IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 19, Issue 2 Ser. II (Mar. – Apr. 2023), PP 46-54
www.iosrjournals.org

Bayesian Modelling Of Churn Confounding Competing Risks With Time-Dependent Covariates Among The Mobile Phone Subscribers In Kenya

Mr. Ndilo Benjamen Fwaru[1], Dr. Leonard Alii[2], Dr. Jerita Mwambi[3], Dr. Otulo Wandera[4]

[1] *Mathematics and Computer, Pwani University, Kenya*
[2] *Mathematics and Computer, Pwani University, Kenya*
[3] *Mathematics and Computer, Pwani University, Kenya*
[4]*Mathematics and Physics, Technical University of Mombasa, Kenya*

Abstract
Background: *Mobile phone service providers operate in an oligopolistic market structure and are experiencing high churn rates. Consequently, the service providers invest heavily in market research trying to develop precise ways of predicting customer churn. Building models to predict customer churn in the mobile phone industry when competing risks are confounded with time-dependent covariates has been the center of focus for many studies recently.*
Materials and Methods: *This paper develops a Bayesian model (Gamma model) using a gamma prior, a penalized likelihood, and a posterior gamma to model churn confounding competing risks with time-dependent covariates and uses data from three mobile phone service providers in Mombasa and Kilifi Counties in Kenya to analyze and evaluate the performance of the model.*
Results: *The paper establishes that the Bayesian model (Gamma model) is a better model for predicting subscriber survival probabilities when competing risks are confounded with time-dependent covariates.*
Conclusion: *The Gamma model can be used to predict subscriber survival probabilities when competing risks are confounded with time-dependent covariates.*
Keyword*: Bayesian; Competing risks; Time-dependent covariates; Churn.*

Date of Submission: 02-04-2023 Date of Acceptance: 13-04-2023

I. Introduction

The telecommunications industry has undergone major changes over the past decades such as the addition of new services, technological advancements, and increased competition due to deregulation (Huang, Huang, and Kechadi, 2011). Despite these changes, customer churn remains a serious issue of concern for most mobile phone service providers. Predicting customer churn has become critical for service providers to protect their loyal customer base, drive business growth, and enhance customer relationship management (Idris and Khan, 2012). The service providers organize a variety of research and campaigns to predict the churners before they churn since it is more cost-effective to retain a subscriber than to acquire a new one. Therefore, the ability to predict whether a subscriber is at high risk of churn is an important additional potential revenue stream for service providers. (*Customer Churn Prediction & Prevention Model*, 2021, December 22).

Bayesian models have been used to understand subscriber behavior and attributes that indicate the risk and timing of a customer's churn because of their ability to estimate the likelihood of occurrences. Two Bayesian algorithms; Bayes Naive and Bayes networks have been used extensively in the literature to model churn in the telecommunications industry.

Shyam (2003) used the Naive Bayes classifier on a data set containing 2000 customers, the model showed 68% accuracy. The main limitation of the model is the feature independence assumption. The technique also

[1]*Mathematics and Computer, Pwani University, Kenya*
[2] *Mathematics and Computer, Pwani University, Kenya*
[3] *Mathematics and Computer, Pwani University, Kenya*
[4] *Mathematics and Physics, Technical University of Mombasa, Kenya*

builds algorithms with low variance, as it is fairly immune to data fluctuations (Heckerman, Geiger Chickering, 1995). Thus, making the predictions to be less accurate than high-variance models.

Brandusoiu and Toderean (2013) compared three data mining techniques: Bayesian networks, logistic regression, and k-nearest neighbors. The Bayesian network provided acceptable accuracy and was very close to the accuracy produced by the other two algorithms. Bayesian networks (Heckerman 1997) appeared at the intersection of artificial intelligence, statistics, and probabilities, and constituted a representation formalism for the data mining process (Pearl 1988). Bayesian networks too have practical limitations on the nature of the distribution and the form of statistical dependence and are not ideal for computing small probabilities. In addition to the above limitations, Bayes naïve and Bayesian networks cannot model churn when competing risks are confounded with time-dependent covariates.

This paper develops a Bayesian model of churn when competing risks are confounded with time-dependent covariates. Mobile phone service providers can use the model to determine the propensity of churn precisely and predict customer churn. The paper also adds up to the existing literature on modeling churn when competing risks are confounded with time-dependent covariates and application in other areas such as employee turnover within a business, the lifetimes of components and equipment, duration of unemployment, and cause of death among patients when competing risks are confounded with time-dependent covariates.

