Q. Inf. Science 3 (8.S372 / 18.996) — Fall 2022

Assignment 4

Due: **Friday**, *Oct 7, 2020 at* ***5pm*** on gradescape.

1. **Gibbs distributions** In this problem we define entropy with log base-e, i.e. ln. Also let $\exp(x) := e^x$.

   (a) Consider a classical system whose state lies in the set Ω. For simplicity assume that Ω is finite. The energy is defined by function $E : \Omega \to \mathbb{R}$. The Gibbs distribution at temperature T is the probability distribution

   $$g_T(x) := \frac{e^{-E(x)/T}}{\sum_{x'\in\Omega} e^{-E(x')/T}}. \tag{1}$$

   For given E, T, define the free energy of a probability distribution p by

   $$F(p) := \mathop{\mathbb{E}}_{x\sim p}[E(x)] - TH(p) = \sum_{x\in\Omega} p(x)[E(x) + T\ln(p(x))] \tag{2}$$

   Prove that g_T is a local minimum of the free energy. There are a few different ways to do this; probably calculus is the most straightforward.

   (b) Now repeat the above exercise quantumly. Let H be a finite-dimensional Hermitian matrix. Define the Gibbs state

   $$\gamma_T := \frac{e^{-H/T}}{\mathrm{tr}[e^{-H/T}]} \tag{3}$$

   and the free energy

   $$F(\rho) := \mathrm{tr}[H\rho] - TS(\rho). \tag{4}$$

   Prove that γ_T is a local minimum of F. *Hint: One way to solve this problem is to use the formula*

   $$\ln(A+B) = \ln(A) + \int_0^\infty \mathrm{d}z\, \frac{1}{A+zI} B \frac{1}{A+B+zI}. \tag{5}$$

   to evaluate the gradient of F. Another approach uses the fact that if U is a unitary matrix, then $\sum_{i,j} |U_{ij}|^2 |i\rangle\langle j|$ is doubly stochastic, meaning that each row and column is a probability distribution.

   (c) Is F concave, convex or neither? Does this tell us anything about whether g_T and γ_T are global minima of F?

(d) For any state ρ, interpret $F(\rho) - F(\gamma_T)$ as a relative entropy. Use this to derive a robust version of (c), showing that even approximate minimizers of F are close to γ_T. You may use without proof the quantum Pinsker inequality $D(\rho\|\sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1^2$; note that this formulation uses entropies defined with the natural log ($D(\rho\|\sigma) = \operatorname{tr}\rho[\ln(\rho) - \ln(\sigma)]$), and that the usual relative entropy has an extra factor of $\frac{1}{\ln 2}$ on the RHS.

2. **Compression with side information.**

(a) *Conditionally typical set.* For a probability distribution p_{XY} define $J^n_{p,\delta}$ to be the jointly typical set: formally $J^n_{p,\delta} := T^n_{p,\delta} \cap (T^n_{p_X,\delta} \times T^n_{p_Y,\delta})$. Given y^n, define the conditionally typical set $J(y^n) := J^n_{p,\delta}(y^n)$ by

$$J(y^n) = \left\{x^n \in X^n : (x^n, y^n) \in J^n_{p,\delta}\right\}. \tag{6}$$

Observe that if $y^n \notin T^n_{p_Y,\delta}$ then $J(y^n)$ is empty. If $y^n \in T^n_{p_Y,\delta}$ then what bounds can you place on $p^n(x^n|y^n)$ for $x^n \in J(y^n)$? Prove that

$$|J(y^n)| \leq \exp(n(H(X|Y) + 2\delta)). \tag{7}$$

(b) Let $(X^n, Y^n) \sim p^n_{XY}$, i.e. each (X_i, Y_i) is drawn independently from p_{XY}. Suppose that Alice knows X^n and Y^n, Bob holds Y^n and Alice wishes to transmit X^n to Bob. Shannon's noiseless coding theorem tells her how to do this using $\approx nH(X)$ bits, but this would not take advantage of the correlations between X^n and Y^n. Show that she can transmit X^n to Bob using $n(H(X|Y) + \delta)$ bits and error ϵ, with $\epsilon, \delta \to 0$ as $n \to \infty$. (Note: the δ in (a) might not be the same δ as the one here.)

(c) Now suppose that Alice knows only X^n and Bob knows Y^n. This is significantly more challenging than the situation in (b). Suppose that Alice uses a random codebook, as we will also see in Shannon's noisy coding theorem. To compress to rate R, Alice uses a random function $E : X^n \to [2^{nR}] := \{1, 2, \ldots, 2^{nR}\}$, meaning that each $E(x^n)$ is chosen independently and uniformly from $[2^{nR}]$. As in the channel coding theorem, E is chosen randomly and then fixed and can be assumed to be known by both parties.

Given message m, Bob decodes by choosing the unique x^n such that $E(x^n) = m$ and $(x^n, Y^n) \in J$, i.e. in the set $E^{-1}(m) \cap J(Y^n)$. If this x^n either doesn't exist or isn't unique, then he declares failure. Let WRONG be the event where

$$E^{-1}(m) \cap J(Y^n) \tag{8}$$

contains a string x^n that is not equal to the correct string X^n. Prove that $p^n(\text{WRONG}) \to 0$ if $R > H(X|Y) + 3\delta$ as $n \to \infty$.

(d) What other errors are possible? By bounding their probabilities show that the coding strategy in (c) can work with error approaching 0 as $n \to \infty$ for any $R > H(X|Y)$.