Répartition complètement aléatoire : Lois, simulations et tests

Jean VAILLANT

Département de Mathématiques et Informatique, U.A.G.

29 Janvier 2010

Plan

Exemples introductifs

Intéressons nous à des occurrences se produisant en dimension 1, 2 ou 3:

- Arrivées de clients sur un serveur dans un laps de temps

Exemples introductifs

Intéressons nous à des occurrences se produisant en dimension 1, 2 ou 3:

- Arrivées de clients sur un serveur dans un laps de temps
- Arbres malades dans une parcelle sylvicole

Exemples introductifs

Intéressons nous à des occurrences se produisant en dimension 1, 2 ou 3:

- Arrivées de clients sur un serveur dans un laps de temps
- Arbres malades dans une parcelle sylvicole
- Accidents de la route dans une zone au cours d'une année.

Exemples introductifs

Intéressons nous à des occurrences se produisant
en dimension 1, 2 ou 3:

- Arrivées de clients sur un serveur dans un laps de temps
- Arbres malades dans une parcelle sylvicole
- Accidents de la route dans une zone au cours d'une année.
- Orchidées dans un massif forestier

Questionnement

- Que signifient les expressions

  - "les occurrences se sont produites au hasard" ?

Questionnement

- Que signifient les expressions

- "les occurrences se sont produites au hasard" ?
- "la répartition des occurrences est complètement aléatoire" ?

Questionnement

- Que signifient les expressions
  - "les occurrences se sont produites au hasard"?
  - "la répartition des occurrences est complètement aléatoire"?
- Comment peut-on tester cette hypothèse statistiquement?

Questionnement

- Que signifient les expressions
  - "les occurrences se sont produites au hasard"?
  - "la répartition des occurrences est complètement aléatoire"?
- Comment peut-on tester cette hypothèse statistiquement?
- Avec quelles données?

Questionnement

- Que signifient les expressions
  - "les occurrences se sont produites au hasard" ?
  - "la répartition des occurrences est complètement aléatoire" ?
- Comment peut-on tester cette hypothèse statistiquement?
- Avec quelles données?
- Avec quelles procédures statistiques?

Conditions définissant une RCA

Conditions définissant une RCA

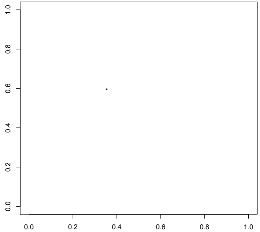
- Homogénéité : Chaque point de l'espace considéré a la même chance de "recevoir une occurrence du phénomène"

Conditions définissant une RCA

- Homogénéité : Chaque point de l'espace considéré a la même chance de "recevoir une occurrence du phénomène"
- Indépendance : les occurrences se réalisent indépendamment l'une de l'autre. Une occurrence n'a pas d'influence sur les autres.

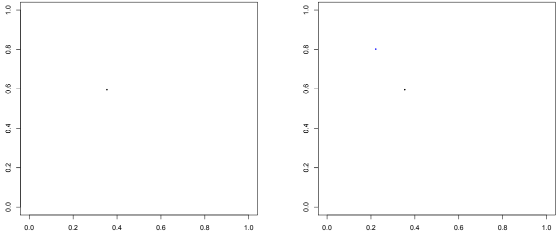
Conditions définissant une RCA

- Homogénéité : Chaque point de l'espace considéré a la même chance de "recevoir une occurrence du phénomène"
- Indépendance : les occurrences se réalisent indépendamment l'une de l'autre. Une occurrence n'a pas d'influence sur les autres.



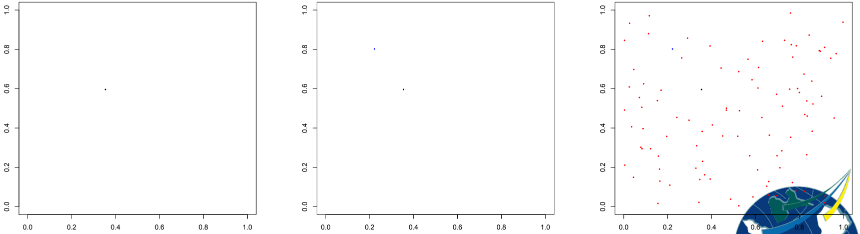
Conditions définissant une RCA

- Homogénéité : Chaque point de l'espace considéré a la même chance de "recevoir une occurrence du phénomène"
- Indépendance : les occurrences se réalisent indépendamment l'une de l'autre. Une occurrence n'a pas d'influence sur les autres.



Conditions définissant une RCA

- Homogénéité : Chaque point de l'espace considéré a la même chance de "recevoir une occurrence du phénomène"
- Indépendance : les occurrences se réalisent indépendamment l'une de l'autre. Une occurrence n'a pas d'influence sur les autres.



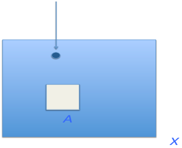
Lois de dénombrement

On a un ensemble de référence X. Pour toute partie (mesurable) A de X, on s'intéresse au nombre d'occurrences N_A dans A.

Lois de dénombrement

On a un ensemble de référence X. Pour toute partie (mesurable) A de X, on s'intéresse au nombre d'occurrences N_A dans A.

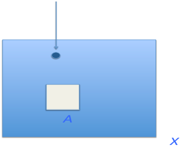
RCA donc position (d'une occurrence dans X) distribuée selon la loi uniforme sur X.



Lois de dénombrement

On a un ensemble de référence X. Pour toute partie (mesurable) A de X, on s'intéresse au nombre d'occurrences N_A dans A.

RCA donc position (d'une occurrence dans X) distribuée selon la loi uniforme sur X.



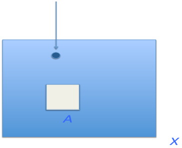
Pour 1 seule occurrence dans X, loi de Bernoulli :

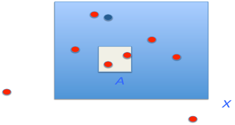
$N_A \in \{0, 1\}$ et $P(N_A = 1) = \dfrac{\nu(A)}{\nu(X)}$, avec ν mesure sur X.

