Aula 5. Processo de Poisson. Exemplos. (Teórica).

Anatoli Iambartsev
IME-USP

Processo de Poisson com diferentes tipos de eventos.

Consideramos um processo de Poisson com taxa λ. Suponha que em cada instante de ocorrência de evento com probabilidade $p_i, i = 1, \ldots, n, (p_1 + p_2 + \cdots + p_n = 1)$ vamos classificar o evento como o evento do tipo i. Por exemplo, considere que as pessoas que entram numa loja formam um processo de Poisson com intensidade λ. Podemos supor que a pessoa que entra na loja pode ser jovem com probabilidade 10%, pode ser adulto com probabilidade 70% e idoso com probabilidade 20%. Neste caso $p_1 = 0.1, p_2 = 0.7$ e $p_3 = 0.2$, notamos que $p_1 + p_2 + p_3 = 1$. Sejam $N_i(t)$ o número de eventos do tipo i que ocorreram até o tempo t.

Processos $\{N_i(t), t \geq 0\}$ são processos de Poisson com taxas $p_i\lambda$ para cada $i = 1, \ldots, n$. Os processos $N_1(t), N_2(t), \ldots, N_n(t)$ são independentes.

Processo de Poisson com diferentes tipos de eventos.

Ex. 1. Suponha que a chegada dos imigrantes em um país forma um processo de Poisson com taxa $\lambda = 1$ pessoa por um dia. Sabe-se que 40% dos imigrantes são europeus, 40% são africanos e 20% são asiáticos. Qual é o tempo médio até que o décimo imigrante asiático chegue?

Solução. Imigrantes asiáticos formam um processo de Poisson com taxa $\lambda_{as} = \lambda \cdot 0.2 = 0.2$ (pessoa por dia) Seja S_{10} o tempo de espera do décimo imigrante asiático. Temos que $\mathbb{E}[S_{10}] = 10\frac{1}{\lambda_{as}} = 10\frac{1}{0.2} = 50$ dias. $\square$

Distribuição condicional do tempo de chegada (Ross, Cap. 5.3.5).

Sabe-se que exatamente um evento ocorreu durante o tempo $(0, t]$. Qual é a distribuição de tempo da ocorrência deste evento? Devido à estacionariedade e independência dos incrementos, espera-se a distribuição uniforme. Realmente,

$$\begin{aligned} F(s) = \mathbb{P}(T_1 \leq s \mid N(t) = 1) &= \frac{\mathbb{P}(T_1 \leq s, N(t) = 1)}{\mathbb{P}(N(t) = 1)} \\ &= \frac{\mathbb{P}(1 \text{ evento em } [0, s]; 0 \text{ eventos em } (s, t])}{\mathbb{P}(N(t) = 1)} \\ &= \frac{\mathbb{P}(1 \text{ evento em } [0, s]\} P\{0 \text{ eventos em } (s, t]\}}{\mathbb{P}(N(t) = 1)} = \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t} \end{aligned}$$

e a densidade

$$(F(s))' = \left(\frac{s}{t}\right)' = \frac{1}{t}, \quad s \in [0, t].$$

O que prova que a distribuição de tempo de ocorrência condicional seja uniforme no intervalo $[0, t]$.

Distribuição condicional do tempo de chegada (Ross, Cap. 5.3.5).

Este resultado pode ser generalizado, mas para isso, temos que introduzir o conceito de estatísticas de ordem. Sejam $Y_1, Y_2, \dots, Y_n$ variáveis aleatórias. Vamos dizer que $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ são estatísticas de ordem que correspondem a $Y_1, Y_2, \dots, Y_n$, se $Y_{(k)}$ é o k-ésimo menor valor entre $Y_1, Y_2, \dots, Y_n$, $k = 1, 2, \dots, n$. Por exemplo, se $n = 3$ e $Y_1 = 4, Y_2 = 5, Y_3 = 1$, então $Y_{(1)} = 1, Y_{(2)} = 4, Y_{(3)} = 5$. Se $Y_i, i = 1, \dots, n$, são independentes identicamente distribuídas com a densidade f, então a densidade conjunta das estatísticas de ordem $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ é dado pela fórmula

$$f(y_1, y_2, \dots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n.$$

Distribuição condicional do tempo de chegada (Ross, Cap. 5.3.5). A inferência de

$$f(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n.$$

segue o seguinte raciocínio:

(i) $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ é igual à $(y_1, y_2, \ldots, y_n)$ se $(Y_1, Y_2, \ldots, Y_n)$ é igual a qualquer uma das $n!$ permutações de $(y_1, y_2, \ldots, y_n)$;

(ii) a densidade de que $(Y_1, Y_2, \ldots, Y_n)$ é igual à $(y_{i_1}, y_{i_2}, \ldots, y_{i_n})$ é $\prod_{j=1}^{n} f(y_{i_j}) = \prod_{j=1}^{n} f(y_j)$, onde $i_1, \ldots, i_n$ é uma permutação dos índices $1, 2, \ldots, n$.

