# A new asymmetric link-based binary regression model to detect Parkinson's disease by using replicated voice recordings

1st Lizbeth Naranjo
*Facultad de Ciencias*
*Universidad Nacional Autónoma de México*
México DF, Mexico
lizbethna@ciencias.unam.mx

2nd Carlos J. Pérez
*Departamento de Matemáticas*
*Universidad de Extremadura*
Cáceres, Spain
carper@unex.es

3rd Jacinto Martín
*Departamento de Matemáticas*
*Universidad de Extremadura*
Badajoz, Spain
jrmartin@unex.es

4th Fernando Calle-Alonso
*Departamento de Matemáticas*
*Universidad de Extremadura*
Cáceres, Spain
fcalonso@unex.es

***Abstract*—Addressing dependent data as independent has become usual for Parkinson's Disease (PD) detection by using features extracted from replicated voice recordings. A binary regression model with an Asymmetric Student $t$ (AST) distribution as link function has been developed in a classification context by taking into account the within-subject dependence. This opens the possibility of handling situations in which the probabilities of the binary response approach 0 and 1 at different rates. The computational issue has been addressed by proposing and using a representation based on a mixture of normal distributions for the AST distribution. This allows to include latent variables to derive a Gibbs sampling algorithm that is used to generate samples from the posterior distribution. The applicability of the proposed approach has been tested with a simulation-based experiment and has been applied to a real dataset for PD detection.**

***Index Terms*—Asymmetric Student $t$, Bayesian binary regression, Gibbs sampling, Parkinson's disease, Voice features.**

## I. Introduction

Parkinson's Disease (PD) is the second most common neurodegenerative disorder and the most common movement disorder. This disease leads to the progressive deterioration of the motor function due to the loss of dopamine. Currently, there is no cure, but there are successful treatments to reduce the symptoms. Addressing early diagnoses of people with PD is a key issue to improve the patients' quality of life.

Vocal impairment has been presented as one of the most likely and earliest signs of the disease [1]. The muscles are affected, and, as a consequence, the voice production does not perform well. This can be measured in an objective way by extracting features from voice recordings and analyzing them properly.

This research has been supported by projects MTM2014-56949-C3-3-R and MTM2017-86875-C3-2-R (MINECO), and projects IB16054 and GR18108 (Junta de Extremadura/European Regional Development Funds, EU).

Addressing dependent data as independent has become usual for PD detection by using voice features, see, for instance, [2]–[5] and references therein. [2] presented one of the most used PD datasets consisting on 22 features extracted from 195 recordings of sustained /a/ phonations. These phonations belonged to only 32 people (24 with PD), having each one six or seven replicated voice recordings. This dataset[1] has been extensively used with classifiers based on independence assumptions. Obviating the dependent nature of the data artificially increases the sample size and leads to a diffuse criterion to decide when a subject should be classified as suffering from PD or healthy, since it usually happens that some voice recordings of the same subject are classified as healthy and some others as disordered.

Since the features were extracted from multiple voice recordings from the same subjects in a concrete time, in principle, the features should be identical for each subject. However, imperfections in technology and the own biological variability result in non-identical replicated features that are more similar to one another than to features from different subjects. Therefore, the underlying within-subject dependence must be properly modelled. For the first time, [6] demonstrates a classification approach for PD detection that takes into account the underlying within-subject dependence by using the dataset provided in [2]. This Bayesian approach was based on a binary logistic regression model. However, a Markov Chain Monte Carlo (MCMC) was not directly implemented. Instead the posterior distribution was generated by using WinBUGS [7]. WinBUGS is a software used to implement MCMC simulations, but it is not possible to know how the generation process is being performed. Later, a Bayesian binary regression approach based on probit model was proposed [8], which

[1] http://archive.ics.uci.edu/ml/datasets/Parkinsons

ISBN 978-90-827970-1-5 © EURASIP 2018

introduces latent variables to provide an augmented framework available for efficient simulation by using Gibbs sampling.

