

## CHANGE POINT PROBLEM PO DESETI LETECH

Jaromír ANTOCH, Marie HUŠKOVÁ<sup>1</sup>

a Daniela JARUŠKOVÁ

MFF UK Praha a FSt ČVUT Praha

**Abstract.** Change point detection and related problems are of vital interest of both applied and theoretical statisticians since more than twenty years. It was ten years ago when during the ROBUST'88 one of the authors of this paper presented the lecture with an analogous contents. However, since that time a lot of work has been done in this field. Therefore, we decided to summarize at least some of the most important results achieved during the last years and present them to the reader.

**Резюме:** Проблема детекции разладки в статистических моделях привлекает интерес теоретических и прикладных статистиков более чем двадцать лет. Десять лет тому назад во время конференции РОБУСТ'88 один из авторов этой статьи прочитал доклад по той же проблематике. С того времени было опубликовано много нового по этой проблематике, вот и почему мы решили собрать хотя бы самые важные (новые) результаты из данной области и предложить из читателю.

### 1. ÚVOD

Před deseti lety byla na programu ROBUSTu'88 přednáška velice podobného zaměření. Problematika detekce změny/změn ve statistických modelech a s ní související úlohy jsou neustále aktuální v tom smyslu, že byla odvozena řada zajímavých teoretických výsledků s přímým použitím v praxi, a zároveň se vyskytují další problémy jež více či méně souvisejí s touto problematikou. Cílem tohoto článku je proto alespoň částečně seznámit čtenáře s vývojem v posledních letech.

Základní výsledky i problematika detekce změn ve statistických modelech budou podrobně vysvětleny na úloze detekce změny v modelu polohy. V dalších odstavcích pak budou stručněji uvedeny výsledky pro modely složitější, ukázány způsoby výpočtu kritických hodnot a pro ilustraci uvedeny dvě názorné aplikace.

### 2. ZMĚNY V MODELU POLOHY

Nejprve budeme uvažovat model

$$(2.1) \quad Y_i = \begin{cases} \mu + e_i, & i = 1, 2, \dots, m, \\ \mu + \delta_n + e_i, & i = m + 1, \dots, n, \end{cases}$$

---

<sup>1</sup>Tento příspěvek vznikl za přispění grantu GA ČR 97/201/1163.

kde  $\mu$ ,  $\delta_n \neq 0$  a  $m$  jsou parametry, a  $e_1, \dots, e_n$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, konečným a nenulovým rozptylem  $\sigma^2$  a konečným absolutním momentem řádu  $2 + \Delta$ ,  $\Delta > 0$ . Dále budeme předpokládat, že  $m = \lfloor n\gamma \rfloor$ ,  $\gamma \in (0, 1]$ , kde  $\lfloor x \rfloor$  označuje celou část čísla  $x$ .

Model (2.1) odpovídá situaci, kdy prvních  $m$  pozorování má střední hodnotu  $\mu$  a od  $(m + 1)$ ního mají střední hodnotu  $\mu + \delta_n$ . Parametr  $\gamma$  (popř.  $\lfloor n\gamma \rfloor$ ) se nazývá *bod změny (change point)*. Parametr  $\delta_n$  odpovídá velikosti změny a obecně může záviset na počtu pozorování  $n$ . Jestliže  $\forall n \geq 1$  platí  $\delta_n = \delta \neq 0$ , budeme hovořit o *pevné změně*. Jestliže předpokládáme, že  $\delta_n \rightarrow 0$  pro  $n \rightarrow \infty$ , budeme mluvit o *lokální změně*. Konečně, je-li  $\delta_n = 0$ , bude se jednat o situaci, kdy nedošlo k *žádné změně*.

Základní úlohou je testovat hypotézu

$$(2.2) \quad H_0 : \gamma = 1 \quad \text{proti alternativě} \quad H_1 : \gamma \in (0, 1),$$

a jestliže hypotézu zamítneme, potom také  $\gamma$ , popřípadě další parametry modelu, odhadnout. V této první úloze vlastně testujeme hypotézu, že *žádná změna nenastala* proti alternativě, že *změna nastala*.

## 2.1. TESTY

Testové statistiky, které se většinou používají, jsou zpravidla odvozeny buď od testu odpovídajícího metodě podílem věrohodností, nebo vycházejí z bayesovského principu. První typ statistik vede k postupům maximálního typu (tzv. *max-type postupy*), zatímco druhý vede k postupům založeným na částečných součtech (tzv. *sum-type postupy*).

Předpokládejme nejprve, že  $e_i$  mají rozdělení  $N(0, \sigma^2)$ ,  $\sigma^2 > 0$  známé. Test podílem věrohodností je založen na statistice

$$(2.3) \quad \max_{1 \leq k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\sigma} \right\},$$

kde

$$S_k = \sum_{i=1}^k (Y_i - \bar{Y}_n), \quad k = 1, \dots, n, \quad \text{a} \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Velké hodnoty této testové statistiky indikují, že nulová hypotéza neplatí, zatímco hodnoty blízké 0 odpovídají platnosti nulové hypotézy. Přesnou kritickou hodnotu nelze bohužel stanovit, nicméně lze nalézt její aproximace.

Tyto aproximace mohou být založeny na limitní distribuční funkci testové statistiky za platnosti  $H_0$ , na simulacích, na metodě bootstrap nebo na různých nerovnostech, blíže viz odstavec 4.

Nechť  $\epsilon > 0$  je malá pevná konstanta. Lze ukázat, že za  $H_0$  platí

$$\begin{aligned} \max_{1 \leq k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\sigma} \right\} &\xrightarrow{P} \infty \quad \text{pro } n \rightarrow \infty, \\ \max_{n\epsilon \leq k < n(1-\epsilon)} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\sigma} \right\} &= O_P(1) \quad \text{pro } n \rightarrow \infty. \end{aligned}$$

Toto jsou nepříjemné vlastnosti, neboť z nich plyne, že dominantní vliv mají pouze členy odpovídající malému  $k$ , nebo  $k$  blízkému  $n$ . Proto byly navrženy takové modifikace, které tyto nepříjemné vlastnosti nemají. Některé z nich uvádíme v dalším textu.

Jestliže  $\sigma^2$  neznáme, což je v praxi téměř pravidlem, musíme jej nahradit některým rozumným odhadem. Ukazuje se, že

$$(2.4) \quad \hat{\sigma}_n^2 = \frac{1}{n-2} \min_{1 \leq k \leq n} \left( \sum_{i=1}^k (Y_i - \bar{Y}_k)^2 + \sum_{i=k+1}^n (Y_i - \bar{Y}_k^o)^2 \right),$$

kde

$$\bar{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i \quad \text{a} \quad \bar{Y}_k^o = \frac{1}{n-k} \sum_{i=k+1}^n Y_i, \quad k = 1, \dots, n-1,$$

je takovým vhodným odhadem, neboť je dobrý při  $H_0$  i  $H_1$ . Pozorný čtenář si jistě všiml faktu, že hodnota  $k$  maximalizující (2.3) je totožná s hodnotou minimalizující (2.4), což může podstatně zjednodušit výpočty. Další možné odhady jsou popsány v poznámce 2.7.

Místo se statistikou (2.3) typicky pracujeme spíše s testovými statistikami tvaru

$$(2.5) \quad T_n = \max_{1 \leq k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\hat{\sigma}_n} \right\},$$

$$(2.6) \quad T_{n0}(\epsilon) = \max_{n\epsilon \leq k < n(1-\epsilon)} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\hat{\sigma}} \right\},$$

$$(2.7) \quad T_{n1}(q) = \max_{1 \leq k \leq n} \left\{ \frac{1}{\sqrt{n}q(k/n)} \frac{|S_k|}{\hat{\sigma}_n} \right\},$$

$$(2.8) \quad T_{n2}(G) = \max_{G < k \leq n} \left\{ \frac{1}{\sqrt{G}\hat{\sigma}_n} |S_k - S_{k-G}| \right\},$$

$$(2.9) \quad T_{n3}(G) = \max_{G < k \leq n} \left\{ \frac{1}{\sqrt{2G} \widehat{\sigma}_n} |S_{k+G} - 2S_k + S_{k-G}| \right\},$$

$$(2.10) \quad T_{n4}(r) = \frac{1}{n} \sum_{k=1}^n \frac{1}{r(k/n)} \frac{S_k^2}{\widehat{\sigma}_n^2},$$

kde  $q(\cdot)$  and  $r(\cdot)$  jsou váhové funkce definované na  $(0, 1)$  a  $G \ll n$ . Časté volby jsou

$$(2.11) \quad q(t) \equiv q_\eta(t) = (t(1-t))^\eta, \quad t \in (0, 1), \quad \eta \in [0, 1/2),$$

$$(2.12) \quad r(t) \equiv r_\eta(t) = (t(1-t))^\eta, \quad t \in (0, 1), \quad \eta \in [0, 1].$$

Citlivost uvedených statistik vůči sledovaným hypotézám je vidět z následujících středních hodnot a rozptylů

$$E_{H_0} S_k = 0, \quad 1 \leq k \leq n,$$

$$(2.13) \quad E_{H_1} S_k = \begin{cases} -\delta_n k \frac{n-m}{n}, & 1 \leq k \leq m, \\ -\delta_n (n-k) \frac{m}{n}, & m < k \leq n, \end{cases}$$

$$\text{var}_{H_0} S_k = \text{var}_{H_1} S_k = \sigma^2 \frac{k(n-k)}{n}, \quad 1 \leq k \leq n,$$

$$E_{H_0} (S_k - S_{k-G}) = 0, \quad G < k \leq n,$$

$$(2.14) \quad E_{H_1} (S_k - S_{k-G}) = \begin{cases} -\delta_n G \frac{n-m}{n}, & G < k \leq m-G, \\ \delta_n \left( k - m - G \frac{n-m}{n} \right), & m-G < k \leq m, \\ \delta_n G \frac{m}{n}, & m < k \leq n. \end{cases}$$

$$E_{H_0} (S_{k+G} - 2S_k + S_{k-G}) = 0, \quad G < k \leq n-G,$$

(2.15)

$$E_{H_1}(S_{k+G} - 2S_k + S_{k-G}) = \begin{cases} 0, & G < k \leq m - G, \\ \delta_n(k + G - m), & m - G < k \leq m, \\ \delta_n(m + G - k), & m < k \leq m + G, \\ 0, & m + G < k < n - G, \end{cases}$$

$$E_{H_0}S_k^2 = \sigma^2 \frac{k(n-k)}{n}, \quad 1 \leq k \leq n,$$

$$E_{H_1}S_k^2 = \sigma^2 \frac{k(n-k)}{n} + \delta_n^2 \left( \frac{\min(k, m)(n - \max(m, k))}{n} \right)^2, \quad 1 \leq k \leq n.$$

Povšimněte si, že absolutní hodnoty  $|E_{H_1}(S_k)|$  a  $|E_{H_1}(S_{k+G} - 2S_k + S_{k-G})|$  jakožto funkce  $k$  dosahují maxima pro  $k = m$ .

Nyní zformulujeme tvrzení o limitním chování uvažovaných testových statistik. Označme přitom

$$(2.16) \quad \begin{cases} Q_{0,1} = \left\{ q \mid q \text{ je neklesající v okolí nuly a jedné, přičemž} \\ \quad \inf_{\eta \leq t \leq 1-\eta} q(t) > 0 \text{ pro všechna } 0 < \eta < 1/2 \right\}, \\ a(x) = \sqrt{2 \log x}, \quad b(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi, \\ I(q, c) = \int_0^1 \frac{1}{t(1-t)} \exp \left\{ -\frac{c q^2(t)}{t(1-t)} \right\} dt. \end{cases}$$

Všimněte si, že váhová funkce (2.11) s  $\eta \in [0, 1/2)$  náleží do  $Q_{0,1}$ .

V následující větě jsou zformulovány výsledky o limitním chování testových statistik při platnosti nulové hypotézy  $H_0$ :

**Věta 2.1.** *Uvažujme model (2.1) a předpokládejme, že platí  $H_0$ . Potom:*

(i) *Pro všechna  $y \in R_1$  platí*

$$(2.17) \quad \lim_{n \rightarrow \infty} P(a(\log n) T_n - b(\log n) \leq y) = \exp \left\{ -2 \exp \{ -y \} \right\}.$$

(ii) *Jestliže  $q \in Q_{0,1}$  a  $I(q, c) < \infty$  pro nějaké  $c > 0$ , pak  $\forall x \in R_1$*

$$(2.18) \quad \lim_{n \rightarrow \infty} P(T_{n1}(q) \leq x) = P\left( \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \leq x \right).$$

Dále  $\forall x \in R_1$  platí

$$(2.18') \quad \lim_{n \rightarrow \infty} P(T_{n0}(\varepsilon) \leq x) = P\left(\sup_{\varepsilon < t < 1-\varepsilon} \frac{|B(t)|}{\sqrt{t(1-t)}} \leq x\right).$$

Ve výrazech (2.18) i (2.18')  $\{B(t); t \in (0, 1)\}$  značí Brownův můstek.

(iii) Jestliže pro  $n \rightarrow \infty$  platí

$$(2.19) \quad G_n \rightarrow \infty \quad \text{a} \quad G_n^{-1} n^{\frac{2}{2+\Delta}} \log n \rightarrow 0,$$

pak  $\forall y \in R_1$

$$(2.20) \quad \lim_{n \rightarrow \infty} P\left(a(n/G_n) T_{n2}(G_n) - b(n/G_n) \leq y\right) = \exp\{-2 \exp\{-y\}\}$$

a

$$(2.21) \quad \lim_{n \rightarrow \infty} P\left(a(n/G_n) T_{n3}(G_n) - b(n/G_n) \leq y\right) = \exp\{-2 \exp\{-y\}\}.$$

(iv) Jestliže  $\inf_t r(t) > 0$  a  $\int_0^1 t(1-t)r^{-2}(t) dt < \infty$ , pak  $\forall x \in R_1$

$$\lim_{n \rightarrow \infty} P(T_{n4}(r) \leq x) = P\left(\int_0^1 \frac{B^2(t)}{r(t)} dt \leq x\right).$$

*Důkaz.* Tvzení (i) a (ii) lze nalézt např. v knihách Csörgő a Horváth (1993, 1997), důkaz (iii) je v článku Chen (1988) a důkaz (iv) v knize Csörgő a Horváth (1997). □

*Poznámka 2.1.* Věta 2.1 patří k základním větám v oblasti detekce změny. Její analogie byly dokázány pro řadu dalších statistik a modelů, takže některé její varianty budou uvedeny později.

Lze ukázat, že při  $H_0$  jsou statistiky  $T_n$  a  $T_{n1}(q)$  pro váhové funkce  $q \in Q_{0,1}$  asymptoticky nezávislé. Dále z věty 2.1 vyplývá, že limitní rozdělení statistik  $T_n, T_{n2}(G)$  a  $T_{n3}(G)$  náleží při  $H_0$  do třídy rozdělení tzv. extrémálních hodnot a nezávisí na distribuční funkci pozorování. K výpočtu hodnot odpovídající limitní distribuční funkce stačí standardní kapesní kalkulačka. Poznamenejme ale, že konvergence je poměrně pomalá a pro praktické nepříliš užitečná.

