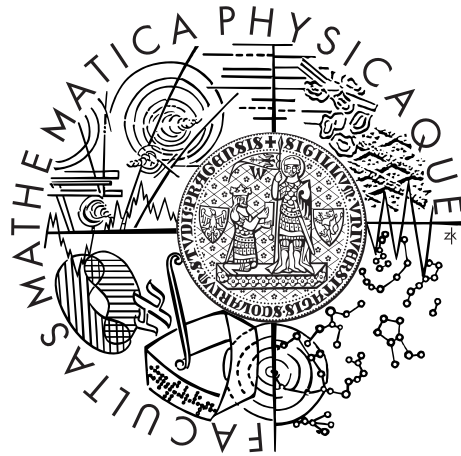


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Martin Vdovičenko

Testování statistických vlastností časových řad derivátových cen

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Matěj Kadavý

Studijní program: Matematika

Studijní obor: Obecná matematika

Praha 2011

Na tomto mieste by som sa rád pod'akoval Mgr. Matějovi Kadavému za odborné vedenie, cenné pripomienky a konzultácie pri písaní mojej bakalárskej práce. Takisto by som sa chcel pod'akovať firme RSJ a.s. za poskytnutie dát.

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 dne

Podpis autora

Název práce: Testování statistických vlastností časových řad derivátových cen

Autor: Martin Vdovičenko

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Matěj Kadavý, Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V tejto práci sa budeme zaoberať Brownovým pohybom a jeho základnými transformáciami. Popíšeme základné vlastnosti jeho trajektórii a ukážeme, že Brownov pohyb je martingal a self-similar proces. Ďalej sa budeme venovať analýze časových radov. Uvedieme grafické nástroje na skúmanie dát a popíšeme teoretické základy vybraných testov normality a nezávislosti. Nakoniec sa budeme zaoberať hypotézou, že z krátkodobého hľadiska môžeme cenu finančných aktív modelovať práve Brownovým pohybom. V štatistickom programe R prevedieme základné štatistické testy na reálnych dátach a rozoberieme získané výsledky.

Klíčová slova: náhodný proces, Brownov pohyb, štatistické testy

Title: Statistical Testing of Time Series Properties of Derivative Prices

Author: Martin Vdovičenko

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Matěj Kadavý, Department of Probability and Mathematical Statistics

Abstract: This work discusses Brownian motion and its basic transformations. The work describes basic properties of its trajectories and shows that Brownian motion is a martingale and a self-similar process. Next, we discuss time series analysis. We introduce graphical tools for analyzing data and we describe theoretical basics of some normality and independence tests. Finally, we consider the hypothesis that in the short run the price of financial assets can be modelled by Brownian motion. We conduct basic statistical tests on real data using the R program and we talk through our results.

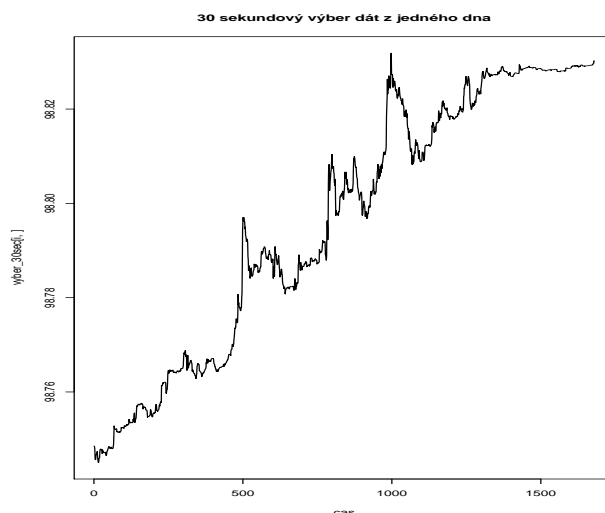
Keywords: stochastic process, Brownian motion, statistical tests

Obsah

Úvod	2
1 Brownov pohyb	3
1.1 Základné definície	3
1.2 Brownov pohyb	4
1.2.1 Kvadratická variácia Brownovho pohybu	5
1.2.2 Martingalová vlastnosť	7
1.2.3 Markovská vlastnosť	8
1.2.4 Princíp reflexie	10
1.2.5 Vlastnosť self-similarity	12
1.2.6 Lévyho modul spojitosti	13
1.3 Geometrický Brownov pohyb	13
2 Testy	16
2.1 Normalita	16
2.1.1 Grafická analýza dát	16
2.1.2 Testy normality	19
2.2 Stacionarita	21
2.3 Nezávislosť	21
2.4 Testovanie self-similarity pre Brownov pohyb	23
3 Testovanie štatistických vlastností	25
3.1 Prvé spracovanie dát	25
3.2 Stacionarita a nezávislosť	26
3.3 Normalita	26
3.4 Výberové momenty	28
3.5 Self-similarity	28
Záver	30
Zoznam použitej literatúry	31

Úvod

Predpokladá sa, že ceny finančných derivátov sa dajú z krátkodobého hľadiska modelovať pomocou Brownovho pohybu, resp. geometrického Brownovho pohybu. Moja práca spočíva v testovaní týchto vlastností na reálnych dátach. Konkrétne na cene podkladového aktíva, ktorého cena je zaznamenávaná v priebehu polroka každú sekundu od 8.00 hod do 22.00 hod (obrázok 1). Jedná sa o referenčnú úrokovú sadzbu Euribor (Euro Interbank Offered Rate), ktorá je obchodovaná na londýnskej burze cenných papierov LIFFE (London International Financial Futures and Option Exchange). Euribor je aritmetický priemer úrokových sadzieb z pôžičiek ponúkaných bankami. Ide teda o obchodovanie futures kontraktov, t.j. o individuálne zjednané termínované kontrakty na budúcu dodávku podkladového aktíva - konkrétne sa jedná o trojmesačné futures. Euribor je na LIFFE jeden z najlikvidnejších kontraktov na svete, preto môžeme predpokladať, že je možné zabezpečiť kontinuálne, prudké a adekvátne premietanie nových, neočakávaných informácií do kurzov cenných papierov, čo je jedna z kľúčových vlastností efektívneho trhu. Na testovanie som používal štatistický program R.



Obrázok 1: 30 sekundový výber z údajov jedného dňa

1. Brownov pohyb

1.1 Základné definície

Na začiatok uvedieme základné definície a pojmy potrebné na zavedenie Brownovho pohybu.

Definícia 1. *Nech (Ω, \mathcal{F}, P) je pravdepodobnostný priestor a nech $T \subset \mathbb{R}$. Potom rodina reálnych náhodných veličín $X = \{X_t, t \in T\}$ definovaných na (Ω, \mathcal{F}, P) sa nazýva stochastický (náhodný) proces.*

Vo všeobecnosti množina T predstavuje čas. Môžeme rozlíšiť dva prípady. Ak je $T = \mathbb{Z}$ množina všetkých celých čísel, tak hovoríme o procese s diskretným časom alebo o časovom rade. V druhom prípade, ak $T = [a, b]$, $-\infty \leq a < b \leq \infty$, hovoríme o procese so spojitým časom. Vo väčšine prípadov však pozorujeme náhodný proces od nejakého počiatočného času, vtedy za T volíme \mathbb{R}^+ .

Na náhodný proces $X = \{X_t, t \in \mathbb{R}^+\}$ teda môžeme pozerat' ako na zobrazenie

$$\begin{aligned} X : \Omega \times \mathbb{R}^+ &\longmapsto \mathbb{R} \\ (\omega, t) &\longmapsto X(\omega, t), \end{aligned}$$

kde $X(\omega, t) := X_t(\omega)$.

Ak zafixujeme $\omega \in \Omega$, tak sa funkcia

$$X(\omega, \cdot) : t \longmapsto X_t(\omega) \quad t \geq 0,$$

nazýva trajektória (realizácia) náhodného procesu pre dané ω a skúmame jej analytické vlastnosti. Pre pevné $t \in \mathbb{R}^+$ je naopak $X(\cdot, t)$ náhodná veličina.

Definícia 2. *Bud' (Ω, \mathcal{F}) merateľný priestor a $\emptyset \neq T \subseteq \mathbb{R}$. Neklesajúci systém $(\mathcal{F}_t, t \in T)$ pod- σ -algebrií \mathcal{F} sa nazýva filtrácia, ak platí*

$$\forall s, t \in T, \quad s \leq t, \quad \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}.$$

Navyše definujeme limitnú σ -algebru

$$\mathcal{F}_\infty = \sigma \left(\bigcup_{t > 0} \mathcal{F}_t \right).$$

Zavedenie filtrácie nám dáva možnosť vedieť, akú informáciu budeme mať v budúcnosti, t.j. v čase t budeme o každom jave A zo σ -algebry \mathcal{F}_t vedieť, či nastal alebo nie. Ak máme na (Ω, \mathcal{F}) definovaný náhodný proces X , tak najjednoduchšia voľba filtrácie $(\mathcal{F}_t, t \in T)$ je taká, ktorá je generovaná samotným procesom X .

Definícia 3. *Kanonickou filtráciou náhodného procesu $X = (X_t, t \in T)$ rozumieme filtráciu $(\mathcal{F}_t^X, t \in T)$, kde*

$$\mathcal{F}_t^X = \sigma(X_s; 0 \leq s \leq t, s \in T).$$

Kanonickú filtráciu interpretujeme tak, že \mathcal{F}_t^X obsahuje všetky možné varianty náhodného procesu do času t .

Definícia 4. *Nech $X = (X_t, t \in T)$ je náhodný proces a $(\mathcal{F}_t, t \in T)$ je filtrácia. Povieme, že náhodný proces X je adaptovaný k filtrácii \mathcal{F}_t , ak náhodná veličina X_t je \mathcal{F}_t -merateľná pre každé $t \in T$.*

Definícia 5. *Nech $(\mathcal{F}_t, t \in T)$ je filtrácia na (Ω, \mathcal{F}, P) . Položme $\mathcal{F}_{t+} = \bigcap_{s>t, s \in T} \mathcal{F}_s$. Potom povieme, že filtrácia $(\mathcal{F}_t, t \in T)$ je sprava spojitá, ak $\mathcal{F}_t = \mathcal{F}_{t+}$ pre všetky $t \in T$.*

1.2 Brownov pohyb

Podrobnejšou konštrukciou Brownovho pohybu sa v tomto texte zaoberať nebudeme. Jeho existenciu dokázal Brown a neskôr aj Lévy. Vznikne postupne z modelu náhodnej prechádzky, v ktorom sa časový úsek približi limitne k nule.

Definícia 6. *Bud' (Ω, \mathcal{F}, P) pravdepodobnostný priestor. Potom náhodný proces $W = (W_t, t \geq 0)$ definovaný na (Ω, \mathcal{F}, P) nazveme Brownov pohyb, ak spĺňa nasledujúce podmienky*

- $W_0 = 0$,
- W_t je P -s.i. spojitý, t.j. $W(\omega, \cdot) : t \rightarrow W_t(\omega), t \geq 0$ je spojitá funkcia,
- pre všetky $0 \leq s_1 \leq s_2 \leq t_1 \leq t_2 \leq \infty$ sú náhodné veličiny $W_{t_2} - W_{t_1}$ a $W_{s_2} - W_{s_1}$ nezávislé,
- $\mathcal{L}(W_t - W_s) = N(0, \sigma^2|t - s|)$ pre všetky $t, s > 0$ a $\sigma > 0$ je konštanta a \mathcal{L} označuje pravdepodobnostné rozdelenie.

