

# Binomni epidemiološki modeli

---

**Blažević, Ivana**

**Master's thesis / Diplomski rad**

**2022**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, Department of Mathematics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:126:850934>

*Rights / Prava:* [In copyright](#) / [Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-12-26**



**mathos**

*Repository / Repozitorij:*

[Repository of School of Applied Mathematics and Informatics](#)



Sveučilište J. J. Strossmayera u Osijeku  
Odjel za matematiku  
Diplomski studij matematike, smjer: Financijska matematika i statistika

Ivana Blažević

## **Binomni epidemiološki lanci**

Diplomski rad

Osijek, 2022.

Sveučilište J. J. Strossmayera u Osijeku  
Odjel za matematiku  
Diplomski studij matematike, smjer: Financijska matematika i statistika

Ivana Blažević

## **Binomni epidemiološki lanci**

Diplomski rad

Mentor: izv. prof. dr. sc. Nenad Šuvak

Osijek, 2022.

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Osnovni pojmovi</b>	<b>3</b>
2.1	Pretpostavke modela . . . . .	6
<b>3</b>	<b>Tipovi binomnih epidemioloških modela</b>	<b>9</b>
3.1	Reed-Frostov model . . . . .	9
3.2	Greenwoodov model . . . . .	10
3.3	Modificirani Reed-Frostov model . . . . .	13
3.4	Generalizirani model binomnih epidemioloških lanaca . . . . .	13
<b>4</b>	<b>Prikaz modela preko Markovljevih lanaca</b>	<b>15</b>
4.1	Reed-Frostov model . . . . .	16
4.2	Greenwoodov model . . . . .	19
4.3	Modificirani Reed-Frostov model . . . . .	21
<b>5</b>	<b>Primjena na stvarnim podacima</b>	<b>23</b>
5.1	Simulacije modela . . . . .	27
<b>6</b>	<b>Literatura</b>	<b>30</b>
<b>7</b>	<b>Životopis</b>	<b>33</b>

# 1 Uvod

Kroz povijest pojavljivale su se razne zaraze koje su značajno utjecale na smrtnost ljudske vrste. Njihovu pojavu, razvoj, uzrok te posljedice koje ostave proučava znanost nazvana epidemiologija. Ako dođe do iznenadnog i naglog širenja neke zaraze na velik broj ljudi u kratkom vremenu onda se ta zaraza naziva epidemijom dok se širenje na više država i kontinenata naziva pandemijom. Na širenje zaraze, točnije epidemiju, utječu brojni čimbenici - način prenošenja, podložnost, uzrok, otpornost osobe, gustoća naseljenosti i drugo. Ono što mnoge epidemiologe zabrinjava je procjena ukupnog broja zaraženih, trajanje epidemije i slično. Zbog toga se, kao nadopuna epidemiološkim modelima, koristi matematičko modeliranje koje zahtijeva jednostavnost i preciznost u opisu epidemije.

Općenito, matematičke modele grubo možemo podijeliti na determinističke i stohastičke. Rezultat modeliranja s obzirom na ulazne podatke je u determinističkom modelu svaki put jednak dok je u stohastičkom različit. Također, u stohastičkom modelu uključena je neizvjesnost odnosno slučajnost neke pojave u epidemiji. Prvi modeli koji su dominirali epidemiologijom bili su upravo deterministički. Međutim, neki nisu davali jednostavna rješenja za svoje diferencijalne jednadžbe što je potaknulo na modeliranje sa stohastičkim svojstvima. Prvi takvi modeli, kao uvod u teoriju stohastičkih modela, bili su modeli **binomnih epidemioloških lanaca**.

Za epidemiju u kojoj je inkubacijsko ili latentno razdoblje male varijabilnosti te zarazno razdoblje kratko, treba koristiti modele diskretnog vremena. Binomni epidemiološki lanci upravo su takvi modeli - modeli u diskretnom vremenu u kojima je latentno razdoblje konstanta. Uvelike se koriste za modeliranje širenja zaraze u zatvorenoj i homogenoj populaciji kao što su kućanstva, vrtići, škole i slično. Pokazali su se korisnima u opisivanju virusnih zaraza poput ospica, vodenih kozica, gripe i obične prehlade.

Dvije najpoznatije klase modela binomnih epidemioloških lanaca uveli su L. Reed i W. H. Frost 1928. godine te M. Greenwood 1931. godine. Razlikuju se u činjenici da vjerojatnost zaraze u Reed-Frostovom modelu ovisi o zaraženim osobama dok u Greenwoodovom modelu to nije slučaj.

J. Gani i D. Jerwood povezuju binomne epidemiološke lance sa stohastičkim procesima s posebnim naglaskom na Markovljeve lance. Markovljev lanac je slučajni proces sa diskretnim skupom stanja i specifičnim svojstvom koje glasi: "prošlost i budućnost međusobno su nezavisne u odnosu na poznatu sadašnjost". Njihov pristup dovodi do formiranja relativno jednostavne metode za izračun distribucije trajanja epidemije i ukupnog broja zaraženih.

S druge strane, N. Becker kombinira i modificira Reed-Frostov i Greenwoodov mo-

del u generalizirani model binomnih epidemioloških lanaca.

U ovome radu opisat ćemo nekoliko klasa binomnih epidemioloških lanaca, a neke od njih primjeniti i simulirati na stvarnim podacima.

## 2 Osnovni pojmovi

Događaje u stvarnosti koje želimo analizirati modeliramo slučajnim varijablama na nekom vjerojatnosnom prostoru. Diskretna i neprekidna slučajna varijabla dva su tipa slučajnih varijabli koje se koriste u praksi.

**Definicija 2.1.** Funkcija  $X : \Omega \rightarrow \mathbb{R}$  je **slučajna varijabla** na vjerojatnosnom prostoru  $(\Omega, \mathcal{F}, \mathbb{P})$  ako vrijedi

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}) \quad (1)$$

pri čemu je  $\mathcal{B}(\mathbb{R})$  **Borelova  $\sigma$ -algebra**.

**Definicija 2.2.** Neka je  $(\Omega, \mathcal{F}, \mathbb{P})$  vjerojatnosni prostor i neka je  $X$  slučajna varijabla. Funkcija  $\mathcal{F} : \mathbb{R} \rightarrow [0, 1]$  koja realnom broju  $x$  pridružuje vjerojatnost da dana slučajna varijabla bude manja ili jednaka tom broju, tj. funkciju

$$F(x) = P\{\omega \in \Omega : X(\omega) \leq x\} = P\{X \leq x\}, \quad (2)$$

zovemo **funkcija distribucije slučajne varijable  $X$** .

Distribucija koja će nama biti od velike važnosti je **binomna distribucija** što možemo naslutiti pomoću samog naziva modela kojeg ćemo analizirati.

**Definicija 2.3.** Neka je  $n \in \mathbb{N}$  i  $p \in \langle 0, 1 \rangle$ . Za slučajnu varijablu koja prima vrijednosti iz skupa  $\{0, 1, 2, \dots, n\}$  s vjerojatnostima

$$p_i = P\{X = i\} = \binom{n}{i} p^i (1-p)^{(n-i)} \quad (3)$$

kažemo da ima **binomnu distribuciju** s parametrima  $n$  i  $p$ , oznaka  $X \sim \mathcal{B}(n, p)$ .

Parametar  $n$  je broj nezavisnih ponavljanja pokusa dok  $p$  označava vjerojatnost uspjeha u jednom izvođenju pokusa.

Kako neku pojavu pratimo kroz vrijeme, treba nam koncept slučajnog procesa.

**Definicija 2.4. Slučajni proces**  $(X_t, t \in T)$  je familija slučajnih varijabli na istom vjerojatnosnom prostoru  $(\Omega, \mathcal{F}, \mathbb{P})$ , pri čemu je  $t$  element parametarskog skupa ili skupa indeksa  $T \subseteq \mathbb{R}$ .

Navedeni **skup indeksa  $T$**  najčešće je modeliran kao vrijeme procesa koje može biti diskretno ili neprekidno. U našoj vrsti modela radi se o diskretnom vremenu, a nešto više o njemu reći ćemo nakon nekoliko bitnih definicija.

**Definicija 2.5.** Neka je  $S$  prebrojiv skup. Slučajni proces  $X = (X_n, n \geq 0)$  definiran na vjerojatnosnom prostoru  $(\Omega, \mathcal{F}, \mathbb{P})$  s vrijednostima u skupu  $S$  je **Markovljev lanac** ako vrijedi

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i)$$

za svaki  $n \geq 0$  i za sve  $i_0, \dots, i_{n-1}, i, j \in S$  za koje su obje uvjetne vjerojatnosti dobro definirane.

Skup stanja  $S$  u prethodnoj definiciji zapravo je skup svih mogućih vrijednosti slučajnih varijabli  $(X_n, n \in \mathbb{N})$ . Kako je on u našoj vrsti modela diskretan, radi se o lancu.

Uz prethodnu definiciju, imamo još jednu koja daje bolji opis Markovljevih lanaca. Prije nego nju i iskažemo, uvodimo sljedeću definiciju.

**Definicija 2.6.** Matrica  $P = (p_{ij}, i, j \in S)$  naziva se **stohastičkom matricom** ako je  $p_{ij} \geq 0$  za sve  $i, j \in S$ , te

$$\sum_{j \in S} p_{ij} = 1, \forall i \in S. \quad (4)$$

**Definicija 2.7.** Neka je  $\lambda = (\lambda_i, i \in S)$  vjerojatnosna distribucija na  $S$ , te neka je  $P = (p_{ij}, i, j \in S)$  stohastička matrica. Slučajni proces  $X = (X_n, n \geq 0)$  definiran na vjerojatnosnom prostoru  $(\Omega, \mathcal{F}, \mathbb{P})$  sa skupom stanja  $S$  je **homogen Markovljev lanac s početnom distribucijom  $\lambda$  i prijelaznom matricom  $P$**  ako vrijedi:

- i)  $\mathbb{P}(X_0 = i) = \lambda_i, \forall i \in S$ ,
- ii)  $\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{ij}, \forall n \geq 0, \forall i_0, \dots, i_{n-1}, i, j \in S$ .

Takav Markovljev lanac možemo još zvati  **$(\lambda, P)$ -Markovljev lancem**.

Markovljev lanac u diskretnom vremenu svrstavamo u klasu procesa za koje vjerojatnost prijelaza iz jednog stanja u drugo zadajemo **funkcijom prijelaznih vjerojatnosti** definiranom pravilom  $p(i, s; j, t) = \mathbb{P}(X_t = j | X_s = i)$ , za stanja  $i, j \in S$  i trenutke  $s, t \in \mathbb{N}_0$  takve da je  $s < t$ .

