

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITA KOMENSKÉHO, BRATISLAVA

MERANIE VÝKONNOSTI STOCHASTICKÝCH
ALGORITMOV PRI HĽADANÍ KOREŇA



KATEDRA MATEMATICKEJ ANALÝZY A NUMERICKEJ
MATEMATIKY
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITA KOMENSKÉHO, BRATISLAVA

MERANIE VÝKONNOSTI STOCHASTICKÝCH ALGORITMOV PRI HĽADANÍ KOREŇA

(bakalárska práca)

SAMUEL HAPÁK

Študijný program: Matematika

Študijný odbor: 9.1.1. Matematika

Vedúci: RNDr. Tomáš Kulich, PhD.

Kód práce: f5e2ded2-63ac-4ed3-a734-02d67baea342

Bratislava, 2011

Čestne prehlasujem, že som túto bakalársku prácu
vypracoval samostatne s použitím citovaných zdrojov.

.....

Abstrakt

SRFP (Stochastic Root Finding Problem), špeciálne jeho binárna verzia, je problém, ktorý sa opakovane zjavuje v rôznych oblastiach vedeckej činnosti. Od roku 1951, kedy Robbins a Monro [1] predstavili prakticky prvý stochastický algoritmus na jeho riešenie, vyšlo množstvo článkov teoreticky analyzujúcich tento a iné algoritmy. Napriek tomu je možné nájsť veľmi málo prác, ktoré by skúmali správanie sa a vhodnosť týchto algoritmov v reálnom nasadení. Teoretické výsledky o asymptotickom správaní sa algoritmov a konvergencii sú nepochybne istým indikátorom jeho kvality, nedokážu však nikdy plne zodpovedať otázku, ako sa bude algoritmus správať v praxi. Preto sme sa v tejto práci zamerali na praktické otestovanie algoritmov pri riešení binárneho SRFP a pomôcť tak pri ich reálnom nasadení.

Merania poukázali na slabiny algoritmov, Stochastická verzia Newtonovho algoritmu [2] sa dokonca ukázala úplne nepoužiteľnou. Po zakomponovaní istých heuristík sa nám ju však podarilo podstatne vylepšiť. V práci sme taktiež predstavili vlastný návrh algoritmu s veľkým potenciálom ďalších zlepšení, ktorý sa za niektorých podmienok ukázal ako výkonnostne najlepší spomedzi algoritmov.

Kľúčové slová: Stochastic root finding problem, Robbins-Monro, Stochastic Newton-Rapshon

Obsah

1	Prehľad existujúcich algoritmov	10
1.1	Robbins-Monro algoritmus	10
1.2	Dôkaz konverencie Robbins-Monro algoritmu	10
1.3	Stochastická Newton-Raphsonová metóda	15
2	Motivácia binárneho SRFP	17
2.1	Prečo ľudia konajú dobro?	17
2.2	Definícia binárneho SRFP	18
3	Stochastická metóda regula falsi	19
3.1	Motivácia	19
3.2	Stochastická verzia	19
3.3	Dôkaz konverencie	21
3.4	Voľba funkcie s	23
4	Metodika testovania	24
4.1	Dodatočné predpoklady pri implementácii algoritmov	24
4.2	Testované parametre	25
4.3	Stochastický Newtonov algoritmus	26
4.4	Voľba zahadzovacej funkcie pre Newtonov algoritmus	27
4.5	Voľba funkcie s pre metódu regula falsi	27
4.6	Distribúcia výsledkov predpovedaných algoritmami	28
4.7	Asymptotické správanie sa algoritmov	29
4.8	Schopnosť vysporiadať sa s monotónnymi oblasťami	29
4.9	Testy na ďalších funkciách	30
5	Záver	31

A	Prehľad testovaných funkcií	32
B	Výsledky meraní	36
B.1	Voľba parametrov pre metódy Newtonova a regula falsi	36
B.2	Tvar distribučnej funkcie	38
B.3	Asymptotické správanie sa algoritmov	43
B.4	Monótonne oblasti	48
B.5	Rôzne funkcie	53

Úvod

Numerické metódy sú vďaka rozmachu počítačovej techniky v dvadsiatom prvom storočí nepostrádateľným nástrojom modernej vedy. Dôležitým reprezentantom sú algoritmy na hľadanie koreňov, používané dennodenne v takmer všetkých oblastiach ľudského skúmania. Kým však Newtonova metóda, metóda regula falsi, či Brentova metóda, poskytujú efektívny spôsob hľadania nulových bodov známej funkcie f , pri skúmaní komplikovaných javov je možné určiť hodnoty f len experimentálne s nezanedbateľnou náhodnou chybou.

Uvedme jednoduchý príklad: analýza účinkov nového chemoterapeutika pri liečbe rakoviny. Pri väčšom množstve látky je šanca na zahubenie rakovinových buniek väčšia, no rovnako pribúdajú aj nežiadúce účinky. Cieľom farmaceutov je určiť množstvo látky x^* , pri ktorom je pravdepodobnosť vyliečenia sa väčšia ako vopred stanovená hodnota p . Z podstaty experimentu nevieme merať funkciu pravdepodobnosti $f(x)$ v závislosti od množstva látky x priamo, výsledok každého merania je binárny – úspech/neúspech. Jedna z možností, ako riešiť rovnicu $f(x) = p$, je uskutočniť veľa experimentov, dostatočne presne určiť funkciu f a následne použiť niektorú z numerických metód. Takýto prístup by však stál život priveľa laboratórnych myší, s čím ochrancovia zvierat nemusia súhlasiť a lobingom v europarlamente by mohli dosiahnuť zákaz nášho výskumu.

V tejto práci ukážeme efektívnejší spôsob riešenia uvedeného problému. Tieto takzvané SRFP (Stochastic root finding problem) algoritmy nevyžadujú znalosť funkcie f , jej koreň hľadajú len na základe náhodných meraní $F(x)$, tie vo všeobecnosti nemusia byť obmedzené na dve hodnoty ako v našom príklade, ak to tak však je, hovoríme o binárnom SRFP. Bližšie si binárne

SRFP predstavíme na konkrétnom probléme z teoretickej biológie v druhej časti práce.

Prvú metódu na riešenie SRFP navrhli v roku 1951 Robbins a Monro [1]. V roku 1977 ju Dan Anbar [2] modifikoval, pričom sa nechal inšpirovať klasickým Newtonovým algoritmom. S oboma metódami sa bližšie zoznámime v prvej časti bakalárskej práce. Následne v tretej časti predstavíme vlastný algoritmus na riešenie binárneho SRFP a v záverečnej štvrtej časti otestujeme výkonnosť všetkých troch algoritmov za rôznych podmienok.

Implementácie algoritmov použité v tejto práci sprístupňujeme pod licenciou GPL vo forme jednoduchého modulu pre Python.

Definície a označenia

<i>Označenie</i>	<i>Vysvetlenie</i>
$\langle X \rangle$	stredná hodnota náhodnej premennej X
$X_n \rightarrow_p X$	$\{X_n\}$ konverguje podľa pravdepodobnosti k X
$f : \mathbb{R} \rightarrow \mathbb{R}$	neznáma funkcia, ktorej koreň hľadáme v SRFP
x^*	koreň f
$F(x)$	funkcia merania, náhodná premenná s vlastnosťou $\langle F(x) \rangle = f(x)$
$P(J)$	pravdepodobnosť javu J
$P(\bigcup_i J_i)$	pravdepodobnosť, že nastane aspoň jeden z javov J_i
$P(\bigcap_i J_i)$	pravdepodobnosť, že nastanú všetky javy J_i
$P(J \mid K)$	pravdepodobnosť, že nastane jav J , ak nastal jav K

1 Prehľad existujúcich algoritmov

1.1 Robbins-Monro algoritmus

Zvoľme $x_0 \in D$ a postupnosť $\{a_n\}_{n=0}^{\infty}$ spĺňajúcu podmienky:

$$\sum_{n=0}^{\infty} a_n = \infty, \quad (1.1)$$

$$\sum_{n=0}^{\infty} a_n^2 < \infty. \quad (1.2)$$

Pomocou rekurentného vzťahu definujeme postupnosť náhodných premenných x_n

$$x_{n+1} = x_n - a_n F(x_n), \quad (1.3)$$

V roku 1951 Robbins a Monro vo svojom článku [1] dokázali, že takto zvolená postupnosť za istých podmienok konverguje podľa pravdepodobnosti k x^* . My si tu uvedieme jemne upravenú verziu ich dôkazu.

1.2 Dôkaz konvergenzie Robbins-Monro algoritmu

Lema 1.1 *Nech f je neklesajúca náhodná premenná ohraničená konštantou $C > 0$, teda*

$$F^2(x) < C^2. \quad (1.4)$$

Potom strednú kvadratickú chybu n -tého kroku

$$b_n = \langle (x_n - x^*)^2 \rangle \quad (1.5)$$

možno vyjadriť v tvare

$$b_n = b_0 + \sum_{i=0}^{n-1} a_i^2 e_i - \sum_{i=0}^{n-1} 2a_i d_i, \quad (1.6)$$

kde

$$e_n = \langle f^2(x_n) + S^2(x_n) \rangle, \quad (1.7)$$

$$d_n = \langle (x_n - x^*)f(x_n) \rangle. \quad (1.8)$$

Dôkaz: Rekurenciu možno prepísať do tvaru [4]

$$x_{n+1} = x_n - a_n f(x_n) - a_n (S(x_n)), \quad (1.9)$$

kde náhodná premenná

$$S(x_n) = F(x_n) - f(x_n). \quad (1.10)$$

Označme

$$b_n = \langle (x_n - x^*)^2 \rangle \quad (1.11)$$

Potom možno b_{n+1} vyjadriť pomocou b_n .

$$b_{n+1} = \langle \langle (x_{n+1} - x^*)^2 \mid x_n \rangle \rangle \quad (1.12)$$

Využitím rekurencie (1.9) výraz zjednodušíme na

$$b_{n+1} = b_n - 2a_n \langle (x_n - x^*)f(x_n) \rangle + a_n^2 \langle f^2(x_n) + S^2(x_n) \rangle, \quad (1.13)$$

kde sme využili fakt, že $\langle S(x_n) \mid x_n \rangle = 0$. Zopakovaním postupu n -krát dostaneme požadované vyjadrenie (1.6). \square

Lema 1.2 *Limita $\lim_{n \rightarrow \infty} b_n$ existuje a suma $\sum_{i=0}^{\infty} a_n d_i$ konverguje.*

Dôkaz: Z (1.4) dostávame podmienku

$$e_n < 2C^2. \quad (1.14)$$

Z podmienky (1.2) potom vyplýva konvergencia sumy

$$\sum_{i=0}^{\infty} a_i^2 e_i. \quad (1.15)$$

Uvážením $b_n \geq 0$ a využitím vzťahu (1.6) dostávame nasledovný odhad pre sumu

$$\sum_{i=0}^n a_n d_i \leq \frac{1}{2} \left(b_0 + \sum_{i=0}^{\infty} a_i^2 e_i \right). \quad (1.16)$$

Funkcia f je neklesajúca a $f(x^*) = 0$, členy postupnosti d_n musia byť preto nezáporné a suma $\sum_{i=0}^{\infty} a_n d_i$ musí konvergovať. Z vyjadrenia (1.6) pre b_n potom plynie existencia limity postupnosti b_n . \square

Všimnime si, že b_n a d_n udávajú v istom zmysle presnosť aproximácie x^* pomocou x_n . To naznačuje, že veľkosť d_n by mohla súvisieť s veľkosťou b_n . Predpokladajme preto, že existuje nezáporná postupnosť k_n dostatočne veľkých, aby

$$\sum_{i=0}^{\infty} a_n k_n = \infty, \quad (1.17)$$

ktoré však odhadujú d_n pomocou b_n vzťahom

$$d_n \geq k_n b_n. \quad (1.18)$$

Nasledovná lema ukazuje, že takýto odhad nám umožní dokázať konvergenciu $x_n \rightarrow_p x^*$.

Lema 1.3 *Nech postupnosť k_n spĺňa podmienky (1.17) a (1.18). Potom $\lim_{n \rightarrow \infty} b_n = 0$ a $x_n \rightarrow_p x^*$.*

Dôkaz: Z podmienky (1.18) vyplýva konvergencia sumy

$$\sum_{n=0}^{\infty} a_n k_n b_n, \quad (1.19)$$

no podľa podmienky (1.17) samotná suma

$$\sum_{n=0}^{\infty} a_n k_n$$

nekonverguje. To znamená, že pre dostatočne veľké n , musí byť väčšina b_n dostatočne malých, aby dokázali zabezpečiť konvergenciu (1.19). Formálne

$$\forall \epsilon > 0, n > 0 \exists k > n : b_k < \epsilon, \quad (1.20)$$

Nula je hromadným bodom konvergentnej postupnosti b_n a preto musí byť tiež jej limitou. Ukázali sme, že

$$\lim_{n \rightarrow \infty} \langle (x_n - x^*)^2 \rangle = 0 \quad (1.21)$$

čo je však postačujúca podmienka pre konvergenciu x_n podľa pravdepodobnosti. \square

Ostáva nám ukázať, že je možné zvoliť vyhovujúcu postupnosť k_n . Predtým si však dokážeme jedno pomocné tvrdenie.

Lema 1.4 *Pre ľubovoľnú kladnú postupnosť a_n je podmienka (1.1) ekvivalentné podmienke*

$$\sum_{n=1}^{\infty} \frac{a_n}{a_1 + a_2 + \cdots + a_{n-1}} = \infty \quad (1.22)$$

Dôkaz: Prvá z implikácií je triviálna, ukážeme preto iba $(1.1) \Rightarrow (1.22)$. Predpokladajme, že

$$\sum_{n=1}^{\infty} \frac{a_n}{a_1 + a_2 + \cdots + a_{n-1}} = H < \infty, \quad (1.23)$$

potom existuje m , že

$$\sum_{n=1}^m \frac{a_n}{a_1 + a_2 + \cdots + a_{n-1}} > H - \frac{1}{2}, \quad (1.24)$$

avšak z (1.1) vieme, že existuje k

$$2 \sum_{n=0}^m < \sum_{n=1}^k. \quad (1.25)$$

Potom však

$$\sum_{n=m+1}^k \frac{a_n}{a_1 + a_2 + \cdots + a_{n-1}} > \frac{\sum_{n=m+1}^k a_n}{\sum_{n=0}^m a_n} > \frac{1}{2} \quad (1.26)$$

To však znamená

$$\sum_{n=1}^k \frac{a_n}{a_1 + a_2 + \cdots + a_{n-1}} > H, \quad (1.27)$$

čo je spor s predpokladom (1.23). \square

Veta 1.5 *Nech f je neklesajúca funkcia spĺňajúca v nejakom δ okolí bodu x^* nerovnosť*

$$M(x - x^*)f(x) > (x - x^*)^2 \quad (1.28)$$

pre nejaké $M > 0$. Nech zároveň F spĺňa podmienku (1.4). Potom postupnosť definovaná rekurenciou (1.3) konverguje podľa pravdepodobnosti k x^* .

Dôkaz: Pre $|x_n - x^*| < \delta$ z (1.28) platí odhad

$$|x_n - x^*| < Mf(x_n) \quad (1.29)$$

Ak je $|x_n - x^*| > \delta$, dostávame z triviálnej nerovnosti

$$|x_n - x^*| < |x_0 - x^*| + C \sum_{i=0}^{n-1} a_i \quad (1.30)$$

odhad

$$|x_n - x^*| < A f(x_n) \sum_{i=0}^{n-1} a_i, \quad (1.31)$$

kde

$$A = M/a_0 + (C + (|x_0 - x^*|)/a_0)M/\delta, \quad (1.32)$$

Prvý člen zároveň zabezpečuje platnosť odhadu aj pre prípad $|x_n - x^*| < \delta$.

