

Test du processus de Poisson homogène par la statistique de Ripley

Gabriel Lang, Eric Marcon

► **To cite this version:**

Gabriel Lang, Eric Marcon. Test du processus de Poisson homogène par la statistique de Ripley. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386800>

HAL Id: inria-00386800

<https://hal.inria.fr/inria-00386800>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TEST DU PROCESSUS DE POISSON HOMOGÈNE PAR LA STATISTIQUE DE RIPLEY

Gabriel Lang & Eric Marcon

*AgroParisTech, Equipe MORSE, UMR MIA518, F-75005 Paris, France,
AgroParisTech, UMR EcoFoG, F-97310 Kourou, France*

Mots-clés principaux: statistique mathématique, statistique des processus

Mots-clés secondaires: processus ponctuels, processus de Poisson, agrégation, test asymptotique

1 Résumé

1.1 Définition et notations

Soit X un processus ponctuel sur \mathbb{R}^2 . On peut caractériser la dépendance entre les positions des points de X en étudiant le processus ponctuel $X^{(2)}$ défini sur $\mathbb{R}^2 \times \mathbb{R}^2$ par tous les couples de points du processus original. L'intensité de ce nouveau processus informe sur la présence simultanée de points dans le processus original. Notons $\rho^{(2)}(x, y)$ cette (appelée densité produit du second ordre). L'intensité du processus de Poisson $\rho(x)$ a pour propriété que $\rho^{(2)}(x, y) = \rho(x)\rho(y)$.

La statistique de Ripley est une méthode d'estimation de la densité $\rho^{(2)}(x, y)$, plus exactement d'estimer l'intégrale sur des ensembles tests du rapport $g(x, y) = \rho^{(2)}(x, y)/\rho(x)\rho(y)$. La fonction $g(x, y)$ caractérise le fait que les points x and y simultanément dans les tirages de X . Si $g(x, y) = 1$, les points apparaissent indépendamment. Si $g(x, y) < 1$, ils ont tendance à s'exclure; si $g(x, y) > 1$, ils apparaissent fréquemment ensemble.

Nous supposons une certaine invariance spatiale du processus. Tout d'abord l'invariance par translation $g(x, y) = g(x - y)$. Nous définissons alors l'intégrale de la fonction g sur un ensemble \mathcal{K} :

$$\mathcal{K}(A) = \int_A g(x)dx.$$

Si nous supposons en plus que le processus is isotrope, nous définissons la fonction K de Ripley par

$$K(r) = \mathcal{K}(B(0, r)).$$

Par exemple, si le processus est processus de Poisson, alors $g(x) = 1$ et $K(r) = \pi r^2$. Nous utilisons l'estimateur de Ripley (Ripley, 1976) de la fonction K comme une statistique de test pour les processus de Poisson. Soit A un ensemble fini de \mathbb{R}^2 , m la mesure de Lebesgue. Soit S be a realisation du processus ; la statistique de Ripley est définie par

$$\hat{K}_A(r) = \frac{1}{m(A)} \sum_{X_i \neq X_j \in S} \frac{\mathbb{I}\{d(X_i, X_j) \leq r\}}{\hat{\rho}(X_i) \hat{\rho}(X_j)}.$$

1.2 Hypothèses

Nous supposons que X est un processus de Poisson homogène sur \mathbb{R}^2 d'intensité inconnue ρ . On choisit l'estimateur $\hat{\rho} = \sqrt{N(N-1)}/n^2$. Les données sont observées sur le carré $A_n = [0, n]^2$ et on fait tendre n vers l'infini. La statistique de Ripley s'exprime comme

$$\hat{K}_{A_n}(r) = \frac{n^2}{N(N-1)} \sum_{X_i \neq X_j \in S} \mathbb{I}\{d(X_i, X_j) \leq r\}.$$

1.3 Biais

Nous calculons le biais dû aux effets de bord :

Proposition 1

$$\mathbb{E}\hat{K}_{A_n}(r) = r^2 \left(\pi - \frac{8r}{3n} + \frac{r^2}{2n^2} \right) \left(1 - e^{-\rho n^2} - \rho n^2 e^{-\rho n^2} \right).$$

1.4 Variance

nous calculons un équivalent de la matrice de covariance de $\hat{K}_{A_n}(r)$.

Proposition 2

$$\begin{aligned} \text{Var}(\hat{K}_{A_n}(r)) &= \frac{2\pi r^2}{\rho^2 n^2} + O\left(\frac{1}{n^3}\right). \\ \text{Cov}(\hat{K}_{A_n}(r), \hat{K}_{A_n}(r')) &= \frac{2\pi r^2}{\rho^2 n^2} + O\left(\frac{1}{n^3}\right). \end{aligned}$$

1.5 Théorème de la Limite Centrale

Nous montrons qu'un vecteur normalisé de statistiques de Ripley pour différents r converge en distribution vers un vecteur normal.

Theorem 1 *Soit d un entier, $0 < r_1 < \dots < r_d$ un ensemble de réels ; Soit $\mathcal{K}_n = (\hat{K}_{A_n}(r_1), \dots, \hat{K}_{A_n}(r_d))$:*

$$n\sqrt{\hat{\rho}}(\mathcal{K}_n - \pi(r_1^2, \dots, r_d^2)) \longrightarrow \mathcal{N}(0, \Sigma)$$

avec

$$\Sigma_{s,t} = \frac{2\pi(r_s^2 \wedge r_t^2)}{\rho}$$

1.6 Applications à la statistique de test

Du TLC, nous déduisons que $T = \Sigma^{-1/2}\mathcal{K}_n$ est asymptotiquement distribué comme $\mathcal{N}(0, I_d)$. Sous l'hypothèse

H_0 : X est un processus de Poisson homogène.

nous utilisons $T^2 = \sum T_i^2$ comme statistique de test avec zone de rejet au niveau α :

$$T^2 > \chi_\alpha^2(d).$$

où $\chi_\alpha^2(d)$ est le quantile $(1 - \alpha)$ de la distribution $\chi^2(d)$.

