**MA273**

1. **Formula totalne verovatnoće. Bajesova formula**

**Formula totalne verovatnoće** se bazira na uslovnoj verovatnoći i dokazuje primenom pravila množenja.

Stav.Ako su A1, A2,…, Anuzajamno disjunktni događaji i važi P(Ai) > 0, za i = 1, 2,…, n i ako je ∑n i=1 Ai = Ω,tada važi:

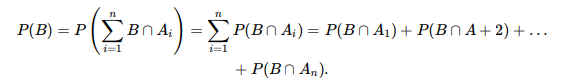


za svako B ∈ F. Prethodna formula se naziva Formula totalne verovatnoće.

Dokaz.Kako je događaj B ⊆ Ω i ∑n i=1 Ai = Ω, imamo da je



Odatle imamo da je



Dalje, iz pravila množenja, imamo da je



Konačno, imamo da je



**Važi da je:**



**Bajesova formula** se još naziva formula verovatnoće hipoteza, odnosno formula aposteriorne verovatnoće i ona se može interpretirati nasledeći način: događaj B može da se realizuje pod različitim uzrocima ili hipotezama A1, A2,…, An. Događaj B se realizovao. Kolika je verovatnoća da je tome uzorak baš Ai , za i = 1, 2,…, n.

1. **Bernulijeva šema**

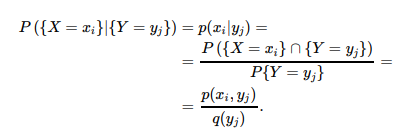
Bernulijeva raspodela slučajne promenljive se još naziva i "zakon 0-1" jer je dihotomog tipa po pitanju ishoda. Dakle, pri izvođenju nekog eksperimenta se neki događaj nije javio ili se javio. Neka slučajna promenljiva X predstavlja realizaciju događaja A u posmatranom eksperimentu. Tada, kažemo da slučajna promenljiva X ima Bernulijevu raspodelu sa parametrom p, 0 < p < 1, i to zapisujemo na sledeći način



Bernulijevi eksperimenti ili Bernulijeva šema predstavlja niz eksperimenata koji se nezavisno jedan od drugog izvode pod istim (neizmenjenim) uslovima, pri čemu svaki od tih eksperimenata predstavlja model slučajne promenljive sa Bernulijevom raspodelom. Bilo koji slučajni događaj, koji je u vezi sa samo jednim određenim eksperimentom (iz niza ponovljenih ekperimenata) se može predstaviti kao slučajan događaj složenog eksperimenta

1. **Dvodimenzionalne slučajne promenljive diskretnog tipa. Marginalne raspodele i nezavisnost.**

Može nas interesovati raspodela jedne slučajne promenljive, recimo X, posedujući informaciju ili pretpostavljajući da je druga slučajna promenljiva “uzela” neku vrednost Y = y. U slučaju kada je dvodimenzionalna slučajna promenljiva (X, Y ) diskretnog tipa imamo



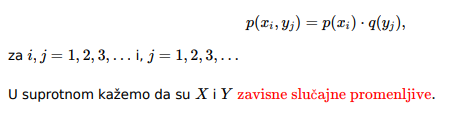
Verovatnoća p(xi |yj ) za i = 1, 2, 3,… predstavlja uslovnu verovatnoću za slučajnu promenljivu X, pri uslovu da je Y = yj.

Da bismo odredili uslovnu raspodelu za X|Y potrebno je da odredimo marginalnu raspodelu za slučajnu promenljivu Y. Marginalna raspodela predstavlja raspodelu samo jedne od slučajnih promenljivih koje se javljaju u dvodimenzionalnoj slučajnoj promenljivoj.



za i = 1, 2, 3,… Verovatnoća p(xi), i = 1, 2, 3,… naziva se marginalna raspodela za slučajnu promenljivu X. Na sličan način se određuje i za Y.

