Pravděpodobnost a statistika; Opakování pro rozpoznávání

Václav Hlaváč

České vysoké učení technické v Praze

Český institut informatiky, robotiky a kybernetiky

160 00 Praha 6, Jugoslávských partyzánů 1580/3

http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

Poděkování: T. Brox, V. Franc, R. Gutierrez-Osuna, M. Navara, M. Urban.

Osnova přednášky:

- Pravděpodobnost × statistika.
- Náhodné jevy.
- Sdružená, podmíněná pravděpodobnost.
- Bayesova věta.
- Distribuční funkce, hustota.
- Charakteristiky náhodných veličin.

Doporučené čtení

- M. Navara: Pravděpodobnost a matematická statistika, skriptum FEL ČVUT, Vydavatelství ČVUT, Praha 2007.
- J. Novovičová: Pravděpodobnost a matematická statistika. Skriptum Fakulty dopravní ČVUT, Vydavatelství ČVUT, Praha 2002.
- A. Papoulis: Probability, Random Variables and Stochastic Processes, McGraw Hill, Edition 4, 2002.
- H. Pishro-Nik: Introduction to probability, statistics, and random processes. Kappa Research LLC, 2014. Freely available at https://www.probabilitycourse.com
- http://mathworld.wolfram.com/
- http://www.statsoft.com/textbook/stathome.html

Pravděpodobnost, motivační příklad

- Los loterie se prodává za **cenu** 2 EUR.
- 1 los z 1000 vyhrává 1000 EUR, ostatní nic. (Tím je dána **hodnota** losu po losování.)
- Za kolik máme prodat los *před* losováním?

- Za 2 EUR by ho koupil jen hlupák. (Nebo ne?)
- Hodnota losu před losováním je $\frac{1}{1000}1000 = 1$ EUR $=$ průměrná hodnota po losování.

Na to je teorie pravděpodobnosti.

Otázka "Loterie": Proč se přesto kupují losy a loterie fungují?

Statistika, motivační příklad

Dosud jsme předpokládali, že parametry pravděpodobnostního modelu jsou známy. To je málokdy splněno.

Příklad (Sportka): Na Sportce se normálně prodělává; jelikož jsou výhry stanoveny podle počtu výherců, je výhodnější sázet jinak než ostatní. K tomu potřebujeme vědět, podle jakého modelu sázejí ostatní.

Příklad (ruleta): U rulety se obě strany zajímají, zda padají všechna čísla se stejnou pravděpodobností, přesněji, jak velké jsou odchylky od rovnoměrného rozdělení. Ale jak to zjistit a jaké je riziko chybných závěrů?

Na to je statistika.

Statistika poskytuje daleko víc: nástroj pro zkoumání světa, pro hledání a ověřování závislostí, které nejsou zjevné.

Pravděpodobnost, statistika

- Pravděpodobnost: Pravděpodobnostní popis $\Longrightarrow$ budoucí chování systému.
  - je teorie (nástroj) pro účelné rozhodování v situacích, kdy výsledek náhodných jevů závisí na okolnostech, které známe jen částečně.
  - Poskytuje model takových systémů a slouží jako nástroj pro kvantifikaci výsledků.
- Statistika: Chování systému $\Longrightarrow$ pravděpodobnostní popis.
  - Statistika je nástroj pro hledání a ověřování pravděpodobnostního popisu reálných systémů na základě jejich pozorování.
  - Poskytuje daleko víc: Nástroj pro zkoumání světa, pro hledání a ověřování závislostí, které nejsou zjevné.
  - Dva typy: Popisná nebo inferenční statistika.
  - Sběr, organizace a analýza dat.
  - Zobecňuje z omezených / konečných vzorků.
  - Odhad parametrů, testování hypotéz, atd.

Náhodné jevy, pojmy

Náhodný pokus

Prostor elementárních jevů je neprázdná množina Ω všech možných výsledků daného pokusu.

Elementární jevy $\omega \in \Omega$ jsou prvky prostoru elementárních jevů (výsledky pokusu).

Jevové pole $\mathcal{A}$ je tvořeno systémem všech podmnožin prostoru elementárních jevů Ω.

Náhodný jev $A \in \mathcal{A}$ je prvkem jevového pole.

Poznámka: pojem náhodného jevu byl zaveden proto, aby bylo možné definovat pravděpodobnost, rozdělení pravděpodobnosti, atd.

Pravděpodobnost, zavedení

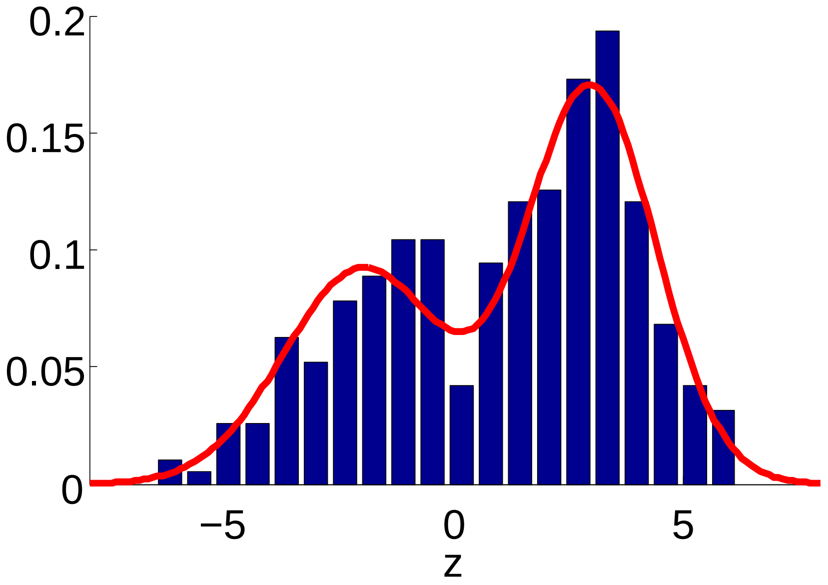
- Klasická (P.S. Laplace, 1812. Dnes se již nepovažuje za definici pravděpodobnosti, ale za metodu odhadu pravděpodobnosti)

$$P(A) \approx \frac{N_A}{N}$$

- Limitní (četnostní)

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

- Axiomatická
(Andreje Nikolajeviče Kolmogorova 1930)



histogram × spojitá hustota pravděpodobnosti

Axiomatická definice pravděpodobnosti, 1930

- Ω - prostor elementárních jevů
- $\mathcal{A}$ - jevové pole

1. $P(A) \geq 0, \quad A \in \mathcal{A}.$
2. $P(\Omega) = 1.$

   Neformálně: Pokaždé, když se uskuteční experiment, poskytne nějaký výsledek.
3. Jestliže $A \cap B = \emptyset$, pak $P(A \cup B) = P(A) + P(B)$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Andrej Nikolajevič Kolmogorov

* 1903, † 1987

Fine, T. (2014). Theories of Probability: An Examination of Foundations. Academic Press.

