1. [10 pts] Information-Projection. The I-projection of a distribution $p$ onto a set of distributions $Q$ is given as:

$$q^* = \arg \min_{q \in Q} D(q \| p)$$

Show that if $Q$ is the linear family $Q_{linear} = \{ q \in \mathcal{P} : E_q[r_i(X)] = \alpha_i \text{ for } i = 1, \ldots, m \}$, then $q^*$ belongs to the exponential family with base distribution $p$, i.e.

$$q^*(x) = \frac{p(x) \exp\{\sum_{i=1}^m \lambda^*_i r_i(x)\}}{\sum_x p(x) \exp\{\sum_{i=1}^m \lambda^*_i r_i(x)\}}$$

where $\lambda^*_i$ are chosen so that $q^* \in Q_{linear}$.

Note: In class, we showed that the maximum entropy distribution constrained to the linear family $Q_{linear}$ also belongs to an exponential family with a uniform base distribution, and thus is a special case of the I-projection.

2. [5 pts] Show that the maximum entropy distribution under first moment $\int x f(x) dx = \mu$ and second moment $\int x^2 f(x) dx = \alpha$ is the Gaussian distribution with mean $\mu$ and variance $\alpha - \mu^2$.

3. [10 pts] Prove that the maximum entropy multi-variate distribution under given marginal distributions is the product distribution.

4. [10 pts] Markov chains

(a) Derive the entropy rate of a stationary Markov chain $\{X_i\}$ with stationary distribution $\mu$ and transition matrix $P$. What transition probabilities maximize the entropy rate?

(b) If the transition matrix is given as

$$P = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix}$$

what value of $p$ maximizes the entropy rate? Comment.

5. [5 pts] Mutual information for correlated normals. Find the mutual information $I(X,Y)$ where $(X,Y) \sim \mathcal{N}(0, [\sigma^2 \rho \sigma^2; \rho \sigma^2 \sigma^2])$. Comment on its value for $\rho = 1, 0, -1$. 
6. [10 pts] Logit Regression and Maximum Conditional Entropy. A common approach for binary classification problem is to assume the logistic model, i.e.

\[
P(Y|X) = \frac{\exp(Y(\beta_0 + \sum_{i=1}^{m} \beta_i r_i(X)))}{1 + \exp(\beta_0 + \sum_{i=1}^{m} \beta_i r_i(X))}
\]

where \( Y \) is a binary random variable that indicates the class, and \( r_1(X), \ldots, r_m(X) \) are random variables that depict the features. Show that the logistic model arises as the solution to the following maximum conditional entropy problem:

\[
\begin{align*}
\max & \quad H(Y|X) \\
\text{s.t.} & \quad P(Y|X) > 0 \\
& \quad \sum_y P(Y = y|X) = 1 \\
& \quad \sum_x p(x) r_i(x) \sum_y y P(Y = y|X = x) = \alpha_i \quad \text{for } i = 1, \ldots, m
\end{align*}
\]