MS

(curs 12)

Procese Poisson

Sistemele coadă sau liniile (firele) de așteptare sunt modele conceptuale ale unor sisteme ce constau din entități ce fac coadă pentru a fi servite.

- Exemple de DES=Discrete Events Systems:
  - sistemele de calcul;
  - rețelele de comunicații;
  - fabricația automată,
  - sistemele trafic.
- Exemple de clienți și servere:
  - sistemul PC sau sistemul multiprocesor: clienții sunt procesele și job-urile, iar serverul este procesorul (procesoarele).
  - sistemul de comunicații de date: clienții sunt pachetele de informații, iar server este procesorul nod de rețea.
  - sistemul bază de date: clienctii sunt cererile de tranzacții (inserare, modificare, ștergere), serverul este serverul bază de date

Noi studiem doar fluxul sosirii clienților la coadă, acesta fiind modelat de un proces Poisson.

Procese stochastice

Proces stochastic

- intuitiv: un fenomen aleator care evoluează în timp.
- formal: o familie de variabile aleatoare (X_t), $t \in T$, definite pe același câmp de probabilitate $(\Omega, \mathcal{K}, P)$, cu valori într-o mulțime S ($S = \{1, 2, \ldots, n\}$, $S =$ sau $S \subseteq$), numită spațiul stărilor.

Un proces stochastic în timp continuu (N_t), $t \geq 0$, ce ia valori în mulțimea numerelor naturale se numește **proces de numărare.**

- N_t v.a. ce indică numărul de evenimente rare ce se produc în intervalul de timp $(0, t]$.
- $N_t - N_s$ este numărul de evenimente ce se produc în intervalul $(s, t]$ $s < t$.

Procese Poisson - definiție

Definitie

Un **proces Poisson** este un proces stochastic de numărare (N_t), $t \geq 0$, ce verifică proprietățile:
1) $N_0 = 0$;
2) Evenimentele rare ce se produc în intervale disjuncte sunt independente, adică pentru orice alegere de momente de timp $t_0 < t_1 < \cdots < t_n$, variabilele aleatoare $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, $\ldots$, $N_{t_n} - N_{t_{n-1}}$ sunt independente;
3) Distribuția de probabilitate a variabilelor $N_{t+s} - N_s$, $s \geq 0$, ce dau numărul de evenimente rare (numărul de clienți ce intră în sistemul coadă) ce se produc în intervalul $(s, s+t]$, este

$$P(N_{t+s} - N_s = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

unde $\lambda > 0$ este un parametru fixat, numit **rata procesului Poisson**.

- Din condiția 3): distribuția de probabilitatea a v.a. $N_{t+s} - N_s$ nu depinde decât de lungimea t a intervalului $(s, s+t]$, nu și de extremități.
- intervalul $(0, t]$ are aceeași lungime ca și intervalul $(s, s+t]$, deci

$$P(N_{t+s} - N_s = k) = P(N_{t+s-s} - N_{s-s} = k) = P(N_t = k).$$

- pentru fiecare $t > 0$, variabila aleatoare $N_t \sim Pois(\lambda t)$.
- din condiția 3)

$$Q(h) = P(N_{t+h} - N_t = 1) = \lambda h + \mathfrak{o}(h),$$

unde $\mathfrak{o}(h)$ este o funcție cu proprietatea că $\lim\limits_{h \to 0} \frac{\mathfrak{o}(h)}{h} = 0$
- probabilitatea ca într-un interval de timp h foarte mic să se producă un singur eveniment este proporțională cu $\lambda > 0$.

- Calcul and

$$\lim_{h\to 0} \frac{Q(h)}{h} = \lim_{h \to 0} \frac{\lambda h + \mathfrak{o}(h)}{h} = \lambda,$$

obținem intensitatea producerii evenimentelor rare contorizate de procesul Poisson sau **rata procesului Poisson**, λ.
- Se mai poate arăta că

$$P(N_{t+h} - N_t \geq 2) = \mathfrak{o}(h).$$

Această relație ilustrează că probabilitatea ca în intervalul $(t, t+h]$, de lungime h, să se producă mai mult de o intrare este $0 + \mathfrak{o}(h)$, adică pentru h foarte mic această probabilitate este aproape 0, iar viteza de variație a probabilității $Q(h) = P(N_{t+h} - N_t \geq 2)$ este $\lim\limits_{h\to 0} Q(h)/h = 0$.
- Cu alte cuvinte, nu se produc simultan două evenimente rare (în sistemul coadă nu intră simultan doi sau mai mulți clienți).

