Podrobný sylabus
EVF 112 - Metody zpracování fyzikálních měření

Ondřej Santolík (http://os.matfyz.cz) 9., 16., 30. listopadu, 7. prosince

1. Nutné minimum z teorie pravděpodobnosti

   - Klasická pravděpodobnost
   - Kolmogorovova axiomatická soustava
   - Podmíněná pravděpodobnost, nezávislé jevy, úplná pravděpodobnost, Bayesův vzorec
   - Náhodná veličina, distribuční funkce, hustota pravděpodobnosti
   - Parametry náhodné veličiny: momenty, střední hodnota, rozptyl, směrodatná odchylka, kvantily, mod
   - Příklady rozdělení: rovnoměrné, normální, binomické, Poissonovo, exponenciální
   - Centrální limitní věta
   - Náhodný vektor, sdružená distribuční funkce
   - Střední hodnota a rozptyl lineární kombinace náhodných veličin, kovariance
   - Pseudonáhodné sekvence, projekční metoda

2. Některé partie z matematické statistiky

   - Základní pojmy
   - Odhady, jejich vychýlenost a rozptyl
   - Odhady střední hodnoty, směrodatné odchylky, mediánu
   - Test shody středních hodnot pro dva výběry, Studentovo rozdělení
   - Lineární korelace
   - Spearmanův korelační koeficient

3. Hledání parametru modelu

   - Druhy experimentálních chyb, šíření chyb
   - Cílová funkce
   - Odhad parametru modelu, jejich směrodatných odchylek, test dobré shody
   - Princip maximální věrohodnosti, metoda χ^2, metoda nejmenších čtverců
   - Prokládání dat jednoduchým lineárním modelem
   - Nelineární modely, Marquardtova metoda
   - Test pomocí simulovaných realizací dat, konfidenční oblasti

4. Spektrální analýza

   - Stacionarita (silná a slabá), ergodicita
   - Konvoluce, odezva lineárního systému

- Fourierova transformace, vlastnosti, normalizace (Parcevalův teorém)
- Fourierova transformace konvoluce a korelace
- Výkonové spektrum (Wiener-Chincinuv teorém)
- Diskrétně vzorkovaná data, Nyquistova kritická frekvence
- Vzorkovací teorém, aliasing
- Diskrétní Fourierova transformace, FFT
- Spektrální analýza, křížová spektra, spektrální matice
- Aplikace váhových oken, překrývání datových segmentů
- Waveletová analýza, aplikace

5. Příklady analýzy dat z umělých družic.

   - Sluneční vítr: Model válcového elektrostatického analyzátoru
   - Aurorální oblast: Vícebodové vlnové měření plazmatických struktur nad polárními zářemi

1 Klasická pravděpodobnost

Klasická pst jevu A:

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

N_A ... počet případů příznivých jevu A, neboli elementární jevy ze kterých se skládá A.

N ... počet všech možných případů, neboli počet prvků prostoru elementárních jevů.

- Arabsky "al zar" kostka. Nejstarší kniha ... 16. stol., Cardano: "De ludo alae"
- 17. stol ... Pascal, Fermat: řešení klas. paradoxů
  - Kolikrát musíme hodit 2 kostkami aby pst, že padne (6,6) byla větší, než 0.5?
  - Dva hráči hrají spravedlivou hru: kdo první vyhraje 6 kol, dostane celou výhru. Hra musí skončit dříve - hráč A vyhrál 5 kol, hráč B 3 kola. Jak rozdělit výhru?

Nevýhoda: zavedena příliš speciálně (diskrétní jevy)

2 Kolmogorovova axiomatická soustava (1929)

- $\Omega\ldots$ Množina elementárních jevů.
- **Jev**: podmnožina $A \subseteq \Omega$ (nemusí platit, že každá podmnožina je jevem).
- $\mathcal{B}\ldots$ Množina všech jevů - systém podmnožin množiny Ω.