Tři Kolmogorovovy axiomy nám neříkají (neposkytují): (a) Kde a kdy se mají použít; (b) Návody nebo postupy pro výpočet pravděpodobností;(c) Vhled do podstaty náhodných procesů.

Pravděpodobnost

je funkce P, která jevům přiřazuje čísla z intervalu $[0,1]$ a splňuje podmínky

(P1) $P(true) = 1$,

(P2) $P\left(\bigvee_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n)$, pokud se jevy $A_n,\ n \in \mathbb{N}$, navzájem vylučují.

Z těchto podmínek vyplývá:

- $P(false) = 0$
- $P(\neg A) = 1 - P(A)$,
- jestliže $A \subseteq B$, pak $P(A) \leq P(B)$.

Poznámka: Pro korektnost je potřeba, aby systém jevů splňoval určité další podmínky.

Odvozené vztahy

- Jestliže $A \subset B$, pak $P(B \setminus A) = P(B) - P(A)$.
  Symbol $\setminus$ označuje množinový rozdíl.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- Statistická nezávislost: $P(A \cap B) = P(A)\, P(B)$
  Slovy: Jevy A a B jsou nezávislé, když znalost o výskytu jevu A nám neříká nic o výskytu jevu B.

Sdružená pravděpodobnost, marginalizace

- Sdružená pravděpodobnost $P(A, B)$, také někdy označovaná $P(A \cap B)$, je pravděpodobnost, že jevy A, B nastanou současně.

- Sdružená pravděpodobnost je symetrická: $P(A, B) = P(B, A)$.

- Marginalizace (neformálně sčítací pravidlo, ignoruje se proměnná(é)):
  $P(A) = \sum_B P(A, B)$ umožňuje vypočítat pravděpodobnost jevu A ze sdružené pravděpodobnosti $P(A, B)$ jako $P(A, B)$ přes všechny možné jevy B.

- Pravděpodobnosti $P(A)$ se říká marginální pravděpodobnost.

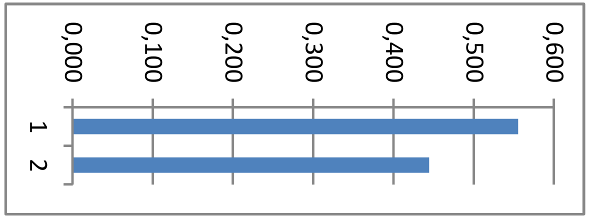
Kontingenční tabulka, marginalizace
Příklad – závod v orientačním běhu

Orienteering competition example, participants

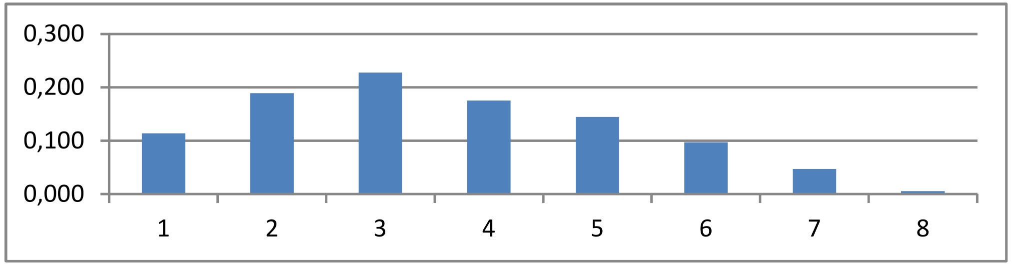
Age	<= 15	16-25	26-35	36-45	46-55	56-65	66-75	>= 76	Sum
Men	22	36	45	33	29	21	12	2	200
Women	19	32	37	30	23	14	5	0	160
Sum	41	68	82	63	52	35	17	2	360

Orienteering competition example, frequency

Age	<= 15	16-25	26-35	36-45	46-55	56-65	66-75	>= 76	Sum
Men	0,061	0,100	0,125	0,092	0,081	0,058	0,033	0,006	0,556
Women	0,053	0,089	0,103	0,083	0,064	0,039	0,014	0,000	0,444
Sum	0,114	0,189	0,228	0,175	0,144	0,097	0,047	0,006	1



Marginal probability P(sex)

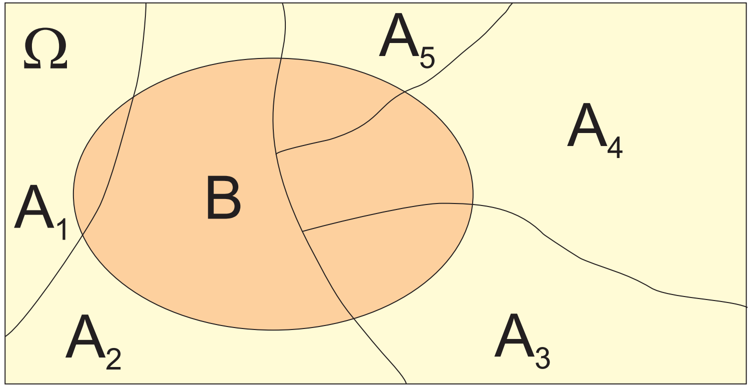


Marginal probability P(Age_group)

Použití rozdělení množiny na části

Když se jevy A_i vzájemně vylučují a rozdělují zcela prostor jevů Ω, tj.

