7. cvičení z STP

28. března - 1. dubna 2022

Příklad 7.1 *Při numerickém výpočtu se reálná čísla zaokrouhlují na jedno desetinné místo. Jaká je pravděpodobnost, že vzdálenost skutečného čísla od zaokrouhleného bude větší než* 0.04*?*

Řešení:
Podobně jako v příkladu o házení mince na nekonečnou mřížku (**Příklad 2.2**) budeme předpokládat, že vstupní hodnoty (tj. čísla, která budeme zaokrouhlovat) pocházejí z nějakého referenčního intervalu $\Omega \subseteq \mathbb{R}$ délky 0.1 (zaokrouhlujeme na jedno desetinné místo), na kterém máme geometrickou pravděpodobnost, např. $\Omega = \langle -0.05,\ 0.05\rangle$.

Označme si teď náhodnou veličinu $X : \Omega \to \mathbb{R}$ jako

$$X = \text{ \textit{“skutečná hodnota” - “zaokrouhlená hodnota”}}$$

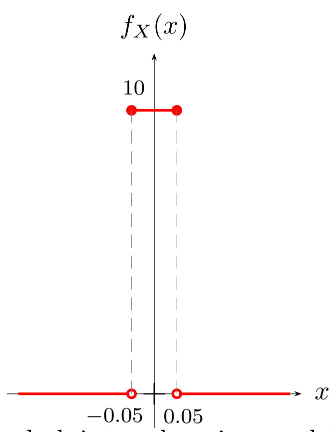
tj. pro $\omega \in \Omega = \langle -0.05,\ 0.05\rangle$ to funguje jako

$$X(\omega) = \begin{cases} \omega - 0 = \omega & \text{pro} \quad -0.05 \le \omega < 0.05, \\ 0.05 - 0.05 = 0 & \text{pro} \qquad \omega = 0.05\ . \end{cases}$$

A nás teď zajímá pravděpodobnost $P\big(|X| > 0.04\big)$. K tomu potřebujeme znát rozdělení veličiny X. Intuitivně tušíme, že X bude mít rovnoměrné rozdělení na své množině hodnot, tj. v intervalu $\langle -0.05,\ 0.05)$. Předpokládejme tedy, že $X \sim \mathrm{Ro}(-0.05,\ 0.05)$ (níže si to pak zdůvodníme). Veličina X má pak spojité rozdělení s hustotou pravděpodobnosti

$$f_X(x) = \begin{cases} \frac{1}{0.05-(-0.05)} = 10 & ,\ -0.05 \le x \le 0.05 \\ 0 & ,\ \text{jinak} \end{cases}$$

s grafem



(Hodnoty f_X v krajních bodech intervalu nejsou podstatné.)

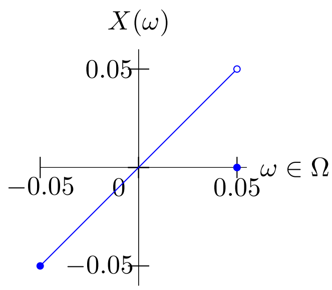
Hledaná pravděpodobnost je tudíž

$$P\Big(|X| > 0.04\Big) = 1 - P\Big(|X| \le 0.04\Big) = 1 - P(-0.04 \le X \le 0.04) =$$

$$= 1 - \int_{-0.04}^{0.04} 10\ dx = 1 - 10 \cdot 0.08 = 0.2\ .$$

- -

Výše zmíněné rovnoměrné rozdělení si ještě můžeme odvodit. Veličinu X si můžeme znázornit tímto grafem:



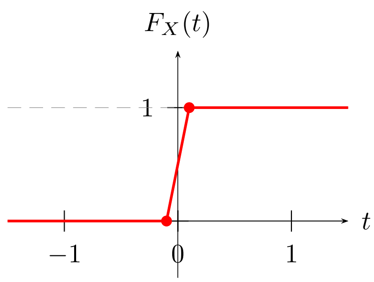
Odsud snadno vidíme, že pro $t \in \langle -0.05,\ 0.05)$ je

$$vol\Big(\{\omega \in \Omega \mid X(\omega) \leq t\}\Big) = vol\Big(\langle -0.05,\ t\rangle\Big) = t + 0.05$$

takže distribuční funkce je

$$F_X(t) = P(X \leq t) = \begin{cases} 0 & , t < -0.05 \\ \frac{vol\{\omega\in\Omega \mid X(\omega)\leq t\}}{vol(\Omega)} = \frac{t+0.05}{0.1} = 10t + 0.5 & , t \in \langle -0.05,\ 0.05) \\ 1 & , t \geq 0.05 \end{cases}$$

s grafem



Distribuční funkce F_X je tedy spojitá a má výše uvedenou hustotu.

