

OMEZENÉ HISTOGRAMY A SBRA

Milan Guštar

Abstract

This paper describes a non-parametric form of random variables description by truncated histograms in SBRA method and a corresponding way of random numbers generation based on the inverse transformation method.

1. Reprezentace náhodných proměnných omezeným rozdělením

Metoda SBRA (Simulation Based Reliability Assessment) využívá pro určování spolehlivosti výpočetních postupů založených na stochastické simulaci a metodě Monte Carlo [4]. K provedení simulací třeba získat náhodná čísla - numerické realizace vstupních náhodných proměnných. Statistická rozdělení těchto proměnných lze definovat několika způsoby. Teoretická rozdělení lze popsat **parametrickým** způsobem, tedy definiční funkcí a hodnotami jejich parametrů [2]. Jelikož rozdělení mnoha reálných veličin neodpovídají dobře žádnému ze známých teoretických rozdělení, je pro jejich popis vhodnější **neparametrická** forma.

Mnoho reálných náhodných veličin nabývá pouze hodnot z omezeného intervalu $[a,b]$, proto lze k jejich popisu použít omezená (useknutá) rozdělení, definovaná pouze na tomto omezeném intervalu (viz obr. 1).



Obr. 1: a) Rozdělení na intervalu $(-\infty, \infty)$, b) Rozdělení na omezeném intervalu $[a,b]$

Kupříkladu veličiny jako délka, hmotnost apod. nemohou nabývat záporných hodnot i když často bývají popsány rozdělením definovaným na nekonečném intervalu $(-\infty, \infty)$. Podobně lze najít i další “rozumná” omezení definičního intervalu mnoha veličin. Například skutečná délka průmyslově vyráběného šroubu s nominální délkou 50 mm nemůže přesáhnout jistou maximální hodnotu ani nemůže být menší než jistá hodnota minimální a to ani v případě že je rozdělení délek tohoto šroubu považováno za normální rozdělení se střední hodnotou $\mu = 50$ mm a s definičním intervalem $(-\infty, \infty)$.

Dalším důvodem k omezení definičního intervalu náhodných proměnných může být požadavek generování těchto veličin pomocí počítače. V takových případech bývá generované rozdělení aproximací rozdělení požadovaného, a to jak z důvodu omezené přesnosti reprezentace čísel, tak i z důvodu jejich omezeného rozsahu. Rozsah

generovaných hodnot je třeba volit s ohledem na přijatelnou odchylku statistických parametrů generovaného rozdělení od hodnot požadovaných. Také je třeba porovnat pravděpodobnost vyloučení hodnot ležících mimo zvolený interval s předpokládaným počtem generovaných hodnot. Např. při generování 10^4 hodnot může být vyloučení hodnot ležících mimo zvolený interval s pravděpodobností 10^{-5} zanedbatelné, ale při generování 10^6 hodnot již může mít vyloučení okrajových hodnot významný vliv. Je však třeba si uvědomit, že pravděpodobnost současného výskytu více vzájemně nezávislých jevů je rovna součinu jednotlivých pravděpodobností jejich výskytu. Tedy např. již pravděpodobnost současného vyloučení dvou hodnot ležících mimo zvolený interval, a tedy zanedbání jejich možné interakce, je při pravděpodobnosti vyloučení každého z jevů 10^{-5} pouze 10^{-10} . Tedy ani značné omezení definičních intervalů generovaných rozdělení nemusí mít v některých případech mít významný vliv na přesnost výsledků.

2. Aproximace omezených rozdělení

Rozdělení definovaná na omezeném intervalu lze popsat neparametrickým způsobem pomocí histogramu. Histogram ve formě sloupcového grafu četností je grafickou reprezentací pravděpodobnostní funkce. Histogram lze snadno vytvořit následujícím postupem. Definiční obor pravděpodobnostní funkce je rozdělen na disjunktní intervaly - třídy. Nad každým intervalem je zkonstruován obdélník o ploše odpovídající pravděpodobnosti v dané třídě.

Vytvoření histogramu lze považovat za aproximaci původního rozdělení. Pokud je rozdělení hodnot uvnitř jednotlivých tříd považováno za rovnoměrné, je originální rozdělení aproximováno rozdělením **po částech rovnoměrným**. Pokud jsou všechny hodnoty uvnitř jednotlivých tříd reprezentovány jedinou hodnotou, je originální rozdělení aproximováno rozdělením **diskrétním** (viz obr. 2).



Obr. 2: Aproximace po částech rovnoměrným a diskrétním rozdělením

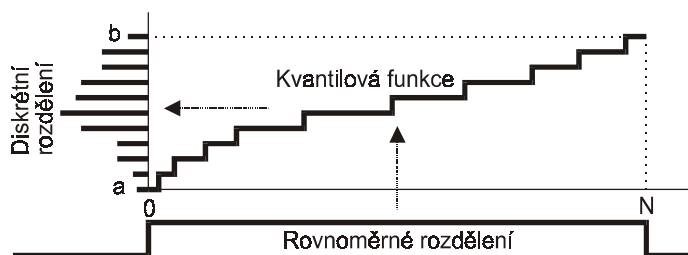
Histogramů je často používáno pro určení statistických rozdělení dat, získaných na základě měření, testů, zkoušek nebo historických záznamů. Histogram je vytvářen tříděním a statistickým zpracováním těchto dat. Získané empirické rozdělení v neparametrické formě bývá obvykle nahrazováno zvoleným rozdělením teoretickým [1]. Pokud je k popisu rozdělení a k jejich následnému generování při simulačních výpočtech používána neparametrická forma, lze jako výsledné rozdělení použít přímo zpracovaný histogram bez nutnosti jeho náhrady rozdělením teoretickým.

