MIE 1516-Project Report

Predicting Non-ambulatory Patient Arrival in Emergency Department(ED) Using Probabilistic Graphical Models

Tahera Yesmin
999049275

April 13, 2020

University of Toronto

Contents

1 Introduction

Prolonged patient waiting times in the Emergency Department (ED) considered a global crisis [1]. In order to alleviate the ED crowding and plan accordingly, various methods have been used to predict the patient arrivals in the ED such as. Queuing methods, forecasting techniques and many more [2]. In the ED, patients are usually classified as (i) ambulatory patients and (ii) non-ambulatory patients. Ambulatory patients are the less severe and can move without support; whereas non-ambulatory patients are more severe patients, not able to walk without support, and who usually requires a quick response in treatment. When a hospital operates close to capacity, specifically during the surge, it is very important to allocate the resources carefully. Therefore, prior prediction on non-ambulatory patients will support for decisions, aware the staff at the ED on their workload and thus will help to reduce the wait time.

2 Objectives

The primary objectives of this project are:

(i) To predict the non-ambulatory patients' arrival rate, using probabilistic graphical models.

(ii) Build several models and evaluate them to identify the best performing one.

3 Data

Arrival rate of ED patients is characterized by daily, weekly and seasonal variations, a degree of inherent unpredictability, and the arriving patients' characteristics such as, age, gender, chief complaints[2]. Hourly information of the number of patients' arrival and the mean age of the patients arrived per hour will be used in this project. Non-ambulatory patient arrival information from September 01, 2017 to November 02, 2017 and their age from one of the hospitals of Ontario,Canada was used in this study. Table 1 reports the descriptive analysis of the variables used in the models.

Table 1: Descriptive analysis of the data

Variables	Mean	Standard deviation	Min	Max
No of patients arrived per hour	2.60	1.86	0 (as no patients arrived at that hour)	10
Mean age of the patients/hour(years)	51.50	26.78	0 (as no patients arrived at that hour)	99.0

Figure 1 shows the total number of patients' arrival per hour in the ED. As illustrated in the figure (Figure 1), the number of patient arrival varies based on the hour of the day. Clearly, the number of patients arrived is lower during 12 am-8 am, then it starts to increase and reach a pick during 10 am-1:00 pm. Afterward there is a decrease during 4pm -6pm, followed by an increase again at 7:00 pm.

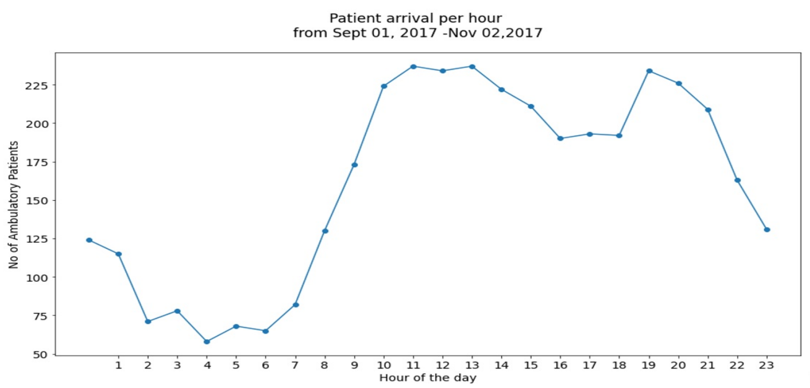


Figure 1: Patient arrival per hour

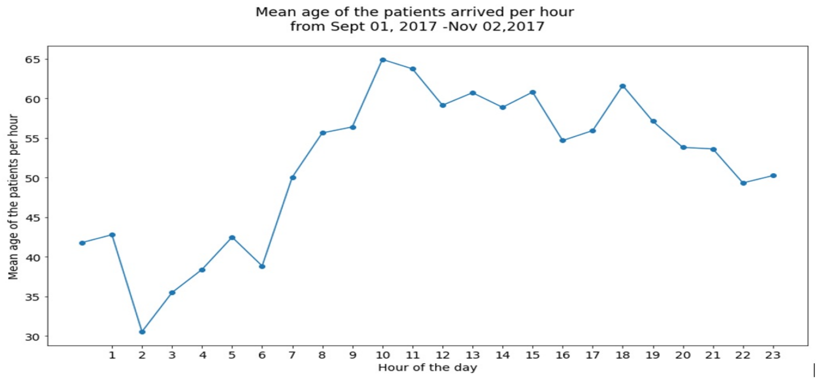


Figure 2: Mean age of the patients arrived per hour

Similarly, Figure 2 illustrates the variation in the mean age of the total number of patients' arrival during different hour of the day. Based on this figure, comparatively older patients come to ED during 10 am to 2:00 pm. The younger patients come for treatments in the later part of the day (10pm- 6am).

