Homework

M SEKOUA

November 2019

Exercice 1

Soit Y_i un echantillon aléatoire simplke d'une loi exponentiele avec une fonction de densité de probabilité :

$$f(y_i|\theta) = \frac{1}{\theta}e^{-\frac{y_i}{\theta}},\ y_i, \theta \in \mathbb{R}_+$$

Trouvez la distribution a posteriori de δ

Soit $\delta = \theta^{-1}$ et soit un a priori sur δ spécifié comme une distribution Gamma avec les hyperparamètres α_0 et β_0 où $\pi(\delta) \propto \delta^{\alpha_0 - 1}e^{-\frac{\delta}{\beta_0}}$.
Par definition de l aposteriori,on a $\pi(\delta|y) \propto f(y|\delta)\pi(\delta)$.
Donc en injectant la densité dans notre approximation, nous obtenons :

$$\begin{aligned}\pi(\delta|y) &\propto \delta e^{-\delta y} \times \delta^{\alpha_0 - 1}e^{-\frac{\delta}{\beta_0}} \\ &\propto \delta^{\alpha_0}e^{-\delta(y+\frac{1}{\beta_0})}\end{aligned}$$

Donc $\delta|y \rightsquigarrow \mathcal{G}(\alpha_0 + 1, \frac{1}{y+\frac{1}{\beta_0}})$

Pour un echantillon de taille n, alors on a alors

$$\delta|y \rightsquigarrow \mathcal{G}(\alpha_0 + n, \frac{1}{(\sum_{i=1}^{n} y_i) + \frac{1}{\beta_0}})$$

Trouvez la distribution a posteriori de θ

On sait que $\delta \rightsquigarrow \mathcal{G}(\alpha_0, \beta_0)$ donc $\theta \rightsquigarrow \mathcal{IG}(\alpha_0, \frac{1}{\beta_0})$.
Pour une facilité de lecture et de manipulation, nous posons $\delta_0 = \frac{1}{\beta_0}$

$$\begin{aligned}\pi(\theta|y) &\propto f(y|\theta)\pi(\theta) \\ &\propto (\frac{1}{\theta})^{\alpha_0+2}e^{-\frac{y+\delta_0}{\theta}}\end{aligned}$$

Donc $\theta|y \rightsquigarrow \mathcal{IG}(\alpha_0 + 1, y + \delta_0)$

Pour un echantillon de taille n, alors on a alors

$$\theta|y \rightsquigarrow \mathcal{IG}(\alpha_0 + n, \sum_{i=1}^{n} y_i + \delta_0)$$

Derivons l'estimateur Bayesien $\tilde{\theta}$ de θ

Pour une fonction de coût quadratique,l'estimateur bayésien $\tilde{\theta}(y)$ est défini comme suit :

$$\tilde{\theta}(y) = \arg\min_{d} \rho(\pi, d|y)$$

où $\rho(\pi, d|y) = \int_\Theta (\theta - d)^2 \pi(\theta|y) d\theta$.
Comme l a posteriori de θ est une inverse gamma, alors les moments d'ordre 1, 2 sont definis. Egalement la marginale de $m(x)$ est bien definie et intégrable des lorsque l'on a $\alpha_0 + n > 2$.
La fonction de coût est convexe et donc admet un minimum. Donc notre fonction $\rho(\pi, d|y)$ admet un minimum.
L'estimateur bayésien pour une a priori donné, verifie la condition suivante

$$\frac{\partial \rho}{\partial \delta}(\pi, \delta, y)\Big|_{\delta = \tilde{\theta}} = 0$$

D'où

$$\begin{aligned}\tilde{\theta}(y) &= \int_\Theta \theta \pi(\theta|y) d\theta \\ &= \frac{\delta_0 + \sum_{i=1}^{n} y_i}{\alpha_0 + n - 1}\end{aligned}$$

L'estimateur bayésien est celui ayant le cout a posteriori minimum.