Lois de dénombrement

On a un ensemble de référence X. Pour toute partie (mesurable) A de X, on s'intéresse au nombre d'occurrences N_A dans A.

RCA donc position (d'une occurrence dans X) distribuée selon la loi uniforme sur X.





Pour 1 seule occurrence dans X, loi de Bernoulli :

$N_A \in \{0, 1\}$ et $P(N_A = 1) = \dfrac{\nu(A)}{\nu(X)}$, avec ν mesure sur X.

Lois de dénombrement

On a un ensemble de référence X. Pour toute partie (mesurable) A de X, on s'intéresse au nombre d'occurrences N_A dans A.

RCA donc position (d'une occurrence dans X) distribuée selon la loi uniforme sur X.

Pour 1 seule occurrence dans X, loi de Bernoulli :

$N_A \in \{0, 1\}$ et $P(N_A = 1) = \dfrac{\nu(A)}{\nu(X)}$, avec ν mesure sur X.

Pour n occurrences dans X, loi binomiale : $N_A \in \{0, n\}$

et $P(N_A = k) = C_n^k p^k (1-p)^{n-k}$ avec $p = \dfrac{\nu(A)}{\nu(X)}$.

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Considérons l'aspect temporel et la RCA dans X sur $[0, t], t > 0$.
Notons $N_X([0, t])$ le nombre d'occurrences sur $[0, t]$ dans X.

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Considérons l'aspect temporel et la RCA dans X sur $[0, t]$, $t > 0$.
Notons $N_X([0, t])$ le nombre d'occurrences sur $[0, t]$ dans X.

Notons T_1 le temps d'attente (aléatoire) de *la première occurrence*.
Notons λ le nombre attendu d'occurrences par unité de temps.

$\lambda = E(N_X([0, t])/t)$= intensité temporelle des occurrences.

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Considérons l'aspect temporel et la RCA dans X sur $[0, t], t > 0$.
Notons $N_X([0, t])$ le nombre d'occurrences sur $[0, t]$ dans X.

Notons T_1 le temps d'attente (aléatoire) de *la première occurrence*.
Notons λ le nombre attendu d'occurrences par unité de temps.

$\lambda = E(N_X([0, t])/t)$= intensité temporelle des occurrences.

$T_1 > t \iff$ aucune occurrence dans $[0, t]$.

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Considérons l'aspect temporel et la RCA dans X sur $[0,t], t>0$.
Notons $N_X([0,t])$ le nombre d'occurrences sur $[0,t]$ dans X.

Notons T_1 le temps d'attente (aléatoire) de *la première occurrence*.
Notons λ le nombre attendu d'occurrences par unité de temps.

$\lambda = E(N_X([0,t])/t)$= intensité temporelle des occurrences.

$$T_1 > t \iff \text{aucune occurrence dans } [0,t].$$

Si $[0,t]$ divisé en m parties égales (m grand), conditions de RCA
alors $P(T_1 > t) \simeq \left(1 - \frac{\lambda t}{m}\right)^m$ (m Bernoulli $\mathcal{B}(\frac{\lambda t}{m})$ indépendantes).

Lois de dénombrement

D'où vient la loi de Poisson (1781-1840) pour le dénombrement de RCA?

Pour une RCA, n n'est pas fixé mais *aléatoire*.

Considérons l'aspect temporel et la RCA dans X sur $[0, t], t > 0$.
Notons $N_X([0, t])$ le nombre d'occurrences sur $[0, t]$ dans X.

Notons T_1 le temps d'attente (aléatoire) de *la première occurrence*.
Notons λ le nombre attendu d'occurrences par unité de temps.

$\lambda = E(N_X([0, t])/t)$= intensité temporelle des occurrences.

$T_1 > t \iff$ aucune occurrence dans $[0, t]$.

Si $[0, t]$ divisé en m parties égales (m grand), conditions de RCA alors $P(T_1 > t) \simeq (1 - \frac{\lambda t}{m})^m$ (m Bernoulli $\mathcal{B}(\frac{\lambda t}{m})$ indépendantes).

$$P(T_1 > t) = \lim_{m \longrightarrow \infty} (1 - \frac{\lambda t}{m})^m = \exp(-\lambda t).$$

Lois de dénombrement

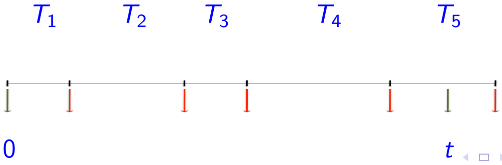
T_1 suit donc une loi exponentielle de paramètre λ car

$$P(T_1 \leq t) = \begin{cases} 1 - \exp(-\lambda t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Lois de dénombrement

T_1 suit donc une loi exponentielle de paramètre λ car

$$P(T_1 \leq t) = \begin{cases} 1 - \exp(-\lambda t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

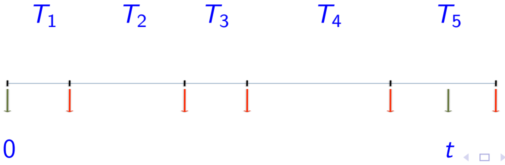


Lois de dénombrement

T_1 suit donc une loi exponentielle de paramètre λ car

$$P(T_1 \leq t) = \begin{cases} 1 - \exp(-\lambda t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Même résultat pour le $k$ième temps d'attente $T_k, k \in \mathbb{N}^*$.



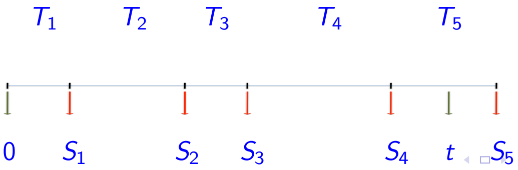
Lois de dénombrement

T_1 suit donc une loi exponentielle de paramètre λ car

$$P(T_1 \leq t) = \begin{cases} 1 - \exp(-\lambda t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Même résultat pour le $k$ième temps d'attente $T_k, k \in \mathbb{N}^*$.

