Show That The Rejection Region Is Of The Form {x X0} {x X1}, Where X0 And X1 Are Determined By C.

Show That The Rejection Region Is Of The Form {x X0} {x X1}, Where X0 And X1 Are Determined By C.

In statistical hypothesis testing, especially when dealing with continuous probability distributions, the concept of a rejection region is fundamental. It determines the set of values for the test statistic that leads us to reject the null hypothesis. A common goal is to understand the structure of this rejection region and how it relates to the critical value(s) determined by the significance level, denoted by C. This article aims to demonstrate that, under appropriate conditions, the rejection region can be expressed as an interval of the form {x | x < X0} or {x | x > X1}, where X0 and X1 are critical points derived from the significance level C. We will explore the theoretical foundations behind this form, the conditions under which it holds, and methods to determine the specific values of X0 and X1.

Understanding the Rejection Region in Hypothesis Testing

Before delving into the specific form of the rejection region, it's essential to understand what a rejection region is and its role in hypothesis testing.

Definition of the Rejection Region

The rejection region, also called the critical region, is the subset of the sample space where, if the observed test statistic falls within it, we reject the null hypothesis (H0). Conversely, if the test statistic lies outside this region, we fail to reject H0. The size or significance level C of the test is the probability of incorrectly rejecting H0 when it is true (Type I error). The rejection region is thus constructed to ensure that the probability of a Type I error does not exceed C.

Significance Level and Critical Values

The significance level C is used to determine critical values, X0 and X1, which demarcate the boundaries of the rejection region. For example, in a two-tailed test, the rejection region is typically split into two parts: one in the lower tail and one in the upper tail, each associated with a specific probability (C/2). The critical values are chosen such that:

- $P(X \leq X0 \mid H0 \text{ is true}) = C/2$
- $P(X \geq X1 \mid H0 \text{ is true}) = C/2$

for a two-tailed test, or

- $P(X \geq X1 \mid H0 \text{ is true}) = C$ (for a one-tailed test)

depending on the nature of the hypothesis.

Mathematical Foundations of the Rejection Region

The structure of the rejection region depends on the distribution of the test statistic under the null hypothesis and the test's design.

Distribution of the Test Statistic

Suppose X is a continuous random variable representing the test statistic with probability density function (pdf) f(x). Under H0, X has a known distribution with cumulative distribution function (CDF) F(x). The properties of F directly influence the shape of the rejection region.

Formulation of the Rejection Region

Given the distribution, the rejection region R at significance level C is typically defined as:

- For a one-tailed test: $R = \{x \mid x \geq X1\}$ or $\{x \mid x \leq X0\}$
- For a two-tailed test: $R = \{x \mid x \leq X0\} \cup \{x \mid x \geq X1\}$

where X0 and X1 are the critical values satisfying:

- $P(X \leq X0) = \alpha$ (for the lower tail)
- $P(X \geq X1) = \beta$ (for the upper tail)

with $\alpha + \beta = C$ (for two tails), or C for a one tail.

Showing That The Rejection Region Is Of The Form $\{x \mid x < X0\}$ or $\{x \mid x > X1\}$

The core of this article is demonstrating that, under standard assumptions, the rejection region can be expressed as one or two intervals of the form $\{x \mid x < X0\}$ and/or $\{x \mid x > X1\}$. The proof relies on properties of the distribution function and the monotonicity of the test statistic's distribution.

Monotonicity of the Distribution Function

Since $F(x) = P(X \leq x)$ is a non-decreasing function, its inverse (the quantile function) is well-defined for points where F is continuous and strictly increasing. This property is crucial because it ensures that critical values X0 and X1 can be identified uniquely for the given significance level C.

Constructing the Rejection Region

Suppose the test is designed to detect deviations in a particular direction, either lower or upper tail, or both.

- For a one-tailed test with significance level C:
- If testing whether the parameter is greater than some value, the rejection region is $R = \{x \mid x > X1\}$, where X1 satisfies $P(X \geq X1) = C$.

- If testing whether the parameter is less than some value, $R = \{x \mid x < X0\}$, with X0 satisfying $P(X \leq X0) = C$.

- For a two-tailed test:
- The rejection region is $R = \{x \mid x < X0\} \cup \{x \mid x > X1\}$, with X0 and X1 chosen such that:
- $P(X \leq X0) = C/2$
- $P(X \geq X1) = C/2$

This configuration ensures that the total probability of the rejection region under H0 is C, matching the significance level.

Determining X0 and X1 Based on C

The determination of the critical values X0 and X1 hinges on the distribution of the test statistic and the significance level C.

Using Quantile Functions

The inverse of the distribution function, known as the quantile function, is instrumental:

- $X0 = F^{-1}(C/2)$ for the lower tail in a two-tailed test.
- $X1 = F^{-1}(1 - C/2)$ for the upper tail in a two-tailed test.
- For a one-tailed test:
- $X0 = F^{-1}(C)$ or $X1 = F^{-1}(1 - C)$, depending on the direction.

