1. kolokvij - Osnove teoretične statistike

20. februar, 2015

1. Vprašanja za ogrevanje
   Kaj velja za asimptotsko gostoto vzorčnih povprečij (spremenljivke X_i so iz neke porazdelitve s končnim povprečjem in varianco)? **Utemeljite** vsako odločitev.

   - Povprečje in mediana sta enaka.
   - Površina pod krivuljo = 1.
   - Krivulja je asimetrična v desno.
   - Standardnemu odklonu pravimo standardna napaka.
   - Interval zaupanja je asimetičen.
   - Velikost vzorca vpliva na vse parametre, s katerimi bi opisali to funkcijo.

2. Nepristranskost
   Na avtobusnem postajališču mimo katerega poteka ena sama avtobusna proga, so izvajali raziskavo o zadovoljstvu potnikov. Med drugim so za vsakega potnika (skupno število je n) zabeležili čas čakanja (od prihoda na postajo do odhoda naslednjega avtobusa). S pomočjo teh podatkov bi radi ocenili, na koliko minut vozi avtobus. Pod predpostavko, da potniki prihajajo na postajo naključno in neodvisno, so vrednosti njihovih časov enakomerno porazdeljene slučajne spremenljivke na intervalu $[0, \theta]$, kjer je θ časovni razmak med prihodi avtobusa.
   Imamo dve ideji, kako oceniti θ:

   (i) Pri oceni bi si pomagali s povprečnim časom čakanja ($\overline{X}$).

   (ii) Vrednost bi ocenili k z največjim časom čakanja (označimo ga z $X_{(n)} = \max_i X_i$).

   Pri spodnjih vprašanjih si pomagajte z naslednjimi rezultati:

   $$E(X_{(n)}) = \frac{n}{n+1}\theta \qquad , \qquad E(X_{(n)}^2) = \frac{n}{n+2}\theta^2$$
   $$Y \sim U[a,b] \quad \Rightarrow \quad var(Y) = \frac{(b-a)^2}{12}$$

   Vprašanja:

   (a) Zapišite nepristransko cenilko za vsako izmed zgornjih idej, za vsako cenilko pokažite, da je zares nepristranska.

   (b) Imamo torej dve nepristranski cenilki za isti parameter. Na podlagi katere lastnosti bi se odločali med njima?

   (c) Izpeljite to lastnost za vsako izmed cenilk in ju primerjajte. Utemeljite, za katero cenilko se boste odločili?

   (d) Kdaj so naše predpostavke smiselne oz. kdaj ne bi bile?

3. Metoda največjega verjetja

   Otroci mečejo žogo na koš, vsak meče toliko časa dokler ne zadene. Naj X_i označuje število zgrešenih metov otroka i. Če imajo otroci pri vsakem metu verjetnost p, da koš zadenejo, in so njihovi meti neodvisni med seboj, je slučajna spremenljivka X_i porazdeljena geometrijsko, $P(X_i = k) = (1-p)^k p$.

   Pri spodnjih vprašanjih vam bo koristil rezultat: $E(X_i) = \frac{1-p}{p}$

   (a) Poiščite cenilko za parameter p po metodi največjega verjetja, intuitivno razložite dobljeno formulo.

   (b) Kaj lahko rečete o doslednosti cenilke?

   (c) Poiščite cenilko za standardno napako po metodi največjega verjetja.

4. Stratificirano vzorčenje

   Raziskovalca A in B sta želela oceniti dosežek na testu znanja matematike na osnovnih šolah ljubljanske regije. Število vseh učencev je N=2000, razdeljeni so med K=30 šol ($i = 1, \ldots K$), na vsaki šoli pa je L_i (L_i je vsaj 30) učencev. Vsaka šola ima seveda lahko drugačno povprečje (označimo ta povprečja z μ_i), predpostavimo pa, da imajo vse enako varianco ($\sigma_i^2 = \sigma^2$).

   Raziskovalec namerava v vzorec zajeti n učencev, odloča se me naslednjima shemama vzorčenja:

   A. vse šole, na vsaki vzorec enake velikosti $k = \frac{n}{K}$

   B. vse šole, kjer je n_i je proporcionalen glede na velikost L_i

   Zapišite:

   (a) Populacijsko povprečje s povprečji na posameznih šolah

   (b) Zapišite cenilki za povprečje za obe shemi z utežmi, ki so proporcionalne glede na L_i. Ali sta taki cenilki nepristranski?

   (c) Zapišite varianci obeh cenilk (pri tem lahko popravek za končnost zanemarite). Za katero shemo bi se odločili?

   (d) Kako bi SE za shemo A ocenili s simulacijami?