Please provide complete and well-written solutions to the following exercises.

Due April 27, 2PM PST, to be uploaded in blackboard as a single PDF document (in the Assignments tab).

Homework 12

Exercise 1. Let $X_1, X_2, \ldots$ be independent random variables, each with exponential distribution with parameter $\lambda = 1$. For any $n \geq 1$, let $Y_n := \max(X_1, \ldots, X_n)$. Let $0 < a < 1 < b$. Show that $\mathbf{P}(Y_n \leq a \log n) \to 0$ as $n \to \infty$, and $\mathbf{P}(Y_n \leq b \log n) \to 1$ as $n \to \infty$. Conclude that $Y_n / \log n$ converges to 1 in probability as $n \to \infty$.

Exercise 2. We say that random variables $X_1, X_2, \ldots$ converge to a random variable X in L_2 if

$$\lim_{n\to\infty} \mathbf{E}\,|X_n - X|^2 = 0.$$

Show that, if $X_1, X_2, \ldots$ converge to X in L_2, then $X_1, X_2, \ldots$ converges to X in probability.

Is the converse true? Prove your assertion.

Exercise 3. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables such that $\mathbf{E}\,|X_1| < \infty$ and $\mathrm{var}(X_1) < \infty$. For any $n \geq 1$, define

$$Y_n := \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Show that $Y_1, Y_2, \ldots$ converges in probability. Express the limit in terms of $\mathbf{E}X_1$ and $\mathrm{var}(X_1)$.

Exercise 4. Let $f, g, h \colon \mathbf{R} \to \mathbf{R}$. We use the notation $f(t) = o(g(t))\ \forall\ t \in \mathbf{R}$ to denote $\lim_{t\to 0} \left|\frac{f(t)}{g(t)}\right| = 0$. For example, if $f(t) = t^3\ \forall\ t \in \mathbf{R}$, then $f(t) = o(t^2)$, since $\lim_{t\to 0} |\frac{f(t)}{t^2}| = \lim_{t\to 0} |t| = 0$. Show: (i) if $f(t) = o(g(t))$ and if $h(t) = o(g(t))$, then $(f + h)(t) = o(g(t))$. (ii) If c is any nonzero constant, then $o(cg(t)) = o(g(t))$. (iii) $\lim_{t\to 0} g(t) o(1/g(t)) = 0$. (iv) $\lim_{t\to 0} o(g(t))/g(t) = 0$. (v) $o(g(t) + o(g(t))) = o(g(t))$.

Exercise 5. This exercise demonstrates that geometry in high dimensions is different than geometry in low dimensions.

Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. Let $||x|| := \sqrt{x_1^2 + \cdots + x_n^2}$. Let $\varepsilon > 0$. Show that for all sufficiently large n, "most" of the cube $[-1, 1]^n$ is contained in the annulus

$$A := \{x \in \mathbf{R}^n \colon (1 - \varepsilon)\sqrt{n/3} \leq ||x|| \leq (1 + \varepsilon)\sqrt{n/3}\}.$$

That is, if $X_1, \ldots, X_n$ are each independent and identically distributed in $[-1, 1]$, then for n sufficiently large

$$\mathbf{P}((X_1, \ldots, X_n) \in A) \geq 1 - \varepsilon.$$

(Hint: apply the weak law of large numbers to $X_1^2, \ldots, X_n^2$.)

Exercise 6 (Confidence Intervals)**.** Among 625 members of a bank chosen uniformly at random among all bank members, it was found that 25 had a savings account. Give an interval of the form $[a,b]$ where $0 \leq a,b \leq 625$ are integers, such that with about 95% certainty, if we sample 625 bank members independently and uniformly at random (from a very large bank membership), then the number of these people with savings accounts lies in the interval $[a,b]$. (Hint: if Y is a standard Gaussian random variable, then $\mathbf{P}(-2 \leq Y \leq 2) \approx .95$.)