Zbog definicije 2.7. za Markovljev lanac bitna je distribucija slučajne varijable  $X_0$  i **funkcija 1-koračnih prijelaznih vjerojatnosti**, za stanja  $i, j \in S$  i  $n \in \mathbb{N}_0$ , definirana pravilom

$$p(i, n; n+1, j) = \mathbb{P}(X_{n+1} = j | X_n = i). \quad (5)$$

Ako (5) ne ovisi o vremenu, radi se o (vremenski) homogenom Markovljevom lancu. Tada, funkcija 1-koračnih prijelaznih vjerojatnosti ovisi samo o stanju  $i \in S$  iz kojeg lanac polazi i stanju  $j \in S$  u kojeg lanac dolazi pa uvodimo oznaku

$$p(i, n; n+1, j) = p_{ij}. \quad (6)$$



Za takav Markovljev lanac 1-koračne prijelazne vjerojatnosti  $p_{ij}, i, j \in S$  zapisujemo formi matrice koju nazivamo **matrica 1-koračnih prijelaznih vjerojatnosti** te ju označavamo s  $P$ . Matrica  $P$  je **stohastička matrica** sa svojstvima iz definicije 2.6. Za  $S = \mathbb{N}$ , matrica 1-koračnih prijelaznih vjerojatnosti ima oblik:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

**Definicija 2.8.** Neka je  $X = (X_n, n \in \mathbb{N}_0)$  Markovljev lanac s prostorom stanja  $S$  i prijelaznom matricom  $P$ . Za  $B \subset S$  definiramo **prvo vrijeme pogađanja** tog skupa kao

$$T_B = \min\{n \geq 0 : X_n \in B\}, \quad (7)$$

uz konvenciju da je  $\min \emptyset = +\infty$ . U slučaju da je  $B = \{j\}$ , za  $j \in S$ , jednostavnije pišemo  $T_j$ .

**Definicija 2.9.** Slučajna varijabla  $T : \Omega \rightarrow \{\mathbb{N}_0\} \cup \{\infty\}$  zove se **vrijeme zaustavljanja** ako je za sve  $n \in \mathbb{N}_0$

$$\{T \leq n\} = \sigma(X_0, \dots, X_n), \quad (8)$$

tj. događaj  $\{T \leq n\}$  ovisi samo o  $X_0, X_1, \dots, X_n$ .

Prethodne dvije definicije ukazuju na činjenicu da je  $T_B$  ujedno i vrijeme zaustavljanja. Prije nego to pokažemo, moramo se prisjetiti pojma *filtracija*.

**Definicija 2.10.** Familija  $\sigma$ -podalgebri  $\mathbb{F} = \{\mathcal{F}_n, n \in \mathbb{N}_0\}$   $\sigma$ -algebre  $\mathcal{F}$  sa svojstvom

$$\mathcal{F}_n \subset \mathcal{F}_{n+1}, \quad \forall n \in \mathbb{N}_0$$

naziva se *filtracija na vjerojatnosnom prostoru*  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definicija 2.11.** Za slučajni proces  $(X_n, n \in \mathbb{N}_0)$  na vjerojatnosnom prostoru  $(\Omega, \mathcal{F}, \mathbb{P})$  s filtracijom  $\mathbb{F} = \{\mathcal{F}_n, n \in \mathbb{N}_0\}$  kažemo da je *adaptiran* na filtraciju  $\mathcal{F}$  ako je

$$\sigma(X_n) \subset \mathcal{F}_n, \quad \forall n \in \mathbb{N}_0.$$

Sada, možemo uočiti da za svaki  $n \in \mathbb{N}_0$  vrijedi

$$\{T_B = n\} = \{X_0 \notin B, \dots, X_{n-1} \notin B, X_n \in B\} \in \mathcal{F}_n.$$

Također, znamo da je proces  $X$   $\mathbb{F}$ -adaptiran pa vrijedi  $\sigma(X_0, \dots, X_n) \subseteq \mathcal{F}_n$ . Dakle,  $T_B$  je vrijeme zaustavljanja.

Neka je binomna slučajna varijabla broj neovisnih "uspjeha" u nizu bacanja novčića. Analogija je ovdje da bacamo novčić (s vjerojatnošću "glava"  $p$ ) za svakog podložnog pojedinca u populaciji. Ako se pri bacanju kao ishod pojavi "glava", podložna osoba postaje zaražena. Inače, ostaje podložna i prelazimo na sljedeću podložnu osobu. Ovo nam objašnjenje služi kao motivacija za bolje razumijevanje binomnih epidemioloških modela.

Kako bismo opisali neki model potrebne su nam pretpostavke. One su, općenito, početak svakog modela. Daju opis svakog faktora koji utječe i kontrolira način širenja zaraze. Osim što pretpostavke moraju točno opisati biološke dijelove zaraze, one služe kao sredstvo za predviđanje pa tako moraju biti i matematički točne. Kao što smo već naveli u uvodu, model binomnih epidemioloških lanaca primjer je modela koja zadovoljava obje vrste kriterija. U sljedećem ćemo ih poglavljju i opisati.

## 2.1 Pretpostavke modela

Neka je  $N$  početni broj jedinki zatvorene i homogene populacije. Homogena populacija ima male varijacije dok je heterogena ona u kojoj pojedinci nisu slični jedni drugima. Npr. populacija ljudi koja je tisućama godina naselila otok s malom migracijom na ili s otoka vjerojatno je relativno homogena ili slična u svojim osobinama. Takva se populacija, u modelu, sastoji od dvije klase osoba:

- $S$  - podložne osobe (*eng.* susceptible); osobe koje nisu zaražene, ali su sposobne dobiti bolest i postati zarazne,
- $I$  - zaražene osobe (*eng.* infected); osobe koje su zaražene i zarazne, tj. mogu prenijeti bolest drugim jedinkama.

Izravno prenošenje zaraze sa zaražene na podložnu osobu naziva se "*adekvatan kontakt*". To znači da će se podložna osoba zaraziti unutar vremenskog intervala  $t$  (latentno razdoblje bolesti) i bit će zarazna samo tijekom sljedećeg vremenskog razdoblja. Pretpostavlja se da se svaka podložna osoba može zaraziti u svakom trenutku s istom vjerojatnošću  $p$  tj. *izbjecí* zarazu s vjerojatnošću  $q = 1 - p$  (vjerojatnost  $q$  bit će nam od veće značajnosti). Također, podložna osoba može se zaraziti neovisno o drugoj podložnoj. Sve osobe jednako su: otporne, sposobne se

zaraziti, prenijeti zarazu. Ove pretpostavke i uvjeti primjenjuju se tijekom cijele epidemije.

Broj osoba koje se zaraze u  $t = 0$  označavamo s  $I_0$ , ( $1 \leq I_0 \leq N$ ), a ostatak  $N - I_0 = S_0$  u istom trenutku predstavlja podložne osobe. Ako stavimo da je  $I_1$  broj novozaraženih u  $t = 1$  onda u modelu binomnog epidemiološkog lanca pretpostavljamo da  $I_1$ , uvjetno na poznato  $S_0 = N - I_0$ , ima binomnu distribuciju s parametrima  $S_0 = N - I_0$  i  $p = 1 - q$  tj.

$$I_1 \sim \mathcal{B}(S_0, 1 - q).$$

**Generalno**,  $S_t$  modelira broj podložnih osoba u trenutku  $t$  dok  $I_t$  modelira broj *novih zaraženih osoba* koji uvjetno na poznatu realizaciju slučajnog vektora  $(N, I_0, I_1, \dots, I_{t-1})$  ima binomnu distribuciju s parametrima  $S_t$  i  $p = 1 - q$ . Dakle,

$$I_{t+1} \sim \mathcal{B}(S_t, 1 - q). \quad (9)$$

Pretpostavlja se da u populaciji nema procesa rađanja ni umiranja (tj. brojnost populacije je konstantna u vremenu) te da broj podložnih osoba nije rastuća funkcija vremena. Odnos varijabli, u modelu binomnih epidemioloških lanaca, dan je sljedećom rekurzivnom formulom:

$$S_t = I_{t+1} + S_{t+1} \quad (10)$$

za  $t = 0, 1, 2, \dots$  i  $N = I_0 + S_0$ .

Kako su bitne novozaražene osobe, prethodno zaražene osobe se uklanjaju iz modela nakon što se zaraza pokaže kroz simptome. Epidemija završava kada  $I_t = 0$  jer se u sljedećem trenutku nitko ne može zaraziti, tj. vrijedi  $I_{t+1} = 0$ . Stoga,  $S_{t+1} = S_t$ .

Vjerojatnost zaraze novozaraženih  $p$  za vrijeme  $\langle t, t + 1 \rangle$  može se razlikovati s obzirom na sljedeće zahtjeve:

1.  $p = 1 - q$  ovisi o broju zaraženih u populaciji u trenutku  $t$
2.  $p = 1 - q$  ne ovisi o broju zaraženih u populaciji u trenutku  $t$

Upravo zbog tih vjerojatnosti, razvila su se dva modela binomnih epidemioloških lanaca. **Reed-Frostov** model temeljen na pretpostavci 1. te **Greenwoodov** model temeljen na 2. U svakom trenutku, broj novozaraženih osoba ima **binomnu distribuciju** s parametrima  $S_t$  (broj podložnih osoba) i  $q = 1 - p$  (vjerojatnost izbjegavanja zaraze).

U sljedećem poglavlju analizirat ćemo četiri modela binomnih epidemioloških lanaca koji uz već navedene karakteristike imaju i neke druge - svake specifične za određeni model. Prije toga, uvodimo jednu važnu napomenu koju ćemo često koristiti.

**Napomena 2.1.** Kroz cijeli model, promatra se realizacija različitih lanaca oblika  $\{I_0 = i_0, I_1 = i_1, \dots, I_T = i_T\}$  koji ćemo zapisivati u obliku  $\{i_0, i_1, \dots, i_T\}$ . Zapis prikazuje broj novozaraženih u svakom trenutku tijekom epidemije počevši s brojem zaraženih u  $t = 0$  pa do budućeg trenutka  $T$ .

## 3 Tipovi binomnih epidemioloških modela

### 3.1 Reed-Frostov model

Reed-Frostov model predstavljen je oko 1928. godine od strane biostatističara L.J. Reeda i epidemiologa W.H. Frosta na John Hopkins Sveučilištu. Prikazan je tada pomoću šarenih loptica i drvenih štapića, ali nije objavljen. To je umjesto njih učinila biostatističarka H. Abbey 1952. godine. Prvu računalnu simulaciju napravili su znanstvenici L. Elveback i A. Varma 1965. godine, a determinističku verziju modela opisao je 1980. godine ruski epidemiolog En'ko. Reed-Frostov model pripada ujedno i stohastičkim **SIR** (*eng.* Susceptible-Infected-Recovered) modelima o kojima nešto više možemo naći u [2],[3].

Vrijede sve pretpostavke opisane u poglavlju 2.1 pa tako znamo da u ovom modelu vjerojatnost zaraze  $p$ , odnosno vjerojatnost izbjegavanja zaraze  $q$ , tijekom  $\langle t, t+1 \rangle$  **ovisi** o broju zaraženih osoba  $I_t$  prisutnih u populaciji. Vjerojatnost adekvatnog kontakta s barem jednom podložnom osobom  $S_t$  označava se s  $(1 - q)^{i_t} = p^{i_t}$  pri čemu je  $q^{i_t}$  vjerojatnost izbjegavanja zaraze.

