Error Exponents of Mismatched Likelihood Ratio Testing

Sangana Sai Venkarta Surya Teja, Gedela Madhu Babu

[1] EE Department, Indian Institute of Technology Hyderabad
[2] EE Department, Indian Institute of Technology Hyderabad

Abstract :

The study on the problem of mismatched likelihood ratio test. We analyze the type-I and II error exponents when the actual distributions generating the observation are different from the distributions used in the test. We derive the worst-case error exponents when the actual distributions generating the data are within a relative entropy ball of the test distributions. In addition, we study the sensitivity of the test for small relative entropy balls.

Index Terms: Likelihood Ratio test, KL Divergence.

1 Introduction

Binary hypothesis testing involves making decisions between two possible distributions based on observed data. In this study, we consider the scenario where observations are generated from two possible distributions P1 and P2. The likelihood ratio test is a common approach for binary hypothesis testing, comparing the likelihoods of observed data under the competing distributions. The research focuses on mismatched likelihood ratio testing, where fixed test distributions are used even when the true distributions are uncertain

2 Theotitical Framework

The likelihood ratio test is formulated to compare the likelihoods of observed data under two competing distributions P1 and P2. The test is derived from the Kullback-Leibler divergence, quantifying the relative entropy between distributions. The error exponents, characterizing the asymptotic behavior of error probabilities, are analyzed to understand the tradeoff between type-I and type-II errors

2.1 Prelimenaries

The problem of mismatched binary hypothesis testing between iid distributions. We analyze the tradeoff between the pairwise error probability exponents when the actual distributions generating the observation are different from the distributions used in the LRT in composite setting.

Consider Binary hypothesis testing problem X =(x1,x2..........,xn) are the observations generated from two probability distributions P_1^n and P_2^n defined on probability simplex $\wp(\chi^n)$ and $P_1^n > 0$ and P_2^n >0 and x ϵ χ Assume P_1^n and P_2^n are product distributions

$$P_1^n(X) = \prod_{i=1}^{n} P_1''(x_i)$$

and

$$P_2^n(X) = \prod_{i=1}^{n} P_2''(x_i)$$

3 Liklihood Ratio Test

Let $\varphi : \chi^n \to \{1, 2\}$ be a hypothesis test that decides which distribution generated the observation X.

Here we consider deterministic tests φ that decide in favor of P_1^n if $A_1 \subset \chi^n$ and P_2^n if $A_2 \subset \chi^n$. The test performance is determined by using pairwise error probabilities.The Type-1 and Type-2 Errors are given below : $\epsilon(\varphi) = \sum_{X \in A_2} P_1^n(x)$ and $\epsilon(\varphi) = \sum_{X \in A_1} P_2^n(x)$ The optimal error probability tradeoff given by: $\alpha\beta = \min_{\phi:2(\phi)\leq\beta} \epsilon(\phi)$ The Likelihood ratio test is for the above is obtained form the Neyman and pearson paper :
$\phi\gamma(x) = P_{n2}(x)/P_{n1}(x) \geq e^{n\gamma}$ The type sequence X =(x1,x2..........,xn) is $\hat{T}_x(a) = N(a|x)$ where N(a|x) is the no of a are there in the sequence. Then the LRT is defined as a function of type of the observation as:
$\phi\gamma(\hat{T}_x) = 1, \quad D(\hat{T}_x||P_1) - D(\hat{T}_x||P_2) \geq \gamma$
Where
D(P ∥ Q) = $\sum_x P(x) \log \frac{P(x)}{Q(x)}$ is the relative entropy between P and Q.
In this paper r, we are interested in the asymptotic exponential decay of the pairwise error probabilities.
$E_0 = \liminf_{n\to\infty} \frac{-1}{n} \log_0(\phi), \quad E_1 = \liminf_{n\to\infty} \frac{-1}{n} \log_1(\phi).$

