Information Theory and Signal Processing
Gastpar, Telatar, Urbanke

Ecole Polytechnique Fédérale, Lausanne: Fall 2018
October 12, 2018

Problem Set 2 — *Due Friday, October 26, before class starts*

For the Exercise Sessions on Oct 12 and 19

Last name	First name	SCIPER Nr	Points

Problem 1: Information Measures for Continuous Random Variables

Recommended Reading: Chapter 8 of the book by T. M. Cover and J. A. Thomas. "Elements of Information Theory," Second Edition, Wiley, 2006. It is available for free download from the EPFL library.

Find the differential entropy $h(X)$ for the case where X is an exponentially distributed random variable of mean $1/\lambda$.

Problem 2: MMSE Estimation

Consider the scenario where $p(x|d) = de^{-dx}$, for $x \geq 0$ (and zero otherwise), that is, the observed data x is distributed according to an exponential with mean $1/d$. Moreover, the desired variable d itself is also exponentially distributed, with mean $1/\mu$.

(a) Find the MMSE estimator of d given x, and calculate the corresponding mean-squared error incurred by this estimator.

(b) Find the MAP estimator of d given x.

Problem 3: Tweedie's Formula

For the special case where $X = D + N$, where N is Gaussian noise of mean zero and variance σ^2, *Tweedie's formula* says that the conditional mean (that is, the MMSE estimator) can be expressed as

$$\mathbb{E}[D | X = x] = x + \sigma^2 \ell'(x), \tag{1}$$

where

$$\ell'(x) = \frac{d}{dx} \log f_X(x), \tag{2}$$

where $f_X(x)$ denotes the marginal PDF of X. In this exercise, we derive this formula.

(a) Assume that $f_{X|D}(x|d) = e^{\alpha dx - \psi(d)} f_0(x)$ for some functions $\psi(d)$ and $f_0(x)$ and some constant α (such that $f_{X|D}(x|d)$ is a valid PDF for every value of d). Define

$$\lambda(x) = \log \frac{f_X(x)}{f_0(x)}, \tag{3}$$

where $f_X(x)$ is the marginal PDF of X, i.e., $f_X(x) = \int f_{X|D}(x|\delta) f_D(\delta) d\delta$. With this, establish that

$$\mathbb{E}\left[D \mid X = x\right] = \frac{1}{\alpha} \frac{d}{dx} \lambda(x). \tag{4}$$

(b) Show that the case where $X = D + N$, where N is Gaussian noise of mean zero and variance σ^2, is indeed of the form required in Part *(a)* by finding the corresponding $\psi(d), f_0(x)$, and α. Show that in this case, we have

$$\frac{f_0'(x)}{f_0(x)} = -\frac{x}{\sigma^2}, \tag{5}$$

and use this fact in combination with Part *(a)* to establish Tweedie's formula.

Problem 4: FIR Wiener Filter

Consider a (discrete-time) signal that satisfies the difference equation $d[n] = 0.5d[n-1] + v[n]$, where $v[n]$ is a sequence of uncorrelated zero-mean unit-variance random variables. We observe $x[n] = d[n] + w[n]$, where $w[n]$ is a sequence of uncorrelated zero-mean random variables with variance 0.5.

(a) (you may skip this at first and do it later — it is conceptually straightforward) Show that for this signal model, the autocorrelation function of the signal $d[n]$ is

$$\mathbb{E}[d[n]d[n+k]] \quad = \quad \frac{4}{3}\left(\frac{1}{2}\right)^{|k|}, \tag{6}$$

and thus the autocorrelation function of the signal $x[n]$ is

$$\mathbb{E}[x[n]x[n+k]] \quad = \quad \begin{cases} \frac{11}{6}, & \text{for } k = 0, \\ \frac{4}{3}\left(\frac{1}{2}\right)^{|k|}, & \text{otherwise.} \end{cases} \tag{7}$$

(b) We would like to find an (approximate) linear predictor $\hat{d}[n+3]$ using only the observations $x[n], x[n-1], x[n-2], \ldots, x[n-p]$. Using the Wiener Filter framework, determine the optimal coefficients for the linear predictor. Find the corresponding mean-squared error for your predictor.

