Generally Altered, Inflated and Truncated Count Distributions

T. W. Yee

University of Auckland, NZ

2018-07-4 @ Besançon

t.yee@auckland.ac.nz
http://www.stat.auckland.ac.nz/~yee

Outline of this document

VGLMs

GLMs are largely confined to 1-parameter distributions from an exponential family, but practical data analysis and regression modelling demands much greater flexibility.

Example: Count regression

Extensions of Poisson regression include:

- negative binomial regression: overdispersion
- zero-inflated Poisson: excess 0s
- zero-altered Poisson: excess 0s
- zero-truncated Poisson: positive distribution
- zero-inflated negative binomial: excess 0s and overdispersion
- zero-altered negative binomial: excess 0s and overdispersion
- zero-truncated negative binomial: positive and overdispersion
- ...

There is a solution! VGLMs extend GLMs to multiple η_j and far beyond.

Borrowed ideas:

- (parameter) link functions $g_j(\theta_j)$,
- IRLS,
- Fisher scoring (EIM is used, not the OIM),
- software (VGAM).

Three 'new' important ideas/infrastructure are

- *constraint matrices*,
- `xij`,
- `multiple responses`.

Others ideas are

- smoothing,
- reduced-rank regression (latent variables).

Vector generalized linear models

Data $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)$ on n independent "individuals".

Definition Conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ depends on $\boldsymbol{\theta}$:

$$f(\mathbf{y}|\mathbf{x}; \boldsymbol{\beta}) \;=\; f(\mathbf{y}, \eta_1, \ldots, \eta_M)$$

for $j = 1, \ldots, M$, and some function f,

$$\begin{aligned} \eta_j &= \eta_j(\mathbf{x}) = \boldsymbol{\beta}_j^T \mathbf{x} \quad (\in (-\infty, \infty)), && (1) \\ \boldsymbol{\beta}_j &= (\beta_{(j)1}, \ldots, \beta_{(j)p})^T, \\ \boldsymbol{\beta} &= (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_M^T)^T. && (2) \end{aligned}$$

Often $\boxed{g_j(\theta_j) = \eta_j}$ for parameters θ_j and link functions g_j.

The formulation is deliberately general so that it encompasses as many distributions and models as possible.

More general $\Longrightarrow$ more useful.

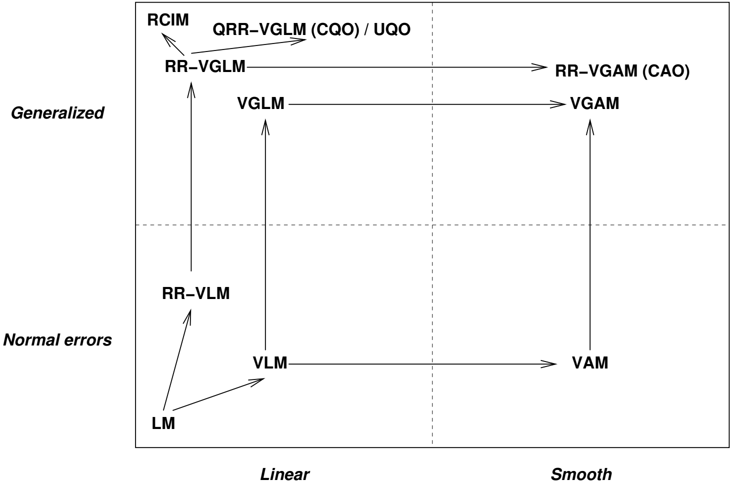

Figure : Flowchart for different classes of models. Legend: **LM** = linear model, **V** = vector, **G** = generalized, **A** = additive, **RR** = reduced-rank, **Q** = quadratic.

t

$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)^T$	Model	R function	Reference
$\mathbf{B}_1^T \mathbf{x}_1 + \mathbf{B}_2^T \mathbf{x}_2 \; (= \mathbf{B}^T \mathbf{x})$	VGLM	`vglm()`	Yee & Hastie (2003)
$\mathbf{B}_1^T \mathbf{x}_1 + \sum_{k=p_1+1}^{p_1+p_2} \mathbf{H}_k \mathbf{f}_k^*(x_k)$	VGAM	`vgam()`	Yee & Wild (1996)
$\mathbf{B}_1^T \mathbf{x}_1 + \mathbf{A}\,\boldsymbol{\nu}$	RR-VGLM	`rrvglm()`	Yee & Hastie (2003)
$\mathbf{B}_1^T \mathbf{x}_1 + \mathbf{A}\,\boldsymbol{\nu} + \begin{pmatrix} \boldsymbol{\nu}^T \mathbf{D}_1 \boldsymbol{\nu} \\ \vdots \\ \boldsymbol{\nu}^T \mathbf{D}_M \boldsymbol{\nu} \end{pmatrix}$	QRR-VGLM	`cqo()`	Yee (2004)
$\mathbf{B}_1^T \mathbf{x}_1 + \sum_{r=1}^{R} \mathbf{f}_r(\nu_r)$	RR-VGAM	`cao()`	Yee (2006)

Table : A summary of VGAM and its framework. The latent variables $\boldsymbol{\nu} = \mathbf{C}^T \mathbf{x}_2$, or $\nu = \mathbf{c}^T \mathbf{x}_2$ if rank $R = 1$. Here, $\mathbf{x}^T = (\mathbf{x}_1^T, \mathbf{x}_2^T)$. Abbreviations: A = additive, C = constrained, L = linear, O = ordination, Q = quadratic, RR = reduced-rank, VGLM = vector generalized linear model.

Table : Some VGAM link functions (grouped approximately by their domains).

t

Function	Link $g_j(\theta_j)$	Domain of θ_j	Link name
`cauchit()`	$\tan(\pi(\theta - \frac{1}{2}))$	$(0, 1)$	cauchit
`cloglog()`	$\log\{-\log(1-\theta)\}$	$(0, 1)$	complementary log-log
`logit()`	$\log \dfrac{\theta}{1-\theta}$	$(0, 1)$	logit
`multilogit()`	$\log \dfrac{\theta_j}{\theta_{M+1}}$	$(0, 1)^M$	multi-logit; $\sum_{j=1}^{M+1} \theta_j = 1$
`probit()`	$\Phi^{-1}(\theta)$	$(0, 1)$	probit (for "probability unit")
`rhobit()`	$\log \dfrac{1+\theta}{1-\theta}$	$(-1, 1)$	rhobit
`loge()`	$\log\ \theta$	$(0, \infty)$	log (logarithmic)
`extlogit()`	$\log \dfrac{\theta - A}{B - \theta}$	(A, B)	extended logit
`explink()`	e^{θ}	$(-\infty, \infty)$	exponential
`identitylink()`	θ	$(-\infty, \infty)$	identity
`loglog()`	$\log\log(\theta)$	$(1, \infty)$	log-log
`logoff(`θ `,` A`)`	$\log(\theta + A)$	$(-A, \infty)$	log with offset

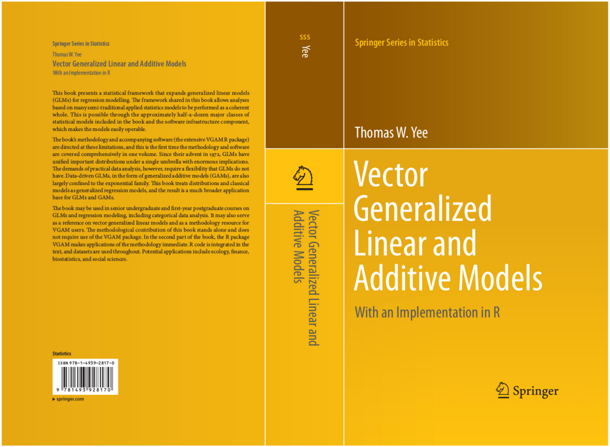

Figure : This appeared about two years ago. . . see Chapter 17 especially.

The scope of VGAM is very broad; it potentially covers

- univariate and multivariate distributions, including count distributions!
- categorical data analysis,
- extreme value analysis,
- quantile and expectile regression,
- time series (with Víctor Miranda),
- survival analysis,
- mixture models,
- nonlinear regression,
- reduced-rank regression,
- ordination, ...

It conveys GLM/GAM-type modelling to a much broader range of multivariate models.

S and GLMs and VGLMs

Use, e.g., `glm(y ~ x2 + x3 + x4, family = binomial, bdata)`

- Family functions: about 6.
- Generic functions include `anova()`, `coef()`, `fitted()`, `plot()`, `predict()`, `print()`, `resid()`, `summary()`, `update()`.

S and GLMs and VGLMs

Use, e.g., `glm(y ~ x2 + x3 + x4, family = binomial, bdata)`

- Family functions: about 6.
- Generic functions include `anova()`, `coef()`, `fitted()`, `plot()`, `predict()`, `print()`, `resid()`, `summary()`, `update()`.

Use, e.g., `vglm(y ~ x2 + x3 + x4, family = negbinomial, ndata)`

- Family functions: 150+.
- Generic functions include `anova()`, `coef()`, `fitted()`, `plot()`, `predict()`, `print()`, `resid()`, `summary()`, `update()`.
- Written in S4, not S3.

Some central VGLM/VGAM concepts are:

- parameter link functions $g_j(\theta_j)$ applied to all parameters,
- multivariate responses, and sometimes multiple responses too,
- linear predictors $\eta_j = \boldsymbol{\beta}_j^T \mathbf{x}$ and additive predictors $\eta_j = \sum_{k=1}^{d} f_{(j)k}(x_k)$,
- constraints on the functions $(\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_p)$,
- η_j-specific covariates (i.e., $\eta_j(\mathbf{x}_{ij})$) via the `xij` facility,
- reduced-rank regression (RRR), latent variables $\boldsymbol{\nu} = \mathbf{C}^T \mathbf{x}_2$, ordination,
- Fisher scoring, iteratively reweighted least squares (IRLS), maximum likelihood estimation, latent variables,
- the VGAM R package.

