

Projektna naloga pri predmetu Statistika

Beno Učakar

Profesor: doc. dr. Martin Raič

1 2. naloga

1.1 Primer (a)

Najprej uvedimo nekaj oznak. Če je $k \in \{1, \dots, 12\}$ število skokov ptic, naj bo S_k frekvenca tega opažanja. Podatke z novimi oznakami predstavimo v spodnji tabeli. Skupno število opaženih skokov označimo z S , število vseh

k	1	2	3	4	5	6	7	8	9	10	11	12
S_k	48	31	20	9	6	5	4	2	1	1	2	1

Tabela 1: Frekvence števila skokov

opažanj pa z N . Velja

$$N = \sum_{k=1}^{12} S_k \quad S = \sum_{k=1}^{12} k S_k.$$

V našem primeru znaša $N = 130$ in $S = 363$.

Želimo poiskati geometrijsko porazdelitev, ki se najbolj prilega tem podatkom. Če je število skokov posameznega ptiča slučajna spremenljivka $K \sim \text{Geom}(p)$, v resnici iščemo cenilko za parameter p . To znamo narediti na vsaj dva načina.

Prvi način: Postopamo po metodi momentov. Spomnimo se, da pričakovana vrednost geometrijske porazdelitve $\text{Geom}(p)$ znaša $\frac{1}{p}$. Zato velja

$$p = \frac{1}{E(K)}.$$

Po metodi momentov $E(K)$ ocenimo s prvim momentom opaženih vrednosti

$$\frac{1}{N} \sum_{k=1}^{12} k S_k = \frac{S}{N},$$

kar nam da cenilko

$$\hat{p} = \frac{N}{S}.$$

Drugi način: Postopamo po metodi največjega verjetja. Verjetnostna funkcija geometrijske porazdelitve $\text{Geom}(p)$ je $P(K = k) = p(1 - p)^{k-1}$. Verjetje lahko torej izrazimo kot

$$L(p \mid S_1, \dots, S_{12}) = p^N (1 - p)^{S - N}.$$

Ko logaritmujemo, dobimo

$$l(p \mid S_1, \dots, S_{12}) = N \ln\left(\frac{p}{1 - p}\right) + S \ln(1 - p).$$

Če parcialno odvajamo po p in malo računamo, ponovno pridemo do cenilke

$$\hat{p} = \frac{N}{S}.$$

V obeh primerih pridemo do iste cenilke. Ta v našem primeru znaša

$$\hat{p} = \frac{130}{363} \approx 0.358.$$

Teorija metode momentov in metode največjega verjetja nam zagotovita, da je ta izbira smiselna. Iskana geometrijska porazdelitev je $\text{Geom}(\hat{p})$.

1.2 Primer (b)

Ob predpostavki, da je $K \sim \text{Geom}(\hat{p})$, poračunamo verjetnosti p_k , da pri enem opazanju pride do k skokov. Pričakovano vrednosti frekvenc določimo kot

$$\hat{S}_k = p_k S.$$

Dobimo spodnjo tabelo.

k	1	2	3	4	5	6
p_k	0.3580	0.2298	0.1476	0.0947	0.0608	0.0390
\hat{S}_k	46.54	29.87	19.19	12.31	7.90	5.07
k	7	8	9	10	11	12
p_k	0.0251	0.0161	0.0103	0.0066	0.0043	0.0027
\hat{S}_k	3.26	2.09	1.34	0.86	0.56	0.35

Tabela 2: Pričakovane frekvence skokov.

Te podatke združimo v črtni grafikon.

1.3 Primer (c)