

## Basic facts

### Common GLM distributions

Given the pdf:

$$f(y; \theta_i, \phi) = \exp\left[\frac{y\theta_i - b(\theta_i)}{\phi/w} + c(y, \phi)\right] \quad (1)$$

We know that

$$E(Y_i) = \mu_i = h(x_i^T \beta) = b'(\theta) \quad (2)$$

Therefore:

$$x_i^T \beta = h^{-1}(b'(\theta)) = g b'(\theta) \quad (3)$$

↑  
canonical link

Distribution	$h(x_i^T \beta) = \mu_i$	$g(\mu_i) = \theta_i$
Binomial	$\frac{\exp[\theta_i]}{1 + \exp[\theta_i]}$	$\log \frac{y}{1-y}$
logit link		
Normal	$\theta$	$g = h$
identity		
Poisson	$\exp[\theta]$	$\log[\mu]$
log		
Gamma	$-\frac{1}{\theta}$	$-\frac{1}{\mu_i}$
inverse		
cloglog	$1 - \exp[-\exp[\theta_i]]$	$\log(-\log(1 - \mu_i))$
cloglog		
Probit	$\Phi(\theta)$	$\Phi^{-1}(\theta)$ (qnorm)
probit		

The big thing about the canonical link is that it expresses  $\theta_i$  as a linear combination of the parameters:  $x_i^T \beta$ .

**Relevance of canonical link:** You can decide which link to use by plotting  $g(\mu_i)$  against the predictor (in case we have only a single predictor  $x$ ).

## Iteratively reweighted least squares

- Specify an initial vector of parameters:  $b^{(m)} = (\beta_0, \dots, \beta_p)^T$ , where initially  $m = 1$ :

```
> ## eta=xbeta:
> eta.i <- -60+35*beetle$conc
```

- Specify a weight matrix  $W$  that depends on current parameter estimates:

Given (proof on p. 83-84):

$$w_{ii} = \frac{n_i \exp[\eta_i]}{(1 + \exp[\eta_i])^2} \quad (4)$$

we can compute  $W$ :

```
> n.i <- beetle$number
> w.i.fn <- function(n.i, eta.i){
+   (n.i*exp(eta.i))/(1+exp(eta.i))^2
+ }
> w.iis <- w.i.fn(n.i, eta.i)
> ##weights matrix:
> W <- diag(as.vector(w.iis))
```

- Specify a vector  $z$  that depends on the current parameter estimates and response values:

$$z_i = \eta_i + \frac{y_i - \mu_i}{\mu_i(1 - \mu_i)} \quad \mu_i = \frac{\exp[\eta_i]}{1 + \exp[\eta_i]} \quad (5)$$

```
> mu.i <- exp(eta.i)/(1+exp(eta.i))
> z.i <- eta.i + ((beetle$propn.death-mu.i))/
+   (mu.i*(1-mu.i))
```

- Compute new estimate of parameters:  $b^{(m+1)} = (X^T W X)^{-1} X^T W z$ :

```
> ##The design matrix:
> col1 <- c(rep(1, 8))
> K <- as.matrix(cbind(col1, beetle$conc))
> ## update coeffs:
> eta.i <- solve(t(X)%*%W%*%X)%*%
+   t(X)%*%W%*%z.i
```

Stop at convergence.

## Residual deviances

- Normal:  $\sum (y_i - \hat{\mu}_i)^2$
- Poisson:  $2 \sum y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)$
- Binomial:

$$-2 \sum_i n_i [y_i \log\left(\frac{\mu_i}{y_i}\right) + (1 - y_i) \log\left(\frac{1 - \mu_i}{1 - y_i}\right)]$$

- Gamma:  $-2 \sum \log\left(\frac{y_i}{\hat{\mu}_i}\right) - \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$

## Testing model fit using pseudo $R^2$ and GLRT

Let the log likelihood of the minimal model (with only an intercept) be:

$$l(\tilde{\mu}, \phi; y) \quad (6)$$

Pseudo  $R^2$  is the proportional improvement in the log-likelihood due to the model under consideration compared to the minimal mode:

$$\frac{l(\tilde{\mu}, \phi; y) - l(\hat{\mu}, \phi; y)}{l(\tilde{\mu}, \phi; y)} \quad (7)$$

