

Stohastični populacijski modeli

9. domača naloga pri predmetu Modelska analiza I

Domen Vaupotič

V nalogi se ukvarjamo s stohastičnimi populacijskimi modeli, ki jih modeliramo s točkovnimi Poissonovimi procesi. Spomnimo se najprej diskretne enoparametrične Poissonove porazdelitve in vzemimo diskretno naključno spremenljivko $X \approx \mathcal{P}(\mu)$:

$$\mathbf{P}(X = x) = \frac{\mu^x e^{-\mu}}{x!}, \quad x \in \mathbb{N}_0,$$

pri čemer je μ parameter porazdelitve. Velja $\mathbf{E}[X] = \text{Var}[X] = \mu$. Za nabor naključnih spremenljivk, porazdeljenih po Poissonovi porazdelitvi, $X_1 \approx \mathcal{P}(\mu_1)$, $X_2 \approx \mathcal{P}(\mu_2)$, ..., $X_n \approx \mathcal{P}(\mu_n)$, velja $\mathbf{E}[X_1 + X_2 + \dots + X_n] = \mu_1 + \mu_2 + \dots + \mu_n$.

1. Eksponentni model populacije

Preprosti eksponentni model

V prvem delu naloge se bomo ukvarjali s preprostim eksponentnim modelom populacije. Zapišimo najprej deterministični zvezni model za eksponentno umiranje populacije:

$$\frac{dN(t)}{dt} = -\beta N(t),$$

pri čemer je $N(t)$ velikost populacije ob času t in β koeficient (hitrost) umiranja.

V resnici je velikost populacije diskretna spremenljivka, prav tako se smrti dogajajo ob dobro definiranih, a naključnih časovnih trenutkih. V ta namen želimo sestaviti *stohastični* model. Najprej diskretizirajmo v časovni spremenljivki, da dobimo diferenčno enačbo:

$$\frac{dN(t)}{dt} = -\beta N(t) \rightarrow N(t + \Delta t) = N(t) - \beta N(t) \Delta t$$

Težava nastopi, ko želimo diskretizirati še v številu osebkov. Število smrti v vsakem trenutku je naključna spremenljivka, za katero poznamo povprečno vrednost $\beta N(t) \Delta t$. Za žrebanje naključne spremenljivke pa moramo poznati njeno porazdelitev. Zamislimo si seveda lahko poljubno diskretno porazdelitev z dano povprečno vrednostjo in katero koli od njih bi lahko uporabili za simulacijo. Odločimo se, da uporabimo tisto, ki je skladna z mehanističnimi predpostavkami modela. Smrti v populaciji so med seboj neodvisni procesi s preprosto Bernoullijevjo porazdelitvijo. Če je število osebkov veliko in verjetnost za smrt majhna, lahko N -kratno Bernoullijevjo porazdelitev dobro aproksimiramo s Poissonovo porazdelitvijo:

$$N(t + \Delta t) = N(t) - \Delta N, \quad \Delta N \approx \mathcal{P}(\beta N(t) \Delta t).$$

Definirajmo najprej generator števil po Poissonovi porazdelitvi:

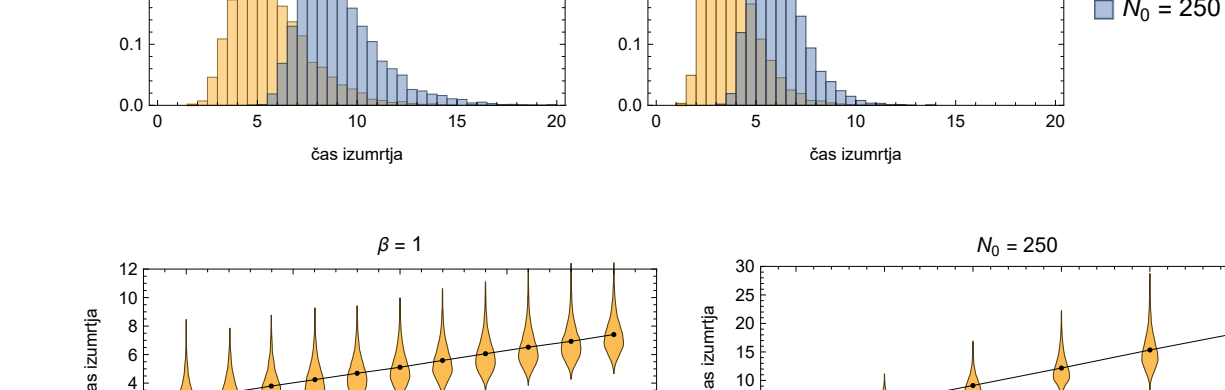
```
Poisson[μ_ /; μ_ > 0] := RandomVariate[PoissonDistribution[μ_]]
Poisson[μ_] := 0
```

Stohastično trajektorijo sedaj izračunamo diskretno od časa 0 in začetne populacije N_0 do zelenega končnega časa s s časovnimi koraki Δt . Pri tem smo pozorni, da ohranjamo populacijo negativno.

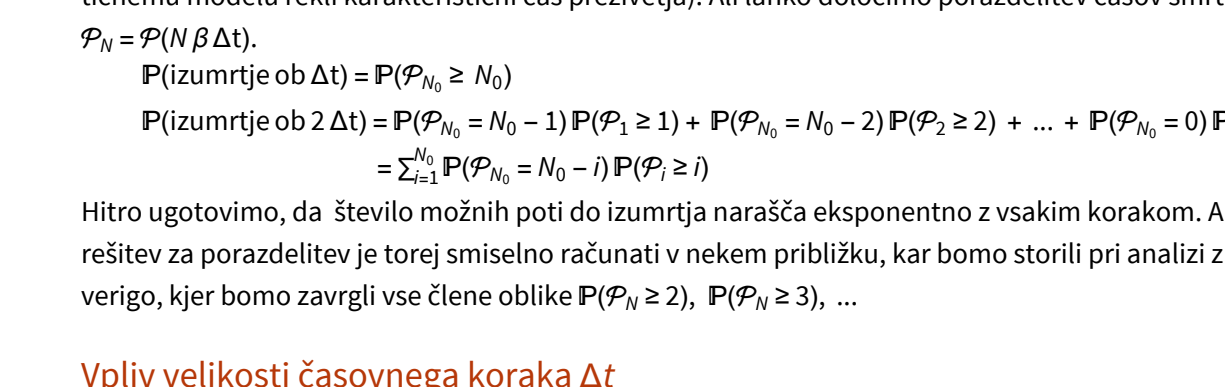
```
trSmrt[μ_ , Δt_ , t_ , dt_ ] := NestList[{
  Max[0, #2] - Poisson[β #2 Δt]
} &, {0, μ}, Floor[t / dt]]
```

Značilne trajektorije

Ogledimo si začetek nekaj tipičnih trajektorij (barvne črte) pri majhni začetni populaciji (25) in veliki začetni populaciji (250) ter deterministični model (črna črta) z isto hitrostjo umiranja.

