AUTRES LOIS, AUTRES MODÈLES : Introduction

- Loi de Poisson
- Loi Binomiale Négative
- Loi Exponentielle et loi Gamma

Plan commun pour chaque loi étudiée

Exemple

Présentation

Hypothèses

Variance

Le modèle dans le cadre GLM

Les écarts aux hypothèses initiales

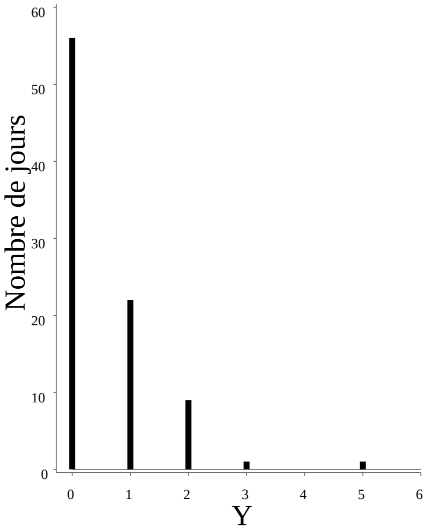
Quasi-vraisemblance, surdispersion

LOI DE POISSON : exemple

Exemple : le merle à plastron

Y	0	1	2	3	4	5	6
Fréquence	56	22	9	1	0	1	0

Y : nombre de merles capturés par jour

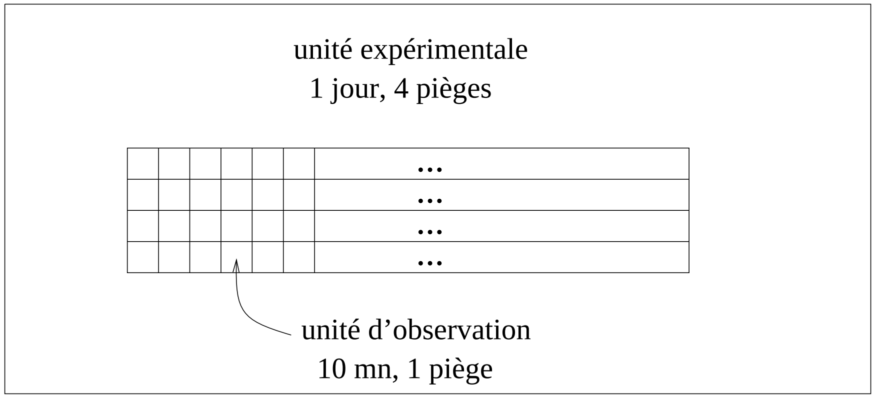


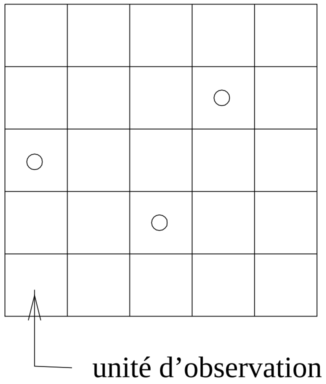
LOI DE POISSON :

unité d'observation

unité expérimentale

merle à plastron





unité expérimentale de taille n

(n unités d'observation)

LOI DE POISSON : présentation

C'est la loi d'une variable aléatoire Y

qui exprime le nombre d'occurrences

d'événements

distribués au hasard

dans un certain laps de temps

ou dans une région donnée de l'espace.

LOI DE POISSON : hypothèses

sur les unités expérimentales

- n, taille des unités expérimentales fixée a priori et *constante*.

sur le phénomène observé

- *événement* compté *bien défini*, *identique* pour toutes les unités expérimentales.
- *comptage exhaustif* par unité expérimentale.
- *un* événement compté *au plus* par unité d'observation.
- *indépendance* des *unités d'observation*.
- *répartition* des événements *au hasard* parmi les unités d'observation d'une unité expérimentale.
- phénomène "*rare*".

LOI DE POISSON : variance

Distribution discrète.

