Advanced dataanalysis and statistical modelling, Week 7

Generalized Linear Models - part III

Jan Kloppenborg Møller, Henrik Madsen
Poul Thyregod

DTU-Compute
Technical University of Denmark
DK-2800 Kgs. Lyngby
jkmo@dtu.dk

March 11, 2018

Overview

The likelihood function

Definition (Likelihood function)

Given the parametric density $f_{\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta^P$, for the observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ the *likelihood function for $\boldsymbol{\theta}$* is the function

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = c(y_1, y_2, \ldots, y_n) f_Y(y_1, y_2, \ldots, y_n; \boldsymbol{\theta})$$

where $c(y_1, y_2, \ldots, y_n)$ is a constant.

- The likelihood function is thus (proportional to) the joint probability density for the actual observations considered as a function of $\boldsymbol{\theta}$.
- Very often it is more convenient to consider the *log-likelihood* function defined as

$$l(\boldsymbol{\theta}; \boldsymbol{y}) = \log(L(\boldsymbol{\theta}; \boldsymbol{y})).$$

Likelihood summary

- Likelihood function $L(\theta) = P_\theta(Y = y)$
- Log likelihood function $\ell(\theta) = \log(L(\theta))$
- Maximum likelihood estimate $\widehat{\theta} = \underset{\theta \in \Theta}{\text{argmax}}\ \ell(\theta)$
- Score function $\ell'(\theta)$
- Observed information matrix $-\ell''(\widehat{\theta})$
- Distribution of the ML estimator $\widehat{\theta} \sim \mathrm{N}(\theta, (-\ell''(\widehat{\theta}))^{-1})$
- Likelihood ratio test $2(\ell_A(\widehat{\theta}_A, Y) - \ell_B(\widehat{\theta}_B, Y)) \sim \chi^2_{\dim(A)-\dim(B)}$
- Dealing with nuisance parameters

Statistical concepts

- Unbiased estimator $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$
- Minimum least square error
  $\mathsf{E}\left[(\widehat{\boldsymbol{\theta}}(\boldsymbol{Y}) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(\boldsymbol{Y}) - \boldsymbol{\theta})^T\right] \leq \mathsf{E}\left[(\tilde{\boldsymbol{\theta}}(\boldsymbol{Y}) - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}(\boldsymbol{Y}) - \boldsymbol{\theta})^T\right]$
- Efficient estimator (relative or absolute (Cramer-Rao))
- Sufficient statistics
- Consistent estimator (convergence of $\hat{\boldsymbol{\theta}}$)
- Dispersion matrix $(\mathrm{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{Y})])$
- Quadratic approximation
- Hypothesis chains (partial tests)
- Profile likelihood

The general linear model - intro

- We will use the term *classical* GLM for the General linear model to distinguish it from GLM which is used for the Generalized linear model.
- The classical GLM leads to a unique way of describing the variations of experiments with a *continuous* variable.
- The classical GLM's include
  - Regression analysis
  - Analysis of variance - ANOVA
  - Analysis of covariance - ANCOVA
- The residuals are assumed to follow a multivariate normal distribution in the classical GLM.

The likelihood and log-likelihood function

- The likelihood function is:

$$L(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y}) = \frac{1}{(\sqrt{2\pi})^n \sigma^n \sqrt{\det(\boldsymbol{\Sigma})}} \exp\left[-\frac{1}{2\sigma^2}\,\mathrm{D}(\boldsymbol{y}; \boldsymbol{\mu})\right]$$

- The log-likelihood function is (apart from an additive constant):

$$\begin{aligned}\ell_{\mu,\sigma^2}(\boldsymbol{\mu}, \sigma^2; \boldsymbol{y}) &= -(n/2)\log(\sigma^2) - \frac{1}{2\sigma^2}\,(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \\ &= -(n/2)\log(\sigma^2) - \frac{1}{2\sigma^2}\,\mathrm{D}(\boldsymbol{y}; \boldsymbol{\mu}).\end{aligned}$$

Estimation of mean value parameters

Under the hypothesis

$$\mathcal{H}_0 : \boldsymbol{\mu} \in \Omega_0 \ ,$$

the maximum likelihood estimate for the set $\boldsymbol{\mu}$ is found as the orthogonal projection (with respect to δ_Σ), $p_0(\boldsymbol{y})$ of $\boldsymbol{y}$ onto the linear subspace Ω_0.