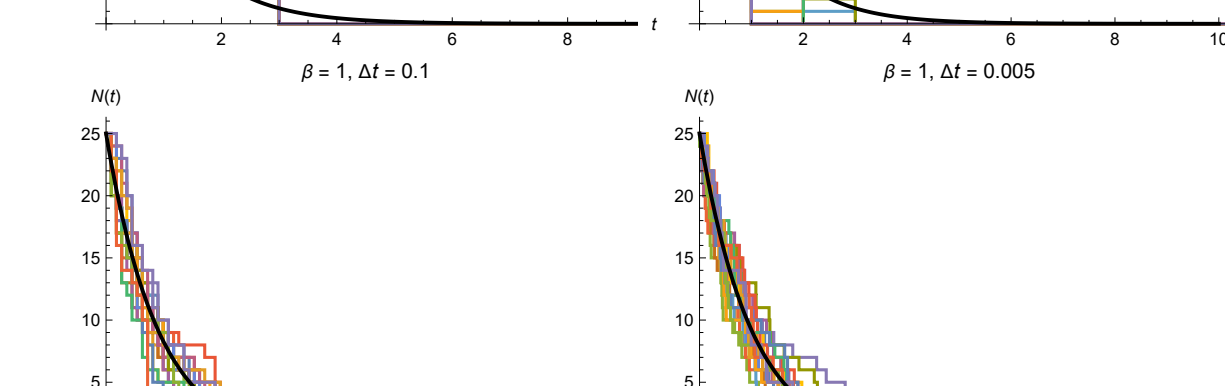


Opazimo, da imajo kvalitativno vse trajektorije podobno obliko in populacija se približno eksponentno zmanjšuje. Ali s povprečenjem različnih trajektorij dobimo isti rezultat kot pri determinističnem modelu? Z zadostno majhnim korakom in zadostno velikim številom povprečnih trajektorij M dobimo zelo dobro ujemanje.



Porazdelitev časa izumrtja

V determinističnem zveznem modelu se populacija eksponentno zmanjšuje, vendar izumre le v limiti $t \rightarrow \infty$, medtem ko v stohastičnem modelu populacija izumre v nekem končnem času. Ogledimo si, kako je porazdeljen čas izumrtja populacije, izračunan na 5000 stohastičnih trajektorijah.



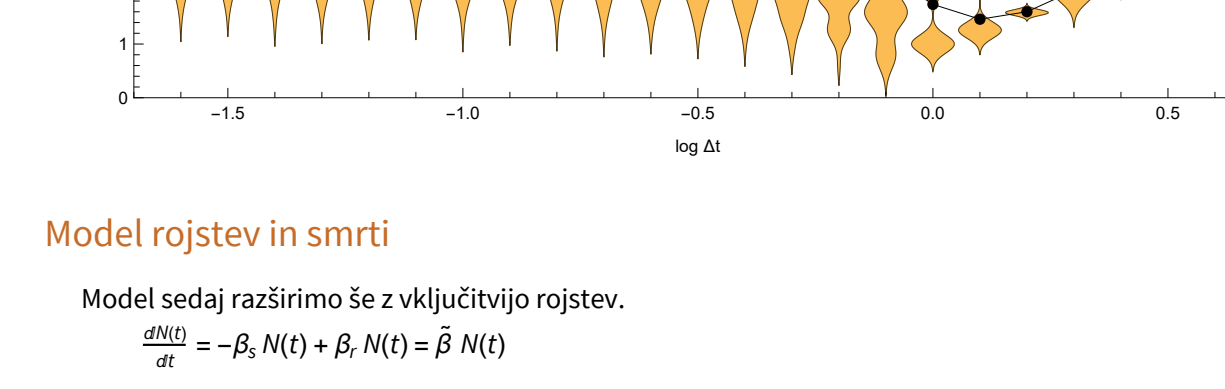
Ugotovimo, da z večanjem začetne populacije pri konstantnem β čas izumrtja eksponentno narašča (kar je povsem pričakovano v eksponentnem modelu) ter da je čas izumrtja sorazmeren z $1/\beta$ (čemer bi determinističnemu modelu rekli karakteristični čas preživetja). Ali lahko določimo porazdelitev časov smrti? Pišimo $\mathcal{P}_N = \mathcal{P}(N \Delta t)$.

$$\mathbf{P}(\text{izumrtje ob } \Delta t) = \mathbf{P}(\mathcal{P}_{N_0} = N_0 - 1) \mathbf{P}(\mathcal{P}_1 \geq 1) + \mathbf{P}(\mathcal{P}_{N_0} = N_0 - 2) \mathbf{P}(\mathcal{P}_2 \geq 2) + \dots + \mathbf{P}(\mathcal{P}_{N_0} = 0) \mathbf{P}(\mathcal{P}_{N_0} \geq N_0)$$
$$= \sum_{i=0}^{N_0} \mathbf{P}(\mathcal{P}_{N_0} = N_0 - i) \mathbf{P}(\mathcal{P}_i \geq i)$$

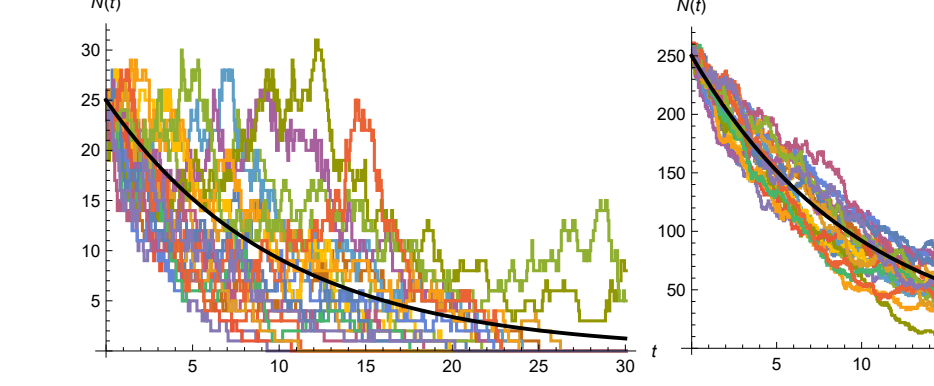
Hitro ugotovimo, da število možnih poti do izumrtja narašča eksponentno z vsakim korakom. Analitično rešitev za porazdelitev je torej smiselno računati v nekem približku, kar bomo storili pri analizi z Markovsko verigo, kjer bomo zavrgli vse člene oblike $\mathbf{P}(\mathcal{P}_i \geq 2)$, $\mathbf{P}(\mathcal{P}_i \geq 3)$, ...

