

Information Theory, Statistics, and Learning

Assignment 1

Ali Abbasi
98105879

December 3, 2022

Contents

1	Contiguity!	3
1.1	3
1.2	3
2	Large Deviation for Log-Likelihood	4
2.1	4
2.2	5
2.3	6
3	Large Deviation for Erlang Distribution	6
3.1	6
3.2	8
3.3	9
4	Chernoff Deviation	9
5	Some Properties of Total Variation	12
5.1	12
5.2	13
5.3	14
5.4	14
6	Bayesian Approach and Properties of $\mathcal{R}(P, Q)$	16
6.1	16
6.2	16
6.3	16
7	Achievable Rates	17
7.1	17
7.2	17
7.3	19

7.4	19
7.5	19

1 Contiguity!

1.1

Proof by contradiction: Assume that sum of type 1 and type 2 errors converges to zero as $n \rightarrow \infty$ and $\mathbb{P} \triangleleft \mathbb{Q}$. We have:

$$P_n(Z = 1) + Q_n(Z = 0) \xrightarrow{n \rightarrow \infty} 0 \quad (1.1)$$

$$\xrightarrow{\substack{P_n \text{ and } Q_n \text{ are positive}}} \begin{cases} P_n(Z = 1) \xrightarrow{n \rightarrow \infty} 0 \\ Q_n(Z = 0) \xrightarrow{n \rightarrow \infty} 0 \end{cases} \implies P_n(Z = 0) \xrightarrow{n \rightarrow \infty} 1 \quad (1.2)$$

Consider sequence of $\{A_n\}_{n=1}^\infty$ such that $\forall n : A_n$ is equal to the event $Z = 0$:

$$Q_n(A_n) \xrightarrow{n \rightarrow \infty} 0 \quad (1.3)$$

$$\xrightarrow{\mathbb{P} \triangleleft \mathbb{Q}} P_n(A_n) \xrightarrow{n \rightarrow \infty} 0 \quad (1.4)$$

$$\implies P_n(Z = 0) \xrightarrow{n \rightarrow \infty} 0 \quad (1.5)$$

which is clearly in contradiction with what we derived in (1.2).

The same argument can be used to show that when $\mathbb{Q} \triangleleft \mathbb{P}$ then sum of type 1 and type 2 errors cannot converge to zero as $n \rightarrow \infty$ (in that case, we use the event $Z = 1$ instead of $Z = 0$ as A_n in the above argument).

1.2

Proof by contradiction: Suppose $\mathbb{P} \triangleleft \mathbb{Q}$ does not hold. Then there exists a sequence of $\{A_n\}_{n=1}^\infty$ such that $Q_n(A_n) \xrightarrow{n \rightarrow \infty} 0$ and $P_n(A_n) \xrightarrow{n \rightarrow \infty} a < \infty$.

We have:

$$\|L_n\|^2 = \langle L_n, L_n \rangle \quad (1.6)$$

$$= \mathbb{E}_{Q_n} [L_n^2] \quad (1.7)$$

$$= \mathbb{E}_{X \sim Q_n} \left[\frac{P_n(x)^2}{Q_n(x)^2} \right] \quad (1.8)$$

$$= \sum_x \frac{P_n(x)^2}{Q_n(x)} \quad (1.9)$$

$$\geq \sum_{x \in A_n} \frac{P_n(x)^2}{Q_n(x)} \quad (1.10)$$

Using Cauchy-Schwarz inequality:

$$\sum_{x \in A_n} \frac{P_n(x)^2}{Q_n(x)} \sum_{x \in A_n} Q_n(x) \geq \left(\sum_{x \in A_n} P_n(x) \right)^2 \quad (1.11)$$

$$\sum_{x \in A_n} \frac{P_n(x)^2}{Q_n(x)} \times Q_n(A_n) \geq P_n(A_n)^2 \quad (1.12)$$

$$\implies \sum_{x \in A_n} \frac{P_n(x)^2}{Q_n(x)} \geq \frac{P_n(A_n)^2}{Q_n(A_n)} \quad (1.13)$$

$$\implies \|L_n\|^2 \geq \sum_{x \in A_n} \frac{P_n(x)^2}{Q_n(x)} \geq \frac{P_n(A_n)^2}{Q_n(A_n)} \quad (1.14)$$

$$\implies \lim_{n \rightarrow \infty} \|L_n\|^2 \geq \lim_{n \rightarrow \infty} \frac{P_n(A_n)^2}{Q_n(A_n)} = \infty \quad (1.15)$$

Which is in contradiction with our hypothesis. Thus $\mathbb{P} \triangleleft \mathbb{Q}$ holds.

2 Large Deviation for Log-Likelihood

We want to show that:

$$\forall x \geq 0, n \in \mathbb{N}: \quad \mathbb{P} \left[\sum_{i=1}^n (Y_i - X_i) \geq nx \right] \leq \exp(-n(\alpha + \frac{x}{2})) \quad (2.1)$$

Where

$$\begin{aligned} X_i &= \log \frac{p(r_i)}{q(r_i)}, \quad r_i \stackrel{\text{i.i.d.}}{\sim} P \\ Y_i &= \log \frac{p(t_i)}{q(t_i)}, \quad t_i \stackrel{\text{i.i.d.}}{\sim} Q \\ \alpha &= -2 \log \mathcal{B}(P, Q) \end{aligned}$$

