Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Alica Lapšanská

Sekvenční intervalové odhady dané délky

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce:	prof. RNDr. Marie Hušková, DrSc.
Studijní program:	Matematika
Studijní obor:	Obecná matematika

Praha 2012

Na tomto mieste by som sa chcela poďakovať vedúcej mojej bakalárskej práce prof. RNDr. Marii Huškovej, DrSc. za odbornú pomoc a cenné rady pri vypracovaní bakalárskej práce a Jurajovi Hríbovi za pomoc s LaTeXom a jazykom R.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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 v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 25.5.2012 Alica Lapšanská

Názov práce: Sekvenční intervalové odhady dané délky

Autor: Alica Lapšanská

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: prof. RNDr. Marie Hušková, DrSc.

Abstrakt: Bakalárska práca sa zaoberá problémom konštrukcie sekvenčných intervalových odhadov danej dĺžky a danej spoľahlivosti. V rámci práce rozoberáme rôzne postupy jeho riešenia. V prvej časti sa venujeme špeciálnemu prípadu výberu z normálneho rozdelenia. Pre známy rozptyl využívame poznatky z nesekvenčnej teórie intervalových odhadov. Pre neznámy rozptyl popisujeme dvojfázový Steinov postup. Ďalej pre rôzne voľby požadovanej dĺžky intervalového odhadu určujeme strednú hodnotu celkového rozsahu výberu, ktorý je náhodnou veličinou. V druhej časti uvažujeme obecne výber z rozdelenia s kladným konečným rozptylom, ktorý je neznámy. Popisujeme modifikovaný Steinov postup a sekvenčný Waldov postup. V závere práce sa pomocou simulácie snažíme zistiť rozdelenie náhodnej veličiny, ktorá určuje celkový rozsah výberu, pre všetky tri postupy používané pri neznámom rozptyle.

Kľúčové slová: sekvenčné odhady, intervaly spoľahlivosti danej dĺžky, rozsah výberu

Title: Bounded length sequential intervals

Author: Alica Lapšanská

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Marie Hušková, DrSc.

Abstract: The Bachelor's thesis concerns the construction of bounded length sequential intervals with predetermined confidence. This paper analyses some methods, which solve this problem. In the first part we deal with a special case of random sample from normal population. For a known variance we use knowledge from nonsequential theory of interval estimation. We describe Stein's two-stage procedure for an unknown variance. Furthermore, we determine expected value of total sample range for various interval lengths. The second part generally considers a random sample from population with unknown finite variance. We present modified Stein's procedure and sequential Wald's procedure. Finally using simulation, we endeavor to find out a distribution of random variable, which corresponds to the sample range in case of unknown variance. We do this for all of the three mentioned procedures.

Keywords: sequential estimation, bounded length sequential confidence intervals, range of sample

Obsah

Úvod

Sekvenčná analýza sa zaoberá riešením štatistických úloh, v ktorých dáta prichádzajú postupne. Po každom novom pozorovaní sa rozhodnujeme, či v pozorovaní budeme pokračovať, alebo ho ukončíme. Rozsah výberu je teda náhodná veličina, ktorú pred začiatkom pozorovania nepoznáme. Jednou zo základných úloh sekvenčnej analýzy je práve konštrukcia sekvenčných intervalových odhadov danej dĺžky a danej spoľahlivosti. V práci sa zameriavame na sekvenčné intervalové odhady pre strednú hodnotu.

Prvú kapitolu venujeme postupom, ktoré sa používajú pri náhodnom výbere z normálneho rozdelenia. Pre prípad neznámeho rozptylu dokazujeme, že neexistuje intervalový odhad strednej hodnoty založený na pevnom rozsahu výberu, ktorý je danej dĺžky a danej spoľahlivosti a zároveň nezávisí na rozptyle. Popisujeme dvojfázový Steinov postup, jeho vlastnosti a aplikáciu na konkrétne dáta. Odhadneme strednú hodnotu rozsahu výberu potrebného na zostrojenie sekvenčného intervalového odhadu danej dĺžky pre strednú hodnotu Steinovým postupom a odvodíme jeho limitné chovanie v závislosti na požadovanej dĺžke intervalu.

Obsahom druhej kapitoly sú postupy, ktoré sa dajú použiť aj v prípade, že nemáme výber z normálneho rozdelenia a odvodenie ich vlastností.

Posledná kapitola je venovaná náhodnej veličine určujúcej rozsah výberu. Čo sa týka jej rozdelenia, je problém presne ho určiť. Našou snahou je získať informácie o asymptotickom rozdelení rozsahu výberu. Simulujeme úlohu konštrukcie sekvenčného intervalového odhadu danej dĺžky a danej spoľahlivosti pre strednú hodnotu pomocou postupov z kapitoly 1 a 2. Uvádzame grafy s rozdelením rozsahu výberu pre rôzne voľby požadovanej dĺžky intervalu a sledujeme správanie pri jej zmenšovaní. Porovnávame výhodnosť jednotlivých postupov na základe strednej hodnoty rozsahu výberu. Sledujeme aj to, či sa výsledky zhodujú s vlastnosťami, ktoré sme pre jednotlivé postupy odvodili.

V texte predpokladáme, že čitateľ je oboznámený so základnými pojmami z matematickej štatistiky, konkrétne so základmi intervalového odhadu.

Kapitola 1

Sekvenčné intervalové odhady danej dĺžky pre náhodné veličiny s normálnym rozdelením

V tejto kapitole sa budeme zaoberať sekvenčnými intervalovými odhadmi danej dĺžky pre špeciálny prípad.

Nech $X_1, X_2, \dots$ je náhodný výber z normálneho rozdelenia so strednou hodnotou μ a rozptylom σ^2, značíme $\mathsf{N}(\mu, \sigma^2)$. Postup pri určovaní rozsahu výberu n sa líši pre prípady, kedy je σ^2 známe a kedy je neznáme.

1.1 Známe σ^2

Nech $X_1, X_2, \dots$ je náhodný výber z $\mathsf{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ je neznámy parameter a $\sigma^2 > 0$ je známy parameter. Hľadáme intervalový odhad parametru μ dĺžky $2d$ s koeficientom spoľahlivosti $1 - \alpha$, kde $d > 0$ a $0 < \alpha < 1$ sú dané.

Na konštrukciu intervalového odhadu s daným koeficientom spoľahlivosti použijeme poznatky z nesekvenčnej teórie intervalových odhadov, viď Anděl (2007, str. 70-73). Aby zostrojený interval bol požadovanej dĺžky, musíme mať výber o rozsahu n, ktorý určíme neskôr.

Vieme, že *výberový priemer* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ je nestranný, konzistentný odhad parametru μ a platí:

$$\overline{X}_n \sim \mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

kde $X \sim \mathcal{L}(X)$ značí, že náhodná veličina X má rozdelenie $\mathcal{L}(X)$ (všeobecné označenie rozdelenia náhodnej veličiny X). Ďalej platí:

$$\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \sim \mathsf{N}(0,1)$$

a teda $\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ nezávisí na μ. Označme u_β β-kvantil normovaného normálneho rozdelenia. Potom

$$\mathsf{P}\left(\frac{|\overline{X}_n - \mu|}{\sigma}\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad \forall \mu.$$

Teda intervalový odhad μ spoľahlivosti $1-\alpha$ je

$$\left(\overline{X}_n - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \overline{X}_n + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right).$$

Tento interval je dĺžky $2u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Ak chceme aby bol najviac dĺžky $2d$, musíme vziať n tak, aby splňovalo nerovnosť

$$u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le d,$$

$$\left(\frac{u_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2 \le n.$$

Pre intervalový odhad dĺžky $2d$ s koeficientom spoľahlivosti $1-\alpha$ parametru μ je potrebný náhodný výber o rozsahu $n(d) = \left\lfloor \left(\frac{u_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2 \right\rfloor + 1$, kde $\lfloor x \rfloor$ je dolná celá časť čísla x. Pre tento výsledok môžeme písať:

$$n(d) \sim d^{-2} \tag{1.1}$$

čo znamená, že pre $d \to 0$ konverguje $n(d)$ ako d^{-2}.

1.2 Neznáme σ^2

Teraz riešme úlohu z kapitoly 1.1 pre nezmány parameter $\sigma^2 > 0$.

Vieme, že *výberový rozptyl* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ je nestranný, konzistentný odhad parametru σ^2 a platí, viď Anděl (2007, str. 74-75), že $\overline{X}_n$ a S_n^2 sú nezávislé náhodné veličiny. Ďalej platí:

$$\frac{(n-1){S_n}^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{1.2}$$

kde χ^2_{n-1} je χ^2-rozdelenie o $(n-1)$ stupňoch voľnosti,

$$\frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \sim t_{n-1},$$

kde t_{n-1} je Studentovo rozdelenie o $(n-1)$ stupňoch voľnosti, a teda $\frac{\overline{X}_n-\mu}{S_n}\sqrt{n}$ nezávisí na μ ani na σ^2. Označme $t_{\beta,n}$ β-kvantil Studentovho rozdelenia o n stupňoch voľnosti. Potom

$$\mathsf{P}\left(\left|\frac{\overline{X}_n - \mu}{S_n}\right|\sqrt{n} < t_{1-\frac{\alpha}{2},n-1}\right) = 1-\alpha \quad \forall \mu, \sigma^2.$$

Teda intervalový odhad μ spoľahlivosti $1-\alpha$ je

$$\left(\overline{X}_n - t_{1-\frac{\alpha}{2},n-1}\frac{S_n}{\sqrt{n}},\ \overline{X}_n + t_{1-\frac{\alpha}{2},n-1}\frac{S_n}{\sqrt{n}}\right). \tag{1.3}$$

Tento interval je dĺžky $2t_{1-\frac{\alpha}{2},n-1}\frac{S_n}{\sqrt{n}}$. Táto dĺžka je však sama náhodnou veličinou, preto nemôžeme postupovať rovnakým postupom ako v predošlom prípade.

1.2.1 Neexistencia intervalového odhadu založeného na pevnom rozsahu výberu

Pre náš problém konštrukcie intervalového odhadu danej dĺžky a danej spoľahlivosti pri výbere z normálneho rozdelenia s neznámimi parametrami μ a σ^2 bolo dokázané, že neexistuje netriviálny intervalový odhad založený na pevnom rozsahu výberu, ktorý by nezávisel na parametri σ^2. Neskôr bol tento záver dokázaný aj pre širšiu skupinu rozdelení. Vychádza sa z toho, že dve rozdelenia s rôznou polohou môžeme priblížiť na ľubovoľne malú vzdialenosť, ak dostatočne zväčšíme parameter určujúci ich rozptyl.

Nech μ a σ^2 sú neznáme parametre. Predpokladajme, že existuje n a interval $(L_n = L_n(\boldsymbol{X}), U_n = U_n(\boldsymbol{X}))$, kde $\boldsymbol{X} = (X_1, \dots, X_n)'$ je vektor náhodných veličín s pevným počtom zložiek, ktorých združená hustota $f(\boldsymbol{x}, \mu, \sigma^2)$ je všade spojitá v $\boldsymbol{x} \in \mathbb{R}^n$, $\mu \in \mathbb{R}$ a $0 < \sigma^2 < \infty$. Predpokladajme, že interval (L_n, U_n) má požadované vlastnosti, tj.

a) $U_n - L_n = 2d$,

b) $\mathsf{P}_{\mu,\sigma^2}(L_n < \mu < U_n) = 1 - \alpha \quad \forall\mu \in \mathbb{R} \quad \forall\sigma^2 > 0.$

Keďže $-\infty < \mu < \infty$, tak $\forall r \geq 2$ existujú $\mu_1, \dots, \mu_r$ splňujúce $\forall\mu_i, \mu_j,\ i, j = 1, \dots r$ platí $|\mu_i - \mu_j| > 2d$. Z a) plynie, že množiny

$$\mathcal{E}_i = \{\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n, \mu_i \in (L_n(\boldsymbol{x}), U_n(\boldsymbol{x}))\},\ i = 1, \dots, r,$$

sú po dvoch disjunktné. Z b) je $\mathsf{P}_{\mu_i,\sigma^2}(\boldsymbol{X} \in \mathcal{E}_i) = 1 - \alpha$, $i = 1, \dots, r$, a z článku Chatterjee (1991, tvrdenie (2.2)) platí

$$\sup_{\mathcal{E}} |\mathsf{P}_{\mu_i,\sigma^2}(\boldsymbol{X} \in \mathcal{E}) - \mathsf{P}_{\mu_1,\sigma^2}(\boldsymbol{X} \in \mathcal{E})| < \frac{1}{2}(1 - \alpha), \quad i = 1, \dots, r.$$

Supremum beží cez všetky merateľné množiny $\mathcal{E} \subset \mathbb{R}^n$. Odtiaľ

$$\mathsf{P}_{\mu_1,\sigma^2}(\boldsymbol{X} \in \mathcal{E}_i) = 1 - \alpha,\ i = 1, \dots, r. \tag{1.4}$$

Keďže množiny $\mathcal{E}_i,\ i = 1, \dots, r$, sú po dvoch disjunktné, tak podľa (1.4) musí platiť $\frac{r(1-\alpha)}{2} < 1$ a to je spor, lebo $r \geq 2$ je ľubovoľné prirodzené číslo. Dokázali sme neexistenciu intervalového odhadu slňujúceho a), b).