When modelling binary response data through regression models, several link functions have been defined. The most popular models are the logistic and normal ones. However, in many applications the overall fit can be significantly improved by using asymmetric or skewed links. [9] considered that the rates at which the probabilities of a given binary response approach 0 or 1 are different. Under this notion, a link is symmetric if the rates are similar, otherwise the link is asymmetric. An asymmetric link can be characterized as positively skewed if the rate approaching 1 is faster than the rate approaching 0, otherwise it is negatively skewed.

Besides the link asymmetry, tail modelling is important to produce values farther from the mean, which is an advantage for robustness. Although the logistic distribution has slightly heavier tails than the normal one, symmetry is also present and they are not flexible enough compared to other distributions for which the tails can be modelled even separately. The Asymmetric Student $t$ (AST) distribution can handle heavy tails and asymmetry simultaneously. There are several parameterizations for this distribution, but the one based on [10] is especially interesting because it can be efficiently integrated in a binary regression model considering replications through a mixture of normal distributions representation.

An AST link-based binary regression addressing replications in a classification context is proposed. The representation of the AST distribution as a mixture of normal distributions and the introduction of latent variables allows to derive a Gibbs sampling algorithm and, therefore, to calculate predictive probabilities to assign the class. A simulation-based experiment shows the potential of the approach by comparing with the other two binary regression approaches addressing replications that have been developed in the scientific literature. The results show the superior performance of the proposed approach and open the possibility of using more flexible links. The approach has also been applied to the PD dataset provided in [2].

The rest of the paper is as organized as follows. Section II describes the probability density function (pdf) of the AST distribution in the chosen parameterization and presents a proposition to represent the AST distribution as a mixture of normal distributions. In Section III, the AST link-based binary regression model is described from a Bayesian viewpoint, and the Gibbs sampling algorithm is derived. Section IV presents a simulation-based experiment to show the model performance when asymmetry is present and compares the proposed approach with symmetric link-based regression models such as probit and logit. Section V applies the approach to a real dataset. A brief conclusion is presented in Section VI. Finally, the proof of the proposition and the full conditional distributions are presented in two appendices.

## II. AST DISTRIBUTION

The AST distribution proposed in [10] is defined by a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$, a skewness parameter $\alpha \in (0,1)$, and two shape parameters $\nu_1 > 0$ and $\nu_2 > 0$ for tail modelling. It is denoted as $Z \sim \mathrm{AST}(\mu, \sigma, \alpha, \nu_1, \nu_2)$ and its pdf is:

$$f_{AST}(z) = \begin{cases} \frac{1}{\sigma}\left[1 + \frac{1}{\nu_1}\left(\frac{z-\mu}{2\sigma\alpha K(\nu_1)}\right)^2\right]^{-\frac{\nu_1+1}{2}} & \text{if } z \le \mu, \\ \frac{1}{\sigma}\left[1 + \frac{1}{\nu_2}\left(\frac{z-\mu}{2\sigma(1-\alpha) K(\nu_2)}\right)^2\right]^{-\frac{\nu_2+1}{2}} & \text{if } z > \mu, \end{cases}$$

where $K(\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}$ and $\Gamma(\cdot)$ is the gamma function.

The density is unimodal with mode given by $\mu$, which is also the $\alpha$-quantile, since $\alpha = \mathrm{P}[Z \le \mu]$. When $\alpha = 1/2$ and $\nu_1 = \nu_2$, the density is symmetric. This parameterization allows to model each tail separately and identify the effect of the skewness and shape parameters.

We have derived a proposition to represent the AST distribution by means of a mixture of normal distributions. The proof is presented in Appendix A.

**Proposition 1.** Let $Z$ be a random variable with pdf represented as

$$f(z) = \alpha \int f(z|\gamma_1) f(\gamma_1) d\gamma_1 + (1-\alpha) \int f(z|\gamma_2) f(\gamma_2) d\gamma_2,$$

where

$$\begin{aligned} Z|\gamma_1 &\sim \mathrm{Normal}\left(\mu,\ [2\sigma\alpha K(\nu_1)]^2 \gamma_1^{-1}\right) I[Z \le \mu], \\ \gamma_1 &\sim \mathrm{Gamma}(\nu_1/2,\ \texttt{rate} = 2/\nu_1), \\ Z|\gamma_2 &\sim \mathrm{Normal}\left(\mu,\ [2\sigma(1-\alpha)K(\nu_2)]^2 \gamma_2^{-1}\right) I[Z > \mu], \\ \gamma_2 &\sim \mathrm{Gamma}(\nu_2/2,\ \texttt{rate} = 2/\nu_2), \end{aligned}$$

with $\alpha \in (0,1)$ and $I[\cdot]$ denotes the indicator function, then $Z \sim \mathrm{AST}(\mu, \sigma, \alpha, \nu_1, \nu_2)$.