4 Method

Probabilistic graphical models (PGM) will be used in this study to construct the prediction models. PGM is the compact specification of joint probability over random variables in a compact way by exploiting the dependencies between them [3]. There are two major types of Graphical Models: Bayesian Networks and Markov Networks.

In this study, a Bayesian Network will be used to construct the models. A Bayesian Network consists of a directed cyclic graph(DAG) and a conditional probability distribution associated with each of the variables [3]. Later, PyMC3 will be used to do the Bayesian statistical inference and for the prediction [4]. The following figure (Figure 3) shows the DAG to illustrate the relationships between the variables of this study.

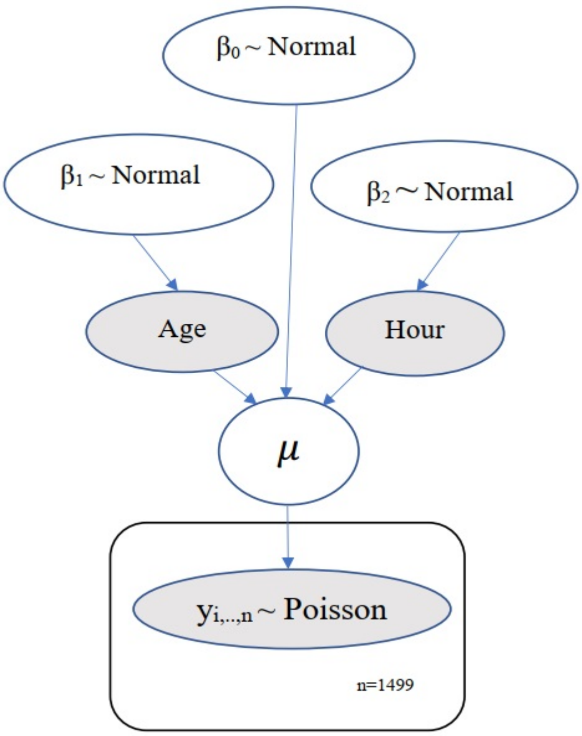


Figure 3: Model to predict patient arrival

In this study, to predict the outcome Y which is patient arrival, it has been considered that Y has a Poisson distribution with an expected value μ [5], which is the exponential function of θ. Whereas, θ is the linear function of the variable "Age" and/or "Hour".

$$Y(patient - arrival) \sim Poisson(\mu)$$
$$\mu = exp(\theta)$$
$$\theta = \beta_0 + \beta_i X_i$$

Here, β_0 is the intercept, β_i are the coefficients of the co-variates, Xi. Since, this study involves Bayesian model, a prior distribution was assigned to each of the unknown variables in the model, which are β_0, β_i. It has been considered that all the linear regression coefficients are normally distributed with mean 0 and a specific standard deviation (sd) for each of the variable.

$$\beta_0(intercept) \sim N(0, sd)$$
$$\beta_i(covariate) \sim N(0, sd)$$

4.1 Models

Three different models have been constructed in this study. The first model was built involving both the variables (Age and Hour), second model was built involving only the variable Age and third model was built with only the variable Hour.

4.1.1 Model-1

Model-1 involves both variables. Therefore, θ is a function of both Age and Hour.β_1 is the coefficient of the co-variate X_1(Age) and β_2 is the coefficient of the co-variate X_2(Hour). Mode-1 architecture is,

$$Y(patient - arrival) \sim Poisson(\mu)$$
$$\mu = exp(\theta)$$
$$\theta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$
$$\beta_0(intercept)(0, 10)$$
$$\beta_1(Age) \sim N(0, 20)$$
$$\beta_2(Hour) \sim N(0, 5)$$

4.1.2 Model-2

In Model-2, θ is a function of the variable Age only. Below is the architecture of Mode-2:

$$Y(patient - arrival) \sim Poisson(\mu)$$
$$\mu = exp(\theta)$$
$$\theta = \beta_0 + \beta_1 X_1; X1 = Age$$
$$\beta_0(intercept) \sim N(0, 10)$$
$$\beta_1 X_1(Age) \sim N(0, 20)$$

4.1.3 Model-3

Similarly, θ is a function the variable Hour, in Model-3. Therefore, Model-3 architecture is:

$$Y(patient - arrival) \sim Poisson(\mu)$$
$$\mu = exp(\theta)$$
$$\theta = \beta_0 + \beta_2 X_2; X2 = Hour$$
$$\beta_0(intercept) \sim N(0, 10)$$
$$\beta_2(Hour) \sim N(0, 5)$$

5 Result

After constructing the models, posterior estimates for the unknown variables(β_0,β_1, β_2) were obtained for each of the model respectively using PyMC3. To obtain that, 2000 samples for 2 chains were drawn from the posterior distribution using Markov Chain Monte Carlo (MCMC) sampling methods. The following section reports the posterior analysis of the models.