Vpliv velikosti časovnega koraka Δt

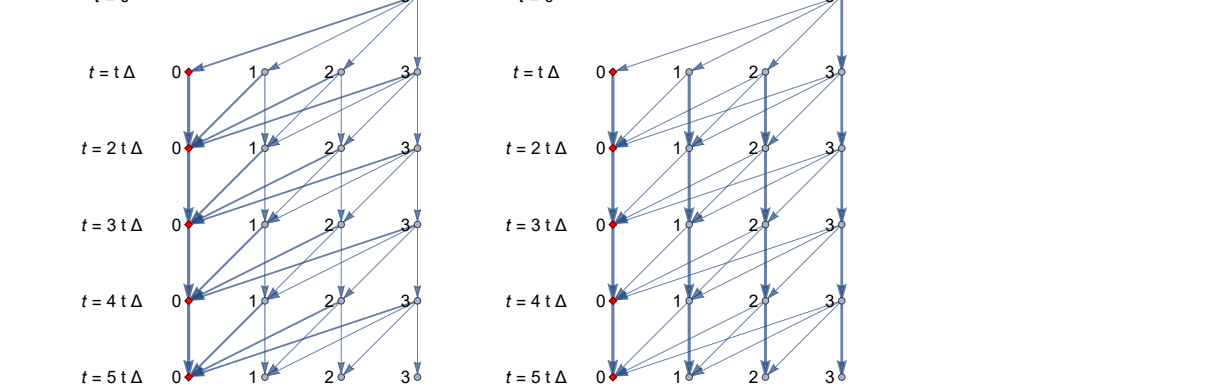
Ogledimo si, kako izbira časovnega intervala Δt vpliva na stohastično simulacijo.



Ko je Δt prevelik, bo že v prvem koraku izumrla vsa populacija, zato iz takšne trajektorije izvem le malo. Kadar je Δt premajhen, pa v večini korakov ne bo prišlo do spremembe, zato bomo po nepotrebnem tratili računski čas. Smiselno je torej izbrati kompromisni Δt . Spodnji graf prikazuje Poissonovo porazdelitev pri dveh različnih Δt .



Na spodnjem grafu vidimo, kako Δt vpliva na porazdelitev časa izumrtja. Opatimo, da se pri zadosti majhnem Δt izračunani čas izumrtja stabilizira. Znatno preveliki Δt povzročijo izumrtje v prvem časovnem koraku (ki postaja vedno daljši). Pri vmesnih vrednostih Δt opazimo, da populacija izumre prehitro.



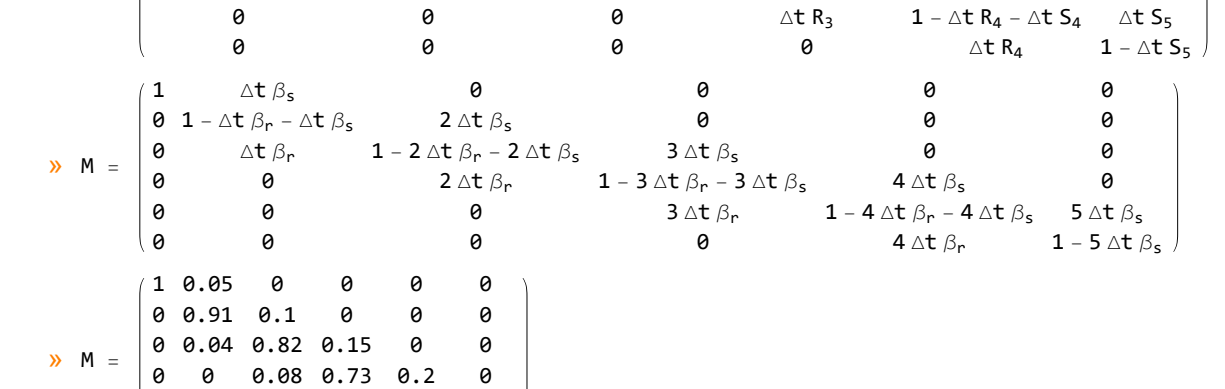
Model rojstev in smrti

Model sedaj razširimo še z vključitvijo rojstev.

$$\frac{dN(t)}{dt} = \beta_s N(t) - \beta_d N(t) = \tilde{\beta} N(t)$$

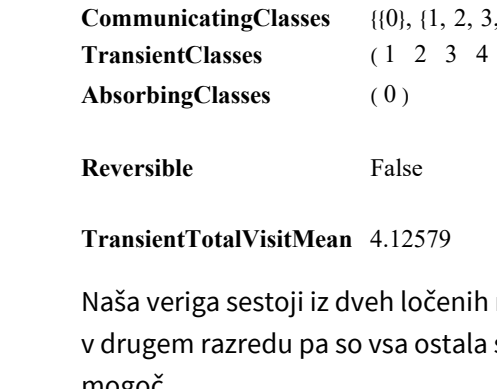
Pozorni smo na to, da je v determinističnem modelu prav zares zgolj en parameter, efektivna stopnja rasti populacije $\tilde{\beta} = \beta_s - \beta_d$. V stohastičnem modelu pa smrti in rojstva obravnavamo ločeno kot dva neodvisna stohastična procesa.

$$N(t + \Delta t) = N(t) - \Delta N_s + \Delta N_r, \quad \Delta N_s \approx \mathcal{P}(\beta_s N(t) \Delta t), \quad \Delta N_r \approx \mathcal{P}(\beta_r N(t) \Delta t).$$



2. Markovska veriga

Ogledimo si začetek vse možne poti trajektorij z začetno populacijo 3 osebkov in prvih 5 časovnih intervalov.



Če izberemo dovolj majhen časovni korak, potem bo večina prehodov takšnih, kjer bo populacija ostala enako velika ali pa se bo spremenila za 1. Tedaj lahko izračun verjetnosti nekoliko poenostavimo, tako da Poissonovo porazdelitev razvijemo do prvega reda v Δt .

$$\mathbf{P}(1 \text{ rojstvo in 1 smrt}) = R_N \Delta t + \mathcal{O}(\Delta t^2)$$
$$\mathbf{P}(1 \text{ rojstvo in 1 smrt}) = 1 - S_N \Delta t - R_N \Delta t + \mathcal{O}(\Delta t^2)$$
$$\mathbf{P}(\text{karkoli drugega}) = \mathcal{O}(\Delta t^2)$$

Pri tem smo uvedli parametra R_N in S_N , ki podajata hitrosti obeh procesov. V našem modelu imamo $R_N = \beta_r N$ in $S_N = \beta_d N$.