Se a distribuição de Y_i é uniforme no intervalo $[0, t]$, então a densidade conjunta das estatísticas de ordem é

$$f(y_1, y_2, \ldots, y_n) = \frac{n!}{t^n}, \quad 0 < y_1 < y_2 < \cdots < y_n < t. \tag{1}$$

Distribuição condicional do tempo de chegada (Ross, Cap. 5.3.5). Podemos provar (Teorema 5.2, Ross)

Sabe-se que $N(t) = n$. A distribuição conjunta de n tempos de ocorrências $S_1, \ldots, S_n$ desses eventos é igual a distribuição de estatísticas de ordem para n variáveis independentes e uniformemente distribuídas no intervalo $[0, t]$.

Para obter a densidade conjunta condicional de $S_1, \ldots, S_n$ dado que $N(t) = n$ notaremos que para $0 < s_1 < \cdots < s_n < t$, o evento $\{S_1 = s_1, S_2 = s_2, \ldots, S_n = s_n, N(t) = n\}$ é equivalente ao evento que os $n + 1$ tempos entre ocorrências satisfazem $T_1 = s_1, T_2 = s_2 - s_1, \ldots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n$. Logo,

$$
\begin{aligned}
f(s_1, \ldots, s_n \mid n) &= \frac{f(s_1, \ldots, s_n, n)}{\mathbb{P}(N(t) = n)} \\
&= \frac{\lambda e^{-\lambda s_1} \lambda e^{-\lambda(s_2 - s_1)} \ldots \lambda e^{-\lambda(s_n - s_{n-1})} e^{-\lambda(t - s_n)}}{e^{-\lambda t}(\lambda t)^n / n!} = \frac{n!}{t^n},
\end{aligned}
$$

onde $0 < s_1 < \cdots < s_n < t$.

$\square$

Paradoxo de tempo de espera.

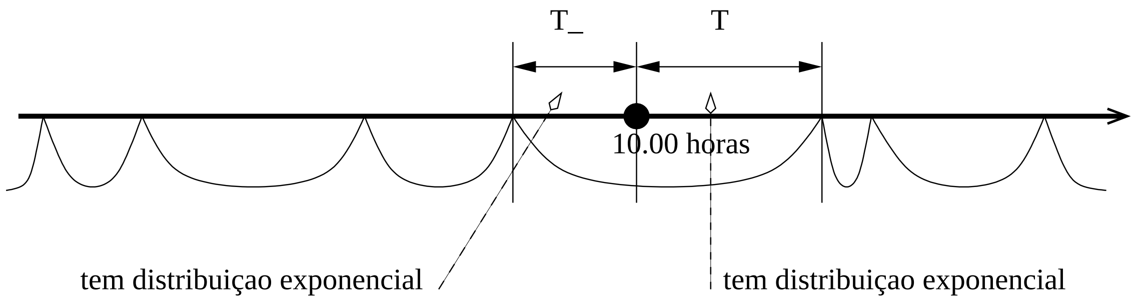
Suponha que as chegadas de ônibus formam um processo de Poisson com intensidade λ ônibus por hora. Suponha que um homem chega às 10:00hs no ponto de ônibus. Seja T o tempo que ele espera o próximo ônibus. Qual é a média do tempo de espera T.

Paradoxo de tempo de espera.

Um raciocínio pode ser o seguinte: O intervalo de tempo entre os ocorrências tem distribuição exponencial com intensidade λ. Logo, o tempo médio de intervalo é $\frac{1}{\lambda}$. Por isso, espera-se que o tempo de espera deveria ser menor do que o tempo médio do intervalo entre as chegadas: $\mathbb{E}(T) < \frac{1}{\lambda}$, mais precisamente, espera-se que a média $\mathbb{E}(T)$ seja igual a $\frac{1}{2\lambda}$, devido à "uniformidade" do momento de chegada do homem num ponto de ônibus dentro do intervalo entre chegadas de ônibus. Mas esse raciocínio está **errado**.