2.1

We use Chernoff bound to prove it:

$$\mathbb{P}[X \geq \epsilon] \leq \frac{\mathbb{E}[\exp(\theta X)]}{\exp(\theta \epsilon)} \quad (\text{Chernoff bound})$$

where $\theta \geq 0$. We have:

$$\mathbb{P} \left[\sum_{i=1}^n (Y_i - X_i) \geq nx \right] \leq \frac{\mathbb{E} [e^{\theta \sum_{i=1}^n (Y_i - X_i)}]}{e^{\theta nx}} \quad (2.2)$$

$$= \frac{\prod_{i=1}^n \mathbb{E} [e^{\theta(Y_i - X_i)}]}{e^{\theta nx}} \quad (2.3)$$

$$= \frac{\prod_{i=1}^n \mathbb{E}_Q [e^{\theta Y_i}] \mathbb{E}_P [e^{-\theta X_i}]}{e^{\theta nx}} \quad (2.4)$$

$$= \frac{\mathbb{E}_Q [e^{\theta Y_1}]^n \mathbb{E}_P [e^{-\theta X_1}]^n}{e^{\theta nx}} \quad (2.5)$$

$$= \exp [-n (\theta x - \log \mathbb{E}_P [e^{-\theta X_1}] - \log \mathbb{E}_Q [e^{\theta Y_1}])] \quad (2.6)$$

$$= \exp [-n (\theta x - \psi_P(-\theta) - \psi_Q(\theta))] \quad (2.7)$$

Using the fact that X s and Y s are independent in (2.3) and (2.4). This inequality is true for all $\theta \geq 0$. So it also holds for the inf value of RHS with respect to θ :

$$\mathbb{P} \left[\sum_{i=1}^n (Y_i - X_i) \geq nx \right] \leq \inf_{\theta} \exp [-n (\theta x - \psi_P(-\theta) - \psi_Q(\theta))] \quad (2.8)$$

$$= \exp \left[-n \times \sup_{\theta} \{ \theta x - \psi_P(-\theta) - \psi_Q(\theta) \} \right] \quad (2.9)$$

$$= \exp [-nF(x)] \quad (2.10)$$

2.2

First we show that $\psi_P(-\theta) + \psi_Q(\theta) \geq 2 \log \mathcal{B}(P, Q)$:

$$\psi_P(-\theta) + \psi_Q(\theta) = \log \mathbb{E}_P [e^{-\theta X_1}] + \log \mathbb{E}_Q [e^{\theta Y_1}] \quad (2.11)$$

$$= \log [\mathbb{E}_P [e^{-\theta X_1}] \mathbb{E}_Q [e^{\theta Y_1}]] \quad (2.12)$$

$$= \log \left[\mathbb{E}_P \left[\left(\frac{dP}{dQ} \right)^{-\theta} \right] \mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\theta} \right] \right] \quad (2.13)$$

$$= \log \left[\mathbb{E}_P \left[\left(\frac{dQ}{dP} \right)^{\theta} \right] \mathbb{E}_Q \left[\left(\frac{dP}{dQ} \right)^{\theta} \right] \right] \quad (2.14)$$

$$= \log \left[\left(\sum_t p(t)^{1-\theta} q(t)^{\theta} \right) \left(\sum_t p(t)^{\theta} q(t)^{1-\theta} \right) \right] \quad (2.15)$$

Using Cauchy Schwarz inequality, we have:

$$\psi_P(-\theta) + \psi_Q(\theta) \geq \log \left[\left(\sum_t \sqrt{p(t)q(t)} \right)^2 \right] \quad (2.16)$$

$$= \log [\mathcal{B}(P, Q)]^2 \quad (2.17)$$

$$= 2 \log \mathcal{B}(P, Q) \quad (2.18)$$

So:

$$-\psi_P(-\theta) - \psi_Q(\theta) \leq -2 \log \mathcal{B}(P, Q) = \alpha \quad (2.19)$$

If we define $f(\theta) \triangleq -\psi_P(-\theta) - \psi_Q(\theta)$, we show that $F(0) = f(\frac{1}{2}) = \alpha$:

$$f(\frac{1}{2}) = -\psi_P(-\frac{1}{2}) - \psi_Q(\frac{1}{2}) \quad (2.20)$$

$$= -\log \left[\left(\sum_t p(t)^{\frac{1}{2}} q(t)^{\frac{1}{2}} \right) \left(\sum_t p(t)^{\frac{1}{2}} q(t)^{\frac{1}{2}} \right) \right] \quad \text{using Eq. ((2.15))} \quad (2.21)$$

$$= -\log [\mathcal{B}(P, Q)]^2 \quad (2.22)$$

$$= -2 \log \mathcal{B}(P, Q) \quad (2.23)$$

$$= \alpha \quad (2.24)$$

$$\left. \begin{array}{l} F(0) = \sup_{\theta \geq 0} f(\theta) \\ \forall \theta \geq 0, \quad f(\theta) \leq f(\frac{1}{2}) \end{array} \right\} \implies F(0) = f(\frac{1}{2}) = \alpha \quad (2.25)$$