The paper is organized into five sections: section one gives the introduction, section two the material and methods, section three explains the results, and the discussion and conclusion are presented in sections four and five respectively.

II. Material And Methods

The paper develops a Bayesian model of churn (Gamma model) to determine customer propensity to churn as well as predict future customer churn when competing risks are confounded with time-dependent covariates.

In developing the model, we assume that customer churn$\sim Gamma(x;\theta)$. The probability distribution function for the Gamma distribution is

$$f(x|\alpha,\lambda)=\frac{\lambda^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}exp(-\lambda x);\ x>0, \qquad [1]$$

and $\theta=(\lambda,\alpha)$ is the vector of unknown parameters. $\alpha>0$ and $\lambda>0$ are the shape and scale parameters respectively. Its cumulative distribution function (CDF) is given by

$$F(x|\theta)=\int_0^{\lambda x}\frac{1}{\Gamma(\alpha)}w^{\alpha-1}e^{-w}dw=\gamma\left(\alpha;\frac{\lambda x}{\Gamma(\alpha)}\right) \qquad [2]$$

where $\gamma(y,x)=\int_0^{x}w^{y-1}e^{-w}dw$ is called lower incomplete gamma function. The survival function is given by

$$S(x|\theta)=1-F(x|\theta)=1-\int_0^{\lambda x}\frac{1}{\Gamma(\alpha)}w^{\alpha-1}e^{-w}dw=1-\frac{\Gamma(\alpha,\lambda x)}{\Gamma(\alpha)} \qquad [3]$$

where $\Gamma(y,x)=\int_x^{\infty}w^{y-1}e^{-w}dw$ is the upper incomplete gamma function. The hazard function is given by

$$h(x|\theta)=\frac{f(x|\theta)}{S(x|\theta)}=\frac{\frac{\lambda^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}exp(-\lambda t)}{\frac{\Gamma(\alpha,\lambda x)}{\Gamma(\alpha)}}=\frac{\lambda^{\alpha}x^{\alpha-1}exp(-\lambda x)}{\Gamma(\alpha,\lambda x)} \qquad [4]$$

Assuming conditional independence across individuals, the full likelihood is the product of [4], across all customers $i=1,\dots,6000$.

The functional form for $F_j(x_j|\cdot)$ is a Gamma distribution, parameterized such that $\theta_j=[\alpha_j,\lambda_j]$, $j=(network\ quality\ churn, service\ quality\ churn, price\ sensitivity\ churn,$ $carrier\ responsiveness\ churn, fraud\ churn)$ represents competing risks. Under this parameterization, α_j is the shape parameter, λ_j is the scale parameter, and Γ is the gamma function which has the formula

$$\Gamma(x_j)=\int_0^{\infty}s_j^{\,x_j-1}e^{-s_j}ds_j$$

The shape parameter, α_j, that affects duration dependence and tail behavior and the scale parameter, λ_j, represents customer's mean "churn time" attributable to risk j, that is, $\alpha_1=15$ and $\lambda_1=5$ reflects a belief that from a total of 15 respondents there were on average 5 churners due to network quality per week. The parameters λ_j and α_j are the only model parameters that require an explicit declaration of prior knowledge.

Assuming that the scale and the shape parameters have gamma priors and are independently distributed, we perform a penalization in the likelihood function of α_j and λ_j using a modified Jeffreys prior (Jeffreys, 1946) as a penalization term of [1] to overcome the difficulty in finding the MLEs of α_j and λ_j and improve the estimates. The penalized likelihood function of α_j and λ_j is given by

$$L_{p_j}(\alpha_j,\lambda_j|x_j)=\left|\sqrt{\alpha_j^{2}\left[\Psi'(\alpha_j)\right]^{2}-\Psi'(\alpha_j)-1}\right|\left(\prod_{j=1}^{5}x_j^{(\alpha_j-1)}\right)\alpha_j(\lambda_j)^{5\alpha_j-1}\left(\Gamma(\alpha_j)\right)^{-5}exp\left(-\lambda_j\sum_{j=1}^{5}x_j\right) \qquad [5]$$

