Pravděpodobnost a statistika

Normální rozdělení a centrální limitní věta

Vilém Vychodil

KMI/PRAS, Přednáška 9

Vytvořeno v rámci projektu 2963/2011 FRVŠ

Přednáška 9: Přehled

1. Náhodné výběry a výběrová rozdělení:
   - náhodný výběr, statistika, výběrové rozdělení,
   - lineární kombinace nezávislých veličin a jejich očekávané hodnoty,
   - moment generující funkce, rozdělení součtů náhodných veličin.
2. Normální rozdělení:
   - funkce hustoty normálního rozdělení,
   - normální a standardní normální rozdělení a jejich vztah,
   - střední hodnota, rozptyl,
   - kvantilová funkce, percentily, kritické hodnoty,
   - Box-Mullerova transformace a generování náhodných čísel,
   - rozdělení χ^2, lineární kombinace nezávislých normálních veličin.
3. Limitní věty a aproximace:
   - centrální limitní věta, aproximace pravděpodobností,
   - aproximace binomického a Poissonova rozdělení normálním rozdělením,
   - Čebyševova nerovnost, zákon velkých čísel.

Náhodný výběr

Abstrakce pojmu **výběr** (PŘEDNÁŠKA 1):

- místo (posloupnosti) konkrétních číselných hodnot $x_1, \ldots, x_n$ uvažujme (posloupnost) náhodných veličin $X_1, \ldots, X_n$;
- má smysl uvažovat sdružené rozdělení $P_{X_1,\ldots,X_n}$ (PŘEDNÁŠKA 8).

Definice (Náhodný výběr, angl.: *random sample*)

Mějme pravděpodobnostní prostor $\langle \Omega, \mathcal{F}, P \rangle$ a n nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$ v prostoru $\langle \Omega, \mathcal{F}, P \rangle$, které mají stejné rozdělení pravděpodobnosti, to jest $P(\{X_i \in A\}) = P(\{X_j \in A\})$ pro každé i, j a $A \in \mathcal{B}$. Označme toto rozdělení P_X. Pak náhodný vektor $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ definovaný $\big(\boldsymbol{X}(\omega)\big)(i) = X_i(\omega)$ se nazývá (**n-prvkový**) **náhodný výběr** z rozdělení P_X.

- Náhodný výběr $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ značíme $\boldsymbol{X} = \langle X_1, \ldots, X_n \rangle$, nebo jen $X_1, \ldots, X_n$;
- posloupnost nezávislých náhodných veličin se stejným rozdělením.

Statistiky a výběrová rozdělení

Definice (Statistika a výběrové rozdělení)

Mějme pravděpodobnostní prostor $\langle \Omega, \mathcal{F}, P \rangle$, náhodný výběr $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ a Borelovskou funkci $g : \mathbb{R}^n \to \mathbb{R}$. Pak náhodnou veličinu $\vartheta : \Omega \to \mathbb{R}$ definovanou $\vartheta = g(\boldsymbol{X})$ nazveme (**výběrová**) **statistika** nebo **výběrová charakteristika** (angl.: *sample statistics*) náhodného výběru $\boldsymbol{X}$. Rozdělení pravděpodobnosti $P_\vartheta : \mathcal{B} \to [0, 1]$ veličiny ϑ nazýváme **výběrové rozdělení**, angl.: *sampling distribution*.

Poznámky:

- Z definice složené funkce pro statistiku ϑ máme $\vartheta(\omega) = g(\boldsymbol{X}(\omega)) \in \mathbb{R}$;
- $P_\vartheta(A) = P(\{g(\boldsymbol{X}) \in A\}) = P_{\boldsymbol{X}}(\{g \in A\})$ pro každé $A \in \mathcal{B}$.
- pro konkrétní výběr $x_1, \ldots, x_n$ je $g(x_1, \ldots, x_n)$ konkrétní hodnota;
- pro náhodný výběr $\boldsymbol{X}$ je funkce g náhodná veličina v $\langle \mathbb{R}^n, \mathcal{B}^n, P_{\boldsymbol{X}} \rangle$.

Příklad (Výběr, náhodný výběr a související pojmy)

Náhodný výběr:

- n-prvková posloupnost $X_1, \ldots, X_n$ nezávislých náhodných veličin;
- lze chápat jako náhodný vektor $\boldsymbol{X}\colon \Omega \to \mathbb{R}^n$, kde $\boldsymbol{X}(\omega)(i) = X_i(\omega)$;
- díky nezávislosti *lze vyjádřit $P_{\boldsymbol{X}}$ z marginálních rozdělení P_{X_i}.*

Výběr:

- n-prvková posloupnost reálných čísel $x_1, \ldots, x_n$;
- vznikla *jedním pozorováním hodnot veličin náhodného výběru.*

Výběrová statistika (charakteristika):

- $\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ (průměr náhodného výběru);
- $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$ (rozptyl náhodného výběru).

Výběrové rozdělení:

- $P_{\overline{X}}, P_{S^2}, \ldots$ (závisí na μ a σ^2 výchozích X_i a na velikosti náhodného výběru n).

Věta (Očekávané hodnoty součinu nezávislých veličin)

Mějme nezávislé náhodné veličiny $X_1, X_2, \ldots, X_n$ *v pravděpodobnostním prostoru* $\langle \Omega, \mathcal{F}, P \rangle$ *a předpokládejme, že jejich očekávané hodnoty* $E(X_i)$ *existují. Potom*

$$E(X_1 \cdot X_2 \cdots X_n) = E(X_1) \cdot E(X_2) \cdots E(X_n).$$

Důkaz.

Pro diskrétní $X_1, \ldots, X_n$ s pravděpodobnostními funkcemi $f_{X_1}, \ldots, f_{X_n}$ můžeme s využitím nezávislosti náhodných veličin a distributivity psát:

$$\begin{aligned} E(X_1 \cdot X_2 \cdots X_n) &= \sum\nolimits_{x_1 \in C_1} \cdots \sum\nolimits_{x_n \in C_n} x_1 \cdots x_n \cdots f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \\ &= \sum\nolimits_{x_1 \in C_1} \cdots \sum\nolimits_{x_n \in C_n} x_1 \cdots x_n \cdots f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \sum\nolimits_{x_1 \in C_1} x_1 \cdot f_{X_1}(x_1) \cdots \sum\nolimits_{x_n \in C_n} x_n \cdot f_{X_n}(x_n). \end{aligned}$$

Pro absolutně spojité náhodné veličiny lze prokázat analogicky. $\square$

Lineární kombinace nezávislých náhodných veličin

Věta (Střední hodnoty a rozptyl lineárních kombinací)

Mějme nezávislé náhodné veličiny $X_1, \ldots, X_n$, *které mají střední hodnoty* $\mu_{X_1}, \ldots, \mu_{X_n}$ *a rozptyly* $\sigma^2_{X_1}, \ldots, \sigma^2_{X_n}$. *Pak pro libovolná* $a_1, \ldots, a_n \in \mathbb{R}$ *platí*

$$\mu_Y = \sum_{i=1}^{n} a_i \cdot \mu_{X_i}, \qquad \sigma^2_Y = \sum_{i=1}^{n} a_i^2 \cdot \sigma^2_{X_i},$$

kde $Y = \sum_{i=1}^{n} a_i \cdot X_i$ *je lineární kombinace* $X_1, \ldots, X_n$.

Pro náhodný výběr $X_1, \ldots, X_n$ z rozdělení se střední hodnotou μ a rozptylem σ^2:

$$\boxed{\mu_{\overline{X}} = \mu \quad \text{a} \quad \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}} \quad \text{pro} \quad \boxed{\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}}.$$

Důležité související otázky:

- *Jaké má* $\overline{X}$ *(průměr náhodného výběru) rozdělení?*
- *Pro jaká* n *je* $\overline{x}$ *(výběrový průměr) dobrým odhadem* μ *(střední hodnota)?*

Důkaz.

