Oblig. 1 STK1100 våren 2026

Matematisk institutt, Universitetet i Oslo

Obligatorisk oppg. 1: **STK1100 – Sannsynlighetsregning og statistisk modellering**

Med: **Nils Lid Hjort og Emil Aas Stoltenberg**

Innleveringsfrist: **5. mars, 2026, 14:00 i Canvas**

Instruksjoner: Du velger selv om du skriver besvarelsen for hånd og scanner besvarelsen, eller om du skriver løsningen direkte inn på datamaskin, for eksempel ved bruk av LaTeX. Besvarelsen skal leveres som én pdf-fil. Scannede ark må være godt lesbare. Besvarelsen skal inneholde navn, emne og oblignummer. I oppgaver der du blir bedt om å programmere, må du legge programkoden inn i pdf-en sammen med resten av besvarelsen.

Samarbeid og alle slags hjelpemidler er tillatt, men den innleverte besvarelsen skal være skrevet av deg og reflektere din forståelse av stoffet. Oppgi gjerne hvilke hjelpemidler du har brukt. Er vi i tvil om du virkelig har forstått det du har levert inn, kan vi be deg om en muntlig redegjørelse.

Oppgave 1. Betrakt mengden $\Gamma = \{h, u, n, d, r, e\}$. Tenk på Γ som en hatt med lapper med bokstaver på som vi trekker fra. Anta at alle lappene i hatten har like stor sjanse for å bli trukket. Du trekker tre lapper med tilbakelegg.

(a) Hva er sannsynligheten for at samme bokstav bli trukket alle tre gangene?

(b) Hva er sannsynligheten for at minst to av trekkene gir samme bokstav?

(c) Hva er sannsynligheten for ingen av de tre trekkene gir samme bokstav?

(d) Eksperimentet i denne oppgaven er fullt ut forståelig, også for folk som ikke tar STK1100. Men vi tar STK1100: Lag et sannsynlighetsrom $(\Omega, \mathcal{A}, \Pr)$ vi kan bruke for å gi en matematisk presis beskrivelse av eksperimentet og hendelsene i denne oppgaven.

Oppgave 2. La $\Gamma = \{h, u, n, d, r, e\}$ være som i Oppgave 1, og innfør også $\Psi = \{p, r, o, s, e, n, t\}$. Definer

$$\Omega = \{(\omega_1, \omega_2) \colon \omega_1 \in \Gamma, \omega_2 \in \Psi\}.$$

(a) Hvor mange delmengder av Ω er det i potensmengden 2^Ω?

(b) Se på funksjonen Pr definert ved

$$\Pr(A) = \frac{|A|}{42}, \quad \text{for } A \in 2^\Omega,$$

der $|A|$ = antall elementer i A. Vis at Pr er et sannsynlighetsmål på $(\Omega, 2^\Omega)$, slik at $(\Omega, 2^\Omega, \Pr)$ er et sannsynlighetsrom.

(c) Konstruer to hendelser A og B, som verken er tomme eller lik Ω, slik at A og B er uavhengige.

(d) Konstruer tre hendelser A og B og C, som verken er tomme eller lik Ω, slik at A, B, C er uavhengige.

(e) Hvor mange uavhengige hendelser er det mulig å konstruere på dette sannsynlighetsrommet?

(f) Lar det seg gjøre å konstruere tre hendelser A og B og C, som verken er tomme eller lik Ω, slik at A, B, C er parvis uavhengige, men ikke uavhengige?

Oppgave 3. Vi fortsetter med sannsynlighetsrommet $(\Omega, 2^\Omega, \Pr)$ fra Oppgave 2. La A og B være de to uavhengige hendelsene du konstruerte i Oppgave 2(c), og definer den tilfeldige variabelen

$$X(\omega) = I_A(\omega) + I_B(\omega), \quad \omega \in \Omega.$$

Her er I_A indikatorfunksjonen til hendelsen A, det vil si at $I_A(\omega) = 1$ når ω er i A, og $I_A(\omega) = 0$ når ω ikke er i A; og I_B er indikatorfunksjonen til hendelsen B.

