Problem Set 2
Information Measures

For the Exercise Sessions on Sept 17 and Oct 1 — **Due: Tue, October 7, 10am, on Moodle**

1 Problems for Class

We will solve these problems together in class on Tuesday, Sept 16, and on Tuesday, Sept 23.

Problem 1: Entropy and pairwise independence

Consider three binary random variables X, Y, Z. Each of the three random variables is uniformly distributed, but they are not independent. However, we know that they are *pairwise* independent. That is, X and Y are independent, and X and Z are independent, and Y and Z are independent.

(a) What is $H(X,Y)$?

(b) Give a lower bound to the value of $H(X,Y,Z)$.

(c) Give an example that achieves this bound.

Solution 1. *(a)* Since X, Y, Z are pairwise independent fair flips, $H(X) = H(Y) = H(Z) = 1$, and $H(Y|X) = H(Y)$. Therefore, using the chain rule for entropy, $H(X,Y) = H(X) + H(Y|X) = H(X) + H(Y) = 2$.

(b) Using the chain rule for entropy, we can write: $H(X,Y,Z) = H(X,Y) + H(Z|X,Y) \geq H(X,Y) = 2$, which holds since (conditional) entropy is non-negative, thus $H(Z|X,Y) \geq 0$.

(c) Let $Z = X + Y \mod 2$, then $H(Z|X,Y) = 0$ and $H(X,Y,Z) = H(X,Y)$.

Problem 2: Conditional KL divergence

We saw in class that a *probability kernel* $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is a matrix $P_{Y|X} = P_{Y|X}(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}$ such that $P_{Y|X}(y|x) \geq 0$, and for each $x \in \mathcal{X}, \sum_y P_{Y|X}(y|x) = 1$. Let $P_X \in \Pi(\mathcal{X})$ be a probability distribution on $\mathcal{X}$. We define the *conditional KL divergence* between two probability kernels $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ and $Q_{Y|X} : \mathcal{X} \to \mathcal{Y}$ given P_X to be

$$D(P_{Y|X} \| Q_{Y|X} | P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x) \| Q_{Y|X}(\cdot|x))$$

where for every x, $D(P_{Y|X}(\cdot|x) \| Q_{Y|X}(\cdot|x))$ is the standard KL divergence between the two distributions $P_{Y|X}(\cdot|x)$ and $Q_{Y|X}(\cdot|x)$ over $\mathcal{Y}$.

(a) *(Chain rule of the KL divergence)* Show that

$$D(P_{X,Y} \| Q_{X,Y}) = D(P_X \| Q_X) + D(P_{Y|X} \| Q_{Y|X} | P_X)$$

where $P_{X,Y}$ and $Q_{X,Y}$ are two joint distributions on $\mathcal{X} \times \mathcal{Y}$ such that $P_{X,Y}(x,y) = P_X(x) P_{Y|X}(y|x)$ and $Q_{X,Y}(x,y) = Q_X(x) Q_{Y|X}(y|x)$.

(b) Using (a), show that

$$D(P_{Y|X}\|Q_{Y|X}|P_X) = D(P_{X,Y}\|Q_{X,Y})$$

where $P_{X,Y}(x,y) = P_X(x)P_{Y|X}(y|x)$ and $Q_{X,Y}(x,y) = P_X(x)Q_{Y|X}(y|x)$.

(c) *(Conditioning increases divergence)* Using (b) and the Data Processing Inequality seen in class, show that

$$D(P_Y\|Q_Y) \le D(P_{Y|X}\|Q_{Y|X}|P_X)$$

where $P_Y(y) = \sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}(y|x)$ and $Q_Y(y) = \sum_{x\in\mathcal{X}} P_X(x)Q_{Y|X}(y|x)$.

