Qu'est-ce que la loi Gamma ?

Application à l'analyse de la série criminelle d'Andrej Tchikatilo

Yvonnick Noël

Université Européenne de Bretagne, CRPCC-LPE

MODEVAIIA, Camaret, Juin 2013

Sommaire

Le tueur en série Andrei Chikatilo

- Andrei Chikatilo est un **tueur en série** ukrainien qui a sévi dans la région de Rostov entre 1978 et 1990.
- Sur cette période, on lui attribue la liste impressionnante de **53 victimes**, de sexe masculin ou féminin, de 7 à 35 ans.
- Il a commencé comme professeur de russe en école en 1971, mais sera renvoyé en 1981 après plusieurs plaintes de violences contre les enfants.

Premiers meurtres

- En septembre 1978, il déménage pour la ville proche de Shakhty et y commet son **premier meurtre** deux mois plus tard.
- Bien qu'il ait fait partie de la liste des suspects, et qu'il ait été vu par des témoins, c'est un **autre agresseur** sexuel qui sera arrêté pour ce meurtre (puis exécuté en 1983 sous la pression de la famille).
- Il commet son deuxième meurtre en septembre 81, puis son troisième en juin 82.
- A partir de là, la série s'accélère : de juillet à décembre 82, il va tuer 5 fois de suite, des victimes entre 9 et 19 ans.

Constitution d'une équipe spéciale

- Le contexte historique est particulier : par idéologie, la Russie soviétique considérait que les tueurs en série étaient un produit du capitalisme et ne pouvaient exister dans la société socialiste.
- En janvier 1983, une équipe spéciale de la police de Moscou a été dépêchée sur Rostov pour diriger l'enquête. Tchikatilo ne tue plus **pendant 6 mois**.
- En février 1984, il fait l'objet d'une accusation de vol sur son lieu de travail et il fait l'objet d'une surveillance.

Première arrestation

- Il a été arrêté en septembre 1984, alors qu'il essayait d'emmener une fillette, mais relâché en décembre à cause d'une **erreur d'analyse de groupe sanguin** qui semblait l'innocenter (fait rarissime, son groupe sanguin salivaire est différent de celui détectable dans le sperme).
- A la suite de cette relaxe, il déménage à Novocherkassk près de Moscou et trouve un autre travail. Il **ne tue plus** jusqu'à l'été 1985, où il fit deux nouvelles victimes.
- Reconnaissant le **mode opératoire** (lacérations), la police se remobilise.

La traque

- En novembre 1985, la police reprit l'**enquête à zéro** :
  - on examina tous les relevés de vol Rostov-Moscou (mais Chikatilo ne voyageait qu'en train),
  - un procureur spécial fut désigné pour s'occuper de cette seule affaire,
  - on interrogea des dizaines de tueurs connus,
  - un profiler analysa les heures et localisations des meurtres.
- Les **journaux** se faisaient écho des moindres avancées de l'enquête, et Chikatilo les lisaient avec soin. Dans cette période (d'août 1985 à mai 1987), il reste inactif.

L'arrestation

- A partir de 1987, il **recommença à tuer** régulièrement, en prenant la précaution de tuer à distance de chez lui, lors de déplacements professionnels, en changeant son mode opératoire (démembrement plutôt que lacérations) et le type de victimes (des garçons plutôt que des filles).
- Il a finalement été **arrêté** en 1990 grâce à un effectif policier important déployé près des gares de la région (la plupart des meurtres avaient eu lieu à proximité de gares) puis exécuté en 1994.

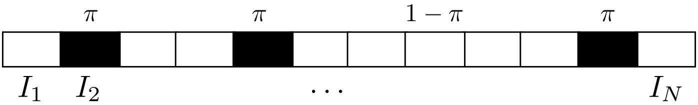
Questions sur le rythme de la séquence

Nous allons chercher à traiter les questions suivantes :

- Pouvons-nous donner du sens à la notion de **pulsion meurtrière** ?
- Si oui, comment cette pulsion meurtrière se manifeste-t-elle **dans le temps** ?
  - en situation de contraintes minimales (pas de poursuites ou d'arrestation), le **débit** des meurtres est-il constant ? Fluctuant ? Exponentiel ?
- Après une arrestation ou un contrôle policier :
  - quelle est la **capacité de contrôle** du tueur sur sa pulsion (délai sans meurtre) ?
  - Si le flux meurtrier reprend après une phase de contrôle, retrouve-t-il son **débit antérieur** ?

Modèle en intervalles

- On peut découper le flux temporel en **intervalles de temps** de largeurs constantes, au sein desquels on observe l'apparition (ou non) d'un événement cible.

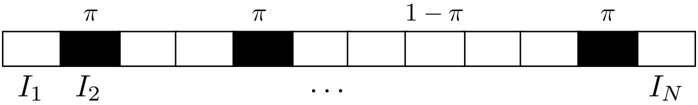


- Dans une première approche, nous supposons **constante la probabilité** d'apparition de l'événement à chaque instant.

- **Interprétation** : on attend en moyenne 1 apparition du comportement tous les $\frac{1}{\pi}$ instants. **Exemple** : $\pi = \frac{1}{4}$ (1 apparition en moyenne tous les 4 instants).

Loi sur les nombres d'apparitions

- On cherche la probabilité d'avoir ainsi $X = 3$ apparitions du comportement sur 12 unités de temps.
- Si le débit est constant, la probabilité d'avoir **exactement ces 3 apparitions** aux instants 2, 5 et 11 est $\pi^3(1-\pi)^9$.



- Mais il y a en tout C_{12}^3 séquences équivalentes avec 3 apparitions.
- On a donc au final $P(X = 3|\pi) = C_{12}^3\pi^3(1-\pi)^9$.

Loi binomiale temporelle

- Le modèle du nombre d'apparition en temps discret est donc simplement **binomial**.
- Le **nombre espéré** d'apparitions de l'événement de probabilité π sur N instants est $E(X) = N\pi$. **Exemple** avec $\pi = \frac{1}{4}$, sur 12 instants on attend en moyenne $\frac{12}{4} = 3$ apparitions.
- Pouvons-nous utiliser cette analyse pour écrire un modèle $P(X = k)$ en **temps continu** ? Il suffit d'augmenter le nombre N d'intervalles, pour un nombre $E(X) = N\pi$ maintenu constant ($N \to \infty$ suppose que $\pi \to 0$).

