NMST434 CVIČENIA: OTÁZKY A ODPOVEDE

STANISLAV NAGY

M- a Z-odhady (týždeň 7)

Otázka. *V prvom cvičení, hneď prvý príklad ste napísali, že $\bar{X}$ je odhad riešenia rovnice* $\mathrm{E}(X_1 - \mu) = 0$. *Čomu ale nerozumiem - pretože, ako to chápem ja, tak M-odhad identifikuje síce μ_x, avšak toto μ_x práve minimalizuje funkciu ρ, nie? V skratke - nerozumiem to prepojenie toho M-odhadu a tej strednej hodnoty.*

Maximálne vierohodné odhady sú špeciálnym prípadom M-odhadov. Nie všetky M-odhady vychádzajú z maximálne vierohodných odhadov, pre M-odhad si môžem zvoliť akúkoľvek funkciu ρ (resp. ψ pre Z-odhad), a typicky M-odhady priamo konštruujem tak, že začnem s funkciou ψ s rozumnými vlastnosťami (napr. obmedzená).

Vzťahy MLE a M-odhadov. V prvom príklade (to je $N(\mu, 1)$ v mojich poznámkach) začnem z MLE odhadu. Vidím, že MLE odhad v modele $N(\mu, 1)$ je definovaný ako M-odhad s funkciou $\rho(x, t) = (x - t)^2$, teda M-odhad identifikuje (odhaduje) parameter rozdelenia daný ako $\arg\min_t \mathrm{E}(X - t)^2$. Vieme že toto $\arg\min$ je práve $t = \mathrm{E}\, X$, takže skutočne M-odhad odhaduje parameter strednej hodnoty. To, že M-odhad identifikuje riešenie rovnice $\mathrm{E}(X_1 - \mu) = 0$ dostávam ak formulujem MLE ako Z-odhad pre $\psi(x, t) = (x - t)$. Vtedy skutočne Z-odhad identifikuje riešenie rovnice $\mathrm{E}\, \psi(X, t) = \mathrm{E}(X - \mu) = 0$, a riešením je opäť $\mu = \mathrm{E}\, X$.

Predpoklad, že dáta sú z $N(\mu, 1)$ ale vôbec nemusím využívať. Ak by som predpokladal akékoľvek rozdelenie X, a začal priamo s M-odhadom daným funkciou $\rho(x, t) = (x - t)^2$, vidím, že pre túto funkciu dostanem ako M-odhad $\bar{X}$ (výberový priemer), a tento M-odhad identifikuje parameter $\mathrm{E}\, X$, teda strednú hodnotu. To všetko vychádza priamo z teórie M-odhadov, bez toho aby som akokoľvek použil že som motivačne začal s normálnym rozdelením, alebo s MLE.

To, že ak začnem s MLE pre parameter t, a potom MLE odhad vyjadrím ako M-odhad, a nakoniec sa ukáže že M-odhad skutočne identifikuje parameter t, platí vždy (za splnenia predpokladov MLE, napr. toho že skutočné rozdelenie pochádza z modelu ktorý používam). To je vidieť z toho, že ak $f(x, t)$ je združená hustota náhodného výberu tak MLE definujem ako Z-odhad daný ako riešenie (f' je derivácia f podľa t)

$$f'(x, t)/f(x, t) = 0$$

takže $\psi(x, t) = f'(x, t)/f(x, t)$, a Z-odhad identifikuje parameter $\mathrm{E}\, f'(X, t)/f(X, t) = \int \frac{f'(x,t)}{f(x,t)} f(x, t) \,\mathrm{d}\, x = \int f'(x, t) \,\mathrm{d}\, x$, a ten posledný integrál bude 0 ak zameníte poradie derivácie a integrálu. Takže skutočne, za platnosti predpokladaného modelu, váš M- (resp. Z-) odhad správne identifikuje parameter t.

Výhoda toho, že sa na MLE pozeráte ako na M-odhad je tá, že v odvodení as. rozdelenia M-odhadov nepredpokladáte platnosť modelu v ktorom ste začínali pri MLE. As.

E-mail address: `nagy@karlin.mff.cuni.cz`.