Soit $S_n = \sum_{k=1}^{n} T_k$ la date de la $n$ième occurrence,



Lois de dénombrement

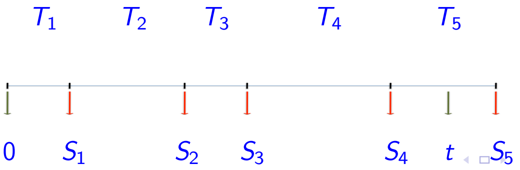
T_1 suit donc une loi exponentielle de paramètre λ car

$$P(T_1 \leq t) = \begin{cases} 1 - \exp(-\lambda t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Même résultat pour le $k$ième temps d'attente $T_k, k \in \boldsymbol{N}^*$.

Soit $S_n = \sum_{k=1}^{n} T_k$ la date de la $n$ième occurrence,

$$P(N_X([0,t]) = n) = P(S_n \leq t < S_{n+1}).$$



Lois de dénombrement

Reste à calculer, pour tout n entier naturel, $P(N_X([0,t]) = n)$

c'est-à-dire $P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k)$ où les T_k sont indépendants

identiquement distribués selon la loi exponentielle de paramètre λ.

Lois de dénombrement

Reste à calculer, pour tout n entier naturel, $P(N_X([0,t]) = n)$

c'est-à-dire $P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k)$ où les T_k sont indépendants identiquement distribués selon la loi exponentielle de paramètre λ.

- Cas $n = 0$: déjà fait car équivaut à $T_1 > t$.

Lois de dénombrement

Reste à calculer, pour tout n entier naturel, $P(N_X([0,t]) = n)$ c'est-à-dire $P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k)$ où les T_k sont indépendants identiquement distribués selon la loi exponentielle de paramètre λ.

- Cas $n = 0$: déjà fait car équivaut à $T_1 > t$.
- Cas $n > 0$:

$$P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k) = \int_{0<s_n\leq t<s_{n+1}} \left(\prod_{k=1}^{n+1} \lambda e^{-\lambda t_k}\right) dt_1 \cdots dt_{n+1}$$

$$= \int_{0<s_1<\cdots<s_n\leq t<s_{n+1}} \lambda^{n+1} e^{-\lambda s_{n+1}} ds_1 \cdots ds_{n+1} \text{ (en posant } s_n = \sum_{k=1}^{n} t_k)$$

$$= \lambda^n \left(\int_{0<s_1<\cdots<s_n\leq t} ds_1 \cdots ds_n\right) \int_t^{+\infty} \lambda e^{-\lambda s_{n+1}} ds_{n+1}$$

$$= \lambda^n \frac{t^n}{n!} e^{-\lambda t} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Lois de dénombrement

Reste à calculer, pour tout n entier naturel, $P(N_X([0,t]) = n)$

c'est-à-dire $P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k)$ où les T_k sont indépendants identiquement distribués selon la loi exponentielle de paramètre λ.

- Cas $n = 0$: déjà fait car équivaut à $T_1 > t$.
- Cas $n > 0$:

$$P(\sum_{k=1}^{n} T_k \leq t < \sum_{k=1}^{n+1} T_k) = \int_{0<s_n\leq t<s_{n+1}} \left(\prod_{k=1}^{n+1} \lambda e^{-\lambda t_k}\right) dt_1 \cdots dt_{n+1}$$

$$= \int_{0<s_1<\cdots<s_n\leq t<s_{n+1}} \lambda^{n+1} e^{-\lambda s_{n+1}} ds_1 \cdots ds_{n+1} \text{ (en posant } s_n = \sum_{k=1}^{n} t_k)$$

$$= \lambda^n \left(\int_{0<s_1<\cdots<s_n\leq t} ds_1 \cdots ds_n\right) \int_t^{+\infty} \lambda e^{-\lambda s_{n+1}} ds_{n+1}$$

$$= \lambda^n \frac{t^n}{n!} e^{-\lambda t} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

donc $N_X([0,t])$ suit la loi $\mathcal{P}(\lambda t)$.

Lois de dénombrement

Revenons au spatial : soit A une partie de l'espace X.

Sachant $N_X([0,t]) = n$, $N_A([0,t])$ suit une binomiale $\mathcal{B}(n, \frac{\nu(A)}{\nu(X)})$.

Lois de dénombrement

Revenons au spatial : soit A une partie de l'espace X.

Sachant $N_X([0,t]) = n,\ N_A([0,t])$ suit une binomiale $\mathcal{B}(n, \dfrac{\nu(A)}{\nu(X)})$.

Pour tout entier naturel k,

$$P(N_A([0,t]) = k) = \sum_{n\in\mathbb{N}} P(N_A([0,t]) = k | N_X([0,t]) = n) P(N_X([0,t]) = n)$$

$$= \sum_{n\in\mathbb{N}} C_n^k \left(\frac{\nu(A)}{\nu(X)}\right)^k \left(1 - \frac{\nu(A)}{\nu(X)}\right)^{n-k} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \frac{(\lambda t \frac{\nu(A)}{\nu(X)})^k}{k!} e^{-\lambda t \frac{\nu(A)}{\nu(X)}}$$

$$= \frac{(\tilde{\lambda} t \nu(A))^k}{k!} e^{-\tilde{\lambda} t \nu(A)}$$

Lois de dénombrement

Revenons au spatial : soit A une partie de l'espace X.

Sachant $N_X([0,t]) = n$, $N_A([0,t])$ suit une binomiale $\mathcal{B}(n, \dfrac{\nu(A)}{\nu(X)})$.