Passage à la limite

- On a :

$$\begin{aligned} P(X=k) &= C_N^k \pi^k (1-\pi)^{N-k} \\ &= \frac{N!}{k!(N-k)!} \pi^k (1-\pi)^{N-k} \end{aligned}$$

- On peut réécrire :

$$\begin{aligned} P(X=k) &= \frac{1}{k!} \left[\left(\frac{N}{N}\right) \left(\frac{N-1}{N}\right) \left(\frac{N-2}{N}\right) \cdots \left(\frac{N-k+1}{N}\right) \right] \\ &\quad \times (N\pi)^k \left[1 - \frac{N\pi}{N}\right]^{N-k} \end{aligned}$$

- Pour $N \to \infty$, les fractions $\frac{N-j}{N} \to 1$

Passage à la limite

- On a donc :

$$P(X = k) = \lim_{N \to \infty} \frac{1}{k!} \times \mu^k \frac{\left[1 - \frac{\mu}{N}\right]^N}{\left[1 - \frac{\mu}{N}\right]^k}$$

- Pour $N \to \infty$, $\left[1 - \frac{\mu}{N}\right]^k \to 1$:

$$P(X = k) = \lim_{N \to \infty} \frac{1}{k!} \times \mu^k \left[1 - \frac{\mu}{N}\right]^N$$

- En mathématiques, on sait que :

$$\lim_{N \to \infty} \left(1 + \frac{x}{N}\right)^N = \exp(x)$$

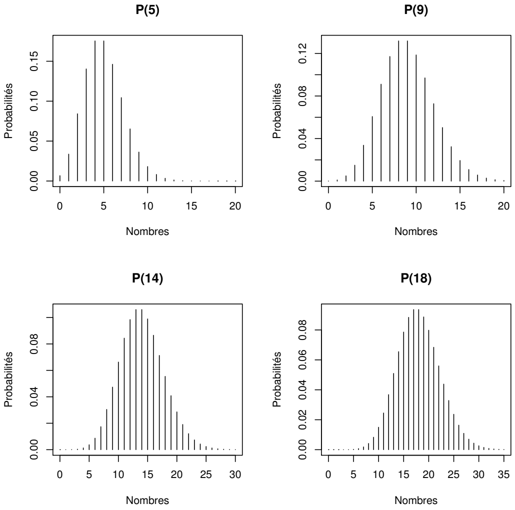
La loi de Poisson

- Pour $N \to \infty$, $\left[1 - \frac{\mu}{N}\right]^N \to \exp(-\mu)$. Au final on trouve :

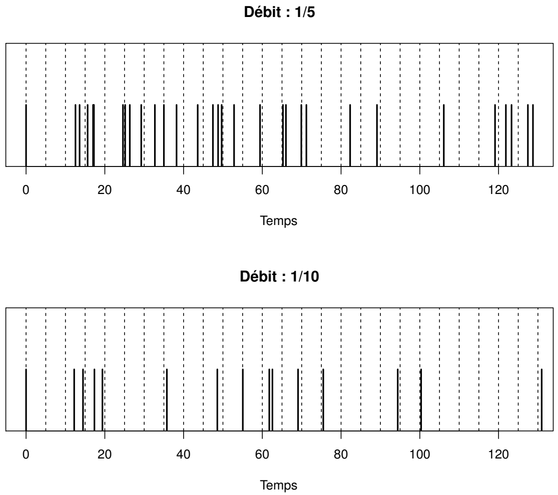
$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu}$$

- Cette expression est celle de la **loi de Poisson** sur un comptage. On note $X \sim Pois(\mu)$.

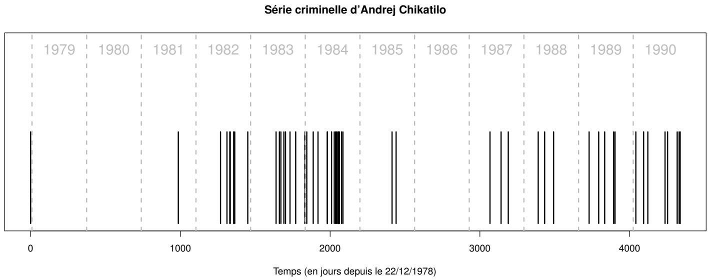
Représentation graphique



Processus poissonien à débit constant



Série criminelle d'Andrej Chikatilo

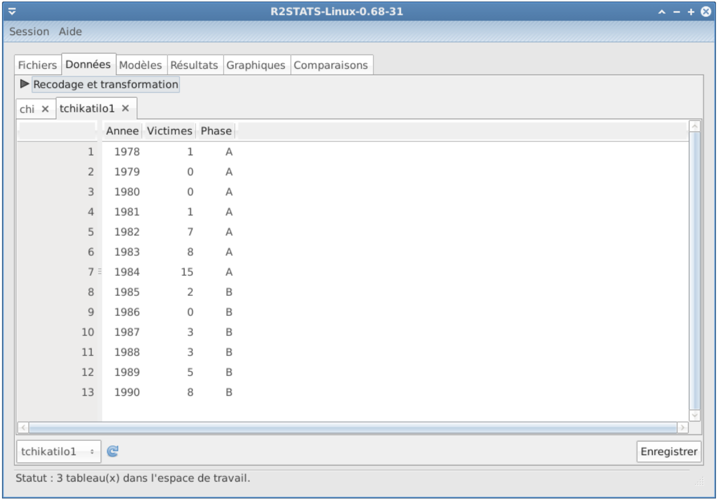


Un modèle de débit constant

- On examine les **nombres annuels** de victimes du tueur entre 1978 et 1990.
- Nous cherchons à détecter une éventuelle **structure dans son débit** criminel.
- On teste un premier modèle poissonien M_0 qui affirme que le débit meurtrier (nombre de victimes) pour l'année j est **stable avec les années** (débit constant) :

