MS-A0501 Todennäköisyyslaskennan ja tilastotieteen peruskurssi mallivastaukset

Date of exam: 17.10.2022 **Solved by**: mathwiz.ai

Disclaimer:
These problems are solved by mathwiz.ai, a service leveraging the latest generative AI technologies. This means the solutions come with inherent limitations:

1. **Understanding:** The AI doesn't understand mathematics like humans do; it predicts answers based on historical data patterns.

2. **Reliability and Context:** The AI's responses can be contextually inappropriate or inconsistent, and thus contain errors.

3. **Creativity:** AI applies known patterns, it does not invent new mathematical methods.

Use these solutions as guides, not definitive answers. Despite occasional errors, the solutions can be useful for providing context around the problems. For verifiable accurate solutions, seek guidance from a qualified educator. At the time of solving these problems, the mathematical capabilities of AIs were at the level of an advanced university mathematics student. If you find errors in the solutions, feel free to inform us. Contact information: mathwizcontact@gmail.com.

Problem 1

1. Arvioidaan, että eräässä populaatiossa 1% :Ila on harvinainen tauti. Kyseisen taudin toteamiseksi on kehitetty testi, joka ei kuitenkaan ole täysin luotettava. Jos henkilöllä on tauti, testi antaa tautin viittaavan tuloksen todennäköisyydellä 0.95 . Jos henkilöllä ei ole tautia, testi antaa tautiin viittaavan tuloksen todennäköisyydellä 0.06 . a) Mikä on todennäköisyys että väestöstä satunnaisesti poimittu henkilö saa tautiin viittaavan tuloksen? (3p.) b) Mikä on todennäköisyys että tietyllä henkilöllä on tauti, jos testi antaa tautiin viittaavan tuloksen? (3p.)

Problem 1 solution

Käytetään Bayesin teoreemaa ongelman ratkaisemiseksi.

Olkoon tapahtuma A se, että henkilöllä on tauti, ja tapahtuma B se, että testi antaa tautiin viittaavan tuloksen. Haluamme löytää todennäköisyydet $P(B)$ ja $P(A|B)$.

Annetut todennäköisyydet ovat:

1. $P(A) = 0.01$, eli taudin esiintyvyys populaatiossa. 2. $P(B|A) = 0.95$, eli testin positiivinen tulos, kun henkilöllä on tauti. 3. $P(B|\overline{A}) = 0.06$, eli testin positiivinen tulos, kun henkilöllä ei ole tautia.

a) Lasketaan todennäköisyys $P(B)$ eli todennäköisyys, että satunnaisesti valittu henkilö saa tautiin viittaavan tuloksen. Käytetään kaavaa:

$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$

Missä $P(\overline{A}) = 1 - P(A) = 0.99$, eli todennäköisyys, että henkilöllä ei ole tautia.

Sijoitetaan arvot kaavaan:

$P(B) = (0.95)(0.01) + (0.06)(0.99) = 0.0095 + 0.0594 = 0.0689$

b) Lasketaan todennäköisyys $P(A|B)$ eli todennäköisyys, että henkilöllä on tauti, jos testi antaa tautiin viittaavan tuloksen. Käytetään Bayesin teoreemaa:

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Sijoitetaan arvot kaavaan:

$P(A|B) = \frac{(0.95)(0.01)}{0.0689} = \frac{0.0095}{0.0689} \approx 0.1378$

Vastaukset:

a) Todennäköisyys, että väestöstä satunnaisesti poimittu henkilö saa tautiin viittaavan tuloksen, on noin 0.0689.

b) Todennäköisyys, että tietyllä henkilöllä on tauti, jos testi antaa tautiin viittaavan tuloksen, on noin 0.1378.

Problem 2

2. Jatkuva satunnaismuuttuja T noudattaa jakaumaa, jonka tiheysfunktio on

$$f_T(t) = \begin{cases} \frac{3}{4} \cdot t^2(2-t), & \text{kun } 0 \leq t \leq 2, \\ 0 & \text{muten.} \end{cases}$$

a) Laske $T:$n odotusarvo $\mu_T = \mathbb{E}(T)$ (1p.) ja keskihajonta $\sigma_T = \sqrt{\mathrm{Var}(T)}$(2p.).
b) Jatkuva satunnaismuuttuja $U = T_1 + T_2 + \ldots + T_{100}$, missä satunnaismuuttujat T_i ovat riippumattomat ja noudattavat $T : n$ jakaumaa.