Axiómy:

1. $\Omega \in \mathcal{B}$.
2. $A \in \mathcal{B} \Longrightarrow \Omega \backslash A$ (doplněk do Ω) $\in \mathcal{B}$. Odtud $\emptyset \in \mathcal{B}$.
3. $A_1, A_2, \ldots A_n \in \mathcal{B}$ (konečný nebo spočetný systém jevů)
$$\Longrightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{B}.$$
   Odtud
$$\bigcap_{i=1}^{n} A_i \in \mathcal{B}.$$

neboli $\mathcal{B}$ je množinová σ-algebra.

Každý jev $A \in \mathcal{B}$ má nezápornou pst $P(A)$

Axiómy:

4. $P(\Omega) = 1.$
5. $P(\emptyset) = 0.$
6. Je-li $A_i, i = 1 \dots n$ systém po dvou disjunktních jevů (tj. žádné dva nemohou nastat zároveň):

$$A_i \cap A_j = \emptyset, \forall i \neq j, i, j = 1 \dots n$$

pak

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

3 Podmíněná pravděpodobnost

Buď $\{\Omega, \mathcal{B}, P\}$ pstní pole a $B \in \mathcal{B}$ takové, že $P(B) > 0$. Potom na systému jevů

$$C = \{A \cap B\}, A \in \mathcal{B}$$

lze definovat podmíněnou pravděpodobnost (pst jevu A za podmínky B):

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

<u>Vlastnosti:</u>

1. $0 \leq P(A|B) \leq 1$.
2. $A_i \ldots$ po dvou disjunktní jevy $A_i \cap A_j = \emptyset, \forall i \neq j$:

$$\Longrightarrow P(\bigcup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B).$$

3. jev B implikuje jev A, neboli

$$B \subset A \Longrightarrow P(A|B) = 1.$$

4. Matematicky triv. z definice podmíněné psti (ne tak prakticky):

$$P(A \cap B) = P(A|B)\, P(B).$$

Nezávislé jevy

Jevy A a B jsou nezávislé :
$$P(A \cap B) = P(A)\ P(B)$$

Souvislost s podmíněnou pstí:
$$P(A|B) = P(A).$$

Úplná pravděpodobnost

$B_1, \ldots, B_n$ jsou po dvou disjunktní jevy takové, že
$$\Omega = \bigcup_{i=1}^{n} B_i,$$
tj. tvoří úplný soubor jevů ($P(\bigcup B_i) = \sum P(B_i) = 1$) a A je libovolný jev, pak
$$P(A) = \sum_{i=1}^{n} P(B_i)\ P(A|B_i).$$

Důkaz:
$$A = \bigcup_{i=1}^{n} (B_i \cap A)$$

4 Bayesův vzorec

neboli věta o pravděpodobnosti hypotéz (inverzní pravděpodobnosti):

$B_1, \ldots, B_n$ je úplný soubor jevů a A je libovolný jev, pak

$$P(B_k|A) = \frac{P(A|B_k)\ P(B_k)}{\sum_{i=1}^{n} P(A|B_i)\ P(B_i)}$$

kde $P(B_k|A)$ jsou aposteriorní pravděpodobnosti hypotéz B_k, a $P(B_k)$ jsou jejich apriorní pravděpodobnosti.

→ **Oprava pravděpodobností hypotéz B_i nastane-li jev A.**

5 Náhodná veličina

Náhodná veličina ξ je reálná funkce definovaná na množině elementárních jevů Ω (zobrazení $\Omega \to \mathbb{R}$) taková, že množiny všech $\omega \in \Omega$ splňujících

$$\{\omega : \xi(\omega) \leq a\} \in \mathcal{B} \quad \forall a \in \mathbb{R},$$

neboli takové množiny jsou prvky množiny jevů (jsou jevy)

<u>mají tudíž pravděpodobnost</u>.

Distribuční funkce náhodné veličiny

je reálnou funkcí reálné proměnné

$$F_\xi(x) = P(\xi \leq x)$$

neboli je pravděpodobností, že $\xi \leq x$.

