HY / Matematiikan ja tilastotieteen laitos
Tilastollinen päättely II, kevät 2017
14.3.2017 – Ratkaisuehdotuksia

1. Olkoon θ positiivinen parametri, ja asetetaan

$$f(y;\theta) = \begin{cases} 2\theta^{-1} y \exp(-y^2/\theta), & \text{kun } \; y > 0 \\ 0, & \text{muuten} \end{cases}$$

Oletetaan, että $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat kukin yllä mainittua jakaumaa. Muodosta tämän mallin uskottavuusfunktio sekä määritä suurimman uskottavuuden estimaatti $\widehat{\theta}$.

Ratkaisu: Huomautus! Alla oleva ehdotus on hyvin seikkaperäinen kuvaus jokaisesta askeleesta. En luonnollisestikaan esitä, että tämä olisi *mallivastaus* vaan ainoastaan vastaus, jossa pyrin kirjoittamaan jokaisen (tehtävän kannalta merkityksellisen) kohdan kirjoitettamaan näkyviin. Vaikenen muista kohdista :)

Tehtävänä on on siis selvittää

a) mallin uskottavuusfunktio $L(\theta; \mathbf{y}) = c(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}; \theta)$ sekä
b) suurimman uskottavuuden *estimaatti* $\widehat{\theta}(\mathbf{y})$ eli su-estimaatin.

Jälkimmäinen on uskottavuusfunktion tai log-uskottavuusfunktion $\ell(\theta; \mathbf{y})$ maksimikohta, joten tarvitsemme ensin uskottavuusfunktion L ja siten tilastollisen mallin määräävän yhteistiheysfunktion $f_{\mathbf{Y}}$.

Voimme lähestyä tehtävää seuraavalla tavalla:

c) Määräämme tilastollisen mallin yhteistiheysfunktion $f_{\mathbf{Y}}$
d) Määräämme uskottavuusfunktion $L(\theta; \mathbf{y})$
e) Määräämme log-uskottavuusfunktion $\ell(\theta; \mathbf{y})$
f) Tarkastelemalla derivaattaa $\ell'(\theta; \mathbf{y})$ ja sen käytöstä etsimme maksimikohdan eli su-estimaatin.

Nyt riippumattomuuden ja potenssin laskusääntöjen nojalla

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}; \theta) &= \prod_{i=1}^{n} f(y_i; \theta) = \prod_{i=1}^{n} \Big(2\theta^{-1} y_i \exp(-y_i^2/\theta) \mathbf{1}\{\, y_i > 0 \,\} \\ &= 2^n \theta^{-n} \Big(\prod_{i=1}^{n} y_i\Big) \exp(-\sum_{i=1}^{n} y_i^2/\theta) \mathbf{1}\{\, y_i > 0 \text{ kun } i = 1, \ldots, n \,\} \end{aligned}$$

joten olemme määränneet yhteistiheysfunktion $f_{\mathbf{Y}}$.

Erään uskottavuusfunktion saamme tästä jättämällä pelkästään aineistosta riippuvat tekijät (sekä myös vakiot) pois, joten päättelemme, että

$$\begin{aligned} L(\theta;\ \mathbf{y}) &= \theta^{-n} \exp(-\sum_{i=1}^{n} y_i^2/\theta) \mathbf{1}\{\, y_i > 0 \text{ kun } i = 1, \ldots, n \,\} \\ &= \theta^{-n} \exp(-t(\mathbf{y})/\theta) \mathbf{1}\{\, y_i > 0 \text{ kun } i = 1, \ldots, n \,\} \end{aligned}$$

kun merkintöjen lyhentämiseksi esittelemme tunnusluvun $t(\mathbf{y}) = \sum_i y_i^2$. Olemme nyt määränneet pyydetyn mallin uskottavuusfunktion. Koska voimme olettaa vallan hyvin, että aineisto on järkevä, niin voimme ihan hyvillä mielin myös unohtaa aineistosta riippuvan indikaattorifunktion ja vain lyhyesti kirjoittaa uskottavuusfunktion olevan

$$L(\theta;\ \mathbf{y}) = \theta^{-n} \exp(-t(\mathbf{y})/\theta)$$

Jatkamme kohti su-estimaattia. Jos aineisto $\mathbf{y}$ ei ole järkevä (eli jollakin i on havainto $y_i \leq 0$), niin su-estimaattia ei ole yksikäsitteisenä olemassa vaan kaikki parametrin arvot ovat yhtä uskottavia. Joten oletamme jatkossa, että aineisto on järkevä eli oletamme $y_i > 0$ jokaisella i, joten indikaattoriosan voi unohtaa myös. Määräämällä log-uskottavuusfunktion $\ell(\theta; \mathbf{y})$ ja suoraan laskemalla

$$\ell(\theta;\ \mathbf{y}) = \log L(\theta; \mathbf{y}) = -n \log \theta - t(\mathbf{y})/\theta.$$

