

# Hawkesovi procesi

---

**Pich-Aguilera, Luis**

**Master's thesis / Diplomski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:306772>

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

*Download date / Datum preuzimanja:* **2024-09-18**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)



**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Luis M. Pich-Aguilera Barroso

**HAWKESOVI PROCESI**

Diplomski rad

Voditelj rada:  
doc. dr. sc. Hrvoje Planinić

Zagreb, srpanj, 2024.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

*Onome bez koga ničega ne bi ni bilo, a onda i svima onima koji su me poticali da  
podignem jedra. Zbogom!  
(Igoru također na pisanju zahvalnice).*

# Sadržaj

Sadržaj	iv
Uvod	1
<b>I Osnovna teorija vremenskih točkovnih procesa</b>	<b>2</b>
<b>1 Točkovni procesi</b>	<b>3</b>
1.1 Točkovni i brojeći procesi . . . . .	3
1.2 Svojstva brojećih procesa . . . . .	6
<b>2 Poissonovi procesi</b>	<b>8</b>
2.1 Homogeni Poissonovi procesi . . . . .	9
2.2 Intenzitet $\lambda$ . . . . .	13
2.3 Nehomogeni Poissonovi procesi . . . . .	16
<b>II Hawkesovi procesi</b>	<b>18</b>
<b>3 Uvjetna funkcija intenziteta</b>	<b>19</b>
<b>4 Osnove Hawkesovih procesa</b>	<b>23</b>
4.1 Struktura grananja . . . . .	25
<b>5 Algoritmi simuliranja</b>	<b>30</b>
5.1 Homogeni Poissonov proces . . . . .	30
5.2 Nehomogeni Poissonov proces i <i>thinning</i> algoritam . . . . .	32
5.3 Hawkesovi procesi . . . . .	36

<b>III Statističko modeliranje Hawkesovim procesima</b>	<b>39</b>
<b>6 Prilagodba parametara Hawkesovih procesa</b>	<b>40</b>
<b>7 Primjer modeliranja Hawkesovim procesima na društvenim mrežama</b>	<b>42</b>
7.1 Označeni Hawkesov proces za širenje informacije . . . . .	43
7.2 Procjena parametara . . . . .	44
7.3 Rezultati i zaključci . . . . .	47
<b>IV Dodatci</b>	<b>48</b>
<b>A Neki matematički rezultati</b>	<b>49</b>
<b>B R kodovi u radu</b>	<b>51</b>
<b>Bibliografija</b>	<b>60</b>

# Uvod

Vremenski točkovni procesi predstavljaju temeljni alat u analizi slučajnih događaja koji se javljaju kroz vrijeme. Ovaj diplomski rad se fokusira na teoriju i primjenu točkovnih procesa, s posebnim naglaskom na Poissonove i Hawkesove procese. Poissonovi procesi, kao jedan od najosnovnijih i najvažnijih modela u teoriji točkovnih procesa, omogućuju detaljno razumijevanje događaja koji se javljaju nezavisno jedan od drugog, uz konstantnu stopu intenziteta. S druge strane, Hawkesovi procesi, koji su prirodno proširenje Poissonovih, uvode koncept samo-pobude (*self-excitement*), omogućujući modeliranje događaja u klasterima gdje svaki događaj može utjecati na buduće događaje.

Hawkesovi procesi su posebno zanimljiva klasa točkovnih procesa koju je početkom sedamdesetih godina prošlog stoljeća uveo A. G. Hawkes za modeliranje seizmičkih događaja. Od tada su našli primjenu u raznim područjima, od modeliranja potresa do financijske analize. Njihova ključna karakteristika je uvjetna funkcija intenziteta koja predstavlja uvjetnu gustoću vjerojatnosti pojave događaja u neposrednoj budućnosti s obzirom na prošlost.

Ovaj rad pruža pregled osnovnih pojmova i teorije točkovnih procesa, nakon čega se detaljno uvodi Hawkesov proces, njegova funkcija intenziteta događaja, te sheme za simulaciju događaja i procjenu parametara. Također je prikazan praktičan primjer modeliranja širenja informacija na društvenim mrežama pomoću Hawkesovih *self-excitement* procesa, gdje se analizira dinamika retweet kaskada. Dizajn funkcijske jezgre i rezultati procjene parametara i predikcije popularnosti detaljno su opisani. (Primjer je iz [9]).

Također je uključen dodatak s nekim matematičkim rezultatima i R kodovima korištenim za simulaciju i analizu, čime se omogućuje repliciranje i daljnje istraživanje prezentiranih modela.

## **Dio I**

# **Osnovna teorija vremenskih točkovnih procesa**



# Poglavlje 1

## Točkovni procesi

Statistička analiza se često bavi pojavama koje se događaju u vremenu ili prostoru na slučajnan način. Za kvantifikaciju i modeliranje takvih pojava koriste se točkovni procesi.

Neformalno možemo definirati točkovni proces kao kolekciju vremensko indeksiranih slučajnih varijabli. To je zapravo prirodan način modeliranja stvarnih događaja poput dolazaka klijenata u trgovinu, ili slučajevi nasilja i fizičkog obračunavanja između suparničkih bandi gdje svaki slučaj se razlikuje po veličini i intenzitetu [1]. Još možemo proširiti model kako bi previdio mogući val kriminalnih odmazdi izazvanih zbog tog pojedinog slučaja.

Stohastička procedura za generiranje podataka je temeljni pristup mnogih vjerojatnosnih modela na kojima gradi moderno strojno učenje [2], jednostavno jer odgovara ponašanju nekih sustava iz stvarnog svijeta kao što smo prethodno opisali.

### 1.1 Točkovni i brojeći procesi

Među najpoznatijim točkovnim procesima je Poissonov proces koji ćemo izbliza proučavati, ne samo kao primjer (klasičnog) točkovnog procesa nego i zbog uske povezanosti s Hawke-sonim procesima, koji također spadaju pod tu vrstu točkovnih procesa.

Točkovni proces bismo mogli definirati kao niz sortiranih slučajnih varijabli  $T_1, T_2, \dots$  u  $\mathbb{R}_+$  koje nazivamo *događaji*.

Primjerice, u prethodno opisanom scenariju klijenata u trgovinu imamo da su dolasci u određenom vremenskom intervalu  $[0, T]$  (uvijek gledamo zatvorene intervale) i svaka pojava novog kupca je zabilježena s  $T_i$  gdje je  $i$  označava  $i$ -tog kupca i vrijednost od  $T_i$  je upravo vrijeme dolaska  $i$ -tog kupca (ponekad pišemo funkcijski:  $T(i) = \alpha$ , gdje je  $\alpha \in [0, T]$ . Čitamo:  *$i$ -ti kupac je došao u trenutku  $\alpha$* ).

Drugu reprezentaciju istog procesa možemo postići pomoću vremena međudolazaka

(eng. *inter-arrival times*), odnosno slučajnih varijabli  $\mathcal{T} = (\tau_1, \tau_2, \dots)$  gdje je  $\tau_i = T_i - T_{i-1}$ , a još definiramo  $T_0 = 0$ . (Po prethodnom primjeru, sad bi  $\tau_i = \beta$  glasilo *i-ti kupac je došao  $\beta$  jedinica vremena nakon zadnjeg kupca*).

**Definicija 1.1.1.** Neka je  $\{T_i, i \in \mathbb{N}\}$  niz nenegativnih slučajnih varijabli tako da  $\forall i \in \mathbb{N}$  vrijedi  $T_i < T_{i+1}$ , tada skup  $\{T_i, i \in \mathbb{N}\}$  nazivamo **točkovnim procesom** u  $\mathbb{R}_+$ .

Ponekad u literaturi ih nazivaju vremenskim točkovnim procesom jer se radi o nizu nad vremenskim nenegativnim realnim poljem. Kad su točki u dvodimensionalnom polju, tada se radi o prostornom točkovnom procesu.

**Napomena 1.1.2.** Uočimo da smo u definiciji onemogućili slučaj pojave dvaju različitih događaja istovremeno. Odnosno da zaista vrijedi strogi uređaj  $<$ , te ne postoji  $i \in \mathbb{N}$  za koji bi vrijedio  $T_i = T_{i+1}$ . Takvi točkovni procesi se nazivaju *jednostavni*.

Svakom točkovnom procesu pridružujemo njegov **brojeći proces**  $\{N(t), t \geq 0\}$  gdje je

$$N(t) := \sum_{i \in \mathbb{N}} \mathbb{1}_{\{t \geq T_i\}} \quad (1.1)$$

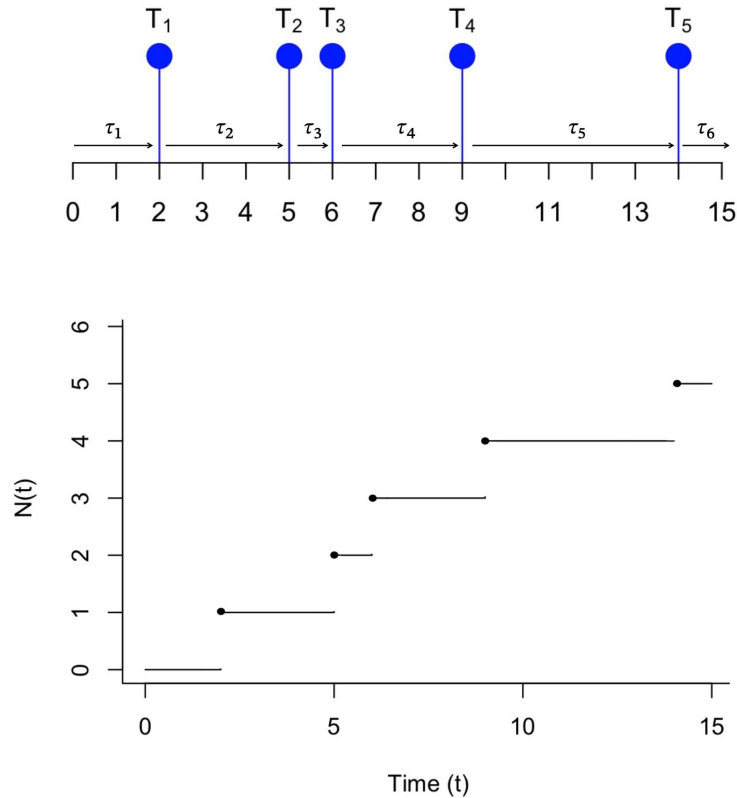
$\mathbb{1}_{\dots}$  je indikatorska funkcija sa  $\Omega$  u  $\{0, 1\}$  koja vraća 1 ako je  $\omega \in C$ , gdje je  $C$  događaj, odnosno 0 ako nije. Preciznije:

$$\mathbb{1}_C(\omega) = \begin{cases} 1 & \text{ako } \omega \in C \\ 0 & \text{inače} \end{cases} \quad (1.2)$$

Imamo dakle da je  $N(t)$  zapravo broj događaja do trenutka  $t$  gdje je  $t \in [0, T]$ . Zato se  $N(t)$  povećava za jedan nakon svakog događaja  $T_i$ . Eksplicitno možemo napisati:

$$N(t) = \begin{cases} 0 & \text{ako } 0 \leq t < T_1 \\ 1 & \text{ako } T_1 \leq t < T_2 \\ 2 & \text{ako } T_2 \leq t < T_3 \\ \dots & \\ n & \text{ako } T_n \leq t < T_{n+1} \\ \dots & \end{cases} \quad (1.3)$$

Lako možemo vidjeti ekvivalenciju između jednog i drugog procesa. To jest, da su skup događaja  $T_1, T_2, \dots$  (točkovni proces) i njegov odgovarajući brojeći proces ekvivalentni prikazi istog procesa.



Slika 1.1: *Gornja*: točkovni proces.  $T_i$  označava  $i$ -ti događaj dok je s  $\tau_i$  označen vrijeme međudolazaka. *Donja*: njegov ekvivalentni brojeći proces gdje na  $y$  osi stoje vrijednosti  $N(t)$  kroz vrijeme.

**Primjer 1.1.3.** Uzmimo primjer iz slike 1.1 gdje je zadan jedan točkovni proces. Vidimo da u intervalu od  $[0, 7]$  broj događaja je konačan (ima ih 3 da budemo precizni) i vrijedi  $T_1 = 2, T_2 = 5$  i  $T_3 = 6$ , gdje su  $T_i \in \mathcal{T}$ .

S druge strane, brojeći proces  $N(t)$  je prikazan na grafu i njegova reprezentacija je zadana preko slučajnih varijabli  $T_i$ . Recimo za  $t = 5$ ,

$$N(5) = \sum_{i \in \mathbb{N}} \mathbb{1}_{\{5 \geq T_i\}} = \mathbb{1}_{\{5 \geq T_1\}} + \mathbb{1}_{\{5 \geq T_2\}} + \mathbb{1}_{\{5 \geq T_3\}} = 1 + 1 + 0 = 2 \quad (1.4)$$

Što je upravo broj događaja do trenutka  $t = 5$ .

## 1.2 Svojstva brojećih procesa

Uzmimo sad nekoliko primjera iz stvarnog života gdje bismo mogli iskoristiti brojeće procese i pomoću kojih se jasno vidi svojstva koja ćemo opisati.

Ranije na početku poglavlja smo govorili o primjeru klijenata u trgovini (njihovi dolasci u određenom vremenskom intervalu). Uvijek promatramo vremenske intervale jer tako su uvijek zadane u problemima koje modeliramo iz stvarnog svijeta. U tom slučaju imamo za bilo koji  $t$  iz  $[0, T]$  broj  $N(t)$  daje informaciju o broju kupaca koji su stigli do trenutka  $t$ . Svakim dolaskom kupca povećava se broj za 1. Drugi primjer može biti gledanje jednog videa na mrežnoj usluzi *YouTube*. Tada je  $\{N(t), t \geq 0\}$  brojeći proces gdje  $N(t)$  označava broj *views*-a ili pregleda do trenutka  $t$ .

Iz prošlih primjera uočimo da brojeći proces  $\{N(t), t \geq 0\}$  u potpunosti opisuju točkovni proces, odnosno cijeli zadani sustav. Naime definirali smo točkovni proces kao niz nenegativnih slučajnih varijabli  $\{T_i, i \in \mathbb{N}\}$  gdje su  $T_i$  događaji u vremenu. Time smo izvadili iz sustava informaciju o točnoj pojavi svakog događaja i o vremenima između njih, opisana pomoću  $\mathcal{T} = (\tau_1, \tau_2, \dots)$ . Zaista smo zaključili da vrijedi

$$\tau_i = T_i - T_{i-1} \quad (1.5)$$

uz dodatno definiranje vrijednosti  $T_0$  i  $T_{N+1}$  gdje je  $N$  broj događaja u promatranom intervalu.

Sada vidimo kako možemo odrediti  $T_i = t$ . Izraz odgovara  $i$ -tom događaju i određuje vrijeme dolaska (u trenutku  $t$ ). Iz toga znamo da će  $N(t)$  postići vrijednost  $i$  u određenom trenutku budući da svakim dolaskom povećava se za 1, onda  $i - ti$  dolazak je povećao funkciju do  $i$ -te vrijednosti (pod pretpostavkom  $N(0) = 0$ , o tome ćemo posvetiti napomenu na kraju ovog zaključka). Dakle ako  $i < N(t)$  za bilo koji  $t \in [0, T]$ , tada znamo da se dogodio  $T_i$ , a vrijednost  $t$  dobivamo upravo kao najmanji  $u \in [0, T]$  za koji vrijedi  $N(u) = i$ .

Time smo pokazali što smo htjeli. Formalno zapisujemo u obliku propozicije:

**Propozicija 1.2.1.** *Brojeći proces potpuno opisuje točkovni proces i obratno, odnosno oba su ekvivalentna.*

Mogli bismo i matematički zapisati jednostavno za  $\{T_i, i \in \mathbb{N}\}$  u  $[0, T]$  vrijedi

$$0 < T_1 < T_2 < \dots < T_{N(t)} < T \quad (1.6)$$

Veza između ta dva procesa može biti uočljivija iz sljedeće druge perspektive. Uzmimo vremena dolazaka, dobivena kao

$$T_n = \sum_{i=1}^n \tau_j \quad (1.7)$$

gdje je  $T_n$  vrijeme  $n$ -tog događaja.

**Napomena 1.2.2.** Ponekad se u definiciji brojećeg procesa dodaju uvjeti  $N(0) = 0$  i konačnost kao pretpostavke [3]. U našim daljnjim promatranjima ćemo zaista to i pretpostaviti makar smo odlučili to izostaviti u definiciji.

Sada možemo na temelju definicije formalno napisati osnovna svojstva brojećeg procesa.

**Korolar 1.2.3.** *Neka je  $\{N(t), t \geq 0\}$  brojeći proces, tada slijedi:*

- a)  $N(t) \geq 0$ , za sve  $t \geq 0$ ,
- b)  $N(t)$  je prirodan broj (ili 0),
- c) ako je  $s < t$ , tada vrijedi  $N(s) \leq N(t)$ ,

pri čemu za sve  $s < t$  imamo da je  $N(t) - N(s)$  jednak broju događaja u poluotvorenom intervalu  $\langle s, t \rangle$ .

Prije nego nastavimo dalje, napomenimo kako točkovni procesi mogu biti karakterizirani po različitim svojstvima prirasta.

**Definicija 1.2.4.** (Nezavisnost prirasta) *Točkovni proces ima nezavisne priraste ako su slučajne varijable*

$$N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1}) \quad (1.8)$$

nezavisne za svaki izbor od  $t_1, t_2, \dots, t_n$  gdje je  $t_i \geq 0$  za koje vrijedi

$$t_1 < t_2 < \dots < t_n \quad (1.9)$$

**Definicija 1.2.5.** (Stacionarnost prirasta) *Točkovni proces ima stacionarne priraste ako broj događaja u intervalu  $\langle s, s+t \rangle$  ima istu distribuciju za sve  $s$ -ove, to jest ako distribucija od prirasta*

$$N(t+h) - N(t) \quad (1.10)$$

ovisi samo o duljini  $h$ .

## Poglavlje 2

# Poissonovi procesi

Siméon Denis Poisson <sup>1</sup> je francuski matematičar i fizičar rođen 21. lipnja 1781. godine u Loiretu. Sin jednog francuskog pukovnika, studirao je u École Polytechnique u Parizu krajem 19. stoljeća i ubrzo se isticao kao vrstan matematičar i briljatan um – njegove prve publikacije su još za vrijeme studija. Imao je priliku susretati u znanstvenim krugovima Pariza ličnosti poput Joseph Louis Lagrange ili Pierre-Simon Laplace.

Njegova famozna distribucija daje ime najvažnijim točkovnim procesima: Poissonovi procesi. Podjeljeni su u homogene i nehomogene u ovisnosti o intenzitetu  $\lambda$ .

U ovom poglavlju ćemo isto govoriti i o funkcijama gustoće, distribucije, svojstvu zaboravljivosti i još nekoliko pojedinosti koje nam trebaju. Tijekom rada ćemo još povezivati teoriju s primjerima iz stvarnog svijeta gdje se primijenjuju i homogeni i nehomogeni Poissonovi procesi.



Slika 2.1: *Siméon Denis Poisson i Poissonov proces*

---

<sup>1</sup>Francusko izgovaranje: [pwason].

## 2.1 Homogeni Poissonovi procesi

Postoji mnogo ekvivalentnih definicija homogenih Poissonovih procesa, ova je jedna od tih.

