

STATISTICS 538, Lecture #5

Deviance

November 8, 2010

A closer look at deviance

Write

Saturated model
 n parameters for
 n datapoints

p parameters
↓

$$2\{\max l(\theta_1, \dots, \theta_n) - \max l(\beta)\} = \frac{D}{\phi}$$

GOF

from exp. family

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

D = Deviance

ϕ = Dispersion

D/ϕ = scaled deviance

approx

$\sim \chi^2_{n-p}$

if model is correct

Bernoulli, Poisson : $\phi = 1$

Normal : $\phi = \sigma^2$

Deviance residuals

Since deviance is like RSS... $\sim \chi^2_{n-p}$ = sum of squared normals

define i -th residual as signed square-root of i -th term in deviance

e.g. Binomial

$$r_i = \sqrt{2 \left[y_i \log(y_i/\hat{\mu}_i) + (n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right) \right]}$$

$$\times \begin{cases} +1 & \text{if } y_i > \hat{\mu}_i \\ -1 & \text{if } y_i \leq \hat{\mu}_i \end{cases}$$

hence LM diagnostics \rightarrow GLM diagnostics

Deviance residuals ex.

$n_i = 5, i=1, \dots, n=40$

```
n <- 40; x <- rnorm(n); ni <- rep(5, n)
```

y_1, \dots, y_{40}

```
y <- rbinom(n, size=ni, prob=expit(-1 + 0.75*x))
```

$\beta_0 \quad \beta_1$

```
ft <- glm(cbind(y,ni-y)~x, family=binomial)
```

```
plot(x, resid(ft, type="deviance"),ylim=4*c(-1,1))
```

```
plot(x, resid(ft,type="pearson"),ylim=4*c(-1,1))
```

$\begin{pmatrix} y_1 & n_1 - y_1 \\ \vdots & \vdots \\ y_{40} & n_{40} - y_{40} \end{pmatrix}$

Residuals versus Explanatory Variable

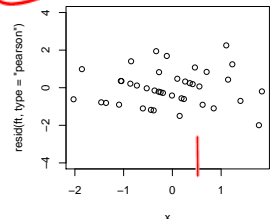
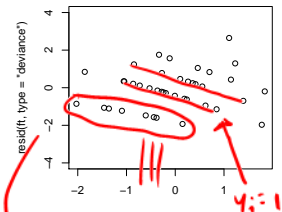
$n=40$

residuals
more
useful?

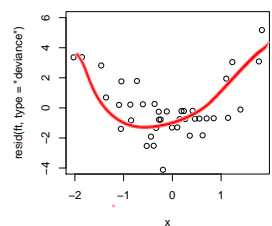
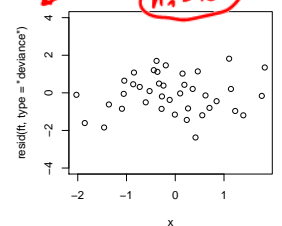
deviance
resids $n_i=5$

Pearson resids

granularity!



$n_i=15$



detects
a problem!

Deviance residuals ex., continued

```
ni <- rep(15, n)
```

```
y <- rbinom(n, size=ni, prob=expit(-1 + 0.75*x))
```

```
ft <- glm(cbind(y,ni-y)~x, family=binomial)  
plot(x, resid(ft, type="deviance"),ylim=4*c(-1,1))
```

```
y <- rbinom(n, size=ni, prob=expit(-1 + 0.75*x^2))
```

```
ft <- glm(cbind(y,ni-y)~x, family=binomial)  
plot(x, resid(ft, type="deviance"),ylim=c(-4,6))
```

*fitting a
wrong
model*

Deviance similar to Pearson's Chi-squared statistic?

