STATS 305B: Methods for Applied Statistics I **Winter 2016/17**

Lecture 1: January 9

Lecturer: Jonathan Taylor *Scribes: Kenneth Tay*

1.1 Distributions for Categorical Data

In this lecture, if π is the parameter for the family of distributions, $f(y \mid \pi)$ will refer to the density of y given π, while $L(\pi \mid y)$ refers to the likelihood function of π given y.

The 3 distributions that are commonly used in categorical data are the *binomial*, *multinomial* and *Poisson* distributions.

1.1.1 Poisson-Multinomial Connection

Theorem 1.1 *Suppose $1 \leq i \leq k$, and Y_i's are independent random variables such that $Y_i \sim Poisson(\lambda_i)$. Then*

$$(Y_1, \dots, Y_k)\Big| \sum_{i=1}^k Y_i \sim \textit{Multinom}\left(\sum_{i=1}^k Y_i, \left(\frac{\lambda_1}{\sum \lambda_i}, \dots, \frac{\lambda_k}{\sum \lambda_i}\right)\right).$$

1.2 Exponential Families

An **exponential family** is a parametrized collection of distributions, denoted by P_γ. (γ could be one- or multi-dimensional, and is called the "natural parameter" of the family.) To define an exponential family, we need:

1. A reference measure/distribution m on $\mathbb{R}^k$, and
2. A sufficient statistic T, which is a function from Ω, the space which the P_γ live on, to $\mathbb{R}^k$.

We can define the exponential family by specifying its density w.r.t. to the the reference measure:

$$\frac{dP_\gamma}{dm}(\omega) \propto \exp(\gamma \cdot T(\omega)), i.e. \frac{dP_\gamma}{dm}(\omega) = \frac{\exp(\gamma \cdot T(\omega))}{\int_\Omega \exp(\gamma \cdot T(\omega'))m(d\omega')},$$

where the denominator of the last fraction is just the integrating constant to ensure that P_γ is a probability distribution. If we denote the denominator by $\exp[\Lambda(\gamma)]$, then we have a simpler expression:

$$\frac{dP_\gamma}{dm}(\omega) = \exp[\gamma \cdot T(\omega) - \Lambda(\gamma)].$$

In exponential families, $\Lambda(\cdot)$ is a convex function. This makes computation a lot easier.

1.2.1 Binomial as an Exponential Family

Let $\bar{m}$ be the counting measure on $\{0, \dots, n\}$. Define the reference measure m by its density relative to $\bar{m}$:

$$\frac{dm}{d\bar{m}}(y) = \binom{n}{y}.$$

Then, if we take the sufficient statistic $T(y) = y$, we obtain densities

$$\frac{dP_\gamma}{d\bar{m}}(y) = \frac{\binom{n}{y} e^{\gamma y}}{\sum_{j=0}^{n} \binom{n}{j} e^{\gamma j}},$$

i.e. $P_\gamma \sim \text{Binom}\left(n, \frac{e^\gamma}{1+e^\gamma}\right)$. Thus, the canonical link function (i.e. to get from the usual binomial parameter π to γ) is $\gamma(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.

1.2.2 Poisson as an Exponential Family

Define the reference measure $m(A) = \sum_{j\geq 0} \frac{1_A(j)}{j!}$, and define sufficient statistic $T(y) = y$. Then if we have $\frac{dP_\gamma}{dm}(y) \propto \exp(\gamma y)$, we can compute the integrating constant

$$e^{\Lambda(\gamma)} = \sum_{j\geq 0} \frac{e^{\gamma y}}{y!} = e^{e^\gamma}.$$

Putting it altogether, this implies that $P_\gamma \sim \text{Poisson}(e^\gamma)$. Thus, the canonical link function (i.e. to get from the usual Poisson parameter λ to γ) is $\gamma(\lambda) = \log \lambda$.

1.2.3 Regression with Poisson Observations

Suppose Y_i are count data, with $Y_i \overset{ind.}{\sim} \text{Poisson}(\lambda_i) = \text{Poisson}(e^{\gamma_i})$. Then the negative log-likelihood is given by

$$\begin{aligned} -\log L(\gamma \mid Y) &= \sum_{i=1}^{n} \Lambda(\gamma_i) - \gamma_i y_i \\ &= \sum_{i=1}^{n} e^{\gamma_i} - \gamma_i y_i \end{aligned}$$

The natural regression model would be to have $\gamma_i = x_i^T \beta$. With this choice, the negative log-likelihood can be viewed as a function of β:

$$-\log L(\beta \mid Y) = \sum_{i=1}^{n} e^{x_i^T \beta} - \left(x_i^T \beta\right) y_i.$$

This is convex in β, and hence has computational advantages.