Semiparametric Vector Generalized Linear Models

Estimation and Computation

Gabriel Dennis

Honours Talk

Supervisor: Dr Alan Huang

28th of May 2021

Outline of talk

Table of Contents

VGLMs

A VGLM (Vector Generalized Linear Model) aims to generalize GLMs to multivariate responses, however, multivariate generalizations of common families of probability distributions (e.g., poisson, gamma, ...) are difficult to construct.

Motivating Example (Butterfly Dataset)

This dataset contains 66 counts of 14 species of butterflies in Boulder Colorado, USA.

Table 1: Snippet of Butterfly Counts for the 3 most common species

site	Pieris.rapae	Colias.philodice	Colias.eurytheme	building	vegetation	habitat
1	0	0	0	2.1	10.8	mixed
2	1	1	2	2.1	10.8	mixed
3	1	1	2	19.8	1.7	mixed
4	0	2	1	5.3	0	mixed
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

We would like to fit a joint model, as species could complement or compete with each other within a site.

Motivating Example (Butterfly Dataset)

Current approach is to fit independent poisson regression models for each species, using geographic features of each site as covariates.

We would like to fit a joint model, as species could complement or compete with each other within a site.

However, no multivariate generalisation of the poisson distribution allows for both positive and negatively correlated response components.

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship
- observed data often exhibit over-or under-dispersion relative to the postulated model

Problems Modelling with parametric GLMs and VGLMs

Common problems that occur with parametric modelling

- requires correct specification of the error distribution
- requires correct specification of the mean-variance relationship
- observed data often exhibit over-or under-dispersion relative to the postulated model

For multivariate data, problems arise when specifying a parametric distribution and the response covariance structure. For most cases there is no standard parametric model which could produce a given dataset.

Table of Contents

Vector Generalisation

This model is a vector generalisation of a Semiparametric Generalized Linear Model [1], which is based on a exponential tilt reformulation of the joint distribution from which the data originates for which the response distribution F remains unspecified.

[1] (P. J. Rathouz and Gao 2009; Huang and P. Rathouz 2012; Huang 2014)

Vector Generalisation

This model is a vector generalisation of a Semiparametric Generalized Linear Model [1], which is based on a exponential tilt reformulation of the joint distribution from which the data originates for which the response distribution F remains unspecified.

This allows arbitrary nonparametric response distribution parameter F and the mean model parameters $\boldsymbol{\beta}$ to be simultaneously estimated without loss of information.

[1](P. J. Rathouz and Gao 2009; Huang and P. Rathouz 2012; Huang 2014)

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

Proposed Model

Given data

$$(\boldsymbol{Y}_i, \boldsymbol{X}_i) \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^K \times \mathbb{R}^q, i = 1, \ldots, n$$

Assume these are independent samples originating from some multivariate exponential family. The joint density can be written in the form

$$dF_i(\boldsymbol{y}) = \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y}), \quad i = 1, \ldots, n$$

where $dF_i(\boldsymbol{y})$ is an exponential tilt of the response distribution $dF(\boldsymbol{y})$, with the amount of tilting $\boldsymbol{\theta}$ determined by the mean $\boldsymbol{\mu}(\boldsymbol{X}_i^T \boldsymbol{\beta})$ of the observation $\boldsymbol{Y}_i$.

Model Tilt Parameters

The normalising constants $b_i = b(\boldsymbol{X}_i, \boldsymbol{\beta}, F)$ are defined to be

$$b(\boldsymbol{X}_i, \boldsymbol{\beta}, F) = -\log\left\{\int_{\mathcal{Y}} \exp\{\boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y})\right\},$$

Model Tilt Parameters

The normalising constants $b_i = b(\boldsymbol{X}_i, \boldsymbol{\beta}, F)$ are defined to be

$$b(\boldsymbol{X}_i, \boldsymbol{\beta}, F) = -\log\left\{\int_{\mathcal{Y}} \exp\{\boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y})\right\},$$

and the vector of tilt parameters $\boldsymbol{\theta}_i = \boldsymbol{\theta}(\boldsymbol{X}, \boldsymbol{\beta}, F) \in \mathbb{R}^K$ is the implicit solution to

$$\mu_{(k)}(\boldsymbol{X}_{(k)}^T \boldsymbol{\beta}_{(k)}) = \int_{\mathcal{Y}} y_{(k)} \exp\{b_i + \boldsymbol{\theta}_i^T \boldsymbol{y}\} dF(\boldsymbol{y}), \quad k = 1, \ldots, K$$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F|\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F|\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

