International Journal of Data Science and Analysis
2023; 9(2): 43-49
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20230902.13
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Bayesian Kernel Machine Analysis of Communicable Diseases Distribution in Machakos County, Kenya

Cecilia Mbithe Titus*, Anthony Wanjoya, Thomas Mageto

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

mbimaingi@gmail.com (Cecilia Mbithe Titus), awanjoya@gmail.com (Anthony Wanjoya), tttmageto@gmail.com (Thomas Mageto)

*Corresponding author

To cite this article:

Cecilia Mbithe Titus, Anthony Wanjoya, Thomas Mageto. Bayesian Kernel Machine Analysis of Communicable Diseases Distribution in Machakos County, Kenya. *International Journal of Data Science and Analysis*. Vol. 9, No. 2, 2023, pp. 43-49.
doi: 10.11648/j.ijdsa.20230902.13

Received: September 21, 2023; **Accepted**: October 20, 2023; **Published**: October 31, 2023

Abstract: Effective management and control of communicable diseases are paramount for healthcare managers. Data-driven analysis plays a crucial role in understanding and curbing the spread of such diseases. While numerous epidemiological models have been developed to explain disease spread, many lack the incorporation of prior and posterior probabilities. In this research, we introduce a novel model called BKMR, designed to analyze and predict communicable disease occurrences in Machakos County. This study underscores the significance of data-driven approaches and outlines a plan to evaluate prediction accuracy through empirical analysis, with a particular focus on comparing BKMR with existing models using the R statistical software. We highlight the differences between estimated parameters and actual observations, emphasizing aspects not present in the training dataset. Our findings demonstrate that BKMR outperforms the Poisson regression model, offering greater flexibility and robustness. Moreover, it provides the ability to quantify uncertainty in model parameters, enhancing the capacity to make inferences about the real world. This research has substantial implications for healthcare management and disease control efforts in Machakos County.

Keywords: Communicable Diseases, Data Driven Analysis, Health-Care Managers, Epidemiological Models, Prior Probabilities, Posterior Probabilities, BKMR, Poisson Distribution

1. Introduction

Kernel functions introduce a class of machine learning for non-linear classification that is flexible, able to accommodate different types of data and detects different types of relations. The models are upper hand given that they simulate joint responses and detects relations amongst groups of discrete variables. Most cases, the event likelihood is found using the average rate at which it happens. Many times, the characteristics surrounding the event are associated with the rate. To merge the influence of component characteristics and prior information of failure, a model is proposed as an approach to an estimation which is more accurate of the occurrence rate of a communicable disease. The intent of this study is to consolidate prior data with historical record and ascribe the information to produce the rate of occurrence probability distribution of the Communicable diseases in Machakos County. The communicable diseases (i.e. HIV, TB and Malaria), will continue to be a crucial global public health issue. 2015, 1.3 million people passed on from HIV related causes and 1.8 million passed on from TB which included 0.4 million amid people with HIV globally. Cause of death, by communicable diseases in Kenya was reported at 48.29 % in 2019, according to the World Bank collection of development indicators. HIV prevalence is estimated at 14 per 1000 people in 2019, higher than the global 0.24 per 1000 people. TB incidence is estimated at 140 out of hundred, 000 populace in 2019 which is also higher than worldwide incidence, 132 out of hundred, 000 populace. Malaria incidence is estimated at 95.7 per 1,000 in 2019, which is 1.7 times higher than the worldwide incidence of 57.4 per millennium population. With these statistics, this study aims to confound itself in using this prior information to predict future communicable disease occurrence using the Poisson model.

1.1. Statement of the Problem

The algorithms for detecting nonlinear relations have started

to appear with heuristic tools like decision trees and ANN (Artificial Neural Networks) with applications across many disciplines like healthcare, geotechnical engineering and atmospheric sciences among many others. The main extension of the Kernel techniques building on the basic model is the non-Gaussian model, which improves the accuracy predictions of certain matters. This is where model parameters Gaussian distribution should not be taken for granted.

