Processos de Poisson espaciais e Cox

Yuri Passuelo
PPGE - FEARP/USP
Ribeirão Preto, SP

03/04/2024

Sumário

Introdução e Objetivos

- Introduzir os conceitos basicos do que são processos pontuais, Poisson e de Cox
- Realizar simulações e demonstrações gráficas e vizuais dos principais conceitos desses processos se utilizando do pacote estatístico `spatstats`, do software `R`

Disclaimer

Todos os scripts e códigos utilizados nessa apresentação ficaram disponíveis por meio do seguinte link:
- `github.com/yuripassuelo/estat_aplicada`

Processos pontuais

- Antes de adentrar especificamente dentro dos processos de *Poisson* vamos descrever alguns conceitos básicos de um processo pontual.
- Quando dizemos um processo pontual estamos falando da geração de eventos aleatórios que podem se dar por exemplo em uma ou n dimensões com $n \geq 2$.

Uma dimensão Processos por exemplo que mensuramos no tempo, número de passagens num posto de fronteira por minuto, número de clientes que passam por uma bomba por hora.

Duas Dimensões ou Mais Processos que ocorrem no espaço, por exemplo distribuição do número de arvores numa região . . .

Processos pontuais

- Normalmente os processos pontuais em uma dimensão tem a característica da unica dimensão a ser analisada ser o tempo, e portanto eventos como : Número de ligações em uma central telefônica, número de registros policiais ao longo de um dia
- Sendo assim nossos eventos seriam observados através de uma reta continua, que seria justamente o tempo mensurado em segundos, minutos, horas. . .
- A partir da observação desses eventos no tempo podemos por exemplo, a partir de um momento 0 registrar o momento de ocorrências/realização desses eventos
- Além disso podemos registrar o número de ocorrências desses eventos até determinado ponto no tempo, ou calcular o intervalo de tempo entre realizações desses eventos.

Processos Pontuais

$$I_{i,t} = \begin{cases} 1 & \text{se } T_i \leq t \\ 0 & \text{c.c} \end{cases}$$

$$N_t = \sum_{i=1}^{\infty} I_i \text{ - Soma do número de ocorrências até o intervalo t}$$

Representamos por N_t o número de realizações de um determinado processo pontual em uma dimensão até determinado ponto t no tempo.

$$N_{[a,b)} = N_b - N_a$$

e Usamos a notação $N_[a, b)$ para calcular o número de realizações desse evento entre os momentos b e a com $b > a$

Processos Pontuais

- No caso bidimensional, aonde estariamos em um processo pontual espacial, nossa unidade de trabalho seriam ocorrências em uma região/área, e portanto ao invés de contabilizar número de ocorrências no tempo podemos pegar o número de ocorrências desses eventos em uma área/região

Processos de Poisson

- Os processos de Poisson são basicamente processos pontuais, porém seu processo gerador tem por trás uma distribuição de Poisson

Processos de Poisson - Unidimensionais

1. O número de realizações de um evento em um intervalo de tempo, depende de forma proporcional do tamanho/duração desse intervalo

$$E(N_{[a,b)}) = \beta(b - a)$$

Aqui $\beta > 0$ é uma constante, que pode ser denominada como intensidade

2. Uma vez que os intervalos sejam disjuntos, ou seja, não se sobreponham, ou seja : $a_1 < b_1 < a_2 < b_2 < ... < a_n < b_n$, então serão variáveis independentes.
3. A probabilidade de duas ou mais ocorrências dentro de um intervalo de tempo tem ordem assintoticamente menor que o tamanho do intervalo

Processos de Poisson - Unidimensionais

A partir dessas premissas, segue a definição de um processo de Poisson Unidimensional

1. O número de ocorrências de eventos dentro de um intervalo $N_[a, b)$ segue uma distribuição de *Poisson* e tal que $\lambda = \beta(b - a)$, ou seja a intensidade é proporcional ao tamanho do intervalo

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

2. Se os intervalos são disjuntos dois a dois (disjuntos em pares), portanto a soma de realizações de eventos desses processos serão independentes.