5.1 Posterior Analysis

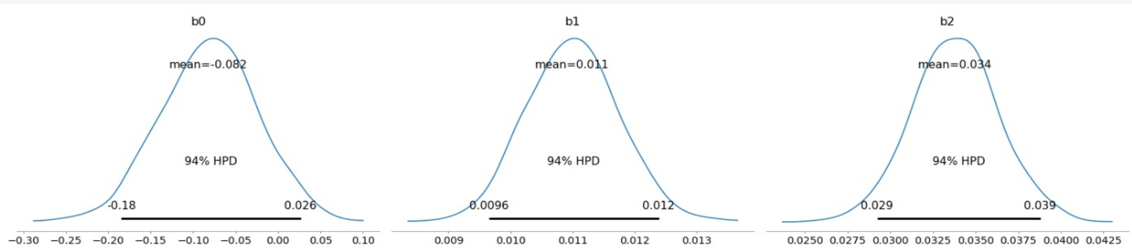
With the help of the traceplot function of PyMC3 and the package Arviz [4], the summary of the posterior analysis and the trace plots of models have been obtained.

5.1.1 Model-1

Figure 4a, shows the mean values of the β_0, β_1, β_2 and the trace plots (Figure 4b) show the Highest-Posterior Density (HPD) interval for Model-1 . An HPD is the shortest interval containing a given portion of the probability density. From Figure 4a, it has been reported that the intercept (β_0) is negative for Model-1, where as the mean of β_1 is 0.011 and β_2 is 0.034. Trace plots also show the smoothed histogram (using kernel density estimation) of the marginal posteriors of each stochastic random variable, samples of the the Markov chain plotted in sequential order for each of the chains, and the pair-plots of the unknown variables. These plots are shown in the Appendix.

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
b0	-0.082	0.057	-0.185	0.026	0.002	0.001	1346.0	1207.0	1354.0	1283.0	1.0
b1	0.011	0.001	0.010	0.012	0.000	0.000	1333.0	1315.0	1344.0	1142.0	1.0
b2	0.034	0.003	0.029	0.039	0.000	0.000	1818.0	1818.0	1820.0	1455.0	1.0

(a) Summary of the posteriors



(b) Highest-Posterior Density (HPD) interval of the posterior

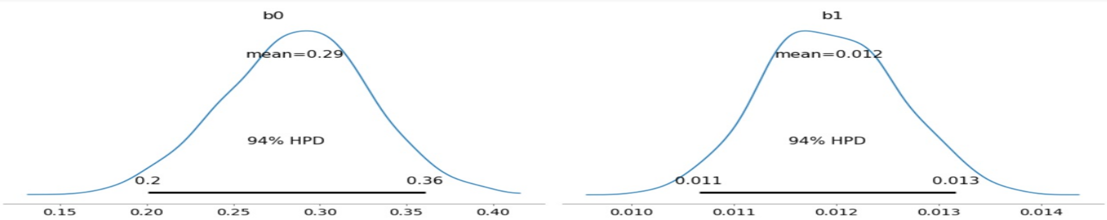
Figure 4: Posterior analysis of Model-1

5.1.2 Model-2

Figure 5, shows the results obtained from Model-2 for the co-efficients β_0 and β_1. The mean value of β_0 in 0.286, and for β_1 is 0.012. Figure 5b, shows the HPD interval of the unknown variables from this model. Similar to Model-1, the pair plot (β_0 and β_1) and additional posterior trace plots are provided in Appendix for Model-2, as well.

