



Državni univerzitet u Novom Pazaru
Departman za prirodno-matematičke nauke

Atif Avdović

**Funkcije zone i kvantil-zone i primene u
kontroli kvaliteta i testiranju normalnosti**

Doktorska disertacija

Novi Pazar, 2024.



State University of Novi Pazar


Department of Science and Mathematics

Atif Avdović

**Zone and Quantile-Zone Functions and
Applications in Quality Control and
Normality Testing**

Doctoral Dissertation

Novi Pazar, 2024.

	Državni univerzitet u Novom Pazaru Departman za prirodno-matematičke nauke Vuka Karadžića 9, 36300 Novi Pazar	Datum i broj
	KLJUČNA DOKUMENTACIJA	

Redni broj, RBR:			
Identifikacioni broj, IBR:			
Tip dokumentacije, TD:	Monografska dokumentacija		
Tip zapisa, TZ:	Tekstualni štampani materijal		
Vrsta rada, VR:	Doktorska disertacija		
Autor, AU:	Atif Avdović		
Mentori, MN:	Prof. dr Vesna Jevremović		
Naslov rada, NR:	Funkcije zone i kvantil-zone i primene u kontroli kvaliteta i testiranju normalnosti		
Jezik publikacije, JP:	Srpski jezik		
Jezik izvoda, JI:	Srpski jezik		
Zemlja publikovanja, ZP:	Srbija		
Uže geografsko područje, UGP:	Novi Pazar		
Godina, GO:	2024. godina		
Izdavač, IZ:	Državni univerzitet u Novom Pazaru		
Mesto i adresa, MA:	Vuka Karadžića 9, 36300 Novi Pazar		
Fizički opis rada, FO: (poglavlja/strana/citata/tabela/slika/grafika/priloga)	5 poglavlja/194 strane/133 citata/17 tabela/48 slika i grafikona/7 priloga		
Naučna oblast, NO:	Matematika		
Predmetna odrednica/ključne reči, PO:	zona, kvantil-zona, normalna raspodela, kontrolne karte, testovi normalnosti		
UDK:			
Čuva se, ČU:	Biblioteka Državnog univerziteta u Novom Pazaru		
Važna napomena, VN:			
Izvod, IZ:			
Datum prihvatanja teme, DP:	30.08.2024.		
Datum odbrane, DO:			
Članovi komisije, KO:	Mentor:	Prof. dr Vesna Jevremović	
	Predsednik, član:	Prof. dr Diana Dolićanin Đekić	
	Član:	Prof. dr Bojana Milošević	
	Član:	Prof. dr Predrag Popović	
	Član:	Doc. dr Marko Obradović	

Sadržaj:

Predgovor.....	8
Sažetak	9
Abstract.....	11
1. Uvod.....	13
2. Slučajni i pseudoslučajni brojevi.....	17
2.1. Generatori slučajnih brojeva	17
2.2. Primeri generatora pseudoslučajnih brojeva	19
Metoda sredine kvadrata	19
Prost linearni kongruentni generator.....	19
Drugi kongruentni generatori.....	21
Proširenje linearnog kongruentnog generatora	22
Mersenov tvister.....	28
2.3. Testiranje slučajnosti generisanih brojeva	28
Test tačaka rasta	29
Test tačaka zaokreta	30
Test koraka.....	31
Test serijske korelacije.....	32
DIEHARD testovi	33
2.4. Analiza performansi i poređenja.....	34
3. Modeliranje slučajnih veličina	36
3.1. Metode modeliranja diskretnih slučajnih veličina	36
Metoda inverzne funkcije	38
Metoda prihvatanja i odbacivanja.....	39

Metoda mere uniformnosti.....	40
3.2. Metode modeliranja apsolutno neprekidnih slučajnih veličina	41
Metoda diskretizacije	43
Metoda inverzne funkcije	43
Metoda prihvatanja i odbacivanja.....	45
Metoda mere uniformnosti.....	47
3.3. Neke važne raspodele i njihovo modeliranje	48
Modeliranje indikatora događaja – Bernulijeve raspodele	49
Modeliranje beta raspodele	49
Modeliranje binomne raspodele.....	51
Modeliranje Burove raspodele	52
Modeliranje eksponencijalne raspodele	53
Modeliranje gama raspodele	54
Modeliranje Gumbelove raspodele	55
Modeliranje χ^2 (hi-kvadrat) raspodele.....	56
Modeliranje Košijeve raspodele	57
Modeliranje lognormalne raspodele	58
Modeliranje Laplasove raspodele	59
Modeliranje logističke raspodele	59
Normalna raspodela i njeno modeliranje	60
Modeliranje Paretove raspodele.....	68
Modeliranje Studentove raspodele.....	69
Modeliranje Tjukijeve raspodele	69
Modeliranje Vejbulove raspodele	70
4. Kontrola kvaliteta i kontrolne karte	72
4.1. Empirijska funkcija raspodele.....	72
4.2. Kontrolne karte	74

4.3. Šuhartova X kontrolna karta	77
4.4. Druge kontrolne karte	80
R karta.....	80
S karta	81
CUSUM kontrolna karta – Kontrolna karta kumulativnih suma.....	82
EWMA kontrolna karta.....	84
4.5. Kontrolne karte – novi pristupi	85
EFR karta	86
Kontrolna karta kvantil-zona	90
Q-Q kontrolna karta	97
Kontrolna karta p-vrednosti	102
4.6. Analiza poređenja performansi	105
5. Testovi normalnosti	107
5.1. Moć i p-vrednost testa	108
5.2. Primeri testova saglasnosti – normalnosti	113
Test Kolmogorova	114
Lilifors test.....	114
Pirsonov χ^2 test.....	115
Šapiro-Vilk test.....	116
Anderson-Darling test.....	116
Test zona	117
Test kvantil-zona.....	140
5.3. Analiza moći i poređenja testova normalnosti	159
Analiza moći i poređenja testova zona i kvantil-zona za prostu nultu hipotezu.....	162
Analiza moći i poređenja testova zona i kvantil-zona za složenu nultu hipotezu	167
Zaključak	174

Reference.....	176
Prilozi	186

Predgovor

Izrada ove doktorske disertacije bila je dug i izazovan proces koji ne bi bio moguć bez pomoći i podrške mnogih ljudi i institucija. Ovaj rad je rezultat višegodišnjeg istraživanja, truda, i predanosti, ali takođe i podrške koju sam imao tokom ovog putovanja.

Prvo, želim da izrazim svoju duboku zahvalnost mojoj mentorki, prof. dr Vesni Jevremović, čija stručnost, nesebična podrška, praćenje i mentorstvo su bili od neprocenjive vrednosti. Njena motivacija, strpljenje i spremnost da podeli svoje znanje učinili su ovaj rad mogućim. Bez njene pomoći, ovaj rad ne bi bio završen.

Zahvaljujem se i članovima komisije, prof. dr Diani Dolićanin Đekić, prof. dr Bojani Milošević, prof. dr Predragu Popoviću i doc. dr Marku Obradoviću, na njihovim dragocanim savetima, konstruktivnim kritikama i sugestijama koje su značajno unapredile kvalitet ove disertacije.

Posebnu zahvalnost dugujem svojim roditeljima, bratu, sestri, supruzi i sinu koji su bili moja najčvršća podrška tokom čitavog akademskog putovanja. Njihovo strpljenje, ljubav i vera u mene dali su mi snagu da istrajem i da se sa svakim izazovom suočim sa zadovoljstvom.

Veliko hvala i mojim profesorima, prijateljima, kolegama sa univerziteta i vrednim studentima, koji su uvek bili tu da me podrže, ohrabre i podsete me da moj rad ima svrhu. Vaša podrška, kako akademska tako i emotivna, bila je neprocenjiva. Isto važi za sve moje prijatelje van univerziteta.

Zahvaljujem se i Državnom univerzitetu u Novom Pazaru na doprinosu mom istraživačkom radu. Takođe, želim da izrazim zahvalnost svim državnim univerzitetima Srbije koji su svojom saradnjom sa DUNP-om omogućili da od njihovih profesora učim, a i da sa njima saradujem.

Na kraju, želim da zahvalim svima koji su na bilo koji način doprineli ovom istraživanju, bilo kroz formalnu ili neformalnu podršku. Vaša pomoć je bila dragocena i od velikog značaja za uspeh ovog rada.

Ova disertacija je posvećena svima vama.

Atif Avdović

Sažetak

Ovaj rad (disertacija) je rezultat dosadašnjeg istraživanja autora na teme generatora pseudoslučajnih brojeva, modeliranja slučajnih veličina (posebno normalne raspodele), kontrole kvaliteta i testiranja normalnosti. Tekst je pokušaj da se kroz detaljan pregled literature, prikaz postojeće teorije i njene dopune i empirijske verifikacije rezultata da doprinos spomenutim oblastima. Iako rezultati iz oblasti generatora pseudoslučajnih brojeva i modeliranja slučajnih veličina nisu zanemarljivi glavni rezultat ovog istraživanja su predlozi novih kontrolnih karata i testova saglasnosti. Funkcije zone i kvantil-zone predstavljaju osnovu istraživanja, a njihova primena je u definisanju značajnih elemenata kontrolnih karata i testova normalnosti i sredstvo je za njihovo adekvatno tumačenje.

Rad se sastoji od šest poglavlja.

Prvo poglavlje daje uvod u materiju kroz detaljan pregled literature iz svake oblasti obrađene u radu.

Drugo poglavlje se bavi pseudoslučajnim brojevima, generatorima pseudoslučajnih brojeva i testiranjem slučajnosti nizova generisanih brojeva. Ovde je predloženo proširenje linearnog kongruentnog generatora, analizirane njegove osobine i izvršena simulaciona analiza poređenja predloženog generatora sa najčešće korišćenim generatorom. Rezultati pokazuju da predloženi generator zadržava sve osobine slučajnosti linearnog kongruentnog generatora, ali period biva višestruko povećan, što ga čini konkurentnim sa većim brojem generatora koji se sreću u praksi.

U nastavku, tj. u trećem poglavlju, obrađene su metode modeliranja slučajnih veličina i izložene posebno za diskretne i posebno za neprekidne slučajne veličine. Treće poglavlje sadrži i kratak pregled definicija i osobina za svaku raspodelu od važnosti za istraživanje, sa ilustracijama metoda modeliranja algoritmima prilagođenim tim raspodelama. Normalna raspodela je kao najznačajnija za ovo istraživanje najviše obrađena, prikazano je najviše algoritama njenog modeliranja, a uveden je i jedan novi dobijen kao rezultat istraživanja, čiji je postupak zasnovan na funkciji zone. Takođe, u ovom poglavlju su formalizovani algoritmi modeliranja diskretne slučajne veličine metodom prihvatanja i odbacivanja, i modeliranja apsolutno neprekidne slučajne veličine metodom diskretizacije.

U četvrtom poglavlju dati su teorijski prikazi o empirijskoj funkciji raspodele i najčešće korišćenim kontrolnim kartama kao što su Šuhartova kontrolna karta, R i S karte, CUSUM karta i EWMA karta. Prikazane su i nove kontrolne karte uvedene ovim istraživanjem, EFR kartom, kartom kvantil-zona, Q-Q kartom i kartom p-vrednosti. Tumačenje kontrolnih karata je obogaćeno korišćenjem funkcija zone i kvantil-zone, a neke od njih su i zasnovane na tim funkcijama. Dodatno, u ovom poglavlju data je analiza poređenja performansi korišćenjem prosečnog broja uzoraka pre detekcije prvog defekta (ARL – average run length) i dobijene kontrolne karte su se pokazale konkurentnim, a u nekim slučajevima i boljim od postojećih.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Konačno, u petom poglavlju, dat je kratak prikaz test-statistika i procedura testiranja postojećim, najčešće korišćenim testovima saglasnosti, prikaz teorijske zasnovanosti za dva testa normalnosti uvedena ovim istraživanjem, tj. testova zona i kvantil-zona čije su test-statistike bazirane na funkcijama zone i kvantil-zone, analiza moći ta dva nova testa za razne alternativne raspodele i analiza poređenja moći sa prikazanim postojećim testovima. Testovi uvedeni u ovom radu su u mnogim slučajevima pokazali bolje performanse od inače korišćenih testova. Kada važi suprotno, uvedeni testovi se mogu smatrati konkurentnim postojećim.

Ceo rad je praćen detaljnim tabeliranjem svih empirijski dobijenih rezultata i grafičkim ilustracijama svih pogodnih elemenata istraživanja.

Na kraju je dat zaključak, prikazane su reference koje su korišćene u istraživanju i MATLAB kodovi svih programa (algoritama) kojim je vršena empirijska verifikacija rezultata.

Abstract

This manuscript is result of long-lasting research in topics of pseudorandom numbers generators, random variables modelling (mostly normal distribution), quality control and normality testing. Through detailed literature review, relevant theory display and empirical verification of results, this manuscript's text attempts to contribute to these topics. Though the results in pseudorandom numbers generators and random variables modelling are not neglectable, this research's main result consists of introducing new control charts and normality tests. Zone and Quantile-zone functions give the idea behind this research, they are applied in defining main control charts and normality tests and are used as a tool in their interpretation.

This thesis consists of five sections.

The first section provides with an introduction to the subject through a detailed literature review from each area that this manuscript deals with.

The second section deals with pseudorandom numbers, pseudorandom number generators, and testing the randomness of generated numbers. Here, an extension of the linear congruential generator is proposed, its properties analysed, and a simulation analysis comparing the modified generator using the extension with the most commonly used generator is performed. The results show that the generator obtained in this way retains all the properties of randomness of the linear congruential generator, but its period is significantly larger, making it competitive with other high-period generators.

In the following, i.e., in the third section, methods for modelling random variables are discussed, separately for discrete and continuous random variables.

The third section contains a brief overview of definitions and properties for each distribution of importance for the research, as well as modelling methods using algorithms adjusted to these distributions. The normal distribution, being the most important for this research, is most extensively discussed, with the presentation of various algorithms for its modelling, including a new one resulting from this research, based on the zone function. Also, in this section, algorithms for discrete random variable modelling via acceptance/rejection method, and continuous random variable modelling via discretization method, have been formally introduced.

The fourth section provides a theoretical overview regarding the empirical distribution function, commonly used control charts such as Shewhart's control chart, R and S charts, CUSUM chart, and EWMA chart, and new control charts introduced in this research, EFR chart, quantile-zone chart, Q-Q chart, and p-value chart, whose interpretation is enhanced using zone and quantile-zone functions, which are the basis for some of these charts. Additionally, this section includes a performance comparison analysis using average run length (ARL),

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

where the obtained control charts proved to be competitive to, and in some cases, even better than the existing ones.

Finally, in the fifth section, a brief overview of test statistics and procedures for commonly used goodness of fit or normality tests is given, as well as the theory behind two normality tests introduced by this research, the zone test and the quantile-zone test, whose test statistics are based on zone and quantile-zone functions, an analysis of the power of these two new tests for various alternative distributions, and a comparison of power with the presented existing tests. Introduced tests showed better performance than the commonly used ones in many of considered cases. When the opposite holds, these test can be considered competitive with the existing ones.

The entire work has detailed tabulation following all empirically obtained results and graphical illustrations of all relevant research elements.

The study is concluded with a additional section, the references used in the research are presented, and appendices containing MATLAB codes of all programs used for empirical verification of the results.

1. Uvod

Linearni kongruentni generator (LKG), poznat i kao prost linearni kongruentni generator ili Lemerov kongruentni generator, koji je predstavio Lemer (1949), pokazao se kao najčešće korišćeni generator slučajnih brojeva (Eichenauer, Grothe, Lehn & Topuzoğlu, 1987; Gentle, 2002, str. 11). Razlog tome su karakteristike LKG-a kao što su karakteristike slučajnosti generisanih brojeva, jednostavnost za programiranje i brzina. Značaj ovih karakteristika je detaljno razmatran (Devroye, 1986). Ono što je već inicijalno problem ovog generatora jesu male vrednosti perioda, pa je taj problem najdetaljnije razmatran. Naročito je razmatran izbor modula (Avdović, 2023).

Knut (1998, str. 12-15) je analizirao LKG i dobio rezultate koji pokazuju da je optimalni modul za multiplikator oblika $2^k + 1$ prost broj¹, iako modul oblika 2^p , ($p > k$) najbrže generiše brojeve. Ovakav izbor modula i multiplikatora su takođe detaljno razmatrani (Park & Muller, 1988). Naime, Park i Miler (1988) tvrde da bi LKG sa modulom $2^{31} - 1$ i multiplikatorom 16807 trebalo uzeti kao minimalni standard u pogledu perioda i slučajnosti. Međutim, Tang (2005) tvrdi da se dobra slučajnost i period LKG-a postižu za vrednosti modula deset najvećih prostih brojeva manjih od 2^{31} . Osim toga, Fišman i Mur (1986) pružili su listu od 414 optimalnih multiplikatora za ovaj modul. Park i Miler (1988) su zaključili da je stvaran broj tih multiplikatora 410, a broj 414 u originalnom radu naveden je usled greške u brojanju.

Modul 2^{32} se koristi zbog jednostavnosti generisanja brojeva. Dodatna prednost ovog modula je ta da sa neparnim priraštajem i multiplikatorom ekvivalentnim $\pm 3 \bmod 8$ LKG dostiže maksimalan period (Marsaglia, 2003). Fišman (1990) je izvršio iscrpnu analizu ovog modula. Zbog potrebe za većim periodom i boljim kvalitetom generisanih brojeva, kao i usled boljih performansi računara, performanse LKG-a su se poboljšale povećanjem modula sa inače korišćenog $2^{31} - 1$ na veće, poput $2^{61} - 1$, i sličnih (Wu, 1997).

Teorijska zasnovanost ovog generatora inicijalno nije bila razvijena. Tom problemu su kasnije istraživači posvetili puno pažnje u mnogim radovima (Fuller, 1976), kao i udžbenicima (Knuth, 1998, str. 16-21). Prema Parku i Mileru (1988), najkompletnija teorija razvijena je za multiplikativne i mešovite linearne kongruentne generatore.

Iako je istraživanje metoda generisanja slučajnih brojeva relativno dugo ono se ne može proglasiti završenim zbog stalne potrebe za modeliranjem velikih skupova podataka usled eksponencijalnog rasta broja oblasti primene. To se može videti u skoro objavljenim udžbenicima (Mulder, Mermoud, Lenders & Tellenbach, 2023), a novi generatori slučajnih brojeva, bilo da su zasnovani na modularnoj redukciji ili ne (Goualard, 2022), još uvek se razvijaju, razmatraju, testiraju i upoređuju sa drugim poznatim generatorima (Jacak, Józwiak,

¹ Prost broj kao modul rezultuje sa generatorom najvećeg perioda među generatorima sa multiplikatorom oblika $2^k + 1$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Niemczuk & Jacak, 2021). U tom smislu, deo istraživanja ovog rada predstavlja važno poboljšanje široko korišćenog i proučavanog generatora slučajnih brojeva koji i dalje ima značaj za primenu, kao i za dalja naučna istraživanja kao što su analiza performansi kontrolnih karata (Jevremović & Avdović, 2020a, 2020b), analiza moći testova saglasnosti (Avdović & Jevremović, 2022, 2023), modeliranje slučajnih veličina i dr.

Postojeća i dobro uspostavljena rešenja u vezi sa diskretnim raspodelama, poput binomne raspodele, geometrijske raspodele i Puasonove raspodele, i dalje se često koriste (Hogg, McKean & Craig, 2019). Novi modeli često se razvijaju kao rešenja teorijskih i empirijskih problema (Aboraya, Yousof, Hamedani & Ibrahim, 2020; Gillariose i dr., 2021; Avdović & Jevremović, 2022; Fabiano, Gardašević-Filipović, Mirkov, Todorčević & Radenović, 2022; Hassan, Shalhaf, Bilal & Rashid, 2020; Vaz & Torres, 2021; Almetwally i dr., 2021). U poslednjih nekoliko godina, puno pažnje posvećeno je opisu i prognoziranju obeležja COVID-19 (Gillariose i dr., 2021; Vaz & Torres, 2021; Almetwally i dr., 2021). Međutim, druge oblasti primene, poput medicine i poljoprivrede (Aboraya, Yousof, Hamedani & Ibrahim, 2020), kontrole kvaliteta i testiranja normalnosti (Jevremović & Avdović, 2022, 2023), genetike i biologije (Vaz & Torres, 2021), indeksa života, tačnosti čitanja i inteligencije (Korkmaz, Chesneau & Korkmaz, 2021), vremena do otkazivanja uređaja i elektronskih komponenti (Aboraya i dr., 2020; Gillariose i dr., 2021; Korkmaz i dr., 2021), modeliranja zavisnosti od starosti (Nafidi, Bahij, Gutiérrez-Sánchez & Achchab, 2020), i drugih, nisu zanemarene.

Budući da normalna raspodela opisuje najveći broj izučavanih obeležja u prirodi, društvu, medicini, inženjeringu itd., tema od značaja su i raspodele povezane s normalnom raspodelom (Avdović & Jevremović 2022; Korkmaz i dr., 2021; Nafidi i dr., 2020). Takođe, pouzdanost parametarskih metoda statističke analize zavisi od normalnosti referentnih varijabli. Normalna raspodela i njene osobine, grafičke i druge metode preliminarne analize normalnosti, kao i testovi normalnosti, su analizirani, razrađeni i često korišćeni (Thode, 2002). S obzirom na njenu važnost, čini se da je istraživanje normalne raspodele, posebno u vezi s testiranjem normalnosti, i dalje otvoreno polje za istraživanje.

Rastući značaj kontrole kvaliteta kao teme istraživanja je primetan kako u teoriji tako i u primeni (Jevremović & Avdović, 2020a, 2020b; Ajadi, Zwetsloot & Tsui, 2021; Hu, Sun, Xie, & Tang, 2022; Owens & Rioborue, 2022; Kotb & El-Ashkar, 2020; Veljković, Elfaghihe & Jevremović, 2015).

Upotreba verovatnoće i statistike u kontroli kvaliteta, tj. statistička kontrola procesa (statistical process control) datira iz 1924. godine kada je Šuhart predstavio svoj kontrolni dijagram zasnovan na kvantilima za normalnu raspodelu, tačnije na pravilu 3σ (Montgomery, 2009). Od tada do danas, kontrolni dijagrami, tj. kontrolne karte su korišćeni, poboljšavani i menjani. Šuhartova \bar{X} karta kontroliše prosečnu vrednost obeležja određenog željenom karakteristikom procesa, pretpostavljajući da to obeležje ima normalnu raspodelu. Uskoro su u kontroli kvaliteta dodati R i S dijagrami (Minić, 2018), zatim CUSUM (Page, 1954) i EWMA (Roberts, 1959) dijagrami kojima je moguće pratiti male promene u proizvodnom procesu. Druge statističke metode su takođe korišćene u kontroli kvaliteta bilo da je ta disciplina bila

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

poznata kao takva ili ne. Neke od tih metoda su kutijasti dijagram (box plot) (Praveen, Delhi Narendran, Pavithran & Thirumalai, 2017), Q-Q dijagram (Hui i dr., 2018), ANOVA (analiza varijanse) (Dura & Isac, 2006), testiranje normalnosti (Jevremović & Avdović, 2020a), empirijska funkcija raspodele (Jevremović i Avdović, 2020b; Bakir, 2012), kao i mnoge druge metode (Montgomery, 2009). Novi radovi tretiraju primenu kontrole kvaliteta u raznim disciplinama (Owens & Rioborue, 2022; Kotb & El-Ashkar, 2020; Neuburger, Walker, Sherlaw-Johnson, van der Meulen & Cromwell, 2017; Veljković i dr., 2015). Često su izučavana i dodatna unapređivanja EWMA karte ili rešenja bazirana na njoj (Ajadi i dr., 2021; Hu i dr., 2022). Takođe, predloženo je rešenje kontrole kvaliteta za podatke merene ordinalnom skalom (Ottensstreuer, Weiß & Testik, 2023).

Mnogi izvori vrše i analize poređenja performansi, tj. prosečnog broja uzoraka kontrole pre detekcije defekta kod kontrolnih karata (videti npr. Gupta, 2021; Adegoke, Smith, Anderson, Sanusi & Pawley, 2019; Muslim Raza & Butt, 2018; Rahardja, 2014). Nove kontrolne karte se uglavnom upoređuju sa onim najčešće korišćenim kao što su Šuhartova \bar{X} karta (videti npr. Oakland, 2003), CUSUM karta (Page, 1954) i EWMA karta (Roberts, 1959) i njihove eventualne modifikacije (videti npr. Montgomery, 2009).

Kao što je opšte poznato grafičke metode preliminarne analize normalnosti kao što su Q-Q dijagram, histogram, kutijasti dijagram (box plot) i sl. mogu biti korisno sredstvo pomoću kojeg se može steći uvid u saglasnost raspodele obeležja iz kojeg je izvučen uzorak sa nekom pretpostavljenom raspodelom. Ipak, kako su u pitanju metode koje nemaju formalan standard na osnovu kojeg se hipoteza normalnosti, ili saglasnosti generalno, može odbaciti, one su korisne u veoma ograničenom broju slučajeva. Istraživači diskutuju da je najbolja varijanta analize saglasnosti raspodele primena testa saglasnosti uz prikaz neke od grafičkih metoda radi boljeg uvida u zasnovanost zaključka. Štaviše, neki testovi su razvijeni po tom principu (Shapiro & Francia, 1972; Öztürk & Dudewicz, 1992).

U naučnim i stručnim istraživanjima i dalje se proučavaju i koriste poznate statističke tehnike među kojima su moć testa i p-vrednost. Moć testa i p-vrednost su rezultat izračunavanja po odgovarajućim formulama, na osnovu kojih se dobijaju bitne informacije o zaključku testiranja. Problem koji se javlja je da se pri tumačenju ovih pojmova često pravi puno grešaka, naročito u istraživanjima u oblasti društvenih nauka (videti npr. Lakens, 2021; Ioannidis, 2019; Vidgen & Yasserli, 2016; Goodman, 2008). Takođe se dovoljno opširno i jasno matematičko izlaganje p-vrednosti i moći testa uglavnom ne može naći u radovima koji su iz oblasti matematičke statistike, već u radovima iz oblasti medicine i psihologije (videte npr. McShane, Gal, Gelman, Robert, & Tackett, 2019; Ioannidis, 2019; Wang, Zhou, Wang, Tu & Feng, 2019; Greenland et al., 2016; Dahiru 2008; Goodman, 2008). Naknadne posledice navedenog su i da se p-vrednost pogrešno definiše (Reinhart, 2015, p. 8; Fischeti, 2018, p. 140; Upton & Cook, 2006, p. 200; Cabrera & McDougal, 2002, p. 76; Konold & Fan, 2010). Štaviše, mogu se navesti liste primera grešaka u definisanju i korišćenju p-vrednosti i razumevanju moći testa (Greenland i dr., 2016). Stoga će u ovom radu jedan deo biti posvećen rešavanju navedenih problema kroz preciznije matematičke obrade datih pojmova.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Testovi normalnosti su često razmatrana statistička metoda i njihove osobine, prednosti, mane i analize moći i poređenja detaljno su razrađeni (Avdović & Jevremović 2022; Thode, 2002; Arnastauskaitė, Ruzgas & Bražėnas, 2021; Thadewald & Büning, 2007). Analiza moći u slučajevima manje uobičajenih alternativnih raspodela je takođe izučavana (Noughabi, 2018; Sürücü, 2008; Gel, Miao & Gastwirth, 2007). Trenutno, najčešće korišćeni testovi normalnosti su test Kolmogorova i χ^2 test, zajedno sa Šapiro-Vilk testom i Anderson-Darling testom (Thode, 2002, str. 99-112; Arnastauskaitė i dr., 2021; Sürücü, 2008; Slakter, 1965; Noughabi, 2018; Razali & Wah, 2011; Ahmad & Khan, 2015; Boyerinas, 2018). To ukazuje na važnost jednostavnosti implementacije i brzine testiranja, pa se stoga koriste i testovi sa relativno nižom moći.

U ovom i drugim radovima pokazalo se značajnim vršiti odvojene analize i studije za slučajeve kada su parametri raspodele iz nulte hipoteze poznati i kada nisu poznati te se ocenjuju primenom odgovarajućih ocena. Naime, kod χ^2 testa broj stepeni slobode raspodele test-statistike se smanjuje povećanjem broja parametara nulte raspodele koji se ocenjuju (Jevremović, 2014, str. 256-258), što takođe zavisi i od izbora ocene koja se koristi. Treba imati na umu i to da je raspodela test statistike granična raspodela. U slučajevima kada raspodelu test-statistike nije moguće opisati za ocenjene parametre (ili uopšte), raspodela biva određena simulacijama. U mnogim studijama se pokazalo da se raspodela tada značajno menja. Promena raspodele biva takva da test-statistika bolje detektuje odstupanja od proste nulte raspodele. Razlog tome može biti činjenica da ocene parametara, kada raspodela test statistike ne zavisi od njih, za uzorak izvučen iz nulte raspodele ipak mogu odstupati od pretpostavljene vrednosti, a kako se odstupanje detektuje na osnovu uzorka, uzimanje vrednosti ocene za vrednost parametra daje precizniji rezultat. Samim tim, moć testa, pri testiranju proste nulte hipoteze, se povećava kada se parametri ocenjuju (Avdović & Jevremović, 2022, 2023; Stephens, 1975; Lilliefors, 1967). Novi testovi i dalje se predlažu, primenjuju, unapređuju i analiziraju kako bi se odredile i istakle njihove prednosti u odnosu na postojeće (Avdović & Jevremović 2022, 2023; Arnastauskaitė i dr., 2021; Sürücü, 2008; Gel i dr., 2007; Bakshaev, 2009; Coin, 2008; Desgagné & Lafaye de Micheaux, 2017). Kod nekih testova zasnovanost rezultata zahteva komplikovana matematička izvođenja (Sürücü, 2008; Anderson & Darling, 1952; Cochran, 1952; Obradović, Jovanović & Milošević, 2014).

2. Slučajni i pseudoslučajni brojevi

U ovom poglavlju biće definisani i okarakterisani slučajni i pseudoslučajni brojevi, metode njihovog veštačkog dobijanja, tj. generisanja ili simuliranja, i načini provere tih metoda preko osobina koje dobijeni brojevi treba da imaju. Važnost ovog poglavlja sledi iz činjenice da su slučajni brojevi neophodni za Monte Karlo simulacije (a i druge tipove simulacija), tj. modeliranje obeležja raznih raspodela. O modeliranju raspodela i njihovom značaju za istraživanje ovog rada diskutovano je u posebnom poglavlju. Dodatno, deo rezultata ovog istraživanja jeste novi generator pseudoslučajnih brojeva koji će biti izložen i čije će performanse biti analizirane i poređene sa performansama drugih često primenjivanih generatora.

2.1. Generatori slučajnih brojeva

Pri dobijanju suštinski važnih rezultata istraživanja kao što su izračunavanja i poređenja moći testova ili prosečnih dužina niza kontrolnih tačaka pre detekcije narušenog kvaliteta koriste se simulacije raznih modela. Pod modelima se uglavnom podrazumevaju posebne raspodele verovatnoća jer one opisuju skoro svako obeležje koje je od značaja u prirodi i društvu. Da bi se modelirali uzorci tih raspodela potrebni su nizovi slučajnih brojeva ili adekvatno generisani nizovi pseudoslučajnih brojeva.

Definicija 2.1. Slučajni brojevi su nezavisno realizovane vrednosti slučajne veličine X sa uniformnom raspodelom $\mathcal{U}(0,1)$.

Slučajni brojevi se mogu dobiti korišćenjem nekog fizičkog aparata: kockica, simetričnog novčića, ruleta, brojača radioaktivnog zračenja... koji se naziva **generatorom slučajnih brojeva**. Sledeći stavovi daju neke osobine slučajnih brojeva.

Stav 2.1. Važe sledeća tvrđenja.

- a. *Ako slučajna veličina X ima uniformnu $\mathcal{U}(0,1)$ raspodelu, tada slučajna veličina $1 - X$ takođe ima $\mathcal{U}(0,1)$ raspodelu.*
- b. *Ako slučajna veličina X ima uniformnu $\mathcal{U}(0,1)$ raspodelu, tada slučajna veličina*

$$Y = a + (b - a)X$$

ima uniformnu $\mathcal{U}(a, b)$ raspodelu.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 2.2. Neka je $n \in \mathbb{N}$ proizvoljno i neka je X slučajna veličina sa uniformnom raspodelom $\mathcal{U}(0,1)$. Slučajna veličina $Y = \{nX\}$, gde $\{\cdot\}$ predstavlja decimalni deo broja, takođe ima $\mathcal{U}(0,1)$ raspodelu.

Definicija 2.2. Pseudoslučajni brojevi su brojevi iz intervala $(0,1)$ koji se dobijaju na osnovu jednog ili više unapred određenih pravila korišćenjem jednog ili više poznatih (zadatih) brojeva. Algoritam kojim se ti brojevi dobijaju je **generator pseudoslučajnih brojeva**, a broj ili brojevi koji se za to koriste čine **seme** generatora. Broj elemenata semena je **red** generatora.

Generatori su često formulisani tako da se pomoću njih dobijaju brojevi iz nekog intervala koji nije $(0, 1)$, ali se odgovarajućom transformacijom lako mogu dobiti brojevi iz $(0, 1)$ ili bilo kojeg drugog intervala. Takođe, generatori pseudoslučajnih brojeva se konvencionalno nazivaju generatorima slučajnih brojeva.

Generatori pseudoslučajnih brojeva se formalno mogu opisati rekurentnom formulom

$$X_{i+1} = f(X_i, X_{i-1}, \dots, X_{i-k+1}), \quad X_i \in \mathbb{R}, i = 0, 1, 2, \dots$$

gde je $(X_{k-1}, X_{k-2}, \dots, X_1, X_0)$ seme generatora, a $k \in \mathbb{N}$ red generatora.

Pseudoslučajni brojevi se mogu dobiti na brži i jednostavniji način od slučajnih brojeva, a, za dobro izabran generator, imaju (skoro) sve osobine koje imaju slučajni brojevi. Prednošću pseudoslučajnih brojeva se smatra i to što se već dobijeni niz pseudoslučajnih brojeva može ponovno generisati ukoliko je to potrebno, tj. "eksperiment" se može ponoviti u identičnim uslovima.

Definicija 2.3. Neka je

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+j}, \dots \quad (2.1)$$

neki niz pseudoslučajnih brojeva generisanih nekim generatorom. Ako se u (2.1) desi da je $x_{n+s_j} = x_{n-t+j}$, $j \in \{1, \dots, t\}$, $(\forall s \in \mathbb{N})$ gde su t i n neki prirodni brojevi takvi da $t \leq n$, tada je (za seme (x_1, x_2, \dots, x_k)) broj n period tog generatora.

Period generatora se u opštem slučaju razlikuje za različita semena, a mora biti barem 10^9 da bi taj generator bio prihvaćen kao svrsishodan. Čak se i taj period smatra relativno kratkim zbog brzine kojom se pomoću računara generišu svi članovi do perioda (Gentle, 2002, str. 32-33).

2.2. Primeri generatora pseudoslučajnih brojeva

Metoda sredine kvadrata

Metoda sredine kvadrata je algoritam generisanja pseudoslučajnih brojeva, objavljen 1946. godine od strane mađarsko-američkog matematičara Džona Fon Nojmana².

Definicija 2.4. Algoritam *metode sredine kvadrata* dat je formulom:

$$X_{i+1} = \{10^{-2k} \cdot [10^{3k} \cdot X_i^2]\}, X_i \in (0,1), i \in \mathbb{N}$$

gde je $\{\cdot\}$ decimalni, $[\cdot]$ ceo deo broja X_i , a broj X_i ima oblik $0, \varepsilon_1 \varepsilon_2 \dots \varepsilon_{2k}$, $k \in \mathbb{N}$. Za seme se uzima proizvoljan broj $X_0 \in (0,1)$ sa odgovarajućim brojem cifara. Broj $n = 2k \in \mathbb{N}$ predstavlja broj decimala generisanih brojeva (seme se takođe bira tako da ima n decimala).

Kako ima 10 cifara, a n mesta, period ne može biti duži od 10^n .

Ovaj algoritam se može prilagoditi i za generisanje slučajnih cifara. Naime, n slučajnih cifara se može dobiti tako što proizvoljan prirodan broj (seme) koji ima n cifara treba kvadrirati, a zatim iz dobijenog broja izdvojiti srednjih n cifara. Ponavljanje ovog postupka se okončava generisanjem onoliko brojeva koliko iznosi period generatora koji, pak, zavisi od semena (pa može biti i vrlo kratak).

Prost linearni kongruentni generator

Derik Henri Lemer³ je 1948. godine predložio **prost linearni kongruentni generator** kao "izvor slučajnih brojeva". U nizovima ovog generatora svaki član određuje sledeći kao argument proste linearne funkcije čijoj vrednosti se određuje ostatak pri deljenju zadatim prirodnim brojem, tj. primenjuje se kongruencija po modulu.

Definicija 2.5. *Prost linearni kongruentni generator* dat je formulom:

$$X_i = (a \cdot X_{i-1} + c) \bmod m, \quad 0 \leq X_i < m; \quad (2.2)$$

² John Von Neumann (1903-1957), američki matematičar mađarskog porekla.

³ Derrick Henry Lehmer (1905-1991), američki matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

gde važi $a, m, i, X_i \in \mathbb{N}$, a broj a je **množilac**, c **priraštaj**, a broj m **modul** generatora.

Generator (2.2) je **multiplikativni kongruentni generator** kada je $c = 0$, tj. ako je dat formulom

$$X_i = (a \cdot X_{i-1}) \bmod m, \quad 0 \leq X_i \leq m$$

gde važi $a, m, i, X_i \in \mathbb{N}$. U oba slučaja se za seme bira proizvoljan broj $X_0 \in \mathbb{N}$. Rezultujući niz brojeva je **Lemerov niz**.

Najbitnije osobine prostog linearnog kongruentnog generatora date su sledećim stavovima (Knuth, 1998, str. 17-19).

Teorema 2.1. Generator (2.2) ima maksimalan period m , ako su ispunjeni sledeći uslovi:

1. Brojevi m i c su uzajamno prosti.
2. Svaki prost delilac p broja m je delilac i broja $a - 1$.
3. Broj $a - 1$ je deljiv sa 4 kada je m deljiv sa 4.

Definicija 2.6. Broj a je **primitivni koren modula m** ako u jednačini

$$a^k = 1 \bmod m$$

važi $k = \phi(m)$, gde je ϕ Ojlerova⁴ funkcija (broj prostih brojeva manjih od m).

Teorema 2.2. (Ojler-Ferma) Ako su a i m uzajamno prosti, onda važi

$$a^{\phi(m)} = 1 \bmod m.$$

Lema 2.1. Ako za multiplikativni LKG $X_i = (aX_{i-1}) \bmod m$ važi $a^k = 1 \bmod m$, onda period ovog generatora nije veći od k .

Napomena 2.1. Važna činjenica koja direktno sledi je ta da je za prost broj m broj primitivnih korena modula m jednak $\phi(m) = m - 1$, a u slučaju da je a jedan od tih brojeva, onda multiplikativni LKG dostiže svoj maksimalni period koji je upravo $\phi(m) = m - 1$.

⁴ Leonhard Paul Euler (1707-1783), švajcarski matematičar i fizičar.

Drugi kongruentni generatori

Višestruko rekurzivni multiplikativni kongruentni generator dat je formulom

$$X_i = (a_1X_{i-1} + a_2X_{i-2} + \dots + a_kX_{i-k}) \bmod m$$

za $k > 1$, gde su $a_1, \dots, a_k, m, X_i \in \mathbb{N}$.

Fibonačijev generator sa raskorakom dat je formulom:

$$X_i = (X_{i-j} + X_{i-k}) \bmod m, \quad 0 \leq X_i \leq m$$

gde je $i \geq j, i \geq k, k > j + 1, i, j, k \in \mathbb{N}$ i $m, X_i \in \mathbb{N}$.

Linearni kongruentni generator može biti uopšten u **Knutov⁵ kvadratni kongruentni generator** koji je dat formulom:

$$X_{i+1} = (dX_i^2 + aX_i + c) \bmod m$$

gde su X_{i+1}, d, a, c i m prirodni brojevi $0 \leq X_i \leq m$.

Uz odgovarajući izbor koeficijenata d, a i c ovaj generator može dostići maksimalan period koji iznosi m , samo ako je m prost broj (Avdović, 2019, str. 21).

Blum⁶-Blum⁷-Šab⁸ generator je dat formulom:

$$X_{i+1} = X_i^2 \bmod m$$

gde važi da su X_i i m prirodni brojevi, $0 \leq X_i \leq m$ i $i > 1, i \in \mathbb{N}$.

Generatori **sabiranja sa prenošenjem** i **množenja sa prenošenjem** su dva generatora koja koriste ideju LKG-a. Razvijanju ovih generatora najviše je doprineo Džordž Marsalja⁹ (Gentle, 2002, str. 35-36).

Generator **sabiranja sa prenošenjem** dat je jednačinom

⁵ Donald Ervin Knuth (1938-), američki informatičar i matematičar.

⁶ Leonore Blum (1942-), američka informatičarka i matematičarka.

⁷ Manuel Blum (1938-), venezuelanski informatičar.

⁸ Michael Shub (1943-), američki matematičar.

⁹ George Marsaglia (1924-2011), američki matematičar i informatičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$X_{i+1} = (X_{i-s} + X_{i-r} + c_i) \bmod m, \quad i \geq s > r$$

gde su $X_j \in \{0, 1, \dots, m-1\}$, $j \in \mathbb{N}$, $m, s, r \in \mathbb{N}$, $c_0 = 0$ i $c_{j+1} = 0$ ako je $X_{i-s} + X_{i-r} + c_i < m$, a $c_{j+1} = 1$ u suprotnom. Brojevi c_i nazivaju se prenošenjem.

Generator **množenja sa prenošenjem** dat je jednačinom

$$X_{i+1} = (aX_i + c_i) \bmod m, \quad i \geq s > r$$

gde su $a, X_j \in \{0, 1, \dots, m-1\}$, $j \in \mathbb{N}$, $m \in \mathbb{N}$, $c_0 = 0$ i $c_{j+1} = 0$ ako je $X_{i-s} + X_{i-r} + c_i < m$, a $c_{j+1} = 1$ u suprotnom.

Proširenje linearnog kongruentnog generatora

Rezultati koji slede su rezultati autora. Prikazani su definicija proširenja LKG-a, kao i različite varijante LKG-a sa proširenjem i iskazane su i dokazane neke osobine tih generatora koje se tiču slučajnosti generisanih brojeva i perioda generatora.

Definicija 2.7. LKG sa proširenjem je generator dat rekurentnom formulom

$$X_i = Y(aX_{i-1} + c, k(X_{i-1})) \bmod m,$$

gde je $Y: \mathbb{N}^2 \rightarrow \mathbb{N}$, linearna funkcija po X_{i-1} , $a, c \neq 0$ i

$$k(X_{i-1}) = \sum_{j=1}^i I(X_j = X_{i-1}),$$

gde je I indikator događaja. Funkcija k je **proširenje LKG-a** (Avdović, 2023).

Drugim rečima, proširenje modifikuje rekurentnu formulu LKG-a uzimajući u obzir broj prethodnih pojavljivanja argumenta (zadnjeg generisanog broja). Uticaj proširenja na LKG se može diskutovati kroz sledeće slučajeve.

- (i) **LKG sa aditivnim proširenjem** dat je rekurentnom formulom

$$X_i = (a \cdot X_{i-1} + c \cdot k(X_{i-1})) \bmod m.$$

- (ii) **LKG sa multiplikativnom proširenjem** dat je rekurentnom formulom

$$X_i = (a \cdot X_{i-1} \cdot k(X_{i-1}) + c) \bmod m.$$

- (iii) **Prošireni LKG** dat je rekurentnom formulom

$$X_i = k(X_{i-1}) \cdot (a \cdot X_{i-1} + c) \bmod m.$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Uslov $a, c \neq 0$ uveden je jer u suprotnom nastaju slučajevi koje nema potrebe razmatrati. Na primer, za $c = 0$ generator (i) postaje multiplikativni LKG. U slučajevima (ii) i (iii), za $c = 0$, jednom kada se dobije $X_i = 0$, $i = 0, 1, 2, \dots$ važi $X_{i+j} = 0$, $j \in \mathbb{N}$. Štaviše, može se zanemariti i slučaj u kojem je $a = 1$ i $c = 1$ (istovremeno), jer tada generisani brojevi nisu uniformno raspodeljeni (Knuth, 1998, str. 16-17).

Stav 2.3. *Neka su X_{i-1} , X_i i X_{i+1} brojevi generisani generatorom (i), (ii) ili (iii). Bar dva su međusobno različita za bilo koje $i \in \mathbb{N}$ i bilo koji izbor parametara generatora.*

Dokaz. Neka su brojevi generisani generatorom (i).

Neka važi suprotno, tj. neka važi

$$X_{i_0-1} = X_{i_0} = X_{i_0+1}$$

za neko $i_0 \in \mathbb{N}$. Tada je

$$X_{i_0} = (a \cdot X_{i_0-1} + c \cdot k(X_{i_0-1})) \bmod m = X_{i_0-1}$$

kao i

$$X_{i_0+1} = (a \cdot X_{i_0} + c \cdot k(X_{i_0})) \bmod m = (a \cdot X_{i_0-1} + c \cdot (k(X_{i_0-1}) + 1)) \bmod m = X_{i_0-1}.$$

Što znači da mora važiti $c = 0$ što je nemoguće jer je suprotno uslovima definicije generatora sa proširenjem. Do protivrečnosti je dovela pretpostavka da su neka tri uzastopna broja generisana generatorom (i) jednaka pa mora važiti suprotno, tj. da su bar dva međusobno različita.

Neka su sada brojevi generisani generatorom (ii).

Neka važi suprotno, tj. neka važi

$$X_{i_0-1} = X_{i_0} = X_{i_0+1}$$

za neko $i_0 \in \mathbb{N}$. Tada je

$$X_{i_0} = (a \cdot X_{i_0-1} \cdot k(X_{i_0-1}) + c) \bmod m = X_{i_0-1}$$

kao i

$$X_{i_0+1} = (a \cdot X_{i_0} \cdot k(X_{i_0}) + c) \bmod m = (a \cdot X_{i_0-1} \cdot (k(X_{i_0-1}) + 1) + c) \bmod m = X_{i_0-1}.$$

Što znači da mora važiti $a = 0$ što je nemoguće jer je suprotno uslovima definicije generatora sa proširenjem. Do protivrečnosti je dovela pretpostavka da su neka tri uzastopna broja

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

generisana generatorom (ii) jednaka pa mora važiti suprotno, tj. da su bar dva međusobno različita.

Neka su sada brojevi generisani generatorom (iii).

Neka važi suprotno, tj. neka važi

$$X_{i_0-1} = X_{i_0} = X_{i_0+1}$$

za neko $i_0 \in \mathbb{N}$. Tada je

$$k(X_{i_0-1}) \cdot (a \cdot X_{i_0-1} + c) \bmod m = X_{i_0-1}$$

kao i

$$(k(X_{i_0}) + 1) \cdot (a \cdot X_{i_0-1} + c) \bmod m = X_{i_0-1}.$$

Oduzimanjem dve poslednje jednakosti, dobija se

$$(a \cdot X_{i_0-1} + c) \bmod m = 0$$

Onda može nastupiti jedan od dva slučaja. Prvi slučaj je da je

$$(X_{i_0-1} = 0 \wedge c = 0) \vee (a = 0 \wedge c = 0) \vee (X_{i_0-1} = 0 \wedge a = 0 \wedge c = 0)$$

što je opet suprotno uslovu definicije generatora sa proširenjem.

Drugi slučaj je da važi

$$a \cdot X_{i_0-1} + c = qm, q \in \mathbb{N} \Leftrightarrow X_{i_0-1} = \frac{qm - c}{a} \in \{1, \dots, m - 1\}.$$

Slučaj sa $X_{i_0-1} = q^*m, q^* \in \mathbb{N}$ se odbacuje jer se dobija $c = 0$. Tada važi i

$$X_{i_0} = k(X_{i_0-1}) \cdot (a \cdot X_{i_0-1} + c) \bmod m = k(X_{i_0-1}) \cdot \left(a \cdot \frac{qm - c}{a} + c\right) \bmod m = 0.$$

To konačno znači $X_{i_0-1} \neq X_{i_0}$, što je suprotno pretpostavci da su jednaki. Do protivrečnosti je dovela pretpostavka da su neka tri uzastopna broja generisana generatorom (iii) jednaka pa mora važiti suprotno, tj. da su bar dva međusobno različita. ■

Sledeći stav ukazuje na to da će se jednom generisan par brojeva (X, Y) opet pojaviti u istom redosledu, ako se posle X bar m puta generiše neki broj različit od Y .

Stav 2.4. Neka su X i Y dva broja uzastopno generisana generatorom (i), (ii) ili (iii), i neka je

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned} k_X &= \min\{d \in \mathbb{N} \mid (aX + dc) \bmod m = Y\}, \text{ u slučaju (i),} \\ k_X &= \min\{d \in \mathbb{N} \mid (aXd + c) \bmod m = Y\}, \text{ u slučaju (ii),} \\ k_X &= \min\{d \in \mathbb{N} \mid d(aX + c) \bmod m = Y\}, \text{ u slučaju (iii).} \end{aligned}$$

Tada je $(X, Y) = (X_{i-1}, X_i)$ za neko $i \in \mathbb{N}$ ako i samo ako je

$$k(X_{i-1}) = lm + k_X, \quad l \in \mathbb{N}.$$

Dokaz. Generator (i).

(\Rightarrow) Neka su $(X, Y) = (X_{i-1}, X_i)$ parovi brojeva generisanih uzastopno.

Tada važi

$$X_i = (a \cdot X_{i-1} + c \cdot k(X_{i-1})) \bmod m \Leftrightarrow Y = (a \cdot X + c \cdot k(X_{i-1})) \bmod m.$$

Kako je $k(X_{i-1}) \in \mathbb{N}$, to postoje $l \in \mathbb{N}_0$ i $r \in \{0, 1, \dots, m-1\}$ takvi da važi $k(X_{i-1}) = lm + r$, gde je r najmanji takav broj. Sledi

$$Y = (a \cdot X + c \cdot (lm + r)) \bmod m = (a \cdot X + c \cdot r) \bmod m$$

pa zbog minimalnosti r i zadnje jednakosti sledi $k_X = r$, što je i trebalo dokazati.

(\Leftarrow) Neka je $X_{i-1} = X$, za neko $i \in \mathbb{N}$, i neka je $k(X_{i-1}) = lm + k_X$, $l \in \mathbb{N}_0$. Tada

$$\begin{aligned} X_i &= (a \cdot X_{i-1} + c \cdot k(X_{i-1})) \bmod m \\ &= (a \cdot X_{i-1} + c \cdot (lm + k_X)) \bmod m \\ &= (a \cdot X_{i-1} + c \cdot k_X) \bmod m \\ &= Y \end{aligned}$$

pa je $X_i = Y$, jer je generatorom (i) nakon X generisan Y , a nakon X_{i-1} generisan je X_i . Dakle, $(X, Y) = (X_{i-1}, X_i)$, što je i trebalo dokazati.

Generator (ii).

(\Rightarrow) Neka su $(X, Y) = (X_{i-1}, X_i)$ parovi brojeva generisanih uzastopno.

Tada važi

$$X_i = (a \cdot X_{i-1} \cdot k(X_{i-1}) + c) \bmod m \Leftrightarrow Y = (a \cdot X \cdot k(X_{i-1}) + c) \bmod m.$$

Kako je $k(X_{i-1}) \in \mathbb{N}$, to postoje $l \in \mathbb{N}_0$ i $r \in \{0, 1, \dots, m-1\}$ takvi da važi $k(X_{i-1}) = lm + r$, gde je r najmanji takav broj. Sledi

$$Y = (a \cdot X \cdot (lm + r) + c) \bmod m = (a \cdot X \cdot r + c) \bmod m$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

pa zbog minimalnosti r i zadnje jednakosti sledi $k_X = r$, što je i trebalo dokazati.

(\Leftarrow) Neka je $X_{i-1} = X$, za neko $i \in \mathbb{N}$, i neka je $k(X_{i-1}) = lm + k_X$, $l \in \mathbb{N}_0$. Tada

$$\begin{aligned} X_i &= (a \cdot X_{i-1} \cdot k(X_{i-1}) + c) \bmod m \\ &= (a \cdot X_{i-1} \cdot (lm + k_X) + c) \bmod m \\ &= (a \cdot X_{i-1} \cdot k_X + c) \bmod m \\ &= Y \end{aligned}$$

pa je $X_i = Y$, jer je generatorom (ii) nakon X generisan Y , a nakon X_{i-1} generisan je X_i . Dakle, $(X, Y) = (X_{i-1}, X_i)$, što je i trebalo dokazati.

Generator (iii).

(\Rightarrow) Neka su $(X, Y) = (X_{i-1}, X_i)$ parovi brojeva generisanih uzastopno.

Tada važi

$$X_i = k(X_{i-1}) \cdot (a \cdot X_{i-1} + c) \bmod m \Leftrightarrow Y = k(X_{i-1}) \cdot (a \cdot X + c) \bmod m.$$

Kako je $k(X_{i-1}) \in \mathbb{N}$, to postoje $l \in \mathbb{N}_0$ i $r \in \{0, 1, \dots, m-1\}$ takvi da važi $k(X_{i-1}) = lm + r$, gde je r najmanji takav broj. Sledi

$$Y = (lm + r) \cdot (a \cdot X + c) \bmod m = r \cdot (a \cdot X + c) \bmod m$$

pa zbog minimalnosti r i zadnje jednakosti sledi $k_X = r$, što je i trebalo dokazati.

(\Leftarrow) Neka je $X_{i-1} = X$, za neko $i \in \mathbb{N}$, i neka je $k(X_{i-1}) = lm + k_X$, $l \in \mathbb{N}_0$. Tada

$$\begin{aligned} X_i &= k(X_{i-1}) \cdot (a \cdot X_{i-1} + c) \bmod m \\ &= (lm + k_X) \cdot (a \cdot X_{i-1} + c) \bmod m \\ &= k_X \cdot (a \cdot X_{i-1} + c) \bmod m \\ &= Y \end{aligned}$$

pa je $X_i = Y$, jer je generatorom (iii) nakon X generisan Y , a nakon X_{i-1} generisan je X_i . Dakle, $(X, Y) = (X_{i-1}, X_i)$, što je i trebalo dokazati. ■

Napomena 2.2. Činjenica da je $k_X \in \{0, 1, \dots, m-1\}$ sledi iz njegove minimalnosti za uslov dat stavom 2.4. Na primer, kod generatora (i), pretpostavivši suprotno dobija se da moraju postojati $l \in \mathbb{N}$ i $r \in \{0, 1, \dots, m-1\}$ tako da je $k_X = lm + r$. Tada

$$Y = (aX + k_X c) \bmod m = (aX + (lm + r)c) \bmod m = (aX + rc) \bmod m$$

Dakle, $r < k_X$ i zadovoljava uslov stava 2.4, što je nemoguće jer je k_X najmanji takav broj. Zaključak se donosi analogno za generatore (ii) i (iii).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stavovi 2.3 i 2.4 ukazuju da proširenje LKG-a ne narušava slučajnost generisanih brojeva. Sledeća teorema precizira kakav uticaj proširenje ima na period rezultujućeg generatora.

Teorema 2.3. Neka je $p \neq 2$ prost broj, c i p^n , $n \in \mathbb{N}$ uzajamno prosti i

$$h(X_{i-1}) = \begin{cases} k(X_{i-1}), & k(X_{i-1}) \neq p \\ k(X_{i-1}) + 1, & k(X_{i-1}) = p \end{cases} \quad i \in \mathbb{N}.$$

Generator dat rekurentnom formulom

$$X_i = \left((p^l + 1) \cdot X_{i-1} + c \cdot h(X_{i-1}) \right) \bmod p^n, \quad (2.3)$$

gde je $l \in \{1, \dots, n-1\}$, ima minimalnu dužinu perioda $sp^n + d_i$, a maksimalnu dužinu perioda $sp^n + \sum_{j=1}^v d_j$, gde su $s = \sum_{j=1}^{p^n-1} I(j \neq p^l)$, d_i period generatora datog sa $(p^l + 1) \cdot X_{i-1} \bmod p^n$ za seme $X_0 = i$, $i \in \{0, 1, \dots, p^n - 1\}$, i v broj različitih varijanti tog generatora dobijenih za svako od mogućih semena i .

Dokaz. Za $a = p^l + 1$ broj $a - 1 = p^l$ deli p^n . Kako su c i p^n uzajamno prosti, svi brojevi oblika $(j \cdot c) \bmod p^n$, $j = 1, \dots, p^n - 1$, $j \neq p^l$, $l \in \{1, \dots, n-1\}$ su takođe uzajamno prosti sa p^n , pa su i brojevi oblika $(h(X_{i-1}) \cdot c) \bmod p^n$. Uslov $p \neq 2$ povlači da 4 ne deli ni $a - 1 = p^l$ ni p^n . Stoga generator (2.3) za fiksireno $h(X_{i-1})$ zadovoljava uslove teoreme 2.1, pa svaki od tih generatora dobijenih za različite $h(X_{i-1})$ dostiže maksimalni period od p^n za bilo koje seme. Generator (2.3) formira se od s takvih generatora primenjenih jedan za drugim. Kada svaki od tih generatora nakon nekog indeksa $i - 1$ dostigne svoj maksimalni period p^k , dobija se $h(X_{i-1}) = p^k$. Odatle su svi predstojeći generisani brojevi rezultat generatora

$$(p^l + 1) \cdot X_{i-1} \bmod p^n \quad (2.4)$$

čiji je period d_i za seme $X_0 = i$, $i \in \{0, 1, \dots, p^k - 1\}$. Dakle, ako je $X_{sp^n + d_i} = X_{sp^n}$, period je dostignut i jednak upravo $sp^n + d_i$ jer je seme generatora (2.4) ponovljeno. U suprotnom, taj se postupak nastavi dok ne dođe to takvog slučaja. Ako se u v navrata dobije različito seme generatora (2.4), onda ih više ne može biti, pa će period generatora (2.3) biti dostignut. To znači da period ne može biti veći od $sp^n + \sum_{j=1}^v d_j$. ■

Napomena 2.3. Preciznije informacije o broju d_i zavise od brojeva a i m , tj. $p^l - 1$ i p^n , a mogu se dobiti razmatranjem na osnovu definicije 2.5, teoreme 2.1, leme 2.1 i napomene 2.1.

Napomena 2.4. Često se više generatora kombinuje u jedan čime se period produžava i dobiju se kvalitetniji nizovi brojeva (Gentle, 2002, str. 46-51).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Mersenov¹⁰ tvister

Mersenov tvister (Mersenne Twister) je generator pseudoslučajnih brojeva konstruisan od strane Macumota i Nišimure (1998). Ovaj generator se pokazao korisnim jer njegove varijante imaju velike periode, a generisani brojevi zadovoljavaju karakteristike slučajnosti i algoritam izvršavanja je jednostavan što garantuje brzinu generisanja (Shema, 2012). Verzija Mersenovog Tvistera koja je dostupna u mnogim programskim jezicima, poznata kao MT19937, ima izuzetno dug period od $2^{19937} - 1$. Ove karakteristike čine ovaj algoritam jednim od najčešćih generatora u programskim jezicima i paketima.

Ono što je negativna strana ovog generatora je to što je formalni opis algoritma, a samim tim i kod kojim bi se programirao, veoma komplikovan. Stoga će i ovde biti izostavljen. Takođe, varijanta MT19937 je ona koja se uglavnom jedina koristi, pa uprkos dužini perioda i brzini generisanja, drugi generatori imaju prednost jer je moguće koristiti onu od njihovih varijanti koja je najprikladnija za neko istraživanje, najčešće u smislu dužine perioda.

2.3. Testiranje slučajnosti generisanih brojeva

Brojevi dobijeni generatorom pseudoslučajnih brojeva teorijski ne mogu biti zaista slučajni, ali ti brojevi mogu imati ista svojstva koja imaju slučajni brojevi. Svojstva slučajnih brojeva se u većini slučajeva svode na to da su oni uniformno, tj. ravnomerno raspodeljeni na posmatranom intervalu i da su raspodeljeni nezavisno jedan od drugog. Testiranje tih osobina se može vršiti na dva načina.

Prvi način jeste testiranje saglasnosti raspodele obeležja sa uniformnom raspodelom. Poznat je i razrađen veliki broj testova saglasnosti kojima se to može uraditi kao što su test Kolmogorova-Smirnova, Pirsonov¹¹ χ^2 test i drugi. Mnogo detaljnije o ovome biće u narednim poglavljima ovog rada pa će dalja razrada ove teme dotle biti odložena.

Drugi način jeste korišćenje tzv. testova slučajnosti koji testiraju hipotezu da je dobijeni uzorak prost slučajan uzorak. U ovom poglavlju su navedeni primeri dva testa slučajnosti, a više primera i informacija se može naći kod Knuta (1998, str. 41-93) ili Džentla (2002, str. 61-87). Zajedničko za testove koji slede je to da je test-statistika diskretna slučajna veličina, čiji se parametri raspodele ili raspodela u potpunosti mogu odrediti, ali se u praksi koristi aproksimacija normalnom raspodelom.

¹⁰ Marin Mersenne (1588-1648), francuski matematičar.

¹¹ Karl Pearson (1857-1936), engleski matematičar (biostatističar).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Test tačaka rasta

Definicija 2.8. Neka je dat prost slučajan uzorak X_1, X_2, \dots, X_n , $n \in \mathbb{N}$. Element X_j , $j \in \{1, \dots, n-1\}$ je tačka rasta ako je $X_j < X_{j+1}$.

Na osnovu definicije, u nizu može biti najviše $n-1$ tačka rasta, kada je niz vrednosti rastući, a najmanje 0, kada je niz vrednosti nerastući.

Test-statistika testa tačaka rasta je

$$U = \sum_{j=1}^{n-1} I(X_j < X_{j+1}),$$

gde je $I(X_j < X_{j+1})$ **indikator rasta**. Dakle, U broji tačke rasta u polaznom uzorku. Može se dokazati

$$E(U) = \frac{n-1}{2} \quad \text{i} \quad D(U) = \frac{n+1}{12}.$$

Na osnovu centralne granične teoreme (tj. nekih njenih modifikacija za slabo korelisane uzorke), za dovoljno veliko n važiće (Turković, 2021, str. 22; Mališić & Jevremović, 2008, str. 183; Kanji, 1993, str. 105)

$$U \sim \mathcal{N}\left(\frac{n-1}{2}, \frac{n+1}{12}\right).$$

Kritična oblast se određuje iz uslova

$$P_{H_0}(U \leq c_1) = P_{H_0}(U \geq c_2) = \frac{\alpha}{2},$$

pa je data sa $W_n = (-\infty, c_1] \cup [c_2, +\infty)$.

Budući da kritičnu oblast u ovom obliku nije moguće direktno odrediti, koristi se normirana statistika U , tj.

$$U^* = \frac{U - \frac{n-1}{2}}{\sqrt{\frac{n+1}{12}}} \sim \mathcal{N}(0,1).$$

Dakle, kritična oblast se određuje iz uslova

$$P_{H_0}(|U^*| \geq c) = \alpha$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

a ima oblik $W = (-\infty, -c] \cup [c, +\infty)$. Ako se za uzorak X_1, X_2, \dots, X_n realizovana vrednost test-statistike U (U^*), nalazi unutar kritične oblasti W_n (W), nulta hipoteza se odbacuje.

Test tačaka zaokreta

Definicija 2.9. Tačka zaokreta uzorka X_1, X_2, \dots, X_n je element tog uzorka $X_i, i \in \{2, \dots, n-1\}$ za koji važi $X_{i-1} < X_i \wedge X_i > X_{i+1}$ ili $X_{i-1} > X_i \wedge X_i < X_{i+1}$. Za uzorak obima $n \in \mathbb{N}$ može biti najmanje 0, a najviše $n-2$ tačke zaokreta.

Zakoni raspodele indikatora događaja " X_i je tačka zaokreta uzorka X_1, X_2, \dots, X_n " dati su sa:

$$I_i: \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad i \in \{2, \dots, n-1\}.$$

Test-statistika je:

$$U = I_2 + I_3 + \dots + I_{n-1}.$$

Dakle, U predstavlja broj tačaka zaokreta uzorka koji se testira.

Može se dokazati da je $E(U) = \frac{2}{3}(n-2)$ i $D(U) = \frac{16n-29}{90}$ (Mališić, Jevremović, 2008, str. 180-182), a na osnovu centralne granične teoreme, za uzorak dovoljno velikog obima ($n > 15$) važi (Kanji, 1993, str. 104)

$$U \sim N\left(\frac{2}{3}(n-2); \frac{16n-29}{90}\right).$$

Kritična oblast se određuje iz uslova:

$$P_{H_0}(U \leq c_1) = P_{H_0}(U \geq c_2) = \frac{\alpha}{2}$$

pa je data sa $W_n = (-\infty, c_1] \cup [c_2, +\infty)$.

Budući da kritičnu oblast u ovom obliku nije moguće direktno odrediti, koristi se normirana statistika U , tj.

$$U^* = \frac{U - \frac{2}{3}(n-2)}{\sqrt{\frac{16n-29}{90}}} \sim N(0,1).$$

Dakle, kritična oblast se određuje iz uslova

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$P_{H_0}(|U^*| \geq c) = \alpha$$

a ima oblik $W = (-\infty, -c] \cup [c, +\infty)$. Ako se za uzorak X_1, X_2, \dots, X_n realizovana vrednost test-statistike $U (U^*)$, nalazi unutar kritične oblasti $W_n (W)$, nulta hipoteza se odbacuje.

Test koraka

Da bi se osnovni pojmovi mogli definisati, za dobijeni uzorak X_1, X_2, \dots, X_n potrebno je svakom njegovom elementu po nekom pravilu dodeliti znak "+" ili "-" (ili neka druga dva simbola). U ovom radu će se koristiti sledeće pravilo:

"Ako važi $X_i < X_{i+1}; i \in \{1, \dots, n - 1\}$, elementu X_i dodeliti znak "-", a u suprotnom znak "+". Ako važi $X_i = X_{i+1}$, elementu X_i dodeliti znak elementa X_{i-1} . Ako je $X_1 = X_2$ uporediti taj član niza sa prvim članom različitim od njega. Element X_n uporediti sa X_1 na isti način" (Avdović, 2019, str. 26).