Pour tout entier naturel k,

$$P(N_A([0,t]) = k) = \sum_{n\in\mathbb{N}} P(N_A([0,t]) = k | N_X([0,t]) = n) P(N_X([0,t]) = n)$$

$$= \sum_{n\in\mathbb{N}} C_n^k \left(\frac{\nu(A)}{\nu(X)}\right)^k \left(1 - \frac{\nu(A)}{\nu(X)}\right)^{n-k} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \frac{(\lambda t \frac{\nu(A)}{\nu(X)})^k}{k!} e^{-\lambda t \frac{\nu(A)}{\nu(X)}}$$

$$= \frac{(\tilde{\lambda} t \nu(A))^k}{k!} e^{-\tilde{\lambda} t \nu(A)}$$ donc $N_A([0,t])$ suit la loi de Poisson $\mathcal{P}(\tilde{\lambda} t \nu(A))$

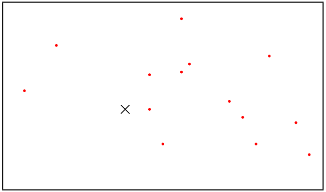
avec $\tilde{\lambda} = \dfrac{E(N_X([0,t]))}{t\nu(X)} = \dfrac{E(N_X([0,1]))}{\nu(X)} = \dfrac{\lambda}{\nu(X)}$.

$\tilde{\lambda}$ est l'intensité spatio-temporelle des occurrences

ou encore l'intensité spatiale par unité de temps.

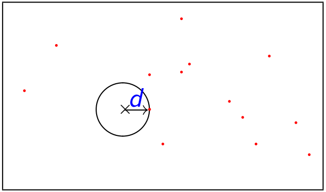
Loi de distance au plus proche voisin

Soit x un point quelconque de $\mathbb{R}^2$, et d un élément de $\mathbb{R}_+$.



Loi de distance au plus proche voisin

Soit x un point quelconque de $\mathbb{R}^2$, et d un élément de $\mathbb{R}_+$.
D_x la distance de x à la plus proche occurrence vérifie

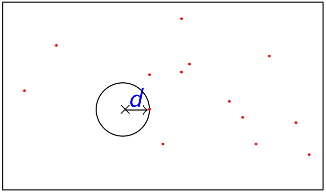


Loi de distance au plus proche voisin

Soit x un point quelconque de $\mathbb{R}^2$, et d un élément de $\mathbb{R}_+$.
D_x la distance de x à la plus proche occurrence vérifie

$$P(D_x \leq d) = 1 - P(D_x > d) = 1 - \exp(-\tilde{\lambda}\pi d^2)$$

$\forall x \in \mathbb{R}^2,\ D_x$ suit la loi de Rayleigh d'espérance $\dfrac{1}{2\sqrt{\tilde{\lambda}}}$.

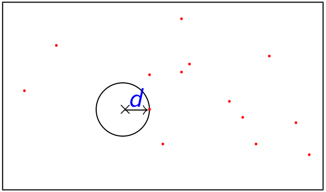


Loi de distance au plus proche voisin

Soit x un point quelconque de $\mathbb{R}^2$, et d un élément de $\mathbb{R}_+$.
D_x la distance de x à la plus proche occurrence vérifie

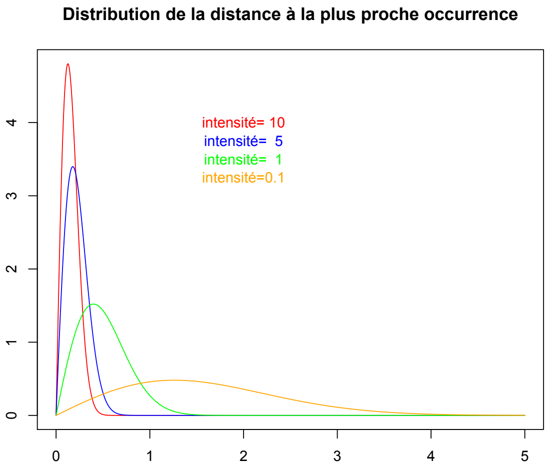
$$P(D_x \leq d) = 1 - P(D_x > d) = 1 - \exp(-\tilde{\lambda}\pi d^2)$$

$\forall x \in \mathbb{R}^2,\ D_x$ suit la loi de Rayleigh d'espérance $\dfrac{1}{2\sqrt{\tilde{\lambda}}}$.



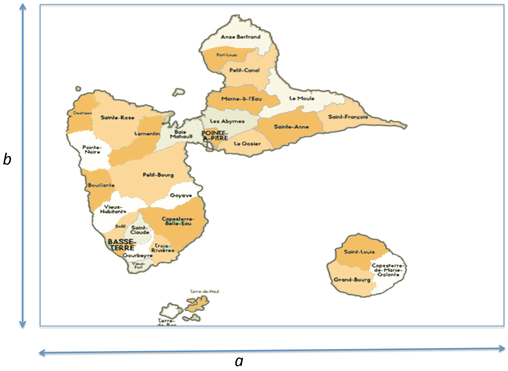
Si $X \subset \mathbb{R}^2$ borné, attention!!
x en bordure et d grand $\Rightarrow B(x,d) \not\subseteq X$.

Loi de distance au plus proche voisin



Algorithme de simulation d'1 RCA spatiale

On suppose X borné, inclus dans rectangle de longueur a et de largeur b. Pour une unité de temps, $\tilde{\lambda}$ est l'intensité spatiale (nombre attendu d'occurrences dans X par unité de surface).



Algorithme de simulation d'1 RCA spatiale

Note sur la Méthode de simulation par inversion

Notons $\mathcal{U}_{[0,1]}$ la loi uniforme sur $[0,1]$.

Soit F une fonction de répartition d'une loi de probabilité sur $\boldsymbol{R}$ avec F inversible, soit U une variable aléatoire de loi $\mathcal{U}_{[0,1]}$,

Algorithme de simulation d'1 RCA spatiale

Note sur la Méthode de simulation par inversion

Notons $\mathcal{U}_{[0,1]}$ la loi uniforme sur $[0, 1]$.

Soit F une fonction de répartition d'une loi de probabilité sur $\boldsymbol{R}$ avec F inversible, soit U une variable aléatoire de loi $\mathcal{U}_{[0,1]}$,

- alors la variable aléatoire $F^{-1}(U)$ est de loi définie par F.

Algorithme de simulation d'1 RCA spatiale

Note sur la Méthode de simulation par inversion

Notons $\mathcal{U}_{[0,1]}$ la loi uniforme sur $[0,1]$.