Vlastnosti distribuční funkce náhodné veličiny

- neklesající;
- $\lim_{x \to -\infty} F_\xi(x) = 0$;
- $\lim_{x \to \infty} F_\xi(x) = 1$;
- spojitá zprava;
- nejvýše spočetně bodů nespojitosti;

Pro diskrétní náhodné veličiny

$$F_\xi(x) = \sum_{i, x_i \leq x} P(\xi = x_i)$$

Příklad - součet dvou kostek:

- Náhodná veličina ξ = součet dvou kostek.
- $x_i = 2, 3, \ldots, 6, 7, 8, \ldots, 12.$
- $P(\xi = x_i) = \frac{1}{36}, \frac{2}{36}, \ldots, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \ldots, \frac{1}{36}.$
- $F_\xi(x_i) = \frac{1}{36}, \frac{3}{36}, \ldots, \frac{15}{36}, \frac{21}{36}, \frac{26}{36}, \ldots, \frac{36}{36}.$

Hustota pravděpodobnosti

Je-li ξ spojitá náhodná veličina, pak

$$F_\xi(x) = \int\limits_{-\infty}^{x} f(u)du,$$

kde $f(u)$ je hustota pravděpodobnosti distribuční funkce F_ξ.

6 Parametry náhodné veličiny

Střední hodnota náhodné veličiny

Též „očekávaná hodnota“ neb „matematická naděje“.

- Spojitá náhodná veličina:
$$\langle \xi \rangle = E(\xi) = \int_{-\infty}^{\infty} u\ f(u)\ du.$$
obecně Lesbequeův-Stieltjesův integrál s mírou danou pravděpodobností $\longrightarrow$
- Diskrétní náhodná veličina:
$$\langle \xi \rangle = E(\xi) = \sum_i x_i P(\xi = x_i).$$

Střední hodnota funkce náhodné veličiny

Spojitá n.v. : $E\Big(\phi(\xi)\Big) = \int_{-\infty}^{\infty} \phi(u)\ f(u)\ du.$ Diskrétní n.v. : $E\Big(\phi(\xi)\Big) = \sum_i \phi(x_i)\ P(\xi = x_i).$

Momenty

$$\phi(\xi) = \xi^k,$$

Neboli ϕ je mocninná funkce: k-tý moment

$$m_k(\xi) = E(\xi^k) = \int_{-\infty}^{\infty} u^k \; f(u) \; du.$$

Pozn.: $m_1(\xi) = E(\xi)$ je střední hodnota.

Centrální momenty

$$M_k(\xi) = E\Big([\xi - E(\xi)]^k\Big) = \int_{-\infty}^{\infty} [u - E(\xi)]^k \; f(u) \; du.$$

Rozptyl

$$\mathrm{var}(\xi) = D(\xi) = \sigma^2(\xi) = M_2(\xi) = \int_{-\infty}^{\infty} [u - E(\xi)]^2 \; f(u)\, du; \qquad \Big(= \sum [x_i - E(\xi)]^2 \, P(\xi = x_i) \text{ pro diskrétní n.v.}\Big)$$

neboli rozptyl je 2. centrální moment. σ je směrodatná odchylka (standard deviation).

Kvantily

p-kvantil náhodné veličiny ξ (neboli jejího rozdělení) je číslo x_p takové, že

$$F(x_p) = p$$

- $x_{0.5}$ ... medián;
- $x_{0.25(0.75)}$ ... dolní (horní) kvartil;
- $x_{\ell/10}$ ... ℓ-tý decil;
- $x_{\ell/100}$ ... ℓ-tý percentil;

Mod

Mod je číslo x_m takové, že hustota pravděpodobnosti $f(x_m)$ má maximum:

$$f'(x_m) = 0,\ \ f''(x_m) < 0.$$

$\longrightarrow$ Lokální maximum $\longrightarrow$ modů může být více (příklad: bimodální rozdělení).

