

STVARNI TEMELJI NORMALNE DISTRIBUCIJE

Dobromir Bonacin

Kineziološki fakultet Univerziteta u Travniku, BiH

Izvorni znanstveni rad

Sažetak

U članku je pokazano da aproksimacija normalne (Gaussove) distribucije Laplaceovom funkcijom nije ni blizu jedini način opisa prirodnog grupiranja pojava oko srednje vrijednosti, a da svakako nije ni najprecizniji. Razvojem Pascalovog trokuta moguće je dobiti znatno kvalitetniji i precizniji opis normalne distribucije koji ne pokazuje grešku u okolini srednje vrijednosti. Na jednostavnom numeričkom primjeru sa samo 101 retkom trokuta savršeno je pojašnjeno da normalna distribucija ima svoja ista pravila i u situacijama s najmanjim brojem događaja u polju događaja, te da će ta pravila važiti na svakoj, pa i na infinitezimalnoj razini. Pretpostavljeno je kako je time moguće rekonstruirati samu srž nastanka bilo kojih pojava u svijetu koje se ponašaju na opisani način. Tako, pokazatelje u ovom radu možemo identificirati kao nezamjenjive za razumijevanje stvarnih temelja normalne distribucije, a sve zajedno predstavlja svojevrsnu Teoriju interakcije elementarnih pojava, s dosegom u najsloženije oblike koji se uopće mogu pojaviti, a da realno predstavljaju pojave kojih se "ponašanje" može podvesti pod normalnu distribuciju.

Ključne riječi: *distribucija, normalitet, Pascalov trokut, Teorija interakcije*

Uvod

Statistika kao disciplina artikulira se kao dio matematike i služi se različitim oblicima simboličkih alata za utvrđivanje ili opisivanje pojava iz realnog svijeta. Brojni su pristupi ovoj logici ponuđeni kroz povijest kao i brojne definicije, i sve one odražavaju aktualno stanje spoznaja u trenutku definicije. Tako npr. Serdar i Šošić (1981) navode da se radi o opisivanju, upoznavanju, istraživanju, uspoređivanju i analiziranju pojava. Mijanović (2005) navodi da nas zanima zakonomjernost u pojavama kojima smo okruženi.

Jakob (1970) napominje da se u proučavanju odigrava dekriptiranje prirode, a statistika sistematski zanemaruje pojedinosti dok se ne dođe do određenog pravila. Ovo je moguće jednostavno zato jer je očito da su sve pojave u Univerzumu međusobno povezane (Bonacin, 2005), iako je ovo mnogo očitije u društvenim nego prirodnim znanostima (Mijanović, 2005). Međutim, varijacije koje primjećujemo nipošto ne znače da do zakonitosti ne možemo doći, jer raznolikosti oblika nisu raspoređene slučajno. Svaki element neke cjeline se nadovezuje na druge kako bi se osigurala harmonija cjeline. Svaka modifikacija strukture vrši utjecaj na druge elemente.

One pojave koje ne mogu postojati zajedno – isključuju se, a sve se varijacije usaglašuju i integriraju (Jacob, 1970). Ovo ne znači ništa drugo nego da u pojavama koje proučavamo nema slučajnosti. To što ponekad nismo u stanju egzaktno utvrditi stanje neke pojave, nikako ne znači da neko svojstvo neke pojave nema status, kao što to navodi kritika Heisenbergove kopenhaške interpretacije (Strnad, 1985; Bonacin, 2004 a). U takvom kontekstu postaje jasno zašto će empirijska tehnologija uvijek zaostajati za dijalektičkim načelima indukcije i dedukcije, jer se radi o prirodnoj tendenciji uma da slobodno kreira, a da se ne govori o tehnologiji metričke definicije pojava što je isključivo problem tehnologije ili, ono najvažnije, kako bi se moglo empirijski djelovati treba postojati dijalektički okvir unutar kojega se eksperiment izvršava (Piaget, 1972). Zanemarivanje dijalektičkih načela i prenaplašene pozitivističke tendencije u znanosti često udalje istraživače od problema te se u prevelikoj mjeri započinju baviti alatima umjesto pojavama, te se zaboravlja da suštinski uopće ne postoje aritmetičke sredine, medijani, standardne devijacije, korelacije, faktori, integrali, Heisenbergova neodređenost, Planckova konstanta, Paulijev princip i slično.