Date: 5. mája 2020.

rozptyl je daný sandwichovým odhadom, ktorý je zložitejší, ale platí aj pri porušení predpokladov MLE. Takže získavate možnosť, ako napr. získať MLE odhady, ale v prípade, že platnosť základných predpokladov nie je splnená, stále dokážete konzistentne odhadovať as. rozdelenie M-odhadu a robiť testy.

Otázka. *Prvá otázka generuje vlastne túto otázku - tá identifikovateľnosť znamená, že sa snažíme nájsť odhady s určitými vlastnosťami? Napr. v normálnom rozdelení, sigma nijak neovplyvňuje tú identifikovateľnosť ? Resp. obecne nejaký rušivý parameter ?*

Identifikovateľnosť znamená, že ak chcem odhadovať (alebo testovať) o parametre θ, a test založím na M-odhade danom funkciou $\rho(x,t)$, tak $\theta = \arg\min_t \mathrm{E}\,\rho(X,t)$. Takže moje M-odhady skutočne zodpovedajú parametru, ktorý ma zaujímal na začiatku. To musím vždy overiť, pretože inak M-odhady ktoré dostávam budú v skutočnosti odhadovať niečo iné ako parameter θ ktorý ma zaujíma, a sú pre môj problém úplne nanič.

V MLE pre normálne rozdelenie $N(\mu,\sigma^2)$ (druhý príklad na prvej strane poznámok) vidíte, že funkcia $\rho(x,\mu,\sigma^2) = (x-\mu)^2/\sigma^2 + \log(\sigma^2)$. Pre ľubovoľnú hodnotu σ^2 bude tento výraz minimalizovaný v μ v strednej hodnote práve vtedy ak $\mathrm{E}(X-\mu)^2$ je minimalizované, takže opäť identifikujeme $\mu = \mathrm{E}\,X$ a všetko je OK. Pre druhú zložku, po zderivovaní rho podľa σ^2 a jednoduchej úprave dostávam že druhý parameter je identifikovaný ako riešenie $\mathrm{E}(X-\mu)^2 - \sigma^2 = 0$, z prvej zložky viem že μ identifikuje $\mathrm{E}\,X$, a teda riešenie v σ^2 je skutočne $\mathrm{E}(X-\mathrm{E}\,X)^2$, rozptyl. Opäť OK.

Otázka. *Trochu nerozumiem, ako si tú domácu úlohu 10 formulovať. Ja sa to snažím nejak spojiť s vaším príkladom z cvičení - príklad estimaton of location and scatter.*

Nie, v DÚ žiadny scatter parameter nie je. Iba location parameter. Máte daný nejaký model, tj nejaké rozdelenie s parametrami ϵ a ν, a tri rôzne odhady parametra polohy ktoré sa dajú vyjadriť ako M-odhady. Vašou úlohou je vyjadriť tieto odhady ako M-odhady, a použiť vety 9 alebo 10 na odvodenie as. rozdelenia týchto M-odhadov, za platnosti tohto predpokladaného modelu. Potom chcem nech sa zamyslíte aké parametre vo vašom modele tieto tri rôzne M-odhady identifikujú. Ide o rovnaký parameter, alebo odhadujete vždy niečo iné? A nakoniec, ak ide o rovnaký parameter, ktorý z odhadov sa zdá byť najlepší v zmysle najmenšieho as. rozptylu? Samozrejme v závislosti na parametroch modelu ϵ, ν, a možných ďalších parametroch ktoré sa pri výpočte objavia.

Nakoniec sa zamyslite prečo sa zdá, že pre nejaké hodnoty ϵ, ν je najlepší ten alebo ten odhad. Dávajú tieto asymptotické výsledky nejaký zmysel? Máte k týmto výsledkom nejakú intuíciu?

Otázka. *Chcel som Vás poprosiť o hint, ako sa generujú zmesi rozdelení.*

Zmesi rozdelení sa typicky generujú dvojfázovo. Najprv si generujete náhodný výber z $Alt(p)$ rozdelenia, a každej zložke ak dostanete 0 generujete X z distribučnej funkcie F_0, ak dostanete 1 generujete X z distribučnej funkcie F_1. Overte si, že výsledné rozdelenie z ktorého generujete X má (nepodmienene) distribučnú funkciu $p\,F_1 + (1-p)\,F_0$.