Příklad 7.2 *Prodejna očekává dodávku nového zboží v určitý den buď dopoledne v době od* 8 *do* 10 *hodin nebo odpoledne od* 14 *do* 15 *hodin. Uskutečnění dodávky je stejně možné kdykoliv během těchto časových úseků. Veličina* X *bude představovat čas dodání zboží.*

(a) Určete rozdělení veličiny X *a její distribuční funkci* F_X.

(b) Jaká je pravděpodobnost, že zboží bude dodáno v době od 8:30 *do* 8:45*? Jaká bude pravděpodobnost, že zboží bude dodáno až po* 9:00*?*

(c) Průměrně v kolik hodin bude zboží dodáno?

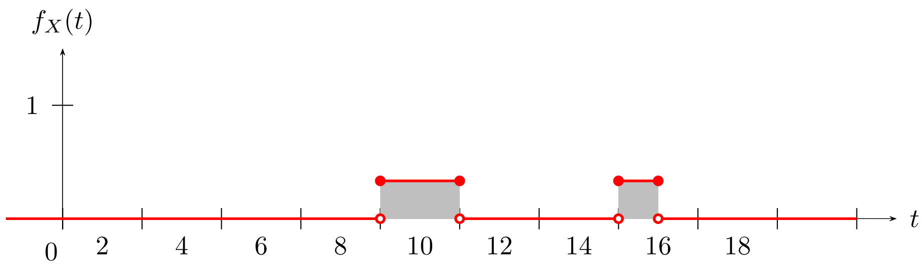
Řešení:
(a) Veličina X (měřená v hodinách) má rovnoměrné rozdělení na sjednocení intervalů $\langle 8, 10\rangle$ a $\langle 14, 15\rangle$. Její hustota f_X je tedy dána jako

$$f_X(t) = \begin{cases} c & , t \in \langle 8,\, 10\rangle \cup \langle 14,\, 15\rangle \\ 0 & , \text{jinak} \end{cases}$$

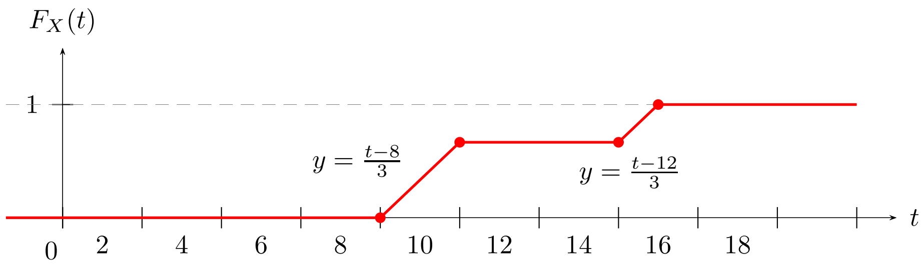
pro vhodné c, které určíme z podmínky, že (šedá) plocha pod hustotou je rovna 1:

$$1 = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(t)\ \mathrm{d}t = \int_8^{10} c\ \mathrm{d}t + \int_{14}^{15} c\ \mathrm{d}t = 3c$$

tedy $c = \frac{1}{3}$.



Funkci F_X nyní už snadno napočítáme pomocí integrování hustoty f_X (anebo ještě jednodušeji z obrázku hustoty: F_X poroste s lineárním přírůstkem o směrnici $\frac{1}{3}$ tam, kde je nenulová hustota, a jinde bude funkce F_X konstantní):



$$F_X(t) = \int_{-\infty}^{t} f_X(u)\ \mathrm{d}u = \begin{cases} \int_{-\infty}^{t} 0\ \mathrm{d}u = & 0 & , t \leq 8 \\ \int_8^t \frac{1}{3}\ \mathrm{d}u = & \frac{t-8}{3} & , t \in \langle 8,\, 10\rangle \\ \frac{2}{3} + \int_{10}^t 0\ \mathrm{d}u = & \frac{2}{3} & , t \in \langle 10,\, 14\rangle \\ \frac{2}{3} + \int_{14}^t \frac{1}{3}\ \mathrm{d}u = & \frac{t-12}{3} & , t \in \langle 14, 15\rangle \\ & 1 & , t \geq 15 \end{cases}$$