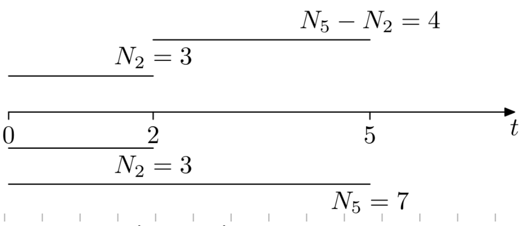
Exemplu

Clienții sosesc la o coadă conform unui proces Poisson cu rata de 6 clienți pe oră. Fie N_t numărul de clienți sosiți până la momentul t, inclusiv. Să se calculeze, exploatând proprietățile procesului Poisson, următoarele probabilități:

a) Probabilitatea ca până la momentul $t = 5$ să sosească 3 clienți, adică $P(N_5 = 3)$;

b) $P(N_2 = 3, N_5 = 7)$;

c) $P(N_5 = 7 | N_2 = 3)$.



Rezolvare: V.a.$N_t \sim Pois(\lambda = 6t)$, deci

$$P(N_t = k) = e^{-6t}\frac{(6t)^k}{k!}, \ k \in .$$

În particular, pentru a) avem

$$P(N_5 = 3) = e^{-6 \cdot 5}\frac{(6\cdot 5)^3}{3!}.$$

b) Evenimentul $(N_2 = 3, N_5 = 7)$ coincide cu evenimentul $(N_2 - N_0 = 3, N_5 - N_2 = 4)$ (vezi Fig.1). Conform propr. 2) din definiție, pentru diviziunea

$$t_0 = 0 < t_1 = 2 < t_2 = 5$$

a intervalului de timp $[0, 5]$ v.a. $N_2 - N_0$, $N_5 - N_2$ sunt independente, deci

$$\begin{aligned} P(N_2 - N_0 = 3, N_5 - N_2 = 4) &= P(N_2 - N_0 = 3)P(N_5 - N_2 = 4) \\ &= P(N_2 = 3)P(N_3 = 4) \\ &= e^{-6\cdot 2}\frac{12^3}{3!}e^{-6\cdot 3}\frac{18^4}{4!}. \end{aligned}$$

Relația proces Poisson - distribuția exponențială

- deoarece nu intră la coadă simultan doi sau mai mulți clienți (nu se produc simultan două evenimente rare), momentele aleatoare ale sosirilor/producerilor evenimentelor rare în intervalul de timp $[0, T]$ sunt distincte, $0 < t_1 < t_2 < \cdots < t_{n-1} < t_n \leq T$
- v.a. N_t- numărul clienților sosiți până la momentul T, ia valorile:

$$\begin{aligned} N_t &= 0, && \text{pentru } t \in [0, t_1), \\ N_t &= 1, && \text{pentru } t \in [t_1, t_2), \\ &\vdots \\ N_t &= n-1, && \text{pentru } t \in [t_{n-1}, t_n). \end{aligned}$$

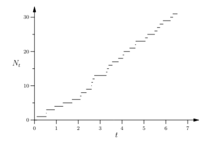


Figure: Vizualizarea procesului Poisson de rată = 5 pe intervalul [0, 7]

- Momentele de timp ale intrării clienților în sistem sunt aleatoare, deci și lungimile intervalelor dintre două sosiri consecutive sunt aleatoare.
- notăm X- variabila aleatoare ce dă lungimea intervalelor inter–sosiri
- Variabila aleatoare inter–sosiri asociată unui proces Poisson de rată λ are distribuția exponențială de parametru $\theta = 1/\lambda$:

$$X \sim \text{Exp}(\theta = 1/\lambda)$$

.

- Timpul mediu între două sosiri consecutive, cu rata sosirilor λ, este

$$M(X) = 1/\lambda,$$

- Notăm X_k v.a. ce dă lungimea intervalului de timp dintre sosirea clientului $k-1$ și k.
- X_k sunt independente și identic distribuite, $X_k \sim \text{Exp}(1/\lambda)$.
- $M(X_k) = 1/\lambda$ și $\sigma^2(X_k) = 1/\lambda^2$.
- $T_n = X_1 + X_2 + \cdots + X_n$ dă momentul sosirii celui de-al n-lea client în sistem.
- Suma a n variabile aleatoare independente și identic distribuite $Exp(\theta)$ este o variabilă aleatoare ce are distribuția de probabilitate n–Erlang,
- deci, momentul intrării în sistem a clientului n, T_n, este o variabilă aleatoare n–Erlang.
- Momentul mediu al sosirii este

$$M(T_n) = M(X_1) + M(X_2) + \cdots + M(X_n) = n/\lambda,$$

- dispersia acestuia este

$$\sigma^2(T_n) = \sigma^2(X_1) + \sigma^2(X_2) + \cdots + \sigma^2(X_n) = n/\lambda^2$$

Spamurile sosesc în inbox-ul adresei `student.absent@yahoo.com` cu o rată $\lambda_s = 2$/oră, iar în mod independent, mailurile uzuale (obișnuite) cu o rată $\lambda_u = 5$/oră.
a) Care este probabilitatea ca `student.absent` să nu primească nici un spam de la ora 12 la ora 15?
b) La ora 8 dimineața, când a accesat inbox-ul, un student nu a găsit nici un mail obișnuit nou. Care este timpul mediu ce trece până la sosirea a 4 mailuri obișnuite?
Rezolvare: a)

