A new link function for the prediction of binary variables

Gloria Gheno
Free University of Bolzano-Bozen, Italy
gloriagheno@libero.it

Abstract

If there are no heavy sanctions in place to prevent it, the problem of the cancellation of appointments can lead to huge economic losses and can have a significant impact on underutilized resources of healthcare facilities. A good model to predict the appointment cancellations could be an effective solution to this problem. Therefore, a new Bayesian method is proposed to estimate accurately the probability of the cancellation of visits to healthcare institutions based on specific factors such as age. This model uses the regression for binary variables, linking the explanatory variables to the probability of appearance at a previously made appointment with a new weighted function and estimating the parameters with the Bayesian method. The goodness of the new method is demonstrated by applying it to a real case and by comparing it to other methodologies. Therefore, the advantages of the proposed method are exposed and possible real-world applications are described.

Keywords: Bayesian method, binary variables, cancellation prediction, heath care, link function.

JEL classification: C11, C50, C51, I11.
DOI:10.2478/crebss-2018-0014

Received: June 12, 2018
Accepted: October 25, 2018

Introduction

A problem in the health sector is represented by too many cancellations and by the absence of the patient to the medical visits, which cause huge economic losses to the organization and the underutilization of the structure. Often, indeed, the cancellation is too close to the date of the appointment, which enables clinics to replace the patient (Alaeddini et al., 2015). One possibility to solve this problem is open access, which schedules the appointment on the same day of the call regardless of the reason for the visit (O'Hare, Corlett, 2004). Another possibility is overbooking, which establishes an additional number of patients every day based on the rate of cancellations or absences recorded in the organization (Daggy et al., 2010, LaGanga, Lawrence, 2007). According to Lee et al. (2013), overbooking performs better than open access in general. However, overbooking leads to patient dissatisfaction, caused by the increase in the waiting time, and to the increment of the costs for clinic overtime. The choice of the number of appointments to be added in excess and their most profitable temporal collocation are essential to optimize the trade-off between the number of additional patients and the consequent increase in the waiting times and in the costs. LaGanga and Lawrence (2007), starting from the rate of presence of patients, analyse various overbooking policies so as to decide how many appointments have to be inserted and how they have to be included in the program of visits, considering the waiting time and the reduction of the costs. Their analysis points out that the "interappointment times" method works just as well as other methods, which are more complex but more attractive for the clinics interested in minimizing overtime to the detriment of patient waiting times. However, as evidenced by Daggy et al. (2010), although some studies have examined the factors, which lead to the cancellation of the visit or to the absence of the patient, few introduce the probabilities, so obtained, in the system for planning clinical visits so as to improve them further. For example, LaGanga and Lawrence's paper considers the average rate of presence in the organization but without considering the peculiarities of the patients. Instead, an example of improvement of the system for planning clinical visits linked to the personal characteristics of patients is the work proposed by Samorani and LaGanga (2015). Indeed, they study the optimal overbooking using the prediction of an absence based on the individual characteristics of the patient and on the date of the appointment. Therefore, an accurate prediction of no-show and cancellation probability becomes fundamental for any scheduling system (Alaeddini et al., 2015) and for this reason, this paper proposes a new method to predict the probability of presence based on the age of the person, one of the individual factors, which mainly influences it. The factors, which influence the no-show rate, vary across clinics (Kopach et al. 2006), but if their and other studies are considered, age remains always one of the fundamental factors. This new method, improving the estimation of the rate of presence, improves the scheduling methods making them more accurate.