Inými slovami Brownov pohyb je náhodný proces, ktorý má normálne rozdelené nezávislé prírastky s nulovou strednou hodnotou a rozptylom rovným dĺžke časového úseku vynásobeného nejakou konštantou σ . Ak je $\sigma = 1$, hovoríme o tzv. štandardnom Brownovom pohybe. K Brownovmu pohybu $(W_t, t \geq 0)$ sa viaže pojem *Brownova filtrácia* $(\mathcal{F}_t, t \in T)$. Je to filtrácia, ku ktorej je Brownov pohyb $(W_t, t \geq 0)$ adaptovaný a navyše pre každé $0 \leq t < s$ je prírastok $W_s - W_t$ nezávislý na σ -algebre \mathcal{F}_t . Od tejto chvíle, ak budeme uvažovať Brownov pohyb na nejakom pravdepodobnostnom priestore s filtráciou, tak máme na mysli práve Brownovu filtráciu.

Veta 1. *Nech $W = (W_t, t \geq 0)$ je Brownov pohyb. Potom platia nasledujúce vlastnosti:*

- (časová homogenita). *Pre každé $s > 0$, je proces $(W_{t+s} - W_s, t \geq 0)$ Brownov pohyb,*
- (symetria). *Proces $(-W_t, t \geq 0)$ je Brownov pohyb,*
- (scaling). *Pre každé $c > 0$ je proces $(cW_{t/c^2}, t \geq 0)$ Brownov pohyb,*
- (časová inverzia). *Proces X definovaný ako $X_0 = 0$ a $X_t = tW_{1/t}$ pre $t > 0$ je Brownov pohyb.*

Procesy definované vyššie sa nazývajú transformácie Brownovho pohybu. Vlastnosť scaling sa voľne prekladá ako zmena merítka.

Dôsledok 1. *Nech $W = (W_t, t \geq 0)$ je Brownov pohyb, potom*

$$\lim_{t \rightarrow \infty} \frac{W_t}{t} = 0 \quad \text{s.i.}$$

Dôkaz. Vieme, že proces X definovaný v štvrtom bode predošlej vety je Brownov pohyb. Po preznačení potom platí

$$\lim_{t \rightarrow \infty} \frac{W_t}{t} = \lim_{t \rightarrow \infty} X_{1/t} = 0.$$

□

Táto vlastnosť sa dá dokázať aj priamo použitím silného zákona veľkých čísel.

1.2.1 Kvadratická variácia Brownovho pohybu

Ďalej zavedieme pojem kvadratická variácia a ukážeme, že je pre Brownov pohyb skoro iste nenulová.

Definícia 7. *Nech $f(t)$ je funkcia definovaná pre $0 \leq t \leq T$. Potom kvadratická variácia funkcie f do času T je*

$$[f, f](T) = \lim_{\|\Pi\| \rightarrow 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2,$$

kde $\Pi = \{t_0, t_1, \dots, t_n\}$ je delenie intervalu $[0, T]$, $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ a $\|\Pi\| = \max \{t_{i+1} - t_i, i = 0, \dots, n-1\}$.

Väčšina funkcií, s ktorými sa bežne stretávame, má spojité derivácie. Dá sa ukázať, že ak f je funkcia so spojitou prvou deriváciou, tak jej kvadratická variácia je nulová.

Tvrdenie 1. *Ak má funkcia spojitú prvú deriváciu, tak jej kvadratická variácia do času T je nulová.*

Dôkaz. Budeme postupovať podľa [6] str. 101. Uvažujme Π delenie intervalu $[0, T]$. Z Lagrangeovej vety o strednej hodnote, ktorej podmienky sú triviálne splnené, existuje v každom intervale (t_j, t_{j+1}) bod t_j^* taký, že platí

$$f'(t_j^*) = \frac{f(t_{j+1}) - f(t_j)}{t_{j+1} - t_j}.$$

Potom môžeme písať

$$\sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 = \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j)^2 \leq \|\Pi\| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j),$$

a preto

$$\begin{aligned} [f, f](T) &\leq \lim_{\|\Pi\| \rightarrow 0} \left[\|\Pi\| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j) \right] \\ &= \lim_{\|\Pi\| \rightarrow 0} \|\Pi\| \cdot \lim_{\|\Pi\| \rightarrow 0} \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j) = \lim_{\|\Pi\| \rightarrow 0} \|\Pi\| \cdot \int_0^T |f'(t)|^2 dt = 0. \end{aligned}$$

Pričom posledná rovnosť plynie z faktu, že $f'(t)$ je spojitá funkcia, preto je integrál konečný. \square

Avšak pre trajektórie Brownovho pohybu platí, že jeho kvadratická variácia cez nejaký interval je rovná jeho dĺžke.

Veta 2. *Bud' W Brownov pohyb. Potom $[W, W](T) = T$ pre všetky $T > 0$ skoro iste.*

Dôkaz. Postup je založený na [6] str. 102. Nech $\Pi = \{t_0, t_1, \dots, t_n\}$. Pre lepšiu orientáciu budeme v tomto dôkaze písať $W(t)$ namiesto W_t . Definujme náhodnú veličinu

$$Q_\Pi = \sum_{j=0}^{n-1} (W(t_{j+1}) - W(t_j))^2.$$

Chceme dokázať, že Q_Π konverguje skoro iste k T , keď $\|\Pi\| \rightarrow 0$. Ak by sme dokázali túto konvergenciu v L^2 , tak z Čebyševovej nerovnosti a zo silného zákona veľkých čísel plynie konvergencia skoro iste. Na dokázanie konvergencie v L^2 potrebujeme dokázať, že $E Q_\Pi = T$ a $\text{var } Q_\Pi \rightarrow 0$, keď $\|\Pi\| \rightarrow 0$. Keďže z definície Brownovho pohybu máme $E [W(t_{j+1}) - W(t_j)] = 0$, $\text{var} [W(t_{j+1}) - W(t_j)] = t_{j+1} - t_j$ a Q_Π je súčtom nezávislých veličín, tak platí

$$\begin{aligned} E Q_\Pi &= \sum_{j=0}^{n-1} E [W(t_{j+1}) - W(t_j)]^2 \\ &= \sum_{j=0}^{n-1} \text{var} [W(t_{j+1}) - W(t_j)] = \sum_{j=0}^{n-1} (t_{j+1} - t_j) = T. \end{aligned}$$

Ďalej vieme, že

$$E [W(t_{j+1}) - W(t_j)]^4 = 3(t_{j+1} - t_j)^2,$$

$$\begin{aligned} \text{var} [(W(t_{j+1}) - W(t_j))^2] &= E \left[((W(t_{j+1}) - W(t_j))^2 - (t_{j+1} - t_j))^2 \right] \\ &= 2(t_{j+1} - t_j)^2. \end{aligned}$$

Takže dostávame

$$\text{var } Q_\Pi = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \leq \sum_{j=0}^{n-1} 2\|\Pi\|(t_{j+1} - t_j) = 2\|\Pi\|T.$$

Teda sme dospeli k tomu, že $\lim_{\|\Pi\| \rightarrow 0} \text{var} (Q_\Pi) = 0$, takže $\lim_{\|\Pi\| \rightarrow 0} (Q_\Pi) = E Q_\Pi = T$. \square

Táto zaujímavá vlastnosť Brownovho pohybu hovorí, že kvadratická variácia Brownovho pohybu nezávisí na voľbe trajektórie, resp. je pre všetky trajektórie na danom intervale rovnaká. Ďalej z dokázaných tvrdení platí, že Brownov pohyb nemá spojitú prvú deriváciu. Dokonca platí, že Brownov pohyb nie je v žiadnom bode diferencovateľný skoro iste. Narozdiel od kvadratickej variácie sa dá dokázať, že na každom intervale má Brownov pohyb nekonečnú (totálnu) variáciu.

Definícia 8. Nech $f(t)$ je funkcia definovaná pre $0 \leq t \leq T$. Potom totálna variácia funkcie f do času T je

$$V_0^T(f) = \lim_{\|\Pi\| \rightarrow 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)],$$

kde $\Pi = \{t_0, t_1, \dots, t_n\}$ je delenie intervalu $[0, T]$, $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ a $\|\Pi\| = \max \{t_{i+1} - t_i, i = 0, \dots, n-1\}$.

Dôsledok 2. Brownov pohyb má nekonečnú totálnu variáciu.

Dôkaz. Predpokladajme pre spor konečnosť kvadratickej variácie. Potom máme

$$\begin{aligned} [W, W](T) &= \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|^2 \\ &\leq \max_{1 \leq j \leq n-1} |W(t_{j+1}) - W(t_j)| \cdot \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| \\ &\leq V_0^T(W) \max_{1 \leq j \leq n-1} |W(t_{j+1}) - W(t_j)|. \end{aligned}$$

Keďže W má spojitú trajektóriu skoro iste, tak

$$\max_{1 \leq j \leq n-1} |W(t_{j+1}) - W(t_j)| \xrightarrow{\|\Pi\| \rightarrow 0} 0,$$

a z konečnosti prvej variácie vyplýva, že

$$\sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|^2 \rightarrow 0 \quad \text{s.i.,}$$

čím sme dospeli k sporu a tým je tvrdenie dokázané. □

1.2.2 Martingalová vlastnosť

Definícia 9. Nech (Ω, \mathcal{F}, P) je pravdepodobnostný priestor s filtráciou $(\mathcal{F}_t, t \in T)$. Uvažujme integrovateľný adaptovaný náhodný proces $X = (X_t, t \in T)$ vzhľadom k $(\mathcal{F}_t, t \in T)$. Potom povieme, že proces X je

- \mathcal{F}_t -martingal, ak $E[X_t | \mathcal{F}_s] = X_s$, pre všetky $s \leq t$ a $s \in T$ skoro iste,
- \mathcal{F}_t -submartingal, ak $E[X_t | \mathcal{F}_s] \geq X_s$, pre všetky $s \leq t$ a $s \in T$ skoro iste,
- \mathcal{F}_t -supermartingal, ak $E[X_t | \mathcal{F}_s] \leq X_s$, pre všetky $s \leq t$ a $s \in T$ skoro iste.

V definícii martingalu sa vyskytuje podmienka $E[X_t | \mathcal{F}_s] = X_s$, pre všetky $s \leq t$ a $s \in T$, ktorú interpretujeme tak, že pri informácii v čase s je X_s verným odhadom veličiny X_t . Ak na obe strany podmienky aplikujeme strednú hodnotu E , tak vidíme, že v každom čase t má martingal konštantnú strednú hodnotu EX_t , podobne supermartingaly nerastúcu a submartingaly neklesajúcu.

Veta 3. Brownov pohyb $W = (W_t, t \geq 0)$ definovaný na pravdepodobnostnom priestore (Ω, \mathcal{F}, P) s filtráciou $(\mathcal{F}_t, t \geq 0)$ je \mathcal{F}_t -martingal.

Dôkaz. V tomto dôkaze využijeme základné vlastnosti podmienenej strednej hodnoty. Majme $0 \leq s \leq t$. Potom

$$\begin{aligned} \mathbb{E} [W_t | \mathcal{F}_s] &= \mathbb{E} [(W_t - W_s) + W_s | \mathcal{F}_s] \\ &= \mathbb{E} [(W_t - W_s) | \mathcal{F}_s] + \mathbb{E} [W_s | \mathcal{F}_s] \\ &= \mathbb{E} [W_t - W_s] + W_s \\ &= W_s \end{aligned}$$

Druhá rovnosť plynie z linearity podmienenej strednej hodnoty, tretia rovnosť vyplýva z toho, že náhodná veličina $W_t - W_s$ je nezávislá na \mathcal{F}_s a W_s je \mathcal{F}_s -merateľná. Posledná rovnosť je daná nulovou strednou hodnotou prírastkov Brownovho pohybu. \square

Pomocou kvadratickej variácie a martingalov sa dá Brownov pohyb definovať ekvivalentne s našou definíciou. Takzvaná Lévyho charakterizácia hovorí, že Brownov pohyb $W = (W_t, t \geq 0)$ je skoro iste spojitý martingal a platí $W_0 = 0$ a $[W, W](T) = T$ pre všetky $T \geq 0$.

