Institut für Statistik
Gerhard Tutz, Moritz Berger Wintersemester 2014/15

Formelsammlung zur Vorlesung Kategoriale Daten

29.01.2015

Diese Formelsammlung darf in der Klausur verwendet werden. Eigene handschriftliche Notizen und Ergänzungen dürfen, ausschlielich auf den bedruckten Vorderseiten, eingefügt werden. Es drfen keine zustzlichen Bltter eingeheftet werden.

Contents

1 Distributions

1.1 Binomial distribution

The variable y is binomially distributed, $y \sim B(n,\pi)$, if the probability mass function is given by

$$f(y) = \begin{cases} \binom{n}{y}\pi^y(1-\pi)^{n-y} & y \in \{0,1,\dots,n\} \\ 0 & \text{otherwise}. \end{cases}$$

The parameters are $n \in \mathbb{N}, \pi \in [0,1]$. One obtains

$$E(y) = n\pi,\ \mathit{var}(y) = n\pi(1-\pi).$$

1.2 Multinomial distribution

The vector $\boldsymbol{y}^T = (y_1,\dots,y_k)$ is multinomially distributed, $\boldsymbol{y} \sim M(n,(\pi_1,\dots,\pi_k))$, if the probability mass function is given by

$$f(y_1,\dots,y_k) = \begin{cases} \frac{n!}{y_1!\cdot\ldots\cdot y_k!} \ \pi_1^{y_1}\cdot\ldots\cdot\pi_k^{y_k} & y_i \in \{0,\dots,n\},\ \sum_i y_i = n \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{\pi}^T = (\pi_1,\dots,\pi_k)$ is a probability vector, i.e. $\pi_i \in [0,1],\ \sum_i \pi_i = 1$.
One has

$$E(y_i) = n\pi_i, \quad \mathit{var}(y_i) = n\pi_i(1-\pi_i),$$
$$\mathit{cov}(y_i,y_j) = -n\pi_i\pi_j, i \neq j.$$

1.3 Poisson distribution

A variable y follows a Poisson distribution, $y \sim Po(\lambda)$, if the probability mass function is given by

$$f(y) = \begin{cases} \frac{\lambda^y}{y!}e^{-\lambda} & y = 0,1,2,\dots \\ 0 & \text{otherwise}. \end{cases}$$

The parameter $\lambda > 0$ determines

$$E(y) = \lambda,\ \mathit{var}(y) = \lambda.$$

1.4 Poisson gamma distribution

The variable $y \in \{0, 1, \ldots\}$ follows a Poisson gamma distribution, $PoGa(\mu, a, b)$, (also known as negative binomial distribution), if the mass function is given by

$$f(y) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \left(\frac{b}{b+\mu}\right)^a \left(\frac{\mu}{b+\mu}\right)^y,\ y \in \{0, 1, \ldots\}$$

where $\mu > 0, a > 0, b > 0$. One has

$$E(y) = \mu\frac{a}{b},\ \ \textit{var}(y) = \mu\frac{a}{b}\left(1 + \frac{\mu}{b}\right).$$

For $a \in \mathbb{N}$ the distribution is exclusively called negative binomial distribution, $NB(a, \pi)$, with mass function

$$f(y) = \binom{y+a-1}{y}\pi^a(1-\pi)^y,\ y \in \{0, 1, \ldots\}$$

and $\pi = b/(b+\mu)$. Here a denotes the number of successes and y the number of failures.

1.5 Gamma distribution

A random variable y is gamma distributed, $y \sim \Gamma(\nu, \alpha)$, if it has density function

$$f(y) = \begin{cases} 0 & y \le 0 \\ \frac{\alpha^\nu}{\Gamma(\nu)} & y^{\nu-1}e^{-\alpha y} \quad y > 0 \end{cases}$$

where for $\nu > 0$, $\Gamma(\nu)$ is defined by

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1}e^{-x}dx.$$

One has

$$E(y) = \frac{\nu}{\alpha},\ \ \textit{var}(y) = \frac{\nu}{\alpha^2}.$$

For $\nu = 1$ the exponential distribution is a special case.

2 Loglinear Models

2.1 Two-way tables

In general, two-way $(I \times J)$-contingency tables with I rows and J columns may be described by

$X_{ij} =$ counts in cell (i, j),
$X_A \in \{1, \ldots, I\}$ representing the rows,
$X_B \in \{1, \ldots, J\}$ representing the columns.

