

# Estimation de l'intensité d'un processus ponctuel appliqué à la vision de nouveaux-nés



Antoine Sétif,  
sous la direction d'Adeline Leclercq-Samson et de Jean-François Coeurjolly

# Plan de la soutenance

- Motivations du stage
  - Une collaboration entre 2 laboratoires (LJK, BabyLab)
  - Relation entre données oculométriques et processus ponctuels spatiaux
- Rappels sur les processus ponctuels spatiaux
  - Processus ponctuel de Poisson et intensité
  - Estimation paramétrique d'un processus ponctuel de poisson
- Application au jeu de données
  - Description et visualisation du jeu de données
  - Analyse de la distribution des processus ponctuels
  - Définition des covariables spatiales
  - Analyse des données par un modèle log-linéaire et tests

## Une collaboration entre 2 laboratoires (LJK, BabyLab)

Le laboratoire Jean Kuntzmann (**LJK**) est structuré en 3 départements :

- Le département "*Géométrie-Image*"
- Le département "*Modèles et Algorithmes Déterministes*"
- Le département "*Probabilités/Statistique*" : probabilistes, statisticiens et spécialistes de l'analyse des données et du traitement du signal.

Le centre d'études grenoblois du nourrisson et du jeune enfant (**BabyLab**) :

- Spécialisé dans le développement des capacités sensorielles et de la motricité.
- Techniques d'études reposant sur l'observation visuelle, auditive ou motrice.

## Relation entre mouvements oculaires et processus ponctuels spatiaux

- Vision oculaire : succession de mouvements rapides (saccades) et de fixations.
- A chaque fixation est associée une entité géographique.
- Modèle statistique adapté : processus ponctuel spatiaux (interaction géométrie stochastique et statistique spatiale).
- Statistique pour les processus ponctuels : modèle linéaire généralisé (GLM), modèle mixte (GLMM), package spatstat.

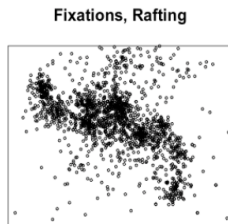
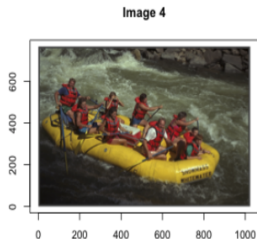


FIGURE: Fixations associées à l'image Rafting

# Processus ponctuel de Poisson (PPP) et intensité

## Propriétés :

Soit  $X$  suivant une loi de Poisson( $\rho$ ), observable dans  $W \subseteq S$ , alors :

- L'estimateur naturel de l'intensité  $\rho$  dans le cas Poisson homogène est :  $\hat{\rho} = \frac{N(W)}{|W|}$
- La densité d'un processus Poisson inhomogène est donnée par :

$$f : N_{lf} \rightarrow \mathbb{R}$$
$$x = \{x_1, \dots, x_n\} \rightarrow e^{|W| - \int_W \rho(u) du} \prod_{u \in x} \rho(u)$$

où  $N_{lf}$  est l'espace contenant l'ensemble des configurations localement finies.

## Estimation paramétrique de l'intensité d'un PPP

### Cas modèle linéaire généralisé :

Soit la fonction d'intensité  $\rho(u) = \rho_\theta(u) = e^{(\theta^\top z(u))}$  où :

- $\theta$  est le vecteur des paramètres  $\in \mathbb{R}^d$
- $z(u) = (z_1(u), \dots, z_p(u))^\top$  le vecteur des covariables spatiales.

La vraisemblance d'un modèle de Poisson, pour  $x = \{x_1, \dots, x_n\}$  sur  $W$  est :

$$L_W(\theta) = e^{|W| - \int_W \rho_\theta(u) du} \prod_{i=1}^{n(X)} \rho_\theta(x_i)$$

Soit la log-vraisemblance suivante :

$$\log L_W(\theta) = |W| - \int_W \rho_\theta(u) du + \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) = \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \int_W \rho_\theta(u) du$$

## Estimation paramétrique de l'intensité d'un PPP

### Approche Berman-Turner

L'estimation du maximum de vraisemblance de  $\theta$  est obtenue en maximisant :

$$\log L_W(\theta) = \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \int_W \rho_\theta(u) du$$

Berman et Turner (1992) ont développé un programme en utilisant des méthodes de quadrature. L'idée étant de discrétiser l'intégrale par une somme pondérée.