7 Příklady diskrétních rozdělení

Binomické rozdělení

Diskrétní náhodná veličina (ξ): počet výskytů nějakého jevu v n nezávislých pokusech, je-li pravděpodobnost výskytu tohoto jevu v každém jednotlivém pokuse rovna π.

Pravděpodobnost pro $\xi = x$, kde $x = 0 \dots n$:

$$P(\xi = x) \quad = \quad \binom{n}{x} \pi^x \ (1-\pi)^{n-x}.$$

- Střední hodnota: $E(\xi) = n\pi$;
- Rozptyl: $\mathrm{var}(\xi) = n\pi(1-\pi)$.

Poissonovo rozdělení

Limita binomického rozdělení pro velký počet *nezávislých* pokusů, klesá-li pravděpodobnost výskytu daného jevu v každém jednotlivém pokuse jako λ/n ($n \to \infty$, $n\pi \to \lambda$).
Pravděpodobnost pro $\xi = x$, kde $x = 0, 1, \ldots$:

$$P(\xi = x) \quad = \quad \exp(-\lambda)\frac{\lambda^x}{x!}$$

- Střední hodnota: $E(\xi) = \lambda$;
- Rozptyl: $\mathrm{var}(\xi) = \lambda \;\longrightarrow\; \sigma(\xi) = \sqrt{\lambda}$;
- Relativní chyba: $\sigma(\xi)/E(\xi) = 1/\sqrt{\lambda}$;
- Vlastnost: náhodné veličiny $\xi_1, \xi_2 \ldots \xi_k$ mají Poissonova rozdělení s $\lambda_1, \lambda_2 \ldots \lambda_k$.
  Potom $\sum \xi_i$ má Poissonovo rozdělení s $\sum \lambda_i$.

Praktické příklady náhodných veličin s Poissonovým rozdělením:

- Počet náhodných čísel padnoucích do zadaného podintervalu ($\to$ počet překlepů v textu).
- Počet nezávislých událostí, které nastanou v zadaném časovém intervalu τ: $\lambda = I\tau$ ($\to$ detekce částic, rozpad částic, telefonní hovory za den na dané číslo, zboží určitého druhu prodané za den).
- Počet bodů ve čtverci (nebo jiném ohraničeném útvaru) náhodně umisťovaném na plochu na níž jsou body předem náhodně rozmístěny ($\to$ malé částice či krvinky v zorném poli mikroskopu).

8 Příklady spojitých rozdělení

Rovnoměrné rozdělení

Konstantní hustota pravděpodobnosti na intervalu $\langle a, b\rangle$:

$$f(u) = 0 \;\text{ pro }\; (x < a) \vee (x > b); \qquad f(u) = \frac{1}{b-a} \;\text{ pro }\; (x \geq a) \wedge (x \leq b).$$

Distribuční funkce:

$$F(x) = 0 \;\text{ pro }\; x \leq a \qquad F(x) = 1 \;\text{ pro }\; x \geq b; \qquad F(x) = \frac{x-a}{b-a} \;\text{ pro }\; (x \geq a) \wedge (x \leq b).$$

- Střední hodnota:

$$E(\xi) = \int_a^b \frac{u}{b-a}\, du = \frac{1}{2}(a+b).$$

- Rozptyl:

$$\mathrm{var}(\xi) = \int_a^b \left(u - \frac{1}{2}(a+b)\right)^2 \frac{1}{b-a}\, du = \frac{1}{12}(b-a)^2.$$

Normální rozdělení

Hustota pravděpodobnosti:

$$f(u) = n_{\mu,\sigma}(u) = \frac{1}{\sigma\sqrt{2\pi}} \ \exp\left(-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right)$$

- Střední hodnota: $E(\xi) = \mu$
- Rozptyl: $\mathrm{var}(\xi) = \sigma^2$

Distribuční funkce:

