Homework 5

Due: 8:00pm (PDT) Thursday, Nov 10th, 2022

For Problems 1 and 2, review your undergrad probbility for pairs of random variables and/or read Chapter 4 of Prof. Kim's note

1. Problem 4.5 of Prof. Kim's note.
2. Problem 4.11 of Prof. Kim's note.
3. *Fair vs Unfair Coin*:
   (a) Consider a coin with the probability of head $p = 0.3$. Provide an estimate for the probability that after 1000 flipping of this coin we observe between 250 and 350 heads.
   (b) Repeat the above part for a fair coin with the probability of head $p = 0.5$, i.e., provide an estimate for the probability that after 1000 flipping of this coin we observe between 250 and 350 heads.
   (c) If a third party mixes the two coins, how can you distinguish between the two coins?
4. Consider the simple pandemics model $X_{k+1} = W_k X_k$ where $\{W_k\}$ is an i.i.d. sequence that is uniformly distributed over $[0.4, 3]$.
   (a) Find the CDF and PDF of $\log W_k$.
   (b) Find $\mathbb{E}[\log W_k]$. What can you say about $\lim_{k\to\infty} X_k$?
   (c) Let $X_1 = 1$. Generate 100 sample paths $\{X_k\}$ for $1 \leq k \leq 200$ and plot them using the regular linear plot (command:`plot`) and log-plot (command:`semilogy`) and store all the 100 sample paths for a later use.
   (d) Theoretically explain your observation.
   (e) For a level $\alpha \geq 1$, define the random variable $T_\alpha = \min\{k \geq 1 \mid X_k \geq \alpha\}$. Such random variables are called *stopping times*. In this case, it is the first time that the pandemics hits α people. Show that T_α is indeed a random variable.
   (f) Let $\alpha = 10^4$. Familiarize yourself with the command `histogram`. Plot the histogram of T_{10^4} for the 100 sample paths.
   (g) Based on the observed data, provide an estimate for $\mathbb{E}[T_{10^4}]$.
5. For a random vector $\mathbf{X} = (X_1, \ldots, X_n)$, we define its covariance matrix to be the $n \times n$ matrix C with
$$C_{ij} = \mathbb{E}[(X_i - \bar{X}_i)(X_j - \bar{X}_j)],$$
   where $\bar{X}_i = \mathbb{E}[X_i]$. Show that a covariance matrix C is always positive semi-definite. Can we say that it is always positive definite?

   hint: An $n \times n$ symmetric matrix A is called positive semi-definite if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$. If the inequality holds as a strict inequality for all non-zero x, it is called positive definite.
6. In the class, we looked at the LMMSE estimator of X given $Y = X + Z$, where $X, Z \sim \mathcal{N}(0, 1)$ are independent. We found out that in this case the LMMSE estimator is $\hat{X} = \frac{1}{2}Y$. Now, suppose that we have k independent measurements $Y_i = X + Z_i$ for $i = 1, \ldots, k$ of X where $X, Z_1, \ldots, Z_k \sim \mathcal{N}(0, 1)$ are all independent. Find the LMMSE estimator of X given $Y_1, \ldots, Y_k$ using the provided formula
$$\hat{X} = \mathbf{Cov}(X, Y)\mathbf{Cov}^{-1}(Y, Y)(Y - \bar{Y}) + \bar{X}.$$
   You don't need to prove the above identity.

7. Suppose that $\{X_k\}$ is an i.i.d. random process with finite mean $\mathbb{E}[X_k] = \mu$. By the strong law of large numbers we know that

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu \quad \text{a.s.},$$

where $S_n = \sum_{k=1}^{n} X_k$. Show that if we further have a finite variance, i.e., $\mathbf{Var}(X_k) = \sigma^2 < \infty$, then $\lim_{n\to\infty} \frac{S_n}{n} \overset{L_2}{\to} \mu$. In other words,

$$\lim_{n\to\infty} \|\frac{S_n}{n} - \mu\|^2 = \lim_{n\to\infty} \mathbf{Var}(\frac{S_n}{n}) = 0.$$

This is known as the weak law of large numbers.