Évolution du nombre d’incidents dans les mines de charbon (inférence Bayésienne)

GMMA 106

Cas d'étude

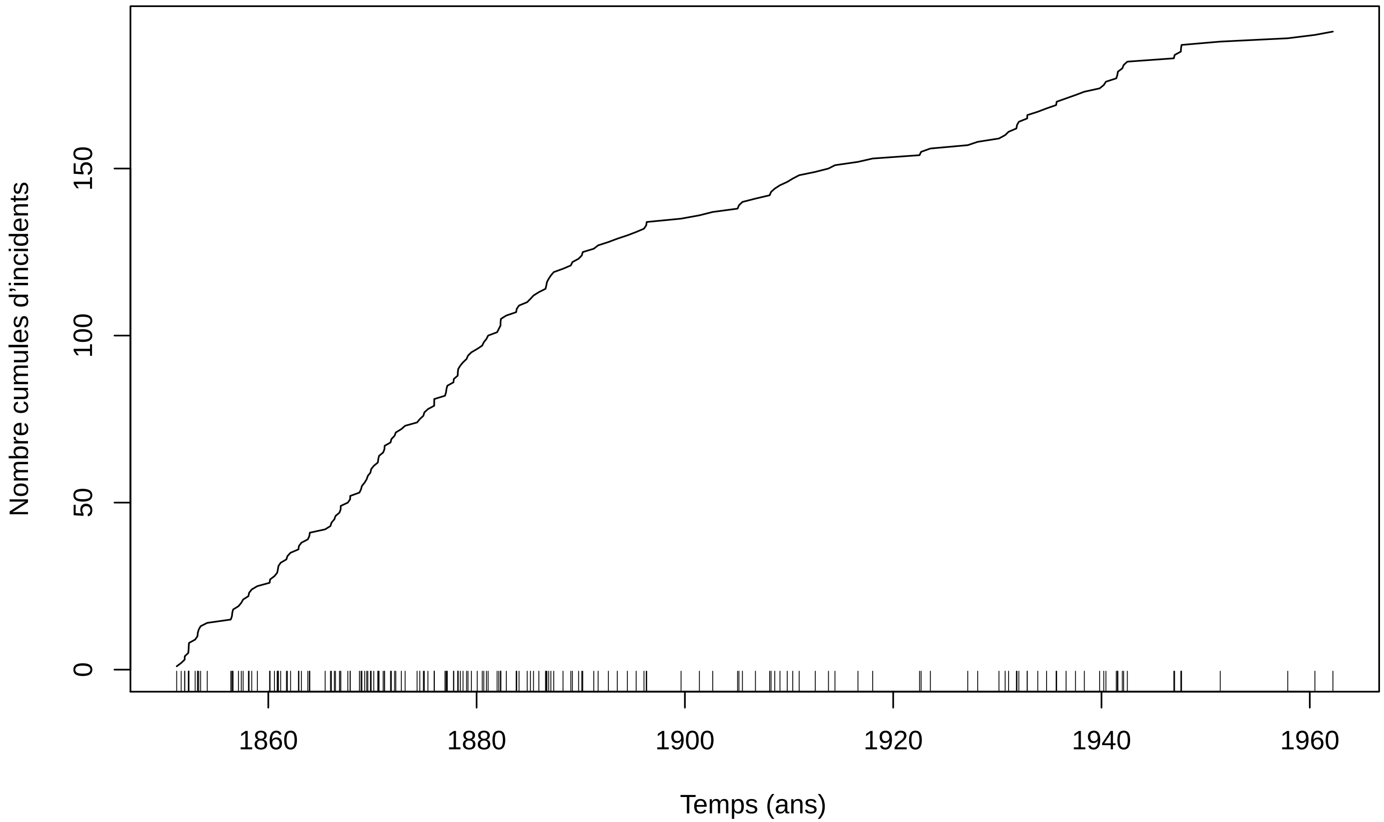

Figure 1: *Évolution du nombre d'incidents dans les mines de charbon au Royaume–Uni.*

- ☐ Quelles remarques/constats pouvez vous faire à la vue de la Figure 1 ?

Les données

```
> library(SMPracticals);data(coal)
> coal
        date
1   1851.203
2   1851.632
3   1851.969
4   1851.975
.
.
188 1951.405
189 1957.883
190 1960.489
191 1962.220
```

date Date de l'incident grave (+ de 10 morts)

Objectifs et éléments parcourus

- Notre objectif est de proposer un modèle statistique pour modéliser l'évolution du nombre d'incidents graves dans les mines de charbon.
- Ceci nous permettra de croiser les objets suivants :
  - Loi a priori, a posteriori, loi conjuguée
  - Méthodes MCMC: Metropolis–Hastings, Échantillonneur de Gibbs
  - Processus de Poisson