Solution 2. *(a)*

$$\begin{aligned}
D(P_{XY}\|Q_{XY}) &= \sum_{x,y} P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{Q_{XY}(x,y)} \\
&= \sum_{x,y} P_X(x)P_{Y|X}(y|x)\log\frac{P_X(x)P_{Y|X}(y|x)}{Q_X(x)Q_{Y|X}(y|x)} \\
&= \sum_{x,y} P_X(x)P_{Y|X}(y|x)\log\frac{P_X(x)}{Q_X(x)} + \sum_{x,y} P_X(x)P_{Y|X}(y|x)\log\frac{P_{Y|X}(y|x)}{Q_{Y|X}(y|x)} \\
&= D(P_X\|Q_X) + \sum_x P_X(x)D(P_{Y|X}(\cdot|x)\|Q_{Y|X}(\cdot|x)) = D(P_X\|Q_X) + D(P_{Y|X}\|Q_{Y|X}|P_X).
\end{aligned}$$

(b)

$$D(P_{XY}\|Q_{XY}) = D(P_X\|P_X) + D(P_{Y|X}\|Q_{Y|X}|P_X) = D(P_{Y|X}\|Q_{Y|X}|P_X).$$

(c) Define the kernel

$$W(\tilde{y}|x,y) = \begin{cases} 1, & \text{if } \tilde{y} = y, \\ 0, & \text{otherwise.} \end{cases}$$

Then we have $P_{\tilde{Y}}(\tilde{y}) = \sum_{x,y} P_{XY}(x,y)W(\tilde{y}|x,y) = P_Y(\tilde{y})$ and $Q_{\tilde{Y}}(\tilde{y}) = \sum_{x,y} Q_{XY}(x,y)W(\tilde{y}|x,y) = Q_Y(\tilde{y})$. Hence, we have

$$D(P_{Y|X}\|Q_{Y|X}|P_X) = D(P_{XY}\|Q_{XY}) \ge D(P_{\tilde{Y}}\|Q_{\tilde{Y}}) = D(P_Y\|Q_Y).$$

where the equality follows from part (b) and the inequality follows from DPI.

2 The Homework

The next three problems are the core of this homework. You work out solutions and turn them in. Problem 3 is suitable for Sept 17. Problems 4 and 5 are suitable for Oct 1.

Problem 3: Entropy and Geometry

Suppose X, Y and Z are random variables.

(a) Show that $H(X) + H(Y) + H(Z) \ge \frac{1}{2}\big[H(X,Y) + H(Y,Z) + H(Z,X)\big]$.

(b) Show that $H(X,Y) + H(Y,Z) \ge H(X,Y,Z) + H(Y)$.

(c) Show that

$$2\big[H(X,Y) + H(Y,Z) + H(Z,X)\big] \ge 3H(X,Y,Z) + H(X) + H(Y) + H(Z).$$

(d) Show that $H(X,Y) + H(Y,Z) + H(Z,X) \geq 2H(X,Y,Z)$.

(e) Suppose n points in three dimensions are arranged so that their their projections to the xy, yz and zx planes give n_{xy}, n_{yz} and n_{zx} points. Clearly $n_{xy} \leq n$, $n_{yz} \leq n$, $n_{zx} \leq n$. Use part (d) show that

$$n_{xy}n_{yz}n_{zx} \geq n^2.$$

Solution 3. *(a)* By the sub-addivitity of Entropy we know that

$$\begin{aligned} H(X,Y) &\leq H(X) + H(Y) \\ H(Y,Z) &\leq H(Y) + H(Z) \\ H(X,Z) &\leq H(X) + H(Z). \end{aligned}$$

Adding the three inequalities together we retrieve:

$$H(X) + H(Y) + H(Z) \geq \frac{1}{2}\left(H(X,Y) + H(Y,Z) + H(Z,X)\right).$$

(b) It is easier to show

$$H(X,Y) + H(Y,Z) - (H(X,Y,Z) + H(Y)) \geq 0.$$

Indeed we have that:

$$H(X|Y) - H(X|Y,Z) = I(X;Z|Y) \geq 0.$$

(c) Applying *(b)*, but inverting the roles of X, Y, Z we get:

$$\begin{aligned} H(X,Y) + H(Y,Z) &\geq H(X,Y,Z) + H(Y) \\ H(Y,Z) + H(Z,X) &\geq H(Y,Z,X) + H(Z) \\ H(Y,X) + H(X,Z) &\geq H(Y,X,Z) + H(X). \end{aligned}$$