(b) Zajímají nás pravděpodobnosti $P\big(8.5 \leq X \leq 8.75\big)$ a $P\big(X \geq 9\big)$. Ty můžeme zjistit buď jednoduše z geometrické pravděpodobnosti (která je jen jiným vyjádřením rovnoměrného rozdělení):

$$P\big(8.5 \leq X \leq 8.75\big) = \frac{vol\big(\langle 8.5,\, 8.75\rangle\big)}{vol\big(\langle 8,\, 10\rangle \cup \langle 14,\, 15\rangle\big)} = \frac{0.25}{2+1} = \frac{1}{12} \doteq 0.083$$

$$P\big(X \geq 9\big) = \frac{vol\big(\langle 9,\, 10\rangle \cup \langle 14,\, 15\rangle\big)}{vol\big(\langle 8,\, 10\rangle \cup \langle 14,\, 15\rangle\big)} = \frac{1+1}{2+1} = \frac{2}{3} \doteq 0.667$$

nebo pomocí hustoty f_X a (ještě lépe) pomocí distribuční funkce F_X, kterou už máme určenou:

$$P\big(8.5 \leq X \leq 8.75\big) = P(X \leq 8.75) - P(X < 8.5) \overset{\text{spoj. rozd.}}{=} P(X \leq 8.75) - P(X \leq 8.5) =$$

$$= F_X(8.75) - F_X(8.5) = \frac{8.75-8}{3} - \frac{8.5-8}{3} = \frac{1}{12}$$

$$P\big(X \geq 9\big) = 1 - P(X < 9) \overset{\text{spoj. rozd.}}{=} 1 - P(X \leq 9) = 1 - \frac{9-8}{3} = \frac{2}{3}\,.$$

(c) Zde chceme zjistit $E(X)$. Máme

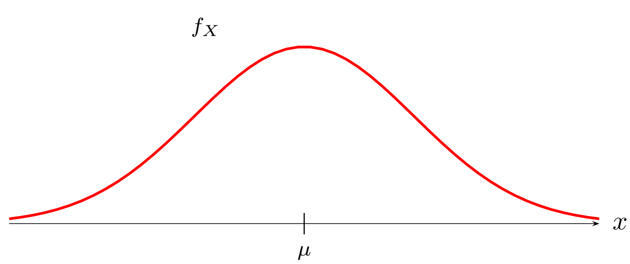
$$E\big(X\big) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ \mathrm{d}t = \int_8^{10} t \cdot \tfrac{1}{3}\ \mathrm{d}t + \int_{14}^{15} t \cdot \tfrac{1}{3}\ \mathrm{d}t =$$

$$= \Big[\tfrac{t^2}{6}\Big]_8^{10} + \Big[\tfrac{t^2}{6}\Big]_{14}^{15} = \frac{100 - 64 + 225 - 196}{6} = \frac{65}{6} = 10\,\mathrm{hod}\,50\,\mathrm{min}\,.$$

Průměrná doba, kdy bude zboží dodáno tedy je v 10:50. To se může zdát na první pohled divné, když v tento čas přece zboží vůbec nemůže být (fyzicky) dodáno, ale tento výsledek jen vyjadřuje, že střední čas dodání bude blíž k dopolednímu rozmezí než k odpolednímu (a jak moc blízko to bude). Podobně, u alternativního rozdělení $\mathrm{Alt}(p)$ je střední hodnota také rovna $p \in (0,1)$, přestože hodnoty alternativního rozdělení jsou jen 0 a 1.

Poznámky k normálnímu rozdělení:
Veličina X má *normální* (neboli Gaussovo) rozdělení $N(\mu, \sigma^2)$ (kde $\mu \in \mathbb{R}$ a $\sigma > 0$), jestliže má hustotu

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{pro} \quad x \in \mathbb{R}\,.$$



Je to tedy spojité rozdělení, $E(X) = \mu$, $\mathrm{var}(X) = \sigma^2$ a oborem hodnot veličiny X je celá reálná osa. Všimněme si ještě, že hustota f_X je symetrická vzhledem ke středu μ a proto platí $F_X(\mu) = \frac{1}{2}$.