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
b0	0.286	0.043	0.201	0.361	0.002	0.001	560.0	552.0	558.0	752.0	1.01
b1	0.012	0.001	0.011	0.013	0.000	0.000	572.0	572.0	570.0	865.0	1.01

(a) Summary of the posteriors



(b) Highest-Posterior Density (HPD) interval of the posterior

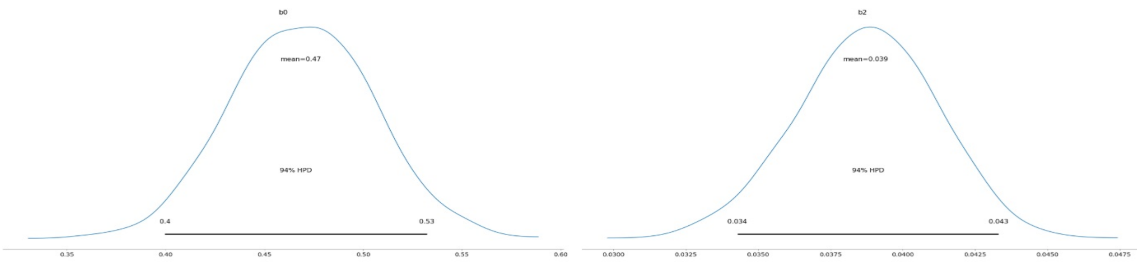
Figure 5: Posterior analysis of Model-2

5.1.3 Model-3

The results obtained from Model-3 show (Figure 6) that the mean value for β_0 is 0.470 and β_2 is 0.039. The pair plot of β_0 and β_2 and other additional posterior trace plots are provided in Appendix.

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
b0	0.470	0.035	0.405	0.539	0.001	0.001	988.0	988.0	988.0	1131.0	1.0
b2	0.039	0.002	0.034	0.043	0.000	0.000	950.0	947.0	945.0	1083.0	1.0

(a) Summary of the posteriors



(b) Highest-Posterior Density (HPD) interval of the posterior

Figure 6: Posterior analysis of Model-3

6 Model Comparisons

6.1 Log-likelihood

The likelihood of any model and its parameters given the data set is equal to the probability of the data set given the model and its parameters.The number of patient-arrivals arrival per hour (Y) can be expressed as the bellow:

$$Y = \{y_1, y_2, ..., y_N\}$$

Here, $y_1, y_2, ..., y_N$ = number of patient arrival in each hour.
So, the total likelihood is the product of the likelihood for each point,

$$l(Y(data) \mid \Theta) = \prod_{i=1}^{N} l(y_i \mid \Theta)$$

Here, Θ is the model parameters0ypothesis, and so the log-likelihood is:

$$\log l(Y(data) \mid \Theta) = \sum_{i=1}^{N} \log l(y_i \mid \Theta)$$

Since the log-likelihood increases with the increase of samples, the mean of the log-likelihood has been shown in the figure (Figure 7) below. Figure 7, shows that Model-1, has the highest value of the log-likelihood than the other two models, which implies, Model-1 is the best fit.

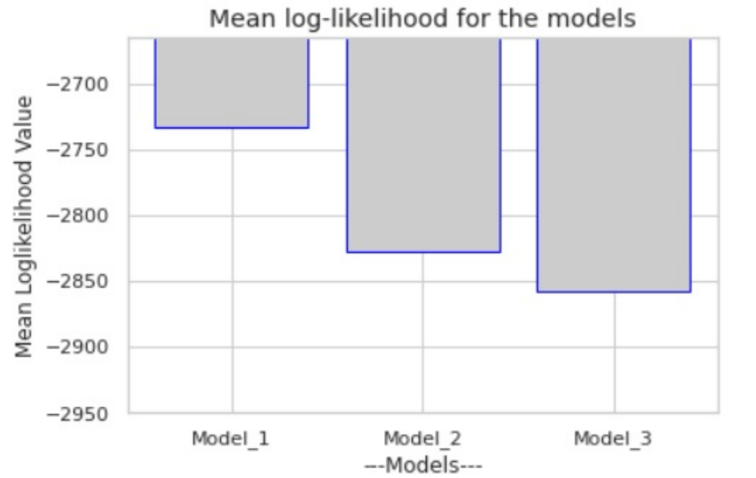


Figure 7: Model evaluation with log-likelihood

6.2 Mean Absolute Error(MAE)

MAE had been calculated for each of the models. The error calculated using the below formula where, predicted values of patient-arrivals per hour(y_{pred}) had been just obtained for 1499 patient-arrivals.

$$MAE = (\frac{1}{n}) \sum_{i=1}^{n} |y_i - y_{pred_i}|$$

Here, y_i = number of patient arrival in each hour from the data.
Here, y_{predi} = predicted number of patient arrival in each hour.

As Figure 8 shows, that Model-1 has the MAE of 1.77, whereas Model-2 and Model-3 has 1.82, 1.83 respectively. This analysis also shows that Model-1 fits the data best.