By using Sanov's Theorem from Elements of Information theory by thomas cover the optimal error exponent tradeoff (E1, E2), attained by the likelihood ratio test, can be shown to be:

$$E_0 = \min_{Q \in \mathcal{Q}_1} D(Q||P_1)$$
$$E_1 = \min_{Q \in \mathcal{Q}_2} D(Q||P_2)$$

where

$$\mathcal{Q}_1 = \{Q \in P(X) : D(Q||P_1) - D(Q||P_2) \geq \gamma\}$$
$$\mathcal{Q}_2 = \{Q \in P(X) : D(Q||P_1) - D(Q||P_2) \leq \gamma\}\,.$$

The minimizing distribution is the tilted distribution

$$Q_\lambda(x) = \frac{P_1^{1-\lambda}(x) P_2^{\lambda}(x)}{\sum_{a \in X} P_1^{1-\lambda}(a) P_2^{\lambda}(a)}, \quad 0 \leq \lambda \leq 1$$

Equivalently, the dual expressions can be derived by substituting the minimizing distribution into the Lagrangian yielding :

$$E_0 = \max_{\lambda \geq 0} \left\{ \lambda\gamma - \log \left(\sum_{x \in X} P^0_{1-\lambda}(x) P^1_{\lambda}(x) \right) \right\}$$

$$E_1 = \max_{\lambda \geq 0} \left\{ -\lambda\gamma - \log \left(\sum_{x \in X} P^0_{\lambda}(x) P^1_{1-\lambda}(x) \right) \right\}.$$

4 Results

The optimal error exponent tradeoff E1 attained by LRT :
Objective:
Minimize D(Q $\|P_1$)
subject to the constraint $D(Q\|P_1) - D(Q\|P_2) >= \gamma$.

Step 1: Lagrangian Setup
Construct the Lagrangian function incorporating the constraint with a Lagrange multiplier
$\lambda : \mathcal{L}(Q, \lambda) = D(Q\|P_1) - \lambda\left(D(Q\|P_1) - D(Q\|P_2) - \gamma\right)$
Expanding this gives:

$$\mathcal{L}(Q, \lambda) = \sum_x Q(x) \log \frac{Q(x)}{P_1(x)} - \lambda \left(\sum_x Q(x) \log \frac{Q(x)}{P_1(x)} \right.$$

$$\left. - \sum_x Q(x) \log \frac{Q(x)}{P_2(x)} - \gamma \right.$$

Step 2: Simplification
Simplify the terms:

$$\mathcal{L}(Q, \lambda) = (1 - \lambda) \sum_x Q(x) \log \frac{Q(x)}{P_1(x)} + \lambda \sum_x Q(x) \log \frac{Q(x)}{P_2(x)} + \lambda\gamma$$

Step 3: First Order Condition
To find the stationary points, set the derivative of L with respect to Q(x) to zero:

$$\frac{\partial \mathcal{L}}{\partial Q(x)} = (1-\lambda)\left(\log Q(x) - \log P_1(x) + 1\right) + \lambda\left(\log Q(x) - \log P_2(x) + 1\right) = 0$$

This simplifies to:

$$\log Q(x) = \lambda \log P_2(x) + (1 - \lambda) \log P_1(x)$$

Step 4: Solve for Q(x)
Exponentiating both sides to solve for Q(x):

$$Q(x) = \exp\left(\lambda \log P_2(x) + (1 - \lambda) \log P_1(x)\right)$$

$$Q(x) = \frac{P_1^{1-\lambda}(x) P_2^{\lambda}(x)}{Z}$$

where Z is a normalization factor ensuring Q is a probability distribution.

Step 5: Normalization and Parameter Adjustment
The final form of Q(x) should include normalization and can be adjusted to:

$$Q_\lambda(x) = \frac{P_1^{1-\lambda}(x) P_2^{\lambda}(x)}{\sum_{a \in X} P_1^{1-\lambda}(a) P_2^{\lambda}(a)}$$

This final expression is exactly the tilted distribution.
The derivation using the Lagrangian shows how the distribution that minimizes the Kullback-Leibler divergence to P_1 under the constraint on the difference in divergences to P_1 and P_2 can be found. This method effectively combines the intuitive aspects of exponential tilting with the formalism of constrained optimization.

