

Estatística Espacial (MI418) / Geoestatística (ME907)

Guilherme Ludwig

2019-02-05

Função K de Ripley

Processos de Neyman-Scott

Fry plot

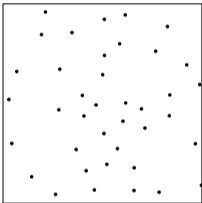
Os *fryplots* são uma representação simples de interação entre pontos. O objetivo é revelar estruturas de interação, mas os *fryplots* também podem revelar anisotropia.

Para construir um *fryplot* a partir de um padrão de pontos $x = \{x_1, \dots, x_n\}$, usa-se o seguinte algoritmo:

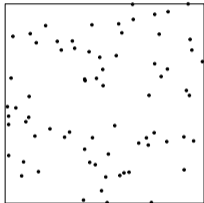
1. Para $i = 1, \dots, n$, centramos a janela na coordenada i .
2. Registramos os outros pontos onde eles ficarem com relação à i .

Exemplos de pontos

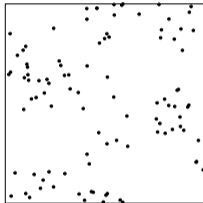
Processo Hardcore (repulsivo)



Processo de Poisson

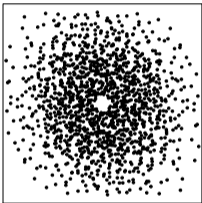


Processo de Thomas (atractivo)

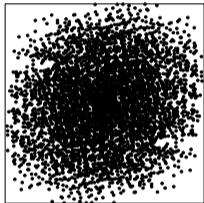


Exemplos de fryplot

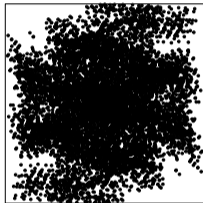
Processo Hardcore (repulsivo)



Processo de Poisson



Processo de Thomas (atrativo)



Função K

A função K de um processo é definida como

$$K(r) = \frac{1}{\rho} \mathbb{E} \left(\sum_{\xi, \eta \in X: \xi \neq \eta} \mathbf{1}\{\|\xi - \eta\| < r\} \right),$$

com estimador

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{x_i, x_j \in X: x_i \neq x_j} \mathbf{1}\{\|x_i - x_j\| < r\} e_{ij}(r),$$

onde $e_{ij}(r)$ é uma correção do viés da janela. Por exemplo, uma correção simples é

$$e_{ij}(r) = \frac{n(n-1)}{\sum_{kl} e_{kl}(r)} \mathbf{1}\{\{\|x_i - x_j\| < r\} \in W\}.$$

Interpretação

O processo de Poisson homogêneo tem função K dada por

$$K_{\text{Poisson}}(r) = \pi r^2.$$

- ▶ Se $K(r) > \pi r^2$, então o processo mostra atração entre pontos à distância r .
- ▶ Se $K(r) < \pi r^2$, então o processo mostra repulsão entre pontos à distância r .

Função L

O processo de Poisson homogêneo tem função K dada por

$$K_{\text{Poisson}}(r) = \pi r^2.$$

É comum encontrar na literatura a função L , dada por

$$L(r) = \sqrt{\frac{K(r)}{\pi}}.$$

Porém, como vamos usar a função K_θ para construir contrastes com relação a \hat{K} , eu costumo evitar usar L .

A função K e a correlação g

A função K é diretamente conectada à função g de correlação entre pares, na seguinte forma:

$$g(r) = \frac{K'(r)}{2\pi r},$$

onde $K'(r)$ é a derivada de K com respeito a r (veja Baddeley et al., 2015, p. 225).

A idéia é que a função K represente o número esperado de pontos ao redor (com distância r) de um ponto fixado, enquanto a correlação entre pares é o número esperado de pontos à distância exatamente r . A função K é mais estável para ser usada para contrastes, ainda que a função g seja perfeitamente interpretável como uma medida de correlação.

Função K

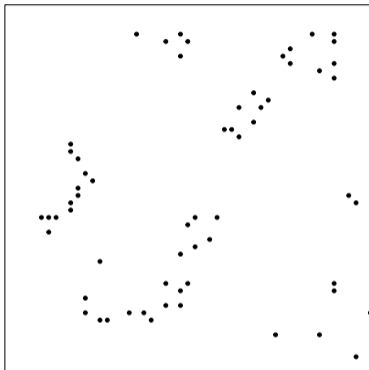
Na prática assumir que $\rho(\mathbf{s}) = \rho$ (constante) é muito restritivo. Há propostas de correção, como por exemplo a de Baddeley et al. (2000), que é

$$K_{\text{inhom}}(r) = \frac{1}{|W|} \mathbb{E} \left(\sum_{\xi, \eta \in X: \xi \neq \eta} \frac{\mathbf{1}\{\|\xi - \eta\| < r\}}{\rho(\xi)\rho(\eta)} \right).$$

É claro, é preciso estimar ρ .