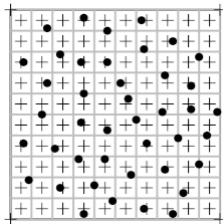


FIGURE: Exemple de schéma en quadrature

Données (•), dummy points (+) et bordures des cellules (lignes grises)

## Estimation paramétrique de l'intensité d'un PPP

L'intégrale peut alors être approchée par la quadrature :

$$\int_W \rho_\theta(u) du \approx \sum_{j=1}^m \rho_\theta(u_j) w_j$$

La log vraisemblance peut alors être exprimée par :

$$\log L_W(\theta) \approx \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \sum_{j=1}^m \rho_\theta(u_j) w_j$$

Soit  $z_j$  une indicatrice des données observées et  $y_j = z_j/w_j$ , alors :

$$\log L_W(\theta) \approx \sum_{j=1}^m (y_j \log \rho_\theta(u_j) - \rho_\theta(u_j)) w_j$$

→ Log vraisemblance pondérée Poissonnienne : fonction lien, la fonction log.

Attention : La variable  $y_j$  n'est pas de type entière → **quasi-vraisemblance**



## Description du jeu de données

Expérience : 5 groupes, 6 images

- 44 bébés de 3 mois, 47 de 6 mois, 44 de 9 mois, 46 de 12 mois et 44 adultes.
- 6 images de scènes naturelles.

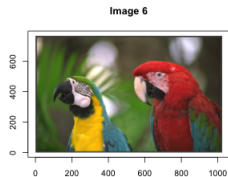
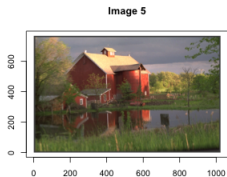
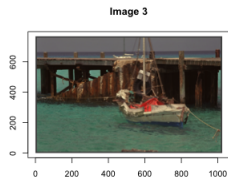
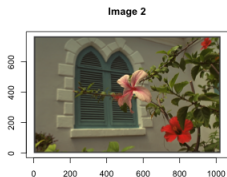
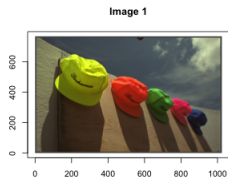


FIGURE: Les 6 images de l'expérience

## Visualisation du jeu de données

- Images en format **bitmap**
- Concentration de fixations autour de zones bien précises.

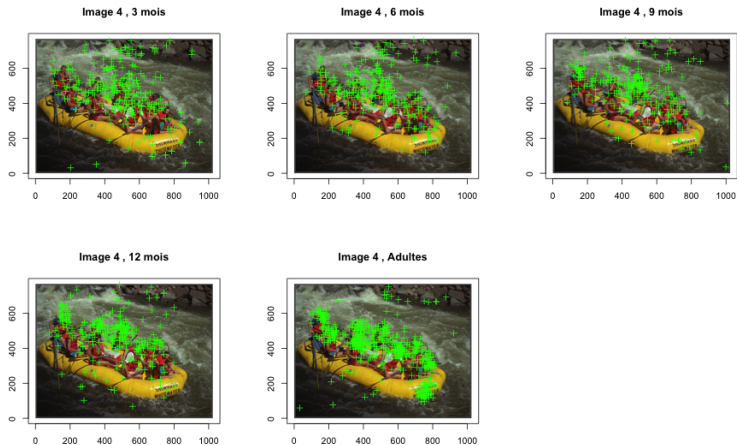
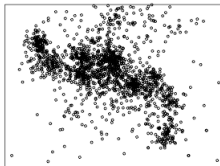


FIGURE: Fixations par groupe d'âge, image Rafting

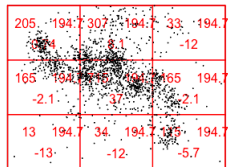
## Distribution du processus ponctuel spatial (1)

- Test de l'homogénéité poissonnienne :

