Error Exponents of Mismatched Likelihood Ratio Testing

Convex Optimization Project

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

BY:
SANGANA SAI VENKATA SURYA TEJA(EE23MTECH11032)
GEDELA MADHU BABU(EE23MTECH11027)

Contents:

Abstract:

We study 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.

Introduction and Preliminaries:

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:

$X=(x_1,x_2.........,x_n)$ are observations generated from two probability distributions P_1^n and P_2^n defined on probability simplex ƥ(χ^n) , $P_1^n > 0$ and $P_2^n > 0$ and $x \in \chi$.

We assume P_1^n and P_2^n are product distributions $P_1^n (X) = \prod_{i=1}^{n} P_1(x_i)$ and $P_2^n (X)$.

Likelihood Ratio Test :

Let ϕ: $\chi^n \rightarrow \{1,2\}$ be hypothesis test that decide which distribution generated the observation X.

Here we consider deterministic tests ϕ that decide in favor of P_1^n if A_1 subset of χ^n and P_2^n if A_2 subset of χ^n.

The test performance is determined in by using pair wise error probabilities:

$$\epsilon_1(\phi) = \sum_{\boldsymbol{x} \in \mathcal{A}_2} P_1^n(\boldsymbol{x}), \qquad \epsilon_2(\phi) = \sum_{\boldsymbol{x} \in \mathcal{A}_1} P_2^n(\boldsymbol{x}).$$

The hypothesis test is said to be optimal whenever it achieves the optimal error probability tradeoff given by:

$$\alpha_\beta = \min_{\phi : \epsilon_2(\phi) \le \beta} \epsilon_1(\phi).$$

Likelihood Ratio Test :

The LRT is defined as :

$$\phi_\gamma(\boldsymbol{x}) = \mathbb{1}\left\{\frac{P_2^n(\boldsymbol{x})}{P_1^n(\boldsymbol{x})} \geq e^{n\gamma}\right\} + 1.$$

It attains optimal trade off for every γ.

The LRT can also be expressed as a function of type of the observation $T\hat{}_x$ as:

$$\phi(\hat{T}_{\boldsymbol{x}}) = \mathbb{1}\{D(\hat{T}_{\boldsymbol{x}}\|P_0) - D(\hat{T}_{\boldsymbol{x}}\|P_1) \geq \gamma\},$$

Where $T\hat{}_x$ (a) = N(a/x)/n, N(a/x) is no., of occurrence of a symbol in the given sequence.

And D(P||Q) = $\sum_\chi P \log\left(\frac{P}{Q}\right)$, is called KL divergence.

The asymptotical delay of pair wise error probabilities as n tends to infinity:

$$E_0 \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log\epsilon_0(\phi), \quad E_1 \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log\epsilon_1(\phi).$$

Likelihood Ratio Test :

By using sanov's Theorem the optimal error exponent tradeoff (E_1, E_2) attained by LRT :

$$E_1(\phi_\gamma) = \min_{Q \in \mathcal{Q}_1(\gamma)} D(Q\|P_1),$$
$$E_2(\phi_\gamma) = \min_{Q \in \mathcal{Q}_2(\gamma)} D(Q\|P_2),$$

Where,

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

The minimizing distribution is the tilted distribution

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

where γ satisfies $-D(P_1||P_2) <= \gamma <= D(P_2||P_1)$.

Likelihood Ratio Test :

In this case λ is the solution of $D(Q_\lambda \| P_0) - D(Q_\lambda \| P_1) = \gamma.$

By using the Tilted(minimizing) distribution expression we can find the dual of optimal error exponent tradeoff (E_1 ,E_2) by solving it through Lagrangian :

$$E_1(\phi_\gamma) = \max_{\lambda \geq 0} \lambda\gamma - \log\Big(\sum_{x \in \mathcal{X}} P_1^{1-\lambda}(x) P_2^{\lambda}(x)\Big),$$
$$E_2(\phi_\gamma) = \max_{\lambda \geq 0} -\lambda\gamma - \log\Big(\sum_{x \in \mathcal{X}} P_1^{\lambda}(x) P_2^{1-\lambda}(x)\Big).$$

