DpGLM: Semiparametric Bayesian Inference for a GLM Using Inhomogeneous Normalized Random Measures

Abstract

We introduce an instance of a varying weight dependent Dirichlet process (DDP) model to implement a semi-parametric GLM. The model extends the recently developed semi-parametric generalized linear model (SPGLM) by adding a nonparametric Bayesian prior on the centering distribution of the GLM. We show that the resulting model takes the form of an inhomogeneous normalized random measure that arises from exponential tilting of a normalized random measure. Building on familiar posterior simulation methods for mixtures with respect to normalized random measures we introduce modification to implement posterior simulation in the resulting semi-parametric GLM model.

1 Introduction

We introduce a semi-parametric Bayesian extension of the semi-parametric GLM introduced in Rathouz and Gao [2009]. Under the proposed model the marginal distribution conditional on a given covariate takes the form of an inhomogeneous normalized random measure (NRM) (Regazzini et al., 2003). The joint model (across covariates x) is a variation of the popular dependent Dirichlet process (DDP) model (MacEachern, 2000; Quintana et al., 2022), replacing the marginal DP by an exponentially tilted DP with varying weights across covariates. We discuss the model construction, including a formal statement of the mentioned representations as NRM and DDP model, and characterize the posterior law. Appropriate extensions of the results in James et al. [2009] allows for straightforward posterior simulation. We validate the proposed model with a simulation study and illustrate it with an application.

SPGLM We first briefly review the semi-parametric GLM introduced in Rathouz and Gao [2009]. Consider a GLM

$$y \sim p(y \mid x) = p_x(y) \propto \exp(\theta_x y)\widetilde{\mu}(y) \tag{1}$$

with response $y \in \mathcal{Y} \subset \mathbb{R}$ and a p-dimensional covariate vector $x \in \mathcal{X}$ and (log) normalization constant

$$b(\theta_x) = \log \int_{\mathcal{Y}} \exp(\theta_x y)\widetilde{\mu}(dy). \tag{2}$$

In anticipation of the upcoming discussion we allow $\widetilde{\mu}(y)$ to be an un-normalized positive measure, implying a baseline density $f_{\widetilde{\mu}} = \widetilde{\mu}/\widetilde{\mu}(\mathcal{Y})$ in the GLM (1). Whilein the classical GLM the baseline distribution is assumed to be in a parametric family, in the semi-parametric SPGLM model the random measure $\widetilde{\mu}(y)$ itself becomes an unknown parameter. As in the classical GLM, we introduce a linear predictor $\eta = x^T\beta$, [Q: generalise η?] and a link function g to implicitly define θ by requiring $\lambda = \mathbb{E}(y \mid x) = g^{-1}(\eta)$. That is,

$$\lambda(x) = \mathbb{E}(y \mid x) = b'(\theta_x) = \frac{\int_{\mathcal{Y}} y \exp(\theta_x y)\widetilde{\mu}(dy)}{\int_{\mathcal{Y}} \exp(\theta_x y)\widetilde{\mu}(dy)}. \,, \tag{3}$$

Noting that $b'(\theta)$ is a strictly increasing function of θ we have $\theta_x = b'^{-1}(\lambda; \widetilde{\mu}) = \theta(\beta, \widetilde{\mu}, x)$. Here we added $\widetilde{\mu}$ to the arguments of b'^{-1} to highlight the dependence on $\widetilde{\mu}$. Alternatively, when we want to highlight dependence on β and x, indirectly through λ, we write $\theta(\beta, \widetilde{\mu}, x)$.

The defining characteristic of the SPGLM is a nonparametric baseline or reference distribution $f_{\widetilde{\mu}}$ that replaces a parametric specification in the classical GLM such as binomial or Poisson distribution. Keeping $f_{\widetilde{\mu}}$ nonparametric instead, the analyst needs to only specify the linear predictor and link function, even avoiding a variance function, leaving model specification less onerous than even with quasilikehood (QL) models, while still yielding a valid likelihood function. Beyond the initial introduction of the SPGLM by Rathouz and Gao [2009], which focused primarily on the finite support case, Huang [2014] characterized the SPGLM in the infinite support case, and Maronge et al. [2023] discussed the use with outcome-dependent or generalized case-control sampling. Despite these developments, there are still many important gaps in the literature. These include inference for application-driven functionals of the fitted models such as exceedance probabilities, which are crucial in clinical diagnostic (Paul et al., 2021); natural hazard detection (Kossin et al., 2020); financial risk management (Taylor and Yu, 2016); or in general, any decision-making setting. These inference problems are not straightforward to address with maximum likelihood based approaches. We address these gaps by developing a non-parametric Bayesian (BNP) extension of the SPGLM. In this BNP model we introduce $\widetilde{\mu}$ as an (un-normalized) positive random measure, implicitly defining an exponentially tilted DP prior for p_x.

Roadmap In Section 2.1 we introduce the proposed semiparametric Bayesian extension of the SPGLM, and characterize it as a variation of the popular DDP model in Section 2.2, and in Section 2.3 we show a representation of the implied marginal for one covariate as an inhomogeneous NRM. In Section 3 we characterize the posterior distribution under the DP–SPGLM by showing it to be conditionally conjugate given auxiliary variables similar to the construction used in James et al. [2009]. Section 4 summarizes a simulation study. Section 5 discusses an application, and Section 6 concludes with a final discussion.