Processos de Poisson - Unidimensionais

3. Os intervalos de tempo entre a ocorrência dos eventos, denominados como S_i seguem uma distribuição exponencial com parâmetro $\beta > 0$

$$P(S_i \leq s) = 1 - e^{-\beta s}$$

4. Os intervalos entre as realizações são independentes, ou seja, o intervalo de tempo entre a 2^a e a 3^a realização é independente da duração entre a 1^a e a 2^a realização.
5. A duração do i-ésimo intervalo segue uma distribuição Gamma com parâmetro $\alpha = i$ e $\beta = \beta$

$$f(x, \alpha = i, \beta = \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

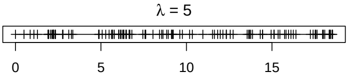
Processos de Poisson - Unidomensionais

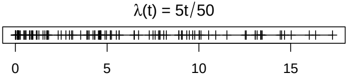
- Processo de Poisson Não Homogeneo
- Até aqui vimos o caso dos processos de Poisson que possuiam intensidade uniforme, ou seja, para cada intervalo no tempo de tamanho $b - a$, o número de realizações esperadas nesse intervalo seria o mesmo, independente em que intervalo da reta ocorresse
- O processo não Homogeneo por exemplo coloca que a intensidade é função do tempo

$$E[N_{[a,b)}] = \int_a^b \beta(t)dt$$

Processos de Poisson - Unidimensionais

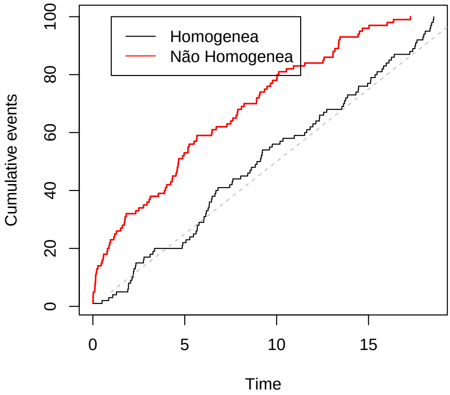
- Abaixo temos dois exemplos de Processos Unidimensionais de Poisson que foram simulados, o primeiro usa um homogeneo comum com $\lambda = 5$ e outro processo em que $\lambda(t) = \frac{5t}{50}$





Processos de Poisson - Unidimensionais

- Enquanto o processo homogeneo basicamente segue uma tendência media no tempo o não homogeneo apresenta uma variação ao longo do tempo.



Processos de Poisson - Espaciais

Podemos generalizar o que vimos em uma dimensão para duas dimensões, e aplicar de forma semelhante, conceitos e definições de $\mathbb{R}$ para $\mathbb{R}^d$, de forma inicial vamos olhar para o $\mathbb{R}^2$

- Processo bi-dimensional ocorre dentro do $\mathbb{R}^2$, se definirmos um conjunto fechado B tal que $B \subset \mathbb{R}^2$, então o número de ocorrências dentro do conjunto B seguirá uma *Poisson* com $\lambda = \beta\lambda_2(B)$[1]

[1] $\beta > 0$ e $\lambda_2(B)$ é a intensidade da área

Processos de Poisson - Espaciais

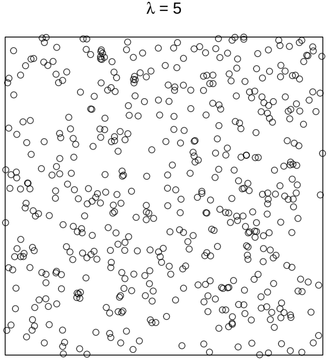
- Da mesma forma que no caso unidimensional, se tivermos regiões ditas $B_1, B_2, ..., B_n$ tais que essas regiões sejam disjuntas duas as duas, ou seja não haja intersecção entre as áreas, então o número de eventos ocorridos em cada uma dessas áreas denotadas por $N_{B_1}, N_{B_2}, ..., N_{B_n}$ serão independentes.

$$P(N(B_i) = n_i, i = 1, 2, ..., k) = \prod_{i=1}^{k} \frac{(\lambda|B_i|)_i^n e^{-\lambda|B_i|}}{n_i!}$$

- Aqui $|B_i|$ denota a área da região B_i

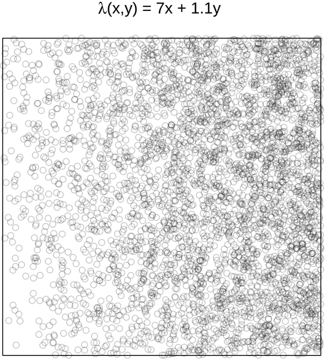
Procesos de Poisson - Espaciais

- Abaixo um exemplo de um processode Poisson espacial com Intensidade $\lambda = 5$.