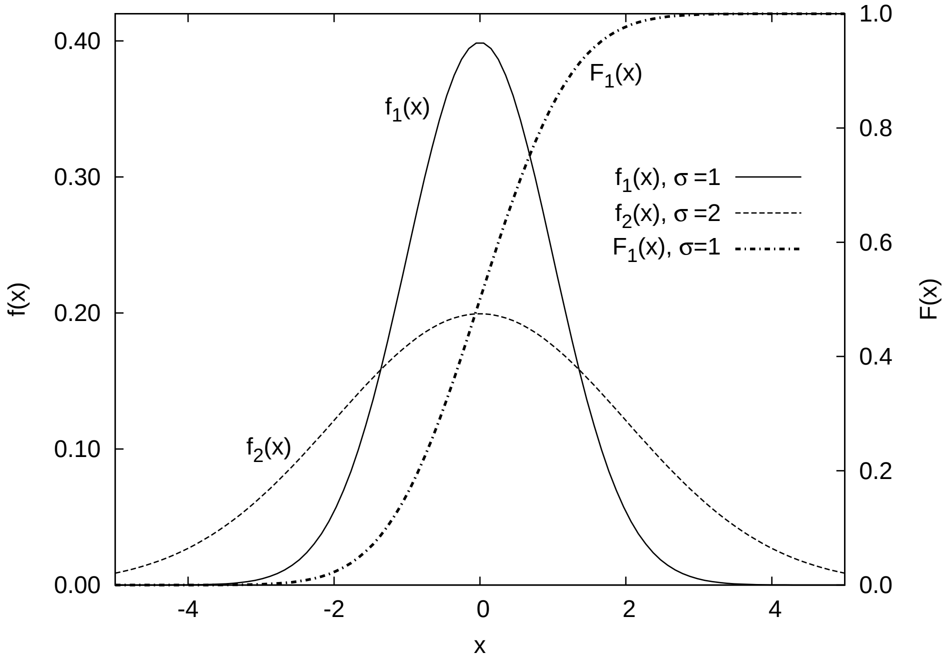
$$F(x) = N_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-\mu}{\sigma}} \exp\left(-\frac{t^2}{2}\right) \ dt,$$

kde $\mathrm{erfc}(y)$ je funkce chyb a $\Phi(y)$ je Laplaceova funkce:

$$\mathrm{erfc}(y) = \frac{2}{\sqrt{\pi}} \int\limits_0^y \exp\left(-t^2\right) \ dt; \qquad \Phi(y) = \sqrt{\frac{2}{\pi}} \int\limits_0^y \exp\left(-\frac{t^2}{2}\right)$$

$$\Phi(a) = P(-a\sigma < \xi - \mu < a\sigma) :$$

$$\Phi(0.67) = 0.5, \quad \Phi(1) = 0.68, \quad \Phi(1) = 0.68, \quad \Phi(2) = 0.95, \quad \Phi(3) = 0.997$$


0.40
0.30
0.20
0.10
0.00
f(x)
-4
-2
0
2
4
x
1.0
0.8
0.6
0.4
0.2
0.0
F(x)
f1(x)
F1(x)
f2(x)
f1(x), σ =1
f2(x), σ =2
F1(x), σ=1

Exponenciální rozdělení

Hustota pravděpodobnosti:

$$f(u) = \frac{1}{\delta} \exp\left(-\frac{u}{\delta}\right) \quad \text{pro} \quad (\delta > 0,\ x \geq 0).$$

Distribuční funkce:

$$F(x) = 1 \ - \ \exp\left(-\frac{x}{\delta}\right).$$

- Střední hodnota: $E(\xi) = \delta$.
- Rozptyl: $\mathrm{var}(\xi) = \delta^2$.
- Vlastnost: rozdělení bez paměti $\quad P(\xi > a + x \mid \xi > a) = P(\xi > x)$.
- Příklad: délky časových intervalů mezi nezávislými událostmi.

Rozdělení χ^2

Rozdělení veličiny

$$\chi^2 = \sum_{i=1}^{\nu} \frac{(\xi_i - E(\xi_i))^2}{\mathrm{var}(\xi_i)},$$

kde každá ξ_i je normálně rozdělená.