Tým sme ukázali platnosť nerovnosti

$$(x_n - x^*)^2 < f(x_n)(x_n - x^*)A \sum_{i=0}^{n-1} a_i \quad (1.33)$$

a naši kandidáta pre k_n

$$k_n = \left(A \sum_{i=0}^{n-1} a_i \right)^{-1} \quad (1.34)$$

Stačí ukázať, že podmienka (1.17) je splnená, teda

$$\sum_{n=1}^{\infty} \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}} = \infty \quad (1.35)$$

to však na základe dokázanej lemy vyplýva z (1.1). \square

Dôsledok 1.6 *Nech f je neklesajúca funkcia, ktorá má v bode x^* nenulovú deriváciu a nech F spĺňa podmienku (1.4). Potom postupnosť x_n definovaná rekurenciou*

$$x_{n+1} = x_n + F(x_n)k/n, \quad (1.36)$$

kde x_0 a k sú pevne zvolené reálne čísla, konverguje podľa pravdepodobnosti k x^ .*

Dôkaz: Rekurencia z tvrdenia je zodpovedá rekurencii (1.3) s $a_n = k/n$. Je jednoduché overiť, že takto definované a_n vyhovuje podmienkam (1.1) aj (1.2). Stačí teda overiť vlastnosti f , tá má v bode x^* deriváciu $f'(x^*) > 0$. Existuje teda δ okolie x^* na ktorom f spĺňa (1.28) s konštantou $M = 2/f'(x^*)$ a môžeme uplatniť vetu 1.5. \square

1.3 Stochastická Newton-Raphsonová metóda

V roku 1958 Sacks skúmal [3] asymptotické správanie Robbins-Monro algoritmu s postupnosťou $a_n = k/n$. Sacksovi sa podarilo ukázať, že za splnenia istých bežných podmienok konverguje postupnosť $n^{1/2}(x_n - x^*)$ k náhodnej premennej s nulovou strednou hodnotou a disperziou $k^2\sigma^2/(2A\alpha - 1)$, kde $\alpha > 0$ je derivácia funkcie f v bode x^* a

$$\sigma^2 = \lim_{x \rightarrow x^*} \text{Var } F(x). \quad (1.37)$$

Je ľahké overiť, že optimálna hodnota k je z hľadiska rýchlosti konverencie α^{-1} . Problémom však je, že v praxi môžeme bez predošlej znalosti funkcie f hodnotu α len tipovať. To inšpirovalo v roku 1977 Dana Anbara k navrhnutiu metódy [2], ktorá by odhadovala α počas iterovania Robbins-Monro algoritmu. V ním navrhovanom algoritme sa α odhaduje lineárnym fitom metódou najmenších štvorcov nameraných hodnôt $F(x_n)$. Keďže však prvé hodnoty sú spravidla ďaleko od stredu bodu x^* , kde nás derivácia zaujíma, s pribúdajúcim množstvom hodnôt F sa oplatí prvé merania do výpočtu α nezahrnúť.

Takto upravený Robbins-Monro algoritmus teda iteruje rekurentným vzťahom

$$x_{n+1} = x_n + \frac{F(x_n)}{n\alpha_{m(n),n}}, \quad (1.38)$$

kde $\alpha_{m(n),n}$ označuje odhad derivácie berúci do úvahy body $x_{m(n)}$, $x_{m(n)+1}$, \dots , x_n . Autor v článku [2] dokázal, že takto zvolená postupnosť konverguje pre $m(n) = o(\log(n)^{1/(2+\epsilon)})$ pre všetky $\epsilon > 0$.

2 Motivácia binárneho SRFP

Algoritmy, ktoré sme si zatiaľ predstavili, boli navrhnuté pre ľubovoľný experiment s ohraničenými hodnotami výsledkov. To je nepochybne veľmi užitočná vlastnosť týchto algoritmov, v praxi sa však vo veľkej miere stretávame so situáciami, ktoré majú binárny charakter. Študenti testujú či sa im podarí obhájiť bakalársku prácu, alebo nepodarí, pyrotechnika zaujíma či bomba vybuchne, alebo nevybuchne a lekára trápí, či sa pacient uzdraví, alebo zomrie. Preto sme sa rozhodli v tejto práci zamerať na aplikáciu SRFP algoritmov na problémy s binárnymi výsledkami. Doteraz sme uviedli len veľmi všeobecné príklady reálnych aplikácií, na konkrétnom príklade si preto teraz predvedieme užitočnosť binárneho SRFP pri riešení otázky

2.1 Prečo ľudia konajú dobro?

Na riešenie tohto problému sa podujali Tomáš Kulich a Jaroslav Flegr [5]. Ako vo svojej práci uvádzajú, pôvod altruistického chovania, teda chovania užitočného pre spoločnosť ako celok, avšak nevýhodného pre jedinca, bol dlho orieškom pre evolučných biológov. Skupine s altruistickými jedincami sa ako celku darí lepšie než ostatným skupinám, no altruistický jedinci v skupine sú oproti sebcm znevýhodnení a postupne vymierajú. Pred nejakou dobou vznikla hypotéza, že ak altruizmus kontroluje viacero na sebe nezávislých génov, ktorých vhodná kombinácia vytvorí altruistického jedinca, môže v spoločnosti existovať stabilná rovnováha medzi altruistami a sebcami.

Autori sa pokúsili overiť toto tvrdenie numerickou simuláciou jednoduchého modelu veľkej metapopulácie, ktorá sa členila na množstvo menších populácií. Monitorovali chovanie sa metapopulácie v priebehu množstva generácií, kde každá generácia pozostávala z troch fáz: prirodzený výber, mig-

rácia a mutácia. Samotný prirodzený výber zachovával počet členov celej metapopulácie, no zvýhodňoval altruistické populácie pred sebeckými. Počet potomkov narodených do populácie bol úmerný počtu členov populácie a podielu altruistov v nej. V momente, kedy niektorá populácia klesla v počte členov pod dvoch, automaticky ju pohltila najväčšia existujúca populácia.

V rámci samotných populácií sa však potomok narodil s väčšou pravdepodobnosťou sebeckému jedincovi než altruistovi. Veľkosť zvýhodnenia sebcov v rámci skupiny a veľkosť zvýhodnenia altruistických skupín pred ostatnými skupinami bola popísaná parametrami (α, β) a v práci bolo skúmané, pre akú najmenšiu hodnotu parametra α je aspoň polovičná pravdepodobnosť nájsť po N generáciách v metapopulácii aspoň nejaké malé množstvo ε altruistických jedincov.

V tomto prípade sa jednalo o binárny problém závislý od parametra α , funkcia pravdepodobnosti $f(\alpha)$ nebola známa a jej jednotlivé merania boli časovo veľmi náročné. Nameranie pár stoviek hodnôt zabralo týždeň počítania tridsaťjadrového clusteru.

Podobné problémy, pri ktorých treba riešiť binárne SRFP, sa objavujú v rôznych oblastiach vedeckej činnosti a je preto oprávnené zaoberať sa výkonnosťou SRFP algoritmov pri ich riešení.

2.2 Definícia binárneho SRFP

V ďalšom budeme za binárne SRFP považovať SRFP pri ktorom F môže nadobúdať len hodnoty ± 1 . Od skúmaných funkcií f preto budeme prirodzene považovať splnenie podmienky $|f(x)| < 1$ pre všetky x . Takto definované binárne SRFP nestráca na všeobecnosti, pretože ako čitateľ ľahko nahliadne, transformáciou $aF + b$ možno tento problém previesť na ľubovoľné binárne SRFP.

3 Stochastická metóda regula falsi

3.1 Motivácia

Spomedzi numerických metód hľadania koreňa patrí regula falsi k najstarším. Prvé zmienky o nej sa vyskytujú už v treťom storočí pred našim letopočtom v indickom diele Vaishali Ganit.

Regula falsi funguje podobne ako metóda polenia intervalov. Štartuje s dvoma bodmi a_0, b_0 takými, že $f(a_0)$ a $f(b_0)$ sú opačných znamienok. Za predpokladu spojitosti analyzovanej funkcie f je potom v intervale $[a_0, b_0]$ zaručená existencia koreňa. Metóda produkuje postupnosť intervalov $[a_k, b_k]$ aproximujúcich hodnotu koreňa. V k -tom kroku sa spočíta hodnota

$$c_k = \frac{f(b_k)a_k - f(a_k)b_k}{f(b_k) - f(a_k)} \quad (3.1)$$

Možno ľahko nahliadnuť, že c_k je priesečníkom x -ovej osi a spojnice bodov $(a_k, f(a_k))$ a $(b_k, f(b_k))$. V nasledovnom kroku sa jeden z bodov a_k, b_k nahradí bodom c_k a druhý sa ponechá tak, aby bolo naďalej splnené, že a_{k+1} a b_{k+1} majú opačné znamienka. Tento postup sa iteruje dokým nedostaneme dostatočne presnú aproximáciu koreňa.

3.2 Stochastická verzia

Celú metódu budeme formulovať pre rastúcu f a binárny experiment s hodnotami ± 1 . Ak by sme pustili originálnu verziu regula falsi na SRFP, algoritmus by síce k nejakej hodnote konvergoval (v každom kroku by sa interval zmenšil na polovicu), ale nebolo zaručené (a ani pravdepodobné), že táto hodnota by bola koreňom skúmanej funkcie. Aby sme si toto poistili, metódu jemne modifikujeme tak, aby v každej iterácii merala hodnoty v daných bodoch viackrát a aby tento počet stúpал s každou iteráciou. Takto algoritmus získa

vždy dostatočne presné aproximácie hodnôt funkcie a bude zaručené konvergencia ku koreňu.

Nasleduje podrobný popis fungovania algoritmu:

- Rastúca funkcia $s(k)$ udáva závislosť počtu meraní od čísla iterácie k
- V každej iterácii algoritmus disponuje poľom trojíc (x, y, c) , kde x sú všetky doteraz skúmané body, y sú aproximácie hodnoty $f(x)$ vytvorené z c meraní. Toto pole trojíc je usporiadané podľa hodnôt x . Ďalej si udržiava informáciu o práve analyzovanom bode x_{cur} .
- Algoritmu štartuje s dvoma hodnotami x_0, x_1 , pre ktoré je zaručené, že $f(x_0) < 0$ a $f(x_1) > 0$ a v poli sú zaradené ako $(x_0, -1, \infty)$ a $(x_1, 1, \infty)$. Práve analyzovaný bod x_{cur} je inicializovaný na x_0 .
- V k -tej iterácii algoritmus odmeria $s(k)$ -krát hodnotu v bode x_{cur} . Na základe znamienka priemeru usúdi, či sa jedná o bod vpravo alebo vľavo od hľadaného koreňa. Ďalej vezme najbližší bod x vľavo resp. vpravo od x_{cur} . K nemu príslušná hodnota y má určite opačné znamienko ako hodnota práve spočítaná hodnota y_{cur} .
- V takto vybratom bode x algoritmus odmeria hodnotu toľkokrát, aby počet meraní v tomto bode bol aspoň $s(k)$. V prípade, že sa znamienko y týmito meraniami zmení, algoritmus nastaví novú hodnotu x_{cur} na x a celý postup zopakuje.
- Ak hodnota y ani po dodatočných meraniach nezmení znamienko, je zaručené, že y a y_{cur} majú opačné znamienka. Podľa vzťahu (3.1) vypočítame novú hodnotu x_{cur} a prejdeme do $(k + 1)$. iterácie.

3.3 Dôkaz konverencie

Ukážeme, že popísaný algoritmus konverguje podľa pravdepodobnosti ku koreňu hľadanej funkcie f . Najskôr však uvedieme niekoľko pomocných tvrdení.

Lema 3.1 *Nech Y_1, Y_2, \dots sú nezávislé Bernoulliho náhodné premenné so strednou hodnotou $p > \frac{1}{2}$ a $S_n = \frac{1}{n} \sum_{i=1}^n Y_i$. Potom pre ľubovoľné ε existuje také k , že:*

$$P\left[\bigcup_{i=k}^{\infty} (S_i \leq \tfrac{1}{2})\right] < \varepsilon \quad (3.2)$$

Dôkaz: Podľa Chernoffovej nerovnosti platí

$$P(S_i \leq \tfrac{1}{2}) < \exp[-2n(p - \tfrac{1}{2})^2]. \quad (3.3)$$

Podľa union bound nerovnosti potom možno (3.2) odhadnúť ako

$$\sum_{i=k}^{\infty} \exp[-2n(p - \tfrac{1}{2})^2] \quad (3.4)$$

Pre pravdepodobnosť z (3.2) dostávame odhad

$$P < \frac{\exp[-2k(p - \tfrac{1}{2})^2]}{1 - \exp[-2(p - \tfrac{1}{2})^2]}. \quad (3.5)$$

Kedže $p - \frac{1}{2} > 0$

$$\lim_{k \rightarrow \infty} \frac{\exp[-2k(p - \tfrac{1}{2})^2]}{1 - \exp[-2(p - \tfrac{1}{2})^2]} = 0. \quad (3.6)$$

Čím je dôkaz dokonaný. \square

Dôsledok 3.2 *Pre každé $\varepsilon > 0$ a $d > 0$ existuje $k \in \mathbb{N}$, že ak v k -tej alebo neskoršej iterácii algoritmu platilo*

$$|x_{\text{cur}} - x^*| < d, \quad (3.7)$$

tak pravdepodobnosť, že pre nejakú neskoršiu iteráciu by táto podmienka prestala platiť, je menšia ako ε .

Veta 3.3 *Nech $s(k) \geq k$. Potom v stochastickej verzii regula falsi konverguje x_{cur} k x^* takmer určite.*

Dôkaz: Nech $d > 0$. Označme β menšiu z hodnôt $|f(x^* - d)|$ a $|f(x^* + d)|$. Keďže f je rastúca, tak pre x mimo intervalu $[x^* - d, x^* + d]$ platí

$$P[F(x) = \text{sgn } f(x)] \geq \frac{1}{2}(\beta + 1), \quad (3.8)$$

Horeuvedený vzťah popisuje pravdepodobnosť dostať správne znamienko pri jednom meraní v bode x .

V k -tej iterácii je každá použitá hodnota aproximovaná priemerom $s(k) > k$ hodnôt. Podľa Chernoffovej nerovnosti je preto pravdepodobnosť name-rať v k -tej iterácii nesprávne znamienko pre x mimo intervalu $[x^* - d, x^* + d]$ menšia ako

$$\exp[-\frac{1}{2}k\beta^2]. \quad (3.9)$$

Hodnôt x mimo intervalu $[x^* - d, x^* + d]$ je v k -tej iterácii najviac k . Pravdepodobnosť, že by sme v tejto iterácii pre niektorú z nich určili nesprávne znamienko je preto menšia ako

$$k \exp[-\frac{1}{2}k\beta^2]. \quad (3.10)$$

Nepochybne vieme ku každému $\varepsilon > 0$ nájsť také k , že horeuvedený výraz bude menší ako ε . To však znamená, že v $(k + 1)$. iterácii je s pravdepodobnosťou $1 - \varepsilon$ bod x_{cur} od x^* vzdialený menej ako d .

Potiaľto sme dokázali konvergenciu podľa pravdepodobnosti k x^* . S využitím dôsledku (3.2) však dostávame silnejšiu konvergenciu takmer isto.

