



LEARN  
INTERACT  
GROW

# Chapter 5: Generalized Linear Models

by Curtis Gary Dean, FCAS, MAAA, CFA

Ball State University: Center for Actuarial Science and Risk Management

# **My Interest in Predictive Modeling**

---

- ❖ **1989 article in *Science***
- ❖ **“Clinical Versus Actuarial Judgment”**
- ❖ **Summarized in 1990 in *Contingencies***

# “Clinical Versus Actuarial Judgment”

---

- ❖ *“In the clinical method the decision-maker combines or processes information in his or her head.”*
- ❖ *“In the actuarial or statistical method the human judge is eliminated and conclusions rest solely on empirically established relations between data and the condition or event of interest.”*

# “Clinical Versus Actuarial Judgment”

---

- ❖ *“...with a sample of about 100 studies and the same outcome obtained in almost every case, it is reasonable to conclude that the actuarial advantage is not exceptional but general and likely encompasses many of the unstudied judgment tasks.”*

# “Clinical Versus Actuarial Judgment”

---

- ❖ *“To be truly actuarial, interpretations must be both automatic (that is, prespecified or routinized) and based on empirically established relations.”*
- ❖ Gary’s statement: “This is predictive modeling (predictive analytics).”

# “Clinical Versus Actuarial Judgment”

---

- ❖ *“Even when given an information edge, the clinical judge still fails to surpass the actuarial method; in fact, access to additional information often does nothing to close the gap between the two methods.”*

# Why Use Generalized Linear Models?

---

- ❖ Can readily see link between predictors and outcomes
- ❖ Useful statistical tests for coefficients and fit of model
- ❖ Easier to explain than some other methods
- ❖ Software is widely available

# Classical Multiple Linear Regression

---

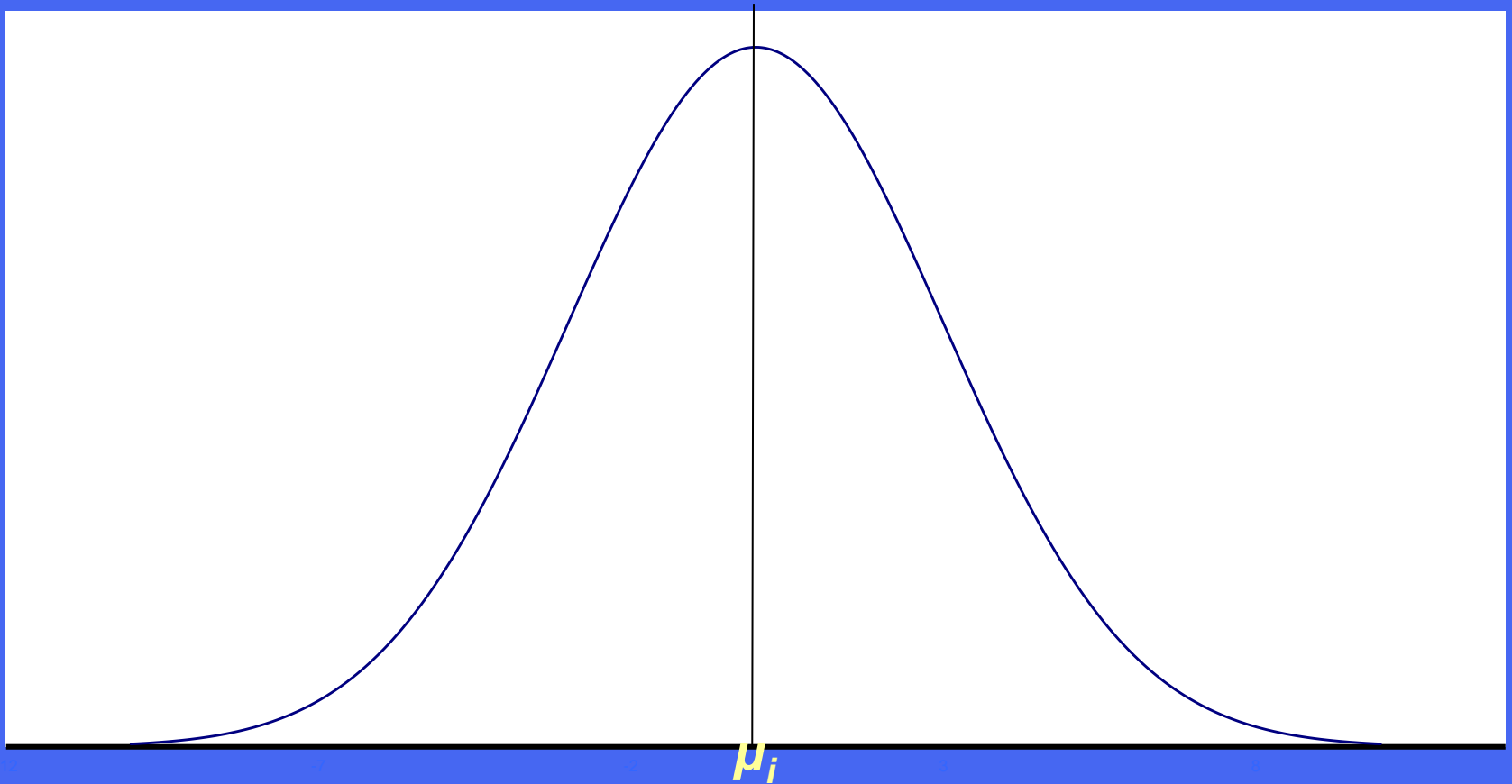
❖  $\mu_i = E[Y_i] = a_0 + a_1 X_{i1} + \dots + a_m X_{im}$