Hustota pravděpodobnosti:

$$f(u) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\ u^{\nu/2-1}\ \exp\left(-\frac{u}{2}\right),$$

kde ν je počet stupňů volnosti a $\Gamma(z) = \int\limits_0^\infty t^{z-1} \exp(-t) dt$.

- Střední hodnota: $E(\xi) = \nu$.
- Rozptyl: $\mathrm{var}(\xi) = 2\nu$.
- Vlastnost: pro velká ν konverguje k normálnímu rozdělení.

9 Centrální limitní věta

Lindbergova-Lévyho formulace

$\{\xi_n\}$, $n \in \mathbb{N}$ je posloupnost vzájemně **nezávislých** náhodných veličin s **tímtéž** rozdělením,

se střední hodnotou μ a s konečným rozptylem $\sigma^2 < \infty$.

Potom, pro $x \in \mathbb{R}$

$$\lim_{n\to\infty} P\left\{ \frac{1}{\sigma\sqrt{n}} \left(\sum_{1}^{n} \xi_i - n\mu \right) \leq x \right\} = N_{0,1}(x),$$

t.j. k distribuční fci normálního rozdělení se

- střední hodnotou $\mu = 0$
- a rozptylem: $\sigma^2 = 1$

Ljapunovova formulace centrální limitní věty

Při podmínce

$$\lim_{n\to\infty} \frac{\sqrt[3]{\sum E\left(|\xi_i - \mu_i|^3\right)}}{\sqrt{\sum \sigma_i^2}}$$

pro 3. absolutní centrální moment nemusejí být rozdělení náhodných veličin ξ_i všechna stejná a k $N_{0,1}$ konverguje pro $n \to \infty$ distribuce náhodné proměnné

$$\frac{\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

Aplikace centrální limitní věty

Součet většího počtu náhodných veličin ξ_i rovnoměrně rozdělených na intervalu $< 0, 1 >$ aproximuje normální rozdělení.

Pro každé

- střední hodnota $\mu = 0.5$
- rozptyl: $\sigma^2 = \frac{1}{12}$

Potom, z Lindbergovy-Lévyho formulace centrální limitní věty, náhodná veličina

$$\sqrt{\frac{12}{n}} \left(\sum_1^n \xi_i - \frac{n}{2} \right)$$

konverguje pro $n \to \infty$ k náhodné veličině s rozdělením $N_{0,1}$.

Tvrdí se občas, že stačí $n = 6$:

$$\sqrt{2} \left(\sum_1^6 \xi_i - 3 \right)$$

10 Náhodný vektor

$\vec{\xi} = [\xi_1, \xi_2, \ldots, \xi_n]$ ... složen z náhodných veličin $\xi_1, \xi_2, \ldots, \xi_n$.

Sdružená distribuční funkce

$$F(x_1, x_2, \ldots, x_n) = P\{(\xi_1 \leq x_1) \wedge (\xi_2 \leq x_2) \wedge \ldots \wedge (\xi_n \leq x_n)\}$$

Sdružená hustota pravděpodobnosti

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n$$

Střední hodnota funkce náhodného vektoru

$$E[h(\xi_1, \xi_2, \ldots, \xi_n)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h(u_1, u_2, \ldots, u_n) f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n$$

Střední hodnota lineární kombinace náhodných veličin

$$h(\xi_1, \xi_2, \ldots, \xi_n) = \sum_1^n a_i \xi_i + b$$

$$\begin{aligned}
E\left[\sum_1^n a_i \xi_i + b\right] &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \left[\sum_1^n a_i u_i + b\right] f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n \\
&= \sum_1^n\ a_i\ \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u_i\ f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n \\
&\quad +b\ \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n = \\
&= \sum_1^n\ a_i\ E(\xi_i) + b
\end{aligned}$$