Data: $(\mathbf{x}_i, \mathbf{y}_i)$ indepenently, for $i = 1, \ldots, n$.

- $\mathbf{x}_i$ is a vector of explanatory variables. Sometimes we drop the i and write $\mathbf{x} = (x_1, \ldots, x_p)^T$ with $x_1 = 1$ denoting an intercept.
- $\mathbf{y}_i$ is a (possibly vector) response.

To fit a regression model involving parameters θ_j, VGLMs model each parameter, transformed if necessary, as a linear combination of the explanatory variables, viz.

$$g_j(\theta_j) \ = \ \eta_j \ = \ \boldsymbol{\beta}_j^T \mathbf{x} \ = \ \beta_{(j)1} \, x_1 + \cdots + \beta_{(j)p} \, x_p, \qquad j = 1, \ldots, M, \tag{3}$$

where g_j is a *parameter link function*. Potentially *every* parameter is modelled using *all* explanatory variables x_k. It's *model-driven*.

Data-driven VGAMs extend (3) to

$$g_j(\theta_j) \ = \ \eta_j \ = \ \sum_{k=1}^{d} f_{(j)k}(x_k), \qquad j = 1, \ldots, M, \tag{4}$$

i.e., an additive model for each parameter. The functions $f_{(j)k}$ are merely assumed to be smooth and are estimated by smoothers such as splines. In (4) $f_{(j)1}(x_1) = \beta_{(j)1}$ is simply an intercept.

The Negative Binomial and Constraint Matrices

$$\Pr(Y = y; \mu, k) \;=\; \binom{y+k-1}{y} \left(\frac{\mu}{\mu+k}\right)^{y} \left(\frac{k}{k+\mu}\right)^{k} \tag{5}$$

with $\mu, k > 0$. The variance function is

$$\mathrm{Var}(Y) \;=\; \mu + \frac{\mu^2}{k} \;>\; \mu. \tag{6}$$

Easy to fit the *NB-H*:

$$\begin{aligned} \log \mu &= \eta_1 &= \boldsymbol{\beta}_1^T \mathbf{x}, \\ \log k &= \eta_2 &= \boldsymbol{\beta}_2^T \mathbf{x}. \end{aligned} \tag{7}$$

```
vglm(y ~ x2 + ... + xp, negbinomial(zero = NULL), data = ndata)
```

But a simpler model (*NB-2*) is:

$$\begin{aligned} \log \mu &= \eta_1 = \boldsymbol{\beta}_1^T \mathbf{x} \\ \log k &= \eta_2 = \beta^*_{(2)1}. \end{aligned} \qquad (8)$$

We say k is *intercept-only*. Easy to fit since `zero = "size"` is default:

```
vglm(y ~ x2 + ... + xp, negbinomial, data = ndata)
```

With "$*$" denoting the parameters that are estimated and $p = 3$,

$$\begin{aligned} \eta_1(\mathbf{x}_i) &= \beta^*_{(1)1} + \beta^*_{(1)2}\, x_{i2} + \beta^*_{(1)3}\, x_{i3}, \\ \eta_2(\mathbf{x}_i) &= \beta^*_{(2)1}, \end{aligned}$$

so

$$\begin{aligned} \boldsymbol{\eta}(\mathbf{x}_i) = \begin{pmatrix} \eta_1(\mathbf{x}_i) \\ \eta_2(\mathbf{x}_i) \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta^*_{(1)1} \\ \beta^*_{(2)1} \end{pmatrix} x_{i1} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \beta^*_{(1)2}\, x_{i2} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \beta^*_{(1)3}\, x_{i3} \\ &= \sum_{k=1}^{3} \mathbf{H}_k\, \boldsymbol{\beta}^*_{(k)}\, x_{ik}. \end{aligned} \qquad (9)$$

The constraint matrices $\mathbf{H}_k$ can be inputted, else defined via `zero`.

Q: How can I estimate μ from a random sample from a NB distribution with $\mathrm{Var}(Y) \propto \mu$? (aka *NB-1*)

Hints: use $\mathrm{Var}(NB(\mu, k)) = \mu + \mu^2/k$ and

```
> args(negbinomial)

function (zero = "size", parallel = FALSE, deviance.arg = FALSE,
    type.fitted = c("mean", "quantiles"), percentiles = c(25,
        50, 75), mds.min = 0.001, nsimEIM = 500, cutoff.prob = 0.999,
    eps.trig = 1e-07, max.support = 4000, max.chunk.MB = 30,
    lmu = "loge", lsize = "loge", imethod = 1, imu = NULL, iprobs.y = NULL,
    gprobs.y = ppoints(6), isize = NULL, gsize.mux = exp(c(-30,
        -20, -15, -10, -6:3)))
NULL
```

A:

Make $k \propto \mu$. Let $k = c_1 \mu$ for some scalar c_1. Then $\sigma^2 = (1 + c_1^{-1})\mu$.
Use $\mathbf{H}_1 = \mathbf{I}_2$ and $\mathbf{H}_2 = \cdots = \mathbf{1}_2$.

Even easier:

```
vglm(y ~ x2 + x3, negbinomial(parallel = TRUE, zero = NULL), data = ndata)
```

Table : How VGAM can fit negative binomial variants and other associated models. Note: k is an intercept-only (scalar) unless specified. Nomenclature follows Hilbe (2011) but "NB-G" refers to the geometric variant.

t

NB variant	Var(Y)		VGAM family
NB-1	$\phi\,\mu$	`vglm()`	`negbinomial(parallel = TRUE, zero = NULL)`
NB-2	$\mu + \mu^2/k$	`vglm()`	`negbinomial()`
NB-C	$\mu + \mu^2/k$	`vglm()`	`negbinomial("nbcanlink")`
NB-H	$\mu + \mu^2/k(\mathbf{x})$	`vglm()`	`negbinomial(zero = NULL)`
NB-P (RR-NB)	$\mu(\mathbf{x}) + \delta_1\,\mu(\mathbf{x})^{\delta_2}$	`rrvglm()`	`negbinomial(zero = NULL)`
NB-G	$\mu + \mu^2$	`vglm()`	`negbinomial.size(size = 1)`
Poisson	μ	`vglm()`	`negbinomial.size(size = Inf)`
Poisson	μ	`vglm()`	`poissonff()`

In general for VGLMs, we represent the models as

$$\boldsymbol{\eta}(\mathbf{x}_i) = \sum_{k=1}^{p} \boldsymbol{\beta}_{(k)}\, x_{ik} = \sum_{k=1}^{p} \mathbf{H}_k\, \boldsymbol{\beta}^*_{(k)}\, x_{ik} \tag{10}$$

$$= \sum_{k=1}^{p} \mathrm{diag}(x_{ik1}, \ldots, x_{ikM})\, \mathbf{H}_k\, \boldsymbol{\beta}^*_{(k)} \quad \text{(aka xij)} \tag{11}$$

where

- $\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_p$ are known *constraint matrices*.
  Properties: of full column-rank (i.e., rank `ncol`$(\mathbf{H}_k)$), known and fixed finite elements, prespecified. Usually the elements are 0s and 1s.
- $\boldsymbol{\beta}^*_{(k)}$ is a vector containing a possibly reduced set of regression coefficients.

No constraints at all $\Longrightarrow$ all $\mathbf{H}_k = \mathbf{I}_M$ and $\boldsymbol{\beta}^*_{(k)} = \boldsymbol{\beta}_{(k)}$ (aka *trivial constraints*).

VGLM Algorithm†

Models with log-likelihood $\sum_{i=1}^{n} \ell_i\{\eta_1(\mathbf{x}_i), \ldots, \eta_M(\mathbf{x}_i)\}$, where $\eta_j = \boldsymbol{\beta}_j^T \mathbf{x}_i$. Then

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \eta_j} \mathbf{x}_i \quad \text{and} \quad \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_j \, \partial \boldsymbol{\beta}_k^T} = \sum_{i=1}^{n} \frac{\partial^2 \ell_i}{\partial \eta_j \, \partial \eta_k} \mathbf{x}_i \, \mathbf{x}_i^T .$$

Newton-Raphson algorithm:

$$\boldsymbol{\beta}^{(a+1)} = \boldsymbol{\beta}^{(a)} + \mathcal{J}\left(\boldsymbol{\beta}^{(a)}\right)^{-1} \mathbf{U}\left(\boldsymbol{\beta}^{(a)}\right)$$

written in iteratively reweighted least squares (IRLS) form is

$$\begin{aligned} \boldsymbol{\beta}^{(a+1)} &= \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} \, \boldsymbol{\beta}^{(a)} + \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{W}^{-1} \mathbf{u} \\ &= \left(\mathbf{X}_{\text{VLM}}^T \, \mathbf{W}^{(a)} \, \mathbf{X}_{\text{VLM}}\right)^{-1} \mathbf{X}_{\text{VLM}}^T \, \mathbf{W}^{(a)} \, \mathbf{z}^{(a)} . \end{aligned}$$

Let $\mathbf{z} = (\mathbf{z}_1^T, \ldots, \mathbf{z}_n^T)^T$ and $\mathbf{u} = (\mathbf{u}_1^T, \ldots, \mathbf{u}_n^T)^T$, where $\mathbf{u}_i$ has jth element $(\mathbf{u}_i)_j = \partial \ell_i / \partial \eta_j$, and $\mathbf{z}_i = \boldsymbol{\eta}(\mathbf{x}_i) + \mathbf{W}_i^{-1} \mathbf{u}_i$ (*adjusted dependent vector* or *pseudo-response*).