Exemplo: árvores tropicais

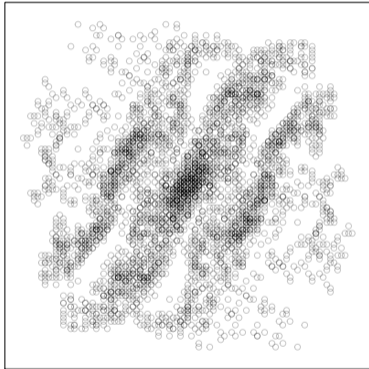
California Redwood Trees



Árvores do tipo *California Redwood*, originalmente no artigo Strauss (1975). Fonte: `spatstat::redwood`.

Exemplo: California Redwood

Fry plot of redwood

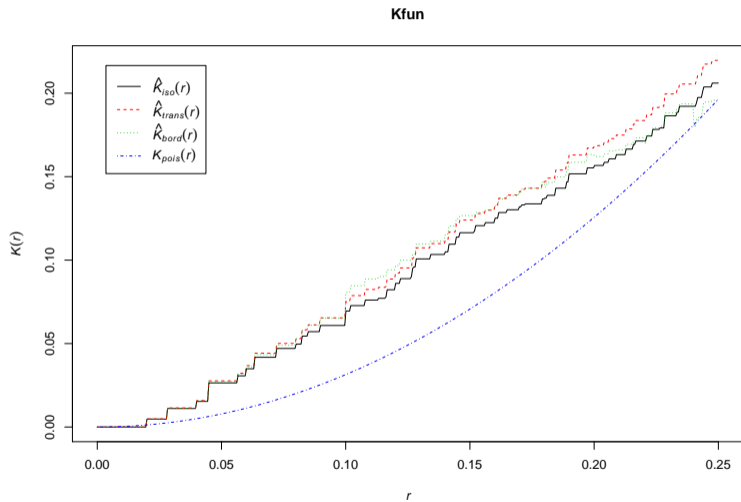


Exemplo: California Redwood

```
(Kfun <- Kest(redwood))
```

```
## Function value object (class 'fv')
## for the function r -> K(r)
## .....
##      Math.label      Description
## r      r            distance argument r
## theo   K[pois](r)    theoretical Poisson K(r)
## border hat(K)[bord](r) border-corrected estimate of K(r)
## trans  hat(K)[trans](r) translation-corrected estimate of K(r)
## iso    hat(K)[iso](r) isotropic-corrected estimate of K(r)
## .....
## Default plot formula:  .~r
## where "." stands for 'iso', 'trans', 'border', 'theo'
## Recommended range of argument r: [0, 0.25]
## Available range of argument r: [0, 0.25]
```

Exemplo: California Redwood



Exemplo: California Redwood

```
eKfun <- envelope(redwood)
```

```
## Generating 99 simulations of CSR ...
```

```
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,
```

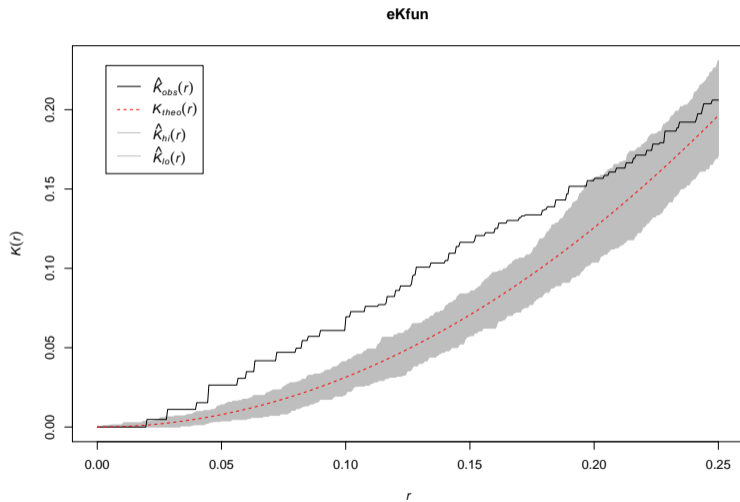
```
## 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57,
```

```
## 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95,
```

```
##
```

```
## Done.
```

Exemplo: California Redwood



Processos de Neyman-Scott

Um processo de Neyman-Scott modela fenômenos de *atração*. Alguns livros chamam os processos de Neyman-Scott de *cluster processes*. A construção de um modelo de Neyman-Scott é simples:

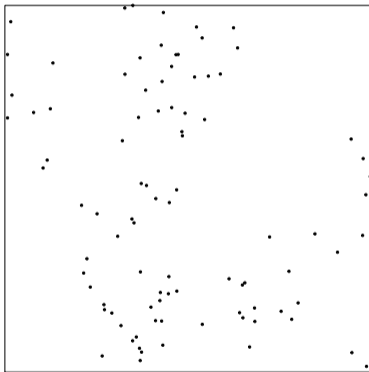
- ▶ Gere um processo de Poisson¹ Y_c , com intensidade κ_c , normalmente chamado de *parent process*.
- ▶ Para cada ponto y_1, y_2, \dots, y_n na realização de Y_c , gere um processo de Poisson X_c com intensidade ρ_c (*child process*).
- ▶ Apague os pontos Y_c , defina $X = \cup_{c=1}^n X_c$.

Então X é um processo de Neyman-Scott.

¹Rigorosamente não precisa ser Poisson nem homogêneo, mas por simplicidade...

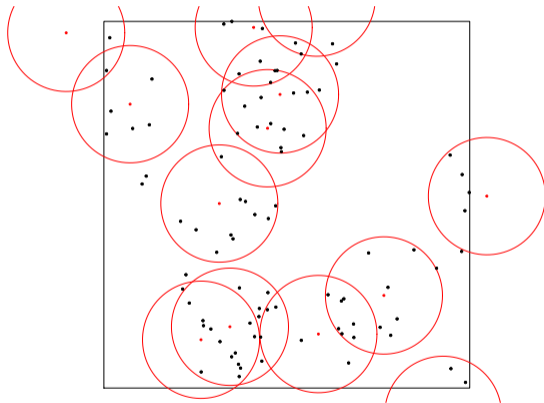
Exemplo

Thomas' Process



Exemplo

Thomas' Process



Processos de Neyman-Scott

Alguns processos famosos:

- ▶ **Thomas**: *Parent* homogêneo com intensidade κ , com *child* sendo um kernel Gaussiano centrado no *parent*, com desvio-padrão σ e número médio de pontos μ (`rThomas`).
- ▶ **Matérn Cluster**: *Parent* homogêneo com intensidade κ , com *child* sendo um kernel uniforme centrado no *parent*, de raio σ e número médio de pontos μ (`rMatClust`).
- ▶ Outros modelos incluem **Cauchy radial** (`rCauchy`) e **gamma radial** (`rVarGamma`).

Note que o *parent* pode ser não-homogêneo mas na prática é muito difícil estimá-lo assim, e pode levar a confundimentos. Regressão é usualmente feita em μ .

Estimação

Na prática, a verossimilhança de um processo de Neyman-Scott é muito difícil de ser escrita.

Porém, as funções K são fáceis. Por exemplo, o processo de Thomas com parâmetros κ, σ, μ tem função K teórica dada por

$$K_{\kappa, \sigma, \mu}(r) = \pi r^2 + \left(1 - \exp\left\{-\frac{r^2}{4\sigma^2}\right\}\right) / \kappa,$$

com intensidade teórica $\kappa\mu$.

As estimativas $\hat{\kappa}, \hat{\sigma}$ de mínimo contraste são soluções de

$$\hat{\kappa}, \hat{\sigma} = \arg \min_{\kappa, \sigma} \int_0^{\infty} \ell(K_{\kappa, \sigma}(r) - \hat{K}(r)) dr,$$

onde ℓ é algum critério de divergência (por exemplo $\ell(r) = r^2$) e $\hat{\mu} = n(X)/(\hat{\kappa}|W|)$.

Estimação

```
kppm(redwood ~ 1, "Thomas", method="c")
```

```
## Stationary cluster point process model
## Fitted to point pattern dataset 'redwood'
## Fitted by maximum second order composite likelihood
## rmax = 0.25
## weight function: Indicator(distance <= 0.125)
##
## Uniform intensity:    62
##
## Cluster model: Thomas process
## Fitted cluster parameters:
##      kappa      scale
## 28.02839670  0.03538136
## Mean cluster size:  2.212042 points
```

Regressão

Regressão para processos de Neyman-Scott é baseada no contraste perfilado.

```
slope <- bei.extra$grad
model <- kppm(bei ~ slope, "Thomas", method = "c")
coef(summary(model))
```

##	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
## (Intercept)	-5.390553	0.1224604	-5.630571	-5.150535	***	-44.018729
## slope	5.022021	1.1727197	2.723533	7.320510	***	4.282371

Referências I

- Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC Press.
- Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.
- Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62(2):467–475.