Processos de Poisson - Espaciais

- Assim como no caso Unidimensional, no caso Espacial (dimensão com $d \geq 2$) temos as situações de processos de Poission não homogeneos, aonde por exemplo a intensidade varia de acordo com a região



Processos de Poisson - Espaciais

Intensidade A estimação da intensidade de um processo de poisson pode se dividir de duas formas.

1. Estimação de Processos Homogeneos
2. Estimação de Processos Não Homogeneos

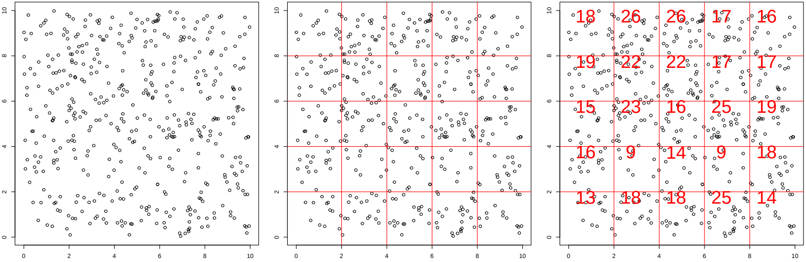
Estimação de Processos Homogeneos

- Quando tratamos de processos Homogeneos dentro de um contexto espacial no $\mathbb{R}^2$ o λ nos da uma noção de intensidade média por unidade de área, assim de dividirmos um espaço no $\mathbb{R}^2$ em áreas de mesmo tamanho, e simularmos um processo de Poisson nesse espaço, é esperado que cada região nesse espaço tenha o mesmo número de eventos dados por $\lambda * |B|$ [2] (Assim como ja observamos nas definições)

[2] Aqui $|B|$ denota a área do subconjunto B

Processos de Poisson - Espaciais

- O exemplo abaixo demonstra uma forma de estimação. Aqui geramos um processo de Poisson com $\lambda = 5$ em uma região de área $10x10$.
- Dividimos a região em uma 25 áreas de tamanho $2x2$, com o que vimos seria esperado que cada região tivesse um número de ocorrências esperados de $n = 5 * 2^2 = 20$, o efetivo acabou sendo 18.08, número muito próximo mas nao exato.



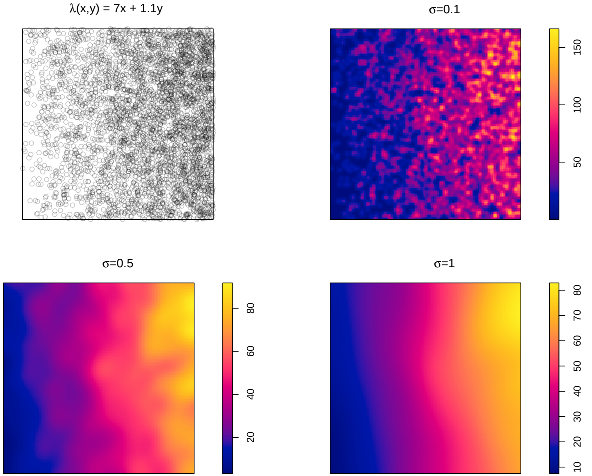
Processos de Poisson - Espaciais

Estimação de Processos Não Homogeneos Quando os processos são não Homogeneos, ou seja a intensidade varia de acordo com a posição no espaço, são utilizadas funções *Kernel*

- Os estimadores de Kernel, são funções que basicamente pegam os pontos distribuidos no nosso espaço gerados pelo nosso processo e buscam suavizar essa distribuição.
- Como temos alguns pontos usamos uma função que projetaria a distribuição de pontos baseado na distribuição efetiva dos pontos

Processos de Poisson - Espaciais

- Abaixo temos a simulação de um processo de Poisson não Homogeneo, temos o plot dos pontos no espaço e também o plot da densidade de *Kernel* com diferentes . . . ,



Processos de Poisson - Espaciais

Teste de Homogeneidade Como vimos é possivel estimar a densidade de um processo com base em um procedimento simples, que é justamente dividindo a regiao em n áreas e calcular o número médio de ocorrencia de pontos sob a área para justamente "estimar" a intensidade.

