Todennäköisyyslaskennan ja tilastotieteen peruskurssi
MS-A050x — Tentti
Matematiikan ja systeemianalyysin laitos — 2.12.2024
Aalto-yliopisto — J Kohonen

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

Sallitut välineet: kirjoitusvälineet, laskin (symbolinen ja graafinen OK), enintään A4-kokoinen käsin kierjoitettu lunttilappu, jossa molemmat puolet saa käyttää.

Arviointiohje. Yleinen tarkastusohje. Tehtävien osista lasketaan pisteet tarvittaessa puolen pisteen tarkkuudella ja lasketaan yhteen. Pisteitä ei pyöristetä kokonaisluvuiksi.

T1 Avaruusluotaimessa on tietojen tallentamista varten neljä kovalevyä. Avaruuden ankarissa oloissa kullakin levyllä on todennäköisyys 0.1 hajota ennenaikaisesti (luotaimen suunniteltuna toiminta-aikana), toisista levyistä riippumatta.

(a) Levyt yhdistetään RAID0-järjestelmäksi, joka lakkaa toimimasta jos yksikin levy hajoaa. Mikä on todennäköisyys, että järjestelmä lakkaa toimimasta? Vastaus ainakin neljällä desimaalilla. **(2p)**

(b) Levyt yhdistetään RAID10-järjestelmäksi, jossa levyt 1 ja 2 muodostavat parin, ja levyt 3 ja 4 muodostavat parin. Järjestelmä toimii, jos kummassakin parissa ainakin jompikumpi levy toimii. Mikä on todennäköisyys, että järjestelmä lakkaa toimimasta? Vastaus ainakin neljällä desimaalilla. **(2p)**

(c) Levyt yhdistetään RAID6-järjestelmäksi, joka toimii kunhan ainakin kaksi levyistä toimii. Mikä on todennäköisyys, että järjestelmä lakkaa toimimasta? Vastaus ainakin neljällä desimaalilla. **(2p)**

Ratkaisu.

(a) Kullakin levyllä on todennäköisyys $1 - 0.1 = 0.9$ toimia (komplementtisääntö). Tn että kaikki toimivat on 0.9^4. Tn että ainakin yksi hajoaa on $1 - 0.9^4 = \mathbf{0.3439}$.

(b) **Tapa 1.** Parissa (1,2) ainakin jompikumpi toimii tn:llä $1 - 0.1^2 = 0.99$ (komplementti sille, että molemmat hajoavat). Samoin parissa (3,4). Tn, että järjestelmä toimii, on 0.99^2, ja komplementtisäännöllä tn että järjestelmä lakkaa toimimasta on $1 - 0.99^2 = 0.0199$.

Tapa 2. Merkitään tapahtumat A = parin (1,2) molemmat levyt hajoavat, ja B = parin (3,4) molemat levyt hajoavat. Tällöin $P(A) = P(B) = 0.1^2 = 0.01$. Yleisen yhteenlaskusäännön sekä tapahtumien A ja B riippumattomuuden perusteella

$$P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.01 + 0.01 - 0.0001 = 0.0199.$$

(c) Tilannetta voidaan ajatella toistokokeena: Hajoavien levyjen määrän on binomijakautunut parametrein $n = 4$ ja $p = 0.01$. Järjestelmä lakkaa toimimasta, jos levyjä hajoaa enemmän kuin 2, toisin sanoen 3 tai 4. Tämän todennäköisyys on

$$\binom{4}{3} \cdot 0.1^3 \cdot 0.9^1 + 0.1^4 \approx \mathbf{0.0037}.$$

Arviointiohje.

(a) Jos vain laskettu yhteen tai kerrottu $4 \cdot 0.1 = 0.4$, niin a-kohdasta 0.5 pistettä (= hiukan oikeaa ideaa, mutta ei huomioitu että levyjä voi hajota useampikin). Jos muuten oikein mutta unohdettu viimeinen komplementti (ts. vastataan kysymykseen "millä tn:llä toimii"), niin 0.5p sakko.

(b) Tavassa 2, jos on vain laskettu yhteen $P(A) + P(B) = 0.02$ eikä mitenkään huomioitu termiä $P(A \cap B)$, niin 1p sakko. Tapahtumien A ja B riippumattomuuteen ei tarvitse eksplisiittisesti vedota, jos on osannut laskea oikein.

(c) Päättelyn voi tehdä monellakin tavalla, kaikki pätevät päättelyt kelpaavat. Binomijakaumaa tai toistokoetta ei tarvitse nimeltä mainita. Jos huomioitu vain tapaus "3 levyä rikkoutuu" mutta ei tapausta "4 levyä rikkoutuu", niin 0.5p sakko.