2.3

$$F(x) = \sup_{\theta \geq 0} \{x\theta + f(\theta)\} \quad (2.26)$$

$$\geq x\theta + f(\theta) \Big|_{\theta=\frac{1}{2}} \quad (2.27)$$

$$= \frac{x}{2} + F(0) \quad (2.28)$$

So we have:

$$\mathbb{P} \left[\sum_{i=1}^n (Y_i - X_i) \geq nx \right] \leq \exp [-nF(x)] \quad (2.29)$$

$$\leq \exp \left[-n \left(\frac{x}{2} + F(0) \right) \right] \quad (2.30)$$

$$\leq \exp \left[-n \left(\frac{x}{2} + \alpha \right) \right] \quad (2.31)$$

3 Large Deviation for Erlang Distribution

3.1

We are going to solve this part using moment generating function and Laplace transform! We denote moment generating function, of random variable X , its pdf and Laplace transform of its pdf

respectively as $M_X(t)$, $f_X(x)$ and $\mathcal{F}_X(s)$. We know:

$$M_X(t) = \mathbb{E} [e^{tX}] \quad (3.1)$$

$$= \int_0^\infty e^{tx} f_X(x) dx \quad (3.2)$$

$$= \mathcal{F}_X(-t) \quad (3.3)$$

So we will use Laplace transform and its inverse to convert pdf to moment generating function and vice versa. We have:

$$Y_i \stackrel{\text{i.i.d.}}{\sim} \exp(\lambda) \quad (3.4)$$

$$X = \sum_{i=1}^n Y_i \sim \text{Erlang}(n, \lambda) \quad (3.5)$$

$$M_X(t) = \mathbb{E} [e^{tX}] \quad (3.6)$$

$$= \mathbb{E} [e^{t \sum_{i=1}^n Y_i}] \quad \text{i.i.d.} \quad (3.7)$$

$$= \prod_{i=1}^n \mathbb{E} [e^{tY_i}] \quad (3.8)$$

$$= \prod_{i=1}^n M_{Y_i}(t) \quad (3.9)$$

$$= M_{Y_1}(t)^n \quad (3.10)$$

$$f_{Y_1} = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 \quad (3.11)$$

$$= \lambda e^{-\lambda x} u(x) \quad (3.12)$$

$$\mathcal{F}_{Y_1}(s) = \lambda \times \frac{1}{s + \lambda} = \frac{\lambda}{s + \lambda} \quad (3.13)$$

$$M_{Y_1}(t) = \mathcal{F}_{Y_1}(-t) = \frac{\lambda}{-t + \lambda} \quad (3.14)$$

$$\implies M_X(t) = \left(\frac{\lambda}{-t + \lambda} \right)^n \quad (3.15)$$

$$\implies \mathcal{F}_X(s) = \frac{\lambda^n}{(s + \lambda)^n} \quad (3.16)$$

$$\mathcal{L}^{-1} \left[\frac{1}{(s + \lambda)^n} \right] = \frac{x^{n-1}}{(n-1)!} e^{-\lambda x} \quad (3.17)$$

$$\implies f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad (3.18)$$

3.2

We will use inequality we had in large deviation to prove this part.

$$X_i \stackrel{\text{i.i.d.}}{\sim} \exp(1) \quad (3.19)$$

$$\mathbb{P} \left[\sum_{i=1}^n X_i \geq n\xi \right] \stackrel{t \geq 0}{=} \mathbb{P} \left[t \sum_{i=1}^n X_i \geq tn\xi \right] \quad (3.20)$$

$$= \mathbb{P} \left[e^{t \sum_{i=1}^n X_i} \geq e^{tn\xi} \right] \quad (3.21)$$

$$\leq e^{-tn\xi} \mathbb{E} [e^{tX_1}]^n \quad (3.22)$$

$$= \exp [-n(t\xi - \log M_{X_1}(t))] \quad (3.23)$$

$$\log M_{X_1}(t) = \log \frac{1}{-t+1} = -\log(1-t) \quad (3.24)$$

$$\implies \mathbb{P} \left[\sum_{i=1}^n X_i \geq n\xi \right] \leq \exp [-n(t\xi + \log(1-t))] \quad (3.25)$$

Since it holds for all $t \geq 0$, it also holds for t that makes RHS infimum:

$$\mathbb{P} \left[\sum_{i=1}^n X_i \geq n\xi \right] \leq \inf_{t \geq 0} \exp [-n(t\xi + \log(1-t))] \quad (3.26)$$

$$= \exp \left[-n(\sup_{t \geq 0} \{t\xi + \log(1-t)\}) \right] \quad (3.27)$$

We can find supremum value by taking derivative of it:

$$\frac{\partial}{\partial t} [t\xi + \log(1-t)] = \xi - \frac{1}{1-t} = 0 \quad (3.28)$$

$$\implies t = 1 - \frac{1}{\xi} \quad (t \geq 0 \implies \xi \geq 1) \quad (3.29)$$

$$\implies \sup_{t \geq 0} \{t\xi + \log(1-t)\} = \xi - \log \xi - 1 \quad (3.30)$$

$$\implies \mathbb{P} \left[\sum_{i=1}^n X_i \geq n\xi \right] \leq \exp [-n(\xi - \log \xi - 1)] \quad \text{for } \xi \geq 1 \quad (3.31)$$

3.3

This part can be proved in a very similar way; we only need to multiply inequality in the \mathbb{P} with a negative t in order to reverse the inequality.

$$\mathbb{P} \left[\sum_{i=1}^n X_i \leq n\xi \right] \stackrel{t \leq 0}{=} \mathbb{P} \left[t \sum_{i=1}^n X_i \geq tn\xi \right] \quad (3.32)$$

$$= \mathbb{P} \left[e^{t \sum_{i=1}^n X_i} \geq e^{tn\xi} \right] \quad (3.33)$$