Pour une unité expérimentale de taille n

$$Y \sim \text{Poisson}(np)$$

$$\boxed{\text{Var}(Y) = \text{E}(Y)}$$

np : effectif moyen.

n : taille des unités expérimentales

p : probabilité constante d'occurrence de l'événement
par unité d'observation.

LOI DE POISSON :

comme approximation

de la loi Binomiale

La loi $\mathcal{B}(n,p)$ tend vers une loi ($\mathcal{P}(np)$)

quand

- n tend vers l'infini (n grand)
- et p tend vers 0 (événement "rare").

LOI DE POISSON :

le modèle dans le cadre GLM

Y : nombre d'événements
sur une unité expérimentale,

$$Y \sim \mathcal{P}$$

lien canonique : lien log,

$$\boxed{\mathrm{Var}(Y) = \mathrm{E}(Y)}$$

LOI DE POISSON :

les écarts aux hypothèses initiales

Taille des unités expérimentales non constante

- Modélisation de la loi de n
- Introduire n sous forme d'un "offset"

Unités d'observation non indépendantes

- Introduire un facteur de dispersion
- Modélisation de la dépendance

Unités expérimentales non indépendantes

- Introduire un facteur de dispersion
- Prise en compte de la dépendance (GEE, ...)

LOI DE POISSON :

les écarts aux hypothèses initiales (suite)

Non répartition des événements
au hasard
ou
fréquence non constante

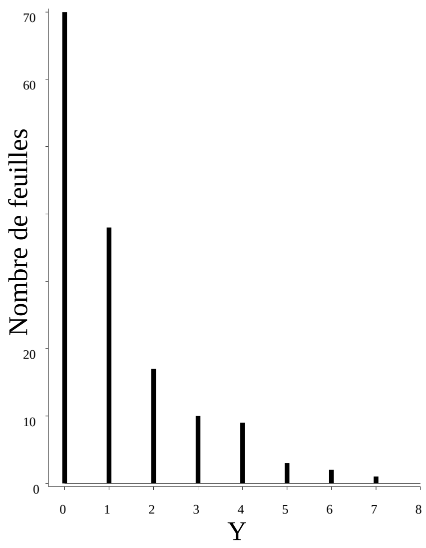
- Redéfinir l'unité expérimentale
- Introduire un facteur de dispersion
- Utiliser un facteur aléatoire
- Prendre en compte les facteurs de variation

LOI BINOMIALE NÉGATIVE : exemple

Exemple : la tétranyque rouge du pommier

Y	0	1	2	3	4	5	6	7	8+
Nombre de feuilles	70	38	17	10	9	3	2	1	0

Y : nombre d'individus par feuilles



LOI BINOMIALE NÉGATIVE : présentation

Présentation classique :

C'est la loi d'une variable aléatoire Y' qui exprime

le nombre d'essais indépendants nécessaires pour obtenir k occurrences d'un événement

ayant une probabilité p constante d'être observé dans chaque essai.

Autre présentation :

C'est aussi la loi d'une variable aléatoire Y qui exprime

le nombre de non événements nécessaires pour observer k occurrences d'un événement.

LOI BINOMIALE NÉGATIVE : hypothèses

sur les unités expérimentales

Y : nombre d'unités d'observation aléatoire.

sur le phénomène observé

- *événement* compté *bien défini*, *identique* pour toutes les unités expérimentales
- *comptage exhaustif* par unité expérimentale
- *un* événement compté *au plus* par unité d'observation
- *indépendance* des *unités d'observation*
- *répartition* des événements *au hasard* parmi les unités d'observation d'une unité expérimentale

LOI BINOMIALE NÉGATIVE : variance

Distribution discrète.

$$Y' \sim \mathcal{BN}(k, p)$$

Y' : nombre d'unités d'observation qu'il faut regarder dans une unité expérimentale pour obtenir k événements, chaque événement ayant la probabilité p d'être obtenu.

$$\mathrm{E}(Y') = \frac{k}{p}$$

$$\mathrm{Var}(Y') = \frac{k(1-p)}{p^2}$$

$$\boxed{\mathrm{Var}(Y') > \mathrm{E}(Y')}$$

On peut aussi voir la relation sous la forme :

$$\mathrm{Var}(Y') = \frac{\mathrm{E}(Y')^2}{k} - \mathrm{E}(Y')$$

LOI BINOMIALE NÉGATIVE :

autre présentation

$$\text{Soit } Y = Y' - k$$

Y : nombre de non-événements qu'il faut observer pour observer k événements.

