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.

Solution: The equation of the straight line connecting the two points $(-M, e^{-M})$ and (M, e^M) on the curve $y = e^x$ is

$$y = \frac{e^M + e^{-M}}{x} + \left(\frac{e^M - e^{-M}}{2M} \right) x = \cosh M + \left(\frac{\sinh M}{M} \right) x$$

where $\sinh M = \frac{e^M - e^{-M}}{2}$ is the hyperbolic sine function. Since the function $y = e^x$ is a convex function, it lies under the the straight line on the interval $[-M, M]$, hence

$$e^x \leq \cosh M \left(\frac{\sinh M}{M} \right) x. \tag{1}$$

The above inequality can also be proved as follows. We write $x \in [-M, M]$ as a convex combination of M and $-M$ as follows:

$$x = \frac{1}{2}\left(1 + \frac{x}{M}\right) \cdot M + \frac{1}{2}\left(1 - \frac{x}{M}\right) \cdot (-M).$$

From the convexity of the function e^x we have

$$e^x \leq \frac{1}{2}\left(1 - \frac{x}{M}\right) e^M + \frac{1}{2}\left(1 - \frac{x}{M}\right) e^{-M} = \cosh M + \left(\frac{\sinh M}{M} \right) x.$$

Now let $M = \|X\|_\infty$. Because $|X| \leq M$ almost surely from (1) we have

$$e^X \leq \cosh M + \left(\frac{\sinh M}{M} \right) X$$

almost surely. Taking the expectation and using the assumption that $\mathbb{E}X = 0$ we obtain $\mathbb{E}e^X \leq \cosh M$.

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).

Solution: We will apply the identity in the problem to the bounded smooth function $f(x) = e^{itx}$ (for a fixed $t \in \mathbb{R}$). The left-hand side is equal to $\lambda e^{it} \phi(t)$, where $\phi(t) = \mathbb{E}e^{itX}$ is the characteristic function of X. The right-hand side is $\mathbb{E}[Xe^{itX}] = -i\phi'(t)$. Note that under the assumption $\mathbb{E}|X| < \infty$ the function $\phi(t)$ is continuously differentiable and the differentiation under the expectation is justified by the dominated convergence theorem. Therefore $\phi'(t) = i\lambda e^{it} \phi(t)$.

Let $g(t) = \phi(t)e^{-\lambda(e^{it}-1)}$. We have

$$g'(t) = \left\{\phi'(t) - i\lambda e^{it} \phi(t)\right\} e^{-\lambda(e^{it}-1)} = 0.$$

Hence $g(t)$ is a constant, *i.e.*, $g(t) = g(0) = 1$. It follows that the characteristic function of the random variable X is $\phi(t) = e^{\lambda(e^{it}-1)}$. This is the characteristic function of the distribution Poisson(λ). By the uniqueness theorem for characteristic functions the distribution of X must be 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 $x = 0$.

Solution: It is clear that $\mathbb{P}_a\{\tau_0 < \infty\} = 1$ and 0 for $p = 0$ and $p = 1$, respectively. We assume that $0 < p < 1$. Let $q = 1 - p$ for simplicity. Consider the function

$$f(x) = \left(\frac{q}{p}\right)^x.$$

It is easy to verify that $f(S_n)$ is a martingale. For each integer $b > a$ let $\tau_b = \inf\{n : S_n = b\}$ and $\tau = \tau_0 \wedge \tau_b$. Then the stopped martingale $f(S_{n\wedge\tau})$ is bounded. By the martingale convergence theorem, it must converge almost surely. Since $\mathbb{P}_a\{\tau < \infty\} = 1$, the walk $S_{n\wedge\tau}$ stops at either a or b and we have

$$f(a) = \mathbb{P}\{\tau_0 < \tau_b\} + f(b)\mathbb{P}\{\tau_0 > \tau_b\}.$$

Note that $f(0) = 1$. If $0 \le p > 1/2$, then letting $b \to \infty$ we have $\tau_b \to \infty$ and $f(b) \to 0$, hence $\mathbb{P}\{\tau_0 < \infty\} = f(a)$. If $p < 1/2$, then $f(b) \to \infty$ as $b \to \infty$ and we must have $\mathbb{P}\{\tau_0 > \tau_b\} \to 0$. This means that $\mathbb{P}\{\tau_0 < \infty\} = 1$.

For the case $p = 1/2$, we can use the same argument with the function $f(x) = x$. We have $a = b\mathbb{P}\{\tau_0 > \tau_b\}$ or $\mathbb{P}\{\tau_0 > \tau_b\} = a/b$. Letting $b \to \infty$ we have again $\mathbb{P}\{\tau_0 < \infty\} = 1$. We have

$$\mathbb{P}\{\tau_0 < \infty\} = \begin{cases} 1 & 0 \le p \le 1/2, \\ \left(\frac{1-p}{p}\right)^a & 1/2 < p \le 1. \end{cases}$$

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.