Nezavisnost dve slučajne promenljive diskretnog tipa se određuje iz odgovarajućih marginalnih raspodela. Neka je dvodimenzionalna slučajna promenljiva (X, Y) diskretnog tipa. Slučajne promenljive X i Y su nezavisne slučajne promenljive ako i samo ako



1. **Koeficijent linearne korelacije**

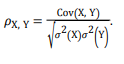
Koeficijent korelacije predstavlja broj kojim se na odredjeni način meri stepen zavisnosti slučajnih promenljivih X i Y. Zamislimo da neki eksperiment ponavljamo mnogo puta i u svakom ponavljanju registrujemo brojčane vrednosti (x, y) koje uzima dvodimenzionalna slučajna promenljiva (X, Y).

Posmatrajući te slučajeve možemo da zaključimo koja je najmanja zavisnost izmedju slučajnih promenljivih X i Y.

Koeficijent korelacije je definisan preko varijansi i kovarijansi slučajnih promenljivih X i Y. Kovarijansa se definiše na sledeći način:



za slučajne promenljive X i Y koeficijent korelacije, u oznaci, 𝜌X, Y , definiše se sa:



Koeficijent proste linearne korelacije može uzeti vrednost samo iz intervala [ − 1, 1]. Ovaj koeficijent ima sledeće osobine:

1) |𝜌X, Y | ≤ 1,

2) |𝜌X, Y | = 1 ako i samo ako je Y = 𝛼X + 𝛽, gde je 𝛼>0 za 𝜌 = 1, a 𝛼<0 za 𝜌=-1

3) ako su slučajne promenljive X i Y nezavisne slučajne promenljive, tada je 𝜌X, Y = 0 (obrnuto ne mora da važi).

1. **Puasonova raspodela**

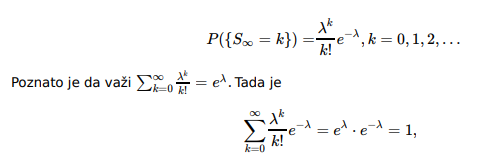
Puasonova raspodela je jedan granični slučaj Binomne raspodele i primenjuue se kada je n ⋅ p ≤ 10.

Radi bržeg izračunavanja binomnih verovatnoća kada je n veliko, od interesa je posmatrati njihovo asimptotsko ponašanje kada n neograničeno raste. Jedna od aproksimacija kada n → +∞ jeste Puasonova aproksimacija .

Stav. Ako u Binomnoj raspodeli n ⋅ pn → λ > 0, kada n → +∞, tada važi da je



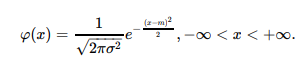
Prethodnim stavom je data aproksimacija ( n k )p k q n−k ≈ λ k k! e −λ za veliko n, gde se uzima da je λ ≈ n ⋅ p. Procena greške (koju ovde ne navodimo) pokazuje da se za primene zadovoljavajuća tačnost dobija već kada je n reda nekoliko desetina, a n ⋅ p ≤ 10. Na osnovu prethodnog važi da je verovatnoća događaja A, mala, pa se prethodni stav naziva Stav o malim verovatnoćama.



pa slučajna promenljiva S∞ uzima vrednosti u prebrojivom skupu RS∞ = {0, 1, 2,…}. Dakle, ona je slučajna raspodela diskretnog tipa. Raspodela verovatnoća za S∞ zove se Puasonova raspodela sa parametrom λ > 0 i to ćemo zapisivati S∞ : P(λ).

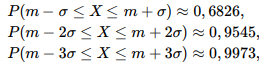
1. **Normalna raspodela**

Granični slučaj binomne raspodele kada n → +∞ jeste i normalna raspodela. U opštem slučaju za slučajnu promenljivu Z kažemo da ima normalnu raspodelu ili Gausovu raspodelu sa parametrima m ( m je realan broj) i σ 2 ( σ 2 je pozitivan broj), ako ima gustinu oblika



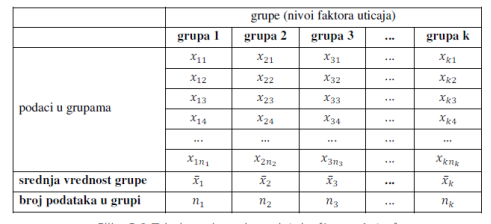
To ćemo zapisivati Z : N (m, σ 2 ). Dakle, ova raspodela zavisi od dva parametra m i σ. Grafik koji odgovara funkciji φ(x) se naziva Gausova kriva i ona je simetrična u odnosu na pravu x = m.