Exercise 7 (Hypothesis Testing)**.** Suppose we run a casino, and we want to test whether or not a particular roulette wheel is biased. Let p be the probability that red results from one spin of the roulette wheel. Using statistical terminology, "$p = 18/38$" is the null hypothesis, and "$p \neq 18/38$" is the alternative hypothesis. (On a standard roulette wheel, 18 of the 38 spaces are red.) For any $i \geq 1$, let $X_i = 1$ if the i^{th} spin is red, and let $X_i = 0$ otherwise.

Let $\mu := \mathbf{E}X_1$ and let $\sigma := \sqrt{\mathrm{var}(X_1)}$. If the null hypothesis is true, and if Y is a standard Gaussian random variable

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}\right| \geq 2\right) = \mathbf{P}(|Y| \geq 2) \approx .05.$$

To test the null hypothesis, we spin the wheel n times. In our test, we reject the null hypothesis if $|X_1 + \cdots + X_n - n\mu| > 2\sigma\sqrt{n}$. Rejecting the null hypothesis when it is true is called a type I error. In this test, we set the type I error percentage to be 5%. (The type I error percentage is closely related to the p-value.)

Suppose we spin the wheel $n = 3800$ times and we get red 1868 times. Is the wheel biased? That is, can we reject the null hypothesis with around 95% certainty?

Exercise 8. Suppose random variables $X_1, X_2, \ldots$ converge in probability to a random variable X. Prove that $X_1, X_2, \ldots$ converge in distribution to X.

Then, show that the converse is false.

ALL EXERCISES BELOW ARE OPTIONAL. THEY WILL NOT BE GRADED.

Exercise 9 (Optional)**.** Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. For any $n \geq 1$, define

$$S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

The Central Limit Theorem says that S_n converges in distribution to a standard Gaussian random variable. We show that S_n does not converge in probability to any random variable. The intuition here is that if S_n did converge in probability to a random variable Z, then when n is large, S_n is close to Z, $Y_n := \frac{\sqrt{2}S_{2n} - S_n}{\sqrt{2}-1}$ is close to Z, but S_n and Y_n are independent. And this cannot happen.

Proceed as follows. Assume that S_n converges in probability to Z.

- Let $\varepsilon > 0$. For n very large (depending on ε), we have $\mathbf{P}(|S_n - Z| > \varepsilon) < \varepsilon$ and $\mathbf{P}(|Y_n - Z| > \varepsilon) < \varepsilon$.

- Show that $\mathbf{P}(S_n > 0, Y_n > 0)$ is around $1/4$, using independence and the Central Limit Theorem.
- From the first item, show $\mathbf{P}(S_n > 0 | Z > \varepsilon) > 1 - \varepsilon$, $\mathbf{P}(Y_n > 0 | Z > \varepsilon) > 1 - \varepsilon$, so $\mathbf{P}(S_n > 0, Y_n > 0 | Z > \varepsilon) > 1 - 2\varepsilon$.
- Without loss of generality, for ε small, we have $\mathbf{P}(Z > \varepsilon) > 4/9$.
- By conditioning on $Z > \varepsilon$, show that $\mathbf{P}(S_n > 0, Y_n > 0)$ is at least $3/8$, when n is large.

Exercise 10 (Optional)**.** Let $X_1, X_2, \ldots$ be random variables that converge almost surely to a random variable X. That is,

$$\mathbf{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Show that $X_1, X_2, \ldots$ converges in probability to X in the following way.

- For any $\varepsilon > 0$ and for any positive integer n, let

$$A_{n,\varepsilon} := \bigcup_{m=n}^{\infty} \{\omega \in \Omega \colon \ |X_m(\omega) - X(\omega)| > \varepsilon\}.$$

  Show that $A_{n,\varepsilon} \supseteq A_{n+1,\varepsilon} \supseteq A_{n+2,\varepsilon} \supseteq \cdots$.
- Show that $\mathbf{P}(\cap_{n=1}^{\infty} A_{n,\varepsilon}) = 0$.
- Using Continuity of the Probability Law, deduce that $\lim_{n\to\infty} \mathbf{P}(A_{n,\varepsilon}) = 0$.

