



INSTITUTT FOR MATEMATISKE FAG

TMA4265 - STOKASTISK MODELLERING

Prosjekt 1

Av :
Alper Yarenbasi

Oktober, 2025

Table of Contents

List of Figures	i
List of Tables	i
1 Oppgave 1	1
2 Oppgave 2	5

List of Figures

1 Resultater	2
2 SIR plot	3
3 Histogram for 1000 simuleringer	4
4 95% CI's	4
5 10 plots av Poissonprosess	6
6 10 plots av	6

List of Tables

1 Oppgave 1

(a) S–I–R som Markovkjede

Vi modellerer en person som kan være i tre tilstander: S (susceptible/mottakelig), I (infected/smittet) og R (recovered/immune). La

$$X_n \in \{0, 1, 2\}, \quad n = 0, 1, 2, \dots$$

angi tilstanden ved døgn n , med koding $0 = S$, $1 = I$, $2 = R$. Per døgn antas faste sannsynligheter:

$$\beta \in (0, 1): S \rightarrow I, \quad \gamma \in (0, 1): I \rightarrow R, \quad \alpha \in (0, 1): R \rightarrow S.$$

Markov-egenskap og tids-homogenitet

Gitt nåværende tilstand $X_n = i$ bestemmes fordelingen til X_{n+1} utelukkende av i (ikke hele historien), fordi overgangene per døgn styres av de faste sannsynlighetene (α, β, γ) . Dermed gjelder Markov-egenskapen

$$\Pr(X_{n+1} = j | X_0, \dots, X_n = i) = \Pr(X_{n+1} = j | X_n = i),$$

og kjeden er tids-homogen (stasjonære overganger).

Overgangsmatrise

Med tilstandsrekkefølge $(S, I, R) = (0, 1, 2)$ er endøgns overgangsmatrise

$$P = \begin{pmatrix} 1 - \beta & \beta & 0 \\ 0 & 1 - \gamma & \gamma \\ \alpha & 0 & 1 - \alpha \end{pmatrix}.$$

Begrunnelse: Fra S blir man smittet med sannsynlighet β og ellers værende i S ; direkte $S \rightarrow R$ tillates ikke. Fra I blir man frisk med sannsynlighet γ og ellers værende i I ; direkte $I \rightarrow S$ tillates ikke. Fra R mister man immunitet med sannsynlighet α og ellers værende i R ; direkte $R \rightarrow I$ tillates ikke. Hver rad summerer til 1.

Parameterfortolkning: β : infeksjonssannsynlighet per døgn for en mottakelig (S). γ : tilfrisknings-sannsynlighet per døgn for en smittet (I). α : sannsynlighet per døgn for å miste immunitet for en rekongalesent (R).

(b) long-mean antall dager per år

Hvorfor finnes en limiterende fordeling? Tilstandsrommet er endelig $\{S, I, R\}$. Kjeden er *irreducibel* fordi $S \rightarrow I$ ($\beta > 0$), $I \rightarrow R$ ($\gamma > 0$), $R \rightarrow S$ ($\alpha > 0$) og dermed kan alle tilstander nå hverandre. Den er *aperiodisk* fordi hver tilstand har en positiv selvsløyfe ($P_{SS} = 1 - \beta > 0$, $P_{II} = 1 - \gamma > 0$, $P_{RR} = 1 - \alpha > 0$). For endelige kjeder betyr *irreducibel + aperiodisk* at kjeden er *regular*, og da finnes en unik stasjonær (og dermed limiterende) fordeling π med $P^n \rightarrow \pi$ (Teorem fra [rep.8.pdf](#)).