**Definicija 2.1.1. (Poissonov proces)** Neka je  $(\tau_i)_{i \geq 1}$  niz nezavisnih i jednako distribuiranih (od sada nadalje: n.j.d.)<sup>2</sup> eksponencijalnih slučajnih varijabli s parametrom  $\lambda$  i događaji  $T_n = \sum_{i=1}^n \tau_i$ . Tada točkovni proces  $\{T_i, i \in \mathbb{N}\}$  zovemo Poissonov proces s intenzitetom  $\lambda$ .

Prije nego nastavimo, kratka digresija koja će služiti za podsjetnik. Ako je čitatelj upoznat s pojmovima distribucije, funkcije gustoće i Poissonovih varijabli, može ovaj dio slobodno preskočiti.

Ako je  $X$  slučajna varijabla, funkciju  $F_X : \mathbb{R} \rightarrow [0, 1]$  definiranu s

$$F_X(x) = \mathbb{P}(X \leq x) \quad (x \in \mathbb{R}) \quad (2.1)$$

zovemo funkcija distribucije od  $X$ . (eng. *cumulative distribution function*, CDF).

Ako imamo diskretnu slučajnu varijablu koja poprima vrijednosti u skupu  $\mathbb{Z}$ , njenu vjerojatnosnu funkciju mase  $f_X : \mathbb{Z} \rightarrow [0, 1]$  definiramo kao

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \quad (2.2)$$

(eng. *probability mass function*, PMF).

Nadalje, kažemo da slučajna varijabla  $X$  ima Poissonovu distribuciju ako poprima vrijednosti u  $\mathbb{N}_0$  te joj je vjerojatnosna funkcija mase oblika

$$f(x; \lambda) = \mathbb{P}(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad (2.3)$$

za  $x \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ . Pišemo  $X \sim \text{Poiss}(\lambda)$ .

Vraćamo se na definiciju Poissonovog procesa. Rečeno je da je niz međusobno jednako distribuiranih *eksponencijalnih* slučajnih varijabli, dakle nameće se pitanje gdje je Poisson u cijeloj priči i zašto je proces nazvan po njemu. Odgovor je u  $N(t)$ . U Poissonovom procesu slučajne varijable jesu eksponencijalne, ali prirasti imaju Poissonovu razdiobu, odnosno  $N(t)$  je Poissonova slučajna varijabla.

**Lema 2.1.2.** Neka su  $X_1, \dots, X_n$  nezavisne eksponencijalne slučajne varijable s parametrom  $\lambda$  (vrijedi  $X_i \sim \text{Exp}(\lambda)$ ), tada  $X_1 + \dots + X_n$  ima Gamma( $n, \lambda$ ) distribuciju, odnosno

$$X_1 + \dots + X_n \sim \Gamma(n, \lambda). \quad (2.4)$$

<sup>2</sup>U engleskoj literaturi: Independent identical distributed, često i.i.d. ili IID

Ova lema će nam biti potrebna za dokaz propozicije 2.1.4. Još ćemo napisati jedan važan rezultat koji ćemo koristiti u daljnjim razmatranjima:

**Propozicija 2.1.3.** *Neka je točkovni proces  $\{T_1, T_2, \dots\}$  homogeni Poissonov proces s intenzitetom  $\lambda$  gdje su mu vremena međudolazaka  $\{\tau_1, \tau_2, \dots\}$ . Nadalje neka je  $T'_1 = T_{N_t+1} - t$  vrijeme preostalo nakon trenutka  $t$  pa sve do idućeg (prvog) događaja  $T_{N_t+1}$ , na isti način definiramo  $T'_k = T_{N_t+k} - t$  za  $k \geq 2$  i još definiramo  $\tau'_1 = T'_1$  i  $\tau'_i = \tau_{N_t+i}$  za  $i \geq 2$ . (vidjeti sliku 2.2). Tada vrijedi da je  $(T'_1, T'_2, \dots)$  ponovno homogeni Poissonov proces s intenzitetom  $\lambda$  i nezavisan od  $(N_t, T_1, T_2, \dots, T_{N_t})$ .*

*Skica dokaza.* Tvrdnja slijedi jer je događaj

$$A = \{N_t = n, T_1 = t_1, \dots, T_n = t_n\} \quad (2.5)$$

za  $t_1 < t_2 < \dots < t_n \leq t$ , isti kao i događaj

$$\{\tau_{n+1} > t - t_n, T_1 = t_1, \dots, T_n = t_n\}, \quad (2.6)$$

a budući da su  $T_1, \dots, T_n$  nezavisni od  $\tau_{n+1}$ , imamo da je

$$\mathbb{P}(T'_1 > s \mid A) = \mathbb{P}(\tau_{n+1} - (t - t_n) > s \mid \tau_{n+1} > t - t_n) = \mathbb{P}(\tau_{n+1} > s) = e^{-\lambda s}. \quad (2.7)$$

U predzadnjem koraku smo iskoristili svojstvo zaboravljivosti od eksponencijalne razdiobe. To jest, uvjetno na  $\tau_{n+1} > u$ ,  $\tau - u$  ponovno ima istu eksponencijalnu razdiobu. Dakle,  $T'_1$  je nezavisna od prošlosti (to jest, od svih događaja oblika  $A$ ) i ima opet eksponencijalnu distribuciju s parametrom  $\lambda$ . Uvjetno na  $A$  imamo da je  $\tau'_k = \tau_{n+k}$ ,  $k \geq 2$ , pa tvrdnja slijedi budući da su vremena međudolazaka  $(\tau_i)_{i \geq 1}$  n.j.d.

□

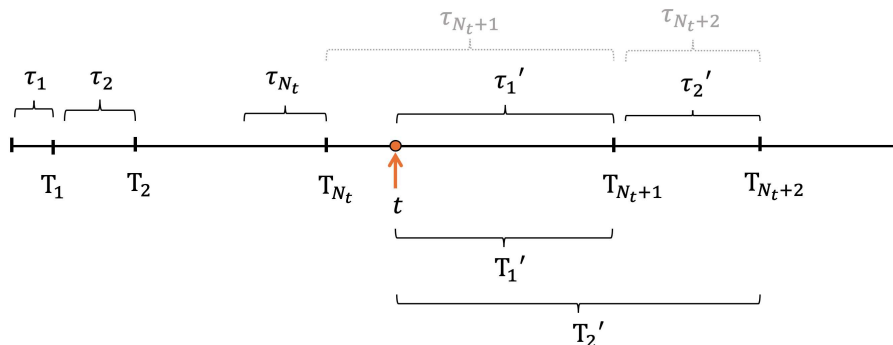
Svojstvo homogenog Poissonovog procesa iz prethodne propozicije dakle slijedi iz *svojstva zaboravljivosti* eksponencijalne razdiobe<sup>3</sup>. Iz njega praktički slijede sva ostala svojstva. Riječima: prošlost ne igra nikakvu ulogu u određivanju novih događaja, odnosno vremena novog neposrednog događaja nakon određenog trenutka  $t$  (do na trenutno stanje u kojem se nalazimo).

Sad idemo iskazati i dokazati važnu ekvivalenciju definicije Poissonovog procesa.

---

<sup>3</sup>eng. *memorylessness property*





Slika 2.2: Novi Poissonov proces od trenutka  $t$ , gdje je  $\tau_{N_t+i} = \tau'_i$  i  $T_{N_t+i} = T'_i + t$  za  $i \geq 2$ .

**Propozicija 2.1.4.** *Točkovni proces  $\{T_1, T_2, \dots\}$  je homogeni Poissonov proces s intenzitetom  $\lambda$  ako i samo ako vrijedi sljedeće:*

1.  $N(0) = 0$ .
2.  $\{N(t) : t \geq 0\}$  ima nezavisne priraste.
3. Za sve  $0 \leq s < t$ ,  $N\langle s, t \rangle \sim \text{Pois}(\lambda(t - s))$ , specijalno, prirasti su stacionarni.

*Dokaz.* Dokazat ćemo smjer ( $\Rightarrow$ ). Neka je točkovni proces  $\{T_1, T_2, \dots\}$  homogeni Poissonov proces s intenzitetom  $\lambda > 0$ . Prvi uvjet je više iz tehničkih razloga i prokomentirali smo ga u napomeni 1.2.2. Dakle slijedi iz pretpostavke definicije brojećeg procesa (naznačuje da proces počinje u  $t = 0$ ).

Što se druge točke tiče, prisjetimo se da po definiciji točkovni proces ima nezavisne priraste ako za sve  $n \geq 1$  i sve  $0 \leq s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n$  vrijedi da su  $\{N(I_i)\}_{i=1, \dots, n}$ ,  $I_i = \langle s_i, t_i \rangle$  međusobno nezavisni. Dodatno, definirali smo  $N_t = \max\{n \geq 0 : T_n \leq t\}$  (izrazi  $N(t)$  i  $N_t$  su ekvivalentni) i još smo podesili  $T_0 = 0$  i  $T_n = \tau_1 + \dots + \tau_n$ ,  $\tau_n \sim \text{Exp}(\lambda)$  n.j.d. slučajne varijable.

Za dva različita intervala  $\langle s_1, t_1 \rangle$  i  $\langle s_2, t_2 \rangle$  definirana kao maloprije ( $0 \leq s_1 < t_1 < s_2 < t_2$ ) njihova nezavisnost slijedi direktno iz prethodne propozicije. Naime, za  $t$  uzimamo točno  $t_1$ , prema propoziciji 2.1.3 ( $T'_1, T'_2, \dots$ ) su nezavisne sa  $N_t, T_1, T_2, \dots, T_{N_t}$  i vrijedi da  $N(\langle s_1, t_1 \rangle)$  ovisi isključivo o  $\{T_1, T_2, \dots, T_{N_t}\}$  dok  $N(\langle s_2, t_2 \rangle)$  jedino o  $\{T'_1, T'_2, \dots\}$ , dakle zaista su  $N(\langle s_1, t_1 \rangle)$  i  $N(\langle s_2, t_2 \rangle)$  nezavisni kad su skupovi  $\{T_1, T_2, \dots, T_{N_t}\}$  i  $\{T'_1, T'_2, \dots\}$

disjunktni. Sad ponovnom upotrebom ovog postupka možemo proširiti tvrdnju na konačno mnogo intervala. Bitno je da su intervali disjunktni kako bi ovo prošlo.

Nedostaje dokazati 3. tvrdnju. Najprije ćemo pokazati da je  $N_t \sim \text{Poiss}(\lambda t)$ ,  $\forall t > 0$ . Za to moramo pokazati da vrijedi

$$\mathbb{P}(N_t = n) = \mathbb{P}(\text{Poiss}(\lambda t) = n), \quad \forall n \geq 0 \quad (2.8)$$

znamo da su sljedeći događaji ekvivalentni:

$$N_t = n \iff T_n \leq t \text{ i } T_{n+1} > t. \quad (2.9)$$

Iz toga slijedi

$$\begin{aligned} \mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t < T_{n+1}) = \mathbb{P}(T_n \leq t, \tau_{n+1} > t - T_n) \\ &= \int_0^{\infty} \mathbb{P}(u \leq t, \tau_{n+1} > t - u \mid T_n = u) f_{T_n}(u) du \\ &\stackrel{\text{nez.}}{=} \int_0^t \mathbb{P}(\tau_{n+1} > t - u) f_{T_n}(u) du. \end{aligned} \quad (2.10)$$

Gdje smo koristili nezavisnost varijabli  $T_n$  i  $\tau_{n+1}$ . Znamo da je  $T_n = \tau_1 + \dots + \tau_n$  i da su  $\tau_i \sim \text{Exp}(\lambda)$  pa koristeći lemu 2.1.2 dobijemo da je  $T_n$  gamma distribuirana slučajna varijabla i funkcija gustoće nam je poznata (A.0.1):

$$f_{T_n}(u) \stackrel{T_n \sim \Gamma}{=} \frac{\lambda^n u^{n-1} e^{-\lambda u}}{\Gamma(n)} = \frac{\lambda^n u^{n-1} e^{-\lambda u}}{(n-1)!}, \quad (2.11)$$

zato što vrijedi

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx = (n-1)! \quad (2.12)$$

za sve  $n \in \mathbb{N}$  (dokaz ide indukcijom po  $n$ . Više u A.0.2).

S druge strane,  $\tau_i \sim \text{Exp}(\lambda)$  dakle znamo njenu funkciju distribucije i gustoće.

$$\mathbb{P}(\tau_{n+1} \geq x) = 1 - \mathbb{P}(\tau_{n+1} < x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} \quad (2.13)$$

Što daje sve skupa iz (2.10), (2.11) i (2.13)

$$\int_0^t \mathbb{P}(\tau_{n+1} > t - u) f_{T_n}(u) du = \int_0^t \frac{\lambda^n u^{n-1}}{(n-1)!} e^{-\lambda u} e^{-\lambda(t-u)} du = \frac{\lambda^n t^n}{n!} e^{-\lambda t}, \quad (2.14)$$

dakle  $N_t \sim \text{Poiss}(\lambda t)$ .

Konačno tvrdnja  $N(\langle s, t \rangle) = N_t - N_s \sim \text{Poiss}(\lambda(t - s))$  slijedi iz

$$N_t - N_s \sim N_{s-t} \sim \text{Poiss}(\lambda(t - s)), \quad (2.15)$$

pri čemu činjenica da  $N_t - N_s$  ima istu distribuciju kao i  $N_{s-t}$  slijedi iz Prop. 2.1.3.

□

Još jedno zgodno svojstvo Poissonovih procesa je da uvjetno na  $N(T) = n$ , događaji  $\{T_1, T_2, \dots, T_n\}$  imaju istu distribuciju kao i  $\{U(1), \dots, U(n)\}$  pri čemu su  $U(1) < U(2) < \dots < U(n)$  dobiveni sortiranjem n.j.d. slučajnih varijabli  $U_1, \dots, U_n$  koje imaju uniformu distribuciju na  $[0, T]$ .

## 2.2 Intenzitet $\lambda$

Kao što smo vidjeli ranije, vremena međudolazaka  $\tau_i$  su nezavisna i svako od njih slijedi eksponencijalnu distribuciju s parametrom  $\lambda$  (za niz  $(\tau_i)_{i \geq 1}$  gdje prvi događaj se događa u trenutku  $\tau_1$ , drugi se događa  $\tau_2$  jedinica vremena nakon prvog, itd.)

Ovdje, oznaka  $f_\tau(t)$  označava funkciju gustoće slučajne varijable  $\tau$  koja uzima vrijednosti označene s  $t$ . Funkcija gustoće za eksponencijalnu je

$$f_\tau(t) = \begin{cases} \lambda e^{-\lambda t} & \text{ako } t \geq 0 \\ 0 & \text{ako } t < 0 \end{cases} \quad (2.16)$$

za  $\lambda > 0$ . Sada očekivanje od  $\tau$  možemo dobiti u zatvorenoj formi:

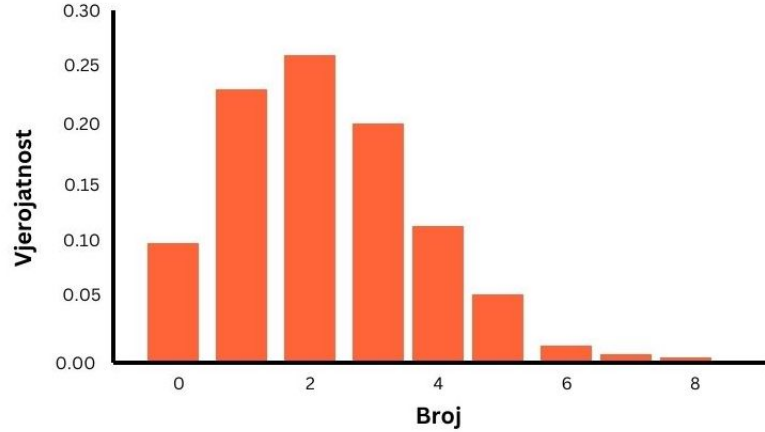
$$\begin{aligned} \mathbb{E}_\tau[\tau] &= \int_{-\infty}^{\infty} t f_\tau(t) dt = \lambda \int_0^{\infty} t e^{-\lambda t} dt \\ &= \lambda \left[ -\frac{t}{\lambda} e^{-\lambda t} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt \\ &= 0 + \left[ -\frac{1}{\lambda} e^{-\lambda t} \right]_0^{\infty} = \frac{1}{\lambda} \end{aligned}$$

Dakle, ako je očekivano vrijeme između događaja  $\mathbb{E}[\tau] = \frac{1}{\lambda}$ , to znači da je  $\lambda$  zapravo očekivani broj događaja u jedinici vremena.

Dodatno, možemo pokazati da vrijedi

$$\lambda = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h} \quad (2.17)$$

što daje dodatnu interpretaciju intenziteta  $\lambda$ .

Slika 2.3: Poissonova distribucija, za  $\lambda = 2.24$ 

U homogenom procesu  $\lambda$  je konstanta, ali to ne mora biti tako. Može biti funkcija kroz vrijeme, zapisana kao  $\lambda(t)$ . U tom slučaju govorimo o *nehomogenom Poissonovom procesu*. S njima ćemo se pozabaviti u poglavlju 2.3.

Uvest ćemo sada dodatnu notaciju, koja se ponekad naziva i Landauovim simbolim po njemačkom matematičaru Edmundu Landauu.<sup>4</sup>

**Definicija 2.2.1.** *Neka su  $f : \mathbb{R} \rightarrow \mathbb{R}$  i  $g : \mathbb{R} \rightarrow \mathbb{R}$  dvije funkcije i neka je  $g$  pozitivna u otvorenoj okolini od  $a \in \mathbb{R}$ , to jest za  $x \in (a - \xi, a + \xi) \setminus \{a\}$  za neki  $\xi > 0$ . Kažemo da je  $f$  'malo  $o$ ' od  $g$  u okolini od  $a$  i pišemo  $f(x) = o(g(x))$  kada  $x \rightarrow a$  ako*

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 0. \quad (2.18)$$

Budući da se simbol  $h$  često koristi za predstavljanje broja koji se približava nuli, u ovom radu ćemo reći da funkcija  $f(\cdot)$  je  $o(h)$  ako

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0. \quad (2.19)$$

Drugim riječima, da bi funkcija  $f(\cdot)$  bila  $o(h)$ , potrebno je da  $f(h)$  teži prema nuli brže nego  $h$ .

<sup>4</sup>Drugi Landauov simbol je veliko O notacija. S istim pretpostavcima kao i kod definicije 2.2.1, kažemo da je  $f$  'veliko O' od  $g$  u okolini od  $a$  i pišemo  $f(x) = O(g(x))$  kada  $x \rightarrow a$  ako  $\limsup_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| < \infty$ .

**Propozicija 2.2.2.** Za homogeni Poissonov proces  $\{T_1, T_2, \dots\}$  s intenzitetom  $\lambda$  vrijedi:

$$(i) \mathbb{P}(N\langle t, t+h \rangle = 1) = \lambda h + o(h)$$

$$(ii) \mathbb{P}(N\langle t, t+h \rangle \geq 2) = o(h)$$

*Dokaz.* Najprije pokazujemo svojstvo (ii).