Also, $\mathbf{W} = \text{Diag}(\mathbf{W}_1, \ldots, \mathbf{W}_n)$, $\quad (\mathbf{W}_i)_{jk} = -\dfrac{\partial^2 \ell_i}{\partial \eta_j \, \partial \eta_k}$,

$\mathbf{X}_{\text{VLM}} = (\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T)^T$, $\quad \mathbf{X}_i = \text{Diag}(\mathbf{x}_i^T, \ldots, \mathbf{x}_i^T) = \mathbf{I}_M \otimes \mathbf{x}_i^T$.
Then $\boldsymbol{\beta}^{(a+1)}$ is the solution to

$$\mathbf{z}^{(a)} \;=\; \mathbf{X}_{\text{VLM}}\, \boldsymbol{\beta}^{(a+1)} + \boldsymbol{\varepsilon}^{(a)}, \qquad \text{Var}(\boldsymbol{\varepsilon}^{(a)}) \;=\; {\mathbf{W}^{(a)}}^{-1}.$$

Fisher scoring:

$$(\mathbf{W}_i)_{jk} \;=\; -E\left[\frac{\partial^2 \ell_i}{\partial \eta_j \, \partial \eta_k}\right]$$

usually results in slower convergence but is preferable because the *working weight matrices* are positive-definite over a larger parameter space.

Note: packages such as pscl call `optim()`, which is inferior both algorithmically and for inference.

Some Notes

1. To handle various link functions use:

$$
\begin{aligned}
\frac{\partial \ell}{\partial \eta_j} &= \frac{\partial \ell}{\partial \theta_j} \frac{\partial \theta_j}{\partial \eta_j}, \\
\frac{\partial^2 \ell}{\partial \eta_j^2} &= \frac{\partial \ell}{\partial \theta_j} \frac{\partial^2 \theta_j}{\partial \eta_j^2} + \left(\frac{\partial \theta_j}{\partial \eta_j}\right)^2 \frac{\partial^2 \ell}{\partial \theta_j^2}, \\
\frac{\partial^2 \ell}{\partial \eta_j \, \partial \eta_k} &= \left\{ \frac{\partial^2 \ell}{\partial \theta_j \, \partial \theta_k} - \frac{\partial \ell}{\partial \theta_k} \frac{\partial \theta_k}{\partial \eta_k} \frac{\partial^2 \eta_k}{\partial \theta_j \, \partial \theta_k} \right\} \frac{\partial \theta_j}{\partial \eta_j} \frac{\partial \theta_k}{\partial \eta_k}, \quad j \neq k,
\end{aligned}
$$

Fisher scoring for most VGAM family functions reduce to

$$
-E\left[\frac{\partial^2 \ell_i}{\partial \eta_j^2}\right] = -E\left[\frac{\partial^2 \ell_i}{\partial \theta_j^2}\right] \cdot \left(\frac{\partial \theta_j}{\partial \eta_j}\right)^2, \tag{12}
$$

$$
(\mathbf{W}_i)_{jk} = -E\left[\frac{\partial^2 \ell_i}{\partial \eta_j \, \partial \eta_k}\right] = -E\left[\frac{\partial^2 \ell_i}{\partial \theta_j \, \partial \theta_k}\right] \cdot \frac{\partial \theta_j}{\partial \eta_j} \frac{\partial \theta_k}{\partial \eta_k}, \quad j \neq k. \tag{13}
$$

2. With trivial constraints, `model.matrix(fit, type = "vlm")` returns

$$\mathbf{X}_{\text{VLM}} = \left((\mathbf{X}\,\mathbf{e}_1) \otimes \mathbf{H}_1 \,\middle|\, \cdots \,\middle|\, (\mathbf{X}\,\mathbf{e}_p) \otimes \mathbf{H}_p \right) = \mathbf{X}_{\text{LM}} \otimes \mathbf{I}_M.$$

With the `xij` facility,

$$\mathbf{X}_{\text{VLM}} = \begin{pmatrix} \mathbf{X}^{\#}_{(11)}\mathbf{H}_1 & \cdots & \mathbf{X}^{\#}_{(1p)}\mathbf{H}_p \\ \vdots & & \vdots \\ \mathbf{X}^{\#}_{(n1)}\mathbf{H}_1 & \cdots & \mathbf{X}^{\#}_{(np)}\mathbf{H}_p \end{pmatrix}. \tag{14}$$

EIMs

The *expected information matrix (EIM)* is crucial when fitting VGLMs. For these, $E(\partial\ell/\partial\theta) = 0$ sometimes helps.

Examples:

- Positive-Poisson(λ):

$$\left(1 + \frac{1}{e^{\lambda} - 1}\right)\left(\frac{1}{\lambda} - \frac{1}{e^{\lambda} - 1}\right).$$

- Zero-inflated Poisson($\phi,\ \lambda$) with $\Pr(Y = 0) = \phi + (1-\phi)\, e^{-\lambda}$:

$$\begin{pmatrix} \dfrac{1 - e^{-\lambda}}{(1-\phi)(\phi + (1-\phi)\, e^{-\lambda})} & \dfrac{-e^{-\lambda}}{\phi + (1-\phi)\, e^{-\lambda}} \\ \dfrac{-e^{-\lambda}}{\phi + (1-\phi)\, e^{-\lambda}} & \dfrac{1-\phi}{\lambda} - \dfrac{\phi(1-\phi)\, e^{-\lambda}}{\phi + (1-\phi)\, e^{-\lambda}} \end{pmatrix}. \qquad (15)$$

- $\mathcal{M}_{tbh}$ for capture-recapture positive-Bernoulli data:

$$-E\left(\frac{\partial^2 \ell}{\partial p_{cj}^2}\right) = \frac{1}{(1-p_{cj})^2(1-Q_{1:\tau})}\left\{\frac{Q_{1:j}}{p_{cj}} - \frac{Q_{1:\tau}}{1-Q_{1:\tau}}\right\},$$

$$-E\left(\frac{\partial^2 \ell}{\partial p_{rj}^2}\right) = \frac{1-Q_{1:j}/(1-p_{cj})}{p_{rj}(1-p_{rj})(1-Q_{1:\tau})},$$

$$-E\left(\frac{\partial^2 \ell}{\partial p_{cj}\,\partial p_{ck}}\right) = \frac{-\dfrac{\partial Q_{1:\tau}}{\partial p_{cj}}\dfrac{\partial Q_{1:\tau}}{\partial p_{ck}}}{(1-Q_{1:\tau})^2} - \frac{\dfrac{\partial^2 Q_{1:\tau}}{\partial p_{cj}\,\partial p_{ck}}}{(1-Q_{1:\tau})}, \qquad j \neq k,$$

where $\partial Q_{1:\tau}/\partial p_{cj} = -Q_{1:\tau}/(1-p_{cj})$
and $\partial^2 Q_{1:\tau}/(\partial p_{cj}\,\partial p_{ck}) = Q_{1:\tau}/\{(1-p_{cj})(1-p_{ck})\}$.

EIMs—ZI-Distributions

Consider

$$\Pr(Y = y) = \begin{cases} \phi + (1-\phi)\, f(0), & y = 0; \\ (1-\phi)\, f(y), & y = 1, 2, 3, \ldots. \end{cases} \quad (16)$$

and let $A = \phi + (1-\phi) f(0) = \Pr(Y = 0)$ where $f(0) = \Pr(Y^* = 0)$ where Y^* denotes the random variable from the parent distribution.

(a) Let ℓ_{i+} and ℓ_{i0} be parts of the log-likelihood corresponding to positive y_i and $y_i = 0$ respectively. Then

  (i) $\partial\ell_{i0}/\partial\phi = [1 - f(0)]/A$ and $\partial\ell_{i0}/\partial\theta = [(1-\phi)/A]\,\partial f(0)/\partial\theta$,

  (ii) $\partial\ell_{i+}/\partial\phi = -1/(1-\phi)$ and $\partial\ell_{i+}/\partial\theta = \partial\ell_i^*/\partial\theta$.

(b) The EIM for a zero-inflated 1-parameter discrete distribution is

$$\begin{pmatrix} \frac{1}{A} \frac{(1 - f(0))}{(1 - \phi)} & \frac{1}{A} \cdot \frac{\partial f(0)}{\partial \theta} \\ \frac{1}{A} \cdot \frac{\partial f(0)}{\partial \theta} & J \end{pmatrix} \tag{17}$$

where a conditional expectation is used for $J =$

$$-E\left[\frac{\partial^2 \ell_i^*}{\partial \theta^2}\right] \cdot (1 - A) - \left[\frac{1 - \phi}{A}\right] \cdot \left\{A \frac{\partial^2 f(0)}{\partial \theta^2} - (1 - \phi)\left(\frac{\partial f(0)}{\partial \theta}\right)^2\right\}. \tag{18}$$

Hint: $E(Y) = E(Y^*)/(1 - f(0))$ for ℓ_{i+}.

(c) Using (17), the EIM for the ZIP distribution is (15). It is positive-definite for $(\phi, \lambda) \in (0, 1) \times (0, \infty)$.

(d) One can use (17) to derive the EIM for the ZI-binomial and ZI-geometric, etc.