**Example:** To compute pseudo  $R^2$ , we need the AIC value of the models. Recall that

$$AIC = -2l + 2p \Leftrightarrow l = p - \frac{1}{2} AIC \quad (8)$$

## Binomial distribution

$$\binom{n}{ny} p^{ny} (1 - p)^{n - ny}, \quad B_i(ny, \frac{\exp[\theta]}{1 + \exp[\theta]}) \quad \mu = \frac{\exp[\theta]}{1 + \exp[\theta]}.$$

## Logit link

$$f(y; \theta_i, \phi) = \exp\left[\frac{y\theta_i - b(\theta_i)}{\phi/w} + c(y, \phi)\right]$$

$$1. b(\theta) = \log(1 + \exp[\theta])$$

$$(a) b'(\theta) = \mu = \frac{\exp[\theta]}{1 + \exp[\theta]}$$

$$(b) b''(\theta) = \mu(1 - \mu)$$

$$2. c(y, \phi) = \log\binom{n}{ny} \text{ and } \phi = 1, w = n.$$

$$3. \text{ The model: } \log\left[\frac{\mu}{1-\mu}\right] = \beta_0 + \beta_1 x = \eta, h(\eta) = \mu \Leftrightarrow \mu = \frac{\exp[\eta]}{1 + \exp[\eta]} \text{ and } g(\mu) = \eta \Leftrightarrow \log\left[\frac{\mu}{1-\mu}\right] = \eta.$$

$$4. \text{ Mean and Variance: } E(Y) = \mu, \text{Var}(Y) = b''(\theta)a(\phi) = \frac{\phi}{w}V(\mu), \text{ where } V(\mu) = \mu(1 - \mu).$$

$$5. \text{ Residual deviance:}$$

**Maximal model:**  $\mu_i^\diamond = y_i$ . Recall that:

$$\begin{aligned} \ell &= \log f_i(y; \theta_i, \phi) \\ &= \log \exp[w_i \frac{y\theta_i - b(\theta_i)}{\phi} + c(y, \phi)] \\ &= w_i \frac{y\theta_i - b(\theta_i)}{\phi} + c(y, \phi) \end{aligned} \quad (9)$$

For the maximal model:

$$\ell(y; \theta_i^\diamond, \phi) = w_i \frac{y\theta_i^\diamond - b(\theta_i^\diamond)}{\phi} + c(y, \phi) \quad (10)$$

For the model under consideration:

$$\ell(y; \hat{\theta}_i, \phi) = w_i \frac{y\hat{\theta}_i - b(\hat{\theta}_i)}{\phi} + c(y, \phi) \quad (11)$$

Then, scaled deviance for a model  $\mu_i = h(x_i^T \beta)$  is defined as:

$$\begin{aligned} S(y, \hat{\mu}) &= -2[\ell(\hat{\mu}, \phi, y) - \ell(\hat{\mu}^\diamond, \phi, y)] \\ &= 2 \sum_i \left[ \frac{w_i}{\phi} [y(\theta_i^\diamond - \hat{\theta}_i)] - b(\theta_i^\diamond) + b(\hat{\theta}_i) \right] \end{aligned} \quad (12)$$

Note that  $\phi S(y, \hat{\mu})$  depends only on the data (including  $w_i$ ). This is called residual deviance or deviance.

$$D(y, \hat{\mu}) = \phi S(y, \hat{\mu}) = 2 \sum_i w_i [y(\theta_i^\diamond - \hat{\theta}_i) - b(\theta_i^\diamond) + b(\hat{\theta}_i)] \quad (13)$$

Asymptotically,  $S(y, \hat{\mu})$  has a  $\chi_{n-p}^2$  distribution.

$$6. \text{ Pearson residuals:}$$

These are approx.  $N(0, 1)$ .

$$\begin{aligned} e_{P,i} &= \sqrt{w_i} \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \\ &= \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)/w_i}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(Y_i)/\phi}} \end{aligned} \quad (14)$$