(a) Hvilke $x \in \mathbb{R}$ er slik at hendelsen $\{X = x\}$ har positiv sannsynlighet?

(b) Regn ut forventningen og variansen til X?

(c) Finn sannsynlighetsmassefunksjonen (pmf-en) til X. Finn også den kumulative fordelingsfunksjonen (cdf-en) til X og lag en skisse av den (evt. en helt presis skisse i Python eller R).

(d) I Python eller R, for hver $n \geq 4$ opp til et passende stort tall, gjør uavhengige trekninger av $X_1, \ldots, X_n$ der hver av disse har samme fordeling som X. La $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ være det empiriske gjennomsnittet av disse. Lag et plot med n på x-aksen og dine simulerte $\bar{X}_n$ på y-aksen.

(e) Forsatt i Python eller R: La $\varepsilon > 0$ være et lite tall som du velger selv, f.eks. $\varepsilon = 0.01$, men prøv noen forskjellige små ε. For en gitt n, forsøk å simuler deg frem til sannsynligheten for hendelsen $\{|\bar{X}_n - \mathrm{E}\, X| \geq \varepsilon\}$. Gjør dette for alle n langs x-aksen i plottet ditt fra (d), og lag et nytt plott med n på x-aksen og dine simulerte anslag på sannsynlighetene til $\{|\bar{X}_n - \mathrm{E}\, X| \geq \varepsilon\}$ på y-aksen. I samme figur, plott også funksjonen $n \mapsto \mathrm{Var}(X)/(\varepsilon^2 n)$. Kommenter.

(f) Til slutt, kan du forklare hvorfor, eller vise at, X er en tilfeldig variabel?

Oppgave 4. I denne oppgaven skal vi forstå mer av hva som foregår i Oppgave 3(d)&(e).

(a) La X være en ikke-negativ tilfeldig variabel med forventning $\mathrm{E}\, X = \mu$. Forklar at for envher hendelse A gjelder det at

$$X \geq X I_A,$$

altså at $X(\omega) \geq X(\omega) I_A(\omega)$ for alle ω. Betrakt hendelsen $A = \{X \geq \varepsilon\}$ der $\varepsilon > 0$ er et eller annet (lite) tall, og vis at

$$\Pr(X \geq \varepsilon) \leq \frac{\mathrm{E}\, X}{\varepsilon}. \tag{1}$$

Forklar at hendelsene $\{X \geq \varepsilon\}$ og $\{X^2 \geq \varepsilon^2\}$ er samme hendelse, og at dette leder til ulikhetene $\Pr(X \geq \varepsilon) \leq \mathrm{E}\, X^2/\varepsilon^2$ for enhver $\varepsilon > 0$, såfremt $\mathrm{E}\, X^2$ er endelig.

(b) La nå Y være en tilfeldig variabel med forventning μ og varians σ^2. Vis at for enhver (liten) $\varepsilon > 0$ har vi at

$$\Pr(|Y - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

(c) La $Y_1, \ldots, Y_n$ være uavhengige tilfeldige variable med forventning μ og varians σ^2. Utled fra det over og det du vet om forventningen og variansen til $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i$ at

$$\Pr(|\bar{Y}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}, \quad \text{for enhver } \varepsilon > 0. \tag{2}$$

(d) Hva mener den norske befolkingen om norsk EU-medlemskap? (Du behøver ikke svare på det.) Vi stoler mer på en spørreundersøkelse som har spurt 1000 tilfeldig utvalget mennesker om 'Ja' eller 'Nei' til EU, enn på en spørreundersøkelse som har spurt 10 tilfeldig utvalget mennesker om det samme. Forklar hvorfor ulikheten i (2) gir oss rett i å stole mer på den første enn den andre spørreundersøkelsen.

