

Inférence bayésienne pour l'estimation des paramètres des erreurs de reconstruction à l'aide d'une distribution gamma multivariée

Ibrahim TRAORE^{1,2} Imen MEGDICHE² Jérôme MARQUET-DOLÉAC³ Lotfi CHAARI¹

¹Toulouse INP, IRIT, Université de Toulouse, France

²IRIT, ISIS Castres, INUC, Université de Toulouse, France

³Institut de Formation en Psychomotricité, Toulouse NeuroImaging Center, UMR 1214 - INSERM, Université de Toulouse, France

Résumé – L'utilisation des modèles génératifs pour la détection d'anomalies repose sur l'analyse de l'erreur de reconstruction, définie par le carré de la différence entre les données observées et générées. Cependant, un défi critique persiste : comment modéliser efficacement la distribution des erreurs de reconstruction, dont les paramètres restent généralement inconnus. Les approches classiques supposant une distribution normale multivariée sont inadaptées, car les erreurs sont positives et asymétriques. Cet article propose une nouvelle modélisation des erreurs de reconstruction pour pallier cette limitation. Nous introduisons une distribution adaptée aux caractéristiques de ces erreurs, en l'occurrence une distribution gamma multivariée, mieux à même de capturer leur asymétrie et leur positivité intrinsèques. Pour estimer ses paramètres, un modèle bayésien hiérarchique est proposé, combinant un échantillonnage de Gibbs et une méthode d'exécution parallèle afin d'optimiser les performances computationnelles. Cette approche permet de traiter de grands volumes de données tout en garantissant une estimation robuste et un temps de convergence raisonnable.

Abstract – The use of generative models for anomaly detection relies on analyzing the reconstruction error, defined as the squared difference between observed and generated data. However, a critical challenge remains: how to effectively model the distribution of reconstruction errors, whose parameters are generally unknown. Traditional approaches assuming a multivariate normal distribution are inadequate, as the errors are positive and asymmetric. This paper proposes a novel modeling of reconstruction errors to address this limitation. We introduce a distribution that better fits the characteristics of these errors, namely a multivariate gamma distribution, which is more capable of capturing their asymmetry and intrinsic positivity. To estimate its parameters, we propose a hierarchical Bayesian model, combining Gibbs sampling with a parallel execution method to optimize computational performance. This approach enables the handling of large volumes of data while ensuring robust estimation and a reasonable convergence time.

1 Introduction

Une anomalie est une donnée qui présente un comportement inhabituel à un moment donné, comparativement aux autres valeurs de la série temporelle [1]. Les anomalies peuvent être univariées ou multivariées, selon qu'elles affectent une ou plusieurs variables dépendantes du temps. Dans le cas multivarié, il n'est pas nécessaire que toutes les variables soient affectées [1]. Ainsi, une approche multivariée des erreurs est plus appropriée car elle permet une détection plus fine. La détection d'anomalies dans les séries temporelles multivariées repose souvent sur des modèles génératifs qui reconstruisent les données d'entrée afin d'identifier les écarts anormaux. Ces modèles exploitent un principe clé : l'erreur quadratique (ou absolue) de reconstruction, définie comme le carré de la différence entre les données observées et leur version générée, est minimale pour les comportements normaux et élevée pour les anomalies. Cette approche, efficace pour capturer les dépendances temporelles complexes [1], génère en sortie un vecteur d'erreurs dont la distribution statistique détermine le seuil de détection. Cependant, un défi critique persiste : comment modéliser efficacement la distribution des erreurs de reconstruction, dont les paramètres restent généralement inconnus. Une estimation imprécise de ces paramètres entraîne des seuils de détection non fiables, augmentant les risques de faux positifs ou négatifs. Certains travaux, comme ceux de [2, 3], supposent une distribution normale multivariée pour

