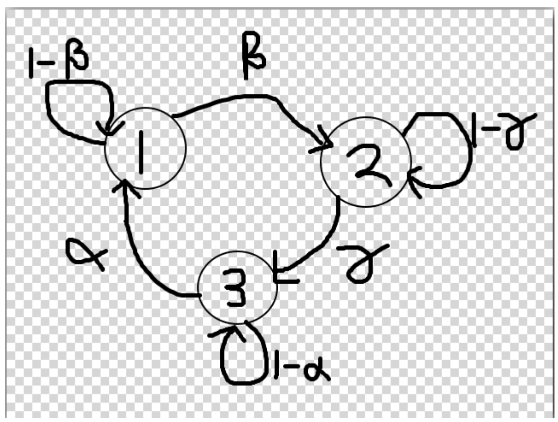
Stokastisk modellering prosjekt 1

Ole Kristian Skogly og Jo Andersson Stokke

Oktober 2020

1 Oppgave 1



1a) { $X_n : n = 0, 1, ...\}$ er en Markov kjede siden hver hendelse avhenger kun av forrige hendelse. P er en "transition" sannsynlighetsmatrise siden man alltid kan forbli i den samme tilstanden eller gå over til neste tilstand.

1b) { $X_n : n = 0, 1, ...\}$ er en "irreducible" Markov kjede fordi alle tilstander kommuniserer med hverandre(se firgur). Får kun en "equivalence class" som er $\{1, 2, 3\}$, der 1,2,3 henholdsvis tilhører tilstandene "susceptible" (S), "infected"(I) og "recovered and immune"(R). Ekvivalent klassen er "recurrent" fordi vi kan komme tilbake til en tilstand med sannsynlighet 1 etter å ha forlatt den. Hver tilstand har periode 1 og er derfor aperiodisk fordi det ikke er noen krav for hvor lang tid det tar fra en går fra en state til en annen. Evt kan man se at gcd($P^1[1], P^2[1], P^3[1], ...$) =1.

1c) Vi antok her at $\beta = 0,05$, $\gamma = 0,10$ og $\alpha = 0,01$. Regner ut først den forventede tiden til en "susceptible" person blir "infected": Har at dette er en geometrisk fordeling med forventning $1/P$, altså $E[P_{12}]$=$1/\beta = 1/0,05 = 20 dager$.Kunne også regnet dette ut ved å bruke first step analysis:

$V_i = E[T|x_0 = i]$. Setter $V_2 = 0$.Får da følgende tre ligninger:

$V_1 = 1 + \sum_{j=0}^{3} V_j P_{1j} = 1 + V_1 P_{11} + V_2 P_{12} + V_3 P_{13} = 1 + 0,95 V_1.$

$V_2 = 1 + \sum_{j=0}^{3} V_j P_{2j} = 1 + V_1 P_{21} + V_2 P_{22} + V_3 P_{23} = 1 + 0.1 V_3.$

$V_3 = 1 + \sum_{j=0}^{3} V_j P_{3j} = 1 + V_1 P_{31} + V_2 P_{32} + V_3 P_{33} = 1 + 0,01 V_1 + 0,99 V_3$

(1)

Løser ligningssystemet og får at V_1= 20 dager. Dette stemmer godt med utregningen over, der forventningsverdien ble regnet ut dirkete.

Ønsker videre å regne ut forventet tid det tar til en "susceptible" person blir "infected": Bruker igjen at det er en geometrisk fordeling med forventning $1/P$, altså $E[P_{13}]$=$1/\beta$ + $1/\gamma = 1/0,05 + 1/0,10 = 30 dager$.

Kan også regne ut dette ved å bruke first step analysis: $V_i = E[T|x_0 = i]$. Setter V_3= 0. Får da følgende tre ligninger:

$V_1 = 1 + \sum_{j=0}^{3} V_j P_{1j} = 1 + V_1 P_{11} + V_2 P_{12} + V_3 P_{13} = 1 + 0.95 V_1 + 0,05 V_2.$

$V_2 = 1 + \sum_{j=0}^{3} V_j P_{2j} = 1 + V_1 P_{21} + V_2 P_{22} + V_3 P_{23} = 1 + 0.9 V_2$

$V_3 = 1 + \sum_{j=0}^{3} V_j P_{3j} = 1 + V_1 P_{31} + V_2 P_{32} + V_3 P_{33} = 1 + 0,1 V_1$

(2)

Løser ligningssystemet som disse ligningene gir og får at $V_1 = 30$ dager, som stemmer godt med det som ble regnet ut direkte.