- Esse mesmo procedimento pode ser utilizado para testar Homogeneidade dos processos de Poisson. A chamada Contagem quadratica permite testar se nosso processo tem intensidade Homogenea.

$$\chi^2 = \sum_i \frac{(O(a_i) - E(a_i))^2}{E(a_i)}$$

Aonde $O(a_i)$ é o número de pontos observados na área a_i, e $E(a_i)$ é o número de pontos esperados na área a_i

Processos de Poisson - Espaciais

Usando o exemplo citado testaremos a homogeneidade com áreas de $2x2$, no teste realizados temo p-valor de 0.1561, portanto, não rejeitamos a hipotese nula de homogeneidade.

```
quadrat.test( sim_poss_1, 2, 2 )
```

```

    Chi-squared test of CSR using quadrat counts

data:  sim_poss_1
X2 = 6.8142, df = 3, p-value = 0.1561
alternative hypothesis: two.sided

Quadrats: 2 by 2 grid of tiles
```

Função K

A função K é uma forma de tentar mensurar essa suposta aleatoriedade do processo estudado.

Dado um conjunto de pontos num espaço em $\mathbb{R}^2$, a função K nos retorna o número esperado de pontos quando traçamos uma bola de raio d em qualquer um dos pontos do nosso conjunto.

$$K(r) = \frac{E(\text{Pontos na bola de raio r})}{\lambda}$$

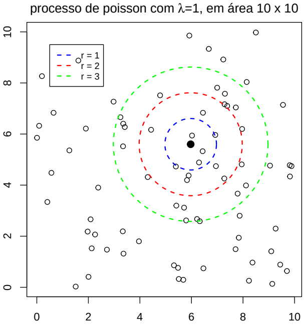
Aonde r é o raio e λ a intensidade do processo

Para um processo de *Poisson* homogeneo em $\mathbb{R}^2$ temos que a função K é dada por

$$K(r) = \pi r^2$$

Função K

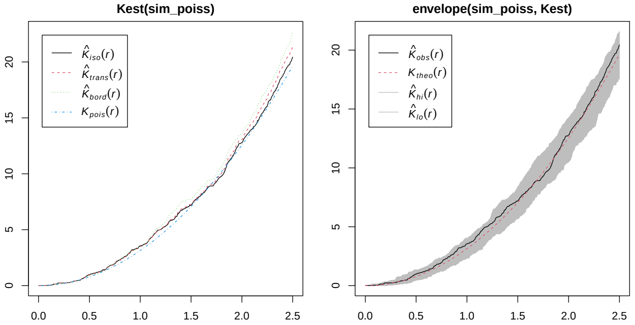
O exemplo abaixo se trata da simulação de um processo de Poisson numa região $10x10$ em um processo com $\lambda = 1$.



Como simulamos um processo homogeneo de poisson, então é possivel ver no gráfico que a função K observada se assemelha em muito ao do que seria do processo de Poisson teórico, que tem formula representada por πr^2.

Função K

- Exemplo abaixo mostra a função K para o processo gerado, na legenda vemos o que seria a função K para um processo de Poisson e outros processos, é possivel ver que como nossos dados foram simulados a partir de um processo de Poisson as curvas ficam muito próximas.



Processos espaciais de Poisson

Estimação Para a estimação de processos de Poisson sejam em uma ou mais dimensões o principal metodo se dá por meio da aplicação da MLE (Maximum Likelihood Estimation).

- Para além da MLE, também existem outros métodos[3] como:

1. Coarse quadrature approximation
2. Fine pixel approximation
3. Regressão Logística Condicional

[3] Aqui apresentamos os métodos associados e implementados no pacote `spatstat`

Processos de Cox

- Até agora, observando todo arcabouço dos processos de Poisson espaciais, vimos que uma das principais premissas utilizadas é a da indepêndencia dos eventos observados, o que muitas vezes se não na maioria das vezes se mostra quase que impraticavel em dados do mundo real.
- Os processos de Cox são muito parecidos com os processos de Poisson, porém contam com uma diferença que é o efeito aleátorio sobre a intensidade da distribuição $\Lambda(u)$, aonde aqui u é usado para se referênciar a uma área/ponto no espaço.
- Os efeitos aleatórios sobre a intensidade podem ser gerados por diversos mecanismos, desde que $\Lambda(u)$ seja positivo e integravel sobre conjuntos fechados

Processos de Cox

- Para além dos efeitos aleátorios dentro da intensidade também podem ser consideradas efeitos de outras variáveis sobre a distribuição espacial por exemplo:
  - Quando se analisa a distribuição espacial de tipos de arvores em uma floresta e alguns tipos de arvores são mais presentes em regiões com determinada altitude ou tipo de solo
  - Distribuição da criminalidade em determinadas regiões/ruas de uma cidade em função por exemplo de menor policiamento ou isolamento geográfico.