New link function and parameters estimation

When a researcher analyses a dichotomous response variable Y, i.e. which can take only the values 0 or 1, he assumes that it is distributed as a Bernoulli with probability p of having success, therefore he will have $Y = 1$ with probability p. In general, the best way to understand and to estimate p is to assume that it is a function, called link, of other variables, defined as explanatory. In traditional models the function, which links the explanatory variable X to the probability p, is monotone. The logit link function is the most used in many fields of applied statistics because the interpretation of its parameters is simple and straightforward (Ntzoufras, 2008). When the value x_i of the explanatory variable X is observed, the probability of having a success is given by the following logit function:

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}, \qquad (1)$$

where the subscript i defines the i-th observation and the parameter β determines if the probability is increasing or decreasing with respect to the explanatory variable X. If the parameter β is positive, the function is increasing, if it is negative, the function is decreasing. In this paper a new function, which links the probability to the variable age in a non-monotone way, is proposed because many empirical analyses of data find that the trend, based on age, of the probability of being present at the clinical appointment is not monotone, but it follows an irregular trend (Davies et al., 2016, Gebhart, 2017, Chua, Chow, 2018). This non-monotone function is the following

$$p_i = \underbrace{\left|\frac{\gamma x_i}{1+\gamma x_i}\right|}_{w_i} \underbrace{e^{-|\alpha+\beta x_i|}}_{f_1} + \underbrace{\left|\frac{1}{1+\gamma x_i}\right|}_{1-w_i} \underbrace{\left(\frac{2x_i+\omega}{x_i{}^2+\tau+1}\right)^2}_{f_2}, \tag{2}$$

with the parameters γ, ω between 0 and 1 and τ greater than 0 and with the constraint that p is between 0 and 1, including the extremes. The weight W depends on the explanatory variable X and therefore w_i is the weight associated with the observed value x_i. The weight determines the importance of the function 1 (f_1) and of the function 2 (f_2) and with the increase in the variable X the weight of the first function increases ($\partial W/\partial X>0$) and that of the second decreases. For example, if x_i=0 is observed, the probability p_i is exactly f_2.

In the case where the variable X is always greater than or equal to 0, the relation (2) can be simplified by removing some absolute values:

$$p_i = \underbrace{\left(\frac{\gamma x_i}{1+\gamma x_i}\right)}_{w_i} \underbrace{e^{-|\alpha+\beta x_i|}}_{f_1} + \underbrace{\left(\frac{1}{1+\gamma x_i}\right)}_{1-w_i} \underbrace{\left(\frac{2x_i+\omega}{x_i{}^2+\tau+1}\right)^2}_{f_2}. \tag{3}$$

Both functions are not monotone. For example, if the parameter β is greater than 0 and the parameter α is less than 0, the function f_1 is increasing for the values of the regressor X between 0 and -α/β and decreasing for the values greater than -α/β. The function f_2 is increasing for the values of the regressor X between 0 and g, while it is decreasing for the values greater than g. Therefore, the value g results as given in (4):

$$g = \frac{-\omega+\sqrt{\omega^2+4(\tau+1)}}{2}. \tag{4}$$

Estimation method

The Bayesian method, which considers the parameters as random variables, is recommended to estimate the parameters of formula (3). Similarly, to the classic Bayesian model, a prior distribution is assigned for the parameters. From the data, the researcher obtains a posterior distribution, from which it is possible to calculate the average, which represents one of the punctual estimations of the parameter, together with the mode and the median. A very simple example is used to remember the Bayesian operating method. If the researchers want to estimate the probability p directly from the observed data of the variable Y, without considering any explanatory variable X, generally they assume that Y has a Bernoulli distribution with parameter p and that this latter follows the Beta distribution with parameters (a,b), where a and b are selected by them on the basis of their knowledge and on the basis of readings of previous studies. Then, in Bayesian theory the Beta distribution (a,b) becomes a prior distribution. To find an estimation of the parameter p, it is necessary to calculate the posterior distribution, which is proportional to the product of the likelihood function and the prior distribution. Under the assumption of having I_s observations y_i of the dichotomous variable Y, the posterior distribution of p is proportional to:

$$\underbrace{\prod_{i=1}^{I_s} p^{y_i}(1-p)^{1-y_i}}_{likelihood\ function} \underbrace{\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}}_{prior\ distribution}, \tag{5}$$

and then it is equal to:

$$\frac{p^{\sum_{i=1}^{I_s} y_i+a-1}(1-p)^{I_s-\sum_{i=1}^{I_s} y_i+b-1}}{B(\sum_{i=1}^{I_s} y_i+a, I_s-\sum_{i=1}^{I_s} y_i+b)} = Beta(\underbrace{\sum_{i=1}^{I_s} y_i + a}_{a^1}, \underbrace{I_s - \sum_{i=1}^{I_s} y_i + b}_{b^1}). \tag{6}$$