## III. THE APPROACH

### A. Hierarchical binary model with replications

Suppose that $n$ independent binary random variables $Y_1, \ldots, Y_n$ are observed, where $Y_i$ is Bernoulli distributed with success probability $\mathrm{P}(Y_i = 1) = p_i$, $i = 1, \ldots, n$. The probabilities $p_i$ are related to a set of covariates $\boldsymbol{x}_i$, where $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iK})'$ is a $K \times J$ matrix of a set of $K$ covariates which have been measured with $J$ replications. Suppose that $\boldsymbol{x}_{ij} = (x_{i1j}, \ldots, x_{iKj})'$ is the $j$-th replication of the unknown covariates vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iK})'$ and assume that they have a linear relationship, i.e., they follow an additive measurement error model structure [11]. This leads to the following hierarchical model:

$$\begin{aligned} Y_i &\sim \mathrm{Bernoulli}(p_i), \\ p_i &= \Psi(\beta_0 + \boldsymbol{w}_i' \boldsymbol{\beta}), \\ \boldsymbol{x}_{ij} &= \boldsymbol{w}_i + \boldsymbol{\varepsilon}_{ij}, \\ \varepsilon_{ikj} &\sim \mathrm{Normal}(0, \sigma_k^2), \end{aligned}$$

where $\beta_0$ is the intercept parameter, $\boldsymbol{\beta}$ is a $K$-dimensional vector of unknown parameters, $\Psi(\cdot)$ is a known nonnegative and nondecreasing function ranging between 0 and 1, and $\sigma_k^2$ is the variance related to the replicates of the $k-$th covariate.

ISBN 978-90-827970-1-5 © EURASIP 2018

In the Bayesian methodology, the initial information about the parameters is elicited through a prior distribution, which is combined with the likelihood to provide a posterior distribution that contains all the information about the model. Then, in order to complete the hierarchical model, the prior distributions must be defined. For the regression parameters $\beta_k$, normal distributions are usually assumed, inverse Gamma distributions can be chosen for the variance parameters $\sigma_k^2$ as conjugate, and, finally, the latent variables $w_{ik}$ can also be considered as normal distributions, that is:

$$\begin{aligned}\beta_0 &\sim \text{Normal}(b_0, B_0),\\ \beta_k &\sim \text{Normal}(b_k, B_k),\\ \sigma_k^2 &\sim \text{InverseGamma}(s_k, r_k),\\ w_i &\sim \text{Normal}_K(\mu, \Omega),\end{aligned}$$

with $k = 1, \ldots, K$.

### *B. Binary regression model with AST-link*

We assumed an AST-based link model, $p_i = \Psi(\beta_0 + \boldsymbol{w}_i'\boldsymbol{\beta})$, where $\Psi(\cdot)$ depends on the cdf of the distribution $\text{AST}(0, 1, \alpha, \nu_1, \nu_2)$.

Based on the idea of introducing latent variables of [12], independent variables $Z_1, \ldots, Z_n$ are introduced in the model, where $Z_i$ given $\boldsymbol{w}_i$, $\beta_0$, $\boldsymbol{\beta}$, $\alpha$, $\nu_1$ and $\nu_2$ is distributed as $\text{AST}(\beta_0 + \boldsymbol{w}_i'\boldsymbol{\beta}, 1, \alpha, \nu_1, \nu_2)$, and it is defined $Y_i = 1$ if $Z_i > 0$ and $Y_i = 0$ if $Z_i \leq 0$. Besides, by introducing the latent variables $\gamma_{i1}$ and $\gamma_{i2}$, the AST distribution of $Z_i$ can be represented as a mixture of normal distributions by Proposition 1.