Rozptyl lineární kombinace náhodných veličin

$$
\begin{aligned}
D\left[\sum_{i=1}^{n} a_i\xi_i + b\right] &= E\left\{\left[\sum_{i=1}^{n} a_i\xi_i + b - E\left(\sum_{i=1}^{n} a_i\xi_i + b\right)\right]^2\right\} \\
&= E\left\{\left[\sum_{i=1}^{n} a_i\xi_i - \sum_{i=1}^{n} a_i E(\xi_i)\right]^2\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{cov}(\xi_i, \xi_j)
\end{aligned}
$$

Kovariance

$$\mathrm{cov}(\xi_i, \xi_j) = E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\}$$

platí, že kovariance náhodné proměnné samé se sebou je jejím rozptylem:

$$\mathrm{cov}(\xi_i, \xi_i) = E\left\{[\xi_i - E(\xi_i)]^2\right\} = D(\xi_i)$$

Nezávislé náhodné veličiny

Sdružená distribuční funkce:

$$\begin{aligned}F(x_1, x_2, \ldots, x_n) &= P\{(\xi_1 \leq x_1) \wedge (\xi_2 \leq x_2) \wedge \ldots \wedge (\xi_n \leq x_n)\} \\ &= P(\xi_1 \leq x_1) \cdot P(\xi_2 \leq x_2) \cdot \ldots \cdot P(\xi_n \leq x_n) \\ &= F(x_1) \cdot F(x_2) \cdot \ldots \cdot F(x_n)\end{aligned}$$

Podobně pro sdruženou hustotu pravděpodobnosti:

$$f(x_1, x_2, \ldots, x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$$

Střední hodnota součinu nezávislých náhodných veličin

$$\begin{aligned}E(\xi_1 \cdot \xi_2 \cdot \ldots \cdot \xi_n) &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u_1 \cdot u_2 \cdot \ldots \cdot u_n \; f(u_1, u_2, \ldots, u_n) \; du_1 \; du_2 \; \ldots \; du_n \\ &= E(\xi_1) \cdot E(\xi_2) \cdot \ldots \cdot E(\xi_n)\end{aligned}$$

Kovariance nezávislých náhodných veličin

$$\operatorname{cov}(\xi_i, \xi_j) = E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\} = E[\xi_i - E(\xi_i)] \cdot E[\xi_j - E(\xi_j)] = 0$$

Rozptyl lineární kombinace nezávislých náhodných veličin

$$D\left[\sum_{i=1}^{n} a_i \xi_i + b\right] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \operatorname{cov}(\xi_i, \xi_j) = \sum_{i=1}^{n} a_i^2 D(\xi_i)$$

Kovarianční matice

$$C = \begin{pmatrix} \mathrm{cov}(\xi_1, \xi_1) & \cdots & \mathrm{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(\xi_n, \xi_1) & \cdots & \mathrm{cov}(\xi_n, \xi_n) \end{pmatrix}$$

- symetrická z definice
- na hlavní diagonále $\mathrm{cov}(\xi_i, \xi_i) = D(\xi_i)$
- pro nezávislé veličiny pouze diagonální (zbytek 0)
- pro lineárně závislé veličiny ($\xi_i = a_i \xi$):

$$\mathrm{cov}(a_i\xi, a_j\xi) = E\left\{[a_i\xi - E(a_i\xi)][a_j\xi - E(a_j\xi)]\right\} = a_i a_j D(\xi) = \pm\sqrt{D(\xi_i)D(\xi_j)}$$

Koeficient lineární korelace náhodných veličin ξ_i a ξ_j

$$\rho(\xi_i, \xi_j) = \frac{\mathrm{cov}(\xi_i, \xi_j)}{\sqrt{D(\xi_i)D(\xi_j)}}$$

Pro nezávislé veličiny

$$\rho(\xi_i, \xi_j) = 0$$

Pro lineárně korelované veličiny (závislé s kladným koeficientem):

$$\xi_i = a\xi_j, a > 0 : \rho(\xi_i, \xi_j) = 1$$

Pro lineárně antikorelované veličiny (závislé se záporným koeficientem):

$$\xi_i = a\xi_j, a < 0 : \rho(\xi_i, \xi_j) = -1$$

11 Pseudonáhodné sekvence, projekční metoda

12 Základní pojmy matematické statistiky

- Základní statistický soubor... hypotetický nekonečný soubor dat vzniklých za stejných podmínek; má prvky.
- Výběr... konečná část statistického souboru; má rozsah = počet prvků.
- Znak... vyšetřovaná vlastnost = náhodná proměnná.

