



Универзитет у Београду
Математички факултет

БЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ)

ЈУН 2 - 2. октобар 2025. године

1. Нека је низ случајних величина (X_n) такав да за сваки природан број n важи да је $EX_n = \mu$ и $DX_n = \sigma^2$. Ако је

$$\text{cov}(X_i, X_{i+h}) = \begin{cases} 1 - \frac{h}{m+1}, & h = 1, 2, \dots, m, \\ 0, & h > m, \end{cases}$$

за $i \in \mathbb{N}$, где је m познат природан број, испитати да ли за низ случајних величина (X_n) важи слаби закон великих бројева.

2. Обележје X има негативну биномну $\mathcal{NB}(r, p)$ расподелу са законом расподеле

$$P\{X = x\} = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x \in \{r, r+1, \dots\}.$$

На основу узорка обима n , методом максималне веродостојности одредити оцену непознатог параметра $\frac{1}{p}$, а затим испитати непристрасност и постојаност добијене оцене.

3. У платном систему тестира се нови софтвер за обраду трансакција. На случајан начин изабрано је 12 трансакција и мерено је време њихове обраде у секундама. Добијено је да је просечно време обраде трансакције једнако 3.49 секунди, а да је узорачка дисперзија 0.16. Познато је да је на ранијем тестирању старог софтвера за случајно изабраних 13 трансакција добијено да је просечно време обраде трансакције 3.82 секунде, а да је узорачка дисперзија 0.19. Претпостављајући да времена обраде трансакција и старим и новим софтвером имају нормалне расподеле и да су им дисперзије једнаке, са прагом значајности 0.05 испитати да ли је просечно време обраде трансакције на старом софтверу веће од просечног времена обраде трансакције на новом софтверу.

Решења задатака

1. Нека је $S_n = \sum_{i=1}^n X_k$. Пошто је $ES_n = n\mu$, желимо да испитамо да ли $\frac{S_n}{n} \xrightarrow{P} \mu$, $n \rightarrow \infty$. Нека је $\varepsilon > 0$ произвољно. Тада на основу Чебишевљево неједнакости следи да је

$$P\left\{\left|\frac{S_n - ES_n}{n}\right| \geq \varepsilon\right\} = P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = P\{|S_n - n\mu| \geq n\varepsilon\} \leq \frac{DS_n}{n^2\varepsilon^2}.$$

Одредимо сада дисперзију од S_n . Важи да је

$$\begin{aligned} DS_n &= E(S_n - ES_n)^2 = E\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2 = E\left(\sum_{i=1}^n (X_i - EX_i)\right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n E(X_i - EX_i)(X_j - EX_j) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) = \sum_{i=1}^n DX_k + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \\ &= \sum_{i=1}^n DX_k + 2 \sum_{h=1}^m \sum_{i=1}^{n-h} \text{cov}(X_i, X_j) = n\sigma^2 + 2 \sum_{h=1}^m (n-h) \left(1 - \frac{h}{m+1}\right) \\ &= n\sigma^2 + 2 \left[n \sum_{h=1}^m \left(1 - \frac{h}{m+1}\right) - \sum_{h=1}^m h \left(1 - \frac{h}{m+1}\right) \right] \\ &= n\sigma^2 + 2a \left[n \left(m - \frac{1}{m+1} \sum_{h=1}^m h\right) - \left(\sum_{h=1}^m h - \frac{1}{m+1} \sum_{h=1}^m h^2\right) \right] \\ &= n\sigma^2 + 2 \left[n \left(m - \frac{1}{m+1} \cdot \frac{m(m+1)}{2}\right) - \left(\frac{m(m+1)}{2} - \frac{1}{m+1} \cdot \frac{m(m+1)(2m+1)}{6}\right) \right] \\ &= n\sigma^2 + 2 \left[n \frac{m}{2} - \left(\frac{m(m+1)}{2} - \frac{m(2m+1)}{6}\right) \right] = n\sigma^2 + 2 \left[\frac{nm}{2} - \frac{m(m+2)}{6} \right] \\ &= n(\sigma^2 + m) - \frac{1}{3}m(m+2). \end{aligned}$$

Одавде следи да је

$$P\left\{\left|\frac{S_n - ES_n}{n}\right| \geq \varepsilon\right\} \leq \frac{\sigma^2 + m}{n\varepsilon^2} - \frac{m(m+2)}{n^2\varepsilon^2} \rightarrow 0, n \rightarrow \infty.$$

Према томе, $\frac{S_n}{n} \xrightarrow{P} \mu$, $n \rightarrow \infty$, па важи слаби закон великих бројева.

