Lecture 25: The Dispersion Parameter

Sometimes the exponential family is written in the form

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - B(\theta)}{\phi} + C(y,\phi)\right\}, \tag{1}$$

where $B(\cdot)$ and $C(\cdot,\cdot)$ are known functions, and the range of Y does not depend on θ or ϕ. In this formulation, we call θ the *canonical parameter*, and ϕ the *dispersion parameter*. If the distribution is parameterized in terms of the mean of Y, μ, so that $\theta \equiv g(\mu)$ for some function g, then $g(\mu)$ is the canonical link.

If ϕ is known, then (1) agrees with the usual definition of the 1-parameter exponential family in canonical form.

NOTE: For a random variable Y with distribution of the form (1),

$$\mu \equiv \mathrm{E}[Y] = B'(\theta)$$

and

$$\mathrm{Var}[Y] = B''(\theta)\phi \equiv V(\mu)\phi.$$

Here V is called the *variance function*. Thus, the variance function is equal to $B''(\theta)$ for exponential families.

NOTE: The dispersion parameter and variance function are unique only up to a constant, since

$$\mathrm{Var}[Y] = \phi V(\mu) = (c\phi)[V(\mu)/c] = \phi' V'(\mu)$$

for any constant c, where ϕ' is now considered the dispersion parameter, and $V'(\mu)$ is now considered the variance function. However, by convention, we would take both ϕ and V to be positive, since they're contributing to the variance, which is always positive.

Example: Normal distribution

The density of a random variable Y with $\mathrm{N}(\mu,\sigma^2)$ distribution can be written as

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\ &= \exp\left\{\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}[y^2/\sigma^2 + \log(2\pi\sigma^2)]\right\}. \end{aligned}$$

So, $B(\mu) = \mu^2/2$ and $C(y,\sigma) = -\frac{1}{2}[y^2/\sigma^2 + \log(2\pi\sigma^2)]$. The dispersion parameter is σ^2. The canonical parameter is μ, and $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma^2$. By convention (and in S-PLUS), we take $V(\mu) = 1$ and $\phi = \sigma^2$.

Example: Gamma distribution

The density of a random variable Y with Gamma(α, ν) distribution, $0 < \alpha, \nu, y < \infty$, can be written as

$$\begin{aligned} f_Y(y) &= \frac{y^{\nu-1}\alpha^{\nu}e^{-y\alpha}}{\Gamma(\nu)} \\ &= \exp\{-y\alpha + \nu\log\alpha + (\nu-1)\log y - \log\Gamma(\nu)\} \\ &= \exp\left\{\frac{y(-\alpha/\nu) - [-\log\alpha]}{1/\nu} + (\nu-1)\log y - \log\Gamma(\nu)\right\}. \end{aligned}$$

Letting $\theta \equiv -\alpha/\nu$ and $\phi \equiv 1/\nu$,

$$f_Y(y) = \exp\left\{\frac{y\theta - [-\log(-\theta)]}{\phi} - \log(\phi)/\phi + (1/\phi - 1)\log y - \log\Gamma(1/\phi)\right\}$$

Therefore, the Gamma distribution is in the exponential family with $B(\theta) = -\log(-\theta)$ and dispersion parameter $\phi \equiv 1/\nu$. This definition of ϕ is conventional, and is used by S-PLUS. Since

$$\mu \equiv \mathrm{E}[Y] = B'(\theta) = -\frac{1}{\theta}$$

and

$$\mathrm{Var}[Y] = B''(\theta)\phi = \frac{\phi}{\theta^2} = \phi\mu^2,$$

this definition then implies that $V(\mu) = \mu^2$.

The easiest way to find the canonical link for the Gamma distribution is to parameterize it in terms of its mean. We can compute Thus,

$$f_Y(y) = \exp\left\{\frac{y\left(-\frac{1}{\mu}\right) - \log(\mu)}{\phi} - \log(\phi)/\phi + (1/\phi - 1)\log y - \log\Gamma(1/\phi)\right\}.$$

Therefore, the canonical link function is $g(\mu) = -\frac{1}{\mu}$.