1.2.3 Markovská vlastnosť

Definícia 10. *Povieme, že reálny náhodný proces $(X_t, t \in T)$ je Markovov proces, ak pre ľubovoľné $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$ platí*

$$P(X_{t_n} \leq x | X_{t_{n-1}}, \dots, X_{t_0}) = P(X_{t_n} \leq x | X_{t_{n-1}}) \quad \text{s.i.} \quad (1.1)$$

pre každé $x \in \mathbb{R}$.

Podmienke (1.1) v definícii hovoríme *markovská vlastnosť*. Hovoríme, že správanie sa Markovovho reťazca je „bez pamäťové“, t.j. nezávisí na jeho predošlom vývoji. V každom stave je nastanie ďalších stavov nezávislé od predošlých stavov.

Vo vete 1 sme zaviedli vlastnosť časovej homogenity, že proces $(W_{t+s} - W_s, t \geq 0)$ je Brownov pohyb, markovská vlastnosť navyše tvrdí, že tento proces je nezávislý na trajektóriách Brownovho pohybu $(W_t, t \geq 0)$ do času s , t.j. na procese $(W_r, 0 \leq r \leq s)$. Aby naše tvrdenie bolo matematicky korektné, prevedieme ho do reči filtrácií.

Veta 4 (Markovská vlastnosť). *Nech (Ω, \mathcal{F}, P) je pravdepodobnostný priestor s filtráciou $(\mathcal{F}_t, t \geq 0)$ a na ňom definovaný štandardný Brownov pohyb $W = (W_t, t \geq 0)$. Pre každé $s > 0, t \geq 0$ definujme proces $W_t^* = W_{t+s} - W_s$ a nech $(\mathcal{F}_t^*, t \geq 0)$ je jeho filtrácia. Potom pre každé $t > 0$ sú σ -algebry \mathcal{F}_s a \mathcal{F}_t^* nezávislé.*

Na to, aby sme mohli vysloviť silnejšie tvrdenie (*silnú markovskú vlastnosť*), musíme zaviesť náhodný čas. Náhodný čas definovaný na pravdepodobnostnom priestore (Ω, \mathcal{F}, P) je \mathcal{F} -merateľná reálna náhodná veličina s hodnotami v $[0, \infty]$. Teraz zavedieme dôležitú triedu náhodných časov - markovský čas.

Definícia 11. *Nech $(\mathcal{F}_t, t \in T)$ je filtrácia a $\tau : \Omega \rightarrow T \cup \{\infty\}$. Náhodnú veličinu τ nazveme markovský čas (stopping time), ak $[\tau \leq t] \in \mathcal{F}_t$ pre všetky $t \in T$.*

Už vieme, že σ -algebra \mathcal{F}_t obsahuje všetky nahromadené informácie do času t . Z definície markovského času teda vyplýva, že o jave $[\tau \leq t]$ vieme v čase t povedať, či nastal alebo nie. Markovský čas τ je teda náhodná veličina, ktorá vyjadruje čas nastania nejakej udalosti (počet zákazníkov v obchode prekročí určitý počet, cena akcie dosiahne určitú hladinu, atď’).

Ďalej definujeme čas τ_m , kedy Brownov pohyb W prvýkrát dosiahne hladinu m (*first passage time to level m*) ako

$$\tau_m = \min \{t \geq 0, W_t = m\}.$$

Ak Brownov pohyb nikdy nedosiahne hladinu m , položíme $\tau_m = \infty$.

Tvrdenie 2. Čas prvého dosiahnutia hladiny m , vyššie definovaný τ_m , je markovský čas.

Dôkaz. Máme ukázať, že jav $[\tau_m \leq t]$ je prvkom σ -algebry \mathcal{F}_t . Definujme náhodnú veličinu

$$M_t := \max_{0 \leq s \leq t} W_s.$$

Potom jav $[\tau_m \leq t]$ je totožný s javom $[M_t \geq m]$. Ukážeme, že $[M_t \geq m] \in \mathcal{F}_t$. Zo spojitosti Brownovho pohybu plynie

$$[M_t \geq m] = \bigcup_{s \in Q, 0 \leq s \leq t} [W_s > m].$$

Vieme, že každý z javov $[W_s \geq m]$ v spočetnom zjednotení (množina Q je spočetná) patrí do \mathcal{F}_t . Preto aj spočetné zjednotenie patrí do \mathcal{F}_t . \square

Veličinu definovanú v dôkaze

$$M_t := \max_{0 \leq s \leq t} W_s.$$

nazveme maximum Brownovho pohybu do času t .

Ešte dodajme, že každému markovskému času τ priradíme σ -algebru \mathcal{F}_τ , ktorú definujeme ako množinu všetkých udalostí B takých, že $B \cap [\tau \geq t] \in \mathcal{F}_t$, t.j. \mathcal{F}_τ obsahuje všetky možné udalosti do času τ .

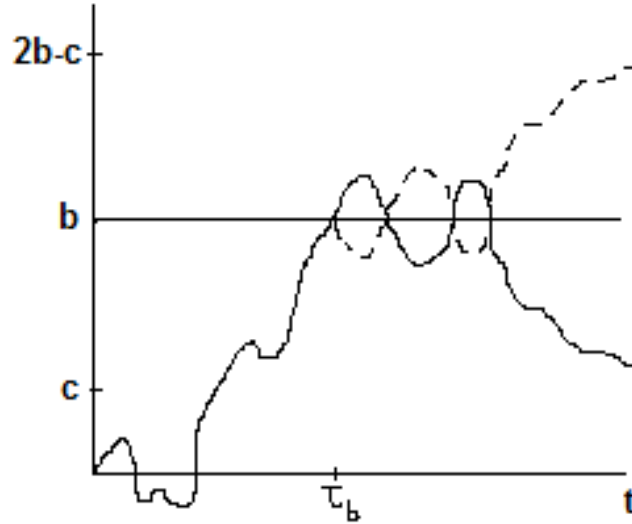
Veta 5 (Silná markovská vlastnosť). *Nech (Ω, \mathcal{F}, P) je pravdepodobnostný priestor s filtráciou $(\mathcal{F}_t, t \geq 0)$ a na ňom definovaný štandardný Brownov pohyb $W = (W_t, t \geq 0)$. Nech τ je markovský čas definovaný pomocou filtrácie $(\mathcal{F}_t, t \geq 0)$. Nech σ -algebra prislúcha k τ , potom pre $t \geq 0$ definujeme proces*

$$W_t^* = W_{t+\tau} - W_\tau,$$

nech $(\mathcal{F}_t^*, t \geq 0)$ je filtrácia procesu $(W_t^*, t \geq 0)$. Potom

- $(W_t^*, t \geq 0)$ je štandardný Brownov pohyb,
- σ -algebry \mathcal{F}_τ a \mathcal{F}_t^* sú pre všetky $t > 0$ nezávislé.

Silná markovská vlastnosť teda hovorí, že Brownov pohyb začne v čase τ „odznovu“, nezáleží na tom, ako sa vyvíjal do času τ , t.j. dá sa povedať, že „nemá pamäť“.



Obrázek 1.1: Princíp reflexie Brownovho pohybu

1.2.4 Princíp reflexie

Na tomto mieste si najprv intuitívne predstavíme princíp reflexie (zrkadlenia), ako ho uvádzajú autori v [6] str. 100-101. Uvažujme Brownov pohyb W a nejakú hladinu b . Zrejme platí

$$P[\tau_b < t] = P[\tau_b < t, W_t > b] + P[\tau_b < t, W_t < b] \quad (1.2)$$

Pretože javy $[\tau_b < t, W_t > b]$ a $[W_t > b]$ sú totožné, z prvého člena na pravej strane automaticky vyplýva $P[\tau_b < t, W_t > b] = P[W_t > b]$. V druhom člene máme situáciu $\tau_b < t$ a $W_t < b$, teda v nejakom čase pred t už dosiahol bod b a zvyšok času „šiel“ do bodu c , kde $c \leq b$. Zo symetrie Brownovho pohybu so začiatkom v bode b , je rovnako „pravdepodobné“, že v čase t by sa Brownov pohyb nachádzal v bode $2b - c$. Intuitívne sme dospeli k záveru, že každá trajektória Brownovho pohybu má rovnakú „pravdepodobnosť“ ako jej zrkadlový obraz (obrázok 1.1, čiarkovaná krivka). Slová pravdepodobnosť sú v úvodzovkách preto, lebo majú len intuitívny význam. Je zrejmé, že každá trajektória Brownovho pohybu nastáva s nulovou pravdepodobnosťou. Teda máme

$$P[\tau_b < t, W_t < b] = P[\tau_b < t, W_t > b] = P[W_t > b].$$

Z vyššie uvedených vzťahov plynie na záver konečný vzťah

$$P[\tau_b < t] = 2P[W_t > b].$$

To znamená, že ak je $\tau_b < t$ tak je rovnako pravdepodobné, že W_t bude nad hladinou b alebo pod hladinou b . Toto odvodenie má však len motivačný charakter. Naša argumentácia sa opiera o skutočnosť, že Brownov pohyb má silnú markovskú vlastnosť, inými slovami, že Brownov pohyb W začal v čase τ_b odznovu, resp. nezávisí na vývoji pred časom τ_b .

Tvrdenie 3. *Nech W je štandardný Brownov pohyb, $m > 0$ a τ_b čas prvého dosiahnutia b . Potom platí*

$$P[\tau_b < t] = 2P[W_t > b].$$

Dôkaz. Vieme, že $W_{\tau_b} = b$. Potom

$$P[\tau_b < t, W_t < b] = P[\tau_b < t, W_{\tau_b+(t-\tau_b)} - W_{\tau_b} < 0],$$

Položme $s = (t - \tau_b)$. Vieme už, že τ_b je markovský čas, potom zo silnej markovskej vlastnosti Brownovho pohybu vyplýva, že $W_{\tau_b+s} - W_{\tau_b}$ je Brownov pohyb nezávislý na \mathcal{F}_{τ_b} . O prvom člene vo vzťahu (1.2) sme už pojednali v texte pred dôkazom. Vieme, že ak $[W_t < b]$, tak $\tau_b < t$ a preto náhodná veličina $W_t - W_{\tau_b}$ má normálne rozdelenie so strednou hodnotou 0. Teda zo symetrie normálneho rozdelenia plynie, že

$$P[W_t - W_{\tau_b} < 0 | \tau_b < t] = P[W_t - W_{\tau_b} > 0 | \tau_b < t] = \frac{1}{2}.$$

Z definície podmienkej strednej hodnoty dostaneme

$$P[\tau_b < t, W_t < b] = P[W_t - W_{\tau_b} < 0 | \tau_b < t] P[\tau_b < t] = \frac{1}{2} P[\tau_b < t].$$

Dosadením do vzťahu (1.2) dostaneme

$$P[\tau_b < t] = P[W_t > b] + \frac{1}{2} P[\tau_b < t],$$

z čoho ihneď plynie naše tvrdenie. □

Práve dokázané tvrdenie nám ukazuje význam markovského času. Ak by sme τ_b nahradili náhodným časom μ , nie markovským, tak tvrdenie stroskotá na tom, že $W_{\mu+s} - W_{\mu}$ nemusí byť Brownov pohyb nezávislý na \mathcal{F}_{μ} . Takýmto príkladom môže byť napríklad čas prvého dosiahnutia maxima na nejakom intervale. Jednoducho povedané informácia, ktorú máme do času t , nestačí na to, aby sme vedeli povedať, či bolo dosiahnuté maximum. Potrebujeme vedieť aj, čo sa bude „diať“ po čase t .