The joint posterior distribution for α_j and λ_j, is given by

$$P(\alpha_j,\lambda_j) \propto \frac{\left|\sqrt{\alpha_j{}^2[\Psi'(\alpha_j)]^2-\Psi'(\alpha_j)-1)}\right|}{\alpha_j\lambda_j}\left(\prod_{j=1}^{5} x_j{}^{(\alpha_j-1)}\right)(\lambda_j)^{5\alpha_j}\left(\Gamma(\alpha_j)\right)^{-5} exp\left(-\lambda_j\sum_{j=1}^{5} x_j\right)$$

$$P(\alpha_j,\lambda_j) = \frac{1}{d_1(x_j)}\left|\sqrt{\alpha_j{}^2[\Psi'(\alpha_j)]^2-\Psi'(\alpha_j)-1}\right|\left(\prod_{j=1}^{5} x_j{}^{(\alpha_j-1)}\right)(\alpha_j)^{-1}(\lambda_j)^{5\alpha_j-1}\left(\Gamma(\alpha_j)\right)^{-5} exp\left(-\lambda_j\sum_{j=1}^{5} x_j\right)$$

Where $d_1(x_j)$ is the posterior density and is given by

$$d_1(x_j) = \int_A \left|\sqrt{\alpha_j{}^2[\Psi'(\alpha_j)]^2-\Psi'(\alpha_j)-1}\right|\left(\prod_{j=1}^{5} x_j{}^{(\alpha_j-1)}\right)(\alpha_j)^{-1}(\lambda_j)^{5\alpha_j-1}\left(\Gamma(\alpha_j)\right)^{-5} exp\left(-\lambda_j\sum_{j=1}^{5} x_j\right)d\theta$$

Where $\theta = (\alpha_j,\lambda_j)$ and $A = \{(0,\infty)\times(0,\infty)\}$ is the parameter space of θ.
The full conditional posterior distributions for α_j and λ_j are given as follows:

$$P_M(\lambda_j|\alpha_j,x_j) = \frac{\prod_{j=1}^{5} x_j{}^{\alpha_j}}{\sum_{j=1}^{5} x_j}$$

$$P_M(\lambda_j|\alpha_j,x_j) = \left(\sqrt{\alpha_j{}^2[\Psi'(\alpha_j)]^2-\Psi'(\alpha_j)-1}\right)\frac{\Gamma(5\phi_j)}{\left(\Gamma(\phi_j)\right)^5}\frac{\prod_{j=1}^{5} x_j{}^{\alpha_j}}{\sum_{j=1}^{5} x_j}$$

$$P_M(\lambda_j|\alpha_j,x_j) = gamma\left(\alpha_j,\sum_{j=1}^{5} x_j{}^{\alpha_j}\right) \qquad [6]$$

The posterior parameters are $\hat{\alpha}_j = \alpha_j$ and $\hat{\lambda}_j = \sum_{j=1}^{5} x_j{}^{\alpha_j}$

The Bayesian algorithm specified, we set specific priors for the parameters and determine the posterior estimates of the parameters. The posterior distribution can then be used to determine customer propensity to churn as well as predict future customer churn when competing risks are confounded with time-dependent covariates.

To analyze and evaluate the performance of the model, a churn dataset with 5-time dependent variables (age, marital status, occupation, education level, and residence) and 5 competing risks (network quality, service quality, price sensitivity, carrier responsiveness, and fraud) from November 2003 to July 2019 are used. The weekly churn rate for this period is used to determine the extent of customer churn in Safaricom PLC, Airtel Networks Limited, and Telkom Kenya Limited.

The population size for the study included all present and past active mobile subscriptions in Kenya. The sample size (6000) was calculated by using the stratified sampling technique as well as the random sampling method of Yamane, (1967). The latest national census data of 2019 was used as a sampling frame to identify the subscribers.

Secondary data was gathered through close-ended questionnaires to find out if subscribers churned based on; subscriber, residence, age group, marital status, occupation, and education. The customers were followed up for 830 weeks. The minimum follow-up time was 0 weeks and the maximum was 830 weeks.

III. Result

1128 subscribers (18.80%) churned during the follow-up time with 678 (60.11%), 956 (84.75%), and 1093 (96.90%) churns occurring within 207 weeks, 265 weeks, and 623 weeks of line activation, respectively. 70% (4200 respondents) of the data set, randomly selected, was used as the training data set, and 30% (1800 respondents) of the data set to analyze the performance of the model, as the testing data set to evaluate the performance of the model.

The gamma prior parameters were set to $\alpha = 5$ and $\lambda = 150$. Figure 1 shows that the distribution of churn is possibly the gamma distribution, with parameters $\alpha = 3.918$ and $\lambda = 159.36$.

