ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

ΑΔΣ

Advances in Decision Sciences

Volume 28
Issue 3
September 2024

亞洲大學
ASIA UNIVERSITY

SCIENTIFIC & BUSINESS WORLD

Published by Asia University, Taiwan and Scientific and Business World

Newton-Raphson Method: Overview and Applications

Sal Ly
Faculty of Mathematics and Statistics, Ton Duc Thang University
Ho Chi Minh City, Vietnam
Email: lysal@tdtu.edu.vn

Kim-Hung Pho
Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
Email: phokimhung@tdtu.edu.vn

Shin-Hung Pan
Department of Information Management
Chaoyang University of Technology, Taiwan
Email: vincentpan@cyut.edu.tw

Wing-Keung Wong
Department of Finance, Fintech Center, and Big Data Research Center,
Asia University, Taiwan, and
Department of Medical Research, China Medical University Hospital, Taiwan, and
Business, Economic and Public Policy Research Centre, Hong Kong Shue Yan University, and
The Economic Growth Centre, Nanyang Technological University, Singapore
***Corresponding author* Email:** wong@asia.edu.tw

Received: April 29, 2024; First Revision: July 10, 2024;

Last Revision: August 11, 2024; Accepted: September 7, 2024;

Published: March 30 ,2025

Abstract

Purpose: This paper provides a comprehensive overview of the Newton-Raphson method (NRM) and illustrates the use of the theory by applying it to diverse scientific fields.

Design/methodology/approach: This study employs a systematic approach to analyze the key characteristics of the NRM that facilitate its broad applicability across numerous scientific disciplines. We thoroughly explore its mathematical foundations, computational advantages, and practical implementations, emphasizing its versatility as a problem-solving tool.

Findings: The findings of this paper include a detailed examination of the NRM, demonstrating its efficacy in solving non-linear equations, systems of equations, and optimization problems. The study further highlights the relevance of NRM in addressing complex challenges within probability, statistics, applied mathematics, and other related fields.

Originality/value: This study contributes to the existing literature by providing a comprehensive and in-depth analysis of the NRM's diverse applications. It effectively bridges the gap between theoretical understanding and practical utilization, thereby serving as a valuable resource for researchers and practitioners seeking to leverage the NRM in their respective domains.

Practical implications: This research showcases the practical utility of the NRM through two illustrative case studies: optimizing loudspeaker placement for COVID-19 public health communication and determining the submersion depth of a floating spherical object in water. Additionally, the paper demonstrates the NRM's extensive use in estimating parameters of probability distributions and regression models. It highlights its significance across various areas within Decision Sciences, including applied mathematics, finance, and education. This paper contributes both a theoretical overview and a display of diverse practical applications of the NRM.

Keywords: Newton-Raphson method, Application, real problems, Mathematics.

JEL classifications: A10, G00, G31, O32

Article type: Research article

1 Introduction

Mathematics is a fundamental discipline with widespread real-world applications, particularly in today's rapidly evolving technological landscape. Its importance is reflected in its inclusion across all levels of education, leading to significant and ongoing scientific research. Among the core topics within mathematics, studying equations and systems of equations is paramount, underpinning theoretical advancements and practical applications. The Newton-Raphson Method (NRM), named after Isaac Newton and Joseph Raphson, is a highly effective and widely adopted iterative technique for solving such problems (Dedieu, 2015). Renowned for its precision and efficiency, the NRM has become an essential tool in mathematics and across numerous scientific fields. It provides rapid solutions to complex equations and drives engineering, economics, and decision sciences innovation.

The NRM is a well-established numerical method for approximating solutions to equations involving real-valued functions. Given a real-valued function, the process begins by calculating its derivative and selecting an initial guess. Provided the initial value satisfies the assumptions of the NRM, such as the continuity and differentiability of the function, the method iteratively refines the approximation. Each iteration yields a closer estimate of the true solution by leveraging the relationship between the function and its derivative. This process continues until the difference between successive approximations falls below a predefined tolerance level, ensuring a solution of desired precision and reliability. Notably, the NRM is also known as the tangent method, a concept further explored in Section 2.2. The method's applicability extends to complex functions and linear and nonlinear equations systems (Dwyer, et al., 2009). Furthermore, the NRM can be refined and tailored for applications across various scientific disciplines, and it is a frequently utilized algorithm in computational software designed to find optimal solutions (Truong, et al., 2019a). Consequently, the NRM stands as a powerful and valuable scientific tool.

While direct applications and improvements of the NRM to address related problems or practical issues have been extensively documented, researched, and discussed in the literature—as evidenced by works such as Smietanski (2007), Morini, et al. (2010), Smietanski (2011), Gatilov (2014), Gaudreau, et al. (2015), Zhou and Zhang (2020), Zhou, et al. (2021), Gobet and Grangereau (2022), and Hassan and Moghrabi (2023)—there remains a gap in the literature for a comprehensive study specifically discussing the breadth of the NRM's applications in practice and within the domain of Decision Sciences.

This paper aims to fill this gap by presenting a complete and comprehensive overview of the NRM, offering a clear and detailed perspective on the method's theory and application. We present the core formulas and the iterative approach, followed by detailed illustrations to facilitate understanding and optimal application of the NRM. Subsequently, we introduce, discuss, and formally present a range of applications, including two practical examples. We further explore the method's relevance within probability, statistics, applied mathematics, finance, and education.

The structure of this paper is as follows: Section 2 provides a detailed presentation of the NRM, including its foundational formula, iterative approach, and illustrative examples in both univariate and multivariate contexts. Section 3 introduces two practical problems where the NRM is applied, demonstrating its utility in real-world scenarios. Sections 4 and 5 explore specific applications of the

NRM within probability and statistics, focusing on parameter estimation for probability distributions and regression models. Section 6 reviews the applications of the NRM in Decision Sciences, encompassing areas such as applied mathematics, finance, and education. Finally, the concluding section summarizes this work's key findings and contributions.

2 Background of the Newton-Raphson method (NRM)

2.1 A historical note on the NRM

The Newton-Raphson Method (NRM), named in honor of the eminent mathematicians Isaac Newton (1643-1727) and Joseph Raphson (1648-1715), boasts a rich history of utilization spanning several centuries. Due to its early development and evolution, pinpointing the precise origins of the method remains a challenge. Initially conceived as a technique for solving equations with real-valued parameters, the NRM has since demonstrated remarkable adaptability, being extended to handle complex functions and systems of equations. This versatility has solidified the NRM's position as a fundamental tool in theoretical mathematical research and various applied scientific disciplines. For readers seeking a more in-depth exploration of the method's historical development, underlying construction, and graphical interpretations, readers may consult the scholarly works of Chen (2000), Coleman, et al. (2003), Agarwal, et al. (2006), Gatilov (2014), Gaudreau, et al. (2015), Truong, et al. (2019b), Pho (2022), Na, et al. (2023), and Doikov and Nesterov (2024). These references offer a comprehensive perspective on the method's evolution and foundational principles.

In the subsequent sections, our focus shifts to presenting and analyzing the NRM's most essential and widely employed formulas. While this paper refrains from a detailed examination of the algorithmic construction of the NRM, a topic extensively covered in prior studies, we encourage interested readers to consult the works of Truong, et al. (2019b) and Pho (2022) for in-depth discussions. Herein, to promote accessibility and ease of comprehension, we will examine the NRM in two distinct contexts: the univariate case involving single equations and the multivariate case concerning systems of equations. This structured approach will facilitate a clearer understanding of the NRM's application across different problem types.

2.2 Univariate Case

The general recurrence relation of the NRM to find the root (solution) of an equation $f(x) = 0$ in the univariate case is given as follows (note that $f: A \to \mathbb{R}$, with $A \subseteq \mathbb{R}$):

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0,1,2,\ldots; \tag{1}$$

where f represents the objective function and f' denotes the first derivative of f. Here, we can start by x_0 serving as the initial guess, x_1 is the first approximation obtained by substituting x_0 into the Equation (1), and x_n is the next approximation derived using x_{n-1}, continuing iteratively. The parameter n typically refers to the number of iterations performed. We repeat the general Equation in Equation (1) until the difference between two adjacent solutions is less than a given small number ϵ, for example, until $|x_{n+1} - x_n| < \epsilon = 0.0001$.

Figure 1 illustrates a recurrence process of the NRM in the one-dimensional setting for the case of a decreasing function $f(x)$ with the true solution (root) at point A. Letting x_0 as the starting value, we can calculate x_1 using Equation (1). Continuing this process, we use x_1, to find x_2 via the same formula. This iterative approach leads to an important observation: the values obtained in subsequent iterations get progressively closer to the correct solution. Specifically, x_1 is closer to the true solution than x_0, and x_2 is closer than x_1. This process continues until the calculated value is very close to or coincides with the correct solution.