Dôsledok 3. Pre $b \neq 0$ je $P[\tau_b < \infty] = 1$.

Dôkaz. Zo spojitosti v t platí, že $P[\tau_b < t] = P[\tau_b \leq t]$. Už vieme, že javy $[M_t \geq b]$ a $[\tau_b \leq t]$ sú totožné. Potom

$$P[M_t \geq b] = 2P[W_t > b] = 2P\left[\frac{W_t}{\sqrt{t}} > \frac{b}{\sqrt{t}}\right] = 2\left(1 - \Phi\left(\frac{b}{\sqrt{t}}\right)\right). \quad (1.3)$$

Ak $t \rightarrow \infty$, tak hodnota distribučnej funkcie bude konvergovať k 1/2, preto

$$\lim_{t \rightarrow \infty} P[M_t \geq b] = 1.$$

Jav $[\tau_b < \infty]$ je zjednotením javov, cez všetky t , že maximum do času t je väčšie ako b . To je rastúca postupnosť javov (každá množina je nadmnožinou tej predošlej). Keďže $[M_t > b]$ je rastúca postupnosť javov, tak stačí spraviť zjednotenie cez prirodzené čísla, lebo pre každé t existuje prirodzené číslo n také, že $n > t$ a tým padom $[M_n > b]$ je nadmnožinou $[M_t > b]$. Preto môžeme písať

$$P[\tau_b < \infty] = P\left(\bigcup_{t>0} [M_t > b]\right) = P\left(\bigcup_{n \in \mathbb{N}} [M_n > b]\right)$$

a prejsť k limite

$$P\left(\bigcup_{n \in \mathbb{N}} [M_n > b]\right) = \lim_{n \rightarrow \infty} P[M_n > b] = 1,$$

čím sme dokázali naše tvrdenie. \square

Toto tvrdenie hovorí, že Brownov pohyb dosiahne akúkoľvek hladinu v konečnom čase, resp. maximum M_t do času t prekročí ľubovoľnú hranicu, ak $t \rightarrow \infty$.

Takisto môžeme zo vzťahu (1.3) vypočítať hustotu markovského času τ_b :

$$f_{\tau_b}(t) = \frac{\partial}{\partial t} P[\tau_b \leq t] = 2 \frac{\partial}{\partial t} \Phi\left(\frac{b}{\sqrt{t}}\right) = \frac{|b|}{t\sqrt{2\pi t}} e^{-\frac{b^2}{2t}}, \quad t \geq 0.$$

Nie je ťažké ukázať, že sa princíp reflexie dá vyjadriť aj v tvare

$$P[M_t \geq b, W_t \leq c] = P[W_t \geq 2b - c] \quad c \leq b, b > 0.$$

Na záver uvedieme základný dôsledok princípu reflexie Brownovho pohybu W . Proces W^b definovaný ako $W_t^b = W_t$ pre $t < \tau_b$ a $W_t^b = 2a - W_t$ pre $t \geq \tau_b$ je Brownov pohyb.

1.2.5 Vlastnosť self-similarity

Definícia 12. *Nech $H > 0$. Reálny náhodný proces $(X_t, t \in T)$ nazývame H -self-similar, ak platí*

$$\mathcal{L}(c^H X_{t_1}, c^H X_{t_2}, \dots, c^H X_{t_n}) = \mathcal{L}(W_{ct_1}, W_{ct_2}, \dots, W_{ct_n})$$

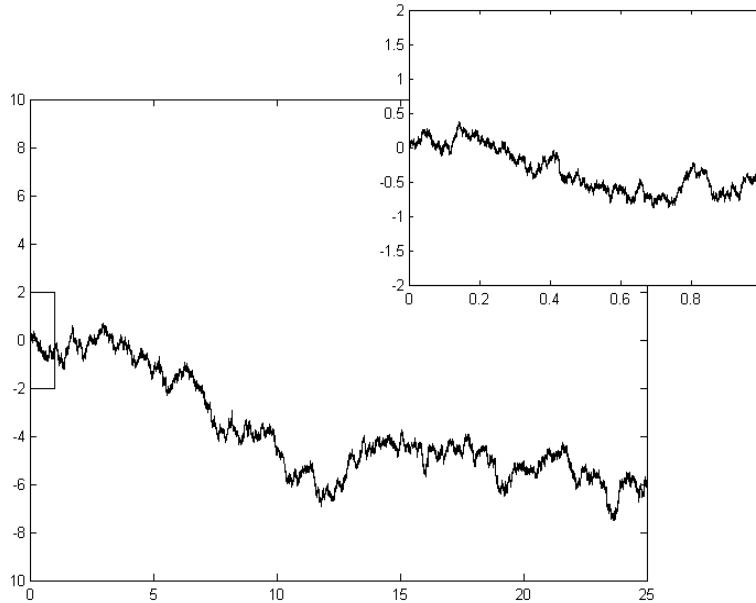
pre všetky $c > 0$, $n \in \mathbb{N}$ a $t_i \geq 0$, kde $i = 1, \dots, n$.

Symbol \mathcal{L} značí konečne rozmerné rozdelenie a rovnosť v definícii je rovnosťou rozdelení náhodných procesov. Číslu H sa hovorí Hurstov parameter. Ako už názov napovedá, táto trieda procesov vykazuje známky akejsi „sebepodobnosti“. My ukážeme, že aj Brownov pohyb je self-similar proces.

Tvrdenie 4. *Brownov pohyb $W = (W_t, t \geq 0)$ je $\frac{1}{2}$ -self-similar proces.*

Dôkaz. Z vety 1 vieme, že pre každé $c > 0$ je proces $(cW_{t/c^2}, t \geq 0)$ Brownov pohyb. Teda aj proces $(c^{-1/2}W_{ct}, t \geq 0)$ je Brownov pohyb. Vieme teda, že veličiny $c^{-1/2}W_{ct}$ sú pre každé $c > 0$ a $t \geq 0$ normálne rozdelené so strednou hodnotou 0, a ako sa dá rýchlo ukázať, s rozptylom t . Z toho plynie, že W je $\frac{1}{2}$ -self-similar proces. \square

Táto vlastnosť Brownovho pohybu hovorí, že ak budeme Brownov pohyb sledovať v akýchkoľvek malých časových intervaloch, jeho štruktúra (tvar trajektórie) bude rovnaká. Vid'. obrázok 1.2.



Obrázek 1.2: Vlastnosť self-similarity Brownovho pohybu

1.2.6 Lévyho modul spojitosti

Veta 6 (Zákon iterovaného logaritmu (A. Hinčin (1933))). *Nech W_t je Brownov pohyb, potom pre skoro všetky $\omega \in \Omega$ je*

$$\limsup_{t \rightarrow 0} \frac{|W_t(\omega)|}{\sqrt{2t \log \log(1/t)}} = 1.$$

Definícia 13. *Funkcia g sa nazýva modul spojitosti pre funkciu $f : [0, T] \rightarrow \mathbb{R}$, ak $0 \leq s < t \leq T$ a $|t - s| \leq \delta$ implikuje $|f(t) - f(s)| \leq g(\delta)$ pre dostatočne malé $\delta > 0$.*

Vďaka zákonu iterovaného logaritmu je každý modul spojitosti pre Brownov pohyb W_t na uzavretom intervale $[0, 1]$ najmenej $\sqrt{2\delta \log \log(1/\delta)}$ a najviac $c\delta^\gamma$, $\gamma > 0$ a c je konštanta. Lévy však zaviedol presný modul spojitosti.

Veta 7 (Lévyho modul spojitosti). *Nech $g : (0, 1] \rightarrow (0, \infty)$ a $g(\delta) = \sqrt{2\delta \log(1/\delta)}$, potom*

$$P \left[\limsup_{\delta \rightarrow 0} \frac{1}{g(\delta)} \max_{\substack{0 \leq s < t \leq 1 \\ t-s \leq \delta}} |W_t - W_s| = 1 \right] = 1.$$

Obe vety a definícia sú uvedené v [5] aj s dôkazmi. Ak predpokladáme, že súbor dát sa riadi Brownovým pohybom, tak Lévyho modul spojitosti je užitočný nástroj na tzv. odfiltrovanie dát od skokov.

1.3 Geometrický Brownov pohyb

Pri modelovaní ceny akcie musíme počítať s trenodovým vývojom procesu (driftom) a nejakým faktorom neistoty (volatilitou). Častokrát preto pracujeme s procesom

$$Y_t = \mu t + \sigma W_t, \quad t > 0.$$

Takto definovaný náhodný proces nazveme Brownov pohyb s driftom μ a volatilitou σ . Ako sme ukázali v predchádzajúcej kapitole, platí, že

$$P[W_t > 0] = P[W_t < 0] = \frac{1}{2}.$$

To znamená, že ak by sme chceli modelovať cenu akcie, mohla by byť aj záporná, čo na finančnom trhu možné nie je. Preto potrebujeme zaviesť transformáciu Brownovho pohybu, ktorá bude všade kladná.

Definícia 14. *Nech μ a $\sigma > 0$ sú konštanty a nech W_t je Brownov pohyb. Potom*

$$S_t = S_0 \exp \left\{ \sigma W_t + \left(\mu - \frac{1}{2} \sigma^2 \right) t \right\}$$

nazveme geometrický Brownov pohyb. Číslo μ nazývame drift a číslo σ nazývame variancia geometrického Brownovho pohybu.

Geometrický Brownov pohyb S_t je lognormálne rozdelený s parametrami

$$E S_t = S_0 e^{\mu t}, \quad \text{var } S_t = S_0^2 e^{2\mu t} \left(e^{\sigma^2} - 1 \right).$$

Nech S_0, \dots, S_n sú namerané ceny podkladového aktíva v časoch t_0, t_1, \dots, t_n . Predpokladajme, že tieto ceny sa riadia geometrickým Brownovým pohybom $S_t = S_0 \exp\{\sigma W_t + (\mu - \frac{1}{2}\sigma^2)t\}$. Potom definujme náhodné veličiny R_t tzv. *log returns*, ako

$$R_t = \log(S_{t+1}) - \log(S_t) = \log \left(\frac{S_{t+1}}{S_t} \right).$$

Z tohto modelu vyplýva, že

$$\log \left(\frac{S_{t_{i+1}}}{S_{t_i}} \right) = \sigma (W_{t_{i+1}} - W_{t_i}) + \left(\mu - \frac{1}{2} \sigma^2 \right) (t_{i+1} - t_i),$$

Pri vysokofrekvenčných dátach sa však driftová zložka vďaka Lévyho modulu spojitosti ($\mu = \frac{\sigma^2}{2}$) zanedbáva. Potom vidíme, že náhodné veličiny R_{t_i} a $\sigma(W_{t_{i+1}} - W_{t_i})$ majú rovnaké konečne dimenzionálne rozdelenie, t.j.

$$\mathcal{L}(R_{t_i}) = \mathcal{L}(\sigma[W_{t_{i+1}} - W_{t_i}]).$$

Navyše sú to pre každé i nezávislé, rovnako rozdelené náhodné veličiny s normálnym rozdelením $N(0, \sigma^2(t_{i+1} - t_i))$. Toto tvrdenie vyplýva zo skutočnosti, že prírastky Brownovho pohybu $W(t)$ majú normálne rozdelenie $N(0, t_{i+1} - t_i)$. Vďaka tomuto faktoru sa často namiesto skutočných cien skúmajú práve ich logaritmy, teda log returns.

Zaved'me tzv. *exponenciálny martingal* vzhľadom k σ ako

$$Z_t = \exp \left\{ \sigma W_t - \frac{1}{2} \sigma^2 t \right\},$$

kde $W = (W_t, t \geq 0)$ je Brownov pohyb a σ je konštanta.