Approksimoi $\mathbb{P}(U \in [115, 130])$ käyttämällä normaaliapproksimaatiota. (3p.)

Problem 2 solution

a) Lasketaan ensin odotusarvo $\mu_T = \mathbb{E}(T)$. Odotusarvo määritellään jatkuvan satunnaismuuttujan tapauksessa seuraavasti:

$\mu_T = \int_{-\infty}^{\infty} t f_T(t) dt$

Koska $f_T(t)$ on nolla, kun $t < 0$ tai $t > 2$, voimme rajoittaa integraalin välille $[0, 2]$:

$\mu_T = \int_0^2 t \left(\frac{3}{4} t^2 (2-t) \right) dt$

$\mu_T = \frac{3}{4} \int_0^2 t^3 (2-t) dt$

Suoritetaan integraali:

$\mu_T = \frac{3}{4} \left[\int_0^2 2t^3 dt - \int_0^2 t^4 dt \right]$

$\mu_T = \frac{3}{4} \left[\frac{1}{2} t^4 \Big|_0^2 - \frac{1}{5} t^5 \Big|_0^2 \right]$

$\mu_T = \frac{3}{4} \left[\frac{1}{2}(16) - \frac{1}{5}(32) \right] = \frac{3}{4} \left[8 - \frac{32}{5} \right] = \frac{3}{4} \cdot \frac{8}{5} = \frac{6}{5}$

Nyt lasketaan varianssi $\mathrm{Var}(T) = \mathbb{E}(T^2) - \mu_T^2$. Ensin lasketaan $\mathbb{E}(T^2)$:

$\mathbb{E}(T^2) = \int_{-\infty}^{\infty} t^2 f_T(t) dt = \int_0^2 t^2 \left(\frac{3}{4} t^2 (2-t) \right) dt$

$\mathbb{E}(T^2) = \frac{3}{4} \int_0^2 t^4 (2-t) dt$

Suoritetaan integraali:

$$\mathbb{E}(T^2) = \frac{3}{4}\left[\int_0^2 2t^4 dt - \int_0^2 t^5 dt\right]$$

$$\mathbb{E}(T^2) = \frac{3}{4}\left[\frac{2}{5}t^5\Big|_0^2 - \frac{1}{6}t^6\Big|_0^2\right]$$

$$\mathbb{E}(T^2) = \frac{3}{4}\left[\frac{2}{5}(32) - \frac{1}{6}(64)\right] = \frac{3}{4}\left[\frac{64}{5} - \frac{64}{6}\right] = \frac{3}{4}\cdot\frac{64}{30} = \frac{16}{5}$$

Nyt voimme laskea varianssin:

$$\mathrm{Var}(T) = \mathbb{E}(T^2) - \mu_T^2 = \frac{16}{5} - \left(\frac{6}{5}\right)^2 = \frac{16}{5} - \frac{36}{25} = \frac{44}{25}$$

Lopuksi lasketaan keskihajonta:

$$\sigma_T = \sqrt{\mathrm{Var}(T)} = \sqrt{\frac{44}{25}} = \frac{2\sqrt{11}}{5}$$

b) Käytetään normaaliapproksimaatiota arvioimaan todennäköisyys $\mathbb{P}(U \in [115, 130])$. Koska $U = T_1 + T_2 + \cdots + T_{100}$, voimme käyttää keskeisen raja-arvo teoreemaa ja todeta, että U noudattaa likimain normaalijakaumaa, jonka odotusarvo on $\mu_U = 100\mu_T$ ja varianssi on $\mathrm{Var}(U) = 100\,\mathrm{Var}(T)$.