Toto rozdělení je limitním rozdělením, které aproximuje součty nezávislých stejně (nebo podobně) rozdělených veličin (více později v Centrální limitní větě). Typicky se tedy objevuje u veličin, jejichž hodnoty jsou ovlivněny mnoha drobnými odchylkami (např. u chyb měření, výšky člověka apod.)

U zmíněné výšky člověka (která může být samozřejmě jen kladná) nebo u veličin s hodnotami omezenými na nějaký interval, je přesto použití normálního rozdělení (které může nabývat libovolných hodnot) přiměřené.

Je to tím, že u dané veličiny Y předpokládáme aproximaci pomocí normálního rozdělení obvykle jen ve vhodném okolí kolem střední hodnoty $\mu := E(Y)$. Je to podobná situace, jako když aproximujeme funkci pomocí jejího Taylorova polynomu v okolí daného bodu.

Přesněji to vystihuje toto tvrzení:

Věta: Nechť Y je veličina s hustotou f_Y, střední hodnotou μ a rozptylem $\sigma^2 \neq 0$. Nechť $X \sim N(\mu, \sigma^2)$. Jestliže se hustoty f_X a f_Y rovnají na nějakém intervalu $(a, b) \subseteq \mathbb{R}$ takovém, že $\mu \in (a, b)$ a pokud $F_Y(\mu) = \frac{1}{2}$, pak

$$F_Y(t) = F_X(t) \quad \text{pro všechna} \;\; t \in (a, b) \; .$$

Značení: Pro náhodnou veličinu X s konečnou střední hodnotou a konečným rozptylem, položme

$$\operatorname{norm}(X) := \frac{X - E(X)}{\sqrt{\operatorname{var}(X)}} \; .$$

Speciálně tedy vidíme, že $E\big(\operatorname{norm}(X)\big) = 0$ a $\operatorname{var}\big(\operatorname{norm}(X)\big) = 1$.

Platí: Pro takovouto veličinu X a konstanty $a > 0$ a $b \in \mathbb{R}$ je

$$\operatorname{norm}(aX + b) = \operatorname{norm}(X) \; .$$

Důležité vlastnosti normálního rozdělení:

- $X \sim N(\mu, \sigma^2) \;\Rightarrow\; \operatorname{norm}(X) = \frac{X-\mu}{\sigma} \sim N(0, 1)$ (je to tzv. normované normální rozdělení s hodnotami v tabulkách) dist. funkce pro $N(0, 1)$ se značí Φ.

  V tomto případě pak máme $F_X(t) = P(X \leq t) = P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{=\operatorname{norm}(X)} \leq \frac{t-\mu}{\sigma}\Big) = \Phi\Big(\frac{t-\mu}{\sigma}\Big)$ pro všechna $t \in \mathbb{R}$.
- hustota $f_{N(0,1)}$ je sudá funkce $\;\Rightarrow\;$ $\Phi(t) + \Phi(-t) = 1$ pro všechna $t \in \mathbb{R}$.
- Nechť $X_i \sim N(\mu_i, \sigma_i^2)$, pro $i = 1, 2$, jsou nezávislé. Pak $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (tj. speciálně součet nezávislých normálních rozdělení je zase normální.)

Pro vybraná čísla $t \geq 0$ se dají hodnoty Φ najít ve statistických tabulkách. Pro záporná čísla si pak pomůžeme vztahem $\Phi(-t) = 1 - \Phi(t)$.

Pro lepší představu o tom, jakou roli pro veličinu s normálním rozdělením $X \sim N(\mu, \sigma^2)$ hraje směrodatná odchylka σ se používá tzv.

pravidlo tří-sigma (`https://cs.wikipedia.org/wiki/Pravidlo_t\%C5\%99\%C3\%AD\_sigma`)

které je ovšem čistě jen technickou pomůckou:

Jestliže si budeme počítat pravděpodobnosti

$$P(|X - \mu| \leq k \cdot \sigma) = P\Big(\Big|\underbrace{\frac{X-\mu}{\sigma}}_{\sim N(0,1)}\Big| \leq k\Big) = \Phi(k) - \Phi(-k) = 2 \cdot \Phi(k) - 1 \quad \text{pro} \quad k = 1, 2, 3, \ldots$$