Adding the three inequalities together gives us *(c)*.
(d) By sub-addivity again, we have that:

$$H(X,Y,Z) \leq H(X) + H(Y) + H(Z). \tag{1}$$

Using (1) in *(c)* we retrieve

$$\begin{aligned} 2[H(X,Y) + H(Y,Z) + H(X,Z)] &\geq 3H(X,Y,Z) + H(X) + H(Y) + H(Z) \\ &\geq 3H(X,Y,Z) + H(X,Y,Z) \\ &= 4H(X,Y,Z). \end{aligned}$$

(e) Let $\{(x_i, y_i, z_i) : i = 1, \ldots, n\}$ be our set of points. Suppose that X, Y, Z are random variables representing the components of the n points with respect to the x, y, z axes. Furthemore, suppose that three random variables are such that $\Pr((X,Y,Z) = (x_i, y_i, z_i)) = 1/n$ for every $1 \leq i \leq n$. This implies that

$$H(X,Y,Z) = \log n. \tag{2}$$

Consequently the random couples $(X,Y), (X,Z), (Y,Z)$ represent the projections of the points respectively, on the xy, xz and yz axes. We can thus say that

$$H(X,Y) \leq \log n_{xy} \tag{3}$$
$$H(X,Z) \leq \log n_{xz} \tag{4}$$
$$H(Y,Z) \leq \log n_{yz}. \tag{5}$$

Using (2),(3),(4),(5) in *(d)* we retrieve the following:

$$\log(n_{xy}n_{xz}n_{yz}) \geq H(X,Y) + H(Y,Z) + H(X,Z) \geq 2H(X,Y,Z) = 2\log n.$$

Which is equivalent to:

$$(n_{xy}n_{xz}n_{yz}) \geq n^2.$$

Problem 4: Variational characterization of mutual information

Let X and Y be two random variables over finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ with joint probability distribution P_{XY}, and let $I(X;Y)$ be their mutual information.

(a) Show that for every function $f(X,Y)$ such that $E_{P_X P_Y}[e^{f(X,Y)}]$ is finite,

$$I(X;Y) \geq \mathbb{E}_{P_{XY}}[f(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{f(X,Y)}]].$$

(b) Show that there is a function $\tilde{f}(X,Y)$ such that $E_{P_X P_Y}[e^{f(X,Y)}]$ is finite and

$$I(X;Y) = \mathbb{E}_{P_{XY}}[\tilde{f}(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{\tilde{f}(X,Y)}]].$$

(c) Conclude that

$$I(X;Y) = \sup_f \mathbb{E}_{P_{XY}}[f(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{f(X,Y)}]]$$

where the sup is over all functions f such that $E_{P_X P_Y}[e^{f(X,Y)}]$ is finite.

Solution 4. *(a)*

$$\begin{aligned}\mathbb{E}_{P_{XY}}[f(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{f(X,Y)}]] &= \mathbb{E}_{P_Y}[\mathbb{E}_{P_{X|Y}}[f(X,Y)] - \log \mathbb{E}_{P_X}[e^{f(X,Y)}]] \\ &\leq \mathbb{E}_{P_Y}[D(P_{X|Y}\|P_X)] = I(X;Y)\end{aligned}$$

where the inequality is due to the Donsker-Varadhan form of the KL divergence seen in class.

(b) Pick $f(x,y) = \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$. For this choice of f, $E_{P_X P_Y}[e^{f(X,Y)}]$ is finite and simple substitution shows that $E_{P_{XY}}[f(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{f(X,Y)}]] = I(X;Y)$.