This study shall expand on non-symmetric models by using a more robust model hinged on the Gamma conjugate prior which gives significant speedup computation time than the old Bayesian kernel technique that depend on either simulation or optimization to find the solution to analyze communicable disease frequency in Machakos.

1.2. Significance of the Study

The Kernel techniques have gained an extensive use in different research fields. This study shall employ methods as a similarity metric to improve predictive accuracy of Bayesian methodologies for count data. It shall also present uncertain predictions to longstanding classification tools for analyzing communicable disease frequency.

Particularly, the matrix given by the computation of inner product and the implicit mapping shall be used to consolidate information from the attributes for communicable disease numbers into the model fitting of the posterior distribution. This will in turn help health-care providers understand the communicable diseases dynamics as a road-map for efficient health-care provision.

1.3. Objectives

1.3.1. General Objective

The General objective of the study shall be to analyze communicable diseases frequency in Machakos County using Machine Learning Models.

1.3.2. Specific Objectives

This study will assess on the following specific objectives;

Evaluate communicable disease frequency using Poisson Methodology.

Estimate model constants of the Poisson distribution.

Compare the attainment of the proposed model to the generalized linear model for count data (Poisson distribution model).

Perform model diagnostics to determine the model of good fit and use the selected model to do prediction.

2. Literature Review

2.1. Introduction

This is the second chapter of the thesis. It explores the empirical literature to substantiate the need for this research.

2.2. Empirical Literature Review

Bobb [1] proposed a BKMR for predicting the impact of health in the mixture of multiple pollutants. They introduced Bayesian kernel as a route to learn about mixtures. The relationship between the health outcome and the mixture is modeled using a flexible function, which is a kernel function. In the advanced setup, variable selection for a new hierarchical approach was merged to pinpoint the constituents that were important and explain the interrelated makeup of the mixture. Modeling research demonstrated the achievement of BKMR in predicting the dose-outcome relationship and identifying the constituents accountable for health consequences. They demonstrated those features using toxicology and epidemiology employments.

Areti [2] developed a modeling approach in the mixture for Bayesian for Public health monitoring capable to assess the risk of disease in time and space and be able to identify anomalies as well. The data used was for hospital admissions on asthmatics on areas to compute the anticipated number of cases taking into account the sex and age traits by use of expected rate method and get the posterior estimates from the Common model, which were therefore used to simulate the variable outcome. They found that proposed model accomplishment was unsatisfactory during the anomalous time trend was stable since it assigned a relatively high fraction of false indications.

Baroud & Barkar [3] proposed a BKMR approach to model the importance of resilience based network. They applied a beta BKMR in order to figure out the resilience of essential infrastructure networks by assessing the historical data on the component's characteristics as well as the resilience contribution of each component in the network using pre-existing knowledge. They used their proposed approach to research the gravity of 4 dams and locks alongside the Mississippi River Navigation System. The most predictive power was attained with a consistent pre-existing allocation. By use of Bayesian posterior distribution and a multi-criteria decision analysis method, they identified five dams and locks with most impact on the system robustness.

Irvene & Hollingsworth [4] designed a Kernel-density estimator for a flexible epidemiological model using approximate Bayesian computation. They developed an approximate, which was adaptive Bayesian calculation scheme to fit various epidemiological data that is relevant with a few hyper-parameter tweaking using an adaptive system that is tolerant. Theirs was an implementation of a Kernel smoothing method to apprehend the multi-dimensional and dispersed data, and weigh this technique against the standard of Bayesian statistics. An application of this procedure was complex individual based Lymphatic filariasis model which is a parasitic infection. They found that a Self-adjusting approximate Bayesian calculation plan with a distance metric and general summary is able to perform model calibration for a diversity of disease outbreak data. In this, no prior experience is required, which is theoretical and is reachable by the divergent Health data research community.

