tyto hodnoty:

$$51.3 \;\; 52.2 \;\; 51.8 \;\; 53.4 \;\; 52.6 \;\; 50.9 \;\; 52.3 \;\; 53.1 \;\; 54.3$$

přičemž jde o výběr z normálního rozdělení se známým rozptylem $\sigma^2 = 1.2$. Pomocí výpočtu dosažené významnosti (tedy p-hodnoty) příslušného testu ověřte, zda zjišťovaný obsah cukru je spolehlivě větší (na hladině významnosti $\alpha = 0.0075$) než 51.5

Testujeme nulovou hypotézu

$$H_0 : \mu \leq 51.5, \quad H_1 : \mu > 51.5$$

Z naměřených hodnot $\overline{x} = 52.43$, $n = 6, \sigma^2 = 1.2, \mu_0 = 51.5$:

$$\frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = 2.556$$

Pokud $z_\alpha = 2.556$, potom $\alpha = 0.0053$, což je nejmenší hladina významnosti na které může být H_0 odmítnuta, tedy P-hodnota.

Jelikož $0.0053 < 0.0075$, můžeme H_0 odmítnout pro dané α. Avšak jak vidíme, nemůžeme H_0 odmítnout na hladině významnosti 0.005.

Použitím Věty 22.6 obdržíme následující dvouvýběrový test rovnosti středních hodnot na hladině významnosti α.

Věta 22.6. *Předpokládáme, že máme dva nezávislé výběry $X = (X_1, \ldots, X_{n_1})$ z rozdělení $N(\mu_1, \sigma^2)$ a $Y = (Y_1, \ldots, Y_{n_2})$ z rozdělení $N(\mu_2, \sigma^2)$. Označme $\overline{X}, \overline{Y}, S_1^2, S_2^2$ výběrové průměry a rozptyly,*

$$S^{*2} = \frac{1}{n_1 + n_2 - 2}((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2).$$

Potom nulovou hypotézu H_0 rovnosti $\mu_1 = \mu_2$ (oproti H_1, $\mu_1 \neq \mu_2$) zamítáme na hladině významnosti na α pokud platí

$$|\overline{X} - \overline{Y}| \geq t_{1-\alpha/2, n_1+n_2-2} S^* \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \;\; \sigma^2 \text{ neznámé},$$

$$|\overline{X} - \overline{Y}| \geq u_{1-\alpha/2} \sigma \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \;\; \sigma^2 \text{ známé}.$$

23. χ^2 TEST DOBRÉ SHODY

Předpokládáme, že proměnná X_1 může nabývat konečně mnoha hodnot z množiny $\{h_1, \dots, h_k\}$ s kladnými pravděpodobnostmi $\{p_1, \dots, p_k\}$. Nechť $\{q_1, \dots, q_k\}$ jsou nezáporné konstanty $\sum q_j = 1$. Naším cílem je otestovat proti sobě hypotézy

$$H_0 : p_j = q_j, \ \forall j = 1, \dots, k$$

$$H_1 : p_j \neq q_j, \text{ pro některé } j$$

Předpokládejme, že při testování vyšel náhodný výběr $X = (X_1, \dots, X_n)$ takový, že celkový počet hodnot h_j byl N_j, $\sum_{j=1}^{k} N_j = n$.

Věta 23.1. *Statistika*

$$Q = \sum_{j=1}^{k} \frac{(N_j - nq_j)^2}{nq_j}$$

splňuje následující vlastnost. Je-li H_0 splněno a $n \to \infty$ potom distribuční funkce Q konverguje k distribuční funkci $\chi^2(k-1)$.