In general, the punctual estimation of the parameter is set equal to average, to median or to mode. The prior distribution and the likelihood are conjugated if the posterior distribution has the same form as the prior distribution (Gill, 2002). In general, the researchers can find exactly the posterior distribution analytically only if the conjugacy subsists. When the posterior distribution is difficult or impossible to handle analytically, the Monte Carlo method is among those advisable, in particular in its form called Markov Chain Monte Carlo (MCMC), also implemented by Winbugs software (Gill, 2002, Carlin, Louis, 2008, Kéry, 2010). This methodology creates a chain, sequentially sampling the parameter values from a stationary distribution, which corresponds to the posterior joint distribution of interest (Carlin, Louis, 2008). The state of the chain after a large number of iterations ($2n$ = number of iterations) is used as a sample of the searched distribution.

A problem of MCMC is the number of iterations necessary to achieve the convergence to the stationary distribution. Therefore, one of the essential analyses is to test that the values come from a stationary distribution, i.e. that the Markov chain is convergent (Kéry, 2010). In the empirical analyses, one of the most frequently applied controls is the test proposed by Brooks, Gelman and Rubin (Brooks, Gelman, 1998, Gelman et al., 2004), which can be used only if two or more chains are available (m = number of chains ≥2). The statistic of this test, called corrected scale reduction factor, is based on “between chain variance” (B/n) and “within chain variance” (W) and is equal to:

$$CorrectedSRF = \frac{df+3}{df+1}\frac{\overbrace{\frac{n-1}{n}W+\frac{m+1}{m}\frac{B}{n}}^{V}}{W} = \frac{df+3}{df+1}PSRF, \tag{7}$$

where $df \approx 2V/var(V)$. The convergence is demonstrated if the square root of the statistic is less than 1.2 or more restrictively less than 1.1. In the multivariate case, i.e. in presence of several parameters to be estimated, the test can be performed by calculating both the corrected SRF for each factor and the following summary statistic:

$$MultivariatePSRF = max_e \frac{e'\left(\overbrace{\frac{n-1}{n}W+\frac{m+1}{m}\frac{B}{n}}^{V}\right)e}{e'We}. \tag{8}$$

To use the Bayesian method, it is necessary to select the prior distributions for the parameters of the new link function to calculate the posterior distributions. The parameters α and β have a prior distribution, which is not very informative so as not to add a priori information, while the parameters ω and γ have a Beta, since their values are between 0 and 1, and the parameter τ a Gamma, being always greater than 0

$$\begin{gathered}\alpha \sim \text{flat prior}\\ \beta \sim \text{flat prior}\\ \gamma \sim \text{Beta}(2,2)\\ \omega \sim \text{Beta}(0.5,0.5)\\ \tau \sim \text{Gamma}(1,0.5)\quad .\end{gathered}\tag{9}$$

The prior distributions of the parameters α and β are flat priors, which in Winbugs can be approximated by a Normal distribution with mean 0 and variance 1,000,000. Very often, the Bayesian logistic regression requires a flat prior, used to provide as little information as possible on the parameters (Gill, 2002). In my model, as in the traditional Bayesian logistic regression, the calculation of the posterior distribution is not simple and it requires the use of the Markov Chain Monte Carlo method.