\vdots

$$\leq \exp \left[-n(\sup_{t \leq 0} \{t\xi + \log(1-t)\}) \right] \quad (3.34)$$

Again by taking derivative, we can find supremum value, which yields the same formula for t . The only difference is that this time ξ should be less than 1 in order to make t negative:

$$t = 1 - \frac{1}{\xi} \quad (t \leq 0 \implies \xi \leq 1) \quad (3.35)$$

So we have:

$$\implies \mathbb{P} \left[\sum_{i=1}^n X_i \leq n\xi \right] \leq \exp [-n(\xi - \log \xi - 1)] \quad \text{for } \xi \leq 1 \quad (3.36)$$

4 Chernoff Deviation

We are going to prove the following theorem:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [I \neq \hat{I}] = \min_{0 \leq s \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(x)}{Q(x)} \right)^s \right] \quad (4.1)$$

First we show that LHS \leq RHS:

$$\mathbb{P}[I \neq \hat{I}] = \frac{1}{2}(1 - d_{TV}(P^{\otimes n}, Q^{\otimes n})) \quad (4.2)$$

$$= \frac{1}{2} \sum_{X^n} \min(P(X^n), Q(X^n)) \quad (4.3)$$

$$\leq \sum_{X_1 \dots X_n} P(X_1 \dots X_n)^s Q(X_1 \dots X_n)^{1-s} \quad (4.4)$$

$$= \sum_{X_1 \dots X_n} \prod_{i=1}^n P(X_i)^s Q(X_i)^{1-s} \quad (4.5)$$

$$= \prod_{i=1}^n \sum_{X_i} P(X_i)^s Q(X_i)^{1-s} \quad (4.6)$$

$$= \prod_{i=1}^n \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right] \quad (4.7)$$

$$= \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right]^n \quad (4.8)$$

$$\implies \frac{1}{n} \log \mathbb{P}[I \neq \hat{I}] \leq \min_{0 \leq s \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right] \quad \forall s : 0 \leq s \leq 1 \quad (4.9)$$

Now we need to show that LHS \geq RHS.

We proved Chernoff inequality both in the class and in (3.23):

$$P \left[\sum_i X_i \geq n\delta \right] \leq \exp \left(-n \{ \sup_{\lambda \geq 0} \lambda \delta - \psi_X(\lambda) \} \right) \quad (4.10)$$

Where $\psi_X(\lambda) = \log \mathbb{E}[e^{\lambda X}]$ is the logarithm of the moment generating function of X .

We also proved in the class that for $n \rightarrow \infty$, this bound is tight and we have:

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log P \left[\sum_i X_i \geq n\delta \right] = \inf_{R: \mathbb{E}_R[X] \geq \delta} D_{KL}(R||P) = \sup_{\lambda \geq 0} \lambda \delta - \psi_X(\lambda) \quad (4.11)$$

(we are not going to prove it here again)

We will use this bound to prove what the problem asks for.

$$\mathbb{P}[I \neq \hat{I}] = \frac{1}{2} P(dQ > dP) + \frac{1}{2} Q(dP > dQ) \quad (4.12)$$

$$\geq \sqrt{P(dQ > dP) Q(dP > dQ)} \quad (4.13)$$

$$\implies \frac{1}{n} \log \mathbb{P}[I \neq \hat{I}] \geq \frac{1}{2n} \log P(dQ > dP) + \frac{1}{2n} \log Q(dP > dQ) \quad (4.14)$$

$$(4.15)$$

Using Arithmetic and Geometric Mean Inequality for the last inequality.

Now we drive an expression using Chernoff-bound for each of the terms $P(dQ > dP)$ and $Q(dP >$

dQ):

$$P(dQ > dP) = P \left[\frac{Q(X^n)}{P(X^n)} > 1 \right] \quad (4.16)$$

$$= P \left[\log \frac{Q(X^n)}{P(X^n)} > 0 \right] \quad (4.17)$$

$$= P \left[\sum_i \log \frac{Q(X_i)}{P(X_i)} > 0 \right] \quad (4.18)$$

Using (4.11) and choosing $\tau = 0$, we get:

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log P(dQ > dP) = \sup_{\lambda_1 \geq 0} -\log \mathbb{E}_P \left[e^{\lambda_1 \log \frac{Q(X)}{P(X)}} \right] \quad (4.19)$$

$$\implies \lim_{n \rightarrow \infty} \frac{1}{n} \log P(dQ > dP) = \inf_{\lambda_1 \geq 0} \log \mathbb{E}_P \left[e^{\lambda_1 \log \frac{Q(X)}{P(X)}} \right] \quad (4.20)$$

$$= \inf_{\lambda_1 \geq 0} \log \mathbb{E}_P \left[\left(\frac{Q(X)}{P(X)} \right)^{\lambda_1} \right] \quad (4.21)$$

In a similar way, we can get:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log Q(dP > dQ) = \inf_{\lambda_2 \geq 0} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda_2} \right] \quad (4.22)$$

We can easily show that RHSs of (4.21) and (4.22) are equal:

$$\inf_{\lambda_1 \geq 0} \log \mathbb{E}_P \left[\left(\frac{Q(X)}{P(X)} \right)^{\lambda_1} \right] = \inf_{\lambda_1 \geq 0} \log \mathbb{E}_Q \left[\frac{P(X)}{Q(X)} \left(\frac{Q(X)}{P(X)} \right)^{\lambda_1} \right] \quad (\text{change of measure}) \quad (4.23)$$

$$= \inf_{\lambda_1 \geq 0} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{1-\lambda_1} \right] \quad (4.24)$$

$$= \inf_{\lambda_2 \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda_2} \right] \quad (\text{change of variable}) \quad (4.25)$$