V praxi požadujeme aby $N_j \geq 5, \forall j$. Hypotézu H_0 tedy odmítneme na hladině významnosti α, pokud

$$Q > \chi^2_{\alpha, k-1}$$

♣ Příklad. Budeme testovat hypotézu rozdělení krevních skupin. Teoretické rozdělení je následující:

$$A\,\frac{1}{3},\ \ B\,\frac{1}{8},\ \ AB\,\frac{1}{24},\ \ O\,\frac{1}{2}$$

Naměřené hodnoty ve vzorku 6004 osob jsou následující:

$$A\,2162,\ \ B\,738,\ \ AB\,228,\ \ O\,2876$$

Dosazením do formule vyjde

$$Q = \frac{(2162 - 2001.3)^2}{2001.3} + \frac{(738 - 750.5)^2}{750.5} + \frac{(228 - 250.2)^2}{250.2} + \frac{(2876 - 3002.0)^2}{3002.0} = 20.37$$

V tomto případě p-hodnota rozdělení χ^2_3 je $1.42 \cdot 10^{-4}$.

Test lze použít i pro spojitá rozdělení. V tomto případě je nutné rozdělit soubor dat do skupin s přesně určenou pravděpodobností.

♣ Pro soubor dat která jsou měřením teplot za určité období

$$22, 26, 19, 18, 17, 19, 20, 26, 25, 16, 20, 18, 16, 26, 25, 23, 20, 21, 23, 23$$

otestujte na hladině významnosti $\alpha = 0,05$ hypotézu

$$H_0: \text{ výběr je z rovnoměrného rozdělení na intervalu } (15, 26)$$

$$H_1: \text{ rozdělení není rovnoměrné}$$

Volíme šířku třídy $d = 3$, tedy třídy $h_1 = \{15, 16, 17\}, \dots h_4 = \{24, 25, 26\}$. Máme $n = 20, k = 4, p_i = \frac{1}{4}$, $np_i = 5$. dostaneme

$$n_1 = 3,\ n_2 = 7,\ n_3 = 5,\ n_4 = 5$$

$$Q = \sum_{j=1}^{4} \frac{(n_j - 5)^2}{5} = 1.6$$

kvantil $\chi^2_{0.95,3} = 7.81$ takže H_0 nezamítáme.

24. Metoda nejmenších čtverců-lineární regrese

Pokud každý výsledek experimentu sestává z dvojice čísel (x_i, y_i), lineární regrese je metoda nalezení koeficientů β_0, β_1 v lineární závislosti

$$y = \beta_0 + \beta_1 x$$

taková, aby výraz

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

byl minimální.

Jak lze snadno zjistit řešením minimizačního problému,

$$\mathrm{grad} Q = 0$$

$$\sum_{i=1}^{n} y_i - \beta_0 + \beta_1 x_i = 0, \quad \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x_i) x_i = 0$$

má řešení

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Teoretické odvození této metody (dupač p.137)

Předpokládáme, že Y_i jsou nezávislé veličiny se střední hodnotou

$$EY_i = \beta_0 + \beta_1 x_i, \quad i + 1, \ldots, n$$

se společným rozptylem σ^2, a x_i jsou známé hodnoty, alespoň dvě různé.

Věta 24.1. *Předpokládejme, že x_j jsou známé hodnoty*

Hlavni věty

náhodna procházka $S_n = \sum_{j=1}^n X_j$, X_j Bernoulli $X_j = 0, 1$ s pravděpodobnosti $\frac{1}{2}$.

Stepan p. 319:

$$P[S_n = n - 2k] = \binom{n}{k} 2^{-n}, \quad -n \le k \le n$$

Věta 24.2. *Lokalni limitni veta*

$$\sup_{-n\le k\le n} |\frac{\sqrt{n}}{2} P[S_n = n - 2k] - \phi(\frac{n-2k}{\sqrt{n}})| \to 0$$

kde $\phi = f_{N(0,1)}$.