Bobb [5] developed a statistical computing software to analyze combined health effects exposures through BKMR. They introduced the BKMR R package (R programming language) and demonstrated visualizing ways for complex exposure response functions and estimation for concise and

informative overviews of scientific research. They illustrated a BKMR probit regression package for dichotomous outcomes and described a BKMR efficient version that leverages Gaussian process regression plan. This is a newly developed software with an extended methodology and integrated tools, making BKMR accessible widely in epidemiological employments where numerous determinants have intricate health effects.

Jinkai [6] did Kernel analysis research on Dirichlet processes mixture models. They combined kernels spectral domains with non-parametric models for Bayesian. They presented the infinite mixture kernel by combining the analyzed spectral domain with the hierarchical Bayesian model. Dirichlet mixture processes methods were used by changing the component's number according to the size of data. They did a lot of experiments on their model and it showed competitive output than the bounded version. Their discoveries were ambitious with popular employed kernels and the solidity of spectral for the IM kernel provided an additional data exploration.

Luong [7] proposed a model based on Kernel to model aperiodic phenomena in Hierarchical Bayesian models. Theirs was a way that combined current Bayesian Dynamic LM with a kernel smoothing for time series regular patterns. The procedure was utilized to model the piezo metric pressure under a dam and the Tamar Bridge traffic volume. The output showed that the method proposed succeeded in simulation both the non-motionless and motionless repeating patterns case studies. It was also versatile, calculation proficient, capable to handle observations that are collected at irregular time intervals and self-adaptive to changing conditions.

Nishiyama [8] researched on sum rule kernel model, which is a Gaussian process inference with predictive models. They introduced a novel method, the Mb KSR, to integrate a kernel Bayesian inference with a probabilistic model. They considered probabilistic filtering in a model that is phase space, where there be a well-fitting probabilistic model for the state evolution. Their method combined data driven non parametric inference for the sampling process and model driven inference for the state change. They further on validated their proposed method with real data experiments and synthetic, the latter being the image-based problem robot localization in automation. This illustrated the proposed hybrid approach proficiency.

Marion [9] researched on Bayesian semi-supervised clustering to work on various strongly linked exposures, a Censored Survival Output and proposed a hierarchical model that comprises an essential zero-truncated DP mixture as an attribution model component. They extended the Bayesian PRM models by presuming a surplus disease risk sub-model. They implemented a highly tailored Metropolis in the Gibbs sampling that included the tag switching steps. They indicated that PRM methods are promising tools for exosome study thus opening methods for research in the type of chance models.

Yadpirun [10] proposed a Poisson distribution parameter estimation which is a Bayesian approach. He considered the Poisson parameter to be a random variable for which he used knowledge of prior distributions in order to change the estimation theory. He proposed 2 Bayesian tactics for estimation of the Poisson parameter by obtaining the updated distribution beneath quadratic loss functions or squared error loss. They compared the output with Empirical Bayes and classical methods through uncertainty simulations. They used the average square to test metric for contrasting the methods for direct estimation. His results revealed that Bayesian methods are the best for point prediction when the precise parameter figure was small.

Arti [11] researched on BKMR application of decision-making processes in medical diagnosis. They found that Kernel average is a is a model-agnostic way to compute conditional probability and that they are needed to synthesize inference rules for uncertainty. They evaluated Bayes estimators for different categories of assumptions i.e. the Gaussian processes, Laplacian processes and vector machines. They found the Laplacian and Gaussian kernels as to determining the uniquely probability but incapable to determine Polynomial kernels.

Jiadong [12] proposed the Bayesian statistics in practice on the foretelling and prediction of cancer. They arranged Bayesian technique based on sophisticated genetic data for cancer risk assessment using the multi-omics integration method. They found that the technology of the multi-omics integration mostly faced a key challenge of effectively integrating the relation between many platforms thus insinuating a need for further research.

Yuxia [13] researched on the alliance between the indicators of depression in adults aged 55 and over and multiple dietary patterns in Northern China. They used the Principal Component Analysis (PCA), Linear Regression, Quantile g calculation (QGcomp), Weighted Quantile Sum Regression and BKMR. In the representation for the dietary patterns, they noticed that the fruit based diet and vegetable had a markedly inhibitory consequence on women and the diet for egg and milk had a strong inhibitory influence on men. They further noticed that the when all aspects for dietary were surpassing 55% there was a considerable positive correlation in-between several food habits and liability to depression. There was an appreciable correlation amidst obesity risk and meat. An inverse correlation between milk & egg, and fruits & vegetables was also established.

