BIOS 7345: Advanced Regression Analysis I

Andrew J. Spieker, Ph.D.

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 12: Exponential families

Version: 09/17/2023

Table of Contents

Recall:

- Suppose we're willing to make parametric assumptions about Y. Under suitable regularity conditions, theory justifies this procedure:
  1. Determine the likelihood function:

$$\mathcal{L}(\boldsymbol{\theta}; Y_1, \dots, Y_N) = \prod_{i=1}^{N} f(Y_i; \boldsymbol{\theta})$$

  2. Determine the log-likelihood:

$$\ell(\boldsymbol{\theta}; Y_1, \dots, Y_N) = \log \mathcal{L}(\boldsymbol{\theta}; Y_1, \dots, Y_N) = \sum_{i=1}^{N} \log f(Y_i; \boldsymbol{\theta})$$

  3. Determine $\boldsymbol{\theta}$ such that observed data had maximal probability:

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}; Y_1, \dots, Y_N) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}; Y_1, \dots, Y_N) = \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(Y_i; \boldsymbol{\theta}) \overset{\text{SET}}{=} \mathbf{0}.$$

Recall:

- When we set the derivative of the log-likelihood equal to zero, we've created a set of unbiased estimating equations (called the *score* equations in likelihood world) by the following extremely useful fact:

$$\mathrm{E}\left[\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}_0; Y)\right] = \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}; Y)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] = \mathbf{0},$$

  where $\boldsymbol{\theta}_0$ marks the true value of the parameter.
- The mean of the individual contributions to the log-likelihood will converge to zero.
- The particular solution to the score equations in a data set is the maximum likelihood estimate, $\widehat{\boldsymbol{\theta}}$.

Recall:

- Further, under (even more) suitable regularity conditions that we won't concern ourselves with in this course, we have:

$$\text{Var}[\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}_0; Y)] = \text{E}[\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}_0; Y)\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}_0; Y)^T] = -\text{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(Y; \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right]$$

Moving forward:

- Regression involving a very nice family of distributions will allow us to form a unifying theory for generalized linear models.
- From here on, we will assume that all of the regularity conditions necessary for the prior results to hold are satisfied.

Table of Contents

Ideas: A very nice family of distributions

- Suppose that Y has a density function that can be written in the following very nice form:

$$f_Y(y;\theta,\phi) \quad = \quad \exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right].$$

- We term θ the "natural parameter" or "canonical parameter."
- We term ϕ the "nuisance parameter."

Example: Normal distribution

- If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then:

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

- The natural parameter is given by $\theta = \mu$.
- The nuisance parameter is given by $\phi = \sigma^2$.

Example: Bernoulli distribution

- If $Y \sim \text{Bernoulli}(p)$, then:

$$\begin{aligned} f_Y(y;\mu,\sigma) &= p^y(1-p)^{1-y} \\ &= \vdots \\ &= \exp\left[y\log\left(\frac{p}{1-p}\right) - \log\left(\frac{1}{1-p}\right)\right]. \end{aligned}$$

- The natural parameter is given by $\theta = \text{logit}(p)$.
- The "nuisance parameter" is given by $\phi = 1$.

Example: Poisson distribution

- If $Y \sim \text{Poisson}(\lambda)$, then:

$$
\begin{aligned}
f_Y(y;\lambda) &= \frac{\lambda^y}{y!}\exp(-\lambda) \\
&= \vdots \\
&= \exp(y\log(\lambda) - \lambda - \log(y!))
\end{aligned}
$$

- The natural parameter is given by $\theta = \log(\lambda)$.
- The "nuisance parameter" is given by $\phi = 1$.

Other examples:

- Binomial distribution.
- Exponential distribution.
- Gamma distribution.
- Inverse Gaussian distribution.

Non-examples:

- Weibull distribution.
- Uniform distribution.
- Beta distribution.
- t-distribution.

Beautiful math:

- If Y has a density function in the "nice form" given by:

$$f_Y(y;\theta,\phi) \;=\; \exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right].$$

- Then the log-likelihood for a single observation is given by:

$$\ell(\theta,\phi;Y) \;=\; \frac{Y\theta - b(\theta)}{\phi} + c(Y,\phi).$$

- The score for a single observation is given by:

$$\dot{\boldsymbol{\ell}}(\theta,\phi;Y) \;=\; \begin{bmatrix} \frac{\partial}{\partial\theta}\ell(\theta,\phi;Y) \\ \\ \frac{\partial}{\partial\phi}\ell(\theta,\phi;Y) \end{bmatrix} = \begin{bmatrix} \frac{Y-b'(\theta)}{\phi} \\ \\ \ldots \end{bmatrix}.$$

Beautiful math:

- Because $\mathrm{E}[\dot{\boldsymbol{\ell}}(\theta_0, \phi_0; Y)] = \mathbf{0}$ (i.e., element-wise), we find that

$$\mathrm{E}\left[\frac{Y - b'(\theta_0)}{\phi_0}\right] = 0 \Rightarrow \mathrm{E}[Y] = b'(\theta_0).$$

- Key point the mean of Y depends upon θ, but *not* on the nuisance parameter!