Ønsker så regne ut forventet tid det tar for en person og gjennomgå en hel cycle. Igjen kan forventing E[P_{23}] = $1/\gamma$ regnes ut ved at det er en geometrisk fordeling med parameteren p=γ

Tilsvarende kan gjøres med E[P_{31}] = $1/\alpha$ regnes ut ved at det er en geometrisk fordeling med parameteren p=α E[P_{12} + P_{23} + P_{31}] = $1/\beta + 1/\gamma + 1/\alpha = 1/0.05 + 1/0.01 + 1/0.1 = 130$ dager.

1d) Simulerer med vedlagt python kode. Resultatene er som følger:

Forventing P_12	Forventing P_23	Forventing P_31
18.531690140845072	10.126760563380282	100.34397163120568
18.93719806763285	9.942028985507246	103.7420924574209 2
19.151785714285715	10.030357142857143	101.11690647482014
19.953257790368273	10.18271954674221	99.17974322396576
19.803944315545245	10.125290023201856	97.22780373831776
20.104145601617795	10.223458038422649	99.0478615071283
19.869488536155202	10.216049382716049	98.77264653641208
20.099610894941634	10.243579766536964	97.52978056426332
20.218793828892007	10.1164095371669	97.74929378531074
20.146435452793835	10.122029543994861	98.8143596377749

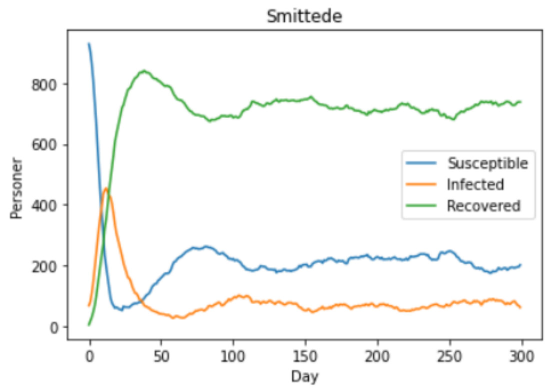
Dette stemmer veldig godt med teoretiske resultater fra oppgave 1c.

1e) S_n er ikke alene en markovkjede, da informasjon om S_n ikke er tilstrekkelig for å kunne si noe om S_{n+1}, da S_{n+1} både avhenger av S_n , R_n og I_n i og med at β er en funksjon av I_n

I_n er ikke alene en markovkjede, da informasjon om I_n ikke er tilstrekkelig for å kunne si noe om I_{n+1}, da I_{n+1} både avhenger av I_n og S_n

Dette er også opplagt ved å bemerke seg at P er irreduserbar, og kan nettop derfor ikke deles inn i mindre markovkjeder.

1f)



Ser utifra figuren (som viser simuleringen) at antall personer "recovered" går opp i intervallet 0-50dager, mens antall personer "susceptible" går ned. Ser også at vi i det samme intervallet har at personer som er "infected" får en tidlig økning før antallet går ned igjen. I intervallet 50-300 dager har antallet stablisert seg mer og vi har ikke lenger en ekstrem endring. Utifra denne modellen vil derfor det i starten være en "smittebombe" som etterhvert går mere under kontroll da flere blir immune. Dette stemmer godt med hva som ser ut til å ofte være tilfellet ved epidimier. Trenger ikke se lenger enn COVID19 for å se en lignende eksplosjon i starten. Til hvilken grad stabilisering oppnår ved såkalt "herdeimunitet" gjennstår å se, men om det er ihvertfall resultatet om vi velger å bruke modellen vi har brukt i denne oppgaven.

1g)

```
State of Y_n for the last 10 days of the 100 year period
[216.  57. 727.]
[220.  52. 728.]
[220.  53. 727.]
[212.  58. 730.]
[213.  59. 728.]
[211.  61. 728.]
[214.  57. 729.]
[210.  58. 732.]
[214.  56. 730.]
[210.  53. 737.]
```

Tabellen over viser simuleringen vi fikk ved å se på en simulering over 100 år, der vi har tatt med de 10 siste dagene. Etter å ha sett på simuleringen så ser vi at i det lange løp så forventer vi at ca. 21,4% personer vil være "susceptible", ca. 5,1% personer vil være "infected" og ca. 73% personer vil være

1h) I utklippet nedenfor står simuleringen for forventet maksimal antall "infected" personer og forventet tid dette vil ta. Kan også se at dette stemmer overens med grafen som ble simulert i oppgave 1f.

```
Forventet max infected fra 1000 simuleringer er: 447.681 personer
Forventet min tid til max infected fra 1000 simuleringer er: 11.736 dager
```