Trouvez l'erreur quadratique moyen de l'estimateur Bayésien $\tilde{\theta}$

Par definition l'erreur quadratique moyen est definie comme suit :

$$EQM(\delta) = \int_{\mathcal{X}} (\delta - \theta)^2 f(y_1, ..., y_n|\theta) dy_1 ....dy_n$$

Remarque: On observe que $\mathbb{E}(\sum_{i=1}^{n} y_i|\theta) = n\theta$
Pour l'estimateur bayésien $\tilde{\theta}$,

$$\theta - \frac{\sum_{i=1}^{n} y_i + \delta_0}{\alpha_0 + n - 1} = \frac{n\theta - \sum_{i=1}^{n} y_i}{\alpha_0 + n - 1} + \frac{\alpha_0 - 1}{\alpha_0 + n - 1}\theta - \frac{\delta_0}{\alpha_0 + n - 1}$$

Donc

$$EQM(\tilde{\theta}) = \frac{n\theta^2}{(\alpha_0 + n - 1)^2} + \frac{(\alpha_0 - 1)^2\theta^2}{(\alpha_0 + n - 1)^2} + \frac{\delta_0^2}{(\alpha_0 + n - 1)^2} - 2\frac{(\alpha_0 - 1)\delta_0\theta}{(\alpha_0 + n - 1)^2}$$

Pour l'estimateur du maximum de vraisemblance $\bar{Y}$,

$$\bar{Y} - \theta = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)$$

Comme les y_i sont independantes et identiquement distribuées,alors

$$EQM(\bar{Y}) = \frac{\theta^2}{n}$$

Pour l'estimateur $\hat{\theta}$,

$$\hat{\theta} - \theta = \frac{1}{n+1}\sum_{i=1}^{n}(y_i - \theta) - \frac{\theta}{n+1}$$

Donc

$$EQM(\hat{\theta}) = \frac{\theta^2}{(n+1)}$$

Choix de l'estimateur apres calcul de risque quadratique moyen

Lorsque $\alpha_0 = 1$, on preferait l'estimateur $\hat{\theta}$ à partir de $n \geq 2$

Calcul du risque a posteriori des estimateurs $\bar{Y}$ et $\hat{\theta}$

On definit le risque a posteriori comme suit :

$$\mathbb{E}^{\pi}(\delta - \theta)^2 = \int_{\Theta}(\delta - \theta)^2\pi(\theta|y)d\theta$$

Pour $\alpha_0 = 1$,

$$\theta|y \rightsquigarrow \mathcal{IG}(n+1, \sum_{i=1}^{n} y_i + \delta_0)$$

Pour l'estimateur $\tilde{\theta}$, on a

$$\mathbb{E}^\pi(\tilde{\theta} - \theta)^2 = \frac{(\sum_{i=1}^n y_i + \delta_0)^2}{n^2(n-1)^2}$$

Pour l'estimateur du maximum de vraisemblance $\bar{Y}$, on a donc :

$$\mathbb{E}^\pi(\bar{Y} - \theta)^2 = \frac{(\sum_{i=1}^n y_i + \delta_0)^2}{n^2(n-1)^2} + \frac{\delta^2}{n^2}$$

Pour l'estimateur $\hat{\theta}$,

$$\mathbb{E}^\pi(\hat{\theta} - \theta)^2 = \frac{(\sum_{i=1}^n y_i + \delta_0)^2}{n^2(n-1)^2} + (\frac{1}{n+1}\bar{Y} + \frac{\delta}{n})^2$$

Le meilleur risque au sens du risque bayesien est l'estimateur bayesien $\tilde{\theta}$

Exercice 2

Soit $y_1, ..., y_T$ une série chronologique de reponses de comptage indépendentes, générées de la façcon suivante :

$$y_t|\gamma, \delta, \lambda \sim \begin{cases} \mathbb{P}_o(\gamma) & \text{si } t \leq \lambda \\ \mathbb{P}_o(\delta) & \text{si } t > \lambda \end{cases}$$

On suppose de plus que

$$\begin{aligned} \gamma &\rightsquigarrow \mathcal{G}(a_1, a_2) \\ \delta &\rightsquigarrow \mathcal{G}(d_1, d_2) \\ \lambda &\rightsquigarrow \mathcal{U}_{1,..T} \end{aligned}$$

Derivons la fonction de vraisemblance pour ce modèle

Definissons la vraisemblance

$$\begin{aligned} l(y|\gamma, \delta, \lambda) &= (\prod_{t=1}^{\lambda} \mathbb{P}_o(y_t|\gamma))(\prod_{t=\lambda+1}^{T} \mathbb{P}_o(y_t|\delta))1_{y_t \in \mathbb{N}} \\ &= (\prod_{t=1}^{\lambda} \gamma^{y_t} \frac{e^{-\gamma}}{y_t!})(\prod_{t=\lambda+1}^{T} \delta^{y_t} \frac{e^{-\delta}}{y_t!})1_{\min_t y_t \in \mathbb{N}} \end{aligned}$$

On a les paramètres γ, δ car λ correspond plutôt à un paramètre de changement de changement. Nous derivons donc la log-vraisemblance par rapport γ et δ.