Theorem (ML estimates of mean value parameters)

For hypothesis of the form

$$\mathcal{H}_0 : \boldsymbol{\mu}(\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta}$$

the maximum likelihood estimated for $\boldsymbol{\beta}$ is found as a solution to the normal equation

$$\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

If $\boldsymbol{X}$ has full rank, the solution is uniquely given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

Properties of the ML estimator

Theorem (Properties of the ML estimator)

For the ML estimator we have

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_k(\boldsymbol{\beta}, \sigma^2 \, (\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1})$$

Unknown $\boldsymbol{\Sigma}$

Notice that it has been assumed that $\boldsymbol{\Sigma}$ is known. If $\boldsymbol{\Sigma}$ is unknown different possiblities exist

- The relaxation algorithm described in Madsen (2008) [a].
- Likelihood based method, either direct or by restricted maximum likelihood (eg. `gls` in the the `nlme` package of R)

[a] Madsen, H. (2008) Time Series Analysis. Chapman, Hall

Deviance table

Source	f	Deviance	Test statistic, F
Model versus hypothesis	$m_1 - m_0$	$\|\|p_1(\boldsymbol{y}) - p_0(\boldsymbol{y})\|\|^2$	$\dfrac{\|\|p_1(\boldsymbol{y}) - p_0(\boldsymbol{y})\|\|^2/(m_1 - m_0)}{\|\|\boldsymbol{y} - p_1(\boldsymbol{y})\|\|^2/(n - m_1)}$
Residual under model	$n - m_1$	$\|\|\boldsymbol{y} - p_1(\boldsymbol{y})\|\|^2$	
Residual under hypothesis	$n - m_0$	$\|\|\boldsymbol{y} - p_0(\boldsymbol{y})\|\|^2$	

Table: Deviance table corresponding to a test for model reduction as specified by $\mathcal{H}_0$. For $\Sigma = \boldsymbol{I}$ this corresponds to an analysis of variance table, and then 'Deviance' is equal to the 'Sum of Squared deviations (SS)'

Residual plots

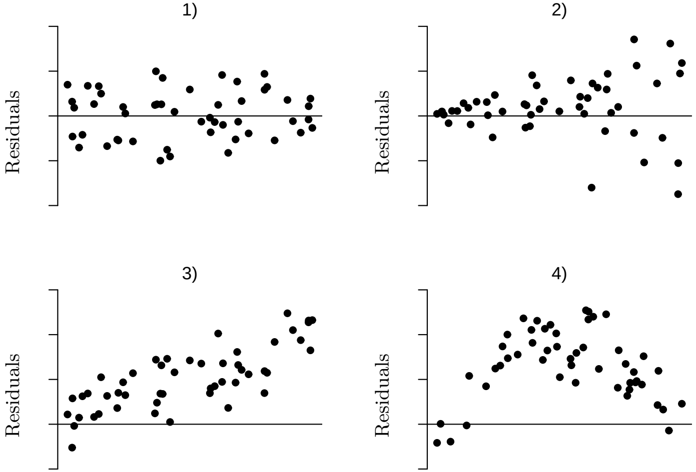

Figure: Residual plots

Other concepts

- Collinearity
- Polynomial / non-parametric regression
- Residial analysis
  - Standardlized residuals
  - Studentlized residuals
  - Leverage / Cook's distance
- Prediction

Types of response variables

i Count data ($y_1 = 57$, …, $y_n = 59$ accidents) - Poisson distribution.

ii Binary response variables ($y_1 = 0, y_2 = 1$, …, $y_n = 0$), or proportion of counts ($y_1 = 15/297$, …, $y_n = 144/285$) - Binomial distribution.

iii Count data, waiting times - Negative Binomial distribution.

iv Multiple ordered categories "Unsatisfied", "Neutral", "Satisfied" - Multinomial distribution.

v Count data, multiple categories.

vi Continuous responses, constant variance ($y_1 = 2.567$, …, $y_n = 2.422$) - Normal distribution.

vii Continuous positive responses with constant coefficient of variation - Gamma distribution.

viii Continuous positive highly skewed - Inverse Gaussian.