One way to estimate the densities $dF(\boldsymbol{Y}_i)$ is by replacing them with the histogram estimators p_i, which are assigned to values in the observed support $\{\boldsymbol{Y}_i \in \mathbb{R}^k | i = 1, 2, \ldots, n\}$

Semiparametric Extension

This joint density has the log-likelihood

$$\ell(\boldsymbol{\beta}, F|\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} \left\{ \log dF(\boldsymbol{Y}_i) + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i \right\}$$

One way to estimate the densities $dF(\boldsymbol{Y}_i)$ is by replacing them with the histogram estimators p_i, which are assigned to values in the observed support $\{\boldsymbol{Y}_i \in \mathbb{R}^k | i = 1, 2, \ldots, n\}$

This produces empirical log-likelihood.

$$\ell(\boldsymbol{\beta}, \boldsymbol{p}) = \sum_{i=1}^{n} \log p_i + b_i + \boldsymbol{\theta}_i^T \boldsymbol{Y}_i$$

Estimation of β and $\boldsymbol{p}$

$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{p}})$ are then the joint maximisers of the empirical log likelihood.

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{p}}) = \arg\max \ell(\boldsymbol{\beta}, \boldsymbol{p})$$

Estimation of β and $\boldsymbol{p}$

Subject to empirical analogous of the normalising constraints

$$1 = \sum_{i=1}^{n} p_i \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\}, \quad j = 1, 2, 3, \ldots, n$$

Estimation of β and $\boldsymbol{p}$

Subject to empirical analogous of the normalising constraints

$$1 = \sum_{i=1}^{n} p_i \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\}, \quad j = 1, 2, 3, \ldots, n$$

and the mean constraints

$$\mu_{(k)}(\boldsymbol{X}_{(k)j}^T \boldsymbol{\beta}_{(k)}) = \sum_{i=1}^{n} p_i Y_{(k)i} \exp\{b_j + \boldsymbol{\theta}_j^T \boldsymbol{Y}_i\} \quad j = 1, \ldots, n, k = 1, \ldots, K$$

Parameter Inference

There have been some results shown for the conducting parameter inference in the univariate case of this model (Huang 2014; Huang and P. Rathouz 2017), which generalise to this model.

Parameter Inference

For testing the null hypothesis

$$H_0 : \beta_{(k)j} = 0,$$
$$k = 1, \ldots, K, j = 1, \ldots, q_k$$

Parameter Inference

For testing the null hypothesis

$$H_0 : \beta_{(k)j} = 0,$$
$$k = 1, \ldots, K, j = 1, \ldots, q_k$$

a Wald test can be used

$$\frac{\hat{\beta}_{(k)j}}{\sqrt{\hat{\Sigma}_{(k)j,(k)j}}} \sim t_{n-q_k}$$

Parameter Inference

For testing the null hypothesis

$$H_0 : \beta_{(k)j} = 0,$$
$$k = 1, \ldots, K, j = 1, \ldots, q_k$$

a Wald test can be used

$$\frac{\hat{\beta}_{(k)j}}{\sqrt{\hat{\Sigma}_{(k)j,(k)j}}} \sim t_{n-q_k}$$

where the standard errors are derived from the empirical log-likelihood

$$\hat{\Sigma} = \left\{ \mathrm{Var}\left[\nabla \ell(\boldsymbol{\beta}, \boldsymbol{p}) \right] \right\}^{-1}$$

Parameter Inference

For Point Hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}^*)\} \sim \chi^2_Q, \quad Q = \sum_{k=1}^{K} q_k$$

Parameter Inference

For Point Hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}^*)\} \sim \chi^2_Q, \quad Q = \sum_{k=1}^{K} q_k$$

Composite Hypothesis $H_0 : M\boldsymbol{\beta} = \gamma$,

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\hat{\boldsymbol{\beta}}_0)\} \sim \chi^2_r$$

Parameter Inference

For Point Hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}^*)\} \sim \chi^2_Q, \quad Q = \sum_{k=1}^{K} q_k$$