□

3.4 Voľba funkcie s

Predošlý výsledok ukazuje, že algoritmus takmer isto konverguje k x^* pre ľubovoľnú rozumnú funkciu s . Napriek tomu má zmysel zamýšľať sa nad jej voľbou, pretože tá bude mať veľký vplyv na rýchlosť algoritmu. Pre príliš malé s sa bude algoritmus často myliť a bude preto často merať v nesprávnych bodoch, pre príliš veľké s však naopak bude merať zbytočne veľa hodnôt v každom bode.

Dobrá voľbu s môžeme skúsiť odhadnúť nasledovnou úvahou. V blízkom okolí koreňa funkcie bude vzdialenosť od koreňa približne priamo úmerná veľkosti funkcie v danom bode. Za optimálnych podmienok môžeme očakávať, že v priemere sa každou iteráciou vzdialenosť od koreňa zmenší na polovicu, tým pádom pre veľkosť funkcie v k -tej iterácii platí približne vzťah $f(x_k) \propto 2^{-k}$. Podľa Chernoffovej nerovnosti potrebujeme na určenie znamienka v k -tej iterácii počet meraní úmerný 2^{2k} . To nasvedčuje, že z asymptotických dôvodov, je rozumné zaoberať sa funkciami, ktoré sú exponenciálne od k . V praxi treba opäť zvoliť istý kompromis medzi asymptotikou a chovaním v prvých krokoch, kedy funkciu nemožno aproximovať priamkou. Vtedy, hlavne v prípadoch kedy sú na krajoch dlhé oblasti hodnôt blízky k 1 a -1, je priveľké s plytvaním. Sú tu ďalšie možnosti zlepšovania algoritmu tak, aby si s volil dynamicky na základe svojich meraní, ale tie idú mimo rozsah tejto práce.

4 Metodika testovania

4.1 Dodatočné predpoklady pri implementácii algoritmov

V praxi je nemožné efektívne hľadať koreň bez akýchkoľvek ďalších informácií v celom \mathbb{R} . Algoritmy síce konvergujú k riešeniu, avšak na nekonečne dlhom intervale im to môže trvať veľmi veľmi dlho. Preto je potrebné odhadnúť nejaký konečný interval $[a, b]$, na ktorom sa koreň bude nachádzať. Nejaké obmedzenia sú v skutočnosti dané už samotnou povahou experimentu, napríklad je jasné, že nemožno nadopovať pacienta dávkou lieku hmotnosti prevyšujúcej hmotnosť pacienta – nezmesť sa. Akékoľvek ďalšie informácie môžu algoritmy pri binárnom experimente potešiť, ale nie sú nutnosťou. Preto, aby bola súťaž férová, stanovili sme nasledovné jednotné rozhranie pre všetky testované algoritmy

- funkciu merania F (vracia hodnoty ± 1);
- konečný interval $[a, b]$ v ktorom je zaručené nájsť koreň.

Bolo by nerozumné, keby niektorý z algoritmov skúšal testovať aj body mimo zadaného intervalu $[a, b]$. Okrem toho, že to určite nepomôže k výsledku, hrozí, že funkcia merania F nebude mimo tohto intervalu definovaná. Preto sú všetky testované algoritmy upravené tak, aby v prípade snahy opustiť interval skončili v jeho príslušnom krajnom bode.

Niekoľko krátkych poznámok k implementácii jednotlivých algoritmov:

- Robbins-Monro algoritmus začína v $x_0 = \frac{1}{2}(a + b)$ a konštantu k zvolí ako $k = b - a$. Je to najrozumnejší odhad prevrátenej hodnoty derivácie f , aký sa zo zadaných informácií dá spraviť.

- Stochastická Newtonova metóda bude tiež začínať v strede intervalu a odhad sklonu si už dopočítava sama z výsledkov meraní v jednotlivých bodoch. Na začiatku vychádza z predpokladu, že v bodoch a , b sa postupne namerali hodnoty -1 a 1 . Nie je úplne jasné, ako zvolíť zahadzovaciu funkciu $m(n)$, budeme preto testovať túto metódu pre rôzne varianty.
- Ako neskôr ukážeme, originálna Stochastická Newtonova metóda sa v praxi chová veľmi neefektívne. Pokúsili sme sa niektoré zdroje tejto neefektivity odstrániť implementovaním niekoľkých heuristik. Bližšiu diskusiu o týchto heuristikách a zdroji neefektív v Newtonovej metóde nájdete v tejto kapitole.
- Stochastická metóda regula falsi bude vychádzať z predpokladu, že v krajných bodoch boli s nekonečnou presnosťou namerané hodnoty ± 1 . Vhodnú voľbu funkcie $s(k)$, určujúcej počet meraní v k -tej iterácii, preskúmame neskôr.

4.2 Testované parametre

Hlavným parametrom, od ktorého sa naše merania efektivity odvíjajú, je šírka 95 % intervalu spoľahlivosti za zadaných podmienok. Všetky merania, ak nie je povedané inak, sa zopakovali 1 000 krát. Aby algoritmy nemohli profitovať z uhádnutia správneho výsledku tipom na stred intervalu a aby merania mali väčšiu validitu pri reálnych podmienkach, pri meraní so zadanou šírkou intervalu sme pri jednotlivých meraniach volili hranice intervalov $[a, b]$ tak, koreň prešiel postupne rovnomerne cez vnútorných 80% intervalu.

Algoritmy sme porovnávali v nasledujúcich oblastiach:

- Tvar distribučnej funkcie výsledkov predpovedaných algoritmi

- Schopnosť vyrovnať sa s dlhými monotónnymi úsekmi (napr. sgn)
- Asymptotická rýchlosť konverencie
- Testy na niekoľkých náhodne zvolených funkciách

4.3 Stochastický Newtonov algoritmus

Stochastický Newtonov algoritmus vo svojej originálnej podobe z práce Dan Anbara [2] je pre účely binárneho experimentu prakticky nepoužiteľný. Ako možno vidieť na B.29, prítomnosť dlhých monotónnych intervalov ho úplne vyvedie z miery. Na lineárnych úsekoch B.21 sa už chová podstatne lepšie, avšak stále prudko zaostáva za pôvodným Robbins-Monro algoritmom, ktorý sa snažil vylepšiť.

Analýzou správania sa tohto algoritmu sme odhalili jeden z častých zdrojov jeho neefektivity. Algoritmus občas spravil extrémne malý odhad sklonu skúmanej funkcie f a to malo za následok prudký skok preč z práve analyzovaného miesta zväčša na niektorú z hraníc intervalu $[a, b]$. Tým pádom algoritmus stratil všetku práčne spočítanú informáciu a začínal odznova.

Aby sme tento efekt eliminovali, pridali sme do algoritmu podmienku, ktorá testovala, či sa algoritmus nesnaží pohnúť o väčšiu vzdialenosť, než je priemerná dĺžka jeho kroku od začiatku vykonávania. V prípade, že áno, tak sa tá iterácia nahradila pohybom podľa Robbins-Monro metódy.

To viedlo k podstatnému zlepšeniu algoritmu v prípade dlhých monotónnych úsekov (B.29), nezanedbateľné zlepšenie nastalo i pri lineárnych úsekoch (B.21), avšak nie také prudké ako pri prvej z možností. Pri lineárnych úsekoch sa algoritmu síce stále nepodarilo prekonať Robbins-Monro algoritmus, avšak pri dlhých monotónnych úsekoch sa ukázala jeho opodstatnenosť.

V ďalších porovnaniach z dôvodov prudkej nepoužiteľnosti originálneho algoritmu budeme pracovať vždy s verziou doplnenou o heuristiky.

4.4 Voľba zahadzovacej funkcie pre Newtonov algoritmus

Vo svojom článku [2] Dan Anbar používal $m(n) = o(\log(n)^{1/(2+\epsilon)})$. To je pre praktické účely takmer konštanta, napríklad $\log(1000)^{1/2} < 3$. Je preto namieste otázka, či v praxi nemožno dosiahnuť lepšie výsledky s pomocou inak zvolenej funkcie $m(n)$. Zamerali sme sa na hľadanie zahadzovacej funkcie v tvare $m(n) = n^\gamma$ a preskúmali sme správanie sa pre rôzne γ z intervalu $[0, 1]$. Výsledky pre lineárne a monotónne úseky sú vynesené do grafov B.1 a B.2. Z pochopiteľných dôvodov je pri lineárnych úsekoch najvýhodnejšie nezahadzovať vôbec, zaujímavé však je, že rozdiel medzi takmer úplným zahadzovaním a nezahadzovaním nie je priepastný. Na druhej strane, pri funkcii signum si až o niekoľko rádov lepšie viedli algoritmy, ktoré zahadzovali viac. Hodnota γ musí byť opäť nejakým kompromisom medzi týmito dvoma extrémnymi prípadmi. Keďže však vplyv v druhom prípade bol síce veľmi výrazný, ale očakáva sa, že funkcia bude v prvom móde bežať väčšinu svojho času. Je preto rozumné sa viac prikloniť niekam do stredu. Preto vo svojich ďalších meraniach budeme používať hodnotu $\gamma = 0,4$

4.5 Voľba funkcie s pre metódu regula falsi

Ako sme už v predošlej kapitole odôvodnili, je rozumné tipovať funkciu s v tvare $s(k) = c^k$. Odkúšali sme rôzne hodnoty c z intervalu $[1, 5]$. Výsledky pre lineárne a monotónne úseky sú vynesené do grafov B.3 a B.4. Pri monotónnych úsekoch je očakávateľné, že najvýhodnejšia funkcia s bude čo

najpomalšie rastúca. Pri lineárnych úsekoch, sa v súlade s teóriou z predošlej kapitoly, ukazujú ako najvýhodnejšie hodnoty s $c > 4$, avšak rozdiely nie sú podstatne veľké. Veľká hodnota c však znamená obrovský nárast počtu meraní v monotónnych úsekoch. Keďže medzi týmito dvoma extrémami treba zvoliť nejaký kompromis, v ďalších meraniach budeme používať hodnotu $c = 1,3$.

4.6 Distribúcia výsledkov predpovedaných algoritmami

Tvar distribučnej funkcie výsledkov jednotlivých algoritmov je ich zaujímavou vlastnosťou. Zo znalosti jediného čísla, štandardnej odchýlky, je možné získať všetky podstatné informácie, napríklad šírku intervalov spoľahlivosti pre ľubovoľný percentil.

Ak je algoritmus dobrý, jeho výsledok by nemal silno závisieť od malého počtu meraní. Jedno pokazené meranie by vtedy znamenalo silný posun vo výsledku a bolo by určite rozumnejšie takúto kritickú časť premerať viackrát. Možno teda očakávať, že hodnota výsledku bude faktor podstatne závislý od mnohých náhodných premenných. V takých prípadoch čakáme, že distribučná funkcia bude podobná Gaussovej funkcii normálneho rozdelenia. Aby sme toto preverili, zostrojili sme pre jednotlivé algoritmy graf rozdelenia výsledkov po 100 krokoch pri meraní sigmoidy na rôznych intervaloch. Napriek tomu, že niektoré grafy (B.11) vyzerajú pomerne gaussovsky, gaussovka, vytvorená len na základe informácie o štandardnej odchýlke ich vôbec nefituje. Po odstránení približne 5 % krajných hodnôt sa štandardná odchýlka prudko zmení a z nej vytvorené gaussovka už hodnoty pekne fituje. Z toho môžeme vyvodiť nasledovný záver: 95 % hodnôt možno pomerne dobre považovať za gaussovsky distribuované, avšak štandardnú odchýlku treba rátať bez 5 % najodchýlenejších hodnôt.

Je zaujímavé všimnúť si, že pre malé dĺžky intervalov sú výsledky všetkých algoritmov gaussovské, no pre najväčšiu dĺžku má Robbins-Monro algoritmus podstatne iný tvar.

4.7 Asymptotické správanie sa algoritmov

Pri skúmaní asymptotického správania vychádzame z predpokladu, že po veľkom počte krokov sa všetky algoritmy dostanú do oblasti, ktorú možno linearizovať. Preto sme asymptotické chovanie skúmali na lineárnych funkciách. Správanie sa algoritmov sa nezmení, v prípade, že všetky zadané dĺžky vynásobíme rovnakým faktorom. Algoritmy totiž nijako nevychádzajú z absolútnej dĺžky, ale z dĺžok relatívnych k šírke zadaného intervalu. Preto má zmysel obmedziť sa pri lineárnych funkciách len na jednu pevnú šírku intervalu a meniť iba sklon intervalov a s rôznymi sklonmi. Z grafov možno vidieť, že až po pomerne veľké sklony (B.18) je Robbins-Monro algoritmus najefektívnejší. Pri väčších sklonoch sa už samozrejme začínajú prejavovať výhody ďalších dvoch algoritmov, ale vtedy je už skúmaná funkcia na väčšine intervalu monotónna, keďže nesmie nikde presiahnuť hodnoty ± 1 .

4.8 Schopnosť vysporiadať sa s monotónnymi oblasťami

Často sa stane, že experimentátor na začiatku určí o niekoľko rádov väčší interval, ako interval, na ktorom skutočne dochádza k nejakej zmene funkcie. Je preto dobré, keď sa vedia algoritmy rýchlo a efektívne vysporiadať s dlhými oblasťami samých 1 a -1. Túto schopnosť sme testovali na dokonalom prototype takej funkcie, funkcii sgn . Vo výsledkoch (B.28) sa prejavila

silná stránka Newtonovej a regula falsi metód, keď za nimi Robbins-Monro o niekoľko rádov zaostával.

Rovnako zaujímavé, i keď trochu menej časté prípady, sú keď funkcia je síce dlho monotónna, avšak jedná sa o oblasti iných ako extrémnych hodnôt. Vtedy je situácia komplikovanejšia, lebo algoritmy môžu pri malom počte meraní dostať mylnú predstavu o hodnotách funkcie. Vyskúšali sme preto efektivitu algoritmov aj na niekoľkých funkciách tvaru $\lambda \operatorname{sgn}$, pre parameter $\lambda \in (0, 1]$. Ako sa postupne hodnota parametra zmenšovala, postupne rástol počet meraní potrebných na to, aby sa prejavila výhodnosť zvyšných dvoch metód oproti Robbins-Monro algoritmu (B.25). Vo všetkých meraniach sa viedlo metóde regula falsi lepšie ako Newtonovej, rozdiely však neboli drastické, väčšinou dosahovali faktor 2, nikdy nepresiahli faktor 10.

4.9 Testy na ďalších funkciách

Na záver sme podrobili algoritmy testu na viacerých rôznych funkciách (B.30, B.31, B.32, B.33). Ako je z meraných dát vidno, presnosti určenia koreňa boli veľmi podobné pre všetky tri algoritmy.

5 Záver

V úvode práce sme si predstavili SRF problém a zoznámili sa s algoritmami na jeho riešenie i praktickú motiváciu. Uviedli sme dôkaz konverencie pre vôbec prvý SRFP algoritmus od Robbinsa a Monroa [1] a zoznámili sa so zaujímavým pokusom o jeho vylepšenie od Dan Anbara [2]. V tretej kapitole sme predstavili vlastnú stochastickú adaptáciu na metódu regula falsi a ukázali, že za podmienok binárneho experimentu rieši SRFP takmer isto.