ces erreurs quadratiques. Or, cette hypothèse est inadaptée : ces erreurs quadratiques sont toujours positives avec une distribution rarement symétrique ou centrée. Nous proposons un cadre statistique adapté pour modéliser ces erreurs. Nous introduisons une distribution gamma multivariée, adaptée aux valeurs positives et à l'hétérogénéité des séries temporelles, et développons une méthode bayésienne avec un échantillonnage stochastique pour estimer ses paramètres. Les contributions de cet article sont : (1) une nouvelle distribution pour modéliser les erreurs de reconstruction dans les réseaux de neurones appliqués aux séries temporelles multivariées ; (2) un modèle bayésien avec échantillonnage de Gibbs pour estimer les paramètres d'une distribution gamma multivariée ; et (3) une méthode d'exécution parallèle pour accélérer l'échantillonnage et traiter efficacement de grands volumes de données. Le reste de l'article est structuré de la manière suivante : le paragraphe 2 présente les travaux connexes, le paragraphe 3 décrit notre modèle ainsi que la méthode d'estimation, et le paragraphe 4 est consacrée aux expérimentations réalisées pour valider notre approche.

2 Travaux connexes

Malhotra et al. [2, 3] ont supposé, dans leur approche de détection d'anomalies dans les séries temporelles, que les erreurs de reconstruction suivent une distribution normale multivariée, estimant les paramètres par la méthode du maximum de vraisem-

blance (MLE). Les erreurs de reconstruction sont quantifiées à l'aide de mesures couramment utilisées telles que l'erreur absolue moyenne (MAE) et l'erreur quadratique moyenne (MSE). Soit $X = \{x_t\}_{t \in T}$, une série temporelle multivariée, définie comme un ensemble ordonné de vecteurs k -dimensionnels, chacun étant enregistré à un instant spécifique $t \in T$, et composé de k observations réelles : $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ où x_t est un point. L'erreur de reconstruction \mathcal{L} est définie par :

$$\begin{aligned}\mathcal{L}_{MAE}(x_t, x'_t) &= \|x_t - x'_t\|_1 \\ \mathcal{L}_{MSE}(x_t, x'_t) &= \|x_t - x'_t\|_2^2,\end{aligned}\quad (1)$$

où x'_t représente la version reconstruite de x_t .

Ces métriques produisent des valeurs positives, et leur distribution est généralement asymétrique à droite, comme l'illustre la figure 1 où le MSE est illustré pour les trois dimensions spatiales (X , Y et Z) sur un exemple de données.

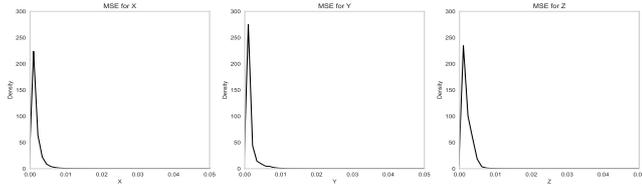


FIGURE 1 : Distribution de l'erreur de reconstruction.

L'hypothèse de normalité n'est donc pas pertinente. Notre objectif est de proposer une distribution adaptée aux erreurs de reconstruction des réseaux de neurones pour les séries temporelles multivariées. Les distributions gamma et exponentielle, flexibles et adaptées aux données asymétriques et positives, semblent pertinentes pour modéliser ces erreurs. La distribution exponentielle étant un cas particulier de la distribution gamma, une distribution gamma multivariée apparaît comme le choix le plus approprié pour représenter ces erreurs.

L'estimation des paramètres d'une distribution gamma multivariée présente des défis méthodologiques, car la méthode du maximum de vraisemblance ne fournit pas d'expressions analytiques simples. L'approche bayésienne apparaît donc comme une alternative prometteuse, explorée dans certains travaux pour des variantes de distributions gamma multivariées, comme un mélange de distributions gamma [4] ou un modèle gamma matriciel [5]. Les modèles de mélange de distributions gamma supposent que les données proviennent d'une combinaison de plusieurs lois gamma distinctes et les modèles gamma matriciels sont conçus pour capturer la structure de dépendance entre des variables organisées sous forme de matrices. Cependant, ces variantes ne conviennent pas aux erreurs de reconstruction car ces erreurs ne suivent ni une structure de mélange ni une organisation matricielle. Nous avons donc adopté la distribution proposée dans [6], plus adaptée à notre contexte. L'inférence bayésienne donne souvent lieu à des lois *a posteriori* de forme compliquée. Le recours à des techniques d'échantillonnage stochastique type MCMC (*Monte Carlo par Chaînes de Markov*) est souvent nécessaire (échantillonneur de Gibbs ou de Metropolis-Hastings).