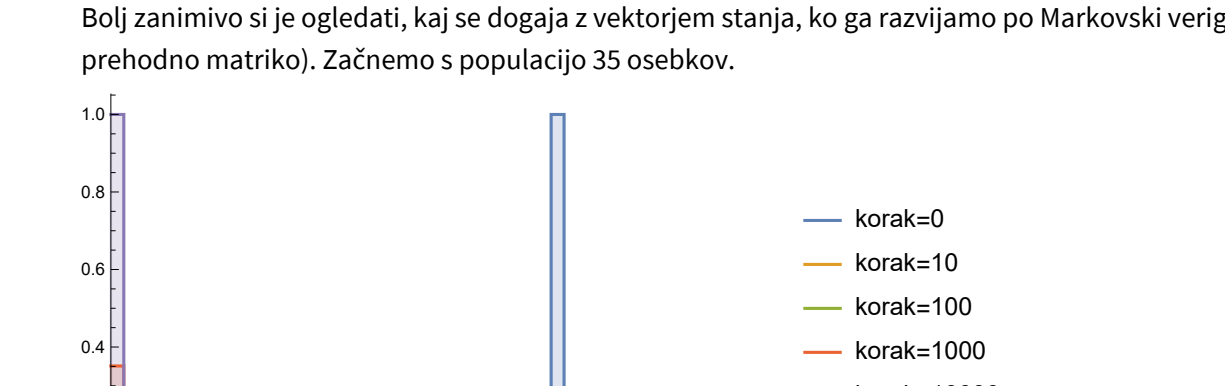
Proces bomo sedaj modelirali s končno Markovsko verigo, ki vsebuje vektor stanj x_i in matriko prehodov M_{ij} :

$$x_i(t) = \mathbf{P}(N(t) = i)$$
$$x_i(t + \Delta t) = M_{ij} x_j(t)$$

Ker bo veriga končna, bo obstajala največja velikost populacije. Ogledimo si, kako izgleda matrika prehodov za populacijo z največ 5 osebki s parametri $\Delta t = 0.1$, $\beta_s = 0.4$ in $\beta_d = 0.5$.

$$M = \begin{pmatrix} 1 - \Delta t R_0 - \Delta t S_0 & \Delta t S_1 & 0 & 0 & 0 \\ \Delta t R_0 & 1 - \Delta t R_1 - \Delta t S_1 & \Delta t S_2 & 0 & 0 \\ 0 & \Delta t R_1 & 1 - \Delta t R_2 - \Delta t S_2 & \Delta t S_3 & 0 \\ 0 & 0 & \Delta t R_2 & 1 - \Delta t R_3 - \Delta t S_3 & \Delta t S_4 \\ 0 & 0 & 0 & \Delta t R_3 & 1 - \Delta t R_4 - \Delta t S_4 \end{pmatrix}$$
$$M = \begin{pmatrix} 1 & \Delta t \beta_s & 0 & 0 & 0 \\ \Delta t \beta_d & 1 - \Delta t \beta_s - \Delta t \beta_d & \Delta t \beta_s & 0 & 0 \\ 0 & \Delta t \beta_d & 1 - \Delta t \beta_d - 2 \Delta t \beta_s & \Delta t \beta_s & 0 \\ 0 & 0 & \Delta t \beta_d & 1 - 3 \Delta t \beta_s - 3 \Delta t \beta_d & \Delta t \beta_s \\ 0 & 0 & 0 & \Delta t \beta_d & 1 - 4 \Delta t \beta_s - 4 \Delta t \beta_d \end{pmatrix}$$
$$M = \begin{pmatrix} 1 & 0.05 & 0 & 0 & 0 \\ 0 & 0.91 & 0.1 & 0 & 0 \\ 0 & 0.04 & 0.82 & 0.15 & 0 \\ 0 & 0 & 0.08 & 0.73 & 0.2 \\ 0 & 0 & 0 & 0.12 & 0.64 \end{pmatrix}$$

Markovsko verigo lahko narišemo tudi z grafom, v katerem vozlišča predstavljajo posamezno stanje (število osebkov), robovi pa prehode med stanji, utežene s pripadajočo verjetnostjo prehoda. Začetna velikost populacije naj bodo 3 osebeki.



Ogledimo si nekaj osnovnih lastnosti te Markovske verige.

Communicating Classes $\{(0), (1, 2, 3, 4, 5)\}$ Sets of states such that all states within a class are accessible from each other

Transient Classes $\{(1, 2, 3, 4, 5)\}$ Classes of states that are visited finitely many times with probability one

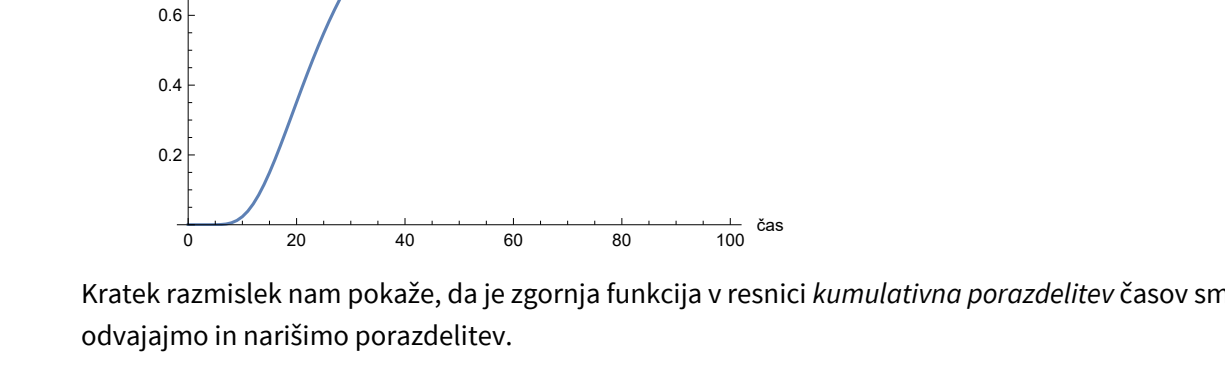
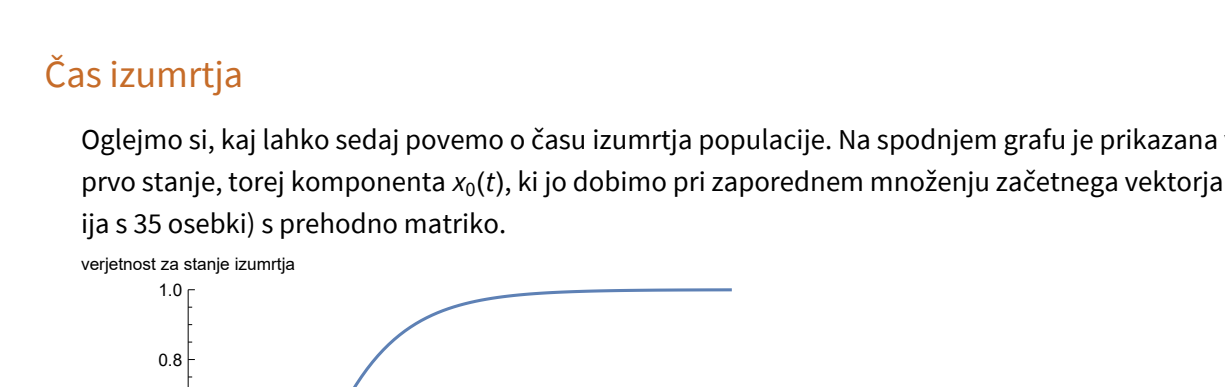
Absorbing Classes $\{(0)\}$ The classes in the chain with a single state from which it is impossible to transition to another state

