

Generalised Regression Models

Sagar Udasi

July 11, 2024

Foreword

Welcome to the World of Generalized Regression Models!

Imagine you're trying to understand how different ingredients affect the taste of a soup. You might start by changing one ingredient at a time and tasting the result. This simple idea - seeing how changes in one thing affect another - is at the heart of regression analysis, a fundamental tool in statistics.

Now, picture a master chef who can adjust multiple ingredients simultaneously, accounting for how they interact, and even predicting how the soup will taste in different conditions. This is akin to what *Generalized Regression Models (GRMs)* do with data. GRMs are like a *Swiss Army knife* for statisticians. They allow us to explore relationships in data that don't follow simple patterns. Think of trying to predict how often it will rain next month. It's not just about temperature or humidity alone - it's a complex interplay of many factors. GRMs give us a way to approach such intricate problems without missing the relationships which we would otherwise miss.

What makes GRMs special is their flexibility. They can handle all sorts of data - whether you're counting rare animals in a forest, predicting the odds of winning a game, or analyzing how quickly a medicine takes effect. This versatility has made GRMs invaluable in fields ranging from ecology to economics, from healthcare to social sciences.

This book aims to explore the world of GRMs, breaking down complex ideas into more digestible pieces. While it can't cover everything - statistics is a vast and ever-evolving field - we hope it will provide a helpful guide for those looking to understand and use these powerful tools.

Whether you're a student, a professional, or simply curious about how we make sense of the world through data, there's something here for you. GRMs may not be a household name, but their impact on our understanding of the world around us is profound.

So let's embark on this journey together, exploring the fascinating world of Generalized Regression Models - where the complexity of real-life meets the elegance of statistical theory.

Contents

1	Prerequisites	1
1.1	Discrete Probability Distribution	3
1.1.1	Binomial Distribution	3
1.1.2	Geometric Distribution	6
1.1.3	Poisson Distribution	10
1.1.4	Negative Binomial Distribution	11
1.2	Continuous Probability Distribution	12
1.2.1	Normal Distribution	12
1.2.2	Chi-squared Distribution	12
1.2.3	t Distribution	12
1.2.4	F Distribution	12

Prerequisites

Consider a scenario where you have a pair of distinguishable fair dice. We are interested in the distribution of the sum of the numbers appearing on the top faces of these dice when they are rolled. Let's denote this sum by the random variable S . The possible values of S range from 2 to 12, since each die can roll a number between 1 and 6.

The sample space of all possible outcomes (or microstates) when rolling two dice consists of 36 pairs, because each of the 6 faces of the first die can pair with any of the 6 faces of the second die. Therefore, the sample space Ω is:

$$\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$$

where i and j represent the numbers on the top faces of the two dice.

Probability of a Particular Score

The possible values of S (the macrostates) are the integers from 2 to 12. For each possible value of S , we can calculate the probability $P(S = k)$ by counting the number of pairs (i, j) that sum to k and dividing by the total number of pairs (36). Let's define $S = i + j$.

$$P(S = k) = \begin{cases} \frac{k-1}{36} & \text{for } 2 \leq k \leq 7 \\ \frac{13-k}{36} & \text{for } 8 \leq k \leq 12 \end{cases}$$

The most common score, or the mode of this distribution, is 7. This is because there are more pairs that sum to 7 than any other number. Specifically, there are 6 such pairs:

$$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$$

Thus, the probability of rolling a 7 is:

$$P(S = 7) = \frac{6}{36} = \frac{1}{6}$$

Uniform Distribution in Microstates Leading to Non-Uniform Distribution in Macrostates

An interesting observation in this example is that even though each individual pair of dice outcomes (microstates) has an equal probability of $1/36$ (a uniform distribution), the distribution of the sum (macrostate) is not uniform. This is because the number of microstates that contribute to each macrostate varies.