2. Претпоставимо да имамо реализован узорак $(x_1, x_2, \dots, x_n)$. Функција веродостојности је

$$L(p) = L(p|x_1, x_2, \dots, x_n) = \prod_{i=1}^n \binom{x_i-1}{r-1} p^r (1-p)^{x_i-r}, \quad p \in [0, 1], \quad x_1, \dots, x_n \in \{r, r+1, \dots\},$$

тј.

$$L(p) = p^{nr} (1-p)^{\sum_{i=1}^n x_i - nr} \prod_{i=1}^n \binom{x_i-1}{r-1}$$

па је

$$\ln L(p) = nr \ln p + \left(\sum_{i=1}^n x_i - nr\right) \ln(1-p) + \ln \prod_{i=1}^n \binom{x_i-1}{r-1}.$$

Желимо да пронађемо p које максимизује $\ln L(p)$, те стога одређујемо стационарну тачку:

$$\frac{\partial}{\partial p} \ln L(p) = \frac{nr}{p} - \frac{\sum_{i=1}^n x_i - nr}{1-p} = 0 \implies p = \frac{nr}{\sum_{i=1}^n x_i} = \frac{r}{\bar{x}_n}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је $\frac{\partial}{\partial p} \ln L(p)$ позитивна за $p < \frac{r}{\bar{x}_n}$, а негативна за $p > \frac{r}{\bar{x}_n}$, па је функција $\frac{\partial}{\partial p} \ln L(p)$ растућа за $p < \frac{r}{\bar{x}_n}$, а опадајућа за $p > \frac{r}{\bar{x}_n}$. Према томе, $\frac{r}{\bar{x}_n}$ је тачка у којој $\ln L(p)$ достиже максимум. Према томе, на основу овог узорка, оцена методом максималне веродостојности је

$$\hat{p} = \frac{r}{\bar{X}_n}.$$

Како је оцена методом максималне веродостојности инваријантна у односу на трансформацију Борел-мерљивом функцијом, одавде закључујемо да је оцена непознатог параметра $\frac{1}{p}$ дата са

$$T_p = \frac{\sum_{i=1}^n X_i}{nr} = \frac{\bar{X}_n}{r}.$$

Испитајмо сада да ли је дата оцена непристрасна. Како је $X \in \mathcal{NB}(r, p)$, то важи да је $EX = \frac{r}{p}$ и $DX = \frac{r(1-p)}{p^2}$. Сада имамо да је

$$E(T_p) = E\left(\frac{\bar{X}_n}{r}\right) = \frac{1}{nr} \cdot \frac{nr}{p} = \frac{1}{p},$$

тј. оцена T_p јесте непристрасна оцена параметра $\frac{1}{p}$. Даље, имамо да је

$$D(T_p) = D\left(\frac{\bar{X}_n}{r}\right) = \frac{1}{r} \frac{1}{n} \frac{r(1-p)}{p^2} = \frac{1-p}{nrp^2}.$$

Остаје још да испитамо постојаност ове оцене. Постојаност подразумева да оцена конвергира у вероватноћи ка параметру $\frac{1}{p}$. Нека је $\varepsilon > 0$ произвољно:

$$P\left\{\left|T_p - \frac{1}{p}\right| > \varepsilon\right\} \leq \frac{E\left(T_p - \frac{1}{p}\right)^2}{\varepsilon^2} = \frac{D(T_p)}{\varepsilon^2} = \frac{1-p}{nrp^2\varepsilon^2},$$

при чему је коришћена Чебишовљева неједнакост. Дакле,

$$0 \leq P\left\{\left|T_p - \frac{1}{p}\right| > \varepsilon\right\} \leq \frac{1-p}{nrp^2\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty,$$

па следи да $P\left\{\left|T_p - \frac{1}{p}\right| > \varepsilon\right\} \rightarrow 0$, када $n \rightarrow \infty$ и то важи за свако $\varepsilon > 0$. Закључујемо да је T_p постојана оцена за параметар $\frac{1}{p}$.

- Нека је X обележје које представља време обраде трансакција старим софтвером, а Y обележје које представља време обраде трансакција новим софтвером. Претпоставља се да обележје X и Y имају нормалну расподелу са једнаким дисперзијама. Дакле, претпостављамо да

$$X \in \mathcal{N}(\mu_1, \sigma^2), \quad Y \in \mathcal{N}(\mu_2, \sigma^2).$$

На основу добијених узорака обима $n_1 = 13, n_2 = 12$, желимо да тестирамо хипотезу $H_0(\mu_1 = \mu_2)$ против алтернативе $H_1(\mu_1 > \mu_2)$.

На основу услова задатка имамо да је $\bar{x}_{13} = 3.82, \bar{y}_{12} = 3.49, \bar{s}_{13}^2(X) = 0.19, \bar{s}_{12}^2(Y) = 0.16$. Даље, знамо да

$$T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{n_1 \bar{S}_{n_1}(X) + n_2 \bar{S}_{n_2}(Y)}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)} \in t_{n_1 + n_2 - 2}.$$

Користимо тест статистику

$$T = \frac{\bar{X}_{13} - \bar{Y}_{12}}{\sqrt{13 \bar{S}_{13}(X) + 12 \bar{S}_{12}(Y)}} \sqrt{143.52},$$

која, при H_0 , има Студентову t_{23} расподелу.

Пошто је расподела статистике T симетрична око нуле, формирамо критичну област на следећи начин

$$W = \{T \geq c\},$$

где критичну вредност c налазимо из таблице за Студентову расподелу:

$$0.05 = P_{H_0}\{T \geq c\} = \frac{1}{2}P_{H_0}\{|T| \geq c\},$$

односно

$$0.1 = P_{H_0}\{|T| \geq c\},$$

па је $c = 1.714$. Реализована вредност тест статистике једнака је

$$\frac{3.82 - 3.49}{\sqrt{13 \cdot 0.19 + 12 \cdot 0.16}} \sqrt{143.52} = 1.887.$$

Како реализована вредност упада у критичну област, то одбацујемо H_0 и прихватамо алтернативу да нови софтвер ради брже.