Solution: We first show that X and Y have the same variance $\sigma^2 = \mathbb{E}|X|^2 = \mathbb{E}|Y|^2$. By hypothesis, we have

$$\begin{aligned} \mathbb{E}|X|^2 &= \cos^2\theta\,\mathbb{E}|X|^2 + \sin^2\theta\,\mathbb{E}|Y|^2, \\ \mathbb{E}|Y|^2 &= \sin^2\theta\,\mathbb{E}|X|^2 + \cos^2\theta\,\mathbb{E}|Y|^2. \end{aligned} \tag{2}$$

Subtracting the two equations we have

$$(1 - \cos 2\theta)\left(\mathbb{E}|X|^2 - \mathbb{E}|Y|^2\right) = 0.$$

If $\cos 2\theta \ne 1$ we have $\mathbb{E}|X|^2 = \mathbb{E}|Y|^2$; otherwise $\cos^\theta = \sin^2\theta = 1/2$ and the equality obviously holds.

From (2) by induction we can show that

$$X \sim \sum i = 1^{2^n} a_{ni} Z_i$$

where $\{Z_i\}$ are independent and have the same distribution as either X or Y, say $X_i \sim X$ for $i \in I$ and $Z_i \sim Y$ for $i \in J$. Each coefficient has the form $a_{ni} = \pm \cos^k \theta \sin^{n-k} \theta$ for some k. Since θ is not a multiple of 90 degrees, there is a constant $0 < \lambda < 1$ such that $|a_{ni}| \leq \lambda^n$. Furthermore, since X and Y have the same variance we have

$$\sum_{i=1}^{2^n} |a_{ni}|^2 = 1.$$

We show that Lindeberg's condition is satisfied for the sum $\sum_{i=1}^{2^n} a_{ni} Z_i$. Indeed, by Chebyshev's inequality

$$
\begin{aligned}
\sum_{i=1}^{2^n} \mathbb{E}\left\{ |a_{ni} z_i|^2 \; ; |a_{ni} Z_i| \geq \epsilon \right\} \quad & \leq \quad \left(\sum_{i \in I} |a_{ni}|^2 \right) \mathbb{E}\left\{ |X|^2 \; ; |X| \geq \frac{\epsilon}{\lambda^n} \right\} + \left(\sum_{i \in J} |a_{ni}|^2 \right) \mathbb{E}\left\{ |Y|^2 \; ; |Y| \geq \frac{\epsilon}{\lambda^n} \right\} \\
& \leq \quad \frac{\lambda^n}{\epsilon} \left\{ \left(\sum_{i \in I} |a_{ni}|^2 \right) \mathbb{E}|X|^2 + \left(\sum_{i \in J} |a_{ni}|^2 \right) \mathbb{E}|Y|^2 \right\} \\
& = \quad \frac{\lambda^n}{\epsilon} \mathbb{E}|X|^2 \to 0.
\end{aligned}
$$

Now by the Lindeberg central limit theorem, the above sum converges in distribution to a normal distribution, hence X (and also Y by the same argument) are normal random variables.

We can also work with characteristic functions without using the Lindeberg central limit theorem. We have

$$\phi_X(t) = \prod_{i \in I} \phi_X(a_{ni} t) \prod_{i \in J} \phi_Y(a_{ni} t).$$

Since X and Y have mean zero and the same variance σ^2

$$\phi_X(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2) = e^{-\sigma^2 t^2 / 2 + o(t^2)}.$$

The same equality holds for $\phi_Y(t)$. It follows from the fact that a_{ni}'s tend to zero uniformly as $n \to \infty$ and $\sum_{i=1}^{2^n} |a_{ni}|^2 = 1$ that

$$\phi_X(t) = \prod_{i=1}^{2^n} e^{-|a_{ni}|^2 \sigma^2 / 2 + o(|a_{ni}|^2 t^2)} = e^{-\sigma^2 t^2 + o(t^2)} \to e^{-\sigma^2 t^2 / 2}.$$

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.

Solution: Let X_i denote the covariate vector of unit i, and $X = (X_1, \ldots, X_N)^\top$ denote the covariate matrix for all N units. Let z_i denote the treatment allocation for unit i, which equals 1 if the unit is assigned to drug A and 0 otherwise, and $z \in \{0, 1\}^N$ denote the treatment allocation for all N units.