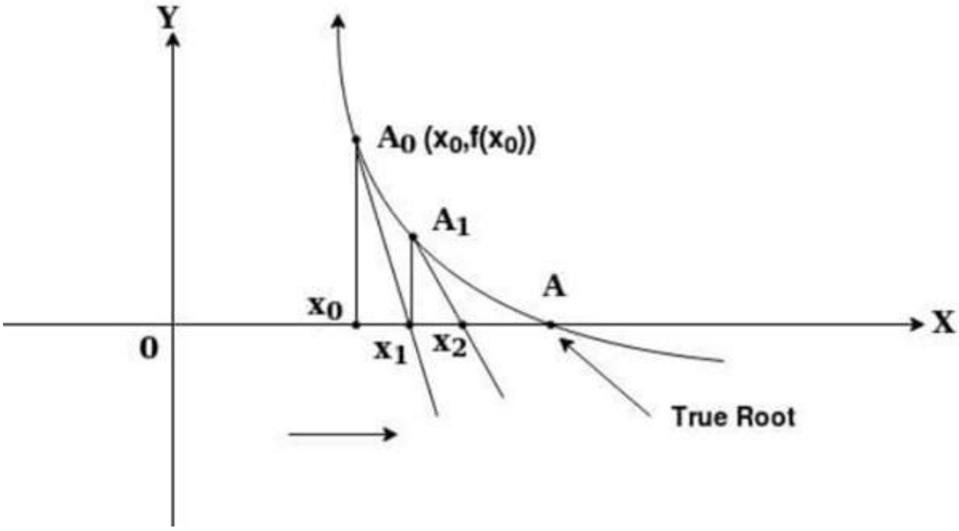

Figure 1: The recurrence process of the NRM for a decreasing univariate function.

Figure 2 provides another overview of how the NRM works for the case of an increasing function $f(x)$. It can be briefly concluded that the values obtained in later iterations are getting closer to the true solution.

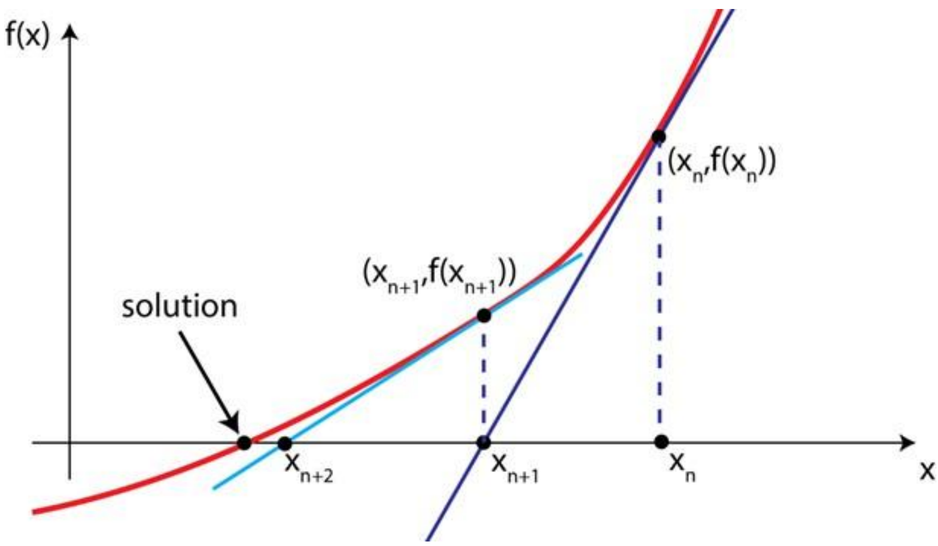

Figure 2: Another overview of how the NRM works for an increasing univariate function.

As mentioned earlier, the NRM is also known as the tangent method. Figure 3 provides an intuitive visualization of why the NRM is called the tangent method. Looking at Figure 3, we can see that after choosing an initial value x_0, the point $(x_0, f(x_0))$ is generated. The tangent line at this point intersects the horizontal axis at x_1. After obtaining x_1, the process continues: the tangent line at x_1 intersects the horizontal axis at the next point x_2, and this iterative process continues until the desired solution is reached. Thus, each iteration to find the optimal solution in the NRM requires a tangent line, which is why the NRM is sometimes called the tangent method.

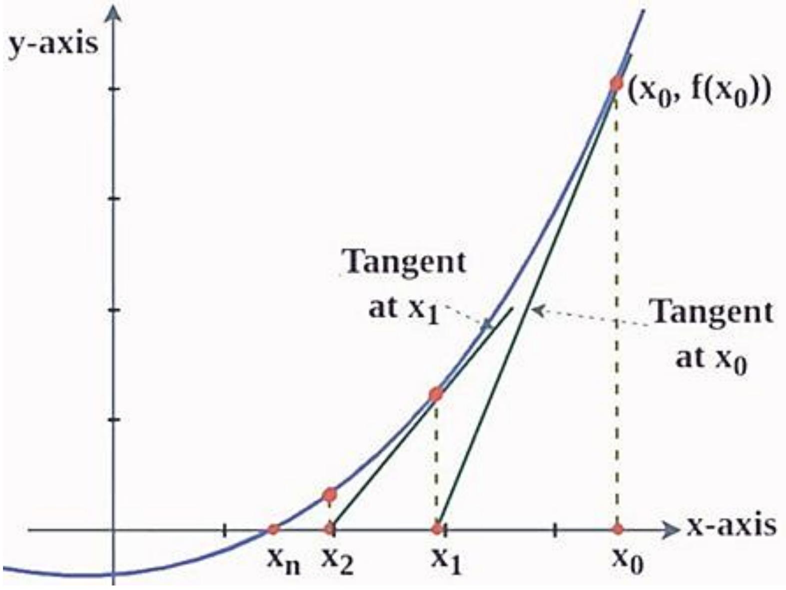

Figure 3: Visualization of the NRM as a tangent method.

In particular, there are cases where solving the equation *f'(x) = 0* is necessary (e.g., when solving optimization problems *f(x) → min* or *f(x) → max*, or finding critical points of the function f(x)). In such scenarios, the NRM can still be applied:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \qquad n = 0,1,2,\dots; \tag{2}$$

where $f"(x)$ denotes the second derivative of the function $f(x)$.

2.3 Multivariate Case

In this section, we consider a more general case that we want to find the root (solution) of a system of equations (i.e., d variables and d nonlinear equations):

$$\begin{cases} f_1(x_1, \dots, x_d) = 0, \\ \dots \\ f_d(x_1, \dots, x_d) = 0. \end{cases}$$

The general formula for the NRM can be expressed as follows:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \left[\boldsymbol{J_f}\left(\boldsymbol{x}^{(n)}\right)\right]^{-1} \boldsymbol{f}(\boldsymbol{x}^{(n)}), \tag{3}$$

where $\boldsymbol{x}^{(0)} = \left(x_1^{(0)}, \dots, x_d^{(0)}\right)^T$ is an initial value, $\boldsymbol{f} = (f_1, \dots, f_d)^T$, and $\boldsymbol{J_f}(\boldsymbol{x})$ is the Jacobian matrix with each element defined by $[\boldsymbol{J_f}(\boldsymbol{x})]_{ij} = \frac{\partial f_i}{\partial x_j}(\boldsymbol{x}),\ i, j = 1, \dots, d.$

Similarly, for solving an unconstrained optimization problem ($f(x_1, \dots, x_d) \to min$ or $f(x_1, \dots, x_d) \to max$), the NRM is applied to find the root of the gradient equation $\nabla f(\boldsymbol{x}) = 0$ for the multivariable function $f(x_1, \dots, x_d)$. The iterative formula is:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \left[\boldsymbol{H}_f(\boldsymbol{x}^{(n)})\right]^{-1} \nabla f(\boldsymbol{x}^{(n)}), \tag{4}$$

where

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}; \frac{\partial f(\boldsymbol{x})}{\partial x_2}; \dots; \frac{\partial f(\boldsymbol{x})}{\partial x_d}\right)^T,$$

$$\boldsymbol{H}_f(x) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 x_d} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d^2} \end{bmatrix}.$$

We note that $\nabla f(\boldsymbol{x})$ and $\boldsymbol{H}_f(\boldsymbol{x})$ are the gradient vector and Hessian matrix of $f(\boldsymbol{x})$, respectively. Readers may read, for example, Gaudreau, et al. (2015), Truong, et al. (2019b), Pho (2022), and Doikov and Nesterov (2024) for more information.

