Tilastomatematiikka

Ensimmäinen välikoe 13.02.2020

1. Olkoon $X \sim N(100, 5^2)$ normaalijakautunut satunnaismuuttuja ja olkoot

$$A = \{X < 90\} \quad \text{ja} \quad B = \{X > 110\}$$

tapahtumia.

a) Laske todennäköisyydet $\mathbb{P}(A)$ ja $\mathbb{P}(B)$.

b) Laske todennäköisyys $\mathbb{P}(A \cup B)$. Ovatko tapahtumat A ja B riippumattomia?

2. Satunnaismuuttujalla X on jakauma

x	0	1	2
$\mathbb{P}(X = x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

a) Laske muuttujan X odotusarvo ja mediaani. Satunnaismuuttujan X mediaani määritellään lukuna Md_X, joka kuuluu X:n arvojoukkoon ja jolle pätee

$$\mathbb{P}(X < \mathrm{Md}_X) < \frac{1}{2} \quad \text{ja} \quad \mathbb{P}(X \geq \mathrm{Md}_X) \geq \frac{1}{2}.$$

(4p)

b) Anna esimerkki satunnaismuuttujasta Y, jolla on sama arvojoukko kuin satunnaismuuttujalla X ja jolle $\mu_Y = \mathrm{Md}_Y$. (2p)

3. Viruksen tunkeutumista soluun voidaan ajatella satunnaisilmiönä, jota voidaan mallintaa Poisson-jakauman avulla. Tällöin virusinfektion MOI-suhde ilmoittaa kuinka monta viruspartikkelia pääsee keskimäärin tunkeutumaan soluun. Oletetaan, että solu altistuu virusinfektiolle, jonka MOI-suhde on 2.1.

a) Millä todennäköisyydellä soluun tunkeutuu vähintään yksi viruspartikkeli?

b) Laske a)-kohdan todennäköisyys normaalijakauma-approksimaatiolla. Käytä jatkuvuuskorjausta. Mitä voit sanoa approksimaation tarkkuudesta?

Kaavoja

Todennäköisyyden ominaisuuksia

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

Odotusarvoja ja varriansseja

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

$$
\begin{aligned}
&r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}};\; s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y});\; s_{xx} = s_x^2;\\
&y = a + bx; \quad b = \frac{s_{xy}}{s_{xx}}; \quad a = \overline{y} - b\overline{x};\\
&s_r^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = \frac{n-1}{n-2}(1 - r^2)s_{yy};\\
&\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))), \quad \sigma_{XX} = \sigma_X^2;\\
&\qquad = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\\
&\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}
\end{aligned}
$$

Tehtävien ratkaisuperiaatteet

1. a) Koska $X \sim N(100, 5^2)$, saadaan standardoimalla $Z = \frac{X-\mu}{\sigma} = \frac{X-100}{5} \sim N(0,1)$. Tapahtumalle A saadaan todennäköisyydeksi

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(X < 90) = \mathbb{P}\left(\frac{X-100}{5} < \frac{90-100}{5}\right) && \text{(standardointi)}\\
&= \mathbb{P}(Z < -2) = \Phi(-2) && (F_Z(z) = \Phi(z))\\
&= 1 - \Phi(2) && \text{(symmetriaominaisuus)}\\
&\approx 1 - 0.9772 \approx 0.023. && \text{(taulukosta)}
\end{aligned}$$

Symmetrian nojalla $\mathbb{P}(B) = \mathbb{P}(A)$.

b) Koska $A \cap B = \{X > 110\} \cap \{X < 90\} = \emptyset$, on

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \overset{\text{a)}}{=} 2 \cdot \mathbb{P}(A) \approx 0.046.$$

Koska $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B) \approx 0.023^2$, A ja B *eivät ole riippumattomat.*

2. a) Kaavoista saadaan diskreetin jakauman odotusarvoksi

$$\mathbb{E}(X) = \sum_{k=0}^{2} k \cdot \mathbb{P}(X = k) = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} = \frac{4}{3}.$$

Koska $\mathbb{P}(X < 1) = \mathbb{P}(X = 0) = \frac{1}{6}$ ja $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = \frac{5}{6} \geq \frac{1}{2}$, niin mediaanin määritelmän mukaan $\mathrm{Md}_X = 1$.

b) Sekä odotusarvo että mediaani ovat keskilukuja. Ne ovat samat, jos jakauma on symmetrinen ”keskikohtansa” suhteen. Erityisen symmetrinen jakauma on diskreetti tasajakauma Y, jolle $\mathbb{P}(Y = k) = \frac{1}{3}$ kaikilla $k = 0, 1, 2$. Tällöin $\mathbb{E}(Y) = 1 = \mathrm{Md}_Y$, joten diskreetti tasajakauma on eräs esimerkki.

3. a) Meitä kiinnostaa satunnaismuuttuja

$$X = \text{”soluun tunkeutuneiden viruspartikkelien lukumäärä”} \sim \mathrm{Poi}(a).$$

Poisson-jakautuneen muuttujan $X \sim \mathrm{Poi}(a)$ odotusarvo on a, joten tehtävänannon mukaan $X \sim \mathrm{Poi}(2.1)$. Kysytään todennäköisyyttä

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) \overset{\text{kaavat}}{=} 1 - \mathrm{e}^{-2.1}\frac{2.1^0}{0!} \approx 0.88.$$

b) Koska $X \sim \mathrm{Poi}(2.1)$, niin kaavojen mukaan $\mathbb{E}(X) = 2.1$ ja $\mathrm{Var}(X) = 2.1$, joten standardoimalla saadaan

$$Z = \frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} = \frac{X - 2.1}{\sqrt{2.1}} \overset{\text{norm.appr.}}{\sim} N(0,1).$$

Kysytyksi todennäköisyydeksi saadaan

$$\begin{aligned}
\mathbb{P}(X \geq 1) &= \mathbb{P}(X \geq \frac{1}{2}) && \text{(jatkuvuuskorjaus)}\\
&= \mathbb{P}\left(\frac{X-2.1}{\sqrt{2.1}} \geq \frac{0.5-2.1}{\sqrt{2.1}}\right) \approx \mathbb{P}(Z \geq -1.10) && \text{(standardointi)}\\
&= 1 - \Phi(-1.10) = \Phi(1.10) && \text{(symmetriaominaisuus)}\\
&\approx 0.86. && \text{(taulukosta)}
\end{aligned}$$

Approksimaatio toimii kohtuullisen hyvin, vaikka $a = 2.1 \approx np$ ei ole erityisen suuri.