Let $\phi(X, z)$ denote the pre-determined criterion for balanced covariate means, which equals 1 if the allocation z is acceptable and 0 otherwise. By construction, the criterion is invariant when we switch treatment and control groups, *i.e.*, $\phi(X, z) = \phi(X, 1 - z)$. Note that under a completely randomized experiment (CRE) with half of the units assigned to each treatment group, Z and $1 - Z$ follows the same distribution. This implies that, under the randomization (*i.e*, the CRE with balance criterion ϕ),

$$Z \mid \phi(X, Z) = 1 \sim 1 - Z \mid \phi(X, 1 - Z) = 1 \sim 1 - Z \mid \phi(X, Z) = 1.$$

Therefore, under re-randomization, Z and $1 - Z$ must have the same distribution. Consequently, for any $1 \le i \le N$,

$$\mathbb{E}\left(Z_i \mid X, \phi(X, Z) = 1\right) = \mathbb{E}\left(1 - Z_i \mid X, \phi(X, Z) = 1\right) = 1 - \mathbb{E}\left(Z_i \mid X, \phi(X, Z) = 1\right),$$

which immediately implies that $\mathbb{E}\left(Z_i \mid X, \phi(X, Z) = 1\right) = 0.5$.

We consider potential outcomes $Y_i(1) = Y_i(0) = \alpha + \beta^\top X_i$ that are some linear function of the covariate. Obviously, the treatment effect of drug A versus B on outcome Y is zero. Let $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$ be the observed outcome for unit i. The estimator of this causal effect based on the samples means of Y in groups A and B is

$$\hat{\tau} = \frac{1}{N/2} \sum_{i=1}^{N} Z_i Y_i - \frac{1}{N/2} \sum_{i=1}^{N} (1 - Z_i) Y_i = \frac{1}{N/2} \sum_{i=1}^{N} Z_i Y_i(1) - \frac{1}{N/2} \sum_{i=1}^{N} (1 - Z_i) Y_i(0).$$

Under re-randomization, $\hat{\tau}$ has expectation

$$
\begin{aligned}
\mathbb{E}(\hat{\tau} \mid X, \phi(X, Z) = 1) &= \frac{2}{N} \sum_{i=1}^{N} \mathbb{E}(Z_i \mid X, \phi(X, Z) = 1) \cdot Y_i(1) - \frac{2}{N} \sum_{i=1}^{N} \{1 - \mathbb{E}(Z_i \mid X, \phi(X, Z) = 1)\} \cdot Y_i(0) \\
&= \frac{2}{N} \sum_{i=1}^{N} \frac{1}{2} \cdot Y_i(1) - \frac{2}{N} \sum_{i=1}^{N} \left(1 - \frac{1}{2}\right) \cdot Y_i(0) \\
&= \frac{1}{N} \sum_{i=1}^{N} Y_i(1) - \frac{1}{N} \sum_{i=1}^{N} Y_i(0) \\
&= 0.
\end{aligned}
$$

Therefore, $\hat{\tau}$ is an unbiased estimator for the A versus B causal effect on the outcome Y.

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

Solution: We consider an experiment with three units, among which two will be assigned to drug A and the remaining one to drub B. Suppose each unit has a scalar covariate with covariate values $(X_1, X_2, X_3) = (0, 0, 10)$ and the potential outcomes $Y_i(1) = Y_i(0) = X_i$ for $1 \le i \le 3$.

We consider re-randomization with the following balance criterion:

$$\phi(X, z) = 1\left\{ \left| \frac{1}{2} \sum_{i=1}^{N} z_i X_i - \sum_{i=1}^{N} (1 - z_i) X_i \right| \le 8 \right\},$$

i.e., a treatment allocation is acceptable if and only if the absolute value of the difference between covariate means in groups A and B are less than or equal to 8. Consequently, for all $\binom{3}{2} = 3$ treatment allocations, only 2 are acceptable:

$z = (z_1, z_2, z_3)^\top$	$\hat{\tau}_X \equiv \frac{1}{2} \sum_{i=1}^N z_i X_i - \sum_{i=1}^N (1 - z_i) X_i$	$\phi(X, z)$
$(1, 1, 0)$	-10	0; Not acceptable
$(1, 0, 1)$	5	1; Acceptable
$(0, 1, 1)$	5	1; Acceptable

By the definition of potential outcomes, $\hat{\tau}$ is the same as $\hat{\tau}_X$ in the above table. From the above table, under re-randomization with the balance criterion ϕ, $\hat{\tau}$ has probability 1 to be 5. Note that the true casual effect of A versus B on outcome Y is zero. Therefore, in this example, $\hat{\tau}$ is not unbiased for the casual effect.