O raciocínio **certo** neste caso é usar a propriedade de distribuição exponencial: no momento quando o homem chega no ponto de ônibus, o intervalo de tempo do último ônibus até o prôximo, que tem a distribuição exponencial, já "esqueceu" quanto tempo passou depois do último ônibus. Por isso, o tempo até o próximo ônibus tem distribuição exponencial com intensidade λ, o que significa que o homem vai esperar em média o tempo $\frac{1}{\lambda}$: $\mathbb{E}(T) = \frac{1}{\lambda}$, mas não $\frac{1}{2\lambda}$.

Paradoxo de tempo de espera.

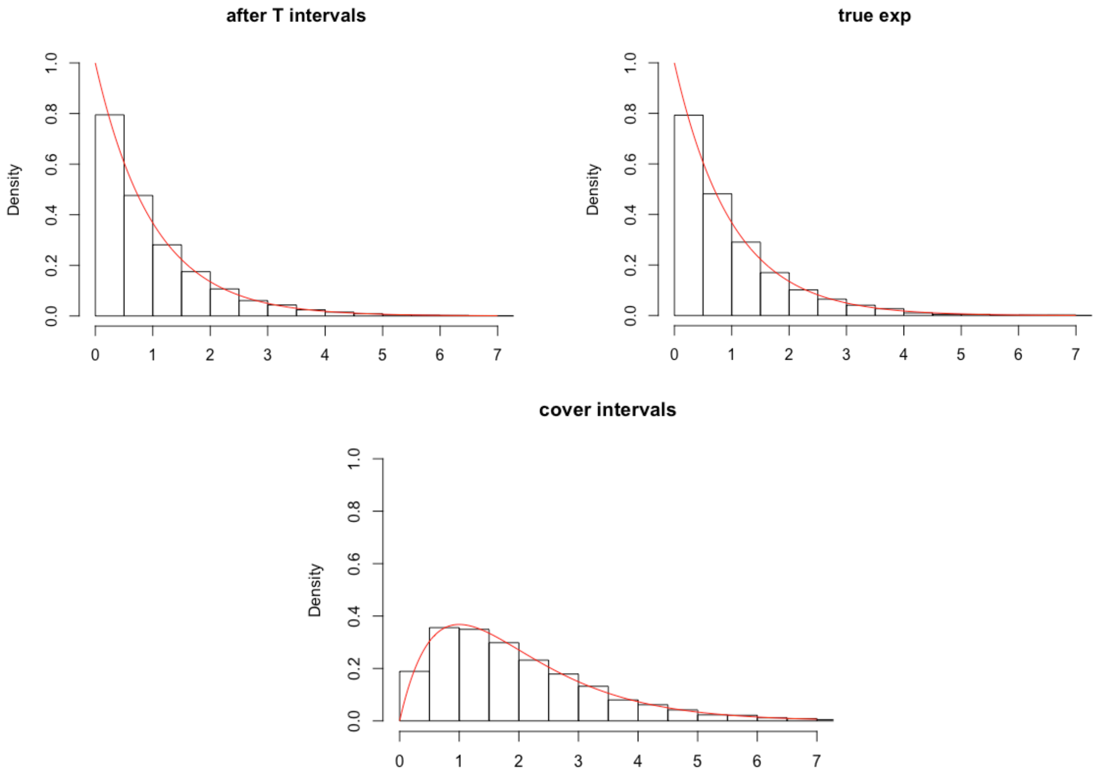


Instante 10 horas cobre-se com "mais chances" com um intervalo "largo" - soma de duas variáveis independentes $T_{-} + T$ e exponencialmente (com a taxa λ) distribuídas. Então, o intervalo que vai cobrir o instante 10hs tem distribuição gamma com os parâmetros 2 e λ. Essa distribuição é mais "larga" de que uma exponencial com a mesma taxa λ.

Paradoxo de tempo de espera.

O mesmo resultado vale para o tempo T_{-} que passou da partida do último ônibus até o momento da chegada do homem no ponto do ônibus (até às 10:00hs). A distribuição de T_{-} é exponencial com a mesma taxa λ. Este fato leva à seguinte conclusão: a distribuição do intervalo entre os ônibus cujo inécio é menor do que 10:00hs e o final é maior do que 10:00hs não tem distribuição exponencial, mas sim, a distribuição gamma com os parâmetros 2 e λ (soma das duas variáveis exponenciais independentes com o mesmo parâmetro λ).

Paradoxo de tempo de espera.