Poor man's EIM: Under regularity conditions, the ith EIM is

$$\mathrm{Var}\left(\frac{\partial \ell_i}{\partial \boldsymbol{\theta}}\right) = E\left[\left(\frac{\partial \ell_i}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ell_i}{\partial \boldsymbol{\theta}}\right)^T\right]. \tag{19}$$

If it exists, an `r`-type function might be used to generate random score vectors.

VGAM family functions such as `zanegbinomial()` and `zinegbinomial()` use this technique.

The VGAM Package for R

The Central Functions of VGAM:

`vglm()`	Vector generalized linear models.
`vgam()`	Vector generalized additive models.
`rrvglm()`	Reduced-rank vector generalized linear models.
`cqo()`	Constrained quadratic (Gaussian) ordination (QRR-VGLM).
`cao()`	Constrained additive ordination (RR-VGAM).

Others:

`grc()`	Goodman's RC(R) model.
`rcim()`	Row-column interaction models.

Modular construction, flexible, easy to use and useful.

See the `DESCRIPTION` file.

```
Package: VGAM
Version: 1.0-6
Date: 2018-07-03
Title: Vector Generalized Linear and Additive Models
Author: Thomas W. Yee <t.yee@auckland.ac.nz>
Maintainer: Thomas Yee <t.yee@auckland.ac.nz>
Depends: R (>= 3.4.0), methods, stats, stats4, splines
Suggests: VGAMdata, MASS, mgcv
Description: An implementation of about 6 major classes of
    statistical regression models. At the heart of it are the
    vector generalized linear and additive model (VGLM/VGAM)
    classes, and the book "Vector Generalized Linear and
    Additive Models: With an Implementation in R" (Yee, 2015)
    <DOI: 10.1007/978-1-4939-2818-7>
    gives details of the statistical framework and VGAM package.
    Currently only fixed-effects models are implemented,
    i.e., no random-effects models. Many (150+) models and
    distributions are estimated by maximum likelihood estimation
    (MLE) or penalized MLE, using Fisher scoring. VGLMs can be
    loosely thought of as multivariate GLMs. VGAMs are data-driven
    VGLMs (i.e., with smoothing). The other classes are RR-VGLMs
    (reduced-rank VGLMs), quadratic RR-VGLMs, reduced-rank VGAMs,
    RCIMs (row-column interaction models)---these classes perform
    constrained and unconstrained quadratic ordination (CQO/UQO)
    models in ecology, as well as constrained additive ordination
    (CAO). Note that these functions are subject to change;
    see the NEWS and ChangeLog files for latest changes.
License: GPL-3
URL: https://www.stat.auckland.ac.nz/~yee/VGAM
```

Table : Some VGAM generic functions applied to a model called `fit`.

Function	Value
`coef(fit)`	$\widehat{\beta}^*$
`coef(fit, matrix = TRUE)`	$\widehat{\mathbf{B}}$
`constraints(fit, type = "lm")`	$\mathbf{H}_k, \ k = 1, \ldots, p$
`deviance(fit)`	Deviance $D = \sum_{i=1}^{n} d_i$
`fitted(fit)`	$\widehat{\mu}_{ij}$ usually
`logLik(fit)`	Log-likelihood $\sum_{i=1}^{n} w_i \, \ell_i$
`model.matrix(fit, type = "lm")`	LM model matrix $(n \times p)$
`model.matrix(fit, type = "vlm")`	VLM model matrix $\mathbf{X}_{\text{VLM}}$
`predict(fit)`	$\widehat{\eta}_{ij}$

Table : Some VGAM generic functions applied to a model called `fit`.

Function	Value
`predict(fit, type = "response")`	$\widehat{\mu}_{ij}$ usually
`resid(fit, type = "response")`	$y_{ij} - \widehat{\mu}_{ij}$
`resid(fit, type = "deviance")`	$\text{sign}(y_i - \hat{\mu}_i)\,\sqrt{d_i}$
`resid(fit, type = "pearson")`	$\mathbf{W}_i^{-\frac{1}{2}}\,\mathbf{u}_i$
`resid(fit, type = "working")`	$\mathbf{z}_i - \boldsymbol{\eta}_i = \mathbf{W}_i^{-1}\,\mathbf{u}_i$
`vcov(fit)`	$\widehat{\text{Var}}(\widehat{\boldsymbol{\beta}})$
`weights(fit, type = "prior")`	w_i (`weights` argument)
`weights(fit, type = "working")`	$w_i\mathbf{W}_i$ (in matrix-band format)

Some Computational and Implementational Details†

- Is S4 object-orientated and very modular—simply have to write a VGAM "family function".
- ```
  > args(vglm.control)
  
  function (checkwz = TRUE, Check.rank = TRUE, Check.cm.rank = TRUE,
      criterion = names(.min.criterion.VGAM), epsilon = 1e-07,
      half.stepsizing = TRUE, maxit = 30, noWarning = FALSE, stepsize = 1,
      save.weights = FALSE, trace = FALSE, wzepsilon = .Machine$double.eps^0.75,
      xij = NULL, ...)
  NULL
  ```
- Parameter link functions, e.g.,
  - $\log\theta$ for $0 < \theta$,
  - $\text{logit}\ \theta$ for $0 < \theta < 1$.
  - $\log(\theta - 1)$ for $1 < \theta$.
- *Half-step sizing*.
- `@validparams` slot.
- Good initial values, e.g., self-starting VGAM family functions.
- Numerical linear algebra based on orthogonal methods, e.g., QR method in LINPACK. Yet to do: use LAPACK.
- B-splines, not the Reinsch algorithm.
  ```

Example: ZIP and Professors of Statistics in NZ

Loosely,

$$\Pr(Y = y; \phi, \lambda) \;\; = \;\; \phi \Pr(Y = 0) + (1 - \phi)\,\text{Poisson}(\lambda).$$

where $0 < \phi < 1$ and $\lambda > 0$.

The default for `zipoisson()` is

$$\boldsymbol{\eta} \;=\; \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \;=\; \begin{pmatrix} \operatorname{logit} \phi \\ \log \lambda \end{pmatrix}.$$

Here, $Y =$ the number of first-author publications in MathSciNet.

```
> data(profs.nz, package = "VGAMdata")
> fit.zip <- vglm(pub1stAuthor ~ log(2014 - firstyear), trace = TRUE,
                  crit = "coef", zipoisson, data = profs.nz)

VGLM    linear loop  1 :  coefficients =
 3.35934349,  0.63102852, -1.68545205,  0.68534221
VGLM    linear loop  2 :  coefficients =
 2.76416615, -0.34883684, -1.59704166,  0.93825124
VGLM    linear loop  3 :  coefficients =
 2.69471617, -0.39875167, -1.58808340,  0.95142286
VGLM    linear loop  4 :  coefficients =
 2.68764904, -0.39781236, -1.58611724,  0.95118385
VGLM    linear loop  5 :  coefficients =
 2.68760285, -0.39784682, -1.58610346,  0.95119298
VGLM    linear loop  6 :  coefficients =
 2.68759881, -0.39784620, -1.58610223,  0.95119282
VGLM    linear loop  7 :  coefficients =
 2.68759881, -0.39784622, -1.58610223,  0.95119283

> coef(fit.zip, matrix = TRUE)

