Error Exponents of Mismatched Likelihood Ratio Testing:-

Given an observation, $X = (x_1, x_2, \dots x_n) \in \mathcal{X}^n$ which is generated from two distributions P_1^n & P_2^n. (defined on probability simplex $\mathcal{P}(\mathcal{X})$)

P_1^n & P_2^n are product distributions i.e.,

$$P_1^n(X) = \prod_{i=1}^{n} P_1(x_i) \qquad \text{where } P_1(x) > 0 \text{ \&}$$
$$P_2^n(X) = \prod_{i=1}^{n} P_2(x_i) \qquad P_2(x) > 0.$$

and $x \in \mathcal{X}$.

→ Likelihood Ratio Test (LRT) $\phi : \mathcal{X}^n \to \{1, 2\}$

$H_1 : \phi(\mathcal{X}^n) \sim P_1^n$

$H_2 : \phi(\mathcal{X}^n) \sim P_2^n$

for H_1, $x \in A_1$ & for H_2; $x \in A_2$ i.e., $\mathcal{X}^n \setminus A_1$

we know that type-I error & Type-II error

i.e., $P_r(H_2/H_1) = \epsilon_1(\phi)$ i.e., $P_r(H_1/H_2) = \epsilon_2(\phi)$.

$$\epsilon_1(\phi) = \sum_{x \in A_2} P_1^n(x) \to ① \qquad \epsilon_2(\phi) = \sum_{x \in A_1} P_2^n(x) \to ②$$

– The optimal error probability tradeoff,

$$\alpha_\beta = \min_{\phi : \epsilon_2(\phi) \le \beta} \epsilon_1(\phi) \quad ; \quad \beta \in [0, 1]$$

→ The likelihood Ratio, denoted $\Lambda(x)$, is the ratio of the likelihood of observation under P_2^n to the likelihood under P_1^n.

$$\Lambda(x) = \frac{P_2^n(x)}{P_1^n(x)}$$

Apply log on both sides i.e.,

$$r(x) = \log \Lambda(x) = \log\left(\frac{P_2^n(x)}{P_1^n(x)}\right)$$

Let the threshold r such that if $r(x) \geq nr$ we decide in favour of P_2^n.

$$\phi_r(x) = \begin{cases} 1 & \text{if } r(x) \geq nr \\ 0 & \text{o.w.} \end{cases}$$

$$\phi_r(x) = \begin{cases} 1 & \text{if } \frac{P_2^n(x)}{P_1^n(x)} \geq e^{nr} \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore \phi_r(x) = \mathbb{1}\left\{ \frac{P_2^n(x)}{P_1^n(x)} \geq e^{nr} \right\} \rightarrow (3)$$

Let us find the type of the sequence,

$$\hat{T}_x(a) = \frac{N(a/x)}{n}, \quad N(a/x) \text{ is no. of occurance of } a \text{ in } x.$$

we know that Def. of KL divergence, i.e.,

$$D(P\|Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

LRT under two different probability distributions P_1 & P_2. If $\hat{T}_x(x)$ represents empirical distribution of observed data x.

$$\frac{P_2(x)}{P_1(x)} = e^{n(D(\hat{T}_x\|P_1) - D(\hat{T}_x\|P_2))}$$

$$\Rightarrow \frac{P_2(x)}{P_1(x)} = e^{n\left(\sum_{x \in \mathcal{X}} \hat{T}_x(a) \log \frac{\hat{T}_x(a)}{P_1(a)} - \sum_{x \in \mathcal{X}} \hat{T}_x(a) \log \frac{\hat{T}_x(a)}{P_2(a)}\right)}$$

$$\Rightarrow \frac{P_2(x)}{P_1(x)} = \prod_{x \in \mathcal{X}} \left(\frac{\hat{T}_x(a)}{P_1(x)}\right)^{n\hat{T}_x(x)} \times \prod_{x \in \mathcal{X}} \left(\frac{P_2(x)}{\hat{T}_x(x)}\right)^{n\hat{T}_x(x)}$$

By interpreting in equality,

$$\frac{P_2(x)}{P_1(x)} \geq e^{r}$$

This implies that if product of likelihood ratios is greater than (or) equal to e^{r}, the likelihood of P_2 compared to P_1 is sufficiently high based on the observed data, leading to a decision in favour of P_2 over P_1 in likelihood Ratio Test

Therefore, the likelihood Ratio Test is equivalent to comparing the likelihood ratio of P_2 to P_1 to a threshold e^r, which is a common way of express the test in terms of KL-divergence

$$\therefore \phi_r(x) = \mathbb{1}\{D(T_x \| P_1) - D(T_x \| P_2) \geq r\} \rightarrow (4)$$