The stein Regime is defined as the highest error exponent under one hypothesis when the error probability under other hypothesis is at most some fixed ε ϵ (0,1/2).

$$E_2^{(\epsilon)} \triangleq \sup \big\{ E_2 \in \mathbb{R}_+ : \exists \phi, \exists n_0 \in \mathbb{Z}_+ \text{ s.t. } \forall n > n_0$$
$$\epsilon_1(\phi) \leq \epsilon \quad \text{and} \quad \epsilon_2(\phi) \leq e^{-nE_2} \big\}.$$

The optimal $E_2^{(\epsilon)}$, given by [3]

$$E_2^{(\epsilon)} = D(P_1 \| P_2),$$

Likelihood Ratio Test :

The optimal $E_2^{(\varepsilon)}$ can be achieved by setting the threshold in LRT to get

$$\gamma = -D(P_1||P_2) + C_2/n\hat{}0.5.$$

Where C_2 is a constant that depends on the distributions of P_1 , P_2 and ε.

2. Mismatched LRT:

Now, here in mismatched LRT in place of actual distributions we consider Test distribution with threshold Ϋ and LRT of mismatched given by:

$$\hat{\phi}_{\hat{\gamma}}(\hat{T}_{\boldsymbol{x}}) = \mathbb{1}\{D(\hat{T}_{\boldsymbol{x}}\|\hat{P}_1) - D(\hat{T}_{\boldsymbol{x}}\|\hat{P}_2) \geq \hat{\gamma}\}$$

$P\hat{}_1$ and $P\hat{}_2$ are test distributions where both are greater than zero.

Likelihood Ratio Test :

For fixed $\hat{P}_1$ and $\hat{P}_2$ ϵ ƥ(χ) the optimal error exponent tradeoff given by:

$$\hat{E}_1(\hat{\phi}_{\hat{\gamma}}) = \min_{Q \in \hat{\mathcal{Q}}_1(\hat{\gamma})} D(Q\|P_1)$$
$$\hat{E}_2(\hat{\phi}_{\hat{\gamma}}) = \min_{Q \in \hat{\mathcal{Q}}_2(\hat{\gamma})} D(Q\|P_2)$$

Where

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

Because two composite setting we are having two different distributions , so we are calculating two different Tilted distributions.

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

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

Likelihood Ratio Test :

The dual expressions for Type-1 and Type-2 errors are:

$$\hat{E}_1(\hat{\phi}_{\hat{\gamma}}) = \max_{\lambda \geq 0} \lambda\hat{\gamma} - \log\Big(\sum_{x\in\mathcal{X}} P_1(x)\hat{P}_1^{-\lambda}(x)P_2^{\lambda}(x)\Big),$$

$$\hat{E}_2(\hat{\phi}_{\hat{\gamma}}) = \max_{\lambda \geq 0} -\lambda\hat{\gamma} - \log\Big(\sum_{x\in\mathcal{X}} P_1^{\lambda}(x)P_2(x)\hat{P}_2^{-\lambda}(x)\Big).$$

3. Mismatched LRT with Uncertainty :

Here we analyze, the worst case error exponents tradeoff when the actual distributions P_1, P_2 are close to the mismatched test distributions P^$_1$, P^$_2$.

$$P_1 \in \mathcal{B}(\hat{P}_1, R_1), \quad P_2 \in \mathcal{B}(\hat{P}_2, R_2)$$

$$\mathcal{B}(Q,R) = \{P \in \mathcal{P}(\mathcal{X}) : D(Q\|P) \leq R\}$$

Likelihood Ratio Test :

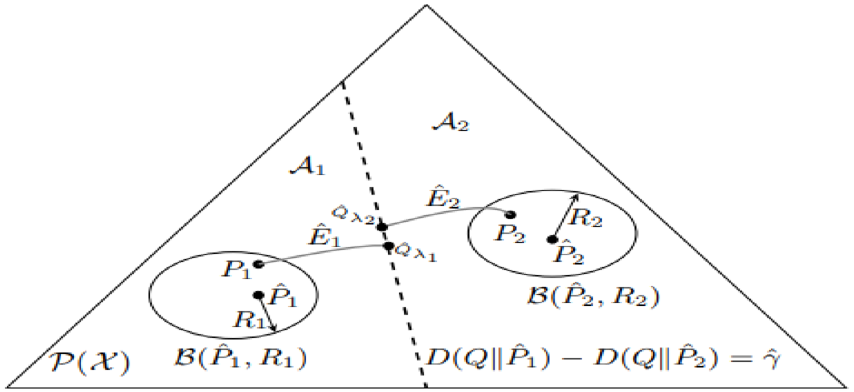


Fig. 1. Mismatched likelihood ratio test over distributions in D-balls.