(c) By (a) we know that $\sup_f \mathbb{E}_{P_{XY}}[f(X,Y)] - \mathbb{E}_{P_Y}[\log \mathbb{E}_{P_X}[e^{f(X,Y)}]]$ is a lower bound on $I(X;Y)$. By (b) we know that the bound can be achieved with $f(x,y) = \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$. This proves that the bound is actually an equality.

Problem 5: Geometrical interpretation of mutual information

We saw in class that a *probability kernel* $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is a matrix $P_{Y|X} = P_{Y|X}(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}$ such that $P_{Y|X}(y|x) \geq 0$, and for each $x \in \mathcal{X}, \sum_y P_{Y|X}(y|x) = 1$. Let $P_X \in \Pi(\mathcal{X})$ be a probability distribution on $\mathcal{X}$. We define the *conditional KL divergence* between two probability kernels $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ and $Q_{Y|X} : \mathcal{X} \to \mathcal{Y}$ given P_X to be

$$D(P_{Y|X}\|Q_{Y|X}|P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x)\|Q_{Y|X}(\cdot|x))$$

where for every $x \in \mathcal{X}$, $D(P_{Y|X}(\cdot|x)\|Q_{Y|X}(\cdot|x))$ is the standard KL divergence between the two distributions $P_{Y|X}(\cdot|x)$ and $Q_{Y|X}(\cdot|x)$ over $\mathcal{Y}$. *Hint:* Recall Problem 2 above (the problem we solved together in class).

(a) Let X and Y be two random variables with joint distribution $P_{XY} = P_X P_{Y|X}$. Show that

$$I(X;Y) = \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x) \| P_Y)$$

where P_Y is the marginal distribution of Y. This formula shows that the mutual information can be interpreted as a weighted average of the distances between the conditional distributions $P_{Y|X}(\cdot|x)$ and the marginal distribution P_Y.

(b) Show that for any distribution Q_Y on $\mathcal{Y}$,

$$I(X;Y) = D(P_{Y|X} \| Q_Y | P_X) - D(P_Y \| Q_Y).$$

You can think of this formula as a KL equivalent of the classical $I(X;Y) = H(Y) - H(Y|X)$.

(c) Show that

$$I(X;Y) = \min_{Q_Y} D(P_{Y|X} \| Q_Y | P_X).$$

According to this formula, the minimizing Q_Y can be interpreted as the "center of gravity" of the conditional distributions $P_{Y|X}(\cdot|x)$, and the mutual information as its radius.

Solution 5. All the results can be proved working directly with the definitions of KL divergence and mutual information. The following is a simple solution that makes use of the results of the Problem that we solved together in class (Problem 2 on this Homework Set).

(a)

$$I(X;Y) = D(P_X P_{Y|X} \| P_X P_Y) = D(P_{Y|X} \| P_Y | P_X) = \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x) \| P_Y),$$

where the second equality is due to Problem 2(b).

(b)

$$\begin{aligned} D(P_Y \| Q_Y) + I(X;Y) &= D(P_Y \| Q_Y) + D(P_{X|Y} \| P_X | P_Y) \\ &= D(P_{XY} \| P_X Q_Y) \\ &= D(P_{Y|X} \| Q_Y | P_X) \end{aligned}$$

where the first equality is due to part (a) by exchanging the roles of X and Y, the second equality is due to the chain rule of the KL divergence (Problem 2(a)), and the third equality is again due to Problem 2(b).

(c) By part (b) we know that $I(X;Y) \leq D(P_{Y|X} \| Q_Y | P_X)$ for every Q_Y, since $D(P_Y \| Q_Y) \geq 0$. Hence, $I(X;Y) \leq \min_{Q_Y} D(P_{Y|X} \| Q_Y | P_X)$. The equality is achieved by picking $Q_Y = P_Y$, for which $D(P_{Y|X} \| Q_Y | P_X) = D(P_{Y|X} \| P_Y | P_X) = I(X;Y)$.

3 Additional Problems

For the next three problems, you do not need to turn in solutions. For the exam, we do expect you know how to solve these problems. So, you may keep these problems around and tackle them in preparation for the Midterm and/or Final exam. Or you may solve them now.

