Dalhousie University

Department of Statistics

New Frontiers in Statistical Analysis: Broadly Applicable Universal Inference Procedures

Author:
Jingyi Wang

Supervised by:
Dr. Edward Susko

Contents

1 Introduction

In the field of statistical inference, especially when constructing confidence intervals (CI), analysts frequently deal with complex datasets that traditional methods can not deal with. Frequently likelihood methods test hypotheses and contrast confidence sets using that likelihood ratio statistics have approximate χ^2 distributions with large samples or that estimators have approximately normal distributions. However, the regularity conditions for these results do not hold for complex datasets. A prime example that regularity conditions do not hold is found in mixture models. These models are instrumental in analyzing heterogeneous data, which is often the case when data originates from multiple underlying sources or populations. For instance, consider analyzing fish lengths from two different age groups, Age 1 and Age 2. Without direct age information, each fish's length is potentially drawn from one of two distinct distributions, each representing a different age group. Universal inference deals with regularity condition violations like those of mixture models.

This thesis explores advanced methodologies in statistical inference, focusing on universal confidence sets, hypothesis testing methods, derandomization, and their applications in regression analysis. The universal inference was recently developed in Wasserman et al (2020). A conservative confidence interval (CI) is constructed such that the probability, denoted as $P_\theta[\theta \in CI]$, is at least $1-\alpha$, but it might be much larger. The consequence of this conservatism is that the confidence interval might be wider than necessary. Conservative universal confidence sets and tests apply without asymptotic calculus and for irregular models, making them a crucial tool in statistical analysis. Universal inferences have some advantages like they do not require large samples and do not depend on regular conditions. However, they also have potential disadvantages, such as loss of power. The thesis describes the main way in which universal inference is implemented and reviews the main results. In addition, it illustrates the use of universal inference in a regression setting, contrasting with traditional analyses. The thesis also proves that the result extends to cases involving nuisance parameters and to independent but not identical cases, like regression. It concludes with a simulation study examining CI proportions in a Poisson setting.

Through a combination of theoretical exploration and empirical analysis, this thesis aims to provide a deeper understanding of these advanced statistical methodologies and their significance in data analysis and inference.

2 Universal and Cross-Fit Confidence Set

This section describes how universal confidence sets are constructed. Some randomization is introduced in the sets. The cross-fit and derandomized approaches are discussed that reduce the variance due to randomization but require additional computation.

The universal confidence set for a parameter θ is constructed based on data divided into two groups, D_0 and D_1. The steps involved in this process are:

1. The data is divided into two groups, D_0 and D_1.
2. An estimator $\widehat{\theta}_1$ is derived from D_1. This estimator can be any function estimator, like the maximum likelihood estimator, method of moments estimator, Bayes estimator, etc.
3. The likelihood ratio between $L_0(\widehat{\theta}_1)$ and $L_0(\theta)$ is compared, where $L_0(\theta)$ is the likelihood function based on dataset D_0 for parameter true value θ, and $L_0(\widehat{\theta}_1)$ is the likelihood of dataset D_0 evaluated at $\widehat{\theta}_1$. The split likelihood-ratio statistic is calculated as:

$$T_n(\theta) = \frac{L_0(\widehat{\theta}_1)}{L_0(\theta)}$$

4. The universal confidence set is defined as:

$$C_n = \{\theta \in \Theta : T_n(\theta) \leq \frac{1}{\alpha}\}$$

The main result of the fundamental theorem of the universal confidence set of Wasserman et al. (2020) is that generally

$$P_\theta[\theta \in C_n] \geq 1 - \alpha$$

Traditional Likelihood Ratio Method: Unlike the universal method, people usually use chi-squared distribution to obtain a confidence interval based on a likelihood ratio. In the regular model, this confidence interval can be expressed as:

$$\{\theta_0 : 2 \log \frac{L(\widehat{\theta})}{L(\theta_0)} \leq \chi^2_{\alpha,n-1}\}$$

Note:

- $\widehat{\theta}$: This denotes the maximum likelihood estimate of the parameter θ.
- $n - 1$: This is used to represent the degrees of freedom in the chi-squared distribution, where n is the sample size.
- α: The significance level of the test.
- $L(\theta)$: This is the likelihood based on all of the data.