2 The DP–SPGLM Model and Related Literature

2.1 A Bayesian semiparametric SPGLM

We extend (1)–(3) to a Bayesian inference model by adding a prior model for all unknown parameters, including in particular the baseline density $f_{\widetilde{\mu}}(\cdot) \equiv \widetilde{\mu}/\widetilde{\mu}(\mathcal{Y})$. Prior models for random probability measures like $f_{\widetilde{\mu}}$ are known as non-parametric Bayesian models (BNP). The most widely used BNP model is the Dirichlet process (DP) prior introduced in the the seminal work of Ferguson [1973]. The DP prior is characterized by two parameters: a concentration parameter α and a base distribution G_0. We write $F \sim \mathrm{DP}(\alpha G_0)$ or $F \sim \mathrm{DP}(\alpha, G_0)$. One of the many defining properties of the DP is the stick-breaking representation of Sethuraman [1994] for $G \sim \mathrm{DP}(\alpha G_0)$ as

$$G \equiv \sum_{h=1}^{\infty} \omega_h \delta_{m_h}(\cdot) \tag{4}$$

with atoms $m_h \overset{\text{iid}}{\sim} G_0$, and weights $\omega_h = v_h \prod_{\ell<h}(1 - v_\ell)$, where $v_h \overset{\text{iid}}{\sim} Be(1, \alpha)$. We use the terms atoms and locations interchangeably throughout the paper, and similarly for weights and jumps.

An alternative defining property of the DP prior is as a normalized completely random measure. A completely random measure (CRM) is a random measure $\widetilde{\mu}$ with the property that the random measure for any two non-overlapping events A, B are independent, that is $\widetilde{\mu}(A) \perp \widetilde{\mu}(B)$ when $A \cap B = \emptyset$ (Kingman, 1967). A CRM is characterized by its Laplace transform $\mathbb{E}[\exp\{\int h(y)\widetilde{\mu}(dy)\}]$, which in turn is completely characterized by the Lévy intensity $\nu(ds, dy)$ that appears in the Lévy-Khintchine representation

$$\mathbb{E}\left[e^{-\int_{\mathcal{Y}} h(y)\mu(dy)}\right] = \exp\left[-\int_{R^+ \times \mathcal{Y}} \left\{1 - e^{-sf(y)}\right\} \nu(ds, dy)\right]. \tag{5}$$

If ν factors as $\nu(s, y) = \rho(s)\, G_0(y)$ the CRM is known as a homogeneous CRM. Regazzini et al. [2003] introduced the wide class of normalized random measures (NRM) by defining a BNP prior for a random probability measure $f_{\widetilde{\mu}}$ as $\widetilde{\mu}/\widetilde{\mu}(\mathcal{Y})$ with a CRM $\widetilde{\mu}$. The DP prior is one example of an NRM prior as a normalized gamma CRM with Lévy intensity

$$\nu(ds, dy) = \frac{e^{-s}}{s} ds \cdot \alpha G_0(dy) \tag{6}$$

for a $DP(\alpha, G_0)$. We use a gamma CRM for $\widetilde{\mu}$ in the SPGLM (1), with base measure G_0 on the support $\mathcal{Y}$ and concentration parameter α. This implies a DP prior on the baseline density $f_{\widetilde{\mu}}$. We add a normal prior on β to complete a prior model

$$\widetilde{\mu} \sim \text{ gamma CRM}(\nu) \text{ with } \nu(ds, dy) = \frac{e^{-s}}{s} ds \cdot \alpha G_0(dy)$$
$$\beta \sim \text{ MVN}(\mu_\beta, \Sigma_\beta) \tag{7}$$

This implies a prior on $\mathcal{F} = \{p_x : x \in \mathcal{X}\}$. We add two more extensions. First, we add a convolution with a continuous kernel $K(\cdot)$ to define a continous sampling model for y. Using a symmetric kernel $K(\cdot)$, this does not change the GLM mean regression structure, as $\mathbb{E}(y_i \mid x_i) = \mathbb{E}_{z_i \mid x_i} \mathbb{E}(y_i \mid x_i, z_i) = g^{-1}(x_i' \beta)$. Second, we add an additional layer for θ_x. The latter is for computational reasons that will become clear when we discuss posterior simulation.

For reference we state the complete hierarchical model:

$$y_i \mid z_i \sim K(y_i \mid z_i); \tag{8}$$
$$z_i \mid x_i = x, \theta_x, \widetilde{\mu} \sim p_x(z_i) \propto \exp(\theta_x z_i) \widetilde{\mu}(z_i)$$
$$\theta_x \mid \beta, \widetilde{\mu} \sim p(\theta_x \mid \widetilde{\theta}_x), \text{ with } b'(\widetilde{\theta}_x) = g^{-1}(x'\beta) = \lambda(x)$$
$$\widetilde{\mu} \sim \text{ gamma CRM}(\nu), \text{ with } \nu(ds, dz) = \frac{e^{-s}}{s} ds \cdot \alpha G_0(dz)$$
$$\beta \sim \text{ MVN}(\mu_\beta, \Sigma_\beta) \tag{9}$$

We refer to the proposed model (8) as DpGLM. Also, we refer to $\widetilde{\mu}_i = \widetilde{\mu}(z; \theta_{x_i}) := \exp(\theta_{x_i} z) \widetilde{\mu}(z)$ as the *titled CRM*, with tilting parameter θ_{x_i}.

2.2 A varying weights DDP

MacEachern [2000] first introduced the *dependent* Dirichlet process (DDP) by extending the DP model to a family of random distributions $\{G_x : x \in \mathcal{X}\}$. The construction starts by assuming marginally, for each x, a DP prior for each $G_x = \sum w_{xh} \delta_{m_{xh}}$. The desired dependence can then be accomplished by using shared $w_{xh} = w_h$ and defining a dependent prior for $\{m_{xh},\ x \in X\}$ while maintaning independence across h, as required for the marginal DP prior. This defines the *common weights DDP*. Alternatively one can use common atoms m_h with a dependent prior on varying weights $\{w_{xh},\ x \in X\}$ (*common atoms DDP*), or use varying weights and atoms. See, for example, Quintana et al. [2022] for a review of the many difference specific instances of DDP models. A commonly used version are common weights and Gaussian process (GP) priors for $\{m_{xh},\ x \in X\}$, independently across h (MacEachern, 2000).