Therefore, we can write:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [I \neq \hat{I}] \geq \lim_{n \rightarrow \infty} \frac{1}{2n} \log P(dQ > dP) + \frac{1}{2n} \log Q(dP > dQ) \quad (4.26)$$

$$= \frac{1}{2} \inf_{\lambda \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda} \right] + \frac{1}{2} \inf_{\lambda \geq 0} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda} \right] \quad (4.27)$$

We can show that both terms on the RHS are equal and minimized with a λ between 0 and 1:

$$\frac{\partial^2}{\partial \lambda^2} \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda} \right] = \frac{\partial}{\partial \lambda} \mathbb{E}_Q \left[\log \frac{P(X)}{Q(X)} \left(\frac{P(X)}{Q(X)} \right)^{\lambda} \right] \quad (4.28)$$

$$= \mathbb{E}_Q \left[\left(\log \frac{P(X)}{Q(X)} \right)^2 \left(\frac{P(X)}{Q(X)} \right)^{\lambda} \right] \geq 0 \quad (4.29)$$

Thus, the function $f(\lambda) = \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right]$ is convex.

Moreover, we have:

$$f(0) = \mathbb{E}_Q [1] = 1 \quad (4.30)$$

$$f(1) = \mathbb{E}_Q \left[\frac{P(X)}{Q(X)} \right] = \sum_X P(X) = 1 \quad (4.31)$$

Therefore f takes values less than or equal to 1 for $0 \leq \lambda \leq 1$ and larger (or equal) values out of this range. So f is minimized in range $0 \leq \lambda \leq 1$ and $\inf_\lambda f(\lambda) = \inf_{0 \leq \lambda \leq 1} f(\lambda) = \min_{0 \leq \lambda \leq 1} f(\lambda)$.

$$\xrightarrow{s=\lambda} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [I \neq \hat{I}] \geq \min_{0 \leq s \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right] \quad (4.32)$$

Using (4.9) and (4.32) we can see:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [I \neq \hat{I}] = \min_{0 \leq s \leq 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right] \quad (4.33)$$

Or equivalently:

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [I \neq \hat{I}] = \max_{0 \leq s \leq 1} - \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^s \right] \quad (4.34)$$

5 Some Properties of Total Variation

5.1

Consider numbers a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n .

We define numbers c_0, c_1, \dots, c_n as:

$$c_i = (a_1 \cdots a_i)(b_{i+1} \cdots b_n) \quad \text{for } 0 \leq i \leq n \quad (5.1)$$

Thus c_0 and c_n are defined as:

$$c_0 = b_1 \cdots b_n \quad (5.2)$$

$$c_n = a_1 \cdots a_n \quad (5.3)$$

We have:

$$c_i - c_{i-1} = (a_i - b_i)(a_1 \cdots a_{i-1})(b_{i+1} \cdots b_n) \quad (5.4)$$

$$\implies a_1 \cdots a_n - b_1 \cdots b_n = c_n - c_0 \quad (5.5)$$

$$= \sum_{i=1}^n (c_i - c_{i-1}) \quad (5.6)$$

$$= \sum_{i=1}^n (a_i - b_i)(a_1 \cdots a_{i-1})(b_{i+1} \cdots b_n) \quad (5.7)$$

$$\implies |a_1 \cdots a_n - b_1 \cdots b_n| \leq \sum_{i=1}^n |a_i - b_i|(a_1 \cdots a_{i-1})(b_{i+1} \cdots b_n) \quad (5.8)$$

Using this inequality, we can prove what we want:

$$d_{TV}(P^{\otimes n}, Q^{\otimes n}) = \frac{1}{2} \sum_{x_1, \dots, x_n} |P(x_1, \dots, x_n) - Q(x_1, \dots, x_n)| \quad (5.9)$$

$$\leq \sum_{i=1}^n \frac{1}{2} \sum_{x_i} |P(x_i) - Q(x_i)| \times \sum_{\hat{x}_i}^n \xi_i \quad (5.10)$$

Where:

$$\hat{x}_i = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad \text{for } 1 \leq i \leq n \quad (5.11)$$

and:

$$\xi_i = P(x_1) \cdots P(x_{i-1}) Q(x_{i+1}) \cdots Q(x_n) \quad (5.12)$$

So $\sum_{\hat{x}_i}^n \xi_i$ is sum over all possible values of product of probability mass functions which is equal to 1. So we have:

$$d_{TV}(P^{\otimes n}, Q^{\otimes n}) \leq \sum_{i=1}^n \frac{1}{2} \sum_{x_i} |P(x_i) - Q(x_i)| \quad (5.13)$$

$$= \sum_{i=0}^n d_{TV}(P, Q) \quad (5.14)$$

5.2

$$Y = g(X) \quad (5.15)$$

$$d_{TV}(P_Y, Q_Y) = \frac{1}{2} \sum_y |P_Y(y) - Q_Y(y)| \quad (5.16)$$

$$= \frac{1}{2} \sum_y |P_X(g^{-1}(y)) - Q_X(g^{-1}(y))| \quad (5.17)$$

$$\stackrel{x=g^{-1}(y)}{=} \frac{1}{2} \sum_x |P_X(x) - Q_X(x)| \quad (5.18)$$

$$= d_{TV}(P_X, Q_X) \quad (5.19)$$

5.3

$$d_{TV}(P_0 \otimes Q, P_1 \otimes Q) = \frac{1}{2} \sum_{x,y} |P_0(x)Q(y) - P_1(x)Q(y)| \quad (5.20)$$