Cross-Fit Likelihood-Ratio Statistic and Confidence Interval

There is some arbitrariness in the choice of test and training data. The cross-fit likelihood-ratio statistic (cross-fit LRS) adjusts for this to some extent. Therefore, cross-fit LRS is a significant statistical tool. It is formulated by averaging

$T_n(\theta)$ and $T_n^{swap}(\theta)$, with $T_n^{swap}(\theta)$ being the statistic computed after interchanging the datasets D_0 and D_1. The cross-fit LRS is denoted as $S_n(\theta)$ and is defined by:

$$S_n(\theta) = \frac{T_n(\theta) + T_n^{swap}(\theta)}{2}$$

The cross-fit confidence set, distinct from the universal confidence set, is then defined as:

$$C_n = \{\theta \in \Theta : S_n(\theta) \leq \frac{1}{\alpha}\}$$

Derandomization

The random choice of D_0 and D_1 introduces randomness to the universal sets. This variability can be reduced by using the cross-fit confidence set and reduced even more by iterating the process. To get precise confidence intervals, we can use derandomization techniques developed in Wasserman et al. (2020), which can help in creating samples or procedures that are less variable. Suppose we repeatedly randomly split the data into D_0 and D_1. Let $T_{n,1}, \ldots, T_{n,B}$ denote the split likelihood ratio states for each of these B choices of D_0 and D_1.

Let

$$\overline{T_n}(\theta) = \frac{1}{B} \sum_{j=1}^{B} T_{n,j}(\theta)$$

- $T_n(\theta)$: This represents the split likelihood-ratio statistic used in statistical inference.
- $\overline{T_n}(\theta)$: This is the average or mean of the figures $T_{n,j}(\theta)$ obtained from B samples.

In conclusion, by employing derandomization techniques in the process of generating split likelihood-ratio statistics, we can effectively reduce the variability inherent in random sampling, leading to more precise and reliable confidence intervals in statistical analysis.

3 Universal Hypothesis Testing Methods

Universal hypothesis testing is a fundamental aspect of statistical inference, providing methods to assess the validity of hypotheses using observed data. This section delves into various methodologies within this domain, particularly focusing on testing the null hypothesis in diverse statistical scenarios.

Let $\textcircled{H}_0 \subset \textcircled{H}$ be a possible composite null set and consider testing

$$H_0 : \theta \in \textcircled{H}_0 \quad \text{versus} \quad \theta \notin \textcircled{H}_0.$$

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad \theta \neq \theta_0.$$

Duality in Universal Hypothesis Testing

The duality between confidence set construction and hypothesis testing is discussed here. It allows us to construct hypothesis tests using the confidence set procedures discussed in the previous section. The dual test of a confidence set procedure rejects $H_0 : \theta = \theta_0$ if θ_0 is not in the confidence set. It rejects $H_0 : \theta \in \textcircled{H}_0$ if there is no $\theta \in \textcircled{H}_0$ that is in the confidence set.

The origin of this duality is a special kind of symmetry: if an event occurs, its opposite event does not occur, and vice versa. The duality test is that if a parameter value θ_0 lies outside a $(1-\alpha) \times 100\%$ confidence interval for a parameter θ, then a hypothesis test at level α will reject the null hypothesis $H_0 : \theta = \theta_0$. Conversely, if θ_0 lies within this confidence interval, the hypothesis test will not reject H_0.

Consider a universal hypothesis testing framework where we test

$$H_0 : \theta \in \textcircled{H}_0 \quad \text{versus} \quad \theta \notin \textcircled{H}_0.$$

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad \theta \neq \theta_0.$$

using a confidence region approach.