Normalna raspodela ima najveći značaj među raspodelama verovatnoća. Veoma je važno napomenuti da za normalnu raspodelu N(m, σ 2 ) važi tzv. σ -pravilo:



1. **Jednofaktorska analiza varijanse – ANOVA**

Podaci koji pripadaju različitim grupama i iz kojih su izračunate srednje vrednosti za svaku grupu, mogu da se predstave preko naredne tabele.



Kod testiranja hipoteze za dve srednje vrednosti postoji jedna nulta i samo jedna alternativna hipoteza, a kod analize varijanse postoji jedna nulta i više alternativnih hipoteza. Nulta i alternativna hipoteza koje mogu da se postave kod analize varijanse (za najjednostavniji primer sa tri grupe podataka) su:

Nulta hipoteza : H0 : m1 = m2 = m3.

Alternativne hipoteze :

1) H1 : m1 ≠ m2 ≠ m3(sve srednje vrednosti se razlikuju jedna od druge)

2) H1 : m1 ≠ m2 = m3 ≠ m1,(srednja vrednost u grupi 1 razlikuje od srednjih vrednosti u grupama 2 i 3, a srednje vrednosti u grupama 2 i 3 se ne razlikuju među sobom)

3) H1 : m1 ≠ m2 ≠ m3 = m1(srednja vrednost u grupi 2 razlikuje od srednjih vrednosti u grupama 1 i 3, a srednje vrednosti u grupama 1 i 3 se ne razlikuju među sobom)

4) H1 : m1 = m2 ≠ m3 ≠ m1(srednja vrednost u grupi 3 razlikuje od srednjih vrednosti u grupama 1 i 2, a srednje vrednosti u grupama 1 i 2 se ne razlikuju među sobom)

5) H1 : najmanje jedno m ≠ mi , (i = 1, 2, 3).

6) H1 : nisu sve m jednake

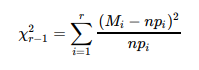
Ako se posle testiranja prihvati nulta hipoteza zaključak je da se srednje vrednosti ne razlikuju, odnosno da nezavisno promenljiva (faktor uticaja) nema efekta. Ako se prihvati jedna od alternativnih hipoteza zaključak je da ispitivani faktor uticaja ima efekta na srednje vrednosti onih grupa podataka koje se značajno razlikuju.

Za testiranje prethodno pomenutih hipoteza koristi se F-test. Osnovna pretpostavka kod analize varijanse je upoređivanje dva tipa varijansi (ili varijacija), varijansa između grupa upoređuje se sa veličinom koja se naziva varijansa unutar grupa, da bi se ocenila razlika između srednjih vrednosti. Mere varijacije se dobijaju "razdvajanjem" ukupne varijacije na varijaciju koja je posledica ispitivanog faktora uticaja (varijacija između grupa) i slučajnu varijaciju (varijacija unutar grupa).

1. **. 𝝌 𝟐 -test saglasnosti.**

Pri ispitivanju raspodela obeležja X iz uzorka, do sada smo pretpostavljali da nam je poznat tip raspodele, npr. X : N (m, σ 2 ), a da treba testirati hipoteze koje se odnose na parametre m i σ 2 . χ 2 test ili Pirsonov test, kako se još naziva, se koristi kada se testira nulta hipoteza H0(Fx = F0), gde je F0 data funkcija raspodele, a Fx funkcija raspodele obeležja X, protiv alternativne hipoteze H1(Fx ≠ F0).