YES!

$$X^2 = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \approx X_{n-p}^2 \text{ if model correct}$$

parameters

e.g. Poisson

$$D = 2 \sum_{i=1}^n \left\{ y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i) \right\}$$

$$\approx 2 \sum_{i=1}^n \left[\cancel{y_i(\hat{\mu}_i)} + \overset{w_i(y_i)}{w_i(\hat{\mu}_i)} \cancel{(y_i - \hat{\mu}_i)} + \frac{1}{2} \underbrace{w_i(\hat{\mu}_i)}_{\frac{1}{\hat{\mu}_i}} (y_i - \hat{\mu}_i)^2 \right]$$

matches ✓

Dispersion unknown?

Rough intuition: estimate it to make the scaled deviance compatible with χ_{n-p}^2 :

$$\hat{\phi} = \frac{D}{n-p}$$

*next lecture
- overdispersion
+ quasi-likelihood*

Certainly works for normal models - applicability elsewhere???

Investigate the chi-sq approx. to sampling dist. of deviance ... meaning of 'large-sample'?

$n_i = 5$

$n = 5$

```
### start with 5 trials at each of 5 design points  
x <- rep(1:5, each=5)
```

```
set.seed(17); dvres <- matrix(0, 100, 2)
```

simulate 100 datasets

```
for (lp in 1:100) {  
  y <- rbinom(length(x), size=1, prob=expit(-2+0.4*x))  
  fit.a <- glm(y~x, family="binomial")  
  fit.b <- glm(table(x,1-y)~unique(x), family="binomial")  
  dvres[lp,] <- c(fit.a$dev, fit.b$dev)  
}
```

100 Bernoulli responses

x=1
x=2

2	3
3	2
...	...

successes
failures

E.g., for first simulated dataset

```
> summary(fit.a)
```

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	5.410	2.172	2.490	0.0128	*
x	-1.280	0.545	-2.349	0.0188	*

```
---
```

```
Residual deviance: 20.198 on 23 degrees of freedom
```

```
> summary(fit.b)
```

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	5.410	2.172	2.490	0.0128	*
unique(x)	-1.280	0.545	-2.349	0.0188	*

```
---
```

```
Residual deviance: 1.7335 on 3 degrees of freedom
```

Compare actual sampling dist. and χ^2 approx

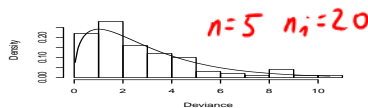
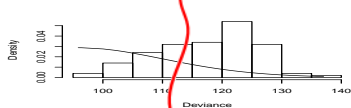
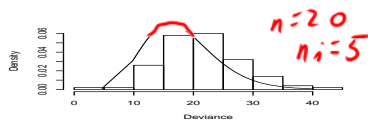
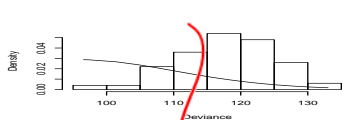
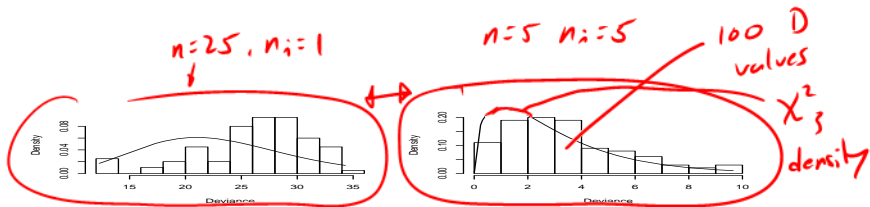
```
hist(dvres[,1], prob=T,  
     main="", xlab="Deviance", ylab="Density")  
points(sort(dvres[,1]),  
       dchisq(sort(dvres[,1]), length(x)-2), type="l")  
 $\chi^2$  density  
hist(dvres[,2], prob=T,  
     main="", xlab="Deviance", ylab="Density")  
points(sort(dvres[,2]),  
       dchisq(sort(dvres[,2]), length(unique(x))-2), type="l")
```

Consider more data in two different senses

```
### now 5 trials at each of twenty design points  
x <- rep(seq(from=1,to=5,length=20), each=5)  
...
```

```
### now 20 trials at each of 5 design points  
x <- rep(1:5, each=20)  
...
```

Simulation results



χ^2 approx. to deviance
is large n_i approx, not large n approx!