Věta 24.3.

$$\overline{\lim} S_n = +\infty, \quad \underline{\lim} S_n = -\infty$$

Věta 24.4. *Zakon velkych čísel*

$$\lim_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}+\delta}} = 0, \quad s.j\ pro \delta > 0.$$

Věta 24.5. *Zakon iterovaneho logaritmu*

$$\overline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{2n \log\log(n)}} = 1, \quad s.j \ \underline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{2n \log\log(n)}} = -1, \quad s.j$$

Věta 24.6. *(Berry Essen)*

$$\sup_{x\in\mathbb{R}} |P[\frac{S_n}{\sqrt{n}} < x] - \Phi(x)| \le \frac{0.8}{n^{\frac{1}{2}}}$$

kde $\Phi = F_{N(0,1)}$

Věta 24.7. *pravděpodobnost velkych odchylek*

$$\lim \frac{\log P[S_n \ge na]}{n} = \log f(a), \ a \in (0,1)$$

kde

$$f(a) = \frac{1}{2}((\frac{1+a}{1-a})^{\frac{1-a}{2}} + (\frac{1-a}{1+a})^{\frac{1+a}{2}}$$

Označme

$$M_n = \max_{1\le k\le n} S_k, \quad T_n = \mathrm{card}\{1 \le k \le n : S_k > 0\}$$

Věta 24.8.

$$P[\frac{M_n}{\sqrt{n}} \le x] \to 2\Phi(x) - 1$$

Věta 24.9. *Zakon arcusinu*

$$P[\frac{T_n}{n} < x] \to \frac{2}{\pi} \arcsin\sqrt{x}, \quad x \in (0,1)$$

general sums

Věta 24.10. *Zakon velkych čísel (slaby)*

Nechť $X_1, \dots, X_n$ jsou stejně rozdelene nezávislé veličiny, jejíchz střední hodnota je μ a variance σ^2. Potom platí pro každe $\varepsilon > 0$:

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Věta 24.11. *Centralni limitni veta.*

Nechť $X_1, \dots, X_n$ jsou nezávislé stejně rozdelene veličiny se střední hodnotou μ a varianci σ^2. Nechť distribuční funkce proměnné $\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$ ma tvar

$$G_n(x) = P(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x)$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{t^2} dt.$$

Potom $G_n(x) \to N(x)$ konverguje v každem bode pro $n \to \infty$.

Věta 24.12. *(Berry Essen)*

Nechť X_j jsou nezávislé veličiny s nulovou střední hodnotou, $D(X_j) = \sigma_j^2$. Označme $s_n^2 = \sum_{j=1}^n D(X_j)$. Potom

$$\sup |G_n(x) - N(x)| \leq \frac{1}{s_n^3} \sum E(|X_j|^3)$$

Cviceni 1 tyden-množinove operace, kombinatorika

1. Kolik různých slov lze sestavit ruznym usporadanim pismen ABCD?
2. Kolik různých slov lze sestavit ruznym usporadanim pismen AABC?
3. Kolik slov o peti pismenech lze sestavit použitím pismen z množiny $\{A, B, C\}$?
4. Kolik slov o peti pismenech, která obsahuji prave jedno pismeno A, lze sestavit použitím pismen z množiny $\{A, B, C\}$?
5. Kolik slov o peti pismenech, která obsahuji alespon jedno pismeno A, lze sestavit použitím pismen z množiny $\{A, B, C\}$?
6. Kolik slov o peti pismenech lze sestavit použitím pismen z množiny $\{A, B, C\}$ tak, aby každe pismeno melo aspoň jedno zastoupeni?
7. Hazime kostkou 4 krat po sobe. Najdete pravděpodobnost ze:

a. všechna čísla jsou navzájem ruzna.

b. dve čísla jsou stejna a zbyla dve ruzna.

c. 3 jsou stejna a 1 ruzne

8. ůkazte ze počet nezapornych celočíselnych reseni rovnice $x_1 + \cdots + x_n = k$ je roven $\binom{n+k-1}{n-1}$.
9. Najdete počet kladnych celočíselnych reseni rovnice $x_1 + \cdots + x_n = k$.
10. Kolika zpusoby lze rozdelit k darku n-detem ($k \geq n$), pokud každe dite ma aspoň jeden darek?
11. Kolika zpusoby lze rozdelit 5 cernych a 5 bilych kouli do tri navzájem různých krabic?
12. Kolika zpusoby lze rozdelit 5 kouli různých barev do tri navzájem různých krabic?

