Laskuharjoitus 4A

15:05 » H4AT3a) Saisiko vähän selvennystä lausekkeelle f_{Y|Y}(y|y';\theta)..? Ovatko aineistot y ja y' peräti samoja vai satunnaisvektorin Y kaksi eri havaittua aineistovektoria?
16:53 » 15:05: merkitään hetkeksi W = Y ja w = y'. 1) Mitä olisi f_{Y | W}{y | w; \theta) ilmaistuna ehdollisen ptnf:n määritelmän avulla? (TN2 luku 8.1)... – PetteriP (*)
16:54 » ... 2) Miten tämä ilmaistaisiin W:n ja (Y,W):n (y)ptnf:ien avulla? 3) Lopuksi vain nimeät uudestaan? Eli lyhyesti: y ja y' ovat ennakkoon kaksi _mahdollisesti_ eri aineistoa (voivat tosin olla myös samat). – PetteriP (*)
16:58 » Kiitos vinkistä, taisin ymmärtää :)
01:39 » Pitäisikö muuten H4AT1a:ssa viittauksen aiempaan tehtävään olla harjoituksen 1B tehtävään 3 (ei tehtävään 2, kuten tehtävänannossa nyt lukee, jossa oli eri malli)?
07:24 » 01:39: kyllä, jatkon pitäisi olla tehtävään 3. Korjaan tuon vielä tänään. – PetteriP (*)
12:27 » Onko H4AT1 tuo thetan su-estimaattori nyt siis vain theta-hatun n:äs potenssi?
12:30 » vastaan itselleni. Selitys löytyikin matskuista, en aluksi huomannut :) t.12:27
13:45 » Saisiko vinkkiä H4at4?
14:01 » 13:45: Itse aloitin seuraavasti: H2BT5:n mukaan jono (x1, ..., xn) on jokin kiinteä jono. Olennaista on, että se ei riipu aineistosta. Nyt pitäisi kirjoittaa tiheysfunktio sellaisessa muodossa, että se on kahden funktion tulo, olkoon ne vaikka g ja h. g on funktio t(y):stä ja betasta (aineisto ei esiinny sen lausekkeessa suoraan, ainoastaan t(y):n välityksellä), ja h on funktio y:stä (beta ei esiinny sen lausekkeessa lainkaan). (...)
14:02 » (...) Tästä eteenpäin tarvitsee nähdäkseni seuraavia huomioita: 1) kun jono (x1, ..., xn) on tunnettu, voidaan n\mean{y} kirjoittaa t(y):n avulla viittaamatta aineistoon; 2) potenssin/eksponenttifunktion laskusäännöt.
14:05 » ... ja t(y) kannattaa siis toki poimia suoraan vanhoista vastauksista (http://wiki.helsinki.fi/pages/viewpage.action?pageId=197657276&preview=/197657276/212096725/ratk2b.pdf), ja ehkä vähimmällä kirjoittamisella selviää käyttämällä siitä muotoa n\mean{y}/n\mean{x} eikä summamerkintöjä.
14:09 » vau :) suuret kiitokset 14:01-14:05, eiköhän se näin hyvin avuin lähde selviämään
14:19 » ja selvisikin :) Tuhannet kiitokset vielä! :)
15:48 » 12:27: kuten 12:30 huomasitkin, tuo yläindeksissä ollut (n) viittasi siis siihen, että aineiston koko on n. Tällöin on mukavampi puhua käytöksestä suurilla n, kun estimaattorin riippuvuus luvusta n on ylipäätään näkyvissä. – PetteriP (*)
15:53 » ... mainiota että löysit selityksen materiaalista :) Ja hyvä että kysyit, sillä moni muukin saattaa merkintää pohtia. :) – PetteriP (*)
15:53 » 13:35: saitkin mainioita vertaistukea, kiitos valtavasti 14:01-14:05 :) – PetteriP (*)