Definicija 2.10. Korak je svaki neprekidan niz simbola "+" (ili "-") dodeljenih elementima uzorka, koji prethode ili slede nizu simbola "-" (ili "+").

Neka je broj znakova "+" u uzorku n_1 , a broj znakova "-" n_2 .

Test-statistika K je slučajna veličina koja predstavlja broj koraka u uzorku. Raspodela statistike K data je sa (Vuković, 2007, str. 229)

$$P(K = k) = \begin{cases} \frac{2 \binom{n_1-1}{\frac{k}{2}-1} \binom{n_2-1}{\frac{k}{2}-1}}{\binom{n_1+n_2}{n_1}}, & k \text{ je paran broj} \\ \frac{\binom{n_1-1}{\frac{k-1}{2}} \binom{n_2-1}{\frac{k-3}{2}} + \binom{n_1-1}{\frac{k-3}{2}} \binom{n_2-1}{\frac{k-1}{2}}}{\binom{n_1+n_2}{n_1}}, & k \text{ je neparan broj} \end{cases}$$

gde $k = 1, \dots, l$ za $l = \begin{cases} n, & n \text{ je paran broj} \\ n - 1, & n \text{ je neparan broj} \end{cases}$. Izračunavanjem se dobija (Siegel, 1956, str. 56)

$$E(K) = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad \text{i} \quad D(K) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

pa na osnovu centralne granične teoreme za uzorak dovoljno velikog obima važi

$$K \sim \mathcal{N} \left(\frac{2n_1n_2}{n_1 + n_2} + 1, \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \right)$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

tj. važi

$$K^* = \frac{K - \left(\frac{2n_1n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}} \sim \mathcal{N}(0,1)$$

Dakle kritična oblast za statistiku K^* se određuje iz tablica za normiranu normalnu raspodelu, a iz uslova

$$P_{H_0}(|K^*| \geq c) = \alpha,$$

i ima oblik $W = (-\infty, -c] \cup [c, +\infty)$.

Ako se za uzorak X_1, X_2, \dots, X_n realizovana vrednost test-statistike K^* nalazi unutar kritične oblasti W , nulta hipoteza se odbacuje.

Test serijske korelacije

Ovim testom se proverava da li postoji statistički značajna korelacija među elementima datog uzorka.

Definicija 2.11. Koeficijent serijske korelacije sa korakom k je

$$r_k = \frac{\frac{1}{n-k} \sum_{j=1}^{n-k} X_j X_{j+1} - A_k B_k}{\sqrt{\frac{1}{n-k} \sum_{j=1}^{n-k} (X_j - A_k)^2 \cdot \frac{1}{n-k} \sum_{j=1}^{n-k} (X_{j+1} - B_k)^2}}$$

gde su A_k, B_k dati sa

$$A_k = \frac{1}{n-k} \sum_{j=1}^{n-k} X_j \quad \text{i} \quad B_k = \frac{1}{n-k} \sum_{j=1}^{n-k} X_{j+1}.$$

Za velike serije (nizove, uzorke) i malo k može se koristiti sledeći izraz:

$$r'_k = \frac{\frac{1}{n-k} \sum_{j=1}^{n-k} (X_j - \bar{X}_n)(X_{j+1} - \bar{X}_n)}{\sqrt{\frac{1}{n-k} \sum_{j=1}^{n-k} (X_j - \bar{X}_n)^2 \cdot \frac{1}{n-k} \sum_{j=1}^{n-k} (X_{j+1} - \bar{X}_n)^2}}$$

kao i jednostavniji od njega

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$r'_k = \frac{\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)(X_{j+1} - \bar{X}_n)}{\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

gde je $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ (Turković, 2021, str. 30).

Pokazuje se da je

$$E(r_1) \approx -\frac{1}{n} \quad \text{i} \quad D(r_1) \approx \frac{n^3 - 2n^2 + 2}{n^2(n^2 - 1)}$$

i to se može koristiti pri testiranju hipoteze slučajnosti. Za velike vrednosti n pri hipotezi slučajnosti raspodela veličine r_1 je asimptotski $\mathcal{N}(0,1)$ raspodela, a koeficijent korelacije r_k ima približno normalnu raspodelu $\mathcal{N}\left(\frac{-1}{n-k-1}, \frac{n-k-2}{(n-k-1)^2}\right)$. Dakle, za veliko n i malo k , raspodela slučajne veličine r_k je približno normalna raspodela $\mathcal{N}\left(0, \frac{1}{n}\right)$ (Jevremović & Mališić, 2008, str. 187-189).

Kritična oblast je oblika $W = (-\infty, -c] \cup [c, +\infty)$, a njene granice se određuju iz uslova

$$P_{H_0}(|r_k| \geq c) = \alpha.$$

DIEHARD testovi

DIEHARD predstavlja skup od osamnaest testova slučajnosti razvijenih od strane Marsalje (1995). Većina ovih testova može se primeniti nad brojevima iz intervala $(0, 2^{31} - 1)$, tj. testiraju hipotezu da je dobijeni uzorak zaista uzorak ravnomerno raspodeljenih prirodnih brojeva u tom intervalu. Ovi testovi, umesto p-vrednosti rezultiraju vrednošću funkcije raspodele test-statistike za realizovanu vrednost te statistike kao vrednost argumenta. Neki testovi daju jednu takvu vrednost, dok drugi testovi zapravo izvode test mnogo puta i porede raspodelu test-statistike sa teorijskom raspodelom test-statistike pri tačnoj nultoj hipotezi; odnosno, neki testovi su tzv. testovi drugog reda. U nastavku su ukratko opisani neki od tih testova, a kompletan opis svih testova tog skupa se može naći kod Džentla (2002, str. 80-83).

Test razmaka između rođendana (birthday spacing test) je statistički test koji se koristi za analizu raspodele razmaka između „rođendana“ u nizu podataka. Ideja ovog testa potiče iz poznatog paradoksa rođendana, gde je često iznenađujuće koliko ljudi treba da se okupi kako bi postojala relativno visoka verovatnoća da bar dvoje ljudi ima rođendan istog dana. U ovom testu, uzimaju se brojevi iz niza i analiziraju se razmaci između uzastopnih u uzorku. Cilj je da se proceni da li su ovi razmaci raspodeljeni na način koji bi bio očekivan u slučaju rasporeda slučajnih brojeva. Ako se razmaci ponašaju više kao da su slučajno raspoređeni, to može ukazivati na određeni stepen nasumičnosti u nizu.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Test preklapajućih permutacija dužine 5 (overlapping 5-permutation test), ili OPERM5 test ima sledeći postupak testiranja. Posmatra se niz od milion 32-bitnih slučajnih celih brojeva. Svaki skup od pet uzastopnih brojeva može biti u jednom od 120, tj. $5!$ rasporeda. Kako se posmatraju hiljade prelazaka iz jednog rasporeda u drugi, formiraju se kumulativni zbrojevi brojeva pojavljivanja svakog rasporeda. Zatim, kvadratna forma slabog inverza (Fountain, 2002) kovarijacione matrice reda 120×120 daje test ekvivalentan testu odnosa verodostojnosti (Nikulín, 2020) da 120 brojeva dolazi iz određene (asimptotski) normalne raspodele sa izvesnom 120×120 kovarijacionom matricom (sa rangom 99). Ova verzija koristi 1000000 brojeva, dva puta. Moguće je da ovaj test ima nerešene greške u dostupnim softverima i ponuđenim kodovima koje dovode do slabih performansi.

Test retkih pojavljivanja parova (overlapping-pairs-sparse-occupancy test) bitova se primenjuje tako što se parovi bitova formiraju preklapanjem susednih bitova u binarnom nizu. Fokus je na popunjenosti retkih konfiguracija ovih parova. Konkretni detalji testa uključuju formiranje parova bitova uzimajući u obzir određene preklapajuće pozicije bitova unutar binarnog niza. Zatim se test bavi raspodelom popunjenosti ovih parova, sa naglaskom na retkim konfiguracijama. Cilj je proceniti da li se posmatrana raspodela značajno razlikuje od onoga što bi se očekivalo u zaista slučajnom nizu.

Test minimalnih rastojanja (minimum distance test) sprovodi se analiziranjem minimalnih rastojanja između parova tačaka u nizu brojeva. Cilj je da se proceni da li su ova minimalna rastojanja saglasna sa raspodelom koja se dobija u slučaju prostog slučajnog uzorka. Ako je raspodela minimalnih udaljenosti slična raspodeli koja bi se dobila u slučaju potpuno nasumičnih tačaka, to ukazuje da je generisani niz praktično prost slučajan uzorak.

DIEHARD testovi su često primenjivani jer su za sve njih dostupna tehnička rešenja, pa pri proveru slučajnosti brojeva, istraživač odmah može primeniti više testova slučajnosti i time uštediti vreme koje bi potrošio samostalnim programiranjem drugih testova. Ipak, ove testove prati puno ograničenja od kojih su neka komplikovani postupci primene, dugotrajno testiranje kod velikih uzoraka, niska vrednost moći i postupci slični postojećim. Stoga je preporučljivo koristiti i prethodno navedene testove koji su moćni i, osim u slučaju testa serijske korelacije, veoma brzi.

2.4. Analiza performansi i poređenja

U nastavku biće izvršeno generisanje nizova pseudoslučajnih brojeva različitih obima korišćenjem LKG-a sa aditivnim proširenjem ilustrovanog u odeljku 2.2 i Mersenovog tvister (MT19937) generatora kao najčešće korišćenog. Generator MT19937 je u dosadašnjim studijama pokazao najbolje performanse (Matsumoto & Nishikura, 1998). Štaviše, autori konstatuju da se i Mersenov Tvister i drugi poznati generatori kao što je Marsaljin (2003) „xorshift“ generator zapravo svode na generatore sa modularnom redukcijom (Panneton & L'Ecuyer, 2005; Brent, 2004).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Dodatno, za dobijene nizove brojeva izračunate su i tabelirane p-vrednosti testova tačaka rasta i tačaka zaokreta radi ispitivanja slučajnosti generisanih brojeva. Takođe, za LKG sa proširenjem odabrani su parametri koji zadovoljavaju uslove teoreme 2.3 tako da period bude duži od traženog obima uzorka. Naknadno je data i diskusija poređenja brzine ovih generatora. P-vrednosti izračunate su za 10 000 simulacija nizova brojeva generisanih navedenim generatorima. Simulacije su programirane u MATLABu.

Tabela 2.1. Prosečne p-vrednosti testova slučajnosti za nizove brojeva generisane LKG-om sa proširenjem i MT19937.¹²

Generator	Test	<i>n</i>				
		10	100	1000	10000	100000
LKG sa proširenjem	Test tačaka rasta	0,5060	0,5057	0,5072	0,5011	0,5103
	Test tačaka zaokreta	0,4948	0,4941	0,4972	0,4906	0,4844
Mersenov Tvister	Test tačaka rasta	0,4614	0,4987	0,5000	0,5032	0,5060
	Test tačaka zaokreta	0,4786	0,4954	0,4956	0,5021	0,4945

Nakon izvršenih simulacija dobijeni rezultati su prikazani tabelom 2.1. Ti rezultati ukazuju da i LKG sa proširenjem i Mersenov Tvister generišu uzorke brojeva koji zadovoljavaju hipotezu slučajnosti pri testiranju. Štaviše, pokazuje se da je kvalitet brojeva generisanih LKGom sa proširenjem, u smislu slučajnosti, isti ili bolji u odnosu na MT19937.

U smislu brzine za LKG sa aditivnim proširenjem performanse su slične za $n < 10000$, dok za veće obime uzorka, MT19937 ima blagu prednost jer LKG sa proširenjem mora prolaziti i porediti zadnji generisani element sa ostalima. Ta se teškoća može lako prevazići uz bolji način programiranja proširenja. Naime, ako se svakom celom broju od 0 do $m - 1$ dodeli proširenje 1 koje se povećava za 1 nakon svakog novog generisanja tog broja, onda se brzina LKG-a sa proširenjem drastično povećava i nema primetnih razlika u brzini generisanja u odnosu na ostale najčešće korišćene generatore.

¹² U oba slučaja, testiranje saglasnosti raspodele simuliranih p-vrednosti sa uniformnom raspodelom na intervalu $[0, 1]$ (bilo kojim testom) rezultira prihvatanjem hipoteze saglasnosti.

3. Modeliranje slučajnih veličina

U prethodnom poglavlju je prikazano na koje načine se mogu generisati pseudoslučajni brojevi, a zatim i testirati hipoteza slučajnosti, tj. hipoteza da je dobijeni uzorak prost slučajan uzorak. U slučaju prihvatanja nulte hipoteze može se zaključiti i da je dobijeni uzorak brojeva izvučen iz obeležja koje je raspodeljeno po uniformnoj raspodeli. Međutim, često je potrebno modelirati slučajnu veličinu čija raspodela nije uniformna (nekad su čak potrebne permutacije, kombinacije ili varijacije neke grupe brojeva). Štaviše, pokazaće se da većina metoda modeliranja slučajnih veličina zahteva poznavanje jedne ili više realizacija slučajne veličine sa uniformnom $\mathcal{U}(0, 1)$ raspodelom. Dakle, za potrebe daljeg izlaganja, može se smatrati da je dat prost slučajan uzorak U_1, U_2, \dots, U_n potrebnog obima $n \in \mathbb{N}$ uniformno raspodeljen na intervalu $(0, 1)$.

Ovo poglavlje je veoma bitno za rad jer je modeliranje slučajnih veličina od suštinske važnosti za empirijske verifikacije rezultata raznih statističkih metoda, pa i onih koje su rezultat istraživanja autora (kontrolne karte i testovi normalnosti). Iako mnogi softveri, a i programski jezici nude gotove funkcije (kodove) za direktno modeliranje uzoraka željenih obima, to nije slučaj za sve raspodele. Štaviše, za većinu raspodela korišćenih u ovom istraživanju programe je napisao autor. Iako nisu korišćeni svi algoritmi koji su izloženi, mnogi jesu, a zbog metodološke potpunosti teksta i eventualne motivacije za dalja istraživanja, bilo ih je potrebno prikazati.

U nastavku će biti razmotrene metode za modeliranje slučajnih veličina u opštem slučaju i navedeni posebni postupci za modeliranje slučajnih veličina (sa raspodelama) koje su korišćene u rezultatima ovog rada. U odeljku 3.3 ovog poglavlja, data je i nova metoda modeliranja normalne raspodele koja je rezultat istraživanja autora, i ukazano je na neke njene prednosti i nedostatke.

Pored navedenih metoda dostupne su i u raznim slučajevima se koriste i metoda stohastičke kolokacije (Ivanović, 2018), butstrep¹³ metoda, metoda testa permutacija (Carsey & Harden, 2014) i druge.

3.1. Metode modeliranja diskretnih slučajnih veličina

Iako modeliranje diskretnih slučajnih veličina neće biti konkretno primenjeno u ovom istraživanju, već je to slučaj samo sa modeliranjem neprekidnih slučajnih veličina, upravo se

¹³ Bootstrap. Novi uzorak se reuzorkuje na osnovu jednog postojećeg tako što se kao simulacija, tj. modelirani uzorak uzima varijacija sa ponavljanjem elemenata polaznog uzorka gde se svaki element varijacije dobija slučajnim izborom jednog elementa tog uzorka (Carsey & Harden, 2014).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

modeliranje neprekidnih slučajnih veličina može vršiti odgovarajućom diskretizacijom, tj. modeliranjem prikladno izabrane diskretne slučajne veličine. Naime, izborom dovoljno guste podele intervala u kojem se mogu naći modaliteti neprekidne slučajne veličine, i modeliranjem diskretne slučajne veličine čiji su modaliteti čvorovi te podele, sa dovoljnom preciznošću se može modelirati i neprekidna slučajna veličina.

Definicija 3.1. *Diskretna slučajna veličina je svaka slučajna veličina sa konačno ili prebrojivo mnogo vrednosti (modaliteta). Za one sa prebrojivo mnogo vrednosti kaže se i da su elementarne slučajne veličine.*

Zakon raspodele je pravilo po kojem se svakom modalitetu slučajne veličine dodeli verovatnoća tako da zbir verovatnoća svih modaliteta bude 1.

Neka je slučajna veličina X diskretna sa konačno mnogo modaliteta x_1, x_2, \dots, x_n , $n \in \mathbb{N}$ i neka tim modalitetima u zakonu raspodele redom odgovaraju verovatnoće p_1, p_2, \dots, p_n . Tada se za zapis zakona raspodele slučajne veličine X može koristiti

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \quad \sum_{i=1}^n p_i = 1 \quad (3.1)$$

ili $P(X = x_k) = p_k, k = 1, \dots, n$.

U slučaju da je X elementarna slučajna veličina sa modalitetima $x_1, x_2, \dots, x_k, \dots, k \in \mathbb{N}$ kojima u zakonu raspodele redom odgovaraju verovatnoće $p_1, p_2, \dots, p_k, \dots$ onda je zapis tog zakona raspodele

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}, \quad \sum_{k=1}^{\infty} p_k = 1 \quad (3.2)$$

ili $P(X = x_k) = p_k, k \in \mathbb{N}$.

U oba slučaja važi $x_k < x_{k+1}$, za svako dostupno k .

Elementarne slučajne veličine se mogu aproksimirati, sa unapred zadatom tačnošću, diskretnim slučajnim veličinama (sa konačno mnogo vrednosti).

Definicija 3.2. *Neka je slučajna veličina X data sa (3.2). Zasečena slučajna veličina slučajne veličine X , za proizvoljno mali broj $\delta > 0$, je slučajna veličina X^* čiji je zakon raspodele dat sa:*

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$X^*: \begin{pmatrix} x_1 & x_2 & \dots & x_{n_0-1} & x_{n_0} \\ p_1 & p_2 & \dots & p_{n_0-1} & p_{n_0}^* \end{pmatrix}, \quad (3.3)$$

gde je $n_0 \in \mathbb{N}$ određeno sa $\sum_{i=n_0+1}^{+\infty} p_i < \delta$, a $p_{n_0}^* = 1 - \sum_{i=1}^{n_0-1} p_i$.

Zbog beskonačnog broja modaliteta, modeliranje slučajnih veličina oblika datog sa (3.2) nije moguće čak ni računom, ali su u većini slučajeva verovatnoće tih modaliteta u zakonu raspodele nakon nekog $n_0 \in \mathbb{N}$ zanemarljivo male. Tada se za potrebe modeliranja slučajne veličine date sa (3.2) formira njoj zasečena slučajna veličina data sa (3.3) za izabrano n_0 .

Metoda inverzne funkcije

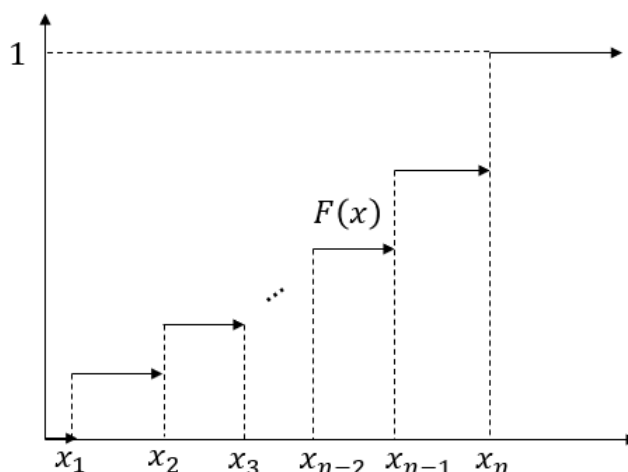
Definicija 3.3. Funkcija raspodele slučajne veličine X je funkcija $F_X: \mathbb{R} \rightarrow (0, 1)$ definisana sa:

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Vidi se da je analitički zapis funkcije F_X za slučajne veličine date sa (3.1) i (3.2)

$$F_X(x) = P(X \leq x) = \sum_{i=1}^k p_i, \quad x \in [x_k, x_{k+1})$$

gde u slučaju (3.1) $k \in \{1, \dots, n\}$, a u slučaju (3.2) $k \in \mathbb{N}$. Videti sliku 3.1.



Slika 3.1. Funkcija raspodele diskretne slučajne veličine.

Algoritam 3.1. Postupak modeliranja slučajne veličine X , zadate sa (3.1), je sledeći.

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Neka je x_k iz zakona raspodele slučajne veličine X za koje važi

$$F_X(x_{k-1}) < u \leq F_X(x_k)$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

(jer je F_X neprekidna samo sa desne strane).

3. Tada je x_k realizovana (modelirana) vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Metoda prihvatanja i odbacivanja

Neka su date slučajne veličine

$$X_\alpha: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_{\alpha 1} & p_{\alpha 2} & \dots & p_{\alpha n} \end{pmatrix}, \quad \alpha \in \mathcal{A},$$

gde je $\mathcal{A} \sim (0, 1)$ (\mathcal{A} ekvipotentan sa $(0, 1)$) jer \mathcal{A} ima isti kardinalni broj kao i skup slučajnih veličina čije su moguće vrednosti x_1, x_2, \dots, x_n sa različitim raspodelama. Raspodele se razlikuju shodno varijaciji verovatnoća koje mogu biti u njenom zakonu. Dakle, $|\mathcal{A}| = |(0, 1)| = c$ (c -kontinuum). Ipak, kako je tehnički izvodljivo raditi samo sa brojevima konačnog broja decimala, broj decimala verovatnoća $p_{\alpha 1}, p_{\alpha 2}, \dots, p_{\alpha n}, \alpha \in \mathcal{A}$ se ograničava na neki konačan broj. Drugim rečima, razmatraju se slučajne veličine $X_s, s \in \mathbb{N}$ čije verovatnoće imaju konačan (izabran) broj decimala.

Neka je l najveći broj ne-nula decimala u decimalnom zapisu verovatnoća iz zakona raspodele slučajne veličine X . Tada su sve verovatnoće oblika

$$p_{\alpha j} = 0, \varepsilon_1 \varepsilon_2 \dots \varepsilon_l, \quad j = 1, \dots, n; \alpha \in \mathcal{A}, \varepsilon_i \in \{0, 1, \dots, 9\}, i = 1, \dots, l. \quad (3.4)$$

Skup svih slučajnih veličina čiji zakoni raspodele imaju samo verovatnoće oblika (3.4) je konačan, tj. takvih slučajnih veličina ima konačno mnogo (Avdović, 2019, str. 38-39). Neka ih je $r, r \in \mathbb{N}, r < +\infty$. Tada, na osnovu svih pretpostavki za X datu sa (3.1) postoji jedinstveno $s \in \{1, \dots, r\}$ tako da je

$$X = X_s: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_{s1} & p_{s2} & \dots & p_{sn} \end{pmatrix}.$$

Verovatnoće iz njenog zakona raspodele se mogu zapisati kao

$$P(X_s = x_j) = p_{sk} = \frac{a_{sk}}{a_{s1} + a_{s2} + \dots + a_{sn}}, \quad a_{sk} \in \mathbb{R}; k = 1, \dots, n.$$

Primer takve reprezentacije je $a_{sk} = p_{sk}; k = 1, \dots, n$, ili bilo koji izbor brojeva a_{sk} tako da je $a_{sk}: (a_{s1} + a_{s2} + \dots + a_{sn}) = p_{sk}: 1$.

Neka je $a^* = \max\{a_{sj}, j = 1, \dots, n\}$. Konačno, neka su dobijene vrednosti

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$P_{sj} = \frac{a_{sj}}{a^*}, \quad j = 1, \dots, n.$$

Predstojeći algoritam predstavlja prilagođenu verziju algoritma koji je originalno predložen (Rajasekaran & Ross, 1993) za modeliranje slučajnih veličina čija se raspodela menja pri protoku vremena (indeks slučajne veličine). Da bi se oznake u originalnom, a i u ovom prilagođenom algoritmu bile razumljive, potrebno je odgovarajuće teorijsko rasuđivanje koje je dato prethodnim pasusima. Ova modifikacija algoritma se takođe može smatrati doprinosom autora. Ipak, ideja potiče iz kratkog teksta koji je Džentl dao u svojoj knjizi (Gentle, 2002, str. 119-120).

Algoritam 3.2. Postupak modeliranja slučajne veličine X zadate sa (3.1) je sledeći.

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $d = [nu]$ (funkcija $[\cdot]$ predstavlja gornju (veću) celobrojnu vrednost realnog broja).
3. Izračunati $t = d - nu$.
4. Proveriti da li važi $t \leq P_{sd}$.
 - a. Ako važi, uzeti x_d kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, ponoviti postupak sa drugim brojem iz $\mathcal{U}(0, 1)$ raspodele.

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Metoda mere uniformnosti

Stav 3.1. Neka slučajna veličina (U, V) ima uniformnu raspodelu nad pravougaonikom $\{(u, v): 0 \leq u \leq 1; -1 \leq v \leq 1\}$. Ako je $X_0 = s\frac{U}{V} + a, s > 0$, gde se $s, a \in \mathbb{R}$ biraju tako da važi $g_{X_0}(x_k) \geq P(X = x_k)$, tada važe sledeća tvrđenja.

Gustina raspodele slučajne veličine X_0 je

$$g_{X_0}(x) = \begin{cases} \frac{1}{4s}, & a - s \leq x \leq a + s \\ \frac{1}{4(x-a)^2}, & -\infty < x < a - s, a + s < x < +\infty \end{cases}.$$

Gustina uslovne raspodele $Y|X_0$ (za datu vrednost $X_0 = x$), gde je $Y = U^2$ data je sa

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$g_{Y|X_0}(y|x) = \begin{cases} 1, & a - s \leq x \leq a + s, 0 \leq y \leq 1 \\ \frac{(x - a)^2}{s^2}, & x > |a + s|, 0 \leq y \leq \frac{s^2}{(x - a)^2} \\ 0, & y > \max\left\{1, \frac{s^2}{(x - a)^2}\right\} \end{cases}$$

Uslovna raspodela slučajne veličine Y pri uslovu $X_0 = x$ je uniformna na intervalu $(0, 4sg_{X_0}(x))$

(Stadlober, 1990, 1991; Gentle, 2002, str. 130-131.; Avdović, 2019, str. 40).

Stadlober (1990) je predložio sledeći algoritam za modeliranje diskretnih slučajnih veličina, koji je zasnovan na izloženim činjenicama.

Algoritam 3.3. Postupak modeliranja slučajne veličine X zadate sa (3.1) je sledeći.

1. Neka su u i v brojevi, međusobno nezavisno generisani iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $x = \left\lceil a + \frac{s(2v-1)}{u} \right\rceil$.
3. Izračunati $y = u^2$.
4. Proveriti da li važi $y \leq P(X = x_k)$, za neko $k = 1, \dots, n$ (iz zakona raspodele za X).
 - a. Ako važi, uzeti x_k kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, ponoviti postupak sa drugim brojem iz $\mathcal{U}(0, 1)$ raspodele.

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

3.2. Metode modeliranja apsolutno neprekidnih slučajnih veličina

Definicija 3.4. *Apsolutno neprekidna slučajna veličina je slučajna veličina X čija je funkcija raspodele funkcija $F_X(x)$, za koju postoji integrabilna funkcija $g_X(x)$, $x \in \mathbb{R}$, takva da za svako $x \in \mathbb{R}$ važi:*

1. $g_X(x) \geq 0$,
2. $F_X(x) = \int_{-\infty}^x g_X(t) dt$.

Funkcija $g_X(x)$ je **gustina raspodele** slučajne veličine X .

Definicija 3.5. *Neka je $g_X(x) \neq 0$, $x \in (a, b)$ (može biti $(a, b) = \mathbb{R}$). **Zasečena slučajna veličina** apsolutno neprekidne slučajne veličine X , za proizvoljno mali broj $\delta > 0$, je slučajna veličina X^* sa gustinom raspodele*

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$g_{X^*}(x) = \begin{cases} g_X(x) + \frac{\delta}{b' - a'}, & x \in (a', b'), \\ 0, & x \notin (a', b') \end{cases}$$

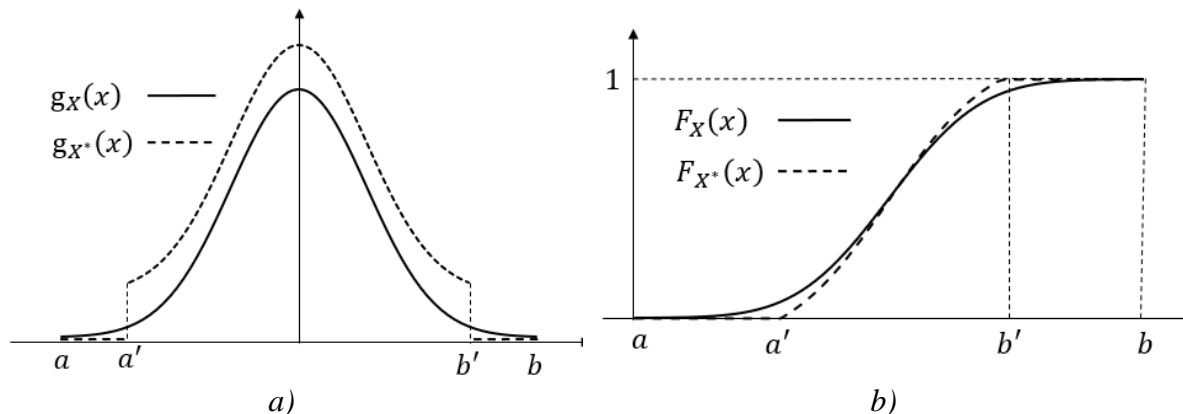
gde za $a' > a$ i $b' < b$ važi

$$1 - \int_{a'}^{b'} g_X(t) dt = \delta. \quad (3.5)$$

Lako je izračunati da je

$$F_{X^*}(x) = \begin{cases} 0, & x \in (a, a') \\ F_X(x) - F_X(a') + \frac{\delta(x - a')}{b' - a'}, & x \in (a', b') \\ 1, & x \in (b', b) \end{cases}$$

Ako su a i b konačni brojevi podrazumeva se da je gustina $g_X(x)$ jednaka nuli van intervala (a, b) .



Slika 3.2. Poređenje funkcionalnih karakteristika slučajne veličine X i njoj zasečene slučajne veličine X^* ; a) Gustina raspodele; b) Funkcija raspodele.

Modeliranje zasečene slučajne veličine X^* slučajne veličine X , zapravo i nije modeliranje tražene slučajne veličine X , ali se u definiciji 3.5 δ bira tako da razlika u raspodelama bude zanemarljivo mala. Ovo modeliranje može biti važno u slučajevima kada je direktno modeliranje X veoma komplikovano ili čak nemoguće.

Napomena 3.1. Zasečena slučajna veličina se može definisati množenjem gustine polazne slučajne veličine odgovarajućom konstantom (Johnson, Kotz & Balakrishnan, 1994, str. 156-157). Ipak, iako neznatno, ovakva definicija uzrokuje promenu oblika slučajne veličine. Za oznake već korišćene u ovom delu rada, gustina i funkcija raspodele ovako definisane zasečene slučajne veličine su

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$g_{X^*}(x) = \begin{cases} \frac{1}{F(b') - F(a')} g_X(x), & x \in (a', b') \\ 0, & x \notin (a', b') \end{cases},$$

$$F_{X^*}(x) = \begin{cases} 0, & x \in (a, a') \\ \frac{F_X(x) - F_X(a')}{F(b') - F(a')}, & x \in (a', b') \\ 1, & x \in (b', b) \end{cases}.$$

Metoda diskretizacije

Neka za slučajnu apsolutno neprekidnu slučajnu veličinu X važi da je $g_X(x) \neq 0$ za $x \in (a, b)$. Tada je i $F_X(x) \neq 0$ za $x \in (a, b)$. Ako za slučajnu veličinu X iz definicije 3.4 izaberemo niz modaliteta $x_1, x_2, \dots, x_n \in (a, b)$ tada se slučajna veličina X može približno modelirati modeliranjem slučajne veličine

$$Y: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

gde su $p_1 = P(Y = x_1) = F_X(x_1)$, $p_i = P(Y = x_i) = F_X(x_i) - F_X(x_{i-1})$, za $i = 2, \dots, n-1$ i $p_n = P(Y = x_n) = 1 - F(x_{n-1})$.

Da bi se obezbedila što veća preciznost ove metode potrebno je da modaliteti budu raspodeljeni tako da broj različitih modaliteta slučajne veličine Y unutar nekog intervala $(c, d) \subset (a, b)$, u odnosu na n , bude proporcionalan verovatnoći $P(X \in (c, d))$. Takođe, najbolje je da prvi i zadnji modalitet slučajne veličine Y budu izabrani kao $x_1 = a$ i $x_n = b$. Povećanjem broja modaliteta slučajne veličine Y povećava se i preciznost metode, a ukoliko je preciznost određena nekom tačnošću $\varepsilon > 0$ onda se broj modaliteta bira tako da je

$$\sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)| < \varepsilon.$$

Detaljno opisan algoritam metode diskretizacije se ne može naći u (pretraženoj) literaturi, pa se formalan opis metode diskretizacije takođe može smatrati doprinosom autora.

Metoda inverzne funkcije

Neka za funkciju raspodele $F_X(x)$ apsolutno neprekidne slučajne veličine X postoji inverzna funkcija $F_X^{-1}(u)$, za svako $u \in (0, 1)$. Metoda inverzne funkcije se svodi na rešavanje jednačine $F_X(x) = u$ po nepoznatoj x , a u je slučajan broj (generisan iz $\mathcal{U}(0, 1)$ raspodele) koji se koristi za modeliranje.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 3.2. Neka je data slučajna veličina $U: \mathcal{U}(0, 1)$. Tada slučajna veličina $F_X^{-1}(U)$ ima raspodelu slučajne veličine X (tj. funkciju raspodele $F_X(x)$).

Najjednostavniji (drugi algoritmi koji koriste inverznu funkciju podrazumevaju kombinacije sa drugim metodama) način modeliranja metodom inverzne funkcije je dat sledećim algoritmom. Budući da je izuzetno jednostavan, a istovremeno i dovoljno precizan za slučajne veličine čiji je analitički izraz inverzne funkcije raspodele poznat, ovaj način modeliranja neprekidnih slučajnih veličina je ujedno i najprimenjivaniji.

Za slučajne veličine za koje nije moguće odrediti eksplicitni izraz inverzne funkcije raspodele ili za one za koje je teško ili dugotrajno izračunati vrednosti inverzne funkcije, metodu inverzne funkcije je moguće koristiti za brojeve iz statističkih tablica. Ako se vrednost inverzne funkcije za željeni argument ne nalazi u tablici, ona se može aproksimirati metodom polovljenja intervala sa vrednostima inverzne funkcije za dva argumenta iz tablice koji su najbliži željenom argumentu.

Algoritam 3.4. Slučajna veličina X (iz definicije 3.4) može se modelirati na sledeći način:

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $x = F_X^{-1}(u)$.
3. Uzeti x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Sledeća definicija daje tip slučajnih veličina koje se mogu primeniti pri modeliranju realnih pojava. Biće izložen i postupak njihovog modeliranja.

Definicija 3.6. Neka su date slučajne veličine X_1, \dots, X_n i brojevi p_1, \dots, p_n takvi da važi $p_j \in (0, 1), j = 1, \dots, n$, i $\sum_{i=1}^n p_i = 1$. **Mešavina slučajnih veličina** je slučajna veličina X za koju važi

$$P(X = X_1) = p_1, \dots, P(X = X_n) = p_n.$$

Važi $F_X(x) = \sum_{i=1}^n p_i F_{X_i}(x)$, gde su $F_{X_j}(x)$ funkcije raspodele slučajnih veličina X_j , $j = 1, \dots, n$.

Neka je za funkcije raspodele $F_{X_j}(x), j = 1, \dots, n$, moguće odrediti inverznu funkciju $F_{X_j}^{-1}(u), u \in (0, 1)$, svakog od njih.

Algoritam 3.5. Modeliranje slučajne veličine X iz definicije 3.6 metodom inverzne funkcije se može vršiti na sledeći način.

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

2. Odrediti $k \in \{1, \dots, n-1\}$ takvo da važi $\sum_{i=1}^k p_i \leq u < \sum_{i=1}^{k+1} p_i$. Ako je $k = n$ treba da važi $\sum_{i=1}^k p_i \leq u \leq 1$.
3. Izračunati $x = F_{X_k}^{-1}(u)$.
4. Uzeti x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Metoda prihvatanja i odbacivanja

Primena ove metode modeliranja apsolutno neprekidne slučajne veličine X sastoji se u tome da je dat uzorak vrednosti (brojeva) odgovarajućeg obima neke slučajne veličine Y tako da vrednosti gustina za slučajne veličine X i Y budu približne za iste vrednosti argumenta. Algoritmi ove metode se uglavnom svode na selekciju poduzorka uzorka Y takvog da njegova raspodela bude statistički saglasna sa raspedelom obeležja X . Veći broj odbačenih vrednosti ukazuje na manju preciznost metode ili loš izbor slučajne veličine Y .

Definicija 3.7. Neka je (a, b) interval na kom su vrednosti gustina raspodele $g_X(x)$ i $g_Y(x)$, slučajnih veličina X i Y , redom, različite od nule. Ako postoji realan broj $C \geq 1$ takav da važi $C g_Y(x) \geq g_X(x)$ ($\forall x \in (a, b)$), **majorantna funkcija** ili **obvojnica** slučajne veličine X je funkcija $C g_Y(x)$. **Majorantna gustina** slučajne veličine X je gustina raspodele $g_Y(x)$ slučajne veličine Y .

Broj $\frac{1}{C}$ je **mera efikasnosti** postupka određenog slučajnom veličinom Y . Što je $\frac{1}{C}$ bliže 1 to je procedura efikasnija.

Stav 3.3. Ako za vrednosti y i u , redom iz raspodele slučajne veličine Y i $\mathcal{U}(0, 1)$, važi $u \leq \frac{g_X(y)}{C g_Y(y)}$, tada je y realizacija slučajne veličine X .

Algoritam 3.6. Slučajna veličina X (iz definicije 3.4) se može modelirati na sledeći način.

1. Neka je y broj generisan iz raspodele slučajne veličine Y .
2. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
3. Proveriti da li važi $u \leq \frac{g_X(y)}{C g_Y(y)}$.
 - a. Ako važi, uzeti y kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, odbaciti y i ponoviti postupak sa drugim brojevima iz raspodele slučajne veličine Y i $\mathcal{U}(0, 1)$.

Postupak ponoviti sa drugim brojevima iz raspodele slučajne veličine Y i $\mathcal{U}(0, 1)$ dok se ne dobije uzorak traženog obima.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Uz modifikaciju algoritma 3.6 može se dobiti sledeći algoritam čija primena rezultira preciznijim modeliranjem posmatrane slučajne veličine.

Algoritam 3.7. Neka je moguće odrediti $F_Y^{-1}(u)$, $u \in (0, 1)$, i neka je ona diferencijabilna funkcija. Slučajna veličina X (iz definicije 3.4) se može modelirati na sledeći način.

1. Neka su u i v brojevi, međusobno nezavisno generisani iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $x = F_Y^{-1}(u)$
3. Proveriti da li važi $v \leq \frac{g_X(F_Y^{-1}(u))F_Y^{-1'}(u)}{C}$.
 - a. Ako važi, uzeti x kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, odbaciti u i v , i ponoviti postupak sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele.

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Često je nemoguće ili vrlo teško odabrati odgovarajuću slučajnu veličinu Y pa se algoritmi metode prihvatanja i odbacivanja zadaju njihovim zasnivanjem na sledećem stavu.

Stav 3.4. Neka se $g_X(x)$, $x \in (a, b)$ može predstaviti kao

$$g_X(x) = C g_Y(x) h(x),$$

gde su $C \geq 1$ realan broj, $0 \leq h(x) \leq 1$ neka realna funkcija i $g_Y(x)$, $x \in (a, b)$ gustina raspodele neke slučajne veličine Y . Tada, za slučajnu veličinu $U \sim \mathcal{U}(0, 1)$ važi

$$g_Y(x|U \leq h(Y)) = g_X(x).$$

Algoritam 3.8. Slučajna veličina X (iz definicije 3.4) se može modelirati na sledeći način.

1. Neka je y broj generisan iz raspodele slučajne veličine Y .
2. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
3. Proveriti da li važi $u \leq h(y)$.
 - a. Ako važi, uzeti y kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, odbaciti y i ponoviti postupak sa drugim brojevima iz raspodela slučajne veličine Y i $\mathcal{U}(0, 1)$.

Postupak ponoviti sa drugim brojevima iz raspodela slučajne veličine Y i $\mathcal{U}(0, 1)$ dok se ne dobije uzorak traženog obima.

Da bi sva tri navedena algoritma bila dovoljno efikasna, izbor majorantne funkcije mora biti dobar, tj. C treba da bude što bliže jedinici, a Y mora dovoljno dobro aproksimirati X (između njih mora postojati statistička saglasnost), tako da se njene vrednosti lako mogu generisati (modelirati).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Džon fon Nojman (1951) pokazuje da se za Y može uzeti odgovarajuća uniformno raspodeljena slučajna veličina.

Stav 3.5. Neka je $g_X(x) \leq M$, $x \in (a, b)$. Neka su x i y redom vrednosti slučajnih veličina sa raspodelama $\mathcal{U}(a, b)$ i $\mathcal{U}(0, M)$. Ako važi $y \leq g_X(x)$, tada je x realizovana vrednost slučajne veličine X (von Neuman, 1951).

Algoritam 3.9. Slučajna veličina X (iz definicije 3.4) se može modelirati na sledeći način.

1. Neka je x broj generisan iz $\mathcal{U}(a, b)$ raspodele.
2. Neka je y broj generisan iz $\mathcal{U}(0, M)$ raspodele.
3. Proveriti da li važi $y \leq g_X(x)$.
 - a. Ako važi, uzeti x kao realizovanu vrednost slučajne veličine X .
 - b. Ako ne važi, odbaciti x i ponoviti postupak sa drugim brojevima iz raspodela $\mathcal{U}(a, b)$ i $\mathcal{U}(0, M)$.

Postupak ponoviti sa drugim brojevima iz raspodela $\mathcal{U}(a, b)$ i $\mathcal{U}(0, M)$ dok se ne dobije uzorak traženog obima.

Metoda mere uniformnosti

Primena ove metode uslovljena je korišćenjem odgovarajuće zakonitosti koja važi za $\frac{V}{U}$, gde je dvodimenzionalna slučajna veličina (U, V) **uniformno raspodeljena** na skupu

$$S = \left\{ (u, v) : 0 \leq u \leq \sqrt{h\left(\frac{v}{u}\right)} \right\},$$

za nenegativnu integrabilnu funkciju h . Spomenuta zakonitost data je sledećim stavom.

Stav 3.6. Gustina raspodele slučajne veličine $\frac{V}{U}$ je proporcionalna funkciji h .

Praktično je bolje odrediti nadskup \bar{S} skupa S kao neku jednostavnu figuru, a zatim iskoristiti metodu prihvatanja i odbacivanja, gde se vrednosti unutar \bar{S} , a izvan S , odbacuju. Takav skup naziva se **majorantnim skupom** skupa S .

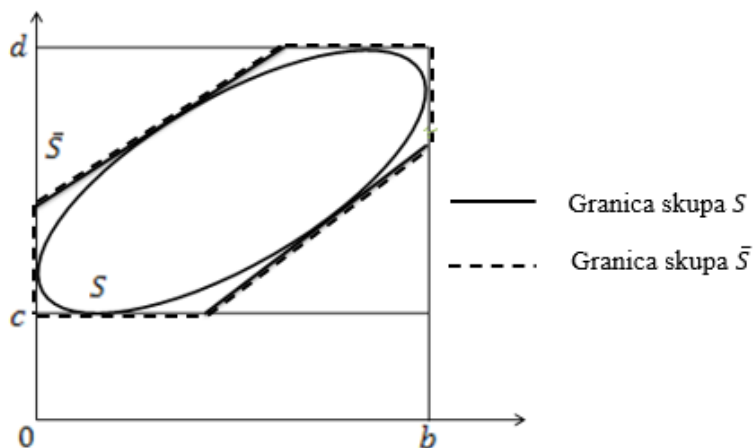
Ako su oblasti vrednosti funkcija $h(x)$ i $x^2 h(x)$ unutar oblasti S , tada je moguć izbor skupa \bar{S} dat sa

$$\bar{S} = \{(u, v) : 0 \leq u \leq b, c \leq v \leq d\}$$

gde su $b = \sup\{\sqrt{h(x)} : x \in \mathbb{R}\}$, $c = \inf\{\sqrt{h(x)} : x \in \mathbb{R}\}$, $d = \sup\{x\sqrt{h(x)} : x \in \mathbb{R}\}$ (Gentle 2002, str. 129). Za geometrijsku interpretaciju videti sliku 3.3.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Što je mera \bar{S} veća u odnosu na meru S to je opisani postupak manje efikasan. Tada je bolje izabrati drugi majorantni skup \bar{S}' , površine približne površini skupa S , kao što je ilustrovano na slici 3.3 (Avdović, 2019, str. 49).



Slika 3.3. Geometrijska ilustracija majorantnog skupa.

Algoritam 3.10. Slučajna veličina X (iz definicije 3.4) se može modelirati na sledeći način.

1. Neka su u i v brojevi, međusobno nezavisno generisani iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $u' = bu$ i $v' = c + (d - c)v$.
3. Izračunati $x = \frac{v'}{u'}$.
4. Proveriti da li važi $u'^2 \leq h(x)$.
 - a. Ako važi, uzeti x kao realizovanu vrednost slučajne veličine $Y = \frac{v}{u}$.
 - b. Ako ne važi, odbaciti x i ponoviti postupak sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele.

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

3.3. Neke važne raspodele i njihovo modeliranje

Posebni slučajevi raspodela verovatnoća koji su bili od važnosti za ovo istraživanje izloženi su u ovoj sekciji uz definicije, bitne osobine, veze sa drugim raspodelama i primere primena tih raspodela. Naime, za potrebe empirijske verifikacije rezultata ovog istraživanja biće potrebno modelirati razne raspodele. Prevažodno, najvažnije je modeliranje normalne raspodele, ali se i modeliranje ostalih raspodela izloženih u ovom poglavlju pokazalo važnim. Konkretna primena sastoji se u korišćenju modeliranih vrednosti u simulacijama raspodele odgovarajućih statistika, moći testova zasnovanih na tim statistikama ili određivanje prosečnog broja kontrolisanih uzoraka pre prvog detektovanog defekta kod kontrole kvaliteta. Dodatno, deo rezultata ovog istraživanja jeste novi algoritam modeliranja normalne raspodele koji će biti

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

izložen i čije će performanse biti analizirane i upoređivane sa performansama drugih često primenjivanih algoritama.

Modeliranje indikatora događaja – Bernulijeve¹⁴ raspodele

Neka je dat slučajni događaj A i neka je verovatnoća njegove realizacije jednaka p .

Definicija 3.8. *Indikator događaja A je slučajna veličina I_A sa zakonom raspodele*

$$I_A: \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad p = P(A).$$

Bernulijeva raspodela je raspodela indikatora događaja.

Važi $E(I_A) = p$, $D(I_A) = p(1-p)$.

Algoritam 3.11. Modeliranje indikatora događaja A , $I_A: \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, može se ostvariti metodom inverzne funkcije na sledeći način:

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Proveriti da li važi $u < 1-p$.
 - a. Ako važi, uzeti 0 kao realizovanu vrednost slučajne veličine I_A .
 - b. Ako ne važi, uzeti 1 kao realizovanu vrednost slučajne veličine I_A .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Najpoznatija primena Bernulijeve raspodele je u definisanju i opisivanju binomne raspodele (Jevremović, 2014, str. 98-99, 101).

Modeliranje beta raspodele

Definicija 3.9. *Beta raspodela (dvoparametarska), sa parametrima $a, a > 0$, i $b, b > 0$ (parametri oblika), je raspodela slučajne veličine X čija je gustina*

$$g(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad 0 \leq x \leq 1,$$

¹⁴ Jacob Bernoulli (1654-1705), švajcarski matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

gde je $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$, $\alpha > 0, \beta > 0$, beta funkcija, za koju važi $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Koristi se oznaka $X \sim B(a, b)$.

Funkcija raspodele slučajne veličine X je tada:

$$F(x) = \frac{B_x(a, b)}{B(a, b)} = I_x(a, b),$$

gde su $B_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$ i $I_x(\alpha, \beta) = \frac{B_x(\alpha, \beta)}{B(\alpha, \beta)}$ nepotpune beta funkcije.

Važi $E(X) = \frac{a}{a+b}$, $D(X) = \frac{ab}{(a+b)^2(a+b+1)}$ (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 34, 366-367).

Stav 3.7. Neka su $U_1 \sim \mathcal{U}(0, 1)$ i $U_2 \sim \mathcal{U}(0, 1)$ međusobno nezavisne slučajne veličine. Ako za $X_1 = U_1^{\frac{1}{a}}$ i $X_2 = U_2^{\frac{1}{b}}$ važi $X_1 + X_2 \leq 1$, tada slučajna veličina $X = \frac{X_1}{X_1 + X_2}$ ima $B(a, b)$ raspodelu (Jöhnk, 1964).

Algoritam 3.12. Slučajna veličina $X \sim B(a, b)$ može se modelirati na sledeći način.

1. Neka su u_1 i u_2 dva broja, međusobno nezavisno generisana iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $x = \frac{u_1^{\frac{1}{a}}}{u_1^{\frac{1}{a}} + u_2^{\frac{1}{b}}}$.
3. Proveriti da li važi $u_1^{\frac{1}{a}} + u_2^{\frac{1}{b}} \leq 1$.
 - a. Ako važi, uzeti x kao realizovanu vrednost slučajne veličine.
 - b. Ako ne važi, u_1 i u_2 odbaciti i ponoviti postupak s drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele.

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Stav 3.8. Neka su $X_1 \sim \Gamma(a, 1)$ i $X_2 \sim \Gamma(b, 1)$ uzajamno nezavisne slučajne veličine. Tada slučajna veličina $X = \frac{X_1}{X_1 + X_2}$ ima $B(a, b)$ raspodelu.

Algoritam 3.13. Postupak modeliranja slučajne veličine $X \sim B(a, b)$ je sledeći.

1. Neka je x_1 broj generisan iz $\mathcal{G}(a, 1)$ raspodele.
2. Neka je x_2 broj generisan iz $\mathcal{G}(b, 1)$ raspodele.
3. Izračunati $x = \frac{x_1}{x_1 + x_2}$.
4. Uzeti x kao realizovanu vrednost slučajne veličine X .

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Postupak ponoviti sa drugim vrednostima iz $\mathcal{G}(a, 1)$ i $\mathcal{G}(b, 1)$ raspodela dok se ne dobije uzorak traženog obima.

Beta raspodela se primenjuje u meteorologiji, biologiji, hidrologiji, komunikacijama, operacionim istraživanjima, klimatologiji, Bajesijanskoj statistici itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 37).

Modeliranje binomne raspodele

Neka se izvodi n eksperimenata pod istim uslovima i nezavisno jedan od drugog i neka je verovatnoća realizacije posmatranog događaja A u svakom od tih eksperimenata konstantna i jednaka p .

Definicija 3.10. Binomna raspodela sa parametrima n i p je raspodela slučajne veličine X koja je jednaka broju realizacija događaja A pri opisanim uslovima. Koristi se oznaka $X \sim \mathcal{B}(n, p)$.

Zakon raspodele slučajne veličine X dat je sa

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Iz definicije sledi da za slučajnu veličinu X sa binomnom raspodelom važi

$$X = I_1 + I_2 + \dots + I_n, \quad (3.6)$$

gde su $I_j, j = 1, \dots, n$ međusobno nezavisni indikatori događaja A .

Važi $E(X) = np$ i $D(X) = np(1 - p)$.

Kako za $X \sim \mathcal{B}(n, p)$ važi (3.6) posmatraće se indikatori $I_j, j = 1, \dots, n$ događaja A u svakom od n eksperimenata.

Algoritam 3.14. Postupak modeliranja slučajne veličine $X \sim \mathcal{B}(n, p)$ metodom inverzne funkcije je sledeći:

1. Neka su u_1, u_2, \dots, u_n brojevi generisani iz $\mathcal{U}(0, 1)$ raspodele.
2. Neka su i_1, i_2, \dots, i_n redom modelirane vrednosti slučajnih veličina I_1, I_2, \dots, I_n korišćenjem brojeva u_1, u_2, \dots, u_n metodom ilustrovanom u algoritmu 3.11.
3. Izračunati $x = i_1 + i_2 + \dots + i_n$.
4. Uzeti broj x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugom n -torkom brojeva iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Algoritmi drugih metoda modeliranja za binomnu raspodelu u nastavku neće biti navođeni zbog često složenih postupaka izračunavanja verovatnoće p .

Binomna raspodela se primenjuje u hidrologiji i statistici pri opisivanju drugih raspodela i metoda statističke analize (Popović & Blagojević, 2003, str. 29-32; Popović, 2003, str. 121-124).

Modeliranje Burove raspodele

Definicija 3.11. *Burova*¹⁵ (dvoparametarska tipa XII) **raspodela** sa parametrima $a > 0$ i $b > 0$ (parametri oblika) je raspodela slučajne veličine X čija je gustina raspodele

$$g(x) = \frac{abx^{a-1}}{(1+x^a)^{b+1}}, \quad x > 0.$$

Koristi se oznaka $X \sim \text{Burr}(a, b)$.

Funkcija raspodele slučajne veličine X je tada

$$F(x) = 1 - (1 + x^a)^{-b}, \quad x > 0,$$

pa je njoj inverzna funkcija

$$F^{-1}(u) = \left(\frac{1}{\frac{1}{u^b} - 1} \right)^{\frac{1}{a}}, \quad 0 \leq u < 1,$$

čijom se primenom u metodi inverzne funkcije može modelirati X .

Važi

$$E(X) = bB\left(b - \frac{1}{a}, 1 + \frac{1}{a}\right) \text{ i}$$

$$D(X) = b\left(B\left(b - \frac{2}{a}, 1 + \frac{2}{a}\right) - B\left(b - \frac{1}{a}, 1 + \frac{1}{a}\right)\right).$$

Modeliranje slučajne veličine X se može ostvariti i korišćenjem sledećeg stava.

¹⁵ Irving Wingate Burr (1908-1989), američki statističar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 3.9. Neka je data slučajna veličina $Y \sim \mathcal{E}\left(\frac{1}{b}\right)$. Tada $X = (e^Y - 1)^{\frac{1}{a}} \sim \mathcal{Burr}(a, b)$.

Iz ovog stava se može zaključiti da se eksponencijalna raspodela primenjuje u opisivanju Burove raspodele. Burova raspodela se primeniti u opisivanju gama raspodele (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 66).

Modeliranje eksponencijalne raspodele

Eksponencijalna raspodela nije predmet ovog istraživanja, ali, kako će se pokazati, eksponencijalna raspodela često bude raspodela slučajne veličine sa majorantnom gustinom kod metode prihvatanja i odbacivanja. Kako se druge raspodele mogu opisati eksponencijalnom raspodelom, ona se može koristiti pri njihovom direktnom modeliranju. Na primer, u Boks-Milerovom postupku koristi se za modeliranje normalne raspodele. Stoga je bitno izložiti i algoritme modeliranja ove raspodele.

Definicija 3.12. Eksponencijalna raspodela je raspodela slučajne veličine X sa gustinom

$$g_X(x) = \begin{cases} 0, & x \leq 0 \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}, \quad \lambda > 0.$$

Funkcija raspodele slučajne veličine X je tada:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Parametar λ je parametar razmere.¹⁶

$$\text{Važi } E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}.$$

Inverzna funkcija funkcije eksponencijalne raspodele sa parametrom λ je

$$F_X^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u), \quad 0 < u < 1. \quad (3.7)$$

Da bi metoda inverzne funkcije bila uspešno primenjena, broj u mora biti generisan iz $\mathcal{U}(0, 1)$ raspodele. Umesto (3.7) se može koristiti $F_X^{-1}(u) = -\frac{1}{\lambda} \ln u$, $0 < u < 1$, što sledi iz tvrđenja a) stava 2.1.

¹⁶ Naziva se i parametrom stope otkaza ili parametrom smrtnosti.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Algoritam 3.15. Postupak modeliranja slučajne veličine $X \sim \mathcal{E}(\lambda)$ sledeći.

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Izračunati $x = -\frac{1}{\lambda} \ln u$.
3. Uzeti x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Drugi načini modeliranja eksponencijalne raspodele (Reuven, 1981, str. 68-69) obično zahtevaju više slučajnih brojeva za seme pa prethodni algoritam očigledno ima prednost zbog brzine modeliranja.

Eksponencijalna raspodela se primenjuje u biologiji, fizici, sistemima masovnog opsluživanja, elektrotehnici, meteorologiji, hidrologiji, teoriji pouzdanosti itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 124)

Modeliranje gama raspodele

Definicija 3.13. *Gama raspodela (dvoparametarska), sa parametrima a (parametar oblika) i b (parametar skaliranja), je raspodela slučajne veličine X čija je gustina*

$$g(x) = \frac{1}{b\Gamma(a)} \left(\frac{x}{b}\right)^{a-1} e^{-\frac{x}{b}}, \quad x \geq 0,$$

gde je $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ gama funkcija. Koristi se oznaka $X \sim \Gamma(a, b)$.

Funkcija raspodele slučajne veličine X je tada

$$F(x) = \frac{1}{\Gamma(a)} \gamma\left(a, \frac{x}{b}\right), \quad x \geq 0,$$

gde je $\gamma(\alpha, \beta) = \int_0^\beta t^{\alpha-1} e^{-t} dt$ nepotpuna gama funkcija.

Parametar b je parametar skaliranja.

Važi $E(X) = ab$, $D(X) = ab^2$ (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 98, 363, 365).

Metoda inverzne funkcije je neprimenljiva zbog nepostojanja eksplicitnog analitičkog izraza inverzne funkcije funkcije gama raspodele.

Gama raspodela ima tzv. reproduktivno svojstvo dato sledećim stavom.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 3.10. Ako su $X_i, i = 1, \dots, n$ međusobno nezavisne slučajne veličine sa raspedelama $\Gamma(a_i, b), i = 1, \dots, n$, gde je $\sum_{i=1}^n a_i = a$, tada slučajna veličina $X = \sum_{i=1}^n X_i$ ima $\Gamma(a, b)$ raspodelu.

Ukoliko su poznate vrednosti slučajnih veličina $X_i, i = 1, \dots, n$, njihovim sabiranjem će se modelirati vrednosti slučajne veličine $X: \Gamma(a, b)$.

Stav 3.11. Neka je $a = m, m \in \mathbb{N}$. Tada se slučajna veličina $X \sim \Gamma(m, b)$ može predstaviti kao suma m međusobno nezavisnih slučajnih veličina sa $\mathcal{E}\left(\frac{1}{b}\right)$ raspodelom koje se mogu dobiti metodom inverzne funkcije iz $\mathcal{U}(0, 1)$ raspodele. Dakle,

$$X = \sum_{i=1}^m -b \ln U_i = -b \ln \prod_{i=1}^m U_i,$$

gde su $U_i: \mathcal{U}(0, 1), i = 1, \dots, m$ (Reuven, 1981, str. 61-62).

Definicija 3.14. Erlangova¹⁷ raspodela sa parametrima m i b je raspodela slučajne veličine $X \sim \Gamma(m, b)$ gde je $m \in \mathbb{N}$. Koristi se oznaka $X \sim \mathcal{E}r(m, b)$.

Algoritam 3.16. Slučajna veličina $X \sim \Gamma(m, b)$, tj. $X \sim \mathcal{E}r(m, b)$ se modelira na sledeći način

1. Neka su u_1, \dots, u_m brojevi generisani iz raspodele $\mathcal{U}(0, 1)$.
2. Izračunati $x = -b \ln \prod_{i=1}^m u_i$.
3. Uzeti x kao realizovanu vrednost slučajne veličine X .

Nakon što se jedna tražena vrednost modelira, postupak se ponavlja dok se ne dobije uzorak traženog obima.

Gama raspodela se primenjuje u finansijama, hidrologiji, meteorologiji, ekonomiji itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 105).

Modeliranje Gumbelove raspodele

Definicija 3.15. Gumbelova¹⁸ (dvoparametarska) raspodela sa parametrima $a > 0$ (parametar razmere) i $b \in \mathbb{R}$ (parametar lokacije) je raspodela slučajne veličine X čija je gustina raspodele

¹⁷ Agner Krarup Erlang (1878-1929), danski matematičar.

¹⁸ Emil Julius Gumbel (1891-1966), nemački matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$g(x) = ae^{-a(x-b)-e^{-a(x-b)}}, \quad x \in \mathbb{R}.$$

Koristi se oznaka $X \sim Gum(\lambda, a)$.

Funkcija raspodele slučajne veličine X je

$$F(x) = e^{-e^{-a(x-b)}}, \quad x \in \mathbb{R},$$

pa je njoj inverzna funkcija

$$F^{-1}(u) = \frac{b - \ln(-\ln u)}{a}, \quad 0 < u < 1$$

čijim se korišćenjem u metodi inverzne funkcije može modelirati X . Važi $E(X) = b + \gamma \frac{1}{a}$, gde je $\gamma \approx 0.577$ Ojlerova konstanta i $D(X) = \frac{\pi^2}{6a^2}$.

Gumbelova raspodela se primenjuje pri određivanju ekstremnih vrednosti slučajnih nizova, u hidrologiji, meteorologiji itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 111).

Modeliranje χ^2 (hi-kvadrat) raspodele

Definicija 3.16. χ^2 raspodela sa n stepeni slobode je raspodela slučajne veličine:

$$X = X_1^2 + X_2^2 + \dots + X_n^2$$

gde su X_1, X_2, \dots, X_n međusobno nezavisne slučajne veličine koje sve imaju $\mathcal{N}(0, 1)$ raspodelu. Koristi se oznaka $X \sim \chi_n^2$.

Gustina i funkcija raspodele slučajne veličine X su tada:

$$g(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad \text{i}$$

$$F(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \gamma\left(\frac{n}{2}, \frac{x}{2}\right) & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Važi $E(X) = n$, $D(X) = 2n$ (Jevremović, 2014, str. 118-119).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Iz definicije 3.16 sledi da se modeliranjem n (međusobno nezavisnih realizacija) vrednosti slučajne veličine sa standardnom normalnom raspodelom, a zatim sabiranjem njihovih kvadrata, modelira jedna vrednost slučajne veličine $X \sim \chi_n^2$.

Najpoznatija primena χ^2 raspodele je kroz χ^2 test. (Popović, 2003, str. 112-120).

Modeliranje Košijeve raspodele

Definicija 3.17. *Košijeva*¹⁹ (dvoparametarska) **raspodela** sa parametrima $a, a > 0$, (parametar lokacije) i $b, b > 0$ (parametar skaliranja), je raspodela slučajne veličine X čija je gustina:

$$g(x) = \frac{b}{\pi(b^2 + (x - a)^2)}, \quad x \in \mathbb{R}.$$

Koristi se oznaka $X \sim \mathcal{C}(a, b)$.

Funkcija raspodele slučajne veličine X je tada:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x - a}{b}, \quad x \in \mathbb{R}.$$

Očekivanje i disperzija slučajne veličine X ne postoje jer integral $\int_{-\infty}^{+\infty} xg(x) dx$ divergira.

Inverzna funkcija Košijeve raspodele je

$$F_X^{-1}(u) = a - \frac{b}{\operatorname{tg} \pi u}, \quad u \in (0, 1)$$

pa se modeliranje uzorka željenog obima metodom inverzne funkcije dobija izračunavanjem vrednosti inverzne funkcije funkcije raspodele za uzorak slučajnih brojeva istog obima.

Stav 3.12. a) Ako su date dve međusobno nezavisne slučajne veličine $Z_1 \sim \mathcal{N}(0, 1)$ i $Z_2 \sim \mathcal{N}(0, 1)$, tada $X = \frac{Z_1}{Z_2}$ ima Košijevu raspodelu $\mathcal{C}(0, 1)$. b) Ako $Z \sim \mathcal{C}(0, 1)$ tada $X \sim \mathcal{C}(a, b)$, gde je $X = bZ + a$.

Algoritam 3.17. Postupak modeliranja slučajne veličine $X \sim \mathcal{C}(a, b)$ je sledeći.

1. Neka su z_1 i z_2 dva međusobno nezavisno generisana broja iz raspodele $\mathcal{N}(0, 1)$.
2. Izračunati $x = b \frac{z_1}{z_2} + a$.

¹⁹ Augustin Louis Cauchy (1789-1857), francuski matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

3. Uzeti x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{N}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Stav 3.13. *Ako su date međusobno nezavisne slučajne veličine $U, V \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$ i ako važi $U^2 + V^2 \leq \frac{1}{4}$, tada $X \sim \mathcal{C}(0, 1)$, ako je $X = \frac{U}{V}$.*

Algoritam 3.18. Postupak modeliranja slučajne veličine $X \sim \mathcal{C}(a, b)$ je sledeći.

1. Neka su u i v dva međusobno nezavisno generisana broja iz raspodele $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$.
2. Proveriti $u^2 + v^2 \leq \frac{1}{4}$
 - a. Ako važi, izračunati $x = b \frac{u}{v} + a$.
 - b. Ako ne važi, ponovimo postupak sa drugim vrednostima iz $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$.
3. Uzeti x kao realizovanu vrednost slučajne veličine X .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$ raspodele dok se ne dobije uzorak traženog obima.

Košijeva raspodela se primenjuje u mehanici, elektrotehnici, antropologiji, teoriji mere, finansijama i analizama rizika (Alzaatreh, Lee, Famoye & Ghosh, 2016).

Modeliranje lognormalne raspodele

Definicija 3.18. *Lognormalna raspodela sa parametrima μ i σ^2 je raspodela slučajne veličine Y za koju $\ln Y$ ima $\mathcal{N}(\mu, \sigma^2)$ raspodelu.*

Očigledno $\ln Y = X$, tj. $Y = e^X$, gde $X: \mathcal{N}(\mu, \sigma^2)$.

Modeliranje lognormalne raspodele se vrši tako što navedenim postupcima modelira slučajna veličina Z sa standardnom normalnom raspodelom, pa se modelirane vrednosti transformišu u normalnu raspodelu sa željenim parametrima računajući $X = \mu + \sigma Z$, a zatim se broj e^X uzima kao modelirana vrednost slučajne veličine sa lognormalnom raspodelom.

