VGAM Family Functions for Univariate Distributions

Beta Version 0.5-15

© Thomas W. Yee

Department of Statistics,
University of Auckland,
New Zealand
yee@stat.auckland.ac.nz
http://www.stat.auckland.ac.nz/∼yee

January 16, 2004

Contents

[Important note: This document and code is not yet finished, but should be completed one day . . .]

1 Introduction

This document describes in detail VGAM family functions for fitting data to univariate distributions.

For general theory and instructions about the VGAM software library see the VGAM *User Manual*. S-PLUS's `glm()` and `gam()` are described in Chambers and Hastie (1993), where as VGAM family functions for GLMs and GAMs are documented elsewhere.

Table 1 summarized the models currently implemented by VGAM and described here.

Table 1: VGAM *family functions documented in this article.*

`beta2()`	2-parameter Beta distribution
`beta4()`	4-parameter Beta distribution
`betabin()`	Beta-binomial distribution
`chisq()`	χ^2 distribution
`gamma1()`	1-parameter gamma distribution
`gamma2()`	2-parameter gamma distribution
`geometric()`	Geometric distribution
`gpoisson()`	Generalized Poisson distribution
`gumbel()`	Gumbel distribution
`neg.binomial()`	Negative binomial (preferred)
`negbin.mu()`	Negative binomial
`negbin.mn()`	Negative binomial (p.373 of McCullagh and Nelder (1989))
`neg.binomial.k()`	Negative binomial (k known)
`normal1()`	Normal (μ unknown, σ known)
`skewnorm1()`	Skewed normal
`Zeta()`	Zeta distribution

2 Dispersion Models

A reproductive dispersion model with positional parameter μ and dispersion parameter σ^2 is a family of distributions whose probability density functions are of the form

$$f(y;\mu,\sigma^2) \;=\; a(y;\sigma^2)\,\exp\left\{-\frac{1}{2\sigma^2}\,d(y;\mu)\right\}. \tag{1}$$

where $a \geq 0$ is a suitable function, and d is a unit deviance. Equation (1) is said to be of standard form. Jørgensen (1997) discusses many models within a dispersion model framework. The VGLM framework is too general to take advantage of the special structure of dispersion models, however, many of the models discussed in that book have been implemented in VGAM. See Table 2. Other dispersion models are discussed in the GLM framework.

Distribution	Density function $f(y;\boldsymbol{\theta})$	Range of y	Range of $\boldsymbol{\theta}$[a]	Mean	VGAM family function
Negative binomial	$\binom{y+k-1}{y} p^k (1-p)^y$	$\{0,1,2,\ldots\}$	$0<p<1,\ k>0$	$\frac{k(1-p)}{p}$	`neg.binomial`
Negative binomial	$\binom{y+k-1}{y} \left(\frac{\mu}{\mu+k}\right)^y \left(\frac{k}{k+\mu}\right)^k$	$\{0,1,2,\ldots\}$	$\mu>0,\ k>0$	μ	`negbin.mu`
Negative binomial	$\binom{y+k-1}{y} \alpha^y (1+\alpha)^{-(y+k)},$	$\{0,1,2,\ldots\}$	$\alpha>0,\ k>0$	$k\alpha$	`negbin.mn`
Hyperbolic secant	$\frac{\exp\{\theta y + \log(\cos(\theta))\}}{2\cosh(\pi y/2)}$	$(-\infty,\infty)$	$\theta\in(-\pi/2,\pi/2)$	$\tan\theta$	`hyper.secant`
Hyperbolic secant	$\frac{\cos(\theta)}{\pi} u^{-\frac{1}{2}+\theta/\pi}(1-u)^{-\frac{1}{2}-\theta/\pi}$	$(0,1)$	$\theta\in(-\pi/2,\pi/2)$	$\frac{1}{2}+\frac{\theta}{\pi}$	`hyper.secant1`
Inverse binomial	$\frac{\lambda\,\Gamma(2y+\lambda)}{\Gamma(y+1)\,\Gamma(y+\lambda+1)}\{\rho(1-\rho)\}^y\rho^\lambda$	$\{0,1,2,\ldots\}$	$\frac{1}{2}<\rho<1,\ \lambda>0$	$\frac{\lambda(1-\rho)}{2\rho-1}$	`invbinomial`
Reciprocal inverse Gaussian	$\sqrt{\frac{\lambda}{2\pi y}}\exp\left\{-\frac{\lambda(y-\mu)^2}{2y}\right\}$	$(0,\infty)$	$\lambda>0$	μ	`rig`
Leipnik	$\frac{\{y(1-y)\}^{-\frac{1}{2}}}{\text{Beta}(\frac{\lambda+1}{2},\frac{1}{2})}\left[1+\frac{(y-\mu)^2}{y(1-y)}\right]^{-\frac{\lambda}{2}}$	$(0,1)$	$\lambda>-1$	μ	`leipnik`
Generalized Poisson	$\frac{\theta(\theta+y\lambda)^{y-1}}{y!}\exp(-y\lambda-\theta),$	$\{0,1,2,\ldots\}$	$0\le\lambda<1,\ \theta>0$	$\frac{\theta}{1-\lambda}$	`genpoisson`
Simplex	$\frac{\exp\left\{-\frac{1}{2\sigma^2}\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right\}}{\sqrt{2\pi\sigma^2\{y(1-y)\}^3}}$	$(0,1)$	$\sigma>0$	μ	`simplex`

Table 2: *Dispersion models implemented in* VGAM.

[a] $-\infty<\theta_j<\infty$ and θ_j real-valued is assumed unless otherwise stated.