Naproti tomu statistiky  $T_{n0}(\varepsilon), T_{n1}(q)$  a  $T_{n4}(r)$  mají v limitě rozdělení jako funkcionály Brownova můstku, jejichž rozdělení je známo jen pro některé

velmi speciální váhové funkce  $q(\cdot)$ . Bližší informace lze nalézt ve čtvrtém odstavci.

*Poznámka 2.2.* Lze ukázat, že testy založené na uvažovaných testových statistik jsou konzistentní. Podrobné informace lze nalézt např. v Horváth, Hušková a Serbinowska (1997).

## 2.2. ODHADY

Parametry v modelu (2.1) lze odhadnout například metodou nejmenších čtverců. Odhady  $\hat{\mu}_n, \hat{\delta}_n, \hat{m}$  parametrů  $\mu, \delta_n$  a  $m$  jsou definovány jako řešení minimalizační úlohy

$$\min \left\{ \sum_{i=1}^k (Y_i - \mu)^2 + \sum_{i=k+1}^n (Y_i - \mu - \delta)^2 \mid k \in \{1, \dots, n\}, \mu, \delta \in R_1 \right\}.$$

Mají-li náhodné chyby  $e_i$  rozdělení  $N(0, \sigma^2)$ , pak se odhady metodou nejmenších čtverců shodují s odhady metodou maximální věrohodnosti, podrobnosti viz např. Gombay a Horváth (1996). Přímým výpočtem lze snadno ověřit, že

$$(2.22) \quad \hat{\mu}_n = \bar{Y}_{\hat{m}} \quad \text{a} \quad \hat{\delta}_n = \bar{Y}_{\hat{m}}^o - \bar{Y}_{\hat{m}},$$

kde  $\hat{m}$  je řešením minimalizační úlohy

$$(2.23) \quad \min \left\{ \sum_{i=1}^k (Y_i - \bar{Y}_k)^2 + \sum_{i=k+1}^n (Y_i - \bar{Y}_k^o)^2 \mid k \in \{1, \dots, n-1\} \right\}.$$

Má-li tato minimalizační úloha více řešení, volíme řešení s minimální hodnotou. Přímým výpočtem lze dále ověřit, že  $\hat{m}$  lze ekvivalentně definovat jako

$$(2.24) \quad \begin{aligned} \hat{m} &= \arg \max \left\{ \sqrt{\frac{n}{k(n-k)}} |S_k| \mid k = 1, \dots, n-1 \right\} \\ &= \arg \max \left\{ \frac{k(n-k)}{n} (\bar{Y}_k - \bar{Y}_k^o)^2 \mid k = 1, \dots, n-1 \right\}. \end{aligned}$$

Přirozenou modifikací odhadu  $\hat{m}$  jsou pro  $\eta \in [0, 1/2)$  odhady

$$(2.25) \quad \hat{m}(\eta) = \arg \max \left\{ \left( \frac{n}{k(n-k)} \right)^\eta |S_k| \mid k = 1, \dots, n-1 \right\}.$$

Souvislost s testovými statistikami je zde zřejmá.

Dalším typem odhadů parametru  $m$  jsou odhady založené na diferencích klouzavých součtů (tzv. *MOSUM – moving sums*). Definujeme je následovně:

$$(2.26) \quad \hat{m}(G) = \arg \max \left\{ |S_{k+G} - 2S_k + S_{k-G}| \mid k = G+1, \dots, n-G \right\}.$$

Jednoduchou představu o očekávaných vlastnostech tohoto odhadu lze získat z (2.15).

Nyní zformulujeme základní věty o chování odhadů bodu změny.

**Věta 2.2 (pevná změna).** *Předpokládejme, že model (2.1) platí,  $\delta_n = \delta \neq 0$ ,  $m = \lfloor n\gamma \rfloor$ ,  $\gamma \in (0, 1)$ . Pak  $\hat{m} - m$  má v limitě stejné rozdělení jako*

$$\arg \max \left\{ \delta V_j - \frac{\delta^2 |j|}{2} \mid j = 0, 1, 2, \dots \right\},$$

kde

$$V_j = \begin{cases} 0, & j = 0, \\ \sum_{i=j}^0 Z_i, & j = -1, -2, \dots, \\ -\sum_{i=1}^j Z_i, & j = 1, 2, \dots, \end{cases}$$

kde  $\{Z_i\}$  je posloupnost nezávislých stejně rozdělených náhodných veličin takových, že  $Z_i$  má rozdělení jako  $e_i$ .

*Důkaz.* Lze nalézt v Antoch a Hušková (1998). □

**Věta 2.3 (lokální změna).** *Uvažujme model (2.1). Necht'  $m = \lfloor n\gamma \rfloor$ ,  $\gamma \in (0, 1)$  a pro  $n \rightarrow \infty$*

$$(2.27) \quad \delta_n \rightarrow 0 \quad \text{a} \quad \frac{|\delta_n| \sqrt{n}}{\sqrt{\log \log n}} \rightarrow \infty.$$

*Pak pro  $n \rightarrow \infty$*

$$(2.28) \quad \frac{\delta_n^2}{\sigma^2} (\hat{m}_1(\eta) - m) \xrightarrow{\mathcal{D}} V_{\gamma, \eta}, \quad \eta \in [0, 1/2].$$



Jestliže navíc platí (2.19), pak pro  $n \rightarrow \infty$

$$(2.29) \quad \frac{\delta_n^2}{6\sigma^2} (\widehat{m}(G) - m) \xrightarrow{\mathcal{D}} V_0,$$

kde

$$(2.30) \quad V_{\gamma, \eta} = \arg \max \left\{ W(t) - |t| \left[ ((1-\gamma)(1-\eta) + \eta\gamma) I\{t < 0\} + ((1-\eta)\gamma + \eta(1-\gamma)) I\{t \geq 0\} \right]; t \in R_1 \right\},$$

$$(2.31) \quad V_0 = \arg \max \left\{ W(t) - |t|; t \in R_1 \right\},$$

$\{W(t), t \in R_1\}$  je dvoustranný Wienerův proces a  $I\{A\}$  označuje indikátor množiny  $A$ .

*Důkaz.* Lze najít v Antoch, Hušková a Veraverbeke (1995). □

*Poznámka 2.3:* Připomeňme, že dvoustranný Wienerův proces  $\{W(t), t \in R_1\}$  je definován následovně

$$W(t) = \begin{cases} W_1(-t), & t < 0, \\ W_2(t), & t \geq 0, \end{cases}$$

kde  $\{W_1(t), t > 0\}$  a  $\{W_2(t), t > 0\}$  jsou nezávislé Wienerovy procesy.

Pro úplnost uvedeme limitní chování odhadů bodu změny, jestliže změna nenastala. Pozorný čtenář si jistě povšimne, že chování odhadů je v tomto případě zcela jiné než v případě, kdy změna nastala.

**Věta 2.4 (žádná změna).** *Nechť  $Y_1, \dots, Y_n$  splňují model (2.1) a  $m = n$ , tj. nenastala žádná změna. Potom pro libovolné  $\eta \in [0, 1/2]$  platí*

$$\frac{\widehat{m}(\eta)}{n} \xrightarrow{\mathcal{D}} U(\eta) \quad \text{pro } n \rightarrow \infty,$$

kde  $U(\eta)$  je náhodná veličina s rozdělením:

(i) je-li  $\eta = 1/2$ , pak

$$P(U(1/2) = 0) = P(U(1/2) = 1) = 1/2;$$

(ii) je-li  $\eta \in [0, 1/2)$ , PAK  $U(\eta)$  má rozdělení jako

$$\arg \max \left\{ |B(s)| (s(1-s))^{-\eta} \mid s \in R_1 \right\},$$

kde  $\{B(t), t \in (0, 1)\}$  je Brownův můstek.

Jestliže je navíc splněno (2.19), pak

$$\frac{\widehat{m}(\eta)}{n} \xrightarrow{\mathcal{D}} U^o \quad \text{pro } n \rightarrow \infty,$$

kde  $U^o$  má rovnoměrné rozdělení  $R(0, 1)$ . □

Nyní se krátce zmíníme o vlastnostech odhadů  $\widehat{\mu}$ ,  $\widehat{\delta}_n$  a  $\widehat{\sigma}_n^2$  parametrů  $\mu$ ,  $\delta_n$  and  $\sigma^2$ , jež byly definovány v (2.22), (2.23) a (2.4).

**Věta 2.5.** *Nechť jsou splněny předpoklady věty 2.3 a necht'  $\widehat{m}$  je odhad bodu změny  $m$  s vlastností  $\widehat{m} - m = O_P(\delta_n^{-2})$ . Potom:*

- i)  $\widehat{\mu}_n = \overline{Y}_{\widehat{m}}$  a  $\widehat{\delta}_n = \overline{Y}_{\widehat{m}}^o - \overline{Y}_{\widehat{m}}$  jsou  $\sqrt{n}$ -konzistentní odhady parametrů  $\mu$  a  $\delta_n$ .  
Navíc  $\sqrt{m}(\widehat{\mu}_n - \mu)/\sigma$  a  $\sqrt{n-m}(\widehat{\delta}_n - \delta_n)/\sigma$  mají asymptoticky  $N(0, 1)$  rozdělení a jsou v limitě nezávislé.
- ii) Odhad  $\widehat{\sigma}_n^2$  je  $n^\beta$ -konzistentní odhad rozptylu  $\sigma^2$  pro nějaké  $\beta > 0$ .

*Poznámka 2.4.* Limitní rozdělení ostatních odhadů bodu změny při pevné změně lze nalézt v článku Antoch a Hušková (1998).

*Poznámka 2.5.* Je-li  $\delta_n = \delta \neq 0$  (pevná změna), pak limitní rozdělení odhadů  $\widehat{m}$ ,  $\widehat{m}(\eta)$  a  $\widehat{m}(G)$  závisí na distribuční funkci  $F$  náhodných chyb  $e_i$ , zatímco v případě lokálních změn (resp.  $\delta = 0$ ) jsou odhady tzv. „distribution free“, tj. limitní rozdělení na  $F$  nezávisí. Podrobnosti viz Antoch a Hušková (1998).

*Poznámka 2.6.* Tvrzení věty 2.2 zůstává v platnosti, jestliže parametry  $\delta_n$  and  $\sigma^2$  nahradíme vhodnými odhady. Speciálně, parametry  $\delta_n$  a  $\sigma^2$  mohou být nahrazeny odhady  $\widehat{\delta}_n$  a  $\widehat{\sigma}_n^2$  definovanými v (2.23) či (2.4).

*Poznámka 2.7.* Hodnoty  $\sigma^2$  lze též odhadnout pomocí

$$\widehat{\sigma}_n^2 = \frac{1}{n-2} \left( \sum_{i=1}^{\widehat{m}} (Y_i - \widehat{\mu})^2 + \sum_{i=1+\widehat{m}}^n (Y_i - \widehat{\mu} - \widehat{\delta})^2 \right),$$

kde  $\widehat{m}$ ,  $\widehat{\mu}$  a  $\widehat{\delta}$  je „libovolná vhodná kombinace“ výše uvedených odhadů.

*Poznámka 2.8.* Explicitní vyjádření distribuční funkce náhodných veličin  $V_{\eta, \gamma}$ ,  $\eta \in [0, 1/2]$ ,  $\gamma \in (0, 1)$ , a  $V_0$  je známé. Bylo odvozeno nezávisle na sobě několika autory, viz např. Bhattacharya a Brockwell (1976), Shepp (1979), Yao (1987) či Ferger (1994). Sthryn (1996) ukázal, že pro  $c_1 > 0$  a  $c_2 > 0$

$$(2.32) \quad V(c_1, c_2) = \arg \max \left\{ W(s) - |s| \left( c_1 I\{s \leq 0\} + c_2 I\{s > 0\} \right); s \in R_1 \right\}$$

má hustotu

$$(2.33) \quad p(t; c_1, c_2) = \begin{cases} q(-t; c_1, c_2), & t \leq 0, \\ q(t; c_2, c_1), & t > 0, \end{cases}$$

kde pro  $t > 0$

$$q(t; c_1, c_2) = 2c_1(c_1 + 2c_2) \exp \{2c_2(c_1 + c_2)t\} \Phi(- (c_1 + 2c_2)\sqrt{t}) - 2c_1^2 \Phi(-c_1\sqrt{t}).$$

Odpovídající distribuční funkce může být spočítána, jestliže si uvědomíme, že pro  $t > 0$

$$\begin{aligned} \int_0^t q(x; c_1; c_2) dx &= \frac{c_1}{c_1 + c_2} + \frac{2c_1}{\sqrt{2\pi}} \sqrt{t} \exp \{ -c_1^2 t/2 \} \\ &+ \frac{c_1(c_1 + 2c_2)}{c_2(c_1 + c_2)} \exp \{ 2c_2(c_1 + c_2)t \} \Phi(- (c_1 + 2c_2)\sqrt{t}) \\ &- \left[ 2c_1^2 t + \frac{c_1^2 + 2c_2^2 + 2c_1c_2}{c_2(c_1 + c_2)} \right] \Phi(-c_1\sqrt{t}). \end{aligned}$$

Tento výsledek lze využít při konstrukci intervalových odhadu bodu změny  $m$ . Tabulky pro  $c_1 = c_2 = 1/2$  jsou v článku Horváth, Hušková a Serbinowska (1997). Poznamenejme nicméně, že přes explicitní tvar hustoty i distribuční funkce vlastní výpočet není pro  $c_1 \neq c_2$  z numerického hlediska příliš jednoduchý, neboť:

- 1) jenom v málokterém software je dostatečně přesně (spolehlivě) naprogramován výpočet hodnot distribuční funkce  $\Phi(x)$  standardního normálního rozdělení pro velké hodnoty  $x$ ;
- 2) je nutné počítat součiny velmi velkých a velmi malých čísel, což může způsobovat netušené problémy.

### 2.3. POSTUPY ZALOŽENÉ NA $M$ - A $R$ -STATISTIKÁCH

Nyní si uvedeme testy a odhady založené na  $R$ - a  $M$ -statistikách. Budeme přitom (z hlediska požadavků na znalost rozdělení chyb) uvažovat model poněkud obecnější než (2.1), tj. budeme předpokládat, že  $Y_1, \dots, Y_n$  splňují

$$(2.1') \quad Y_i = \begin{cases} \mu + e_i, & i = 1, \dots, m, \\ \mu + \delta_n + e_i, & i = m + 1, \dots, n, \end{cases}$$

kde  $\mu, \delta_n$  a  $m$  jsou neznámé parametry a  $e_1, \dots, e_n$  jsou nezávislé stejně rozdělené náhodné veličiny se společnou distribuční funkcí  $F$ .