Napokon, obilježje modela je uvjetna vjerojatnost da, ako u trenutku  $t$  imamo  $s_t$  podložnih i  $i_t$  novozaraženih osoba, onda u  $t + 1$  imamo  $i_{t+1}$  novozaraženih osoba te glasi:

$$\begin{aligned} \mathbb{P}(I_{t+1} = i_{t+1} | I_t = i_t, S_t = s_t) &= \binom{s_t}{i_{t+1}} q^{i_t(s_t - i_{t+1})} (1 - q^{i_t})^{i_{t+1}} \\ &= \frac{s_t!}{i_{t+1}! (s_t - i_{t+1})!} q^{i_t(s_t - i_{t+1})} (1 - q^{i_t})^{i_{t+1}}, \quad \forall i_t \in \mathbb{N}_0. \end{aligned} \quad (11)$$

Pomoću rekurzivne formule  $I_{t+1} = S_t - S_{t+1}$  možemo (11) pisati kao:

$$\mathbb{P}(I_{t+1} = i_{t+1} | I_t = i_t, S_t = s_t) = \frac{s_t!}{i_{t+1}! s_{t+1}!} q^{i_t s_{t+1}} (1 - q^{i_t})^{i_{t+1}}.$$

Kako vrijedi da broj podložnih osoba u  $t + 1$ , uvjetno na broj podložnih i novozaraženih u  $t$ , ima binomnu distribuciju s parametrima  $s_t$  i  $q^{i_t}$  lako se izračuna sljedeće:

$$E(S_{t+1} | S_t = s_t, I_t = i_t) = s_t q^{i_t},$$

$$E(I_{t+1} | S_t = s_t, I_t = i_t) = s_t - s_t q^{i_t} = s_t (1 - q^{i_t}),$$

$$\text{Var}(S_{t+1} | S_t = s_t, I_t = i_t) = s_t (1 - q^{i_t}) q^{i_t} = \text{Var}(I_{t+1} | S_t = s_t, I_t = i_t).$$

Sada ćemo na primjeru pokazati kako bi ovaj model funkcionirao. Kako se svi naši modeli koriste u malim zatvorenim populacijama (kućanstva, škole, vrtići, razne grupe, ...) odabran je primjer kućanstva.

**Primjer 1.** Neka se kućanstvo sastoji od tri osobe pri čemu je jedna zaražena. Vjerojatnost da se osoba zarazi je  $p = 0.7$ . Ako sve podložne osobe u trenutku  $t = 0$  dođu u "adekvantan kontakt" s tom jednom zaraženom onda su u  $t = 1$  one sve zaražene. Odnosno, imali bismo  $I_1 = 2$  i  $S_1 = 0$  što karakterizira kraj epidemije. Vjerojatnost takvog događaja u ovom modelu je:

$$\mathbb{P}(I_1 = 2 | I_0 = 1, S_0 = 2) = \frac{2!}{2! 0!} q^{1 \cdot 0} (1 - q^1)^2 = (1 - q)^2 = p^2 = 0.49.$$

S druge strane, očekivanje broja podložnih  $S_t$  i zaraženih osoba  $I_t$  jednako je:

$$E(S_1 | S_0 = 2, I_0 = 1) = s_0 q^{i_0} = 2q = 2 \cdot 0.3 = 0.6,$$

$$E(I_1 | S_0 = 2, I_0 = 1) = s_0 (1 - q^{i_0}) = 2(1 - q) = 2 \cdot 0.7 = 1.4.$$

U sljedećoj tablici dani su svi putevi tj. realizacije lanca zaraze i pripadne vjerojatnosti, gdje koristimo notaciju uvedenu u Napomeni 2.2.

realizacija $\{i_0, i_1, \dots, i_T\}$	vjerojatnost lanca	$p = 0.7$	ukupan broj zaraženih
$\{1, 0\}$	$q^2$	0.09	1
$\{1, 1, 0\}$	$2pq^2$	0.126	2
$\{1, 1, 1\}$	$2p^2q$	0.294	3
$\{1, 2\}$	$p^2$	0.49	3

Tablica 1: Vjerojatnosti u Reed-Frostovom modelu sa  $S_0 = 2$  i  $I_0 = 1$

Ako bismo imali 1000 malih kućanstava kao što je primjerice naše, u njih  $1000 \cdot 0.09 = 90$  zaraza se s početnog zaraženog ne bi proširila ni na koga. Njih  $1000 \cdot 0.126 = 126$  rezultiralo bi s ukupno dvije zaražene osobe dok bismo u  $1000 \cdot (0.294 + 0.490) = 784$  kućanstava očekivali da će se sve podložne zaraziti.

## 3.2 Greenwoodov model

Ovaj model nastao je 1931. godine. Za razliku od Reed-Frost modela, u Greenwoodovom modelu vjerojatnost zaraze  $p$  odnosno vjerojatnost izbjegavanja zaraze  $q = 1 - p$  **ne ovisi** o broju zaraženih osoba  $I_t$ .

Vrijede sve pretpostavke iz poglavlja 2.1. Tako, uvjetna vjerojatnost da, ako u

trenutku  $t$  imamo  $s_t$  podložnih i  $i_t$  novozaraženih osoba, onda u  $t + 1$  imamo  $I_{t+1}$  novozaraženih osoba te glasi:

$$\begin{aligned}\mathbb{P}(I_{t+1} = i_{t+1} | I_t = i_t, S_t = s_t) &= \binom{s_t}{i_{t+1}} q^{(s_t - i_{t+1})} (1 - q)^{i_{t+1}} \\ &= \frac{s_t!}{i_{t+1}! (s_t - i_{t+1})!} q^{(s_t - i_{t+1})} (1 - q)^{i_{t+1}}, \quad \forall i_t \geq 1, s_t \geq i_{t+1}.\end{aligned}\quad (12)$$

Pomoću rekurzivne formule  $I_{t+1} = S_t - S_{t+1}$  možemo (12) zapisati kao:

$$\mathbb{P}(I_{t+1} = i_{t+1} | I_t = i_t, S_t = s_t) = \frac{s_t!}{i_{t+1}! s_{t+1}!} q^{s_{t+1}} (1 - q)^{i_{t+1}}. \quad (13)$$

Primjetimo kako se u (12) i (13) ne pojavljuje  $i_t$  što je specifično za ovaj model. Naime, potreban nam je samo broj novozaraženih  $i_{t+1}$  za izračun vjerojatnosti. Ovo ga čini jednostavnijim modelom od Reed-Frostovog.

Uz rekurzivnu formulu vrijedi i da broj podložnih osoba u  $t + 1$  uvjetno na broj zaraženih i podložnih u  $t$  ima binomnu distribuciju s parametrima  $s_t$  i  $q$  pa imamo:

$$E(S_{t+1} | S_t = s_t) = q s_t,$$

$$E(I_{t+1} | S_t = s_t) = s_t - s_t q = (1 - q) s_t,$$

$$\text{Var}(S_{t+1} | S_t = s_t) = \text{Var}(I_{t+1} | S_t = s_t) = q(1 - q) s_t.$$

Pokažimo i ovdje na primjeru kućanstva kako model funkcionira.

**Primjer 2.** Neka se kućanstvo sastoji od tri osobe pri čemu je jedna zaražena. Vjerojatnost da se osoba zarazi je  $p = 0.7$ . Vjerojatnost istog događaja kao i u Primjeru 1. u ovom modelu bio bi:

$$\mathbb{P}(I_1 = 2 | I_0 = 1, S_0 = 2) = \frac{2!}{2! 0!} q^0 (1 - q)^2 = p^2 = (0.7)^2 = 0.49.$$

Uz to, očekivani broj podložnih  $S_t$  i zaraženih osoba  $I_t$  jednako je:

$$E(S_1 | S_0 = 2) = 2q = 0.6 \quad E(I_1 | (S_0 = 2)) = 2(1 - q) = 1.4.$$

U sljedećoj tablici dani su svi putevi tj. realizacije lanaca zaraze i pripadna vjerojatnost.

$\{i_0, i_1, \dots, i_T\}$	vjerojatnost lanca	$p = 0.7$	ukupan broj zaraženih
$\{1,0\}$	$q^2$	0.09	1
$\{1,1,0\}$	$2pq^2$	0.126	2
$\{1,1,1\}$	$2p^2q$	0.294	3
$\{1,2\}$	$p^2$	0.49	3

Tablica 2: Vjerojatnosti u Greenwoodovom modelu sa  $S_0 = 2$  i  $I_0 = 1$

Možemo vidjeti da imamo **iste** vrijednosti kao i za Reed-Frostov model u Primjeru 1.

Međutim, kada bismo započeli s **dvije** zaražene osobe imali bismo drugačiju situaciju. U tom slučaju postoje samo dva događaja koja se mogu realizirati.

1. U  $t = 1$  i dalje imamo samo jednu podložnu osobu bez ijedne novozaražene jer ne dođe do "adekvatnog kontakta". Vjerojatnost takvog događaja je:

$$\mathbb{P}(I_1 = 0 | I_0 = 2, S_0 = 1) = \frac{1!}{0! 1!} q^1 (1 - q)^0 = q = 0.3.$$

2. U  $t = 1$  dođe do "adekvatnog kontakta" i podložna osoba se zarazi. Vjerojatnost takvog događaja je:

$$\mathbb{P}(I_1 = 1 | I_0 = 2, S_0 = 1) = \frac{1!}{1! 0!} q^0 (1 - q)^1 = p = 0.7.$$

U Reed-Frostovom modelu, za isti događaj, imali bismo sljedeće vjerojatnosti:

1.  $\mathbb{P}(I_1 = 0 | I_0 = 2, S_0 = 1) = \frac{1!}{0! 1!} q^{1 \cdot 2} (1 - q)^0 = q^2 = 0.09.$
2.  $\mathbb{P}(I_1 = 1 | I_0 = 2, S_0 = 1) = \frac{1!}{1! 1!} q^0 (1 - q^2)^1 = 1 - q^2 = 0.91.$

Usporedbu vjerojatnosti pri realizaciji lanaca u oba modela možemo vidjeti u sljedećoj tablici.

lanac	Reed-Frost model	Greenwood model
$\{2,0\}$	$q^2$	$q$
$\{2,1\}$	$1 - q^2$	$p$

Tablica 3: Usporedba vjerojatnosti za oba modela u kućanstvu s  $S_0 = 1$  i  $I_0 = 2$

Kada bismo u ovom slučaju uzeli 1000 tročlanih kućanstava u kojima su dvije zaražene osobe u  $t = 0$ , očekivali bismo da se epidemija ne proširi u  $1000 \cdot 0.3 = 300$  kućanstava dok bismo u njima  $1000 \cdot 0.7 = 700$  očekivali da se proširi, ali i da stane jer u  $t = 1$  nestane podložnih osoba.



### 3.3 Modificirani Reed-Frostov model

Reed-Frost model modificiran je tako što podložne osobe nakon što se zaraze mogu ponovno postati podložne. Dakle, ne bivaju maknute iz modela kao prije. Takav model sličan je klasi **SIS** (*eng.* Susceptible-Infected-Susceptible) modela o kojima također nešto više možemo naći u [2],[3].

