11 Rovnoměrné a normální rozdělení psti

Břetislav Fajmon, UMAT FEKT, VUT Brno

Této přednášce odpovídá část kapitoly 13 ze skript [1] a vše, co se nachází v kapitole 7 sbírky úloh [2] – tuto kapitolu 7 sbírky úloh je vhodné vzít jako cvičení tohoto tématu.

Doporučuji přijít také na poslední přednášku příští týden, protože bude probírána statistika – téma, které nebude procvičeno ve cvičeních odborného základu, ale u zkoušky by se mohlo vyskytnout.

Dnes se budeme zabývat dvěma popisy spojité veličiny. Jedná se o spojité veličiny, konkrétně o

- **rovnoměrné spojité rozdělení**
- **normální rozdělení**

11.1 Rovnoměrné spojité rozdělení

Už minulý týden bylo zmíněno jedno důležité rozdělení, a sice rozdělení exponenciální. Existuje ovšem ještě jedno jednodušší spojité rozdělení, a sice rovnoměrné (spojité) rozdělení. Rovnoměrně spojitě rozdělená veličina je veličina nabývající hodnot z jistého intervalu $\langle a; b\rangle$, a to jakákoli hodnota je stejně častá jako ty ostatní, tj. její hustotou je úsečka na tomto intervalu.

Je to vlastně rozdělení reprezentující geometrickou pravděpodobnost (viz předn. číslo 7), kde pravděpodobnost naměření jistých hodnot počítáme jako délku intervalu.

11.2 Normální rozdělení pravděpodobnosti

Dnes se budeme zabývat zejména rozdělením normálním. Jeho název zřejmě pochází z toho, že hodně veličin praxe lze jím dosti dobře popsat (nikoli ze tvaru její hustoty pravděpodobnosti – ten je naopak značně hrozivý):

$$f(t) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}. \tag{1}$$

Dále lze zjistit výpočtem podle příslušných vzorců (viz přednáška 8), že

$$EX = \mu, \quad DX = \sigma^2 \tag{2}$$

(jedná se přesně o parametry, které se vyskytují ve

vzorci hustoty $f(t)$). Dále distribuční funkci musíme počítat integrací, protože jiný vzorec pro její výpočet neexistuje (o tom bude víc řečeno za chvilinku):

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt. \tag{3}$$

Normální rozdělení se stalo slavným díky tomu, co říká tzv.
centrální limitní věta:

Pokud $X_1, X_2, \dots X_N$ **jsou nezávislé stejně rozdělené veličiny (rozdělení psti může být libovolné, diskrétní či spojité, ale pro všech** N **veličin stejné) se střední hodnotou** $EX_i = \mu$ **a rozptylem** $DX_i = \sigma^2$**, pak veličina**

$$Y := \sum_{1}^{N} X_i$$

má pro dostatečně velké N **rozdělení normální se střední hodnotou** $N \cdot \mu$ **a rozptylem** $N \cdot \sigma^2$**.**

Jaký je důsledek této matematické věty?

Ten, že celá řada veličin je ovlivněna součtem mnoha vlivů – proto řadu veličin lze popsat normálním rozdělením.

Příklad 11.1. **Veličina** Y **udává výšku borovice (v metrech) v daném lese. Průměrná výška (**EY**) je** 50 **metrů. Vezměme nyní konkrétní strom, jeho výška je** 54 **metrů. Co způsobilo, že vyrostl o čtyři metry nad průměr? Řada různých vlivů:**

- **Stromek byl zasazen v obzvlášť příznivém období roku, což způsobilo, že vyrostl o jeden metr nad průměr.**

- **Místo, kde strom roste, získává zdroje hnojiva navíc, což vede k růstu o** 2,3 **m nad průměr.**
- **Nešťastnou náhodou byl stromek při sazení nalomen, což způsobilo, že vyrostl o** 1,4 **m méně, než by mohl.**
- **Strom má dobré místo na slunci, což mu pomohlo vyrůst o** 2 **m nad průměr.**
- **Skupina jedinců antagonistického hmyzu si vybrala strom za svůj domov, což mu vzalo šance vyrůst 0** 0,6 **m výš, než ostatní stromy.**
- **Atd. Zkrátka a dobře, vychýlení o čtyři metry nad průměr je dáno součtem všech těchto mož-**

ných kladných i záporných vlivů. Protože těchto vlivů je poměrně dost, výslednou výšku stromů danou součtem všech těchto vlivů lze s velkou přesností popsat normálním rozdělením.