Duality Proof

We now show that a $(1-\alpha) \times 100\%$ confidence set procedure gives an α-level dual test. Suppose $c(x)$ is a $(1-\alpha) \times 100\%$ confidence region. In other words, $P_\theta(\theta \in c(x)) \geq 1 - \alpha$. Define the test function $\phi_{\textcircled{H}_0}(x)$ as:

$$\phi_{\textcircled{H}_0}(x) = \begin{cases} 1 & \text{if } \textcircled{H}_0 \cap c(x) = \emptyset \text{ (reject } H_0), \\ 0 & \text{otherwise (do not reject } H_0). \end{cases}$$

The probability of rejecting H_0 under θ_0 (θ_0 in $\textcircled{H}_0$)is given by:

$$\begin{aligned} P_{\theta_0}[\text{reject } H_0] &= P_{\theta_0}[\textcircled{H}_0 \cap c(x) = \emptyset] \\ &= P_{\theta_0}[\theta_0 \notin c(x)] \\ &= 1 - P_{\theta_0}[\theta_0 \in c(x)] \\ &\leq 1 - (1 - \alpha) \\ &= \alpha \end{aligned}$$

Thus, $\phi_{\theta_0}(x)$ is an α level test of $H_0 : \theta = \theta_0$.

Given an α level test $\phi_{\theta_0}(x)$ of $H_0 : \theta = \theta_0$, let

$$c(x) = \{\theta_0 : \phi_{\theta_0}(x) = 0\}$$

Then the probability that θ lies in $c(x)$ is:

$$\begin{aligned} P_\theta[\theta \in c(x)] &= P_\theta[\phi_\theta(x) = 0] \\ &= 1 - P_\theta[\phi_\theta(x) = 1] \\ &\geq 1 - \alpha \end{aligned}$$

Therefore, $c(x)$ is a $(1-\alpha) \times 100\%$ confidence region (CR).

3.1 Testing H_0 Using Confidence Interval Approach

One can determine the result of the dual test of the universal confidence set without actually determining the confidence set as we now show. This approach involves comparing the intersection of the CI with the null hypothesis set H_0. The dual test rejects $H_0 : \theta \in \Theta_0$ if The CI approach: rejects H_0 if

$$CI \cap \Theta_0 = \emptyset$$

We now simplify this condition:
$\leftrightarrow$:

$$\theta \notin CI \quad (\forall \theta \in \Theta_0)$$

$\leftrightarrow$:
Let $\widehat{\theta}_1$ be any estimator constructed from D_1,

$$\frac{L_0(\hat{\theta}_1)}{L_0(\theta)} > \frac{1}{\alpha} \quad (\forall \theta \in \Theta_0)$$

$\leftrightarrow$:

$$\min_{\theta \in \Theta_0} \left\{ \frac{L_0(\hat{\theta}_1)}{L_0(\theta)} \right\} > \frac{1}{\alpha}$$

$\leftrightarrow$:
Let $\hat{\theta}_0 := \arg\max_{\theta \in \Theta_0} L_0(\theta)$ be the maximum likelihood estimator under H_0 constructed from D_0. Then the dual test rejects if and only if

$$\frac{L_0(\hat{\theta}_1)}{L_0(\hat{\theta}_0)} = \min_{\theta \in \Theta_0} \left\{ \frac{L_0(\hat{\theta}_1)}{L_0(\theta)} \right\} > \frac{1}{\alpha}$$

3.1.1 Note

This approach indicates that if the null set Θ_0 does not contain any values that make the likelihood ratio exceed the threshold $\frac{1}{\alpha}$, it suggests insufficient evidence to support H_0, leading to its rejection.

4 Validity of universal inference confidence intervals

Theorem 1 of Universal inference by Wasserman et et al. (2020) gives the main result for the universal confidence set.
Let

- D_0: $y_1, \ldots, y_n$
- D_1: $y_{n+1}, \ldots, y_m$

Then, assuming $y_1, \ldots, y_m$ are independent and identically distributed(iid). Theorem 1 delves into the properties of a specific type of confidence set, denoted as C_n, in the realm of both regular and irregular statistical models.

Theorem 1: C_n is a finite-sample valid $(1-\alpha)$ confidence set for θ, meaning that $P_\theta(\theta \in C_n) \geq 1 - \alpha$.

The statement says that if C_n is a confidence set constructed in such a way that, under $y_1, \ldots, y_m$ are independent and identically, it has a finite-sample property of correctly containing the true parameter θ^* with a probability of at least $1 - \alpha$.