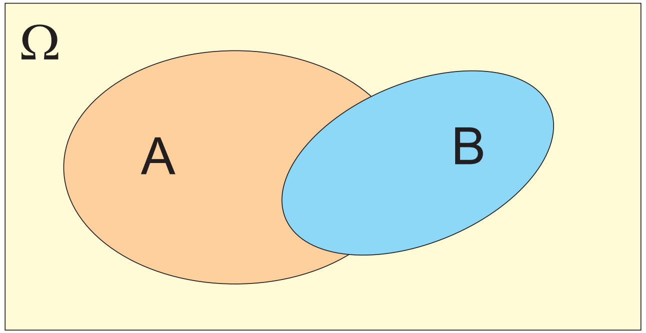
$$A_i \cap A_j = \emptyset,\ \text{for}\ \forall\, i,j\,,\quad \bigcup_{i=1,\ldots,n} A_i = \Omega\,,\quad \text{potom}\ P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$



Podmíněná pravděpodobnost

- Mějme pravděpodobnostní popis systému daný sdruženou pravděpodobností $P(A, B)$.
- Dostaneme-li dodatečnou informaci, že nastal jev B, změní se naše znalost o pravděpodobnosti jevu A na

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



což je *podmíněná pravděpodobnost* jevu A za podmínky B.

- Podmíněná pravděpodobnost je definována pouze pro $P(B) \neq 0$.
- Pravidlo součinu: $P(A, B) = P(A|B)\, P(B) = P(B|A)\, P(A)$.
- Ze symetrie sdružené pravděpodobnosti a pravidla součinu lze odvodit Bayesovu větu. (Uvedeme pro obecnější formulaci pro více než dva jevy).

Vlastnosti podmíněné pravděpodobnosti

- $P(true|B) = 1$, $P(false|B) = 0$.
- Pokud $A = \bigvee_{n\in\mathbb{N}} A_n$ a jevy $A_1, A_2, \ldots$ se navzájem vylučují, pak
  $P\left(A|B\right) = \sum_{n\in\mathbb{N}} P(A_n|B)$.
- Jevy A, B jsou *nezávislé*, právě když $P(A|B) = P(A)$.
- Pokud $B \Rightarrow A$, pak $P(A|B) = 1$. Pokud $B \Rightarrow \neg A$, pak $P(A|B) = 0$.

- Jevy B_i, $i \in I$, tvoří *úplný systém jevů*, jestliže se navzájem vylučují a $\bigvee_{i\in I} B_i = true$.
- Úplný systém jevů má tu vlastnost, že nastane právě jeden z nich.

Příklad: podmíněná pravděpodobnost

Jeden hod kostkou. 

Jaká je pravděpodobnost, že padne číslo > 3 (jev A) za podmínky, že padlo liché číslo (jev B).

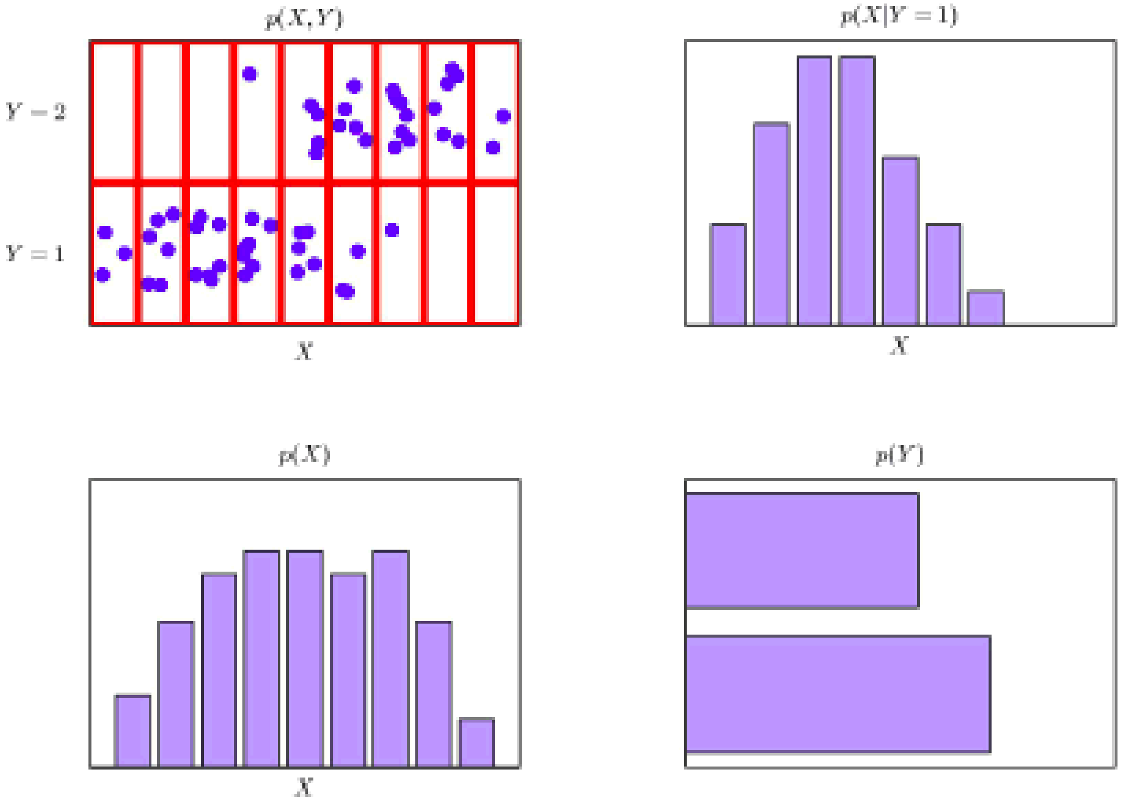
$$\Omega = \{1,2,3,4,5,6\}\,, \quad A = \{4,5,6\}\,, \quad B = \{1,3,5\}$$

$$P(A) = P(B) = \frac{1}{2}$$

$$P(A,B) = P\left(\{5\}\right) = \frac{1}{6}$$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Sdružená a podmíněná pravděpodobnost, příklad



Courtesy T. Brox

Věta o úplné pravděpodobnosti

Nechť B_i, $i \in I$, je úplný systém jevů a $\forall i \in I \colon P(B_i) \neq 0$.