The observed contingency table has the form

$$\begin{array}{cc|cccc|c} & & & X_B & & & \\ & & 1 & 2 & \ldots & J & \\ \hline & 1 & X_{11} & X_{12} & \ldots & X_{1J} & X_{1+} \\ & 2 & X_{21} & \ddots & & \vdots & \vdots \\ X_A & \vdots & \vdots & & \ddots & \vdots & \\ & I & X_{I1} & \ldots & & X_{IJ} & X_{I+} \\ \hline & & X_{+1} & \ldots & & X_{+J} & \end{array}$$

where $X_{i+} = \sum_{j=1}^J X_{ij}, X_{+j} = \sum_{i=1} X_{ij}$ denote the marginal counts. The subscript "$+$" denotes the sum over that index.

Poisson and Multinomial settings

Let X_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ follow independent Poisson distributions, $X_{ij} \sim Po(\lambda_{ij})$. Then the *conditional distribution* of $(X_{11}, \ldots, X_{IJ})$ given $n = \sum_{i,j} X_{ij}$ is multinomial. More concrete, one has

$$\left((X_{11}, \ldots, X_{IJ})|\sum_{i,j} X_{ij} = n\right) \sim M(n, (\lambda_{11}/\lambda, \ldots, \lambda_{IJ}/\lambda)),$$

with $\lambda = \sum_{i,j} \lambda_{ij}$.

Multinomial and Product-Multinomial settings

Let $(X_{11}, \ldots, X_{IJ})$ have multinomial distribution, $(X_{11}, \ldots, X_{IJ}) \sim M(n, (\pi_{11}, \ldots, \pi_{IJ}))$. By conditioning on the row margins $n_{i+} = \sum_j X_{ij}$ one obtains the product-multinomial distribution with probability mass function

$$f(x_{11}, \ldots, x_{IJ}) = \prod_{i=1}^{I} \frac{n_{i+}!}{x_{i1}! \cdot \ldots \cdot x_{iJ}!} \pi_{1|i}^{x_{i1}} \cdot \ldots \cdot \pi_{J|i}^{x_{iJ}}$$

where $\pi_{j|i} = \pi_{ij}/\sum_j \pi_{ij} = \pi_{ij}/\pi_{i+}$. Thus the cell counts of one row given n_{i+} have multinomial distribution,

$$(X_{i1}, \ldots, X_{iJ}) \sim M(n_{i+}, (\pi_{1|i}, \ldots, \pi_{J|i}))$$

Loglinear Model for Two-Way Tables

$$\log(\mu_{ij}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{AB(i,j)}$$

Constraints:

$$\sum_{i=1}^{I} \lambda_{A(i)} = \sum_{j=1}^{J} \lambda_{B(j)} = \sum_{i=1}^{I} \lambda_{AB(ij)} = \sum_{j=1}^{J} \lambda_{AB(ij)} = 0,$$

or

$$\lambda_{A(I)} = \lambda_{B(J)} = \lambda_{AB(iJ)} = \lambda_{AB(Ij)} = 0.$$

Additional constraints for multinomial tables:

$$\sum_{i,j} e^{\lambda_0} e^{\lambda_{A(i)}} e^{\lambda_{B(j)}} e^{\lambda_{AB(ij)}} = n$$

Additional constraints for product-multinomial tables:

$$\sum_{j=1}^{J} e^{\lambda_0} e^{\lambda_{A(i)}} e^{\lambda_{B(j)}} e^{\lambda_{AB(ij)}} = n_{i+}, \quad i = 1, \ldots, I \quad (\text{for } n_{i+} \text{ fixed}),$$

$$\sum_{i=1}^{I} e^{\lambda_0} e^{\lambda_{A(i)}} e^{\lambda_{B(j)}} e^{\lambda_{AB(ij)}} = n_{+j}, \quad j = 1, \ldots, J \quad (\text{for } n_{+j} \text{ fixed}).$$

2.2 Three-way tables

Three-Way Tables

X_A	X_B	X_C: 1	2	$\cdots$	K	
1	1	X_{111}	X_{112}	$\cdots$	X_{11K}	X_{11+}
	2	X_{121}	X_{122}			$\vdots$
	$\vdots$	$\vdots$				
	J	X_{1J1}	$\cdots$		X_{1JK}	X_{1J+}
2	1	X_{211}	X_{212}	$\cdots$	X_{21K}	X_{21+}
	2	X_{221}	X_{222}			$\vdots$
	$\vdots$	$\vdots$				
	J	X_{2J1}	$\cdots$		X_{2JK}	X_{2J+}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
I	1	X_{I11}	X_{I12}	$\cdots$	X_{I1K}	X_{I1+}
	2	X_{I21}	X_{I22}			$\vdots$
	$\vdots$	$\vdots$				
	J	X_{IJ1}	$\cdots$		X_{IJK}	X_{IJ+}