Note that the result of Theorem 1 applies without additional regularity conditions. In other words, it applies to both regular and irregular models. In regular models, the traditional likelihood ratio method can help us get the confidence interval, but for the irregular models, it fails. . In addition, it is difficult for us to approximate the distribution. The traditional likelihood ratio test also assumes sample sizes were large. The universal inference is applied with finite samples.

4.1 The Case for Independent but Not Identically Distributed Data

While Theorem 1 effectively addresses the construction and validity of confidence sets in regular models under the assumption of identically and independently distributed data, extending these concepts to scenarios involving independent but not identically distributed data poses additional challenges. We now consider an extension of Theorem 1 to accommodate such scenarios, particularly focusing on cases where the traditional likelihood ratio method may also be applicable.

For this model, we define the density of Y_i as $p_{i,\theta}(y)$, where the parameter vector θ and its components are defined as follows:

$$\begin{aligned} &\theta = (\zeta, \lambda), \\ \text{where} \quad &\theta \text{ is the parameter vector,} \\ &\lambda \text{ represents nuisance parameters,} \\ &\zeta \text{ represents the parameter of interest.} \end{aligned}$$

We use $\hat{\lambda}(\zeta)$ to denote the maximum likelihood estimator of the nuisance parameter λ, holding the parameter of interest, ζ, fixed.

Now we explain how the method has coverage of at least $1 - \alpha$, as claimed by *Theorem 1*:

Because the distribution is independent,

$$L_0(\hat{\theta}_1) = \prod_{i \in D_0} p_{i,\hat{\theta}_1}(Y_i)$$

$$L_0(\zeta, \widehat{\lambda}(\zeta))) = \prod_{i \in D_0} p_{i,\zeta,\widehat{\lambda}(\zeta)}(Y_i)$$

Let $\widehat{\theta}_1$ be some estimate based on D_1 alone, The

$$\begin{aligned} E_\theta\left[\frac{L_0(\widehat{\theta}_1)}{L_0(\zeta, \widehat{\lambda}(\zeta))}\middle| D_1\right] &= E_\theta\left[\frac{\prod_{i\in D_0} p_{i,\widehat{\theta}_1}(Y_i)}{\prod_{i\in D_0} p_{i,\zeta,\widehat{\lambda}(\zeta)}(Y_i)}\middle| D_1\right] \\ &= \int_A \frac{\prod_{i\in D_0} p_{i,\widehat{\theta}_1}(y_i)}{\prod_{i\in D_0} p_{i,\zeta,\widehat{\lambda}(\zeta)}(y_i)} \cdot \prod_{i\in D_0} p_{i,\zeta,\lambda}(y_i)\, dy_1 \dots dy_n \\ &= \int_A \frac{\prod_{i\in D_0} p_{i,\zeta,\lambda}(y_i)}{\prod_{i\in D_0} p_{i,\zeta,\widehat{\lambda}(\zeta)}(y_i)} \cdot \prod_{i\in D_0} p_{i,\widehat{\theta}_1}(y_i)\, dy_1 \dots dy_n \end{aligned} \tag{1}$$

where A is the support of P_θ.

- D_0: $y_1, \dots, y_n$
- D_1: $y_{n+1}, \dots, y_m$

Since $\widehat{\lambda}(\zeta)$ is the maximum likelihood estimator of λ holding ζ fixed, by definition for all $y \in D_0$,

$$\frac{\prod_{i\in D_0} p_{i,\zeta,\lambda}(y_i)}{\prod_{i\in D_0} p_{i,\zeta,\widehat{\lambda}(\zeta)}(y_i)} \leq 1 \tag{2}$$

$$\begin{aligned} \prod_{i\in D_0} \int p_{i,\widehat{\theta}_1}(y_i)\, dy_i &= \int p_{1,\widehat{\theta}_1}(y_1)\, dy_1 \dots \int p_{n,\widehat{\theta}_1}(y_n)\, dy_n \quad (p_{i,\widehat{\theta}_1} \geq 0) \\ &= \int \dots \int p_{1,\widehat{\theta}_1}(y_1) \dots p_{n,\widehat{\theta}_1}(y_n)\, dy_1 \dots dy_n = 1 \end{aligned}$$

