WIENEROV PROCES S POSUNUTÍM; TESTY HYPOTÉZY O PARAMETRI POSUNUTIA

Andrea Kvitkovičová
UNIVERZITA KARLOVA V PRAZE
Matematicko - fyzikální fakulta, Katedra pravděpodobnosti a matematické statistiky
e-mail: akvitkovicova@centrum.cz

Abstrakt : *Uvažujme spojitý náhodný proces $\{M_t;\ t \geq 0\}$, u ktorého nesledujeme celý jeho priebeh, ale len čas, v ktorom prekročí hranicu $A > 0$. Na základe tejto informácie chceme testovať hypotézu $H_0 : M_t = w_t$, kde w_t je Wienerov proces, proti alternatíve $H_1 : M_t = w_t + at$, kde $a > 0$ je známe. Príspevok je zameraný na konštrukciu a vlastnosti dvoch testov založených na Neymanovej a Pearsonovej leme. V prvom prístupe použijeme rozdelenie náhodnej veličiny $\tau_A = \inf\{t \geq 0; M_t \geq A\}$, ktoré je za nulovej hypotézy známe a za alternatívnej hypotézy ho vypočítame pomocou Girsanovovej vety. V druhom prístupe sa problém prevedie na test hypotézy o parametri binomického rozdelenia pre náhodnú veličinu $B_n = \sum_{i=1}^{n} I\left\{\max_{0\leq t\leq T} M_t^i \geq A\right\}$.*

Výpočet hustoty náhodnej veličiny τ_A

Nech w_t je Wienerov proces na pravdepodobnostnom priestore (Ω, F, P), nech $A > 0$, $a > 0$. Potom náhodná veličina

$$\tau_A^w = \inf\{t \geq 0; w_t \geq A\}$$

má (podľa [2], str. 41) rozdelenie s Waldovou hustotou

$$f(s) = \frac{1}{\sqrt{2\pi}} \frac{A}{s^{\frac{3}{2}}} \exp\left\{-\frac{A^2}{2s}\right\} I\{s > 0\}.$$

Vypočítajme rozdelenie náhodnej veličiny

$$\widetilde{\tau}_A = \inf\{t \geq 0; w_t + at \geq A\}.$$

Veta (Girsanov) : *Nech w_t je d-dimenzionálny Wienerov proces na pravdepodobnostnom priestore (Ω, F, P). Nech $T \in [0, \infty)$ a nech b je F_t-adaptovaný, $F \otimes B((0,\infty))$-merateľný proces s hodnotami v $\mathbf{R}^\mathbf{d}$ spĺňajúci :*

$$\int_0^K b_s^2\, ds < \infty$$

(P - skoro všade) pre $\forall\ K < \infty$ a $E\rho_T\ (b) = 1$, kde

$$\rho_t(b) = \exp\left\{\int_0^t b_s\, dw_s - \frac{1}{2}\int_0^t (b_s)^2\, ds\right\}.$$

Na merateľnom priestore (Ω, F) definujme mieru $\widetilde{P}(d\omega) = \rho_T(b)(\omega)P(d\omega)$. Potom $(\Omega, F, \widetilde{P})$ je pravdepodobnostný priestor a

$$\widetilde{w}_t = w_t - \int_0^t b_s\, ds$$

je d-dimenzionálny Wienerov proces na pravdepodobnostnom priestore $(\Omega, F, \widetilde{P})$ pre $t \leq T$.