Pak pro každý jev A platí

$$P(A) = \sum_{i \in I} P(B_i)\, P(A|B_i)\,.$$

Důkaz:

$$P(A) = P\left(\left(\bigvee_{i \in I} B_i\right) \wedge A\right) = P\left(\bigvee_{i \in I} (B_i \wedge A)\right)$$

$$= \sum_{i \in I} P(B_i \wedge A) = \sum_{i \in I} P(B_i)\, P(A|B_i)\,.$$

Bayesova věta

(Thomas Bayes *1702 - †1761)

Nechť B_i, $i \in I$, je úplný systém jevů a $\forall i \in I$: $P(B_i) \neq 0$.

Pak pro každý jev A splňující $P(A) \neq 0$ platí $P(B_i|A) = \dfrac{P(B_i)\, P(A|B_i)}{\sum_{i \in I} P(B_i)\, P(A|B_i)}$,

kde $P(B_i|A)$ je aposteriorní pravděpodobnost; $P(B_i)$ je apriorní pravděpodobnost a $P(A|B_i)$ jsou známé podmíněné pravděpodobnosti (též věrohodnosti) jevu A, když známe pozorování B_i.

Důkaz (s využitím věty o úplné pravděpodobnosti):

$$P(B_i|A) = \frac{P(B_i \wedge A)}{P(A)} = \frac{P(B_i)\, P(A|B_i)}{\sum_{i \in I} P(B_i)\, P(A|B_i)}\,.$$

Význam Bayesovy věty

- Bayesova věta je základním nástrojem strojového učení (rozpoznávání). Známe (pozorujeme) jevy B_i, kde $i \in I$ představuje rozdělení prostoru jevů na části. Předpokládejme, že nastane jev A. Bayesova věta dovoluje optimálně odhadnout, který z jevů B_i nastal.
- Podmíněné pravděpodobnosti $P(A|B_i)$ (někdy se jim říká věrohodnosti) se odhadují pomocí experimentů nebo ze statistického modelu.
- Když známe věrohodnosti $P(A|B_i)$, můžeme určit aposteriorní pravděpodobnosti $P(B_i|A)$, které slouží k optimálnímu odhadu, který jev z B_i nastal.
- K výpočtu aposteriorní pravděpodobnosti $P(B_i|A)$ musíme znát *apriorní* pravděpodobnosti $P(B_i)$.
- Neformálně: *aposteriorní* $\propto$ (*apriorní* $\times$ *podmíněné pravděpodobnosti*) jevu při známých pozorováních.

- Podobně definujeme podmíněné rozdělení náhodné veličiny, podmíněnou hustotu spojité náhodné veličiny apod.

Maximálně věrohodný odhad (ML) a odhad s maximální aposteriorní pravděpodobností (MAP)

Opišme Bayesovu větu z průsvitky 19

$$P(B_i|A) = \frac{P(B_i)\,P(A|B_i)}{\sum_{i\in I} P(B_i)\,P(A|B_i)}\,.$$

- Apriorní pravděpodobnosti je pravděpodobnost $P(B_i)$, která nevyužívá znalost z pozorování (experimentů).
- Věrohodnost $P(A|B_i)$ (podmíněná pravděpodobnost jevu A za podmínky B_i) ohodnocuje kandidáta na výstupu měření. Postupu hledání takového ohodnocení výstupu, které maximalizuje věrohodnost, se říká postup maximální věrohodnosti, zkráceně ML postup.
- Aposteriorní pravděpodobnost $P(B_i|A)$ je pravděpodobnost jevu B_i, když se vezme v úvahu pozorování (měření). Postupu, který maximalizuje aposteriorní pravděpodobnost, se říká postup maximální aposteriorní pravděpdobnosti, zkráceně MAP postup.

Podmíněná nezávislost

Náhodné jevy A, B jsou podmíněně nezávislé za podmínky C, jestliže

$$P(A \wedge B|C) = P(A|C)\, P(B|C)\,.$$

Podobně definujeme podmíněnou nezávislost více jevů, náhodných veličin apod.

Nezávislé jevy

Jevy A, B jsou nezávislé $\Leftrightarrow P(A \cap B) = P(A)\, P(B)$.

Příklad

Jeden hod kostkou. Jevy $A > 3$, jev B liché číslo. Jsou jevy nezávislé? Spočteme pravděpodobnosti na obou stranách rovnice a ověříme rovnost.

$$\Omega = \{1,2,3,4,5,6\}\,, \quad A = \{4,5,6\}\,, \quad B = \{1,3,5\}$$

$$P(A) = P(B) = \frac{1}{2}$$

$$P(A \cap B) = P(\{5\}) = \frac{1}{6}$$

$$P(A)\, P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$P(A \cap B) \neq P(A)\, P(B) \Leftrightarrow \text{jevy jsou závislé.}$$

Náhodná veličina

- Náhodná veličina je libovolná funkce $X:\Omega \rightarrow \mathbb{R}$, kde Ω je prostor elementárních jevů.
- Proč byl zaveden pojem náhodné veličiny? Dovoluje pracovat s pojmy jako distribuční funkce, hustota pravděpodobnosti, matematické očekávání (střední hodnota), atd..
- Existují dva základní typy náhodných veličin:
  - Diskrétní – mají spočitatelně hodnot. *Příklady: vrhací kostka, počet aut, které ulicí projela za hodinu.*

    Diskrétní pravděpodobnost je dána $P(X = a_i) = p(a_i)$, $i = 1, \ldots$, $\sum_i p(a_i) = 1$.
  - Spojité – hodnoty jsou z intervalu, tedy z nekonečného množství hodnot. *Example: výška člověka.*

    Spojitá pravděpodobnost je dána distribuční funkcí nebo hustotou pravděpodobnosti.

Distribuční funkce náhodné veličiny

Distribuční funkce náhodné veličiny X je funkce $F\colon X \to [0,1]$ je definovaná pomocí $F(x) = P(X \le x)$, kde P je pravděpodobnost.

