```
Residual deviance: 20.943 on 13 degrees of freedom
```

The standard errors are not correct in this output and further calculation, described in Smyth, Huele, and Verbyla (2001), would be necessary. This would result in somewhat larger standard errors (about twice the size), but the two factors would still be significant.

7.4 Quasi-Likelihood

Suppose that we are able to specify the link and variance functions of the model for some new type of data, but that we do not have a strong idea about the appropriate distributional form for the response. For example, suppose that we specify an identity link and constant variance. This would be typical in the standard regression setting. We can use least squares to estimate the regression parameters. If we want to do some inference, then formally we need to assume a Gaussian distribution for the errors (or equivalently, the response). We know that the inference is fairly robust to nonnormality especially as the sample size gets larger. The important part of the model specification is the link and variance; the outcome is less sensitive to the distribution of the response.

The same effect holds for other GLMs. Provided we have a larger sample, the results are not sensitive to smaller deviations from the distributional assumptions. The link, variance and independence assumptions are far more important. Now suppose that we were to specify a link and variance function combination that does not correspond to any of the standard GLMs. An examination of the fitting procedure for GLMs reveals that only the link and variance functions are used and no distributional assumptions are necessary. This opens up new modeling possibilities because one might well be able to suggest reasonable link and variance functions but not know a suitable distribution.

Computation of $\hat{\beta}$ and standard errors is often not enough and some form of inference is required. To compute a deviance, we need a likelihood and to compute a likelihood we need a distribution. At this point, we need a suitable substitute for a likelihood that can be computed without assuming a distribution.

Let Y_i have mean μ_i and variance $\phi V(\mu_i)$. We assume that Y_i are independent. We define a score, U_i:

$$U_i = \frac{Y_i - \mu_i}{\phi V(\mu_i)}$$

Now:

$$EU_i = 0$$

$$\text{var } U_i = \frac{1}{\phi V(\mu_i)}$$

$$-E\frac{\partial U_i}{\partial \mu_i} = -E\frac{-\phi V(\mu_i) - (Y_i - \mu_i)\phi V'(\mu_i)}{[\phi V(\mu_i)]^2} = \frac{1}{\phi V(\mu_i)}$$

These properties are shared by the derivative of the log-likelihood, l'. This suggests that we can use U in place of l'. So we define:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt$$

The intent is that Q should behave like the log-likelihood. We then define the log *quasi-likelihood* for all n observations as:

$$Q = \sum_{i=1}^{n} Q_i$$

The usual asymptotic properties expected of maximum likelihood estimators also hold for quasi-likelihood-based estimators as may be seen in McCullagh (1983).

Notice that the quasi-likelihood depends directly only on the variance function and that the choice of distribution also determines only the variance function. So the choice of variance function is associated with the random structure of the model while the link function determines the relationship with the systematic part of the model.

For the variance functions associated with the members of the exponential family distribution, the quasi-likelihood corresponds exactly to the log-likelihood. However, there is an advantage to using the quasi-likelihood approach for models with variance functions corresponding to the binomial and Poisson distribution. The regular GLMs assume $\phi = 1$ whereas the corresponding *quasi-binomial* and *quasi-Poisson* GLMs allow for the dispersion ϕ to be a free parameter which is useful in modeling overdispersion. One curious possibility is that some choices of $V(\mu)$ may not correspond to a known, or even any, distribution.

$\hat{\beta}$ is obtained by maximizing Q. Everything proceeds as in the standard GLMs except for the estimation of ϕ since the likelihood approach is not reliable here. We recommend:

$$\hat{\phi} = \frac{X^2}{n-p}$$

Although quasi-likelihood estimators are attractive because they require fewer assumptions, they are generally less efficient than the corresponding regular likelihood-based estimator. So if you have information about the distribution, you are advised to use it.

The inferential procedures are similar to those for standard GLMs. Recall that the regular deviance for a model is formed from the difference in log-likelihoods for the model and the saturated model:

$$D(y,\hat{\mu}) = -2\phi \sum_i (l(\hat{\mu}_i|y_i) - l(y_i|y_i))$$

so by analogy the quasi-deviance is $-2\phi Q$ because the contribution from the saturated model is zero. The ϕ cancels, so the quasi-deviance is just:

$$Q = -2\sum_i \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt$$

In Allison and Cicchetti (1976), data on the sleep behavior of 62 mammals is presented. Suppose we are interested in modeling the proportion of sleep spent dreaming as a function of the other predictors: the weight of the body and the brain, the lifespan, the gestation period and the three constructed indices measuring vulnerability to predation, exposure while sleeping and overall danger:

```
> data(mammalsleep)
> mammalsleep$pdr <- with(mammalsleep, dream/sleep)
> summary(mammalsleep$pdr)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   0.118   0.176   0.186   0.243   0.462  14.000
```

We notice that the proportion of time spent dreaming varies from zero up to almost half the time. A normal model seems inappropriate while transformations are problematic. We attempt to model the proportion response directly. A logit link seems sensible since the response is restricted between zero and one. Furthermore, we might expect the variance to be greater for moderate values of the proportion μ and less as μ approaches zero or one because of the nature of the measurements. This suggests a variance function of the approximate form $\mu(1-\mu)$. This corresponds to the binomial GLM with the canonical logit link and yet the response is not binomial. We propose a quasi-binomial:

```
> modl <- glm(pdr ~
log(body)+log(brain)+log(lifespan)+log(gestation)
          +predation+exposure+danger,
family=quasibinomial, mammalsleep)
```

where we have logged many of the predictors because of skewness. Since we now have a free dispersion parameter, we must use *F*-tests to compare models:

```
> dropl(modl,test="F")
Single term deletions
               Df Deviance F value Pr (F)
<none>               1.57
log(body)       1     1.78    4.51 0.041
log(brain)      1     1.59    0.33 0.568
log(lifespan)  -1     1.65    1.79 0.189
log(gestation)  1     1.62    1.15 0.292
predation       1     1.57    0.10 0.749
exposure        1     1.58    0.32 0.575
danger          1     1.58    0.31 0.579
```