1.2.2 Steinov postup

Použijeme metódu, ktorú sformuloval Stein (1945), viď Ghosh (1991, str. 24-26) alebo Rao (1978, str. 531-532).

Krok I Nech $n(\geq 2)$ je predom zvolený rozsah prvého výberu. Nech $X_1, \dots, X_n$ je prvý náhodný výber z $\mathsf{N}(\mu, \sigma^2)$, kde μ, σ^2 sú neznáme parametre a nech $z > 0$ je číslo, ktoré určíme neskôr. Z prvého výberu vypočítame S_n^2 a určíme najmenšie prirodzené číslo N tak, aby splňovalo $N > n$ a $N \geq \frac{S_n^2}{z}$.
Teda

$$N = max\left\{n + 1, \left\lfloor \frac{S_n^2}{z} \right\rfloor + 1\right\}, \tag{1.5}$$

kde $\lfloor x \rfloor$ je dolná celá časť čísla x.

Potom nájdeme $1 + N - n$ čísel a_0, $a_{n+1}, \dots, a_N$ splňujúcich rovnice

a) $a_0 + \sum_{i=n+1}^{N} a_i = 1$,

b) $\frac{a_0^2}{n} + \sum_{i=n+1}^{N} a_i^2 = \frac{z}{S_n^2}$.

Poznamenajme, že pre čísla splňujúce rovnosť a) nadobúda ľavá stana v rovnici b) hodnoty z polouzavretého intervalu $\left[\frac{1}{N}, \infty\right)$. Pre N ale platí $\frac{z}{S_n^2} \geq \frac{1}{N}$, a teda sústava rovníc a), b) má riešenie.

Krok II Vezmeme $X_{n+1}, \dots, X_N$ druhý výber z $\mathsf{N}(\mu, \sigma^2)$ o veľkosti $N - n$. V tomto kroku definujeme náhodnú veličinu

$$Y_N := a_0 \overline{X}_n + \sum_{i=n+1}^{N} a_i X_i \tag{1.6}$$

a dokážeme o nej následujúce tvrdenie.

Veta 1.1. *Platí*

$$T = \frac{Y_N - \mu}{\sqrt{z}} \ \sim \ t_{n-1}.$$

Dôkaz. Najprv si uvedomíme, že N je náhodná veličina závislá na S_n^2. Teda aj náhodná veličina Y_N je závislá na S_n^2. Vypočítame podmienené rozdelenie Y_N za podmienky S_n^2, ako je definované v učebnom texte Lachout (2004, Definice 9.1). Vieme, že Y_N je súčtom dvoch nezávislých náhodných veličín $a_0 \overline{X}_n$ a $\sum_{i=n+1}^{N} a_i X_i$. Z vety o rozdelení lineárnej kombinácie iid (nezávislých, rovnako rozdelených) náhodných veličín s normálnym rozdelením, viď Anděl (2007, str. 63-65), platí, že podmienené rozdelenie $a_0 \overline{X}_n$ za podmienky S_n^2 je $\mathsf{N}\left(a_0 \mu, \frac{a_0^2 \sigma^2}{n}\right)$ a podmienené rozdelenie $\sum_{i=n+1}^{N} a_i X_i$ za podmienky S_n^2 je $\mathsf{N}\left(\sum_{i=n+1}^{N} a_i \mu, \sum_{i=n+1}^{N} a_i^2 \sigma^2\right)$. Teda podmienené rozdelenie Y_N za podmienky S_N^2 je

$$\mathsf{N}\left\{\left(a_0 + \sum_{i=n+1}^{N} a_i\right)\mu, \left(\frac{a_0^2}{n} + \sum_{i=n+1}^{N} a_i^2\right)\sigma^2\right\}.$$

A z predpokladov a), b) na konštanty a_0, $a_{n+1}, \dots, a_N$ platí, že podmienené rozdelenie Y_N za podmienky S_n^2 je $\mathsf{N}\left(\mu, \frac{z}{S_n^2}\sigma^2\right)$.

Ak chceme dokázať, že náhodná veličina T má t-rozdelenie o $n - 1$ stupňoch voľnosti, tak stačí ukázať, viď Anděl (2007, Věta 4.22), že platí $T = \frac{U}{\sqrt{V}} \sqrt{n-1}$, kde U a V sú nezávislé náhodné veličiny, $U \sim \mathsf{N}(0, 1)$ a $V \sim \chi^2_{n-1}$.

Položme

$$U = \frac{Y_N - \mu}{\sqrt{z}} \frac{S_n}{\sigma},$$

$$V = \frac{(n-1){S_n}^2}{\sigma^2}.$$

Využitím toho, že $\overline{X}_n$ a S_n^2 sú nezávislé, dostávame, že podmienené rozdelenie U za podmienky S_n^2 je $\mathsf{N}(0,1)$. Podmienené rozdelenie nezávisí na podmienke, a teda U a S_n^2 sú nezávislé. Teda

$$U = \frac{Y_N - \mu}{\sqrt{z}} \frac{S_n}{\sigma} \sim \mathsf{N}(0,1)$$

a z tvrdenia (1.2)

$$V = \frac{(n-1){S_n}^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Odtiaľ U, V nezávislé a platí:

$$\frac{U}{\sqrt{V}}\sqrt{n-1} = \frac{\frac{Y_N-\mu}{\sqrt{\frac{z}{S_n^2}\sigma^2}}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2}}}\sqrt{n-1} = \frac{Y_N - \mu}{\sqrt{z}} = T.$$

Dokázali sme, že T má t-rozdelenie o $n-1$ stupňoch voľnosti.

$\square$

Podľa Vety 1.1 potom platí: $\mathsf{P}\left(|Y_N - \mu| < \sqrt{z}\, t_{1-\frac{\alpha}{2},n-1}\right) = 1-\alpha.$
Ak na začiatku zvolíme z tak, aby splňovalo rovnosť

$$\sqrt{z}\, t_{1-\frac{\alpha}{2},n-1} = d, \tag{1.7}$$

potom bude platiť: $\mathsf{P}(Y_N - d < \mu < Y_N + d) = 1-\alpha \quad \forall \mu, \sigma^2.$
Teda $Y_N \mp d$ je intervalový odhad s koeficientom spoľahlivosti $1-\alpha$ a dĺžkou $2d$ pre parameter μ.

Ukážme si použitie Steinovho postupu na následujúcom príklade, ktorý je uvedený v knihe Wonnacot (1993, Cvičení 8-9).

Príklad 1.1. U náhodného výberu štyridsiatich automobilov bola sledovaná rýchlosť (v km/h):

49	83	58	65	68	60	76	86	74	53
71	74	65	72	64	42	62	62	58	82
78	64	55	87	56	50	71	58	57	75
58	86	64	56	45	73	54	86	70	73

Predpokladajme, že merania tvoria náhodný výber z normálneho rozdelenia $\mathsf{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ sú neznáme parametre označujúce strednú hodnotu a rozptyl rýchlosti, akou idú automobily meraným úsekom. Zostrojte intervalový odhad o spoľahlivosti $0,95$ pre strednú rýchlosť všetkých automobilov, ktoré prešli daným úsekom.

Najprv zostrojme nesekvenčný intervalový odhad pomocou vzorca (1.3). Máme $n = 40$, $\alpha = 0,05$. Vypočítame potrebné údaje:

$$\overline{X}_{40} = 66, \quad S_{40}^2 = 139,923.$$

Z tabuliek pre kvantily t-rozdelenia máme $t_{0,975;40} = 2,021$. Intervalový odhad pre μ o spoľahlivosti $0,95$ je $(62,22; 69,78)$. Jeho dĺžka je $7,56$.

Teraz použijme Steinov postup na zostrojenie intervalových odhadov o spoľahlivosti $0,95$ a dĺžok 9 a 8. Teda $d_1 = 4,5$ a $d_2 = 4$.

Krok I Zvoľme $n = 10$ ako rozsah prvého výberu. Vypočítame potrebné údaje:

$$\overline{X}_{10} = 67,2; \quad S_{10}^2 = 155,733$$

Vypočítame konštanty z_1, z_2 zo vzorca (1.7), kde $t_{0,975;9} = 2,262$: $z_1 = 3,957$; $z_2 = 3,127$. Označme

$$m_1 := \left\lfloor \frac{S_{10}^2}{z_1} \right\rfloor + 1 = 40,$$

$$m_2 := \left\lfloor \frac{S_{10}^2}{z_2} \right\rfloor + 1 = 50.$$

Z (1.5) máme $N_1 = 40$ a $N_2 = 50$. Podľa prvých desiatich pozorovaní, sme v prvom kroku Steinovho postupu dostali výsledok, že ak chceme zostrojiť intervalový odhad dĺžky 8, musíme urobiť ešte ďalších 40 pozorovaní. Pre intervalový odhad dĺžky 9 je potrebné urobiť ďalších 30 pozorovaní. V porovnaní s výsledkom nesekvenčného odhadu potrebuje Steinov postup pre túto konkrétnu realizáciu dát väčší rozsah výberu na zostrojenie intervalového odhadu väčšej dĺžky. Pamätajme, že rozsah výberu N je náhodná veličina a tento konkrétny prípad nám nič nehovorí o jej vlastnostiach.

Zvoľme teraz $n = 20$ ako rozsah prvého výberu a pridajme jednu dĺžku $d_3 = 3,5$. Vypočítajme potrebné údaje:

$$\overline{X}_{20} = 66,2; \quad S_{20}^2 = 114,6$$

Ďalej z tabuliek pre kvantily t-rozdelenia je $t_{0,975;19} = 2,093$ a zo vzorca (1.7) máme $z_1 = 4,6226$, $z_2 = 3,6524$, $z_3 = 2,7964$. Potom

$$m_1 := \left\lfloor \frac{S_{20}^2}{z_1} \right\rfloor + 1 = 25,$$

$$m_2 := \left\lfloor \frac{S_{20}^2}{z_2} \right\rfloor + 1 = 32,$$

$$m_3 := \left\lfloor \frac{S_{20}^2}{z_3} \right\rfloor + 1 = 41.$$

Teda $N_1 = 25$; $N_2 = 32$ a $N_3 = 41$. Vidíme, že pre inak zvolený rozsah prvého výberu, sa viditeľne zmenšil rozsah druhého výberu. Dokonca ak by sme chceli zostrojiť intervalový odhad dĺžky 7, potrebovali by sme urobiť len o jedno pozorovanie viac, než sme použili na zostrojenie intervalového odhadu dĺžky $7,56$ pomocou nesekvenčného postupu. My máme k dispozícii pevný počet pozorovaní, a preto zostrojíme dva intervalové odhady dĺžky $2d_1$ a $2d_2$.

1: Najprv chceme intervalový odhad μ dĺžky 9. Hľadáme $N_1 - n + 1 = 6$ konštánt $a_0,\ a_{21},\ a_{22}, \dots, a_{25}$ splňujúcich rovnice a), b). Zvoľme $a_1 := a_{21} = \dots = a_{25}$. Potom riešime sústavu rovníc:

$$
\begin{aligned}
a_0 + 5a_1 &= 1 & &\Rightarrow a_0 = 1 - 5a_1 \\
\frac{a_0^2}{20} + 5a_1^2 &= 0,04
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow (1 - 5a_1)^2 + 100a_1^2 &= 0,8 \\
125a_1^2 - 10a_9 + 0,2 &= 0
\end{aligned}
$$

Vezmeme $a_0 = 0,8;\ a_1 = 0,04$.

Krok II Vypočítame Y_N z (1.6): $Y_{25} = 66,46$. Teda intervalový odhad s koeficientom spoľahlivosti $0,95$ a dĺžkou 9 pre μ je $(61,96; 70,96)$. Pre tento konkrétny súbor dat je interval, ktorý sme zhotovili pomocou nesekvenčného postupu, obsiahnutý v intervale zo Steinovho postupu.

2: Teraz chceme intervalový odhad μ dĺžky 8. Hľadáme $N_2 - n + 1 = 13$ konštánt $a_0,\ a_{21},\ a_{22}, \dots, a_{32}$ splňujúcich rovnice a), b). Zvoľme $a_1 := a_{21} = \dots = a_{32}$. Potom riešime sústavu rovníc:

$$
\begin{aligned}
a_0 + 12a_1 &= 1 & &\Rightarrow a_0 = 1 - 12a_1 \\
\frac{a_0^2}{20} + 12a_1^2 &= 0,032
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow (1 - 12a_1)^2 + 240a_1^2 &= 0,64 \\
384a_1^2 - 24a_9 + 0,36 &= 0
\end{aligned}
$$

Vezmeme $a_0 = 0,55;\ a_1 = 0,0375$.