Product-Multinomial settings

$n_{i++} = X_{i++}$ is fixed and

$$(X_{i11}, \ldots, X_{iJK}) \sim M(n_{i++}, (\pi_{i11}, \ldots, \pi_{iJK})), \qquad (1)$$

where $\pi_{ijk} = P(X_B = j, X_C = k | X_A = i)$. An example of the second variant (two design variables) is obtained by letting X_A and X_B be design variables, i. e. $n_{ij+} = X_{ij+}$ is fixed and

$$(X_{ij1}, \ldots, X_{ijK}) \sim M(n_{ij+}, (\pi_{ij1}, \ldots, \pi_{ijK})). \qquad (2)$$

where $\pi_{ijk} = P(X_C = k | X_A = i, X_B = j)$.

Loglinear Model for Three-Way Tables

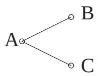
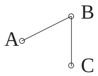
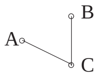
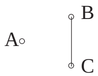
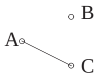
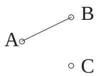
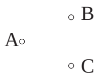
$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AB(i,j)} + \lambda_{AC(ik)} + \lambda_{BC(jk)} + \lambda_{ABC(ijk)}$$

Constraints:

$$\sum_i \lambda_{A(i)} = \sum_j \lambda_{B(j)} = \sum_k \lambda_{C(k)} = 0,$$
$$\sum_i \lambda_{AB(ij)} = \sum_j \lambda_{AB(ij)} = \sum_i \lambda_{AC(ik)} = \sum_k \lambda_{AC(ik)} = \sum_j \lambda_{BC(jk)} = \sum_k \lambda_{BC(jk)} = 0,$$
$$\sum_i \lambda_{ABC(ijk)} = \sum_j \lambda_{ABC(ijk)} = \sum_k \lambda_{ABC(ijk)} = 0,$$

or

$$\lambda_{A(I)} = \lambda_{B(J)} = \lambda_{C(K)} = 0,$$
$$\lambda_{AB(i,J)} = \lambda_{AB(I,j)} = \lambda_{AC(iK)} = \lambda_{AC(Ik)} = \lambda_{BC(jK)} = \lambda_{BC(Jk)} = 0,$$
$$\lambda_{ABC(Ijk)} = \lambda_{ABC(iJk)} = \lambda_{ABC(ijK)} = 0.$$

	Loglinear Model	Regressors of Logit-Model
AB/AC X_B, X_C conditionally independent, given X_A		$1, x_A$
AB/BC X_A, X_C conditionally independent, given X_B		$1, x_B$
AC/BC X_A, X_B conditionally independent, given X_C		$1, x_A, x_B$
A/BC X_A independent of (X_B, X_C)		$1, x_B$
AC/B (X_A, X_C) independent of X_B		$1, x_A$
AB/C (X_A, X_B) independent of X_C		1
$A/B/C$ X_A, X_B, X_C are dependent		1

AB/AC	$\lambda_{ABC} = \lambda_{BC} = 0$	$P(X_B, X_C \mid X_A) = P(X_B \mid X_A)P(X_C \mid X_A)$ X_B, X_C conditionally independent given X_A
AB/BC	$\lambda_{ABC} = \lambda_{AC} = 0$	$P(X_A, X_C \mid X_B) = P(X_A \mid X_B)P(X_C \mid X_B)$ X_A, X_C conditionally independent given X_B
AC/BC	$\lambda_{ABC} = \lambda_{AB} = 0$	$P(X_A, X_B \mid X_C) = P(X_A \mid X_C)P(X_B \mid X_C)$ X_A, X_B conditionally independent given X_C
A/BC	$\lambda_{ABC} = \lambda_{AB} = \lambda_{AC} = 0$	$P(X_A, X_B, X_C) = P(X_A)P(X_B, X_C)$ X_A jointly independent of (X_B, X_C)
AC/B	$\lambda_{ABC} = \lambda_{AB} = \lambda_{BC} = 0$	$P(X_A, X_B, X_C) = P(X_A, X_C)P(X_B)$ (X_A, X_C) jointly independent of X_B
AB/C	$\lambda_{ABC} = \lambda_{AC} = \lambda_{BC} = 0$	$P(X_A, X_B, X_C) = P(X_A, X_B)P(X_C)$ (X_A, X_B) jointly independent of X_C
$A/B/C$	$\lambda_{ABC} = \lambda_{AB} = \lambda_{AC} =$ $\lambda_{BC} = 0$	$P(X_A, X_B, X_C) = P(X_A)P(X_B)P(X_C)$ X_A, X_B, X_C are independent