In addition, prior distributions for the parameters related to the AST distribution must be defined, i.e:

$$\begin{aligned}\alpha &\sim \text{Beta}(a_1, a_2),\\ \nu_1 &\sim \text{Gamma}(e_1, d_1),\\ \nu_2 &\sim \text{Gamma}(e_2, d_2).\end{aligned}$$

### *C. Exploring the posterior distribution*

The likelihood function of the model proposed in the previous subsections considering the observed and the latent variables is given by

$$\begin{aligned}&\mathcal{L}(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta_0, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \alpha, \nu_1, \nu_2 \mid \boldsymbol{y}, \boldsymbol{x})\\ &= \prod_{i=1}^{n} \Bigg\{ f_Y(y_i|z_i) f_Z(z_i|\gamma_{i1}, \gamma_{i2}, \boldsymbol{w}_i, \beta_0, \boldsymbol{\beta}, \alpha, \nu_1, \nu_2)\\ &\times f_{\gamma_1}(\gamma_{i1}|\nu_1) f_{\gamma_2}(\gamma_{i2}|\nu_2) \left[\prod_{j=1}^{J} f_{\boldsymbol{X}}(\boldsymbol{x}_{ij}|\boldsymbol{w}_i, \boldsymbol{\sigma}^2)\right] f_{\boldsymbol{W}}(\boldsymbol{w}_i) \Bigg\}\end{aligned}$$

The joint posterior distribution of the latent variables $\boldsymbol{z}$, $\boldsymbol{w}$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, and the parameters $\beta_0$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$, $\alpha$, $\nu_1$ and $\nu_2$ is obtained by using the likelihood function and the prior distributions, and it is given by

$$\begin{aligned}&\pi(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta_0, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \alpha, \nu_1, \nu_2 \mid \boldsymbol{y}, \boldsymbol{x})\\ &\propto \mathcal{L}(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \beta_0, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \alpha, \nu_1, \nu_2 \mid \boldsymbol{y}, \boldsymbol{x})\\ &\times \pi(\beta_0)\pi(\boldsymbol{\beta})\pi(\sigma_1^2)\cdots\pi(\sigma_K^2)\pi(\alpha)\pi(\nu_1)\pi(\nu_2).\end{aligned}$$

The posterior distribution previously presented is analytically intractable, so an MCMC algorithm must be applied to generate from the posterior distribution. The idea of introducing latent variables and the mixture representation of AST of Proposition 1, make possible the development of an efficient Gibbs sampling algorithm with easy-to-generate full conditional distributions, that are necessary to implement the iterative process. The full conditional distributions are presented in Appendix B.

## IV. Simulation-based experiment

A simulation-based experiment has been conducted to validate the proposed approach. The generated datasets are based on the motivating problem on PD detection. Specifically, the covariates $\boldsymbol{x}_{ij}$ have not been simulated, and we have used the four most relevant acoustic features according to [2], i.e., Harmonic-to-Noise Ratio (HNR), Recurrence Period Detrended Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Entropy (PPE). The covariates $\boldsymbol{w}_i$, that are used to simulate, are computed by averaging their $J = 6$ or $J = 7$ replications. The following values were considered for the regression parameters: $\beta_0 = 3.3$ and $\boldsymbol{\beta} = (0.02, -1.2, 0.7, 4.6)$. The data structure keeps the same, but the responses $y_i$ (0 for healthy subjects and 1 for those with PD) have been simulated according to an AST distribution with a very different rate at which probabilities approach 0 or 1. Specifically, an $\text{AST}(0, 1, 0.75, 0.5, 5)$ distribution has been considered to calculate the probabilities $p_i$, following the next process: generate $u_i \sim \text{U}(0, 1)$, if $p_i \geq u_i$, then $y_i = 1$, else $y_i = 0$. A total of 100 datasets for each link have been generated for posterior average purpose.