Rekurzivna formula bitna za ovaj model sada je

$$N = I_t + S_t.$$

Kako i dalje promatramo broj novoizaraženih, broj podložnih osoba se više ne promatra. Prema tome, za vjerojatnost da u  $t + 1$  imamo  $I_{t+1}$  bitno je samo znati koliko je  $I_t$  u  $t$ . Takva uvjetna vjerojatnost iznosi:

$$\mathbb{P}(I_{t+1} = i_{t+1} | I_t = i_t) = \binom{N - i_t}{i_{t+1}} (1 - q^{i_t})^{i_{t+1}} q^{i_t(N - i_t - i_{t+1})}, \quad i_t + i_{t+1} \leq N.$$

U ovome modelu, epidemija završava isključivo u slučaju  $I_t = 0$  jer će podložne osobe  $S_t$  uvijek postojati. Kako se i dalje radi o binomnoj distribuciji, imamo sljedeće vrijednosti:

$$E(I_{t+1} | I_t = i_t) = (N - i_t)(1 - q^{i_t}),$$

$$\text{Var}(I_{t+1} | I_t = i_t) = (N - i_t)(1 - q^{i_t})q^{i_t}.$$

### 3.4 Generalizirani model binomnih epidemioloških lanaca

Primjer 2. ukazao je na jaku sličnost između Reed-Frost i Greenwood modela. Upravo iz tog razloga nastaje model koji ih "spaja". Naime, 1981. godine Neils Becker kombinira Reed-Frostov model s Greenwoodovim te dobiva *generalizirani model binomnih epidemioloških lanaca*. Vidjet ćemo da je ovaj model zapravo sažetak svega prije navedenog i definiranog. Koristimo se istim oznakama kao i do sada, uz početne vrijednosti  $I_0 = a$  i  $S_0 = k$  te imamo sljedeće:

$$\mathbb{P}(S_{t+1} = x, I_{t+1} = s - x | S_t = s, I_t = i) = \binom{s}{x} q_i^x (1 - q_i)^{s-x}$$

za  $x = 0, 1, \dots, s$ .

Vjerojatnost bilo koje realizacije lanca zaraženih osoba  $\{a, i_1, i_2, \dots, i_T\}$  za vrijeme  $t = 0, 1, \dots, T$  iznosi:

$$\mathbb{P}(\{a, i_1, i_2, \dots, i_T\}) = \frac{s_0!}{i_1! i_2! \dots i_T! s_{T+1}!} \prod_{t=0}^T q_{i_t}^{s_{t+1}} p_{i_t}^{i_{t+1}}, \quad (14)$$

gdje  $i_0 = a, s_t = k - i_1 - i_2 - \dots - i_t$  za  $t = 1, 2, \dots, T$ .

Ovim modelom moguće je pratiti razvoj epidemije kroz realizacije lanaca. Recimo, za realizaciju lanca  $\{1, 2, 1, 0\}$  u peteročlanom kućanstvu s jednim zaraženim imali bismo:

$$\begin{aligned} & \mathbb{P}(S_1 = 2 | I_0 = 1, S_0 = 4) \mathbb{P}(S_2 = 1 | I_1 = 2, S_1 = 2) \mathbb{P}(S_3 = 0 | I_2 = 1, S_2 = 1) \\ &= \mathbb{P}(\{1, 2, 1, 0\}) = \binom{4}{2} q_1^2 p_1^2 \binom{2}{1} q_2 p_2 \binom{1}{0} q_1^0 p_1^0 = 12 q_1^3 p_1^2 q_2 p_2, \end{aligned}$$

gdje je  $p_i = 1 - q_i$  za  $i = 1, 2, \dots$

Sada zamjenom  $q_i = q^i$  imamo Reed-Frostov model, točnije:

$$\mathbb{P}(\{a, i_1, i_2, \dots, i_k\}) = \frac{s_0!}{i_1! i_2! \dots i_k! s_{k+1}!} \prod_{j=0}^{k-1} (1 - q^{i_j})^{i_{j+1}} q^{\sum_{j=0}^k i_j} s_{j+1}.$$

Zamjenom  $q_i = q$ , pak, dobivamo Greenwoodov model:

$$\mathbb{P}(\{a, i_1, i_2, \dots, i_k\}) = \frac{s_0!}{i_1! i_2! \dots i_k! s_{k+1}!} p^{\sum_{j=1}^k i_j} q^{\sum_{j=1}^{k+1} s_j}.$$

Ovakav račun može se činiti kompliciran, no olakšan je pomoću **rekurzivnih formula** za oba modela.

Neka je  $N$  veličina populacije s  $a$  početnim brojem zaraženih osoba i  $j$  ukupnim brojem slučajeva za koje vrijedi ( $a \leq j \leq N$ ). Vjerojatnost novih  $j - a$  slučajeva, uključujući  $k$  zaraženih, označavamo s  ${}_k P_{N-a, j-a}$ .

Sada, rekurzivna formula za vjerojatnost da imamo  $k$  zaraženih u Greenwoodovom modelu glasi:

$${}_a P_{N, j} = \sum_{k=1}^{j-a} \binom{N-a}{k} p^b q^{N-a-k} {}_k P_{N-a, j-a},$$

gdje vrijedi  ${}_a P_{N, a} = q^{N-a}$ . Za Reed-Frostov model, s druge strane, imamo sljedeću formulu:

$${}_a P_{N, j} = \sum_{k=1}^{j-a} \binom{N-a}{k} (1 - q^a)^k q^{a(N-a-k)} {}_k P_{N-a, j-a},$$

gdje  ${}_a P_{N, a} = q^{a(N-a)}$ .

Svaku fazu epidemije prati binomna distribucija s  $k$  zaraženih i  $N - a - k$  podložnih osoba.

Generalizirani model bit će nam od iznimne važnosti kada ćemo modelirati na stvarnim podacima. K tome je još dobar uvod u sljedeće poglavlje u kojem ćemo opisati svaki naš model pomoću Markovljevih lanaca.

## 4 Prikaz modela preko Markovljevih lanaca

Matematičari J. Gani i D. Jerwood 1971. godine razradili su oba modela pomoću Markovljevih lanaca te tako pojednostavili praćenje epidemije. Daju jednostavniju metodu za računanje **funkcije izvodnice vjerojatnosti** vremena trajanja epidemije  $T$ .

**Definicija 4.1.** Neka je  $X$  diskretna slučajna varijabla s vrijednostima u skupu  $\mathbb{N}_0$  te neka je njena distribucija dana pravilom  $p_k = \mathbb{P}(X = k)$ ,  $k \in \mathbb{N}_0$ . **Funkcija izvodnica vjerojatnosti** slučajne varijable  $X$  je funkcija  $g(x)$  definirana kao:

$$g_X(t) = E[t^X] = \sum_{k=0}^{\infty} t^k p_k$$

za realne brojeve  $t$  za koje  $E[|t^X|] = \sum_{k=0}^{\infty} |t^k| p_k < \infty$ .

Realizacija epidemije tada postaje promatrani slijed vrijednosti formuliranih preko Markovljevih lanaca. Taj slijed završava kada nema više zaraznih osoba u populaciji dok je skup stanja konačan jer ne postoje više podložne osobe koje se mogu zaraziti.

Oni Reed-Frostov i Greenwoodov model gledaju kao klasu Markovljevih lanaca  $(X_t, t \in \mathbb{N})$  sa skupom stanja  $S \in \{0, 1, 2, \dots, k\}$  i matricom prijelaznih vjerojatnosti  $M = \{m_{ij}, i, j \in S\}$  dimenzije  $(k + 1) \times (k + 1)$  čije su vrijednosti na dijagonali strogo pozitivne te vrijedi

$$0 \leq m_{ij} \leq 1, \quad i \neq j,$$

$$0 \leq m_{jj} \leq 1, \quad i, j = 0, 1, \dots, k.$$

Ovdje slučajna varijabla  $X_t$  predstavlja broj podložnih osoba u epidemiji pri čemu epidemija završava kada je  $X_t = X_{t-1}$ . Slučajna varijabla  $T$  definirana je kao  $T = \min\{t > 0 : X_t = X_{t-1}\}$  te i dalje modelira vrijeme trajanja epidemije.

Matricu  $P$  potrebnu za računanje funkcije izvodnice vjerojatnosti dobit ćemo pomoću matrice  $M$ . Uz nju, uvodimo matrice  $A_i$  i  $R$  čije ćemo vrijednosti prikazati kasnije, za sada služe kao motivacija u računanju funkcije izvodnice vjerojatnosti.

Za  $t > 0$  i  $j = 0, 1, 2, \dots, k$  vrijedi:

$$\mathbb{P}(T = t, X_t = j | X_0 = i) = \mathbb{P}(T = t, X_t = X_{t-1} = j | X_0 = i) = (P^{t-1})_{ij} m_{jj} \quad (15)$$

Zbrajanjem (15) preko  $0 \leq j \leq k$  dobivamo:

$$\mathbb{P}(T = t | X_0 = i) = \sum_{j=0}^k \mathbb{P}(T = t, X_t = X_{t-1} = j | X_0 = i) = \sum_{j=0}^k (P^{t-1})_{ij} m_{jj}$$

$$= A_i P^{t-1} R, \quad 0 \leq t < \infty. \quad (16)$$

Za proizvoljne  $0 \leq \theta, \phi \leq 1$  vrijedi:

$$E[\theta^T \phi^{X_t} | X_0 = i] = \sum_{t=1}^{\infty} \sum_{j=0}^k \theta^t \phi^j A_i P^{t-1} R,$$

gdje je  $A_i = [0, 0, \dots, 0, 1]$  vektor redak s jedinicom na  $(i+1)$ -tom mjestu.  $R$  je oblika  $[m_{00}, m_{11}, \dots, m_{kk}]^T$ .  $P = \{p_{ij}, i, j \in S\}$  je  $(k+1) \times (k+1)$  matrica za koju vrijedi:

$$p_{ij} = \begin{cases} m_{ij}, & i \neq j \\ 0, & i = j. \end{cases}$$

Sada,

$$E[\theta^T \phi^{X_t} | X_0 = i] = \sum_{t=1}^{\infty} \theta^t A_i (P(\phi))^{t-1} R,$$

pri čemu je  $A_i = [0, 0, \dots, \phi^i]$  vektor redak sada s  $\phi^i$  na  $(i+1)$ -tom mjestu.  $P(\phi)$  je  $(k+1) \times (k+1)$  matrica koja na poziciji  $(i, j)$  ima vrijednosti  $p_{ij} \phi^{j-i}$ .

Napokon, **funkcija izvodnica vjerojatnosti** slučajne varijable  $T$  uvjetno na  $X_0$  može se zapisati kao:

$$E[\theta^T \phi^{X_t} | X_0 = i] = A_i(\phi)(I - \theta P(\phi))^{-1} \theta R. \quad (17)$$

Matrica  $I$  je  $(k+1) \times (k+1)$  jedinična matrica, a kako vrijedi  $|\theta P| < 1$  onda postoji inverz od (17).