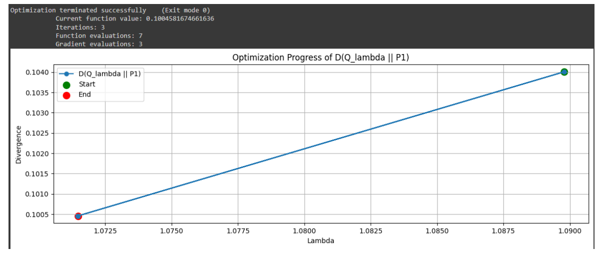

Figure 1. Used tilted distribution to find Optimum Divergence and Optimum lambda Value for Two Distributions(Prabalilities).

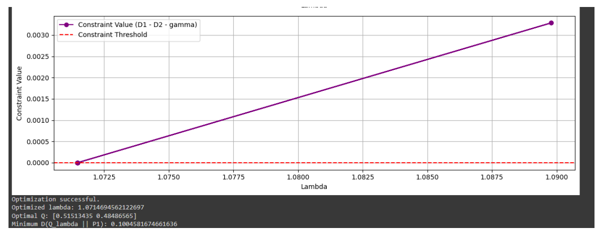

Figure 2. Plot of the Constrain Value when minimizing the Divergence:

5 Mismatched Likelihood Ratio Testing

Let $P_1(x)$ and $P_2(x)$ be the test distributions used in the likelihood ratio test with threshold γ given by
$\phi_{\hat{\gamma}}(\hat{T}_x) = \quad D(\hat{T}_x||\hat{P}_1) - D(\hat{T}_x||\hat{P}_2) \geq \hat{\gamma}$
The optimal error exponents:

$$\hat{E}_1(\hat{\phi}_{\hat{\gamma}}) = \min_{Q \in \hat{Q}_1(\hat{\gamma})} D(Q||P_1)$$

$$\hat{E}_2(\hat{\phi}_{\hat{\gamma}}) = \min_{Q \in \hat{Q}_2(\hat{\gamma})} D(Q||P_2)$$

where

$$\hat{Q}_1(\hat{\gamma}) = \left\{ Q \in P(X) : D(Q||\hat{P}_1) - D(Q||\hat{P}_2) \geq \hat{\gamma} \right\}$$

$$\hat{Q}_2(\hat{\gamma}) = \left\{ Q \in P(X) : D(Q||\hat{P}_1) - D(Q||\hat{P}_2) \leq \hat{\gamma} \right\}.$$

The minimizing distributions in (18) and (19) are given by

$$\hat{Q}^1_\lambda(x) = \frac{P_1(x)\left(\hat{P}_1^{-\lambda}(x)\hat{P}_2^{\lambda}(x)\right)}{\sum_{a \in X} P_1(a)\left(\hat{P}_1^{-\lambda}(a)\hat{P}_2^{\lambda}(a)\right)}, \quad \lambda_1 \geq 0,$$

$$\hat{Q}^2_\lambda(x) = \frac{P_2(x)\left(\hat{P}_2^{-\lambda}(x)\hat{P}_1^{\lambda}(x)\right)}{\sum_{a \in X} P_2(a)\left(\hat{P}_2^{-\lambda}(a)\hat{P}_1^{\lambda}(a)\right)}, \quad \lambda_2 \geq 0,$$

This tilted distribution equation are solved using same lagrangian equation as solved above.

6 Conclusion

Mismatched likelihood ratio testing is a powerful tool for binary hypothesis testing in scenarios where the true distributions are uncertain. Through rigorous analysis and optimization techniques, this approach provides robust and reliable decision-making capabilities. Further research could explore extensions to more complex scenarios and practical implementations in real-world systems

References:

Error Exponents of Mismatched Likelihood Ratio Testing Author(s): Boroumand, Parham; Guillén i Fàbregas, Albert
Mismatched Binary Hypothesis Testing: Error Exponent Sensitivity Boroumand, Parham; Guillén i Fàbregas, Albert