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 2)}{h} &= \lim_{h \rightarrow 0} \frac{1 - \mathbb{P}(N\langle t, t+h \rangle = 0) - \mathbb{P}(N\langle t, t+h \rangle = 1)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h} - \lambda h e^{-\lambda h}}{h} \\ &= \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h}}{h} - \lim_{h \rightarrow 0} \lambda e^{-\lambda h} = \lambda - \lambda = 0, \end{aligned}$$

što koristeći 2.19 daje:

$$0 = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 2)}{h} \iff \mathbb{P}(N\langle t, t+h \rangle \geq 2) = o(h). \quad (2.20)$$

Za svojstvo (i) idemo najprije dokazati izraz iz (2.17) koji smo prokomentirali, ali ni smo dokazali.

$$\lambda = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h}, \quad (2.21)$$

Pozivanjem na 3. svojstvo propozicije 2.1.4 imamo da je  $N\langle t, t+h \rangle \sim \text{Poiss}(\lambda h)$ , dakle vrijedi:

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h} = \lim_{h \rightarrow 0} \frac{1 - \mathbb{P}(N\langle t, t+h \rangle = 0)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h}}{h} = \lambda.$$

Time je dokazan gornji izraz (2.21). Sada konačno pomoću njega i 2.19 dobijemo svojstvo (i):

$$\begin{aligned} \mathbb{P}(N\langle t, t+h \rangle = 1) = \lambda h + o(h) &\iff \mathbb{P}(N\langle t, t+h \rangle \geq 1) - \mathbb{P}(N\langle t, t+h \rangle \geq 2) = \lambda h + o(h) \\ &\stackrel{(2.20)}{\iff} \mathbb{P}(N\langle t, t+h \rangle \geq 1) - \lambda h = o(h) + o(h) = o(h) \\ &\stackrel{\text{Def. 2.2.1}}{\iff} \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1) - \lambda h}{h} = \lim_{h \rightarrow 0} \left( \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h} - \lambda \right) = 0, \end{aligned}$$

a desni izraz je istinit upravo iz jednakosti (2.21). Time je dokazana propozicija.  $\square$

## 2.3 Nehomogeni Poissonovi procesi

Vrijeme je da prijedemo na malo složenije procese. Vidjeli smo da u homogenim Poissonovim procesima, događaji stižu slučajno s konstantnim intenzitetom  $\lambda$ . Prirodno je pretpostaviti da dolasci ipak ovise o nekoliko varijabli koje utječu na njih, kao što bi primjerice bio vremenski raspon u trgovini. Nastavljajući alegoriju gdje smo dolaske kupaca poistovjećivali s našim događajima u promatranom homogenom Poissonovom procesu, pretpostavili smo da su dolasci nezavisni i jednakim intenzitetom kroz dan. Međutim, u stvarnom životu to nije tako. Naime u 7 ujutro najčešće ima manje pojava nego u 11 ili u 17 popodne, kad je jedan veliki broj građana je završio s poslom (drugi primjer je prometna gustoća u takozvanim *rush hours* koja je veća u određenim periodima nego u ostatku dana). Ovo možemo oblikovati funkcijski po  $\lambda$  koji ovisi o vremenu  $t$ , umjesto da bude konstanta. Ovako se intezitet mijenja kroz vrijeme i stvara trenutke gdje je vjerojatnost veća nego u drugim trenucima. Upravo taj Poissonov proces čiji intenzitet nije konstanta nego je

$$\lambda = \lambda(t), \quad (2.22)$$

gdje je  $t \in [0, T]$ , nazivamo *nehomogenim Poissonovim procesom*. Razlika između jednog i drugog je prikazana na slici 2.4. Mogli bismo redefinirati homogene Poissonove procese kao specijalni slučaj nehomogenih, kad je funkcija intenziteta zapravo konstanta.

**Definicija 2.3.1.** Za danu funkciju  $\lambda : [0, \infty) \rightarrow [0, \infty)$ , točkovni proces  $\{T_1, T_2, \dots\}$  je (nehomogeni) Poissonov proces ako vrijedi:

1.  $N(0) = 0$ .
2.  $\{N(t) : t \geq 0\}$  ima nezavisne priraste.
3. Za sve  $0 \leq s < t$ ,  $N\langle s, t \rangle \sim \text{Poiss}(\int_s^t \lambda(u) du)$ .

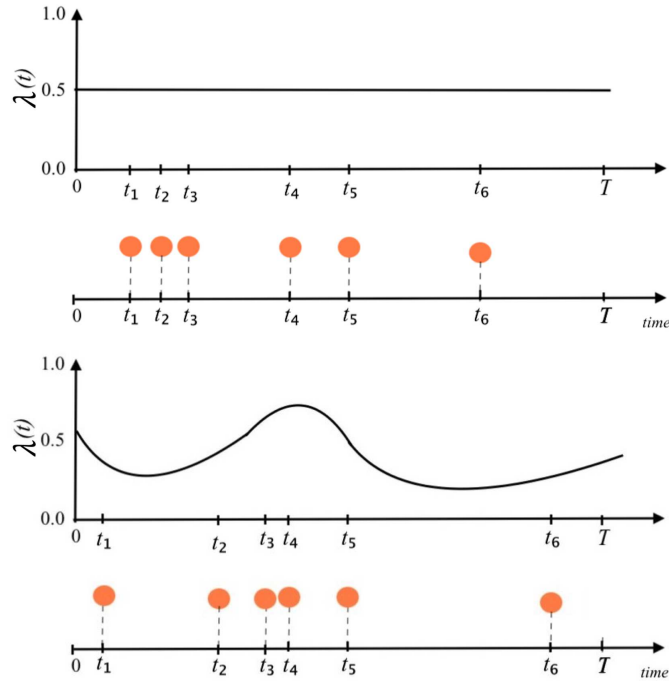
Sada homogeni Poissonov proces očigledno dobijemo ako stavimo  $\lambda(t) := \lambda$  za sve  $t \geq 0$ .

S druge strane, na sličan način pokazujemo kao ranije da za sve  $t > 0$  vrijedi

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h}. \quad (2.23)$$

Naime, iz svojstva 3. slijedi:

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}(N\langle t, t+h \rangle \geq 1)}{h} = \lim_{h \rightarrow 0} \frac{1 - \mathbb{P}(N\langle t, t+h \rangle = 0)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-\int_t^{t+h} \lambda(u) du}}{h} = \lambda(t),$$



Slika 2.4: Homogeni proces (*gornji graf*) s konstantom  $\lambda = 0.5$ . Događaji su jednako vjerojatni po svoj dužini segmenta  $[0, T]$ . S druge strane, za nehomogeni proces (*donji graf*) intenzitet je funkcija  $\lambda(t)$  koja ovisi o vremenu. Tamo gdje je vjerojatnost veća, tamo se najčešće pojavljuju događaji. U oba procesa je prikazano šest događaja  $\{t_1, t_2, \dots, t_6\}$ \* u vremenu (strelica s oznakom *time*) i iznad odgovarajuća funkcija intenziteta.

\* U ovoj skripti smo se odlučili za sljedeću notaciju: kad se radi o slučajnim varijablama tada pišemo velikim slovom, dakle  $\{T_1, T_2, \dots, T_6\}$ . No kad te varijable su zapravo točke konkretne (pojave ili događaji), tada ćemo ih malim slovom napisati:  $\{t_1, t_2, \dots, t_6\}$  kao jedna konkretna realizacija kojom se dalje bavimo do kraja primjera.

budući da je integral neprekidna funkcija te je derivacija integrala u  $t$  jednaka  $\lambda(t)$  za sve  $t > 0$ . (Za dodatna pojašnjenja, pogledati A.0.3).

Analogno kao u homogenom slučaju, pomoću izraza (2.23) slijedi propozicija 2.2.2 za nehomogene procese:

$$\mathbb{P}(N(t, t + h) = 1) = \lambda(t)h + o(h), \tag{2.24}$$

Što je važan rezultat za daljnje nadograđivanje funkcije intenziteta, kao što ćemo uskoro vidjeti.

**Dio II**

**Hawkesovi procesi**



## Poglavlje 3

# Uvjetna funkcija intenziteta

Pod nekim dodatnim uvjetima, točkovni proces  $\{T_1, T_2, \dots\}$  možemo potpuno karakterizirati pomoću uvjetne funkcije intenziteta. Označena kao  $\lambda^*(t)$  predstavlja trenutnu stopu dolaska novih događaja u vremenu  $t$ , s obzirom na povijest prethodnih događaja. U nastavku ćemo preskočiti neke tehničke detalje, oni se mogu pronaći u [20] i [21].

U Hawkesovim procesima ćemo vidjeti još jedan korak složenosti: ne samo da  $\lambda$  neće biti konstanta nego zapravo neće ni biti unaprijed definirana ili zadana kao da bi se radilo o krivulji ili funkciji koja ovisi samo o vremenu (vidi sliku 2.4 za takav primjer), nego se svaki put mijenja s obzirom na nove događaje ili dolaske. Povijest događaja utječe na buduće, odnosno što je bilo do trenutka  $t$  nije zanemarivo. Primijetite da je ovo zapravo suprotno od svojstva zaboravljivosti u kojem vjerojatnost novih događaja ne ovisi o prijašnjem stanju nego isključivo sadašnjem.

**Definicija 3.0.1.** *Uvjetnu funkciju intenziteta definiramo kao*

$$\lambda^*(t) := \lim_{h \rightarrow 0} \frac{\mathbb{P}\{N_{t+h} - N_t = 1 \mid \mathcal{H}(t)\}}{h}, \quad (3.1)$$

*pri čemu  $\mathbb{P}(\dots \mid \mathcal{H}(t))$  predstavlja vjerojatnost uvjetno na informacije (povijest) do trenutka  $t$ .*

*Povijest  $\mathcal{H}(t)$  uvijek uključuje broj događaja koji su se dogodili  $N(t)$ , kao i točna vremena  $T_1, \dots, T_{N(t)}$ .*

Bitno je naglasiti da je  $(\lambda^*(t), t \geq 0)$  slučajan proces jer za sve  $t \geq 0$ ,  $\lambda^*(t)$  ovisi o  $N(t)$  i  $T_1, \dots, T_{N(t)}$ . Strogo formalno,  $(\mathcal{H}(t), t \geq 0)$  se definira kao familija  $\sigma$ -algebri, to jest *filtracija*, ali to izlazi van okvira ovog rada. Uбудuće ćemo uvjetnu funkciju intenziteta neformalno pisati kao  $\lambda^*(t) := \mathbb{P}\{N(t+dt) - N(t) = 1 \mid \mathcal{H}(t)\}$ .

Budući da preko  $\lambda^*(t)$  tražimo vjerojatnost događaja u trenutku  $t$ , na klasičan način gledamo intervale proizvoljno malene, a to postizemo puštajući limes u nulu. I gornji i

donji izraz gledaju vjerojatnost da razlika u brojećoj funkciji  $N$  bude 1 u tom intervalu  $[t, t + h]$  (u drugom izrazu,  $h = dt$  kad  $h \rightarrow 0$ ), što znači da je u tom intervalu stigao događaj koji je uzrokovao povećanje  $N$  za 1.

Vrijedno je naglasiti da  $\lambda^*(t)$  ovisi samo o informacijama o brojećom procesu u prošlosti, što znači da je  $\mathcal{H}_t$ -imjeriv. Štoviše, primijetite da  $\lambda^*(t)$  može biti samo nenegativan.

Bitno je napomenuti da nam povijest  $\mathcal{H}(t)$  daje informaciju ne samo o vremenskim trenucima kad je događaj postignut nego i kad nije. Pogotovo nam govori da se u intervalu između zadnjeg događaja pa sve do promatranog trenutka  $t$  nije dogodio nikakav događaj. (interval  $\langle s, t \rangle$  pri čemu je  $s = T_{N(t)}$  ne sadrži nikakav događaj). To nas odmah vodi do *funkcija hazarda*.

Prije nego što ih definiramo, trebamo učiniti jedan korak nazad. Funkcija hazarda je funkcija povezana s takozvanim distribucijama preživljavanja (eng. *survival distributions*). U tim modelima promatramo slučajne varijable koje se ne odnose na sam događaj nego na vrijeme čekanja do trenutka zbivanja. Primjeri mogu biti:

- udio stanovništva koji će preživjeti određene godine života.
- vrijeme između primjene cjepiva i razvoja infekcije.
- trajanje rada stroja do kvara.

Općenitije, promatramo vrijeme života (*lifetime*) preostalo do fatalnog zbivanja (bilo kakvog), npr. u području biologije, strojarstva, ekonomije ili sociologije.

Kao u većini slučajeva postoje različiti pristupi i definicije. Mi ćemo samo ukratko definirati osnovne pojmove (slijedit ćemo natuknice iz [6]) jer, iako je povezana tema s našom, nije glavni predmet ovog rada. (za više o tome, vidjeti [7]).

**Definicija 3.0.2.** *Neka je  $T$  slučajna varijabla i  $F(\cdot)$  njena funkcija distribucije, funkciju  $S(t) := 1 - F(t) = \mathbb{P}(T > t)$  nazivamo funkcijom preživljavanja. (eng. survival function).*

Zaputili smo se u cijelu ovu priču nakon što smo opazili dodatnu informaciju u uvjetnoj funkciji intenziteta  $\lambda^*(t)$ : nije bilo događaja u segmentu  $\langle T_{N(t)}, t \rangle$ , to jest neposredno prije trenutka koji promatramo. Ovo možemo gledati kao preživljavanje do trenutka  $t$ . Samo pažljivo s interpretacijom: nigdje nije rečeno da se u trenutku  $t$  dogodilo nešto, nego da se ništa nije dogodilo do prije trenutka  $t$ . Sad nas zanima vjerojatnost da se u tom trenutku dogodi nešto, ili prirodnim jezikom "koliko nam je još preostalo do tog fatalnog, ali neizbježnog zbivanja" što može biti prirodna smrt, infekcija ili kvar (nastavljajući s gore navedenim primjerima).

**Definicija 3.0.3.** Za fiksne  $i \in \mathbb{N}_0$  i trenutke  $0 < t_1 < \dots < t_{i-1}$ , funkciju hazarda definiramo kao

$$\phi_i(t | t_1, \dots, t_{i-1}) = \mathbb{P}(N[t, t + dt] = 1 | N_t = i - 1, T_1 = t_1, \dots, T_{i-1} = t_{i-1}) \quad (3.2)$$

za sve  $t > t_{i-1}$ , gdje  $dt$  označava

$$:= \lim_{h \rightarrow 0} \frac{\mathbb{P}(N[t, t + h] = 1 | T_i \geq t, T_1 = t_1, \dots, T_{i-1} = t_{i-1})}{h}. \quad (3.3)$$

Uočimo da možemo zapisati funkciju hazarda na ekvivalentan način:

$$\phi_i(t | t_1, \dots, t_{i-1}) = \mathbb{P}(N[t, t + dt] = 1 | T_i \geq t, T_1 = t_1, \dots, T_{i-1} = t_{i-1}).$$

Sada možemo prikazati našu uvjetnu funkciju intenziteta pomoću njih:

$$\lambda^*(t) = \phi_i(t | T_1, \dots, T_{i-1}), \quad (3.4)$$

ako je  $t \in (T_{i-1}, T_i]$ . Tu se onda jasno vidi da je  $\lambda^*(t)$  slučajna varijabla.

Sad kad smo uveli uvjetnu funkciju intenziteta  $\lambda^*(t)$ , nužnu za definiranje Hawkesovog procesa, idemo pokazati da je ona zapravo poopćenje funkcije intenziteta  $\lambda(t)$  iz definicije Poissonovog procesa. Odnosno da u slučaju homogenog Poissonovog procesa vrijedi

$$\lambda^*(t) = \mu, \quad (3.5)$$

gdje je  $\mu$  konstanta, a u slučaju nehomogenog Poissonovog procesa

$$\lambda^*(t) = \lambda(t), \quad (3.6)$$

Za svaki  $t > 0$ .

Kako bismo to dokazali, treba nam nezavisnost prirasta koja nam kaže da ono što se u Poissonovim procesima događa nakon trenutka  $t$  ne ovisi o  $\mathcal{H}(t)$ .

Naime, za homogeni Poissonov proces s intenzitetom  $\lambda$  imamo da je

$$\begin{aligned} \phi_i(t | t_1, \dots, t_{i-1}) &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(t, t + h] = 1 | N(t) = i - 1, T_1 = t_1, \dots, T_{i-1} = t_{i-1})}{h} \\ [nezavisnost prirasta: prop. 2.1.3] &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(t, t + h] = 1)}{h} \stackrel{Prop. 2.2.2}{=} \lambda, \end{aligned}$$

za svaki  $t > 0$ .

S druge strane, za nehomogeni Poissonov proces s funkcijom intenziteta  $\lambda(t)$  na potpuno analogni način slijedi tvrdnja, jedino umjesto propozicije 2.2.2 imamo zadnju jednakost iz jednadžbe (2.24) koju smo pokazali. Odnosno:

$$\phi_i(t \mid t_1, \dots, t_{i-1}) = \dots = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(t, t+h] = 1)}{h} \stackrel{(2.24)}{=} \lambda(t), \quad (3.7)$$

za svaki  $t > 0$ .

## Poglavlje 4

# Osnove Hawkesovih procesa

U poglavlju 2 smo vidjeli homogeni i nehomogeni Poissonov proces. U Poissonovom procesu događaji dolaze neovisno jedan o drugome uz konstantnu stopu  $\lambda$ . U nehomogenom Poissonovom procesu događaji dolaze neovisno jedan o drugome, ali ovaj put uz promjenjivu stopu  $\lambda(t)$ .

Međutim, u stvarnosti nije uvijek slučaj da događaji dolaze neovisno jedan o drugome. Znamo, na primjer, da nam seizmička aktivnost u prošlosti može pomoći u predviđanju budućih potresa. Dakle, u slučaju potresa vidimo da pojava jednog događaja utječe na pojavu sljedećeg događaja. Što je isto kao reći da dolazak događaja povećava vjerojatnost budućih događaja u bliskoj budućnosti. Slučaj potresa nije jedini: kaskada retweetova na društvenim mrežama je drugi primjer [9]. Tamo ćemo otkriti da čim se dogodi retweet, funkcija intenziteta se povećava, dakle vjerojatnost budućih događaja se povećava.

U ovom poglavlju ćemo predstaviti klasu procesa u kojoj stopa dolazaka događaja ipak ovisi o prošlosti, odnosno ovise jedan o drugome. Takvi procesi se nazivaju *self-exciting* procesi, a najpoznatiji među njima je *Hawkesov proces*.

*Self-exciting* proces je točkovni proces u kojem dolazak događaja uzrokuje povećanje uvjetne funkcije intenziteta  $\lambda^*(t)$ . Njoj smo posvetili poglavlje 3. U *self-exciting* procesima, uvjetna funkcija intenziteta se povećava ako se dogodi novi događaj i time je pojava novog događaja vjerojatnija. Pogledati sliku 4.1 za primjer realizacije uvjetne funkcije intenziteta jednog takvog procesa.

Godine 1971. Hawkes je predložio novi *self-exciting* model, koji se pokazao vrlo praktičnim i još uvijek se koristi. Radi se o procesu koji eksplicitno ovisi o svim prethodno nastalim događajima, tako da neposredno nakon svakog događaja funkcija intenziteta raste i odmah krene padati eksponencijalno.

**Definicija 4.0.1. (Hawkesov proces)** Neka je  $\{T_1, T_2, \dots\}$  točkovni proces s pripadajućom povijesti  $\mathcal{H}(t)$ ,  $t \geq 0$ . Kažemo da je proces Hawkesov ako mu je uvjetna funkcija intenziteta

oblika

$$\lambda^*(t) = \lambda_0 + \sum_{i:t>T_i} \mu(t - T_i), \quad (4.1)$$

gdje su  $\lambda_0 > 0$ ,  $\mu : \langle 0, \infty \rangle \rightarrow [0, \infty]$  i gdje su  $T_i < t$  svi događaji do trenutka  $t$ .