Now, we substitute (2) in (1)

$$E_\theta\left[\frac{L_0(\widehat{\theta}_1)}{L_0(\zeta, \widehat{\lambda}(\zeta))}\middle| y(D_1)\right] \leq \int \prod_{i\in D_0} p_{i,\widehat{\theta}_1}(y_i)\, dy_1 \dots dy_n = 1$$

Let $C_n = \{\zeta : \frac{L_0(\widehat{\theta}_1)}{L_0(\zeta, \widehat{\lambda}(\zeta))} \leq \frac{1}{\alpha}\}$. Applying Markov's inequality, we find an upper bound for the probability that ζ is not in the confidence set C_n:

$$
\begin{aligned}
P_\theta(\theta \notin C_n) &= P_\theta(T_n(\theta) > \frac{1}{\alpha}) \\
&\leq \alpha E_\theta[T_n(\theta)] \\
&= \alpha E_\theta \left[\frac{L_0(\widehat{\theta}_1)}{L_0(\zeta, \widehat{\lambda}(\zeta))} \right] \\
&\leq \alpha.
\end{aligned}
$$

This establishes that $P_\theta(\theta \in C_n) \geq 1 - \alpha$, validating that C_n is a finite-sample valid $(1 - \alpha)$ confidence set for θ, as claimed in Theorem 1.

5 Real data illustration and simulation

5.1 Statistical Analysis and Illustration

In the extended phase of our statistical exploration, we conducted a comparative analysis of two methodologies for estimating confidence intervals of Poisson-distributed data sets. The traditional approach capitalizes on the Central Limit Theorem, approximating the distribution of the sample mean for a Poisson random sample with a normal distribution to calculate the confidence intervals, provided the sample mean is sufficiently large. This conventional method is computationally efficient and is widely accepted when the sample size justifies the normal approximation. Conversely, the universal method, a robust alternative, does not presuppose a normal distribution and instead employs a likelihood ratio test to construct the intervals, an approach that is advantageous with smaller sample sizes or when the distribution deviates significantly from normality.

Our simulation investigation encompassed the generation of fifty datasets of Poisson-distributed random variables, each with a lambda parameter set at ten and comprising one hundred observations. We observed that the universal method typically produced broader intervals, with an average width of 2.31, compared to the traditional method's narrower mean interval width of 1.24. Such an observation suggests that the universal method may provide a more conservative estimation, potentially reducing the risk of type I errors. Additionally, while the traditional method's intervals contained the true mean parameter in 98% of cases, underscoring its practical reliability under classical conditions, the universal method demonstrated even more remarkable performance, containing the true mean parameter in 100% of the cases.

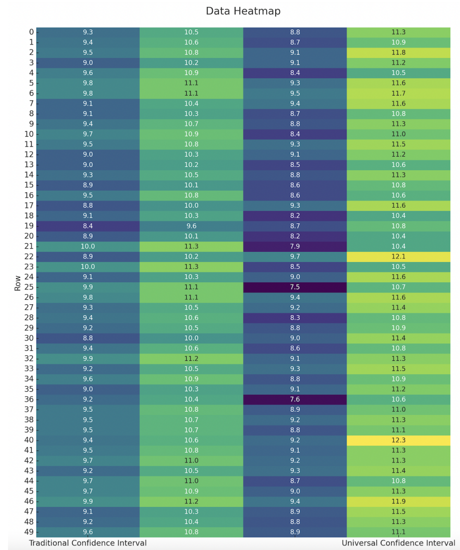

Figure 1: Confidence Interval Widths: Traditional vs. Universal Methods.

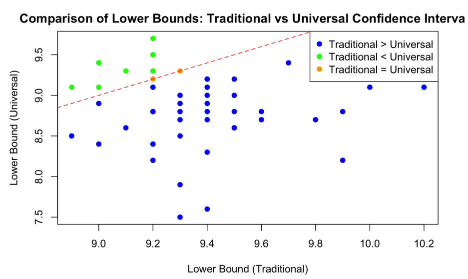

Figure 2: Comparison of Lower Bounds: Traditional vs Universal Confidence Interval.