Soit F une fonction de répartition d'une loi de probabilité sur $\boldsymbol{R}$ avec F inversible, soit U une variable aléatoire de loi $\mathcal{U}_{[0,1]}$,

- alors la variable aléatoire $F^{-1}(U)$ est de loi définie par F.
- La loi exponentielle de paramètre θ rentre dans ce cadre :

  Pour $t \geq 0,\ F(t) = 1 - \exp(-\theta t) = y$ implique $t = -\frac{1}{\theta}\log(1-y)$.

  Autre remarque : U de loi $\mathcal{U}_{[0,1]}$ équivaut à $1-U$ de loi $\mathcal{U}_{[0,1]}$.

Algorithme de simulation d'1 RCA spatiale

Note sur la Méthode de simulation par inversion

Notons $\mathcal{U}_{[0,1]}$ la loi uniforme sur $[0,1]$.

Soit F une fonction de répartition d'une loi de probabilité sur $\boldsymbol{R}$ avec F inversible, soit U une variable aléatoire de loi $\mathcal{U}_{[0,1]}$,

- alors la variable aléatoire $F^{-1}(U)$ est de loi définie par F.
- La loi exponentielle de paramètre θ rentre dans ce cadre :

  Pour $t \geq 0,\ F(t) = 1 - \exp(-\theta t) = y$ implique $t = -\dfrac{1}{\theta}\log(1-y)$.

  Autre remarque : U de loi $\mathcal{U}_{[0,1]}$ équivaut à $1-U$ de loi $\mathcal{U}_{[0,1]}$.
- Par conséquent, U de loi $\mathcal{U}_{[0,1]}$ implique

  $-\dfrac{1}{\theta}\log(U)$ de loi exponentielle de paramètre θ.

Algorithme de simulation d'1 RCA spatiale

Posons $\theta = \tilde{\lambda} ab$, nombre attendu d'occurrences dans le rectangle d'étude.

- Principe de simulation :

  - Générer des réalisations de loi exponentielle de paramètre θ jusqu'à ce que leur somme soit supérieure à 1.

  - Générer des réalisations de loi uniforme sur le rectangle d'étude.

Algorithme de simulation d'1 RCA spatiale

Posons $\theta = \tilde{\lambda} ab$, nombre attendu d'occurrences dans le rectangle d'étude.

- Principe de simulation :

  - Générer des réalisations de loi exponentielle de paramètre θ jusqu'à ce que leur somme soit supérieure à 1.

  - Générer des réalisations de loi uniforme sur le rectangle d'étude.

- Algorithme de simulation :

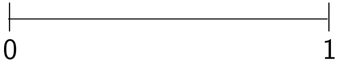
  - Générer u_i selon $\mathcal{U}_{[0,1]}$ et calculer $t_i = -\frac{1}{\theta}\log(u_i)$ jusqu'à ce que $t_1 + \cdots + t_i > 1$. On note n le nombre de t_i avant la règle d'arrêt.

  Génération des positions des occurrences

  - Générer $v_1, \cdots, v_{2n}$ selon $\mathcal{U}_{[0,1]}$.

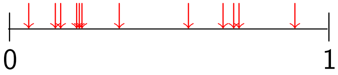
  (av_{2i-1}, bv_{2i}) est la position de la $i$ème occurrence.

Algorithme de simulation d'1 RCA spatiale



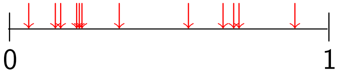
Simulations des temps d'attente selon la loi exponentielle

Algorithme de simulation d'1 RCA spatiale

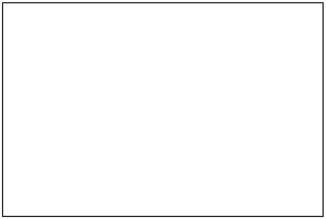


Simulations des temps d'attente selon la loi exponentielle

Algorithme de simulation d'1 RCA spatiale

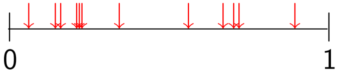


Simulations des temps d'attente selon la loi exponentielle

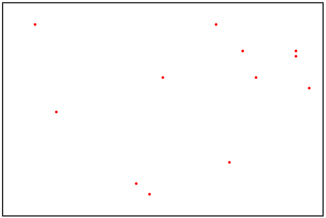


Simulations des positions spatiales selon la loi uniforme sur le rectangle

Algorithme de simulation d'1 RCA spatiale

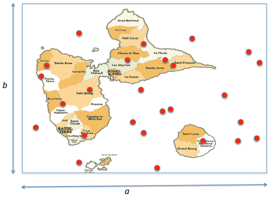


Simulations des temps d'attente selon la loi exponentielle

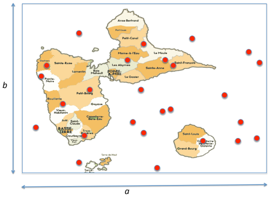


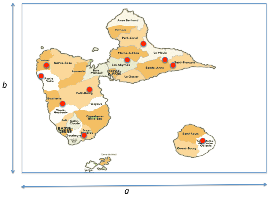
Simulations des positions spatiales selon la loi uniforme sur le rectangle

Algorithme de simulation d'1 RCA spatiale

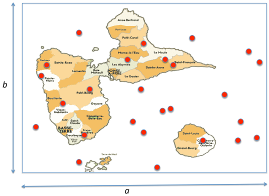


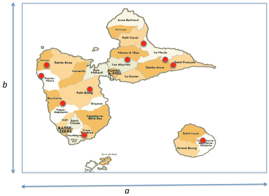
Algorithme de simulation d'1 RCA spatiale

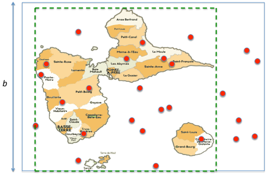




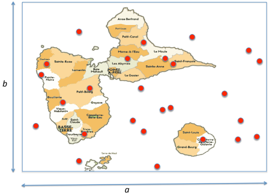
Algorithme de simulation d'1 RCA spatiale

