Cvičení z `NSTP097`
9.12.2008

Úvod do `R`, pravděpodobnostní rozdělení a limitní věty

Úvodní nastavení. Na disku `H` si založte adresář na toto cvičení, např. `NSTP097`. Ze stránky `www.karlin.mff.cuni.cz/~dosla/education/` si stáhněte soubory `uvod.R` a `cviceni1.R`.

Otevřete si program `R` a pomocí File/Open script... si načtěte do `R` soubor `uvod.R`. Odtud budeme postupně společně spouštět jednotlivé příkazy pomocí `CTRL+R`. Kdo už s `R` **umí pracovat**, tak může řešit jednotlivé úkoly **samostatně** a nemusí tyto soubory používat.

Nejprve se s pomocí kódu v `uvod.R` ve stručnosti seznámíme s tím, jak provádět některé základní operace v `R` (přiřazení do proměnné, práce s vektory apod.). Řešení následujících úkolů najdete v souboru `cviceni1.R`.

Úloha 1: Hustoty a distribuční funkce. R má zabudované funkce pro výpočet hustot, distribučních funkcí a kvantilových funkcí pro řadu běžných rozdělení. Taktéž umí generovat náhodné výběry z těchto rozdělení.

Například pro exponenciální rozdělení Exp(λ):
`dexp(x,lambda)` počítá hustotu v bodě `x` nebo ve vektoru bodů `x`,
`pexp(x,lambda)` počítá distribuční funkci v bodě `x` nebo ve vektoru bodů `x`,
`rexp(n,lambda)` generuje náhodný výběr o rozsahu n,
`qexp(p,lambda)` počítá hodnotu kvantilové funkce v bodě `p`

(a) Nakreslete hustotu rozdělení Exp(4).
(b) Nakreslete graf distribuční funkce Exp(4).
(c) Vygenerujte náhodný výběr o délce 25 pozorování z Exp(4).
  - Nakreslete si histogram tohoto výběru a porovnejte s hustotou z bodu a).
  - Spočtěte průměr z tohoto výběru a porovnejte jej se střední hodnotou rozdělení Exp(4). Čemu je rovna střední hodnota rozdělení Exp(λ)?

(d) Zopakujte tento postup pro další rozdělení: N(0,1), N(-2,4), Γ(2,0.5), Γ(0.5,0.1), C(0,1). Hustoty, distribuční funkce, a náhodné výběry dostaneme takto:

N(μ,σ^2)	Hustota:	`dnorm(x,mi,sigma)` (Pozor! `sigma` je $\sqrt{\sigma^2}$!)
	D.f:	`pnorm(x,mi,sigma)`
	Náh. výběr:	`rnorm(n,mi,sigma)`
$\Gamma(a,p)$	Hustota:	`dgamma(x,rate=a,shape=p)`
	D.f:	`pgamma(x,rate=a,shape=p)`
	Náh. výběr:	`rgamma(n,rate=a,shape=p)`
C(a,b)	Hustota:	`dcauchy(x,a,b)`
	D.f:	`pcauchy(x,a,b)`
	Náh. výběr:	`rcauchy(n,a,b)`

(Poznámka: podobně jako u exponenciálního rozdělení ještě existují funkce `qnorm`, `qgamma`, atd, které počítají kvantily.)

(e) Cauchyho rozdělení je známé pro svoje „těžké chvosty“. Poznali byste od oka výběr z Cauchyho rozdělení od výběru z normálního rozdělení? Jak?
(f) Porovnejte do jednoho obrázku hustoty normálního $\mathsf{N}(0,1)$ a Cauchyho $\mathsf{C}(0,1)$ rozdělení.
(g) Nakreslete si stejným způsobem obrázek porovnávající hustotu exponenciálního rozdělení se střední hodnotou 2, gama rozdělení se střední hodnotou 2 a parametrem $p = 0.5$ a gama rozdělení se střední hodnotou 2 a parametrem $p = 2$.
(Poznámka: Zvolte vhodně rozsah bodů `x`, v nichž se hustoty počítají a kreslí.)

Úloha 2: Průměry ze vzrůstajícího počtu pozorování. Nyní budeme sledovat, jak se mění vlastnosti průměru v závislosti na počtu pozorování. Vzpomeňte si, co víte z přednášky o asymptotických vlastnostech $\bar{X}_n$.

Uvažujme gama rozdělení Γ(0.5,2).

(a) Nejdříve pro představu nakreslete hustotu tohoto rozdělení.
(b) Nagenerujte 50 výběrů z tohoto rozdělení, každý o velikosti $n = 25$ pozorování.
   - Spočtěte průměry ve všech těchto výběrech a vypište je. Jaká je střední hodnota Γ(0.5,2)?
   - Spočtěte odhad rozptylu průměrů. Jaký je skutečný rozptyl průměrů z 25 pozorování s rozdělením Γ(0.5,2)? (Víte z teorie.)
   - Nakreslete histogram průměrů.
     Jaké přesné rozdělení má $n \cdot \bar{X}_n$?
(c) Zopakujte bod b) pro rozsahy výběrů $n = 250$ a $n = 2500$, případně i vyšší hodnoty, z téhož rozdělení. (Poznámka: Volte vždy stejný rozsah na ose x.)
Jak se chovají průměry při rostoucím počtu pozorování? Která věta teoreticky zdůvodňuje výsledky, které vidíte?
(d) Vzpomeňte si na předpoklady této věty. Vyberte si rozdělení, které je nesplňuje, a ukažte, co se děje v takovém případě.

Úloha 3: Průměry ze vzrůstajícího počtu pozorování II. O chování průměru $\bar{X}_n$ při $n \to \infty$ toho víme z teorie ještě více. Uvažujme opět např. gama rozdělení Γ(0.5,2).

(a) Podobně jako v předchozí úloze nagenerujte 50 výběrů z tohoto rozdělení, o velikosti $n = 25$ a spočtěte průměry v těchto výběrech. Nakreslete si histogram veličiny Z_n, kde

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma},$$

kde μ a σ^2 jsou střední hodnota a rozptyl rozdělení Γ(0.5,2).
(b) Zopakujte totéž pro $n = 250$, $n = 2500$. Co soudíte o asymptotickém rozdělení veličiny Z_n při $n \to \infty$? Která věta tyto závěry teoreticky zdůvodňuje?
(c) Zamyslete se, jaké jsou předpoklady této věty, a najděte rozdělení, které je nesplňuje.

Další (nepovinné) úkoly.

(a) Nakreslete si do jednoho obrázku hustoty exponenciálního rozdělení s parametry $\lambda = 4$, 2, 1 a 1/2. Interpretujte, co se „děje“ s hustotou, jestliže hodnota λ klesá, resp. roste.
(b) V úkolu 2 a 3 jsme ukazovali platnost dvou důležitých vět na gama rozdělení. Proveďte totéž pro exponenciální rozdělení s parametrem $\lambda = 5$.
(c) Spočtěte (teoreticky) medián exponenciálního rozdělení Exp(λ). Pro hodnotu $\lambda = 4$ spočtěte tuto hodnotu také pomocí `R` a porovnejte.
(d) Pro výběr z Exp(λ) z úkolu 1 porovnejte teoretickou a empirickou distribuční funkci v jednom obrázku. Proveďte pro zvyšující se rozsahy výběru. Co vidíte?