Kolik navzájem různých slov lze sestavit prerovnanim pismen AABBBC?

jaká je pravděpodobnost ze SPZ skladajici se ze 6-číslic ma každou číslici jinou?

jaká je pravděpodobnost ze pri zamichani baliku bridgeovych karet vyjdou všechna srdce navrch?

Kolik kombinaci s opakovanim (tj. souboru prvku které se mohou pakovat a u nichz nezáleží na poradi) délky k lze vytvorit ze skupiny n prvku, jestlize každý prvek ma byt zastoupen aspoň jednou? ($\binom{k-1}{n-1}$)

Kolik kombinaci s opakovanim (tj. souboru prvku které se mohou pakovat a u nichz nezáleží na poradi) délky k lze vytvorit ze skupiny n prvku? ($\binom{n+k-1}{n-1}$)

Kolik je posloupnosti $a_1 \leq a_2 \leq a_3$ setavajicich z čísel z množiny $\{1, \ldots, 7\}$?

Hazime n krat minci, na jedne strane je číslo 0 na druhe 1. jaká je pravděpodobnost ze průměr výsledku hodu je roven $\frac{1}{2}$, resp. nejvýše $\frac{1}{2}$? Uziti symetrie.

Vytahneme náhodne osm ponozek ze supliku ve kterém je 10 paru ponozek. jaká je pravděpodobnost ze jsme vytahli prave n paru ponozek?

Cviceni 2 a 3 tyden-pravděpodobnostní prostory a kombinatorika

1. jaká je pravděpodobnost ze ve skupine n osob maji alespon dva lide narozeniny ve stejný den?

2. Pokerova rozlozeni jsou následujíchich typu:

1. Flush: 5 karet teze barvy
2. Flushova postupka: 5 po sobe jdoucich karet teze barvy.
3. 4 stejně karty+1
4. full house: 3+2 stejně
5. 3 stejně a zbytek dve ruzne
6. postupka (ne nutne stejně barevnych karet).

Spocitejte pravděpodobnost jednotlivych rozlozeni.

3. Listky očíslovane 1,2,3,4 jsou náhodne vlozene do obalek opatrenych stejnými čísly. jaká je pravděpodobnost ze všechna čísla obalek odpovidaji vlozenym číslum?

5. jaká je pravděpodobnost ze pri náhodnem rozmisteni n muzu a n zen kolem kulateho stolu sedi vedle každeho muze po obou stranach zeny?

6. V sade n vyrobku je 6 vadnych. výběreme náhodne k z cele sady. jaká je pravděpodobnost ze vzorek obsahuje prave jeden vadny jestlize výběr je

a. bez navraceni

b. s navracenim

7. Hazime minci dokud nepadne dvakrát po sobe totez. jaká je pravděpodobnost ze hra bude mit sudy počet hodu?

8. Tahame náhodne karty z balicku 32 karet. jaká je pravděpodobnost ze n-ta sejmuta karta bude mit mensi hodnotu nez první? Ze mezi prvními n-kartami budou aspoň 2 esa?

9. výběreme náhodne 3 čísla z množiny $\{1,\ldots,n\}$. jaká je pravděpodobnost ze zadna dve z nich nejsou po sobe jdouci?

10. výběreme 3 náhodne body na kružnici. jaká je pravděpodobnost ze tvori ostrouhly trojuhelnik?

11. Uvnitr koule o polomeru 1 je vybráno náhodne n bodu. jaká je pravděpodobnost ze nejblizssi bod ma vzdálenost k počátku nejvýše r?