Physical systems/natural phenomena tend to reside in the most probable macrostate because it has the largest number of contributing microstates. In our dice example, the sum of 7 is the most probable macrostate because it has the highest number of pairs adding up to it. Here are some examples.

Imagine you have a large jar filled with small, identical balls. You shake the jar, and the balls move around randomly. Now, think of the balls as gas molecules in a closed container. The positions and velocities of these molecules represent the microstates of the system. In thermodynamics, the macroscopic properties of the gas, such as temperature, pressure, and volume, are what we observe and measure. These properties are known as macrostates. Each macrostate can be realized by a vast number of different arrangements of the molecules (microstates).

Entropy is a measure of the number of possible microstates that correspond to a particular macrostate. It quantifies the level of disorder or randomness in the system. The more ways there are to arrange the molecules without changing the macroscopic properties, the higher the entropy.

$$S = k_B \ln(\Omega)$$

Here, S is the entropy, k_B is the Boltzmann constant, and Ω is the number of microstates corresponding to a given macrostate. This equation tells us that entropy increases logarithmically with the number of microstates.

The second law of thermodynamics states that the entropy of an isolated system will tend to increase over time, reaching a maximum at equilibrium. This means that over time, the system will naturally progress towards the state with the highest entropy, because this state has the largest number of possible microstates.

At equilibrium, the gas molecules are most likely found in the state with the highest entropy. This is because there are more ways (microstates) for the system to be in this state than in any other. For example, if the gas molecules can move around freely, they will spread out to fill the container evenly. This uniform distribution corresponds to the macrostate with the highest entropy, as it can be achieved by the largest number of different microstates.

In finance, the principle can be observed in the context of portfolio diversification. Investors diversify their portfolios to spread risk and increase the probability of achieving a desired return. Each investment in the portfolio represents a microstate, and the overall performance of the portfolio represents the macrostate.

Case with the Indistinguishable Dice

Since the dice are indistinguishable, the order in which the numbers appear does not matter. For example, the outcome (1, 2) is the same as (2, 1).

The possible outcomes (microstates) for the two dice can be listed as:

$$(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ (3, 3), (3, 4), (3, 5), (3, 6), (4, 4), (4, 5), (4, 6), (5, 5), (5, 6), (6, 6)$$

There are a total of 21 distinct microstates. The number of macrostates still remain the same, a set of integers from 2 to 12.

1.1 Discrete Probability Distribution

Consider an experiment where a biased coin is tossed N times. Let X denote the random variable representing the number of heads observed. The probability of getting a head in a single toss is denoted by p , and thus the probability of getting a tail is $q = 1 - p$. The sample space for this experiment consists of all possible outcomes of N coin tosses. Each toss is independent, and the outcome of each toss is either a head or a tail. Therefore, the sample space has 2^N possible outcomes.

The random variable X represents the number of heads obtained in N tosses. The possible values of X are $\{0, 1, 2, \dots, N\}$.

1.1.1 Binomial Distribution

Definition 1.1. The binomial distribution describes the probability of having exactly k successes (heads in this case) in N independent Bernoulli trials (coin tosses), each with success probability p .

The probability mass function of X , the number of heads, is given by:

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

where $\binom{N}{k}$ is the binomial coefficient.

Expected Value and Variance

Definition 1.2. The expected value represents the average value of X over all possible outcomes, weighted by their probabilities.

The expected value $E[X]$ of X is calculated as:

$$E[X] = \sum_{k=0}^N k \cdot P(X = k)$$

where $P(X = k)$ is the probability of getting exactly k heads in N tosses. Hence,

$$E[X] = \sum_{k=0}^N k \cdot \binom{N}{k} p^k (1 - p)^{N-k}$$

This represents the average number of heads obtained over a large number of Bernoulli trials.

Definition 1.3. The variance is the measure of how much the values of X deviate from the mean $E[X]$.

