

# IRM - Sclérose en plaques

**Bruno Lecoutre**

ERIS, Laboratoire de Mathématiques Raphaël Salem

UMR 6085 C.N.R.S. et Université de Rouen

Avenue de l'Université, BP 12, 76801 Saint-Etienne-du-Rouvray

bruno.lecoutre@univ-rouen.fr

Internet : <http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris>

## Modèle binomial négatif

### Analyses séparées pour chaque visite

#### Modèle d'échantillonnage

Pour chaque visite  $j$ , distribution binomiale négative de paramètres  $\varphi_j$  et  $\omega_j$

$$y_{i<j>} | \varphi_j, \omega_j \sim \text{BinNeg}(\varphi_j, \omega_j)$$

On note  $n_j$  le nombre de sujets et  $T_j = \sum y_{i<j>}$  la somme des observations.

$$p(y_{i<j>} | \varphi_j, \omega_j) = \frac{\Gamma(y_i + \omega_j)}{\Gamma(y_i + 1)\Gamma(\omega_j)} \varphi_j^{\omega_j} (1 - \varphi_j)^{y_{i<j>}}$$

d'où la vraisemblance

$$v(\varphi_j, \omega_j | (y_{i<j>})) \propto \left( \prod_{i<j>} \frac{\Gamma(y_i + \omega_j)}{\Gamma(y_i + 1)} \right) \Gamma(\omega_j)^{-n_j} \varphi_j^{n_j \omega_j} (1 - \varphi_j)^{T_j}$$

#### Distributions initiales et finales

On considère les distributions initiales indépendantes :

- de densité  $p(\omega_j)$  pour  $\omega_j$
- *Bêta*( $a_0, b_0$ ) pour  $\varphi_j$

On en déduit les distributions finales :

- marginale pour  $\omega_j$

$$p(\omega_j | (y_{i<j>})) \propto p(\omega_j) \left( \prod_{i<j>} \Gamma(y_i + \omega_j) \right) \Gamma(\omega_j)^{-n_j} \int_0^1 \varphi_j^{n_j \omega_j + a_0 - 1} (1 - \varphi_j)^{T_j + b_0 - 1} d\varphi_j$$

d'où

$$p(\omega_j | (y_{i<j>})) \propto p(\omega_j) \left( \prod_{i<j>} \Gamma(y_i + \omega_j) \right) \frac{\Gamma(\omega_j)^{-n_j} \Gamma(n_j \omega_j + a_0)}{\Gamma(n_j \omega_j + T_j + a_0 + b_0)}$$

- conditionnelle à  $\omega_j$  pour  $\varphi_j$

$$\varphi_j | \omega_j, (y_{i<j>}) \sim \text{Bêta}(n_j \omega_j + a_0, T_j + b_0)$$

La distribution marginale de  $\omega_j$  peut être obtenue par une méthode “classique” d’intégration numérique. Pour obtenir la distribution des paramètres  $\varphi_j$ , on la génère par une méthode d’acceptation-rejet. On génère pour chaque  $\omega_j$  obtenu la distribution conditionnelle de  $\varphi_j$  pour obtenir la distribution marginale de  $\varphi_j$ .

On en déduit la distribution de paramètres dérivés tel que  $\rho_j = 1/\omega_j$  ou la moyenne  $\mu_j$  de la distribution d’échantillonnage, ou encore des combinaisons linéaires des  $\mu_j$ .

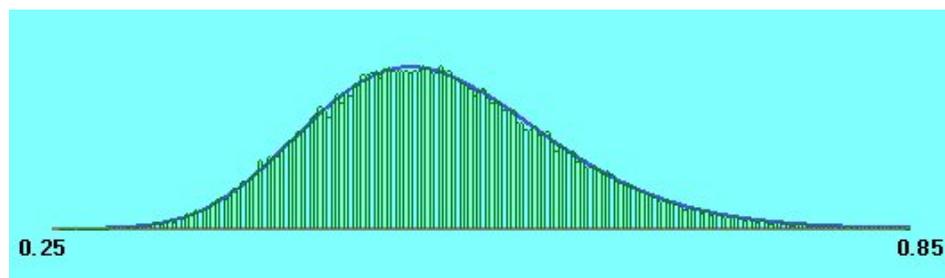
### Application numérique

On considère les distributions initiales indépendantes

- uniforme pour  $\log(\omega_j)$  :  $p(\omega_j) \propto 1/\omega_j$
- *Bêta*(0,0.5) (Jeffreys) pour  $\varphi_j$

On effectue 100 000 tirages

On obtient par exemple la distribution finale de  $\omega_3$  (visite = 300). La figure suivante montre la courbe obtenue à partir de la formule de la densité, la constante de proportionnalité étant calculée par intégration numérique et la courbe obtenue par simulation. On remarquera la qualité de l’approximation par simulation.