Type 0: Saturated Model

$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AB(i,j)} + \lambda_{AC(ik)} + \lambda_{BC(jk)} + \lambda_{ABC(ijk)}.$$

Type 1: No three-factor interaction

$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AB(ij)} + \lambda_{AC(ik)} + \lambda_{BC(jk)}.$$

Type 2: Only two two-factor interactions contained

$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AC(ik)} + \lambda_{BC(jk)}.$$

Type 3: Only one two-factor interaction contained

$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{BC(jk)}.$$

Type 4: Main effects model

$$\log(\mu_{ijk}) = \lambda_0 + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)}.$$

2.3 Inference for Loglinear Models

Likelihood and log-likelihood

Let all the Parameters be collected in one parameter vector γ. From the likelihood function

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \frac{\mu_{ijk}^{x_{ijk}}}{x_{ijk}!} e^{-\mu_{ijk}}$$

one obtains the log-likelihood

$$l(\boldsymbol{\gamma}) = \sum_{i,j,k} x_{ijk} \log(\mu_{ijk}) - \sum_{i,j,k} \mu_{ijk} - \sum_{i,j,k} \log(x_{ijk}!).$$

With μ_{ijk} parameterized as the saturated loglinear model one obtains by rearranging terms (and omitting constants)

$$\begin{aligned} l(\gamma) &= n\lambda_0 + \sum_i x_{i++}\lambda_{A(i)} + \sum_j x_{+j+}\lambda_{B(j)} + \sum_k x_{++k}\lambda_{C(k)} \\ &+ \sum_{i,j} x_{ij+}\lambda_{AB(ij)} + \sum_{i,k} x_{i+k}\lambda_{AC(ik)} + \sum_{j+k} x_{+jk}\lambda_{BC(jk)} \\ &+ \sum_{i,j,k} x_{ijk}\lambda_{AB(ijk)} - \sum_{i,j,k} \exp(\lambda + \lambda_{A(i)} + \ldots + \lambda_{ABC(ijk)}). \end{aligned}$$

Testing and Goodness of fit

For models with an intercept the deviance has the form

$$D = 2\sum_i x_i \log\left(\frac{x_i}{\hat{\mu}_i}\right).$$

When considering goodness-of-fit an alternative is Pearson's χ^2

$$\chi^2_P = \sum_i \frac{(x_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

For fixed N, both statistics have approximate χ^2-distribution if the assumed model holds and means μ_i are large. The degrees of freedom are $N - p$ where p is the number of estimated parameters.

3 Regression Models

3.1 Binary Response

$$\pi(\boldsymbol{x}_i) = P(y_i = 1|\boldsymbol{x}_i) = h(\boldsymbol{x}_i^T\boldsymbol{\beta})$$

Likelihood and Fisher Matrix for grouped data (with N groups)

$$l(\boldsymbol{\beta}) = \sum_{i=1}^N n_i \left\{\bar{y}_i \log\left(\frac{\pi(\boldsymbol{x}_i)}{1-\pi(\boldsymbol{x}_i)}\right) + \log(1-\pi(\boldsymbol{x}_i))\right\}$$

$$F(\boldsymbol{\beta}) = E(-\partial l^2(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T) = \sum_{i=1}^N n_i \frac{\partial h(\boldsymbol{x}_i^T\boldsymbol{\beta})/\partial\eta}{h(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-h(\boldsymbol{x}_i^T\boldsymbol{\beta}))}\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

For the logit model which has canonical link one obtains the simpler Fisher matrix

$$F(\boldsymbol{\beta}) = \sum_{i=1}^N n_i h(\boldsymbol{x}_i^T\boldsymbol{\beta})(1-h(\boldsymbol{x}_i^T\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

Alternative links:

Probit model

$$\pi(\boldsymbol{x}) = \phi(\boldsymbol{x}^T\boldsymbol{\beta});\ \ \phi^{-1}(\pi(\boldsymbol{x})) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Complementary log-log model

$$\pi(\boldsymbol{x}) = 1 - \exp(-\exp(\boldsymbol{x}^T\boldsymbol{\beta}));\ \ \log(-\log(1-\pi(\boldsymbol{x}))) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Log-log model

$$\pi(\boldsymbol{x}) = \exp(-\exp(-\boldsymbol{x}^T\boldsymbol{\beta}));\ \ -\log(-\log(\pi(\boldsymbol{x}))) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Exponential distribution or Complementary log model

$$\pi(\boldsymbol{x}) = 1 - \exp(-\boldsymbol{x}^T\boldsymbol{\beta});\ \ -\log(1-\pi(\boldsymbol{x})) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Exponential or log-link model

$$\pi(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\boldsymbol{\beta});\ \ \log(\pi(\boldsymbol{x})) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Cauchy model

$$\pi(\boldsymbol{x}) = \tan^{-1}(\boldsymbol{x}^T\boldsymbol{\beta})/\pi + 1/2;\ \ \tan(\pi(\pi(\boldsymbol{x}) - 1/2)) = \boldsymbol{x}^T\boldsymbol{\beta}.$$

Explanatory value

Likelihood-Ratio-Index

$$R^2_{MF} = 1 - \frac{\log(L(\hat{\boldsymbol{\beta}}))}{\log(L(\hat{\beta}_0))}$$

Cox & Snell (1989)

$$R^2_{LR} = 1 - \left(\frac{L(\hat{\beta}_0)}{L(\hat{\boldsymbol{\beta}})}\right)^{2/n}$$

Efrons measure

$$R^2_E = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Mittlere Vorhersagedifferenz

$$\lambda = \frac{1}{n_1}\sum_{y_i=1}\hat{\pi}_i - \frac{1}{n_2}\sum_{y_i=0}\hat{\pi}_i$$

Kendalls

$$\tau_a = \frac{N_c - N_d}{n(n-1)/2}$$

Overdispersion: Beta-binomial Model

$$P(y_i; n_i, a_i, b_i) = \int P(y_i | D_i = \vartheta_i) p(\vartheta_i) d\vartheta_i$$
$$= \int \binom{n_i}{y_i} \vartheta_i^{y_i} (1-\vartheta_i)^{n_i - y_i} \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \vartheta_i^{a_i - 1} (1-\vartheta_i)^{b_i - 1} d\vartheta_i$$
$$= \binom{n_i}{y_i} \frac{(a_i + y_i - 1)_{y_i} (b_i + n_i - y_i - 1)_{n_i - y_i}}{(a_i + b_i + n_i - 1)_{n_i}}.$$

Overdispersion: Quasi-likelihood

$$E(y_i) = n_i \pi_i = n_i h(\mathbf{x}_i^T \beta),$$
$$var(y_i) = n_i \pi_i (1 - \pi_i) \phi.$$

3.2 Multinomial Response

Generic Multinomial Logit Model

$$P(Y = r | \boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{\beta}_r)}{\sum_{s=1}^k \exp(\boldsymbol{x}^T \boldsymbol{\beta}_s)}$$

with optional side constraints

$$\boldsymbol{\beta}_k = (0, \ldots, 0)^T \quad \text{reference category } k$$
$$\boldsymbol{\beta}_{r_0} = (0, \ldots, 0)^T \quad \text{reference category } r_0$$
$$\sum_{s=1}^k \boldsymbol{\beta}_s = (0, \ldots, 0)^T \quad \text{symmetric side constraint}$$

Multinomial logit model mit reference category *k*

$$\log \left(\frac{P(Y = r|\boldsymbol{x})}{P(Y = k|\boldsymbol{x})} \right) = \boldsymbol{x}^T \boldsymbol{\beta}_r, \quad r = 1, \ldots, k-1,$$

or

$$P(Y = r|\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{k-1} \exp(\boldsymbol{x}^T \boldsymbol{\beta}_s)}, \quad r = 1, \ldots, k-1,$$
$$P(Y = k|\boldsymbol{x}) = \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\boldsymbol{x}^T \boldsymbol{\beta}_s)}.$$

Multinomial logit model with category specific covariates (reference category *k*)

$$P(Y = r|\boldsymbol{x}, \{\boldsymbol{w}_j\}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{\beta}_r + \boldsymbol{v}_r^T \boldsymbol{\alpha})}{1 + \sum_{s=1}^{k-1} \exp(\boldsymbol{x}^T \boldsymbol{\beta}_s + \boldsymbol{v}_s^T \boldsymbol{\alpha})}, \quad r = 1, \ldots, k-1$$
$$P(Y = k|\boldsymbol{x}, \{\boldsymbol{w}_j\}) = \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\boldsymbol{x}^T \boldsymbol{\beta}_s + \boldsymbol{v}_s^T \boldsymbol{\alpha})}$$

or

$$\log \left(\frac{P(Y = r|\boldsymbol{x}, \{\boldsymbol{w}_i\})}{P(Y = k|\boldsymbol{x}, \{\boldsymbol{w}_i\})} \right) = \boldsymbol{x}^T \boldsymbol{\beta}_r + \boldsymbol{v}_r^T \boldsymbol{\alpha}, \quad r = 1, \ldots, k-1$$

