S.-T. Yau College Student Mathematics Contests 2021

Probability and Statistics

Solve every problem.

Part I: Probability

Problem 1. Suppose that a sequence $\{X_n\}$ of real-valued random variables converges to X in distribution and there are positive constants r and C such that $\mathbb{E}|X_n|^r \le C$ for all n. Show that

$$\lim_{n\to\infty} \mathbb{E}|X_n|^s = \mathbb{E}|X|^s$$

for all $0 < s < r$.

Solution: $X_n \to X$ in distribution if and only if $\lim_{n\to\infty} \mathbb{E}_x f(X_n) = \mathbb{E} f(X)$ for all bounded continuous functions f. Let M be a fixed large positive number and define

$$f_M(x) = \begin{cases} |x|^s, & |x| < M; \\ M^s, & |x| \ge M. \end{cases}$$

Then f_M is a bounded continuous function and $0 \le f_M(x) \uparrow |x|^s$ as $M \uparrow \infty$. We have

$$\lim_{n\to\infty} \mathbb{E} f_M(X_n) = \mathbb{E} f_M(X).$$

On the other hand,

$$0 \le \mathbb{E}|X_n|^s - \mathbb{E} f_M(X_n) \le \mathbb{E}\left[|X_n|^s;\ |X_n| \ge M\right] \le \frac{1}{M^{r-s}} \mathbb{E}\left[|X_n|^r;\ |X_n| \ge M\right] \le \frac{C}{M^{r-s}}.$$

Hence

$$\mathbb{E} f_M(X_n) \le \mathbb{E}|X_n|^s \le \mathbb{E} f_M(X_n) + \frac{C}{M^{r-s}}.$$

Letting $n \to \infty$ and then $M \uparrow \infty$ we have $\lim_{n\to\infty} \mathbb{E}|X_n|^s = \mathbb{E}|X|^s$. In the second limit we use the monotone convergence theorem.

Problem 2. Let $p(x, y)$ be the (one-step) transition function of a Markov chain on a discrete state space S and $p_n(x, y)$ be the n-step transition function. Show that for any positive integers L and N and any two states x and y we have

$$\sum_{n=L}^{N+L} p_n(x, y) \le \sum_{n=0}^{N} p_n(y, y).$$

Solution: The intuition of this problem is that for counting how many steps the chain starting from x to be at y the chain needs to spend time to reach y first.

By the Markov property at L we have for $n \ge L$,

$$p_n(x, y) = \mathbb{P}_x\{X_n = y\} = \mathbb{E}_x p_{n-L}(X_L, y).$$

If the assertion holds for $L = 0$, then ($l = n - L$)

$$\sum_{n=L}^{N+L} p_n(x, y) = \mathbb{E}_x \left[\sum_{l=0}^{N} p_l(X_L, y)\right] \le \sum_{l=0}^{N} p_l(y, y).$$

Thus it is enough to show for the case $L = 0$. By the strong Markov property at T_y, we have

$$p_n(x,y) = \mathbb{P}_x\left\{X_n = y,\ T_y \le n\right\} = \mathbb{E}_x\left[p_{n-T_y}(y,y); T_y \le n\right].$$

It follows that ($l = n - T_y$)

$$\sum_{n=0}^{N} p_n(x,y) = \sum_{n=0}^{N} \mathbb{E}_x\left[p_{n-T_y}(y,y); T_y \le n\right] = \mathbb{E}_x\left[\sum_{l=0}^{N-T_y} p_l(y,y); T_y \le N\right] \le \sum_{l=0}^{N} p_n(y,y).$$

Problem 3. Let $\{X_n\}$ be an independent, identically distributed sequence of random variables with the symmetric Bernoulli distribution

$$\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = \tfrac{1}{2}.$$

Let $S_n = \sum_{i=1}^{N} X_i$ be the partial sum. Show that for all $\alpha > \frac{1}{2}$,

$$\mathbb{P}\left\{\lim_{n\to\infty} \frac{S_n}{n^\alpha} = 0\right\} = 1.$$

Solution: Let k be a positive integer. By Markov's inequality we have

$$\mathbb{P}\left\{\left|\frac{S_n}{n^\alpha}\right|^{2k} \ge \frac{1}{n}\right\} \le n^{-(2k\alpha-1)}\mathbb{E}|S_n|^{2k}.$$

We have

$$\mathbb{E}|S_n|^{2k} = \sum_{1\le i_1,\dots,i_{2k}\le n} \mathbb{E}\left[X_{i_1}X_{i_2}\cdots X_{i_{2k}}\right].$$

Since the sequences are independent and each random variable has mean zero, each expectation is zero unless the indices $i_1, i_2, \dots, i_{2k}$ can be grouped into k pairs of identical indices. The number of such groups is at most $(2k)! \cdot n^k$. Hence

$$\mathbb{P}\left\{\left|\frac{S_n}{n^\alpha}\right|^{2k} \ge \frac{1}{n}\right\} \le C_k n^{-(2k\alpha-k-1)}.$$

Choosing $k > 2/(2\alpha - 1)$ we have $2k\alpha - k - 1 > 1$. By the Borel-Cantelli lemma, the probability that $|S_n|/n^\alpha \ge n^{-1}$ happens infinitely often is zero, hence $\lim_{n\to\infty} S_n/n^\alpha = 0$ with probability one.