12. Rozdame trem hracum po peti kartach z 32. jaká je pravděpodobnost ze aspoň jeden z nich dostane eso? Ze všichni 3 dostanou eso?

Vypoctete pravděpodobnost náhodných rozdělení 3 různých kouli do tri různých krabic. Nyni predpokládejme ze koule jsou nerozlisitelne, coz vede k pravděpodobnostnímu rozdělení pro tri stejene koule do tri různých krabic. V prirode ale platí pro nekteré castice jine pravidlo, napr. pro fotony platí experimentalni fakt ze každe rozdělení 3 stejných fotonu do tri různých krabic ma stejnou pravděpodobnost.

Cviceni 4 a 5 tyden. podminena pravděpodobnost a nezavislost

1. Spocitejte $P(A \cup B)$ pokud vite ze $P(A) = \frac{1}{3}$ a $P(B|A^c) = \frac{1}{4}$.

2. Spocitejte $P(B)$ pokud vite ze $P(A \cup B) = \frac{2}{3}$ a $P(A^c|B^c) = \frac{1}{2}$.

3. Dva nezávislé jevy A, B splňuji $P(B|A \cup B) = \frac{2}{3}$ a $P(A|B) = \frac{1}{2}$. Najdete $P(B)$.

4. jaká je pravděpodobnost ze 2 deti v rodine jsou oba chlapci, pokud vime ze aspoň jeden je chlapec? Ze starsi z nich je chlapec?

5. součet dvou hodu kostkou je 8. jaká je pravděpodobnost ze aspoň jeden z hodu byl 6?

6. Hodíme 3 kostkami. jaká je pravděpodobnost ze padla aspoň jedna setka jestlize vime ze všechny hody byly ruzne?

7. Pri hodu 10 kostkami padla aspoň jedna 1. jaká je pravděpodobnost ze padly aspoň 2 jednicky?

8. Hodíme 2 kostkami. A-první číslo je sude, B-druhe číslo je liche, C-součet hodu je lichy. ůkazte ze jevy jsou po dvou nezávislé, ale nejsou nezávislé.

9. Kolikrat musíme hodit dvěma kostkami, aby pravděpodobnost ze padla aspoň jedna 6 byla větší nez $\frac{1}{2}$?

10. V krabici je jeden listek s číslem 1 a dva listky s číslem 2. Vytahneme n krat po sobe listek z krabice, a potom jej zase vratime zpet. jaká je pravděpodobnost ze součet všech tazenych čísel je delitelny číslem n?

11. Uvazme hru kdy hrac hodi ferovou kostkou. Pokud padne 3 nebo mene, prohrava. Jinak výběre náhodne tolik karet z balicku na stole kolik mu padlo na minci. Vyhrava pokud jedna z karet je eso. jaká je pravděpodobnost ze hrac vyhraje? Pokud vyhral, jaká je pravděpodobnost ze hodil kostkou číslo 6?

12. Mame n minci a mezi nimi dve falesne, na kterých pada panna s pravděpodobnosti p. výběreme náhodne jednu minci a Hodíme ji desetkrat. Pokažde padne panna. jaká je pravděpodobnost ze mince je falesna?

13. Mame 2 kostky. Na jedne jsou 4 steny bile a dve cerne, na druhe jsou 3 cerne a tri bile. náhodne výběreme kostku a Hodíme s ni destkrat za sebou. Pokažde padne bila. jaká je pravděpodobnost ze hazime prvi kostkou? předpokladejem ze dvakrát padla bila. jaká je pravděpodobnost ze i potreti padne bila?

14. V krabici je 2n míčků. náhodne výběreme skupinu míčků. Je-li jejích počet sudy vratime ji do krabice, jinak je nechame venku. Podruhe sahname do krabice a výběreme skupinu míčků. S jakou pravděpodobnosti jsme podruhe vybrali 4 míčky?