3 Some Real-World Applications of the Newton-Raphson Method

As one of humanity's oldest scientific disciplines, mathematics continues to demonstrate unparalleled relevance and applicability, especially in the contemporary era. Modern mathematical innovation accelerates, leading to new fields and challenging established paradigms. Beyond its traditional applications in astronomy, physics, and mechanics, mathematics has become increasingly indispensable in various disciplines, including chemistry, biology, machine learning, artificial intelligence (AI), and the social sciences. Indeed, nearly every scientific field now relies upon mathematical tools for analysis and problem-solving.

Despite this pervasive applicability of mathematics, concerns have been raised regarding the limited availability of applied mathematicians, particularly in developing nations such as Vietnam. This issue has been repeatedly highlighted by experts at various scientific conferences, as documented by Blum and Niss (1991). Such a shortage significantly restricts the effective application of mathematical principles to socio-economic development. In transitioning to a multi-sector commodity economy with a socialist-oriented market mechanism, Vietnam is experiencing a growing demand for sophisticated mathematical tools across diverse sectors, including banking, finance, and technology.

A key question thus arises: How can mathematics, and specifically powerful techniques like the NRM, be most effectively leveraged in practical contexts? As an iterative method recognized for its optimization capabilities, the NRM offers considerable potential for addressing complex real-world challenges. To illustrate its practical utility and broad applicability, this paper presents concrete examples showcasing the method's diverse and impactful applications across various fields. These examples demonstrate the NRM's capability to solve complex problems and encourage greater utilization of mathematical techniques to address real-world challenges.

3.1 Optimizing Loudspeaker Placement for COVID-19 Public Health Communication

The Coronavirus disease, caused by the SARS-CoV-2 virus, represents a global pandemic of unprecedented scale, with profound and far-reaching impacts on societies worldwide. Addressing this multifaceted crisis necessitates innovative and multifaceted solutions, including implementing effective public health communication strategies. One such strategy involves deploying loudspeakers to disseminate anti-epidemic information to convey critical updates and educational materials related to COVID-19. These loudspeakers are pivotal tools for informing and educating the public regarding preventive measures, policy updates, and essential health guidelines, as noted by Tuan et al. (2022).

Consider a scenario within a rural region or district where an anti-epidemic loudspeaker requires optimal positioning to ensure comprehensive signal coverage across four key locations: A(37, 21, 10), B(30, 30, 30), C(10, 45, 25), and D(30, 70, 10). The primary objective is determining the precise spatial coordinates for the loudspeaker placement such that the sound intensity is equivalent across all specified locations. This approach ensures that local leaders and public health managers can effectively communicate real-time updates and crucial information to residents in an equitable manner. The solution to this optimization problem involves applying mathematical techniques to identify the optimal spatial location for the loudspeaker. Specifically, we will use NRM to solve this problem.

Figure 4: The anti-epidemic propaganda loudspeaker of COVID-19.

Solution:

Emitted sound waves propagate from a loudspeaker such that points equidistant from the source receive the same sound intensity. Thus, the center of the loudspeaker acts as the center of a spherical surface passing through points A, B, C, and D, which are illustrated in Figure 5.

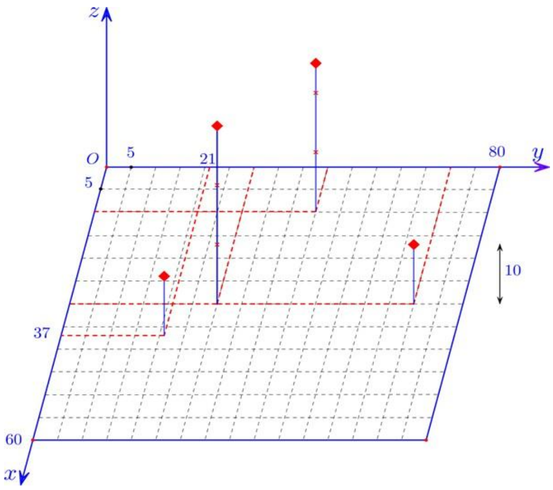

Figure 5: The illustration of $A, B, C,$ and D on the Oxyz coordinate.

The general spherical equation with the center is $I(a_1, a_2, a_3)$ has the following form:

$$(S): x_1^2 + x_2^2 + x_3^2 - 2a_1x_1 - 2a_2x_2 - 2a_3x_3 + a_4 = 0.$$

Given that (S) passes through points $A(37; 21; 10), B(30; 30; 30), C(10; 45; 25),$ and $D(30; 70; 10)$, we can set up the following system of equations:

$$\begin{cases} 37^2 + 21^2 + 10^2 - 2a_1.37 - 2a_2.21 - 2a_3.10 + a_4 = 0, \\ 30^2 + 30^2 + 30^2 - 2a_1.30 - 2a_2.30 - 2a_3.30 + a_4 = 0, \\ 10^2 + 45^2 + 25^2 - 2a_1.10 - 2a_2.45 - 2a_3.25 + a_4 = 0, \\ 30^2 + 70^2 + 10^2 - 2a_1.30 - 2a_2.70 - 2a_3.10 + a_4 = 0. \end{cases}$$

This system of equations can be simplified to:

$$\begin{cases} 74a_1 + 42a_2 + 20a_3 - a_4 = 1910, \\ 60a_1 + 60a_2 + 60a_3 - a_4 = 2700, \\ 20a_1 + 90a_2 + 50a_3 - a_4 = 2750, \\ ; 60a_1 + 140a_2 + 20a_3 - a_4 = 5900. \end{cases}$$

Solving this system of equations, we obtain the following solution:

$$\begin{cases} a_1 = 30, \\ a_2 = 45, \\ a_3 = 10, \\ a_4 = 2400. \end{cases}$$

Therefore, the center of the spherical surface (S) is identified as I(30; 45; 10), indicating that the optimal spatial position for the anti-epidemic loudspeaker is located at the coordinates (30; 45; 10). In the context of rapid technological advancements, using computational tools to efficiently and precisely resolve systems of equations is crucial. In such computations, the NRM serves as a primary algorithm for approximating solutions, as supported by the work of Truong, et al. (2019a). While this specific example reduced to a linear system, applying the NRM becomes necessary when dealing with systems containing higher-order terms or exhibiting substantial non-linearity.

Regarding economic implications, the precise determination of the optimal loudspeaker location allows for minimizing the material (e.g., cabling) required to connect the central hub to the four key locations. This optimization reduces operational expenses through efficient planning and enhances the overall economic viability of the deployment. This section demonstrated how principles related to NRM applications arise in solving practical problems like optimizing loudspeaker placement for public health communication. The following section introduces another application, solving a nonlinear equation for a different real-world problem.

3.2 Determining the Submersion Depth of a Floating Ball in Water

The Mekong Delta, situated in the southernmost region of Vietnam, is widely recognized for its abundant rice production and diverse seafood resources. Furthermore, the region is known for unique cultural practices, including water balloon games and vibrant floating markets (see Figure 6). The most prominent floating market in this area has garnered international recognition, as noted by Gutkin (2012).

Figure 6: Floating market in the Mekong Delta, Vietnam.

Given the cultural significance of water-related activities in the Mekong Delta, this paper introduces the practical problem of determining the submersion depth of a floating spherical object. This "floating ball problem" (Figure 7) presents an application of mathematical principles to understand everyday phenomena and leads to a non-linear equation solvable by NRM.

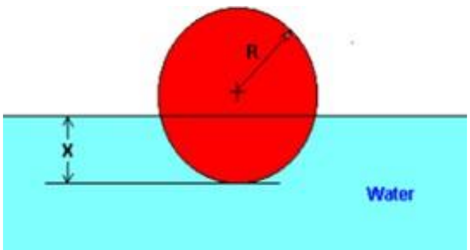

Figure 7: Floating ball problem.

Assuming a floating ball has a specific gravity of 0.5 (N/m^3) and a radius of 7 cm, we need to determine the depth to which the ball is submerged when floating in the water. The equation for the depth x in meters that the ball is submerged underwater is given by:

$$20x^5 + 12x^4 + 21.35x^3 + 11.9x^2 + 1.35x - 0.1 = 0.$$

We shall use the NRM with four iterations to find the depth x to which the ball is submerged.

Solution:

Since the radius of the floating ball is 7 cm, its diameter is 14 cm (0.14 m). As shown in Figure 8, the submersion depth x must be within the range [0, 0.14] m.

Let $f(x) = 20x^5 + 12x^4 + 21.35x^3 + 11.9x^2 + 1.35x - 0.1 = 0$. The graph of f(x) is plotted in Figure 8.