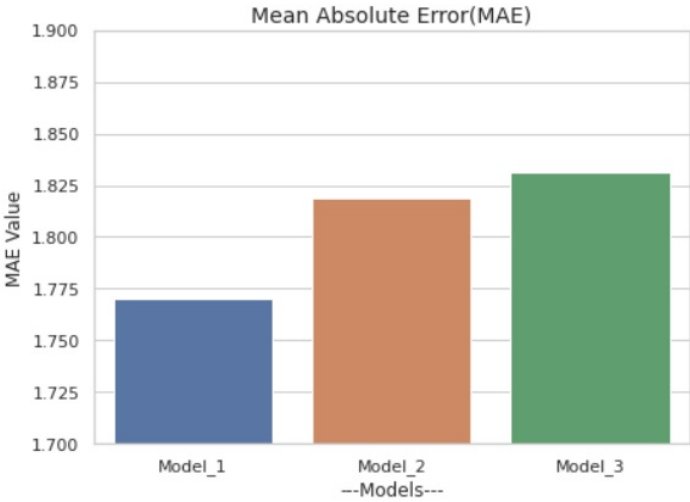


Figure 8: Model evaluation with mean absolute error

6.3 Widely Applicable Information Criterion (WAIC)

The next matrix applied to compare the models is, WAIC [6]. WAIC has two terms: one that measures how well the data fits the model and one penalizing complex models. Figure 9 compares the three models of this study based on the WAIC.

	rank	waic	p_waic	d_waic	weight	se	dse	warning	waic_scale
model_1	0	-2725.4	3.19968	0	1	31.9395	0	False	log
model_2	1	-2821.97	2.55215	96.5676	1.76862e-25	31.8513	13.7736	False	log
model_3	2	-2852.89	2.29136	127.484	1.63136e-39	31.756	17.3309	False	log

Figure 9: Model evaluation with WAIC

Based on Figure 9,

(i) The first column ranks the models, which shows that Model-1 performs better than the other two models.
(ii) Second column contains the values of the WAIC.
(iii) The third column is the estimated effective number of parameters. Models with higher number of parameters are flexible to fit the data but can be overfitted too.Thus, p-waic is a penalization term, which measures the flexibility of each model in fitting the data.
(iv) d-waic is the relative difference between the value of WAIC for the top-ranked model and the value of WAIC for each model.
(v) Weight can be interpreted as the probability of each model (among the compared models) given the data.
(vi) The sixth column is the standard error for the WAIC computations.
(vii) The seventh column is the standard error of the differences between the values of the WAIC of the models.
(viii) Finally, last column named “warning”, which is false here.

6.4 Posterior Predictive Checks of the Models

Once the posterior is computed, it is possible to use the posterior, $p(\Theta \mid Y)$,to generate predictions, y_{pred}, based on the data (Y) and the estimated parameters (Θ) [7]. In this section, the mean fit for all the three models have been plotted. Figure 10 shows that in most of the time Model-1 performs better than the other two models. Though Model-2 performs in a similar manner of Model-1 but Model-3 did not fit the data well.