Lasketaan μ_U ja σ_U:

$$\mu_U = 100\mu_T = 100 \cdot \frac{6}{5} = 120$$

$$\sigma_U = \sqrt{\mathrm{Var}(U)} = \sqrt{100\,\mathrm{Var}(T)} = \sqrt{100 \cdot \frac{44}{25}} = 4\sqrt{11}$$

Nyt voimme standardisoida U:

$$Z = \frac{U - \mu_U}{\sigma_U}$$

Haluttu todennäköisyys on $\mathbb{P}(U \in [115, 130]) = \mathbb{P}(115 \le U \le 130)$. Standardisoidussa muodossa tämä on:

$$\mathbb{P}\left(\frac{115-120}{4\sqrt{11}} \le Z \le \frac{130-120}{4\sqrt{11}}\right) = \mathbb{P}\left(-\frac{5}{4\sqrt{11}} \le Z \le \frac{10}{4\sqrt{11}}\right)$$

Käyttämällä normaalijakauman taulukoita tai laskinta, voimme arvioida tämän todennäköisyyden:

$$\mathbb{P}\left(-\frac{5}{4\sqrt{11}} \le Z \le \frac{10}{4\sqrt{11}}\right) \approx \Phi\left(\frac{10}{4\sqrt{11}}\right) - \Phi\left(-\frac{5}{4\sqrt{11}}\right) \approx 0.9545 - 0.0455 = 0.909$$

Vastaukset:

a) T:n odotusarvo on $\mu_T = \frac{6}{5}$ ja keskihajonta on $\sigma_T = \frac{2\sqrt{11}}{5}$.

b) Todennäköisyys, että $U \in [115, 130]$, on noin 0.909, kun käytetään normaaliapproksimaatiota.

Problem 3

3. Diskreetti satunnaismuuttuja X, jonka perusjoukko on $S_X = \{0, 1, 2, \ldots\}$ ja tiheysfunktio $f_X(k) = \mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$, kun $k \in S_X$, missä $\lambda > 0$, sanotaan Poisson-jakautuneeksi parametrilla λ. Tätä merkitään $X \sim Po(\lambda)$. Johda kaava $\lambda : n$ suurimman uskottavuuden estimaatille λ^*_{ML}, joka saadaan satunnaismuuttujan X riippumattomasta otoksesta $\bar{x} = \{x_1, x_2, \ldots, x_n\}$.(6p.

Problem 3 solution

Suurimman uskottavuuden estimaatin löytämiseksi tarkastellaan ensin uskottavuusfunktiota. Koska otos $x_1, x_2, \ldots, x_n$ on riippumaton, voimme kirjoittaa uskottavuusfunktion $L(\lambda)$ seuraavasti:

$L(\lambda) = f_X(x_1, \lambda) f_X(x_2, \lambda) \cdots f_X(x_n, \lambda)$

Koska $f_X(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$, voimme kirjoittaa uskottavuusfunktion muodossa:

$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!}$

Yleensä on helpompi työskennellä logaritmisen uskottavuusfunktion kanssa, joten otetaan molemmat puolet logaritmi:

$\log L(\lambda) = \sum_{i=1}^n \log \left(e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!} \right)$

$\log L(\lambda) = \sum_{i=1}^n (-\lambda + x_i \log \lambda - \log(x_i!))$

Nyt haluamme löytää λ, joka maksimoi $\log L(\lambda)$. Tätä varten otetaan derivaatta $\log L(\lambda)$ suhteessa λ:

$\frac{d}{d\lambda} \log L(\lambda) = \sum_{i=1}^n \left(-1 + \frac{x_i}{\lambda} \right)$

Jotta löydämme maksimin, asetetaan derivaatta nollaksi ja ratkaistaan λ:

$\sum_{i=1}^n \left(-1 + \frac{x_i}{\lambda} \right) = 0$

$\sum_{i=1}^n \frac{x_i}{\lambda} = n$

$\frac{1}{\lambda} \sum_{i=1}^n x_i = n$

$\lambda = \frac{\sum_{i=1}^n x_i}{n}$

Nyt olemme johdettu kaava suurimman uskottavuuden estimaatille λ:

$\lambda^*_{ML} = \frac{\sum_{i=1}^n x_i}{n}$

Problem 4

4. Teemu Teekkari tutki erästä Poisson-jakautunutta satunnaismuuttujaa $X \sim Po(\theta)$, missä tuntematon parametri $\theta > 0$. Hän ajatteli $\theta : n$ arvon olevan eräs jatkuva satunnaismuuttuja Θ, jonka perusjoukko $S_\Theta =]0, \infty[$ ja jonka priori-jakaumaksi hän arvioi

$$p_0(\theta) = \begin{cases} 2e^{-2\theta} & ,\theta \in S_\Theta, \\ 0 & ,\theta \notin S_\Theta. \end{cases}$$