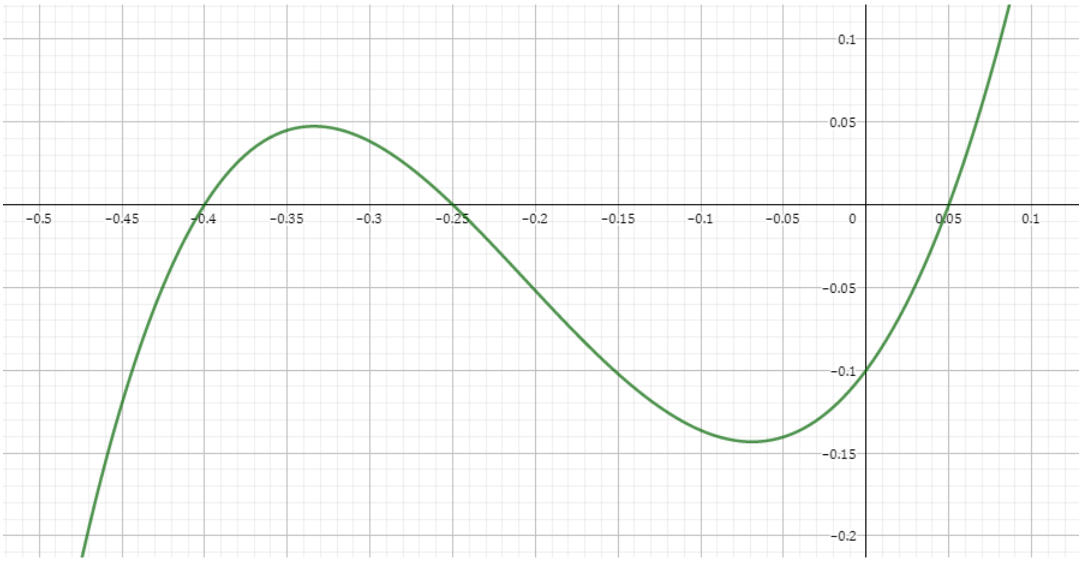

Figure 8: The graph of *f*(x).

The starting value is chosen as $x_0 = \frac{0+0.14}{2} = 0.07$.

We have $f(x) = 20x^5 + 12x^4 + 21.35x^3 + 11.9x^2 + 1.35x - 0.1$,

and thus, $f'(x) = 100x^4 + 48x^3 + 64.05x^2 + 23.8x + 1.35$.

To determine the solution of $f(x) = 0$ using the NRM, we construct the following sequence:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{20x_n^5 + 12x_n^4 + 21.35x_n^3 + 11.9x_n^2 + 1.35x_n - 0.1}{100x_n^4 + 48x_n^3 + 64.05x_n^2 + 23.8x_n + 1.35}.$$

Using this starting value $x_0 = 0.07$, we get:

$$x_1 = 0.05194,\ f(x_1) = 0.00532,$$
$$x_2 = 0.05002,\ f(x_2) = 5.7 * 10^{-5},$$
$$x_3 = 0.05,\ f(x_3) = 6.8 * 10^{-9},$$
$$x_4 = 0.05,\ f(x_4) = 1.1 * 10^{-16}.$$

Since the third and fourth iterations yield the same value up to the eighth decimal place, we can conclude that $x = 0.05$ is a solution to $f(x) = 0$. This is also consistent with the graph of $f(x)$. Thus, the depth x to which the ball is submerged underwater is approximately $0.05m = 5cm$.

Statistics is a significant subsection of mathematics where the NRM can be used to solve several related problems (Truong, et al., 2019b). Among these, estimating unknown parameters of distribution functions and regression models are two important applications. The next two sections discuss these problems in detail: Section 4 covers parameter estimation for distributions, and Section 5 addresses parameter estimation for regression models.

4 Applying the NRM to Estimate Parameters of Probability Distributions

Statistics is a scientific discipline intrinsically connected to daily life, founded on numerical data, visual representations, and observed phenomena. Within statistical analysis, probability distribution functions hold a central position (Bakouch, et al., 2014). Distribution functions are often classified based on the number of parameters they incorporate, ranging from single-parameter distributions to those involving multiple parameters. Each distribution is defined by its unique set of parameters, which characterize a family of distributions sharing a common functional form but differing in complexity based on parameter count.

For distributions characterized by a single parameter, estimation is generally straightforward, often achievable using basic algebraic manipulations, thus typically obviating the need for numerical techniques like the Newton-Raphson Method (NRM). However, for distributions constructed with two or more parameters, the mathematical formulations become significantly more complex, frequently rendering direct analytical parameter estimation infeasible. In such cases, the NRM becomes essential, as elementary operations are insufficient to derive closed-form solutions. The estimation challenge increases with the number of parameters, further necessitating numerical methods like the NRM.

In the subsequent subsections, we explore parameter estimation for probability distribution functions, addressing scenarios where the NRM is not required and others where its application is crucial. We integrate the Maximum Likelihood Estimation (MLE) method to determine the characteristic parameters (Myung, 2003). MLE is a widely adopted statistical method for estimating unknown parameters from observed data. By identifying parameter values that maximize the likelihood function, MLE provides optimal parameter estimates, which are maximum likelihood estimates. This discussion highlights situations where the NRM is necessary versus those where simpler alternatives suffice.

The process of performing MLE for estimating parameters in statistics can be briefly summarized as follows: In general, we assume that $X_1, \dots, X_n$ are random variables derived from the population X associated with the probability distribution function (PDF) $f(x, \eta)$ with η is a parameter that needs to be estimated. The likelihood function (LF) $L(\eta)$ is defined as the joint PDF of $X_1, \dots, X_n$:

$$L(\eta) = \prod_{i=1}^{n} f(x_i;\ \eta). \tag{5}$$

The value η that maximizes the likelihood function $L(\eta)$ is typically called the MLE of η, and it is usually denoted by $\hat{\eta}$ such that:

$$\hat{\eta} = \arg \sup_{\eta \in \Omega} \mathrm{L}(\eta) = \arg \sup_{\eta \in \Omega} \log(\mathrm{L}(\eta)), \tag{6}$$

where Ω is the sample space. The algorithm for MLE can be briefly described as follows:

Step 1: Suppose that $(x_1, \dots, x_n)$ is a sample of size n collected from the population X with the PDF $f(x, \eta)$.

Step 2: Compute the log-likelihood $\log(L(\eta))$ of $(x_1, \dots, x_n)$ as mentioned in Equation (5).

Step 3: Determine $\hat{\eta}$ such that $\log(L(\hat{\eta}))$ reaches its maximum value, as mentioned in Equation (6).

4.1 Distributions Defined by a Single Parameter

Probability distributions characterized by a single parameter generally exhibit relatively straightforward mathematical formulations. Consequently, MLE can often estimate this parameter directly without requiring iterative numerical techniques. Well-known examples include the Bernoulli, Poisson, Geometric, Exponential, and Bell distributions (Pho, et al., 2019). We focus on the Bernoulli and Poisson distributions to illustrate cases where MLE yields a closed-form solution, rendering the NRM unnecessary.

4.1.1 Bernoulli Distributions (BD)

The Bernoulli distribution (BD) is a fundamental discrete probability distribution that plays a vital and significant role in the statistical literature. Named in honor of the renowned Swiss mathematician Jacob Bernoulli, this distribution is characterized by its capacity to assume only two possible values: either 0 or 1. Conventionally, the value 1 is often associated with the probability of a successful outcome, whereas the value 0 represents the probability of a failure. Despite its simplicity in only taking two distinct values, the BD is remarkably versatile due to its applicability in encoding a wide range of real-world scenarios. Specifically, if an event or problem occurs, progresses, or results in a success, it can be encoded as 1; otherwise, it is encoded as 0. Consequently, this distribution is frequently encountered in diverse real-world events or problems, serving as a basis for more complex models.

Suppose $X = (X_1, \dots, X_n)$ are Bernoulli distributed random variables, i.e., $X_i \sim \text{Ber}(\eta)$, $i = 1, 2, \dots, n$. The probability mass function (PMF) of X_i can then be expressed as follows (Bobkov, 1997):

$$f\,(x_i;\,\eta) = \eta^{x_i}(1 - \eta)^{1-x_i}, \qquad \text{with } x_i \in \{0, 1\}.$$

Thus, the LF of the BD is defined by:

$$L(\eta|X) = \prod_{i=1}^{n} f(x_i;\eta) = \prod_{i=1}^{n} \eta^{x_i}(1-\eta)^{1-x_i} = \prod_{i=1}^{n} \eta^{x_i} . \prod_{i=1}^{n} (1-\eta)^{1-x_i}$$
$$= \eta^{\sum_{i=1}^{n} x_i} . (1-\eta)^{n-\sum_{i=1}^{n} x_i}.$$

Hence, we have the log-likelihood function:

$$\log\big(L(\eta|X)\big) = \log\big[\eta^{\sum_{i=1}^{n} x_i} . (1-\eta)^{n-\sum_{i=1}^{n} x_i}\big]$$
$$= \sum_{i=1}^{n} x_i . \log(\eta) + \left(n - \sum_{i=1}^{n} x_i\right) . \log(1-\eta) . \tag{7}$$

The MLE of η is the solution (root) of the following equation:

$$\frac{\partial}{\partial \eta}\log(L(\eta|X)) = 0.$$

This can be briefly rewritten as follows:

$$\frac{\sum_{i=1}^{n} x_i}{\eta} - \frac{(n - \sum_{i=1}^{n} x_i)}{1-\eta} = 0.$$

We further simplify it as: $\sum_{i=1}^{n} x_i = n\eta$.

