VII International Conference on Mathematics and its Applications in Science and Engineering
(ICMASE 2026)

ICMASE 2026

A POISSON SAULEH HIDDEN MARKOV MODEL FOR OVERDISPERSED AND MULTIMODAL COUNT DATA WITH APPLICATIONS TO PUBLIC HEALTH

Bamigbala Olateju Alao [1,*], Saidu S. Abdulkadir, Adesupo A. Akinrefon

[1] *Department of Statistics, Federal University Wukari, Nigeria*

ABSTRACT

Count data modelling is of significant importance across the biological and physical sciences. Although the Poisson distribution is the conventional approach for modelling count events, its inherent assumption of equal mean and variance limits its practical applicability, particularly in the presence of overdispersion [1]. Mixed Poisson models such as the Negative Binomial and Generalized Poisson distributions were developed to address overdispersion, they remain inadequate for count data exhibiting non-normality, serial dependence, multimodality, non-stationarity, and skewness [1, 3]. To overcome these limitations, the Poisson Sauleh distribution (PSD) was developed and incorporated into a Hidden Markov Model (HMM) framework. The parameters of the HMM are estimated via the Expectation-Maximization (EM) algorithm [2, 4]. Datasets were simulated with sample sizes ranging from 50 to 1,000 across one to five states alongside two real-world datasets on maternal mortality (MMD) and acute encephalitis syndrome (AESD), and both data are confirmed to exhibit multimodality and serial dependence based on Hartigan Dip and Ljung-Box tests. The simulation results indicate that model accuracy improves with larger sample sizes and two-state models consistently yield the lowest AIC and BIC values. When applied to the real datasets, the 2-State Poisson Sauleh HMM outperformed all competing models, recording the lowest AIC and BIC values of 245.16 and 252.04, and 251.66 and 260.03 for MMD and AESD, respectively, with Viterbi algorithm-based classification method for distinguishing low and high case counts in both datasets. The results suggests that the PSD-HMM is a good fit for count data characterised by overdispersion, serial dependence, non-stationarity, and multimodality.

***K*eywords** Hidden Markov Model · Poisson Sauleh · Overdispersion · EM Algorithm · Count Data

References

[1] Ambily, J., Tunny, S. , Sebastian, G. (2020). Modelling time series of count data using hidden markov models. Stochastic Modeling and Applications. 24 (1), pp. 25-33.

[2] Bamigbala, A. O., Abdulkadir, S. S., Akinrefon, A. A., Danjuma, J. (2025). Poisson Sauleh distribution and its properties. International Journal of Development Mathematics, 2(1), 246–266. https://doi.org/10.62054/ijdm/0202.14.

[3] George, J. , Thomas, S. (2019). Negative Binomial Hidden Markov Model for AES count data. Think India Journal. 22 (14), pp. 14448-14451.

[4] Zucchini, W., MacDonald, I. L., Langrock, R. (2017). Hidden markov models for time series: An introduction using r. Chapman; Hall/CRC.

[*] *Corresponding Author's E-mail: bamigbala@fuwukari.edu.ng*