ECE 563: Information Theory (Fall 2017)

Maxim Raginsky

Homework 3

Assigned September 21, 2017; due September 28, 2017

Reading:

- Polyanskiy and Wu, Sections 4.1, 4.4–4.7, 5.1, 5.3–5.6

---

1. Consider the *Z-channel* with binary input and output alphabets, $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and with transition probabilities $P_{Y|X}(1|1) = 1$ and $P_{Y|X}(0|0) = P_{Y|X}(1|0) = 1/2$.

   (a) Find the capacity
   $$C = \max_{P_X} I(X;Y).$$

   (b) Let P_Y^* be the unique capacity-achieving output distribution. Explicitly compute the values of $D(P_{Y|X=0}\|P_Y^*)$ and $D(P_{Y|X=1}\|P_Y^*)$. How do they compare to the value of C?

2. Consider the additive-noise channel $X \to X+N$, where $N \sim \text{Uniform}[-1,1]$ is independent of X. Find the capacity
   $$\max_{X:|X|\le A \text{ a.s.}} I(X;X+N),$$
   where $A \in \mathbb{N}$ is a given integer-valued peak amplitude constraint on X, and find the maximizing distribution on X.

   Hint: Use the golden formula with a well-chosen Q_Y.

3. Recall that a channel $P_{Y|X}$ can be thought of as a collection of probability distributions $\{P_{Y|X=x} : x \in \mathcal{X}\}$. We say that a collection of distributions $\{Q_1,\ldots,Q_N\}$ on $\mathcal{Y}$ is an *ε-cover* of $P_{Y|X}$ if
   $$\sup_{x\in\mathcal{X}} \min_{1\le j\le N} D(P_{Y|X=x}\|Q_j) \le \varepsilon$$
   — in other words, if, for each $x \in \mathcal{X}$, we can find some Q_j, such that $D(P_{Y|X=x}\|Q_j) \le \varepsilon$. Let $N(\varepsilon)$ denote the minimal cardinality of an ε-cover.

   (a) Prove that, for every ε, we have the inequality
   $$\max_{P_X} I(X;Y) =: C \le \varepsilon + \log N(\varepsilon).$$

   Hint: Let $N = N(\varepsilon)$, consider the minimal ε-cover $\{Q_1,\ldots,Q_N\}$, and apply the saddle-point characterization to $Q_Y := \frac{1}{N}\sum_{j=1}^N Q_j$.

(b) Prove that, in fact, this bound is tight:

$$C = \inf_{\varepsilon \geq 0} (\varepsilon + \log N(\varepsilon)).$$

Hint: When is $N(\varepsilon) = 1$?

4. Let X and Y be jointly distributed random variables, where X takes values in $\mathcal{X}$ with $M = |\mathcal{X}| < \infty$. Suppose that we observe Y and wish to determine the value of X. However, instead of coming up with a single estimate $\widehat{X}$, we generate a *list* $G(Y) \subseteq \mathcal{X}$. We then say that an error has occurred if the list $G(Y)$ does not contain X. Consider the error probability

$$\varepsilon_G := \mathbf{P}[X \notin G(Y)].$$

Suppose that $|G(Y)| \leq N$ almost surely. Prove the following generalization of Fano's inequality:

$$H(X|Y) \leq (1 - \varepsilon_G) \log N + \varepsilon_G \log M + h(\varepsilon_G),$$

where $h(\cdot)$ is the binary entropy.

Hint: Consider the random variable $E := \mathbf{1}_{\{X \notin G(Y)\}}$. Expand the conditional entropy $H(E, X|Y)$ in two ways using the chain rule.

5. Let $\mathbb{X} = (X_1, X_2, \ldots)$ be a discrete-time random process with discrete alphabet $\mathcal{X}$. The *entropy rate* of $\mathbb{X}$ is defined as

$$H(\mathbb{X}) := \lim_{n \to \infty} \frac{H(X^n)}{n},$$

if the limit exists. Suppose that $\mathbb{X}$ is a *stationary process*, so all of its finite-dimensional marginal distributions are invariant with respect to shifts – that is, for all $n, t_1, \ldots, t_n, k \in \mathbb{N}$, the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+k}, \ldots, X_{t_n+k})$ have the same distribution.

(a) Show that the entropy $H(X^n)$ is *subadditive*, i.e.,

$$H(X^{m+n}) \leq H(X^m) + H(X^n).$$

(b) Use the result from part (a) to prove that

$$H(\mathbb{X}) = \inf_{n \geq 1} \frac{H(X^n)}{n}.$$