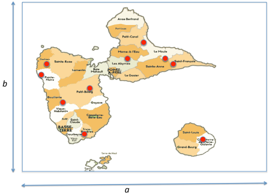




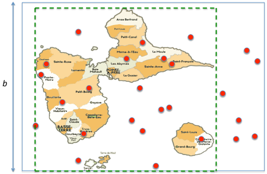


Algorithme de simulation d'1 RCA spatiale





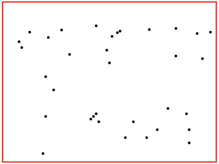
(1) Restriction de RCA est RCA, de même intensité. (2) Ne pas surdimensionner le rectangle instrumental



Type de données observées

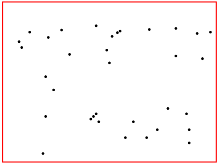
Type de données observées

- Observation complète : toutes les positions d'occurrence sont enregistrées



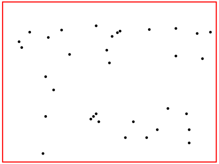
Type de données observées

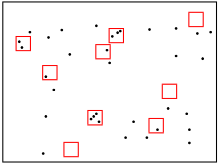
- Observation complète : toutes les positions d'occurrence sont enregistrées
- Observation partielle : deux méthodes usuelles



Type de données observées

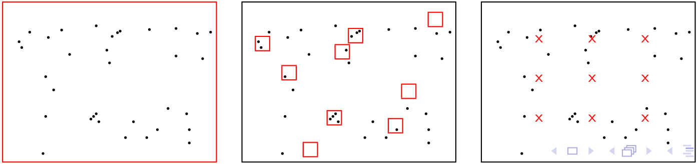
- Observation complète : toutes les positions d'occurrence sont enregistrées
- Observation partielle : deux méthodes usuelles
  - comptage des occurrences dans des unités d'observation disposées selon une procédure d'échantillonnage spatial





Type de données observées

- Observation complète : toutes les positions d'occurrence sont enregistrées
- Observation partielle : deux méthodes usuelles
  - comptage des occurrences dans des unités d'observation disposées selon une procédure d'échantillonnage spatial
  - mesures de distance entre des points d'échantillonnage spatial et la plus proche occurrence observée



Test rapport variance-moyenne

Critère de test basé sur : loi de Poisson $\Rightarrow$ variance $=$ espérance.

Pour des comptages $N_1, \cdots N_k$ dans k unités de même taille s,

Intensité estimée par $\overline{N}/s$ où $\overline{N} = \sum_{i=1}^{k} N_i/k$ $=$ moyenne des comptages.

Calcul du rapport variance sur moyenne observées :

$$\widehat{I}_d = \frac{\sum_{i=1}^{k} (N_i - \overline{N})^2}{(k-1)\overline{N}}$$ appelé indice de dispersion.

Il estime le rapport variance sur espérance I_d de la loi de dénombrement.

Si RCA, $(k-1)\widehat{I}_d$ suit χ^2_{k-1}, loi du Khi-deux à $k-1$ degrés de liberté.

Test rapport variance-moyenne

Critère de test basé sur : loi de Poisson $\Rightarrow$ variance $=$ espérance.

Pour des comptages $N_1, \cdots N_k$ dans k unités de même taille s,

Intensité estimée par $\overline{N}/s$ où $\overline{N} = \sum_{i=1}^{k} N_i/k$ $=$ moyenne des comptages.

Calcul du rapport variance sur moyenne observées :

$$\widehat{I}_d = \frac{\sum_{i=1}^{k}(N_i - \overline{N})^2}{(k-1)\overline{N}}$$ appelé indice de dispersion.

Il estime le rapport variance sur espérance I_d de la loi de dénombrement.

Si RCA, $(k-1)\widehat{I}_d$ suit χ^2_{k-1}, loi du Khi-deux à $k-1$ degrés de liberté.

Règle de décision au niveau de signification α :

- Rejet de l'hypothèse de RCA si $(k-1)\widehat{I}_d > \chi^2_{k-1,1-\alpha}$
- Non rejet de l'hypothèse RCA si $(k-1)\widehat{I}_d \leq \chi^2_{k-1,1-\alpha}$

où $\chi^2_{k-1,1-\alpha}$ fractile d'ordre $1-\alpha$ de la loi χ^2_{k-1}.

Exemple numérique

Nombre d'unités observées $k = 12$

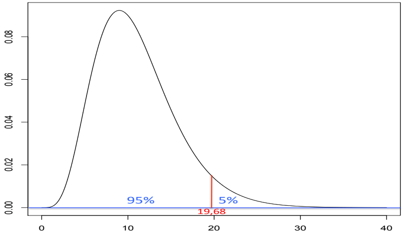
Indice de dispersion calculé $\widehat{I_d}$ =1,73

Niveau de signification choisi α =0,05

Critère de test $(k-1)\widehat{I_d}$ =19,03

Fractile seuil $\chi^2_{11;0,95}$ =19,68

Conclusion : RCA non rejetée au niveau 5%.



Test des distances

Mesures de distance $D_1, \cdots, D_k$ de la plus proche occurrence à k points d'échantillonnage.

Estimation par maximum de vraisemblance de l'intensité : $\dfrac{1}{\dfrac{\pi}{k}\displaystyle\sum_{i=1}^{k} D_i^2}$.

Calcul de l'indice de Johnson-Zimmer : $\widehat{I}_{JZ} = \dfrac{(k+1)\displaystyle\sum_{i=1}^{k} D_i^4}{\left(\displaystyle\sum_{i=1}^{k} D_i^2\right)^2}$

proche de 2 sous RCA (car D_i^2 suit la loi exponentielle).

Critère de test : $Z = \dfrac{\widehat{I}_{JZ} - 2}{\sqrt{\dfrac{4(k-1)}{(k+2)(k+3)}}}$

Règle décisionnelle

Règle de décision au niveau de signification α :

- Rejet de l'hypothèse de RCA si $|Z| > u_{1-\alpha/2}$
- Non rejet de l'hypothèse RCA si $|Z| \leq u_{1-\alpha/2}$

où $u_{1-\alpha/2}$ fractile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$.

