STK-IN1050 Statistikk for informatikere, V26

Obligatorisk oppgave 3 av 4

Innleveringsfrist: Torsdag, 16. april 2026 kl 14:30 i Canvas (canvas.uio.no).

Instruksjoner

- De obligatoriske oppgavene i dette kurset har fokus på læring og ikke testing. For å få godkjent oppgave er det derfor nok å levere en **blank pdf-fil**, men du må minst levere dette! Hvis du ønsker tilbakemelding på det du har gjort, bør du følge instruksjonene nedenfor. Innkluder gjerne også noen setninger i begynnelsen om hvilke deler du spesielt ønsker tilbakemelding på.
- Du velger selv hvilken programvare du bruker til å skrive oppgaven (for eksempel LaTeX eller Jupyter Notebook). Besvarelsen skal leveres som **én pdf-fil**.
- Besvarelsen skal inneholde navn, emne og oblignummer.
- Det forventes at man har en klar og ryddig besvarelse med tydelige og konkrete begrunnelser. Husk å inkludere alle relevante plott og figurer.
- Følg oppsettet i den foreslåtte malen for besvarelse av hver deloppgave.
- Legg ved programkoden og levere den sammen med resten av besvarelsen. Det er viktig at programkoden du leverer inneholder et kjøreeksempel, slik at det er lett å se hvilket resultat programmet gir. Skriv kommentarer til hver kommando i programkoden du legger ved, dvs. en kort forklaring av hva som skjer når kommandoen brukes.
- Samarbeid og alle slags hjelpemidler er tillatt, men den innleverte besvarelsen skal være skrevet av deg og reflektere din forståelse av stoffet.
- Merk at man har **ett forsøk** på å få oppgaven godkjent.

Søknad om utsettelse av innleveringsfrist

Ved udokumenterte årsaker kan du benytte egenmelding for å få innvilget fire kalenderdager utsettelse på en obligatorisk oppgave. Send i så fall e-post til faglærer (thordist@math.uio.no) innen midnat på innleveringsdagen.

Ved dokumenterte årsaker kan du søke om inntil en uke utsettelse totalt på en obligatorisk oppgave på bakgrunn av gyldig dokumentasjon. Se lenken nedenfor for prosedyre for dette.

https://www.uio.no/studier/eksamen/obligatoriske-aktiviteter/mn-math-oblig.html

Retningslinjer

For å få adgang til avsluttende eksamen i dette emnet, må man bestå alle obligatoriske oppgaver i ett og samme semester. For fullstendige retningslinjer for innlevering av obligatoriske oppgaver, se www.uio.no/studier/admin/obligatoriske-aktiviteter/mn-math-oblig.html

Oppgave 1: Opinionsdynamikker på et nettverk

Oppsett

Vi har N personer koblet sammen i et nettverk. Hver person har til enhver tid én av to mulige meninger: 0 eller 1.

Nettverket beskriver hvem som er naboer (hvem som kan påvirke hvem). Vi bruker en sirkel som nettverksstruktur. Personene nummereres $0, 1, \ldots, N-1$, og person i har naboene $(i-1) \bmod N$ og $(i+1) \bmod N$. Det betyr at hver person har nøyaktig to naboer. Person i hvor $i \in \{1, 2, \ldots, N-2\}$ er nabo med personer $(i-1)$ og $(i+1)$, og personer 0 og $N-1$ er også naboer med hverandre i tillegg til at person 0 er nabo med person 1 og person $N-1$ er nabo med person $N-2$.

Vi skal modellere følgende dynamikk:

- I hvert tidssteg:
  - Velg én tilfeldig person $i \in \{0, 1, \ldots, N-1\}$.
  - Velg deretter tilfeldig én av de to naboene til person i.
  - Person i kopierer (adopterer) naboens mening.
- Prosessen fortsetter til alle har samme mening (konsensus, dvs. alle 0 eller alle 1), eller til vi har nådd maks antall tidssteg.

Bruk følgende parametre:

- Antall personer: $N \in \{10, 25\}$
- Startandel for mening 1: $f_1 \in \{0.3, 0.5, 0.7\}$
- Maks antall tidssteg: $T_{\max} = 10000$

Oppgaver

a) Implementer modellen:
   1) Implementer nettverksstrukturen;
   2) Gi hver person en tilfeldig mening 0 eller 1, med sannsynlighet f_1 for mening 1;
   3) Simuler én realisering fra modellen:
      1) Start fra en intialtilstand;
      2) Utfør tidssteg inntil konsensus er oppnådd, eller maks antall tidssteg er nådd;
      3) Registrer sluttmeningen (0 eller 1) og antall tidssteg.

b) For hver kombinasjon av $N \in \{10, 25\}$ og $f_1 \in \{0.3, 0.5, 0.7\}$, kjør $R = 200$ simuleringer. Vi skal nå estimere to størrelser:
   1) Sannsynlighet p for at mening 1 vinner konsensus:
      - Definer indikatorvariabel:

$$X_r = \begin{cases} 1, & \text{hvis konsensus i simulering } r \text{ er mening 1,} \\ 0, & \text{ellers} \end{cases}$$

      - Definer estimator for p:

$$\hat{p} = \frac{1}{R} \sum_{r=1}^{R} X_r.$$

- Beregn et estimat for p og bruk bootstrap til å beregne et 95% konfidensintervall for p.