3 Negative Binomial Distribution

Many different models give rise to this distribution, and consequently there are several parameterizations. We write this discrete distribution as

$$f(y;p,k) \;=\; \binom{y+k-1}{y} p^k\,(1-p)^y, \quad 0<p<1, \quad k>0, \quad y=0,1,2,\ldots, \tag{2}$$

where

$$\binom{a}{b} \;=\; \frac{a(a-1)\times\cdots\times(a-b+1)}{b!}$$

for a positive real a and a positive integer b, and $\binom{a}{0}=1$. Then the mean and variance of Y are $\mu = k\,(1-p)/p$ and $k\,(1-p)/p^2 = \mu+\mu^2/k$ respectively. Note that k is real.

The negative binomial is related to several other distributions. When k is an integer, the distribution is sometimes called the *Pascal distribution*, and the geometric distribution corresponds to $k=1$. If k is integer, then the negative binomial distribution can be thought of as modelling $Y=$ the number of failures before k successes in successive Bernoulli trials. The negative binomial distribution provides an example of a variance function containing an unknown parameter that is not a dispersion parameter.

The Poisson distribution, another special case, occurs when $k\to\infty$. The log series distribution occurs when zeros are missing and $k\to 0^+$. The negative binomial distribution has become popular because it is a more flexible alternative to the Poisson distribution. A good treatment of the negative binomial distribution is Johnson et al. (1993).

Little use of the negative binomial distribution seems to be have been made in practice, partly because of the problem that the linear predictor is a function of a parameter of the variance function: the distribution has the form of a GLM with canonical link

$$\eta \;=\; \log(1-p) \;=\; \log\left(\frac{\mu}{\mu+k}\right), \tag{3}$$

and variance function $V=\mu(1+\mu/k)$. However, this problem is readily handled within the VGAM framework, though VGAM does **not** use the canonical link function above as the default. The reason is that if $\eta>0$ then $p=1-e^\eta<0$ resulting in a range error. Instead we let $\eta_1=\mathrm{logit}(p)$ as the default link. However, (3) may be fitted by specifying `neg.binomial(link=logc)` or `neg.binomial(link=canonical)`.

Additionally, we let $\eta_2=\log k$. Then

$$\begin{aligned}\ell &= \sum_{i=1}^n y_i\log(1-p_i)+k_i\,\log p_i+\log(y_i+k_i-1)!-\log(k_i-1)!-\log y_i!\\ &= \sum_{i=1}^n y_i\log(1-p_i)+k_i\log p_i+\log\Gamma(y_i+k_i)-\log\Gamma(k_i)-\log\Gamma(y_i+1)\,.\end{aligned}$$

Although the partial derivatives involve the digamma function $\psi(z)=\Gamma'(z)/\Gamma(z)$ and trigamma function $\psi'(z)$, one does not actually need these functions implemented—see below for details.

Here are some notes.

Distribution	Density function $f(y;\boldsymbol{\theta})$	Range of y	Range of $\boldsymbol{\theta}$[a]	Mean	VGAM family function
Chi	$\frac{y^{\nu-1}\exp\{-y^2/2\}}{2^{(\nu/2)-1}\Gamma(\nu/2)}$	$(0,\infty)$	$\nu > 0$	$\Gamma((\nu+2)/2)/\Gamma(\nu/2)$	`chi`
Noncentral χ^2	$\frac{\exp\{-\frac{1}{2}(y+\lambda)\}}{2^{\nu/2}}\sum_{k=0}^{\infty}\frac{y^{\nu/2+k-1}\lambda^k}{\Gamma(\nu/2+k)2^{2k}k!}$	$(0,\infty)$	$\lambda > 0$	$\nu+\lambda$	`ncchisq`

Table 3: *More models implemented in* VGAM.

[a] $-\infty < \theta_j < \infty$ and θ_j real-valued is assumed unless otherwise stated.

1. If k is known a priori then one can use the family `neg.binomial.k()`. It has both ℓ and deviance D. This is useful because `neg.binomial()` has only ℓ because an expression for D does not exist. Then it is possible to compare `neg.binomial.k()` and `neg.binomial()` models.

2. By default, k is modelled as a constant (as `neg.binomial(zero=2)`). As k is positive, a number of links are provided, including the reciprocal link, which may be more numerically stable than just k or even $\log\, k$.

3. The expected Hessian (rather than the observed Hessian) is computed. One has, because y_i is integer-valued,

$$\frac{\partial^2 \ell_i}{\partial k^2} \;=\; \psi'(y_i + k) - \psi'(k) \;=\; -\sum_{r=0}^{y_i-1} (k+r)^{-2},$$

   where $\psi'(z)$ is the trigamma function (the digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$). Its expected value is computed numerically using the relation $\psi'(z+1) = \psi'(z) - z^{-2}$. Practical examples have suggested the observed Hessian is often not positive-definite, thus being more problematic.

   Also, $\psi(1) = -\gamma$, $\psi(z+1) = \psi(z) + z^{-1}$,

$$\frac{\partial \ell_i}{\partial k} \;=\; \log p + \psi(y_i + k) - \psi(k) \;=\; \log p + \sum_{r=0}^{y_i-1} (k+r)^{-1}.$$

   as $\psi(y_i) = -\gamma + \sum_{r=1}^{y_i-1} r^{-1}$, $y_i \geq 2$, where $\gamma \approx 0.5772\ldots$ is Euler's constant.

4. `glm.nb()` in the MASS library, which is similar to VGAM, can only model k as a scalar parameter.

5. `negbin.mn()` has been written to implement the parameterization of p.373 of McCullagh and Nelder (1989), viz.

$$f(y; \alpha, k) \;=\; \binom{y+k-1}{y} \alpha^y \, (1+\alpha)^{-(y+k)}, \quad \alpha > 0, \quad k > 0, \quad y = 0, 1, 2, \ldots. \tag{4}$$

   However, (2) has the advantage that probabilities have more link functions that can be specified, where as only $\log(\alpha)$ would be feasible in (4). It is related to (2) by $p = (1+\alpha)^{-1}$. Consequently, $\text{logit}\, p \equiv -\log\, \alpha$.