3. Research Methodology

This section discusses the methodology to be applied in the research. It bestows a review of the Bayesian Kernel Methods for the analysis of communicable diseases frequency in Machakos County. A mention of the research data applied in the study is also given.

3.1. Research Data

This study will use a secondary data obtained from KNBS and The Demographic Health Surveys Program on the spread of communicable diseases. A major concern shall be on HIV/AIDS, Tuberculosis and Malaria.

3.2. Bayesian Kernel Model for Count Data

For an $m \times d$ X matrix with corresponding rows m data with each d properties, a function, $t(X)$ is defined as a variate random vector m to map input data $t(X)$ into a class. Conventional Kernel methods use Gaussian methods where the valued vector function t is presumed to follow a normal multivariate distribution with mean $E(t(X)) = 0$ and matrix.

Covariance of $Cov(t(X)) = K$, where K is matrix element and +ve definite.

K_{ij} is the function of kernel $K(X_i, X_j)$ between the i^{th} and j^{th} data. The normal multivariate distribution for the recognition of t is;

$$P(t) = \frac{1}{\sqrt{(2\pi)^m}} (det\, K)^{-\frac{1}{2}} exp(-\frac{1}{2} t^T K^{-1}\, t) \quad (1)$$

In order to model communicable diseases frequency, this study shall use Bayesian Kernel Methods for count data. These are machine learning methodologies introduced to prove probabilistic solutions as opposed to deterministic solutions for research problems.

3.2.1. Count Data Modeling

The Poisson distribution and Negative Binomial distribution are models for count data and are defined here in respectively as;

$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!}, \hat{\lambda} = e^{\beta_i X} \quad (2)$$

And

$$P(y) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \left(\frac{k\lambda}{1+k\lambda}\right)^y \left(\frac{1}{1+k\lambda}\right)^{\frac{1}{k}} \quad (3)$$

Where k is the dispersion parameter and λ is believed to have a gamma distribution. A multivariate distribution shall be applied to model various discrete data as follows:

$$Yk_i/\lambda k_i \underset{\sim}{iid} Pois\,(\lambda k_i)$$

$$\left(log(\lambda k_i)\right)_{i=1,..,4}/\mu, T \underset{\sim}{iid} N_4\,(\mu, T)$$

$$\mu_i \underset{\sim}{iid} N(0, 0.0001)$$

$$T \sim Wishart\,(R, 4) \quad (4)$$

$N_m(\mu, T)$ Is m dimensional normal multivariate distribution. The mean vector is given by μ and T is the reciprocal of the co-variance matrix. Hyper-parameters R and $\pi = 4$ are acquainted.

3.2.2. Poisson Bayesian Kernel Model

The Poisson distribution shall approximate the occurrence rate, $\lambda > 0$ which ensues from a prior Gamma distribution with constants $\alpha > 0$ & $\beta > 0$, as follows:

$$P(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{(-\beta\lambda)} \quad (5)$$

The likelihood function shall be obtained as;

$$L = \prod_{i=1}^m P(y_i) = \prod_{i=1}^m \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \Rightarrow \frac{\lambda_i^{\sum_{i=1}^m y_i} e^{-m\lambda_i}}{\prod_{i=1}^m y_i!} \quad (6)$$

Defining $\alpha^* = \sum_{i=1}^m x_i + \alpha$ and $\beta^* = m + \beta$, the posterior distribution shall be given as;

$$P\,(\lambda/x) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{(-\beta\lambda)}\right)\left(\lambda^{\sum_{i=1}^m y_i} e^{-m\lambda}\right)$$

$$= \frac{\lambda^{\sum_{i=1}^m y_i + \alpha - 1} e^{-\lambda(m+\beta)} (n+\beta)^{\sum_{i=1}^m y_i + \alpha}}{\Gamma(\sum_{i=1}^m y_i + \alpha)}$$

$$= Gamma\,(\alpha^*, \beta^*) \quad (7)$$

For

$$\alpha^* = KY + \alpha$$

$$\beta^* = KV + \beta \quad (8)$$

Where K is $m \times m$ kernel matrix, Y is the $m \times 1$ output data vector related with m value of X & V is the $m \times 1$ vector, with ones.