Les non-événements comptés par Y ont une probabilité $(1-p)$ constante d'être observés dans chaque essai.

$$Y \sim \mathcal{BN}(k, p)$$

$$\mathrm{E}(Y) = \frac{k}{p}(1-p)$$

$$\mathrm{Var}(Y) = \frac{k(1-p)}{p^2}$$

$$\boxed{\mathrm{Var}(Y) > \mathrm{E}(Y)}$$

On peut aussi voir la relation sous la forme :

$$\mathrm{Var}(Y) = \mathrm{E}(Y) + \mathrm{E}(Y)^2/k$$

LOI BINOMIALE NÉGATIVE

$Var(Y) = E(Y)$	loi de poisson	nb d'occurrences d'événements rares indépendants
$Var(Y) > E(Y)$	loi Binomiale Négative	nb d'occurrences d'événements non indépendants (agrégation)

LOI BINOMIALE NÉGATIVE

comme alternative à la loi de Poisson

- si $(Y|N = n) \sim \mathcal{P}(np)$
- si p constant
- si $N \sim$ Gamma, $\mathrm{E}(N) = \mathrm{Var}(N) = b$

alors

$$Y \sim \mathcal{BN}(b, \tfrac{1}{1+p})$$

$$\mathrm{E}(Y) = bp \qquad \mathrm{Var}(Y) = bp(p+1)$$

$\Rightarrow$ surdispersion $= (p+1)$

LOI BINOMIALE NÉGATIVE

comme alternative à la loi de Poisson

- si $(Y|P = p) \sim \mathcal{P}(np)$
- si n constant
- si $P \sim$ Gamma, $\mathrm{E}(P) = \mathrm{Var}(P) = c$

alors

$$Y \sim \mathcal{BN}(c, \tfrac{1}{1+n})$$

$$\mathrm{E}(Y) = cn \qquad \mathrm{Var}(Y) = cn(1+n)$$

$\Rightarrow$ surdispersion $= (1+n)$

LOI BINOMIALE NÉGATIVE

comme alternative à la loi de Poisson

La loi Binomiale Négative est un

modèle mathématique

qui s'ajuste très bien à des données

provenant de phénomènes

ayant un caractère "contagieux"

et à des données "agrégées" en général.

exemple :

comptage d'organismes parasites, avec
taux de croissance dépendant de l'effectif
et d'un taux d'immigration constant.

LIENS D'APPROXIMATION ENTRE LOIS

BINOMIALE

POISSON

BINOMIALE NÉGATIVE

$$\begin{array}{ccccc}
\mathcal{B}(n,p) & \longrightarrow & \mathcal{P}(np) & \longrightarrow & \mathcal{BN} \\
 & \begin{cases} n \to \infty \\ p \to 0 \end{cases} & & \begin{array}{r} n \sim \text{Gamma} \\ \text{ou } p \sim \text{Gamma} \\ \text{ou agrégation} \end{array} &
\end{array}$$

LOI BINOMIALE NÉGATIVE

le modèle dans le cadre GLM

Y : nombre de non-événements observés

lien log

$$\boxed{\mathrm{Var}(Y) = \mathrm{E}(Y) + \mathrm{E}(Y)^2/k}$$

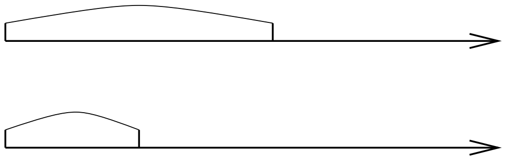
Où

k est soit connu *a priori*,

soit estimé en même temps que $\mathrm{E}(Y)$.

