

**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Anna Švarcová

**Limitní věty pro závislé náhodné  
veličiny**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D.

Studijní program: Obecná matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Tímto bych chtěla především poděkovat vedoucímu mé práce doc. RNDr. Danielovi Hlubinkovi, Ph.D. za velmi užitečné rady, ochotu, vstřícnost a skvělou spolupráci. Dále bych chtěla poděkovat své rodině a přátelům, že byli příčinou mého smíchu. Zvláště mé díky patří Vendule Rusé za hodiny, které se mnou trávila v knihovně, a za rady, které mi poskytla.

Název práce: Limitní věty pro závislé náhodné veličiny

Autor: Anna Švarcová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V předložené práci se zabýváme centrální limitní větou pro závislé náhodné veličiny. Nejdříve si zopakujeme základní znění věty a ilustrujeme si ji na příkladu jejího využití. Poté zavedeme definici strong mixing condition, jež nám umožní dokázat centrální limitní větu i pro závislé náhodné veličiny. Dále se zaměříme na předpoklady, které jsou nutné k platnosti věty. Většinu práce se budeme zabývat jejím důkazem. Na závěr si předvedeme zmíněnou větu na příkladu, který nám pomůže lépe pochopit hlavní myšlenku důkazu. Daný příklad nasimulujeme pro konkrétní hodnoty posloupností zdefinovaných ve znění věty.

Klíčová slova: centrální limitní věta, závislé náhodné veličiny, strong mixing condition

Title: Limit theorems for dependent random variables

Author: Anna Švarcová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Daniel Hlubinka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In presented work we will discuss the central limit theorem for dependent random variables. First of all, we brush up the basic version of the theorem and we illustrate it by an example of its use. Then we introduce a definition of the strong mixing condition that allows us to prove the theorem even for dependent random variables. Next, we focus on the assumptions which are essential for the validity of the theorem. The biggest part of the work we deal with its proof. Last of all, we illustrate this theorem with an example which helps us to better understand the main idea of the proof. We simulate this example for specific value of sequences that we define in the wording of the theorem.

Keywords: central limit theorem, dependent random variables, strong mixing condition

# Obsah

Úvod	2
<b>1 Centrální limitní věta</b>	<b>3</b>
1.1 Znění CLV . . . . .	3
1.2 Příklad . . . . .	3
<b>2 CLV a závislé náhodné veličiny</b>	<b>5</b>
2.1 Strong mixing condition . . . . .	5
2.2 CLV a strong mixing condition . . . . .	6
2.3 Příklad . . . . .	15
2.3.1 Simulace příkladu . . . . .	18
Závěr	24
Seznam použité literatury	25
Seznam obrázků	26

# Úvod

V této bakalářské práci se budeme zabývat centrální limitní větou pro závislé náhodné veličiny. Standardní centrální limitní věta je jeden ze základních nástrojů jak pravděpodobnosti tak i statistiky, pro niž je nezávislost jeden z hlavních předpokladů. Nicméně my si zde pomocí tzv. *strong mixing condition* dokážeme, že není nutnou podmínkou pro platnost této věty. Základním podkladem práce je článek Rosenblatt (1956), ve kterém je daná problematika skromně popsána. Tudíž hlavním cílem práce je ji detailněji zpracovat a platnost věty důkladně dokázat.

V první kapitole si připomeneme standardní centrální limitní větu pro nezávislé náhodné veličiny, která je nám již známa ze základního kurzu pravděpodobnosti a matematické statistiky. Pro lepší pochopení si zde zmíníme jeden motivační příklad. Jedno z jejího využití je určení intervalového odhadu pro námi daný parametr. Ve statistice se též hojně využívá pro určení rozdělení testové statistiky a následného kritického oboru.

Co se druhé kapitoly týče, tvoří většinu práce. V té formálně zavedeme značení, která jsou potřeba k zadefinování *strong mixing condition*, načež ji dáme do souvislosti s důkazem centrální limitní věty pro závislé náhodné veličiny. Poté se seznámíme s podrobným důkazem této věty, která má důležité předpoklady na strukturu závislosti. Výsledkem této kapitoly je tudíž tvrzení, že asymptotickou normalitu lze dokázat i pro náhodné veličiny, které jsou závislé.

V poslední podkapitole si dané výsledky představíme na příkladu tzv. *klouza-vých průměrů*. Ten nám přiblíží právě myšlenku dokazované věty. V poslední části závěrečné kapitoly nalezneme simulaci výše zmíněného příkladu pro konkrétní hodnoty posloupností představených v průběhu důkazu. Získané poznatky dále porovnáme.

# 1. Centrální limitní věta

Nejdříve si zde připomene definici konvergence v distribuci pro náhodné veličiny, která je základní pro znění centrální limitní věty. Následnou definici a znění věty nalezneme ve skriptech Prokešová (2023).

**Definice 1** (konvergence v distribuci). *Bud'  $\{Y_n\}_{n=1}^{\infty}$  posloupnost náhodných veličin na libovolných pravděpodobnostních prostorech a bud'  $Y$  také náhodná veličina. Řekneme, že náhodné veličiny  $Y_n$  konvergují v distribuci k náhodné veličině  $Y$ , pokud*

$$\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x)$$

*v každém bodě  $x \in \mathbb{R}$ , ve kterém je  $F_Y$ , distribuční funkce náhodné veličiny  $Y$ , spojitá. Značíme  $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ .*

## 1.1 Znění CLV

**Věta 1** (Centrální limitní věta). *Budte  $X_1, X_2, \dots$  vzájemně nezávislé, stejně rozdělené náhodné veličiny,  $\mathbf{E}X_i = \mu \in \mathbb{R}$ ,  $\text{var} X_i = \sigma^2 \in (0, \infty)$ ,  $\forall i \in \mathbb{N}$ . Potom pro náhodné veličiny*

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n\sigma^2}}$$

*platí  $Z_n \xrightarrow[n \rightarrow \infty]{d} Z$ , kde  $Z \sim \mathcal{N}(0,1)$ .*

Věta nám vlastně říká, že součet rozdílů náhodných veličin a jejich středních hodnot, který je vhodně podělen, má asymptoticky normované normální rozdělení.

Ekvivalentně můžeme náhodné veličiny  $Z_n$  zapsat jako

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

kde  $\bar{X}_n$  značí výběrový průměr náhodných veličin  $X_1, X_2, X_3, X_4, \dots$ , tzn.  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Tedy platí:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0,1).$$

## 1.2 Příklad

Pro lepší představu si zmíníme jeden příklad na využití centrální limitní věty.

Mějme náhodné veličiny  $X_1, X_2, \dots \sim \text{Alt}(p)$  s parametrem  $p \in (0,1)$ , které jsou nezávislé a stejně rozdělené. Tedy se jedná o rozdělení, pro něž platí

$$P(X_i = 1) = p \quad \text{a} \quad P(X_i = 0) = 1 - p.$$

Střední hodnota je  $\mathbf{E}X_i = p$  a rozptyl  $\text{var} X_i = p(1 - p)$  pro  $\forall i \in \mathbb{N}$ .

Víme, že  $\bar{X}_n$  je nestranný a konzistentní odhad parametru  $p$ . Máme tedy z definice konzistentního odhadu

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} p,$$

tudíž výběrový průměr konverguje v pravděpodobnosti k parametru  $p$ .

Využitím věty o spojité transformaci s funkcí  $g(x) = x(1 - x)$  získáme

$$\bar{X}_n(1 - \bar{X}_n) \xrightarrow[n \rightarrow \infty]{P} p(1 - p).$$

Z poznatku výše plyne opět pomocí věty o spojité transformaci následující:

$$\sqrt{\frac{p(1-p)}{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n \rightarrow \infty]{P} 1.$$

Z centrální limitní věty dostáváme

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0,1).$$

Platí tedy

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \cdot \sqrt{\frac{p(1-p)}{\bar{X}_n(1-\bar{X}_n)}} = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0,1)$$

z Cramer-Sluckého věty. Výsledek můžeme dále využít ke zkonstruování následujícího asymptotického intervalového odhadu střední hodnoty

$$\left( \bar{X}_n \mp u_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} \right),$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$ -tý kvantil normovaného normálního rozdělení.



## 2. CLV a závislé náhodné veličiny

Z minulé kapitoly víme, že důležitými předpoklady standardní centrální limitní věty jsou nezávislost a stejné rozdělení náhodných veličin. My si však v té následující ukážeme, že nezávislost není nutnou podmínkou pro platnost CLV a že asymptotickou normalitu lze dokázat i pro závislé náhodné veličiny. Avšak důležité jsou určité předpoklady na strukturu závislosti. Pro důkaz využijeme tzv. *strong mixing condition*. Proto si nejdříve zavedeme potřebné značení a poté zadefinujeme danou podmínku.

V této kapitole budeme vycházet především z článku Rosenblatt (1956), který si zde rozebereme více do hloubky. Na závěr si uvedenou teorii předvedeme na příkladu tzv. *klouzavých průměrů*.