The variance $\text{Var}(X)$ of X is defined as:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

What's the physical interpretation of this? From the above expression, you can see that the variance can't be negative (*because it's the expected value of some quantity that is squared*). Its value can't be zero. Because if variance is 0, it means it is a sure variable (*where its value is always equal to the mean*) and is no longer random. Any variable that can take multiple values, must have a non-zero positive variance.

Generating Functions

Important: For a detailed exploration of generating functions, refer to Appendix 1.2.4.

For the distribution of X , the generating function $f(z)$ is defined as:

$$f(z) = \sum_{n=0}^N P(X = n) \cdot z^n$$

where $P(X = n)$ is the probability mass function of X . For our case, $P(n)$ representing the probability of obtaining exactly n heads in N tosses.

For the binomial distribution, where $P(n) = \binom{N}{n} p^n q^{N-n}$, the generating function becomes:

$$f(z) = (pz + q)^N$$

Definition 1.4. The r -th moment $E[X^r]$ of X is defined as the r 'th derivative of the generating function $f(z)$ at $z = 1$:

$$E[X^r] = \left. \frac{d^r}{dz^r} f(z) \right|_{z=1}$$

Using this definition of moments, we can easily find the mean and the variance of the binomial distribution.

Mean and Variance of Binomial Distribution using Generating Function

The mean μ (expected value $E[X]$) can be found using the first derivative of $f(z)$ at $z = 1$:

$$\mu = \left. \frac{d}{dz} f(z) \right|_{z=1}$$

For $f(z) = (pz + q)^N$,

$$\frac{d}{dz}f(z) = Np(pz + q)^{N-1}$$

So,

$$\mu = Np$$

The variance σ^2 (variance $\text{Var}(X)$) can be found using the second derivative of $f(z)$ at $z = 1$:

$$\sigma^2 = \left. \frac{d^2}{dz^2}f(z) \right|_{z=1} + \mu - \mu^2$$

For $f(z) = (pz + q)^N$,

$$\frac{d^2}{dz^2}f(z) = Np(N-1)(pz + q)^{N-2}$$

So,

$$\sigma^2 = Np(N-1)p$$

Definition 1.5. The relative fluctuation RF of a random variable X is defined as the ratio of its standard deviation to its mean:

$$\text{RF} = \frac{\sigma}{\mu}$$

where σ is the standard deviation and μ is the mean (expected value) of X .

Substituting the expressions for σ and μ :

$$\text{RF} = \frac{\sigma}{\mu} = \frac{\sqrt{Np(1-p)}}{Np}$$

Therefore, the relative fluctuation RF of a binomially distributed random variable X is:

$$\text{RF} = \sqrt{\frac{q}{N \cdot p}}$$

The important thing to notice here is that - the term $\frac{1}{\sqrt{N}}$ in RF indicates how this quantity decreases as N , the number of trials or degrees of freedom in a system, increases. This implies, the values relatively get closer and closer to mean, as the value of N increases. This characteristic is not unique to the binomial distribution but is typical in equilibrium systems.

In such contexts, N often scales with the system's complexity, such as the number of particles in a gas or molecules in a substance. For large N , like Avogadro's number for particles, $N^{-1/2}$ becomes very small (approximately 10^{-12}), implying that RF diminishes significantly.

This feature is pivotal because in equilibrium, thermodynamics focuses on average values of macroscopic properties and often disregards fluctuations. When N is large, the relative fluctuation RF becomes negligible compared to the mean value μ , supporting the validity of thermodynamic principles under these conditions. Refer to the example 1.2.

1.1.2 Geometric Distribution

Consider a coin with probability p of landing heads (H) and probability $q = 1 - p$ of landing tails (T) each time it is tossed. Our experiment consists of repeatedly tossing the coin until we obtain heads for the first time. We are interested in finding the probability P_n that this first occurrence of heads happens on the $(n + 1)$ -th toss.

