Statistički praktikum 1 - 36. zadatak

Hermina Petric Maretić

22. siječnja 2014.

Opis zadatka

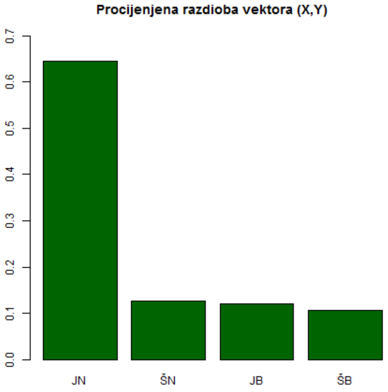
- List indijanske penjačice može biti šaren ili jednoličan, te istovremeno blijed ili normalne boje.
- Postavljeno je pitanje jesu li te dvije karakteristike nezavisne.
- Prijašnji eksperimenti su pokazali da će mladica s vjerojatnosti $1/4$ biti blijeda (a normalne boje s vjerojatnosti $3/4$). Isto tako lišće mladica ce s vjerojatnosti $1/4$ biti šareno (a s vjerojatnosti $3/4$ jednolično).
- Na slučajan način prikupljen je uzorak od 290 mladica indijske penjačice.

	normalne boje	blijedo
jednolično	187	35
šareno	37	31

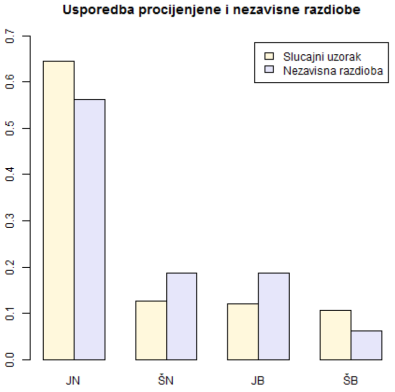
(a) zadatak

- Neka je X stanje prve karakteristike (šareno ili lišće jednolične boje), a Y stanje druge karakteristike (blijedo lišće ili normalne boje).
- Na osnovi danog uzorka, procijenite zajedničku razdiobu varijabli X i Y, tj. procijenite razdiobu vektora (X,Y) i grafički je predstavite.
- Također, graficki je usporedite s razdiobom u kojoj su X i Y nezavisne, a marginalne razdiobe su u skladu s prijašnjim saznanjima.

Razdioba vektora (X,Y)



Usporedba procijenjene i nezavisne razdiobe



(b) zadatak

- Testirajte da li su X i Y nezavisne s pretpostavljenim marginalnim razdiobama kao u tekstu zadatka.
- Postavljamo hipoteze:
- H_0= X i Y su nezavisne s pretpostavljenim marginalnim razdiobama
- H_0= X i Y nisu neazavisne s pretpostavljenim marginalnim razdiobama
- Za testiranje koristimo Pearsonov χ^2-test o pripadnosti distribuciji s testnom statistikom
  $H = \sum_{i=1}^{r} \sum_{j=1}^{c} \dfrac{(f_{i,j} - f'_{i,j})^2}{f'_{i,j}} \sim \chi^2(r * c - d - 1)$

(b) zadatak

- Kritično područje: $[7.815, \infty\rangle$
- Vrijednost testne statistike: h=25.096
- Odbacujemo nultu hipotezu na razini značajnosti od 5%
- P-vrijednost: $1.47 * 10^{-5}$

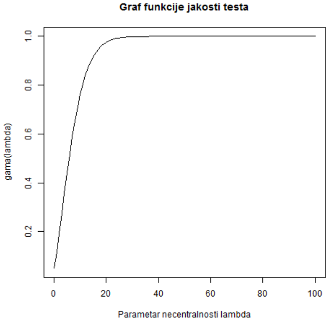
(c) zadatak

- Odredite jakost testa iz (b) uz značajnost 5%
- Postavljamo hipoteze:
- $H_0 : p = p^{(0)}$
- $H_1 : p = p^{(0)} - \frac{1}{\sqrt{n}}\delta$, za neki $\delta \neq 0$

(c) zadatak

- Testna statistika je
  $H = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{i,j} - f'_{i,j})^2}{f'_{i,j}} \sim \chi^2(r * c - d - 1)$
- Ako vrijedi hipoteza H_1, tada je $H \sim A_{\chi^2}(k-1, \lambda)$, gdje je
  $\lambda = \sum_{k=1}^{k} \frac{\delta_i^2}{p_i^0}$ parametar necentralnosti.
- Tada funkciju jakosti Pearsonovog χ^2-testa možemo prikazati ovako:
  $\gamma(\lambda) = P(H \geq \chi^2(k-1) | p = p^0 - \frac{1}{\sqrt{n}} * \delta) =$
  $1 - P_{\chi^2(k-1,\lambda)}(\chi^2_\alpha(k-1))$

(c) zadatak



- Minimum se postiže u točki 0 i iznosi 0.05. To proizlazi iz činjenice da hipoteza H_0 govori upravo da je $\lambda = 0$ i iz činjenice da test ima razinu značajnosti 5%.