   This parameterization also appears in, e.g., p.199 of Johnson et al. (1993), p.375 of Cameron and Trivedi (1998). For this, Y has mean $k\alpha$ and variance $k\alpha(1+\alpha)$.

6. Yet another popular parameterization (especially in ecology) is

$$f(y; \mu, k) \;=\; \binom{y+k-1}{y} \left(\frac{\mu}{\mu+k}\right)^y \left(\frac{k}{k+\mu}\right)^k \quad \mu, \; k > 0, \quad y = 0, 1, 2, \ldots. \tag{5}$$

   For this, Y has mean μ and variance $\mu(1 + \mu/k)$. The parameter k is known as the *dispersion* parameter. The family function `negbin.mu()` has been written for this. This often requires half-stepsizing at early iterations, but should work ok for most data. If `crit="coeff"` is specified then you

may need to use `step=0.5` to take half-steps because the k parameter is hard to estimate. Some computational details are given in Lawless (1987).

The parameters μ and k are independent, and the confidence region for k is extremely skewed so that its standard error is often of no practical use. The parameter k^{-1} has been used as a measure of aggregation; positive values imply individuals tend to aggregate into groups, and negative values imply individuals tend to disperse. The negative binomial is an example of a contagious distribution.

Parameterization (5) is amenable with QRR-VGLMs (Yee, 2004), therefore can be used in the form `rrvglm(..., fam=negbin.mu, Quadratic=T)`.

VGAM has the capabilities of fitting additive models to negative binomial data using `vgam()`; see Thurston et al. (2000).

3.1 Example

In the following we consider an example from Bliss and Fisher (1953). From each of 6 McIntosh apple trees in an orchard that had been sprayed, 25 leaves were randomly selected. On each of the leaves, the number of adult female European red mites were counted.

```
> y = 0:7; w = c(70, 38, 17, 10, 9, 3, 2, 1)
> fit = vglm(y ~ 1, neg.binomial, wei=w)
> summary(fit)

Call:
vglm(formula = y ~ 1, family = neg.binomial, weights = w)

Pearson Residuals:
             Min        1Q  Median      3Q    Max
logit(p) -5.2399 -4.641981 -3.1257 0.98617 4.1725
  log(k) -8.9258  0.094723  2.0885 4.18016 5.1629

Coefficients:
                  Value Std. Error   t value
(Intercept):1 -0.112566    0.30192 -0.372840
(Intercept):2  0.024293    0.28077  0.086522

Number of linear predictors:  2

Names of linear predictors: logit(p), log(k)

Dispersion Parameter for neg.binomial family:   1

Log-likelihood: -222.4372 on 14 degrees of freedom

Number of Iterations: 10

> k = eta2theta(coef(fit)[2], fit$misc$link[2])    # Estimate of k
> p = k / (k + fitted(fit))
```

```
> logpdf = y * log(1-p) + k * log(p) + lgamma(y+k) - lgamma(k) - lgamma(y+1)
> pdf = exp(logpdf)
> print( cbind(y, w, fitted=pdf*sum(w)), dig=1)
     y  w fitted
[1,] 0 70   69.5
[2,] 1 38   37.6
[3,] 2 17   20.1
[4,] 3 10   10.7
[5,] 4  9    5.7
[6,] 5  3    3.0
[7,] 6  2    1.6
[8,] 7  1    0.8

# chisquare statistic (pooled so that expected counts >= 5)
# Bliss and Fisher, p.183, gives P=0.48
> ppdf = c(pdf[1:5], sum(pdf[6:8]))
> pw = c(w[1:5], sum(w[6:8]))
> fv = ppdf*sum(pw)
> 1-pchisq(sum((pw-fv)^2 / fv), df=length(fv)-1-length(coef(fit)))
[1] [1] 0.4726398
```

One has $\hat{k} = 1.025$ and $\hat{p} = 0.472$.

To check whether the negative binomial fits the data, we pool the $y \geq 5$ categories so that the expected value is ≥ 5. A chi-square test for the goodness of fit then yields a chi-square statistic of 2.52 on 3 degrees of freedom, giving a p-value of 0.48. Thus the data are quite consistent with the negative binomial distribution. In contrast, a Poisson distribution does not fit as it can be shown that a chi-square test gives a test statistic of 82.4 on 5 degrees of freedom, hence an extremely small p-value.

4 Beta-binomial model

The beta-binomial model, a random-effects binomial model, is commonly used for the analysis of teratological data. The classical example of this type of experiment involves a group of n pregnant rats which are randomized and exposed to a chemical. The ith rat gives birth to a litter of size n_i, of which y_i^* are malformed or die within a specified time period. Table 4 gives a dataset of this description. The scientific objective is to determine whether the risk of malformation differs between groups. Here, there were 32 pregnant rats which were randomized to receive a placebo or a chemical, and sacrificed prior to the end of gestation or pregnancy. Each fetus was examined and a binary response indicating the presence or absence of a particular malformation was recorded.