$$= \frac{1}{2} \sum_{x,y} |P_0(x) - P_1(x)|Q(y) \quad (5.21)$$

$$= \frac{1}{2} \sum_x |P_0(x) - P_1(x)| \sum_y Q(y) \quad (5.22)$$

$$= \frac{1}{2} \sum_x |P_0(x) - P_1(x)| \quad (5.23)$$

$$= d_{TV}(P_0, P_1) \quad (5.24)$$

5.4

First we calculate the total variation distance between two gaussian distribution $\mathcal{N}(0, 1)$ and $\mathcal{N}(\mu, 1)$ where μ is positive:

$$1 - d_{TV}(\mathcal{N}(0, 1), \mathcal{N}(\mu, 1)) = \int_{-\infty}^{\infty} \min(\mathcal{N}(0, 1), \mathcal{N}(\mu, 1)) \quad (5.25)$$

$$\implies d_{TV}(\mathcal{N}(0, 1), \mathcal{N}(\mu, 1)) = 1 - \int_{-\infty}^{\infty} \min(\mathcal{N}(0, 1), \mathcal{N}(\mu, 1)) \quad (5.26)$$

$$= 1 - \int_{-\infty}^{\frac{\mu}{2}} \mathcal{N}(\mu, 1) + \int_{\frac{\mu}{2}}^{\infty} \mathcal{N}(0, 1) \quad (5.27)$$

$$= 1 - 2 \int_{\frac{\mu}{2}}^{\infty} \mathcal{N}(0, 1) \quad (5.28)$$

$$= 1 - 2 \int_{\frac{\mu}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (5.29)$$

$$= 1 - 2\Phi\left(\frac{\mu}{2}\right) \quad (5.30)$$

Now we generalize what we proved for two multivariate gaussian distributions $X \sim \mathcal{N}(0, \Sigma)$ and $Y \sim \mathcal{N}(\theta, \Sigma)$.

We know that if X is a multivariate gaussian, then $AX + b$ is a gaussian distribution too (where A is a matrix and b is a vector). We also know that Σ is a positive definite matrix. So we can diagonalize it as:

$$\Sigma = SAS^{-1} \quad (5.31)$$

By defining $W \triangleq \Lambda^{-1/2} S^T$, we find mean and covariance of new gaussian distributions WX and WY :

$$\text{Cov}(X) = \mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X]^T = \Sigma \quad (5.32)$$

$$\text{Cov}(Y) = \mathbb{E}[YY^T] - \mathbb{E}[Y]\mathbb{E}[Y]^T = \Sigma \quad (5.33)$$

$$\mathbb{E}[WX] = W\mathbb{E}[X] = 0 \quad (5.34)$$

$$\mathbb{E}[WY] = W\mathbb{E}[Y] = W\theta \quad (5.35)$$

$$\text{Cov}(WX) = \mathbb{E}[WX(WX)^T] - \mathbb{E}[WX]\mathbb{E}[WX]^T \quad (5.36)$$

$$= W\mathbb{E}[XX^T]W^T - W\mathbb{E}[X]\mathbb{E}[X]^TW^T \quad (5.37)$$

$$= W\Sigma W^T \quad (5.38)$$

$$= \Lambda^{-1/2} S^T S \Lambda S^T \Lambda^{-1/2} \quad (5.39)$$

$$= \Lambda^{-1/2} \Lambda \Lambda^{-1/2} \quad (5.40)$$

$$= I \quad (5.41)$$

$$\xrightarrow{\text{similarly}} \text{Cov}(WY) = I \quad (5.42)$$

Now we want to convert them to gaussian distribution where all elements of their mean vector except the first one is zero. So we define Q as orthonormal matrix achieved by doing QR decomposition on the matrix $[W\theta \mid I]$ (I got this idea from Kasra Khoshjoo):

$$\mathbb{E}[QWX] = QW\mathbb{E}[X] = 0 \quad (5.43)$$

$$\mathbb{E}[QWY] = QW\mathbb{E}[Y] = QW\theta = \|QW\theta\|_{q_1} \quad (5.44)$$

$$\text{Cov}(QWX) = Q\text{Cov}(WX)Q^T = I \quad (5.45)$$

$$\text{Cov}(QWY) = Q\text{Cov}(WY)Q^T = I \quad (5.46)$$

Where q_1 is the first column of Q .

By defining $g(X) = QWX$, we can calculate the total variation distance between X and Y as follow:

$$d_{TV}(X, Y) \stackrel{\text{part 2}}{=} d_{TV}(QWX, QWY) \quad (5.47)$$

$$= d_{TV}(\mathcal{N}(0, I), \mathcal{N}(\|QW\theta\|_{q_1}, I)) \quad (5.48)$$

$$= d_{TV}(\mathcal{N}(0, 1) \otimes \mathcal{N}(0_{n-1}, I_{n-1}), \mathcal{N}(\|QW\theta\|, 1) \otimes \mathcal{N}(0_{n-1}, I_{n-1})) \quad (5.49)$$

$$\stackrel{\text{part 3}}{=} d_{TV}(\mathcal{N}(0, 1), \mathcal{N}(\|QW\theta\|, 1)) \quad (5.50)$$

$$= 1 - 2\Phi\left(\frac{\|QW\theta\|}{2}\right) \quad (5.51)$$

The idea behind this part was introducing a new basis for \mathbb{R}^n which was columns of Q and calculate the total variation distance between two multivariate gaussian distributions in this new basis.