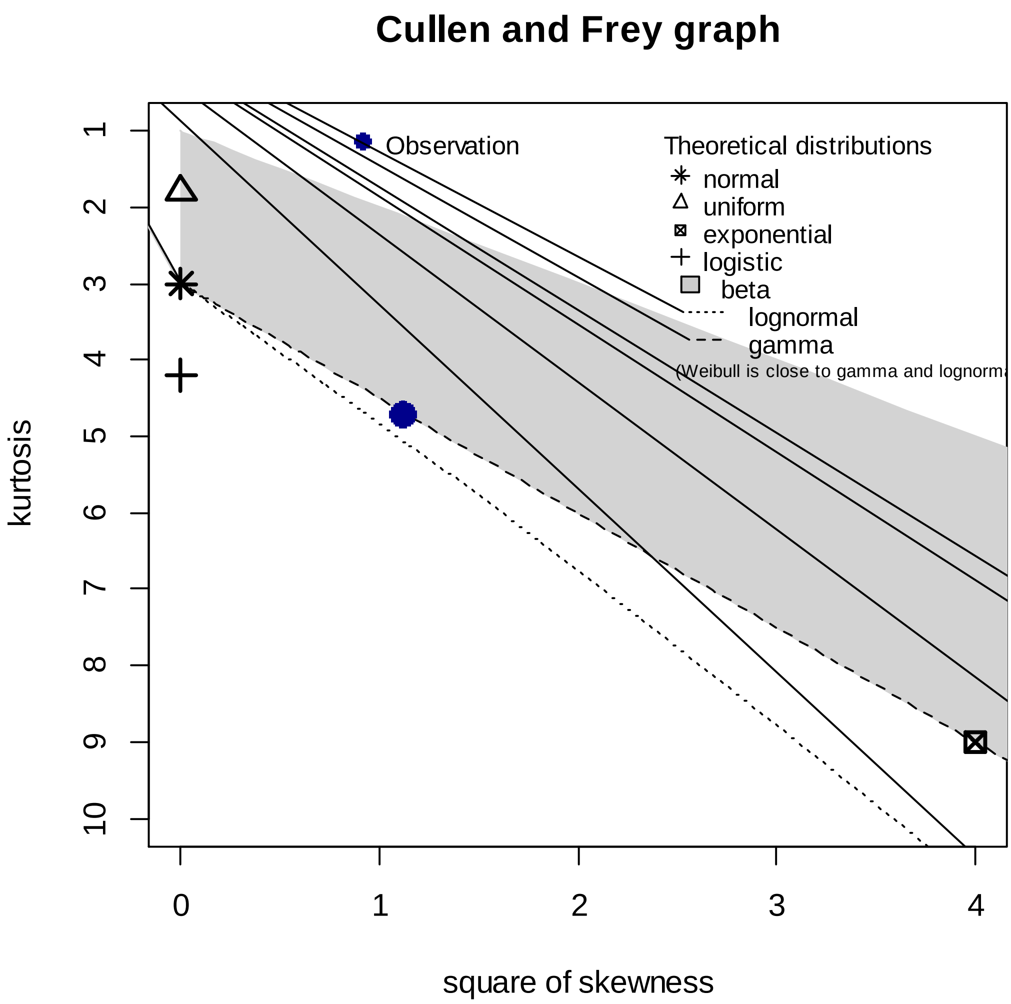

Figure 1 Culley and Frey Graph for the churn data

In addition, figure 2 fit for the gamma also shows that the churn data set could be well fit using a gamma distribution.