The beta-binomial model fitted by VGAM is as follows. It is assumed that a random malformation probability Π_i in cluster i comes from a beta distribution with mean μ_i. Given $\Pi_i = \pi_i$, the number of malformations Y_i^* within the ith cluster follows a Binomial(n_i, π_i) distribution. The model

$$\begin{pmatrix} \log\left(\dfrac{\mu_i}{1-\mu_i}\right) \\ \log\left(\dfrac{1+\rho_i}{1-\rho_i}\right) \end{pmatrix} = \begin{pmatrix} \eta_1(\boldsymbol{x}_i) \\ \eta_2(\boldsymbol{x}_i) \end{pmatrix} = \boldsymbol{\eta}(\boldsymbol{x}_i), \tag{6}$$

called the *extended* beta-binomial model (Prentice, 1986), provide transformations that restrict μ_i and the intracorrelation ρ_i to the intervals $(0,1)$ and $(-1,1)$ respectively. Here, $\rho_i = \mathrm{Corr}(Z_{ij}, Z_{ik})$ denote the intralitter correlation (an exchangeable correlation structure is assumed) where $Y_i^* = \sum_{j=1}^{n_i} Z_{ij}$, $Z_{ij} = 0$ or 1 is binary. A correlation parameter ρ that is positive/negative gives rise to over-/under-dispersed data respectively. The log-likelihood is given by $\ell = \sum_{i=1}^{n} \ell_i$ where

$$\ell_i = \log\binom{n_i}{y_i^*} + \sum_{r=0}^{y_i^*-1} \log\left(\mu_i + \frac{r\,\rho_i}{1-\rho_i}\right) + \sum_{r=0}^{n_i-y_i^*-1} \log\left(1-\mu_i + \frac{r\,\rho_i}{1-\rho_i}\right) - \sum_{r=0}^{n_i-1} \log\left(1 + \frac{r\,\rho_i}{1-\rho_i}\right).$$

Furthermore,

$$\begin{aligned} E(Y_i^*) &= n_i\,\mu_i, \\ \rho_i &= \frac{P(Z_{ij}=1, Z_{ik}=1) - \mu_i^2}{\mu_i(1-\mu_i)}, \\ \mathrm{Var}(Y_i^*) &= n_i\,\mu_i(1-\mu_i)\left\{1 + (n_i-1)\rho_i\right\}\ . \end{aligned} \tag{7}$$

Table 4: *Toxological experiment data. The numerator is the number with birth defects out of the size of the litter (denominator).*

t	
Control	13/13, 12/12, 9/9, 9/9, 8/8, 8/8, 12/13, 11/12, 9/10, 9/10, 8/9, 11/13, 4/5, 5/7, 7/10, 7/10
Treated	12/12, 11/11, 10/10, 9/9, 10/11, 9/10, 9/10, 8/9, 8/9, 4/5, 7/9, 4/7, 5/10, 3/6, 3/10, 0/7

However, like `Binomial()`, the response variable Y is a *proportion* rather than the number of successes. Thus $Y_i = Y_i^*/n_i$, $E(Y_i) = \mu_i$, the logit link $g(\mu) = \log(\mu/(1-\mu))$ is the default, and the n_i are assimulated as weights.

Note that $\Pi \sim$ Beta(α, β) means the probability density function is

$$f(\pi) \;=\; \frac{1}{B(\alpha,\beta)}\,\pi^{\alpha-1}\,(1-\pi)^{\beta-1}, \quad \alpha>0, \quad \beta>0, \quad 0<\pi<1.$$

Then $E(\Pi) = \alpha/(\alpha+\beta)$ and $\mathrm{Var}(\Pi) = \alpha\,\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$. It then follows that $\mu = \alpha/(\alpha+\beta)$ and $\rho = (\alpha+\beta+1)^{-1}$.

By default, the software fits ρ not as a function of the covariates: `betabin(zero=2)`, meaning η_2 is modelled as a constant only. VGAM also allows a different specification of the probability link function η_1, e.g., probit and complementary log-log links instead of logit in (6). Additionally, it is rare for $\rho < 0$, hence logit, probit and complementary log-log links are also provided for ρ.

Practical experience has shown that the convergence of the estimation process depends much upon the initial value of ρ. If VGAM fails to converge, users are directed to using the `coefstart`/`etastart`/`mustart` options.

Here are some more notes.

1. In general, one difficulty with fitting the beta-binomial distribution with IRLS is that the $\mathbf{W}_i$ are not all positive-definite. To fix this problem, if an intercept model is fitted, then the $\mathbf{W}_i$ are replaced by their weighted mean. This results in ordinary Newton-Raphson, and `object$misc$pooled.weight` is `TRUE`.

2. The beta-binomial model was proposed by Williams (1975) and has been used widely, e.g., to model the incidence of non-infectious diseases in a household, the number of malformed fetuses in a litter, and the number of chromosomal aberrant cells among repeated samples for an individual.

3. Initially we assume all covariates are cluster-specific, and that interest focuses on the proportion, μ, of successful responses in a cluster.

4. Because of (7), ρ is known as the over-dispersion parameter. A litter effect is typically reflected by a positive value of ρ. Prentice (1986) pointed out that the correlation coefficient ρ need not be positive as previously thought. He showed that the lower bound is

$$\max\left\{\frac{-\mu_i}{n_i-\mu_i-1}, \frac{-(1-\mu_i)}{n_i+\mu_i-2}\right\}.$$

5. Writing $\theta_i = \rho_i/(1-\rho_i) = 1/(\alpha_i+\beta_i)$, then $\mathrm{Var}(Y_i^*) = n_i\mu_i(1-\mu_i)(n_i\theta_i+1)/(1+\theta_i)$. Thus, the binomial model is a special case of the beta-binomial model with $\theta_i = 0$. Note that choosing `link.rho="logit"` means $\theta = \exp(\eta_2)$.

   It is usual to test for homogeneity of proportions (in the presence of common dispersion): $H_0 : \mu_1 = \cdots = \mu_n$ versus H_0 : not all the μ_i are equal, with the assumption that $\theta_1 = \cdots = \theta_n = \theta$ which is unknown and unspecified. If H_0 is not rejected, then it is often assumed that $\alpha_1 = \cdots = \alpha_n$ ($= \alpha$, say) and $\beta_1 = \cdots = \beta_n$ ($= \beta$, say), and then α and β are estimated by maximum likelihood.

6. The beta-binomial model has been shown to provide a much better fit to many data sets than the simple binomial model. However, the beta-binomial assumption is mainly used for mathematical convenience rather than on any biological basis.