To su samo simboličke konstrukcije koje nam pomažu u razumijevanju pojava ili dijela pojava kojima smo okruženi (Bonacin, 2004 b). Izrada alata (pa i simboličkih) direktno je vezana uz metodologiju, no to nimalo ne umanjuje značenje indukcije kao epistemološke kategorije već je još više naglašava. Statistika je način sagledavanja pojava na neki znanstveno prihvatljiv i reproducibilan način, ali nakon utvrđivanja zakonitosti u istraživanim pojavama, cjelokupni proces nipošto nije završio, jer se samo radi o utvrđivanju i prepoznavanju prirodnih zakonitosti, pa indukciju potpuno vjerodostojno možemo označiti kao istraživanje i prepoznavanje.

U tom dijelu statistika je nezamjenjiva kao znanstvena disciplina. Moglo bi se reći da se bavi nepoznatim pojavama, koje pokušava objasniti, što direktno znači da se, baveći se statistikom na svim razinama, bavimo znanošću, tj. ugradnjom poznatoga u nepoznate pojave i time osiguravamo veći stupanj poznatosti zakonitosti ili dijela zakonitosti. Međutim, nakon prepoznavanja pravila (zakonitosti), slijedi drugi dio ukupnosti našeg djelovanja, a to je konstrukcija novih ili transformacija postojećih entiteta (pojava) u čemu se integriraju naša prethodno stečena znanja. U ovom dijelu, esencijalne znanosti u suštini više nema, već se radi o tehnologiji, odnosno našoj sposobnosti da prepoznata pravila upotrijebimo za rješavanje praktičnih zadataka (Bonacin, 2004 b; Bonacin, 2005).

Iz svega toga je, naravno, jasno, da u konstrukciji neće biti moguće direktno ugraditi inducirane pojmove i zakone, jer ako su zaista zakoni, oni su simbolički definirani, pa se stvaraju nove tehnološke kreacije koje realno egzistiraju u stvarnom svijetu, što dovodi do određenih razlika u realizaciji koje nastaju samom činjenicom da novostvorene konstrukcije nisu esencijalni zakoni, već njihovi svojevrsni emanati (Klaić, 1962), iz čega nastaje poznata divergencija: teorija – praksa. Iz navedenog je potpuno jasan izniman značaj spoznavanja, indukcije ili statističkog zaključivanja, jer će pogrešno utvrđena pravila imati direktne negativne reperkusije na konstrukciju, što znači da će utjelovljenje dedukcijskih ideja u konstrukcijska djelovanja biti opterećeno nepoznatim dijelom koji induksijski nije identificiran. Zbog realnosti stvarnog svijeta, konstrukcija uvijek sadrži dio kontaminacije koja nije imanentna induciranoj zakonitosti, ali ako je uz to još i zakonitost nedovoljno točno utvrđena posljedice na kvalitetu konstrukcije su sigurno predvidljive.

Cilj

Problem ovog rada, naizgled, najlakše bi bilo uvesti u uže metodološke okvire, jer bi tada bilo dosta jednostavno prikazati eventualna lokalna poboljšanja u izgradnji ograničenih alata ili prepoznavanju pojava u okvirima statističke teorije, kombinatorike i teorije vjerojatnosti.