Pogledajmo izbliza ovu uvjetnu funkciju intenziteta. Najprije vidimo da se radi o zbroju dva dijela: konstanta  $\lambda_0$  i suma funkcije  $\mu$ . Konstanta predstavlja pozadinski intenzitet procesa (engl. *background intensity*), koji ne ovisi o povijesti niti prethodnim događajima. Kao što ćemo vidjeti u 4.1, Hawkesov proces može se promatrati kao proces u kojem postoje događaji "imigranata" i "potomaka" <sup>1</sup>. Pozadinski intenzitet određuje dolazak događaja imigranata, a njihov dolazak je neovisan o prethodnim događajima koji su se dogodili. Ponekad umjesto konstante može biti i funkcija  $\lambda_0(t) : \mathbb{R} \mapsto \mathbb{R}_+$  (*base intensity function*).

S druge strane imamo drugi dio funkcije koji karakterizira Hawkesov proces i čini ga bude *self-exciting*: funkcija  $\mu(\cdot)$ , takozvana jezgrena funkcija (eng. *memory kernel*). Ona modulira učinak koji događaj  $T_i$  ima nad uvjetnom funkcijom intenziteta u trenutku  $t$ . Jasno vidimo da u slučaju  $\mu(\cdot) = 0$  zapravo imamo homogeni Poissonov proces, zato ćemo taj slučaj izostaviti.

Najčešće je funkcija  $\mu(\cdot)$  padajuća kako bi događaji bliži u vremenu imali veći utjecaj u trenutnom intenzitetu događaja dok daleki u vremenu imaju sve manje i manje.

## Odabir jezgre $\mu$

Funkcija  $\mu(\cdot)$  može biti različitih oblika, ne nužno padajućih. Međutim, mi ćemo kratko predstaviti dva poznata odabira funkcije: eksponencijalna jezgra i tzv. *power-law* jezgra <sup>2</sup>. ([9], str. 8)

*Eksponencijalna jezgra* je popularna padajuća funkcija koju Hawkes koristio u svojem radu [10]. Oblik glasi:

$$\mu(t) = \alpha e^{-\beta t}, \quad (4.2)$$

gdje je  $\alpha \geq 0$ ,  $\beta > 0$  i  $\alpha < \beta$  (vidi sliku B.1). Tada je uvjetna funkcija intenziteta

$$\lambda^*(t) = \lambda_0 + \sum_{i:t>T_i} \alpha e^{-\beta(t-T_i)}. \quad (4.3)$$

Dakle pri dolasku događaja, funkcija se povećava za  $\alpha$  no kroz vrijeme to povećavanje se smanjuje eksponencijalno sa stopom  $\beta$ .

<sup>1</sup>eng. *immigrants and offsprings*

<sup>2</sup>eng. *exponential kernel and power-law kernel*

*Power-law jezgra* je kao i eksponencijalna jezgra tipični odabir za Hawkesove procese, a često se upotrebljava u seizmologiji i analizi društvenih mreža:

$$\mu(t) = \frac{K}{(c + t)^p}, \quad (4.4)$$

gdje uvjetna funkcija intenziteta glasi:

$$\lambda^*(t) = \lambda_0 + \sum_{i:t>T_i} \frac{K}{(c + (t - T_i))^p}, \quad (4.5)$$

za tri konstante  $K$ ,  $c$  i  $p$ . Kasnije ćemo ih objasniti u našem primjeru (slika B.2).

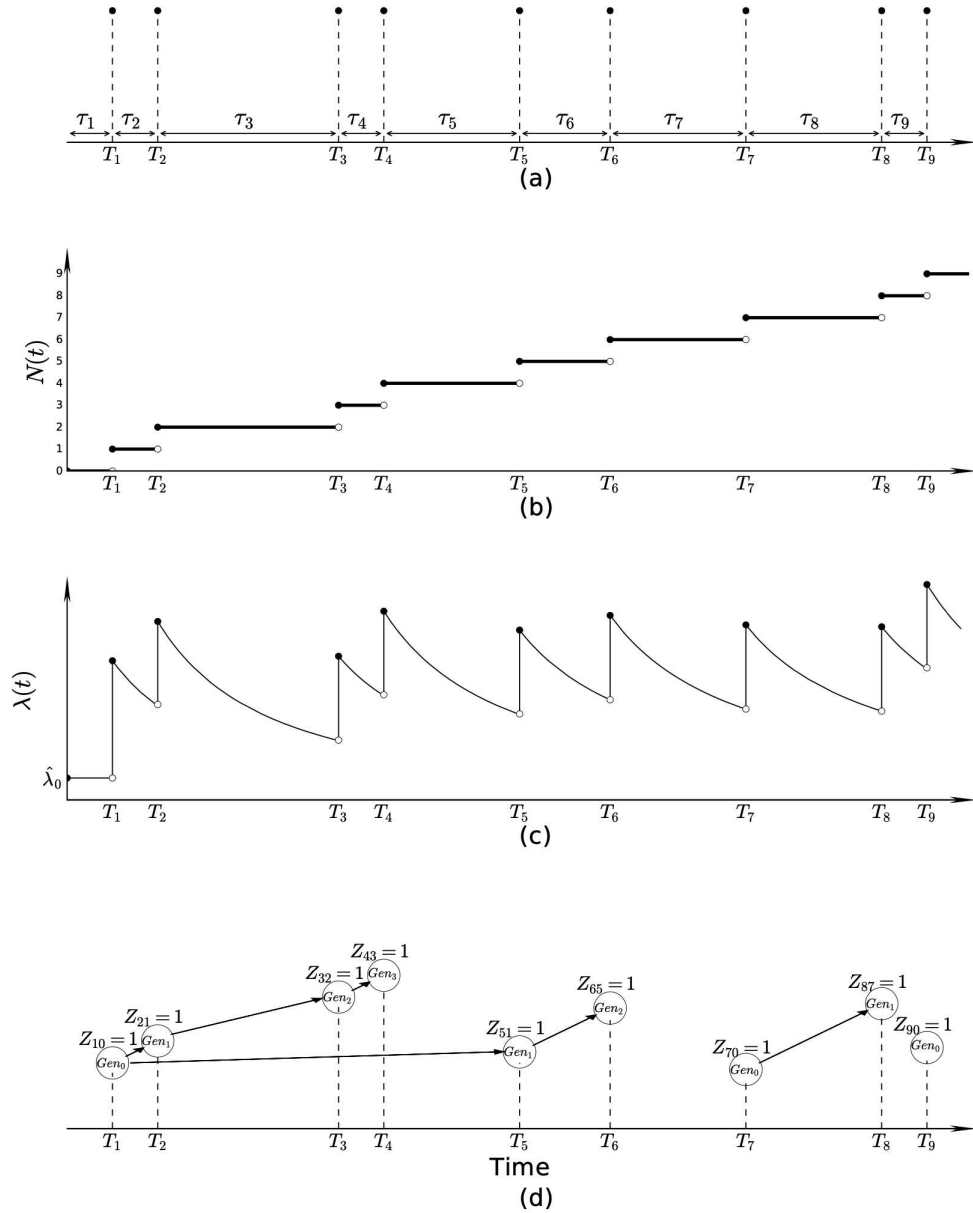
Na slici 4.1 može se detaljno vidjeti jedan Hawkesov proces (prikazana su vremena  $T_1, \dots, T_9$ , skupa sa svojim pripadajućim vremenima međudolazaka  $\tau_1, \dots, \tau_9$  i pripadajući brojeći proces  $N(t)$ ), i dodatno prikazuje uvjetnu funkciju intenziteta. Naime, bez informacije o funkciji intenziteta, grafovi pod (a) i (b) lako bi mogli pripadati homogenom ili nehomogenom Poissonovom procesu. No odmah vidimo da se funkcija intenziteta  $\lambda(t)$  mijenja kroz vrijeme, odnosno mijenja se odmah nakon pojave događaja i smanjuje se sve do pojave sljedećeg događaja.

## 4.1 Struktura grananja

Hawkesov proces se može interpretirati kao Poissonov klaster proces (za više tehničkih detalja, vidjeti [26]). U ovoj interpretaciji, događaji u Hawkesovom procesu dijele se u dvije kategorije: imigranti i potomci. Kao što sami nazivi sugeriraju, imigrantski događaji dolaze neovisno od drugih događaja (sjetimo se, njih određuje prvi dio uvjetne funkcije intenziteta:  $\lambda_0$ ), dok su potomci potaknuti prethodnim događajima u procesu. Potomci se strukturiraju u klasteru, povezani s određenim imigrantskim događajem. Ovaj način strukturiranja događaja naziva se *struktura grananja* [9].

Postoje tri razloga za objašnjavanje ovog alternativnog pogleda na Hawkesov proces. Prvo, ova nova interpretacija olakšava dokazivanje određenih svojstava Hawkesovog procesa, poput stacionarnosti procesa. Drugo, na temelju strukture grananja može se osmisлити algoritam za simulaciju Hawkesovog procesa. Treće, struktura grananja može omogućiti modeliranje određenih situacija iz stvarnog života.

U ostatku ovog odjeljka detaljno ćemo razmotriti strukturu grananja i izračunati dvije ključne veličine: faktor grananja – očekivani broj događaja koji su izravno potaknuti određenim događajem u Hawkesovom procesu – i očekivani ukupni broj događaja u klasteru potomaka. Obje ove veličine postaju vrlo važne kada se Hawkesovi procesi primjenjuju na praktična područja, poput društvenih mreža na internetu.



Slika 4.1: (Preuzeta iz [9]) Hawkesov proces s eksponencijalnom jezgrom. (a) Prikazana su prvih devet vremena događaja.  $T_i$  predstavljaju vremena događaja, dok  $\tau_i$  predstavljaju vremena međudolazaka događaja. (b) Brojeći proces kroz vrijeme  $N_t$  se povećava za jednu jedinicu u svakom vremenu događaja  $T_i$ . (c) Funkcija intenziteta kroz vrijeme. Primijetite kako svaki događaj izaziva skok, nakon čega slijedi eksponencijalni pad (d) Skrivena ili neopažena struktura grananja Hawkesovog procesa. Svaki krug predstavlja jedan događaj koji se dogodio u  $T_i$ , strelice predstavljaju odnos korijen-potomak.  $Gen_i$  specificira generaciju događaja, s  $i = 0$  za imigrante ili  $i > 0$  za potomke.  $Z_{ij}$  su slučajne varijable, tako da je  $Z_{i0} = 1$  ako je događaj  $T_i$  imigrant, a  $Z_{ij} = 1$  ako je događaj  $T_i$  potomak događaja  $T_j$ .



## Primjer strukture grananja

Vremena dolazaka imigranata tvore homogeni Poissonov proces s parametrom intenziteta  $\lambda_0$ , dok su dolasci potomaka generirani nehomogenim Poissonovim procesima čija je funkcija inteziteta određena funkcijom  $\mu(\cdot)$ . Na ovaj način bismo mogli razdvojiti Hawkesov proces na dva dijela: homogeni dio i nehomogene dijelove (jer ih ima više). Time smo postigli našu strukturu grananja: imigranti su nezavisni jedan od drugoga, a iz svakoga od tih imigranata izvire jedan nehomogeni proces potomaka. Najčešće gledamo po generacijama: nulta generacija su imigranti, prva generacija prvi potomci pojedinog imigranta, itd.

Promatrajući dalje sliku 4.1 pod (d) vidimo jedan primjer strukture grananja devet događaja Hawkesovog procesa kroz vrijeme (isti proces koje smo prokomentirali ranije). Vremena događaja  $T_i$  su označena krugovima, a odnosi 'roditelj-potomak' između događaja prikazani su strelicama. Uvodimo slučajne varijable  $Z_{ij}$ , gdje je  $Z_{i0} = 1$  ako je događaj  $i$  imigrant, a  $Z_{ij} = 1$  ako je događaj  $i$  potomak događaja  $j$  (oznake slijede [9]). Tekst u svakom krugu označava generaciju kojoj događaj pripada, tj.  $\text{Gen}_k$  označava  $k$ -tu generaciju. Imigranti su označeni kao  $\text{Gen}_0$ , dok generacije  $\text{Gen}_k, k > 0$  označavaju njihove potomke. Na primjer,  $T_2$  i  $T_5$  su neposredni potomci imigranta  $T_1$ , tj. matematički izraženo kao  $Z_{21} = 1, Z_{51} = 1$  i  $Z_{10} = 1$ .

Potomci povezani s određenim roditeljem dolaze prema nehomogenom Poissonovom procesu s intenzitetom  $\mu(\cdot)$ , tj.  $T_2$  i  $T_5$  su realizacije događaja koji dolaze iz nehomogenog Poissonovog procesa s intenzitetom  $\mu(t - T_1)$  za  $t > T_1$ . Događaj koji proizvodi potomka opisuje se kao neposredni predak ili korijen potomka,  $T_7$  je neposredni predak  $T_8$ . Događaji koji su izravno ili neizravno povezani s imigrantom tvore klaster potomaka povezan s tim imigrantom, npr.  $T_1$  je imigrant, a  $T_2, T_3, T_4, T_5$  i  $T_6$  tvore njegov klaster potomaka. Slično,  $T_7$  i  $T_8$  tvore drugi klaster. Konačno,  $T_9$  je klaster sam po sebi. (Vidi sliku 4.2).

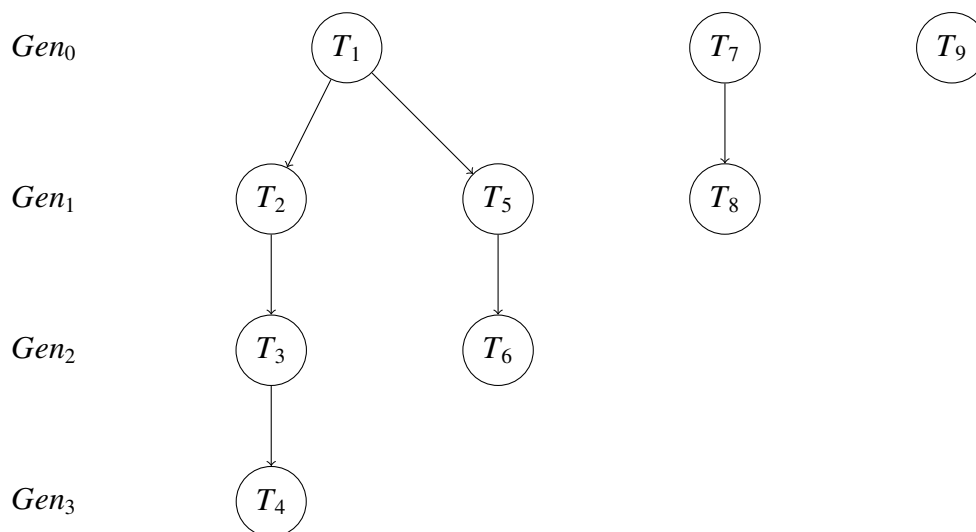
## Faktor grananja

Faktor grananja (eng. *branching factor*) predstavlja važan broj koji opisuje Hawkesove procese. Definicija je jednostavna: faktor grananja  $n^*$  je očekivani broj izravnih potomaka jednog događaja u procesu. (Na grafu 4.1  $T_2$  je direktni potomak od  $T_1$ , ali  $T_3$  nije).

Ako se u trenutku  $t \geq 0$  pojavi događaj, onda njegovi direktni potomci čine nehomogeni Poissonov proces s funkcijom inteziteta  $\lambda(s) := \mu(s - t)$  za  $s \geq t$ . To znači da je ukupan broj direktnih potomaka zapravo ukupan broj događaja tog Poissonovog procesa u intervalu  $[t, \infty)$ , a to je Poissonova slučajna varijabla s očekivanjem

$$\int_t^\infty \mu(s - t) ds = \int_0^\infty \mu(s) ds. \quad (4.6)$$

Taj broj dakle ne ovisi o  $t$ , te ga označavamo s  $n^*$ .



Slika 4.2: Primjer Hawkesovog procesa slike 4.1 s generacijama događaja. Korijeni su imigranti ( $Gen_0$ ), ostali su potomci.

Intuitivno, faktor grananja  $n^*$  opisuje broj događaja u procesu ili, neformalno, *virality* ili utjecaj jedne poruke u društvenim mrežama. Dodatno, faktor grananja utječe na vjerojatnost da klaster ukupno sadrži konačan broj potomaka. Naime, za  $n^* < 1$  proces je u podkritičnom režimu: broj događaja u svakom klasteru je konačan s vjerojatnosti 1. Za  $n^* > 1$ , tada proces je superkritičnom režimu gdje  $\lambda(t)$  raste i za svaki klaster postoji pozitivna vjerojatnost da sadrži beskonačan broj potomaka. Intuitivno, proces eksplodira i nikad ne završava. ([9], str. 9).

U slučaju eksponencijalne jezgre, gdje  $\mu(t) = \alpha e^{-\beta t}$  (4.2), faktor grananja iznosi

$$n^* = \int_0^{\infty} \alpha e^{-\beta s} ds = \alpha \left[ -\frac{e^{-\beta s}}{\beta} \right]_{s=0}^{s=\infty} = \frac{\alpha}{\beta} \quad (4.7)$$

### Očekivani broj događaja u klasteru potomaka

Ako se nalazimo u slučaju kada je  $n^* < 1$ , tada moguće je dobiti bolju aproksimaciju veličine svakog klastera pojedinog imigranta.

Neka je  $B_0$  broj imigranata u jednom klasteru, dakle  $B_0 = 1$ . To je populacija u 0-toj generaciji ( $Gen_0$ ).  $B_1$  definiramo kao direktne potomke od imigranata, dakle  $Gen_1$ , njih ima  $B_{1,1}$ . Zatim  $B_2 = \sum_{j=1}^{B_1} B_{2,j}$ , što su direktni potomci članova prve generacije. Ovako

dobijemo  $i$ -ti član

$$B_i = \sum_{j=1}^{B_{i-1}} B_{i,j}, \quad (4.8)$$

gdje je dakle  $B_{i,j}$  broj jedinki u  $i$ -toj generaciji koji su potomci  $j$ -te jedinke iz  $(i-1)$ -ve populacije. Ključno je ovdje da su varijable  $(B_{i,j})_{i,j=1,2,\dots}$  nezavisne i jednako distribuirane, te sve imaju  $\text{Poiss}(n^*)$  distribuciju. Takav proces se zove *Galton-Watsonov proces*. Primijetimo da se (4.8) radi o slučajnoj sumi jer je sama varijabla  $B_{i,j}$  slučajna isto kao i  $B_{i-1}$ .

Zanima nas očekivani broj događaja u  $\text{Gen}_i$ , odnosno  $A_i := \mathbb{E}[B_i]$ . Tada za očekivani broj događaja u klasteru vrijedi:

$$N_\infty := \mathbb{E} \left[ \sum_{i=0}^{\infty} B_i \right] = \sum_{i=0}^{\infty} \mathbb{E}[B_i] = \sum_{i=0}^{\infty} A_i. \quad (4.9)$$

Sad da bismo izračunali očekivani broj potomaka  $A_i$ , koristit ćemo (4.8). No prije pokazujemo jedno svojstvo očekivanja koje ćemo koristiti.