16:27 » Saisiko H4AT1 jotain vinkkiä tuon fisherin laskentaan? Pitääkö siinä laskea theta^2 odotusarvo integroimalla?
16:29 » Theta ei ole satunnainen, vaan aineisto on
16:53 » 16:27: Kuten 16:29 sanoi, theta ei odotusarvolausekkeen E_{theta} [-l''(theta; Y)] sisällä ole satunnaismuuttujan roolissa. Se on tuntematon, mutta kiinteä parametri.
16:55 » Itse yritän tulkita Fisherin informaatiota i(theta) = E_{theta} [-l''(theta; Y)] päässäni suunnilleen seuraavasti: 1) "-l''(theta; Y)" on jokin satunnaisvektorista Y saatu muunnos. Olemme saaneet sille lausekkeen laskemalla logaritmisen uskottavuusfunktion toisen derivaatan. Nyt vain korvaamme lausekkeessa aineiston y sitä vastaavalla satunnaisvektorilla Y ja sijoitamme lausekkeen odotusarvon sisään. (...)
16:58 » 2) Sitten laskemme odotusarvoa tälle muunnokselle olettaen, että Y:n jakaumassa parametrina on jokin "tuntematon mutta kiinteä" luku theta, joka kuuluu parametriavaruuteen. Tavallaan "annamme thetan juosta" koko parametriavaruuden läpi. Ja esimerkiksi millä tahansa theta > 0 vakiosatunnaismuuttujan theta * 5 odotusarvo on theta * 5.

17:02 » Ja tällaisia laskuja on tehty ainakin tehtävissä H3A1, H2B1, H2A2, joten niiden ratkaisuehdotuksista voi katsoa myös mallia. :)
17:06 » Nyt pääsin kärryille! Erehdyin ajattelemaan, että theta olis satunnaismuuttuja odotusarvoa laskiessa. Kiitos avusta!

17:10 » miten H4 tehtävässä 5 pitäisi hyödyntää faktorointikriteeriä? tasajakauman tiheysfunktio ei taida lainkaan riippua aineistosta...
17:13 » 17:10: Tasajakauman tiheysfunktiota määriteltäessä usein unohtuu, että se ei ole (sama) vakio kaikkialla. :) Kokeile kirjoittaa tiheysfunktio indikaattorifunktion avulla, niin huomaat, että ehkä tiheysfunktio f(y; theta) sittenkin riippuu myös y:stä jotenkin. :) Indikaattorifunktion k(y, theta) saattaa sitten pystyä jakamaan hauskasti palasiin k1 ja k2, joista k1 riippuu vain aineistosta (ei parametrista) ja k2 parametrista + aineistosta vain tunnusluvun kautta.
17:14 » ...selventävä lisävihje: mikä olisi välin (0, 1) tasajakauman tiheysfunktion f(y) arvo pisteissä y = -3 tai y = 100?
22:18 » 16:27 (ja varmaan 17:06): eli kuten 16:29-17:02 mainiosti neuvoikin, niin theta ei ole satunnaismuuttuja. Kun Fisherin informaatiota lasketaan, yritämme laskea odotusarvon $E_\theta(-l''(\theta; Y))$ missä aineisto Y on satunnaisvektori, eli jos $g(y) = -l''(\theta; y)$, niin laskemme sv:n Y muunnoksen g(Y) odotusarvoa, eli $E_\theta\, g(Y)$:tä. – PetteriP (*)
22:22 » 17:10: Kuten 17:13 jo totesi, kannattaa miettiä indikaattorifunktioita. Mieti myös, miten seuraava väite voidaan esittää $y_{(n)}$:n avulla: $y_i < \theta$ kaikilla i
22:25 » Oliko H4A5 tarkoitus saada siis sellainen tilanne, jossa sekä funktiossa h että funktiossa g esiintyy indikaattorifunktiota?
22:25 » 17:10: tuo 17:13:n vihje on tärkeä, sillä vaikka tasajakauman Tas(a,b) tiheysfunktio on vakio, niin se tarkoittaa että $f(y) = 1/(b-a) * 1\{a < y < b\}$ (tai $1\{a \leqq y \leqq b\}$). Aivan kuten 17:13 vihjasikin, niin aineistoriippuvuus näkyy nimenomaan näiden indikaattorifunktioiden kautta. – PetteriP (*)
22:27 » ... ja aivan kuten 22:22 mainitsi, kannattaa miettiä, miten $\{ y_i < \theta$ kaikilla i $\}$ voidaan esittää $y_{(n)} = \max(y_1,...,y_n)$:n avulla. Tällöin ollaan jo aika "lähellä" :) – PetteriP (*)
22:32 » 22:25: funktio h on vallan yksinkertainen ja funktiossa g esiintyy indikaattorifunktio. :) – PetteriP (*)
22:34 » 22:32: Kiitos.
09:17 » 22:25 (ja 22:34): tarkentaisin vastaustani 22:32. Tuo "vallan yksinkertainen" tarkoittaa myös indikaattorifunktiota. Ajattelin kirjoittaaa "vallan yksinkertainen indikaattori", mutta hups... – PetteriP (*)