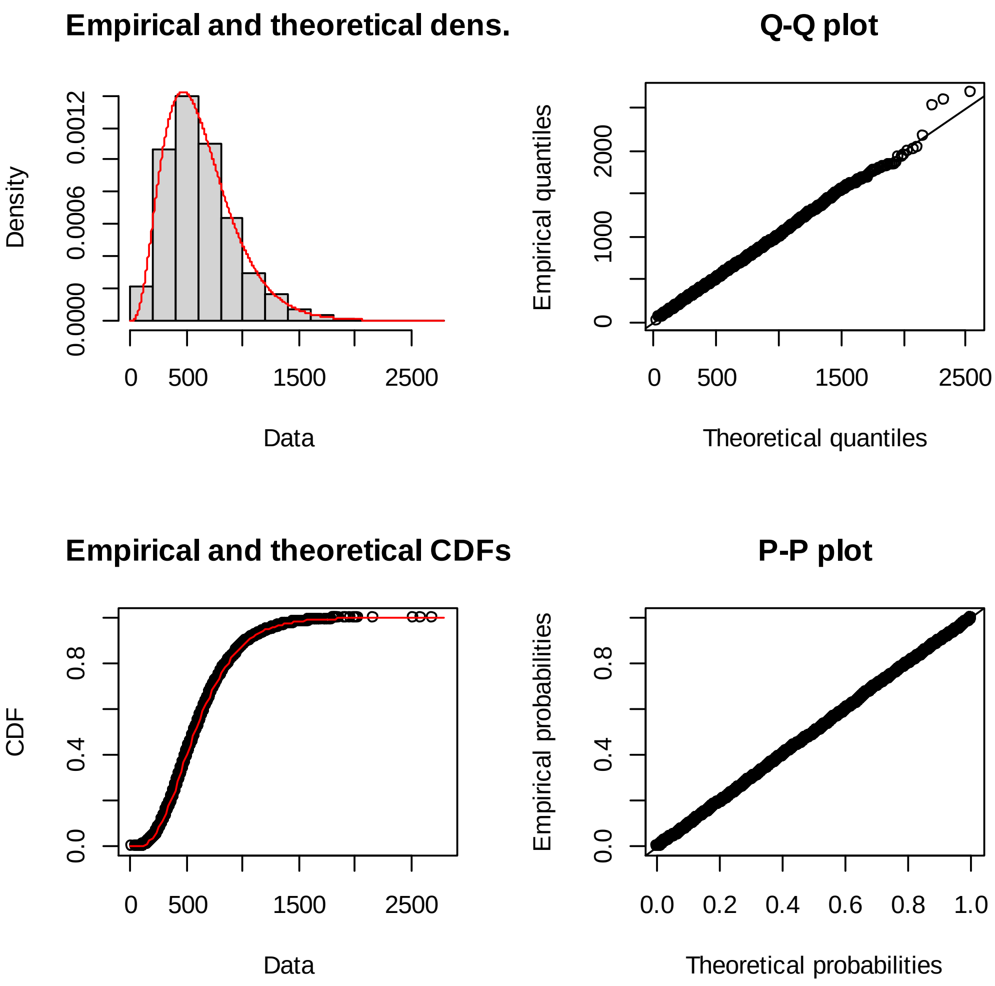

Figure 2 Fit for the churn data set

Using the Bayes algorithm, [5] and [6], we have the training dataset posterior $gamma(4.1379,154.0832)$. In Figure 3, the empirical and theoretical probability densities and the cumulative densities as well as the Q-Q and P-P plots of the fit for the training dataset posterior shows that $gamma(4.1379,154.0832)$ fits the training dataset well and can be used to model customer churn.