To understand the situation better, let us consider what must happen for heads to appear for the first time on the $(n + 1)$ -th toss:

1. The first n tosses must all result in tails.
2. The $(n + 1)$ -th toss must result in heads.

Definition 1.6. The probability of the first n tosses all being tails is q^n since each tail has a probability q and the events are independent. The probability of the $(n + 1)$ -th toss being heads is p .

Since these events are independent, we multiply their probabilities to get the probability P_n :

$$P_n = q^n \cdot p = (1 - p)^n \cdot p$$

This distribution of probabilities is known as the *geometric distribution*. It is called so because the probabilities form a geometric progression. The sample space of the random variable n , which represents the number of tails before the first head, consists of all non-negative integers $\{0, 1, 2, \dots\}$.

Let's check if the probability distribution is normalized. To show that the geometric distribution is normalized, we need to show that the sum of all probabilities equals 1. Mathematically, this means proving:

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} q^{k-1} p = 1$$

Let's compute the sum:

$$\sum_{k=1}^{\infty} q^{k-1} p = p \sum_{k=1}^{\infty} q^{k-1}$$

Recognize that $\sum_{k=1}^{\infty} q^{k-1}$ is a geometric series with the first term $a = 1$ and common ratio $r = q$:

$$\sum_{k=1}^{\infty} q^{k-1} = \sum_{n=0}^{\infty} q^n$$

The sum of an infinite geometric series is given by $\frac{1}{1-r}$ for $|r| < 1$:

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} = \frac{1}{p}$$

Therefore:

$$p \sum_{k=1}^{\infty} q^{k-1} = p \cdot \frac{1}{p} = 1$$

Generating Function of the Geometric Distribution

The probability generating function $f(z)$ for the geometric distribution is:

$$f(z) = \sum_{k=0}^{\infty} P(X = k)z^k = \sum_{k=0}^{\infty} q^k p z^k$$

$$f(z) = p \sum_{k=0}^{\infty} (qz)^k$$

This is a geometric series with first term $a = 1$ and common ratio $r = qz$:

$$\sum_{k=0}^{\infty} (qz)^k = \frac{1}{1-qz} \quad \text{for } |qz| < 1$$

Thus:

$$f(z) = p \cdot \frac{1}{1-qz} = \frac{p}{1-qz}$$

Moments of the Geometric Distribution

The mean $E[X]$ can be derived using the generating function $f(z)$:

$$E[X] = f'(1)$$

$$f'(z) = \frac{pq}{(1-qz)^2}$$

$$f'(1) = \frac{pq}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}$$

Hence, the mean is:

$$E[X] = \frac{q}{p}$$

The variance $\text{Var}(X)$ can be derived using the second moment $E[X^2]$:

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

First, we find $E[X^2]$:

$$E[X^2] = f''(1) + f'(1)$$

$$f''(z) = \frac{2pq^2}{(1 - qz)^3}$$

$$f''(1) = \frac{2pq^2}{(1 - q)^3} = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2}$$

Thus:

$$E[X^2] = \frac{2q^2}{p^2} + \frac{q}{p}$$

Now we can calculate the variance:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p}\right)^2$$

Simplify:

$$\text{Var}(X) = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q^2}{p^2} + \frac{q}{p} = \frac{q(q + p)}{p^2} = \frac{q}{p^2}$$

Therefore, the variance is:

$$\text{Var}(X) = \frac{q}{p^2}$$

Hence, the relative fluctuation is:

$$\text{Relative Fluctuation} = \frac{\frac{\sqrt{q}}{p}}{\frac{q}{p}} = \frac{\sqrt{q}}{q} = \frac{1}{\sqrt{q}}$$

Example 1.1. Time to Default in Credit Risk Modeling.

In credit risk analysis, one of the key metrics is the probability of default (PD) over a given time horizon. Let's consider a simple model where we assume that for each period (e.g., a year), there is a constant probability p that the borrower will default. The probability that the borrower does not default in a given period is then $q = 1 - p$.

We are interested in the random variable X , which represents the number of periods until default occurs. This scenario perfectly fits the geometric distribution.

