

Une application biomédicale avec WinBUGS et SAS/BGENMOD

Gérard Derzko & Stéphanie Galtier

21 juin 2007- AppliBUGS

Objectif de la présentation

Comparer, sur la base d'une application bio-médicale (étude clinique de phase II) simplifiée, quelques caractéristiques des logiciels

WINBUGS

SAS/v9.1.3

dans un modèle linéaire généralisé appliqué à des dénombrements (non Poissonniens).

Une application dans la sclérose en plaques

- En phase 2, l'imagerie par résonance magnétique (IRM) permet de dénombrer des lésions dans le cerveau liées à la sclérose en plaque (SP).
- Un traitement efficace doit être capable de faire diminuer ce nombre. Examen IRM mensuels pendant 1 an.
- L'analyse statistique des dénombrements de lésions au cours du temps utilise un modèle linéaire généralisé. La fonction de lien et la distribution restent à définir
- [Certaines données peuvent être manquantes du fait d'un processus non-ignorable].

Modèle de distribution (Sormani)

First the data is assumed globally exchangeable, and denoted simply Y . Sormani noticed that Y distribution is closely fitted with a negative binomial distribution. This property is first checked on the available data. This model can be viewed as a particular compound Poisson-Gamma distribution, as follows. Notations are borrowed from Sormani : :

$$\Pr(Y = y \mid \lambda) = \frac{\lambda^y}{y!} \exp(-\lambda)$$

if λ was fixed, actually $\lambda = \mu\epsilon$, where ϵ is Gamma(θ, θ)-distributed

$$\Pr(\epsilon \mid \theta) = \left(\frac{\theta}{\Gamma(\theta)} \right) (\theta\epsilon)^{(\theta-1)} \exp(-\theta\epsilon)$$

which gives (intégrant out in ϵ)

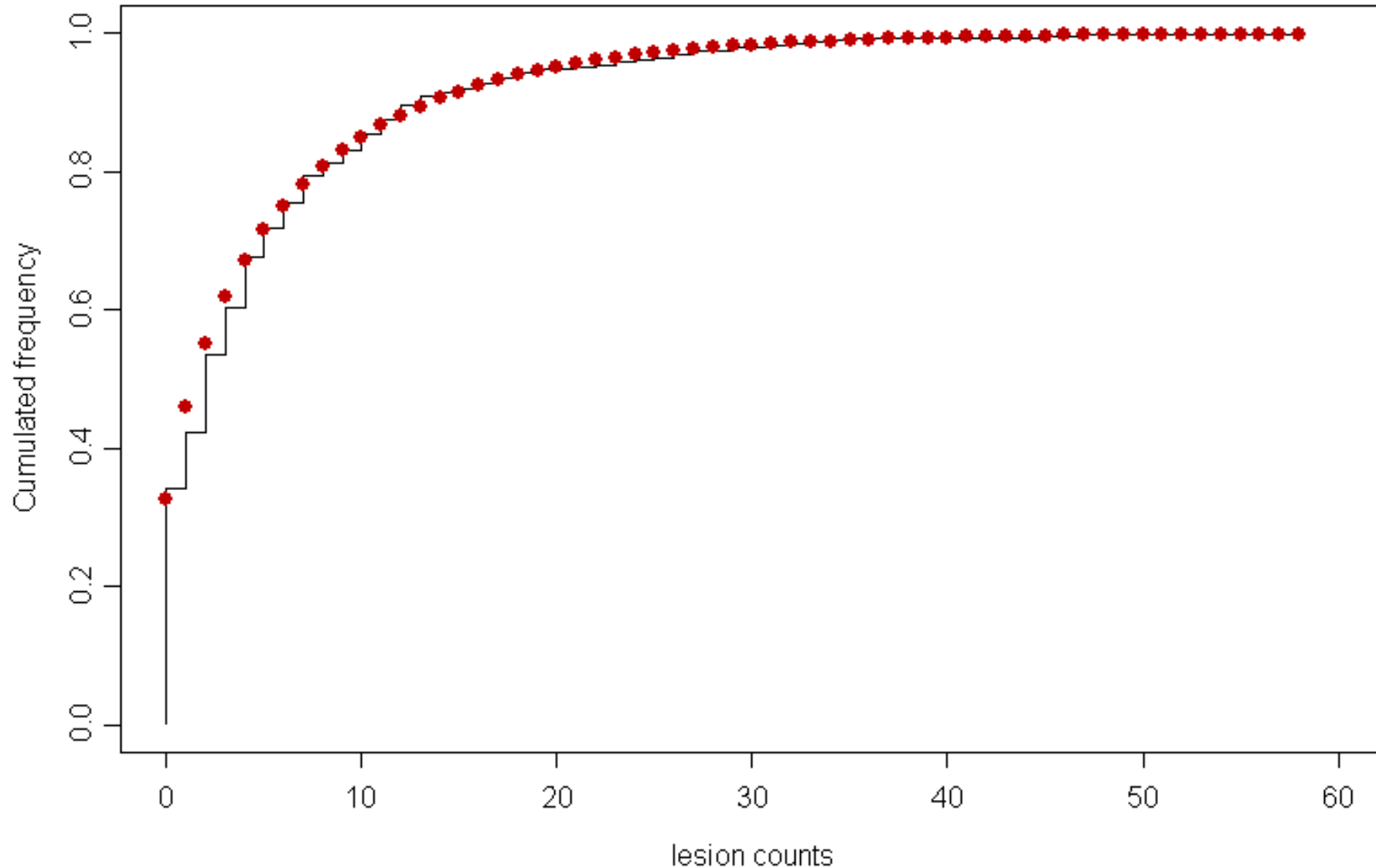
$$\Pr(Y = y \mid \mu, \theta) = \frac{\Gamma(y + \theta)}{\Gamma(\theta) y!} \frac{\theta^\theta \mu^y}{(\theta + \mu)^{(\theta+y)}} = \binom{\theta + y - 1}{\theta - 1} \left(\frac{1}{1 + \mu/\theta} \right)^\theta \left(\frac{\mu/\theta}{\mu/\theta + 1} \right)^y$$

Estimation des paramètres (moments)

[months]	[sample]	[$\hat{\mu}$]	[$\hat{\theta}$]	[\hat{k}]
3	118	4.483051	0.5549833	1.801856
4	118	5.525424	0.5762341	1.735406
5	117	4.769231	0.4430305	2.257181
6	113	4.849558	0.4634707	2.157634
7	114	4.596491	0.4998777	2.000489
8	105	4.628571	0.3459110	2.890917
9	88	5.056818	0.4048780	2.469880
10	78	5.884615	0.3908061	2.558814
11	74	5.081081	0.4536349	2.204416

