Stockholms universitet

Epidemispridning i skolan

Elin Kull

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2018:11
Matematisk statistik
Juni 2018

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematisk statistik
Stockholms universitet
Kandidatuppsats **2018:11**
http://www.math.su.se/matstat

Epidemispridning i skolan

Elin Kull*

Juni 2018

Sammanfattning

I denna uppsats undersöks genom simuleringar hur epidemispridning beter sig i skolan. Hur påverkar storleken på skolklasserna storleken på sjukdomsutbrotten? Hur är storleken på utbrotten beroende av varandra över tid? Två modeller för elevomsättningen på skolan jämförs, en mer realistisk modell där eleverna omsätts årskursvis, och en slumpmässig modell där eleverna i antal omsätts motsvarande en årskurs i taget men slumpmässigt utvalda. Det visar sig att utbrotten blir betydligt större då skolklasserna är större. Det visar sig även vara en tydlig skillnad mellan de två modellerna för elevomsättingen. Utbrotten blir större vid simuleringar där den realistiska modellen för elevomsättningen tillämpas än där den slumpmässiga modellen tillämpas.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: elin_musik@live.se. Handledare: Pieter Trapman och Daniel Ahlberg.

Förord

Det här är en kandidatuppsats i matematisk statistik om 15 högskolepoäng skriven vid matematiska instutitionen på Stockholms Universitet. Jag vill tacka mina handledare Pieter Trapman och Daniel Ahlberg vars idéer, inspiration och handledning under arbetets gång har betytt mycket.

Innehåll

1 Introduktion **3**
1.1 Arbetets upplägg 3

2 Reed-Frostmodellen **3**

3 Smittspridning i skolan **5**
3.1 Sannolikheten att en elev smittas av någon annan elev på skolan 5
3.2 Uppdelade S-, I- och R-värden 6
3.3 Introduktion av utbrotten 8

4 2 olika modeller för årsbyte **9**
4.1 Den realistiska modellen 9
4.2 Den slumpmässiga modellen 9
4.3 Jämförelse mellan modellerna 9

5 Reproduktionstalet $\mathbf{R_0}$ **10**

6 Simuleringar och resultat **11**
6.1 Smittspridningsparametrarna 11
6.2 10 elever/klass 11
6.3 20 elever/klass 11
6.4 Smittspridningsparametrarnas påverkan 13
6.4.1 $\mathbf{R_0}$ 15
6.5 Antalet smittade elever i en årskurs över hela skoltiden 15
6.5.1 Monte Carlo-simuleringar 18
6.6 Antalet immuna elever som slutar 20

7 Diskussion **22**
7.1 Slutsatser 22
7.2 Förslag till ytterligare frågeställningar 22

1 Introduktion

Skolan är en plats där många barn samlas och därför en gynnsam miljö för sjukdomar att spridas i. Det är därför intressant att undersöka hur smittspridning beter sig och vad som påverkar den i denna miljö. I det här arbetet tittar vi på hur storleken på flera utbrott av en sjukdom är beroende av varandra över tid. Vi undersöker även hur storleken på skolklasserna och hur sjukdomarnas smittsamhet påverkar storleken på utbrotten. För att kunna göra det här sätter vi upp en modell för ett utbrottsförlopp i en skola där det tas hänsyn till att elever som går i samma klass generellt sett smittar varandra med en högre sannolikhet än elever som inte går i samma klass, och att elever som går i samma årskurs generellt sett smittar varandra med en högre sannolikhet än elever som går i olika årskurser.

När vi sätter upp en modell vill vi att den ska vara så enkel som möjligt men utan att det sker på bekostnad av att resultaten vi får då vi använder modellen kommer allt för långt ifrån verkligheten. Då vi vill undersöka hur storleken på utbrotten förändras över tid behöver vi en modell för omsättningen av elever på skolan. Här har vi valt att jämföra två olika modeller, en där eleverna omsätts årskursvis och en där eleverna omsätts slumpvis. Då de omsätts slumpvis väljs slumpmässigt ett antal elever motsvarande en årskurs ut som får sluta, alla elever på skolan väljs med lika stor sannolikhet. I båda modellerna byts de elever som slutar ut mot nya elever som alla är mottagliga för smittan. Om resultaten mellan dessa modeller inte skiljer sig nämnvärt åt är användning av den slumpvisa modellen att föredra. Det framförallt för att den slumpvisa modellen är enklare att räkna på men även för att den skulle kunna tillämpas mer generellt för att t.ex. simulera smittspridningen på en arbetsplats. Utifrån dessa modeller genomförs sedan simuleringar.

1.1 Arbetets upplägg

I avsnitt 2 presenteras Reed-Frost modell för spridning av smittsamma sjukdomar. I avsnitt 3 används Reed-Frost modell till utveckling av en modell som bättre motsvarar sjukdomsspridningen på en skola. I avsnitt 4 presenteras den realistiska och den slumpmässiga modellen för vilka elever som börjar och slutar på skolan ett nytt år. I avsnitt 5 ges en kort presentation av reproduktionstalet R_0. I avsnitt 6 beskrivs de simuleringar som har gjorts och vad de gav för resultat. I det 7:e och avslutande avsnittet diskuteras resultaten och vad som skulle kunna undersökas framöver inom samma område men som inte har täckts av detta arbete.

2 Reed-Frostmodellen

En av de enklaste modellerna för spridning av en smittsam sjukdom i en liten grupp av individer kallas för Reed-Frostmodellen. Namnet Reed-Frost

kommer från de två grundarna Lowell Reed och Wade Hampton Frost som formulerade modellen år 1928, se (1). Modellen är en så kallad kedje-binomial SIR-modell. Det betyder att det finns tre möjliga tillstånd som individerna i modellen kan befinna sig i. Bokstäverna S, I och R står för var sitt tillstånd. Individerna är först mottagliga (**S***usceptible*) för sjukdomen. Om en individ blir smittad kommer denna först vara smittsam, och kallas för en smittspridare (**I***nfective*), för en viss tid och sedan tillfriskna och bli immun, motsvarande borttagen (**R***emoved*).

I modellen används oftast diskreta tidssteg. Eftersom det är en liten grupp av individer antas alla individer ha lika mycket kontakt med varandra och därför också sprida en smitta mellan varandra med samma sannolikhet p. En individ som blir smittad antas vara smittad under ett tidsteg. Under det tidsteget smittar denna individ varje annan mottaglig individ med sannolikheten p oberoende av varandra. Antalet mottagliga individer under ett tidssteg t betecknas med S_t, anatalet smittspridare betecknas med I_t och antalet immuna eller borttagna individer betecknas med R_t. Till följande tidssteg $t+1$ blir

$$I_{t+1} \sim \mathrm{Bin}(S_t, 1-(1-p)^{I_t})$$

eftersom det finns S_t mottagliga individer som kan bli smittade. Var och en av dessa individer blir smittad av någon annan smittad individ med sannolikheten $1-(1-p)^{I_t}$. $1-p$ är sannolikheten att en mottalig individ inte blir smittad av en specifik smittspridare. Eftersom kontakten mellan en mottaglig individ och de olika smittspridarna är oberoende av varandra är sannolikheten att en mottaglig individ <u>inte</u> blir smittad av någon smittspridare $(1-p)^{\text{antalet smittspridare}} = (1-p)^{I_t}$. Eftersom händelsen att en mottaglig individ blir smittad av någon annan smittspridare är komplementhändelsen till att en mottaglig individ <u>inte</u> blir smittad av någon smittspridare är sannolikheten för att en mottaglig individ blir smittad av någon smittspridare $1-(1-p)^{I_t}$. Att I_{t+1} är en binomialfördelad slumpvariabel är anledningen till att modellen kallas för kedje-binomial. Vidare blir antalet mottagliga och antalet immuna i det kommande tidssteget

$$S_{t+1} = S_t - I_{t+1}$$

och

$$R_{t+1} = R_t + I_t$$

eftersom antalet mottagliga individer minskar med det antalet som blir smittade och antalet immuna individer ökar med antalet smittade från föregående tidssteg som nu tillfrisknar. Se (2, s. 4-6) för en närmare beskrivning av modellen.