T2 Taksimatkan pituus kilometreinä on jatkuva satunnaismuuttuja X, jonka tiheysfunktio on $f(x) = 1/10$, kun $0 \leq x \leq 10$, ja 0 muualla. Jos matka on enintään 5 kilometriä, matkan hinta on 10 euroa. Jos matka ylittää 5 kilometriä, hinta on $3X - 5$ euroa.

(a) Laske matkan pituuden odotusarvo integroimalla. **(1p)**

(b) Laske hinnan odotusarvo. **(2p)**

(c) Laske hinnan keskihajonta. **(2p)**

(d) Tehdään 100 matkaa, joiden pituudet ovat toisistaan riippumatta edellämainitun jakauman mukaiset. Laske kokonaishinnan odotusarvo ja keskihajonta. **(1p)**

Ratkaisu.

(a)

$$
\begin{aligned}
\mathbb{E}(X) &= \int_0^{10} x\, f(x)\, dx \\
&= \int_0^{10} (x/10)\, dx \\
&= \Big/_0^{10} (x^2/20) \\
&= 100/20 = \mathbf{5} \text{ kilometriä.}
\end{aligned}
$$

(b)

$$\begin{aligned}
\mathbb{E}(Y) &= \int_0^5 10\, f(x)\, dx \;+\; \int_5^{10} (3x-5)\, f(x)\, dx \\
&= \int_0^5 1\, dx \;+\; \int_5^{10} (0.3x-0.5)\, dx \\
&= \Big/_0^5 (x) \;+\; \Big/_5^{10} \big((0.15x^2-0.5x)\big) \\
&= 5 + (0.15\cdot 10^2 - 0.5\cdot 10) - (0.15\cdot 5^2 - 0.5\cdot 5) \\
&= 5 + (15-5) - (3.75-2.5) \\
&= \mathbf{13.75} \text{ euroa.}
\end{aligned}$$

(c) Käytetään kaavaa $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$. Tässä tarvittava $\mathbb{E}(Y)$ laskettiin jo b-kohdassa. Lasketaan sitten

$$\begin{aligned}
\mathbb{E}(Y^2) &= \int_0^5 10^2\, f(x)\, dx \;+\; \int_5^{10} (3x-5)^2\, f(x)\, dx \\
&= \int_0^5 (100/10)\, dx \;+\; \int_5^{10} \big((9x^2-30x+25)/10\big)\, dx \\
&= \int_0^5 10\, dx \;+\; \int_5^{10} (0.9x^2-3x+2.5)\, dx \\
&= \Big/_0^5 (10x) \;+\; \Big/_5^{10} \big((0.3x^3-1.5x^2+2.5x)\big) \\
&= 50 + (300-150+25) - (37.5-37.5+12.5) \\
&= 212.5 \text{ neliöeuroa.}
\end{aligned}$$

Tästä saadaan

$$\mathrm{Var}(Y) = 212.5 - 13.75^2 \approx 23.4375 \text{ neliöeuroa}$$

ja edelleen

$$\mathrm{SD}(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{23.4375} \approx \mathbf{4.84} \text{ euroa.}$$

(d) Merkitään matkojen hintoja $Y_1, Y_2, \ldots, Y_{100}$ ja niiden summaa S. Odotusarvon ja varianssin yhteenlaskukaavojen perusteella

$$\begin{aligned}
\mathbb{E}(S) &= \sum_{i=1}^{100} \mathbb{E}(Y_i) = \sum_{i=1}^{100} 13.75 = 100\cdot 13.75 = \mathbf{1375} \text{ euroa} \\
\mathrm{Var}(S) &= \sum_{i=1}^{100} \mathrm{Var}(Y_i) = \sum_{i=1}^{100} 23.4375 = 100\cdot 23.4375 = 2343.75 \text{ neliöeuroa} \\
\mathrm{SD}(S) &= \sqrt{\mathrm{Var}(S)} = \sqrt{2343.75} \approx \mathbf{48.4} \text{ euroa.}
\end{aligned}$$

Arviointiohje.

(a) Pelkästään laskemalla päätepisteiden keskiarvon saa 0p, koska nimenomaan tehtävänä on suorittaa integraali.

(b) 0p, jos käytetty matkan odotusarvoa 5 km ja laskettu sen mukainen hinta 10 euroa. 1p, kun on osattu muodostaa oikea integraali ja jakaa se paloihin (välit 0–5 ja 5–10). Laskuvirheistä sakkoa 0–1p tilanteen mukaan (1p jos saatu lukuarvo järjetön eli mahdollisten hintojen $[10, 25]$ ulkopuolella).

(c) Pelkästään kaavan $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2$ muistamisesta (muistilapulta) ilman laskuja ei saa pisteitä. Jos menetelmä on oikea, niin laskuvirheistä 0–1p sakkoa tilanteen mukaan (1p jos varianssi tai keskihajonta on saatu negatiiviseksi). Jos sotkettu keskihajonta ja varianssi: max 1p c-kohdasta.