(podľa [3], str. 207)
Zvoľme $T > 0$ pevné. Pre $d = 1$ a $b = -a$ z Girsanovovej vety plynie, že ak definujeme $\widetilde{P}(d\omega) = \exp\left\{-aw_T - \frac{1}{2}a^2T\right\} P(d\omega)$, potom $\widetilde{w}_t = w_t + at$ je pre $t \leq T$ Wienerov proces na pravdepodobnostnom priestore $(\Omega, F, \widetilde{P})$.
Potom pre $0 < s < T$ platí :

$$P(\widetilde{\tau}_A \leq s) = P(\max_{0\leq t\leq s}(w_t + at) \geq A) =$$

$$EI\left\{\max_{0\leq t\leq s}(w_t + at) \geq A\right\} \exp\left\{aw_T + \frac{1}{2}a^2T\right\} \exp\left\{-aw_T - \frac{1}{2}a^2T\right\} =$$

$$E\left\{E\left[I\left\{\max_{0\leq t\leq s} w_t \geq A\right\} \exp\left\{a(w_T - w_s) - \frac{1}{2}a^2T\right\} \exp\{aw_s\} \Big| F_s\right]\right\} =$$

$$EI\left\{\max_{0\leq t\leq s} w_t \geq A\right\} \exp\left\{aw_s - \frac{1}{2}a^2 s\right\} =$$

$$\exp\{2aA\} \int_{-\infty}^{\frac{-A-as}{\sqrt{s}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz + \int_{\frac{A-as}{\sqrt{s}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz$$

A teda

$$\frac{d}{ds}\left(P\left(\widetilde{\tau}_A \leq s\right)\right) = \frac{1}{\sqrt{2\pi}} \frac{A}{s^{\frac{3}{2}}} \exp\left\{-\frac{1}{2s}(A - as)^2\right\}$$

pre $0 < s < T$. Keďže tento výraz nezávisí na T, dostávame pre náhodnú veličinu $\widetilde{\tau}_A$ na intervale $(0, \infty)$ hustotu

$$\tilde{f}(s) = \frac{1}{\sqrt{2\pi}} \frac{A}{s^{\frac{3}{2}}} \exp\left\{-\frac{1}{2s}(A - as)^2\right\}.$$

Zostavenie testu

Problém : *Na základe času, v ktorom jedna realizácia náhodného procesu $\{M_t;\ t \geq 0\}$ prekročí hranicu $A > 0$, kde A môžeme voliť, chceme testovať hypotézu*
$H_0 : M_t = w_t$ proti alternatíve
$H_1 : M_t = w_t + at$, kde $a > 0$ je známe.
a) proces sledujeme tak dlho, ako potrebujeme
b) proces sledujeme len do času $T > 0$.

Lema (Neyman a Pearson) : *Nech P_0 a P_1 sú dve rozdelenia pravdepodobností na merateľnom priestore (Ω, F), ktoré majú hustoty p_0 a p_1 vzhľadom na mieru μ.*
Potom pre ľubovoľné $\alpha \in (0,1)$ existuje test hypotézy $H_0 : P = P_0$ proti alternatíve $H_1 : P = P_1$ charakterizovaný testovou funkciou Ψ spĺňajúcou :
$\Psi(x) = 1$ pre $p_1(x) > k\ p_0(x)$
$\Psi(x) = 0$ pre $p_1(x) < k\ p_0(x)$,
kde konštanta k a hodnoty $\Psi(x)$ na $\{x : p_1(x) = k\ p_0(x)\}$ sú určené tak, že platí : $E_0\Psi(X) = \alpha$. Test charakterizovaný Ψ je najsilnejším testom H_0 proti H_1 na hladine menšej, nanajvýš rovnej α.

(podľa [1], str. 30)
Pre A pevné je úloha ekvivalentná testu hypotézy
H_0 : čas prekročenia hladiny A má hustotu f voči Lebesgueovej miere λ
proti alternatíve
H_1 : čas prekročenia hladiny A má hustotu $\tilde{f}$ voči λ.
Pre jednoduchosť ďalej predpokladajme, že $\alpha < \frac{1}{2}$. Keďže α je hladina testu, nie je tento predpoklad nijako obmedzujúci. Podľa Neymanovej a Pearsonovej lemy zamietame hypotézu H_0 v prospech hypotézy H_1 práve vtedy, keď

$$\frac{\tilde{f}(s)}{f(s)} = \exp\left\{-\frac{1}{2}\left(a^2 s - 2aA\right)\right\} \geq k,$$