Hän päätti ottaa otoksen $\bar{X} = \{X_1, X_2, \ldots, X_n\}$, laskea sen perusteella posteriori-jakauman $p_1(\theta \mid \bar{X})$ ja käyttää θ : n estimattina posteriori-jakauman moodia, eli sellaista parametrin θ : n arvoa $\theta_M(\bar{X})$, jossa posteriori-jakauman tiheysfunktio $p_1(\theta \mid \bar{X})$ saavuttaa suurimman arvonsa.
Johda kaava tälle estimaatille $\theta_M(\bar{x})$, joka saadaan otoksesta $\bar{x} = \{x_1, x_2, \ldots, x_n\}$. (6p.)
(Huom: Jos $g(t)$ on positivinen funktio ja $c > 0$ on vakio, niin $g(t)$ ja $c \cdot g(t)$ saavuttavat suurimman arvonsa samassa pisteessä.)

Problem 4 solution

Bayesin teoreeman mukaan posteriori-jakauma $p_1(\theta|\bar{X})$ voidaan laskea seuraavasti:

$p_1(\theta|\bar{X}) = \frac{f(\bar{X}|\theta)p_0(\theta)}{f(\bar{X})}$

Missä $f(\bar{X}|\theta)$ on otoksen $\bar{X}$ yhteinen tiheysfunktio ehdolla θ, ja $f(\bar{X})$ on otoksen $\bar{X}$ marginaalinen tiheysfunktio. Koska $X_i \sim Po(\theta)$, niin $f(X_i|\theta) = e^{-\theta} \cdot \frac{\theta^{x_i}}{x_i!}$. Koska otos on riippumaton, niin yhteinen tiheysfunktio on:

$f(\bar{X}|\theta) = \prod_{i=1}^n f(X_i|\theta) = \prod_{i=1}^n e^{-\theta} \cdot \frac{\theta^{x_i}}{x_i!} = e^{-n\theta} \cdot \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$

Nyt voimme kirjoittaa posteriori-jakauman muodossa:

$p_1(\theta|\bar{X}) = \frac{e^{-n\theta} \cdot \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \cdot 2e^{-2\theta}}{f(\bar{X})}$

$p_1(\theta|\bar{X}) \propto e^{-(n+2)\theta} \cdot \theta^{\sum_{i=1}^n x_i}$

Huomaa, että emme tarvitse marginaalista tiheysfunktiota $f(\bar{X})$, koska se ei riipu θ:sta ja etsimme vain suurimman arvon saavuttavaa θ:ta.

Nyt haluamme löytää $\theta_M(\bar{x})$, joka maksimoi $p_1(\theta|\bar{X})$. Tätä varten otetaan derivaatta $p_1(\theta|\bar{X})$ suhteessa θ ja asetetaan se nollaksi:

$\frac{d}{d\theta} p_1(\theta|\bar{X}) \propto \frac{d}{d\theta} \left(e^{-(n+2)\theta} \cdot \theta^{\sum_{i=1}^n x_i} \right) = 0$

Tämän derivaatan laskeminen suoraan voi olla hankalaa, joten otetaan molemmat puolet logaritmi:

$\log p_1(\theta|\bar{X}) \propto -(n+2)\theta + (\sum_{i=1}^n x_i) \log \theta$

Nyt otetaan derivaatta logaritmisen funktion suhteen:

$\frac{d}{d\theta} \log p_1(\theta|\bar{X}) = -(n+2) + \frac{\sum_{i=1}^n x_i}{\theta}$

Asetetaan derivaatta nollaksi ja ratkaistaan θ:

$-(n+2) + \frac{\sum_{i=1}^n x_i}{\theta} = 0$

$\theta = \frac{\sum_{i=1}^n x_i}{n+2}$

Nyt olemme johdettu kaava estimaatille $\theta_M(\bar{x})$:

$\theta_M(\bar{x}) = \frac{\sum_{i=1}^n x_i}{n+2}$