* Asymptotic exponential decay means, $n \to \infty$. then pairwise error probability,

$$E_1 \triangleq \lim_{n\to\infty} \inf -\frac{1}{n} \log \epsilon_1(\phi) \rightarrow (5)$$

$$E_2 \triangleq \lim_{n\to\infty} \inf -\frac{1}{n} \log \epsilon_2(\phi) \rightarrow (6)$$

$$\left[\because A f p = \frac{1}{n} \log \frac{1}{\epsilon_{rror}} = -\frac{1}{n} \log \epsilon_{rror}\right].$$

→. Let all the observations are iid, then optimal Exponential trade off $(E_1 \& E_2)$ is defined as,

$$E_2(E_1) \triangleq \sup \{E_2 \in R_+ ; \exists \phi, \exists n \in Z_+ \text{ s.t. } \forall n > n_0 \; \epsilon_1(\phi) \leq e^{-nE_1} \text{ and } \epsilon_2(\phi) \leq e^{-nE_2}\} \rightarrow (7)$$

* <u>Sanov's Theorem</u>: let $x_1, x_2 \ldots x_n$ be iid $\sim Q(x)$. Let $E \subseteq P$ be a set of probability distribution, then,

$$Q^n(E) = Q^n(E \cap P_n) \leq (n+1)^{|X|} 2^{-nD(P^* \| Q)}.$$

where $P^* = \arg\min D(P \| Q)$ is the distribution in E i.e, the closest to Q in relative entropy. If in addition, the set E is the closure of its interior then

$$\frac{1}{n} Q^n(E) \longrightarrow -D(P^* \| Q).$$

By using theorem and eq ⑤ & ⑥ we can write.

Optimal error tradeoff (E_1, E_2) is,

$$E_1(\phi_r) = \min_{Q \in \mathcal{Q}_1(r)} D(Q \| P_1) \rightarrow ⑧$$

$$E_2(\phi_r) = \min_{Q \in \mathcal{Q}_2(r)} D(Q \| P_2) \rightarrow ⑨$$

where, $\mathcal{Q}_1(r) = \{ Q \in \mathcal{D}(\mathcal{X}) : D(Q \| P_1) - D(Q \| P_2) \geq r \}$ ⑩

$$\mathcal{Q}_2(r) = \{ Q \in \mathcal{D}(\mathcal{X}) : D(Q \| P_1) - D(Q \| P_2) \leq r \} \rightarrow ⑪$$

The minimizing distribution of 8 & 9 we get the tilted distribution, to reduce the value of r.

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

if $\lambda = 0$ then $Q_\lambda(x) = \dfrac{P_1(x)}{\sum_{a \in \mathcal{X}} P_1(a)}$

if $\lambda = 1$ then, $Q_\lambda(x) = \dfrac{P_2(x)}{\sum_{a \in \mathcal{X}} P_2(a)}$

This tilted distributions are used to adjust the decision boundary based on the likelihood ratio of observation under the two distributions being compared.

This adjustment helps in optimizing the trade off between Type-I & Type-II error in hypothesis.

The minimizing distribution in E_1 & E_2 is the tilted distribution, when ever (r) satisfies

$$-D(P_1 \| P_2) \leq r \leq D(P_2 \| P_1).$$

In this case α is the solution of

$$D(\phi_\alpha \| P_1) - D(\phi_\alpha \| P_0) = \gamma.$$

Here the parameter α is choosen such that the relative entropy between $\phi_\alpha(x)$ & $P_1(x)$ is equal to γ, where γ is with in the rang determined by relative entropy between $P_1(x)$ & $P_2(x)$.

This ensures that the likelihood ratio test remains valid with in specified range γ.

→ If $\gamma < -D(P_1 \| P_2)$, indicating a very low likelihood of second hypothesis then it is optimal set $\phi_\alpha(x) = P_1(x)$ resulting $E_1(\phi) = 0$

→ if $\gamma > D(P_2 \| P_1)$ viceversa i.e $\phi_\alpha(x) = P_2(x)$

resulting $E_2(\phi) = 0$.

$$E_1(\phi_y) = \min_{Q \in Q_1(y)} D(Q \| P_1) \quad \text{—— (1)}$$

$$E_2(\phi_y) = \min_{Q \in Q_2(y)} D(Q \| P_2) \quad \text{—— (2)}$$

where

$$Q_1(y) = \{ Q \in P(X) : D(Q\|P_1) - D(Q\|P_2) \geq y \}$$

$$Q_2(y) = \{ Q \in P(X) : D(Q\|P_1) - D(Q\|P_2) \leq y \}.$$

For (1) :-

Langrangian function:-

$$L(Q, \lambda) = D(Q\|P_1) - \lambda \left(D(Q\|P_1) - D(Q\|P_2) - y \right)$$

$$= \sum_x Q(x) \log \frac{Q(x)}{P_1(x)} - \lambda \left(\sum_x Q(x) \log \frac{Q(x)}{P_1(x)} - \sum_x Q(x) \log \frac{Q(x)}{P_2(x)} - y \right)$$

$$= \sum_x Q(x)(1-\lambda) \log \frac{Q(x)}{P_1(x)} + \lambda \sum_x Q(x) \log \frac{Q(x)}{P_2(x)} + \lambda y$$

