Tilastollinen päättely I
3. harjoitukset, 5. vko 2004

3.1. Aineistossa Odotus (`http://mtl.uta.fi/tilasto/TILTA33_05/Aineistot/`) on annettu 300:n erään sairaalan synnytysosastolle synnyttämään saapuneen äidin odotusajat (tunteina) saapumisesta synnytyksen alkamiseen.

(a) Laske aineistosta odotusajan keskiarvo ja varianssi.

(b) Tutkitaan, noudattaako odotusaika gammajakaumaa. Oletetaan, että otoksesta lasketut keskiarvo ja varianssi ovat hyvät arviot vastaaville populaation parametreille. Estimoi gammajakauman parametrit asettamalla otoskeskiarvo ja -varianssi yhtäsuuriksi vastaavien populaation parametrien kanssa.

(c) Piirrä aineistosta histogramma ja samaa kuvioon estimoimasi gammajakauman tiheysfunktio.

3.2. Olkoon $X_1, \ldots, X_{50}$ otos binomijakaumasta $\mathrm{Bin}(5, \pi)$, missä π on tuntematon parametri. Otos on tiivistetty oheiseen taulukkoon:

x	0	1	2	3	4	5
Frekvenssi	6	10	11	13	6	1

(a) Laske π:n suurimman uskottavuuden estimaatti.

(b) Laske todennäköisyyden $P(X \geq 3)$ suurimman uskottavuuden estimaatti (Vihje: SU-estimaatin invarianssi, alaluku 7.8 ja JU alaluku 2.7).

3.3. Oletetaan, että puustotauti A on levinnyt satunnaisesti ja tasaisesti yli laajan metsäalueen ja sairastuneita puita on keskimäärin λ kappaletta hehtaaria kohden. Kymmenellä satunnaisesti valitulla neljän hehtaarin koealalla havaittujen sairaiden puiden lukumäärät olivat: $0, 1, 3, 0, 0, 2, 2, 0, 1, 1$. Määritä λ:n suurimman uskottavuuden estimaatti, kun sairastuneiden puiden lukumäärä koealalla noudattaa Poissonin jakaumaa.

3.4. Olkoon $Y_1, \ldots, Y_n$ otos normaalijakaumasta $\mathrm{N}(\mu, 1)$.

(a) Piirrä samaan kuvaan logaritmoidun uskottavuusfunktion kuvaajat seuraavissa tapauksissa:

i. $n = 20$ ja $\bar{y} = 3$

ii. $n = 40$ ja $\bar{y} = 3$

(b) Mikä parametrin μ suurimman uskottavuuden estimaatti?

(c) Määritä logaritmoidun uskottavuusfunktion 2. derivaatta ja sen arvo pisteessä $\mu = \bar{y}$.

(JU, H1.1, s. 169)

3.5. Määritä tehtävän 4 tapauksessa parametrin μ 14.7%:n uskottavuusväli, kun $n = 20$ ja $n = 40$. Määritä myös μ:n 95%:n luottamusväli, kun $n = 20$. Vertaa luottamusväliä vastaavaan uskottavuusväliin. (JU, H1.1, s. 169)

3.6. Olkoon $Y_1, \ldots, Y_n$ otos Poissonin jakaumasta Poi(μ).

(a) Näytä, että $l(\mu) = -n\mu + n\bar{y}\,\log(\mu)$ ja että

(b) potenssisarjakehitelmän katkaiseminen 2. asteen termin jälkeen johtaa likiarvoon

$$l^*(\mu) = n\bar{y}[\log(\bar{y}) - 1] - \frac{n(\mu - \bar{y})^2}{2\bar{y}}$$

3.7. Piirrä funktioiden $l(\mu)$ ja $l^*(\mu)$ kuvaajat samaan kuvioon, kun

(a) $n = 10$ ja $\bar{y} = e$

(b) $n = 10$ ja $\bar{y} = 25$

Valitse pystyasteikko niin, että käyrien huipun arvo on 0. Millä alueella likiarvo on hyvä?

3.8. Erään geneettisen teorian mukaan verityyppien MM, NM ja NN suhteelliset osuudet suuressa populaatiossa ovat $\theta^2, 2\theta(1-\theta)$ ja $(1-\theta)^2$, missä $0 \leq \theta \leq 1$ on tuntematon parametri.