Exponential families of distributions

Definition (An exponential dispersion family)

A family of probability densities which can be written on the form

$$f_Y(y;\theta) = c(y,\lambda)\exp(\lambda\{\theta y - \kappa(\theta)\})$$

is called an *exponential dispersion family* of distributions. The parameter $\lambda > 0$ is called the *precision parameter*.

The Generalized Linear Model

Definition (The generalized linear model)

Assume that $Y_1, Y_2, \dots, Y_n$ are mutually independent, and the density can be described by an exponential dispersion model with the same variance function $V(\mu)$.

A *generalized linear model* for $Y_1, Y_2, \dots, Y_n$ describes an affine hypothesis for $\eta_1, \eta_2, \dots, \eta_n$, where

$$\eta_i = g(\mu_i)$$

is a transformation of the mean values $\mu_1, \mu_2, \dots, \mu_n$.

The hypothesis is of the form

$$\mathcal{H}_0 : \boldsymbol{\eta} - \boldsymbol{\eta}_0 \in L,$$

where L is a linear subspace $\mathbb{R}^n$ of dimension k, and where $\boldsymbol{\eta}_0$ denotes a vector of *known off-set values*.

Specification of a generalized linear model

a) Distribution / Variance function:
Specification of the distribution – or the *variance function* $V(\mu)$.

b) Link function:
Specification of the *link function* $g(\cdot)$, which describes a function of the mean value which can be described linearly by the explanatory variables.

c) Linear predictor:
Specification of the linear dependency

$$g(\mu_i) = \eta_i = (\boldsymbol{x}_i)^T \boldsymbol{\beta}.$$

d) Precision (optional):
If needed the precision is formulated as *known individual weights*, $\lambda_i = w_i$, or as a *common dispersion parameter*, $\lambda = 1/\sigma^2$, or a *combination* $\lambda_i = w_i/\sigma^2$.

Properties of the ML estimator

Theorem (Asymptotic distribution of the ML estimator)

Under the hypothesis $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ *we have asymptotically*

$$\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sqrt{\sigma^2}} \in N_k(\mathbf{0}, \boldsymbol{\Sigma}),$$

where the dispersion matrix $\boldsymbol{\Sigma}$ *for* $\widehat{\boldsymbol{\beta}}$ *is*

$$\mathrm{D}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\Sigma} = [\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}) \boldsymbol{X}]^{-1}$$

with

$$\boldsymbol{W}(\boldsymbol{\beta}) = \operatorname{diag}\left\{ \frac{w_i}{[g'(\mu_i)]^2 V(\mu_i)} \right\},$$

Analysis of deviance table

Source	f	Deviance	Mean deviance	Goodness of fit interpretation
Model $\mathcal{H}_{null}$	$k-1$	$\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)}{k-1}$	$G^2(\mathcal{H}_{null} \mid \mathcal{H}_1)$
Residual (Error)	$n-k$	$\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)}{n-k}$	$G^2(\mathcal{H}_1)$
Corrected total	$n-1$	$\mathrm{D}\left(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}_{null}\right)$		$G^2(\mathcal{H}_{null})$

Table: Initial assessment of goodness of fit of a model $\mathcal{H}_0$. $\mathcal{H}_{null}$ and $\widehat{\boldsymbol{\mu}}_{null}$ refer to the *minimal model*, i.e. a model with all observations having the same mean value.

Overview

1. Summary of course (so far)

2. GLM - Examples
    - Example 1: Germination of Orobanche
    - Exampe 2: Accident rates
    - Example 3: The Challenger disaster
    - Example 4: Customer satisfaction for bus passengers
    - Example 5: Empirical variances for normally distributed observations

Example 1: Germination of Orobanche [1] (Binomial, Overdispersion)

Orobanche is a genus of parasitic plants without chlorophyll that grows on the roots of flowering plants. An experiment was made where a bach of seeds of the species Orobanche aegyptiaca was brushed onto a plate containing an extract prepared from the roots of either a bean or a cucumber plant. The number of seeds that germinated was then recorded. Two varieties of Orobanche aegyptiaca namely O.a. 75 and O.a. 73 were used in the experiment.