Nyt voimmekin lähteä hakemaan log-uskottavuusfunktion maksimia ja derivoidaan siksi parametrin θ suhteen

$$\ell'(\theta) = -n/\theta + t(\mathbf{y})/\theta^2.$$

Maksimit ja minimit löydämme etsimällä nollakohdat, eli laskemalla

$$\ell'(\theta) = -n/\theta + t(\mathbf{y})/\theta^2 = 0 \iff -n\theta + t(\mathbf{y}) = 0 \iff \theta = t(\mathbf{y})/n.$$

Löysimme yhden nollakohdan, joka voisi siis olla etsitty log-uskottavuusfunktion maksimikohta. Tämän voimme varmistaa kahdella tapaa: tarkastelemalla derivaatan merkkiä tai toisen derivaatan avulla. Derivaatan merkin tarkastelu hieman yleistää nollakohdan hakemista

$$\ell'(\theta) \geq 0 \iff -n/\theta + t(\mathbf{y})/\theta^2 \geq 0 \iff -n\theta + t(\mathbf{y}) \geq 0 \iff \theta \leq t(\mathbf{y})/n.$$

Havaitsemme siis, että log-uskottavuusfunktio ℓ kasvaa, kun $\theta \leq t(\mathbf{y})/n$ ja vähenee, kun $\theta \geq t(\mathbf{y})/n$, joten

$$\theta = \widehat{\theta}(\mathbf{y}) = t(\mathbf{y})/n = \sum_{i=1}^{n} y_i^2/n$$

on haettu su-estimaatti.

Olisimme voineet päätellä kohdan $t(\mathbf{y})/n$ maksimikohdaksi tarkastelemalla toista derivaatan merkkiä tässä kohdassa. Nyt

$$\ell''(\theta) = n/\theta^2 - 2t(\mathbf{y})/\theta^3 = \theta^{-3}(n\theta - 2t(\mathbf{y}))$$

joten

$$\ell''(t(\mathbf{y})/n) = (t(\mathbf{y})/n)^{-3}\Big(nt(\mathbf{y})/n - 2t(\mathbf{y})\Big) = t(\mathbf{y})^{-3}n^3 \cdot (-t(\mathbf{y})) = -t(\mathbf{y})^{-2}n^3 < 0,$$

mikä kertoo luentojen mukaan, että $t(\mathbf{y})/n$ on maksimikohta, minkä tosin tiesimmekin jo.

2. Olkoon $(t_1, t_2, t_3) = (1, 2, 3)$, olkoon Y_1, Y_2 ja Y_3 kolme riippumatonta eksponenttijakautunutta satunnaismuuttujaa ja $Y_i \sim \text{Exp}(t_i/\mu)$ missä $\mu > 0$. Näytä, että estimaattori

$$T = \frac{Y_1 + 2Y_2 + 3Y_3}{3}$$

on parametrin μ harhaton estimaattori. Onko T täystehokas estimaattori? (Perustele vastauksesi).

Ratkaisu: Huomautus! Alla oleva ehdotus on hyvin seikkaperäinen kuvaus jokaisesta askeleesta. En luonnollisestikaan esitä, että tämä olisi *mallivastaus* vaan ainoastaan vastaus, jossa jokainen (tarvittava) kohta on kirjoitettu näkyviin.