3. Generování rozdělení

Volba generátoru náhodných čísel a způsob jeho implementace ovlivňuje rychlost výpočtu i přesnost výsledků. Generování náhodných čísel s definovaným statistickým rozdělením se obvykle provádí ve dvou krocích. Nejprve je pomocí primárního generátoru generována posloupnost náhodných, vzájemně nezávislých čísel s rovnoměrným rozdělením. Z této posloupnosti je vhodnou transformací vytvořena posloupnost čísel s požadovaným rozdělením. Volba vhodné transformace a způsob její implementace je ve většině případů kompromisem mezi dosaženou přesností aproximace požadovaného rozdělení, rychlostí výpočtu a paměťovými nároky.

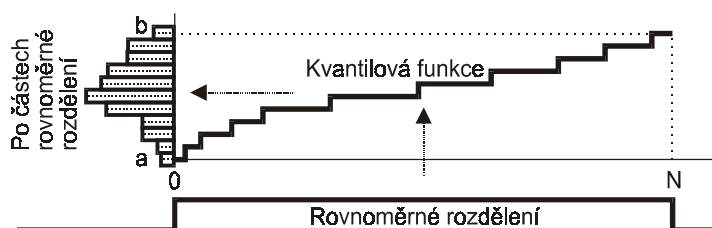
Pro generování obecného diskrétního rozdělení i rozdělení po částech spojitého lze využít metodu inverzní transformace [3]. Pokud generované rozdělení nabývá pouze hodnot z omezeného intervalu, lze potřebné hodnoty inverzní distribuční funkce (kvantilové funkce) tabelovat. Toho lze využít při implementaci rychlých algoritmů pro generování požadovaného rozdělení.

Diskrétní rozdělení má schodovitou distribuční funkci, ke které neexistuje funkce inverzní. Pokud však tuto funkci považujeme za relaci, lze vytvořit relaci inverzní. Z ní lze vhodnou úpravou vytvořit kvantilovou funkci vhodnou pro použití v metodě inverzní transformace [5]. Po tabelaci této funkce lze generovat diskrétní rozdělení velmi snadno. Vstupní celočíselná hodnota z rovnoměrného rozdělení tvoří ukazatel do tabulky hodnot kvantilové funkce. Tato tabulka může obsahovat přímo hodnoty generovaného rozdělení (viz obr. 3).



Obr. 3 : Generování diskrétního rozdělení

Po částech rovnoměrné rozdělení lze získat modifikací výše popsané metody. Nejprve je výše uvedeným postupem vygenerována hodnota z diskrétního rozdělení vyjadřujícího relativní četnosti hodnot v jednotlivých třídách po částech rovnoměrného rozdělení. Vygenerovaná hodnota určuje číslo použité třídy. Pro tuto třídu je následně vygenerována hodnota ze spojitého rovnoměrného rozdělení s rozsahem odpovídajícím hranicím třídy. Takto získané hodnoty jsou realizacemi náhodné veličiny s daným po částech rovnoměrným rozdělením (viz obr. 4).



Obr. 4 : Generování po částech rovnoměrného rozdělení

4. Souhrn a závěry

Výše popsaná reprezentace náhodně proměnných veličin **neparametrickými omezenými rozděleními** je použita jako základní metoda popisu vstupních náhodných proměnných v metodě SBRA (Simulation-Based Reliability Assessment) [4]. Naznačený postup generování náhodných proměnných byl implementován v programech pro určování spolehlivosti metodou SBRA AnthillTM, M-StarTM atd. Zvyšování výkonnosti dostupných počítačů umožňuje efektivní aplikaci metody SBRA pro řešení stále se rozšiřujícího spektra problémů. Podrobnější informace včetně seznamu publikací, podávajících přehled o hlavních aplikacích metody SBRA, lze najít například na internetové adrese www.itam.cas.cz/sbra.

Oznámení

Příspěvek byl vypracován s podporou GA ČR (projekt č. 103/01/1410) a agentury Leonardo da Vinci, Brusel, (projekt č. CZ/98/1/82502/PI/I.1.1.a/FPI).

Literatura

- [1] SILVERMAN, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
- [2] DUMMER, M., *Introduction to statistics*, VŠB-TU Ostrava, 1998.
- [3] HURT, J., *Simulační metody*, SPN Praha, 1982.
- [4] MAREK, P., GUŠTAR, M., ANAGNOS, T., *Simulation-Based Reliability Assessment for Structural Engineers*, CRC Press, Inc., Boca Raton, Florida, 1995.
- [5] POPELA, P., *Generování náhodných čísel pro technické aplikace*, 1997.