O. aegyptiaca 75				*O. aegyptiaca* 73			
Bean		Cucumber		Bean		Cucumber	
y	n	y	n	y	n	y	n
10	39	5	6	8	16	3	12
23	62	53	74	10	30	22	41
23	81	55	72	8	28	15	30
26	51	32	51	23	45	32	51
17	39	46	79	0	4	3	7
		10	13				

[1] Modelling binary data, David Collett

The model

We shall assume that the number of seeds that germinated y_i in each independent experiment followers a binomial distribution:

$$
\begin{aligned}
& y_i \sim \mathsf{Bin}(n_i, p_i) \ , \ \text{where} \\
& \mathsf{logit}(p_i) = \mu + \alpha(\mathsf{root}_i) + \beta(\mathsf{variety}_i) + \gamma(\mathsf{root}_i, \mathsf{variety}_i)
\end{aligned}
$$

Overdispersion?

The deviance is to big. Possible reasons are:

- Incorrect linear predictor
- Incorrect link function
- Outliers
- Influential observations
- Incorrect choose of distribution

To check this we need to look at the residuals! If all the above looks ok the reason might be over-dispersion.

Overdispersion

- In the case of over-dispersion the variance is larger than expected for the given distribution.
- When data are *overdispersed*, a *dispersion parameter*, σ^2, should be included in the model.
- We use $\mathrm{Var}[Y_i] = \sigma^2 V(\mu_i)/w_i$ with σ^2 denoting the overdispersion.
- Including a dispersion parameter does not affect the estimation of the mean value parameters $\boldsymbol{\beta}$.
- Including a dispersion parameter does affect the standard errors of $\boldsymbol{\beta}$.
- The distribution of the test statistics will be influenced.

Possible reasons for overdispersion

Nothing in the plots indicate that the model is not reasonable. We conclude that the big residual deviance is because of overdispersion.

In binomial models overdispersion can often be explained by variation between the response probabilities or correlation between the binary responses. In this case it might because of:

- The batches of seeds of particular spices germinated in a particular root extract are not homogeneous.
- The batches were not germinated under similar experimental conditions.
- When a seed in a particular batch germinates a chemical is released that promotes germination in the remaining seeds of the batch.

Overdispersion - some facts

- The residual deviance cannot be used as a goodness of fit in the case of overdispersion.
- In the case of overdispersion an F-test should be used in stead of the χ^2 test. The test is not exact in contrast to the Gaussian case.
- When fitting a model to overdispersed data in R we use `family = quasibinomial` for binomial data and `family = quasipoisson` for Poisson data.
- The families differ from the binomial and poisson families only in that the dispersion parameter is not fixed at one, so they can model over-dispersion.

Model results

Probability of germination is $\frac{e^{-0.512}}{1+e^{-0.512}} \approx 37\%$ on bean roots.

Probability of germination is $\frac{e^{-0.512+1.0574}}{1+e^{-0.512+1.0574}} \approx 63\%$ on cucumber roots.

The odds ratio becomes:

$$\frac{\text{odds}(\text{Germination}|\text{Cucumber})}{\text{odds}(\text{Germination}|\text{Bean})} \approx 2.88$$

with confidence interval from 1.9 to 4.4.

Consider The model

Will still assume that the number of seeds that germinated y_i in each independent experiment follow a binomial distribution:

$$y_i \sim \text{Bin}(n_i, p_i) \ , \ \text{where}$$
$$\text{logit}(p_i) = \mu + \alpha(\text{root}_i) + \beta(\text{variety}_i) + \gamma(\text{root}_i, \text{variety}_i) + B_i$$

Where $B_i \sim N(0, \sigma^2)$

Notice B_i is unobserved

In some sense this model does exactly what we need.

Can we even handle such a model? Yes! Wait for next chapter...

Exampe 2: Accident rates [2]

Events that may be assumed to follow a Poisson distribution are sometimes recorded on units of different size. For example number of crimes recorded in a number of cities depends on the size of the city. Data of this type are called *rate data*.

If we denote the measure of size with t, we can model this type of data as:

$$\log\left(\frac{\mu}{t}\right) = \boldsymbol{X}\boldsymbol{\beta}$$

and then

$$\log(\mu) = \log(t) + \boldsymbol{X}\boldsymbol{\beta}$$

[2] Generalized linear models, Ulf Olsson

Accident rates

The data are accidents rates for elderly drivers, subdivided by sex. For each sex, the number of person years (in thousands) are also given.