Taas tehtävässä on kaksi osaa:

a) tulisi näyttää estimaattorin harhattomuus ja

b) tulisi selvittää, onko estimaattori täystehokas

Määritelmän mukaan estimaattori T on parametrin μ harhaton estimaattori, jos

$$\mathbb{E}_\mu T = \mu \text{ jokaisella } \ \mu > 0.$$

Tämä seuraa mukavasti laskemalla, sillä odotusarvon lineaarisuuden nojalla

$$\mathbb{E}_\mu T = \mathbb{E}_\mu \frac{Y_1 + 2Y_2 + 3Y_3}{3} = \frac{\mathbb{E}_\mu Y_1 + 2\mathbb{E}_\mu Y_2 + 3\mathbb{E}_\mu Y_3}{3}$$

Koska tehtävänannon mukaan $Y_i \sim \mathrm{Exp}(t_i/\mu)$, niin $\mathbb{E}_\mu Y_i = \mu/t_i$, joten

$$\mathbb{E}_\mu T = \frac{\mu + 2\mu/2 + 3\mu/3}{3} = \frac{\mu + \mu + \mu}{3} = \mu$$

eli T on harhaton (eli näytimme kohdan a)).

Kohta b) vaatii hieman lisää. Ensiksi tiedämme, että informaatioepäyhtälön nojalla

$$\operatorname{var}_\mu U \geq \frac{1}{i(\mu)}$$

on voimassa kaikilla parametrin μ harhattomilla estimaattoreilla ja estimaattori U on täystehokas, jos epäyhtälön alaraja tavoitetaan, eli jos

$$\operatorname{var}_\mu U = \frac{1}{i(\mu)}.$$

Toinen tapa ajatella on että estimaattori on täystehokas, jos sen teho on 100%, kun teho määritellään osamääränä

$$\operatorname{teho} U = \frac{1/i(\mu)}{\operatorname{var}_\mu U} \cdot 100\%.$$

Riippumatta ajattelutavasta, on meidän siis selvittävä:

c) mikä on $\operatorname{var}_\mu T$?

d) mikä on $i(\mu)$?

e) ja onko $\operatorname{var}_\mu T = 1/i(\mu)$?

Aloitetaan kohdasta c). Nyt riippumattomuuden ja varianssin laskusääntöjen mukaan

$$\begin{aligned}\operatorname{var}_\mu T &= \operatorname{var}_\mu \Big(\tfrac{1}{3}(Y_1 + 2Y_2 + 3Y_3)\Big) = \tfrac{1}{9}\Big(\operatorname{var}_\mu Y_1 + \operatorname{var}_\mu(2Y_2) + \operatorname{var}_\mu(3Y_3)\Big) \\ &= \tfrac{1}{9}\Big(\operatorname{var}_\mu Y_1 + 4\operatorname{var}_\mu Y_2 + 9\operatorname{var}_\mu Y_3\Big).\end{aligned}$$

Koska tehtävänannon mukaan $Y_i \sim \mathrm{Exp}(t_i/\mu)$, niin $\mathrm{var}_\mu Y_i = \mu^2/t_i^2$, joten

$$\begin{aligned}\mathrm{var}_\mu T &= \tfrac{1}{9}\Big(\mathrm{var}_\mu Y_1 + 4\,\mathrm{var}_\mu Y_2 + 9\,\mathrm{var}_\mu Y_3\Big) \\ &= \tfrac{1}{9}\Big(\mu^2 + 4\mu^2/2^2 + 9\mu^2/3^2\Big) = \tfrac{1}{9}(\mu^2+\mu^2+\mu^2) = \frac{\mu^2}{3}.\end{aligned}$$

Olemme siis selvittäneet kysymyksen c).

Kohdan d) selvittämiseen tarvitsemme Fisherin informaation (sekä aineistosta havaitun informaation) määritelmää

$$i(\mu) = \mathbb{E}_\mu j(\mu;\mathbf{Y}) = \mathbb{E}_\mu(-\ell''(\mu;\mathbf{Y}))$$

Tarvitsemme siis vielä tilastollisen mallin $f_\mathbf{Y}$, jotta voimme selvittää log-uskottavuusfunktion ℓ. Nyt tehtävänannon mukaan

$$\begin{aligned}f_\mathbf{Y}(\mathbf{y};\mu) &= \prod_{i=1}^{3} f_{Y_i}(y_i;\mu) = \prod_{i=1}^{3} \frac{t_i}{\mu}\exp(-t_i y_i/\mu)\mathbf{1}\{\, y_i > 0\,\} \\ &= c(\mathbf{y})\mu^{-3}\exp(-\sum_{i=1}^{3} t_i y_i/\mu)\end{aligned}$$

missä $c(\mathbf{y})$ riippuu vain aineistosta. Voimme siten valita uskottavuusfunktioksi