Lognormalna raspodela se primenjuje u klimatologiji, ekonomiji, telekomunikacijama, kontroli kvaliteta, geologiji, medicini, biohemiji, meteorologiji itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 243).

Modeliranje Laplasove raspodele

Definicija 3.19. Laplasova (dvoparameterska) raspodela sa parametrima $\lambda > 0$ (parametar skaliranja) i $a \geq 0$ (parametar lokacije) je raspodela slučajne veličine X čija je gustina raspodele

$$g(x) = \frac{1}{2\lambda} e^{-\frac{|x-a|}{\lambda}}, \quad x \in \mathbb{R}.$$

Koristi se oznaka $X \sim \text{Lap}(\lambda, a)$.

Funkcija raspodele slučajne veličine X je

$$F(x) = \begin{cases} \frac{1}{2} e^{-\frac{x-a}{\lambda}}, & x \leq a \\ 1 - \frac{1}{2} e^{-\frac{x-a}{\lambda}}, & x > a \end{cases},$$

pa je njoj inverzna funkcija

$$F^{-1}(u) = \begin{cases} a + \lambda \ln u, & 0 < u \leq \frac{1}{2} \\ a - \lambda \ln u, & \frac{1}{2} < u \leq 1 \end{cases}$$

čijim se korišćenjem u metodi inverzne funkcije može modelirati X . Važi $E(X) = 0$ i $D(X) = 2\lambda^2$.

Takođe, modeliranje slučajne veličine X se može ostvariti korišćenjem sledećeg stava.

Stav 3.14. Neka su date nezavisne slučajne veličine $Y_1, Y_2 \sim \mathcal{E}\left(\frac{1}{\lambda}\right)$ i $U_1, U_2 \sim \mathcal{U}(0, 1)$. Tada $X_2 = Y_1 - Y_2 \sim \text{Lap}(\lambda, 0)$ i $X_2 = \frac{U_1}{U_2} \sim \text{Lap}(1, 0)$ kad je $U_2 \neq 0$.

Laplasova raspodela se primenjuje u finansijama (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 185).

Modeliranje logističke raspodele

Definicija 3.20. Logistička (dvoparameterska) raspodela sa parametrima $a \in \mathbb{R}$ (parametar lokacije) i $b > 0$ (parametar skaliranja) je raspodela slučajne veličine X čija je gustina raspodele

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$g(x) = \frac{e^{-\frac{x-a}{b}}}{b \left(1 + e^{-\frac{x-a}{b}}\right)^2}, \quad x \in \mathbb{R}.$$

Koristi se oznaka $X \sim \mathcal{Lg}s(a, b)$.

Funkcija raspodele slučajne veličine X je

$$F(x) = \frac{e^{-\frac{x-a}{b}}}{b \left(1 + e^{-\frac{x-a}{b}}\right)^2}, \quad x \in \mathbb{R},$$

pa je njoj inverzna funkcija

$$F^{-1}(u) = a + b \ln \frac{u}{1-u}, \quad 0 < u < 1$$

čijim se korišćenjem u metodi inverzne funkcije može modelirati X .

$$\text{Važi } E(X) = a \text{ i } D(X) = \frac{b^2\pi^2}{3}.$$

Logistička raspodela se primenjuje u ekonomiji i demografiji (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 205).

Normalna raspodela i njeno modeliranje

Testiranje normalnosti raspodele je centralna tema ovog rada, pa će biti izloženo više informacija o normalnoj raspodeli kako bi pristup temi bio što potpuniji. Normalna raspodela je model koji opisuje ubedljivo najveći broj obeležja koji su tema naučno-stručnih istraživanja skoro svake discipline (Martin, 1985, str. 106, 242; Wasserman, 2005, str. 28).

Definicija 3.21. Slučajna veličina X ima **normalnu (Gausovu²⁰) raspodelu** ako joj je gustina raspodele data sa

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

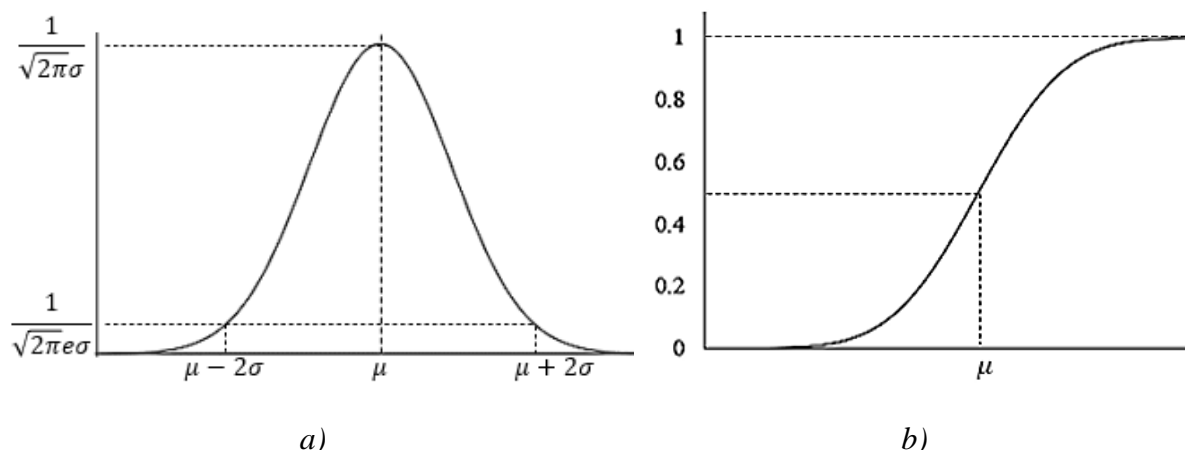
gde su $\mu \in \mathbb{R}$ i $\sigma \in \mathbb{R}^+$ fiksirani. Koristi se oznaka $X \sim \mathcal{N}(\mu, \sigma^2)$.

²⁰ Johann Carl Friedrich Gauss (1777-1855), nemački matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Funkcija normalne raspodele data je sa

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R}.$$



Slika 3.4. a) Grafik gustine raspodele za normalnu raspodelu. b) Grafik funkcije raspodele za normalnu raspodelu.

Stav 3.15. Za slučajnu veličinu $X \sim \mathcal{N}(\mu, \sigma^2)$ važe sledeća tvrđenja.

- a. Matematičko očekivanje i disperzija su redom $E(X) = \mu$ i $D(X) = \sigma^2$.
- b. Mod i medijana su redom $M_o(X) = m$ i $M_e(X) = m$.
- c. Centralni momenti neparnog reda su svi jednaki nuli, tj. $C_{2k+1} = 0$, $k \in \mathbb{N}_0$.
- d. Centralni momenti parnog reda su $C_{2k} = (2k - 1)!! \sigma^{2k}$, $k \in \mathbb{N}$.
- e. Koeficijenti asimetrije i spoljštenosti su oba jednaki nuli, tj. $KA(X) = \frac{C_3}{\sigma^3} = 0$ i $KS(X) = \frac{C_4}{\sigma^4} = 0$.

Stav 3.16. (Osobine normalne raspodele) Za slučajnu veličinu $X \sim \mathcal{N}(\mu, \sigma^2)$ važe sledeća tvrđenja

- a. Lokalni maksimum grafika gustine normalne raspodele je u tački $M_{\max}\left(\mu, \frac{1}{\sqrt{2\pi}\sigma}\right)$.
- b. Grafik gustine normalne raspodele je simetričan u odnosu na pravu datu jednačinom $x = \mu$.
- c. Grafik funkcije raspodele je centralno simetričan u odnosu na tačku $M\left(\mu, \frac{1}{2}\right)$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

- d. Prevojne tačke grafika gustine normalne raspodele su $P_{1,2} = \left(\mu \pm 2\sigma, \frac{1}{\sqrt{2\pi e}\sigma}\right)$.
- e. Prevojna tačka grafika funkcije raspodele je $P\left(\mu, \frac{1}{2}\right)$.
- f. Horizontalna asimptota grafika gustine normalne raspodele data je jednačinom $y = 0$.
- g. Horizontalne asimptote grafika funkcije raspodele date su jednačinama $y = 0$ i $y = 1$.

Stav 3.17. Ako slučajna veličina X ima normalnu $\mathcal{N}(\mu, \sigma^2)$ raspodelu, tada slučajna veličina $Z = \frac{X-\mu}{\sigma}$ ima normalnu $\mathcal{N}(0,1)$ raspodelu.

Sledeće tvrđenje se može smatrati polaznom tačkom daljeg istraživanja u ovom radu, što se može videti u narednim poglavljima (4 i 5).

Stav 3.18. Za slučajnu veličinu $X \sim \mathcal{N}(\mu, \sigma^2)$ važi pravilo „3-sigma“²¹, tj.

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0.6826, \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 0.9544, \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 0.9973. \blacksquare \end{aligned}$$

Kako integral

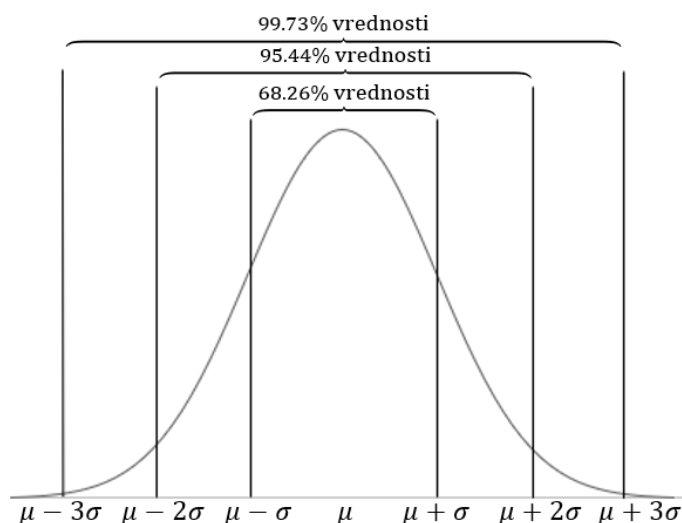
$$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

ne može da se izrazi preko elementarnih funkcija, vrednosti funkcije raspodele, zaokružene na četiri decimale, su izračunate za standardnu normalnu raspodelu $\mathcal{N}(0, 1)$ za vrednosti od -3,6 do 3,6 (na dve decimale) i smeštene u tablice koje se i dalje često koriste, a svi matematički i statistički softveri nude mogućnost aproksimacije funkcije normalne raspodele za proizvoljan realan broj ili parametre. Korišćenjem jedne od tih mogućnosti izračunavanja se može proveriti tvrđenje stava 3.18.

Konkretno za korišćenje tablica, bez promene vrednosti verovatnoće, može se koristiti stav 3.17 i iz tablice videti da stav 3.18 važi. Tablice normalne, a i drugih poznatih raspodela, se mogu naći u bilo kom udžbeniku ili zbirci zadataka iz matematičke statistike.

²¹ U literaturi se navodi i kao „6-sigma“.

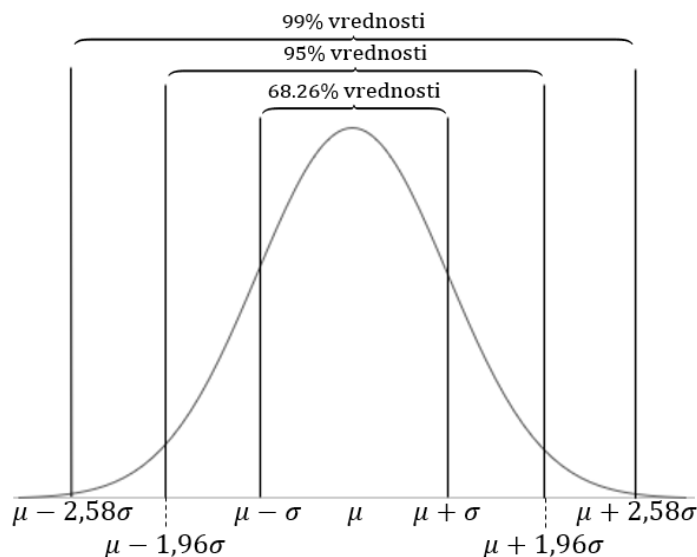
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 3.5. Pravilo 3σ .

Pravilo 3-sigma se može dati i u prilagođenom obliku. Termin „prilagođeno“ se koristi jer su u ovoj varijanti tvrdjenja multiplikatori standardne devijacije prilagođeni na vrednosti izabrane tako da rezultujuće verovatnoće budu redom 0,6826 (koja se često zaokružuje na 0,68 (Minić, 2018)), 0,95 i 0,99 redom. Dakle,

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0,6826, \\ P(\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma) &= 0,95, \\ P(\mu - 2,58\sigma \leq X \leq \mu + 2,58\sigma) &= 0,99. \end{aligned}$$



Slika 3.6. Pravilo 3σ u prilagođenom obliku.

Iako je ovde dat ovaj način prilagođavanja pravila 3-sigma, drugi načini se mogu koristiti. Izbor množioca standardne devijacije se može birati na proizvoljan način, ili tako da rezultujuće verovatnoće budu neke druge. Na primer, umesto množioca 1, za koji je

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

odgovarajuća verovatnoća 0,6826, može se staviti množilac za koji će odgovarajuća verovatnoća biti 0,7 ili 0,75.

Stav 3.19. *Ako je $X_1, \dots, X_n, n \in \mathbb{N}$ prost slučajan uzorak obeležja X sa normalnom raspodelom, onda su uzoračka sredina $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ i uzoračka disperzija $\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ međusobno nezavisne (Geary, 1936; Lukacs, 1942). Drugim rečima, nezavisnost uzoračke sredine i uzoračke disperzije u klasi neprekidnih raspodela važi samo za normalnu raspodelu.*

Sledeće tvrđenje nije direktno vezano za temu ovog istraživanja ali se u ostalim izvorima neizostavno spominje sa osobinama normalne raspodele. Stoga će i ovde biti navedeno bez dokaza.

Teorema 3.1. (Muavr²²-Laplasova²³ teorema) *Neka slučajna veličina S_n ima binomnu $B(n, p)$ raspodelu. Tada za verovatnoće iz zakona raspodele za S_n važi*

$$\lim_{n \rightarrow +\infty} P(S_n = k) = g(k), \quad k \in \{0, 1, \dots, n\}.$$

gde je $g(x) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(x-np)^2}{2np(1-p)}}$, $x \in \mathbb{R}$.

Teorema 3.2. (Centralna granična teorema) *Neka su X_1, X_2, \dots, X_n međusobno nezavisne slučajne veličine sa istom raspodelom sa očekivanjem μ i disperzijom σ^2 . Tada slučajna veličina $Z_n = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$ konvergira u raspodeli (Jevremović, 2014, str. 132) ka slučajnoj veličini $Z \sim \mathcal{N}(0, 1)$, tj. $Z_n \xrightarrow{D} Z$.*

Centralna granična teorema je od suštinskog značaja u matematičkoj statistici (Borovkov, 2013, str. 185).

Kako na osnovu stava 3.17 sledi da ako $Z \sim \mathcal{N}(0, 1)$, onda $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$, to će za modeliranje normalne raspodele za bilo koje vrednosti parametara biti dovoljno modelirati standardnu normalnu raspodelu $\mathcal{N}(0, 1)$.

Metodu inverzne funkcije nije moguće koristiti direktno jer eksplicitan izraz funkcije raspodele, a samim tim i njoj inverzne funkcije, zahtevaju dugotrajna izračunavanja. Metoda inverzne funkcije se može primeniti korišćenjem statističkih tablica. Takođe, ponuđeno je više drugih korisnih načina modeliranja standardne normalne raspodele.

²² Abraham de Moivre (1667-1754), francuski matematičar.

²³ Pierre-Simon, Marquis de Laplace (1749-1827), francuski matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Boks²⁴-Miler²⁵ postupak, ili još **metoda polarnih koordinata**, je izuzetno efikasan način modeliranja dveju međusobno nezavisno realizovanih vrednosti standardne normalne raspodele $Z \sim \mathcal{N}(0, 1)$. Zasnovan je na sledećoj teoremi.

Teorema 3.3. *Neka su date nezavisne slučajne veličine $U \sim \mathcal{U}(0, 1)$ i $Y \sim \mathcal{E}(1)$. Tada su*

$$\begin{aligned} Z_1 &= \sqrt{2Y} \cos 2\pi U, \\ Z_2 &= \sqrt{2Y} \sin 2\pi U, \end{aligned}$$

međusobno nezavisne slučajne veličine sa standardnom normalnom raspodelom.

Algoritam 3.19. Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ Boks-Milerovim postupkom je sledeći.

1. Neka je u broj generisan iz $\mathcal{U}(0, 1)$ raspodele.
2. Neka je y broj generisan iz $\mathcal{E}(1)$ raspodele.
3. Izračunati $z_1 = \sqrt{2y} \cos 2\pi u$ i $z_2 = \sqrt{2y} \sin 2\pi u$.
4. Uzeti z_1 i z_2 kao realizovane vrednosti slučajne veličine Z .

Postupak ponoviti sa drugim brojevima iz raspodela $\mathcal{U}(0, 1)$ i $\mathcal{E}(1)$ dok se ne dobije uzorak traženog obima.

Iako moderni softveri izračunavaju vrednosti trigonometrijskih funkcija veoma brzo, kod većih obima uzoraka koje treba generisati to i dalje može biti problem. Boks-Milerov postupak se ipak može prilagoditi u bržu varijantu datu sledećim algoritmom.

Algoritam 3.20. Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ je sledeći.

1. Neka su u_1 i u_2 brojevi nezavisno generisani iz $\mathcal{U}(0, 1)$ raspodele.
2. Neka je $r^2 = u_1^2 + u_2^2$.
 - a. Ako je $r^2 < 1$, onda se $z_1 = u_1 \sqrt{-\frac{2 \ln r^2}{r^2}}$ i $z_2 = u_2 \sqrt{-\frac{2 \ln r^2}{r^2}}$ uzimaju kao nezavisno realizovane vrednosti slučajne veličine Z .
 - b. Ako je $r^2 \geq 1$, onda se u_1 i u_2 odbacuju i postupak se ponavlja sa novim brojevima generisanim iz raspodele $\mathcal{U}(0, 1)$.

Postupak ponoviti sa drugim brojevima iz raspodele $\mathcal{U}(0, 1)$ dok se ne dobije uzorak traženog obima.

²⁴ George Edward Pelham Box (1919-2013), engleski statističar.

²⁵ Mervin Edgar Muller (1928-2018), američki informatičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Definicija 4.6 i stav 4.1 ovog rada daju matematičku zasnovanost sledećeg algoritma za modeliranje normalne raspodele $\mathcal{N}(0, 1)$.

Algoritam 3.21. (Algoritam zona) Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ je sledeći.

1. Neka je u broj generisan iz raspodele $\mathcal{U}(0, 1)$, i v generisan iz raspodele $\mathcal{U}(-1, 1)$.
2. Ako važi $0 \leq |v| \leq 0.6826$, onda se $z = \text{sgn } v \cdot (1 - u)$ uzima kao realizacija slučajne veličine Z .
3. Ako važi $0.6826 < |v| \leq 0.9544$, onda se $z = \text{sgn } v \cdot (2 - u)$ uzima kao realizacija slučajne veličine Z .
4. Ako važi $0.9544 < |v| \leq 0.9973$, onda se $z = \text{sgn } v \cdot (3 - u)$ uzima kao realizacija slučajne veličine Z .
5. Ako važi $|v| > 0.99$, onda se $z = \text{sgn } v \cdot (3 + u)$ uzima kao realizacija slučajne veličine Z .

Ako uslov iz bilo kojeg od ovih koraka nije zadovoljen onda se prelazi na sledeći. Postupak ponoviti sa drugim brojevima raspodela $\mathcal{U}(0, 1)$ i $\mathcal{U}(-1, 1)$ dok se ne dobije uzorak traženog obima.

Prednosti algoritma 3.21 su te što jedan od uslova koji se proveravaju sigurno važi pa se nijedan par slučajnih brojeva ne odbacuje, izračunavanja su vrlo jednostavna pa je uticaj na brzinu generisanja minimalan. Negativna strana bi mogla biti to što se može desiti da se proveravaju četiri uslova pre nego se broj generiše, mada je najverovatnija provera samo jednog od uslova. Ograničenje generisanih brojeva na interval $(-4, 4)$ ne predstavlja problem jer je verovatnoća generisanja broja izvan tog intervala vrlo bliska nuli, a iako se takvi brojevi u uzorku dobiju uglavnom se smatraju spoljašnjim vrednostima. Ograničenje ovog algoritma je to što su za $n > 2000$ uzorci modeliranih brojeva više odstupaju od normalne raspodele. Naime, pri testiranju saglasnosti sa normalnom raspodelom p-vrednost opada (za $n > 2000$ ispod 0,1).

Algoritam je brz što se može proveriti simulacijama. Naime, u proseku se dobija da ovaj algoritam generiše normalno raspodeljene vrednosti za 1,02 vremena koje je potrebno ziggurat algoritmu koji se danas smatra najbržim.²⁶

Na sličan način za formiranje ovakvog algoritma se mogu koristiti funkcija zone data definicijom 5.4, okarakterisana stavom 5.3.

Sledeći algoritam je zasnovan na metodi prihvatanja i odbacivanja gde normalnost modeliranih vrednosti sledi iz rasuđivanja datog kod Avdovića (2019, str. 77-78).

²⁶ Simulacije su vršene na računaru "ACER Extensa EX215-54 (Black) Full HD IPS, I7-1165G7, 8GB, 512GB SSD (NX.EGJEX.00L)" preko programskog paketa MATLAB.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Algoritam 3.22. Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ je sledeći.

1. Neka su y_1 i y_2 brojevi, međusobno nezavisno generisani iz $\mathcal{E}(1)$ raspodele.
2. Proveriti da li važi $y_2 \geq \frac{(y_1-1)^2}{2}$.
 - a. Ako važi, uzeti novi broj u generisan iz $\mathcal{U}(0, 1)$ raspodele.
 - b. Ako ne važi, ponoviti postupak sa druga dva broja iz $\mathcal{E}(1)$ raspodele.
3. Proveriti da li važi $u \geq 0,5$.
 - a. Ako važi, uzeti $z = -y_1$ kao realizovanu vrednost slučajne veličine Z .
 - b. Ako ne važi, uzeti $z = y_1$ kao realizovanu vrednost slučajne veličine Z .

Postupak ponoviti sa drugim brojevima iz raspodela $\mathcal{E}(1)$ i $\mathcal{U}(0, 1)$ dok se ne dobije uzorak traženog obima.

Da su vrednosti modelirane sledećim algoritmom normalno raspodeljene posledica je centralne granične teoreme, a obrazloženje je dato kod Jevremović (2014, str. 154, 156).

Algoritam 3.23. Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ je sledeći.

1. Neka su u_1, u_2, \dots, u_{12} međusobno nezavisno generisani brojevi iz raspodele $\mathcal{U}(0, 1)$.
2. Izračunati $\sum_{i=1}^{12} u_i - 6$.
3. Uzeti z kao realizovanu vrednost slučajne veličine Z .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Sledeći algoritam predstavlja jednu varijantu metode mere uniformnosti. Izvođenje rezultata dato je kod Knuta (1998, str. 130-132).

Algoritam 3.24. Postupak modeliranja slučajne veličine $Z \sim \mathcal{N}(0, 1)$ je sledeći.

1. Neka su u_1 i u_2 međusobno nezavisno generisani brojevi iz raspodele $\mathcal{U}(0, 1)$.
2. Izračunati $x = \sqrt{\frac{\pi}{8}} \ln \frac{1+u_1}{1-u_1}$.
3. Proveriti da li važi $u_2 \leq 0,5$.
 - a. Ako važi, uzeti $z = -x$ kao realizovanu vrednost slučajne veličine Z .
 - b. Ako ne važi, uzeti $z = x$ kao realizovanu vrednost slučajne veličine Z .

Postupak ponoviti sa drugim brojevima iz $\mathcal{U}(0, 1)$ raspodele dok se ne dobije uzorak traženog obima.

Pored navedenih algoritama često korišćeni algoritam za modeliranje normalne raspodele je tzv. „**ziggurat**“ algoritam (Marsaglia & Tsang, 2000). Ovaj algoritam je sam po sebi jednostavan i brz, ali zahteva puno pripremnih objašnjenja i podešavanja (Ivanović, 2018, str. 39-42), stoga će ovde biti izostavljen. Sva detaljna objašnjenja i podešavanja data su u originalnom radu Marsalje i Tsanga (2000).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

U MATLABu pre verzije 5 za generisanje normalno raspodeljenih brojeva korišćen je Box-Miler algoritam, a od verzije 5 pa nadalje koristi se ziggurat algoritam (Moler, 2004).

Modeliranje Paretove raspodele

Definicija 3.22. *Paretova*²⁷ (dvoparametarska) **raspodela** sa parametrima $a > 0$ (parametar oblika) i $b > 0$ (parametar lokacije) je raspodela slučajne veličine X čija je gustina raspodele

$$g(x) = \frac{ab^a}{x^{a+1}}, \quad x \geq b.$$

Koristi se oznaka $X \sim \mathcal{P}ar(a, b)$.

Funkcija raspodele slučajne veličine X je tada

$$F(x) = 1 - \left(\frac{b}{x}\right)^a, \quad x \geq b,$$

pa je njoj inverzna funkcija

$$F^{-1}(u) = \frac{b}{(1-u)^{\frac{1}{a}}}, \quad 0 \leq u < 1,$$

čijim se korišćenjem u metodi inverzne funkcije može modelirati X .

Važi

$$E(X) = \frac{ab}{a-1}, \quad a > 1 \text{ i}$$

$$D(X) = \frac{ab^2}{(a-2)(a-1)^2}, \quad a > 2.$$

Modeliranje slučajne veličine X se može ostvariti i korišćenjem sledećeg stava.

Stav 3.20. *Neka je data slučajna veličina $Y \sim \mathcal{E}(1)$. Tada $X = be^{\frac{Y}{a}} \sim \mathcal{P}ar(a, b)$.*

Paretova raspodela se primenjuje u ekonomiji, astronomiji, meteorologiji, lingvistici, energetici itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 276).

²⁷ Vilfredo Federico Damaso Pareto (1848-1923), italijanski ekonomista.

Modeliranje Studentove raspodele

Definicija 3.23. Studentova²⁸ raspodela sa n stepeni slobode je raspodela slučajne veličine:

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}}$$

za međusobno nezavisne slučajne veličine $Y \sim \mathcal{N}(0, 1)$ i $Z \sim \chi_n^2$. Koristi se oznaka $X \sim t_n$.

Gustina raspodele slučajne veličine X je tada:

$$g(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \frac{n}{2}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$

Zbog komplikovanosti izraza funkcije raspodele i činjenice da za modeliranje raspodele slučajne veličine X njeno poznavanje nije potrebno, ona se neće navoditi.

$$\text{Važi } E(X) = 0, D(X) = \begin{cases} \infty, & n \leq 2 \\ \frac{n}{n-2}, & n > 2 \end{cases} \text{ (Jevremović, 2014, str. 120).}$$

Definicija 3.23 omogućava modeliranje realizovane vrednosti slučajne veličine sa Studentovom raspodelom t_n nakon modeliranja dve međusobno nezavisne realizacije slučajnih veličina koje, redom, imaju raspodele $\mathcal{N}(0, 1)$ i χ_n^2 na osnovu formule definicije 3.23.

Studentova raspodela se primenjuje u opisivanju normalne, χ^2 , Košijeve i drugih raspodela (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 322).

Modeliranje Tjukijeve raspodele

Definicija 3.24. Tjukijeva²⁹ (jednparametarska) raspodela sa parametrom $\lambda > 0$ (parametar oblika) je raspodela slučajne veličine

$$X = \frac{U^\lambda - (1-U)^\lambda}{\lambda}, \quad \lambda \neq 0$$

²⁸ William Sealy Gosset (1876-1937), engleski hemičar i statističar.

²⁹ John Wilder Tukey (1915-2000), američki matematičar i statističar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

gde slučajna veličina U ima uniformnu raspodelu na $(0,1)$. Koristi se oznaka $X \sim \mathcal{J}(\lambda)$.

Funkcija i gustina raspodele se u opštem slučaju ne mogu odrediti iako je bilo rezultata u posebnim slučajevima (Johnson, Kotz & Balakrishnan, 1994, str. 39-42).

Važi

$$E(X) = 0, i$$

$$D(X) = \frac{2}{(\lambda + 1)(2\lambda + 1)} - \frac{2}{\lambda^2} \left(B(\lambda + 1, \lambda + 1) - \frac{1}{(\lambda + 1)^2} \right), \quad \lambda \neq 0.$$

Korišćenjem Definicije 3.24 Tjukijeve raspodele ona se lako modelira uz postojeći uzorak slučajnih brojeva. Kada je to pogodno, modeliranje slučajne veličine X se može vršiti i korišćenjem sledećeg stava.

Stav 3.21. Neka je data slučajna veličina $X \sim \mathcal{J}(\lambda)$. Tada za

- $\lambda = -1, X \sim \mathcal{C}(0, \pi)$ približno;
- $\lambda = 0,14, X \sim \mathcal{N}(0; 2,142)$ približno;
- $\lambda = 1, X \sim \mathcal{U}(-1, 1)$;
- $\lambda = 2, X \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$.

Ovim stavom su dati primeri primene Tjukijeve raspodele u opisivanju nekih drugih raspodela.

Modeliranje Weibulove raspodele

Definicija 3.25. Weibulova³⁰ (dvoparametarska) raspodela sa parametrima $a, a > 0$, (parametar lokacije) i $b, b > 0$, (parametar skaliranja) je raspodela slučajne veličine X čija je gustina:

$$g(x) = \begin{cases} \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Koristi se oznaka $X: \mathcal{W}(a, b)$.

³⁰ Ernst Hjalmar Waloddi Weibull (1887-1979), švedski inženjer i matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Funkcija raspodele slučajne veličine X je tada:

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{b}\right)^a}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

Važi $E(X) = b\Gamma\left(1 + \frac{1}{a}\right)$, $D(X) = b^2\left(\Gamma\left(1 + \frac{2}{a}\right) - \left(\Gamma\left(1 + \frac{1}{a}\right)\right)^2\right)$ (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 71-72).

Inverzna funkcija funkcije Vejbulove raspodele je

$$F_X^{-1}(u) = b(-\ln(1-u))^{\frac{1}{a}}, \quad u \in (0, 1)$$

čijim se korišćenjem u metodi inverzne funkcije može modelirati X .

Modeliranje je moguće vršiti i korišćenjem veze $\mathcal{W}(1, b) \sim \mathcal{E}(b)$.

Vejbulova raspodela se primenjuje u meteorologiji, teoriji pouzdanosti, komunikacijama, klimatologiji itd. (Đorić, Mališić, Jevremović & Nikolić-Đorić, 2007, str. 74).

4. Kontrola kvaliteta i kontrolne karte

Kontrola kvaliteta, tj. kontrolne karte predstavljaju jedan od glavnih delova ovog istraživanja. Naime, u ovom poglavlju su izloženi neki od važnih primera kontrolnih karata sa osvrtom na njihove osobine i performanse. Dodatno, izloženi su primeri kontrolnih karata koje su rezultat istraživanja autora, a čiji je cilj da budu bolja alternativa postojećim rešenjima. Shodno tome, prikazana je i analiza performansi svake od novih karata sa analizom poređenja sa performansama postojećih kontrolnih karata.

Na kraju ovog poglavlja izvršena je komparativna analiza performansi izloženih postojećih kontrolnih karata i novih kontrolnih karata koje su rezultat ovog istraživanja. Performanse se mere korišćenjem prosečnog broja uzorka pre prve detekcije defekta, tj. ARL-a za raspodelu u odnosu na koju se proces kontroliše i za raspodele sa pomerenim vrednostima očekivanja.

4.1. Empirijska funkcija raspodele

Ovo istraživanje se bavi testovima normalnosti i kontrolnim kartama. Neki od testova pa, i kontrolnih karata zasnovani su na empirijskoj funkciji raspodele. To znači da se empirijska funkcija raspodele koristi u njihovom definisanju i opisivanju. Najpre su dati postojeći i već istraženi rezultati iz oblasti kontrole kvaliteta, dok su poslednja dva odeljka (4.4 i 4.5) ovog poglavlja u potpunosti rezultat istraživanja autora.

Definicija 4.1. Neka je dat prost slučajnan uzorak X_1, \dots, X_n obeležja (slučajne veličine) X . Funkcija (slučajna veličina) definisana sa

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), \quad x \in \mathbb{R}$$

je **empirijska funkcija raspodele** uzorka X_1, \dots, X_n . Ako je k_x broj elemenata uzorka koji su manji ili jednaki od broja $x \in \mathbb{R}$, onda je $F_n^*(x) = \frac{k_x}{n}$.

$I(X_i \leq x)$ je indikator događaja

$$\{X_i \leq x\} = \{\omega \in \Omega: X_i(\omega) \leq x\}, \quad i = 1, \dots, n$$

čija je raspodela određena funkcijom raspodele $F(x)$ obeležja X i data sa

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$I(X_i \leq x): \begin{pmatrix} 0 & 1 \\ 1 - F(x) & F(x) \end{pmatrix}.$$

Empirijska funkcija raspodele je stepenasta i sa desne strane neprekidna funkcija sa vrednostima u $[0,1]$ i najviše prebrojivo mnogo tačaka prekida prve vrste.

Stav 4.1. *Empirijska funkcija raspodele za izabran realan broj $x \in \mathbb{R}$ ima binomni zakon raspodele dat sa*

$$P\left(F_n^*(x) = \frac{k}{n}\right) = \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}, \quad k = 1, \dots, n.$$

Sledeća tvrđenja su najvažnije osobine empirijske funkcije raspodele zbog čega se često smatraju jednim od nabitnijih elemenata matematičke statistike generalno (Popović, 2003, str. 17), a samim tim i ovog istraživanja.

Stav 4.2. *Neka je dat prost slučajan uzorak X_1, \dots, X_n obeležja X sa funkcijom raspodele $F(x)$ čija je empirijska funkcija raspodele $F_n^*(x)$ i neka je $x \in \mathbb{R}$. Važi*

$$P\left(\lim_{n \rightarrow \infty} F_n^*(x) = F(x)\right) = 1.$$

Teorema 4.1. *(Glivenko³¹-Kanteli³² teorema ili centralna teorema matematičke statistike) Neka je dat prost slučajan uzorak X_1, \dots, X_n obeležja X sa funkcijom raspodele $F(x)$ čija je empirijska funkcija raspodele $F_n^*(x)$ i neka je $x \in \mathbb{R}$ proizvoljno. Tada važi*

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| = 0\right) = 1.$$

Osobine empirijske funkcije raspodele se koriste za izvođenje rezultata datih sledećim teoremama.

Teorema 4.2. *(Teorema Kolmogorova³³) Neka je dat prost slučajan uzorak X_1, \dots, X_n apsolutno neprekidnog obeležja X sa funkcijom raspodele $F(x)$ čija je empirijska funkcija raspodele $F_n^*(x)$ i neka su $x, t \in \mathbb{R}, t > 0$ proizvoljni. Tada statistika*

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|$$

³¹ Валерій Іванович Гливінко (1897-1940), sovjetski (ukrajinski) matematičar.

³² Francesco Paolo Cantelli (1875-1966), italijanski matematičar.

³³ Андрій Николаевич Колмогоров (1903-1987), sovjetski (ruski) matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

ima raspodelu za koju važi

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n \leq t) = K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 t^2}.$$

Raspodela čija je funkcija $K(t)$ naziva se **raspodelom Kolmogorova** (Vrbik, 2020).

Teorema 4.3. (Teorema Smirnova³⁴) Neka su dati prosti slučajni uzorci obima n_1 i n_2 redom, apsolutno neprekidnog obeležja X i neka je $x \in \mathbb{R}$ proizvoljno. Tada statistika

$$D_{n_1, n_2} = \sup_{x \in \mathbb{R}} |F_{n_1}^*(x) - F_{n_2}^*(x)|$$

ima raspodelu za koju važi

$$\lim_{n_1, n_2 \rightarrow \infty} P\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_{n_1, n_2} \leq t\right) = K(t).$$

Dokaz Smirnovljeve teoreme dat je kod Felera (1948).

Ove teoreme omogućavaju formiranje pojasa poverenja u kojem će se sa unapred zadatom verovatnoćom naći grafik funkcije raspodele obeležja iz kojeg je uzorak izvučen.

Takođe, na osnovu ovih teorema konstruisani su testovi Kolmogorova-Smirnova za jedan i dva uzorka, o čemu će biti reči u poglavlju 5.

4.2. Kontrolne karte

Definicija 4.2. Kontrola kvaliteta je postupak ili niz postupaka čija je svrha da obezbedi da proizvodni proces ili izvršavanje usluge podleže unapred određenim uslovima kvaliteta ili da zadovolji zahteve klijenta (Bakir, 2012).

Neka je X obeležje koje numerički karakteriše proces koji se kontroliše. Statistička kontrola kvaliteta se najčešće vrši kontrolnim kartama.

Neka je $X_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, m$, $m, n \in \mathbb{N}$ serija n uzoraka obima m obeležja X . Broj m je obično mali, jer su proizvodni procesi u jednoj seriji u vremenskom intervalu u kojem se uzorkovanje vrši takvi da proizvedeni proizvodi daju približno iste vrednosti za X .

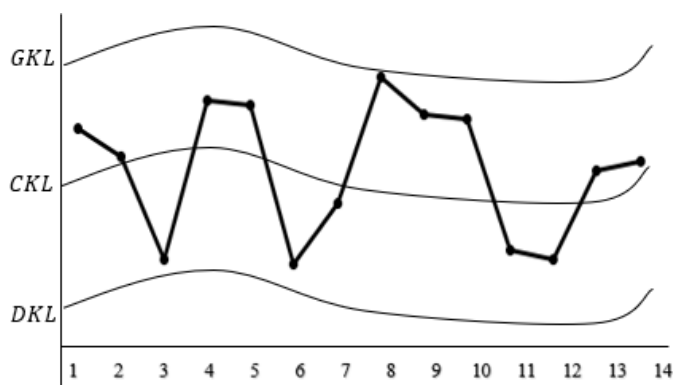
³⁴ Николай Васильевич Смирнов (1900-1966), sovjetski (ruski) matematičar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Definicija 4.3. *Kontrolna statistika K_m je svaka statistika čija raspodela³⁵ dovoljno dobro opisuje proizvodni proces koji u potpunosti (idealno) zadovoljava standarde kvaliteta. **Centralna kontrolna linija (CKL)** je grafik funkcije³⁶ čije vrednosti predstavljaju, za kvalitet kontrolnog procesa, idealan slučaj vrednosti kontrolne statistike. **Gornja i donja kontrolna linija (GKL i DKL)** su grafici funkcija čije su vrednosti granice intervala u kojem K_m treba da bude da bi kvalitet proizvoda u datoj seriji bio pod kontrolom (tj. da bi postavljeni standardi bili zadovoljeni).*

Sve tri kontrolne linije su određene raspedlom kontrolne statistike.

Definicija 4.4. *Kontrolna karta predstavlja statistički dijagram kojim se grafički predstavljaju vrednosti kontrolne statistike u svakom od n uzoraka obima m . Osnovni elementi kontrolne karte su centralna, donja i gornja kontrolna linija i linija promena kontrolne statistike. Videti sliku 4.1.*



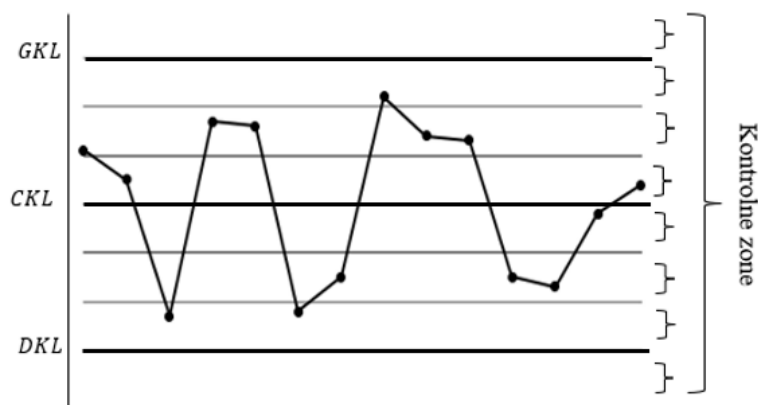
Slika 4.1. Primer izgleda kontrolne karte.

Uslovi zadovoljenja željenog nivoa kvaliteta su varijabilni jer zavise od puno faktora. Samim tim, mogu nastupiti slučajevi u kojima nije dovoljno da linija promena bude unutar kontrolnih linija. Za precizniju kontrolu procesa nekad je korisno formirati dodatne kontrolne linije koje pojas unutar gornje i donje kontrolne linije dele na **kontrolne zone**. Broj promena u svakoj od zona daje više informacija o kvalitetu kontrolnog procesa. Videti sliku 4.2.

³⁵ Najčešće je u pitanju normalna raspodela.

³⁶ Kontrolne linije su uglavnom opisane neprekidnim funkcijama.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.2. Kontrolna karta sa kontrolnim zonama.

Razvijen je veliki broj kontrolnih karata, mada se puno njih zasniva na manjem broju kontrolnih karata aktivno i redovno primenjuje što ukazuje na potrebu za razvojem novih kontrolnih karata koje pružaju pouzdanije rezultate.

Napomena 4.1. U nastavku će pri skiciranju kontrolnih linija oblika $y = f(x)$ u legendama kontrolnih karata konvencionalno biti korišćena samo oznaka $f(x)$.

Definicija 4.5. *Prosečan broj uzoraka pre detekcije defekta procesa*, u oznaci **ARL** (*average run length*), u slučaju procesa pod kontrolom je vrednost

$$ARL_0 = \frac{1}{\alpha}, \quad \alpha = P(K_m \notin [DKL, GKL] \mid \text{Proces je pod kontrolom.}),$$

(α je verovatnoća da u slučaju procesa pod kontrolom, kontrolna karta određena statistikom K_m otkrije defekt). U slučaju procesa koji nije pod kontrolom to je vrednost

$$ARL_\Delta = \frac{1}{1 - \beta}, \quad \beta = P(K_m \in [DKL, GKL] \mid \text{Proces nije pod kontrolom.}),$$

(β je verovatnoća da u slučaju procesa koji nije pod kontrolom kontrolna karta određena statistikom K_m ne otkrije defekt).

4.3. Šuhartova³⁷ \bar{X} kontrolna karta

Neka je $X_{i,j}, i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$ serija n uzoraka obima m obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$ i neka je za svaki od uzoraka $X_{i,1}, \dots, X_{i,m}, i \in \{1, \dots, n\}$ izračunata uzoračka sredina $\bar{X}_i = \frac{1}{m} \sum_{j=1}^m X_{i,j}$, a zatim i $\bar{X} = \frac{1}{n} \sum_{i=1}^n \bar{X}_i$. Neka je

$$\hat{\sigma}_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (X_{i,j} - \bar{X}_i)^2}, \quad i \in \{1, \dots, n\}$$

$$\text{i } \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i.$$

Definicija 4.6. Šuhartova \bar{X} kontrolna karta je kontrolna karta kod koje su

$$CKL: y = \bar{X}, \quad GKL: y = \bar{X} + \frac{3\hat{\sigma}}{\sqrt{n}}, \quad DKL: y = \bar{X} - \frac{3\hat{\sigma}}{\sqrt{n}}$$

Kontrolna statistika je upravo uzoračka sredina uzorka iz serije. Drugim rečima, na kontrolnu kartu se nanose tačke $(i, \bar{X}_i), i = 1, \dots, n$. Dodatno, mogu se formirati kontrolne zone korišćenjem linija $KL_1: y = \bar{X} \pm \frac{\hat{\sigma}}{\sqrt{n}}$ i $KL_2: y = \bar{X} \pm \frac{2\hat{\sigma}}{\sqrt{n}}$, i to

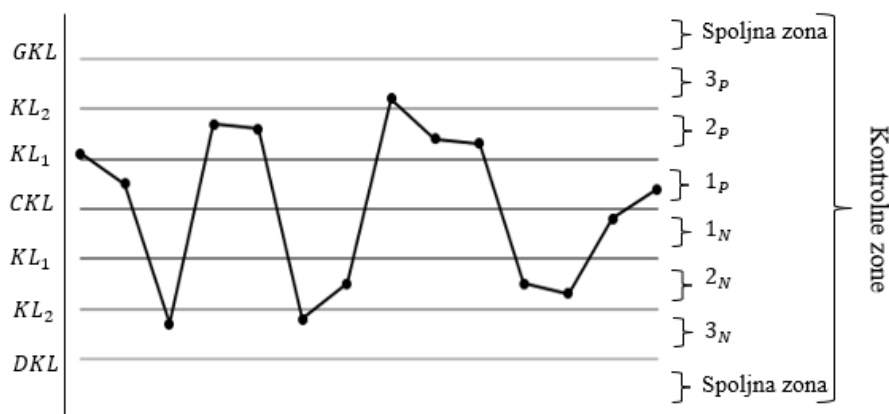
$$\begin{aligned} \text{zona1}_P &= \left(\bar{X}, \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} \right], & \text{zona1}_N &= \left(\bar{X}, \bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} \right], \\ \text{zona2}_P &= \left(\bar{X} + \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + \frac{2\hat{\sigma}}{\sqrt{n}} \right], & \text{zona2}_N &= \left(\bar{X} - \frac{2\hat{\sigma}}{\sqrt{n}}, \bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} \right], \\ \text{zona3}_P &= \left(\bar{X} + \frac{2\hat{\sigma}}{\sqrt{n}}, \bar{X} + \frac{3\hat{\sigma}}{\sqrt{n}} \right], & \text{zona3}_N &= \left(\bar{X} - \frac{3\hat{\sigma}}{\sqrt{n}}, \bar{X} - \frac{2\hat{\sigma}}{\sqrt{n}} \right]. \end{aligned}$$

Na slici 4.3 dat je grafički prikaz kontrolne karte formirane na opisan način.

Umesto broja 3 u $y = \bar{X} \pm \frac{3\hat{\sigma}}{\sqrt{n}}$ može se staviti $z_{\frac{\alpha}{2}}$ za koje je $P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ za željeni prag značajnosti. U izloženom slučaju korišćeno je 3σ pravilo, tj. $\alpha = 0.0027$.

³⁷ Walter Andrew Shewhart (1891-1967), američki fizičar, inženjer i statističar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.3. \bar{X} kontrolna karta.

Bitno je opet napomenuti da je uslov primene ove kontrolne karte normalna raspodela obeležja X , i to sa vrednostima parametara koje odgovaraju postavljenom standardu. Ukoliko nije poznato da li je raspodela saglasna sa pretpostavljenom onda bi ovako postavljena kontrolna karta mogla dati pogrešne rezultate.

Rešenje tog problema može biti zamena \bar{X} sa μ i $\hat{\sigma}$ sa σ (Jevremović & Avdović, 2020b). Ipak, kvalitet procesa je inicijalno upitan ako proizvođač ne može garantovati proces stabilan i u smislu raspodele (Oakland, 2003, str. 26-29).

Pouzdanost rezultata, tj. matematička zasnovanost ove kontrolne karte sledi iz pravila 3σ i centralne granične teoreme. Samim tim, za neznatno preciznije rezultate može se koristiti i prilagođena varijanta pravila 3σ (Oakland, 2003, str. 88-95).

Glavni pokazatelj specijalnog uzroka varijacije je tačka koja predstavlja karakteristiku ocenjenog uzorka, a koja se nalazi van kontrolne granice. Pored ovog pravila, Međunarodna zajednica za standarde usvojila je preporuke po kojima se donosi odluka o tome da li je proces pod kontrolom ili van kontrole (Prokić, 2019, str. 14; Oakland, 2003, str. 125, 158). Osnovna pravila koja su usvojena za detekciju neslučajnog ponašanja procesa dao je Western Electric 1956. godine u Statistical Quality Control Handbook, a to su

- jedna tačka je izvan kontrolnih granica;
- četiri od pet uzastopnih tačaka su izvan zone 1 od centralne linije;
- dve od tri uzastopne tačke su izvan granica upozorenja (zona 2);
- osam uzastopnih tačaka je sa iste strane centralne linije.

Pored ovih pokazatelja postoji još par karakteristika koje ukazuju na specifično ponašanje procesa i na postojanje defekta, i to

- šest tačaka u nizu su u rastućem ili opadajućem poretku;
- petnaest uzastopnih tačaka su koncentrisane oko centralne linije;
- četrnaest tačaka konstantno osciluju gore-dole ;
- osam uzastopnih tačaka na udaljenosti većoj od $\pm 1\sigma$ od centralne linije;

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

- neobična ili neslučajna šema dobijena od tačaka sa podacima;
- jedna ili više tačaka blizu kontrolne granice.

Definicija 4.7. Funkcija zone uzorka X_1, \dots, X_n , $n \in \mathbb{N}$ za normalnu raspodelu $\mathcal{N}(\mu, \sigma^2)$ data je sa

$$\text{zone}(x) = \begin{cases} 1; & \mu - \sigma \leq x < \mu + \sigma \\ 2; & \mu - 2\sigma \leq x < \mu - \sigma \vee \mu + \sigma < x \leq \mu + 2\sigma \\ 3; & \mu - 3\sigma \leq x < \mu - 2\sigma \vee \mu + 2\sigma < x \leq \mu + 3\sigma \\ 3.1947; & x < \mu - 3\sigma \vee x > \mu + 3\sigma \end{cases} \quad x \in \mathbb{R}.$$

Stav 4.3. Ako je $X \sim \mathcal{N}(\mu, \sigma^2)$, onda

$$\text{zone}(X_j): \begin{pmatrix} 1 & 2 & 3 & 3.1947 \\ 0,6826 & 0,2718 & 0,0429 & 0,0027 \end{pmatrix}$$

$$i E(\text{zone}(X_j)) = 1,3635.$$

Dokaz. Dobijanje raspodele obeležja $\text{zone}(X_j)$ čigledno sledi iz pravila 3σ . Izračunavanje očekivanja dato je sa

$$E(\text{zone}(X_j)) = 1 \cdot 0,6826 + 2 \cdot 0,2718 + 3 \cdot 0,0429 + 3,1947 \cdot 0,0027 = 1,3635. \blacksquare$$

Avdović i Jevremović su za \bar{X} kontrolnu kartu definisali vrednost A za procenu odstupanja proizvodnog procesa od onog procesa koji je model stanja pod kontrolom koji je definisan odgovarajućim standardom. Vrednost A data je formulom

$$A = \frac{1}{n} \sum_{i=1}^n \text{zone}(X_i).$$

Napomena 4.2. Izbor vrednosti 3,1947 nije obuhvaćen pravilom 3σ , ali prati isti princip jer važi

$$F^{-1}\left(\frac{F(3) + \lim_{x \rightarrow +\infty} F(x)}{2}\right) = 3,1947$$

kao i

$$F^{-1}\left(\frac{\lim_{x \rightarrow -\infty} F(x) + F(-3)}{2}\right) = -3,1947$$

gde je sa F data funkcija normalne raspodele $\mathcal{N}(0,1)$. Izbor bilo koje druge vrednosti je moguć.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 4.4. Važi $E(A) = 1.3635$.

Dokaz.

$$E(A) = E\left(\frac{1}{n} \sum_{i=1}^n \text{zone}(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(\text{zone}(X_i)) = \frac{1}{n} \cdot n \cdot 1.3405 = 1.3405. \blacksquare$$

Realizacija statistike A može se porediti sa vrednošću 1,3635 kako bi se stekao uvid u odstupanje uzorka koji se kontroliše od stanja pod kontrolom.

Statistika A je dodefinisana u sledećem poglavlju, data je njena raspodela i karakteristike te raspodele. Ovaj oblik statistike A može biti korišćen u poređenju \bar{X} kontrolne karte sa kontrolnim kartama kvantil-zona i Q-Q kartom o čemu će biti reči u narednim odeljcima.

Statistika T data formulom

$$T = \frac{1}{n} \sum_{i=1}^n I(\text{zone}(X_i) = 3,1947)$$

(gde je I indikator događaja) može biti od značaja jer računa relativnu frekvenciju broja uzoraka sa defektom, tj. koji odstupaju od standardom zadatog kvaliteta. Pomnožena sa 100% daje frekvenciju iste karakteristike u procentima. Time ova statistika daje dodatne informacije o procesu koji se kontroliše.

Uzevši u obzir da je kod Šuhartove \bar{X} kontrolne karte korišćeno pravilo 3σ , verovatnoća da se tačka kontrolne linije, za proces pod kontrolom, nađe van kontrolnih granica je $\alpha = 0,0027$. Odatle je $ARL_0 = \frac{1}{\alpha} \approx 370$.

4.4. Druge kontrolne karte

R karta

R karta ili kontrolna karta za raspon varijacije je kontrolna karta čija je konstrukcija slična principu konstrukcije \bar{X} karte. CKL je ovde prava $y = \bar{R}$ gde je $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$ sa $R_i = \max_{j=1, \dots, m} X_{i,j} - \min_{j=1, \dots, m} X_{i,j}$, $i = 1, \dots, n$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kontrolne granice GKL i DKL su redom $y = B_3\bar{R}$ i $y = B_4\bar{R}$ gde su koeficijenti B zbog jednostavnosti korišćenja izračunati i dati u tabliama. Primer tablica se može naći kod Prokić (2019, str. 23) i Minić³⁸ (2018, str. 116).

Kontrolna statistika je raspon uzorka iz serije, tj. na kontrolnu kartu se nanose tačke $(i, R_i), i = 1, \dots, n$.

R karta se obično koristi za uzorke malih obima ($m \leq 12$).

S karta

S karta se konstruiše slično kao prethodne sa CKL $y = \bar{S}$, gde je $\bar{S} = \frac{1}{n}\sum_{i=1}^n S_i$ za

$$S_i = \sqrt{\frac{1}{m}\sum_{j=1}^m (X_{i,j} - \bar{X}_i)^2}, \quad i = 1, \dots, n$$

dok su GKL i DKL redom $y = D_3\bar{S}$ i $y = D_4\bar{S}$ gde su koeficijenti D zbog jednostavnosti korišćenja izračunati i dati u tablicama. Primer tablica se može naći kod Prokić (2019, str. 23) i Minić³⁹ (2018, str. 116).

Kontrolna statistika je standardna devijacija uzorka iz serije, tj. na kontrolnu kartu se nanose tačke $(i, S_i), i = 1, \dots, n$.

S karta se obično koristi za uzorke većih obima ($m > 15$).

Pored R i S karti po uzoru na \bar{X} kartu se formiraju i mnoge druge. Primeri tih karata su kontrolna karta za broj neispravnih proizvoda, kontrolna karta za proporciju neispravnih proizvoda, kontrolna karta za broj nedostataka na elementima kontrole i kontrolna karta za broj nedostataka po jedinici mere.

Pored kontrolnih karata po uzoru na Šuhartovu \bar{X} kartu postoje i druge kontrolne karte koje koriste drugačije principe konstrukcije mada se sve najčešće svode na korišćenje osobina normalne raspodele. Neke od tih karata su CUSUM karta, EWMA karta i Frešeova karta (Minić, 2018, str. 108-115).

³⁸ Ovde su koeficijenti označeni sa B_1 i B_2 .

³⁹ Ovde su koeficijenti označeni sa C_1 i C_2 .

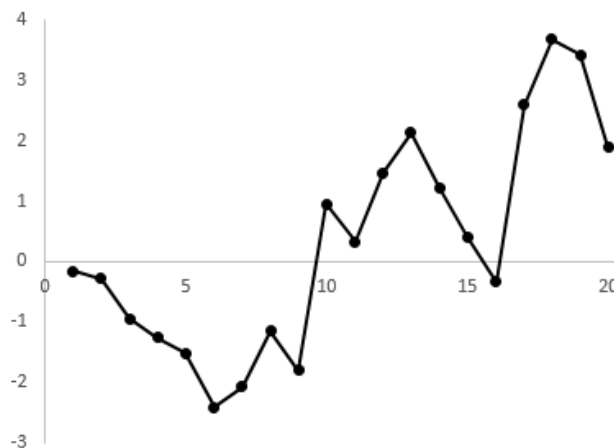
CUSUM kontrolna karta – Kontrolna karta kumulativnih suma

Nedostatak kontrolnih karata Šuhartovog tipa jeste što one često ne detektuju male promene u očekivanju kontrolisane veličine. Pa kada se pri kontrolisanju procesa izostave uslovi detekcije određenog broja tačaka sa jedne strane centralne linije, može se desiti da je proces sa promenjenim očekivanjem i dalje unutar kontrolnih linija kontrolne karte Šuhartovog tipa (Montgomery, 2009, str. 400). Jedna od metoda kojim se ovaj nedostatak može nadomestiti bez uvođenja novih uslova kontrole jesu kumulativne sume, tj. CUSUM šema uvedena od strane Pejdzha (1954). Naime, CUSUM šema uzima u obzir više informacija tako što se za svaki uzorak računa kumulativna suma razlika prosečnih vrednosti i vrednosti očekivanja raspodele u odnosu na koju se proces kontroliše, i to do uzorka koji se u referentnom računanju razmatra. U kontekstu ove metode se očekivanje raspodele kontrole naziva ciljnom vrednošću (target value).

Definicija 4.8. Neka su date iste postavke kao kod Šuhartove \bar{X} kontrolne karte. **CUSUM šemu** čine tačke oblika (i, C_i) , $i = 1, \dots, n$ gde su

$$C_i = \sum_{j=1}^i (\bar{X}_j - \mu), \quad i = 1, \dots, n.$$

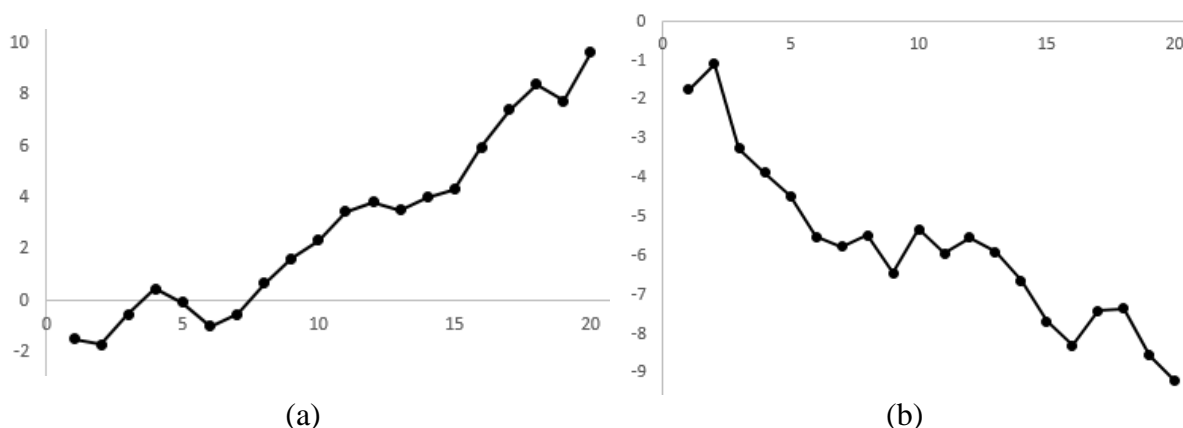
Vrednost C_i se naziva **kumulativnom sumom odstupanja do i -tog uzorka** (uključujući i -ti uzorak).



Slika 4.4. Kumulativne sume uzoraka.

Onda kada je očekivanje raspodele obeležja koje karakteriše proces koji se kontroliše povećano, linija koja povezuje tačke (i, C_i) , $i = 1, \dots, n$ ima rastući trend sa većinom tačaka iznad x-ose. U slučaju da očekivanje odgovara ciljnoj vrednosti pomenute tačke se ravnomerno raspodeljuju sa obe strane prave $y = 0$. U slučaju smanjene vrednosti očekivanja linija koja povezuje te tačke će imati opadajući trend sa većinom tačaka ispod x-ose.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.5. Kumulativne sume uzoraka iz raspodele sa (a) povećanim očekivanjem obeležja kontrole; (b) smanjenim očekivanjem obeležja kontrole.

Za ovu svrhu ova šema je naročito korisna za uzorke obima 1, tj. može poslužiti kao preliminarna analiza testiranja hipoteze $H_0(E(X) = \mu)$. Rezultati su evidentni i za male promene vrednosti očekivanja (Montgomery, 2009, str. 401).

CUSUM šema se kao takva ne može koristiti kao kontrolna karta jer nema definisane kontrolne granice. Ipak odgovarajući uslovi detekcije defekata koji ukazuju da proces nije pod kontrolom se mogu definisati i tako formirati CUSUM kontrolna karta. Postoje dva načina formiranja CUSUM kontrolne karte, a to su Tabelarna, tj. algoritamska CUSUM karta i CUSUM karta sa V-maskom (Barnard, 1959).

Definicija 4.9. Tabelarnu, tj. algoritamsku CUSUM kontrolnu kartu čine dve izlomljene linije koje nastaju spajanjem, za prvu liniju tačaka oblika (i, C_i^+) , $i = 1, \dots, n$, a za drugu liniju tačaka oblika (i, C_i^-) , $i = 1, \dots, n$, gde su

$$C_i^+ = \max\{0, \bar{X}_i - (\mu + K) + C_{i-1}^+\}, \quad C_0^+ = 0,$$

$$C_i^- = \max\{0, (\mu - K) - \bar{X}_i + C_{i-1}^-\}, \quad C_0^- = 0;$$

K je tzv. referentna vrednost ili dozvoljeno odstupanje, a najčešće se bira

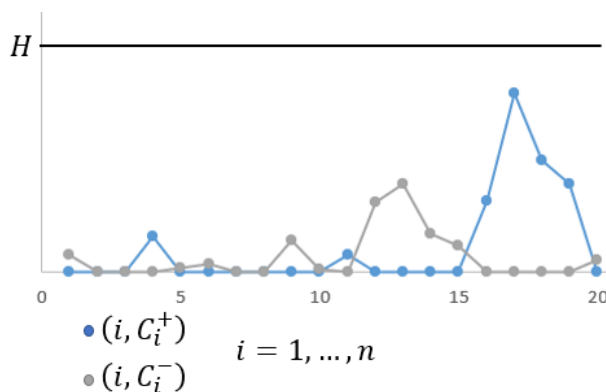
$$K = \frac{|(\mu + \delta\sigma) - \mu|}{2} = \frac{\delta\sigma}{2},$$

gde je δ promena vrednosti očekivanja koju treba detektovati (tačnije, koja se želi detektovati).

Ako se bilo koji od događaja „ $C_i^+ > H$ “, tj. „tačka (i, C_i^+) je iznad prave $y = H$ “ i „ $C_i^- > H$ “, tj. „tačka (i, C_i^-) je iznad prave $y = H$ “, za bar jedno $i = 1, \dots, n$, onda posmatrani proces nije pod kontrolom. Najčešće se bira $H = 5\sigma$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Vrednosti C_i^+ i C_i^- se redom nazivaju **jednostrana gornja i jednostrana donja kumulativna suma odstupanja**. Parametar H naziva se **parametrom odluke**. Parametri H i K , tj. δ , se mogu birati shodno potrebi istraživanja, standarda kvaliteta, zahtevima klijenata i sl.; a mogu se birati i tako da optimizuju vrednosti ARL_0 i ARL_Δ .



Slika 4.6. CUSUM kontrolna karta

Preporučene vrednosti parametara daju se kao k i h za $K = k\sigma$ i $H = h\sigma$. Vrednosti $k = \frac{1}{2}$, $h = 4$ ili $h = 5$ rezultiraju najboljim performansama ove kontrolne karte, tj. optimalnim vrednostima za ARL-ove.

Drugi način definisanja CUSUM kontrolne karte, tj. uslova za detekciju defekata, je korišćenjem tzv. **V-mask** uvedene od strane Barnarda (1959), ili njene modifikovane varijante koju je predložio Lukas (1973). Iako je ovaj način kontrolisanja procesa grafički ilustrativniji, komplikovan je za formiranje i manje precizan pa često biva korišćen u svrhe ilustracije umesto u praktičnoj primeni. Stoga će ovde biti izostavljen. Postupak formiranja CUSUM karte sa V-maskom, nedostaci primene, veze sa tabelarnom varijantom karte i primeri za oba slučaja mogu se naći kod Montgomerija (2009, str. 400-417) i Ouklenda (2003, str. 227-242).

EWMA kontrolna karta

Kontrolna karta pokretnih proseka sa eksponencijalnim opterećenjem (exponentially weighted moving average) je još jedna kontrolna karta koja može poslužiti kao alternativa Šuhartovoj \bar{X} kontrolnoj karti. Predložena je od strane Robertsa (1959). Performanse ove karte su približne performansama CUSUM kontrolne karte, a i ova kontrolna karta je posebno pogodna za uzorke obima 1, tj. za individualna posmatranja, mada je primenljiva i za uzorke većeg obima.

Neka su pretpostavke iste kao u slučaju Šuhartove \bar{X} kontrolne karte.

Definicija 4.10. *EWMA kontrolna karta je kontrolna karta čije su gornja kontrolna linija, centralna linija i donja kontrolna linija redom date sa*

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$GKL: y = \mu + L\sigma \sqrt{\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2j})},$$

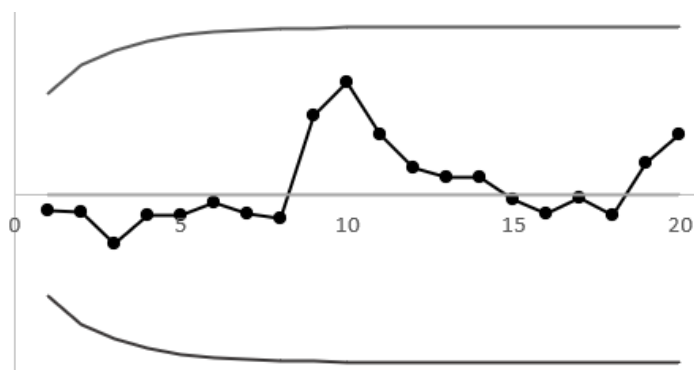
$$CL: y = \mu,$$

$$DKL: y = \mu + L\sigma \sqrt{\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2j})},$$

gde je L širina kontrolnih granica, a $\lambda \in (0, 1]$ parametar opterećenja. Kontrolna statistika ove kontrolne karte je statistika proseka sa eksponencijalno opterećenim promenama datim sa

$$Z_i = \lambda \bar{X}_i + (1 - \lambda)Z_{i-1}, \quad i = 1, \dots, n$$

za $Z_0 = \mu$. Linija promena dobija se povezivanjem tačaka oblika (i, Z_i) , $i = 1, \dots, n$.