Stationary distribution: La $\pi = (\pi_S, \pi_I, \pi_R)$ med $\pi = \pi P$ og $\pi_S + \pi_I + \pi_R = 1$. Strukturen $S \rightarrow I \rightarrow R \rightarrow S$ gir

$$\pi_S \beta = \pi_I \gamma = \pi_R \alpha.$$

Dermed er $\pi_S : \pi_I : \pi_R \propto \frac{1}{\beta} : \frac{1}{\gamma} : \frac{1}{\alpha}$,

Fra oppgaveteksten bruker vi $\beta = 0.01$, $\gamma = 0.10$, $\alpha = 0.005$. Da er

$$\frac{1}{\beta} = 100, \quad \frac{1}{\gamma} = 10, \quad \frac{1}{\alpha} = 200, \quad \text{sum} = 310,$$

slik at

$$\pi_S = \frac{100}{310} = \frac{10}{31} \approx 0.32258, \quad \pi_I = \frac{10}{310} = \frac{1}{31} \approx 0.03226, \quad \pi_R = \frac{200}{310} = \frac{20}{31} \approx 0.64516.$$

Long-run mean per år: På årsbasis (365 dager) blir de forventede dagene i hver tilstand

$$\mathbb{E}[\text{dager i } S] = 365 \pi_S \approx 117.74, \quad \mathbb{E}[\text{dager i } I] = 365 \pi_I \approx 11.77, \quad \mathbb{E}[\text{dager i } R] = 365 \pi_R \approx 235.48.$$

(c) Simulering og 95% konfidensintervall for π

Anta $X_0 = S$ (mottakelig ved tid 0). Vi simulerer $T = 20 \cdot 365 = 7300$ døgn. Deretter bruker vi de siste 10 år (3650 døgn) for å estimere long-run mean.

For hvert døgn trekkes neste tilstand kategorisk fra raden i P som svarer til dagens tilstand.

Fra de siste 10 årene (3650 døgn) beregner vi estimatorene

$$\hat{\pi}_S = \frac{1}{3650} \sum \mathbf{1}\{X_n = S\}, \quad \hat{\pi}_I = \frac{1}{3650} \sum \mathbf{1}\{X_n = I\}, \quad \hat{\pi}_R = \frac{1}{3650} \sum \mathbf{1}\{X_n = R\}.$$

Vi gjentar prosedyren uavhengig $n_{\text{runs}} = 30$ ganger (forskjellige seed) for å kvantifisere usikkerhet.

Konfidensintervall: La $\bar{\pi}$ og s være henholdsvis gjennomsnitt og utvalgsstandardavvik av de 30 estimatene for en gitt komponent. Et tosidig 95% t -intervall blir

$$\bar{\pi} \pm t_{0.975, 29} \frac{s}{\sqrt{30}}, \quad t_{0.975, 29} = \text{scipy.stats.t.ppf}(0.975, 29).$$

Seed: Vi bruker `np.random.default_rng(seed)` med *fast base_seed* og *base_seed+run* per kjøring. Dette gir uavhengige 30 simuleringer, men gjør resultatet reproducerbar hvert kjøring av programmet. Hvilket tall seedet er har ikke noe å si. Brukte 4265 på grunn av emnekoden :)

Resultater:

```
pi_S: mean=0.35595, 95% CI=(0.31679, 0.39512)
      days/year: mean=129.92, 95% CI=(115.63, 144.22)

pi_I: mean=0.03358, 95% CI=(0.02977, 0.03739)
      days/year: mean=12.26, 95% CI=(10.87, 13.65)

pi_R: mean=0.61047, 95% CI=(0.56961, 0.65132)
      days/year: mean=222.82, 95% CI=(207.91, 237.73)
```

Figure 1: Resultater

Alle tallene fra oppgave (b) ligger innenfor de estimerte 95%-intervallene fra simuleringen, noe som bekrefter konsistens mellom simulering og utregning.

(d) Er prosessene Markov?

La $N = 1000$ og

$$\beta_n = 0.5 \frac{I_n}{N}, \quad \gamma = 0.10, \quad \alpha = 0.005.$$

På ett døgn skjer tre typer endringer: Nye smittede blant de mottakelige: $\text{Bin}(S_n, \beta_n)$. Nye friske blant de smittede: $\text{Bin}(I_n, \gamma)$. Tap av immunitet blant de immune: $\text{Bin}(R_n, \alpha)$.

Her betyr $\text{Bin}(m, p)$ "antall treff" når vi prøver m ganger og hvert forsøk lykkes med sannsynlighet p . Dermed

$$S_{n+1} = S_n - \text{Bin}(S_n, \beta_n) + \text{Bin}(R_n, \alpha),$$

$$I_{n+1} = I_n + \text{Bin}(S_n, \beta_n) - \text{Bin}(I_n, \gamma),$$

$$R_{n+1} = R_n + \text{Bin}(I_n, \gamma) - \text{Bin}(R_n, \alpha).$$

En prosess er Markov dersom “*for å forutsi neste steg holder det å vite dagens tilstand*” (hele fortiden trengs ikke).

