

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Scenario generation methods for discrete data

Autor: Bc. Ondřej Komora

SHRNUTÍ OBSAHU PRÁCE

Práce se zabývá problematikou generování scénářů za účelem stochastické optimalizace v případě, že stochasticita je generovaná náhodným vektorem s mnohorozměrným diskrétním rozdělením. Práce vychází z přístupu pomocí copulí, který byl prezentovaný v *M. Kaut. A copula-based heuristic for scenario generation. Computational Management Science, 11:503–516, 2014.* Tento přístup (navržený pro spojitá data) však modifikuje na diskrétní data s využitím tzv. *checker-board* copule, která představuje přirozenou volbu copule v případě diskrétních marginálních rozdělení.

Tématem práce jsou přístupy ke generování scénářů pro stochastickou optimalizaci v případech, kde uvažovaná náhodnost je diskrétního typu. Toto téma byla naplněno.

Vlastní příspěvek Autor přehledně popsal konstrukci výše zmíněné *checker-board* copule a ukázal, jak tato copule souvisí s randomizací, které diskrétní náhodný vektor převede na spojitý náhodný vektor, pro který je již copule jednoznačně určena.

Dále modifikoval metodu generování scénářů pomocí copule (spojitého rozdělení) z výše uvedeného článku *Kaut (2014)* pro diskrétní rozdělení. Toto rozšíření pak ilustroval na optimalizačním problému batohu, kdy do hry vstupuje náhodnost prostřednictvím toho, zda daná položka bude k dispozici a také prostřednictvím náhodných cen.

Ocenil bych také ilustrační příklady a grafy.

Matematická úroveň. Z matematického hlediska považuji práci za spíše obtížnější. Tato obtížnost spočívá zejména v technické náročnosti práce s inverzemi distribučních funkcí diskrétních náhodných veličin a následně pak v konstrukci *checker-board* copule. To vyžadovalo značnou pečlivost. Nicméně práce se nevyhnuly některé nepřesnosti či nejasnosti.

Práce se zdroji. Na kolik jsem schopen posoudit, tak všechny použité zdroje jsou řádně citovány.

Formální úprava. Formální úroveň práce je velmi dobrá, množství překlepů je malé. Práce obsahuje jisté drobné nedostatky v citacích a také některé rovnice by se daly „vysázet“ přehledněji.

CELKOVÉ HODNOCENÍ PRÁCE

Práci považuji za nadprůměrnou. Autor očividně pracoval velmi pečlivě. Pouze na některých místech by si práce zasloužila více explicitních vysvětlení, aby byla přístupná i širšímu okruhu čtenářů.

PŘIPOMÍNKY A DOTAZY

1. str. 4⁸: Autor využívá značení $\text{supp } X$, které však vysvětluje až o několik řádků níže.
2. str. 10: Definice výběrové verze Kendallova korelačního koeficientu není správně. Dále výběrovou verzi Spearmanova korelačního koeficientu bych zavedl jinak. Použitý vzorec je výpočetní tvar, který platí pouze tenkrát, když v datech nejsou shody.

3. str. **10** Theorem 1.4: Tato věta je sice matematicky pravdivá, nicméně s její následnou interpretací bych byl opatrnější. Problém je ten, že definice populačních/teoretických verzí Kendallova a Spearmanova korelačního koeficientu (viz Definition 1.4) není vhodná pro případ, že některá z náhodných veličiny X, Y má diskrétní rozdělení, viz např. Nešlehová, J. (2007). *On rank correlation measures for non-continuous random variables. Journal of Multivariate Analysis, 98(3), 544–567.*
4. str. **15** poslední odstavec: Není jasné, co autor rozumí slabou limitou (*weak limit*) odhadu copule \hat{C}_m^* .
5. str. **24**: Mám pocit, že problém je již v tom, že funkce $C(u,v)$, kterou autor spočítal pomocí vzorce (1.2) není copule. Takže je otázka, zda by vůbec mohlo dávat smysl používat tuto funkci v algoritmu z Kapitoly 2.
6. Kapitola 4.2: Některé věci v této kapitole nejsou jasné. Jaký byl rozsah výběru historických dat? Jaká byla volba parametrů normálního rozdělení μ a Σ (str. 37)? Co jsou přesně *input data* (str. 39)? Jak je určeno η (představující skutečné rozdělení)?
7. str. **40** Figure 4.1: Vzhledem k definici *in-sample stability* ST_n mně není jasné, jak je možné, že tato míra je klesající, když počet scénářů roste.
8. str. **41** Figure 4.3: K tomuto grafu je zapotřebí znát optimální řešení x^* . V práci nějak nemohu najít, jak bylo toto optimální řešení nalezeno.

OTÁZKY K OBHAJOBĚ

Připravte si odpovědi na připomínky č. 2, 4, 6, 7 a 8.

ZÁVĚR

Domnívám, že práce splňuje nároky kladené na diplomovou práci na programu Pravděpodobnost, matematická statistika a ekonometrie a doporučuji ji za ni uznat.

doc. Ing. Marek Omelka, Ph.D.
 KPMS MFF UK
 1. června 2024