Cependant, ces méthodes sont coûteuses en calculs, notamment en raison de l'étape d'acceptation/rejet (lorsqu'elle est nécessaire) qui nécessite d'évaluer le rapport de vraisemblance sur l'ensemble des données à chaque itération. Cette complexité computationnelle limite leur applicabilité aux grands jeux de données, car elles impliquent un balayage complet des

observations à chaque mise à jour des paramètres. Pour accélérer l'échantillonnage MCMC sur de grandes bases de données, certaines approches utilisent des minibatches [7]. Cependant, ces méthodes se révèlent souvent inefficaces, car elles peuvent nécessiter de traiter l'intégralité des données en une seule itération et peuvent introduire un biais [7]. Nous proposons une approche d'exécution parallèle qui exploite le multithreading pour améliorer les performances computationnelles. Plus de détails sont fournis dans la section suivante.

3 Le modèle d'estimation

Soit $\mathbf{y} = (y_1, y_2, \dots, y_K) \in \mathbb{R}_+^K$ un vecteur d'erreurs de reconstruction de dimension K . Nous supposons que \mathbf{y} est une réalisation de la variable aléatoire Y et que les N mesures d'erreur $(\mathbf{y}^n)_{1 \leq n \leq N}$ sont i.i.d. (indépendantes et identiquement distribuées). En notant $y_0 = 0$, la densité de probabilité de la n ème observation de \mathbf{y} peut être formulée ainsi [6] :

$$f(y_1^n, \dots, y_K^n) = \frac{\prod_{k=1}^K (y_k^n - y_{k-1}^n - \lambda_k)^{(\alpha_k - 1)} \exp\left(-\frac{y_K^n - \sum_{k=1}^K \lambda_k}{\gamma}\right)}{\gamma^{\bar{\alpha}} \prod_{k=1}^K \Gamma(\alpha_k)} \quad (2)$$

où $\alpha_k > 0$, $\gamma > 0$, $\lambda_k \in \mathbb{R}$, $\lambda_k < y_k^n$ et $\bar{\alpha} = \sum_{k=1}^K \alpha_k$.

La vraisemblance est par conséquent donnée par :

$$L(Y; \theta) = \prod_{n=1}^N f(y_1^n, \dots, y_K^n), \quad (3)$$

où $\theta = (\alpha, \gamma, \lambda)$, $\lambda = \{\lambda_1, \dots, \lambda_K\}$ et $\alpha = \{\alpha_1, \dots, \alpha_K\}$. La construction d'un modèle bayésien hiérarchique nécessite la définition de distributions *a priori* pour les autres paramètres du modèle. Nous supposons que les paramètres α_k suivent *a priori* des distributions exponentielles avec des paramètres p_k , que le paramètre γ suit une distribution *a priori* inverse gamma avec des paramètres w et v , tandis que les paramètres λ_k suivent des distributions *a priori* normales tronquées négativement avec des paramètres μ_k et σ_k :

$$\begin{aligned}\alpha_k &\sim \mathcal{G}(1, p_k) \\ \gamma &\sim \mathcal{IG}(w, v) \\ \lambda_k &\sim \mathcal{N}_{]-\infty, 0[}(\mu_k, \sigma_k^2).\end{aligned}\quad (4)$$

Ainsi, nous avons un modèle hiérarchique où la loi jointe *a posteriori* est donnée par :

$$\begin{aligned}f(\theta | Y) &\propto f(Y | \theta) \pi(\theta) \\ &\propto \prod_{n=1}^N \frac{\prod_{k=1}^K (y_k^n - y_{k-1}^n - \lambda_k)^{(\alpha_k - 1)} \exp\left(-\frac{y_K^n - \sum_{k=1}^K \lambda_k}{\gamma}\right)}{\gamma^{\bar{\alpha}} \prod_{k=1}^K \Gamma(\alpha_k)} \\ &\quad \times \left(\frac{1}{\gamma}\right)^{(w+1)} \exp\left(-\frac{v}{\gamma}\right) \prod_{k=1}^K \exp\left(-p_k \alpha_k - \frac{(\lambda_k - \mu_k)^2}{2\sigma_k^2}\right).\end{aligned}\quad (5)$$