Therefore, we obtain the following: $\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} x_i$. To verify whether $\hat{\eta}$ is the MLE or not, we need to check if the corresponding second derivative of the log-likelihood function defined in Equation (7) at $\eta = \hat{\eta}$ is negative, i.e. $\hat{\eta}$ is a maximum point:

$$\frac{\partial^2}{\partial \eta^2}\log\left(L(\eta|X)\right) = -\frac{\sum_{i=1}^{n} x_i}{\eta^2} - \frac{(n - \sum_{i=1}^{n} x_i)}{(1-\eta)^2}$$

$$= \frac{(-1+2\eta-\eta^2)\sum_{i=1}^{n} x_i - \eta^2(n - \sum_{i=1}^{n} x_i)}{\eta^2\ (1-\eta)^2} = \frac{(-1+2\eta)\sum_{i=1}^{n} x_i - \eta^2 n}{\eta^2\ (1-\eta)^2}$$

$$= \frac{(-1+2\eta)\, n\eta - \eta^2 n}{\eta^2\ (1-\eta)^2} = \frac{-n\eta(1-\eta)}{\eta^2\ (1-\eta)^2} = \frac{-n}{\eta\,(1-\eta)}.$$

Thus, the log-likelihood function in Equation (7) at $\eta = \hat{\eta}$ is always negative. This confirms that $\hat{\eta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the MLE of the BD.

4.1.2 Poisson Distributions (PD)

The Poisson distribution (PD) is a well-established discrete probability distribution named in honor of the eminent French mathematician Siméon Denis Poisson. This distribution is frequently employed to model the occurrence of events at discrete intervals and to characterize the average rate of successful occurrences of an event within a defined period. In this study, the mean value of the PD is represented by the parameter η. The PD exhibits broad applicability across diverse measurement units, including time intervals, distances, areas, and volumes. A prominent example of its real-world application lies in modeling the nuclear decay of atoms, a process that occurs randomly but with a predictable average rate.

Suppose $X = (X_1, \dots, X_n)$ are Poisson distributed random variables, i.e., $X_i \sim \text{Poi}(\eta)$, for $i = 1, 2, \dots, n$. The PMF of X_i can then be expressed as follows (Consul & Shoukri, 1984):

$$\mathbb{P}(X = x) = \frac{e^{-\eta}.\eta^{x}}{x!}, x \in \mathbb{N}; \qquad (8)$$

where $\eta > 0$, and $\mathbb{N}$ is the set of all natural numbers. The LF $L(\eta|X)$ is given by:

$$L(\eta|X) = \prod_{i=1}^{n} \frac{e^{-\eta}.\eta^{x_i}}{x_i!} = \frac{e^{-n\eta}.\eta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

The log-likelihood function can be presented as follows:

$$\log(L(\eta|X)) = \log \frac{e^{-n\eta}.\eta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} = -n\eta + \sum_{i=1}^{n} x_i \log(\eta) - \log\left(\prod_{i=1}^{n} x_i!\right).$$

To find the MLE of η, we take the first partial derivative of $\log(L(\eta|X))$ with respect to η:

$$\frac{\partial}{\partial \eta} \log L(\eta|X) = -n + \sum_{i=1}^{n} x_i \,.\frac{1}{\eta}. \quad (9)$$

The solution (root) of Equation (9) is:

$$\eta = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Next, we check the second partial derivative of $\log(L(\eta|X))$:

$$\frac{\partial^2}{\partial \eta^2} \log(L(\eta|X)) = -\sum_{i=1}^{n} x_i \,.\frac{1}{\eta^2} < 0.$$

This indicates that $\hat{\eta} = \frac{\sum_{i=1}^{n} x_i}{n}$ is the MLE of the Poisson distribution.

These examples illustrate that for some single-parameter distributions, MLE provides a simple closed-form estimate. However, even for single-parameter cases, if the likelihood equation becomes complex, numerical methods like NRM might be needed.

4.2 Distributions Defined by Two Parameters

Probability distributions characterized by two parameters often possess intricate mathematical formulations. While MLE is commonly used, deriving analytical solutions for the parameters is frequently not achievable. In such scenarios, numerical methods, particularly the NRM, become essential for approximating the MLEs. Examples include the Weibull, Zero-Inflated Poisson (ZIP), Beta, Gamma, Normal, Log-normal, Laplace, and Gumbel distributions (Pho, et al., 2019). We examine the Weibull and ZIP distributions to illustrate the complexities requiring NRM.

4.2.1 *Weibull Distributions (WD)*

The Weibull distribution (WD) is a versatile continuous probability distribution that plays a crucial role in modeling various phenomena, particularly the characteristics of event occurrences or the time intervals between events. This distribution is also frequently employed in reliability analysis and in estimating the average lifespan of devices, as evidenced by the work of Vasudeva Rao, et al. (1994).

Named in honor of the distinguished Swedish mathematician Waloddi Weibull, the WD is characterized by its flexibility and its broad applicability across diverse fields. To facilitate the calculation of the likelihood function (LF) of the WD, the probability density function (PDF) of the WD is presented as follows (Cohen, 1965):

$$f(x;\mu,\eta) = \left(\frac{\mu}{\eta}\right) x^{\mu-1}\exp\left(\frac{-x^{\mu}}{\eta}\right); \;\; x \geq 0, k > 0, \eta > 0.$$

Given $X_i \sim Wei(\eta, \mu)$, $i = 1, 2, ..., n$, and letting $X = (X_1, \dots, X_n)$, the LF of X is defined by:

$$\begin{aligned} L(\eta,\mu|X) &= \prod_{i=1}^{n}\left(\frac{\mu}{\eta}\right) x_i^{\mu-1}\exp\left(-\frac{x_i^{\mu}}{\eta}\right) \\ &= \left(\frac{\mu}{\eta}\right)^n.\left(\prod_{i=1}^{n} x_i\right)^{\mu-1}.\left[\prod_{i=1}^{n}\exp\left(-\frac{x_i^{\mu}}{\eta}\right)\right] \\ &= \left(\frac{\mu}{\eta}\right)^n.\left(\prod_{i=1}^{n} x_i\right)^{\mu-1}.\exp\left[-\sum_{i=1}^{n}\left(\frac{x_i^{\mu}}{\eta}\right)\right] \\ &= \left(\frac{\mu}{\eta}\right)^n.\left(\prod_{i=1}^{n} x_i\right)^{\mu-1}.\exp\left(\frac{-\sum_{i=1}^{n} x_i^{\mu}}{\eta}\right). \end{aligned} \tag{10}$$

The log-likelihood function can be determined by:

$$\begin{aligned} \log\big(L(\eta,\mu|X)\big) &= n\log\left(\frac{\mu}{\eta}\right) + (\mu-1)\sum_{i=1}^{n}\log(x_i) - \frac{\sum_{i=1}^{n} x_i^{\mu}}{\eta} \\ &= n\log(\mu) - n\log(\eta) + (\mu-1)\sum_{i=1}^{n}\log(x_i) - \frac{1}{\eta}\sum_{i=1}^{n} x_i^{\mu}. \end{aligned} \tag{11}$$

By taking the first partial derivatives of the function defined by Equation (11) with respect to μ and η, and then setting these expressions to zero, we obtain:

$$\frac{\partial \log\big(L(\eta,\mu|X)\big)}{\partial\mu} = \frac{\eta}{\mu} + \sum_{i=1}^{n}\log(x_i) - \frac{1}{\eta}\sum_{i=1}^{n} x_i^{\mu}\log(x_i) = 0. \tag{12}$$

and

$$\frac{\partial \log\big(L(\eta,\mu|X)\big)}{\partial\eta} = -\frac{n}{\eta} + \frac{1}{\eta^2}\sum_{i=1}^{n} x_i^{\mu} = 0. \tag{13}$$

From Equation (13), we can attain: $\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\mu}$. By using the value $\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\mu}$ substituted in Equation (12), we have

$$\frac{1}{\mu} + \frac{1}{n}\sum_{i=1}^{n}\log(x_i) - \frac{\sum_{i=1}^{n} x_i^{\mu}\log(x_i)}{\sum_{i=1}^{n} x_i^{\mu}} = 0. \tag{14}$$

As can be seen, Equation (14) is quite complicated and cannot be solved directly with basic algorithms. One of the most common methods to find μ in Equation (14) is to use the NRM, see Bhattacharya and Bhattacharjee (2010) for more details. We note that once we obtain $\hat{\mu}$, then $\hat{\eta}$ can be computed based on the formula:

$$\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{\mu}}.$$

Next, we examine the well-known ZIP distribution, which has numerous practical applications.