Příklad 11.2. Veličina Y udává výsledek zkoušky z matematiky. Vezmeme nyní výsledek zkoušky jednoho konkrétního studenta, který se jmenuje Honza. Co naň mělo vliv?

- Honza měl den před zkouškou chřipku – to snížilo jeho výkon o deset bodů.
- Honza si něco tipl a náhodou to trefil – přidalo mu to 2 body.

- **Honza chyběl na klíčové přednášce a neměl u zkoušky její kopii – přišel o 5 bodů.**
- **Profesor byl v dobré náladě a přidal při opravování Honzovi tři body zadarmo.**
- **Profesor zjistil, že student počítá střední hodnotu diskrétní veličiny pomocí integrálu (nebo naopak střední hodnotu spojité veličiny pomocí sumy), a to jej rozladilo.**
- **Atd. Opět vidíme, že výsledek Honzovy zkoušky je dán součtem většího počtu navzájem nezávislých náhodných vlivů, a tedy jej lze s velkou přesností popsat normálním rozdělením.**

Následující důsledek centrální limitní věty je někdy nazýván
<u>Moivre-Laplaceova věta</u>:

Speciálně i binomické rozdělení $Bi(n,p)$ lze pro dostatečně velké n dobře popsat (= aproximovat, nahradit) normálním rozdělením se stejnou střední hodnotou a rozptylem.

Uvažujme například veličinu Y, která udává počet líců při sto hodech korunou. Tato veličina má binomické rozdělení s parametry $n = 100$, $p = 0{,}5$. Tuto veličinu Y lze také vyjádřit jako součet veličin X_1, X_2, …, X_{100}, kde každé z rozdělení X_i má parametry $n = 1$, $p = 0{,}5$ (rozdělení X_i se někdy nazývá <u>alternativní rozdělení</u>, protože příslušná veličina

může nabývat jen dvou alternativ: 0 (= padl rub mince) a 1 (= padla koruna horní stranou nahoru)).

Veličina $Y = \sum_1^{100} X_i$ **má rozdělení binomické (**$EY = 50$**,** $DY = 25$**), ale díky centrální limitní větě lze tutéž veličinu popsat též normálním rozdělením se stejnou střední hodnotou (**$\mu = 50$**) a se stejným rozptylem (**$\sigma^2 = 25$**).**

Na konkrétní příklad nahrazení binomického rozdělení normálním i na další výpočty se podívejme v dalším oddílku.

11.3 Výpočty pravděpodobností normálního rozdělení

Je potřeba říci, že výpočet integrálů z hustoty normálního rozdělení je neobyčejně náročný, a z toho důvodu se provádějí při výpočtech následující dva kroky:

1. **Veličina Y, jejíž rozdělení je normální s parametry $EY = \mu$, $DY = \sigma^2$, se převede na veličinu**

$$U = \frac{Y - \mu}{\sigma} \tag{4}$$

(druhou mocninu jsem nezapomněl, ve jmenovateli zlomku je skutečně pouze $\sqrt{DY} = \sigma$). Pro veličinu U lze spočítat, že platí

$$EU = 0, \quad DU = 1.$$

2. **Hodnoty distribuční funkce této <u>normované = standardizované</u> normální veličiny se jednou provždy spočetly do tabulky – místo klasického označení F tuto distribuční funkci u**

veličiny U standardně označujeme Φ:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt. \tag{5}$$

Tabulku uvedenou na str. 216-217 skript [1] si vytiskněte a přineste na zkoušku, budete pomocí ní počítat příklady s normálním rozdělením.

Podobné tabulky jsou také jedním z důvodů, proč se pojem distribuční funkce zavádí – v tabulce jsou právě uvedeny hodnoty pravděpodobností $\int_{-\infty}^{x} f(t)dt$.