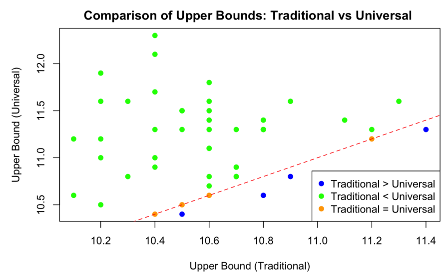

Figure 3: Comparison of Upper Bounds: Traditional vs Universal Confidence Interval.

In Figures 1, 2, and 3, we present a visual comparison between the lower and upper bounds of confidence intervals obtained through traditional and universal methods. The scatter plots illustrate individual instances of the estimated bounds for fifty groups of Poisson-distributed data.

For the lower bounds, as depicted in Figure 2, the universal method tends to yield smaller values than the traditional approach, which is consistent with a more conservative interval estimation. This is visually represented by the concentration of points above the line of equality, where traditional bounds are lower than the universal ones.

Conversely, Figure 3 showcases the upper bounds, where a similar trend is observable. The universal method frequently results in a larger upper bound compared to the traditional method, which can be inferred from the majority of points being located above the dashed line that signifies equivalence between the two methods.

The divergence from the line of equality in both figures highlights the universal method's tendency to provide broader confidence intervals. This is in alignment with the quantitative findings discussed earlier, where the universal method's average interval width was notably greater than that of the traditional method. Such broader intervals from the universal method could be advantageous in maintaining the actual parameter within the bounds, thus potentially reducing the likelihood of Type I errors.

The detailed analysis of these plots reinforces the conclusion that while the traditional method is efficient and reliable under certain conditions, the universal method offers a more cautious approach, which may be preferable in situations where the assumption of normality is questionable or when dealing with small sample sizes.

5.2 Regression Analysis

In regression models, managing nuisance parameters is crucial for making accurate inferences about the parameters of interest. This section outlines practical approaches for dealing with these parameters.

5.2.1 Model Framework

In cases that an interval for one of β_j is of interest, $\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}...\beta_p$ and σ^2 are all nuisance parameters. We now consider confidence interval construction for one of the coefficients β. The standard linear regression model is formulated as:

$$y_i = \beta_0 + \beta^T x_i + \epsilon_i, \quad \epsilon_i \sim \text{iid } \mathcal{N}(0, \sigma^2),$$

5.3 Independent but not Identically Distributed Observations in Nested Regression Models

5.3.1 Context and Objective

In this part of the extension, we explore Universal Confidence Intervals within nested regression models (Using "Real-Estate Data", featuring 20 rows of numerical data entries for residential properties, categorized into 'Total dwelling size' in hundreds of square feet, 'Assessed value' in thousands of dollars, and 'Selling price' also in thousands of dollars, with each row representing a different property. The data set is described in Section Multivariate Linear Regression Models of Johnson and Wichern (2007), particularly focusing on models that deviate from the assumption of independent but not identically distributed observations. Our primary objective is to test hypotheses comparing two nested linear regression models: Model 0 (a simpler model) and Model 1 (a full model).

5.3.2 Hypothesis Testing Approach

The hypothesis testing is designed to evaluate the significance of the additional parameter β_2 in Model 1. The testing procedure includes:

- **Hypotheses:**
  - Null Hypothesis (H_0): $\beta_2 = 0$
  - Alternative Hypothesis (H_a): $\beta_2 \neq 0$
- **Testing Methodology:**
  - A split universal inference approach is used for the hypothesis test.
  - The likelihood ratio statistic is employed to compare Model 1 and Model 0.

5.3.3 Nested Regression Models

Model 1 (Full Model) Model 1 is formulated with the assumption that the errors ϵ_i are normally distributed with a mean of zero and a constant variance σ^2. The model is expressed as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

where y_i represents the selling price $(y_1, \ldots, y_n, y_{n+1}, \ldots, y_m)$, x_{i1} represents the total dwelling size, and x_{i2} represents the assessed value for the i-th observation. The coefficients β_1 and β_2 correspond to the total dwelling size and assessed value, respectively.