LOI EXPONENTIELLE : présentation

La loi Exponentielle est utilisée comme modèle pour la variable aléatoire à valeur positive donnant le *temps d'attente* pour l'apparition d'un phénomène aléatoire.



La loi Exponentielle peut servir de modèle pour une variable aléatoire à valeur positive correspondant à la *distance* séparant des événements répartis aléatoirement le long d'un itinéraire.

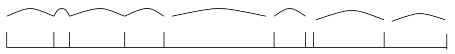
LOI EXPONENTIELLE : présentation

La loi Exponentielle s'applique si l'inverse

d'une variable suivant une loi de Poisson

présente une signification

physique, biologique ou autre.



unité expérimentale
de taille n

nombre d'événements $\sim \mathcal{P}(np)$

Y : écarts entre événements successifs.

$$Y \sim Exp(p)$$

$E(Y) = 1/p$: écart moyen entre événements

$$V(Y) = \frac{1}{p^2}$$

LOI GAMMA : présentation

La loi Gamma

est une généralisation de la loi Exponentielle

elle est utilisée comme modèle

pour la variable aléatoire à valeur positive

donnant le temps séparant

l'apparition de deux événements

séparés par un nombre fixé d'événements.

Quand les deux événements sont successifs

on retrouve la loi Exponentielle.

LOIS GAMMA ET EXPONENTIELLE : hypothèses

Nombre d'événements $\sim \mathcal{P}(np)$

$$\Longleftrightarrow$$

écart entre événements successifs

$$\begin{aligned} T_l - T_{l-1} &\sim \text{ Exponentielle}(p) \\ T_1 &\sim \text{ Exponentielle}(p) \end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned} T_l - T_{l-k} &\sim \text{ Gamma}(p, k) \\ T_k &\sim \text{ Gamma}(p, k) \end{aligned}$$

donc

mêmes hypothèses que pour la loi de Poisson

LOI GAMMA : variance

Distribution continue

$$Y \sim \text{Gamma}(p, k)$$

$$\mathrm{E}(Y) = \frac{k}{p} \qquad \mathrm{Var}(Y) = \frac{k}{p^2}$$

$$\boxed{\mathrm{Var}(Y) = [\mathrm{E}(Y)]^2/k}$$

$$\text{Gamma}(p, 1) = \text{Exponentielle}(p)$$

LOIS GAMMA ET EXPONENTIELLE

le modèle dans le cadre GLM

loi continue

Y : temps ou distance,

ou autre.

lien canonique : fonction “inverse”,

$$\boxed{\mathrm{Var}(Y) = \mathrm{E}(Y)^2/k}$$

AUTRES LOIS : récapitulatif

lois discrètes :

Binomiale	$\mathrm{Var}(Y) < \mathrm{E}(Y)$
Poisson	$\mathrm{Var}(Y) = \mathrm{E}(Y)$
Binomiale Négative	$\mathrm{Var}(Y) > \mathrm{E}(Y)$

lois continues :

Gaussienne	$\mathrm{Var}(Y) = \mathrm{Constante}$
Exponentielle	$\mathrm{Var}(Y) = \mathrm{E}(Y)^2$
Gamma	$\mathrm{Var}(Y) = \mathrm{E}(Y)^2/k$

quasi-vraisemblance :

modéliser la relation entre $\mathrm{Var}(Y)$ et $\mathrm{E}(Y)$

sans faire le choix d'une loi,

méthode utilisée en cas de surdispersion.