Un estimateur du maximum a posteriori (MAP) peut être appliqué. Cependant, la forme de la loi jointe dans l'équation (5) ne permet pas d'avoir des estimateurs explicites. Comme alternative, une approximation numérique de cette loi reste possible en ayant recours à des techniques d'échantillonnage stochastique. Dans ce sens, plusieurs méthodes ont été proposées, comme l'échantillonnage par rejet, l'échantillonnage par importance et les méthodes de type MCMC. Dans cet

article, notre inférence bayésienne repose sur un échantillonneur de Gibbs. Cette technique MCMC génère des vecteurs de paramètres aléatoires convergeant asymptotiquement vers la distribution *a posteriori* sous des conditions vérifiables [4]. Les vecteurs sont produits à partir des distributions conditionnelles détaillées ci-dessous :

Loi conditionnelle pour le paramètre α_k :

$$f(\alpha_k | \alpha_{-k}, \gamma, \lambda, \mathbf{y}) \propto \Gamma(\alpha_k)^{-N} \exp(-\alpha_k S_k), \quad (6)$$

où $S_k = p_k + N \log(\gamma) - \sum_{n=1}^N \log(y_k^n - y_{k-1}^n - \lambda_k)$ et α_{-k} désigne le vecteur α sans la k ème composante.

Loi conditionnelle pour le paramètre γ :

$$f(\gamma | \alpha, \lambda, \mathbf{y}) \propto \frac{\exp\left(-\frac{v + \sum_{n=1}^N y_K^n - N \sum_{k=1}^K \lambda_k}{\gamma}\right)}{\gamma^{N\alpha + w + 1}} \quad (7)$$

$$\propto \text{InverseGamma}\left(w + N \sum_{k=1}^K \alpha_k, v + \sum_{n=1}^N y_K^n - N \sum_{k=1}^K \lambda_k\right).$$

Loi conditionnelle pour le paramètre λ_k :

$$f(\lambda_k | \alpha, \gamma, \lambda_{-k}, \mathbf{y}) \propto \quad (8)$$

$$\exp\left(\frac{N\lambda_k}{\gamma} - \frac{(\lambda_k - \mu_k)^2}{2\sigma_k^2}\right) \prod_{n=1}^N (y_k^n - y_{k-1}^n - \lambda_k)^{(\alpha_k - 1)}.$$

Réaliser des tirages aléatoires à partir de la loi conditionnelle du paramètre γ est simple. En revanche, générer des tirages à partir de la loi conditionnelle de α_k est moins évident. Nous avons utilisé une méthode par acceptation/rejet comme dans [4] avec une loi de proposition exponentielle. Quant à la distribution conditionnelle *a posteriori* de λ_k , elle ne ressemble pas à une forme connue. Une étape de proposition/acceptation devra être exécutée. Vu que $\lambda_k \in \mathbb{R}$ et $\lambda_k < y_k$, nous considérons une loi de proposition Gaussienne tronquée à valeurs négatives, $q \sim \mathcal{N}_{]-\infty, 0]}(\lambda_k, \sigma_k^2)$. Pour un candidat λ^* proposé, on définit la probabilité r avec laquelle λ^* sera accepté comme :

$$r = \min\left(1, \frac{f(\lambda^*)}{f(\lambda_k)} \times \frac{q(\lambda_k | \lambda^*)}{q(\lambda^* | \lambda_k)}\right).$$