6 Bayesian Approach and Properties of $\mathcal{R}(P, Q)$

6.1

We use Maximum Likelihood approach to find τ .

$$E = \pi_0 P(Z = 1) + \pi_1 Q(Z = 0) \quad (6.1)$$

$$= \pi_0 - \pi_0 P(Z = 0) + \pi_1 Q(Z = 1) \quad (6.2)$$

$$= \sum_x \pi_0 + (-\pi_0 P(x) + \pi_1 Q(x)) Z(0|x) \quad (6.3)$$

Where E is error value.

We want to maximize the above expression with respect to Z . We have:

$$\begin{cases} -\pi_0 P(x) + \pi_1 Q(x) < 0 \equiv \frac{P(x)}{Q(x)} > \frac{\pi_1}{\pi_0} \rightarrow Z(0|x) = 0 \\ -\pi_0 P(x) + \pi_1 Q(x) \geq 0 \equiv \frac{P(x)}{Q(x)} \leq \frac{\pi_1}{\pi_0} \rightarrow Z(0|x) = 1 \end{cases} \quad (6.4)$$

So $\tau = \frac{\pi_1}{\pi_0}$.

6.2

We know that $\beta_\alpha = \alpha^2$. Hence the point (α, α^2) lies on the lower curve of \mathcal{R} . And we know that minimum value of error is achieved for one of points on the lower curve. So we find the point with minimum error value on the lower by setting derivative of error with respect to α to zero.

$$E = \pi_0 P(Z = 1) + \pi_1 Q(Z = 0) \quad (6.5)$$

$$E \text{ on lower curve} = \pi_0(1 - \alpha) + \pi_1 \beta_\alpha \quad (6.6)$$

$$= \pi_0(1 - \alpha) + \pi_1 \alpha^2 \quad (6.7)$$

$$\xrightarrow{\text{setting derivative to zero}} -\pi_0 + 2\pi_1 \alpha = 0 \quad (6.8)$$

$$\alpha = \frac{\pi_0}{2\pi_1} \quad (6.9)$$

6.3

α is a probability value. So it is always between 0 and 1.

$$0 \leq \frac{\pi_0}{2\pi_1} \leq 1 \quad (6.10)$$

$$\pi_0 \leq 2\pi_1 = 2 - 2\pi_0 \quad (6.11)$$

$$\pi_0 \leq \frac{2}{3} \quad (6.12)$$

Note that if $\pi_0 > \frac{2}{3}$, then $\pi_0(1 - \alpha) + \pi_1 \beta_\alpha$ has no critical point and the minimum must be on one of the border points (e.g. $\alpha = 0$ or $\alpha = 1$). In this case, $\alpha = 1$ yields a smaller error value than $\alpha = 0$.

7 Achievable Rates

7.1

We will show that e^{-nE_0}, e^{-nE_1} region (we denote that with \mathcal{R}) is a convex region.

$$(x1, y1), (x2, y2) \in \mathcal{R} \xrightarrow{?} (\lambda x1 + (1 - \lambda)x2, \lambda y1 + (1 - \lambda)y2) \in \mathcal{R} \quad (7.1)$$

If we have test1 which achieves errors $x1, y1$ and test2 which achieves errors $x2, y2$, then we can introduce a new test which is as follows:

We toss a coin with probability λ (a Bernoulli experiment) and if it is head, we use test1 and if it is tail, we use test2. So the type one and type two errors of this new test are $\lambda x1 + (1 - \lambda)x2$ and $\lambda y1 + (1 - \lambda)y2$ respectively. So the point $(\lambda x1 + (1 - \lambda)x2, \lambda y1 + (1 - \lambda)y2)$ is in \mathcal{R} and we have shown that \mathcal{R} is a convex region.

7.2

(Note: I am using $-D(Q||P) \leq \theta \leq D(P||Q)$ which is different from what the problem says. Because this is needed for the proof. Moreover, if $X \sim P$ then $\sum_i \log \frac{P(X_i)}{Q(X_i)} \rightarrow nD(P||Q)$ and if $X \sim Q$ then $\sum_i \log \frac{Q(X_i)}{P(X_i)} \rightarrow -nD(Q||P)$. Thus choosing $\tau = n\theta$ between $-nD(Q||P)$ and $nD(P||Q)$ seems a good choice intuitively!)
Again, using Chernoff bound (see (4.10)) we have:

$$\pi_{1|0} = P(Z = 1) \quad (7.2)$$

$$= P\left(\log \frac{P(X^n)}{Q(X^n)} < n\theta\right) \quad (7.3)$$

$$= P\left(\sum_{i=1}^n \log \frac{P(X_i)}{Q(X_i)} < n\theta\right) \quad (7.4)$$

$$= P\left(\lambda \sum_{i=1}^n \log \frac{P(X_i)}{Q(X_i)} > \lambda n\theta\right) \quad (\lambda < 0) \quad (7.5)$$

$$\leq \exp(-n\{\sup_{\lambda < 0} \lambda\theta - \log \mathbb{E}_P \left[e^{\lambda \log \frac{P(X)}{Q(X)}} \right]\}) \quad (7.6)$$