Composite Hypothesis $H_0 : M\boldsymbol{\beta} = \gamma$,

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\hat{\boldsymbol{\beta}}_0)\} \sim \chi^2_r$$

For finite samples the following adjustment can be made

$$2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\hat{\boldsymbol{\beta}}_0)\} \sim rF_{r,n-Q}$$

Table of Contents

Code

Currently the model is fit computationally in MATLAB using non-linear constrained optimization to maximise the empirical log-likelihood. [2]

[2] The code is available at GitHub https://github.com/gden173/vspglm

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i
- nK tilt parameters $\boldsymbol{\theta}_i$

Constraints

For n observations of a K dimensional response $\boldsymbol{Y}_i \in \mathbb{R}^K$ the optimization simultaneously solves for $Q + n(2 + K)$ parameters.

- Q regression parameters $\boldsymbol{\beta}$
- n probability masses p_i
- n normalising constants b_i
- nK tilt parameters $\boldsymbol{\theta}_i$

which are subject to $nK + n$ mean and normalising constraints.

Table of Contents

Butterfly Dataset

Contains 66 counts of 14 species of butterfly in Boulder Colorado, USA.

Table 2: Snippet of Butterfly Counts for the 3 most common species

site	Pieris.rapae	Colias.philodice	Colias.eurytheme	building	vegetation	habitat
1	0	0	0	2.1	10.8	mixed
2	1	1	2	2.1	10.8	mixed
3	1	1	2	19.8	1.7	mixed
4	0	2	1	5.3	0	mixed
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Butterfly Dataset

We can fit this model directly to the counts of the each butterfly species using separate mean models, without the need to specify any association between species.

Butterfly Dataset

We can fit this model directly to the counts of the each butterfly species using separate mean models, without the need to specify any association between species.

$$\mu_{(k)i} = \mathbb{E}[Y_{(k)i}|\boldsymbol{X}_i] = \exp\{\boldsymbol{X}_i^T \boldsymbol{\beta}_{(k)}\},$$
$$i = 1, 2, \ldots, 66, \quad k = 1, 2 \ldots, 14$$

Butterfly Dataset

```
butterfly = fit_vspglm(...
         ["colias.philodice ~ (building,urban,habitat)",...
         "pieris.rapae ~ (building, urban, habitat)", ...
         "colias.eurytheme ~ (building, urban, habitat)"], ...
         tbl, {'log', 'log', 'log'})
```

Figure 1: Formula Syntax to fit 3 separate unconstrained models using the same covariates and log link functions

Butterfly Dataset

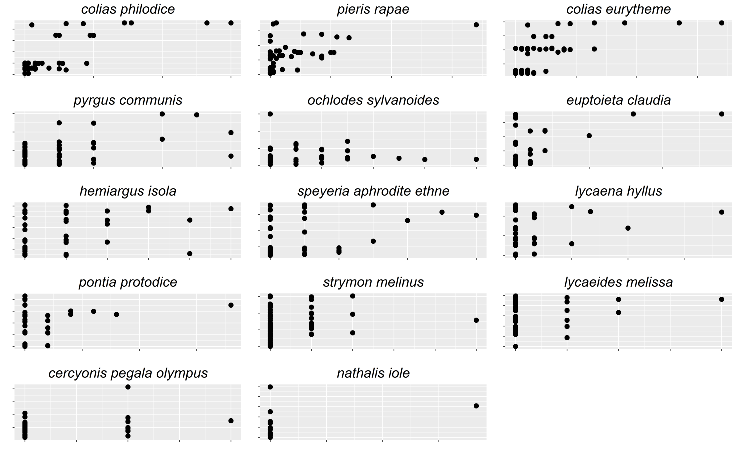

Figure 2: Predictions (y-axis) vs Observed counts (x-axis)

