

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK-FYS1110 — Sannsynlighetsregning og statistisk modellering

Eksamensdag: 7. juni - 2024

Tid for eksamen: 09.00–13.00.

Oppgavesettet er på 6 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Godkjent kalkulator  
Formelsamling for STK1100

Kontroller at oppgavesettet er komplett før  
du begynner å besvare spørsmålene.

### Oppgave 1

Vi har at  $X$  er Poisson-fordelt, dvs.

$$P(X = x) = \frac{(\lambda v_0)^x}{x!} e^{-\lambda v_0}, \quad x = 0, 1, 2, \dots$$

(a)  $E(X) = \lambda v_0$  og  $V(X) = \lambda v_0$ . Vi kan tolke  $\lambda$  som forventet antall E.coli-bakterier per liter vann.

(b) Med  $\lambda = 3$  og  $v_0 = 1$  finner vi

$$P(X = 0) = e^{-3} = 0.0498.$$

Med  $\lambda = 3$  skal vi finne  $v_0$  slik at det er minst 0.9975 i sannsynlighet for at  $P(X \geq 1)$ :

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-3v_0} \geq 0.9975$$

som gir  $v_0 \geq 1.997$ , dvs. prøven må være mer enn 1.997 liter.

(c) La  $X_i$  være antall bakterier i prøve  $i$ ,  $i = 1, 2, \dots, 10$ . Vi har

$$P(X_i = 0) = e^{-0.3} = 0.741.$$

Bruker uavhengigheten til å multiplisere sannsynlighetene i siste linje:

$$\begin{aligned} P(\text{alarm}) &= P(\text{minst én } X_i \geq 1) \\ &= 1 - P(\text{ingen } X_i \geq 1) \\ &= 1 - P(\text{alle } X_i = 0) \\ &= 1 - 0.741^{10} = 0.9502. \end{aligned}$$

(Fortsettes på side 2.)

Dette er egentlig identisk med situasjonen i a) der man ser etter minst én bakterie i en liter vann.

Også mulig å identifisere dette som en binomisk situasjon,  $Y \sim \text{bin}(n, p)$  med  $n = 10$ ,  $p = 1 - P(X_i = 0) = 1 - e^{-0.3} = 0.259$ , og finne  $P(\text{alarm}) = P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - 0.259)^{10} = 0.9502$

- (d) Har  $E(X_1) = V(X_1) = \lambda v_1$  og  $E(X_2) = V(X_2) = \lambda v_2$ . Reglene for forventning og varians til lineærkomb. av uavhengige stokastiske variabler gir

$$E(\hat{\lambda}) = \frac{E(X_1 + X_2)}{v_1 + v_2} = \frac{\lambda(v_1 + v_2)}{v_1 + v_2} = \lambda$$

$$V(\hat{\lambda}) = \frac{V(X_1 + X_2)}{(v_1 + v_2)^2} = \frac{\lambda(v_1 + v_2)}{(v_1 + v_2)^2} = \frac{\lambda}{v_1 + v_2}$$

Standardfeilen til  $\hat{\lambda}$  er derfor  $\sqrt{\frac{\lambda}{v_1 + v_2}}$ .

## Oppgave 2

La  $X$  og  $Y$  være kontinuerlige stokastiske variabler med simultantetthet

$$f(x, y) = \begin{cases} \frac{1}{17}x(x + y) & \text{når } 0 \leq x \leq 2 \text{ og } 0 \leq y \leq 3 \\ 0 & \text{ellers} \end{cases}$$

- (a) Det enkleste er å beregne

$$P(Y \geq X) = \frac{1}{17} \int_0^2 \int_x^3 (x^2 + xy) dy dx = \frac{11}{17}.$$

Kan også finnes ved

$$P(Y \geq X) = \frac{1}{17} \int_0^2 \int_0^y (x^2 + xy) dx dy + \frac{1}{17} \int_2^3 \int_0^2 (x^2 + xy) dx dy = \frac{11}{17}$$

(tegn figur over området for å se integrasjonsgrensene!).

- (b) Når  $0 \leq y \leq 3$  har vi at

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{17} \int_0^2 (x^2 + xy) dx = \frac{1}{17} \left[ \frac{1}{3}x^3 + \frac{1}{2}x^2y \right]_{x=0}^2 = \frac{1}{17} \left( \frac{8}{3} + 2y \right)$$

slik at vi får

$$f_Y(y) = \begin{cases} \frac{2}{17} \left( \frac{4}{3} + y \right) & \text{når } 0 \leq y \leq 3 \\ 0 & \text{ellers} \end{cases}$$

(Fortsettes på side 3.)