*R-postupy*, tj. postupy vycházející z pořadových testů, jsou založeny na jednoduchých lineárních pořadových statistikách

$$(2.34) \quad S_{k,R} = \sum_{i=1}^k \left( a_n(R_i) - \bar{a}_n \right), \quad k = 1, \dots, n,$$

kde  $R_1, \dots, R_n$  jsou pořadí odpovídající pozorováním  $Y_1, \dots, Y_n$ ,  $a_n(1), \dots, a_n(n)$  jsou skóry a

$$\bar{a}_n = \frac{1}{n} \sum_{i=1}^n a_n(i).$$

$R$ -testy a  $R$ -odhady získáme, jestliže v (2.5)–(2.10) a (2.24)–(2.26) nahradíme částečné součty  $S_k$  částečnými součty  $S_{k,R}$ , zatímco veličinu  $\sigma^2$  veličinou

$$\hat{\sigma}_R^2 = \frac{1}{n-1} \sum_{i=1}^n \left( a(i) - \bar{a}_n \right)^2.$$

**Věta 2.6.** *Nechť distribuční funkce  $F$  je spojitá a nechť existuje monotonní funkce  $\phi$  definovaná na  $(0, 1)$  taková, že pro nějaké  $\Delta > 0$*

$$\int_0^1 |a_n(1 + [un]) - \phi(u)|^{2+\Delta} du \rightarrow 0 \quad \text{pro } n \rightarrow \infty,$$

$$\int_0^1 \phi(u) du = 0 \quad \text{a} \quad 0 < \int_0^1 |\phi(u)|^{2+\Delta} du < \infty.$$

Potom věta 2.1 zůstává v platnosti, jestliže  $S_k$ ,  $k = 1, \dots, n$ , a  $\sigma^2$  nahradíme statistikami  $S_{k,R}$ ,  $k = 1, \dots, n$ , resp.  $\hat{\sigma}_R^2$ .

Jestliže je navíc Fisherova informace nenulová a konečná, tj.

$$0 < \int_{-\infty}^{\infty} \frac{(f'(x))^2}{f(x)} dx < \infty,$$

věta 2.3 zůstává v platnosti, jestliže podíl  $\delta_n^2/\sigma^2$  nahradíme podílem

$$\frac{b_R^2 \delta_n^2}{\hat{\sigma}_R^2}, \quad \text{kde} \quad b_R = - \int_{-\infty}^{\infty} \phi(F(x)) f'(x) dx.$$

*Důkaz.* Lze nalézt v Hušková (1997) či Gombay a Hušková (1998).

□

Uvažované předpoklady jsou poněkud silnější než například u vět o limitním rozdělení pořadových statistik pro dvouvýběrový problém. Poznamenejme nicméně, že podmínky jsou splněny pro všechny běžně používané skóry. Na druhé straně tvrzení platí i v obecnějších modelech. Další informace lze nalézt např. v článcích Antoch a Hušková (1998), Gombay, Hušková (1998) a Hušková (1997).

Dále obrátíme pozornost k *M-postupům*. Tyto postupy jsou založeny na částečných součtech *M*-reziduálů

$$(2.35) \quad S_{k,M} = \sum_{i=1}^k \psi(Y_i - \hat{\mu}_n^M), \quad k = 1, \dots, n,$$

kde  $\psi(x)$ ,  $x \in R_1$ , je skórová funkce s vlastností

$$\int_{-\infty}^{\infty} \psi(x) dF(x) = 0$$

a  $\hat{\mu}_n^M$  je *M*-odhad generovaný skórovou funkcí  $\psi$ . Tento odhad je zpravidla definovaný jako řešení rovnice

$$(2.36) \quad \sum_{i=1}^n \psi(Y_i - t) = 0.$$

Poznamenejme, že lze též užít některý asymptoticky ekvivalentní odhad.

Označme

$$\begin{aligned} \sigma^2(\psi) &= \int_{-\infty}^{\infty} \psi^2(x) dF(x), \\ \hat{\sigma}_M^2(\psi) &= \frac{1}{n-1} \sum_{i=1}^n \psi^2(Y_i - \hat{\mu}_n^M), \\ \lambda(t) &= - \int_{-\infty}^{\infty} \psi(x-t) dF(x), \quad t \in R_1, \\ \kappa(t) &= \int_{-\infty}^{\infty} \psi^2(x-t) dF(x), \quad t \in R_1. \end{aligned}$$

**Věta 2.7.** Uvažujme model (2.1). Necht' skórová funkce  $\psi$  je monotonní, funkce  $\lambda(t)$  má spojitou první derivaci v okolí bodu 0,  $\lambda(0) = 0$  a  $\lambda'(0) \neq 0$ . Funkce  $\kappa(t)$  necht' je spojitá v okolí 0,  $\kappa(0) > 0$  a pro nějaké  $D > 0$

$$\sup_{|t| < D} \int_{-\infty}^{\infty} |\psi(u+t)|^{2+\Delta} dF(u) < \infty.$$

Potom věta 2.1 zůstává v platnosti, jestliže  $S_k$ ,  $k = 1, \dots, n$ , a  $\sigma^2$  nahradíme statistikami  $S_{k,M}$ ,  $k = 1, \dots, n$ , a  $\hat{\sigma}_M^2(\psi)$ .

Věta 2.3 též zůstává v platnosti, jestliže podíly  $\delta_n^2/\sigma^2$  nahradíme podíly  $\lambda'(0)^2 \delta_n^2/\hat{\sigma}_M^2(\psi)$ .

*Důkaz.* Lze nalézt v Hušková (1997). □

Většina poznámek (příslušně zmodifikovaných), jež byly uvedeny v sekcích 2.1 a 2.2, platí i pro situace uvažované v této sekci.

Existuje řada dalších postupů založených na různých statistikách užívaných v neparametrické a robustní statistice, např. postupy založené na  $U$ -statistikách, statistikách Kolmogorovova–Smirnovova typu či založené na různých vzdálenostech. Bližší informace lze nalézt např. v Dümbgen (1991), Ritov (1990), Ferger (1994a, 1994b, 1995), Brodsky a Darkhovskiy (1993) či Horváth a Qi-Man Shao (1995).

## 2.4. ZÁVISLÁ POZOROVÁNÍ

Uvažujme model

$$(2.37) \quad \begin{aligned} Y_t &= \mu + e_t, & t = 1, 2, \dots, m, \\ &= \mu + \delta_n + e_t, & t = m + 1, \dots, n, \end{aligned}$$

kde  $\mu$ ,  $\delta_n \neq 0$  a  $m(< n)$  jsou neznámé parametry a  $\{e_t\}_{t=1}^{\infty}$  je lineární proces. Necht'

$$(2.38) \quad e_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j}, \quad t = 1, 2, \dots,$$

kde  $\{\varepsilon_t\}_{t=-\infty}^{\infty}$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, rozptylem  $\sigma^2 > 0$ ,  $E|\varepsilon_t|^{2+\Delta} < \infty$  pro nějaké  $\Delta > 0$  a váhy necht'  $\{w_j\}_{j=0}^{\infty}$  splňují

$$(2.39) \quad \sum_{j=0}^{\infty} j |w_j| < \infty, \quad \sum_{j=0}^{\infty} w_j \neq 0.$$

Předpokládejme, že  $m = \lfloor n\gamma \rfloor$  pro nějaké  $\gamma \in (0, 1]$ . Uvažujme stejné úlohy jako v předchozích odstavcích. Pro test hypotézy  $H_0$  proti  $H_1$  můžeme použít testové statistiky  $T_n, T_{n0}, T_{n1}(q), T_{n2}(G)$  i  $T_{n4}(r)$ , pokud v nich (kde zapotřebí) nahradíme odhad  $\hat{\sigma}_n^2$  odhadem

$$(2.39) \quad \tilde{\sigma}_n^2(L) = \hat{R}(0) + 2 \sum_{k=1}^L \left(1 - \frac{k}{L}\right) \hat{R}(k), \quad L < n,$$

kde

$$\hat{R}(k) = \frac{1}{n} \left\{ \sum_{t=1}^{\hat{m}-k} (Y_t - \bar{Y}_{\hat{m}}) (Y_{t+k} - \bar{Y}_{\hat{m}}) + \sum_{t=\hat{m}+1}^{n-k} (Y_t - \bar{Y}_{\hat{m}}^o) (Y_{t+k} - \bar{Y}_{\hat{m}}^o) \right\}, \quad k \geq 0,$$

a jako odhady bodu změny  $m$  uijeme, analogicky modelu (2.1), některý z odhadů definovaných vztahy (2.22)–(2.24). Veličina  $\tilde{\sigma}_n^2(L)$  je odhadem  $\sigma_w^2 = \sigma^2 \left( \sum_{j=0}^{\infty} w_j \right)^2$ .

Jsou-li splněny předpoklady (2.36)–(2.38) a platí-li  $H_0$ , je proces  $\{X_t\}$  stacionární druhého řádu se spektrální hustotou  $g(\cdot)$ , takže asymptotický rozptyl  $\sqrt{n} \bar{Y}_n$  je  $2\pi g(0) = \sigma_w^2$ , podrobnosti viz Brockwell a Davis (1991), kapitola 7.2.

**Věta 2.8.** Předpokládejme, že (2.37)–(2.39) jsou splněny, a necht' pro  $n \rightarrow \infty$

$$L \equiv L_n \rightarrow \infty \quad \text{a} \quad \frac{L^2 \log L}{n} \rightarrow 0.$$

Potom věty 2.1 a 2.3 zůstávají v platnosti, jestliže  $\hat{\sigma}_n^2$  nahradíme  $\tilde{\sigma}_n^2(L)$ .

Při platnosti  $H_0$  dále platí

$$\tilde{\sigma}_n^2(L) - \sigma_w^2 = O_p \left( \frac{1}{L} + \sqrt{\frac{L^2 \log L}{n}} \right) \quad \text{pro} \quad n \rightarrow \infty.$$

Jestliže navíc předpokládáme platnost (2.27), potom platí

$$\tilde{\sigma}_n^2(L) - \sigma^2 = o_p(1) \quad \text{pro } n \rightarrow \infty.$$

Důkaz. Lze nalézt v Antoch, Hušková a Prášková (1998). □

*Poznámka 2.9.* Věta 2.8 říká, že v případě závislých pozorování můžeme použít tytéž testové statistiky a odhady jako v případě nezávislých pozorování, jenom je nutné použít jinou standardizaci, tj.  $\sigma^2$  a  $\tilde{\sigma}_n^2$  zaměnit  $\sigma_w^2$  a  $\tilde{\sigma}^2(L)$ . Simulace ukazují, že kvalita odhadu  $\tilde{\sigma}_n^2(L)$  závisí na  $L$  a na síle závislosti. Odhad je rozumný jen tehdy, je-li závislost slabá, v případě silnější závislosti jeho kvalita závisí volbě  $L$  a na tom, zda platí  $H_0$  nebo  $H_1$ .

Podrobnější informace lze nalézt v článcích Antoch, Hušková a Prášková (1997), Bai (1994), Brodsky a Darkhovskiy (1993), Horváth a Qi-Man Shao (1997) či Horváth a Kokoszka (1997). V posledních dvou uvedených článcích je uvažován tzv. model s dlouhou pamětí *long range dependence model*.

## 2.5. VÍCENÁSOBNÉ ZMĚNY

Uvažujme model

$$(2.40) \quad Y_i = \mu_j + e_i, \quad [n\gamma_{j-1}] < i \leq [n\gamma_j], \quad j = 1, \dots, q+1,$$

kde  $\mu_j \in R_1$ ,  $\mu_j \neq \mu_{j+1}$ ,  $j = 1, \dots, q$ ,  $0 = \gamma_0 < \gamma_1 < \dots < \gamma_{q+1} = 1$ , a  $e_1, \dots, e_n$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, kladným rozptylem  $\sigma^2$  a konečným absolutním momentem řádu  $2 + \Delta$  pro nějaké  $\Delta > 0$ . Hodnoty  $[n\gamma_j]$ ,  $j = 1, \dots, q$ , jsou body změny. Jejich počet  $q$  může být známý nebo neznámý, v druhém případě předpokládáme, že existuje horní mez (kterou označíme  $q_0$ ) pro  $q$  je známa. V dalším textu této kapitoly též budeme používat značení  $m_j = [n\gamma_j]$ ,  $j = 1, \dots, q$ .

Pro test nulové hypotézy „nenastala žádná změna“ proti alternativní hypotéze „nastala aspoň jedna změna“ lze použít kteroukoli testovou statistiku uvažovanou v sekcích 2.1 nebo 2.3. Vyjdeme-li z principu metody podílem věrohodností, dostaneme statistiky, které mají poněkud větší sílu, jsou ale početně velice náročné.

Uvedeme si tři metody odhadu bodů změny a jejich počtu. První z nich navrhl Yao (1988). Jedná se o modifikaci tzv. Schwarzova kritéria pro odhad dimenze regrese, která umožňuje odhadnout parametry v našem případě. Přesněji řečeno, Yao definoval odhad  $\hat{q}$  jako řešení minimalizační úlohy

$$\min \left\{ \frac{1}{2} n \log \tilde{\sigma}_q^2 + q \log n; \quad q = 1, \dots, q_0 \right\},$$



kde

$$\tilde{\sigma}_q^2 = \min \left\{ \sum_{j=1}^q \sum_{i=n_{j-1}+1}^{n_j} (Y_i - \mu_j)^2 \mid \mu_j \in R_1, j = 0, \dots, q, n_0 = 0 < n_1 < \dots < n_q < n_{q+1} = n \right\}.$$

Odhady parametrů  $\mu_j$  a  $n_j$  pak získáme jako řešení minimalizační úlohy

$$\min \left\{ \sum_{j=1}^{\hat{q}} \sum_{i=n_{j-1}+1}^{n_j} (Y_i - \mu_j)^2 \mid \mu_j \in R_1, j = 0, \dots, \hat{q}, n_0 = 0 < n_1 < \dots < n_{\hat{q}} < n_{\hat{q}+1} = n \right\}.$$

Náročnost výpočtu je zřejmá. Mají-li pozorování normální rozdělení, pak jsou takto získané odhady konzistentní a shodují se odhady metodou maximální věrohodnosti. Pokud mají pozorování jiné rozdělení, je nutné kritérium modifikovat. Tento problém se těší v současné době značné pozornosti, pro další informace, viz Chung-Bow Lee (1995a, 1995b).

Další dvě metody vycházejí z odhadů navržených pro případ jedné změny. Přímým výpočtem lze ověřit, že v modelu (2.40) pro  $[n\gamma_{v-1}] < k \leq [n\gamma_v]$ ,  $v = 1, \dots, q+1$  platí

$$(2.41) \quad ES_{k,LS} = \sum_{j=1}^v ([n\gamma_j] - [n\gamma_{j-1}])(\mu_j - \bar{\mu}) + (k - [n\gamma_v])(\mu_v - \bar{\mu}),$$

zatímco pro  $j = 1, \dots, q$

$$(2.42) \quad E(S_{k+G,LS} - 2S_{k,LS} + S_{k-G,LS}) = \begin{cases} 0, & |k - [n\gamma_j]| > G, \\ (\mu_j - \mu_{j-1})([n\gamma_j] - k + G), & [n\gamma_j] - G \leq k \leq [n\gamma_j], \\ (\mu_j - \mu_{j-1})(k - [n\gamma_j] + G), & [n\gamma_j] \leq k \leq [n\gamma_j] + G. \end{cases}$$

Lokální extrémy v obou případech jsou dosaženy pro  $k = [n\gamma_j]$ ,  $j = 1, \dots, q$ . Toto vede k následujícím dvěma typům odhadů.