Example: Poisson distribution

The distribution of a random variable Y with a Poisson(λ) distribution can be written as

$$\begin{aligned} f_Y(y) &= \frac{e^{-\lambda}\lambda^y}{y!} \\ &= \exp\{y\log\lambda - \lambda - \log y!\} \\ &= \exp\left\{\frac{y\log\lambda - \lambda}{1} - \log y!\right\}. \end{aligned}$$

So, the canonical parameter is $\log\lambda$. In this case, the dispersion parameter is 1. It is for this reason that S-PLUS gives the message

```
(Dispersion Parameter for Poisson family taken to be 1 )
```

in the summary output when we fit a Poisson GLM.

Example: Binomial distribution

The distribution of a random variable Y with Binomial(n,p) distribution can be written as

$$\begin{aligned} f_Y(y) &= \begin{pmatrix} n \\ y \end{pmatrix} p^y(1-p)^{n-y} \\ &= \exp\left\{ y\log\left(\frac{p}{1-p}\right) + n\log(1-p) + \log\begin{pmatrix} n \\ y \end{pmatrix} \right\} \\ &= \exp\left\{ \frac{y\log\left(\frac{p}{1-p}\right) + n\log(1-p)}{1} + \log\begin{pmatrix} n \\ y \end{pmatrix} \right\}. \end{aligned}$$

So, the canonical parameter is $\log\left(\frac{p}{1-p}\right)$. Like in the Poisson case, in this case, the dispersion parameter is also 1. It is for this reason that S-PLUS gives the message

```
(Dispersion Parameter for Binomial family taken to be 1 )
```

in the summary output when we fit a Binomial GLM.

Extending the GLM Framework to Allow for a Dispersion Parameter

In most situations, ϕ is unknown. In this case, the distribution (1) does not fit into the usual GLM framework (where we assume that the distribution is in the 1-parameter exponential family). However, it turns out that we can still use the GLM framework to model observations with distribution (1). In particular, we treat ϕ as known and common to all observations. Then, we let μ (and hence θ) vary among observations in the usual way, i.e. for observations $Y_1, \ldots, Y_n$, we assume that

$$g(\mu_i) = \sum_{j=1}^{p} x_{ij}\beta_j.$$

The difference between this model and usual GLMs is that, in addition to estimating the β_j's, we will now need to estimate ϕ as well. We can obtain such an estimate using the Pearson chi-squared statistic and its asymptotic properties.

Estimating the Dispersion Parameter

We have defined the Pearson chi-squared statistic in the special case where Y_i has a Poisson or binomial distribution. In general, this statistic is defined as

$$X^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)},$$

where $\text{Var}[Y_i] = V(\mu_i)\phi$.

Exercise: Check that this definition is consistent with that given for Poisson or binomial data.

The *scaled* Pearson chi-squared statistic is defined as

$$X_s^2 = \frac{X^2}{\phi}.$$

It turns out that, if the model is specified correctly,

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

asymptotically, where n is the sample size and p is the number of unknown regression coefficients (the β_j's) in the model.

Since the mean of a χ^2_{n-p} random variable is $n-p$, we can use the approximation $X_s^2 \approx n-p$, and hence the estimator

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

Note that this is *not* the MLE of ϕ (it is actually a *moment estimator*). However, it has some nice properties not shared by the MLE.

Example: Normal distribution (cont.)

For independent observations $Y_1, \ldots, Y_n$ with $Y_i \sim \text{N}(\mu_i, \sigma^2)$,

$$\begin{aligned} \text{Var}[Y_i] &= V(\mu_i)\phi \\ &\equiv 1 \cdot \sigma^2 \end{aligned}$$

so $V(\mu_i) \equiv 1$ and

$$X^2 = \sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2.$$

Therefore,

$$\hat{\phi} = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2}{n-p}.$$

This is the usual unbiased estimator of σ^2. (The MLE, which has n rather than $n-p$ in the denominator, is biased.)

Example: Gamma distribution (cont.)

For independent observations $Y_1, \ldots, Y_n$ with $Y_i \sim$Gamma(θ_i, ν),

$$\begin{aligned}
\text{Var}[Y_i] &= B''(\theta)\phi \\
&= \frac{\phi}{\theta^2} \\
&\equiv \mu_i^2 \phi
\end{aligned}$$

so $V(\mu_i) \equiv \mu_i^2$ and

$$X^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2}.$$

Therefore,

$$\hat{\phi} = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 (n-p)}.$$

Since moment estimators also have the invariance property (like MLEs), we can estimate ν by

$$\hat{\nu} = \frac{1}{\hat{\phi}}.$$