3.3 Multinomial Response: Ordered Categories

Cumulative Type Model, dichotomization into groups

$$[1, \ldots, r | r+1, \ldots, k] \qquad y_r = \begin{cases} 1 & Y \in \{1, \ldots, r\} \\ 0 & Y \in \{r+1, \ldots, k\} \end{cases}$$

Sequential Type Model, dichotomization given $Y \geq r$

$$1, \ldots, [r | r+1, \ldots, k] \qquad y_r = \begin{cases} 1 & Y = r \\ 0 & Y > r \end{cases} \quad \text{given } Y \geq r$$

Adjacent Type Model, dichotomization given $Y \in \{r, r+1\}$

$$1, \ldots, [r | r+1], \ldots, k] \qquad y_r = \begin{cases} 1 & Y = r \\ 0 & Y \geq r+1 \end{cases} \quad \text{given } Y \in \{r, r+1\}$$

Threshold (simple Cumulative) Model

$$P(Y \leq r|\boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T \boldsymbol{\gamma}),$$

or

$$P(Y = r|\boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T \boldsymbol{\gamma}) - F(\gamma_{0,r-1} + \boldsymbol{x}^T \boldsymbol{\gamma}),$$

where $-\infty = \gamma_{00} \leq \gamma_{01} \leq \cdots \leq \gamma_{0k} = \infty$.

Sequential Model

$$P(Y = r | Y \geq r, \boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T \boldsymbol{\gamma}), \quad r = 1, \ldots, k,$$

or

$$P(Y = r|\boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T \boldsymbol{\gamma}) \prod_{i=1}^{r-1} (1 - F(\gamma_{0i} + \boldsymbol{x}^T \boldsymbol{\gamma})).$$

Cumulative Logit Model (Proportional Odds Model)

$$P(Y \leq r|\boldsymbol{x}) = \frac{\exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})}{1 + \exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})},$$

or

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}.$$

Cumulative Probit Model

$$P(Y \leq r|\boldsymbol{x}) = \Phi(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}).$$

Cumulative Maximum Extreme Value (Gumbel) Model

$$P(Y \leq r|\boldsymbol{x}) = \exp(-\exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})),$$

or

$$\log(-\log(P(Y \leq r|\boldsymbol{x})) = \gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}.$$

Cumulative Minimum Extreme Value Model (Proportional Hazards Model, Gompertz Model)

$$P(Y \leq r|\boldsymbol{x}) = 1 - \exp(-\exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})),$$

or

$$\log(-\log(P(Y > r|\boldsymbol{x})) = \gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma},$$

or

$$P(Y = r|Y \geq r, \boldsymbol{x}) = 1 - \exp(-\exp(\tilde{\gamma}_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})),$$

with $\gamma_{0r} = \log(\sum_{i=1}^{r}\exp(\tilde{\gamma}_{0i}))$, $r = 1, \ldots, k$.

Sequential Logit Model (Continuation Ratio Logits Model)

$$P(Y = r|Y \geq r, \boldsymbol{x}) = \frac{\exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}_r)}{1 + \exp(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma})},$$

or

$$\log\left(\frac{P(Y = r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}.$$

Generalized Cumulative Model

$$P(Y \leq r|\boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}_r), \quad r = 1, \ldots, k.$$

Generalized Sequential Model

$$P(Y = r|Y \geq r, \boldsymbol{x}) = F(\gamma_{0r} + \boldsymbol{x}^T\boldsymbol{\gamma}_r), \quad r = 1 \ldots, k.$$

3.4 Common structure

Basic structure of multicategorical models

$$g(\boldsymbol{\pi}_i) = \mathbf{Z}_i\boldsymbol{\beta}, \quad \boldsymbol{\pi}_i = h(\mathbf{Z}_i\boldsymbol{\beta}),$$

with

$$\begin{aligned} \boldsymbol{g} &= (g_1, \ldots, g_q)^T : \mathbb{R}^q \longrightarrow \mathbb{R}^q, \quad h = g^{-1}, \\ \boldsymbol{\pi}_i &= (\pi_{i1}, \ldots, \pi_{iq})^T, \quad \pi_{ir} = P(Y = r|\mathbf{x}_i). \end{aligned}$$

