

# Problem Set 2

Information and Coding Theory

February 12, 2021

CLAYTON SEITZ

**Problem 0.1.** Find tight upper and lower bounds on two extremely biased coins where the first coin is distributed according to

$$P = \begin{cases} 0 & \epsilon \\ 1 & 1 - \epsilon \end{cases}$$

and the second is distributed according to

$$Q = \begin{cases} 0 & 2\epsilon \\ 1 & 1 - 2\epsilon \end{cases}$$

**Solution.** I will assume that distinguishing the two coins means that, given a sequence of  $n$  flips, we can say whether it is coin  $P$  or coin  $Q$  in at least  $\frac{9}{10}n$  trials, on average. To start, we will bound the KL-Divergence between the distributions  $P$  and  $Q$

$$\begin{aligned} D(P||Q) &= \epsilon \log \frac{\epsilon}{2\epsilon} + (1 - \epsilon) \log \frac{1 - \epsilon}{1 - 2\epsilon} \\ &= (1 - \epsilon) \log \frac{1 - \epsilon}{1 - 2\epsilon} - \epsilon \\ &= (1 - \epsilon) \log \left( 1 + \frac{\epsilon}{1 - 2\epsilon} \right) - \epsilon \\ &\leq (1 - \epsilon) \left( \frac{\epsilon}{1 - 2\epsilon} \right) - \epsilon \\ &= \frac{\epsilon^2}{1 - 2\epsilon} \end{aligned}$$

So we have found an upper bound for the KL-Divergence i.e. there exists a distribution  $P$  Now, for a sequence  $\bar{x}$  drawn from the product distribution  $Q^n$ , we can bound the probability that  $\bar{x}$  is of a particular type  $P$

$$\frac{1}{(n+1)} \cdot 2^{-nD(P||Q)} \leq \Pr_{\bar{x} \sim Q^n} [T(P_{\bar{x}}) = T(P)] \leq 2^{-nD(P||Q)}$$

where  $T(P)$  denotes the type of an empirical distribution  $P$ . For the lower bound on  $n$ , the arbitrary type  $T(P)$  becomes any  $\bar{x} \sim Q^n$  such that the number of heads in  $\bar{x}$  is at least  $n(1 - \epsilon)$ .

$$\begin{aligned} \Pr_{\bar{x} \sim Q^n} \left[ \sum_i \bar{x}_i \geq n(1 - \epsilon) \right] &\leq \sum_j 2^{-nD(P_j||Q)} \\ &= \sum_j 2^{-nD(P||Q)} \\ &= s \cdot 2^{-nD(P||Q)} \end{aligned}$$

where the empirical distribution  $P_j$  has more than  $n(1 - \epsilon)$  heads and  $s$  is the total number of them. At this point we want to substitute a desirable error rate of  $1/10$  for the LHS and solve for the  $n$  that satisfies this inequality.

$$s \cdot 2^{-nD(P||Q)} \geq \frac{1}{10}$$

taking the logarithm of both sides gives

$$n \geq \log(10 \cdot s) \frac{1 - 2\epsilon}{\epsilon^2}$$

showing that  $n$  is  $O(\frac{1-2\epsilon}{\epsilon^2})$ . Now for the upper bound on  $n$ , we have

$$\left( \frac{s}{n+1} \right) 2^{-nD(P||Q)} \leq \frac{1}{10}$$

by again taking logarithms and converting to base  $e$  is approximately

$$n(\ln 2 - \frac{\epsilon^2}{1 - 2\epsilon}) \leq \frac{\ln 10 \cdot s}{\ln 2}$$

Giving an upper bound on  $n$  such that our classifier performs well

$$n \leq \frac{1}{(\ln 2 - \frac{\epsilon^2}{1 - 2\epsilon})} \cdot \frac{\ln(10 \cdot s)}{\ln 2}$$

■

**Problem 0.2.** Show that  $0 \leq \mathbf{JSD}(P, Q) \leq 1$

**Solution.**

$$\mathbf{JSD}(P, Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$$

The lower bound must be true because  $D(P||M) \geq 0$  and  $D(Q||M) \geq 0$ . These bounds can be seen directly from the fact that the ratio  $\frac{P(x)}{P(x)+Q(x)}$  must lie between 1 and 2. To be precise, consider just one of the terms

$$\begin{aligned} D(P||M) &= \frac{1}{2} \sum_{x \sim P} P(x) \log \frac{P(x)}{M(x)} \\ &= \frac{1}{2} \sum_{x \sim P} P(x) \log \frac{2P(x)}{P(x) + Q(x)} \\ &\leq \frac{1}{2} \sum_{x \sim P} P(x) \log 2 = \frac{1}{2} \end{aligned}$$

Therefore,  $\mathbf{JSD}(P, Q) \leq 1$ .