Slika 4.7. EWMA kontrolna karta.

Parametri L i λ se analogno parametrima CUSUM karte biraju shodno potrebama kontrole koje se mogu nametnuti ili tako da performanse karte budu bolje. Optimalni izbori parametara obično podrazumevaju izbor $\lambda \in [0,05, 0,25]$, a vrednosti $\lambda = 0.05$, $\lambda = 0.1$ i $\lambda = 0.2$ su popularni izbori. Obično se koriste manje vrednosti parametra λ pri detekciji manjih promena u vrednosti proseka raspodele u odnosu na koju se proces kontroliše. Vrednosti L obično podrazumevaju vrednost blisku broju 3 kada je λ veće, a vrednost L se smanjuje za manje λ (Montgomery, 2009, str. 423).

4.5. Kontrolne karte – novi pristupi

U okviru dosadašnjih istraživanja u oblasti kontrole kvaliteta autor je predložio i publikovao nekoliko novih kontrolnih karata. Njihove osobine su detaljno analizirane kako pojedinačno, tako i u poređenju sa postojećim kartama koje se koriste u praksi.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kod kontrolnih karata koje su date u nastavku teksta, ponuđene kontrolne granice su ilustrativnog karaktera. Optimalan izbor kontrolnih granica diskutovan je u analizi poređenja performansi kontrolnih karata na kraju poglavlja.

EFR karta

EFR karta je karta bazirana na empirijskoj funkciji raspodele (Jevremović & Avdović, 2020b). Ideja potiče iz teoreme Kolmogorova (Teorema 4.2). Naime, za zadato $\alpha \in (0,1)$, ako je neki uzorak X_1, \dots, X_n izvučen iz raspodele $F(x)$, $x \in \mathbb{R}$, onda će za empirijsku funkciju raspodele, kada statistika D_n ne zavisi od parametara raspodele zadate sa F^{40} , važiti

$$F(x) - d_{n,\alpha} \leq F_n^*(x) \leq F(x) + d_{n,\alpha} \quad (4.1)$$

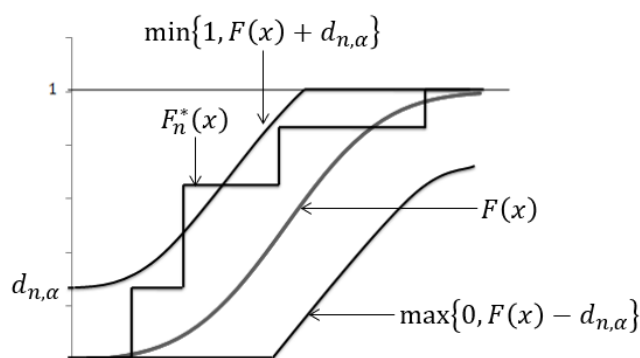
sa verovatnoćom $1 - \alpha$, gde je $P(D_n \geq d_{n,\alpha}) = \alpha$ za D_n dato Teoremom 2.2, tj. D_n ima raspodelu Kolmogorova (Avdović & Jevremović, 2020b). Tada, na osnovu definicije empirijske funkcije raspodele i nejednakosti (4.1) sledi

$$\max\{0, F(x) - d_{n,\alpha}\} \leq F_n^*(x) \leq \min\{1, F(x) + d_{n,\alpha}\}, \quad x \in \mathbb{R}.$$

Drugim rečima, empirijska funkcija raspodele se nalazi u pojasu

$$[\max\{0, F(x) - d_{n,\alpha}\}, \min\{1, F(x) + d_{n,\alpha}\}], \quad x \in \mathbb{R} \quad (4.2)$$

(videti sliku 4.8).



Slika 4.8. Empirijska funkcija raspodele u odnosu na pojas (4.2) – EFR karta.

⁴⁰ Raspodela statistike pri istoj raspodeli obeležja X se u slučaju pojedinih raspodela (koje ovde neće biti razmatrane) može menjati ako se promene vrednosti parametara (Babu & Rao, 2004).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Formiranje datog pojasa i primena u slici 4.8 je očigledno izvedena iz test-statistike testa Kolmogorova. Stoga, grafička primena dobijenog pojasa može biti korišćena kao preliminarna metoda saglasnosti raspodele obeležja iz kojeg je uzorak izvučen sa nekom zatom raspodelom datom sa $F(x)$. Kako test Kolmogorova odbacuje nultu hipotezu sa pragom značajnosti α ako je d_n (realizovana vrednost test-statistike) veća od kritične vrednosti $d_{n,\alpha}$, to će se isti zaključak doneti ako bar jedna od vrednosti $F_n^*(X_i)$, $i = 1, \dots, n$ nije u pojasu (4.2). Grafički, nulta hipoteza se odbacuje sa pragom značajnosti α ako je bar jedan levi kraj horizontalnih linija grafika empirijske funkcije raspodele izvan grafika granica pojasa (4.2).

Može se formirati statistika

$$T = 1 - \frac{1}{n} \sum_{i=1}^n I(\max\{0, F(X_i) - d_{n,\alpha}\} \leq F_n^*(X_i) \leq \min\{1, F(X_i) + d_{n,\alpha}\})$$

koja meri stepen odstupanja od nulte hipoteze. Tumačenje statistike T bi moglo biti dato sa „ $T \cdot 100\%$ vrednosti uzorka uzrokuje odbacivanje nulte hipoteze sa pragom značajnosti α “. Ovo može biti značajno pri testiranju normalnosti na uzorcima velikog obima gde je moć testa veća i najmanja odstupanja čak i pojedinačnih elemenata uzorka od modela nulte hipoteze mogu dovesti do toga da test odbaci nultu hipotezu. Ako bi od značaja bilo kontrolisati takve slučaje do unapred zatomog, za istraživanje prihvatljivog, procenta, onda bi vrednost T mogla poslužiti svrsi.

Primer 4.1. Neka je dat uzorak 3,04; 3,12; 3,12; 3,22. Razmatra se empirijska funkcija raspodele datog uzorka u odnosu na pojas (4.2) određen funkcijom $F(x)$ normalne raspodele $\mathcal{N}(3,2; 0,06^2)$. Istovremeno, biće dat zaključak o testiranju hipoteze $H_0: X \sim \mathcal{N}(3,2; 0,06^2)$ sa pragom značajnosti $\alpha = 0,05$, gde je X obeležje iz kojeg je uzorak izvučen.

Empirijska funkcija datog uzorka je

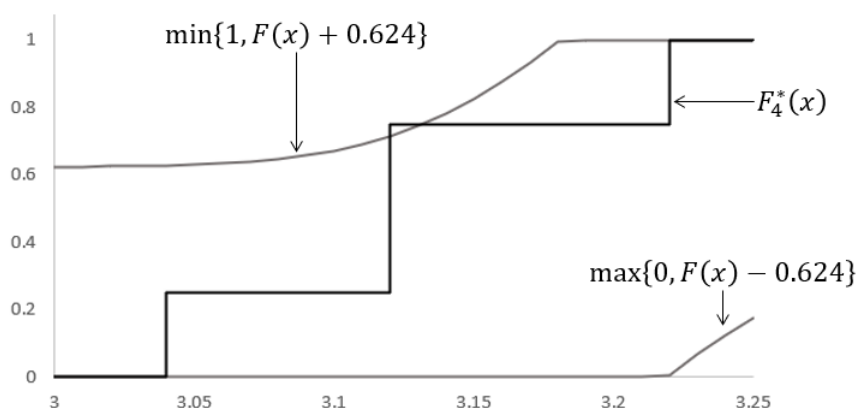
$$F_4^*(x) = \begin{cases} 0; & x < 3,04 \\ 0,25; & 3,04 \leq x < 3,12 \\ 0,75; & 3,12 \leq x < 3,22 \\ 1; & x \geq 3,22 \end{cases} \quad (4.3)$$

Pojas (4.2) dat je sa

$$\begin{aligned} & [\max\{0, F(x) - d_{4;0,05}\}, \min\{1, F(x) + d_{4;0,05}\}] \\ & = [\max\{0, F(x) - 0.624\}, \min\{1, F(x) + 0.624\}], \quad x \in \mathbb{R}. \end{aligned} \quad (4.4)$$

Interpretacija slike 4.8 u ovom primeru data je sledećom slikom.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.9. Empirijska funkcija raspodele (4.3) u odnosu na pojas (4.4).

Kako se cela empirijska funkcija raspodele ne nalazi unutar pojasa (4.4), to test Kolmogorova sa pragom značajnosti α , odbacuje hipotezu $H_0: X \sim \mathcal{N}(3,2; 0,06^2)$. Tačka $(3.12, F_4^*(3.12))$, tj. levi kraj treće horizontalne linije na grafiku empirijske funkcije raspodele se nalazi iznad granice $\max\{0, F(x) - d_{4;0,05}\}$ pojasa (4.4).

Izračunavanjem vrednosti statistike T dobija se $T = 0,25$, pa 25% uzorka dovodi do odbacivanja hipoteze normalnosti raspodele sa pragom značajnosti $\alpha = 0,05$. ■

Neka je $X_{i,j}, i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$ serija n uzoraka obima m obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$ i neka je za svaki od uzoraka $X_{i,1}, \dots, X_{i,m}, i \in \{1, \dots, n\}$ izračunata uzoračka sredina

$$\bar{X}_i = \frac{1}{m} \sum_{j=1}^m X_{i,j}, \quad i = 1, \dots, n. \quad (4.5)$$

Na osnovu centralne granične teoreme, uzorak (4.5) je izvučen iz obeležja $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Neka je F funkcija raspodele tog obeležja i F_n^* empirijska funkcija raspodele uzorka dobijenih uzoračkih sredina.

Definicija 4.11. *EFR karta je kontrolna karta čija je linija promena dobijena spajanjem tačaka $(\bar{X}_i, F_n^*(\bar{X}_i)), i = 1, \dots, n$, donja kontrolna linija je grafik $\max\{0, F(x) - d_{n,\alpha}\}$ kao, a gornja kontrolna linija je grafik $\min\{1, F(x) + d_{n,\alpha}\}$.*

Umesto linije promena može se koristiti i grafik empirijske funkcije raspodele gde se posmatraju levi krajevi horizontalnih linija, tj. tačke promena. Slika 4.8 ilustruje i EFR kartu. Grafik funkcije raspodele $\mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right)$ je centralna kontrolna linija.

Tumačenje EFR karte u odnosu na \bar{X} kartu se menja shodno činjenici da EFR nema isti broj kontrolnih zona, tj. kontrolne linije između DKL, CKL i GKL. Naime, postojanje defekta

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

može biti proglašeno ako važi bar jedna od sledećih tvrdnji (po analogiji sa pravilima Western Electric-a iz 1956.).

- jedna tačka je izvan kontrolnih granica;
- osam uzastopnih tačaka je sa jedne strane centralne linije.
- šest tačaka u nizu su u rastućem ili opadajućem položaju;
- četrnaest tačaka konstantno osciluju gore-dole ;
- neobična ili neslučajna šema dobijena od tačaka sa podacima;
- jedna ili više tačaka blizu kontrolne granice.

Iako manjak zona kod kontrolnih karata ukazuje na manju senzitivnost detekcije defekta, u ovom slučaju je senzitivnost EFR karte bolja od svih prethodno navedenih karti. Ilustracija toga data je sledećim primerom.

Statistika T primenjena nad kontrolnom kartom može izbrojati koliko uzoraka, tj. koji procenat uzoraka, poseduje defekt.

Primer 4.2. Mereno je procentualno učešće proteina u pakovanjima mleka od 1000ml. Dobijeni su sledeći rezultati

Tabela 4.1. Serije uzoraka merenja učešća proteina u mleku.

Redni broj serije	Elementi uzorka svake od serija			
1	3,04	3,12	3,12	3,22
2	3,09	3,13	3,21	3,18
3	3,10	3,18	3,21	3,18
4	3,04	3,11	3,17	3,06
5	3,13	3,12	3,11	3,07
6	3,15	3,05	3,14	3,18
7	3,11	3,21	3,22	3,13
8	3,06	3,07	3,17	3,22
9	3,05	3,19	3,18	3,20
10	3,08	3,20	3,21	3,09
11	3,05	3,14	3,22	3,08
12	3,19	3,18	3,21	3,06

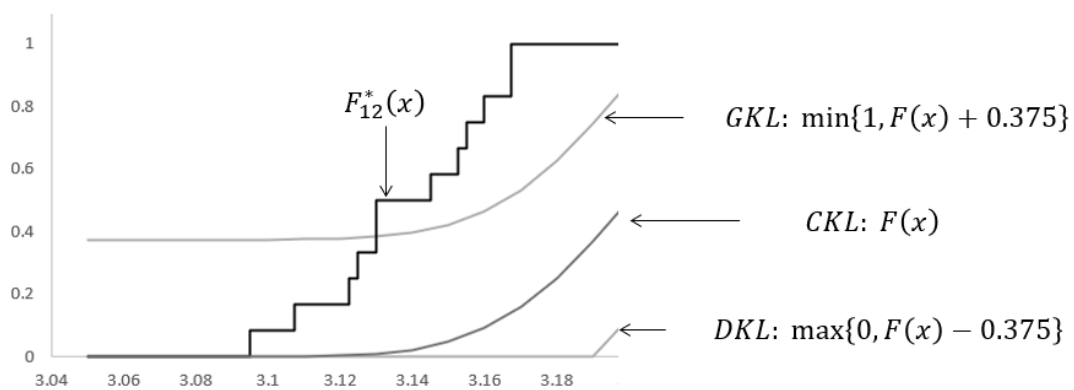
Postavljen je standard da ako je proizvodni proces pod kontrolom onda je procentualno učešće proteina u mleku rasodeljeno po normalnoj raspodeli $\mathcal{N}(3,2; 0,06^2)$. U ovom primeru ilustrovana je kontrola proizvodnog procesa mleka primenom EFR karte.

Izračunavanjem uzoračke sredine svakog od uzoraka redom dobija se uzorak 3,1250; 3,1525; 3,1675; 3,0950; 3,1075; 3,1300; 3,1675; 3,1300; 3,1550; 3,1450; 3,1225; 3,1600, čija je empirijska funkcija raspodele

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$F_{12}^*(x) = \begin{cases} 0; & x < 3,095 \\ 0,08; & 3,095 \leq x < 3,1075 \\ 0,17; & 3,1075 \leq x < 3,1225. \\ \vdots & \vdots \\ 1; & x \geq 3,1675 \end{cases}$$

Za $\alpha = 0,05$ dobija se $d_{12;0,05} = 0,375$ a $F(x)$ je funkcija normalne raspodele $\mathcal{N}\left(3,2; \frac{0,06^2}{4}\right)$. Skiciranjem kontrolnih granica, tj. grafika krajeva pojasa (4.2) i grafika $F_{12}^*(x)$ dobija se sledeća kontrolna karta.



Slika 4.10. EFR kontrolna karta za podatke date tabelom 4.1.

Kako empirijska funkcija raspodele, a samim tim i linija promena, nije cela unutar pojasa, to kontrolisani proizvodni proces nije pod kontrolom. Izračunavanjem T dobija se $T = 0.6667$ pa 66,67% uzorka ukazuje na defekt u kvalitetu proizvodnog procesa. ■

Napomena 4.3. U nastavku će se skiciranje kontrolnih linija određenih empirijskom funkcijom raspodele vršiti spajanjem tačaka čija pozicija se na kontrolnoj karti kontroliše.

Kako važe (4.1) i (4.2), to se vrednost $d_{n,\alpha}$, tj. kontrolne granice, mogu podesiti tako da se dobije isti ARL_0 kao u slučaju Šuhartove \bar{X} kontrolne karte. Za $\alpha = 0,0027$ izbor odgovarajućeg $d_{n,\alpha}$ zavisice od n . Prednost ove kontrolne karte jeste što korisnik može izabrati α , tj. ARL, zatim jednostavno odrediti granicu koja mu za kontrolisanje treba. Optimalan izbor parametara, tj. granica ove kontrolne karte diskutovan je u poslednjem odeljku ovog poglavlja.

Kontrolna karta kvantil-zona

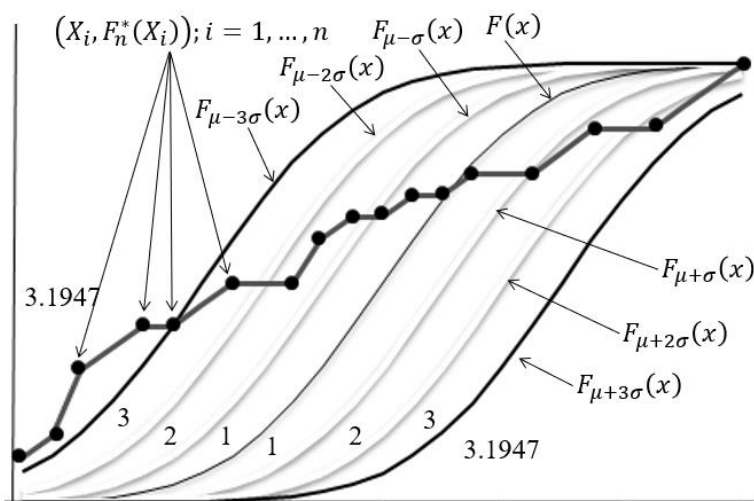
Ova kontrolna karta je kao i EFR karta zasnovana na empirijskoj funkciji raspodele. Razlika je u tome što se pri konstrukciji kontrolne karte kvantil-zona koristi analogija sa pravilom 3σ . Naime, ako je uzorak $X_1, \dots, X_n, n \in \mathbb{N}$ izvučen iz obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$, $F_n^*(x)$, $x \in \mathbb{R}$ njegova empirijska funkcija raspodele i ako su funkcije $F(x)$, $F_{\mu \pm i\sigma}(x)$, $x \in \mathbb{R}$ redom funkcije raspodela $\mathcal{N}(\mu, \sigma^2)$, $\mathcal{N}(\mu \pm i\sigma, \sigma^2)$ za $i = 1, 2, 3$, onda će važiti

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$F_{\mu+3\sigma}(x) < F_{\mu+2\sigma}(x) < F_{\mu+\sigma}(x) < F(x) < F_{\mu-\sigma}(x) < F_{\mu-2\sigma}(x) < F_{\mu-3\sigma}(x) \quad (4.6)$$

za svako $x \in \mathbb{R}$, a na osnovu stava 4.2 $F_n^*(x) \xrightarrow{P} F(x)$ ($n \rightarrow +\infty$), pa će (4.6) sa određenom preciznošću biti zadovoljeno i kada je F_n^* umesto F .

Nejednakost (4.6) može biti korišćenja za uvođenje još jedne metode preliminarne analize normalnosti. Kada se elementi nejednakosti (4.6) skiciraju zajedno sa tačkama $(X_j, F_n^*(X_j))$, $j = 1, \dots, n$, dobija se sledeći dijagram.



Slika 4.11. Dijagram preliminarne analize normalnosti korišćenjem kvantil-zona.

Slika 4.11 ukazuje da se može definisati sledeća funkcija.

Definicija 4.12. Funkcija kvantil-zone uzorka X_1, \dots, X_n , $n \in \mathbb{N}$ data je formulom

$$q_{zone}(x) = \begin{cases} 1; & F_{\mu+\sigma}(x) \leq F_n^*(x) < F(x) \vee F(x) < F_n^*(x) \leq F_{\mu-\sigma}(x) \\ 2; & F_{\mu+2\sigma}(x) \leq F_n^*(x) < F_{\mu+\sigma}(x) \vee F_{\mu-\sigma}(x) < F_n^*(x) \leq F_{\mu-2\sigma}(x) \\ 3; & F_{\mu+3\sigma}(x) \leq F_n^*(x) < F_{\mu+2\sigma}(x) \vee F_{\mu-2\sigma}(x) < F_n^*(x) \leq F_{\mu-3\sigma}(x) \\ 3,1947; & F_n^*(x) < F_{\mu+3\sigma}(x) \vee F_n^*(x) > F_{\mu-3\sigma}(x) \end{cases} \quad x \in \mathbb{R}.$$

Statistika

$$V = \frac{1}{n} \sum_{i=1}^n q_{zone}(X_i),$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

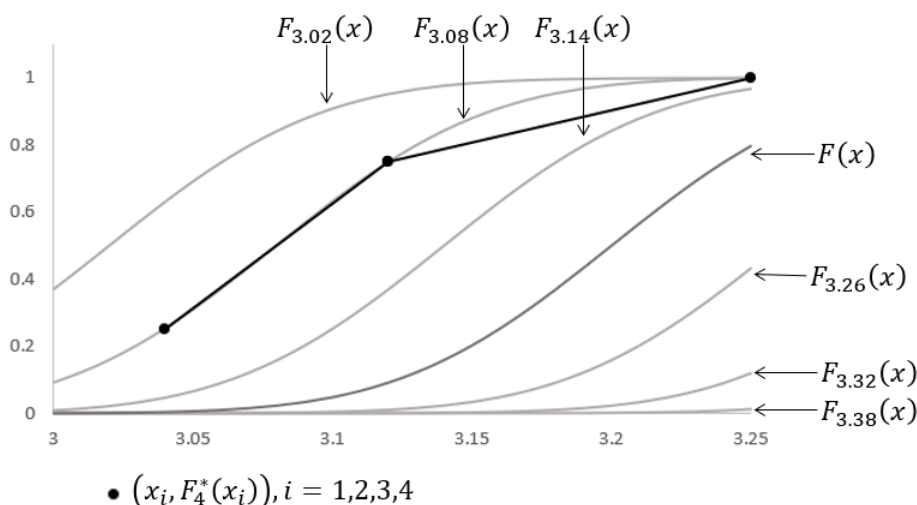
očigledno predstavlja procenu saglasnosti obeležja polaznog uzorka sa normalnom raspodelom.

Raspodela statistike V je od interesa jer se može iskoristiti za testiranje normalnosti, što je tema sledećeg poglavlja, gde je data iscrpnija analiza statistike V i njene primene.

Konkretno, bitno je naglasiti da iako eksplicitnu vrednost $E(V)$ ili $D(V)$ nije moguće odrediti, simulacijama se može videti da je za bilo koji obim uzorka $E(V) < 1,5$. Štaviše, ta vrednost se sve manje razlikuje od 1, kako se n povećava (videti tabele 5.6, 5.7, 5.8 i 5.9). Takođe, standardna devijacija je manja od 0,5 za najmanje obime uzorka, a smanjuje se kako se n povećava. Isto važi za slučaj u kojem se parametri ocenjuju. Ova informacija je dovoljna za preliminarnu analizu normalnosti ili kontrole kvaliteta.

Primena preliminarne analize normalnosti primenom dijagrama 7.7 i statistike V je ilustrovana sledećim primerom.

Primer 4.3. Nad podacima iz primera 4.1 i sa istom raspodelom sa kojom se upoređuje saglasnost raspodele uzorka formiran je dijagram dat slikom 4.11 i dobijen je sledeći grafički rezultat.



Slika 4.12. Dijagram preliminarne analize normalnosti korišćenjem kvantil-zona dobijen obradom podataka primera 4.1.

Najbolja saglasnost se dobija kada se tačke oblika $(x_i, F_4^*(x_i)), i = 1,2,3,4$, nalaze unutar pojasa $[F_{3,14}(x), F_{3,26}(x)]$, $x \in \mathbb{R}$ i to sa obe strane grafika funkcije $F(x)$, $x \in \mathbb{R}$. Na dijagramu datom slikom 4.12 vidi se da za posmatrane podatke to nije slučaj.

Računanjem statistike V dobija se

$$V = \frac{1}{4}(q_{zone}(3.04) + 2 \cdot q_{zone}(3.12) + q_{zone}(3.22)) = \frac{1}{3}(2 + 2 \cdot 2 + 3,1947) = 2,2987,$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

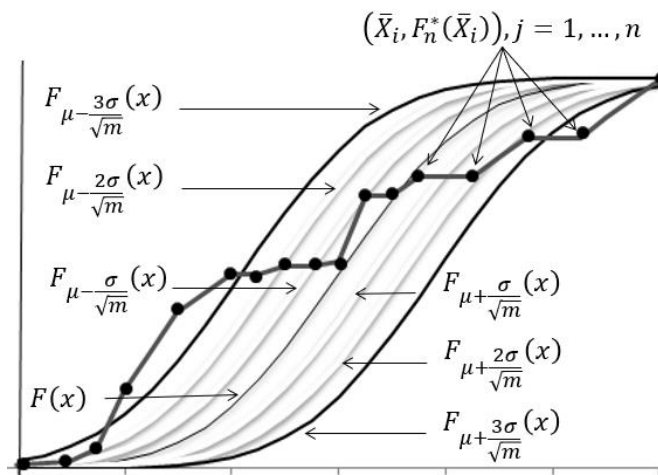
pa preliminarna analiza normalnosti ukazuje na odstupanje raspodele obeležja uzorka od posmatrane normalne raspodele. ■

Neka je $X_{i,j}, i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$ serija n uzoraka obima m obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$ i neka je za svaki od uzoraka $X_{i,1}, \dots, X_{i,m}, i \in \{1, \dots, n\}$ izračunata uzoračka sredina data po formuli (4.5) za svako $i = 1, \dots, n$. Na osnovu centralne granične teoreme, uzorak (4.5) je izvučen iz obeležja $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right)$. Prilagođavanjem (4.6) dobija se

$$F_{\mu+\frac{3\sigma}{\sqrt{m}}}(x) < F_{\mu+\frac{2\sigma}{\sqrt{m}}}(x) < F_{\mu+\frac{\sigma}{\sqrt{m}}}(x) < F(x) < F_{\mu-\frac{\sigma}{\sqrt{m}}}(x) < F_{\mu-\frac{2\sigma}{\sqrt{m}}}(x) < F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x) \quad (4.7)$$

gde je u ovom slučaju $F(x), x \in \mathbb{R}$ funkcija raspodele obeležja \bar{X} .

Na osnovu stava 4.2 $F_n^*(x) \xrightarrow{P} F(x) (n \rightarrow +\infty)$, pa (4.7) sa određenom preciznošću treba da bude zadovoljeno i kada je F_n^* umesto F . Skicirajući granice pojaseva određenih nejednakošću (4.7) i liniju promena dobijenu povezivanjem tačaka $(\bar{X}_i, F_n^*(\bar{X}_i)), i = 1, \dots, n$ dobija se sledeći dijagram.



Slika 4.13. Kontrolna karta kvantil-zona.

Definicija 4.13. Kontrolna karta kvantil-zona je kontrolna karta čija je linija promena dobijena spajanjem tačaka $(\bar{X}_i, F_n^*(\bar{X}_i)), i = 1, \dots, n$, donja kontrolna linija je grafik $F_{\mu+\frac{3\sigma}{\sqrt{m}}}(x)$, a gornja kontrolna linija je grafik $F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)$ (Jevremović & Avdović, 2020a).

Kontrolnoj karti kvantil-zona mogu se dodeliti kontrolne zone koje su određene kontrolnim granicama, kontrolnom linijom F i graficima funkcija $F_{\mu \pm \frac{j\sigma}{\sqrt{m}}}(x), j = 1, 2$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Analogno slučaju sa preliminarnom analizom normalnosti statistika V može se koristiti kao procena odstupanja posmatranog proizvodnog procesa od modela pod kontrolom, i to na isti način. U primeru 4.4 mogućnost primene je ilustrovana.

Tumačenje kontrolne karte kvantil-zona ostvaruje se na isti način kao kod \bar{X} karte.

Statistika T u slučaju ove kontrolne karte data je formulom

$$T = \frac{1}{n} \sum_{i=1}^n I(q_{zone}(\bar{X}_i) = 3,1947).$$

Koristi se analogno kao kod prethodno navedenih karata, tj. grafičkih metoda preliminarne analize normalnosti.

Ono što je otežavajuća okolnost kod ove kontrolne karte jeste ta što se moraju uvesti izvesna ograničenja da bi rezultati bili precizni. Naime, kod maksimalne vrednosti u uzorku uvek važi $q_{zone}(\bar{Y}_i) = 3,1947, i = l + 1, \dots, n$, tj. vrednost kontrolne statistike za maksimalni element uzorka je uvek iznad GKL. Za male obime uzorka, broj vrednosti kontrolne statistike iznad GKL, ili ispod DKL može biti i veći od 1. Iako se u kontroli kvaliteta pretežno radi sa uzorcima malog obima, može se pretpostaviti da će $F_n^*(x)$ biti izvan pojasa

$$\left[F_{\mu + \frac{3\sigma}{\sqrt{m}}}(x), F_{\mu - \frac{3\sigma}{\sqrt{m}}}(x) \right], \quad x \in \mathbb{R},$$

ređe nego što je to slučaj za normalno raspodeljene vrednosti u odnosu na interval $(\mu - 3\sigma, \mu + 3\sigma)$. Stoga se za ovu kontrolnu kartu mora koristiti i vrednost T , a i V radi što puzdanijih rezultata. Kako normalno raspodeljen uzorak ima samo jednu maksimalnu vrednost, to za proces pod kontrolom važi $T = \frac{1}{n}$. Radi procene procenta defektnih uzoraka, može se računati

$$T^* = T - \frac{1}{n}$$

ili

$$T^* = T - \frac{n_{\max}}{n}$$

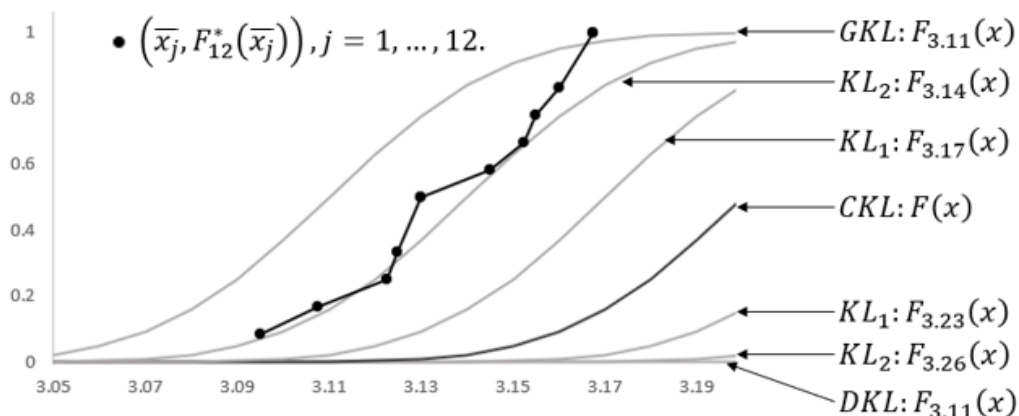
ako je poznato da su uzorci normalno raspodeljeni ali sa pomerenim očekivanjem.

Takođe, kontrolor procesa treba da proceni preveliku vrednost broja n_{\max} ili da, ako je to moguće ustanoviti, odredi standardizovani dozvoljeni maksimum tog broja.

Iz navedenih rasuđivanja sledi matematička zasnovanost kontrolne karte kvantil-zona.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Primer 4.4. Korišćeni su isti podaci kao u primeru (4.2), tj. podaci iz tabele 4.1. Kako su potrebna izračunavanja data u primeru (4.2), kontrolna karta kvantil-zona data je direktno.



Slika 4.14. Kontrolna karta kvantil-zona primenjena nad podacima tabele 4.1.

Više pravila tumačenja kontrolnih karata ukazuje na postojanje defekata, a samim tim i primetno narušen kvalitet proizvodnog procesa. Jedna tačka je iznad gornje kontrolne linije, deset tačaka je u trećoj zoni kontrolne karte i sl.

Računanjem realizacije statistike V dobija se

$$V = \frac{1}{12} (3 + 3 + 2 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3,1947 + 3,1947) = 2,9491,$$

što poređenjem sa očekivanom vrednošću, tj. onoj koja odgovara procesu pod kontrolom

$$E(V) = \frac{(n - n_{\max}) \cdot 1.3635 + n_{\max} \cdot 3.1947}{n} = \frac{(12 - 2) \cdot 1.3635 + 2 \cdot 3.1947}{12} = 1.6687$$

potvrđuje zaključak donet tumačenjem kontrolne karte. Isto važi ako se uzme $n_{\max} = 1$.

Vrednost statistike A za iste podatke je

$$A = \frac{1}{12} (3 + 2 + 2 + 3,1947 + 3,1947 + 3 + 2 + 3 + 2 + 2 + 3 + 2) = 2,5324.$$

Manja vrednost statistike A ukazuje na manju senzitivnosti \bar{X} karte u odnosu na kartu kvantil-zone.

Za ovu kontrolnu kartu je $T = \frac{2}{12} = 0,166$, odakle je $T^* = 0.083$, pa je 8,3% uzoraka defektno. ■

Za ovako definisane kontrolne granice sledeći stav može dati vrednost ARL_0 .

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Stav 4.5. Neka važe iste pretpostavke kao za Šuhartovu \bar{X} kontrolnu kartu. Tada

$$\alpha = 2P(D_n > 0.8664).$$

gde je D_n ima raspodelu Kolmogorova.

Dokaz. Sledeći niz jednakosti dokazuje stav.

$$\begin{aligned} \alpha &= P\left(F_n^*(\bar{X}_i) \notin \left[F_{\mu+\frac{3\sigma}{\sqrt{m}}}(\bar{X}_j), F_{\mu-\frac{3\sigma}{\sqrt{m}}}(\bar{X}_j)\right]\right) \\ &= P\left(F_n^*(x) \notin \left[F_{\mu+\frac{3\sigma}{\sqrt{m}}}(x), F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)\right]\right) \\ &= P\left(F_n^*(x) < F_{\mu+\frac{3\sigma}{\sqrt{m}}}(x) \vee F_n^*(x) > F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)\right) \\ &= 2P\left(F_n^*(x) > F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)\right) \\ &= 2P\left(|F_n^*(x)| > \left|F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)\right|\right) \\ &= 2P\left(|F_n^*(x) - F(x)| > \left|F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x) - F(x)\right|\right) \\ &= 2P\left(\sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| > \sup_{x \in \mathbb{R}} \left|F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x) - F(x)\right|\right) \\ &= 2P\left(\sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| > \left|F_{\mu-\frac{3\sigma}{\sqrt{m}}}\left(-\frac{1.5\sigma}{\sqrt{m}}\right) - F\left(-\frac{1.5\sigma}{\sqrt{m}}\right)\right|\right) \\ &= 2P(D_n > 0.8664). \blacksquare \end{aligned}$$

Dakle, α zavisi od n , tj. broja uzoraka koji se kontrolišu, i kod ove kontrolne karte. Međutim, varijabilitet koji nastaje promenama $n \in \mathbb{N}$ nema značajan uticaj na performanse ove kontrolne karte. Naime, za $n = 35$ dobije se $ARL_0 = 58.1$, a za $n = 10000$, $ARL_0 = 60.1$. Stoga, za dovoljno veliko n ARL_0 se može aproksimirati brojem 60.

Kao što je već rečeno, broj 3 u pojasu

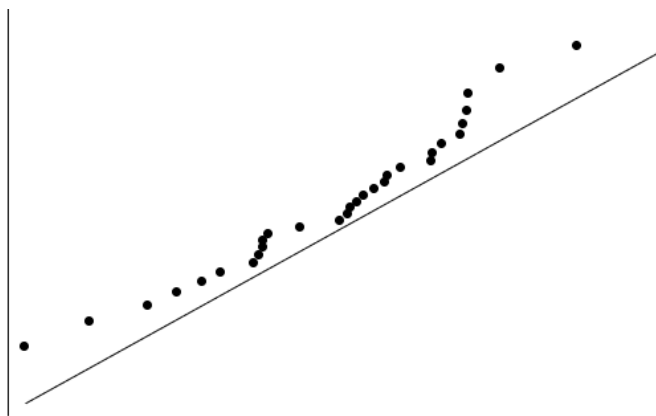
$$\left[F_{\mu+\frac{3\sigma}{\sqrt{m}}}(x), F_{\mu-\frac{3\sigma}{\sqrt{m}}}(x)\right], \quad x \in \mathbb{R},$$

sledi iz analogije sa pravilom 3σ i ilustrativnog je karaktera. Sužavanjem pojasa na odgovarajući novi dobiće se mnogo bolji rezultati ove kontrolne karte.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Q-Q kontrolna karta

Poznato je da je Q-Q dijagram grafička metoda preliminarne analize normalnosti. Formira se tako što se u pravougli koordinatni sistem unose tačke oblika $\left(F^{-1}\left(\frac{j-0.5}{n}\right), Y_j\right)$, $j = 1, \dots, n - 1$, za dobijeni uzorak X_1, \dots, X_n , $n \in \mathbb{N}$ i njegove statistike poretka Y_1, \dots, Y_n , iz obeležja X čija raspodela se upoređuje s normalnom, i funkciju F normalne raspodele $\mathcal{N}(\mu, \sigma^2)$, pa se te tačke posmatraju u odnosu na pravu $y = x$ (Jevremović & Avdović, 2020a). Ako se tačke skupljaju oko prave, kaže se da je raspodela približno normalna. U suprotnom, u slučaju tačaka koje se rasipaju sa strana prave, postoje izvesna odstupanja. Pozicije skiciranja se mogu menjati, a predložene su mnoge zavisno od potreba i konteksta istraživanja. Ovaj izbor je najčešće korišćen jer je najjednostavniji, a sam izbor nema veliki uticaj na preciznost zaključka (Thode, 2002, str. 21).



Slika 4.15. Q-Q dijagram

Ako se opet iskoristi ideja pravila 3σ preciznije informacije o normalnosti raspodele obeležja X se mogu doneti i na osnovu Q-Q dijagrama dodavanjem (grafika) linija $y = x \pm i\sigma$, $i = 1, 2, 3$. Naime, važi

$$x - 3\sigma < x - 2\sigma < x - \sigma < x < x + \sigma < x + 2\sigma < x + 3\sigma$$

za bilo koje $x \in \mathbb{R}$, pa i x koje je izvučeno iz normalno raspodeljenog obeležja, tj. na primer $F^{-1}\left(\frac{j-0.5}{n}\right)$, $j = 1, \dots, n$. Vrednosti $F^{-1}\left(\frac{j-0.5}{n}\right)$ pozicioniraju tačku tako da za Y_j treba da važi da je približno jednako, a u idealnom slučaju i jednako $F^{-1}\left(\frac{j-0.5}{n}\right)$.

Važi

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$F^{-1}\left(\frac{j-0.5}{n}\right) - 3\sigma < F^{-1}\left(\frac{j-0.5}{n}\right) - 2\sigma < F^{-1}\left(\frac{j-0.5}{n}\right) - \sigma < Y_j < F^{-1}\left(\frac{j-0.5}{n}\right) + \sigma < F^{-1}\left(\frac{j-0.5}{n}\right) + 2\sigma < F^{-1}\left(\frac{j-0.5}{n}\right) + 3\sigma \quad (4.8)$$

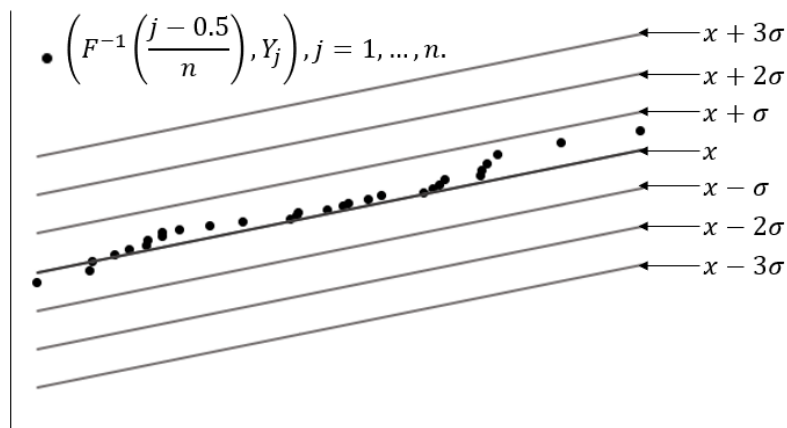
sa eventualnim odstupanjima čije su verovatnoće nastanka precizirane pravilom 3σ .

U ovom slučaju se svaki element $Y_j, j = 1, \dots, n$ uzorka Y_1, \dots, Y_n posmatra samo u odnosu na $F^{-1}\left(\frac{j-0.5}{n}\right)$. Stoga, ako je dat niz uređenih normalno raspodeljenih uzoraka, $Y_{k,i}, k = 1, \dots, n, i = 1, \dots, d, d \in \mathbb{N}$ po raspodeli $\mathcal{N}(\mu, \sigma^2)$, onda je uzorak $Y_{j,i}, i = 1, \dots, d, j \in \{1, \dots, n\}$ približno raspodeljen po raspodeli $\mathcal{N}\left(F^{-1}\left(\frac{j-0.5}{n}\right), \sigma^2\right)$, tj. na osnovu centralne granične teoreme treba da važi

$$\lim_{d \rightarrow +\infty} \frac{1}{d} \sum_{i=1}^d Y_{j,i} \approx F^{-1}\left(\frac{j-0.5}{n}\right), \quad j \in \{1, \dots, n\}.$$

Standardna devijacija se ne menja jer su uzoračka disperzija i očekivanje uzajamno nezavisni. Dakle, iz pravila 3σ važi (4.8) sa odstupanjima preciziranim istim pravilom.

Odatle se može zaključiti da se Q-Q diagram može formirati i na način predstavljen sledećom slikom.



Slika 4.16. Q-Q dijagram sa dodatnim linijama.

Korišćenje Q-Q dijagrama za preliminarnu analizu normalnosti je opšte poznata statistička metoda koja je ilustrovana u raznim slučajevima, pa će primer stoga ovde biti izostavljen.

Neka je $X_{i,j}, i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$ serija n uzoraka obima m obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$ i neka je za svaki od uzoraka $X_{i,1}, \dots, X_{i,m}, i \in \{1, \dots, n\}$ izračunata uzoračka sredina

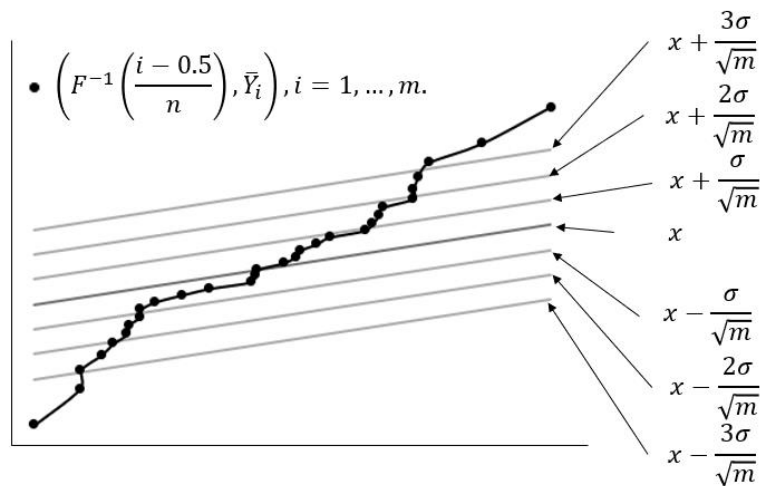
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

data po formuli (4.5) za svako $i = 1, \dots, n$. Ako je $X \sim \mathcal{N}(\mu, \sigma^2)$, onda je uzorak dobijenih srednjih vrednosti izvučen iz obeležja $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right)$, i neka je sada F funkcija raspodele obeležja \bar{X} .

Modifikacijom (4.8) i slike 4.16 dobija se sledeća definicija.

Definicija 4.14. *Q-Q kontrolnu kartu za seriju uzoraka $X_{i,j}$, $i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$ je Q-Q dijagram dat slikom 4.16 primenjen nad uzorkom (4.5), za funkciju F raspodele $\mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right)$. Drugim rečima, to je kontrolna karta čija je linija promena dobijane spajanjem tačaka $\left(F^{-1}\left(\frac{j-0.5}{n}\right), Y_j\right)$, $j = 1, \dots, n$, gornja kontrolna linija je prava $y = x + \frac{3\sigma}{\sqrt{m}}$, centralna linija je prava $y = x$, a donja kontrolna linija prava $y = x - \frac{3\sigma}{\sqrt{m}}$.*

Modifikacija slike 4.16 data je sledećom slikom.



Slika 4.17. Q-Q kontrolna karta.

Pravila interpretacije Q-Q kontrolne karte mogu biti ista kao u slučaju \bar{X} kontrolne karte jer se time uzimaju u obzir odstupanja predviđena pravilom 3σ .

Neka su za sledeće skupove date oznake.

$$S_1(i) = \left[\sigma F^{-1}\left(\frac{i-0.5}{n}\right) - \sigma, \sigma F^{-1}\left(\frac{i-0.5}{n}\right) + \sigma \right],$$

$$S_2(i) = \left[\sigma F^{-1}\left(\frac{i-0.5}{n}\right) - 2\sigma, \sigma F^{-1}\left(\frac{i-0.5}{n}\right) - \sigma \right] \cup \left[\sigma F^{-1}\left(\frac{i-0.5}{n}\right) + \sigma, \sigma F^{-1}\left(\frac{i-0.5}{n}\right) + 2\sigma \right],$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$S_3(i) = \left[\sigma F^{-1} \left(\frac{i-0.5}{n} \right) - 3\sigma, \sigma F^{-1} \left(\frac{i-0.5}{n} \right) - 2\sigma \right] \cup \left(\sigma F^{-1} \left(\frac{i-0.5}{n} \right) + 2\sigma, \sigma F^{-1} \left(\frac{i-0.5}{n} \right) + 3\sigma \right],$$

$$S_{3,1947}(i) = \left(-\infty, \sigma F^{-1} \left(\frac{i-0.5}{n} \right) - 3\sigma \right) \cup \left(\sigma F^{-1} \left(\frac{i-0.5}{n} \right) + 3\sigma, +\infty \right).$$

Na osnovu (4.8) može se definisati sledeća funkcija.

Definicija 4.15. Funkcija *Q-Q zone* $Q_{zone}: \{X_1, \dots, X_n\} \rightarrow \{1; 2; 3; 3,1947\}$ data je formulom

$$Q_{zone}(X_i) = \begin{cases} 1, & Y_j \in S_1(j) \\ 2, & Y_j \in S_2(j) \\ 3, & Y_j \in S_3(j) \\ 3.1947, & Y_j \in S_{3,1947}(j) \end{cases}$$

gde je $Y_j = X_i$ za uzorak X_1, \dots, X_n sređen u neopadajući poredak Y_1, \dots, Y_n .

Statistika

$$Q = \frac{1}{n} \sum_{i=1}^n Q_{zone}(X_i),$$

očigledno predstavlja procenu saglasnosti obeležja polaznog uzorka sa normalnom raspodelom. U slučaju procesa koji je pod idealnom kontrolom, tačke su raspodeljene duž centralne linije i dobija se $Q = 1$.

Stav 4.6. Ako je $X \sim \mathcal{N}(\mu, \sigma^2)$ onda

$$Q_{zone}(X_j): \begin{pmatrix} 1 & 2 & 3 & 3.1947 \\ 0,6826 & 0,2718 & 0,0429 & 0,0027 \end{pmatrix},$$

$$E(Q_{zone}(X_j)) = 1,3635 \text{ i } E(Q) = 1.3635.$$

Dokaz. Sledi iz (4.8) i pratećih rasuđivanja. ■

Statistika T u slučaju ove kontrolne karte je

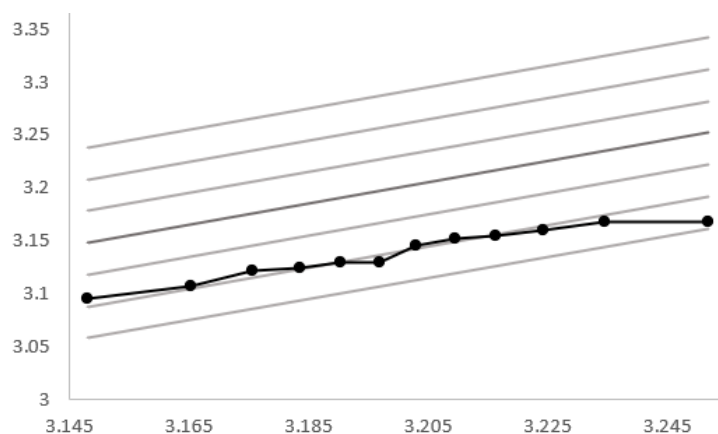
$$T = \frac{1}{n} \sum_{i=1}^n I(Q_{zone}(X_i) = 3,1947).$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Koristi se analogno kao kod prethodno navedenih karti, tj. grafičkih metoda preliminarne analize normalnosti.

Primer 4.5. Korišćeni su isti podaci kao u primeru 5.2, tj. podaci iz tabele 4.1. U primeru 5.2 data su izračunavanja prosečnih vrednosti za seriju uzoraka obeležja koje se kontroliše. Uzorak uređen u neopadajući poredak je 3,0950; 3,1075; 3,1225; 3,1250; 3,1300; 3,1300; 3,1450; 3,1525; 3,1550; 3,1600; 3,1675; 3,1675.

Za dobijeni uzorak obima 12 i funkciju F normalne raspodele $\mathcal{N}\left(3,2; \frac{0,06^2}{4}\right)$, izračunavaju se vrednosti $F^{-1}\left(\frac{j-0,5}{12}\right)$, $j = 1, \dots, 12$. Dobija se Q-Q kontrolna karta data sledećom slikom.



Slika 4.18. Q-Q kontrolna karta za podatke date tabelom 4.1.

Cele linija promena se nalazi ispod centralne kontrolne linije pa proizvodni proces nije pod kontrolom.

Izračunavanjem Q dobija se

$$Q = \frac{1}{12} (2 + 2 + 2 + 2 + 3 + 3 + 2 + 2 + 3 + 3 + 3 + 3) = 2,5.$$

Kako je u prethodnim primerima dobijeno $A = 2,5324$ i $V = 2,9491$, zaključujemo da je Q-Q kontrolna karta za dati primer registrovala defekte u manjoj meri od \bar{X} kontrolne karte i od kontrolne karte kvantil-zone. Kod ove kontrolne karte dobijeno je $T = 0$. ■

Napomena 4.4. Vrednosti statistika A i Q su kompatibilne zbog iste raspodele slučajnih veličina iz uzorka nad kojim su primenjene funkcije *zone* i Q_{zone} . Samim tim, one su i uporedive. To nije slučaj sa statistikom V jer zavisnost elemenata uzorka koji se dobija primenom funkcije q_{zone} utiče na vrednost očekivanja. Pritom, ne može se napraviti zadovoljavajuća analogija sa pravilom 3σ .

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kontrolna karta p-vrednosti

Ideja ove karte je u suštini izračunavanje p-vrednosti pri testiranju normalnosti raspodele obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$ iz kojeg su dobijeni uzorci obima m posmatrane serije n uzoraka $X_{i,j}$, $i = 1, \dots, n, j = 1, \dots, m, n, m \in \mathbb{N}$. Test koji se može koristiti za testiranje može biti proizvoljno izabran. Najbolje je izabrati neki koji je brz, a i dalje moćan. Primer takvog testa je test Kolmogorova (Jevremović & Avdović, 2020a). Kako se Šapiro-Vilk test pokazao najboljim kod često korišćenih testova normalnosti, naročito za male uzorke, to i on može poslužiti svrsi uprkos tome što je sporiji za primenu (Avdović & Jevremović, 2022). Drugi mogući izbori su Anderson-Darling test ili neki novi, na primer test zona (Avdović & Jevremović, 2023), ili test kvantil-zona (Avdović & Jevremović, 2022).

Kako je pri ovom postupku za donošenje odluke da proces od značaja nije pod kontrolom dovoljno dobiti bar jednu p-vrednost manju od izabranog praga značajnosti α , to se dobija situacija u kojoj nema potrebe za formiranjem kontrolne karte. Ipak, formiranje kontrolne karte opisane u nastavku može dati više informacija o procesu i detektovati defekt u nekim slučajevima koji ne bi bili lako uočljivi na ovaj način.

Neka je T_m funkcija koja određuje test-statistiku izabranog testa i W_α kritična oblast određena pragom značajnosti

$$\alpha = P_{X \sim \mathcal{N}(\mu, \sigma^2)}(T_m(X_1, \dots, X_m) \in W_\alpha).$$

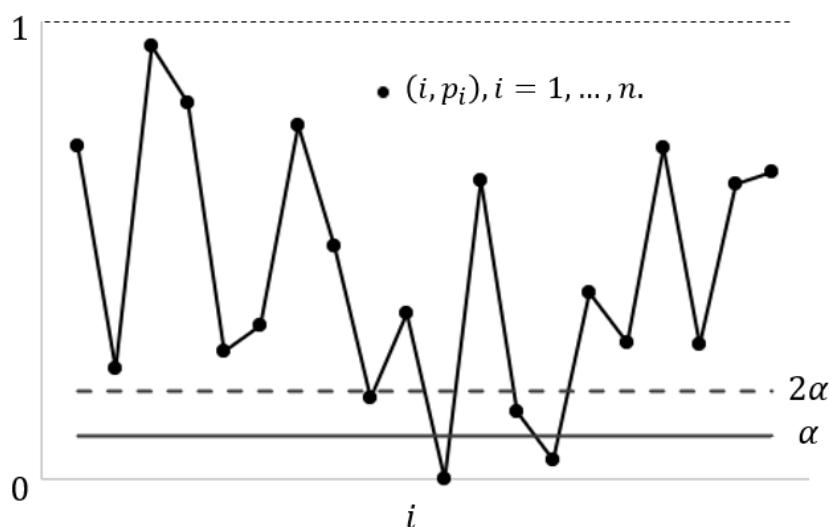
Definicija 4.16. Linija promena kontrolne karte p-vrednosti određena je tačkama (i, p_i) , $i = 1, \dots, n$, gde su

$$p_i = \inf\{\alpha \in (0,1) | T_m(X_{i1}, \dots, X_{im}) \in W_\alpha\},$$

tj. p-vrednosti testa određenog statistikom T_m , čija je donja kontrolna linija $y = \alpha$ (Slika 4.15).

Ova kontrolna karta ima samo jednu kontrolnu liniju jer je odbijanje hipoteze uslovljeno time da je $p < \alpha$. Sa druge strane, granični slučajevi u kojima je p veće od α ali blizu mogu ukazivati na odbijanje hipoteze ako se podaci reuzorkuju (Reinhart, 2015, str. 8). Kako je $\alpha = 0,05$ najčešći prag značajnosti (konvencionalno usvojen standard) i istraživači diskutuju na moguće odbacivanje hipoteze čak i u slučajevima $p < 0,09$ ili $p < 0,1$, ovde je predloženo dodavanje dodatne kontrolne linije $y = 2\alpha$. Ovo može biti od značaja u slučajevima kada su sve tačke iznad DKL, ali ispod nove kontrolne linije.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.19. Kontrolna karta p-vrednosti.

Prilagođavanjem standarda tumačenja \bar{X} karte dobijaju se sledeća pravila interpretacije kontrolne karte p-vrednosti, tj. pravila detekcije defekata.

- jedna tačka je ispod kontrolne granice;
- četiri od pet uzastopnih tačaka su između kontrolnih linija;

Pored ovih pokazatelja postoji još par karakteristika koje ukazuju na specifično ponašanje procesa i na postojanje defekta

- šest tačaka u nizu su u rastućem ili opadajućem poretku;
- petnaest uzastopnih tačaka su koncentrisane oko kontrolne linije 2α ;
- četrnaest tačaka konstantno osciluju gore i dole ;
- neobična ili neslučajna šema dobijena od tačaka sa podacima;
- jedna ili više tačaka blizu kontrolne granice.

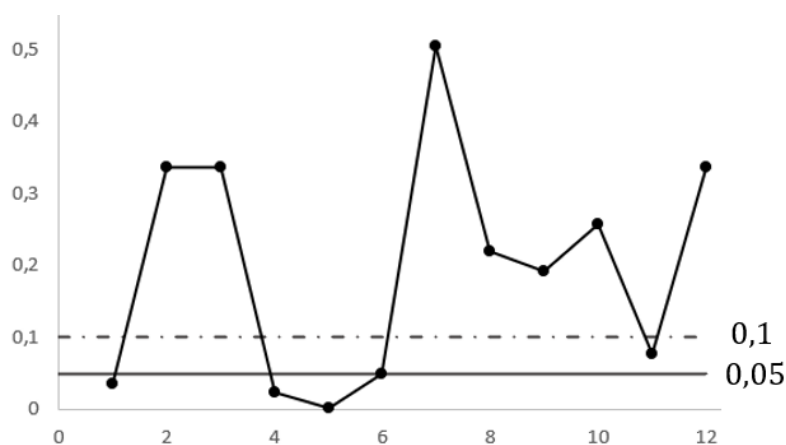
Statistika T u slučaju ove kontrolne karte je

$$T = \frac{1}{n} \sum_{i=1}^n I(p_i \leq \alpha).$$

Koristi se analogno kao kod prethodno navedenih karti, tj. grafičkih metoda preliminarne analize normalnosti.

Primer 4.5. Za podatke iz primera 5.2, tj. iz tabele 4.1 ilustrovana je primena kontrolne karte p-vrednosti. Uslovi kontrole proizvodnog procesa su takođe isti.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 4.20. Kontrolna karta p-vrednosti primenjena nad podacima tabele 4.1.

Kontrolna karta p-vrednosti detektuje 4 uzorka sa defektom, tj. koji nisu raspodeljeni po standardom zadatoj raspodeli $\mathcal{N}(3,2; 0.06^2)$. Dakle, proizvodni proces nije pod kontrolom. Preciznije, dobijeno je $T = \frac{4}{12} = 0,333$, tj. 33,3% uzoraka sadrži defekt. ■

Za potrebe izračunavanja ARL-a kontrolne karte p-vrednosti koristi se sledeća teorema.

Teorema 4.3. *Ako je pri tačnoj nultoj hipotezi raspodela test-statistike koja se koristi neprekidna tada se p-vrednost tog testa raspodeljuje po uniformnoj raspodeli $\mathcal{U}(0,1)$ (Wasserman, 2004, str. 158).*

Posledica 4.1. *Za kontrolnu kartu p-vrednosti sa kontrolnom granicom $y = \alpha$ važi $ARL_0 = \frac{1}{\alpha}$.*

Kako je $ARL_{\Delta} = \frac{1}{1-\beta}$, to ARL u slučaju kada proces nije pod kontrolom očigledno zavisi od moći testa koji se koristi u kontrolnoj karti p-vrednosti. Moć testa zavisi direktno od alternativne hipoteze, a u ovom slučaju biće razmatrani najčešći slučajevi procesa koji nije pod kontrolom, a to su slučajevi pomerenog očekivanja, tj. proseka. Dakle, razmatra se raspodela procesa koja je normalna, sa istom disperzijom, ali pomerenim očekivanjem (zbog simetričnosti normalne raspodele nije važno da li je pomeranje očekivanja povećanje ili smanjenje).

Performanse kontrolne karte p-vrednosti očigledno zavise od vrednosti moći testa čije se p-vrednosti uzimaju kao tačke kontrolne linije kao i obima uzoraka koji se uzimaju za kontrolu. Dva su razloga zašto je kontrolnu kartu p-vrednosti bolje koristiti samo u izuzetnim slučajevima, a to su niska moć za male obime uzoraka i preduga izračunavanja za više uzoraka koji se kontrolišu. Dakle, ovu kartu bi trebalo koristiti samo u slučajevima kada je dostupno malo uzoraka relativno velikog obima ($m \geq 10$).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

4.6. Analiza poređenja performansi

U nastavku date su vrednosti ARL_0 za Šuhartovu \bar{X} kontrolnu kartu, CUSUM kontrolnu kartu, EWMA kontrolnu kartu, EFR kontrolnu kartu, kontrolnu kartu kvantil-zona, Q-Q kontrolnu kartu. Kontrolne R i S karta su izostavljene iz ove analize jer one za cilj imaju kontrolu varijabilnosti procesa od interesa. Šuhartova \bar{X} , CUSUM i EWMA kontrolna karta kontrolišu očekivanje, tj. prosek obeležja kojim se proces kontroliše uzimajući u obzir dopustivu varijabilnost od tog očekivanja. Isto se obezbeđuje primenama EFR karte, kontrolne karte kvantil-zona i Q-Q kontrolne karte.

Tabela 4.2. Vrednosti ARL_0 (prva vrsta) i ARL_Δ (ostale vrste) za kontrolne karte Šuhartovu \bar{X} , CUSUM i EWMA kartu za odabrane vrednosti parametara (Rahardja, 2014; Montgomery, 2009).

Pomeranje očekivanja μ	\bar{X}	CUSUM $k = \frac{1}{2},$ $h = 4$	CUSUM $k = \frac{1}{2},$ $h = 5$	EWMA $L = 2,962,$ $\lambda = 0,2$	EWMA $L = 2,615,$ $\lambda = 0,05$
0	370,4	168	465	500	500
0,25	281,2	74,2	139	150	84,1
0,5	155,2	26,6	38,0	41,8	28,8
0,75	81,2	13,1	17,0	18,2	16,4
1	43,9	8,4	10,4	10,5	11,4
1,5	15,0	4,8	5,8	5,5	7,1
2	6,3	3,3	4,0	3,7	5,2
2,5	3,2	2,6	3,1	2,9	4,2
3	2,0	2,2	2,6	2,4	3,5
4	1,2	1,7	2,0	1,9	2,7

Na osnovu tabele 4.2 može se zaključiti da Šuhartova \bar{X} karta dobro detektuje proces pod kontrolom. Međutim performanse te karte su slabije u odnosu na druge kada je očekivana vrednost obeležja na osnovu kojeg se proces kontroliše pomerena za male vrednosti (manje ili jednake 1). CUSUM kontrolna karta ima bolje performanse. Za proces pod kontrolom, boljom se pokazala CUSUM karta sa parametrima $k = \frac{1}{2}$ i $h = 5$. Za proces kod kojeg je očekivanje obeležja koje se koristi pomereno CUSUM karta sa parametrima $k = \frac{1}{2}$ i $h = 4$ brže detektuje defekt u odnosu na Šuhartovu \bar{X} kartu. Kontrolna karta EWMA sa parametrima $L = 2,962$ i $\lambda = 0,2$ ima najveći $ARL_0 = 500$, dok za pomeranje očekivanja veće od 0,75 ima ARL_Δ manji od svih perthodno spomenutih karata. Za očekivanje pomereno za manje od 1, najbolje performanse pokazuje kontrolna karta EWMA sa parametrima $L = 2,615$ i $\lambda = 0,05$. Za očekivanje pomereno za 1 ili više, razlike u performansama svih analiziranih karata osim Šuhartove \bar{X} karte su zanemarljive.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Tabela 4.3. Vrednosti ARL_0 (prva vrsta) i ARL_Δ (ostale vrste) za kontrolne karte EFR, Kvantil-zona i Q-Q kartu za odabrane vrednosti parametara..

Pomeranje očekivanja μ	EFR karta $d_{n; 0,0027}$	EFR karta $d_{n; 0,002}$	Karta kvantil-	
			zona $F_{\mu \pm \frac{3\sigma}{\sqrt{m}}}$	Q-Q karta
0	370,4	500	60	370,4
0,25	370,4	17,8	60	281,2
0,5	1,7	1,4	60	155,2
0,75	1,4	1,3	60	81,2
1	1,4	1,3	3,2	43,9
1,5	1,4	1,3	1	15,0
2	1,4	1,3	1	6,3
2,5	1,4	1,3	1	3,2
3	1,4	1,3	1	2,0
4	1,4	1,3	1	1,2

Najbolje vrednosti ARLa daje EFR kontrolna karta. Naime, za kontrolne granice sa parametrom $d_{n; 0,0027}$ ARL_0 je isti kao kod Šuhartove kontrolne karte, dok pomeranje očekivanja za 0,25 ova kontrolna karta ne detektuje. Za pomeranja očekivanja veća ili jednaka od 0,5, ova kontrolna karta ima vrednost ARL_Δ manju od 2. Ista kontrolna karta za vrednost parametra $d_{n; 0,002}$ daje još bolje rezultate i za pomeranje očekivanja od 0,25 daje vrednost ARL_Δ 17,8. Za veća pomeranja ARL je niži od 2.

Kod kontrolne karte kvantil-zona defekt se ne detektuje za pomeranja manja ili jednaka 0,75, međutim, ako je pomeranje jednako 1, onda je $ARL=3,2$, dok je za veća pomeranja $ARL=1$, što predstavlja bolji rezultat u odnosu na ostale kontrolne karte. Eventualne modifikacije granica ove kontrolne karte mogle bi dovesti do boljih rezultata. S obzirom na preciznost aproksimacije funkcije raspodele empirijskom funkcijom raspodele, sužavanje pojasa kontrole za ovu kontrolnu kartu neće značajno smanjiti ARL_0 , a može značajno poboljšati performance u slučaju ARL_Δ . Stoga se uz adekvatan izbor kontrolnih granica i ova karta može smatrati najboljim izborom.

Q-Q kontrolna karta pokazuje iste performanse kao i Šuhartova kontrolna karta ((4.8) i stav 4.6).

Konačan komentar analize poređenja ukazuje na to da je EFR karta najbolja kontrolna karta za kontrolisanje procesa određenih normalno raspodeljenim obeležjima. U slučaju kada je odstupanje od normalne raspodele veće u smislu očekivanja, kontrolna karta kvantil-zona je takođe dobar izbor. Kao što je već rečeno, i odgovarajuće varijante EWMA karte imaju dobre performanse.

5. Testovi normalnosti

Pored kontrolnih karata i testovi saglasnosti, tj. normalnosti čine oblast u kojoj rezultati ovog istraživanja daju značajne rezultate. Rezultati predstavljaju uvođenje dva nova testa normalnosti sa detaljnim obrazloženjem njihove teorijske zasnovanosti, kao i izvršenim Monte Karlo simulacijama da bi se dobile tablice raspodela ili tablice moći tih testova u zavisnosti od raznih alternativnih hipoteza, tj. raspodela. Dodatno, analiza poređenja dobijenih testova sa postojećim najčešće korišćenim je takođe data.