Den første observatoren sier spesielt noe om hvor alvorlig en smitte kan være i omfang. Et stort maks-antall smittede på et tidspunkt forteller at taket for skade kan være stor. Den andre observatoren sier derimot nærmere hvor raskt viruset sprer seg og noe om virusets dødlighet. Hvorfor? Om et virus er av det dødligere slaget, så vil hosten ikke kunne spre viruset videre like lenge og like effektivt, og viruset vil ikke overleve for lenge. Dette impliserer at maks smittede vil nås relativt tidlig. Er denne observatoren høy derimot, så kan det si noe om virusets evne til å holde hosten levende, og maks antall smittede vil da nås relativt sent. Stemmer dette, så kan disse to observatorene sammen beskrive en epidimi ganske godt.

2 Oppgave 2

2a) Skal regne ut P(X>100) til X $\sim$ pois(λt).

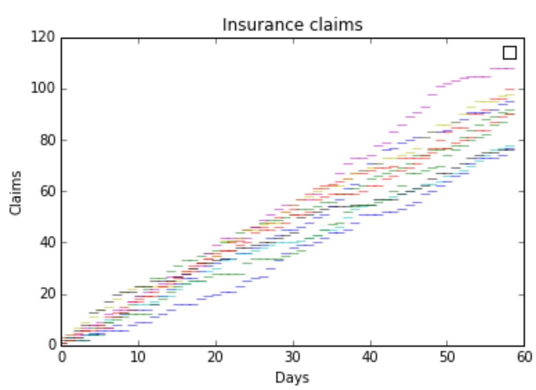
Det er da naturlig å bruke den kumulative distribusjonsfunksjonen til poissonfordelingen.

$$P(X <= x) = CDF_{P}oisson(\lambda t) = \sum_{i=0}^{x} e^{-\lambda t}(\lambda t)^{i}/i! \tag{3}$$

$$P(X > 100) = 1 - P(X <= 100) = 1 - \sum_{i=0}^{x} e^{-\lambda t}(\lambda t)^{i}/i! \tag{4}$$

Implementerer denne CDFen i python med intensiteten λ = 1.5 og dager t = 59. Simulerer så med vedlagt kode 1000 ganger. Holder styring på hver gang simulasjonen endte med over 100 forsikringskrav, og regner ut gjennomsnittet med å dele på antall simulasjoner. Ut detter følgende teoretisk forventing og gjennomsnitt fra 1000 simulasjoner.

```
Teoretisk forventning:
0.102822200365
Gjennomsnitt fra simulation:
0.105
```



2b) Får oppgitt at N $\sim$ pois(λt)
$C_i \sim$ exp(γ)
og at $\lambda = 1.5$, t = 59 og at $\gamma = 10$
Bruker kjente formler for forventing og varianse til poiss og exp distribusjoner.
E[N] = λt, Var[N] = λt, E[C_i] = $1/\gamma$, Var[C_i] = $1/\gamma^2$
og veldig viktig at N og C_i er uavhengige stokatiske variabler.
Da sier loven av total forventning:

$$E[C] = E[E[\sum_{i=0}^{N}[C_i|N=n]] = E[\sum_{i=0}^{N} E[C_i|N=n]] = E[N]E[C_i] = \lambda t/\gamma = 8.85 mill.NOK \tag{5}$$

Og loven av total Varianse sier at:

$$Var[C] = E[Var[\sum_{i=0}^{N}[C_i|N=n]]+Var[E[\sum_{i=0}^{N}[C_i|N=n]] = E[NVar[C_i]+Var[NE[C_i]$$

$= \mathrm{E}[\mathrm{N}/\gamma^2]+Var[N/\gamma^2] = E[N]/\gamma^2+Var[N]/\gamma^2 = 2\lambda t/\gamma^2 = 1.77 mill.NOK$(6)
Nedenfor har vi simulasjone av forventning og varians:

Forventning:	Varianse
8.912124010038037	1.791869535694675
8.832131344304821	1.8546589043154817
8.970750024052737	1.8018402141991563
8.893548952820082	1.7654488050294004
8.83097977769668	1.8054418708477884
8.810980749817313	1.8299218007868718
8.870105056551816	1.7591496541252238
8.808729295018104	1.6994648310773237
8.950495979026734	1.780143942453343
8.820280695023696	1.6579963365587977

Vi ser ved å sammenligne simulasjonene med beregningene at dette stemmer godt overens med hverandre.