### 2.1 Strong mixing condition

Teorii o strong mixing condition budeme čerpat z Bradley (2005) a Wikipedia contributors (2023). Mějme posloupnost náhodných veličin  $X = (X_k, k \in \mathbb{Z})$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, \mathbb{P})$ . Zavedeme značení

$$\mathcal{F}_a^b = \sigma(X_k, a \leq k \leq b),$$

kde  $-\infty \leq a \leq b \leq \infty$ . Jedná se tedy o  $\sigma$ -algebru, která je generovaná náhodnými veličinami  $X_a, X_{a+1}, \dots, X_{b-1}, X_b$ .

Dále si označíme funkci

$$\alpha(s) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A) \cdot \mathbb{P}(B)| : -\infty < t < \infty, A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+s}^{\infty}\},$$

kde  $\mathcal{F}_{-\infty}^t = \sigma(\dots, X_{t-1}, X_t)$  a  $\mathcal{F}_{t+s}^{\infty} = \sigma(X_{t+s}, X_{t+s+1}, \dots)$ .

O posloupnosti náhodných veličin  $X = (X_k, k \in \mathbb{Z})$  řekneme, že splňuje *strong mixing condition*, pokud

$$\alpha(s) \rightarrow 0 \quad \text{pro } s \rightarrow \infty.$$

Takovou posloupnost nazýváme *strong mixing*.

Nyní se omezíme na indexy  $k \in \mathbb{N}$ . Pro obecné indexy bychom místo  $1, 2, \dots, i$  používali  $\dots, i-2, i-1, i$ . Ekvivalentně posloupnost náhodných veličin splňuje strong mixing condition, pokud

$$\left| \mathbb{P}[(X_1, \dots, X_i) \in A \cap (X_{i+m}, X_{i+m+1}, \dots) \in B] - \right. \quad (2.1)$$

$$\left. - \mathbb{P}[(X_1, \dots, X_i) \in A] \cdot \mathbb{P}[(X_{i+m}, X_{i+m+1}, \dots) \in B] \right| \leq f(m)$$

pro všechna  $i, m \in \mathbb{N}$ , množiny  $A, B$  a funkci  $f(m) \rightarrow 0$  pro  $m \rightarrow \infty$ .

Můžeme si ji představit následujícím způsobem. Pokud vzdálenost mezi množinami indexů je větší, pak rozdíl pravděpodobnosti průniku těchto množin a součin jejich pravděpodobnosti se blíží k nule. Musíme podotknout, že v našem případě vzdáleností máme na mysli rozdíl indexů náhodných veličin. V případě (2.1) tedy vzdálenost mezi množinami indexů  $A$  a  $B$  je  $m$ . Tuto variantu podmínky můžeme nalézt v LEHMANN (1999).

## 2.2 CLV a strong mixing condition

**Věta 2.** *Mějme náhodné veličiny  $X_1, X_2, \dots$  a předpokládejme, že jsou centrované (mají nulovou střední hodnotu). Necht' dále jsou splněné následující předpoklady:*

$$\mathbb{E}\left(\sum_{j=a}^b X_j\right)^2 \sim h(b-a) \quad (2.2)$$

a pro každé  $\delta > 0$

$$\mathbb{E}\left(\sum_{j=a}^b X_j\right)^{2+\delta} = o\left(h(b-a)^{1+\frac{\delta}{2}}\right) \quad (2.3)$$

pro  $b-a \rightarrow \infty$ , kde  $h$  je neklesající funkce  $m$  a  $h(m) \rightarrow \infty$  pro  $m \rightarrow \infty$ .

Dále necht' je splněna strong mixing condition. Pak pro posloupnost náhodných veličin  $S_1, S_2, \dots$ , které jsou zdefinované jako  $S_n = \sum_{j=1}^n X_j$ , platí:

$$\frac{S_n}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{d} Z,$$

kde  $k, p_n$  a  $q_n$  jsou posloupnosti přirozených čísel splňující  $k(p_n + q_n) = n$ ,  $p_n, q_n, k \rightarrow \infty, \frac{q_n}{p_n} \rightarrow 0$  pro  $n \rightarrow \infty$  a  $Z$  má normované normální rozdělení.

Symbolem  $\sim$  máme na mysli, že podíl pravé a levé strany konverguje k jedné. Toto značení budeme využívat ve zbytku práce. Nyní se podíváme na náznak důkazu. Nejdříve si zavedeme následující značení:

$$\begin{aligned} S_n &= \sum_{j=1}^n X_j, \\ U_r &= \sum_{j=(r-1)(p_n+q_n)+1}^{rp_n+(r-1)q_n} X_j, \\ V_r &= \sum_{j=rp_n+(r-1)q_n+1}^{r(p_n+q_n)} X_j, \end{aligned}$$

kde  $r = 1, \dots, k$ .

Mějme tedy:

$r = 1$ :

$$U_1 = \sum_{j=1}^{p_n} X_j \quad \text{a} \quad V_1 = \sum_{j=p_n+1}^{p_n+q_n} X_j,$$

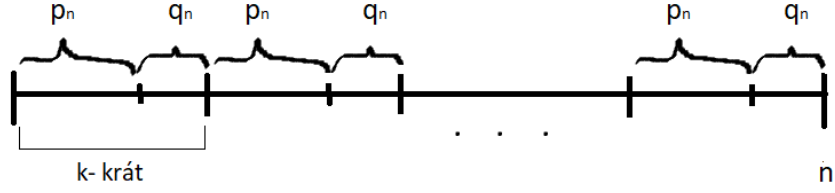
$r = 2$ :

$$U_2 = \sum_{j=p_n+q_n+1}^{2p_n+q_n} X_j \quad \text{a} \quad V_2 = \sum_{j=2p_n+q_n+1}^{2(p_n+q_n)} X_j,$$

$\vdots$

$r = k$ :

$$U_k = \sum_{j=(k-1)(p_n+q_n)+1}^{kp_n+(k-1)q_n} X_j \quad \text{a} \quad V_k = \sum_{j=kp_n+(k-1)q_n+1}^{k(p_n+q_n)} X_j.$$



Obrázek 2.1: Rozepsání součtu náhodných veličin po částech.

Jedná se tedy rozepsání součtu  $S_n = \sum_{j=1}^n X_j$  na jednotlivé části, které jsou zavedeny ve znění výše. Množinu indexů  $1, \dots, n$  rozdělíme na  $k$  disjunktních množin po sobě jdoucích indexů, které mají velikost  $(p_n + q_n)$  (viz obrázek 2.1). Součty prvních  $p_n$  indexů v každém podintervalu označíme  $U_r$  a součty přes zbylé indexy značíme  $V_r$  pro  $r = 1, \dots, k$ . Všimněme si, že části  $q_n$  obsahují asymptoticky zanedbatelné množství indexů oproti  $p_n$ . Dále platí

$$\frac{S_n}{\sqrt{kh(p_n)}} = \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} + \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}}.$$

Budeme tedy dokazovat konvergenci v distribuci pro

$$\sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} + \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{d} Z. \quad (2.4)$$

V první části důkazu ukážeme, že

$$\sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{P} 0.$$

V té druhé předvedeme platnost

$$\sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{d} Z.$$

Tedy z Cramer-Sluckého věty dostaneme (2.4).

*Důkaz.* V důkazu budeme využívat značení, která jsme si zavedli výše. Nejdříve si ukážeme, že

$$\sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{P} 0. \quad (2.5)$$

$$\mathbb{E} \left| \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \right| = \mathbb{E} \left| \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \right|^2 \leq \sqrt{\mathbb{E} \left| \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \right|^2} \leq \sum_{r=1}^k \frac{\sqrt{\mathbb{E} |V_r|^2}}{\sqrt{kh(p_n)}},$$

kde první nerovnost plyne z Jensenovy nerovnosti pro funkci  $g(x) = \sqrt{x}$ . Zároveň poslední získáváme z Minkowského nerovnosti aplikovanou na střední hodnotu.

Dále máme z vlastnosti  $V_r$  a (2.2)

$$\sum_{r=1}^k \frac{\sqrt{\mathbb{E}|V_r|^2}}{\sqrt{kh(p_n)}} \sim \sum_{r=1}^k \frac{\sqrt{h(q_n)}}{\sqrt{kh(p_n)}} \sim \sqrt{\frac{kh(q_n)}{h(p_n)}}.$$

Pokud tedy vybereme posloupnosti  $k, q_n, p_n \rightarrow \infty$  pro  $n \rightarrow \infty$ , které také splňují

$$\frac{kh(q_n)}{h(p_n)} \rightarrow 0, \quad (2.6)$$

dostáváme

$$\mathbb{E} \left| \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \right| \xrightarrow{n \rightarrow \infty} 0.$$

Konvergenci v pravděpodobnosti budeme dokazovat z definice. Označme si  $Y = \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}}$  a necht'  $\varepsilon > 0$ .

$$\mathbb{P}(|Y| > \varepsilon) = \mathbb{E}(1[|Y| > \varepsilon]) \leq \frac{\mathbb{E}|Y|}{\varepsilon} \sim \frac{kh(q_n)}{h(p_n)} \frac{1}{\varepsilon} \xrightarrow{n \rightarrow \infty} 0.$$

Tedy platí (2.5).

