Tilastomatematiikka

Ensimmäinen välikoe 16.02.2023

1. Laske todennäköisyys $\mathbb{P}(A \cup B)$, kun

   a) $\mathbb{P}(A) = 0.5$ ja $\mathbb{P}(B) = 0.3$ sekä tapahtumien A ja B oletetaan olevan riippumattomat.

   b) $A = \{X \leq 9\}$ ja $B = \{X \geq 12\}$, missä X on normaalijakautunut satunnaismuuttuja parametreilla $\mu = 10$ ja $\sigma = 1$.

   Piirrä tehtävänannon mukaiset Vennin diagrammit.

2. Erään ydinvoimalan kunnossapidosta vastaavien insinöörien täytyy usein tarkistaa jäähdytysjärjestelmässä olevien putkien sisäpinnan korroosio. Koska putkien sisäpinnan tutkiminen ei suoraan ole mahdollista, käytetään epäsuoraa testiä putkia rikkomatta. Testi ei ole virheetön, vaan siinä on seuraavat virhemahdollisuudet

   -testitulos on negatiivinen, vaikka putkessa on korroosiota, 30 % todennäköisyydellä.

   -testitulos on positiivinen, vaikka putkessa ei ole korroosiota, 20 % todennäköisyydellä.

   Oletetaan, että putkissa on korroosiota 10 % todennäköisyydellä.

   a) Millä todennäköisyydellä testitulos on positiivinen? Piirrä tilannetta havainnollistava puukaavio.

   b) Millä todennäköisyydellä putkessa on korroosiota, jos testitulos on positiivinen?

3. Digitaalisessa viestintäkanavassa yksittäinen bitti vastaanotetaan virheellisenä todennäköisyydellä 10^{-2} muista biteistä riippumatta. Jos lähetetään 2023 bittiä, niin millä todennäköisyydellä vastaanotetaan korkeintaan 16 virheellistä bittiä?

   a) Määrää tilanteeseen sopiva satunnaismuuttuja ja anna laskentakaava, jolla kysytty todennäköisyys voidaan laskea. (2p)

   b) Laske kysytty todennäköisyys normaalijakauma-approksimaatiolla. Käytä jatkuvuuskorjausta. (4p)

   c) Mitä voit sanoa approksimaation tarkkuudesta? (1p)

Tehtävien ratkaisuperiaatteet

1. a) Kaavakokoelman ja riippumattomuuden mukaan

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= 0.5 + 0.3 - 0.5 \cdot 0.3 = 0.65 = 65\,\%.\end{aligned}$$

b) Koska $X \sim N(10, 1)$, niin tapahtuman A todennäköisyydeksi saadaan

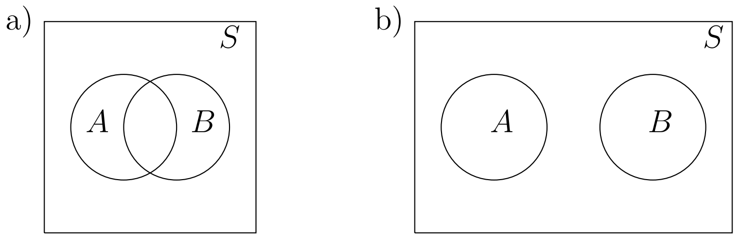
$$\begin{aligned}\mathbb{P}(A) = \mathbb{P}(X \le 9) &= \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le \frac{9-10}{1}\right) && \text{(standardointi)} \\ &= \mathbb{P}(Z \le -1) = \Phi(-1) = 1 - \Phi(1) && \text{(symmetriaominaisuus)} \\ &\approx 0.1587 && \text{(taulukosta)}\end{aligned}$$

Vastaavalla tavalla saadaan

$$\begin{aligned}\mathbb{P}(A) = \mathbb{P}(X \ge 12) &= \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \ge \frac{12-10}{1}\right) && \text{(standardointi)} \\ &= \mathbb{P}(Z \ge 2) = 1 - \mathbb{P}(Z \le 2) = 1 - \Phi(2) && (Z \text{ on jatkuva sm.}) \\ &\approx 0.0228 && \text{(taulukosta)}\end{aligned}$$

Koska A ja B ovat erilliset, on kysytty todennäköisyys

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \approx 18\,\%.$$



Kuva 1: Vennin diagrammit tehtävään 1

Kuvassa on oleellista, etteivät a)-kohdan tapahtumat voi olla erilliset, kun taas b)-kohdan tapahtumat ovat erilliset.