Exemple numérique

Nombre de points d'échantillonnage $k = 100$

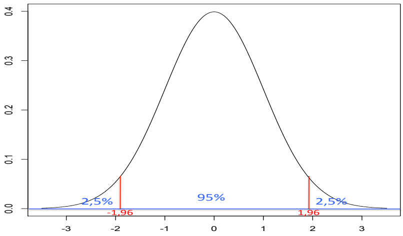
Indice de Johnson-Zimmer calculé $\widehat{I}_{JZ}$ =2,41

Niveau de signification choisi α =0,05

Critère de test Z =(2,41-2)/0,194 = 2,113

Fractile seuil $u_{0,975}$ =1,96

Conclusion : RCA rejetée au niveau 5%.



Test de Monte-Carlo

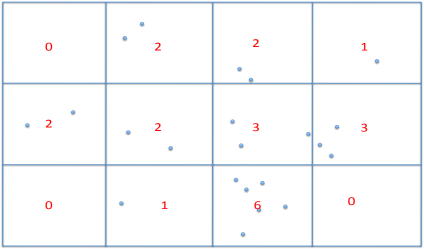
Données : carte complète avec n occurrences.

Principe : approcher la loi de probabilité, sous RCA, d'un critère pertinent C, par simulations numériques.

- Simuler N fois RCA en redistribuant les n occurrences uniformément
- Calculer pour chacune des N simulations le critère de test C
- Comparer la valeur C_{obs} calculée à partir des données à la distribution des valeurs simulées de C
- Rejeter RCA si C_{obs} est **parmi les valeurs critiques** de la distribution simulée.

UNIVERSITE DES ANTILLES ET DE LA GUYANE

Exemple numérique

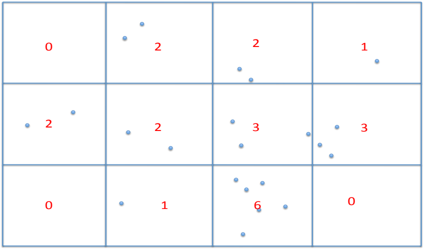


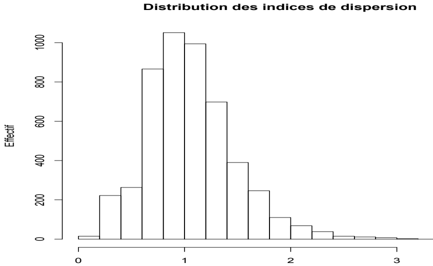
Nombre d'occurrences $n = 22$; nombre de quadrats $k = 12$

Critère observé : rapport variance-moyenne $\widehat{I_d}$ =1,713

Niveau de signification α =0,05; Nombre de simulations N =5000

Exemple numérique





Nombre d'occurrences $n = 22$; nombre de quadrats $k = 12$

Critère observé : rapport variance-moyenne $\widehat{I_d}$ =1,713

Niveau de signification α =0,05; Nombre de simulations N =5000

Rang du critère observé parmi les simulés : 4622

Probabilité critique : 0,0758 = (5001-4622)/5001

Conclusion : RCA n'est pas rejetée au niveau 5%.

Exemple numérique

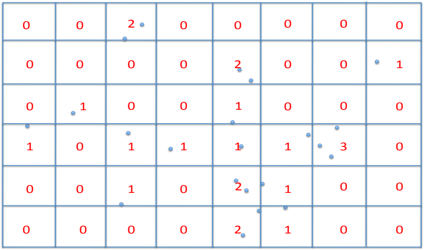
0	0	2	0	0	0	0	0
0	0	0	0	2	0	0	1
0	1	0	0	1	0	0	0
1	0	1	1	1	1	3	0
0	0	1	0	2	1	0	0
0	0	0	0	2	1	0	0

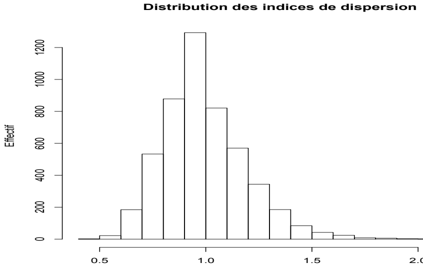
Nombre d'occurrences $n = 22$; nombre de quadrats $k = 48$

Critère observé : rapport variance-moyenne $\widehat{I_d}$ =1,229

Niveau de signification α =0,05; Nombre de simulations N =5000

Exemple numérique





Nombre d'occurrences $n = 22$; nombre de quadrats $k = 48$

Critère observé : rapport variance-moyenne $\widehat{I_d}$ =1,229

Niveau de signification α =0,05; Nombre de simulations N =5000

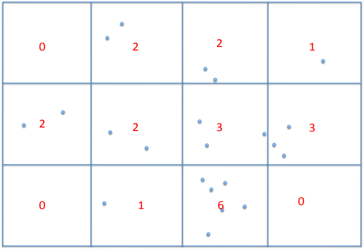
Rang du critère observé parmi les simulés : 4325

Probabilité critique : 0,1352 = (5001-4325)/5001

Conclusion : RCA n'est pas rejetée au niveau 5%.

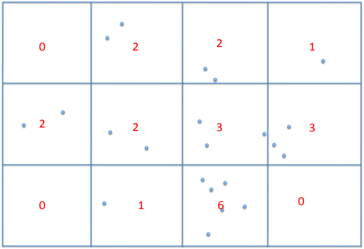
Exemple numérique

Autre critère possible : autocorrélation spatiale (indice de Moran).
But : Mieux prendre en compte la disposition spatiale.



Exemple numérique

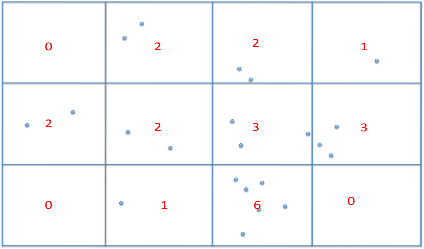
Autre critère possible : autocorrélation spatiale (indice de Moran).
But : Mieux prendre en compte la disposition spatiale.



0	0	2	2
0	2	2	3
1	1	3	6

Deux configurations avec comptages identiques
mais avec une disposition spatiale différente!!!

