Modélisation Charges sinistres

Chapitre 4 : Inteférence d'une distribution composée

I - Problème

Soit $X = \sum_{i=1}^{N} U_i$

. N est une va de comptage de loi $P_N(\cdot, \theta_N)$

. $(U_i)_{i\geq 1}$ forme une suite iid de va positives indépendante de N et de loi $f_U(\cdot ; \theta_U)$

Soient $x_1, \dots, x_t$ un échantillon d'observations iid de X sur t périodes d'exercice

Comment inférer la valeur des paramètres $\theta = (\theta_N, \theta_U)$ à partir de ces observations

Il s'agit d'un problème de données incomplète. Les données complètes comprennent la fréquence et les montants de sinistres soit

$$(n_s, \vec{u}_s) = \{ n_s, (u_{s,1}, \dots, u_{s,n_s}) \}, \; s = 1, \dots, t$$

Supposons que $N \sim Geom(q)$ avec $\mathbb{P}(N=k) = (1-q)q^k$, $k \geq 0$

et $U \sim Exp(\delta)$ avec $f_U(x) = \dfrac{e^{-x/\delta}}{\delta}$

Méthode par maximum de vraisemblance, méthode des moments.

II – Approche fréquentiste

1/ Maximum de vraisemblance

Fonction de vraisemblance avec $x = (x_1, \dots, x_t)$

$$L(x, q, \delta) = (1-q)^{t_0} \prod_{s=1}^{t-t_0} f_X^+(x_s^+)$$

où t_0 est le nombre de zéros dans mes données et $x_1^+, \dots, x_{t-t_0}^+$

$$f_X^+(x) = ?$$

$$= \sum_{n=1}^{+\infty} \mathbb{P}(N=n) f_U^{*n}(x)$$

$$= \sum_{n=1}^{\infty} (1-q)q^n \frac{e^{-x/\delta} x^{n-1}}{(n-1)!\,\delta}$$

série exp $\sum \frac{x^n}{n!} \dots$

$$= \frac{(1-q)q}{\delta} e^{-x\frac{(1-q)}{\delta}}$$

$$L(x,q,\delta) = (1-q)^{t_0}\left[\frac{(1-q)q}{\delta}\right]^{t-t_0} \exp\left[-\frac{1-q}{\delta}\sum_{s=1}^{t-t_0} x_s^{\,t}\right]$$

$$= (1-q)^t \left(\frac{q}{\delta}\right)^{t-t_0} \exp\left[-\frac{(1-q)}{\delta}\sum_{s=1}^{t-t_0} x_s^{\,t}\right]$$

On cherche à maximiser

$$\ell(x,q,\delta) = \log(L(x,q,\delta))$$

$$= t\log(1-q) + (t-t_0)\log q + (t-t_0)\log\delta - \frac{(1-q)}{\delta}\sum_{s=1}^{t-t_0} x_s^{\,t}$$

On résout $\begin{cases} \dfrac{\partial \ell}{\partial q} = 0 \\ \dfrac{\partial \ell}{\partial \delta} = 0 \end{cases}$ pour trouver $(\hat{q},\hat{\delta}) = \operatorname{argmax}(x,q,\delta)$

On a

$$\frac{d\ell}{dq} = -\frac{t}{1-q} + \frac{t-t_0}{q} + \frac{\sum_{s=1}^{t-t_0} x_s^{\,t}}{\delta} = 0 \quad (1)$$

$$\frac{d\ell}{d\delta} = \frac{t-t_0}{\delta} + (1-q)\,\delta^{-2}\sum_{s=1}^{t-t_0} x_s^{\,t} = 0 \quad (2)$$

Cela donne $\hat{q} = \frac{t - t_0}{t}$ et $\hat{\delta} = \frac{t_0}{t} \frac{1}{t - t_0} \sum_{s=1}^{t-t_0} z_s^+$

2) Approche Bayesienne

(Inference Bayesienne repose sur la loi a posteriori)

$$p(\theta|x) = \frac{L(x,\theta)\,p(\theta)}{\int L(x,\theta)\,p(\theta)\,d\theta} = \frac{L(x,\theta)\,p(\theta)}{L(x)}$$

de θ sachant les données x. On obtient $p(\theta|x)$ en mettant à jour la loi a priori $p(\theta)$ via la vraisemblance $L(x,\theta)$.

La constante de normalisation $L(x)$ est souvent difficile à calculer. On utilise alors des algo de simulation MCMC pour échantillonner la loi a posteriori.

Metropolis - Hasting

Pour $i = 1, \dots I$

① Initialisation : Si $i = 1$ alors $\theta_1 \sim p(\theta)$ sinon étape 2

② Perturbation : si $i > 1$ $\theta^* = \theta_{i-1} + \varepsilon_i$ avec $\varepsilon \sim \mathcal{N}$-Normal$(0, \Sigma)$

③ Acceptation Rejet, soit $U \sim \text{Unif}([0,1])$

- Si $\min\left(1, \dfrac{L(x,\theta^*)p(\theta^*)}{L(x,\theta_{i-1})\,p(\theta_{i-1})}\right) > u$ alors $\theta_i = \theta^*$

- Sinon $\theta_i = \theta_{i-1}$

Le résultat est une suite $\theta_1, \dots, \theta_I$ distribué suivant $p(\theta, x)$

Prenons des lois a priori uniforme et indépendantes sur les paramètres avec

$$p(\theta) = p(q) \cdot p(\delta) = \frac{1}{b_q - a_q} \mathbb{1}_{[a_q, b_q]} \cdot \frac{1}{b_\delta - a_\delta} \mathbb{1}_{[a_\delta, b_\delta]}$$