$$L(\mu;\mathbf{y}) = \mu^{-3}\exp(-w(\mathbf{y})/\mu)$$

ja log-uskottavuusfunktioksi $\ell(\mu;\mathbf{y})$

$$\ell(\mu;\mathbf{y}) = -3\log\mu - w(\mathbf{y})/\mu,$$

missä tunnusluku $w(\mathbf{y}) = \sum_{i=1}^{3} t_i y_i$. Derivoimalla kahdesti parametrin μ suhteen näemme, että

$$\begin{aligned}\ell'(\mu;\mathbf{y}) &= -3/\mu + w(\mathbf{y})/\mu^2, \quad \text{ja} \\ \ell''(\mu;\mathbf{y}) &= 3/\mu^2 - 2w(\mathbf{y})/\mu^3.\end{aligned}$$

joten Fisherin informaatio on siten

$$\begin{aligned}i(\mu) &= \mathbb{E}_\mu(-3/\mu^2 + 2w(\mathbf{Y})/\mu^3) = -3/\mu^2 + 2\mathbb{E}_\mu\ (Y_1 + 2Y_2 + 3Y_3)/\mu^3 \\ &= -3/\mu^2 + 2\cdot 3\mu/\mu^3 = 3/\mu^2\end{aligned}$$

Siispä havaitsemme, että

$$\mathrm{var}_\mu T = \frac{\mu^2}{3} = \frac{1}{i(\mu)},$$

joten estimaattori T on täystehokas.

3. Olkoon $Y_1,\ldots,Y_n \sim \mathrm{P}(\mu)$ Poisson-jakautuneita ja riippumattomia havaintoja vastaavia satunnaismuuttujia. Etsi tässä mallissa parametrille μ reaaliarvoinen tyhjentävä tunnusluku.

   Ratkaisu: Huomautus! Alla oleva ehdotus on hyvin seikkaperäinen kuvaus jokaisesta askeleesta. En luonnollisestikaan esitä, että tämä olisi *mallivastaus* vaan ainoastaan vastaus, jossa jokainen (tai lähes jokainen) kohta on kirjoitettu näkyviin.

   Määritelmän mukaan tunnusluku $T = t(\mathbf{Y})$ on parametrin μ tyhjentävä tunnusluku, jos satunnaisvektorin $\mathbf{Y}$ ehdollinen jakauma ehdolla $T = t$ ei koskaan riipu parametrista μ. Tiedämme lisäksi, että tämä on yhtäpitävää *faktorointikriteerin* kanssa, jonka mukaan tunnusluku $t(\mathbf{Y})$ on tyhjentävä tunnusluku jos ja vain jos jollakin tunnusluvulla $h(\mathbf{y})$ ja jollakin $g(t;\mu)$ voimme jakaa yhteistiheysfunktion tuloksi

$$f_{\mathbf{Y}}(\mathbf{y};\mu) = h(\mathbf{y})g(t(\mathbf{y});\mu)$$

kaikilla aineistoilla $\mathbf{y}$ ja $\mu > 0$

Tätä varten tarvitsemme tilastollisen mallin määräävän yhteispistetodennäköisyysfunktion $f_{\mathbf{Y}}$. Koska $Y_i \sim \mathrm{P}(\mu)$ ja havainnot ovat riippumattomia, niin

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y};\mu) &= \prod_{i=1}^{n} \left(e^{-\mu}\ \mu^{y_i}/y_i!\mathbf{1}\{\ y_i = 0,1,2,\dots\}\right) \\ &= e^{-n\mu}\Big(\prod_{i=1}^{n} y_i!\Big)^{-1}\mu^{\sum y_i}\mathbf{1}\{\ y_i = 0,1,2,\dots \text{ jokaisella } i\,\} \end{aligned}$$

Jos merkitsemme $t(\mathbf{y}) = \sum_{i=1}^{n} y_i$ ja yhdistämme kaikki vain aineistosta riippuvat tekijät yhteen, eli jos merkitsemme

$$h(\mathbf{y}) = \Big(\prod_{i=1}^{n} y_i!\Big)^{-1}\mathbf{1}\{\ y_i = 0,1,2,\dots \text{ jokaisella } i\,\},$$

voimme kirjoittaa edellisen muodossa

$$f_{\mathbf{Y}}(\mathbf{y};\mu) = e^{-n\mu}h(\mathbf{y})\mu^{t(\mathbf{y})} = h(\mathbf{y})g(t(\mathbf{y});\mu)$$

kun merkitsemme $g(t;\mu) = e^{-n\mu}\mu^{t}$. Olemme siten saaneet hajotettua yhteispistetodennäköisyysfunktion $f_{\mathbf{Y}}$ faktorointikriteerin mukaiseksi tuloksi, joten voimme todeta, että tunnusluku $t(\mathbf{y}) = \sum y_i$ on parametrin μ tyhjentävä tunnusluku.