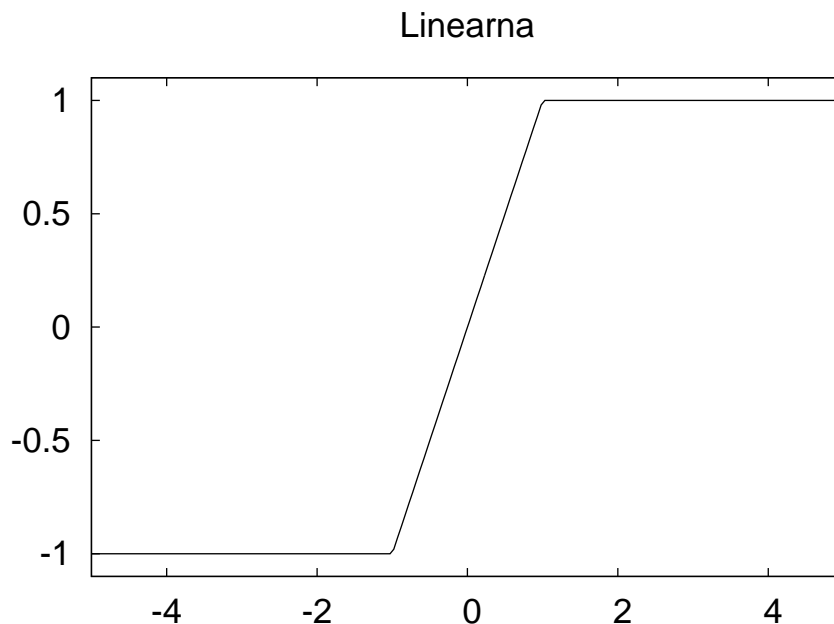
V záverečnej kapitole sme všetky tri algoritmy podrobili dôkladnému testovaniu, kedy sme preskúmali ich asymptotické vlastnosti, rýchlosť počiatočnej konverencie a ďalšie parametre. Ukázalo sa, že Stochastická adaptácia Newtonovej metódy [2] od Dana Anbara je pre účely binárneho experimentu v pôvodnom tvare nepoužiteľná, avšak po drobnej modifikácii sa stala veľmi výkonnou pri niektorých inštanciách problému. Takto modifikovaná Newtonova metóda aj metóda regula falsi mali veľmi podobné charakteristiky, v meraniach sa však ukazovalo, že regula falsi je vo väčšine prípadov jemne výkonnejšia. Tieto dve metódy vhodne dopĺňala metóda Robbins-Monro [1], ktorá bola jednoznačne najvýkonnejšou metódou pokiaľ sa jednalo o konvergenciu na úseku okolo koreňa, kde bolo možné funkciu dobre lineárne aproximovať. Zvyšným dvom metódam sa skôr darilo v prípade prudko meniacich sa funkcií, na ktoré sa vedeli lepšie adaptovať.

A Prehľad testovaných funkcií

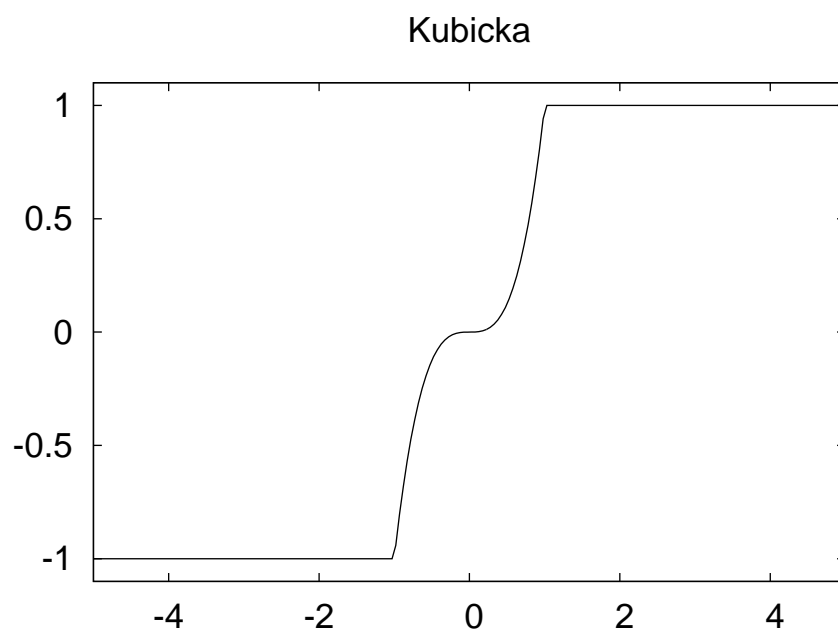
Počas testovania bolo použitých niekoľko funkcií, ich grafy a vlastnosti uvádzame v tejto prílohe

Názov	Definícia	Derivácia v bode 0
Lineárna	x	1
Kubická	x^3	0
Sigmoida	$2[1 + \exp(-x)]^{-1} - 1$	0,5
Cauchyho	$2\pi^{-1} \arctan(x)$	$2\pi^{-1} \approx 0,63$
Erf	$\pi^{-1/2} \int_{-x}^x \exp(-x^2) dx$	$2\pi^{-1/2} \approx 1,13$
Sgn	$\text{sgn}(x)$	∞

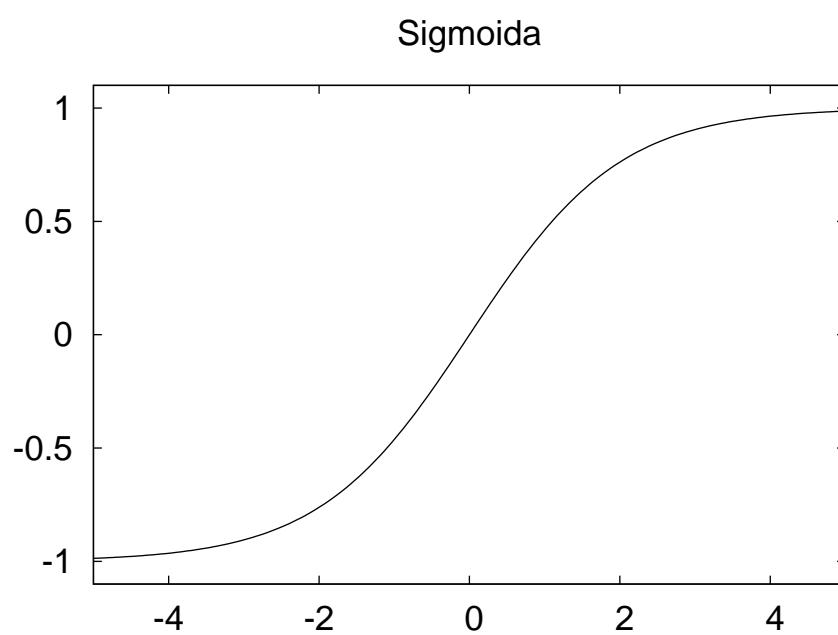
Tabuľka A.1: Vlastnosti testovaných funkcií



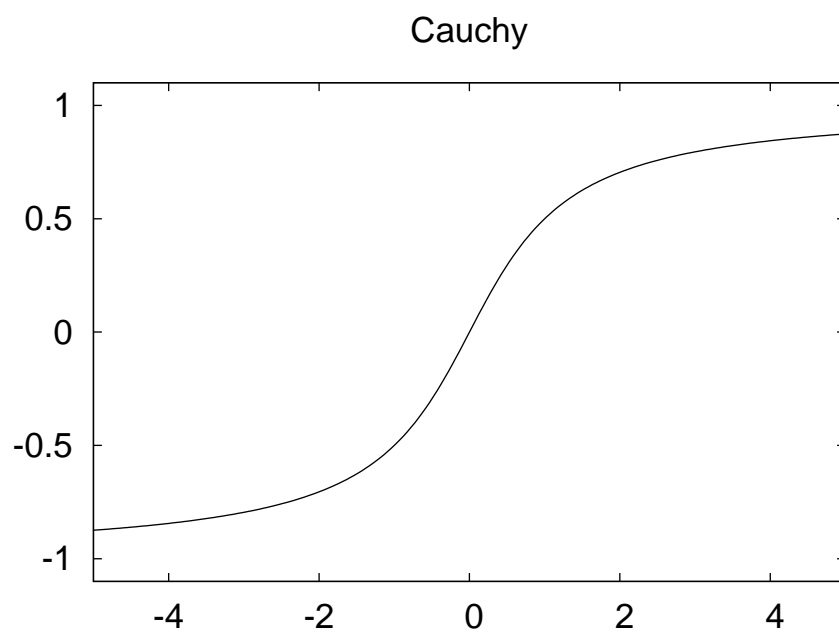
Obr. A.1: Priebeh lineárnej funkcie



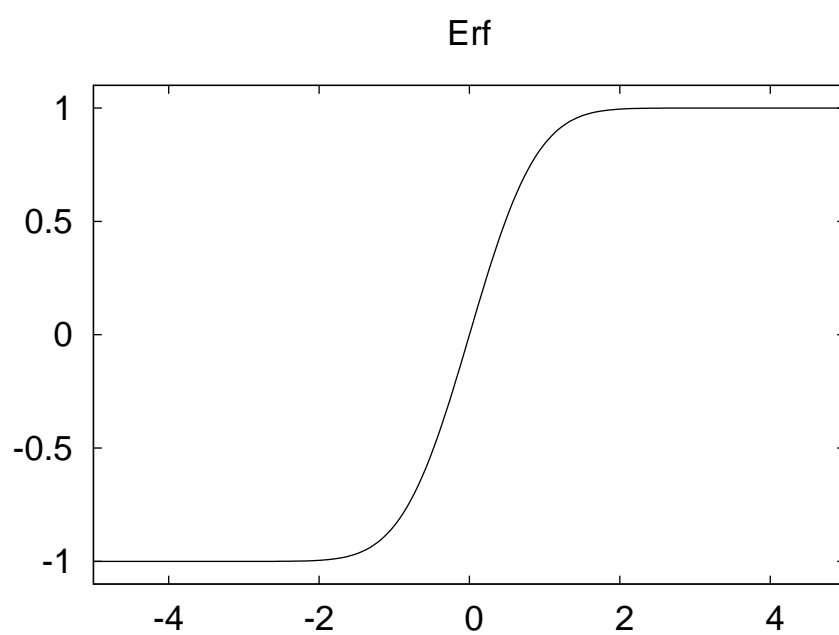
Obr. A.2: Priebeh kubickej funkcie



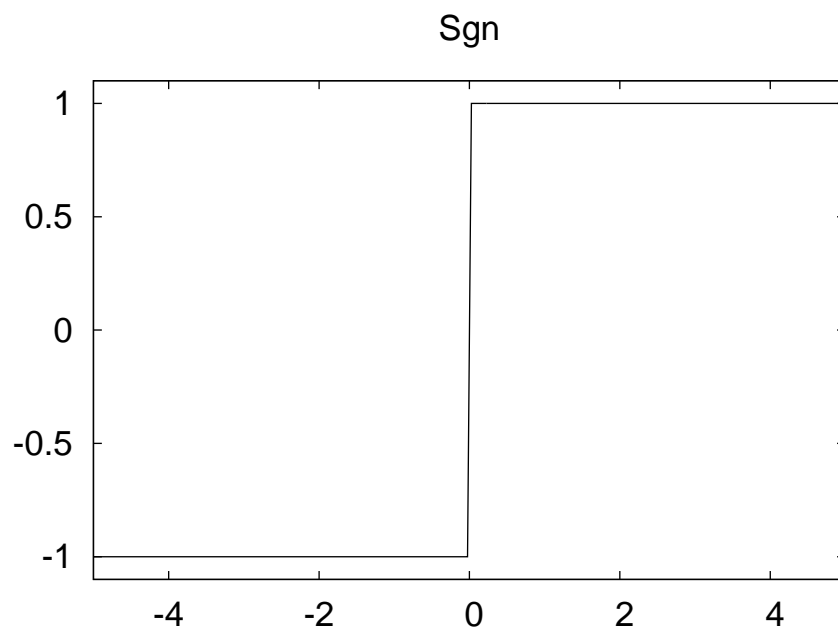
Obr. A.3: Priebeh sigmoidy



Obr. A.4: Pribeh Cauchyho funkcie



Obr. A.5: Pribeh Erf

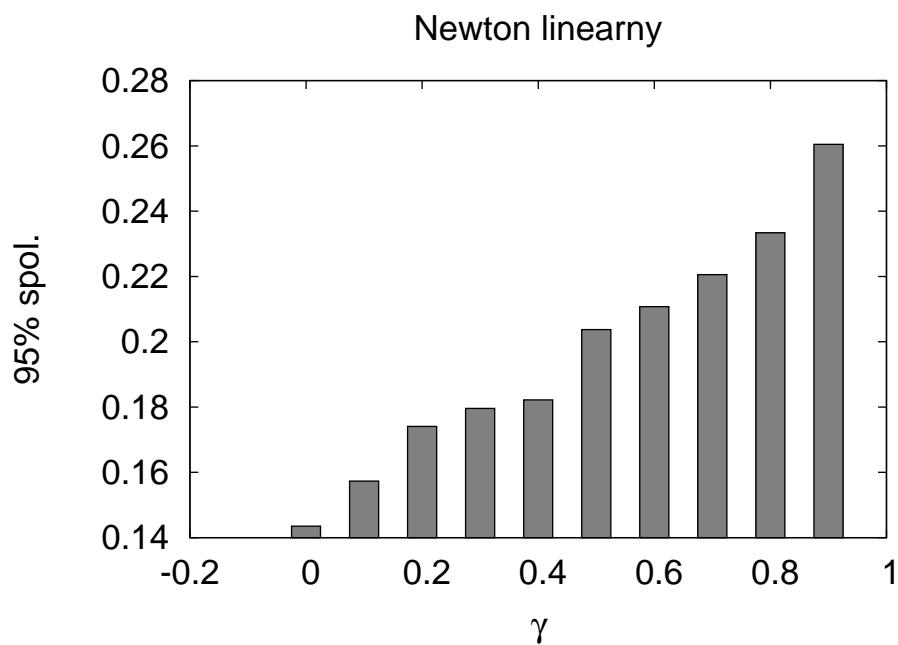


Obr. A.6: Priebeh Sgn

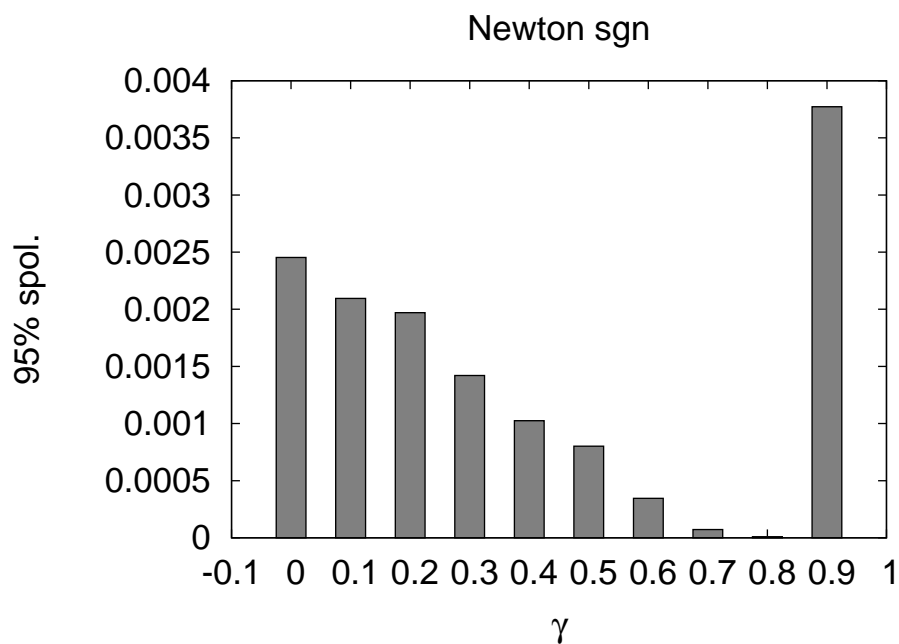
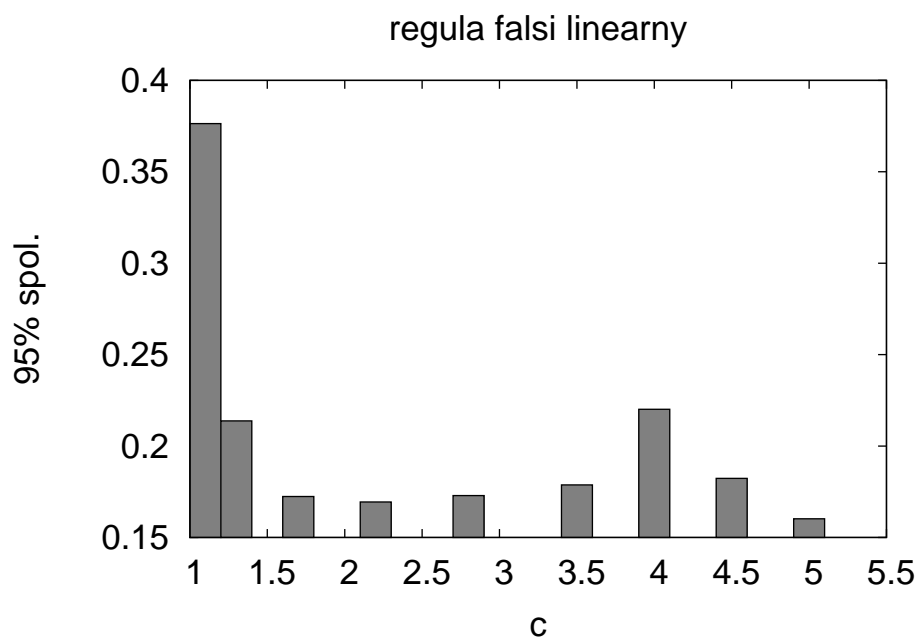
B Výsledky meraní

