Pár „výpisků z Lehmanna“

Daniel Hlubinka

Univerzita Karlova
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

Mikuklášské setkání 2020

STATISTICKÉ ROZHODOVÁNÍ

- Naše každodenní činnost prováděná na základě dostupné informace a zkušeností.
- Otázky, na které hledáme odpovědi mohou být od celkem otevřených (kdy a jaké koupit auto), přes jasné otázky (kudy jet z Prahy do Ostravy v pondělí ráno), až po cílené experimenty.
- Čím přesnější je otázka a máme pro ni dostupná data, tím snáze můžeme formulovat hypotézu a alternativu a ty statisticky otestovat.
- Každé statistické rozhodování je optimalizační problém závislý na ztrátové funkci.

Test jako rozhodovací problém

- Uvažujeme dva výroky o jednom parametru (parametr může být klidně i funkce, ne jenom skalár).
- Hypotéza/alternativa - dva navzájem se vylučující (a doplňující) výroky *v rámci modelu*.
- Rozhodnutí je hodnota $d \in (d_0, d_1)$. Rozhodnutí d_0 znamená, že nadále budeme pracovat s hypotézou, rozhodnutí d_1, že nadále pracujeme s alternativou.

Test jako optimalizační problém

- Ztrátová funkce je speciální, ztráta při správném rozhodnutí je 0, ztráta při špatném rozhodnutí je 1.
- Rozhodovací pravidlo, které minimalizuje riziko (střední hodnotu ztrátové funkce) pro všechna možná rozdělení v rámci modelu, není možné obvykle sestrojit.
- Proto se obvykle uvažuje pouze test s předem daným omezením pro pravděpodobnost chyby 1. druhu—riziko zamítnutí platné hypotézy. Známe jako *hladinu významnosti testu*.
- Mezi těmito testy se hledá optimální pro riziko nezamítnutí platné hypotézy. Pravděpodobnost chyby druhého druhu, $1-$síla testu.

Eficience / vydatnost

- Model a optimální test jsou provázané.
- Co když model neplatí? Robustní varianty testů.
- Jak moc ztratíme, když použijeme jiný test a přitom model platí?
- Dobrý test by měl být konzistentní, tedy síla pro každou alternativu roste s rozsahem výběru (limitně k 1).
- Pokud pro pevnou alternativu a předepsanou sílu potřebujeme rozsah výběru n_1 pro test ϕ_1 a n_2 pro test ϕ_2, pak n_1/n_2 je indikátor „výkonnosti“ testu ϕ_2 vůči testu ϕ_1.

FORMÁLNÍ ZAVEDENÍ (PITMANOVY) ARE

- Uvažujme hypotézu $H_0 : \theta = \theta_0$ proti $H_1 : \theta > \theta_0$.
- Testová statistika T_n testu ϕ nechť splňuje

$$n^{1/2}\big(T_n - \mu(\theta_n)\big) \to N(0, \sigma^2)$$

  pro libovolnou $n^{1/2}(\theta_n - \theta_0) = O(1)$. Funkce μ je diferencovatelná (stačí zprava v θ_0). Označme $n(\theta)$ rozsah výběru k dosažení síly β při hladině α.
- Je-li $\tilde{T}_n$ jiná (asymptoticky normální) testová statistika, pak

$$\lim_{\theta \searrow \theta_0} \frac{n(\theta)}{\tilde{n}(\theta)} = \left(\frac{\tilde{\mu}'(\theta_0)/\tilde{\sigma}}{\mu'(\theta_0)/\sigma} \right)^2$$

  je asymptotická relativní eficience testu $\tilde{\phi}$ vůči testu ϕ.
- Všimněme si, že výraz vpravo nezávisí na α a β, je důsledkem asymptotické normality.

Jednovýběrové testy polohy

- Uvažujme diferencovatelnou hustotu f symetrickou kolem 0.
- Uvažované rozdělení má hustotu $f(x-\theta)$.
- Testujeme $H_0 : \theta = 0$ proti $H_1 : \theta > 0$.
- Použijeme t-test, znaménkový test a jednovýběrový Wilcoxonův test.
- Označme T_n, S_n a W_n příslušné testové statistiky. Pro všechny máme lokální asymptotickou normalitu.