(c) We would like to find a linear denoiser $\hat{d}[n]$ using *all* of the samples $\{x[k]\}_{k=-\infty}^{\infty}$. Find the filter coefficients and give a formula for the incurred mean-squared error.

Problem 5: Bounding The Exploration Bias

(a) Let $X_1, X_2, \ldots, X_n \sim$ i.i.d. $\mathcal{N}(0,1)$. Let $Y = \operatorname{argmax}_i X_i$ and $T \in \{1, 2, \ldots, n\}$ is such that $P_{T|Y}(t|y) = \begin{cases} p, & t = y \\ \frac{1-p}{n-1}, & t \neq y \end{cases}$ for some $p \in [0,1]$.

1. Compute $I(X;T)$ where $X = (X_1, X_2, \ldots, X_n)$. (Hint: write $I(X;T) = H(T) - H(T|X)$. What is the marginal distribution of T?)
2. Compute the mutual information and J_∞ bounds on the exploration bias $|\mathbb{E}[X_T]|$.
3. For $p = 0.3$, plot the bounds as a function of n.
4. For $n = 10$, plot the bounds as a function of p.

(b) Let $X_1, \ldots, X_4 \sim$ i.i.d. $\mathcal{N}(0,1)$ and $X_5 \sim \mathcal{N}(0,4)$. Let Y and T be as in part (a) with $p = 0.3$.

1. Show that $\mathbf{Pr}(Y=5) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{8\pi}}(1-Q(x))^4 e^{-x^2/8}dx$ (where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$), and find a corresponding numerical approximation (using Mathematica, for example).
2. Using the previous numerical approximation, find the marginal distributions P_Y and P_T.
3. Compute the mutual information and J_∞ bounds on the exploration bias $|\mathbb{E}[X_T]|$. How do the bounds compare with the case $X_5 \sim \mathcal{N}(0,1)$?

Problem 6: Gibbs Algorithm

Let $\mathcal{X}$ be the sample space, $\mathcal{W}$ the hypothesis space, and let $\ell : \mathcal{W} \times \mathcal{X} \to \mathbb{R}_+$ be a corresponding loss function. On a dataset $D = (X_1, X_2, \ldots, X_n)$, the empirical risk for a hypothesis w is given by $L_D(w) = \frac{1}{n}\sum_{i=1}^n \ell(w, X_i)$. We saw in class that $I(D;W)$ can be used to bound the generalization error. Hence, we can use it as a *regularizer* in empirical risk minimization.

(a) First, show that given any joint distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$ and marginal distribution Q on $\mathcal{Y}$, $D(P_{XY}||P_X P_Y) \le D(P_{XY}||P_X Q)$.

Since we cannot directly compute $D(P_{DW}||P_D P_W)$, we will use $D(P_{DW}||P_D Q)$ as a proxy, where Q is a distribution on $\mathcal{W}$.

(b) Let

$$P^\star_{W|D} = \operatorname*{argmin}_{P_{W|D}} \left(\mathbb{E}[L_D(W)] + \frac{1}{\beta} D(P_{DW}||P_D Q) \right).$$

1. Show that

$$\min_{P_{W|D}} \left(\mathbb{E}[L_D(W)] + \frac{1}{\beta} D(P_{DW}||P_D Q) \right) = \mathbb{E}_D \left[\min_{P_{W|D=d}} \left(\mathbb{E}[L_d(W)] + \frac{1}{\beta} D(P_{W|D=d}||Q) \right) \right].$$

2. Show that the minimizer on the right-hand side $P^\star_{W|D=d}$ is given by

$$P^\star_{W|D=d} = \frac{e^{-\beta L_d(w)} Q(w)}{\mathbb{E}_Q\left[e^{-\beta L_d(W)}\right]}.$$

   This is known in the literature as the Gibbs algorithm. (Hint: Write $\mathbb{E}[\beta L_d(W)] = \mathbb{E}[\log e^{\beta L_d(W)}]$, combine with the KL divergence term and use non-negativity of KL divergence.)

3. Show that $P^\star_{W|D=d}$ is $2\beta/n$-differential private if $\ell \in [0,1]$.