Processos de Cox

Random Fields Como dito os processos de cox possuem uma um efeito aleatório, e portanto a intensidade representada $\Lambda(u)$ varia no espaço, por exemplo:

Aqui temos um componente aleatório espacial,

Processos de Cox

Exemplos de Modelos de Processos de Cox:

- Log Gaussian Cox Process (LGCP)

$$g(u) = log\Lambda(u)$$

No modelo LGCP modelamos nosso $\Lambda(u)$ como uma função exponencial de $g(u)$, aonde $g(u)$ é um processo Gaussiano, e dai deriva ser um modelo log Gaussiano de cox.

A vantagem de trabalhar com um processo Gaussiano é que esse processo pode ser completamente identificado por meio da sua função de **média** e de **Covariância**

- **Média**: $\mu(u) = E(g(u))$
- **Covariância:** $\Sigma(u,v) = Cov(g(u), g(v)) = E(g(u)g(v)) - \mu(u)\mu(v)$[4]

[4] Relemebrando que aqui u e v são posições no espaço.

Processos de Cox

- Em um LGCP temos algumas restrições e não podemos usar qualquer função de média e covariância, algumas condições precisam ser atendidas:
  - Matrix de covariância ser positivamente definida
- Modelo da função de covariância segue seguinte forma[5]:

$$C(u, v) = C_0(r) = \sigma^2 R(r/\alpha)$$

- Principais funções de Covariância

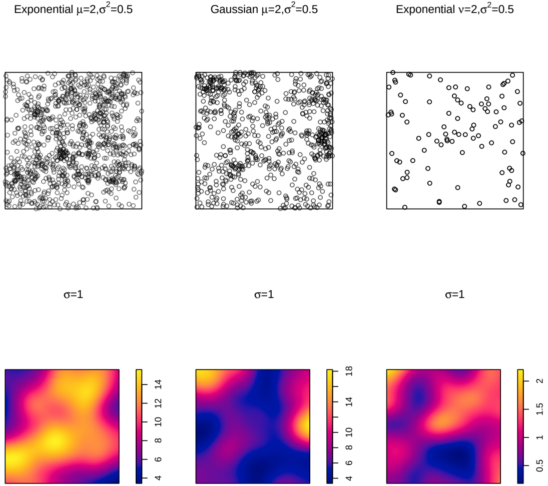
1. **Gaussiana**: $R(r) = \exp(r^2)$
2. **Exponencial**: $R(r) = \exp(-r/\alpha)$
3. **Matern**[6]: $R(r) = 2^{1-v}\Gamma(v)^{-1}(2\sqrt{2v}r)^v K_v(2\sqrt{2v}r)$

[5] Aqui $r \geq 0$ é colocado como como uma distância entre pontos

[6] Aqui v é um parâmetro de ordem para uma função K_v chamada função de Bessel

Processos de Cox

- Abaixo temos simulações de um *Log Gaussian Cox Process* usando diferentes funções de covariâncias, para além disso temos o *Smooth* da intensidade utilizando um $\sigma = 1$



Processos de Cox

Estimação Diferentemente dos processos de Poisson, os processos de Cox não podem ser estimados diretamente por MLE, e por isso contam com metodos próprios, abaixo temos alguns dos principais métodos disponíveis[7].

1. Metodo do Mínimo Contraste
2. Palm Likelihood
3. Composite Likelihood

Para além dos métodos também é comum a utilização de Inferência Bayesiana

[7] Os metodos de estimação aqui mencionados são os utilizados no `spatstat`

Processos de Cox

- Quando falamos da estimação para a restrição dos LGCP *Log Gaussian Cox Process* entramos em uma outra gama de modelos para estimação, no geral também se utiliza inferência Bayesiana mas os principais métodos podem ser sumarizados em:

1. MCMC *(Markov Chain Monte Carlo)*
2. INLA *(Integrated Nested Laplace Approximation)*

- Cada metodo terá suas vantagens e desvantagens na estimação, alguns podem ser mais eficientes em problemas complexos e outros mais precisos, não abordaremos com profundidade nessa apresentação os metodos de estimação.