This study shall use the Radius Basis kernel Function (RBF) where $k(x_i, x_j)$ is a single record in matrix K showing the function of kernel amidst the characteristic of i^{th} and j^{th} data given as;

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

3.3. Goodness of Fit Measures

To assess model performance, the log likelihood and deviance residuals must be utilized to identify the model's capability to apprehend data patterns.

3.3.1. Deviance Residuals

Deviance finds the difference in the log likelihood function amidst the saturated model, and fitted model. It is represented as;

$$D_p = 2 \times \sum_{i=1}^m y_i \log\left(\frac{y_i}{\hat{\lambda}_i}\right) - \left(y_i - \hat{\lambda}_i\right) \quad (10)$$

will be used to evaluate how fitted figures representing the occurrence rate observed in Poisson BKM are compared to the generalized linear models.

3.3.2. Log-Likelihood Function

The functional value of log likelihood represents the observed data joint probability as a parameter function of interest that shall be for this study. The higher the value of the function, the better the model thus being able to capture the data patterns using the parameters estimated. The Log-Likelihood shall be given as;

$$l(\hat{\lambda}/y) = \sum_{i-1}^m \left[y_i \ln(\hat{\lambda}_i) - \hat{\lambda}_i - \ln(y_i!)\right] \quad (11)$$

where y is the data value and the rate estimated for a specific data point.

3.4. Prediction Accuracy

Validating the prediction of the models, some metrics shall be assessed to evaluate the unseen data error.

3.4.1. Root Mean Square Error

RMSE shall be obtained as;

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2} \quad (12)$$

3.4.2. Normalized Root Mean Square Error

The NRMSE shall be obtained as;

$$NRMSED = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{sd(Y_i)} \quad (13)$$

$$NRMSEM = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}_i\right)^2}}{Y_{max} - Y_{min}} \quad (14)$$

3.4.3. Mean Absolute Error

The MAE shall be obtained as;

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{\lambda}_i\right| \quad (15)$$

4. Results and Discussions

This is the fourth phase of this project and it gives the results and discussions.

4.1. Exploratory Data Analysis

Poisson regression will be used to predict the dependent variable (Population of Machakos county) given the following independent variables; HIV/Aids, Tuberculosis and Malaria infections in the county.

The study uses the Bayesian Kernel Model to apply prior information in order to analyze and predict the frequency of these communicable diseases in the county.

4.2. Descriptive Statistics

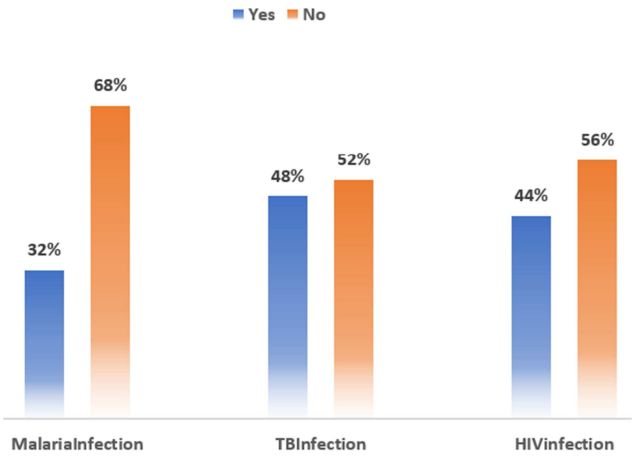

Figure 1. *Descriptive Statistics.*

Based on a sample of 392 from Machakos County, those infected by Malaria, Tuberculosis and HIV were 32%, 48% and 44% respectively. This indicates a significant number of the communicable diseases within the region.