Tvrdenie 5 (Exponenciálny martingal). *Nech $W = (W_t, t \geq 0)$ je Brownov pohyb s filtráciou $(\mathcal{F}_t, t \in T)$. Potom proces $Z = (Z_t, t \geq 0)$, definovaný vyššie, je martingal.*

Dôkaz. Pre $0 \leq s \leq t$ máme

$$\begin{aligned} \mathbb{E} [Z_t | \mathcal{F}_s] &= \mathbb{E} \left[\exp \left\{ \sigma W_t - \frac{1}{2} \sigma^2 t \right\} \middle| \mathcal{F}_s \right] \\ &= \mathbb{E} \left[\exp \sigma(W_t - W_s) \cdot \exp \left\{ \sigma W_s - \frac{1}{2} \sigma^2 t \right\} \middle| \mathcal{F}_s \right] \\ &= \exp \left\{ \sigma W_s - \frac{1}{2} \sigma^2 t \right\} \cdot \mathbb{E} [\exp \{ \sigma(W_t - W_s) \} | \mathcal{F}_s], \end{aligned}$$

príčom sme z podmienenej strednej hodnoty vyňali \mathcal{F}_s -merateľnú náhodnú veličinu (vlastnosť podmienenej strednej hodnoty). Z definície Brownovho pohybu vieme, že náhodná veličina $\exp \{ \sigma(W_t - W_s) \}$ je nezávislá na σ -algebre \mathcal{F}_s a využijeme ďalšiu vlastnosť podmienenej strednej hodnoty

$$\mathbb{E} [\exp \{ \sigma(W_t - W_s) \} | \mathcal{F}_s] = \mathbb{E} [\exp \{ \sigma(W_t - W_s) \}].$$

Vieme, že náhodná veličina $W_t - W_s$ má normálne rozdelenie so strednou hodnotou 0 a rozptylom $t - s$, preto ľahko vypočítame strednú hodnotu na pravej strane. Pre prehľadnosť označíme $Y = W_t - W_s$ a $c = t - s$.

$$\begin{aligned} \mathbb{E} [e^{\sigma(W_t - W_s)}] &= \int e^{\sigma x} \frac{1}{\sqrt{2\pi c}} e^{-\frac{x^2}{2c}} dx \\ &= e^{\frac{\sigma^2 c}{2}} \int \frac{1}{\sqrt{2\pi c}} e^{-\frac{(x - \sigma c)^2}{2c}} dx = e^{\frac{\sigma^2 c}{2}}, \end{aligned}$$

posledná rovnosť plynie z toho, že argument integrálu predstavuje hustotu normálneho rozdelenia $N(\sigma c, c^2)$. Teda na záver dostávame podmienku z definície martingalu

$$\mathbb{E} [Z_t | \mathcal{F}_s] = \exp \left\{ \sigma W_s - \frac{1}{2} \sigma^2 s \right\} = Z_s.$$

□

Dôsledok tohto tvrdenia je, že na výpočet strednej hodnoty nám stačí znalosť hodnoty v počiatkovom čase.

$$\mathbb{E} S_t e^{-\mu t} = \mathbb{E} \left[\mathbb{E} \left[S_t e^{-\mu t} \middle| \mathcal{F}_{s=0} \right] \right] = \mathbb{E} [S_0 e^{-\mu \cdot 0}] = S_0.$$

2. Testy

Hypotézu, že ceny nášho podkladového aktíva sa dajú modelovať pomocou Brownovho pohybu, musíme overiť. Preto by sme mali overiť, že pre všetky časové úseky $0 = t_0 < t_1 < \dots < t_m$ sú prírastky $W_{t_1} = W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_m} - W_{t_{m-1}}$ nezávislé, stacionárne a majú normálne rozdelenie $N(0, t_{i+1} - t_i)$. Pojem stacionarity bude uvedený nižšie. V ďalšom texte uvediem postupne základné princípy fungovania testov:

- normality,
- stacionarity,
- nezávislosti.

2.1 Normalita

Už názov napovedá, že normálne rozdelenie je jedno z najbežnejších rozdelení v štatistike, ale aj v prírode. Normálne rozdelenie so strednou hodnotou μ a rozptylom $\sigma^2 > 0$ je definované hustotou

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}.$$

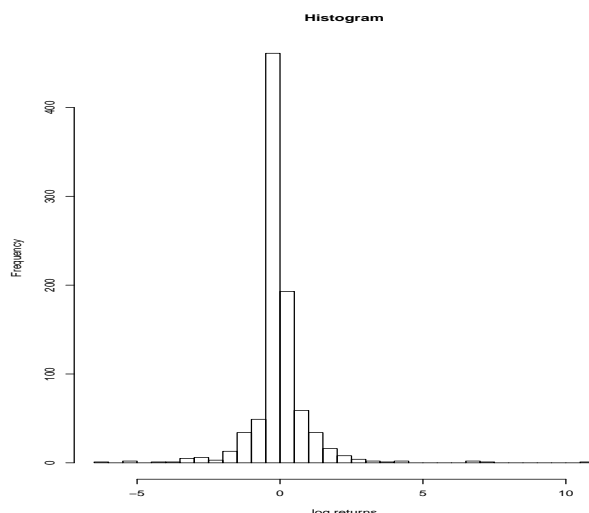
Častokrát štatistickí normalitu skumaných dát mlčky predpokladajú kvôli jej výhodným vlastnostiam. Ak sa ukáže opak, že dáta sa neriadia priamo normálnym rozdelením, snažíme sa dáta nejakým spôsobom transformovať (odfiltrovanie, log returns a iné transformácie) tak, aby tieto transformácie už pochádzali z normálneho rozdelenia. Spôsobov overovania, testovania normality je veľmi veľa, preto je dôležité pristupovať ku každej vzorke dát zvlášť. Niektoré metódy sú vhodnejšie pre rozsiahlejšie výbery, niektoré testujú „zhodu“ rozdelení, niektoré skúmajú šikmost', atď. Avšak, ako sa ukázalo, nie všetky metódy sú rovnako účinné, niektoré môžu dokonca viesť k chybným záverom.

2.1.1 Grafická analýza dát

Pri štatistickej analýze dát je dôležité mať o dátach nejakú vizuálnu predstavu. Na vytvorenie tejto predstavy nám poslúžia grafy, ktoré nám taktiež dajú prvotnú predstavu o tvare hustoty, o jej symetrii, resp. nesymetrii, o tom, kde je najviac koncentrovaná, takisto nám veľa napovedia o šikmosti a špicatosti rozdelenia, z ktorej dáta pochádzajú. Samotné testy normality sú veľmi dobrým nástrojom pre formálne zamietnutie hypotézy o normalite na nejakej hladine spoľahlivosti α , avšak neposkytnú nám žiadnu informáciu o dôvode zamietnutia.

Histogram

Histogram je azda najbežnejší typ grafu na modelovanie hustoty rozdelenia, z ktorého dáta pochádzajú. Dáva nám prvý intuitívny náhľad na rozloženie dát, na symetriu a šikmost'. Konštruuje sa spôsobom, že číselnú os si rozdelíme na



Obrázek 2.1: Histogram zostavený z 30 sekundového výberu log returns

rovnako široké intervaly a nad každým z intervalov vykreslíme stĺpec, pričom výška stĺpca je daná početnosťou skúmaných dát v danom intervale. Je dôležité voliť správnu šírku intervalu, lebo inak sa môže výpovedná hodnota histogramu stratit'. Z histogramu 2.1 vieme vyčítať, že hlavná časť dát je koncentrovaná okolo nuly. Graf je symetrický až na druhý najvyšší stĺpec, práve ten ovplyvňuje najviac šikmost', ktorá bude tým pádom kladná (výberová šikmost' vyšla 1.87688). Rozdelenie, ktorého šikmost' $\alpha_3 = E(X - EX)^3 / \sigma^3$ je rôzna od nuly, sa volá šikmé. Normálne rozdelenie má šikmost' 0.

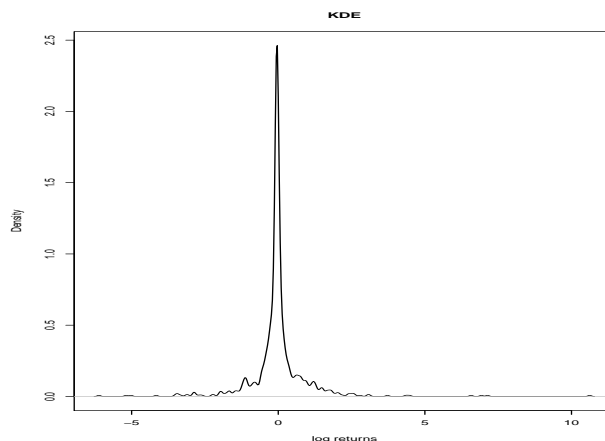
Odhad hustoty pomocou jadra

Nevýhodou histogramov je, že nie sú hladké a závisia na dĺžke intervalu a jeho koncových bodoch. Práve pre tieto tri nevýhody je niekedy výhodnejšie použiť *odhad hustoty pomocou jadra, resp. kernelový odhad*. Tento postup odhadu hustoty sa radí medzi neparametrické odhady hustoty. Je založený na myšlienke, že do každého nášho pozorovaného bodu je umiestnená symetrická funkcia, tzv. jadro K . Tým sa vyhneme problému s fixovanými hranicami intervalov. Potom je odhad našej hustoty

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

kde $\int K(t)dt = 1$ a h je parameter.

Jadro K musí byť symetrická funkcia. Najčastejšie používané sú jadrá: trojuhelníkové, rovnomerné a gaussovské (normálne). Aj my budeme používať práve gaussovské jadro, ktoré má tvar $K(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$. Príklad je uvedený na obrázku 2.2, pre rovnaké dáta ako pre histogram. Problém správneho zvolenia parametru h je zložitejší, zaoberá sa ním napríklad T. Duong v jeho prácach.



Obrázek 2.2: Odhad hustoty pomocou jadra

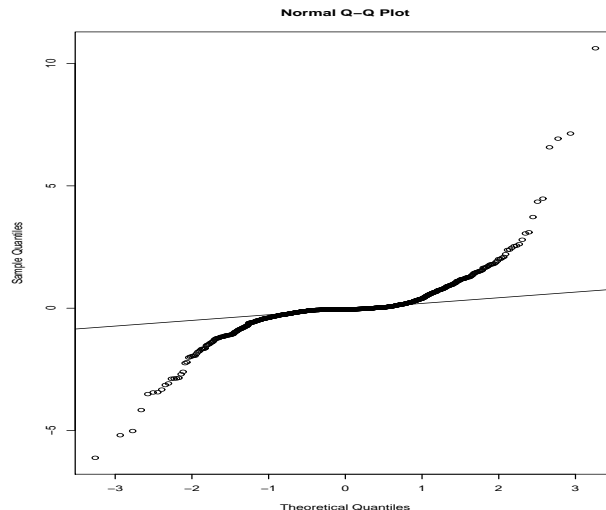
Kvantil-kvantilový graf

Ako nám už názov napovedá, kvantil-kvantilový graf (Q-Q graf) je graf, ktorý porovnáva dve rozdelenia pomocou ich kvantilov. Q-Q graf patrí medzi tzv. pravdepodobnostné grafy, ktoré porovnávajú namerané hodnoty s teoretickými hodnotami nejakého rozdelenia. Jeho konštrukcia spočíva v tom, že si vytvoríme poradovú štatistiku z našich dát $x_{(1)}, \dots, x_{(n)}$, teda ich usporiadame od najmenšieho prvku po najväčší. Toto sú tzv. empirické kvantily, určujú y -ové súradnice. K nim ešte potrebujeme pridať x -ovú zložku. Preto si najprv musíme zaviesť pojem kvantilu.