Når  $0 \leq x \leq 2$  har vi at

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{17} \int_0^3 (x^2 + xy) dy = \frac{1}{17} \left[ x^2 y + \frac{1}{2} xy^2 \right]_{y=0}^3 = \frac{1}{17} (3x^2 + \frac{9}{2}x),$$

slik at vi får

$$f_X(x) = \begin{cases} \frac{3}{17}x(x + \frac{3}{2}) & \text{når } 0 \leq x \leq 2 \\ 0 & \text{ellers} \end{cases}$$

$X$  og  $Y$  kan ikke være uavhengige, siden vi enkelt ser fra formlene at

$$f(x, y) \neq f_X(x) \cdot f_Y(y).$$

(c) Når  $0 \leq y \leq 3$  har vi at

$$F_Y(y) = \int_{-\infty}^y f_Y(t) dt = \frac{2}{17} \int_0^y (\frac{4}{3} + t) dt = \frac{2}{17} \left[ \frac{4}{3}t + \frac{1}{2}t^2 \right]_0^y$$

slik at vi får

$$F_Y(y) = \begin{cases} 0 & \text{når } y < 0 \\ \frac{2}{17}(\frac{4}{3}y + \frac{1}{2}y^2) & \text{når } 0 \leq y \leq 3 \\ 1 & \text{når } y > 3 \end{cases}$$

Vi finner medianen  $\nu$  ved å løse andregradsligningen som oppstår når vi setter  $F_Y(\nu) = \frac{1}{2}$ .

## Oppgave 3

(a) Vi har at

$$L(\sigma) = \prod_{i=1}^n f(x_i|\sigma) = \prod_{i=1}^n \frac{1}{2\sigma} \exp(-|x_i|/\sigma) = \frac{1}{2^n \sigma^n} \exp(-\sum_{i=1}^n |x_i|/\sigma)$$

og

$$\ell(\sigma) = -n \log(2) - n \log(\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |x_i|.$$

Dermed er

$$\frac{\partial}{\partial \sigma} \ell(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n |x_i|$$

og setter vi lik 0, får vi

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n |x_i|.$$

(Fortsettes på side 4.)

(b) Vi har at

$$\begin{aligned} E[\hat{\sigma}] &= E[|X_i|] = \int_{-\infty}^{\infty} \frac{1}{2\sigma} |x| \exp(-|x|/\sigma) dx \\ &= \int_0^{\infty} \frac{1}{\sigma} x \exp(-x/\sigma) dx = \frac{1}{\sigma} \sigma^2 \Gamma(2) = \sigma \end{aligned}$$

dvs forventningsrett.

Vi har videre at

$$\begin{aligned} E[|X|^2] &= \int_{-\infty}^{\infty} \frac{1}{2\sigma} x^2 \exp(-|x|/\sigma) dx = \int_0^{\infty} \frac{1}{\sigma} x^2 \exp(-x/\sigma) dx \\ &= \frac{1}{\sigma} \sigma^3 \Gamma(3) = 2\sigma^2 \end{aligned}$$

Dermed er

$$V[|X|] = E[|X|^2] - E[|X|]^2 = 2\sigma^2 - \sigma^2 = \sigma^2$$

(c) Vi har at  $\hat{\sigma}$  er et gjennomsnitt, og ifølge sentralgrenseteoremet vil da  $\hat{\sigma}$  være tilnærmet normalfordelt når  $n$  er stor.

Vi har da at

$$\Pr(-z_{\alpha/2} < \frac{\hat{\sigma} - \sigma}{\sigma/\sqrt{n}} < z_{\alpha/2}) \approx 1 - \alpha$$

Nå er

$$\begin{aligned} \frac{\hat{\sigma} - \sigma}{\sigma/\sqrt{n}} &< z_{\alpha/2} \\ &\Downarrow \\ \hat{\sigma} - \sigma &< z_{\alpha/2} \sigma / \sqrt{n} \\ &\Downarrow \\ \hat{\sigma} &< (1 + z_{\alpha/2} / \sqrt{n}) \sigma \\ &\Downarrow \\ \frac{\hat{\sigma}}{1 + z_{\alpha/2} / \sqrt{n}} &< \sigma \end{aligned}$$

Tilsvarende vil den andre ulikheten gi

$$\frac{\hat{\sigma}}{1 - z_{\alpha/2} / \sqrt{n}} > \sigma$$

som da gir det ønskede intervallet.

(Fortsettes på side 5.)