Tvrzení prokážeme pomocí linearity E. První tvrzení je zřejmé, protože

$$\mu_Y = E(Y) = E\left(\sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n} E(a_i \cdot X_i) = \sum_{i=1}^{n} a_i \cdot \mu_{X_i}.$$

Hodnotu rozptylu σ_Y^2 veličiny Y vyjádříme podobně:

$$\begin{aligned}\sigma_Y^2 &= E\big((Y - \mu_Y)^2\big) = E\Big(\big(\textstyle\sum_{i=1}^{n}(a_i \cdot X_i) - \sum_{i=1}^{n}(a_i \cdot \mu_{X_i})\big)^2\Big) \\ &= E\Big(\big(\textstyle\sum_{i=1}^{n} a_i(X_i - \mu_{X_i})\big)^2\Big) = E\big(\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i - \mu_{X_i})(X_j - \mu_{X_j})\big).\end{aligned}$$

Užitím faktu, že E je lineární operátor dostáváme

$$\sigma_Y^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \cdot E\big((X_i - \mu_{X_i})(X_j - \mu_{X_j})\big).$$

Pokud ale $i \neq j$, pak z nezávislosti X_i a X_j dostáváme

$$\begin{aligned}E\big((X_i - \mu_{X_i})(X_j - \mu_{X_j})\big) &= E(X_i - \mu_{X_i}) \cdot E(X_j - \mu_{X_j}) \\ &= \big(E(X_i) - \mu_{X_i}\big)\big(E(X_j) - \mu_{X_j}\big) = \big(\mu_{X_i} - \mu_{X_i}\big)\big(\mu_{X_j} - \mu_{X_j}\big) = 0.\end{aligned}$$

Odtud dostáváme $\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \cdot E\big((X_i - \mu_{X_i})^2\big) = \sum_{i=1}^{n} a_i^2 \cdot \sigma_{X_i}^2$. □

Příklady (Lineární kombinace náhodných veličin)

Mějme nezávislé náhodné veličiny X a Y se středními hodnotami $\mu_X = \mu_Y$ a rozptyly $\sigma_X^2 = \sigma_Y^2$. Potom pro $Z = X + Y$ platí

$$\mu_Z = \mu_X - \mu_Y = 0,$$
$$\sigma_Z^2 = \sigma_X^2 + (-1)^2 \cdot \sigma_Y^2 = \sigma_X^2 + \sigma_Y^2 = 2 \cdot \sigma_X^2.$$

Pokud je $X_1, \ldots, X_{10}$ je desetiprvkový náhodný výběr z rozdělení se střední hodnotou $\mu = 5$ a rozptylem $\sigma^2 = 120$, pak pro průměr $\overline{X}$ náhodného výběru:

$$\overline{X} = \frac{X_1 + \cdots + X_{10}}{10} = \sum_{i=1}^{10} \frac{1}{10} \cdot X_i$$

platí

$$\mu_{\overline{X}} = \sum\nolimits_{i=1}^{10} \frac{1}{10} \cdot 5 = 5, \qquad \sigma_Y^2 = \sum\nolimits_{i=1}^{10} \frac{1}{100} \cdot 25 = \frac{250}{100} = 2.5.$$

Generující funkce

Definice (Moment generující funkce)

Mějme náhodnou veličinu X v pravděpodobnostním prostoru $\langle \Omega, \mathcal{F}, P \rangle$. Pak se reálná funkce M_X, která je definovaná v okolí bodu 0 předpisem

$$M_X(t) = E\big(e^{t \cdot X}\big) = \int e^{t \cdot X} \, \mathrm{d}P.$$

nazývá **moment generující funkce** veličiny X, angl.: *moment generating function*.

Významné vlastnosti:

- Lze chápat pro diskrétní i absolutně spojité veličiny;
- tvar funkce obvykle jednodušší než distribuční funkce;
- **jednoznačně určuje rozdělení** pravděpodobnosti veličiny X:

  Pokud mají X a Y stejné M_X a M_Y, pak jsou P_X a P_Y stejné.

  Důkaz vyžaduje pokročilé techniky z matematické analýzy (viz Billingsley).

Vyjádření momentů pomocí generující funkce

Mějme spojitou náhodnou veličinu X s hustotou f_X. Pak M_X je rovna

$$M_X(t) = E\big(e^{t \cdot X}\big) = \int_{-\infty}^{\infty} e^{t \cdot x} \cdot f_X(x)\,\mathrm{d}x\,.$$

Poznámka: Použitím Leibnitzova pravidla pro derivování pod integrálem máme:

$$M'_X(t) = \int_{-\infty}^{\infty} e^{t \cdot x} \cdot x \cdot f_X(x)\,\mathrm{d}x, \qquad M''_X(t) = \int_{-\infty}^{\infty} e^{t \cdot x} \cdot x^2 \cdot f_X(x)\,\mathrm{d}x.$$

Důsledek: Pomocí hodnot $M_X^{(n)}$ v bodě 0 lze vyjádřit $n$tý moment, například:

$$M'_X(0) = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x = E(X), \quad M''_X(0) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\,\mathrm{d}x = E(X^2).$$

Pokud je X diskrétní veličina, kde $P_X(C) = 1$ pro spočetnou $C \subseteq \mathbb{R}$, pak

$$M_X(t) = E\big(e^{t \cdot X}\big) = \sum_{x \in C} e^{t \cdot x} \cdot f_X(x)\,,$$

kde f_X je pravděpodobnostní funkce; rovněž platí $M_X^{(n)}(0) = E(X^n)$.

Věta (Generující funkce lineárních kombinací náhodných veličin)

Mějme nezávislé náhodné veličiny $X_1, \ldots, X_n$*, které mají moment generující funkce* $M_{X_1}, \ldots, M_{X_n}$*. Pak pro libovolná* $a_1, \ldots, a_n \in \mathbb{R}$ *platí, že*

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i \cdot t)$$

je moment generující funkce náhodné veličiny $Y = \sum_{i=1}^{n} a_i \cdot X_i$.

Důkaz.

Rozepsáním definičního vztahu a s využitím vlastností exponenciální funkce máme

$$M_Y(t) = E\big(e^{tY}\big) = E\big(\exp\big(t \cdot \textstyle\sum_{i=1}^{n} a_i X_i\big)\big) = E\big(\textstyle\prod_{i=1}^{n} \exp(t a_i X_i)\big).$$

Jelikož jsou $X_1, \ldots, X_n$ nezávislé, pak jsou i $\exp(t a_1 X_1), \ldots, \exp(t a_n X_n)$ nezávislé:

$$M_Y(t) = E\big(\textstyle\prod_{i=1}^{n} \exp(t a_i X_i)\big) = \textstyle\prod_{i=1}^{n} E(\exp(t a_i X_i)).$$

Užitím $M_{X_i}(t) = E(\exp(t X_i))$ dostáváme $M_{X_i}(a_i \cdot t) = E(\exp(t a_i X_i))$ pro každé $i = 1, \ldots, n$, z čehož už požadované tvrzení okamžitě plyne. $\square$

Důsledky pro náhodné výběry a jejich průměry

Mějme náhodný výběr $X_1, \ldots, X_n$ z rozdělení jehož moment generující funkce je M.

Potom platí:

1. Moment generující funkce $Y = \sum_{i=1}^{n} X_i$ je ve tvaru

$$\boxed{M_Y(t) = \prod_{i=1}^{n} M(t) = \big(M(t)\big)^n}\,.$$

2. Moment generující funkce $\overline{X} = \sum_{i=1}^{n} \frac{1}{n} \cdot X_i$ je ve tvaru

$$\boxed{M_{\overline{X}}(t) = \prod_{i=1}^{n} M\left(\frac{t}{n}\right) = \left(M\left(\frac{t}{n}\right)\right)^n}\,.$$

Související otázka:

- Jak vypadá rozdělení $P_{\overline{X}}$ odpovídající $M_{\overline{X}}$ (pro $n \to \infty$)?

Příklady (Rozdělení součtů nezávislých náhodných veličin)

Mějme nezávislé náhodné veličiny $X_1, \ldots, X_n$, které mají Bernoulliho (alternativní) rozdělení s parametrem p. Potom moment generující funkce každé z X_i je ve tvaru $M(t) = (1-p) + pe^t$. Užitím předchozí věty pro $Y = \sum_{i=1}^n X_i$ platí,

$$M_Y(t) = M(t)^n = \big((1-p) + pe^t\big)^n,$$

což je moment generující funkce binomického rozdělení $b(n,p)$.

Pokud jsou $X_1, \ldots, X_n$ nezávislé a mají Poissonovo rozdělení s parametrem $\lambda = 1$, pak je $M(t) = \exp(e^t - 1)$. Pro $Y = \sum_{i=1}^n X_i$ máme

$$M_Y(t) = M(t)^n = \big(\exp(e^t - 1)\big)^n = \exp\big(n(e^t - 1)\big),$$

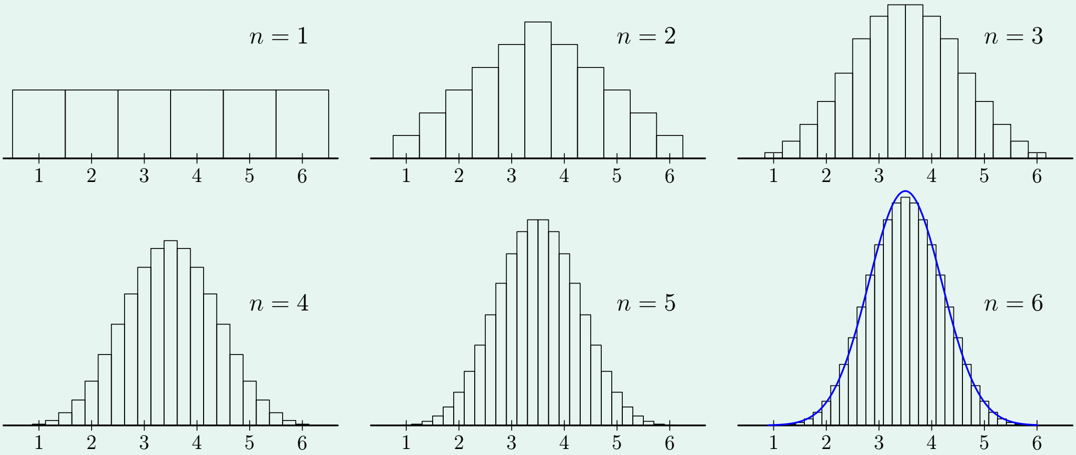
to jest Y má Poissonovo rozdělení s parametrem $\lambda = n$.