Mimochodem, u normálně rozdělené veličiny uvedená primitivní (= zintegrovaná) funkce je tzv. vyšší funkcí, kterou nelze vyjádřit jinak než nekonečnou řadou (a tu integrovat člen po členu), nebo je možné spočítat integrál numericky pomocí metod uvedených v první části tohoto předmětu (složená lichoběžníková nebo Simpsonova metoda).

Podrobněji viz str. 220 skript – generování hodnot normálního rozdělení na počítači.

Ukažme si oba kroky výpočtu s normálním rozdělením na příkladu:

Příklad 11.3. **Veličina Y udává velikost proudu v dané části obvodu. Její rozdělení je normální s parametry $EY = \mu = 10$ mA, $DY = \sigma^2 = 4$ (mA)^2, tj. směrodatná odchylka $\sigma = 2$ mA.**

a) **Určete pravděpodobnost, že měření proudu překročí 13 mA.**

b) **Určete pravděpodobnost, že naměřený proud bude v intervalu $\langle 9; 11 \rangle$ mA.**

c) **Pravděpodobnost, že naměřený proud je větší než y_0, je rovna $0{,}6$. Určete y_0.**

Řešení:

$$P(Y > 13) = P(\frac{Y-10}{2} > \frac{13-10}{2}) = P(U > 1{,}5) =$$
$$= 1 - \Phi(1{,}5) \doteq 1 - 0{,}9332 = 0{,}0668.$$

Úpravu nerovnosti $Y > 13$ jsme provedli takovou, abychom cíleně na jedné straně nerovnosti dostali výraz $\frac{Y-\mu}{\sigma}$, a ten pak podle 4 nahradíme proměnnou U. Hodnotu $\Phi(1{,}5)$ jsme nalezli v tabulce, kterou po takto provedené úpravě můžeme použít.

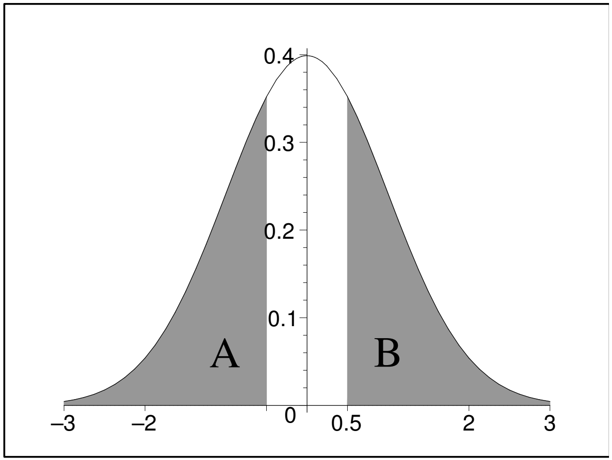
ad b) Podobně postupujeme při jakémkoli užití výpočtu pravděpodobnosti u normální veličiny:

$$P(9 < Y < 11) = P(\frac{9-10}{2} < \frac{Y-10}{2} < \frac{11-10}{2}) =$$
$$= P(-0{,}5 < U < 0{,}5) = \Phi(0{,}5) - \Phi(-0{,}5) \doteq 0{,}6915 - (1 - 0{,}6915)$$
$$= 0{,}383.$$

Φ záporných hodnot není uvedeno – místo toho využíváme symetrie hustoty $f(x)$, čili pro distribuční funkci ϕ platí

$$\Phi(-u) = 1 - \Phi(u)$$

tento fakt je ilustrován na obrázku:



ad c) Stejně postupujeme i při „inverzním" procesu, kdy je známa pravděpodobnost a hledáme hodnotu y_0 na vodorovné ose:

$$\begin{aligned} P(Y > y_0) &= 0{,}6 \\ P\left(\frac{Y-10}{2} > \frac{y_0-10}{2}\right) &= 0{,}6 \\ P\left(U > \frac{y_0-10}{2}\right) &= 0{,}6 \\ 1-\Phi\left(\frac{y_0-10}{2}\right) &= 0{,}6 \\ 0{,}4 &= \Phi\left(\frac{y_0-10}{2}\right) \end{aligned}$$