Međutim, tada bi izostala transparentija na epistemološka i gnoseološka područja filozofije i znanosti općenito, čiji je značaj u uvodu jasno naglašen. Iz strogo praktičnih razloga, prečesto se zadovoljavamo "dovoljnom preciznošću" i "prihvatljivom razinom istine", generirajući simboličke (matematičke) konstrukcije "dostatne točnosti". Međutim, zaboravlja se da se na tako utvrđenim "istinama" i "modelima" gradi dalje, pri čemu je u sve daljnje razine konstrukcije ugrađen upravo onoliko istinit rezultat koliko je istinita početna inducirana zakonitost, ali jednako i početna inducirana greška. Problem ovog rada je, dakle, definicija potpune točnosti određene statističke simbolike "masovnih pojava". S tim u vezi, predmet ovog rada definiran je na gotovo najnižoj mogućoj razini, a to je generiranje zakonitosti elementarnih pojava, npr. interakcije među pojavama poput Leibnizovih monada (Grlić, 1967), uz tu razliku što se njegove monade podvrgavaju vlastitim zakonima, bez interakcija s drugima, a što je, kako smo vidjeli (Bonacin, 2004 a) nemoguće, jer bi tada bilo Svemira koliko i monada, bez ikakvih složenijih oblika. Na razni pojedinačnih pojava i njihovih interakcija, u simboličkoj statistici, podaci se svrstavaju, grupiraju i sl., čime se formira tzv. distribucija podataka.

Još Gauss je početkom 19. stoljeća pokazao da se masovne pojave u realnom svijetu podvrgavaju određenim pravilima, iz čega je izveo poznata pravila o familiji kontinuiranih distribucija vjerojatnosti (Dixon & sur., 1969; Bailey, 1971; Cruz, 2002). Međutim, iako je ova distribucija davno simbolički definirana, do danas nije pojašnjen njen temeljni uzročni karakter, odnosno na čemu je suštinski utemeljena, i to je cilj ovog rada. Ovo iz razloga što se matematička definicija uzima kao aksiom, opet uz zanemarivanje da se kod simboličke definicije radi samo o aproksimaciji, a ne i stvarnim pojavama.

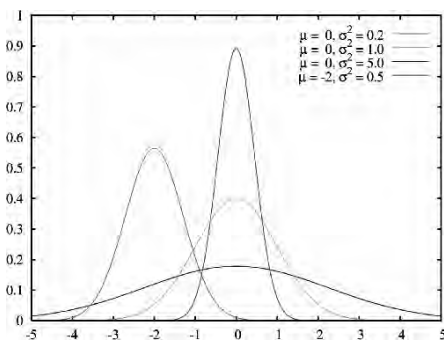
Uopće nema sumnje da je "normalna" distribucija prihvatljiva u životu, ali je isto tako potpuno sigurno da nikad nije objašnjeno što stoji u njenim samim temeljima. Ovaj rad će pokazati upravo to.

Model

U velikom skupu pojedinačnih podataka koji opisuju neku pojavu moguće je prepoznati polje događaja kao skup svih događaja koji odgovaraju jednom pokusu (Ivković, 1980). Događaji koji ulaze u takvo polje moraju zadovoljiti određene uvjete, kako bi bili izvjesni, za razliku od nemogućih koji se ne nalaze u tom polju događaja (Pauše, 1978). Iako se u pokusu sa slučajnim ishodom ne može govoriti o realizaciji ili nerealizaciji slučajnog događaja, pri većem broju ponovljenih pokusa zapaža se određena pravilnost koja leži u temelju Teorije vjerojatnosti i Bernoulijevog zakona velikih brojeva, te garantira njenu suglasnost sa zakonima objektivne stvarnosti (Pauše, 1978; Ivković, 1980; Serdar & sur., 1981; Mijanović, 2005). Primjećeno je da se podaci, naročito masovnih pojava grupiraju oko središnje vrijednosti, s opadanjem učestalosti prema zonama ekstremno velikih i ekstremno malih vrijednosti. Ova pojava opisana je kao normalna (Gaussova) distribucija uz parametre (μ , σ) koji su dovoljni za njen potpuni simbolički opis uz primjenu Laplaceove funkcije