(d) Vi kan også bruke at siden  $\hat{\sigma}$  er konsistent, så kan argumentere med at

$$\frac{\hat{\sigma} - \sigma}{\sigma/\sqrt{n}} \approx \frac{\hat{\sigma} - \sigma}{\hat{\sigma}/\sqrt{n}}$$

som da også vil være tilnærmet normalfordelt. Da har vi

$$\begin{aligned} \Pr(-z_{\alpha/2} < \frac{\hat{\sigma} - \sigma}{\hat{\sigma}/\sqrt{n}} < z_{\alpha/2}) &\approx 1 - \alpha \\ &\Downarrow \\ \Pr(-z_{\alpha/2}\hat{\sigma}/\sqrt{n} < \hat{\sigma} - \sigma < z_{\alpha/2}\hat{\sigma}/\sqrt{n}) &\approx 1 - \alpha \\ &\Downarrow \\ \Pr(-\hat{\sigma} - z_{\alpha/2}\hat{\sigma}/\sqrt{n} < -\sigma < -\hat{\sigma} + z_{\alpha/2}\hat{\sigma}/\sqrt{n}) &\approx 1 - \alpha \\ &\Downarrow \\ \Pr(\hat{\sigma} + z_{\alpha/2}\hat{\sigma}/\sqrt{n} > \sigma > \hat{\sigma} - z_{\alpha/2}\hat{\sigma}/\sqrt{n}) &\approx 1 - \alpha \end{aligned}$$

som da gir det alternative intervallet.

Siden vi har  $\alpha = 0.05$  så ønsker vi at konfidensintervallet skal dekke den sanne parameter 95% av gangene vi bruker dette intervallet. For  $n = 10$  så vil det første intervallet dekke det 95.9% av gangene mens det andre intervallet kun dekker det 91.1% av gangene, som antyder at det første intervallet er best.

Når  $n = 100$  så oppfører begge intervallene seg svært nærme målnivået.

## Oppgave 4

(a) Feilpropagering gir ved første ordens Taylorutvikling osv:

$$u_d^2 = \left( \frac{\partial d}{\partial v} u_v \right)^2 + \left( \frac{\partial d}{\partial \alpha} u_\alpha \right)^2 \quad (\text{s1})$$

Ved å derivere  $d$  mhp.  $v$  og  $\alpha$  får vi dermed:

$$u_d^2 = \left( \frac{2v \sin 2\alpha}{g} u_v \right)^2 + \left( \frac{2v^2 \cos 2\alpha}{g} u_\alpha \right)^2 \quad (\text{s2})$$

$$\rightarrow u_d = \frac{2v}{g} \sqrt{(\sin 2\alpha u_v)^2 + (v \cos 2\alpha u_\alpha)^2} \quad (\text{s3})$$

(b) 1. Fra likning (1) har vi:

$$v = \sqrt{\frac{dg}{\sin 2\alpha}} \quad (\text{s4})$$

(Fortsettes på side 6.)

Ved innsetting fås  $v = 15.2 \text{ m/s}$ .

2. Vi bruker (s3), setter inn for  $v$ ,  $\alpha$ ,  $u_v$ ,  $u_\alpha$  og  $g$ , og får  $u_d = 2.1 \text{ m}$ .

3. Presentasjon av usikkerhet skal normalt kun ha ett signifikant siffer, så rekkevidde inkludert usikkerhet er  $20 \pm 2 \text{ m}$ .

- (c) Vi ønsker å finne sannsynligheten for at prosjektilet faller *utenfor* intervallet  $[\mu - 1, \mu + 1]$  (siden målet har en utstrekning på  $\pm 1 \text{ m}$ ).  $\mu$  er  $20 \text{ m}$ , men dette tallet trenger vi ikke. Vi ønsker altså å estimere

$$P_{bom} = 2P(D > \mu + 1) \quad (\text{s5})$$

Vi introduserer

$$Z = \frac{D - \mu}{\sigma} \quad (\text{s6})$$

som er standardnormalfordelt. Ved innsetting blir sannsynligheten  $P_{bom} = 2P(Z > \frac{1}{\sigma})$ . Siden  $\sigma = u_d \approx 2 \text{ m}$ , skal man dermed beregne:

$$P_{bom} = 2P(Z > 0.5) = 2[1 - P(Z \leq 0.5)] \quad (\text{s7})$$

Vi ser fra tabellen at  $P(Z \leq 0.5) = 0.69$ , slik at  $P_{bom} = 2 \times 0.31 = 0.64$ . Det er altså 64% sjanse for at prosjektilet bommer under de gitte betingelsene.