Krok II Vypočítame Y_N z (1.6): $Y_{32} = 66,2225$. Teda intervalový odhad s koeficientom spoľahlivosti $0,95$ a dĺžkou 8 pre μ je $(62,2225; 70,2225)$. Pre danú realizáciu dat je horný odhad zostrojený nesekvenčným postupom obsiahnutý v intervale zo Steinovho postupu, ale dolný odhad v ňom už neleží.

△

1.3 Stredná hodnota a aproximácia náhodnej veličiny N

V Steinovom postupe je veľkosť výberu vyjadrená náhodnou veličinou N, o ktorej toho zatiaľ veľa nevieme. Z definície (1.5) môžeme vypočítať strednú hodnotu veľkosti výberu potrebnú na zostrojenie intervalu spoľahlivosti pre μ pomocou Steinovho postupu. Keďže platí

$$\frac{(n-1){S_n}^2}{\sigma^2} \sim \chi^2_{n-1},$$

máme

$$\begin{aligned}\mathsf{E}_{\sigma^2}(N) &= (n+1)\,\mathsf{P}\left(\frac{S_n^2}{z} \le n+1\right) + \sum_{i=n+1}^{\infty}(i+1)\,\mathsf{P}\left(i < \frac{S_n^2}{z} \le i+1\right)\\ &= (n+1)\,\mathsf{P}\left(\chi^2_{n-1} \le \frac{(n+1)(n-1)z}{\sigma^2}\right) + \\ &+ \sum_{i=n+1}^{\infty}(i+1)\,\mathsf{P}\left(\frac{i(n-1)z}{\sigma^2} < \chi^2_{n-1} \le \frac{(i+1)(n-1)z}{\sigma^2}\right).\end{aligned}$$

Ak je $n < \frac{S_n^2}{z}$, tak platí

$$\frac{S_n^2}{z} \le N \le \frac{S_n^2}{z} + 1. \tag{1.8}$$

Označme

$$e(\sigma^2) = (n+1)\,\mathsf{P}\left(\chi^2_{n-1} \le \frac{(n+1)(n-1)z}{\sigma^2}\right) + \frac{\sigma^2}{z}\,\mathsf{P}\left(\chi^2_{n-1} > \frac{(n+1)(n-1)z}{\sigma^2}\right).$$

Potom odhad strednej hodnoty je:

$$e(\sigma^2) < \mathsf{E}_{\sigma^2}(N) < e(\sigma^2) + \mathsf{P}\left(\chi^2_{n-1} > \frac{(n+1)(n-1)z}{\sigma^2}\right).$$

Pri malých hodnotách n a $\frac{z}{\sigma^2}$ je $\mathsf{E}_{\sigma^2}(N)$ približne $\frac{\sigma^2}{z}$ a z (1.7) platí: $d \to 0 \;\Rightarrow\; z \to 0$. Teda pre $d \to 0$ platí $\mathsf{E}_{\sigma^2}(N) \to 0$. Pre pevné $d > 0$ a $n \to \infty$ je z (1.5) $N = n$, a teda $\mathsf{E}_{\sigma^2}(N) \to \infty$.

Podľa úvahy (1.1) skúsme voliť $n = n(d)$ v závislosti na d. Zvoľme $n = \left[\frac{1}{d}\right]$, kde $[x]$ je celé číslo, ktoré vznikne zaokrúhlením reálneho čísla x. Náhodná veličina $N(d)$ je tiež závislá na d. Potom dosadením do (1.8) a z (1.7) máme:

$$\begin{aligned} S^2_{\left[\frac{1}{d}\right]}\; t_{1-\frac{\alpha}{2},\left[\frac{1}{d}\right]-1} d^{-2} \le N(d) \le S^2_{\left[\frac{1}{d}\right]}\; t_{1-\frac{\alpha}{2},\left[\frac{1}{d}\right]-1} d^{-2} + 1,\\ S^2_{\left[\frac{1}{d}\right]}\; t_{1-\frac{\alpha}{2},\left[\frac{1}{d}\right]-1} \le N(d)d^2 \le S^2_{\left[\frac{1}{d}\right]}\; t_{1-\frac{\alpha}{2},\left[\frac{1}{d}\right]-1} + d^2\end{aligned}$$

Dokážeme, že pre $d \to 0$ konverguje ľavá i pravá strana *skoro iste*[1] k číslu $\sigma^2 u^2_{1-\frac{\alpha}{2}}$. Zo *silného zákona veľkých čísel (SZVČ)*, viď Dupač (2005, Věta 4.8), plynie: $S_n^2 \xrightarrow[n\to\infty]{s.i.} \sigma^2$, lebo

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mathsf{E}\,X_i + \mathsf{E}\,X_i - \overline{X}_n)^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mathsf{E}\,X_i)^2 + \frac{1}{n-1}\sum_{i=1}^{n}(\mathsf{E}\,X_i - \overline{X}_n)^2\\ &+ \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mathsf{E}\,X_i)(\mathsf{E}\,X_i - \overline{X}_n)\end{aligned}$$

[1] Jav A nastáva *skoro iste* práve vtedy, keď $\mathsf{P}(A) = 1$. Ďalej budeme značiť skratkou s.i.

a jednotlivé sčítance sa správajú nasledovným spôsobom

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mathsf{E}\,X_i)^2 \xrightarrow[n\to\infty]{\text{SZVČ},s.i.} \sigma^2,$$
$$\frac{1}{n-1}\sum_{i=1}^{n}(\mathsf{E}\,X_i - \overline{X}_n)^2 = (\mu - \overline{X}_n)^2 \xrightarrow[n\to\infty]{\text{SZVČ},s.i.} 0,$$
$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mathsf{E}\,X_i)(\mathsf{E}\,X_i - \overline{X}_n) = (\mu - \overline{X}_n)(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\text{SZVČ},s.i.} 0.$$

Pre kvantily platí pre $n \to \infty$ aproximácia $t_{\beta,n} \sim u_\beta$, viď Dupač (2005, str. 108). Ak vezmeme za $n = n(d) = \left[\frac{1}{d}\right]$, potom pre $d \to 0$ konverguje $n \to \infty$.

Dokázali sme, že platí

$$N(d)d^2 \xrightarrow[d\to 0]{s.i.} \sigma^2 u^2_{1-\frac{\alpha}{2}}. \tag{1.9}$$

Kapitola 2

Sekvenčné intervalové odhady danej dĺžky - všeobecný prípad

Uvažujme teraz všeobecný prípad náhodného výberu $X_1, X_2, \ldots$ z rozdelenia s hustotou $f(x;\theta)$, kde θ je neznámy parameter. Nech $\mathsf{E}\, X_i = \mu$, $\mathsf{var}\, X_i = \sigma^2$, kde $0 < \sigma^2 < \infty$ a $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ (množinový kartézsky súčin). Všetky výsledky dosiahnuté v tejto kapitole chápeme asymptoticky pre $d \to 0$.

2.1 Modifikovaný Steinov postup

Steinov postup môžeme použiť aj v prípade, že nemáme zaručené normálne rozdelenie, ale n je dosť veľké. Plynie to z *centrálnej limitnej vety (CLV)*, viď Dupač (2005, Věta 4.12). Nech Φ je distribučná funkcia normovaného normálneho rozdelenia. Potom platí

$$\mathsf{P}\left(\frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \le x\right) \xrightarrow[n\to\infty]{CLV} \Phi(x) \quad \forall x \in \mathbb{R}$$

Podľa výsledkov z kapitoly 1.3 zvoľme rozsah prvého výberu v závislosti na d, teda $n = n(d)$, a predpokladajme $d \to 0$. Zostrojme intervalový odhad danej dĺžky $2d$ o asymptotickej spoľahlivosti $1 - \frac{\alpha}{2}$.

Krok I Pre $n(d) = \left[\frac{1}{d}\right]$ spravíme prvý výber $X_1, \ldots, X_{n(d)}$ z rozdelenia s hustotou $f(x;\mu,\sigma^2)$, kde μ, σ^2 sú neznáme parametre. Vypočítame $S^2_{n(d)}$ a určíme najmenšie prirodzené číslo $N(d)$ tak, že

$$N(d) = max\left\{n(d)+1, \left\lfloor\frac{S^2_{n(d)}}{z}\right\rfloor + 1\right\}, \tag{2.1}$$

kde $\lfloor x \rfloor$ je dolná celá časť čísla x.

Potom nájdeme $1 + N(d) - n(d)$ čísel a_0, $a_{n(d)+1}, \ldots, a_{N(d)}$ splňujúcich rovnice

a) $a_0 + \sum_{i=n(d)+1}^{N(d)} a_i = 1$,

b) $\frac{a_0^2}{n(d)} + \sum_{i=n(d)+1}^{N(d)} a_i^2 = \frac{z}{S^2_{n(d)}}$.

Krok II Ďalej spravíme druhý výber $X_{n(d)+1}, \ldots, X_{N(d)}$ o veľkosti $N(d)-n(d)$. Definujeme náhodnú veličinu

$$Y_{N(d)} := a_0 \overline{X}_{n(d)} + \sum_{i=n(d)+1}^{N(d)} a_i X_i \tag{2.2}$$

a označme

$$T_{N(d)} = \frac{Y_{N(d)} - \mu}{\sqrt{z}}. \tag{2.3}$$

Pre túto náhodnú veličinu bude platiť nasledujúca veta.

Veta 2.1. *Platí*

$$\mathsf{P}\left(T_{N(d)} \leq x\right) \xrightarrow[d\to 0]{} \Phi(x) \quad \forall x \in \mathbb{R}.$$

Túto vetu dokážeme pomocou vety, ktorú aj s dôkazom nájdeme v knihe Rényi (1972, Věta 2).

Veta 2.2. *Nech $X_1, \ldots, X_i, \ldots$ sú náhodné veličiny s rovnakým rozdelením, pre ktoré platí $\mathsf{E}\, X_i = 0$ a $\mathsf{var}\, X_i = 1$. Nech pre náhodné veličiny*

$$Z_m = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} X_i$$

platí vzťah $\lim_{m\to\infty} \mathsf{P}(Z_m \leq x) = \Phi(x)$. Predpokladajme, že $\{N_i\}_{i=1}^{\infty}$ je postupnosť náhodných veličín nadobúdajúcich za svoje hodnoty iba prirodzené čísla a platí

$$\frac{N_i}{i} \xrightarrow[i\to\infty]{P} c,$$

tj. $\lim_{i\to\infty} \mathsf{P}\left(\frac{N_i}{i} = c\right) = 1$, kde c je kladná konštanta. Potom platí

$$\lim_{i\to\infty} \mathsf{P}(Z_{N_i} \leq x) = \Phi(x).$$

Dôkaz Vety 2.1. Zavedieme náhodné veličiny

$$U_i = \frac{Y_i - \mu}{\sqrt{z}} \frac{S_{n(d)}}{\sigma}, \quad i \geq n(d),$$

kde Y_i je definované ako v (2.2), kde berieme $i = N(d)$. Potom sú náhodné veličiny $U_{n(d)}, U_{n(d)+1}, \ldots$ rovnako rozdelené a z dôkazu Vety 1.1 platí $\mathsf{E}\, U_i = 0$ a $\mathsf{var}\, U_i = 1$, $i \geq n(d)$. Ďalej definujeme veličiny

$$V = \frac{(n(d)-1){S_{n(d)}}^2}{\sigma^2},$$

$$T_i = \frac{U_i}{\sqrt{V}} \sqrt{n(d)-1}, \quad i \geq n(d).$$

Z Vety 1.1 platí $T_i = \frac{Y_i - \mu}{\sqrt{z}} \sim t_{i-1}$, pre $i \geq n(d)$. Označme H_i distribučnú funkciu Studentovho rozdelenia o i stupňoch voľnosti. O nej vieme, že platí:

$$\lim_{i\to\infty} H_i(x) = \Phi(x), \quad x \in \mathbb{R},$$

viď Dupač (2005, Věta 5.5). Teda

$$\lim_{d\to 0} \mathsf{P}\left(T_{n(d)} \leq x\right) = \Phi(x).$$

Za Z_{N_i} z Vety 2.2 vezmeme náhodnú veličinu $T_{N(d)}$. Potom z výsledku (1.9), konvergencia skoro iste, plynie konvergencia v pravdepodobnosti, viď Rényi (1972, str. 333), a teda sú splnené predpoklady Vety 2.2 a z nej plynie platnosť Vety 2.1.

□

Teda z Vety 2.1 platí: $\mathsf{P}\left(\left|Y_{N(d)} - \mu\right| < \sqrt{z}\, u_{1-\frac{\alpha}{2}}\right) \doteq 1 - \alpha$.
Zvoľme z tak, aby splňovalo rovnicu

$$\sqrt{z}\, u_{1-\frac{\alpha}{2}} = d. \tag{2.4}$$

Platí: $\mathsf{P}(Y_{N(d)} - d < \mu < Y_{N(d)} + d) = 1 - \alpha \quad \forall(\mu, \sigma^2) \in (\mathbb{R}, \mathbb{R}^+)$.
Interval $(Y_{N(d)} - d, Y_{N(d)} + d)$ je intervalový odhad o asymptotickej spoľahlivosti $1 - \alpha$ a dĺžkou $2d$ pre parameter μ.