Definicija 5.1. *Test saglasnosti je svaki neparametarski test koji testira hipotezu o saglasnosti raspodele obeležja iz kojeg je dobijen uzorak sa nekom pretpostavljenom (nultom) raspodelom. Ako je nulta raspodela normalna, kaže se da je u pitanju test normalnosti.*

Neki od najpoznatijih testova o kojima će biti reči jesu test Kolmogorova, Lilifors⁴¹ test, Pirsonov χ^2 test, Šapiro⁴²-Vilk⁴³ test i Anderson⁴⁴-Darling⁴⁵ test.

Testovi saglasnosti uglavnom testiraju generalne slučajevne hipoteze saglasnosti, a neki samo hipotezu normalnosti. Ako test ispituje saglasnost sa samo jednom raspodelom to ne predstavlja ograničenje testa, što je pokazano sledećim stavom. U slučaju kada će potrebna modifikacija za drugu raspodelu zahtevati značajno duža izračunavanja, onda je bolje izabrati neki drugi test.

Stav 5.1. *Neka je X_1, \dots, X_n , $n \in \mathbb{N}$ uzorak izvučen iz obeležja $X \sim F_X(x)$, $x \in \mathbb{R}$, a $F_Y(x)$, $x \in \mathbb{R}$ funkcija raspodele obeležja Y koja ima inverznu funkciju. Uzorak*

$$Y_i = F_Y^{-1}(F_X(X_i)), \quad i = 1, \dots, n$$

je uzorak obeležja Y .

⁴¹ Hubert Whitman Lilliefors (1928-2008), američki statističar.

⁴² Samuel Sanford Shapiro (1930-), američki statističar i inženjer.

⁴³ Martin Bradbury Wilk (1922-2013), kanadski statističar.

⁴⁴ Theodore Wilbur Anderson (1918-2016), američki matematičar i statističar.

⁴⁵ Donald Allan Darling (1915-2014), američki statističar.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Dokaz. Ako je $X_i, i = 1, \dots, n$ uzorak sa raspodelom kojoj odgovara funkcija F_X , onda će $F_X(X_i), i = 1, \dots, n$ biti uzorak ravnomerno raspodeljenih brojeva na intervalu $[0,1]$, što znači da će $F_Y^{-1}(F(X_i)), i = 1, \dots, n$ biti uzorak raspodele kojoj odgovara funkcija F_Y . ■

Dakle, testiranjem saglasnosti raspodele obeležja uzorka $Y_i = F_Y^{-1}(F_X(X_i)), i = 1, \dots, n$ sa raspodelom određenom sa F_Y , zapravo se testira saglasnost raspodele obeležja uzorka $X_i, i = 1, \dots, n$ sa raspodelom određenom sa F_X . Na primer, korišćeni test testira saglasnost raspodele obeležja uzorka samo sa normalnom raspodelom. Ako je potrebno testirati saglasnost raspodele obeležja uzorka sa, na primer, eksponencijalnom raspodelom, onda se to može uraditi korišćenjem stava 5.1.

Takođe, nulte raspodele ne moraju biti u potpunosti određene. Naime, ako je poznato koja je nulta raspodela može se desiti da vrednosti njenih parametara nisu precizirane. Samim tim, ocenjivanje parametara je za ovu temu od velike važnosti. Dodatno, ocenjivanje parametara nekad narušava postavljeni model tako da raspodela test-statistike ne bude određiva. U tom slučaju, a kada raspodela test statistike ne zavisi od parametara raspodele nulte hipoteze, Monte Karlo simulacijama velikog obima se formiraju tablice raspodele koje se koriste pri testiranju (Milošević, 2021, 126-127).

5.1. Moć i p-vrednost testa

Neka je X_1, \dots, X_n prost slučajan uzorak iz obeležja $X \sim F(x), x \in \mathbb{R}$.

Definicija 5.2. Moć testa $1 - \beta$ je verovatnoća da pri tačnoj alternativnoj hipotezi H_1 test-statistika T posmatranog testa bude unutar kritične oblasti W , tj.

$$1 - \beta = P(T(X_1, \dots, X_n) \in W | H_1).$$

Drugim rečima, moć testa je verovatnoća da se pri tačnoj alternativnoj hipotezi ona i prihvati.

Kada su u pitanju neparametarski testovi, a naročito testovi saglasnosti, moć testa se ne može neposredno izračunati i zavisi od alternativne hipoteze. Korišćenjem Monte Karlo simulacija moć testa se može aproksimirati dovoljno precizno da da informaciju o pouzdanosti rezultata testa za izabrani prag značajnosti α , obim uzorka n i alternativnu hipotezu, tj. alternativnu raspodelu (Avdović & Jevremović, 2022). Već sa 10.000 simulacija proces ili njegov dobijeni trend se može smatrati stabilnim (Ritter, Schoelles, Quigley & Klein, 2011).

Algoritam 5.1. Aproksimacija moći testa saglasnosti sa test-statistikom T za prag značajnosti α kojim je određena kritična oblast $W \subset \mathbb{R}$ se postiže na sledeći način.

1. Modelirati uzorak x_1, \dots, x_n obeležja X sa alternativnom raspodelom.
 - a. Modeliranje vršiti Monte-Karlo metodom ako raspodela test-statistike ne zavisi od parametara raspodele obeležja X .

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

- b. Modeliranje vršiti butstrep metodom ako raspodela test-statistike zavisi od parametara raspodele obeležja X (Babu & Rao, 2004).
2. Izračunati $T(x_1, \dots, x_n)$.
 - a. Ako je nulta hipoteza prosta, T se izračunava za zadate vrednosti parametara.
 - b. Ako je nulta hipoteza složena, parametri se ocenjuju na osnovu uzorka, i T se izračunava za dobijene ocene korišćene kao vrednost parametara.
3. Ponoviti prva dva koraka m puta i time dobiti uzorak $y_{n,1}, \dots, y_{n,m}$ realizovanih vrednosti test-statistike T (što je m veće to će aproksimacija biti preciznija).
4. Odrediti empirijsku funkciju raspodele uzorka $y_{n,1}, \dots, y_{n,m}$

$$F_m^*(x) = \frac{1}{m} \sum_{i=1}^m I(y_{n,i} \leq x), \quad x \in \mathbb{R}.$$

5. Izračunati moć testa po formuli
 - a. $1 - \beta = F_m^*(c)$ ako je kritična oblast oblika $W = (-\infty, c]$, $c < +\infty$.
 - b. $1 - \beta = 1 - F_m^*(c)$ ako je kritična oblast oblika $W = [c, +\infty)$, $c > -\infty$.
 - c. $1 - \beta = F_m^*(c_1) + (1 - F_m^*(c_2))$ ako je kritična oblast oblika $W = (-\infty, c_1] \cup [c_2, +\infty)$, $-\infty < c_1 < c_2 < +\infty$.⁴⁶

Postoje i drugačiji pristupi u analizi pouzdanosti zaključaka koji se izabranim testom saglasnosti dobijaju. Naime, može se poći od praga značajnosti, a samim tim i poznate kritične oblasti, željene moći testa i nakon toga dobiti minimalni obim uzorka za koji će test imati unapred zadanu vrednost moći. Takav način je komplikovaniji i daje manji uvid u konzistentnost pouzdanosti testa jer se ne mogu videti promene koje nastaju u moći promenom obima uzorka. Takva informacija je bitna jer iako postoje najčešće prihvatljive vrednosti moći one i dalje zavise od istraživanja, primene dobijenih rezultata, subjektivnih tendencija istraživača i dr. (Reinhart, 2015, str. 15-31).

Kada su u pitanju definicija i interpretacija p-vrednosti testa one mogu biti date za uopšteni slučaj nulte hipoteze. Taj slučaj podrazumeva zadavanje nulte hipoteze nekim svojstvom više obeležja koja mogu biti zavisna ili nezavisna. Samim tim, testiranje se obavlja na osnovu više uzoraka koji ne moraju biti istog obima. Ipak za potrebe ovog istraživanja i rada biće razmatrana samo hipoteza o saglasnosti raspodele sa nekom zadanom raspodelom, a najčešće normalnom raspodelom, tj.

$$H_0: \text{„Raspodela obeležja } X \text{ data je funkcijom } F(x)\text{“}.$$

Jasno je da je kod nekih testova potrebno dodatno precizirati hipotezu.

Definicija i razmatranje osobina p-vrednosti u literaturi je često nedovoljno precizirano pa se zbog toga u primeni često nalaze greške u interpretaciji (Rafi & Greenland, 2020; Vidgen

⁴⁶ Kritične vrednosti, tj. konačne granice kritične oblasti se često nazivaju i veličinama efekta (effect size).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

& Yasseri, 2016). Stoga će se to ovde detaljnije argumentovati, ali samo u odnosu na navedenu nultu hipotezu, jedno obeležje i jedan uzorak $X_1, \dots, X_n, n \in \mathbb{N}$ tog obeležja.

Neka je dalje T funkcija koja određuje test-statistiku izabranog testa saglasnosti, α prag značajnosti, tj. verovatnoća greške prve vrste, da pri tačnoj nultoj hipotezi test rezultira zaključkom da se ona odbacuje, W kritična oblast, tj. skup vrednosti test-statistike za koju se nulta hipoteza odbacuje, a konkretno W_α kritična oblast određena pragom značajnosti α .

Definicija 5.3. *p-vrednost testa datog statistikom T data je sa*

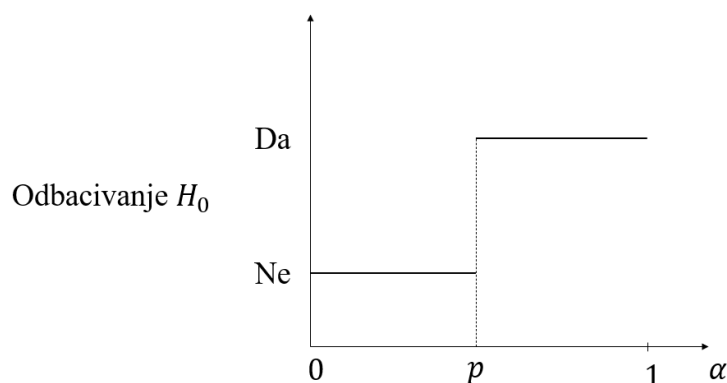
$$p = \inf\{\alpha \in (0,1) | T(X_1, \dots, X_n) \in W_\alpha\},$$

Dakle, p-vrednost je najmanji prag značajnosti za koji će se testom određenim statistikom T nulta hipoteza odbaciti (Wasserman, 2004, str. 156).

Na osnovu definicije p-vrednosti dolazi se do osnovnog pravila tumačenja. Ako je

- $\alpha \geq p$ nulta hipoteza se odbacuje;
- $\alpha < p$ nulta hipoteza se prihvata

za prag značajnosti α . Radi preciznosti je potrebno napomenuti to da zaključak koji se donosi testom u nekim slučajevima ne mora biti tačan, pa je bolje komentar tumačenja dati kao „na osnovu testiranja dobijenog uzorka dobija se zaključak da nultu hipotezu treba prihvatiti (odbaciti) sa pragom značajnosti α “. Grafička interpretacija definicije p-vrednosti data je na slici 5.1.



Slika 5.1. Grafički predstavljena definicija p-vrednosti.

P-vrednost se može definisati i kao

$$p = \sup\{\alpha \in (0,1) | T(X_1, \dots, X_n) \notin W_\alpha\}.$$

tj. kao najveći prag značajnosti za koji će test određen statistikom T prihvatiti nultu hipotezu. Ova formulacija u nešto drugačijem obliku data je kod nekih autora (Merkle & Vasić, 1995, str. 220; Petrović, 2018, str. 159).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kritična oblast testa može imati sledeće oblike

- KO1. $W_\alpha = (-\infty, c], c \in \mathbb{R};$
 KO2. $W_\alpha = [c, +\infty), c \in \mathbb{R}$
 KO3. $W_\alpha = (-\infty, c_1] \cup [c_2, +\infty), c_1, c_2 \in \mathbb{R};$
 KO4. $W_\alpha = (-\infty, -c] \cup [c, +\infty), c \in \mathbb{R}.$

U slučajevima kada je statistika T ograničena, odozdo ili odozgo, umesto beskonačnosti stavljaju se odgovarajuće granice. Kako se izračunavanje p-vrednosti za te specijalne slučajeve suštinski ne razlikuje, neće biti tretirani zasebno.

Stav 5.2. *Ako je kritična oblast testa datog statistikom T oblika*

$$KO1, \text{ onda je } p = \sup_{X \sim F(x)} P(T(X_1, \dots, X_n) \leq T(x_1, \dots, x_n));$$

$$KO2, \text{ onda je } p = \sup_{X \sim F(x)} P(T(X_1, \dots, X_n) \geq T(x_1, \dots, x_n));$$

$$KO3, \text{ onda je } p = 2 \min \left\{ \sup_{X \sim F(x)} P \left(T(X_1, \dots, X_n) \underset{\geq}{\leq} T(x_1, \dots, x_n) \right) \right\};$$

$$KO4, \text{ onda je } p = \sup_{X \sim F(x)} P(T(X_1, \dots, X_n) \geq |T(x_1, \dots, x_n)|),$$

gde je x_1, \dots, x_n realizovani uzorak.

Dokaz. Sledi direktno iz definicije.

Realizovani uzorak ili veličina efekta ne moraju biti takvi da verovatnoća stava 5.2 dostiže supremum. Stoga je ovaj stav moguće primeniti samo za dobijanje aproksimacije p-vrednosti. Sledeća posledica precizira postupak kojim se to postiže.

Posledica 5.1. *Ako je kritična oblast testa datog statistikom T oblika*

$$KO1, \text{ onda je } p \approx P(T(X_1, \dots, X_n) \leq T(x_1, \dots, x_n) | X \sim F(x));$$

$$KO2, \text{ onda je } p \approx P(T(X_1, \dots, X_n) \geq T(x_1, \dots, x_n) | X \sim F(x));$$

$$KO3, \text{ onda je } p \approx 2 \min \left\{ P \left(T(X_1, \dots, X_n) \underset{\geq}{\leq} T(x_1, \dots, x_n) | X \sim F(x) \right) \right\};$$

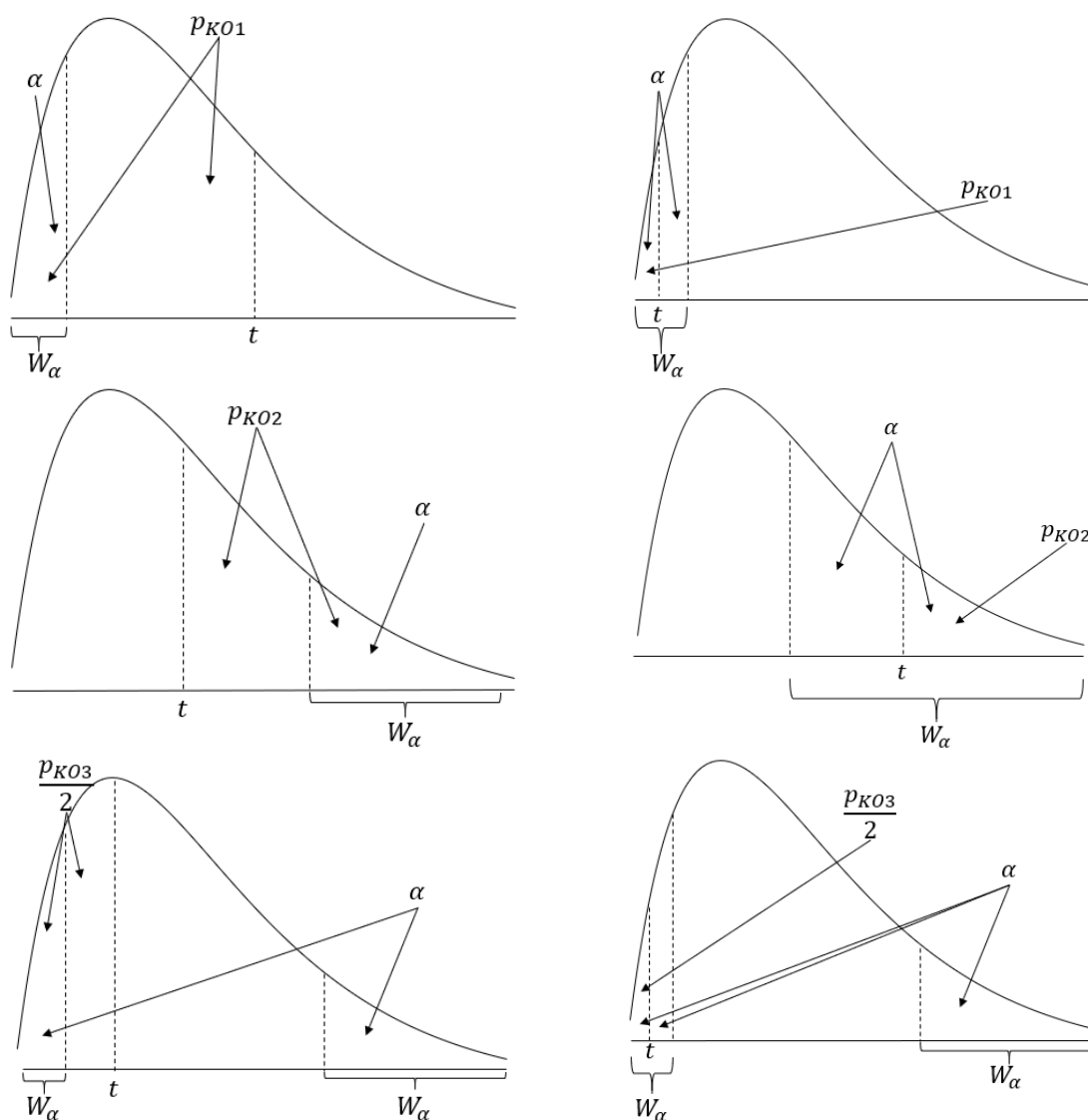
$$KO4, \text{ onda je } p \approx P(T(X_1, \dots, X_n) \geq |T(x_1, \dots, x_n)| | X \sim F(x)).$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

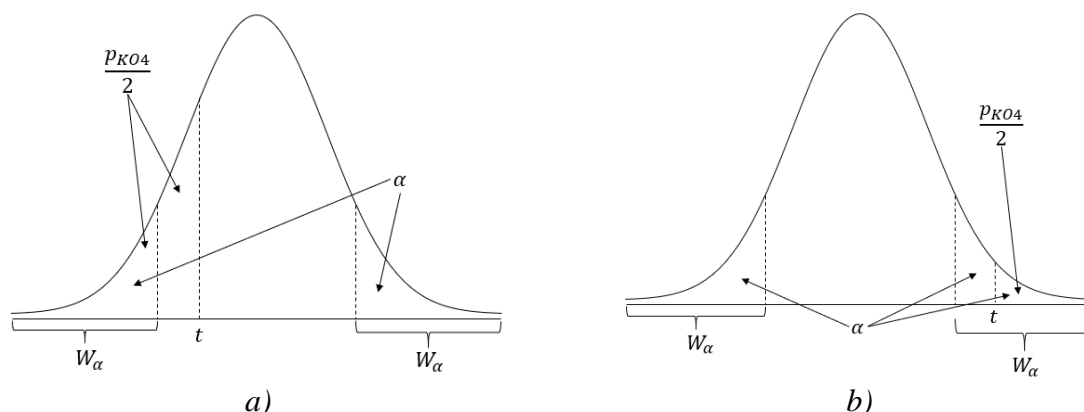
gde je x_1, \dots, x_n realizovani uzorak.

Na osnovu navedenog može se reći da se p-vrednost približno izračunava kao verovatnoća da prost slučajni uzorak odstupa od svojstva nulte hipoteze bar onoliko koliko je to slučaj za njegovu realizaciju. Ovu formulaciju posledice 5.1, mnogo autora koristi kao definiciju p-vrednosti (Cabrera & McDougal, 2002, str. 76; Reinhart, 2015, str. 8; Hogg, McKean & Craig, 2019, str. 280; Fischetti, 2018, str. 140; Upton & Cook, 2006, str. 200), što može biti jedan od razloga brojnih grešaka u tumačenju i razumevanju p-vrednosti u literaturi.

Za grafičku interpretaciju posledice 5.1, datu na slici 5.2 je $t = T(x_1, \dots, x_n)$, a W_α kritična oblast.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.2. p -vrednost za razne oblike kritične oblasti. a) Prihvatanje H_0 ; b) Odbacivanje H_0 .

Takođe, autori p -vrednost definišu i kao meru kritične oblasti čija je granica veličina efekta (realizovana vrednost test-statistike) (Popović, 2003, str. 86).

Kako je p -vrednost dobijena posledicom 5.1 procena, tj. aproksimacija prave p -vrednosti, ne može se znati da li je zaključak koji se donosi istinski pouzdan. Vaserman (2004, str. 156) daje pravilo tumačenja koje je bolje od onog koje je već navedeno (Reinhart, 2015, str. 39). Dato tumačenje je za prag značajnosti 0,05 jer je to najčešće korišćena vrednost. U primenjenoj statistici se uglavnom koristi kao standardizovana još otkako je uvedena od strane Fišera (Dahiru, 2008). Naravno, izbor praga značajnosti treba prilagoditi kontekstu i potrebnoj preciznosti istraživanja (Betensky, 2019).

Tabela 5.1. Tumačenje p -vrednosti.

p	Dokaz protiv H_0
$< 0,01$	Veoma jak
$0,01 \leq p < 0,05$	Jak
$0,05 \leq p \leq 0,1$	Slab
$> 0,1$	Slab ili nepostojeći

U literaturi se primer često koristi kao sredstvo uvođenja, tj. definisanja p -vrednosti (Reinhart, 2015, str. 8; Hogg, McKean & Craig, 2019, str. 279-280). Primere treba koristiti isključivo kao dodatnu ili pomoćnu nastavnu metodu.

5.2. Primeri testova saglasnosti – normalnosti

U radu će biti navedeni neki novi testovi normalnosti i dato njihovo poređenje sa klasičnim testovima normalnosti koji su navedeni u ovom odeljku, test Kolmogorova, Lilifors test, Pirsonov χ^2 test, Šapiro-Vilk test i Anderson-Darling test.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Test Kolmogorova

U stranoj literaturi se test Kolmogorova često naziva testom Kolmogorova-Smirnova za jedan uzorak. U ovom radu će ovaj test biti nazivan testom Kolmogorova jer se u domaćoj literaturi naziv „test Kolmogorova-Smirnova“ odnosi na test za dva uzorka.

Nulta hipoteza data je sa H_0 : „Raspodela obeležja X data je funkcijom $F(x)$ “. Funkcija F je funkcija nulte raspodele, a hipoteza se umesto sa F može formulirati sa gustinom raspodele, tipično korišćenom oznakom raspodele, ili prosto rečima. Ako nulta raspodela zavisi od vrednosti parametara da bi se test Kolmogorova primenio vrednosti parametara moraju biti precizirane (Massey, 1951).

Test-statistika ovog testa data je sa

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|.$$

Poznato je da ima raspodelu Kolmogorova opisanu Teoremom 4.2. Kritična oblast je oblika $W = [c, 1]$, gde se c određuje iz uslova $P(D_n \geq c | H_0) = \alpha$, za željeni prag značajnosti α . Kritična vrednost c se iz jednačine date teoremom 4.2. teško određuje pa su za raspodelu Kolmogorova date statističke tablice u svakom udžbeniku ili zbirci iz matematičke statistike, a sada i svaki softver za statistiku ima opciju primene ovog testa.

Kritična oblast test-statistike testa Kolmogorova je oblika KO_2 , pa ako je d_n realizovana vrednost test-statistike D_n , na osnovu posledice 5.1 p-vrednost se približno izračunava kao

$$p = P(D_n \geq d_n | H_0).$$

Lilifors test

Lilifors test je prilagođeni test Kolmogorova kada vrednosti parametara nulte raspodele nisu poznate već se ocenjuju odgovarajućim statistikama. Često se kaže da je ovaj test „test Kolmogorova sa Liliforsovom korekcijom“. Konkretno, u svom radu iz 1967. godine Lilifors je predložio modifikaciju testa Kolmogorova za testiranje normalnosti raspodele sa nepoznatim vrednostima parametara μ i σ^2 koji se ocenjuju redom statistikama $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ i $\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Naravno, raspodela test-statistike D_n tada nije raspodela Kolmogorova. Novodobijena raspodela nije određiva u smislu dobijanja njene funkcije i gustine raspodele, već je dobijena Monte Karlo simulacijama test-statistike a Lilifors ju je predstavio tablicom (Lilliefors, 1967).

Jedan od sporednih zaključaka ovog istraživanja je da se moć testa saglasnosti, pri prostoj nultoj hipotezi, povećava kada se parametri ocene. To se pokazalo i kod ove dve varijante suštinski istog testa (Avdović & Jevremović, 2022).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Bitno je napomenuti da se test Kolmogorova može koristiti standardnim postupkom tako što se ocenjene vrednosti parametra uzmu kao poznati parametri, ali da će tada korišćenje raspodele Kolmogorova dati manje precizne ili netačne rezultate (Babu & Rao, 2004).

Analogno testu Kolmogorova, p-vrednost približno se izračunava sa

$$p = P(D_n \geq d_n | H_0),$$

gde D_n u ovom slučaju ima Lilifors raspodelu (dobijenu simulacijama za složenu nultu hipotezu) pa je d_n dobijeno iz tablica te raspodele.

Pirsonov χ^2 test

Nulta hipoteza data je sa H_0 : „Raspodela obeležja X je $F(x)$ “. Test-statistika χ^2 testa data je sa

$$\chi_U^2 = \sum_{j=1}^r \frac{(M_j - np_j)^2}{np_j}$$

gde je skup S vrednosti obeležja X , podeljen u r disjunktivnih podskupova $S_j, j = 1, \dots, r, M_j = \sum_{i=1}^n I(X_i \in S_j)$ i $p_j = P(X_i \in S_j), j = 1, \dots, r$ (Cochran, 1952; Jevremović, 2014, str. 256-258).

Prednost ovog testa je da je granična raspodela test-statistike jednostavnija od raspodele nekih drugih. Naime, raspodela test-statistike χ_U^2 je pod određenim uslovima, detaljno opisanim u radu Kohrana (1952), χ_{r-s-1}^2 (χ^2 raspodela sa $r - s - 1$ stepeni slobode) gde je s broj nepoznatih parametara nulte raspodele, koji se pri računanju test-statistike ocenjuju metodom koja minimizira vrednost test statistike (Milošević, 2021, str. 136-137). Kohran (1952) nagoveštava da sve metode ocenjivanja daju ocene koje su asimptotski iste kao ocene metodom maksimalne verodostojnosti, a referiše da je Pirson u originalnom radu tvrdio da se raspodela ne menja značajno (u smislu krajnje odluke) ako se broj stepeni slobode ne umanjuje za s .

Kritična oblast je oblika $W = [c, +\infty)$, gde se c određuje iz uslova $P(\chi_U^2 \geq c | H_0) = \alpha$, za željeni prag značajnosti α . Za χ^2 raspodelu vrednosti se ne računaju direktno već se uzimaju iz odgovarajućih tablica. (Jevremović, 2014, str. 256-258). Sada se u skoro svakom statističkom softveru može primeniti χ^2 test.

Kritična oblast test-statistike χ^2 testa je oblika KO2, pa ako je χ_u^2 realizovana vrednost test-statistike χ_U^2 , na osnovu posledice 5.1 p-vrednost se približno izračunava kao

$$p = P(\chi_U^2 \geq \chi_u^2 | H_0).$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Šapiro-Vilk test

Nulta hipoteza data je sa H_0 : „Raspodela obeležja X je normalna“. Test-statistika Šapiro-Vilk testa data je sa

$$W = \frac{(\sum_{i=1}^n a_i Y_i)^2}{(n-1)S_n^2}$$

gde su $Y_i, i = 1, \dots, n$ statistike poretka (uzorak uređen u neopadajući niz), $(a_1, \dots, a_n) = \frac{m^T V^{-1}}{C}$ za $m^T = (E(Y_1), \dots, E(Y_n))$, $V = [\sigma_i \sigma_j \rho_{ij}]_{i,j=1}^n$ (σ_i standardna devijacija Y_i , a ρ_{ij} koeficijent korelacije Y_i i Y_j) i $C = \|V^{-1}m\|$ (gde je $\|\cdot\|$ Euklidova norma matrice).

Kritična oblast je oblika $W_{crit} = [0, c]$ gde se kritična vrednost c određuje iz uslova

$$P(W \leq c | H_0) = \alpha.$$

Ukoliko parametri raspodele nisu precizirani oni se ocene izabranim statistikama, a zatim se dobijene ocene tretiraju kao poznati parametri.

Za raspodelu statistike W takođe nema eksplicitno određenih funkcije i gustine raspodele već su kritične vrednosti tabelirane za razne pragove značajnosti, a izračunate putem Monte-Karlo simulacija (Shapiro & Wilk, 1965).

Grubo rečeno, test statistika W Šapiro-Vilk testa je mera linearnosti (straightness) linije dobijene spajanjem tačaka Q-Q dijagrama (Yang & Berdine, 2021).

Kritična oblast test-statistike Šapiro-Vilk testa je oblika KO1, pa ako je w realizovana vrednost test-statistike χ_{ij}^2 , na osnovu posledice 5.1 p-vrednost se približno izračunava kao

$$p = P(W \leq w | H_0).$$

Anderson-Darling test

Nulta hipoteza data je sa H_0 : „Raspodela obeležja X je $F(x)$ “. Test-statistika Anderson-Darling testa data je sa

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} (\ln(F(Y_i)) + \ln(1 - F(Y_{n+1-i})))$$

gde su $Y_i, i = 1, \dots, n$ statistike poretka (uzorak uređen u neopadajući niz) (Anderson & Darling, 1954). Iako izraz test-statistike nije složen, teorijski postupci koji garantuju rezultat su komplikovani i apstraktni (Anderson & Darling, 1952).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kritična oblast je oblika $W = [c, +\infty)$, gde se c određuje iz uslova $P(A^2 \geq c | H_0) = \alpha$, za željeni prag značajnosti α . Raspodela test-statistike se takođe određuje Monte Karlo simulacijama, a kritične vrednosti se uzimaju iz odgovarajućih tablica. (Marsaglia, 2004).

Ocenjivanje parametara se može vršiti na više načina zavisno od toga sa koliko parametara je raspodela određena, a zatim koliko ih je poznato a koliko ne. Ocene parametara se mogu dobijati metodom maksimalne verodostojnosti ili metodom najmanjih kvadrata. Raspodelu statistike i metode ocene parametara Stivens (1976) je detaljno diskutovao. Ako je za neki element uzorka, vrednost funkcije F jednaka nuli ili jedinici A^2 se ne može izračunati (Stephens, 1974).

Kritična oblast test-statistike Anderson-Darling testa je oblika KO_2 , pa ako je a^2 realizovana vrednost test-statistike A^2 , na osnovu posledice 5.1 p-vrednost se približno izračunava kao

$$p = P(A^2 \geq a^2 | H_0).$$

Test zona

Test-statistika testa je zapravo statistika A koja je korišćena kao procena stanja procesa pod kontrolom kod Šuhartove \bar{X} karte. Statistika je bazirana na pravilu 3σ , tj. funkciji zone date definicijom 4.4. U definiciji 4.4 korišćeni su brojevi 1, 2, 3 i 3,1947 pri definiciji funkcije. Kako raspodela statistike A zavisi direktno od raspodele koja se dobija primenom funkcije zone na elemente uzorka, u nastavku će biti korišćeni brojevi 1, 1,95, 2,58 i 2,81, tj. prilagođena varijanta pravila 3σ . Razlog tome je činjenica da sa prvim skupom brojeva dva različita uzorka mogu dati iste vrednosti statistike A . Na primer, važi $1+1+3=1+2+2$. Stoga je bolje uvesti novu definiciju funkcije zone. Drugačija definicija je korišćena u prethodnom poglavlju radi ilustracije više mogućnosti. Dodatno, navedeni nedostatak biva nadomešten pravilima tumačenja kontrolnih karti.

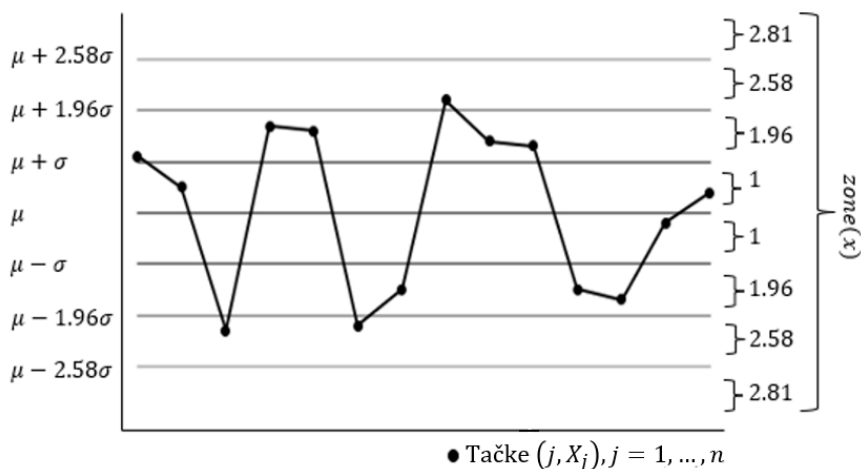
Neka je $X_1, \dots, X_n, n \in \mathbb{N}$ prost slučajan uzorak obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$.

Definicija 5.4. Funkcija zone normalne raspodele $\mathcal{N}(\mu, \sigma^2)$ data je sa

$$zone(x) = \begin{cases} 1, & \mu - \sigma \leq x \leq \mu + \sigma \\ 1,96, & \mu - 1,96\sigma \leq x < \mu - \sigma \vee \mu + \sigma < x \leq \mu + 1,96\sigma \\ 2,58, & \mu - 2,58\sigma \leq x < \mu - 1,96\sigma \vee \mu + 1,96\sigma < x \leq \mu + 2,58\sigma \\ 2,81, & x < \mu - 2,58\sigma \vee x > \mu + 2,58\sigma \end{cases} \quad x \in \mathbb{R}.$$

Na sledećoj slici je grafički prikazan način definisanja funkcije zone.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.3. Grafička ilustracija funkcije zone.

Napomena 5.1. Izbor vrednosti 2,81 nije obuhvaćen pravilom 3σ , ali prati isti princip jer važi

$$F^{-1}\left(\frac{F(2,58) + \lim_{x \rightarrow +\infty} F(x)}{2}\right) = 2,81$$

kao i

$$F^{-1}\left(\frac{\lim_{x \rightarrow -\infty} F(x) + F(-2,58)}{2}\right) = -2,81$$

gde je sa F data funkcija normalne raspodele $\mathcal{N}(0,1)$. Izbor bilo koje druge vrednosti je moguć.

Ideja za testiranje normalnosti se sastoji u tome da se funkcija zone primeni na polazni uzorak i tako se dobije uzorak mera odstupanja od očekivane vrednosti μ . Tako dobijen uzorak u ovom slučaju je

$$zone(X_1), \dots, zone(X_n). \quad (5.1)$$

Prilagođena varijanta pravila 3σ povlači sledeći stav.

Stav 5.3. Raspodela obeležja zone(X) data je sledećim zakonom raspodele

$$zone(X): \begin{pmatrix} 1 & 1,96 & 2,58 & 2,81 \\ 0,68 & 0,27 & 0,04 & 0,01 \end{pmatrix}.$$

Dokaz. Tvrđenje sledi jer važi

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0,6826; \\ P(\mu - 1,96\sigma \leq X < \mu - \sigma) + P(\mu + \sigma < X \leq \mu + 1,96\sigma) &= 0,2673; \\ P(\mu - 2,58\sigma \leq X < \mu - 1,96\sigma) + P(\mu + 1,96\sigma < X \leq \mu + 2,58\sigma) &= 0,0401; \\ P(X > \mu + 2,58) + P(X < \mu - 2,58) &= 0,0099. \blacksquare \end{aligned}$$

Definicija 5.5. Za statistiku

$$A = \frac{1}{n} \sum_{i=1}^n zone(X_i)$$

kaže se da ima *raspodelu zona sa n stepeni slobode*, u oznaci $A \sim zone_n$.

Teorema 5.1. *Raspodela zona sa n stepeni slobode data je sledećim zakonom raspodele*

$$\begin{aligned} P\left(A = \frac{1}{n} \sum C\left(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}}\right)\right) \\ = \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} \cdot 0,68^{n_1^{\{k\}}} \cdot 0,27^{n_{1,96}^{\{k\}}} \cdot 0,04^{n_{2,58}^{\{k\}}} \cdot 0,01^{n_{2,81}^{\{k\}}}, \\ k = 1, \dots, \binom{n+3}{n} \end{aligned}$$

gde je

$$\begin{aligned} C(n_1, n_{1,96}, n_{2,58}, n_{2,81}) \\ = \left\{ \underbrace{1, 1, \dots, 1}_{n_1}, \underbrace{1,96, 1,96, \dots, 1,96}_{n_{1,96}}, \underbrace{2,58, 2,58, \dots, 2,58}_{n_{2,58}}, \underbrace{2,81, 2,81, \dots, 2,81}_{n_{2,81}} \right\} \end{aligned}$$

$i \sum C\left(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}}\right)$ je suma elemenata tih skupova. Broj k je redosled referentne sume u nizu svih suma sredenih u neopadajući poredak.

Dokaz. Da bi se raspodela adekvatno opisala potrebno je koristiti kombinacije sa ponavljanjem elemenata skupa $\{1, 1,96, 2,58, 2,81\}$. Razlog je to što uzorak (5.1) čini upravo takvu kombinaciju jer redosled elemenata ne utiče na vrednost statistike A . Dakle, modaliteti statistike A su sume elemenata kombinacija sa ponavljanjem klase n skupa $\{1, 1,96, 2,58, 2,81\}$. To znači da je broj modaliteta obeležja (statistike) A

$$\binom{4+n-1}{n} = \binom{n+3}{n} = \frac{(n+3)!}{n! \cdot 3!}.$$

Navedene kombinacije su oblika

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$C(n_1, n_{1,96}, n_{2,58}, n_{2,81}) = \left\{ \underbrace{1, 1, \dots, 1}_{n_1}, \underbrace{1,96, 1,96, \dots, 1,96}_{n_{1,96}}, \underbrace{2,58, 2,58, \dots, 2,58}_{n_{2,58}}, \underbrace{2,81, 2,81, \dots, 2,81}_{n_{2,81}} \right\},$$

pa je modalitet, tj. vrednost statistike A dobijena iz takve kombinacije

$$a_k = \frac{1}{n} \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}}), \quad k = 1, \dots, \binom{n+3}{n}.$$

Redosled elemenata u uzorku ne utiče na vrednost statistike A , ali, kako se taj redosled može menjati za više modaliteta, to on utiče na verovatnoću dobijanja referentne vrednosti. Stoga se za izačunavanje verovatnoća zakona raspodele zona moraju brojati permutacije sa ponavljanjem dužine n sačinjene od elemenata skupa $\{1, 1,96, 2,58, 2,81\}$. Tačnije, broje se permutacije elemenata skupova $C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})$, jer je svaki jednoznačno vezan sa modalitet $a_k, k = 1, \dots, \binom{n+3}{n}$. Broj permutacija takvog skupa je

$$\frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!}.$$

Nezavisnost elemenata polaznog uzorka povlači i nezavisnost elemenata uzorka (5.1), pa za $s_k \in \{1, 1,96, 2,58, 2,81\}, k = 1, \dots, n$ važi

$$\begin{aligned} P(\text{zone}(X_1) = s_1; \text{zone}(X_2) = s_2; \dots; \text{zone}(X_n) = s_n) \\ = P(\text{zone}(X_1) = s_1) \cdot P(\text{zone}(X_2) = s_2) \cdots P(\text{zone}(X_n) = s_n). \end{aligned}$$

Tada je na osnovu navedenog, i stava 5.3

$$\begin{aligned} P(nA = \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})) \\ = P\left(\sum_{i=1}^n \text{zone}(X_i) = \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})\right) \\ = \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} P\left(\bigcap_{i=1}^n \text{zone}(X_i) = C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})(i)\right) \\ = \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} P\left(\bigcap_{i=1}^{n_1^{\{k\}}} \text{zone}(X_i) = 1 \cap \bigcap_{i=n_1^{\{k\}}}^{n_{1,96}^{\{k\}}} \text{zone}(X_i) = 1,96 \right. \\ \left. \cap \bigcap_{i=n_{1,96}^{\{k\}}}^{n_{2,58}^{\{k\}}} \text{zone}(X_i) = 2,58 \cap \bigcap_{i=n_{2,58}^{\{k\}}}^{n_{2,81}^{\{k\}}} \text{zone}(X_i) = 2,81\right) \end{aligned}$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned}
 &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} \prod_{i=1}^{n_1^{\{k\}}} P(\text{zone}(X_i) = 1) \prod_{i=n_1^{\{k\}}}^{n_{1,96}^{\{k\}}} P(\text{zone}(X_i) \\
 &= 1,96) \prod_{i=n_{1,96}^{\{k\}}}^{n_{2,58}^{\{k\}}} P(\text{zone}(X_i) = 2,58) \prod_{i=n_{2,58}^{\{k\}}}^{n_{2,81}^{\{k\}}} P(\text{zone}(X_i) = 2,81) \\
 &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} \cdot 0,68^{n_1^{\{k\}}} \cdot 0,27^{n_{1,96}^{\{k\}}} \cdot 0,04^{n_{2,58}^{\{k\}}} \cdot 0,01^{n_{2,81}^{\{k\}}}.
 \end{aligned}$$

Konačno se dobija

$$\begin{aligned}
 P(nA = \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})) &= P\left(A = \frac{1}{n} \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})\right) \\
 &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} \cdot 0,68^{n_1^{\{k\}}} \cdot 0,27^{n_{1,96}^{\{k\}}} \cdot 0,04^{n_{2,58}^{\{k\}}} \cdot 0,01^{n_{2,81}^{\{k\}}}
 \end{aligned}$$

što dokazuje teoremu. ■

Posledica 5.2. Neka je

$$p_k = \frac{n!}{n_1^{\{k\}}! \cdot n_{1,96}^{\{k\}}! \cdot n_{2,58}^{\{k\}}! \cdot n_{2,81}^{\{k\}}!} \cdot 0,68^{n_1^{\{k\}}} \cdot 0,27^{n_{1,96}^{\{k\}}} \cdot 0,04^{n_{2,58}^{\{k\}}} \cdot 0,01^{n_{2,81}^{\{k\}}},$$

$$k = 1, \dots, \binom{n+3}{n}.$$

Funkcija raspodele obeležja $A \sim \text{zone}_n$ je

$$F(x) = P(A \leq x) = \begin{cases} 0; & x < a_1 \\ \sum_{i=1}^k p_i, & a_k \leq x < a_{k+1}, k = 1, \dots, \binom{n+3}{n} - 1 \\ 1; & x \geq a_{\binom{n+3}{n}} \end{cases}$$

gde je $a_k = \frac{1}{n} \sum C(n_1^{\{k\}}, n_{1,96}^{\{k\}}, n_{2,58}^{\{k\}}, n_{2,81}^{\{k\}})$, $k = 1, \dots, \binom{n+3}{n}$.

Posledica 5.3. Raspodela statistike A se ne menja pri promeni parametara raspodele obeležja X kada su ti parametri poznati.

Dokaz. Na osnovu stava 5.3, raspodela obeležja $\text{zone}(X_i)$, $i = 1, \dots, n$ je

$$\text{zone}(X): \begin{pmatrix} 1 & 1,96 & 2,58 & 2,81 \\ 0,68 & 0,27 & 0,04 & 0,01 \end{pmatrix},$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

što sledi iz pravila 3σ . Ako bi raspodela X opet bila normalna, ali sa promenjenim parametrima, to bi pravilo 3σ i tada važi, ali za nove parametre. Stoga raspodela $zone(X_i), i = 1, \dots, n$ ostaje nepromenjena, a samim tim i raspodela statistike A . ■

Izračunavanje $F(x), x \in \mathbb{R}$ ili $F^{-1}(p), p \in [0,1]$ je komplikovano čak i za programiranje pa je praktičnije izdvojiti tabelu raspodele, tj. kvantila za razne verovatnoće i obime uzorka. Raspodela $zone_n$ je očigledno diskretna pa $F^{-1}(p)$ ne postoji uvek. Takođe, za veće n broj modaliteta je veći, pa formiranje tabele korišćenjem posledice 5.2 postaje dodatno otežano. Stoga će za formiranje tabele raspodele biti korišćene Monte-Karlo simulacije. Za svako n , primenom programskog paketa MATLAB sprovedeno je po 100000 simulacija statistike A nad modeliranim uzorcima normalne raspodele.

Lilifors (1967) i Stivens (1974) su ukazali na to da se raspodela test-statistike može primetno promeniti kada se parametri raspodele obeležja iz nulte hipoteze ocenjuju. Stoga je ovde simuliranje raspodele statistike A vršeno odvojeno za slučaj kada su parametri normalne raspodele poznati (testiranje za prostu nultu hipotezu) i za slučaj kada nisu poznati pa se ocenjuju (testiranje za složenu nultu hipotezu). U radu u kojem je raspodela zone uvedena, raspodela oba slučaja je korišćena za testiranje proste nulte hipoteze (Avdović & Jevremović, 2023), stoga je doprinos ovog rada i taj što će se performanse testa zona analizirati i za složenu nultu hipotezu.

Simulacije su vršene sa parametrima 0 i 1 normalne raspodele, ali se promenom parametara rezultat ne menja, što posledica 5.3 i tvrdi.

Tabela 5.2. Raspodela zona – Poznati parametri. $F(x) = P(A \leq x); A \sim zone_n$.

n	p						
	0,01	0,025	0,05	0,1	0,15	0,2	0,5
5	*	*	*	*	1,0011	1,0339	1,3840
10	*	1,0038	1,0315	1,0870	1,1840	1,1897	1,3393
20	1,0925	1,1391	1,1440	1,1914	1,2199	1,2385	1,3334
30	1,1280	1,1600	1,1918	1,2220	1,2447	1,2557	1,3370
50	1,1728	1,2016	1,2233	1,2428	1,2620	1,2744	1,3366
100	1,2228	1,2394	1,2552	1,2716	1,2840	1,2936	1,3366
200	1,2552	1,2680	1,2789	1,2916	1,3004	1,3072	1,3376
300	1,2703	1,2806	1,2897	1,2999	1,3070	1,3127	1,3375
500	1,2853	1,2933	1,3003	1,3084	1,3140	1,3184	1,3378
1000	1,3002	1,3061	1,3111	1,3170	1,3210	1,3241	1,3378
1500	1,3077	1,3121	1,3162	1,3209	1,3242	1,3267	1,3379
2000	1,3112	1,3154	1,3190	1,3231	1,3259	1,3281	1,3378

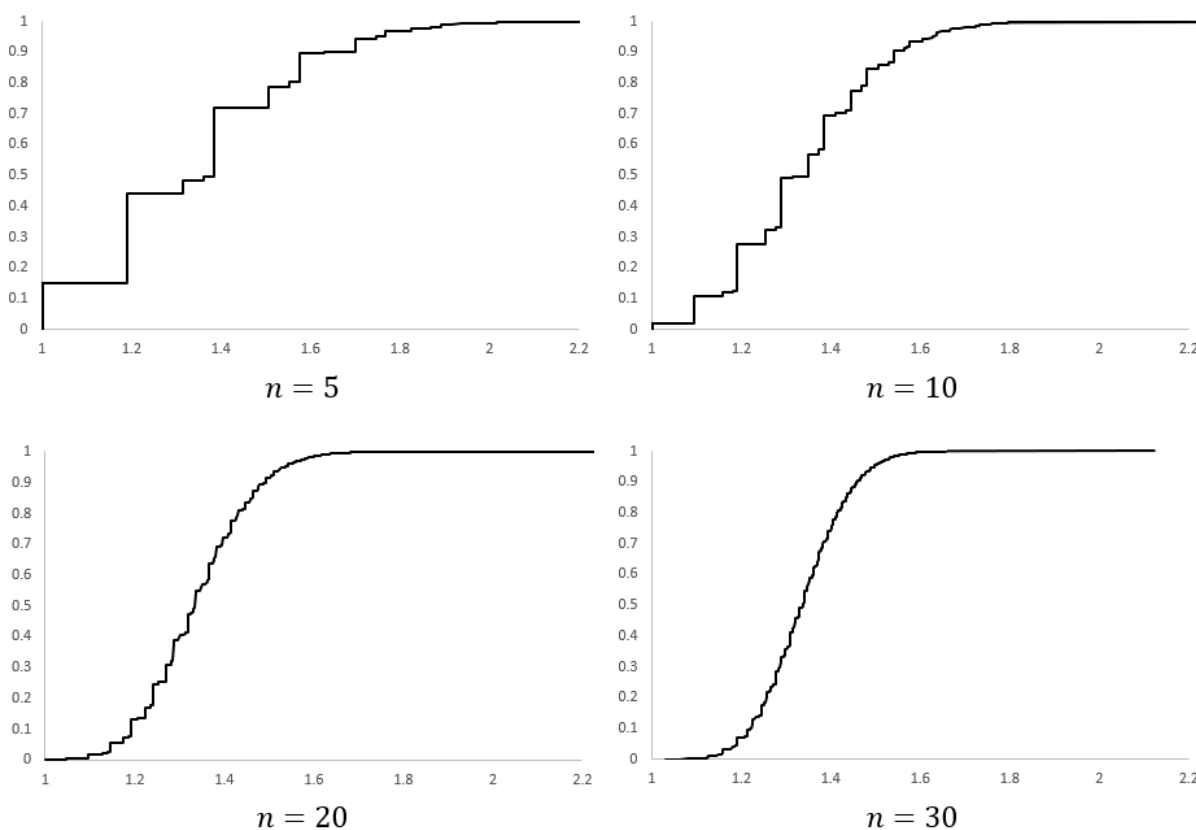
*Svaki modalitet ima verovatnoću veću od zadate.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

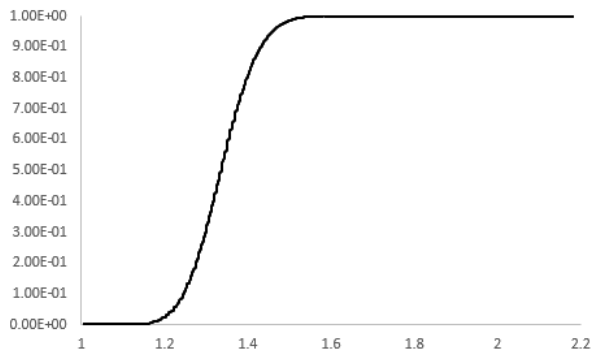
Tabela 5.2. Nastavak.

n	p					
	0,8	0,85	0,9	0,95	0,975	0,99
5	1,5540	1,5651	1,6382	1,7421	1,8198	1,9201
10	1,4749	1,5080	1,5420	1,6270	1,6890	1,7518
20	1,4303	1,4612	1,4940	1,5365	1,5730	1,6206
30	1,4148	1,4359	1,4573	1,4970	1,5290	1,5647
50	1,3990	1,4134	1,4325	1,4596	1,4846	1,5149
100	1,3811	1,3912	1,4038	1,4235	1,4405	1,4602
200	1,3684	1,3756	1,3849	1,3984	1,4108	1,4249
300	1,3627	1,3686	1,3759	1,3872	1,3968	1,4084
500	1,3572	1,3671	1,3675	1,3760	1,3835	1,3925
1000	1,3515	1,3547	1,3588	1,3647	1,3699	1,3762
1500	1,3490	1,3516	1,3548	1,3597	1,3641	1,3690
2000	1,3476	1,3499	1,3527	1,3569	1,3605	1,3650

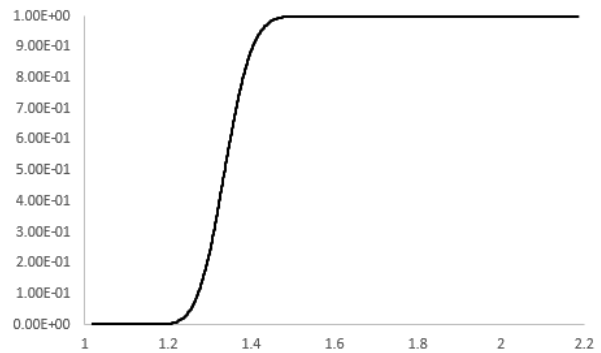
Sledećom slikom dat je grafik funkcije raspodele zona za razne obime uzorka.



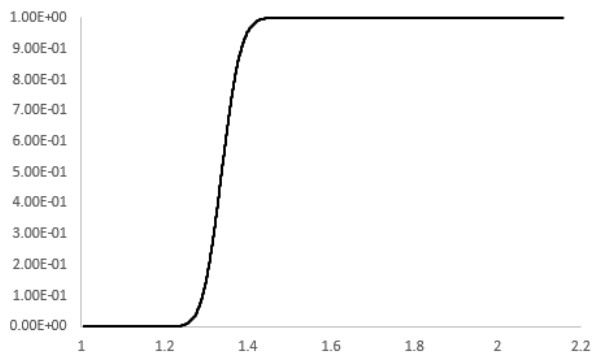
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



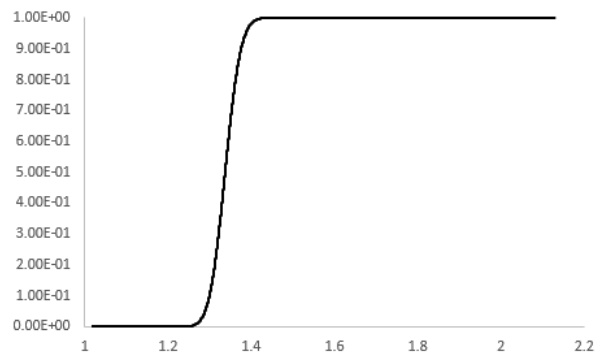
$n = 50$



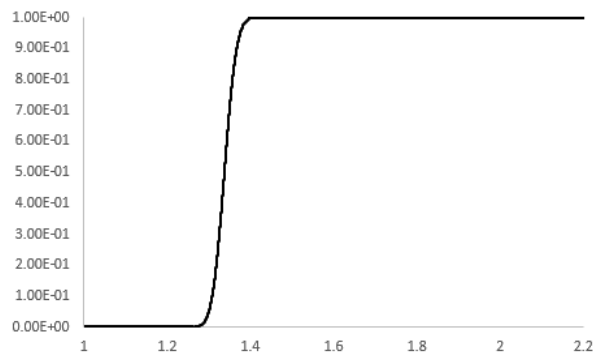
$n = 100$



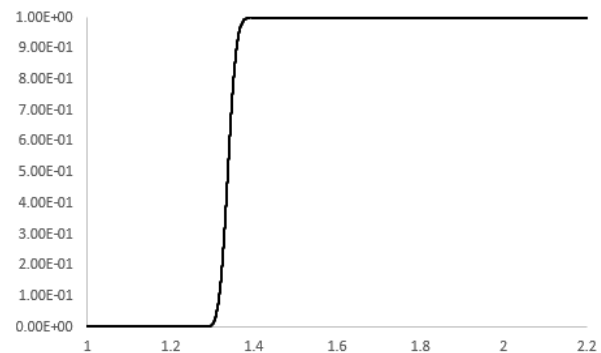
$n = 200$



$n = 300$

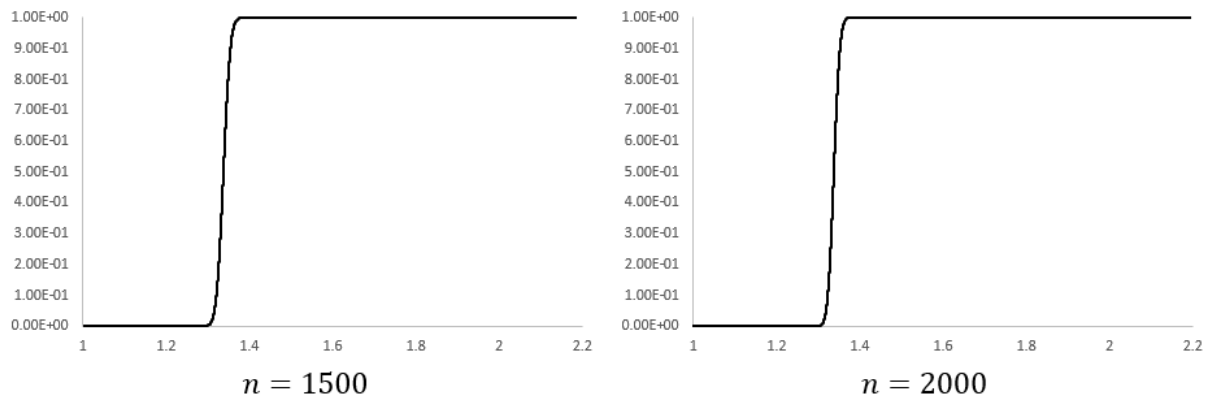


$n = 500$



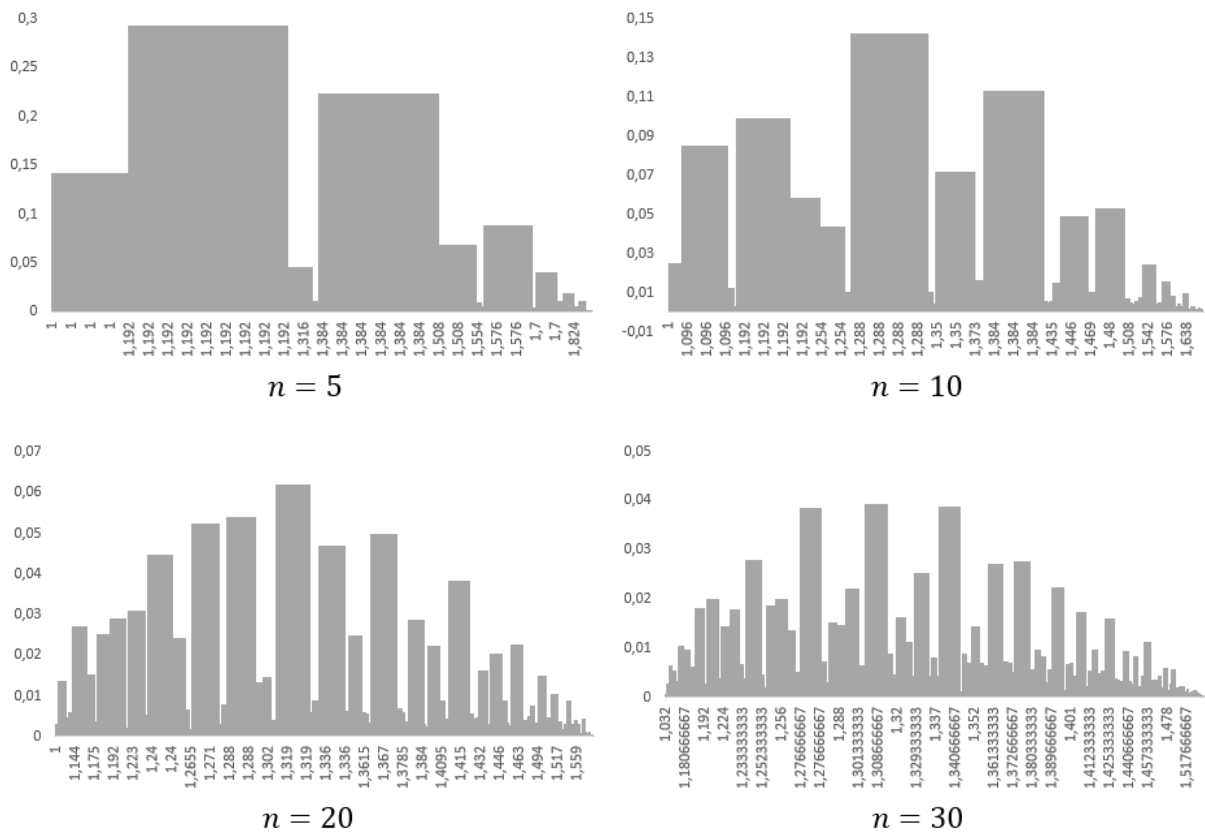
$n = 1000$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

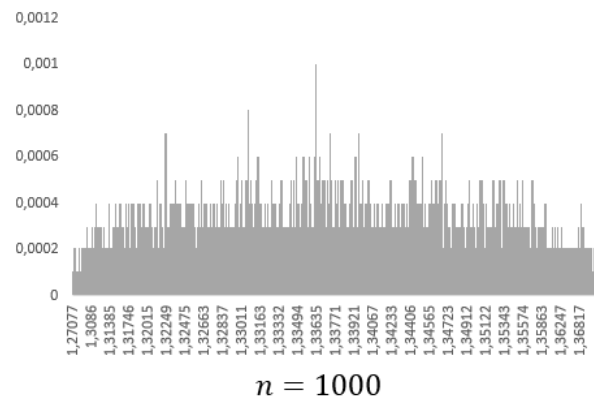
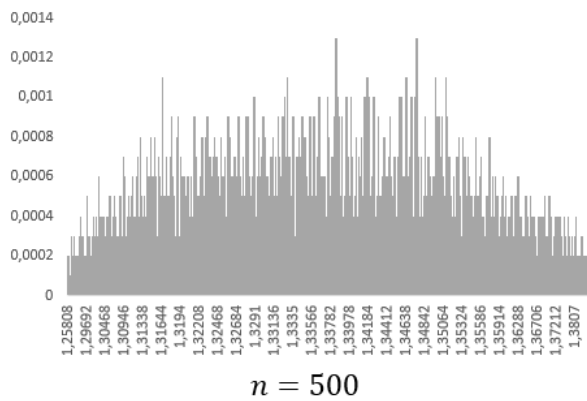
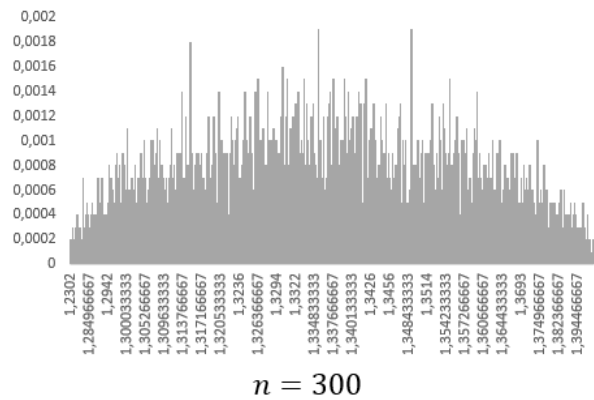
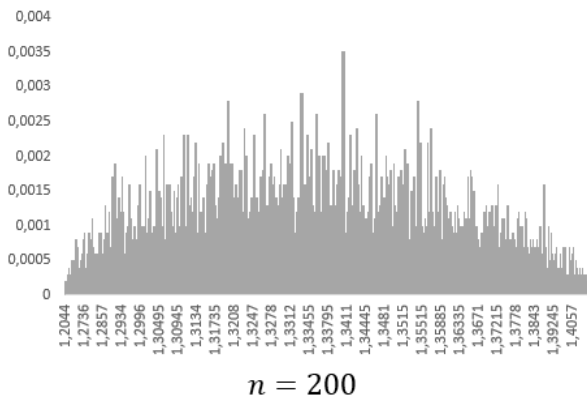
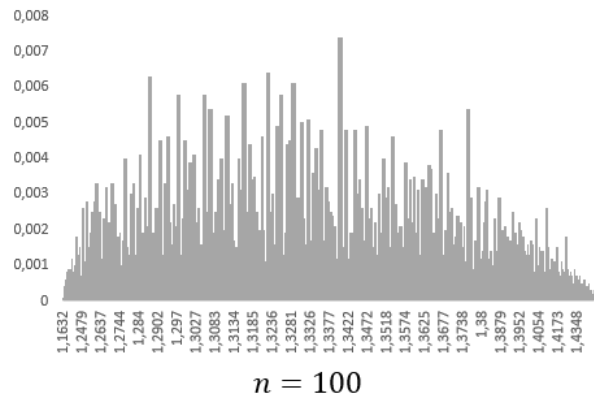
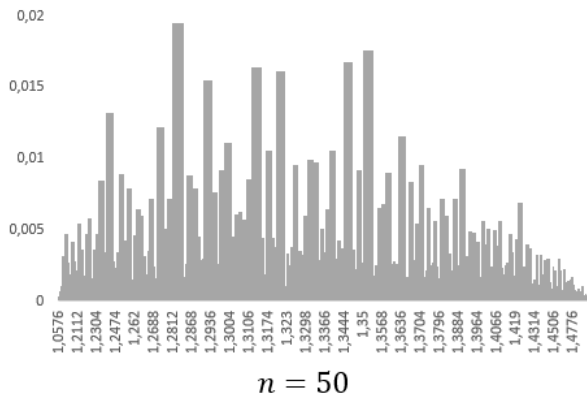


Slika 5.4. Grafik funkcije raspodele zone_n za poznate parametre normalne raspodele.

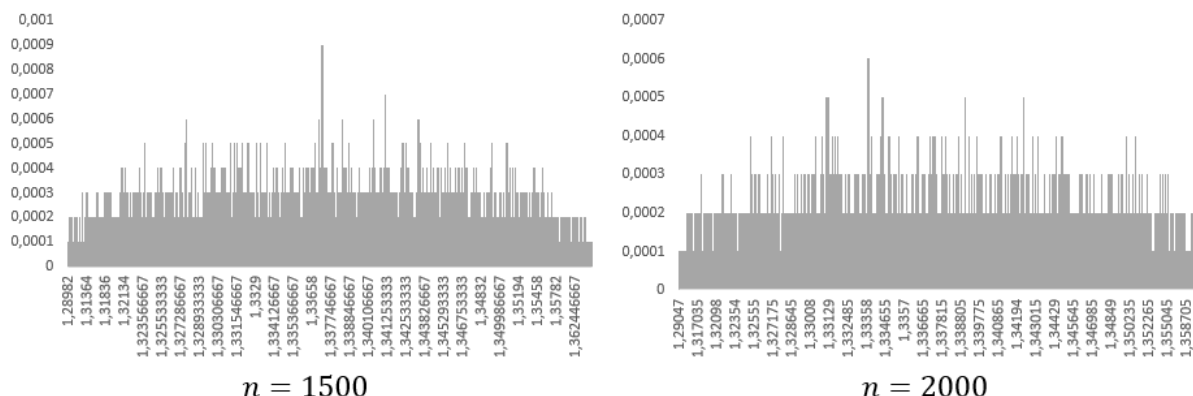
Sledećom slikom dat je grafički prikaz zakona raspodele zona, takođe za razne obime uzorka.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.5. Zakon raspodele zone_n dat histogramom za poznate parametre normalne raspodele.