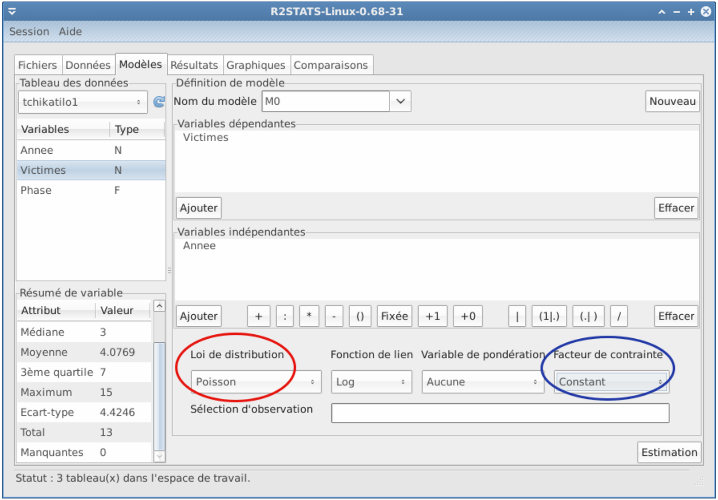
$$M_0 : \mu_j = \mu, \forall j$$

Présentation des données

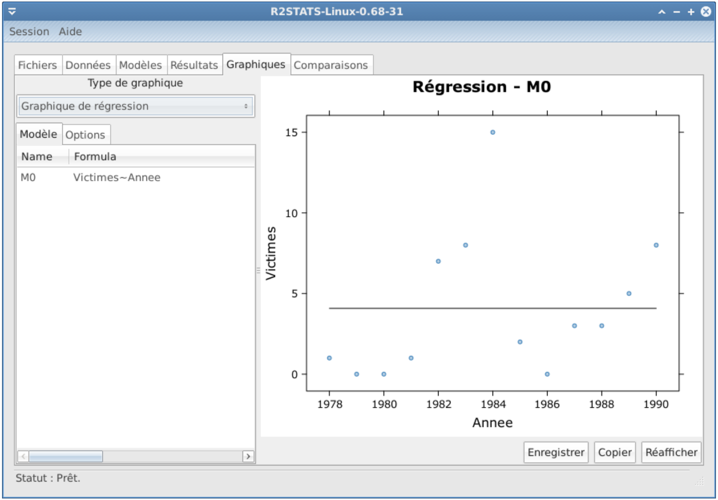


	Annee	Victimes	Phase
1	1978	1	A
2	1979	0	A
3	1980	0	A
4	1981	1	A
5	1982	7	A
6	1983	8	A
7	1984	15	A
8	1985	2	B
9	1986	0	B
10	1987	3	B
11	1988	3	B
12	1989	5	B
13	1990	8	B

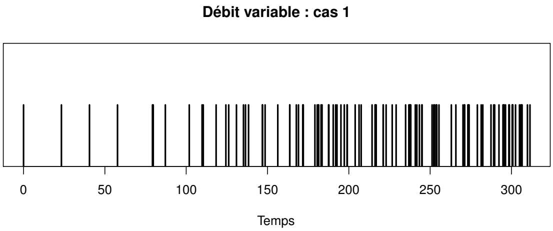
Définition sous R2STATS

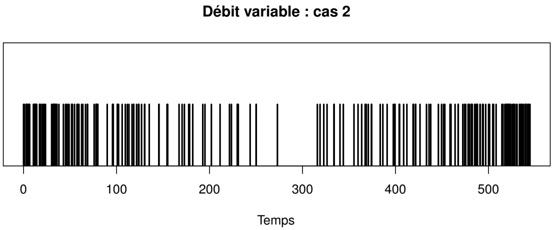


Représentation graphique du modèle



Processus poissonien à débit variable





Fonction de régression exponentielle

- Pour prendre en compte une variation du débit, on peut construire un sous-modèle structural qui place le paramètre de moyenne (nombre d'événements attendus par unité de temps) sous la dépendance de **variables explicatives**.
- Comme la moyenne d'un comptage **ne peut être négative**, on écrit :

$$\mu = \exp\left[\beta_0 + \beta_1 X\right] \text{ ou encore } \ln \mu = \beta_0 + \beta_1 X$$

(plutôt que $\mu = \beta_0 + \beta_1 X$ qui ne garantirait pas la positivité).

Modèle de régression poissonienne

- Le modèle complet s'écrit :

$$Y|X \sim Pois(\mu)$$
$$\ln \mu = \beta_0 + \beta_1 X$$

- On parle de **régression poissonienne** avec **lien logarithmique** (ou fonction de réponse exponentielle).

- On parle aussi de **modèle log-linéaire**, car il est linéaire au lien logarithmique près.

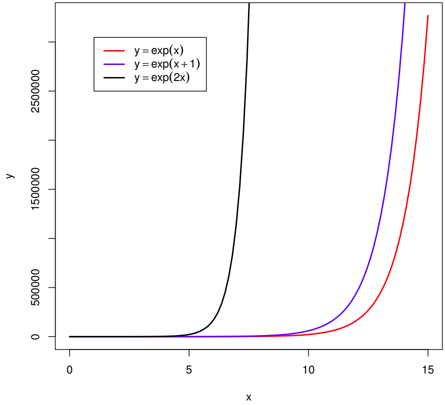
Un modèle de débit croissant

- Si nous concevons la séquence criminelle comme un processus pulsionnel avec **auto-renforcement**, on peut s'attendre à un débit qui ne cesse d'augmenter avec le temps ($\beta_1 > 0$) :

$$\mu_t = \exp\left[\beta_0 + \beta_1 t\right]$$

- Dans cette expression, les **paramètres structuraux** β_0 et β_1 jouent un rôle de position horizontale et de vitesse de croissance.
- On teste ce second modèle M_1 qui affirme que le débit meurtrier **croît avec les années**.

