POSUDOK OPONENTA BAKALÁRSKEJ PRÁCE

Název: Zamítací metoda pro generování vzorků ze složitých rozdělení
Autor: Tomáš Rajtmayer

ZHRNUTIE OBSAHU PRÁCE

Predkladaná bakalárska práca predstavuje dve metódy generovania (pseudo-)náhodných čísel zo zadaných rozdelení: inverznú a zamietaciu metódu. Zamietacia metóda je následne použitá v simulačnom teste normality, pri testovaní, či pravdepodobnostné rozdelenie s hustotou danou profilom hory Říp by sa dalo považovať za normálne rozdelenie.

CELKOVÉ ZHODNOTENIE PRÁCE

Zadanie práce, tak ako bolo formulované, bolo splnené. Teoretická časť (kapitoly 1 a 2) je štandardným učebnicovým materiálom v kurzoch Monte Carlo metód; autorovo podanie sleduje knihu Devroya, dôkazy prevzaté odtiaľ sú rozpísané do o niečo väčších podrobností. Aplikovaná časť (kapitola 3) je spracovaná korektne; pri jej čítaní ma napadlo, v akom smere by sa teoretická časť mohla trochu rozšíriť, čo mi dalo tiež podnet na dve otázky k obhajobe (viď sekcia nižšie).

Práca je spracovaná pomerne starostlivo a na kultivovanej úrovni, autor preukázal schopnosť písať matematický text s vysvetľujúcim komentárom a príslušnými referenciami. Autor očividne pozná zásadu, že aj odsadené formuly sú súčasťou gramatických viet v texte, a tým pádom musia v prípade potreby obsahovať interpunkciu — no nie vždy túto zásadu dôsledne dodržiava. Jazykové prehrešky som nezaznamenal, preklepy sa samozrejme nájdu (napr. prvý riadok na str 4), ale sú zriedkavé.

OTÁZKA K OBHAJOBE

Predpokladajme, že namiesto hustoty $f(x)$ máme danú funkciu $p(x)$, ktorá nemusí byť hustotou pravdepodobnosti; vieme len, že je *nejakým* kladným násobkom *nejakej* hustoty pravdepodobnosti, teda že

$$p(x) = Kh(x), \quad \text{kde} \quad 0 < K < \infty \text{ a } h(x) \text{ je hustota pravdepodobnosti}$$

— no nepoznáme K ani $h(x)$. Samozrejme, príslušné K sa v princípe dá nájsť integrovaním funkcie $p(x)$, no takáto úloha nemusí byť riešiteľná analyticky, a numericky môže byť príliš náročná; budeme teda odteraz predpokladať, že takýto postup nie je v danej situácii prakticky možný.

Popísaná situácia nastáva typicky v bayesovskej štatistike, keď aposteriórne rozdelenie dostaneme veľmi jednoducho ako násobok apriórneho a vierohodnosti — až na konštantu K, ktorej numerický výpočet môže znamenať náročný výpočet, v prípade väčšieho počtu odhadovaných parametrov mnohorozmerného, integrálu. Toto motivovalo vývoj simulačnej metodológie MCMC založenej na Markovových reťazcoch, v ktorej sa znalosť normujúcej konštanty nevyžaduje a algoritmy sú schopné generovať postupnosť náhodných čísel s hustotou $h(x)$ bez znalosti K, pracujúc priamo s funkciou $p(x)$. Skutočnosť, že takúto metodológiu bolo treba vyvinúť, podsúva záver, že klasické algoritmy, popisované v predkladanej bakalárskej práci, túto schopnosť nemajú — inými slovami, nedokážu pracovať výlučne s funkciou $p(x)$, potrebujú poznať aj K.

Predpokladajme teda, že by sme chceli generovať náhodné čísla s hustotou $h(x)$, poznajúc iba funkciu $p(x)$, pomocou niektorej z metód uvedených v bakalárskej práci. V prípade inverznej metódy by sme museli najsamprv odvodiť distribučnú funkciu prislúchajúcu hustote $h(x)$ — na to by sme museli zvládnuť výpočet príslušných integrálov, a v takom prípade by očividne nebol problém vypočítať aj konštantu K. Takže inverznú metódu môžeme z našej diskusie vylúčiť (v mnohorozmernom prípade je aj tak nepoužiteľná).

Otázka je: ako je to v prípade zamietacej metódy? Pokúste sa vo svojej odpovedi podať jednak stručné (primerane možnostiam obhajoby) matematické zdôvodnenie, jednak krátku diskusiu možných praktických (analyticko-numerických) aspektov. Súvisí táto otázka nejako s kapitolou 3 (simulácia "rozdelenia Řípu")?

Záver

Prácu odporúčam uznať ako bakalársku prácu na MFF UK.

Edmonton, 26. 8. 2024

Ivan Mizera
KPMS MFF UK
imizera@me.com