Općenito vrijedi za  $X_1, X_2, \dots$  n.j.d. slučajne varijable

$$\mathbb{E} \left[ \sum_{j=1}^n X_j \right] = \sum_{j=1}^n \mathbb{E}[X_j] = n \cdot \mathbb{E}[X_1], \quad (4.10)$$

jer je  $\mathbb{E}[X_j] = \mathbb{E}[X_1]$ , za sve  $j \in \mathbb{N}$ . Još može se pokazati za slučajnu sumu da je za slučajnu varijablu  $N$  nezavisnu od  $X_1, X_2, \dots$

$$\mathbb{E} \left[ \sum_{j=1}^N X_j \right] = \mathbb{E}[N] \cdot \mathbb{E}[X_1], \quad (4.11)$$

Konačno možemo izračunati naš očekivani broj događaja u  $\text{Gen}_i$ :

$$A_i = \mathbb{E}[B_i] \stackrel{(4.11)}{=} \mathbb{E}[B_{i-1}] \cdot \mathbb{E}[B_{i,1}] = \mathbb{E}[B_{i-1}] \cdot n^* = A_{i-1} n^*. \quad (4.12)$$

Znajući da je  $A_0 = 1$ , izvedemo:

$$A_i = A_{i-1} n^* = A_{i-2} (n^*)^2 = \dots = A_0 (n^*)^i = (n^*)^i, \quad i \geq 1 \quad (4.13)$$

tada dobivamo očekivanu vrijednost veličine svakog klastera imigranata  $N_\infty$  kao zbroj konvergentnog geometrijskog reda (pod pretpostavkom da je  $n^* < 1$ ):

$$N_\infty = \sum_{i=0}^{\infty} A_i = \sum_{i=0}^{\infty} (n^*)^i = \frac{1}{1-n^*}, \quad \text{gdje je } n^* < 1. \quad (4.14)$$

# Poglavlje 5

## Algoritmi simuliranja

U ovom dijelu rada ćemo se fokusirati na metode simuliranja Hawkesovih procesa. Da bismo to postigli najprije ćemo govoriti o simuliranju (nehomogenih) Poissonovih procesa. Iako ćemo navesti nekoliko algoritama, fokus će biti na tzv. algoritmu stanjivanja (eng. *thinning algorithm*) koji ćemo kasnije upotrijebiti. Zadnji dio poglavlja ćemo posvetiti podešavanju parametara Hawkesovih procesa.

Teorijska svojstva koja smo ispitali u poglavlju 2 mogu se ilustrirati korištenjem simulacijskih metoda. Osim toga, simulacijske metode mogu nam pomoći da bolje razumijemo kako proces funkcionira. Tako ćemo u ovom poglavlju pogledati simulaciju homogenih i nehomogenih Poissonovih procesa. Čitavo poglavlje prati shemu iz [11], a kodovi u R jeziku su priloženi u Dodatcima (pod B), zajedno s prikazima grafova i rezultata.

### 5.1 Homogeni Poissonov proces

Prvi proces koji ćemo simulirati je homogeni Poissonov proces. Pretpostavimo da želimo simulirati Poissonov proces s intenzitetom  $\lambda$  na intervalu  $[0, T]$ . Simulacija bi nam tada trebala dati popis vremena događaja  $\{T_1, \dots, T_n\}$  svih  $n$  događaja koji su se dogodili između 0 i  $T$ , gdje je  $n = N(T)$ .

Umjesto simulacije vremena događaja  $\{T_1, \dots, T_n\}$ , simulirat ćemo međudolazna vremena  $\{\tau_1, \dots, \tau_n\}$ , budući da je distribucija vremena međudolazaka poznata (podsjetimo se, vrijeme  $n$ -tog događaja procesa je tada jednostavno zbroj prvih  $n$  vremena međudolazaka). To ćemo generirati pomoću *inverzne transformacijske metode*.

Ta metoda koristi činjenicu da ako je  $U$  uniformna slučajna varijabla na intervalu  $(0, 1)$ , tada je slučajna varijabla

$$\tau = -\frac{1}{\lambda} \ln(1 - U) \quad (5.1)$$

eksponencijalno distribuirana s parametrom  $\lambda$ . Naime,

$$\begin{aligned} F_{\tau}(x) &= \mathbb{P}\left(-\frac{1}{\lambda} \ln(1 - U) \leq x\right) \\ &= \mathbb{P}(\ln(1 - U) \geq -\lambda x) \\ &= \mathbb{P}(1 - U \geq e^{-\lambda x}) \\ &= \mathbb{P}(U \in [e^{-\lambda x}, 1]) \\ &= 1 - \mathbb{P}(U \leq e^{-\lambda x}) = 1 - e^{-\lambda x}, \end{aligned}$$

gdje smo iskoristili činjenicu da je  $U \sim \text{Unif}(0, 1)$  zapravo ekvivalentno s  $1 - U \sim \text{Unif}(0, 1)$ . Dakle zaista  $\tau \sim \text{Exp}(\lambda)$ .

Tada se generiranje međudolaznih vremena pomoću uniformnih slučajnih varijabli može opisati sljedećim koracima:

---

**Algorithm 1** Generiranje vremena međudolazaka

---

**Require:**  $[0, T]$

**Ensure:**  $(\tau_1, \dots, \tau_n)$

1:  $N = 0$

2:  $sum = 0$

3: polje međudolazaka  $\mathcal{T}$ .

4: **while**  $sum < T$  **do**

5:     1. Generiraj uniformnu slučajnu varijablu  $U$  na intervalu  $\langle 0, 1 \rangle$ .

6:     2. Izračunaj međudolazno vrijeme  $\tau = -\frac{1}{\lambda} \ln(1 - U)$ .

7:     3.  $sum = sum + \tau$

8:     4. Ako je  $sum < T$ , tada spremi  $\tau$  u polju vremena međudolazaka  $\mathcal{T}$ .

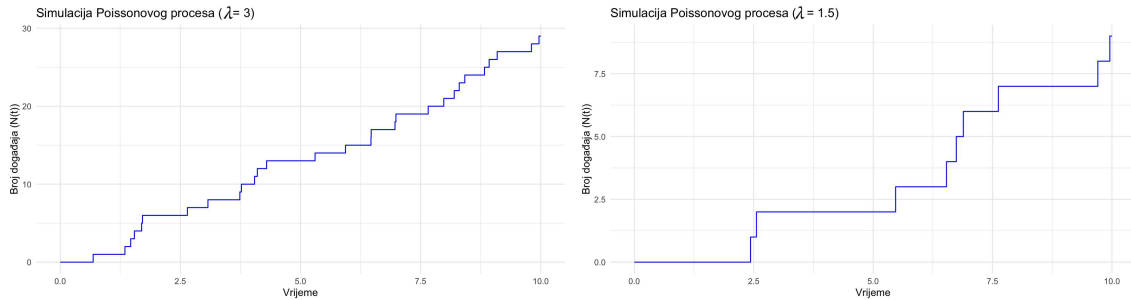
9: **end while**

---

Sada imamo sve što nam treba za simuliranje homogenog Poissonovog procesa. Izvođenjem algoritma u programskom jeziku  $R$  dobijemo naš traženi proces.

Uspoređujući obje slike 5.1, grafički možemo uočiti neke stvari koje smo primijetili u teoriji. Može se vidjeti da su međudolazna vremena procesa s intenzitetom  $\lambda = 3$  kraća od međudolaznih vremena procesa s intenzitetom  $\lambda = 1.5$ . Budući da su međudolazna vremena procesa s intenzitetom  $\lambda = 3$  kraća, očekujemo više događaja u istom vremenu za proces s većim intenzitetom. Ovo su bile dvije stvari koje smo uočili, budući da veći intenzitet implicira više događaja, a time i kraća međudolazna vremena.

Sada koristeći simulaciju možemo provjeriti svojstva Poissonovog procesa. Najzanimljivije za provjeru je treće svojstvo propozicije 2.1.4. Ono kaže da  $N(s + t) - N(s)$  slijedi Poissonovu slučajnu varijablu s parametrom  $\lambda t$ . Za  $s = 0$  imamo



Slika 5.1: U procesu s  $\lambda = 3$  smo uočili 28 događaja dok u drugom procesu s  $\lambda = 1.5$  smo primili 19. Zaista vremena međudolazaka (horizontalni segmenti) su bitno kraća u prvom nego u drugom procesu. Kod naveden u B.0.1

$$\mathbb{P}(N(t) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \tag{5.2}$$

odnosno

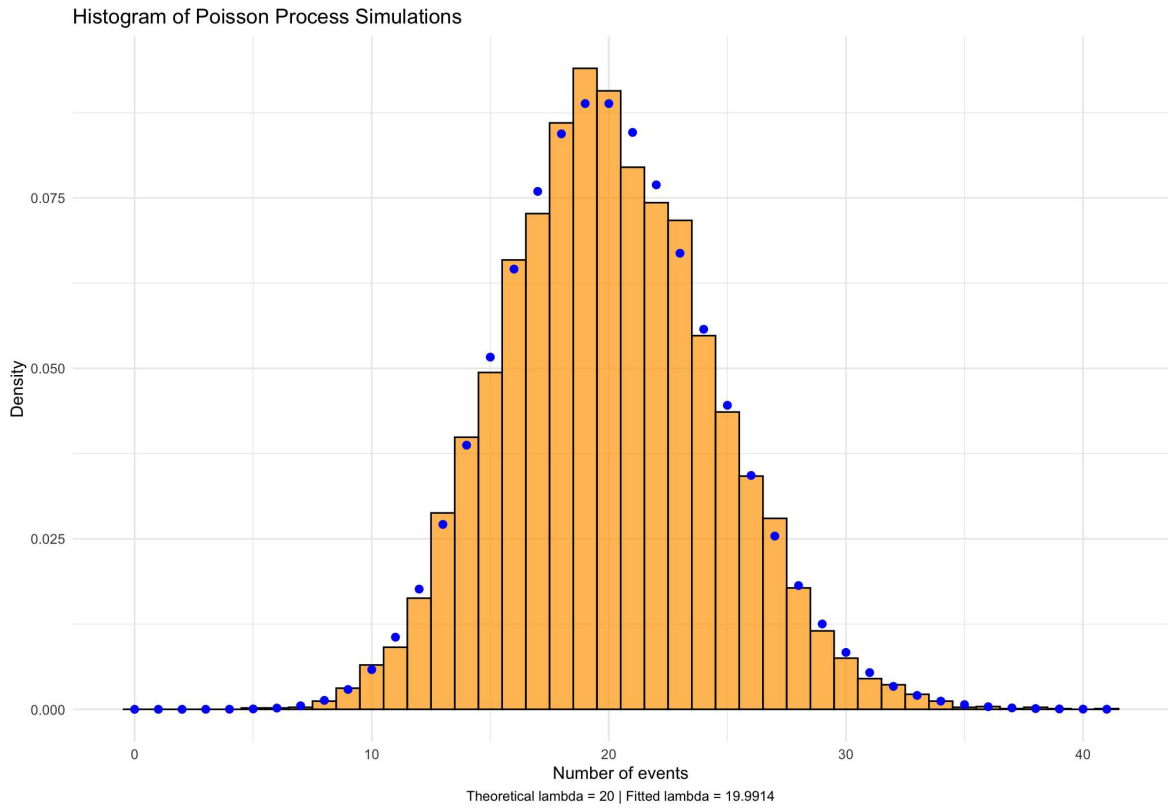
$$N(t) \sim \text{Poi}ss(\lambda t). \tag{5.3}$$

Da bismo prpovjerali gornju tvrdnju, Poissonov proces s intenzitetom  $\lambda = 2$  simuliramo 10.000 puta na intervalu  $[0, 10]$  (R kod priložen u B.0.2). Za svaku simulaciju bilježimo broj događaja koji su se dogodili na kraju simulacije. Histogram dobivenog uzorka je prikazan na slici 5.1.

Prema teoremu, trebali bismo pronaći da  $N(T)$  slijedi Poissonovu distribuciju s parametrom  $\lambda \cdot T = 2 \cdot 10 = 20$ , odnosno da je  $N(10) \sim \text{Poi}ss(\lambda \cdot 10)$ , što se čini da zaista je slučaj.

## 5.2 Nehomogeni Poissonov proces i *thinning* algoritam

Pretpostavimo da želimo simulirati Poissonov proces s intenzitetom  $\lambda(t)$  na intervalu  $[0, T]$ . Simulacija bi nam tada trebala dati niz vremena događaja  $\{T_1, \dots, T_n\}$  svih  $n$  događaja koji su se dogodili između 0 i  $T$ . Isto kao kod homogenog, generirat ćemo zapravo vremena međudolazaka  $(\tau_1, \dots, \tau_n)$ . Problem kod nehomogenog procesa je što ta vremena ne slijede neku specifičnu distribuciju kao što su slijedili kod homogenog, jer u svakom trenutku ovisi o intenzitetu  $\lambda(t)$  koji ne mora biti isti cijelo vrijeme.



Slika 5.2: Poissonov proces s parametrom  $\lambda = 4$  simuliran je 10.000 puta na intervalu  $[0, 10]$ . Histogram u kojem je prikazan broj događaja  $N(10)$  koji su se dogodili na kraju svake simulacije, plave točke odgovaraju Poissonovoj distribuciji. Kod naveden u B.0.2.

**Teorem 5.2.1.** *Razmotrimo homogeni Poissonov proces  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_n$  s intenzitetom  $\tilde{\lambda}$  u intervalu  $\langle 0, T \rangle$ . Pretpostavimo da za  $0 \leq t \leq T$  vrijedi  $0 \leq \lambda(t) \leq \tilde{\lambda}$ . Za  $k = 1, 2, \dots, n$ , uklanjamo točku  $\tilde{T}_k$  s vjerojatnošću  $1 - \lambda(\tilde{T}_k)/\tilde{\lambda}$ ; tada preostale točke formiraju nehomogeni Poissonov proces na  $[0, T]$  s funkcijom inteziteta  $\lambda$ .*

*Dokaz.* Pogledati [12] teorem 4.2. □

Sada pomoću ovog teorema dobivamo algoritam simuliranja nehomogenog Poissonovog procesa:

1. Odaberemo konstantu  $\tilde{\lambda}$  tako da  $\lambda(t) \leq \tilde{\lambda}$  za sve  $t$  u  $[0, T]$ .
2. Simuliramo homogeni Poissonov proces s intenzitetom  $\tilde{\lambda}$  na intervalu  $[0, T]$ .

3. Za svaki generirani događaj  $T_i$ , generiramo uniformnu slučajnu varijablu  $U$  na intervalu  $\langle 0, 1 \rangle$ . Ako  $U \leq \frac{\lambda(T_i)}{\tilde{\lambda}}$ , prihvaćamo  $T_i$  kao događaj nehomogenog Poissonovog procesa. Inače, odbacujemo  $T_i$ .

---

**Algorithm 2** Simuliranje nehomogenog Poissonovog procesa
 

---

**Require:**  $T, \lambda(t)$

**Ensure:**  $\{T_1, \dots, T_n\}$

```

1: Inicijalizirati  $n = m = 0, T_0 = S_0, \tilde{\lambda} = \sup_{0 \leq t \leq T} \lambda(t)$ ;
2: while  $S_m < T$  do
3:   Generiraj  $u \sim \text{Unif}(0, 1)$ ;
4:   Inicijaliziraj  $\tau = -\ln(u)/\tilde{\lambda}$ ;
5:   Stavi  $S_{m+1} = S_m + \tau$ ;
6:   Generiraj  $D \sim \text{Unif}(0, 1)$ ;
7:   if  $D \leq \lambda(S_{m+1})/\tilde{\lambda}$  then
8:      $T_{n+1} = S_{m+1}$ ;
9:      $n = n + 1$ ;
10:  end if
11:   $m = m + 1$ ;
12: end while
13: if  $T_n \leq T$  then
14:  return  $\{T_1, \dots, T_n\}$ ;
15: else
16:  return  $\{T_1, \dots, T_{n-1}\}$ ;
17: end if

```

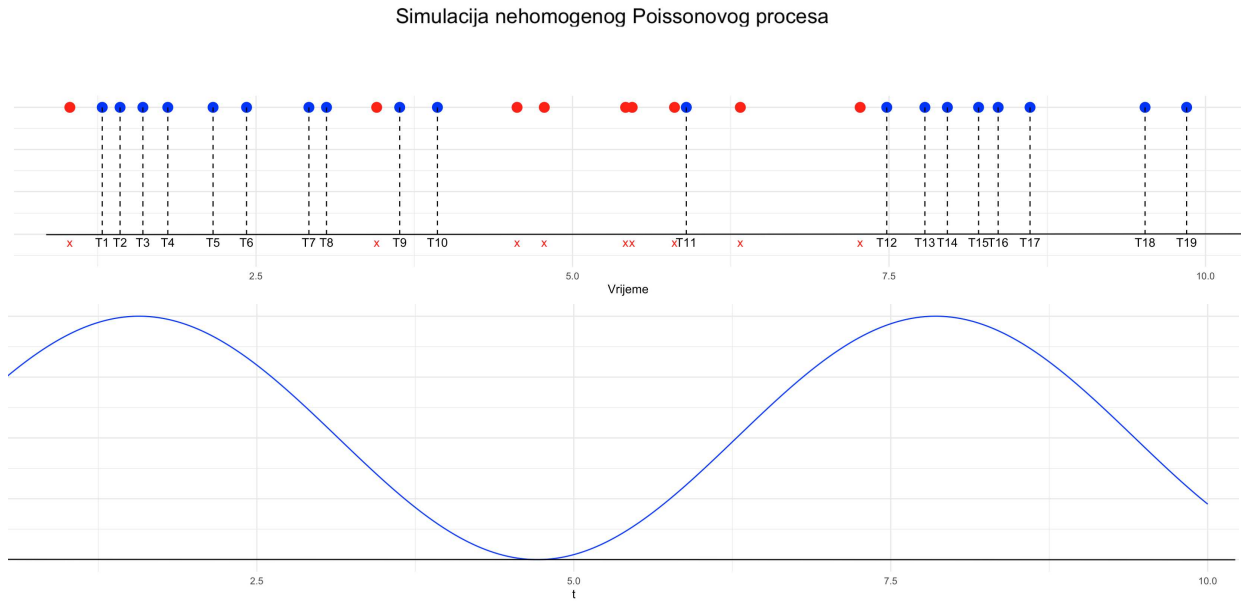
---

*Detaljnije objašnjenje algoritma 2:*

Za simulaciju nehomogenog Poissonovog procesa, trebamo dati kao ulaznu vrijednost funkciju intenziteta  $\lambda(t)$  i krajnje vrijeme  $T$  intervala preko kojeg simuliramo proces. Za inicijalizaciju, trebamo pogledati tri različite stvari: brojače za broj događaja, trenutni razmotreni događaj i vrijednosti  $\tilde{\lambda}$ , koja će predstaviti intenzitet pomoćnog homogenog Poissonovog procesa. Za broj događaja i trenutnog događaja trebamo specifične varijable za homogeni ( $n$  i  $S_0$ ) i nehomogeni ( $m$  i  $T_0$ ) Poissonov proces. Oba brojača počinju od 0. Također,  $S_0 = T_0 = 0$ , budući da se nijedan događaj nije dogodio u vremenu  $t = 0$ . Prema teoremu 5.2.1 trebamo imati  $\tilde{\lambda} \geq \lambda(t)$  za sve vrijednosti  $t$  preko kojih simuliramo, pa biramo  $\tilde{\lambda} = \sup_{0 \leq t \leq T} \lambda(t)$ . Ovi koraci su svi u 1. retku.