Vlastnosti:

1. $F(x)$ je neklesající funkce, tj. pro $\forall$ dvojici $x_1 < x_2$ platí $F(x_1) \le F(x_2)$.
2. $F(X)$ je zprava spojitá, tj. platí $\lim\limits_{h\to 0^+} F(x+h) = F(x)$.
3. - Pro každou distribuční funkci platí $\lim\limits_{x\to-\infty} F(x) = 0$ a $\lim\limits_{x\to\infty} F(x) = 1$. Zapsáno zkráceně: $F(-\infty) = 0$, $F(\infty) = 1$.
   - Jestliže jsou možné hodnoty $F(x)$ z intervalu (a,b), pak $F(a) = 0$, $F(b) = 1$.

Každou funkci splňující předchozí tři vlastnosti můžeme pokládat za distribuční funkci.

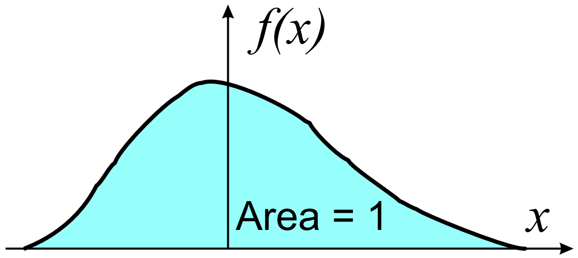
Spojitá distribuční funkce a hustota

- Distribuční funkce F se nazývá absolutně spojitá, jestliže existuje nezáporná funkce f (hustota pravděpodobnosti) a platí

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \qquad \text{pro každé } x \in X.$$

- Hustota pravděpodobnosti splňuje

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1\,.$$



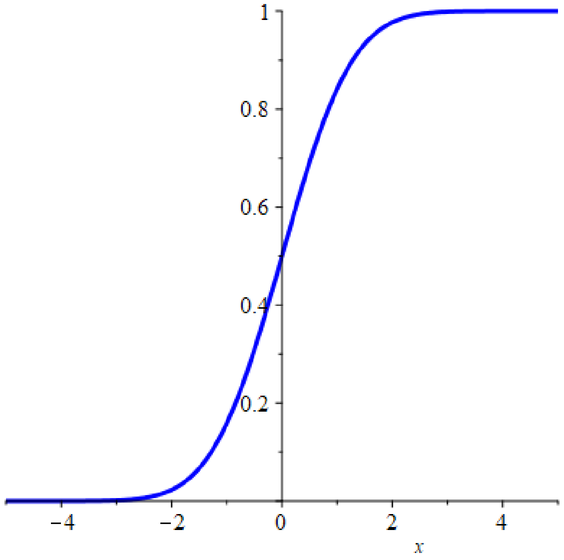
- Existuje-li derivace $F(x)$ v bodě x, potom $F'(x) = f(x)$.
- Pro $a, b \in \mathbb{R}$, $a < b$ platí $P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$.

Příklad, normální (= gaussovské) rozdělení

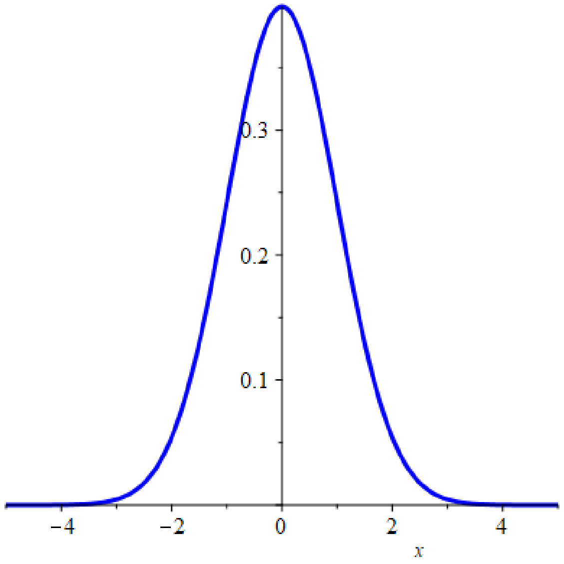
Zobrazení pro $\mu = 0$, $\sigma = 1$

$F(x)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{-x^2}{2\sigma^2}}$$



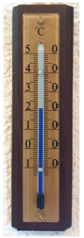
Distribuční funkce



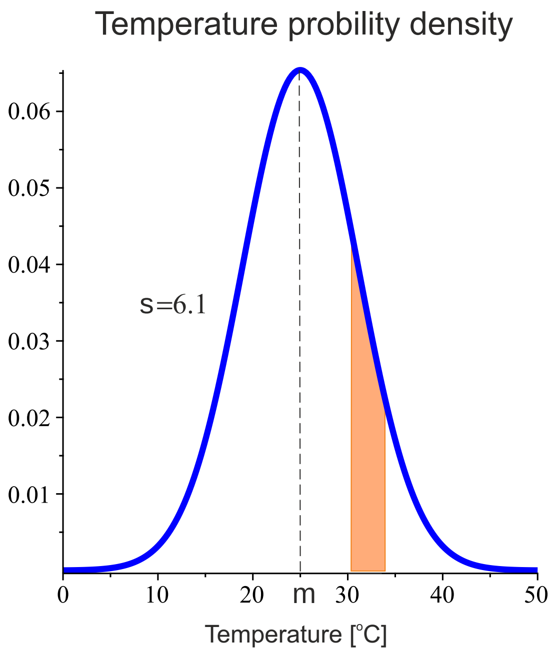
Hustota pravděpodobnosti

Příklad: rozdíl mezi pravděpodobností a hustotou pravděpodobnosti

Předpokládejme teploměr s gaussovským rozložením chyb, $\mu = 25°$C and $\sigma = 6.1°$C.



- Otázka 1: Jaká je pravděpodobnost, že změřená teplota je přesně 31.5°C?
- Odpověď 1: Tato pravděpodobnost je v limitě nula.
- Otázka 2: Jaká je pravděpodobnost, že měřená teplota je v intervalu mezi 30°C and 31°C?
- Odpověď 2: Tato pravděpodobnost je dána plochou pod hustotou pravděpodobnosti (také funkce hustoty pravděpodobnosti), což je podle výpočtu pro obrázek vpravo zhruba 0,11.