První typ odhadů získáme následovně. Nejprve určíme

$$(2.43) \quad \tilde{m}_{(1)} = \arg \max \left\{ \sqrt{\frac{n}{k(n-k)}} |S_{k,LS}|; k = 1, \dots, n-1 \right\}.$$

Potom rozdělíme pozorování do dvou skupin  $Y_1, \dots, Y_{\tilde{m}_{(1)}}$  a  $Y_{\tilde{m}_{(1)}+1}, \dots, Y_n$  a nalezneme odhad v každé skupině podle (2.43). Celý postup pak opakujeme tak dlouho, až testy budou indikovat, že v dané skupině nedošlo ke změně střední hodnoty, přesněji řečeno, po každém kroku testujeme hypotézu zda v nově získané skupině došlo ke změně. Aplikujeme test založený na

$$\max_{k \in D} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\hat{\sigma}_n} \left| \sum_{i \in D} \left( Y_i - \frac{1}{\#D} \sum_{j \in D} Y_j \right) \right| \right\},$$

kde  $D$  označuje indexy odpovídající sledované skupině,  $\#D$  je její kardinalita a  $\hat{\sigma}_n^2$  je vhodný odhad  $\sigma^2$ . Kritické hodnoty získáme z věty 2.1, přičemž hladinu  $\alpha_n$  zvolíme tak, aby  $\lim_{n \rightarrow \infty} \alpha_n = 0$ . Odhady získané tímto postupem, který navrhla Vostrikova (1981), jsou konzistentní.

Vztahy (2.42) vedou k následující metodě odhadu. Nalezneme všechny dvojice indexů  $v_j, w_j$  takových, že  $w_j - v_j \geq G/2$  a takových, že pro  $k = v_j, \dots, w_j$

$$\begin{aligned} |S_{k+G,LS} - 2S_{k,LS} + S_{k-G,LS}| &\leq D(n, G, \alpha_n) \\ |S_{v_j-1+G,LS} - 2S_{v_j-1,LS} + S_{v_j-1-G,LS}| &\geq D(n, G, \alpha_n) \\ |S_{w_j+1+G,LS} - 2S_{w_j+1,LS} + S_{w_j+1-G,LS}| &\geq D(n, G, \alpha_n) \end{aligned}$$

kde

$$\begin{aligned} D(n, G, \alpha_n) &= \sigma_n \sqrt{2G} (2 \log(n/G) + (1/2) \log \log(n/G) - \\ &\quad - (1/2) \log(\pi/9) - \log \log(1 - \alpha_n)^{-1}) (\sqrt{2 \log(n/G)})^{-1} \end{aligned}$$

a  $\alpha_n \rightarrow 0$  pro  $n \rightarrow \infty$ . Toto znamená, že nalezneme všechny segmenty, kde je překročena určitá kritická hodnota. Kritická hodnota odpovídá testu hypotézy „žádná změna“ proti alternativě „změna nastala“ založené na testové statistice

$$\max_{G < k < n_G} \left\{ |S_{k+G,LS} - 2S_{k,LS} + S_{k-G,LS}| \right\}$$

s asymptotickou hladinou  $\alpha_n$ . Odhad počtu změn je  $\hat{q}$  a indexy  $k \in [v_j, w_j]$ , pro které je dosaženo maximum na množině indexů  $\{v_j, w_j\}$  jsou odhady jednotlivých bodů změn.

Druhá a třetí metody jsou poměrně jednoduché a obzvláště třetí může být též použita pro diagnostiku. Potřebujeme najít vhodný odhad  $\sigma^2$ . Je-li horní mez pro počet změn  $q_0$ , pak můžeme  $\sigma^2$  odhadnout následovně:

$$\sigma_n^2 = \frac{1}{n - q_0 - 1} \min_{1 \leq k_1 < \dots < k_{q_0} \leq n} \left\{ \sum_{i=1}^{k_1} (Y_i - \bar{Y}_{k_1})^2 + \sum_{i=k_1+1}^{k_2} (Y_i - \bar{Y}_{k_1, k_2})^2 + \dots + \sum_{i=k_{q_0}+1}^n (Y_i - \bar{Y}_{k_{q_0}, n})^2 \right\},$$

kde

$$\bar{Y}_{k,s} = \frac{1}{s - k} \sum_{i=k+1}^s Y_i.$$

Jiná možnost je použít odhad

$$\tilde{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=2}^n (Y_i - Y_{i-1})^2.$$

Zvláštní pozornost byla v literatuře věnována tzv. epidemickým alternativám, které odpovídají modelu (2.40) s  $q = 2$  and  $\mu_1 = \mu_3 \neq \mu_2$ . Další informace lze nalézt v článcích Bhattacharya a Brockwell (1976), Yao Q. (1993), Antoch a Hušková (1996) či Hušková (1995).

### 3. DETEKCE ZMĚNY V REGRESNÍM MODELU

Nejprve se budeme zabývat případem, kdy se mohou měnit jen regresní parametry. V závěru sekce se pak zmíníme o případu, kdy ke změně může dojít v regresních parametrech a/nebo měřítku. Budeme přitom uvažovat následující model:

$$(3.1) \quad Y_i = \begin{cases} \mathbf{x}_i^T \boldsymbol{\beta} + e_i, & 1 \leq k \leq m, \\ \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\delta}_n + e_i, & m < k \leq n, \end{cases}$$

kde  $1 < m \leq n$  je neznámý bod změny,  $\boldsymbol{\beta} \in R_p$  a  $\boldsymbol{\delta}_n \in R_p$  jsou parametry,  $\mathbf{x}_i \in R_p$  jsou známé regresní vektory a  $e_1, \dots, e_n$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, rozptylem  $\sigma^2 > 0$ , a  $E|e_i|^{2+\Delta} < \infty$  pro nějaké  $\Delta > 0$ .

Předpokládáme, že  $\boldsymbol{\delta}_n \neq \mathbf{0}$ , tj. aspoň jedna složka je nenulová. Je-li v (3.1)  $1 < m < n$ , pak model zachycuje změnu v regresním parametru. Mezi základní úlohy patří, podobně jako v modelu polohy, testovat

$$H_0 : m = n \quad \text{proti} \quad H_1 : m < n,$$

a jestliže platí  $H_1$ , pak odhadnout  $m$ .

Budeme používat značení

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{1,k} \\ \mathbf{X}_{k+1,n} \end{pmatrix},$$

kde

$$\mathbf{X}_{1,k} = \begin{pmatrix} \mathbf{x}_1^T \\ \dots \\ \mathbf{x}_k^T \end{pmatrix}, \mathbf{X}_{k+1,n} = \begin{pmatrix} \mathbf{x}_{k+1}^T \\ \dots \\ \mathbf{x}_n^T \end{pmatrix}, \mathbf{Y}_{1,k} = \begin{pmatrix} Y_1 \\ \dots \\ Y_k \end{pmatrix}, \mathbf{Y}_{k+1,n} = \begin{pmatrix} Y_{k+1} \\ \dots \\ Y_n \end{pmatrix}.$$

Dále označíme pro  $p \leq k \leq n - p$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{1,k} &= (\mathbf{X}_{1,k}^T \mathbf{X}_{1,k})^{-1} \mathbf{X}_{1,k}^T \mathbf{Y}_{1,k}, \\ \hat{\boldsymbol{\beta}}_{k+1,n} &= (\mathbf{X}_{k+1,n}^T \mathbf{X}_{k+1,n})^{-1} \mathbf{X}_{k+1,n}^T \mathbf{Y}_{k+1,n}, \\ \hat{\sigma}_{1,k}^2 &= \frac{1}{k} \|\mathbf{Y}_{1,k} - \mathbf{X}_{1,k} \hat{\boldsymbol{\beta}}_{1,k}\|^2, \\ \hat{\sigma}_{k+1,n}^2 &= \frac{1}{n-k} \|\mathbf{Y}_{k+1,n} - \mathbf{X}_{k+1,n} \hat{\boldsymbol{\beta}}_{k+1,n}\|^2. \end{aligned}$$

Je zřejmé, že  $\hat{\boldsymbol{\beta}}_{1,k}$  a  $\hat{\boldsymbol{\beta}}_{k+1,n}$  jsou odhady metodou nejmenších čtverců regresních parametrů založené na prvních  $k$ , resp. posledních  $n - k$ , pozorováních. Dále  $\hat{\sigma}_{1,k}^2$  a  $\hat{\sigma}_{k+1,n}^2$  jsou odhady rozptylu založené na prvních  $k$ , resp. posledních  $n - k$ , pozorováních.

Při konstrukci testů a odhadů v uvažovaném regresním modelu lze vycházet ze stejných principů jako v modelu polohy (2.1). Získáme tak celou řadu testů a odhadů. Jistou komplikaci působí matice plánu experimentu  $\mathbf{X}_{1,n}$ , neboť limitní rozdělení testových statistik i odhadů je známo pouze za poměrně silných předpokladů kladených na  $\mathbf{X}_{1,n}$ .

Vyjdeme-li z předpokladu, že  $Y_i$  mají normální rozdělení a rozptyl  $\sigma^2$  neznáme, dostaneme metodou podílem věrohodností testovou statistiku

$$(3.2) \quad V_{n0} = \max_{p < k < n-p} -n \log \left( \frac{\frac{k}{n} \hat{\sigma}_{1,k}^2 + \frac{n-k}{n} \hat{\sigma}_{k+1,n}^2}{\hat{\sigma}_{1,n}^2} \right).$$

Další testové statistiky jsou založeny na vektorech částečných součtů

$$(3.3) \quad \mathbf{S}_k = (S_{k1}, \dots, S_{kp})^T = \sum_{i=1}^k \mathbf{x}_i (Y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{1,n}), \quad k = 1, \dots, n$$

nebo částečných součtech reziduálů

$$(3.4) \quad S_{k0} = \sum_{i=1}^k \left( Y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_{1,n} \right), \quad k = 1, \dots, n.$$

Testové statistiky pak můžeme zkonstruovat analogicky jako v jednorozměrném případě. Dostaneme

$$(3.5) \quad V_n = \max_{p \leq k \leq n} \left\{ \frac{1}{\widehat{\sigma}_n} \sqrt{\frac{n}{k(n-k)} \mathbf{S}_k^T (\mathbf{X}_{1,n}^T \mathbf{X}_{1,n})^{-1} \mathbf{S}_k} \right\},$$

$$(3.6) \quad V_{n1}(q) = \max_{p \leq k \leq n} \left\{ \frac{1}{\widehat{\sigma}_n} \frac{\sqrt{\mathbf{S}_k^T (\mathbf{X}_{1,n}^T \mathbf{X}_{1,n})^{-1} \mathbf{S}_k}}{q(k/n)} \right\},$$

$$(3.7) \quad V_{n2}(G) = \max_{G < k < n} \left\{ \frac{1}{\widehat{\sigma}_n} \frac{1}{\sqrt{G}} \sqrt{(\mathbf{S}_k - \mathbf{S}_{k-G})^T (\mathbf{X}_{1,n}^T \mathbf{X}_{1,n})^{-1} (\mathbf{S}_k - \mathbf{S}_{k-G})} \right\},$$

$$(3.8) \quad V_{n3}(G) = \max_{G < k < n} \left\{ \frac{1}{\widehat{\sigma}_n} \frac{1}{\sqrt{2G}} \sqrt{\mathbf{S}_{kG}^T (\mathbf{X}_{1,n}^T \mathbf{X}_{1,n})^{-1} \mathbf{S}_{kG}} \right\},$$

kde  $\mathbf{S}_{kG} = \mathbf{S}_{k+G} - 2\mathbf{S}_k + \mathbf{S}_{k-G}$ ,  $q(\cdot)$  je váhová funkce uvažovaná v sekci 2 a  $\widehat{\sigma}_n^2$  je vhodný odhad  $\sigma^2$ . Například lze použít

$$\widehat{\sigma}_n^2 = \frac{1}{n-p} \min_{1 \leq k \leq n} \left\{ k\widehat{\sigma}_k^2 + (n-k)\widehat{\sigma}_{n-k}^2 \right\}.$$

Testy založené na částečných součtech  $S_{k0}$  definovaných (3.4) dostaneme, jestliže v (2.5)–(2.9) nahradíme  $S_k$  statistikou  $S_{k0}$ .

Položme

$$b_p(x) = 2 \log x + \frac{p}{2} \log \log n - \log \Gamma\left(\frac{p}{2}\right).$$

Předpokládejme, že

$$(3.9) \quad \lim_{n \rightarrow \infty} \max \left\{ \frac{(\log(j-i))^\kappa}{j-i} \left\| \mathbf{X}_{i,j}^T \mathbf{X}_{i,j} - \mathbf{Q}(j-i) \right\| \mid \right. \\ \left. 1 < i < j \leq n, |j-i| \geq b_n \right\} = 0,$$

kde  $\mathbf{Q}$  je pozitivně definitní matice,  $\kappa > 0$ ,  $\{b_n\}$  je libovolná posloupnost s vlastností  $1 \leq b_n \leq n$  a  $b_n \rightarrow \infty$  pro  $n \rightarrow \infty$ .

Nyní zformulujeme základní větu o limitním chování testových statistik při platnosti hypotézy  $H_0$ . Jde v podstatě o zobecnění věty 2.1.

**Věta 3.1.** *Nechť pozorování  $Y_1, \dots, Y_n$  vyhovují modelu (3.1) s  $m = n$ , tj. platí  $H_0$ , a necht' platí (3.9).*

(i) *Potom pro  $\forall y \in R_1$  platí*

$$\lim_{n \rightarrow \infty} P\left(a(\log n) V_{n0} \leq (b_p(\log n) + y)\right) = \exp\{-2 \exp\{-y\}\},$$

$$\lim_{n \rightarrow \infty} P\left(a(\log n) V_n \leq (b_p(\log n) + y)\right) = \exp\{-2 \exp\{-y\}\}.$$

(ii) *Jestliže navíc  $q \in Q_{0,1}$  a  $I(q, c) < \infty$  pro nějaké  $c > 0$ , pak  $\forall x \in R_1$*

$$\lim_{n \rightarrow \infty} P\left(V_{n1}(q) \leq x\right) = P\left(\sup_{0 < t < 1} \frac{\|\mathbf{B}(t)\|}{q(t)} \leq x\right),$$

kde  $\{\mathbf{B}(t) = (B_1(t), \dots, B_p(t))^T; t \in (0, 1)\}$ ,  $\{B_j(t); t \in (0, 1)\}$  jsou nezávislé Brownovy můstky.

(iii) *Jestliže platí (2.19), pak  $\forall y \in R_1$  platí*

$$\lim_{n \rightarrow \infty} P\left(a(n/G) V_{n2}(G) \leq (b_p(n/G) + y)\right) = \exp\{-2 \exp\{-y\}\},$$

a

$$\lim_{n \rightarrow \infty} P\left(a(n/G) V_{n3}(G) \leq (b_p(n/G) + y \log(3/2))\right) = \exp\{-2 \exp\{-y\}\}.$$

*Důkaz.* Tvrzení (i) a (ii) lze nalézt v článku Horváth (1995). Důkaz tvrzení (iii) je v práci Steinebach a Eastwood (1996). □

Předpoklady kladené na matici plánu experimentu  $\mathbf{X}_{1,n}$  jsou velice silné. Některá tvrzení ve větě lze ukázat i za slabších předpokladů. Informace lze nalézt např. v pracích Hušková (1997), Kim a Siegmund (1989) či Ploberger and Krämer (1992).

Limitní rozdělení uvažovaných testových statistik při platnosti hypotézy  $H_0$  nezávisí ani na rozdělení náhodných chyb  $e_1, \dots, e_n$  ani na matici plánu experimentu  $\mathbf{X}_{1,n}$ . Hodnoty distribučních funkcí statistik  $V_{n0}, V_{n2}, V_{n3}(G)$  a  $V_n$  mohou být snadno určeny použitím kapesní kalkulačky, konvergence je však značně pomalá a prakticky pouze informativní.