15. Mame testovaci metodu pro vyrobek která nefunguje vždy. Pokud je vyrobek defektni, test ůkaze ze je defektni s pravděpodobnosti 0.9. Pokud je vyrobek funkcni, test ůkaze rekne ze je defektni s pravděpodobnosti 0.2. Nyni otestujeme 3 vyrobky s výsledkem testu funkcni. jaká je pravděpodobnost ze všechny vyrobky jsou skutecen funkcni?

16. Mame 3 modre, 4 zelene a 9 cervenych kouli. Koule náhodne rozmistime do 3 krabic. jaká je pravděpodobnost ze v každe krabici jsou koule všech tri barev?

Motivacni Příklad: ruska ruleta. v bubinku jsou 2 naboje vedle sebe, pri prvním zmacknuti byla komora prazdna. Je lepsi zmacknout znovu nebo náhodne pretocit?

Hodíme dvěma kostkami. jaká je pravděpodobnost ze součet hodu je 4 za předpokladu ze jeden z hodu je 3?

Hodíme 3 kostky. jaká je pravděpodobnost ze padla aspoň jedna 6 vime-li ze padla navzájem ruzna čísla?

Mame tri krabice K_1, K_2, K_3, ve kterých je n_j bilych a m_j cernych míčků.

b. Z krabice K_1 výběreme náhoden micek a vlozime ho do K_2. Potom výběreme náhodne micek z K_2 a vlozime ho do K_1. jaká je pravděpodobnost ze slozeni K_1 a K_2 zustala stejna?

c. Z krabice K_1 preHodíme náhodný micek do K_2, potom z K_2 do K_3 a nakonec z K_3 do K_1. jaká je pravděpodobnost ze slozeni krabic zustala stejna?

Cviceni 6, 7 a 8 tyden, náhodná veličiny

Najdete pravděpodobnostní rozdělení (nebo hustotu), distribuční funkci, střední hodnotu a rozptyl veličiny zadane rozdělením

$$P(X=-3)=\frac{1}{4}, P(X=1)=\frac{2}{4}, P(X=3)=\frac{1}{4}$$

1. Rozdelim balicek 32 mariasovych karet na dve poloviny, X je počet sedmicek v prave hromadce karet.

2. V krabici je devet bilych míčků a jeden cerny. Budeme náhodne vybirat míčky z krabice dokud nevytahneme cerny. X je počet tahu ve hre.

3. V krabici je 20 zarovek, z nich 3 jsou vadne. výběreme náhodne 5 zarovek, X je počet vadnych zarovek ve výběru.

4. Hazim kostkou dokud nepadne 6. X je počet hodu.

5. Na tramvajive zastavce je 6 cestujicich, kteri nahidne nastoupi do jednech ze 4 dveri. X je počet cestujicich kteri nastoupi poslednimi dvermi.

6. V krabici je 5 cernych a 5 bilych kouli. Tahame postupně koule bez vraceni. X je první tah ve kterém jsme vytahli bilou kouli.

7. Kolika kostkami najednou musíme hodit aby pravděpodobnost ze padne aspoň jedna 1 byla alespon $\frac{1}{2}$?

8. Najdete distribuční funkci, střední hodnotu a rozptyl veličiny X, která je zadana hustotou $f_X(x)=\frac{1}{2}$ na $[-1,1]$ a nula jinde.

9. X ma distribuční funkci $F_X(x)=\frac{1}{2}+\frac{1}{\pi}\arctan(x)$. Najdete její hustotu a střední hodnotu.

10. Na sfere jsou vybrány náhodne dva body, X je jejích vzdálenost.

11. Na sfere je vybrán náhodne bod P, X je délka projekce vektoru OP na rovinu xy.

12. Na sfere je vyznacen pevny bod A a vybrán náhodne bod P, X je délka vektoru AP.

13. Na sfere je vyznacen pevny bod A a vybrán náhodný vektor v tak ze bod P je prunikem vektoru v z bodu A a sfery. X je délka vektoru AP.