$$= \exp(-n\{\sup_{\lambda < 0} \lambda\theta - \log \phi_P(\lambda)\}) \quad (7.7)$$

$$= \exp(-n\phi_P^*(\theta)) \quad (7.8)$$

Similarly for $\pi_{0|1}$:

$$\pi_{0|1} = Q(Z = 0) \quad (7.9)$$

$$= Q\left(\log \frac{P(X^n)}{Q(X^n)} > n\theta\right) \quad (7.10)$$

$$= Q\left(\sum_{i=1}^n \log \frac{P(X_i)}{Q(X_i)} > n\theta\right) \quad (7.11)$$

$$= Q\left(\lambda \sum_{i=1}^n \log \frac{P(X_i)}{Q(X_i)} > \lambda n\theta\right) \quad (\lambda > 0) \quad (7.12)$$

$$\leq \exp(-n\{\sup_{\lambda > 0} \lambda\theta - \log \mathbb{E}_Q \left[e^{\lambda \log \frac{P(X)}{Q(X)}} \right]\}) \quad (7.13)$$

$$= \exp(-n\{\sup_{\lambda > 0} \lambda\theta - \log \phi_Q(\lambda)\}) \quad (7.14)$$

$$= \exp(-n\phi_Q^*(\theta)) \quad (7.15)$$

Note that the supremum is taken over $\lambda < 0$ and $\lambda > 0$ respectively, instead of $\lambda \in \mathbb{R}$ which is in the problem statement. We will show that they are equivalent with the problem statements since their supremum occurs at $\lambda < 0$ and $\lambda > 0$ respectively.

Defining $f(\lambda) = \lambda\theta - \log \phi_P(\lambda)$, we have:

$$f(\lambda) = \lambda\theta - \log \mathbb{E}_P \left[e^{\lambda \log \frac{P(X)}{Q(X)}} \right] \quad (7.16)$$

$$= \lambda\theta - \log \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \quad (7.17)$$

We will show that $f(\lambda)$ takes its maximum in the range $[-1, 0]$.

$$f'(\lambda) = \theta - \frac{1}{\mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right]} \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \quad (7.18)$$

$$f''(\lambda) = - \frac{\mathbb{E}_P \left[\left(\log \frac{P(X)}{Q(X)} \right)^2 \left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right] - \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \left(\frac{P(X)}{Q(X)} \right)^\lambda \right]^2}{\mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right]^2} \quad (7.19)$$

Using Cauchy-Schwarz inequality, we have:

$$\mathbb{E}_P \left[\left(\log \frac{P(X)}{Q(X)} \right)^2 \left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \geq \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \left(\frac{P(X)}{Q(X)} \right)^\lambda \right]^2 \quad (7.20)$$

So $f''(\lambda) \leq 0$ for all $\lambda \in \mathbb{R}$ and f is a concave function. Moreover:

$$f'(0) = \theta - \frac{1}{1} \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \right] \quad (7.21)$$

$$= \theta - D(P||Q) \leq 0 \quad (7.22)$$

$$f'(-1) = \theta - \frac{1}{\mathbb{E}_P \left[\frac{Q(X)}{P(X)} \right]} \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)} \frac{Q(X)}{P(X)} \right] \quad (7.23)$$

$$= \theta - \frac{1}{1} \mathbb{E}_Q \left[\log \frac{P(X)}{Q(X)} \right] \quad (7.24)$$

$$= \theta + D(Q||P) \geq 0 \quad (7.25)$$

f is concave, $f'(0) \geq 0$ and $f'(-1) \leq 0$ \rightarrow f takes its maximum in range $[-1, 0]$

So we have: $\sup_{\lambda < 0} f(\lambda) = \sup_{\lambda \in [-1, 0]} f(\lambda) = \sup_{\lambda \in \mathbb{R}} f(\lambda)$. The same argument can be applied to $\phi_Q^*(\theta) = \sup_{\lambda \in \mathbb{R}} \lambda\theta - \log \phi_Q(\lambda)$ to show that it takes its maximum in range $[0, 1]$.

7.3

In the previous section, we showed that the pair of $\phi_P^*(\theta)$ and $\phi_Q^*(\theta)$ rates are achievable. Now we will show that $\phi_Q^*(\theta)$ is equal to $\phi_P^*(\theta) + \theta$.

$$\phi_Q^*(\theta) = \sup_{\lambda \in \mathbb{R}} \lambda\theta - \log \phi_Q(\lambda) \quad (7.26)$$

$$= \sup_{\lambda \in \mathbb{R}} \lambda\theta - \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \quad (7.27)$$

$$= \sup_{\lambda \in \mathbb{R}} \lambda\theta - \log \mathbb{E}_P \left[\frac{Q(X)}{P(X)} \left(\frac{P(X)}{Q(X)} \right)^\lambda \right] \quad (\text{change of measure}) \quad (7.28)$$

$$= \sup_{\lambda \in \mathbb{R}} \lambda\theta - \log \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda-1} \right] \quad (7.29)$$

$$\xrightarrow{\lambda' = \lambda - 1} = \sup_{\lambda' \in \mathbb{R}} (\lambda' + 1)\theta - \log \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda'} \right] \quad (7.30)$$

$$= \theta + \sup_{\lambda' \in \mathbb{R}} \lambda'\theta - \log \mathbb{E}_P \left[\left(\frac{P(X)}{Q(X)} \right)^{\lambda'} \right] \quad (7.31)$$

$$= \theta + \phi_P^*(\theta) \quad (7.32)$$

7.4

7.5