Důsledky:

- X s rozdělením $b(n,p)$ je součtem n nezávislých Bernoulliho veličin;
- X s Poissonovým rozdělením s $\lambda \in \mathbb{N}$ je součtem n nezávislých veličin s Poissonovým rozdělením s parametrem 1.

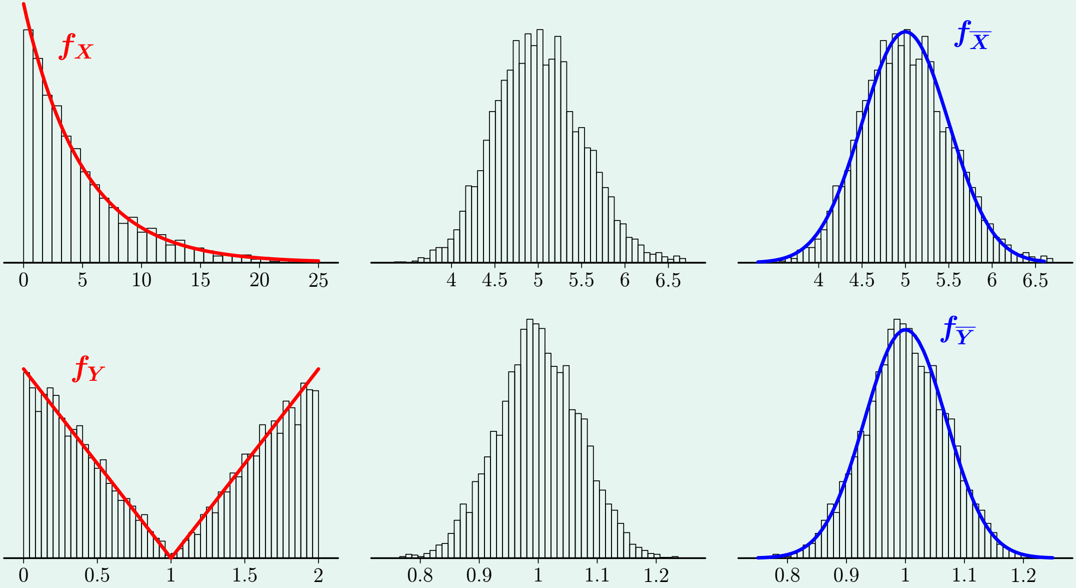
Příklad (Motivace: rozdělení průměru náhodného výběru)

Hodíme n šestistěnných kostek a spočítáme průměrnou hodnotu teček, které padnou na kostkách (pozorujeme tak náhodný výběr $X_1, \ldots, X_n$ a jeho průměr $\overline{X}$).



Otázka: *Jaké má $\overline{X}$ rozdělení?*

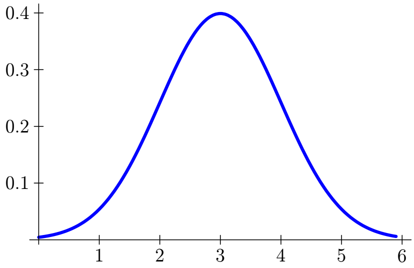
Příklad (Motivace: simulace náhodných veličin a jejich průměrů)



Normální rozdělení: význam a historické pozadí

Rozdělení pravděpodobnosti absolutně spojitých náhodných veličin.

Funkce hustoty je symetrická a ve tvaru „zvonu“:



Poznámky:

- významné pro součty identicky distribuovaných a nezávislých náhodných veličin,
- limitní věty (*centrální limitní věta*);
- Abraham de Moivre, Carl Friedrich Gauss, Marquis Pierre Simon de Laplace, ...

Věta

Mějme $a \in \mathbb{R}$ *a* $b \in (0, \infty)$. *Pak zobrazení* $f\colon \mathbb{R} \to \mathbb{R}$ *dané předpisem*

$$f(x) = \frac{1}{b \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right), \qquad \textit{pro každé } x \in \mathbb{R},$$

je funkce hustoty.

Důkaz (začátek).

Z definice f plyne, že $f(x) > 0$ pro každé $x \in \mathbb{R}$. Dále f je zjevně spojitá funkce, což znamená, že je Borelovská. Stačí tedy ověřit, že $\int f \, \mathrm{d}m = 1$.

To znamená ukázat, že I definované

$$I = \int_{-\infty}^{\infty} f(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{b \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right) \mathrm{d}x,$$

nabývá hodnoty 1.

Důkaz (pokračování).

Substitucí $g(x) = x \cdot b + a$ získáváme

$$I = \int_{-\infty}^{\infty} f(g(x)) \cdot g'(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

Dále máme $I > 0$. To znamená, že pro prokázání $I = 1$ stačí dokázat $I^2 = 1$. Máme

$$I^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2 = \frac{1}{2 \cdot \pi} \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2.$$

Z poslední rovnosti vyjádříme I^2 pomocí dvojného integrálu:

$$\begin{aligned} I^2 &= \frac{1}{2 \cdot \pi} \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right) \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,\mathrm{d}y\right) \\ &= \frac{1}{2 \cdot \pi} \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cdot e^{-\frac{y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2 \cdot \pi} \cdot \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2 + y^2}{2}\right)\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Důkaz (dokončení).

Transformací dvojného integrálu

$$I^2 = \frac{1}{2 \cdot \pi} \cdot \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2 + y^2}{2}\right) \mathrm{d}x \, \mathrm{d}y$$

do polárních souřadnic pomocí substituce $x = \varrho \cdot \cos\varphi$, $y = \varrho \cdot \sin\varphi$ dostáváme pro transformovanou oblast $T = [0, \infty) \times [0, 2\pi)$:

$$I^2 = \frac{1}{2 \cdot \pi} \cdot \iint_{T} \exp\left(-\frac{(\varrho \cdot \cos\varphi)^2 + (\varrho \cdot \sin\varphi)^2}{2}\right) \cdot \begin{vmatrix} \cos\varphi & -\varrho \sin\varphi \\ \sin\varphi & \varrho \cos\varphi \end{vmatrix} \mathrm{d}\varrho \, \mathrm{d}\varphi.$$

Zjednodušením a využitím faktu, že $-e^{-\frac{x^2}{2}}$ je primitivní funkce k $x \cdot e^{-\frac{x^2}{2}}$ máme

$$I^2 = \frac{1}{2 \cdot \pi} \cdot \int_0^{2 \cdot \pi} \int_0^{\infty} e^{-\frac{\varrho^2}{2}} \cdot \varrho \, \mathrm{d}\varrho \, \mathrm{d}\varphi = \frac{1}{2 \cdot \pi} \cdot \int_0^{2 \cdot \pi} 1 \, \mathrm{d}\varphi = \frac{1}{2 \cdot \pi} \cdot 2 \cdot \pi = 1.$$

To znamená, že $I = 1$, tedy f je funkce hustoty. $\square$

Normální rozdělení pravděpodobnosti

Definice (Normální rozdělení, angl.: *normal distribution*)

Spojitá náhodná veličina X má **normální rozdělení** pravděpodobnosti pokud existují konstanty $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ tak, že její funkce hustoty f_X je ve tvaru

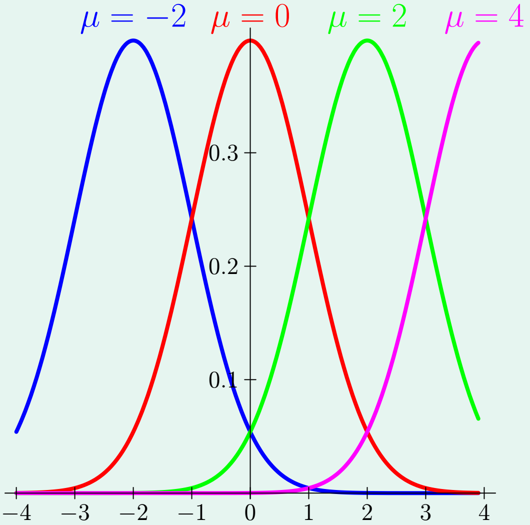
$$f_X(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right), \qquad \text{pro každé } x \in \mathbb{R}.$$

Zkráceně zapisujeme, že X má rozdělení $N(\mu, \sigma^2)$. Pokud má X rozdělení $N(0,1)$ pak říkáme, že X má **standardní normální rozdělení** pravděpodobnosti.