Propriétés de la fonction exponentielle



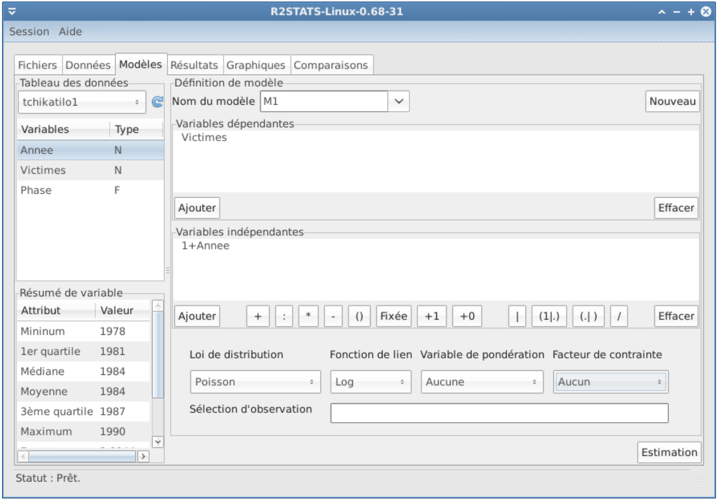
- **Ajouter** une constante à X revient à **multiplier** Y par une constante :

$$\exp(x+1) = \exp(x) \times \exp(1)$$

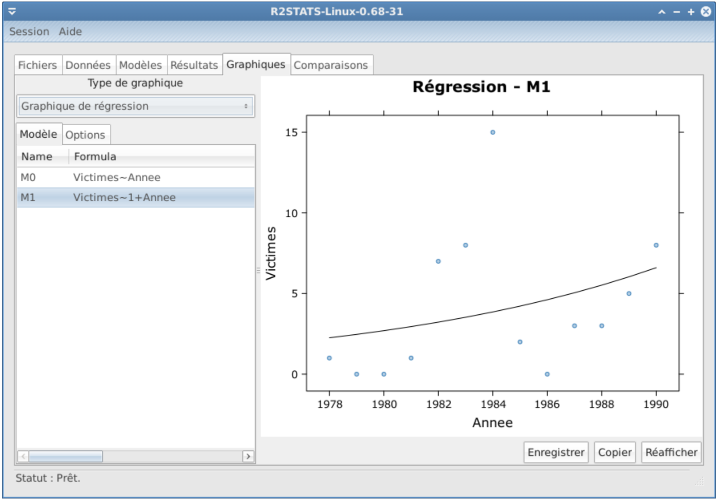
- **Multiplier** X par une constante revient à élever Y à une certaine **puissance** :

$$\exp(2x) = [\exp(x)]^2$$

Modèle exponentiel simple



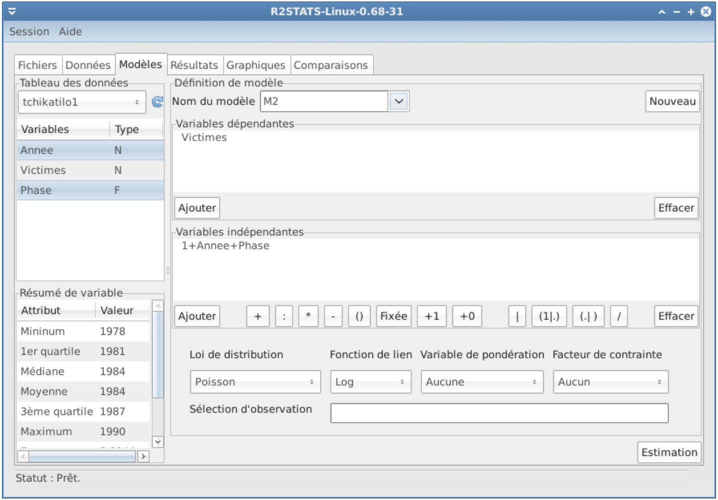
Représentation graphique



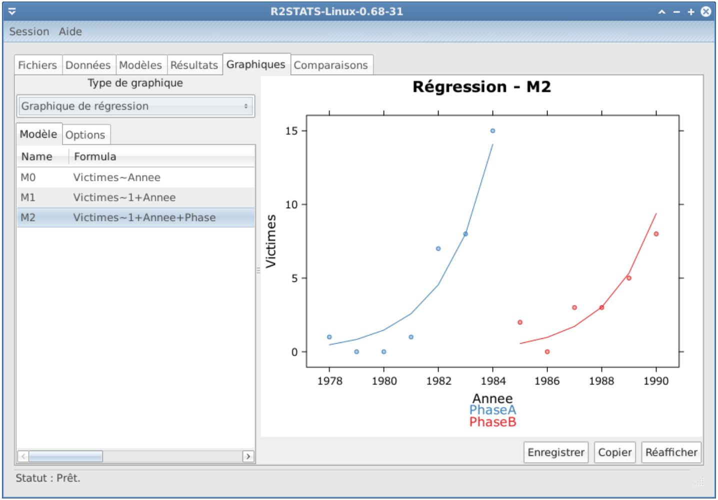
Un modèle en phases

- On peut tester un modèle qui prend en compte un découpage de la carrière criminelle en deux phases : avant et après son **arrestation** en septembre 1984.
- Deux possibilités :
  - on considère qu'après sa libération (deux mois plus tard), il redémarre une série criminelle **à zéro**, identique à celle qu'il a initiée en 1978.
  - on peut aussi supposer que la **vitesse d'augmentation** du débit est différente dans la deuxième phase.