2. Merkitään

$$\begin{aligned}K &= \text{"putkessa on korroosiota"}, \\ +/- &= \text{"testitulos on positiivinen/negatiivinen"}.\end{aligned}$$

Tiedetään todennäköisyydet $\mathbb{P}(K) = 0.1$, $\mathbb{P}(-|K) = 0.3$ ja $\mathbb{P}(+|\overline{K}) = 0.2$.

a) Kysytään todennäköisyyttä $\mathbb{P}(+)$. Kokonaistodennäköisyyden mukaan

$$\begin{aligned}\mathbb{P}(+) &= \mathbb{P}(+|K)\mathbb{P}(K) + \mathbb{P}(+|\overline{K})\mathbb{P}(\overline{K}) = (1 - \mathbb{P}(-|K))\mathbb{P}(K) + \mathbb{P}(+|\overline{K})(1 - \mathbb{P}(K)) \\ &= 0.7 \cdot 0.1 + 0.2 \cdot 0.9 = 0.25,\end{aligned}$$

eli testitulos on positiivinen 25 % todennäköisyydellä.

c) Kysytään todennäköisyyttä $\mathbb{P}(K|+)$. Bayesin kaavan ja edellä lasketun kokonaistodennäköisyyden mukaan

$$\mathbb{P}(K|+) = \frac{\mathbb{P}(+|K)\mathbb{P}(K)}{\mathbb{P}(+)} = \frac{0.7 \cdot 0.1}{0.25} = 28\,\%.$$

3. a) Merkitään $X =$ "vastaanotettujen virheellisten bittien lkm.", joka noudattaa binomijakaumaa $\mathrm{Bin}(2023, 0.01)$. Kysytään todennäköisyyttä $\mathbb{P}(X \leq 16)$, joka binomijakauman pistetodennäköisyyksien laskukaavan mukaan on

$$\mathbb{P}(X \leq 16) = \sum_{k=0}^{16} \mathbb{P}(X = k) = \sum_{k=0}^{16} \binom{2023}{k} 0.01^k 0.99^{2023-k} \approx 20.5\,\%.$$

Viimeisen arvion saa vaikkapa Excelistä, mutta sitähän tässä ei kysytty.

b) Normaalijakauma-approksimaatiota varten tarvitaan tunnusluvut $\mathbb{E}(X) = np = 2023 \cdot 0.01 = 20.23$ ja $\mathrm{Var}(X) = np(1-p) = 2023 \cdot 0.01 \cdot 0.99 = 20.0277$. Kysytyksi todennäköisyydeksi saadaan

$$\begin{aligned}
\mathbb{P}(X \leq 16) &= \mathbb{P}(X \leq 16.5) && \text{(jatkuvuuskorjaus)}\\
&= \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \leq \frac{16.5 - 20.23}{\sqrt{20.0277}}\right) && \text{(standardointi)}\\
&\approx \Phi(-0.83) && \text{(normaaliapproksimaatio)}\\
&= 1 - \Phi(0.83) && \text{(symmetriaominaisuus)}\\
&\approx 20.3\,\%. && \text{(taulukosta).}
\end{aligned}$$

c) Jos on laskenut tarkan todennäköisyyden, niin voi todeta, että arvio on aika tarkka. Jos tarkkaa todennäköisyyttä ei ole laskenut, voi todeta, että "n on aika suuri", jolloin approksimaatio toimii hyvin. Lisäksi esimerkiksi $np(1-p) \approx 20 > 9$, joten tämänkin arvion mukaan approksimaation pitäisi toimia.

Kaavoja

Todennäköisyyden ominaisuuksia

$$
\begin{aligned}
&\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),\\
&\mathbb{P}(A \backslash B) = \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B),\\
&\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A),\\
&\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B),\\
&\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}\\
&\mathbb{P}(B) = \sum_{k=1}^{n} \mathbb{P}(B|A_k)\mathbb{P}(A_k)
\end{aligned}
$$

Odotusarvoja ja variansseja

Ptnf. tai tf.	$\mu_X := \mathbb{E}(X)$	$\sigma_X^2 := \mathrm{Var}(X)$
$\mathbb{P}(X = x)$	$\sum_x x\mathbb{P}(X = x)$	$\sum_x (x - \mu_X)^2 \mathbb{P}(X = x)$
$f_X(x)$	$\int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$	$\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) \mathrm{d}x$
$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
$p(1-p)^{x-1}$	$1/p$	$(1-p)/p^2$
$\frac{a^x}{x!} \mathrm{e}^{-a}$	a	a
$1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$
$\theta \mathrm{e}^{-\theta x}$	$1/\theta$	$1/\theta^2$
$\frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

$$
\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y), \quad \mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2
$$

Eräitä testimuuttujia

$$
\begin{aligned}
&\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ (likimain, kun ”n on suuri”)},\\
&\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},\\
&\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t_{n+m-2},\\
&\sqrt{n-1}\, s_x \frac{\frac{S_{xY}}{s_{xx}} - \beta}{S_r} \sim t_{n-2}
\end{aligned}
$$

Regressio, korrelaatio ja kovarianssi

$$r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}};\ s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y});\ s_{xx} = s_x^2;$$

$$y = a + bx; \quad b = \frac{s_{xy}}{s_{xx}}; \quad a = \overline{y} - b\overline{x};$$

$$s_r^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = \frac{n-1}{n-2}(1-r^2)s_{yy};$$

$$\begin{aligned}\sigma_{XY} &= \mathrm{Cov}(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))), \quad \sigma_{XX} = \sigma_X^2; \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

$$\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$