J. Gani i D. Jerwood distribuciju (16) nazivaju *Markovljevom geometrijskom distribucijom* jer jako slični klasičnoj geometrijskoj koja je

$$P(X = x) = p^{t-1} q, \quad 0 < p = 1 - q < 1, \quad t = 1, 2, \dots$$

čija funkcija izvodnica glasi:

$$g(t) = (1 - tp)^{-1} tq.$$

## 4.1 Reed-Frostov model

U ovome modelu radi se o *dvodimenzionalnom* Markovljevom lancu  $((S_t, I_t), t \in \mathbb{N}_0)$ . Kao skup stanja  $S$  uzima se  $\{0, 1, 2, \dots, k\} \times \{0, 1, 2, \dots, k\}$ . Matrica  $M$  je dimenzije  $(k+1)^2 \times (k+1)^2$  te se sastoji od  $(k+1)^2$  podmatrica  $(m_{ij})_{rs}$ ; svaka

dimenzije  $(k+1) \times (k+1)$ . Za  $r, s \in \{0, 1, 2, \dots, k\}$ , matrica  $(m_{ij})_{rs}$  definirana je kao:

$$(m_{ij})_{rs} = \begin{cases} \binom{s+j}{j} (1-q)^{ij} q^{js}, & r = s+j \\ 0 & \text{inače.} \end{cases}$$

$R$  je  $k(k+1) \times 1$  vektor stupac nastao spajanjem dijagonala matrica  $[(m_{ij})_{10}, (m_{ij})_{20}, \dots, (m_{ij})_{k0}]$ , a  $P(\phi)$  je  $k(k+1) \times k(k+1)$  matrica koja se sastoji od  $k^2$  podmatrica  $(P(\phi)_{ij})_{rs}$  dimenzije  $(k+1) \times (k+1)$ . Za  $r, s \in \{0, 1, 2, \dots, k\}$ , matrica  $(P(\phi)_{ij})_{rs}$  onda je definirana kao

$$(P(\phi)_{ij})_{rs} = (m_{ij})_{rs} \phi^{r-s}.$$

Sve gornje vrijednosti dosta su komplicirane, no ono najbitnije možemo zapisati na sljedeći način:

- *Prijelazne vjerojatnosti* u matrici  $M$  iz stanja  $(s_t, i_t)$  u stanje  $(s_{t+1}, i_{t+1})$  su binomne te iznose:

$$\begin{aligned} \mathbb{P}(S_{t+1} = s_t - i_{t+1}, I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t) &= p_{(s,i)_{t+1}, (s,i)_t} = p_{s_{t+1}, s_t} \\ &= \binom{s_t}{s_{t+1}} q^{i_t s_{t+1}} (1 - q^{i_t})^{s_t - s_{t+1}} = \frac{s_t!}{i_{t+1}! (s_t - i_{t+1})!} q^{i_t (s_t - i_{t+1})} (1 - q^{i_t})^{i_{t+1}}. \end{aligned}$$

- *Skup stanja*  $S$  je  $S = \{0, 1, \dots, S_0\}^2$ .
- *vrijeme zaustavljanja*  $T$  jednako je:

$$T = \inf_{t \geq 0} \{t : S_t I_t = 0\}. \quad (18)$$

- *veličina*  $W$  epidemije je slučajna varijabla koja modelira ukupan broj ikada zaraženih osoba za trajanja epidemije i iznosi:

$$W = S_0 - S_T. \quad (19)$$

- Realizacija epidemije je i dalje lanac  $\{i_0, i_1, \dots, i_T\}$  čija je vjerojatnost dana umnoškom uvjetnih binomnih vjerojatnosti kao:

$$\begin{aligned} &\mathbb{P}(I_1 = i_1 | I_0 = i_0, S_0 = s_0) \cdots \mathbb{P}(I_T = i_T | I_{T-1} = i_{T-1}, S_{T-1} = s_{T-1}) \\ &= \prod_{t=0}^{T-1} \binom{s_t}{i_{t+1}} q^{i_t (s_t - i_{t+1})} (1 - q^{i_t})^{i_{t+1}}. \end{aligned} \quad (20)$$

**Primjer 3.** Neka se kućanstvo sastoji od pet osoba od kojih je jedna zaražena dok ostale četiri nisu. Započinjemo u  $t = 0$  te gledamo koliko se osoba zarazi u sljedećim trenutcima. Jedna od opcija je zaraza svih članova pa u  $t = 1$  imamo sljedeću vjerojatnost:

$$\mathbb{P}(S_1 = 0, I_1 = 4 | I_0 = 1, S_0 = 4) = \frac{4!}{4!0!} q^{1 \cdot 0} (1 - q^1)^4 = (1 - q)^4 = p^4.$$

Ostale realizacije događaja dobivamo analognim postupkom pomoću (18)-(20) čije su vrijednosti i vjerojatnosti dane u Tablici 4.

$\{i_0, i_1, \dots, i_T\}$	$T$	$W$	$p_{(s,i)_{t+1}, (s,i)_t}$
{1}	0	0	$q^4$
{1,1}	1	1	$4pq^6$
{1,1,1}	2	2	$12p^2q^7$
{1,2}	1	2	$6p^2q^6$
{1,1,1,1}	3	3	$24p^3q^7$
{1,1,2}	2	3	$12p^3q^6$
{1,2,1}	2	3	$12p^3q^5(1 + q)$
{1,3}	1	3	$4p^3q^4$
{1,1,1,1,1}	4	4	$24p^4q^6$
{1,1,1,2}	3	4	$12p^4q^5$
{1,1,2,1}	3	4	$12p^4q^4(1 + q)$
{1,2,1,1}	3	4	$12p^4q^4(1 + q)$
{1,2,2}	2	4	$6p^4q^2(1 + q)^2$
{1,1,3}	2	4	$4p^4q^3$
{1,3,1}	2	4	$4p^4q(1 + q + q^2)$
{1,4}	1	4	$p^4$

Tablica 4: Sve realizacije lanaca Reed-Frostovog modela pri zarazi u peteročlanom kućanstvu s  $I_0 = 1$

Distribucija *ukupnog* broja slučajeva u epidemiji tj. distribucija slučajne varijable  $W$  dobiva se zbrajanjem svih vjerojatnosti za određeni lanac. Stoga, imamo sljedeću tablicu.

Broj slučajeva	Distribucijske vrijednosti
1	$\mathbb{P}(W = 0) = q^4$
2	$\mathbb{P}(W = 1) = 4pq^6$
3	$\mathbb{P}(W = 2) = 6p^2q^6(1 + 2q)$
4	$\mathbb{P}(W = 3) = 4p^3q^4(1 + 3q + 6q^2 + 6q^3)$
5	$\mathbb{P}(W = 4) = p^4(1 + 4q + 10q^2 + 20q^3 + 30q^4 + 36q^5 + 24q^6)$

Tablica 5: Distribucija veličine epidemije u peteročlanom kućanstvu s  $I_0 = 1$

## 4.2 Greenwoodov model

U ovome modelu ne zanima nas broj podložnih osoba modeliran s  $X_t$  pa će (17) ovisiti samo o vremenu trajanja epidemije tj. o slučajnoj varijabli  $T$ . Matrica  $R$  sada je  $R = [1, q, q^2, \dots, q^k]^T$ , a za  $M = \{m_{ij}\}$  vrijedi:

$$m_{ij} = \begin{cases} \binom{i}{j} p^{i-j} q^j, & 0 \leq j \leq i \\ 0, & 0 < j \leq k. \end{cases}$$

i  $P = M - R$ . Radi se o Markovljevom lancu  $(I_t, t \in \mathbb{N}_0)$  sa skupom stanja  $S = \{0, 1, 2, \dots, S_0 = N - I_0\}$  i matricom prijelaznih vjerojatnosti  $M = \{m_{ij}, i, j \in S\}$ . Dakle, funkcija izvodnica vjerojatnosti, u ovom modelu jednaka je:

$$E[\theta^T | X_0 = i] = \sum_{t=0}^{\infty} A_i \theta^{t-1} P^{t-1} \theta R = A_i (I - \theta P)^{-1} \theta R, \quad 0 \leq \theta \leq 1. \quad (21)$$

Dobivena je pomoću  $S_0 + 1$  od  $(I - \theta P(\phi))^{-1} \theta R$  gdje je  $P(\phi)$  matrica dimenzije  $(S_0 + 1) \times (S_0 + 1)$  čije su  $(i, j)$  vrijednosti jednake  $p_{ij} \phi^{i-j}$ .

Ukupan broj zaraženih jednak je  $S_0 - S_T = (S_0 - S_1) + (S_1 - S_2) + \dots + (S_{T-1} - S_T)$ .

U ovom modelu jednostavno je izračunati očekivano vrijeme trajanja epidemije dok za ostale to nije slučaj. Naime, ono se dobiva derivacijom (21) u koju unosimo  $\theta = 1$  za  $i$  podložnih osoba u početnom trenutku  $t = 0$ . Imamo:

$$E[T | X_0 = i] = A_i (I - P)^{-2} R,$$

$$\text{Var}(T | X_0 = i) = 2A_i P (I - P)^{-3} R + A_i (I - P)^{-2} R - A_i (I - B)^{-2} R^2.$$

Analogno zapisu u Reed-Frostovom modelu, ovdje vrijedi:

- *Prijelazne vjerojatnosti* iz stanja  $(s_t, i_t)$  u stanje  $(s_{t+1}, i_{t+1})$  su binomne i iznose:

$$\begin{aligned} \mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t) &= p_{(s,i)_{t+1}, (s,i)_t} = p_{s_{t+1}, s_t} = \binom{s_t}{s_{t+1}} q^{s_{t+1}} (1 - q)^{s_t - s_{t+1}} \\ &= \frac{s_t!}{s_{t+1}! (s_t - s_{t+1})!} q^{s_{t+1}} (1 - q)^{s_t - s_{t+1}} = \frac{s_t!}{s_{t+1}! (s_t - s_{t+1})!} q^{s_{t+1}} p^{s_t - s_{t+1}}. \end{aligned}$$

- *Skup stanja* je  $S = \{0, 1, \dots, S_0\}$  gdje  $s_0 > 0$  i  $i_0 > 0$ .
- *vrijeme zaustavljanja*  $T$  je  $T = \{0, 1, \dots, S_0\}$ .

- *veličina*  $W$  epidemije ista je kao i u Reed-Frostovom modelu:

$$W = S_0 - S_T, \quad W \in \{0, 1, \dots, s_0\}$$

pri čemu je  $\max W = s_0$ .

- Lanac  $\{i_0, i_1, \dots, i_T\}$  je realizacija epidemije za kojeg vrijedi:

$$\mathbb{P}(S_1 = s_1 | S_0 = s_0) \mathbb{P}(S_2 = s_2 | S_1 = s_1) \cdots \mathbb{P}(S_T = s_T | S_{T-1} = s_{T-1})$$

$$= \prod_{t=0}^{T-1} \binom{s_t}{s_{t+1}} q^{s_{t+1}} p^{s_t - s_{t+1}}.$$

**Primjer 4.** Neka se kućanstvo sastoji od pet osoba od kojih je jedna zarazna dok ostale četiri nisu. Započinjemo u  $t = 0$  te gledamo koliko se osoba zarazi u sljedećim trenutcima. Ako bismo sada uzeli situaciju da se zaraze tri osobe, odnosno samo jedna osoba izbjegne zarazu, u  $t = 1$  imali bismo sljedeće:

$$\mathbb{P}(S_1 = 1 | S_0 = 4) \mathbb{P}(S_2 = 1 | S_1 = 1) = \frac{4!}{3! 1!} q^1 (1 - q)^{4-1} \frac{1!}{1! 0!} q^1 (1 - q)^{1-1} = 4q^2 p^3.$$

Ostale realizacije dane su u Tablici 6.