The hyperparameters used for the latent variables and prior distributions of the probit, logit and AST links were given by $\beta_0 \sim \text{Normal}(0, 100)$, $\beta_k \sim \text{Normal}(0, 100)$, $1/\sigma_k^2 \sim \text{Gamma}(0.01, 0.01)$ and $w_{ik} \sim \text{Normal}(0, 100)$. Besides, the prior distributions for the skewness and shape parameters of the AST link are given by $\alpha \sim \text{Beta}(1, 1)$, $\nu_1 \sim \text{Gamma}(1, 1)$ and $\nu_2 \sim \text{Gamma}(1, 1)$.

The posterior estimates have been obtained by using the Gibbs sampling algorithm described in Appendix A. A total of 200,000 iterations with a burn-in of 50,000 and a thinning period of 30 generated values have been considered for each chain. With these specifications, all the chains generated appear to have converged. The convergence analysis has been performed by using `BOA` package [13].

For each generated dataset, a cross-validation has been used to assess the generalization performance of the model with probit, logit and AST links. Specifically, a stratified sampling to choose 75% for the training subset and 25% for the testing subset has been considered. Note that the classifier learns from 18 PD and 6 healthy individuals (training subset), and the parameters are applied to predict the outcome of 6 PD and 2 healthy subjects (testing subset). The model parameters are determined using the training subset, and errors are computed using the testing subset. The following notation is considered for each iteration in the cross-validation scheme: TP

ISBN 978-90-827970-1-5 © EURASIP 2018

(True Positive), TN (True Negative), FP (False Positive), FN (False Negative), accuracy rate=$(TN+TP)/n$, sensitivity = $TP/(TP+FN)$, specificity =$TN/(TN+FP)$, and precision =$TP/(TP+FP)$. This process is performed 100 times and the results are then averaged.

After the cross-validation has been performed for each dataset, the accuracy metrics from the 100 datasets have been averaged and presented in Table I. One-way ANOVA reported statistically significant differences among the accuracy means of the three methods (p-value=0.006), providing AST link-based approach the best result with an accuracy mean of 81.75%. This means around 3% and 7% more than the ones obtained with probit and logit models, respectively.

TABLE I
ACCURACY RATES AND OTHER INDICATORS FOR THE MODELS BASED ON THE 100 SIMULATED DATASETS.

| **Logit** | Mean | SD |
|---|---|---|
| Accuracy | 0.7437 | 0.1698 |
| Sensitivity | 0.7800 | 0.2022 |
| Specificity | 0.6350 | 0.3817 |
| Precision | 0.8784 | 0.1262 |
| **Probit** | Mean | SD |
| Accuracy | 0.7875 | 0.1623 |
| Sensitivity | 0.8300 | 0.1909 |
| Specificity | 0.6600 | 0.3752 |
| Precision | 0.8931 | 0.1156 |
| **AST** | Mean | SD |
| Accuracy | 0.8175 | 0.1563 |
| Sensitivity | 0.8783 | 0.1722 |
| Specificity | 0.6350 | 0.3883 |
| Precision | 0.8885 | 0.1179 |

Figure 1 shows some graphics for the same randomly chosen dataset fitted by logit, probit and AST models. The symbol $\times$ denotes values for the responses $y_i$ and the symbol $\bullet$ denotes the estimated probabilities $\widehat{p}_i$ with the three models, and the lines are the simulated probabilities $p_i$. The flexibility to model the tails makes that AST link-based model provides the best results.

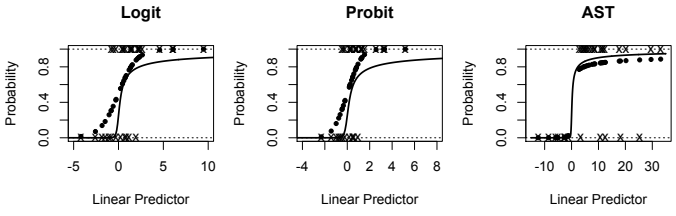


Fig. 1. Responses, estimated and simulated probabilities for one randomly chosen dataset with logit, probit and AST models.