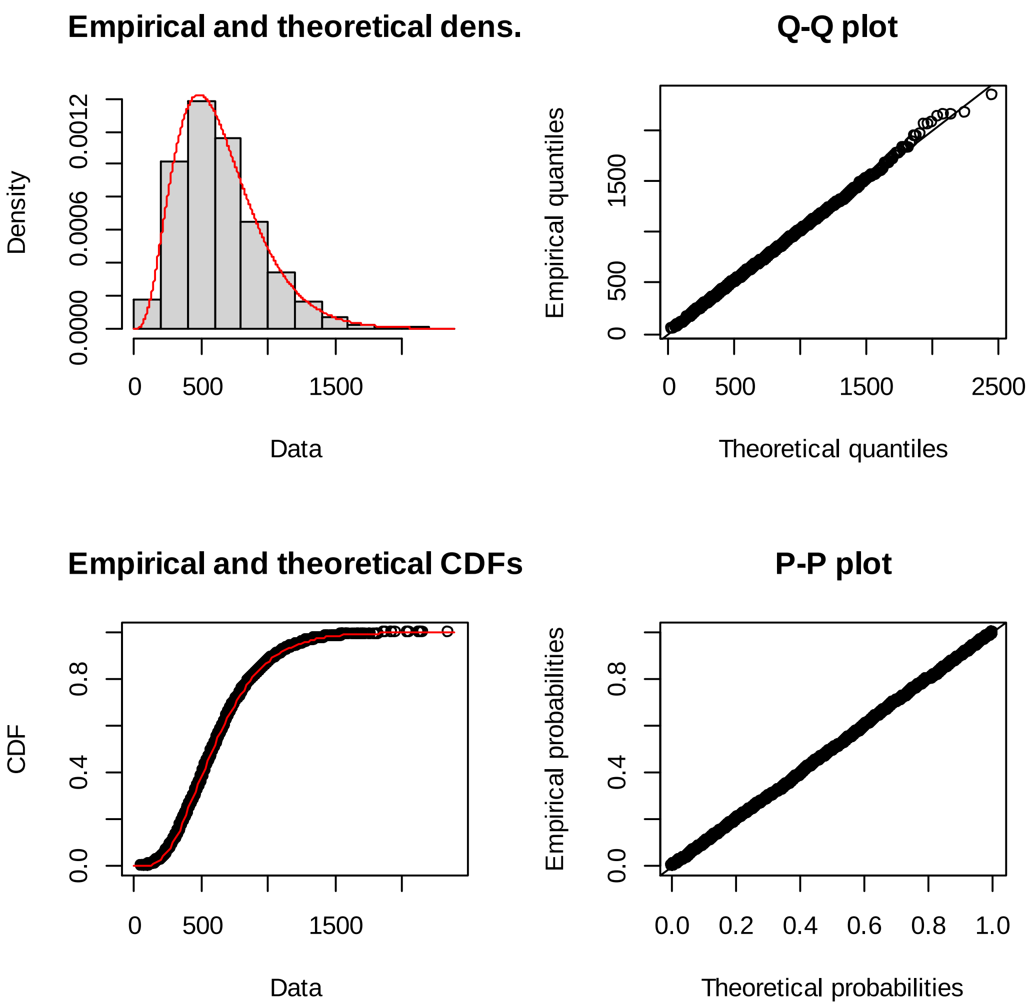

Figure 3 Fit for the training dataset posterior

The maximum likelihood estimates of the parameters α and λ are $\hat{\alpha}_{MLE} = 4.1379$ and $\hat{\lambda}_{MLE} = 154.0832$ with the corresponding asymptotic variances as 0.00603 and 1.5952×10^{-8}, respectively. Using these asymptotic variances, we obtain the 95% credible intervals for α and λ as $(4.0602, 4.2156)$ and $(151.14, 157.14)$, respectively.

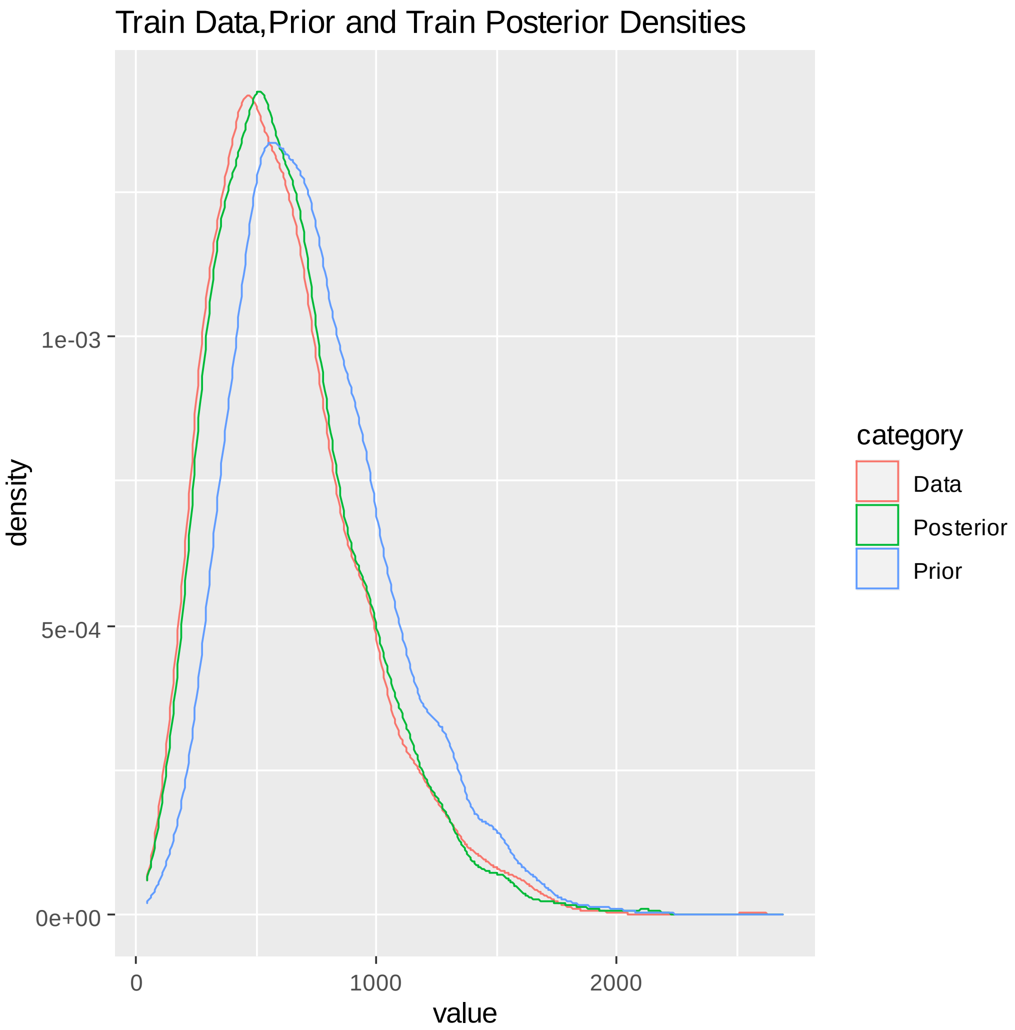

Figure 4 Comparison plots for the training data, Prior and Posterior

Figure 4 shows a comparison of the graphs of the training churn data set, prior and posterior. The graphs reveal that the Bayesian model developed is a best-fit model for churn when competing risks are confounded with time-dependent covariates.