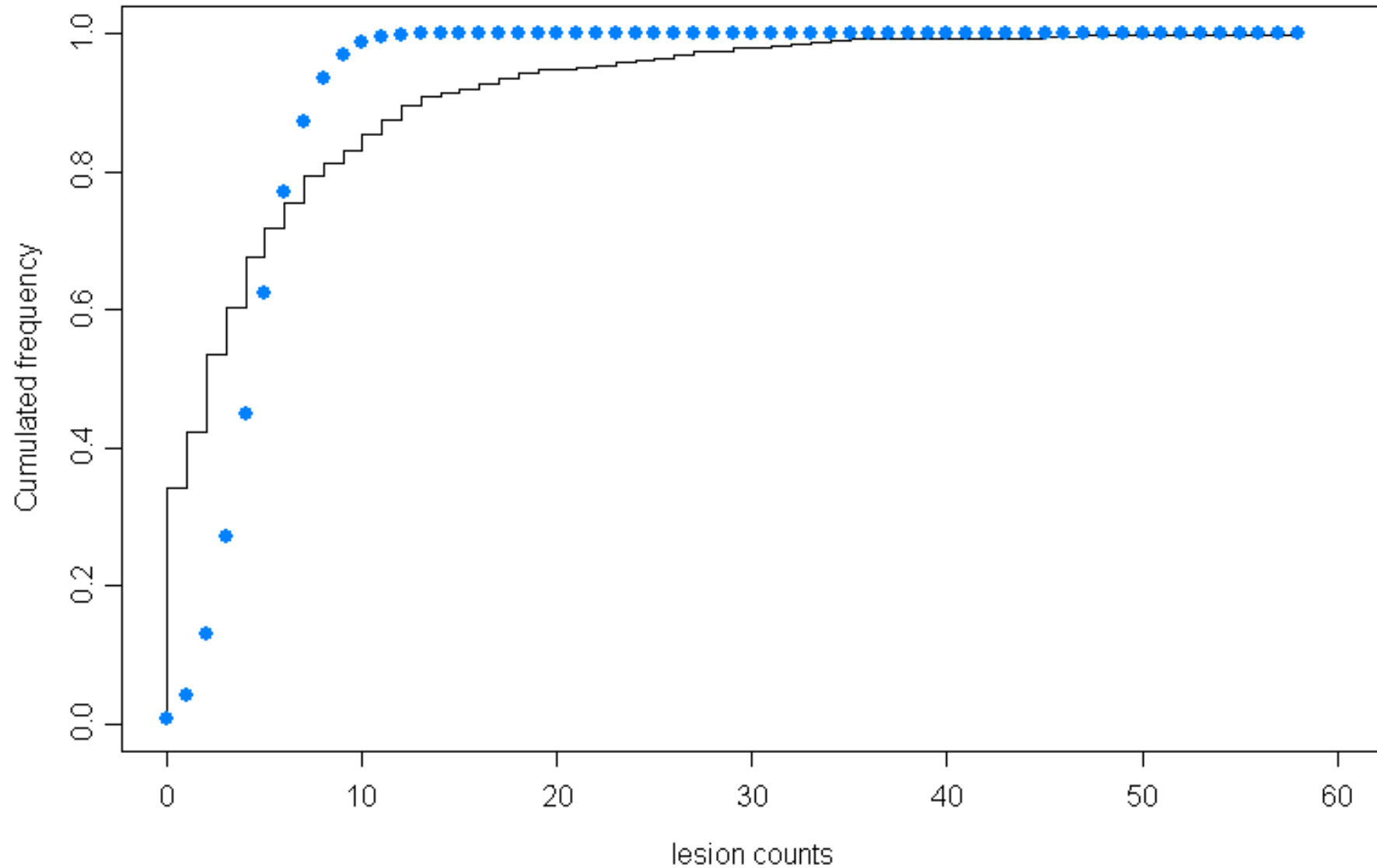
Modèle binomial négatif

MRI in MS : EDF versus Neg. Bin. estimation (red dots)



Modèle de Poisson

MRI in MS : EDF versus Poisson estimation (blue dots)



Modèle Linéaire Généralisé

$$Y_i \sim NB(\mu_i, k) \text{ avec } f(y) = \frac{\Gamma\left(y + \frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \cdot \frac{(k\mu)^y}{(1+k\mu)^{y+\frac{1}{k}}}$$

$$\mu_i = E(Y_i | X_i = x_i) = \alpha + \beta \cdot x_i$$

$$V(Y_i | X_i = x_i) = \mu_i + k\mu_i^2$$

Paramétrisation

WinBUGS

$$Y \sim NB(p, r) : f(y) = \frac{(y+r-1)!}{y!(r-1)!} p^r (1-p)^y \quad \text{avec } E(Y) = \frac{r(1-p)}{p}, \quad V(Y) = \frac{r(1-p)}{p^2}$$

$$\Rightarrow p = \frac{r}{r + E(Y)}$$

BGENMOD

$$Y \sim NB(\mu, k) : f(y) = \frac{\Gamma\left(y + \frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \cdot \frac{(k\mu)^y}{(1+k\mu)^{y+\frac{1}{k}}} \quad \text{avec } E(Y) = \mu, \quad V(Y) = \mu + k\mu^2$$

Écriture sous WinBUGS

```
model{
  for (i in 1:N){
    Y[i]~dnegbin(p[i],r)
    mu[i]<-alpha+beta*v[i]
    p[i]<-r/(r+mu[i])
  }
  r<- 1/dispersion
  alpha~dnorm(0,1.0E-6)
  beta~dnorm(0,1.0E-6)
  dispersion~dgamma(0.001,0.001)
}
```

```
list( N=925,
Y=c(0,0,5,6,3,0,0,0,0,0,4,26,25,10,23,19,0,0,0,0,0,0,0,0,0,0,3,0,9,15,8,8,6,6,5,5,2,0,0,0,
,0,0,0,0,0,4,6,3,0,0,3,0,4,2,10,7,8,4,1,0,2,2,4,0,2,7,0,0,0,0,2,0,1,0,4,3,3,3,1,0,0,0,0,5,
2,0,2,0,9,6,2,3,
```