In the proposed DP–SPGLM approach (8), dependence is introduced naturally through the weights w_{xh} while keeping atoms m_h constant across x. Starting from the stick-breaking representation (4) for a (single) DP prior we define $p_x(z)$ as follows:

$$p_x(z) = \exp\{\theta_x z - b(\theta_x)\}\, \widetilde{\mu}(z) = \exp\{\theta_x z - b(\theta_x)\} \sum_{h=1}^{\infty} \omega_h \delta_{m_h}(z)$$
$$= \sum_{h=1}^{\infty} [\exp\{\theta_x z - b(\theta_x)\}\, \omega_h]\, \delta_{m_h}(z) = \sum_{h=1}^{\infty} w_{xh} \delta_{m_h}(z),$$

where $w_{xh} = \exp\{\theta_x z - b(\theta_x)\}\, \omega_h$, depends on x implicitly through θ_x. The model defines a variation of a DDP model using comon atoms and varying weights. However, the exponential tilting in (**??**) defines a marginal prior p_x beyond a DP model, as we shall discuss next in more detail.

2.3 The marginal model

The implied marginal model $p_x(z)$ for given covariate x in (8) can be shown to be an NRM again. This is seen by noting that the Laplace transform of p_x takes the form of (5) again, allowing us to recognize the NRM by inspection of the Lévy intensity in (8).

Proposition 1 *[Nieto-Barajas et al., 2004][1] Consider the DP-SPGLM with implied marginal* $p_x(z) \propto \exp(\theta_x z)\widetilde{\mu}(z)$*, assuming a gamma CRM* (6)*, i.e.,* $f_{\widetilde{\mu}} \sim DP(\alpha, G_0)$ *and given* θ_x*. Then* p_x is a non-homogeneous normalized random measure (NRM) with Lévy intensity,

$$\nu(ds, dz) = \frac{1}{s}\, e^{-s/\exp(\theta_x\, z)} ds \cdot \alpha G_0(dz) \tag{10}$$

The Lévy intensity ν in (10) characterizes an inhomogeneous NRM, with the distributions of the jumps depending on the locations as $\rho(ds \mid z) = \frac{e^{-s/\exp(\theta_x z)}}{s} ds$.

The use of the DP prior for $f_{\widetilde{\mu}}$ made the result in (10) particularly simple, allowing a closed form expression. A similar result, albeit not necessarily in closed form anymore, is true under any other NRM prior for $f_{\widetilde{\mu}}$. For example, Lijoi et al. [2007] argue for the richer class of normalized generalized gamma, which includes the DP as a special case. One common reason to consider alternatives to the DP prior is the lack of flexibility in modeling the random partition implied by ties of a sample from a DP random measure. In the context of (8) the discrete nature of $\widetilde{\mu}(\cdot)$ gives rise to ties of the z_i. Under $\theta_x = 0$ the random partition characterized by the configuration of ties is known as the Chinese restaurant process. It is indexed by a single hyperparameter, α. De Blasi et al. [2013], for example, argue that the nature of this random partition is too restrictive for many applications. However, in the context of the DPGLM the random partition is not an inference target, and we shall never interpret the corresponding clusters, leaving the DP prior as an analytically and computationally appealing prior choice for $\widetilde{\mu}$.

[1] E: please read and make sure it's the right ref - think so

The BNP prior for $p_x(z)$ and the kernel in the first two levels of the DPGLM model (8) define a variation of popular BNP mixture models. The use of the particular NRM with Lévy intensity (10) arises naturally in the context of the GLM-style regression with the exponential tilting. Posterior simulation for BNP mixtures with NRM priors on the mixing measure is discussed, for example, in Argiento et al. [2010], Barrios et al. [2013] or Favaro and Teh [2013]. However, the GLM regression introduces a complication by applying different exponential tilting for each unique covariate x_i. This leads to some variations in the posterior characterization and the corresponding posterior simulation algorithms. We next discuss those changes.

3 Posterior characterization

Let $\mathcal{D}_n$ denote the observed data $\{x_i, y_i\}_{i=1}^n$, with $x_i \in \mathcal{X} \subset R^p$ and $y_i \in \mathcal{Y} \subset R$, and (8) adds the latent variables z_i. For simplicity we write θ_i for θ_{x_i}, and define T_i as the total mass for the tilted CRM $\widetilde{\mu}_i$ as $T_i = \int_{\mathcal{Y}} \exp(\theta_i z)\widetilde{\mu}(dz)$. We can then adapt the results from James et al. [2009; Section 2] to characterize the posterior distribution under the DP-GLM model (8).

We first introduce a data augmentation with auxiliary variables u_i, using one auxiliary variable for each unique covariate vector x_i. For the moment we assume n unique covariate vectors (and shall comment later on simple modifications to accommodate the more general case). We define

$$u_i \sim \gamma_i / T_i$$

for independent exponential random variables γ_i, implying $p(u_i \mid T_i) = \mathrm{Ga}(1, T_i)$. We first state the conditional posterior for $\mathbf{u} = (u_1, \ldots, u_n)$, conditional on z_i, but marginalizing w.r.t. $\widetilde{\mu}$ (and thus T_i).