Empirical analysis

The new method is demonstrated using 691 appointments observed in a hospital clinic in the second quarter of 2016. The distance to clinic does not affect the probability of being present at the visit because all of the patients, analyzed in this paper, belong to the same neighbourhood. Figure 1 describes the non-monotone trend of the probability of being present based on age groups.

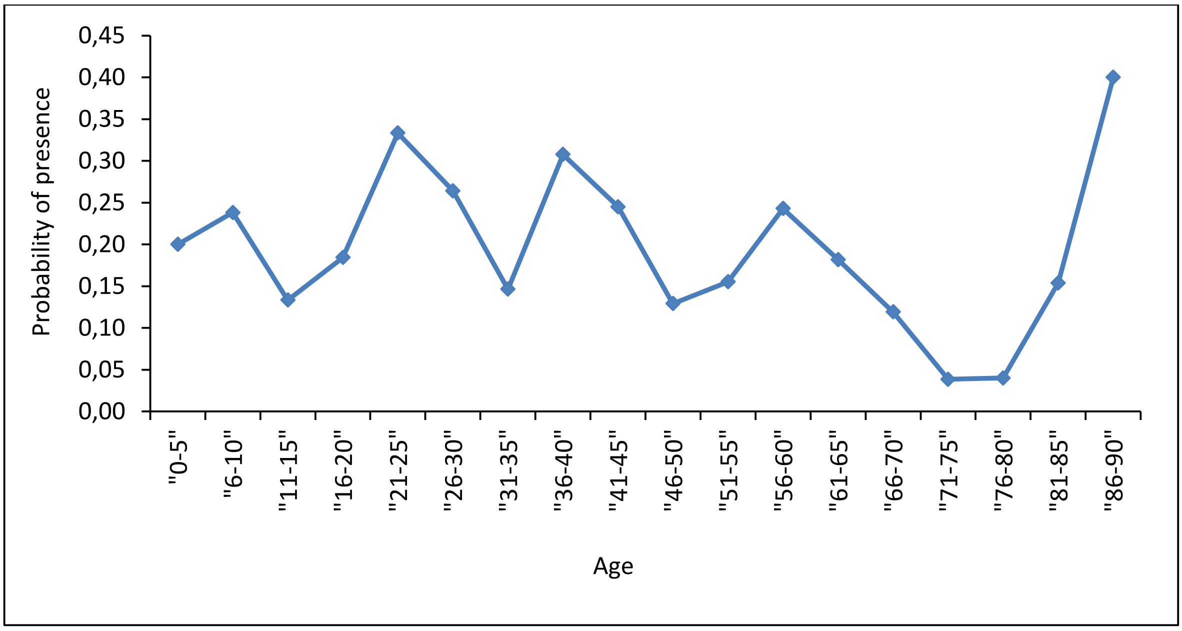

Figure 1 Empirical probability of presence

This paper analyses how age affects the probability of being present at the medical examination. Therefore, the variable of interest Y is dichotomous and it assumes value 1 if the patient is present and value 0 otherwise. The explanatory variable X, which may influence the probability of being present, is represented by the factor age. The probability p is linked to the explanatory variable age by function (3), being X, by its very nature, greater than or equal to 0. The zero represents the children until 12 months. The parameters α, β, γ, ω and τ are estimated using Winbugs software. The Markov Chain Monte Carlo method is used to obtain the posterior distributions from the prior distributions and the likelihood function. Two Markov chain are calculated such that Brooks-Gelman-Rubin test can be applied. For any chain, it is necessary to select the initial values of the parameters, which are respectively (0, -

0.1, 0.5, 0.5, 20) and (0.5, -0.6, 0, 2, 10). They are chosen considering the proprieties of the link function and the prior distributions. The number of analysed iterations are respectively 10000, 15000, 20000, 40000 and 50000. The average is used as punctual estimation of the parameters. The estimations of the parameters for the various iterations are shown in Table 1. Before interpreting them, it is necessary to study if the Markov Chains have reached a stable distribution, using Brooks-Gelman-Rubin diagnostic statistics. For example, the number of iterations necessary for convergence depends on the complexity of the model.