résultats généraux

$$E(Y) = \mu$$

$$\boxed{g(\mu) = X\theta = \eta \qquad \text{loi de } Y?}$$

loi	lien canonique	fct. de var. V
Normale	μ	1
Binomiale	$\log(\frac{\mu}{1-\mu})$	$\mu(1-\mu)$
Poisson	$\log(\mu)$	μ
Gamma	$\frac{1}{\mu}$	μ^2
Gaussienne inverse	$\frac{1}{\mu^2}$	μ^3
quelconque	$g(\mu)$	$V(\mu)$

	logit	probit	cloglog	id.	log	sqrt	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
Normale				D				
Binomiale	D	•	•					
Poisson				•	D	•		
Gamma				•	•		D	
Gauss. inv.								D
quelconque	•	•	•	•	•	•	•	•

Généralisation : la famille exponentielle

- densité $f(y;\theta,\phi) \in$ famille exponentielle si :

$$\log(f(y;\theta,\phi)) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)$$

- θ : **paramètre canonique**

- ϕ : paramètre de dispersion (connu)

- $E(Y) = \mu = b'(\theta)$

- $Var(Y) = b''(\theta)a(\phi)$

- $b''(\theta)$ **fonction de variance**

application : **loi binomiale**
(loi d'une proportion X)

- $f(x;p) = \mathcal{C}_n^{nx}\, p^{nx}(1-p)^{n-nx}$
  $\log f(x;p) = n(x \log \dfrac{p}{1-p} + \log(1-p)) + \log \mathcal{C}_n^{nx}$
- $\theta = \log(\frac{p}{1-p})$ lien canonique entre p et θ
- $b(\theta) = -\log(1-p) = \frac{1}{1+\exp\theta}$
- $\phi = 1/n$, $a(\phi) = \phi$
- $c(x,\phi) = \log(\mathcal{C}_n^{x})$
- $E(X) = \mu = \dfrac{\exp\theta}{1+\exp\theta} = p$
- $Var(X) = b''(\theta)a(\phi) = \mu(1-\mu)/n$
  $V(\mu) = \mu(1-\mu)$ fonction de variance

application : **loi normale**

- $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x-\mu)^2/2\sigma^2)$

- $\log f(x;\mu,\sigma^2) = \frac{x\mu-\mu^2/2}{\sigma^2} - 1/2(x^2/\sigma^2 + \log(2\pi\sigma^2))$

- $\theta = \mu$ **lien canonique** (identité)

- $b(\theta) = \theta^2/2$

- $\phi = \sigma^2$, $a(\phi) = \phi = \sigma^2$

- $c(x,\phi) = -1/2(x^2/\sigma^2 + \log(2\pi\sigma^2))$

- $E(X) = \mu$

- $Var(X) = b''(\theta)a(\phi) = 1 \times \sigma^2$
  $V(\mu) = 1$ **fonction de variance**

Si surdispersion

$$V(Y) = \sigma^2 V(\mu)$$

σ^2 inconnu

on n'est plus dans le cadre GLM,

maximisation d'une fonction $\rightarrow \widehat{\theta}$

quasi-vraisemblance

Si surdispersion

$$V(Y) = \sigma^2 V(\mu)$$

σ^2 paramètre de surdispersion

$$\boxed{D \sim_{as} \sigma^2 \chi^2(m-k)}$$

$$\boxed{D_2 - D_1 \sim_{as} \sigma^2 \chi^2(dl_2 - dl_1)}$$

1. $\sigma^2 = 1$: tests du χ^2
2. σ^2 inconnu : tests de Fisher

$$\frac{(D_2 - D_1)/(dl_2 - dl_1)}{D_0/dl_0} \sim F(dl_2 - dl_1, dl_0)$$

Si surdispersion

Maximisation d'un critère $\Rightarrow \widehat{\theta}$

$$\boxed{\widehat{\theta} \sim_{as} \mathcal{N}(\theta, \sigma^2 (X'WX)^{-1})}$$

σ^2 remplacé par $\widehat{\sigma}^2$

$\rightarrow$ Test de nullité d'un paramètre

$$\frac{\widehat{\theta}_j}{\widehat{se(\widehat{\theta}_j)}} \sim_{as} t(dl)$$