To validate the performance of the model, we run the model on the test data. Figure 5 shows a comparison of the graphs of the test churn data set, prior and posterior.

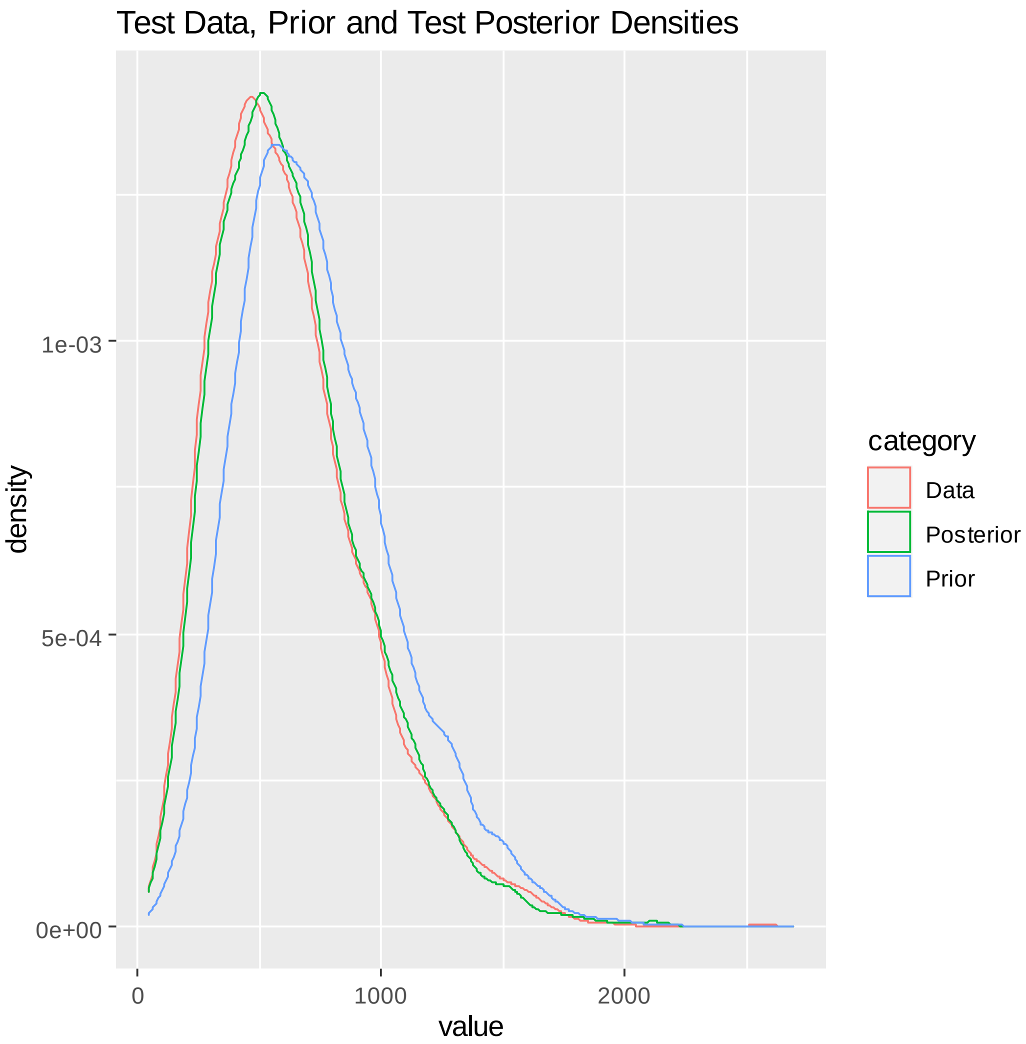

Figure 5 Comparison plots for the test data, Prior and Posterior

Table 1 shows that the Bayesian model is the best model for fitting customer churn when competing risks are confounded with time-dependent covariates since its log-likelihood test statistic is maximum and the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) test statistic is minimum.