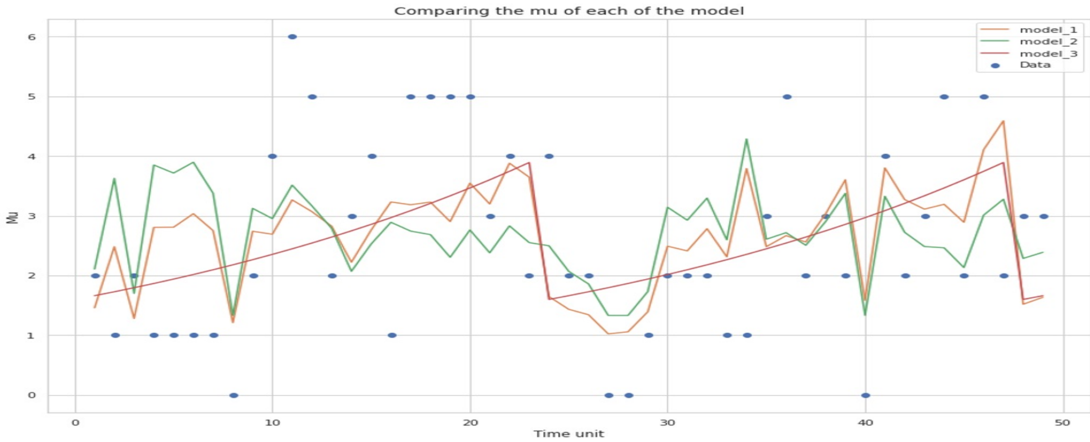


Figure 10: Posterior predictive checks of the models

7 Conclusion

Probabilistic graphical models (Bayesian network), have been used in this study to predict the patient arrival per hour for non-ambulatory patients in an emergency department of a hospital. Three different models have been constructed. Model-1 predicts the patient arrival based on the hour of the day and mean age of the patients of that hour. Model-2 predicts the patient arrival using only the age information, whereas Model-3 uses only the hour of the day. It has been considered here that, Patient arrival follows a Poisson distribution whose rate depends on the linear regression of the variables(age and/or hour), The priors of the unknown variables (β_0,β_1, β_2) are considered normally distributed with mean 0. After constructing the model, posterior estimates of the co-efficients (β_0,β_1, β_2) were obtained for each of the model respectively using PyMC3 using Markov Chain Monte Carlo (MCMC) sampling methods. After that, each of the model was evaluated by matrices such as log-likelihood, mean absolute error, WAIC and posterior predictive checks and has been reported that Model-1 involving both the variables age and hour, performs better than the other two models. Below figure (Figure 11) shows the actual number of patient arrival in hour 1 to 50 with the predicted number of patient arrival from Model-1.

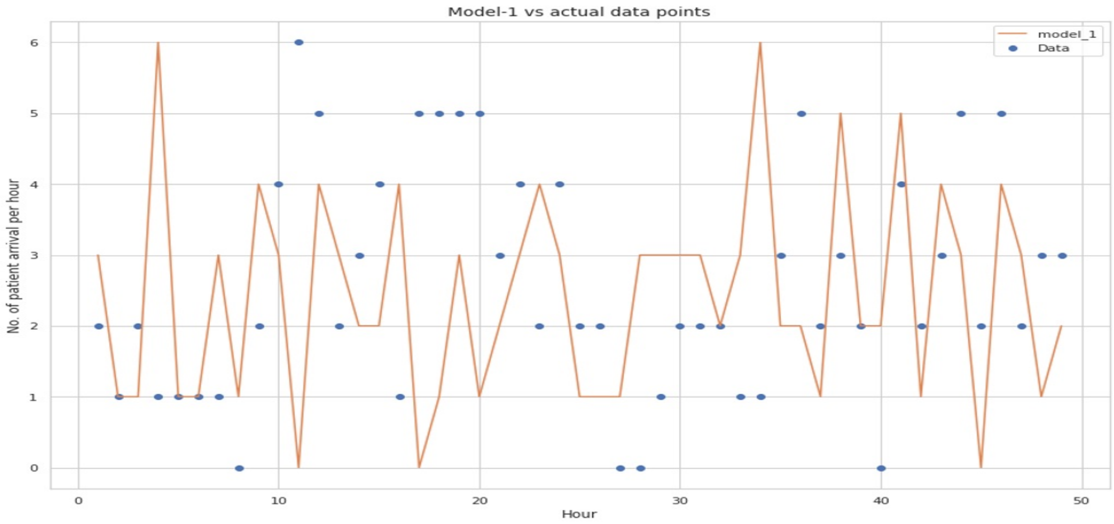


Figure 11: Actual patient arrival data vs predictions from Model-1

Though, Model-1 performs better, however, involving other information such as month, day of the week, gender might have improved the prediction further.

8 References

[1] M. L. McCarthy, S. L. Zeger, R. Ding, D. Aronsky, N. R. Hoot, and G. D. Kelen, "The challenge of predicting demand for emergency department services," *Academic Emergency Medicine*, vol. 15, no. 4, pp. 337–346, 2008.

[2] S. Au-Yeung, U. Harder, E. McCoy, and W. Knottenbelt, "Predicting patient arrivals to an accident and emergency department," *Emergency medicine journal*, vol. 26, no. 4, pp. 241–244, 2009.