                      logit(pstr0) loge(lambda)
(Intercept)                  2.688      -0.3978
log(2014 - firstyear)       -1.586       0.9512
```

VGAMs

VGAMs allow additive-model extensions to all η_j in a VGLM, i.e., from

$$\eta_j(\mathbf{x}) \;=\; \beta_{(j)1}\, x_1 + \cdots + \beta_{(j)p}\, x_p \;=\; \boldsymbol{\beta}_j^T \mathbf{x}$$

to M additive predictors:

$$\eta_j(\mathbf{x}) \;=\; f_{(j)1}(x_1) + \cdots + f_{(j)p}(x_p),$$

a sum of arbitary smooth functions. Equivalently,

$$\begin{aligned} \boldsymbol{\eta}(\mathbf{x}) &= \mathbf{f}_1(x_1) + \cdots + \mathbf{f}_p(x_p) \\ &= \mathbf{H}_1\, \mathbf{f}_1^*(x_1) + \cdots + \mathbf{H}_p\, \mathbf{f}_p^*(x_p) \end{aligned} \qquad (20)$$

for *constraint matrices* $\mathbf{H}_k$ (default: $\mathbf{H}_k = \mathbf{I}_M$).

- $\mathbf{H}_1, \ldots, \mathbf{H}_p$ are known and of full-column rank,
- $\mathbf{f}_k^* = (f_{(1)k}^*(x_k), f_{(2)k}^*(x_k), \ldots)^T$ contains possibly a reduced set of component functions. Starred quantities are estimated. They are centered for identifiability.

Estimation†

There are 2 types of VGAMs currently implemented now:

1. Based on O-splines, (vector) backfitting. See Hastie and Tibshirani (1990). Can be considered *1st-generation* GAMs.
2. Based on P-splines, minimization of UBRE/GCV, mgcv. See Wood (2006). Can be considered *2nd-generation* GAMs.
   Automatic smoothing parameter selection!

A Short History of GAMs. . . †

1. *First generation GAMs*: Hastie and Tibshirani, late-1980s/early-1990s, gam.

   Main ideas: smoothing splines, backfitting.

2. *Second generation GAMs*: Wood, early-2000s onwards, mgcv.

   Main ideas: "Penalized B-splines" (Eilers and Marx, 1996). Smoothing parameter selection is much easier, as well as inference.

3. *First generation VGAMs*: Yee and Wild (1996), VGAM.

   Main ideas: vector (smoothing) splines, vector backfitting.

4. *Second generation VGAMs*: VGAM has a rudimentary implementation of P-spline VGAMs. Should be refined by the end of this year. Joint work with Chanatda Somchit and Chris Wild. See `sm.os()` and `sm.ps()`.
   A comparison between O-splines and P-splines is in Wand and Ormerod (2008).

Reduced-Rank VGLMs

What are Latent Variables?

They have various meanings in statistics.

- The word latent means ‘concealed’, ‘dormant’, ‘hidden’.
- Often a random variable which cannot be measured directly or an unobserved or latent trait, e.g., quality of life, business confidence, morale, happiness.
- They are inferred from other variables that are observed (directly measured), e.g., for quality of life, we can measure wealth, employment, environment, physical and mental health, leisure time and social belonging.
- Often a linear combination of the explanatory variables, e.g.,

$$\nu = \mathbf{c}^T \mathbf{x} \tag{21}$$

  For the $\nu =$ quality of life example, $\mathbf{x} = ($wealth, employment, environment, physical and mental health, leisure time, social belonging$)^T$.
- I call the **c** the *constrained coefficients*. Also called *loadings* or *weights*.
- It reduces the dimensionality of the model, e.g., from p to 1.
- Used in many disciplines, e.g., psychology, economics, medicine, ecology, social sciences.

RR-VGLMs and QRR-VGLMs

Suppose we have $R = 1$ latent variable, ν (nu). Something like

$$g_1(\theta_1) = \eta_1 = c_1 + \texttt{poly}(\texttt{nu}, 1), \quad (22)$$
$$g_1(\theta_1) = \eta_1 = c_1 + \texttt{poly}(\texttt{nu}, 2) \quad (23)$$

are *RR-VGLMs* and *QRR-VGLMs* respectively.

They allow a *monotonic* and *unimodal* responses wrt ν respectively.

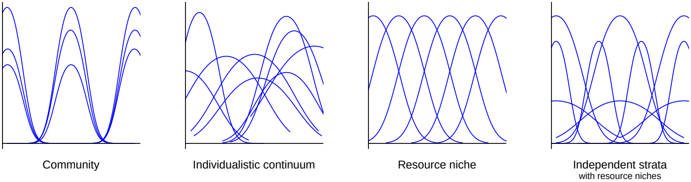

Figure : Four hypothetical models for community ecology, along a gradient.

Motivation: if M and p are large then $\mathbf{B} = (\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_M)$ is "too big" (Mp elts).

Idea: approximate part of $\mathbf{B}$ by a lower rank matrix.

Partition $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}$ accordingly.

RR-VGLMs:

$$\boldsymbol{\eta} = \mathbf{B}_1^T \mathbf{x}_1 + \mathbf{A}\,\mathbf{C}^T \mathbf{x}_2 \quad \text{i.e.,} \tag{24}$$

$$\mathbf{B}_2 = \mathbf{C}\,\mathbf{A}^T, \qquad (p_2 \times R) \times (R \times M) \text{ in dimension.} \tag{25}$$

$$\boldsymbol{\nu} = \mathbf{C}^T \mathbf{x}_2 \qquad \text{is a } R\text{-vector of } \textit{latent variables} \tag{26}$$

Often $R = 1$, 2, or 3 . . . . Latent variables are an important concept in ecology, e.g., direct and indirect gradient analysis.

Roles:

- **C** can be considered as choosing the 'best' predictors from a linear combination of the original predictors.
- **A** are regression coefficients of the new predictors $\boldsymbol{\nu}$.

Example: *Stereotype model* = *RR-multinomial logit model* (Anderson, 1984).

A Simple and Important Result

RR-VGLMs are VGLMs where the constraint matrices are unknown and to be estimated.

Estimation by an *alternating algorithm* exploits this.

Two-parameter Rank-1 RR-VGLMs

These allow 2 parameters to be tractably coupled (tied/linked together).

For Variant I,

$$\begin{pmatrix} g_1(\theta_1) \\ g_2(\theta_2) \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \mathbf{B}_1^T \mathbf{x}_1 + \mathbf{A}\,\mathbf{C}^T \mathbf{x}_2 \tag{27}$$

$$= \begin{pmatrix} \beta_{(1)1} \\ \beta_{(2)1} \end{pmatrix} + \begin{pmatrix} 1 \\ a_{21} \end{pmatrix} \nu \tag{28}$$

where $\nu = \mathbf{c}^T \mathbf{x}_2$. Here, $a_{11} \equiv 1$ by corner constraints, and a_{21} is to be estimated.

We have the RRR applied to all x_k except possibly for x_1, and $M = 2$. Then

$$g_1(\theta_1) = \eta_1 = \beta_{(1)1} + \mathbf{c}^T \mathbf{x}_2, \tag{29}$$

$$\begin{aligned} g_2(\theta_2) = \eta_2 &= \beta_{(2)1} + a_{21}\left(\mathbf{c}^T \mathbf{x}_2\right) \\ &= \left(\beta_{(2)1} - a_{21} \cdot \beta_{(1)1}\right) + a_{21}\,\eta_1 \\ &= t_1 + a_{21} \cdot \eta_1, \text{ say.} \end{aligned} \tag{30}$$

Actually, there are two variants:

1. Variant I: $\mathbf{H}_1 = \mathbf{I}_2$, $\mathbf{H}_k = (1, a_{21})^T$ for $k = 2, \ldots, p$.
2. Variant II: $\mathbf{H}_k = (1, a_{21})^T$ for all k. So $t_1 = 0$ in (30).

Equation (28) is central here. Then

$$\theta_2 = g_2^{-1}(t_1 + a_{21} \cdot g_1(\theta_1)), \qquad \text{or equivalently,} \tag{31}$$

$$\eta_2 = t_1 + a_{21} \cdot \eta_1. \tag{32}$$

The table on Slide 47 gives a small selection of popular link functions for the g_j.

	identitylink	log	logit	power
identitylink	$t_1 + a_{21}\ \theta_1$	$K_1 \cdot K_2^{\theta_1}$	$\dfrac{K_1 \cdot K_2^{\theta_1}}{1 + K_1 \cdot K_2^{\theta_1}}$	$(t_1 + a_{21}\ \theta_1)^{1/s_2}$
log	$t_1 + a_{21} \log\ \theta_1$	$K_1 \cdot {\theta_1}^{a_{21}}$	$\dfrac{K_1 \cdot {\theta_1}^{a_{21}}}{1 + K_1 \cdot {\theta_1}^{a_{21}}}$	$(t_1 + a_{21} \log\ \theta_1)^{1/s_2}$
logit	$t_1 + a_{21}\ \log \dfrac{\theta_1}{1 - \theta_1}$	$K_1 \left(\dfrac{\theta_1}{1 - \theta_1}\right)^{a_{21}}$	$\dfrac{K_1 \left(\dfrac{\theta_1}{1-\theta_1}\right)^{a_{21}}}{1 + K_1 \left(\dfrac{\theta_1}{1-\theta_1}\right)^{a_{21}}}$	$\left(t_1 + a_{21}\ \log \dfrac{\theta_1}{1 - \theta_1}\right)^{1/s_2}$
power	$t_1 + a_{21}\ \theta_1^{s_1}$	$K_1 \cdot K_2^{\theta_1^{s_1}}$	$\dfrac{K_1 \cdot K_2^{\theta_1^{s_1}}}{1 + K_1 \cdot K_2^{\theta_1^{s_1}}}$	$\left(t_1 + a_{21}\ \theta_1^{s_1}\right)^{1/s_2}$

Table : Expressions for θ_2 as a function of θ_1 for several parameter link functions; the cells are $\theta_2 = g_2^{-1}(t_1 + a_{21} \cdot g_1(\theta_1))$. The rows are for θ_1, columns for θ_2. The parameters t_1, K_j and a_{21} are unknown and to be estimated, and K_j are positive. Notes: (i) For Variant II, $t_1 = 0$ (so $K_1 = 1$), and for Variant I, $t_1 \neq 0$ and to be estimated. (ii) $K_1 = \exp(t_1)$, $K_2 = \exp(a_{21})$ and $K_3 = \exp(K_1)$. (iii) The power link is $g_j(\theta_j) = \theta_j^{s_j}$ for a prespecified s_j.

Some Examples

1.
```
rrvglm(cbind(y1, y2) ~ x2 + x3, poissonff, data = pdata)
```

   results in the coupling $\mu_2(\mathbf{x}) = K_1 \cdot [\mu_1(\mathbf{x})]^{a_{21}}$.

2. RR-negative binomial distribution, RR-NB.

   A response $Y \sim \text{NB}(\mu, k)$ has $V(\mu) = \mu + \mu^2/k$. With the default $\boldsymbol{\eta} = (\log \mu,\ \log k)^T$ we have

$$V(\mu) \ = \ \mu + \delta_1 \, \mu^{\delta_2} \qquad (33)$$

   where the δ_j are to be estimated.

   Called a NB-P (Greene, 2008). Includes NB-1 ($V(\mu) \propto \mu$) and NB-2 as special cases.

   A confidence interval for δ_2 does not cover 1 or 2 $\Longrightarrow$ neither NB-1 or NB-2. Thus under- and over-dispersion relative to a Poisson and negative binomial distribution can be detected automatically.

3. RR-zero-inflated Poisson distribution, RR-ZIP.

Liu and Chan (2010) call this 'new' methodology a *COZIGAM* ('*constrained zero-inflated generalized additive model*'). It has

$$\mathrm{Pr}(Y = y) \;=\; I(y = 0) \cdot \phi + (1 - \phi)\, e^{-\mu} \mu^{y} / y!$$

with $\boldsymbol{\eta} = (\log \mu,\ \mathrm{logit}\,\phi)^{T}$ and $\eta_1 = t_1 + a_{21}\,\eta_2$.

Examples

- Trawl survey studies where the spatio-temporal aggregation of fish due to schooling means the probability of a positive catch is likely a monotonic function of the mean.
- Some grasshopper species' abundances affected by swarming due to suitable environmental conditions becoming available.

Example: Arizona Cardiovascular Patients

Hilbe (2011) describes a data set where 3589 patients entering a hospital in Arizona, USA, in 1991, to receive one of two standard cardiovascular treatments.

The variables in `azpro` are

- `los` length of (hospital) stay (response; probably in days),
- `procedure` (0 = PTCA, 1 = CABG)
- `sex` (1 = M, 0 = F),
- `admit` (0 = elective, 1 = urgent/emergency), and
- `age75` (0 if age $<$ 75 years, otherwise 1).