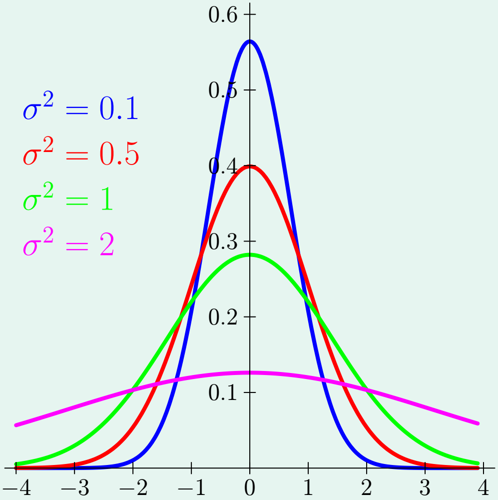
Poznámky:

- Korektnost definice plyne z přechodí věty (f_X je vskutku funkce hustoty);
- označení parametrů μ a σ^2 není náhodné (uvidíme dále);
- standardní normální rozdělení má velký teoretický i praktický význam.

Příklad (Vliv parametrů μ a σ^2 na tvar funkce hustoty)



$\sigma^2 = 1$



$\mu = 0$

Věta (Moment generující funkce normálního rozdělení)

Nechť X *má rozdělení* $N(\mu, \sigma^2)$*. Potom platí:*

$$M_X(t) = \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right).$$

Důkaz (začátek).

Z definice M_X dostáváme:

$$\begin{aligned}
M_X(t) = E\left(e^{t \cdot X}\right) &= \int_{-\infty}^{\infty} e^{t \cdot x} \cdot f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{e^{t \cdot x}}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(t \cdot x - \frac{(x-\mu)^2}{2 \cdot \sigma^2}\right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot \left(x^2 - 2 \cdot x \cdot (\mu + \sigma^2 \cdot t) + \mu^2\right)\right) \mathrm{d}x.
\end{aligned}$$

Důkaz.

V exponentu z předešlého výrazu

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot \left(x^2 - 2 \cdot x \cdot (\mu + \sigma^2 \cdot t) + \mu^2\right)\right) \mathrm{d}x$$

vyjádříme $x^2 - 2 \cdot (\mu + \sigma^2 \cdot t) \cdot x + \mu^2$ jako rozdíl od doplňku na čtverec, to jest:

$$x^2 - 2 \cdot x \cdot (\mu + \sigma^2 \cdot t) + \mu^2 = \left(x - (\mu + \sigma^2 \cdot t)\right)^2 - 2 \cdot \mu \cdot \sigma^2 \cdot t - \sigma^4 \cdot t^2.$$

Odtud dostáváme, že

$$M_X(t) = \exp\left(\frac{2 \cdot \mu \cdot \sigma^2 \cdot t + \sigma^4 \cdot t^2}{2 \cdot \sigma^2}\right) \cdot \int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x - (\mu + \sigma^2 \cdot t))^2}{2 \cdot \sigma^2}\right) \mathrm{d}x.$$

Předchozí integrál je ale ve tvaru $\int_{-\infty}^{\infty} f_Y(y) \,\mathrm{d}y$, kde f_Y je funkce hustoty spojité náhodné veličiny s rozdělením $N(\mu + \sigma^2 \cdot t, \sigma^2)$, tudíž musí být roven 1. Odtud:

$$M_X(t) = \exp\left(\frac{2 \cdot \mu \cdot \sigma^2 \cdot t + \sigma^4 \cdot t^2}{2 \cdot \sigma^2}\right) = \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right).$$

□

Střední hodnota a rozptyl

Mějme spojitou náhodnou veličinu X s rozdělením $N(\mu, \sigma^2)$. Potom:

Derivace moment generující funkce:

$$M'_X(t) = (\mu + \sigma^2 \cdot t) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right), \qquad M'_X(0) = \mu,$$

$$M''_X(t) = \left((\mu + \sigma^2 \cdot t)^2 + \sigma^2\right) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right), \qquad M''_X(0) = \mu^2 + \sigma^2.$$

Střední hodnota $\boldsymbol{\mu_X}$ a rozptyl $\boldsymbol{\sigma_X}$:

$$\mu_X = E(X) = M'_X(0) = \boxed{\mu},$$

$$\sigma_X^2 = E(X^2) - E(X)^2 = M''_X(0) - \left(M'_X(0)\right)^2 = (\mu^2 + \sigma^2) - \mu^2 = \boxed{\sigma^2}.$$

Důsledek: *Normální rozdělení je dáno svou střední hodnotou a rozptylem.*

Standardní normální rozdělení a distribuční funkce

Pokud má náhodná veličina X rozdělení $N(0,1)$, potom je její funkce hustoty

$$f_X(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right) = \boxed{\frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}}\,.$$

Distribuční funkce X se obvykle značí Φ. To jest:

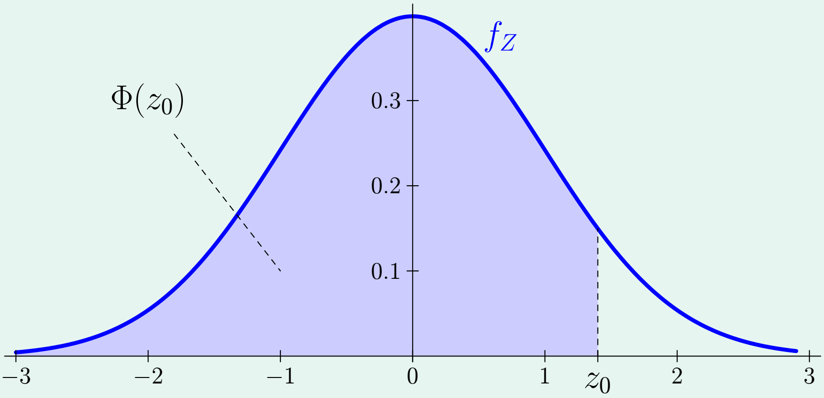
$$\boxed{\Phi(a) = P(\{X \leq a\}) = \int_{-\infty}^{a} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}\, \mathrm{d}x}\,.$$

Poznámky:

- Φ *není elementární funkce* (její hodnoty se numericky aproximují);
- $\Phi(a) = \frac{1}{2} \cdot \left(\frac{2}{\sqrt{\pi}} \cdot \int_{0}^{a} e^{-x^2}\, \mathrm{d}x\right) = \frac{1}{2} \cdot \left(1 + \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)\right)$;
- erf je **chybová funkce**; ve standardu C99 je funkce `erfl` (viz `math.h`).

Příklad (Vlastnosti distribuční funkce Φ)

Mějme náhodnou veličinu Z s rozdělením $N(0,1)$ a hustotou f_Z.



Základní vlastnosti Φ:

- $\Phi(-z) = 1 - \Phi(z)$, $\Phi(z) = 1 - \Phi(-z)$;
- $P(\{Z \leq z\}) = \Phi(z) = 1 - \Phi(-z) = 1 - P(\{Z \leq -z\}) = P(\{Z > -z\})$.

Kvantilová funkce standardního normálního rozdělení

Podle definice kvantilové funkce (Přednáška 7) dostáváme, že

$$\Phi^{-}(p) = \inf\{z \in \mathbb{R} \,|\, p \leq \Phi(z)\}\,.$$

Protože je Φ absolutně spojitá a rostoucí, existuje k ní inverzní funkce Φ^{-1}.
V důsledku máme, že $\Phi^{-}(p) = \Phi^{-1}(p)$ pro každé $p \in (0,1)$. Odtud platí

$$p = \Phi\bigl(\Phi^{-1}(p)\bigr) = \int_{-\infty}^{\Phi^{-1}(p)} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{z^2}{2}}\,\mathrm{d}z\,.$$

Poznámky:

- Φ^{-1} (funkce **probit**) *není elementární funkce* (numericky se aproximuje);
- $\Phi^{-1}(p) = \sqrt{2} \cdot \mathrm{erf}^{-1}(2p - 1)$, kde erf^{-1} je **inverzní chybová funkce**;
- omezená možnost generovat (pseudo)náhodná čísla pomocí metod založených na inverzní transformaci (Přednáška 8).

Věta (O vztahu $N(\mu, \sigma^2)$ a standardního normálního rozdělení)

Pokud má X *rozdělení* $N(\mu, \sigma^2)$, *pak má* $Z = \dfrac{(X - \mu)}{\sigma}$ *rozdělení* $N(0, 1)$.

Důkaz.