Processos de Cox

Exemplos A fim de deixar o conteúdo de maneira mais palpavel apresentaremos exemplos mais aplicados sobre Processos de Cox, usaremos bases do pacote spatstats do R

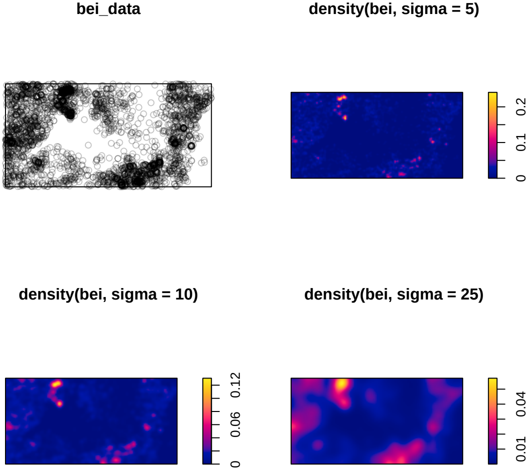
Em anexo a essa apresentação deixaremos um Notebook com algumas instruções em que será possível replicar todas as simulações e aplicações realizadas nessa apresentação.

Nosso problema Inicial será a analise de um dado espacial do pacote `spatstat`, no `R`, a base em especifico se chama `bei` e contém dados relacionados ao padrão de crescimento de arvores do tipo em uma região - Para além da distribuição espacial por meio de pontos temos dados de:

1. Elevação do Terreno
2. Inclinação do terreno

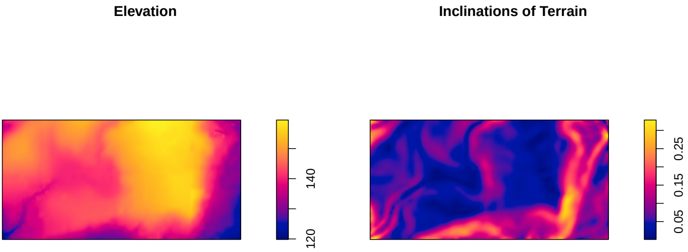
Processos de Cox

- Abaixo temos gráficos que mostram a distribuição por coordenadas dos dados, e além disso a densidade calculada pela função *Kernel*



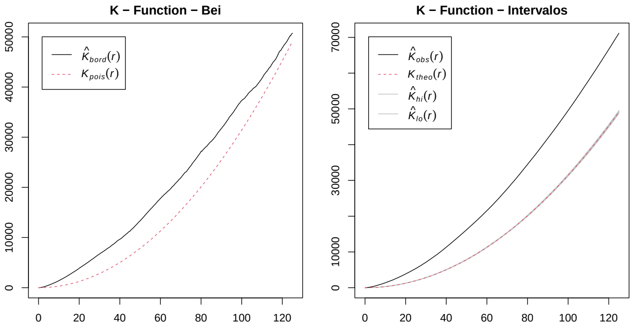
Processos de Cox

- Algumas covariadas apresentadas no *data-set*, no caso inclinação do terreno e Altitude:



Processos de Cox

- Aqui usaremos a função K para avaliar a homogeneidade dos dados, pela função fica claro que se trata de uma distribuição bem diferente dos casos de uma poisson teorica.



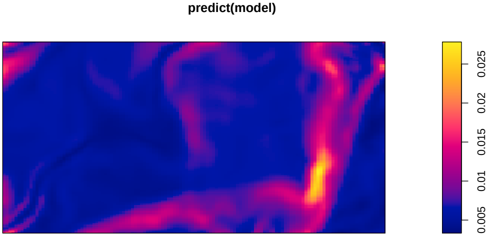
Processos de Cox

```
model  <-
  kppm(bei ~ elev + grad, "LGCP",
       model="exp",
       data = bei.extra)

print( model )
```

```
Inhomogeneous Cox point process model
Fitted to point pattern dataset 'bei'
Fitted by minimum contrast
    Summary statistic: inhomogeneous K-function

Log intensity:  ~elev + grad

Fitted trend coefficients:
(Intercept)        elev        grad
-8.56355220  0.02143995  5.84646680

Cox model: log-Gaussian Cox process
    Covariance model: exp
Fitted covariance parameters:
      var     scale
 1.580813 48.311832
Fitted mean of log of random intensity: [pixel image]
```