→ Derivative w.r.t $Q(x)$

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

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

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

$$\therefore Q(x) = P_2(x)^{\lambda} \, P_1(x)^{(1-\lambda)}$$

∴ The minimizing distribution after Normalization

Tilted distribution :- $Q_\lambda(x) = \dfrac{P_1^{1-\lambda}(x)\, P_2^{\lambda}(x)}{\sum_{a \in \mathcal{X}} P_1^{1-\lambda}(a)\, P_2^{\lambda}(a)}$; $0 \le \lambda \le 1$

whenever y satisfies $-D(P_1 \| P_2) \le y \le D(P_2 \| P_1)$

→ In this case λ is the solution of

$$D(Q_\lambda \| P_1) - D(Q_\lambda \| P_2) = y.$$

→ If $y < -D(P_1 \| P_2)$, the optimal distribution is $Q_\lambda(x) = P_1(x)$ and $F_1(\phi_y) = 0$

$$E_2^*(\phi_1) = \max_{\lambda \ge 0} -\lambda\gamma - \log\left(\sum_{x\in\mathcal{X}} P_1^{\lambda}(x)\, P_2^{1-\lambda}(x)\right).$$

* The Stein regime refers to the scenario where the error exponent under one hypothesis E_2 is maximized ~~by~~ while ensuring that the error probability under the other hypothesis is below a certain threshold

$0 < \epsilon < 1/2$

$$E_2^{(\epsilon)} \triangleq \sup\left\{ E_2 \in \mathbb{R}_+ : \exists \phi, \exists n_0 \in \mathbb{Z}_+ \text{ s.t. } \forall n > n_0, \; E_1(\phi) \le \epsilon \text{ and } E_2(\phi) \le e^{-nE_2} \right\}$$

The optimal $E_2^{(\epsilon)} = D(P_1 \| P_2)$.

Here $E_2^{(\epsilon)}$ denotes the highest error exponent under hypothesis such that error probability under hypothesis 1 is less than or equal to ϵ.

$$\gamma = -D(P_1\|P_2) + \sqrt{c_2/n}$$ where c_2 is constant depending on P_1 & P_2 & n = sample size

(ii) Mismatched Likelihood ratio testing:

Let $\hat{P}_1(x)$ & $\hat{P}_2(x)$ be the test distributions used in the likelihood Ratio test with threshold $\hat{\gamma}$ given by

$$\phi_{\hat{\gamma}}(\hat{T}_x) = \mathbb{1}\left\{ D(\hat{T}_x \| \hat{P}_1) - D(\hat{T}_x \| \hat{P}_2) \ge \hat{\gamma} \right\}.$$

$$\hat{P}_1(x) > 0 \;\&\; \hat{P}_2(x) > 0 \quad \forall x \in \mathcal{X}.$$

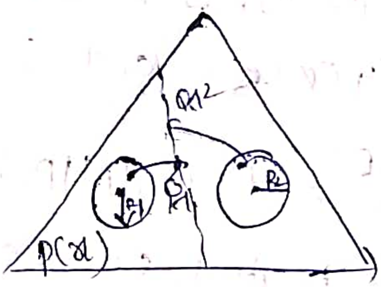
Similarly as above we calculate every equation by changing actual probabilities with test distributions (distribution)

(iii) Mismatched LRT with uncertainty :

We analyze worst case error exponents tradeoff when the actual distributions P_1, P_2 are close mismatched distributions $\hat{P}_1, \hat{P}_2$.

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

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



Here in this simplex, from fig in paper The LRT divides the probability space into two decision regions based on hyperplane

Then mathematical representation

$$D(Q \| \hat{P}_1) - D(Q \| \hat{P}_2) = \gamma$$

From this, we aim to find the least favorable distribution P_1^L & P_2^L within D-ball, which achieve the lowest error exponent $E_1^L(R_1)$ & $E_2^L(R_2)$.

(iv) Mismatched LRT sensitivity :

Here we study how worst case error exponent $E_1^L(R_1), E_2^L(R_2)$ behave, when R_1 & R_2 are small for this we use theorem 4.