3.5 Inference

Log Likelihood

$$n_i\mathbf{p}_i \sim M(n_i, \boldsymbol{\pi}_i), i = 1, \ldots, g$$

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{g}\left\{\sum_{r=1}^{q} n_i p_{ir}\log(\pi_{ir}/(1 - \pi_{i1} - \cdots - \pi_{iq})) + \log(1 - \pi_{i1} - \cdots - \pi_{iq})\right\} + \log(c(p_{i1} \ldots, p_{iq}))$$

Score function

$$s(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \sum_{i=1}^{g}\mathbf{Z}_i^T\mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})),$$

where $\mathbf{D}_i(\boldsymbol{\beta}) = \partial h(\mathbf{Z}_i\boldsymbol{\beta})/\partial\boldsymbol{\eta}, \boldsymbol{\Sigma}_i(\boldsymbol{\beta}) = cov(\mathbf{p}_i) = \frac{1}{n_i}\left\{\mathrm{diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i\boldsymbol{\pi}_i^T\right\}$

Fisher matrix

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{g}\mathbf{Z}_i^T\mathbf{W}_i(\boldsymbol{\beta})\mathbf{Z}_i,$$

where $\mathbf{W}_i(\beta) = \mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i(\boldsymbol{\beta})^{-1}\mathbf{D}_i(\boldsymbol{\beta})$

Goodness-of-fit Tests

Pearson Statistic

$$\chi^2_P = \sum_{i=1}^{g} \chi^2_P(\boldsymbol{p}_i, \hat{\boldsymbol{\pi}}_i),$$

$$\text{with} \quad \chi^2_P(\boldsymbol{p}_i, \hat{\boldsymbol{\pi}}_i) = n_i \sum_{r=1}^{k} (p_{ir} - \hat{\pi}_{ir})^2 / \hat{\pi}_{ir}.$$

Deviance

$$\chi^2_D = \sum_{i=1}^{g} \chi^2_D(\boldsymbol{p}_i, \hat{\boldsymbol{\pi}}_i),$$

$$\text{with} \quad \chi^2_D(\boldsymbol{p}_i, \hat{\boldsymbol{\pi}}_i) = 2n_i \sum_{r=1}^{k} \pi_{ir} \log\left(\frac{p_{ir}}{\hat{\pi}_{ir}}\right).$$

4 Regression Analysis of count data

4.1 Basic Model

$$P(Y = y) = \begin{cases} \frac{\lambda^y}{y!} e^{-\lambda} & \text{for } y \in \{0, 1, 2, \dots\} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mu_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}), \quad \log(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$$

Log-likelihood $l(\boldsymbol{\beta})$, score function $s(\boldsymbol{\beta})$ and Fisher matrix $F(\boldsymbol{\beta})$:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \boldsymbol{x}_i^T \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) + \log(y_i!),$$

$$s(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i (y_i - \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})),$$

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}).$$

Goodness-of-fit for Poisson Regression Model

$$D = 2 \sum_{i=1}^{N} y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right),$$

$$\chi^2_P = \sum_{i=1}^{N} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

For $\mu_i \to \infty$ one obtains the approximation

$$D, \chi^2_P \sim \chi^2(N - p).$$

4.2 Modelling of Overdispersion

(1) Quasi-Likelihood Methods

$$\mu_i = h(x_i^T \beta)$$

Estimation equation

$$\sum_{i=1}^{n} \boldsymbol{x}_i \frac{\partial \mu_i}{\partial \eta} \frac{y_i - \mu_i}{v(\mu_i)} = 0$$

(2) Negative binomial Gamma-Poisson Model

$$\begin{aligned} P(y_i) &= \int f(y_i|b_i) f(b_i) \boldsymbol{d}\, b_i \\ &= \int \left(e^{-b_i \mu_i} \frac{(b_i \mu_i)^{y_i}}{y_i!} \right) \left(\frac{\nu^\nu}{\Gamma(\nu)} b_i^{\nu-1} e^{-\nu b_i} \right) \boldsymbol{d}\, b_i \\ &= \frac{\Gamma(y_i + \nu)}{\Gamma(\nu)\Gamma(y_i + 1)} \left(\frac{\mu_i}{\mu_i + \nu} \right)^{y_i} \left(\frac{\nu}{\mu_i + \nu} \right)^{\nu}. \end{aligned}$$

$$E(y_i) = \mu_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}), \quad var(y_i) = \mu_i + \frac{1}{\nu} \mu_i^2.$$