Table 1 Estimated parameters

	iterations				
	10000	15000	20000	40000	50000
α	-1.05	-1.033	-1.015	-1.031	-1.038
β	-0.01129	-0.01157	-0.01182	-0.01154	-0.01142
γ	0.4853	0.4764	0.4676	0.4739	0.4779
ω	0.7134	0.7171	0.7175	0.7185	0.7192
τ	3.251	3.273	3.28	3.271	3.271

Table 2 $\sqrt{CorrectedSRF}$

	iterations				
	10000	15000	20000	40000	50000
α	1.012582	1.021191	1.03211	1.000087	1.0008
β	1.005074	1.014746	1.015258	0.999997	1.000528
γ	1.001081	1.000516	1.000685	1.000026	0.999995
ω	1.00017	1.000382	1.000415	1.000141	1.000046
τ	0.999975	0.999995	1.000042	1.00002	0.999999

Table 3 $Multiple\ PSRF$

	iterations				
	10000	15000	20000	40000	50000
Multiple PSRF	1.005031	1.005492	1.009169	1.000621	1.000162

The chains reach the stationary already at 10000 iterations, indeed the square root of the corrected SRF is less than 1.1 (Table 2) and also the multiple PSRF is less than 1.1 (Table 3).

The new methods and others present in literature

The goodness of the model developed in this paper is verified by its comparison with the same models present in the literature. The new method is compared to the following methods: the traditional Bayesian logistic regression, in which the parameters of the link function (1) have a flat prior distribution; the traditional Bayesian probit regression, in which the prior distributions are flat distributions; the Frequentist logistic and probit regressions (McCullagh, Nelder, 1989), in which the parameters are estimated from the maximum likelihood; and the Bayesian logistic and probit regressions proposed by Gelman (Gelman et al., 2008), in which the prior distributions are Cauchy distributions.

The difference between the regression described in this paper and the traditional Bayesian logistic one consists exclusively in the function, which links the probability p_i to the explanatory variable x_i, which is represented by the patient age in this specific case. The link function of the new method is represented by formula (3), being $x_i \geq 0$, while the corresponding of the traditional logistic regression from (1). Both models use the Bayesian estimation method and they choose a flat prior distribution for the common parameters, i.e. for α and β. In the estimation process, the same initial values of the parameters α and β are chosen for both of the methodologies.

The difference between the new method and the Frequentist logistic one consists both in the chosen link function and in the estimation method. In the Frequentist logistic regression, the link between the probability of being present at the visit p_i is linked to the variable age by equation (1). In the Frequentist theory the parameters are considered unknown fixed values and therefore in the logistic Frequentist regression the parameters α and β of equation (1) are estimated maximizing the likelihood function as given in (10):

$$likelihood\ function = \prod_{i=1}^{I_s} p_i{}^{y_i}(1-p_i)^{1-y_i}, \tag{10}$$

where the probability p_i is given by function (1). The frequentist probit method differs from the frequentist logit only for the following link function:

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta x_i} e^{-z^2} dz, \tag{11}$$

while the estimation of the parameters α and β is obtained again by maximizing the likelihood function (10) where, however, p_i is given by equation (11) rather than by (1). In both of the frequentist models, the parameter β is significant (its p-value is less than 0.05 in both cases) and therefore the variable age affects the presence of the patient at the visit.