Butterfly Dataset

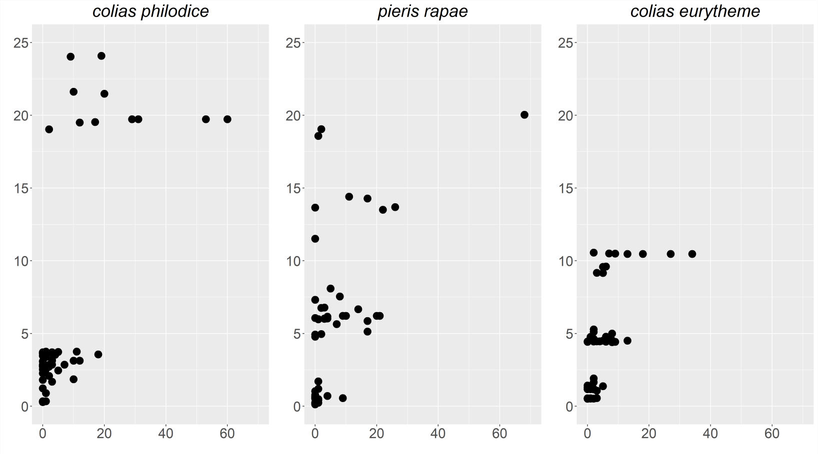

Figure 3: Predictions (y-axis) vs Observed counts (x-axis) for the 3 butterfly species with the greatest number of observed counts

Burns Injury Dataset

Contains 981 observations of patients admitted to hospital with 3rd degree burns.

Table 3: Snippet of Burns Dataset

	age	**burns**	**death**
1	3	6.9	1
2	39	7.4	1
3	42	3.9	1
4	21	6.1	1
⋮	⋮	⋮	⋮

Burns Injury Data Set

Response vector has two components $\boldsymbol{Y} = (Y_{\text{death}}, Y_{\text{burn}})$, which are binary and continuous, and a single covariate $\boldsymbol{X} = (X_{\text{age}})$.

Response vector has two components $\boldsymbol{Y} = (Y_{\text{death}}, Y_{\text{burn}})$, which are binary and continuous, and a single covariate $\boldsymbol{X} = (X_{\text{age}})$.
This suggests the two marginal mean models

$$\mu_{(1)} = \mathbb{E}[\text{death}|\text{age}] = \frac{\exp\{\beta_{(1)1} + \beta_{(1)2}\text{age}\}}{1 + \exp\{\beta_{(1)1} + \beta_{(1)2}\text{age}\}}$$

$$\mu_{(2)} = \mathbb{E}[\text{burns}|\text{age}] = \beta_{(2)1} + \beta_{(2)2}\text{age}$$

Burns Injury Data Set

Response vector has two components $\boldsymbol{Y} = (Y_{\text{death}}, Y_{\text{burn}})$, which are binary and continuous, and a single covariate $\boldsymbol{X} = (X_{\text{age}})$.
This suggests the two marginal mean models

$$\mu_{(1)} = \mathbb{E}[\text{death}|\text{age}] = \frac{\exp\{\beta_{(1)1} + \beta_{(1)2}\text{age}\}}{1 + \exp\{\beta_{(1)1} + \beta_{(1)2}\text{age}\}}$$

$$\mu_{(2)} = \mathbb{E}[\text{burns}|\text{age}] = \beta_{(2)1} + \beta_{(2)2}\text{age}$$

The code for fitting this model is

```
model = fit_vspglm(["death~age","burns~age"],tbl,{'logit','id'})
```

Burn Injury Data Set

This data set has been analysed several times using a variety of VGLMS, with Huang (2017) using GEEs to fit these mean models.

Burn Injury Data Set

This data set has been analysed several times using a variety of VGLMS, with Huang (2017) using GEEs to fit these mean models.

GEE model parameters

$$\hat{\mu}_{(1)} = \frac{\exp\{-3.6891 + 0.0508\text{age}\}}{1 + \exp\{-3.6891 + 0.0508\text{age}\}}$$

$$\hat{\mu}_{(2)} = 6.7118 + 0.0035\text{age}$$

Burn Injury Data Set

This data set has been analysed several times using a variety of VGLMS, with Huang (2017) using GEEs to fit these mean models.

