

Basic Framework of Hypothesis Testing and Its Application in Maximum Likelihood Function and Regression Models

Weitong Sun

*Department of Mathematics and Physics, Xi'an Jiaotong Liverpool University, Suzhou, China
Weitong.Sun22@student.xjtlu.edu.cn*

Abstract: The research field of this essay is probability and statistics, mainly focusing on the content of hypothesis testing. This paper discusses the most basic framework and mathematical explanation behind hypothesis testing by introducing the concept of sample space. It also discusses how to construct hypothesis testing. The two testing methods are the following: one is to calculate test statistics, and the other is to calculate the corresponding p-value. It then makes a judgment based on the calculated test statistics and p-value, whether null hypothesis need to be refused or not. The main aim is to solve problems of maximum likelihood estimation and how to reasonably test and summarize coefficients in linear regression models through the explanation and utilization of hypothesis testing. The comparative argumentation method is used. The conclusion of the paper is that through hypothesis testing and Lagrangian multiplier testing, it is possible to test whether there are specific constraints in the model and quickly test whether a certain term in the regression model is necessary or not. The F test can also be used to verify whether multiple coefficients can coexist.

Keywords: Hypothesis test, Lagrangian multiplier test, regression model, p-value.

1. Introduction

In the field of statistics, hypothesis testing is an important research method when faced with specific scientific questions. These examples include determining whether the survival rate of wheat grown on farms is higher than 50%, and whether the average lifespan of cars exceeds 10 years. Many scientific studies today involve hypothesis testing in analysis.

Hypothesis testing plays a crucial role in statistical signal processing, particularly in evaluating the significance of transformations applied to complex data. The quaternion fractional Fourier transform (1DQFRFT) provides a novel approach to analyzing multidimensional signals by bridging probability theory and signal processing. By utilizing 1DQFRFT, hypothesis testing can be employed to assess whether transformed signal characteristics exhibit significant statistical differences, aiding in classification and anomaly detection. This method enhances the robustness of statistical inference in noisy environments, making it valuable for biomedical signal processing and communication systems [1]. Hypothesis testing plays a crucial role in assessing probabilistic models, particularly in risk measures and game theory. The researchers extend the use of convex-ranged probabilities to risk assessment frameworks and investigate its uniqueness. By refining probability structures, hypothesis testing can be used to validate assumptions about risk measures, ensuring robustness in financial modeling. Additionally, in game theory, hypothesis testing helps evaluate strategic decision-making

under uncertainty. These findings contribute to a deeper understanding of statistical inference in economic and financial systems, where probability-based decision-making plays a fundamental role [2]. Hypothesis testing is essential in evaluating treatment effects in clinical trials and observational studies. By taking into account confounding variables, the adjusted win ratio (AWR) with inverse probability of treatment weighting (IPTW) improves statistical inference. By applying hypothesis testing, researchers can determine whether the treatment effect observed through AWR is statistically significant. This approach ensures that treatment comparisons account for baseline imbalances, leading to more reliable conclusions. The integration of IPTW with hypothesis testing strengthens causal inference, making it a valuable tool in medical research for assessing the effectiveness of interventions [3]. Hypothesis testing serves as a fundamental bridge between statistics and probability theory, providing a rigorous framework for making statistical inferences. The connection between mathematics, probability, and statistics is evident in the formulation of null and alternative hypotheses, significance levels, and p-values. Probability distributions underpin the theoretical foundation of hypothesis testing, allowing researchers to quantify uncertainty and assess statistical significance. This interplay ensures that hypothesis testing remains a critical tool for data-driven decision-making in various disciplines, from scientific research to economics. Understanding these mathematical principles enhances the robustness and accuracy of statistical inference [4].

This essay will analysis the background about hypothesis testing and discuss how hypothesis testing make role in maximum likelihood estimation and linear regression model.