1. Processus de Poisson

Processus ponctuels

- Nos données sont en fait des dates aléatoires que l'on modélisera par un processus ponctuel
- Un processus ponctuel $\mathcal{P}$ est simplement une collection de points (aléatoires)
- Pour tout ensemble (convenable) $\mathcal{A}$, le processus ponctuel est identifié par la mesure

$$\mathcal{P}_n(\mathcal{A}) = \sum_{j=1}^{n} \delta_{X_j}(\mathcal{A}),$$

  où les X_j sont les points et n est le nombre total de point–éventuellement aléatoire.
- $\mathcal{P}_n$ est une mesure ponctuelle aléatoire appelée mesure de comptage
- Le processus ponctuel le plus connu est le processus (ponctuel) de Poisson

Définition 1. Soient $\mathcal{X} \subset \mathbb{R}^d$ et une fonction $\Lambda(\mathcal{A}) \geq 0$ définie pour tout mesurable $\mathcal{A} \subset \mathcal{X}$. Un processus de Poisson $\mathcal{P}$ sur $\mathcal{X}$ ayant pour mesure d'intensité Λ vérifie

- Le nombre de points de $\mathcal{P}$ tombant dans $\mathcal{A}$, noté $N(\mathcal{A})$, suit une loi de Poisson de moyenne $\Lambda(\mathcal{A})$;
- Si $\mathcal{A}, \mathcal{B} \subset \mathcal{X}$ sont disjoints, alors $N(\mathcal{A})$ et $N(\mathcal{B})$ sont indépendantes.

Si $\mathcal{A} = [a_1, x_1] \times \cdots \times [a_d, x_d]$, et si

$$\lambda(x_1, \ldots, x_d) = \frac{\partial^d \Lambda(\mathcal{A})}{\partial x_1 \cdots \partial x_d}$$

existe, alors λ est appelée la fonction d'intensité de $\mathcal{P}$.

Remarque. Lorsque $\Lambda(\mathcal{A}) \propto |\mathcal{A}|$ on parle de processus de Poisson homogène.