Nyní si zdefinujeme následující značení, která nám budou usnadňovat zápis:

$$G_{r,n}(x) = \mathbb{P} \left( \frac{U_r}{\sqrt{kh(p_n)}} \leq x \right),$$

$$A(r, n, m_r, \delta) = \left\{ m_r \delta < \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1) \delta \right\},$$

kde  $m_r \in \mathbb{Z}$  a  $\delta > 0$ . Funkce  $G_{r,n}(x)$  značí distribuční funkci náhodné veličiny  $\frac{U_r}{\sqrt{kh(p_n)}}$ .

Poznamenejme, že

$$\sum_{(m_1 + \dots + m_k + k) \delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) \leq \mathbb{P} \left( \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \leq x \right) \quad (2.7)$$

$$\leq \sum_{(m_1 + \dots + m_k) \delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right). \quad (2.8)$$

Ukážeme si pouze platnost nerovnosti (2.7). Důkaz (2.8) je obdobný.

$$\begin{aligned} & \sum_{(m_1 + \dots + m_k + k) \delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) \\ &= \sum_{(m_1 + \dots + m_k + k) \delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k \left\{ m_r \delta < \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1) \delta \right\} \right) \\ &= \mathbb{P} \left( \bigcup_{(m_1 + \dots + m_k + k) \delta \leq x} \bigcap_{r=1}^k \left\{ m_r \delta < \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1) \delta \right\} \right) \end{aligned}$$

$$\leq \mathbf{P} \left( \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \leq x \right).$$

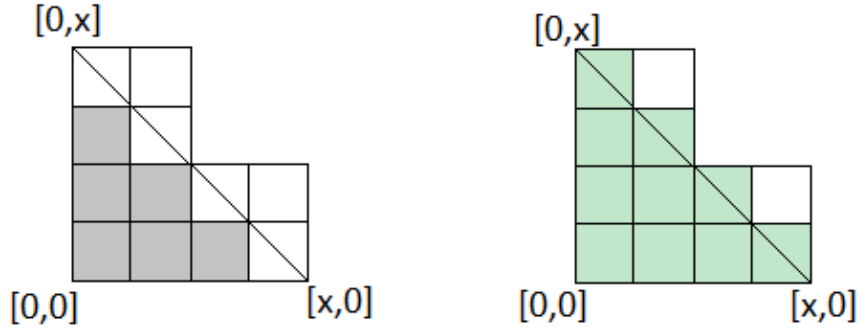
Suma

$$\sum_{(m_1+\dots+m_k+k)\delta \leq x} \mathbf{P} \left( \bigcap_{r=1}^k \left\{ \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1)\delta \right\} \right)$$

nám říká, že

$$\forall r \in \{1, \dots, k\} \quad \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1)\delta \quad \text{pro} \quad \sum_{r=1}^k (m_r + 1)\delta \leq x.$$

Zároveň pro všechny možnosti výběru  $m_1, m_2, \dots, m_k$  jsou  $\bigcap_{r=1}^k A(r, n, m_r, \delta)$  disjunktní. Tedy nám platí rovnosti a nerovnosti výše. Nyní si ukážeme, proč by měla druhá nerovnost (2.8) platit a celkovou myšlenku dokazovaných nerovností. V případě, že bychom měli pouze dva sčítance, dostáváme následující obrázek.



Obrázek 2.2: Obsah trojúhelníku  $[0,0], [x,0], [0,x]$  využit při nerovnostech.

Jedná se o obsah trojúhelníku  $[0,0], [x,0], [0,x]$ , kde jednotlivé čtverce reprezentují  $\frac{U_r}{\sqrt{kh(p_n)}}$  pro jednotlivé možnosti výběru  $m_1, m_2, \dots, m_k$  (je zde vidět, že jsou disjunktní). Diagonála nám určuje hodnotu hranice  $x$ . Na obrázku vlevo je znázorněná právě již dokázaná první nerovnost. Počítáme tedy s celými čtverci, které jsou obsaženy v trojúhelníku.

Co se týče pravého, ten nám přibližuje případ zbylé nerovnosti (2.8). Z platnosti

$$\begin{aligned} & \sum_{(m_1+\dots+m_k)\delta \leq x} \mathbf{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) \\ &= \mathbf{P} \left( \bigcup_{(m_1+\dots+m_k)\delta \leq x} \bigcap_{r=1}^k \left\{ m_r \delta < \frac{U_r}{\sqrt{kh(p_n)}} \leq (m_r + 1)\delta \right\} \right) \end{aligned}$$

můžeme vidět, že

$$(m_1 + \dots + m_k)\delta \leq x \quad \implies \exists \quad m_1, m_2, \dots, m_k : (m_1 + \dots + m_k + 1)\delta > x.$$

Tedy sčítáme-li přes všechny  $k$ -tice  $m_1, m_2, \dots, m_k$ , které mají vlastnost výše, pak mezi nimi nalezneme takové, že  $(m_1 + \dots + m_k + 1)\delta > x$ . To jsou právě ty čtverce na druhém obrázku, které jsou trojúhelníkem  $[0,0], [x,0], [0,x]$  zasažené.

**Lemma 3.** Pro značení, která jsme si zavedli výše, platí

$$\mathbb{P} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| > t_k \right) \leq \varepsilon, \quad (2.9)$$

pokud  $t_k = \sqrt{\frac{k}{\varepsilon}}$  a pro dostatečně velká  $n$ .

*Důkaz.* Necht  $\varepsilon > 0$ . K důkazu využijeme Čebyševovu nerovnost. Z ní vyplývá následující:

$$\begin{aligned} \mathbb{P} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| > t_k \right) &= \mathbb{P} \left( \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 > t_k^2 \right) \\ &\leq \mathbb{E} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \frac{\varepsilon}{k}. \end{aligned} \quad (2.10)$$

Dále nám platí, že

$$\mathbb{E} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \leq \mathbb{E} \left[ \left( \sum_{r=1}^k \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \right].$$

Nerovnost získáváme, jelikož maximum přes nezáporné hodnoty je menší nebo rovno než součet těchto hodnot. Poté máme

$$\begin{aligned} &\sqrt{\mathbb{E} \left[ \left( \sum_{r=1}^k \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \right]} \leq \sum_{r=1}^k \frac{\sqrt{\mathbb{E} |U_r|^2}}{\sqrt{kh(p_n)}} \\ &= \sum_{r=1}^k \frac{\sqrt{h(p_n)(1 + o_n(1))}}{\sqrt{kh(p_n)}} = \sum_{r=1}^k \frac{\sqrt{(1 + o_n(1))}}{\sqrt{k}} = \sqrt{k(1 + o_n(1))}, \end{aligned}$$

kde  $o_n(1)$  značí posloupnost jdoucí k nule pro  $n \rightarrow \infty$ . První nerovnost nám plyne z Minkowského nerovnosti a první rovnost z předpokladu (2.2). To jest pro dostatečně velká  $n$  dostáváme

$$\mathbb{E} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \leq \mathbb{E} \left[ \left( \sum_{r=1}^k \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \right)^2 \right] \leq k.$$

Tudíž získáváme z (2.10)

$$\mathbb{P} \left( \max_{r=1, \dots, k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| > t_k \right) \leq \varepsilon.$$

□

**Lemma 4.**

$$\left| \sum_{(m_1+\dots+m_k)\delta\leq x} \mathbb{P}\left(\bigcap_{r=1}^k A(r, n, m_r, \delta)\right) - \sum_{(m_1+\dots+m_k)\delta\leq x} \prod_{r=1}^k \mathbb{P}(A(r, n, m_r, \delta)) \right| \quad (2.11)$$

$$< k \left(\frac{2t_k}{\delta}\right)^k f(q_n) + 2\varepsilon,$$

pro  $\varepsilon > 0$ .