Proposition 2 *Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and $\mathbf{z} = (z_1, \ldots, z_n)$. Then, the complete conditional of $\mathbf{u}$ is given by*

$$p(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{z}) \propto \exp\left\{-\int_{\mathcal{Y}} \ln\left[1 + \sum_{i=1}^{n} u_i \exp(\theta_i z)\right] G_n(dz)\right\},$$

where $G_n = \alpha G_0 + \sum_{i=1}^n \delta_{z_i}$.

The proof is implied as part of the proof for the next result. As mentioned, the discrete nature of $\widetilde{\mu}$ introduces ties in z_i. Let $\{z_1^\star, \ldots, z_k^\star\}$ denote the unique values among the currently imputed $\{z_1, \ldots, z_n\}$, with multiplicity $\{n_1^\star, \ldots, n_k^\star\}$. Then G_n in Result 2 can be written as $G_n = \alpha G_0 + \sum_{\ell=1}^k n_\ell^\star \delta_{z_\ell^\star}$. Clearly, $\sum_{\ell=1}^k n_\ell^\star = n$.

We next characterize the posterior of $\widetilde{\mu}$ given $\mathbf{u}$ and $\boldsymbol{\theta}$.

Proposition 3 *Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, and $\mathbf{z} = (z_1, \ldots, z_n)$ with k unique values $z^\star_\ell$, $\ell = 1, \ldots, k$, with multiplicities $n^\star_\ell$. Then $\widetilde{\mu}$ includes atoms at the $z^\star_\ell$ with random probability masses J_ℓ. Letting $\widetilde{\mu}^o$ denote the remaining part of $\widetilde{\mu}$ we have*

$$\widetilde{\mu} \mid \mathbf{u}, \mathbf{z}, \boldsymbol{\theta} \overset{d}{=} \widetilde{\mu}^o + \sum_{\ell=1}^{k} J_\ell \delta_{z^\star_\ell} \ ,$$

where

1. $\widetilde{\mu}^o \overset{d}{=} CRM(\nu^o)$ *with Lévy intensity* $\nu^o(ds, dz) = \frac{1}{s}\, e^{-\left(1+\sum_{i=1}^n u_i \exp(\theta_i z)\right)s} ds \cdot \alpha G_0(dz)$

2. *Let* $\Psi(z^\star_\ell; \mathbf{u}, \boldsymbol{\theta}) = 1 + \sum_{i=1}^n u_i \exp(\theta_i z^\star_\ell)$*. Then*

$$p\left(J_\ell \mid \mathbf{u}, \boldsymbol{\theta}, z^\star_\ell, n^\star_\ell\right) \propto s^{n^\star_\ell - 1}\, e^{-\left(1+\sum_{i=1}^n u_i \exp(\theta_i z^\star_\ell)\right)s} \equiv Ga\left(n^\star_\ell,\ \Psi(z^\star_\ell; \mathbf{u}, \boldsymbol{\theta})\right).$$

$\ell = 1, \ldots, k$

Result 3 shows that given $\mathbf{z}, \boldsymbol{\theta}$ and auxilliary variables $\mathbf{u}$, *a posteriori* $\widetilde{\mu}$ is again a CRM. To be precise, it is a sum of two components. One part is an inhomogeneous CRM $\widetilde{\mu}^o = \sum_{\ell=1}^\infty \widetilde{J}_\ell \delta_{\widetilde{z}_\ell}$ with Lévy intensity ν^o. The random locations $\widetilde{z}_\ell$ and jumps $\widetilde{J}_\ell$ can be generated using, for example, the Ferguson and Klass [1972] algorithm. The second component is finite discrete measure with gamma distributed random jumps J_ℓ at fixed atoms $z^\star_\ell$.

Finally, in the general case with possible ties of the covariate vectors x_i, one could still use the same results, with n auxiliary variables u_i. Alternatively, the following construction could be used with fewer auxiliary variables. Let ξ_j, $j = 1, \ldots, J$ denote the unique covariate combinations with multiplicities a_j. Let then T_j denote the normalization constant under covariate $x = \xi_j$. Similar results as above hold, starting with latent variables $u_j \sim \mathrm{Ga}(a_j, T_j)$.

[Peter: Updated till this point.]

4 Simulation Studies

We proceed with simulation studies to evaluate the operating characteristics of the DP-GLM model. Our investigation addresses the following key questions:

(Q1) How does the model perform in terms of predictive accuracy when estimating the baseline density, $f_{\widetilde{\mu}}(y)$, under various scenarios?

(Q2) Do the credible intervals for $f_{\widetilde{\mu}}(y)$ achieve coverage rates close to their nominal levels?

(Q3) In scenarios where the response is independent of predictors, does $\theta_{x;n} := [\theta_x \mid \mathcal{D}_n]$ converge in probability to zero, or alternatively, do the credible intervals for θ_x attain nominal coverage rates?

(Q4) Do the credible intervals for β_j parameters attain nominal coverage? How is their predictive accuracy?

We consider a data generating mechanism where the response y is sampled from the Speech Intelligibility dataset (for dataset details, see Section 5).

- Null case (setting I): Let $f_{\widetilde{\mu}}^{(kde)}$ denote the kernel density estimate based on the response data from Speech Intelligibility dataset (ignoring the covariates). We consider $f_{\widetilde{\mu}}^{(kde)}$ as the simulation truth for the baseline density $f_{\widetilde{\mu}}$. Covariates are generated as: $x_{1i} = 1, x_{2i} \sim Unif(a_1, a_2), x_{3i} \sim Unif(b_1, b_2)$, where we take $a_1 = 1, a_2 = 3, b_1 = -1, b_2 = 1$. We sample y independent of x i.e, $y_i \sim f_{\widetilde{\mu}}^{(kde)}$. We use $\mathcal{D}_n$ to refer the observed data $\{x_i, y_i\}_{i=1}^n$. This setting aims to address Q1−Q3.

