Todennäköisyyslaskennan ja tilastotieteen peruskurssi
Matematiikan ja systeemianalyysin laitos
Aalto-yliopisto

Tentti
11.12.2017
L Leskelä

Ohje: Vastaa lyhyesti ja ytimekkäästi, mutta perustele ratkaisusi. Pelkkä lukuarvo vastauksena ei anna pisteitä. Kokeessa on 4 tehtävää, jokaisesta saa 0–6 pistettä. Merkitse jokaiseen vastauspaperiin:

- Kurssin nimi ja koodi
- SUKUNIMI, ETUNIMET ja OPISKELIJANUMERO (tikkukirjaimin)
- Koulutusohjelma ja vuosikurssi
- Päivämäärä ja nimikirjoitus

Sallitut apuvälineet: Mellinin tilastolliset taulukot, laskin ja a4-muistilappu (käsin kirjoitettu, tekstiä vain toisella puolella, oikeassa yläkulmassa oma nimi, ei tarvitse palauttaa)

T1 Laatikossa on kolme noppaa: yksi 4-sivuinen ja kaksi 6-sivuista. Kaikki nopat on numeroitu normaaliin tapaan ykkösestä eteenpäin. Ystäväsi valitsee laatikosta yhden nopan umpimähkään ja heittää sitä kerran näyttämättä noppaa sinulle. Merkitään

$$S = \text{valitun nopan sivujen lukumäärä},$$
$$T = \text{nopanheiton tulos}.$$

(a) Määritä satunnaismuuttujien S ja T yhteisjakauma listaamalla kaksiulotteiseen taulukkoon parin (S, T) mahdollisten arvojen todennäköisyydet. **(2p)**

(b) Ovatko S ja T stokastisesti riippuvat vai riippumattomat? Perustele miksi. **(1p)**

(c) Määritä satunnaismuuttujan T jakauma ja odotusarvo. **(2p)**

(d) Ystäväsi paljastaa, että nopanheiton tulos on 3. Mikä on tämän tiedon valossa todennäköisyys, että heitetty noppa oli 6-sivuinen? **(1p)**

T2 Kun satunnaisotannalla valitun 2500 suomalaisen aikuisen pituudet (cm) mitattiin, saatiin keskiarvoksi 170.9 ja otoskeskihajonnaksi 11.4. Määritä mittaustulosten perusteella kaikkien suomalaisten aikuisten keskipituudelle μ:

(a) piste-estimaatti, **(1p)**

(b) likiarvoinen luottamustason 99% luottamusväli. **(2p)**

(c) Jos yo. mittaus uusittaisiin kymmenen kertaa suuremmalle satunnaisotokselle, niin miten paljon kapeampi luottamusväli todennäköisesti saataisiin 99% luottamustasolla? **(1p)**

(d) Onko (b)-kohdan luottamusvälin määrittämisessä tarpeen olettaa, että suomalaisaikuisten pituudet koko väestössä noudattavat likimain normaalijakaumaa? **(1p)**

(e) Selitä, mitä tarkoittaa harhaton estimaattori. Onko (a)-kohdassa raportoimasi piste-estimaatti laskettu harhattomalla estimaattorilla? **(1p)**

T3 Tieverkon sulana pitämiseen on varastoitu suolaa 200 cm lumimäärän varalle. Yksittäisen talvipäivän aikana lunta sataa keskimäärin 4.5 cm, keskihajonnan ollessa 2.5 cm. Tulevien talvipäivien lumimäärät oletetaan toisistaan riippumattomiksi.

(a) Laske 50 talvipäivän aikana sataneen lumimäärän odotusarvo ja keskihajonta. **(2p)**

(b) Onko yo. tietojen pohjalta mahdollista laskea, millä todennäköisyydellä yksittäisenä talvipäivänä sataa lunta yli 12 cm? Jos on, ilmoita vastaus. Jos ei, selitä miksi ei. **(1p)**

(c) Laske normaaliapproksimaation avulla arvio todennäköisyydelle, että suola riittää 50 talvipäiväksi. **(1p)**

(d) Arvioi normaaliapproksimaatiota käyttäen, kuinka paljon enemmän suolaa pitäisi varata, jotta sitä riittäisi 50 talvipäiväksi vähintään 99% todennäköisyydellä? **(1p)**

(e) Mitkä ylläolevista laskelmista pitävät paikkansa myös silloin, kun talvipäivien riippumattomuusoletuksesta luovutaan? **(1p)**

T4 Tuntematon datalähde tuottaa stokastisesti riippumattomia $\{0,1\}$-arvoisia satunnaislukuja tiheysfunktiona

$$f(x_i \mid \theta) = \begin{cases} 1-\theta, & x_i = 0, \\ \theta, & x_i = 1, \end{cases}$$

missä $\theta \in (0,1)$ on tuntematon parametri. Datalähteestä on havaittu luvut $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, $x_4 = 1$, $x_5 = 1$. Auta Ronaldia, Karlia ja Thomasia estimoimaan parametrin θ arvo näiden havaintojen pohjalta.

(a) Ronald päättää käyttää suurimman uskottavuuden estimaattia. Määritä parametrin uskottavuusfunktio ja laske sen avulla suurimman uskottavuuden estimaatti datajoukolle $x = (x_1, \ldots, x_5)$. **(2p)**

(b) Karlin käyttämän estimaatin arvo määräytyy ratkaisemalla θ yhtälöstä

$$\mathbb{E}(X \mid \theta) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

jonka vasemmalla puolella on jakaumaa $f(x_i \mid \theta)$ noudattavan satunnaismuuttujan odotusarvo ja oikealla puolella havaintojen keskiarvo. Laske tämän estimaatin arvo datajoukolle x. **(2p)**

(c) Thomas tulkitsee tuntemattoman parametrin satunnaismuuttujaksi ja valitsee priorijakaumaksi yksikkövälin tasajakauman tiheysfunktiona

$$p(\theta) = \begin{cases} 1, & \theta \in (0,1), \\ 0, & \text{muuten.} \end{cases}$$

Thomas päättää käyttää estimaattina posteriorijakauman odotusarvoa. Laske tämän estimaatin arvo datajoukolle x. **(2p)**