M- a Z-odhady (týždeň 8)

Otázka. *Ve vašich poznámkách k M- a Z-odhadům na webu na straně 6 dole - symboly X a Y používáte jako označení generického vektoru z něhož pochází ten náhodný výběr? Tedy Y je náh. veličina, X je náh vektor délky p? Pak při výpočtu $\psi(x,y,\beta) = -2x(y - x^{\mathsf{T}}\beta)$ platí, že násobím číslo (-2) krát vektor (x) krát číslo $(y - x^{\mathsf{T}}\beta)$? A podobně ve výpočtu Sigma matice: $X(Y - X^{\mathsf{T}}\beta)(Y - X^{\mathsf{T}}\beta)^{\mathsf{T}}X^{\mathsf{T}}$ je vlastně vektor*číslo*číslo*vektor?*

Áno, X a Y sú generické náhodné veličiny, tj X je stĺpcový vektor s dimenziou povedzme p, a Y je skalár. Ďalej presne ako píšete - ψ je stĺpcový vektor p funkcií, a Σ je $p \times p$ matica.

Otázka. *V úkolu HW11, můžeme předpokládat homoskedasticitu — tj.* $\mathrm{var}(Y|X) = \mathrm{var}(\varepsilon|X) = \sigma^2$*? Nebo máme obecně předpokládat, že* $\mathrm{var}(Y|X) = \sigma^2(X)$*, tedy nějaká funkce (tj. heteroskedasticita)? Pokud nemáme homoskedasticitu, jak můžu porovnat odhad s odhadem LAD, který má tvar využívající hustotu* ε *(tedy pro normální rozdělení i její rozptyl). Napadá mě pouze použít odhad ve stejném tvaru, jen použít podmíněnou hustotu* $\varepsilon|X$*, čímž pak tato hustota může mít rozptyl* $\sigma^2(X)$*. Uvažuju správně?*

V predpokladoch homoskedasticita nie je, takže ak riešenie dokážete odvodiť aj bez nej, nepredpokladajte ju. Ak sa to ukáže ako zásadný problém, môžete si príklad zjednodušiť a pridať rozumné ďalšie predpoklady. V porovnaní s LAD sa bude stačiť obmedziť na homoskedastický model, čo bude špeciálny prípad výsledkov z (i)-(iii). Nemusíte odvodzovať sandwichové odhady pre heteroskedastické LAD.

Otázka. *K zadání HW12: Pochopil jsem to správně, že v (i) máme 2 rovnice a máme určit kdy* $a = \alpha$ *a zároveň* $b = \beta$*? Potom v úloze (ii) tedy máme pouze 1 rovnici (derivovanou podle b, resp a) a máme pro ni určit, kdy* $b = \beta$ *resp.* $a = \alpha$*? Postupuji podobně jako ve vašich poznámkách na straně 7, využívám derivace, protože funkce* ρ *je konvexní a identifikuju pomocí toho vztahu pro* ψ*. Je tahle úvaha správně, nebo jsem špatně zadání pochopil?*

Áno, v časti (i) máte určiť podmienky za ktorých platia obe rovnice zároveň. V časti (ii) slabšie podmienky, kedy stačí, aby platila jedna z rovníc. Postup bude podobný ako v poznámkach na str. 7.

EM-algoritmus (týždeň 10)

Otázka. *HW14: Čo je myslené ako* measure of uncertainty*? Je to niečo ako metrika, ktorá má zobraziť úspešnosť/neúspešnosť klasifikátora?*

Ak sa bod x rozhodnete zaradiť do zmesi i, dokážete nejak číselne vyjadriť ako „istý“ si týmto zaradením ste? Ak to dokážete nejakým spôsobom kvantifikovať (napr. u pozorovania x som si na 99% istý že patrí do skupiny 1) tak táto charakteristika je to, čo myslíme „measure of uncertainty“.