

VÝSLEDKY SIMULACE

SIMULACE S KONEČNÝM HORIZONTEM A SIMULACE DLOUHODOBÉHO CHOVÁNÍ.
REPLIKAČNÍ METODA, METODA SKUPINOVÝCH PRŮMĚRŮ, REGENERATIVNÍ METODA.
METODA SPOLEČNÝCH NÁHODNÝCH ČÍSEL A JEJÍ VYUŽITÍ PŘI SROVNÁNÍ VARIANT.

ANALÝZA VÝSLEDKŮ SIMULACE

Analýza výsledků simulace může být komplikovaná tím, že vstupní charakteristiky simulačního modelu jsou náhodné veličiny, takže i výstupy náhodné veličiny, proto je uvádíme jako bodový či intervalový odhad.

Výsledky se mohou významně lišit z následujících důvodů:

1. Vlivem náhodného kolísání: abychom eliminovali vliv náhodného kolísání, můžeme využívat vždy stejnou posloupnost náhodných čísel (v R je na to funkce `set.seed`), což ale není vždy proveditelné, nebo můžeme využít některé metody matematické statistiky (průměry, odchylky...);
2. Vlivem různých počátečních podmínek: lze vynechat počáteční pozorování, nebo eliminovat jejich vliv replikační metodou, metodou skupinových průměrů či regenerativní metodou;
3. Vlivem experimentálních zásahů: ty samozřejmě většinou naopak eliminovat nechceme.

Problém je, že v simulaci často pracujeme s daty, která jsou nestacionární (jejich rozdělení se mění v čase) či autokorelovaná (když jeden zákazník čeká dlouho, bude zákazník po něm nejspíš také čekat dlouho). Řešení autokorelace závisí na tom, zda jde o simulaci s konečným horizontem, či o simulaci dlouhodobého chování.

SIMULACE S KONEČNÝM HORIZONTEM A SIMULACE DLOUHODOBÉHO CHOVÁNÍ

1. **U simulace s konečným horizontem** známe počáteční stav a událost, která simulaci ukončuje (čas uzavření ochodu či pobočky, počet průchozích prvků – např. vyrobených výrobků, náhodná událost – vyčerpání zdrojů apod.). Problém autokorelace řešíme replikační metodou: simulační běh několikrát opakujeme a bodový a intervalový odhad stanovíme na základě průměrných hodnot v jednotlivých bězích. Průměry z nezávislých replikací jsou totiž nezávislé, takže problém autokorelace zde nehrozí. Metoda může být ale poměrně výpočetně náročná.

Když chceme odhadnout střední hodnotu μ pro veličinu X , postupujeme následovně. Uvažujme, že máme k replikací ($i = 1, 2, \dots, k$) a v i -té replikaci máme n_i veličin. V jednotlivých replikacích můžeme spočítat odhad střední hodnoty veličiny X jako:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad i = 1, 2, \dots, k.$$

kde x_{ij} je j -tá hodnota veličiny X v i -té replikaci. Bodový odhad střední hodnoty μ přes všech k replikací spočítáme jako:

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i$$

Protože střední hodnoty z jednotlivých replikací jsou nezávislé veličiny, nebezpečí autokorelace už nehrozí. Intervalový odhad získáme jako

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, k-1} \frac{s}{\sqrt{k}}$$

kde s^2 je výběrový rozptyl stanovený podle vzorce:

$$s^2 = \frac{1}{k-1} \sum_{i=1}^k (\bar{x} - \bar{x}_i)^2$$

2. U **simulace s dlouhodobého chování** neexistuje žádná událost, která by ji ukončila. Cílem je analýza běžného, dlouhodobého provozu. Musíme se tedy nějak vypořádat s počátečním, přechodným (tranzientním) stavem, což je stav, než se systém dostane do svého běžného provozu (na počátku mohou být fronty prázdné apod.). Můžeme použít opět **replikační metodu**, ale musíme v ní vynechat několik počátečních pozorování. Pokud chceme odhadnout střední hodnotu μ pro veličinu X , provedeme opět několik simulačních běhů. Z každého i -tého běhu (replikace) získáme odhad střední hodnoty náhodné veličiny X :

$$\bar{x}_i = \frac{1}{n_i - l} \sum_{j=l-1}^{n_i} x_{ij} \quad i = 1, 2, \dots, k$$

kde k je počet replikací, x_{ij} je j -tá hodnota náhodné veličiny X v i -té replikaci, přičemž vynecháme l počátečních pozorování, n_i je počet pozorování v i -té replikaci.