Základem výše uvedené věty je, že za určitých předpokladů kladených na matici  $\mathbf{X}_{1,n}$  (předpoklady mohou být i slabší než (3.9)) při platnosti  $H_0$  proces

$$\mathbf{V}_n(t) = \frac{1}{\sigma} (\mathbf{X}_n^T \mathbf{X}_n)^{-1/2} \sum_{i=1}^{[nt]} \mathbf{x}_i e_i, \quad t \in (0, 1),$$

konverguje v  $\mathcal{D}(0, 1)$  k procesu  $\mathbf{W}(t) = (W_1(t), \dots, W_p(t))$ ,  $t \in (0, 1)$ , kde  $\{W_j(t); t \in (0, 1)\}$ ,  $j = 1, \dots, p$ , jsou nezávislé standardní Wienerovy procesy.

Problém změny regresní závislosti je jedním z často uvažovaných problémů v ekonometrii. V této oblasti jsou studovány hlavně testy založené na částečných součtech reziduálů (3.4), přičemž náhodné chyby  $e_1, \dots, e_n$  mohou být i závislé. Další informace lze nalézt např. v Krämer and Ploberger (1992), Jandhyala a Mac Neill (1989, 1991, 1992), Mac Neill (1978) či Michels a Trenkler (1990). Mac Neill et al (1990) se zabývali též Bayesovskými statistikami pro detekci změny v regresi.

V závěru sekce se zmíníme o případě, kdy změna může nastat v regresních parametrech a/nebo v měřítku. Budeme uvažovat následující model:

$$(3.10) \quad Y_i = \begin{cases} \mathbf{x}_i^T \boldsymbol{\beta} + e_i, & 1 \leq k \leq m \\ \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\delta}_n + e_i, & m < k \leq n, \end{cases}$$

kde  $1 < m \leq n$  je neznámý bod změny,  $\boldsymbol{\beta}, \boldsymbol{\delta}_n$  jsou parametry,  $\mathbf{x}_i \in R_p$  jsou známé regresní vektory a

$$e_i = \begin{cases} \sigma \epsilon_i, & 1 \leq k \leq m \\ (\sigma + h_n) \epsilon_i, & m < k \leq n, \end{cases}$$

kde  $\epsilon_1, \dots, \epsilon_n$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, jednotkovým rozptylem,  $E|\epsilon|^\nu < \infty$  pro nějaké  $\nu > 4$ ,  $\sigma > 0$ ,  $h_n$  jsou parametry.

Předpokládáme, že  $(\boldsymbol{\delta}_n, h_n) \neq \mathbf{0}$ , tj. aspoň jedna složka je nenulová. Je-li  $1 < m < n$  v uvažovaném modelu, pak model zachycuje změnu v regresním parametru a/nebo v měřítku.

Při konstrukci testů a odhadů obvykle se vychází z metody podílem věrohodností (pro testy), resp. z metody maximální věrohodnosti (pro odhady), a z předpokladu, že jednotlivá pozorování mají normální rozdělení. Přímý (ale dlouhý) výpočet vede k testové statistice

$$Z_n = \max_{p \leq k \leq n-p} -2 \log \boldsymbol{\Lambda}_k,$$

kde

$$\boldsymbol{\Lambda}_k = \frac{\widehat{\sigma}_{1,k}^k \widehat{\sigma}_{k+1,n}^{n-k}}{\widehat{\sigma}_{1,n}^n}, \quad p \leq k \leq n-p.$$

**Věta 3.2.** Jsou-li splněny předpoklady věty 3.1, pak

$$\lim_{n \rightarrow \infty} P\left(a(\log n)Z_n^{1/2} \leq t + b_{p+1}(\log n)\right) = \exp\{-2e^{-t}\} \quad t \in R_1.$$

*Důkaz.* Lze nalézt v Horváth (1995). □

**Věta 3.3.** Necht'  $Y_1, \dots, Y_n$  vyhovují modelu (3.10), necht' platí (3.9) a

$$\begin{aligned} \|\delta_n\| &\rightarrow 0, & n\|\delta_n\|^2 / \log \log n &\rightarrow \infty, \\ h_n &\rightarrow 0, & nh_n^2 / \log \log n &\rightarrow \infty. \end{aligned}$$

Pak

$$\frac{\left(\frac{1}{\sigma^2} \delta_n^T \mathbf{Q} \delta_n + \frac{1}{2} \left(\frac{h_n}{\sigma^2}\right)^2\right)^2}{\frac{1}{\sigma^2} \delta_n^T \mathbf{Q} \delta_n + \frac{h_n}{\sigma^3} \bar{\mathbf{x}}_n \delta_n m_3 + \left(\frac{h_n}{2\sigma^2}\right)^2 (m_4 - 1)} (\hat{m} - m) \xrightarrow{\mathcal{D}} V_{1/2,1/2}.$$

kde  $\bar{\mathbf{x}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ ,  $m_j = E\epsilon^j$ ,  $j = 2, 3$ .

*Důkaz.* Lze nalézt v Horváth, Hušková a Serbinowska (1997). □

Věta pokrývá i případ, kdy je buď  $\delta_n = \mathbf{0}$  nebo  $h_n = 0$ . Tvzení zůstává v platnosti, jestliže  $\delta_n$  a  $h_n$  nahradíme vhodnými odhady, což umožňuje zkonstruovat aproximaci pro intervalový odhad bodu změny  $m$ .

#### 4. INTERVALY SPOLEHLIVOSTI A VYBRANÉ KRITICKÉ HODNOTY

V tomto odstavci se soustředíme na testy popsané v sekci 2.1 a odhady popsané v sekci 2.2, ale většinu postupů je možné použít po přirozené modifikaci i pro situace posané v dalších částech článku.

##### 4.1. JAK KONSTRUOVAT INTERVALY SPOLEHLIVOSTI

Pokud se týče konstrukce intervalových odhadů bodu změny  $m$ , záleží na tom zda uvažujeme lokální změnu nebo pevnou změnu. V případě tzv. pevné změny totiž limitní rozdělení závisí na distribuční funkci původních pozorování, kterou obvykle neznáme. Z věty 2.3 a poznámky 2.6 dostaneme následující  $100(1 - \alpha)\%$  intervalový odhad pro bod změny  $m$ .



$$(3.11) \quad \left( \hat{m}(\eta) - \hat{v}_{\eta,\gamma}(\alpha/2) \frac{\hat{\sigma}_n^2}{\hat{\delta}_n^2}, \hat{m}(\eta) + \hat{v}_{\eta,\gamma}(1 - \alpha/2) \frac{\hat{\sigma}_n^2}{\hat{\delta}_n^2} \right),$$

kde  $\hat{v}_{\eta,\gamma}(\alpha)$  je  $100(1 - \alpha)\%$ ní kvantil náhodné veličiny  $V_{\eta,\gamma}$ , přičemž za  $\eta$  dosadíme jeho odhad.

Zmiňme ještě možnost využití bootstrapu pro konstrukci přibližných intervalů spolehlivosti. Popíšeme přitom dvě základní bootstrapové schemata, přičemž první z nich je použitelné právě pro model (2.1), zatímco druhé lze použít i pro modely, kdy pozorování před změnou a po změně mají rozdělení  $F_1$ , resp.  $F_2$ ,  $F_1 \neq F_2$ . Využijeme přitom odhadů  $m$  zavedených v sekci 2, přičemž označíme pomocí  $\hat{m}$  libovolný z těchto odhadů.

### Bootstrapové výběrové schéma I.

Nalezneme odhad  $\hat{m}$  bodu změny  $m$  (přičemž lze použít kterýkoli z odhadů zavedených v sekci 2) a spočteme jak odhadnuté reziduály

$$\tilde{E}_i = \begin{cases} Y_i - \bar{Y}_{\hat{m}}, & i = 1, \dots, \hat{m}, \\ Y_i - \bar{Y}_{\hat{m}}^0, & i = \hat{m} + 1, \dots, n, \end{cases}$$

tak centrované reziduály

$$\hat{E}_i = \tilde{E}_i - \frac{1}{n} \sum_{i=1}^n \tilde{E}_i, \quad i = 1, \dots, n.$$

Potom realizujeme bootstrapový výběr  $E_1^*, \dots, E_n^*$  odpovídající empirické distribuční funkci veličin  $\hat{E}_1, \dots, \hat{E}_n$ , a konstruujeme bootstrapová pozorování

$$Y_i^* = \begin{cases} \bar{Y}_{\hat{m}} + E_i^*, & i = 1, \dots, \hat{m}, \\ \bar{Y}_{\hat{m}}^0 + E_i^*, & i = \hat{m} + 1, \dots, n. \end{cases}$$

Z hodnot  $Y_1^*, \dots, Y_n^*$  spočteme bootstrapový odhad  $\hat{m}^*$  (při výpočtu  $\hat{m}$  nahradíme původní pozorování  $Y_1, \dots, Y_n$  jejich bootstrapovými verzemi).

### Bootstrapové výběrové schéma II.

Nejprve nalezneme odhad  $\hat{m}$  bodu změny  $m$ , přičemž lze použít kterýkoli z odhadů uvedených v sekci 2.3. Dále realizujeme bootstrapové výběry, přesněji řečeno realizujeme nezávislé náhodné výběry  $Y_1^{**}, \dots, Y_{\hat{m}}^{**}$  a  $Y_{\hat{m}+1}^{**}, \dots, Y_n^{**}$  odpovídající empirickým distribučním funkcím založeným na hodnotách  $Y_1, \dots, Y_{\hat{m}}$ , resp.  $Y_{\hat{m}+1}, \dots, Y_n$ . Z hodnot  $Y_1^{**}, \dots, Y_n^{**}$  spočteme (bootstrapový) odhad  $\hat{m}^{**}$ .

Lze ukázat, že (bootstrapová) distribuční funkce odhadu  $\hat{m}^*$  (popř.  $\hat{m}^* - \hat{m}$ ) je konzistentním odhadem distribuční funkce odhadu  $\hat{m}$  (popř.  $\hat{m} - m$ ), takže můžeme bootstrapovou distribuční funkci použít jako aproximaci skutečné distribuční funkce odhadu  $\hat{m}$  (popř.  $\hat{m} - m$ ). Odtud můžeme získat i aproximaci pro kvantily a následně pro intervaly spolehlivosti. Další podrobnosti lze nalézt např. v Hinkley a Schechtman (1987), Dümbgen (1991), Antoch Hušková a Veraverbeke (1995) či Antoch a Hušková (1998).

#### 4.2. JAK POČÍTAT VYBRANÉ KRITICKÉ HODNOTY

Nalézt aproximace pro kritické hodnoty testů popsaných v předchozí části článku lze několika způsoby.

- (1) Můžeme využít limitního rozdělení testových statistik (za platnosti nulové hypotézy), jsou-li dostupné hodnoty kvantilů limitního rozdělení.
- (2) Použití Bonferroniho nerovnosti a limitního rozdělení, což vede ke konzervativnímu testu.
- (3) Simulovat limitní rozdělení, neboť toto rozdělení nezávisí na rozdělení původních náhodných veličin.

Jak je zřejmé z věty 2.1, spočítat limitní kritické hodnoty pro statistiky  $T_n$ ,  $T_{n2}(G)$  a  $T_{n3}(G)$  není žádný problém a lze k tomu užít obyčejnou kapesní kalkulačku. Naproti tomu s výpočtem kritických hodnot pro ostatní statistiky již problémy jsou, mnohdy více než značné.

Jak bylo uvedeno v odstavci 2.1, statistiky  $T_{n0}(\varepsilon)$ ,  $T_{n1}(q)$  a  $T_{n4}(r)$  mají v limitě rozdělení jako funkcionály Brownova můstku. Aproximace těchto limitních rozdělení jsou (pokud víme) známy pouze pro některé speciální volby váhových funkcí, jež uvádíme dále.

##### Statistika $T_{n0}(\varepsilon)$

- James et al. (1987) pro velké hodnoty  $x$  odvodili následující aproximaci

$$\begin{aligned} \lim_{n \rightarrow \infty} P(T_{n0}(\varepsilon) > x) &= P\left(\sup_{\varepsilon < t < 1-\varepsilon} \frac{|B(t)|}{\sqrt{t(1-t)}} > x\right) \\ &\approx x e^{-x^2/2} \sqrt{2/\pi} \left( \left(1 - \frac{1}{x^2}\right) \log \frac{1-\varepsilon}{\varepsilon} + \frac{2}{x^2} \right). \end{aligned}$$

- James et al. (1987) navrhli v případě normálně rozdělených veličin  $Y_i$  pro velké hodnoty  $x$  následující aproximaci

$$P\left(\max_{n\varepsilon \leq k \leq n(1-\varepsilon)} T_{n0}(\varepsilon) > x\right) \approx \sqrt{\frac{2n}{\pi}} \int_{x/\sqrt{n}}^1 (1-y^2)^{n/2-2} dy$$

$$+ \frac{x\sqrt{2}}{\sqrt{\pi}} \left(1 - \frac{x^2}{n}\right)^{n/2-2} \int_{x\sqrt{\varepsilon/(n-x^2)(1-\varepsilon)}}^{x\sqrt{(1-\varepsilon)/((n-x^2)\varepsilon)}} \frac{1}{y} \nu\left(y + \frac{x^2}{(n-x^2)y}\right) dy,$$

kde

$$\nu(y) = \frac{2}{y^2} \exp\left\{-2 \sum_{n=1}^{\infty} \frac{1}{n} \Phi\left(-\frac{1}{2}y\sqrt{n}\right)\right\} \approx \exp\{-0.583y\}, \quad y > 0.$$

- V případě, kdy můžeme  $\sigma$  považovat za známé, lze pro velké hodnoty  $x$  užít následující aproximaci

$$P\left(\max_{n\varepsilon \leq k \leq n(1-\varepsilon)} \sqrt{\frac{n}{k(n-k)}} \frac{|S_k|}{\sigma} > x\right) \approx 2\left(1 - \Phi(x) + x\phi(x) \int_{x\sqrt{\varepsilon/(n(1-\varepsilon))}}^{x\sqrt{(1-\varepsilon)/(n\varepsilon)}} \frac{1}{y} \nu\left(y + \frac{x^2}{ny}\right) dy\right).$$

### Statistika $T_{n1}(q)$

- Jestliže v případě statistiky  $T_{n1}(q)$  volíme  $q(t) = 1$  pro  $t \in (0, 1)$ , potom ve větě 2.1 dostaneme rozdělení absolutní hodnoty Brownova můstku (které je dobře známé z Kolmogorovova-Smirnovova testu), tj.  $\forall x \in R_1$

$$\lim_{n \rightarrow \infty} P(T_{n1}(q) > x) = P\left(\sup_{0 < t < 1} |B(t)| > x\right) = 2 \sum_{n=1}^{\infty} (-1)^{n+1} e^{-2n^2 x^2}.$$

### Statistika $T_{n4}(r)$

- Volíme-li v případě statistiky  $T_{n4}(r)$   $r(t) = 1$  pro  $t \in (0, 1)$ , potom z věty 2.1 a prací Andersona a Darlinga (1952, 1954) dostaneme  $\forall x \in R_1$

$$\lim_{n \rightarrow \infty} P(T_{n4}(r) > x) = P\left(\int_0^1 B^2(t) dt > x\right) = 1 -$$

$$-\frac{\sqrt{2}}{\pi^{3/2}\sqrt{x}} \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)}{\Gamma(j+1)} \sqrt{2j + \frac{1}{2}} \exp\left\{-\frac{(4j+1)^2}{16x}\right\} K_{1/4}\left(\frac{(4j+1)^2}{16x}\right),$$

kde  $K_{1/4}(\cdot)$  označuje modifikovanou Besselovu funkci druhého druhu (v Matlabu se jedná o funkci *besselk*( $nu, z$ ), v Mathematice o funkci *BesselK*). Tento výsledek byl odvozen v práci Andersona a Darlinga (1952), viz též Kiefer (1960), strana 423.