Koska $t(\mathbf{y})$ on lisäksi reaaliarvoinen, olemme löytäneet erään reaaliarvoisen parametrin μ tyhjentävä tunnusluvun. Huomaamme myös, että mikä tahansa tunnusluku $\alpha t(\mathbf{y})$ jollakin vakiolla $\alpha \in \mathbb{R}$ kävisi myös vallan mainioisti, joten esimerkiksi myös otoskeskiarvo $\overline{y}$ kävisi myös.

4. Oletetaan, että havaintoja vastaavat satunnaismuuttujat $Y_1,\dots,Y_n$ ovat riippumattomia ja noudattava kukin samaa jatkuvaa jakaumaa, jonka tiheysfunktio on gammajakauman erikoistapaus

$$f(y;\theta) = \tfrac{1}{2}\theta^{-3}y^2\exp(-y/\theta)\ \mathbf{1}\{\,y > 0\,\}$$

ja θ on positiivinen parametri.

a) Määrää mallin Fisherin informaatio $i(\theta)$. (vihje: satunnaismuuttujan Y_i odotusarvo saadaan gammajakauman avulla suoraan *ilman* integrointia)

b) Halutaan testata nollahypoteesia $H_0\colon \theta = \theta_0$ kaksisuuntaista vastahypoteesia $H_1\colon \theta \neq \theta_0$ vastaan. Johda Raon pistemäärätestisuure.

Ratkaisu: Huomautus! Alla oleva ehdotus on hyvin seikkaperäinen kuvaus jokaisesta askeleesta. En luonnollisestikaan esitä, että tämä olisi *mallivastaus* vaan ainoastaan vastaus, jossa jokainen (tai lähes jokainen) kohta on kirjoitettu näkyviin.

Tehtävän a)-kohdan periaate on sama kuin tehtävässä 2, eli määrätään ensin loguskottavuusfunktio, sitten määrätään tämän avulla aineistosta havaittu informaatio ja lopuksi lasketaan sen odotusarvo eli Fisherin (odotettu) informaatio.

Koska havaintoja y_i vastaavat satunnaismuuttujat ovat riippumattomia, joten

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y};\theta) &= \prod_{i=1}^n f(y_i;\theta) = \prod_{i=1}^n \Big(\tfrac{1}{2}\theta^{-3} y_i^2 \exp(-y_i/\theta)\ \mathbf{1}\{\, y_i > 0 \,\}\Big) \\ &= c(\mathbf{y})\theta^{-3n} \exp(-\sum_{i=1}^n y_i/\theta) = c(\mathbf{y})\theta^{-3n} \exp(-n\overline{y}/\theta) \end{aligned}$$

missä $c(\mathbf{y})$ riippuu vain aineistosta ja $\overline{y}$ on otoskeskiarvo. Kunhan aineisto $\mathbf{y}$ on järkevä (eli $y_i > 0$ kullakin i), voimme todeta, että eräs uskottavuusfunktio on

$$L(\theta;\mathbf{y}) = \theta^{-3n} \exp(-n\overline{y}/\theta)$$

ja log-uskottavuusfunktio on

$$\ell(\theta;\mathbf{y}) = -3n\log\theta - n\overline{y}/\theta.$$

Derivoimalla kahdesti parametrin θ suhteen saamme lausekkeet

$$\ell'(\theta;\mathbf{y}) = -3n/\theta + n\overline{y}/\theta^2 \quad \text{ja} \quad \ell''(\theta;\mathbf{y}) = 3n/\theta^2 - 2n\overline{y}/\theta^3$$