In particular, we are interested in the least favorable distributions $P^L{}_1$, $P^L{}_2$ in $\beta(P^{\wedge}{}_1, R_1)$ and $\beta(P^{\wedge}{}_2, R_2)$ and these distributions achieving lowest error exponents $E^L{}_1(R_1)$, $E^L{}_2\,(R_2)$.

The least favorable exponents $E^L{}_1(R_1)$, $E^L{}_2\,(R_2)$ are defines as :

$$\hat{E}_1^L(R_1) = \min_{P_1 \in \mathcal{B}(\hat{P}_1, R_1)} \min_{Q \in \hat{\mathcal{Q}}_1(\hat{\gamma})} D(Q\|P_1),$$
$$\hat{E}_2^L(R_2) = \min_{P_2 \in \mathcal{B}(\hat{P}_2, R_2)} \min_{Q \in \hat{\mathcal{Q}}_2(\hat{\gamma})} D(Q\|P_2),$$

Here $Q^{\wedge}{}_1$(Ÿ) and $Q^{\wedge}{}_2$(Ÿ) are same as defined earlier.

Likelihood Ratio Test :

Here the corresponding error exponent pair satisfies:

$$\hat{E}_1^L(R_1) \leq \hat{E}_1(\hat{\phi}_{\hat{\gamma}}), \quad \hat{E}_2^L(R_2) \leq \hat{E}_2(\hat{\phi}_{\hat{\gamma}}).$$

Now by minimizing the optimization problem is with convex optimizing distributions:

$$Q_{\lambda_1}^L(x) = \frac{P_1^L(x)\hat{P}_1^{-\lambda_1}(x)\hat{P}_2^{\lambda_1}(x)}{\sum_{a\in\mathcal{X}} P_1^L(a)\hat{P}_1^{-\lambda_1}(a)\hat{P}_2^{\lambda_1}(a)},$$
$$P_1^L(x) = \beta_1 Q_{\lambda_1}^L(x) + (1-\beta_1)\hat{P}_1(x),$$

Where λ_1>=0 , 0<= β_1<=1 are chosen such that:

$$D(Q_{\lambda_1}^L \| \hat{P}_1) - D(Q_{\lambda_1}^L \| \hat{P}_2) = \hat{\gamma},$$
$$D(\hat{P}_1 \| P_1^L) = R_1,$$

When

$$\max_{P_1 \in \mathcal{B}(\hat{P}_1, R_1)} D(P_1 \| \hat{P}_1) - D(P_1 \| \hat{P}_2) \leq \hat{\gamma}.$$

Otherwise we can find the least favorable distribution $P^L_1 \in \mathcal{B}(P\hat{}_1, R_1)$ such that $E\hat{}_1(\phi\hat{}_\gamma) = 0$.

Likelihood Ratio Test :

Similarly, the optimization of E^L_2 (R_2) is convex with optimizing distributions :

$$Q^L_{\lambda_2}(x) = \frac{P^L_2(x)\hat{P}_2^{-\lambda_2}(x)\hat{P}_1^{\lambda_2}(x)}{\sum_{a\in\mathcal{X}} P^L_2(a)\hat{P}_2^{-\lambda_2}(a)\hat{P}_1^{\lambda_2}(a)},$$

$$P^L_2(x) = \beta_2 Q^L_{\lambda_2}(x) + (1-\beta_2)\hat{P}_2(x),$$

where $\lambda_2 \geq 0, 0 \leq \beta_2 \leq 1$ *are chosen such that*

$$D(Q^L_{\lambda_2}\|\hat{P}_2) - DQ^L_{\lambda_2}\|\hat{P}_1) = \hat{\gamma},$$

$$D(\hat{P}_2\|P^L_2) = R_2,$$

whenever,

$$\min_{P_2\in\mathcal{B}(\hat{P}_2,R_2)} D(P_2\|\hat{P}_1) - D(P_2\|\hat{P}_2) \geq \hat{\gamma}.$$

Otherwise we can find the least favorable distribution P^L_2 ϵ $\beta(P\hat{}_2, R_2)$ such that $E\hat{}_2(\phi\hat{}_\gamma)$ =0.