Sada smo u mogućnosti simulirati nehomogeni Poissonov proces. Nastavljamo generirati nove točke, sve dok  $S_m < T$ , dakle sve dok  $m$ -ti događaj homogenog Poissonovog procesa nije veći od krajnje točke našeg simulacijskog intervala (2. redak). Vidjeli smo gore





Slika 5.3: Poissonov nehomogeni proces (*gornja slika*) s funkcijom intenziteta  $\lambda(t) = 2 + \sin(t)$ , gdje je  $t \in [0, T]$ . (*donja*) (crvene točke su odbijane, plave su prihvaćene).

da možemo simulirati novo vrijeme međudolazaka  $\tau$  tako što prvo generiramo slučajnu varijablu  $u \sim \text{Uni}(0, 1)$  i poslije koristimo (5.1) (3. i 4. redak). Novi događaj  $S_{m+1}$  homogenog Poissonovog procesa tada je jednak starom događaju  $S_m$  plus generirano vrijeme međudolazaka  $\tau$  (5. redak). Sada možemo koristiti teorem 5.2.1 kako bismo provjerili odgovara li događaj homogenog Poissonovog procesa i događaju u nehomogenom Poissonovom procesu. Da bismo to učinili, prvo generiramo slučajnu varijablu  $D \sim \text{Uni}(0, 1)$  (6. redak), nakon čega provjeravamo je li  $D \leq \lambda(S_{m+1})/\tilde{\lambda}$  (7. redak). Ako je zadnji uvjet točan, novogenerirani događaj također pripada nehomogenom Poissonovom procesu, dakle  $T_{n+1} = S_{m+1}$  (8. redak), pa također moramo ažurirati naš brojač (9. redak). Ako je  $D > \lambda(S_{m+1})/\tilde{\lambda}$ , jednostavno podižemo naš brojač homogenog Poissonovog procesa (11. redak) i ponovo započinjemo korake. Međutim, ako sada  $S_m \geq T$ , simulacija je završena. Jedina stvar koju trebamo provjeriti je je li posljednje pronađeno vrijeme događaja nehomogenog Poissonovog procesa  $T_n$  veće od  $T$  (13. redak). Ako je  $T_n \leq T$ , jednostavno vraćamo  $\{T_1, \dots, T_n\}$  (14. redak). Ako je  $T_n > T$ , vraćamo  $\{T_1, \dots, T_{n-1}\}$  (16. redak). Ovaj algoritam je uzet iz [11] (algoritam 2).

Na temelju algoritma 2, napisan je R kod za simulaciju nehomogenog Poissonovog procesa. Kod imate u dodatku B.0.3.

Na slici 5.2 može se vidjeti simulacija nehomogenog Poissonovog procesa s intenzi-

tetom  $\lambda(t) = 2 + \sin(t)$ . Iz grafa možemo vidjeti da se u simulacijskom intervalu  $[0, T]$  dogodilo 19 događaja (28 događaja u homogenom, od kojeg smo odbacili 9 da generiramo nehomogeni proces).

Na kraju napomenimo da ova metoda odbacivanja točaka nije optimalna jer generiramo više točaka nego što ih poslije koristimo. Mi ćemo ju koristiti za simuliranje Hawkesov proces, no postoji druga tehnika koju su predstavili Dassios i Zhao [13] i bolje je optimizirana za takve procese.

### 5.3 Hawkesovi procesi

Ima puno različitih načina kako možemo simulirati Hawkesov proces, mi ćemo detaljno proći kroz jedan koji upotrebljava prethodno promatrani *thinning* algoritam. Prvi put je algoritam predložio Ogata [14] 1981. godine.

Podsjećamo da su Hawkesovi procesi tip točkovnih procesa koji se koriste za modeliranje događaja kod kojih svaki događaj može uzrokovati dodatne događaje u budućnosti (vjerojatnije u bližoj nego u dalekoj budućnosti).

Funkcija intenziteta  $\lambda^*(t)$  je oblika

$$\lambda^*(t) = \lambda_0 + \sum_{i: t < T_i} \mu(t - T_i) \quad (5.4)$$

gdje je  $\lambda_0$  osnovni intenzitet, a  $\mu(\cdot)$  funkcija jezgre (*memory kernel*, vidi definiciju 4.0.1 za podsjetnik).

Teorija iza simulacije Hawkesovih procesa uključuje generiranje imigranata i njihovih potomaka koristeći jezgru  $\mu$ . Najčešće korištena funkcija je eksponencijalna jezgra:

$$\mu(t) = \alpha e^{-\beta t} \quad (5.5)$$

gdje su  $\alpha$  i  $\beta$  parametri procesa, ali također je popularna i power-law jezgra:

$$\mu(t) = K(c + t)^{-p}, \quad (5.6)$$

gdje su  $K$ ,  $c$  i  $p$  konstante. Tu ćemo koristiti za našu simulaciju zbog pogodnosti s modeliranjem društvenih mreža. U Dodacima je prikaz obje funkcije (B.0.4 i B.0.5).

Napokon predstavimo Ogatin algoritam za simulaciju Hawkesovih procesa, koji se sastoji od sljedećih koraka:

1. Inicijalizacija početnih uvjeta
2. Postavljanje gornje ograde na intezitet  $\lambda^*$ .

3. Generiranje vremena međudolazaka za sljedeći događaj.
4. Thinning algoritam: prihvatanje ili odbijanje.
5. Ponavljanje koraka 2-4 dok se ne dosegne željeni vremenski interval.

Idemo sada izbliza proći kroz algoritam 3 (preuzet iz [9], algoritam 1). Za bilo koji ograničeni  $\lambda^*(t)$  možemo naći konstantu  $\bar{\lambda}$  tako da je  $\lambda^*(t) \leq \bar{\lambda}$  u zadanom vremenskom intervalu (3. redak). Konkretno, vidimo da je za Hawkesove procese, s pretpostavkom da je jezgra  $\mu(t)$  monotono padajuća, između dva uzastopna vremena događaja  $[T_i, T_{i+1})$  upravo  $\lambda^*(T_i)$  gornja međa intenziteta događaja.

---

**Algorithm 3** Ogatin algoritam simuliranja Hawkesovog procesa koristeći thinning

---

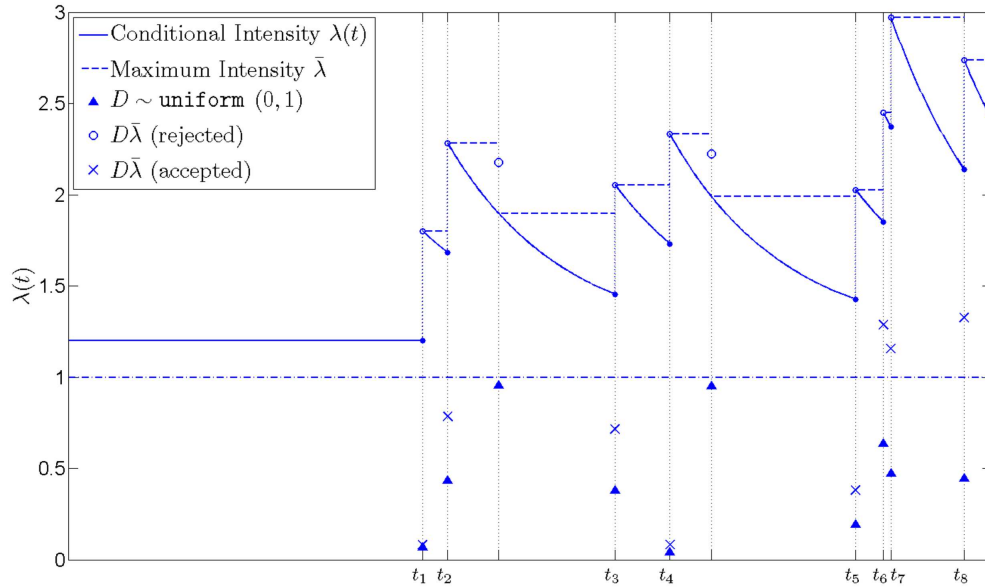
**Require:** vrijeme  $N$  i  $\lambda^*(t)$  za Hawkesov proces kao u (5.4)

**Ensure:**  $\{T_1, \dots, T_n\}$

- 1: Inicijalizirati vrijeme  $T = 0$  i brojač  $i = 1$ ;
  - 2: **while**  $i < N$  **do**
  - 3:     Postavi gornju među intenziteta  $\bar{\lambda} = \lambda^*(T)$  (koristeći (5.4));
  - 4:     Generiraj vrijeme međudolazaka  $\tau$ :  $u \sim \text{Unif}(0, 1)$  i  $\tau = -\ln(u)/\bar{\lambda}$ ;
  - 5:     Ažuriraj vrijeme:  $T = T + \tau$ ;
  - 6:     Generiraj  $D \sim \text{Unif}(0, 1)$ ;
  - 7:     **if**  $D \leq \lambda^*(T)/\bar{\lambda}$  **then**
  - 8:         Prihvati primjerak:  $T_i = T$  i  $i = i + 1$ ; ▷ *inače odbij*
  - 9:     **end if**
  - 10: **end while**
- 

Algoritam je isti kao i za nehomogeni Poissonov proces uz konstantno ažuriranje gornje međe  $\bar{\lambda}$  jer se funkcija inteziteta mijenja svaki put kad dobijemo događaj.

Napomenimo da čak i ako je vrijeme međudolazaka odbijeno, brojač vremena  $T$  se i dalje ažurira. U nastavku pokazujemo graf gdje su prikazani svi parametri i jasno vidimo utjecaj događaja na uvjetnu funkciju intenziteta  $\lambda^*(\cdot)$ , te kako pomoću nehomogenog Poissonovog procesa možemo simulirati Hawkesov proces.



Slika 5.4: Primjer simulacije nehomogenog Hawkesovog procesa s eksponencijalnom jezgrom  $\mu(t) = \alpha e^{-\beta t}$  gdje su  $\alpha = 0.6$ ,  $\beta = 0.8$  i osnovna intenzivnost  $\mu_0 = 1.2$ , koristeći algoritam 3. Sve točke koji su kandidati za prihvaćanje su označene na x-osi među kojima su one prihvaćene kao točke za Hawkesov proces označene sa  $t_k$  gdje  $k = 1, 2, \dots, 8$ . Uvjetna funkcija intenziteta  $\lambda(t)$  je prikazana kao puna krivulja. Konstantna  $\bar{\lambda}$  korištena u svakoj iteraciji unutar “while” petlje na liniji 2 algoritma 3 prikazana je isprekidanim linijama. Na svakoj isprekidanoj liniji, okomitoj na x-osi, su nacrtane dvije točke: donje su označene trokutima i predstavljaju vrijednosti uniformne slučajne varijable  $D$  (raspon od 0 do 1) generirane na liniji 6 algoritma; gornje predstavljaju vrijednosti  $D\bar{\lambda}$  (raspon od 0 do  $\bar{\lambda}$ ) i označene su križićima za one prihvaćene (kada  $D\bar{\lambda} \leq \lambda(\cdot)$ ) ili kružićima za one odbijene. Slika je preuzeta iz [12].

## **Dio III**

# **Statističko modeliranje Hawkesovim procesima**

## Poglavlje 6

# Prilagodba parametara Hawkesovih procesa

Jedan važan dio modeliranja koristeći Hawkesove procese je prilagodba njegovih parametara, odnosno izabrati pogodan odabir kako bi proces najbolje odgovarao podacima koje se pokušava modelirati.

Za odabir parametara maksimizirati ćemo takozvanu funkciju vjerodostojnosti<sup>1</sup>. U konačnici, ako promatramo Hawkesov proces s eksponencijalnim jezgrom, tada obično moramo definirati konstantu  $\lambda_0$  (pozadinski intenzitet) koji smo definirali u (5.4) te parametre  $\alpha$  i  $\beta$  iz (5.5). U tu svrhu trebamo također imati promatrane rezultate modela koji pokušamo definirati.

### Funkcija vjerodostojnosti za Hawkesov proces

Prafrazirajući [3], vjerojatnosna funkcija gustoće i funkcija vjerodostojnosti su zapravo dvije strane iste kovanice. Opisujemo proces preko  $\theta$ , koji predstavlja skup potrebnih parametara preko kojih je proces definiran. U slučaju Hawkesovog procesa s eksponencijalnim jezgrom:  $\theta = (\lambda_0, \alpha, \beta)$ . Dok funkcija gustoće  $f(t; \theta)$  pretpostavlja da je model, specificiran s  $\theta$ , fiksiran i izračunava "vjerojatnost" opažanja  $t$ , ipak ako se interpretacija ove funkcije preokrene tako da je opažanje  $t$  fiksirano, a parametar modela  $\theta$  varira, tada se naziva funkcija vjerodostojnost i označava se s  $L$ :

$$L(\theta; t) = f(t; \theta). \quad (6.1)$$

Obično je logaritam ove funkcije lakše maksimizirati, što označavamo s

$$l(\theta; t) = \ln(L(\theta; t)). \quad (6.2)$$

---

<sup>1</sup>eng. *likelihood*

Radi jednostavnosti, nećemo stalno pisati  $t$  nego  $L(\theta) = L(\theta; t)$ ,  $l(\theta) = l(\theta; t)$ ,

Neka je  $\{t_1, t_2, \dots, t_n\}$  realizacija točkovnog procesa na  $[0, T]$ , s uvjetnim intenzitetom  $\lambda^*$ , tada vrijedi da je funkcija vjerodostojnosti nad parametrima  $\theta$  oblika

$$L(\theta) = \prod_{i=1}^n \lambda^*(t_i) e^{-\int_0^T \lambda^*(t) dt}. \quad (6.3)$$

Za put do izraza, možete pogledati [9] (str. 13).

Za procjenu parametara se koristi MLE metoda (najčešće je maksimizacijski problem težak, pa se koriste numerički algoritmi). Glavna ideja je maksimizirati funkciju vjerodostojnosti (6.3) s obzirom na  $\theta$  nad prostorom  $\Theta$ , odnosno naći  $\hat{\theta} = \arg \max_{\theta \in \Theta} l(\theta)$ . Ideja u pozadini kod logaritmiranja je činjenica da je jeftinije maksimizirati sumu nego produkt. Na kraju dobijemo pogodniji izraz za maksimiziranje:

$$l(\theta) = \log L(\theta) = - \int_0^T \lambda^*(t) dt + \sum_{i=1}^{N(T)} \log \lambda^*(T_i). \quad (6.4)$$

Negativna log-vjerodostojnost može biti minimizirana s optimizacijskim paketima za nelinearne objekte. Više detalja može se naći u [15].

## Poglavlje 7

# Primjer modeliranja Hawkesovog procesa na društvenim mrežama

Postoji mnogo različitih područja gdje se primjenjuju Hawkesovi procesi. Kratko navodimo nekoliko primjera.

- *Seizmološki modeli.* Alan Hawkes je 70tih uveo ovaj model u svrhu proučavanja seizmološke vibracije. Pojava potresa u regiji izaziva jedan veliki niz drugih manje jakih potresa u kratkom vremenu i eksponencijalno se smanjuje kroz vrijeme. Ove su procese dalje proučavali u seizmološkom kontekstu između ostalih Y. Ogata [14] i L. Adamopoulos [16].
- *Financijski modeli.* Kad promatramo utjecaj kupoprodaja na tržištu, jedan događaj (bilo kupnje bilo prodaje) može izazvati jedan niz posljedica na tržištu. U [17] je objašnjeno kako kad trgovac želi trgovati velikom narudžbom, svoje namjere ne želi otkrivati tržištima, kako cijena ne bi “krenula protiv njega”. Opasnost se temelji na činjenici da kada se roba kupi, cijena će porasti; a kad se proda, cijena će pasti. Ioane Muni Toke u [18] pokazuje da pomoću LOB (*Limit Order Book*) možemo uskladiti povoljan Hawkesov model za različite strategije.
- *Društvene mreže.* 2015. godine MacKinlay u svom radu [19] analizira koji su uzroci uspjeha u pojedinim slučajevima na društvenim mrežama. Zaključuje da viralna aktivnost<sup>1</sup> može biti djelomično *egzogena*, potaknuta utjecajima vanjskog svijeta, a djelomično *endogena*, potaknuta njihovom vlastitom prošlošću. Hawkesovi procesi mogu biti pogodni za modeliranje njih.

---

<sup>1</sup>U kontekstu društvenih mreža “viralna aktivnost” opisuje svojstvo proširenja koje ima slika, video, tweet, poruka itd. u medijima. eng. *virality*



Sada ćemo detaljnije promatrati jednu primjenu Hawkesovog procesa upravo iz društvenih mreža. Primjer je uzet iz [9]. Ideja je preformulirati širenje informacije u jednoj društvenoj mreži kao *self-excitement* točkovni proces.

## 7.1 Označeni Hawkesov proces za širenje informacije

U prethodnim cjelinama smo dali teoretsku podlogu za rad s Hawkesovim procesima. Najprije smo općenito o točkovnim procesima govorili u 1.1 i 1.2, zatim smo prešli u Poissonove procese, homogene i nehomogene (cjeline 2.1 i 2.3) i za kraj smo govorili o uvjetnoj funkciji intenziteta i naravno Hawkesovom procesu (cjeline 3 i 4), a posebno o odabiru funkcije jezgre  $\mu(\cdot)$  koja je ključna za proces. U konačnici smo opisivali metode simuliranja događaja u Hawkesovom procesu (među ostalim) i kako prilagoditi parametre modela (poglavlje III).

Međutim, za naš posebni primjer moramo malo nadograditi Hawkesov model. Kad je u pitanju modeliranje društvenih pojava, uočavamo razlike u difuziji medija u mreži. Konkretnije, kad se radi o tweetovima<sup>2</sup>, ne postiže isti rezultat jedan tweet od poznate ličnosti ili od bilo koje prosječne osobe. Naime ovaj prvi prouzrokuje dakako veći odgovor na mreži nego ovaj drugi. A budući da ćemo mi modelirati reakcije (odgovore, retweetove<sup>3</sup>) na pojedini tweet, ta informacija je dakako bitna.

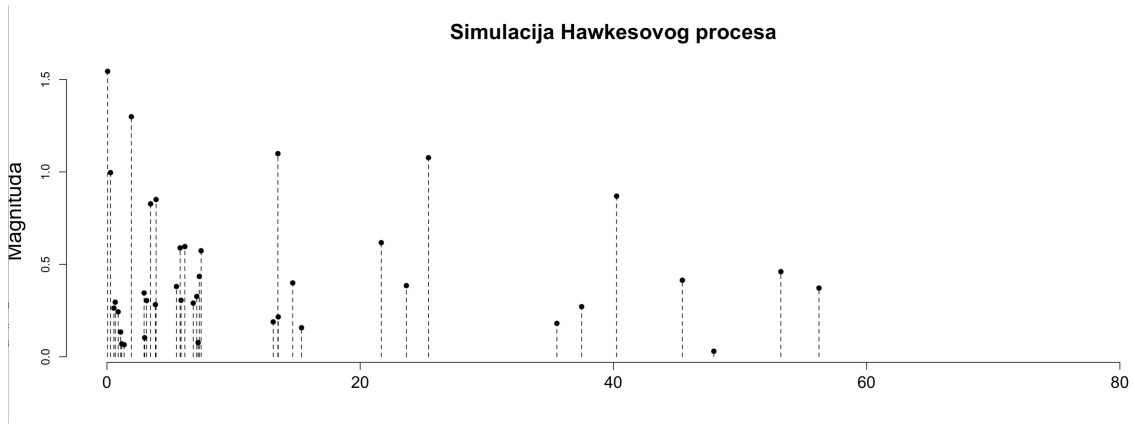
Drugim riječima, moramo malo preoblikovati našu funkciju intenziteta –odgovornu za nastanak novih događaja– tako da uzima u obzir tko šalje ili odgovara, što ćemo matematički modelirati kao *magnitudu utjecaja* (eng. *magnitude of influence*). Pozivajući se na (4.1), imamo

$$\lambda^*(t) = \lambda_0 + \sum_{i:T_i < t} \mu_{m_i}(t - T_i). \quad (7.1)$$

U našem primjeru 7.2, originalni tweet je jedini imigrant u procesu, sve ostale su potomci (retweetovi) tog početnog tweeta. Zato ćemo koristiti  $\lambda_0 = 0, \forall t > 0$ . Štoviše, proces čija funkcija intenziteta uzima u obzir magnitudu utjecaja zove se **označeni Hawkesov proces**. Oznaka  $m$  označava magnitudu događaja s obzirom na utjecaj osobe koja piše. Početni tweet je  $T_0 = 0$  i označen s  $m_0$ . Svaki sljedeći tweet ima oznaku  $m_i$  u događaju  $T_i$  (našem primjeru  $m_i$  biti broj pratitelja koje ta osoba ima na platformi  $X$ ). Jedna realizacija takvog procesa je prikazana na slici 7.1.