Table 1 Test statistics for the churn data, Prior and Posterior

Test statistic	**Churn data**	**Prior**	**Posterior**
Log-likelihood	-42492	-42972	-42404
AIC	84989	85949	84811
BIC	85003	85962	84824

The model can also be used to predict the propensity of churn. Table 2, shows the probabilities of customer churn for network quality confounding time-dependent covariates.

Table 2 Customer churn for network quality at different times

Time (weeks)	Probability of network quality churn at the specified time
260	92.15%
520	59.28%
780	27.92%

The table shows that 92.15%, 59.28%, and 27.92% of network quality churns will occur within 5 years, 10 years, and 15 years of line activation. The model can also be used to determine the propensity of churn within a given interval, for example, the probability of network quality churn between 300 weeks and 450 weeks of line activation is 19.10%.

IV. Discussion

Customer churn in the mobile telecommunications industry when competing risks are confounded with time-dependent covariates can be modeled using a Bayesian model, a gamma prior, $f(x|\alpha,\lambda) = \frac{\lambda^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}exp(-\lambda x)$; $x > 0$, a penalized likelihood function given by $L_{p_j}(\alpha_j,\lambda_j|x_j) = \left|\sqrt{\alpha_j{}^2[\Psi'(\alpha_j)]^2 - \Psi'(\alpha_j) - 1}\right| \left(\prod_{j=1}^{5} x_j{}^{(\alpha_j-1)}\right) \alpha_j(\lambda_j)^{5\alpha_j-1}(\Gamma(\alpha_j))^{-5} exp\left(-\lambda_j \sum_{j=1}^{5} x_j\right)$ and a posterior density, $P_M(\lambda_j|\alpha_j,x_j) = gamma\left(\alpha_j, \sum_{j=1}^{5} x_j{}^{\alpha_j}\right)$.

The Gamma model was only used to predict the probabilities of customer churn for network quality, the model can also be used to determine the probabilities of customer churn for other competing risks.

The mobile phone service providers should worry more about network quality churn and be regressive in their customer retention efforts and campaign in the first 5 years of line activation of a customer and invest comparatively less to customers who have been with the operator for 15 years since line activation. This is because the probabilities in the earlier years of line activation are higher compared to the later years.

V. Conclusion

We assert that the Bayesian model (Gamma model), with a 95% credible interval, is a better model for predicting subscribers' survival probabilities when competing risks are confounded with time-dependent covariates since the log-likelihood test statistic is maximum and the AIC and BIC test statistics are minimum and the posterior fitted well both for the training data and the test data.

References

[1]. Brandusoiu, I. B. and Toderean, G. 2013: *Predicting Churn in Mobile Telecommunications Industry. ACTA Technica Napocensis Electronics and Telecommunications*, p. 54(3).

[2]. *Customer Churn Prediction & Prevention Model*. (2021, December 22). Optimove. Retrieved January 15, 2022, from https://www.optimove.com/resources/learning-center/customer-churn-prediction-and-prevention#:%7E:text=Churn%20prediction%20modeling%20techniques%20attempt%20to%20understand%20the,to%20the%20success%20of%20any%20proactive%20retention%20efforts.

[3]. Heckerman, D. 1997: Bayesian Networks for Data Mining. *Data Mining Knowledge Discovery*, p. 1.

[4]. Huang, Y., Huang, B., and Kechadi, M. T. 2011: A rule-based method for customer churn prediction in telecommunication services. In: *Advances in Knowledge Discovery and Data Mining*, Springer, p. 411–422.

[5]. Idris, A. and Khan, A. 2012: Customer churn prediction for telecommunication: Employing various features selection techniques and tree-based ensemble classifiers. In: *Multitopic Conference (INMIC), 2012 15th International*, p. 23–27.

[6]. Jeffreys, H. 1946: An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London.* Series A, Mathematical and physical sciences. 186 (1007): 453 – 461. doi:10.1098/rspa.1946.0056.

[7]. Pearl, J. 1988: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

[8]. Shyam, V. N. 2003: Customer churn analysis in the wireless industry: A data mining approach. In: *Proceedings-annual meeting of the decision sciences institute*, p. *561*.

[9]. Yamane, T. 1967: *Statistics: An Introductory Analysis*, 2nd Edition, New York: Harper and Row.

Mr. Ndilo Benjamen Fwaru. "Bayesian Modelling Of Churn Confounding Competing Risks With Time-Dependent Covariates Among The Mobile Phone Subscribers In Kenya." *IOSR Journal of Mathematics (IOSR-JM)*, 19(2), (2023): pp. 46-54.