16:24 » Millaisia malleja lauseen 3.6.5 säännöllisyysehdot käytännössä sulkevat pois? Mitä tuo "riittävien säännöllisyysehtojen vallitessa" itse asiassa tarkoittaa? Kaikki kohdan 2.5.2 ehdot vai ehkä vähemmän?
22:59 » 16:32: tätä kysymystä tarkastellaan tarkemmin Tilastollinen päättely III -kurssilla :) mutta oletusten 2.5.2. lisäksi tarvitaan riittävästi oletuksia, että su-estimaattori on ainakin tarkentuva, ja sitten todistustavasta riippuen erilaisia (heikompia/vahvempia) oletuksia informaation sekä pistemäärän suppenemisesta... – PetteriP (*)
22:59 » ... Eräitä (ei kovin tarkkoja) ehtoja voi "haistella" katsomalla todistusluonnoksia, mistä seurais oletus että $l'''(...)$ on mukavasti "hallittavissa", mutta tarkemmin väliarvolausetta tarkastelemalla tämän voi tehdä heikommilla oletuksilla... – PetteriP (*)
23:00 » ... mutta säännöllisen mallin 2.5.2. oletukset tarvitaan kyllä kaikki (koska näitä käytetään apuna pistemäärän jakaumasuppenemisessa apulauseen kautta), joten esim. havainnot tasajakaumasta eivät toteuta näitä ehtoja. – PetteriP (*)
23:05 » ... siis riippumattomat havainnot. Lisäksi oletimme koko ajan että havainnot ovat samoin jakautuneita ja riippumattomia :) – PetteriP (*)

13:05 » Vielä tuohon H4AT5:seen varmistaisin, vaikka tuosta ollaankin puhuttu paljon alla joku kirjoitti, että " Indikaattorifunktion k(y, theta) saattaa sitten pystyä jakamaan hauskasti palasiin k1 ja k2, joista k1 riippuu vain aineistosta (ei parametrista) ja k2 parametrista + aineistosta vain tunnusluvun kautta." eli tämähän on se homman "ydin". Tehtävän on jo ratkaissut niin moni, että voisin kysäistä, että eikös sen funktion, joka ei riipu parametrista voi valita kahdella eri tapaa?
13:06 » tai no, ainakin kahdella tapaa olisi ehkä oikeammin sanottu
13:28 » en ymmärrä pitääkö h4at5:sessa käyttää sitä tiheysfunktiota, joka selvitettiin derivoimalla h3bt4:sessa vai sitä tiheysfunktiota, joka on kurssimateriaalissa, eli f_Y(y;theta)= 1/theta^n ?
13:34 » 13:28: H3BT4:ssä laskettiin tiheysfunktio SU-estimaattorille hattu theta, joka on sm:n Y muunnos. H4AT5:ssä tarkastellaan Y:n jakaumaa eli sen tiheysfunktiota. Huomasithan kuitenkin, että tiheysfunktio on f_Y(y; theta) = 0, kun y < 0 tai y > theta? Tällä on tehtävän kannalta olennainen merkitys. :)
13:37 » 13:05: Ihan suoraan määritelmää ei voi soveltaa. SU-estimaatti tarkoittaa pistettä, jossa uskottavuusfunktio saa suurimman arvonsa. SU-estimaatti on siis eräs aineistosta laskettu tunnusluku. Tyhjentävyys taas puhuu ehdollisesta jakaumasta ja parametrista. Tehtävä ei ole pitkä, mutta jotenkin sinun pitäisi kytkeä tyhjentävyys ja uskottavuusfunktio toisiinsa. :) Tässä kannattaa miettiä, olisiko määritelmän lisäksi käytettävissä jokin lause, jolla kytkeminen onnistuisi.
13:41 » 13:34 Kiitos! :) ja juu, huomasin toki.Laiskuuttani en sitä kirjoittanut. Ratkaisin ensin tehtävän tuon H3BT4:sen tf:lle ja sain ihan "tolkullisen vastauksen" siihenkin, mutta sitten en jotenkin ollutkaan enää ihan perillä, että mitä tässä haettiinkaan :D
13:41 » Eli sinänsä tuo h4at5 on varsin simppeli kun huomaa jujun
15:22 » 13:05,13:06: tehtävään ei ole täysin _yksikäsitteistä_ ratkaisua :) – PetteriP (*)
15:22 » 13:41: mainiota :) – PetteriP (*)
15:34 » 13:05 kuten 13:37 sanoikin, niin että theta-hattu on jokin tunnusluku aineistosta riippuva a(y) (tämäkin on osa tehtävänannon oletusta...) Nyt pitäisi päätellä, että itse asiassa löytyy jokin kuvaus b, jolle a(y) = b(t(y)). Tarkkaa lauseketta ei tarvitse vain päätelmä, että sellainen löytyy (ja tähän tarvitset jotain lausetta). Määritelmä tosiaankin puhuu vain ehdollisista jakaumista jotain hieman järkeilyä tarvitaan mukaan :) – PetteriP (*)