Données sous WinBUGS ↔ Splus & R

```
x=read.table("U:/fichiers_GD/mri/mri.txt",header=T)
visit=x[,2]
irnbet1=x[,3]
visit=as.numeric(visit)
irnbet1=as.numeric(irnbet1)
dput(list(visit=visit),file="U:/visit.txt")
dput(list(irnbet1=irnbet1),file="U:/irnbet1.txt")
```

```
structure(list(irnbet1 = c(0, 0, 5, 6, 3, 0, 0, 0, 0, 0, 4, 26,
25, 10, 23, 19, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 9, 15, 8, 8,
6, 6, 5, 5, 2, 0, 0, 0, 0, 0, 0, 0, 0, 4, 6, 3, 0, 0, 3, 0, 4,
2, 10, 7, 8, 4, 1, 0, 2, 2, 4, 0, 2, 7, 0, 0, 0, 0, 2, 0, 1,
0, 4, 3, 3, 3, 1, 0, 0, 0, 0, 5, 2, 0, 2, 0, 9, 6, 2, 3, 0, 0,
0, 0, 0, 0, 0, 0, 0, 2, 14, 5, 4, 6, 5, 0, 0, 2, 8, 3, 0, 5,
0, 18, 11, 0, 0, 32, 22, 0, 3, 0, 0, 0, 2, 4, 1, 0, 0, 4, 4,
4, 11, 7, 7, 1, 9, 2, 13, 19, 7, 6, 7, 14, 7, 8, 8, 8, 7, 0,
9, 11, 14, 23, 3, 0, 15, 12, 12, 2, 5, 2, 3, 10, 20, 1, 3, 0,
4, 0, 0, 1, 5, 2, 0, 0, 1, 0, 0, 4, 4, 2, 0, 12, 1, 3, 1, 3,
0, 0, 4, 2, 0, 0, 3, 0, 0, 3, 0, 7, 0, 0, 0, 2, 0, 0, 0, 0, 0,
```

Écriture sous BGENMOD

```
ods rtf;  
ods graphics on;  


---

proc bgenmod data = d;  
MODEL IRNBET1 = VISID / dist = NB link = id;  
BAYES  
    INITIAL=init  
    /*seed=3*/  
    NBI=2000  
    NMC=5000  
    THIN=1  
    COEFFPRIOR=Normal /*Uniform ou Jeffreys*/  
    DISPERSIONPRIOR=Gamma /*Igamma ou Improper*/  
    DIAGNOSTICS=(autocorr ess heidelberger gelman geweke raftery)  
    PLOTS=(trace autocorr density)  
    SUMMARY=(descriptive interval corr);  
  
ODS OUTPUT PosteriorSample = posterior;  
run;  


---

ods graphics off;
```

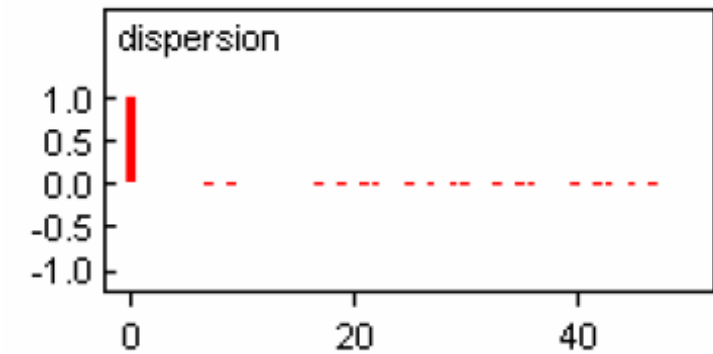
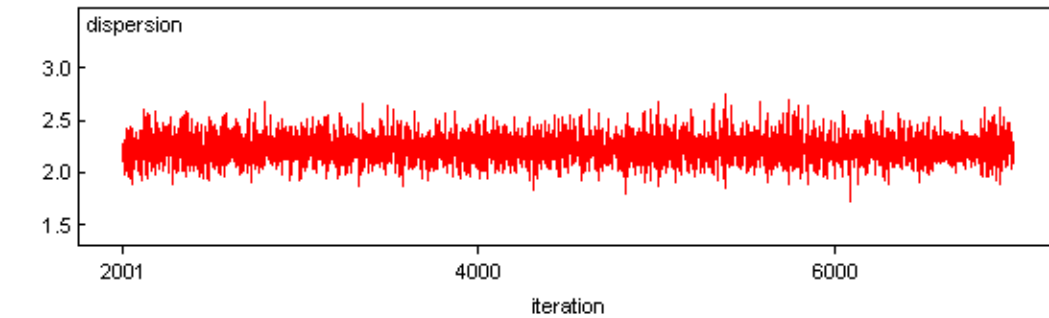
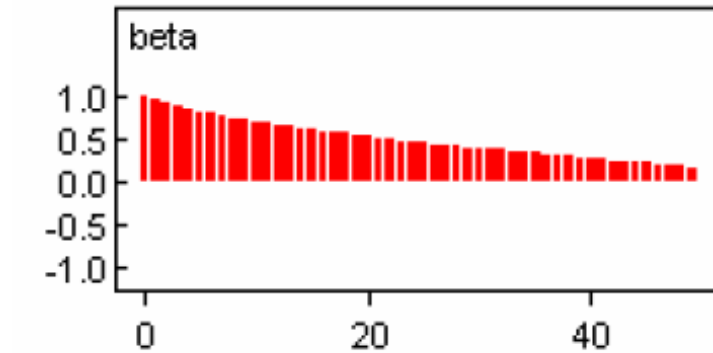
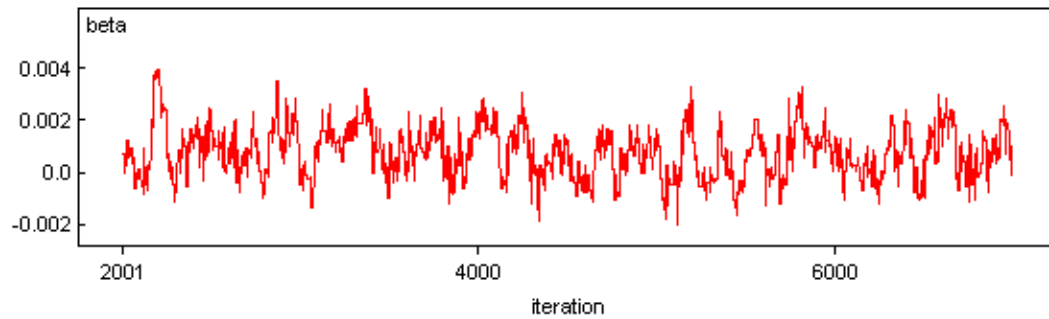
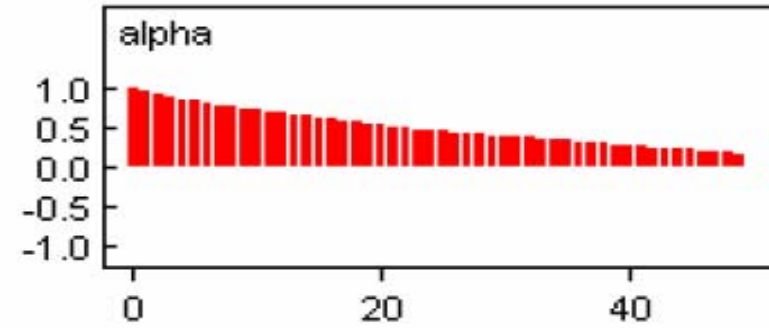
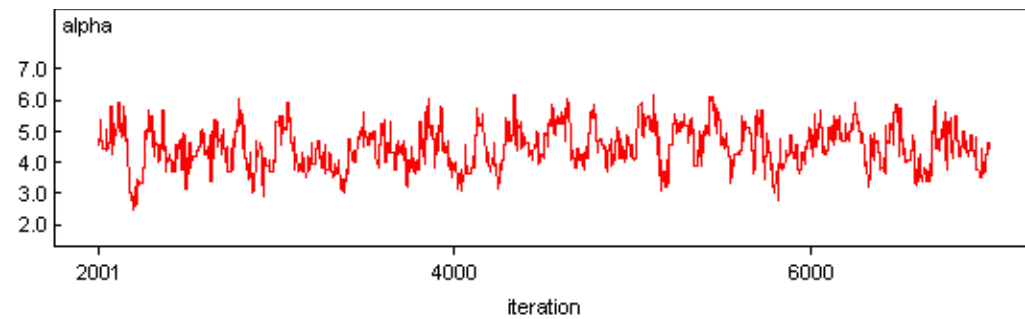
A priori BGENMOD

