Information Theory
Sheet 0

1. You are given 12 balls, all indistinguishable by size and appearance. All balls have the same weight except for one that is either lighter or heavier. You are given a balance scale and can put any number of balls on the left and right pan of the scale. There are three possible readings: left and right have same weight, left is lighter than right, right is lighter than left.

   (a) Find a strategy to determine the odd ball with as few readings as possible (Hint: think about the "information gained" in one reading).

   (b) Represent your strategy as a tree. At each node in the tree, record how much information was gained and how much information is missing.

   (c) How much information is gained if we weigh 6 balls against 6 balls in the first reading; how much information is gained, if we weigh 4 balls against 4 balls?

2. Pick your prelims lecture notes on probability and recall the

   (a) definition of a random variable (discrete and continuous),

   (b) the law of large numbers and the central limit theorem,

   (c) the definition of independence and conditional probability.

3. Recall Lagrange multipliers: given $f : U \subset \mathbb{R}^n \to \mathbb{R}$ and $g_1, \ldots, g_k : U \to \mathbb{R}$ we wish to minimize $f(x)$ subject to constraints $g_i(x) = 0$ for $i = 1, \ldots, k$. Assuming sufficient smoothness of $f, g_1, \ldots, g_k$ one can introduce the Lagrangian $\mathcal{L}(x, \lambda_1, \ldots, \lambda_k) := f(x) - \sum_{i=1}^k \lambda_i g_i(x)$ and find a minimizer $x \in U$ by solving $\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0$, $i = 1, \ldots, k$.

   (a) Give an informal argument why this works. One way to think about this (wlog $k = 1$) is to consider $g^{-1}(0)$, $f^{-1}(v)$ for $v \in \mathbb{R}$ as surfaces. Start with v being much bigger than the minimum constrained by $g(x) = 0$ and visualize what happens to the surfaces as v decreases and approaches a constrained minimum. What does this tell us about the relation of the gradient ∇f to ∇g?

   (b) Pick up your favourite analysis textbook and look up the formal proof and the assumptions on $f, g_1, \ldots, g_k$.

   (c) Let $X_1, \ldots X_n$ be independent, real valued variables with $\mathbb{E}[X_i] = \mu$, $Var(X_i) = \sigma_i^2$. Find $c_1, \ldots, c_n$ that minimize $Var\left(\sum_{i=1}^n c_i X_i\right)$ subject to $\mathbb{E}\left[\sum_{i=1}^n c_i X_i\right] = \mu$ for given $\mu \in \mathbb{R}$.

4. Let X be a real-valued random variable.

   (a) Assume additionally that X is non-negative. Show that for every $x > 0$
   $$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X]}{x}.$$
   Find a random variable for which above estimate is sharp.

   (b) Let X be a random variable of mean μ and variance σ^2. Show that $\mathbb{P}(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$.

   (c) Let $(X_n)_{n \geq 1}$ be a sequence of identically distributed, independent random variables with mean μ and variance σ^2. Show that for every $\epsilon > 0$
   $$\lim_{m \to \infty} \mathbb{P}\left(\left|\frac{1}{m} \sum_{n=1}^m X_n - \mu\right| > \epsilon\right) = 0$$

   (d) Let X be uniformly distributed on $\left[0, \frac{\pi}{2}\right]$. Find the density of $Y = \sin X$.