*Důkaz.* Pravděpodobnost všech množin  $\bigcap_{r=1}^k A(r, n, m_r, \delta)$ , pro které je

$$\max_{r=1,\dots,k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| > t_k,$$

je nejvýše  $\varepsilon$ . K výsledku jsme dospěli pomocí lemma 3 (2.2), jelikož oba členy odhadneme právě  $\varepsilon$ . Mějme nyní posloupnosti množin, pro které platí

$$\max_{r=1,\dots,k} \left| \frac{U_r}{\sqrt{kh(p_n)}} \right| \leq t_k.$$

Opakovanou aplikací strong mixing condition dostáváme:

$$\left| \mathbb{P}\left(\bigcap_{r=1}^k A(r, n, m_r, \delta)\right) - \prod_{r=1}^k \mathbb{P}(A(r, n, m_r, \delta)) \right| \leq (k-1)f(q_n) < kf(q_n), \quad (2.12)$$

kde funkce  $f(m)$  je ze znění (2.1). Abychom dokázali (2.12), využijeme indukci. Pro  $k = 2$  se jedná o samotnou strong mixing condition. Budeme předpokládat, že (2.12) platí pro  $k$  a chceme ukázat platnost pro  $k + 1$ . Pro snadnější zápis budeme značit  $A_r = A(r, n, m_r, \delta)$ .

$$\left| \mathbb{P}\left(\bigcap_{r=1}^{k+1} A_r\right) - \prod_{r=1}^{k+1} \mathbb{P}(A_r) \right| =$$

$$\left| \mathbb{P}\left(A_{k+1} \cap \bigcap_{r=1}^k A_r\right) - \mathbb{P}(A_{k+1}) \cdot \mathbb{P}\left(\bigcap_{r=1}^k A_r\right) + \mathbb{P}(A_{k+1}) \cdot \mathbb{P}\left(\bigcap_{r=1}^k A_r\right) - \prod_{r=1}^{k+1} \mathbb{P}(A_r) \right|$$

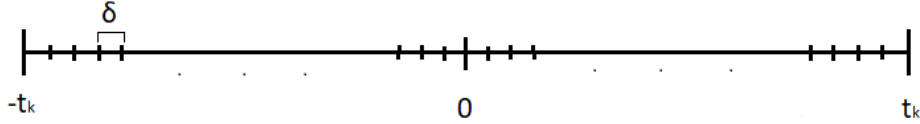
$$\leq \left| \mathbb{P}\left(A_{k+1} \cap \bigcap_{r=1}^k A_r\right) - \mathbb{P}(A_{k+1}) \cdot \mathbb{P}\left(\bigcap_{r=1}^k A_r\right) \right|$$

$$+ \left| \mathbb{P}(A_{k+1}) \cdot \mathbb{P}\left(\bigcap_{r=1}^k A_r\right) - \prod_{r=1}^{k+1} \mathbb{P}(A_r) \right| \leq f(q_n) + (k-1)f(q_n) |\mathbb{P}(A_{k+1})|$$

$$\leq f(q_n) + (k-1)f(q_n) \cdot 1 = kf(q_n) < (k+1)f(q_n).$$

Druhou nerovnost získáváme díky tomu, že vzdálenost mezi jednotlivými  $U_1, U_2, \dots, U_k$  je právě  $q_n$ , a indukčnímu předpokladu. Tedy nám platí (2.12).

Z obrázku 2.3 vidíme, že pro právě jedno  $\frac{U_r}{\sqrt{kh(p_n)}}$  máme, že takových množin je  $\frac{2t_k}{\delta}$ . My však počítáme s  $\bigcap_{r=1}^k A(r, n, m_r, \delta)$ , tedy množin takového tvaru je  $\left(\frac{2t_k}{\delta}\right)^k$ . Z poznatků výše nám vyplývá naše nerovnost.



Obrázek 2.3: Interval pro jeden tvar množiny.

□

Nyní si zadefinujeme náhodné veličiny  $\widetilde{U}_r$ , kde  $\widetilde{U}_r$  má stejné rozdělení jako náhodná veličina  $U_r$  pro  $r = 1, 2, \dots, k$  (mají i stejné vlastnosti), avšak na rozdíl od nich jsou vzájemně nezávislé. Označme si konvoluci distribučních funkcí  $G_{1,n}(x), \dots, G_{k,n}(x)$ , kde se jedná postupně o distribuční funkce náhodných veličin  $\widetilde{U}_1, \widetilde{U}_2, \dots, \widetilde{U}_k$ , následovně:

$$G_{1,n} * \dots * G_{k,n}(x).$$

Zároveň platí, že

$$\begin{aligned} G_{1,n} * \dots * G_{k,n}(x) &\leq \sum_{(m_1 + \dots + m_k)\delta \leq x} \prod_{r=1}^k \mathbf{P}(A(r, n, m_r, \delta)) \\ &\leq G_{1,n} * \dots * G_{k,n}(x + k\delta), \end{aligned} \quad (2.13)$$

$$\begin{aligned} G_{1,n} * \dots * G_{k,n}(x - k\delta) &< \sum_{(m_1 + \dots + m_k + k)\delta \leq x} \prod_{r=1}^k \mathbf{P}(A(r, n, m_r, \delta)) \\ &\leq G_{1,n} * \dots * G_{k,n}(x). \end{aligned} \quad (2.14)$$

K důkazu vlastností výše budeme využívat nerovnosti (2.7) a (2.8), kde si musíme uvědomit, že  $\widetilde{U}_1, \widetilde{U}_2, \dots, \widetilde{U}_k$  jsou nezávislé, tudíž

$$\sum_{(m_1 + \dots + m_k)\delta \leq x} \mathbf{P}\left(\bigcap_{r=1}^k A(r, n, m_r, \delta)\right) = \sum_{(m_1 + \dots + m_k)\delta \leq x} \prod_{r=1}^k \mathbf{P}(A(r, n, m_r, \delta))$$

a obdobně pro součet přes  $(m_1 + \dots + m_k + k)\delta \leq x$ .

Dokážeme si pouze (2.13) a (2.14) se dokáže analogicky. Konvoluce distribučních funkcí nezávislých náhodných veličin je rovna rozdělení součtu těchto veličin, proto nám platí

$$\begin{aligned} G_{1,n} * \dots * G_{k,n}(x) &= \mathbf{P}\left(\sum_{r=1}^k \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} \leq x\right) \leq \sum_{(m_1 + \dots + m_k)\delta \leq x} \mathbf{P}\left(\bigcap_{r=1}^k A(r, n, m_r, \delta)\right) \\ &= \sum_{(m_1 + \dots + m_k)\delta \leq x} \prod_{r=1}^k \mathbf{P}(A(r, n, m_r, \delta)). \end{aligned}$$

Zároveň máme pro  $z = x + k\delta$  z (2.7)

$$G_{1,n} * \dots * G_{k,n}(z) = \mathbf{P}\left(\sum_{r=1}^k \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} \leq z\right) \geq \sum_{(m_1 + \dots + m_k + k)\delta \leq z} \mathbf{P}\left(\bigcap_{r=1}^k A(r, n, m_r, \delta)\right)$$



$$= \sum_{(m_1+\dots+m_k)\delta \leq x} \prod_{r=1}^k \mathbf{P}(A(r, n, m_r, \delta)).$$

Obdržíme tedy nakonec (2.13).

Dále dostáváme rovnosti

$$\begin{aligned} \text{var} \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} &= \frac{1}{kh(p_n)} \text{var} \left( \sum_{j=(r-1)(p_n+q_n)+1}^{rp_n+(r-1)q_n} X_j \right) \\ &= \frac{1}{kh(p_n)} \mathbf{E} \left( \sum_{j=(r-1)(p_n+q_n)+1}^{rp_n+(r-1)q_n} X_j \right)^2 \sim \frac{1}{kh(p_n)} h(p_n) = \frac{1}{k} \\ &\Rightarrow \text{var} \sum_{r=1}^k \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} = k \text{var} \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} \sim 1, \end{aligned}$$

což nám plyne z toho, že  $\widetilde{U}_r$  pro  $r = 1, \dots, k$  jsou nezávislé. Zároveň z předpokladů víme, že

$$\mathbf{E} \left( \sum_{j=(r-1)(p_n+q_n)+1}^{rp_n+(r-1)q_n} X_j \right) = 0.$$

Náhodné veličiny  $\widetilde{U}_r$  pro  $r = 1, \dots, k$  jsou tedy nezávislé a splňují podmínky výše. Ze standardní centrální limitní věty nám plyne

$$G_{1,n} * \dots * G_{k,n}(x) = \mathbf{P} \left( \sum_{r=1}^k \frac{\widetilde{U}_r}{\sqrt{kh(p_n)}} \leq x \right) \xrightarrow[k \rightarrow \infty]{} \Phi(x),$$

kde  $\Phi(x)$  je distribuční funkce normovaného normálního rozdělení.

Nyní chceme, aby platilo  $k, p_n, q_n \rightarrow \infty, kp_n \sim n, k\delta \rightarrow 0$  pro  $n \rightarrow \infty$  tak, aby

$$k \left( \frac{2t_k}{\delta} \right)^k f(q_n) \rightarrow 0, \quad (2.15)$$

$$\frac{kh(q_n)}{h(p_n)} \rightarrow 0.$$

Podmínku (2.15) získáváme z lemmatu 2 a druhá je potřeba dokázat z první části důkazu (2.6).