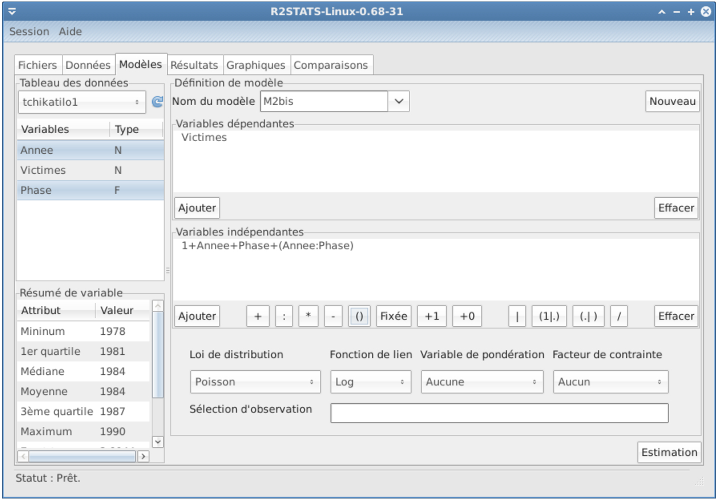
Modèle en phases (vitesse identique)



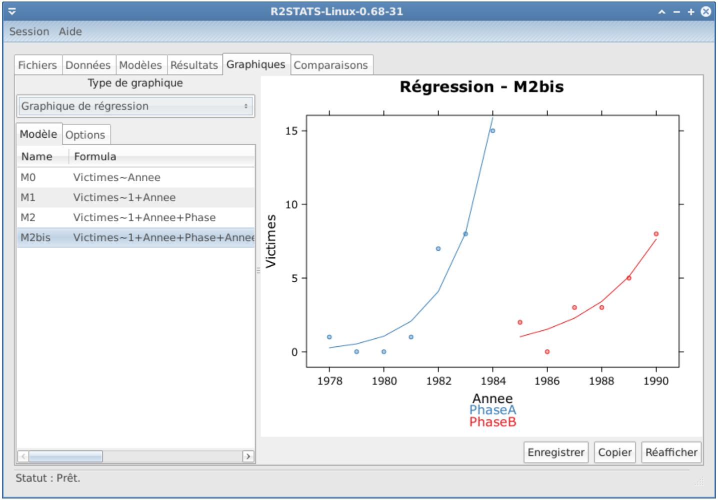
Modèle en phases (vitesse identique)



Modèle en phases (vitesse différente)



Modèle en phases (vitesse différente)



Comparaison des modèles

R2STATS-Linux-0.68-31

Session Aide

Fichiers | Données | Modèles | Résultats | Graphiques | Comparaisons

Constante	-1343.474	272.163	-4.936	0.000
Annee	0.679	0.137	4.944	0.000
PhaseB	544.059	399.117	1.363	0.173
Annee:PhaseB	-0.276	0.201	-1.372	0.170

Dispersion 1

Comparaison et résumé de modèles

Table d'analyse de la variance/déviance

	Formula	Constraint	Distribution	Link
M0	Victimes~Annee	Constant	Poisson	Log
M1	Victimes~1+Annee	Aucun	Poisson	Log
M2	Victimes~1+Annee+Phase	Aucun	Poisson	Log
M2bis	Victimes~1+Annee+Phase+Annee:Phase	Aucun	Poisson	Log

	Ddl rés.	Dév. res.	Ddl diff.	Rap.Vr.	Pr(>Chi2)	AIC	BIC	Expl.(%)
M0	12	58.110				92.429	94.400	0.00000
M1	11	52.353	1	5.757	0.016419	88.672	92.612	0.09908
M2	10	12.709	1	39.644	0.000000	51.028	56.939	0.78130
M2bis	9	10.833	1	1.876	0.170823	51.152	59.033	0.81358

Effacer

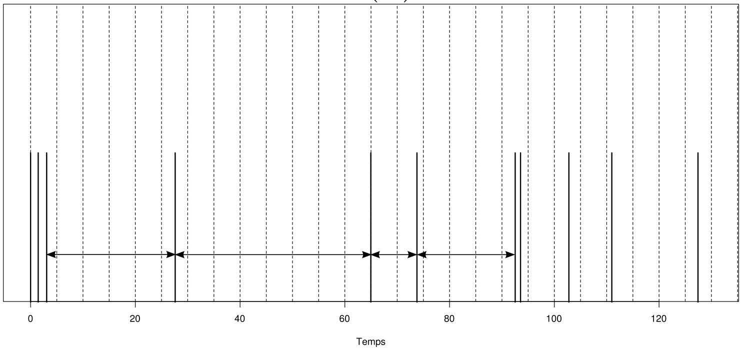
Statut : Prêt.

Conclusions provisoires

- La pulsion criminelle de Chikatilo **n'est pas** à débit constant.
- Le modèle exponentiel suggère plutôt un processus d'**auto-renforcement**, le crime appelant le crime.
- L'allure de ce schéma d'évolution n'est **pas affectée** par une arrestation : il redémarre à l'identique.
- Si on extrapole, il est donc souhaitable d'**arrêter tôt** ce type de personnage...

Loi exponentielle sur les temps inter-réponse

$$Exp\left(\frac{1}{10}\right)$$



Loi exponentielle sur les temps inter-réponse

- Si l'apparition d'un comportement dans le temps suit une loi de Poisson, les **temps qui séparent** ses apparitions suivent une loi exponentielle.
- On la note $Exp(\lambda)$.

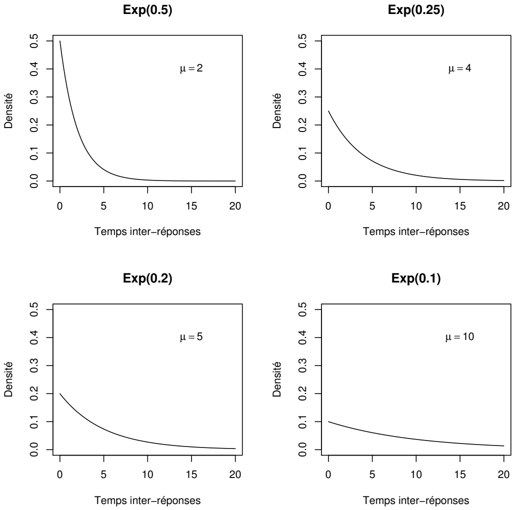
$$f(t|\lambda) = \lambda e^{-\lambda t}$$

Paramètre : λ est le **taux** (ou débit) d'apparition de l'événement-cible par unité de temps, et

$$\mu = \frac{1}{\lambda}$$

est le **temps moyen** pour le voir apparaître.