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

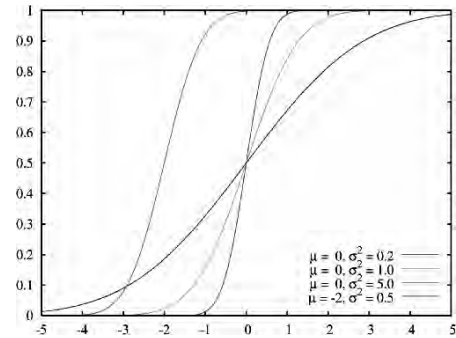
za pozitivne vrijednosti argumenta x, jer se integral normalne aproksimacije, kako se smatra, ne može izračunati jednostavnim postupkom.



Grafikon 1. Funkcije gustoće

Tako se dobiju funkcije gustoće (grafikon 1) i distribucije (grafikon 2) koje opisuju skup podataka na temelju očekivanih vrijednosti za parametre (μ , σ) (aritmetička sredina i standardna devijacija).

Ovo su sasvim sigurno vrlo kvalitetno utemeljene aproksimacije, ali je još uvijek nejasno kako nastaju stvarni pokazatelji.



Grafikon 2. Funkcije distribucije

Rezultati i rasprava

Poslužimo li se jednom drugom varijantom kombinatorike, možemo doći do iznenađujućih rezultata. Naime, većina istraživača Teorije kombinatorike i srodnih područja, poznaje tzv. Pascalov trokut koji je zadan kao niz brojeva, pri čemu se u n-tom retku Pascalovog trokuta nalaze binomni koeficijenti n-tog reda n=0, 1, 2, 3, ..., i to poredani po razredu k=0, 1, 2, ..., n. Može se vidjeti da je svaki element, osim rubnih, zbroj dvaju elemenata koji se nalaze s lijeve i desne strane u retku iznad (grafikon 3.). Teorem kojim je zadan Pascalov trokut također je dobro poznat matematičarima i teoretičarima statistike i kombinatorike.

Teorem glasi:

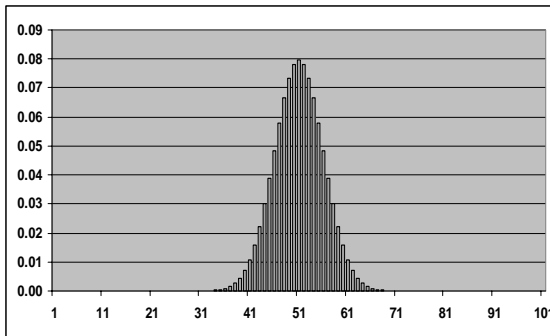
$$\binom{n}{k} = \binom{n}{n-k}, \quad \forall k, n \in \mathbb{N} \cup \{0\}, \quad k \leq n,$$

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}, \quad \forall k, n \in \mathbb{N} \cup \{0\}, \quad k < n.$$

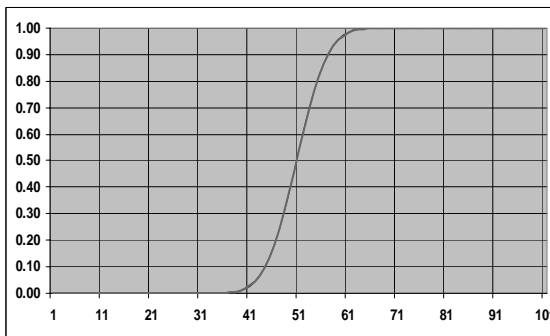
Druga tvrdnja ovog teorema daje Pascalov trokut na slijedeći način:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
:
:
:

U tablici 1 nalazi se primjer podataka za razvijeni red od 101 retka Pascalovog trokuta koji pokazuju da se ovaj red ponaša gotovo potpuno sukladno integralu normalne distribucije i Laplaceovoj funkciji, što znači da je moguće direktno dobiti vjerojatnosti bilo kojeg retka za koji nije potrebno računati aproksimativne vrijednosti Laplaceove funkcije. To također znači da je izbjegnuta aproksimacija i podaci se mogu dobiti direktno. Već na ovom primjeru sa svega 101 retkom dobivena je suladnost podataka, a naravno za veći broj redaka (1000 ili 10000), još bi se preciznije izračunali svi potrebni elementi, do točnosti koja višestruko prelazi točnost aproksimacije.