Nejdříve předvedeme důkaz (2.15). Položme  $\delta = \frac{1}{k^2}$ . Vidíme, že stále platí  $k\delta \rightarrow 0$ . Z našeho označení

$$t_k = \sqrt{\frac{k}{\varepsilon}}$$

plyne

$$k \left( \frac{2t_k}{\delta} \right)^k = k^{\frac{5k}{2}+1} \frac{2^k}{\varepsilon^{\frac{k}{2}}} \leq k^{5k} C^k,$$

kde  $C = \frac{2}{\sqrt{\varepsilon}}$ . Člen  $C^k$  jde k nekonečnu pomaleji než  $k^{5k}$ , proto se budeme zaobírat rychleji rostoucím členem. Chceme tedy

$$k^{5k} f(q_n) = e^{5k \log k + \log f(q_n)} < e^{k^2 + \log f(q_n)} \rightarrow 0.$$

První nerovnost dostáváme díky platnosti  $5k \log k \ll k^2$ . Tedy obdržíme

$$k^2 + \log f(q_n) < 0 \quad \Rightarrow \quad k^2 < -\log f(q_n) \quad \Rightarrow \quad k < \sqrt{-\log f(q_n)}.$$

Víme z toho, jakým způsobem jsme si zadefinovali funkci  $f(x)$ , že  $f(q_n) \xrightarrow{n \rightarrow \infty} 0$ . Proto mínus je pod odmocninou správné (funkce  $\log(x)$  je na okolí 0 záporná).

Tedy pro  $k$

$$k < \sqrt{-\log f(q_n)},$$

naše podmínka (2.15) bude splněna a rozdíl (2.11) bude menší než  $2\varepsilon$  pro  $\varepsilon > 0$  a  $n, k, q_n, p_n \rightarrow \infty$ . Lemma 4 nám vlastně říká, že menšenec a menšitel se nám bude lišit o libovolně malé  $\tilde{\varepsilon} > 0$ .

Pro důkaz (2.6) budeme vycházet z našeho předpokladu (2.2). Víme, že

$$h(n) \sim \mathbf{E} \left( \sum_{j=1}^n X_j \right)^2.$$

Zároveň z vlastností, které jsme si zmínili v úvodu důkazu, platí tyto rovnosti a nerovnosti:

$$\begin{aligned} \sqrt{\mathbf{E} \left( \sum_{j=1}^n X_j \right)^2} &= \sqrt{\mathbf{E} \left( \sum_{r=1}^k U_r + \sum_{r=1}^k V_r \right)^2} \leq \sqrt{\mathbf{E} \left( \sum_{r=1}^k U_r \right)^2} + \sqrt{\mathbf{E} \left( \sum_{r=1}^k V_r \right)^2} \\ &\leq \sum_{r=1}^k \sqrt{\mathbf{E} \left( \sum_{j=(r-1)(p_n+q_n)+1}^{rp_n+(r-1)q_n} X_j \right)^2} + \sum_{r=1}^k \sqrt{\mathbf{E} \left( \sum_{j=rp_n+(r-1)q_n+1}^{r(p_n+q_n)} X_j \right)^2} \\ &\sim \sum_{r=1}^k \sqrt{h(p_n)} + \sum_{r=1}^k \sqrt{h(q_n)} = k \left( \sqrt{h(p_n)} + \sqrt{h(q_n)} \right) \leq k \left( 2\sqrt{h(p_n)} \right). \end{aligned}$$

První a druhá nerovnost vyplývá z Minkowského nerovnosti a vyjádření  $U_r$  a  $V_r$  podle našeho zadefinování. Dále jsme využili předpokladu (2.2) a toho, že funkce  $h(n)$  je rostoucí pro  $n \rightarrow \infty$  a  $\frac{q_n}{p_n} \rightarrow 0$  (posloupnost  $q_n$  roste pomaleji). Máme

$$h(n) \sim \mathbf{E} \left( \sum_{j=1}^n X_j \right)^2 \leq 4k^2 h(p_n).$$

To jest obdržíme

$$\frac{1}{h(p_n)} \leq \frac{4k^2}{h(n)},$$

a tím pádem

$$\frac{kh(q_n)}{h(p_n)} \leq \frac{4k^3 h(q_n)}{h(n)}.$$

Tedy podmínka (2.6) je splněna pro  $h(q_n) = o\left(\frac{h(n)}{k^3}\right)$ , jelikož platí

$$\frac{o\left(\frac{h(n)}{k^3}\right)}{\frac{h(n)}{k^3}} \xrightarrow{n \rightarrow \infty} 0.$$

Nyní si shrneme poznatky, které jsme během důkazu ukázali. Budeme předpokládat, že posloupnosti splňují  $k, p_n, q_n \rightarrow \infty, kp_n \sim n, k\delta \rightarrow 0$  pro  $n \rightarrow \infty$  a platí podmínky (2.15), (2.6) za zmíněných předpokladů. Poté

$$\begin{aligned}
& \mathbb{P} \left( \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \leq x \right) \leq \sum_{(m_1+\dots+m_k)\delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) \\
& < \sum_{(m_1+\dots+m_k)\delta \leq x} \prod_{r=1}^k \mathbb{P}(A(r, n, m_r, \delta)) + \tilde{\varepsilon} \leq G_{1,n} * \dots * G_{k,n}(x + k\delta) + \tilde{\varepsilon}, \\
& \mathbb{P} \left( \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \leq x \right) \geq \sum_{(m_1+\dots+m_k+k)\delta \leq x} \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) \\
& > \sum_{(m_1+\dots+m_k+k)\delta \leq x} \prod_{r=1}^k \mathbb{P}(A(r, n, m_r, \delta)) - \tilde{\varepsilon} \geq G_{1,n} * \dots * G_{k,n}(x - k\delta) - \tilde{\varepsilon}.
\end{aligned}$$

Postupně používáme lemma 4, (2.13), (2.14) a (2.10).

Dále jsme si dokázali, že platí následující:

$$G_{1,n} * \dots * G_{k,n}(x + k\delta) + \tilde{\varepsilon} \longrightarrow \Phi(x) + \tilde{\varepsilon},$$

$$G_{1,n} * \dots * G_{k,n}(x - k\delta) - \tilde{\varepsilon} \longrightarrow \Phi(x) - \tilde{\varepsilon},$$

kde  $\Phi(x)$  je distribuční funkce normovaného normálního rozdělení a  $\tilde{\varepsilon} > 0$  libovolné (tedy i libovolně malé). Musíme zde též podotknout, že  $k\delta \rightarrow 0$  a  $\Phi(x)$  je spojitá funkce. Dále využijeme větu o dvou strážnicích. Nakonec nám tedy plyne

$$\sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

Spolu s (2.5) a Cramér-Sluckého větou získáváme

$$\frac{S_n}{\sqrt{kh(p_n)}} = \sum_{r=1}^k \frac{U_r}{\sqrt{kh(p_n)}} + \sum_{r=1}^k \frac{V_r}{\sqrt{kh(p_n)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0,1).$$

□

## 2.3 Příklad

V této kapitole si zmíníme příklad posloupnosti náhodných veličin, které splňují strong mixing condition a centrální limitní větu pro závislé náhodné veličiny dokázané výše.

Uvažujme náhodné veličiny  $Y_{-l}, Y_{-l+1}, Y_{-l+2}, \dots$ , které jsou nezávislé stejně rozdělené nabývající hodnot  $\pm 1$  a  $l \in \mathbb{N}$ . Pro pravděpodobnosti platí

$$\mathbb{P}(Y_i = 1) = \frac{1}{2} \quad \text{a} \quad \mathbb{P}(Y_i = -1) = \frac{1}{2} \quad \forall i \in \{-l, -l+1, \dots\}.$$

Nyní si definujeme náhodné veličiny následujícím způsobem:

$$X_i = \sum_{j=i-l}^{i+l} Y_j.$$

Náhodné veličiny  $X_1, X_2, X_3, \dots$  se nazývají *klouzavé průměry* a jsou závislé. Pro lepší představu si zde rozepíšeme pár členů pro  $l = 2$ .

$$X_1 = \sum_{j=1-2}^{1+2} Y_j = Y_{-1} + Y_0 + Y_1 + Y_2 + Y_3,$$

$$X_2 = \sum_{j=2-2}^{2+2} Y_j = Y_0 + Y_1 + Y_2 + Y_3 + Y_4,$$

$$X_3 = \sum_{j=3-2}^{3+2} Y_j = Y_1 + Y_2 + Y_3 + Y_4 + Y_5,$$

$$X_4 = \sum_{j=4-2}^{4+2} Y_j = Y_2 + Y_3 + Y_4 + Y_5 + Y_6,$$

$$X_5 = \sum_{j=5-2}^{5+2} Y_j = Y_3 + Y_4 + Y_5 + Y_6 + Y_7,$$

$$X_6 = \sum_{j=6-2}^{6+2} Y_j = Y_4 + Y_5 + Y_6 + Y_7 + Y_8.$$

Vidíme, že pro  $l = 2$  jsou  $X_1$  a  $X_m$  nezávislé pro  $m \geq 6$ .

Pro obecné  $l \in \mathbb{N}$  máme, že  $X_i$  je nezávislá s náhodnými veličinami  $X_m$  pro  $m \geq i + 2l + 1$ . Znovu se podíváme na ekvivalentní znění strong mixing condition:

$$\left| \mathbb{P}[(X_1, \dots, X_i) \in A \cap (X_{i+m}, X_{i+m+1}, \dots) \in B] - \right.$$

$$\left. - \mathbb{P}[(X_1, \dots, X_i) \in A] \cdot \mathbb{P}[(X_{i+m}, X_{i+m+1}, \dots) \in B] \right| \leq f(m),$$

pro všechna  $i, m \in \mathbb{N}$ , množiny  $A, B$  a funkci  $f(m) \rightarrow 0$  pro  $m \rightarrow \infty$ .