4.3. Correlation Matrix for Covariates

```
     z1   z2   z3   z4
z1 1.00 0.12 0.95 0.04
z2 0.12 1.00 0.32 0.04
z3 0.95 0.32 1.00 0.05
z4 0.04 0.04 0.05 1.00
```

Figure 2. *Correlation Matrix for Covariates.*

Where;

Z_1 = Machakos County Sample $n = 392$

Z_2 = Malaria Infection

Z_3 = TB Infection

Z_4 = HIV Positive Status

Based on the output, there is a high correlation amid Machakos county population and TB Infection.

4.4. Goodness of Fit

Deviance Residuals

```
Null deviance: 29560  on 391  degrees of freedom
Residual deviance: 29382  on 388  degrees of freedom
```

Figure 3. *Goodness of Fit.*

4.5. Count Data Modeling

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     5.249148   0.005496 955.143  < 2e-16 ***
MalariaInfection1  0.043932   0.007704   5.703 1.18e-08 ***
TBInfection1    -0.119454   0.015451  -7.731 1.07e-14 ***
HIVinfection1    0.168236   0.015524  10.837  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 29560  on 391  degrees of freedom
Residual deviance: 29382  on 388  degrees of freedom
AIC: 32064
```

Figure 4. *Poisson distribution modeling in R.*

$$L_n(\lambda) = 5.2491 + 0.043932\text{MalariaInfection} - 0.119454\text{TBInfection} + 0.168236\text{HIVInfection}$$

For every unit increase in Machakos population the Malaria Infection increased by 0.043932, for every unit increase in Machakos population the TB Infection reduces by 0.119454 and for every unit increase in Machakos population the HIV Infection increases by 0.168236.

Interpretation of parameter estimates

Holding TB Infection and HIV Infection constant, Machakos County is 4% more likely to have Malaria infection. Holding Malaria Infection and HIV Infection constant, Machakos County is 12% less likely to have TB infection. Holding TB Infection and Malaria Infection constant, Machakos County is 17% more likely to have HIV Infection.

Fitting Bayesian Kernel Model

The fundamental formula behind Bayesian Machine Learning is Bayes' theorem:

$$P(H/D) = (P(D/H) * P(H)/P(D))$$

Where;

$P(H/D)$ Is the hypothesis H posterior probability due to the data observed D.

$P(D/H)$ Is data D likelihood given hypothesis, H.

$P(H)$ Is hypothesis H prior probability.

$P(D)$ Is the observed data D probability.

Investigating Model Convergence

We virtually inspect how various parameter values change as the sample runs.

The trace plot shows the values of μ generated over time. The blue line shows the true value of μ. The trace plot is initially quite variable, but it eventually converges to the true value of μ. This means that the samples are being able to be generated from the posterior distribution of μ. The probability of accepting the proposal determined by the current value, the proposed value and posterior distribution of μ.

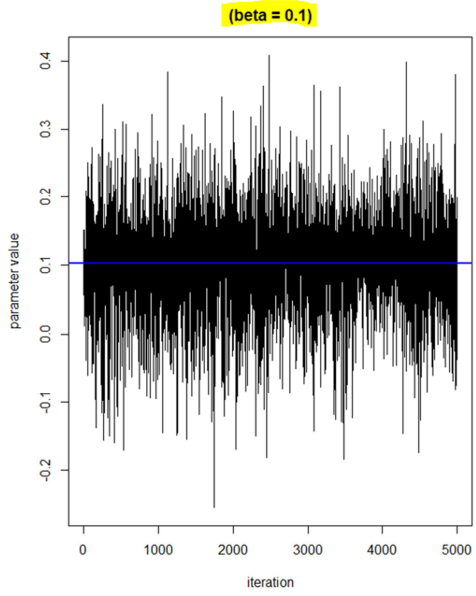

Figure 5. *Parameter Value β.*

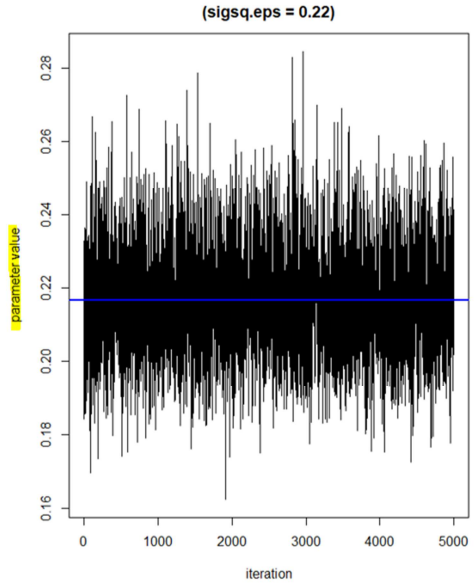

Figure 6. *Parameter Value μ.*