Représentation graphique

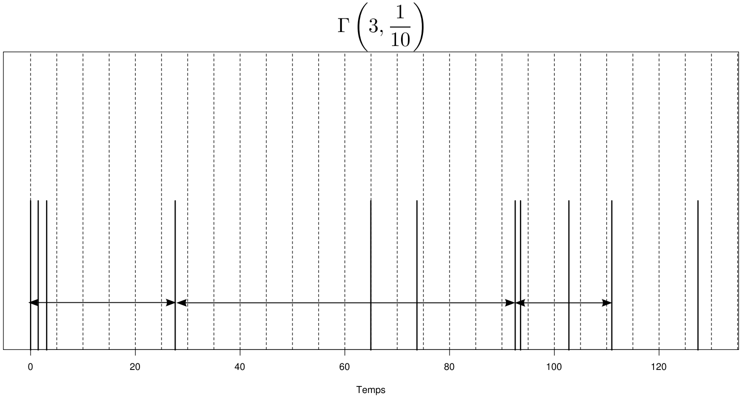


La loi Gamma

- La loi exponentielle est un cas particulier d'une loi plus générale appelée **loi Gamma**.
- La loi Gamma décrit la distribution du temps d'attente **jusqu'à ce que la s-ième apparition** (poissonienne) d'un comportement survienne.
- C'est donc la loi d'un **temps total**, somme de temps issus d'une loi exponentielle de même débit.
- Elle est notée $\Gamma(s, \lambda)$.

$$f(t) = \frac{\lambda}{\Gamma(s)} (\lambda t)^{s-1} e^{-\lambda t}$$

Processus de Poisson et loi Gamma

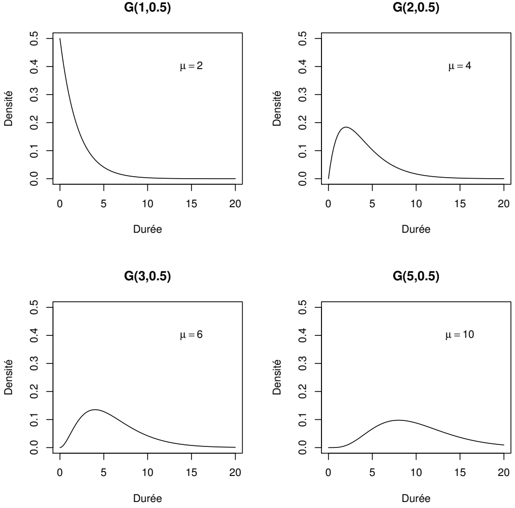


Interprétation des paramètres

- λ est le taux ou **débit de réponse** (exemple : une valeur de $\frac{1}{10}$ signifie un débit d'une réponse toutes les 10 unités de temps).
- s est le **nombre de réponses** qu'on attend (exemple : une valeur de 3 signifie qu'on modélise le temps jusqu'à l'apparition de trois réponses consécutives).
- Le **temps d'attente moyen** sera donc de :

$$\mu = s \times \frac{1}{\lambda} = 3 \times 10 = 30 \text{ unités de temps}$$

Représentation graphique



Régression Gamma (lien inverse)

- Comme nous l'avons fait pour les modèles poissoniens, nous pouvons construire un modèle du **débit de réponse** à partir d'une ou de plusieurs variables indépendantes.
- Comme la moyenne d'un Gamma s'écrit :

$$\mu = \frac{s}{\lambda}$$

on peut définir un modèle de régression dit **inverse** :

$$\mu = \frac{s}{\lambda} = \frac{s}{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}$$

Régression Gamma (lien inverse)

- Le modèle de **régression inverse**

$$\mu = \frac{s}{\lambda} = \frac{s}{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}$$

peut aussi s'écrire :

$$\begin{aligned} \frac{1}{\mu} &= \frac{\beta_0}{s} + \frac{\beta_1}{s} X_1 + \ldots + \frac{\beta_p}{s} X_p \\ &= \beta_0^* + \beta_1^* X_1 + \ldots + \beta_p^* \end{aligned}$$

Régression Gamma (lien log)

Nous pouvons aussi observer que si $\mu = \frac{s}{\lambda}$, alors :

$$\ln \mu = \ln s - \ln \lambda$$

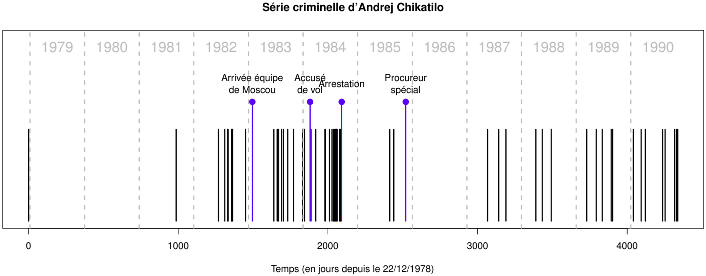
Si nous pensons que l'effet des variables sur le temps d'attente est non plus additif mais multiplicatif (effet d'apprentissage ou d'accumulation) :

$$\begin{aligned} \lambda &= e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 X_2} \ldots e^{\beta_p X_p} \\ &= e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p} \end{aligned}$$