(d) Odotusarvosta 0.5p ja varianssista 0.5p. Hyväksytään myös perustelu, jossa yksittäisen hinnan odotusarvo on kerrottu 100:lla ja keskihajonta $\sqrt{100}$:lla. Jos varianssi ja keskihajonta sotkettu, ei pisteitä kyseisestä laskusta. Samoin jos keskihajonnat on laskettu yhteen tai kerrottu 100:lla, ei pisteitä kyseisestä laskusta.

T3 Luennoijalla on viisi 6-sivuista noppaa, joista neljä on tavallisia (tulokset $1, \ldots, 6$ ovat yhtä todennäköiset) ja viides on painotettu siten, että tuloksen i todennäköisyys on $i/21$, kun $i = 1, \ldots, 6$. Luennoija on poiminut yhden nopista satunnaisesti ja heittänyt sitä seitsemän kertaa saaden tulosjonon $(6, 6, 6, 1, 6, 6, 5)$. Merkitään $\Theta = 1$ jos poimittu noppa on tavallinen ja $\Theta = 2$ jos se on painotettu.

(a) Laske todennäköisyys saada juuri tämä tulosjono, jos noppa oli tavallinen. **(1p)**

(b) Laske todennäköisyys saada juuri tämä tulosjono, jos noppa oli painotettu. **(1p)**

(c) Määritä parametrin Θ posteriorijakauma. **(3p)**

(d) Käyttäen saatua posteriorijakaumaa, määritä todennäköisyys sille, että ei saada yhtään kuutosta, kun poimittua noppaa heitetään vielä neljä kertaa. **(1p)**

Ratkaisu.

(a)
$$(1/6)^7 \approx 0.000003572$$

(b)
$$6 \cdot 6 \cdot 6 \cdot 1 \cdot 6 \cdot 6 \cdot 5/21^7 \approx 0.000021587$$

(c) Käytetään Bayes-laatikkoa.

Θ	priori	uskottavuus	tulo	posteriori
1	0.80	0.000003572	0.000002858	$0.398289 \approx \mathbf{0.398}$
2	0.20	0.000021587	0.000004317	$0.601711 \approx \mathbf{0.602}$

(d) Noppa on joko tavallinen tai painotettu, c-kohdassa lasketuilla todennäköisyyksillä. Jos se on tavallinen, niin neljä ei-kuutosta tulee todennäköisyydellä $(5/6)^4$. Jos se on painotettu, niin neljä ei-kuutosta tulee todennäköisyydellä $(1 - 6/21)^4 = (15/21)^4$. Osituskaavan nojalla
$$0.398 \cdot (1/6)^4 + 0.602 \cdot (15/21)^4 \approx 0.1921 + 0.1566 \approx \mathbf{0.3487}.$$

Arviointiohje.

- Jos sotkettu mukaan binomikertoimia tms. niin 0.5p. Jos kertomisen sijaan yhteenlaskettu, niin 0.5p.
- Jos sotkettu mukaan binomikertoimia tms. niin 0.5p. Jos kertomisen sijaan yhteenlaskettu, niin 0.5p.
- Selkeä Bayes-laatikko riittää perusteluksi, ei tarvita erikseen kaavoja (mutta niilläkin voi laskea). Jos priori unohtui tai väärin (esim. 1/2 ja 1/2): 1p sakko. Jos normalisointi unohtui tai väärin tehty: 1p sakko.

- Jos käytetty osituskaavaa, mutta laskettu vain 2 tapausta erikseen eikä laskettu yhteen = 0.5p. Jos laskettu vain toinen tapauksista, tai muuta täysin väärää = 0p. Jos laskettu yhden kuutosen tn ja korotettu neljänteen potenssiin, tähän tapaan niin 0.5p (osattu kuitenkin osituskaava, vaikka potenssiin korotus väärässä kohdassa).

T4 Eräässä fysikaalisessa kokeessa hiukkaset hajoavat satunnaisesti siten, että kunkin kahden peräkkäisen hajoamisen väliaika on jatkuvasti jakautunut tiheysfunktiolla $f(x) = \lambda e^{-\lambda x}$, kun $x > 0$. Tässä $\lambda > 0$ on tuntematon parametri. Kolme peräkkäistä havaittua väliaikaa olivat millisekunteina $\vec{x} = (2.5, 3.4, 1.1)$.

(a) Johda λ:n uskottavuusfunktio havaintojen perusteella. Laske sen perusteella λ:n suurimman uskottavuuden estimaatti. Perustele huolellisesti, että kyseessä on uskottavuusfunktion maksimikohta. **(3p)**

(b) Parametria pidetään jatkuvana satunnaismuuttujana Λ, jonka priorijakauman tiheys on $f_\Lambda(\lambda) = \lambda e^{-\lambda}$, kun $\lambda > 0$. Johda Λ:n posteriorijakauman tiheysfunktio (voit jättää normalisointivakion ratkaisematta). Laske posteriorijakauman moodi eli se λ:n arvo, jossa posterioritiheys on suurin. **(3p)**

Ratkaisu.