$\{I_n\}$ er **ikke Markov**. Selv om vi kjenner I_n , holder ikke det for å bestemme fordelingen til I_{n+1} , fordi antall nye smittede er $\text{Bin}(S_n, \beta_n)$ og avhenger av S_n . To situasjoner kan ha samme $I_n = 50$ men ulike S_n :

$$\text{A: } S_n = 900 \quad \text{vs.} \quad \text{B: } S_n = 100.$$

Fordelingen til I_{n+1} gitt I_n avhenger også av $S_n \Rightarrow$ ikke Markov.

$\{(S_n, I_n)\}$ er **Markov**. Kjenner vi (S_n, I_n) , så er $R_n = N - S_n - I_n$ bestemt. Parameterne i oppdateringen over (β_n) , samt “antall forsøk” i hver binomial) avhenger dermed bare av (S_n, I_n) . Altså: fordelingen til (S_{n+1}, I_{n+1}) avhenger kun av $(S_n, I_n) \Rightarrow$ Markov.

$\{(S_n, I_n, R_n)\}$ er **Markov**. Dette følger direkte av oppdateringene: gitt (S_n, I_n, R_n) bestemmes fordelingen til neste trinn fullt ut av disse tre tallene (ingen ekstra historikk trengs).

(e) Hva skjer i populasjonen over tid? (0–50 vs. 50–300)

Vi simulerer en befolkning på $N = 1000$ med start $(S_0, I_0, R_0) = (950, 50, 0)$. Hver dag kan

- noen mottakelige (S) bli smittet (I),
- noen smittede (I) bli friske/immune (R),
- noen immune (R) miste immunitet og bli mottakelige (S) igjen.

Sannsynligheten for at en mottakelig smittes den dagen er *proposjonal med hvor mange som er smittet*:

$$\beta_n = 0.5 \cdot \frac{I_n}{N}, \quad \text{mens } \gamma = 0.10 \text{ (tilfriskning)} \text{ og } \alpha = 0.005 \text{ (tap av immunitet)} \text{ er faste.}$$

Intuisjon: jo flere som er smittet, desto lettere er det å bli smittet.

Resultat på simulering

Med samme kode/seed som i figuren fikk vi:

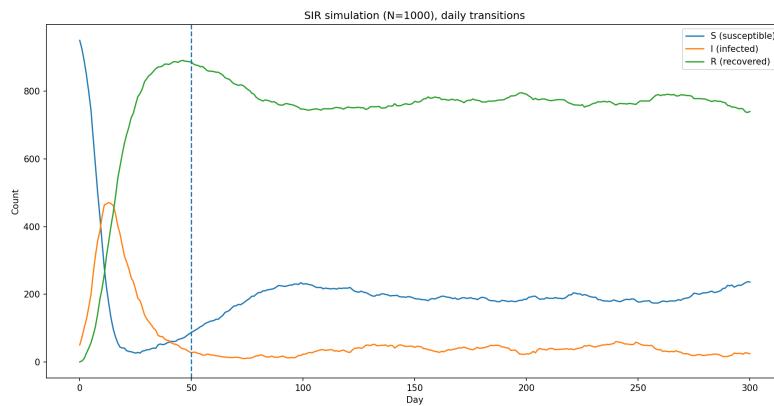


Figure 2: SIR plot

Hvorfor øker I raskt i starten (0–50)? I starten er nesten alle mottakelige ($S_0 = 950$). Da er det mange som *kan* bli smittet, og siden smitteraten per person øker med antall smittede, får vi en selvforsterkende effekt: flere smittede \Rightarrow lettere å smitte flere. Derfor stiger I bratt og når en tidlig topp (her rundt dag 13).