- Point masses (setting II): Let $f_{\widetilde{\mu}}^{(Beta)}$ denote the Beta(a, b) density estimate based on the response data from Speech Intelligibility dataset (ignoring the covariates). We consider $f_{\widetilde{\mu}}^{(Beta)}$ as the simulation truth for $f_{\widetilde{\mu}}$, with additional point masses at $y = 0$ and $y = 1$. The rest is same as in Setting I. Apart from Q1−Q3, the primary objective here is to assess whether the model accurately estimates the point masses.

- Regression (setting III): We consider the same framework as in Setting I, with one modification: the sampling y is now dependent on x. Specifically, we sample $y_i \sim p(y_i \mid x_i) \propto \exp(\theta_{x_i} y_i) f_{\widetilde{\mu}}^{(kde)}(y_i)$, where $\theta_x \sim \text{Normal}(\widetilde{\theta}_x, \sigma_\theta^2)$. Here, $\widetilde{\theta}_x = b'^{-1}(g^{-1}(\eta_x))$, with $\eta_x = x^T\beta$. We set $\beta_1 = 0.1, \beta_2 = 0.25, \beta_3 = 0.75$. This setting aims to address Q1, Q2 and Q4.

5 Application to Speech Intelligibility Data

We further study performance of the proposed model on a speech intelligibility dataset for typically developing (TD) children from 30 to 96 months of age. The study included $n = 505$ TD children, wherein *mean speech intelligibility* was measured as the proportion of words correctly transcribed by the evaluating adult listeners. The *mean speech intelligibility* (MSI) was recorded separately for *single-word* and *multi-word* utterances, and we refer them as SW-MSI and MW-MSI, respectively. For further details on the dataset, we refer to (Mahr et al., 2020; Hustad et al., 2021).

5.1 Study of single word intelligibility

We use a $3 - df$ natural cubic splines on age for modeling logit mean intelligibility.

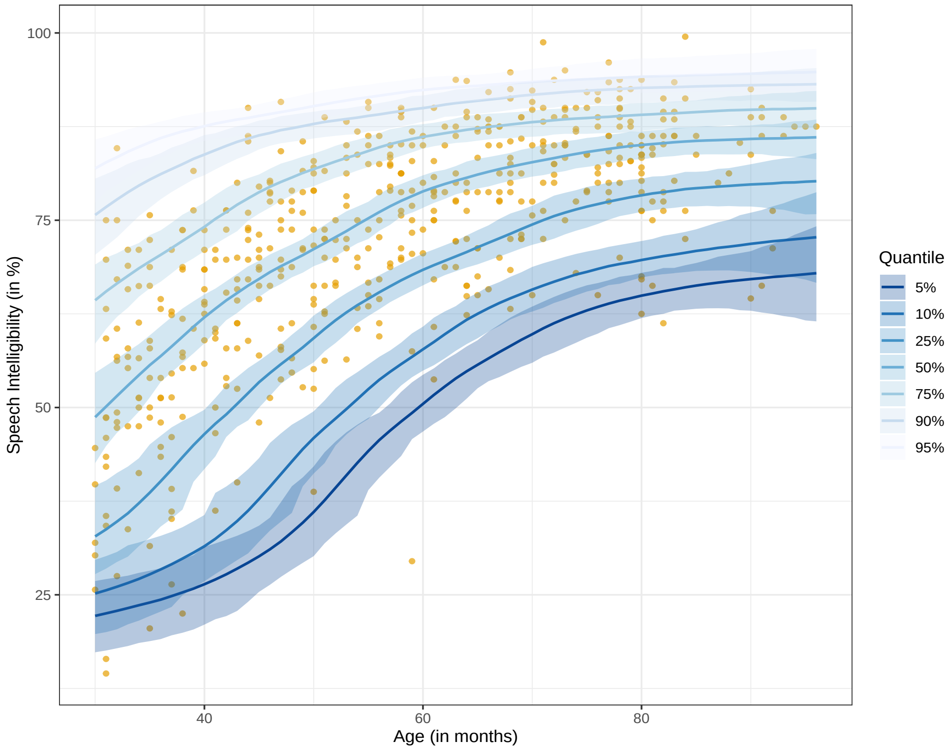

Figure 1: Quantile regression, $q_\alpha(\boldsymbol{x})$, with 90% point-wise uncertainty intervals. $K = unif(z-c, z+c)$, $c = 0.025$ and $G_0 = unif(0,1)$. Intelligibility type: *single-word*.