A nyní pozor – musíme v tabulce ϕ najít, ve kterém bodě nabývá funkce Φ hodnotu $0{,}4$. Musíte hledat v

tom správném sloupci tabulky. Pokud hledáte v tom správném sloupci tabulky, tak

Pokud hledáte v tom správném sloupci tabulky, tak hodnotu $\frac{y_0-10}{2}$, že $\phi(\frac{y_0-10}{2}) = 0{,}4$, v tabulce nenajdete, protože $\frac{y_0-10}{2} < 0$ a v tabulce jsou jen hodnoty kladné.

Znovu musíme užít vzorce $\Phi(-u) = 1 - \Phi(u)$, ale nyní jaksi naopak, protože u – nebo vlastně $\frac{y_0-10}{2}$ – je nyní záporné, nikoli kladné: MUSÍME UMĚLE VYROBIT POMOCÍ TOHOTO VZORCE V ZÁVORCE MINUS, PROTOŽE $(-\frac{y_0-10}{2})$ je kladné číslo – pak teprve lze $(-\frac{y_0-10}{2})$ najít v tabulce.

$$
\begin{aligned}
0{,}4 &= \Phi\left(\frac{y_0 - 10}{2}\right) \\
1 - 0{,}4 &= \Phi\left(\frac{10 - y_0}{2}\right) \\
0{,}25 &\doteq \frac{10 - y_0}{2} \\
y_0 &= 10 - 0{,}5 = 9{,}5.
\end{aligned}
$$

Příklad 11.4. Vypočtěte pravděpodobnost, že veličina X s rozdělením $No(\mu, \sigma^2)$ leží v intervalu a) $\langle \mu - \sigma; \mu + \sigma \rangle$, b) $\langle \mu - 2\sigma; \mu + 2\sigma \rangle$, c) $\langle \mu - 3\sigma; \mu + 3\sigma \rangle$.

Výsledek: a] 0,6827; b] 0,9545; c] 0,9973.

ad c: Většina hodnot veličiny X leží tedy v intervalu $< \mu_x - 3\sigma_x, \mu_x + 3\sigma_x >$. Veličina X nabude hodnoty z tohoto intervalu s pravděpodobností $99{,}7$ (= tzv. pravidlo tří sigma).

Příklad 11.5. **Firma vyrábí balíčky ořechů po** 200**ks, přičemž** $\frac{3}{4}$ **oříšků jsou burské a** $\frac{1}{4}$ **lískové, dokonale se promíchají, a pak se teprve sypou do balíčků. Jestliže koupíme jeden balíček ořechů, jaká je pravděpodobnost, že počet lískových ořechů je v intervalu** $< 47; 56 >$**?**

<u>Řešení:</u> Náhodná veličina X **udávající počet lískových ořechů v jednom balíčku má rozdělení** $Bi(N = 200, p = 0{,}25)$**, čili** $\mu_x = 50$**,** $\sigma_x^2 = 37{,}5$**. Přímý výpočet**

$$P\ (47 \leq X \leq 56) = P(X = 47) + P(X = 48) + \cdots + P(X = 56) =$$
$$= \binom{200}{47} 0{,}25^{47} 0{,}75^{153} + \binom{200}{48} 0{,}25^{48} 0{,}75^{152} + \cdots +$$

$$+\binom{200}{56}0{,}25^{56}0{,}75^{144} = 0,572$$

byl určen pomocí robustní kalkulačky, která má funkci pro obecnou sumu a také funkci pro vyčíslení kombinačních čísel. Při náhradě daného binomického rozdělení normálním rozdělením se stejnou střední hodnotou a rozptylem ($\sigma_x^2 = 37{,}5 \implies \sigma_x \doteq 6{,}12$) dostaneme výsledek:

$$P(47 \leq X \leq 56) = P\left(\frac{47-50}{6{,}12} \leq U \leq \frac{56-50}{6{,}12}\right) =$$
$$\Phi(0{,}98) - \Phi(-0{,}49) = \Phi(0{,}98) - (1 - \Phi(0{,}49)) \doteq 0{,}524.$$