2.2 Sekvenčný Waldov postup

Tento postup je popísaný v skriptách Hušková (1982, str. 109).

Buď $\{X_i\}_{i=1}^{+\infty}$ postupnosť nezávislých rovnako rozdelených náhodných veličín, X_i má hustotu $f(x;\theta)$, kde $\theta \in (\underline{\theta}, \overline{\theta})$, θ je neznámy parameter. Riešme úlohu konštrukcie sekvenčného intervalového odhadu pre parameter θ danej dĺžky $2d$ s daných koeficientom spoľahlivosti $1 - \alpha$, kde $d > 0$ je známe.

Pri riešení tejto úlohy používame trojnú postupnosť $\{L_n, U_n, \varphi_n\}$, kde dvojica $(L_n = L_n(X_1, \ldots, X_n), U_n = U_n(X_1, \ldots, X_n))$ je nesekvenčný intervalový odhad parametru θ a $\varphi_n = \varphi_n(X_1, \ldots, X_n)$ je náhodná veličina určujúca pravidlo pre ukončenie výberu. Veličina φ_n nadobúda hodnoty z $\{0, 1\}$, pričom ak $\varphi_n = 1$, tak ukončíme výber, inak pokračujeme vo výbere. Sekvenčný intervalový odhad pre parameter θ je potom interval

$$\left(L_N(X_1, \ldots, X_N), U_N(X_1, \ldots, X_N)\right),$$

kde $N = \min\{n;\ \varphi_n(X_1, \ldots, X_n) = 1\}$. Náhodná veličina N je rozsah konečného výberu. Očakávame, že tento intervalový odhad má koeficient spoľahlivosti $1-\alpha$, ak platí:

$$\mathsf{P}\left(\theta \in (L_N, U_N); \theta\right) = 1 - \alpha.$$

Predpokladajme, že štatistiky L_n a U_n, $n = 1, 2, \ldots$, majú nasledujúce vlastnosti:

i) $L_n < U_n$ s.i.,

ii) $\sqrt{n}\,(U_n - L_n) \xrightarrow[n\to\infty]{s.i.} 2Au_{1-\frac{\alpha}{2}}, \quad A > 0,$

iii) $\sqrt{n}\,(U_n - \theta) - Z_n A - Au_{1-\frac{\alpha}{2}} \xrightarrow[n\to\infty]{s.i.} 0,$

kde Z_n je štandardizovaný súčet iid náhodných veličín s konečným druhým momentom.

Definujeme pravidlo pre ukončenie výberu, ktoré závisí na voľbe d:

$$N(d) = \min\{n;\ U_n - L_n \leq 2d\}, \quad d > 0. \tag{2.5}$$

Potom hľadaný sekvenčný intervalový odhad pre parameter θ dĺžky menšej alebo rovnej $2d$ je $(L_{N(d)}, U_{N(d)})$.

Za predpokladu, že náhodné veličiny $\{N(d)\ d^2;\ d > 0\}$ sú rovnomerne integrovateľné, tj. $\lim_{n\to\infty}\sup_{0<d<d_0} \mathsf{E}\, N(d)\ d^2\ \mathbb{I}_{[N(d)d^2>n]} = 0$, kde

$$\mathbb{I}_{[N(d)d^2>n]} = \begin{cases} 1, & \text{ak } N(d)d^2 > n, \\ 0, & \text{inak,} \end{cases}$$

platia ďalšie vlastnosti.

Veta 2.3. *Nech sú splnené vyššie uvedené predpoklady. Potom pre všetky $\theta \in (\underline{\theta}, \overline{\theta})$ platí:*

a) $N(d) < +\infty$ *s.i.,* $\mathsf{E}_\theta\, N(d) < +\infty$ *pre všetky* $d > 0$,
$N(d)$ *je nerastúca funkcia premennej* d,
$\lim_{d\to 0} N(d) = +\infty$ *s.i.,* $\lim_{d\to 0} \mathsf{E}_\theta\, N(d) = +\infty$,

b) $\lim_{d\to 0} N(d)d^2 = A^2 u_{1-\frac{\alpha}{2}}{}^2$ *s.i.,*
$\lim_{d\to 0} \mathsf{E}_\theta\, N(d)d^2 = A^2 u_{1-\frac{\alpha}{2}}{}^2$,

c) $\lim_{d\to 0} P\left(L_{N(d)} \leq \theta \leq U_{N(d)}\right) = 1 - \alpha$.

Dôkaz. Viď Hušková (1982, str. 110-111).

□

Tvrdenie a) nám zaručuje, že tento postup bude pri danom d ukončený s pravdepodobnosťou 1. Ďalej hovorí, že pri $d \to 0$ rastie rozsah výberu $N(d)$ nad všetky medze a tvrdenie b) hovorí, že rastie rádovo rovnako rýchlo ako d^{-2}. Tvrdenie c) hovorí, že intervalový odhad je o asymptotickej spoľahlivosti $1 - \alpha$.

Nech $X_1, \ldots, X_n$ je náhodný výber z rozdelenia so strednou hodnotou μ a rozptylom $0 < \sigma^2 < +\infty$, kde μ a σ^2 sú neznáme parametre. Chceme intervalový odhad s koeficientom spoľahlivosti $1-\alpha$ a dĺžkou menšou alebo rovnou $2d$ pre parameter μ. Zvoľme

$$\begin{aligned} L_n &= \overline{X}_n - u_{1-\frac{\alpha}{2}}\ \frac{S_n}{\sqrt{n}}, \\ U_n &= \overline{X}_n + u_{1-\frac{\alpha}{2}}\ \frac{S_n}{\sqrt{n}}, \\ N(d) &= \min\left\{n;\ U_n - L_n \leq 2d\right\}. \end{aligned} \tag{2.6}$$

Sekvenčný intervalový odhad dĺžky menšej alebo rovnej $2d$ o približnej spoľahlivosti $1-\alpha$ pre μ je tvaru

$$\left(\overline{X}_{N(d)} - \frac{S_{N(d)} u_{1-\frac{\alpha}{2}}}{\sqrt{N(d)}},\ \overline{X}_{N(d)} + \frac{S_{N(d)} u_{1-\frac{\alpha}{2}}}{\sqrt{N(d)}}\right).$$

Dokážeme, že pre túto voľbu trojnej postupnosti sú splnené predpoklady Vety 2.2 a teda okrem iného platí, že pre $d \to 0$ má tento intervalový odhad asymptotickú spoľahlivosť $1-\alpha$.

Predpoklad (i) platí z toho, že

$$u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} > 0\ s.i.$$

vzhľadom k tomu, že $\sigma^2 > 0$. Zo SZVČ plynie: $S_n^2 \xrightarrow[n\to\infty]{s.i.} \sigma^2$, čo sme dokázali v kapitole 1.3. Odtiaľ platí aj $S_n \xrightarrow[n\to\infty]{s.i.} \sigma$. V (ii) je

$$\sqrt{n}(U_n - L_n) = 2u_{1-\frac{\alpha}{2}} S_n \xrightarrow[n\to\infty]{\text{SZVČ},s.i.} 2\sigma u_{1-\frac{\alpha}{2}}.$$

Teda označme $A = \sigma > 0$. Definujme $Z_n = \frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \mu)$. Potom je výraz v predpoklade (iii) rovný výrazu:

$$\begin{aligned} \sqrt{n}(U_n - \mu) - \frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \mu)\ \sigma - \sigma u_{1-\frac{\alpha}{2}} &= \\ = (\overline{X}_n - \mu)(\sqrt{n} - n) + u_{1-\frac{\alpha}{2}}(S_n - \sigma) &\xrightarrow[n\to\infty]{\text{SZVČ},s.i.} 0. \end{aligned}$$

Nakoniec dokážeme, že náhodné veličiny $\{N(d)\ d^2;\ d > 0\}$ sú rovnomerne integrovateľné. Na to stačí ukázať, že $\lim_{n\to\infty} \mathsf{E}\, N(d)\ d^2\ \mathbb{I}_{[N(d)d^2>n]} = 0$ rovnomerne pre $0 < d < d_0$. Platí

$$\mathsf{E}\, N(d)\ d^2\ \mathbb{I}_{[N(d)d^2>n]} = \sum_{k>\frac{n}{d^2}} \mathsf{P}(N(d) > k) d^2.$$

Z definície $N(d)$ a z Čebyševovej nerovnosti, viď Dupač (2005, Věta 2.10 (ii)) platí:

$$\mathsf{P}(N(d) > k) = \mathsf{P}(S_k u_{1-\frac{\alpha}{2}} > d\sqrt{k}) \leq \frac{\mathsf{E}\, S_k^4 u_{1-\frac{\alpha}{2}}^4}{d^4 k^2} \leq \frac{C}{d^4 k^2},$$

kde C je konštanta nezávislá na d ani k. Plynie to z toho, že $\frac{(k-1)S_k^2}{\sigma^2} \sim \chi^2_{k-1}$ a platí, viď Anděl (2007, str. 27):

$$\begin{aligned}
\mathsf{E}\left(\frac{(k-1)S_k^2}{\sigma^2}\right) &= k-1, \quad \mathsf{var}\left(\frac{(k-1)S_k^2}{\sigma^2}\right) = 2(k-1),\\
\mathsf{E}\left(\frac{(k-1)S_k^2}{\sigma^2}\right)^2 &= \mathsf{var}\left(\frac{(k-1)S_k^2}{\sigma^2}\right) + \mathsf{E}\left(\frac{(k-1)S_k^2}{\sigma^2}\right) = 2(k-1) + (k-1)^2\\
&= (k-1)(k+1).
\end{aligned}$$

Keďže $u^4_{1-\frac{\alpha}{2}}$ je konštanta, platí

$$\begin{aligned}
\mathsf{E}\, S_k^4 u^4_{1-\frac{\alpha}{2}} = u^4_{1-\frac{\alpha}{2}}\, \mathsf{E}\, S_k^4 &= u^4_{1-\frac{\alpha}{2}} \left(\frac{\sigma^2}{k-1}\right)^2 \mathsf{E}\left[\frac{(k-1)S_k^2}{\sigma^2}\right]^2\\
&= u^4_{1-\frac{\alpha}{2}} \sigma^4 \left(\frac{k+1}{k-1}\right)
\end{aligned}$$

a $\frac{k+1}{k-1} \leq 3$, pre $k > 1$. Teda celá stredná hodnota sa dá odhadnúť konštantou C.

Dosadíme

$$\mathsf{E}\, N(d)\ d^2\ \mathbb{I}_{[N(d)d^2 > n]} \leq c \sum_{k > \frac{n}{d^2}} d^{-2} k^{-2} \leq \frac{c}{n}$$

a dostávame

$$\lim_{n\to\infty} \mathsf{E}\, N(d)\ d^2\ \mathbb{I}_{[N(d)d^2 > n]} \leq \lim_{n\to\infty} \frac{c}{n} = 0.$$

Dokázali sme, že pre našu špeciálnu voľbu trojnej postupnosti $\{L_n, U_n, \varphi_n\}$ platí Veta 2.3.

Kapitola 3

Vlastnosti náhodnej veličiny N

Venujme teraz pozornosť náhodnej veličine $N(d)$, ktorá označuje celkový rozsah výberu potrebný na konštrukciu sekvenčného intervalového odhadu dĺžky d a spoľahlivosti $1-\alpha$. Aké má rozdelenie a strednú hodnotu?

Simulujme úlohu konštrukcie intervalového odhadu parametru μ danej dĺžky $2d$ a spoľahlivosti $1-\alpha$. Cieľom tejto simulácie je získať informácie o rozdelení náhodnej veličiny $N(d)$. Vytvoríme histogram z nameraných hodnôt a zároveň spočítame aj ich strednú hodnotu a medián. Tieto hodnoty sú na konci tejto kapitoly v prehľadných Tabuľkách 3.1 a 3.2. Zvoľme $\alpha = 0,05$ a za d postupne vezmime hodnoty $d_1 = 0,5$; $d_2 = 0,3$; $d_3 = 0,1$; $d_4 = 0,05$.