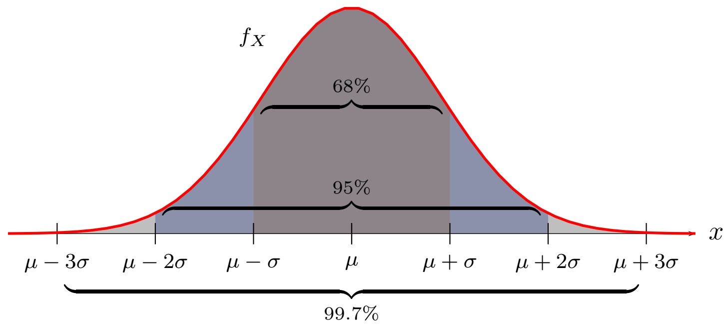
dostaneme postupně

$$P(|X - \mu| \leq \sigma) = 2 \cdot \Phi(1) - 1 \doteq 2 \cdot 0.8413 - 1 = 0.6826 \doteq 68\%$$

$$P(|X - \mu| \leq 2 \cdot \sigma) = 2 \cdot \Phi(2) - 1 \doteq 2 \cdot 0.9772 - 1 = 0.9544 \doteq 95\%$$

$$P(|X - \mu| \leq 3 \cdot \sigma) = 2 \cdot \Phi(3) - 1 \doteq 2 \cdot 0.99865 - 1 = 0.9973 \doteq 99.7\%$$

Pro vyšší hodnoty, tj. $k \geq 4$ už jsou pravděpodobnosti vpodstatě rovny 1, takže se v praxi příliš nepoužívají (záleží samozřejmě na zvolené přesnosti).



Příklad 7.3 *Výška dětí v 1. třídě je náhodná veličina $X \sim N(130\,\text{cm},\ 36\,\text{cm}^2)$. Jaká je pravděpodobnost, že náhodně vybrané dítě bude*

(a) větší než 136 cm,

(b) menší než 118 cm,

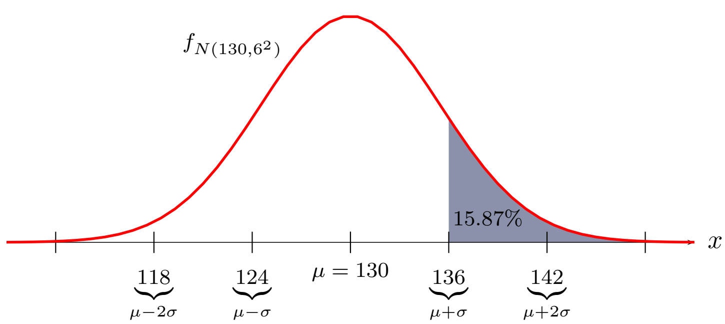
(c) mít výšku mezi 127 a 133 cm?

Řešení:
Nyní tedy máme $X \sim N(130,\ 36)$. Pro jednodušší zápis si ještě označme $Z := \text{norm}(X)$.

(a)

$$P(X > 136) = P\left(\underbrace{\frac{X-130}{\sqrt{36}}}_{Z} > \underbrace{\frac{136-130}{\sqrt{36}}}_{1}\right) = P(Z > 1) = 1 - P(Z \leq 1) =$$

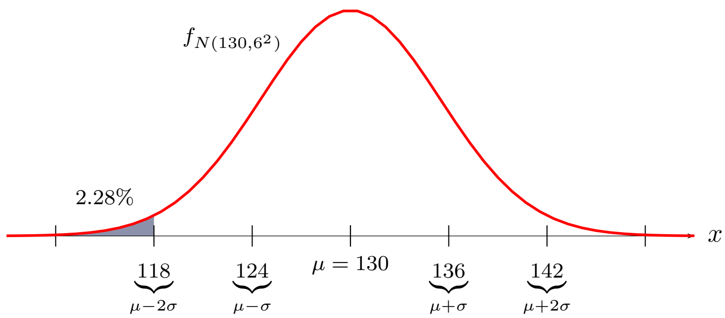
$$= 1 - \Phi(1) \doteq 1 - 0.8413 = 0.1587\ .$$



