TILASTOMATEMATIIKKA

Harjoitus 4 ratkaisut, kevät 2018

1. Laitteen käyttämättömyysajan jakauman tiheysfunktio on

$$f(t) = \begin{cases} \theta^2 t \mathrm{e}^{-\theta t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Jakauman parametrin θ likiarvon määräämiseksi tehtiin havaintoja monen samanlaisen laitteen käyttämättömyysajoista, jotka voidaan olettaa toisistaan riippumattomiksi. Havainnoiksi saatiin $t_1, t_2, \ldots, t_n$. Määrää parametrin θ likiarvon laskukaava suurimman uskottavuuden (maximum likelihood) menetelmällä.

Ratkaisu: Muodostetaan uskottavuusfunktio

$$L(\theta) = f(t_1; \theta) \cdots f(t_n; \theta) = \theta^2 t_1 \mathrm{e}^{-\theta t_1} \cdots \theta^2 t_n \mathrm{e}^{-\theta t_n} = \theta^{2n} t_1 \cdots t_n \mathrm{e}^{-\theta(t_1 + \cdots + t_n)}$$

Derivaataksi saadaan tulon derivoimissäännöllä

$$\begin{aligned} L'(\theta) &= 2n\theta^{2n-1} t_1 \cdots t_n \mathrm{e}^{-\theta(t_1+\cdots+t_n)} - \theta^{2n} t_1 \cdots t_n (t_1 + \cdots + t_n) \mathrm{e}^{-\theta(t_1+\cdots+t_n)} \\ &= \theta^{2n-1} t_1 \cdots t_n \mathrm{e}^{-\theta(t_1+\cdots+t_n)} (2n - \theta(t_1 + \cdots + t_n)). \end{aligned}$$

Minimikohdaksi ja siten suurimman uskottavuuden estimaatiksi saadaan

$$\hat{\theta} = \frac{2n}{t_1 + \cdots + t_n} = \frac{2}{\bar{t}},$$

missä $\bar{t} = \frac{1}{n} \sum_{k=1}^{n} t_k$ on havaintojen aritmeettinen keskiarvo.

2. Järven kirjolohikannan suuruutta arvioidaan seuraavalla menetelmällä (capture-recapture method). Ensimmäisessä vaiheessa pyydetään 500 kirjolohta, jotka merkitään ja lasketaan vapaaksi. Toisessa vaiheessa pyydetään 200 kirjolohta ja lasketaan merkittyjen kirjolohien lukumäärä X. Tavallisesti oletetaan, että X noudattaa binomijakaumaa $\mathrm{Bin}(n,p)$, missä p on todennäköisyys sille, että toisessa vaiheessa pyydetty kirjolohi on merkitty.

   a) Kuinka suuri on n? Laske parametrin p arvo, jos tiedetään, että kirjolohien lukumäärä järvessä on $N = 2000$ tai $N = 5000$.

   b) Kokeessa havaittiin 40 kirjolohta. Määrää järven kirjolohikannan 95% luottamusväli.

Ratkaisu:

a) Nyt X = ”toisessa vaiheessa pyydetty kirjolohi on merkitty” $\sim \mathrm{Bin}(n,p)$, joten $n = 200$ ja p on tuntematon parametri. Koska p ilmoittaa järvessä olevien kirjolohien suhteellisen osuuden, niin p tunnetaan, jos järvessä olevien kirjolohien määrä N tunnetaan. Erityisesti, jos $N = 2000$, niin
$$p = \frac{500}{N} = \frac{500}{2000} = \frac{1}{4}.$$
Jos taas $N = 5000$, niin
$$p = \frac{500}{5000} = \frac{1}{10}.$$

b) Nyt N on tuntematon, jota pitäisi estimoida capture-recapture-menetelmällä. Estimoidaan suhteellista osuutta p estimaattorilla
$$p^* = \frac{X}{n},$$
jolle $\mathrm{E}(p^*) = p$ ja $\sigma_{p^*} = \sqrt{\frac{p(1-p)}{n}}$. Standardisoimalla saadaan testimuuttujaksi
$$Z = \frac{p^* - \mathrm{E}(p^*)}{\sigma_{p^*}} = \frac{\frac{X}{200} - p}{\sqrt{\frac{p(1-p)}{200}}} \sim N(0,1) \quad \text{likimain,}$$
sillä $n = 200$ on ”suuri”.
Määrätään kaksisuuntainen symmetrinen luottamusväli. Sitä varten määrätään normaalijakauman taulukosta luku r_0 siten, että
$$P(-r_0 < Z \leq r_0) = \Phi(r_0) - \Phi(-r_0) = 2\Phi(r_0) - 1 = 0.95 \Leftrightarrow \Phi(r_0) = 0.975 \overset{\text{taulukosta}}{\Rightarrow} r_0 = 1.96.$$
Muokataan epäyhtälö
$$-1.96 \leq Z \leq 1.96$$
muotoon
$$\frac{X}{200} - 1.96\sqrt{\frac{p(1-p)}{200}} \leq p \leq \frac{X}{200} + 1.96\sqrt{\frac{p(1-p)}{200}}.$$
Epäyhtälön ylä- ja alarajalla estimoidaan havainnoista $p = \hat{p} = \frac{40}{200} = \frac{1}{5}$ ja $\widehat{X} = 40$, josta saadaan suhteellisen osuuden p luottamusväliksi
$$I_p = [0.145, 0.255].$$
Koska toisaalta $p = \frac{500}{N}$, saadaan kirjolohien lukumäärälle N luottamusväli
$$I_N = [1957, 3459].$$