Hvorfor går I ned og blir liten senere (50–300)? Etter den tidlige toppen har mange allerede vært smittet og gått over til R , så det er færre mottakelige igjen. Da blir nye smittede per dag få, mens en fast andel av de smittede blir friske hver dag. Resultatet er at I faller og holder seg lavt. Samtidig lekker noen fra R tilbake til S (tap av immunitet), så systemet stabiliserer seg:

- S legger seg rundt et moderat nivå (i vår bane omtrent 200),
- I holder seg lavt (rundt 30–40 i snitt),
- R er størst og endrer seg sakte.

(f) Topp og tidspunkt for topp over mange simuleringer

Vi simulerer $n = 1000$ uavhengige simuleringer av populasjonsmodellen (samme parametre som i oppg (e) og samme kode).

Resultat

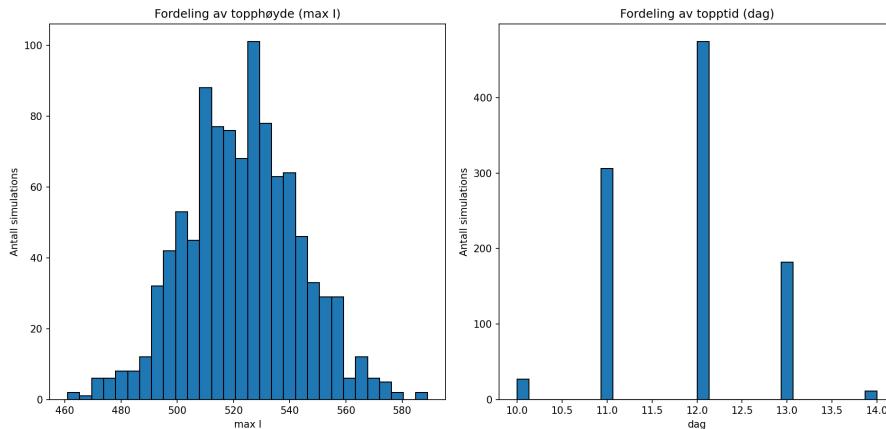


Figure 3: Histogram for 1000 simuleringer

Antall simulations: 1000
Maks smittede I (forventet): 523.14 95% CI: (521.89, 524.39)
Topptid (dag) (forventet): 11.84 95% CI: (11.80, 11.89)

Figure 4: 95% CI's

Tolkning: Den korte topptiden skyldes at andelen mottakelige er høy i starten, slik at smitten vokser bratt, deretter flater den ut når antallet mottakelige reduseres. Tallene er konsistente med observasjonene i (e), men gir nå et kvantitativt bilde av variasjonen mellom simuleringer.

(g) Effekt av vaksinering ($V = 0, 100, 600, 800$)

Oppsett. Vi antar livslang immunitet for V vaksinerte ved tid 0 (de deltar ikke i S–I–R-dynamikken). Blant de *uvaksinerte* introduseres $I_0 = 50$ smittede, resten er mottakelige. Vi simulerer 0–300 døgn med samme parametre som (e)–(f).

Figuren viser $I(t)$ for $V \in \{0, 100, 600, 800\}$. Økt V gir lavere topp og ofte tidligere topp, samt hurtigere demping etter toppen.

Resultat

Vi kjørte $n = 1000$ uavhengige baner for hver V og registrerte topphøyde og første topptid. Estimater:

```
Antall simuleringer per case: 1000

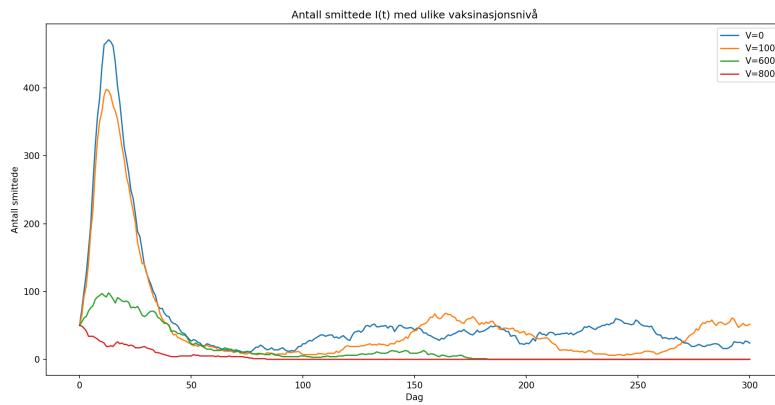
V= 0: max I = 523.14 (95% CI: 521.89, 524.39)
      topptid = 11.84 (95% CI: 11.80, 11.89)

V=100: max I = 438.28 (95% CI: 437.04, 439.51)
      topptid = 12.67 (95% CI: 12.61, 12.73)

V=600: max I = 96.81 (95% CI: 96.04, 97.59)
      topptid = 15.86 (95% CI: 15.59, 16.13)

V=800: max I = 51.20 (95% CI: 51.08, 51.33)
      topptid = 0.98 (95% CI: 0.87, 1.08)

Endring vs V=0 for V=100: E[max I]=-84.86, E[topptid]=0.83 dager
Endring vs V=0 for V=600: E[max I]=-426.33, E[topptid]=4.01 dager
Endring vs V=0 for V=800: E[max I]=-471.94, E[topptid]=-10.87 dager
```