3 Smittspridning i skolan

3.1 Sannolikheten att en elev smittas av någon annan elev på skolan

När vi simulerar smittspridningen på skolan utgår vi från Reed-Frost modell för spridning av en smittsam sjukdom. I Reed-frost modell antas det att smittan sprids med samma sannolikhet mellan samtliga individer. På en skola med flera klasser och årskurser vore detta dock inte ett rimligt antagande. Vi antar istället att elever som går i samma klass smittar varandra med en sannolikhet, att elever som går i samma årskurs men i olika klasser smittar varandra med en annan något mindre sannolikhet och att elever som går i olika årskurser smittar varandra med en tredje och ytterligare något mindre sannolikhet.

Sannolikheten att en enskild elev blir smittad av någon annan elev på skolan i ett tidsteg under epidemin är lika med sannolikheten för unionen av händelserna att eleven blir smittad av någon elev som går i samma klass, någon elev som går i samma årskurs men i en annan klass eller någon elev som går i en annan årskurs. Vi betecknar händelsen att en elev blir smittad av någon klasskamrat med H_k, händelsen att en elev blir smittad av någon elev som går i samma årskurs men i en annan klass med H_a och händelsen att en elev blir smittad av någon elev i en annan årskurs med H_r. Vi betecknar sannolikheten att en elev blir smittad av någon klasskamrat med $\mathbb{P}(H_k)$, sannolikheten att en elev blir smittad av någon elev som går i samma årskurs men i en annan klass med $\mathbb{P}(H_a)$ och sannolikheten att en elev blir smittad av någon elev som går i en annan årskurs med $\mathbb{P}(H_r)$. Enligt principen om inklusion och exklusion är sannolikheten för att en elev blir smittad av någon annan elev på skolan, $\mathbb{P}(H_k \cup H_a \cup H_r)$, lika med

$$\mathbb{P}(H_k) + \mathbb{P}(H_a) + \mathbb{P}(H_r) - (\mathbb{P}\left(H_k \cap H_a\right) + (\mathbb{P}\left(H_k \cap H_r\right) + (\mathbb{P}\left(H_a \cap H_r\right)) +$$

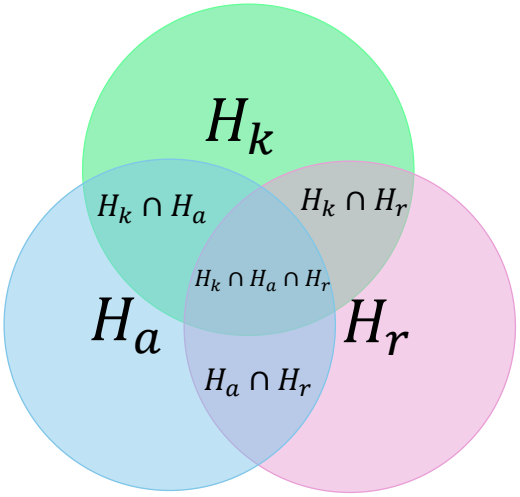
$$\mathbb{P}\left(H_k \cap H_a \cap H_r\right).$$

Eftersom $\mathbb{P}(H_k)$, $\mathbb{P}(H_a)$ och $\mathbb{P}(H_r)$ är oberoende av varandra är sannolikheterna för snitten av de olika händelserna lika med produkterna av sannolikheterna var för sig, vilket ger oss att

$$\mathbb{P}(H_k \cup H_a \cup H_r) =$$

$$\mathbb{P}(H_k)+\mathbb{P}(H_a)+\mathbb{P}(H_r)-(\mathbb{P}\left(H_k\right) \cdot \mathbb{P}\left(H_a\right) + \mathbb{P}\left(H_k\right) \cdot \mathbb{P}\left(H_r\right) + \mathbb{P}\left(H_a\right) \cdot \mathbb{P}\left(H_r\right)) +$$

$$\mathbb{P}\left(H_k\right) \cdot \mathbb{P}\left(H_a\right) \cdot \mathbb{P}\left(H_r\right).$$



Figur 1: Sannolikheten att en elev blir smittad av någon annan elev på skolan i ett tidssteg är lika med $\mathbb{P}(H_k \cup H_a \cup H_r)$

3.2 Uppdelade S-, I- och R-värden

Eftersom sannolikheterna för att klasskamrater smittar varandra, elever som går i samma årskurs men i olika klass smittar varandra och elever som går i olika årskurs smittar varandra skiljer sig åt får vi för tidsstegen i ett utbrottsförlopp inte bara ett S-värde, ett I-värde och ett R-värde för hela skolan, utan även S-,I- och R-värden för varje årskurs och varje klass. Vi låter A_A beteckna antalet årskurser på skolan. Då kan vi beteckna S-, I- och R-värdena för årskurserna vid tidssteg t med

$$S_t^i,\ I_t^i \text{ och } R_t^i \text{ där } i = 1, 2, ..., A_A$$

där i står för den årskurs som värdena tillhör. Vidare låter vi A_K beteckna antalet klasser i varje årskurs och kan då beteckna S-, I- och R-värdena för varje klass med

$$S_t^{i,j},\ I_t^{i,j} \text{ och } R_t^{i,j} \text{ där } i = 1, 2, ..., A_A \text{ och } j = 1, 2, ..., A_K$$

där i står för den årskurs och j står för den klass i den årskursen som värdena tillhör.

Då antalet smittade elever i en klass uppdateras från ett tidssteg t till tidssteg $t + 1$ får vi att

$$I_{(t+1)}^{i,j} \sim \text{Bin}(S_t^{i,j}, \mathbb{P}(H_k \cup H_a \cup H_r)^{i,j})$$

där $S_t^{i,j}$ är antalet mottagliga elever i klass i, j under det föregående tidssteget och alltså potentiella blivande smittspridare till detta tidssteg, och

$\mathbb{P}(H_k \cup H_a \cup H_r)^{i,j}$ är sannolikheten att en mottaglig elev i denna klass blir smittad av någon annan elev på skolan, vilken är indexerad med i, j eftersom den kommmer att skilja sig åt mellan de olika klasserna. $\mathbb{P}(H_k \cup H_a \cup H_r)^{i,j}$ beräknas med hjälp av den formel som beskrivs i avsnitt 3.1 och

$$\begin{aligned}
\mathbb{P}(H_k)^{i,j} &= 1 - (1 - p_k)^{I_t^{i,j}} \\
\mathbb{P}(H_a)^{i,j} &= 1 - (1 - p_a)^{(I_t^i - I_t^{i,j})} \\
&\text{och} \\
\mathbb{P}(H_r)^{i,j} &= 1 - (1 - p_r)^{(I_t - I_t^i)} \\
&\text{där}
\end{aligned}$$

$$\begin{aligned}
p_k &= \text{sannolikheten att en elev blir smittad av en specifik klasskamrat} \\
p_a &= \text{sannolikheten att en elev blir smittad av en specifik elev som går} \\
&\quad \text{i samma årskurs men i en annan klass} \\
p_r &= \text{sannolikheten att en elev blir smittad av en specifik elev som går} \\
&\quad \text{i en annan årskurs} \\
I_t^{i,j} &= \text{antalet smittade elever i klass } i, j \text{ under tidssteg } t \\
I_t^i - I_t^{i,j} &= \text{antalet smittade elever som går i årskurs } i \\
&\quad \text{men inte i klass } i, j \text{ under tidssteg } t \\
I_t - I_t^i &= \text{antalet smittade elever som går i någon annan årskurs} \\
&\quad \text{än årskurs } i \text{ under tidssteg } t
\end{aligned}$$

Vidare får vi precis som i Reed Frost modell att

$$\begin{aligned}
S_{(t+1)}^{i,j} &= S_t^{i,j} - I_{(t+1)}^{i,j} \\
&\text{och} \\
R_{(t+1)}^{i,j} &= R_t^{i,j} + I_t^{i,j}
\end{aligned}$$

Och vi får att de uppdaterade S-, I- och R-värdena för årskurserna och hela skolan blir

$$S_{(t+1)}^i = \sum_{j=1}^{A_K} S_{(t+1)}^{i,j}, \qquad I_{(t+1)}^i = \sum_{j=1}^{A_K} I_{(t+1)}^{i,j}, \qquad R_{(t+1)}^i = \sum_{j=1}^{A_K} R_{(t+1)}^{i,j}$$

och

$$S_{(t+1)} = \sum_{i=1}^{A_A} S^i_{(t+1)}, \qquad I_{(t+1)} = \sum_{i=1}^{A_A} I^i_{(t+1)}, \qquad R_{(t+1)} = \sum_{i=1}^{A_A} R^i_{(t+1)}.$$

Eftersom antalet mottagliga, smittade och immuna elever i en årskurs är lika med summorna av antalet mottagliga, smittade och immuna elver i varje klass i årskursen, och antalet mottagliga, smittade och immuna elever i skolan är lika med summorna av antalet mottagliga, smittade och immuna elever i varje årskurs på skolan.