U nastavku je dat detaljniji opis raspodele zona za slučaj kada su parametri raspodele obeležja X poznati.

Posledica 5.4. Očekivanje i disperzija obeležja zone(X) su redom $E(\text{zone}(X)) = 1,3405$ i $D(\text{zone}(X)) = 0,2655$.

Dokaz. Sledi direktno iz stava 5.3. ■

Stav 5.4. Očekivanje, disperzija, koeficijent asimetrije i koeficijent spljoštenosti obeležja A sa raspodelom zone_n su redom

$$E(A) = 1,3405, D(A) = \frac{0,2655}{n},$$

$$\text{Skew}(A) = \frac{35,2148n^2 - 0,0015n + 1,0636}{\sqrt{n}} i$$

$$\text{Kurt}(A) = \frac{-0,0071n^3 + 375,8936n^2 + 543,6709n - 779,4681}{n}.$$

Dokaz. Koristeći osnovne osobine očekivanja i disperzije dobijaju se navedeni rezultati.

$$E(A) = E\left(\frac{1}{n} \sum_{i=1}^n \text{zone}(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(\text{zone}(X_i)) = \frac{1}{n} \cdot n \cdot 1,3405 = 1,3405.$$

$$D(A) = D\left(\frac{1}{n} \sum_{i=1}^n \text{zone}(X_i)\right) = \frac{1}{n^2} \sum_{i=1}^n D(\text{zone}(X_i)) = \frac{1}{n^2} n \cdot 0,2655 = \frac{0,2655}{n}.$$

Iz $D(A) = E(A^2) - (E(A))^2$ sledi i

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$E(A^2) = \frac{0,2655}{n} + 1,7969.$$

Na osnovu stava 5.3 dobija se

$$E\left(\left(\text{zone}(X_k)\right)^3\right) = 3,6218, \quad k = 1, \dots, n$$

Broj varijacija oblika (i, j) , $i, j \in \{1, \dots, n\}$, $i \neq j$ je $n(n-1)$, a broj kombinacija oblika $\{i, j, k\} \subset \{1, \dots, n\}$ je $\binom{n}{3} = \frac{n!}{6(n-3)!}$. Sledi

$$\begin{aligned} E(A^3) &= E\left(\left(\frac{1}{n} \sum_{i=1}^n \text{zone}(X_i)\right)^3\right) \\ &= \frac{1}{n^3} \left(\sum_{i=1}^n E\left(\left(\text{zone}(X_i)\right)^3\right) + 3 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E\left(\left(\text{zone}(X_i)\right)^2\right) E\left(\text{zone}(X_j)\right) \right. \\ &\quad \left. + 6 \sum_{\{i,j,k\} \subset \{1,\dots,n\}} E\left(\text{zone}(X_i)\right) E\left(\text{zone}(X_j)\right) E\left(\text{zone}(X_k)\right) \right) \\ &= \frac{1}{n^3} \left(\sum_{i=1}^n 3,6218 + 3 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n 2,0624 \cdot 1,3405 + 6 \sum_{\{i,j,k\} \subset \{1,\dots,n\}} 1,3405^3 \right) \\ &= \frac{1}{n^3} \left(3,6218 \cdot n + 3 \cdot n(n-1) \cdot 2,0624 \cdot 1,3405 + 6 \binom{n}{3} 1,3405^3 \right) \\ &= \frac{2,4088 \cdot n^2 + 1,0675 \cdot n + 0,1455}{n^2}. \end{aligned}$$

Stavljanjem dobijenih $E(A)$, $D(A)$, $E(A^2)$ i $E(A^3)$ u formulu za koeficijent asimetrije

$$\begin{aligned} \text{Skew}(A) &= E\left(\frac{A - E(A)}{\sqrt{D(A)}}\right)^3 \\ &= \frac{E(A^3) - E(A)\left(3E(A^2) - 2(E(A))^2\right)}{(D(A))^{\frac{3}{2}}} \end{aligned}$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned}
 &= \frac{1}{\left(\frac{0,2655}{n}\right)^{\frac{3}{2}}} \left(\frac{1}{n^2} (2,4088 \cdot n^2 + 1,0675 \cdot n + 0,1455) \right. \\
 &\qquad \qquad \qquad \left. - 1,3405 \left(3 \frac{0,2655}{n} + 1,7969 - 2 \cdot 1,7969 \right) \right) \\
 &= \frac{35,2148 \cdot n^2 - 0,0015 \cdot n + 1,0636}{\sqrt{n}}.
 \end{aligned}$$

Analogno, dobija se

$$E\left((zone(X_k))^4\right) = 7,0604, \quad k = 1, \dots, n$$

kao i

$$\begin{aligned}
 E(A^4) &= E\left(\left(\frac{1}{n} \sum_{i=1}^n zone(X_i)\right)^4\right) \\
 &= \frac{1}{n^4} \left(\sum_{i=1}^n E\left((zone(X_i))^4\right) + 4 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E\left((zone(X_i))^3\right) E\left(zone(X_j)\right) \right. \\
 &\qquad \qquad \qquad + 6 \sum_{\{i,j\} \subset \{1,\dots,n\}} E\left((zone(X_i))^2\right) E\left((zone(X_j))^2\right) \\
 &\qquad \qquad \qquad + 12 \sum_{i=1}^n \sum_{\substack{\{j,k\} \subset \{1,\dots,n\} \\ j,k \neq i}} \left((zone(X_i))^3\right) E\left(zone(X_j)\right) E\left(zone(X_k)\right) \\
 &\qquad \qquad \qquad \left. + 24 \sum_{\{i,j,k,l\} \subset \{1,\dots,n\}} E\left(zone(X_i)\right) E\left(zone(X_j)\right) E\left(zone(X_k)\right) E\left(zone(X_l)\right) \right) \\
 &= \frac{1}{n^4} \left(n \cdot 7,0604 + 4 \cdot n(n-1) \cdot 3,6218 \cdot 2,0624 + 6 \binom{n}{2} 2,0624^2 + 12 \cdot n \binom{n-1}{2} \right. \\
 &\qquad \qquad \qquad \left. \cdot 3,6218 \cdot 1,3405^2 + 24 \binom{n}{4} \cdot 1,3405^4 \right) \\
 &= \frac{3,2290 \cdot n^3 + 29,3619 \cdot n^2 + 39,1090 \cdot n - 54,9525}{n^3}.
 \end{aligned}$$

Stavljanjem dobijenih $E(A)$, $D(A)$, $E(A^2)$, $E(A^3)$ i $E(A^4)$ u formulu za koeficijent spljoštenosti dobija se

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned}
\text{Kurt}(A) &= E\left(\frac{A - E(A)}{\sqrt{D(A)}}\right)^4 \\
&= \frac{E(A^4) - 4E(A^3)E(A) + 6E(A^2)(E(A))^2 - 3(E(A))^4}{(D(A))^2} \\
&= \left(\frac{3,2290 \cdot n^3 + 29,3619 \cdot n^2 + 39,1090 \cdot n - 54,9525}{n^3} - 4 \frac{2,4088 \cdot n^2 + 1,0675 \cdot n + 0,1455}{n^2} + 6 \left(\frac{0,2655}{n} + 1,7969\right) 1,3405^2 - 3 \cdot 1,3405^4\right) \left(\frac{n^2}{0,2655^2}\right) \\
&= \frac{-0,0071 \cdot n^3 + 375,8936 \cdot n^2 + 543,6709 \cdot n - 779,4681}{n}. \blacksquare
\end{aligned}$$

Kada su parametri raspodele nepoznati, mogu se ocenjivati metodom maksimalne verodostojnosti. Kako je $X \sim \mathcal{N}(\mu, \sigma^2)$, ocene za μ i σ^2 su redom $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ i $\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Ipak, za ocenu σ^2 uzima se $\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ jer je to nepristrasna ocena za σ^2 , za razliku od \bar{S}_n^2 .

Teorema 5.2. *Raspodela statistike A , kada su parametri obeležja X nepoznati i umesto njih se koriste ocene \bar{X}_n i \tilde{S}_n^2 , se ne menja pri promeni parametara obeležja X .*

Dokaz. Neka je raspodela obeležje X normalna $\mathcal{N}(\mu_1, \sigma_1^2)$ gde su parametri μ_1 i σ_1^2 nepoznati i X_1, \dots, X_n prost slučajan uzorak obeležja X . Tada $\bar{X}_n \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n}\right)$ i $\frac{(n-1)\tilde{S}_n^2}{\sigma_1^2} \sim \chi_{n-1}^2$ (Jevremović, 2014, str. 212). Neka su $\hat{\mu} = \bar{X}_n$ i $\hat{\sigma}^2 = \tilde{S}_n^2$ i

$$\begin{aligned}
I_1 &= [\hat{\mu} - \hat{\sigma}, \hat{\mu} + \hat{\sigma}], \\
I_2 &= [\hat{\mu} - 1,96\hat{\sigma}, \hat{\mu} - \hat{\sigma}] \cup (\hat{\mu} + \hat{\sigma}, \hat{\mu} + 1,96\hat{\sigma}), \\
I_3 &= [\hat{\mu} - 2,58\hat{\sigma}, \hat{\mu} - 1,96\hat{\sigma}] \cup (\hat{\mu} + 1,96\hat{\sigma}, \hat{\mu} + 2,58\hat{\sigma}), \\
I_4 &= (-\infty, \hat{\mu} - 2,58\hat{\sigma}) \cup (\hat{\mu} + 2,58\hat{\sigma}, +\infty).
\end{aligned}$$

Tada se za

$$\text{zone}_{\hat{\mu}, \hat{\sigma}}(x) = \begin{cases} 1, & x \in I_1 \\ 1,96, & x \in I_2 \\ 2,58, & x \in I_3 \\ 2,81, & x \in I_4 \end{cases}, \quad x \in \mathbb{R}.$$

dobija uzorak $\text{zone}(X_j), j = 1, \dots, n$.

Neka su dalje $Z_j = \frac{X_j - \hat{\mu}}{\hat{\sigma}}, j = 1, \dots, n, a_1 = 1, a_2 = 1,96, a_3 = 2,58, a_4 = 2,81$ i

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned} I'_1 &= [-1, 1], \\ I'_2 &= [-1,96, -1) \cup (1, 1,96], \\ I'_3 &= [-2,58, -1,96) \cup (1,96, 2,58], \\ I'_4 &= (-\infty, -2,58) \cup (2,58, +\infty). \end{aligned}$$

Tada je za $j = 1, \dots, n$ i $k = 1, 2, 3, 4$

$$\begin{aligned} P(\text{zone}_{\hat{\mu}, \hat{\sigma}}(X_j) = a_k) &= P(X_j \in I_k) = P(\hat{\sigma}Z_j + \hat{\mu} \in I_k) = P(Z_j \in I'_k) \\ &= P\left(\text{zone}_{0,1}(Z_j) = \frac{a_k - \hat{\mu}}{\hat{\sigma}}\right). \end{aligned}$$

Neka je sada raspodela obeležje X normalna $\mathcal{N}(\mu_2, \sigma_2^2)$, gde su parametri μ_2 i σ_2^2 nepoznati ali različiti od μ_1 i σ_1^2 ($(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$). Analogno je $\bar{X}_n \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n}\right)$ i $\frac{n\bar{S}_n^2}{\sigma_2^2} \sim \chi_{n-1}^2$ i za $Y_j = \frac{X_j - \hat{\mu}}{\hat{\sigma}}, j = 1, \dots, n$ važi

$$P(\text{zone}_{\hat{\mu}, \hat{\sigma}}(X_j) = a_k) = P\left(\text{zone}_{0,1}(Y_j) = \frac{a_k - \hat{\mu}}{\hat{\sigma}}\right).$$

Kako je raspodela X u odnosu na raspodele \bar{X}_n i \bar{S}_n^2 ista i u prvom i u drugom slučaju, to se raspodele Z_j i Y_j poklapaju za sve $j = 1, \dots, n$. Odatle sledi da se poklapaju i raspodele za $\text{zone}_{0,1}(Z_j)$ i $\text{zone}_{0,1}(Y_j)$, a time je raspodela $\text{zone}_{\hat{\mu}, \hat{\sigma}}(X_j)$, pa i raspodela za A ista za oba slučaja. ■

Raspodela koja se empirijski može koristiti se može odrediti samo simulacijama (jednostavnije je nego koristiti posledicu 5.2). Kako raspodela statistike A ne zavisi od parametara obeležja X ni u slučaju kada su oni nepoznati i ocenjuju se (teorema 5.2), tablica raspodele se može dobiti primenom Monte Karlo simulacija. U nastavku data je tabela u kojoj su date vrednosti statistike A za ocenjene parametre, razne obime uzorka i kvantile (verovatnoće).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Tabela 5.3. Raspodela zona – Ocenjeni parametri. $F(x) = P(A \leq x); A \sim zone_n$.

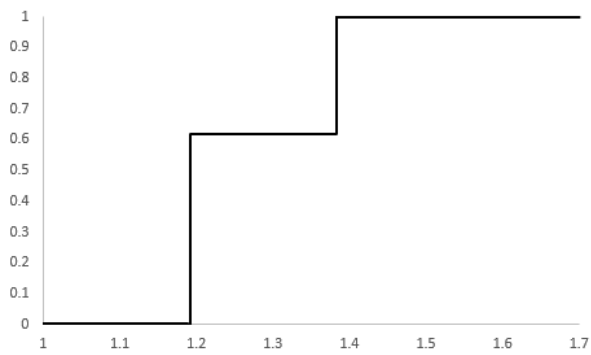
<i>n</i>	<i>p</i>						
	0,01	0,025	0,05	0,1	0,15	0,2	0,5
5	1,0049	1,0123	1,0246	1,0493	1,0739	1,0985	1,2280
10	1,1338	1,1828	1,1890	1,1986	1,2216	1,2446	1,2874
20	1,1990	1,2204	1,2345	1,2710	1,2710	1,2825	1,3257
30	1,2203	1,2410	1,2540	1,2730	1,2843	1,2860	1,3293
50	1,2508	1,2608	1,2722	1,2868	1,2970	1,3050	1,3320
100	1,2750	1,2846	1,2925	1,3021	1,3089	1,3145	1,3366
200	1,2936	1,3008	1,3067	1,3134	1,3180	1,3216	1,3371
300	1,3019	1,3076	1,3125	1,3179	1,3217	1,3247	1,3375
500	1,3100	1,3144	1,3182	1,3225	1,3254	1,3277	1,3377
1000	1,3184	1,3213	1,3240	1,3270	1,3290	1,3307	1,3377
1500	1,3219	1,3244	1,3266	1,3291	1,3307	1,3320	1,3378
2000	1,3241	1,3263	1,3281	1,3302	1,3317	1,3328	1,3378

Tabela 5.3. Nastavak.

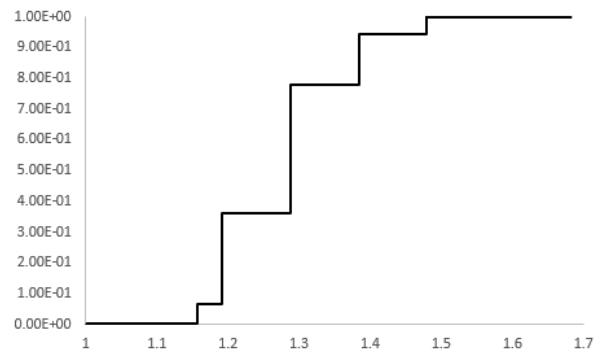
<i>n</i>	<i>p</i>					
	0,8	0,85	0,9	0,95	0,975	0,99
5	1,3263	1,3426	1,3590	1,3753	1,3836	1,4947
10	1,3694	1,3758	1,3821	1,4570	1,4695	1,4769
20	1,3813	1,3831	1,4150	1,4278	1,4314	1,4640
30	1,3727	1,3820	1,3933	1,4047	1,4253	1,4480
50	1,3636	1,3704	1,3828	1,3964	1,4088	1,4212
100	1,3586	1,3636	1,3704	1,3800	1,3868	1,3987
200	1,3530	1,3564	1,3611	1,3676	1,3735	1,3812
300	1,3503	1,3533	1,3570	1,3623	1,3674	1,3731
500	1,3477	1,3501	1,3528	1,3572	1,3608	1,3651
1000	1,3448	1,3464	1,3484	1,3514	1,3541	1,3572
1500	1,3436	1,3449	1,3466	1,3491	1,3512	1,3538
2000	1,3428	1,3440	1,3455	1,3476	1,3495	1,3516

Sledećom slikom dat je grafik funkcije raspodele zona za razne obime uzorka.

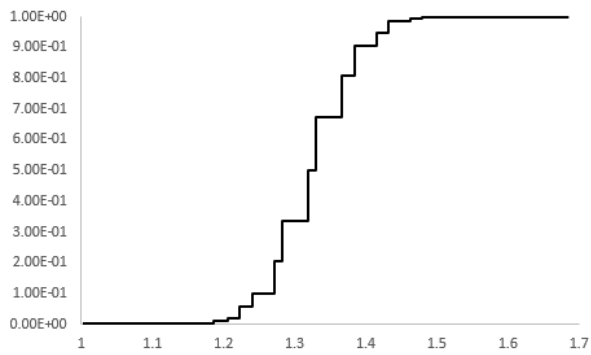
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



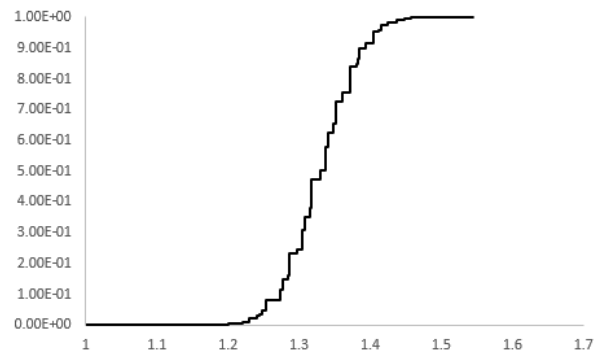
$n = 5$



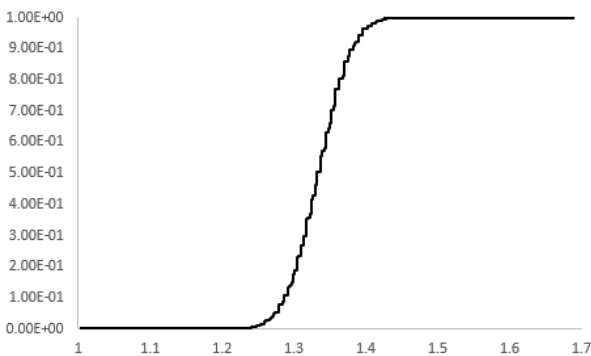
$n = 10$



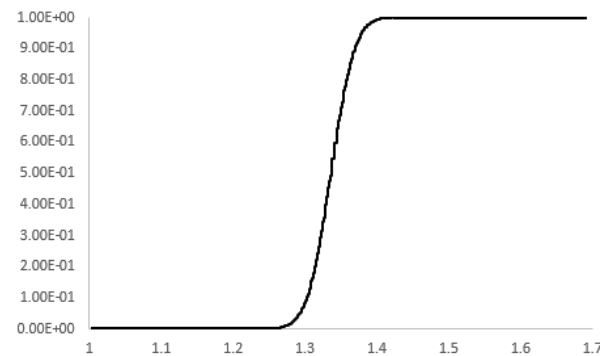
$n = 20$



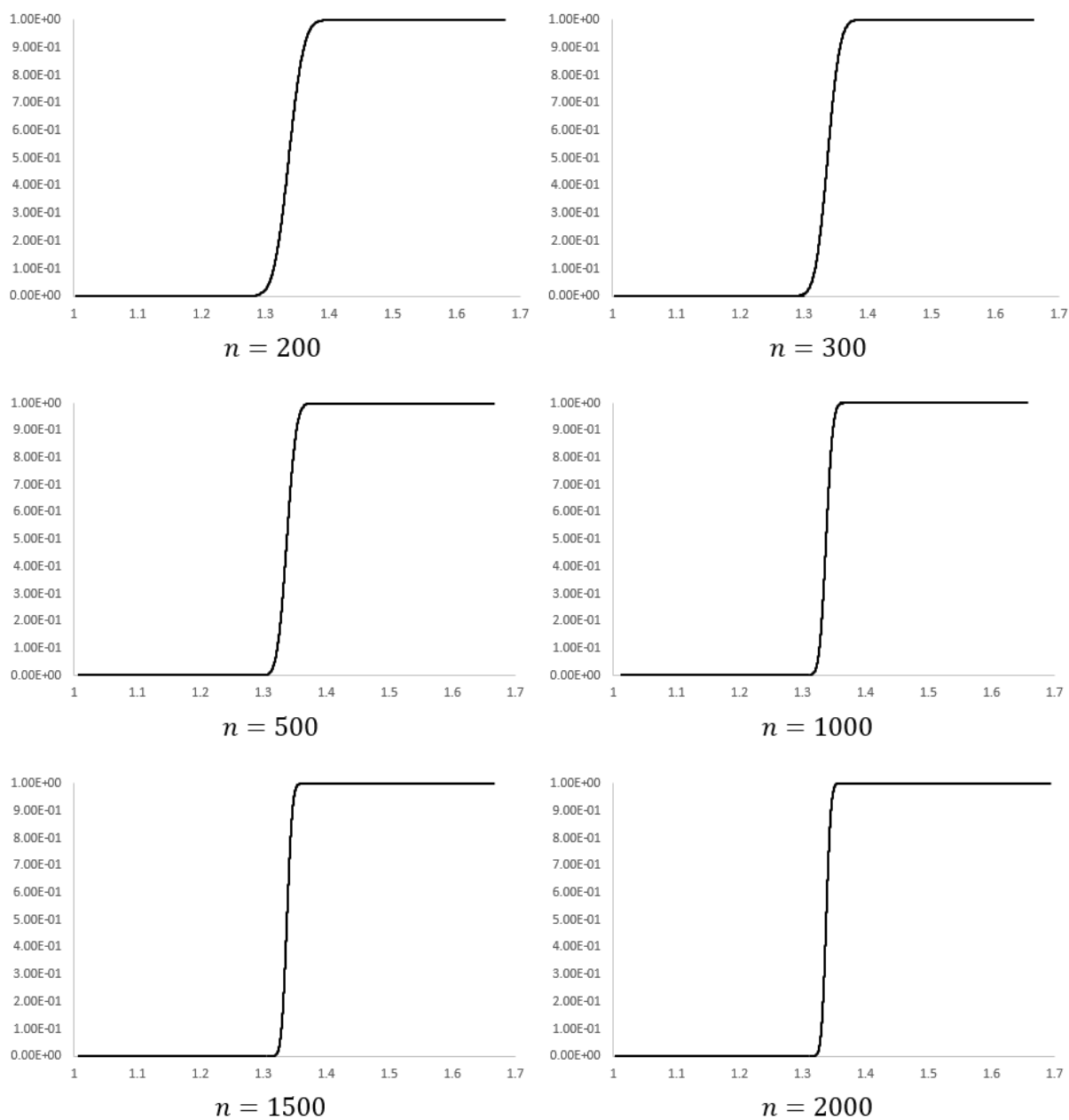
$n = 30$



$n = 50$



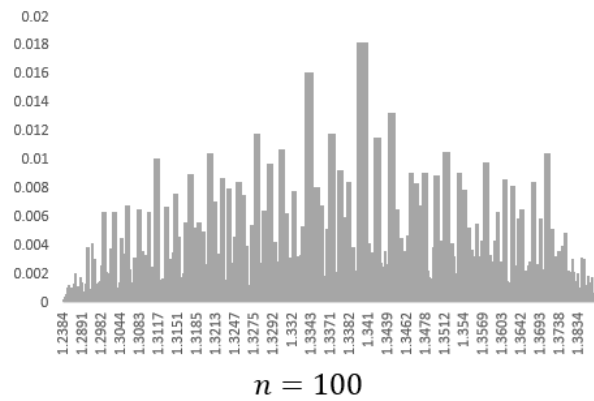
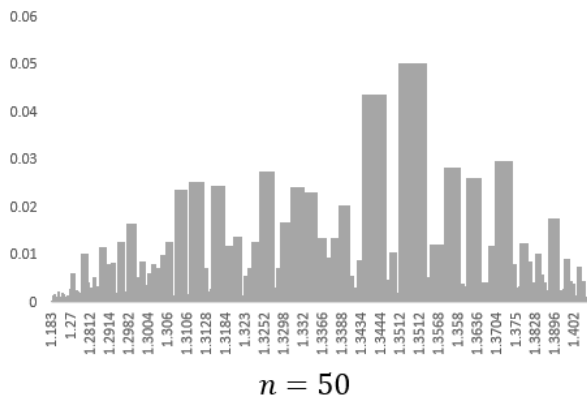
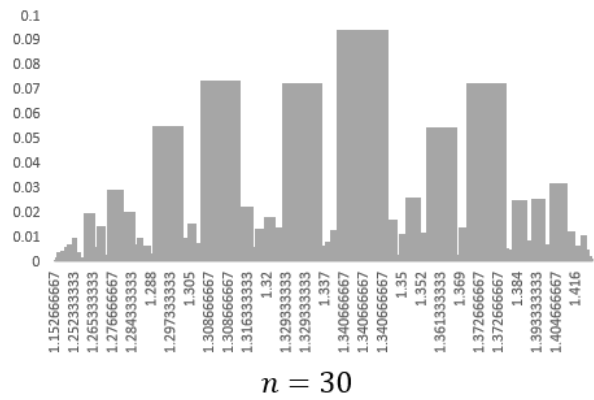
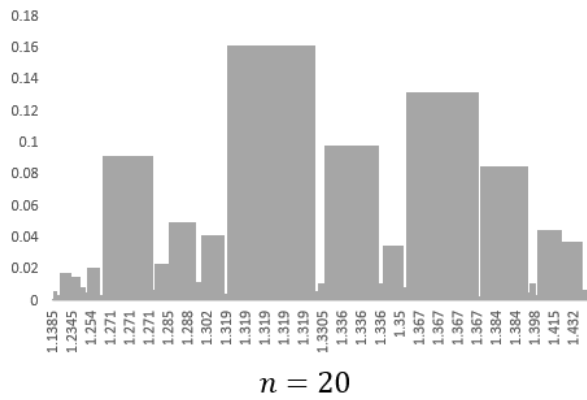
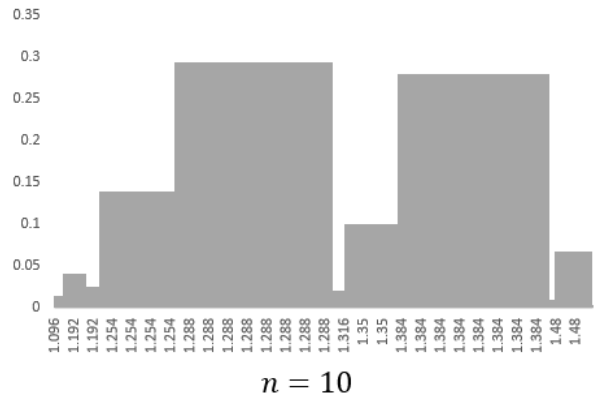
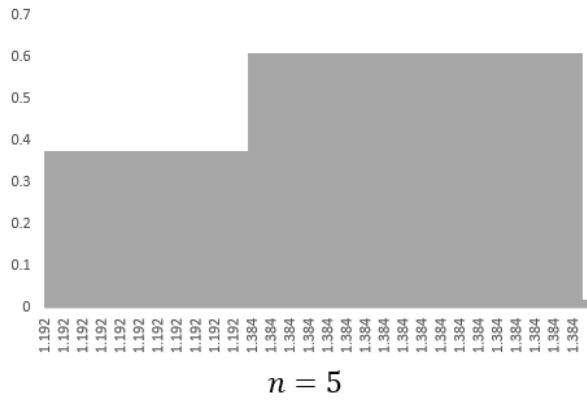
$n = 100$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

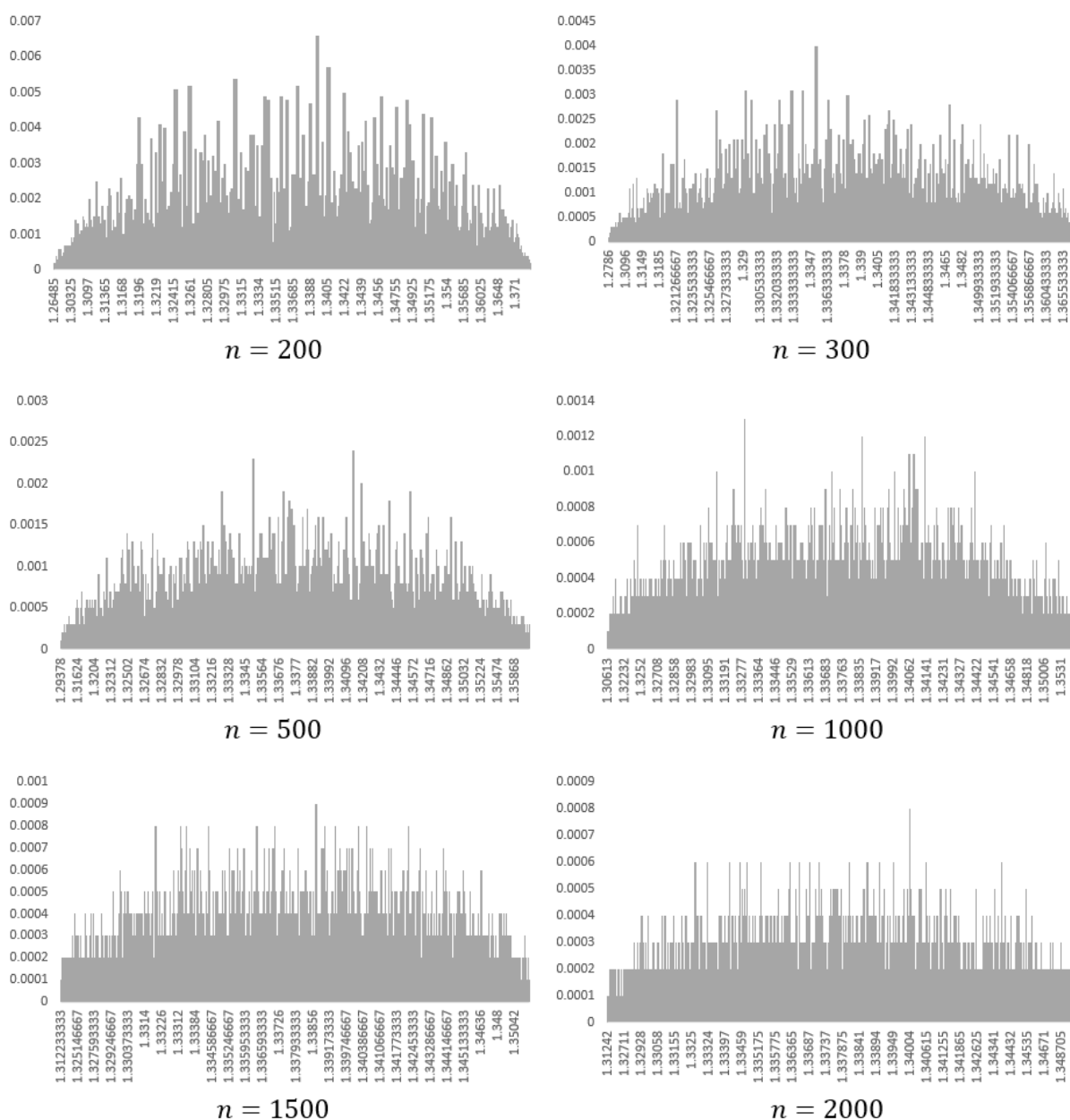
Slika 5.6. Grafik funkcije raspodele zone_n za ocenjene parametre normalne raspodele.

Sledećom slikom dat je grafički prikaz zakona raspodele zona, takođe za razne obime uzorka.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.7. Zakon raspodele zone $_n$ dat histogramom za ocenjene parametre normalne raspodele.

Napomena 5.2. Grafici funkcije raspodele su za različite obime uzorka formirani tako da su dati u odnosu na isti deo ravni pa se mogu upoređivati, tj. može se videti promena raspodele sa povećavanjem n . Kod zakona raspodele to nije moguće jer je za uzorke manjeg obima verovatnoća svakog modaliteta drastično veća u odnosu na slučajeve za veće obime uzorka, pa prikaz u odnosu na isti deo ravni za veće uzorke ne bi bio vidljiv. Dakle, mogu se upoređivati ako se uzmu u obzir dimenzije naznačene na osama.

Posmatranjem slika 5.4, 5.5, 5.6 i 5.7 može se videti da kada su parametri nepoznati i bivaju ocenjeni na osnovu uzorka, raspodela statistike A se značajno menja. Odstupanje tih

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

raspodela se smanjuje pri povećanju obima uzorka. U oba slučaja, povećanjem n disperzija A se smanjuje.

U nastavku je dato izlaganje postupka testiranja testom čija je test-statistika statistika A . Analogno nazivu raspodele statistike A , za test se kaže da je **test zona**. Kako opisivanje raspodele statistike A zavisi direktno od normalnosti raspodele polaznog uzorka X_1, X_2, \dots, X_n , tj. obeležja X , to je nulta hipoteza testa data sa

$$H_0: \text{„Raspodela obeležja } X \text{ je } \mathcal{N}(\mu, \sigma^2)\text{.“}$$

ako su od važnosti i vrednosti parametara μ i σ^2 , što je prosta varijanta nulte hipoteze, a u suprotnom

$$H_0: \text{„Raspodela obeležja } X \text{ je normalna.“}$$

što predstavlja složenu varijantu nulte hipoteze. Testiranje u oba slučaja biće vršeno za alternativnu hipotezu

$$H_1: \text{„Raspodela obeležja } X \text{ nije normalna.“}$$

koja je negacija nulte hipoteze u obe navedene varijante.

Kritična oblast može biti oblika KO1, KO2 ili KO3, u zavisnosti od toga da li istraživač ima informaciju ili potrebu za testiranjem protiv raspodela čije su frekvencije modaliteta veće oko proseka (KO1), veće na repovima raspodele (KO2), ili su frekvencije pak veće nego što bi trebalo i oko proseka i na repovima (KO3).

KO1 je oblika $W = [1, c]$ i određena je uslovom $P(A \leq c|H_0) = \alpha$, a p-vrednost testa se tada procenjuje sa $P(A \leq a|H_0)$.

KO2 je oblika $W = [c, 2,81]$ i određena je uslovom $P(A \geq c|H_0) = \alpha$, a p-vrednost testa se tada procenjuje sa $P(A \geq a|H_0)$.

KO3 je oblika $W = [1, c_1] \cup [c_2, 2,81]$, gde su c_1 i c_2 određeni uslovom

$$P(A \leq c_1|H_0) = P(A \geq c_2|H_0) = \frac{\alpha}{2}.$$

P-vrednost testa se tada procenjuje sa $p = 2 \min\{P(A \leq a|H_0), P(A \geq a|H_0)\}$.

„Idealna saglasnost“ (slučaj kad se dobije $zone(X_j) = 1, j \in \{1, \dots, n\}$ dovoljan broj puta) je slučaj u kojem treba odbaciti H_0 . Naime, raspodela za koju je više od 68,26% modaliteta na rastojanju ne većem od jedne standardne devijacije od proseka ne može biti normalna. U tom slučaju disperzija je suviše mala da bi se raspodela mogla opisati gustinom normalne raspodele. Takođe, takve raspodele su okarakterisane pozitivnim vrednostima koeficijenta spljoštenosti (Avdović & Jevremović, 2022).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Na primer, uzorak dat sa

$$\mu - \left(1 - \frac{j}{n}\right) \sigma; n \in \mathbb{N}, j = 0, \dots, n;$$

ima realizovanu vrednost test-statistike $a = 1$, ali je ovaj uzorak uniformno raspodeljen, tj. raspodela obeležja iz kojeg je uzorak izvučen ne odgovara normalnoj.

Dakle, i ekstremno male i ekstremno velike vrednosti statistike A ukazuju da je raspodela obeležja uzorka alternativna normalnoj.

Primer 5.1. Dato je 48 pakovanja od po 1000ml mleka i merena su procentualna učeća određenog proteina u mleku. Dobijeni su podaci

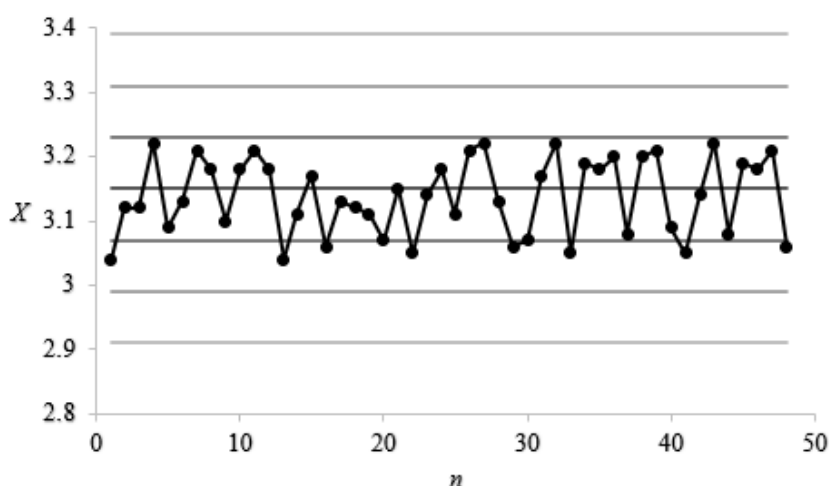
3,04, 3,12, 3,12, 3,22, 3,09, 3,13, 3,21, 3,18, 3,10, 3,18, 3,21, 3,18, 3,04, 3,11, 3,17, 3,06, 3,13, 3,12, 3,11, 3,07, 3,15, 3,05, 3,14, 3,18, 3,11, 3,21, 3,22, 3,13, 3,06, 3,07, 3,17, 3,22, 3,05, 3,19, 3,18, 3,20, 3,08, 3,20, 3,21, 3,09, 3,05, 3,14, 3,22, 3,08, 3,19, 3,18, 3,21, 3,06 (tabela 4.1).

Cilj je testirati da li raspodela učešća proteina statistički značajno odstupa od raspodele

- a) *Prvi slučaj*: $\mathcal{N}(3,15, 0,08^2)$,
- b) *Drugi slučaj*: koja je normalna bez obzira na vrednost parametara,

i to sa pragom značajnosti $\alpha = 0,05$.

Prvi slučaj. Prikazivanjem elemenata uzorka na grafikonu po uzoru na \bar{X} kontrolnu kartu dobija se bolji i jasniji prikaz izračunavanja vrednosti funkcije zone za svaki element uzorka, a time i realizovane vrednosti a test-statistike A . Dobija se sledeća slika.



Slika 5.8. Raspodela uzorka u odnosu na zone određene zadatim očekivanjem 3,15 i disperzijom $0,08^2$.

Izračunavanjem, dobija se $a = 1,16$.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Granice kritične oblasti dobijaju se iz uslova

$$P(A \leq c_1|H_0) = P(A \geq c_2|H_0) = 0,025$$

na osnovu tabele 5.2, i metodom polovljenja intervala. Dobija se kritična oblast $W = [1, 1,1974] \cup [1,4890, 2,81]$. Kako $a \in W$, nulta hipoteza se odbacuje, tj. raspodela učešća proteina statistički odstupa od raspodele $\mathcal{N}(3,15, 0,08^2)$.

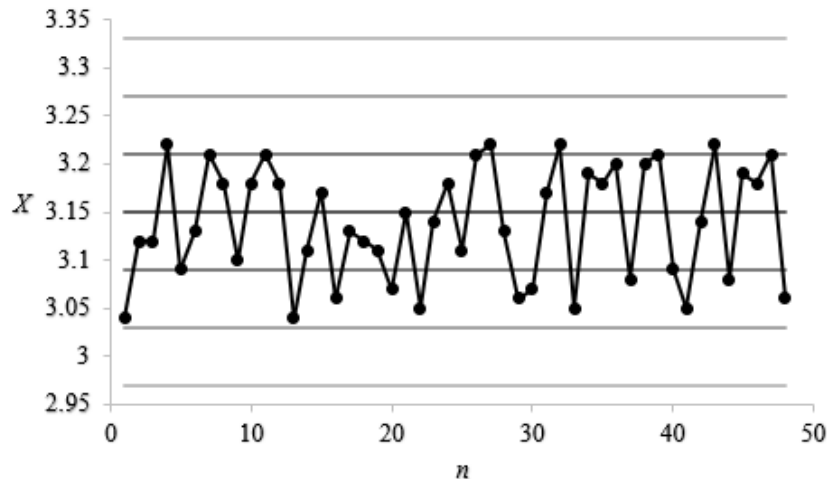
Približnim izračunavanjem p-vrednosti na osnovu tabele 5.2 i metode polovljenja intervala, dobija se

$$p = 2 \min\{P(A \leq 1,16|H_0), P(A \geq 1,16|H_0)\} = \min\{0,01; 0,99\} = 0,02 < 0,05,$$

pa postoji jak dokaz protiv H_0 , tj. da raspodela statistički značajno odstupa od $\mathcal{N}(3,15, 0,08^2)$.

Ovom slučaju bi odgovaralo i testiranje za KO1 jer je vrednost test statistike ekstremno mala. Odluka bi bila ista.

Drugi slučaj. Ovde su vrednosti parametara nepoznate pa se ocenjuju na osnovu uzorka. Izračunato je $\bar{x}_n = 3,14$ i $\tilde{s}_n^2 = 0,06^2$, pa se testira saglasnost raspodele učešća proteina sa raspedelom $\mathcal{N}(3,14, 0,06^2)$. Analogno prethodnom slučaju, dobija se sledeća slika.



Slika 5.9. Raspodela uzorka u odnosu na zone određene ocenjenim očekivanjem 3,14 i disperzijom $0,06^2$.

Izračunavanjem, dobija se $a = 1,42$.

Granice kritične oblasti dobijaju se iz uslova

$$P(A \leq c_1|H_0) = P(A \geq c_2|H_0) = 0,025$$

na osnovu tabele 5.3 (koristi se raspodela pri ocenjivanju parametara), i metodom polovljenja intervala. Dobija se kritična oblast $W = [1, 1,2588] \cup [1,4105, 2,81]$. Kako $a \in W$, nulta

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

hipoteza se odbacuje, tj. raspodela učešća proteina statistički odstupa od raspodele $\mathcal{N}(3,14, 0,06^2)$.

Približnim izračunavanjem p-vrednosti na osnovu tabele 5.3 i metode polovljenja intervala, dobija se

$$p = \min\{P(A \leq 1,42|H_0), P(A \geq 1,42|H_0)\} = \min\{0; 1\} = 0 < 0,05,$$

pa postoji veoma jak dokaz protiv H_0 , tj. da raspodela statistički značajno odstupa od $\mathcal{N}(3,14, 0,06^2)$.

U prvom slučaju može se videti da je odstupanje od prosečnih vrednosti premalo pa nije zadovoljeno pravilo 3σ za zadato očekivanje i disperziju. Stoga se realizovana vrednost test-statistike nalazi u levom repu kritične oblasti. U drugom slučaju, sa slike 5.9, može se videti da je ocenjivanje parametara optimalno pozicioniralo očekivanje i disperziju u odnosu na elemente uzorka, međutim struktura uzorka i dalje ne odgovara modelu normalne raspodele. ■

Test kvantil-zona

Test-statistika testa je zapravo statistika V koja je korišćena kao procena stanja procesa pod kontrolom kod kontrolne karte kvantil-zona. Statistika je bazirana na pravilu 3σ , tj. funkciji kvantil-zone date definicijom 4.12. U definiciji 4.12 korišćeni su brojevi 1, 2, 3 i 3,1947 pri definiciji funkcije kvantil-zone. Kako raspodela statistike V zavisi direktno od raspodele koja se dobija primenom funkcije zone na elemente uzorka, u nastavku će biti korišćeni brojevi 1, 1,95, 2,58 i 2,81, tj. prilagođena varijanta pravila 3σ . Razlog tome je činjenica da sa prvim skupom vrednosti dva različita uzorka mogu dati iste vrednosti statistike V . Na primer, važi $1+1+3=1+2+2$. Stoga je bolje uvesti novu definiciju funkcije kvantil-zone. Drugačija definicija je korišćena u prethodnom poglavlju radi ilustracije više mogućnosti. Dodatno, navedeni nedostatak biva nadomešten pravilima tumačenja kontrolnih karti.

Neka je $X_1, \dots, X_n, n \in \mathbb{N}$ prost slučajan uzorak obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$.

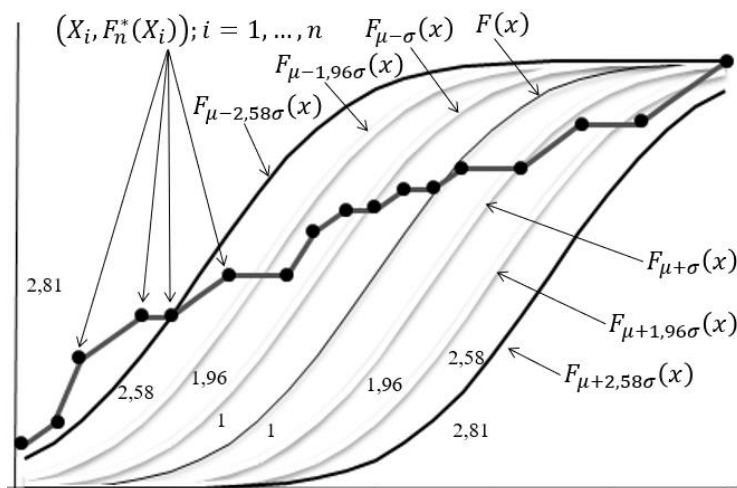
Definicija 5.6. Funkcija kvantil-zone normalne raspodele $\mathcal{N}(\mu, \sigma^2)$ za uzorak X_1, \dots, X_n data je sa

$$q_{zone}(x) = \begin{cases} 1, & F_{\mu+\sigma}(x) \leq F_n^*(x) < F(x) \vee F(x) < F_n^*(x) \leq F_{\mu-\sigma}(x) \\ 1,96, & F_{\mu+1,96\sigma}(x) \leq F_n^*(x) < F_{\mu+\sigma}(x) \vee F_{\mu-\sigma}(x) < F_n^*(x) \leq F_{\mu-1,96\sigma}(x) \\ 2,58, & F_{\mu+2,58\sigma}(x) \leq F_n^*(x) < F_{\mu+1,96\sigma}(x) \vee F_{\mu-1,96\sigma}(x) < F_n^*(x) \leq F_{\mu-2,58\sigma}(x) \\ 2,81, & F_n^*(x) < F_{\mu+2,58\sigma}(x) \vee F_n^*(x) > F_{\mu-2,58\sigma}(x) \end{cases} \quad x \in \mathbb{R},$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

gde su $F_{\mu \pm h\sigma}(x)$, $x \in \mathbb{R}$, $h \in \{1, 1,96, 2,58, 2,81\}$ funkcije odgovarajućih raspodela $\mathcal{N}(\mu \pm h\sigma, \sigma^2)$, $F(x)$ funkcija raspodele $X \sim \mathcal{N}(\mu, \sigma^2)$ i $F_n^*(x)$ empirijska funkcija raspodele polaznog uzorka.

Na sledećoj slici je grafički prikazan način definisanja funkcije kvantil-zone.



Slika 5.10. Grafička ilustracija definicije funkcije kvantil-zone sa prilagođenom varijantom pravila 3σ .

Razlog izbora vrednosti 2,81 je isti kao onaj dat napomenom 5.1, tj. isti zaključak važi kod raspodele kvantil-zona.

Analogno testu zona, ideja za testiranje normalnosti se sastoji u tome da se funkcija kvantil-zone primeni na polazni uzorak i tako se dobije uzorak mera odstupanja od očekivane vrednosti μ . Tako dobijen uzorak u ovom slučaju je

$$q_{zone}(X_1), \dots, q_{zone}(X_n). \quad (5.2)$$

Sa druge strane ista analogija pravila 3σ nije primenljiva na raspodelu elemenata uzorka što je precizirano sledećim stavom.

Stav 5.5. Neka su Y_1, \dots, Y_n statistike poretka polaznog uzorka X_1, \dots, X_n , $n \in \mathbb{N}$ obeležja $X \sim \mathcal{N}(\mu, \sigma^2)$, i neka je $Y_j = Y_n$, $j \in \{l+1, \dots, n\}$ za $l \in \{1, \dots, n-1\}$. Raspodela obeležja $q_{zone}(X_i)$, $i \in \{1, \dots, n\}$ data je sledećim zakonom raspodele za odgovarajuće Y_i

$$q_{zone}(Y_i): \begin{pmatrix} 1 & 1,96 & 2,58 & 2,81 \\ p_1(n) & p_2(n) & p_3(n) & p_4(n) \end{pmatrix}, \quad i \in \{1, \dots, l\},$$

$$P(q_{zone}(Y_i) = 2,81) = 1, \quad i \in \{l+1, \dots, n\},$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

gde $p_1(n) > p_2(n) > p_3(n) > p_4(n)$ i $\lim_{n \rightarrow +\infty} p_1(n) = 1$ i $\lim_{n \rightarrow +\infty} p_j(n) = 0, j = 2, 3, 4$.

Dokaz. Za $i \in \{1, \dots, l\}$ stav sledi na osnovu teoreme 4.1 Glivenka i Kantelija. Naime, uvek važi $F_{\mu+\sigma}(x) < F(x) < F_{\mu-\sigma}(x)$ (promenom parametra lokacije menja se i lokacija funkcije raspodele), a kako na osnovu teoreme 4.1 važi $F_n^*(x) \xrightarrow{P} F(x) (n \rightarrow +\infty), x \in \mathbb{R}$, to nakon nekog $n \in \mathbb{N}$ važi $F_{\mu+\sigma}(x) < F_n^*(x) < F_{\mu-\sigma}(x)$. Za manje vrednosti n u retkim slučajevima se stoga mogu realizovati i vrednosti 1,96, 2,58 i 2,81 funkcije q_{zone} .⁴⁷

Sa druge strane, ako je $i \in \{l + 1, \dots, n\}$ onda, kako je $Y_i = Y_n = \max\{X_1, \dots, X_n\}$ to je

$$F_n^*(Y_i) = \frac{1}{n} \sum_{j=1}^n I(Y_j \leq Y_i) = \frac{n}{n} = 1.$$

Tada je $F_n^*(Y_i) > F_{\mu-3\sigma}(x)$ za bilo koje $x \in \mathbb{R}$, tj. $q_{zone}(Y_i) = 2,81$ kad god je $i \in \{l + 1, \dots, n\}$, a to je ekvivalentno sa

$$P(q_{zone}(Y_i) = 2,81) = 1, \quad i \in \{l + 1, \dots, n\}. \blacksquare$$

Definicija 5.7. Statistika

$$V = \frac{1}{n} \sum_{i=1}^n q_{zone}(X_i)$$

ima raspodelu kvantil-zona sa n stepeni slobode, u oznaci $V \sim q_{zone}_n$.

Posledica 5.5. Raspodela statistike V ne zavisi od parametara raspodele obeležja X kada su parametri raspodele obeležja X poznati.

Dokaz. Na osnovu stava 5.5, raspodela obeležja $q_{zone}(X_i), i = 1, \dots, n$, a samim tim ni raspodela statistike V , ne zavisi od izbora parametara obeležja X osim možda u smislu veličine n za koje će se raspodele poklapati za različite parametre. Međutim, kako je konvergencija $p_1(n) \rightarrow 1 (n \rightarrow +\infty)$ posledica konvergencije statistike D_n ka raspodeli Kolmogorova, koja ne zavisi od izbora parametara obeležja X , to ni ta konvergencija, a samim tim ni $p_j(n) \rightarrow 0 (n \rightarrow +\infty), j = 1, 2, 3$ (jer zbirno daju 1), nisu određene izborom parametara obeležja X . ■

S obzirom na to da se uzorak (5.2) dobija korišćenjem empirijske funkcije raspodele, to su njegovi elementi uzajamno zavisni. Stoga se ne može eksplicitno odrediti ni zakon ni funkcija raspodele statistike V , pa se koristi 100000 Monte Karlo simulacija raspodele statistike V i

⁴⁷ Simulacijama se može proveriti da je već za $n = 5$, izuzev za maksimalnu vrednost, realizacija vrednosti različite od 1 reda od 3% slučajeva.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

dobijaju se tabele raspodele. I kod ove statistike pravi se razlika između slučajeva sa poznatim i nepoznatim i ocenjenim parametrima, mada će se pokazati da se za velike obime uzorka raspodele poklapaju.

Tabela 5.4. Raspodela kvantil-zona – Poznati parametri. $F(x) = P(V \leq x); V \sim q_{zone_n}$.

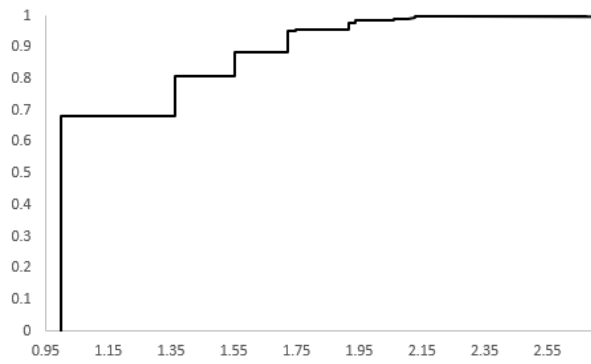
<i>n</i>	<i>p</i>						
	0,01	0,025	0,05	0,1	0,15	0,2	0,5
5	1,3620	1,3620	1,3620	1,3620	1,3620	1,3620	1,3620
10	1,1810	1,1810	1,1810	1,1810	1,1810	1,1810	1,1810
20	1,0905	1,0905	1,0905	1,0905	1,0905	1,0905	1,0905
30	1,0603	1,0603	1,0603	1,0603	1,0603	1,0603	1,0603
50	1,0362	1,0362	1,0362	1,0362	1,0362	1,0362	1,0362
100	1,0181	1,0181	1,0181	1,0181	1,0181	1,0181	1,0181
200	1,0091	1,0091	1,0091	1,0091	1,0091	1,0091	1,0091
300	1,0060	1,0060	1,0060	1,0060	1,0060	1,0060	1,0060
500	1,0036	1,0036	1,0036	1,0036	1,0036	1,0036	1,0036
1000	1,0018	1,0018	1,0018	1,0018	1,0018	1,0018	1,0018
1500	1,0012	1,0012	1,0012	1,0012	1,0012	1,0012	1,0012
2000	1,0009	1,0009	1,0009	1,0009	1,0009	1,0009	1,0009

Tabela 5.4. Nastavak.

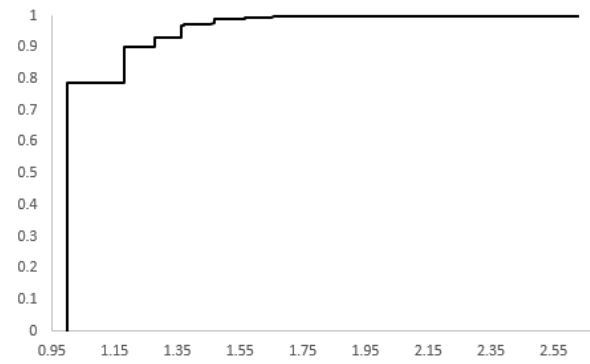
<i>n</i>	<i>p</i>					
	0,8	0,85	0,9	0,95	0,975	0,99
5	1,5540	1,5540	1,7460	1,7460	1,9380	2,1300
10	1,2770	1,2770	1,2770	1,3730	1,4690	1,5650
20	1,0905	1,0905	1,1385	1,1385	1,1695	1,1865
30	1,0603	1,0603	1,0603	1,0923	1,0923	1,1243
50	1,0362	1,0362	1,0362	1,0554	1,0554	1,0554
100	1,0181	1,0181	1,0181	1,0181	1,0277	1,0277
200	1,0091	1,0091	1,0091	1,0091	1,0139	1,0139
300	1,0060	1,0060	1,0060	1,0060	1,0092	1,0092
500	1,0036	1,0036	1,0036	1,0036	1,0055	1,0055
1000	1,0018	1,0018	1,0018	1,0018	1,0018	1,0028
1500	1,0012	1,0012	1,0012	1,0012	1,0012	1,0018
2000	1,0009	1,0009	1,0009	1,0009	1,0009	1,0014

Sledećom slikom dat je grafik funkcije raspodele kvantil-zona za razne obime uzorka, a na osnovu Monte Karlo simulacija. Slika se ne menja promenom parametara polazne normalne raspodele.

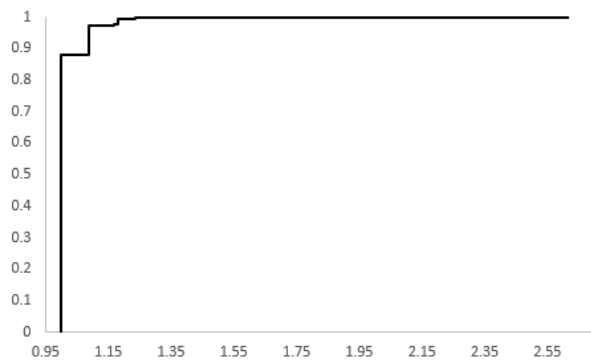
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



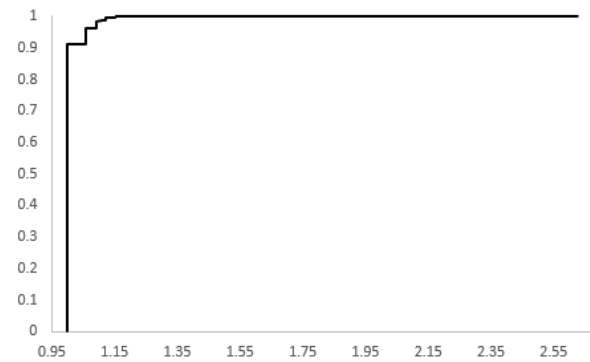
$n = 5$



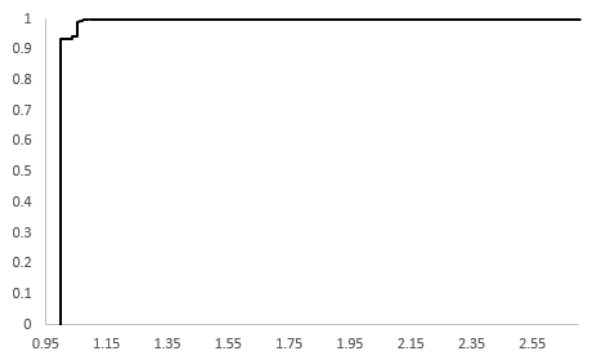
$n = 10$



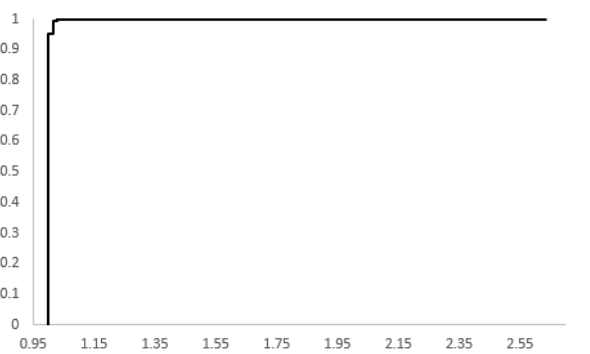
$n = 20$



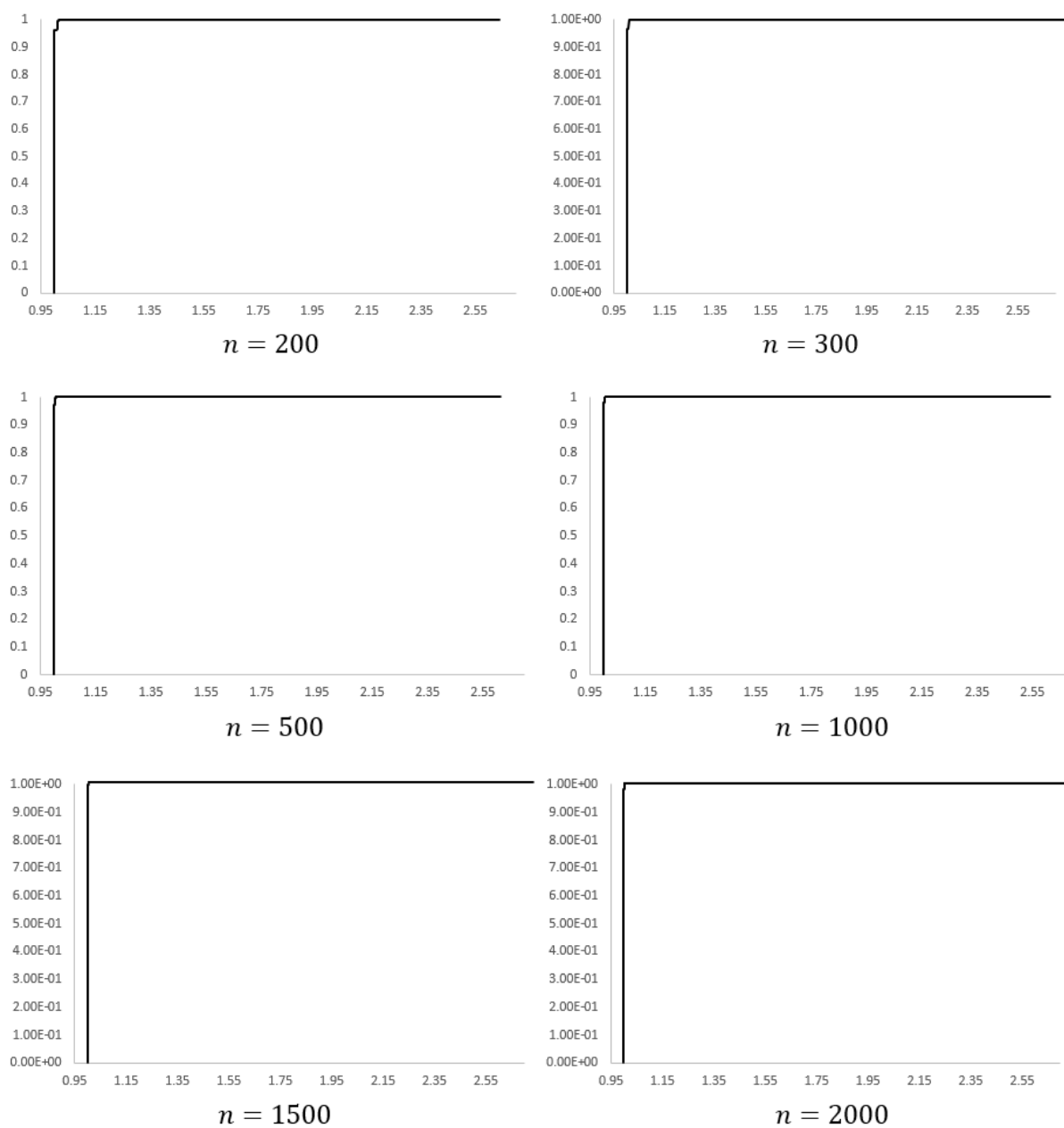
$n = 30$



$n = 50$



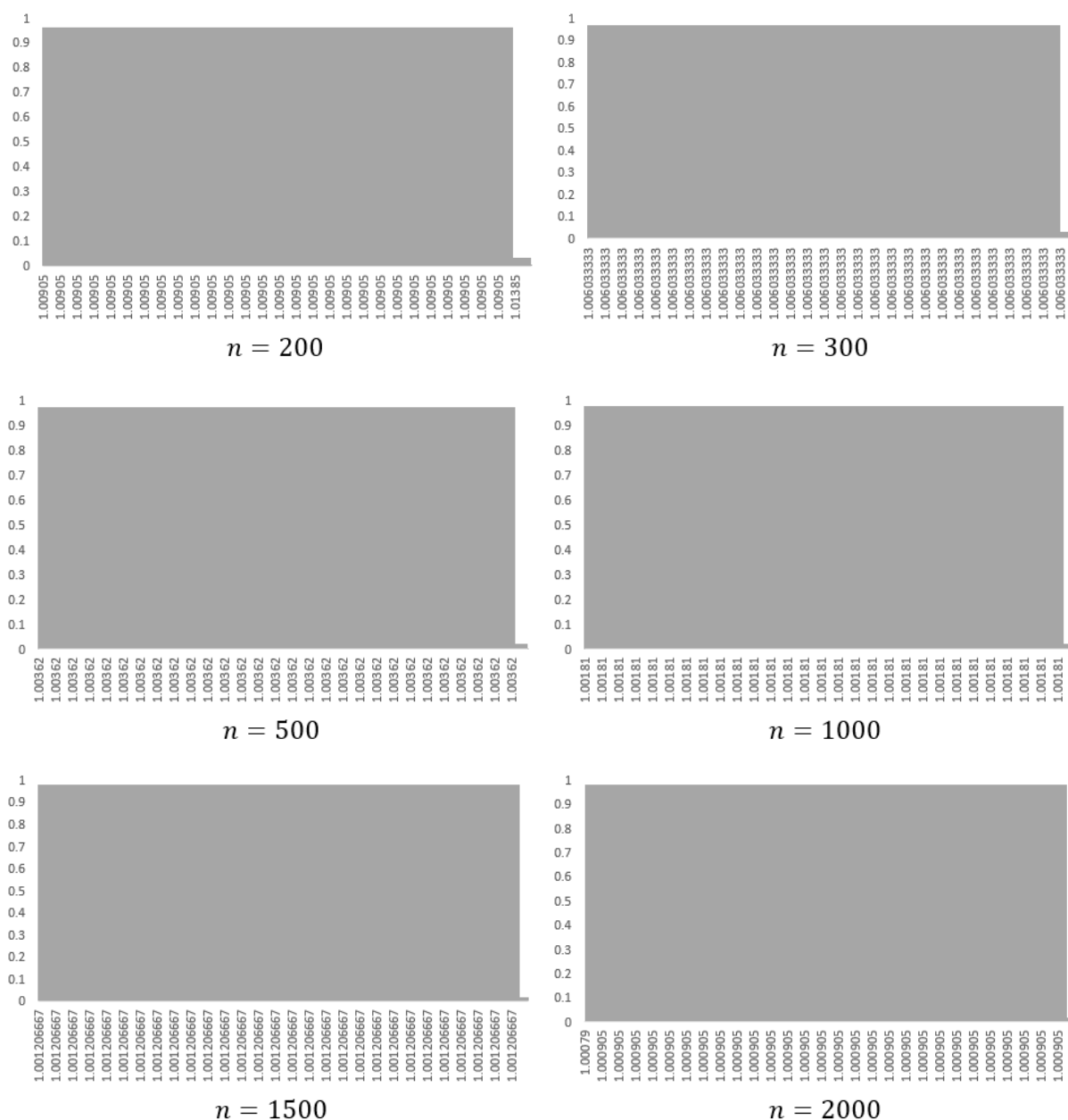
$n = 100$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Slika 5.11. Grafik funkcije raspodele q_{zone_n} za poznate parametre normalne raspodele.