čo je splnené práve vtedy, keď

$$s \leq \frac{2aA - 2\ln k}{a^2} = d,$$

kde d budeme voliť tak, aby sme dosiahli požadovanú hladinu testu α. Položme teda

$$\alpha = P_{H_0}(\text{zamietneme } H_0) = \int_0^d \frac{1}{\sqrt{2\pi}} \frac{A}{s^{\frac{3}{2}}} \exp\left\{-\frac{1}{2s}A^2\right\} ds.$$

Odtiaľ plynie, že

$$d = \frac{A^2}{u_{1-\frac{\alpha}{2}}^2},$$

kde $u_{1-\frac{\alpha}{2}}$ je kvantil normálneho rozdelenia, teda pre náhodnú veličinu $X \sim N(0,1)$ platí :

$$P\left(X \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

Najsilnejší test hypotézy H_0 proti alternatíve H_1 na hladine menšej, nanajvýš rovnej α je charakterizovaný testovou funkciou

$$\Psi(s) = I\left\{s \leq \frac{A^2}{u_{1-\frac{\alpha}{2}}^2}\right\}.$$

Pre silu testu $\beta(A, a)$ platí :

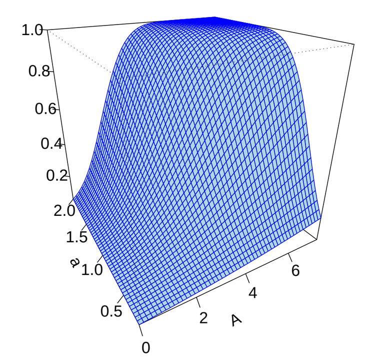
$$\beta(A, a) = P_{H_1}(\text{zamietneme } H_0) =$$

$$\exp\{2Aa\}\, \Phi\left(\frac{-u_{1-\frac{\alpha}{2}}^2 - Aa}{u_{1-\frac{\alpha}{2}}}\right) + \Phi\left(\frac{-u_{1-\frac{\alpha}{2}}^2 + Aa}{u_{1-\frac{\alpha}{2}}}\right),$$

kde $\Phi(x)$ je distribučná funkcia normovaného normálneho rozdelenia. Derivovaním $\beta(A, a)$ podľa A zistíme, že sila testu je pre pevné a rastúcou funkciou A. Keďže

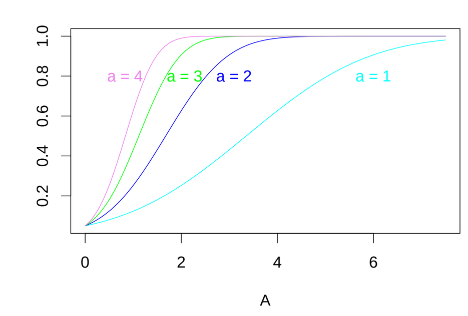
$$\lim_{A\to\infty} \beta(A, a) = 1,$$

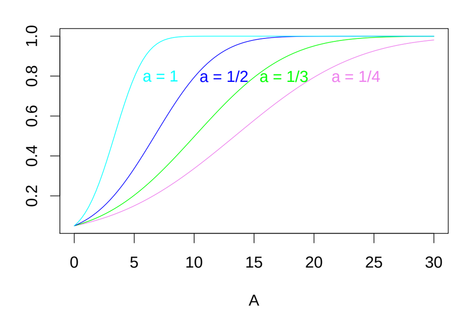
môžeme voľbou A dostatočne veľkého dosiahnuť ľubovoľne veľkú silu testu.



Sila testu ako funkcia dvoch premenných a a A pre $\alpha = 0.05$.
Pre malé hodnoty testovanej smernice a rastie sila testu so zväčšujúcim sa A pomalšie než pre väčšie hodnoty smernice a.

Pre väčšiu názornosť sa pozrime na silu testu ešte ako na funkciu jednej premennej A pre niektoré konkrétne hodnoty a a pre $\alpha = 0.05$.





