

Pokroky matematiky, fyziky a astronomie

Tomáš Mrkvička

Globální obálkové testy aneb jak otestovat vhodnost statistického modelu na základě funkcionální charakteristiky

Pokroky matematiky, fyziky a astronomie, Vol. 62 (2017), No. 1, 17–23

Persistent URL: <http://dml.cz/dmlcz/146720>

Terms of use:

© Jednota českých matematiků a fyziků, 2017

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

Globální obálkové testy aneb jak otestovat vhodnost statistického modelu na základě funkcionální charakteristiky

Tomáš Mrkvička, České Budějovice

Abstrakt. Obálkové metody představují populární nástroj pro testování hypotéz o vhodnosti statistického modelu. Tyto testy graficky porovnávají funkci $T : I \rightarrow \mathbb{R}$ vypočtenou ze statistických dat s jejím protějškem získaným simulacemi. Chyba prvního druhu α , tj. pravděpodobnost zamítnutí platné hypotézy, je obvykle kontrolována pouze pro fixní hodnotu $r \in I$, zatímco funkce T je definována na intervalu hodnot I . V tomto článku představíme nový globální obálkový test, který umožňuje kontrolovat chybu prvního druhu současně pro všechny hodnoty r z intervalu I pomocí tzv. globální obálky, která je přímo provázána s výslednou p -hodnotou testu. Použití obálkového testu ilustrujeme na příkladu zkoumání interakcí mezi částicemi v rovině.

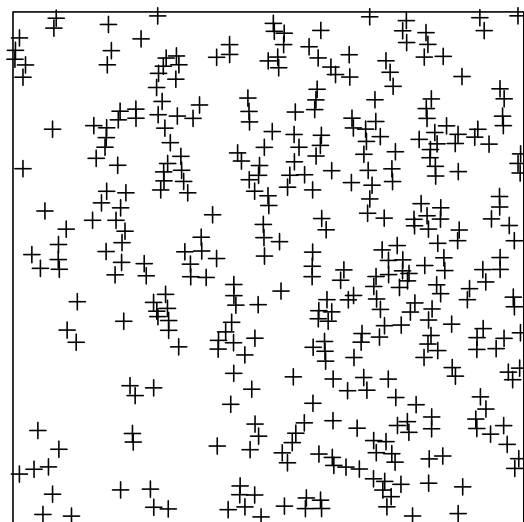
1. Úvod

Při analýze statistických dat se v současné době pracuje s velmi komplexními matematickými modely. Obvykle chceme otestovat, zda zvolený model skutečně odpovídá naměřeným datům. Abychom mohli takový test provést, musíme nejprve zvolit testovou statistiku, která je schopná identifikovat, zda naše data odpovídají modelu. V nejjednodušších případech se jako testová statistika používá jednorozměrná charakteristika (číslo spočtené z dat, např. testová statistika T u Studentova t -testu). Ta ovšem bývá v dnešní době často nedostačující, a proto je nutné uvažovat statistiky vícerozměrné (vektorové), nebo dokonce funkcionální (tj. funkce $T : I \rightarrow \mathbb{R}$, kde $I \subset \mathbb{R}^d$). Při jejich použití se test vlastně skládá z řady samostatných jednorozměrných podtestů, které jsou vzájemně korelované (hovoří se o problému mnohonásobného porovnávání).

V tomto článku shrneme některé metody, které řeší problém testování vhodnosti modelu při použití vícerozměrné nebo funkcionální charakteristiky jako testovací statistiky. Tyto metody navíc poskytují tzv. globální obálky, které umožňují grafickou interpretaci výsledku testu.

V článku ukážeme, jak provádět globální obálkové testy na příkladu bodového procesu, tyto metody se ovšem dají využít v mnoha jiných oblastech, např. v časových řadách, v geostatistice, k testování modelů distribučních funkcí, v analýze funkcionálních dat atd. Bodovým procesem rozumíme náhodnou konfiguraci bodů v prostoru [4]

Doc. RNDr. TOMÁŠ MRKVIČKA, Ph.D., Katedra aplikované matematiky a informatiky, Ekonomická fakulta JU v Českých Budějovicích, Studentská 13, 370 01 České Budějovice, e-mail: mrkvicka.toma@gmail.com