Problem 4. Let $X^n = \left\{X_{ij}\right\}$ be an $n \times n$ random matrix whose entries are independent and identically distributed random variables with the symmetric Bernoulli distribution

$$\mathbb{P}\{X = 0\} = \mathbb{P}\{X = 1\} = \tfrac{1}{2}.$$

Let $p_n = \mathbb{P}\{\det X_N \text{ is odd}\}$. Show that $\lim_{n\to\infty} p_n > 0$.

Solution: Consider the (algebraic) field $\mathbf{Z}/2\mathbf{Z} = \{0,1\}$ with two elements, so we need $\det X_n \ne 0$ in $\mathbf{Z}/2\mathbf{Z}$, which means that p_n is the probability that the n rows of the random matrix are independent.

Each row has 2^n choices. There are $2^n - 1$ choices for the first row X_1 (the zero column is not allowed). The linear span for the first row has two vectors (the zero vector and X_1), and we cannot choose from this linear span, hence the number of choices for X_2 is $2^n - 2$. In general after the first l rows are fixed, the linear span of these rows has 2^l vectors,

and we cannot choose X_{l+1} from this linear span, hence the number of choices for X_{l+1} is $2^n - 2^l$. It follows that the total number of choices for the matrix $\det X^n$ with nonvanishing determinant is

$$(2^n - 1) \cdot (2^n - 2) \cdot (2^n - 2^2) \cdots (2^n - 2^{n-1}).$$

The total number of choices for the random matrix X^n is $(2^n)^n = 2^{n^2}$ and the probability distribution is uniform, hence

$$p_n = \prod_{l=1}^{n} \left(1 - \frac{1}{2^l}\right) \to \prod_{l=1}^{\infty} \left(1 - \frac{1}{2^l}\right) > 0.$$

Part II: Statistics

Problem 5. You have been asked to help design a randomized trial of a new drug, call it drug A, to be used in place of the current drug, call it drug B, for a particular medical condition. The budget is fixed to have 1000 patients treated with A and 1000 treated with drug B. The issue is how to do the allocation of patients, because we have many pre-randomization measurements on each patient, roughly 200, such as blood pressure recordings, age, sex, and a large collection of genetics measurements. Obviously it is desirable to have the A group similar to the B group with respect to all pre-treatment covariates and non-linear functions of them that are expected to influence the effectiveness of the drugs with respect to the outcome variables.

Complete (or simple) randomization does this in expectation, but with many covariates, some covariates will not be balanced between the A and B groups in any one single randomized allocation. Standard blocking used in traditional experimental design can force balance on a few covariates, but the designer of drug A wants to have an experimental design that creates balance on many covariates, and feels that you, as a modern applied mathematician/statistician, should be able to do this.

Describe a class of methods that achieves this goal where each patient has a positive probability of receiving drug A and a positive probability of receiving drug B. Provide enough detail that you are describing an explicit algorithm.

Solution: Let X_i denote the covariate vector of unit i, which can include the pre-treatment measurements as well as their non-linear functions, and let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ denote the convariate matrix for all $n = 2000$ 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 $\boldsymbol{z} \in \{0,1\}^n$ denote the treatment allocation for all n units.

Let $\phi(\boldsymbol{X}, \boldsymbol{z})$ denote a pre-determined covariate balance criterion, which equals 1 if the allocation $\boldsymbol{z}$ is acceptable and 0 otherwise. Moreover, we construct the criterion ϕ such that it is invariant when we switch treatment and control groups, *i.e.*, $\phi(\boldsymbol{X}, \boldsymbol{z}) = \phi(\boldsymbol{X}, 1 - \boldsymbol{z})$. We then consider the following re-randomization design to randomly allocate the patients to groups A and B:

- **(i)** Completely randomize the patients to groups A and B, with 1000 patients within each group.
- **(ii)** If the balance criterion is satisfied (*i.e.*, $\phi(\boldsymbol{X}, \boldsymbol{Z}) = 1$), proceed to step **(iii)**; otherwise, return to step **(i)**
- **(iii)** Conduct the experiment using the final randomization obtained in step **(ii)**.

Note that under the completely randomized experiment (CRE) with the same number of units assigned to each treatment group, $\boldsymbol{Z}$ and $1 - \boldsymbol{Z}$ follows the same distribution. This implies that, under re-randomization described above (*i.e.*, the CRE with balance criterion ϕ,

$$\boldsymbol{Z} \mid \phi(\boldsymbol{X}, \boldsymbol{Z}) = 1 \sim \mathbf{1} - \boldsymbol{Z} \mid \phi(\boldsymbol{X}, 1 - \boldsymbol{Z}) = 1 \sim \mathbf{1} - \boldsymbol{Z} \mid \phi(\boldsymbol{X}, \boldsymbol{Z}) = 1.$$

Thus, under re-randomization, $\boldsymbol{Z}$ and $\mathbf{1} - \boldsymbol{Z}$ must have the same distribution. Consequently, for any $1 \le i \le N$,

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

which immediately implies that $\mathbb{E}(Z_i \mid \boldsymbol{X}, \phi(\boldsymbol{X}, \boldsymbol{Z}) = 1) = 0.5$. Therefore, under our proposed design, each patient has a positive probability (*i.e.*, $\frac{1}{2}$) to receive either drug A or drug B.