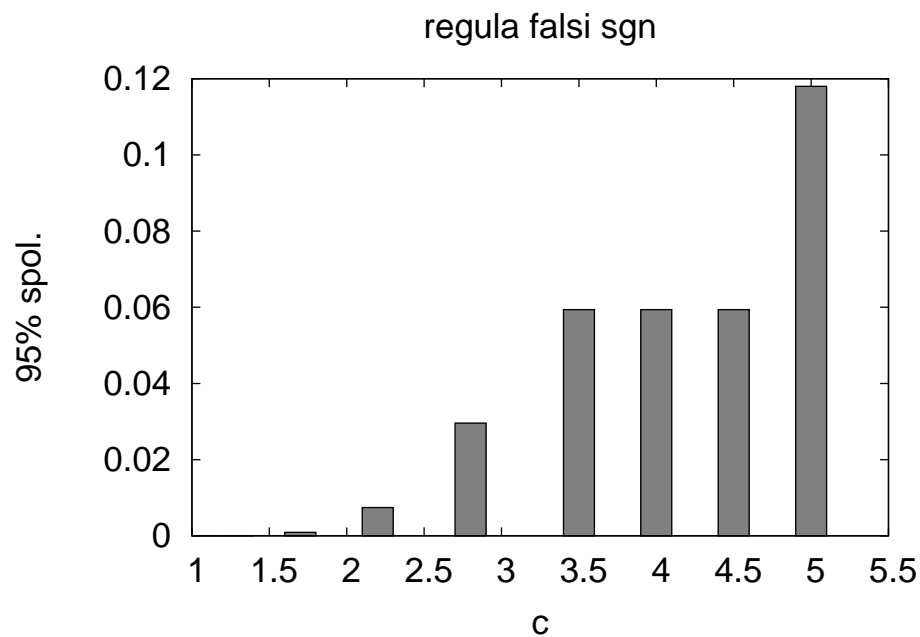
B.1 Voľba parametrov pre metódy Newtonova a regula falsi

Nasledovné výsledky ukazujú šírku 95 % intervalu spoľahlivosti po 1 000 krokoch algoritmu. Šírka počiatočného intervalu bola vo všetkých meraniach nastavená na 1.



Obr. B.1: Šírka intervalu spoľahlivosti pre rôzne hodnoty γ pre lineárnu funkciu

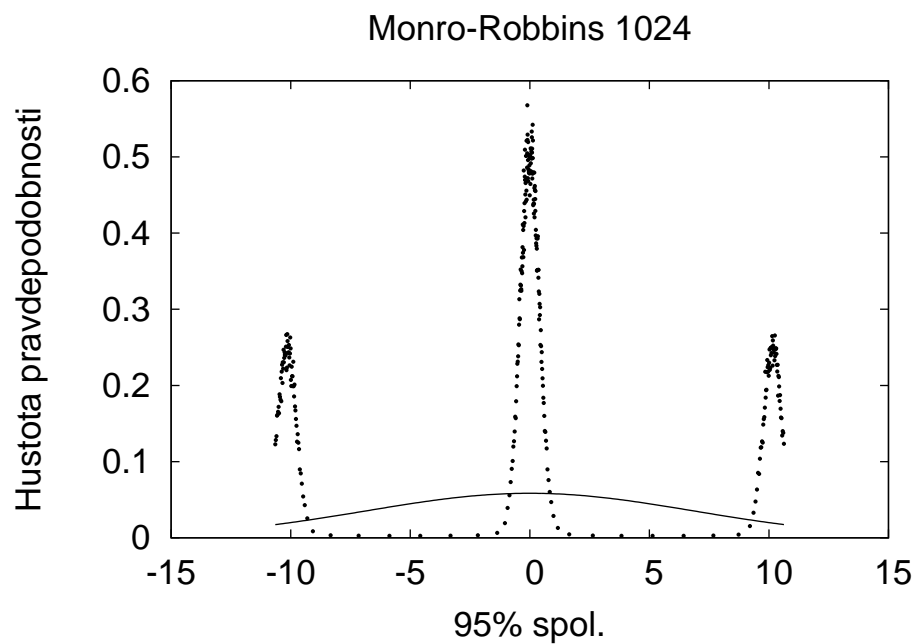
Obr. B.2: Šírka intervalu spoľahlivosti pre rôzne hodnoty γ pre funkciu sgnObr. B.3: Šírka intervalu spoľahlivosti pre rôzne hodnoty c pre lineárnu funkciu.



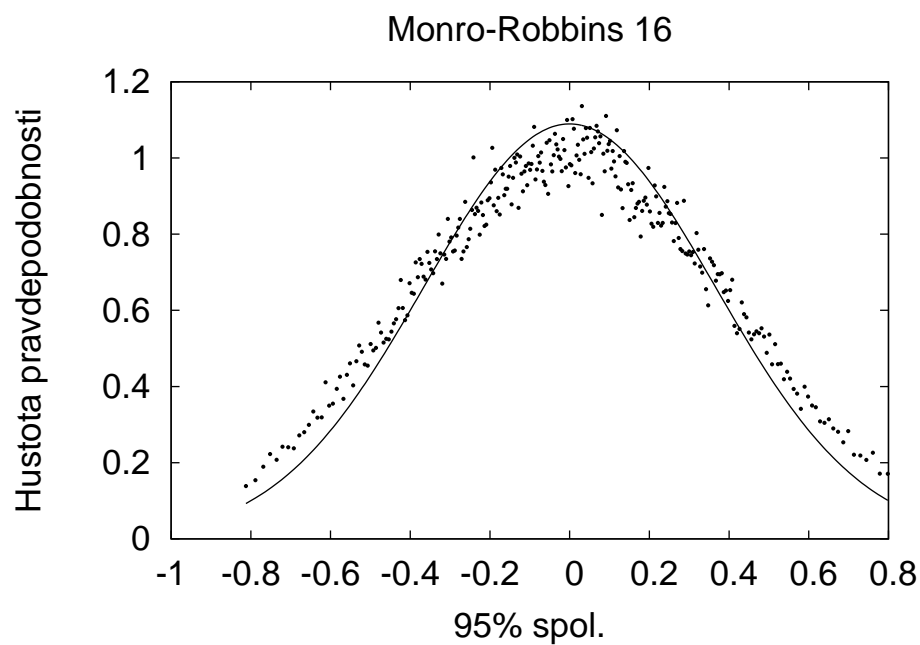
Obr. B.4: Šírka intervalu spoľahlivosti pre rôzne hodnoty c pre funkciu sgn.

B.2 Tvar distribučnej funkcie

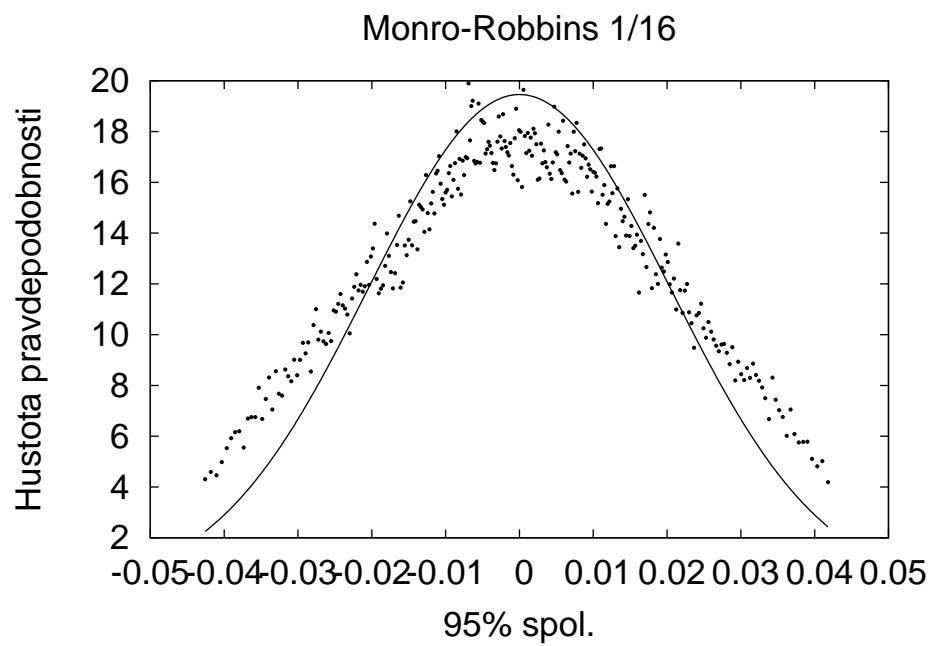
Pri meraniach boli použité tri šírky intervalov: 1024, 16 a $1/16$, algoritmy hľadali koreň sigmoidy. Do grafov sú vynesené ich predpovede po 100 krokoch a funkcia normálneho rozdelenia so štandardnou odchýlkou rovnou štandardnej odchýlke spočítanej z 95 % najmenej odchylených dát.



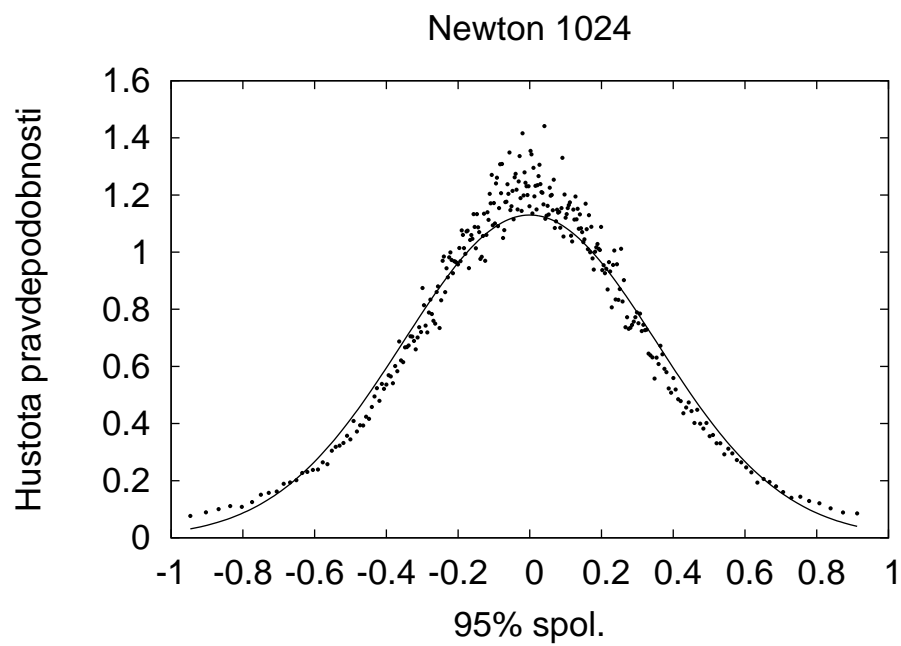
Obr. B.5: Distribúcia predpovedí Robbins-Monro algoritmu na intervale šírky 1024



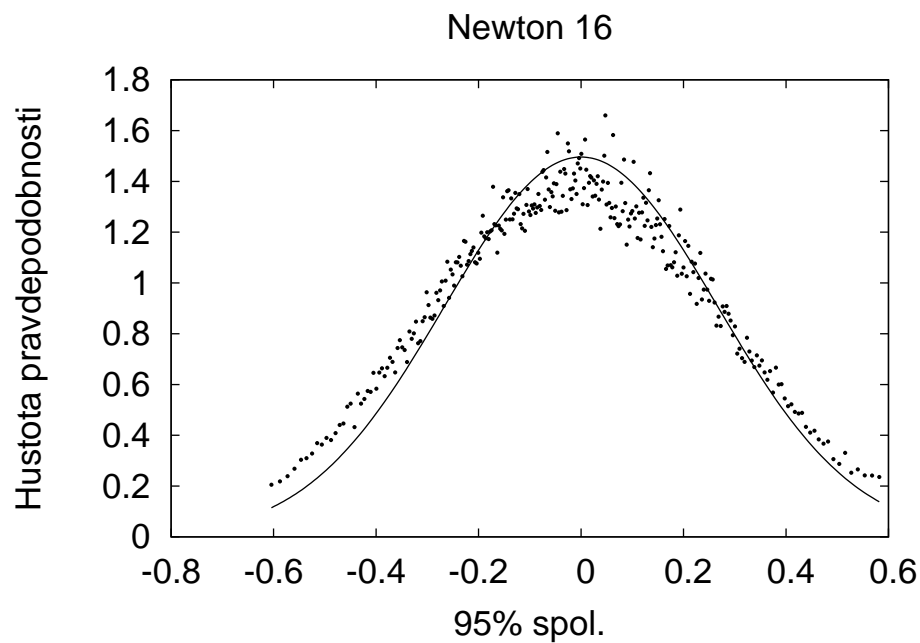
Obr. B.6: Distribúcia predpovedí Robbins-Monro algoritmu na intervale šírky 16



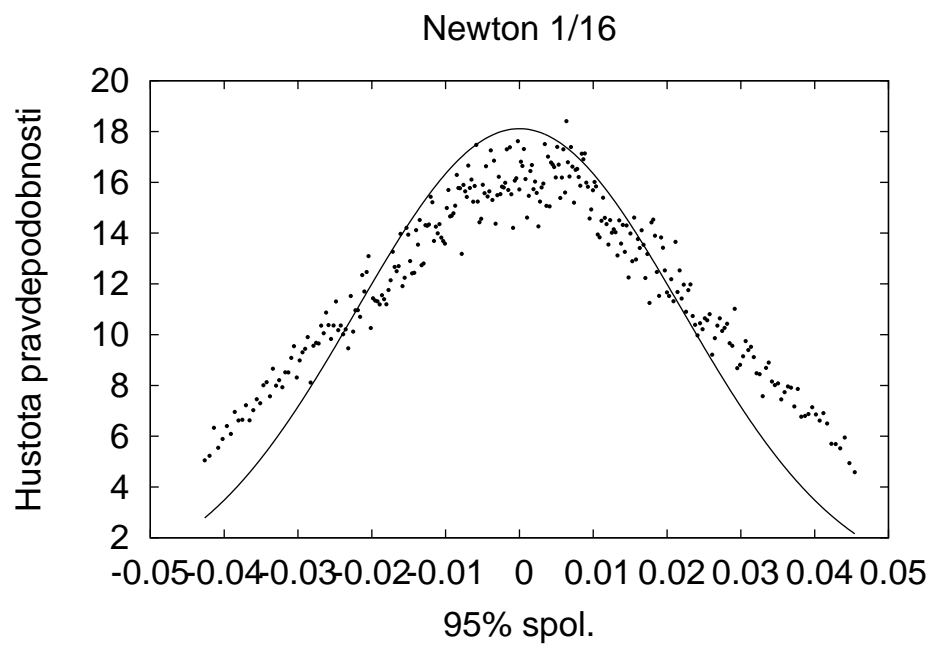
Obr. B.7: Distribúcia predpovedí Robbins-Monro algoritmu na intervale šírky $1/16$