joten aineistosta havaittu informaatio $j(\theta;\mathbf{y})$ on

$$j(\theta;\mathbf{y}) = -3n/\theta^2 + 2n\overline{y}/\theta^3.$$

Fisherin informaatioksi $i(\theta)$ saamme siten

$$i(\theta) = \mathbb{E}_\theta j(\theta;\mathbf{Y}) = \mathbb{E}_\theta\Big(-\frac{3n}{\theta^2} + \frac{2n\overline{Y}}{\theta^3}\Big) = -\frac{3n}{\theta^2} + \frac{2n}{\theta^3}\mathbb{E}_\theta\overline{Y},$$

missä ensimmäinen yhtäsuuruus oli määritelmä Fisherin informaatiolle ja viimeisessä sovelsimme vain lineaarisuutta ja tietoa $\mathbb{E}_\theta\beta = \beta$ vakioilla β. Vielä pitäisi laskea otoskeskiarvon odotusarvo

$$\mathbb{E}_\theta\ \overline{Y} = \mathbb{E}_\theta\ \Big(n^{-1}\ \sum_{i=1}^n Y_i\Big) = n^{-1}\ \sum_{i=1}^n \mathbb{E}_\theta Y_i = n^{-1}\ \sum_{i=1}^n \mathbb{E}_\theta Y_1 = \frac{n\mathbb{E}_\theta Y_1}{n} \ = \mathbb{E}_\theta Y_1$$

ja tässä laskussa sovelsimme jälleen kerran lineaarisuutta sekä tietoa, että satunnaismuuttujat Y_i ovat samoin jakautuneita, joten niiden odotusarvot ovat samoja. Vielä tulisi laskea satunnaismuuttujan Y_1 odotusarvo. Tehtävänannossa on kerrottu satunnaismuuttujan tiheysfunktio, joten voisimme käyttää määritelmää ja laskea

$$\mathbb{E}_\theta\ Y_1 = \int_0^\infty \tfrac{1}{2}\theta^{-3}y^3\exp(-y/\theta)\mathrm{d}y$$

Jos emme tuntisi ylläolevaa tiheysfunktiota paremmin, voisimme laskea tämän muuttujanvaihdolla $x = y/\theta$ ja sen jälkeen integroimalla osittain muutaman kerran. Tunnistamme kyllä ylläolevan integraalin integroitavan funktion olevan vakiokerrointa vaille gammajakautuneen satunnaismuuttujan tiheysfunktioksi, joten paljon parempi tapa olisi integroida tilastotieteilijän tapaan, eli etsiä sopiva vakio $c > 0$, jotta

$$\mathbb{E}_\theta\ Y_1 = c \cdot \mathbb{P}_\theta\ (Z > 0) = c.$$

missä Z on jokin gammajakautunut satunnaismuuttuja. Mutta vielä helpompi on todeta, että tehtävänannonkin mukaan $Y_1 \sim \mathrm{G}(\kappa,\lambda)$ joillakin $\kappa > 0$ ja $\lambda > 0$, joten tiedämme

$$\mathbb{E}_\theta\ Y_1 = \kappa/\lambda,$$

eli kunhan selvitämme mitkä κ ja λ ovat, on odotusarvo selvitetty. Koska tehtävänannon tiheysfunktio ja yleinen gammajakauman tiheysfunktio ovat

$$f(y;\theta) \propto y^2 \exp(-y/\theta) \quad \text{ja} \quad f(y;\kappa,\lambda) \propto y^{\kappa-1} \exp(-\lambda y)$$

joten päättelemme, että tiheysfunktiot ovat samat, jos ja vain jos $\kappa - 1 = 2$ ja $\lambda = 1/\theta$. Siispä $\kappa = 3$ ja $\lambda = 1/\theta$ eli $Y_1 \sim \mathrm{G}(3, 1/\theta)$ ja siten

$$\mathbb{E}_\theta\, Y_1 = \frac{\kappa}{\lambda} = \frac{3}{1/\theta} = 3\theta.$$

Olemme nyt saaneet Fisherin informaation viimeisen palasen laskettua, joten

$$i(\theta) = -\frac{3n}{\theta^2} + \frac{2n}{\theta^3}\mathbb{E}_\theta\overline{Y} = -\frac{3n}{\theta^2} + \frac{2n}{\theta^3}\mathbb{E}_\theta Y_1 = -\frac{3n}{\theta^2} + \frac{2n \cdot 3\theta}{\theta^3} = \frac{3n}{\theta^2}$$

eli löysimme a)-kohdassa kysytyn Fisherin informaation.