Pre dolný odhad sily testu môžeme použiť niektoré z odhadov pre distribučnú funkciu normovaného normálneho rozdelenia. Napríklad podľa Fellerových nerovností pre $x > 0$ platí :

$$1 - \frac{1}{x}\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \leq \Phi(x) \leq 1 - \left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

(podľa [4], str. 318).
Odtiaľ plynie, že pre $aA > u_{1-\frac{\alpha}{2}}^2$ platí :

$$\beta(A, a) >$$

$$1 - \frac{1}{\sqrt{2\pi}} \frac{u_{1-\frac{\alpha}{2}}^3}{u_{1-\frac{\alpha}{2}}^2 + aA} \left(\frac{1}{aA - u_{1-\frac{\alpha}{2}}^2} + \frac{1}{\left(aA + u_{1-\frac{\alpha}{2}}^2\right)^2}\right) \exp\left\{-\frac{1}{2u_{1-\frac{\alpha}{2}}^2}\left(aA - u_{1-\frac{\alpha}{2}}^2\right)^2\right\}.$$

Tento odhad je ale pre niektoré hodnoty a, A a α príliš hrubý. Presnejšiu aproximáciu pre tieto hodnoty a, A a α získame buď pomocou iných odhadov distribučnej funkcie normovaného normálneho rozdelenia alebo pomocou štatistického softwaru.

Návrh testu :
a) proces sledujeme tak dlho, ako potrebujeme :
Pre požadovanú hladinu testu α a silu testu β nájdeme A dosť veľké tak, aby test mal aspoň danú silu. Potom sledujeme realizáciu procesu $\{M_t;\ t \geq 0\}$ do času

$$T = \frac{A^2}{u_{1-\frac{\alpha}{2}}^2}.$$

Ak dovtedy prekročí hranicu A, zamietame nulovú hypotézu v prospech alternatívnej. Inak nulovú hypotézu nezamietame.
b) proces sledujeme len do času $T > 0$:
Pre požadovanú hladinu testu α volíme $A = u_{1-\frac{\alpha}{2}}\sqrt{T}$. Ak do času T proces prekročí hranicu A, zamietneme nulovú hypotézu v prospech alternatívnej; inak nulovú hypotézu nezamietame.

Test navrhnutý pre prípad b) má silu

$$\beta = \exp\left\{2au_{1-\frac{\alpha}{2}}\sqrt{T}\right\} \Phi\left(-u_{1-\frac{\alpha}{2}} - \sqrt{T}a\right) + \Phi\left(-u_{1-\frac{\alpha}{2}} + \sqrt{T}a\right),$$

čo je maximum, ktoré sa za daných podmienok dalo dosiahnuť.

Definícia : *Nech $\Lambda = \{P_\theta, \theta \in \Theta\}$ je systém dominovaný σ-konečnou mierou μ, kde $\Theta \subseteq \mathbf{R}$. Hovoríme, že Λ má monotónny pomer vierohodnosti, keď existuje reálna štatistika $T(x)$ tak, že pre ľubovoľné $\theta < \theta'$ sú rozdelenia P_θ a $P_{\theta'}$ navzájom rôzne a pomer $\frac{p_{\theta'}(x)}{p_\theta(x)}$ ich hustôt je neklesajúcou funkciou $T(x)$.*