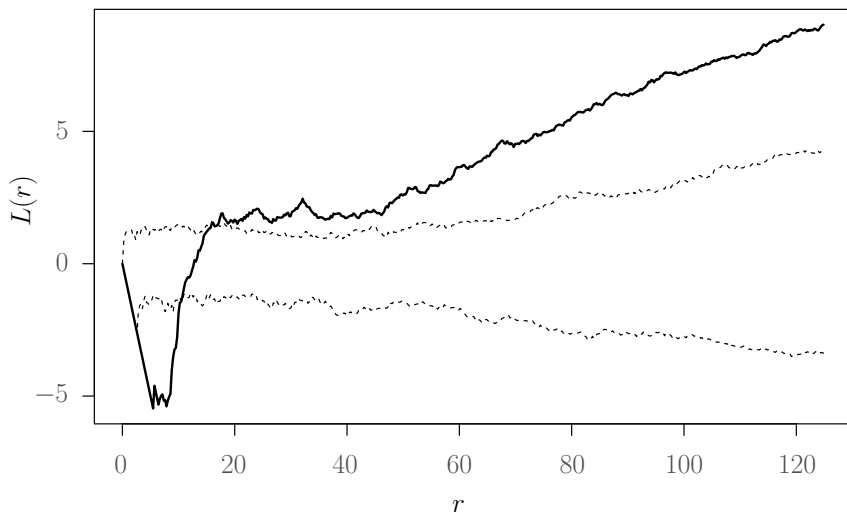
Obr. 1. Vzorek nitromembránových částic pozorovaný v okně velikosti 512×512 pixelů ($336 \text{ nm} \times 336 \text{ nm}$)

a pro jednoduchost budeme uvažovat jen rovinný bodový proces. Příklad realizace bodového procesu je na obr. 1.

2. Jednorozměrný Monte Carlo test

Obr. 1, který v dalším textu použijeme jako ilustrační příklad, ukazuje prostorové umístění nitromembránových částic mitochondriálních membrán z HeLa buněk, a to konkrétně jednoho vzorku kontrolní skupiny převzatého ze studie [3]. HeLa buňky jsou nádorové buňky děložního hrdla. Otázka, která nás zajímá, je, zda částice mezi sebou interagují, případně jaký typ interakce vykazují. Jako model, který může tuto otázku zodpovědět, zvolíme Poissonův bodový proces, který popisuje body, jež jsou úplně náhodně rozmístěné v prostoru. To znamená, že počet bodů v libovolné podmnožině B uvažované oblasti má Poissonovo rozdělení s parametrem daným jako λ -násobek obsahu množiny B , kde λ je tzv. intenzita procesu (vyjadřuje střední počet bodů v množině o jednotkovém obsahu). Dále platí, že počet bodů v množině B je nezávislý na počtu bodů v jiných, disjunktních množinách. Tento model nevykazuje žádné interakce mezi částicemi. Nulová hypotéza, kterou budeme testovat, je, zda uvedená data mohou být modelována Poissonovým bodovým procesem.

Nejpoužívanější funkcionální charakteristikou bodových procesů je tzv. Ripleyho K -funkce, $K(r) = \mathbb{E}(N_r)/\lambda$, kde N_r je počet bodů do vzdálenosti r od náhodného bodu, \mathbb{E} označuje střední hodnotu a λ je intenzita procesu. Normovaná, centrovaná K -funkce se označuje jako L -funkce a je definována vztahem $L(r) = \sqrt{K(r)/\pi} - r$. Pro Poissonův proces platí $L(r) \equiv 0$. Jestliže pro malá r je $L(r) > 0$, pak lze ve vzorku identifikovat shluky bodů. Pokud naopak pro malá r platí $L(r) < 0$, pozorujeme ve vzorku odpuzující chování bodů.



Obr. 2. Čárkovaně je zobrazena hranice obálky, tj. maximum a minimum spočtené z 99 simulací L -funkce. Plná čára zobrazuje L -funkci spočtenou z dat na obrázku 1.

Jako jednorozměrnou testovou statistiku můžeme použít např. hodnotu $L(R)$ pro konkrétní vzdálenost R . Monte Carlo test pak proběhne tak, že nasimulujeme s Poissonových bodových procesů a z nich spočteme s testových statistik $L_1(R), \dots, L_s(R)$. Patří-li pak $L(R)$ mezi 2,5 % největších nebo 2,5 % nejmenších hodnot, zamítneme nulovou hypotézu na hladině významnosti 5 % (tj. s chybou prvního druhu 5 %).

V uvedeném příkladě je zjevným nedostatkem fakt, že bereme v úvahu pouze jednu vzdálenost R . Vezmeme-li v úvahu více hodnot R , nastane problém s vícenásobným porovnáváním. A právě řešení tohoto problému se budeme dále věnovat.