Fixations, Rafting



Test quadrat



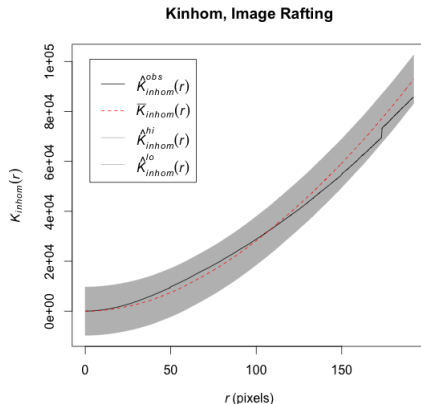
```
quadrat.test(unmark(X_Im4),nx=3)

##
## Chi-squared test of CSR using quadrat counts
## Pearson X2 statistic
##
## data: unmark(X_Im4)
## X2 = 1934.24, df = 8, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 3 by 3 grid of tiles
```

- On rejette l'hypothèse d'homogénéité poissonnienne.
- Même conclusion pour toutes les images.

## Distribution du processus ponctuel spatial (2)

- Test de l'inhomogénéité poissonnienne : appel à la fonction  $K_{inhom}$  de Ripley.
- Simulation d'un certain nombre de processus de Poisson inhomogène ( $H_0$ ).



- Notre procesus peut être considéré comme un processus de Poisson inhomogène.
- Raisonnement applicable à toutes les images, à tous les groupes d'âge.

## Covariables spatiales - Cartes de saillance

- Une haute saillance signifie une forte probabilité de fixation.
- Carte de saillance "moyenne" construite à partir de 6 niveaux de saillance :

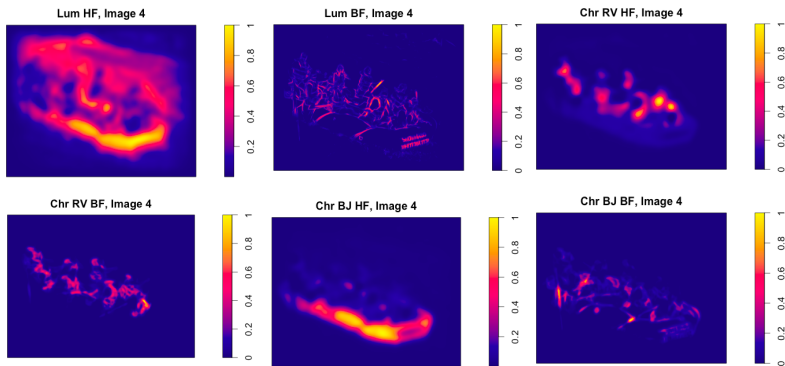
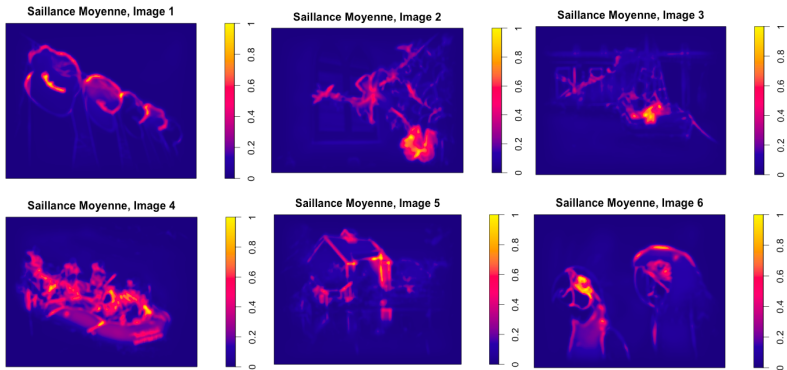


FIGURE: 6 cartes de saillances, image Rafting

## Covariables spatiales - Cartes de saillance

- A partir de ces 6 cartes, construction de 6 cartes de saillance "moyenne" :



- Les cartes de saillance ont été normalisées.

## Covariables spatiales - Zones cognitives

- Recherche d'une covariable spatiale permettant de mesurer la variabilité d'acuité visuelle entre les 4 groupes d'âges des bébés.
- Zone cognitive : zone qui a attiré le plus l'attention d'un adulte.
- Zones cognitives construites en estimant non paramétriquement l'intensité du groupe adultes (via la fonction **density.ppp** de R).
- Création d'une carte binaire à partir des 5% pixels les plus intenses.