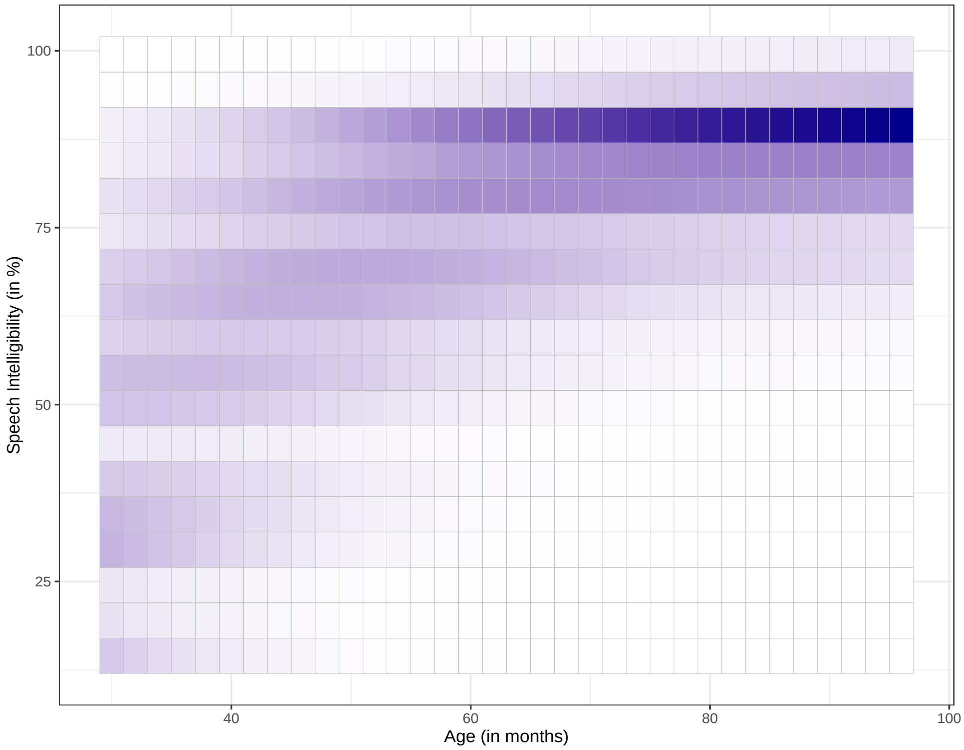

Figure 2: Heatmap for fitted probabilities, $\widehat{p}(y \mid x)$. y: Intelligibility (in %) and x: Age (in months). Here 'white' to 'blue' represents an increase in $\widehat{p}(y \mid x)$. Intelligibility type: *single-word*.

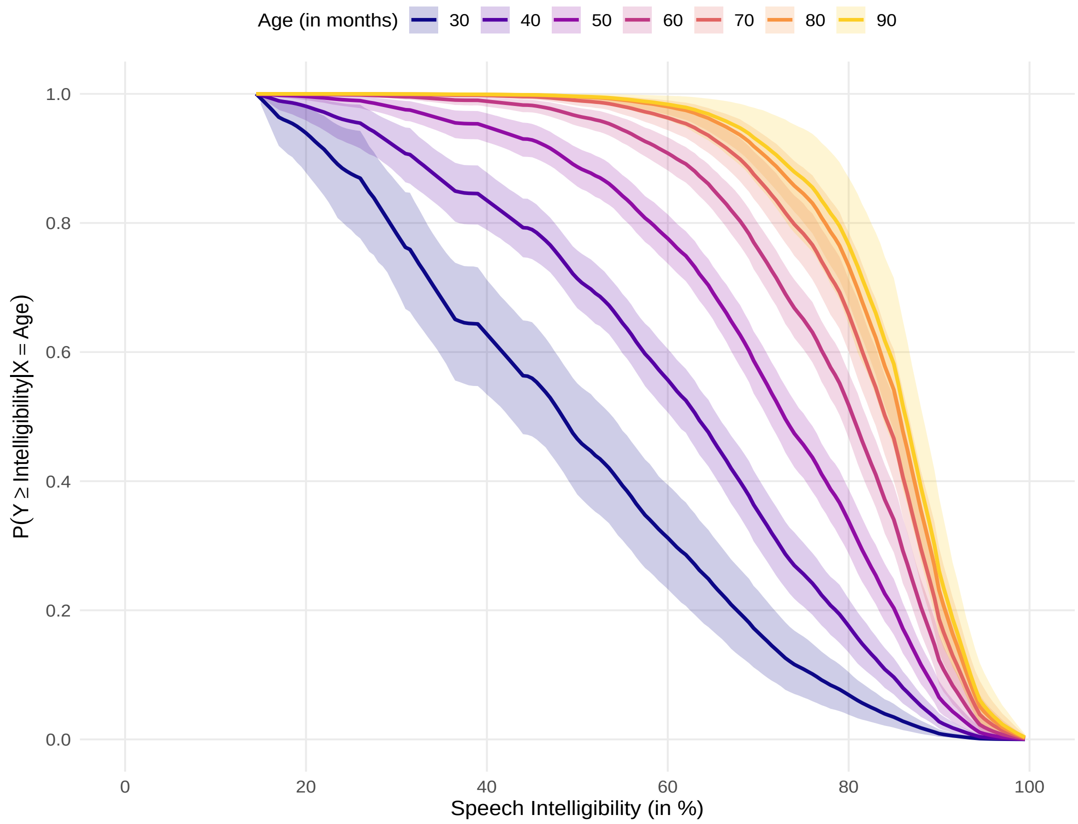

Figure 3: Estimate for exceedance probabilities, $\widehat{p}(y \geq y_0 \mid x)$, with 90% point-wise uncertainty intervals.. y: Intelligibility (in %) and x: Age (in months). Intelligibility type: *single-word*.