3.3 Introduktion av utbrotten

En epidemi kommer inte att bryta ut på skolan förens någon elev blir smittad av en sjukdom utanför skolan för att sedan sprida den vidare bland övriga elever på skolan. I det här arbetet vill vi undersöka hur flera utbrott över tid på en skola påverkar varandras storlek. Att det skulle kunna finnas en påverken utbrotten emellan är rimligt att anta eftersom att det efter ett utbrott blir immuna elever på skolan som under sin resterande tid på skolan inte kommer att vara potentiella smittspridare i något kommande utbrott. Då det första utbrottet introduceras antar vi att samtliga elever på skolan är mottagliga för smittan. För de kommande utbrotten kommer en andel av eleverna att vara immuna mot smittan sedan tidigare utbrott. Varje utbrott introduceras genom att vi låter varje mottaglig elev på skolan bli smittad av sjukdomen utanför skolan med en sannolikhet p_u. Sannolikheten att någon elev på skolan blir smittad utifrån kommer då att vara beroende av antalet elever på skolan som är mottagliga för smittan före utbrottets start. Detta kan göra så att vissa år kanske t.o.m. kommer att hoppas över då så många elever har blivit immuna sedan tidigare utbrott att det inte finns tillräckligt många mottagliga elever på skolan för att en epidemi ska kunna bryta ut.

Eftersom samtliga elever på skolan kommer att vara mottagliga inför det första utbrottet kommer samtliga I- och R-värden att vara lika med 0 vid tiden 0. Med tiden 0 menar vi här början av ett nytt år just innan ett nytt utbrott ska bryta ut. S-värdena vid tiden 0 kommer att vara lika med det totala antalet elever på skolan, i årskurserna och i klasserna. Inför de följande utbrotten kommer samtliga I-värden att vara 0 eftersom inga elever kommer att vara smittade av sjukdomen mellan utbrotten. Däremot kommer både S- och R-värdena att övertas från de sista tidsstegen i de närmaste föregående utbrotten. Med ett tidssteg menas den tid det tar för en smittad individ att smitta andra individer och själv tillfriskna, vilka inte kommer att vara längre än att varje utbrott hinner dö ut innan ett nytt år. I båda fallen kommer vid tiden 1 S-, I- och R-värdena för klasserna att uppdateras enligt följande

$$\begin{aligned} I_1^{i,j} &\sim \mathrm{Bin}(S_0^{i,j}, p_u) \\ S_1^{i,j} &= S_0^{i,j} - I_1^{i,j} \\ R_1^{i,j} &= R_0^{i,j} + I_0^{i,j} \end{aligned}$$

S-, I- och R-värdena för årskurserna och hela skolan kommer att beräknas genom att S-, I- och R-värdena för klasserna summeras.

För följande tidssteg (2 till utbrottens slut) beräknas S-, I- och R-värdena enligt beskrivningen i avsnitt 3.1 – 3.2.

4 2 olika modeller för årsbyte

4.1 Den realistiska modellen

Om vi vill åstadkomma simuleringar som ska komma så nära verkliga smittspridningsförlopp i skolan som möjligt så är det rimligast att låta alla elever som går på skolan gå där så många år som det finns årskurser på skolan eftersom det är så länge som de allra flesta elever går i skolan i verkligheten. I den realistiska modellen låter vi samtliga elever som börjar på skolan börja i den lägsta årskursen. Efter varje år som går förflyttas de en årskurs uppåt. En elev har alltså alltid både samma klasskamrater över alla skolår och samma elever som går i samma årskurs som de själva. Då eleverna har gått ett år i den högsta årskursen slutar de på skolan.

4.2 Den slumpmässiga modellen

För både den realistiska och den slumpmässiga modellen genomförs simuleringar av utbrotten enligt beskrivningen i avsnitt 3.1 – 3.3. Men i den slumpmässiga modellen går inte alla elever i skolan i precis så många år som det finns årskurser i skolan och slutar efter den högsta årskursen. Vid slutet av varje år väljs istället slumpmässigt ett antal elever motsvarande en hel årskurs ut som ersätts med nya elever som inte tidigare har utsatts för smittan som har spridits på skolan och därmed är mottagliga för den.

4.3 Jämförelse mellan modellerna

Inom den matematiska statistiken finns från matematikern George E. P. Box följande kända citat:

"*All models are wrong, but some are useful*" (3, s. 792)

Alla modeller är fel, men några är användbara. Generellt sätt när man sätter upp en modell så strävar man efter att göra den så enkel men ändå så användbar som möjligt. Om det finns delar av en modell som gör den mycket mer komplicerad men som bara påverkar resultaten i marginalerna så är det att föredra att avstå från dessa delar. Skillnaden mellan simuleringar när man utgår från den realistiska modellen i avsnitt 4.1 och den slumpmässiga modellen i avsnitt 4.2 skulle kunna förväntas vara relativt små eftersom det i de både modellerna slutar lika många elever efter varje år, och även om vissa elever enligt den slumpmässiga modellen kommer att gå i skolan i något färre

antal år än vad det finns årskurser i skolan och andra elever något fler antal år så kommer eleverna där ändå i genomsnitt att gå i skolan lika länge som i den realistiska modellen. Då de finns sammanhang då tillämpning av den slumpmässiga modellen gör beräkningarna lättare än om man istället tillämpar den realistiska modellen är dessa två modeller intressanta att jämföra. Då det i detta fall är möjligt att tillämpa de båda modellerna är det intressant att se hur stort fel vi får genom att tillämpa den slumpmässiga modellen istället för den realistiska modellen. Förlorar man mycket information om man i andra sammanhang för att åstadkomma en förenkling tillämpar den slumpmässiga modellen istället för den realistiska. En motsvarande situation där den slumpmässiga modellen kan tillämpas är t.ex. om man istället för att undersöka epdiemispridning i en skola undersöker epidemispridning i ett helt samhälle som istället för årskurs- och klassindelningar av individer kan ha t.ex. generationsindelningar. Precis som att alla elever befinner sig i en årskurs under lika lång tid så befinner sig alla individer i ett samhälle i en generation under lika lång tid.

En till aspekt som gör jämförelsen av den realistiska och den slumpmässiga modellen intressant är att den slumpmässiga modellen påminner mer om hur individer börjar och slutar på en arbetsplats. På en arbetsplats är det inte typiskt att alla individer jobbar lika länge och precis som på en skola så kan de finnas grupper av anställda som ser varandra mer frekvent och därför sprider en smitta mellan varandra med en högre sannolikhet. Det gör att jämförelsen mellan de två modellerna också skulle kunna ses som en jämförelse mellan hur epidemispridning beter på en skola och hur den beter sig på en arbetsplats.