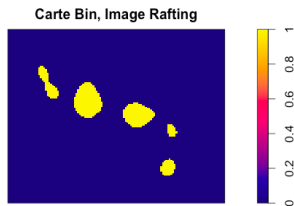


FIGURE: Carte binaire, top 5%, image rafting

# Analyse par un modèle log-linéaire

## Présentation des objets R utilisés

```
## [[1]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 309
## 2 (ppp) 6mois  (im) (im) 338
## 3 (ppp) 9mois  (im) (im) 299
## 4 (ppp) 12mois (im) (im) 323
## 5 (ppp) Adulte (im) (im) 632
##
## [[2]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 265
## 2 (ppp) 6mois  (im) (im) 279
## 3 (ppp) 9mois  (im) (im) 246
## 4 (ppp) 12mois (im) (im) 318
## 5 (ppp) Adulte (im) (im) 561

## [[3]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 260
## 2 (ppp) 6mois  (im) (im) 223
## 3 (ppp) 9mois  (im) (im) 240
## 4 (ppp) 12mois (im) (im) 269
## 5 (ppp) Adulte (im) (im) 583
##
## [[4]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 288
## 2 (ppp) 6mois  (im) (im) 292
## 3 (ppp) 9mois  (im) (im) 280
## 4 (ppp) 12mois (im) (im) 334
## 5 (ppp) Adulte (im) (im) 631

## [[5]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 265
## 2 (ppp) 6mois  (im) (im) 276
## 3 (ppp) 9mois  (im) (im) 228
## 4 (ppp) 12mois (im) (im) 355
## 5 (ppp) Adulte (im) (im) 609
##
## [[6]]
## Hyperframe:
##      ppp  group ZCBin sail  n
## 1 (ppp) 3mois  (im) (im) 288
## 2 (ppp) 6mois  (im) (im) 294
## 3 (ppp) 9mois  (im) (im) 274
## 4 (ppp) 12mois (im) (im) 354
## 5 (ppp) Adulte (im) (im) 549
```

FIGURE: Hyperframes construites



# Analyse par un modèle log-linéaire

## Présentation des objets R utilisés

Par exemple, pour le groupe adultes sur l'image rafting :

```
str(X_adults_Im[[4]])

## List of 6
## $ window      :List of 4
## ..$ type      : chr "rectangle"
## ..$ xrange: num [1:2] 0 1024
## ..$ yrange: num [1:2] 0 768
## ..$ units :List of 3
## .. ..$ singular : chr "pixels"
## .. ..$ plural    : chr "pixels"
## .. ..$ multiplier: num 1
## .. ..- attr(*, "class")= chr "units"
## ..- attr(*, "class")= chr "owin"
## $ n           : int 631
## $ x           : int [1:631] 443 519 368 365 222 213 154 386 367 687 ...
## $ y           : num [1:631] 348 452 421 451 487 501 582 544 531 315 ...
## $ markformat: chr "vector"
## $ marks       : Factor w/ 43 levels "Sujet 181","Sujet 182",...: 1 1 1 1 1 1 1 1 1 1
## - attr(*, "class")= chr "ppp"
```

# Analyse par un modèle log-linéaire

## Ecriture du modèle

### Objectif :

Evaluer les différences d'intensités (différences d'emplacement des fixations) entre les 4 groupes d'âge des bébés.

Pour chaque image, on construit le modèle suivant :

$$\log \rho(u, m, \theta) = \theta_1 \text{Saillance}(u) + \sum_{m'=1}^4 \left( \theta_0^{m'} + \theta_1^{m'} \text{ZCBin}(u) \right) \mathbb{1}_{m'=m}$$

où :

- $u$  est un point de  $R^2$
- $m$  est un groupe d'âge
- *Saillance* et *ZCBin* sont les covariables spatiales précédemment décrites

# Analyse par un modèle log-linéaire

## Fonction MPPM du package spatstat

- Application de la fonction MPPM :

```
fit <- list()
for(i in 1:6) {
  fit[[i]] <- mppm(ppp~sail+group*ZCBin-ZCBin,data=hyp.Im[[i]])
}
```