Na obr. 2 je pro úplnost zobrazena průzkumná metoda, jež se často používala před zavedením globálních obálek. Hranici obálky zde tvoří maximum a minimum z 99 funkcí L získaných simulacemi Poissonova procesu. Pro pevně zvolené R platí, že pokud $L(R)$ neleží v intervalu určeném obálkou, pak můžeme zamítnout nulovou hypotézu na hladině významnosti $\alpha = 2\%$ (dle jednorozměrného Monte Carlo testu spočteného na základě 99 simulací). Problém nastává, jestliže se graf pozoruje jako celek a ze skutečnosti, že $L(r)$ je mimo zobrazenou obálku pro nějaké $r \in I$, se vyvozují závěry o zamítnutí nulové hypotézy. Takové tvrzení může mít hladinu významnosti α značně velikou; konkrétní příklady ukazují, že α může narůst až přes 60 %. Toto zjištění rozpoutalo vlnu studií, které vedly k zavedení globálních obálek.

3. Pořadový obálkový test

Pořadový obálkový test zavedený poprvé v [1] poskytuje globální obálku a zároveň p -hodnotu testu. Globální obálka vytvořená v rámci ilustračního příkladu je zobrazena na obr. 3. Vidíme zde testovou funkci spočtenou na základě zkoumaných dat (plná čára) a 95% globální obálku spočtenou ze simulací testové funkce za předpokladu nulové hypotézy. Jestliže funkce spočtená z dat vystupuje z globální obálky, dostáváme

důkaz o neplatnosti nulové hypotézy na hladině významnosti 5%. Navíc zde můžeme pozorovat, pro které hodnoty argumentu r jsou data mimo globální obálku.

Pořadový obálkový test je plně neparametrický, tudíž není senzitivní na změny rozdělení v testové statistice $T(r)$ pro různá $r \in I$ (kde I je množina hodnot r , které vstupují do testu). Všechny hodnoty r , které bereme v potaz, mají v testu stejnou váhu. Tato vlastnost umožňuje konstrukci testu na základě několika testových funkcí zároveň, jak bylo navrženo v článku [2]. V tomto spojeném testu mají všechny testové funkce stejnou váhu, což vede k mnoha zajímavým aplikacím testu.

Popišme nyní v krátkosti zmíněný test. Detaily je možné dohledat v originálních publikacích. Předpokládejme, že máme testovou funkci $T_1 : I \rightarrow \mathbb{R}$, která je spočtená z dat, a soubor simulací $T_2, \dots, T_{s+1} : I \rightarrow \mathbb{R}$ získaných za předpokladu nulové hypotézy. Definujme k -tou horní a dolní pořadovou obálku jako

$$T_{\text{low}}^{(k)}(r) = \min_{i=1, \dots, s+1}^k T_i(r), \quad T_{\text{upp}}^{(k)}(r) = \max_{i=1, \dots, s+1}^k T_i(r), \quad k = 1, \dots, \lfloor s/2 \rfloor,$$

kde \min^k a \max^k označují k -té nejmenší a největší hodnoty. Potom *extrémní pořadí* R_i je největší k takové, že

$$T_{\text{low}}^{(k)}(r) \leq T_i(r) \leq T_{\text{upp}}^{(k)}(r) \quad \text{pro všechna } r \in I.$$

Velká hodnota R_i říká, že funkce T_i je centrální funkcí v souboru $\{T_1, \dots, T_{s+1}\}$, zatímco malá hodnota R_i znamená extrémnost.

Přesné pořadí R_1 mezi všemi hodnotami R_1, \dots, R_{s+1} nemůžeme vždy určit, neboť některé hodnoty se mohou shodovat. Můžeme ale definovat nejmenší p -hodnotu, označovanou jako liberální, a největší p -hodnotu, označovanou jako konzervativní (symbol **1** označuje indikátorovou funkci):

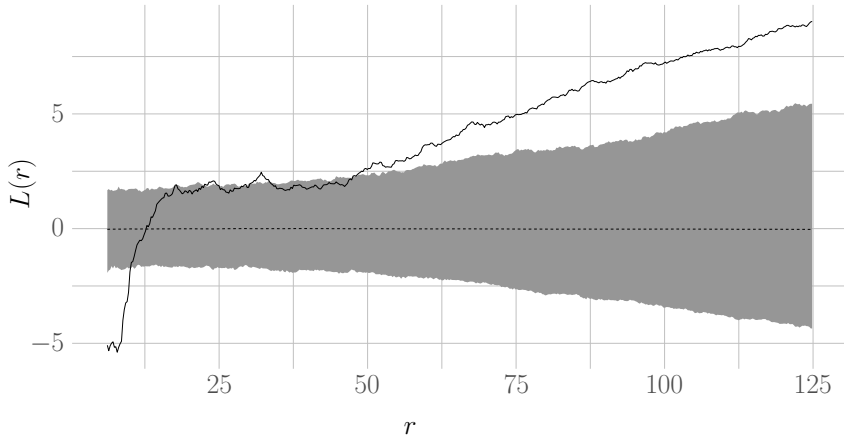
$$p_- = \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_i < R_1) \quad \text{a} \quad p_+ = \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_i \leq R_1).$$