Rozdíl mezi pravděpodobností a věrohodností (1)

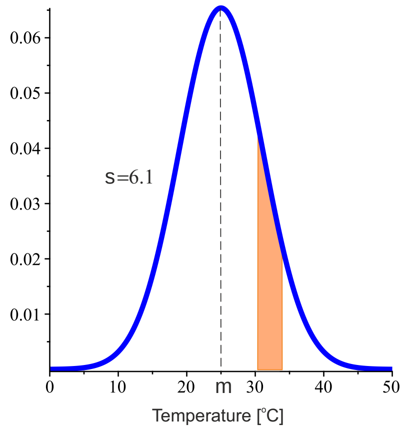
- Ve statistice je věrohodnostní funkce (stručně věrohodnost) pravděpodobnost, že určitý zafixovaný výsledek byl generovaný náhodným rozdělením s určitými, ale neznámými parametry.
- Pravděpodobnost předvídá budoucí výsledek (událost) za předpokladu zafixované hodnoty parametru (hodnot parametrů).
- Uvažujme pravděpodobnostní model s parametry Θ. Model $p(x|\Theta)$ interpretujeme dvojím způsobem a má dvě jména.
  - Pravděpodobnost X při daných parametrech Θ.

    Pravděpodobnost se vypočte jako plocha pod zadanou hustotou pravděpodobnosti.
  - Věrohodnost parametrů Θ za předpokladu, že jsme pozorovali x.

    Věrohodnost je hodnota v ose y pro danou hodnotu x pravděpodobnostního rozdělení. Rozdělení je možné posunout a odečíst příslušnou věrohodnost. V příkladu na obrázku vpravo dole:

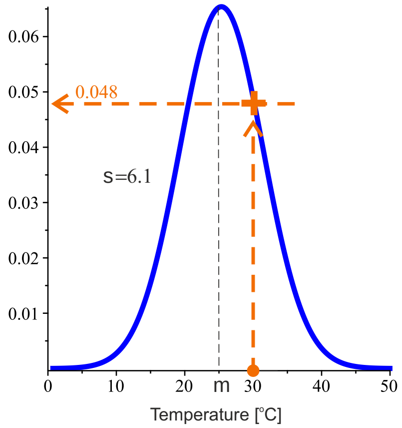
    $L = p(\text{Gaussian}, \mu = 25, \sigma = 6.1|\text{temperature} = 30^\circ\text{C}) = 0.048.$

probability



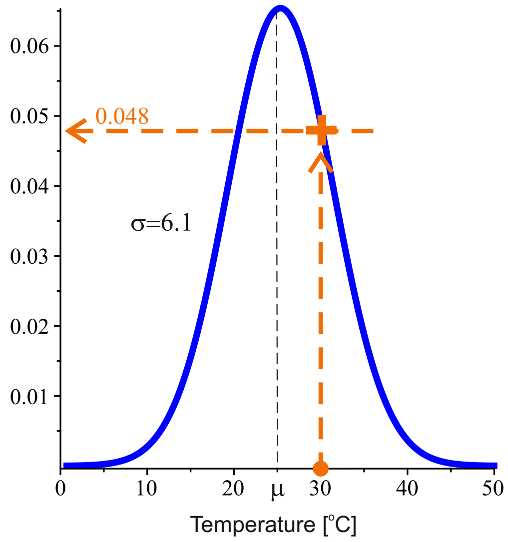
versus

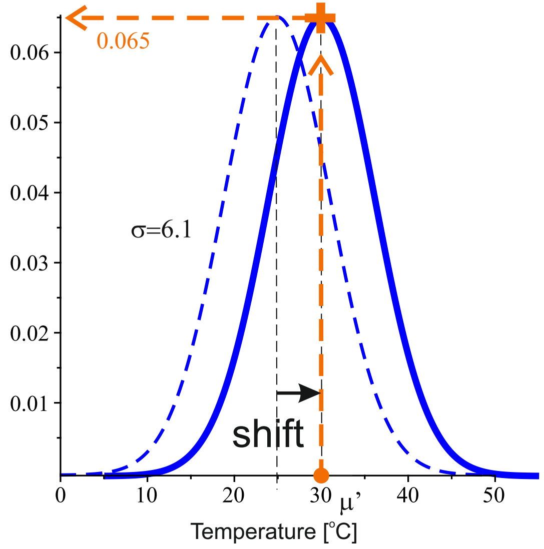
likelihood



Rozdíl mezi pravděpodobností a věrohodností (2)

- Připomínáme z předchozí průsvitky: Předpokládáme gaussovské rozdělení chyb při měření teploty určitým teploměrem s $\mu = 25°\mathrm{C}$ and $\sigma = 6.1°\mathrm{C}$.
- Změřili jsme teplotu $30°\mathrm{C}$. Odpovídají věrohodnost se odhaduje, viz obrázek vlevo, $L = p(\text{Gaussian}, \mu = 25, \sigma = 6.1 \,|\, \text{temperature} = 30°\mathrm{C}) = 0,048.$
- Když posuneme rozdělení, aby $\mu' = 30$, viz obrázek vpravo, nová věrohodnost bude $0,065$. Hodnotu na pravé straně podmíněné pravděpodobnosti jsme zafixovali.





Zákon velkých čísel

Zákon velkých čísel říká, že při velkém počtu nezávislých pokusů je možné téměř jistě očekávat, že relativní četnost se bude blížit teoretické hodnotě pravděpodobnosti (střední hodnota se bude přibližovat k matematickému očekávání).

- Gerolamo Cardano (Italský mathematik, * 1501, † 1576) vyslovil bez důkazu, že přesnost empirických statistik se zlepšuje se zvětšováním počtu pokusů.
- Jakob Bernoulli, Ars Conjectandi: Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis, 1713, Chapter 4.