$\{i_0, i_1, \dots, i_T\}$	$T$	$W$	$p_{(s,i)_{t+1}, (s,i)_t}$
$\{1\}$	0	0	$q^4$
$\{1,1\}$	1	1	$4pq^6$
$\{1,1,1\}$	2	2	$12p^2q^7$
$\{1,2\}$	1	2	$6p^2q^4$
$\{1,1,1,1\}$	3	3	$24p^3q^7$
$\{1,1,2\}$	2	3	$12p^3q^5$
$\{1,2,1\}$	2	3	$12p^3q^4$
$\{1,3\}$	1	3	$4p^3q^2$
$\{1,1,1,1,1\}$	4	4	$24p^4q^6$
$\{1,1,1,2\}$	3	4	$12p^4q^5$
$\{1,1,2,1\}$	3	4	$12p^4q^4$
$\{1,2,1,1\}$	3	4	$12p^4q^3$
$\{1,2,2\}$	2	4	$6p^4q^2$
$\{1,1,3\}$	2	4	$4p^4q^3$
$\{1,3,1\}$	2	4	$4p^4q$
$\{1,4\}$	1	4	$p^4$

Tablica 6: Sve realizacije lanaca Greenwoodovog modela pri zarazi u peteročlanom kućanstvu s  $I_0 = 1$

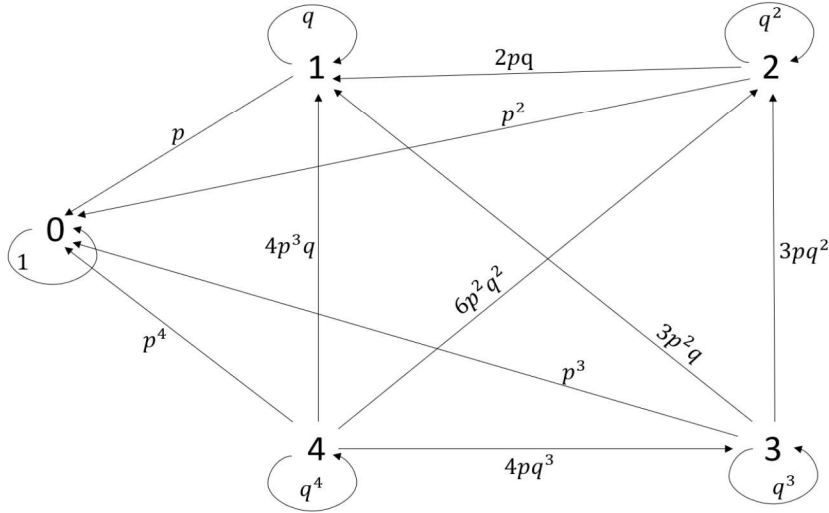
Analogno Reed-Frostovom modelu imamo sljedeću tablicu:



Broj slučajeva	Distribucijske vrijednosti
1	$\mathbb{P}(W = 0) = q^4$
2	$\mathbb{P}(W = 1) = 4pq^6$
3	$\mathbb{P}(W = 2) = 6p^2q^4(1 + 2q^3)$
4	$\mathbb{P}(W = 3) = 4p^3q^2(1 + 3q + 3q^3 + 6q^5)$
5	$\mathbb{P}(W = 4) = p^4(1 + 4q + 6q^2 + 16q^3 + 12q^4 + 12q^5 + 24q^6)$

Tablica 7: Distribucija veličine epidemije u peteročlanom kućanstvu s  $I_0 = 1$

Skup stanja u ovom primjeru je  $S = \{0, 1, 2, 3, 4\}$  pri čemu je  $p_{s_{t+1}, s_t}$  vjerojatnost dolaženja u određeno stanje. Sva stanja, također, mogu biti kraj epidemije kada se dođe u njih. Na sljedećoj slici prikazana je realizacija epidemije za  $S_0 = 4$ ,  $I_0 = 1$ , ali ovaj put kroz realizacije lanca  $\{s_0, s_1, \dots, s_T\} = \{4, 3, 2, 1, 0\}$ . Npr. lanac  $\{i_0, i_1, i_2\} = \{1, 1, 2\}$  sada postaje  $\{s_0, s_1, s_2\} = \{4, 3, 1\}$ .



Slika 1: Vjerojatnosti realizacije lanca  $\{s_0, s_1, \dots, s_T\}$  u peteročlanom kućanstvu za  $S_0 = 4$  u Greenwodovom modelu

### 4.3 Modificirani Reed-Frostov model

Ovdje se, također, radi o Markovljevom lancu  $(I_t, t \in \{0, 1, \dots, n\})$ , ali je skup stanja ovaj put jednak  $S = \{0, 1, \dots, N\}$ . Matrica prijelaznih vjerojatnosti  $M = \{m_{ij}, i, j \in S\}$  dimenzije  $N \times N$  definirana je kao:

$$m_{ij} = \begin{cases} \binom{N-i}{j} q^{i(N-i-j)} (1-q)^j, & i+j \leq N \\ 0, & i+j > N. \end{cases}$$

Kako u ovom modelu znamo da epidemija završava isključivo za  $I_t = 0$ , broj podložnih osoba nije bitan. Zbog toga  $X_t$  postaje slučajna varijabla koja modelira

broj **zaraženih** osoba u trenutku  $t$ . Zapis (15) postaje:

$$\mathbb{P}(T = t, X_t = 0 | X_0 = i) = \sum_{j=1}^{N-1} (P^{t-1})_{ij} m_{j0} = A_i P^{t-1} R$$

pri čemu je  $A_i$  vektor redak s jedinicom na  $i$ -tom mjestu. Matrica  $R$  je vektor stupac  $[m_{10}, m_{20}, \dots, m_{N-1,0}]^\tau$ , a za  $P = \{p_{ij}, i, j \in S\}$  dimenzije  $(N-1) \times (N-1)$  vrijedi sljedeće:

$$p_{ij} = \begin{cases} m_{ij}, & i + j \leq N - 1 \\ 0, & i + j > N - 1. \end{cases}$$

Funkcija izvodnica vjerojatnosti  $T$ , uvjetno na  $X_0$  je oblika:

$$E[\theta^T | X_0 = i] = A_i \left( \sum_{t=1}^{\infty} \theta^t P^{t-1} \right) R = A_i (I - \theta P)^{-1} \theta R$$

za  $0 \leq \theta \leq 1$  s jediničnom matricom  $I$  veličine  $(N-1) \times (N-1)$ .

## 5 Primjena na stvarnim podacima

Prvi koji su pokazali funkcionalnost modela bili su **M. A. Heasman** i **D. D. Reid** 1961. godine. Prikupili su podatke o trajanju obične prehlade u londonskim peteročlanim kućanstvima. Obitelji s dva roditelja i troje djece koji žive u različitim kvartovima posjećivane su u dvotjednim intervalima. Početni datum prehlade te broj infekcija gornjih dišnih puteva zabilježavao se na vremenskim grafikonima. Upravo ti grafikoni pružili su detaljan uvid u bolest te omogućili procjenu novog slučaja prehlade pomoću lanaca. Zabilježena su 664 slučaja u dvije godine.

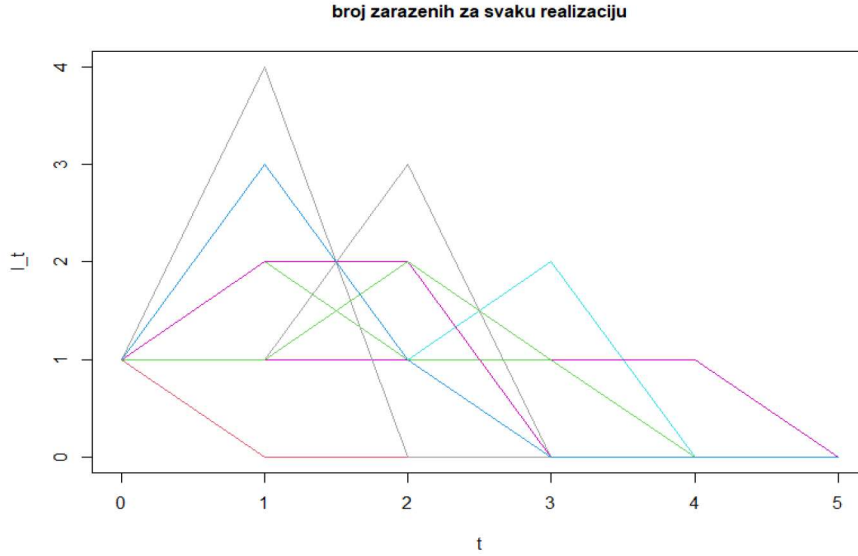
Kako se radi o malom kućanstvu, sve pretpostavke modela mogu biti zadovoljene pa su ovi podaci idealni za ispitivanje valjanosti modela binomnih epidemioloških lanaca.

Promatrane vrijednosti koje su nam Heasman i Reid pružili dane su u sljedećoj tablici dok su svi putevi lanaca za peteročlano kućanstvo prikazani na Slici 2.

$\{i_0, i_1, \dots, i_T\}$	<i>Stvarna frekvencija</i>
{1}	423
{1,1}	131
{1,1,1}	36
{1,2}	24
{1,1,1,1}	14
{1,1,2}	8
{1,2,1}	11
{1,3}	3
{1,1,1,1,1}	4
{1,1,1,2}	2
{1,1,2,1}	2
{1,2,1,1}	3
{1,2,2}	1
{1,1,3}	2
{1,3,1}	0
{1,4}	0

Tablica 8: Heasmanove i Reidove promatrane frekvencije podataka

Iz tablice se može iščitati kako se u najviše slučajeva nitko nije zarazio od jednog prehladenog. S druge strane, nisu ni zabilježeni slučajevi da su se svi članovi kućanstva odmah zarazili u sljedećem trenutku.



Slika 2: Realizacija puteva lanca  $\{i_0, i_1, \dots, i_T\}$

Prvi model koji ćemo procjenjivati je *generalizirani model binomnih epidemioloških lanaca*, što se dalo i naslutiti jer, kao što smo rekli, on "spaja" dva modela u jedan. Za procjenjivanje parametara potrebne su metode procjene od kojih ćemo mi koristiti jednu od najpoznatijih - metodu maksimalne vjerodostojnosti.

Kako je nama parametar  $q$  od veće važnosti nego  $p$ , upravo njega procjenjujemo. Vjerojatnost izbjegavanja zaraze prilikom adekvatnog kontakta sa  $i$ -tom zaraženom osobom bilo koje generacija označavamo s  $q_i$ ,  $i = 1, 2, \dots, k-1$ . Postupak traženja procjenitelja maksimalne vjerodostojnosti započinjemo s definiranjem funkcije vjerodostojnosti i njenog logaritma gdje će  $m_j$  označavati ukupan broj jedinki koje su bile u adekvatnom kontaktu s  $j$ -tom zaraženom dok  $x_j$  označava ukupan broj jedinki koje su izbjegli zarazu. Oznaka  $C$  je konstanta, a postupak je sljedeći:

$$\mathbb{L}(q_1, \dots, q_{k-1}) = \prod_{j=1}^{k-1} \binom{s_j}{x_j} q_j^{x_j} (1 - q_j)^{s_j - x_j}$$

$$l(q_1, \dots, q_{k-1}) = \ln \mathbb{L}(q_1, \dots, q_{k-1}) = \sum_{j=1}^{k-1} [\ln \binom{s_j}{x_j} + x_j \ln q_j + (s_j - x_j) \ln(1 - q_j)] + C$$

$$\frac{d}{dq} l(q_1, \dots, q_{k-1}) = \sum_{j=1}^{k-1} \frac{x_j - q_j s_j}{1 - q_j} = 0 \Rightarrow \hat{q}_j = \frac{x_j}{s_j} \quad (22)$$

za  $j = 1, 2, \dots, k-1$ .