4.2.2 Zero-inflated Poisson Distributions (ZIPD)

The Zero-Inflated Poisson distribution (ZIPD) is a specialized discrete probability distribution that plays a critical role in statistical modeling. The ZIPD is primarily employed to model numerical datasets with a disproportionately high frequency of zero values. Such datasets are frequently encountered in various real-world scenarios, including, but not limited to, the number of earthquakes, meteorites, or rainstorms recorded in regions where these events occur infrequently. This specific type of distribution is valuable in accurately capturing the excess zeros that are not well-modeled by standard Poisson distributions.

Given $X_i \sim \text{ZIP}(\eta, \mu)$, for $i = 1, 2, ..., n$, the PMF of X_i is given by:

$$\mathbb{P}(X = x) = \begin{cases} \eta + (1-\eta)e^{-\mu} & if \;\; x = 0, \\ (1-\eta)\frac{e^{-\mu}\,\mu^x}{x!} & if \;\; x \neq 0, x \in \mathbb{N}. \end{cases} \quad (15)$$

Equation (15) is widely referenced in the literature and is the most frequently cited formula for regression analyses involving zero-inflated data. This formula was first introduced in Singh's study in 1963. To facilitate future calculations, we denote:

$$M = \sum_{i=1}^{n} I\,(x_i = 0)\,.$$

Thus, the likelihood function $L(\eta, \mu|X)$ can be briefly described as

$$L(\eta,\mu|X) = [\eta + (1-\eta)e^{-\mu}]^M \prod_{x_i \neq 0}^{} (1-\eta)e^{-\mu}\,\frac{\mu^{x_i}}{x_i!}.$$

From the above, the log-LF can be written as follows:

$$\begin{aligned} log(\text{L}(\eta,\mu|\text{X})) = \text{M}.\,log(\eta + (1-\eta).\,\text{e}^{-\mu}) + (n-M)log(1-\eta) \quad (16) \\ -(n-M).\,\mu + log(\mu).\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} log(X_i!)\,. \end{aligned}$$

By taking the first partial derivatives of $\log(L(\eta,\mu|X))$ with respect to (w.r.t.) η:

$$\frac{\partial}{\partial\eta}\log(L(\eta,\mu|X)) = M\frac{\partial}{\partial\eta}\log(\eta + (1-\eta)e^{-\mu}) + (n-M)\frac{\partial}{\partial\eta}\log(1-\eta) \quad (17)$$

$$= M\frac{1-e^{-\mu}}{\eta+(1-\eta)e^{-\mu}} - (n-M)\frac{1}{1-\eta}.$$

The first partial derivatives of $\log(L(\eta,\mu|X))$ w.r.t μ has the following form:

$$\frac{\partial}{\partial\mu}\log(L(\eta,\mu|X)) = M\frac{\partial}{\partial\mu}\log(\eta+(1-\eta).e^{-\mu}) + (n-M) + \sum_{i=1}^{n} x_i \frac{\partial}{\partial\mu}\log(\mu) \qquad (18)$$

$$= M\frac{1-e^{-\mu}}{\eta+(1-\eta)e^{-\mu}} - (n-M) + \sum_{i=1}^{n} x_i \,.\frac{1}{\mu}.$$

It has been observed that using the MLE to determine the parameters of the ZIPD requires solving a system of equations formed by Equations (17) and (18). This system of equations is highly complex, making it nearly impossible to find solutions quickly using conventional methods. One intelligent and popular approach to solving this problem is the NRM (Wagh & Kamalja, 2018).

5 Applying the NRM to Estimate Parameters of Regression Models

Regression models constitute a fundamental component of statistical analysis and are extensively explored in specialized statistical literature. A prevalent approach for developing these models involves utilizing corresponding probability distribution functions as a basis. Like the categorization of distribution functions, regression models are often classified based on the number of parameter vectors they incorporate. However, the process of estimating the characteristic vector parameters of regression models differs fundamentally from the estimation of parameters in probability distributions. Given that the general formulation of regression models includes both independent and dependent variables, most regression models necessitate the application of numerical methods such as the Newton-Raphson Method (NRM) to estimate their parameter vectors. Fortunately, with the rapid advancements in science and technology, this complex process has been considerably simplified by the availability of built-in functions in various computational software packages. Among these tools, R statistical software stands out as a widely used, freely accessible platform that offers a diverse range of functions—including maxLik, optim, and nleqslv—designed to compute Maximum Likelihood Estimates (MLEs) efficiently (Truong, et al., 2019a).

For clarity and organization, we will categorize regression models into two distinct types: those constructed with a single parameter vector and those constructed with two parameter vectors. This approach facilitates a structured analysis of the complexities involved in each model type. We will focus our attention on the logistic regression model and the Zero-Inflated Poisson regression model, to streamline the analysis and provide representative examples of each category. By examining these two models, we will demonstrate the application of the NRM in practical regression analysis and offer insight into the broader applicability of this technique.

5.1 Regression Models Constructed with a Single Parameter Vector

Regression models formulated using a single parameter vector, with the logistic model (LM) serving as a prominent example, are widely recognized and have been extensively analyzed in the statistical literature. The LM, initially introduced by Cox (1958), is particularly well-suited for modeling binary outcome variables, wherein the dependent variable assumes one of two values: either 0 or 1. Despite this binary nature, the LM can be effectively applied across a diverse range of real-world scenarios by appropriately encoding relevant events. For instance, phenomena such as occurrences, successes, achievements, or specific conditions—such as being sick, satisfied, or exhibiting a particular trait—can be coded as 1. Conversely, their absence or the negation of these states is typically coded as 0. This straightforward encoding mechanism renders the LM highly versatile and applicable to numerous real-life situations where binary outcomes are interesting.

We first denote Y as a binary (dichotomous) outcome variable, and the covariates are denoted as X and Z. The general formula of the LM is given as follows (Pho & McAleer, 2021):

$$P(Y_{\mathrm{i}} = 1|X_i, Z) = \mathrm{H}(\mu_0 + \mu_1^{\mathrm{T}} X_{\mathrm{i}} + \mu_2^{\mathrm{T}} Z_i) = \mathrm{H}(\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}) = \frac{e^{\mu^T \mathcal{X}_i}}{1+e^{\mu^T \mathcal{X}_i}}, \tag{19}$$

where $H(x) = (1 + \mathrm{e}^{-x})^{-1}$ is the logistic function.

We let $\mathcal{X}_i = (1, X_i^T, Z_i^T)^T$, and $\mu = (\mu_0, \mu_1^T, \mu_2^T)^T$ is a parameter vector that we need to estimate.

The log-LF of μ is $\ell(\mu) = \log[L(\mu)] = \sum_{i=1}^{n} \ell_i(\mu)$ described as follows:

$$\begin{aligned}
\ell(\mu) &= \log[L(\mu)] \\
&= \log \prod_{i=1}^{n} \left[\left(\frac{\mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}}{1 + \mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}} \right)^{Y_i} \left(1 - \frac{\mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}}{1 + \mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}} \right)^{1-Y_i} \right] \\
&= \sum_{i=1}^{n} Y_i \left[\log\left(\mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}\right) - \log\left(1 + \mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}\right) \right] - \sum_{i=1}^{n} (1 - Y_i) \left[\log\left(1 + \mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}\right) \right] \\
&= \sum_{i=1}^{n} \left[Y_i (\mu^T \mathcal{X}_i) - \log\left(1 + e^{\mu^T \mathcal{X}_i}\right) \right].
\end{aligned}$$

The estimating score function (ESF) is the first derivative of the log-LF, which can be briefly stated as follows:

$$U_{n(\mu)} = \frac{1}{\sqrt{n}} \frac{\partial \ell(\mu)}{\partial \mu} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_i \left(Y_i - \frac{e^{\mu^T \mathcal{X}_i}}{1+e^{\mu^T \mathcal{X}_i}} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(\mu), \tag{20}$$

where

$$S_{\mathrm{i}}(\mu) = \mathcal{X}_i \left(Y_i - \frac{\mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}}{1+\mathrm{e}^{\mu^{\mathrm{T}} \mathcal{X}_{\mathrm{i}}}} \right). \tag{21}$$

We note that $E[\boldsymbol{U_n}(\mu)] = 0$, meaning that $\boldsymbol{U_n}(\mu)$ is an unbiased ESF. Additionally, we can get the MLE $\hat{\mu}$ of μ by solving $\boldsymbol{U_n}(\mu) = 0$.