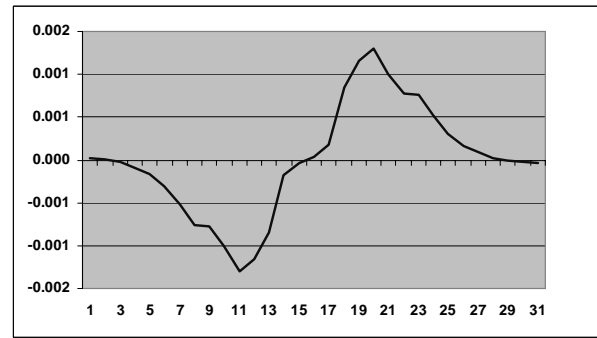


Grafikon 3. Gustoća iz Pascalovog trokuta



Grafikon 4. Distribucija iz Pascalovog trokuta

Iako se podaci gotovo potpuno slažu, ipak podaci dobiveni Laplaceovom funkcijom pokazuju mala, ali primjetna odstupanja od podataka dobivenih iz Pascalovog trokuta. Te razlike su najviše izražene u okolini srednje vrijednosti skupa podataka. Kako se vidi na grafikonu 5 (i u tablici 1) uobičajeni model normalne (Gaussove) distribucije pokazuje grešku koja se jasno vidi. Ta greška je utoliko manja ukoliko se više približavamo ekstremnim vrijednostima. Rekapitulirajmo ono što je posebno značajno u ovom novom modelu, a to su dvije činjenice: **A.** "idealna" normalna distribucija zadana je određenim brojem podataka u nekoj kategoriji, neovisno o tome radi li se o malom ili velikom broju podataka pa tako npr. slijed frekvencija (1, 4, 6, 4, 1) opisuje "idealno" normalno ponašanje.



Grafikon 5. Greška Laplaceove funkcije

Isto kao i slijed (1, 6, 15, 20, 15, 6, 1) ili slijed (1, 14, 91, 364, 1001, 2002, 3003, 3432, 3003, 2002, 1001, 364, 91, 14, 1). Ove frekvencije opisuju segmente diskretnih odsječaka, koje će važiti na svakoj, pa i na infinitezimalnoj razini. **B.** Na temelju dobivenih pokazatelja, činjenica je da je moguće sasvim realno pretpostaviti kako interakcijom početnih elemenata (ma kakvi bili) prema prikazanim načelima, možemo identificirati sve složenije oblike i pojave koje se "ponašaju" po tzv. normalnoj distribuciji, ili drugim riječima, možemo rekonstruirati samu srž nastanka bilo kojih pojava u Svijetu. Tako, ove pokazatelje možemo identificirati kao nezamjenjive za razumijevanje stvarnih temelja normalne distribucije ili usložnjavanja bilo kojih pojava. Sve zajedno predstavlja svojevrsnu **Teoriju interakcije elementarnih pojava**, sa dosegom u najsloženije oblike koji se uopće mogu pojaviti, a da realno predstavljaju pojave kojih se "ponašanje" može podvesti pod normalnu distribuciju. Kako znamo, pojava koje se ponašaju drugačije, uopće nema, a i one za koje se obično drži da su drugačije, najčešće nisu prepoznate kao takve, jer je tehnologija njihove analize najčešće opterećena našim neznanjem ili lošim tehničkim rješenjima. Naime, tko to uopće promišlja da je slijed frekvencija (1, 2, 1) upravo kompletni izraz normalne distribucije u slučaju kad ima svega 4 događaja koje se promatra, ili ako imamo npr. 16384 događaja, tada u polju događaja, frekvencije 1, 14, 91, ..., 91, 14, 1, opisuju upravo idealnu normalnu distribuciju. Broj mogućih događaja u polju događaja za idealne situacije je potencija broja 2. Do sada su to bili postulati koji su uvelike koketirali s područjem metafizike, a od sad ćemo uvijek znati da to nije ništa drugo nego na neki način izveden razvoj u red. Interakcije između srodnih pojava, a u skladu s pravilima Univerzuma, čine da se pojave očituju na način koji opisujemo normalnom distribucijom, jer da nemaju baš nikakve međusobne veze, tada bi sve pojave Univerzuma bile uniformno prepoznate, što znači da ne bi bilo ničega, pa niti ovog teksta i vas koji to čitate.

