Lecture 6: GLM Theory — Moments and Model Structure

MATH3823 Generalised Linear Models

Richard P Mann

MATH3823 Generalised Linear Models

Reading

Course notes: Chapter 3, Sections 3.4–3.6

`www.richardpmann.com/MATH3823`

Recall: Exponential Family Form

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

Last lecture: We derived this form for Poisson, Binomial, and Normal.

Today:

1. Derive moments from $b(\theta)$
2. The systematic component (linear predictor)
3. The link function

Moments of Exponential Family Distributions

Proposition

For a distribution in exponential family form:

$$\boxed{\mathbb{E}[Y] = b'(\theta), \qquad \mathrm{Var}[Y] = b''(\theta) \cdot \phi}$$

Key insight: The mean and variance are determined entirely by derivatives of $b(\theta)$.

This is why the exponential family is so powerful — once we know $b(\theta)$, we know the moments.

Proof Sketch: Deriving the Mean

Starting point: All probability densities integrate to 1:

$$\int f(y;\theta,\phi)\,dy = 1$$

Differentiate both sides w.r.t. θ:

$$\frac{\partial}{\partial\theta}\int f(y)\,dy = 0$$

Move derivative inside:

$$\int \frac{\partial f}{\partial\theta}\,dy = \int f(y)\cdot\frac{y-b'(\theta)}{\phi}\,dy = 0$$

Rearranging:

$$\mathbb{E}[Y] = b'(\theta)$$

Proof Sketch: Deriving the Variance

Differentiate again w.r.t. θ:

Taking the second derivative of $\int f(y)\,dy = 1$ and simplifying gives:

$$\mathbb{E}\left[\left(\frac{Y - b'(\theta)}{\phi}\right)^2\right] = \frac{b''(\theta)}{\phi}$$

Since $\mathbb{E}[Y] = b'(\theta)$:

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

Therefore:

$$\mathrm{Var}[Y] = b''(\theta) \cdot \phi$$

Verifying Moments: Poisson

Recall: $\theta = \log \lambda$, $b(\theta) = e^\theta$, $\phi = 1$

Mean:

$$\mathbb{E}[Y] = b'(\theta) = \frac{d}{d\theta} e^\theta = e^\theta = \lambda \quad \checkmark$$

Variance:

$$\mathrm{Var}[Y] = b''(\theta) \cdot \phi = e^\theta \cdot 1 = \lambda \quad \checkmark$$

Note: For Poisson, $\mathbb{E}[Y] = \mathrm{Var}[Y] = \lambda$ (equidispersion).

Verifying Moments: Binomial

Recall: $\theta = \operatorname{logit}(p)$, $b(\theta) = m\log(1+e^\theta)$, $\phi = 1$

Mean:

$$b'(\theta) = m \cdot \frac{e^\theta}{1+e^\theta} = mp \quad \checkmark$$

Variance:

$$\begin{aligned} b''(\theta) &= m \cdot \frac{e^\theta \cdot (1+e^\theta) - e^\theta \cdot e^\theta}{(1+e^\theta)^2} = m \cdot \frac{e^\theta}{(1+e^\theta)^2} \\ &= mp(1-p) \quad \checkmark \end{aligned}$$

Note: Variance depends on mean — this is the *mean-variance relationship*.

Verifying Moments: Normal

Recall: $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^2$

Mean:

$$\mathbb{E}[Y] = b'(\theta) = \frac{d}{d\theta}\frac{\theta^2}{2} = \theta = \mu \quad \checkmark$$

Variance:

$$\mathrm{Var}[Y] = b''(\theta) \cdot \phi = 1 \cdot \sigma^2 = \sigma^2 \quad \checkmark$$

Note: For Normal, variance is independent of mean.

The Variance Function

Define: The **variance function** $V(\mu)$ by:

$$\mathrm{Var}[Y] = V(\mu) \cdot \phi$$

Distribution	Mean μ	Variance function $V(\mu)$
Poisson	λ	μ
Binomial	mp	$\mu(1 - \mu/m)$
Normal	μ	1
Gamma	α/λ	μ^2

The variance function characterizes how spread changes with the mean.