- Notăm cu (N_t^u), respectiv (N_t^s) procesul Poisson al intrării email-urilor uzuale, respectiv a spam-urilor în inbox.
- Se cere $P(N_{15}^s - N_{12}^s = 0)$.
- Avem: $P(N_t^s - N_\tau^s = 0) = P(N_{t-\tau}^s = 0)$,
- Deci,

$$P(N_{15}^s - N_{12}^s = 0) = P(N_3^s = 0) = e^{-\lambda_s 3}(\lambda_s 3)^0/0! = e^{-6} \approx 0.0025.$$

b) Fie X_1, X_2, X_3, X_4 variabilele aleatoare inter–sosiri pentru procesul Poisson (N_t^u), al intrării mesajelor uzuale în inbox, de rată $\lambda_u = 5$.

- Variabila aleatoare ce dă momentul sosirii celui de-al patrulea mail uzual (non-spam) este $T_4 = X_1 + X_2 + X_3 + X_4$.
- Deoarece variabilele aleatoare X_1, X_2, X_3, X_4 sunt identic distribuite, $X_k \sim \text{Exp}(1/\lambda_u)$
- $M(T_4) = 4/\lambda_u = 4/5$ ore

Cunoscând distribuția de probabilitate a inter–sosirilor putem genera momentele sosirilor astfel:

$$
\begin{aligned}
t_0 &= 0, \\
t_k &= t_{k-1} + x_k, \; k = 1, 2, \ldots
\end{aligned}
$$

unde x_k sunt observații asupra variabilelor aleatoare $X_k \sim \mathrm{Exp}(1/\lambda)$, $k \in \mathbb{N}^*$.
O observație asupra variabilei aleatoare $X \sim \mathrm{Exp}(\theta)$ se generează astfel:
u= urand();
$x = -\theta * \ln(1 - u)$;

Simulare proces Poisson

Algoritm de simulare pe un interval fixat $[0, T]$ a momentelor producerii evenimentelor contorizate de un proces Poisson (momentele sosirilor clienților în sistemul coadă), de rată λ
SimulProcPoissonλ, T
$k = 0$; $t_0 = 0$;
u=urand();
$\mathbf{x} = -(\mathbf{1}/\lambda) * \ln(\mathbf{1} - \mathbf{u})$;
k=k+1;
$\mathbf{t_k} = \mathbf{t_{k-1}} + \mathbf{x}$;
$(\mathbf{t_k} \leq \mathbf{T})$;
return $t_0, t_1, \ldots$

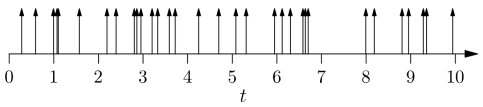


Figure: Simularea intrărilor într-un sistem coadă în intervalul $[0, 10]$, $\lambda = 3$.

Suprapunerea mai multor procese Poisson independente

Date două fluxuri de intrare ale mailurilor, unul cu rata λ_u și unul cu rata λ_s, care este rata fluxului global?

Observație: Dacă X, Y sunt v.a. indep., $X \sim \mathrm{Poiss}(\lambda_1)$, $Y \sim \mathrm{Poiss}(\lambda_2)$, atunci $X + Y \sim \mathrm{Poiss}(\lambda_1 + \lambda_2)$

Proprietate

Dacă $(N_t^1), (N_t^2), \ldots, (N_t^n)$ sunt n procese Poisson independente de rate $\lambda_1, \lambda_2, \ldots, \lambda_n$, atunci procesul $(N_t = N_t^1 + N_t^2 + \cdots + N_t^n)$ este de asemenea un proces Poisson de rată $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Suprapunerea proceselor Poisson poate fi interpretată ca o coadă constituită din combinarea a n cozi.

Intervalul inter–sosirilor în procesul Poisson rezultat din suprapunerea a n procese Poisson independente de rate $\lambda_1, \lambda_2, \ldots, \lambda_n$ este $X \sim \mathrm{Exp}(1/(\lambda_1 + \lambda_2 + \cdots + \lambda_n))$.

Ramificarea unui proces Poisson în n subprocese

Proprietate

Dacă (N_t) este un proces Poisson de rată λ și $p_1, p_2, \ldots, p_n \in (0, 1)$ cu $p_1 + p_2 + \cdots + p_n = 1$, atunci subfluxurile generate sunt procese Poisson independente având ratele $p_1\lambda, p_2\lambda, \ldots, p_n\lambda$.

Observație: Se spune că fluxul principal se ramifică în n subfluxuri. Evident că suprapunerea subfluxurilor dă fluxul principal de rată $= p_1\lambda + p_2\lambda + \cdots + p_n\lambda = (p_1 + p_2 + \cdots + p_n)\lambda$.

Remarcă

O astfel de situație apare în modelarea și simularea rețelelor de cozi, în care un client după ce este servit are mai multe opțiuni: alege la întâmplare alte servere din rețea sau iese din rețea.