  Různé typy znaků (náh. proměnných):

  - Nominální - diskrétní množina bez daného pořadí. Příklad: národnost.
  - Ordinální - diskrétní množina s daným pořadím. Příklad: školní známky.
  - Spojité. Příklad: teplota. Převedení na ordinální znak - rozdělení do tříd. Třídním znakem pa může být střed intervalu.

- Statistika($\widehat{\vartheta}$) ... funkce znaků prvků výběru = náhodná proměnná.

  Z hodnot statistiky $\widehat{\vartheta}$ chceme udělat závěr o příslušném parametru ϑ základního statistického souboru.

  Statistika $\widehat{\vartheta}$ (t.j. náh. proměnná) je odhadem příslušného parametru ϑ základního statistického souboru na základě prvků výběru.

13 Odhady, jejich vychýlenost a rozptyl

- Vychýlenost odhadu (statistiky)

$$b[\widehat{\vartheta}] = E[\widehat{\vartheta}] - \vartheta.$$

  Asymptoticky nevychýlený odhad:

$$\lim_{N\to\infty} b[\widehat{\vartheta}] = 0.$$

- Rozptyl odhadu (statistiky)

$$D[\widehat{\vartheta}] = E\{[\widehat{\vartheta} - E(\widehat{\vartheta})]^2\} = E[(\widehat{\vartheta})^2] - [E(\widehat{\vartheta})]^2.$$

  Většinou závisí na rozsahu (počtu prvků) výběru. Mělo by platit:

$$\lim_{N\to\infty} D[\widehat{\vartheta}] = 0.$$

14 Odhad střední hodnoty

Nechť v základním statistickém souboru má nějaký znak střední hodnotu μ a rozptyl σ^2.

Výběrový aritmetický průměr . . . odhad střední hodnoty

$$\widehat{\mu} = \frac{1}{N}\left(\underbrace{X_1 + X_2 + \cdots + X_N}\right)$$

prvky výběru: náhodné proměnné. N. . . rozsah výběru.

a) Vychýlenost aritmetického průměru

$$E[\widehat{\mu}] = E\left(\frac{\sum_{i=1}^{N} X_i}{N}\right) = \frac{1}{N}\sum_{i=1}^{N} E(X_i) = \frac{1}{N} N\mu = \mu$$

ARITMETICKÝ PRŮMĚR $\widehat{\mu}$ JE NEVYCHÝLENÝM ODHADEM STŘEDNÍ HODNOTY μ.

b) Rozptyl aritmetického průměru

$$
\begin{aligned}
D[\widehat{\mu}] &= E\left\{[\widehat{\mu} - E(\widehat{\mu})]^2\right\} = E\left[\left(\frac{\sum_{i=1}^{N} X_i}{N} - \mu\right)^2\right] \\
&= \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} X_i - \mu\right)^2\right] = \frac{1}{N^2} E\left[\sum_{i=1}^{N} (X_i - \mu)^2\right] + \frac{1}{N^2} E\left[\sum \sum\nolimits_{i \neq j} (X_i - \mu)(X_j - \mu)\right] = \\
&= \frac{1}{N^2} N\sigma^2 + \frac{1}{N^2} \sum \sum\nolimits_{i \neq j} \underbrace{E(X_i - \mu)}_{0} \underbrace{E(X_j - \mu)}_{0} \quad \ldots \quad \text{nezávislé náh. proměnné}
\end{aligned}
$$

$$= \boxed{\frac{1}{N}\sigma^2}$$

15 Odhad rozptylu

16 Odhad mediánu

střední absolutní odchylka