(d) zadatak

- Prema jednoj drugoj teoriji, razdioba od (X,Y) je oblika

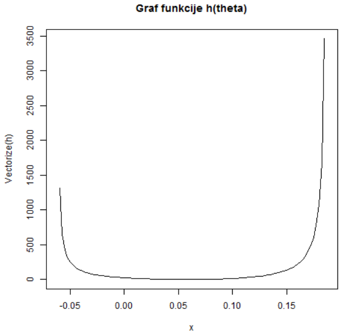
X Y	normalne boje	blijedo
jednolično	${}^9/_{16} + \theta$	${}^3/_{16} - \theta$
šareno	${}^3/_{16} - \theta$	${}^1/_{16} + \theta$

gdje je H nepoznati parametar. Odredite parametarski prostor i pomoću χ^2-testa testirajte tu hipotezu, pri čemu nepoznati parametar procjenite minimum χ^2-metodom.

(d) zadatak

- Parametarski prostor: $\theta \in [-1/16, 3/16]$
- Minimum χ^2-metoda se temelji na tome da za funkciju $h(\theta) : \Theta \to R$, $h(\theta) = \sum_{i=1}^{k} \sum_{j=1}^{c} \dfrac{(f_{i,j} - f'_{i,j}(\theta))^2}{f'_{i,j}(\theta)}$ tražimo vrijednost $\hat{\theta}$ u kojoj ona postiže minimum.

(d) zadatak



- Kao što i otprilike vidimo na grafu, minimum se postiže u točki 0.0586, pa je baš to naš $\hat{\theta}$.

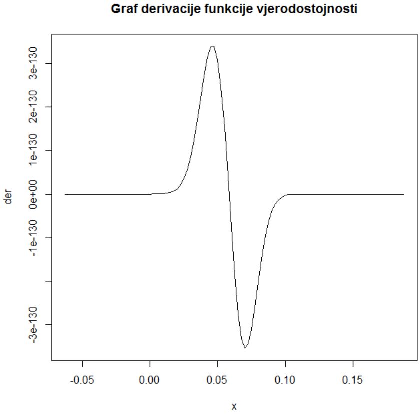
(d) zadatak

- Testiramo hipotezu da je razdioba našeg slučajnog vektora iz zadanog modela
- Iskoristimo procijenjeni parametar θ i provodimo χ^2-test s 2 (d=1 jer imamo jedan procijenjeni parametar) stupnja slobode.
- P-vrijednost: 0.6371446
- P-vrijednost je velika, pa ne možemo odbaciti hipotezu o pripadnosti zadanom modelu.

(e) zadatak

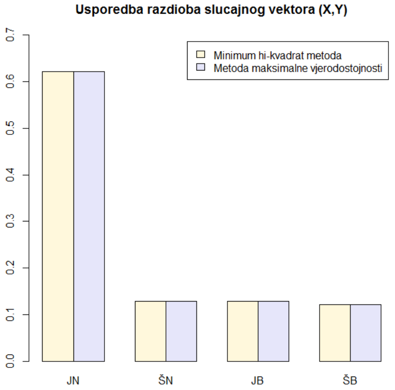
- Procijenite nepoznati prametar iz modela u (d) metodom maksimalne vjerodostojnosti i usporedite ga s minimum χ^2-procjenom
- Definiramo funkciju vjerodostojnosti $L : \Theta \to R$, sa $L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. Tražimo θ koji maksimizira ovu funkciju i zovemo ga procjenom metodom maksimalne vjerodostojnosti.

(e) zadatak



- $\hat{\theta} = 0.05840407$

(e) zadatak



Usporedba razdioba slucajnog vektora (X,Y)
Minimum hi-kvadrat metoda
Metoda maksimalne vjerodostojnosti
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
JN
ŠN
JB
ŠB

Hvala na pažnji!