Experimental Design and Statistics - AGA47a

M. Maciak (Czech University of Life Sciences, Prague)

Lab Session 5 - Summer Term 2015

1 Prehled nekterych prikazu

- generatory pro diskretne nahodne veliciny: `rbinom()`, `rpois()` a `rgeom()`;
- generatory pro spojite nahodne veliciny: `runif`, `rexp` a `rnorm()`;
- krome toho existuje mnoho dalsich generatoru pro diskretni i spojite nahodne veliciny (pro jejich vyhledani pouzite *google* a pak help v `R`ku;
- funkce pro vykresleni empirickych pravdepodobnostnich a distribucnich funkci: `hist()`, `density`, `ecdf()`, `stepfun()`;
- nektere graficke nastroje (uzitecne funkce): `plot()`, `boxplot()`, `barplot()`, `pie()`, atd.;
- graficke funkce, ktere vyzaduji jiz otevreny graficky interface: `abline()`, `lines()`, `points()`, `text()`, `legend()`, a dalsi;

2 Gaussove rozdelenie v Rku

- nahodny generator pro generovani hodnot z normalneho rozdeleni: `rnorm()`;
- krome toho existuji funkce `pnorm()` - pro kumulativni distribucni funkci, `dnorm()` - pro pravdepodobnostni funkci (hustotu) a `qnorm()` pro kvantilovou funkci;
- pouzite help v `R`ku a zjistite, jak pouzivat tyhle funkce namisto tabulovanych kritickych hodnot;
  (napr. tu Standard Normal Table)
- vyhoda oproti tabulovanym hodnotam je, ze muzeme primo ziskat kriticke hodnoty aj pro Gaussova rozdeleni, ktore nejsou standardizovana - do funkci `pnorm()`, `dnorm()` a `qnorm()` staci specifikovat dodatecne parametre `mean = ...` a `sd = ...`;
- najdete prislusne kvantily a specifikujte 90 % a 95 % celkove plochy pod standardnou normalni Gaussovou krivkou (specifikujte plochu, ktera je symetricka, napravo a nalevo);
- **Dulezite:** v teorii pouzivame znaceni $X \sim N(\mu, \sigma^2)$, kde μ znaci stredni hodnotu nahodnej veliciny X a $\sigma^2 > 0$ je rozptyl, avsak v softwaru `R` se vyzaduje specifikovat stredni hodnotu μ a smerodatnu chybu namisto rozptylu (odmocnina z rozptylu):

```
> rnorm(n = 10, mean = 0, sd = 2) ### generuje 10 hodnot z rozdeleni N(0, 4)
>                                 ### sd na druhou je 4 = rozptyl
```

- vykreslite teoretickou krivku standardneho normalniho rozdeleni $N(0, 1)$:

```
> xseq <- seq(-10,10, length=5000)
> yvalues <- (1/sqrt(2 * pi)) * exp(- xseq^2 / 2)
> plot(yvalues ~ xseq, type = "l", col = "red")
```

- porovnajte teoretickou krivku s krivkou odhadnutou na zaklade nahodneho vyberu:

```
> sampleSize <- 100
> sample <- rnorm(sampleSize, 0, 1)
> lines(density(sample), col = "blue", lty = 2)
```

- pomoci predchoziho kodu skuste ruzne hodnoty pro velikost nahodneho vyberu (parametr `sampleSize`) a otestujte, jak se empiricka hustoto priblizuje teoreticke, kdyz `sampleSize` $\longrightarrow \infty$;
- vyzkousejte analogycky postup pro nejake obecne normalne rozdeleni s nenulovou stredni hodnotou a rozptylem $\sigma^2 \neq 1$;

3 Centralni Limitni Veta (CLT) - Priklad

- pokuste se nasimulovat nasledujici experiment (jednoducha symetricka mince, pravdepodobnost jedne strany je 50 %);
- nahodny experiment opakujte dostatecne mnoho krat (napr. 100 krat);
  Jak vypada histogram pro nejaky nahodny vyber $X1, \ldots, X_n$?

```
> coinFlip <- rbinom(100,1,0.5)
> hist(coinFlip, col="lightblue")
```

- pouzite Centralni Limitni Vetu (CLT) a overte, ze pro narustajici velikost nahodneho vyberu bude stredni hodnota skutecne konvergovat k standardnimu normalnimu rozdeleni $N(0,1)$:

```
> N <- 100 ### sample size
> clt <- NULL
>
> for (i in 1:5000){
>   coinFlip <- rbinom(N,1,0.5)
>   mean <- sum(coinFlip) / 100
>   clt <- c(clt, (mean - 0.5)/0.5)
> }
>
> hist(clt, col="lightblue")
> plot(density(clt), col="red")
```

- vyzkousejte jine generatory nahodnych velicin (napr.. Poissonovo rozdeleni, nebo Exponencialni rozdeleni) a overte, ze vhodne standardizovany prumer ma skutecne (v limite, pro velikost vyberu dostatecne velkou) standardne normalne rozdeleni $N(0,1)$;
- Jaky parameter odhadujeme pomoci vyberoveho prumeru, kdyz predpokladame, ze $X \sim Poiss(\lambda)$ (pro nejake $\lambda > 0$)?
- Jaky parameter odhadujeme pomoci vyberoveho prumeru, kdyz predpokladame, ze $X \sim Exp(\lambda)$ (pro nejake $\lambda > 0$)?