The probability mass function of the geometric distribution in this context is:

$$P(X = k) = q^{k-1}p, \quad k = 1, 2, 3, \dots$$

where k is the number of periods until default.

Now, let's derive some important results for credit risk modeling using this distribution:

1. Probability of Default within n Periods: The cumulative probability of default within n periods is:

$$\begin{aligned} P(X \leq n) &= 1 - P(X > n) \\ &= 1 - q^n \\ &= 1 - (1 - p)^n \end{aligned}$$

This is a crucial result in credit risk modeling, as it allows risk managers to estimate the probability of default over any given time horizon.

2. Expected Time to Default: The expected value of a geometric distribution is $1/p$. In our context:

$$E[X] = \frac{1}{p}$$

This gives us the average time to default, which is an important metric for assessing the long-term risk of a loan or bond.

3. Survival Function: The survival function, which gives the probability that the borrower has not defaulted by time t, is:

$$S(t) = P(X > t) = q^t = (1 - p)^t$$

This function is used in pricing credit-sensitive instruments and in calculating expected loss.

4. Hazard Rate: The hazard rate, which represents the instantaneous default probability given survival up to time t, is constant in this model:

$$h(t) = \frac{f(t)}{S(t)} = \frac{pq^{t-1}}{q^t} = p$$

where $f(t)$ is the probability mass function. This constant hazard rate is a key property of the geometric distribution and implies a "memoryless" default process.

5. Value at Risk (VaR): For a given confidence level α , the VaR in terms of time to default is:

$$\text{VaR}_\alpha = \left\lceil \frac{\log(1 - \alpha)}{\log(1 - p)} \right\rceil$$

where $\lceil \cdot \rceil$ denotes the ceiling function. This gives the number of periods within which default will occur with probability α .

6. Expected Shortfall: The Expected Shortfall (ES), also known as Conditional VaR, for a continuous approximation of the geometric distribution is:

$$\text{ES}_\alpha = \frac{\text{VaR}_\alpha + \frac{1}{p}}{1 - \alpha}$$

This measure provides the expected time to default given that default occurs within the VaR time horizon.

7. Duration of a Default-Risky Zero-Coupon Bond: For a zero-coupon bond with maturity T and face value 1, assuming a constant risk-free rate r , the duration D is:

$$D = \frac{\sum_{t=1}^T t \cdot e^{-rt} \cdot p(1-p)^{t-1}}{\sum_{t=1}^T e^{-rt} \cdot p(1-p)^{t-1}}$$

This duration measure incorporates both interest rate risk and default risk, providing a more comprehensive measure of the bond's sensitivity to market changes.

The geometric distribution's application in credit risk modeling demonstrates its utility in financial mathematics. Its simplicity allows for tractable calculations, while still capturing the essential nature of default risk as a series of independent trials over time. However, it's important to note that this model assumes a constant default probability over time, which may not always be realistic. More sophisticated models might incorporate time-varying default probabilities or dependencies between defaults of different entities.

1.1.3 Poisson Distribution

Example 1.2. Fluctuations of Gas Molecules. Consider a classical ideal gas of N molecules in thermal equilibrium inside a container of volume V . The molecules move randomly, undergoing elastic collisions with each other. The average number density of the particles is $\rho = \frac{N}{V}$.

Now, let's focus on a smaller sub-volume v within the container. The number of molecules n present in this sub-volume is a fluctuating quantity. At any given instant, n can range from 0 to N . We want to find the probability $P(n)$ that there are exactly n molecules in the sub-volume.

The a priori probability that a molecule resides within the sub-volume v is $\frac{v}{V}$, because each molecule has an equal chance of being in any volume element within the container. Therefore, the probability that a molecule is outside v is $1 - \frac{v}{V}$.