14. Na kružnici k o stredu $(0,1)$ a polomeru 1 je dan bod $A=(0,2)$. Zvolíme náhodne bod B na k a najdeme prusecik P primky AB s osou x. X je délka OP.

15. V jednotkovem čtverci ABCD výběreme náhodný bod P, X je plocha trojuhelniku ABP.

16. V jednotkovem čtverci ABCD výběreme náhodný bod P, X je vzdálenost P od hranice čtverce.

17. Rozrezeme usecku $[0,1]$ na $n+1$ dilu pomoci n náhodných bodu. ůkazte ze každý interval deleni ma stejnou distribuční funkci délky a najdete tuto funkci.

18. Autobus ve kterém je 9 cestujicich zastavuje pouze na známeni v 5 zastavkach. Cestujici vystupuji náhodne a na sobe nezávislé. X je počet zastavek.

19. náhodna procházka. Bod je v case $t=0$ v počátku číslne osy. V case $n+1$ se posune o jedna doprava nebo doleva s pravděpodobnosti $\frac{1}{2}$ z předchozí pozice v case n. X je cas prvního návratu bodu zpet do počátku.

20. Rozdelime náhodne k míčků do n krabic. X je počet neprazdnych krabic.

Karetni hra. první hrac dostane 1 kartu a druhý hrac 2 karty. Pokud první hrac dostane eso, druhý zaplatí 10 korun; pokud dostane krale druhý zaplatí 5 korun. Pokud první hrac ma cervenou barvu a druhý nikoli, druhý zaplatí 1 korunu. V ostatních případech první hrac platí druhemu 1 korunu. Zjistete střední hodnotu vyher.

Rozrezeme usecku $[0,1]$ na $n+1$ dilu pomoci n náhodných bodu. ůkazte ze každý interval deleni ma stejnou distribuční funkci délky.

Cviceni 9, 10 a 11. tyden, transformace pravděpodobnostních veličin, nezavislost a spolecne rozdělení veličin

Problem: k zadane veličine X a funkci $g : I\!R \to I\!R$ vytvorime novou promennou $Y = g \circ X$.

1. X je uniformne rozdelena na množine $\{1, \dots, 6\}$, $g(x) = \sin(\frac{\pi}{2}x)$, $h(x) = \cos(\frac{\pi}{2}x)$. Najdete rozdělení $Y = gX, Z = hX, Y^2 + Z^2$.

2. X je uniformne rozdelena na $[2, 4]$ najdete rozdělení $aX + b$, $a > 0, b \in I\!R$. Totez pro obecnou promennou.

3. X ma hustotu rozdělení $f_X(x) = \frac{3}{4}x(2 - x)$ pro $0 \leq x \leq 2$ (jinak nula). Najit distribuční funkci X, a tez $\sqrt{X}$ a hustotu.

4. spojitá veličina X s hustotou f_X ma pouze pozitivni hodnoty, $Y = \frac{1}{X}$. Vyjadrete pravděpodobnostní hustotu a rozdělení Y v zavislosti na rozdělení veličiny X.

5. spojitá veličina X, $Y = e^X$. Vyjadrete pravděpodobnostní hustotu a rozdělení Y v zavislosti na rozdělení veličiny X.

6. Dokazte ze součet dvou nezávislých Poissonovych proměnných je opet Poissonova promenna.

7. X, Y jsou nezávislé stejnoměrně rozdelene veličiny na $[-\frac{1}{2}, \frac{1}{2}]$. Najdete rozdělení $X + Y$, $X - Y$, XY.

8. Pokud X, Y jsou nezávislé veličiny s rozdělením $N(0, 1)$. Najdete rozdělení $\sqrt{X^2 + Y^2}$.

9. X, Y jsou nezávislé veličiny s rozdělením $N(0, 1)$. Dokazte ze rozdělení $X + Y, X - Y$ jsou nezávislá. Dokazte ze rozdělení $X^2 + Y^2$, $\frac{X}{Y}$ jsou nezávislá.