As in the Frequentist method, also in the Bayesian one the difference between the logit and the probit regressions consists only in the link function, which in the first is represented by function (1), while in the second by (11). The prior distributions of the probit model remain flat as in that logit and the initial values for the parameters are the same chosen for the logistic one. The standardization of the explanatory variable is used to avoid overflow in the estimation of the traditional Bayesian probit model. The methods proposed by Gelman differ from the new method for the link function and for the choice of the prior distributions of the parameters α and β, which have two Cauchy distributions. In Gelman's methods, the link function is represented by equation (1) in the logit regression and by (11) in the probit one, while in the new model by function (3). In the logistic model proposed by Gelman, the intercept α follows a Cauchy distribution with center 0 and scale 10, while the coefficient of the regressor, i.e. β, follows a Cauchy distribution with center 0 and scale 2.5. Instead, in the probit model proposed by Gelman, the prior distribution of the intercept is a Cauchy with center 0 and scale 16, while the parameter β follows a Cauchy distribution with center 0 and scale 4. Gelman chooses the distribution of Cauchy, because he considers it better than the normal for its characteristic shape. In a Normal, the values around zero are more probable and the extreme ones are less probable if they are compared to those of a Cauchy (Gelman et al., 2008).

Comparison

The comparison among the methods is developed using the mean square error of the predictor $\hat{Y}$, given by

$$MSE = \frac{\sum_{i=1}^{I_S}(\hat{y}_i - y_i)^2}{I_S}, \tag{12}$$

where $\hat{y}_i$ is the estimated value and y_i is the observed value. In the regressions with binary Y, $\hat{y}_i$ is exactly the same as $\hat{p}_i$ and therefore the MSE is called Brier score (Rufibach, 2010). This statistic is chosen to compare the different methodologies because it measures the difference between the observed value and the estimated one. The MSE values of the various methods are shown in Table 4, but only for the new method and the traditional Bayesian ones (logit and probit) the MSE values are calculated using the different numbers of MCMC iterations.

Table 4 MSE

		iterations				
		10000	15000	20000	40000	50000
New method		0.154564	0.154576	0.154604	0.154585	0.154579
Bayesian Logistic		0.155365	0.155366	0.155366	0.155367	0.155367
Bayesian Probit		0.155352	0.155351	0.155351	0.155351	0.155351
Frequentist Logistic	0.155373					
Frequentist Probit	0.155355					
Gelman's Logistic	0.155373					
Gelman's Probit	0.155355					

The convergence test proposed by Brooks, Gelman and Rubin (Brooks, Gelman, 1998) shows that the parameters estimated by the traditional Bayesian models come from a stationary distribution already with 10000 iterations, indeed the square root of the corrected SRF is less than 1.1 for both parameters and also the multiple PSRF is less than 1.1 (Table 5 and Table 6).

Table 5 $\sqrt{\text{CorrectedSRF}}$and MultiplePSRF for traditional Bayesian logistic model

	iterations				
	10000	15000	20000	40000	50000
$\sqrt{\text{CorrectedSRF}}$					
α	1.000801	1.000051499	1.00008	1.00002	1.000014
β	1.00072	1.000219976	1.000176	1.00006	0.999992
MultiplePSRF	1.001557	1.000547	1.000295	1.000038	0.9999807

Table 6 $\sqrt{\text{CorrectedSRF}}$and MultiplePSRF for traditional Bayesian probit model

	iterations				
	10000	15000	20000	40000	50000
$\sqrt{CorrectedSRF}$					
α	1.000082	1.000145	1.000235	1.000025	1.00004
β	1.00007	1.000039	1.000135	0.999992	0.999992
$MultiplePSRF$	1.000199	1.000383	1.000529	1.000054	1.000064

The Frequentist logit regression and the Gelman's logit method have the same MSE approximated to the sixth digit after the point, the same equivalence is true for the Frequentist probit, and the Gelman's probit. The parameters of the two methods, indeed, are very similar as shown in Table 7. The new method has always a lower MSE and therefore it estimates better the variable Y, i.e. the presence at a clinic appointment.

Table 7 Estimated parameters

	α	β
Frequentist Logistic	-1.020047	-0.009556
Gelman's Logistic	-1.019845	-0.009555
Frequentist probit	-0.627001	-0.005556
Gelman's probit	-0.626984	-0.005556

The MSE statistic evaluates the model fit. From a Frequentist perspective, the new model has 3 extra parameters and then it is more complex than the others present in literature. However, according Gelman et al. (2014), in a Bayesian analysis the number of parameters of a model can be different from the effective one, called p_D, which depends on data.