The worst-case achievable error exponents of mismatched LRT for data distributions in D-ball oer essentially the minimum relative entropy between two sets of probability distributions. Specifically, Relative entropy between $\beta(P\hat{}_1, R_1)$ and $Q\hat{}_2(\ddot{\Upsilon})$ gives $E^L_1(R_1)$ and similarly for $E^L_2(R_2)$

Likelihood Ratio Test :

4. Mismatched LRT Sensitivity:

We will know how the worst case error exponents ($E\text{^}^{L}_{1}$,$E\text{^}^{L}_{2}$) behave when the D-ball radii R_1 and R_2 are small. For this we derive taylor expansion of the worst case error exponents. This approximation can also be interpreted as the worst case sensitivity of the test i.e., how does the test perform when actual distributions are very close to the mismatched distributions.

Theorem 4. *For every $R_i \geq 0$, $\hat{P}_i \in \mathcal{P}(\mathcal{X})$ for $i = 1, 2$, and*

$$-D(\hat{P}_1 \| \hat{P}_2) \leq \hat{\gamma} \leq D(\hat{P}_2 \| \hat{P}_1), \tag{45}$$

we have

$$\hat{E}_i^L(R_i) = E_i(\hat{\phi}_{\hat{\gamma}}) - S_i(\hat{P}_1, \hat{P}_2, \hat{\gamma})\sqrt{R_i} + o(\sqrt{R_i}), \tag{46}$$

where

$$S_i^2(\hat{P}_1, \hat{P}_2, \hat{\gamma}) = 2Var_{\hat{P}_i}\left(\frac{\hat{Q}_\lambda(X)}{\hat{P}_i(X)}\right) \tag{47}$$

and $\hat{Q}_\lambda(X)$ is the minimizing distribution in (10) *for test $\hat{\phi}_{\hat{\gamma}}$.*

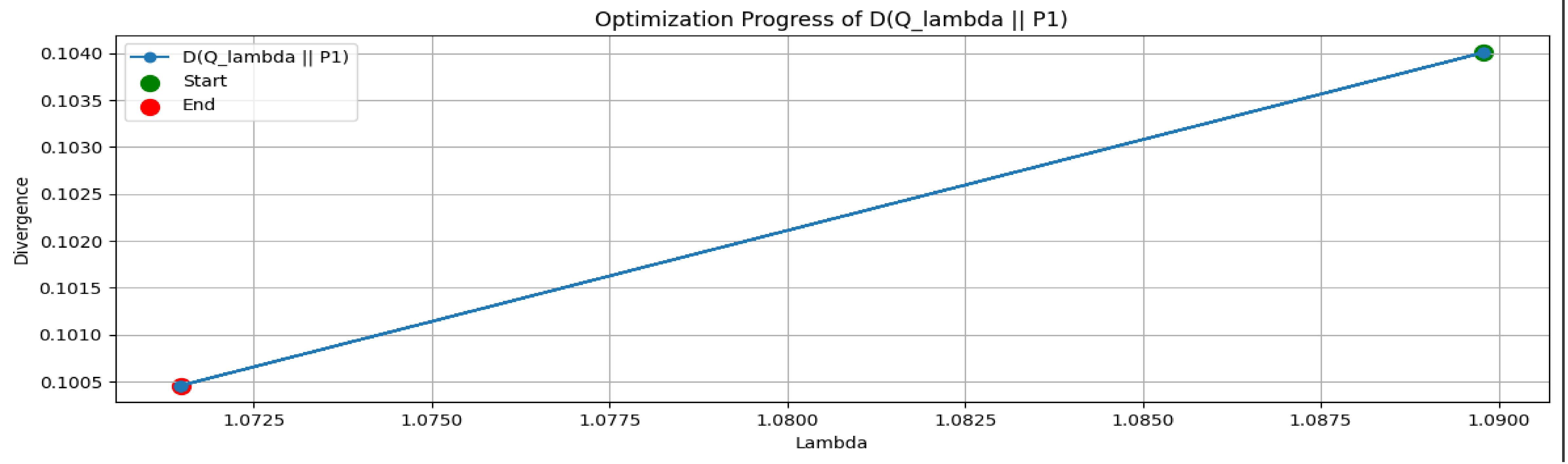
Results :

