1 Rules

2. Write rigorously, prove all claims.
3. You can use notes and textbooks.
4. All exercises are 10 points.

2 Exercises

1. Let $X \in \{0, 1\}$ and let $Y$ be a nonnegative integer-valued random variable with joint distribution
   \[ P_{XY}(i, j) = \alpha 2^{-i-2j} \]
   where $\alpha$ is a normalization constant. Find $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $D(P_{Y|X=0}||P_{Y|X=1})$ and $D(P_{Y|X=1}||P_{Y|X=0})$.

2. Let $X$ be distributed according to the exponential distribution with mean $\mu > 0$, i.e., with density $p(x) = \frac{1}{\mu} e^{-x/\mu} 1_{\{x \geq 0\}}$. Let $\delta$. Compute the divergence $D(P_{X+\delta}||P_X)$.

3. Let $(X, Y)$ be uniformly distributed in the unit $\ell_p$-ball $B_p \triangleq \{(x, y) : |x|^p + |y|^p \leq 1\}$, where $p \in (0, \infty)$. Also define the $\ell_\infty$-ball $B_\infty \triangleq \{(x, y) : |x| \leq 1, |y| \leq 1\}$.
   1. Compute $I(X; Y)$ for $p = 1/2$, $p = 1$ and $p = \infty$.
   2. (Bonus) What do you think $I(X; Y)$ converges to as $p \to 0$. Can you prove it?

4. Let $X$ and $Y$ have finite alphabets. Let $C(P_{Y|X}) = \max_{P_X} I(X; Y)$ be the capacity of $P_{Y|X}$.
   1. Is $P_X \mapsto H(P_X)$ strictly concave?
   2. Fix $P_{Y|X}$. Is $P_X \mapsto I(X; Y)$ strictly concave?
   3. Fix $P_{Y|X}$ with $C(P_{Y|X}) > 0$. Is $P_X \mapsto I(X; Y)$ strictly concave?
   4. Fix $P_X$ with $H(P_X) > 0$. Is $P_{Y|X} \mapsto I(X; Y)$ strictly convex?
   5. Is $P_{XY} \mapsto I(X; Y)$ convex, concave, or neither?
   6. Is $P_{Y|X} \mapsto C(P_{Y|X})$ convex, concave or neither?

5. Let \{\(Y_k, k = 0, \ldots\)\} be a binary stationary Markov process defined as follows: Let $Y_0$ be a binary equiprobable random variable, and
   \[
P_{Y_{k+1}|Y_k}(b|a) = \begin{cases} 
1 - \delta & b = a \\
\delta & b \neq a
\end{cases}
\]
   Find $I(Y_0; Y_n)$. At what speed does $I(Y_0; Y_n)$ vanish with $n$?
6 (Finiteness of entropy) We have shown that any \( N \)-valued random variable \( X \), with \( \mathbb{E}[X] < \infty \) has \( H(X) \leq \mathbb{E}[X] h(1/\mathbb{E}[X]) < \infty \). Next let us improve this result.

1. Show that \( \mathbb{E}[\log X] < \infty \Rightarrow H(X) < \infty \).

Moreover, show that the condition of \( X \) being integer-valued is not superfluous by giving a counterexample.

2. Show that if \( k \mapsto P_X(k) \) is a decreasing sequence, then \( H(X) < \infty \Rightarrow \mathbb{E}[\log X] < \infty \).

Moreover, show that the monotonicity of pmf is not superfluous by giving a counterexample.

7 Consider the hypothesis testing problem:

\[
H_0 : X_1, \ldots, X_n \text{i.i.d.} \sim P = \mathcal{N}(0, 1), \\
H_1 : X_1, \ldots, X_n \text{i.i.d.} \sim Q = \mathcal{N}(\mu, 1).
\]

Questions:

1. Compute the Stein exponent.

2. Compute the tradeoff region \( \mathcal{E} \) of achievable error-exponent pairs \((E_0, E_1)\). Express the optimal boundary in explicit form (eliminate the parameter).

3. Identify the divergence-minimizing geodesic \( P^{(\lambda)} \) running from \( P \) to \( Q \), \( \lambda \in [0, 1] \). Verify that \((E_0, E_1) = (D(P^{(\lambda)} || P), D(P^{(\lambda)} || Q)), 0 \leq \lambda \leq 1 \) gives the same tradeoff curve.