Nech je distribučná funkcia F spojitá a rýdzo monotónna a nech $\beta \in (0, 1)$, tak číslo $F^{-1}(\beta)$ sa nazýva β -kvantil tohto rozdelenia. Náš prípad sa týka distribučnej funkcie normálneho rozdelenia $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\{-\frac{t^2}{2}\} dt$. Na prvý pohľad je najprirodzenejšie vziať za x -ovú zložku kvantily normálneho rozdelenia v bodoch $p_i = i/n$, kde $i = 1, \dots, n$. Avšak pre $n/n = 1$ sa kvantil blíži k nekonečnu. Aby sme sa tomu vyhli, Kimball (1960) navrhuje počítat kvantily v bodoch $p_i = (i - 0.5)/n$ alebo $i/(n + 1)$. Potom Q-Q graf zobrazuje dvojice $(\Phi^{-1}(p_i), x_{(i)})$.

Ak je teoretické rozdelenie podobné tomu empirickému, tak graf by mal ležať približne na priamke $y = x$. Ak je medzi rozdeleniami nejaká lineárna závislosť, tak graf bude ležať na priamke, nie nutne na $y = x$. Ak pozorujeme na grafe nejakú systematickú odchýlku od priamky, tak s najväčšou pravdepodobnosťou sa empirické rozdelenie nebude zhodovať s tým teoretickým.

Ďalej vieme z Q-Q grafu pojednať aj o „chvostoch“ rozdelenia. Rozdelenie môže mať ťažké (dlhé) alebo ľahké (krátke) chvosty. Táto vlastnosť je daná špicatnosťou grafu. Špicatnosť $\alpha_4 = E(X - EX)^4/\sigma^4$ normálneho rozdelenia je 3. Ak má rozdelenie špicatnosť väčšiu ako 3, tak povieme, že toto rozdelenie má ťažké chvosty, v opačnom prípade $\alpha_4 < 3$, má ľahké chvosty. Samozrejme, že ak má nejaké rozdelenie $\alpha_4 = 3$, neznamená to v žiadnom prípade, že toto rozdelenie je normálne. V prípade ľahkých chvostov má Q-Q graf tvar písmena S. V prípade ťažkých chvostov má rozdelenie tvar ako na obrázku 2.3, ktorý znázorňuje Q-Q graf pre naše dáta.



Obrázek 2.3: Kvantil - kvantilový graf

2.1.2 Testy normality

Grafy nám dávajú len určitú indíciu o distribúcii, preto v štatistike potrebujeme formálnejšie a objektívnejšie metódy na overovanie normality. Toto nám ponúkajú testy normality. Je ich veľmi veľa a delia sa do viacerých kategórií, napr. testy založené na regresii, momentové testy, testy dobrej zhody atď.

Shapiro-Wilkov test normality

Už konštrukcia Q-Q grafu nám kladie otázku, či je medzi hodnotami x_i a $\Phi^{-1}(p_i)$ lineárna závislosť. Shapiro-Wilkov test patrí medzi regresné testy, ktoré sú tvaru $S < k$, kde S je štatistika a k je vhodná konštanta. V podstate je založený na porovnaní odhadu smerodatnej odchýlky našich dát oproti teoretickému odhadu smerodatnej odchýlky normálneho rozdelenia. Ak naše dáta x_i pochádzajú z normálneho rozdelenia $N(\mu, \sigma^2)$, tak potom

$$x_i = \mu + \sigma z_i,$$

kde z_i pochádzajú z normovaného normálneho rozdelenia. Thode (vid'. [4]) uvádza nasledovný postup konštrukcie štatistiky W . Po utvorení poradovej štatistiky $x_{(1)}, \dots, x_{(n)}$ utvoríme vektor stredných hodnôt \mathbf{w} , kde $w_i = E(x_{(i)})$. Ďalej nech V je variančná matica poradovej štatistiky $x_{(1)}, \dots, x_{(n)}$, tak potom najlepší nes-tranný lineárny odhad σ určený metódou najmenších štvorcov je

$$b = a'x,$$

kde $a = (w'V^{-1})/(w'V^{-1}V^{-1}w)^{1/2}$ a $a'a = 1$. Potom je štatistika W definovaná takto:

$$W = \frac{b^2}{(n-1)s^2},$$

kde s^2 je výberový rozptyl. Častejší tvar štatistiky W sa uvádza

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{(x_{(i)} - \bar{x})^2},$$

kde \bar{x} je výbový priemer dát. Problémom tejto štatistiky však je, že prvky V sú presne známe len do rozsahu 20, pre rozsah 21 až 50 odhadli prvky a_i Shapiro a Wilk (1965). Royston (1982a, 1982c) zas ponúka transformáciu W pre rozsah výberu $7 < n < 2000$

$$z = ((1 - W)^\lambda - \mu_y) / \sigma_y.$$

Momentové testy

Majme náhodný výber x_1, \dots, x_n . Potom definujeme k -ty výbový moment

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k.$$

Najpoužívanejšie momentové testy su odvodené od štatistiky

$$g_k = \frac{m_k}{m_{k/2}^{k/2}}.$$

Ak k zvolíme nepárne, tak test nazveme nepárny momentový test. Tieto testy merajú nesymetriu rozdelenia, najbežnejšie volíme $k = 3$ a veličinu $a_3 := g_3$ nazveme výbová šikmosť. Pre najpoužívanejší párný momentový test sa volí $k = 4$, meria symetrické ochýlky od normálneho rozdelenia. Veličinu $a_4 := g_4$ nazývame výbová špicatosť. Ak je x_1, \dots, x_n výber z normálneho rozdelenia, potom

$$\begin{aligned} \text{E } a_3 &= 0, & \text{E } a_4 &= 3 - \frac{6}{n+1}, \\ \text{var } a_3 &= \frac{6(n-2)}{(n+1)(n+3)}, & \text{var } a_4 &= \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}. \end{aligned}$$

Anděl, vid'. [1] str. 275, uvádza, že a_3 a a_4 majú asymptoticky normálne rozdelenie.

Test šikmosti. Test proti alternatíve, že výber pochádza z nejakého nesymetrického rozdelenia, je založený na a_3 . Pre veľké n využijeme asymptotickú normalitu. Vypočítame veličinu

$$U_3 = \frac{a_3}{\sqrt{\text{var } a_3}}.$$

Ako Anděl ďalej uvádza, ak $|U_3| \geq u(\frac{\alpha}{2})$, kde $u(\cdot)$ je kvantil normálneho rozdelenia, tak zamietame hypotézu, že sa jedná o výber z normálneho rozdelenia.

Test špicatosti je založený na štatistike a_4 . Pre malé hodnoty n nájdeme zase kritické hodnoty v tabuľkách. Pre veľké hodnoty n sa zase využije asymptotická normalita. Vypočíta sa veličina

$$U_4 = \frac{a_4 - \text{E } a_4}{\sqrt{\text{var } a_4}}$$

a hypotézu normality zase zamietneme ak $|U_4| \geq u(\frac{\alpha}{2})$. D'Agostino (1990) uvádza ďalšie vylepšenia tohto postupu.

Andersonov - Darlingov test normality

Ďalej budeme používať na testovanie normality našich dát Andersonov - Darlingov test. Patrí medzi tzv. testy dobrej zhody. Je založený na štatistike

$$A = -n - \frac{1}{n} \sum_{i=1}^n [2i - 1][\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})],$$

kde $p_{(i)} = \Phi(\frac{x_{(i)} - \bar{x}}{s})$, Φ je distribučná funkcia normovaného normálneho rozdelenia $N(0, 1)$, \bar{x} je výberový priemer a s smerodatná odchýlka našich dát. Často sa však používa modifikovaná štatistika $Z = A(1 + 0.75/n + 2, 25/n^2)$.

2.2 Stacionarita

Definícia Brovnovho pohybu hovorí, že jednotlivé prírastky majú byť stacionárne. Vo všeobecnosti existujú dva druhy stacionarity. V analýze časových radov je $(r_t, t \geq 0)$ časový rad *silno stacionárny*, ak združené rozdelenie $(r_{t_1}, \dots, r_{t_k})$ má totožné rozdelenie ako $(r_{t_1+t}, \dots, r_{t_k+t})$, pre všetky t a k a t_1, \dots, t_k , t.j. združené rozdelenie $(r_{t_1}, \dots, r_{t_k})$ je v čase nemenné. Časový rad $(r_t, t \geq 0)$ sa nazýva *slabo stacionárny*, ak $E r_t = \mu$ a $\text{cov}(r_t, r_{t-\ell}) = \gamma_\ell$, kde μ je konštanta a ℓ je prirodzené číslo. Inými slovami slabá stacionarita znamená, že je stredná hodnota r_t v čase nemenná a kovariancia medzi r_t a $r_{t-\ell}$ závisí len na ℓ (nie na t), teda $\text{cov}(r_t, r_{t-\ell})$ sa tiež v čase nemení. Číslo ℓ sa nazýva *lag* a γ_ℓ je lag- ℓ autokovariancia r_t .

Z oboch definícií priamo vyplýva, že ak r_t má konečné prvé dva momenty a je silno stacionárny, tak je aj slabo stacionárny. Ďalej ak má r_t normálne rozdelenie, dá sa ukázať, že slabá a silná stacionarita sú ekvivalentné.

My vieme, že pre Brownov pohyb $(W_t, t \geq 0)$ platí, že pre časový prírastok Δ má veličina W_Δ normálne rozdelenie. Inými slovami združené rozdelenie prírastkov má vlastnosť silnej stacionarity. V literatúre sa pri analýze cien finančných derivátov takmer vždy predpokladá ich slabá stacionarita. My sa pokúsime tento predpoklad overiť na prírastkoch našich dát. Na testovanie použijeme najpopulárnejší KPSS test, ktorého bližší teoretický základ popisovať nebudeme.

2.3 Nezávislosť

Mieru lineárnej závislosti medzi dvomi náhodnými veličinami X a Y vyjadruje korelačný koeficient

$$\rho_{x,y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{E(X - EX)(X - EX)}{\sqrt{E(X - EX)^2(Y - EY)^2}}.$$

Dá sa ukázať, že $-1 \leq \rho_{x,y} \leq 1$, pričom ak $\rho_{x,y} = 0$ sú nahodné veličiny nekorelované. Ak sú nahodné veličiny X a Y nezávislé, tak sú taktiež nekorelované,

naopak to ale neplatí. Dá sa dokázať, že ak X a Y majú normálne rozdelenie, tak nekorelovanosť je ekvivalentná s nezávislosťou.

Predpokladajme, že máme náhodný výber $(x_1, y_1), \dots, (x_T, y_T)$, potom výberový korelačný koeficient definujeme

$$\hat{\rho}_{x,y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}},$$

kde \bar{x} a \bar{y} sú výberové priemery x_1, \dots, x_T resp. y_1, \dots, y_T .

Autokorelačná funkcia ACF

Predpokladajme teraz, že máme časový rad $\{r_t\}$, ktorý je slabo stacionárny. Ak chceme skúmať závislosť medzi r_t a $r_{t-\ell}$, vypočítame ich korelačný koeficient, ktorý sa nazýva *lag- ℓ autokorelácia* r_t a často sa značí ρ_ℓ . Pri predpoklade slabej stacionarity je

$$\rho_\ell = \frac{\text{cov}(r_t, r_{t-\ell})}{\sqrt{\text{var}(r_t)\text{var}(r_{t-\ell})}} = \frac{\text{cov}(r_t, r_{t-\ell})}{\text{var}(r_t)} = \frac{\gamma_\ell}{\gamma_0},$$

kde sa využila vlastnosť slabej stacionarity, pre $\ell = 0$ je $\gamma_0 = \text{cov}(r_t, r_{t-0}) = \text{var } r_t$ a $\text{var } r_t = \text{cov}(r_t, r_t) = \text{cov}(r_{t-\ell}, r_{t-\ell}) = \text{var } r_{t-\ell} = \gamma_0$. Potom ρ_1, ρ_2, \dots sa nazýva *autokorelačná funkcia - ACF*. Má tri zrejme vlastnosti: a) $\rho_0 = 1$, b) $\rho_\ell = \rho_{-\ell}$, c) $-1 \leq \rho_\ell \leq 1$. Z toho vyplýva, že slabo stacionárny rad $\{r_t\}$ je sériovo nekorelovaný práve vtedy, keď $\rho_\ell = 0$ pre všetky $\ell > 0$.