3. Lentokoneen siipipäällysteen niiteille on asetettu laatuvaatimukseksi vähintään 400 N (Newton) odotettavissa oleva vetolujuus ainakin 99 % varmuudella. Laatua testattiin aika ajoin ottamatta 20 niitin umpimähkäinen otos ja vetorasittamalla niitit poikki. Erään otoksen vetorasitustulokset olivat

$$401, 405, 390, 385, 399, 360, 378, 408, 410, 365, 400, 402, 383, 388, 389, 403, 395, 397, 402, 391.$$

Minkä rajan yläpuolella niittien odotettavissa oleva vetolujuus on mittaustulosten mukaan 99 % varmuudella? Toteutuiko niittien laatuvaatimus? Niittien vetolujuutta voidaan pitää normaalijakautuneena suureena.

Ratkaisu: Koska hajonta σ on tuntematon, käytetään t-jakautunutta testimuuttujaa

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

missä $\overline{X}$ on havaintojen aritmeettinen keskiarvo, μ on vetolujuuden odotusarvo. Olemme kiinnostuneita yksisuuntaisesta luottamusvälistä $[a, \infty[$, missä a saadaan määrättyä ehdosta

$$P(T \leq r_0) = 0.99.$$

Koska havaintoja on yhteensä 20, niin vapausasteiden lukumäärä on $n - 1 = 19$. Taulukosta saadaan

$$r_0 = 2.539.$$

Epäyhtälöstä

$$T \leq r_0$$

saadaan

$$\mu \geq \mu_l = \overline{X} - r_0 S/\sqrt{n}.$$

Estimoidaan $\overline{X}$ ja S otoksesta otoskeskiarvolla $\overline{x} = 392.55$ ja otoshajonnalla $s = 13.379$. Sijoittamalla saadut estimaatit alarajaan μ_l saadaan alarajaksi $a = \widetilde{\mu}_l \approx 385$ ja siten luottamusväliksi

$$I_\mu = [385, \infty[.$$

Koska luottamusvälin alaraja on laatuvaatimuksen alapuolella, ei laatuvaatimus toteudu tämän otoksen perusteella.

4. Tutkittiin henkilöiden ruumiinlämpöä mittaamalla 130 henkilön ruumiinlämpö. Mittausten tuloksiksi saatiin $\overline{x} = 36.805$ ja $s = 0.407$ yksikkönä [°C]. Määrää ruumiinlämmön odotusarvon 95% luottamusväli käyttämällä

   a) t-jakaumaa ja sille annettua taulukkoa.

   b) t-jakaumaa laskemalla kynnysarvo graafisella laskimella tai vaikkapa Excelillä.

   c) normaalijakaumaa, missä on arvioitu $\sigma \approx s$.

   Mitä voit näiden tietojen perusteella sanoa normaalijakauman käytöstä odotusarvon estimoinnissa suurilla n:n arvoilla?

Ratkaisu:

$$Z = \frac{\overline{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

missä $n = 130$, $\overline{x} = 36,805$ ja $s = 0,407$. Määrätään ruumiinlämmön odotusarvon 95%:n luottamusväli vaatimalla, että on voimassa

$$P(-t_p \leq Z \leq t_p) \geq 0,95$$

Tällöin saadaan odotusarvon μ luottamusväliksi laskettua

$$\begin{aligned} -t_p \leq \frac{\overline{x} - \mu}{s/\sqrt{n}} &\leq t_p \\ -t_p \cdot \frac{s}{\sqrt{n}} \leq \overline{x} - \mu &\leq +t_p \cdot \frac{s}{\sqrt{n}} \\ \overline{x} - t_p \cdot \frac{s}{\sqrt{n}} \leq \mu &\leq \overline{x} + t_p \cdot \frac{s}{\sqrt{n}} \end{aligned}$$

johon voidaan sijoittaa tehtävän arvot $n = 130$, $\overline{x} = 36,805$ ja $s = 0,407$