❖  $Y_i$  is Normally distributed random variable with constant variance  $\sigma^2$

❖ Want to estimate  $\mu_i = E[Y_i]$  for each  $i$



# Response $Y_i$ has Normal Distribution



# Problems with Traditional Model

---

- ❖ Number of claims is discrete
- ❖ Claim sizes are skewed to the right
- ❖ Probability of an event is in  $[0,1]$
- ❖ Variance is not constant across data points  $i$
- ❖ Nonlinear relationship between  $X$ 's and  $Y$ 's

# Generalized Linear Models - GLMs

---

- ❖ Fewer restrictions
- ❖  $Y$  can model number of claims, probability of renewing, loss severity, loss ratio, etc.
- ❖ Large and small policies can be put into one model
- ❖  $Y$  can be nonlinear function of  $X$ 's
- ❖ Classical linear regression model is a special case

# Generalized Linear Models - GLMs

---

- ❖  $g(\mu_i) = a_0 + a_1 X_{i1} + \dots + a_m X_{im}$
- ❖  $g(\cdot)$  is the link function
- ❖  $E[Y_i] = \mu_i = g^{-1}(a_0 + a_1 X_{i1} + \dots + a_m X_{im})$ 
  - $Y_i$  can be Normal, Poisson, Gamma, Binomial, Compound Poisson, ...
  - Variance can be modeled

# Exponential Family of Distributions – Canonical Form

---

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

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

$$\text{Var}[Y] = b''(\theta)a(\phi)$$

$\theta$  is the parameter of interest !

$\phi$  is often called a nuisance parameter.