Illustration

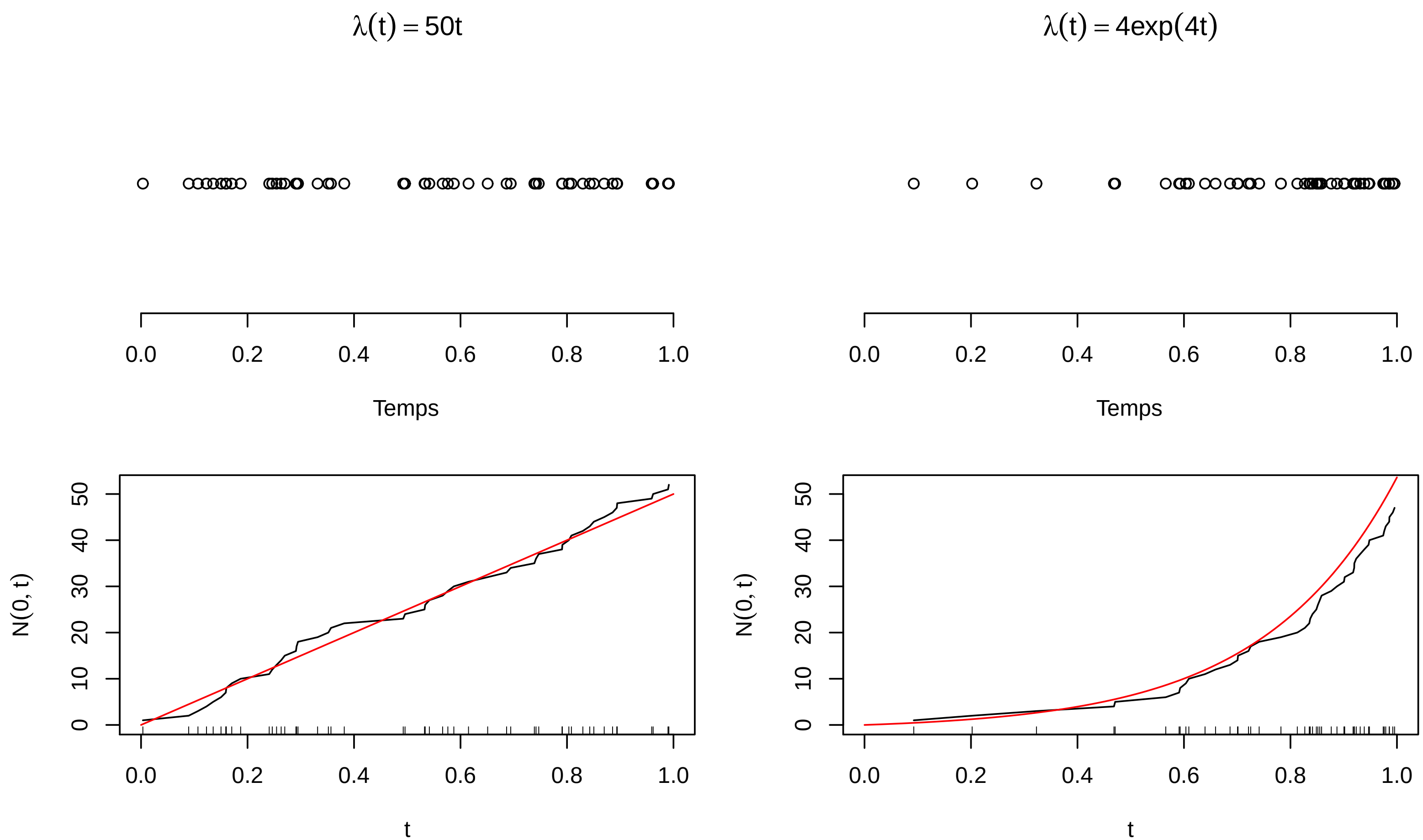

Figure 2: *Deux processus de Poisson sur* $(0, 1)$*. La ligne du haut représente les points; celle du bas le processus de comptage. Les courbes rouges représentent la fonction* $t \mapsto \Lambda\{(0, t)\}$.

Soient $t_1, \dots, t_n$ les dates d'occurrences issues d'un processus de Poisson (non homogène) de fonction d'intensité $\lambda(t)$, $t \in [0, L]$. La vraisemblance s'écrit alors

$$\left\{\prod_{j=1}^{n} \lambda(t_j)\right\} \times \exp\left\{-\int_0^L \lambda(u) \mathrm{d}u\right\}.$$

Exercice 1. Montrez le.
Astuce : Soit T la variable aléatoire représentant le temps entre un point w et le point suivant. Alors
$\Pr[T > t] = \Pr[N\{(w, w+t)\} = 0]$.

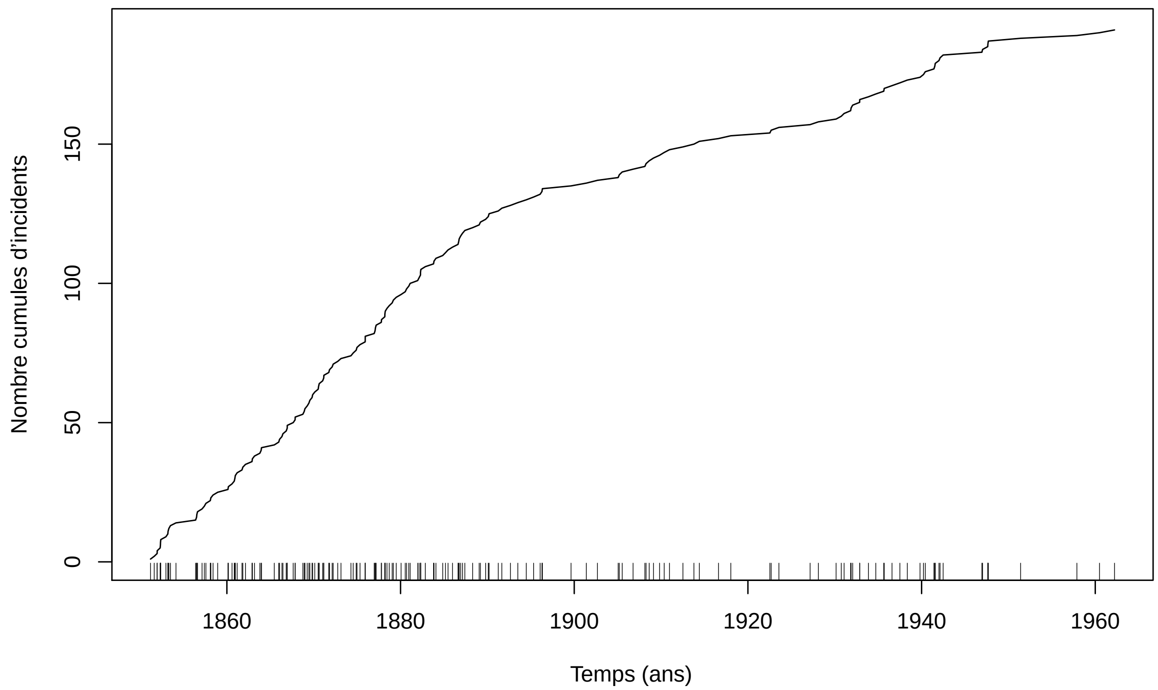

Figure 3: *Évolution du nombre d'incidents dans les mines de charbon au Royaume–Uni.*

□ Quelles remarques/constats pouvez vous faire ?

2. Le monde Bayésien

- Considérons un modèle statistique $f(y \mid \theta)$ où θ est inconnu.
- L'approche Bayésienne consiste a traiter θ comme une variable aléatoire. θ admet donc une loi !!!
- A partir d'une loi a priori $\pi(\theta)$ pour θ et le théorème de Bayes on a

$$\pi(\theta \mid y) = \frac{f(y \mid \theta)\pi(\theta)}{\int f(y \mid \theta)\pi(\theta)\mathrm{d}\theta} \propto f(y \mid \theta)\pi(\theta),$$

  où $\pi(\theta \mid y)$ est appelée la loi a posteriori
- Éventuellement on peut s'intéresser à la loi prédictive a posteriori d'une variable Z

$$f(z \mid y) = \int f(z \mid y, \theta)\pi(\theta \mid y)\mathrm{d}\theta.$$

Difficultés

- Le monde Bayésien est souvent critiqué du point de vu conceptuel car il place au même plan les données et la loi a priori
- D'un point de vue pratique la plus grande difficulté réside dans le calcul d'intégrales
- Bien qu'il existe plusieurs approches possible, ici nous allons voir les approches dîtes de Monte Carlo par Chaîne de Markov (MCMC)
- L'idée des approches MCMC consiste à :