- Nechť σ_f^2 je (konečný) rozptyl rosdělení s hustotou f.
- Uvažujme posloupnost lokálních alternativ $\theta_n = h/n^{1/2}$.
- Platí

$$n^{1/2} T_n \to N(h/\sigma_f, 1)$$
$$n^{1/2} S_n \to N(hf(0), 1/4)$$
$$n^{1/2} W_n \to N(2h \int f^2 dx, 1/3)$$

- Odtud $e_{S,t}(f) = \big(2f(0)\sigma_f\big)^2$ a $e_{W,t}(f) = 12\sigma_f^2 \big(\int f^2 dx\big)^2$.

Porovnání

- Pro normální rozdělení je t-test podstatně lepší než znaménkový, $e_{S,t}(N) \doteq 0,637$, ale jen mírně lepší než Wilcoxonův, $e_{W,t}(N) \doteq 0,955$.
- Pro Laplaceovo rozdělení je t-test podstatně horší než znaménkový, $e_{S,t}(L) = 2$, je horší i než Wilcoxonův test, $e_{W,t}(L) = 3/2$.
- Wilcoxonův test pro symetrické diferencovatelné hustoty není nikdy o moc horší než t-test, $e_{W,t}(f) \geq 0,864$.
- V normálním rozdělení tedy platí, že pokud k dosažení určité síly potřebujeme pro Wilcoxonův test 1000 pozorování, u t-testu je potřeba 955, tedy téměř stejné množství. U Laplaceova rozdělení je ale na 1000 pozorování pro Wilcoxonův test potřeba 1500 pozorování pro t-test.

Dvouvýběrové testy polohy

- Předpokládejme dvě rozdělení F a G, pro něž platí $G(x) = F(x - \theta)$.
- Opět uvažujme test $H_0 : \theta = 0$ proti $H_1 : \theta > 0$.
- Porovnáme opět t-test s Wilcoxonovým dvouvýběrovým a van der Waerdenovým testem (neparametrický test s normálními skóry).
- Předpokládejme stejný rozsah výběrů.
- Wilcoxonův test si nevede špatně. Pro normální rozdělení se stejnými rozptyly platí $e_{W,t}(N) \doteq 0,955$ a i zde $e_{W,t}(f) \geq 0,864$ pro všechna rozdělení.
- Test s normálními skóry je ještě lepší, $e_{vdW,t}(N) = 1$ pro normální rozdělení (nepřekvapivě), vždy ale platí $e_{vdW,t}(f) \geq 1$, což je zajímavý výsledek.

Jak „poplést“ t-test

- Pro symetrická rozdělení je t-test obvykle velmi dobrý.
- Vezměme šikmá rozdělení F a G a posuňme je proti sobě, například $G(x) = F(x - \theta)$, takže $\mathsf{E}_G = \mathsf{E}_F - \theta$. Udělejme výběr X_i z F a pak výběr Y_i z $\gamma G + (1 - \gamma)U$, kde γ je dost malá hodnota a rozdělení U volíme tak, aby $\mathsf{E}_{\gamma G + (1-\gamma)U}$ bylo co nejblíže E_F.
- Test o „shodě rozdělení“ pak časti vyjde tak, že t-test hypotézu nezamítá a Wilcoxonův test hypotézu zamítá.
- Ale zmátli jsme opravdu t-test? Ne, protože mezi středními hodnotami opravdu významný rozdíl není.
- Uvědomme si však, že jsme se dostali mimo rámec modelu pro Wilcoxonův i t-test.

OBECNĚJŠÍ POJETÍ ARE

- U Pitmanovy ARE vycházíme z asymptotické normality.
- Obecně ARE závisí na hladině i požadované síle: $e_{\tilde{\phi},\phi}(\alpha, \beta, \theta)$.
- Zkoumat eficienci se všemi třemi volnými parametry je náročné.
- Obvykle se uvažují dva parametry pevné a třetí limitně se blížící k nějaké hodnotě.
- Bahadurova eficience (pro $\alpha \to 0$) a Hodgesova-Lehmannova eficience (pro $\beta \to 1$) jsou nicméně považovány za nepříliš významné v pro praktické porovnání testů. Proto se nejčastěji využívá Pitmanova ($\theta \to \theta_0$).