Konzistentný odhad ρ_ℓ je definovaný nasledovne ako

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq \ell < T - 1.$$

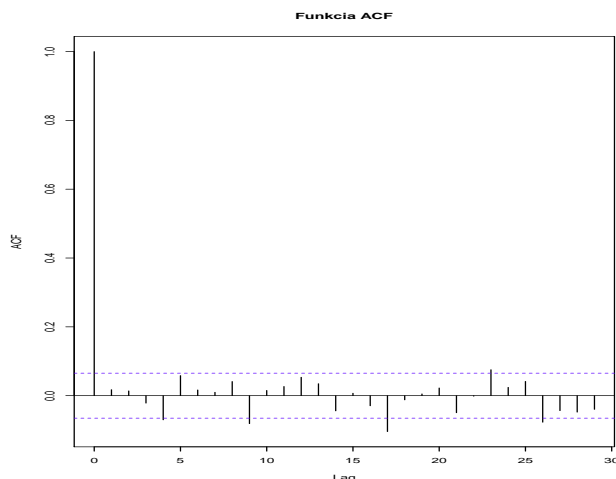
Funkcia $\hat{\rho}_1, \hat{\rho}_2, \dots$ sa potom nazýva výberová autokorelačná funkcia. Jej graf sa často používa na subjektívne vyhodnotenie autokorelácie radu. Graf funkcie ACF je tvorený bodmi (ℓ, ρ_ℓ) (vid'. obrázok 2.4, na obrázku sú spracované rovnaké dáta, ako v kapitole 2.1).

Na obrázku sú okrem hodnôt ρ_ℓ zobrazené aj dve rovnobežné priamky, ktoré zobrazujú 95%-ný interval spoľahlivosti, že rad pre dané ℓ vykazuje signifikantné známky autokorelácie. Z grafu sa dá teda vycítať, že okrem $\ell = 17$, nevykazuje rad žiadne významnejšie známky autokorelácie.

Boxov - Piercov test

Pre formálnejší postup použijeme test nezávislosti v časovom rade, ktorý predstavili Box a Pierce v roku 1970. Zaoberá sa testovaním nulovosti autokorelácií v r_t . Testuje hypotézu $H_0 : \rho_1 = \dots = \rho_m = 0$, proti $H_1 : \rho_i \neq 0$, pre nejaké $i \in \{1, \dots, m\}$. Je založený na štatistike

$$Q_m^* = T \sum_{l=1}^m \hat{\rho}_l^2.$$



Obrázek 2.4: Graf funkcie ACF

Potom Q_m^* má asymptoticky rozdelenie χ_m^2 . Box a Ljung (1978) modifikovali štatistiku Q_m^* , ktorá je silnejšia pre testovanie konečných výberov

$$Q_m = T(T+2) \sum_{\ell=1}^m \frac{\hat{\rho}_\ell^2}{n-\ell}.$$

2.4 Testovanie self-similarity pre Brownov pohyb

Ako sme ukázali v prvej kapitole Brownov pohyb je $\frac{1}{2}$ -self-similar proces. Ukážeme, že je úplne popísaný v čase 1. Podmienku pre self-similarity môžeme napísať v tvare

$$\mathcal{L}(W_{ct}, t \geq 0) = \mathcal{L}(c^H W_t, t \geq 0).$$

Ak zvolíme $c = 1/t$, tak dostaneme

$$\mathcal{L}(t^H W_1) = \mathcal{L}(W_t).$$

Pre distribučnú funkciu potom platí

$$F_t(x) = P[W_t \leq x] = P[t^H W_1 < x] = P[W_1 \leq x/t^H] = F_1(x/t^H).$$

Ak F_t má hustotu ρ_t , tak po derivovaní dostaneme

$$\rho(x) = \frac{\partial}{\partial x} F_t(x) = \frac{\partial}{\partial x} F_1(x/t^H) = \frac{1}{t^H} \rho_1(x/t^H).$$

Teda pre $x = 0$ máme

$$\rho_t(0) = \frac{\rho_1(0)}{t^H} \quad \forall t > 0. \quad (2.1)$$

Testovať self-similarity pomocou odvodených vzťahov budeme nasledovne. Keďže zo stacionarity prírastkov Brownovho pohybu vieme, že $\mathcal{L}(W_{t_\Delta} - W_t) = \mathcal{L}(W_\Delta)$, tak hustotu budeme odhadovať iba z W_Δ . Najprv odhadneme pomocou

kernelového odhadu hustoty $\hat{\rho}_t$ pre rôzne t v nule. Potom zafixujeme jeden z časov Δ . Po zlogaritmovaní rovnice (2.1), dostávame

$$\log \hat{\rho}_t(0) = H \log \frac{\Delta}{t} + \log \hat{\rho}_\Delta(0).$$

Na záver pomocou lineárnej regresie budeme skúmať túto závislosť a odhadneme parameter H pomocou metódy najmenších štvorcov, ktorý by mal výjsť približne 0.5. Za maticu X zvolíme maticu zloženú z dvoch stĺpcov. Prvý bude $\log \frac{\Delta}{t}$ a druhý $\log \hat{\rho}_\Delta(0)$. Potom regresný koeficient pri prvom stĺpci je odhadovaný parameter H .

3. Testovanie štatistických vlastností

3.1 Prvé spracovanie dát

Naše dáta sú uložené v textovom súbore, ktorý obsahuje dva stĺpce: v prvom je časový údaj a v druhom je cena aktíva. My budeme naše testy aplikovať na jednotlivé dni. Testy budeme robiť na 9 náhodne vybraných dňoch. V každý deň sa cena meria každú sekundu od 8.00 hod do 22.00 hod, čo činí 50400 pozorovaní na jeden deň. Skúmané vlastnosti však nebudeme testovať len na celom dni, ale aj na rôznych výberoch. V prvom rade to budú výbery z dlhších časových úsekov, konkrétne: 1, 2, 15 a 30 sekundové výbery. Musíme si pri každom teste dávať pozor, aby rozsah nášho výberu neklesol pod minimálny rozsah výberu daného testu. Ďalej si musíme uvedomiť, že v čase 14.30 hod (európskeho času) sa otvára burza v USA, kedy môže prísť k výkyvom cien. Preto budeme testovať aj výbery v časoch od 8.00 hod do 14.30 hod a od 14.30 hod do 22.00 hod. Preto si v programe R najprv načítame ceny po dňoch.

```
> pocet_dni <- 9
> den <- matrix(nrow = pocet_dni, ncol = 50400)
> den[1, ] <- fp[6634231:6684630]
> den[2, ] <- fp[5341833:5392232]
> den[3, ] <- fp[1576847:1627246]
> den[4, ] <- fp[1324850:1375249]
> den[5, ] <- fp[1375250:1425649]
> den[6, ] <- fp[1425650:1476049]
> den[7, ] <- fp[2271643:2322042]
> den[8, ] <- fp[2322043:2372442]
> den[9, ] <- fp[2372442:2422841]
```

Jednotlivé výbery budú uložené do matíc typu *vyber_čsec_xx*, kde **č** znamená, koľko sekundový výber uskutočňujeme, a na mieste **xx** je buď **eu**, čo znamená výber do 14.30 hod, **usa** po 14.30 hod, alebo nič, čo znamená celý deň.

```
> vyber_1sec <- matrix(nrow = pocet_dni, ncol = 50399)
> vyber_2sec <- matrix(nrow = pocet_dni, ncol = 25200)
> vyber_15sec <- matrix(nrow = pocet_dni, ncol = 3360)
> vyber_30sec <- matrix(nrow = pocet_dni, ncol = 1680)
> vyber_1sec_eu <- matrix(nrow = pocet_dni, ncol = 23400)
> vyber_2sec_eu <- matrix(nrow = pocet_dni, ncol = 11700)
> vyber_15sec_eu <- matrix(nrow = pocet_dni, ncol = 1560)
> vyber_30sec_eu <- matrix(nrow = pocet_dni, ncol = 780)
> vyber_1sec_usa <- matrix(nrow = pocet_dni, ncol = 26999)
> vyber_2sec_usa <- matrix(nrow = pocet_dni, ncol = 13500)
> vyber_15sec_usa <- matrix(nrow = pocet_dni, ncol = 1800)
> vyber_30sec_usa <- matrix(nrow = pocet_dni, ncol = 900)
```

Na uloženie jednotlivých výberov do matice, používame vektory *vyber_čsec_xx*. Výbery z celého dňa (Analogicky uložíme výbery do 14.30 hod a po 14.30 hod.):

```
> for (i in 1:pocet_dni) vyber_1sec[i, ] <- den[i, ][v_1sec]
> for (i in 1:pocet_dni) vyber_2sec[i, ] <- den[i, ][v_2sec]
> for (i in 1:pocet_dni) vyber_15sec[i, ] <- den[i, ][v_15sec]
> for (i in 1:pocet_dni) vyber_30sec[i, ] <- den[i, ][v_30sec]
```

3.2 Stacionarita a nezávislosť

Ako prvú otestujeme stacionaritu prírastkov pomocou KPSS testu. Všetky p-hodnoty vo všetkých výberoch nám vychádzajú približne 0.1, čo znamená, že môžeme na hladine 5% predpokladať, že sa jedná o výber zo stacionárneho rozdelenia.

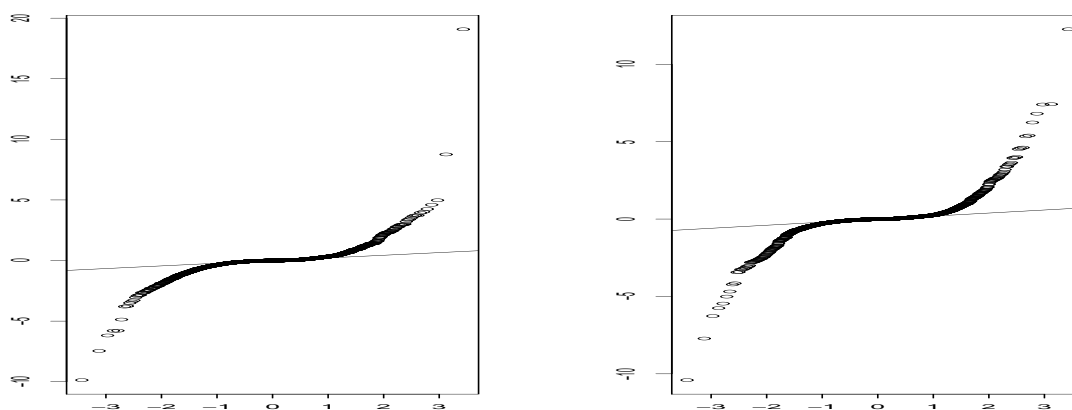
Pri testovaní nezávislosti časového radu použijeme Boxov-Piercov test. Pri testovaní výberov pri frekvenciách 1 a 2 sekundy dosiahnutá p-hodnota nepodporuje vo všetkých prípadoch nulovú hypotézu, že časový rad je nezávisle rozdelený, t.j. že jednotlivé koeficienty korelácie sa pohybujú okolo 0. Avšak pri výbere s frekvenciou 30 sekúnd už z tabuľky 3.1 vidíme, že už vo viacerých dňoch prekročila p-hodnota predpísanú hranicu 5%.