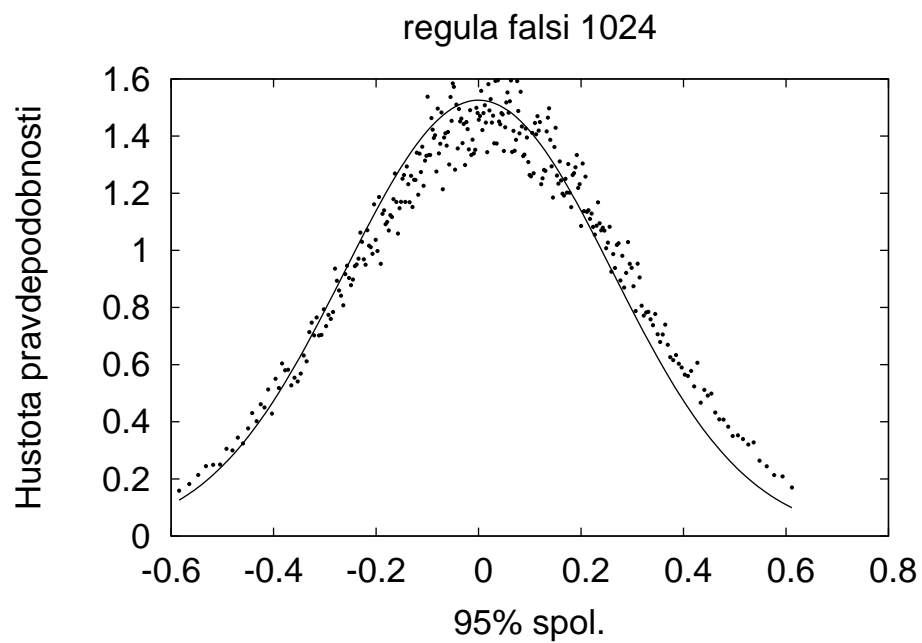
Obr. B.8: Distribúcia predpovedí Newtonovho algoritmu na intervale šírky 1024



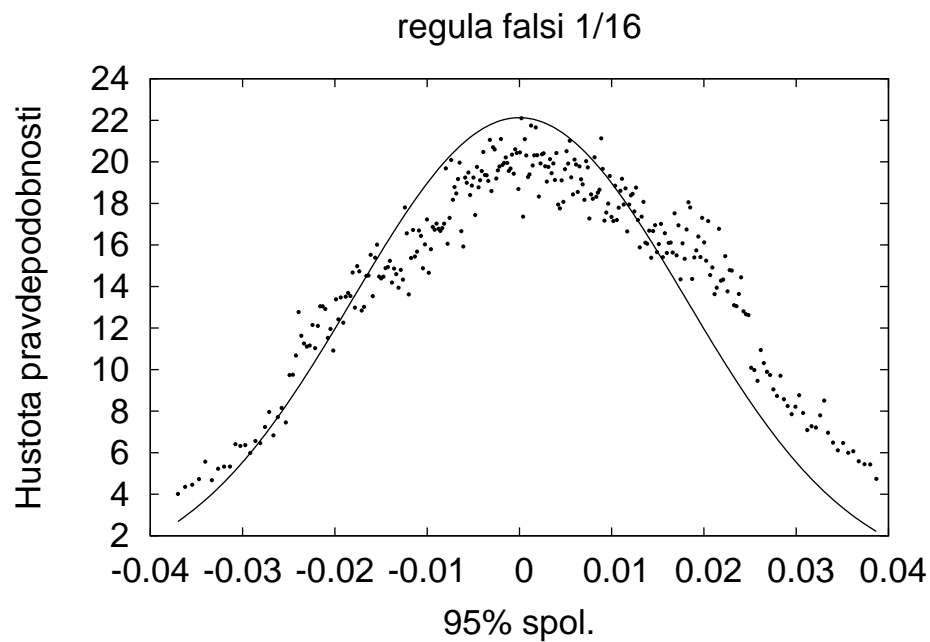
Obr. B.9: Distribúcia predpovedí Newtonovho algoritmu na intervale šírky 16



Obr. B.10: Distribúcia predpovedí Newtonovho algoritmu na intervale šírky 1/16



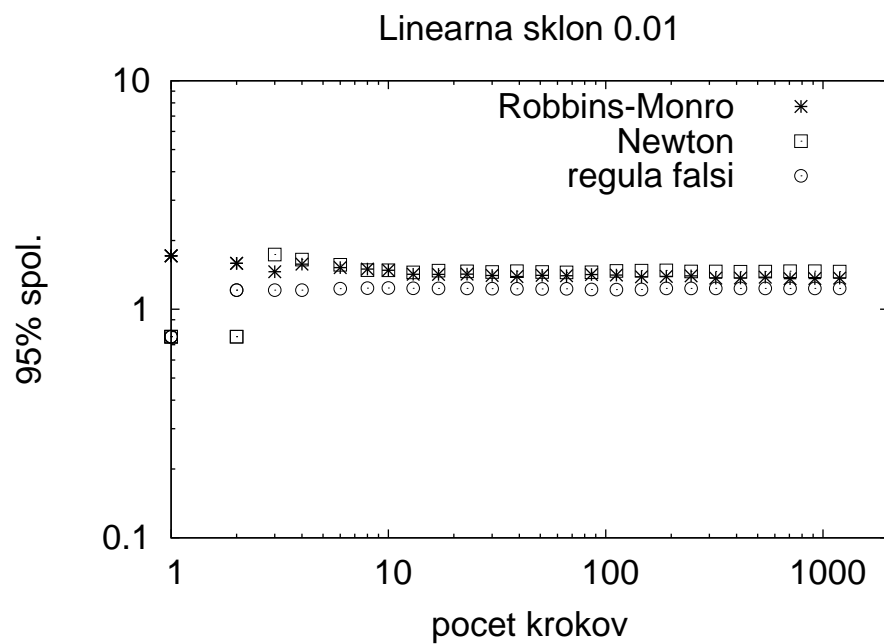
Obr. B.11: Distribúcia predpovedí regula falsi na intervale šírky 1024



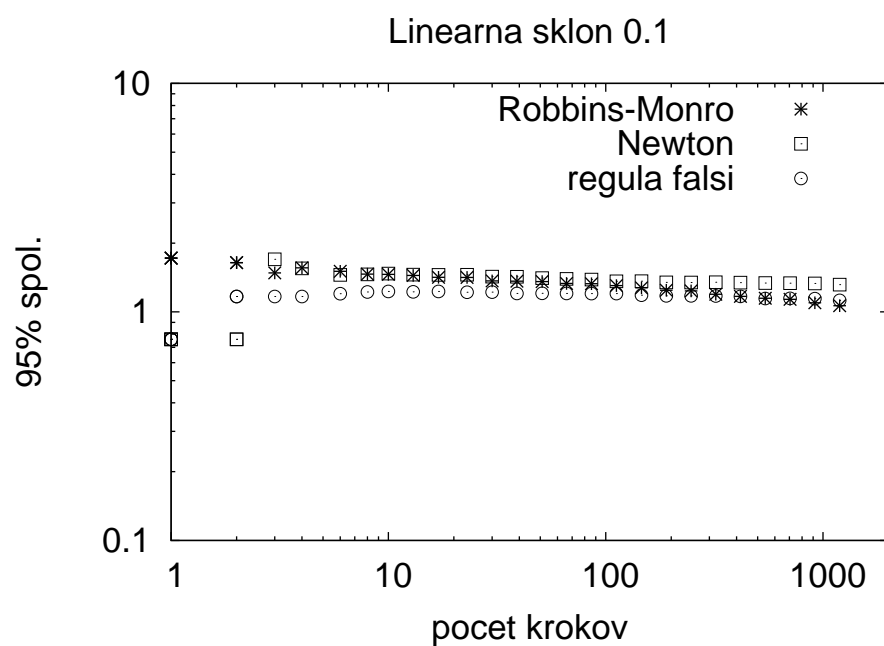
Obr. B.12: Distribúcia predpovedí regula falsi na intervale šírky 1/16

B.3 Asymptotické správanie sa algoritmov

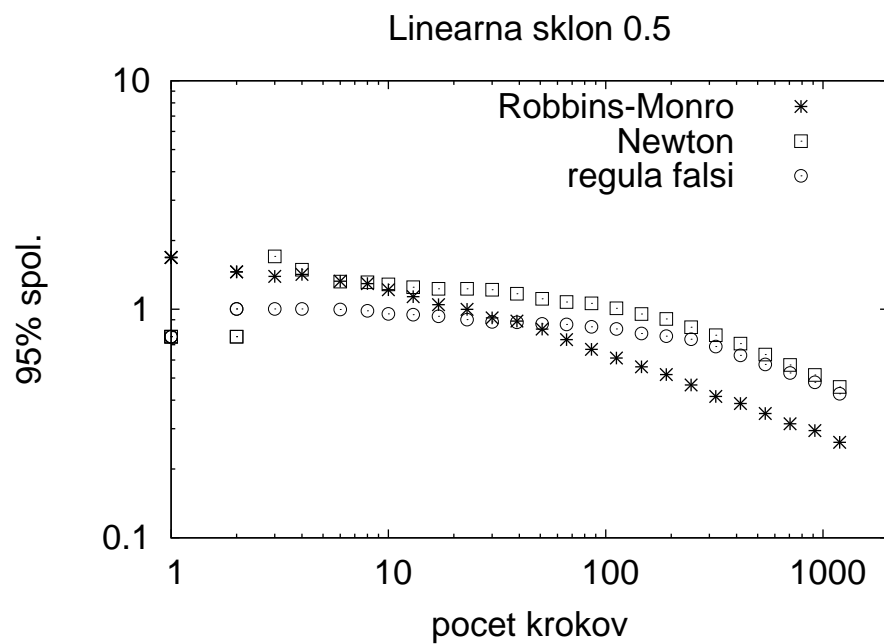
Merali sme lineárnu funkciu na intervale dĺžky 1 so sklonmi 0,01 až 100. Do grafov sme vyniesli šírku ich 95 % intervalu spoľahlivosti v závislosti od počtu krokov. Spravili sme aj jedno porovnanie originálnej Newtonovej metódy s jej vylepšenou verziou.



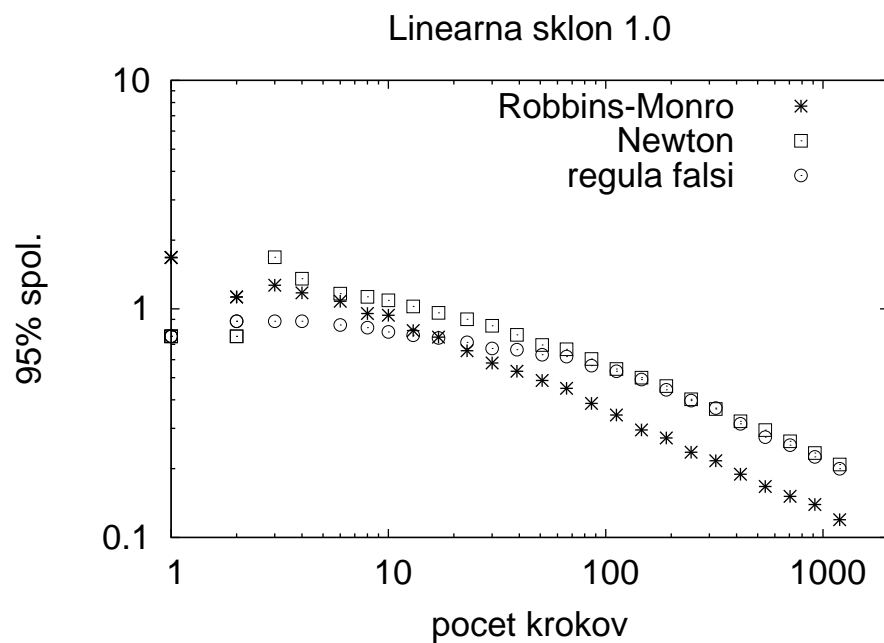
Obr. B.13: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 0,01



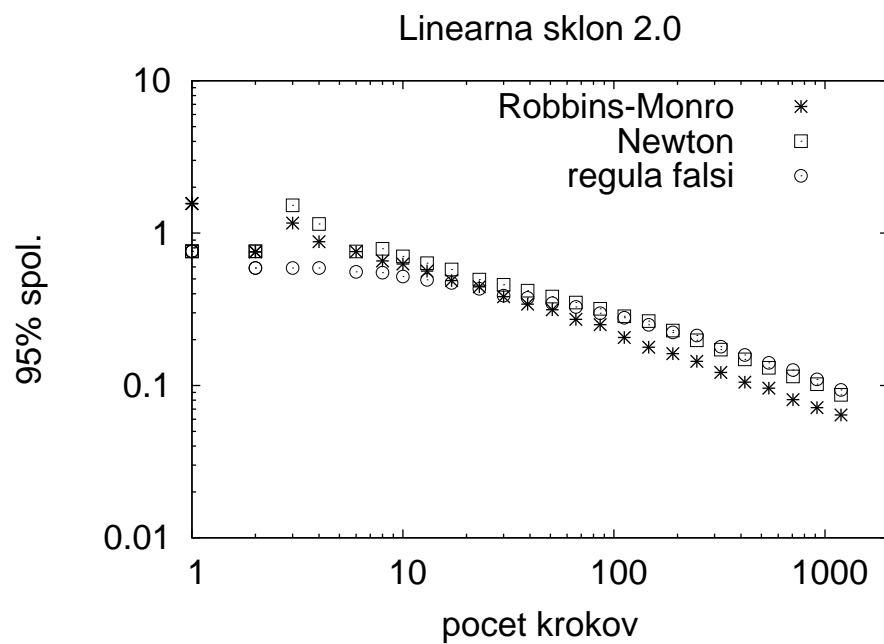
Obr. B.14: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 0,1



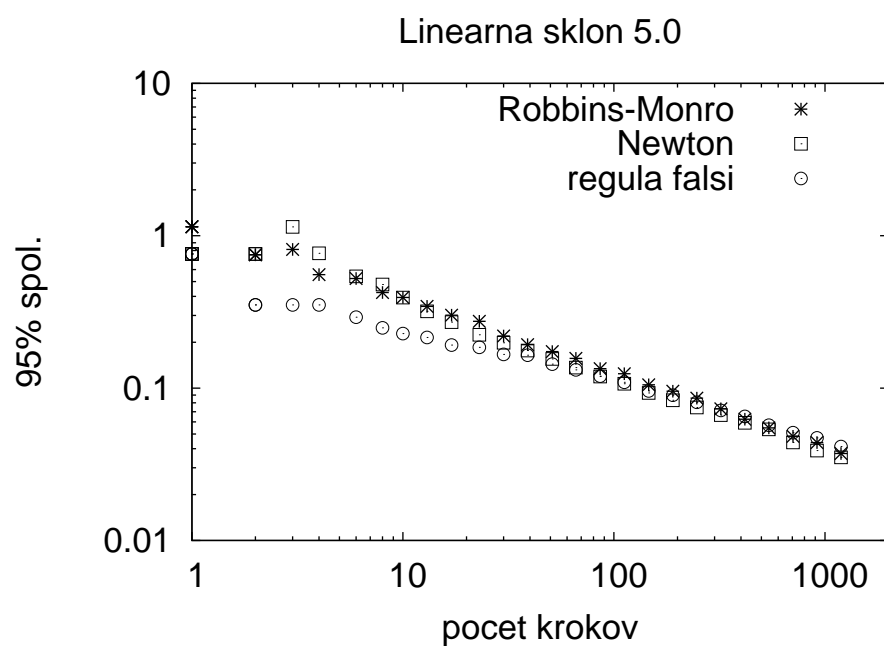
Obr. B.15: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 0,5