Problem 6: Axiomatic definition of entropy

Let $(p_1, p_2, \ldots, p_m)$ be such that $p_i \geq 0$ for $i = 1, \ldots, m$ and $\sum_i p_i = 1$. Let

$$H_m(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i \tag{6}$$

be the entropy of $(p_1, p_2, \ldots, p_m)$.

(a) *(Grouping property)* Prove that

$$H_m(p_1, p_2, p_3, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Also prove it for grouping p_i and p_j for any arbitrary pair of indices (i, j). This property models the fact that the uncertainty in choosing among m objects should be equal to the uncertainty in first choosing a subgroup of the objects, and then choosing an object in the selected subgroup.

(b) Prove that if a sequence of functions F_m of probability vectors $(p_1, p_2, \ldots, p_m)$, is such that for every $m \geq 2$,

1. $F_m(p_1, p_2, \ldots, p_m)$ is continuous in the p_i's,
2. $F_m(p_1, p_2, \ldots, p_m)$ satisfies the grouping property (a),
3. $F_m(\frac{1}{m}, \ldots, \frac{1}{m}) = \log m$

then F_m must be equal to the entropy (6) (under the usual convention $0 \log 0 = 0$).

Hint: Suppose that the $p_i's$ are rational, i.e., $p_i = \frac{n_i}{n}$ for some positive integers $\{n_i\}_{i=1,\ldots,m}$. Show using (a) recursively that

$$F_n\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = F_m\left(\frac{n_1}{n}, \ldots, \frac{n_m}{n}\right) + \sum_i \frac{n_i}{n} F_{n_i}\left(\frac{1}{n_i}, \ldots, \frac{1}{n_i}\right).$$

Solution 6. *(a)* Using (6), we can rewrite the right-hand side as

$$\begin{aligned}
&H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) \\
&\quad = -(p_1 + p_2) \log(p_1 + p_2) - \sum_{i=3}^{m} p_i \log p_i + (p_1 + p_2)\left(-\frac{p_1}{p_1 + p_2} \log \frac{p_1}{p_1 + p_2} - \frac{p_2}{p_1 + p_2} \log \frac{p_2}{p_1 + p_2}\right) \\
&\quad = -(p_1 + p_2) \log(p_1 + p_2) - \sum_{i=3}^{m} p_i \log p_i - p_1 \log p_1 - p_2 \log p_2 + (p_1 + p_2) \log(p_1 + p_2) \\
&\quad = -\sum_{i=1}^{m} p_i \log p_i = H(p_1, p_2, p_3, \ldots, p_m).
\end{aligned}$$

(b) It can be proved by induction that the grouping property holds for grouping an arbitrary number of elements. Hence, using it recursively on $F\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, we get

$$F\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) = F\left(\frac{m_1}{m}, \ldots, \frac{m_k}{m}\right) + \sum_i \frac{m_i}{m} F\left(\frac{1}{m_i}, \ldots, \frac{1}{m_i}\right).$$

Using property 3 on $F\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$ and on each $F\left(\frac{1}{m_i}, \ldots, \frac{1}{m_i}\right)$, we get

$$\log m = F\left(\frac{m_1}{m}, \ldots, \frac{m_k}{m}\right) + \sum_i \frac{m_i}{m} \log m_i.$$

Rearranging the last equation gives

$$F\left(\frac{m_1}{m}, \ldots, \frac{m_k}{m}\right) = -\sum_i \frac{m_i}{m} \log \frac{m_i}{m}.$$

This proves the result for every rational probability vector. By using the continuity of F (property 1), we can extend the result to any probability vector.

Problem 7: Entropy and combinatorics

Let $n \geq 1$ and fix some $0 \leq k \leq n$. Let $p = \frac{k}{n}$ and let $T_p^n \subset \{0,1\}^n$ be the set of all binary sequences with exactly np ones (assume that np is an integer).