À l'itération t , $\lambda_k^{(t)} = \lambda^*$ avec une probabilité r . En cas de rejet du candidat λ^* , le tirage est répété. Puisqu'on a

$$\frac{q(\lambda_k | \lambda^*)}{q(\lambda^* | \lambda_k)} = \frac{\varphi\left(\frac{\lambda_k - \lambda^*}{\sigma_k}\right)}{\Phi\left(\frac{b - \lambda_k}{\sigma_k}\right) - \Phi\left(\frac{a - \lambda_k}{\sigma_k}\right)} \times \frac{\Phi\left(\frac{b - \lambda^*}{\sigma_k}\right) - \Phi\left(\frac{a - \lambda^*}{\sigma_k}\right)}{\varphi\left(\frac{\lambda^* - \lambda_k}{\sigma_k}\right)}, \quad (9)$$

avec $a = -\infty$, $b = 0$, $\Phi(x) = \frac{1}{2}(1 + \text{erf}(x/\sqrt{x}))$ et $\varphi(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\xi^2)$, par simplification, on obtient :

$$r = \min\left(1, \frac{\pi(\lambda^*)}{\pi(\lambda_k)} \times \frac{\Phi\left(\frac{b - \lambda^*}{\sigma_k}\right) - \Phi\left(\frac{a - \lambda^*}{\sigma_k}\right)}{\Phi\left(\frac{b - \lambda_k}{\sigma_k}\right) - \Phi\left(\frac{a - \lambda_k}{\sigma_k}\right)}\right). \quad (10)$$

Dans l'équation (9), φ est symétrique et représente la fonction de densité de probabilité de la loi normale standard, alors que Φ représente sa fonction de répartition cumulative.

L'algorithme 1 décrit le processus d'échantillonnage de Gibbs parallélisé via multithreading. À chaque itération t , la variable $\gamma^{(t+1)}$ est d'abord générée séquentiellement, puis diffusée à tous les threads. Ces derniers échantillonnent en

parallèle les K couples (α_k, λ_k) , rendus indépendants grâce au fait que chaque α_k ne dépend que de γ et λ_k . Cette parallélisation conserve la validité statistique de l'échantillonneur de Gibbs tout en tirant parti des architectures multi-cœurs, ce qui la rend particulièrement adaptée aux simulations Monte Carlo à grande échelle.

Algorithme 1 : Échantillonnage de Gibbs parallélisé avec multithreading

Data : T (nombre d'itérations), K (nombre de composantes), $n_{threads}$ (nombre de threads)
Result : Échantillons $\{\alpha^{(t)}, \gamma^{(t)}, \lambda^{(t)}\}_{t=1}^T$
Initialiser $\alpha^{(0)}, \gamma^{(0)}, \lambda^{(0)}$;
for $t = 0$ **to** $T - 1$ **do**
 Étape 1 (Séquentielle) :
 $\gamma^{(t+1)} \leftarrow$ Échantillonner selon $f(\gamma | \alpha^{(t)}, \lambda^{(t)})$;
 Étape 2 (Parallèle) :
 Diffuser $\gamma^{(t+1)}$ à tous les threads;
 Répartir les K composantes en $n_{threads}$ partitions;
 for chaque thread $i = 1, 2, \dots, n_{threads}$ **do in parallel**
 foreach $k \in \text{partition}_i$ **do**
 $\alpha_k^{(t+1)} \leftarrow$ Échantillonner selon $f(\alpha_k | \gamma^{(t+1)}, \lambda_k^{(t)})$;
 $\lambda_k^{(t+1)} \leftarrow$ Échantillonner selon $f(\lambda_k | \alpha_k^{(t+1)}, \gamma^{(t+1)})$;
 end
 end
 Synchroniser tous les threads ;
end
return $\{\alpha^{(t)}, \gamma^{(t)}, \lambda^{(t)}\}_{t=1}^T$;

4 Résultats expérimentaux

Nous allons évaluer ce modèle d'estimation dans le cadre de la détection de l'agitation motrice chez les enfants atteints du trouble déficit de l'attention avec hyperactivité (TDAH)[8], en utilisant des données d'accélérométrie. L'objectif est d'identifier des séquences de mouvements d'intensité élevée et inadaptée à partir de ces données. Dans un premier temps, nous validons notre méthode sur des données simulées avant de l'appliquer à des données réelles.

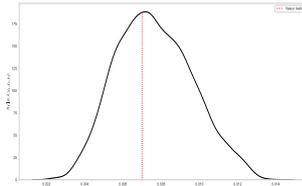


FIGURE 2 : Loi conditionnelle de γ .