Ovaj test se zasniva na χ 2 raspodeli, pa zbog toga nosi takvo ime. Inače u statistiku ga je uveo Pirson, pa se često naziva i Pirsonov test. Neka je (X1, X2,…, Xn) prost slučajni uzorak obeležja X na osnovu koga testiramo nultu hipotezu H0(Fx = F0). Postupak testiranja se izvodi tako što, najpre, realna osa podeli na r disjunktnih intervala S1, S2,…, Sr,tj. Si ∩ Sj = ∅, za i ≠ j,i, j ∈ {1, 2,…, r} tako da važi R = S1 ∪ S2 ∪ … ∪ Sr.Označimo sa M1, M2,…, Mr slučajne promenljive koje predstavljaju broj vrednosti iz uzorka (X1, X2,…, Xn), (r < n) koje se javljaju iz skupova S1, S2,…, Sr , tim redom. Označimo, dalje, sa pi = PH0 {X ∈ Sk}, i = 1, 2,…r. Dakle, pi je verovatnoća da vrednost obeležja X pripada skupu Si , pod uslovom da je hipoteza H0 tačna. Slučajna promenljiva Mi : B(n; pi), i = 1, 2,…r. Njeno matematičko očekivanje, s obzirom da se radi o Binomnoj raspodeli je E(Mi) = n ⋅ pi . Može se dokazati da sledeća test-statistika

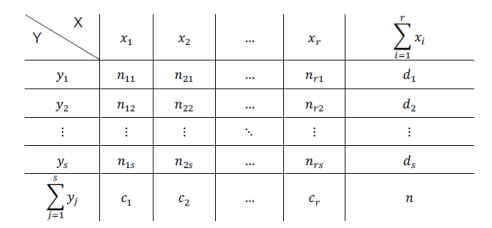


za veliko n (n → +∞) približno ima χ 2 r−1 raspodelu, ako je hipoteza H0 tačna. Naime, ako je H0 tačna za dati prag značajnosti α, onda je odstupanja Mk od E(Mk) malo, za k = 1, 2,…, r, tj. kritična oblast je [χ 2 r−1;α , +∞), gde broj χ 2 r−1;α čitamo iz tablice za χ 2 raspodelu.

Ako funkcija raspodele F0 zavisi od l parametara (l > 1), tada je uz uslov da je hipoteza H0 tačna, data test statistika, za veliko n, je približno jednaka χ 2 r−l−1 raspodeli. Dakle, tada je granica kritična vrednost koja χ 2 r−l−1;α koja se određuje iz tablice za χ 2 raspodelu sa r − l − 1 stepeni slobode. Parametri raspodele F0 se ocenjuju metodom maksimalne verodostojnosti i oni predstavljaju njihove postojane ocene, o čemu smo već govorili. U nastavku ćemo se podsetiti postojanih ocena parametara za raspodele koje ćemo koristiti.

1. **𝝌 𝟐 -test nezavisnosti.**

anje hipoteze H0 da su dva obeležja X i Y istog uzorka (ili istobrojnih različitih uzoraka) nezavisna. Dakle, dva nezavisna uzorka koja se testiraju uzeta su iz jednog skupa i testira se povezanost između dva obeležja. Praktično ceo skup,odnosno uzorak može da se podeli na dva dela: na grupu jedinica koje „imaju“ i grupu jedinica koje „nemaju“ neko obeležje. Neka je (X1, Y1); (X2, Y2),…,(Xn, Yn) uzorak. Neka je pritom skup vrednosti obeležja X dat sa R(X) = {a1, a2,…, ar} i skup vrednosti obeležja Y , dat sa R(Y ) = {b1, b2,…, bs}. Skup vrednosti obeležja (X, Y ) je dat sa R(X, Y ) = {(ai , bj ) ∣ 1 ≤ i ≤ r, 1 ≤ j ≤ s}. Dalje, neka su: ∙ nij frekvencija od (ai , bj ) u uzorku, ∙ ci (marginalna) frekvencija od ai u uzorku, ∙ dj (marginalna) frekvencija od bj u uzorku i važi da je ci = ∑s j=1 nij , dj = ∑r i=1 nij . Ti podaci se daju u tablici kontigencije (povezanosti):



Označimo, takođe, sledeće verovatnoće

pij = P(X = ai , Y = bj), 1 ≤ i ≤ r, 1 ≤ j ≤ s,

pi = P(X = ai), 1 ≤ i ≤ r,

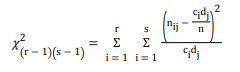
qj = P(Y = bj), 1 ≤ j ≤ s.