Assuming independence among molecules (no correlation between their locations), the probability $P(n)$ follows a binomial distribution:

$$P(n) = \binom{N}{n} \left(\frac{v}{V}\right)^n \left(1 - \frac{v}{V}\right)^{N-n}$$

This distribution has $p = \frac{v}{V} = \frac{\rho v}{N}$, where ρ is the number density $\frac{N}{V}$. The mean number of molecules in the sub-volume v is $\langle n \rangle = \rho v$.

The relative fluctuation RF in n , defined as $\left[\frac{1-\rho v/N}{\rho v}\right]^{1/2}$, characterizes how much n fluctuates relative to its mean.

In the thermodynamic limit where $N \rightarrow \infty$ and $V \rightarrow \infty$ while keeping $\rho = \frac{N}{V}$ fixed, the binomial distribution describing $P(n)$ tends towards the Poisson distribution. This transition occurs because as N and V increase indefinitely, the discrete nature of the binomial distribution smoothens out, resembling the continuous Poisson distribution.

Specifically,

1. The mean number of molecules $\langle n \rangle = \rho v$ remains unchanged in this limit, reflecting the average density within the sub-volume v .
2. The relative fluctuation RF, which measures the standard deviation relative to the mean, scales as $(\rho v)^{-1/2}$. This indicates that as the system size increases, fluctuations in the number of molecules within v become proportionally smaller compared to the mean value.

For a volume $v = 1 \text{ m}^3$ of air at standard temperature and pressure:

- The number density of air molecules at STP is approximately $\rho \approx 2.7 \times 10^{25} \text{ m}^{-3}$. This value is derived from Avogadro's number and the molar volume of a gas at STP.

$$\rho = \frac{N_A}{22.4 \times 10^{-3} \text{ m}^3} \approx 2.7 \times 10^{25} \text{ m}^{-3}$$

- Using $v = 1 \text{ m}^3$, the relative fluctuation RF is:

$$\text{RF} \approx (\rho v)^{-1/2} = (2.7 \times 10^{25} \times 1 \text{ m}^3)^{-1/2} \approx 6.1 \times 10^{-13}$$

This tiny relative fluctuation underscores how negligible the fluctuations are compared to the mean number of molecules, demonstrating the applicability of statistics/thermodynamics in describing macroscopic systems.

1.1.4 Negative Binomial Distribution

Relative Fluctuation

The relative fluctuation is defined as the standard deviation divided by the mean value:

$$\begin{aligned} \text{Relative Fluctuation} &= \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]} \\ &= \frac{\sqrt{q/p^2}}{q/p} \\ &= \frac{\sqrt{q}/p}{q/p} \\ &= \frac{\sqrt{q}}{q} \\ &= \frac{1}{\sqrt{q}} \end{aligned}$$

Therefore, we have shown that the relative fluctuation is $1/\sqrt{q}$.

In conclusion, we have derived and verified all the requested properties of the Geometric Distribution using its generating function and moments:

1. The mean value is q/p
2. The variance is q/p^2
3. The relative fluctuation is $1/\sqrt{q}$

These results provide important insights into the behavior and characteristics of the Geometric Distribution, which is crucial for its applications in various fields, including finance and risk analysis.

1.2 Continuous Probability Distribution

1.2.1 Normal Distribution

1.2.2 Chi-squared Distribution

1.2.3 t Distribution

1.2.4 F Distribution

Appendix

Generating Functions

[1]

In the book *102 Combinatorial Problems*, authored by Titu Andreescu and Zuming Feng to train the USA team for the *International Math Olympiad*, there is a seemingly innocent question, Problem No. 10 of Chapter 2:

Find the number of subsets of $1, \dots, 2000$, the sum of whose elements is divisible by 5.

If we try to write a brute-force program to find all subsets whose sum is divisible by 5 and count them, it's impossible to do so because the 2^{2000} is a massively huge number.

Bibliography

- [1] Grant Sanderson (3Blue1Brown). Olympiad level counting (generating functions). YouTube Video, 2022.