Problem 6. You are given the results of a randomized experiment of two drugs, A and B. The experiment was not conducted in the usual way, however, but rather by allocating patients by a machine-learning algorithm under which each patient has a positive probability of receiving A and of receiving B; moreover the algorithm is completely specified and is built to create better than random balance on the covariates.

(a) Can unbiased estimates of the causal effect of drug A versus B be found, and if so, show why.

(b) Can exact small sample, non-parametric inferences for the causal effect in part **(a)** be derived, based solely on the randomization distribution of some statistic? For example, can we find exact significance levels under a sharp null hypothesis? If so, outline how to accomplish this goal.

Solution:

(a) Yes. For each patient i, let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes under drug A and drug B, Z_i be the treatment assignment indicator, where $Z_i = 1$ if the patient receives drug A and 0 otherwise, and $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$ be the observed outcome.

For each patient i, let $e_i = \Pr(Z_i = 1)$ denote the probability of receiving drug A. Then $1 - e_i$ is the probability of receiving drug B for patient i. Because the algorithm for allocating the patients are completely specified and each patient has a positive probability of receiving A and of receiving B, e_i is known for all i and $e_i \in (0, 1)$. We can construct the following unbiased estimator for the average causal effect of drug A versus B for patients in the experiment:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{e_i}Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{1 - Z_i}{1 - e_i}Y_i.$$

Below we show that $\hat{\tau}$ is unbiased for the average causal effect $\tau \equiv n^{-1}\sum_{i=1}^{n}(Y_i(1) - Y_i(0))$. By definition we have

$$\begin{aligned}
\mathbb{E}\hat{\tau} =& \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{e_i}Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{1 - Z_i}{1 - e_i}Y_i\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i}{e_i}Y_i(1) - \frac{1}{n}\sum_{i=1}^{n}\frac{1 - Z_i}{1 - e_i}Y_i(0)\right) \\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}(Z_i)}{e_i}Y_i(1) - \frac{1}{n}\sum_{i=1}^{n}\frac{1 - \mathbb{E}(Z_i)}{1 - e_i}Y_i(0) = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i}{e_i}Y_i(1) - \frac{1}{n}\sum_{i=1}^{n}\frac{1 - e_i}{1 - e_i}Y_i(0) \\
=& \frac{1}{n}\sum_{i=1}^{n}Y_i(1) - \frac{1}{n}\sum_{i=1}^{n}Y_i(0) = \tau.
\end{aligned}$$

Therefore, $\hat{\tau}$ is unbiased for τ.

(b) Yes. For example we consider the sharp null hypothesis of no treatment effect:

$$H_0 : Y_i(1) = Y_i(0), \quad i = 1, 2, \ldots, n.$$

Let $t(\boldsymbol{Z}, \boldsymbol{Y})$ denote a general test statistic, based on which we test the null hypothesis H_0.

First, we can impute all the unknown potential outcomes based on the observed data and the null hypothesis H_0. Specifically, under H_0, $Y_i(1) = Y_i(0) = Y_i$, *i.e.*, both potential outcomes are the same as the observed one.

Second, we can know the distribution of the test statistic under the null hypothesis H_0. Because the algorithm for allocating the patients are completely specified, we know the distribution of the treatment assignment (*i.e.*, treatment assignment mechanism). Let $\boldsymbol{W}$ denote a random vector following this treatment assignment mechanism. Under H_0, the test statistic follows the same distribution as

$$t\left(\boldsymbol{W}, \boldsymbol{W} \circ \boldsymbol{Y}(1) + (\boldsymbol{1} - \boldsymbol{W}) \circ \boldsymbol{Y}(0)\right),$$

where $\boldsymbol{Y}(1)$ and $\boldsymbol{Y}(0)$ denote the potential outcome vectors for all units, and $\circ$ denotes element-wise multiplication.

Third, we can then evaluate the tail probability of the null distribution of the test statistic at its observed value to get the p-value. Specifically, let

$$G(c) = \Pr\left\{t\left(\boldsymbol{W}, \boldsymbol{W} \circ \boldsymbol{Y}(1) + (\boldsymbol{1} - \boldsymbol{W}) \circ \boldsymbol{Y}(0)\right) \geq c\right\}$$

denote the tail probability of the null distribution. Then the p-value is

$$p = G(t(\boldsymbol{Z}, \boldsymbol{Y})),$$

which is valid for testing H_0.

Note that, for a general sharp null hypothesis, we can obtain a valid p-value in a similar way.