Theoretical homework - all work to be done by yourself

1. A coin with $p(\text{head}) = \theta$ is flipped until the first head occurs. Let $X$ denote the number of flips required. Find the entropy $H(X) = -\sum_x p(X = x) \log p(X = x)$.

   The following expressions might be useful:
   $$\sum_{n=1}^\infty r^n = \frac{r}{1-r} \quad \text{and} \quad \sum_{n=1}^\infty nr^n = \frac{r}{(1-r)^2}.$$  

   *Hint:* first compute $p(X = x)$ in terms of $\theta$.

2. Consider an $M$-state discrete random variable $X$, and use Jensen’s inequality to show that the entropy of its distribution $p(X)$ satisfies $H[X] \leq \log M$.

3. Evaluate the Kullback-Leibler divergence
   $$\text{KL}(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} \, dx,$$

   between two Gaussians $p(x) = \mathcal{N}(x|\mu,\sigma^2)$ and $q(x) = \mathcal{N}(x|m,s^2)$ (1-dimensional, extra credit: do it for the multi-dimensional case).

4. By applying Jensen’s inequality with $f(x) = -\log x$, show that the arithmetic mean of a set of real numbers is never less than their geometrical mean.

5. Given $N$ i.i.d samples $\{x_n\}_{n=1}^N$ from a normal distribution $\mathcal{N}(x|\mu,\Sigma)$, the maximum likelihood estimate of the mean $\mu_{\text{ML}}$ and the covariance matrix $\Sigma_{\text{ML}}$ are given by
   $$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n,$$
   $$\Sigma_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{\text{ML}})(x_n - \mu_{\text{ML}})^\top.$$

   Show that
   $$\mathbb{E}[\mu_{\text{ML}}] = \mu, \quad \mathbb{E}[\Sigma_{\text{ML}}] = \frac{N-1}{N} \Sigma.$$  

*Extra credit.* Show that a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the end points of the segment, i.e. for every $x, y \in \mathbb{R}^n$,

   $$\int_0^1 f(x + \lambda(y - x)) \, d\lambda \leq \frac{f(x) + f(y)}{2}.$$  

Ideally submit your solution in Latex.