```
> rrnb.azpro <- rrvglm(los ~ procedure + sex + admit + age75,
                    negbinomial(zero = NULL), data = azpro)
> unlist(Confint.rrnb(rrnb.azpro))  # Neither a NB-1 nor NB-2

  a21.hat beta11.hat beta21.hat    CI.a211    CI.a212 CI.delta21 CI.delta22     delta1
  0.58884    1.44277    1.39004    0.40266    0.77502    1.22498    1.59734    0.58247
   delta2 SE.a21.hat
  1.41116    0.09499
```

Table : Estimated NB-2 and RR-NB coefficients and SEs for the azpro data.

	NB-2		RR-NB	
Coefficient	Estimate (SE)	Wald	Estimate (SE)	Wald
(Intercept)	1.418 (0.024)	60.1	1.443 (0.026)	55.9
procedure	0.981 (0.018)	53.6	0.972 (0.019)	50.7
sex	−0.126 (0.019)	−6.6	−0.121 (0.019)	−6.5
admit	0.371 (0.019)	19.5	0.333 (0.021)	16.2
age75	0.120 (0.020)	5.9	0.119 (0.020)	6.1

The fitted NB-2 purports a variance function of

$$V(\widehat{\mu}) \;\approx\; \widehat{\mu} + \frac{\widehat{\mu}^2}{6.24} \;\approx\; \widehat{\mu} + 0.16 \cdot \widehat{\mu}^2.$$

In fact, the fitted RR-NB has

$$V(\widehat{\mu}) \;\approx\; \widehat{\mu} + 0.58 \cdot \widehat{\mu}^{1.41}.$$

The 95% confidence for δ_2 is $[1.22, 1.6]$, therefore NB-1 and NB-2 are not strictly appropriate. This is supported by a likelihood ratio test of the RR-NB versus NB-2 with p-value 1.5×10^{-12}.

The profile log-likelihood as a function of $\widehat{a}_{21}$ (Slide 53) is well approximated by a quadratic. It shows both NB-1 and NB-2 models are inappropriate compared to the LR confidence limits.

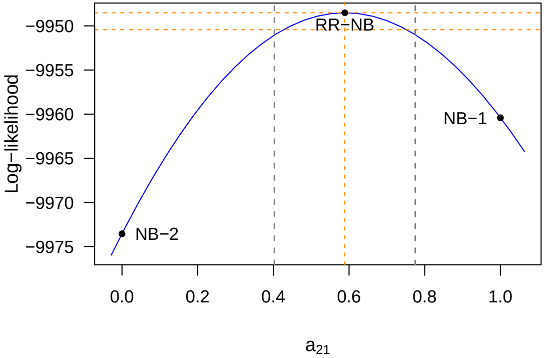

Figure : Profile log-likelihood $\ell(a_{21})$ for `rrnb.azpro`; $\ell(\widehat{a}_{21})$ is the highest point. The MLE and LR confidence limits are the orange horizontal lines. The Wald confidence limits are the grey vertical lines. The point at $a_{21} = 0$ is ℓ for NB-2 (intercept-only for k). The point at $a_{21} = 1$ is ℓ for NB-1.

Some Tables of Family Functions

The following are some distributions currently implemented by VGAM.

Positive Models

t

Distribution	Density function $f(y;\boldsymbol{\theta})$	VGAM family
Positive binomial	$\dfrac{1}{(1-p)^N}\binom{N}{Ny}p^{Ny}(1-p)^{N(1-y)}$	`posbinomial()`
Positive Poisson	$\dfrac{1}{1-e^{-\lambda}}\,\dfrac{e^{-\lambda}\lambda^y}{y!}$	`pospoisson()`
Positive NB	$\dfrac{1}{1-(k/(k+\mu))^k}\binom{y+k-1}{y}\left(\dfrac{\mu}{\mu+k}\right)^y\left(\dfrac{k}{k+\mu}\right)^k$	`posnegbinomial(zero = "size")`
Positive normal	$\dfrac{1}{\sigma}\,\dfrac{\phi((y-\mu)/\sigma)}{[1-\Phi(-\mu/\sigma)]}$	`posnormal()`

Table : Some positive distributions currently supported by VGAM. They are of the form $f(y)/\mathrm{Pr}(Y>0)$, e.g., $f(y)/(1-f(0))$.

Zero-inflated Models

t

Zero-inflated distribution	Probability function $f(y;\boldsymbol{\theta})$	VGAM family function
ZI binomial(ϕ, p)	$I(y=0)\phi+(1-\phi)\times \binom{N}{Ny} p^{Ny}(1-p)^{N(1-y)}$	`zibinomial(zero = NULL)`
ZI geometric(ϕ, p)	$I(y=0)\,\phi+(1-\phi)\,p(1-p)^{y}$	`zigeometric(zero = NULL)`
ZI negative binomial(ϕ, μ, k)	$I(y=0)\phi+(1-\phi)\times \binom{y+k-1}{y}\left(\frac{\mu}{\mu+k}\right)^{y}\left(\frac{k}{k+\mu}\right)^{k}$	`zinegbinomial(zero = "size")`
ZI Poisson(ϕ, λ)	$I(y=0)\phi+(1-\phi)\frac{e^{-\lambda}\lambda^{y}}{y!}$	`zipoisson(zero = NULL)`

Table : Summary of zero-inflated discrete distributions currently supported by VGAM. They are of the form $I(y=0)\phi+(1-\phi)f(y)$ where ϕ is the probability of a structural zero (`"pstr0"`).

t

Zero-inflated distribution	Probability function $f(y; \boldsymbol{\theta})$	VGAM family function
ZI binomial(p, ϕ^*)	$I(y=0)(1-\phi^*)+\phi^* \times \binom{N}{Ny} p^{Ny}(1-p)^{N(1-y)}$	`zibinomialff(zero = "onempstr0")`
ZI geometric(p, ϕ^*)	$I(y=0)\,(1-\phi^*)+\phi^*\; p(1-p)^y$	`zigeometricff(zero = "onempstr0")`
ZI negative binomial(μ, k, ϕ^*)	$I(y=0)\,(1-\phi^*)+\phi^* \times \binom{y+k-1}{y}\left(\frac{\mu}{\mu+k}\right)^y\left(\frac{k}{k+\mu}\right)^k$	`zinegbinomialff(zero = c(-2, -3))`
ZI Poisson(λ, ϕ^*)	$I(y=0)\phi^* + \phi^* \frac{e^{-\lambda}\lambda^y}{y!}$	`zipoissonff(zero = "onempstr0")`

Table : Summary of zero-inflated discrete distributions currently supported by VGAM. They are of the form $I(y=0)(1-\phi^*)+\phi^* f(y)$ where ϕ^* is the complement of the probability of a structural zero (`"onempstr0"`).

One-inflated Positive Models

t

Distribution	Probability function $f(y; \boldsymbol{\theta})$	VGAM family function
OI Pos-Poisson(ϕ, λ)	$I(y=1)\phi + (1-\phi)\dfrac{e^{-\lambda}\lambda^{y}}{(1-e^{-\lambda})y!}$	`oipospoisson(zero = NULL)`
ZOA Beta(s_1, s_2)	$I(y=0)p_0 + I(y=1)p_1 +$ $(1-p_0-p_1)\dfrac{y^{s_1-1}(1-y)^{s_2-1}}{Be(s_1, s_2)}$	`zoabetaR(zero = NULL)`

Table : Summary of one-inflated positive discrete distributions currently supported by VGAM.

Zero-altered Models

t

ZA distribution	Probability function $f(y;\boldsymbol{\theta})$	VGAM family
ZA-binomial(ω, p)	$I(y=0)\,\omega + I(y>0)\,(1-\omega)\,\binom{N}{Ny} p^{Ny}(1-p)^{N(1-y)}$	`zabinomial(zero = NULL)`
ZA-geometric(ω, p)	$I(y=0)\,\omega + I(y>0)\,(1-\omega)\,p(1-p)^{y-1}$	`zageometric(zero = NULL)`
ZA-NB(ω, μ, k)	$I(y=0)\,\omega + I(y>0)\,(1-\omega)\,\dfrac{1}{1-(k/(k+\mu))^{k}} \times \binom{y+k-1}{y}\left(\dfrac{\mu}{\mu+k}\right)^{y}\left(\dfrac{k}{k+\mu}\right)^{k}$	`zanegbinomial(zero = "size")`
ZA-Poisson(ω, λ)	$I(y=0)\,\omega + I(y>0)\,(1-\omega)\,\dfrac{e^{-\lambda}\lambda^{y}}{(1-e^{-\lambda})y!}$	`zapoisson(zero = NULL)`

Table : Summary of zero-altered discrete distributions currently supported by VGAM. They are of the form $I(y=0)\omega + I(y>0)(1-\omega)f(y)/[1-f(0)]$ where $\omega = \Pr(Y=0)$ is `"pobs0"`.

t

ZA distribution	Probability function $f(y; \boldsymbol{\theta})$	VGAM family
ZA-binomial(p, ω^*)	$I(y=0)\,(1-\omega^*) +$ $I(y>0)\,\omega^* \binom{N}{Ny} p^{Ny} (1-p)^{N(1-y)}$	`zabinomialff(zero = "onempobs0")`
ZA-geometric(p, ω^*)	$I(y=0)\,(1-\omega^*) + I(y>0)\,\omega^*\; p(1-p)^{y-1}$	`zageometricff(zero = "onempobs0")`