* 1655 Basel
† 1705 Basel, Švýcarsko

Matematické očekávání

- Relativní četnost náhodného jevu = průměrná hodnota.
- (Matematické) očekávání $E(x)$ = průměrná hodnota při nekonečném počktu pokusů, tj. průměrná hodnota pravděpodobnostního rozdělení.
- Spojitá definice: $E(x) = \mu = \int\limits_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x.$
- Diskrétní definice: $E(x) = \mu = \sum\limits_{x} x\, P(x).$
- Očekávání lze odhadnout z řady vzorků pomocí $E(x) \approx \frac{1}{N} \sum\limits_{i} x_i$. Odhad se stane přesným, když $N \to \infty$ a experimenty jsou statisticky nezávislé.
- Očekávání přes více proměnných: $E_x(x, y) = \int\limits_{-\infty}^{\infty} (x, y)\, f(x)\, \mathrm{d}x$
- Podmíněné očekávání: $E(x|y) = \int\limits_{-\infty}^{\infty} x\, f(x|y)\, \mathrm{d}x.$

Základní charakteristiky náhodné veličiny

Spojité rozdělení	Diskrétní rozdělení

Matematické očekávání

$$E(x) = \mu = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \qquad E(x) = \mu = \sum_x x\, P(x)$$

k-tý obecný moment

$$E(x^k) = \int_{-\infty}^{\infty} x^k\, f(x)\, \mathrm{d}x \qquad E(x^k) = \sum_x x^k\, P(x)$$

k-tý centrální moment

$$\mu_k = \int_{-\infty}^{\infty} (x - E(x))^k\, f(x)\, \mathrm{d}x \qquad \mu_k = \sum_x (x - E(x))^k\, P(x)$$

Rozptyl, disperze, 2. centrální moment

$$D(x) = \int_{-\infty}^{\infty} (x - \mu_x)^2\, f(x)\, \mathrm{d}x \qquad D(x) = \sum_x x^2\, P(x)$$

Standardní odchylka $\sigma(x) = \sqrt{D(x)}$

Centrální limitní věta (1)

Centrální limitní věta poskytuje pravděpodobnostní popis výběrových průměrů. Ty byly vytvořeny z průměrů nekonečného počtu náhodně vybraných vzorků o velikosti N z rodičovské populace. Centrální limitní věta dovoluje předpovědět pravděpodobnostní charakteristiky nezávisle na rozdělení rodičovské populace.

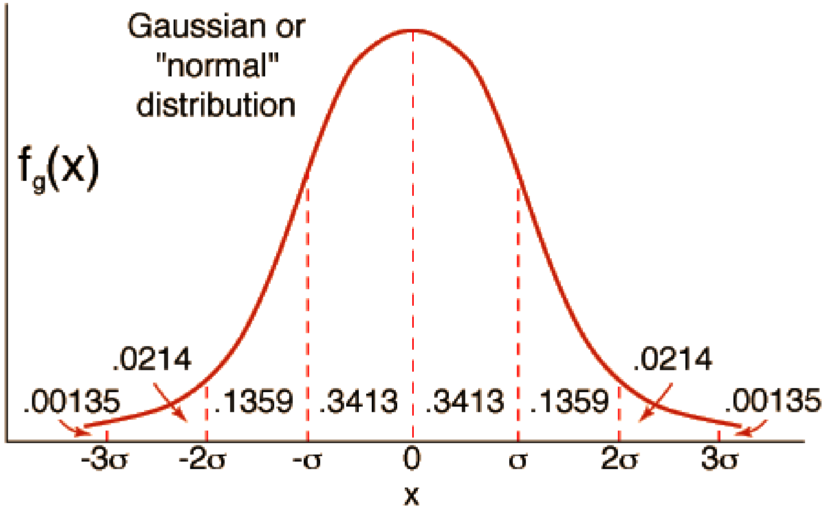
1. Střední hodnota populace výběrových průměrů (tj. která vznikne z výběrových průměrů o N vzorcích náhodně vybíraných z rodičovské pouplace), se rovná střední hodnotě rodičovské populace.

2. Standardní odchylka populace výběrových průměrů se rovná standardní odchylce rodičovské populace dělené druhou odmocninou velikosti vzorků N.

3. Pravděpodobnostní rozdělení výběrových průměrů se bude blížit normálnímu (gausovskému) rozdělení s rostoucí velikostí N vybíraných vzorků.

Centrální limitní věta (2)

- Důsledkem Centrální limitní věty je skutečnost, že po průměrování měření určité veličiny se pravděpodobnostní rozložení těchto průměru bude blížit normálnímu (gaussovskému) rozdělení.
- Uvažujme měřenou veličinu složenou z několika dalších nekorelovaných veličin, které jsou zatíženy šumem různých rozdělení. Pravděpodobnostní rozdělení složené veličiny se bude blížit k normálnímu rozdělení, když bude narůstat počet veličin tvořících složeninu.
- Důsledkem Centrální limitní věty je tudíž i častý výskyt normálního rozložení v úlohách měření.



Centrální limitní věta (3), aplikační pohled

- Pro použití je podstatné, že není potřebné generovat velké množství náhodných výběrů. Stačí pořídit jediný dosti velký náhodný výběr. Díky centrální limitní větě víme, jaké je rozdělení výběrových průměrů. Nemusíme je generovat.
- Co lze považovat za dostatečně velký výběr? Záleží na aplikaci. Ve statistice bývá považováno za dolní hranici 30-50 pozorování. Vzpomeňte na vzorky kolem 1000 respondentů v odhadech volebních výsledku.
- Míra nejistoty hodnoty parametru populace (zde jsme mluvili jen o průměru) se vyjadřuje intervalem spolehlivosti. Viz učebnice statistiky.

Statistický princip filtrace šumu

Uvažujme skoro nejjednodušší statistický model šumu v obraze.

Nechť je každý pixel obrazu zatížen náhodným aditivním šumem:

- statisticky nezávislým,
- s nulovou střední hodnotou μ,
- směrodatnou odchylkou σ.

Mějme i realizací, $i = 1, \ldots n$. Odhad správné hodnoty je

$$\frac{g_1 + \ldots + g_n}{n} + \frac{\nu_1 + \ldots + \nu_n}{n} \,.$$

Výsledkem je náhodná veličina s $\mu' = 0$ a $\sigma' = \sigma/\sqrt{n}$.

Předchozí úvaha má oporu v teorii pravděpodobnosti, a to ve velmi silné a obecné Centrální limitní větě.

Náhodné vektory

- Pojem "náhodný vektor" rozšiřuje pojem "náhodné číslo". Náhodný vektor X je (sloupcový) vektor, který přiřazuje náhodné veličiny $x_1, \ldots, x_n$ výsledkům pokusu, tj. elementárním jevům $\omega \in \Omega$.