Show that  $\mathbf{JSD}(P, Q) \geq \frac{1}{8 \ln 2} \cdot ||P - Q||_1^2$

$$\begin{aligned} \mathbf{JSD}(P, Q) &= \frac{1}{2} [D(P||M) + D(Q||M)] \\ &\geq \frac{1}{4 \ln 2} [||P - M||_1^2 + ||Q - M||_1^2] \\ &= \frac{1}{4 \ln 2} \left[ \left( \sum |P - M| \right)^2 + \left( \sum |Q - M| \right)^2 \right] \\ &= \frac{1}{8 \ln 2} \left[ \left( \sum |P - Q| \right)^2 + \left( \sum |Q - P| \right)^2 \right] \\ &= \frac{1}{8 \ln 2} \cdot ||P - Q||_1^2 \end{aligned}$$

Show that  $\mathbf{JSD}(\mathbf{P}, \mathbf{Q}) \leq \frac{1}{2} \cdot ||P - Q||_1$

$$\begin{aligned}
\mathbf{JSD}_\lambda(P, Q) &= \frac{1}{2} [D(P||M) + D(Q||M)] \\
&= \frac{1}{2} \left[ \sum_{x \sim P} P(x) \log \frac{P(x)}{M(x)} + \sum_{x \sim Q} Q(x) \log \frac{Q(x)}{M(x)} \right] \\
&= \frac{1}{2} \sum_{x \sim \chi} \left[ P(x) \log \frac{2P(x)}{P(x) + Q(x)} + Q(x) \log \frac{2Q(x)}{P(x) + Q(x)} \right] \\
&= \frac{1}{2} \sum_{x \sim \chi} [P(x) + Q(x)] \cdot \frac{P(x)}{P(x) + Q(x)} \log \frac{2P(x)}{P(x) + Q(x)} \\
&\quad + \frac{Q(x)}{P(x) + Q(x)} \log \frac{2Q(x)}{P(x) + Q(x)} \\
&= \frac{1}{2} \sum_{x \sim \chi} (P(x) + Q(x)) \cdot \left( 1 - H \left( \frac{P(x)}{P(x) + Q(x)}, \frac{Q(x)}{P(x) + Q(x)} \right) \right) \\
&\leq \frac{1}{2} \sum_{x \sim \chi} |P(x) - Q(x)|
\end{aligned}$$

Since the entropy is strictly non-negative.

$$\mathbf{JSD}_\lambda(P_1 \dots P_k) = \sum_i \lambda_i D(P_i || M)$$

where  $M = \sum_i \lambda_i P_i$ . Show that

$$0 \leq \mathbf{JSD}_\lambda(P_1 \dots P_k) \leq H(\lambda)$$

As before, the lower bound must be true because  $D(P_i || M) \geq 0$  and  $\lambda$  is non-negative. As for the upper bound,

$$\begin{aligned}
\mathbf{JSD}_\lambda(P_1 \dots P_k) &= \sum_i \lambda_i D(P_i || M) \\
&= \sum_i \lambda_i P_i \log \frac{P_i}{M} \\
&= H\left(\sum_i \lambda_i P_i\right) - \sum_i \lambda_i H(P_i) \\
&= H(\lambda) - \sum_i \lambda_i H(P_i) \\
&\leq H(\lambda)
\end{aligned}$$

■

**Problem 0.3.** *Counting using the method of types*

**Solution.** Sanov's Theorem states that

$$Q^n(E) = (n+1)^r 2^{-nD(P^*||Q)}$$

where  $P^*$  is the distribution in  $E$  that is closest to  $Q$ . Since  $Q$  is a uniform distribution we have that

$$D(P^*||Q) = \log m - H(P)$$

Therefore, if we let  $H^*$  be the entropy of the distribution which has maximum entropy, Sanov's theorem becomes

$$Q^n(E) = (n+1)^r 2^{-n(\log m - H^*)}$$

$$\begin{aligned} |S| &= m^n (n+1)^r 2^{-n(\log m - H^*)} \\ &= (n+1)^r 2^{-nH^*} \end{aligned}$$

and therefore, in general,  $|S| \leq (n+1)^r 2^{-nH^*}$ .