Reversible False Whether or not the chain is reversible, that is, the product of transition rates over any closed loop of states is the same

Transient Total Visit Mean 4.12579 The expected number of states entered before being absorbed

Naša veriga sestoji iz dveh ločenih razredov (classes). V prvem razredu je stanje izumrle populacije (0 osebkov), v drugem razredu pa so vsa ostala stanja. Stanje izumrtja deluje tudi kot absorber, saj prehod iz njega ni več mogoč.

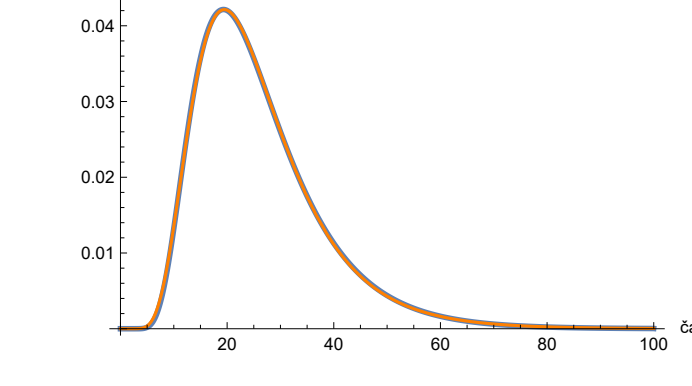
Povečajmo sedaj možno velikost populacije na 50. Pri tem moramo biti nekoliko pozorni: če namreč izberemo prevelik Δt in preveliko možno velikost populacije, se v matriki prehodov pojavijo negativne vrednosti, kar seveda ni dovoljeno. Izbrati moramo torej zadosti majhen Δt . Začetno stanje bo populacija z 20 osebki.



Opazimo, da bodo stanja s številom osebkov manj kot 20 gotovo dosežena, medtem ko verjetnost za porazdelitev večje populacije približno eksponentno pada. Opazimo tudi dve zanimivi stanji: izumrla populacija (stanje 0) ima povprečni čas zadrževanja ∞ , največji možna populacija (stanje 50) pa ima izrazito daljši čas zadrževanja kot populacija 49, kar je robni učinek, ki je nastal zaradi največjega dovoljenega števila populacije.

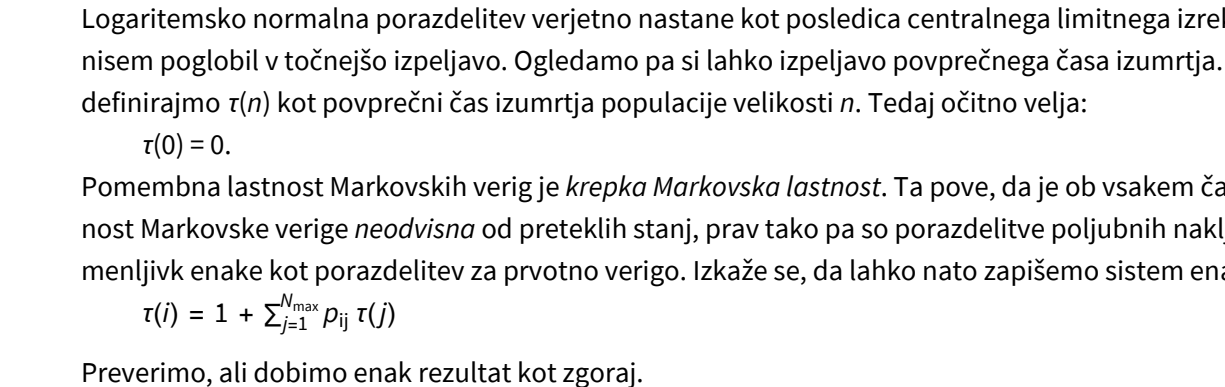
Tipične trajektorije

S podano Markovsko verigo lahko izvajamo simulacije na podoben način kot v prvem delu naloge. Začnemo v začetnem stanju populacije, nato pa prehajamo med stanji z ustreznimi verjetnostmi.

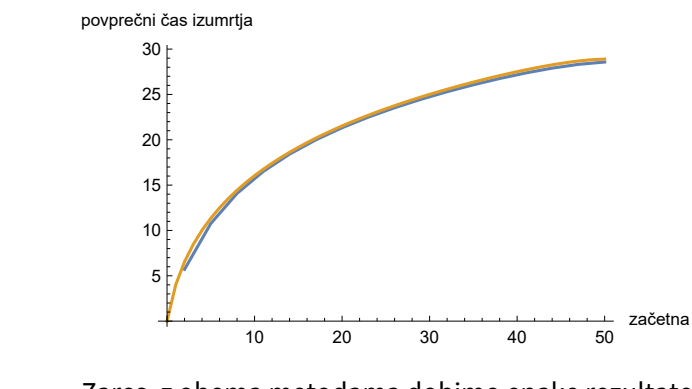


Letenje momentov porazdelitve

Bolj zanimivo si je ogledati, kaj se dogaja z vektorjem stanja, ko ga razvijamo po Markovski verigi (množico s prehodno matriko). Začnemo s populacijo 35 osebkov.