2. Basic framework of hypothesis testing

Hypothesis testing is crucial in quantitative research in social sciences. The standard structures of research papers are usually to propose a possible positive or negative relationship between two variables through theory, and then report the correlation measure between these two variables on the basis of controlling for other possible related factors. Researchers will also conduct a 'null hypothesis' test, which assumes that there is no correlation between these two variables. If the estimated parameters conform to the expected sign and are greater than twice their standard error, people can consider this result to support the theory behind the hypothesis. A hypothesis is a claim regarding specific characteristics (such as mean or variance) of an unknown population distribution. The framework of hypothesis testing consists of two hypotheses: alternative hypothesis (H_A) and null hypothesis (H_0). The significance of hypothesis testing is demonstrated by the fact that it is used to form inferences based on observable data and determine which of the two hypotheses is right [5].

Assuming a sample X_n is chosen at random from the population distribution $f(x, \theta)$, where θ is a parameter which is unknown. In hypothesis testing, parameter space Θ is divided into two disjoint and completely covered subsets Θ_0 and Θ_α i.e., $\Theta_0 \cap \Theta_\alpha = \emptyset$ and $\Theta_0 \cup \Theta_\alpha = \Theta$. The problem is to determine which of these two subsets the true value of Θ belongs to. Obviously, the basic framework of hypothesis testing can be expressed in mathematical language as follows:

$$\mathbb{H}_0: \theta \in \Theta_0 \text{ against } \mathbb{H}_A: \theta \in \Theta_\alpha \quad (1)$$

Subsequently, the author wants to investigate how to use test statistics for hypothesis testing and obtain conclusions corresponding to the data set. The test statistics measure whether the null hypothesis is suitable, and information about the differences between the null hypothesis and the data is also included in the test statistics. The sample space Ω is divided into two disjoint parts: Ω_0 and Ω_1 , where Ω_0 called acceptance region and Ω_1 named critical region or rejection region. Mathematical reasoning shows that Ω_0 must belong to some set $\mathcal{F}(\sigma - \text{field})$, which implies that Ω_1 must also satisfy corresponding conditions, as Ω_0 and Ω_1 are complementary. On the one hand, there is insufficient evidence to reject H_0 , If test statistics calculated from sample data falls within this

region, H_0 is accepted. On the other hand, if test statistics falls within the rejection region, that means reject H_0 .

For example, people can consider about this hypothesis test: $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. Then one can choose a test statistic T_n to separate the sample space of X into two mutually exclusionary and exhaustive regions: $A_n(c) = \{X: T_n \leq c\}$ and $A_n(c) = \{X: T_n > c\}$. The first region, $A_n(c)$, is the acceptance region, and the second, $A_n(c)$, is the rejection region. The cutoff point c is called the critical value. That means, if find that statistics T_n is larger than critical value c , then reject the null hypothesis. Consider another example: Suppose random sample X is follow an IID normally distribution with known parameter μ and unknown parameter σ^2 , hypothesis test is $H_0: \mu = \mu_0$ vs. $H_A: \mu \neq \mu_0$, μ_0 is a known number. Here, $\Theta_0 = \{\mu_0\}$ contains one parameter value μ_0 and Θ_a contains all parameter values on the real line R except μ_0 . Under the null hypothesis,

$$T_n(X_n) = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad (2)$$

is an appropriate test statistics in this issue. The next step is to construct rejection region and acceptance region, $A_n = \{X_n: |T_n(X_n)| \leq z_{\frac{\alpha}{2}}\}$ and $R_n = \{X_n: |T_n(X_n)| > z_{\frac{\alpha}{2}}\}$, where the critical value $z_{\frac{\alpha}{2}}$ is equals to 1.96 at 5% significant level. The acceptance and rejection zones are plotted in Figure 1 according to the normal test statistic value [6].