Now, show that the converse is false. That is, find random variables $X_1, X_2, \ldots$ that converge in probability to X, but where $X_1, X_2, \ldots$ do not converge to X almost surely.

Exercise 11 (Optional, Renewal Theory)**.** Let $t_1, t_2, \ldots$ be positive, independent identically distributed random variables. Let $\mu \in \mathbf{R}$. Assume $\mathbf{E}t_1 = \mu$. For any positive integer j, we interpret t_j as the lifetime of the j^{th} lightbulb (before burning out, at which point it is replaced by the $(j+1)^{st}$ lightbulb). For any $n \geq 1$, let $T_n := t_1 + \cdots + t_n$ be the total lifetime of the first n lightbulbs. For any positive integer t, let $N_t := \min\{n \geq 1 \colon T_n \geq t\}$ be the number of lightbulbs that have been used up until time t. Show that N_t/t converges almost surely to $1/\mu$ as $t \to \infty$. (Hint: by definition of N_t, we have $T_{N_t - 1} < t \leq T_{N_t}$. Now divide the inequalities by N_t and apply the Strong Law.)

Exercise 12 (Optional, Playing Monopoly Forever)**.** Let $t_1, t_2, \ldots$ be independent random variables, all of which are uniform on $\{1, 2, 3, 4, 5, 6\}$. For any positive integer j, we think of t_j as the result of rolling a single fair six-sided die. For any $n \geq 1$, let $T_n = t_1 + \cdots + t_n$ be the total number of spaces that have been moved after the n^{th} roll. (We think of each roll as the amount of moves forward of a game piece on a very large Monopoly game board.) For any positive integer t, let $N_t := \min\{n \geq 1 \colon T_n \geq t\}$ be the number of rolls needed to get t spaces away from the start. Using Exercise 11, show that N_t/t converges almost surely to $2/7$ as $t \to \infty$.

Exercise 13 (Optional, Random Numbers are Normal)**.** Let X be a uniformly distributed random variable on $(0, 1)$. Let X_1 be the first digit in the decimal expansion of X. Let X_2 be the second digit in the decimal expansion of X. And so on.

- Show that the random variables $X_1, X_2, \ldots$ are uniform on $\{0, 1, 2, \ldots, 9\}$ and independent.

- Fix $m \in \{0, 1, 2, \ldots, 9\}$. Using the Strong Law of Large Numbers, show that with probability one, the fraction of appearances of the number m in the first n digits of X converges to $1/10$ as $n \to \infty$.

(Optional): Show that for any ordered finite set of digits of length k, the fraction of appearances of this set of digits in the first n digits of X converges to 10^{-k} as $n \to \infty$. (You already proved the case $k = 1$ above.) That is, a randomly chosen number in $(0, 1)$ is normal. On the other hand, if we just pick some number such that $\sqrt{2} - 1$, then it may not be easy to say whether or not that number is normal.

(As an optional exercise, try to explicitly write down a normal number. This may not be so easy to do, even though a random number in $(0, 1)$ satisfies this property!)

Exercise 14 (Optional)**.** Using the Central Limit Theorem, prove the Weak Law of Large Numbers.

Exercise 15 (Optional)**.** Let $X_1, X_2, \ldots$ be random variables with mean zero and variance one. The Strong Law of Large Numbers says that $\frac{1}{n}(X_1 + \cdots + X_n)$ converges almost surely to zero. The Central Limit Theorem says that $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ converges in distribution to a standard Gaussian random variable. But what happens if we divide by some other power of n? This Exercise gives a partial answer to this question.

Let $\varepsilon > 0$. Show that

$$\frac{X_1 + \cdots + X_n}{n^{1/2}(\log n)^{(1/2)+\varepsilon}}$$

converges to zero almost surely as $n \to \infty$. (Hint: Re-do the proof of the Strong Law of Large Numbers, but divide by $n^{1/2}(\log n)^{(1/2)+\varepsilon}$ instead of n.)