These critical values define the rejection region precisely as an interval (or union of intervals) of the specified form.

Examples with Common Distributions

- Normal Distribution: For a standard normal distribution, the critical values are obtained using the Z-table:
- $X0 = -z\alpha$
- $X1 = z\alpha$
- t-Distribution: For small samples, critical t-values are used, determined from the t-distribution table based on degrees of freedom and significance level.
- Chi-Square or F-Distribution: Critical values are obtained from respective tables, defining the rejection region as upper-tail intervals.

Conclusion

In conclusion, the structure of the rejection region in hypothesis testing can be expressed in the form $\{x \mid x < X0\}$ or $\{x \mid x > X1\}$, where X0 and X1 are critical values derived from the distribution of the test statistic and the significance level C. This form emerges naturally from the monotonic properties of distribution functions, the definition of critical values via quantiles, and the goal of controlling Type I error probability. Understanding this structure allows statisticians to design tests with well-defined rejection regions, making the testing process both rigorous and interpretable. Whether in simple

cases like the normal distribution or more complex distributions, the principle remains the same: the rejection region can always be characterized as an interval (or union of intervals) determined explicitly by the chosen significance level C.

Frequently Asked Questions

What is the significance of the rejection region being of the form {x | x ∈ X0} ∪ {x | x ∈ X1} in hypothesis testing?

This form indicates that the rejection region consists of two separate sets, typically corresponding to the tails of the distribution in a two-sided test. It simplifies the decision rule by defining clear cutoff points X0 and X1, which are determined based on the significance level C.

How do we determine the sets X0 and X1 in the rejection region from the value of C?

X0 and X1 are determined by finding the critical values that partition the distribution into regions where the test statistic leads to rejection or acceptance. These critical values are derived such that the total probability of the rejection region equals C, often by solving for the quantiles of the distribution corresponding to the significance level.

Can you explain why the rejection region is expressed as {x | x ∈ X0} ∪ {x | x ∈ X1} rather than a single continuous interval?

In two-sided hypothesis tests, the rejection region typically includes the extreme values in both tails of the distribution, which form two separate intervals. Therefore, the rejection region is the union of these two sets, rather than a single continuous interval, reflecting the symmetric nature of the test.

What role does the constant C play in defining the sets X0 and X1?

The constant C represents the significance level or the total probability of making a Type I error. It determines the size of the rejection region by setting the combined probability of the two tail regions, X0 and X1, so that their probabilities sum to C, thereby controlling the likelihood of incorrectly rejecting the null hypothesis.

How can the form {x | x ∈ X0} ∪ {x | x ∈ X1} be used to facilitate the computation of critical values in hypothesis testing?

Expressing the rejection region as a union of two sets allows for straightforward calculation of critical values by finding the quantiles of the distribution that correspond to the cumulative probabilities C/2 in each tail. This approach simplifies determining the cutoff points X0 and X1 that define the rejection

region based on the significance level C.

Additional Resources

Rejection Regions in Hypothesis Testing: Characterizing the Form {x ∈ X₀} ∪ {x ∈ X₁}

Introduction

In the landscape of statistical hypothesis testing, the concept of a rejection region (also known as a critical region) plays a pivotal role. It defines the set of all possible data points that lead us to reject the null hypothesis in favor of the alternative. Understanding the structure of this rejection region is fundamental for designing tests with desirable properties, such as optimality, simplicity, and interpretability.

A common and powerful form of the rejection region is one that can be expressed as the union of two subsets, namely:

$$
\text{Rejection Region} = \{ x \in X_0 \} \cup \{ x \in X_1 \}
$$

where X_0 and X_1 are subsets of the sample space, determined explicitly by a constant C.

This review aims to explore why and how the rejection region can be characterized in this form, the underlying principles that lead to this structure, and the mathematical tools involved in establishing this property.

---

1. Understanding the Rejection Region Conceptually

1.1 Definition and Purpose

- The rejection region R is a subset of the sample space $\mathcal{X}$ such that if the observed data x falls within R, we reject the null hypothesis H_0.
- Conversely, if $x \notin R$, we do not reject H_0.

1.2 Connection to Test Functions

- The test function $\phi(x)$, which typically takes values in $[0, 1]$, can be viewed as an indicator of rejection:

$$
\phi(x) =
\begin{cases}
1, & \text{if } x \in R \\
0, & \text{otherwise}
\end{cases}
$$

\]

- In many classical tests, particularly Neyman-Pearson tests, the structure of \(R \) is directly related to likelihood ratios and critical values.