## V. APPLICATION

In this section, the observed responses from the PD dataset are used as well as the four most relevant acoustic features according to [2]. The health status is imbalanced since there are 8 healthy subjects and 24 people suffering from PD. This supports the idea that the success rate could be approaching 0 and 1 in a different way. The same prior distributions, MCMC specifications and cross-validation framework used in the previous section have been considered here. The results are presented in Table II.

TABLE II
ACCURACY RATES AND OTHER INDICATORS FOR THE MODELS BASED ON THE PD DATASET.

| **Logit** | Mean | SD |
|---|---|---|
| Accuracy | 0.8525 | 0.1142 |
| Sensitivity | 0.9433 | 0.1258 |
| Specificity | 0.5800 | 0.3312 |
| Precision | 0.8805 | 0.0919 |
| **Probit** | Mean | SD |
| Accuracy | 0.8550 | 0.1119 |
| Sensitivity | 0.9416 | 0.1261 |
| Specificity | 0.5950 | 0.3232 |
| Precision | 0.8844 | 0.0883 |
| **AST** | Mean | SD |
| Accuracy | 0.8687 | 0.1056 |
| Sensitivity | 0.9516 | 0.1093 |
| Specificity | 0.6200 | 0.3265 |
| Precision | 0.8923 | 0.0899 |

The best accuracy rate has been obtained with the AST link-based approach, being 86.87%, more than 1% over the accuracies of logit and probit approaches. Although the differences are not statistically significant (p-value of one-way ANOVA is 0.536), a relevant fact is that this model also improves the sensitivity, specificity and precision. Also the standard deviations are smaller or of the same magnitude order.

These data do not have a large improvement margin when the appropriate statistical treatment is applied to the replications, since the sample size is very small and the number of healthy subjects is only 8. This greatly affects the specificity, since, for each iteration of the cross-validation, the learning is performed with only 6 healthy people and the training set is composed of only 2 subjects. The best specificity is 62% with the AST link-based approach, whereas the sensitivity is 95.16%. In this context, it is very important to have a high sensitivity to allow the subject to access to an early treatment.

## VI. CONCLUSION

Addressing dependent data as independent has become usual for PD detection by using features extracted from replicated voice recordings. The existing within-subject dependence must be properly modelled. For this task, an AST link-based binary regression model addressing replications is proposed. The inclusion of two types of latent variables has allowed to derive a Gibbs sampling algorithm to generate from the posterior distribution. The good performance of the proposed approach has been shown with a simulation and a real data application. To the best of the authors' knowledge, the proposed approach is the first one addressing replications at the same time that considers an asymmetric link function in a binary regression context. This approach covers a gap in the scientific literature for situations in which the probabilities of a given binary response approach 0 and 1 at different rates.

Although the problem of PD detection based on features extracted from voice recordings has motivated this work, the

ISBN 978-90-827970-1-5 © EURASIP 2018

approach can be applied to other classification contexts where it is necessary to account for the dependent nature of the data in a replicated measure-based design.

## ACKNOWLEDGMENT

Thanks to M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig for distributing the PD dataset.

## APPENDIX

### APPENDIX A. PROOF OF PROPOSITION 1

It is enough to observe that

$$\int f(z|\gamma_1)f(\gamma_1)\mathrm{d}\gamma_1 = \frac{1}{\sigma\alpha}\left[1+\frac{1}{\nu_1}\left(\frac{z-\mu}{2\sigma\alpha K(\nu_1)}\right)^2\right]^{-\frac{\nu_1+1}{2}} I[z\le\mu],$$

$$\int f(z|\gamma_2)f(\gamma_2)\mathrm{d}\gamma_2 = \frac{1}{\sigma(1-\alpha)}\left[1+\frac{1}{\nu_2}\left(\frac{z-\mu}{2\sigma(1-\alpha)K(\nu_2)}\right)^2\right]^{-\frac{\nu_2+1}{2}} I[z>\mu],$$

and therefore

$$f(z)=\alpha\int f(z|\gamma_1)f(\gamma_1)\mathrm{d}\gamma_1+(1-\alpha)\int f(z|\gamma_2)f(\gamma_2)\mathrm{d}\gamma_2$$

is the pdf of the $\mathrm{AST}(\mu,\sigma,\alpha,\nu_1,\nu_2)$.