(3) Zero-inflated counts

With C denoting the class indicator of subpopulations ($C_i = 1$ for responders and $C_i = 0$ for non-responders) one obtains the mixture distribution

$$P(y_i = y) = P(y_i = y | C_i = 1)\pi_i + P(y_i = y | C_i = 0)(1 - \pi_i)$$

$$\begin{aligned} E(y_i) &= \pi_i \mu_i, \\ var(y_i) &= \pi_i \mu_i + \pi_i (1 - \pi_i) \mu_i^2 \\ &= \pi_i \mu_i (1 + \mu_i (1 - \pi_i)). \end{aligned}$$

$$\begin{aligned} \log(\mu_i) &= \boldsymbol{x}_i^T \boldsymbol{\beta}, \\ \text{logit}(\pi_i) &= \boldsymbol{z}_i^T \boldsymbol{\gamma}, \end{aligned}$$

(4) Hurdle Models

$$\begin{aligned} P(y = 0) &= f_1(0), \\ P(y = r) &= f_2(r) \frac{1 - f_1(0)}{1 - f_2(0)}, \quad r = 1, 2, \ldots \end{aligned}$$

$$\begin{aligned} E(y) &= \gamma \mu_2, \\ var(y) &= \sum_{r=1}^{n} r^2 f_2(r) \gamma - \left(\sum_{r=1}^{\infty} r\, f_2(r) \gamma \right)^2 \\ &= \mu_2 (1 + \mu_2) \gamma - \mu_2^2 \gamma^2. \end{aligned}$$

5 Repeated Measurements

5.1 Tests for binary response

General form

		y_2 1	0
y_1	1	n_{11}	n_{10}
	0	n_{01}	n_{00}

Hypotheses

$$H_0 : P(y_2 = 1) - P(y_1 = 1) = 0$$
$$H_1 : P(y_2 = 1) - P(y_1 = 1) \neq 0$$

(1) McNemar

$$M = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \overset{H_0}{\underset{a}{\sim}} \chi^2(1)$$

(2) Binomial test

$$n_{01} | n_{01} + n_{10} = t \overset{H_0}{\sim} B(t, 0.5)$$

(3) Likelihood Ratio test

$$\lambda = 2n_{01} \log\left(\frac{2n_{01}}{n_{01} + n_{10}}\right) + 2n_{10} \log\left(\frac{2n_{10}}{n_{01} + n_{10}}\right) \overset{H_0}{\underset{a}{\sim}} \chi^2(1)$$

5.2 Logistic Regression

(a) **Subject-specific parameter**

$$\text{logit}\left(P(y_{it} = 1|x_t)\right) = \beta_i + x_t\beta, \quad x_t = \begin{cases} 1, & t = 2 \\ 0, & t = 1 \end{cases}$$

Conditional likelihood

$$L_c(\beta) = \prod_{y_{i1}+y_{i2}=1} P\left(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 1\right) = \left(\frac{e^\beta}{1 + e^\beta}\right)^{n_{01}} \left(\frac{1}{1 + e^\beta}\right)^{n_{10}}$$

Hypotheses

$$H_0 : \beta = 0 \qquad H_1 : \beta \neq 0$$

Test statistic

$$\frac{\hat{\beta}}{\sqrt{\frac{1}{n_{01}} + \frac{1}{n_{10}}}} \overset{a}{\sim} N(0, 1)$$

(b) **Regression with covariates**

$$\text{logit}\left(P(y_{it} = 1|\mathbf{x}_{it})\right) = \beta_i + \mathbf{x}_{it}^T\boldsymbol{\gamma}$$

Conditional likelihood

$$\begin{aligned} L_c(\boldsymbol{\gamma}) &= \prod_{y_{i1}+y_{i2}=1} P\left(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 1\right) \\ &= \prod_{y_{i1}+y_{i2}=1} \left(\frac{e^{(\mathbf{x}_{i2}-\mathbf{x}_{i1})^T\boldsymbol{\gamma}}}{1 + e^{(\mathbf{x}_{i2}-\mathbf{x}_{i1})^T\boldsymbol{\gamma}}}\right)^{y_i^*} \left(\frac{1}{1 + e^{(\mathbf{x}_{i2}-\mathbf{x}_{i1})^T\boldsymbol{\gamma}}}\right)^{1-y_i^*} \end{aligned}$$

where

$$y_i^* = \begin{cases} 1, & (y_{i1}, y_{i2}) = (0, 1) \\ 0, & (y_{i1}, y_{i2}) = (1, 0) \end{cases}$$