To explain this concept better, the example made by Gelman et al. (2014), is reported below. A model $y_1, \dots y_{I_s} \sim N(\theta, 1)$ with I_s large and $\theta \sim U(0, \infty)$ is analyzed. This prior distribution can be considered as a non-informative uniform prior distribution. The number of parameters is 1 but the effective one depends on the data. If y is close to 0, p_D is approximately 0.5 but if y is positive and large, the effective number of parameters increases approximately to 1. Then, in Bayesian analysis, to compare more models considering also the complexity, it is advised the use of the DIC statistic because it considers both the model fit and the complexity measured by the effective number of the parameters (Spiegelhalter et al., 2002). The formula of the DIC is given in (13):

$$DIC = \overline{D} + p_D, \tag{13}$$

where $\overline{D}$ is the posterior mean deviance and p_D the effective number of parameters (Spiegelhalter et al., 2002). The DIC is not a good measure when any parameter has a posterior distribution, which is substantially skewed (McElreath, 2016). In the new link function only τ has a skewed distribution while the parameters ω and γ have a symmetric distribution Beta. In the new method and in the Bayesian models the parameters α and β are distributed as a Normal (0, 1,000,000), while in the Gelman's models they are distributed as a Cauchy. These distributions are always symmetric. The values of the DIC are showed in Table 8. Here presented new method has the smallest DIC, then it perform to be the best in results.

Table 8 Comparison using Deviance information criterion (DIC)

Method	DIC
New method	683,713
Bayesian Logistic	685,498
Bayesian Probit	685,311
Gelman's Logistic	688,213
Gelman's Probit	688,123

Table 9 Comparison between the estimated probability and the observed one (EP)

		iterations				
		10000	15000	20000	40000	50000
New method		0.036443	0.036446	0.036464	0.036457	0.036455
Bayesian Logistic		0.036953	0.036944	0.036943	0.036933	0.036934
Bayesian Probit		0.036909	0.036907	0.036908	0.036908	0.036908
Frequentist Logistic	0.036883					
Frequentist Probit	0.036862					
Gelman's Logistic	0.036883					
Gelman's Probit	0.036862					

The comparison between the estimated probability and the observed probability is used to analyze further the different methods

$$EP = \frac{\sum_{i=1}^{I_S}(\hat{p}_i - p_i)^2}{I_s}. \tag{14}$$

The EP values of the various methods are shown in Table 9. The new method continues to have the lowest error (Table 9), resulting better.

Conclusion

For a medical structure, it is fundamental to be able to predict the presence of a patient at an appointment or at a health check, and to determine the optimal number of visits to be included in a day. As showed by many empirical studies, a factor, which influences the rate of presence or absence of the person in a non-monotone way, is age. Consequently, this work proposes a new link function, in which the probability of being present at the visit is linked to the explanatory variable age in a non-monotone way. The parameters of the new link function are estimated using the Bayesian method. The goodness of this new model is demonstrated by applying it to a real case and comparing it to other statistical methods. The new model estimates both the response variable and the probability of presence in a more precise way than the other compared methods. The goals of the future work are to refine further the new link function, for example by considering multiple explanatory variables together and by improving the estimation process of its parameters so as to lower further the MSE or EP statistics, and then to introduce the probabilities estimated by the new method in a system for planning visits.