- 9 coefficients, on ne s'intéresse qu'aux 4 nous permettant de mesurer la variabilité entre les groupes d'âges.

```
summary(fit[[4]])$coef
```

|    |                  |                   |                  |                  |
|----|------------------|-------------------|------------------|------------------|
| ## | (Intercept)      | sail              | group3mois       | group6mois       |
| ## | -8.23463633      | 2.03504193        | -0.01874119      | -0.05988180      |
| ## | group9mois       | group12mois:ZCBin | group3mois:ZCBin | group6mois:ZCBin |
| ## | -0.07942860      | 1.44696303        | 0.85964130       | 1.17161475       |
| ## | group9mois:ZCBin |                   |                  |                  |
| ## | 1.01592721       |                   |                  |                  |

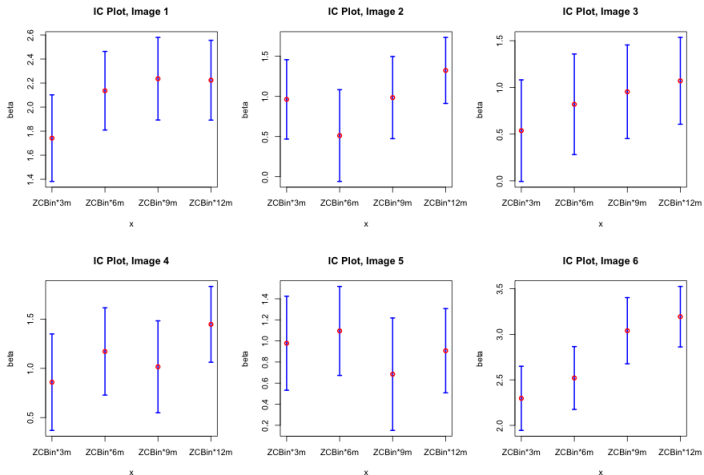
- Méthode d'estimation utilisée par la fonction **mppm** : quasi-vraisemblance.

```
## Call:
## glm(formula = fmla, family = quasi(link = log, variance = mu),
## data = moadf, weights = .mpl.W, subset = (.mpl.SUBSET ==
## "TRUE"), control = glm.control(maxit = 50))
```

# Analyse par un modèle log-linéaire

## IC Plot

A l'aide du package **plotrix**, représentation des coefficients estimés avec leur intervalle de confiance (à 95%) :



## Analyse par un modèle log-linéaire

### Tests

- Différences significatives dans 3 situations ?
  - groupe 12 mois contre groupe 3 mois
  - groupe 9 et 12 mois réunis contre groupe 3 et 6 mois réunis
  - groupe 12 mois contre groupe 3, 6 et 9 mois réunis.
- Pas de tests individuels.
- Réalisation des tests simultanés où l'hypothèse  $H_0$  englobe les 3 conditions (p-values corrigées selon un critère d'ajustement (FWER)).

```
K <- rbind("12m-3m"=c(rep(0,5), 1, -1, 0, 0),
           "9/12m - 3/6m"=c(rep(0,5), 1, -1, -1, 1),
           "12m-3/6/9m"=c(rep(0,5), 3, -1, -1, -1))

K

##           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## 12m-3m      0    0    0    0    0    1   -1    0    0
## 9/12m - 3/6m 0    0    0    0    0    1   -1   -1    1
## 12m-3/6/9m  0    0    0    0    0    3   -1   -1   -1

t <- list()
for(i in 1:6) {
  t[[i]] <- glht(fit[[i]], linfct = K, alternative = c("greater"))
}
```

## Analyse par un modèle log-linéaire

### Résultats des tests multiples

```
## [[5]]
##
## Simultaneous Tests for General Linear Hypotheses
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      -0.07103    0.21975  -0.323  0.819
## 9/12m - 3/6m <= 0 -0.48153    0.33275  -1.447  0.985
## 12m-3/6/9m <= 0  -0.03643    0.53186  -0.068  0.736
## (Adjusted p values reported -- single-step method)
```

```
## [[6]]
##
## Simultaneous Tests for General Linear Hypotheses
##
## Linear Hypotheses:
##              Estimate Std. Error z value  Pr(>z)
## 12m-3m <= 0      0.8966    0.1633   5.489 < 1e-06 ***
## 9/12m - 3/6m <= 0  1.4160    0.2360   6.001 < 1e-06 ***
## 12m-3/6/9m <= 0    1.7235    0.3941   4.373 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

FIGURE: Exemple : résultats tests multiples Image 5 et 6

## Analyse par un modèle log-linéaire mixte

### Effet aléatoire sujet

- Différence d'acuité visuelle entre les sujets d'un même groupe d'âge sur une image donnée ?
- Nécessité de construire une nouvelle hyperframe :