Sledećom slikom dat je grafički prikaz zakona raspodele kvantil-zona, takođe za razne obime uzorka.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.12. Zakon raspodele q_{zone_n} dat histogramom za poznate parametre normalne raspodele.

U nastavku su detaljnije navedene neke osobine raspodele kvantil-zona.

Stav 5.6. Neka je $E(q_{zone}(X)) = E_1$ i $E((q_{zone}(X))^2) = E_2$, $X \sim \mathcal{N}(\mu, \sigma^2)$. Očekivanje i disperzija statistike V sa raspedelom q_{zone_n} su redom

$$E(V) = \frac{(n - n_{max})E_1 + 2,81 \cdot n_{max}}{n},$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$D(V) = \frac{(n - n_{\max})(E_2 + E_1^2 + 5,62 \cdot n_{\max}E_1)}{n^2}.$$

Dokaz. Koristeći osnovne osobine očekivanja dobija se

$$\begin{aligned} E(V) &= E\left(\frac{1}{n} \sum_{i=1}^n q_{zone}(X_i)\right) \\ &= E\left(\frac{1}{n} \sum_{i=1}^n q_{zone}(Y_i)\right) \\ &= \frac{1}{n} \left(\sum_{i=1}^l E(q_{zone}(Y_i)) + \sum_{i=l+1}^n E(q_{zone}(Y_i)) \right) \\ &= \frac{lE_1 + (n - l) \cdot 2,81}{n} \\ &= \frac{(n - n_{\max})E_1 + 2,81 \cdot n_{\max}}{n} \end{aligned}$$

S obzirom na to da se uzorak (5.2) dobija korišćenjem empirijske funkcije raspodele, to su njegovi elementi uzajamno zavisni. Međutim, na osnovu stava 5.5, može se uočiti da važe sledeće jednakosti

$$\begin{aligned} E(q_{zone}(Y_i) \cdot q_{zone}(Y_j)) &= E(q_{zone}(Y_i)) \cdot E(q_{zone}(Y_j)) = E_1^2, \quad i, j \in \{1, \dots, l\}, \\ E(q_{zone}(Y_i) \cdot q_{zone}(Y_j)) &= E((q_{zone}(Y_i))^2) = E(2,81^2) = 7,8961, \\ & \hspace{20em} i, j \in \{l + 1, \dots, n\}, \\ E(q_{zone}(Y_i) \cdot q_{zone}(Y_j)) &= E(q_{zone}(Y_i) \cdot 2,81) \\ &= 2,81 \cdot E(q_{zone}(Y_i)) \\ &= 2,81 \cdot E_1, \quad i \in \{1, \dots, l\}, j \in \{l + 1, \dots, n\}. \end{aligned}$$

Na osnovu navedenog dobija se

$$\begin{aligned} E(V^2) &= E\left(\left(\frac{1}{n} \sum_{i=1}^n q_{zone}(X_i)\right)^2\right) \\ &= E\left(\frac{1}{n^2} \left(\sum_{i=1}^n (q_{zone}(X_i))^2 + 2 \sum_{i \neq j} q_{zone}(X_i) \cdot q_{zone}(X_j) \right)\right) \end{aligned}$$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$\begin{aligned}
 &= \frac{1}{n^2} \left(\sum_{i=1}^l E \left((q_{zone}(Y_i))^2 \right) + \sum_{i=l+1}^n E \left((q_{zone}(Y_i))^2 \right) \right. \\
 &\quad + 2 \left(\sum_{i=1}^l \sum_{\substack{j=1 \\ j \neq i}}^l E \left(q_{zone}(Y_i) \cdot q_{zone}(Y_j) \right) + 2 \sum_{i=1}^l \sum_{j=l+1}^n E \left(q_{zone}(Y_i) \cdot q_{zone}(Y_j) \right) \right. \\
 &\quad \left. \left. + \sum_{i=l+1}^n \sum_{\substack{j=l+1 \\ j \neq i}}^n E \left(q_{zone}(Y_i) \cdot q_{zone}(Y_j) \right) \right) \right) \\
 &= \frac{1}{n^2} \left(l \cdot E_2 + (n-l) \cdot 2,81^2 \right. \\
 &\quad \left. + 2 \left(\binom{l}{2} E_1^2 + 2 \cdot l(n-l) \cdot E_1 \cdot 2,81 + \binom{n-l}{2} 2,81^2 \right) \right) \\
 &= \frac{lE_2 + l(l-1)E_1^2 + 4 \cdot 2,81 \cdot l(n-l)E_1 + 7,8961 \cdot (n-l)^2}{n^2} \\
 &= \frac{1}{n^2} \left((n-n_{\max})E_2 + (n-n_{\max})(n-n_{\max}-1)E_1^2 + 11,24 \cdot n_{\max}(n-n_{\max})E_1 \right. \\
 &\quad \left. + 7,8961 \cdot n_{\max}^2 \right).
 \end{aligned}$$

Sledi

$$\begin{aligned}
 D(V) &= E(V^2) - (E(V))^2 \\
 &= \frac{1}{n^2} \left((n-n_{\max})E_2 + (n-n_{\max})(n-n_{\max}-1)E_1^2 + 11,24 \cdot n_{\max}(n-n_{\max})E_1 \right. \\
 &\quad \left. + 7,8961 \cdot n_{\max}^2 \right) - \left(\frac{(n-n_{\max})E_1 + 2,81 \cdot n_{\max}}{n} \right)^2 \\
 &= \frac{1}{n^2} \left((n-n_{\max})E_2 + (n-n_{\max})(n-n_{\max}-1)E_1^2 + 11,24 \cdot n_{\max}(n-n_{\max})E_1 \right. \\
 &\quad \left. + 7,8961 \cdot n_{\max}^2 - (n-n_{\max})^2 E_1^2 - 2 \cdot E_1 \cdot 2,81 \cdot n_{\max}(n-n_{\max}) - 2,81^2 n_{\max}^2 \right) \\
 &= \frac{1}{n^2} \left((n-n_{\max})E_2 + (n-n_{\max})E_1^2 + 5,62 \cdot n_{\max}(n-n_{\max})E_1 \right) \\
 &= \frac{(n-n_{\max})(E_2 + E_1^2 + 5,62 \cdot n_{\max}E_1)}{n^2}. \blacksquare
 \end{aligned}$$

Napomena 5.3. Kako je normalna raspodela neprekidna, u uzorku normalno raspedljenog obeležja se svaki modalitet uglavnom javlja ne više od jednom, s' tim da

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

se u određenim intervalima nalazi više modaliteta nego u drugima. Samim tim, može se pretpostaviti i da se maksimum javlja ne više od jednom, tj. $n_{\max} = 1$. U tom slučaju važi

$$E(V) = \frac{(n-1)E_1 + 2,81}{n} \quad \text{i} \quad D(V) = \frac{(n-1)(E_2 + E_1^2 + 5,62 \cdot E_1)}{n^2}.$$

Određivanje koeficijenta asimetrije i koeficijenta spoljoštenosti raspodele kvantil-zona zahteva veoma komplikovana izračunavanja koja nisu od praktičnog značaja za postupak testiranja, pa ovde nisu navedeni.

Kod Avdovića i Jevremović (2022) dato je izračunavanje očekivanja i disperzije za slučaj u kojem je empirijska funkcija raspodele raspodeljena po izvesnoj modifikaciji pravila 3σ u kojem je

$$q_{zone}(Y_i): \begin{pmatrix} 1 & 1,96 & 2,58 & 2,81 \\ 0,68 & 0,27 & 0,04 & 0,01 \end{pmatrix}, \quad i \in \{1, \dots, l\}.$$

U opštem slučaju ipak nije moguće znati tačne vrednosti $p_j(n)$, $j \in \{1, 2, 3, 4\}$ zbog zavisnosti elemenata uzorka (5.2).

Iz stava 5.5 se može zaključiti i $\lim_{n \rightarrow +\infty} E_j = 1$, $j \in \{1, 2\}$, tj. $\lim_{n \rightarrow +\infty} E(V) = 1$ i $\lim_{n \rightarrow +\infty} D(V) = 0$.

Kada su parametri raspodele nepoznati, mogu se ocenjivati metodom maksimalne verodostojnosti. Kako je $X \sim N(\mu, \sigma^2)$, ocene za μ i σ^2 su redom $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ i $\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Ipak, za ocenu σ^2 uzima se $\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ jer je to nepristrasna ocena za σ^2 , za razliku od \bar{S}_n^2 .

Teorema 5.3. *Raspodela statistike V se ne menja promenom parametara raspodele obeležja X kada parametri raspodele obeležja X nisu poznati i ocenjuju se.*

Dokaz. Neka je raspodela obeležje X normalna $\mathcal{N}(\mu_1, \sigma_1^2)$ gde su parametri μ_1 i σ_1^2 nepoznati i X_1, \dots, X_n prost slučajan uzorak obeležja X . Tada $F_n^*(x) \sim \mathcal{B}(n, F(x))$, $x \in \mathbb{R}$ gde je $F(x)$ funkcija raspodele obeležja X (Stav 4.1). Neka su $\hat{\mu} = \bar{X}_n$ i $\hat{\sigma}^2 = \tilde{S}_n^2$ i

$$\begin{aligned} I_1(x) &= [F_{\hat{\mu}+\hat{\sigma}}(x), F_{\hat{\mu}-\hat{\sigma}}(x)], \\ I_2(x) &= [F_{\hat{\mu}+1,96\hat{\sigma}}(x), F_{\hat{\mu}+\hat{\sigma}}(x)] \cup (F_{\hat{\mu}-\hat{\sigma}}(x), F_{\hat{\mu}-1,96\hat{\sigma}}(x)], \\ I_3(x) &= [F_{\hat{\mu}+2,58\hat{\sigma}}(x), F_{\hat{\mu}+1,96\hat{\sigma}}(x)] \cup (F_{\hat{\mu}-1,96\hat{\sigma}}(x), F_{\hat{\mu}-2,58\hat{\sigma}}(x)], \\ I_4(x) &= [0, F_{\hat{\mu}+2,58\hat{\sigma}}(x)] \cup (F_{\hat{\mu}-2,58\hat{\sigma}}(x), 1], \end{aligned}$$

gde su $F_{\hat{\mu} \pm h\hat{\sigma}}$, $h \in \{1, 1,96, 2,58, 2,81\}$ funkcije raspodele obeležja sa normalnim raspodelama sa parametrima $\hat{\mu} \pm h\hat{\sigma}$ i $\hat{\sigma}$. Tada se za

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

$$q_{zone_{\hat{\mu},\hat{\sigma}}}(x) = \begin{cases} 1, & x \in I_1(x) \\ 1,96, & x \in I_2(x) \\ 2,58, & x \in I_3(x) \\ 2,81, & x \in I_4(x) \end{cases}, \quad x \in \mathbb{R}.$$

dobija uzorak $q_{zone_{\hat{\mu},\hat{\sigma}}}(X_j), j = 1, \dots, n$. Neka su dalje $Z_j = \frac{X_j - \hat{\mu}}{\hat{\sigma}}, j = 1, \dots, n$, $a_1 = 1, a_2 = 1,96, a_3 = 2,58, a_4 = 2,81$ i

$$\begin{aligned} I'_1(x) &= [F_1(x), F_{-1}(x)], \\ I'_2(x) &= [F_{1,96}(x), F_1(x)] \cup (F_{-1}(x), F_{-1,96}(x)], \\ I'_3(x) &= [F_{2,58}(x), F_{1,96}(x)] \cup (F_{-1,96}(x), F_{-2,58}(x)], \\ I'_4(x) &= [0, F_{2,58}(x)] \cup (F_{-2,58}(x), 1]. \end{aligned}$$

Tada je

$$\begin{aligned} F_{\hat{\mu} \pm h \hat{\sigma}}(X_j) &= \frac{1}{\sqrt{2\pi}\hat{\sigma}} \int_{-\infty}^{X_j} e^{-\frac{(t - \hat{\mu} \mp h \hat{\sigma})^2}{2\hat{\sigma}^2}} dt = \left| \begin{array}{l} t = \hat{\sigma}z + \hat{\mu} \quad t \rightarrow -\infty \quad t = X_j \\ dt = \hat{\sigma}dz \quad z \rightarrow -\infty \quad z = \frac{X_j - \hat{\mu}}{\hat{\sigma}} \end{array} \right| \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_j} e^{-\frac{(z \mp h)^2}{2}} dz = F_{\pm h}(Z_j), j = 1, \dots, n. \end{aligned}$$

Odatle sledi $I_k(X_j) = I'_k(Z_j), k = 1, 2, 3, 4$. Konačno, za $j = 1, \dots, n$ i $k = 1, 2, 3, 4$

$$P(q_{zone_{\hat{\mu},\hat{\sigma}}}(X_j) = a_k) = P(F_n^*(X_j) \in I_k(X_j)) = P(F_n^*(X_j) \in I'_k(Z_j)).$$

Neka je sada raspodela obeležje X normalna $\mathcal{N}(\mu_2, \sigma_2^2)$, gde su parametri μ_2 i σ_2^2 nepoznati ali različiti od μ_1 i σ_1^2 ($(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$). Analogno je $F_n^*(x) \sim \mathcal{B}(n, F(x))$, i za $Y_j = \frac{X_j - \hat{\mu}}{\hat{\sigma}}, j = 1, \dots, n$ važi

$$P(q_{zone_{\hat{\mu},\hat{\sigma}}}(X_j) = a_k) = P(F_n^*(X_j) \in I'_k(Y_j)),$$

a s obzirom da Z_j i Y_j imaju iste raspodele u odnosu na odgovarajuće X (dokaz teoreme 5.2), to je raspodela statistike $q_{zone_{\hat{\mu},\hat{\sigma}}}(X_j)$, a time i statistike V , ista u oba slučaja. ■

Sledi da se približna raspodela može odrediti samo Monte Karlo simulacijama. U nastavku je data tabela u kojoj su prikazane vrednosti statistike V za razne obime uzorka i kvantile (verovatnoće).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Tabela 5.5. Raspodela kvantil-zona – Ocenjeni parametri. $F(x) = P(V \leq x); V \sim q_{zone, n}$.

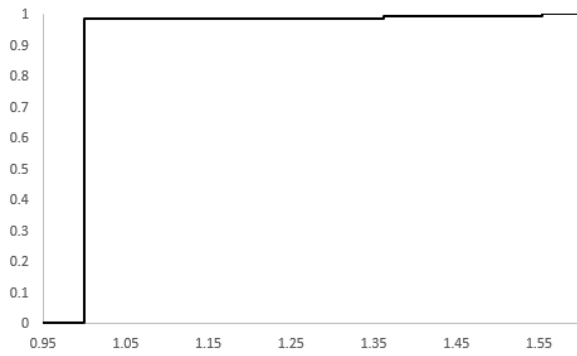
<i>n</i>	<i>p</i>						
	0,01	0,025	0,05	0,1	0,15	0,2	0,5
5	1,3620	1,3620	1,3620	1,3620	1,3620	1,3620	1,3620
10	1,1810	1,1810	1,1810	1,1810	1,1810	1,1810	1,1810
20	1,0905	1,0905	1,0905	1,0905	1,0905	1,0905	1,0905
30	1,0603	1,0603	1,0603	1,0603	1,0603	1,0603	1,0603
50	1,0362	1,0362	1,0362	1,0362	1,0362	1,0362	1,0362
100	1,0181	1,0181	1,0181	1,0181	1,0181	1,0181	1,0181
200	1,0091	1,0091	1,0091	1,0091	1,0091	1,0091	1,0091
300	1,0060	1,0060	1,0060	1,0060	1,0060	1,0060	1,0060
500	1,0036	1,0036	1,0036	1,0036	1,0036	1,0036	1,0036
1000	1,0018	1,0018	1,0018	1,0018	1,0018	1,0018	1,0018
1500	1,0012	1,0012	1,0012	1,0012	1,0012	1,0012	1,0012
2000	1,0009	1,0009	1,0009	1,0009	1,0009	1,0009	1,0009

Tabela 5.5. Nastavak.

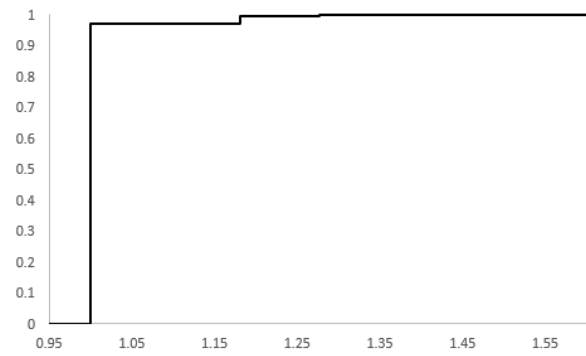
<i>n</i>	<i>p</i>					
	0,8	0,85	0,9	0,95	0,975	0,99
5	1,3620	1,3620	1,3620	1,3620	1,3620	1,5540
10	1,1810	1,1810	1,1810	1,1810	1,2770	1,2770
20	1,0905	1,0905	1,0905	1,0905	1,1385	1,1385
30	1,0603	1,0603	1,0603	1,0603	1,0923	1,0923
50	1,0362	1,0362	1,0362	1,0362	1,0554	1,0554
100	1,0181	1,0181	1,0181	1,0181	1,0277	1,0277
200	1,0091	1,0091	1,0091	1,0091	1,0139	1,0139
300	1,0060	1,0060	1,0060	1,0060	1,0092	1,0092
500	1,0036	1,0036	1,0036	1,0036	1,0055	1,0055
1000	1,0018	1,0018	1,0018	1,0018	1,0028	1,0028
1500	1,0012	1,0012	1,0012	1,0012	1,0018	1,0018
2000	1,0009	1,0009	1,0009	1,0009	1,0014	1,0014

Sledećom slikom dat je grafik funkcije raspodele zona za razne obime uzorka.

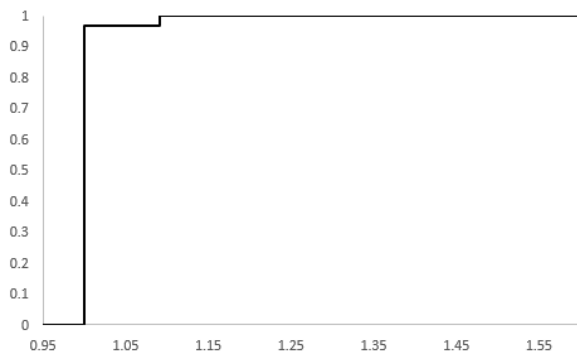
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



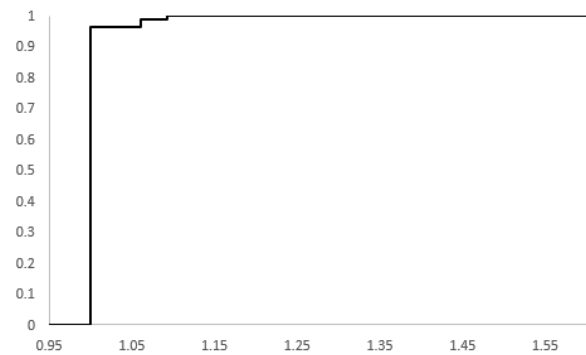
$n = 5$



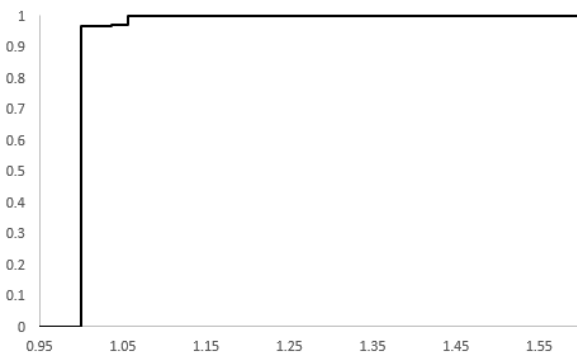
$n = 10$



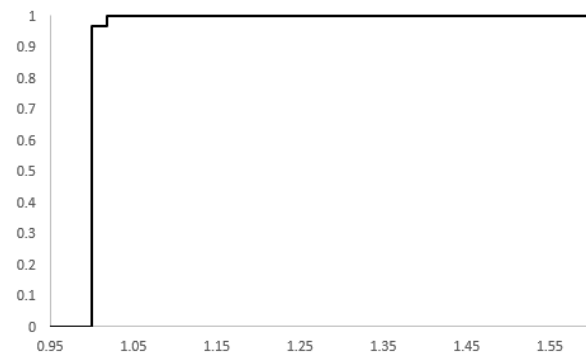
$n = 20$



$n = 30$

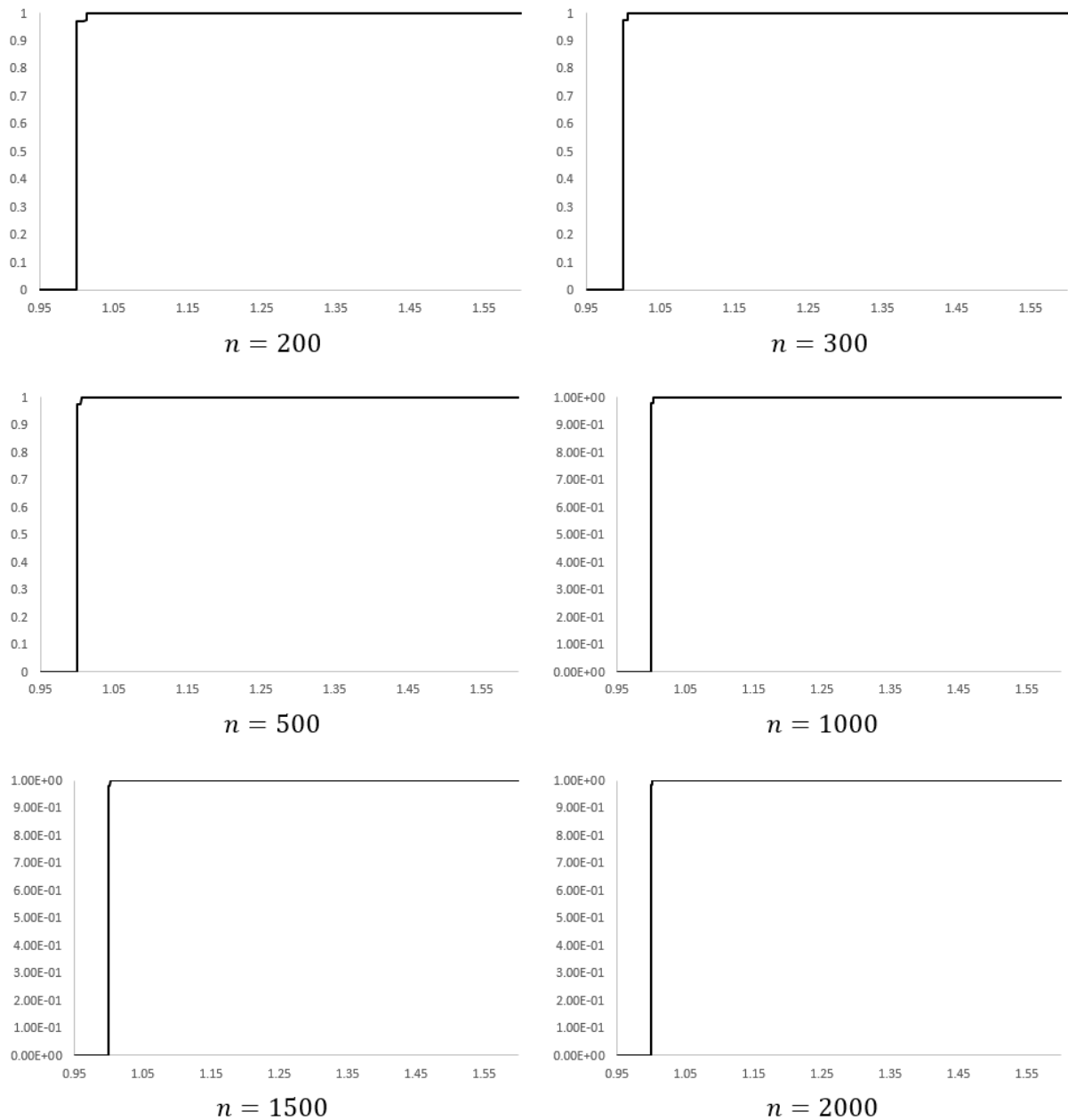


$n = 50$



$n = 100$

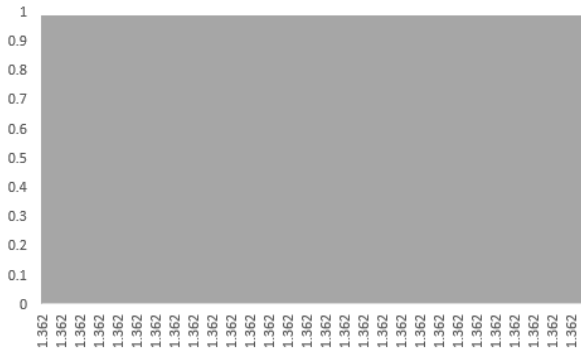
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



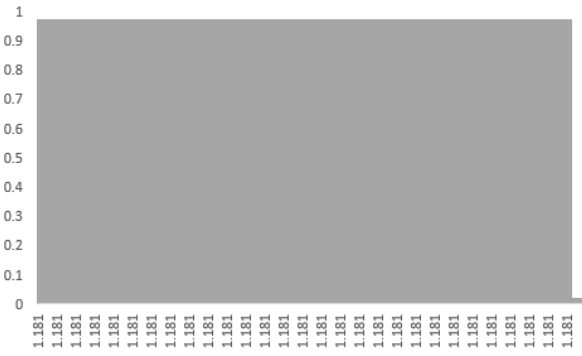
Slika 5.13. Grafik funkcije raspodele q_{zone_n} za ocenjene parametre normalne raspodele.

Sledećom slikom dat je grafički prikaz zakona raspodele zona, takođe za razne obime uzorka.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



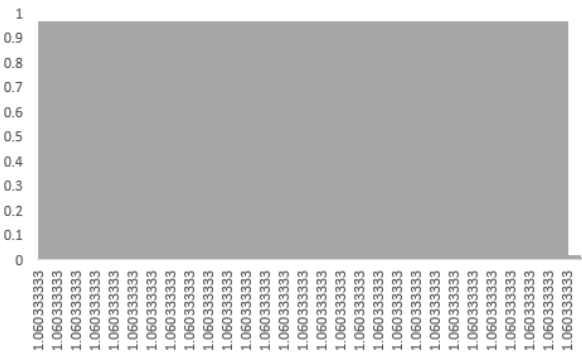
$n = 5$



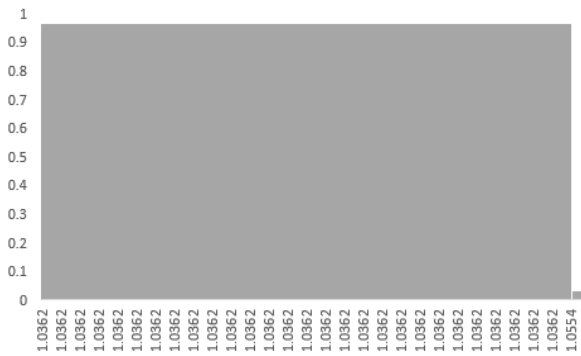
$n = 10$



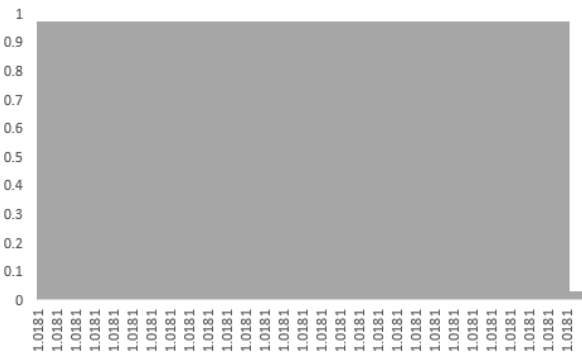
$n = 20$



$n = 30$



$n = 50$



$n = 100$

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Kako određivanje raspodele statistike V zavisi direktno od normalnosti raspodele polaznog uzorka X_1, X_2, \dots, X_n , tj. obeležja X , to je nulta hipoteza testa data sa

$$H_0: \text{„Raspodela obeležja } X \text{ je } \mathcal{N}(\mu, \sigma^2)\text{.“}$$

ako su od važnosti i vrednosti parametara μ i σ^2 , tj. prosta nulta hipoteza, a u suprotnom

$$H_0: \text{„Raspodela obeležja } X \text{ je normalna.“}$$

tj. složena nulta hipoteza. Zbog različitosti načina testiranja, ove varijante nulte hipoteze razmatraće se odvojeno.

Testiranje u oba slučaja biće sa alternativnom hipotezom

$$H_1: \text{„Raspodela obeležja } X \text{ nije normalna.“}$$

jer ona negira obe varijante nulte hipoteze.

Pravilo 3σ ne važi za empirijsku funkciju raspodele i ostale funkcije koje formiraju kvantil-zone. Naime, iako bi uzorak za koji je $F_n^*(X_i) \in [F_{\mu+\sigma}(X_i), F_{\mu-\sigma}(X_i)]$, $i = 1, \dots, n$, $X_i \neq \max\{X_1, \dots, X_n\}$ u nekim slučajevima odgovarao nekoj alternativnoj hipotezi njegova detekcija bi se morala vršiti suženjem kvantil-zona. Stoga se u slučaju ovog testa može koristiti jednostrana kritična oblast $W = [c, 2,81]$, određena uslovom

$$P(V \geq c | H_0) = \alpha.$$

P-vrednost se u tada aproksimira sa $P(V \geq v | H_0)$.

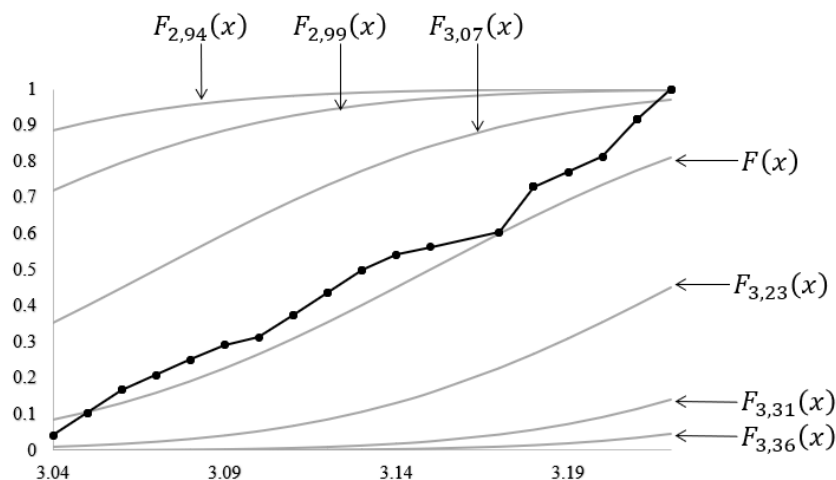
Primer 5.2. Testira se da li raspodela učešća proteina (podaci iz primera 5.1) statistički značajno odstupa od raspodele

- a) *Prvi slučaj*: $\mathcal{N}(3,15, 0,08^2)$,
- b) *Drugi slučaj*: koja je normalna bez obzira na vrednost parametara,

i to sa pragom značajnosti $\alpha = 0,05$.

Prvi slučaj. Prikazivanjem elemenata uzorka na grafikonu po uzoru na kontrolnu kartu kvantil-zona dobija se bolji i jasniji prikaz izračunavanja vrednosti funkcije kvantil-zone za svaki element uzorka, a time i realizovane vrednosti v test-statistike V . Dobija se sledeća slika.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.15. Raspodela uzorka u odnosu na zone određene zadatim očekivanjem 3,15 i disperzijom $0,08^2$.

Izračunavanjem se dobija $v = 1,15$.

Granica kritične oblasti dobija se iz uslova

$$P(V \geq c | H_0) = 0,05$$

na osnovu tabele 5.4, dobija se kritična oblast $W = [1,06, 2,81]$. Kako $v \in W$, nulta hipoteza se odbacuje, tj. raspodela učešća proteina statistički značajno odstupa od raspodele $\mathcal{N}(3,15, 0,08^2)$.

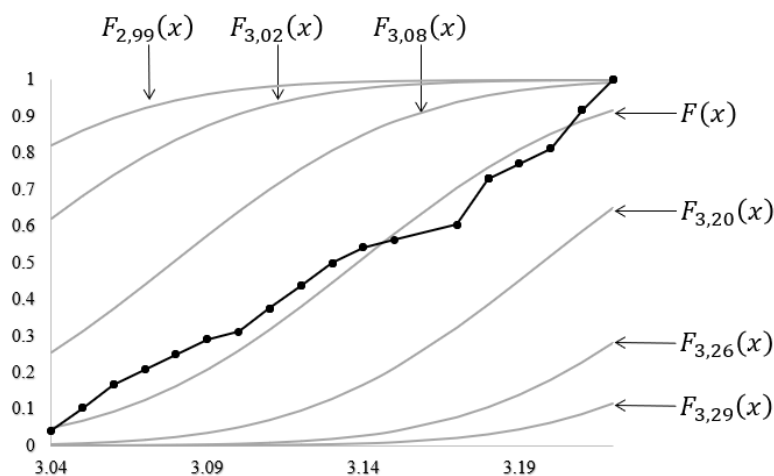
Približnim izračunavanjem p-vrednosti na osnovu tabele 5.4, dobija se

$$p = P(V \geq 1,15 | H_0) = 0 < 0,05,$$

pa postoji veoma jak dokaz protiv H_0 , tj. da raspodela statistički značajno odstupa od $\mathcal{N}(3,15, 0,08^2)$.

Drugi slučaj. Ovde su vrednosti parametara nepoznate pa se ocenjuju na osnovu uzorka. Izračunato je $\bar{x}_n = 3,14$ i $\tilde{s}_n^2 = 0,06^2$, pa se testira saglasnost raspodele učešća proteina sa raspedelom $\mathcal{N}(3,14, 0,06^2)$. Analogno prethodnom slučaju, dobija se sledeća slika.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.16. Raspedela uzorka u odnosu na zone određene ocenjenim očekivanjem 3,14 i disperzijom $0,06^2$.

Izračunavanjem se dobija $v = 1,15$.

Granica kritične oblasti dobija se iz uslova

$$P(V \geq c | H_0) = 0,025$$

na osnovu tabele 5.5, dobija se kritična oblast $W = [1,04, 2,81]$. Kako $v \in W$, nulta hipoteza se odbacuje, tj. raspodela učešća proteina statistički značajno odstupa od raspodele $\mathcal{N}(3,14, 0,06^2)$.

Približnim izračunavanjem p-vrednosti na osnovu tabele 5.5, dobija se

$$p = P(V \geq 1,15 | H_0) = 0 < 0,05,$$

pa postoji veoma jak dokaz protiv H_0 , tj. da raspodela statistički značajno odstupa od $\mathcal{N}(3,14, 0,06^2)$.

Sa slika 5.15 i 5.16, uočava se linija promena sačinjena od 17 povezanih tačaka gde je obim uzorka 48. To znači da ima ponavljajućih vrednosti što može uticati na zaključak. Takođe, može se videti da je ocenjivanje parametara optimalno pozicioniralo očekivanje i disperziju u odnosu na elemente uzorka, međutim struktura uzorka i dalje ne odgovara modelu normalne raspodele. To se može pripisati i tome što broj 3,22 koji je najveći ima frekvenciju 4. ■

5.3. Analiza moći i poređenja testova normalnosti

U ovom odeljku dati su rezultati analize moći testa zona i testa kvantil-zona pošto su oni rezultat istraživanja autora dok su ostali, navedeni u radu, poznati i uobičajeno najčešće korišćeni testovi normalnosti ili saglasnosti (Thode, 2002, str. 99-112). Detaljnije analize moći drugih testova koji su izloženi u okviru poglavlja 5 se mogu naći kod Mejsija (1951), Slektera

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

(1965), Liliforsa (1967) ili Surudžua (2008). U nekim studijama se javljaju slučajevi izračunavanja moći protiv alternativnih raspodela za koje često nije potrebno vršiti testiranje, već se zaključak o odbacivanju hipoteze može doneti na osnovu histograma, Q-Q dijagrama ili druge metode preliminarne analize normalnosti. U ovom radu je prikazano poređenje moći više često korišćenih testova, pored testova zona i kvantil-zona, i to za pojedinačne alternativne raspodele za koje, naročito kod složene nulte hipoteze, metode preliminarne analize normalnosti ne mogu dati pouzdan zaključak o odbacivanju nulte hipoteze. Samim tim je i takva detaljna analiza moći i poređenja takođe rezultat autora.

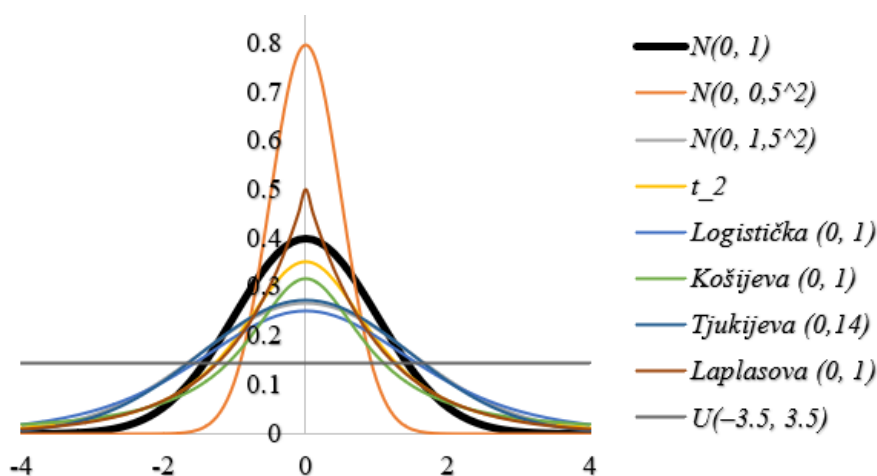
Kao što je već naglašeno u prvom odeljku ovog poglavlja, direktno izračunavanje moći nije moguće. Stoga su korišćene Monte Karlo simulacije raspodele statistike u slučajevima svih alternativnih raspodela. Nakon toga mereno je odstupanje koje dobijena raspodela ima u odnosu na onu simuliranu za nultu raspodelu polaznog obeležja. Praktično se izračunava verovatnoća da test-statistika pri tačnoj alternativnoj raspodeli bude unutar kritične oblasti određene pragom značajnosti $\alpha = 0,05$ (Avdović & Jevremović, 2022, 2023).

Za svaku alternativnu raspodelu vrednost test-statistike za obime polaznog uzorka 10, 20, 30, 50, 100, 200 su simulirane 10 000 puta korišćenjem programskog paketa MATLAB. Koristi se rezultat da 10 000 simulacija daje oblik raspodele od interesa koji se povećanjem broja simulacija ne menja značajno (Ritter i dr., 2011).

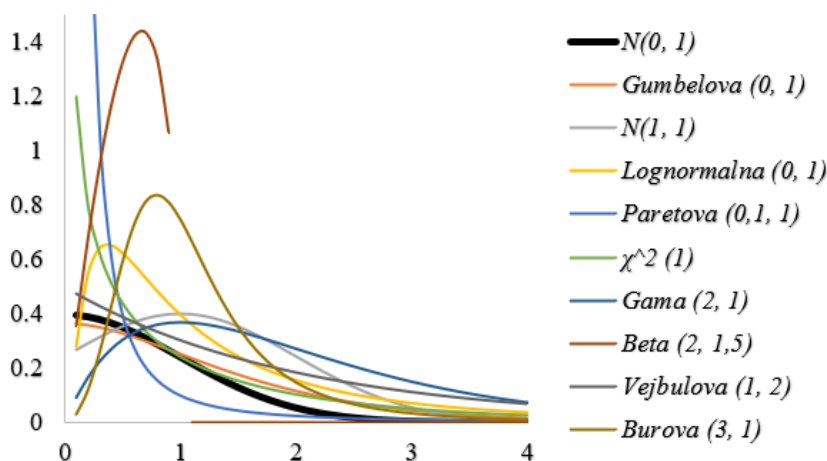
Generatori slučajnih brojeva i metode i algoritmi modeliranja tih raspodela dati su u poglavljima 3, 4 i 5, a mnogi algoritmi su implementirane MATLAB funkcije kao što su “normrnd” za normalnu raspodelu, “chi2rnd” za χ^2 raspodelu, “betarnd” za beta raspodelu, itd. Za raspodele kao što su Laplasova, Košijeva, Gumbelova, Burova i dr. nema gotovih funkcija već se algoritmi modeliranja moraju samostalno programirati po nekoj od izloženih metoda (MATLAB Help Center).

Alternativne raspodele, tj. prikaz rezultata moći za alternativne raspodele, se deli na dve grupe. Posebno za simetrične i posebno za asimetrične alternativne raspodele (Avdović & Jevremović, 2022, 2023; Arnastauskaitė, Ruzgas & Bražėnas, 2021; Noughabi, 2018; Ahmad & Khan, 2015; Razali & Wah, 2011; Sürücü, 2008). Naredne dve slike prikazuju izbor alternativnih raspodela korišćenih u ovom istraživanju.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.17. Grafici gustina simetričnih alternativnih raspodela.



Slika 5.18. Grafici gustina asimetričnih alternativnih raspodela.

Za raspodelu proste nulte hipoteze je izabrana normalna raspodela $\mathcal{N}(0,1)$, a složena nulta hipoteza pretpostavlja normalnost raspodele bez obzira na parametre. Samim tim se kod složene nulte hipoteze parametri ocenjuju. Izbor alternativnih raspodela je takav da su uzete u obzir neke od najčešće razmatranih raspodela alternativnih normalnoj raspodeli, sa parametrima takvim da su odstupanja minimalna ali da se raspodele ipak značajno razlikuju. Umeren broj alternativnih raspodela obrazložen je tendencijom za relevantnom predstavom moći testova jer bi veći broj mogao ili lažno povećavati ili lažno smanjiti predstavu o moći u analizi poređenja. Lažno povećanje bi nastalo u slučaju izbora alternativnih raspodela kod kojih je dovoljno sprovesti bilo koju metodu preliminarne analize normalnosti da bi se utvrdilo značajno odstupanje. Lažno smanjenje je u slučajevima izbora alternativne raspodele koja praktično i nije alternativna, tj. raspodele koja se u odnosu na nultu razlikuje minimalno (Avdović & Jevremović, 2022, 2023).

Kod proste nulte hipoteze izbor parametara ne utiče na opštost rezultata jer su obeležja koja su normalno raspodeljena sa različitim parametrima povezana linearnom vezom. Stoga u slučaju drugih parametara raspodele nulte hipoteze, istom transformacijom obeležja sa

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

alternativnom raspodelom za polaznu nultu raspodelu se dobija alternativna raspodela koja odgovara novoj nultoj raspodeli.

Kako se raspodele statistika A i V razlikuju za slučajeve sa poznatim parametrima (prosta nulta hipoteza) i ocenjenim parametrima (složena nulta hipoteza) nulte normalne raspodele, i vrednosti moći će se razlikovati. Stoga je analiza moći sprovedena posebno za oba slučaja.

Analiza moći i poređenja testova zona i kvantil-zona za prostu nultu hipotezu

Kvalitet performansi zavisi od moći koju studija treba da ima da bi bila dovoljno pouzdana za istraživanje od značaja. Ovde je korišćena minimalna vrednost moći od 0,8 jer se ta vrednost koristi u većini istraživanja (Jenkins & Gauvreau, 2006).

U sledećoj tabeli izložena je analiza moći testova zona i kvantil-zona za simetrične alternativne raspodele, i za slučaj proste nulte hipoteze, tj. one u kojoj su paramteri poznati. Rezultati ta dva testa su upoređeni sa rezultatom za test Kolmogorova.

Tabela 5.6. Empirijski izračunate vrednosti moći testova zona, kvantil-zona i testa Kolmogorova pri $\alpha = 0,05$ za prostu nultu hipotezu (poznate parametre) i simetrične alternativne raspodele.

Raspodela	Test	n					
		10	20	30	50	100	200
$\mathcal{N}(0, 0.5^2)$	Z	0,7242	0,9890	0,9989	1	1	1
	QZ	0,0006	0,1917	0,4570	0,8400	0,9900	1
	K	0,0522	0,1916	0,4442	0,8490	1	1
$\mathcal{N}(0, 1, 5^2)$	Z	0,4737	0,7600	0,9100	0,9946	1	1
	QZ	0,3727	0,7360	0,7676	0,8112	0,9262	1
	K	0,1374	0,1868	0,2352	0,3666	0,6838	0,9726
t_2	Z	0,3525	0,5372	0,6610	0,8188	0,9710	0,9996
	QZ	0,3305	0,7853	0,8592	0,9532	0,9990	1
	K	0,0746	0,0920	0,1056	0,1398	0,2526	0,5864
Logistička (0, 1)	Z	0,6302	0,8544	0,9068	0,9809	1	1
	QZ	0,5337	0,8824	0,9464	0,9804	0,9994	1
	K	0,1550	0,2330	0,3230	0,5116	0,8840	0,9984
Košijeva (0, 1)	Z	0,6530	0,8604	0,9455	0,9917	1	1
	QZ	0,5837	0,9508	0,9920	0,9992	1	1
	K	0,1264	0,2088	0,2756	0,4772	0,8684	1
Tjukijeva (0,14)	Z	0,4428	0,6612	0,8002	0,9286	0,9965	1
	QZ	0,3290	0,5391	0,6130	0,7608	0,9998	1
	K	0,1128	0,1528	0,1980	0,3050	0,5938	0,9344
Laplasova (0, 1)	Z	0,2210	0,8493	0,9626	0,9857	0,9900	1
	QZ	0	0,0154	0,1125	0,3846	0,9308	1

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Raspodela	Test	n					
		10	20	30	50	100	200
	K	1	1	1	1	1	1
$U(-3.5, 3.5)$	Z	0,9541	0,9984	1	1	1	1
	QZ	0,8639	0,9935	0,9991	1	1	1
	K	0,4118	0,7186	0,9080	0,9930	1	1

Z – test zona; QZ – test kvantil-zona; K – test Kolmogorova.

Kod simetričnih alternativnih raspodela test zona se pokazao najmoćnijim protiv alternativnih raspodela normalne sa parametrima 0 i 0,5, normalne sa parametrima 0 i 1,5, Tjukijeve sa parametrom 0,14 i uniformne na intervalu $[-3,5, 3,5]$. Test zona postiže moć od 0,8 pri testiranju protiv

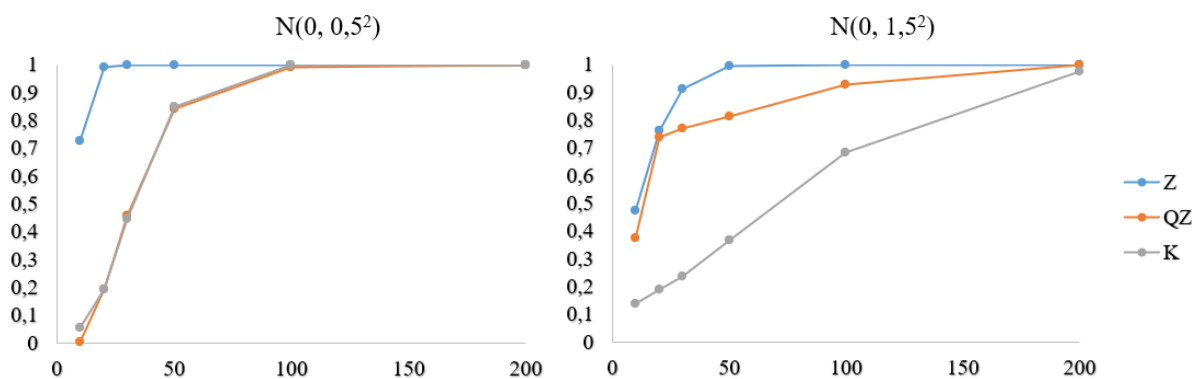
- uniformne raspodele na intervalu $[-3,5, 3,5]$ za $n < 10$,
- normalne sa parametrima 0 i 0,5, logističke sa parametrima 0 i 1, Košijeve sa parametrima 0 i 1 i Laplasove sa parametrima 0 i 1 za $n < 20$,
- normalne sa parametrima 0 i 1,5 i Tjukijeve sa parametrom 0,14 za $n < 30$,
- Studentove raspodele sa 2 stepena slobode za $n < 40$.

Test Kolmogorova je kod razmatranih alternativnih raspodela najmoćniji samo protiv Laplasove raspodele gde moć jednaku 1 već za $n \leq 10$.

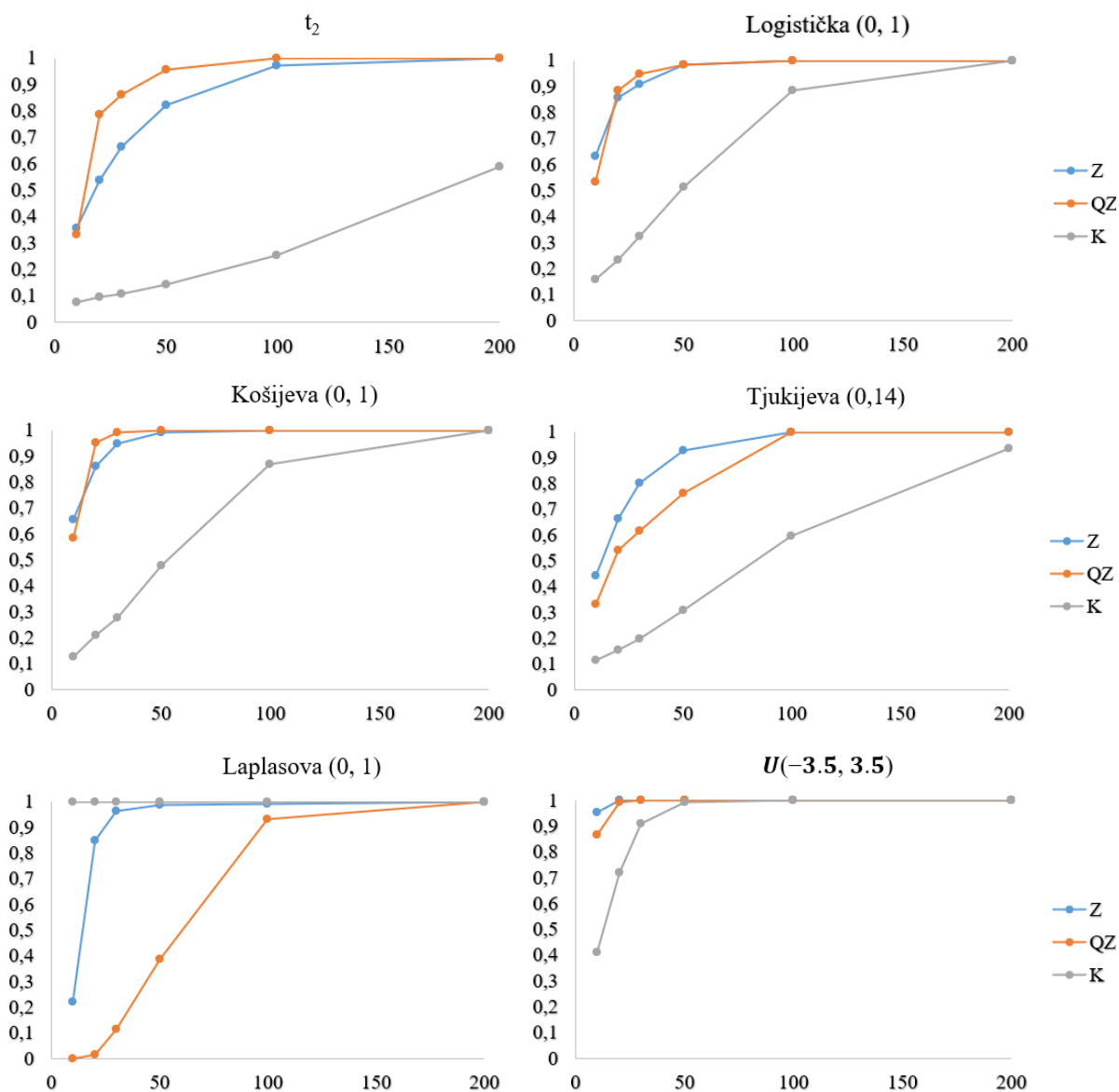
Test kvantil-zona najmoćniji je među tri razmatrana testa pri testiranju protiv alternativnih raspodela, Studentove sa 2 stepena slobode, logističke sa parametrima 0 i 1 i Košijeve sa parametrima 0 i 1.

Analogno interpretaciji kod testa zona, posmatrajći tabelu 5.6, može se zaključiti kod kojeg obima uzorka testovi kvantil-zona i Kolmogorova dostižu moć od 0,8. U tabeli 5.6 i svima koje predstoje, za svaku raspodelu boldiran je najmoćniji test i prva moć veća ili jednaka 0,8 za svaki od testova.

Dobijeni rezultati su grafički prikazani sledećom slikom.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.19. Analiza moći i poređenja za testove zona, kvantil-zona i test Kolmogorova za prostu nultu hipotezu protiv razmatrane grupe simetričnih raspodela.

Dakle, na osnovu razmatranih alternativnih raspodela, testovi zona i kvantil-zona se mogu smatrati moćnijim od često korišćenog testa Kolmogorova za klase simetričnih raspodela pri testiranju proste nulte hipoteze, tj. sa poznatim parametrima nulte normalne raspodele. Generalno posmatrano, test zona ima blagu prednost u odnosu na test kvantil-zona.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Tabela 5.7. Empirijski izračunate vrednosti moći testova zona, kvantil-zona i testa Kolmogorova pri $\alpha = 0,05$ za prostu nultu hipotezu (poznate parametre) i asimetrične alternativne raspodele.

Raspodela	Test	n					
		10	20	30	50	100	200
χ_1^2	Z	0,1744	0,2413	0,2914	0,3678	0,6197	0,7514
	QZ	0,3441	1	1	1	1	1
	K	1	1	1	1	1	1
Gumbelova (0, 1)	Z	0,1986	0,2844	0,3658	0,4797	0,7553	0,9084
	QZ	0,1661	0,5122	0,6077	0,7526	0,9362	0,9994
	K	0,1974	0,3210	0,4696	0,6694	0,9318	0,9996
Burova (3, 1)	Z	0,3735	0,5923	0,7203	0,8939	0,9917	1
	QZ	1	1	1	1	1	1
	K	1	1	1	1	1	1
Paretova (0,1, 1)	Z	1	1	1	1	1	1
	QZ	1	1	1	1	1	1
	K	1	1	1	1	1	1
$\mathcal{N}(1, 1)$	Z	0,3850	0,6215	0,7977	0,9378	0,9974	1
	QZ	0,5233	0,8161	0,8863	0,9400	0,9730	1
	K	0,7664	0,9716	1	1	1	1
Lognormalna (0, 1)	Z	0,4822	0,7294	0,8721	0,9701	0,9998	1
	QZ	0,8529	1	1	1	1	1
	K	1	1	1	1	1	1
Vejbulova (1, 2)	Z	0,8380	0,9783	0,9965	1	1	1
	QZ	0,9442	1	1	1	1	1
	K	1	1	1	1	1	1
Gama (2, 1)	Z	0,9598	0,9984	1	1	1	1
	QZ	1	1	1	1	1	1
	K	1	1	1	1	1	1
Beta (2, 1,5)	Z	1	1	1	1	1	1
	QZ	0,1900	1	1	1	1	1
	K	1	1	1	1	1	1

Z – test zona; QZ – test kvantil-zona; K – test Kolmogorova.

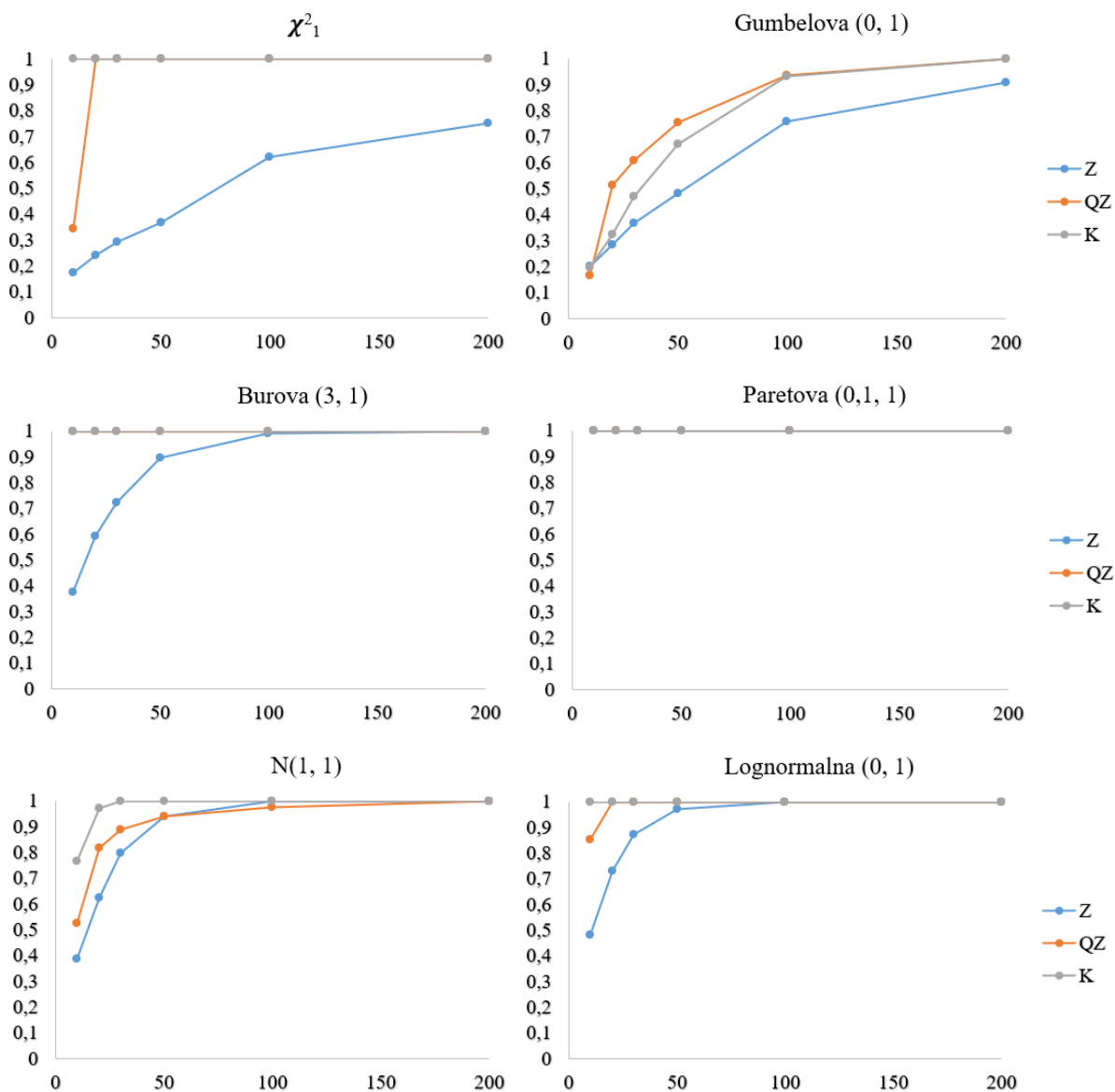
Rezultati ukazuju da je test Kolmogorova moćniji od testova zona i kvantil-zona, ili jednako moćan, u većini razmatranih asimetričnih raspodela. Ipak, rezultati su takvi da odstupanja performansi testova zona i kvantil-zona nisu značajna. Dodatno, iako je bitno da test detektuje odstupanje od normalne raspodele bilo da je prava raspodela simetrična ili asimetrična, u većini slučajeva asimetričnih raspodela za taj zaključak je dovoljan histogram ili neka druga metoda preliminarne analize normalnosti

Test zona najslabije performanse pokazuje pri testiranju protiv χ_1^2 raspodele i Gumbelove raspodele sa parametrima 0 i 1. Moć od 0,8 se za prvu raspodelu postiže sa $n > 200$, a za drugu sa $n \approx 150$. Test kvantil-zona pokazuje dobre performanse kod svih raspodela.

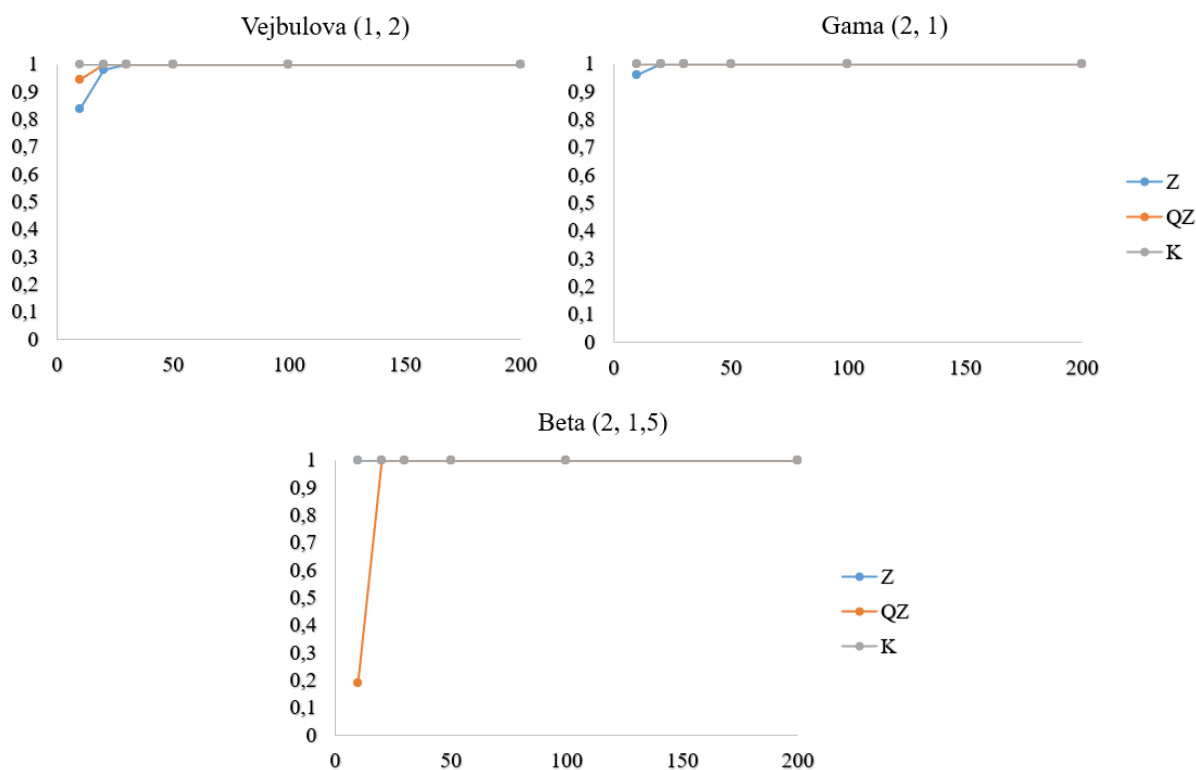
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Samo kod Gumbelove raspodele postiže moć od 0,8 tek sa $n \approx 60$, ali kod ostalih alternativnih raspodela za $n < 20$.

Sledeća slika grafički ilustruje dobijene rezultate i pokazuje dobre performanse testova zona i kvantil-zona u odnosu na test Kolmogorova iako ga ne nadmašuju.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.20. Analiza moći i poređenja za testove zona, kvantil-zona i test Kolmogorova za prostu nultu hipotezu protiv razmatrane grupe asimetričnih raspodela.

Dakle, kod obe klase alternativnih raspodela, testovi zona i kvantil-zona su pri testiranju proste nulte hipoteze, tj. hipoteze normalnosti sa poznatim parametrima, konkurentni testu Kolmogorova, a samim tim i drugim testovima normalnosti. Dodatno, definicija statistike testa zona je prosta što čini primenu ovog testa mnogo bržom i lakšom u odnosu na ostale. Test Kolmogorova i test kvantil-zona imaju približne brzine primene jer se u oba empirijska funkcija raspodele upoređuje sa teorijskom funkcijom raspodele.

Analiza moći i poređenja testova zona i kvantil-zona za složenu nultu hipotezu

U slučaju složene nulte raspodele normalnosti parametri nisu poznati i ocenjuju se uzoračkom sredinom i popravljenom uzoračkom disperzijom. Kao što je to slučaj kod većine testova normalnosti, a i testova zona i kvantil-zona, ocena parametara utiče na promenu raspodele test statistike tako da su njeni kvantili u odnosu na raspodelu bez ocene parametara bliži očekivanju test statistike. Samim tim, pri izračunavanju moći testa kod ove nulte hipoteze za svaki simulirani uzorak se parametri ocenjuju i za te vrednosti parametara i raspodelu test statistike dobijenu ocenjivanjem parametara, se procenjuje moć testa (kao relativna frekvencija simuliranih uzoraka za koje se nulta hipoteza odbacuje – Algoritam 5.1).

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

U sledećoj tabeli dati su rezultati analize moći testova zona, kvantil-zona, Lilifors testa, Pirsonovog χ^2 testa i Anderson-Darling testa za simetrične alternativne raspodele, pri testiranju složene nulte hipoteze, tj. hipoteze normalnosti za nepoznate parametre.

Tabela 5.8. Empirijski izračunate vrednosti moći testova zona, kvantil-zona i testova Liliforsa, χ^2 i Anderson-Darling testa pri $\alpha = 0,05$ za složenu nultu hipotezu (poznate parametre) i simetrične alternativne raspodele.