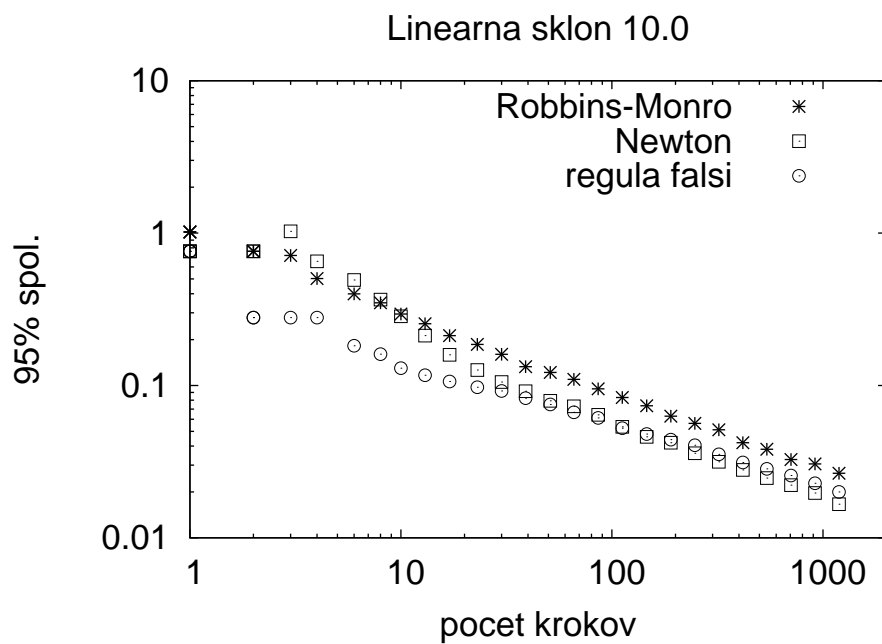
Obr. B.16: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 1,0



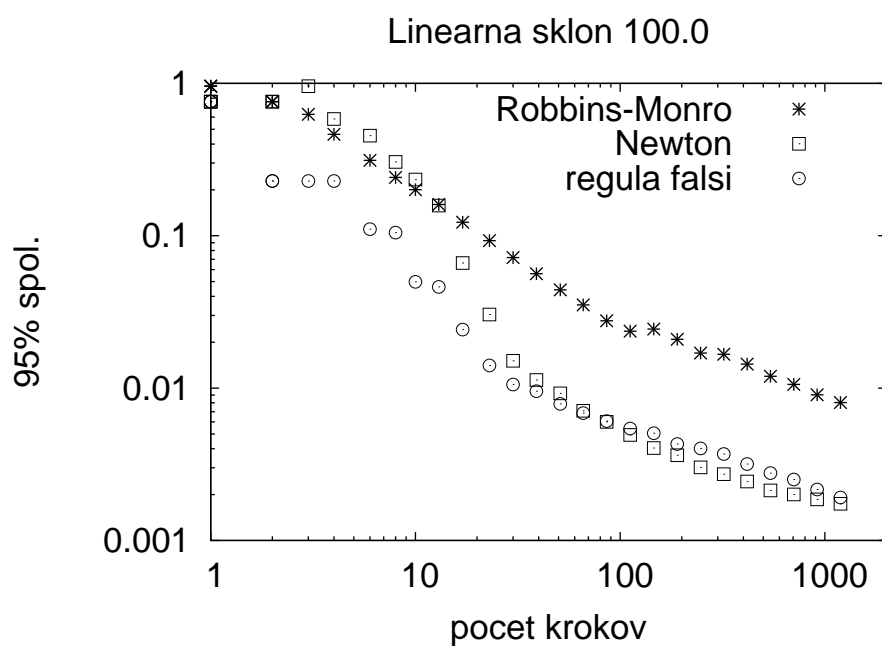
Obr. B.17: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 2,0



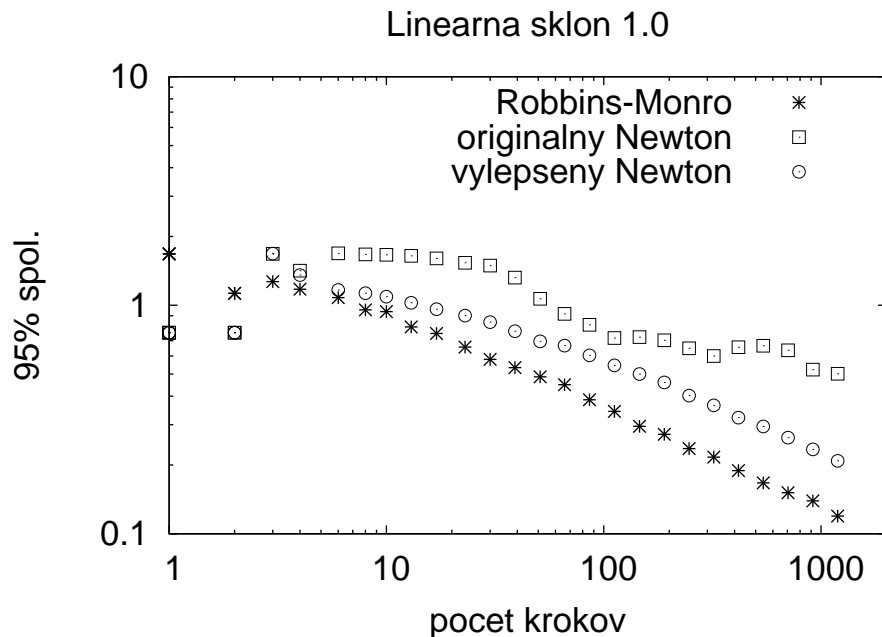
Obr. B.18: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 5,0



Obr. B.19: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 10,0



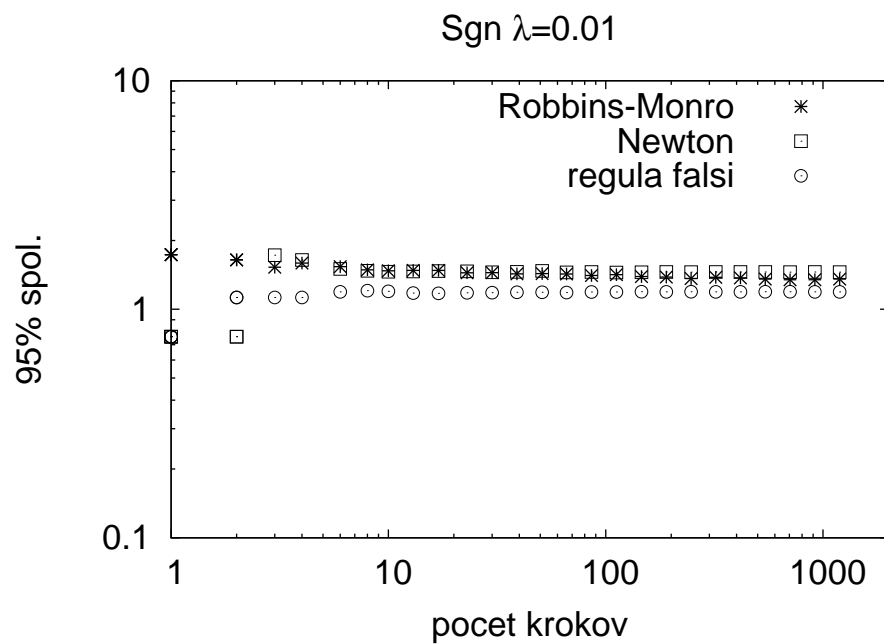
Obr. B.20: Šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 100,0



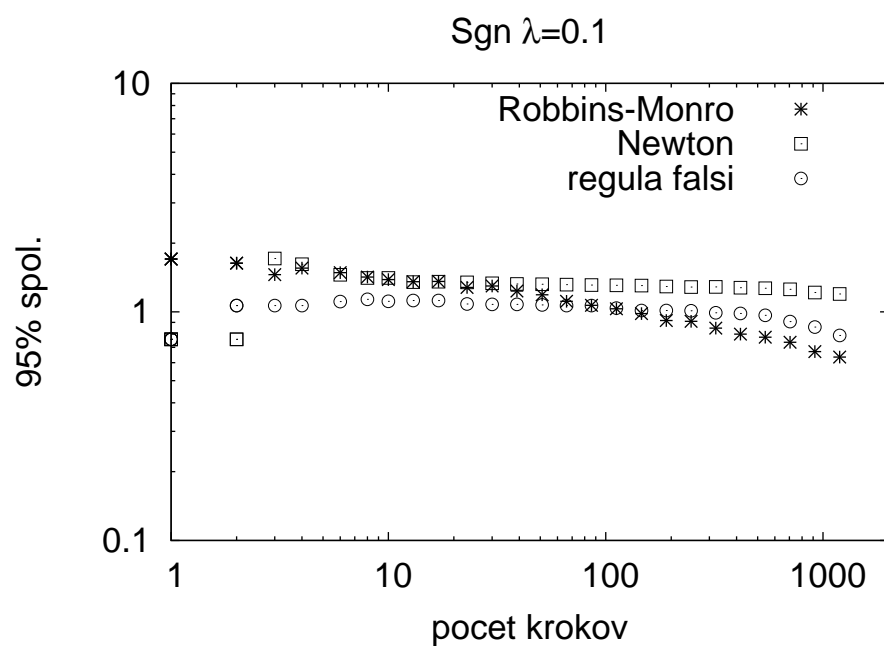
Obr. B.21: Porovnanie originálnej Newtonovej metódy s metódu zlepšenou za pomoci heuristik a s algoritmom Robbinsa-Monroa. Do grafu sú vynesené šírky intervalov spoľahlivosti pre lineárnu funkciu so sklonom 1,0

B.4 Monótonne oblasti

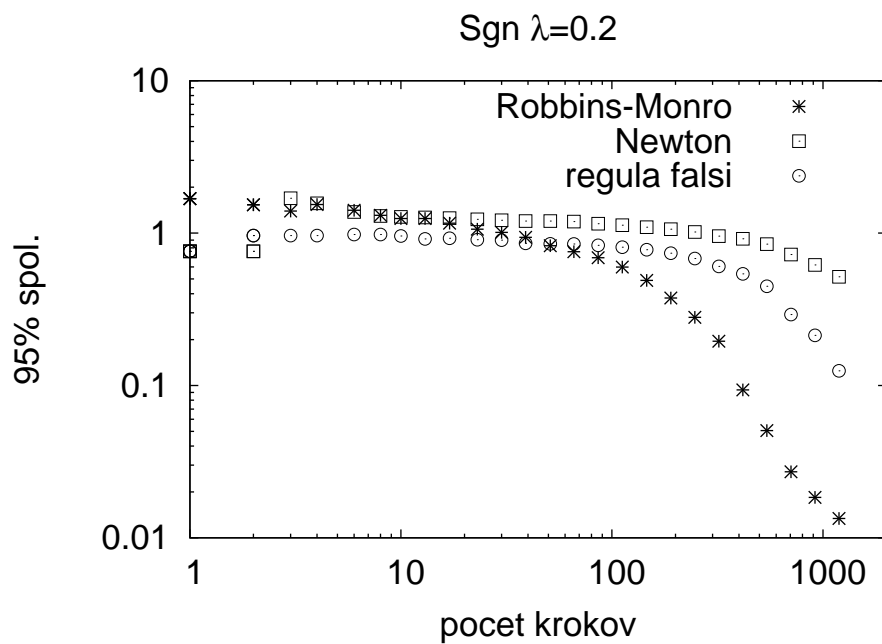
Merali sme $\lambda \operatorname{sgn}$ funkciu na intervale dĺžky 1 s parametrom λ z intervalu $[0,01, 1]$. Do grafov sme vyniesli šírku ich 95 % intervalu spoľahlivosti v závislosti od počtu krokov. Spravili sme aj jedno porovnanie originálnej Newtonovej metódy s jej vylepšenou verziou.



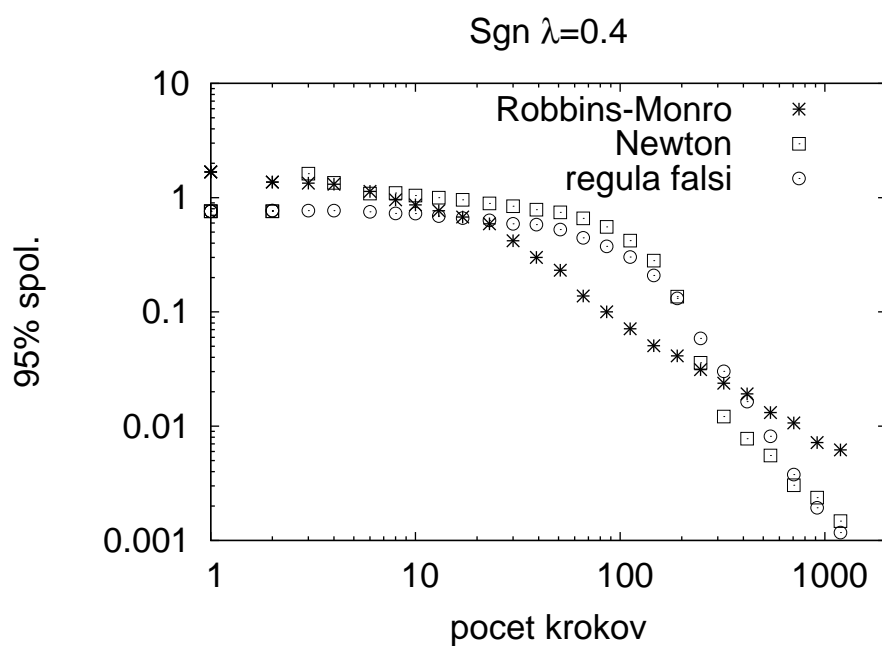
Obr. B.22: Šírky intervalov spoľahlivosti pre 0,01 sgn