Hipoteze su:

H0 : X i Y su nezavisna obeležja, tj da važi : pij = piqj za svako i i j,

H1 : X i Y su zavisna obeležja, tj. postoje i i j takvi da pij ≠ piqj .

Za veliko n, smatrajući da je H0 tačna, koristi se statistika:



i izračunava se ^ 𝜒(r − 1)(s − 1) 2 (vrednosti nij , ci , dj u ovoj formuli shvatamo kao slučajnu promenljivu, jer variraju od uzorka do uzorka). Za dato 𝛼, odredimo 𝜒 (r − 1)(s − 1); 𝛼 2 iz PH0{ 𝜒 {r − 1}(s − 1) 2 > 𝜒 {r − 1}(s − 1); 𝛼 2 } = 𝛼 i kritična oblast je [ 𝜒 [r − 1)[s − 1); 𝛼 2 , + ∞ ). Ako uzoračna vrednost ^ 𝜒(r − 1)(s − 1) 2 ∈ [ 𝜒 [r − 1)[s − 1); 𝛼 2 , + ∞ ) odbacujemo H0 o nezavisnosti obeležja X i Y, tj. ova obeležja su zavisna.

1. **𝝌 𝟐 -test homogenosti**

Ovaj test utvrđuje da li ispitivani nezavisni uzorci pripadaju istom ili su uzeti iz različitih skupova. Kod testa homogenosti postupak izračunavanja je isti, ali on nije identičan sa testom nezavisnosti. Testom nezavisnosti istražujemo razliku između dva obeležja uzoraka uzetih iz jednog skupa, a testom homogenosti ispitujemo razliku između više nezavisnih uzoraka izvučenih iz različitih skupova. Testiranje se bazira na χ 2 raspodeli.

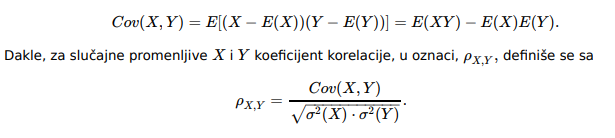
1. **Korelacija. Koeficijent linearne korelacije. Test koreliranost dve slučajne promenljive.**

Pojam stohastika je starogrčka reč i predstavlja sinonim s pojmom slučajnost. On se koristi u situacijama kada se želi istaći da se neka pojava ili više njih posmatraju sa stanovišta slučajnosti.

Osobina korelacije među pojavama jeste da se ispoljava u masi slučajeva. Stoga se u njihovoj osnovi nalazi Zakon velikih brojeva.

Korelacija pretpostavlja stohastičku (verovatnosnu) vezu među pojavama, zbog čega pojave moraju imati osobinu slučajnosti. Najčešće je pretpostavka da slučajne promenljive (pojave) X i Y imaju normalne raspodele i tada govorimo o teoriji Normalne korelacije. Teorija Normalne korelacije pretpostavlja, dakle, postojanje normalnih raspodela korelirajućih slučajnih promenljivih. Uzročne veze među pojavama pretpostavka su postojanja korelacionih veza.

U statističkim ispitivanjima koristi se veoma veliki broj parametara kao mera korelacionih veza. Ipak, najčešće se koristi tzv. koeficijent proste linearne korelacije ili Pirsonov koeficijent, koji služi kao mera linearne zavisnosti promenljivih X i Y . Koeficijent korelacije je definisan preko varijansi i kovarijansi slučajnih promenljivih X i Y . Kovarijansa se definiše na sledeći način



Koeficijent proste linearne korelacije u osnovnom skupu obeležavamo sa ρ, a u uzorku obeležavamo sa rX,Y . Koeficijent proste linearne korelacije može uzeti vrednost samo iz intervala [−1, 1].