- Volíme-li v případě statistiky  $T_{n4}(r)$   $r(t) = 1/(t(1-t))$ ,  $t \in (0, 1)$ , potom z věty 2.1 a prací Andersona a Darlinga (1952, 1954) dostaneme  $\forall x \in R_1$

$$\begin{aligned} \lim_{n \rightarrow \infty} P(T_{n4}(r) > x) &= P\left(\int_0^1 \frac{B^2(t)}{t(1-t)} dt > x\right) \\ &= 1 - \frac{\sqrt{2}}{x} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(j+1/2)(4j+1)}{j!} \exp\left\{-\frac{(4j+1)^2 \pi^2}{8x}\right\} \\ &\quad \cdot \int_0^{\infty} \exp\left\{\frac{x}{8(w^2+1)} - \frac{(4j+1)^2 \pi^2 w^2}{8x}\right\} dw. \end{aligned}$$

Autoři přitom uvádějí, že pro dosažení přesnosti na čtyři desetinná místa stačí vzít pouze člen odpovídající volbě  $j = 0$  a provést numerickou integraci. Vybrané limitní kritické hodnoty jsou uvedeny v práci Andersona a Darlinga (1954).

### 4.3. VYBRANÉ SIMULOVANÉ KRITICKÉ HODNOTY

Tabulky 1.–8. obsahují kritické hodnoty pro statistiky  $T_n$ ,  $T_{n0}(\varepsilon)$  a  $T_{n1}(q)$ , kde  $q(\cdot) \equiv 1$ , získané pomocí simulací. Poznamenejme, že ve vzorcích (2.5)–(2.10) uvádíme testové statistiky pro situaci, kdy  $\sigma$  je neznámé a je třeba je odhadnout. Nicméně, v případě, kdy  $\sigma$  může být považováno za známé, lze použít tytéž statistiky s tím rozdílem, že v nich  $\hat{\sigma}$  nahradíme „známou“ hodnotou  $\sigma$ .

Tabulky 1.–4. obsahují simulované kritické hodnoty pro případ, kdy  $\sigma$  je známé (v simulacích pro jednoduchost zvolili  $\sigma = 1$ ). Naproti tomu tabulky 5.–8. obsahují kritické hodnoty získané pomocí simulací pro případ, kdy je  $\sigma$  neznámé a je odhadnuto pomocí (2.4). V popisu tabulek tento fakt rozlišujeme slovy nestudentizováno/studentizováno.

Jednotlivé tabulky obsahují následující informaci:

- V prvním sloupci je uveden rozsah výběru  $n$ .
- Druhý sloupec obsahuje kritické hodnoty statistiky  $T_n$  počítané dle vztahu (2.5).
- Třetí až pátý sloupec obsahuje kritické hodnoty statistiky  $T_{n0}(\varepsilon_n)$  počítané dle vztahu (2.6) pro různý stupeň useknutí  $\varepsilon$ .
- Šestý sloupec obsahuje kritické hodnoty statistiky  $T_{n1}(q)$  počítané dle vztahu (2.7) pro  $q \equiv 1$ .

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	2.709	2.709	2.639	2.558	1.140
100	2.809	2.783	2.703	2.627	1.166
200	2.892	2.855	2.763	2.682	1.185
300	2.931	2.884	2.780	2.694	1.190
500	2.973	2.916	2.804	2.714	1.198

Tab. 1. 10 % kvantil, nestudentizováno.

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	2.960	2.960	2.900	2.823	1.275
100	3.065	3.040	2.965	2.900	1.302
200	3.143	3.110	3.030	2.950	1.318
300	3.176	3.135	3.042	2.967	1.324
500	3.218	3.169	3.068	2.983	1.333

Tab. 2. 5 % kvantil, nestudentizováno.

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	3.200	3.200	3.144	3.070	1.395
100	3.294	3.275	3.203	3.146	1.425
200	3.371	3.340	3.262	3.194	1.439
300	3.410	3.370	3.286	3.212	1.448
500	3.440	3.397	3.301	3.229	1.465

Tab. 3. 2.5 % kvantil, nestudentizováno.

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	3.486	3.486	3.441	3.370	1.543
100	3.563	3.546	3.490	3.436	1.572
200	3.649	3.623	3.548	3.481	1.590
300	3.684	3.649	3.581	3.510	1.601
500	3.703	3.664	3.587	3.518	1.602

Tab. 4. 1 % kvantil, nestudentizováno.

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	2.857	2.856	2.775	2.683	1.197
100	2.891	2.864	2.778	2.694	1.194
200	2.934	2.893	2.801	2.715	1.198
300	2.961	2.914	2.805	2.718	1.198
500	2.993	2.931	2.820	2.728	1.203

Tab. 5. 10 % kvantil, studentizováno  $s_k$ .

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	3.157	3.157	3.079	2.992	1.344
100	3.164	3.139	3.061	2.984	1.339
200	3.196	3.159	3.071	2.992	1.333
300	3.213	3.172	3.076	2.994	1.337
500	3.241	3.189	3.088	3.003	1.340

Tab. 6. 5 % kvantil, studentizováno  $s_k$ .

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	3.421	3.421	3.359	3.279	1.483
100	3.402	3.383	3.311	3.248	1.467
200	3.428	3.391	3.314	3.245	1.461
300	3.452	3.409	3.324	3.246	1.463
500	3.462	3.420	3.324	3.251	1.468

Tab. 7. 2.5 % kvantil, studentizováno  $s_k$ .

$n$	$T_n$	$T_{n0}(\varepsilon)$			$T_{n1}(q)$
		$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$q \equiv 1$
50	3.747	3.747	3.695	3.625	1.653
100	3.696	3.678	3.615	3.558	1.627
200	3.719	3.690	3.611	3.558	1.611
300	3.737	3.700	3.628	3.555	1.617
500	3.735	3.700	3.617	3.547	1.614

Tab. 8. 1 % kvantil, studentizováno  $s_k$ .

## 5. PŘÍKLADY POUŽITÍ

Tato sekce ukazuje příklady použití výše popsaných metod na dvoje „učebnicová“ reálná data, tj. data o průtocích Nilu v Asuánu a teplotní řadu měřenou v Klementinu, spolu s několika málo výsledky ilustrujícími použití metody bootstrap.

Laskavý čtenář nahlédne porovnáním odstavců 2.1 a 2.2 jež popisují základní testy a odhady, že dále uvedené obrázky týkající se odhadů dávají relevantní informaci též o „průběhu“ testových statistik.

### Data o Nilu

Tato data, zobrazená na obrázku 1., znázorňující průtoky Nilu v Asuánu v letech 1871–1970, podrobnosti viz Cobb (1978) nebo Hinkley a Schechtman (1987). Na obrázcích 2.–13. lze vidět:

- vliv volby  $\eta$  na odhad  $\hat{m}(\eta)$  definovaný vztahem (2.25). Hodnoty  $\eta$  byly položeny 0, 0.25 a 0.5.

- Vliv standardizace. Přitom volba:
  - $s = 1$  odpovídá normalizaci standardní směrodatnou odchylkou spočtenou ze všech dat;
  - $s = 2$  odpovídá normalizaci podle vztahu (2.4);
  - $s = 3$  odpovídá situaci, kdy nenormalizujeme vůbec.
- Rozdíl mezi použitím klasické metody nejmenších čtverců a robustním přístupem, tj. rozdíl mezi volbou  $\psi(x) = x$  a  $\psi(x) = \text{sign}(x)$ .

Ve všech případech můžeme vidět, že odhady detekují možnou změnu okolo roku 1898. Tento výsledek je ve shodě s výsledky jiných autorů. Co se týče způsobu normalizace, je zřejmé, že použití odhadu  $\hat{\sigma}_n^2$  počítaného podle (2.4) je výhodnější než použití standardní směrodatné odchylky. Na druhé straně, výsledky neukazují žádný významný rozdíl mezi použitím odhadu založeném na metodě nejmenších čtverců oproti užití odhadů robustního typu.

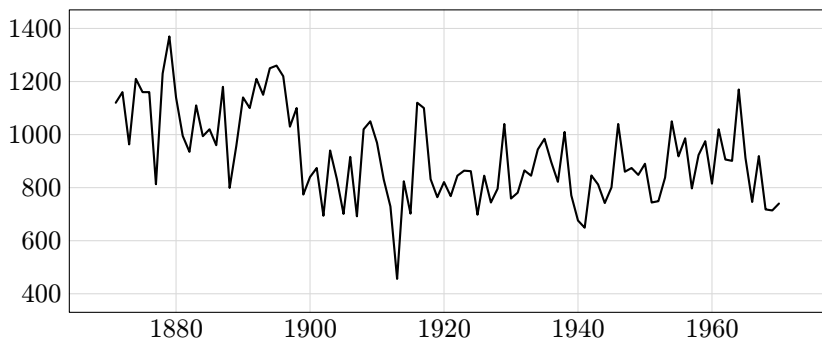
Obrázky 14.–21. ukazují chování odhadů MOSUM typu. Je z nich velmi pěkně vidět vliv dvou základních parametrů, tj. šířky okénka  $G$  a zvolené funkce  $\psi$ , na tvar testových statistik. Okamžitě je z nich zřejmé, že volba  $\psi(x) = \text{sign}(x)$  potřebuje větší hodnoty  $G$  než volba  $\psi(x) = x$  pro to, abychom získali ten samý závěr. Dále je vidět, že v případě, kdy je hodnota  $G$  dostatečně velká, metoda MOSUM též dobře indikuje změnu okolo roku 1898.

Na závěr obrázky 22.–27. ilustrují použití metody bootstrap pro stanovení konfidenčních intervalů pro  $m$ . Jedná se především o bootstrapové aproximace hustoty odhadu  $\hat{m}_{LS}(\eta)$ ,  $\eta \in \{0, 1/2\}$  a jejich rozdílů. Je z nich, mimo jiné, vidět značný rozdíl mezi volbou  $\eta = 0$  a  $\eta = 1/2$ , přičemž tato druhá volba se zdá být mnohem realističtější a doporučení hodná i pro praxi.

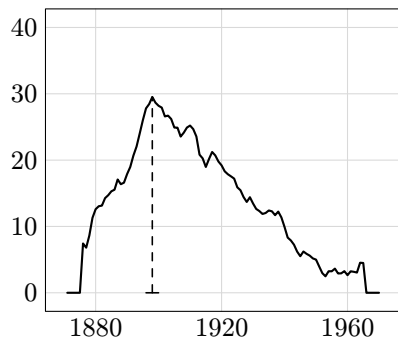
### Data z Klementina

Jako druhý příklad uvádíme teplotní řadu měřenou v Klementinu v letech 1775–1992. Teploty se měří i nyní, data nám ale bohužel nejsou díky neochotě ČHMÚ k dispozici. Podrobný rozbor této řady z pohledu detekce bodu změny lze nalézt v Jarušková (1997). Na obrázcích 29.–34. je ukázáno chování statistik typu MOSUM. Opět je dokumentován vliv volby funkce  $\psi$  a délky okénka  $G$  na výsledný odhad. Závěry jsou naprosto analogické se závěry vyslovenými v předchozím odstavci, tj. volba  $\psi(x) = \text{sign}(x)$  zpravidla potřebuje větší hodnoty  $G$  než volba  $\psi(x) = x$  pro to, abychom získali ten samý závěr. Dále je vidět, že v případě kdy je hodnota  $G$  dostatečně velká, metoda MOSUM dobře indikuje změnu okolo roku 1840. Druhá změna v dvacátých letech tohoto století již není zdaleka tak výrazná, nicméně pečlivá studie dat zobrazených na obr. 28. ukazuje pomalý postupný trend ve zvyšování průměrné roční teploty, který započal v první čtvrtině tohoto století.

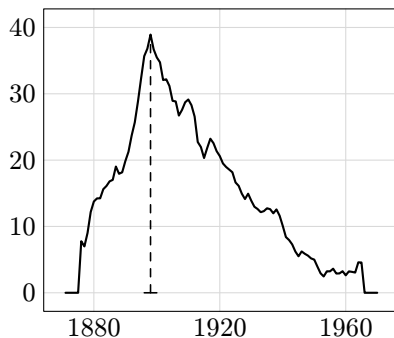




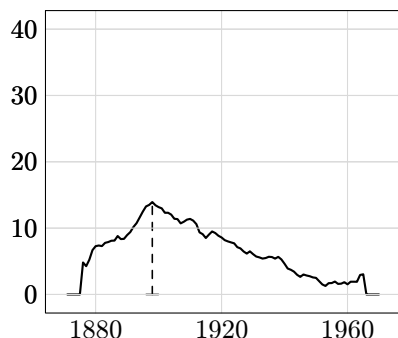
Obr. 1. Data o průtocích Nilu v Asuánu.



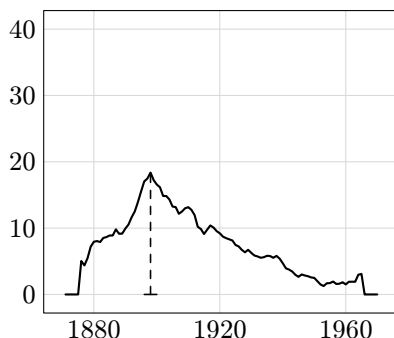
Obr. 2. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0$ ,  $s = 1$ .