For $V = 800$ blir toppen i praksis ved start (dag 0–1), fordi svært få er mottakelige og smitten nesten ikke sprer seg.

Tolkning. Vaksinering reduserer antall mottakelige ved start. Dermed faller smittepresset, som gir lavere topphøyde og, for moderate V , noe senere topp. For høyt V (som 800) sprer smitten seg knapt i det hele tatt, så maksimum ligger helt i starten.

2 Oppgave 2

(a) Antall insurance claims

La $\{X(t) : t \geq 0\}$ være en Poissonprosess med konstant rate $\lambda(t) \equiv 1.5$ (per dag). Mars 1 kl. 00:00 er $t = 59$ dager, så

$$X(59) \sim \text{Poisson}(\mu), \quad \mu = \lambda t = 1.5 \cdot 59 = 88.5$$

Eksakt sannsynlighet. Vi skal beregne $P\{X(59) > 100\}$:

$$P\{X(59) > 100\} = 1 - \Pr\{X(59) \leq 100\} = 1 - \sum_{k=0}^{100} e^{-\mu} \frac{\mu^k}{k!}.$$

Med $\mu = 88.5$ får vi numerisk

$$P\{X(59) > 100\} \approx 0.1028.$$

Simulering

```
mu = 88.5
Exact P[X(59) > 100] = 0.102822
Estimated P[X(59) > 100] from 1000 runs = 0.083000
```

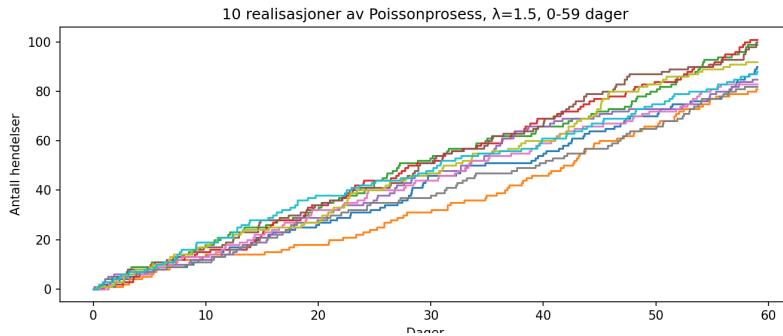


Figure 5: 10 plots av Poissonprosess

Sammenligning og kommentar: Eksakt verdi er $P\{X(59) > 100\} = 0.102822$. Med $n = 1000$ simuleringer fikk vi 0.083, noe lavere enn eksakt verdi, men forventet gitt at det er standardfeil. Øker vi til $n = 10^6$ får vi 0.102362, som ligger svært nær den eksakte verdien.

(b) claim exceed 8 mill. kr

Med $n = 1000$ simuleringer fikk vi $\Pr\{Z(59) > 8\} = 0.743$.

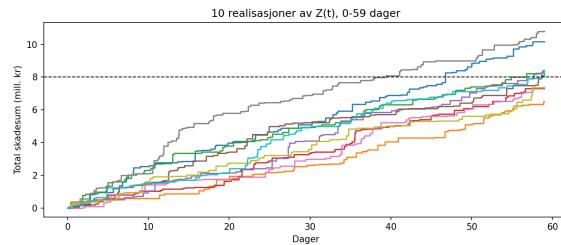


Figure 6: 10 plots av