Z poznatku výše víme, že pro  $m \geq i + 2l + 1$  jsou  $X_m$  nezávislé náhodné veličiny s  $X_i$ . Tedy pravděpodobnosti v podmínce se pro  $m \geq i + 2l + 1$  rovnají a z toho nám plyne, že rozdíl bude roven nule (funkci  $f(m)$  budeme moci položit nule). Výsledek je, že posloupnost náhodných veličin  $X_1, X_2, X_3, \dots$  splňuje naši strong mixing condition.

Nyní se zaměříme na výpočet středních hodnot  $X_1, X_2, X_3, \dots$ :

$$\mathbb{E}Y_j = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0 \quad \forall j \in \{-l, -l+1, \dots\} \quad \implies Y_j \text{ jsou centrované,}$$

$$\mathbb{E}X_i = \mathbb{E} \sum_{j=i-l}^{i+l} Y_j = \sum_{j=i-l}^{i+l} \mathbb{E}Y_j = 0 \quad l, i \in \mathbb{N} \quad \implies X_i \text{ jsou centrované.}$$

V této části budeme uvažovat  $l = 2$  (pokud nebude řečeno jinak) a podíváme se na výše zmíněných prvních šest členů. Pro jejich součet platí

$$X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 1Y_{-1} + 2Y_0 + 3Y_1 + 4Y_2$$

$$+5Y_3 + 5Y_4 + 4Y_5 + 3Y_6 + 2Y_7 + 1Y_8.$$

Dále získáváme, že

$$\begin{aligned} \mathbb{E} \left( \sum_{i=1}^6 X_i \right)^2 &= \mathbb{E} (1Y_{-1} + 2Y_0 + 3Y_1 + 4Y_2 + 5Y_3 + 5Y_4 + 4Y_5 + 3Y_6 + 2Y_7 + 1Y_8)^2 \\ &= \mathbb{E}Y_{-1}^2 + 4\mathbb{E}Y_0^2 + 9\mathbb{E}Y_1^2 + 16\mathbb{E}Y_2^2 + 25\mathbb{E}Y_3^2 + 25\mathbb{E}Y_4^2 + 16\mathbb{E}Y_5^2 \\ &\quad + 9\mathbb{E}Y_6^2 + 4\mathbb{E}Y_7^2 + \mathbb{E}Y_8^2 = (2 \cdot 30 + 2 \cdot 25) \cdot 1. \end{aligned}$$

Druhou rovnost dostáváme díky tomu, že náhodné veličiny  $Y_{-1}, \dots, Y_8$  jsou nezávislé, a poslední vyplývá z toho, že

$$\mathbb{E}Y_i^2 = 1^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} = 1 \quad \forall i \in \{-1, 0, \dots, 8\}.$$

Tedy máme

$$\text{var} \left( \sum_{i=1}^6 X_i \right) = \mathbb{E} \left( \sum_{i=1}^6 X_i \right)^2 - \left( \mathbb{E} \left( \sum_{i=1}^6 X_i \right) \right)^2 = 110.$$

Nyní, když jsme se podívali na součet konkrétního počtu náhodných veličin, budeme mít lepší představu, jak to bude vypadat pro obecný případ. Z předpokladu (2.2) chceme spočítat

$$\mathbb{E} \left( \sum_{i=1}^{p_n} X_i \right)^2 \sim h(p_n).$$

Pro výpočet funkce  $h(p_n)$  využijeme analogický postup, který jsme si ukázali výše. Platí následující

$$\begin{aligned} \sum_{i=1}^{p_n} X_i &= 1Y_{-1} + 2Y_0 + 3Y_1 + 4Y_2 + \overbrace{5Y_3 + \dots + 5Y_{p_n-2}}^{p_n-4} \\ &\quad + 4Y_{p_n-1} + 3Y_{p_n} + 2Y_{p_n+1} + Y_{p_n+2}, \\ \mathbb{E} \left( \sum_{i=1}^{p_n} X_i \right)^2 &= \mathbb{E}Y_{-1}^2 + 4\mathbb{E}Y_0^2 + 9\mathbb{E}Y_1^2 + 16\mathbb{E}Y_2^2 + \overbrace{25\mathbb{E}Y_3^2 + \dots + 25\mathbb{E}Y_{p_n-2}^2}^{p_n-4} \\ &\quad + 16\mathbb{E}Y_{p_n-1}^2 + 9\mathbb{E}Y_{p_n}^2 + 4\mathbb{E}Y_{p_n+1}^2 + \mathbb{E}Y_{p_n+2}^2 = (2 \cdot 30 + (p_n - 4) \cdot 25) \cdot 1, \end{aligned}$$

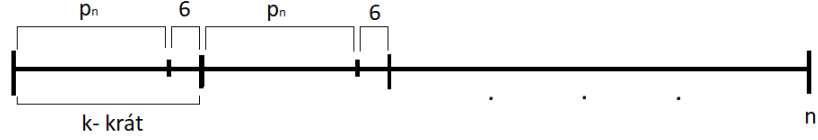
kde zdůvodnění rovností je stejné jako u případu pro prvních šest členů. Dostáváme tedy

$$h(p_n) \sim (60 + (p_n - 4) \cdot 25) \xrightarrow{n \rightarrow \infty} \infty.$$

Na obrázku 2.4 můžeme vidět, jak bychom rozdělili součet  $S_n = \sum_{j=1}^n X_j$  na části, které jsme si zavedli na začátku důkazu věty.

V našem příkladě dostáváme pro  $q_n = 6$

$$U_1 = \sum_{j=1}^{p_n} X_j \quad \text{a} \quad V_1 = \sum_{j=p_n+1}^{p_n+6} X_j,$$



Obrázek 2.4: Příklad.

$$U_2 = \sum_{j=p_n+6+1}^{2p_n+6} X_j \quad \text{a} \quad V_2 = \sum_{j=2p_n+7}^{2(p_n+6)} X_j,$$

⋮

Pevná volba  $q_n = 6$  je zde správná a zároveň

$$\frac{6}{p_n} \xrightarrow{n \rightarrow \infty} 0.$$

Jelikož víme, že  $X_i$  a  $X_m$  jsou nezávislé pro  $m \geq i + 5$ , dostáváme, že  $U_1, U_2, \dots, U_k$  jsou nezávislé náhodné veličiny (viz obrázek 2.4). V lemma 4 nám dokonce platí

$$\left| \mathbb{P} \left( \bigcap_{r=1}^k A(r, n, m_r, \delta) \right) - \prod_{r=1}^k \mathbb{P}(A(r, n, m_r, \delta)) \right| = 0.$$

Z toho nám též plyne (a z poznatků výše), že posloupnost  $p_n$  a  $q_n$  splňují podmínky (2.6) a (2.15). Když se vrátíme do důkazu věty, kterým jsme se zabírali v minulé podkapitole, zaváděli jsme si zde pomocné nezávislé náhodné veličiny  $\tilde{U}_1, \dots, \tilde{U}_k$ . V našem příkladě již  $U_1, \dots, U_k$  jsou nezávislé a můžeme na ně aplikovat standardní centrální limitní větu.

Platí nám tedy centrální limitní věta pro závislé náhodné veličiny  $X_1, X_2, \dots$

$$\frac{S_n}{\sqrt{k(60 + (p_n - 4) \cdot 25)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0,1).$$

### 2.3.1 Simulace příkladu

Nyní výsledky výše představíme na konkrétních příkladech. Mějme  $n = 10000$  a  $q_n = 6$ . Budeme uvažovat tři možnosti volby parametrů  $p_n$  a  $k$ , které by nám měly lépe přiblížit, jaké hodnoty parametrů je vhodné volit. Simulace budeme provádět v programu RStudio.

*První možnost:*

$$k = 100 \implies p_n + q_n = 100 \implies p_n = 94 \implies \frac{q_n}{p_n} = \frac{6}{94},$$

$$\frac{S_n}{\sqrt{100 \cdot (60 + (94 - 4) \cdot 25)}} = \frac{S_n}{\sqrt{231000}} \quad \text{má přibližně } \mathcal{N}(0,1).$$

Vidíme, že podíl  $q_n$  a  $p_n$  je malé číslo a vylučujeme poměrně více úseků s indexy  $q_n$  (jedná se o části, které nám „ruší“ nezávislost).

