Probability and Statistics

Solve every problem.

Part I: Probability

Problem 1. Let X be an essentially bounded random variable with mean zero. Show that

$$\mathbb{E}e^X \leq \cosh \|X\|_\infty,$$

where $\cosh x = \frac{e^x + e^{-x}}{2}$ is the hyperbolic cosine function.

Problem 2. Let λ be a positive number. Suppose that X is a random variable with $\mathbb{E}|X| < \infty$. Suppose that
$$\lambda\mathbb{E}f(X+1) = \mathbb{E}\{Xf(X)\}$$

for all bounded smooth functions. Show that X has the Poisson distribution Poisson(λ).

Problem 3. Consider the random walk

$$S_n = a + X_1 + X_2 + \cdots + X_n,$$

where a is a positive integer and $\{X_i\}$ are independent and identically distributed random variables with a common distribution

$$\mathbb{P}\{X_i = 1\} = p, \quad \mathbb{P}\{X_i = -1\} = 1 - p.$$

Let $\tau_0 = \inf\{n : S_n = 0\}$ be the first time the random walk reaches the state $x = 0$. For all $p \in [0,1]$ find the probability $\mathbb{P}_a\{\tau_0 < \infty\}$ that the random walk will eventually hit the state $x = 0$.

Problem 4. Let $Z = (X,Y)$ be an $\mathbb{R}^2$-valued random variable such that (1) X and Y are independent; (2) both X and Y have mean zero and finite (nonvanishing) second moments; (3) the distribution of Z is invariant under the rotation counter-clockwise around the origin by an angle θ not a multiple of 90 degrees. Show that X and Y must be normal random variables with the same variance.

Part II: Statistics

The following collection of questions concerns the design of a randomized experiment where the N units to be randomized to drug A or drug B are people, for whom we have a large number of background covariates, collectively labelled X (e.g., age, sex, blood pressure, height, weight, occupational status, history of heart disease, family history of heart disease). The objective is to assign approximately half to drug A and half to drug B where the means of each of the X variables (and means of non-linear functions of them, such as squares or products) are close to equal in the two groups. Instead of using classical methods of design, such as blocking or stratification, the plan is to use modern computers to try many random allocations and discard those allocations that are considered unacceptable according to a pre-determined criterion for balanced X means, in particular

an affinely invariant measure such as the Mahalanobis distance between the means of X in the two groups. After an acceptable allocation is found, outcome variables will be measured, and their means will be compared in group A and group B to estimate a treatment effect.

Problem 5.   Prove that if the two groups are of the same size (*i.e.*, $N/2$ for even N), this plan will result in unbiased estimates of the A versus B casual effect based on the sample means of Y in groups A and B, where Y is any linear function of X.

Problem 6.   Provide a counter-example to the assertion that **Problem 5** is true in small samples with odd N.