- Uvažujme náhodný vektor $\mathsf{X} = (x_1, x_2, \ldots x_n)^\top$. Distribuční funkce a hustota pravděpodobnosti se rozšíří takto
  - sdružená distribuční funkce

$$F_{\mathsf{X}}(x) = P_{\mathsf{X}}\left((X_1 \leq x_1) \cap (X_2 \leq x_2) \cap \ldots \cap (X_n \leq x_n)\right)$$

  - sdružená hustota pravděpodobnosti

$$f_{\mathsf{X}}(x) = \frac{\partial^n F_{\mathsf{X}}(x)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

Jednodušší charakteristiky náhodných vektorů: střední vektor, kovarianční matice

- Mějme na paměti, že náhodný vektor plně charakterizuje jeho sdružená distribuční funkce nebo sdružená hustota pravděpodobnosti.
- Podobně jako u náhodných proměnných je praktické používat jednodušší popisné charakterisitiky náhodných vektorů jako
  - Vektor matematického očekávání (odhadovaný střední hodnotou)

$$E(\mathsf{X}) = (E(x_1), E(x_2), \ldots, E(x_n))^\top = \mu = (\mu_1, \mu_2, \ldots, \mu_n)^\top$$

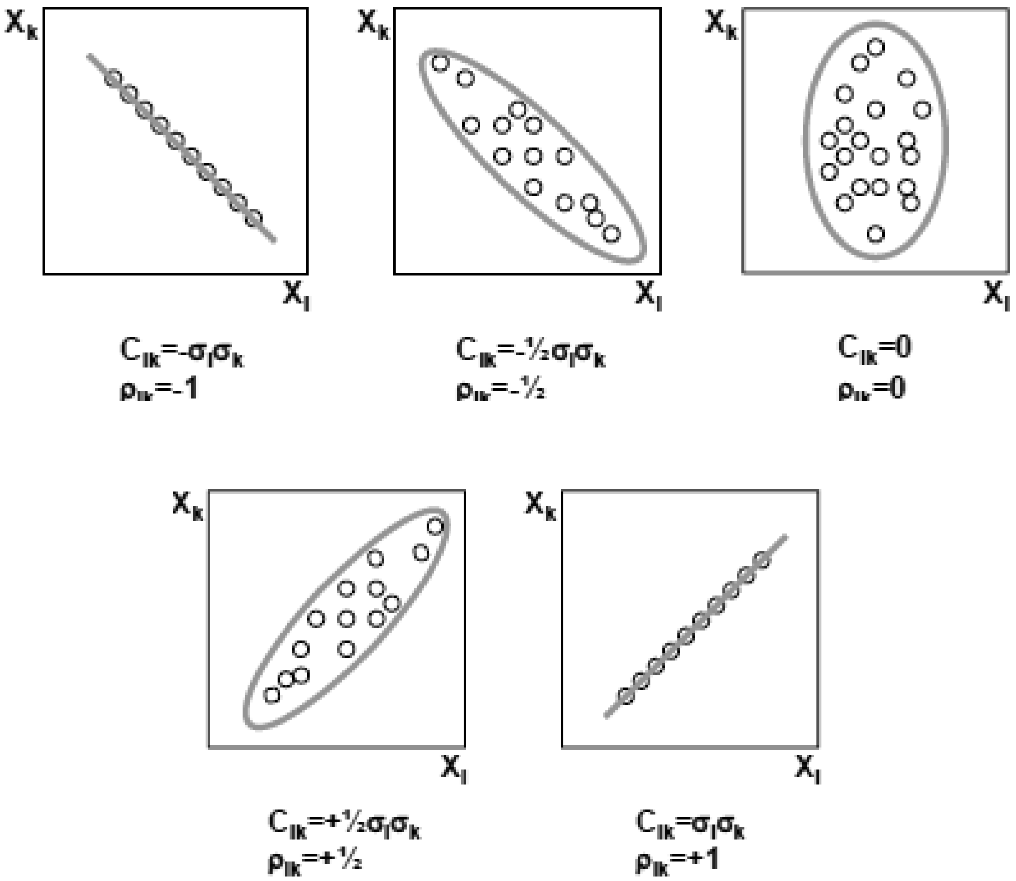
  - Kovarianční matice *(Zobecňuje pojem rozptylu do více dimenzí.)*

$$\Sigma_{\mathsf{X}}(i,k) = \mathrm{cov}(\mathsf{X}) = E((\mathsf{X}-\mu)\,(\mathsf{X}-\mu)^\top) = \begin{bmatrix} \sigma_1^2 & \ldots & c_{1n} \\ \ldots & \ddots & \ldots \\ c_{n1} & \ldots & \sigma_n^2 \end{bmatrix}$$

Kovarianční matice Σ, vlastnosti

- Kovarianční matice naznačuje podobnost, jakou se mění každá možná dvojice prvků náhodného vektoru (jinak, jak se společně mění; angl. co-vary).
- Kovarianční matice má několik důležitých vlastností
  - Kovarianční matice Σ je symetrická (tj. $\Sigma = \Sigma^\top$) a je pozitivně semidefinitní, což znamená, že $x^\star M x \geq 0$ pro všechna $x \in \mathbb{C}$. Značení $x^\star$ znamená komplexně sdružené číslo k číslu x.
  - Když se x_i a x_k obě zvětšují, potom $c_{ik} > 0$.
  - Když x_i klesá a současně x_k roste, potom $c_{ik} < 0$.
  - Když jsou x_i a x_k nekorelované, potom $c_{ik} = 0$.
  - $|c_{ik}| \leq \sigma_i^2$, kde σ_i je standardní odchylka x_i.
  - $c_{ii} = \sigma_i^2 = D(x_i)$.
- Prvky kovarianční matice se mohou vyjádřit jako $c_{ii} = \sigma_i^2$ and $c_{ik} = \rho_{ik}\,\sigma_i\,\sigma_k$. Veličině ρ_{ik} se říká korelační koeficient.

Prvky kovarianční matice, grafická ilustrace



Kvantily, medián

- p-kvantil Q_p: $P(X < Q_p) = p$.
- Medián je p-kvantil pro $p = 0,5$, to je $P(X < Q_p) = 0,5$.

Poznámka: Medián se používá jako náhrada střední hodnoty v robustní statistice.