3.1 Steinov postup

V kapitole 1.3 sme odvodili približné správanie sa strednej hodnoty veličiny N pre Steinov postup. Dostali sme výsledok, že pri malých hodnotách n (rozsah prvého výberu) a $\frac{z}{\sigma^2}$, kde z je definované v (1.7) a σ^2 je rozptyl veličín normálneho rozdelenia, máme približne

$$\mathsf{E}_{\sigma^2}(N) \simeq \frac{\sigma^2}{z}. \tag{3.1}$$

Pri voľbe n v závislosti na d vzťahom

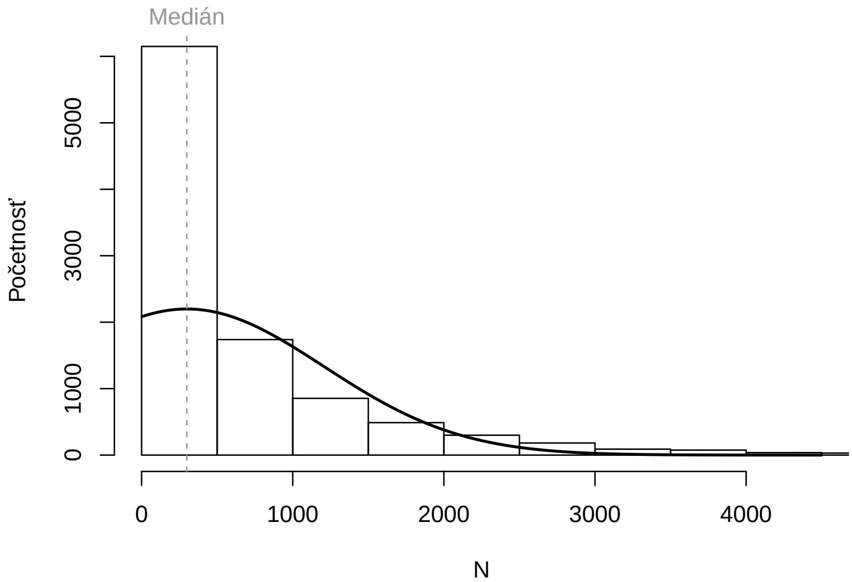
$$n = \left[\frac{1}{d}\right], \tag{3.2}$$

kde $[x]$ je celé číslo, ktoré vznikne zaokrúhlením reálneho čísla x, platí:

$$d \to 0 \quad \Rightarrow \quad n \to \infty.$$

Overme tieto výsledky pomocou simulácie, ktorej skript v jazyku R je uvedený v prílohe 1. Na zistenie hodnoty N použijeme Steinov postup. Rozsah prvého výberu volíme podľa pravidla (3.2) a hodnoty d a α máme určené na začiatku kapitoly 3. Výber robíme z normovaného normálneho rozdelenia.

V simulácii opakujeme krok I zo Steinovho postupu na výpočet celkového rozsahu 10 000 krát a ukladáme si zistené hodnoty N. Z nich následne vykreslíme histogram rozdelenia veličiny N spolu s hustotou normálneho rozdelenia, ktorého stredná hodnota sa zhoduje s mediánom získaných hodnôt veličiny N. Medián (med), priemer (mean) a hodnotu $\frac{\sigma^2}{z} = \frac{1}{z}$ si necháme v danom poradí vypísať pre porovnanie s odvodenou strednou hodnotou (3.1).



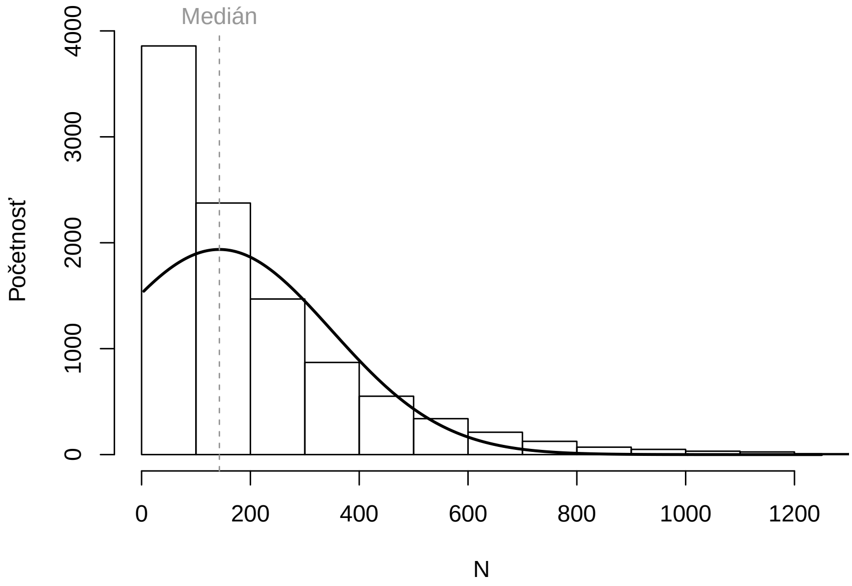
Graf 3.1: Rozdelenie hodnôt N pre výber z $\mathsf{N}(0,1)$ ak $d = 0,5$

Na Grafe 3.1 je histogram hodnôt N zisťovaných pomocou Steinovho postupu, ak $d = 0,5$. Vidíme, že rozdelenie veličiny N sa najviac líši od normálneho rozdelenia práve v okolí mediánu, kde je pravdepodobnosť oveľa väčšia než pri väčších hodnotách veličiny N a toto rozdelenie nie je symetrické. K tomuto grafu máme výstup:

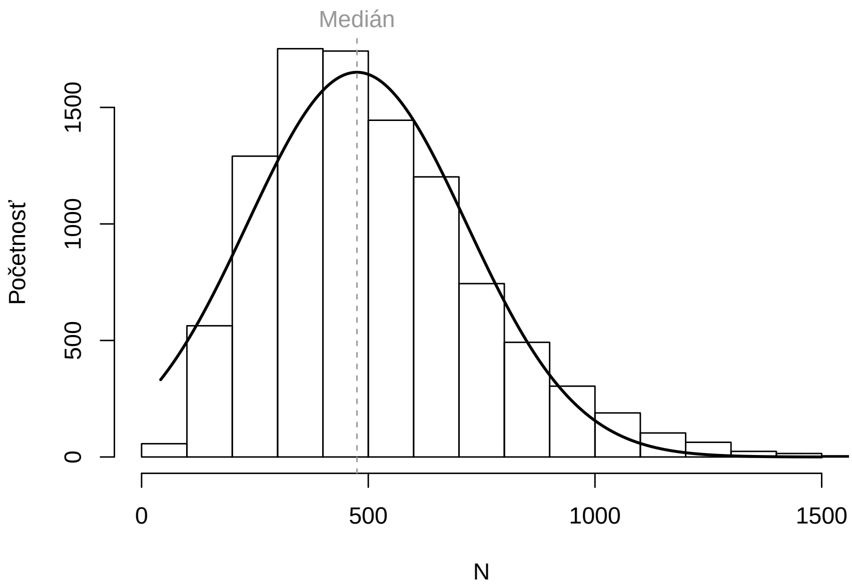
```
> med
[1] 299
> mean
[1] 645.8638
> 1 / z
[1] 645.7697
```

Stredná hodnota a odhad strednej hodnoty sú približne rovnaké, ale medián je menej než polovica z nich. Ku Grafu 3.2, ktorý zobrazuje histogram hodnôt veličiny N ak $d = 0,3$, je výstup:

```
> med
[1] 143
> mean
[1] 205.8439
> 1 / z
[1] 205.7312
```



Graf 3.2: Rozdelenie hodnôt N pre výber z $\mathsf{N}(0,1)$ ak $d = 0,3$



Graf 3.3: Rozdelenie hodnôt N pre výber z $\mathsf{N}(0,1)$ ak $d = 0,1$

Pre $d = 0,1$ je z Grafu 3.3 a výstupu:

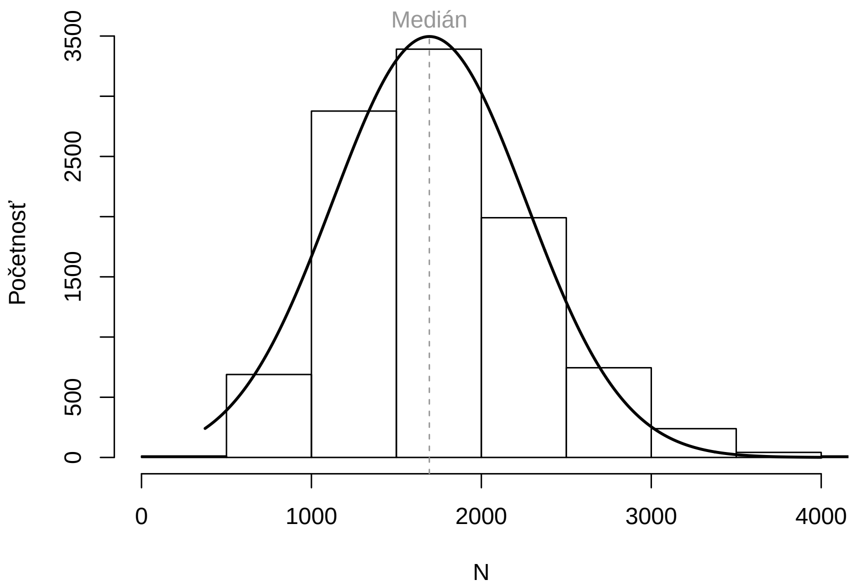
```
> med
[1] 475
> mean
```

```
[1] 512.6912
> 1 / z
[1] 511.6644
```

vidieť, že rozdelenie hodnôt N je s klesajúcim d stále viac podobné normálnemu rozdeleniu. To potvrdzuje aj Graf 3.4 a výstup pre $d = 0,05$:

```
> med
[1] 1683
> mean
[1] 1749.043
> 1 / z
[1] 1740.558
```

a vedie k záveru, že ak d klesá, tak stredná hodnota N rastie a medián sa k nej blíži.



Graf 3.4: Rozdelenie hodnôt N pre výber z $\mathsf{N}(0,1)$ ak $d = 0,05$

V simulácii sme teda dospeli k výsledkom zhodným so závermi kapitoly 1.3, ktoré navyše smerujú k tomu, že pre špeciálnu voľbu n, ako v (3.2), a pre $d \to 0$ má náhodná veličina N asymptoticky normálne rozdelenie.

3.2 Modifikovaný Steinov postup

Pozrime sa na to, ako dopadne naša simulácia, ak nebudeme robiť výber z normálneho rozdelenia.

3.2.1 Laplaceovo rozdelenie

Nech $a \in \mathbb{R}$ a $b > 0$. *Laplaceovo rozdelenie* má hustotu

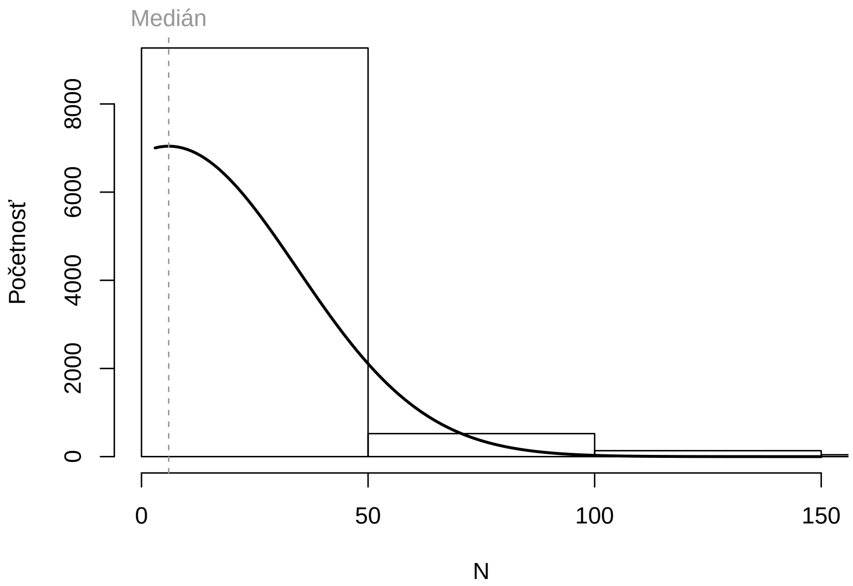
$$f(x) = \frac{1}{2b} exp \left\{ -\frac{|x-a|}{b} \right\}.$$

Toto rozdelenie sa nazýva aj dvojité exponenciálne rozdelenie a značí sa $\mathsf{DEx}(a, b)$. Ak $X \sim \mathsf{DEx}(a, b)$, potom platí

$$\mu = \mathsf{E}\, X = a, \quad \sigma^2 = \mathsf{var}\, X = 2b^2.$$

Zvoľme $a = 0$, $b = \frac{1}{\sqrt{2}}$. Teda $\mu = 0$, $\sigma^2 = 1$.

Upravme simuláciu z kapitoly 3.1 tak, že miesto výberu z normálneho rozdelenia budeme robiť výber z $\mathsf{DEx}\left(0, 1/\sqrt{2}\right)$ a použijeme modifikovaný Steinov postup. Znovu určíme rozsah prvého výberu pravidlom (3.2). R-ový skript k tejto simulácii nájdeme v prílohe 2.