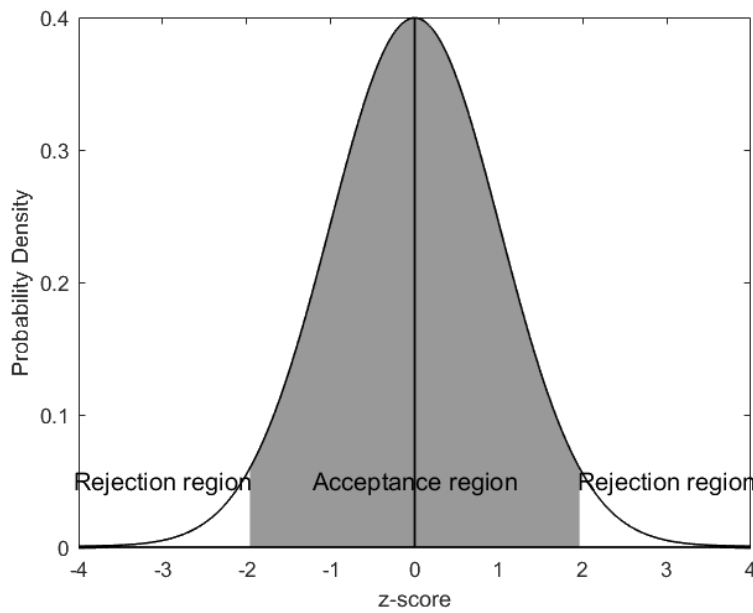


Figure 1: illustration of rejection region and acceptance region

The above is to conduct hypothesis testing by constructing test statistics, and another method is to calculate p-value. The p-value is a statistical measure based on specific assumptions and sample observations, serving as a tool for statistical inference. From the perspective of tool users, the p-value reflects the degree of improbability of the observed data under the null hypothesis. The smaller the p-value, the less likely observed returns are to occur under the null hypothesis. This suggests that there is more evidence against it and, to some extent, justifies rejecting the null hypothesis. From an application standpoint, a smaller p-value indicates that the findings are more significant and the

results are more meaningful. Overall, the likelihood of seeing the current sample result or a more extreme result under the zero-hypothesis assumption is represented by the p-value. For example, if null hypothesis is that the probability of winning a lottery ticket is extremely low, but in reality, a person is observed to have won, then there is reason to doubt the correctness of null hypothesis [7]. This is the p-value's fundamental reasoning. By defining $T(X)$ is a test statistic, the more sufficient the basis for H_1 's truth, the smaller the value of $T(X)$. Therefore, for the sample observation value x , p-value of this test will be defined as:

$$p(x) = \sup_{\theta \in \theta_0} P_{\theta}(T(X) \leq T(x)) \quad (3)$$

Through this definition, one can understand the basic elements of calculating p-value: sample data and probability distribution are required. First, the data set is used to produce test statistics, and the assumed probability distribution is used to determine the p-value. Because sample data is random and variable, p-value is a random variable, and its size includes the degree to which the null hypothesis is rejected using sample data. Usually, p-values can be summarized to varying degrees based on different intervals, There is no evidence against the null hypothesis if the p-value is greater than 0.1; weak evidence against the null hypothesis is found if the p-value is less than 0.05; moderate evidence against the null hypothesis is found if the p-value is between 0.01 and 0.05; strong evidence against the null hypothesis is found for the interval (0.001, 0.01); and very strong evidence against the null hypothesis is found if the p-value is less than or equal to 0.001.

3. Application in maximum likelihood function and regression models

Subsequently, the author wants to investigate some applications of hypothesis testing in maximum likelihood estimation and linear regression. First of all, this part will focus on an important test in maximum likelihood estimation field, which is Lagrangian multiplier test (LM test). The aim of LM test is to test whether the constrained maximum likelihood estimate (MLE) deviates significantly from the unconstrained MLE. In the case where the maximum likelihood estimation satisfies the regularity condition, LM test is asymptotically equivalent to Wald test and LR test, and both of them have the property of asymptotic local optimal invariant test. However, in comparison, the LM test has the smallest computational cost. Therefore, many popular testing methods are based on the LM framework, such as "Wu Hausman's test on unobserved variables" and "Kmenta's test on Cobb Douglas production function [8].