4.1 Données simulées

Nous appliquons notre méthodologie sur des données simulées générées à partir d'une distribution gamma trivariée ($K = 3$). Un échantillon aléatoire de taille $T = 20000$ a été obtenu à partir de cette distribution. La convergence vers la distribution *a posteriori* a été vérifiée à l'aide du diagnostic de Gelman et Rubin [9], en utilisant 10 chaînes indépendantes. Le Tableau 1-(a) présente les valeurs réelles utilisées pour générer les données simulées, ainsi que les valeurs estimées des paramètres. Les Figure 2 et 3 illustrent les distributions marginales *a posteriori* des différents paramètres estimés. Les valeurs estimées sont proches des valeurs réelles des paramètres, attestant de la précision de l'estimation. Dans le Tableau 1-(c), le diagnostic de Gelman et Rubin est réalisé à l'aide de la statistique

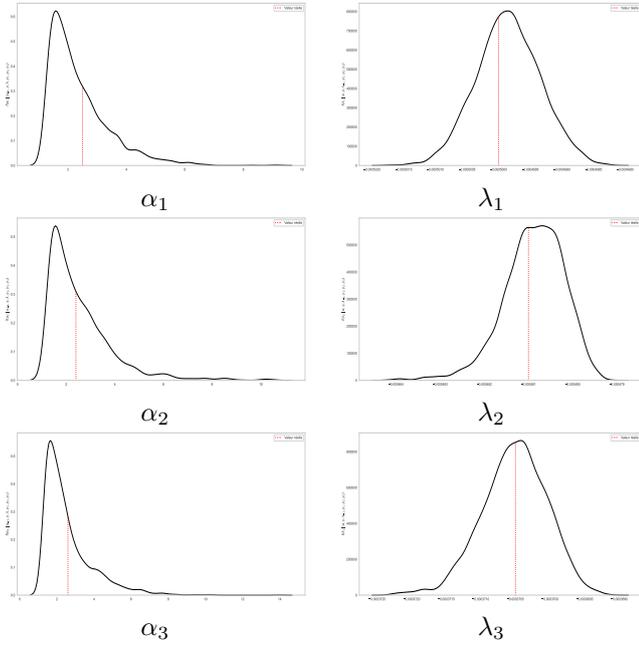


FIGURE 3 : Loïs conditionnelles calculées sur la bases des échantillons simulés après la convergence des chaînes. La droite verticale indique la vérité terrain.

TABLE 1 : Résultats de l’estimation bayésienne et convergence.

Paramètre	α_1	α_2	α_3	γ	λ_1	λ_2	λ_3
(a) Données simulées							
Valeur réelle	2.5	2.4	2.6	0.007	-5	-4.81	-3.705
Valeur estimée	2.4723	2.6428	2.5644	0.0074	-4.998	-4.809	-3.705
(b) Données réelles (erreurs de reconstruction)							
Valeur estimée	0.7940	0.8045	0.7955	0.0478	-5.0204	-4.5064	-4.1250
(c) Diagnostic de convergence (Gelman-Rubin)							
\hat{R}	0.926 01	0.926 00	0.926 02	0.926 11	0.926 40	0.926 47	0.926 51

R qui est un paramètre de réduction d’échelle. Pour chaque paramètre, la valeur de R est proche de 1, ce qui indique une bonne convergence.

4.2 Application à la caractérisation des erreurs de reconstruction

Nous analysons les erreurs de reconstruction de données d’activité physique enregistrées par un accéléromètre tri-axial chez des enfants âgés de 6 à 12 ans [10]. Ces données forment une série temporelle multivariée composée de trois variables ($K = 3$). Le jeu de données d’entraînement, comprenant 6 003 000 observations, a été reconstruit à l’aide du modèle génératif de détection d’anomalies basé sur des modèles de *Transformers* [1]. Le vecteur d’erreur obtenu suit une distribution à paramètres inconnus, que nous estimons en appliquant notre modèle bayésien. L’échantillonnage de Gibbs a été mis en œuvre avec 5 000 itérations. Afin de réduire le temps de calcul sur ce vaste ensemble de données, une exécution parallèle de certaines étapes a été effectuée. Cette optimisation a permis de diviser le temps d’exécution par plus de trois, le faisant passer de 39 minutes et 7 secondes à 11 minutes et 59 secondes sur un ordinateur Apple M2 Pro (10 cœurs). Le Tableau 1-(b) présente les valeurs estimées des paramètres à partir des erreurs de reconstruction. En utilisant ces paramètres, un score