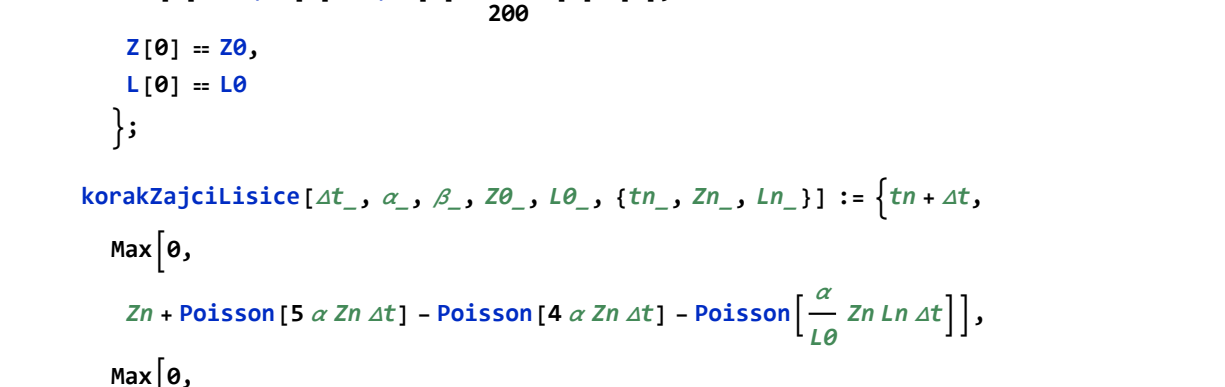
Porazdelitev najprej "sploščuje", povprečje se začne premikati v levo, nato pa se ustali pri umrli populaciji. Obnašanje še bolje vidimo v spodnjem prikazu:



Kakšna je pravzaprav razporeditev? Izkaže se, da je *logaritemska normalna* (Galtonova) porazdelitev:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

pri čemer je povprečni čas take porazdelitve enak $e^{\mu + \frac{\sigma^2}{2}}$.



Ogledimo si, kako je povprečni čas izumrtja odvisen od začetne populacije.



Logaritemska normalna porazdelitev verjetno nastane kot posledica centralnega limitnega izreka, vendar se nismo poglubi točneje izpeljali. Ogledamo pa si lahko izpeljavo povprečnega časa izumrtja. V ta namen definiramo $\langle n(t) \rangle$ kot povprečni čas izumrtja populacije velikosti n . Tedaj očitno velja:

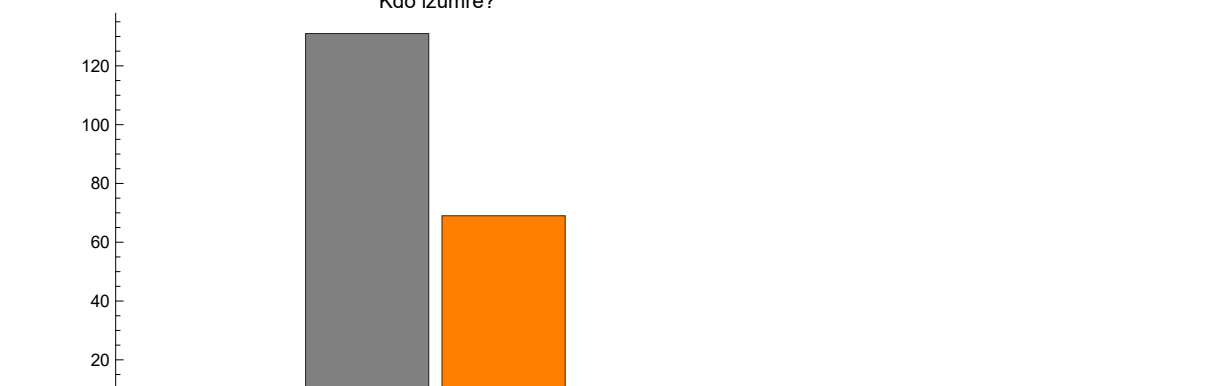
$$\tau(0) = 0.$$

Pomembna lastnost Markovskih verig je *rekurzivna lastnost*. Ta pove, da je ob vsakem času vsa prihodnost Markovske verige neodvisna od preteklih stanj, prav tako pa so porazdelitve poljubnih naključnih spremenljivk enake kot porazdelitve za prvotno verigo. Izkaže se, da lahko nato zapišemo sistem enačb za τ :

$$\tau(i) = 1 + \sum_{j=0}^{i-1} p_{ij} \tau(j)$$

Preverimo, ali dobimo enak rezultat kot zgoraj.

$$\tau(0) = 0$$
$$\tau(1) = 0.01 \tau(0) + 0.982 \tau(2) + 0.008 \tau(3) + 1$$
$$\tau(2) = 0.02 \tau(1) + 0.964 \tau(3) + 0.016 \tau(4) + 1$$
$$\dots$$
$$\tau(50) = 0.49 \tau(49) + 0.118 \tau(50) + 0.392 \tau(51) + 1$$
$$\tau(51) = 0.5 \tau(50) + 0.5 \tau(51) + 1$$



Zares, z obema metodama dobimo enake rezultate. Analitična metoda je seveda znatno hitrejša.

3. Stohastični model zajci-lišice

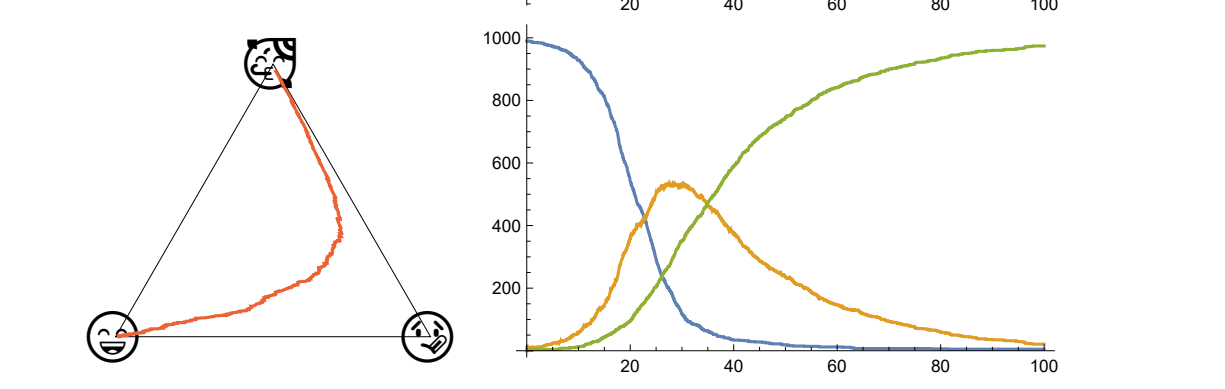
Drugi model, ki ga bomo obravnavali, je klasični model Lotka-Volterra, s katerim opisujemo preprosto dinamiko med plenom in plenikom. Zvezni model zapišemo s sistemom diferencialnih enačb, zaradi česar nam parametrov mogoče omejijo, tako da bo razmerje rodnosti in smrtnosti za zajce 5:4 in za lišice 4:5 ter stacionarno stanje 200 zajcev in 50 lišic.