On trouve par simulation :

Moyenne = 0.5198

Ecart-type = 0.0863

Intervalle de crédibilité 95% : [0.370,0.706]

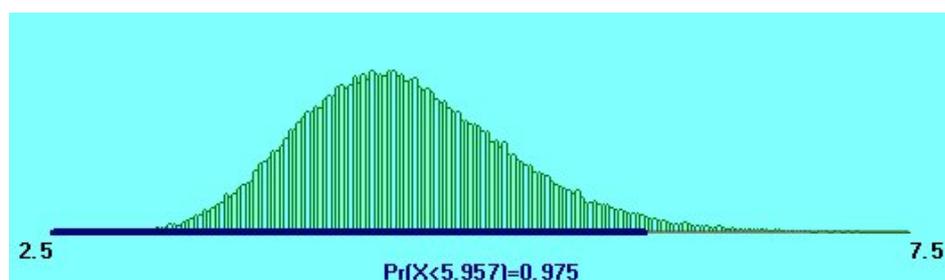
et pour  $\rho_3 = 1/\omega_3$

Moyenne = 1.9771

Ecart-type = 0.3296

Intervalle de crédibilité 95% : [1.416,2.703]

On a par exemple la distribution finale de  $\mu_3$  (visite = 300)



Moyenne = 4.5665

Ecart-type = 0.6320

Intervalle de crédibilité 95% : [3.482,5.957]

visite	moyenne	écart-type	IC 95%
300	4.5678	0.6450	[3.468,5.981]
400	5.6180	0.7711	[4.300,7.327]
500	4.8726	0.7231	[3.648,6.474]
600	4.9614	0.7769	[3.651,6.693]
700	4.7136	0.7579	[3.410,6.414]
800	4.7292	0.7365	[3.490,6.367]
900	5.1772	0.8197	[3.798,7.012]
1000	6.0972	1.1760	[4.202,8.780]
1100	5.2903	1.1454	[3.490,7.942]

### Analyses sur l'ensemble des visites (visite = 300,...1100)

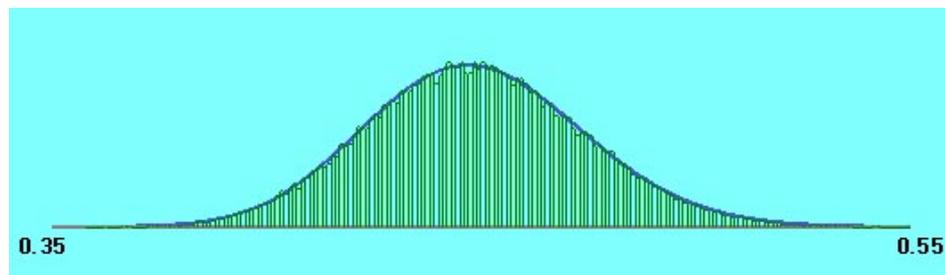
#### Analyses avec égalité des $\omega_j$

On pose pour tout  $j$   $\omega_j = \omega$ . La densité de la distribution finale de  $\omega$  s'obtient comme le produit des densités des  $\omega_j$

On considère les distributions initiales indépendantes

- uniforme pour  $\log(\omega)$  :  $p(\omega) \propto 1/\omega$
- *Bêta*(0, .5) (Jeffreys) pour  $\varphi_j$

La figure suivante montre la courbe de densité de la distribution finale de  $\omega$  obtenue à partir de la formule de la densité, la constante de proportionnalité étant calculée par intégration numérique et la courbe obtenue par simulation (100 000 tirages). On remarquera la qualité de l'approximation par simulation.



On trouve par simulation :

Moyenne = 0.4488

Ecart-type = 0.0256

Intervalle de crédibilité 95% : [0.400,0.501]

et pour  $\rho = 1/\omega$

Moyenne = 2.2353

Ecart-type = 0.1276

Intervalle de crédibilité 95% : [1.997,2.497]

Les valeurs obtenues sont remarquablement proches de celles obtenues dans les analyses avec SAS et WinBUGS.