ZA-NB(ω^*, μ, k)	$I(y=0)\,(1-\omega^*) + I(y>0)\,\dfrac{\omega^*}{1-(k/(k+\mu))^k} \times$ $\binom{y+k-1}{y} \left(\dfrac{\mu}{\mu+k}\right)^y \left(\dfrac{k}{k+\mu}\right)^k$	`zanegbinomialff(zero=-(2:3))`
ZA-Poisson(ω^*, λ)	$I(y=0)\,(1-\omega^*) + I(y>0)\,\omega^* \dfrac{e^{-\lambda}\lambda^y}{(1-e^{-\lambda})y!}$	`zapoissonff(zero = "onempobs0")`

Table : Summary of zero-altered discrete distributions currently supported by VGAM. They are of the form $I(y=0)(1-\omega^*) + I(y>0)\,\omega^*\, f(y)/[1-f(0)]$ where $\omega^* = \Pr(Y>0)$ (`"onempobs0"`).

Zero-inflated, Zero-altered and Positive Models

t

Distribution	Random variates functions
Zero-altered binomial	`[dpqr]zabinom()`
Zero-altered geometric	`[dpqr]zageom()`
Zero-altered negative binomial	`[dpqr]zanegbin()`
Zero-altered Poisson	`[dpqr]zapois()`
Zero-inflated binomial	`[dpqr]zibinom()`
Zero-inflated geometric	`[dpqr]zigeom()`
Zero-inflated negative binomial	`[dpqr]zinegbin()`
Zero-inflated Poisson	`[dpqr]zipois()`
Positive binomial	`[dpqr]posbinom()`
Positive negative binomial	`[dpqr]posnegbinom()`
Positive normal	`[dpqr]posnorm()`
Positive Poisson	`[dpqr]pospois()`

Table : Some of VGAM functions for generating random variates etc. The prefix "d" = density, "p" = distribution function, "q" = quantile function and "r" = random deviates.

More Distributions

t

Distribution	Density function $f(y; \boldsymbol{\theta})$	Range of y	Range of $\boldsymbol{\theta}$	VGAM family
Beta-binomial	$\binom{N}{Ny} \frac{B(\alpha + Ny, \beta + N(1-y))}{B(\alpha, \beta)}$	$0(N^{-1})1$	$0 < \mu < 1, 0 < \rho < 1$	betabinomial()
Geometric	$(1-p)^y p$	$0(1)\infty$	$0 < p < 1$	geometric()
Pólya	$\binom{y+k-1}{y} p^k (1-p)^y$	$0(1)\infty$	$0 < p < 1, k > 0$	polya()
Zeta	$\left[y^{p+1} \zeta(p+1)\right]^{-1}$	$1(1)\infty$	$0 < p < \infty$	zetaff()
Zipf	$y^{-s} / \sum_{i=1}^{N} i^{-s}$	$1, 2, \ldots, N$	$s > 0$	zipf()

Table : More discrete univariate distributions currently supported by VGAM. Others are `dirichlet()` and `dirmultinomial()`.

Positive-Bernoulli Distribution for Capture-Recapture Data

t

Model	$\boldsymbol{\eta}$
$\mathcal{M}_0/\mathcal{M}_h$	$g(p)$
$\mathcal{M}_b/\mathcal{M}_{bh}$	$(g(p_c),\ g(p_r))^T$
$\mathcal{M}_t/\mathcal{M}_{th}$	$(g(p_1),\ \ldots,\ g(p_\tau))^T$
$\mathcal{M}_{tb}/\mathcal{M}_{tbh}$	$(g(p_{c1}),\ \ldots,\ g(p_{c\tau}),\ g(p_{r2}),\ \ldots,\ g(p_{r\tau}))^T$

Model	`family =`
$\mathcal{M}_0/\mathcal{M}_h$	`posbinomial(omit.constant = TRUE)`
	`posbernoulli.b(drop.b = FALSE ~ 0)`
	`posbernoulli.t(parallel.t = FALSE ~ 0)`
	`posbernoulli.tb(drop.b = FALSE ~ 0, parallel.t = FALSE ~ 0)`
$\mathcal{M}_b/\mathcal{M}_{bh}$	`posbernoulli.b()`
	`posbernoulli.tb(drop.b = FALSE ~ 1, parallel.t = FALSE ~ 0)`
$\mathcal{M}_t/\mathcal{M}_{th}$	`posbernoulli.t()`
	`posbernoulli.tb(drop.b = FALSE ~ 0, parallel.t = FALSE ~ 1)`
$\mathcal{M}_{tb}/\mathcal{M}_{tbh}$	`posbernoulli.tb()`

Table : Upper table: the $\boldsymbol{\eta}$ for the 8 Otis et al. (1978) models. Lower table: the relationships between the 8 models and function calls.

Categorical Data Analysis

t

Quantity	Range of j	VGAM Family function
$\mathrm{Pr}(Y = j+1)/\mathrm{Pr}(Y = j)$ $\mathrm{Pr}(Y = j)/\mathrm{Pr}(Y = j+1)$	$1, \ldots, M$ $2, \ldots, M+1$	`acat()` `acat(reverse = TRUE)`
$\mathrm{Pr}(Y > j \mid Y \geq j)$ $\mathrm{Pr}(Y < j \mid Y \leq j)$	$1, \ldots, M$ $2, \ldots, M+1$	`cratio()` `cratio(reverse = TRUE)`
$\mathrm{Pr}(Y \leq j)$ $\mathrm{Pr}(Y \geq j)$	$1, \ldots, M$ $2, \ldots, M+1$	`cumulative()` `cumulative(rev = TRUE)`
$\mathrm{logit}\ \mathrm{Pr}(Y \leq j)$ $\mathrm{logit}\ \mathrm{Pr}(Y \geq j)$	$1, \ldots, M$ $2, \ldots, M+1$	`propodds(rev = FALSE)` `propodds()`
$\log\{\mathrm{Pr}(Y = j)/\mathrm{Pr}(Y = M+1)\}$	$1, \ldots, M$	`multinomial()`
$\mathrm{Pr}(Y = j \mid Y \geq j)$	$1, \ldots, M$	`sratio()`
$\mathrm{Pr}(Y = j \mid Y \leq j)$	$2, \ldots, M+1$	`sratio(reverse = TRUE)`

Table : Quantities defined in VGAM for a categorical response Y taking values $1, \ldots, M+1$. For more details see Yee (2010).

New family functions since Yee (2015)

t

Distribution	Probability function $f(y;\boldsymbol{\theta})$	Support	Range	VGAM family function
OI Pos-binom(ϕ, p)	$I(y=1/N)\phi + (1-\phi)\times \frac{\binom{N}{Ny}\, p^{Ny}(1-p)^{N(1-y)}}{1-(1-p)^{N}}$	$(1/N)(1/N)1$	$0 < p < 1$	`oiposbinomial(zero="")`
OI Pos-Pois(ϕ, λ)	$I(y=1)\,\phi + \frac{(1-\phi)\,e^{-\lambda}\,\lambda^{y}}{(1-e^{-\lambda})\,y!}$	$1(1)\infty$	$0 < \lambda$	`oipospoisson(zero="")`
OI Zipf(ϕ, s; N)	$I(y=1)\,\phi + \frac{1-\phi}{y^{s}\,H_{N,s}}$	$1(1)N$	$0 < s$	`oizipf(zero="")`
OI Zeta(ϕ, s)	$I(y=1)\,\phi + \frac{1-\phi}{y^{s+1}\,\zeta(s+1)}$	$1(1)\infty$	$0 < s$	`oizeta(zero="")`
OI Logarithmic(ϕ, c)	$I(y=1)\,\phi + \frac{(1-\phi)a\cdot c^{y}}{y}$	$1(1)\infty$	$0 < c < 1$	`oilog(zero="")`
ZOA Beta(s_1, s_2)	$I(y=0)\,\omega_0 + I(y=1)\,\omega_1 +$	$[0,1]$	$0 < s_j$	`zoabetaR(zero="")`

Table : New VGAM family functions written since Yee (2015) that pertain to 0- and/or 1-inflation.

t

Distribution	PMF/PDF $f(y; \boldsymbol{\theta})$	Support	Range	VGAM family
OA Zeta(ω, s)	$I[y=1]\,\omega + I[y>1]\,(1-\omega)\,f_{\texttt{otzeta}}(y)$	$1(1)\infty$	$0 < s$	`oazeta()`
OA Pos-Pois(λ)	$I[y=1]\,\omega + I[y>1]\,(1-\omega)\,f_{\texttt{otpospoisson}}(y)$	$1(1)\infty$	$0 < \lambda$	`oapospoisson()`
OA Logarithmic(c)	$I[y=1]\,\omega + I[y>1]\,(1-\omega)\,f_{\texttt{otlog}}(y)$	$1(1)\infty$	$0 < c < 1$	`oalog()`
OT Logarithmic(c)	$f_{\texttt{otlog}} = \dfrac{(-1)\cdot c^{y}}{y\,[c + \log(1-c)]}$	$2(1)\infty$	$0 < c < 1$	`otlog()`
OT Pos-Pois(λ)	$f_{\texttt{otpospoisson}}$	$2(1)\infty$	$0 < \lambda$	`otpospoisson()`
OT zeta(s)	$f_{\texttt{otzeta}} = \dfrac{1}{y^{s+1}\,[\zeta(s+1) - 1 - 2^{-s-1}]}$	$2(1)\infty$	$0 < s$	`otzeta()`
Differenced Zeta(s)	$\left(\dfrac{a}{y}\right)^{s} - \left(\dfrac{a}{1+y}\right)^{s}$	$a(1)\infty$	$0 < s$	`diffzeta()`

Table : Continuation of the previous table. Here, "OT" means one-truncated. Some expressions are omitted for brevity.

Current Work