  1. Construire une chaîne de Markov dont la loi stationnaire est justement la loi a posteriori $\pi(\theta \mid y)$;
  2. Générer une trajectoire $\{\theta_t\}_{t\geq 0}$ de cette chaîne jusqu'à convergence;
  3. Et traiter le résultat comme une réalisation de $\pi(\theta \mid y)$.

Exemple 1. Trouvez la loi jointe a posteriori de la moyenne et la variance d'un échantillon Gaussien de taille n et dont les lois a priori sont $\mu \sim N(\xi, \kappa^{-1})$ et $\sigma^{-2} \sim \Gamma(\alpha, \beta)$.

- □ S'il est possible de simuler selon les loi conditionnelles pleines $\pi(\theta_i \mid \cdots)$ où "$\cdots$" signifie "données et tous les paramètres sauf θ_i, alors on peut utiliser l'échantillonneur de Gibbs :

  1. $n \leftarrow 1; \theta_n \leftarrow \theta_0 = (\theta_{1,0}, \ldots, \theta_{p,0})$;
  2. $\theta_{n+1} \leftarrow \theta_n$;
  3. Pour $j = 1, \ldots, p$: $\theta_{j,n+1} \sim \pi(\theta_{j,n+1} \mid \cdots)$
  4. $n \leftarrow n+1$ et revenir à l'étape 2.

- □ La chaîne ainsi obtenue a pour loi stationnaire $\pi(\theta \mid y)$.

Echantillonneur de Gibbs en R

```
## Darwin's maize data in eighths of an inch
y <- c(49, -67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, -48)
n.obs <- length(y)

xi <- kappa <- alpha <- beta <- 0## Spécification des loi a priori (lois impropres)

n.iter <- 10000## Longueur de la chaîne simulée

## Echantillonneur de Gibbs initialisé à mu = 0, sigma2 = 500
chaine <- matrix(NA, n.iter, 2);chaine[1,] <- c(0, 500)

for (i in 2:n.iter){
  new.mean <- (sum(y) + kappa * chaine[i-1,2] * xi) /
    (n.obs + kappa * chaine[i-1,2])
  new.var <- chaine[i-1,2] / (n.obs + kappa * chaine[i-1,2])
  chaine[i,1] <- rnorm(1, new.mean, sqrt(new.var))
  chaine[i,2] <- 1 / rgamma(1, shape = beta + 0.5 * sum((y - chaine[i,1])^2),
                             rate = alpha + 0.5 * n.obs)
}

chaine[,2] <- sqrt(chaine[,2])## Passage de sigma2 à sigma (plus lisible)
```

Application : Le maïs de Darwin

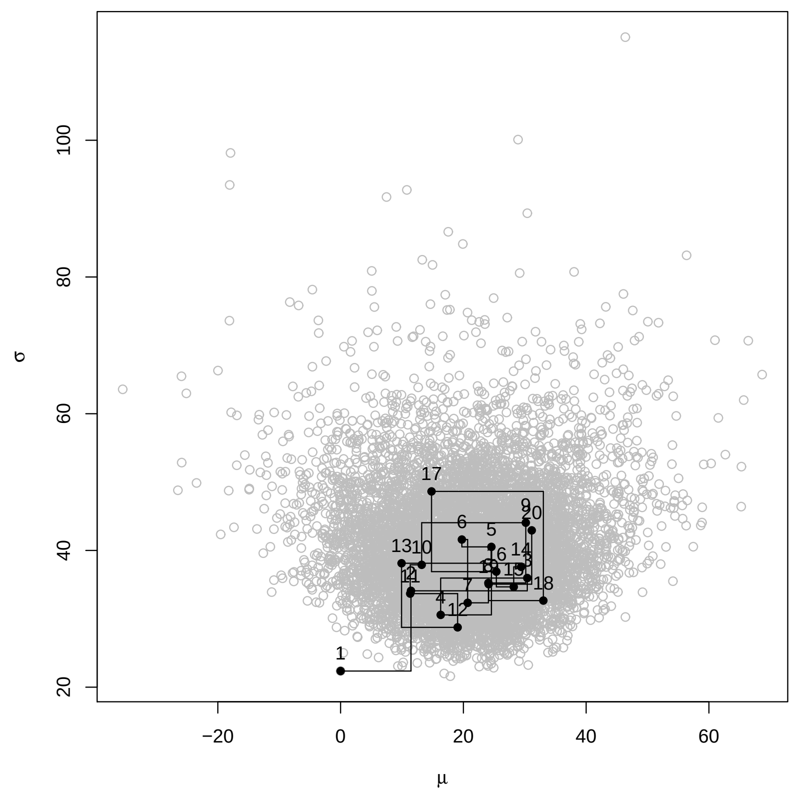

Figure 4: *Représentation d'une chaîne de Markov de longueur 10000 obtenue par l'échantillonneur de Gibbs. Les segments représentent les 20 premières itérations.*