(b)

$$P(X < 118) = P\left(\frac{X-130}{\sqrt{36}} < \frac{118-130}{\sqrt{36}}\right) = P(Z < -2) = \Phi(-2) =$$

$$= 1 - \Phi(2) \doteq 1 - 0.9772 = 0.0228\ .$$



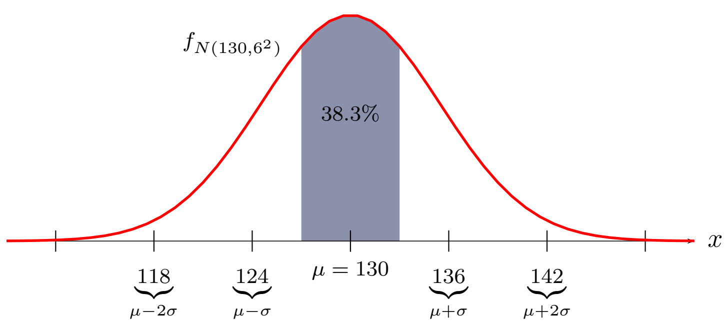
(c)

$$P(127 < X < 133) = P\left(\frac{127-130}{\sqrt{36}} < \frac{X-130}{\sqrt{36}} < \frac{133-130}{\sqrt{36}}\right) =$$

$$= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z \leq -0.5) =$$

$$= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - \Big(1 - \Phi(0.5)\Big) =$$

$$= 2 \cdot \Phi(0.5) - 1 \doteq 2 \cdot 0.6915 - 1 = 0.383\ .$$



Poznamenejme ještě, že hodnoty výšek, které nás zajímaly (tj. 136 cm, 118 cm atd.) se pohybují celkem blízko střední hodnoty $E(X) = 130$ cm, takže předpoklad o normálnosti rozdělení X byl přiměřený.

- -

Výpočty si můžeme i urychlit přímým vzorcem $F_X(t) = \Phi(\frac{t-\mu}{\sigma})$, kde $\mu = 130$ cm a $\sigma = 6$ cm:

(a) $P(X > 136) = 1 - F_X(136) = 1 - \Phi(\frac{136-130}{6}) = \cdots \doteq 0.1587$

(b) $P(X < 118) = F_X(118) = \Phi(\frac{118-130}{6}) = \cdots \doteq 0.0228$

(c) $P(127 < X < 133) = F_X(133) - F_X(127) = \Phi(\frac{133-130}{6}) - \Phi(\frac{127-130}{6}) = \cdots \doteq 0.383$

Příklad 7.4 *Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě* 67 *m a* 75 *m a směrodatné odchylky* 6 *m a* 3 *m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.*

Řešení:
Náhodná veličina

$A =$"*délka hodu Anny*"

má rozdělení $N(67, 6^2)$ a veličina

$B =$"*délka hodu Barbory*"

má rozdělení $N(75, 3^2)$.

Zajímá nás $P(A > B) = P(A-B > 0)$. Protože veličiny A a B jsou nezávislé, tak veličina $Z := A-B$ má také normální rozdělení, a sice

$$Z \sim N(67 - 75, 6^2 + 3^2) = N(-8, 45)\ .$$

Takže

$$P(A > B) = P(Z > 0) = P\left(\underbrace{\frac{Z-(-8)}{\sqrt{45}}}_{\mathrm{norm}(Z)} > \frac{0-(-8)}{\sqrt{45}}\right) = 1 - P\left(\mathrm{norm}(Z) \leq \frac{8}{\sqrt{45}}\right) =$$

$$= 1 - \Phi\left(\tfrac{8}{\sqrt{45}}\right) \doteq 1 - \Phi\,(1.1926) \doteq 1 - 0.883 = 0.117\,.$$

POZOR! Zatímco střední hodnota je lineární zobrazení, tak rozptyl se chová jinak! Konkrétně je to takto:
Nechť X a Y jsou veličiny se střední hodnotou a konečným rozptylem. Pak

- $E(X \pm Y) = E(X) \pm E(Y)$
- $\mathrm{var}(X \pm Y) = \mathrm{var}(X) + \mathrm{var}(Y) \pm 2 \cdot \mathrm{cov}(X, Y)\ (\geq 0)$

Zde $\mathrm{cov}(X, Y)$ je tzv. kovariance (viz poznámky níže). Speciálně, pokud X a Y jsou nezávislé, je $\mathrm{cov}(X, Y) = 0$. Máme tedy:

- X a Y nezávislé $\Rightarrow$ $\mathrm{var}(X \pm Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

Tedy v tomto případě se rozptyly VŽDY sčítají!