Here, the algorithm has been run for 5,000 iterations. The y axis represents values of μ and x axis represents the iteration number. Generally, the trace plot indicates the model is working well and is simulating samples from posterior distribution of μ.

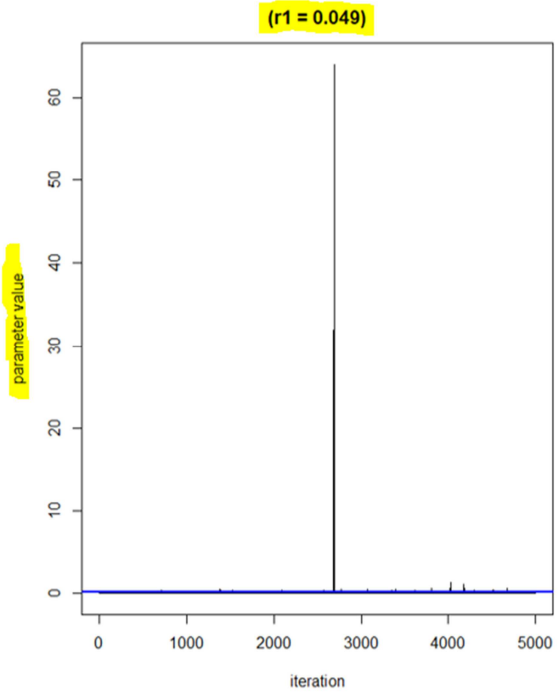

Figure 7. *Parameter Value r.*

The program is generating a random variable X from a distribution with parameter value r. The blue line indicates average value of X after each iteration of the program. The white line indicates the average value of X over that entire program. The black line indicates the parameter value r.

The graph shows the blue line converges to the white line as the number of iterations increases. This implies that the average X value is getting closer to the expected value of X as the program runs longer.

The graph also shows that the convergence is faster when the parameter value r is smaller. This is because the random variable X is less likely to take on extreme values when the parameter value r is smaller.

5. Conclusions and Recommendations

5.1. Summary

Modeling of communicable diseases distribution data, the study gave an application of the count data model and Bayesian Kernel model in which both models fitted the data perfectly well.

Even though both models were fitting the data, BKMR is more flexible and robust method than poisson regression model. It can quantify uncertainty in the model parameters, which is important for making inferences about the real world.

A low deviance indicates that a model fits the data well whereas a high deviance indicates the model does not fit the

data well. In this study, Poisson distribution model has a high deviance since the model is too simple and does not capture the complexity of the data.

This involved the study to choose a more complex model that is the Bayesian Kernel regression model.

5.2. Conclusion

The study has shown high probabilities of there being high increase in communicable diseases in the near future in Machakos County. This is seen with Malaria and HIV Infections as well.

Though there seem to be a negative probability for Tuberculosis Infections in the near future, this does not mean for it to be overlooked during intervention.

5.3. Recommendations

Communicable diseases like Malaria, TB and HIV are seen to pause a great challenge to Eastern part of Kenya and specifically Machakos County. This study suggests that for the country to achieve the vision 2030 of good health, more interventions need to be done on these diseases.

This can be through creating awareness or improving on the awareness campaigns and providing preventive measures like distribution of mosquito nets and fumigation to prevent the spread of malaria, clean water and sanitation to reduce TB infections, and more sensitizing on HIV prevention. Government can also find more donors to assist in funding for more projects in Machakos County that deal with prevention and awareness of these communicable diseases.

References