Je vidět, že chyba od přesného výsledku je v řádu procent (druhé desetinné místo). Pokud bychom

použili korekce (viz následující příklad 6), dostali bychom výsledek $P(46,5 \leq X \leq 56,5) = 0,569$**, jehož odchylka od přesného výsledku je v řádu desetin procenta (třetí desetinné místo).**

Příklad 11.6. Náhodná veličina X **udává počet líců při čtyřech hodech mincí. Vypočteme například pravděpodobnost, že počet líců ve čtyřech hodech bude jeden nebo dva**

a) pomocí $Bi(N = 4, p = 0{,}5)$**;**

b) pomocí normálního rozdělení;

c) *pomocí normálního rozdělení s korekcí.*

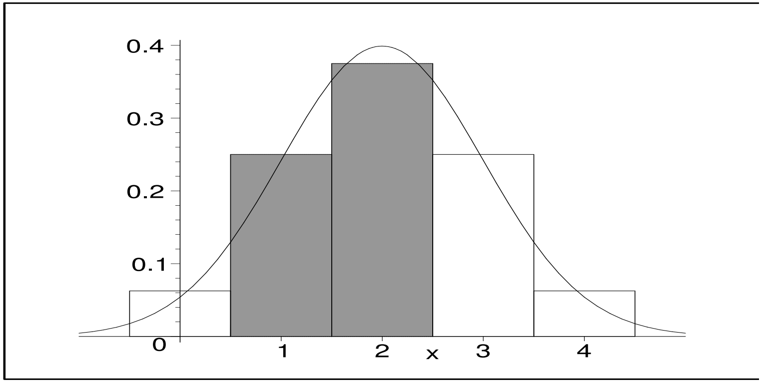
Řešení:

ad a) $P(1 \leq X \leq 2) = p_1 + p_2 = 0{,}25 + 0{,}375 = 0{,}625.$

ad b) *Aproximujme binomické rozdělení normálním rozdělením* $No(\mu_x = Np = 2, \sigma_x^2 = Np(1-p) = 1)$*:*

$$P(1 \leq X \leq 2) = P\left(\frac{1-\mu_x}{\sigma_x} \leq U \leq \frac{2-\mu_x}{\sigma_x}\right) =$$
$$P\left(\frac{1-2}{1} \leq U \leq \frac{2-2}{1}\right) = \Phi(0) - \Phi(-1) = 0{,}341.$$

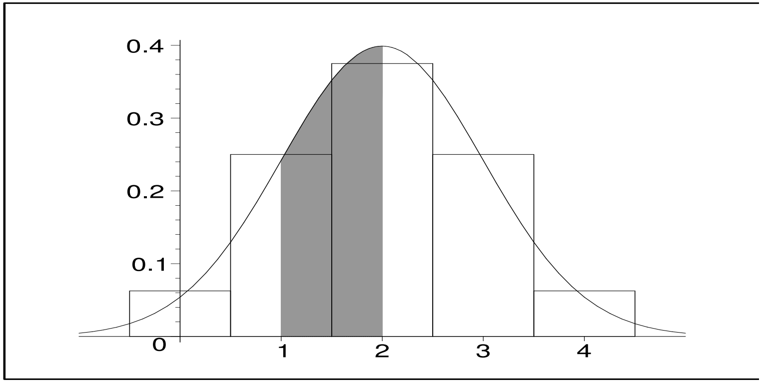
Hodnota z b) se od hodnoty z a) významně liší!! Kde se udála tak velká chyba? V tom, že obsah plochy dvou obdélníků histogramu na obr.1



Obrázek 1: K př. 6 - aproximovaná plocha.