Summary: Exponential Family Moments

Distribution	θ	$b(\theta)$	ϕ	$\mathbb{E}[Y]$	$\mathrm{Var}[Y]$
Poisson	$\log \lambda$	e^θ	1	λ	λ
Binomial	$\mathrm{logit}(p)$	$m \log(1 + e^\theta)$	1	mp	$mp(1-p)$
Normal	μ	$\theta^2/2$	σ^2	μ	σ^2

Key formula: $\mathbb{E}[Y] = b'(\theta)$, $\mathrm{Var}[Y] = b''(\theta)\phi$

The Systematic Component

Now we connect the distribution to covariates.

Linear predictor:

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} = \mathbf{x}_i' \boldsymbol{\beta}$$

In matrix form:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

where:

- $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'$ is the $n \times 1$ vector of linear predictors
- $\mathbf{X}$ is the $n \times p$ design matrix
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the parameter vector

The Link Function

Problem: The mean μ may be constrained:

- Poisson: $\mu > 0$
- Binomial: $0 < \mu/m < 1$

But: The linear predictor $\eta = \mathbf{x}'\boldsymbol{\beta}$ is unconstrained.

Solution: Use a **link function** g such that:

$$\eta = g(\mu) \in \mathbb{R}$$

The inverse $h = g^{-1}$ maps η back to the valid range of μ:

$$\mu = g^{-1}(\eta) = h(\eta)$$

The Canonical Link

Definition: The **canonical link** sets $\eta = \theta$ (the natural parameter).

Since $\mathbb{E}[Y] = \mu = b'(\theta)$, we have:

$$g(\mu) = (b')^{-1}(\mu) = \theta$$

Distribution	Canonical θ	Canonical Link $g(\mu)$	Name
Normal	μ	μ	Identity
Poisson	$\log \lambda$	$\log(\mu)$	Log
Binomial	$\text{logit}(p)$	$\log \frac{\mu/m}{1-\mu/m}$	Logit
Gamma	$-1/\mu$	$-1/\mu$	Reciprocal

Why Use the Canonical Link?

Advantages:

1. **Sufficient statistics:**

$$S_j = \sum_{i=1}^{n} y_i x_{ij}$$

contain all information about β_j.

2. **Simple score equations:** MLE satisfies

$$\sum_{i=1}^{n} (y_i - \hat{\mu}_i) x_{ij} = 0$$

3. **Computational efficiency:** Algorithms converge reliably.

However: Non-canonical links may be more interpretable or provide a better fit.

GLM Summary Diagram

$$\begin{tikzcd}[column sep=huge]
\boxed{\mathbf{X}} \arrow[r, "\eta = \mathbf{X}\boldsymbol{\beta}"] & \boxed{\text{Linear } \eta} \arrow[r, "\mu = g^{-1}(\eta)"] & \boxed{\text{Mean } \mu} \arrow[r, "Y \sim f(\mu)"] & \boxed{\text{Response } Y} \\
\text{Predictors} & \text{Systematic} & \text{Link} & \text{Random}
\end{tikzcd}$$

Flow:

$$\text{Predictors} \xrightarrow{\text{linear model}} \eta \xrightarrow{\text{inverse link}} \mu \xrightarrow{\text{distribution}} Y$$

Normal Linear Model as a GLM

Random component: $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$

Systematic component: $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$

Link function: Identity link, $g(\mu) = \mu$

So:

$$\mu_i = \eta_i = \mathbf{x}_i'\boldsymbol{\beta}$$

This is exactly the normal linear model we studied earlier.

$\Rightarrow$ Normal regression is a *special case* of GLM.

Summary

Key points:

- Moments from $b(\theta)$: $\mathbb{E}[Y] = b'(\theta)$, $\text{Var}[Y] = b''(\theta)\phi$
- Variance function $V(\mu)$ describes mean-variance relationship
- Systematic component: linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$
- Link function: $\eta = g(\mu)$ maps mean to linear predictor
- Canonical link: $g(\mu) = \theta$ (natural parameter)
- Normal regression is a GLM with identity link

Next lecture: More on link functions — logit, probit, complementary log-log.