[3] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

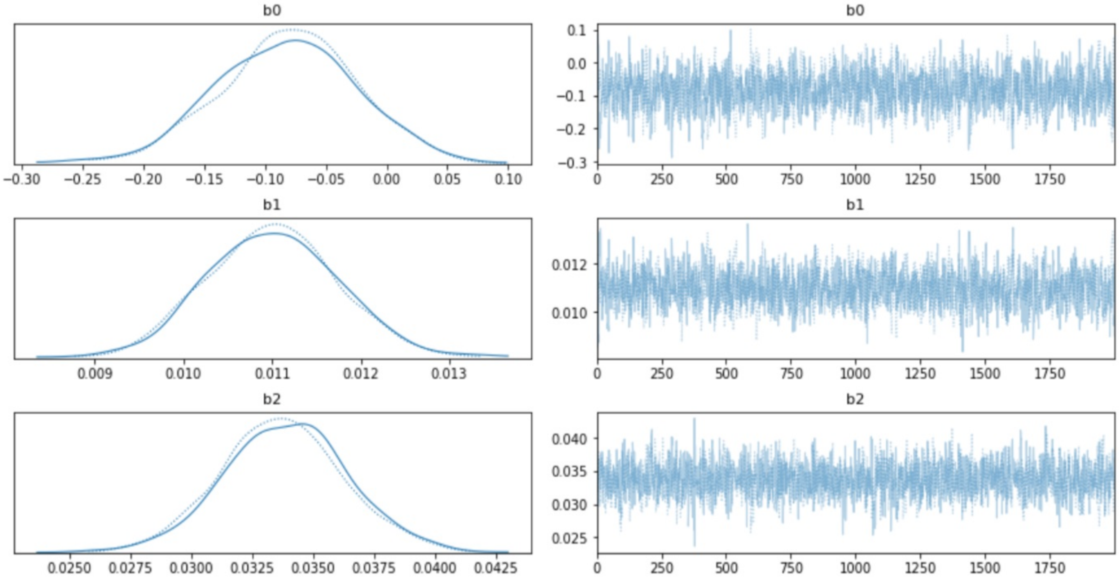
[4] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3," *PeerJ Computer Science*, vol. 2, p. e55, 2016.

[5] S. Chib and R. Winkelmann, "Markov chain monte carlo analysis of correlated count data," *Journal of Business & Economic Statistics*, vol. 19, no. 4, pp. 428–435, 2001.

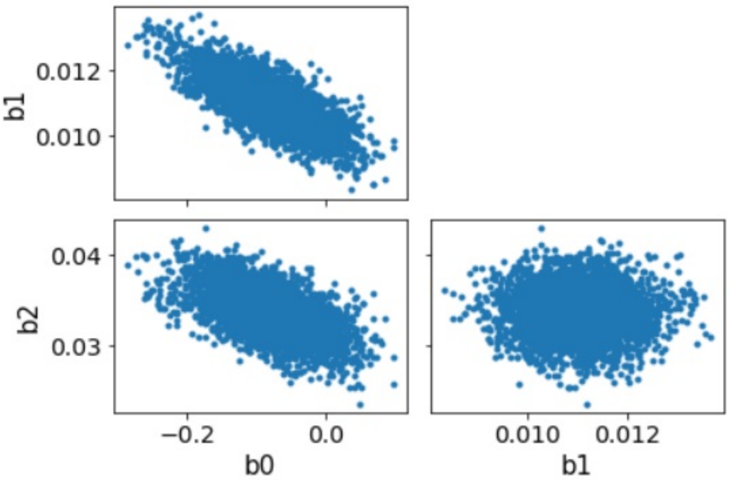
[6] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," *Statistics and computing*, vol. 27, no. 5, pp. 1413–1432, 2017.

[7] O. Martin, *Bayesian analysis with python.* Packt Publishing Ltd, 2016.