5 Reproduktionstalet $\mathbf{R_0}$

Då tiden t för ett utbrott går mot oändligheten finns det två möjliga scenarion.
I: Endast ett fåtal individer blev smittade innan utbrottet dog ut.
II: En stor andel av de mottagliga individerna blev smittade innan utbrottet dog ut.
Mycket jobb har gjorts för att ta reda på vad som påverkar sannolikheten för att antingen bara några få individer drabbas av ett utbrott eller att ett utbrott blir stort. Man har kommit fram till att parmetern R_0, reproduktionstalet, spelar en central roll. R_0 är en beteckning för medelvärdet för antalet individer som en smittad individ smittar i ett tidigt skede av ett utbrott i en population där de flesta av de övriga individerna är mottagliga för smittan. Att ett utbrott blir stort i en population, scenario II, är möjligt om och endast om $R_0 > 1$. För en närmare beskrivning av reproduktionstalet R_0 se (2, s. 6).

6 Simuleringar och resultat

6.1 Smittspridningsparametrarna

Samtliga simuleringar har gjorts med smittspridningsparametrarna 0.003 utifrån skolan, 0.05 inom klassen, 0.005 inom årskursen och 0.0005 över årskurserna. Alla simuleringar har även gjorts med totalt 1000 elever i skolan och 10 årskurser med 100 elever i varje, men de två klassstorlekarna 10 elever/klass och 20 elever/klass har jämförst. För dessa parametrar och elevantal får vi följande R_0-värden:

$$R_0^{10} = 0.05 \cdot 9 + 0.005 \cdot 90 + 0.0005 \cdot 900 = 1.35$$

och

$$R_0^{20} = 0.05 \cdot 19 + 0.005 \cdot 80 + 0.0005 \cdot 900 = 1.8$$

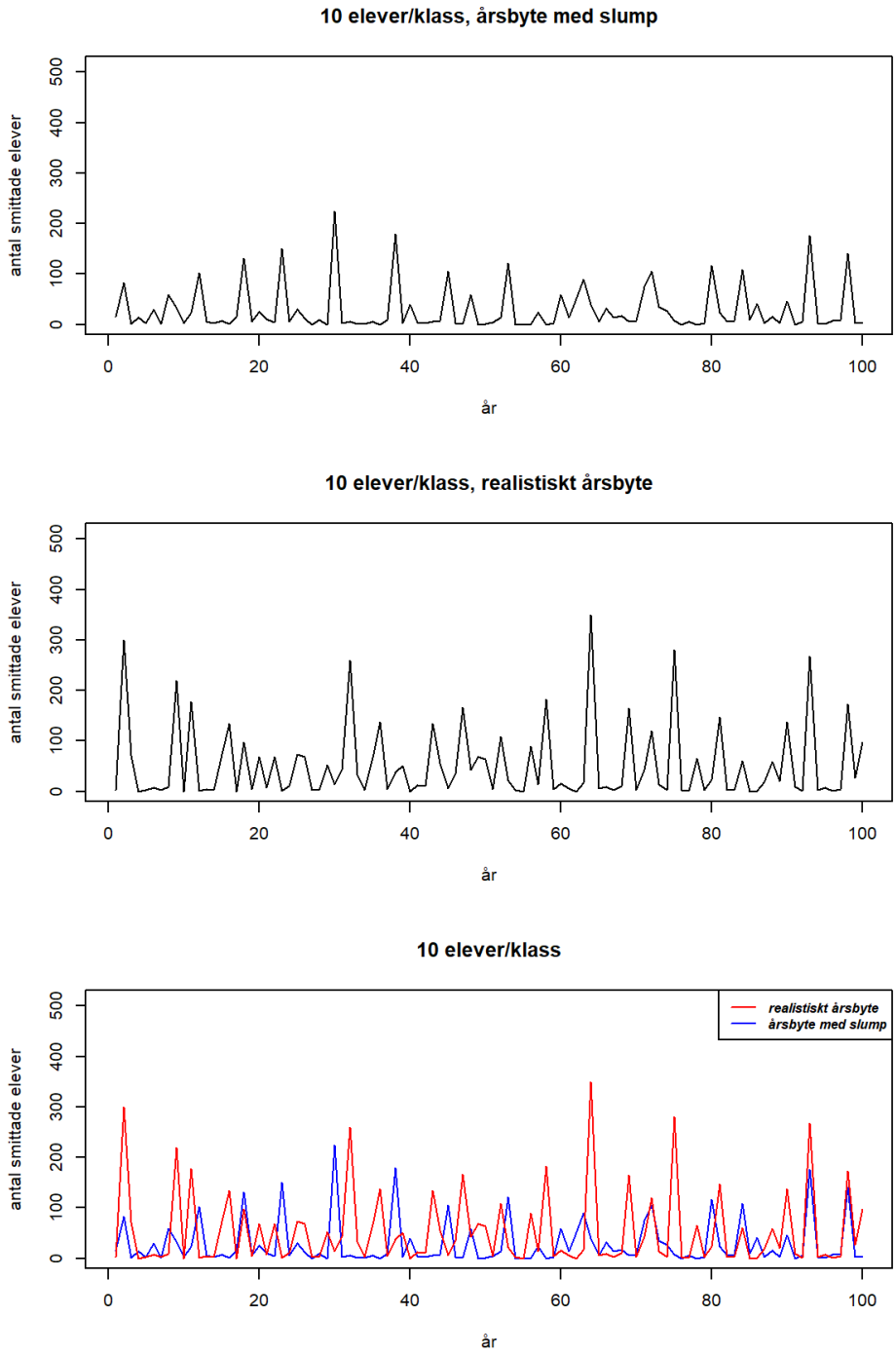
Eftersom om endast en elev på skolan smittas utanför skolan och samtliga av de övriga eleverna på skolan är mottagliga för smittan så har den smittade eleven, beroende av om klasstorlekarna är 10 elever/klass eller 20 elever/klass, 9 eller 19 klasskamrater som denna smittar med sannolikheten 0.05, 90 eller 80 elever som går i samma årskurs men i en annan klass som denna smittar med sannolikheten 0.005 och 900 elever i de andra årskurserna som denna smittar med sannolikheten 0.0005. Eftersom båda R_0-värdena är större än 1 är de möjligt att stora utbrott utvecklas både då klasstorlekarna är 10 elever/klass och då de är 20 elever/klass.

6.2 10 elever/klass

I Figur 2 visas resultaten från simuleringar av 100 utbrott dels då årsbytena har gjorts enilgt den slumpmässiga modellen, och dels då årsbytena har gjorts enligt den realistiska modellen. I båda fallen har klasstorlekarna 10 elever/klass använts och övriga parameterval har gjorts enligt avsnitt 6.1. Samma resultat visas även i en gemensam graf. Vi ser i båda fallen att större utbrott följs av lugnare perioder då inte några eller väldigt få elever blir smittade. I den gemensamma grafen ser vi även att det generellt sett blir större utbrott då den realistiska modellen för årsbyten används än när den slumpmässiga modellen används. Under de 100 simulerade utbrotten där den slumpmässiga modellen för årbyten användes blev totalt 2920 elever smittade. Då den realistiska modellen för årsbyte användes blev totalt 5294 elever smittade.

6.3 20 elever/klass

I Figur 3 visas resultaten från simuleringar av 100 utbrott både för den slumpmässiga modellen för årsbyte och för den realistiska modellen. I båda



Figur 2: 100 simulerade utbrott dels då årsbyte har gjorts enligt den slumpmässiga modellen och även där årsbyte har gjorts enligt den realistiska modellen, både var för sig och i en gemensam graf. Följande smittspridningsparametrar har använts: $p_u = 0.003$, $p_k = 0.05$, $p_a = 0.005$ och $p_r = 0.0005$.

fallen har klasstorlekarna 20 elever/klass använts. Valen av smittspridningsparametrarna har även här gjorts enligt avsnitt 6.1. Vi ser liknande mönster för hur utbrottsstorlekarna förändras över tid som i Figur 2, men med högre toppar, alltså större utbrott. Då den slumpmässiga modellen för årsbyte användes blev totalt 4427 elever smittade. När den realistiska modellen för årsbyte användes blev 7284 elever smittade.