[1] Ali, H., Nwaosu, S. C., Lasisi, K. E & Abdulkadir, A. (2022). Bayesian Information Based Criteria for Poisson Regression Models. *International Journal of Formal Sciences*, 13 (1); 22-40.

[2] Atonio, C. & Dunson, D. B. (2011). Bayesian Kernel Mixtures for Counts. *Journal of AM Statistical Association*, 106 (496); 1528-1539. doi: 10.1198/jasa.2011.tm10552.

[3] Arti, S. (2021). Bayesian Kernel Methods: Applications in Medical Diagnosis Decision-Making Process; A Case Study. *International Journal of Big Data and Analytics in Healthcare*, 6 (1); 26-39. doi: 10.4018/IJBDAH.20210101-oa3.

[4] Areti, B., Bennett, J. E. & Marta, B. (2018). A Bayesian Mixture Modeling Approach for Public Health Surveillance. *Biostatistics*, 00 (00); 1-15. doi: 10.1093/biostatistics/kxy038.

[5] Baroud, H. & Barkar, K. (2018). A Bayesian Kernel Approach to Modelling Resilience-Based Network Component Importance. *Reliability Engineering and System Safety*, 170 (1); 10-19. doi: 10.1016/j.ress.2017.09.022.

[6] Bang sheng, W., Jiang, Y., Xiaoqing, J. & He, L. (2020). Using Three Statistical Methods to Analyze the Association between Exposure to nine Compounds and Obesity in Children and Adolescents: NHANES 2005-2010. *Environmental Health*, 19 (94); 1-13. doi: 10.1186/s12940-020-00642-6.

[7] Bobb, J. F., Henn, B. C., Valeri, L. & Brent, A. C. (2018). Statistical Software for Analyzing the Health Effects of Multiple Concurrent Exposures via Bayesian Kernel Machine Regression. *Environmental Health*, 17 (67); 1-10. doi: 10.1186/s12940-018-0413-y.

[8] Bobb, J. F., Valeri, L., Claus, B. H., David, C. C., Roberto, O. W., Mazumdar, M., Godleski, J. J. & Brent, A. C. (2015). *Biostatistics*, 16 (3); 493-508. doi: 10.1093/biostatistics.kxu058.

[9] Flavio, B. G., Livia, M. D. & Roger, W. C. S. (2022). Exact and Computationally Efficient Bayesian Inference for Generalized Markov Modulated Poisson Process. *Statistics and Computing*, 32 (14); doi: 10.1007/s11222-021-10074-y.

[10] Frenoy, P., Perduca, V., German, C. S., Jean, P. A., Gianluca, S. & Francesca, R. M. (2022). Applications of Two Statistical Approaches (Bayesian Kernel Machine Regression and Principal Component Regression) to Assess Breast Cancer Risk in Association to Exposure to Mixtures of Brominated Flame Retardants. *Environmental Health*, 21 (27); 1-17. doi: 10.1186/s12940-022-00840-4.

[11] Gyeyoon, Y., Yuting, W., Caitlin, G. H. & Megan, E. R. (2022). Exposure to Metal Mixtures in Association with Cardiovascular Risk Factors and Outcomes: A Scoping Review. *Toxics*, 10 (116); 1-32. doi: 10.3390/toxics10030116.

[12] Heba, S. M. (2021). Empirical E-Bayesian Estimation for the Parameter of Poisson distribution. *AIMS Mathematics*, 6 (8); 8205-8220. doi: 10.3934/math.2021475.

[13] Irvine, M. A. & Hollingsworth, D. T. (2018). Kernel-Density Estimation and Approximate Bayesian Computation for Flexible Epidemiological Model Fitting in Python. *Epidemics*, 25 (5); 80-88. doi: 10.1016/j.epidemics.2018.05.009.