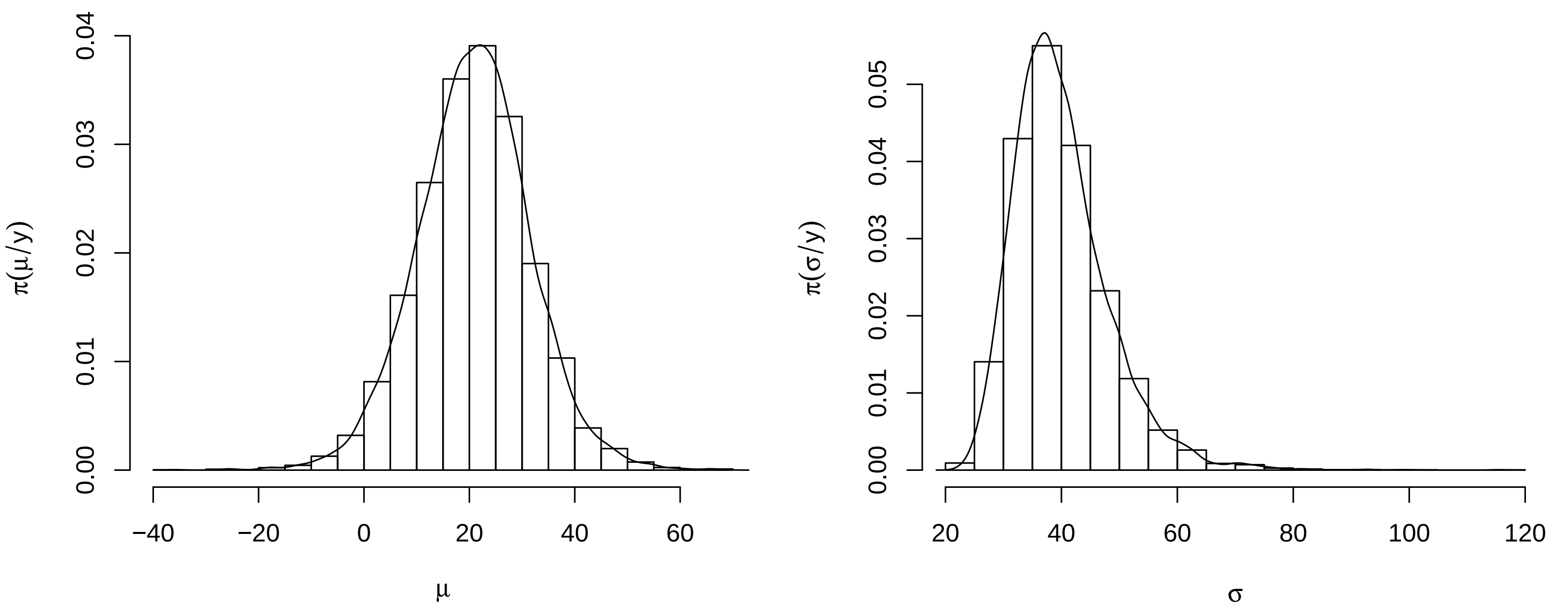

Figure 5: *Histogrammes (et densités) des loi a posteriori marginales pour μ et σ obtenue à partir de 10000 simulations.*