Veta (O monotónnom pomere vierohodnosti) : *Nech systém rozdelení $\Lambda = \{P_\theta, \theta \in \Theta\}$, $\Theta \subseteq \mathbf{R}$ náhodného vektoru $\mathbf{X}$ má monotónny pomer vierohodnosti vzhľadom na štatistiku $T(\mathbf{X})$. Potom existuje rovnomerne najsilnejší test hypotézy $H_0 : \theta \leq \theta_0$ proti alternatíve $H_1 : \theta > \theta_0$ charakterizovaný testovou funkciou Ψ spĺňajúcou :*
$\Psi(\mathbf{x}) = 1$ pre $T(\mathbf{x}) > c$
$\Psi(\mathbf{x}) = 0$ pre $T(\mathbf{x}) < c$,
$\Psi(\mathbf{x}) = \gamma$ pre $T(\mathbf{x}) = c$,
kde konštanty c a γ sú volené tak, aby platilo : $E_{\theta_0}\Psi(\mathbf{X}) = \alpha$. Silofunkcia $\beta(\theta) = E_\theta\Psi(\mathbf{X})$ testu charakterizovaného $\Psi(\mathbf{X})$ je rastúca na množine $\{\theta; \beta(\theta) < 1\}$.

(podľa [1], str. 33, 34)
Uvažujme systém $\Lambda = \{P_a, a \in [0, \infty)\}$ rozdelení náhodnej veličiny $\tau_A = \inf\{t \geq 0; M_t \geq A\}$. Tento systém má monotónny pomer vierohodnosti vzhľadom na štatistiku - τ_A. Podľa vety o monotónnom pomere vierohodnosti teda existuje rovnomerne najsilnejší test hypotézy $H_0 : a = 0$ proti alternatíve $H_1 : a > 0$ charakterizovaný testovou funkciou Ψ spĺňajúcou :
$\Psi(t) = 1$ pre $t \leq c$ a
$\Psi(t) = 0$ pre $t > c$,
kde konštanta c je volená tak, aby platilo : $E_0\Psi(\tau_A) = \alpha$. Z toho vyplýva, že vyššie navrhnutý test sa dá použiť aj pre zloženú jednostrannú alternatívu, a to dokonca ako rovnomerne najsilnejší test. V takomto prípade sa ale nedajú použiť uvedené vzťahy a odhady pre silu testu.

Inou modifikáciou vyššie odvodeného testu je jeho využitie na testovanie jednoduchej hypotézy proti jednoduchej alternatíve, keď máme k dispozícii údaje o n nezávislých realizáciách náhodného procesu $\{M_t;\ t \geq 0\}$. Nulovú hypotézu tentoraz zamietame, keď súčet časov prekročenia hranice A jednotlivými realizáciami je menší, nanajvýš rovný

$$d = \frac{(nA)^2}{u_{1-\frac{\alpha}{2}}^2}.$$

V prípade, že už súčet časov prekročenia hranice A prvými $n_1 < n$ realizáciami procesu je väčší než d, môžeme sa pre nezamietnutie nulovej hypotézy rozhodnúť ešte pred ukončením pokusu. Inak musíme každú realizáciu sledovať až do času $T = d$. Porovnaním tejto modifikácie testu s jeho pôvodnou verziou zistíme, že pri zachovaní tej istej hladiny testu sledujeme prekračovanie n-krát nižšej hranice. V nesekvenčnom prípade však každú z n realizácií sledujeme rovnako dlho, ako sme predtým sledovali realizáciu jedinú.

Testy bez znalosti času prekročenia A

Problém : *Na základe informácie, koľko z n na sebe nezávislých realizácií náhodného procesu $\{M_t;\ t \geq 0\}$ prekročí hranicu $A > 0$ do času T, kde T je dané, chceme testovať hypotézu*
$H_0 : M_t = w_t$ proti alternatíve
$H_1 : M_t = w_t + at$, kde $a > 0$ je známe.