Raspodela	Test	n					
		10	20	30	50	100	200
$\mathcal{N}(0, 0.5^2)$	Z	0,0492	0,0452	0,0498	0,0520	0,0500	0,0546
	QZ	0,0495	0,0482	0,0531	0,0480	0,0449	0,0419
	L	0,0474	0,0510	0,0512	0,0530	0,0514	0,0560
	H	0	0	0,0364	0,0630	0,0556	0,0548
	AD	0,0462	0,0558	0,0504	0,0548	0,0476	0,0474
$\mathcal{N}(10, 15^2)$	Z	0,0471	0,0499	0,0484	0,0515	0,0488	0,0514
	QZ	0,0492	0,0541	0,0476	0,0511	0,0527	0,0497
	L	0,0530	0,0523	0,0509	0,0521	0,0481	0,0506
	H	0	0	0,0343	0,0620	0,0562	0,0574
	AD	0,0525	0,0504	0,0505	0,0520	0,0495	0,0487
Tjukijeva (0,14)	Z	0,0485	0,0483	0,0477	0,0472	0,0489	0,0535
	QZ	0,0500	0,0469	0,0533	0,0486	0,0386	0,0299
	L	0,0496	0,0536	0,0484	0,0494	0,0476	0,0526
	H	0	0	0,0334	0,0580	0,0536	0,0610
	AD	0,0490	0,0540	0,0454	0,0490	0,0478	0,0478
t_2	Z	0,1856	0,4565	0,6068	0,7965	0,9679	0,9994
	QZ	0,2364	0,3726	0,4496	0,6445	0,8003	0,8942
	L	0,2682	0,4502	0,6008	0,7846	0,9628	0,9986
	H	0	0	0,5408	0,7006	0,8856	0,8780
	AD	0,2994	0,5356	0,6726	0,8562	0,9870	0,9999
Logistička (0, 1)	Z	0,0641	0,1229	0,1367	0,1809	0,2808	0,4578
	QZ	0,0812	0,1031	0,1269	0,1615	0,2160	0,3741
	L	0,0762	0,0926	0,0956	0,1096	0,1530	0,2358
	H	0	0	0,0954	0,1036	0,1152	0,1812
	AD	0,0788	0,1056	0,1220	0,1638	0,2398	0,3958
Košijeva (0, 1)	Z	0,3753	0,7483	0,8878	0,9825	1	1
	QZ	0,4819	0,6821	0,7787	0,9082	0,9748	0,9920
	L	0,5648	0,8468	0,9482	0,9950	1	1
	H	0	0	0,7073	0,7758	0,6782	0,4505
	AD	0,6144	0,8786	0,9658	0,9968	1	1
Laplasova (0, 1)	Z	0,5344	0,7933	0,8741	0,9319	0,9731	0,9869
	QZ	0,5000	0,5000	0,5000	0,5176	0,8844	1
	L	0,0712	0,1276	0,2248	0,5744	0,9962	1
	H	0	0	0,1174	0,4530	0,8576	0,9974
	AD	1	1	1	1	1	1

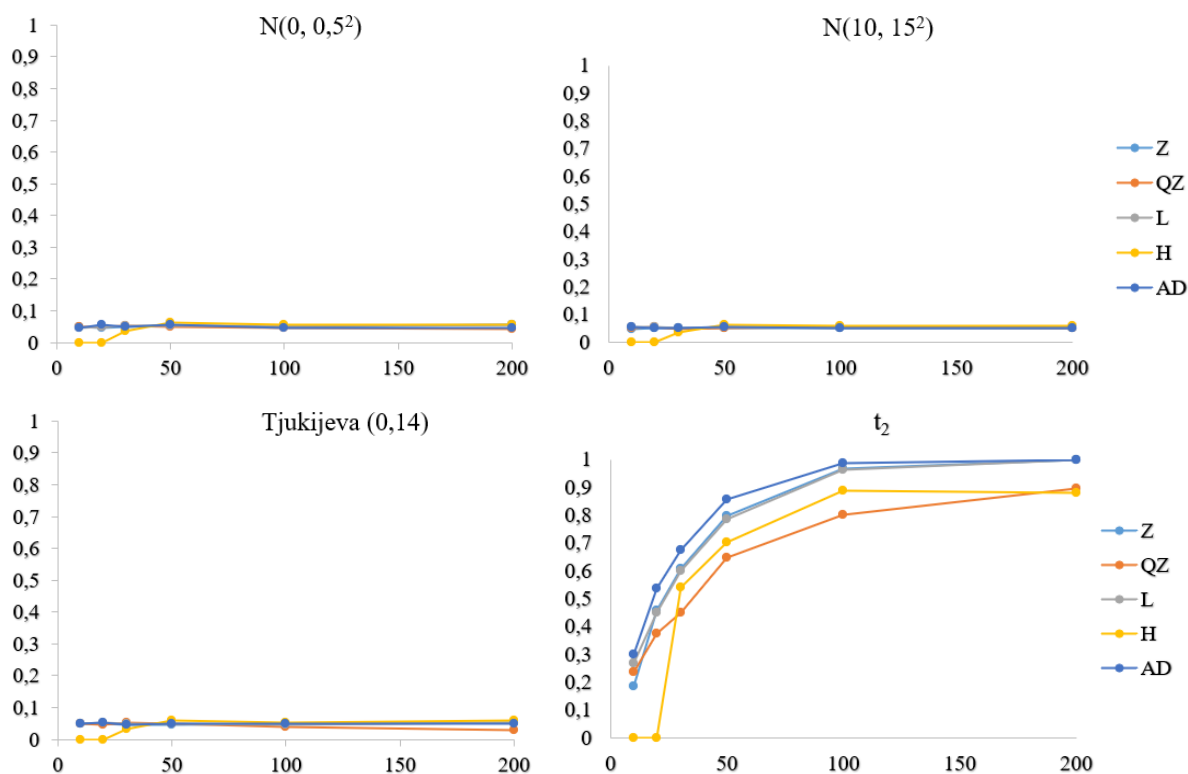
FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Raspodela	Test	n					
		10	20	30	50	100	200
$U(-3.5, 3.5)$	Z	0,1257	0,1748	0,4354	0,5289	0,7833	0,9654
	QZ	0,0629	0,0563	0,0547	1	1	1
	L	0,0606	0,0980	0,1444	0,2566	0,5912	0,9448
	H	0	0	0	0,1750	0,4123	0,8201
	AD	0,0720	0,1614	0,2942	0,5838	0,9484	1

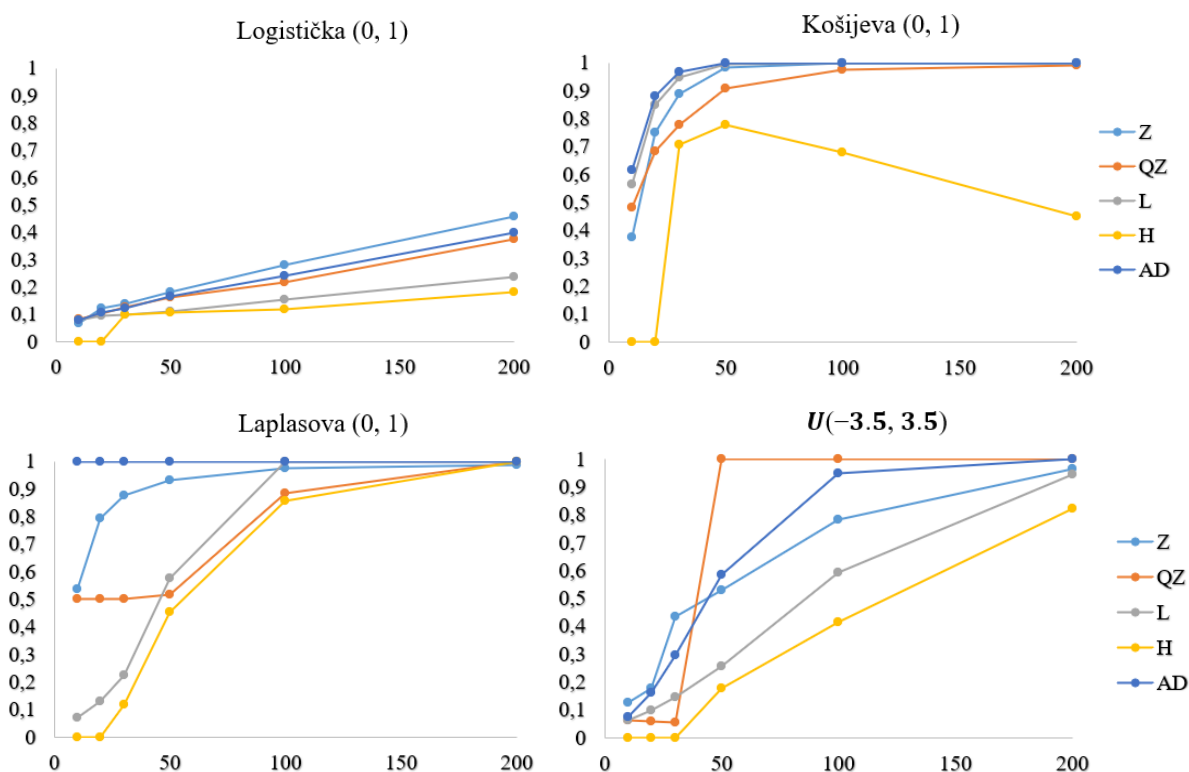
Z – test zona; QZ – test kvantil-zona; L – Lilifors test; H - χ^2 test; AD – Anderson-Darling test.

Normalna raspodela sa parametrima 0 i 0,5, kao i ona sa parametrima 10 i 15, se mogu smatrati nultim raspodelama jer je hipoteza složena pa vrednost parametara ne utiče na zaključak (Teoreme 5.2 i 5.3). Isto važi za Tjukijevu raspodelu sa parametrom 0,14 jer je ista približno ekvivalentna normalnoj raspodeli sa parametrima 0 i 2,142 (Johnson, Kotz & Balakrishnan, 1994, str. 41). Tada, kako se testiranje ovde vrši sa pragom značajnosti od 0,05, moć testa, koja se tada naziva merom testa, treba biti približno 0,05.

Među ostalim simetričnim raspodelama u slučajevima Studentove raspodele sa 2 stepena slobode, Košijeve raspodele sa parametrima 0 i 1 i Laplasove raspodele sa parametrima 0 i 1 najmoćniji test je Anderson-Darling test, a kod logističke raspodele sa parametrima 0 i 1 i uniformne raspodele na intervalu $[-3.5, 3.5]$ najmoćniji je test zona.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.21. Analiza moći i poređenja za testove zona, kvantil-zona, Lilifors testa, Hi-kvadrat testa Anderson-Darling testa za složenu nultu hipotezu protiv razmatrane grupe simetričnih raspodela.

Bitno je napomenuti da se u slučaju simetričnih alternativnih raspodela vrednosti moći za test zona, Anderson-Darling test i Lilifors test ne razlikuju приметно, tj. imaju slične performanse i test zona im je Konkurentan. Test kvantil-zona uglavnom postiže dobre vrednosti moći tek za velike obime uzorka. Pirsonov hi-kvadrat test ima najslabije performanse zato što za uzorke obima manjeg ili jednakog 20 daje nepouzdan rezultate, a i za uzorke većeg obima uglavnom ima najmanju moć.

U sledećoj tabeli data je analogna analiza moći i poređenja svih testova za asimetrične alternativne raspodele.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Tabela 5.9. Empirijski izračunate vrednosti moći testova zona, kvantil-zona i testova Liliforsa, χ^2 i Anderson-Darling testa pri $\alpha = 0,05$ za složenu nultu hipotezu (nepoznate parametre) i asimetrične alternativne raspodele.

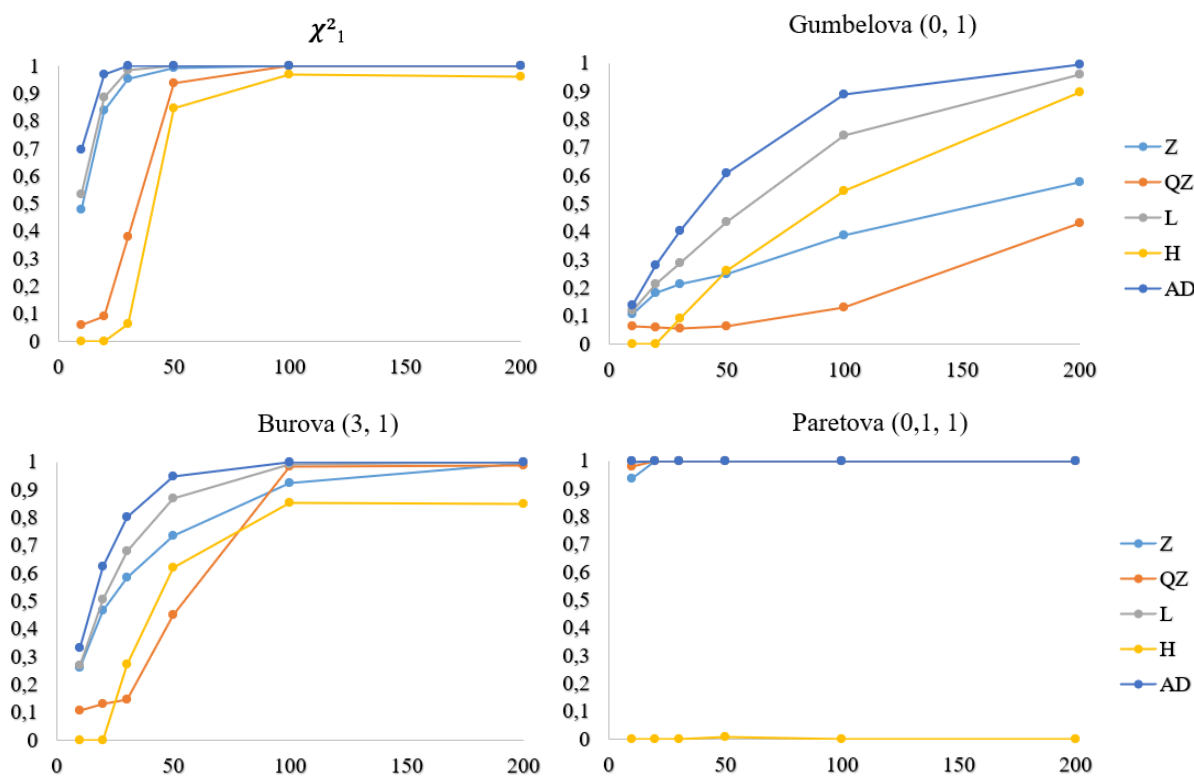
Raspodela	Test	n					
		10	20	30	50	100	200
χ^2_1	Z	0,4759	0,8397	0,9536	0,9936	0,9997	1
	QZ	0,0562	0,0914	0,3805	0,9350	1	1
	L	0,5335	0,8866	0,9836	0,9999	1	1
	H	0	0	0,0624	0,8458	0,9673	0,9587
	AD	0,6943	0,9672	0,9989	0,9999	1	1
Gumbelova (0, 1)	Z	0,1041	0,1821	0,2126	0,2455	0,3860	0,5766
	QZ	0,0622	0,0561	0,0544	0,0623	0,1297	0,4304
	L	0,1163	0,2103	0,2853	0,4318	0,7403	0,9597
	H	0	0	0,0885	0,2582	0,5451	0,8961
	AD	0,1375	0,2798	0,4016	0,6065	0,8893	0,9940
Burova (3, 1)	Z	0,2583	0,4659	0,5819	0,7351	0,9234	0,9935
	QZ	0,1063	0,1274	0,1436	0,4504	0,9819	0,9870
	L	0,2665	0,5027	0,6775	0,8663	0,9900	1
	H	0	0	0,2694	0,6204	0,8513	0,8468
	AD	0,3320	0,6222	0,7987	0,9457	0,9993	1
Paretova (0,1, 1)	Z	0,9360	0,9991	1	1	1	1
	QZ	0,9798	1	1	1	1	1
	L	0,9990	1	1	1	1	1
	H	0	0	0	0,0047	0,0003	0
	AD	0,9999	1	1	1	1	1
Lognormalna (0, 1)	Z	0,4158	0,7420	0,8686	0,9624	0,9986	1
	QZ	0,1557	0,1698	0,4638	0,9047	0,9911	1
	L	0,4544	0,7040	0,9288	0,9955	1	1
	H	0	0	0,2182	0,7210	0,8466	0,7577
	AD	0,5755	0,9048	0,9836	0,9998	1	1
Vejbulova (1, 2)	Z	0,3002	0,5309	0,6252	0,7442	0,9016	0,9852
	QZ	0,0692	0,0751	0,1403	0,6571	0,9987	1
	L	0,3060	0,5879	0,7779	0,9617	0,9999	1
	H	0	0	0,1140	0,4518	0,8664	0,9969
	AD	0,4158	0,7778	0,9323	0,9974	1	1
Gama (2, 1)	Z	0,1567	0,2553	0,2851	0,3456	0,4976	0,6845
	QZ	0,0450	0,0525	0,0601	0,0948	0,5888	0,9943
	L	0,1715	0,3235	0,4723	0,6985	0,9509	0,9996
	H	0	0	0,0726	0,3218	0,6135	0,9407
	AD	0,2171	0,4585	0,6545	0,8898	0,9974	1

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

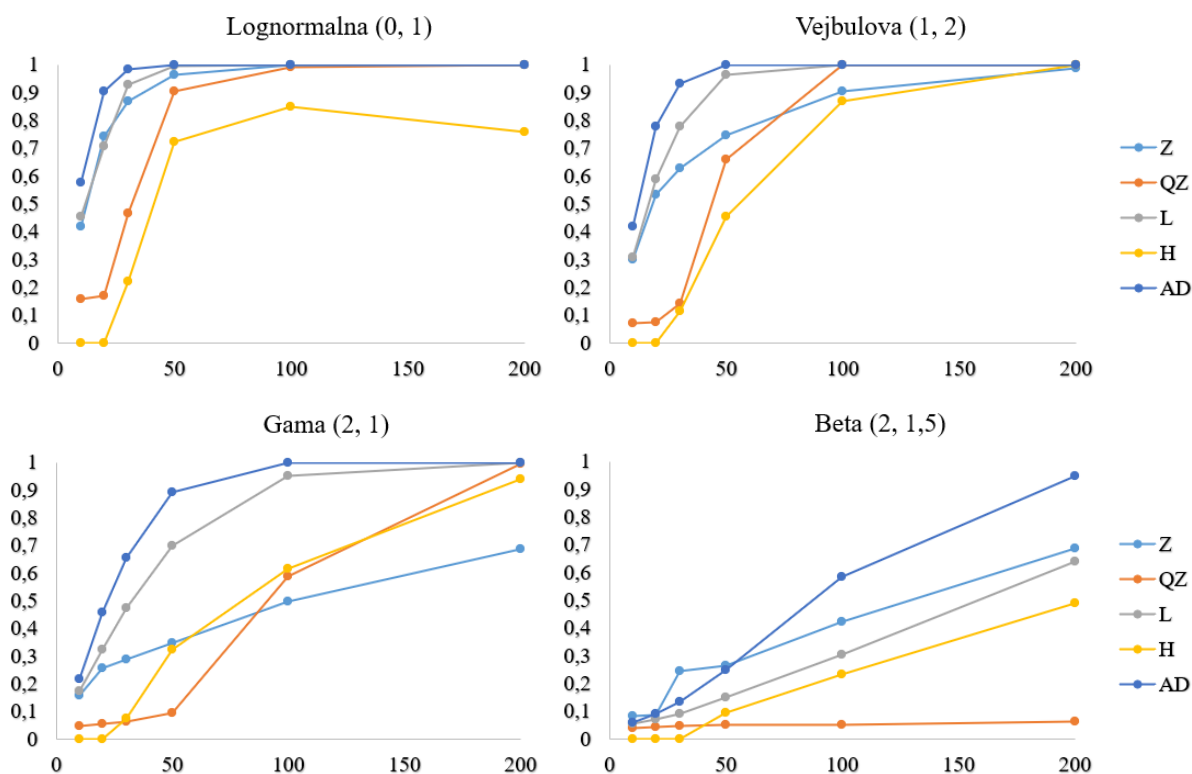
Raspodela	Test	n					
		10	20	30	50	100	200
Beta (2, 1,5)	Z	0,0823	0,0840	0,2432	0,2644	0,4224	0,6846
	QZ	0,0381	0,0413	0,0479	0,0511	0,0499	0,0635
	L	0,0534	0,0694	0,0907	0,1473	0,3019	0,6377
	H	0	0	0,0005	0,0951	0,2318	0,4893
	AD	0,0573	0,0880	0,1343	0,2491	0,5823	0,9452

Z – test zona; QZ – test kvantil-zona; L – Lilifors test; H - χ^2 test; AD – Anderson-Darling test.

Pri analizi moći u testiranju normalnosti protiv asimetričnih alternativnih raspodela Anderson-Darling test se pokazao najmoćnijim za sve raspodele osim za beta raspodelu sa parametrima 2 i 1,5 gde, iako Anderson-Darling test prvi postiže moć od 0,8, za većinu obima uzorka najmoćniji je test zona. Ipak, kao i kod simetričnih raspodela test zona i Lilifors test konkurentni su Anderson-Darling testu. Hi-kvadrat test pokazuje slabije performanse nego u slučaju simetričnih raspodela, a test kvantil-zona je za velike obime uzoraka takođe konkurentan ostalim, s tim da je kod dosta raspodela manje moćan za uzorke malog obima. Ti rezultati su prikazani sledećom slikom.



FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA



Slika 5.22. Analiza moći i poređenja za testove zona, kvantil-zona, Lilifors testa, Hi-kvadrat testa Anderson-Darling testa za složenu nultu hipotezu protiv razmatrane grupe asimetričnih raspodela.

Dakle, testovi zona i kvantil-zona su se za izabrane simetrične alternativne raspodele pokazali konkurentnim ili moćnijim od ostalih, konvencionalno najčešće korišćenih testova normalnosti. U slučaju asimetričnih alternativnih raspodela dolazi do drugačijih rezultata. Test kvantil-zona za poznate parametre se pokazao kao manje moćan u odnosu na konkurentne testove osim u slučaju velikih uzoraka. Uzevši u obzir prosta izračunavanja i njihovu brzinu kod testa zona, njegova primenljivost biva dodatno naglašena.

Sužavanjem zona i kvantil-zona za dva nova testa može doći do povećanja moći testa, jer se time manja odstupanja mogu bolje detektovati. Takođe, povećanjem broja zona, iako bi testovi izgubili na brzini, takođe bi postali osetljiviji na promene u raspodeli i samim tim bi postali moćniji. Ako se analogan pristup primeni na ostale raspodele, testovi zona i kvantil-zona se mogu generalizovati za testiranje saglasnosti sa bilo kojom raspodelom.

Zaključak

Ovaj rad je rezultat istraživanja iz oblasti generatora pseudoslučajnih brojeva, modeliranja slučajnih veličina, kontrole kvaliteta i testiranja normalnosti i kao takav donosi rezultate tim oblastima, a ti su rezultati navedeni u nastavku teksta kao i njihovi doprinosi oblastima istraživanja. Pored toga tekst pruža detaljan pregled literature iz navedenih oblasti kao i mnogih drugih knjiga i članaka vezanih za te oblasti.

Pregled literature ima za cilj da potkrepi postojeća znanja koja su navedena u radu, uputi čitaoce na izvore od značaja, ilustruje razvoj oblasti i ukaže na potrebe njenog dodatnog unapređivanja i osavremenjivanja, i istakne dobijene rezultate sa raznim pojedinostima i teorijskom obradom. Pregled literature je najvećim delom dat u uvodu u rad, mada se i kroz tekst ostalih poglavlja nalazi njegov značajan deo, uglavnom preko kratkog uvoda u svako pojedinačno poglavlje.

U drugom poglavlju, dat je pregled teorije o slučajnim i pseudoslučajnim brojevima i generatorima pseudoslučajnih brojeva. Dati su razni generatori pseudoslučajnih brojeva, sa naročitim osvrtom na linearni kongruentni generator i one koji su dobijeni njegovim unapređenjem, tj. raznim modifikacijama. Kao dodatak odeljku sa generatorima pseudoslučajnih brojeva, dat je generator dobijen kao deo istraživanja autora, tzv. linearni kongruentni generatori sa njegovim proširenjima. Date su definicije proširenja u raznim slučajevima sa osobinama generatora koje preciziraju osobine slučajnosti generisanih brojeva i uticaja proširenja na period linearnog kongruentnog generatora. U posebnom odeljku obrađeni su neki testovi slučajnosti koji se mogu koristiti za testiranje kvaliteta brojeva generisanih izloženim generatorima. Konačno, dato je poređenje linearnog kongruentnog generatora sa aditivnim proširenjem i najčešće korišćene varijante Mersenovog Tvistera testiranjem slučajnosti generisanih brojeva.

U trećem poglavlju obrađeno je modeliranje slučajnih veličina. Prvo su date razne metode modeliranja kako diskretnih, tako i neprekidnih slučajnih veličina. Nakon toga, obrađene su raspodele verovatnoća od značaja za istraživanje kao i algoritmi preko kojih se mogu modelirati slučajne veličine sa tim raspedelama, tj. uzorkovati nizovi brojeva koji odgovaraju tim raspedelama. Dat je detaljan matematički prikaz prilagođenja algoritma prihvatanja i odbacivanja za modeliranje diskretnih slučajnih veličina, i metode diskretizacije za modeliranje raspodela apsolutno neprekidnih slučajnih veličina. Normalna raspodela je zbog svog značaja obrađena detaljnije, a dodatno je ilustrovan i algoritam dobijen u ovom istraživanju. Taj algoritam se može shvatiti kao primena funkcije zone. Ukratko je data i analiza poređenja dobijenog algoritma sa drugim algoritmima modeliranja normalne raspodele.

U četvrtom poglavlju, prvi je poseban odeljak posvećen empirijskoj funkciji raspodele jer je ona veoma značajna za ovaj rad, a posebno za ovo poglavlje. Spomenute su poznate osobine empirijske funkcije raspodele date teoremama Glivenka i Kantelija, Kolmogorova i Smirnova. Poznate, često korišćene kontrolne karte, kao što su Šuhartova kontrolna karta i karte zasnovane na istom principu, CUSUM kontrolna karta i EWMA karta su takođe

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

obrađene. Dodatno, obrazloženo je dobijanje novih kontrolnih karata dobijenih ovim istraživanjem sa raznim unapređenjima tumačenja kontrolnih karata koja između ostalog koriste funkcije zone i kvantil-zone. Dobijene kontrolne karte su EFR (empirijska funkcija raspodele) karta, karta kvantil-zona, Q-Q karta i karta p-vrednosti. Primena svake od novih kontrolnih karata ilustrovana je na primeru. Za kontrolne karte kod kojih je to moguće, istaknuta je mogućnost primene u preliminarnoj analizi normalnosti. Na kraju poglavlja tabelirane su vrednosti ARLa za najznačajnije izložene kontrolne karte i izvršena je analiza tih performansi sa poređenjem. Vrednosti ARLa izračunate su za normalnu raspodelu obeležja na osnovu kojeg se proces kontroliše i za razne slučajeve pomerene vrednosti očekivanja te raspodele. Celo poglavlje praćeno je brojnim ilustrativnim grafikonima. EFR karta je dala najbolje rezultate, a povremeno se dobro pokazala i karta kvantil-zona.

Završno poglavlje se pretežno bavi testiranjem hipoteza saglasnosti, preciznije normalnosti. Inicijalno su izložene osnovne, najvažnije teorijske napomene vezane za uvođenje pojmova moći i p-vrednosti testa i njihovo opisivanje. Kada su u pitanju ove dve metode statističke analize, a naročito p-vrednosti, malo je izvora koji omogućavaju dovoljno dobar pregled teorije iza ovih pojmova kao što nema ni njihove ilustracije niti adekvatnih tehnika tumačenja. Sa druge strane, može se naći mnogo novijih radova koji ističu loše posledice tih pristupa. Jedan deo ovog rada je doprinos rešenju navedenog problema. U ovom delu rada dat je postupak kojim se testovi normalnosti mogu koristiti u testiranju saglasnosti sa bilo kojom raspodelom.

Nakon toga, dati su pregledi najčešće korišćenih testova saglasnosti od važnosti za ovo istraživanje kao što su test Kolmogorova, Lilifors test, Hi-kvadrat test, Anderson-Darling test i Šapiro-Vilk test. Pored toga izvršen je osvrt na neke karakteristike ovih testova koje bi mogle biti kako problematične, tako i od koristi. Pored tih testova izložena su dva nova testa koji su rezultat istraživanja autora. Ti testovi slede iz osobina funkcija zone i kvantil-zone. Date su test-statistike, definisane, opisane i tabelirane njihove raspodele i izvršena detaljna ilustracija grafika gustina i funkcija raspodele rezultujućih test-statistika. Sledeći odeljak nudi analizu moći novih testova i analizu poređenja dobijenih vrednosti moći sa inicijalno izloženim postojećim testovima. Poređenje je izvršeno odvojeno za simetrične i asimetrične alternativne raspodele. Izbor raspodela alternativnih hipoteza je obrazložen, a rezultati su tabelirani i grafički ilustrovani. Testovi zona i kvantil-zona su po uobičajenim pokazateljima konkurentni u odnosu na ostale razmrtrane testove, a u nekim slučajevima i bolji od njih.

Buduća istraživanja se mogu odnositi na unapređivanje i dodatne razrade dobijenih rezultata, dobijanje novih boljih rezultata, na primer sužavanjem zona i kvantil-zona, i povezivanje sa tehnikama multidisciplinarnе primene (npr. Gini indeks), primena karakterizacije raspodela u kontroli kvaliteta i testiranju normalnosti, analiza značaja statistika poretka, senzitivnost testova na spoljašnje vrednosti, testiranje hipoteza nad cenzurisanim podacima, generalizacija rezultata na multivarijantne slučajeve itd.

Reference

1. Aboraya, M.M., Yousof, H., Hamedani G.G., & Ibrahim, M. (2020). A New Family of Discrete Distributions with Mathematical Properties, Characterizations, Bayesian and Non-Bayesian Estimation Methods. *Mathematics*, 8(10):1648. <https://doi.org/10.3390/math8101648>.
2. Adegoke, N. A., Smith, A. N. H., Anderson, M. J., Sanusi, R. A., & Pawley, M. D. M. (2019). Efficient Homogeneously Weighted Moving Average Chart for Monitoring Process Mean Using an Auxiliary Variable. *IEEE Access*, 7, 94021-94032. <https://doi.org/10.1109/ACCESS.2019.2926533>.
3. Ahmad, F., & Khan, R. A. (2015). A power comparison of various normality tests. *Pakistan Journal of Statistics and Operation Research*, 11(3), 331-345. <https://doi.org/10.18187/pjsor.v11i3.845>.
4. Ajadi, J.O., Zwetsloot, I.M., & Tsui, K.-L. (2021). A New Robust Multivariate EWMA Dispersion Control Chart for Individual Observations. *Mathematics*, 9, 1038. <https://doi.org/10.3390/math9091038>.
5. Almetwally, E.M., Abdo, D.A., Hafez, E.H., Jawa, T.M., Sayed-Ahmed, N., & Almongy, H.M. (2021). The new discrete distribution with application to COVID-19 Data. *Results in Physics*, 32, 10498. <https://doi.org/10.1016/j.rinp.2021.104987>.
6. Alzaatreh, A., Lee, C., Famoye, F. & Ghosh, I. (2016). The generalized Cauchy family of distributions with applications. *Journal of Statistical Distributions and Applications*, 3:12. <https://doi.org/10.1186/s40488-016-0050-3>.
7. Anderson, T.W., & Darling, D.A. (1952). Asymptotic Theory of Certain “Goodness of Fit” Criteria Based on Stochastic Processes. *Ann. Math. Stat.* 23: 193–212. <https://doi.org/10.1214/aoms/1177729437>.
8. Anderson, T.W., & Darling, D.A. (1954). A test of goodness of fit. *J. Am. Stat. Assoc.* 49: 765–769. <https://doi.org/10.1080/01621459.1954.10501232>.
9. Arnastauskaitė, J., Ruzgas, T. & Bražėnas, M. (2021). An Exhaustive Power Comparison of Normality Tests. *Mathematics*, 9, 788. <https://doi.org/10.3390/math9070788>.
10. Avdović, A. (2023). Extension of Linear Congruential Generator. *Scientific Publications of the State University of Novi Pazar, Series A, Applied Mathematics, Informatics and Mechanics*, 15, 2, 87-95. <https://doi.org/10.46793/SPSUNP2302.087A>.
11. Avdović, A., & Jevremović, V. (2023). Discrete Parameter-Free Zone Distribution and Its Application in Normality Testing. *Axioms*, 12(12): 1087. <https://doi.org/10.3390/axioms12121087>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

12. Avdović A., & Jevremović, V. (2022). Quantile-Zone Based Approach to Normality Testing. *Mathematics*, 10(11):1828. <https://doi.org/10.3390/math10111828>.
13. Avdović, A. (2019). Metode modeliranja neprekidnih slučajnih veličina i primena u statistici – Master rad. Novi Pazar: Državni univerzitet u Novom Pazaru.
14. Babu, G. J., & Rao, C. R. (2004). Goodness-of-Fit Tests When Parameters Are Estimated. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 66(1), 63–74. <http://www.jstor.org/stable/25053332>.
15. Bakir, S. (2012). A non-parametric Shewhart-type quality control chart for monitoring broad changes in a process distribution. *Int. journal of quality statistics and reliability*, vol. 2012. <https://doi.org/10.1155/2012/147520>.
16. Bakshaev, A. (2009). Goodness of fit and homogeneity tests on the basis of N-distances. *Journal of Statistical Planning and Inference*. 139, 3750–3758. <https://doi.org/10.1016/j.jspi.2009.05.014>.
17. Barnard, G. A. (1959). Control Charts and Stochastic Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 21(2), 239–271. <http://www.jstor.org/stable/2983801>.
18. Betensky, R. A. (2019). The p-Value Requires Context, Not a Threshold. *The American Statistician*, 73:sup1, 115-117. <https://doi.org/10.1080/00031305.2018.1529624>.
19. Borovkov, A. A. (2013). *Probability theory*. London: Springer.
20. Boyerinas, B.M. (2016). Determining the Statistical Power of the Kolmogorov-Smirnov and Anderson-Darling Goodness-of-Fit Tests via Monte Carlo Simulation. *CNA Analysis and Solutions*. https://www.cna.org/CNA_files/PDF/DOP-2016-U-014638-Final.pdf.
21. Brent, R. P. (2004). Note on Marsaglia's Xorshift Random Number Generators. *Journal of Statistical Software*, 11(5), 1–5. <https://doi.org/10.18637/jss.v011.i05>.
22. Cabrera, J., & McDougall, A. (2002). *Statistical Consulting*. New York: Springer.
23. Carsey, T. M., & Harden, J. J. (2014). *Monte Carlo simulation and resampling methods for social science*. SAGE Publications, Inc., <https://doi.org/10.4135/9781483319605>.
24. Cochran, W.G. (1952). The χ^2 Test of Goodness of Fit. *Ann. Math. Statist.* 23(3): 315-345. <https://doi.org/10.1214/aoms/1177729380>.
25. Coin, D. (2008). A goodness-of-fit test for normality based on polynomial regression. *Computational Statistics & Data Analysis*. 52, 2185–2198. <https://doi.org/10.1016/j.csda.2007.07.012>.
26. Dahiru T. (2008). P - value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine*, 6(1), 21–26. <https://doi.org/10.4314/aipm.v6i1.64038>.
27. Desgagné, A. & Lafaye de Micheaux, P. (2017). A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

- kurtosis, using the Rao's score test on the APD family. *Journal of Applied Statistics*. 2017, 45, 2307–2327. <https://doi.org/10.1080/02664763.2017.1415311>.
28. Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer.
 29. Dura, C., & Isac, C. (2006). The ANOVA Method in Quality Audit. *Annals of the University of Petroșani, Mechanical Engineering*, 8, 23-30. <https://www.upet.ro/annals/mechanical/pdf/2006/Annals-Mechanical-Engineering-2006-a3.pdf>.
 30. Đorić, D., Jevremović, V., Mališić, J., & Đorić – Nikolić, E. (2007). *Atlas raspodela*. Beograd: Građevinski fakultet.
 31. Eichenauer, J., Grothe, H., Lehn J., & Topuzoğlu, A. (1987). A multiple recursive non-linear congruential pseudo random number generator. *Manuscripta Mathematica*, 59 (1987), 331–346. <https://doi.org/10.1007/BF01174798>.
 32. Fabiano, N., Gardašević-Filipović, M., Mirkov, N., Todorčević, V., & Radenović, S. (2022). On the Distribution of Kurepa's Function. *Axioms*, 11(8):388. <https://doi.org/10.3390/axioms11080388>.
 33. Feller, W. (1948). On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *Annals of Mathematical Statistics*, 19, 177-189. <https://doi.org/10.1214/aoms/1177730243>.
 34. Fischetti, T. (2018). *R Analiza podataka (II izdanje)*. Beograd: Kompjuter biblioteka.
 35. Fishman, G.S., & Moore, L.R. (1986) An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal of Scientific and Statistical Computing*, 7, 24-45. <https://doi.org/10.1137/0907002>.
 36. Fishman, G.S. (1990). Multiplicative Congruential Random Number Generators with Modulus 2^β : An Exhaustive Analysis for $\beta = 32$ and Partial Analysis for $\beta = 48$. *Mathematics of Computation*, 54:189, 331-344. <https://doi.org/10.1090/S0025-5718-1990-0993929-9>.
 37. Fountain, J. (2002). An introduction to covers for semigroups. U: Gomes, G.M.S. (urednik). *Semigroups, Algorithms, Automata and Languages*. World Scientific. pp. 167–168. <http://www-users.york.ac.uk/~jbf1/coimbra2.pdf>.
 38. Fuller, A.T. (1976). The Period of Pseudo-Random Numbers Generated by Lehmer's Congruential Method. *The Computer Journal*, 19:2, 173–177. <https://doi.org/10.1093/comjnl/19.2.173>.
 39. Geary, R.C. (1936). The Distribution of "Student's" Ratio for Non-Normal Samples. *Supplement to the Journal of the Royal Statistical Society*, 3(2), 178–184. <https://doi.org/10.2307/2983669>.
 40. Gel, Y.R., Miao, W., & Gastwirth, J.L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, 51, 2734–2746. <https://doi.org/10.1016/j.csda.2006.08.022>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

41. Gentle, J.E. (2002). *Random Number Generation and Monte Carlo Methods*. Fairfax, Virginia: Geoge Mason University.
42. Gillariose, J., Balogun, O.S., Almetwally, E.M., Sherwani, R.A.K., Jamal, F., & Joseph, J. (2021). On the Discrete Weibull Marshall–Olkin Family of Distributions: Properties, Characterizations, and Applications. *Axioms*. 10(4):287. <https://doi.org/10.3390/axioms10040287>.
43. Goodman, S. (2008). A dirty dozen: twelve p-value misconceptions. In *Seminars in hematology* (Vol. 45, No. 3, pp. 135-140). WB Saunders. <https://doi.org/10.1053/j.seminhematol.2008.04.003>.
44. Goulard, F. (2022). Drawing random floating-point numbers from an interval. *ACM Transactions on Modeling and Computer Simulation*, 32 (3). <https://hal.science/hal-03282794v4>.
45. Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., & Altman, D.G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*. 31, 337–350. <https://doi.org/10.1007/s10654-016-0149-3>.
46. Gupta, M.M. (2021). Performance Analysis of ARL of Statistical Control Charts For Standard and Modified Ewma. *Matrix Science Mathematic*, 5(2): 42-45. <http://doi.org/10.26480/msmk.02.2021.42.45>.
47. Hassan, A., Shalhaf, G.A., Bilal, S., & Rashid, A. (2020). A New Flexible Discrete Distribution with Applications to Count Data. *J. Stat. Theory Appl.* 19, 102–108. <https://doi.org/10.2991/jsta.d.200224.006>.
48. Hogg, R.V., McKean, J.W., & Craig, A.T. (2019). *Introduction to Mathematical Statistics*. Boston: Pearson.
49. Hui, L., Weijing, W., Caixia, Z., Chunsheng, X., Haiping, D., Xiaocao, T., & Dongfeng, Z. (2018). Heritability and Genome-Wide Association Study of Plasma Cholesterol in Chinese Adult Twins. *Frontiers in Endocrinology*. 9. <https://doi.org/10.3389/fendo.2018.00677>.
50. Hu, X., Sun, G., Xie, F., & Tang, A. (2022). Monitoring the Ratio of Two Normal Variables Based on Triple Exponentially Weighted Moving Average Control Charts with Fixed and Variable Sampling Intervals. *Symmetry*. 14, 1236. <https://doi.org/10.3390/sym14061236>.
51. Ioannidis, J.P.A. (2019). What Have We (Not) Learnt from Millions of Scientific Papers with P Values?, *The American Statistician*, 73:sup1, 20-25. <https://doi.org/10.1080/00031305.2018.1447512>.
52. Ivanović, B. (2018). *Generisanje slučajnih veličina metodom stohastičke kolokacije – Master rad*. Beograd: Matematički fakultet Univerziteta u Beogradu.
53. Jacak, M.M. Józwiak, P., Niemczuk, J., & Jacak, J.E. (2021). Quantum generators of random numbers. *Scientific Reports*, 11:16108. <https://doi.org/10.1038/s41598-021-95388-7>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

54. Jenkins, K., & Gauvreau, K. (2006) Introduction to Clinical Research Design and Analysis. U: Brezinski, M.E. (Ured.) Optical Coherence Tomography. Academic Press. 369-392. <https://doi.org/10.1016/B978-012133570-0/50016-0>.
55. Jevremović, V. & Avdović, A. (2020a). Control Charts Based on Quantiles – New Approaches. Scientific Publications of the State University of Novi Pazar, Series A, Applied Mathematics, Informatics and Mechanics, 12, 2, 99-104. <https://doi.org/10.5937/SPSUNP2002099J>.
56. Jevremović, V. & Avdović, A. (2020b). Empirical Distribution Function as a Tool in Quality Control. Scientific Publications of the State University of Novi Pazar, Series A, Applied Mathematics, Informatics and Mechanics, 12, 1, 37-46. <https://doi.org/10.5937/SPSUNP2001037J>.
57. Jevremović, V. (2014). Verovatnoća i statistika. Beograd: Matematički fakultet.
58. Johnson, N.L., Kotz, S., & Balakrishnan, N. (1994). Continuous Univariate Distributions, Volume 1, 2nd Edition. John Wiley & Sons, Inc.
59. Jöhnk, M.D. (1964). Erzeugung von Betaverteilten und Gammaverteilten Zufallszahlen. *Metrika* 8, 5-15. <https://doi.org/10.1007/BF02613706>.
60. Kanji, G. (1993). 100 Statistical tests. London: Sage Publications Ltd.
61. Knuth, D.E. (1998). The art of computer programming - Volume 2. Boston: Addison-Wesley Professional.
62. Konold, T., & Fan, X. (2010). Hypothesis Testing and Confidence Intervals. In: Peterson, P., Baker, E., & McGaw, B. (Editors) International Encyclopedia of Education (Third Edition). Elsevier Ltd. pp. 216-222. <https://doi.org/10.1016/B978-0-08-044894-7.01337-3>.
63. Korkmaz, M.Ç., Chesneau, C., & Korkmaz, Z.S. (2021). On the Arcsecant Hyperbolic Normal Distribution. Properties, Quantile Regression Modeling and Applications. *Symmetry*. 13, 117. <https://doi.org/10.3390/sym13010117>.
64. Kotb, K.A.M., & El-Ashkar, H.A. (2020) Quality Control for Feedback M/M/1/N Queue with Balking and Retention of Reneged Customers. *Filomat*. 34, 167–174. <https://doi.org/10.2298/FIL2001167K>.
65. Lakens, D. (2021). The practical alternative to the p value is the correctly used p value. *Perspectives on psychological science*, 16(3), 639-648. <https://doi.org/10.1177/1745691620958012>.
66. Lehmer, D.H. (1949). Mathematical methods in large-scale computing unit. U: Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery. Cambridge, Massachusetts: Harvard University Press, pp. 141–146.
67. Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *J. Am. Stat. Assoc.* 62, 399–402. <https://doi.org/10.1080/01621459.1967.10482916>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

68. Lucas, J.M. (1973). A Modified “V” Mask Control Scheme. *Technometrics*, 15(4), 833–847. <https://doi.org/10.2307/1267393>.
69. Lukacs, E. (1942). A Characterization of the Normal Distribution. *The Annals of Mathematical Statistics*, 13(1), 91–93. <http://www.jstor.org/stable/2236166>.
70. Mališić, J. & Jevremović, V. (2008). *Slučajni procesi i vremenske serije*. Beograd: Matematički fakultet.
71. Marsaglia, G. (2004). Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software*. 9 (2): 730–737. <https://doi.org/10.18637/jss.v009.i02>.
72. Marsaglia, G. (2003). Random Number Generators. *Journal of Modern Applied Statistical Methods*, 2:1, 2–13. <https://doi.org/10.56801/10.56801/v2.i.57>.
73. Marsaglia, G. (1995), The Marsaglia Random Number CDROM, including the DIEHARD Battery of Tests of Randomness, Department of Statistics, Florida State University, Tallahassee, Florida. Available at <http://stat.fsu.edu/~geo/diehard.html>.
74. Marsaglia, G., & Tsang, W. W. (2000). The Ziggurat Method for Generating Random Variables. *Journal of Statistical Software*, 5(8), 1–7. <https://doi.org/10.18637/jss.v005.i08>.
75. Marsaglia, G. (2003). Xorshift RNGs. *Journal of Statistical Software*, 8(14), 1–6. <https://doi.org/10.18637/jss.v008.i14>.
76. Martin, D.W. (1985). *Doing Psychology Experiments*. 2nd ed. Monterey, CA: Brooks/Cole.
77. Massey, F.J. Jr. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* 46, 68–78. <http://doi.org/10.1080/01621459.1951.10500769>.
78. MATLAB Help Center. Creating and Controlling a Random Number Stream. Available online: <https://www.mathworks.com/help/matlab/math/creating-and-controlling-a-random-number-stream.html>.
79. Matsumoto, M., & Nishikura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*. 8, 1, 3-30. <https://doi.org/10.1145/272991.272995>.
80. McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, 73(sup1), 235-245. <https://doi.org/10.1080/00031305.2018.1527253>.
81. Merkle, M., & Vasić, P. (1995). *Verovatnoća i statistika – Sa primenama i primerima*. Beograd: Elektrotehnički fakultet.
82. Milošević, B. (2021). *Osnovi statistike*. Beograd: Matematički fakultet Univerziteta u Beogradu.
83. Minić, M. (2018). *Kontrola kvaliteta – kontrolne karte i njihove osobine*. U: Jevremović, V., & Mališić, J. (Urednici) *Odabrana poglavlja matematičke statistike*.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

- Novi Pazar: Državni Univerzitet u Novom Pazaru, Departman za matematičke nauke. 79-116.
84. Moler, C.B. (2004). *Numerical Computing with MATLAB*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).
 85. Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. 6th Edition. New York: John Wiley & Sons.
 86. Mulder, V., Mermoud, A., Lenders, V., & Tellenbach, B. (Urednici) (2023). *Trends in Data Protection and Encryption Technologies*. Cham, New York: Springer Nature Switzerland, Springer Science+Business Media.
 87. Muslim Raza, S. M., & Butt, M. M. (2018). New Shewhart and EWMA Type Control Charts using Exponential Type Estimator with Two Auxiliary Variables under Two Phase Sampling. *Pakistan Journal of Statistics and Operation Research*, 14(2), 367-386. <https://doi.org/10.18187/pjsor.v14i2.1262>.
 88. Nafidi, A., Bahij, M., Gutiérrez-Sánchez, R., & Achchab, B. (2020). Two-Parameter Stochastic Weibull Diffusion Model: Statistical Inference and Application to Real Modeling Example. *Mathematics*, 8, 160. <https://doi.org/10.3390/math8020160>.
 89. Neuburger, J., Walker, K., Sherlaw-Johnson, C., van der Meulen, & Cromwell, D.A. (2017). Comparison of control charts for monitoring clinical performance using binary data. *BMJ Quality & Safety*, 26, 919-928. <http://qualitysafety.bmj.com/>.
 90. Nikulin, M.S. (2020). Likelihood-ratio test. *Encyclopedia of Mathematics*. http://encyclopediaofmath.org/index.php?title=Likelihood-ratio_test&oldid=47635.
 91. Noughabi, H.A. (2018). A Comprehensive Study on Power of Tests for Normality. *Journal of Statistical Theory and Applications*, 17, 647-660, 2214-1766. <https://doi.org/10.2991/jsta.2018.17.4.7>.
 92. Oakland, J.S. (2003). *Statistical Process Control (Fifth edition)*. Oxford: Butterworth-Heinemann.
 93. Obradović, M., Jovanović, M., & Milošević, B. (2014). Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. *Statistics*, 49, 5-1026. <https://doi.org/10.1080/02331888.2014.919297>.
 94. Ottenstreuer, S., Weiß, C.H., & Testik, M.C. (2023). A review and comparison of control charts for ordinal samples. *Journal of Quality Technology*, 55:4, 422-44. <https://doi.org/10.1080/00224065.2023.2170839>.
 95. Owens, A.O., & Rioborue, A.B. (2022). Control Chart and Its Application in Modelling Body Mass Index (BMI) of Students in Delta State Polytechnic, Oghara. *American Journal of Theoretical and Applied Statistics*, 11, 19-26. <https://doi.org/10.11648/j.ajtas.20221101.13>.
 96. Öztürk, A. & Dudewicz, E.J. (1992) A New Statistical Goodness-of-Fit Test Based on Graphical Representation. *Biometrical Journal*, 34, 403-427. <https://doi.org/10.1002/bimj.4710340403>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

97. Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika*. 41, 1-2, 100-115. <https://doi.org/10.1093/biomet/41.1-2.100>.
98. Park, S.K., & Miller, K.W. (1988). Random number generators: Good ones are hard to Find. *Communications of the ACM*, 31:10, 1192–1201. <https://doi.org/10.1145/63039.63042>.
99. Panneton, F., & L'Ecuyer, P. (2005). On the xorshift random number generators. *ACM Transactions on Modeling and Computer Simulation*. 15, 4, 346-361. <https://doi.org/10.1145/1113316.1113319>.
100. Petrović, Lj. (2018). *Teorijska statistika – Teorija statističkog zaključivanja*. Beograd: Ekonomski fakultet.
101. Popović, B. & Blagojević, B. (2003). *Matematička statistika sa primenama u hidrotehnici*. Niš: Izdavačka jedinica Univerziteta u Nišu.
102. Popović, B. (2003). *Matematička statistika i statističko modelovanje*. Niš: Prirodno-matematički fakultet.
103. Praveen, V., Delhi Narendran, T., Pavithran, R., & Thirumalai, C. (2017). Data analysis using box plot and control chart for air quality. U: 2017 International Conference on Trends in Electronics and Informatics (ICEI). Tirunelveli: IEEE. <https://doi.org/10.1109/ICOEI.2017.8300877>.
104. Prokić, S. (2019). *Kontrolne karte i sposobnost procesa sa individualnim merenjima – Master rad*. Niš: Prirodno-matematički fakultet.
105. Rafi, Z., & Greenland, S. (2020). Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Medical Research Methodology* 20(1):244. <https://doi.org/10.1186/s12874-020-01105-9>.
106. Rahardja, D. (2014). Comparison of Individual and Moving Range Chart Combinations to Individual Charts in Terms of ARL after Designing for a Common “All OK” ARL. *Journal of Modern Applied Statistical Methods*. 13, 2(19), 364-378. 10.22237/jmasm/1414815480.
107. Rajasekaran, S., & Ross, K. W. (1993). Fast algorithms for generating discrete random variates with changing distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 3(1), 1-19. <https://doi.org/10.1145/151527.151529>.
108. Reinhart, A. (2015). *Statistics Done Wrong: The Woefully Complete Guide* (1st ed.). No Starch Press.
109. Reuven, Y.R. (1981). *Simulation and the Monte Carlo Method*. Technion, Haifa: John Wiley & Sons, Israel Institute of Technology.
110. Ritter, F.E.; Schoelles, M.J.; Quigley, K.S.; & Klein, L.C. (2011). Determining the Number of Simulation Runs: Treating Simulations as Theories by Not Sampling Their Behavior. *Human-in-the-Loop Simulations*. London: Springer.
111. Roberts, S.W. (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1, 239-250. <http://dx.doi.org/10.1080/00401706.1959.10489860>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

112. Siegel, S. (1956). *Nonparametric Statistics for the behavioral sciences*. New York, Toronto, London: McGraw-Hill Book Company, Inc.
113. Shapiro, S. S., & Francia, R. S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*. American Statistical Association. 67 (337): 215–216. <https://doi.org/10.1080/01621459.1972.10481232>.
114. Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>.
115. Slakter, M.J. (1965). A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Validity. *Journal of the American Statistical Association*, 60:311, 854-858, <https://doi.org/10.1080/01621459.1965.10480833>.
116. Stadlober, E. (1990). The ratio of uniforms approach for generating discrete random variates. *Journal of Computational and Applied Mathematics*, 31, 181–189. [https://doi.org/10.1016/0377-0427\(90\)90349-5](https://doi.org/10.1016/0377-0427(90)90349-5).
117. Stadlober, E. (1991). Binomial variate generation: A method based on ratio of uniforms. *The Frontiers of Statistical Computation, Simulation & Modeling* (edited by P. R. Nelson, E. J. Dudewicz, A. Oztürk, and E. C. van der Meulen), American Sciences Press, Columbus, Ohio, 93–112.
118. Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, 4(2): 357-369. <https://doi.org/10.1214/aos/1176343411>.
119. Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*. 69 (347): 730–737. <https://doi.org/10.2307/2286009>.
120. Sürücü, B. (2008) A power comparison and simulation study of goodness-of-fit tests. *Computers and Mathematics with applications*, 56, 1617–1625. <https://doi.org/10.1016/j.camwa.2008.03.010>.
121. Tang, H.C. (2005) Modulus of Linear Congruential Random Number Generator. *Quality and Quantity*, 39, 413-422. <https://doi.org/10.1007/s11135-004-6788-6>.
122. Thode, H. C. (2002). *Testing for Normality*. New York: Marcel Dekker.
123. Upton, G., & Cook, I. (2006). *Oxford Dictionary of Statistics*. Oxford: Oxford University Press.
124. Vaz, S., & Torres, D.F.M. (2021). A Discrete-Time Compartmental Epidemiological Model for COVID-19 with a Case Study for Portugal. *Axioms*. 10(4):314. <https://doi.org/10.3390/axioms10040314>.
125. Veljkovic, K., Elfaghihe, H., & Jevremovic, V. (2015). Economic Statistical Design of X Bar Control Chart for Non-Normal Symmetric Distribution of Quality Characteristic. *Filomat*. 29, 2325–2338. <https://doi.org/10.2298/FIL1510325V>.
126. Vidgen, B., & Yasseri, T. (2016). P-Values: Misunderstood and Misused. *Frontiers in Physics*, 4:6. <http://dx.doi.org/10.3389/fphy.2016.00006>.

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

127. Von Neumann, J. (1951). Various Techniques Used in Connection with Random Digits. Washington, DC: National Bureau of Standards, Applied Mathematics Series, 12, 36-38.
128. Vrbik, J. (2020). Deriving CDF of Kolmogorov-Smirnov Test Statistic. Applied Mathematics, 11, 227-246. <https://doi.org/10.4236/am.2020.113018>.
129. Vuković, N. (2007). Statističko zaključivanje. Beograd: Fakultet organizacionih nauka.
130. Wang, B., Zhou, Z., Wang, H., Tu, X. M., & Feng, C. (2019). The p-value and model specification in statistics. General psychiatry, 32(3), e100081. <https://doi.org/10.1136/gpsych-2019-100081>.
131. Wasserman, L. (2004). All of Statistics - A Concise Course in Statistical Inference. New York: Springer.
132. Wu, P.C. (1997) Multiplicative, Congruential Random-Number Generators with Multiplier $\pm 2^{k_1} \pm 2^{k_2}$ and Modulus $2^p - 1$. ACM Transactions on Mathematical Software, 23:2, 255–265. <https://doi.org/10.1145/264029.264056>.
133. Yang, S., & Berdine, G. (2021). Normality tests. The Southwest Respiratory and Critical Care Chronicles. 9(37): 87–90. <https://doi.org/10.12746/swrccc.v9i37.805>.

Prilozi

U nastavku su dati prilozi u vidu MATLAB kodova korišćenih u istraživanju. Dati su kodovi za

- generisanje i testiranje slučajnosti brojeva generisanih LKGom sa aditivnim proširenjem,
- generisanje i testiranje slučajnosti brojeva generisanih generatorom MT19937,
- modeliranje normalne raspodele algoritmom zona,
- izračunavanje ARLa za EFR kartu,
- izračunavanje ARLa kontrolne karte kvantil-zona,
- simulacije raspodele statistike A za nultu i alternativne raspodele polaznog obeležja i
- simulacije raspodele statistike V za nultu i alternativne raspodele polaznog obeležja.

Prilog 1. MATLAB kod za generisanje i testiranje slučajnosti brojeva generisanih LKGom sa aditivnim proširenjem.

```
format short
n=100000;
a=50;
c=11;
m=343;
X=[];
for j0=1:m
X(1,1)=j0-1;
U=[];
U(1,1)=X(1,1)/m;
for i=2:n
    h=0;
    for k=1:(i-1)
        if X(k,1)==X(i-1,1)
            h=h+1;
        end
    X(i,1)=mod(h*a*X(i-1,1)+c,m);
    U(i,1)=X(i,1)/m;
end
end
r=0;
for j=1:n-1;
    if U(j,1)<U(j+1,1)
        r=r+1;
    end
end
R=(r-(n-1)/2)/sqrt((n+1)/12);
pr(j0)=normcdf(-abs(R),0,1)+1-normcdf(abs(R),0,1);
z=0;
for j1=1:n-2;
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

```
    if
or(and(U(j1,1)<U(j1+1,1),U(j1+1,1)>U(j1+2,1)),and(U(j1,1)>U(j1+1,1),U(j1+1,1)<U(j1
+2,1)))
    z=z+1;
    end
end
Z=(z-(2/3)*(n-2))/sqrt((16*n-29)/90);
pz(j0)=normcdf(-abs(Z),0,1)+1-normcdf(abs(Z),0,1);
end
prk=mean(pr)
pzk=mean(pz)
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 2. MATLAB kod za generisanje i testiranje slučajnosti brojeva generisanih generatorom MT19937.

```
format short
n=100000;
pr=[];
pz=[];
o=1;
while o<10000
U=rand(n,1);
r=0;
for j=1:n-1;
    if U(j,1)<U(j+1,1)
        r=r+1;
    end
end
R=(r-(n-1)/2)/sqrt((n+1)/12);
pr(o)=normcdf(-abs(R),0,1)+1-normcdf(abs(R),0,1);
z=0;
for j1=1:n-2;
    if
or(and(U(j1,1)<U(j1+1,1),U(j1+1,1)>U(j1+2,1)),and(U(j1,1)>U(j1+1,1),U(j1+1,1)<U(j1
+2,1)))
        z=z+1;
    end
end
Z=(z-(2/3)*(n-2))/sqrt((16*n-29)/90);
pz(o)=normcdf(-abs(Z),0,1)+1-normcdf(abs(Z),0,1);
o=o+1;
end
prk=mean(pr)
pzk=mean(pz)
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 3. MATLAB kod za modeliranje normalne raspodele algoritmom zona.

```
format short
s=1000;
n=1000;
for j=1:s
for i=1:n
u=rand;
v=2*rand-1;
if and(0<=abs(v),abs(v)<=0.6826)
    x(i,j)=(v/abs(v))*(1-u);
elseif and(0.6826<abs(v),abs(v)<=0.9544)
    x(i,j)=(v/abs(v))*(2-u);
elseif and(0.9544<abs(v),abs(v)<=0.9973)
    x(i,j)=(v/abs(v))*(3-u);
elseif abs(v)>0.9973
    x(i,j)=(v/abs(v))*(3+u);
end
end
end

for j1=1:s
    [H,p]=adtest(x(:,j1));
    mp(j1,1)=p;
end
mean(mp)
std(mp)
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 4. MATLAB kod za izračunavanje ARLa za EFR kartu.

```
format long
n=10000;
m=5;
d=0.2130;
t=[];
o=1;
while o<=10000
A=normrnd(4,1,n,m);
a=[];
for i=1:n
    a(i,1)=mean(A(i,:));
end
for j=1:n
    emp=0;
    for k=1:n
        if a(k,1)<=a(j,1)
            emp=emp+1;
        end
    end
    b(j,1)=emp/n;
end
for s=1:n
    if
or(b(s,1)>min(1,normcdf(a(s,1),0,1/sqrt(m))+d),b(s,1)<max(0,normcdf(a(s,1),0,1/sqrt(m))-d))
        t(o,1)=s;
        break
    end
end
o=o+1;
end
mean(t)
length(t)
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 5. MATLAB kod za izračunavanje ARLa kontrolne karte kvantil-zona.

```
format long
n=1000;
m=5;
t=[];
o=1;
while o<=1000
A=normrnd(0,1,n,m);
a=[];
for i=1:n
    a(i,1)=mean(A(i,:));
end
a1=sort(a);
for j=1:n
    emp=0;
    for k=1:n
        if a(k,1)<=a(j,1)
            emp=emp+1;
        end
    end
    b(j,1)=emp/n;
end
c=sort(b);
for s=1:n
    if and(or(c(s,1)>normcdf(a1(s,1), -
3.5/sqrt(m),1/sqrt(m)),c(s,1)<normcdf(a1(s,1),3.5/sqrt(m),1/sqrt(m))),c(s,1)<1)
        t(o,1)=s;
        break
    end
end
o=o+1;
end
mean(t)
length(t)
```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 6. MATLAB kod za simulacije raspodele statistike A za nultu i alternativne raspodele polaznog obeležja.

```

format short
s=10000;
n=[5,10,20,30,50,100,200,300,500,1000,1500,2000];
v=[];
empv=[];
a=[];
b=[];
T=[];
for m=1:12
    A=normrnd(0,1,s,n(m)); % U slučaju druge raspodele programiran je odgovarajući
    algoritam dat u poglavlju 4.
    C=[];
    for k=1:s
        xn=mean(A(m,:)); % Za poznate vrednosti parametara izjednačiti ih sa xn i sn.
        sn=std(A(m,:));
        for h=1:n(m)
            if and(xn-1*sn<=A(k,h),A(k,h)<=xn+1*sn)
                C(k,h)=1;
            elseif or(and(xn-1.96*sn<=A(k,h),A(k,h)<xn-1*sn),
and(xn+1*sn<A(k,h),A(k,h)<=xn+1.96*sn))
                C(k,h)=1.96;
            elseif or(and(xn-2.58*sn<=A(k,h),A(k,h)<xn-1.96*sn),
and(xn+1.96*sn<A(k,h),A(k,h)<=xn+2.58*sn))
                C(k,h)=2.58;
            else
                C(k,h)=2.81;
            end
        end
    end
    for d=1:s
        v(d,m)=mean(C(d,:));
    end
    a(:,m)=sort(v(:,m));
    for k4=1:s
        emp=0;
        for h4=1:s
            if v(h4,m)<v(k4,m) % Za dobijanje aproksimacije zakona raspodele staviti
            "==" umesto "<=".
                emp=emp+1;
            end
        end
        empv(k4,m)=emp/s;
    end
    b(:,m)=sort(empv(:,m));
    T(:,2*m-1)=a(:,m);
    T(:,2*m)=b(:,m);
end
open T

```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

Prilog 7. MATLAB kod za simulacije raspodele statistike V za nultu i alternativne raspodele polaznog obeležja.

```

format short
s=100000;
n=[5,10,20,30,50,100,200,300,500,1000,1500,2000];
v=[];
empv=[];
a=[];
b=[];
T=[];
for m=1:12
    A=normrnd(0,1,s,n(m)); % U slučaju druge raspodele programiran je odgovarajući
    algoritam dat u poglavlju 4.
    for k1=1:s
        for h1=1:n(m)
            emp1=0;
            for t1=1:n(m)
                if A(k1,t1)<=A(k1,h1)
                    emp1=emp1+1;
                end
            end
            B(k1,h1)=emp1/n(m);
        end
    end
    C=[];
    for k2=1:s
        xn=mean(A(m,:)); % Za poznate vrednosti parametara izjednačiti ih sa xn i sn.
        sn=std(A(m,:));
        for h2=1:n(m)
            if and(normcdf(A(k2,h2),xn+sn,sn)<=B(k2,h2),B(k2,h2)<=normcdf(A(k2,h2),xn-
            sn,sn))
                C(k2,h2)=1;
            elseif
            or(and(normcdf(A(k2,h2),xn+1.96*sn,sn)<=B(k2,h2),B(k2,h2)<normcdf(A(k2,h2),xn+sn,s
            n)),and(normcdf(A(k2,h2),xn-sn,sn)<B(k2,h2),B(k2,h2)<=normcdf(A(k2,h2),xn-
            1.96*sn,sn)))
                C(k2,h2)=1.96;
            elseif
            or(and(normcdf(A(k2,h2),xn+2.58*sn,sn)<=B(k2,h2),B(k2,h2)<normcdf(A(k2,h2),xn+1.96
            *sn,sn)),and(normcdf(A(k2,h2),xn-
            1.96*sn,sn)<B(k2,h2),B(k2,h2)<=normcdf(A(k2,h2),xn-2.58*sn,sn)))
                C(k2,h2)=2.58;
            else
                C(k2,h2)=2.81;
            end
        end
    end
    v(k3,m)=mean(C(k3,:));
end
a(:,m)=sort(v(:,m));
for k4=1:s

```

FUNKCIJE ZONE I KVANTIL-ZONE I PRIMENE U KONTROLI KVALITETA I
TESTIRANJU NORMALNOSTI – DOKTORSKA DISERTACIJA

```
emp4=0;
for h4=1:s
    if v(h4,m)<=v(k4,m) % Za dobijanje aproksimacije zakona raspodele staviti
"==" umesto "<=".
        emp4=emp4+1;
    end
end
empv(k4,m)=emp4/s;
end
b(:,m)=sort(empv(:,m));
T(:,2*m-1)=a(:,m);
T(:,2*m)=b(:,m);
end
open T
```