(a) Show that

$$\log |T_p^n| = nh(p) + O(\log_e n)$$

where $h(p) = -p \log_e p - (1-p) \log_e (1-p)$ is the binary entropy function.

Hint: Stirling's approximation states that for every $n \geq 1$,

$$e^{\frac{1}{12n+1}} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq e^{\frac{1}{12n}} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

(b) Let $Q^n = \text{Bernoulli}(q)^n$ be the i.i.d. Bernoulli distribution on $\{0,1\}^n$. Show that

$$\log Q^n[T_p^n] = -nd(p\|q) + O(\log_e n)$$

where $d(p\|q) = p \log_e \frac{p}{q} + (1-p) \log_e \frac{1-p}{1-q}$ is the binary KL divergence.

Solution 7. *(a)* When $p = 0$ or 1, we have $|T_p^n| = 1$, or equivalently $\log |T_p^n| = 0$, so the result holds trivially, since $h(p) = 0$ for $p = 0, 1$. For $p \neq 0, 1$, we have that $|T_p^n| = \binom{n}{np} = \frac{n!}{(np)!(n(1-p))!}$. Using Stirling's approximation on the three factorials we get

$$\frac{1}{\sqrt{2\pi np(1-p)}} p^{-np} (1-p)^{-n(1-p)} e^{\frac{1}{12n+1} - \frac{1}{12np} - \frac{1}{12n(1-p)}} \leq |T_p^n|$$
$$\leq \frac{1}{\sqrt{2\pi np(1-p)}} p^{-np} (1-p)^{-n(1-p)} e^{\frac{1}{12n} - \frac{1}{12np+1} - \frac{1}{12n(1-p)+1}}.$$

By taking the log on each side, we get

$$nh(p) - \frac{1}{2}\log(2\pi np(1-p)) + \frac{1}{12n+1} - \frac{1}{12np} - \frac{1}{12n(1-p)} \leq \log |T_p^n|$$
$$\leq nh(p) - \frac{1}{2}\log(2\pi np(1-p)) + \frac{1}{12n} - \frac{1}{12np+1} - \frac{1}{12n(1-p)+1}.$$

Since $\frac{1}{n} \leq p \leq \frac{n-1}{n}$ and the same holds for $1-p$, we can obtain the following (loose) bounds:

$$-\frac{1}{2}\log n + \frac{1}{2}\log(2\pi) \leq \frac{1}{2}\log(2\pi np(1-p)) \leq \frac{1}{2}\log n + \frac{1}{2}\log(2\pi)$$

$$\frac{1}{12n+1} - \frac{1}{12np} - \frac{1}{12n(1-p)} \geq -2$$

$$\frac{1}{12n} - \frac{1}{12np+1} - \frac{1}{12n(1-p)+1} \leq 1$$

so that we get

$$nh(p) - \frac{1}{2}\log n - \frac{1}{2}\log(2\pi) - 2 \leq \log |T_p^n| \leq nh(p) + \frac{1}{2}\log n - \frac{1}{2}\log(2\pi) + 1$$

i.e., $\log |T_p^n| = nh(p) + O(\log n)$.

(b) We have

$$Q^n[T_p^n] = \binom{n}{np} q^{np} (1-q)^{n(1-p)} = |T_p^n| q^{np} (1-q)^{n(1-p)}$$

and therefore

$$\begin{aligned}\log Q^n[T_p^n] &= \log|T_p^n| + np\log q + n(1-p)\log(1-q)\\ &= nh(p) + np\log q + n(1-p)\log(1-q) + O(\log n)\\ &= -nd(p\|q) + O(\log n)\end{aligned}$$

where in the last step we used (a).

Problem 8: Sum of binomials

Looking at the part (a) previous problem, it can be seen that the entropy function is related to the asymptotic value of the binomial coefficient by:

$$\log\binom{n}{np} = nh(p) + O(\log_e n),$$

for $n \geq 1$ and $0 \leq p \leq 1$, where $h(p) \triangleq -p\log_e p - (1-p)\log_e(1-p)$ is the binary entropy function. We want to derive a similar bound for the sum of binomial coefficients.