### APPENDIX B. FULL CONDITIONAL DISTRIBUTIONS

In order to simplify the notation, define $\eta_i=\beta_0+\boldsymbol{w}_i'\boldsymbol{\beta}$, $V_1=[2\alpha K(\nu_1)]^2$, $V_2=[2(1-\alpha)K(\nu_2)]^2$.

$$[z_i|\cdots] \sim \begin{cases} \mathrm{N}\left(\eta_i, V_1\gamma_{i1}^{-1}\right) I[z_i>0]I[z_i\le\eta_i] \text{ if } y_i=1 \\ \mathrm{N}\left(\eta_i, V_2\gamma_{i2}^{-1}\right) I[z_i>0]I[z_i>\eta_i] \text{ if } y_i=1 \\ \mathrm{N}\left(\eta_i, V_1\gamma_{i1}^{-1}\right) I[z_i\le 0]I[z_i\le\eta_i] \text{ if } y_i=0 \\ \mathrm{N}\left(\eta_i, V_2\gamma_{i2}^{-1}\right) I[z_i\le 0]I[z_i>\eta_i] \text{ if } y_i=0 \end{cases}$$

$$[\gamma_{i1}|\cdots] \sim \mathrm{Gamma}\left(\tfrac{\nu_1+1}{2}, \tfrac{(z_i-\eta_i)^2}{2V_1}+\tfrac{\nu_1}{2}\right)$$

$$[\gamma_{i2}|\cdots] \sim \mathrm{Gamma}\left(\tfrac{\nu_2+1}{2}, \tfrac{(z_i-\eta_i)^2}{2V_2}+\tfrac{\nu_2}{2}\right)$$

Note that if $z_i\le\beta_0+\boldsymbol{w}_i'\boldsymbol{\beta}$ then $\gamma_{i1}>0$ and $\gamma_{i2}=0$, but if $z_i>\beta_0+\boldsymbol{w}_i'\boldsymbol{\beta}$ then $\gamma_{i1}=0$ and $\gamma_{i2}>0$.

$$\begin{aligned}[\boldsymbol{w}_i|\cdots] &\sim \mathrm{Normal}_K(\mu_i^*,\Omega_i^*)\\ [\beta_0|\cdots] &\sim \mathrm{Normal}\left(b_0^*,B_0^*\right)\\ [\sigma_k^2|\cdots] &\sim \mathrm{InverseGamma}\left(s_k^*,r_k^*\right)\end{aligned}$$

where

$$\begin{aligned}\mu_i^* &= \Omega_i^*\Bigg(\boldsymbol{\beta}(z_i-\beta_0)\left\{\tfrac{\gamma_{i1}\mathrm{I}[z_i\le\eta_i]}{V_1}+\tfrac{\gamma_{i2}\mathrm{I}[z_i>\eta_i]}{V_2}\right\}\\ &\quad +\mathrm{diag}\left(\tfrac{1}{\sigma_1^2},\ldots,\tfrac{1}{\sigma_K^2}\right)\sum_{j=1}^{J}\boldsymbol{x}_{ij}+\Omega^{-1}\mu\Bigg),\\ \Omega_i^* &= \Bigg(\boldsymbol{\beta}\boldsymbol{\beta}'\left\{\tfrac{\gamma_{i1}\mathrm{I}[z_i\le\eta_i]}{V_1}+\tfrac{\gamma_{i2}\mathrm{I}[z_i>\eta_i]}{V_2}\right\}\\ &\quad +\mathrm{diag}\left(\tfrac{J}{\sigma_1^2},\ldots,\tfrac{J}{\sigma_K^2}\right)+\Omega^{-1}\Bigg)^{-1},\end{aligned}$$