Nevertheless, using MLE to find a specific solution is challenging because it depends on many variables. In such cases, the NRM should be applied to solve this problem. Researchers can use built-in functions. available in R statistical software to save time on writing code, such as maxLik, optim, nleqslv, and others (Truong, et al., 2019a). We can see that even for the MLE of the simplest model, the LM, it is necessary to use the NRM. In the following subsection, we will consider the most popular model in the family of regression models built through two parameter vectors.

5.2 Regression Models Constructed with Two Parameter Vectors

Among regression models formulated using two parameter vectors, the Zero-Inflated Poisson (ZIP) model stands out as one of the most widely employed and frequently analyzed in the statistical literature. This model proves particularly effective for simulating datasets comprising natural numbers with a disproportionately high incidence of zero values. Illustrative examples include data concerning the number of earthquakes, meteorites, or rainstorms in regions where these events are relatively infrequent. The ZIP model was initially introduced by Lambert (1992) and has since found widespread application in both theoretical and practical contexts. Its unique ability to accommodate excess zero counts makes it suitable for modeling diverse phenomena where standard Poisson models are insufficient.

To facilitate ease of comprehension for the reader, we will provide a concise summary of the most salient formulas associated with the ZIP model. Consistent with the formulation presented in Lukusa, et al. (2016), the log-likelihood function (log-LF) of the ZIP model can be briefly described as follows:

$$\begin{aligned}\log[L\ (\lambda|\mathcal{X})] &= \log\ [L\ (\eta,\mu|\mathcal{X})] \\ &= \sum_{i=1}^{n}\log\left\{I(Y_i=0)\{\log[H(\eta^T\mathcal{X}_i)]-\log\left[H\left(\eta^T X_i+\exp(\mu^T\mathcal{X}_i)\right)\right]\}\right\} \\ &+\sum_{i=1}^{n}\log\{I(Y_i>0)\{\log[1-H(\eta^T X_i)]+[Y_i\mu^T X_i-\exp(\mu^T X_i)-\log(Y_i!)]\}\}.\end{aligned} \tag{22}$$

Here, $I(.)$ is an indicator function, Y is a variable that takes only natural values, and the covariates are denoted as X and Z. We define $\mathcal{X}=(1,X^T,Z^T)^T$, and $\lambda=(\eta^T,\mu^T)^T$ which is a parameter vector that need to be estimated. The ESF is the first derivative of the log-LF, which can be briefly written as follows:

$$U_n(\lambda)=\frac{1}{\sqrt{n}}\frac{\partial\ell(\lambda)}{\partial\lambda}=\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{\partial\ell(\lambda)}{\partial\eta}\\ \frac{\partial\ell(\lambda)}{\partial\mu}\end{pmatrix}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_i(\lambda), \tag{23}$$

where

$$S_i(\lambda)=\partial\ell_i(\lambda)/\partial\lambda=\left(S_{i1}^T(\lambda),S_{i2}^T(\lambda)\right)^T,i=1,\dots,n;$$

$$S_{i1}(\lambda)=\frac{\partial\ell_i(\lambda)}{\partial\eta}=\mathcal{X}_iH\left(\eta^T\mathcal{X}_i+\exp(\mu^T\mathcal{X}_i)\right)\left[I(Y_i=0)-\frac{H(\eta^T\mathcal{X}_i)}{H\left(\eta^T\mathcal{X}_i+\exp(\mu^T\mathcal{X}_i)\right)}\right], \tag{24}$$

$$S_{i2}(\lambda)=\frac{\partial\ell_i(\lambda)}{\partial\mu}=\mathcal{X}_i\{Y_i-[1-H(\eta^T\mathcal{X}_i)\exp(\mu^T\mathcal{X}_i)]\}+\exp(\mu^T\mathcal{X}_i)S_{i1}(\lambda). \tag{25}$$

We note that it is not difficult to prove that $E\left[U_{F,n}(\lambda)\right]=0$ (Lukusa, et al., 2016), hence $U_{F,n}(\lambda)$ is an unbiased ESF. This leads to the idea that we can get the MLE $\hat{\lambda}$ of λ by solving $U_n(\lambda)=0$. Solving this equation is challenging due to the dependence on many factors. In such cases, the NRM should be applied to solve the problem completely. Researchers can use functions available in R statistical software, such as maxLik, optim, nleqslv, and others (Truong, et al., 2019a). It can be seen that finding the MLE of regression models built through one or two parameter vectors often requires numerical methods like the NRM.

6 Other Applications of the NRM in Decision Sciences

6.1 Applied Mathematics

Applied Mathematics (AM) is a highly versatile and practical discipline that employs mathematical methods and models to address complex challenges across diverse fields. These fields include, but are not limited to, science, engineering, economics, finance, insurance, computer science, and various industries. The importance of AM is particularly pronounced in contemporary society, especially within the context of the Fourth Industrial Revolution (Industry 4.0). While this paper has, to this point, discussed specific applications of the Newton-Raphson Method (NRM), it is crucial to highlight the broader significance of NRM within the domain of AM. The NRM is not only applicable to solving moderately complex equations and systems of equations, as previously demonstrated but also extends to tackling considerably more complex mathematical problems across various sub-domains of AM.

It should also be noted that the successful application of the NRM to such intricate problems often necessitates modifications and enhancements to its general formulation, tailoring the method to the specific characteristics of each problem. Some notable examples of complex and intriguing problems within AM, where the NRM has proven useful, include semi-smooth block-triangular systems of equations, quadratic matrix equations, constrained linear least-squares problems, nonsmooth equations, initial value problems, certain singular equations, underdetermined inverse problems, parameter estimation for dynamical systems, power flow problems, nonlinear equations, generalized absolute value equations, and stochastic control problems.

These intricate problems have been the subject of detailed study in the literature. For example, Smietanski (2007) introduced a generalized Jacobian-based NRM for semi-smooth block-triangular systems of equations. Further, Long, et al. (2008) presented an improved NRM incorporating exact line searches to address quadratic matrix equations. Morini, et al. (2010) proposed a reduced NRM for constrained linear least-squares problems. Smietanski (2011) also developed quadrature-based versions of the generalized NRM for addressing nonsmooth equations. Ezquerro, et al. (2012) explored the application of the NRM to initial value problems. Gatilov (2014) investigated using low-rank approximations of the Jacobian matrix within the NRM to address certain singular equations. Improvements to the cluster NRM for underdetermined inverse problems were discussed by Gaudreau, et al. (2015). Moreover, the application of the NRM to parameter estimation for dynamical systems was presented by Xu (2015). Sereeter, et al. (2019) provided a comparative analysis of NRM solvers for power flow problems. Zhou and Zhang (2020) introduced a modified Broyden-like quasi-NRM for nonlinear equations. Zhou, et al. (2021) offered an NRM-based matrix splitting method for a generalized absolute value equation. Lastly, Gobet and Grangereau (2022) demonstrated the use of the NRM for stochastic control problems, and Hassan and Moghrabi (2023) introduced a modified NRM for unconstrained optimization problems.

6.2 Finance

Finance, as an economic construct, reflects the distribution of social wealth through values generated during formation and creation. Its primary function is to allocate monetary resources among economic agents to facilitate the achievement of their objectives under specified conditions. Furthermore, finance illustrates the economic relationships that emerge in allocating financial resources through the generation and utilization of capital. The Newton-Raphson Method (NRM) is frequently employed to identify optimal solutions in various financial contexts. Portfolio optimization is a particularly common area within finance where the NRM finds considerable application.

Numerous studies have explored the application of the NRM to various problems within the financial industry. For example, Chen (2000) introduced using the NRM for stochastic programming in finance. Coleman, et al. (2003) utilized the NRM to price American options. Additionally, Agarwal, et al. (2006) developed algorithms for portfolio management based on the NRM. Mudzimbabwe and Vazquez (2016) presented the NRM for option pricing in scenarios with liquidity switching, among other applications. These studies collectively highlight the versatility of the NRM in addressing diverse financial challenges.

6.3 Education

Education is a fundamental learning process through which knowledge, skills, and values are transmitted from one generation to the next utilizing teaching, training, or research. In its most conventional interpretation, education occurs under the guidance of educators; however, it can also be pursued through self-directed learning. Any experience or activity significantly influencing an individual's cognitive processes, affective responses, or behavioral patterns can be considered educational. Education is a cornerstone of societal advancement, particularly in the current era of industrialization. Higher education plays a critical role in providing students with access to cutting-edge science and technology while guiding them toward making informed career decisions for the future.

Mathematics is an indispensable element of higher education, functioning as a bridge to all other scientific disciplines. The NRM is recognized as a powerful mathematical tool, lauded for its accuracy, speed, and efficiency in finding optimal solutions. Software developers and practitioners extensively utilize it across numerous fields. In Vietnam, the NRM is incorporated into the curricula of most universities and is taught to mathematics majors and students from other disciplines. Notable institutions that include the NRM in their programs are Can Tho University, Vinh University, Sai Gon University, Ton Duc Thang University, and others. The widespread integration of the NRM into educational programs reflects its recognized importance and utility in various academic and professional domains.