<sup>2</sup>U žargonu društvenih mreža, *tweet* znači poruka emitirana u jednoj konkretnoj platformi koja se u početku zvala *Twitter*, a sada za vrijeme pisanja ovog rada je preimenovan u  $X$ .

<sup>3</sup>Retweet: odgovor na tweet.



Slika 7.1: Simulacija označenog Hawkesovog procesa s parametrima  $\kappa = 0.8$ ,  $\beta = 0.6$ ,  $c = 10$ ,  $\theta = 0.8$ , kod dostupan u B.0.6

## 7.2 Procjena parametara

Pokazat ćemo rezultate koristeći obje funkcije jezgre  $\mu_{m_i}(\cdot)$  koje smo definirali u 4, samo preoblikovane za potrebe označenog Hawkesovog procesa.

Eksponecijalna jezgra definirana u (4.2) sada s oznakom  $m$  glasi:

$$\mu_m(\tau) = \kappa m^\beta \theta e^{-\theta\tau}. \quad (7.2)$$

S druge strane, power-law jezgra s oznakom  $m$  glasi:

$$\mu_m(\tau) = \kappa m^\beta (\tau + c)^{-(1+\theta)}. \quad (7.3)$$

Parametri predstavljaju sve što želimo uzimati u obzir za proces.

Parametar	Interpretacija
$\theta$	opisuje koliko brzo događaj bude zaboravljen, odnosno njegovu memoriju. $\theta > 0$ .
$\kappa$	kvaliteta tweeta. Tweetovi visoke kvalitete imaju veću vjerojatnost generiranja više retweetova. $\kappa > 0$ .
$c$	vremenski pomak granične vrijednosti tako da $\phi_m(\tau)$ ostaje ograničen kad je $\tau \simeq 0$ . $c > 0$ .
$m^\beta$	magnituda utjecaja korisnika.

U primjeru (7.3)  $m^\beta$  regulira utjecaj događaja dok je  $(\tau + c)^{-(1+\theta)}$  je zadužen za modeliranje memorije kroz vrijeme.

Ako smo do trenutka  $T$  vidjeli  $n$  retweetova, s pretpostavkom da su  $m_0, m_1, \dots$  nezavisne i jednako distribuirane s gustoćom  $\mathbb{P}(m) = (\alpha - 1) \cdot m^{-\alpha}$ , može se pokazati da je očekivani ukupan broj događaja u kaskadi retweetova jednak

$$N_\infty = n + \frac{\kappa}{1 - n^*} \left( \sum_{i=1}^n \frac{m_i^\beta}{\theta(T + c - t_i)^\theta} \right), \quad (7.4)$$

gdje je

$$n^* = \kappa \frac{\alpha - 1}{\alpha - \beta - 1} \frac{1}{\theta c^\theta}, \quad \text{za } \beta < \alpha - 1 \text{ i } \theta > 0. \quad (7.5)$$

Procjenu smo dobili slično kao u (4.6), uzimajući u obzir i distribuciju od  $m_0$  i (7.1), koristeći funkciju jezgre u (7.3). Uz to još trebamo modelirati društveni utjecaj (Kwak je uveo izraz  $P(m) = (\alpha - 1)m^{-\alpha}$ , gdje je  $\alpha$  dobiven preko velikih količina uzoraka tweetova. [24]). Pogledati [25] za cjelokupni postupak.

## Funkcija vjerodostojnosti

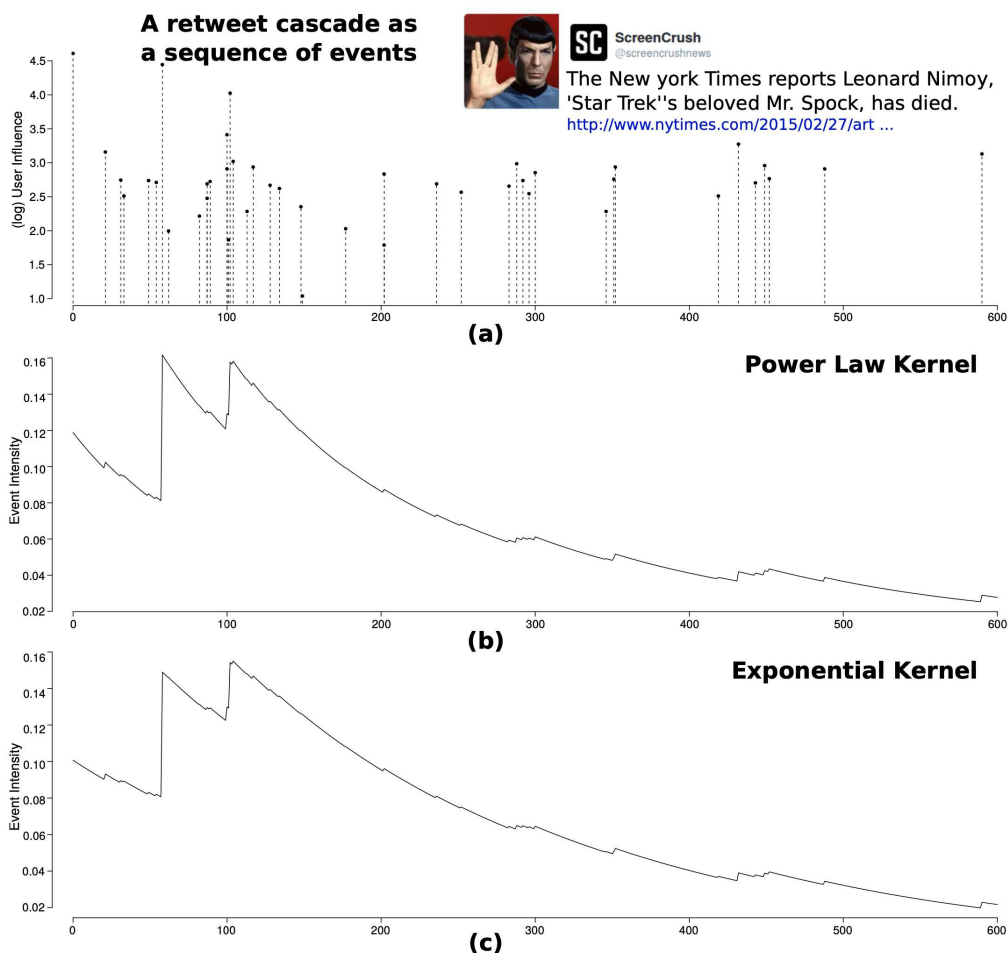
Sada još nedostaje prilagoditi parametre koristeći funkciju vjerodostojnosti koju smo opisali na kraju potpoglavlja 6.

Naš označeni Hawkesov proces ima 4 parametra  $\theta = \{\kappa, \beta, c, \theta\}$ , koje ćemo odrediti koristeći metodu MLE. Uvrstit ćemo (7.3) u opću log-formulu vjerodostojnosti (6.4) i dobit ćemo:

$$\begin{aligned} \mathcal{L}(\kappa, \beta, c, \theta) &= \log \mathbb{P}(\{(m_i, t_i), i = 1, \dots, n\}) \\ &= \sum_{i=1}^n \log(\lambda^*(t_i)) - \int_0^T \lambda(\tau) d\tau \\ &\stackrel{(7.1) \text{ i } (7.3)}{=} \sum_{i=2}^n \log \kappa + \sum_{i=2}^n \log \left( \sum_{t_j < t_i} \frac{(m_j)^\beta}{(t_i - t_j + c)^{1+\theta}} \right) \\ &\quad - \kappa \sum_{i=1}^n (m_i)^\beta \left[ \frac{1}{\theta c^\theta} - \frac{(T + c - t_i)^{-\theta}}{\theta} \right]. \end{aligned} \quad (7.6)$$

Više o tome, pogledati [20], poglavlje 7.2.

Napomenimo još da je (7.6) nelinearna funkcija i za njezino rješavanje koristit ćemo datoteku `Ipop` [22] za jezik R. (Pomoćni kod na B.0.7. Za cijeli projekt: pogledati na github-u <https://github.com/s-mishra/featured-driven-hawkes>)).



Slika 7.2: Primjer kaskade retweetova na jedan članak iz *The New York Times* [?]. (a) Prikaz prvih 600 sekundi kaskade retweetova kao označeni točkovni proces, gdje svakom (re)tweetu odgovara vrijeme događaja. Magnituda odgovara utjecaju pojedinog usera u platformi. U ovom slučaju je određena preko broja pratitelja. (b) Intenzitet događaja  $\lambda(t)$  tijekom vremena, pod pretpostavkom da je točkovni proces Hawkesov proces s power-law jezgrom. Procjene parametara modela metodom maksimalne vjerodostojnosti su  $\{\kappa = 1.00, \beta = 1.01, c = 250.65, \theta = 1.33\}$  s odgovarajućim  $n^* = 0.92$  i predviđenom ukupnom veličinom kaskade od 216. Stvarna veličina kaskade je 219. (c) Intenzitet događaja tijekom vremena za isti proces, ali s eksponencijalnom jezgrom. Parametri prilagođeni za ovu jezgru su  $\{\kappa = 0.0003, \beta = 1.0156, \theta = 0.0054\}$ , odgovarajući  $n^* = 0.997$  i predviđena veličina kaskade je 1603. I primjer i slika su preuzeti iz [9].

### 7.3 Rezultati i zaključci

Na slici 7.2 je prikazana kaskada retweetova nastala nakon imigranta u nultoj sekundi (originalni tweet). Promatrali smo proces u prvih 10 minuta (600 sekunda) i na temelju promatranih podataka smo prilagodili parametre Hawkesovog procesa, i to kako smo pokazali u potpoglavlju 7.2.

Glavna ideja je, nakon što smo ubacili podatke stvarnog procesa, prilagođavati model tako da što bolje opisuje te podatke, u nadi da će tako što bolje predvidjeti ostatak procesa. To postižemo tako što ga, jednom kad smo izračunali parametre na temelju prvih 10 minuta, simuliramo sve dok ne umire. (Znamo da će završiti s vjerojatnošću 1 jer je  $n^* < 1$ ).

Tako smo uzeli prvih 10 minuta, koristeći metodu maksimalne vjerodostojnosti (6.4) smo izračunali da su u slučaju *power-law* jezgre parametri  $\{\kappa = 1.00, \beta = 1.01, c = 250.65, \theta = 1.33\}$  s odgovarajućim  $n^* = 0.92$ . Prema *power-law* jezgri, ovaj tweet ima visoku kvalitetu sadržaja (označeno s  $\kappa$ ) i dugo vrijeme čekanja ( $c$ ), što dovodi do sporog širenja: rezultirajuća kaskada dosegla je 1/4 svoje veličine nakon pola sata, a posljednji tweet je poslan nakon 4 dana [9].

Takav rezultat je malo neobičan za modeliranje s Hawkesovim procesima, ali ima više smisla. U društvenim mrežama odgovori nisu uvijek tako brzi, a vijesti znaju potrajati po nekoliko dana. Zato je *power-law* jezgra prikladnija nego eksponencijalna u ovom slučaju kao što ćemo ubrzo vidjeti.

Koristeći formulu u jednadžbi (7.4), predviđamo očekivanu ukupnu veličinu kaskade  $N_\infty = 216$ , što je vrlo blizu stvarnoj veličini kaskade od 219 tweetova, nakon promatranja samo početnih 10 minuta četverodnevnog širenja na Twitteru. Kada se procjenjuje eksponencijalnom jezgrom, parametri Hawkesovog procesa su  $\{k = 0.0003, \beta = 1.0156, \theta = 0.0054\}$  i odgovarajući faktor grananja je  $n^* = 0.997$ . To proizvodi vrlo neprecizno predviđanje ukupne veličine kaskade od 1603 tweeta, uglavnom zbog visokog  $n^*$ .

**Dio IV**  
**Dodatci**

## Dodatak A

### Neki matematički rezultati

**Primjer A.0.1.** Funkcija gustoće za gamma distribuciju s parametrima  $\alpha, \lambda > 0$  definiramo

$$f(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{za } x \geq 0 \\ 0 & \text{za } x < 0 \end{cases} \quad (\text{A.1})$$

gdje je  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ .

**Primjer A.0.2.** Pokazujemo da vrijedi za bilo koji  $n \in \mathbb{N}$  da je  $\Gamma(n) = (n-1)!$ .

*Dokaz.* Iz definicije  $\Gamma$ , imamo

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1.$$

Zatim, koristeći metodu parcijalne integracije za  $e^{-x}$ ,

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = -x^n e^{-x} \Big|_0^\infty + n \int_0^\infty x^{n-1} e^{-x} dx = n\Gamma(n)$$

jer  $\lim_{x \rightarrow \infty} x^n e^{-x} = 0$ . Integral konvergira za sve pozitivne cijele brojeve  $n$ .

Ovo pokazuje da  $\Gamma(n+1)$  i  $n!$  slijede istu rekurziju i jednaki su za sve  $n$ .

□

**Primjer A.0.3.** Pokazujemo da vrijedi

$$\lim_{h \rightarrow 0} \frac{1 - e^{-\int_t^{t+h} \lambda(u) du}}{h} = \lambda(t).$$

*Dokaz.* Za početak trebamo se prisjetiti analize. Znamo iz definicije da je integral zapravo antiderivacija. Sada iz definicije integrala i neprekidnosti funkcije  $\lambda$  oko  $t$  imamo

$$\lim_{h \rightarrow 0} \frac{\int_t^{t+h} \lambda(u) du}{h} = \lim_{h \rightarrow 0} \frac{F(t+h) - F(t)}{h} = F'(t) = \lambda(t), \quad (\text{A.2})$$

gdje je  $F(x)$  antiderivacija od  $\lambda(x)$ . Rezultat sada slijedi koristeći L'Hospitala:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1 - e^{-\int_t^{t+h} \lambda(u) du}}{h} &\stackrel{L'H}{=} \lim_{h \rightarrow 0} \frac{(\int_t^{t+h} \lambda(u) du)' \cdot e^{-\int_t^{t+h} \lambda(u) du}}{1} \\ &= \lim_{h \rightarrow 0} \left( \int_t^{t+h} \lambda(u) du \right)' \cdot \lim_{h \rightarrow 0} e^{-\int_t^{t+h} \lambda(u) du} \\ &= \lambda(t) \cdot 1 = \lambda(t) \end{aligned} \quad (\text{A.3})$$

□



# Dodatak B

## R kodovi u radu

**Primjer B.0.1.** *Simulacija homogenog Poissonovog procesa s intenzitetom  $\lambda = 3$ :*

```
# Load necessary library
library(ggplot2)
# Parameters
lambda <- 3
T <- 10
num_simulations <- 10000
# Function to simulate a Poisson process using inverse transform sampling
simulate_poisson_process_with_times <- function(lambda, T) {
  times <- numeric()
  events <- 0
  time <- 0
  while (time < T) {
    u <- runif(1)
    tau <- -log(u) / lambda
    time <- time + tau
    if (time < T) {
      events <- events + 1
      times <- c(times, time)
    }
  }
  return(times)
}
# Generate a single Poisson process path with times
single_path_times <- simulate_poisson_process_with_times(lambda, T)
# Create a data frame for plotting
single_path_df <- data.frame(time = c(0, single_path_times),
                             events = c(0, seq_along(single_path_times), length(
                               single_path_times)))
# Plot the single Poisson process path
ggplot(single_path_df, aes(x = time, y = events)) +
  geom_step(color = "blue") +
  labs(title = "Simulacija Poissonovog procesa (lambda = 3)",
        x = "Vrijeme",
        y = "Broj dogadjaja (N(t))") +
  theme_minimal()
```

**Primjer B.0.2.** *Histogram podataka iz 10.000 simulacija (homogenog) Poissonovog procesa za testiranje distribucije:*

```

library(ggplot2)
# Parametri
lambda <- 2
T <- 10
num_simulations <- 10000
expected_lambda <- lambda * T # According to the theorem
# funkcija za simulaciju poissonov proces
simulate_poisson_process <- function(lambda, T) {
  events <- 0
  time <- 0
  while (time < T) {
    u <- runif(1)
    tau <- -log(u) / lambda
    time <- time + tau
    if (time < T) {
      events <- events + 1
    }
  }
  return(events)
}
# vrti simulaciju
set.seed(123) # For reproducibility
event_counts <- numeric(num_simulations)
for (i in 1:num_simulations) {
  event_counts[i] <- simulate_poisson_process(lambda, T)
}
# dodaj Poissonovu distribuciju podacima
fitted_lambda <- mean(event_counts)
# kreiraj histogram
hist_data <- data.frame(events = event_counts)
poisson_prob <- dpois(0:max(hist_data$events), lambda = expected_lambda)
poisson_data <- data.frame(events = 0:max(hist_data$events), probability =
  poisson_prob)
ggplot(hist_data, aes(x = events)) +
  geom_histogram(aes(y = ..density..), binwidth = 1, fill = "orange", color = "
    black", alpha = 0.7) +
  geom_point(data = poisson_data, aes(x = events, y = probability), color = "
    blue", size = 2) +
  labs(title = "Histogram of Poisson Process Simulations",
    x = "Number of events",
    y = "Density",
    caption = paste("Theoretical lambda=", expected_lambda, "| Fitted lambda
    =", round(fitted_lambda, 4))) +
  theme_minimal() +
  theme(plot.caption = element_text(hjust = 0.5))
# Printaj parametar
cat("Fitted Poisson parameter:", fitted_lambda, "\n")

