

UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE NANTES

Mars 2023

Analyse de la variation de salmonelle en fonction de la dose de quinoléine

MINETTE Cloée, BAKHOUCHE Sarah,
MAILLARD Solenn, SAUVAGEOT Aurore

Table des matières

1	Introduction	2
2	Modèle	2
2.1	Choix du modèle	2
2.2	Construction du modèle	3
3	Analyse et interprétation des résultats	4

1 Introduction

Le jeu de données SALM fut étudié par Breslow en 1984. Il contient des données d'essais de mutagénicité de la salmonelle sur trois boîtes de Petri (PLATES) exposées à six doses différentes de quinoléine (DOSES). Dans ce jeu, on rapporte le nombre de colonies révertantes (qui ont mutées et sont revenues au type normal) de Salmonelle TA98 en fonction de la dose de quinoline administrée :

0	10	33	100	333	10000
15	16	16	27	33	20
21	18	26	41	38	27
29	21	33	60	41	42

TABLE 1 – Dose de quinoline administrée (μg)

On peut se donner une représentation graphique de ce jeu de données :

GRAPHE

2 Modèle

2.1 Choix du modèle

Notre jeu de données compte le nombre d'observations d'un certain évènement (révertance) dans une unité d'échantillonnage définie, véhiculant dans un espace discret. Ces données de dénombrement sont souvent soumises à une surdispersion (plus de variation dans les données que prévu). Breslow suggère qu'avec ce type de données, il est efficace d'utiliser des modèles bayésiens de Poisson hiérarchiques à effets aléatoires afin de mieux capturer la surdispersion. C'est pourquoi on décide d'appliquer ce modèle au jeu de données SALM. Il est exprimé comme suit, avec y_{ij} le nombre de colonies révertantes de Salmonelle TA98, x_{ij} la dose donnée aux boîtes j , $i \in \{1, 2, 3, 4, 5, 6\}$ et $j \in \{1, 2, 3\}$:

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\mu_{i,j}) \\ \log(\mu_{ij}) &= \alpha + \beta \log(x_{ij} + 10) + \gamma x_{ij} + \lambda_{ij} \\ \lambda_{i,j} &\sim \text{Normal}(0, \tau) \end{aligned}$$

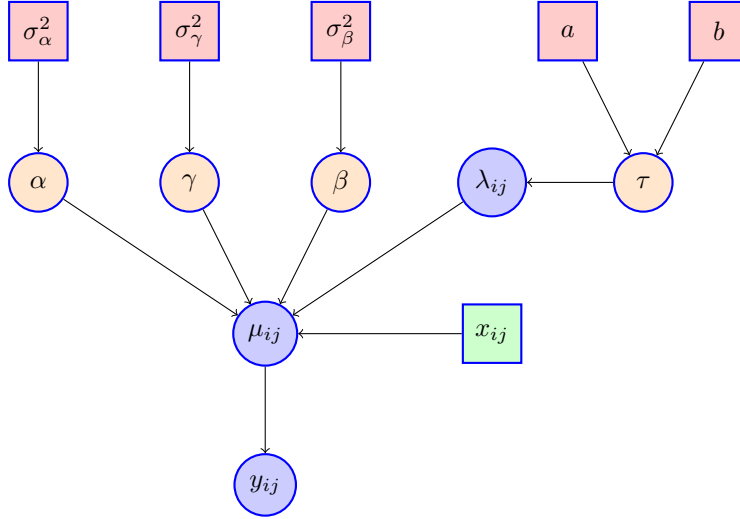
Le paramètre α correspond à l'ordonnée à l'origine (intercept), β modélise l'effet mutagénique de la dose, et τ celui toxique de la dose. Le paramètre λ_{ij} représente les effets aléatoires. Ces paramètres, indépendants, ont des lois a priori non informatives données. On modélise l'effet mutagénique de la quinoline avec $\log(x_{ij} + 10)$, 10 étant la plus petite dose différente de 0.