(e) Vi sier at en følge $Z_1, Z_2, \ldots$ av tilfeldige variable *konvergerer i sannsynlighet* til a (typisk en konstant) dersom $\Pr(|Z_n - a| \geq \varepsilon)$ konvergerer til null, for enhver $\varepsilon > 0$. En versjon av *De store talls lov* (forkortet LLN på engelsk) sier at dersom $Y_1, \ldots, Y_n$ er uavhengige og identisk fordelte tilfeldige variable med endelig varians, så vil det empiriske gjennomsnittet $\bar{Y}_n$ konvergere i sannsynlighet til $\mathrm{E}\, Y_1$. Bevis, eller forklar at du egentlig allerede har bevist, De store talls lov.

Oppgave 5. Oslo kommune besitter to måleinstrumenter kommunen kan bruke til å teste om vannet på Sørenga sjøbad er rent (bading ok) eller skittent (bading ikke ok). Måleinstrument 1 er nitti prosent sikker på å gjenkjenne skittent vann, men, uheldigvis, så er sannsynligheten for at

den feilaktig sier at vannet er skittent tyve prosent. De tilsvarende tallene for måleinstrument 2 er åtti og ti prosent. Ny forskning fra et laboratorium i en kjeller på Blindern viser at vannet på Sørenga er skittent en av hundre julidager.

(a) Den 1. juli er vannet på Sørenga skittent i følge måleinstrument 1. Hva er sannsynligheten for at vannet faktisk er skittent?

(b) Den 1. juli hører du på *Nyhetsmorgen* at vannet på Sørenga er skittent, men Ugo Fermariello sier ikke om det var måleinstrument 1 eller 2 som ble brukt (og hele Oslo kjenner selvfølgelig sensitiviteten og spesifisiteten til begge instrumentene). Du anslår at det er fifty-fifty om de ene eller det andre måleinstrumentet ble brukt. Hva er ditt anslag på sannsynligheten for at vannet på Sørenga er skittent?

(c) Hvilket av de to måleinstrumentene minimerer sannsynligheten for at Oslo kommune kommer med feil anbefaling til byens badere?

Oppgave 6. En tilfeldig variabel Z har en *standard normalfordeling* dersom dens sannsynlighetstetthet (pdf) er

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}z^2), \quad z \in \mathbb{R}.$$

Det er vanlig å bruke symbolet ϕ for pdf-en til standard normalfordelingen, og Φ for cdf-en til standard normalfordelingen, så $\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \phi(u)\,\mathrm{d}u$ for $z \in \mathbb{R}$. En tilfeldig variabel X er normalfordelt med forventning μ og varians σ^2 dersom X har pdf $f(x) = \phi((x-\mu)/\sigma)/\sigma$. Vi skriver $X \sim \mathrm{N}(\mu, \sigma^2)$ (i boka skriver de $\mathrm{N}(\mu, \sigma)$, altså med standaravviket $\sigma = \sqrt{\sigma^2}$ i stedet for variansen σ^2).

(a) La $Z \sim \mathrm{N}(0,1)$, og definer $Y = \mu + \sigma Z$. Vis at $Y \sim \mathrm{N}(\mu, \sigma^2)$.

(b) Vis at $\Phi(z) = 1 - \Phi(-z)$ for alle z. Dette viser det du kan se av en tegning av pdf-en til en normalfordelingen, nemlig at den er symmetrisk rundt sin forventning.

(c) Vis at $\phi'(z) + z\phi(z) = 0$ for alle z.

(d) For $Y \sim \mathrm{N}(\mu, \sigma^2)$, finn et uttrykk for forventningen $\mathrm{E}\,YI\{Y \ge y\}$.

(e) La Z være en standard normalfordelt tilfeldig variabel. Vis at

$$\Pr(|Z| \ge \varepsilon) \le \frac{2\phi(\varepsilon)}{\varepsilon}, \quad \text{for alle } \varepsilon > 0$$

Sammenlign denne ulikhetene med den du får for $\Pr(|Z| \ge \varepsilon)$ fra (1). Hvilken er best?

Slutt