6.4 Smittspridningsparametrarnas påverkan

Nu har vi jämfört de två olika modellerna för en specifik kombination av smittspridningsparametrar. Alla sjukdomar sprids dock inte med en sannolikhet som motsvarar de val av smittspridningsparametrar som vi har gjort. Det är därför intressant att undersöka hur stor del av den observerade skillnaden mellan den realistiska och den slumpmässiga modellen för årsbyte som beror på en faktisk skillnad mellan modellerna, och hur stor del av skillnaden som har uppstått för just de val av smittspridningsparametrar som vi har gjort.

Smittspridningsparametrarnas påverkan på resultatet jämförs med hjälp av simuleringar där följande kombinationer av smittspridningsparametrar används:

$$
\begin{aligned}
p_k &= 0.01 + i \cdot 0.001 \\
p_a &= 0.001 + i \cdot 0.0001 \\
p_r &= 0.0001 + i \cdot 0.00001
\end{aligned}
$$

$$\text{där } i = 0, 1, 2, ..., 160$$

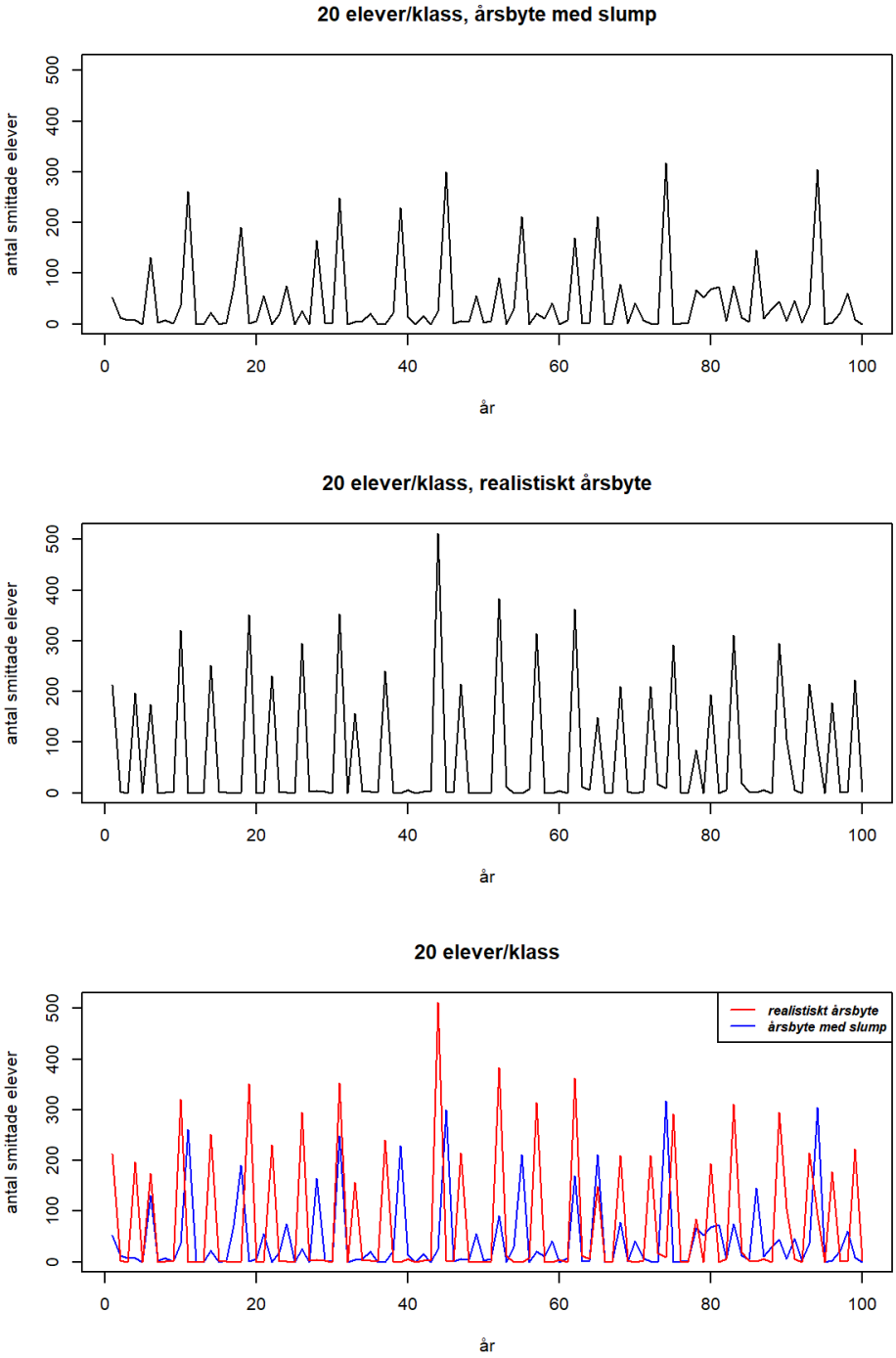
och

p_k = sannolikheten att en elev smittar en klasskamrat

p_a = sannolikheten att en elev smittar en elev som går i samma årskurs men en annan klass

p_r = sannolikheten att en elev smittar en elev som går i en annan årskurs

För samtliga simuleringar används smittspridningsparametern, för sannolikheten att en elev smittas utanför skolan, $p_u = 0.003$.
För varje i-värde simuleras 1000 utbrott med slumpmässigt årsbyte emellan och 1000 utbrott med realistiskt årsbyte emellan. För båda grupperna beräknas medelvärdet för hur många elever som smittas per utbrott. För att tydliggöra skillnaden mellan de två olika modellerna för årsbyte beräknas även differensen mellan och kvoten för de olika medelvärdena. Samma



Figur 3: 100 simulerade utbrott dels då årsbyte har gjorts enligt den slumpmässiga modellen och även där årsbyte har gjorts enligt den realistiska modellen, både var för sig och i en gemensam graf. Följande smittspridningsparametrar har använts: $p_u = 0.003$, $p_k = 0.05$, $p_a = 0.005$ och $p_r = 0.0005$.