The sample framework of LM test is about a following hypothesis test:

$$H_0: g(\theta) = 0 \quad (4)$$

against

$$H_A: g(\theta) \neq 0 \quad (5)$$

where $g(\theta)$ is a q-dimensional function of parameter vector θ , and q equals to the number of restrictions. For instance, if set $\theta = \theta_0$ in null hypothesis, $g(\theta)$ will be $g(\theta) = \theta - \theta_0$ and q equals to 1. Suppose there exist an IID random sample X^n from distribution $f(x, \theta_0)$, where θ_0 is an unknown parameter. Consider about normalized log-likelihood function:

$$\hat{l}(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(x_i, \theta) \quad (6)$$

The author shall define constrained MLE $\hat{\theta}_n = \operatorname{argmax}_{\theta \in \theta_0} \hat{l}(\theta)$ under $g(\theta)$, and also define the Lagrangian function

$$L(\theta, \lambda) = \hat{l}(\theta) + \lambda'g(\theta) \quad (7)$$

where λ is Lagrangian multiplier. Let $\hat{\lambda}_n$ be the max value of λ , subsequently the first order conditions (FOC) is

$$\frac{\partial L(\tilde{\theta}_n, \tilde{\lambda}_n)}{\partial \theta} = \frac{\partial \hat{l}(\tilde{\theta}_n)}{\partial \theta} + G(\tilde{\theta}_n)' \tilde{\lambda}_n = 0 \quad (8)$$

and

$$\frac{\partial L(\tilde{\theta}_n, \tilde{\lambda}_n)}{\partial \lambda} = g(\tilde{\theta}_n) = 0 \quad (9)$$

where $G(\theta)$ is the derivative of $g(\theta)$. By solving this equation one will obtain that

$$G(\tilde{\theta}_n)' \tilde{\lambda}_n = -\frac{\partial \hat{l}(\tilde{\theta}_n)}{\partial \theta} = -\frac{\partial \hat{l}(\theta_0)}{\partial \theta} - \frac{\partial^2 \hat{l}(\tilde{\theta}_\alpha)}{\partial \theta \partial \theta'} (\tilde{\theta}_n - \theta_0) \quad (10)$$

where $\tilde{\theta}_\alpha$ is a value between $\tilde{\theta}_n$ and θ_0 . In addition, when introducing hessian matrix and rewriting this equation, one will obtain that

$$\sqrt{n} \tilde{\lambda}_n = -[G(\tilde{\theta}_b) \hat{H}(\tilde{\theta}_\alpha)^{-1} G(\tilde{\theta}_n)']^{-1} G(\tilde{\theta}_b) \hat{H}(\tilde{\theta}_\alpha)^{-1} \sqrt{n} \frac{\partial \hat{l}(\theta_0)}{\partial \theta} \quad (11)$$

in which $\tilde{\theta}_b$ is a value between $\tilde{\theta}_n$ and θ_0 . By using the central limit theorem, Slutsky theorem, one can define

$$LM = n \tilde{\lambda}_n' G(\tilde{\theta}_n) [-\hat{H}(\tilde{\theta}_n)^{-1}]^{-1} G(\tilde{\theta}_n)' \tilde{\lambda}_n \xrightarrow{d} \chi_q^2 \quad (12)$$

as n goes to infinity. Obviously, if LM test statistics larger than the certain critical value, then one will reject null hypothesis.

The next part is about linear regression model. The basic model is as follows: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Since researchers can never true regression line but can estimate it through the sample data, this is called the fitted regression line, i.e., $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. That is easy to deduce estimators' expectation and variance

$$E(\hat{\beta}_1) = \beta_1 \quad (13)$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (14)$$

$$E(\hat{\beta}_0) = \beta_0 \quad (15)$$

$$\text{Var}(\hat{\beta}_0) = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2 \quad (16)$$

Let V be a chi-squared random variable with v degrees of freedom and Z be a normal random variable. In the event that Z and V are unrelated, the distribution

$$T = \frac{Z}{\sqrt{V/v}} \quad (17)$$

is t-distribution with degrees of freedom is followed.