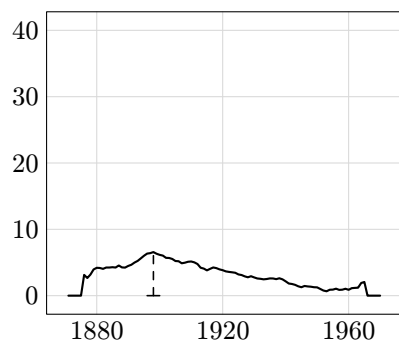
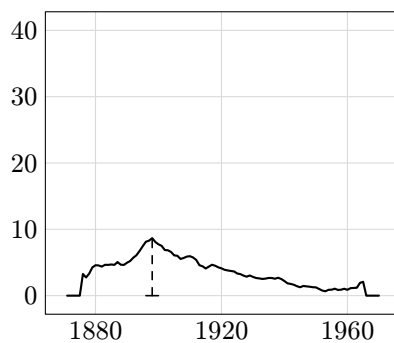
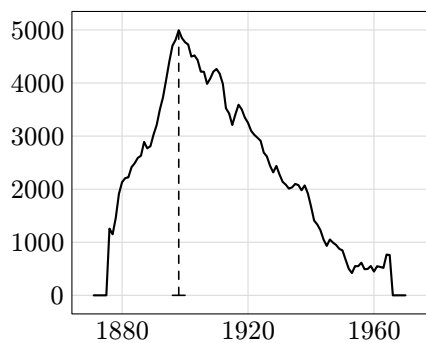
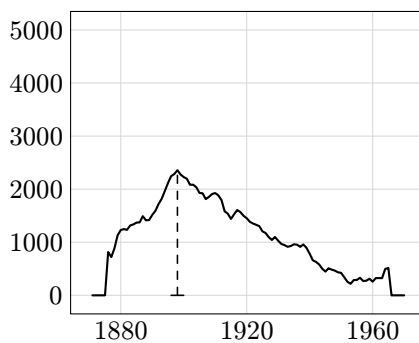
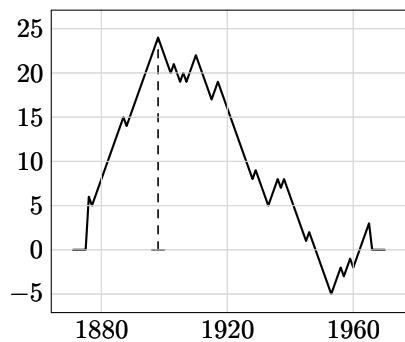
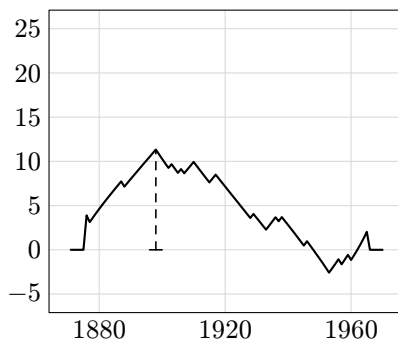
Obr. 3. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0$ ,  $s = 2$ .

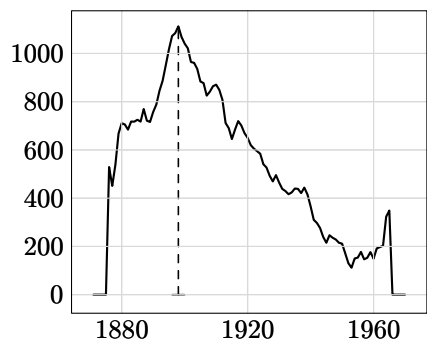


Obr. 4. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.25$ ,  $s = 1$ .

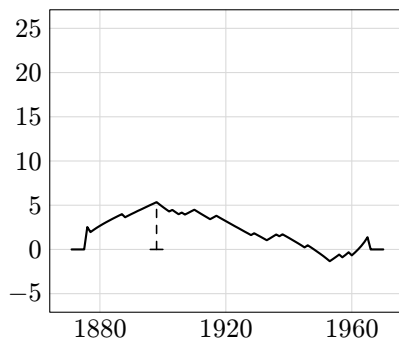


Obr. 5. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.25$ ,  $s = 2$ .

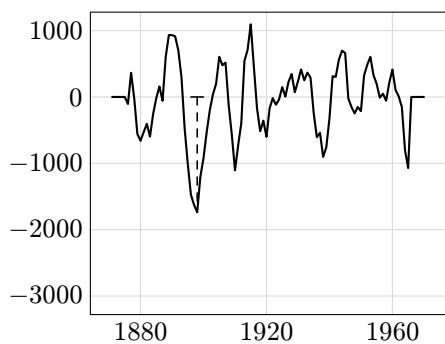
Obr. 6. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.5$ ,  $s = 1$ .Obr. 7. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.5$ ,  $s = 2$ .Obr. 8. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0$ ,  $s = 3$ .Obr. 9. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.25$ ,  $s = 3$ .Obr. 10. Robustní odhad,  $\psi(x) = \text{sign}(x)$ ,  $\eta = 0$ ,  $s = 3$ .Obr. 11. Robustní odhad,  $\psi(x) = \text{sign}(x)$ ,  $\eta = 0.25$ ,  $s = 3$ .



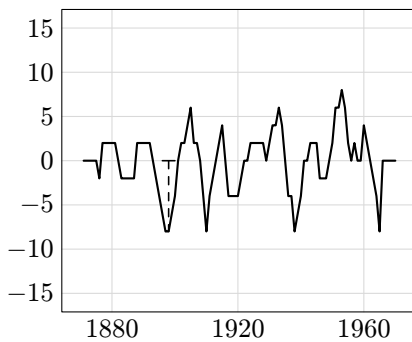
Obr. 12. Odhad  $\hat{m}(\eta)$ ,  $\eta = 0.5$ ,  $s = 3$ .



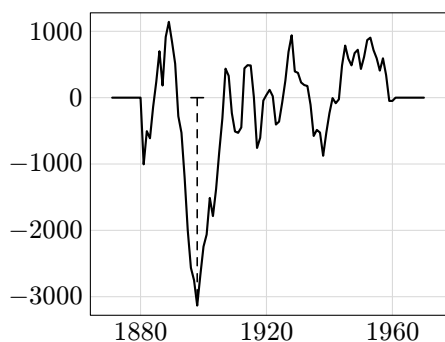
Obr. 13. Robustní odhad,  $\psi(x) = \text{sign}(x)$ ,  $\eta = 0.5$ ,  $s = 3$ .



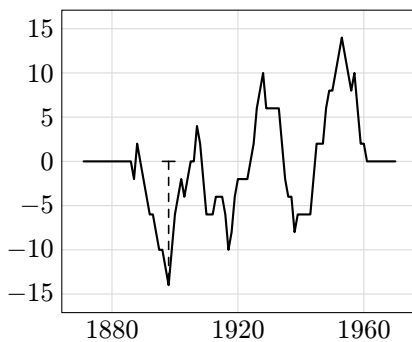
Obr. 14. MOSUM odhad,  $\psi(x) = x$ ,  $G = 5$ ,  $s = 3$ .



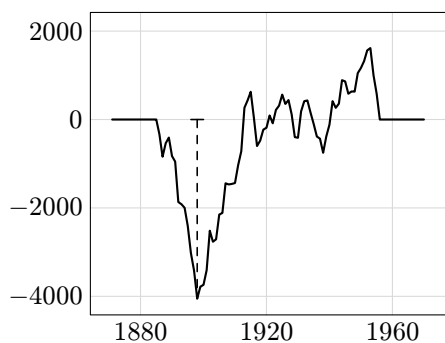
Obr. 15. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  $G = 5$ ,  $s = 3$ .



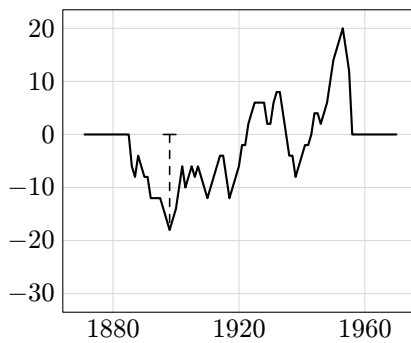
Obr. 16. MOSUM odhad,  $\psi(x) = x$ ,  $G = 10$ ,  $s = 3$ .



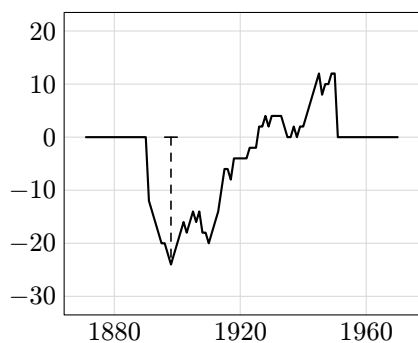
Obr. 17. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  $G = 10$ ,  $s = 3$ .



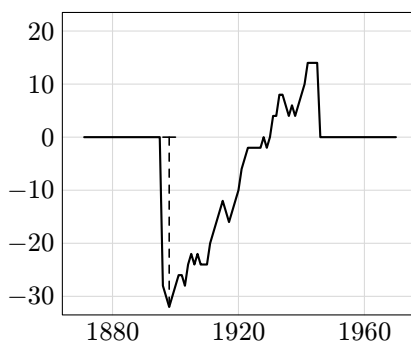
Obr. 18. MOSUM odhad,  $\psi(x) = x$ ,  
 $G = 15$ ,  $s = 3$ .



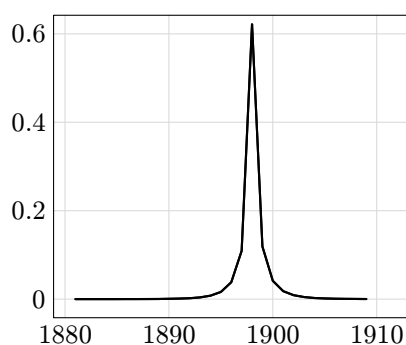
Obr. 19. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  $G = 15$ ,  $s = 3$ .



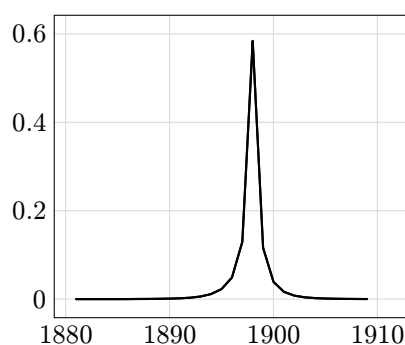
Obr. 20. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  $G = 20$ ,  $s = 3$ .



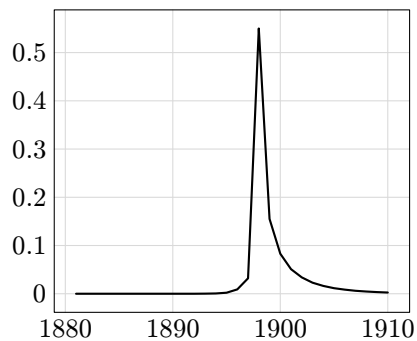
Obr. 21. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  $G = 25$ ,  $s = 3$ .



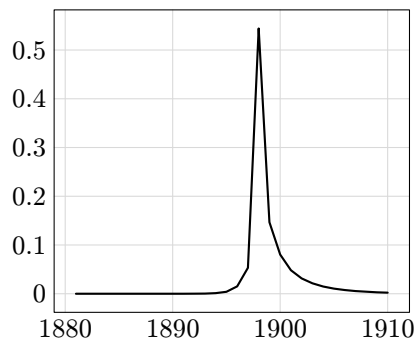
Obr. 22. Bootstrap I. aproximace  
hustoty odhadu  $\hat{m}(1/2)$  de-  
finovaného vztahem (2.24).



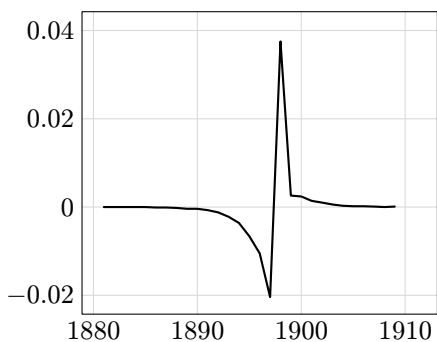
Obr. 23. Bootstrap II. aproximace  
hustoty odhadu  $\hat{m}(1/2)$  de-  
finovaného vztahem (2.24).



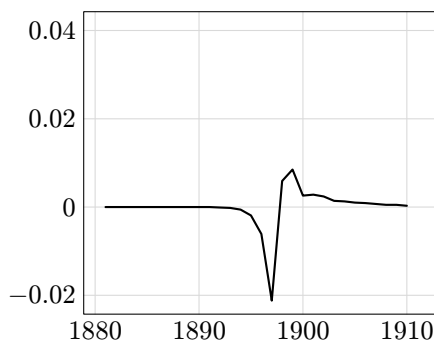
Obr. 24. Bootstrap I. aproximace hustoty odhadu  $\hat{m}(0)$  definovaného vztahem (2.24).



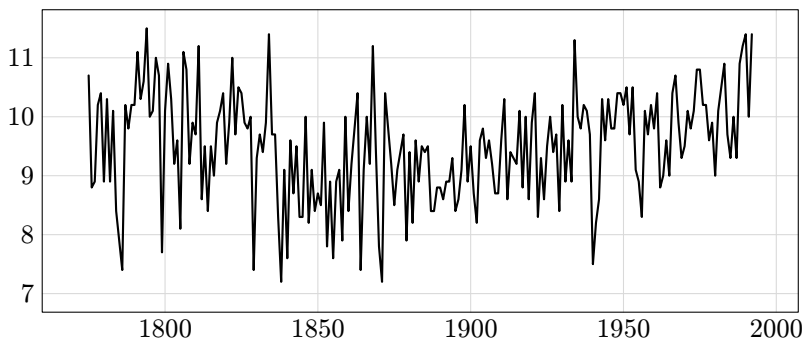
Obr. 25. Bootstrap II. aproximace hustoty odhadu  $\hat{m}(0)$  definovaného vztahem (2.24).



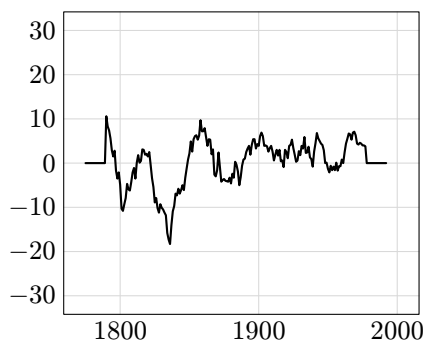
Obr. 26. Rozdíl mezi Bootstrap I. a II. aproximacemi hustoty odhadu  $\hat{m}(1/2)$ .



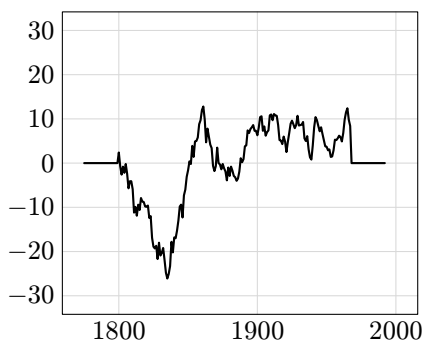
Obr. 27. Rozdíl mezi Bootstrap I. a II. aproximacemi odhadu hustoty  $\hat{m}(0)$ .



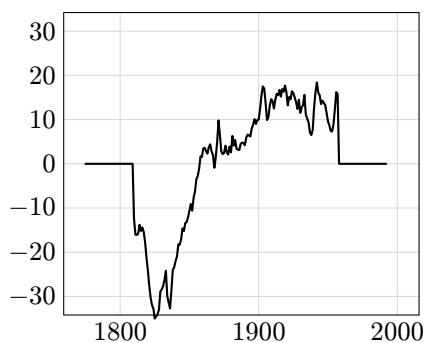
Obr. 28. Data z Klementina.



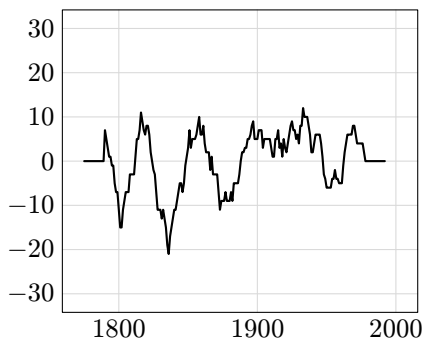
Obr. 29. MOSUM odhad,  $\psi(x) = x$ ,  
 $G = 15$ ,  $s = 3$ .