d’agitation (anomalie) est calculé en estimant la densité de probabilité multivariée des erreurs. Soit $e = \|x_t - x'_t\|_2^2$, le vecteur d’erreur du point x_t . Le point x_t est considéré comme anormal si sa densité est supérieure à un certain seuil τ , c’est à dire : $f(e | \theta) > \tau$.

5 Conclusion

Les architectures basées sur les *Transformers* ont démontré leur efficacité pour la modélisation des séries temporelles multivariées et ont été largement utilisées pour des tâches de détection d’anomalies. La clé d’une détection d’anomalies optimale réside dans le choix de la distribution pour modéliser ces erreurs ainsi que dans l’estimation précise de ses paramètres. Cet article propose une approche bayésienne pour estimer les paramètres de la distribution des erreurs de reconstruction en utilisant une distribution gamma multivariée. La méthodologie repose sur la méthode MCMC avec échantillonnage de Gibbs, et pour permettre son application à grande échelle, nous recourons à une exécution en parallèle. Une étude de simulation portant sur l’estimation des paramètres d’une distribution gamma tri-variée a montré que les valeurs estimées sont très proches des paramètres réels, validant ainsi la pertinence de notre approche. Les résultats expérimentaux confirment la performance et la robustesse de notre modèle, ouvrant des perspectives prometteuses pour la modélisation des erreurs de reconstruction. À l’avenir, nous envisageons d’exploiter cette méthode pour détecter des événements critiques, tels que l’hyperactivité, à partir des données d’activité physique.

Références

- [1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, “A review on outlier/anomaly detection in time series data,” *ACM Comput. Surv.*, vol. 54, no. 3, pp. 56 :1–56 :33, 2022.
- [2] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, “Long short term memory networks for anomaly detection in time series,” in *23rd European Symposium on Artificial Neural Networks, ESANN 2015, Bruges, Belgium, April 22-24, 2015*, 2015.
- [3] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, “Lstm-based encoder-decoder for multi-sensor anomaly detection,” *CoRR*, vol. abs/1607.00148, 2016.
- [4] E. G. Tsionas, “Bayesian inference for multivariate gamma distributions,” *Statistics and Computing*, vol. 14, pp. 223–233, 2004.
- [5] S. Das and D. K. Dey, “On bayesian inference for generalized multivariate gamma distribution,” *Statistics & probability letters*, vol. 80, no. 19-20, pp. 1492–1499, 2010.
- [6] A. Rahayu, Purhadi, Sutikno, and D. D. Prastyo, “Multivariate gamma regression : Parameter estimation, hypothesis testing, and its application,” *Symmetry*, vol. 12, no. 5, p. 813, 2020.
- [7] R. Bardenet, A. Doucet, and C. Holmes, “On markov chain monte carlo methods for tall data,” *Journal of Machine Learning Research*, vol. 18, no. 47, pp. 1–43, 2017.
- [8] J. Marquet-Doleac, M. Biotteau, M. Tardy, M. Broquere, E. Marizy, N. Faure-Marie, N. Lafin, S. Dejean, and Y. Chaix, “Pendah program for parents with children with attention deficit hyperactivity disorder. french adaptation of a behavioral parent training group : pilot study,” *Child Neuropsychology*, vol. 30, no. 1234–1259, 2024.
- [9] A. Gelman and D. B. Rubin, “Inference from iterative simulation using multiple sequences,” *Statistical science*, vol. 7, no. 4, pp. 457–472, 1992.
- [10] I. Traore, I. Megdiche, J. Marquet-Doléac, and L. Chaari, “A cross-validation approach for classifying physical activity intensity : a case study in children with attention deficit/hyperactivity disorder,” in *2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–8, IEEE, 2023.