

Вијећу Природно-математичког факултета

Оцјена подобности теме за израду магистарског рада под називом *Примјена бутстрап метода за рачунање p вриједности теста Колмогоров Смирнова* кандидата Јова Златичанина

На LIX сједници Вијећа ПМФ-а одржаној 9.3.2021. године одлуком број 480/2 именовани смо у Комисију за оцјену подобности теме магистарског рада *Примјена бутстрап метода за рачунање p вриједности теста Колмогоров Смирнова* кандидата Јова Златичанина. Вијећу подносимо следећи извјештај.

1. Подаци о кандидату

Јово Златичанин рођен је 1996. године у Подгорици. Гимназију је похађао у Подгорици и носилац је дипломе Луча. Основне студије на Студијском програму за математику и рачунарске науке, Природно математичког факултета у Подгорици (Универзитет Црне Горе), уписао је 2015. и завршио 2018. године са просјечном оцјеном 8,63. Специјалистичке студије је уписао 2018. године, а завршио 2019. са просјечном оцјеном 9,25; специјалистички рад са насловом "Протокол Дифи-Хелмана за размјену кључева" одбранио је 7.7. 2019. године. Магистарске академске студије на студијском програму Математика и рачунарске науке је уписао у октобру 2019. године. Говори енглески и руски језик.

2. Образложење теме

2.1 Научна област

Тема припада области Математичке статистике (MSC 2010 шифра 62), ужа област Тестирање статистичких хипотеза (MSC 2010 шифра 62Г10).

2.2 Предметрада и научне методе

Поступци којима се примјеном статистичких метода утврђује да ли се, на основу узорка може, и са којом вјероватноћом прихватити претпоставка о конкретној бројчаној вриједности неког параметра су тестирања статистичких хипотеза. Ови тестови могу бити параметарски и непараметарски. Свака претпоставка (хипотеза) о непознатом параметру расподеле назива се параметарска хипотеза, а поступак њеног потврђивања или одбијања на основу узорка је параметарски тест. Статистика које се користи у том поступку је тест статистика. Хипотезе о расподјели обилежја које се не односе на параметре, већ на саму расподелу обележја су непараметарски тестови или тестови сагласности (обилежја са расподелом).

Тест Колмогоров-Смирнова се користи за испитивње сагласности узорка са претпостављеном расподелом и за испитивње једнакости расподеле два обилежја. У једном случају, располаже се једним узорком, док се у другом случају на располагању два независна узорка за свако обилежје посебно. Оба проблема тестирања статистичких хипотеза се посматрају у случају када су расподеле обилежја апсолутно-непрекидне. Претпоставља се да у случају једног простослучајног узорка $(X_1, \dots, X_n)$ обима n из популације чије обилежје X има функцију расподеле F апсолутно-непрекидног типа. Код теста Колмогоров-Смирнова се посматрају одступања између емпиријске функције расподеле F_n и функције расподеле F . Тако, тестирамо нулту хипотезу $H_0(F = F_0)$, гдје је F_0 претпостављена функција расподеле, против алтернативне хипотезе $H_1(F \neq F_0)$.

Нека је α праг значајности и нека је F_n емпиријска функција расподеле која одговара узорку $(X_1, \dots, X_n)$. За тестирање ових хипотеза користимо тест статистику $D_n(F)$ облика

$$D_n(F) = \sup_x |F_n(x) - F(x)|.$$

Тест статистика $D_n(F)$ се зове статистика Колмогоров-Смирнова и има веома интересантне особине. Тако, на основу теореме Гливленко-Кантели познато је да $D_n(F)$ скоро извјесно тежи ка 0 када $n \rightarrow \infty$. Такође, $D_n(F)$ не зависи од функције расподеле F . Дакле, важи

Теорема: Ако је $(X_1, \dots, X_n)$ прост случајан узорак обима n из популације чије обиљежје X има функцију расподеле F апсолутно-непрекидног типа, тада расподела статистике Колмогоров-Смирнова $D_n(F)$ не зависи од функције расподеле F .

Основни проблем који ће се у раду третирати јесте случај када се параметри расподеле оцјењују на основу датог узорка. У том случају хипотеза H_0 не важи, па се одговарајућа p вриједност рачуна примјеном симулација. У раду ће се примјенити бутстрап метода (енгл. bootstrap method). Бутстраповање је метод који се заснива на поновљеном узорковању са понављањем. Бутстрап је једноставан, али веома моћан статистички метод који је Ефрон први пут представио на Стенфорд Универзитету 1977. године. Међутим, и након објављивања рада у часопису *The Annals of Statistics*, 1979. године, бутстрап метод није коришћен јер у то вријеме рачунари нису били довољно развијени тако да није било могуће брзо извођење захтијевних и великих рачуна.

Претпоставиће се да се врши тестирање помоћу Колмогоров Смирнов теста и примјеном бутстрап метода рачунаће се одговарајућа p вриједност, на основу које ће се донијети одговарајући закључак. У поступку рачунања генерисаће се око 10 000 поновљених узорака који ће бити касније коришћени за израчунавање p вриједности. Сва израчунавања биће праћена одговарајућим кодом.

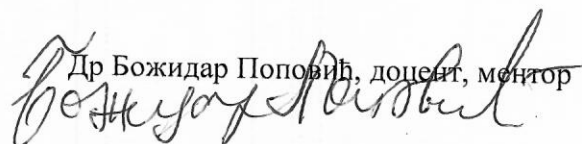
3. Литература

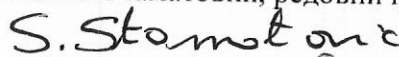
1. B. Efron and R. Tibshirani. An introduction to bootstrap. Chapman and Hall, 1993.
2. A. C. Davison and D. V. Hinkley. Bootstrap methods and their application. Cambridge Series on Statistical and Probabilistic Mathematics.
3. F. Massey. The Kolmogorov-Smirnov test for Goodness of fit in: Journal of the American Statistical Association (1951).
4. B. Vidgen and T. Yasseri. P-Values: Misunderstood and Misused in: Frontiers in Physics 4.6 (2016).

Закључак

Предложена тема има јасно дефинисане циљеве и прецизно наведене методе истраживања. Предлажемо Вијећу да одобри израду магистарског рада под називом *Примјена бутстрап метода за рачунање p вриједности теста Колмогоров Смирнова* кандидата Јова Златичанина.

У Подгорици, 22. март 2021. године


Др Божидар Поповић, доцент, ментор

Др Синиша Стаматовић, редовни професор, члан


Др Горан Попивоца, доцент, члан