5.2 Study of multi word intelligibility

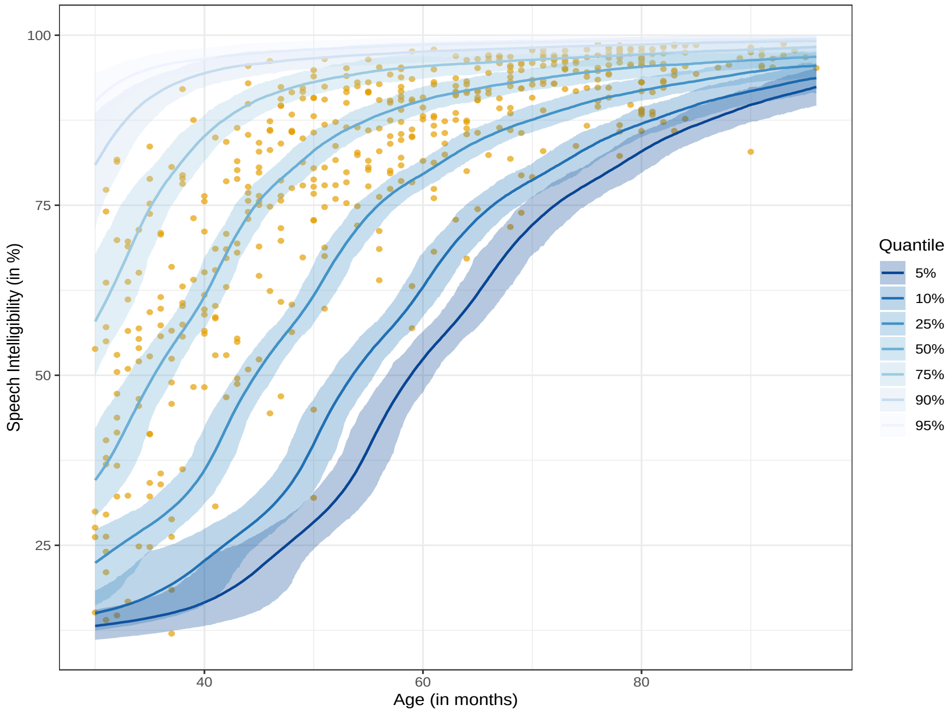

Figure 4: Quantile regression, $q_\alpha(\boldsymbol{x})$, with 90% point-wise uncertainty intervals. $K = unif(z-c, z+c)$, $c = 0.025$ and $G_0 = unif(0,1)$. Intelligibility type: *multi-word*.

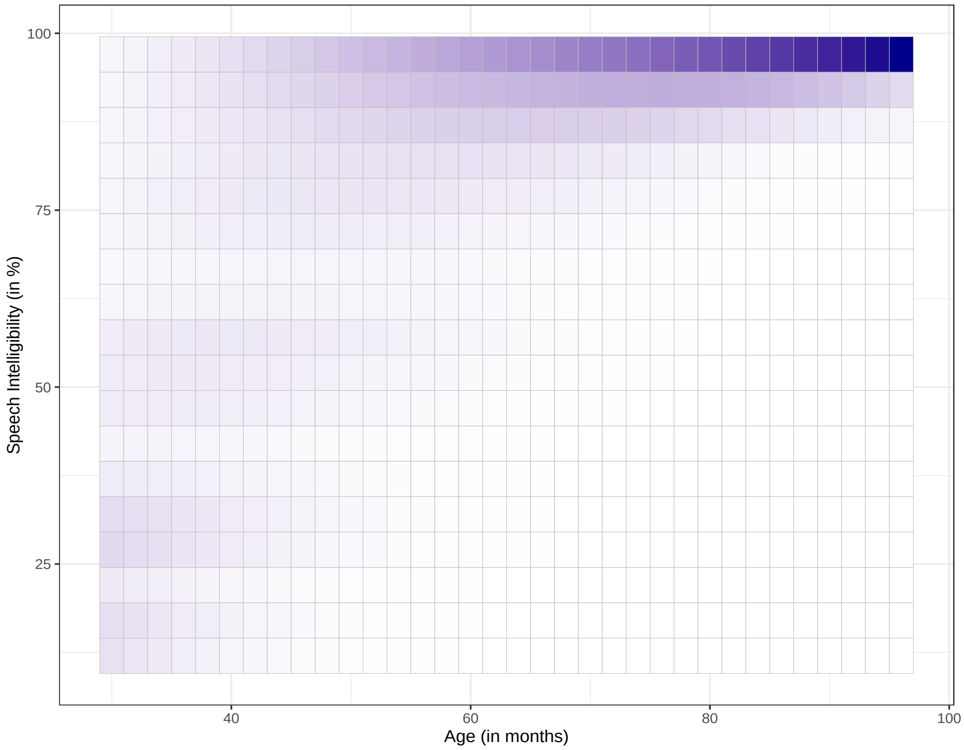

Figure 5: Heatmap for fitted probabilities, $\widehat{p}(y \mid x)$. y: Intelligibility (in %) and x: Age (in months). Here 'white' to 'blue' represents an increase in $\widehat{p}(y \mid x)$. Intelligibility type: *multi-word*.