La log-vraisemblance vaut donc :

$$\mathcal{L}(y|\gamma, \delta, \lambda) = -\sum_{t=1}^{T} \log(y_t!) + \sum_{t=1}^{\lambda}(y_t \log(\gamma) - \lambda\gamma) + \sum_{t=\lambda+1}^{T}(y_t \log(\delta) - \delta(T - \lambda))$$

$$\begin{aligned}
\nabla_{\gamma,\delta}\mathcal{L}(y_t|\gamma,\delta,) = 0 \\
\hat{\gamma} &= \frac{\sum_{t=1} y_t}{\lambda} \\
\hat{\delta} &= \frac{\sum_{t=\lambda+1}^{T} y_t}{T-\lambda}
\end{aligned}$$

Decrire l'algorithme de Gibbs

Pour definir l'algorithme de gibbs, nous avons besoin de definir l'a posteriori $\pi(\gamma, \delta, \lambda|y_t)$ qui nous permettra d'obtenir de determiner les lois des parametres γ, δ et λ, lorsque l'on calcule l a posteriori

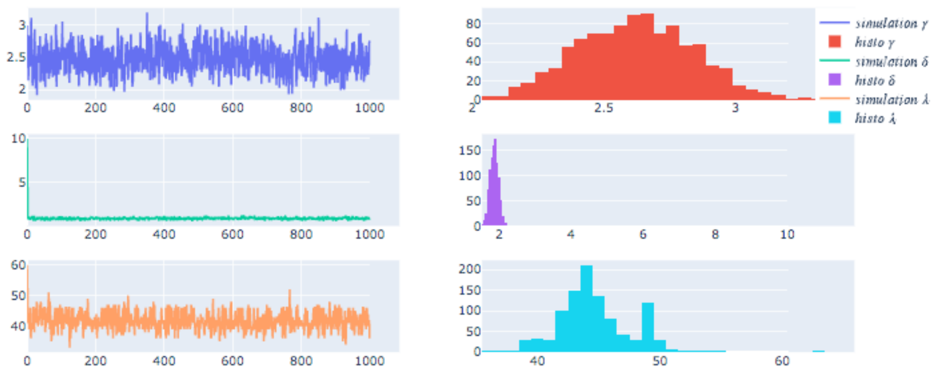
$$\begin{aligned}
\pi(\gamma,\delta,\lambda|y_t) &\propto f(y_t|\gamma,\delta,\lambda)\pi(\gamma,\delta,\lambda) \\
&\propto (\prod_{t=1}^{\lambda} \gamma^{y_t}\frac{e^{-\gamma}}{y_t!})(\prod_{t=\lambda+1}^{T} \delta^{y_t}\frac{e^{-\delta}}{y_t!})\gamma^{a_1-1}e^{-a_2\gamma}\delta^{d_1-1}e^{-d_2\delta}\frac{1}{T} \\
&\propto \gamma^{\sum_{t=1}^{\lambda} y_t+a_1-1}e^{-\gamma(\lambda+a_2)}\delta^{\sum_{t=\lambda+1}^{T} y_t+d_1-1}e^{-\delta(T-\lambda+d_2)}
\end{aligned}$$

Donc pour implementer l'algorithme de Gibbs, on obtient donc

$$\begin{aligned}
\gamma &\rightsquigarrow \mathcal{G}(\sum_{t=1}^{\lambda} y_t + a_1, \lambda + a_2) \\
\delta &\rightsquigarrow \mathcal{G}(\sum_{t=\lambda+1}^{T} y_t + d_1, T - \lambda + d_2) \\
\lambda &\rightsquigarrow Categorical(\frac{\gamma^{\sum_{t=1}^{\lambda} y_t+a_1-1}e^{-\gamma(\lambda+a_2)}\delta^{\sum_{t=\lambda+1}^{T} y_t+d_1-1}e^{-\delta(T-\lambda+d_2)}}{\sum_{\lambda=1}^{T}\gamma^{\sum_{t=1}^{\lambda} y_t+a_1-1}e^{-\gamma(\lambda+a_2)}\delta^{\sum_{t=\lambda+1}^{T} y_t+d_1-1}e^{-\delta(T-\lambda+d_2)}})
\end{aligned}$$

Detection du point de changement

Nous avons simulé les variables du changement de point, des parametres gamma et delta.



La moyenne de la variable λ est 41.65 et l ecart type de la variable λ est 2.84. Donc les changements sont autour de la valeur 42.
Le nombre prevu d'accident avant 2.49 et le nombre d'accident apres est de 0.88.