Výsledkem testu pak není jediná p -hodnota, ale p -interval $(p_-, p_+]$. Pro hladinu spolehlivosti $\alpha \geq p_+$ dojde k zamítnutí nulové hypotézy, zatímco pro $\alpha < p_-$ k jejímu nezamítnutí (viz níže).

$100 \cdot (1 - \alpha)\%$ globální obálka se konstruuje následovně: Necht t_k je počet simulací T_i , $i \in \{1, \dots, s+1\}$, pro které $T_{\text{low}}^{(k)}(r) \leq T_i(r) \leq T_{\text{upp}}^{(k)}(r)$ pro všechna $r \in I$. Pak $100 \cdot (1 - \alpha)\%$ globální obálka je dána kritickými hranicemi $T_{\text{upp}}^{(k^\alpha)}$ a $T_{\text{low}}^{(k^\alpha)}$, kde k^α je největší k , pro něž $t_k / (s+1) \geq 1 - \alpha$.

Grafická interpretace testu, která odpovídá interpretaci dané p -intervalem, je následující:

- Jestliže se graf T_1 nebo jeho část nachází mimo $100 \cdot (1 - \alpha)\%$ globální obálku, potom $p_+ \leq \alpha$ a nulová hypotéza je zamítnuta na předepsané hladině testu α .
- Jestliže graf T_1 leží uvnitř $100 \cdot (1 - \alpha)\%$ globální obálky a nedotýká se její hranice, potom $p_- > \alpha$ a nulová hypotéza není zamítnuta.
- Jestliže graf T_1 leží uvnitř obálky a dotýká se její hranice, potom $p_- \leq \alpha < p_+$ a rozhodnutí je v „šedé zóně“.



Obr. 3. Pořadový obálkový test nulové hypotézy odpovídající tvrzení, že Poissonův proces je vhodným modelem pro nitromembránová data. Test byl proveden na základě 4999 simulací. Šedá oblast ukazuje 95% globální obálku.

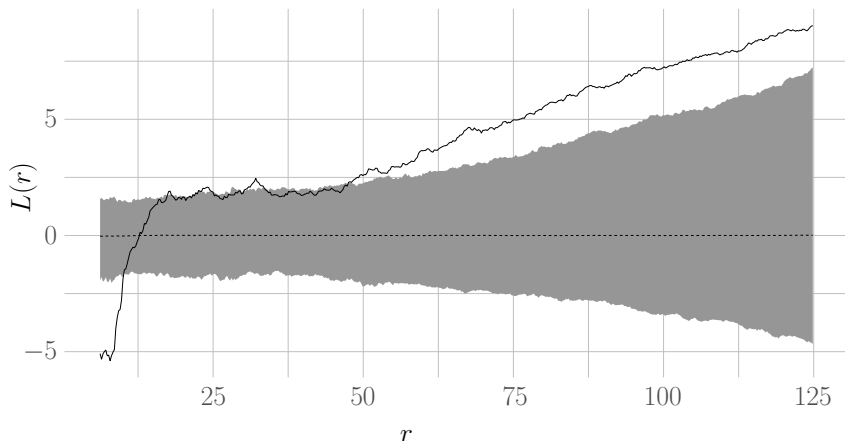
Jeden možný přístup k řešení problému „šedé zóny“ je založen na zjemnění výše definovaného extrémního pořadí R_i ; detaily lze najít v [1].

Obr. 3 ukazuje grafický výsledek pořadového obálkového testu, zkoumajícího vhodnost Poissonova modelu pro data z obr. 1. Vhodnost modelu je jednoznačně zamítnuta, jelikož testová funkce spočtená z dat je mimo globální obálku. Navíc ovšem získáváme informaci o tom, že částice se odpuzují pro velmi malé vzdálenosti (což je zřejmě způsobeno fyzickou velikostí částic) a naopak tvoří shluky v mezičásticových vzdálenostech okolo 75 pixelů. Tyto výsledky jsme prokázali na hladině významnosti $\alpha = 5\%$, což je globální hladina svazující významnost všech podtestů reprezentovaných jednotlivými hodnotami r .

