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

Due March 2, in the discussion section.

Homework 6

Exercise 1. 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

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

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 2 (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, the number of these 625 bank members with savings accounts lies in the interval $[1, 5]$. (Hint: if Y is a standard Gaussian random variable, then $\mathbf{P}(-2 \leq Y \leq 2) \approx .95$.)

Exercise 3 (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 4. 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.

Exercise 5. 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 6. 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 7. Using the Central Limit Theorem, prove the Weak Law of Large Numbers.