9 Appendix

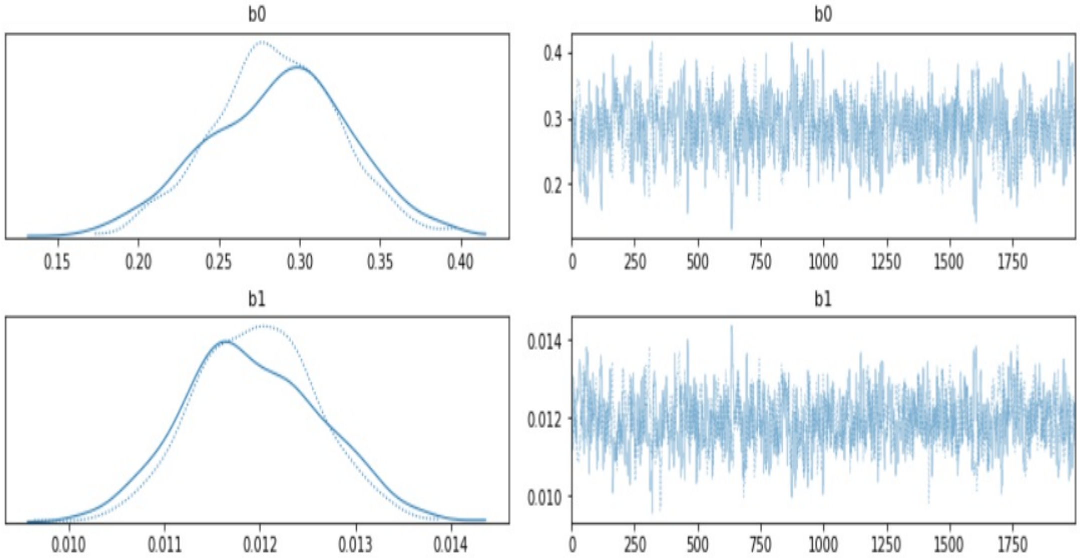


(a) Trace plots the posteriors

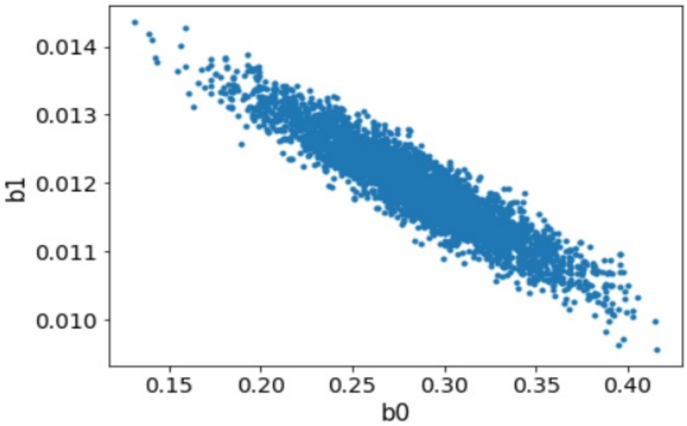


(b) Pair plots of the co-efficient

Figure 12: Additional posterior analysis of Model-1

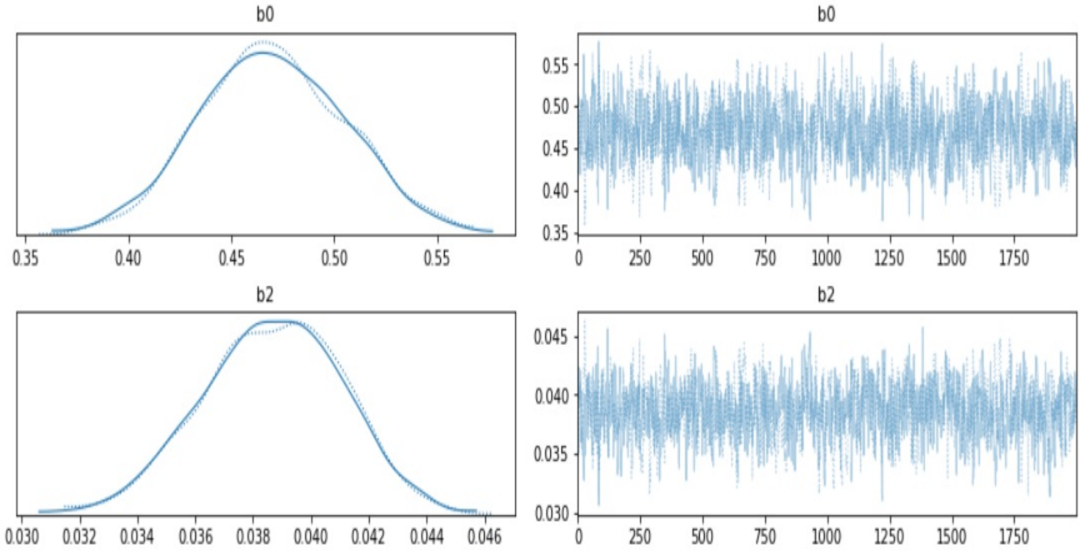


(a) Trace plots the posteriors

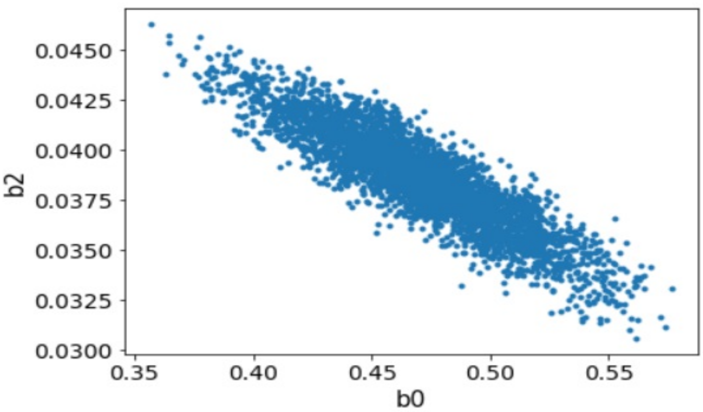


(b) Pair plots of the co-efficient

Figure 13: Additional posterior analysis of Model-2



(a) Trace plots the posteriors



(b) Pair plots of the co-efficient

Figure 14: Additional posterior analysis of Model-3