Independent Normal Prior for Regression Coefficients		
Parameter	Mean	Precision
Intercept	0	1E-6
VISID	0	1E-6

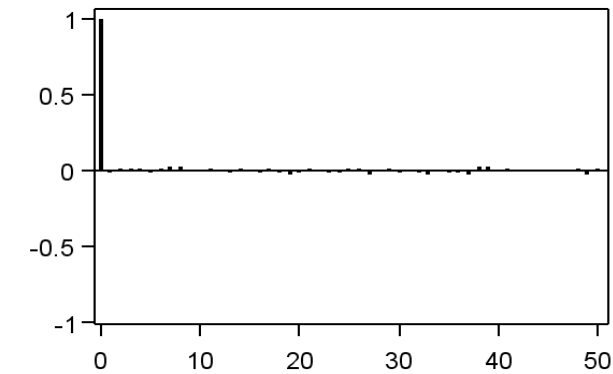
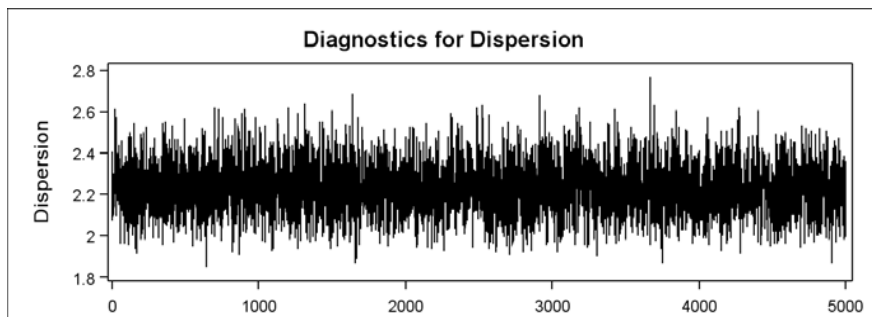
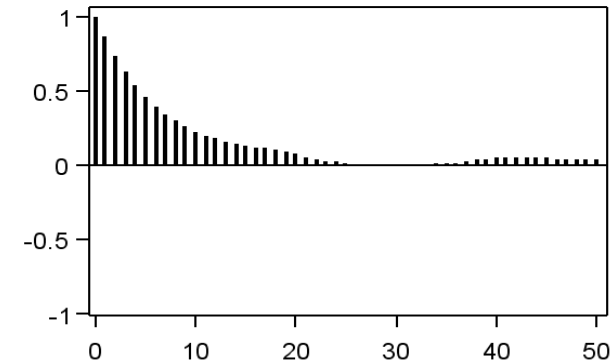
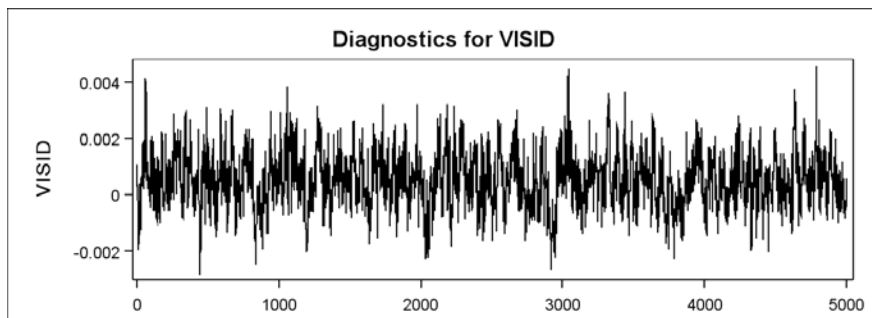
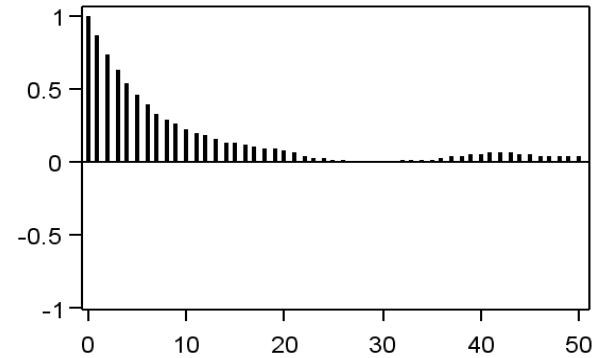
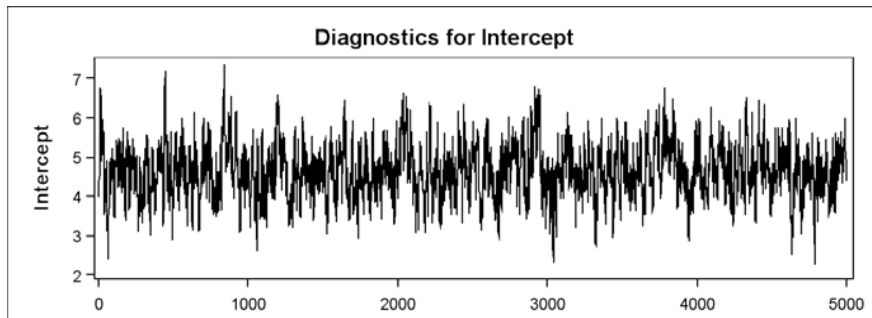
Independent Prior Distributions for Model Parameters			
Parameter	Prior Distribution	Hyperparameters	
		Shape	Inverse Scale
Dispersion	Gamma	0.001	0.001

