1. The midterm is Wednesday, May 12, 3:00pm to 3:50pm, in Boelter 2444.
2. HW due Fri May 7 BY EMAIL to the TA, seanwang@ucla.edu, is problem 4, parts a and b, from the handout. HW should be in PDF or plain text in an email. No lecture Fri May 7!
3. Final exam is Thur June 10, 3-6pm in Math-Science 6229.
4. GLM, continued.

4. Generalized Linear Model. GLM.

Exponential family.

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

where θ ~ mean, and φ ~ variance.

Examples. Normal.

$f(y) = 1/\sqrt{(2\pi\sigma^2)}\ \exp[-(y-\mu)^2/(2\sigma^2)].$

$\quad = \exp[\log(1/\sqrt{(2\pi\sigma^2)})$

$\qquad \times \exp[-y^2/(2\sigma^2) + 2y\mu/(2\sigma^2) - \mu^2/(2\sigma^2)]$

$= \exp[-1/2 \log(2\pi\sigma^2) - y^2/(2\sigma^2) + y\mu/\sigma^2 - \mu^2/(2\sigma^2)]$

$= \mathbf{\exp[(y\mu-\mu^2/2)/\sigma^2 - 1/2 \log(2\pi\sigma^2) - y^2/(2\sigma^2)]}$.

Let $\theta=\mu$, $\varphi=\sigma$, $a(\varphi)=\sigma^2$, $b(\theta) = \mu^2/2 = \theta^2/2$, and

$c(y,\varphi) = -1/2 \log(2\pi\varphi^2) - y^2/(2\varphi^2)$

$= -1/2\log(2\pi\sigma^2) - y^2/(2\sigma^2)$.

Now $f(y) = \exp[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi)]$.

You could also have $\varphi=\sigma^2$, as the Faraway book does on p115.

Strategy.

a) Express the density as an exponential.

b) Look for a term that is y times some function of the mean. That function of the mean should be $\theta/a(\varphi)$.

c) Anything else that depends on θ should be expressed as $-b(\theta)/a(\varphi)$.

d) Anything else needs to depend only on y and φ.

Poisson. $f(y) = e^{-\mu} \mu^y/y!$

$= e^{-\mu} \exp(\log(\mu^y/y!))$

$= \exp[-\mu + \log(\mu\text{^}y) - \log(y!)]$

$= \mathbf{\exp[-\mu + y\log(\mu) - \log(y!)].}$

Note the $y \log(\mu)$ term.

Let $\theta = \log(\mu)$, and $a(\varphi) = 1$.

$$f(y) = \exp\left[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi)\right],$$

where $b(\theta) = \exp(\theta) = \mu$.

$c(y,\varphi) = -\log(y!)$.

So it works out without φ.

This is because the mean, μ, specifies the Poisson distribution. We can simply let $\varphi = 1$.

Uniform? Let θ = center and ϕ = spread to one side.

$f(y) = 1/(2\phi)\ 1(|y-\theta| \leq \phi)$. No way to express that as an exponential family.

Try it yourself for the binomial distribution. See p116.

For a member of the exponential family,

$EY = \mu = b'(\theta)$ and $V(Y) = b''(\theta)a(\phi)$. b'' = variance function.

The link function, g, has $\eta = g(\mu)$. $\eta = x^T \beta$.

The *canonical link* is g such that $\eta = g(\mu) = \theta$. That is, $g(b'(\theta)) = \theta$.

For ordinary regression where Y is normal, for instance, $b(\theta) = \theta^2/2$, so $b'(\theta) = \theta$. So, g is the identity function.

For Poisson regression, $b(\theta) = \exp(\theta)$, so $b'(\theta) = \exp(\theta)$, so if $g(b'(\theta)) = \theta$, then g is the log function. So the canonical link is the log link.

pp.120-121 of the Faraway handout describe general forμlas for deviance for GLMs, and residuals are described on p123.

Pearson residuals are $r_p = (y-\hat{\mu})/\sqrt{V(\hat{\mu})}$, where $V(\mu) = b''(\theta)$. $\sum r_p^2 = \chi^2$, and this is usually asymptotically χ^2-distributed with p degrees of freedom, as $n \to \infty$.

On p126, Faraway suggests plotting $X=\hat{\eta}$ against Y=Pearson or deviance residuals. You should look for nonlinearity and heteroskedasticity. You can also look at partial residuals, see p128-129, and look for outliers, see p129.