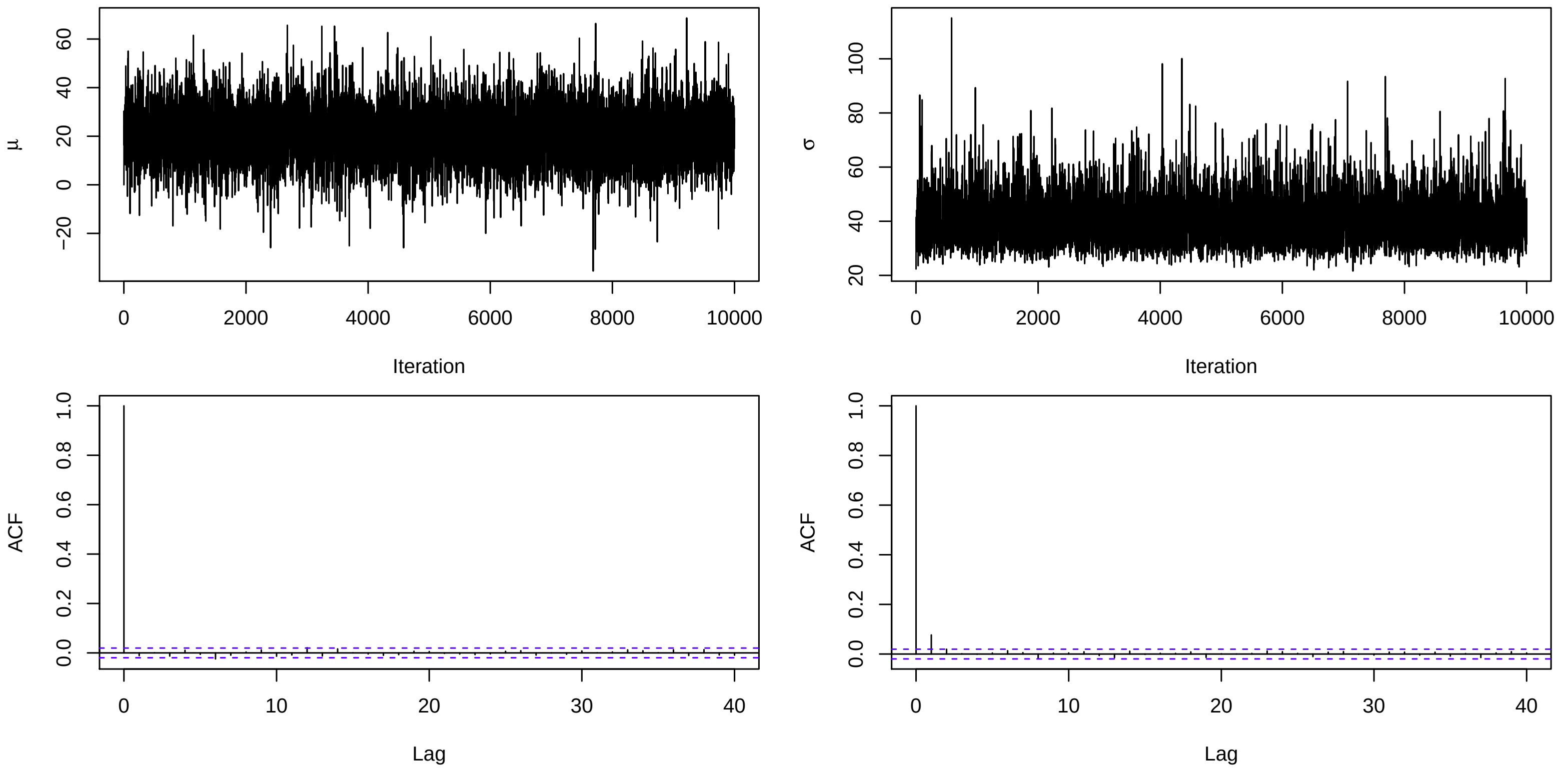

Figure 6: *Séries temporelles pour μ et σ et leur ACF associés. Les séries temporelles semblent être stationnaires et présentent une très faible autcorrélation.*

3. Algorithmes MCMC

Objectifs

- Construire une chaîne de Markov $\{u_t\}_{t\geq 0}$ à valeurs dans $\mathcal{U}$ et dont la loi stationnaire est la loi cible π, i.e.,

$$\Pr[u_t \in \mathcal{A} \mid u_0] \to \pi(\mathcal{A}), \quad t \to \infty, \quad \forall u_0 \in \mathcal{U}, \mathcal{A} \subset \mathcal{U}.$$

- $\{u_t\}_{t\geq 0}$ est dirigée par son noyau de transition $P(u,v)$, i.e., la densité que la chaîne fasse le déplacement $u \mapsto v$, et donc

$$\int P(u,v)\mathrm{dv} = 1, \qquad \forall u \in \mathcal{U}.$$

- Souvent on souhaitera la “detailed balance condition” (réversibilité temporelle)

$$\pi(u)P(u,v) = \pi(v)P(v,u), \qquad u,v \in \mathcal{U}.$$

- Notons u_{-j} toutes les composantes de u sauf u_j
- Nous avons vu que l'échantillonneur de Gibbs consistait à simuler successivement selon les lois conditionnelles pleines $\pi(u_j \mid u_{-j})$
- Le noyau de transition correspondant est

$$P(u, v) = \pi(v_j \mid u_{-j}) 1_{\{u_{-j}=v_{-j}\}}.$$

Exercice 2. Montrer que la "detailed balance condition" est vérifiée.

Metropolis–Hastings

- Un algorithme très simple :
  - Simuler un candidat v pour le prochain état de la chaîne selon une loi de proposition de densité $q(\cdot \mid u)$
  - Calculer la probabilité d'acceptation

$$\alpha(u,v) = \min\left\{1, \frac{\pi(v)q(u \mid v)}{\pi(u)q(v \mid u)}\right\}$$

  - Accepter le candidat comme nouvel état de la chaîne avec probabilité $\alpha(u,v)$, sinon garder la valeur courante.
- Il est suffisant de connaître π qu'à une constante multiplicative près !
- Gibbs est un cas particulier avec $q(u,v) = \pi(v_j \mid u_{-j}) = \pi(v_j \mid v_{-j})$ et donc $\alpha(u,v) = 1$.