(a) Fix $0 \leq p \leq 1/2$ and let $\mathcal{C}$ be the set of all subsets of $\{1, 2, \ldots, n\}$ of size at most np. Let X be a random variable uniformly distributed over $\mathcal{C}$. Show that

$$H(X) \leq nh(p).$$

Hint: Let $(X_1, X_2, \ldots, X_n)$ be a random vector such that for every i, $X_i = 1$ if $i \in X$, and $X_i = 0$ otherwise. Argue that $H(X) = H(X_1, X_2, \ldots, X_n)$.

(b) Using part (a), conclude that

$$\sum_{i=0}^{\lfloor np\rfloor}\binom{n}{i} \leq 2^{nh(p)}.$$

(c) Using part (b), show that if $Z \sim \text{Binomial}(n, p = \frac{1}{2})$, then

$$\Pr\left(\left|Z - \frac{n}{2}\right| \geq c\sigma\right) \leq 2^{1-c^2/2}$$

for every $c \geq 0$, where $\sigma = \frac{\sqrt{n}}{2}$ is the standard deviation of Z. Compare this bound with the Hoeffding inequality for a σ^2-subgaussian random variable we derived in class.

Hint: you can use (without proving it) the bound $h(p) \leq 1 - 2\left(\frac{1}{2} - p\right)^2$.

Solution 8. (a) There is a one-to-one correspondence between X and $(X_1, X_2, \ldots, X_n)$: from the value of X we can uniquely determine the value of $(X_1, X_2, \ldots, X_n)$, and viceversa. Hence, $H(X) = H(X_1, X_2, \ldots, X_n)$. Then,

$$H(X) = H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i) = nH(X_1)$$

where the last equality is due to symmetry.

Now, X takes values from the set C, therefore the expected cardinality of X is less than or equal to np. Since the cardinality of X is equal to the sum of the indicator functions $\sum_{i=1}^n 1(X_i = 1)$, we have $np \geq E(|X|) = E(\sum_{i=1}^n 1(X_i = 1)) = \sum_{i=1}^n E(1(X_i = 1)) = \sum_{i=1}^n P(X_i = 1)$ using the linarity of the expectation and properties of the indicator function.

Then, due to the symmetry, $np\ geq \sum_{i=1}^n P(X_i = 1) = nP(X_1 = 1)$. And, $P(X_1 = 1) \leq p$ follows.

Now, $\Pr(X_1 = 1) \leq p \leq \frac{1}{2}$, and therefore $H(X_1) \leq h(p)$. Hence, $H(X) \leq nh(p)$.

(b)

$$H(X) = \log|\mathcal{C}| = \log \sum_{i=0}^{\lfloor np \rfloor} \binom{n}{i} \leq nh(p).$$

Hence,

$$\sum_{i=0}^{\lfloor np \rfloor} \binom{n}{i} \leq 2^{nh(p)}.$$

(c)

$$\begin{aligned}
\Pr\left(\left|Z - \frac{n}{2}\right| \geq c\frac{\sqrt{n}}{2}\right) &= 2\left(\frac{1}{2}\right)^n \sum_{i=0}^{\left\lfloor n\left(\frac{1}{2} - \frac{c}{2\sqrt{n}}\right)\right\rfloor} \binom{n}{i} \\
&\leq 2^{nh\left(\frac{1}{2} - \frac{c}{2\sqrt{n}}\right) - n + 1} \\
&\leq 2^{n\left(1 - \frac{c^2}{2n}\right) - n + 1} \\
&= 2^{1 - c^2/2}.
\end{aligned}$$

Now, let's consider Hoeffding bound for comparison. Assume we are applying Hoeffding bound to a σ^2-subgaussian random variable Z. We get:

$$\Pr\left(\left|Z - \frac{n}{2}\right| \geq c\sigma\right) \leq 2e^{-\frac{c^2}{2}}$$

Then, the bound that we showed is looser than the the Hoeffding bound.