

Unit 4: Generalized linear models: General theory

Eugene Katsevich

October 26, 2021

Units 1-3 focused on the most common class of models used in applications: linear models. Despite their versatility, linear models do not apply in all situations. In particular, they are not designed to deal with binary or count responses. In Unit 4, we introduced *generalized linear models* (GLMs), a generalization of linear models that encompasses a wide variety of incredibly useful models including logistic regression and Poisson regression.

We'll start Unit 4 by introducing exponential family models (Section 1), a generalization of the Gaussian distribution that serves as the backbone of GLMs. Then we formally define a GLM, demonstrating logistic regression and Poisson regression as special cases (Section 2). Next we discuss maximum likelihood inference in GLMs (Section 3). Finally, we discuss how to carry out statistical inference in GLMs (Section 4).

1 Exponential family distributions

Definition and examples. Let's start with the Gaussian distribution, taking variance $\sigma^2 = 1$ for simplicity. If $y \sim N(\mu, 1)$, then it has density

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \mu)^2\right) = \exp\left(\mu y - \frac{1}{2}\mu^2\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right). \quad (1)$$

Here is a way of generalizing this density:

$$f_\eta(y) = \exp(\eta y - \psi(\eta))h(y). \quad (2)$$

Here η is called the *natural parameter*, ψ is called the *log-partition function*, and h is called the *base measure*. The distribution with density f_η is called a *one-parameter natural exponential family*. Therefore, $y \sim N(\mu, 1)$ is in the exponential family with

$$\eta = \mu, \quad \psi(\eta) = -\frac{1}{2}\eta^2, \quad h(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right). \quad (3)$$

Several other well-known distributions are in the exponential family as well. For example, consider $y \sim \text{Ber}(\pi)$. Then, we have

$$f(y) = \pi^y(1 - \pi)^{1-y} = \exp\left(y \log \frac{\pi}{1 - \pi} + \log(1 - \pi)\right). \quad (4)$$

Therefore, we have $\eta = \log \frac{\pi}{1 - \pi}$, so that $\log(1 - \pi) = -\log(1 + e^\eta)$. It follows that

$$\eta = \log \frac{\pi}{1 - \pi}, \quad \psi(\eta) = \log(1 + e^\eta), \quad h(y) = 1. \quad (5)$$

As another example, consider the Poisson distribution $y \sim \text{Poi}(\mu)$. We have

$$f(y) = e^{-\mu} \frac{\mu^y}{y!} = \exp(y \log \mu - \mu) \frac{1}{y!}. \quad (6)$$

Therefore, we have $\eta = \log \mu$, so that $\mu = e^\eta$. It follows that

$$\eta = \log \mu, \quad \psi(\eta) = e^\eta, \quad h(y) = \frac{1}{y!}. \quad (7)$$

Moments of exponential family distributions. It turns out that the derivatives of the log-partition function ψ give the moments of y . Indeed, let's start with the relationship

$$\int f_\eta(y) dy = \int \exp(\eta y - \psi(\eta)) h(y) dy = 1. \quad (8)$$

Differentiating in η and interchanging the derivative and the integral, we obtain

$$0 = \frac{d}{d\eta} \int f_\eta(y) dy = \int (y - \dot{\psi}(\eta)) f_\eta(y) dy, \quad (9)$$

from which it follows that

$$\dot{\psi}(\eta) = \int \dot{\psi}(\eta) f_\eta(y) dy = \int y f_\eta(y) dy = \mathbb{E}_\eta[y] \equiv \mu_\eta. \quad (10)$$

Thus, the first derivative of the log partition function is the mean of y . Differentiating again, we get

$$\ddot{\psi}(\eta) = \int y(y - \dot{\psi}(\eta)) f_\eta(y) dy = \int y(y - \mu_\eta) f_\eta(y) dy = \int (y - \mu_\eta)^2 f_\eta(y) dy = \text{Var}_\eta[y]. \quad (11)$$

Thus, the second derivative of the log-partition function is the variance of y .

