

Kolokvij 1 - Osnove teoretične statistike

16. februar, 2017

1. Paretova porazdelitev s končno pričakovano vrednostjo ima naslednjo gostoto

$$f(x) = a \frac{b^a}{x^{a+1}} \quad x \geq b, \quad a > 1, b > 0.$$

Naj bo X_1, \dots, X_n vzorec neodvisnih spremenljivk porazdeljenih po Paretovi porazdelitvi, kjer naj bo a neznani parameter in b znana konstanta.

- (a) Zapišite cenilko po metodi momentov za a , če veste, da je pričakovana vrednost spremenljivke X enaka $ab/(a-1)$.
 - (b) Izračunajte cenilko po metodi največjega verjetja za a .
 - (c) Zapišite asimptotski $(1-\alpha)$ interval zaupanja za pravo vrednost a .
 - (d) Komentirajte doslednost cenilke za a po metodi momentov.
2. Imamo k škatel, v katere razporedimo n kroglic. Verjetnost, da damo kroglico v škatlo i , označimo s p_i , X_i označuje število kroglic v i -ti škatli. To porazdelitev imenujemo multinomska porazdelitev, označimo

$$(X_1, \dots, X_k) \sim \text{Multinom}(n, (p_1, \dots, p_k)), \sum_{i=1}^k p_i = 1$$

- (a) Binomska porazdelitev opisuje poseben primer multinomske porazdelitve za $k=2$. Izpeljite formulo za kovarianco med X_1 in X_2 za $k=2$.
3. Pojavnost genotipov posameznega gena je v populaciji porazdeljena v Hardy-Weinbergovem ravnotežju, možni so trije genotipi (AA, Aa, aa) z verjetnostmi $(\theta^2, 2\theta(1-\theta), (1-\theta)^2)$. Naj N_1 označuje število genotipov AA, N_2 število genotipov Aa in N_3 število genotipov aa na vzorcu.
- (a) Izpeljite cenilko za θ po metodi največjega verjetja.
 - (b) Ali je cenilka nepristranska? Pokažite.
 - (c) Izračunajte eksaktno formulo za varianco te cenilke (brez uporabe metode največjega verjetja).

Pri reševanju naloge si lahko pomagata z naslednjim rezultatom. Naj bodo

$$(X_1, \dots, X_k) \sim \text{Multinom}(n, (p_1, \dots, p_k))$$

Potem velja: $X_i \sim \text{Bin}(n, p_i)$, $\text{cov}(X_i, X_j) = -np_i p_j$, za $i \neq j$.

- (d) (vprašanje za bonus točke) Brez uporabe metode največjega verjetja utemeljite, da je cenilka približno normalno porazdeljena. *Namig: definirajte novo slučajno spremenljivko, izrazite cenilko kot vsoto vrednosti te slučajne spremenljivke.*
4. Fizioterapevti bi radi ugotovili, kakšen je povprečen rezultat testa hoje (v sekundah) pri zdravih posameznikih med 30 in 40 leti starosti. V ta namen testirajo 100 naključno izbranih zdravih posameznikov, ki opravijo test hoje po dvakrat. Zabeležijo si vse rezultate in se odločijo da bodo naredili analizo (t.j. določili interval zaupanja za pričakovano vrednost testa hoje) na povprečnih vrednostih vsakega posameznika.

- (a) Povejte, kakšen način vzorčenja so izbrali fizioterapevti in vpeljite oznake oz. zapišite model za en rezultat testa hoje, tako da bo omogočal razlikovanje med povprečjem posameznika in celotnim povprečjem.
- (b) Zapišite (izpeljite), kakšna bi bila varianca ocenjenega skupnega povprečja, če bi se odločili, da analizirajo le prve teste hoje pri vsakem posamezniku.
- (c) Izpeljite, kakšna bo varianca cenilke skupnega povprečja za njihov primer (pri vsakem posamezniku vzamejo povprečje).
- (d) Interpretirajte in primerjajte dobljeni formuli v točkah (b) in (c).
- (e) Napišite psevdo kodo za simulacijo s katero bi preverili, da je varianca ocen približno enaka vaši teoretično izpeljani varianci cenilke.