Beautiful math: Variance (method 1)

- Now, note that

$$\begin{aligned}
\text{Var}[\dot{\boldsymbol{\ell}}(\theta_0, \phi_0; Y)] &= \text{E}[\dot{\boldsymbol{\ell}}(\theta_0, \phi_0; Y)\dot{\boldsymbol{\ell}}(\theta_0, \phi_0; Y)^T] \\
&= \begin{bmatrix} \text{E}\left[\left(\frac{Y - b'(\theta_0)}{\phi_0}\right)^2\right] & \dots \\ \dots & \dots \end{bmatrix}
\end{aligned}$$

- We don't really care about the other three entries, though we could determine them.

Beautiful math: Variance (method 2)

- But on the other hand, note that (letting $\boldsymbol{\theta} = (\theta, \phi)$,

$$\text{Var}[\dot{\boldsymbol{\ell}}(\theta_0, \phi_0; Y)] = -\text{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(Y; \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right]$$

$$= \begin{bmatrix} \frac{b''(\theta_0)}{\phi_0} & \cdots \\ \cdots & \cdots \end{bmatrix}$$

- We don't really care about the other three entries, though we could determine them.

Beautiful math: Variance (method 2)

- Equating the upper-left entries (this step is legit given the regularity conditions of the exponential family), we find that:

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{Y - b'(\theta_0)}{\phi_0}\right)^2\right] &= \frac{b''(\theta_0)}{\phi_0} \\
\Rightarrow \frac{1}{\phi_0^2}\mathrm{E}[(Y - b'(\theta_0))^2] &= \frac{b''(\theta_0)}{\phi_0} \\
\Rightarrow \mathrm{E}[(Y - b'(\theta_0))^2] &= \phi_0 b''(\theta_0) \\
\Rightarrow \mathrm{E}[(Y - \mathrm{E}[Y])^2] &= \phi_0 b''(\theta_0) \\
\Rightarrow \mathrm{Var}[Y] &= \phi_0 b''(\theta_0).
\end{aligned}$$

Mean-variance relationship:

- The variance is given by: $\mathrm{Var}[Y] = \phi_0 b''(\theta_0) \propto b''(\theta_0)$.
- However, recall that $\mathrm{E}[Y] = b'(\theta_0)$, and so $b''(\theta_0)$ marks the relationship between the mean and the variance of Y.

Example: Normal distribution

- If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$f_Y(y; \mu, \sigma) \quad = \quad \exp\left[\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}(y^2/\sigma^2 - \log(2\pi\sigma^2))\right].$$

- The natural parameter is given by $\theta = \mu$.
  - Note: $\mu^2/2 = \theta^2/2 =: b(\theta)$.
- $\phi = \sigma^2$ is a nuisance parameter.
- Therefore, $\mathrm{E}[Y] = b'(\theta) = \theta = \mu$.
- Further, $\mathrm{Var}[Y] = \phi b''(\theta) = \phi \times 1 = \sigma^2$.

Example: Bernoulli distribution

- If $Y \sim \text{Bernoulli}(p)$, then:

$$f_Y(y;\mu,\sigma) \;=\; \exp\left[y\log\left(\frac{p}{1-p}\right) - \log\left(\frac{1}{1-p}\right)\right].$$

- The natural parameter is given by $\theta = \text{logit}(p) \Rightarrow p = \text{expit}(\theta)$.
  - Note: $\log\left(\frac{1}{1-p}\right) = \log\left(\frac{1}{1-\text{expit}(\theta)}\right) = \log(1+\exp(\theta)) =: b(\theta)$.
- The "nuisance parameter" is given by $\phi = 1$.
- Therefore, $\text{E}[Y] = b'(\theta) = \text{expit}(\theta) = p$.
- Further, $\text{Var}[Y] = \phi b''(\theta) = \text{expit}(\theta)[1-\text{expit}(\theta)] = p(1-p)$.

Example: Poisson distribution

- If $Y \sim \text{Poisson}(\lambda)$, then:

$$f_Y(y; \lambda) \quad = \quad \exp(y \log(\lambda) - \lambda - \log(y!))$$

- The natural parameter is given by $\theta = \log(\lambda) \Rightarrow \lambda = \exp(\theta)$.
    - Note: $\lambda = \exp(\theta) =: b(\theta)$.
- The "nuisance parameter" is given by $\phi = 1$.
- Therefore, $\text{E}[Y] = b'(\theta) = \exp(\theta) = \lambda$.
- Further, $\text{Var}[Y] = \phi b''(\theta) = \exp(\theta) = \lambda$.

Where we're headed:

- Suppose that we wish to pose a regression model for outcomes in these nice exponential families.
- Specifically, we want to propose a form for:
  - The family of distributions to which Y belongs.
  - The mean of Y (given X).
- Are some choices "nicer" than others?
- How do we estimate regression parameters having no closed-form solution?
- How do we estimate the variance?
- How do we test hypotheses?

Summary

So far:

- Exponential families.

Summary

Up next:

- Generalized linear models.