Pro distribuční funkci F_Z náhodné veličiny Z platí

$$F_Z(z) = P(\{Z \leq z\}) = P\left(\left\{\frac{X - \mu}{\sigma} \leq z\right\}\right) = P(\{X \leq z \cdot \sigma + \mu\})$$

$$= \int_{-\infty}^{z \cdot \sigma + \mu} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x - \mu)^2}{2 \cdot \sigma^2}\right) \mathrm{d}x.$$

Substitucí $w = \dfrac{x - \mu}{\sigma}$ dostáváme:

$$P(\{Z \leq z\}) = \int_{-\infty}^{z} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{w^2}{2}} \,\mathrm{d}w = \Phi(z).$$

$\square$

Vyjádření hodnot pravděpodobností pro $N(\mu, \sigma^2)$

Pro každé X s rozdělením $N(\mu, \sigma^2)$ máme:

$$
\begin{aligned}
P(\{a \leq X \leq b\}) &= P(\{a - \mu \leq X - \mu \leq b - \mu\}) \\
&= P\left(\left\{\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right\}\right).
\end{aligned}
$$

Jelikož $Z = \dfrac{X - \mu}{\sigma}$ má podle předchozí věty rozdělení $N(0, 1)$, platí

$$
\boxed{P(\{a \leq X \leq b\})} = P\left(\left\{\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right\}\right) = \boxed{\Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)}.
$$

Důsledky:

- distribuční funkce Φ standardního normálního rozdělení postačuje pro vyjádření distribuční funkce pro libovolné $N(\mu, \sigma^2)$;
- náhodnou veličinu X s rozdělením $N(\mu, \sigma^2)$ lze vyjádřit jako $X = Z \cdot \sigma + \mu$.

Vyjádření hodnot distribuční a kvantilové funkce pro $N(\mu, \sigma^2)$

Zobecněním předchozího pozorování pro X s rozdělením $N(\mu, \sigma^2)$ dostáváme:

$$\boxed{F_X(x)} = P(\{X \leq x\}) = P\left(\left\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right\}\right) = \boxed{\Phi\left(\frac{x-\mu}{\sigma}\right)}.$$

Protože je kvantilová funkce F_X^{-1} inverzní funkce k distribuční funkci F_X, použitím předchozího vztahu dostáváme:

$$\Phi^{-1}(p) = \Phi^{-1}\big(F_X(F_X^{-1}(p))\big) = \Phi^{-1}\left(\Phi\left(\frac{F_X^{-1}(p)-\mu}{\sigma}\right)\right) = \frac{F_X^{-1}(p)-\mu}{\sigma}.$$

Vyjádřením $F_X^{-1}(p)$ dostáváme:

$$\boxed{F_X^{-1}(p) = \Phi^{-1}(p) \cdot \sigma + \mu}.$$

Interpretace:

$$\boxed{100p\% \text{ percentil z } N(\mu, \sigma^2) = \mu + (\sigma \text{ krát } 100p\% \text{ percentil z } N(0,1))}.$$

Příklad (Pravděpodobnosti a hodnoty kvantilové funkce pro $N(\mu, \sigma^2)$)

Problém: Mějme náhodnou veličinu X s rozdělením $N(3, 16)$.

Úkoly:

- *Stanovte pravděpodobnost, že X nabude hodnoty z intervalu $[4, 8]$.*
- *Stanovte pravděpodobnost, že X nabude hodnoty větší než 10.*
- *Stanovte kvantil X s hladinou $p = 0.98$.*

Řešení:

$$P(\{4 \leq X \leq 8\}) = P\left(\left\{\frac{4-3}{4} \leq \frac{X-3}{4} \leq \frac{8-3}{4}\right\}\right) = \Phi(1.25) - \Phi(0.25)$$
$$\approx 0.8944 - 0.5987 \approx 0.2957,$$

$$P(\{X \geq 10\}) = 1 - P(\{X \leq 10\}) = 1 - \Phi\left(\frac{10-3}{4}\right) = 1 - \Phi(1.75) \approx 0.04.$$

$$F_X^{-1}(0.98) = \Phi^{-1}(0.98) \cdot 4 + 3 \approx 2.0537 \cdot 4 + 3 \approx 11.2148.$$

Empirické pravidlo pro normální rozdělení

Empirické pravidlo (Přednáška 1)

Uvažujme výběr $x_1, \ldots, x_n$ s výběrovým průměrem $\overline{x}$ a výběrovou směrodatnou odchylkou s. Pokud má histogram tvar „zvonu“ pak

- přibližně $68\,\%$ dat z výběru se nachází v intervalu $(\overline{x} - s, \overline{x} + s)$,
- přibližně $95\,\%$ dat z výběru se nachází v intervalu $(\overline{x} - 2s, \overline{x} + 2s)$,
- přibližně $99.7\,\%$ dat z výběru se nachází v intervalu $(\overline{x} - 3s, \overline{x} + 3s)$.

Formálně: Pokud má X rozdělení $N(\mu, \sigma^2)$, pak

$$P(\{\mu_X - k \cdot \sigma_X \leq X \leq \mu_X + k \cdot \sigma_X\}) = P\left(\left\{\frac{-k \cdot \sigma_X}{\sigma_X} \leq \frac{X - \mu_X}{\sigma_X} \leq \frac{k \cdot \sigma_X}{\sigma_X}\right\}\right),$$

což se rovná $\Phi(k) - \Phi(-k)$. Speciálně pro $k = 1, 2, 3$ dostáváme hodnoty pravděpodobností přibližně 0.6827, 0.9545, 0.9973.

Věta (Box-Mullerova transformace).

Pokud jsou U_1 a U_2 dvě nezávislé náhodné veličiny s rozdělením $U(0,1)$, pak

$$Z_1 = \sqrt{-2\ln U_1} \cdot \cos(2\pi \cdot U_2),$$

$$Z_2 = \sqrt{-2\ln U_1} \cdot \sin(2\pi \cdot U_2),$$

jsou nezávislé a jejich rozdělení je $N(0,1)$. □

Algoritmus (generování pseudonáhodných čísel z $N(0,1)$)

```
(defun standard-normal-random ()
  "Generate two independent standard normal random values."
  (let* ((u1 (random 1L0))
         (u2 (random 1L0))
         (z1 (* (sqrt (* -2 (log u1))) (cos (* 2 pi u2))))
         (z2 (* (sqrt (* -2 (log u1))) (sin (* 2 pi u2)))))
    (values z1 z2)))
```

Rozdělení χ^2

Jedná se o *speciální případ rozdělení* Γ pro parametry

- $\theta = 2$ a
- $\alpha = \frac{r}{2}$, kde r je přirozené číslo.

Z hustoty rozdělení Γ odvodíme hustotu rozdělení χ^2:

Definice (Náhodná veličina s rozdělením χ^2)

Spojitá náhodná veličina X s hustotou f_X má **rozdělení χ^2** (**chi-kvadrát**, angl.: *chi-square distribution*) s **r stupni volnosti** (angl.: *degree of freedom*) pokud $r \in \mathbb{N}$ a f_X je ve tvaru

$$f_X(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right) \cdot 2^{\frac{r}{2}}} \cdot x^{\frac{r}{2}-1} \cdot e^{\frac{-x}{2}} \qquad \text{pokud } x \geq 0,$$

a $f_X(x) = 0$ jinak. Tento fakt značíme, že X má rozdělení $\chi^2(r)$.

Vlastnosti rozdělení $\chi^2(r)$

Vlastnosti $\chi^2(r)$ se odvozují jako speciální případy vlastností rozdělení Γ.

Zjednodušení tvaru distribuční funkce: Mějme náhodnou veličinu W s rozdělením $\chi^2(r)$. Pokud je r (počet stupňů volnosti) sudé číslo, pak je $\alpha = \frac{r}{2}$ přirozené číslo, to znamená, že F_W přejde v distribuční funkce Erlangova rozdělení:

$$F_W(w) = 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda \cdot w)^k \cdot e^{-\lambda \cdot w}}{k!} = 1 - \sum_{k=0}^{\frac{r}{2}-1} \frac{\left(\frac{w}{2}\right)^k \cdot e^{-\frac{w}{2}}}{k!}.$$

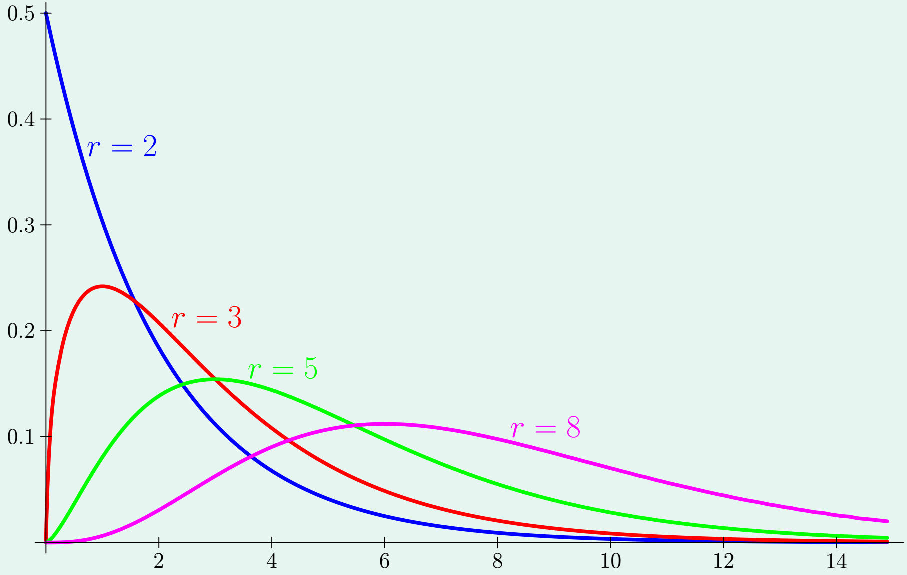
Střední hodnota:

$$\mu_W = \alpha \cdot \theta = \frac{r}{2} \cdot 2 = \boxed{r}\,,$$

Rozptyl:

$$\sigma_W^2 = \alpha \cdot \theta^2 = \frac{r}{2} \cdot 4 = \boxed{2 \cdot r}\,.$$