```

**Primjer B.0.3.** *Prikaz nehomogenog Poissonovog procesa na vremenskom pravcu na intervalu  $[0, T]$ :*

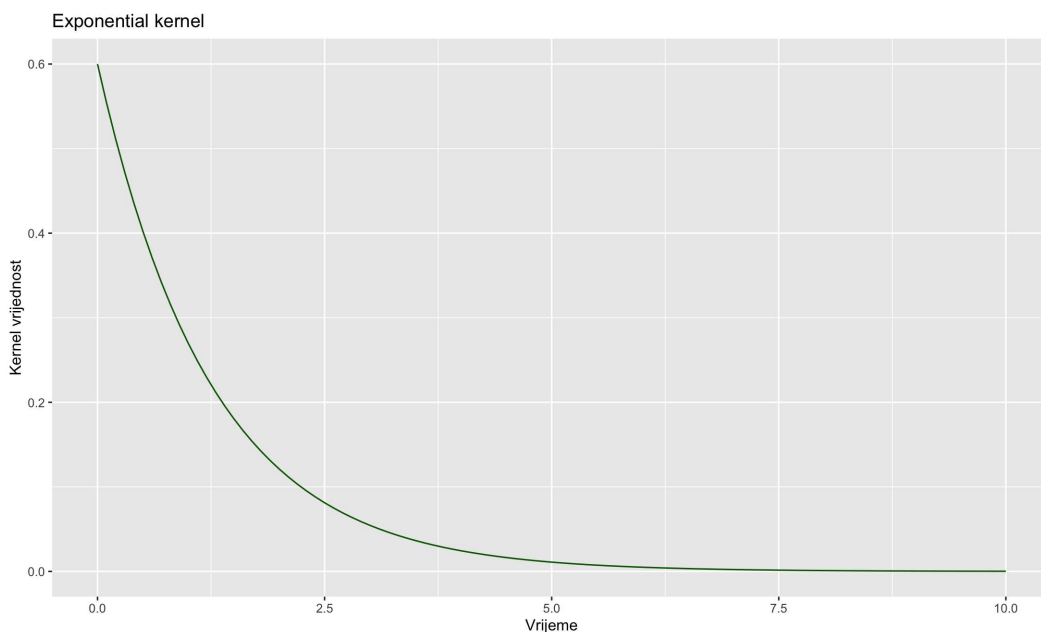
```
# Potrebne knjiznice
library(ggplot2)
# Funkcija za simulaciju nehomogenog Poissonovog procesa
simulate_nhpp <- function(T, lambda_func, lambda_sup) {
  n <- 0
  m <- 0
  S <- c(0)
  T_points <- c()
  rejected_points <- data.frame(time = numeric(), accepted = logical())
  while (S[m + 1] < T) {
    u <- runif(1)
    tau <- -log(u) / lambda_sup
    S <- c(S, S[m + 1] + tau)
    D <- runif(1)
    if (D <= lambda_func(S[m + 2]) / lambda_sup) {
      T_points <- c(T_points, S[m + 2])
      n <- n + 1
      rejected_points <- rbind(rejected_points, data.frame(time = S[m + 2],
        accepted = TRUE))
    } else {
      rejected_points <- rbind(rejected_points, data.frame(time = S[m + 2],
        accepted = FALSE))
    }
    m <- m + 1
  }
  if (T_points[n] <= T) {
    return(list(times = T_points, rejected = rejected_points))
  } else {
    return(list(times = T_points[-n], rejected = rejected_points))
  }
}
# Definiraj parametre
T <- 10
lambda_func <- function(t) { 2 + sin(t) } # Primjer nehomogene funkcije
intenziteta
lambda_sup <- 3 # Gornja granica lambda(t)
# Simuliraj proces
result <- simulate_nhpp(T, lambda_func, lambda_sup)
T_points <- result$times
rejected_points <- result$rejected
# Pripremi podatke za crtanje
accepted_points <- data.frame(time = T_points)
rejected_points <- subset(rejected_points, accepted == FALSE)
# Crtaj rezultate
ggplot() +
  geom_point(data = accepted_points, aes(x = time, y = rep(0.5, length(time))),
    color = "blue", size = 4) +
  geom_segment(data = accepted_points, aes(x = time, xend = time, y = 0.5, yend
    = 0), linetype = "dashed") +
  geom_text(data = accepted_points, aes(x = time, y = -0.05), label = paste0("T"
    , seq_along(T_points)), vjust = -1) +
  geom_point(data = rejected_points, aes(x = time, y = rep(0.5, length(time))),
    color = "red", size = 4) +
```

```
geom_text(data = rejected_points, aes(x = time, y = -0.05), label = "x", color
          = "red", vjust = -1) +
labs(title = "Simulacija nehomogenog Poissonovog procesa",
      x = "Vrijeme",
      y = "") +
theme_minimal() +
theme(axis.title.y = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks.y = element_blank())
```

**Primjer B.0.4.** Prikaz eksponencijalne jezgre:

$$\mu(t) = \alpha e^{-\beta t}, \quad (\text{B.1})$$

s parametrima  $\alpha = 0.6$  i  $\beta = 0.8$ :



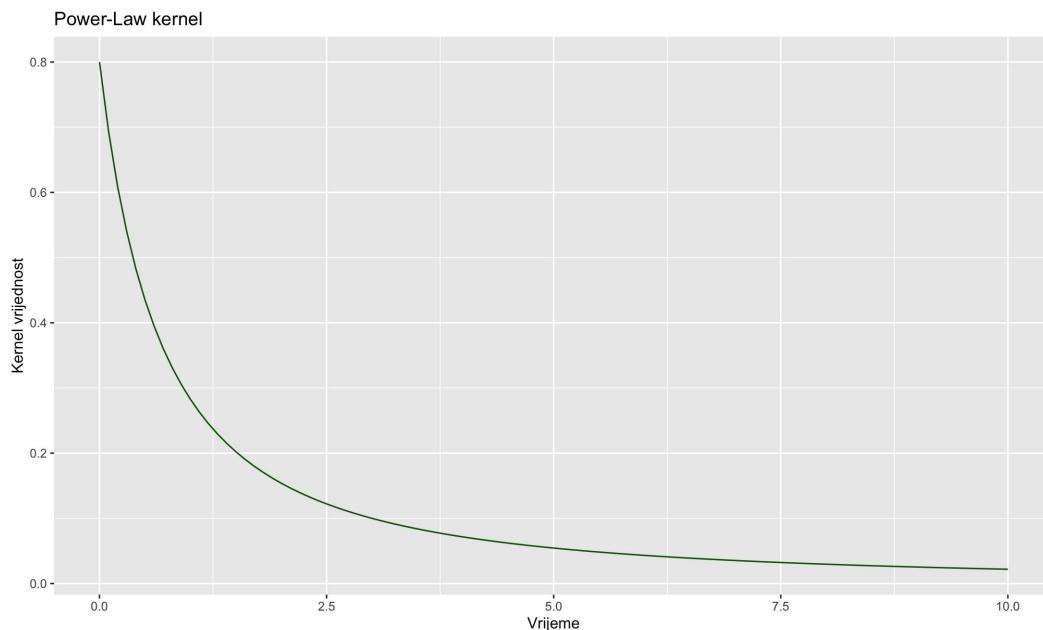
Slika B.1: Eksponencijalna jezgra s parametrima  $\alpha = 0.6$  i  $\beta = 0.8$ .

```
# potrebne libraries
library(ggplot2)
library(dplyr)
# Definiraj eksponencijalnu jezgrinu funkciju kernel function
exp_kernel <- function(t, alpha, beta) {
  alpha * exp(-beta * t)
}
alpha <- 0.6
beta <- 0.8
# Prikaz
t_values <- seq(0, 10, by = 0.1)
kernel_values <- exp_kernel(t_values, alpha, beta)
ggplot(data.frame(t = t_values, kernel = kernel_values), aes(x = t, y = kernel))+
  geom_line(color = "darkgreen") +
  ggtitle("Exponential kernel") +
  xlab("Vrijeme") +
  ylab("Kernel vrijednost")
```

**Primjer B.0.5.** Prikaz power-law jezgre:

$$\mu(t) = K \cdot (c + t)^{-p}, \quad (\text{B.2})$$

s parametrima  $K = 0.8$ ,  $c = 1.0$  i  $p = 1.5$ :



Slika B.2: Eksponencijalna jezgra s parametrima  $\alpha = 0.6$  i  $\beta = 0.8$ .

```
# potrebne datoteke
library(ggplot2)
library(dplyr)
# Definiraj power-law kernel funkcija
power_law_kernel <- function(t, K, c, p) {
  K * (1 + c * t)^(-p)
}
K <- 0.8
c <- 1.0
p <- 1.5
# Prikaz
t_values <- seq(0, 10, by = 0.1)
kernel_values <- power_law_kernel(t_values, K, c, p)
ggplot(data.frame(t = t_values, kernel = kernel_values), aes(x = t, y = kernel)) +
  geom_line(color = "darkgreen") +
  ggtitle("Power-Law kernel") +
  xlab("Vrijeme") +
  ylab("Kernel vrijednost")
```

**Primjer B.0.6.** *Simulacija Hawkesovog procesa (izmijenjena verzija od izvornog koda dostupnog na github-u <https://github.com/s-mishra/featured-driven-hawkes>)*

```
# niz tocaka koji cemo simulirati
t <- seq(from = 0, to = 100, length=1000)
# parametri za jezgru (kernel)
K <- 0.8
beta <- 0.6
c <- 10
theta <- 0.8
# Power Law jezgra
values.PL <- kernelFct(event = event, t = t, K = K, beta = beta, c = c, theta =
  theta, kernel.type='PL')
#simuliramo Hawkesov proces
events1 <- generate_Hawkes_event(K = K, beta = beta, c = c, theta = theta, Tmax
  = 50)
#crtamo ga
pltData <- as.data.frame(events1)
colnames(pltData) <- c("magnitude", "time")
plot(x = pltData$time, y = log(pltData$magnitude, base = 10), type = 'p', col =
  'black', pch=16, bty = "n",
  xaxt = "n", yaxt = "n", xlab = "", ylab = "Magnituda", main = "Simulacija
  Hawkesovog procesa",
  cex.main = 1.8, cex.lab = 1.8, xlim = c(0, 80))
axis(side = 1, cex.axis = 1.4)
segments(x0 = pltData$time, y0 = 0, x1 = pltData$time, y1 = log(pltData$
  magnitude, base = 10), lty = 2)
points(x = pltData$time, y = log(pltData$magnitude, base = 10), type = 'p', col
  = 'black', pch = 16)
axis(side = 2)
```

**Primjer B.0.7.** Pomoćne funkcije za simulaciju Hawkesovog procesa prikazanog u primjeru B.0.6.

```
#' Just to pass the kernel type
kernelFct <- function(event, t, K = 0.024, alpha = 2.016, beta = 0.5, mmin = 1,
  c = 0.001, theta = 0.2, inclusive = T,
  kernel.type='PL') {
  switch(kernel.type,
    PL = .kernelFct.PL(event, t, K = K, alpha = alpha, beta = beta, mmin =
      mmin, c = c, theta = theta, inclusive = inclusive),
    EXP = .kernelFct.EXP(event, t, K = K, alpha = alpha, beta = beta, mmin
      = mmin, theta = theta, inclusive = inclusive))
}
#' Kernel Function for Power Law

.kernelFct.PL <- function(event, t, K = 0.024, alpha = 2.016, beta = 0.5, mmin =
  1, c = 0.001, theta = 0.2, inclusive = T) {
  # the event has 2 components: (magnitude_i, time_i)
  mat_event = matrix(unlist(event), ncol = 2, byrow = F)
  mi = mat_event[,1]
  ti = mat_event[,2]
  # f(p_j) part - virality of a video. Constant for a given video
  fun_f <- K
  # ro(m_i) part - the influence of the user of the event
  fun_ro <- (mi / mmin) ^ beta
  # psi(t, ti) part - the decaying / relaxation kernel
  fun_psi <- 1 / (t - ti + c)^(1+theta)
  val = fun_f * fun_ro * fun_psi
  val[t<ti] = 0
  val[mi<mmin] = 0
  if (!inclusive) {
    val[t == ti] = 0
    val[mi == mmin] = 0
  }
  (val)
}
#' Conditional intensity function, odnosno uvjetna funkcija intenziteta
CIF = function(x, history, ...) {
  subst <- history[history$time <= x,]
  return(sum(kernelFct(event = subst, t = x, ...)))
}
generate_Hawkes_event<- function(K = 0.024, alpha = 2.016, beta = 0.5, mmin = 1,
  c = 0.001, theta = 0.2, M = 10000, Tmax = 10, filename = NULL, kernel.type
  = 'PL') {
  saveInterval = 1000
  history = as.data.frame(matrix(c(M, 0), nrow = 1, byrow = TRUE))
  colnames(history) <- c("magnitude", "time")
  while(history[nrow(history), "time"] <= Tmax){
    t.lambda.max <- history[nrow(history), "time"]
    intensityMax <- CIF(x = t.lambda.max, history = history, K, alpha, beta,
      mmin, c, theta, kernel.type = kernel.type)
    subsCIF <- function(x, ...) CIF(x + history[nrow(history), "time"], ...)
    x <- rnhpoisson(Tmax = NULL, Nmax = 1, LambdaMax = intensityMax, FUN =
      subsCIF, history, K, alpha, beta, mmin, c, theta, kernel.type = kernel.
      type)
    history[nrow(history), "time"] = history[nrow(history), "time"] + x
  }
}
```



```
mag <- generate_user_influence(n = 1, alpha = alpha, mmin = mmin)
# ovdje stvaramo novi event s vremenom koji smo upravo dali (history + x) i
# magnitude
# od mag, poslije priljepimo na history. 3 reda nam treba za to.
event <- matrix( c(mag, history[nrow(history), "time"]), nrow = 1, byrow =
  TRUE)
colnames(event) <- c("magnitude", "time")
history = rbind(history, event)
cat(sprintf("\rCurrent simulation time: %.3f / %.3f (%d events).", history[
  nrow(history), "time"], Tmax, nrow(history)))
if ((nrow(history) %% saveInterval == 0) && (!is.null(filename))) {
  write.table(x = history, file = filename, sep = "\t", row.names = FALSE,
    col.names = TRUE)
}
}
if (!is.null(filename)) {
  write.table(x = history, file = filename, sep = "\t", row.names = FALSE, col
    .names = TRUE)
}
cat(sprintf("\n--> Simulation done!\n"))
return(history)
}
```

# Bibliografija

- [1] Mohler, G., Short, M., Bratingham, P., Schoenberg, F., Tita, G. *Self-exciting Point Process Modeling of Crime*, Journal of the American Statistical Association 106 (493), 2011.
- [2] Bishop, C.M., Nasrabadi, N. M. *Pattern recognition and machine learning*, 2006. (cited on pages 7, 110).
- [3] Laub, Patrick J., Young Lee, and Thomas Taimre. *The elements of Hawkes processes*. New York: Springer, 2021.
- [4] J. F. C. Kingman, *Poisson Processes*, Clarendon Press. 1992. pp. 38–39.
- [5] Laub, P.J.; Taimre, T.; Pollett, P.K. *Hawkes Processes*, 2015.
- [6] BIO 244: Unit 1. *Survival Distributions, Hazard Functions, Cumulative Hazards* <https://web.stanford.edu/lutian/coursepdf/unit1.pdf>.
- [7] Cox DR and Oakes D. *Analysis of Survival Data*, 1984. Chapman and Hall, New York.
- [8] Rick Durrett. *Probability Theory and Examples*, pp. 155-156, 4th edition.
- [9] M.-A. Rizoïu, Young Lee, S. Mishra, L. Xie. *A Tutorial on Hawkes Processes for Events in Social Media*, Preprint arXiv:1708.06401. 2017.
- [10] A.G.Hawkes. *Spectra of some self-exciting and mutually exciting point processes.*, 1971, Biometrika, pp.89-90.
- [11] E.J. van de Winkel. *Parametric Hawkes Processes: Theory, Simulation and Application*, bachelor thesis, Universiteit Utrecht, 2019.
- [12] Y. Chen. *Thinning Algorithm for Simulating Point Processes.*, 2016, <https://www.math.fsu.edu/ychen/research/Thinning%20algorithm.pdf>.

- [13] A. Dassios and H. Zhao. *Exact simulation of Hawkes process with exponentially decaying intensity*, 2013. *Electronic Communications in Probability*, 18: 1–13.
- [14] Y. Ogata. *On Lewis' simulation method for point processes.*, 1981. *IEEE Transactions on Information Theory*, 27(1): 23–31.
- [15] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. 1997. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound- constrained optimization. *ACM Transactions on Mathematical Software* , (TOMS), 23(4): 550–560.
- [16] Adamopoulos, L., Hawkes, *Cluster Models for Earthquakes-regional Comparisons*, Bulletin of the International Statistical Institute, 45 (3), Proc. 39th Session of ISI 454C-461, 1973.
- [17] Muni Toke, I., Pomponio F., *Modeling Trades–Through in a Limit Order Book Using Hawkes Processes*, Economics: The Open-Assessment E-Journal, V. 6, 2012.
- [18] Muni Toke, I. *An Introduction to Hawkes Processes with Application to Finance*, BNP Paribas Chair Meeting, Paris, 2011.
- [19] MacKinlay, D. *Estimating Self-excitation Effects for Social Media Using the Hawkes Process*, Departement Management, Technologie und Oekonomie, Eidgenoessische Technische Hochschule, Zurich, 2015.
- [20] D. J. Daley, D. Vere-Jones. *An introduction to the theory of point processes.*, Vol. I, volume I. 2008.
- [21] D.L.Snyder, M.I.Miller. *Random Point Processes in Time and Space*, 1975, Springer. (6. poglavlje).
- [22] A. Wächter, L. T. Biegler. *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, 2006. *Mathematical Programming*, 106(1): 25–57.
- [23] The New York Times, 27.02.2015. <http://www.nytimes.com/2015/02/27/arts/television/leonard-nimoy-spock-of-star-trek-dies-at-83.html> . ,
- [24] H. Kwak, C. Lee, H. Park, and S. Moon. *What is twitter, a social network or a news media?*, 2010. In WWW '10, pp. 591–600.
- [25] S. Mishra, M.-A. RizoIU, and L. Xie. *Feature Driven and Point Process Approaches for Popularity Prediction*, 2016. F. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM '16. Indianapolis, IN, USA. DOI: 10.1145/2983323.2983812.

[26] Alan G. Hawkes and David Oakes *A Cluster Process Representation of a Self-Exciting Process*, Vol. 11, No. 3 (Sep., 1974), pp. 493-503 (11 pages) Published By: Applied Probability Trust. <https://www.jstor.org/stable/3212693> (zadnji put pogledano: 19.6.2024.)

# Sažetak

Ovaj rad istražuje teoriju i primjenu vremenskih točkovnih procesa, s posebnim naglaskom na Poissonove i Hawkesove procese. Poissonovi procesi omogućuju modeliranje nezavisnih događaja koji se javljaju s konstantnom stopom intenziteta, dok Hawkesovi procesi imaju *self-exciting* svojstvo, gdje svaki događaj povećava stopu budućih događaja.

U prvom dijelu rada razmatraju se osnovni pojmovi i svojstva brojećih procesa, te se uvode homogeni i nehomogeni Poissonovi procesi. Detaljno se analiziraju intenziteti tih procesa i njihova primjena u raznim područjima.

Drugi dio rada posvećen je Hawkesovim procesima. Opisana je njihova struktura, uvjetna funkcija intenziteta i teorija grananja. Na kraju opisujemo simulaciju Hawkesovih procesa koristeći *thinning* algoritam.

Treći dio rada fokusira se na jednu istaknutu primjenu Hawkesovih procesa u modeliranju širenja informacija na društvenim mrežama, s naglaskom na analizu retweet kaskada. Dizajn funkcijske jezgre, procjena parametara i predikcija popularnosti detaljno su opisani.

Rezultati ovog istraživanja doprinose boljem razumijevanju strukture i dinamike događaja kroz vrijeme, te nude značajne uvide za primjenu u različitim znanstvenim i praktičnim područjima.

# Summary

This thesis explores the theory and application of temporal point processes, with a special focus on Poisson and Hawkes processes. Poisson processes allow for the modeling of independent events that occur at a constant rate of intensity, while Hawkes processes introduce the concept of *self-excitement*, where each event increases the rate of future events.

The first part of the thesis examines the basic concepts and properties of counting processes, introducing homogeneous and non-homogeneous Poisson processes. The intensities of these processes and their applications in various fields are analyzed in detail.

The second part of the thesis is dedicated to Hawkes processes. Their structure, conditional intensity function, and branching theory are described. Finally the simulation of Hawkes processes using the *thinning* algorithm is discussed.

The third part of the thesis focuses on one application of Hawkes processes in modeling the spread of information on social networks, with an emphasis on the analysis of retweet cascades, is particularly highlighted. The design of the kernel function, parameter estimation, and popularity prediction are described in detail.

The results of this research contribute to a better understanding of the structure and dynamics of events over time and offer significant insights for application in various scientific and practical fields.

# Životopis

Roden sam 3. siječnja 1999. u Barceloni. Pohađao sam Osnovnu školu Viaro global school u Sant Cugat (grad do Barcelone). Gimnaziju sam upisao i završio u toj istoj školi. Preddiplomski studij Matematika, istraživački smjer, na Prirodoslovno-matematičkom fakultetu u Zagrebu upisujem 2018. godine. Preddiplomski studij završavam 2022. godine, nakon kojeg upisujem diplomski studij Matematika i Računarstvo na istom fakultetu.