Solved Langrange Equation for finding the minimizing distribution:
(i.e tilted distribution)

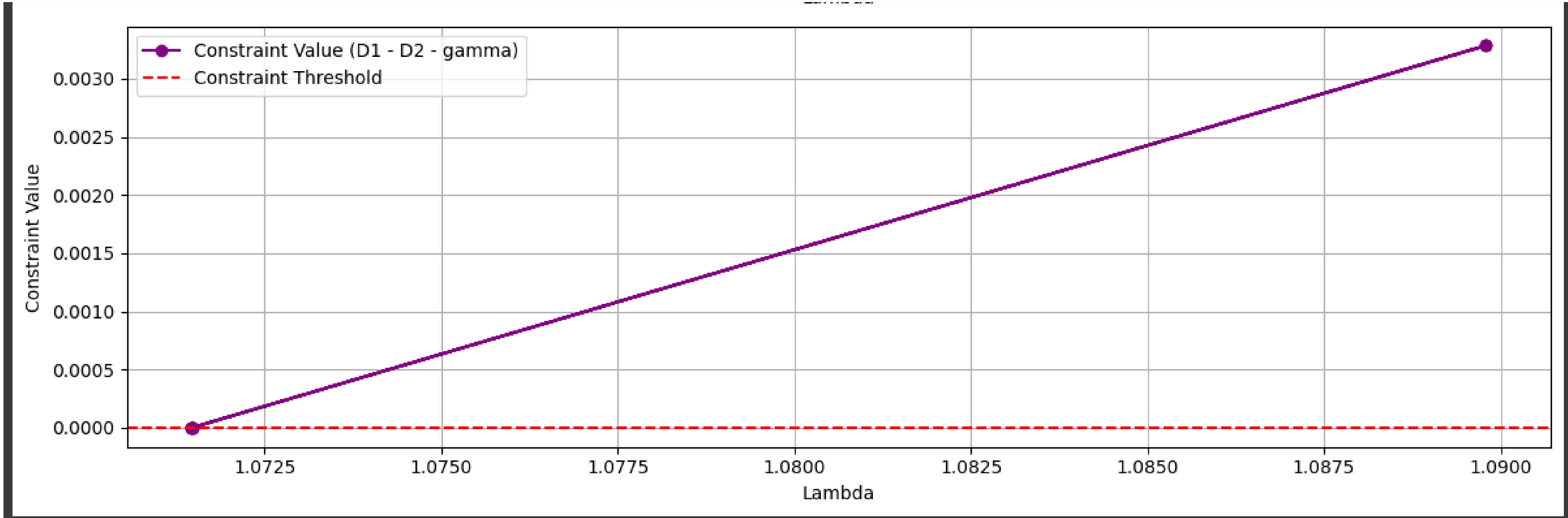
$$Q_\lambda(x) = \frac{P_1^{1-\lambda}(x)P_2^{\lambda}(x)}{\sum_{a\in\mathcal{X}} P_1^{1-\lambda}(a)P_2^{\lambda}(a)}, \quad 0 \leq \lambda \leq 1$$

Used tilted distribution to find Optimum Divergence and Optimum lambda Value for Two Distributions(Prabablities):

```
Optimization terminated successfully    (Exit mode 0)
            Current function value: 0.1004581674661636
            Iterations: 3
            Function evaluations: 7
            Gradient evaluations: 3
```



Plot of the Constrain Value when minimizing the Divergence:




```
Optimization successful.
Optimized lambda: 1.0714694562122697
Optimal Q: [0.51513435 0.48486565]
Minimum D(Q_lambda || P1): 0.1004581674661636
```

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

THANK YOU