**Generalized Linear Models (GLMs)**

Generalized Linear Models (GLMs) are an extension of traditional linear models that allow for response variables that have error distribution models other than a normal distribution. They are widely used in various fields, including machine learning, statistics, and data analysis.

**Structure of a GLM**

The GLM consists of three components:

1. **Random Component**: The dependent variable (Y) follows a certain distribution from the exponential family of distributions (like binomial, Poisson, gamma, etc.)
2. **Systematic Component**: This is the predictor (linear) part of the model. It's essentially a linear combination of the predictors.

η = Xβ

where X is the matrix of predictors, and β is the vector of parameters to be estimated.

1. **Link Function**: This function provides the relationship between the random and systematic components. It links the mean of the distribution function to the predictors. This link function is usually represented as g(µ), where µ is the expected value of the dependent variable.

η = g(µ)

For instance, in logistic regression (a type of GLM), the link function would be the logit function.

**Estimation**

The coefficients in GLMs are estimated using maximum likelihood estimation (MLE).

The log-likelihood function for GLM given data D = {y, X} and parameters β can be expressed as:

L(β | D) = Σ(y\_i \* η\_i - b(η\_i)) + c(y\_i, φ)

where y\_i is the observed data,

η\_i is the linear predictor for observation i,

b(η) is a function derived from the exponential family distribution, and

c(y\_i, φ) is a function that may depend on the data and the dispersion parameter φ.

The score function, or the first derivative of the log-likelihood, is then:

U(β) = dL(β | D) / dβ = X^T \* (y - µ)

where

µ is the expected value of Y and is related to η through the link function.

The second derivative, or the Fisher Information matrix, is:

I(β) = -d^2L(β | D) / dβ^2 = X^T \* W \* X

where

W is a diagonal matrix with elements w\_i = [g'(µ\_i)]^2 / V(µ\_i),

g' is the derivative of the link function, and

V(µ) is the variance function of the distribution.

The MLEs are then the solutions to the score equations, U(β) = 0, which do not typically have closed form solutions for GLMs, unlike in the simpler linear regression case. Instead, these are typically solved using iterative numerical methods, like Newton-Raphson or Fisher scoring methods, which update the estimates at each step as:

β^(t+1) = β^t + (I(β^t))^(-1) \* U(β^t)

**Conclusion**

The Generalized Linear Models provide a flexible framework for modeling a variety of data types, while maintaining the simplicity and interpretability of the linear model. They have become a staple technique in the toolbox of many data scientists and statisticians.

**GLMs vs Linear Models**

Linear models and Generalized Linear Models (GLMs) are two key types of models used in statistics and data science. Each has strengths and is best suited for different types of problems. Here's a comparison between the two:

**Linear Models**

Linear models assume that the relationship between your independent variables and the dependent variable is linear. This means that a unit change in an independent variable will result in a constant change in the dependent variable, no matter the value of the independent variable.

Linear regression models also assume that the residuals (the difference between the observed and predicted values) are normally distributed and that the variance of the residuals is constant across all levels of the independent variables (homoscedasticity).

These assumptions can be quite restrictive. Real world data often doesn't fit these assumptions. For example, if the dependent variable is binary or count data, linear models may produce predictions that don't make sense (like predicting a negative number of events).

**Generalized Linear Models (GLMs)**

GLMs extend linear models by allowing for response variables that have error distribution models other than a normal distribution, and for an arbitrary link function of the response variable and the predictors.

The random component of GLMs can be a range of probability distributions, such as the normal, binomial, Poisson, gamma, etc. This makes GLMs more flexible and able to model a wider range of data, including binary, count, and other types of non-continuous data.

The link function in GLMs allows for the relationship between the dependent variable and the predictors to be non-linear. This means that a unit change in an independent variable does not result in a constant change in the dependent variable. Instead, the change depends on the current value of the independent variable. This can allow for better modeling of certain types of data.

**Summary**

While linear models are simpler and easier to interpret, their applicability is limited by their strict assumptions. GLMs, on the other hand, are a generalization of linear models that remove these limitations, making them a more flexible tool for modeling different types of data. However, with this flexibility comes increased complexity and computational cost.

Therefore, when choosing between linear models and GLMs, it's important to consider the nature of your data and the assumptions each model makes. If your data meet the assumptions of a linear model, it might be the simpler and more interpretable choice. However, if the data do not meet these assumptions, a GLM might be a better choice.