References

1. Alaeddini, A., Yang, K., Reeves, P., Reddy C. K. (2015). A hybrid prediction model for no-shows and cancellations of outpatient appointments. *IIE Transactions on healthcare systems engineering*, Vol. 5, pp. 14-32.
2. Brooks, S. P., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, Vol. 7, No. 4, pp. 434-455.
3. Carlin, B. P., Louis, T. A. (2008). *Bayesian methods for data analysis*. CRC Press, Boca Raton.
4. Chua, S. L., Chow, W. L. (2018). *Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic*. Available at https://journals.sagepub.com/doi/full/10.1177/2010105818793155 [10 June 2018].
5. Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., Turkcan, A., Chakraborty, S., Sands, L. (2010). Using no-show modeling to improve clinic performance. *Health Informatics Journal*, Vol. 16, No. 4, pp. 246-259.
6. Davies, M. L., Goffman, R. M., May, J. H., Monte, R. J., Rodriguez, K. L., Tjader, Y. C., Vargas, D. L. (2016). Large-scale no-show patterns and distributions for clinic operational research. Healthcare, Vol. 4, No. 1, pp. 1-12.
7. Gebhart, T. (2017). *No-Show Management in Primary Care: A Quality Improvement Project*. Available at digitalcommons.ohsu.edu [10 June 2018].
8. Gelman, A, Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). *Bayesian Data Analysis*. CRC/Chapman & Hall, Boca Raton.
9. Gelman, A., Jakulin, A., Pittau, M. G., Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, Vol. 2, No. 4, pp. 1360-1383.
10. Gelman, A., Hwang, J., Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, Vol. 24, No. 6, pp. 997-1016.
11. Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. CRC press, Boca Raton.
12. Kéry, M. (2010). *Introduction to WinBUGS for ecologists*. Academic Press, Burlington.
13. Kopach, R., DeLaurentis, P. C., Lawley, M., Muthuraman, K., Ozsen, L., Rardin, R., Wang, Intravado, P., Qu, X., Willis, D. (2007). Effects of clinical characteristics on successful open access scheduling. *Health care management science*, Vol. 10, No. 2, pp. 111-124.
14. LaGanga, L. R., Lawrence, S. (2007). *Appointment scheduling with overbooking to mitigate productivity loss from no-shows*. Available at http://www.poms.org/conferences/cso2007/talks/06.pdf [10 June 2018].

15.Lee, S., Min, D., Ryu, J., Yih, Y., (2013). A simulation study of appointment scheduling in outpatient clinics: Open access and overbooking. *Simulation*, Vol. 89, No. 12, pp. 1459-1473.

16.McCullagh, P., Nelder, J. (1989). *Generalized Linear Models.* Chapman and Hall, Boca Raton.

17.McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.

18.Ntzoufras, I. (2008). *Bayesian modeling using WinBUGS*. Wiley, Hoboken.

19.O'Hare, C. D., Corlett, J. (2004). The outcomes of open-access scheduling. *Family Practice Management*, Vol. 11, No. 2, pp. 35-38.

20.Rufibach, K. (2010). Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology*, Vol. 63, pp. 938-939.

21.Samorani, M., LaGanga, L. R. (2015). Outpatient appointment scheduling given individual day-dependent no-show predictions. *European Journal of Operational Research*, Vol. 240, No. 1, pp. 245-257.

22.Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 4, pp. 583-639.

About the author

Gloria Gheno is a statistic consultant. She worked on the cancellation problem when she was a teaching assistant at the Free University of Bolzano-Bozen. She was a post-doc researcher at the Ca' Foscari University of Venezia and she collaborated with the Bocconi University and with the University of Sydney. She presented her works at many international conferences and she is a reviewer of two scientific journals. She wrote a few papers and a package for R-project. She received her Master's Degree in Economics in 2002 at the Ca' Foscari University of Venezia, her Master's Degree in Statistics in 2011 at the University of Padova and the PhD in Statistics in 2015 at the University of Padova with scholarship. She obtained her professional international Master's Degree in Economics and Finance in 2004 at the Ca' Foscari University of Venezia. She was a trainee chartered accountant and a trainee auditor. Now she collaborates with M. Garbuio, a senior lecturer at the University of Sydney, and with R. Dussin, a chartered accountant. Author can be contacted at: *gloriagheno@libero.it*.