(a) Pearson chi-sq statistic:

$$X^2 = \sum e_{P,i}^2 \quad (15)$$

This is asymptotically equivalent to the deviance (D) for a model.

$$7. \text{ Deviance residuals For the binomial distribution: Deviance } D = \sum d_i, \text{ where:}$$

$$d_i = -2 \times n_i \left[ y_i \log\left(\frac{\hat{\mu}_i}{y_i}\right) + (1 - y_i) \log\left(\frac{1 - \hat{\mu}_i}{1 - y_i}\right) \right] \quad (16)$$

The  $i$ -th deviance residual is:

$$e_{D,i} = \text{sgn}(y_i - \hat{\mu}_i) \times \sqrt{d_i} \quad (17)$$

Note that  $\sum e_{D,i} = D$ .

$$8. \text{ Log odds and odds ratio}$$

$$(a) \text{ log odds: } \lambda = \log\left[\frac{\mu}{1-\mu}\right].$$

$$\begin{aligned} \text{Var}(\log \lambda) &= \frac{1}{n\mu} + \frac{1}{n(1-\mu)} \\ \text{est. Var}(\log \hat{\lambda}) &= \frac{1}{s} + \frac{1}{n-s} \end{aligned} \quad (18)$$

**Example:** Beetle dataset.

```
> fml<-glm(propn.death~conc,
+          binomial(logit),
+          weights=number,
+          data=beetle)
> #summary(first.beetle.glm)
>
> ## compute log odds of death for
> ## concentration 1.7552:
> x<-as.matrix(c(1, 1.7552))
> #log odds:
> (log.odds<-t(x)%*coef(fml))
      [,1]
[1,] -0.56618
> ### compute CI for log odds:
> ## Get vcov matrix:
> (vcovmat<-vcov(fml))
      (Intercept)      conc
(Intercept)    26.840 -15.0821
conc          -15.082   8.4805
```

```

> ## x~T VCOV x:
> (var.log.odds<-t(x)%*%vcovmat%*%x)Ddds(Control group gets Flu)
      [,1]
[1,] 0.021678
> ##lower
> #log.odds-1.96*sqrt(var.log.odds)
> ##upper
> #log.odds+1.96*sqrt(var.log.odds)
>
> ## variance of log odds using
> ## formula, does not match up
> ## because it's based only
> ## one data point:
> (1/18) + (1/(62-18))
[1] 0.078283

```

(b) **Odds ratio:** Given:

$$\frac{p(Y_i = 1)}{1 - p(Y_i = 1)} = \frac{\mu_i}{1 - \mu_i} \quad (19)$$

taking logs:

$$\log \frac{\mu_i}{1 - \mu_i} = \alpha + \beta x_i \quad (20)$$

Therefore the odds of Y=1 are:

$$\frac{p(Y_i = 1)}{1 - p(Y_i = 1)} = \exp[\alpha + \beta x_i] \quad (21)$$

### Computing odds ratios by hand:

Odds= no. successes / no. failures

**Odds ratio:**

	A1	A2	Totals
B1	w	x	w+x
B2	y	z	y+z
	w+y	x+z	

**Odds ratio (OR):**

$$\frac{w/x}{y/z} \quad (22)$$

**CI's for log(OR):**

$$\log(OR) \pm 1.96 \times se(\log(OR)) \quad (23)$$

where

$$se(\log(OR)) = \sqrt{\frac{1}{w} + \frac{1}{x} + \frac{1}{y} + \frac{1}{z}} \quad (24)$$

To get CI's at odds scale just take exponents.

### 9. Over-dispersion:

$$\hat{\phi} = \frac{D}{n-p} \approx X^2/(n-p) \quad (25)$$

In binomial data, if observations  $Y_i$  have variance greater than that expected from the binomial theorem, then you need to adjust the variance estimate. If the deviance D is greater than N-p (as expected from the fact that it has an approximate

chi-squared deviation and the expectation of a chi-sq distributed random variable is N-p; I think), we should suspect that we have an overdispersion problem. Also, correlated binary responses also lead to overdispersion.