Exercice 3. Construire un algorithme de Metropolis–Hastings dont la loi cible est une $N(0,1)$ et la loi de proposition a pour densité

$$q(v \mid u) = \frac{1}{\sigma}\varphi\left(\frac{v-u}{\sigma}\right).$$

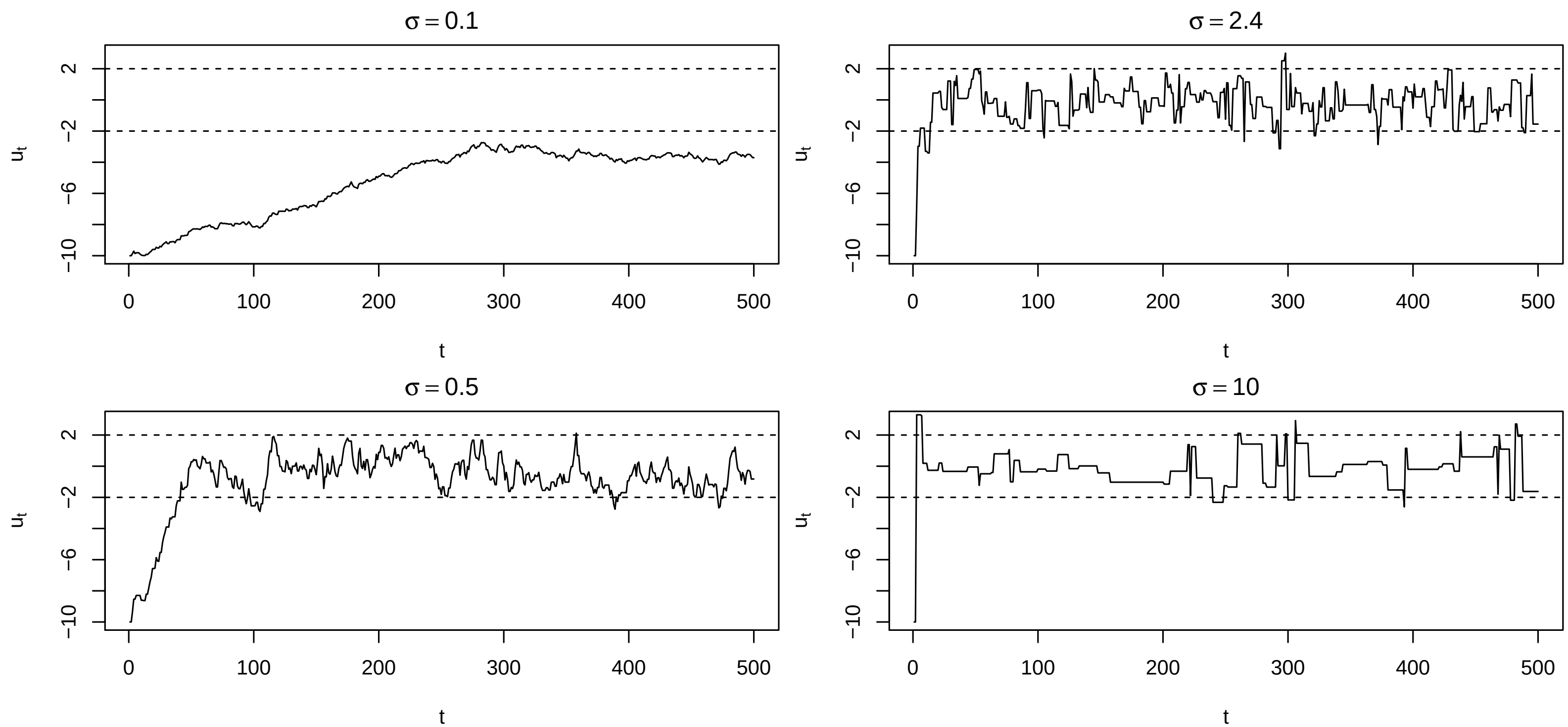

Figure 7: *4 chaines simulées selon Metropolis–Hastings dont la loi cible est une $N(0,1)$, $u_0 = -10$ et $v \sim N(u, \sigma^2)$ avec $\sigma = 0.1, 0.5, 2.4, 10$. Lorsque $\sigma = 0.1, 0.5$ les candidats sont souvent acceptés mais la chaîne évolue lentement. Pour $\sigma = 10$, la chaîne se déplace rarement mais par souvent par de "grands sauts". Le cas $\sigma = 2.4$ semble être un bon compromis. On dit alors que la chaîne mixe bien.*

4. Ce que nous ne verrons pas

Pour les personnes intéressées

Nous avons juste survolé l'inférence Bayésienne. Pour les plus curieux d'entre vous je vous conseille de jeter un oeil à

- Échantillonneur de Gibbs avec mises à jour selon Metropolis–Hastings
- Diagnostic de convergence
- Chaînes de Markov à sauts réversibles
- Modèles hiérarchiques
- Modèles graphiques