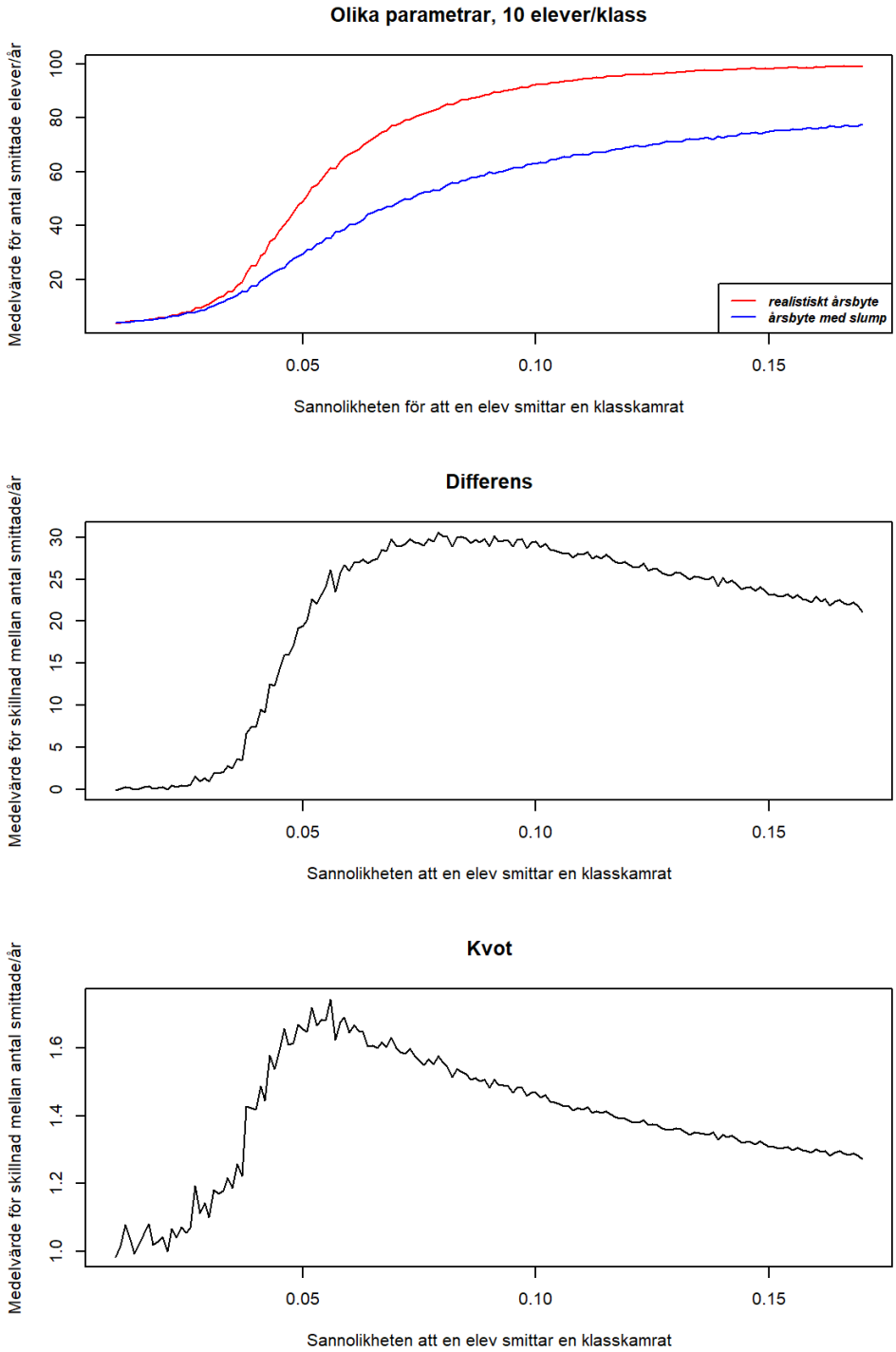
procedur genomförs både med klasstorlekarna 10 elever/klass och klasstorlekarna 20 elever/klass. Resultaten då klasstorlekarna 10 elever/klass användes presenteras i Figur 4 och resultaten då klasstorlekarna 20 elever/klass användes presenteras i Figur 5. I Figur 4 ser vi att skillnaden mellan den slumpmässiga modellen och den realistiska modellen med avseende på kvoten för medelvärdena då klasstorlekarna är 10 elever/klass är som störst ungefär då sannolikheten för att en elev smittar en klasskamrat är 0.05, vilket ju är precis den parameter som användes i de första simuleringarna. I figur 5 där klasstorlekarna är 20 elever/klass är motsvarande kvot som störst för något mindre värden på smittspridningsparametrarna. Vi kan se att medelvärdena i både Figur 4 och 5 förändras på liknande sätt men att förändringen sker fortare då klasstorlekarna är större.

6.4.1 $\mathbf{R_0}$

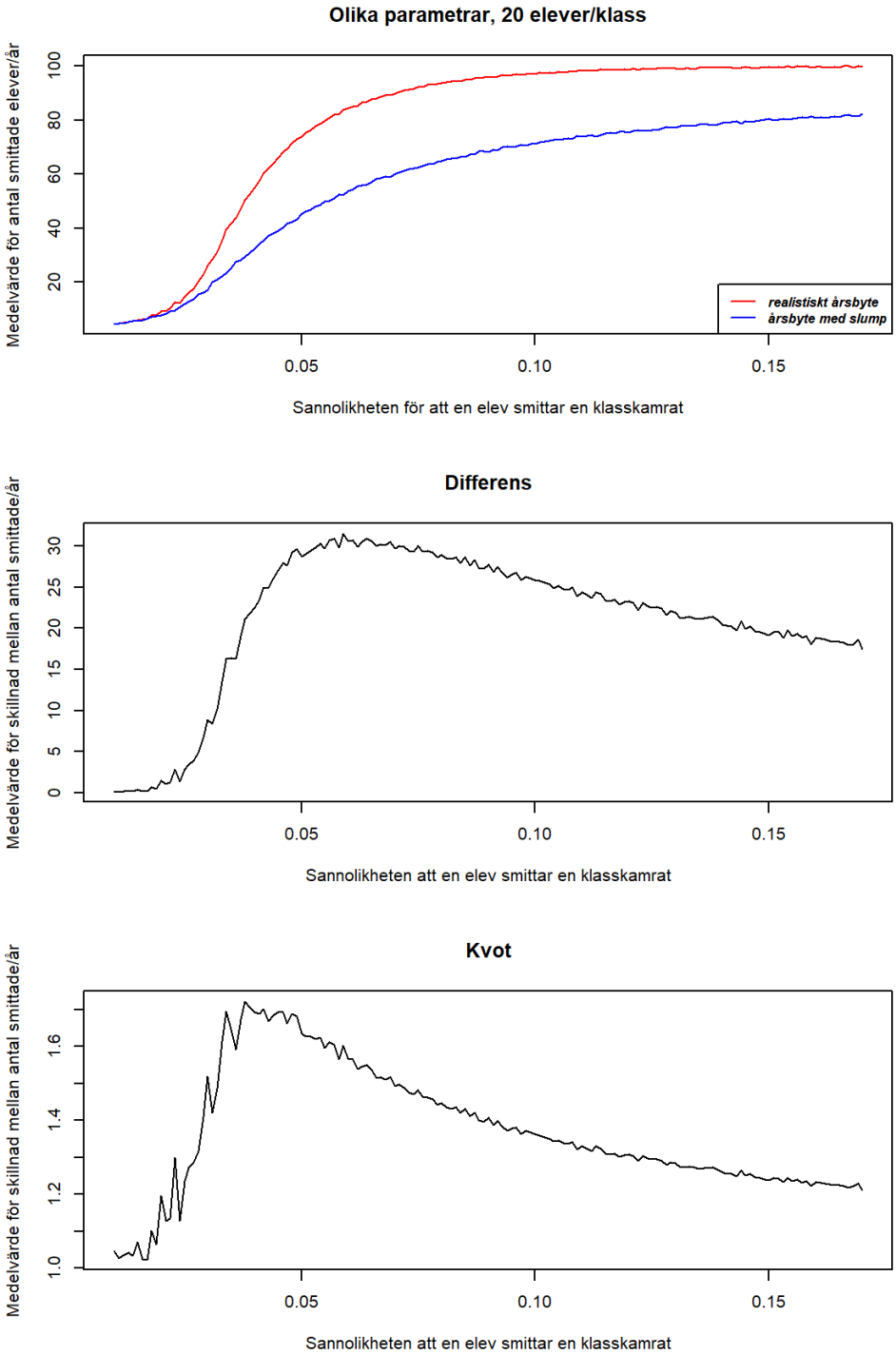
I Figur 6 visas hur värdet på reproduktionstalet R_0 förändras för de två klasstorlekarna 10 elever/klass och 20 elever/klass då kombinationen av smittspridningaparametrar ändras. Värdena på R_0 är beräknade på motsvarande sätt som i början i avsnitt 6.4. Ungefär då $R_0 = 4$ är en kritisk gräns för när sannolikheten är stor att de flesta individerna, $\approx 98\%$, i en population blir smittade under ett utbrott. Värdet 98% har beräknats med hjälp av formel (4.2) på sidan 32 i (2). Denna gräns är därför markerad i figuren. Då klasstorlekarna är 20 elever/klass uppnås R_0-värdet 4 då sannolikheten att en elev smittar en klasskamrat är något större än 0.1. Då klasstorlekarna är 10 elever/klass uppnås den kritiska gränsen $R_0 = 4$ då sannolikheten att en elev smittar en klasskamrat är ungefär 0.15. I Figur 5, "Olika parametrar, 20 elever/klass", syns att medelvärdet för antalet smittade elever/år då modellen för realistiskt årsbyte har använts är nära 100 då sannolikheten att en elev smittar en klasskamrat är något större än 0.1, vilket motsvarar precis att samtliga individer i populationen smittas av sjukdomen eftersom det är 100 nya elever som börjar på skolan varje nytt år. I Figur 4, "Olika parametrar, 10 elever/klass", är medelvärdet för antalet smittade elever/år då modellen för realistist årsbyte har använts nära 100 då sannolikheten att en elev smittar en klasskamrat är ungefär 0.15.