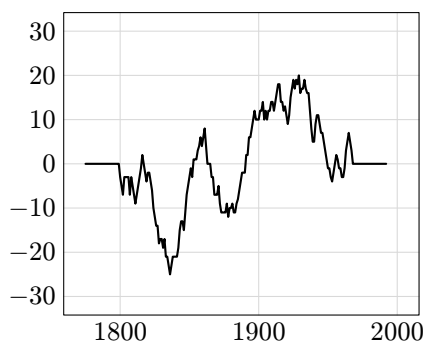
Obr. 30. MOSUM odhad,  $\psi(x) = x$ ,  
 $G = 25$ ,  $s = 3$ .



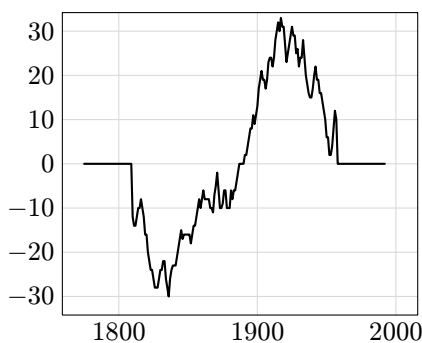
Obr. 31. MOSUM odhad,  $\psi(x) = x$ ,  
 $G = 35$ ,  $s = 3$ .



Obr. 32. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  
 $G = 15$ ,  $s = 3$ .



Obr. 33. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  
 $G = 25$ ,  $s = 3$ .



Obr. 34. MOSUM odhad,  $\psi(x) = \text{sign}(x)$ ,  
 $G = 35$ ,  $s = 3$ .

## LITERATURA

## KNIHY A SBORNÍKY VZTAHUJÍCÍ SE KE STUDOVANÉMU PŘEDMĚTU

- Basseville M. a Benveniste A. (eds.), *Detection of Abrupt Changes in Signals and Dynamic Systems*, Springer Verlag, Heidelberg, 1986.
- Basseville M. a Nikiforov I. V., *Detection of Abrupt Changes. Theory and Applications*, Prentice Hall, London, 1993.
- Broemling L. a Tsurumi H., *Econometrics and Structural Change over Time*, M. Dekker, New York, 1987.
- Brodsky B. S. a Darkhovsky B. E., *Nonparametric Methods in Change-Point Problems*, Kluwer, Dordrecht, 1993.
- Carlstein E., Müller H.-G. a Siegmund D. (eds.), *Change-Point Problem*, IMS Lecture Notes **23**, 1994.
- Csörgő M. a Horváth L., *Limit Theorems in Change-Point Analysis*, J. Wiley, New York, 1997.
- Hackl P., *Testing the Constancy of Regression Relationship over Time*, Vandenhoeck and Ruprecht, Göttingen, 1980.
- Hackl P. (ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, Springer Verlag, New York, 1989.
- Hackl P. a Westlund A. H. (eds.), *Economic Structural Changes. Analysis and Forecastings*, IIASA & Springer Verlag, Heidelberg, 1991.

## ČLÁNKY V ČASOPISECH A SBORNÍCÍCH. POMOCNÉ MONOGRAFIE

- Anderson T. W. and Darling D. A., *Asymptotic theory of certain goodness of fit criteria based on stochastic processes*, Annals of Mathematical Statistics **23** (1952), 193–212.
- Anderson T. W. and Darling D. A., *A test of goodness of fit*, J. of American Statistical Association **49** (1954), 765–769.
- Antoch J. a Hušková M., *Change-point problem*, Computational Aspects of Model Choice (Antoch J., ed.), Physica-Verlag, Heidelberg, 1993, pp. 11–38.
- Antoch J. a Hušková M., *Procedures for detection of multiple changes in series of independent observations*, Proc. of 5th Prague Symp. on Asymptotic Statistics (Hušková M. a Mandl P., eds.), Physica-Verlag, Heidelberg, 1994, pp. 3–20.
- Antoch J. a Hušková M., *Tests and estimators for epidemic alternatives*, Tatra Mountains Math. Publ. **7** (1996), 316–329.
- Antoch J. a Hušková M., *Estimators of changes*, Nonparametrics, Asymptotics and Time Series, A tribute to Madan Lai Puri (Ghosh S., ed.), M. Dekker, New York, 1998.
- Antoch J., Hušková M. a Prášková Z., *Effect of dependence on statistics for determination of change*, J. Stat. Plan. Infer. **60** (1997), 291–310.
- Antoch J., Hušková M. a Veraverbeke N., *Change-point problem and bootstrap*, J. of Nonparametric Statistics **5** (1995), 123–144.
- Bai J., *Least squares estimation of a shift in linear processes*, J. Time Series Analysis **15** (1994), 453–472.
- Bhattacharya P. K. a Brockwell P. J., *The minimum of additive process with applications to signal estimation and storage theory*, Z. Wahrsch. Verw. Gebite **37** (1976), 51–75.
- Brown R. L., Durbin J. a Evans J. M., *Techniques for testing the constancy of regression relationships over time (with discussion)*, JRSS **B 37** (1975), 149–192.
- Carlstein E., *Nonparametric estimation of a change-point*, Ann. Statist. **16** (1988), 188–197.

- Chen X., *Inference in a simple change-point problem*, *Scientia Sinica* **A 31** (1988), 654–667.
- Cobb G. W., *The problem of the Nile. Conditional solution to a change-point problem*, *Biometrika* **65** (1978), 243–251.
- Chow Y. S. a Teicher H., *Probability Theory* (1988), Springer Verlag, New York.
- Chung-Bow Lee, *Estimation of the number of change points in exponential families distributions*, preprint (1995).
- Chung-Bow Lee, *Estimation of the number of change points in normal random variables*, preprint (1995).
- Csörgő M. a Horváth L., *Nonparametric methods for the change-point problem*, *Handbook of Statistics* (Krishnaiah P. R. a Rao C. R., eds.), vol. 7, North Holland, Amsterdam, 1988, pp. 403–425.
- Csörgő M. a Horváth L., *Weighted Approximations in Probability and Statistics* (1993), J. Wiley, New York.
- Darkhovskiy B. E., *A nonparametric method for a posteriori detection of the “disorder” time of a sequence of independent random variables*, *Theory Probab. Appl.* **21** (1976), 178–183.
- Davis R. A., Huang D. a Yao Y. C., *Testing for a change in the parameter values and order of an autoregressive model*, *Ann. Statist.* **23** (1995), 282–304.
- De Long D., *Crossing probabilities for a square root boundary by a Bessel process*, *Commun. in Statist. – Theory and Method.* **A10** (1981), 2197–2213.
- Deshayes J. a Picard D., *Off-line statistical analysis of change point models using nonparametric and likelihood methods*, *Lecture Notes in Control and Information Sciences* (Basseville M. et al., eds.), vol. 77, Springer Verlag, New York, pp. 103–168.
- Dümbgen L., *The asymptotic behavior of some nonparametric change-point estimators*, *Ann. Statist.* **19** (1991), 1471–1495.
- Ferger D., *Change-point estimators in case of small disorders*, *J. Stat. Plan. Infer.* **40** (1994a), 33–49.
- Ferger D., *On the rate of almost sure convergence of Dümbgen change-point estimators*, *Statist. Probab. Letters* **25** (1994b), 27–31.
- Ferger D., *Asymptotic distribution theory of change-point estimators and confidence intervals based on bootstrap approximation*, *Math. Methods of Statistics* **3** (1994c), 362–378.
- Ferger D., *Nonparametric tests for nonstandard change-point problems*, *Ann. Statist.* **23** (1995), 1848–1861.
- Ferger D. a Stute W., *Convergence of change-point estimators*, *Stoch. Proc. Appl.* **42** (1992), 345–351.
- Gombay E. a Horváth L., *Limit theorems for change in linear regression*, *J. Multivariate Anal.* **48** (1994), 43–69.
- Gombay E., Horváth L. a Hušková M., *Estimators and tests for change in variances*, *Statistics & Decisions* **14** (1996), 145–159.
- Gombay E. a Horváth L., *Approximations for the time of change and the power function in change-point models*, *J. Stat. Plan. Infer.* **52** (1996), 43–66.
- Gombay E. a Hušková M., *Rank based estimators of the change-point*, *J. Stat. Plan. Infer.* **67** (1998), 137–154.
- Hinkley D. V., *Inference about the change-point in a sequence of random variables*, *Biometrika* **57** (1970), 1–17.
- Hinkley D. V. a Hinkley E. A., *Inference about the change-point in a sequence of binomial variables*, *Biometrika* **57** (1970), 477–488.
- Hinkley D. a Schechtman E., *Conditional bootstrap methods in the mean-shift model*, *Biometrika* **74** (1987), 85–93.



- Horváth L., *Detecting changes in linear regression*, *Statistics* **26** (1995), 189–208.
- Horváth L., Hušková M. a Serbinovska M., *Estimators for the time of change in linear models*, *Statistics* **27** (1997), 109–130.
- Horváth L. a Kokozska P., *The effect of long-range dependence on change-point estimators*, *J. Stat. Plan. Infer.* **64** (1997), 57–81.
- Horváth L. a Qi-Man Shao, *Darling-Erdős type theorems for sums of Gaussian variables with long range memory*, *Stoch. Proc. Appl.* **63** (1996), 117–137.
- Hušková M., *Estimators for epidemic alternatives*, *Comment. Math. Univ. Carolin.* **36** (1995), 279–291.
- Hušková M., *Limit theorems for rank statistics*, *Statist. Probab. Letters* **30** (1997a), 45–55.
- Hušková M., *Limit theorems for  $M$ -processes via rank statistics processes*, *Advances in Combinatorial Methods with Applications to Probability and Statistics* (Balakrishnan N., ed.), 1997b, pp. 527–533.
- Hušková M., *Multivariate rank statistics processes and change-point analysis*, Volume of Saleh, 1997c.
- Hušková M. a Sen P. K., *Nonparametric tests for shift and change in regression at an unknown time point*, *Statistical Analysis and Forecastings of Economic Structural Change* (Hackl P., ed.), Springer Verlag, New York, 1989, pp. 71–85.
- James B., James K. L. a Siegmund D., *Tests for change-points*, *Biometrika* **74** (1987), 71–84.
- Janhyala V. K. a MacNeill I. B., *Residual partial sum limit process for regression models with applications to detecting parameter changes at unknown times*, *Stoch. Proc. and Their Appl.* **33**, 309–323.
- Jandhyala V. K. a MacNeill I. B., *Tests for parameter changes at unown times in linear regression models*, *J. Stat. Plan. Infer.* **27** (1991), 291–316.
- Jandhyala V. K. a MacNeill I. B., *On testing for the constancy of regression coefficients under random walk and change-point alternatives*, *Econometric Theory* **8** (1992), 501–517.
- Jarušková D., *Detection of change – point in series of river discharges*, *Vodohospodářský časopis* **38** (1990), 501–515.
- Jarušková D., *Vyhledávání změn v průtokových řadách*, *ROBUST'90* (Antoch J. a Jurečková J., eds.), JČSMF, Praha, 1990, pp. 89–96.
- Jarušková D., *Detekce změn v rozptýlenosti*, *ROBUST'92* (Antoch J. a Dohnal G., eds.), JČMF, Praha, 1992, pp. 85–87.
- Jarušková D., *Statistical method for change-point detection using rainfalls – runoffs relationship*, *Vodohospodářský časopis* **41** (1993), 85–99.
- Jarušková D. a Antoch J., *Detection of change in variance*, *Proc. of the 5th Prague Symposium on Asymptotic Statistics* (Mandl P. a Hušková M., eds.), Physica-Verlag, Heidelberg, 1994, pp. 297–301.
- Jarušková D., *Change-point detection in meteorological measurement*, *Monthly Weather Review* **124** (1996), 1535–1543.
- Jarušková D., *Vznik lineárního spojitého trendu*, *ROBUST'96* (Antoch J. a Dohnal G., eds.), JČMF, Praha, 1996, pp. 85–90.
- Jarušková D., *Some problems with application of change-point detection methods to environmental data*, *Environmetrics* **8** (1997), 469–483.
- Jarušková D., *Change-point estimator in gradually changing sequences*, *CMUC* (1998). (v tisku)
- Jarušková D., *Testing appearance of linear trend*, *Statistical Planning and Inference* (1998). (v tisku)

- Kiefer J., *K-sample analogues of the Kolmogorov-Smirnov's and Cramér-von Mises tests*, Annals of Mathematical Statistics **30** (1960), 420–447.
- Kim Hyune-Ju a Siegmund D., *The likelihood ratio test for a change point in simple linear regression*, Biometrika **76** (1989), 409–423.
- MacNeill I.B., *Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times*, Ann. Statist. **6** (1978), 422–433.
- Michels P. a Trenkler G., *Testing the stability of regression coefficients using generalized recursive residuals*, Austr. J. Statist. **22** (1990), 293–312.
- Ritov Y., *Asymptotic efficient estimation of the change-point with unknown distribution*, Ann. Statist. **18** (1990), 1829–1839.
- Schechtman E., *A conservative nonparametric distribution-free confidence bound for the shift in the change-point problem*, Comment. Statist. **A12** (1983), 2455–2464.
- Sen P.K., *Asymptotic theory of some tests for constancy of regression relationships over time*, Statistics **13** (1982), 21–31.
- Sen P.K., *Recursive M-tests for constancy of multivariate regression relationships over time*, Sequential Analysis **3** (1984), 191–211.
- Shepp L. A., *The joint density of the maximum and its location for a Wiener process with drift*, J. Appl. Probab. **16** (1979), 423–427.
- Shorack G. R. a Wellner J. A., *Empirical Processes with Applications to Statistics* (1986), J. Wiley, New York.
- Siegmund D., *Boundary crossing probabilities and statistical applications*, Ann. Statist. **14** (1986), 361–404.
- Siegmund D., *Confidence sets in change-point problems*, Intern. Statist. Review **56** (1988), 31–48.
- Steinebach J. a Eastwood V. R., *Extreme value asymptotics for multivariate renewal processes*, J. Multivariate Analysis **56** (1996), 284–302.
- Stryhn H., *The location of maximum of asymmetric two-sided Brownian motion with triangular drift*, Statist. Probab. Letters **29** (1996), 279–284.
- Wolfe D. A. a Schechtman E., *Nonparametric procedures for the change-point problem*, J. Stat. Plan. Infer. **9** (1984), 389–396.
- Vostrikova L. Y., *Detection of "disorder" in multidimensional random processes*, Sov. Math. Dokl. **24** (1981), 55–59.
- Worsley K. J., *Testing for a two-phase multiple regression*, Technometrics **25** (1983), 35–42.
- Worsley K. J., *Confidence regions and tests for a change-point in a sequence of exponential family random variables*, Biometrika **73** (1986), 91–104.
- Yao Q., *Tests for change-points with epidemic alternatives*, Biometrika **80** (1993), 179–191.
- Yao Y.-C., *Estimating number of change-points via Schwarz' criterion*, Statist. Probab. Letters **6** (1988), 181–189.
- Yao Y.-C., *On asymptotic behavior of a class of nonparametric tests*, Statist. Probab. Letters **19** (1990), 173–177.
- Yao Y.-C., *Approximating the distribution of maximum likelihood estimate of the change-point in a sequence of independent random variables*, Ann. Statist. **15** (1987), 1321–1328.
- Zacks S., *Survey of classical and Bayesian approach to the change point problem. Fixed sample and sequential procedures of testing and sequential procedures of testing and estimation*, Recent Advances in Statistics, Papers in Honor of Herman Chernoff's Sixties Birthday (Rizvi M. H., ed.), Academic Press, New York, 1983, pp. 245–269.