Kohdassa b) tehtävänannossa pyydettiin johtamaan Raon pistemäärätestisuure, kun haluaisimme testata yksinkertaista nollahypoteesia $H_0\colon \theta = \theta_0$, kun vastahypoteesi on kaksisuuntainen $H_1\colon \theta \neq \theta_0$. Luentomonisteen perusteella voimme soveltaa siten kumpaa vain Raon testisuuretta $u^{1/2}(\mathbf{y})$ tai $u(\mathbf{y})$. Jos käyttäisimme ensimmäistä, niin testin havaittu merkitsevyystaso laskettaisiin laskemalla

$$p = \mathbb{P}_{\theta_0}(|\, u^{1/2}(\mathbf{Y})\,| \;\geq |\, u^{1/2}(\mathbf{y})\,|),$$

kun taas käyttämällä jälkimmäistä

$$p = \mathbb{P}_{\theta_0}(u(\mathbf{Y}) \;\geq u(\mathbf{y})),$$

mutta kumpikin antaa luonnollisesti samat havaitut merkitsevyystasot.

Raon pistemäärätestisuure $u^{1/2}$ on luentomonisteen mukaan

$$u^{1/2}(\mathbf{y}) = \frac{\ell'(\theta_0;\mathbf{y})}{\sqrt{i(\theta_0)}}$$

ja $u(\mathbf{y})$ on tämän neliö (eli $u(\mathbf{y}) = (u^{1/2}(\mathbf{y}))^2$)

Näemme siis, että tarvitsemme pistemääärän $\ell'(\theta_0;\mathbf{y})$ sekä Fisherin informaation $i(\theta_0)$, mutta laskimme nämä jo aiemmin, joten

$$u^{1/2}(\mathbf{y}) = \frac{\ell'(\theta_0;\mathbf{y})}{\sqrt{i(\theta_0)}} = \frac{-3n/\theta_0 + n\overline{y}/\theta_0^2}{\sqrt{\frac{3n}{\theta_0^2}}} = \frac{n}{\theta_0^2}(-3\theta_0 + \overline{y}) \cdot \frac{\theta_0}{\sqrt{3n}} = \frac{\sqrt{n}(\overline{y} - 3\theta_0)}{\theta_0\sqrt{3}}$$

Tämä lasku vaati hieman sieventelyjä, mutta ”loppupelleissä” kyseessä oli rationaalilausekkeiden yhteenlasku ja yhteisten tekijöiden ottaminen :) Saamme samalla vaivalla myös toisen Raon pistemäärätestisuureen korottamalla lauseke neliöön, eli

$$u(\mathbf{y}) = \left(u^{1/2}(\mathbf{y})\right)^2 = \left(\frac{\sqrt{n}(\overline{y} - 3\theta_0)}{\theta_0\sqrt{3}}\right)^2 = \frac{n(\overline{y} - 3\theta_0)^2}{3\theta_0^2}$$

Koska tehtävänannossa ei tarkemmin sanottu kumpaa käyttäisimme, olisi vallan hyvin riittänyt laskea kumpi vain :)

5. a) Olkoon $Y_1, \ldots, Y_n$ riippumatton satunnaisotos jakaumasta, jonka tiheysfunktio f riippuu reaaliarvoisesta parametrista θ. Miten määritellään parametrin θ luottamusväli luottamustasolla $1 - \alpha$?

b) Oletaan, että kohdan a) tiheysfunktio f on gammajakauman $\mathrm{G}(2, \theta)$ tiheysfunktio (löytyy tehtäväpaperin takaa). Olkoon $W = (Y_1 + \cdots + Y_n) \cdot 2\theta$. Määrää muotoa $(0, a)$ oleva parametrin θ luottamusväli luottamustasolla $1-\alpha$ satunnaismuuttujan W avulla. (vihje: perustele gammajakauman ominaisuuksien avulla, että $W \sim \chi^2_{4n}$).