	ICI	SAS	winBUGS
Moyenne	2.2353	2.2313	2.227
Ecart-type	0.1276	0.1264	0.1286
IC 95%	[1.997,2.497]	[1.992,2.489]	[1.991,2.492]

On considère les paramètres de la droite de régression ajustée aux  $\mu_j$ ,  $\mu_j = \bar{\mu} + v_j\pi$  ( $v_j = 300j$ ), qui sont des combinaisons linéaires des  $\mu_j$

- l'ordonnée à l'origine  $\bar{\mu} = \sum k_j\mu_j$

avec

$$k_3 = 0.6181152 \quad k_4 = 0.4820934 \quad k_5 = 0.3431388 \quad k_6 = 0.2029296 \quad k_7 = 0.07214609$$

$$k_8 = -0.05516379 \quad k_9 = -0.1476725 \quad k_{10} = -0.2208042 \quad k_{11} = -0.2947827$$

- la pente  $\pi = \sum k_j\mu_j$

avec

$$k_3 = -0.000742724 \quad k_4 = -0.0005368929 \quad k_5 = -0.0003282561 \quad k_6 = -0.0001209855$$

$$k_7 = 0.00007855138 \quad k_8 = 0.0002548757 \quad k_9 = 0.0003671113 \quad k_{10} = 0.000461452$$

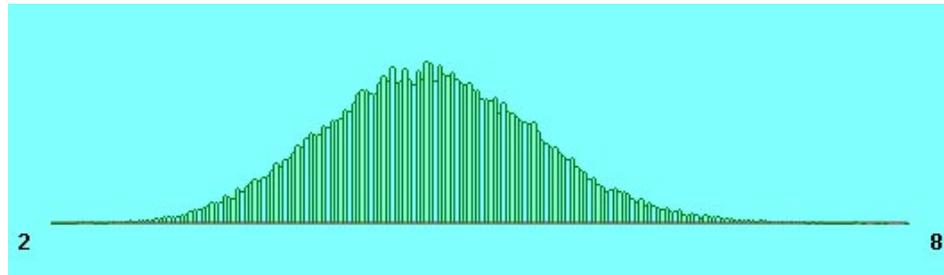
$$k_{11} = 0.0005668683$$

Les distributions finales conditionnelles à  $\omega$  des  $\varphi_j$  sont encore des distributions *Bêta* indépendantes

$$\varphi_j|\omega, (y_{i<j}) \sim \text{Bêta}(n_j\omega + a_0, T_j + b_0)$$

On obtient par simulation (50 000 tirages) - ici avec  $a_0 = 0$  et  $b_0 = 0.5$  - les distributions finales des paramètres  $\bar{\mu}$  et  $\pi$ .

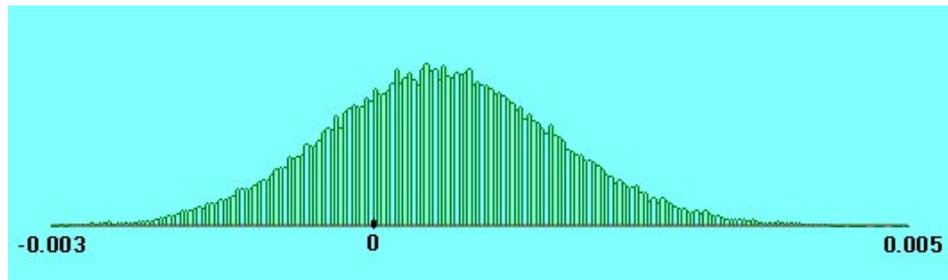
- Distribution finale de  $\bar{\mu}$



La moyenne est quasiment égale à celle obtenue dans l'analyse avec SAS, mais la dispersion est un peu plus grande, ce qui sans doute une conséquence du fait que le modèle utilisé pour ces dernières analyses présuppose la linéarité des espérances des  $\mu_j$  :  $E(\mu_j) = E(\alpha_j) + v_j E(\omega_j)$ .

	ICI	SAS	winBUGS
Moyenne	4.649	4.651	4.485
Ecart-type	0.7583	0.6917	0.6447
IC 95%	[3.212,6.186]	[3.347,6.039]	[3.189,5.711]

- Distribution finale de  $\pi$



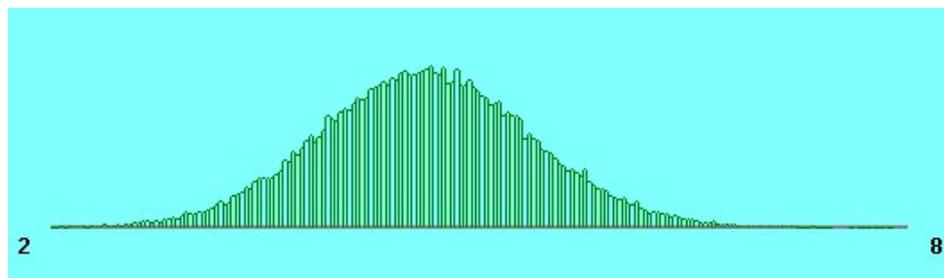
Dans ce cas la moyenne est supérieure à celle obtenue dans l'analyse avec SAS ; elle est en fait... quasiment la moyenne des deux moyennes obtenues avec SAS et WinBUGS, mais cette dernière analyse n'apparaît pas très fiable. Dans ce cas encore la dispersion est un peu plus grande :