Here, we assume that  $Var(Y_i) = \phi \frac{\mu_i(1-\mu_i)}{n_i}$ .

Once the dispersion parameter (e.g., 3.6185, in an example in lecture notes) has been estimated, we adjust the variance:

$$Var(Y_i) = \phi \frac{\mu_i(1-\mu_i)}{n_i} = 3.6185 \frac{\mu_i(1-\mu_i)}{n_i}$$

I.e., standard errors for the coefficients in the quasibinomial are the result of multiplying the regular SE with the square root of the dispersion parameter estimated.

## Poisson

Let the random variable  $X$  count the number of events occurring in the interval. Then under certain reasonable conditions it can be shown that

$$f_X(x) = \mathbb{P}(X = x) = e^{-\mu} \frac{\mu^x}{x!}, \quad x = 0, 1, 2, \dots \quad (26)$$

In GLM setting, there are two situations where we use the Poisson function:

1. **Poisson regression:** The events depend on varying amounts of exposure. Predictors can be categorical or continuous.

2. **Log-linear models:** Exposure is constant. Predictors are usually categorical.

$$f(y; \theta_i, \phi) = \exp\left[\frac{\eta\theta_i - b(\theta_i)}{\phi/w} + c(y, \phi)\right]$$

1.  $b(\theta) = \exp[\theta]$

$$(a) \quad b'(\theta) = \mu = \exp[\theta]$$

$$(b) \quad b''(\theta) = \exp[\theta]$$

2.  $c(y, \phi) = -\log y!$  and  $\phi = 1, w = n$ .

3. The model:  $\log \mu = \beta_0 + \beta_1 x = \eta$ ,  $h(\eta) = \mu \Leftrightarrow \mu = \exp[\eta]$  and  $g(\mu) = \eta \Leftrightarrow \log \mu = \eta$ .

4. Mean and Variance:  $E(Y) = \mu$ ,  $Var(Y) = b''(\theta)a(\phi) = \frac{\phi}{w}V(\mu)$ , where  $V(\mu) = \mu$ .

If  $Y_i$  are independent RVs, each denoting the number of events observed from exposure  $n_i$  (example: numbers of smoking doctors in each age group).

**Offset:** Let  $E(Y_i) = \mu_i = n_i \theta_i$ . Here,  $Y_{-i}$  is a count, and  $\theta_i$  a function of the predictors  $X\beta$ :  $\theta_i = \exp[x_i^T \beta]$ .

Therefore, the GLM is:

$$E(Y_i) = \mu_i = n_i \exp[x_i^T \beta] \quad Y_i \sim Po(\mu_i) \quad (27)$$

The link function is:

$$\log \mu_i = \log n_i + x_i^T \beta \quad (28)$$

where  $\log n_i$ : offset.

**Fitted values:**  $\hat{Y}_i = \hat{\mu}_i = n_i \exp[x_i^T \beta] = e_i$ , where  $e_i$  refers to **expected value** for  $i$ .

Since  $Var(Y_i) = \mu$ ,  $SE(Y_i) = \sqrt{e_i}$ .

- (a) **Pearson residuals:**  $r_i = \frac{o_i - e_i}{\sqrt{e_i}}$ .
- (b) **Chi-squared statistic and  $r_i$ :**  $X^2 = \sum r_i^2 = \sum \frac{(o_i - e_i)^2}{e_i}$ .

- (c) **Deviance:**  $D = 2 \sum [o_i \log(o_i/e_i) - (o_i - e_i)]$ ; note that "for most models"  $\sum o_i = \sum e_i$ , so the last two terms cancel out.

- (d) **Deviance residuals:**  $d_i = \frac{o_i - e_i}{\sqrt{2o_i \log(o_i/e_i) - (o_i - e_i)}} \rightarrow D = \sum d_i^2$ .

- (e) **Likelihood ratio chi-sq statistic:**  $2[l_{current} - l_{min}]$ .

- (f) **Pseudo  $R^2$ :**  $\frac{l_{min} - l_{current}}{l_{min}}$ .