Processos de Cox

```
plot( predict( model ))
```



Processos de Cox

- Agora vamos a um exemplo de manipulação de dados mais "Real", tentaremos olhar dados de queimadas noe stado de São Paulo.

```
library( sf )
# Dados foram baixados do site:
# Dados de origem nacional foram compilados de forma manualmente
# Codigo tratamento dos dados: github.com/yuripassuelo/estat_aplicada/src/processamento.R
path <-
  "C:/Users/yurim/Desktop/Mestrado/1S2024/Estat Aplicada/Trabalho/data/"

base <-
  readRDS( paste0( path, "/inter/inter_data.rds" ) )


base_sp <-
  filter( base, year == 2020, Estado == "SÃO PAULO", RiscoFogo == 1 )

map_sp <-
  geobr::read_state( code_state = "SP")
```

```
Downloading: 8.4 kB     Downloading: 8.4 kB     Downloading: 25 kB     Downloading: 25 kB     Downloading
```

Processos de Cox

- Conversão do mapa para a janela espacial utilizada no pacote *spatstat*

```
# Conversão do Mapa em sf para janela do pacote spatstat
sp_mod <- st_union( map_sp )

sp_flat <- st_transform( sp_mod, crs = 3857 )

sp_owin <- as.owin( sp_flat )
```

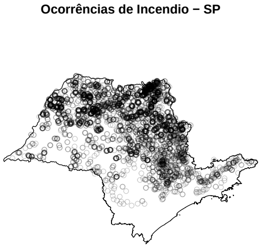
Processos de Cox

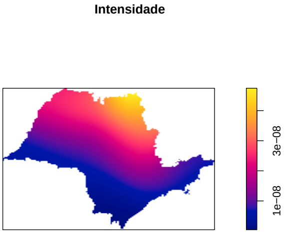
- Conversão dos dados em coordenadas para o formato ppp do pacote spatstat

```
# Conversão dos dados para padrao Spatstat
mod <- base_sp %>%
  mutate( x = Longitude,
          y = Latitude )%>%
  select( x,y,everything() )

points_sp <- mod %>%
  select( x,y ) %>%
  st_as_sf( coords = c("x","y"), crs = 4674 ) %>%
  st_transform( crs = 3857 )%>%
  st_coordinates() %>%
  as.data.frame()

mod_sp <-
  mod %>%
  mutate( x = points_sp$X,
          y = points_sp$Y)

# Criação  de dados em formato ppp
sp_ppp_dta <-
  as.ppp( mod_sp %>% select( x, y ), W = sp_owin  ) %>%
  unique.ppp( ) #biomes_owin )

sp_ppp_dta_extra <-
  as.ppp( mod_sp , W = sp_owin  ) %>%
  unique.ppp( )
```

Processos de Cox

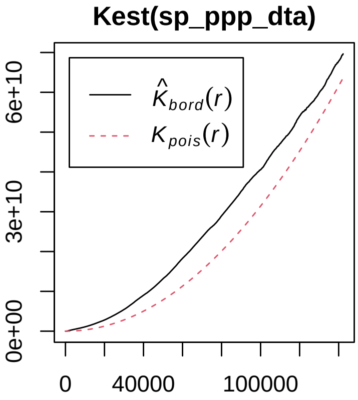
- Dados de queimada e intensidade por *Kernel*





Processos de Cox

- Função K para comparação de com um processos de Poisson



Bibliográfia

Introdução a Teoria:

- BADDLEY, Adrian; BÁRÁNY, Imre; SCHNEIDER, Rolf; WEIL, Wolfgang. Stochastic Geometry: Spatial Point Processes and their Applications. In: STOCHASTIC Geometry: Spatial Point Processes and their Applications. [S. l.]: Springer, 2007. cap. 1,2,3,4.
- KEELER, Paul. Notes on the Poisson point process. -, [s. l.], 2018.

Aplicações Práticas:

- BADDLEY, Adrian; RUBANK, Ege; TURNER, Rolf. Spatial Point Patterns: Methdology and applications in R. In: SPATIAL Point Patterns: Methdology and applications in R. [S. l.]: Chapman & Hall/CRC, 2016. cap. 6,9,10,12.
- MORAGA, Paulo. Spatial Statistics for Data Science: Theory and Practice with R. In: SPATIAL Statistics for Data Science: Theory and Practice with R. [S. l.]: Chapman & Hall/CRC, 2023. cap. 17, 18, 21, 22, 23.