U praksi se javlja još jedan problem, nakon što se odredi koeficijent proste linearne korelacije. Naime, koeficijent proste linearne korelacije se obično izračunava iz uzorka. Postavlja se pitanje njegove značajnosti za celu populaciju, odnosno da li uzorak iz koga je izračunat koeficijent dovoljno reprezentativan za donošenje nepristrasne ocene koeficijenta osnovnog skupa. Dok se to ne utvrdi, dobijena vrednost koeficijenta na osnovu uzorka predstavlja samo hipotezu o vrednosti istog koeficijenta osnovnog skupa. Problem se rešava na sledeći način: Testira sehipoteza da li izračunati prost koeficijent linearne korelacije izuzorka rX,Y predstavlja preciznu ocenu prostog koeficijenta linearne korelacije osnovnog skupa ρX,Y . Ako odgovarajućim testom odbacimo nultu hipotezu, prihvatamo izračunatu vrednost koeficijenta korelacije iz uzorka kao pravu ocenu koeficijenta u osnovnom skupu. Drugim rečima uzorak je reprezentativan, pa dobijeni rezultat može da se uopšti.

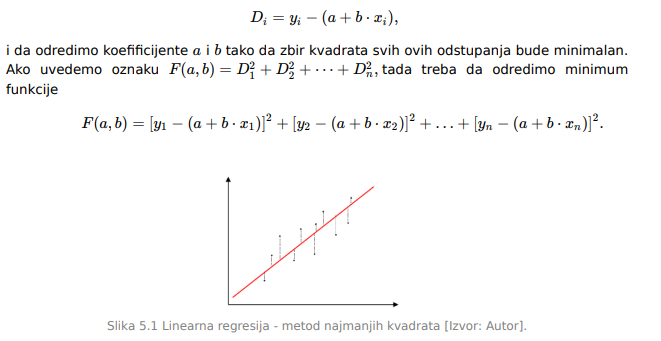
1. **Linearna regresija. Metod najmanjih kvadrata.**

Regresioni model je matematički model koji opisuje vezu između dve ili više promenljivih. Prost regresioni model obuhvata samo dve promenljive: jednu objašnjavajuću i jednu zavisnu. Zavisna promenljiva je promenljiva čije varijacije treba da objasnimo na osnovu kretanja objašnjavajuće promenljive. U regresionom modelu, objašnjavajuća promenljiva se obično obeležava sa X, a zavisna promenljiva sa Y.

Međuzavisnost između dve promenljive u regresionoj analizi izražava se matematičkom jednačinom koja se naziva regresiona jednačina ili regresioni model. Regresioni model kojim izražavamo linearnu vezu između dve promenljive se naziva linearni regresioni model.

Prost linearni regresioni model je oblika: Y = a + b \* X, gde koeficijent b predstavlja koeficijent pravca prave, a a odsečak na Oy osi. Ovaj model se naziva deterministički model.

Postoji egzaktan matematički način kojim se prikazuje najbolje prilagođen pravac linearne veze. Određuje se iz uslova da je zbir kvadrata vertikalnih udaljenosti tačaka od pravca najmanji – metoda najmanjih kvadrata. Osnove metode najmanjih kvadrata predložio je Karl Gaus. U slučaju regresione prave (videti sliku ispod), ideja je da za posmatrani uzorak (x1, y1),(x1, y1),…,(xnyn), uočimo i -to odstupanje, u oznaci Di , (i = 1, 2,…, n) za koje važi



S obzirom da je F(a, b) funkcija dve promenljive, da bismo odredili minimum funkcije potrebno je prvo odrediti stacionarne tačke. Njih određujemo iz sledećih uslova



Kao rezultata ovog procesa, dobijamo da se traženi minimum postiže za