Příklad (Funkce hustoty rozdělení χ^2 s různými stupni volnosti)

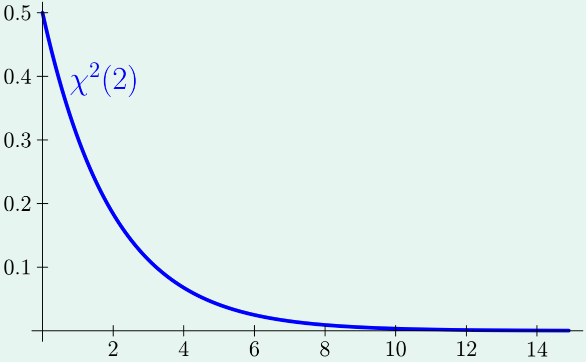


Příklad (Exponenciální rozdělení jako speciální případ χ^2)

Pokud má X exponenciální rozdělení se střední hodnotou 2, pak je f_X ve tvaru

$$f_X(x) = \frac{1}{2} \cdot e^{\frac{-x}{2}} = \frac{x^{\frac{2}{2}-1} \cdot e^{\frac{-x}{2}}}{\Gamma\left(\frac{2}{2}\right) \cdot 2^{\frac{2}{2}}} \qquad \text{pokud } x \geq 0.$$

To znamená, že X má rozdělení $\chi^2(2)$.



Věta (O součtech nezávislých náhodných veličin s rozdělením χ^2)

Mějme nezávislé náhodné veličiny $X_1, \ldots, X_k$*, které mají rozdělení* $\chi^2(r_1), \ldots, \chi^2(r_k)$*. Pak* $Y = X_1 + \cdots + X_k$ *má rozdělení* $\chi^2(r_1 + \cdots + r_k)$*.*

Důkaz.

Tvrzení se prokazuje užitím faktu, že moment generátorová funkce jednoznačně určuje rozdělení. V případě X_i s rozdělením $\chi^2(r_i)$ je M_{X_i} je tvaru

$$M_{X_i}(t) = (1 - 2t)^{-\frac{r_i}{2}} \qquad \text{pro } t < 0.5.$$

Použitím věty o tvaru moment generujících funkcí lineárních kombinací nezávislých náhodných veličin a s využitím vlastností exponenciální funkce dostáváme

$$M_Y(t) = \textstyle\prod_{i=1}^{k} M_{X_i}(t) = \prod_{i=1}^{k}(1 - 2t)^{-\frac{r_i}{2}} = (1 - 2t)^{-(\frac{r_1}{2} + \cdots + \frac{r_k}{2})}.$$

Odtud ale ihned dostáváme, že

$$M_Y(t) = (1 - 2t)^{-(\frac{r_1}{2} + \cdots + \frac{r_k}{2})} = (1 - 2t)^{-(\frac{r_1 + \cdots + r_k}{2})}$$

je moment generující funkce rozdělení $\chi^2(r_1 + \cdots + r_k)$. $\square$

Věta (O vztahu $N(\mu, \sigma^2)$ a χ^2 s jedním stupněm volnosti)

Pokud má X rozdělení $N(\mu, \sigma^2)$, pak má $V = \dfrac{(X-\mu)^2}{\sigma^2}$ rozdělení $\chi^2(1)$.

Důkaz (začátek).

Použitím předchozí věty máme, že $Z = \dfrac{(X-\mu)}{\sigma}$ má rozdělení $N(0,1)$.

Pro distribuční funkce F_V náhodné veličiny V a každé $v \geq 0$ platí:

$$F_V(v) = P(\{V \leq v\}) = P(\{Z^2 \leq v\}) = P(\{-\sqrt{v} \leq Z \leq \sqrt{v}\})$$

To znamená, že pro $v \geq 0$ máme

$$F_V(v) = \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{z^2}{2}} \, \mathrm{d}z = 2 \cdot \int_{0}^{\sqrt{v}} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{z^2}{2}} \, \mathrm{d}z.$$

Důkaz.

Provedeme substituci $z = \sqrt{y}$, to jest dostaneme

$$F_V(v) = \int_0^v \frac{1}{\sqrt{2 \cdot \pi \cdot y}} \cdot e^{-\frac{y}{2}} \,\mathrm{d}y, \qquad \text{pro } y \geq 0.$$

Z definice V je zřejmé, že $F_V(v) = 0$ pro $v < 0$. To znamená, že funkci hustoty f_V můžeme získat z F_V jako důsledek základní věty integrálního počtu:

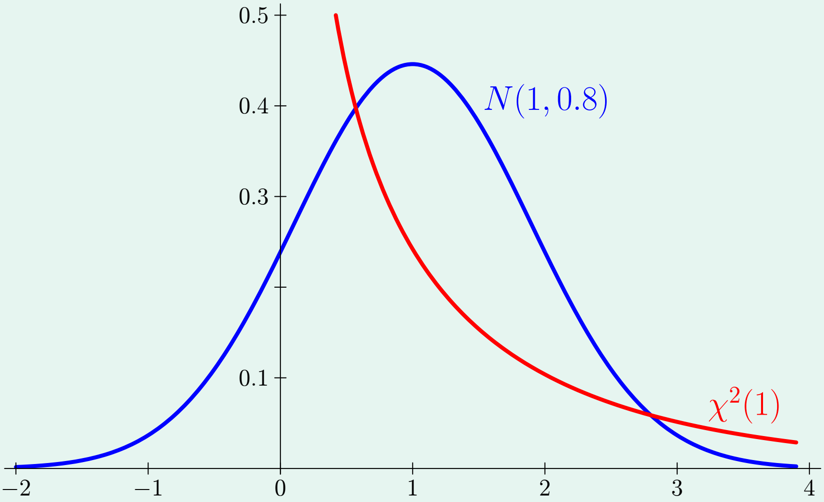
$$f_V(v) = \frac{1}{\sqrt{2 \cdot \pi \cdot v}} \cdot e^{-\frac{v}{2}} = \frac{1}{\sqrt{\pi} \cdot \sqrt{2}} \cdot v^{\frac{1}{2}-1} \cdot e^{-\frac{v}{2}}, \quad v > 0.$$

Nyní stačí ověřit, že $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Jelikož je f_V funkce hustoty, máme:

$$\int_0^\infty \frac{1}{\sqrt{\pi} \cdot \sqrt{2}} \cdot v^{\frac{1}{2}-1} \cdot e^{-\frac{v}{2}} \, dv = 1.$$

Substituce $x = \dfrac{v}{2}$ dává $1 = \dfrac{1}{\sqrt{\pi}} \cdot \int_0^\infty x^{\frac{1}{2}-1} \cdot e^{-x} \,\mathrm{d}x = \dfrac{1}{\sqrt{\pi}} \cdot \Gamma\left(\dfrac{1}{2}\right)$. To jest, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, a tím pádem V má rozdělení $\chi^2(1)$. $\square$

Příklad (Vztah χ^2 rozdělení a normálních rozdělení)



Věta (O součtech čtverců nezávislých normálních veličin)

Mějme nezávislé náhodné veličiny $Z_1, \ldots, Z_n$, které mají standardní normální rozdělení $N(0,1)$ a nezávislé náhodné veličiny $X_1, \ldots, X_n$, které mají normální rozdělení $N(\mu_i, \sigma_i^2)$ pro $i = 1, 2, \ldots, n$. Pak

1. *náhodná veličina $V = Z_1^2 + \cdots + Z_n^2$ má rozdělení $\chi^2(n)$;*
2. *náhodná veličina $W = \sum_{i=1}^{n} \dfrac{(X_i - \mu_i)^2}{\sigma_i^2}$ má rozdělení $\chi^2(n)$.*

Důkaz.