Obrana t-testu

- Občas slýcháme, že pro velké rozsahy dat nevadí porušení normality, protože platí asymptotická normalita a t-test může být použit.
- To je naprostá pravda, t-test opravdu má tvar asymptotického testu pro střední hodnoty (s pomocí Sluckého věty).
- Pořadové testy se složitě vyrovnávají se shodami, t-test můžeme použít i pro výběry z alternativního rozdělení.
- Používejme tedy t-test, ale pro test správné hypotézy proti správné alternativě. Pro normální rozdělení máme přesný test, pro nenormální asymptotický *test shody středních hodnot*. Při porušení předpokladu shody rozptylů lze najít vhodnou úpravu testové statistiky.

A co hladina?

- Velikost *s* testu je supremum pravděpodobnosti zamítnutí (platné) hypotézy. Tedy přes všechny hodnoty v hypotéze.
- V modelu by samozřejmě mělo platit, že $s \leq \alpha$.
- Mějme například hypotézu $H_0 : \mu = 0$ v normálním rozdělení. Velikost t-testu je α.
- Pokud však model neplatí, kritické hodnoty jsou stanovené špatně a velikost testu může nabývat nekontorlovaných hodnot.
- Vezměme třídu $\mathcal{P}_v$ všech rozdělení s konečným rozptylem a označme $\mathcal{P}_{v,0}$ ty s nulovou střední hodnotou.
- Použijme jednovýběrový t-test pro test hypotézy $H_0 : \mu = 0$ proti jednostranné nebo oboustranné alternativě.

NEBO T-TEST NEPOUŽÍVEJME?

- V rámci $\mathcal{P}_{v,0}$ je velikost t-testu 1.
- To ale znamená, že pro libovolné $\eta > 0$ existuje rozdělení s konečným rozptylem a nulovou střední hodnotou, pro které t-test zamítne hypotézu $H_0 : \mu = 0$ s pravděpodobností alespoň $1 - \eta$. To je ale špatná zpráva!
- Standardní protipříklad vychází z extrémně nesymetrického rozdělení.
- Předpoklad symetrie ale také nestačí.
- Teprve symetrie a unimodalita spolu s malou hladinou α zaručí, že velikost t-testu je omezena pravděpodobností zamítnutí nulové hypotézy v případě rovnoměrného rozdělení na intervalu $[-1, 1]$.

Předpoklady jsou důležité. Ale jak je ověřit?

- Lze najít třídu $\mathcal{P}_{UI}$ rozdělení, pro kterou t-test má velikost α. Podmínka pro členství v této třídě je

$$\lim_{\lambda\to\infty} \sup_{F\in\mathcal{P}_{UI}} \mathsf{E}_F \left(\frac{|X-\mu_F|^2}{\sigma_F^2} \mathbf{1}\left\{ \frac{|X-\mu_F|}{\sigma_F} > 0 \right\} \right) = 0.$$

- Není těžké, případně je nějak možné, ověřit tento předpoklad pro určité F.
- Jak ale toto ověřit, když mám v ruce jen „pár“ (čtěme: konečně mnoho) hodnot?

S respektem k datům

- Data jsou obvykle to jediné, čemu můžeme alespoň trochu věřit.
- Můžeme je považovat za náhodný výběr?
- Kde se v nich bere neurčitost, proč jsou náhodná?
- Byla už podobná data analyzována? Jakými prostředky a s jakým výsledkem?
- Co je přesně otázkou, na kterou hledáme v datech odpověď? Jak tuto otázku formulovat v jazyce statistických hypotéz?
- Mohu považovat předpoklady metod za splněné? A co se stane, když ne?
- ...
- Data jsou „jen“ čísla. Konečně mnoho čísel s konečnou přesností. Metoda „nepozná“, odečítá-li hrušky od jablek a dělí je švestkami. Metoda dá výslednou hodnotu (pokud například nedělíme nulou), ale spolehlivost výsledku a jeho vypovídající hodnotu musí zaručit člověk. Alespoň zatím.