Deň	1	2	3	4	5	6	7	8	9
p-hod.	0.88	0.051	4.3e-06	0.022	0.64	0.002	0.057	1.093e-08	0.40

Tabuľka 3.1: P-hodnoty Box-Pierce testu pre 30 sekundový výber

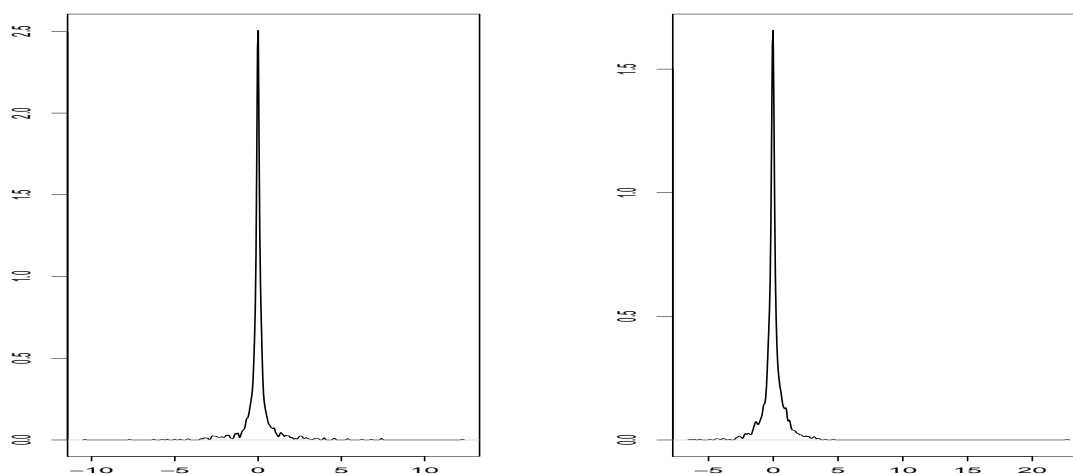
3.3 Normalita

V tejto časti sa budeme zaoberať normalitou prírastkov. Skúmať budeme hneď log-returns. Pre predstavu si necháme vykresliť Q-Q grafy (obrázok 3.1). Uvedieme len 2 grafy z 30 sekundového výberu z celého dňa.



Obrázok 3.1: Q-Q grafy

Tvar grafov nám hovorí, že sa s veľkou pravdepodobnosťou bude jednáť o výber z rozdelenia s ťažkými chvostami. Ďalej si zobrazíme odhad hustoty pomocou jadra (obrázok 3.2).



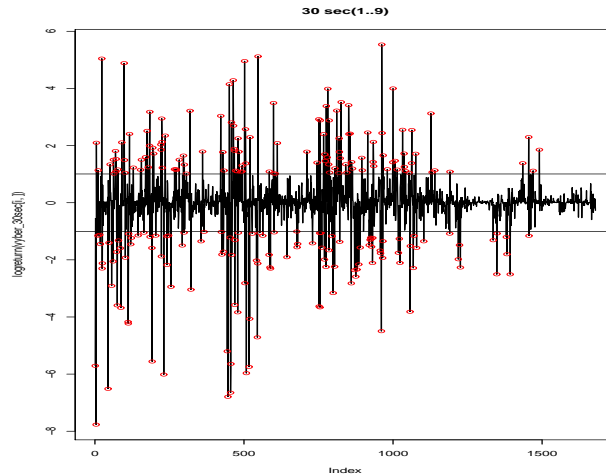
Obrázok 3.2: Odhady hustoty pomocou jadra

Teraz prevedieme testy normality. Konkrétne sa bude jednáť o testy Anderson-Darling a Shapiro-Wilk. Od skúmaných prírastkov log-returns odčítame aj nesetranný odhad strednej hodnoty a vydělíme výberovou smerodajnou odchýlkou vynásobenou odmocninou dĺžky časového intervalu a očakávame, že pôjde o výber z rozdelenia $N(0, 1)$. Konštatujeme, že neboli dosiahnuté žiadne signifikantné p-hodnoty v rôznych výberoch. Testovali sa všetky výbery vo všetkých dňoch.

Skúsime preto odfiltrovať skoky, ktoré by mohli zapríčiniť, že testy normality našu hypotézu nepotvrdzujú. Budeme postupovať nasledovne.

```
> for (i in c(1:pocet_dni)) {
+   for (j in c(1:6)) {
+     hladina <- (alpha1[j] * cas_int^0.47)
+     odfiltrované <- which(abs(logreturn(vyber_30sec[i, ])) >
+       hladina)
+     neodfiltrované <- which(abs(logreturn(vyber_30sec[i,
+       ])) <= hladina)
+     plot(logreturn(vyber_30sec[i, ]), type = "l")
+     points(odfiltrované, logreturn(vyber_30sec[i, ]),
+       [odfiltrované], col = "red")
+     abline(h = c(hladina, -hladina))
+   }
+ }
```

Hladinu sme určili z vety 7. Odfiltrované dáta zobrazuje obrázok 3.3. Ani po pokuse odfiltrovať dáta nedostávame pri testovaní na žiadnej hladine ani pri žiadnom výbere žiadny signifikantný výsledok o tom, že dáta pochádzajú z normálneho rozdelenia. Odfiltrovaných bolo vo väčšine prípadov 10 - 20 % údajov.



Obrázek 3.3: Odffitované dáta

3.4 Výberové momenty

V predchádzajúcej časti sme zistili, že normalitu predpokladať nemôžeme. Skúsime preto skúmať ako sa náš výber líši od normálneho. Porovnáme preto výberové momenty s teoretickými momentmi normálneho rozdelenia a skúsime odhadnúť ich vývoj. Pre jednotlivé dni a výbery sme vypočítali prvých 50 momentov a pomocou regresie sme sa snažili zachytiť trend ich vývoju. Vo všetkých prípadoch so zvyšujúcim sa rádom momentu klesá hodnota momentu prudko k 0. Ako najvhodnejšie sa javí aproximovať trend kubickou regresiou. V tomto prípade sa koeficient determinácie pohybuje okolo 0.5, odhad parametru pri kubickom člene je záporný a zamietame jeho nulovosť na 95% hladine spoľahlivosti.

Pre zaujímavosť uvedieme (tabuľka 3.2) porovnanie prvých šiestich výberových momentov a teoretických momentov z normálneho rozdelenia s parametrami: výberový priemer, výberový rozptyl z dát.

Moment	1.	2.	3.	4.	5.	6.
Výb. mom.	1.8e-07	1.01e-10	2.3e-16	1.7e-19	1.7e-24	6.8e-28
Teor. mom.	1.8e-07	1.01e-10	5.7e-17	3.07e-20	2.9e-26	1.5e-29

Tabuľka 3.2: Výberové a teoretické momenty pre 30 sekundový výber

Momenty boli približne rovnaké aj pre vyššie rády.

3.5 Self-similarity

Asi najzaujímavejší výsledok sme dostali pri skúmaní self-similarity. Postupovali sme podľa postupu uvedeného v kapitole 2. Za časy t sme zvolili časy: 5, 10, 15, 20, 25, ..., 120 sekúnd. Testovali sme celodenné výbery všetkých dní pri fixovaných časoch Δ , ktoré sme zvolili 30, 60 a 90 sekúnd. Testovali sme nasledovne:

```
> casy <- seq(from = 5, to = 120, by = 5)
> i = 2
```



```

> for (j in c(1:24)) {
+   rhot[j] <- kernel(vyber_1sec[i, ], j)
+ }
> k = 6
> for (j in c(1:24)) {
+   logt[j] <- log(casy[k]/casy[j])
+ }
> regrese <- lm(log(rhot) - log(rhot[k]) ~ logt - 1)
> summary(regrese)

```

Call:

```
lm(formula = log(rhot) - log(rhot[k]) ~ logt - 1)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.054013	-0.017371	-0.000818	0.037111	0.112229

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
logt	0.728477	0.009229	78.94	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04294 on 23 degrees of freedom

Multiple R-squared: 0.9963, Adjusted R-squared: 0.9962

F-statistic: 6231 on 1 and 23 DF, p-value: < 2.2e-16

Uviedli sme jeden z výsledkov, konkrétne sa jedná o druhý deň pri $\Delta = 60$. Ostatné výsledky sú veľmi podobné - koeficient determinácie sa blíži k 1 a odhad parametru H kolíše okolo 0.7. To znamená, že do veľkej miery sa potvrdila hypotéza self-similarity.

Záver

Táto práca bola predovšetkým zameraná na bližšie priblíženie teoretických vlastností Brownovho pohybu a ich testovaní. V mnohých teoretických prácach sa stretávame s predpokladom, že na vysoko likvidnom trhu sa proces cien dá modelovať nejakou transformáciou Brownovho pohybu. My sme sa toto tvrdenie pokúsili overiť na reálnej vzorke dát. Konkrétne sme sa zaoberali log returns a ukázali sme, že predpoklad normality prírastkov nie je v našom prípade úplne správny. Skúšali sme testovať rôzne výbery, no dospeli sme k záveru, že rozdelenie prírastkov má výrazne „t'ážšie chvosty“ ako normálne. Taktiež naše testy nepotvrdili nezávislosť prírastkov. Vysvetliť tieto závery by sa dalo tým, že proces obsahuje signifikantný šum, t.j. nejakú náhodnú zložku, ktorá nie je systematická a je ťažko rozpoznateľná. Ďalším cieľom práce by bola snaha túto zložku nejakým spôsobom odfiltrovať.

Dospeli sme však aj k zaujímavým výsledkom. Empirické momenty do značnej miery odpovedali tým teoretickým. Takisto sa nám do značnej miery podarilo potvrdiť hypotézu, že proces má vlastnosť self-similarity a odhad koeficientu sa pohyboval v intervale $(0.5,1)$, pričom vieme, že pre Brownov pohyb je rovný 0.5.

Zoznam použitej literatúry

- [1] ANDĚL, J. *Základy matematické statistiky*. 2. vydání. MATFYZPRESS, Praha, 2007.
- [2] TSAY, R. S. *Analysis of Financial Time Series*. JOHN WILEY & SONS, Inc., USA, 2002.
- [3] SHAPIRO, S., WILK, M., *An analysis of variance test for normality (complete samples)*. Biometrika, 1965.
- [4] THODE, H. C., Jr. *Testing for Normality*. Dekker, New York, 2002.
- [5] KARATZAS, I., SHREVE, S. E., *Brownian motion and stochastic calculus*. Springer, 1991
- [6] SHREVE, S. E., *Stochastic Calculus for Finance II*. Springer Science+Business Media, Inc., USA, 2004
- [7] REVUZ, D., YOR, M., *Continuous martingales and Brownian motion*. Springer, 3rd ed., 1999.
- [8] LALLEY, S., *Prednáška - Stochastic processes II*. University of Chicago, 2011 [online]. Dostupné z: <http://www.stat.uchicago.edu/~lalley/Courses/313/WienerProcess.pdf>
- [9] QUIGLEY, L., *Statistical Analysis of the Log Returns of Financial Assets*. Westfälische Wilhelms-Universität Muenster, BSc in Financial Mathematics, 2008. Dostupné z: <http://www.math.uni-muenster.de/statistik/paulsen/Diplomarbeiten/Quigley.pdf>
- [10] ŠPIRITOVÁ, Z., *Matematické modely v teorii portfolia*. Masarykova Univerzita, dipl. práce, 2006. Dostupné z: http://is.muni.cz/th/63585/prif_m/dip.pdf