Druhá možnost:

$$k = 200 \implies p_n + q_n = 50 \implies p_n = 44 \implies \frac{q_n}{p_n} = \frac{6}{44},$$

$$\frac{S_n}{\sqrt{200 \cdot (60 + (44 - 4) \cdot 25)}} = \frac{S_n}{\sqrt{212000}} \text{ má přibližně } \mathcal{N}(0,1).$$

V druhé možnosti jsme zvolili  $k = 200$ , tudíž dostaneme více náhodných veličin  $U_r$ , měla by tedy lépe fungovat standardní centrální limitní věta. Na rozdíl od prvního příkladu máme podíl  $\frac{q_n}{p_n} = \frac{6}{44}$ , což je již poměrně větší číslo, než jsme měli výše. Zároveň vylučujeme větší počet částí  $q_n$ .

Třetí možnost:

$$k = 50 \implies p_n + q_n = 200 \implies p_n = 194 \implies \frac{q_n}{p_n} = \frac{6}{194},$$

$$\frac{S_n}{\sqrt{50 \cdot (60 + (194 - 4) \cdot 25)}} = \frac{S_n}{\sqrt{240500}} \text{ má přibližně } \mathcal{N}(0,1).$$

Podíl  $\frac{q_n}{p_n}$  a počet vyloučených částí  $q_n$  je v posledním případě nejmenší. Co se náhodných veličin  $U_1, \dots, U_{50}$  týče, máme jich nejméně. Tudíž nám z toho vyplývá, že pro centrální limitní větu, aplikovanou na tyto náhodné veličiny, by tento případ měl být nejméně vhodný ze všech tří.

Z vlastností, které jsme si uvedli v úvodu příkladu, dostáváme

$$\begin{aligned} S_n &= \sum_{j=1}^{10000} X_j = Y_{-1} + 2Y_0 + 3Y_1 + 4Y_2 + 5Y_3 + \dots + 5Y_{9998} \\ &\quad + 4Y_{9999} + 3Y_{10000} + 2Y_{10001} + Y_{10002} \\ &= 5 \cdot \sum_{i=3}^{9998} Y_i + 4(Y_2 + Y_{9999}) + 3(Y_1 + Y_{10000}) + 2(Y_0 + Y_{10001}) + (Y_{-1} + Y_{10002}). \end{aligned}$$

Následně

$$Y_i = 2 \left( V_i - \frac{1}{2} \right) = 2V_i - 1 \quad V_i \sim \text{Alt} \left( \frac{1}{2} \right) \quad \forall i \in \{-1, \dots, 10002\}$$

$$\implies \sum_{i=1}^k Y_i = 2B_k - k \quad , \text{ kde } B_k \sim \text{Bi} \left( k, \frac{1}{2} \right)$$

$$\implies S_n = 5(2B_{9996} - 9996) + 4(2B_2 - 2) + 3(2B_2 - 2) + 2(2B_2 - 2) + (2B_2 - 2).$$

Příklad bude tedy simulován generováním 1000 náhodných veličin  $S_{n_1}, \dots, S_{n_{1000}}$ , které budou určeny binomickým rozdělením. Výpočty provedeme pro všechny tři možnosti uvedené na počátku zadání simulace. Dané výsledky následně porovnáme.

Pro zjednodušení zápisu budeme nadále uvažovat

$$T_i = \frac{S_{n_i}}{\sqrt{231000}},$$

$$R_i = \frac{S_{n_i}}{\sqrt{212000}},$$

$$Q_i = \frac{S_{n_i}}{\sqrt{240500}}$$

pro  $\forall i \in \{1, 2, \dots, 1000\}$ . Poté jednotlivé četnosti spočtených hodnot náhodných veličin

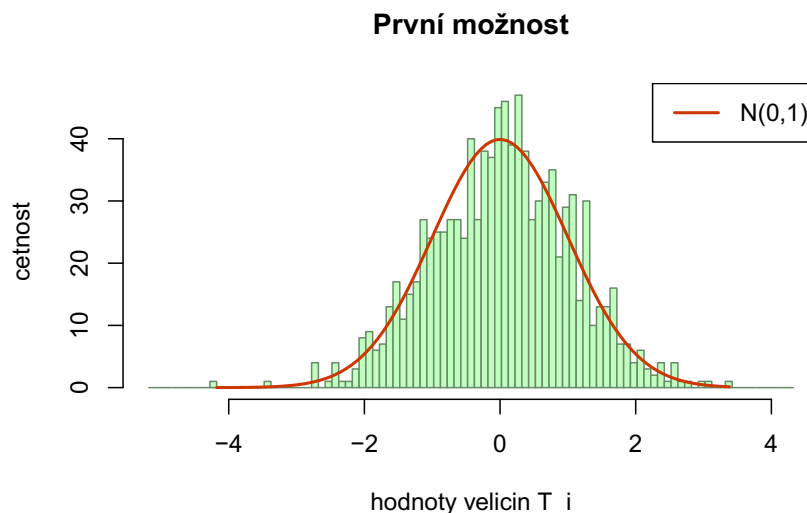
$$T_i, R_i, Q_i \quad \forall i \in \{1, 2, \dots, 1000\}$$

zaneseme do histogramu, ve kterých vykreslíme též graf hustoty normovaného normálního rozdělení (červená křivka v histogramu). Tímto způsobem můžeme nejlépe rozhodnout, zda dané náhodné veličiny mají přibližně rozdělení  $\mathcal{N}(0,1)$ .

Z provedené simulace níže vyplývá, že všechny histogramy pro všechny tři možnosti se nám extrémně neliší. Tedy pro všechny tři možnosti náhodné veličiny  $T_i, R_i, Q_i$  konvergují k normovanému normálnímu rozdělení. Avšak mírný rozdíl je vidět na obrázku 2.6, který nám představuje případ druhý. Lze si všimnout, že histogram je mírně posunutý doprava. Tedy střední hodnota  $ER_i$  bude o něco málo větší než 0 a zároveň než střední hodnoty zbylých náhodných veličin  $T_i, Q_i$ . Co se rozptylu  $R_i$  týče, ten bude také oproti rozptylu  $T_i, Q_i$  větší (viz tabulka 2.1). Dále můžeme vidět z obrázků 2.5, 2.6 a 2.7, že čím blíže jdeme k hodnotě 0, tím se nám histogramy více liší.

	Střední hodnota	Rozptyl
$T_i$	0.0333	1.0269
$R_i$	0.0348	1.0719
$Q_i$	0.0326	1.0063

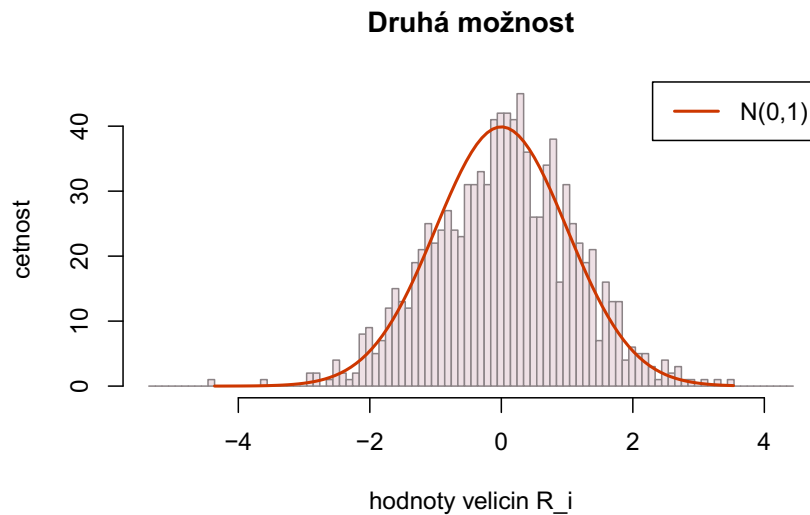
Tabulka 2.1: Tabulka středních hodnot a rozptylů náhodných veličin  $T_i, R_i, Q_i$ .



Obrázek 2.5: Četnost vygenerovaných náhodných veličin  $T_i$ .

Pokud se nyní podíváme na grafy empirických distribučních funkcí náhodných veličin  $T_i, R_i$  a  $Q_i$  (obrázky 2.8, 2.9, 2.10), zjistíme, že velmi dobře aproximují distribuční funkci normovaného normálního rozdělení. Ta je znázorněna v grafu



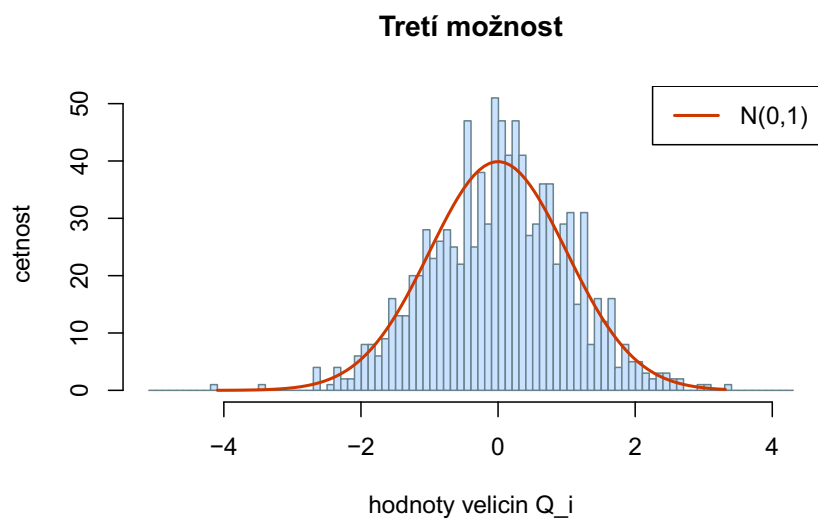


Obrázek 2.6: Četnost vygenerovaných náhodných veličin  $R_i$ .