7 Conclusions

This paper has provided a comprehensive overview of the Newton-Raphson Method (NRM) and illustrated its practical applications across a range of scientific disciplines, with a particular focus on its instrumental role in Decision Sciences. The NRM is a widely recognized and remarkably effective numerical technique for identifying optimal solutions. The primary objective of this work was to provide researchers and practitioners with a detailed analysis of the theoretical foundations and the diverse applications of the NRM, emphasizing its relevance and versatility in addressing real-world problems. While this study emphasizes the application aspects of the NRM, we note that theoretical advancements, including algorithmic refinements, are not explored in depth. These areas represent important directions for future research.

This article does not introduce novel mathematical proofs or overly complex derivations; rather, it presents a comprehensive overview of the NRM and practical demonstrations of its utility across diverse domains. The applications explored include two real-world case studies: optimizing loudspeaker placement for effective COVID-19 public health communication and determining the submersion depth of a floating object in the water. We then extended the discussion to encompass the NRM's application in probability and statistics, specifically focusing on parameter estimation for probability distributions and regression models. Finally, the paper highlighted the NRM’s relevance across various facets of Decision Sciences, such as applied mathematics, finance, and education, thereby underscoring its broad impact and adaptability.

To the best of our knowledge, this study is unique in its combined breadth and depth of coverage, encompassing both a theoretical overview and practical applications of the NRM across diverse fields. This research contributes to the existing body of knowledge by synthesizing information on the NRM, bridging the gap between theory and practice, and providing a valuable resource for researchers and practitioners. This work empowers readers to readily grasp the NRM’s core principles and effectively apply this powerful method to a wide range of challenges. Furthermore, it fosters a deeper appreciation for the method's versatility and encourages further exploration and adaptation of the NRM within various domains.

References

Agarwal, A., Hazan, E., Kale, S., & Schapire, R. E. (2006). Algorithms for portfolio management based on the Newton method. *In Proceedings of the 23rd international conference on Machine learning*, 9-16.

Bakouch, H. S., Jazi, M. A., & Nadarajah, S. (2014). A new discrete distribution. *Statistics*, 48(1), 200-240.

Bhattacharya, P., & Bhattacharjee, R. (2010). A study on Weibull distribution for estimating the parameters. *Journal of Applied Quantitative Methods*, 5(2), 234-241.

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects—State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, 22(1), 37-68.

Bobkov, S. G. (1997). Some extremal properties of the Bernoulli distribution. *Theory of Probability and its Applications*, 41(4), 748-755.

Chen, X. (2000). Newton-type methods for stochastic programming. *Mathematical and Computer Modelling*, 31(10), 89-98.

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics*, 7(4), 579-588.

Coleman, T. F., Li, Y., & Verma, A. (2003). A Newton method for American option pricing. *Cornell University*.

Consul, P. C., & Shoukri, M. M. (1984). Maximum likelihood estimation for the generalized Poisson distribution. *Communications in Statistics-Theory and Methods*, 13(12), 1533- 1547.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2), 215-232.

Dedieu, J. P. (2015). Newton-Raphson Method. *Encyclopedia of Applied and Computational Mathematics*, *6*(7), 1023-1028.

Doikov, N., & Nesterov, Y. (2024). Gradient regularization of Newton method with Bregman distances. *Mathematical programming*, 204(1), 1-25.

Dwyer, J., Barnard, R., Cook, D., & Corte, J. (2009). Iteration of complex functions and Newton's method. *Australian Senior Mathematics Journal*, *23*(1), 9-16.

Ezquerro, J. A., Gonzalez, D., & Hernandez, M. A. (2012). Majorizing sequences for Newton's method from initial value problems. *Journal of Computational and Applied Mathematics*, 236(9), 2246-2258.

Gatilov, S. Y. (2014). Using low-rank approximation of the Jacobian matrix in the Newton–Raphson method to solve certain singular equations. *Journal of Computational and Applied Mathematics*, 272, 8-24.

Gaudreau, P., Hayami, K., Aoki, Y., Safouhi, H., & Konagaya, A. (2015). Improvements to the cluster Newton method for underdetermined inverse problems. *Journal of Computational and Applied mathematics*, 283, 122-141.

Gobet, E., & Grangereau, M. (2022). Newton method for stochastic control problems. *SIAM Journal on Control and Optimization*, 60(5), 2996-3025.

Gutkin, E. (2012). Capillary floating and the billiard ball problem. *Journal of Mathematical Fluid Mechanics*, 14, 363-382.

Hassan, B. A., & Moghrabi, I. A. (2023). A modified secant equation quasi-Newton method for unconstrained optimization. *Journal of Applied Mathematics and Computing*, 69(1), 451-464.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.

Long, J. H., Hu, X. Y., & Zhang, L. (2008). Improved Newton's method with exact line searches to solve quadratic matrix equation. *Journal of Computational and Applied Math- ematics*, 222(2), 645-654.

Lukusa, T. M., Lee, S. M., & Li, C. S. (2016). Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. *Metrika*, 79, 457-483.

Morini, B., Porcelli, M., & Chan, R. H. (2010). A reduced Newton method for constrained linear least-squares problems. *Journal of Computational and Applied Mathe- matics*, 233(9), 2200-2212.

Mudzimbabwe, W., & Vazquez, C. (2016). Newton methods for option pricing with liquidity switching. *Numerical Methods for Scientific Computations and Advanced Applications (NMSCAA'16)*, 58-62.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90-100.

Na, S., Dereziński, M., & Mahoney, M. W. (2023). Hessian averaging in stochastic Newton methods achieves superlinear convergence. *Mathematical Programming*, *201*(1), 473-520.

Pho, K. H. (2022). Improvements of the Newton–Raphson method. *Journal of Computational and Applied Mathematics*, 408, 114106.

Pho, K. H., Ho, T. D. C., Tran, T. K., & Wong, W. K. (2019). Moment Generating Function, Expectation and Variance of Ubiquitous Distributions with Applications in Decision Sciences: A Review. *Advances in Decision Sciences*, 23(2), 65-150.

Pho, K. H., & McAleer, M. (2021). Specification and estimation of a logistic function, with applications in the sciences and social sciences. *Advances in Decision Sciences*, (2), 1-30.

Sereeter, B., Vuik, C., & Witteveen, C. (2019). On a comparison of Newton–Raphson solvers for power flow problems. *Journal of Computational and Applied Mathematics*, 360, 157- 169.

Singh, S. (1963). A note on inflated Poisson distribution. *Journal of the Indian Statistical Association*, 1, 140-144.

Smietanski, M. J. (2007). A generalized Jacobian based Newton method for semismooth block-triangular system of equations. *Journal of Computational and Applied Mathematics*, 205(1), 305-313.

Smietanski, M. J. (2011). Some quadrature-based versions of the generalized Newton method for solving nonsmooth equations. *Journal of Computational and Applied Mathematics*, 235(17), 5131-5139.

Truong, B. C., Nguyen, V. B., Truong, H. V., & Ho, T. D. C. (2019a). Comparison of optim, nleqslv and MaxLik to estimate parameters in some of regression models. *Journal of Advanced Engineering and Computation*, 3(4), 532-550.

Truong, B. C., Van Thuan, N., Hau, N. H., & McAleer, M. (2019b). Applications of the Newton-Raphson method in decision sciences and education. *Advances in Decision Sciences*, (4), 1-28.

Tuan, B. A., Pho, K. H., Pan, S. H., & Wong, W. K. (2022). Applications in Sciences in the prevention of COVID-19. *Advances in Decision Sciences*, 26(4), 1-16.

Vasudeva Rao, A., Dattatreya Rao, A. V., & Narasimham, V. L. (1994). Asymptotically optimal grouping for maximum likelihood estimation of Weibull parameters. *Communications in Statistics-Simulation and Computation*, 23(4), 1077-1096.

Wagh, Y. S., & Kamalja, K. K. (2018). Zero-inflated models and estimation in zero-inflated Poisson distribution. *Communications in Statistics-Simulation and Computation*, 47(8), 2248-2265.

Xu, L. (2015). Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *Journal of Computational and Applied Mathematics*, 288, 33-43.

Zhou, H. Y., Wu, S. L., & Li, C. X. (2021). Newton-based matrix splitting method for generalized absolute value equation. *Journal of Computational and Applied Mathematics*, 394, 113578.

Zhou, W., & Zhang, L. (2020). A modified Broyden-like quasi-Newton method for nonlinear equations. *Journal of Computational and Applied Mathematics*, 372, 112744.