$$\begin{aligned} b_0^* &= B_0^*\left(\tfrac{b_0}{B_0}+\tfrac{m_1}{M_1}+\tfrac{m_2}{M_2}\right),\ B_0^*=\left(\tfrac{1}{B_0}+\tfrac{1}{M_1}+\tfrac{1}{M_2}\right)^{-1},\\ m_1 &= \tfrac{\sum_{i=1}^n\gamma_{1i}(y_i-\boldsymbol{w}_i'\boldsymbol{\beta})I[y_i\le\eta_i]}{\sum_{i=1}^n\gamma_{1i}I[y_i\le\eta_i]},\quad M_1=\tfrac{V_1}{\sum_{i=1}^n\gamma_{1i}I[y_i\le\eta_i]},\\ m_2 &= \tfrac{\sum_{i=1}^n\gamma_{2i}(y_i-\boldsymbol{w}_i'\boldsymbol{\beta})I[y_i>\eta_i]}{\sum_{i=1}^n\gamma_{2i}I[y_i>\eta_i]},\quad M_2=\tfrac{V_2}{\sum_{i=1}^n\gamma_{2i}I[y_i>\eta_i]},\\ s_k^* &= s_k+\tfrac{nJ}{2},\quad r_k^*=r_k+\tfrac{1}{2}\textstyle\sum_{i=1}^n\sum_{j=1}^J\left(x_{ikj}-w_{ik}\right)^2.\end{aligned}$$

The full conditional distributions of $\boldsymbol{\beta}$, $\alpha$, $\nu_1$ and $\nu_2$ are not standard, but they can be easily sampled by using an acceptance-rejection method or the Metropolis-Hastings algorithm.

The final algorithm consists of choosing initial values $\boldsymbol{z}^{(0)}$, $\boldsymbol{w}^{(0)}$, $\boldsymbol{\gamma}_1^{(0)}$, $\boldsymbol{\gamma}_2^{(0)}$, $\beta_0^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\sigma_1^{2(0)},\ldots,\sigma_K^{2\ (0)}$, $\alpha^{(0)}$, $\nu_1^{(0)}$ and $\nu_2^{(0)}$, and iteratively sampling $\boldsymbol{z}^{(l)}$, $\boldsymbol{w}^{(l)}$, $\boldsymbol{\gamma}_1^{(l)}$, $\boldsymbol{\gamma}_2^{(l)}$, $\beta_0^{(l)}$, $\boldsymbol{\beta}^{(l)}$, $\sigma_1^{2(l)},\ldots,\sigma_K^{2\ (l)}$, $\alpha^{(l)}$, $\nu_1^{(l)}$ and $\nu_2^{(l)}$ from their full conditional distributions.

## REFERENCES


[1] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, Elsevier, 2005.

[2] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.

[3] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[4] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with Parkinson's disease through nonlinear dynamics," in *Advances in Nonlinear Speech Processing*, T. Drugman and T. Dutoit, Eds., vol. LNAI 7911 of *Lecture Notes in Artificial Intelligence*, pp. 112–119. Springer-Verlag, 2013.

[5] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, pp. 904–913, 2014.

[6] C. J. Pérez, L. Naranjo, J. Martín, and Y. Campos-Roca, "A latent variable-based Bayesian regression to address recording replication in Parkinson's disease," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO-2014)*, EURASIP, Ed., Lisbon, Portugal, 2014, pp. 1447–1451, IEEE.

[7] I. Ntzoufras, *Bayesian Modeling Using WinBUGS*, Wiley Series in Computational Statistics. Wiley, New Jersey, 2011.

[8] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and J. Martín, "Addressing voice recording replications for Parkinson's disease detection," *Expert Systems With Applications*, vol. 46, pp. 286–292, 2016.

[9] M.-H. Chen, D. K. Dey, and Q.-M. Shao, "A new skewed link model for dichotomous quantal response data," *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1172–1186, 1999.

[10] D. Zhu and J. W. Galbraith, "A generalized asymmetric Student-$t$ distribution with application to financial econometrics," *Journal of Econometrics*, vol. 157, no. 2, pp. 297–305, 2010.

[11] J. P. Buonaccorsi, *Measurement Error: Models, Methods and Applications*, Chapman and Hall/CRC, Boca Raton, Florida, 2010.

[12] J. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.

[13] B. J. Smith, "BOA: an R package for MCMC output convergence assessment and posterior inference," *Journal of Statistical Software*, vol. 21, no. 11, pp. 1–37, 2007.