22:30 » Vinkkiä H4AT3?
22:57 » 22:30: Paras vinkki riippuu varmaan siitä, mihin asti olet päässyt. :) Oletko esim. päässyt ehdollisen todennäköisyyden määritelmän avulla muotoon P(Y = ... ja ...)/P(Y = ...) asti?
22:59 » ...Näemmä tämä keskustelu, josta voi olla apua tuohon vaiheeseen pääsemiseksi, oli jo hautautunut: http://wiki.helsinki.fi/pages/viewpage.action?pageId=197657276&preview=/197657276/212104296/TP2-H4A-presemo.pdf
00:02 » En nyt taida aivan ymmärtää merkintöjä... Liittyykö theta laskun alussa Y.n havaittuun arvoon y' mutta ei y ja pitäisikö se sitten osata pudottaa pois?
02:54 » 00:02: Kannattaa ehkä miettiä TN2-monisteen määritelmästä muotoa f_{Y|Y}(y | y') = P(Y = y, Y = y')/P(Y = y), ks. http://wiki.helsinki.fi/download/attachments/135735948/tn13osa2.pdf?version=1&modificationDate=1414656073419&api=v2 s. 106 kaava (8.1). (...)
02:56 » (...) Jos haluat laittaa thetan tuonne todennäköisyyteen jotenkin, sen voi laittaa P:n alaindeksiin. Ja sen voi lukea "todennäköisyys sille, että Y saa arvon y ja Y saa arvon y', kun Y:n jakauman parametrina on theta". Theta on tässä jokin mielivaltainen, parametriavaruuteen kuuluva luku. Esimerkiksi Bernoulli-jakaumalla B(p), jos parametria ei ole muutoin rajattu, p olisi jokin mielivaltainen luku väliltä 0 < p < 1. Halutaan, että päätelmät pätevät riippumatta p:n todellisesta arvosta.
03:00 » (...) Ja olennainen kysymys tässä on miettiä, mitä tarkoittaa osoittajassa "P(Y = y ja Y = y')". Jos sm Y määriteltäisiin nopanheiton silmälukuna, mikä olisi P(Y = a ja Y = b), kun a = 1 ja b = 2? Entä, kun a = 1 ja b = 1? Löytyisikö tässä jokin säännönmukaisuus, jota voisi kuvata vaikkapa indikaattorifunktiolla?

07:59 » 22:30: kuten 22:59 mainitsi (ja laittoi mainion linkin) on keskustelu "hautautunut", mutta onneksi olin sen jo siirtänyt pysyväistalteen eli kurssisivun "Presemo-keskustelua Harjoituksesta 4A" kohdan täsmäkysymykseen. – PetteriP (*)

08:05 » 00:02: aivan kuten 02:54-03:00 kertookin, niin f_{Y|Y}(y | y'; θ) = P(Y = y, Y = y'; θ) / P(Y = y'; θ). Nyt tarkastelet tuota tapahtumaa {Y = y, Y = y'} eri y:llä ja y':llä (mistä loistavan esimerkin 03:00 antaa)... – PetteriP (*)

08:08 » ... kun lasket näillä a=1,b=2 ja a=1,b=1 esimerkeillä tämän läpi, niin tarkastele kysymystä: "riippuuko saatu ehd. yptnf parametristä θ?" Ja voisiko tämän huomion yleistää muille valinnoille indikaattorifunktion avulla :) – PetteriP (*)