Tvrzení přímo plyne z předchozích vět:

- Z^2 má rozdělení $\chi^2(1)$ pokud má Z rozdělení $N(0,1)$;
- součet n nezávislých veličin s rozdělením $\chi^2(1)$ má rozdělení $\chi^2(n)$;
- pokud má X rozdělení $N(\mu, \sigma^2)$, pak má $(X - \mu)/\sigma^2$ rozdělení $N(0,1)$. □

Věta (O rozdělení lineárních kombinací nezávislých normálních veličin)

Mějme nezávislé náhodné veličiny $X_1, \ldots, X_n$, *které mají normální rozdělení* $N(\mu_i, \sigma_i^2)$ *pro* $i = 1, 2, \ldots, n$. *Pak pro každá reální čísla* $a_1, \ldots, a_n \in \mathbb{R}$ *platí, že lineární kombinace* $Y = \sum_{i=1}^{n} a_i \cdot X_i$ *má normální rozdělení*

$$N\left(\sum_{i=1}^{n} a_i \cdot \mu_i, \sum_{i=1}^{n} a_i^2 \cdot \sigma_i^2\right).$$

Důkaz.

Užitím věty o moment generujících funkcích lineárních kombinací nezávislých náhodných veličin dostáváme

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} \exp\left(\mu_i a_i t + \tfrac{1}{2}\sigma_i^2 a_i^2 t^2\right),$$

protože $M_{X_i} = \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right)$. Potom tedy

$$M_Y(t) = \exp\left(\left(\sum_{i=1}^{n} a_i \mu_i\right) t + \tfrac{1}{2}\left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right) t^2\right),$$

což je moment generující funkce prokazovaného normálního rozdělení. □

Důsledek: Rozdělení průměru náhodného výběru z $N(\mu, \sigma^2)$.

Pokud je $X_1, \ldots, X_n$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$, pak průměr $\overline{X}$ náhodného výběru $X_1, \ldots, X_n$ má rozdělení $N(\mu, \sigma^2/n)$. □

Průměr $\overline{X}$ náhodného výběru $X_1, \ldots, X_n$ z $N(\mu, \sigma^2)$ má normální rozdělení

- se stejnou střední hodnotu μ,
- ale *menším rozptylem* rozptylem σ^2/n.

Příklad (Rozdělení průměru náhodného výběru z $N(\mu, \sigma^2)$)

Pro náhodný výběr $X_1, \ldots, X_{64}$ z rozdělení $N(50, 16)$ máme

$$P(\{49 < X_i < 51\}) = P\left(\left\{\frac{49-50}{4} \leq \frac{X-50}{4} \leq \frac{51-50}{4}\right\}\right) = 0.1974$$

$$P(\{49 < \overline{X} < 51\}) = P\left(\left\{\frac{49-50}{0.5} \leq \frac{X-50}{4} \leq \frac{51-50}{0.5}\right\}\right) = 0.9544.$$

Příklad (Motivace pro centrální limitní větu)

Pro libovolný n-prvkový náhodný výběr $X_1, \ldots, X_n$ z rozdělení se střední hodnotou μ a rozptylem σ^2. Pro průměr $\overline{X}$ tohoto náhodného výběru můžeme zavést W:

$$W = \frac{\sqrt{n}}{\sigma} \cdot (\overline{X} - \mu) = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

Pro každé $n \in \mathbb{N}$ máme z věty o lineárních kombinacích nezávislých veličin

$$\mu_W = E(W) = E\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = \frac{E(\overline{X}) - \mu}{\sigma/\sqrt{n}} = \frac{\mu - \mu}{\sigma/\sqrt{n}} = 0.$$

S využitím předchozího faktu a $\sigma_W^2 = E(W^2) - \mu_W^2$ dostáváme, že

$$\sigma_W^2 = E(W^2) - 0 = E\left(\frac{(\overline{X} - \mu)^2}{\sigma^2/n}\right) = \frac{E\big((\overline{X} - \mu)^2\big)}{\sigma^2/n} = \frac{\sigma^2/n}{\sigma^2/n} = 1.$$

Otázka: *Jak je W distribuované pro $n \to \infty$?*

Centrální limitní věta

Věta (Centrální limitní věta, angl.: *central limit theorem*).

Nechť $\overline{X}$ je průměr n-prvkového náhodného výběru z rozdělení se střední hodnotou μ a rozptylem $\sigma^2 > 0$. Pak

$$W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

má rozdělení $N(0,1)$ pro $n \to \infty$. $\square$

Aplikace: Pokud je n (velikost náhodného výběru) *dost velké*, pak:

$$W = \boxed{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} = \frac{\frac{n}{n}\sum_{i=1}^{n} X_i - n \cdot \mu}{n \cdot \sigma/\sqrt{n}} = \boxed{\frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n} \cdot \sigma}}$$

má *přibližně* standardní normální rozdělení $N(0,1)$.

Příklad (Použití centrální limitní věty)

Problém: Uvažujme průměr $\overline{X}$ patnáctiprvkového náhodného výběru ze spojitého rozdělení, jehož hustota je

$$f_X(x) = \frac{3}{2} \cdot x^2, \qquad \text{pro} \ -1 < x < 1,$$

$f_X(x) = 0$ ve všech ostatních případech.

Úkol: *Pomocí centrální limitní věty odhadněte* $P(\{0.03 \leq \overline{X} \leq 0.15\})$.

Řešení: Střední hodnota a rozptyl X jsou $\mu_X = 0$ a $\sigma_X^2 = 0.6$. Náhodný výběr je patnáctiprvkový, to jest $n = 15$. Pomocí centrální limitní věty dostáváme:

$$\begin{aligned} P(\{0.03 \leq \overline{X} \leq 0.15\}) &= P\left(\left\{\frac{0.03 - 0}{\sqrt{0.6}/\sqrt{15}} \leq \frac{\overline{X} - 0}{\sqrt{0.6}/\sqrt{15}} \leq \frac{0.15 - 0}{\sqrt{0.6}/\sqrt{15}}\right\}\right) \\ &\approx P(\{0.15 \leq W \leq 0.75\}) \approx \Phi(0.75) - \Phi(0.15) \\ &= 0.7734 - 0.5594 = 0.2138. \end{aligned}$$

Aproximace binomického rozdělení normálním rozdělením

Uvažujme diskrétní náhodnou veličinu X, která má rozdělení $b(n, p)$.

Pozorování: X *je součtem* n *nezávislých veličin s Bernoulliho rozdělením*, to jest:

$$X = \sum_{i=1}^{n} X_i,$$

kde $X_1, \ldots, X_n$ jsou nezávislé veličiny, které mají identické Bernoulliho rozdělení s parametrem p. To jest, $\mu_{X_i} = p$ a $\sigma^2_{X_i} = p \cdot (1 - p)$ pro každé $i = 1, \ldots, n$.

Aplikací centrální limitní věty dostáváme, že

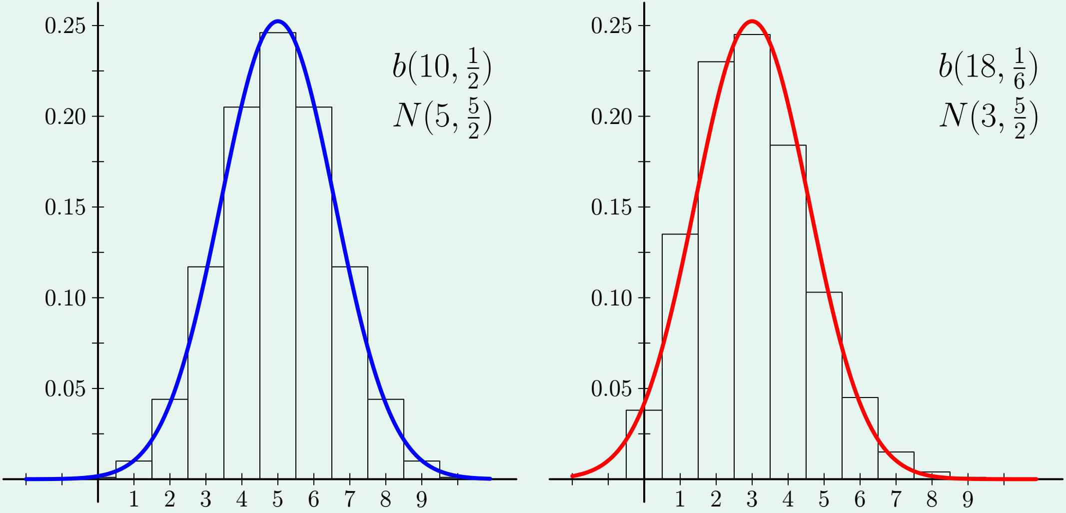
$$W = \frac{X - n \cdot p}{\sqrt{n} \cdot \sqrt{p \cdot (1 - p)}}$$

má rozdělení $N(0, 1)$ pokud $n \to \infty$. Odtud $X = W \cdot \sqrt{n \cdot p \cdot (1 - p)} + n \cdot p$. Pokud je n dost velké, pak má X přibližně rozdělení

$$\boxed{N\big(n \cdot p, n \cdot p \cdot (1 - p)\big)}\,.$$

Příklady (Aproximace binomického rozdělení normálním rozdělením)

Pokud je n velké, pak $b(n,p) \approx N\big(n \cdot p, n \cdot p \cdot (1-p)\big)$.



Aproximace Poissonova rozdělení normálním rozdělením

Uvažujme X, která má Poissonovo rozdělení s parametrem $\lambda = n \in \mathbb{N}$.