Literatura

- Bailey, D.E. (1971). *Probability and statistics: Models for research*. New York: John Wiley & sons.
- Bonacin, D. (2004 a). *Identifikacija restrukturiranja taksona biomotoričkih dimenzija djece uzrasta 7 godina pod utjecajem transformacijskih procesa*. /Doktorska disertacija/, Sarajevo: Fakultet sporta i tjelesnog odgoja.
- Bonacin, D. (2004 b). *Uvod u kvantitativne metode*. Kaštela: Vlastito izdanje.
- Bonacin, D. (2005). Comprehensive continuum. *Homo Sporticus*, 8(2), 16-20.
- Cruz, M. (2002). *Modeling, measuring and hedging operational risk*. New York: John Wiley & sons.
- Dixon, J.W., & Massey, F.J. (1969). *Introduction to statistical analysis*. New York: McGraw Hill HE - III.
- Grlić, D. (1967). *Filozofija (Školski leksikon)*. Zagreb: Panorama.
- Ivković, Z.A. (1980). *Matematička statistika*. Beograd: Naučna knjiga.
- Jacob, F. (1970). *La logique du vivant*. Paris: Gallimard.
- Klaić, B. (1962). *Rječnik stranih riječi, izraza i kratica*. Zagreb: Zora.
- Mijanović, M. (2005). *Statističke metode*. Podgorica: Univerzitet Crne Gore.
- Pauše, Ž. (1978). *Vjerojatnost: Informacija - stohastički procesi*. Zagreb: Školska knjiga.
- Piaget, J. (1972). *Epistemologie des sciences de l'homme*. Paris: Gallimard.
- Serdar, V, & Šošić, I. (1981). *Uvod u statistiku*. Zagreb: Školska knjiga.
- Strnad, J. (1985). *Mala kvantna fizika*. Zagreb: Školska knjiga.

Priljeno: 24.06.2007.
Prihvaćeno: 15.12.2007.

Korespondencija:
dr.Dobromir Bonacin
Kineziološki fakultet
Univerzitet u Travniku, BiH
72270 Travnik, Bosna i Hercegovina
Kalibunar bb
E-mail: dobromir.bonacin@st.t-com.hr

REAL BASICS OF NORMAL DISTRIBUTION

Abstract

It has been shown in this article that the approximation of normal (Gausse's) distribution by Laplace's function is not at all the only way to describe natural grouping of occurrences around the average value and of course that it is not the most precise one. By developing the Pascal's triangle, it is possible to get much better quality and more precise description of normal distribution showing no error around the average value. In a simple numerical sample with only 101 lines of the triangle, it has been perfectly explained that the normal distribution has its own rules even in the situations with the smallest number of events in the scope of events and that those rules will be valid at each and every level and even at an infinite number level. We suppose that while doing so, it is possible to reconstruct the very core of occurrence of any kind of events in the world behaving according to the described way. So, the indicators in this work can be identified as irreplaceable for understanding the practical basis of normal distribution, and this all together makes a specific Theory of interaction of elementary occurrences, reaching the most complex forms that can possibly appear but representing, in reality, the events whose "behaviour" can be described as normal distribution.

Key words: *distribution, normality, Pascal's triangle, Theory of interaction*