Initial Values and Seeds				
Chain	_SEED_	Intercept	Visit identifier (X100)	Dispersion
1	3	4.600793	0.000525	2.219251
2	1964915885	2.506893	-0.00249	1.871084
3	687625807	6.694693	0.003537	2.632203

Graphiques de diagnostic WinBUGS



Graphiques de diagnostic BGENMOD

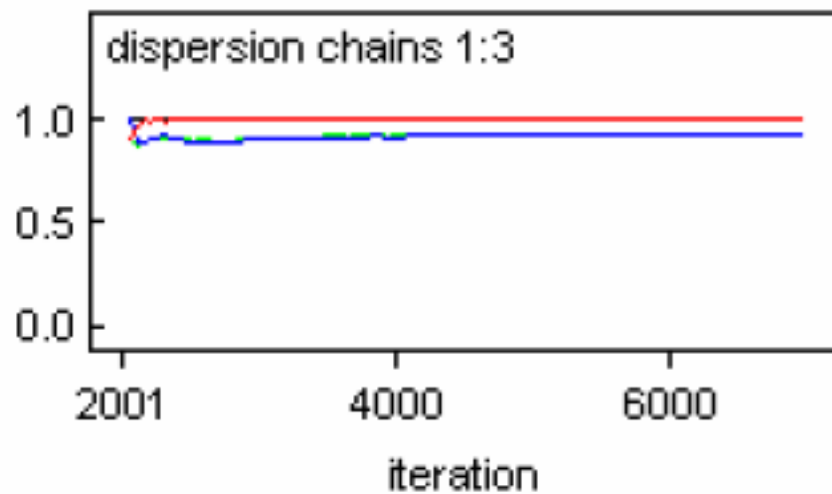
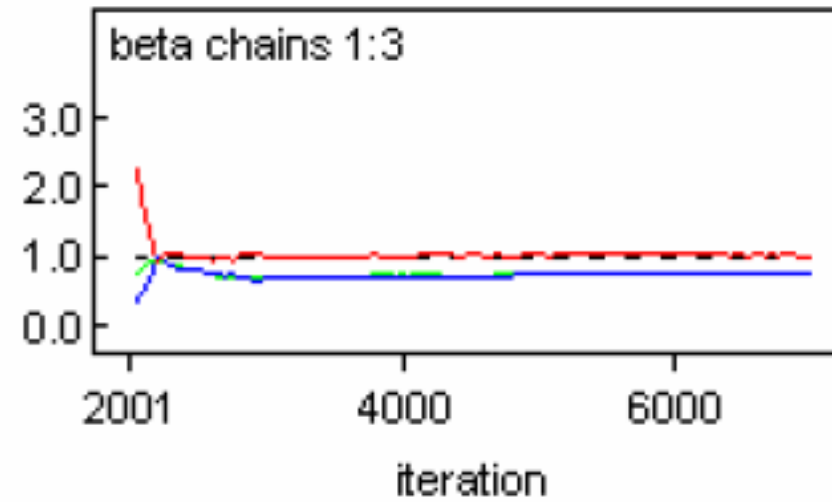
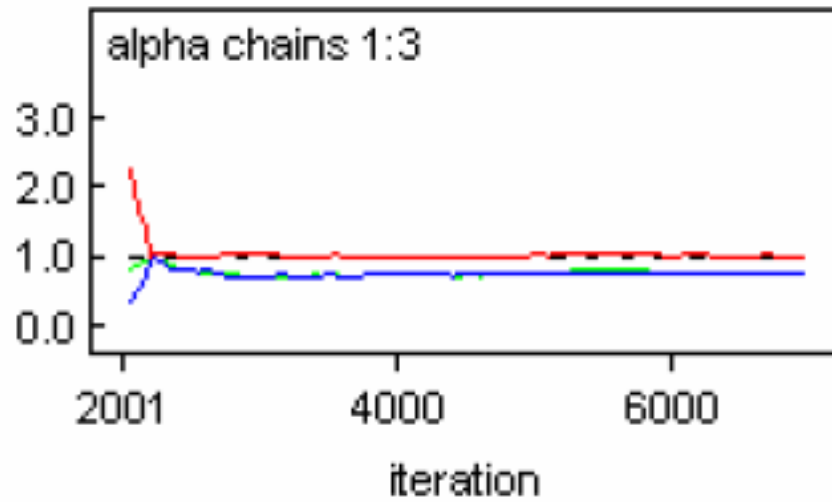


Autocorrélation WinBUGS et BGENMOD

lag		lag		lag		lag		lag	
1	0.9552	11	0.6952	21	0.5082	31	0.3809	41	0.2618
2	0.9178	12	0.675	22	0.4923	32	0.3686	42	0.2494
3	0.8863	13	0.6547	23	0.4769	33	0.3568	43	0.2395
4	0.8597	14	0.6369	24	0.4625	34	0.346	44	0.2286
5	0.8334	15	0.6174	25	0.4505	35	0.3366	45	0.2173
6	0.8077	16	0.5976	26	0.4392	36	0.3263	46	0.2068
7	0.7843	17	0.5787	27	0.4268	37	0.3139	47	0.1962
8	0.7618	18	0.5598	28	0.4144	38	0.3001	48	0.1838
9	0.7392	19	0.541	29	0.4019	39	0.2866	49	0.1711
10	0.7171	20	0.5236	30	0.3907	40	0.2747		

Autocorrelations of the Posterior Samples				
Parameter	Lag1	Lag5	Lag10	Lag50
Intercept	0.8616	0.4598	0.2294	0.0415
VISID	0.8633	0.4662	0.2284	0.0392
Dispersion	-0.0062	-0.0071	-0.0043	0.0178

Brook & Gelman sous WinBUGS (1/2)



Brook & Gelman sous WinBUGS (2/2)

-----80% interval-----

End iteration of bin	Unnormalized		Normalized as plotted		BGR ratio
	of pooled chains	mean within chain	of pooled chains	mean within chain	
→ 2051	2.018	0.8872	0.8296	0.3647	2.275 ←
2101	2.124	1.284	0.8729	0.5278	1.654
2151	2.283	1.675	0.9384	0.6884	1.363
2201	2.342	2.433	0.9628	1.0	0.9628
2251	2.288	2.23	0.9404	0.9167	1.026
2301	2.223	2.152	0.9138	0.8848	1.033
...
6751	1.866	1.84	0.767	0.7564	1.014
6801	1.863	1.842	0.7658	0.7571	1.011
6851	1.854	1.842	0.762	0.7571	1.006
6901	1.867	1.845	0.7676	0.7586	1.012
6951	1.857	1.847	0.7634	0.7591	1.006
7001	1.851	1.842	0.7611	0.7574	1.005 ←