Graf 3.5: Rozdelenie hodnôt N pre výber z $\mathsf{DEx}\left(0, 1/\sqrt{2}\right)$ ak $d = 0,5$

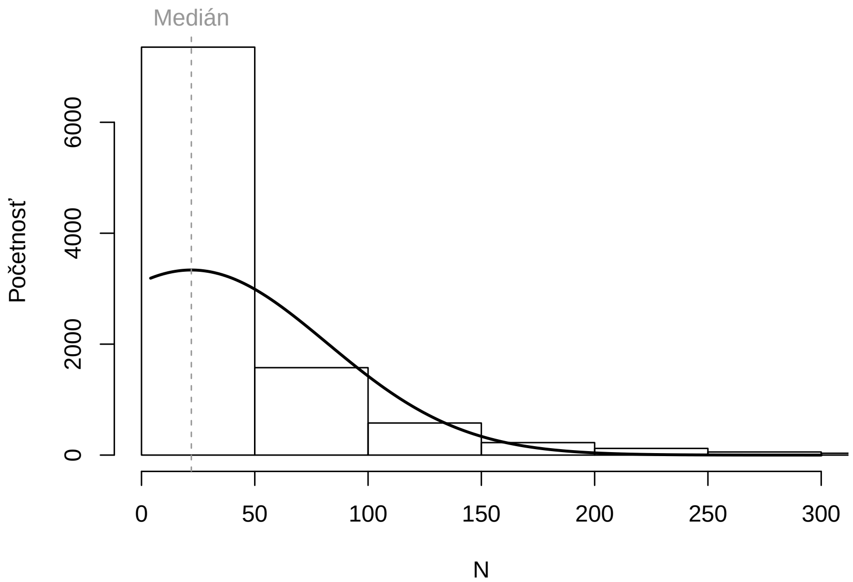
Na Grafe 3.5 a výstupe pre $d = 0,5$:

```
> med
[1] 6
> mean
[1] 16.2736
> 1 / z
[1] 15.3664
```

vidíme, že pri výbere z Laplaceovho rozdelenia stačí na konštrukciu intervalového odhadu pre parameter μ dĺžky $2d$ a spoľahlivosti $0,95$ neporovnatelne menší rozsah výberu, než pri výbere z normálneho rozdelenia. To isté platí aj pre $d = 0,3$ na Grafe 3.6 a vo výstupe:

```
> med
[1] 22
> mean
[1] 43.0336
> 1 / z
[1] 42.68444
```

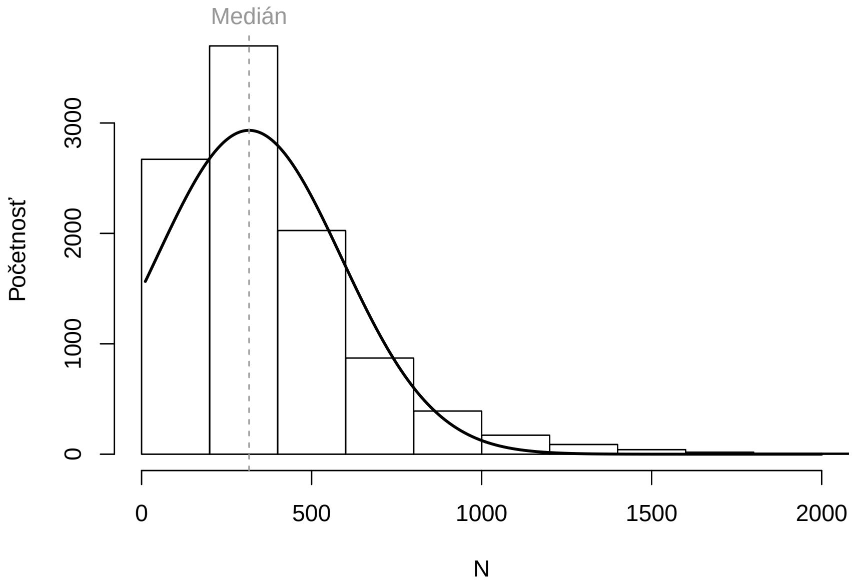
Je to spôsobené tým, že kým v Steinovom postupe využívame kvantily Studentovho rozdelenia, v jeho modifikovanom postupe využívame na základe CLV kvantil normovaného normálneho rozdelenia. Intervalový odhad, ktorý zostrojíme modifikovaným Steinovým postupom potom spĺňa požadovanú spoľahlivoť len asymptoticky.



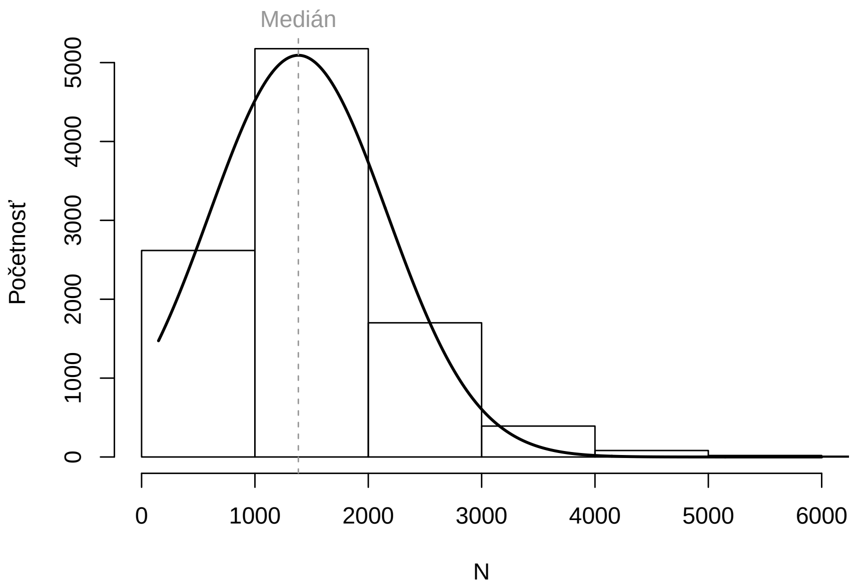
Graf 3.6: Rozdelenie hodnôt N pre výber z DEx $\left(0, 1/\sqrt{2}\right)$ ak $d = 0,3$

Keď porovnáme Graf 3.7 pre $d = 0,1$ a jeho výstup:

```
> med
[1] 316
> mean
[1] 381.3016
> 1 / z
[1] 384.16
```



Graf 3.7: Rozdelenie hodnôt N pre výber z $\mathsf{DEx}\left(0, 1/\sqrt{2}\right)$ ak $d = 0,1$



Graf 3.8: Rozdelenie hodnôt N pre výber z $\mathsf{DEx}\left(0, 1/\sqrt{2}\right)$ ak $d = 0,05$

a Graf 3.8 pre $d = 0,05$ a výstup:

```
> med
[1] 1383
> mean
```

```
[1] 1535.42
> 1 / z
[1] 1536.64
```

s výstupmi a grafmi z kapitoly 3.1, môžeme usúdiť, že pre $d \to 0$ bude mať náhodná veličina N pre výber z Laplaceovho rozdelenia rovnaké asymptotické vlastnosti ako pre výber z normálneho rozdelenia.

3.2.2 Rozdelenie χ^2 o 4 stupňoch voľnosti

Nech $n \geq 1$. *Rozdelenie* χ^2 o n stupňoch voľnosti má hustotu

$$f(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x > 0.$$

Značíme χ^2_n. Ak $X \sim \chi^2_n$, tak platí

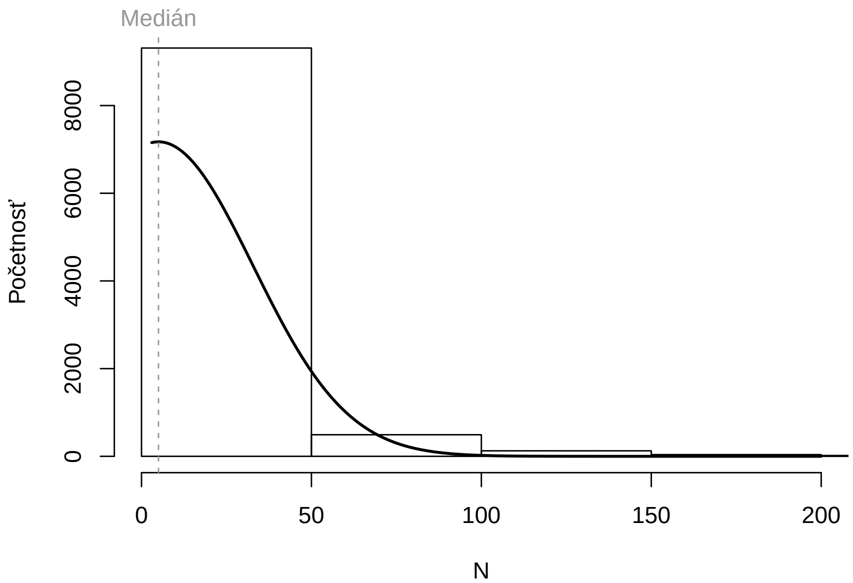
$$\mu = \mathsf{E}\, X = n, \quad \sigma^2 = \mathsf{var}\, X = 2n.$$

Pre voľbu $n = 4$ je $\mu = 4$ a $\sigma^2 = 8$. Označme

$$Z = \frac{X - 4}{\sqrt{8}}.$$

Potom náhodná veličina Z má normované χ^2_4-rozdelenie.

Teraz upravme simuláciu z kapitoly 3.1 tak, že miesto výberu z normálneho rozdelenia vezmeme výber z normovaného χ^2_4-rozdelenia a použijeme modifikovaný Steinov postup a pravidlo (3.2) na určenie rozsahu prvého výberu. R-kový skript k tejto simulácii nájdeme v prílohe 3.



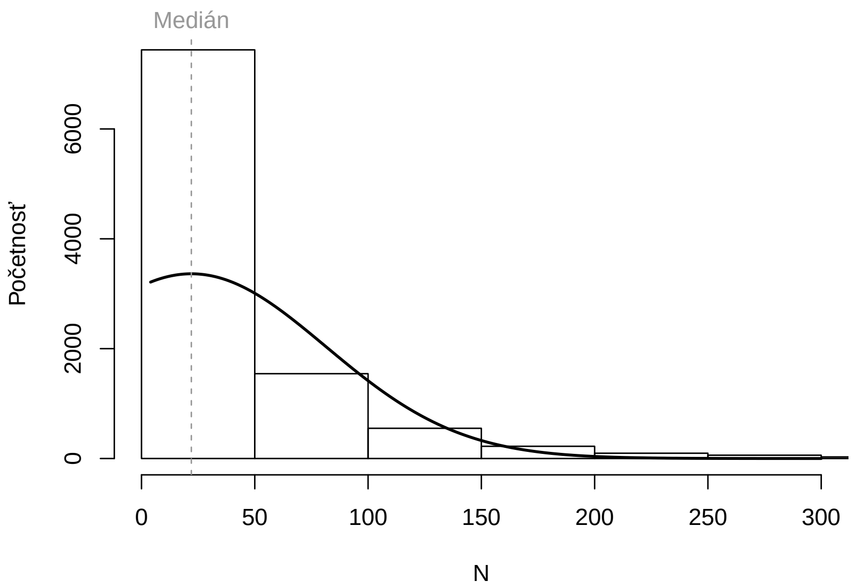
Graf 3.9: Rozdelenie hodnôt N pre výber z normovaného χ^2_4-rozdelenia ak $d = 0,5$

Dostaneme Grafy 3.9 - 3.12 a výstupy pre $d = 0,5$:

```
> med
[1] 5
> mean
[1] 15.84
> 1 / z
[1] 15.3664
```

$d = 0,3$:

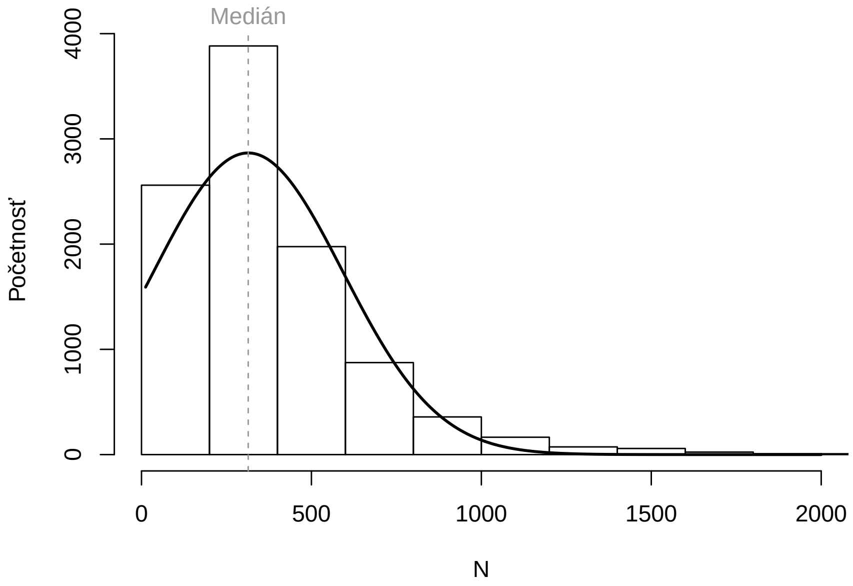
```
> med
[1] 22
> mean
[1] 42.2926
> 1 / z
[1] 42.68444
```