GEE model parameters

$$\hat{\mu}_{(1)} = \frac{\exp\{-3.6891 + 0.0508\text{age}\}}{1 + \exp\{-3.6891 + 0.0508\text{age}\}}$$
$$\hat{\mu}_{(2)} = 6.7118 + 0.0035\text{age}$$

VSPGLM model parameters

$$\hat{\mu}_{(1)} = \frac{\exp\{-3.657 + 0.0509\text{age}\}}{1 + \exp\{-3.657 + 0.0509\text{age}\}}$$
$$\hat{\mu}_{(2)} = 6.7318 + 0.0027\text{age}$$

Burn Injury Data Set

	se($\hat{\boldsymbol{\beta}}$)
$\beta_{\text{death,age}}$	0.0051
$\beta_{\text{burns,age}}$	0.0017

Table 4: GEE corrected standard errors

	se($\hat{\boldsymbol{\beta}}$)
$\beta_{\text{death,age}}$	0.0045
$\beta_{\text{burns,age}}$	0.0019

Table 5: VSPGLM standard errors

Sorbinil Dataset

Contains $n = 41$ observations of itching scores (Y_L, Y_R) for the left and right eye after the application of sorbinil or a placebo, measured on a Likert scale between 0 and 4.

Table 6: Snippet of Sorbinil dataset

$n = 6$		$n = 14$		$n = 14$		$n = 7$	
sorbinil Left	sorbinil Right	sorbinil Left	placebo Right	placebo Left	sorbinil Right	placebo Left	placebo Right
2	2	1	1.5	2.5	2	3	3
1	1	2	2.5	2.5	2.5	2	3
0.5	2	3	1	3	3	2.5	2.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

This model finds a significant effect that sorbinil reduces itching scores by 0.43.

Sorbinil Dataset

We would like to see if sorbinil has a significant effect in reducing itching scores for each eye.

Previous attempts at modeling this dataset using VGLMs standardise these scores to $[0, 1]$ and fit logistic mean models, however, this is unnecessary when using this model.

The following symmetric model was found to be adequate

$$\mu_L = \beta_0 + \beta_1 \mathcal{I}_L, \quad \mu_R = \beta_0 + \beta_1 \mathcal{I}_R$$

This model finds a significant effect that sorbinil reduces itching scores by 0.43.

The code to fit this symmetric model is

```
sorbinil = fit_vspglm(["(Y_L, Y_R) ~ (I_L & I_R)"], tbl, {'id'})
```

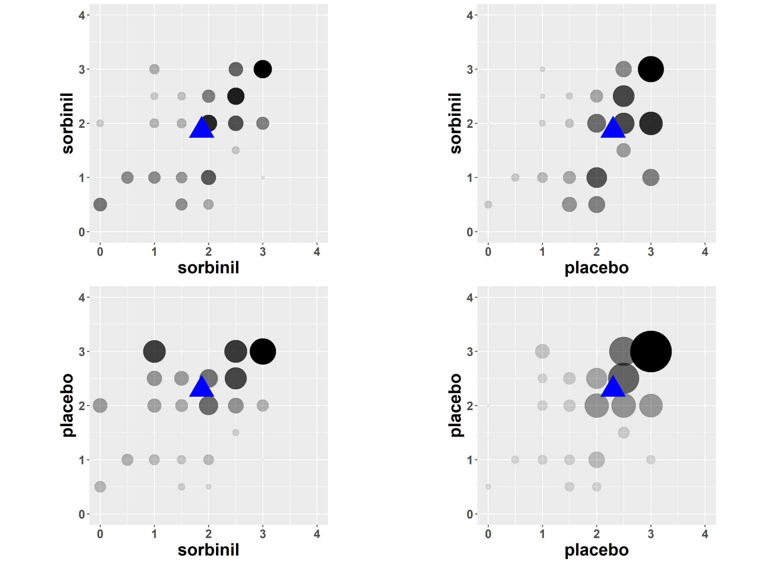

Figure 4: Reweighted Sorbinal pmf at group medians (left = y axis, right = x axis)

Huang, Alan (2014). "Joint estimation of the mean and error distribution in generalized linear models". In: *Journal of the American Statistical Association* 109.505, pp. 186–196.

– (2017). "On generalised estimating equations for vector regression". In: *Australian New Zealand Journal of Statistics* 59.2, pp. 195–213.

Huang, Alan and Paul Rathouz (2017). "Orthogonality of the mean and error distribution in generalized linear models". In: *Communications in Statistics-Theory and Methods* 46.7, pp. 3290–3296.

– (2012). "Proportional likelihood ratio models for mean regression". In: *Biometrika* 99.1, pp. 223–229.