- (g) **Rate ratio:**  $\exp[\beta_i]$ . Shows, for example, that the risk of coronary death (example below) is  $\exp[\beta_i]$  times higher for smokers vs non-smokers, controlling for other factors.

## Contingency tables

### A+B

This is the standard chi-squared analysis, so use that for  $\hat{\mu}$ :

$$\hat{\mu} = e_{jk} = (y_j \cdot y_k) / n \quad (row \times col) / total \quad (29)$$

$$X^2 = \sum_{jk} \frac{(y_{jk} - e_{jk})^2}{e_{jk}} \quad \chi^2_{(J-1)(K-1)} \quad (30)$$

For three-way tables:

1. Ignore C, compute sums for A, B levels, make a two-way table.
2. Compute fitted values for the above two-way table and then partition the values equally to the two levels of C.

### A+B\*C

Algorithm:

1. Ignore A, compute sums of
 

B1	B2
C1	C1
C2	C2

2. Then compute A sums: A1, A2.

3. Use probability  $p = A1 / (A1 + A2)$  and  $1 - p$  to multiply with sums of step 1 to get fitted values.

### (A+B)\*C

Let A be the response.

1. For each level of C, find the proportions of B regardless of A.
2. Then multiply the proportions by the row sums of the A levels.

### Families checklist

1. The **Gaussian family**: identity, log and inverse.
2. The **binomial family**: logit, probit, cauchit (Cauchy CDFs) log and cloglog (complementary log-log).
3. The **Gamma family**: inverse, identity and log.
4. The **Poisson family**: log, identity, and sqrt.
5. The **inverse-gaussian family**:  $1/\mu^2$ , inverse, identity and log.
6. The **quasi family**: logit, probit, cloglog, identity, inverse, log,  $1/\mu^2$  and sqrt, and the function power can be used to create a power link function.

```

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
## the overdispersion models:
quasi(link = "identity", variance = "constant")
## or variance = "mu"
quasibinomial(link = "logit")
quasipoisson(link = "log")

```

Contingency table examples

```

M      D      N      F      D
Y      2      22     4      6
N      8      2      11     2

> counts<-c(2,22,4,6,8,2,11,2)
> G<-factor(rep(c("M", "F"), each=2,2))

> R<-factor(rep(c("Y", "N"), each=4))
> T<-factor(rep(c("N", "D"), 4))
> m1<-glm(counts~G*T, family=poisson)
> #fitted(m1)
> m2<-glm(counts~G*T+R, family=poisson)
> #fitted(m2)
> m3<-glm(counts~(T+R)*G, family=poisson)

> #fitted(m3)
>
> m4<-glm(counts~(G+R)*T, family=poisson)
> fitted(m4)

1      2      3      4      5      6      7      8
2.4 21.0 3.6 7.0 7.6 3.0 11.4 1.0
```

	$\phi$	w	$b(\theta)$	$c(y, \phi)$	$\mu = b'(\theta)$	$b''(\theta)$
Normal $Y \sim N(\theta, \phi)$	$\phi$	1	$\theta^2/2$	$-(y^2/\phi + \log(2\pi\phi))/2$	$\theta$	1
Poisson $Y \sim Po(e^\theta)$	1	1	$e^\theta$	$-\log(y!)$	$e^\theta$	$\mu$
Binomial: $nY \sim Bi(n, e^\theta/(1 + e^\theta))$	1	n	$\log(1 + e^\theta)$	$\log\binom{n}{ny}$	$e^\theta/(1 + e^\theta)$	$\mu(1 - \mu)$
Gamma $Y \sim Ga(\nu, \lambda)^\dagger$	$\phi$	1	$-\log(-\theta)$	$\nu \log \nu + (\nu - 1) \log y - \log \Gamma(\nu)$	$-1/\theta$	$\mu^2$

<sup>†</sup>pdf  $f(y) = \lambda^\nu y^{\nu-1} e^{-\lambda y} / \Gamma(\nu)$  where  $\lambda = -\theta/\phi$  and  $\nu = 1/\phi$ .