$$\mathbf{eqn} = \left\{ \begin{aligned} \mathbf{Z}'[t] &= 5 \alpha \mathbf{Z}[t] - 4 \alpha \mathbf{Z}[t] - \frac{\alpha}{50} \mathbf{Z}[t] \mathbf{L}[t], \\ \mathbf{L}'[t] &= 4 \beta \mathbf{L}[t] - 5 \beta \mathbf{L}[t] + \frac{\beta}{200} \mathbf{Z}[t] \mathbf{L}[t], \\ \mathbf{Z}[0] &= 200, \\ \mathbf{L}[0] &= 50 \end{aligned} \right\}$$

$$\mathbf{korakZajcilisce}[\Delta t, \alpha, \beta, \mathbf{Z0}, \mathbf{L0}, \mathbf{tn}, \mathbf{Zn}, \mathbf{Ln}] := \left\{ \begin{aligned} \mathbf{Max}[\mathbf{0}, \\ \mathbf{Zn} + \mathbf{Poisson}[5 \alpha \mathbf{Zn} \Delta t] - \mathbf{Poisson}[4 \alpha \mathbf{Zn} \Delta t] - \mathbf{Poisson}[\frac{\alpha}{50} \mathbf{Zn} \mathbf{Ln} \Delta t], \\ \mathbf{Max}[\mathbf{0}, \\ \mathbf{Ln} + \mathbf{Poisson}[4 \beta \mathbf{Ln} \Delta t] - \mathbf{Poisson}[5 \beta \mathbf{Ln} \Delta t] + \mathbf{Poisson}[\frac{\beta}{200} \mathbf{Zn} \mathbf{Ln} \Delta t] \end{aligned} \right\}$$

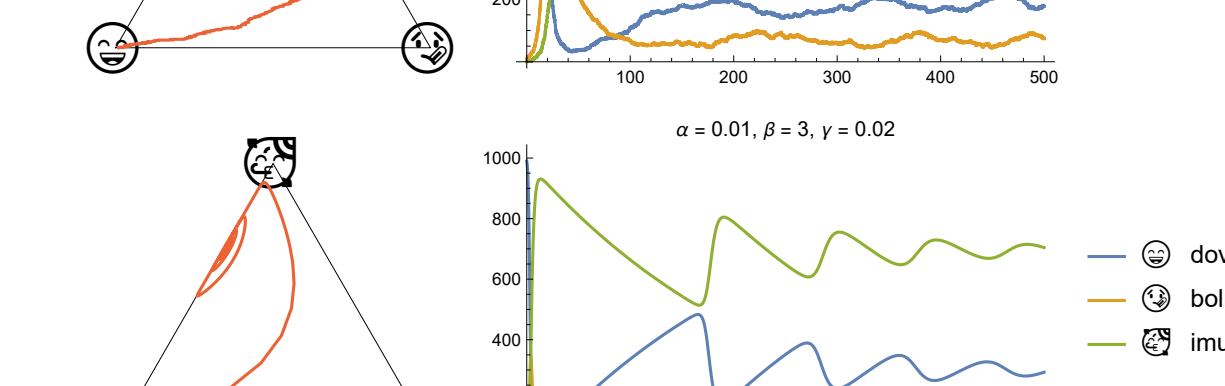
$$\mathbf{trZajcilisce}[\alpha, \beta, \mathbf{Z0}, \mathbf{L0}, \mathbf{tn}, \mathbf{dt}] := \mathbf{NestList}[\mathbf{korakZajcilisce}[\alpha, \beta, \mathbf{Z0}, \mathbf{L0}, \#] \&, \{\mathbf{0}, \mathbf{Z0}, \mathbf{L0}\}, \mathbf{Floor}[\mathbf{t} / \mathbf{dt}]]$$
$$\mathbf{trZajcilisceZmrtje}[\alpha, \beta, \mathbf{Z0}, \mathbf{L0}, \mathbf{tn}, \mathbf{dt}] := \mathbf{NestWhile}[\mathbf{korakZajcilisce}[\alpha, \beta, \mathbf{Z0}, \mathbf{L0}, \#] \&, \{\mathbf{0}, \mathbf{Z0}, \mathbf{L0}\}, \mathbf{\&}[2] > 0 \& \mathbf{\&}[3] > 0 \&]$$

Ogledimo si tipične stohastične trajektorije v faznem prostoru in jih primerjajmo z rešitvami zveznega modela.



V zveznem modelu dobimo stabilne periodične orbite, vendar nobena izmed populacij ne more spontano izumreti. To pa se lahko zgodi v stohastičnem modelu. Ogledimo si porazdelitev časa izumrtja ene izmed vrst.

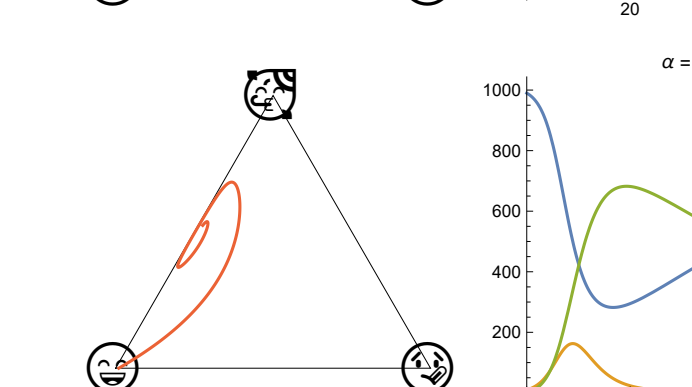
$$\mathbf{dt} = 1 / 20;$$
$$\mathbf{izumrtjeZL} := \mathbf{ParallelTable}[\mathbf{trZajcilisceZmrtje}[\mathbf{1}, \mathbf{1}, \mathbf{200}, \mathbf{50}, \mathbf{dt}], \mathbf{5000}];$$
$$\mathbf{Histogram}[\mathbf{izumrtjeZL}[\mathbf{All}, \mathbf{1}], \mathbf{20}, \text{"PDF"}];$$
$$\mathbf{Frame} = \mathbf{True}, \mathbf{FrameLabel} = \{\text{"Čas izumrtja"}, ""\}, \mathbf{PlotRange} = \mathbf{All}];$$