---

2. The Neyman-Pearson Lemma and Its Role

2.1 Statement of the Lemma

The Neyman-Pearson lemma provides a foundation for deriving the most powerful tests for simple hypotheses:

- Suppose we are testing:

\[
H_0: \text{data follows } f_0(x) \quad \text{vs} \quad H_1: \text{data follows } f_1(x)
\]

- The lemma states that, for a fixed significance level \(\alpha \), the most powerful test rejects \(H_0 \) in favor of \(H_1 \) when the likelihood ratio:

\[
\Lambda(x) = \frac{f_1(x)}{f_0(x)}
\]

exceeds a certain threshold \(C \).

2.2 Rejection Region from the Neyman-Pearson Lemma

- The most powerful rejection region has the form:

\[
R = \{ x \in \mathcal{X} : \Lambda(x) > C \}
\]

- When the likelihood ratio is monotonic in some sufficient statistic, the rejection region simplifies to an inequality involving this statistic.

- The critical value \(C \) is chosen to satisfy the significance level constraint.

---

3. Mathematical Characterization of the Rejection Region

3.1 The Form \(\{x \in X_0\} \cup \{x \in X_1\} \)

- The rejection region can be expressed as the union of two parts, which are often defined via inequalities involving a constant \(C \).

- Key idea: The rejection region is characterized by thresholding some function of the data, leading to sets:

$$
X_0 = \{ x : T(x) \geq C \}
$$

$$
X_1 = \{ x : T(x) \leq C' \}
$$

where $T(x)$ is a test statistic, and C, C' are constants determined by the desired size α or power considerations.

- When the test involves only one threshold C, the rejection region is simply:

$$
R = \{ x : T(x) > C \}
$$

which can be viewed as the union of the set where the test statistic exceeds C and possibly other regions depending on the test's structure.

3.2 Determination of X_0 and X_1 by C

- The sets X_0 and X_1 are defined based on the distributional properties of the test statistic and the critical value C.

- For example:

- $X_0 = \{ x : T(x) \leq C \}$
- $X_1 = \{ x : T(x) > C \}$

resulting in:

$$
R = X_1
$$

- In more complex scenarios, the rejection region may involve two thresholds, such as:

$$
R = \{ x : T(x) < C_1 \} \cup \{ x : T(x) > C_2 \}
$$

where $C_1 < C_2$, creating a rejection region with two parts.

---

4. Criteria and Conditions for the Form

4.1 Monotonicity and Likelihood Ratio Ordering

- The key condition that guarantees the rejection region can be written in the form $\{ x \in X_0 \} \cup \{ x \in X_1 \}$ is the monotone likelihood ratio property (MLRP).
- Definition: A family of distributions $\{f_\theta\}$ has MLRP in a statistic $T(x)$ if, for $\theta_1 > \theta_0$, the likelihood ratio:

$$
\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}
$$

is increasing in $T(x)$.

- When MLRP holds, the most powerful tests are threshold-based on $T(x)$, leading to rejection regions of the form:

$$
R = \{ x : T(x) \geq C \}
$$

- The sets X_0 and X_1 are then naturally defined as the subsets where $T(x)$ lies below or above the threshold.

4.2 Neyman-Pearson and Its Implication

- The Neyman-Pearson lemma provides that the optimal rejection region for simple hypotheses is of the form:

$$
R = \left\{ x : \frac{f_1(x)}{f_0(x)} > C \right\}
$$

- Since the likelihood ratio depends on the data through $T(x)$, the region naturally decomposes into subsets determined by C, i.e.,

$$
R = \{ x : T(x) > C \}
$$

- The complement of the rejection region, R^c, is then:

$$
R^c = \{ x : T(x) \leq C \}
$$

- Accordingly, the entire sample space is partitioned into the union of X_0 and X_1, where:

$$
X_0 = \{ x : T(x) \leq C \}
$$

$$
X_1 = \{ x : T(x) > C \}
$$

making the rejection region:

$$
R = X_1
$$

---

5. Practical Examples and Applications

5.1 Testing the Mean of a Normal Distribution with Known Variance

- Suppose $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$, with known σ^2.
- Null hypothesis: $H_0: \mu = \mu_0$.
- Alternative: $H_1: \mu \neq \mu_0$.

- The test statistic:

$$
T = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}
$$

- Rejection region:

$$
R = \left\{ x : |\bar{X} - \mu_0| > C \right\}
$$

- This can be split into:

$$
X_0 = \left\{ x : \bar{X} \leq \mu_0 + C \right\}
$$
$$
X_1 = \left\{ x : \bar{X} \geq \mu_0 + C \right\}
$$

and similarly for the lower tail.

- Here, the rejection region is explicitly of the form $\{$

Show That The Rejection Region Is Of The Form X X0 X X1

Where X0 And X1 Are Determined By C

Show That The Rejection Region Is Of The Form X X0 X X1 Where X0 And X1 Are Determined By C

Related Articles

- How Do You Tell The Difference Between Pericarditis And Benign Early Repolarization (BER)? (4)
- In Promoting Personality Growth, The Person-centered Perspective Emphasizes Everything Except:
- Command To Find Interface Name Of The Network Interface Device Used To Route The Icmp Ping Packets

Back to Home