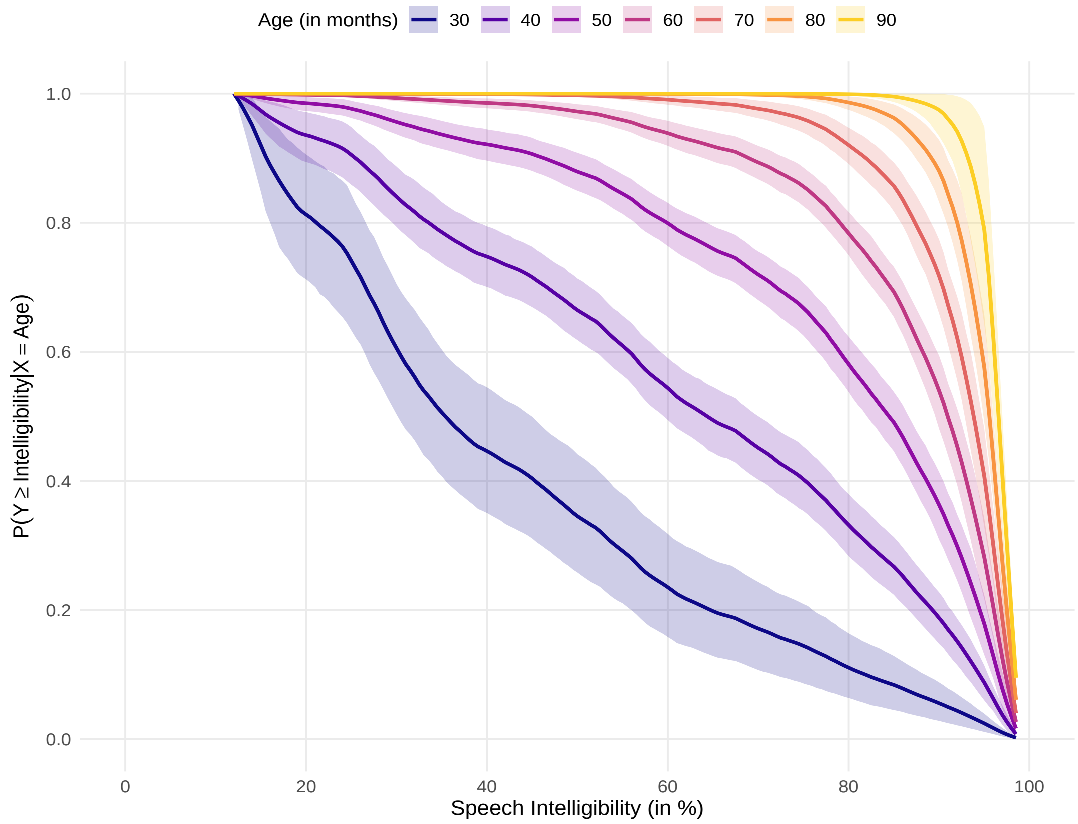

Figure 6: Estimate for exceedance probabilities, $\widehat{p}(y \geq y_0 \mid x)$, with 90% point-wise uncertainty intervals.. y: Intelligibility (in %) and x: Age (in months). Intelligibility type: *multi-word*.

6 Discussion

Supplementary Material

References

Raffaele Argiento, Alessandra Guglielmi, and Antonio Pievatolo. Bayesian density estimation and model selection using nonparametric hierarchical mixtures. *Computational Statistics & Data Analysis*, 54(4): 816–832, 2010.

E. Barrios, L. E. Nieto-Barajas, and I. Prünster. A study of normalized random measures mixture models. *Statistical Science*, page to appear, 2013.

Pierpaolo De Blasi, Stefano Favaro, Antonio Lijoi, Ramsés H Mena, Igor Prünster, and Matteo Ruggiero. Are gibbs-type priors the most natural generalization of the dirichlet process? *IEEE transactions on pattern analysis and machine intelligence*, 37(2):212–229, 2013.

S. Favaro and Y. W. Teh. MCMC for normalized random measure mixture models. *Statistical Science*, page to appear, 2013.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

Thomas S Ferguson and Michael J Klass. A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics*, 43(5):1634–1643, 1972.

Alan Huang. Joint estimation of the mean and error distribution in generalized linear models. *Journal of the American Statistical Association*, 109(505):186–196, 2014.

Katherine C Hustad, Tristan J Mahr, Phoebe Natzke, and Paul J Rathouz. Speech development between 30 and 119 months in typical children i: Intelligibility growth curves for single-word and multiword productions. *Journal of Speech, Language, and Hearing Research*, 64(10):3707–3719, 2021.

Lancelot F James, Antonio Lijoi, and Igor Prünster. Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, 36(1):76–97, 2009.

John Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

James P Kossin, Kenneth R Knapp, Timothy L Olander, and Christopher S Velden. Global increase in major tropical cyclone exceedance probability over the past four decades. *Proceedings of the National Academy of Sciences*, 117(22):11975–11980, 2020.

Antonio Lijoi, Ramses H. Mena, and Igor Prünster. Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):715–740, 2007.

S N MacEachern. Dependent Dirichlet processes. Technical report, Department of Statistics, The Ohio State University, 2000.

Tristan J Mahr, Paul J Rathouz, and Katherine C Hustad. Longitudinal growth in intelligibility of connected speech from 2 to 8 years in children with cerebral palsy: A novel bayesian approach. *Journal of Speech, Language, and Hearing Research*, 63(9):2880–2893, 2020.

Jacob M Maronge, Ran Tao, Jonathan S Schildcrout, and Paul J Rathouz. Generalized case-control sampling under generalized linear models. *Biometrics*, 79(1):332–343, 2023.

Luis E Nieto-Barajas, Igor Prünster, and Stephen G Walker. Normalized random measures driven by increasing additive processes. *Annals of Statistics*, 32:2343–2360, 2004.

Rajib Paul, Oluwaseun Adeyemi, Subhanwita Ghosh, Kamana Pokhrel, and Ahmed A Arif. Dynamics of Covid-19 mortality and social determinants of health: a spatiotemporal analysis of exceedance probabilities. *Annals of Epidemiology*, 62:51–58, 2021.

Fernando A Quintana, Peter Müller, Alejandro Jara, and Steven N MacEachern. The dependent Dirichlet process and related models. *Statistical Science*, 37(1):24–41, 2022.

Paul J Rathouz and Liping Gao. Generalized linear models with unspecified reference distribution. *Biostatistics*, 10(2):205–218, 2009.

Eugenio Regazzini, Antonio Lijoi, and Igor Prünster. Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31(2):560–585, 2003.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

James W Taylor and Keming Yu. Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 179(4):1069–1092, 2016.

Acknowledgement

Acknowledge Igor for Proposition 1.

Appendix