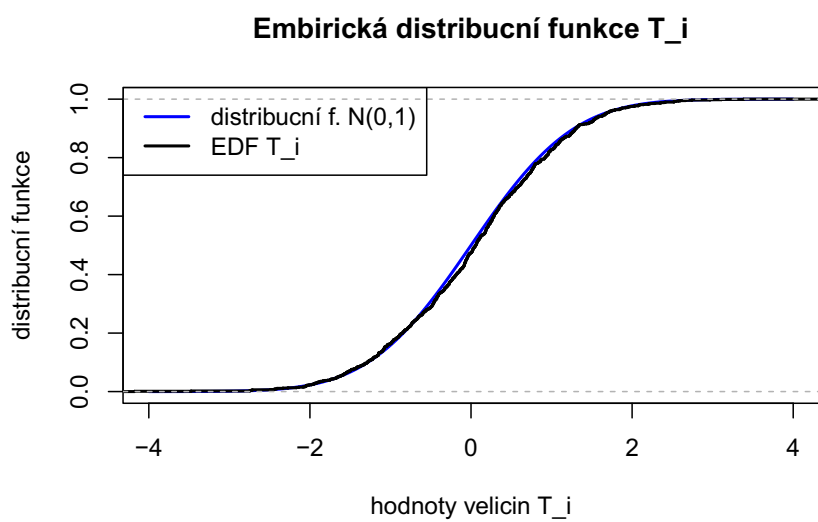
modrou křivkou. Tudiž i zde máme potvrzeno, že náhodné veličiny zmíněné výše konvergují k rozdělení  $\mathcal{N}(0,1)$ .

Z obrázku 2.9 dostáváme, že pro tuto možnost se nám empirická distribuční funkce nejméně přibližuje k distribuční funkci normovaného normálního rozdělení. Pro větší podíl  $\frac{q_n}{p_n}$  nám tudíž náhodné veličiny  $R_i$  pomaleji konvergují k  $\mathcal{N}(0,1)$ , ale i přesto je konvergence velice dobrá.

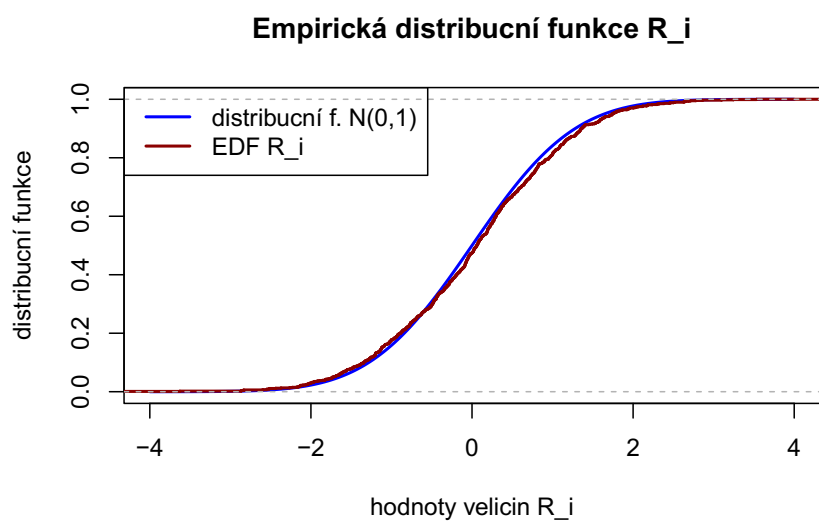
Ze simulace nám vyplývá, že čím menší je podíl  $\frac{q_n}{p_n}$ , tím rychleji nám dané náhodné veličiny konvergují k  $\mathcal{N}(0,1)$ .



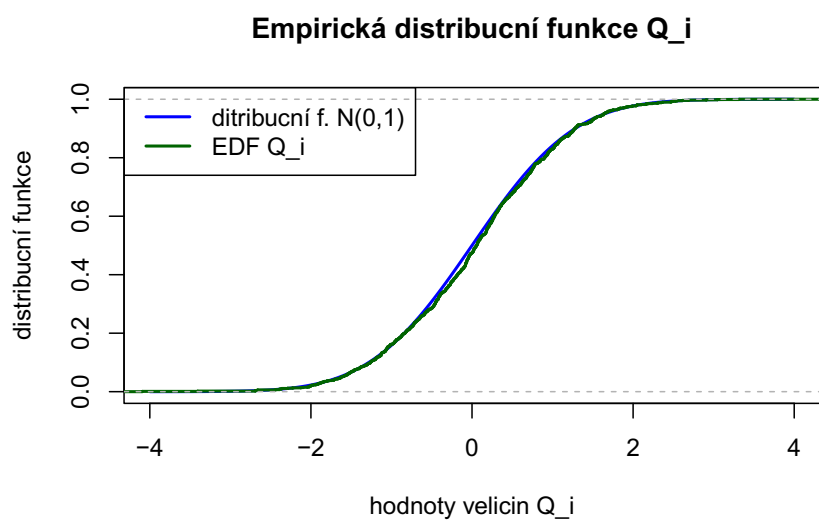
Obrázek 2.7: Četnost vygenerovaných náhodných velicin  $Q_i$ .



Obrázek 2.8: Empirická distribuční funkce vygenerovaných náhodných velicin  $T_i$ .



Obrázek 2.9: Empirická distribuční funkce vygenerovaných náhodných veličin  $R_i$ .



Obrázek 2.10: Empirická distribuční funkce vygenerovaných náhodných veličin  $Q_i$ .

# Závěr

V práci jsme podrobně ukázali platnost centrální limitní věty pro závislé náhodné veličiny. To jest za určitých předpokladů závislé náhodné veličiny mají asymptoticky normované normální rozdělení. Důkaz byl poněkud rozsáhlý, proto jsme si ho rozdělili na jednotlivé části. Článek Rosenblatt (1956), z něhož jsme především čerpali, obsahoval několik málo překlepů a menší chyby, které jsme zároveň opravili.

Dále jsme získané výsledky ilustrovali na konkrétním příkladu klouzavých průměrů. Závěrem práce jsme si nasimulovali tři možnosti volby posloupností, které jsme si zavedli během důkazu. Zjistili jsme, že všechny tři možnosti volby posloupností velice dobře aproximují hustotu normovaného normálního rozdělení. Avšak pokud zvolíme posloupnosti takové, že podíl  $\frac{q_n}{p_n}$  je menší, pak konvergence je rychlejší. Výsledky jsme získaly též z porovnání empirických distribučních funkcí a distribuční funkce normovaného normálního rozdělení.

# Seznam použité literatury

- BRADLEY, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, **2**(none), 107 – 144. doi: 10.1214/154957805100000104. URL <https://doi.org/10.1214/154957805100000104>.
- LEHMANN, E. L. (1999). *Elements of large sample theory*. SPRINGER.
- PROKEŠOVÁ, M. (2023). Nmsa202 pravděpodobnost a matematická statistika. URL <https://www.karlin.mff.cuni.cz/~prokesov/2023/pms.pdf>.
- ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, **42**(1), 43—47. ISSN 0027-8424. doi: 10.1073/pnas.42.1.43. URL <https://europepmc.org/articles/PMC534230>.
- WIKIPEDIA CONTRIBUTORS (2023). Mixing (mathematics) — Wikipedia, the free encyclopedia. URL [https://en.wikipedia.org/w/index.php?title=Mixing\\_\(mathematics\)&oldid=1150799806](https://en.wikipedia.org/w/index.php?title=Mixing_(mathematics)&oldid=1150799806). [Online; accessed 29-June-2023].

# Seznam obrázků

2.1	Rozepsání součtu náhodných veličin po částech. . . . .	7
2.2	Obsah trojúhelníku $[0,0]$ , $[x,0]$ , $[0,x]$ využit při nerovnostech. . . .	9
2.3	Interval pro jeden tvar množiny. . . . .	12
2.4	Příklad. . . . .	18
2.5	Četnost vygenerovaných náhodných veličin $T_i$ . . . . .	20
2.6	Četnost vygenerovaných náhodných veličin $R_i$ . . . . .	21
2.7	Četnost vygenerovaných náhodných veličin $Q_i$ . . . . .	22
2.8	Empirická distribuční funkce vygenerovaných náhodných veličin $T_i$ . . . . .	22
2.9	Empirická distribuční funkce vygenerovaných náhodných veličin $R_i$ . . . . .	23
2.10	Empirická distribuční funkce vygenerovaných náhodných veličin $Q_i$ . . . . .	23