10. (X, Y) je náhodne vybrány bod v trojuhelniku s vrcholy $(0, 0), (0, 1), (1, 0)$. Zjistete jestli X, Y jsou nezávislé nebo nekorelovane.

11. Na usecce $[0, 1]$ je vybrány bod a. X je rovnoměrně rozdelena veličina na intervalu, Y je vzdálenost tohoto náhodneho bodu od a. Zjistete zda X, Y jsou korelovane.

Cviceni 12 . tyden, intervalove odhady

1. Mejme generator který vytvari náhodna čísla $\{0,1\}$ se stejnou pravděpodobnosti. jaká je pravděpodobnost ze mezi 10000 čísly bude cetnost jednicek v rozmezi $[4900,5100]$.

2. Bylo secteno 300 čísel zaokrouhlenych na jedno desetinne misto. Zaokrouhlovaci chyba tedy nepresahuje $300 \times 0.05 = 15$. jaká je pravděpodobnost ze chyba nepresahne 1?

3. Hodím n krat minci. Kolikrat musím hodit aby cetnost η_n jednicek splnovala
$$|\eta_n - \frac{1}{2}| < 0.05$$
s pravděpodobnosti 95 procent? Porovnejte odhady získáne ze zakona velkych čísel a CLV.

4. Pojistovna pojistuje 1000 lidi stejněho veku zivotni pojistkou v cene 1200 Kc na castku 80 000Kc. pravděpodobnost umrti je 1 procento. Analyzujte zisk pojistovny.

5. Vyletni clun ma nosnost 5000kg. Vaha cestujiciho je normalne rozdelena náhodna veličina se střední hodnotou 70kg a rozptylem 4kg. Kolik cestujicich muze cestovat aby pravděpodobnost pretizeni byla mensi nez 1 promile?

6. Pan Novak se procházi mestem o kterém předpokladame ze ma pravouhly system ulic (sever, jih, vychod, zapad). Na každe krizovatce se rozhoduje náhodne kudy dal se stejnou pravděpodobnosti. Urcete kolika krizovatkami musí projit abychom s pravděpodobnosti 95 procen mohli tvrdit ze alespon v petine případu se rozhodl jit na sever.

7. Mame dve mince, ferovou a falesnou která dava pannu s pravděpodobnosti 70 procent. Kolik pokusnych hodu je třeba udelat abychom mohli s jistotou 90 procent urcit zda mince je falesna?

8. Vybirame náhodne cela čísla od 1 do 10. Kolik čísel je třeba vybrat abychom s pravděpodobnosti 95 procent mohli ocekavat ze aspoň jedna tretina vybránych čísel byla prvočísla?

zbytky
Vybirame náhodne z osudi čísla z množiny $\{1, \dots, n\}$ bez vraceni. jaká je pravděpodobnost ze druhým tahem tahneme 2, jestlize jsme prvním tahem tahli 1?

Hazime kostkou tak dlouho dokud nepadne 6. Nepadla-li 6 pri prvním hodu, jaká je pravděpodobnost ze nepadna ani pri pristich dvou hodech? Ω NEBUDE Bernoulliho nezávislé jevy, binarni strom. nybrz $\Omega = \{(a_i) : a_1, \dots, a_n = 0, a_{n+1} = 6\}$. Bez dodatecne informace by otazka byla jaká je pravděpodobnost ze posloupnost bude délky aspoň 3.

Mathematical Institute, Czech Academy of Science, Žitná 25, 115 67 Praha 1, Czech Republic, and Department of Mathematics, Faculty of Electrical Engineering, Czech Technical University in Prague, Zikova 4, 160 00, Prague
E-mail address: hajek@math.cas.cz