(a) Valitaan populaatiosta n:n alkion otos, jossa eri verityyppien lukumäärät olivat x_1, x_2 ja x_3. Esitä $\hat{\theta}$:n lauseke.

(b) Olkoon $n = 100$ ja havaitut frekvenssit $32, 46$ ja 22. Laske $\hat{\theta}$ ja odotettujen frekvenssien estimaattit.

3.9. Heitetään n kertaa sellaista yleistettyä noppaa (suorakulmainen särmiö), että 4 tahoa ovat keskenään yhtä todennäköiset ($p_1 = p_2 = p_3 = p_4 = p$) ja kaksi muuta tahoa (vastakkaiset) ovat vastaavasti keskenään yhtä todennäköiset ($p_5 = p_6 = q$). Oletetaan, että i. taho sattuu x_i kertaa ($i = 1, 2, \ldots, 6$), missä $\sum x_i = n$.

(a) Osoita, että $\hat{\theta} = (3t - 2n)/12n$, kun $p = 1/6 + \theta$ ja $t = x_1 + x_2 + x_3 + x_4$.

(b) Oletetaan, että havaitut frekvenssit ovat $11, 15, 13, 15, 22, 24$. Laske odotettujen frekvenssien estimaatit, kun noppaa koskevat oletukset (malli) oletetaan oikeiksi.

3.10. Hatussa on N arpalippua, jotka on numeroitu peräkkäin $1, 2, \ldots, N$. Hatusta valittiin satunnaisotannalla palauttaen 8 arpalippua, joiden järjestysnumerot olivat $137, 24, 86, 33, 92, 129, 17, 111$. Osoita, että $\hat{N} = 137$.

3.11. Oletetaan, että $\hat{\theta}$ on estimaattien $\hat{\theta}_1$ ja $\hat{\theta}_2$ painotettu keskiarvo. Silloin siis

$$\hat{\theta} = a\,\hat{\theta}_1 + (1 - a)\,\hat{\theta}_2,$$

missä $0 \le a \le 1$.

(a) Osoita, että $\hat{\theta}$:n arvo on $\hat{\theta}_1$:n ja $\hat{\theta}_2$:n välissä.

(b) Olkoon

$$\hat{\theta} = a_1\,\hat{\theta}_1 + \cdots + a_n\,\hat{\theta}_n,$$

missä a_i:t ovat positiivisia ja $\sum a_i = 1$. Osoita, että $\hat{\theta}$ on pienimmän ja suurimman $\hat{\theta}_i$:n välissä.

3.12. Piirrä 3. tehtävän tapauksessa logaritminen normitettu uskottavuusfunktio ja sen perusteella 50%:n ja 10%:n uskottavuusvälit.

3.13. Määritä tehtävässä 9 (yleistetty noppa) arvon $\theta = 0$ (symmetrinen noppa) suhteellinen uskottavuus ja logaritmisen normitetun uskottavuusfunktion arvo pisteessä $\theta = 0$. Onko tämä parametrin arvo uskottava?

3.14. Kun tietyn metallin pitoisuus liuoksessa mitataan, noudattaa mittauksessa syntyvä virhe normaalijakaumaa $\mathrm{N}(0, \sigma^2)$. Jos oikea pitoisuus on μ, niin havaittu pitoisuus $X \sim \mathrm{N}(\mu, \sigma^2)$. Varianssi σ^2 tunnetaan aikaisemmasta kokemuksesta.

(a) Olkoot $x_1, x_2, \ldots, x_n$ tuntemattoman pitoisuuden μ riippumattomat mittaukset. Osoita, että $\hat{\mu} = \bar{x}$ ja logaritminen normitettu uskottavuusfunktio on

$$r(\mu) = -\frac{n}{2\sigma^2}(\bar{x} - \mu)^2, \qquad -\infty < \mu < \infty.$$

(Vihje: Osoita $\sum(x_i - \mu)^2 = \sum(x_i - \bar{x}_i)^2 + n(\bar{x}_i - \mu)^2$.)

(b) Laimennetaan alkuperäisen liuoksen pitoisuus puoleen niin, että pitoisuus on $\mu/2$. Tehdään sitten lisämittaukset $y_1, y_2, \ldots, y_m$. Määritä kaikkiin $n + m$ mittaukseen perustuva μ:n suurimman uskottavuuden estimaatti.