2) Forventet tid T til konsensus:
   - La T_r være tid til konsensus i simulering r (ignorer de som ikke når konsensus før $T_{\max}$)
   - Definer estimator for T:
   $$\bar{T} = \frac{1}{R^*} \sum_{r=1}^{R^*} T_r,$$
   hvor R^* er antall simuleringer som faktisk nådde konsensus.
   - Beregn et estimat for T og bruk bootstrap til å beregne et 95% konfidensintervall for T.

c) Påvirker startandelen f_1 sannsynligheten for at mening 1 vinner? Plott estimatene $\hat{p}$ mot f_1 for fast N, med bootstrap-konfidensintervaller. Kommenter: Virker det som om $\hat{p} \approx f_1$?

d) Plott $\bar{T}$ (med konfidensintervall) som funksjon av f_1. Blir tiden til konsensus typisk lengre eller kortere når vi starter med f_1 nær 0.5 enn når f_1 er veldig lav eller veldig høy? Kommenter eventuelle trender.

e) Diskuter kort modellens begrensninger og mulige utvidelser.

Oppgave 2: Simulering av skogbrann på rutenett

Oppsett

Vi skal simulere en stokastisk modell for skogbrann på et kvadratisk rutenett av størrelse $L \times L$. Hver celle i rutenettet representerer et tre, og kan være én av tre tilstander:

- 0: usvidd tre
- 1: brennende tre
- 2: brent tre

Vi bruker en struktur med 4 naboer: til en celle (i, j) hører naboene $(i-1, j)$, $(i+1, j)$, $(i, j-1)$, $(i, j+1)$, hvis disse ligger innenfor rutenettet. Hvis noen av disse cellene ligger utenfor rutenettet, så har cellen færre en 4 naboer.

Bruk følgende parametre:

- Rutenettstørrelse $L = 50$
- Sannsynlighet for at ild sprer seg fra et brennende tre til et usvidd nabotre i ett tidssteg, $q \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$

Initialtilstand (tid $t = 0$):

- Alle celler settes til 0 (usvidde trær).
- Alle trær i venstre kolonne (kolonne 0) tennes: disse cellene settes til 1 (brenner).

Dynamikk (ett tidssteg): Gitt tilstanden ved tid t, lager vi en ny rutenett-tilstand for tid $t + 1$:

1. For hver celle (i, j):

   a. Hvis cellen brenner (har tilstand 1) ved tid t:

   - For hver av de (opptil) 4 naboene: Hvis naboen er usvidd (har tilstand 0) ved tid t, tar den fyr ved tid $t + 1$ med sannsynlighet q.

- Cellen selv blir brent ved tid $t+1$ (får tilstand 2).

b. Hvis cellen ikke brenner ved tid t:

- Kopiér tilstanden til tid $t+1$, med mindre den blir tent av en brennende nabo (se over).

2. Tiden økes med 1.

Simuleringen stopper når det ikke finnes noen brennende celler (ingen 1-ere).

Definisjoner:

- La $Y = 1$ hvis brannen på et eller annet tidspunkt har tent minst ett tre i høyre kolonne (kolonne $L-1$), ellers $Y = 0$.
- La F være andelen brente trær når simuleringen er ferdig:

$$F = \frac{B}{L^2},$$

der B er antall celler i tilstand 2.

Oppgaver

a) Implementer modellen:
   1) Implementer rutenettet (som liste av lister eller `numpy`-array) og en funksjon som initialiserer tilstanden ved tid 0 for gitt L.
   2) Implementer en funksjon som utfører ett tidssteg i modellen for gitt rutenett og parameter q.
   3) Lag en funksjon som kjører én full simulering fra start til slutt, og returnerer:
      - om brannen har nådd høyre kant eller ikke (verdien til Y)
      - hvor stor andel av trærne som til slutt er brent (verdien til F)

b) Monte Carlo–simuleringer: Vi skal undersøke hvordan brannens oppførsel avhenger av q. For hver verdi av q:

   1) Utfør $R = 200$ uavhengige simuleringer, og lagre utfallet Y_r og F_r for hvert simulering r.
   2) Estimer sannsynligheten p for at brannen når høyre kant ved å bruke estimatet:

$$\hat{p} = \frac{1}{R}\sum_{r=1}^{R} Y_r,$$

   og andelen brente trær F ved gjennomsnittlig simulert andel brente trær:

$$\bar{F} = \frac{1}{R}\sum_{r=1}^{R} F_r.$$

c) Bootstrap og konfidensintervaller: For hver verdi av q:

   1) Bruk bootstrap til å konstruere et 95 % konfidensintervall for p, basert på de R observasjonene $Y_1, \ldots, Y_R$.
   2) Bruk bootstrap til å konstruere et 95 % konfidensintervall for F, basert på de R observasjonene $F_1, \ldots, F_R$.

d) Analyse og diskusjon: Bruk resultatene til å svare på følgende:

1) Plott $\hat{p}$ som funksjon av q, med 95 % konfidensintervaller (for eksempel som vertikale feilstolper). Beskriv hvordan sannsynligheten for at brannen når høyre kant endrer seg når q øker. Finn et (omtrentlig) intervall av q-verdier hvor overgangen går fra "nesten aldri" til "nesten alltid".

2) Plott $\bar{F}$ som funksjon av q, med 95 % konfidensintervaller. Diskuter hvordan den gjennomsnittlige andelen brente trær øker med q. Er det stor variasjon (stor usikkerhet) for enkelte q-verdier?

3) Diskuter kort modellens begrensninger og eventuelle utvidelser.