	Females	Males
No. of accidents	175	320
No. of person years	17.30	21.40

We can model these data using Poisson distribution and a log link and using number of person years as offset.

Residual deviance as goodness of fit - binomial/binary data

- When $\sum_i n_i$ is reasonable large the χ^2-approximation of the residual deviance is usually good and the residual deviance can be used as a goodness of fit.
- The approximation is not particularly good if some of the binomial denominators n_i are very small and the fitted probabilities under the current model are near zero or unity.
- In the special case when n_i, for all i, is equal to 1, that is the data is binary, the deviance is not even approximately distributed as χ^2 and the deviance can not be used as a goodness of fit.

More comments...

- In a binomial setup where all n_i are big the standardized deviance residuals should be close to Gaussian. The normal probability plot can be used to check this.
- In a Poisson setup where the counts are big the standardized deviance residuals should be close to Gaussian. The normal probability plot can be used to check this.
- In a binomial setup where x_i (number of successes) are very small in some of the groups numerical problems sometimes occur in the estimation. This is often seen in very large standard errors of the parameter estimates.

Example 3: The Challenger disaster

On January 28, 1986, Space Shuttle Challenger broke apart 73 seconds into its flight and the seven crew members died. The disaster was due to a disintegration of an O-ring seal in the right rocket booster. The forecast for January 28, 1986 indicated an unusually cold morning with air temperatures around 28 degrees F (-1 degrees C).

The planned launch on January 28, 1986 was launch number 25. During the previous 24 launches problems with the O-ring were observed in 6 cases. A model of the probability for O-ring failure as a function of the air temperature would clearly have shown that given the forecasted air temperature, problems with the O-rings were very likely to occur.

The Challenger disaster

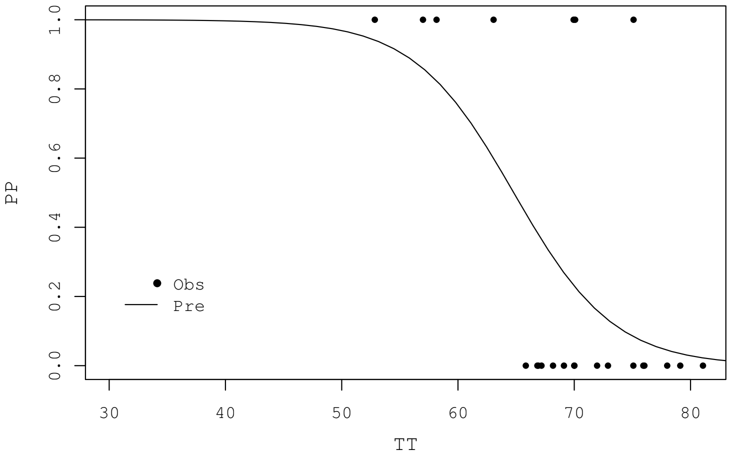

Figure: Observed failure of O-rings in 6 out of 24 launches along with predicted probability for O-ring failure.

Example 4: Customer satisfaction for bus passengers

A widespread method for assessing customer satisfaction is to present a questionnaire to customers. The questionnaire contains a series of items each representing a quality feature of the product in question, and the respondent is asked to indicate his satisfaction by selecting one of the options:

Very dissatisfied	☐
Dissatisfied	☐
Neutral	☐
Satisfied	☐
Very satisfied	☐

Example: Customer satisfaction for bus passengers

How satisfied are you with the punctuality of this bus?

Delay	response, number(%)					
min	V. diss.	Diss.	Neutral	Satis.	V. satis.	Total
0	234(2.3)	559(5.4)	1157(11.2)	5826(56.4)	2553(24.7)	10329
2	41(3.6)	100(8.9)	145(12.9)	602(53.5)	237(21.1)	1125
5	42(7.9)	76(14.3)	89(16.7)	254(47.7)	72(13.5)	533
7	35(14.3)	48(19.7)	39(16.0)	95(38.9)	27(11.1)	244

Table: Customer satisfaction

Example: Customer satisfaction for bus passengers

As the categories are *ordered*, we may meaningfully introduce the *cumulative probabilities*,

$$\Pi_j = p_1 + p_2 + \cdots + p_j \ , \quad j = 1, 2, 3, 4$$

of responding in category j, or lower. For other parameterizations, e.g. continuation logit see Agresti (2002).