Model 0 (Simpler, Nested Model) Model 0, nested within Model 1, also assumes normally distributed errors with mean zero and constant variance σ^2. This model is simpler, as it excludes the parameter β_2 (effectively setting $\beta_2 = 0$):

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

Computing the confidence interval requires the calculation of $T_n(\beta_2)$ for multiple choices of β_2. If the primary interest is in $H_0 : \beta_2 = \beta_{2o}$, one can simply calculate $T_n(\beta_{2o})$ and reject if $T_n(\beta_{2o}) \geq \frac{1}{\alpha}$. A particular hypothesis of interest is $H_0 : \beta_2 = 0$ which corresponds to a test of the simpler model

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

against the full model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

$$T_n(\beta_2) = \frac{L_0(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}, \widehat{\mu})}{L_0(\beta_2, \widehat{\sigma}(\beta_2), \widehat{\mu}(\beta_2), \widehat{\beta_0}(\beta_2), \widehat{\beta_1}(\beta_2))}$$

where:

- Suppose that the number of observations in D_0 is equal to the number of observations in D_1.
- $L_0(\cdot)$ represents the likelihood function based on dataset D_0.
- $\widehat{\beta}, \widehat{\sigma}, \widehat{\mu}$ can be any function estimators (we select maximum likelihood estimator)obtained from dataset D_1.
- β_2 is the true value(according to the condition, it is 0) under H_A, and $\widehat{\sigma}(\beta_2), \widehat{\mu}(\beta_2), \widehat{\beta_0}(\beta_2), \widehat{\beta_1}(\beta_2)$ are the maximum likelihood estimates presented by β_2 based on D_0.

The universal confidence set includes β_2 in the set if:

$$T_n(\beta_2) = \frac{L_0(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}, \widehat{\mu})}{L_0(\beta_2, \widehat{\sigma}(\beta_2), \widehat{\mu}(\beta_2), \widehat{\beta}_0(\beta_2), \widehat{\beta}_1(\beta_2))} \leq \frac{1}{\alpha}$$

Note: The figure4 features $T_n(\beta_2)$ as a black curve against β_2, with a red horizontal line indicating the significance threshold $\frac{1}{\alpha}$. Vertical blue lines mark the confidence interval bounds for β_2. The β_2 values with $T_n(\beta_2)$ under the red line are included in the confidence sets.

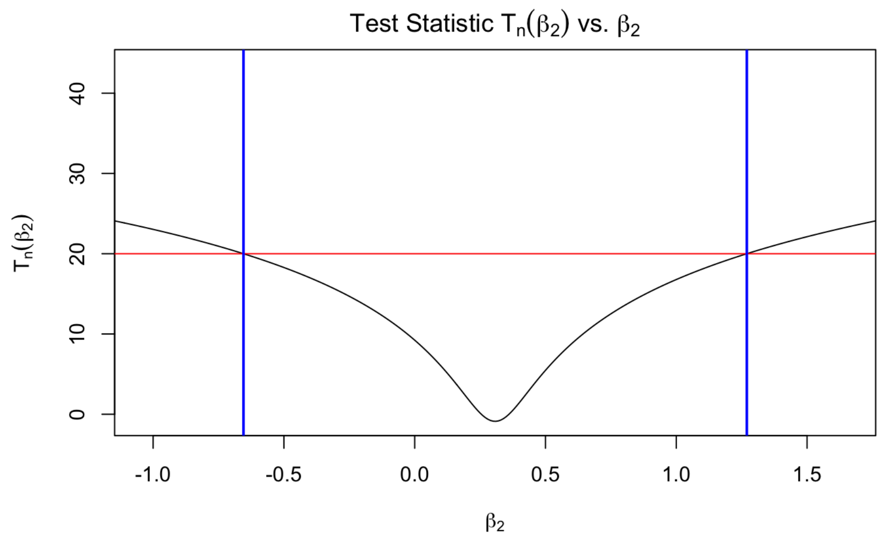

Figure 4: Test Statistic $T_n(\beta_2)$ for a range of β_2 values.

In the context of nested regression models with independent but not identically distributed (non-i.i.d.) observations, the traditional chi-square method, and the universal confidence interval approach offer distinct yet complementary insights. The traditional method, grounded in the assumptions of normality and homoscedasticity, calculates confidence intervals for coefficients by fitting two models: a null model and a full model. This approach yields specific interval estimates, evidenced in our study by the confidence interval for the coefficient of 'Assessed Value' being $[0.13, 0.31]$.