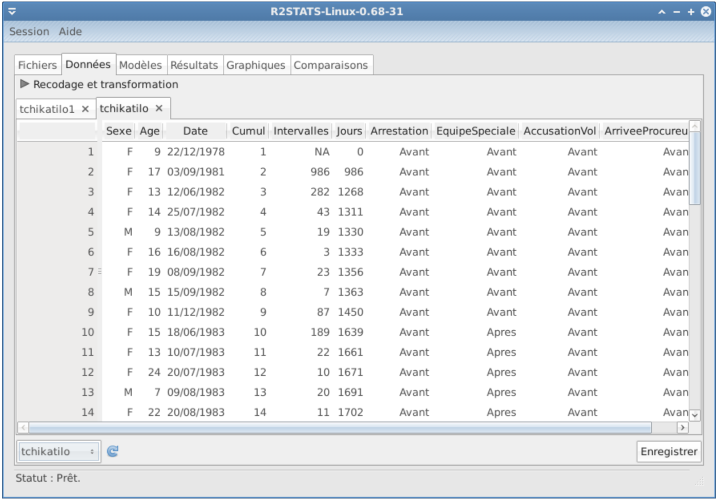
alors on a un modèle loglinéaire :

$$\ln \mu = \beta_0 + \beta_1 X_1 + \ldots + \beta_p$$

Impact des évènements

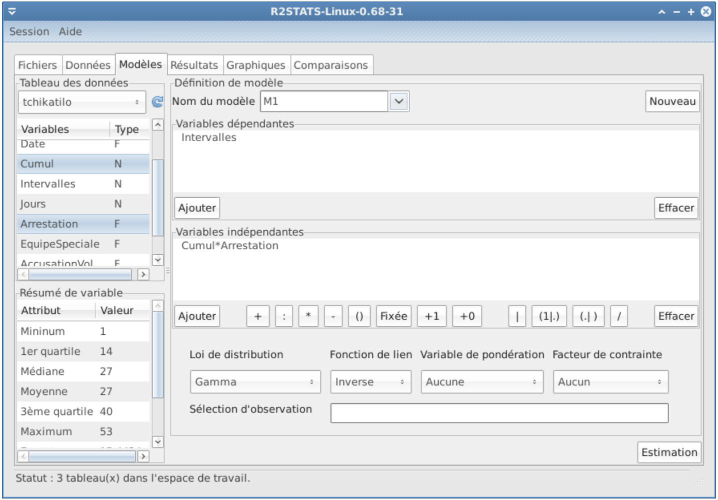


Présentation des données

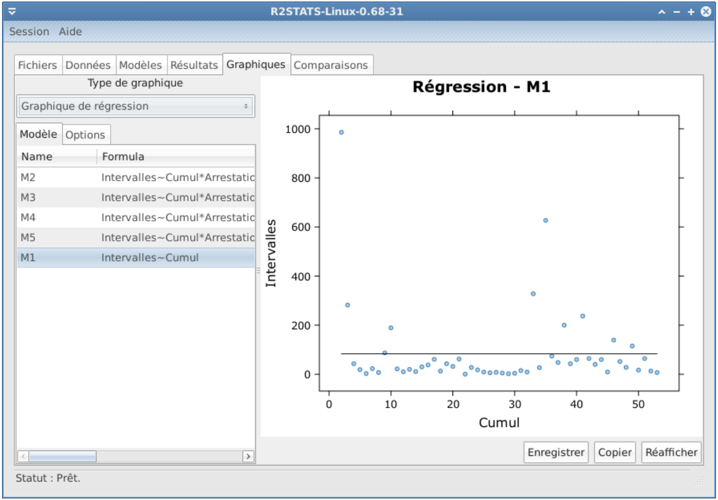


	Sexe	Age	Date	Cumul	Intervalles	Jours	Arrestation	EquipeSpeciale	AccusationVol	ArriveeProcureu
1	F	9	22/12/1978	1	NA	0	Avant	Avant	Avant	Avan
2	F	17	03/09/1981	2	986	986	Avant	Avant	Avant	Avan
3	F	13	12/06/1982	3	282	1268	Avant	Avant	Avant	Avan
4	F	14	25/07/1982	4	43	1311	Avant	Avant	Avant	Avan
5	M	9	13/08/1982	5	19	1330	Avant	Avant	Avant	Avan
6	F	16	16/08/1982	6	3	1333	Avant	Avant	Avant	Avan
7	F	19	08/09/1982	7	23	1356	Avant	Avant	Avant	Avan
8	M	15	15/09/1982	8	7	1363	Avant	Avant	Avant	Avan
9	F	10	11/12/1982	9	87	1450	Avant	Avant	Avant	Avan
10	F	15	18/06/1983	10	189	1639	Avant	Apres	Avant	Avan
11	F	13	10/07/1983	11	22	1661	Avant	Apres	Avant	Avan
12	F	24	20/07/1983	12	10	1671	Avant	Apres	Avant	Avan
13	M	7	09/08/1983	13	20	1691	Avant	Apres	Avant	Avan
14	F	22	20/08/1983	14	11	1702	Avant	Apres	Avant	Avan