4. Compute the Chernoff exponent.

8 Baby Sanov. Let \( \mathcal{X} \) be a finite set. Let \( \mathcal{E} \) be a convex subset of the simplex of probability distributions on \( \mathcal{X} \). Assume that \( \mathcal{E} \) has non-empty interior. Let \( X^n = (X_1, \ldots, X_n) \) be iid drawn from some distribution \( P \) and let \( \pi_n \) denote the empirical distribution, i.e., \( \pi_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \), which is a function of \( X^n \). Our goal is to show that

\[
E \equiv \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{P(\pi_n \in \mathcal{E})} = \inf_{Q \in \mathcal{E}} D(Q || P).
\]  

(1)

a) Define the following set of joint distributions \( \mathcal{E}_n \equiv \{Q_{X^n} : Q_{X_i} \in \mathcal{E}\} \). Show that

\[
\inf_{Q_{X^n} \in \mathcal{E}_n} D(Q_{X^n} || P_{X^n}) = n \inf_{Q \in \mathcal{E}} D(Q || P),
\]

where \( P_{X^n} = P^n \).

b) Consider the conditional distribution \( \tilde{P}_{X^n} = P_{X^n | \pi_n \in \mathcal{E}} \). Show that \( \tilde{P}_{X^n} \in \mathcal{E}_n \).

c) Show that

\[
P(\pi_n \in \mathcal{E}) \leq \exp \left( - n \inf_{Q \in \mathcal{E}} D(Q || P) \right), \quad \forall n.
\]

d) For any \( Q \) in the interior of \( \mathcal{E} \), show that

\[
P(\pi_n \in \mathcal{E}) \geq \exp(-nD(Q || P) + o(n)), \quad n \to \infty.
\]

(Hint: Use data processing as in the proof of the large deviation theorem.)
Conclude (1).

Comment: Benefit of this proof compared to method of types is that it easily extends to infinite alphabets.

Let \( X_j \sim \exp(1) \) be i.i.d. exponential with mean 1. Since MGF \( \Psi_X(\lambda) \) does not exist for all \( \lambda > 1 \), the result

\[
\mathbb{P}\left[ \sum_{j=1}^{n} X_j \geq n\gamma \right] = \exp\{-n\Psi_X(\gamma) + o(n)\}
\]

proven in class does not apply. Show (2) via the following steps:

1. Apply Chernoff argument directly to prove an upper bound:

\[
\mathbb{P}\left[ \sum_{j=1}^{n} X_j \geq n\gamma \right] \leq \exp\{-n\Psi_X^*(\gamma)\}
\]

2. Fix an arbitrary \( A > 0 \) and prove

\[
\mathbb{P}\left[ \sum_{j=1}^{n} X_j \geq n\gamma \right] \geq \mathbb{P}\left[ \sum_{j=1}^{n} (X_j \land A) \geq n\gamma \right],
\]

where \( u \land v = \min(u, v) \).

3. Apply the results shown in class to investigate the asymptotics of the right-hand side of (4).

4. Conclude the proof of (2) by taking \( A \to \infty \).

(Gibbs distribution) Let \( \mathcal{X} \) be finite alphabet, \( f : \mathcal{X} \to \mathbb{R} \) some function and \( E_{\min} = \min f(x) \).

1. Using \( I \)-projection show that for any \( E \geq E_{\min} \) the solution of

\[
H^*(E) = \max \{ H(X) : \mathbb{E}[f(X)] \leq E \}
\]

is given by \( P_X(x) = \frac{1}{Z_j(\beta)} e^{-\beta f(x)} \) for some \( \beta = \beta(E) \).

Comment: In statistical physics \( x \) is state of the system (e.g., locations and velocities of all molecules), \( f(x) \) is energy of the system in state \( x \), \( P_X \) is the Gibbs distribution and \( \beta = \frac{1}{T} \) is the inverse temperature of the system. In thermodynamic equilibrium, \( P_X(x) \) gives fraction of time system spends in state \( x \).

2. Show that \( \frac{dH^*(E)}{dE} = \beta(E) \).

3. Next consider two functions \( f_0, f_1 \) (i.e., two types of molecules with different state-energy relations). Show that for \( E \geq \min_{x_0} f(x_0) + \min_{x_1} f(x_1) \) we have

\[
\max_{\mathbb{E}[f_0(X_0) + f_1(X_1)] \leq E} H(X_0, X_1) = \max_{E_0 + E_1 \leq E} H_0^*(E_0) + H_1^*(E_1)
\]

where \( H_j^*(E) = \max_{\mathbb{E}[f_j(X)] \leq E} H(X) \).

4. Further, show that for the optimal choice of \( E_0 \) and \( E_1 \) in (5) we have

\[
\beta_0(E_0) = \beta_1(E_1)
\]

or equivalently that the optimal distribution \( P_{X_0, X_1} \) is given by

\[
P_{X_0, X_1}(a, b) = \frac{1}{Z_0(\beta)Z_1(\beta)} e^{-\beta(f_0(a) + f_1(b))}
\]
Remark: (7) also just follows from part 1 by taking \( f(x_0, x_1) = f_0(x_0) + f_1(x_1) \). The point here is relation (6): when two thermodynamical systems are brought in contact with each other, the energy distributes among them in such a way that \( \beta \) parameters (temperatures) equalize.