5. Les mains dans le cambouis

Cas d'étude

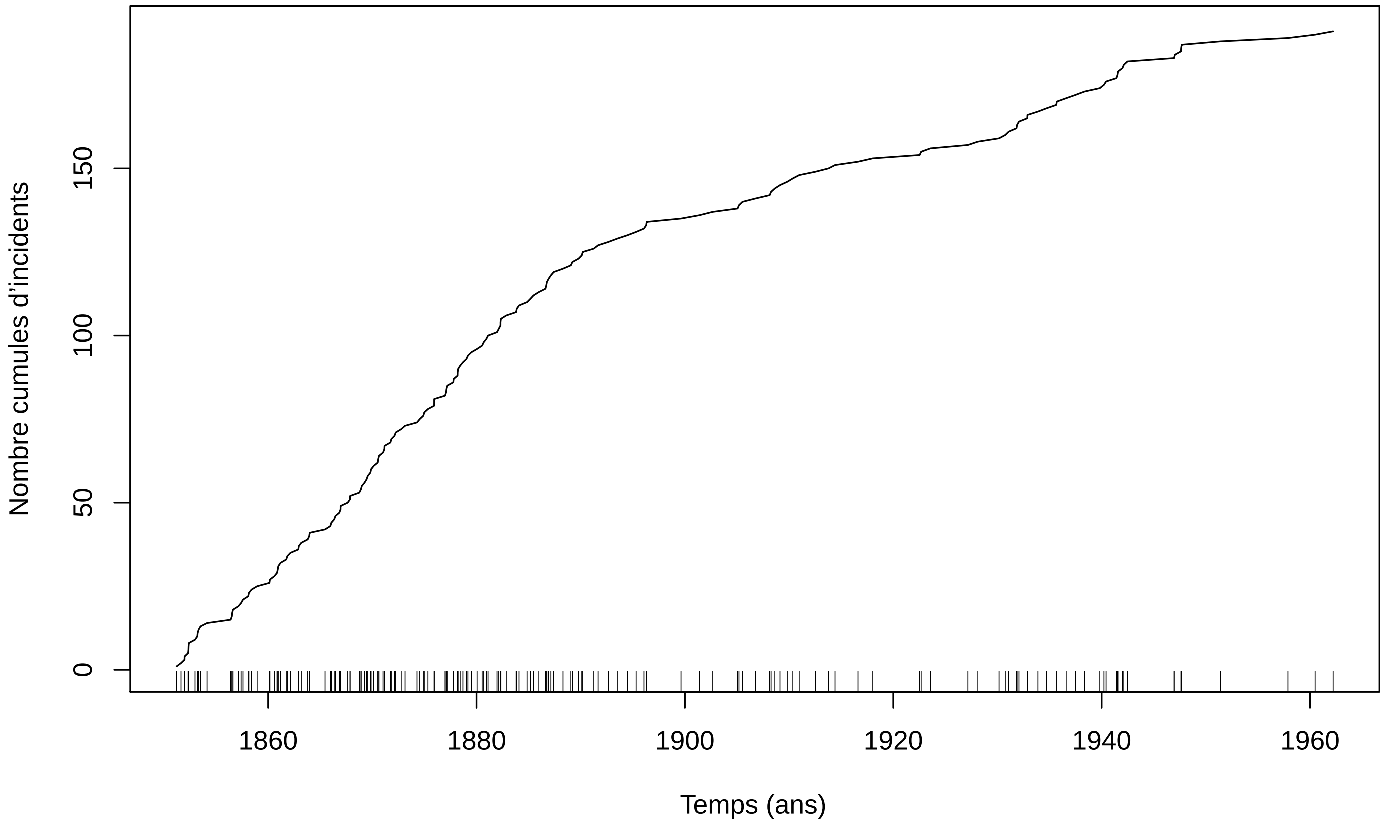

Figure 8: *Évolution du nombre d'incidents dans les mines de charbon au Royaume–Uni.*

- Quel(s) modèle(s) Bayésien(s) proposeriez vous ?

Modèle proposé

- Les données semblent montrer un changement de pente autour de 1900. Devons–nous en tenir compte ?
- Supposons donc que l'intensité de processus de Poisson est

$$\lambda \equiv \lambda(t) = \begin{cases} \lambda_0, & 1851 \leq t \leq 1892 \\ \lambda_1, & 1892 < t < 1962, \end{cases}$$

  et le nombre d'incident avant 1892 est $n_0 = 127$ et après 1900 est $n_1 = 64$.
- On prendra comme lois a priori

$$\lambda_1, \lambda_2 \overset{\text{iid}}{\sim} \mathsf{Gamma}(\alpha, \beta), \qquad \alpha = 1,$$

  et l'on supposera que β est distribuée également selon une loi a priori Gamma$(1, f)$ où $f = (n_1 + n_2)/(1962 - 1851) = 1.72$ (triche !).

Ce que j'aimerai que vous fassiez

- ☐ Calculez les loi conditionnelles pleines pour ce modèle
- ☐ Implémentez un échantillonneur de Gibbs
- ☐ Faire tourner votre code et tirez quelques conclusions à partir de la loi a posteriori
- ☐ Quelles améliorations aimeriez vous faire ?