Generalized Linear Models (GLMs)

An Intuitive, Example-Driven Guide

What is a Generalized Linear Model?

Key Question: How can we predict some outcome (Y) using input features (X), even when Y is *not* normally distributed or *not* continuous?

Ordinary Linear Regression only works when the output is continuous, the relationship is linear, and the noise is Gaussian.

Real-world data often violates this:

• **Binary outcomes:** Did the customer buy? Yes/No

• **Counts:** How many emails did I get today?

• **Probabilities:** Chance of disease?

• **Positive skewed values:** Insurance claims amount

GLMs were invented to generalize linear regression to ANY type of outcome distribution.

The Three Components of a GLM

Every GLM has the same fundamental structure, consisting of three key components:

1. A Probability Distribution for the Output

Y must come from the **exponential family**, which includes:

- Normal (continuous)
- Bernoulli (binary)
- Poisson (counts)
- Gamma (positive continuous)
- Exponential
- Binomial

Type of Output	Suitable Distribution
Yes/No	Bernoulli
Count data	Poisson
Skewed positive values	Gamma
Real valued	Normal

2. A Linear Predictor

You still form a linear combination of inputs:

$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$

This is the **core "linear" part** in *Generalized Linear Model*.

3. A Link Function

The link function **connects the linear predictor (**η**)** to the **mean of the output**:

$g(E[Y]) = \eta$

Different choices of g give different models:

Model	Link Function	Output Distribution

Linear regression	Identity: $g(\mu) = \mu$	Normal
Logistic regression	Logit: $g(\mu) = \log(\mu/(1-\mu))$	Bernoulli
Poisson regression	Log: $g(\mu) = \ln(\mu)$	Poisson
Gamma regression	Inverse: $g(\mu) = 1/\mu$	Gamma

This is the key innovation: The link function lets a linear predictor produce outputs that are *not* linear.

Intuition with Examples

Example 1: Logistic Regression (Binary Classification)

Goal: Predict whether a customer buys (1) or not (0).

Binary Y → use **Bernoulli distribution**. Probabilities must lie in [0,1], but linear output can be any value.

So we connect them with the **logit link**:

$\text{logit}(p) = \ln(p/(1-p)) = \beta_0 + \beta_1 x_1 + \ldots$

Invert it:

$p = 1 / (1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots)})$

What's happening intuitively?

• η is a line
• Logistic function squashes it into a probability
• You get valid predictions for Yes/No outcomes

→ **This is a GLM!**

Example 2: Poisson Regression (Counts)

Predict: How many cars pass a checkpoint in one minute?

Counts → **Poisson distribution**. Counts must be positive, so we need to prevent the linear model from predicting negatives.

Use **log link**:

$\ln(E[Y]) = \beta_0 + \beta_1 x$

Invert:

$E[Y] = e^{(\beta_0 + \beta_1 x)}$

This ensures the rate is always positive.

Example 3: Gamma Regression (Skewed/Positive Data)

Predict: Insurance claim amount (always positive, right-skewed).

Use **Gamma distribution** with **inverse link**:

$1/E[Y] = \beta_0 + \beta_1 x + \ldots$

This ensures positivity.

The Mathematics of GLMs

Let's formalize the GLM framework:

1. Outcome Distribution

Y comes from the exponential family:

$$f(y \mid \theta) = \exp((y\theta - b(\theta))/\phi + c(y, \phi))$$

Where:

- θ = canonical parameter
- ϕ = dispersion (variance)
- $b(\theta)$ controls mean and variance

2. Linear Predictor

$$\eta = X\beta$$

3. Link Function

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

4. Estimation

Parameters β are estimated by **maximum likelihood**, typically via:

- Iteratively Reweighted Least Squares (IRLS)
- Gradient descent
- Newton's method

Why Are GLMs So Important?

GLMs unify many models under one framework:

Model Type	Part of GLM?	Why
Linear regression	✓	Normal dist + identity
Logistic regression	✓	Bernoulli + logit
Poisson regression	✓	Poisson + log link
Softmax regression	✓	Multinomial + logit
Gamma regression	✓	Gamma + inverse

GLMs allow you to:

• Choose an appropriate distribution for your data

• Keep the simplicity of linear models

• Extend to complex outcomes

Final Intuition Summary

A **GLM is a smart extension of linear regression** that:

1. Lets you model **non-normal data**

2. Still uses a **linear combination of features**

3. Uses a **link function** to produce valid outputs

4. Covers logistic regression, Poisson regression, etc. as **special cases**

Key Takeaway: GLMs provide a unified framework for regression and classification that generalizes ordinary linear regression to accommodate diverse response distributions. By combining three components—a probability distribution from the exponential family, a linear predictor, and a link function—GLMs enable principled modeling of binary, count, positive continuous, and other non-Gaussian outcomes while maintaining computational efficiency and interpretability.