

Тестирање хипотеза

5. час

30. март 2016.

Монте Карло тест

Монте Карло методе су методе код којих се употребљавају низови случајних бројева за извршење симулације. Првобитно су познати као "статистичка упрошћавања", али назив "Монте Карло", популаризован од стране првих истраживача у овој области је проистекао из назива чувеног казина у Монаку.

Како су за добијање довољно тачне оцене тражене величине, потребна израчунавања за веома велики број посебних случајева и одговарајућа статистичка обрада огромног нумеричког материјала, то је ефективна примена методе Монте Карло омогућена тек појавом електронских рачунара.

Монте Карло тест

Монте Карло тест се користи када је позната расподела обележја X , а није позната расподела тест статистике (ТС).

Предлаже се да тест статистика буде процењена за симулиране случајне узорке да би се одредила p -вредност теста.

Посматрана вредност тест статистике (ТС) се упоређује са вредношћу тест статистике израчунате у сваком од m симулираних узорака исте величине као дат узорак.

Монте Карло тест

Оцењујемо p - вредност Монте Карло теста као

$$\hat{p} = \frac{r}{m + 1},$$

где је

- r број узорака чија је реализована вредност тест статистике већа или једнака од ТС
- m број симулација.

За ниво значајности α , кад год је $\hat{p} \leq \alpha$ нулта хипотеза ће бити одбијена.

Монте Карло тест - алгоритам

- 1 Рачунамо тест статистику (ТС) на основу полазног узорка.
- 2 Симулирамо m узорака који имају исту расподелу као почетни узорак.
- 3 За сваки симулирани узорак рачунамо тест статистике $T_1, \dots, T_m$.
- 4 Оцењујемо p вредност по формули $\hat{p} = \frac{r}{m+1}$.
- 5 Упоређујемо оцењену $\hat{p}$ вредност са прагом значајности α и доносимо закључак.

Тест Колмогорова

Тестирамо хипотезу $H_0(F_X(x) = F_0(x))$, где је F_0 непрекидна и сви параметри су познати.

Тест статистика је

$$D_n = \sup_x |F_n(x) - F_0(x)|,$$

где је F_n емпиријска функција расподеле добијена на основу почетног узорка.

Статистика $\sqrt{n}D_n$ има расподелу Колмогорова за коју постоје специјалне таблице које се добијају по формули

$$P\{\sqrt{n}D_n \leq x\} = 1 - 2 \sum_{k=0}^{\infty} (-1)^{k-1} e^{-2k^2 x^2}.$$

Тест Колмогорова - алгоритам

- 1 Сортирати податке $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.
- 2 Реализовану вредност тест статистике рачунамо као
$$d_n = \max_x |F_n(x) - F_0(x)|$$
- 3 Упоредјујемо вредност $\sqrt{n}d_n$ са вредношћу из таблице за Колмогорову расподелу и односимо закључак.
На пример, за ниво поверења $\alpha = 0.95$ из таблице добијамо да је $D_n = 1.36$.

Тест Колмогоров-Смирнова

Користи се када имамо два узорка и хоћемо да тестирамо да ли су ти узорци из исте расподеле.

Нулта хипотеза је $H_0(F_X(x) = F_Y(x))$, где не знамо која је то расподела ако је H_0 тачно.

Тест статистика је

$$D_{nn'} = \sup |F_n(x) - F_{n'}(x)|,$$

где су F_n и $F_{n'}$ емпиријске функције расподеле за узорке $(x_1, \dots, x_n)$ и $(y_1, \dots, y_{n'})$.

Хипотеза H_0 се одбацује ако је

$$D_{nn'} > c(\alpha) \sqrt{\frac{n + n'}{nn'}},$$

где се $c(\alpha)$ добија из таблица за Колмогорову расподелу.

Тест Колмогоров-Смирнова - алгоритам

- 1 Сортирати податке $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.
- 2 Сортирати податке $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$.
- 3 Рачунамо емпитијске функције расподеле за узорке $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ и $y_{(1)}, y_{(2)}, \dots, y_{(n)}$.
- 4 Реализовану вредност тест статистике рачунамо као $d_{nn'} = \max_x |F_n(x) - F_{n'}(x)|$
- 5 Упоредимо вредност $d_{nn'} \sqrt{\frac{nn'}{n+n'}}$ са вредношћу из таблице за Колмогорову расподелу и доносимо закључак.
На пример, за ниво поверења $\alpha = 0.95$ из таблице добијамо да је $D_n = 1.36$.

Задаци

1. Нека су $x \leftarrow \text{rnorm}(1000)$, $y \leftarrow \text{rnorm}(1000)$ и $Z = X^2 + Y^2$. Претпоставити коју расподелу има случајна величина Z и тестирати ту претпоставку тестом Колмогорова и Монте Карло тестом.
2. Нека су $x \leftarrow \text{round}((-1/3) * \log(\text{runif}(1000)), 2)$ и $y \leftarrow \text{round}(\text{rexp}(1000, 3), 2)$. Тестирати хипотезу да су X и Y из исте расподеле тестом Колмогоров-Смирнова.
3. Описати неки поступак тестирања хипотеза и написати функцију у R -у која симулира тај поступак. Све кораке детаљно објаснити.