Tentoraz teda nevieme, kedy jednotlivé realizácie hranicu prekročia, ale koľko z nich ju prekročí do času T.
Uvažujme indikátorové náhodné veličiny

$$I\left\{\max_{0\leq t\leq T} M_t^i \geq A\right\}.$$

Pre a, A a T pevné ide o nezávislé rovnako rozdelené náhodné veličiny s alternatívnym rozdelením s parametrom p. Ich súčet

$$B_n = \sum_{i=1}^{n} I\left\{\max_{0\leq t\leq T} M_t^i \geq A\right\}$$

má teda binomické rozdelenie s parametrami n a p. Úlohu teraz môžeme previesť na test hypotézy $H_0 : p = p_0$ proti alternatíve $H_1 : p = p_1$, kde $p_1 > p_0$, konkrétne

$$p_0 = P\left(\max_{0\leq t\leq T} w_t \geq A\right) = 2\Phi\left(-\frac{A}{\sqrt{T}}\right),$$

$$p_1 = P\left(\max_{0\leq t\leq T}(w_t + at) \geq A\right) = \exp\{2aA\}\, \Phi\left(\frac{-A - aT}{\sqrt{T}}\right) + \Phi\left(\frac{-A + aT}{\sqrt{T}}\right).$$

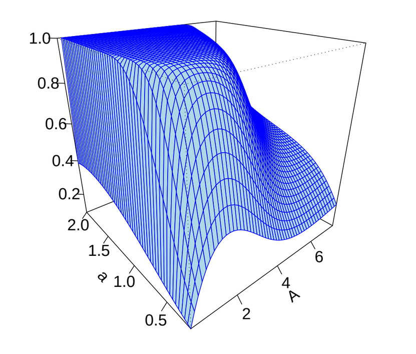
Problém potom riešime niektorým z postupov používaných na testy hypotéz o parametri binomického rozdelenia. Pre n malé použijeme randomizovaný test; pre n veľké aproximujeme binomické rozdelenie normálnym. V druhom prípade tak dostávame test charakterizovaný testovou funkciou

$$\Psi(b) = I\left\{b \geq u_{1-\alpha}\sqrt{np_0(1 - p_0)} + np_0\right\}.$$

Pre silu testu potom platí :

$$\beta(p_0, p_1) = \Phi\left(\frac{\sqrt{n}(p_1 - p_0)}{\sqrt{p_1(1 - p_1)}} - u_{1-\alpha}\frac{\sqrt{p_0(1 - p_0)}}{\sqrt{p_1(1 - p_1)}}\right),$$

kde p_0 a p_1 sú definované vyššie.



Graf závislosti sily testu založeného na aproximácii binomického rozdelenia normálnym na a a A pre $\alpha = 0.05$, $T = 2$, $n = 50$.
Za daných podmienok je možné maximalizovať silu testu voľbou vhodného A. Pre A príliš veľké sú p_0 aj p_1 blízke nule a pre A príliš malé sú p_0 aj p_1 blízke jednej.

Keďže je medzi nimi v týchto prípadoch pomerne malý rozdiel, je ťažké ich od seba odlíšiť, čo spôsobuje nižšiu silu testu. Voľba vhodného A je otázkou jej dobrej numerickej aproximácie.

Systém $\Lambda = \{P_p, p \in [p_0, 1)\}$ binomických rozdelení s parametrom p náhodnej veličiny B má monotónny pomer vierohodnosti vzhľadom na štatistiku B, takže podľa vety o monotónnom pomere vierohodnosti test na hladine α, pri ktorom zamietame nulovú hypotézu práve vtedy, keď je počet realizácií, ktoré prekročili A, veľký, opäť dostávame zároveň aj rovnomerne najsilnejší test jednoduchej hypotézy proti zloženej jednostrannej alternatíve.

Poďakovanie: *Rada by som poďakovala RNDr. Danielovi Hlubinkovi, Ph.D. za podnety, rady a konzultácie, bez ktorých by tento príspevok nebol vznikol. Poďakovanie tiež patrí ČSOB, ktorá umožnila moju účasť na Robuste 2006.*

Literatúra

[1] J.Jurečková. *Testy parametrických hypotéz*, SPN, 1981.
[2] N.V.Krylov. *Introduction to the Theory of Diffusion Processes*, AMS, 1996.
[3] N.V.Krylov. *Introduction to the Theory of Random Processes*, AMS, 2002.
[4] J.Štěpán. *Teorie pravděpodobnosti. Matematické základy*, Academia, 1987.