6.5 Antalet smittade elever i en årskurs över hela skoltiden

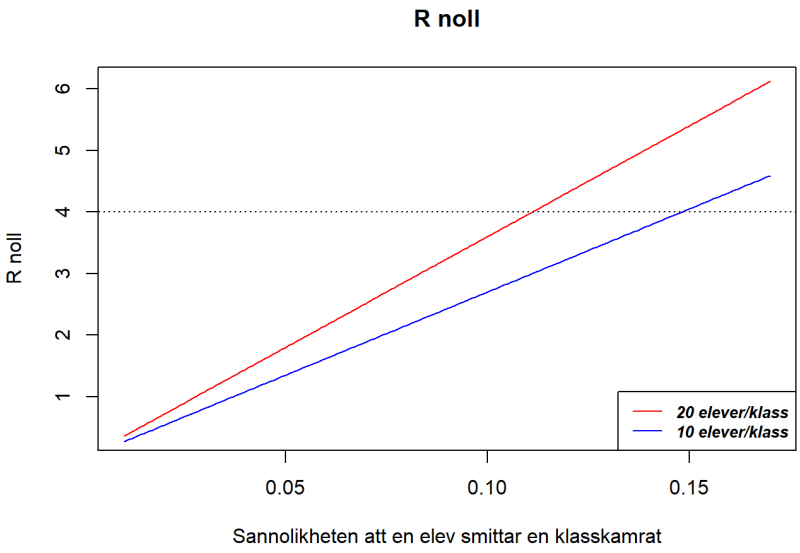
För den realistiska modellen för årsbyte är det intressant att undersöka hur antalet immuna elever i en årskurs förändras under elevernas skoltid. Motsvarande undersökning är svår att göra för den slumpmässiga modellen för årsbyte eftersom inte alla elever går i skolan under lika lång tid enligt den modellen. I Figur 7 visas ett exempel över hur antalet immuna elever förändras, under en hel skoltid på 10 år, i dels en årskurs på en skola där klasstorlekarna är 10 elever/klass och dels en årskurs på en skola där klasstorlekarna är



Figur 4: Visar medelvärden för antalet smittade elever under ett utbrott då årsbyte görs på ett realistiskt sätt och då det görs med slump för olika värden på smittspridningsparametrarna. Visar även differensen och kvoten mellan dessa två.



Figur 5: Visar medelvärden för antalet smittade elever under ett utbrott då årsbyte görs på ett realistiskt sätt och då det görs med slump för olika värden på smittspridningsparametrarna. Visar även differensen och kvoten mellan dessa två.

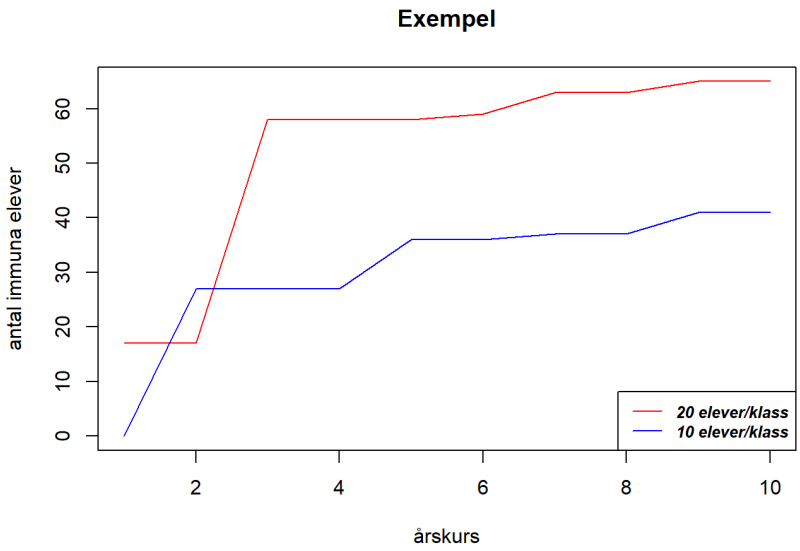


Figur 6: Visar värden på R_0 för de två klasstorlekarna 10 elever/klass och 20 elever/klass då olika kombinationer av smittspridningsparametrar används. Den kritiska gränsen $R_0 = 4$, då sannolikheten att de flesta individerna i populationen blir smittade är hög, är markerad.

20 elever/klass. I exemplet har först 10 utbrott simulerats på båda skolorna motsvarande en kalibreringsperiod för att inte utgångspunkten skulle vara en skola där samtliga elever var mottagliga innan det första utbrottet. Därefter har tio nya utbrott simulerats efter vilka värden för antalet immuna i de årskurser som har följts åt har sparats. Då det är svårt att veta om ett exempel är representativt för hur antalet immuna elever i en årskurs förändras under skoltiden generellt sett är det intressant att även undersöka väntevärdena för hur många elever som är immuna i varje årskurs efter varje år. Approximativa väntevärden kan beräknas med hjälp av Monte Carlo-simuleringar.

6.5.1 Monte Carlo-simuleringar

"Simuleringsstudier, eller Monte Carlo-metoder, används för att skaffa approximativ information om storheter (oftast sannolikheter eller väntevärden) som man inte klarar att beräkna analytiskt."(4, s. 235)



Figur 7: Visar ett exempel över hur antalet immuna elever kan förändras i en årskurs med elever då man följer dem över alla tio år som de går i skolan, dels då klasstorlekarna är 10 elever/klass och dels då klasstorlekarna är 20 elever/klass.

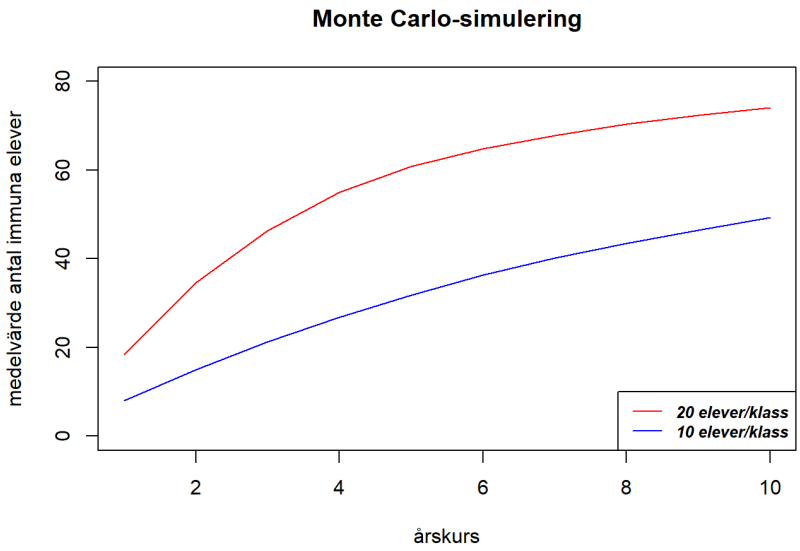
Idén om att approximativt bestämma väntevärden med hjälp av Monte Carlo-metoder bygger på Stora talens lag:

"Låt $X_1, X_2, ...$ vara en följd av oberoende slumpvariabler, alla med samma väntevärde $E(X_i) = \mu$ och standardavvikelse $D(X_i) = \sigma < \infty$.
Låt $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ vara medelvärdet av de n första variablerna.
Då gäller, för godtyckligt $\varepsilon > 0$, att

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \to 0, \qquad \text{då } n \to \infty."(4, \text{s. } 160)$$

Om man skulle kunna simulera oändligt många oberoende likafördelade slumpvariabler med ändlig standardavvikelse skulle man kunna bestämma väntevärdet för dessa genom att beräkna medelvärdet av utfallen. Detta är dock inte möjligt. Men genom att göra många simuleringar kan ändå ett approximativt värde för väntevärdet som är tillräckligt nära de riktiga värdet för att vara intressant bestämmas.