By applying above theory in linear regression model, obviously researchers might derive

$$T = \frac{\widehat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} \quad (18)$$

Additionally, it has $n-2$ degrees of freedom and the t-distribution, where S is called standard error of estimate, and $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ [9]. Subsequently, one will use this statistic in testing null hypothesis: $H_0: \beta_1 = (\beta_1)_0$.

Sometimes people want to test multiple coefficients at once to see if they are significant or not, researchers always use F-test to do so. Assume the multiple regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \quad (19)$$

where Y represents the dependent variable, and the error term, independent variables, and coefficients are all present. Next, one can establish null hypothesis, i.e., $H_0: \beta_1$ and β_2 are both equal to 0. This F-test is calculated as

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(n - k - 1)} \quad (20)$$

where n is the sample size, k is the number of independent variables (here is 2), p is the number of limitations (here is 2), and RSS is the residual sum of squares under the null hypothesis [10]. Then, using degrees of freedom, one can determine the crucial value from the F-distribution table. If it is not zero, people can reject the null hypothesis and conclude that at least one coefficient is not zero.

4. Conclusion

This essay previously covered the fundamentals of hypothesis testing, including the definitions of the alternative and null hypotheses. It separated the sample space into two sections that are mutually exclusive and accepted the corresponding hypothesis when the parameter's true value falls within a specific range. The definition of test statistics and an explanation of how to apply them for judgment were presented simultaneously. By defining acceptance and rejection domains, it can be simply understood that if the test statistics calculated from the given data fall within the rejection domain, the null hypothesis should be questioned because to choose an alternative hypothesis. Similar judgments can also be made using p-value. Subsequently, this article introduces how hypothesis testing plays a role in maximum likelihood estimation. In the LM test, by establishing hypothesis testing, a restrictive MLE is obtained. Through mathematical operations, it is found that the LM statistic follows a chi square distribution. Therefore, the LM statistic can be used to determine whether specific constraints in the model are necessary. Finally, the article discusses how hypothesis testing can be used to test whether coefficients in a regression model are significant, using t-test to test a single coefficient and F-test to simultaneously test whether multiple coefficients are significant together. In the future, more detailed research can be conducted on maximum likelihood estimation, such as LR testing.

References

- [1] Linde, Werner. (2017), *Probability Theory*. De Gruyter.
- [2] Weakliem, David L. (2016), *Hypothesis testing and model selection in the social sciences*. The Guilford Press.
- [3] Wei Lili, Liu Guojun (2020), *Research on hypothesis testing p-value teaching based on degree based thinking*. *Higher Science Education* (05):103-106.
- [4] Astaiza-Gómez, José Gabriel (2020). *Lagrange Multiplier Tests in Applied Research*. *Journal de Ciencia e Ingeniería*, 12(1), 13–19.
- [5] Pesaran, M. Hashem, (2015) *Hypothesis Testing in Regression Models, Time Series and Panel Data Econometrics*, Oxford University Press.
- [6] Xin Yan, & Xiaogang Su. (2009). *Linear Regression Analysis: Theory And Computing*. World Scientific, 41-127.
- [7] Samad, Muhammad Adnan (2025) *Quaternion Fractional Fourier Transform: Bridging Signal Processing and Probability Theory*. *Mathematics*, 13(2), 195.
- [8] Amarante, Massimiliano (2025), *Uniqueness of Convex-Ranged Probabilities and Applications to Risk Measures and Games*. *Mathematics of Operations Research*, 50(1), 743-763.
- [9] Wang, Duolao(2025), *Adjusted Win Ratio Using the Inverse Probability of Treatment Weighting*. *Journal of Biopharmaceutical Statistics*, 35(1), 21–36.
- [10] Proschan MichaelA., Pamela A. Shaw (2025). *Connections between Statistics and Mathematics/Probability*. *The American Statistician*,1–11.