7. An alternative method for the analysis of teratological data is Rao and Scott (1992).

8. Aerts and Claeskens (1997) propose a nonparametric beta-binomial model using the local-likelihood method using local polynomial kernel estimators. It is limited to a single x, and if so, `vgam()` with an `s()` term should return a similar result to theirs.

4.1 Extensions

An extension is to suppose the risk of malformation increases with the teratogen dose x, as characterised, for example, by the parameter β_1 in the model $\mathrm{logit}\,\mu = \beta_0 + \beta_1\, x$. See Table 5 for an example. It is well-known that such data would be very likely to exhibit the so-called *litter effect*, whereby offspring from the same litter tend to respond more alike than offspring from different litters.

More generally, the beta-binomial framework can be extended so that a parametric model may be imposed on the cluster specific means μ_i. For example, μ_i might be assumed to depend on cluster-level explanatory variables $\boldsymbol{x}_i$ through $\mathrm{logit}(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}$. Other extensions include the correlated-binomial model, multiplicative-binomial model, and the extended beta-binomial model of Prentice (1986).

Originally, it was assumed that the beta-binomial distribution required each response from the same cluster to have a common probability μ_i. In the regression set-up, this required the covariates to be the

Table 5: *Low-iron teratology data. n_i denotes litter size, y_i^* the number of dead fetuses, x_i the hemoglobin level, and "Gp" the group number (group 1 is the untreated (low-iron) group, group 2 received injections on day 7 or 10 only, group 3 received injections on days 0 and 7, and group 4 received injections weekly). Source: Moore and Tsiatis (1991).*

t

n_i	y_i^*	x_i	Gp	n_i	y_i^*	x_i	Gp	n_i	y_i^*	x_i	Gp	n_i	y_i^*	x_i	Gp
10	1	4.1	1	9	7	3.1	1	12	12	5.0	1	14	0	12.6	3
11	4	3.2	1	14	14	3.6	1	10	1	8.6	2	14	1	9.5	3
12	9	4.7	1	12	7	4.1	1	3	1	11.1	2	11	0	9.8	3
4	4	3.5	1	11	9	4.8	1	13	1	7.2	2	3	0	16.6	4
10	10	3.2	1	13	8	4.7	1	12	0	8.8	2	13	0	14.5	4
11	9	5.9	1	14	5	4.8	1	14	4	9.3	2	9	2	15.4	4
9	9	4.7	1	10	10	6.7	1	9	2	9.3	2	17	2	14.5	4
11	11	4.7	1	12	10	5.2	1	13	2	8.5	2	15	0	14.6	4
10	10	3.5	1	13	8	4.3	1	16	1	9.4	2	2	0	16.5	4
10	7	4.8	1	10	10	3.9	1	11	0	6.9	2	14	1	14.8	4
12	12	4.3	1	14	3	6.3	1	4	0	8.9	2	8	0	13.6	4
10	9	4.1	1	13	13	4.4	1	1	0	11.1	2	6	0	14.5	4
8	8	3.2	1	4	3	5.2	1	12	0	9.0	2	17	0	12.4	4
11	9	6.3	1	8	8	3.9	1	8	0	11.2	3				
6	4	4.3	1	13	5	7.7	1	11	1	11.5	3				

same for all subunits within a cluster, that is, $\boldsymbol{x}_{i1} = \cdots = \boldsymbol{x}_{in}$. However, Rosner (1984) extended the beta-binomial to allow the covariates to vary within clusters.

To allow η_2 to be a function of the covariates it is necessary to set `betabin(zero=NULL)`.

4.2 Example

In the following we reproduce the results in Williams (1975).

```
> rat = read.table("../data/rat.dat", header=T)
> rat[sample(nrow(rat), 4),]
   trt ri ni
26   1  4  5
18   1 11 11
13   0  4  5
28   1  4  7
> treated = vglm(cbind(ni-ri,ri) ~ 1, betabin, rat, subset=trt==1)
> control = vglm(cbind(ni-ri,ri) ~ 1, betabin, rat, subset=trt==0)
> both = vglm(cbind(ni-ri,ri) ~ 1, betabin(ini=.6), rat, tr=T)
VGLM    linear loop 1: loglikelihood=-121.0913
VGLM    linear loop 2: loglikelihood=-120.6213
VGLM    linear loop 3: loglikelihood=-120.5418
VGLM    linear loop 4: loglikelihood=-120.5405
VGLM    linear loop 5: loglikelihood=-120.5405
> both
Call:
vglm(formula = cbind(ni - ri, ri) ~ 1, family = betabin(ini = 0.6), data = rat, tr = T)

Coefficients:
 (Intercept):1 (Intercept):2
      1.504342     0.4329955

Degrees of Freedom: 64 Total; 62 Residual
Log-likelihood: -120.5405

> fitted(both)
 [1] 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211
 [8] 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211
[15] 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211
[22] 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211 0.8182211
[29] 0.8182211 0.8182211 0.8182211 0.8182211
> # Conduct a likelihood ratio test
> Diff = treated$loglik + control$loglik
> 1-pchisq(2*(Diff - both$loglik), df=2)
[1] 0.021092
```

The null hypothesis tested is $H_0 : \mu_1 = \mu_2,\ \theta_1 = \theta_2$ versus $H_1 : \mu_1 \neq \mu_2,\ \theta_1 \neq \theta_2$. There is quite strong evidence against H_0, that is, there seems a difference between the treatment and control. The estimate of ρ for the treated group can be extracted by

```
> eta2theta(treated$linear[1,2], treated$misc$link[2])
```

```
[1] 0.31801
```

As expected, this is positive. Now

```
> summary(treated)

Call:
vglm(formula = cbind(ni - ri, ri) ~ 1, family = betabin, data = rat, subset = trt == 1)

Coefficients:
              Value Std. Error t value
(Intercept):1 1.044      0.373     2.8
(Intercept):2 0.659      0.263     2.5

Number of linear predictors:  2

Names of linear predictors: logit(mu), log((1+rho)/(1-rho))

Dispersion Parameter for betabin family:   1

Log-likelihood: -64.99 on 30 degrees of freedom

Number of Iterations: 5
```

Because $\eta_2 = 0$ if and only if $\rho = 0$, the t-value of 2.5 provides strong evidence that ρ is non-zero. That is, a ordinary binomial model is not sufficient.