$$36,805 - t_p \cdot 0,035696 \leq \mu \leq 36,805 + t_p \cdot 0,035696$$

a) t-jakauman taulukkoa käyttäen. $t_p = 1,98$, joten saadaan

$$\begin{aligned} 36,805 - 0,0707 \leq \mu &\leq 36,805 + 0,0707 \\ 36,73 \leq \mu &\leq 36,88 \end{aligned}$$

b) t-jakaumalla laskemalla kynnysarvo koneella. $t_p = 1,9785$, joten saadaan

$$\begin{aligned} 36,805 - 0,0706259274 \leq \mu &\leq 36,805 + 0,0706259274 \\ 36,734 \leq \mu &\leq 36,876 \end{aligned}$$

c) normaalijakaumaa, jossa on arvioitu $\sigma \simeq s$. $t_p = 1,96$, joten saadaan

$$\begin{aligned} 36,805 - 0,06996 \leq \mu &\leq 36,805 + 0,06996 \\ 36,74 \leq \mu &\leq 36,87 \end{aligned}$$

Normaalijakaumaa voi hyvin käyttää suurilla n:n arvoilla approksimoimaan t-jakaumaa, sillä t-jakauma lähestyy normaalijakaumaa vapausasteiden kasvaessa.

5. Osakkeiden hinnan mallittamisessa käytetään kuuluisaa Black-Scholesin mallia, jonka mukaan osakkeen hinta ajanhetkellä t on satunnaismuuttuja

$$H(t) = H(0) \cdot \exp\big((b - \frac{1}{2}\sigma^2)t + \sigma W(t)\big),$$

missä $H(0)$ on osakkeen hinta ajanhetkellä $t = 0$, $b, \sigma > 0$ ovat vakioita ja $W(t) \sim N(0, t)$ on satunnaismuuttuja. Oletetaan, että $H(0) = 100$, $b = 0.1$ ja $\sigma = 0.3$ (Vertaa harjoitus 2 tehtävä 5).

a) Laske osakkeen hinnan odotusarvo ajanhetkellä $t = 1$ (**Vihje:** neliöi odotusarvossa esiintyvä eksponentti ja käytä integroinnissa sopivaa sijoitusta).

b) Määrää ajanhetkellä $t = 1$ osakkeen hinnalle väli, jolla hinta on 95% todennäköisyydellä, eli väli $[a_1, a_2]$, jolle $P(a_1 \leq H(1) \leq a_2) = 0.95$, laskemalla normaalijakauman avulla sellainen luku r_0, että $P(-r_0 \leq W(1) \leq r_0) = 0.95$.

Ratkaisu:

a) Osakkeen hinta hetkellä $t = 1$ on

$$H(1) = H(0)e^{(b-\frac{\sigma^2}{2})\cdot 1+\sigma W(1)} = H(0)e^{b-\frac{\sigma^2}{2}}e^{\sigma\cdot Z},$$

missä $Z \sim N(0, 1)$. Merkitään $c = H(0)e^{b-\frac{\sigma^2}{2}}$, jolloin

$$\begin{aligned}
E(H(1)) &= c \cdot E(e^{\sigma\cdot Z}) \\
&= c \int_{-\infty}^{\infty} e^{\sigma x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx \\
&= c \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\sigma)^2+\frac{\sigma^2}{2}} \\
&= c \cdot e^{\frac{\sigma^2}{2}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx}_{=\Phi(\infty)-\Phi(-\infty)=1} \\
&= H(0)e^{b-\frac{\sigma^2}{2}}e^{\frac{\sigma^2}{2}} = H(0)e^b = \underline{110,52}
\end{aligned}$$

b) Muokataan vaadittua ehtoa käyttäen a)-kohdan määritystä c:lle:

$$\begin{aligned}
&P(a_1 \leq H(0)e^{b-\frac{\sigma^2}{2}}e^{\sigma W(1)} \leq a_2) \\
&= P(\frac{a_1}{c} \leq e^{\sigma W(1)} \leq \frac{a_2}{c}) \\
&= P(\frac{1}{\sigma}\ln\frac{a_1}{c} \leq W(1) \leq \frac{1}{\sigma}\ln\frac{a_2}{c}) \\
&= P(-r_0 \leq W(1) \leq r_0) = 0,95
\end{aligned}$$

Koska tässä on $W(1) \sim N(0, 1)$, niin standardisoidun normaalijakauman taulukosta nähdään, että $r_0 = 1,96$, joten

$$a_1 = ce^{-\sigma r_0} \simeq 58,68 \qquad \text{ja} \qquad a_2 = ce^{\sigma r_0} \simeq 190,22$$

Siispä osakkeen hinta hetkellä $t = 1$ on 95%:n todennäköisyydellä välillä $\underline{[58{,}68,\ 190{,}22]}$.