Gelman-Rubin, Geweke, ESS sous BGENMOD

Gelman-Rubin Diagnostics		
Parameter	Estimate	97.5% Bound
Intercept	1.0071	1.0187
VISID	1.0065	1.0173
Dispersion	1.0005	1.0025

Geweke Diagnostics		
Parameter	z	Pr > z
Intercept	-0.8884	0.3743
VISID	0.9741	0.3300
Dispersion	0.8995	0.3684

Effective Sample Size			
Parameter	ESS	Correlation Time	Efficiency
Intercept	371.5	13.4596	0.0743
VISID	369.0	13.5516	0.0738
Dispersion	5000.0	1.0000	1.0000

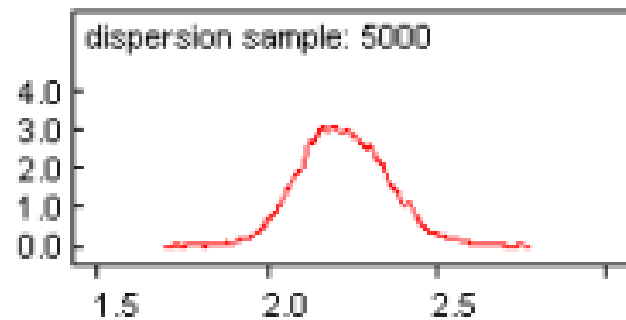
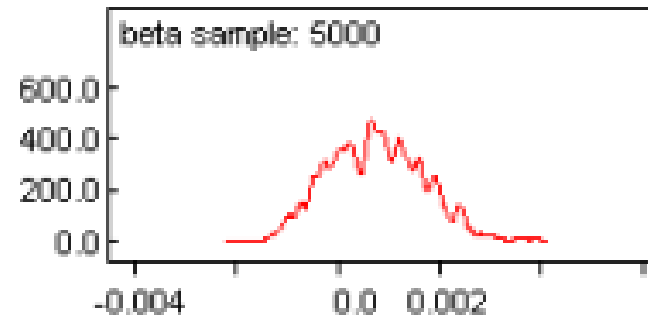
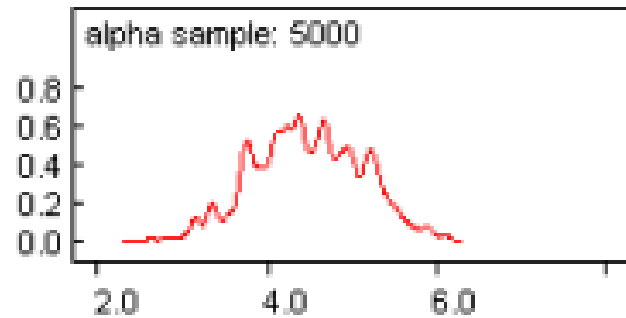
Autres diagnostic sous BGENMOD

Heidelberger-Welsh Diagnostics								
Parameter	Stationarity Test				Half-width Test			
	Cramer-von-Mises		Test Outcome	Iterations Discarded	Half-width	Mean	Relative Half-width	Test Outcome
	Stat	p						
Intercept	0.0709	0.7463	PASSED	0	0.0743	4.6513	0.0160	PASSED
VISID	0.0767	0.7110	PASSED	0	0.000108	0.000524	0.2052	FAILED
Dispersion	0.3505	0.0980	PASSED	0	0.00385	2.2313	0.00173	PASSED

Raftery-Lewis Diagnostics				
(Quantile=0.025 Accuracy=+/-0.005 Probability=0.95 Epsilon=0.001)				
Parameter	Number of Samples			Dependence Factor
	Burn-in	Total	Minimum	
Intercept	15	17513	3746	4.6751
VISID	23	27307	3746	7.2896
Dispersion	2	3866	3746	1.0320

Estimation sous WinBUGS

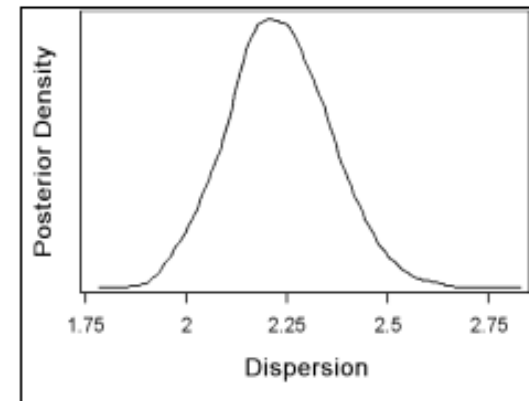
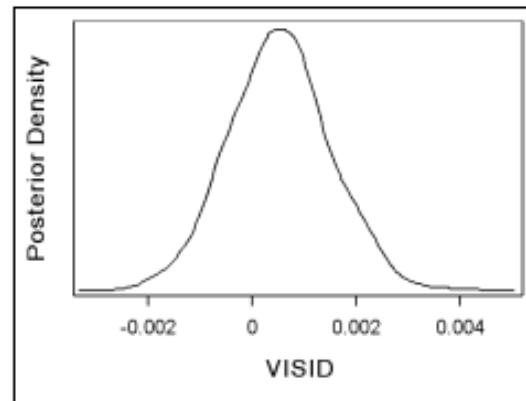
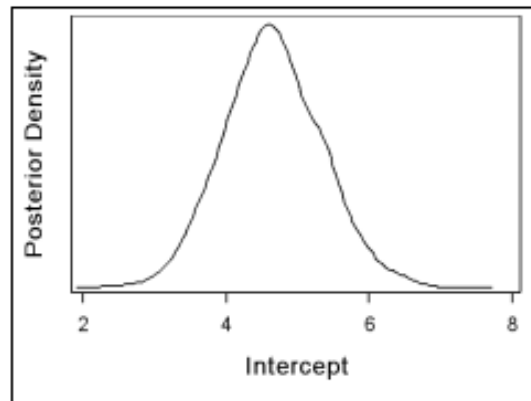
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	4.485	0.6447	0.05628	3.189	4.464	5.711	2001	5000
beta	7.537E-4	9.431E-4	8.216E-5	-9.397E-4	7.404E-4	0.002594	2001	5000
dispersion	2.227	0.1286	0.001686	1.991	2.221	2.492	2001	5000



Estimation sous BGENMOD

Descriptive Statistics of the Posterior Samples						
Parameter	N	Mean	Standard Deviation	Quantiles		
				25%	50%	75%
Intercept	5000	4.6513	0.6917	4.1953	4.6298	5.1083
VISID	5000	0.000524	0.000995	-0.00014	0.000509	0.00115
Dispersion	5000	2.2313	0.1264	2.1453	2.2272	2.3148