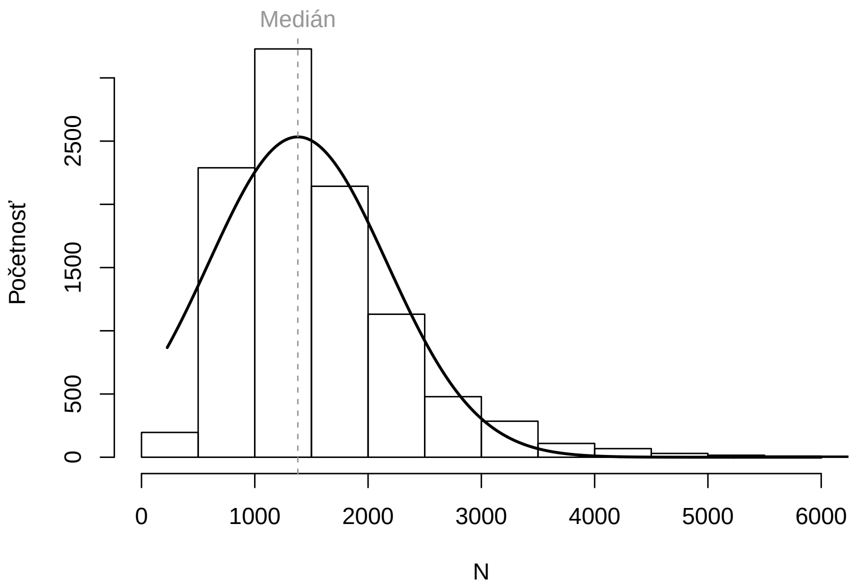
Graf 3.10: Rozdelenie hodnôt N pre výber z normovaného χ^2_4-rozdelenia ak $d = 0,3$

$d = 0,1$:

```
> med
[1] 314
> mean
[1] 383.9397
> 1 / z
[1] 384.16
```



Graf 3.11: Rozdelenie hodnôt N pre výber z normovaného χ^2_4-rozdelenia ak $d = 0,1$



Graf 3.12: Rozdelenie hodnôt N pre výber z normovaného χ^2_4-rozdelenia ak $d = 0,05$

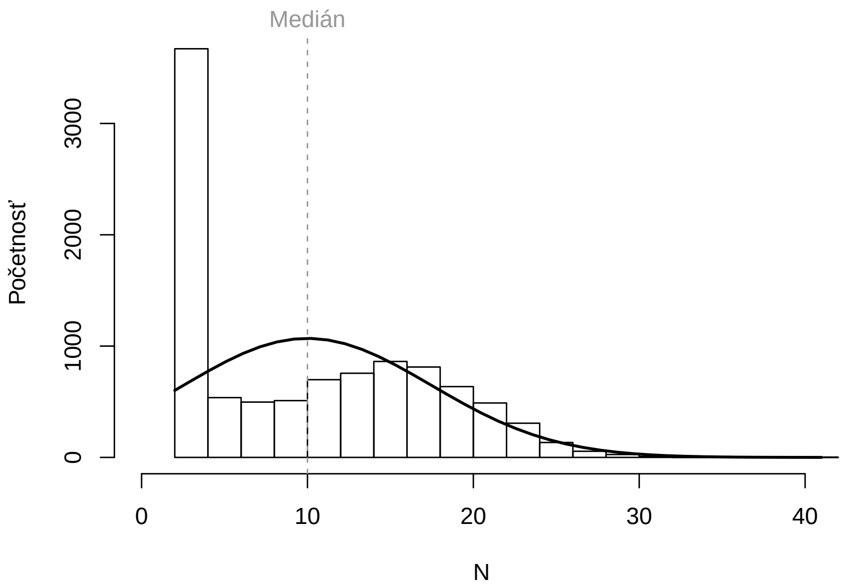
a $d = 0,05$:

```
> med
[1] 1380
> mean
[1] 1539.945
> 1 / z
[1] 1536.64
```

Môžeme si všimnúť značnú podobnosť s grafmi a výstupmi pri výbere z Laplaceovho rozdelenia.

3.3 Sekvenčný Waldov postup

Nakoniec simulujme sekvenčný Waldov postup pre intervalový odhad parametru μ danej dĺžky $2d$ a spoľahlivosti $1-\alpha$. Robíme výber z normovaného normálneho rozdelenia a použijeme špeciálnu voľbu trojnej postupnosti definovanú v (2.6). Parametre α, d a n (rozsah prvého výberu definovaný v (3.2)) vezmeme ako vyššie. Skript v jazyku R je uvedený v prílohe 4.

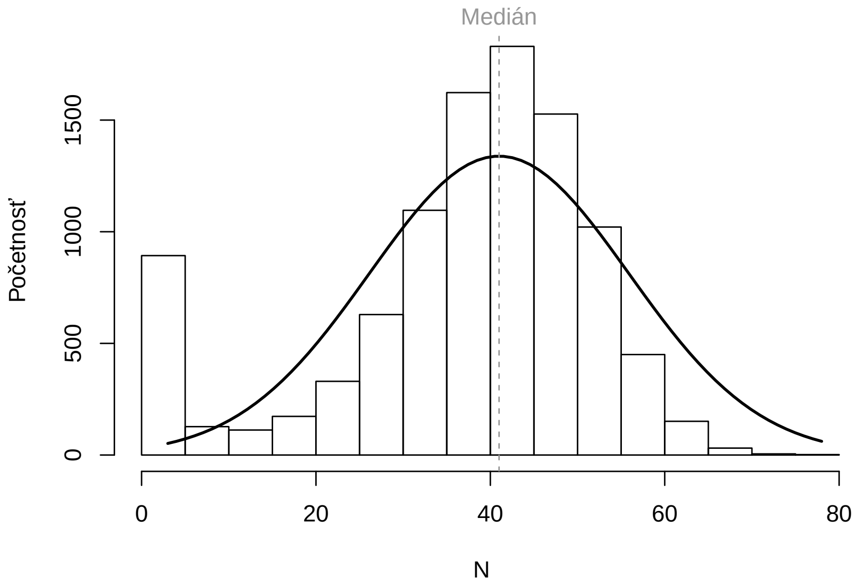


Graf 3.13: Rozdelenie hodnôt N pri sekvenčnom Waldovom postupe ak $d = 0,5$

Pre voľbu $d = 0,5$ je na Grafe 3.13 vidieť, že rozdelenie hodnôt N sa ani trochu nepribližuje normálnemu rozdeleniu a dokonca ani neplatí, že s najväčšou pravdepodobnosťou sú dosahované hodnoty v okolí mediánu, ako to bolo v grafoch pri Steinovom postupe, viď Graf 3.1. Výstup pre $d = 0,5$ je následovný:

```
> med
[1] 10
> mean
[1] 10.1834
```

Vidíme, že tento postup je podľa priemeru hodnôt veličiny N jednoznačne výhodnejší než Steinov postup. V Grafe 3.14 pre $d = 0,3$ majú najväčšiu pravdepodobnosť hodnoty N v okolí mediánu, ale stále je rozdelenie hodnôt veličiny N dosť odlišné od normálneho rozdelenia.



Graf 3.14: Rozdelenie hodnôt N pri sekvenčnom Waldovom postupe ak $d = 0,3$

Výstup pre $d = 0,3$:

```
> med
[1] 41
> mean
[1] 37.5843
```

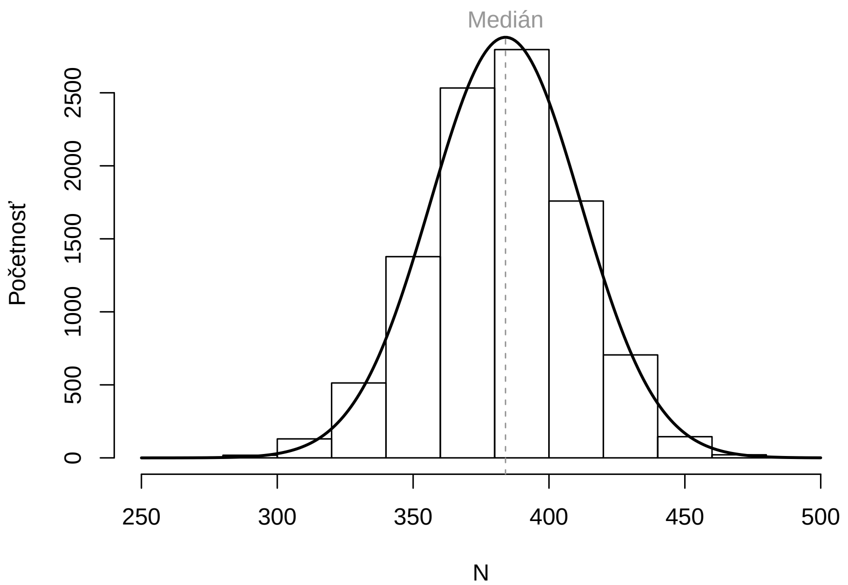
Všimnime si, že na rozdiel od výstupov zo Steinovho postupu a jeho modifikácii je v sekvenčnom Waldovom postupe stredná hodnota menšia než medián hodnôt veličiny N a zároveň sú blízko pri sebe. V Grafoch 3.15 a 3.16 nastala zmena oproti predošlým dvom grafom. Vidíme, že ak $d \to 0$ má náhodná veličina N asymptoticky normálne rozdelenie. Výstupy pre $d = 0,1$:

```
> med
[1] 384
> mean
[1] 383.0298
```

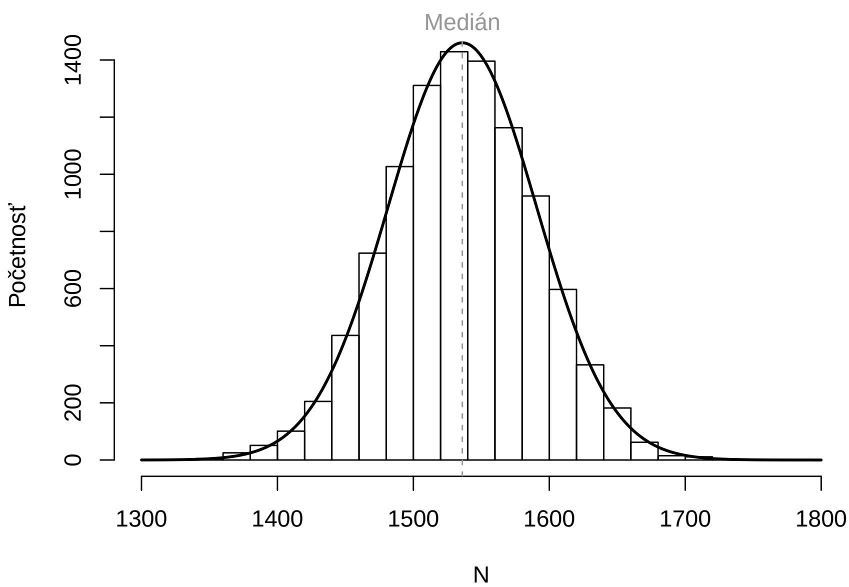
a pre $d = 0,05$:

```
> med
[1] 1536
> mean
[1] 1536.138
```

podkladajú tvrdenie z Vety 2.3 a), že pre $d \to 0$ rastie rozsah výberu N nad všetky medze.



Graf 3.15: Rozdelenie hodnôt N pri sekvenčnom Waldovom postupe ak $d = 0,1$



Graf 3.16: Rozdelenie hodnôt N pri sekvenčnom Waldovom postupe ak $d = 0,05$

3.4 Porovnanie jednotlivých postupov

V Tabuľkách 3.1 a 3.2 môžeme porovnať mediány a stredné hodnoty náhodnej veličiny N pre rôzne postupy a rôzne voľby čísla d.

Je zrejmé, že stredná hodnota N je najmenšia pre sekvenčný Waldov postup, teda je výhodnejší pri riešení našej úlohy. To plynie z toho, že na zostrojenie intervalového odhadu danej dĺžky a danej spoľahlivosti je potrebných priemerne menej pozorovaní než pri Steinovom postupe.

Ďalej je vidieť, že medián a stredná hodnota veličiny N sú pri Steinovom postupe a jeho modifikácii dosť rozdielne. To indikuje asymetriu rozdelenia, ktorú je možno pozorovať na Grafoch 3.1, 3.2, 3.5 - 3.8 a 3.9 - 3.12. Avšak ak $d \to \infty$, tak sa medián veličiny N blíži k jej strednej hodnote.

Pre $d \to 0$ sa stredné hodnoty a mediány náhodnej veličiny N pre rôzne postupy podobajú. Výhoda Steinovho postupu oproti sekvenčnému Waldovmu postupu je vo výpočtovej jednoduchosti.

