Homework exercises, Information Theory and Statistics (2016)

1. (*5p.*) Let X be rv taking on values in the finite alphabet $A = \{a_1, a_2, \dots a_n\}$. Prove that
   (a) $0 \leq H(X) \leq \log n$
   (b) $H(X) = 0$ if and only if X is constant with probability 1
   (c) $H(X)$ is the largest if the distribution of X is the discrete uniform distribution over A.

2. (*5p.*) Find out whether there exists a binary prefix code with code lengths
   (a) 2,3,3,3,4,4,4,4,4,5,5,5
   (b) 2,2,3,3,4,4,4,5,6,6

   If yes, then define such a coding!

3. (*4p.*) Determine the type of the binary sequence 0010000110 and find the exact number of the 10-length long sequences of this type. Compare this with the lower and upper estimate learned for $|\mathcal{T}_{\mathbb{P}}^n|$!

4. (*4p.*) Let $x_1^n = (x_1, \dots, x_n)$ be iid sample from an unknown distribution over the finite set A. Prove that the ML estimate of the unknown distribution is the type of x_1^n.

5. Determine the I-divergence $D(\mathbb{P}\|\mathbb{Q})$ if
   (a) (*1p.*) $\mathbb{P}$ and $\mathbb{Q}$ are Bernoulli distributions with parameters p and q, respectively
   (b) (*2p.*) $\mathbb{P}$ and $\mathbb{Q}$ are binomial distributions over the alphabet $\{0, 1, \dots, n\}$ with parameters p and q, respectively
   (c) (*2p.*) $\mathbb{P}$ and $\mathbb{Q}$ are Poisson distributions with parameters λ and μ, respectively
   (d) (*2p.*) $\mathbb{Q}$ is an arbitrary distributions over a finite set A and $\mathbb{P} = \mathbb{Q}(\cdot|B)$, where $B \subset A$ and $\mathbb{Q}(B) > 0$.

6. (*5p.*) Prove that

$$\frac{1}{n+1} e^{nh(\frac{k}{n})} \leq \binom{n}{k} \leq e^{nh(\frac{k}{n})},$$

   where $h(t) = -t \ln t - (1-t)\ln(1-t)$.

   Hint: use the lemma for the cardinality of type-classes.

7. (*5p.*)

   (a) Let $\mathbb{Q}_1, \dots, \mathbb{Q}_n$ be arbitrary distributions over the finite sets $A_1, \dots, A_n$, and $\mathbb{P}$ be an arbitrary distribution over $A_1 \times \cdots \times A_n$ with marginals $\mathbb{P}_1, \dots, \mathbb{P}_n$. Prove that

$$D(\mathbb{P} \| \mathbb{Q}_1 \times \cdots \times \mathbb{Q}_n) = D(\mathbb{P} \| \mathbb{P}_1 \times \cdots \times \mathbb{P}_n) + \sum_{i=1}^{n} D(\mathbb{P}_i \| \mathbb{Q}_i).$$

   Conclude that among the distributions $\mathbb{P}$ with marginals $\mathbb{P}_1, \dots, \mathbb{P}_n$, the I-divergence $D(\mathbb{P} \| \mathbb{Q}_1 \times \cdots \times \mathbb{Q}_n)$ is minimal if $\mathbb{P} = \mathbb{P}_1 \times \cdots \times \mathbb{P}_n$!

   (b) Let $X_1, \dots, X_n$ be iid rv's over the set $\mathcal{X}$, and let $A \subset \mathcal{X}^n$ be an arbitrary measurable set. Prove that

$$\log \mathrm{Prob}((X_1, \dots, X_n) \in A) \le -\sum_{i=1}^{n} D(\mathbb{P}_i \| \mathbb{Q})$$

   where $\mathbb{Q}$ is the common distribution of X_i's and $\mathbb{P}_i$ is the conditional distribution of X_i under the condition $(X_1, \dots, X_n) \in A$.

   Hint: use the result of part (a) with the following choice of $\mathbb{P}$: it is the conditional joint distribution of $X_1, \dots, X_n$ under the condition $(X_1, \dots, X_n) \in A$.

8. (*5p.*) Let $\mathbb{P}, \mathbb{Q}_1, \mathbb{Q}_2$ be arbitrary distributions over the finite set $\mathcal{X}$. Prove that

$$D(\mathbb{P} \| \mathbb{Q}_1) + D(\mathbb{P} \| \mathbb{Q}_2) \ge -2 \log \sum_{x \in \mathcal{X}} \sqrt{Q_1(x) Q_2(x)},$$

   and that to any $\mathbb{Q}_1, \mathbb{Q}_2$ there exists a $\mathbb{P}$ for which equality holds.

   Hint: calculate the left-hand side (by definition), and apply the log-sum inequality for it.

9. (*6p.*) The variational distance of two distributions $\mathbb{P}$ and $\mathbb{Q}$ on $\mathcal{X}$ is

$$|\mathbb{P} - \mathbb{Q}| = \sum_{x \in X} |P(x) - Q(x)|.$$

   (a) Prove the Pinsker inequality:

$$|\mathbb{P} - \mathbb{Q}| \le \sqrt{2D(\mathbb{P} \| \mathbb{Q})}.$$

Hint: $A := \{x : P(x) \geq Q(x)\}$ and define $p := P(A)$, $q := Q(A)$. First verify that $|\mathbb{P} - \mathbb{Q}| = 2(p - q)$ and that

$$D(\mathbb{P}\|\mathbb{Q}) \geq p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

Then show that for any fixed $0 < q < 1$,

$$f(p) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} - 2(p-q)^2$$

is a convex function of p, and takes its minimum at $p = q$.

(b) Prove the tighter estimate

$$D(\mathbb{P}\|\mathbb{Q}) \geq \frac{1}{2 \ln 2} |\mathbb{P} - \mathbb{Q}|^2$$

and that it is tight in the sense that the ratio of $D(\mathbb{P}\|\mathbb{Q})$ and $|\mathbb{P} - \mathbb{Q}|^2$ can be arbitrarily close to $\frac{1}{2 \ln 2}$.

10. *(4p.)* Apply the EM algorithm to separate the mixture of k multivariate (say, p-variate) Gaussian distributions based on an n-element sample, and estimate the parameters of the components at the same time.