Z těchto odhadů vypočteme celkový průměr a intervaly spolehlivosti. Bodový odhad střední hodnoty μ získáme jako:

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i$$

Intervalový odhad získáme jako:

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, k-1} \frac{s}{\sqrt{k}}$$

Kde s^2 je výběrový rozptyl stanovený podle vzorce:

$$s^2 = \frac{1}{k-1} \sum_{j=1}^k (\bar{x} - x_j)^2$$

Při použití **metody skupinových průměrů** na rozdíl od replikační metody uděláme pouze jeden simulační běh, ale rozdělíme jej na k stejně dlouhých úseků. Výhodou je, že neztrácíme tolik pozorování, protože počáteční stav vynecháme jen jednou. Nevýhodou je, že po sobě jdoucí úseky mohou být zkorelovány, takže riziko autokorelace se zmenší, ale stále tam je, zvláště pokud jsou úseky moc krátké. Delší úseky snižují pravděpodobnost výskytu autokorelace, ale zase tím ztrácíme stupně volnosti, protože počet úseků klesá. Bodový odhad střední hodnoty μ získáme jako průměr v každém z k úseků.

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i$$

Pak pro intervalový odhad platí:

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, k-1} \frac{s}{\sqrt{k}}$$

Regenerativní metoda je podobná metodě skupinových průměrů. Liší se od ní způsobem dělení na úseky, které na rozdíl od předchozí metody nemusí být stejně dlouhé. Další úsek začíná tehdy, když se systém dostane do tzv. regenerativního stavu (např. prázdná fronta). Z toho plyne, že autokorelace bude při použití této metody skutečně nulová. U složitějších systémů může být ale problém tento regenerativní stav najít. Částečným řešením je nacházet stavy, které jsou velmi podobné. Další nevýhoda je, že neznáme celkovou délku simulačního chodu, protože regenerativní cykly mohou být velmi dlouhé (při stanovené délce simulačního běhu nemůžeme zaručit potřebné množství pozorování).

Stejně jako v předchozím případě nejprve musíme vypočítat průměr z každého úseku, ale celkový průměr musí brát v úvahu nestejnou délku úseků. Bodový odhad střední hodnoty μ získáme jako:

$$\bar{x} = \frac{E(y_i)}{E(T_i)}$$

kde y_i je součet veličiny X v i -tém regenerativním cyklu a T_i je délka i -tého cyklu. Interval spolehlivosti získáme jako:

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, k-1} \frac{s}{\bar{T}\sqrt{k}}$$

$$s^2 = s_{11} - 2\bar{x}s_{12} + \bar{x}^2s_{22}$$

kde

$$s_{11} = \frac{1}{k-1} \sum_{i=1}^k (\bar{y} - y_i)^2 \text{ je výběrový rozptyl } y_i$$

$$s_{12} = \frac{1}{k-1} \sum_{i=1}^k (\bar{y} - y_i)(\bar{T} - T_i) \text{ je výběrová kovariance}$$

$$s_{22} = \frac{1}{k-1} \sum_{i=1}^k (\bar{T} - T_i)^2 \text{ je výběrový rozptyl } T_i.$$

JAK REDUKOVAT ROZPTYL VÝSLEDKŮ?

Cílem redukce rozptylu je zmenšení rozptylu výběrového průměru. Patří sem metoda společných náhodných čísel, metoda stratifikovaných výběrů a metoda protikladných veličin.

Metoda společných náhodných čísel spočívá v tom, že simulujeme různé varianty (například provoz s jednou a dvěma pokladnami), ale v každé variantě použijeme stejná náhodná čísla. To znamená, že rozdíly mezi variantami budou systémové, protože vliv náhody se tímto eliminuje. To, co nás skutečně zajímá, je rozdíl mezi variantami. Mějme dvě varianty X, Y a jejich rozdíl označme $Z = X - Y$. Pro rozptyl součtu či rozdílu náhodných veličin platí vztah

$$D(X \pm Y) = D(X) + D(Y) \pm 2\text{cov}(X, Y).$$

Pokud nepoužijeme společná náhodná čísla, kovariance X, Y bude nulová, v opačném případě bude ale kladná. Rozptyl rozdílu mezi variantami X, Y bude tedy v případě použití společných náhodných čísel menší, protože platí:

$$D(X) + D(Y) - 2\text{cov}(X, Y) < D(X) + D(Y)$$

Metoda stratifikovaných výběrů spočívá v zapojení nějakých apriorních informací (časté při výběrových šetřeních), kdy můžeme například předem omezit extrémní varianty, a rozptyl tak snížit.

Metoda protikladných veličin pracuje tak, že se spustí dva simulační běhy, v prvním se použijí náhodná čísla r a ve druhém pak $(1 - r)$. Důvod je ten, že můžeme mít někdy smůlu a vygenerovat velmi velká náhodná čísla. Proto pustíme druhý běh, kde budou tím pádem náhodná čísla naopak malá.

ZDROJE:

Přednášky 4EK421 Simulační modely ekonomických procesů, VŠE Praha, 2013.

Dlouhý, M. a kol.: Simulace podnikových procesů. Computer Press, Brno 2007.