Mediány veličiny N	**d = 0.5**	**d = 0.3**	**d = 0.1**	**d = 0.05**
Steinov postup	299	143	475	1683
Modifikovaný Steinov postup pre Laplaceovo rozdelenie	6	22	316	1383
Modifikovaný Steinov postup pre rozdelenie χ^2	5	22	314	1380
Sekvenčný Waldov postup	10	41	384	1536

Tabuľka 3.1: Medány veličiny N pre rôzne postupy a dĺžky intervalu

Stredné hodnoty veličiny N	**d = 0.5**	**d = 0.3**	**d = 0.1**	**d = 0.05**
Steinov postup	645,8638	205,8439	512,6912	1749,049
Modifikovaný Steinov postup pre Laplaceovo rozdelenie	16,2736	43,0336	381,3016	1535,42
Modifikovaný Steinov postup pre rozdelenie χ^2	15,84	42,2926	383,9397	1539,945
Sekvenčný Waldov postup	10,1834	37,5843	383,0298	1536,138

Tabuľka 3.2: Stredné hodnoty veličiny N pre rôzne postupy a dĺžky intervalu

Záver

V práci sme sa venovali sekvenčným intervalovým odhadom danej dĺžky a danej spoľahlivosti pre strednú hodnotu. Popísali sme rôzne postupy ich konštrukcie a odvodili ich vlastnosti.

Dospeli sme k záveru, že ak sa požadovaná dĺžka intervalu blíži k nule, tak rozsah výberu potrebný na zostrojenie sekvenčného intervalového odhadu danej dĺžky a danej spoľahlivosti rastie nad všetky medze. To sme dokázali pre všetky spomenuté postupy.

Pomocou simulácie jednotlivých postupov sme vytvorili grafy rozdelenia celkového rozsahu výberu a zhodnotili sme výhodnosť jednotlivých postupov vzhľadom na strednú hodnotu rozsahu výberu. Z grafov vyplýva, že čím menšia je požadovaná dĺžka intervalu, tým viac sa rozdelenie celkového rozsahu výberu blíži k normálnemu rozdeleniu.

V teórii sekvenčných intervalových odhadov danej dĺžky nie je známych toľko poznatkov, ako to je u nesekvenčných intervalových odhadov. Pre niektoré problémy však existuje riešenie iba na základe sekvenčnej metódy, ako je to aj v prípade intervalového odhadu strednej hodnoty pri neznámom rozptyle. Ďalej je možné skúmať tento problém a jeho riešenie v lineárnom regresnom modeli.

Zoznam použitej literatúry

ANDĚL, J. (2007). *Základy matematické statistiky.* Matfyzpress, Praha. ISBN 80-7378-001-1.

CHATTERJEE, S. K. (1991). Two-stage and multistage procedures. *Handbook of sequential analysis*, (3), 21–45.

DUPAČ, V.; HUŠKOVÁ, M. (2005). *Pravděpodobnost a matematická statistika.* Univerzita Karlova v Praze - Nakladatelství Karolinum, Praha. ISBN 978-80-246-0009-3.

GHOSH, B. K.; SEN, P. K. (1991). *Handbook of sequential analysis.* Marcel Dekker, New York. ISBN 0-8247-8408-1.

HUŠKOVÁ, M. (1982). *Sekvenční analýza.* SPN, Praha. Skripta.

LACHOUT, P. (2004). *Teorie pravděpodobnosti.* Univerzita Karlova v Praze - Nakladatelství Karolinum, Praha. ISBN 80-246-0872-3.

RAO, C. (1978). *Lineární metody statistické indukce a jejich aplikace.* Academia, Praha. Skripta.

RÉNYI, A. (1972). *Teorie pravděpodobnosti.* Academia, Praha.

WONNACOT, T. H.; WONNACOT, R. J. (1993). *Statistika pro obchod a hospodářství.* Victoria Publishing a.s., Červený Kostelec. ISBN 80-85605-09-0.

Zoznam tabuliek

Zoznam grafov

Zoznam použitých skratiek

$\mathbb{R}$	reálne čísla
$\mathbb{R}^+$	kladné reálne čísla
max	maximum
sup	supremum
$\forall$	všeobecný kvantifikátor
$\lfloor x \rfloor$	dolná celá časť čísla x
$\lvert x \rvert$	absolútna hodnota čísla x
$[x]$	celé číslo, ktoré vznikne zaokrúhlením reálneho čísla x
$\mathbb{I}_{[A]}$	identifikátor javu A
med	medián
mean	stredná hodnota
$\mathsf{E}\,X$	stredná hodnota náhodnej veličiny X
$\mathsf{var}\,X$	rozptyl náhodnej veličiny X
$\mathsf{P}(A)$	pravdepodobnosť javu A
s.i.	skoro iste
iid	nezávislé a rovnako rozdelené
CLV	centrálna limitná veta
SZVČ	silný zákon veľkých čísel
$\mathcal{L}(X)$	všeobecné označenie rozdelenia náhodnej veličiny X
$\sim$	vyjadruje vzťah, napr. $X \sim \mathcal{L}(X)$ značí, že X má rozdelenie $\mathcal{L}(X)$
$\times$	kartézsky súčin
$N(\mu, \sigma^2)$	normálne rozdelenie so strednou hodnotou μ a rozptylom σ^2
Φ	distribučná funkcia normovaného normálneho rozdelenia
u_β	β-kvantil normovaného normálneho rozdelenia
t_n	Studentovo rozdelenie o n stupňoch voľnosti
H_n	distribučná funkcia Studentovho rozdelenia o n stupňoch voľnosti
$t_{\beta,n}$	β-kvantil Studentovho rozdelenia o n stupňoch voľnosti
χ^2_n	χ^2-rozdelenie o n stupňoch voľnosti
$\mathsf{DEx}(a, b)$	Laplaceovo/dvojité exponenciálne rozdelenie s parametrami a,b

Prílohy

Príloha 1.

```
set.seed(123)
d <- 0.05        # polovica dlzky intervalu
n <- round(1 / d, 0)
alfa <- 0.05
# 1-(alfa/2) kvantil t-rozdelenia o n-1 stupnoch volnosti
# zavisi na volbe premennej d
t <- 2.086
z <- (d / t)^2
B <- 10000       # pocet vyberov
N <- numeric(B)  # vektor o B nulach

for(i in 1:B) {
# Urobime vyber n nahodnych velicin z normovaneho normalneho
# rozdelenia a priradime do vektoru X.
X <- rnorm(n, mean = 0, sd = 1)
# Vypocitame N[i] a ulozime do pola.
N[i] <- max(n+1, as.integer(var(X) / z) + 1)
}
# Nechame vypisat median, strednu hodnotu z N a 1/z.
med <- median(N)
med
mean <- mean(N)
mean
y <- 1 / z
y
# Vykreslime histogram pre vektor N a hustotu normalneho rozdelenia
# do grafu
maxN <- max(N)  # horna hranica x-ovej osi v grafe hodnot N
minN <- min(N)  # dolna hranica x-ovej osi v grafe hodnot N
range <- maxN - minN
postscript(file = "histogram.ps", width = 6.5, height = 5,
                horizontal = FALSE)
par(mfrow = c(1,1))
# Histogram
h <- hist(N, xlab="N", ylab="Početnosť", main=NULL, xlim=c(0,maxN))
xfit <- seq(minN, maxN, length=range)
yfit <- dnorm(xfit, mean=med, sd=sd(N))
yfit <- yfit*diff(h$mids[1:2])*length(N)
lines(xfit, yfit, lwd=2) # hustota normalneho rozdelenia
abline(v = med, lty=2, col="#989898") # vertikalna ciara - median
# Oznacime median
mtext("Medián", side=3, line=0, col="#989898",at = med)
dev.off()
```

Príloha 2.

```
set.seed(123)
d <- 0.05          # polovica dlzky intervalu
n <- round(1 / d, 0)
alfa <- 0.05
# 1-(alfa/2) kvantil normovaneho normalneho rozdelenia
u <- 1.960
z <- (d / u)^2
B <- 10000         # pocet vyberov
N <- numeric(B)    # vektor o B nulach

for(i in 1:B) {
# Urobime vyber n nahodnych velicin z Laplaceovho rozdelenia pre a=0,
# b=1/sqrt(2) a priradime do vektoru X.
X <- rlaplace(n, location=0, scale=1/sqrt(2))
# Vypocitame N[i] a ulozime do pola.
N[i] <- max(n+1, as.integer(var(X) / z) + 1)
}
# Nechame vypisat median, strednu hodnotu z N a 1/z.
med <- median(N)
med
mean <- mean(N)
mean
y <- 1 / z
y
# Vykreslime histogram pre vektor N a hustotu normalneho rozdelenia
# do grafu
maxN <- max(N)  # horna hranica x-ovej osi v grafe hodnot N
minN <- min(N)  # dolna hranica x-ovej osi v grafe hodnot N
range <- maxN - minN
postscript(file = "histogram.ps", width = 6.5, height = 5,
                horizontal = FALSE)
par(mfrow = c(1,1))
# Histogram
h <- hist(N, xlab="N", ylab="Početnosť", main=NULL, xlim=c(0,maxN))
xfit <- seq(minN, maxN, length=range)
yfit <- dnorm(xfit, mean=med, sd=sd(N))
yfit <- yfit*diff(h$mids[1:2])*length(N)
lines(xfit, yfit, lwd=2) # hustota normalneho rozdelenia
abline(v = med, lty=2, col="#989898") # vertikalna ciara - median
# Oznacime median
mtext("Medián", side=3, line=0, col="#989898",at = med)
dev.off()
```

Príloha 3.

```
set.seed(123)
d <- 0.05          # polovica dlzky intervalu
n <- round(1 / d, 0)
alfa <- 0.05
# 1-(alfa/2) kvantil normovaneho normalneho rozdelenia
u <- 1.960
z <- (d / u)^2
B <- 10000         # pocet vyberov
N <- numeric(B)    # vektor o B nulach

for(i in 1:B) {
# Urobime vyber n nahodnych velicin z Chi-kvadrat rozdelenia
# o 4 stupnoch volnosti a priradime do vektoru Y. Normovane hodnoty
# nahodnych velicin z vyberu ulozime do vektoru X.
Y <- rchisq(n, 4, ncp=0)
X <- (Y - 4) / sqrt(8)
# Vypocitame N[i] a ulozime do pola.
N[i] <- max(n+1, as.integer(var(X) / z) + 1)
}
# Nechame vypisat median, strednu hodnotu z N a 1/z.
med <- median(N)
med
mean <- mean(N)
mean
y <- 1 / z
y
# Vykreslime histogram pre vektor N a hustotu normalneho rozdelenia
# do grafu
maxN <- max(N)   # horna hranica x-ovej osi v grafe hodnot N
minN <- min(N)   # dolna hranica x-ovej osi v grafe hodnot N
range <- maxN - minN
postscript(file = "histogram.ps", width = 6.5, height = 5,
                horizontal = FALSE)
par(mfrow = c(1,1))
# Histogram
h <- hist(N, xlab="N", ylab="Početnost’", main=NULL, xlim=c(0,maxN))
xfit <- seq(minN, maxN, length=range)
yfit <- dnorm(xfit, mean=med, sd=sd(N))
yfit <- yfit*diff(h$mids[1:2])*length(N)
lines(xfit, yfit, lwd=2) # hustota normalneho rozdelenia
abline(v = med, lty=2, col="#989898") # vertikalna ciara - median
# Oznacime median
mtext("Medián", side=3, line=0, col="#989898",at = med)
dev.off()
```

Príloha 4.

```
set.seed(123)
d <- 0.05          # polovica dlzky intervalu
n  <- round(1/d,0)
alfa <- 0.05
# 1-(alfa/2) kvantil normovaneho normalneho rozdelenia
u <- 1.960
B <- 10000         # pocet vyberov
N <- numeric(B)    # vektor o B nulach

for(i in 1:B) {
# Vygenerujeme nahodnu velicinu z normovaneho normalneho rozdelenia
# a priradime do vektoru X.
X <- rnorm(n, mean = 0, sd = 1)
m <- n                   # rozsah vyberu

while((u*sd(X)/sqrt(m)) > d) {
Y <- rnorm(1, mean = 0, sd = 1)
X <- c(X,Y)
m <- m+1
}
N[i] <- m
}
# Nechame vypisat median, strednu hodnotu z N a 1/z.
med <- median(N)
med
mean <- mean(N)
mean
# Vykreslime histogram pre vektor N a hustotu normalneho rozdelenia
# do grafu
maxN <- max(N)   # horna hranica x-ovej osi v grafe hodnot N
minN <- min(N)   # dolna hranica x-ovej osi v grafe hodnot N
range <- maxN - minN
postscript(file = "histogram.ps", width = 6.5, height = 5,
                horizontal = FALSE)
par(mfrow = c(1,1))
# Histogram
h <- hist(N, xlab="N", ylab="Početnosť", main=NULL, xlim=c(0,maxN))
xfit <- seq(minN, maxN, length=range)
yfit <- dnorm(xfit, mean=med, sd=sd(N))
yfit <- yfit*diff(h$mids[1:2])*length(N)
lines(xfit, yfit, lwd=2) # hustota normalneho rozdelenia
abline(v = med, lty=2, col="#989898") # vertikalna ciara - median
# Oznacime median
mtext("Medián", side=3, line=0, col="#989898",at = med)
dev.off()
```