# Why Exponential Family?

---

- ❖ Distributions in Exponential Family can model a variety of problems
- ❖ Standard algorithm for finding coefficients  $a_0, a_1, \dots, a_m$

# Normal Distribution in Exponential Family

---

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

$$= \exp\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\right) \exp\left(-\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{\mu y - \mu^2 / 2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2}\right)$$

# Normal Distribution in Exponential Family

---

$\theta$

$b(\theta)$

$$f(y; \mu, \sigma^2) = \exp \left( \frac{\mu y - \mu^2 / 2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \right)$$

Let  $\theta = \mu$  and  $a(\phi) = \sigma^2$ ,

then  $b(\theta) = \theta^2 / 2 \rightarrow b'(\theta) = \theta = \mu$

and  $Var[Y] = b''(\theta)a(\phi) = 1 \cdot \sigma^2 = \sigma^2$



# Poisson Distribution in Exponential Family

---

$$\Pr[Y = y] = \frac{\mu^y e^{-\mu}}{y!}$$

$$\Pr[Y = y] = \exp \left\{ \ln \left( \frac{\mu^y e^{-\mu}}{y!} \right) \right\}$$

$\theta$

$$\Pr[Y = y] = \exp \left\{ \frac{(\ln \mu) \cdot y - \mu}{1} - \ln(y!) \right\}$$

# Compound Poisson Distribution

---

- ❖  $Y = C_1 + C_2 + \dots + C_N$
- ❖  $N$  is Poisson random variable
- ❖  $C_i$  are i.i.d. with Gamma distribution
- ❖ This is an example of a Tweedie distribution
- ❖  $Y$  is member of Exponential Family

# Members of the Exponential Family

---

- Normal
- Poisson
- Binomial
- Gamma
- Inverse Gaussian
- Compound Poisson

$$\text{Var}[Y_i] = \Phi V(\mu_i)/w_i$$

---

|                    | <u><math>V(\mu)</math></u> |
|--------------------|----------------------------|
| ❖ Normal           | $\mu^0$                    |
| ❖ Poisson          | $\mu$                      |
| ❖ Binomial         | $\mu(1-\mu)$               |
| ❖ Tweedie          | $\mu^p, 1 < p < 2$         |
| ❖ Gamma            | $\mu^2$                    |
| ❖ Inverse Gaussian | $\mu^3$                    |

# Variance of $Y_i$ and Fit at Data Point $i$

---

- ❖  $\text{Var}(Y_i)$  is big  $\rightarrow$  looser fit at data point  $i$
- ❖  $\text{Var}(Y_i)$  is small  $\rightarrow$  tighter fit at data point  $i$

$$\text{Tightness of fit} \propto \frac{1}{\text{Var}(Y_i)}$$

# Estimating Coefficients $a_1, a_2, \dots, a_m$

---

- ❖ Classical linear regression uses least squares
- ❖ GLMs use Maximum Likelihood Method

# Which Exponential Family Distribution?

---

- ❖ Frequency: Poisson
- ❖ Severity: Gamma
- ❖ Loss ratio: Compound Poisson
- ❖ Pure Premium: Compound Poisson
- ❖ How many policies will renew: Binomial

# What link function?

---

- ❖ Additive model: identity
- ❖ Multiplicative model: natural log
- ❖ Modeling probability of event: logistic



# Chapter 5: Generalized Linear Models

---

- ❖ **Intended as a first exposure to GLMs**
- ❖ **Tried to make it accessible and self-contained**
- ❖ **Hard to squeeze everything into one chapter – at Ball State the topic spans a semester-long course**

# 5.1 Introduction to Generalized Linear Models

---

## 5.1.1 Assumptions of Linear Model

- Shortcomings for actuarial applications

## 5.1.2 Generalized Linear Model Assumptions

## **5.2 Exponential Family of Distributions**

### **5.2.1 The Variance Function and the Relationship between Variances and Means**

## **5.3 Link Functions**

## **5.4 Maximum Likelihood Estimation**

### **5.2.1 Quasi-likelihood**

## **5.5 Generalized Linear Model Review**

## **5.6 Applications**

---

**5.6.1 Modeling Probability of Cross Selling with Logit Link**

**5.6.2 Claim Frequency with Offset**

**5.6.3 Severity with Weights**

**5.6.4 Modeling Pure Premiums or Loss Ratios**

# 5.7 Comparing Models

---

## 5.7.1 Deviance

## 5.7.2 Log-likelihood, AIC, AICC, and BIC

**The End**