Tilastomatematiikka

Ensimmäinen välikoe 14.02.2019

1. Tarkastellaan satunnaiskoetta, jossa heitetään kahta noppaa. Olkoot

$$
\begin{aligned}
A &= \text{”ensimmäisen nopan pisteluku on 6”}, \\
B &= \text{”toisen nopan pisteluku on 1”}, \\
C &= \text{”pistelukujen summa on 8”}
\end{aligned}
$$

satunnaiskokeeseen liittyviä tapahtumia.

a) Esitä joukko-opillisesti tapahtumat A, B ja C.

b) Laske todennäköisyydet $P(A \cup C)$, $P(A \cap C)$ ja $P(A|C)$.

2. Oletetaan, että Kärppien maalimäärä yksittäisessä SM-liigaottelussa on Poisson-jakautunut odotusarvolla 3.4 maalia muista otteluista riippumatta. Runkosarjassa pelataan 60 ottelua. Millä todennäköisyydellä Kärpät tekee runkosarjassa yli 200 maalia?

a) Anna laskentakaava, jolla todennäköisyys voidaan laskea tarkasti. Käytä riippumattomien Poisson-jakautuneiden muuttujien ominaisuutta:

Jos $X_i \sim \mathrm{Poi}(a_i), i = 1, 2, \ldots, n,$ niin $X_1+X_2+\cdots+X_n \sim \mathrm{Poi}(a_1+a_2+\cdots+a_n)$.

Itse todennäköisyyttä ei tarvitse laskea. (2p)

b) Laske kysytty todennäköisyys normaalijakauma-approksimaatiota käyttämällä. (4p)

3. Tutkittiin delaminoitumisen vaikutusta palkin ominaistaajuuteen [Hz]. Mitattiin viiden delaminoituneen palkin ominaistaajuus ja saatiin seuraavat mittaustulokset

$$230.66, 233.05, 232.58, 229.48, 232.58$$

a) Määrää ominaistaajuuden odotusarvolle sopiva piste-estimaatti. (2p)

b) Laske ominaistaajuuden odotusarvon 90% 2-suuntainen luottamusväli. Tukevatko mittaustulokset väitettä, että keskimääräinen ominaistaajuus on 235 Hz? (4p)

Kaavoja

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \backslash B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$$
$$P(\bar{A}) = 1 - P(A)$$
$$P(A|B) = P(A \cap B)/P(B)$$
$$f(x) = 1/(b-a),\ E(X) = (a+b)/2,\ \mathrm{Var}(X) = (b-a)^2/12$$
$$f(x) = \theta e^{-\theta x},\ E(X) = 1/\theta,\ \mathrm{Var}(X) = 1/\theta^2$$
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},\ E(X) = np,\ \mathrm{Var}(X) = np(1-p)$$
$$P(X = k) = p(1-p)^{k-1},\ E(X) = 1/p,\ \mathrm{Var}(X) = (1-p)/p^2$$
$$P(X = k) = \frac{a^k}{k!} e^{-a},\ E(X) = a,\ \mathrm{Var}(X) = a$$
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ (likimain, kun ”}n\text{ on suuri”)}$$
$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$
$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t_{n+m-2};$$
$$r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}};\ s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y});\ s_{xx} = s_x^2;$$
$$y = a + bx; \quad b = \frac{s_{xy}}{s_{xx}}; \quad a = \overline{y} - b\overline{x}$$
$$s_r^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \frac{n-1}{n-2}(1 - r^2) s_{yy}$$
$$\sqrt{n-1}\, s_x \frac{\frac{S_{xY}}{s_{xx}} - \beta}{S_r} \sim t_{n-2};$$
$$E(aX + bY) = aE(X) + bE(Y); \quad \mathrm{Var}(X) = E((X - E(X))^2);$$
$$\sigma_{XY} = \mathrm{Cov}(X,Y) = E((X - E(X))(Y - E(Y))), \quad \sigma_{XX} = \sigma_X^2;$$
$$= E(XY) - E(X)E(Y)$$
$$\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}.$$

Tehtävien ratkaisuperiaatteet

1. a) Satunnaiskoetta voidaan mallintaa vaikkapa järjestetyillä pareilla (i, j), missä i ilmoittaa ensimmäisen ja j toisen nopan silmäluvun. Tällöin kysytyt tapahtumat vastaavat joukkoja

$$A = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\},$$
$$B = \{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1)\},$$
$$C = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}.$$

b) Joukoksi $A \cap C$ saadaan a)-kohdasta $A \cap C = \{(6,2)\}$. Koska kunkin alkeistapahtuman todennäköisyys on $\frac{1}{36}$, niin a)-kohdan mukaan saadaan

$$P(A \cap C) = \frac{1}{36} \Rightarrow P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{6}{36} + \frac{5}{36} - \frac{1}{36} = \frac{5}{18}.$$

Ehdollisen todennäköisyyden määritelmästä taasen saadaan

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}.$$

Toinen (ja yksinkertaisempi) tapa on päätellä suoraan, että C:stä ainostaan yksi alkio $(6, 2)$ on suotuisa A:lle.

2. a) Annetun ominaisuuden perusteella Kärppien maalimäärän runkosarjassa ilmoittaa summamuuttuja

$$S_{60} = X_1 + X_2 + \cdots + X_{60} \sim \mathrm{Poi}(60 \cdot 3.4) = \mathrm{Poi}(204).$$

Kysytty todennäköisyys on siten

$$P(S_{60} > 200) = \sum_{k=201}^{\infty} P(S_{60} = k) = \sum_{k=201}^{\infty} \mathrm{e}^{-204} \frac{204^k}{k!} \approx 59\%.$$

b) Koska taulukoiden mukaan $E(S_{60}) = \mathrm{Var}(S_{60}) = 204$, niin normaalijakauma-approksimaatiolla saadaan

$$Z = \frac{S_{60} - E(S_{60})}{\sigma(S_{60})} = \frac{S_{60} - 204}{\sqrt{204}} \sim N(0,1).$$

Siten

$$P(S_{60} > 200) = 1 - P(S_{60} \leq 200) = 1 - P\left(Z \leq \frac{200 - 204}{\sqrt{204}}\right)$$
$$\approx 1 - \Phi(-0.28) = \Phi(0.28) \approx 61\%$$

ja jatkuvuuskorjauksella 59%.

3. a) Sopiva piste-estimaatti odotusarvolle on havaintojen keskiarvo $\overline{x} = 231.67$.

b) Kyseessä on yhden otoksen 2-suuntainen t-väli, joksi saadaan $[230.21, 233.13]$. Koska väli ei sisällä annettua taajuutta 235 Hz, mittaukset eivät tue väitettä. Luottamusvälin laskennan periaate on oltavä näkyvissä. Pelkkä (laskimella saatu) vastaus ei riitä. Tämä tehtävä voidaan ratkaista kuten esimerkiksi Harjoituksen 4 tehtävä 4 a).