	ICI	SAS	winBUGS
Moyenne	0.000633	0.000524	0.000754
Ecart-type	0.001095	0.000995	0.000943
IC 95%	[-0.00150,+0.00283]	[-0.00141,+0.00248]	[-0.00094,+0.00259]

#### Analyses sans égalité des $\omega_j$

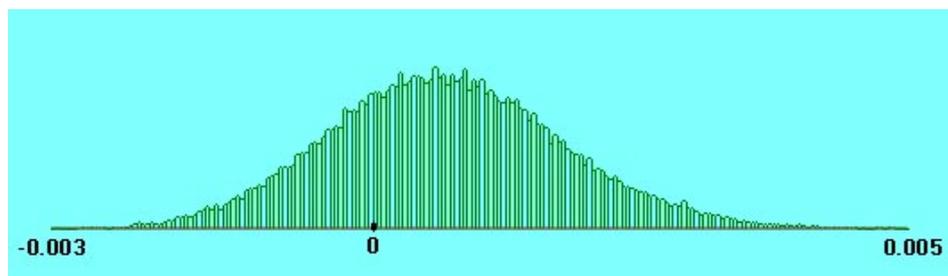
On ne suppose plus l'égalité des  $\omega_j$ . A partir des simulations des  $varphi_j$  effectuées comme dans la première section de manière indépendante, on obtient (50 000 tirages) les distributions finales des paramètres  $\bar{\mu}$  et  $\pi$ . Les résultats sont voisins de ceux obtenus avec égalité des  $\omega_j$ .

- Distribution finale de  $\bar{\mu}$



	sans égalité des $\omega_j$	avec égalité des $\omega_j$
Moyenne	4.649	4.619
Ecart-type	0.7583	0.767
IC 95%	[3.212,6.186]	[3.148,6.110]

- Distribution finale de  $\pi$



	avec égalité des $\omega_j$	sans égalité des $\omega_j$
Moyenne	0.000633	0.000681
Ecart-type	0.001095	0.001122
IC 95%	[-0.00150,+0.00283]	[-0.00144,+0.00299]

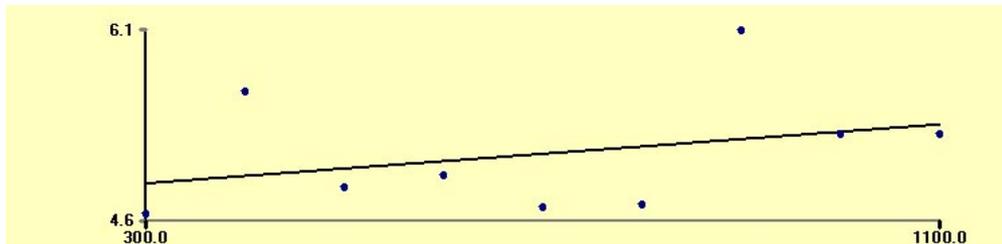
### Retour sur l'hypothèse de linéarité

On peut voir que l'hypothèse de linéarité n'est pas descriptivement très satisfaisante. La droite de régression usuelle ajustée aux moyennes observées est  $4.5739+0.00055v$ , d'où les moyennes observées et ajustées

visite	300	400	500	600	700	800	900	1000	1100
observées	4.483	5.525	4.769	4.798	4.548	4.629	5.057	5.885	5.081
ajustées	4.738	4.793	4.847	4.902	4.957	5.011	5.066	5.121	5.175
écart	-0.255	0.733	-0.078	-0.104	-0.409	-0.383	-0.009	0.764	-0.094

Plutôt que les moyennes observées, il paraît préférable de considérer les estimateurs des  $\mu_j$  (leurs espérances) calculés précédemment :

visite	300	400	500	600	700	800	900	1000	1100
$E(\mu_j)$	4.568	5.618	4.873	4.961	4.714	4.729	5.177	6.097	5.290



### Conclusion

Pas de difficulté pour généraliser cette analyse de manière à prendre en compte différents facteurs expérimentaux.

Pour un exemple sur un modèle voisin (Weibull) :

Mabika B.(1999) - *Analyse bayésienne des données de survie - Application à des essais cliniques en pharmacologie*. Thèse de doctorat de mathématique, Université de Rouen.

Lecoutre B., Mabika B., Derzko G. (2002) - Assessment and monitoring in clinical trials when survival curves have distinct shapes in two groups : a Bayesian approach with Weibull modeling. *Statistics in Medicine*, **21**, 663-674.