Pozorování: X *je součtem* n *nezávislých Poissonových veličin s parametrem* 1:

$$X = \sum\nolimits_{i=1}^{n} X_i,$$

kde $X_1, \ldots, X_n$ jsou nezávislé veličiny, které mají identické Poissonovo rozdělení s parametrem 1. To jest, $\mu_{X_i} = 1$ a $\sigma^2_{X_i} = 1$ pro každé $i = 1, \ldots, n$.

Aplikací centrální limitní věty dostáváme, že

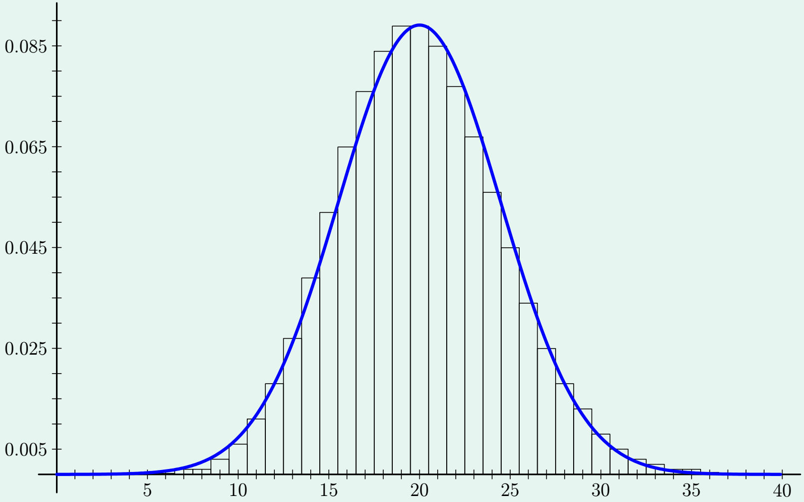
$$W = \frac{X - n \cdot 1}{\sqrt{n} \cdot \sqrt{1}}$$

má rozdělení $N(0, 1)$ pokud $n \to \infty$. Odtud $X = W \cdot \sqrt{n} + n = W \cdot \sqrt{\lambda} + \lambda$. Pokud je n dost velké, pak má X přibližně rozdělení

$$\boxed{N(\lambda, \lambda)}\,.$$

Příklad (Aproximace Poissonova rozdělení normálním rozdělením)

Pokud je λ velké, pak Poissonovo rozdělení s parametrem $\lambda \approx N(\lambda, \lambda)$.



Čebyševova nerovnost

Věta (Пафнутий Львович Чебышёв)

Pokud má náhodná veličina X střední hodnotu μ_X a rozptyl $\sigma_X^2 > 0$, pak pro každé přirozené číslo $k \geq 1$, platí

$$P(\{k \cdot \sigma_X \leq |X - \mu_X|\}) \leq \frac{1}{k^2}.$$

Interpretace:

- *Pravděpodobnost, že pozorovaná hodnota náhodné veličiny X se odchyluje od střední hodnoty veličiny aspoň o k-násobek směrodatné odchylky náhodné veličiny je menší nebo rovna $\frac{1}{k^2}$.*

Poznámky:

- Výsledek *nezávisí* na tom, jaké má X rozdělení. (**!!**)
- Lze aplikovat pro odhady pravděpodobnosti bez znalosti rozdělení (pesimistické).

Důkaz.

Uvažujme, že X je náhodná veličina v prostoru $\langle \Omega, \mathcal{F}, P \rangle$. Pro dané k položme

$$A = \{k \cdot \sigma_X \leq |X - \mu_X|\} = \{\omega \in \Omega \,|\, k \cdot \sigma_X \leq |X(\omega) - \mu_X|\}.$$

Z linearity E můžeme psát

$$\begin{aligned}\sigma_X^2 &= E\big((X - \mu_X)^2\big) = E\big(\mathbf{1}_\Omega \cdot (X - \mu_X)^2\big) = E\big(\mathbf{1}_{A \cup (\Omega - A)} \cdot (X - \mu_X)^2\big) \\ &= E\big(\mathbf{1}_A \cdot (X - \mu_X)^2\big) + E\big(\mathbf{1}_{\Omega - A} \cdot (X - \mu_X)^2\big).\end{aligned}$$

Zobrazení $\mathbf{1}_{\Omega - A} \cdot (X - \mu_X)^2$ nabývá pro všechny $\omega \in \Omega$ nezáporných hodnot, to jest $E\big(\mathbf{1}_{\Omega - A} \cdot (X - \mu_X)^2\big) \geq 0$. Z předchozí rovnosti proto $\sigma_X^2 \geq E\big(\mathbf{1}_A \cdot (X - \mu_X)^2\big)$. Pro každý $\omega \in A$ navíc máme $k \cdot \sigma_X \leq |X(\omega) - \mu_X|$, to znamená

$$\sigma_X^2 \geq E\big(\mathbf{1}_A \cdot (X - \mu_X)^2\big) \geq E\big(\mathbf{1}_A \cdot (k \cdot \sigma_X)^2\big) = k^2 \cdot \sigma_X^2 \cdot E(\mathbf{1}_A) = k^2 \cdot \sigma_X^2 \cdot P(A).$$

Vyjádřením $P(A)$ z předchozí nerovnosti a zkrácením nenulového σ_X^2 dostáváme

$$P(\{k \cdot \sigma_X \leq |X - \mu_X|\}) = P(A) \leq \frac{1}{k^2}.$$

□

Příklad (Dolní odhad pravděpodobnosti bez znalosti rozdělení)

Problém: Náhodná veličina X má rozdělení se střední hodnotou $\mu_X = 25$ a rozptylem $\sigma_X^2 = 16$.

Úkol: *Bez znalosti* P_X *stanovte dolní odhad* $P(\{17 < X < 33\})$.

Řešení:

$$\begin{aligned}
P(\{17 < X < 33\}) &= P(\{17 < X\} \cap \{X < 33\}) \\
&= P(\{17 - 25 < X - 25\} \cap \{X - 25 < 33 - 25\}) \\
&= P(\{-8 < X - 25\} \cap \{X - 25 < 8\}) \\
&= P(\{|X - 25| < 8\}) = P(\{|X - 25| < 2 \cdot \sigma_X\}).
\end{aligned}$$

Použitím Čebyševovy nerovnosti máme $P(\{|X - 25| \geq 2 \cdot \sigma_X\}) \leq \frac{1}{4}$, to znamená

$$P(\{|X - 25| < 2 \cdot \sigma_X\}) = 1 - P(\{|X - 25| \geq 2 \cdot \sigma_X\}) \geq 1 - \frac{1}{4} = 0.75.$$

Dohromady tedy dostáváme, že $P(\{17 < X < 33\}) \geq 0.75$.

Věta (Zákon velkých čísel)

Pro posloupnost náhodných výběrů $X_1, \ldots, X_n$ (kde $n \to \infty$) z rozdělení se střední hodnotou μ a rozptylem σ^2 a jejich průměry $\overline{X}$ platí:

$$\lim_{n\to\infty} P(\{|\overline{X} - \mu| < \varepsilon\}) = 1.$$

Důkaz.

Pro každé $\overline{X}$ platí, že $\mu_{\overline{X}} = \mu$ a $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$. Použitím Čebyševovy nerovnosti pro $\varepsilon = k \cdot \sigma_{\overline{X}} = k \cdot \frac{\sigma}{\sqrt{n}}$ dostáváme

$$0 \leq P(\{\varepsilon \leq |\overline{X} - \mu|\}) \leq \frac{1}{k^2} = \frac{1}{\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma}\right)^2} = \frac{1}{\frac{\varepsilon^2 \cdot n}{\sigma^2}} = \frac{\sigma^2}{\varepsilon^2 \cdot n}.$$

Odtud ihned dostáváme

$$\lim_{n\to\infty} P(\{\varepsilon \leq |\overline{X} - \mu|\}) \leq \lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2 \cdot n} = 0.$$

Z vlastností komplementárních jevů potom $\lim_{n\to\infty} P(\{|\overline{X} - \mu| < \varepsilon\}) = 1.$ $\square$

Přednáška 9: Závěr

Pojmy:

- náhodný výběr, statistika, výběrové rozdělení, lineární kombinace veličin
- (standardní) normální rozdělení, Box-Mullerova transformace, rozdělení χ^2
- centrální limitní věta, aproximace diskrétních rozdělení (binomické, Poissonovo)
- Čebyševova nerovnost, zákon velkých čísel

Použité zdroje:

Billingsley, P.: *Probability and Measure*
John Wiley & Sons; 3. vydání, ISBN 978–0–471–00710–4.

Gentle J. E.: *Random Number Generation and Monte Carlo Methods*
Springer 2004, ISBN 978–0–387–00178–4.

Hogg R. V., Tanis E. A.: *Probability and Statistical Inference*
Prentice Hall; 7. vydání 2005, ISBN 978–0–13–146413–1.