Ratkaisu: Tehtävän a)-kohta on teoriakysymys, joten mitään ei tarvitse laskea :) Aineistosta riippuva parametriavaruuden $\Omega \subset \mathbb{R}$ väli $A(\mathbf{y}) = (a(\mathbf{y}), b(\mathbf{y}))$ on parametrin θ luottamusväli luottamustasolla $1-\alpha$, jos

$$\mathbb{P}_\theta(\theta \in A(\mathbf{Y})) = \mathbb{P}_\theta(a(\mathbf{y}) < \theta < b(\mathbf{Y})) \geq 1 - \alpha$$

kaikilla $\theta \in \Omega$. Väli voi luonnollisesti olla avoin (kuten yllä), suljettu tai puoliavoin. Tehtävänannossa ei kysytty luottamusvälin tulkintaa, mutta toistokokeena tulkittuna luottamusväli luottamustasolla $1-\alpha$ tarkoittaa, että odotamme että keskimäärin vähintään $n \cdot (1-\alpha)$ kertaa tuntematon parametri θ on välillä $A(\mathbf{y})$, kun toistamme kokeen n kertaa.

Tehtävän b)-kohdassa kysytään luottamusväliä luottamustasolla $1-\alpha$, joka on muotoa $(0, a)$. Vihjeen mukaan $W \sim \chi^2_{4n}$ on saranasuure, joten voimme määrätä kysytyn luottamusvälin mukavasti sen avulla. Perustelemme tämän lopuksi. Koska W on saranasuure, niin

$$\mathbb{P}_\theta(b < W < c) = F_W(c) - F_W(b)$$

jokaisella θ, missä F_W on khiin neliön (vapausasteella $4n$) kertymäfunktio. Tärkein havainto on siis se, että tämä kertymäfunktio ei riipu parametrista θ. Vasen puoli voidaan puolestaan kirjoittaa toisessa muodossa

$$\mathbb{P}_\theta(b < W < c) = \mathbb{P}_\theta(b < 2\theta n\overline{Y} < c) = \mathbb{P}_\theta\Big(\frac{b}{2n\overline{Y}} < \theta < \frac{c}{2n\overline{Y}}\Big).$$

Koska haemme muotoa $(0, a)$ olevaa luottamusväliä, on luvuksi b valittava 0 eli $b = 0$. Luku c saadaan siten määräämällä

$$\mathbb{P}_\theta(0 < W < c) = F_W(c) - F_W(0) = F_W(c) \geq 1 - \alpha,$$

jonka voimme mukavasti selvittää kvantiilifunktion $q_W = F_W^{-1}$ avulla

$$c \geq q_W(1-\alpha).$$

Koska pienempi luottamusväli on parempi, valitsemme $c = q_W(1-\alpha)$, joten kysytyksi luottamusväliksi voimme lopulta valita

$$A(\mathbf{y}) = \Big(0, \frac{c}{2n\overline{y}}\Big) = \Big(0, \frac{q_W(1-\alpha)}{2n\overline{y}}\Big).$$

Jätimme perustelun sille, että $W \sim \chi^2_{4n}$ tehtävän loppuun, joten jatkamme nyt sen parissa. Vihjeen mukaan tämä pitäisi selvitä gammajakauman ominaisuuksista, joten tutkimme tehtäväpaperin mukana olleita tietoja.

Tehtäväpaperin tiedoissa kerrotaan, että gammajakaumalla on yhteenlaskuominaisuus, jonka mukaan

$$Z = Y_1 + \cdots + Y_n \sim \mathrm{G}(2 + \cdots + 2, \theta) = \mathrm{G}(2n, \theta)$$

sillä $Y_i \sim \mathrm{G}(2, \theta)$ ja satunnaismuuttujat ovat riippumattomia. Lisäksi tehtäväpaperissa kerrotaan, että jos $c > 0$ ja X on gammajakautunut, niin myös tulo cX on gammajakautunut.

Havaitsemme siten, että

$$W = 2\theta \cdot (Y_1 + \cdots + Y_n) = 2\theta Z \sim \mathrm{G}(2n, \theta/(2\theta)) = \mathrm{G}(2n, \tfrac{1}{2}).$$

Tämä kertoo nyt mukavasti, että W on saranasuure. Lisäksi tiedämme khiin neliön olevan gammajakauman erikoistapaus ja

$$\chi^2_{4n} = \mathrm{G}(4n/2, \tfrac{1}{2}) = \mathrm{G}(2n, \tfrac{1}{2}).$$

Siispä pienellä jakaumapäättelyllä näimme, että $W \sim \mathrm{G}(2n, \tfrac{1}{2}) = \chi^2_{4n}$.