A final note: in terms of software, fitting `treated` and `control` simultaneously by

```
vglm(cbind(ni-ri,ri) ~ trt, betabin, rat)
```

results in numerical problems!

Table 6: *Shell Toxicology Laboratory Data.*

t	
Control	1/12, 1/7, 4/6, 0/6, 0/7, 0/8, 0/10, 0/7, 1/8, 0/6, 2/11, 0/7, 5/8, 2/9, 1/2, 2/7, 0/9, 0/7, 1/11, 0/10, 0/4, 0/8, 0/10, 3/12, 2/8, 4/7, 0/8
Low Dose	0/5, 1/11, 1/7, 0/9, 2/12, 0/8, 1/6, 0/7, 1/6, 0/4, 0/6, 3/9, 0/6, 0/7, 1/5, 5/9, 0/1, 0/6, 3/9
Medium Dose	2/4, 3/4, 2/9, 1/8, 2/9, 3/7, 0/8, 4/9, 0/6, 0/4, 4/6, 0/7, 0/3, 6/13, 6/6, 5/8, 4/11, 1/7, 0/6, 3/10, 6/6
High Dose	1/9, 0/10, 1/7, 0/5, 1/4, 0/6, 1/3, 1/8, 2/5, 0/4, 4/4, 1/5, 1/3, 4/8, 2/6, 3/8, 1/6

Table 7: *Beta-binomial model fitted to the Shell data. Standard errors are in brackets.* θ *is* $(\alpha+\beta)^{-1}$ *and the naïve case corresponds to* $\theta = \rho = 0$*.*

t

Group	Number of Litters	Average Litter Size	Average proportion abnormal	$\hat{\mu}$	$\hat{\theta}$	naïve $\mathrm{se}(\hat{\mu})$
Control	27	7.96	0.135	0.140 (.038)	.273 (.155)	.023
Low Dose	19	7.00	0.135	0.127 (.037)	.118 (.101)	.030
Medium Dose	21	7.19	0.344	0.350 (.068)	.461 (.233)	.039
High Dose	17	5.94	0.228	0.239 (.055)	.128 (.120)	.042

5 Some Generalized Distributions

We now briefly mentions some 'generalized' distributions. Not all are currently implemented.

5.1 Generalized Beta 2

The generalized beta distribution of the second kind has density

$$\frac{|a|\, y^{ap-1}}{b^{ap}\, B(p,q)\, [1+(y/b)^a]^{p+q}}$$

where $a, b, p, q, y > 0$. It has cdf

$$\frac{B_{-(y/b)^a}(p, 1-p-q)}{(-1)^p\, B(p,q)}$$

and moments

$$E(Y^k) = \frac{b^k\, B(p+k/a, q-k/a)}{B(p,q)}.$$

One of the very useful features of the GB2 distribution is that it allows for a large variety of shapes that are nearly lognormal.

5.2 Generalized Lambda

Also known as the *asymmetric lambda* distribution, it exists in two forms of parameterizations: the original form by Ramberg and Schmeiser (1974) and a later formulation by FMKL88 and KDM96.

The density is

$$\psi(y) = \left[\frac{\lambda_3}{\lambda_2} y^{\lambda_3 - 1} + \frac{\lambda_4}{\lambda_2}(1-y)^{\lambda_4 - 1}\right]^{-1}.$$

Its popularity is due to its tremendous flexibility. It has moments

$$E(Y^k) = \int_0^1 \left(\psi^{-1}(u)\right)^k du$$

where

$$\Psi^{-1}(u) = \lambda_1 + \frac{1}{\lambda_2}\left(u^{\lambda_3} - (1-u)^{\lambda_4}\right), \;\; 0 \le u \le 1.$$

The density corresponds to $y = \psi^{-1}(u)$. The first two moment are

$$\begin{aligned} E(Y) &= \frac{1}{\lambda_2}\left(\lambda_1\lambda_2 + \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4}\right), \\ E(Y^2) &= \frac{1}{\lambda_2^2}\left(\lambda_1^2\lambda_2^2 + \frac{1}{1+2\lambda_3} + \frac{1}{1+2\lambda_4} + \frac{2\lambda_1\lambda_2}{1+\lambda_3} - \frac{2\lambda_1\lambda_2}{1+\lambda_4} - 2B(1+\lambda_3, 1+\lambda_4)\right). \end{aligned}$$

5.3 Generalized Poisson

One generalization of the Poisson distribution is the *generalized Poisson* distribution; see the Exercises. It was introduced by Consul and Jain (1973) and studied extensively by Consul (1989). It can be shown

$$\sum_{y=0}^{\infty} \frac{(\theta + y\lambda)^y}{y!} \exp(-y\lambda - \theta) \;=\; \frac{1}{1-\lambda}, \qquad -\lambda_0 < \lambda < 1,$$

where λ_0 solves $\lambda \exp(\lambda) = \exp(-1)$, i.e., $\lambda_0 \approx 0.27846454$. Applications of the GPD can be found in settings where one seeks to describe the distribution of an event that occurs rarely in a short period but where we observe the frequency of its occurrence in longer periods of time. It describes accurately phenomena as diverse as:

- the observed number of industrial accidents and injuries, where a learning effect may be present,
- the spatial distribution of insects, where initial occupation of a spot by a member of the species has an influence on the attractiveness of the spot to other members of the species,
- the number of units of different commodities purchases by consumers, where current sales have an impact on the level of subsequent sales through repeat purchases.