Relationship between mean and natural parameter. The log-partition function ψ induces a connection (10) between the natural parameter η and the mean μ . Because

$$\frac{d\mu}{d\eta} = \frac{d}{d\eta} \dot{\psi}(\eta) = \ddot{\psi}(\eta) = \text{Var}_\eta[y] > 0, \quad (12)$$

it follows that μ is a strictly increasing function of η , so in particular the mapping between μ and η is bijective. Therefore, we can think of equivalently parameterizing the distribution via μ or η . In the context of GLMs (see Section 2), the mean-variance relationship is quantified in terms of the *canonical link function* g , which maps the mean to the natural parameter:

$$\eta = \dot{\psi}^{-1}(\mu) \equiv g(\mu). \quad (13)$$

Relationship between mean and variance. Note that the mean of an exponential family distribution determines its variance (since it determines the natural parameter η). For example, a Poisson random variable with mean μ has variance μ and a Bernoulli random variable with mean μ has variance $\mu(1 - \mu)$. The mean-variance relationship turns out to characterize the exponential family distribution, i.e. an exponential family distribution with mean equal to its variance is the Poisson distribution.

2 Generalized linear models and examples

In this class, the focus is on building models that tie a vector of predictors (\mathbf{x}_{i*}) to a response y_i . For linear regression, the mean of y was modeled as a linear combination of the predictors $\mathbf{x}_{i*}^T \boldsymbol{\beta}$: $\mu = \mathbf{x}_{i*}^T \boldsymbol{\beta}$. Typically, the “right” thing to do is to model the response linearly on the scale of the natural parameter η rather than on the scale of the mean parameter μ . It just happens for linear models (where the underlying distribution is Gaussian) that these two parameters coincide.

Definition. We define $\{(y_i, \mathbf{x}_{i*})\}_{i=1}^n$ as following a generalized linear model based on the exponential family f_η if

$$y_i \stackrel{\text{ind}}{\sim} f_{\eta_i}, \quad \eta_i = \mathbf{x}_{i*}^T \boldsymbol{\beta}. \quad (14)$$

GLMs are often written in terms of their link functions g , which relate the mean of y to the linear predictor $\mathbf{x}_{i*}^T \boldsymbol{\beta}$. When modeling the natural parameter as a linear function in the predictors, as in the definition (14), we get a GLM with *canonical link function* $g = \dot{\psi}^{-1}$:

$$g(\mathbb{E}[y_i]) = \dot{\psi}^{-1}(\mathbb{E}[y_i]) = \mathbf{x}_{i*}^T \boldsymbol{\beta}. \quad (15)$$

Examples. For example, *logistic regression* is the GLM based on the Bernoulli distribution:

$$y_i \stackrel{\text{ind}}{\sim} \text{Ber}(\pi_i); \quad \eta_i = \log \frac{\pi_i}{1 - \pi_i} = \mathbf{x}_{i*}^T \boldsymbol{\beta}. \quad (16)$$

Thus the canonical link function for logistic regression is the *logistic link function* $g(\mu) = \log \frac{\mu}{1-\mu}$. As another example, *Poisson regression* is the GLM based on the Poisson distribution:

$$y_i \stackrel{\text{ind}}{\sim} \text{Poi}(\mu_i); \quad \eta_i = \log \mu_i = \mathbf{x}_{i*}^T \boldsymbol{\beta}. \quad (17)$$

Thus the canonical link function for Poisson regression is the *log link function* $g(\mu) = \log \mu$.

3 Maximum likelihood estimation in GLMs

GLM normal equations. Recall that the least squares estimate $\hat{\boldsymbol{\beta}}$ is also the maximum likelihood estimate. For general GLMs, we also estimate $\boldsymbol{\beta}$ via maximum likelihood. To derive this estimates, let's write down the GLM likelihood and then take a derivative.

Iteratively reweighted least squares. TBD.

4 Inference in GLMs

TBD.