Par ailleurs, nous utiliserons une méthode d'échantillonnage de Gibbs pour calculer les lois de nos estimateurs. L'algorithme de Gibbs est particulièrement adapté à notre jeu de données

puisque celui-ci est multidimensionnel, et qu'il contient une variable latente μ_{ij} (non mesurable directement, mais à la base de la variable observée y_{ij}).

2.2 Construction du modèle

On peut représenter notre modèle par un graphe orienté acyclique (Directed Acyclic Graph (DAG)) :



On suppose que la loi a priori des paramètres α , β , γ est une loi normale de moyenne 0 et de variances respectives σ_α^2 , σ_β^2 et σ_γ^2 fixées. Ce choix permet d'exprimer le manque d'information sur les paramètres de régression. La loi a priori de τ est une inverse Gamma de paramètres ($a = 0.001, b = 0.001$), ce choix peut être justifié par le fait que cette loi est spécifique à la modélisation d'une grande variété de phénomènes se déroulant au cours du temps où par essence, le temps écoulé est une grandeur réelle positive.

On peut en outre justifier ces choix par le calcul des lois conditionnelles de α , β , γ et μ_{ij} .

Nous avons l'échantillonneur de suivant :

$$\begin{aligned}\Pi(y_{ij}|\dots) &\propto \Pi(y_{ij}|\mu_{ij}) \propto \sum_{k=0}^{k=\infty} \frac{\mu_{ij}^k e^{\mu_{ij}}}{k!} \\ \Pi(\mu_{ij}|\dots) &\propto \Pi(\mu_{ij}|\alpha, \beta, \gamma, \lambda_{ij}, x_{ij}) \Pi(y_{ij}|\mu_{ij}) \propto e^{\alpha + \beta \log(x_{ij}+10) + \gamma x_{ij} + \lambda_{ij}} \sum_{k=0}^{k=\infty} \frac{\mu_{ij}^k e^{\mu_{ij}}}{k!} \\ \Pi(x_{ij}|\dots) &\propto \Pi(x_{ij}) \Pi(\mu_{ij}|x_{ij}, \lambda_{ij}, \beta, \gamma, \alpha) \propto \Pi(x_{ij}) e^{\alpha + \beta \log(x_{ij}+10) + \gamma x_{ij} + \lambda_{ij}} \\ \Pi(\lambda_{ij}|\dots) &\propto \Pi(\lambda_{ij}|\tau) \Pi(\mu_{ij}|\lambda_{ij}, \beta, \gamma, \alpha, x_{ij}) \propto \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{\lambda_{ij}^2}{2\tau^2}} \\ \Pi(\alpha|\dots) &\propto \Pi(\alpha) \prod_{ij} \Pi(\mu_{ij}|\alpha, \beta, \gamma, \lambda_{ij}, x_{ij}) \propto \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{\alpha^2}{2\sigma_\alpha^2}} \prod_{ij} e^{\alpha + \beta \log(x_{ij}+10) + \gamma x_{ij} + \lambda_{ij}} \\ &\propto \frac{1}{\sigma_\alpha} e^{-\frac{\alpha^2}{2\sigma_\alpha^2} + n\alpha + \sum_{ij} \beta \log(x_{ij}+10) + \sum_{ij} \gamma x_{ij} + \sum_{ij} \lambda_{ij}}\end{aligned}$$

Pour β et γ , les lois conditionnelles s'écrivent de la même manière que la loi conditionnelle de α avec respectivement σ_β^2 et σ_γ^2 comme variance.

3 Analyse et interprétation des résultats

On peut observer les trajectoires des paramètres, leurs graphes d'autocorrélation et de densité empirique.

GRAPHE ET TABLEAU

Les trajectoires des paramètres nous montrent une stationnarité de nos estimations. On a en outre une faible autocorrélation, décroissante. La distribution empirique des estimateurs est unimodale.