Za Heismanove i Reidove podatke, funkcija vjerodostojnosti je:

$$\mathbb{L}(q_1, q_2, q_3) = C(nq_1^4)^4 23(4nq_1^6 p_1)^{131} \dots (np_1^4)^4$$

$$\mathbb{L}(q_1, q_2, q_3) = Cq_1^{3000}(1 - q_1)^{397}q_2^{70}(1 - q_2)^{18}q_3^3,$$

pa se iz (22) vrijednosti  $s_j$  i  $x_j$  lako daju iščitati. Dakle, vrijedi:

$$x_1 = 3000, \quad x_2 = 70, \quad x_3 = 3, \quad s_1 = 3397, \quad s_2 = 88, \quad s_3 = 3.$$

Procjenitelji su:  $\hat{q}_1 = 0.8831$ ,  $\hat{q}_2 = 0.7955$ ,  $\hat{q}_3 = 1$ . Kako je  $\hat{q}_3$  baziran na samo tri Bernoullijeva pokusa u njega se ne može pouzdati te ga izbacujemo iz procjene. Dakle, ispituju se samo  $\hat{q}_1$  i  $\hat{q}_2$  za koje uzimamo da je  $n = 664$ . Korak po korak, dobivamo sljedeću tablicu.

$\{i_0, i_1, \dots, i_T\}$	Stvarna	Očekivana	Dobivena
{1}	423	$nq_1^4$	403.9
{1,1}	131	$4nq_1^6p_1$	147.3
{1,1,1}	36	$12nq_1^6p_1^2$	45.6
{1,2}	24	$6nq_1^2p_1^2q_2^2$	26.9
{1,1,1,1}	14	$12nq_1^7p_1^3$	10.7
{1,1,2}	8	$12nq_1^7p_1^3q_2$	6.2
{1,2,1}	11	$12nq_1^3p_1^2q_2p_2$	12.2
{1,3}	3	$4nq_1p_1^3q_3$	3.7
{1,1,1,1,1}	4	$24nq_1^6p_1^4$	1.4
{1,1,1,2}	2	$12nq_1^5p_1^4$	0.8
{1,1,2,1}	2	$12nq_1^4p_1^3p_2$	1.6
{1,2,1,1}	3	$12nq_1^2p_1^3q_2p_2$	1.6
{1,2,2}	1	$6nq_1^2p_1^2p_2^2$	1.8
{1,1,3}	2	$4nq_1^3p_1^4$	0.3
{1,3,1}	0	$4nq_1p_1^3p_3$	0.0
{1,4}	0	$np_1^4$	0.1

Tablica 9: Dobivene frekvencije s obzirom na očekivane i stvarne

Kada se sve dobivene frekvencije zbroje, dobivamo 664.1 što je jako blizu 664 pa možemo naslutiti da je ovaj model dobar. Međutim, to moramo ispitati statističkim  $\chi^2$ -testom jer se taj test koristi za usporedbu očekivanih i stvarnih frekvencija. Provodimo ga u Excel programu te dobivamo  $p = 0.077261$  što je veće od 0.05, ne odbacujemo hipotezu o jednakosti očekivanih i stvarnih frekvencija pa na istoj razini značajnosti možemo zaključiti da model dobro opisuje podatke.

Pomoću generaliziranog modela možemo dobiti procjenitelje za Reed-Frostov i Greenwoodov model. Uz zamjenu  $q_i = q$ , funkcija vjerodostojnosti za Greenwoodov model glasi:

$$\mathbb{L}(q) = Cq^{3000}(1 - q)^{397}q^{70}(1 - q)^{18}q^3 = Cq^{3073}(1 - q)^{415},$$

pomoću koje dobivamo  $x = 3000$ ,  $s = 415 + 3073 = 3488$  pa je  $\hat{q} = 0.88102$ . S druge strane, zamjenom  $q_i = q^i$  dobivamo funkciju vjerodostojnosti za Reed-Frostov model:

$$\mathbb{L}(q) = Cq^{3000}(1 - q)^{397}(q^2)^{70}(1 - q^2)^{18}(q^3)^3 = Cq^{3149}(1 - q)^{415}(1 + q)^{18},$$

s kojom, analognim postupkom, dobivamo  $\hat{q} = 0.8838$ . Detaljan postupak može se naći u [4].

$\{i_0, i_1, \dots, i_T\}$	Stvarna	Dobivene (Reed-Frost)	Dobivene (Greenwood)
{1}	423	405.2	400.1
{1,1}	131	147.1	147.8
{1,1,1}	36	45.3	46.5
{1,2}	24	25.6	34.0
{1,1,1,1}	14	10.5	11.1
{1,1,2}	8	6.0	7.1
{1,2,1}	11	12.7	8.1
{1,3}	3	2.5	3.5
{1,1,1,1,1}	4	1.4	1.5
{1,1,1,2}	2	0.8	0.9
{1,1,2,1}	2	1.7	1.0
{1,2,1,1}	3	1.7	1.1
{1,2,2}	1	2.0	0.6
{1,1,3}	2	0.3	0.4
{1,3,1}	0	1.1	0.3
{1,4}	0	0.1	0.1

Tablica 10: Dobivene frekvencije pomoću očekivanih u oba modela

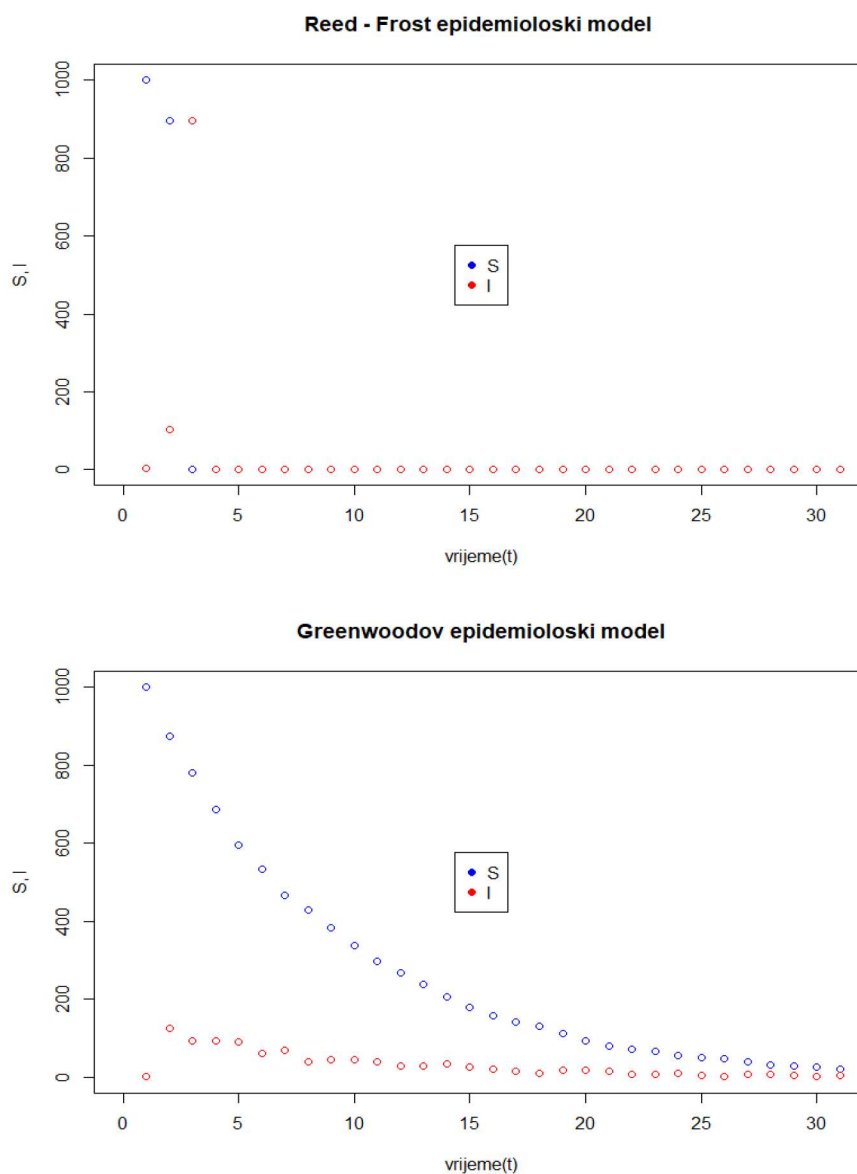
Dobivene frekvencije realizacija lanaca dane u Tablici 10. izračunate su pomoću vjerojatnosti u Primjeru 3 i Primjeru 4 za  $n = 664$ . Njihov zbroj u Reed-Frostovom modelu iznosi 664.0 što je točno  $n$ , no vrijednosti se razlikuju za određene lance pa treba ispitati  $\chi^2$ -testom. Na razini značajnosti od 0.05 možemo zaključiti da je Reed-Frostov model dobar tj. dobro opisuje podatke jer dobivamo  $p = 0.05662$ . S druge strane, ukupan zbroj u Greenwoodovom modelu je 664.1 što bi također sugeriralo na dobar model jer isti takav zbroj imamo i za generalizirani model. No, moramo ispitati statističkim  $\chi^2$ -testom. Dobiva se  $p = 0.016408$  što je manje od 0.05 pa na istoj razini značajnosti zaključujemo da model nije dobar, odnosno u njemu se očekivane frekvencije statistički značajno razlikuju od stvarnih.

Dakle, Heisman i Reid pokazuju kako generalizirani i Reed-Frostov model dobro opisuju podatke dok se u Greenwoodov ne može pouzdati.

## 5.1 Simulacije modela

Simulacije provodimo u R programu čiji je prikaz dan na sljedećim slikama. Neka je  $S_0 = 1000$ ,  $I_0 = 1$  te vjerojatnosti  $q$  koje su jednake procjenama u prethodnom poglavlju.