See also Tuenter (2000).

The *generalized Poisson* distribution is

$$P(Y = y) \;=\; \frac{\theta(\theta + y\lambda)^{y-1}}{y!} \exp(-y\lambda - \theta), \qquad y = 0, 1, \ldots, \quad 0 \le \lambda < 1, \quad \theta > 0.$$

The VGAM family function `genpoisson()` implements this model. It provides the `"logit"`, `"probit"`, `"cloglog"` links for λ, and the `"loge"`, and `"identity"` links for θ.

5.4 Generalized Pareto

The generalized Pareto distribution has density

$$\frac{a\, b^a}{y^{a+1}}$$

where $a \geq b$ and $y > 0$.

The (ordinary) Pareto distribution has density

$$\frac{a}{y^{a+1}}.$$

6 Some Other Distributions

6.1 Log-logistic Distribution

Has pdf

$$\frac{\lambda\kappa(\lambda y)^{\kappa-1}}{[1+(\lambda y)^{\kappa}]^2}$$

where $\lambda, \kappa \in \mathbb{R}$ and $y > 0$. Then $\mu = \pi/(\lambda\kappa\sin(\pi/\kappa))$ and $\mathrm{Var}(Y) = 2\pi/(\lambda^2\kappa\sin(2\pi/\kappa)) - \mu^2$.

Related to this distribution is the Goel-Okumo model whose likelihood function is

$$\exp(-m(y_i))\prod_{i=1}^{n}\lambda(y_i)$$

where $m(t) = a(1-\exp(-bt))$ and $\lambda(t) = ab\exp(-bt)$.

6.2 Gumbel Distribution

This has density

$$ab\exp\left\{-\left(be^{-ay}+ay\right)\right\},$$

where $a, b > 0$ and $y \in \mathbb{R}$. Then

$$\begin{aligned} E(Y) &= \frac{1}{a}\left(\log\, b + C_e\right), \\ E(Y^2) &= \frac{1}{a^2}\left(\log^2 b + 2C_e\log\, b + C_e^2 + \pi^2/6\right). \end{aligned}$$

6.3 Zeta Distribution

The Riemann zeta function is defined by $\zeta(s) = \sum_{k=1}^{\infty} k^{-s},\ \Re(s) > 1$ and the (Riemann) zeta distribution is given in Table 8. VGAM supplies a `zeta()` for computing $\zeta(s)$ for real s. Analytic continuation via $\zeta(s) = 2^s\pi^{s-1}\sin(\pi s/2)\Gamma(1-s)\zeta(1-s)$ means it can be defined for all $\Re(s)$, with $\zeta(1) = \infty$..

For the zeta distribution in Table 8, we have $\mu_r' \equiv E(Y^r) = \zeta(p-r+1)/\zeta(p+1)$ for $p > r$, and if $p \leq r$ then the moment is infinite. For further details see, e.g., pp.465–471 of Johnson et al. (1993) and Abramowitz and Stegun (1992).

The zeta distribution is a long-tailed distribution that is useful for size-frequency data. It is sometimes used in insurance as a model for the number of policies held by a single person in an insurance portfolio. It is also used for the analysis of the frequency of words in long sequences of text. When used in linguistics the zeta distribution is known as the *Ziph* distribution.

t

Distribution	Density function $f(y;\boldsymbol{\theta})$	Range of y	Range of $\boldsymbol{\theta}$[a]	Mean	VGAM family function
Zeta	$1/(y^{p+1}\zeta(p+1))$	$1(1)\infty$	$(0,\infty)$	$\zeta(p)/\zeta(p+1)$, provided $p>1$	`Zeta`

Table 8: *More discrete univariate distributions currently supported by* VGAM*.*

[a] $-\infty < \theta_j < \infty$ and θ_j real-valued is assumed unless otherwise stated.

6.4 Haight's Zeta Distribution

Haight's zeta distribution arises as the limiting distribution to Zipf's conjecture concerning city sizes (Simon, 1955). It is given by

$$P(Y = y) \;=\; (2y-1)^{-\alpha} \;-\; (2y+1)^{-\alpha},\; y = 1, 2, \ldots,\;\; \alpha > 0.$$

It gets its name because

$$E(Y) \;=\; (1 - 2^{-\alpha})\zeta(\alpha)$$

and

$$Var(Y) \;=\; (1 - 2^{1-\alpha})\zeta(\alpha - 1) \;-\; \mu^2$$

are functions of the Riemann zeta function. Nb: For $\alpha \leq 1$ the mean is infinite, and for $\alpha \leq 2$ the variance is infinite. For more information, see p.470 of Johnson et al. (1993).

Haight (1966) used Haight's zeta distribution very successfully on four data sets on word associations. Haight (1966) investigation was concerned with models for word association data (the number of response words elicited by a stimulus word). He applied the Yule, Borel-Tanner and logarithmic distribution.

A related distribution is Haight's harmonic distribution,

$$P(Y = y) \;=\; \frac{1}{2Z}\left\{\left[\frac{2Z}{2y-1}\right] \;-\; \left[\frac{2Z}{2y+1}\right]\right\},$$

where $2Z$ is a positive integer and $y = 1, 2, \ldots, [Z + \frac{1}{2}]$. VGAM cannot fit this model because all unknown parameters must be continuous.

6.5 Log-gamma distribution

One parameterization is

$$f(y) = \frac{\exp\{ky - e^y\}}{\Gamma(k)}$$

for $k > 0$ and $-\infty < y < \infty$.

Reference: (Kotz and Nadarajah, 2000, p.48).