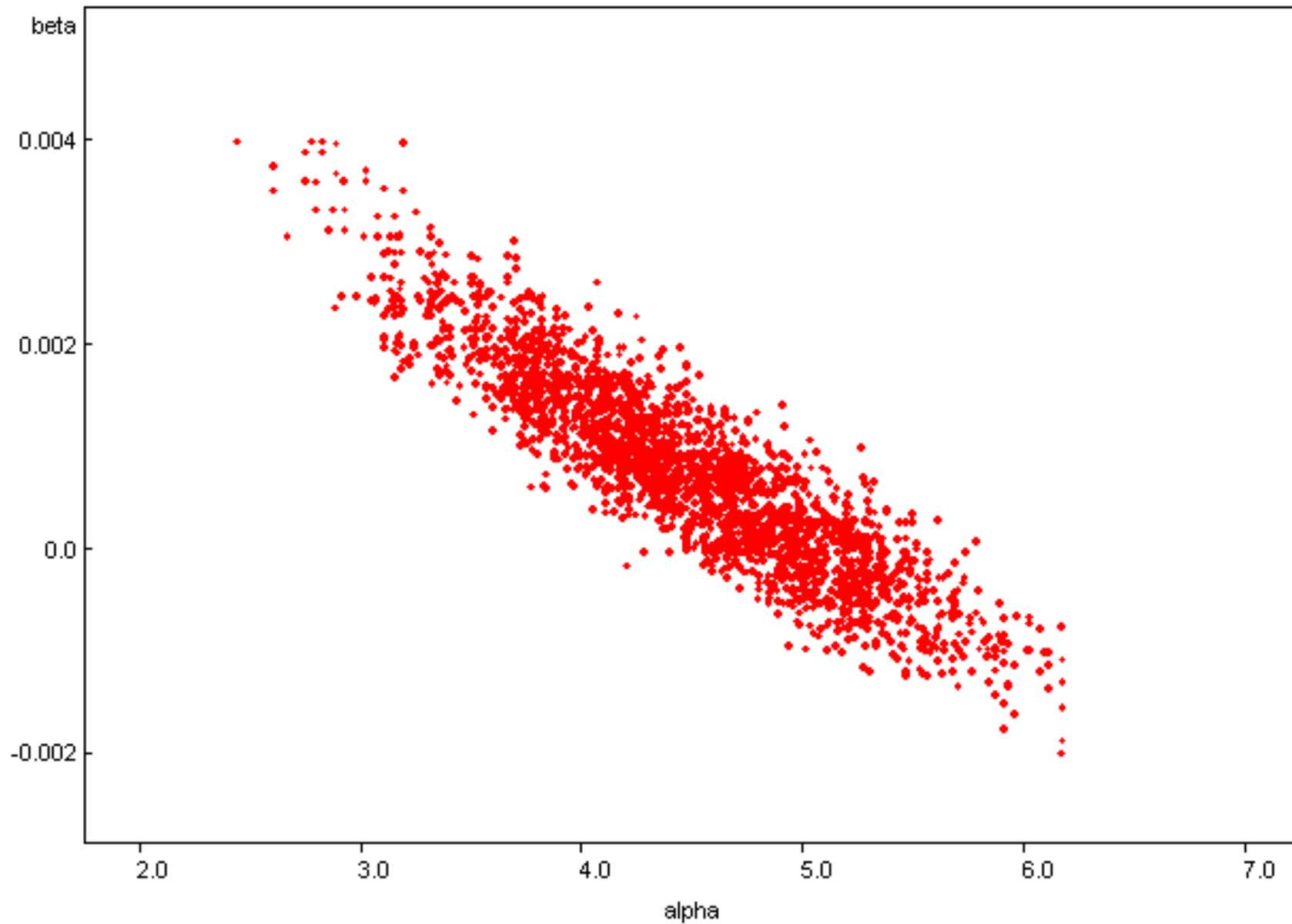
Interval Statistics of the Posterior Samples					
Parameter	Alpha	Credible Interval		HPD Interval	
Intercept	0.050	3.3466	6.0387	3.3414	6.0258
VISID	0.050	-0.00141	0.00248	-0.00142	0.00245
Dispersion	0.050	1.9923	2.4888	1.9903	2.4840