```
hyp.glob[c(1:3,645:647,1342:1344),]

## Hyperframe:
##      ppp      groupe  image   sujet sailmoy ZCBin  n
## 1   (ppp) Baby3month Image1  Sujet1   (im)  (im)  11
## 2   (ppp) Baby3month Image1  Sujet2   (im)  (im)   2
## 3   (ppp) Baby3month Image1  Sujet3   (im)  (im)   7
## 645 (ppp) Baby9month Image3  Sujet107 (im)  (im)   7
## 646 (ppp) Baby9month Image3  Sujet108 (im)  (im)   0
## 647 (ppp) Baby9month Image3  Sujet109 (im)  (im)  11
## 1342 (ppp)      Adult Image6  Sujet222 (im)  (im)  11
## 1343 (ppp)      Adult Image6  Sujet223 (im)  (im)   9
## 1344 (ppp)      Adult Image6  Sujet224 (im)  (im)  12
```

- A l'appel de la fonction, on spécifie que l'on ne souhaite pas prendre en compte les sujets n'ayant aucune fixation sur une image.

```
fit1 <- mppm(ppp~sailmoy+ZCBin, random=~sailmoy+ZCBin|sujet,
  data=subset(hyp.glob[hyp.glob$groupe=="Baby3month" & hyp.glob$image=="Image4"], npoints(ppp)>0))
```

## Analyse par un modèle log-linéaire mixte

### Résultats

- Effets aléatoires des sujets très faibles, quelle que soit la covariable retenue.

```
random.effects(fital)[2:10,]  
  
##           (Intercept)          sailmoy          ZCBin  
## Sujet10  5.736569e-09 -4.050660e-08 -1.030605e-05  
## Sujet11 -6.440928e-09  2.427843e-08  1.825700e-05  
## Sujet12 -2.165114e-09  1.404328e-08  8.600959e-06  
## Sujet13  1.541687e-09  1.202146e-08 -9.173356e-07  
## Sujet14 -2.675526e-09  6.671103e-09  8.591307e-06  
## Sujet15 -1.054407e-09 -6.585642e-08 -8.123355e-07  
## Sujet16 -1.556714e-09 -4.633904e-08 -8.713266e-07  
## Sujet17  5.278193e-10 -5.754879e-09 -9.544204e-07  
## Sujet18 -4.188996e-09 -3.305945e-08  8.517291e-06
```

- Matrice de variance-covariance résiduelle → variabilité résiduelle très faible.

```
## Random effects:  
## Formula: ~sailmoy + ZCBin | sujet  
## Structure: General positive-definite, Log-Cholesky parametrization  
##           StdDev      Corr  
## (Intercept) 2.276325e-05 (Intr) sailmy  
## sailmoy      3.532580e-04 -0.005  
## ZCBin        3.077865e-03 -0.053  0.002  
## Residual     1.000000e+00
```

- Pas de différences significatives entre les sujets d'un même groupe d'âge pour une image donnée.



## Conclusion

- Distribution Poisson inhomogène : l'emplacement des fixations est en partie guidé par les caractéristiques locales de l'image (contrastes, bordures et couleurs).
- En modélisant l'intensité par le modèle log-linéaire suivant,

$$\log \rho(u, m, \theta) = \theta_1 \textit{Saillance}(u) + \sum_{m'=1}^4 \left( \theta_0^{m'} + \theta_1^{m'} \textit{ZCBin}(u) \right) \mathbb{1}_{m'=m}$$

on note un mouvement significatif des fixations (selon l'âge du bébé évoluant) sur certaines images.

- En modélisant l'intensité par un modèle log-linéaire mixte, nous ne décelons pas de différences significatives entre l'acuité visuelle des sujets d'un même groupe d'âge, pour une image donnée.
- Perspectives :
  - D'autres sujets de recherche sur les mouvements de l'oeil pourraient être encore approfondis afin de trouver de meilleures covariables spatiales et obtenir de bons résultats pour toutes les images.
  - Serions-nous être capables d'estimer l'intensité de nos processus ponctuels par un modèle log-linéaire mixte en plaçant l'effet aléatoire sur l'image ?