■

**Problem 0.4.** *Differential entropy of the multivariate Gaussian with  $n$  variables*

$$\phi(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)$$

**Solution.**

$$\begin{aligned} h(x) &= - \int \phi(x) \log \phi(x) dx \\ &= \int \phi(x) \left[ \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log |\Sigma| + \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] dx \\ &= \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log |\Sigma| + \mathbf{E} \left[ \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \\ &= \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log |\Sigma| \end{aligned}$$

Now, we would like to show that

$$|\alpha\Sigma_1 + (1 - \alpha)\Sigma_2| \leq |\Sigma_1|^\alpha \cdot |\Sigma_2|^{1-\alpha}$$

■

**Problem 0.5.** *Chernoff bound for read- $k$  families*

**Solution.** We can modify Shearer's Lemma for the scenario where we are dealing with subsets of *at most* size  $k$ .

$$k \cdot H(X_1 \dots X_n) \geq \sum_{S \in \mathcal{F}} H(X_S)$$

where  $\mathcal{F}$  is a collection of subsets  $S$  of  $[n]$  which are at most of length  $k$  and  $X_S$  is a particular subset. Consider the following modification

$$k \cdot H(X_1 \dots X_n) - k \cdot H(Y_1 \dots Y_n) \geq \sum_{S \in \mathcal{F}} H(X_S) - k \cdot H(Y_1 \dots Y_n)$$

since  $X$  is uniformly distributed we have that

$$\begin{aligned} k \cdot D(Y_1 \dots Y_n || X_1 \dots X_n) &\geq \sum_{S \in \mathcal{F}} H(X_S) - k \cdot H(Y_1 \dots Y_n) \\ &= \sum_{S \in \mathcal{F}} H(X_S) - k \cdot H(Y_1 \dots Y_n) \\ &= k \sum H(X) - k \cdot H(Y_1 \dots Y_n) \\ &\geq k \sum H(X) - \sum H(Y_S) \\ &= \sum_i D(Y_{S_i} || X_{S_i}) \end{aligned}$$

We can see why this last equality is true by using the fact that  $X$  is uniformly distributed and each  $X$  is independent

$$D(Y_S || X_S) = \sum H(X_S) - H(Y_S) = k \sum H(X) - \sum H(Y_S)$$

Since both are uniformly distributed, we can calculate the ratio of the number of subsets where the condition is satisfied to the total number of possible subsets. This can be written in terms of entropies

$$\begin{aligned}
\mathbb{P}_{X_1 \dots X_n} \left[ \sum f_i(X_S) \geq t \right] &= \frac{|A|}{|X|} \\
&= 2^{H(Y_1 \dots Y_n) - H(X_1 \dots X_n)} \\
&= 2^{-D(Y_1 \dots Y_n \| X_1 \dots X_n)}
\end{aligned}$$

when the distributions are uniform. Also, since the expectation

$$\begin{aligned}
D(Y_{S_i} \| X_{S_i}) &\geq D(f_i(Y_{S_i}) \| f_i(X_{S_i})) \\
&= \mathbb{E}_{f_i(Y_{S_i})} \left[ \log \frac{\mathbf{Pr}[f_i(Y_{S_i})]}{\mathbf{Pr}[f_i(X_{S_i})]} \right] \\
&= \mathbf{Pr}[f_i(Y_{S_i}) = 0] \log \frac{\mathbf{Pr}[f_i(Y_{S_i}) = 0]}{\mathbf{Pr}[f_i(X_{S_i}) = 0]} \\
&\quad + \mathbf{Pr}[f_i(Y_{S_i}) = 1] \log \frac{\mathbf{Pr}[f_i(Y_{S_i}) = 1]}{\mathbf{Pr}[f_i(X_{S_i}) = 1]} \\
&= \nu_i \log \frac{\nu_i}{\mu_i} + (1 - \nu_i) \log \frac{1 - \nu_i}{1 - \mu_i} \\
&= D(\nu_i \| \mu_i)
\end{aligned}$$

Finally, we can combine these results and use the convexity of KL-Divergence to show that

$$\begin{aligned}
\mathbb{P}_{X_1 \dots X_n} \left[ \sum_i f_i(X_{S_i}) \geq t \right] &= 2^{-D(Y_1 \dots Y_n \| X_1 \dots X_n)} \\
&\leq 2^{-\frac{1}{k} \sum_i D(Y_{S_i} \| X_{S_i})} \\
&= 2^{-\frac{1}{k} \sum_i D(\nu_i \| \mu_i)} \\
&\leq 2^{-\frac{\tau}{k} D(\mu + \epsilon \| \mu)}
\end{aligned}$$

■