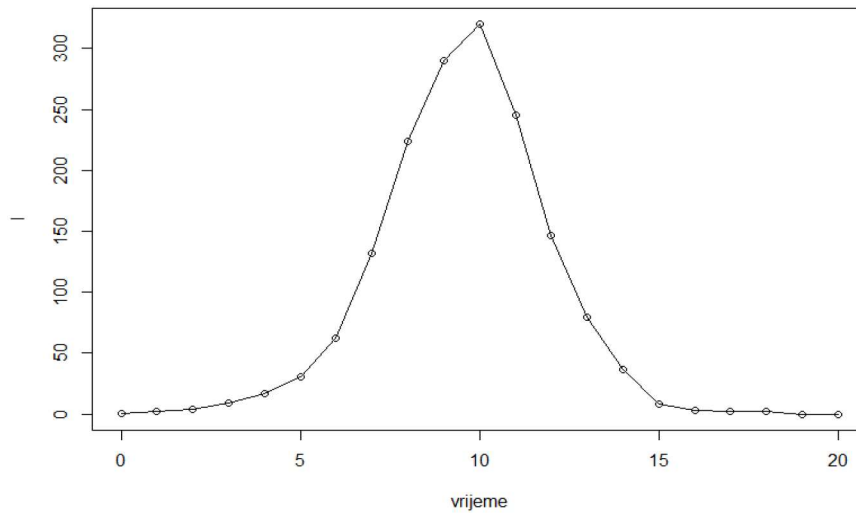
Primjetimo kako u Reed-Frostovom modelu brže ponestane i podložnih i zaraženih osoba nego u Greenwoodovom. Odnosno, epidemija u njemu prije završi.



Slika 3: Simulacija oba modela za veliku populaciju

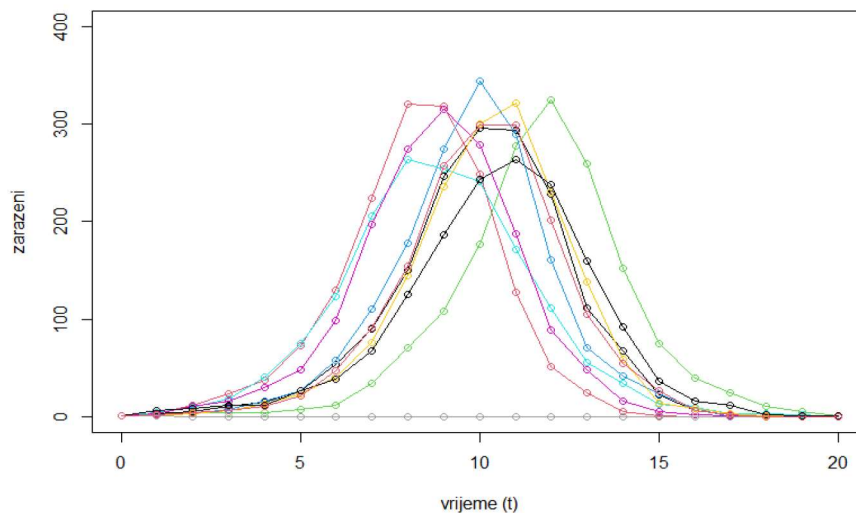
Najpoznatija vjerojatnost koja se uzima za model binomnih epidemioloških lanaca je  $p(t) = 1 - \exp(-\beta I_t)$ . Kada bolje razmislimo, to je i logično jer za epidemije generalno vrijedi "eksponencijalni razvoj". Primjetimo kako se onda i vjerojatnost

u svakom trenutku  $t$  mijenja. Uzet ćemo proizvoljnu vrijednost konstante  $\beta = 0.001$  te za početak napraviti samo jednu simulaciju u dvadeset jedinica vremena. Ovaj put uzimamo  $S_0 = 2000$  i  $I_0 = 1$ .



Slika 4: Jedna simulacija modela binomnih epidemioških lanaca za  $\beta = 0.001$

Kada bismo napravili deset simulacija dobili bismo sljedeći prikaz na kojem se vidi *stohastičnost* modela.

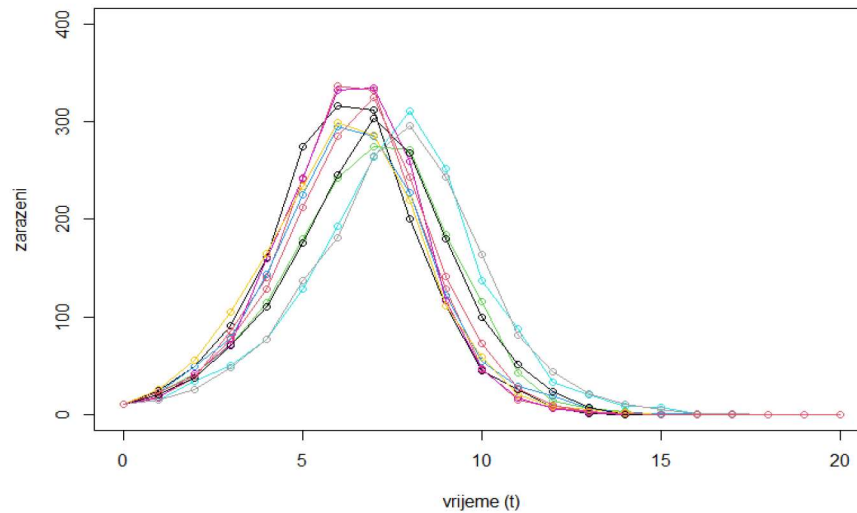


Slika 5: Deset simulacija modela binomnih epidemioških lanaca za  $\beta = 0.001$

Do sada smo uzimali  $I_0 = 1$ , no to ne mora uvijek biti slučaj. Radi usporedbe,



napraviti ćemo deset simulacija u dvadeset jedinica vremena kada je  $I_0 = 10$  dok ostale vrijednosti ostaju iste.



Slika 6: Deset simulacija modela binomnih epidemioških lanaca s  $I_0 = 2$

Uspoređujući Sliku 5. i Sliku 6. može se uočiti kako epidemija prije završava kada imamo veći broj zaraženih u početnom stanju. Također, ima brži rast.

U modelu binomnih epidemioloških lanaca generalno vrijedi da *trajektorije* procesa ili izumiru ili eksplodiraju na eksponencijalni način.

Treba imati na umu da je model binomnih lanaca zapravo gruba aproksimacija stvarnog razvoja epidemije. Jedna od bitnijih pretpostavki je homogenost populacije u kojem se ona savršeno "miješa", odnosno svaka se podložna osoba može zaraziti s istom vjerojatnošću. To za velike populacije poput gradova, država i sl. predstavlja problem. Upravo zato, razvijaju se razni modeli (SIR, SIS, SEIR, ...) koji uključuju rađanje i izumiranje u populaciji. Iako jednostavniji, modeli binomnih epidemioških lanaca korisni su za male grupe ljudi.

## 6 Literatura

- [1] H. ABBEY, *Examination of the Reed-Frost theory of epidemics*, *Human Biology*, vol. 24, no. 3, 1952., <http://www.jstor.org/stable/41449060>
- [2] L. J. S. ALLEN, *An Introduction to Stochastic Epidemic Models In: Brauer, F., van der Driessche, P. and Wu, J., Eds., Mathematical Epidemiology*, Springer, Berlin, 2008.
- [3] L. J. S. ALLEN, *An Introduction to Stochastic Processes with Applications to Biology*, Texas Tech University, Lubbock Texas, Taylor & Francis Group, LLC, 2010.
- [4] N. BECKER, *A General Chain Binomial Model for Infectious Diseases*, *Biometrics*, vol. 37, no. 2, 1981., <https://doi.org/10.2307/2530415>
- [5] M. BENŠIĆ, N. ŠUVAK, *Uvod u vjerojatnost i statistiku*, Sveučilište J.J. Strossmayera u Osijeku, Odjel za matematiku, 2014.
- [6] L. H. CAIROLI, *Chain binomial epidemic models*, Kansas state university, 1988. <https://krex.k-state.edu/dspace/bitstream/handle/2097/23889/LD2668R4STAT1988C35.pdf?sequence=1>
- [7] B. A. DAVIS, *A chain binomial model with immigration*, The Australian National University, 1996.
- [8] S. R. DUNBAR, in *Probability Theory and Stochastic Processes*, University of Nebraska-Lincoln, Lincoln, 2020.
- [9] M.E. HALLORAN, I. M. LONGINI, JR., C. J. STRUCHINER, *Design and Analysis of Vaccine Studies*, Seattle, Rio de Janeiro, 2009.
- [10] M. A. HEASMAN, D. D. REID, *Theory and Observation in family epidemics of the common cold*, Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, 1961., <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1058809/pdf/brjprevsmed00061-0016.pdf>
- [11] J. GANI, D. JERWOOD, *Markov Chain Methods in Chain Binomial Epidemic Models*, vol. 27, no. 3, *Biometrics*, 1971., <https://doi.org/10.2307/2528598>.
- [12] I. M. LONGINI, JR., *A chain binomial model of endemicity*, Cali, Colombia, 1979.

- [13] N. ŠUVAK, *Slučajni procesi I (radni materijali)*, Odjel za matematiku, Sveučilište J.J. Strossmayera u Osijeku, 2020.
- [14] Z. VONDRAČEK, *Markovljevi lanci (web materijali)*, PMF Matematički odsjek, Sveučilište u Zagrebu, 2013.
- [15] Z. VONDRAČEK, *Slučajni procesi (web materijali)*, PMF Matematički odsjek, Sveučilište u Zagrebu, 2018.
- [16] H. J. WEARING, J. M. DRAKE, A. A. KING, *Simulating epidemics in R*, 2012. [https://ms.mcmaster.ca/~bolker/eeid/ecology/EEID2012\\_Simulation.pdf](https://ms.mcmaster.ca/~bolker/eeid/ecology/EEID2012_Simulation.pdf)

## Sažetak

U ovom radu opisana je klasa epidemioloških stohastičkih modela u diskretnom vremenu koji se temelje na binomnoj distribuciji. Kako se radi o diskretnom skupu stanja, modele se može zapisati pomoću Markovljevih lanaca. Modeli se razlikuju u ovisnosti vjerojatnosti izbjegavanja zaraze s zaraženom osobom te o ponovnoj podložnosti osobe nakon zaraze. Model je dobar na malim populacijama dok za velike nije primjenjiv zbog pretpostavke o homogenosti populacije. Znanstvenici Heisman i Reid pokazali su funkcionalnost više klasa modela na stvarnim podacima.

**Ključne riječi:** epidemiološki model, diskretno vrijeme, stohastički model, binomna distribucija, Markovljev lanac, ovisnost, stvarni podaci

## Abstract

This paper describes one of stochastic epidemiological models in discrete time that are based on binomial distribution. Because of discrete state space we can model with Markov chains. The difference between them is based on an assumption whether the probability of escaping an infection depends on the infected person. The model is valid if used only in a small circle of people due to the assumption of homogeneity. Scientists Heisman and Reid made a big contribution when they demonstrated the functionality of different models on real data.

**Keywords:** epidemic model, discrete time, stochastic model, binomial distribution, Markov Chain, dependency, real data

## 7 Životopis

Moje ime je Ivana Blažević, rođena sam 21.12.1996. godine u Novoj Gradišci. Osnovnu školu pohađam od 2003. do 2011. godine u OŠ Marije Jurić Zagorke u Zagrebu nakon čega upisujem II. gimnaziju u Zagrebu. Završavam srednju školu 2015. godine te odmah upisujem preddiplomski studij na Prirodoslovno-matematičkom fakultetu; Odsjek za matematiku u Zagrebu. Godine 2019. stječem naziv prvostupnika edukacije matematike te iste godine upisujem diplomski studij na Odjelu za matematiku Sveučilišta J.J. Strossmayera u Osijeku, smjer Financijska matematika i statistika. Kao učenica osnovne škole redovno sam išla na matematička natjecanja.