We might eliminate predation as the least significant variable. Further sequential backward elimination results in:

```
> modl <- glm(pdr ~ log(body)+log(lifespan)+danger,
          family=quasibinomial, mammalsleep )
> summary(modl)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.4932     0.2913   -1.69  0.09796
log(body)       0.1463     0.0384    3.81  0.00046
log(lifespan)  -0.2866     0.1080   -2.65  0.01126
danger         -0.1732     0.0600   -2.89  0.00615
(Dispersion parameter for quasibinomial family taken to
be 0.040654)
    Null deviance: 2.5088 on 44 degrees of freedom
Residual deviance: 1.7321 on 41 degrees of freedom
AIC: NA
```

Notice that the dispersion parameter is far less than the default value of one that we would see for a binomial. Furthermore, the AIC is not calculated since we do not have a true likelihood. We see that the proportion of time spent dreaming increases for heavier mammals that live less time and live in less danger. Notice that the relatively large residual deviance compared to the null deviance indicates that this is not a particularly well-fitting model.

The usual diagnostics should be performed. Here are two selected plots that have some interest:

```
> ll <- row.names(na.omit(mammalsleept,c(1,6,10,11)]))
> halfnorm(cooks.distance(modi),labs=ll)
> plot(predict (modl),residuals(modl,type="pearsonn),
  xlab="Linear Predictor", ylab="Pearson Residuals")
```

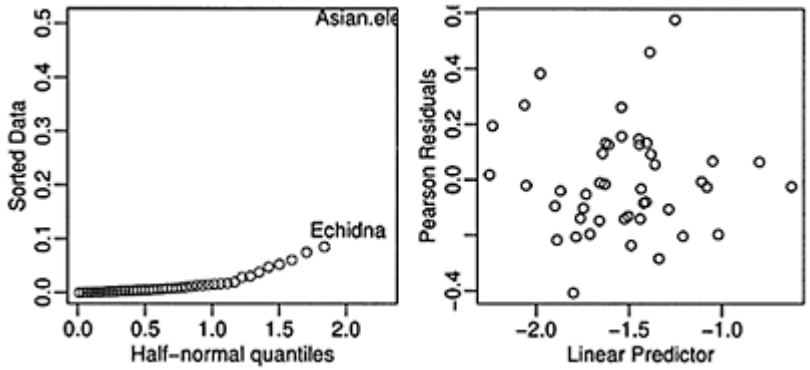

Figure 7.5 *A half-normal plot of the Cook statistics is shown on the left and a plot of the Pearson residuals against the fitted linear predictors is shown on the right.*

In the first panel of Figure 7.5, we see that the Asian elephant is quite influential and a fit without this case should be considered. In the second panel, we see that a pattern of constant variation indicating that our choice of variance function was reasonable. We used the Pearson residuals because these explicitly normalize the raw residuals using the variance function making the check more transparent. Even so, the deviance residuals would have served the same purpose.

Exercises

1. The relationship between corn yield (bushels per acre) and nitrogen (pounds per acre) fertilizer application were studied in Wisconsin. The data may be found in cornnit.

   (a) Using (Gaussian) linear model methods, represent the relationship between the yield response and the nitrogen predictor. You will need to find appropriate transformations for the data. Present a quantitative interpretation for the effect of nitrogen fertilizer on yield.
   (b) Now develop a GLM for the data that does not (explicitly) transform the response. Describe quantitatively the relationship between the response and the predictor and compare it to the linear model you found in the previous question.

2. An experiment was conducted as part of an investigation to combat the effects of certain toxic agents. The survival time of rats depended on the type of poison used and the treatment applied. The data is found in rats.

   (a) Construct a linear model for the data bearing in mind that some transformation of the response may be necessary and that the possibility of interactions needs to be considered. Interpret the effects of the poisons.
   (b) Build an inverse Gaussian GLM for this data. Select an appropriate link function and perform diagnostics to verify your choices. Interpret the effects of the poisons.

3. Components are attached to an electronic circuit card assembly by a wave-soldering process. The soldering process involves baking and preheating the circuit card and then passing it through a solder wave by conveyor. Defects arise during the process. Design is 2^{7-3} with 3 replicates. The data is found in wavesolder. Build a pair of models for the mean and dispersion in the number of defects. Investigate which factors are significant in the two models.
4. Data were collected from 39 students in a University of Chicago MBA class and presented in happy. Happiness was measured on a 10 point scale. The response could be viewed as ordinal, but a quasi-likelihood approach is also possible. Build a quasi-GLM selecting appropriate link and variance functions. Interpret the effect of the predictors.
5. The leafblotch data shows the percentage leaf area affected by leaf blotch on 10 varieties of barley at nine different sites. The data comes from Wedderburn (1974). The data is analyzed in McCullagh and Nelder (1989), which motivates the following questions:

   (a) Fit a quasi-GLM with a logit link and a $\mu(1-\mu)$ variance function. Construct a diagnostic plot that shows that this is not a good choice of variance function.