Rathouz, Paul J. and Liping Gao (2009). "Generalized linear models with unspecified reference distribution". In: *Biostatistics* 10.2, pp. 205–218.

Trivariate Poisson Model

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, \sigma_0^2), \quad \sigma \in \mathbb{R}_+ \\
\boldsymbol{X} &\sim \mathrm{U}[-1, 1]^3 \\
Y_1 | \alpha, \boldsymbol{X}_{(1)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(1)}^T \boldsymbol{\beta}_{(1)} + \alpha)) \\
Y_2 | \alpha, \boldsymbol{X}_{(2)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(2)}^T \boldsymbol{\beta}_{(2)} + 0.5\alpha)) \\
Y_3 | \alpha, \boldsymbol{X}_{(3)} &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(3)}^T \boldsymbol{\beta}_{(3)} - 0.3\alpha))
\end{aligned}
$$

Trivariate Poisson Simulation Results

Table 7: Simulation results for trivariate Poisson model using a sample size of $n = 200$ and $N = 1000$ simulations

$\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}}$	$\lvert\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\rvert$	Errors		$p \leqslant 0.05$	CI		
			$\hat{\sigma}$	$\bar{\mathrm{se}}(\hat{\boldsymbol{\beta}})$		90%	95%	99%
0.4	0.40	0.004	0.13	0.12	0.88	0.89	0.94	0.99
-0.8	-0.79	0.002	0.13	0.14	0.99	0.93	0.96	0.99
0	0.001	0.001	0.13	0.12	0.05	0.89	0.95	0.99

Trivariate Mixed Effects Simulation

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0,\ \sigma_0^2) \\
\boldsymbol{X} &\sim \mathrm{U}[-1, 1]^3 \\
\boldsymbol{Y}_1 | \boldsymbol{X}_{(1)}, \alpha &\sim \mathcal{N}(\boldsymbol{X}_{(1)}^T \boldsymbol{\beta}_{(1)} + \alpha,\ \sigma_1^2) \\
\boldsymbol{Y}_2 | \boldsymbol{X}_{(2)}, \alpha &\sim \mathrm{Pois}(\exp(\boldsymbol{X}_{(2)}^T \boldsymbol{\beta}_{(2)} + 0.5\alpha)) \\
\boldsymbol{Y}_3 | \boldsymbol{X}_{(3)}, \alpha &\sim \mathrm{Gamma}(\lambda,\ \exp(\boldsymbol{X}_{(3)}^T \boldsymbol{\beta}_{(3)} - 0.3\alpha))
\end{aligned}
$$

Trivariate Mixed Effects Simulation Results

Table 8: Simulation results for trivariate mixed effects model using a sample size of $n = 200$ and $N = 1000$ simulations

Margin	$\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}}$	$\lvert\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\rvert$	Errors		$p \leqslant 0.05$	CI		
				$\hat{\sigma}$	$\bar{se}(\hat{\boldsymbol{\beta}})$		90%	95%	99%
Normal	1	1.005	0.0058	0.17	0.18	1	0.90	0.95	0.99
Poisson	-0.5	-0.49	0.0005	0.13	0.13	0.96	0.83	0.88	0.93
Gamma	0.4	0.39	0.004	0.12	0.13	0.85	0.89	0.93	0.97

Multivariate Normal Simulation Results

Table 9: Simulation results for bivariate normal model using sample size of $n = 200$ and $N = 1000$ simulations

β	$\hat{\beta}$	$\lvert\beta - \hat{\beta}\rvert$	Errors		$p \leqslant 0.05$	CI		
			$\hat{\sigma}$	$\bar{se}(\hat{\beta})$		90%	95%	99%
-1	-0.99	0.0004	0.11	0.11	1	0.91	0.95	0.99
0	-0.0004	0.0004	0.11	0.11	0.05	0.90	0.95	0.99
0.5	0.49	0.009	0.14	0.13	0.95	0.88	0.94	0.98
2.2	2.19	0.003	0.14	0.14	1	0.90	0.94	0.98

F simulation results

Table 10: Type 1 errors at significance levels of $0.10, 0.05$ and 0.01 using sample sizes of $n = 75, 150$, and $N = 3000$ simulations

	Type 1 Errors		
n	0.10	0.05	0.01
75	0.119	0.060	0.012
150	0.097	0.050	0.013