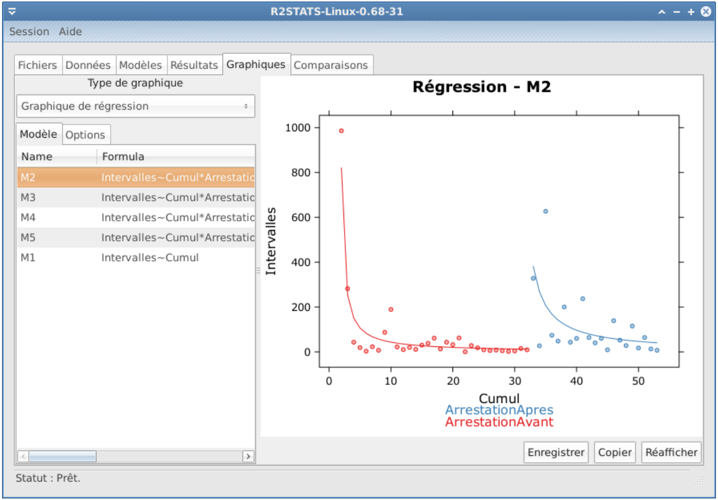
Modèle du délai inter-crime



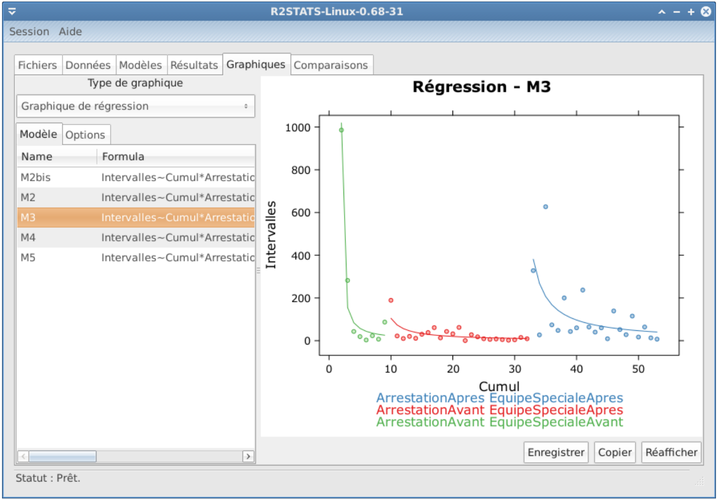
Modèle de débit constant



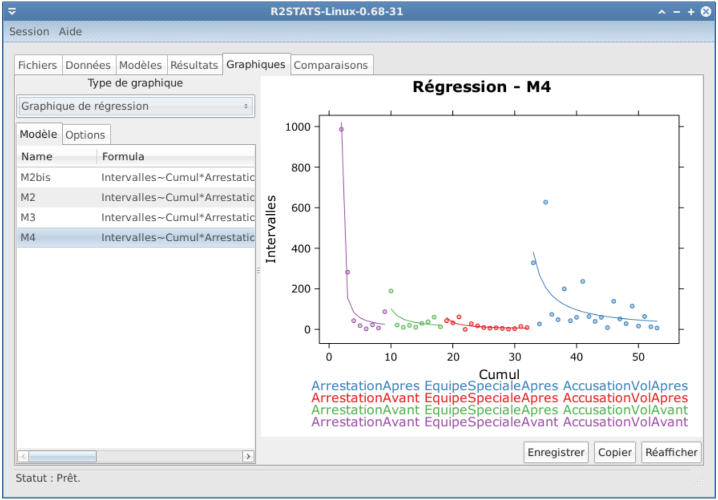
Effet de l'arrestation



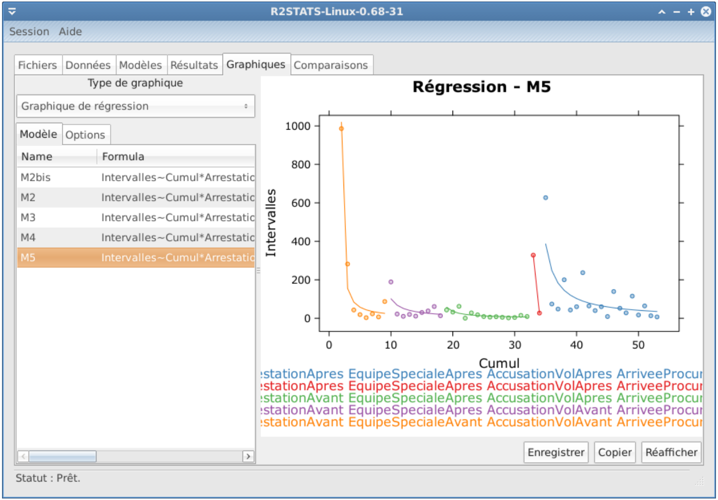
Effet de l'arrivée de l'équipe spéciale de Moscou



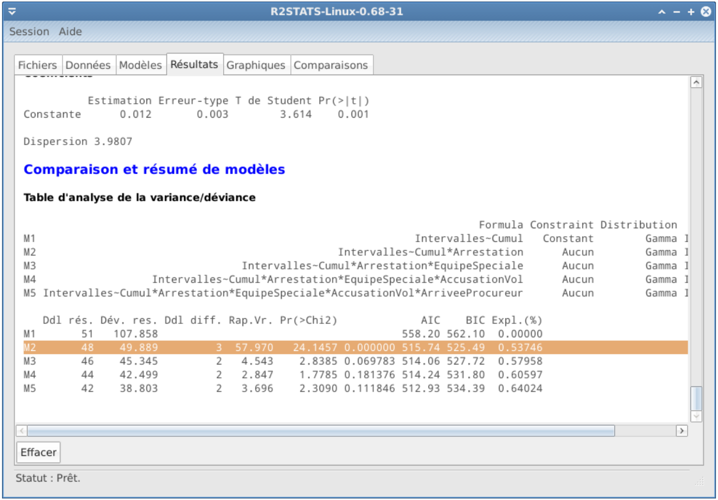
Effet de l'accusation de vol



Effet de l'arrivée du procureur spécial



Comparaison des modèles



	Estimation	Erreur-type	T de Student	Pr(>|t|)
Constante	0.012	0.003	3.614	0.001

Dispersion 3.9807

Comparaison et résumé de modèles

Table d'analyse de la variance/déviance

	Formula	Constraint	Distribution
M1	Intervalles~Cumul	Constant	Gamma I
M2	Intervalles~Cumul*Arrestation	Aucun	Gamma I
M3	Intervalles~Cumul*Arrestation*EquipeSpeciale	Aucun	Gamma I
M4	Intervalles~Cumul*Arrestation*EquipeSpeciale*AccusationVol	Aucun	Gamma I
M5	Intervalles~Cumul*Arrestation*EquipeSpeciale*AccusationVol*ArriveeProcureur	Aucun	Gamma I

	Ddl rés.	Dév. res.	Ddl diff.	Rap.Vr.	Pr(>Chi2)		AIC	BIC	Expl.(%)
M1	51	107.858					558.20	562.10	0.00000
M2	48	49.889	3	57.970	24.1457	0.000000	515.74	525.49	0.53746
M3	46	45.345	2	4.543	2.8385	0.069783	514.06	527.72	0.57958
M4	44	42.499	2	2.847	1.7785	0.181376	514.24	531.80	0.60597
M5	42	38.803	2	3.696	2.3090	0.111846	512.93	534.39	0.64024

Conclusions

- Cette méthode permet d'analyser finement toute production de **comportements répétés en séquence**, ainsi que l'impact d'**événements extérieurs** sur le débit de cette séquence.
- Elle est donc utilisable en psychologie dans des domaines très divers.
- Avec un bon modèle du débit, on peut détecter des temps inter-réponses **étonnamment longs** (victimes inconnues par exemple).