4. Složená hypotéza

Pořadový obálkový test dosahuje přesně přednastavené hladiny α v případě, že nulová hypotéza je jednoduchá, tj. nezávisí na parametrech nebo jsou tyto parametry známy. Výše testovaná nulová hypotéza je ovšem složená, jelikož v rámci testování Poissonova modelu odhadujeme parametr intenzity λ . Provedený test tudíž nemá přesně zvolenou hladinu významnosti 5%, ale je lehce konzervativní. Můžeme říci, že je pouze lehce konzervativní, jelikož procedura pro odhad parametru intenzity (počet pozorovaných bodů vydělený obsahem pozorovaného okna) je pouze slabě korelovaná s procedurou testu založenou na L -funkci, která sleduje počet sousedů typického bodu normovaný právě intenzitou. Tuto konzervativnost testu buď můžeme akceptovat, nebo provedeme tzv. upravený pořadový obálkový test [1], který tento problém vyřeší. Problém tohoto řešení ovšem spočívá v tom, že testová procedura obsahuje dvojnásobný Monte Carlo test, neboli Monte Carlo test vytvořený pro každou simulaci Monte Carlo testu. Taková procedura vyžaduje s^2 simulací, a tudíž je značně výpočetně náročná.

Obr. 4 ukazuje grafický výsledek upraveného pořadového obálkového testu vhodnosti Poissonova modelu pro zkoumaná data. Výsledky jsou ve shodě s neupraveným



Obr. 4. Upravený pořadový obáلكový test nulové hypotézy odpovídající tvrzení, že Poissonův proces je vhodným modelem pro nitromembránová data. Test byl proveden na základě 2499^2 simulací. Šedá oblast ukazuje 95% upravenou globální obáلكu.

testem, jelikož rozdíl mezi oběma variantami je v tomto případě mizivý. Na druhou stranu v případě, že korelace mezi procedurou odhadu a procedurou testu je větší, může být rozdíl mezi oběma variantami testu značný.

5. Diskuze

Pro některé (zejména složené) nulové hypotézy můžou simulace vyžadované testem trvat příliš dlouho. V takovém případě je možné použít aproximace pořadového testu zavedené také v článku [1], které jsou částečně parametrické a aproximují obáلكy pomocí kvantilů rozdělení $T(r)$. Tento test pak vyžaduje minimálně $s = 99$ simulací, zatímco u pořadového testu je doporučováno alespoň $s = 2499$ simulací.

Simulační studie provedené v článku [1] dále ukazují, že pořadový obáلكový test je značně stabilní vzhledem ke změně délky intervalu I , který v testu uvažujeme. Je doporučováno zvolit za I široký interval, který pokryje celou oblast zájmu.

Někdy můžeme vybrat konkrétní testovou funkci, jež je citlivá na konkrétní alternativní model, vůči kterému chceme nulovou hypotézu testovat. Nicméně častěji alternativní model není znám, tudíž neexistuje a priori nejlepší testová funkce dané nulové hypotézy. V takovém případě můžeme testy založit na více testových funkcích, jak bylo ukázáno v článku [2]. Pořadový obáلكový test může být použit pro kombinaci několika testových funkcí zároveň, přičemž vrátí jednu globální p -hodnotu a kombinovanou globální obáلكu pro všechny funkce. Navíc simulační studie ukazují, že kombinovaný pořadový test má jen o málo menší sílu (pravděpodobnost zamítnutí neplatné nulové hypotézy) než pořadový test založený na jedné nejsilnější testové funkci, která ovšem může být určena pouze po provedení konkrétních simulačních studií.

Poděkování. Autor byl finančně podporován Grantovou agenturou České republiky (projekt 16-03708S).

L i t e r a t u r a

- [1] MYLLYMÄKI, M., MRKVIČKA, T., SELJO, H., GRABARNIK, P., HAHN, U.: *Global envelope tests for spatial processes*. J. R. Statist. Soc. B (2016), v tisku, doi: 10.1111/rssb.12172.
- [2] MRKVIČKA, T., MYLLYMÄKI, M., HAHN, U.: *Multiple Monte Carlo testing with applications in spatial point processes*. Stat. Comput. (2016), v tisku, doi: 10.1007/s11222-016-9683-9.
- [3] SCHLADITZ, K., SÄRKKÄ, A., PAVENSTÄDT, I., HAFERKAMP, O., MATTFELDT, T.: *Statistical analysis of intramembranous particles using freeze fracture specimens*. J. Microsc. 211 (2003), 137–153.
- [4] STOYAN, D., KENDALL, W. S., MECKE, J.: *Stochastic geometry and its applications*. 3. vydání, John Wiley, Chichester, 1995.