(a) Uskottavuusfunktio on

$$
\begin{aligned}
L(\lambda) &= f(2.5) \cdot f(3.4) \cdot f(1.1) \\
&= \lambda e^{-\lambda 2.5} \cdot \lambda e^{-\lambda 3.4} \cdot \lambda e^{-\lambda 1.1} \\
&= \lambda^3 e^{-\lambda(2.5+3.4+1.1)} \\
&= \lambda^3 e^{-7\lambda}.
\end{aligned}
$$

Tutkitaan sen logaritmia

$$\ell(\lambda) = \log(L(\lambda)) = 3\log(\lambda) - 7\lambda.$$

Sen derivaatta on

$$\ell'(\lambda) = 3/\lambda - 7,$$

jonka ainoa nollakohta on $3/7$.

Todetaan vielä, että toinen derivaatta

$$\ell''(\lambda) = -3/\lambda^2$$

on negatiivinen (kaikilla $\lambda > 0$ mutta erityisesti pisteessä $3/7$), joten kyseinen piste on ainakin paikallinen maksimikohta.

Päätepisteitä ei tarvitse tutkia, koska λ:n mahdolliset arvot ovat *avoimella* välillä $]0, \infty[$. Log-uskottavuusfunktion ainoa mahdollinen maksimikohta on $\mathbf{3/7 \approx 0.43}$, joka on siis myös uskottavuusfunktion maksimikohta.

(b) Posterioritiheys (jossa normalisointivakiota on merkitty c:llä) on

$$\begin{aligned} f_{\Lambda|\vec{x}} &= c \cdot f_\Lambda(\lambda) \cdot f(2.5) \cdot f(3.4)\ dot f(1.1 \\ &= c \cdot \lambda e^{-\lambda} \cdot \lambda e^{-\lambda 2.5} \cdot \lambda e^{-\lambda 3.4} \cdot \lambda e^{-\lambda 1.1} \\ &= c \cdot \lambda^4 e^{-8\lambda}. \end{aligned}$$

Kuten a-kohdassa, otetaan funktiosta logaritmi

$$\log(c) + 4\log(\lambda) - 8\lambda$$

ja tutkitaan sen derivaattaa

$$4/\lambda - 8.$$

Derivaatan ainoa nollakohta on $\lambda = 0.5$. Koska toinen derivaatta $-4/\lambda^2$ on aina negatiivinen, niin tämä derivaatan nollakohta on funktion maksimikohta. Posteriorimoodi on siis $\lambda = \mathbf{0.5}$.

Arviointiohje.

(a) Uskottavuusfunktion oikeasta lausekkeesta saa 1–1.5p (riippuen siitä onko osattu sieventää eksponenttifunktion ominaisuuksia käyttäen mielekkääseen muotoon $\lambda^3 e^{-7\lambda}$ vai jääty matkan varrelle).

Täysiin pisteisiin vaaditaan *jokin* perustelu sille, että löydetty derivaatan nollakohta on nimenomaan maksimi, esim. toisen derivaatan negatiivisuus, tai lukiotyyliin "merkkikaavio" (1. derivaatta on positiivinen ennen nollakohtaa ja negatiivinen nollakohdan jälkeen). Jos perustelua ei ole ollenkaan (on vain etsitty derivaatan nollakohta), niin 1p sakko. Jos perustelua on yritetty mutta ei ihan onnistuttu, niin 0–0.5p sakkoa. Se, että nollakohtia on vain yksi, ei vielä riitä perusteluksi (funktiohan voisi olla U:n muotoinen eli kyseessä olisi minimi).

Sen täydellisempää perustelua ei vaadita (ei esim. tarvitse erikseen vedota funktion jatkuvuuteen ja derivoituvuuteen eikä johonkin nimeltä mainittuun ääriarvolauseeseen).

Päätepisteiden tutkimista ei tarvita mutta siitä ei myöskään sakoteta.

Maksimin voi etsiä myös suoraan uskottavuutta derivoimalla (tulon derivointisäännöllä, ilman logaritmia). Jos tämä on osattu oikein, tämäkin on sallittu ratkaisutapa, koska logaritmia ei käsketty käyttää.

(b) Posterioritiheyden oikeasta lausekkeesta saa 1–1.5p kuten a-kohdassa. Täysiin pisteisiin tarvitaan riittävä maksimoinnin perustelu kuten a-kohdassa.

Jos maksimointiperustelu oli a-kohdassa kunnossa, niin b-kohdassa sitä ei erikseen vaadita.