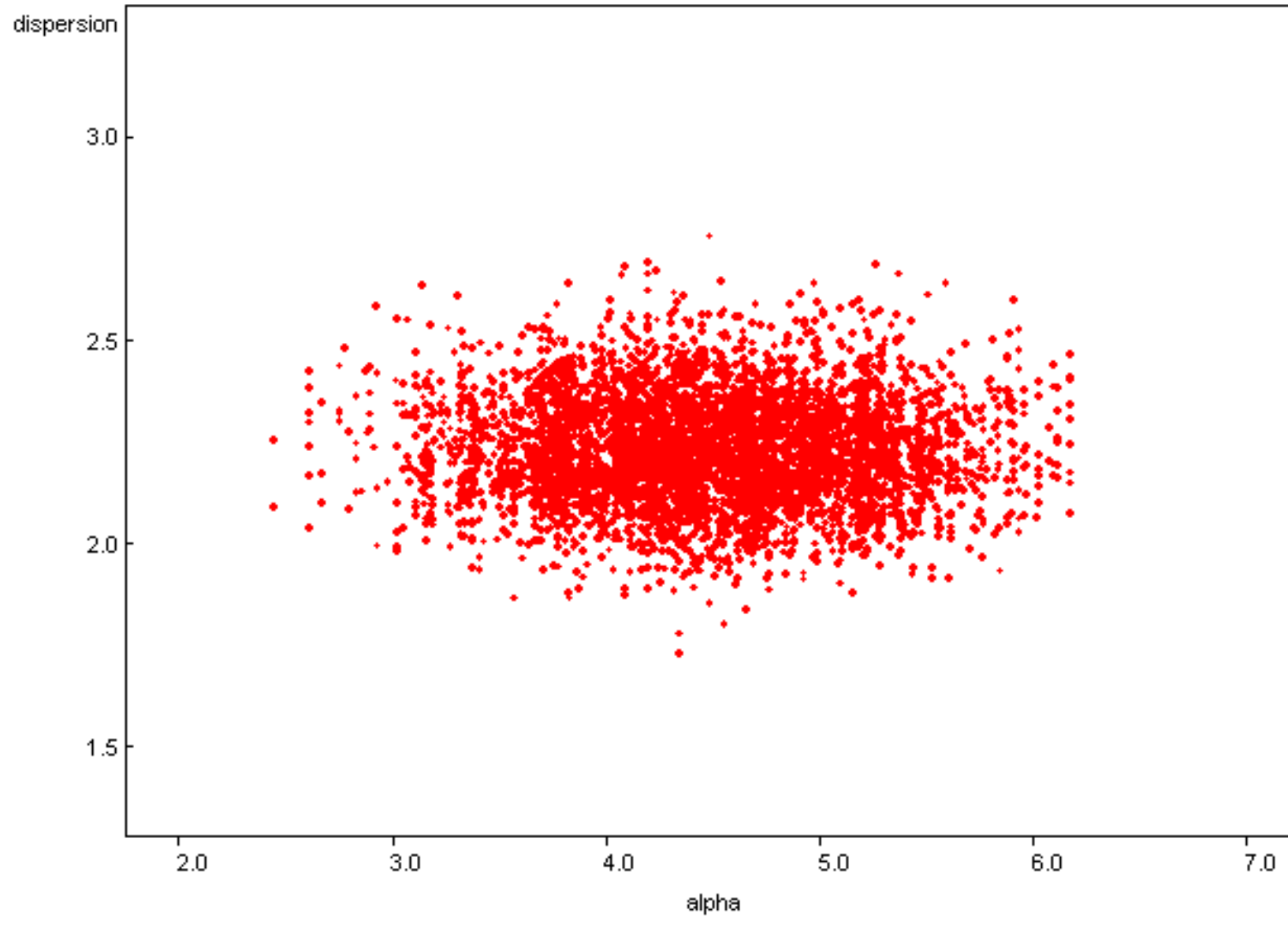
Comparaison

	alpha		beta		dispersion	
	WinBUGS	BGENMOD	WinBUGS	BGENMOD	WinBUGS	BGENMOD
mean	4,485	4,651	0,0007537	0,000524	2,227	2,2313
sd	0,6447	0,6917	0,0009431	0,000995	0,1286	0,1264
2,50%	3,189	3,3466	-0,0009397	-0,00141	1,991	1,9923
median	4,464	4,6298	0,00074	0,000509	2,221	2,2272
97,50%	5,711	6,0387	0,002594	0,00248	2,492	2,4888

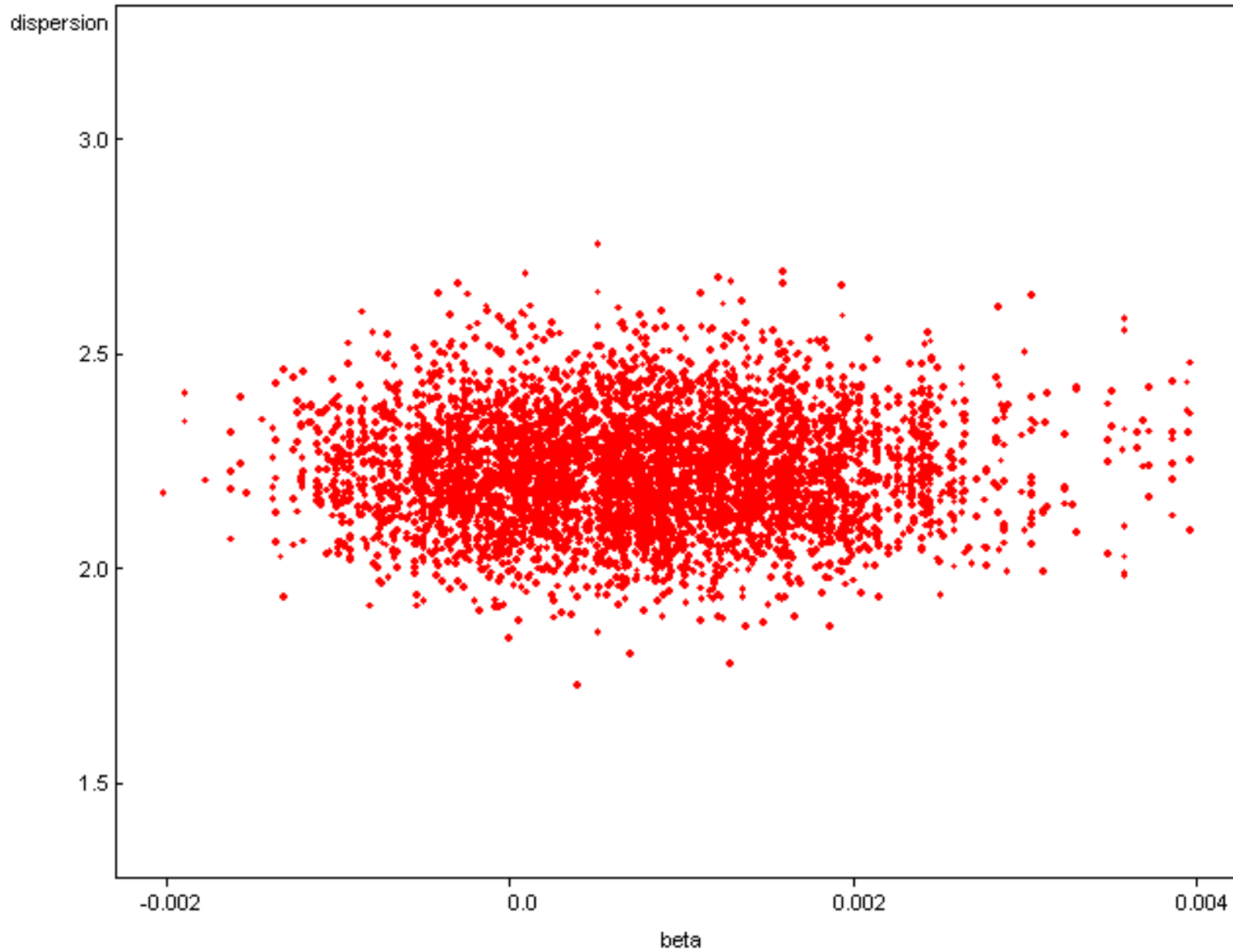
Bivariate posterior scatter plots



Bivariate posterior scatter plots



Bivariate posterior scatter plots



Backup - Estimating model parameters

The estimators below are obtained by the method of moments :

$$\hat{\theta} = \frac{\bar{y}^2}{s^2 - \bar{y}}$$
$$\widehat{\left(\frac{\mu}{\theta}\right)} = \frac{s^2}{\bar{y}} - 1$$

where \bar{y} and s are the sample mean and standard deviation. \bar{y} is also the ML estimator for μ . ML estimation for θ requires numerical resolution.

Overall (total lesion counts between 1 and 9 months (925 counts)), one gets $\hat{\mu} = 4.948108$ (lesions by month), which is interpreted as the average monthly count in the population, and $\hat{\theta} = 0.4534627$, corresponding to $\hat{k} = 1/\hat{\theta} = 2.205253$. The fit is excellent (see graphs MRIBin-neg.pdf).