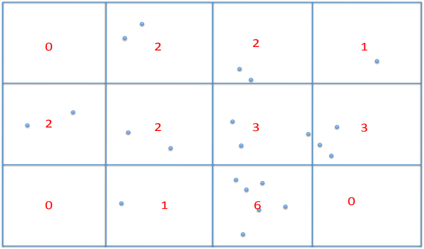
Exemple numérique

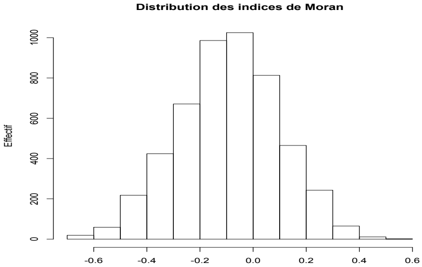


Nombre d'occurrences $n = 22$; nombre de quadrats $k = 12$

Critère observé : indice de Moran $\widehat{I_M} =$ -0,223

Exemple numérique





Nombre d'occurrences $n = 22$; nombre de quadrats $k = 12$

Critère observé : indice de Moran $\widehat{I_M} =$ -0,223

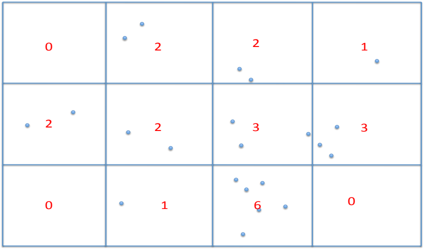
Niveau de signification $\alpha =$0,05; Nombre de simulations $N =$5000

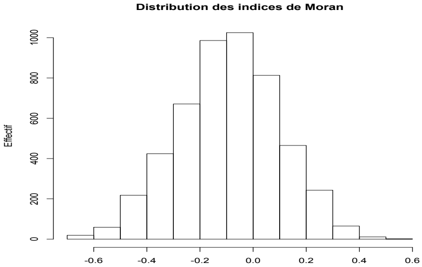
Rang du critère observé parmi les simulés : 1221

Probabilité critique : 0,756

Conclusion : RCA n'est pas rejetée au niveau 5%

Exemple numérique





Nombre d'occurrences $n = 22$; nombre de quadrats $k = 12$

Critère observé : indice de Moran $\widehat{I_M}$ = -0,223 (=0,464 pour cas 2)

Niveau de signification α =0,05; Nombre de simulations N =5000

Rang du critère observé parmi les simulés : 1221 (=4999 pour cas 2)

Probabilité critique : 0,756 (=0,0004 pour cas 2)

Conclusion : RCA n'est pas rejetée au niveau 5%. (Rejet dans cas 2).

Remarques combinatoires

- Nombre total de redistributions de 22 occurrences dans 12 unités :

$$C_{22+11}^{11} = 193536720$$

- Nombre total de redistributions de 22 occurrences dans 48 unités :

$$C_{22+47}^{47} = 5,8867 \times 10^{17}$$

- Nombre total de permutations de 12 unités : $12! = 479001600$

Remarques combinatoires

- Nombre total de redistributions de 22 occurrences dans 12 unités :

$$C_{22+11}^{11} = 193536720$$

- Nombre total de redistributions de 22 occurrences dans 48 unités :

$$C_{22+47}^{47} = 5,8867 \times 10^{17}$$

- Nombre total de permutations de 12 unités : $12! = 479001600$
- Théorie de l'échantillonnage $\Rightarrow$ test fiable pour $N \simeq 1000$.
  Pas besoin d'un taux de sondage élevé.

Récapitulatif des lois usuelles associées à RCA

- Position spatiale d'une occurrence sur X :
  Loi uniforme sur X.
- Nombre d'occurrences dans A partie mesurable de X conditionnellement à n occurrences dans X :
  Loi binomiale de paramètres n et $\dfrac{\nu(A)}{\nu(X)}$.
- Nombre d'occurrences dans A en une unité de temps :
  Loi de Poisson de paramètre $\tilde{\lambda}\nu(A)$
  avec $\tilde{\lambda}=$ intensité spatio-temporelle.
- Temps d'attente d'une occurrence :
  Loi exponentielle d'espérance $1/\lambda$.
- Distance à la plus proche occurrence :
  Loi de Rayleigh d'espérance $1/(2\sqrt{\tilde{\lambda}})$.

Plus formellement

Une RCA est une réalisation d'un processus de Poisson stationnaire (ou homogène)!

Un processus de Poisson homogène $N(.)$ sur un espace mesuré $(X, \mathcal{A}, \nu)$ est une mesure stochastique de comptage (c'est-à-dire mesure à valeurs dans $\mathbb{N}$) vérifiant

(1) $\forall A \in \mathcal{A},\ N(A) \sim \mathcal{P}(\lambda\nu(A))$ où $\lambda \in \mathbb{R}_+^*$ est l'intensité constante du processus

(2) $\forall (A_1, A_2) \in \mathcal{A}^2$, si $A_1 \cap A_2 = \emptyset$ alors $N(A_1)$ et $N(A_2)$ sont indépendants.

(1) et (2) correspondent aux conditions de RCA présentées auparavant; (1) à l'homogénéité, (2) à l'indépendance.

Plus formellement

Une RCA est une réalisation d'un processus de Poisson stationnaire (ou homogène)!

Un processus de Poisson homogène $N(.)$ sur un espace mesuré $(X, \mathcal{A}, \nu)$ est une mesure stochastique de comptage (c'est-à-dire mesure à valeurs dans $\mathbb{N}$) vérifiant

(1) $\forall A \in \mathcal{A},\ N(A) \sim \mathcal{P}(\lambda\nu(A))$ où $\lambda \in \mathbb{R}_+^*$ est l'intensité constante du processus

(2) $\forall (A_1, A_2) \in \mathcal{A}^2$, si $A_1 \cap A_2 = \emptyset$ alors $N(A_1)$ et $N(A_2)$ sont indépendants.

(1) et (2) correspondent aux conditions de RCA présentées auparavant; (1) à l'homogénéité, (2) à l'indépendance.

MERCI DE VOTRE ATTENTION !