För vardera klasstorlek 10 elever/klass och 20 elever/klass har 1000 utbrott med realistiskt årsbyte emellan simulerats. Efter varje utbrott har värden för antalet immuna i varje årskurs sparats. Efter alla utbrott har medelvärdena, som är våra approximativa väntevärden, för dessa beräknats. Resultatet presenteras i Figur 8. I figuren syns en tydlig skillnad mellan hur



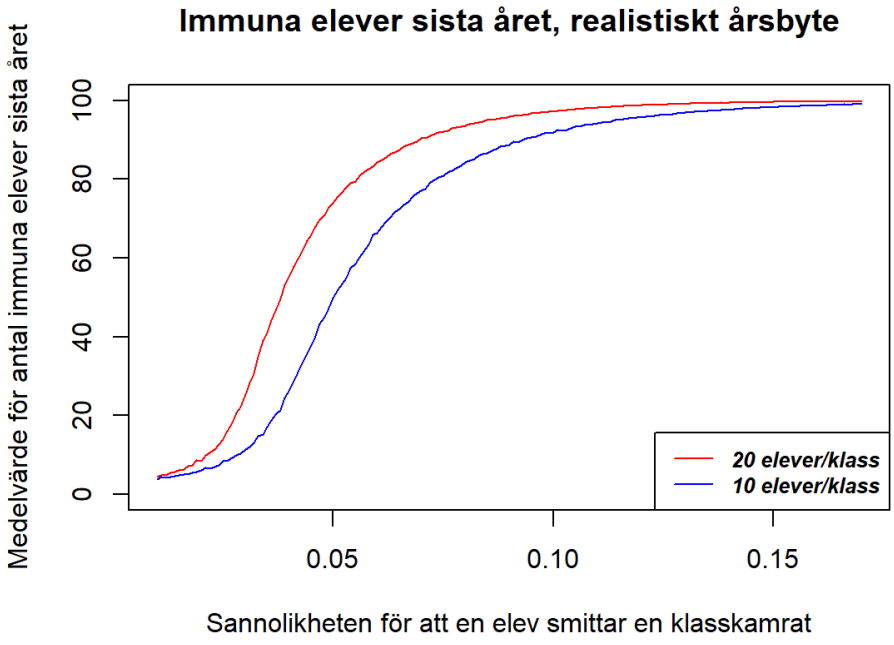
Figur 8: Visar approximativa väntevärden, som har beräknats med hjälp av Monte Carlo-simuleringar, för hur många elever som kommer att vara immuna i en årskurs efter varje år under hela skoltiden

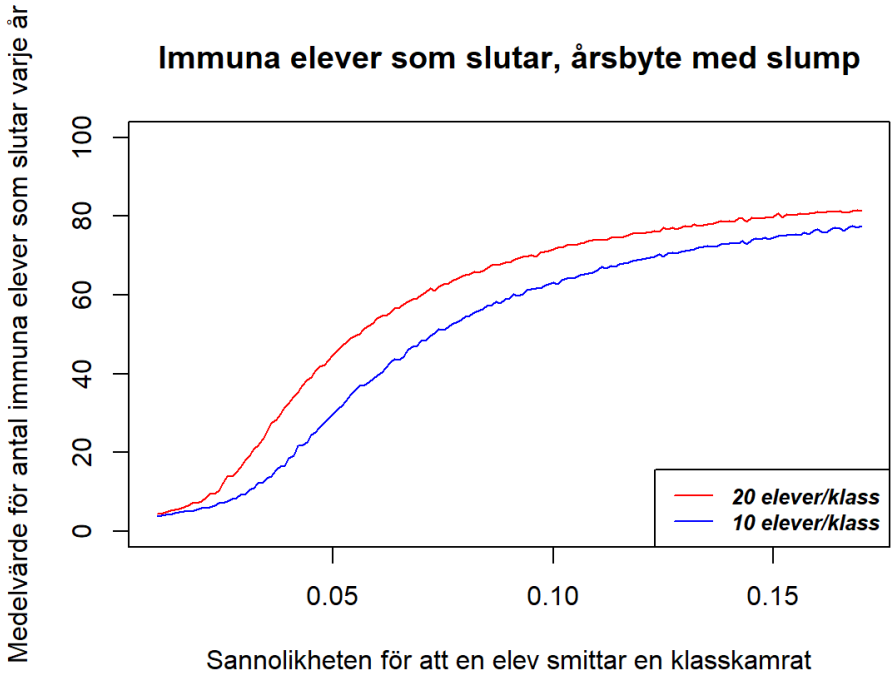
många elever som blir smittade då klasstorlekarna är 10 elever/klass och då de är 20 elever/klass. Observera att värdena är för de diskreta tidsstegen, figuren säger inget om hur många elever som har blivit immuna efter ett halvt år i en årskurs. Eftersom kurvorna har ett jämt utseende, det förekommer inte några plötsliga hopp mellan några årkurser, är det stor sannolikhet att de approximativa väntevärdena är nära de verkliga väntevärdena. Att värdena är strikt växande för högre årskurser är också en indikation på att de är rimliga. Alla elever som blir immuna i en årskurs kommer också att vara immuna under de kommande åren i skolan.

6.6 Antalet immuna elever som slutar

För att ta reda på en förklaring till varför det blir en så tydlig skillnad mellan hur många elever som i genomsnitt blir smittade under ett utbrott då den realistiska modellen för årsbyte används och då den slumpmässiga modellen för årsbyte används, undersöks genomsnitten för hur många immuna elever som slutar på skolorna varje år.

Vi ser att dessa figurer är väldigt lika Figur 4 och Figur 5 som visar hur många elever som i genomsnitt blir smittade varje år. Alla elever som blir smittade kommer på lång sikt någon gång att lämna skolan. På lång sikt blir också därför dessa värden lika med varandra.





Figur 9: Visar approximativa värden för genomsnitten för hur många elever som är immuna av de som slutar på en skola varje år. Både den realistiska modellen för årsbyte, den slumpmässiga modellen för årsbyte och olika kombinationer av smittspridningsparametrar jämförs.

7 Diskussion

7.1 Slutsatser

Vi har sett en tydlig skillnad mellan de två modellerna för årsbyte. Utbrotten blir större då den realistiska modellen för årsbyte tillämpas än när den slumpmässiga modellen tillämpas. Detta skule kunna bero på att så många elever hinner bli smittade när de går 10 år i skolan att att vissa elever går ännu längre i skolan enligt den slumpmässiga modellen inte kan kompensera för de elever som går i skolan en kortare tid och därmed inte löper lika stor risk att bli smittade. Antalet elever som skulle bli smittade under ett verkligt utbrott ligger troligtvis någonstans mellan resultaten för de två modellerna eftersom det både finns elever som byter skola och elever som går om eller hoppar över någon årskurs.

7.2 Förslag till ytterligare frågeställningar

Hur är väntevärdet för hur många elever som blir smittade under ett utbrott beroende av hur många elever som är mottagliga för smittan innan utbrottet? När är detta väntevärde lika med väntevärdet för antalet immuna elever som slutar på skolan samma år? Blir skillnaden mellan de två modellerna för årsbyte den samma för en skola med färre årskurser?

Referenser

[1] Maria Deijfen: "Epidemics and vaccination on weighted graphs", *Mathematical biosciences* **232**.1 (2011): 57-65. https://www.sciencedirect.com/science/article/pii/S0025556411000678

[2] Håkan Andersson och Tom Britton (2000): "Stochastic Epidemic Models and Their Statistical Analysis", *Springer*

[3] George E. P. Box (1976): "Science and Statistics", *Journal of the American Statistical Association* **71** http://www-sop.inria.fr/members/Ian.Jermyn/philosophy/writings/Boxonmaths.pdf

[4] Sven Erick Alm och Tom Britton (2008): "Stokastik: Sannolikhetsteori och statistikteori med tillämpningar", *Liber*