3rd homework set, Due: April 21, 10:15

1. (1p.) Consider the following family of candidate distributions on $\mathcal{X} = \{1, \ldots, k\}$: the distributions of form $\mathbb{P}(x) = c \cdot \exp(t_1 x + t_2 x^2)$. Given a sample $\mathbf{x} = (x_1, \ldots, x_n)$, denote by $\mathbb{P}^*$ the maximum likelihood estimate provided that it exists. Assume that each symbol of $\mathcal{X}$ occurs in the sample $\mathbf{x}$. Show that the maximum likelihood estimate exists in this case. Specify linear set $\mathcal{L}$ of distributions on $\mathcal{X}$ such that $\mathbb{P}^*$ is equal to the I-projection of the uniform distribution onto $\mathcal{L}$.

2. (3p.)

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

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

   Conclude that among the distributions $\mathbb{P}$ with marginals $\mathbb{P}_1$, ..., $\mathbb{P}_n$ 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 $X_1, \ldots, X_n$ be iid random variables over the set $\mathcal{X}$, and let $A \subset \mathcal{X}^n$ be an arbitrary measurable set. Prove that

   $$\log Prob\left((X_1, \ldots, X_n) \in A\right) \leqslant -\sum_{i=1}^{n} D(\mathbb{P}_i||\mathbb{Q}) \tag{2}$$

   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 $(X_1, \ldots, X_n) \in A$.

   Hint: use problem 1a of the first homework set and the result of part (a) with the following choice of $\mathbb{P}$: it is the conditional joint distribution of $X_1, \ldots, X_n$ under the condition $(X_1, \ldots, X_n) \in A$.

3. (3p.) For a finite set $\mathcal{X}$, consider on $\mathcal{X} \times \mathcal{X}$ those distributions whose two marginals are equal. Let $\mathcal{L}$ denote the family of these distributions on $\mathcal{X} \times \mathcal{X}$.

   (a) Determine the I-projection to $\mathcal{L}$ of $\widetilde{\mathbb{Q}} = \mathbb{Q}_1 \times \mathbb{Q}_2$.

   Hint: Show first that the I-projection $\widetilde{\mathbb{P}}$ has to be of product form, $\widetilde{\mathbb{P}} = \mathbb{P} \times \mathbb{P}$ (use problem 1a of this homework set), so that $\mathrm{D}(\widetilde{\mathbb{P}}||\mathbb{Q}_1 \times \mathbb{Q}_2) = \mathrm{D}(\mathbb{P}||\mathbb{Q}_1) + \mathrm{D}(\mathbb{P}||\mathbb{Q}_2)$. Writing the right side as one sum, show via the log-sum inequality that its minimum is attained when $\mathbb{P}(x) = c\sqrt{\mathbb{Q}_1(x)\mathbb{Q}_2(x)}$.

   (b) Apply the result of (a) to testing the null hypothesis that an unknown distribution on $\mathcal{X} \times \mathcal{X}$ has equal marginals, when the alternative hypothesis is that this distribution is $\mathbb{Q}_1 \times \mathbb{Q}_2$. Show that for tests with first kind error probability bounded by any $0 < \varepsilon < 1$, the best possible exponent of the second kind error is $-2\log\sum_{x\in\mathcal{X}}\sqrt{\mathbb{Q}_1(x)\mathbb{Q}_2(x)}$. Express this exponent via the Hellinger divergence of $\mathbb{Q}_1$ and $\mathbb{Q}_1$ (f-divergence with $f(t) = 1 - \sqrt{t}$).

4. (3p.) Let Ξ be the log-linear family of distributions on $\Omega = \bigtimes_{i=1}^{d} \{1, \ldots, r_i\}$ with interactions $\gamma \in \Gamma$ where $\Gamma = \{\{1,2\}, \{2,3\}, \ldots, \{d-1, d\}\}$ (taking for Q the uniform distribution on Ω).

   (a) Show that $\mathbb{P} \in \mathcal{P}(\Omega)$ with $S(\mathbb{P}) = \Omega$ belongs to Ξ if and only if it corresponds to a Markov chain, i.e., it equals the joint distribution of random variables X_1,..., X_d such that for each $3 \leqslant j \leqslant d$ the conditional distribution of X_j on the condition $X_1 = x_1$,..., $X_{j-1} = x_{j-1}$ does not depend on x_1, ..., x_{j-2}.

   Hint: Show first the following two statements:

   - If $\mathbb{P} \in \Xi$, i.e., $\mathrm{Prob}(X_1 = x_1, \ldots, X_d = x_d) = \prod_{i=1}^{d-1} B_i(x_i, x_{i+1})$, then for X_1,..., X_d with joint distribution $\mathbb{P}$

     $$\mathrm{Prob}(X_d = x_d | X_1 = x_1, \ldots, X_{d-1} = x_{d-1}) = \frac{\mathrm{Prob}(X_1 = x_1, \ldots, X_d = x_d)}{\sum_{x'_d \in \{1,\ldots,r_d\}} \mathrm{Prob}(X_1 = x_1, \ldots, X_d = x'_d)}$$

     does not depend on x_1, ..., x_{d-2}.

- The $\{1,\ldots,d-1\}$ marginal of P, given by the sum in the denominator above, belongs to the log-linear family of distributions on $\Omega' = \bigtimes_{i=1}^{d-1} \{1,\ldots,r_i\}$ with interactions $\{1,2\}$, ..., $\{d-2,d-1\}$.

(b) Draw the conclusion that among all distributions $\mathbb{P} \in \mathcal{P}(\Omega)$ with prescribed marginals $\mathbb{P}^{1,2}$, $\mathbb{P}^{2,3}$, ..., $\mathbb{P}^{d-1,d}$, that with largest entropy $\mathrm{H}(\mathbb{P})$ is the joint distribution of the Markov chain X_1,..., X_d with X_i, X_{i+1} having joint distribution $\mathbb{P}^{i,i+1}$, for $i = 1,\ldots,d-1$.

Hint: Use Problem 2 of the second homework set