2 Abstract

2.1 Definition and notations

Consider a point process X over \mathbb{R}^2 . To characterize the dependence of the position of points in X , one may consider the point process $X^{(2)}$ defined on $\mathbb{R}^2 \times \mathbb{R}^2$ by all the couple of points of the original process. The density of this new process gives information on the simultaneous presence of points in the original process. Denote $\rho^{(2)}(x, y)$ this density (called the second order product density). The Poisson process of density $\rho(x)$ will have the property that $\rho^{(2)}(x, y) = \rho(x)\rho(y)$.

The Ripley statistic is a way to estimate the density $\rho^{(2)}(x, y)$. Precisely it is a way to estimate the integral on test sets of the ratio $g(x, y) = \rho^{(2)}(x, y)/\rho(x)\rho(y)$. The function $g(x, y)$ characterizes the fact that the points x and y appear simultaneously in the samples of X . If $g(x, y) = 1$, the points appear independently. If $g(x, y) < 1$, they tend to exclude each other; if $g(x, y) > 1$, they appear more frequently together.

Now we assume some spatial invariance on the point process. The first property is the translation invariance $g(x, y) = g(x - y)$. In order to estimate the function g , we define its integral as the set function \mathcal{K} :

$$\mathcal{K}(A) = \int_A g(x)dx$$

If we also assume that the point process is isotropic, we define Ripley K -function as

$$K(r) = \mathcal{K}(B(0, r))$$

For example, if the process is a Poisson process then $g(x) = 1$ and $K(r) = \pi r^2$. We use the Ripley estimator of the K function as a test for Poisson processes. We define the Ripley statistic (Ripley, 1976) that estimates the K function. Let A be a finite subset of the \mathbb{R}^2 , m be the Lebesgue measure. Assume that the intensity ρ is known. Let S be a realization of the point process; Ripley statistic is defined by

$$\widehat{K}_A(r) = \frac{1}{m(A)} \sum_{X_i \neq X_j \in S} \frac{\mathbb{I}\{d(X_i, X_j) \leq r\}}{\widehat{\rho}(X_i) \widehat{\rho}(X_j)}$$

2.2 Assumptions

We assume that X is an homogeneous Poisson process on \mathbb{R}^2 with unknown intensity ρ . Choose $\hat{\rho} = \sqrt{N(N-1)}/n^2$. We consider that the data are available on the square $A_n = [0, n]^2$ and let n go to infinity. Ripley statistic is expressed as

$$\hat{K}_{A_n}(r) = \frac{n^2}{N(N-1)} \sum_{X_i \neq X_j \in S} \mathbb{I}\{d(X_i, X_j) \leq r\}.$$

2.3 Bias

We compute the bias due to the edge effect

Proposition 3

$$\mathbb{E}\hat{K}_{A_n}(r) = r^2 \left(\pi - \frac{8r}{3n} + \frac{r^2}{2n^2} \right) \left(1 - e^{-\rho n^2} - \rho n^2 e^{-\rho n^2} \right).$$

2.4 Variance

We compute an equivalent of the covariance matrix of $\hat{K}_{A_n}(r)$.

Proposition 4

$$\begin{aligned} \text{Var}(\hat{K}_{A_n}(r)) &= \frac{2\pi r^2}{\rho^2 n^2} + O\left(\frac{1}{n^3}\right). \\ \text{Cov}(\hat{K}_{A_n}(r), \hat{K}_{A_n}(r')) &= \frac{2\pi r^2}{\rho^2 n^2} + O\left(\frac{1}{n^3}\right). \end{aligned}$$

2.5 Central Limit Theorem

We show that a normalized vector of Ripley's statistics for different r converges in distribution to a normal vector.

Theorem 2 *Let d be an integer, $0 < r_1 < \dots < r_d$ a set of reals and define $\mathcal{K}_n = (\hat{K}_{A_n}(r_1), \dots, \hat{K}_{A_n}(r_d))$:*

$$n\sqrt{\hat{\rho}}(\mathcal{K}_n - \pi(r_1^2, \dots, r_d^2)) \longrightarrow \mathcal{N}(0, \Sigma)$$

where

$$\Sigma_{s,t} = \frac{2\pi(r_s^2 \wedge r_t^2)}{\rho}$$

2.6 Applications to test statistics

From the CLT, we deduce that $T = \Sigma^{-1/2}\mathcal{K}_n$ is asymptotically $\mathcal{N}(0, I_d)$ distributed. Under the hypothesis

H_0 : X is an homogeneous Poisson process.

we use $T^2 = \sum T_i^2$ as a test statistic with rejection zone for the level α :

$$T^2 > \chi_\alpha^2(d).$$

where $\chi_\alpha^2(d)$ is the $(1 - \alpha)$ -quantile of the $\chi^2(d)$ distribution.

Bibliographie

- [1] Moller, J., Waagepetersen, R.(2004) *Statistical inference and simulation for spatial point processes*, Chapman and Hall.
- [2] Ripley, B. D. (1976) The second-order analysis of stationary point processes. *J. Appl. Probab.* 13, 255-266.
- [3] Ripley, B. D. (1979) Tests of ‘randomness’ for spatial point patterns. *J. Roy. Statist. Soc. B* 41, 368-374.