Echo[Row[{ "povp. življ. doba = ", N[Mean izumrtjeZL[All, 1], 4]}]]];

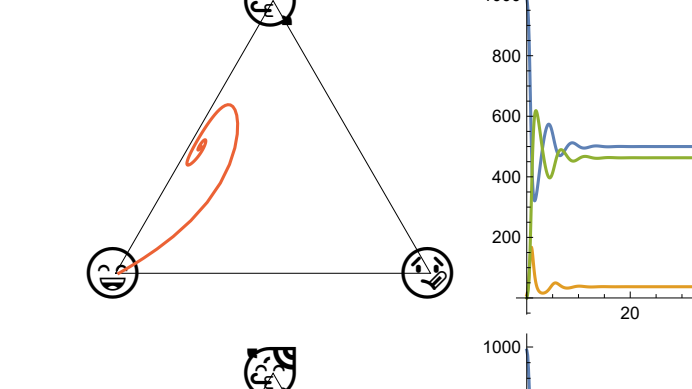
povp. življ. doba = 12.20

Zanimivo je tudi, da dvakrat pogosteje izumrejo zajci.



4. Stohastični model epidemije

V zadnjem delu naloge se ukvarjamo s triparametričnim modelom epidemije ZBIZ (*SIRIS*), kot smo za spoznali pri četrti domači nalogi.



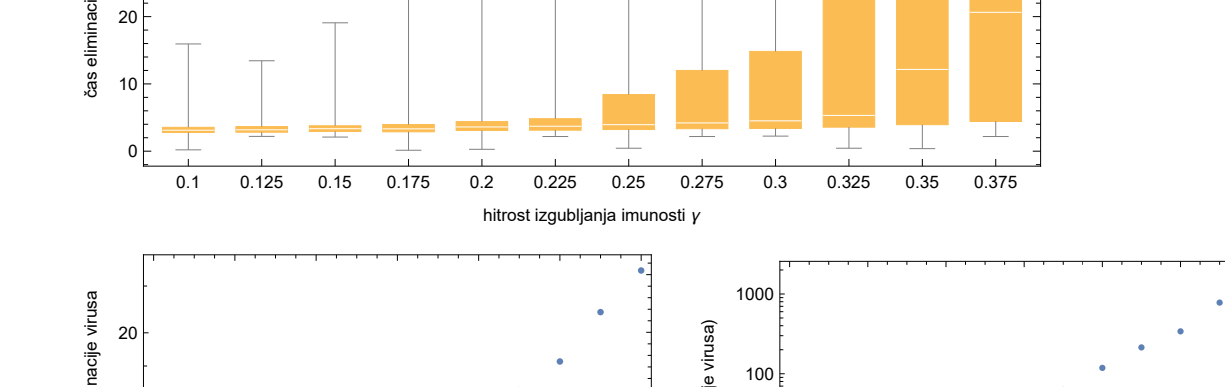
Populacijo razdelimo v tri razdelke: *zdravi*, *bolni* in *imuni*. Prehode med posameznimi razdelki podajajo trije parametri: α , β in γ .

$$\mathbf{eqnZBIZ} = \left\{ \begin{aligned} \mathbf{Z}'[t] &= -\alpha \mathbf{Z}[t] \mathbf{B}[t] + \gamma \mathbf{I}[t], \\ \mathbf{B}'[t] &= \alpha \mathbf{Z}[t] \mathbf{B}[t] - \beta \mathbf{B}[t], \\ \mathbf{I}'[t] &= \beta \mathbf{B}[t] - \gamma \mathbf{I}[t], \\ \mathbf{Z}[0] &= \mathbf{Z0}, \\ \mathbf{B}[0] &= \mathbf{B0}, \\ \mathbf{I}[0] &= \mathbf{I0} \end{aligned} \right\}$$

Ogledimo si še stohastični model. Pri tem smo pozorni, da vsakega izmed členov, ki se sicer pojavi dvakrat v naboru diferencialnih enačb, naključno žrebamo le enkrat.

$$\mathbf{korakZBIZ}[\Delta t, \alpha, \beta, \gamma, \{tn_, Zn_, Bn_, In_, Jn_.. \}] := \mathbf{Block}[\{ \mathbf{p1} = \mathbf{Poisson}[\alpha \mathbf{Zn} \mathbf{Bn} \Delta t], \mathbf{p2} = \mathbf{Poisson}[\beta \mathbf{Bn} \Delta t], \mathbf{p3} = \mathbf{Poisson}[\gamma \mathbf{In} \Delta t] \}, \{ \mathbf{tn} += \Delta t, \mathbf{Max}[\mathbf{0}, \mathbf{Zn} - \mathbf{p1} + \mathbf{p2}], \mathbf{Max}[\mathbf{0}, \mathbf{Bn} + \mathbf{p1} - \mathbf{p3}], \mathbf{Max}[\mathbf{0}, \mathbf{In} + \mathbf{p3} - \mathbf{p2}] \}];$$

Ogledimo si nekaj tipičnih trajektorij, prikazanih na trojnem faznem diagramu in kot časovni potek. Za vsak nabor parametrov so prikazane numerične rešitve diferencialnih enačb in stohastične simulacije. V vseh nadaljnjih primerih bo začetna populacija vsebovala 99 % zdravih, 1 % bolnih osebkov in nič imunih.



Opazimo, da dobimo s stohastičnim modeliranjem v splošnem primerljive rezultate. Razlika se pojavi le tedaj, ko število bolnih pada na 0 in virus iz populacije izginje. To se čisto zgodi v stohastičnem modelu, zaradi česar ne moremo dobiti situacije s kratkotrajnim dolgotrajnim izbruhom, kot se lahko zgodi pri determinističnem zveznem modelu.

Izračunajmo statistiko časov eliminacije virusa. Smiselno je omejiti območje parametrov ter izbrati take, pri katerih lahko do eliminacije v doglestem času sploh pride.

Vpliv velikosti populacije na čas eliminacije

Zgornji graf je razmeroma zanimiv, saj se v srednje velikih populacijah bolezen združuje v povprečju dlje kot v majhnih in velikih.

Vpliv trajanja imunosti na čas eliminacije