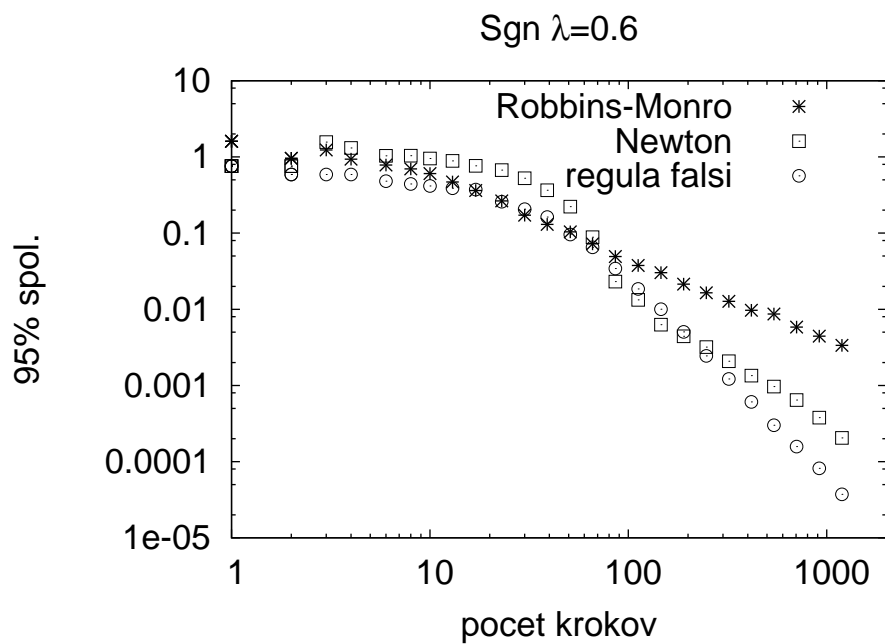
Obr. B.23: Šírky intervalov spoľahlivosti pre 0,1 sgn



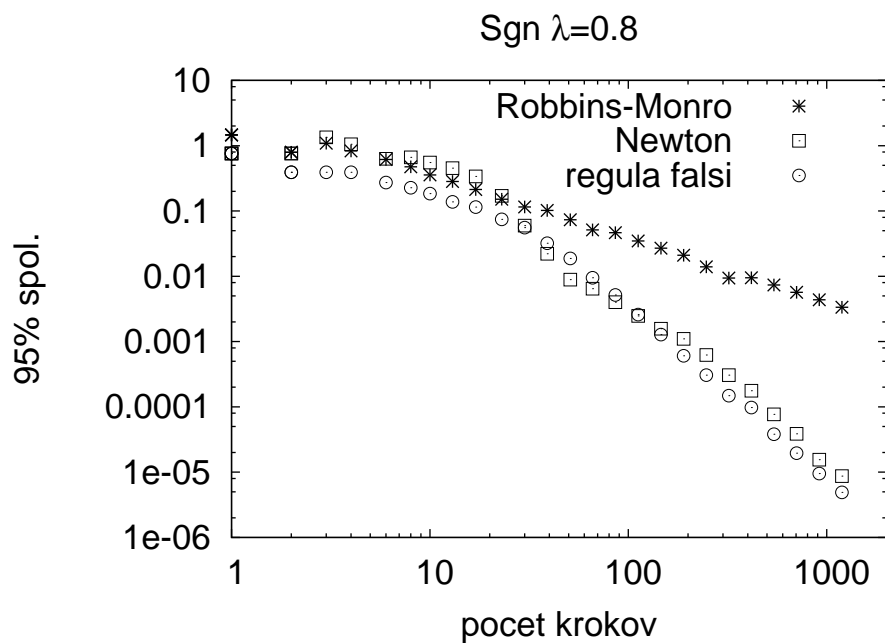
Obr. B.24: Šírky intervalov spoľahlivosti pre 0,2 sgn



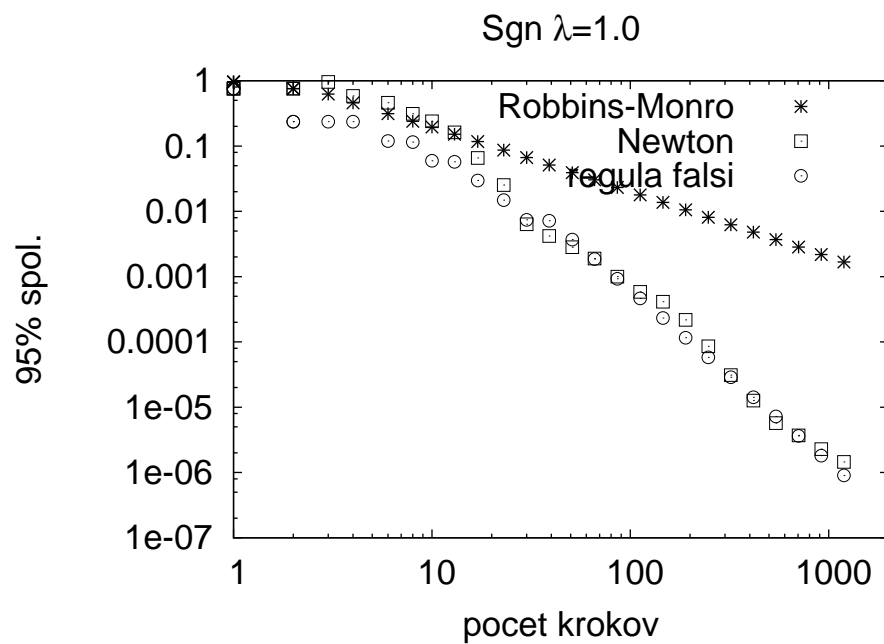
Obr. B.25: Šírky intervalov spoľahlivosti pre 0,4 sgn



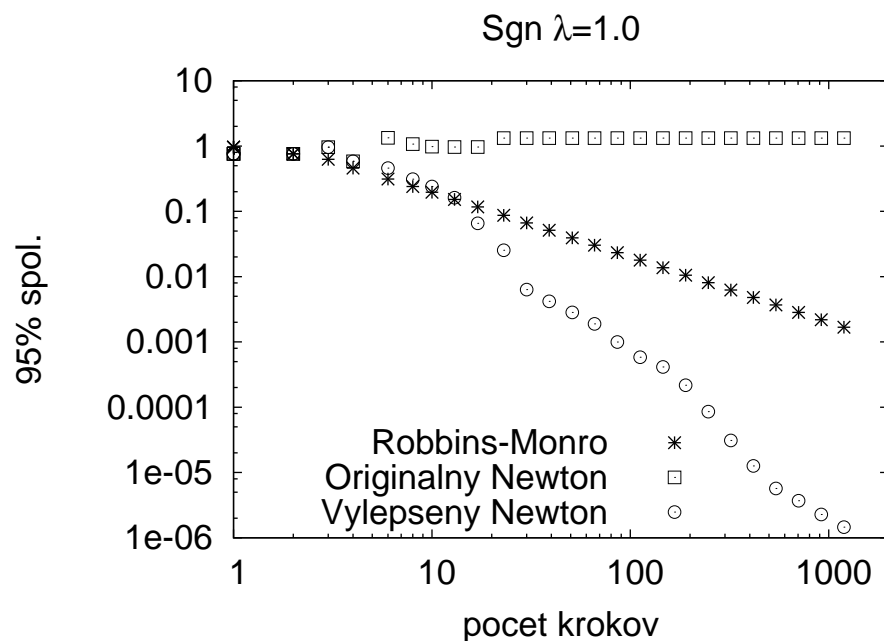
Obr. B.26: Šírky intervalov spoľahlivosti pre 0,6 sgn



Obr. B.27: Šírky intervalov spoľahlivosti pre 0,8 sgn



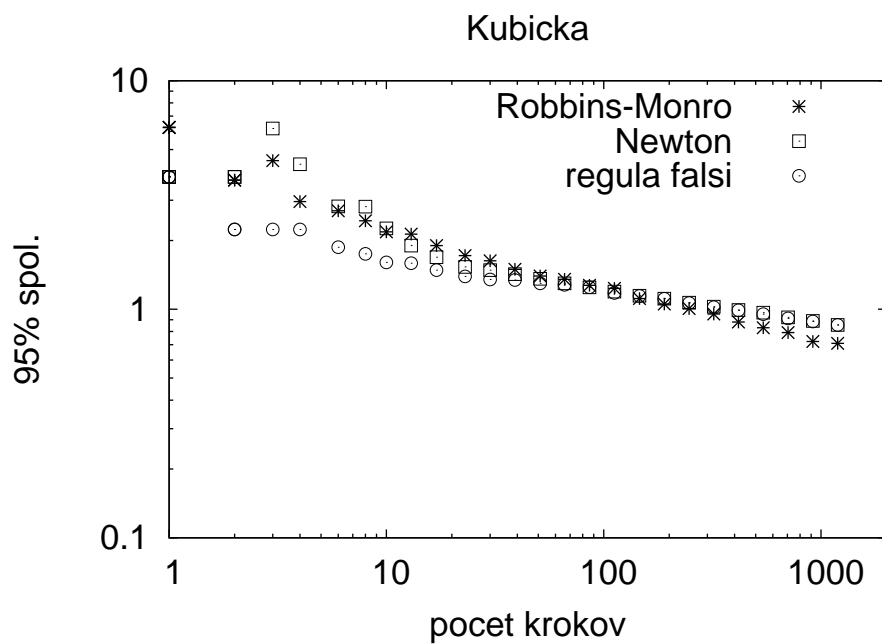
Obr. B.28: Šírky intervalov spoľahlivosti pre 1,0 sgn



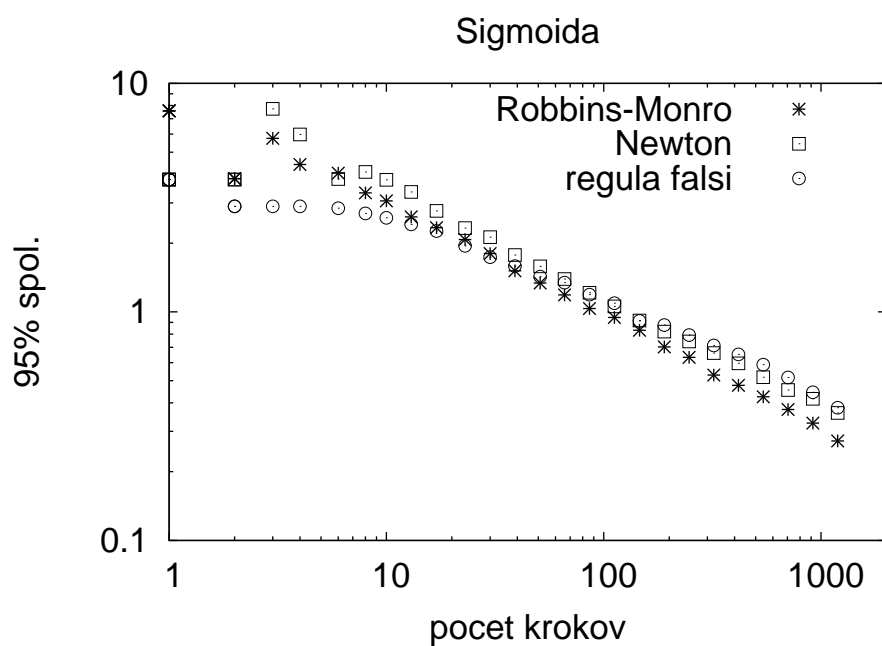
Obr. B.29: Porovnanie originálnej Newtonovej metódy s metódu zlepšenou za pomoci heuristik a s algoritmom Robbinsa-Monroa. Do grafu sú vynesené šírky intervalov spoľahlivosti pre $1,0 \text{ sgn}$

B.5 Rôzne funkcie

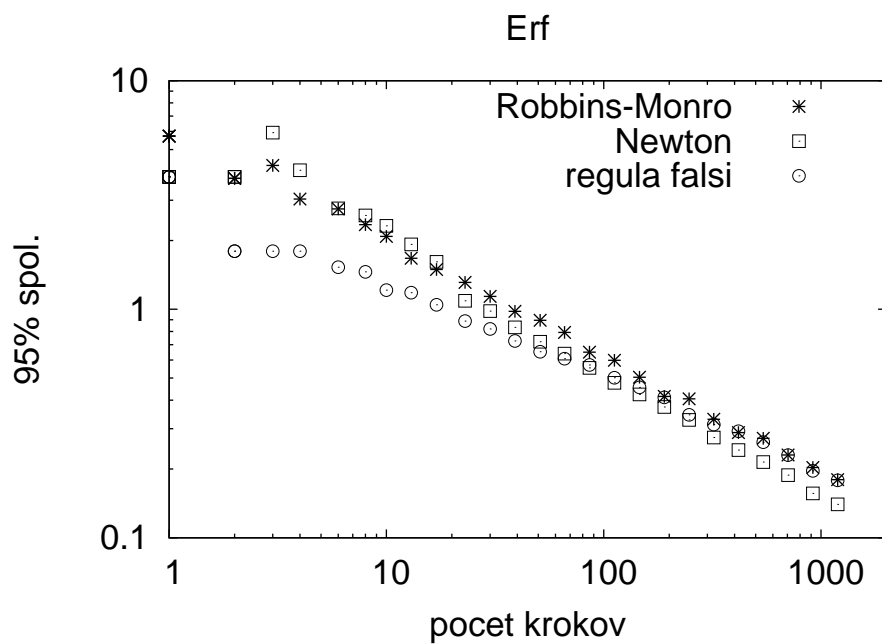
Na záver výsledky meraní na funkciách z tabuľky A.1. Všetky funkcie boli merané na intervale dĺžky 5 a do grafov sme vyniesli šírku ich 95 % intervalu spoľahlivosti.



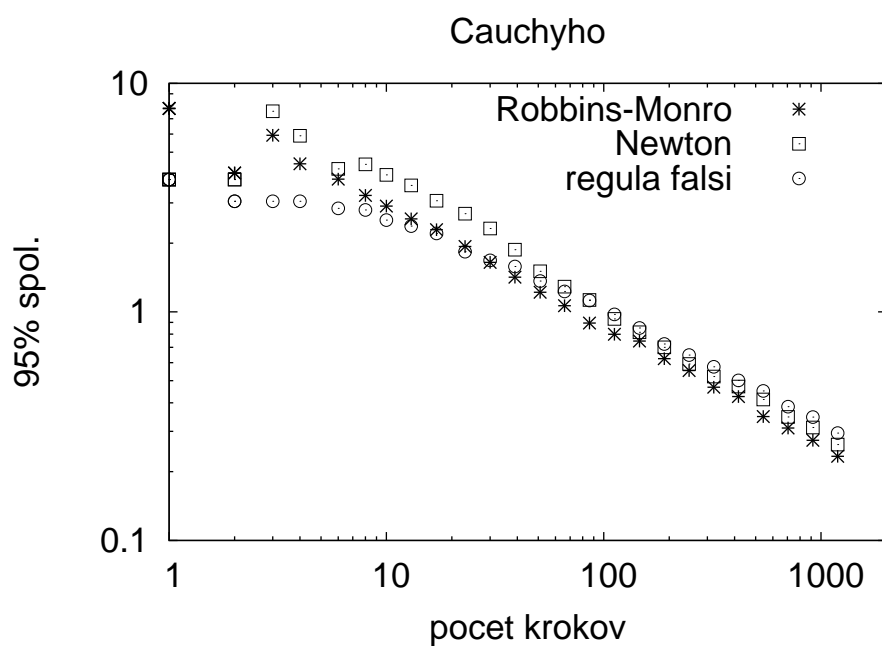
Obr. B.30: Šírky intervalov spoľahlivosti pre kubickú funkciu



Obr. B.31: Šírky intervalov spoľahlivosti pre sigmoidu



Obr. B.32: Šírky intervalov spoľahlivosti pre erf



Obr. B.33: Šírky intervalov spoľahlivosti pre Cauchyho funkciu

Literatúra

- [1] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* 22, 400–407
- [2] D. Anbar (1978). A Stochastic Newton-Raphson Method. *JSPI* 2, 153–163
- [3] Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* 29, 373–405
- [4] Pasupathy, R. and Kim, S. (2011). The stochastic root-finding problem: Overview, solutions, and open questions. *TOMACS* 22/3
- [5] Kulich, T. and Flegr, J. (2010). Positive effects of multiple gene control on the spread of altruism by group selection *Výjde v Journal of Theoretical Biology* id: 10.1016/j.jtbi.2011.05.017 Draft dostupný na <http://arxiv.org/abs/1002.4204>