Consider the *logistic model* for the cumulative probabilities

$$\log\left(\frac{\Pi_j(t)}{1 - \Pi_j(t)}\right) = \alpha_j + \beta_j t \quad , j = 1, \ldots, 4$$

corresponding to

$$\Pi_j(t) = \frac{\exp(\alpha_j + \beta_j t)}{1 + \exp(\alpha_j + \beta_j t)}$$

Each cumulative logit includes data from all 5 categories.

Example: Cumulated data table

How satisfied are you with the punctuality of this bus?

Delay	Cumulated response, number(%)				
min	V. diss.	Diss.	Neutral	Satis.	Total
0	234(2.3)	793(7.7)	1950(18.9)	7776(75.3)	10329
2	41(3.6)	141(12.5)	286(25.4)	888(78.9)	1125
5	42(7.9)	118(22.1)	207(38.8)	461(86.5)	533
7	35(14.3)	83(34.0)	122(50.0)	217(88.9)	244

Table: Customer satisfaction

Example 5: Empirical variances for normally distributed observations

An experiment aiming at the determination of washing power for a detergent consists in uniformly staining sheets of cloth, and subsequently washing them with the detergent and afterward the cleanness is determined by means of measurements of the reflection.

Example 5: Empirical variances for normally distributed observations

In the experiment 3, 5, and 7 sheets were washed simultaneously. The table shows the empirical variances for each of these trials.

Number of sheets, x_i	3	5	7
Sum of squares	2.5800	4.8920	4.9486
Degrees of freedom, f_i	2	4	6
Empirical variance, s_i^2	1.2900	1.2230	0.8248

Table: The empirical variances for each of the trials.

Empirical variances for normally distributed observations

Assuming original data is normal we have $X_i = \frac{S_i^2 f_i}{\sigma^2} \sim \chi^2(f_i)$, i.e.

$$f_{X_i}(x) = \frac{1}{2\Gamma(f_i/2)} \left(\frac{x}{2}\right)^{f_i/2-1} e^{-x/2}$$

using the change of variable formula we get

$$\begin{aligned} f_{S_i^2}(s^2) =& \frac{f_i}{2\sigma^2\Gamma(f_i/2)} \left(\frac{s^2 f_i}{2\sigma^2}\right)^{f_i/2-1} e^{-s^2 f_i/(2\sigma^2)} \\ =& \frac{1}{\beta_i\Gamma(\alpha_i)} \left(\frac{s^2}{\beta_i}\right)^{\alpha_i-1} e^{-s^2/\beta_i} \end{aligned}$$

with $\alpha_i = f_i/2$ and $\beta_i = \frac{2\sigma^2}{f_i}$, hence this is a gamma distribution with parameters α_i and β_i.

Empirical variances for normally distributed observations

Rewriting in the fom of the exponential dispersion family

$$\begin{aligned}
f_{S_i^2}(s^2) =& \frac{1}{\beta_i \Gamma(\alpha_i)} \left(\frac{s^2}{\beta_i}\right)^{\alpha_i - 1} e^{-s^2/\beta_i} \\
=& e^{-s^2/\beta_i - \log(\beta_i) - \log(\Gamma(\alpha_i)) + (\alpha_i - 1)\log(s^2) - (\alpha_i - 1)\log(\beta_i)} \\
=& a(s_i^2, \alpha_i) e^{-s^2/\beta_i - \alpha_i \log(\beta_i)} \\
=& a(s_i^2, \alpha_i) e^{\alpha_i (s^2 \theta_i - \log(-1/\theta_i))}
\end{aligned}$$

with $\theta_i = -\frac{\alpha_i}{\beta_i}$, hence this is in the form of the exponential dispersion family. E.g. we get $\kappa' = -(\frac{1}{\theta})^{-1} \frac{1}{\theta^2} = -\frac{1}{\theta} = \frac{\beta_i}{\alpha_i} = \frac{2\sigma^2}{f_i} \frac{f_i}{2} = \sigma^2$.
Hence we can use a GLM model with weights (α_i).

Summary