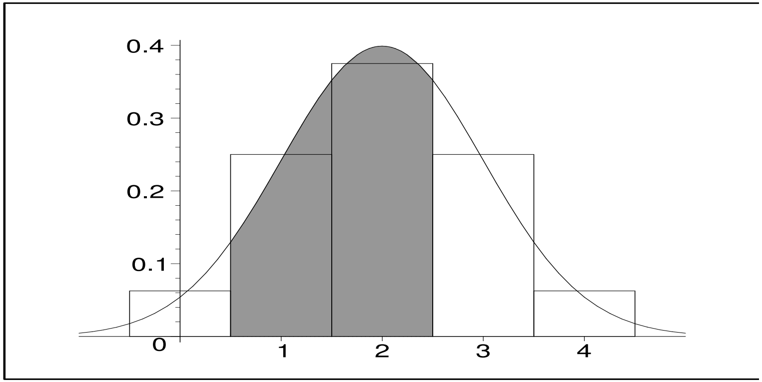
jsme aproximovali pomocí obsahu plochy na obr. 2,

nikoliv pomocí šrafované plochy na obr. 3.



Obrázek 2: K př. 6 - nevhodná aproximace Bi pomocí No.

Aproximační chyba se zmenší, pokud výpočet pravděpodobnosti $P(t_1 \leq X \leq t_2)$ **pomocí** Bi **nahradíme obsahem podgrafu hustoty** No **na intervalu stejné délky, tj. pravděpodobností**



Obrázek 3: K př. 6 - vhodná aproximace Bi pomocí No užitím korekce.

$P(t_1 - 0{,}5 \leq X \leq t_2 + 0{,}5)$. **Toto rozšíření intervalu o** $0{,}5$ **na obou stranách nazýváme korekcí.**

ad c) *V našem příkladu dostaneme užitím korekce:*

$$
\begin{aligned}
&P(1-0{,}5 \leq X \leq 2+0{,}5) = \\
&= P\left(\frac{1-0{,}5-2}{1} \leq U \leq \frac{2+0{,}5-2}{1}\right) = \\
&= \Phi(\frac{1}{2}) - \Phi(-\frac{3}{2}) = 0{,}624,
\end{aligned}
$$

což je docela dobrá aproximace přesné hodnoty $0{,}625$.

Je vidět, že pomocí korekce lze popsat binomické rozdělení normálním i pro malá N.

Otázky k opakování

Z dnešní přednášky vyvstávají tři otázky k opakování pro studenty, jejichž odpovědi je záhodno mít s sebou na písemky a znát zpaměti:

otázka č. 19. **Normální rozdělení – význačné spojité rozdělení číslo 2: jaké jsou jeho základní vlastnosti?**
Odpověď:

a) $EX = \mu$, $DX = \sigma^2$;

b) Hustota psti je dána vztahem

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

Distribuční funkce je dána vztahem

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

c) **Pokud veličina X má normální rozdělení s parametry $EX = \mu, \quad DX = \sigma^2$, pak veličina**

$$U := \frac{X - \mu}{\sigma}$$

má normální rozdělení s parametry $EU = 0$, $DU = 1$ a distribuční funkcí Φ, jejíž hodnoty jsou spočítány v tabulce.

otázka č. 20. Co říká centrální limitní věta?

Odpověď: Pokud $X_1, X_2, \dots X_N$ jsou nezávislé

stejně rozdělené veličiny (rozdělení psti může být libovolné, diskrétní či spojité, ale pro všech N veličin stejné) se střední hodnotou $EX_i = \mu$ a rozptylem $DX_i = \sigma^2$, pak veličina

$$Y := \sum_1^N X_i$$

má pro dostatečně velké N rozdělení normální se střední hodnotou $N \cdot \mu$ a rozptylem $N \cdot \sigma^2$.

otázka č. 21. Jak se počítají pravděpodobnosti normálního rozdělení pomocí tabulky distribuční funkce Φ? Například X má rozdělení normální pro $\mu = 5$ a $\sigma^2 = 16$ Vypočtěte pravděpodobnost,

že a) $X < 4$**; b)** $X > 8$**; c)** $P(X \in (3; x_1)) = 0{,}9$ – **určete hodnotu** x_1**.**

Literatura

[1] Fajmon, B., Růžičková, I.: Matematika 3. Skriptum FEKT 2003 (identifikační číslo v informačním systému VUT: MAT103). http://www.rozhovor.cz/souvislosti/matematika3.pdf.

[2] Hlavičková, I., Hliněná, D.: Sbírka úloh z pravděpodobnosti. Skriptum FEKT 2008.