Paradoxo de tempo de espera. Anexo (R code)

```
T=100          # tempo de chegada de passageiro
N=10000        # número de simulações
lambda=1       # taxa de distribuição exponencial
jj<-vector()   # comprimento de intervalo que "cobre" o ponto T
xi<-vector()   # tempo de espera do próximo ônibus

for(i in 1:N){     # start for: ciclo de simulações
  S=rexp(lambda)
    while(S<T){    # start while: simulamos Poisson até ultrapassar T
    jump=rexp(lambda)
    S<-S+jump
  } #end while
  xi[i]=S-T;
  jj[i]=jump;
} #end for
```

Paradoxo de tempo de espera. Anexo (R code)

Visualização de resultados de simulação: histogramas

```
br=seq(0,27,by=0.5)
hist(xi,probability = T,main = "after T intervals",
     xlim = c(0,7),breaks=br,ylim=c(0,1),xlab = " ")
x=seq(0,7,by=0.01)
lines(x,exp(-x),col="red")

hist(jj,probability = T,main = "cover intervals",
     xlim = c(0,7),breaks=br,ylim=c(0,1),xlab = " ")
lines(x,x*exp(-x),col="red")

hist(rexp(N,lambda),probability = T,main = "true exp",
     xlim = c(0,7),breaks=br,ylim=c(0,1),xlab = " ")
lines(x,exp(-x),col="red")
```

Coleção completa (Ross, Exemplo 5.12).

Temos m diferentes tipos de cupons. Em cada instante, uma pessoa escolhe um cupom independentemente dos outros tipos de cupons já escolhidos. Supomos que ela escolhe um cupom do tipo j com probabilidade p_j, em que $\sum_{j=1}^{m} p_j = 1$. Seja N o número de cupons coletados necessários para completar uma coleção que tem pelo menos um cupom de cada tipo. Ache a média de N.

Coleção completa (Ross, Exemplo 5.12). Solução.

Seja N_j o número de tentativas para obter o primeiro cupom do tipo j. Para cada j, o número N_j tem distribuição geométrica com probabilidade de sucesso p_j. O número N pode ser representado agora pela formula

$$N = \max_{1 \le j \le m} N_j.$$

Infelizmente, N_j's não são independentes. Mas o problema pode ser reduzido a achar a esperança de uma variável aleatória que é o máximo de variáveis *independentes* . Para isso, supomos que a pessoa escolhe um cupom em instantes de tempos que formam um processo de Poisson com taxa $\lambda = 1$. O evento que ocorre no processo de contagem classifica-se como evento do tipo j se foi escolhido um cupom do tipo j. Seja $N_j(t)$ o número do cupons escolhidos até o instante t. Sabemos que $N_j(t)$, para qualquer j, forma um processo de Poisson com a taxa $\lambda p_j = p_j$.

Coleção completa (Ross, Exemplo 5.12). Solução.

Para j qualquer processo $N_j(t)$ é processo de Poisson com a taxa $\lambda p_j = p_j$. Seja X_j o tempo da primeira ocorrência do evento do tipo j e seja

$$X = \max_{1 \le j \le m} X_j$$

o tempo quando a coleção de cupons vai estar completa. Os tempos X_j's são independentes e exponencialmente distribuídos com taxas p_j, logo obtemos a distribuição de X:

$$\begin{aligned}\mathbb{P}(X < t) &= \mathbb{P}(\max X_j < t) \\ &= \mathbb{P}(X_j < t,\ \text{para } j = 1, \ldots, m) = \prod_{j=1}^{m} \Big(1 - e^{-p_j t}\Big).\end{aligned}$$

Coleção completa (Ross, Exemplo 5.12). Solução.

Tendo a distribuição de X: $F_X(t) = \prod_{j=1}^{m}\left(1 - e^{-p_j t}\right)$, obtemos a esperança

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)dt = \int_0^\infty \left(1 - \prod_{j=1}^{m}\left(1 - e^{-p_j t}\right)\right)dt.$$

Falta achar a esperança de N. Para isto, basta notar que

$$X = \sum_{i=1}^{N} T_i,$$

em que T_i são os intervalos de tempos entre sucessivas ocorrências de eventos no processo de Poisson com taxa 1. Logo,

$$\mathbb{E}(X) = \mathbb{E}(N)\mathbb{E}(T_1) = \mathbb{E}(N).$$

□

References:

[Ross] S.Ross. *Introduction to Probability Models.* 6th edition, Academic Press, 1997.