Applying the split universal inference approach for hypothesis testing between Model 0 and Model 1 in our study, the likelihood ratio test statistic $T_n(\theta)$ highlighted the significance of the additional parameter β_2 in the complex model. From the Figure 4, the derived confidence interval for the β of 'Assessed Value'(β_2)was $[-0.65, 1.27]$, which includes zero, indicating that the null hypothesis $H_0 : \beta_2 = 0$ cannot be confidently rejected based on this interval alone.

The Figure 4 representation of the test statistic for the null hypothesis ($\beta_2 = 0$) within the universal confidence interval, which encompasses zero, suggests that the potential effect of 'Assessed Value' on 'Selling Price' is not statistically significant at the chosen confidence level. This observation contrasts with the confidence interval obtained from the traditional chi-square method, which does not include zero, implying a significant effect of 'Assessed Value' on 'Selling Price'.

In conclusion, although Model 1, which includes 'Total Dwelling Size', may offer a perspective on factors affecting 'Selling Price', our analysis suggests that the statistical evidence does not robustly support the significance of β_2. Further investigation, possibly with a larger sample size or additional data, might be necessary to make a more definitive statement regarding the role of 'Total Dwelling Size' in predicting 'Selling Price'.

6 Conclusion

This thesis investigated the new and broadly applicable universal inference procedures. We delved deep into universal inference, a methodological framework designed to transcend the limitations of traditional statistical techniques, especially in scenarios characterized by complex data structures and stringent assumptions.

The exploration began by highlighting the strengths of universal inference. Its principal merit lies in its adaptability to various data scenarios, including those where conventional methods struggle due to the absence of large samples or regularity conditions. In addition, it does not depend on regularity conditions. This quality of universal inference makes it a formidable tool in statistical analysis, allowing for robust interpretations even in the most intricate data landscapes.

Further, we provided proof of the main universal inference results when data are independent but not identically distributed and when there are nuisance parameters. An area where this is important is regression analysis because dependence on x_i and the fact that x_i is treated as fixed makes the observations independent.

However, the journey also revealed a critical aspect of universal inference – its conservative nature in constructing confidence intervals. While this conservatism enhances the reliability of statistical conclusions, guarding against Type I errors, it also introduces a trade-off by widening the confidence intervals. Such a feature, though beneficial in certain contexts, may limit the precision of the estimates, a fact that was evident in our empirical investigations.

Our foray into the realm of Poisson distribution further illuminated this trade-off. We get the traditional method's intervals containing the true mean parameter in 98% of cases and 100% by the universal inference method. While the traditional methods showed a near-perfect success rate in containing the true mean parameter, the universal method, with a slightly lower success rate, offered a more conservative and, thus, potentially more reliable approach. We observed that the universal method typically produced broader intervals, with an average width of 2.31, compared to the traditional method's narrower mean interval width of 1.24. The broader intervals, though at a cost to precision, remained informative and valuable, especially in contexts where traditional assumptions were unmet.

As we conclude, it's evident that the journey of universal inference is still unfolding, with numerous opportunities for exploration and refinement in domains such as time-series analysis and spatial data. There's a vital need to balance the conservatism of universal inference with the pursuit of precision in estimates.

Addressing computational challenges, especially in big data, is crucial. This includes tackling computational issues for efficiency, exploring derandomization, and investigating optimal strategies for data splitting. Questions like 'How best to split the data?' and the effectiveness of different proportion splits, such as 50-50 or 90-10, are fundamental. Such investigations will improve practicality and enhance the accuracy and reliability of universal inference across various applications.

In essence, this thesis does not mark an end, but rather a beginning – a prologue to what could be a transformative chapter in the field of statistical inference. The universal inference approach, with its unique blend of flexibility and robustness, stands poised to redefine the boundaries of data analysis and open up new frontiers in the understanding of broadly applicable universal inference procedures.

7 References

- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

- Wasserman, L., et al. (2020). Universal Inference. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), 16880-16890. https://doi.org/10.1073/pnas.1912789117