```
> fit1 <- vglm(y ~ x2 + x3, genapoisson(alter    = c(a1, a2)), data = gdata)
> fit2 <- vglm(y ~ x2 + x3, genipoisson(inflate  = c(i1, i2)), data = gdata)
> fit3 <- vglm(y ~ x2 + x3, gentpoisson(truncate = c(t1, t2)), data = gdata)
```

Generally-truncated Poisson Distribution

$$\Pr(Y = y;\ \lambda) \quad = \quad \frac{e^{-\lambda}\,\lambda^{y}}{y!\,[1 - \sum_{t \in \mathcal{T}} f(t)]}, \qquad y = 0(1)\infty \setminus \mathcal{T}, \tag{34}$$

where the `truncate` argument contains the set $\mathcal{T}$, and $f(y) = e^{-\lambda}\lambda^{y}/y!$ is the usual Poisson PMF, and $0 < \lambda$.

VGAM models (34) by

$$g(\lambda) \quad = \quad \eta \quad = \quad \boldsymbol{\beta}^{T}\mathbf{x}, \tag{35}$$

where $g = \log$ by default.

Some notes:

1. Fitted values:

$$E[Y] \quad = \quad \frac{\mu_* - \sum_{t \in \mathcal{T}} \; t \cdot f(t)}{1 - \sum\limits_{t \in \mathcal{T}} f(t)}$$

   where $\mu_* = \lambda$ is the mean of the parent distribution.

2. Its EIM is easily computed, hence its VGAM implementation. Once implemented, all VGAM infrastructure is available, e.g., constraints-on-the-function, `xij` facility, smoothing and additive models, residuals and diagnostics, inference.

3. The `dpqr`-type functions have been or will be written: `dgentpois()`, `pgentpois()`, `qgentpois()`, `rgentpois()`.

4. In practice it is a good idea to restrict $|\mathcal{T}|$ to be no more than 3 or 4, since the model can be easily misused.

Generally-altered Poisson Distribution

$$\Pr(Y = y; \lambda) \quad = \quad \begin{cases} \omega_1, & y = a_1, \\ \quad\vdots & \quad\vdots \\ \omega_L, & y = a_L, \\ \dfrac{(1 - \omega_1 - \cdots - \omega_L)\, f(y)}{1 - \sum\limits_{a \in \mathcal{A}} f(a)}, & y = 0(1)\infty \setminus \mathcal{A}, \end{cases} \tag{36}$$

where the `alter` argument contains the set $\mathcal{A}$, and $f(y) = e^{-\lambda}\lambda^y/y!$ is the usual Poisson PMF, and $0 < \lambda$.

Each of the L probabilities $\omega_1, \ldots, \omega_L$ are modelled as a logistic regression by default. The ordinary Poisson mean parameter is also modelled like a Poisson regression. Thus VGAM has $M = L + 1$ linear predictors for this model, and VGAM models (36) by something like

$$
\begin{aligned}
g_1(\omega_1) &= \eta_1 = \boldsymbol{\beta}_1^T \mathbf{x}, \\
&\vdots \quad \vdots \\
g_1(\omega_L) &= \eta_L = \boldsymbol{\beta}_L^T \mathbf{x}, \\
g_2(\lambda) &= \eta_{L+1} = \boldsymbol{\beta}_{L+1}^T \mathbf{x},
\end{aligned} \tag{37}
$$

where $g_1 = \mathrm{logit}$ and $g_2 = \log$ by default.

Loosely speaking, the probabilities ω_j allow both inflation and deflation relative to its usual value, for at support value a_j. So in one way generally-altered distributions are more flexible than generally-inflated distributions, although the latter does allow for a milder form of deflation.

Some notes:

1. Fitted values:

$$E[Y] \;\; = \;\; \sum_{a \in \mathcal{A}} a \cdot f(a) + \frac{1 - \sum_{a \in \mathcal{A}} \omega_a}{1 - \sum_{a \in \mathcal{A}} f(a)} \left[\mu_* - \sum_{a \in \mathcal{A}} a \cdot f(a) \right],$$

   where $\mu_* = \lambda$ is the mean of the parent distribution.

2. The notes about the generally-truncated Poisson distribution in Slide 68 apply similarly.

Generally-inflated Poisson Distribution

$$\Pr(Y = y; \lambda) = \begin{cases} \phi_1 + (1 - \phi_1 - \cdots - \phi_L)\, f(i_1), & y = i_1, \\ \quad\vdots & \quad\vdots \\ \phi_L + (1 - \phi_1 - \cdots - \phi_L)\, f(i_L), & y = i_L, \\ (1 - \phi_1 - \cdots - \phi_L)\, f(y), & y = 0(1)\infty \setminus \mathcal{I}, \end{cases} \tag{38}$$

where the `inflate` argument contains the set $\mathcal{I}$, and $f(y) = e^{-\lambda}\lambda^y/y!$ is the usual Poisson PMF, and $0 < \lambda$.

Each of the L mixing probabilities $\phi_1, \ldots, \phi_L$ are modelled as a logistic regression by default. The ordinary Poisson mean parameter is also modelled like a Poisson regression. Thus VGAM has $M = L + 1$ linear predictors for this model, and VGAM models (38) by something like

$$\begin{aligned} g_1(\phi_1) &= \eta_1 = \boldsymbol{\beta}_1^T \mathbf{x}, \\ &\vdots \quad \vdots \\ g_1(\phi_L) &= \eta_L = \boldsymbol{\beta}_L^T \mathbf{x}, \\ g_2(\lambda) &= \eta_{L+1} = \boldsymbol{\beta}_{L+1}^T \mathbf{x}, \end{aligned} \tag{39}$$

where $g_1 = \text{logit}$ and $g_2 = \log$ by default.

Note that each inflated value of the support has two sources of probability: the first is from the mixing probability ϕ_j and the second from the Poisson or parent distribution. The first type might be called "*structural*", such as a structural 0 for a zero-inflated Poisson distribution.

It is possible to allow for *deflation* whereby some of the ϕ_j are negative. As long as each probability in (38) is positive, this model is perfectly valid.

Some notes:

1. Fitted values:

$$E[Y] \;=\; \sum_{i \in \mathcal{I}} i \cdot f(i) + \left(1 - \sum_{i \in \mathcal{I}} \phi_i\right) \mu_*,$$

   where $\mu_* = \lambda$ is the mean of the parent distribution.

2. The notes about the generally-truncated Poisson distribution in Slide 68 apply similarly.

Future Work

There is a lot one can do for future work, e.g.,

1. Develop some real-world applications.
2. Repeat the work for other distributions, e.g., binomial, logarithmic, negative binomial, zeta, Zipf.

Concluding Remarks

1. The VGLM/VGAM/. . . framework provides a large umbrella to fit many classical regression models, including count distributions.
2. Crucially, the EIM is needed. (Please send me any!)
3. Collaborations with real-life problems are welcome.

Concluding Remarks

1. The VGLM/VGAM/... framework provides a large umbrella to fit many classical regression models, including count distributions.
2. Crucially, the EIM is needed. (Please send me any!)
3. Collaborations with real-life problems are welcome.
4. There is still a lot more work to be done...

Concluding Remarks

1. The VGLM/VGAM/... framework provides a large umbrella to fit many classical regression models, including count distributions.
2. Crucially, the EIM is needed. (Please send me any!)
3. Collaborations with real-life problems are welcome.
4. There is still a lot more work to be done...

Fin

References

- Anderson, J. A., 1984. Regression and ordered categorical variables. J. Roy. Statist. Soc. Ser. B 46 (1), 1–30, with discussion.
- Hastie, T. J., Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall, London.
- Hilbe, J. M., 2011. Negative Binomial Regression, 2nd Edition. Cambridge University Press, Cambridge, UK; New York, USA.
- Liu, H., Chan, K. S., 2010. Introducing COZIGAM: An R package for unconstrained and constrained zero-inflated generalized additive model analysis. J. Statist. Soft. 35 (11), 1–26.
- Wand, M. P., Ormerod, J. T., 2008. On semiparametric regression with O'Sullivan penalized splines. Australian & New Zealand Journal of Statistics 50 (2), 179–198.
- Wood, S. N., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall, London.

Yee, T. W., 2010. The VGAM package for categorical data analysis. J. Statist. Soft. 32 (10), 1–34.
URL http://www.jstatsoft.org/v32/i10/

Yee, T. W., 2015. Vector Generalized Linear and Additive Models: With an Implementation in R. Springer, New York, USA.

Yee, T. W., Wild, C. J., 1996. Vector generalized additive models. J. Roy. Statist. Soc. Ser. B 58 (3), 481–493.