Exercises

1. Go through the derivation of the beta-binomial model. In particular, show that the marginal distribution of Y_i^* is

$$P(Y_i^* = y_i^*) \;=\; \binom{n_i}{y_i^*} B(y_i^* + \alpha_i, n_i - y_i^* + \beta_i)/B(\alpha_i, \beta_i).$$

2. Table 6 gives some experimental data from Paul (1982) and Table 7 gives the results of a beta-binomial model fitted to the data. Obtain these results using `betabin()`. Hint: you'll need to use the delta method, and the results will differ slightly. Comment on the resulting standard errors compared to the naïve standard errors. Test for a treatment effect—both using a likelihood ratio test and a Wald test.

3. Write a VGAM family function for the truncated beta-binomial distribution, defined as a beta-binomial distribution with zeros excluded.

4. Can you obtain the expected Hessian for the generalized Poisson distribution? If so, implement that in `genpoisson`.

5. The *inverse binomial* distribution of Yanagimoto (1989) is defined by the probability function for $y \in \{0, 1, 2, \ldots\}$,

$$P(Y = y) \;=\; \frac{\lambda\,\Gamma(2y + \lambda)}{\Gamma(y+1)\,\Gamma(y + \lambda + 1)}\;\{\rho(1-\rho)\}^y\,\rho^\lambda,$$

   where $\lambda > 0$ and $\frac{1}{2} < \rho < 1$. It is a special case of the generalized negative binomial distribution of Jain and Consul (1971). Write a VGAM family function called `invbinomial()` to implement this model. Provide the `"logit"`, `"probit"`, `"cloglog"` links for ρ (this is not perfect!), and the `"loge"`, and `"identity"` links for λ.

6. Van Dorp and Kotz (2002) describe the Standard Two-Sided Power Distribution STSP(θ, n) with density function

$$f(y; \theta, n) = \begin{cases} n(y/\theta)^{n-1}, & 0 < y \le \theta; \\ n((1-y)/(1-\theta))^{n-1}, & \theta \le y < 1, \end{cases}$$

   where $0 \le \theta \le 1$ and $n > 0$ is not necessarily an integer. Show that $E(Y) = ((n-1)\theta + 1)/(n+1)$, $\mathrm{Var}(Y) = (n - 2(n-1)\theta(1-\theta))/((n+2)(n+1)^2)$. Write a VGAM family function to implement this distribution. Allow for the the logit, probit, cloglog and identity links to θ, and the log and identity links to n. Run it on the following data: 0.34, 0.395, 0.413, 0.42, 0.423, 0.429, 0.465, 0.513, 0.564, 0.588. Show that $\hat{\theta} = 0.423$ and $\hat{n} = 8.399$.

7. Modify your answer to the previous exercise to allow for parameters a and b, where a and b are fixed and given. Then $W \sim$ TSP(a, m, b, n) where $W = (b-a)Y + a$.

References

Abramowitz, M., Stegun, I. A. (Eds.), 1992. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover Publications Inc., New York, reprint of the 1972 edition.

Aerts, M., Claeskens, G., 1997. Local polynomial estimation in multiparameter likelihood models. Journal of the American Statistical Association 92, 1536–1545.

Bliss, C., Fisher, R. A., 1953. Fitting the negative binomial distribution to biological data. Biometrics 9, 174–200.

Cameron, A. C., Trivedi, P. K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge.

Chambers, J. M., Hastie, T. J. (Eds.), 1993. Statistical Models in S. Chapman & Hall, New York.

Consul, P. C., 1989. Generalized Poisson Distributions: Properties and Applications. Marcel Dekker.

Consul, P. C., Jain, G. C., 1973. A generalization of the Poisson distribution. Technometrics 15, 791–799.

Haight, F. A., 1966. Some statistical problems in connection with word association data. Journal of Mathematical Psychology 3, 217–233.

Jain, G. C., Consul, P. C., 1971. A generalized negative binomial distribution. SIAM Journal on Applied Mathematics 21, 501–513.

Johnson, N. L., Kotz, S., Kemp, A. W., 1993. Univariate Discrete Distributions, 2nd Edition. Wiley, New York.

Jørgensen, B., 1997. The Theory of Dispersion Models. Chapman & Hall, London.

Kotz, S., Nadarajah, S., 2000. Extreme Value Distributions: Theory and Applications. Imperial College Press, London.

Lawless, J. F., 1987. Negative binomial and mixed Poisson regression. The Canadian Journal of Statistics 15, 209–225.

McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London.

Moore, D. F., Tsiatis, A., 1991. Robust estimation of the variance in moment methods for extra-binomial and extra-poisson variation. Biometrics 47, 383–401.

Paul, S. R., 1982. Analysis of proportions of affected foetuses in tetralogical experiments. Biometrics 38, 361–370.

Prentice, R. L., 1986. Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. Journal of the American Statistical Association 81, 321–327.

Rao, J. N. K., Scott, A. J., 1992. A simple method for the analysis of clustered binary data. Biometrics 48, 577–585.

Rosner, B., 1984. Multivariate methods for clustered binary data with more than one level of nesting. Biometrics 84, 373–380.

Simon, H. A., 1955. On a class of skew distribution functions. Biometrika 42, 425–440.

Thurston, S. W., Wand, M. P., Wiencke, J. K., 2000. Negative binomial additive models. Biometrics 56 (1), 139–144.

Tuenter, H. J. H., 2000. On the generalized Poisson distribution. Statistica Neerlandica 54, 374–376.

Van Dorp, J. R., Kotz, S., 2002. The standard two-sided power distribution and its properties: with applications in financial engineering. The American Statistician 56, 90–99.

Williams, D. A., 1975. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31, 949–952.

Yanagimoto, T., 1989. The inverse binomial distribution as a statistical model. Communications in Statistics: Theory and Methods 18, 3625–3633.

Yee, T. W., 2004. A new technique for maximum likelihood canonical Gaussian ordination. Ecology (submitted) .