Assignment 8

CSCI244-Randomness and Computation
Assigned April 15
Due Thursday, April 25

April 20, 2019

Problems about the Normal Distribution and the Central Limit Theorem. For some of these problems, you will need to supply a number. Remember that you can use `norm.cdf` and `norm.ppf` from the package `scipy.stats` to compute the cumulative distribution function and the inverse cumulative distribution function, respectively, for the standard normal random variable. The first three problems were inspired by similar exercises in the book by Grinstead and Snell.

1 Limits

Suppose we make n Bernoulli trials, where $p = 0.8$. Let S_n denote the sum of the outcomes, and $A_n = S_n/n$. Determine the following limits. Justify your answer. In every case, the normal approximation to the binomial distribution is sufficient to answer the question, and in most instances you don't have to compute anything.

(a) $\lim_{n\to\infty} P(A_n = 0.8)$

(b) $\lim_{n\to\infty} P(0.7n < S_n < 0.9n)$

(c) $\lim_{n\to\infty} P(0.7n < S_n < 0.8n + 0.5\sqrt{n})$

(d) $\lim_{n\to\infty} P(0.7n < S_n < 0.9n)$

(e) $\lim_{n\to\infty} P(A_n > 0.7)$

(f) $\lim_{n\to\infty} P(A_n > 0.7 - \frac{1}{n}))$

(g) $\lim_{n\to\infty} P(A_n > 0.7 - \frac{1}{\sqrt{n}})$

2 Grades

A student's performance in a course is based on the scores on 20 brief assignments, each of which is graded on a scale from 0 to 100. Nobody's perfect, not even the professor in the course, and it turns out that the grade he gives on an assignment will vary as much a ± 5 points from the 'correct' grade, with a probability of $\frac{1}{10|k|}$ for an error of $k \neq 0$ points. Thus if we let X denote the random variable denoting the error in the grade on a random assignment, we will have

$$P(X = 5) = P(X = -5) = \frac{1}{50},$$

$$P(X = 4) = P(X = -4) = \frac{1}{40},$$

$$P(X = 3) = P(X = -3) = \frac{1}{30},$$

$$P(X = 2) = P(X = -2) = \frac{1}{20},$$

$$P(X = 1) = P(X = -1) = \frac{1}{10}.$$

It follows that the probability that the grade is exactly correct ($P(X = 0)$) is $1 - s$, where

$$s = 2(\frac{1}{10} + \frac{1}{20} + \frac{1}{30} + \frac{1}{40} + \frac{1}{50}) = 0.45667.$$

No big deal—the grade is exactly right more than half the time, and the small errors in favor of the student balance out those to the detriment of the student.

The student's final grade is also a number between 0 and 100. Thus the error in the computation of this grade is

$$(X_1 + \cdots + X_{20})/20,$$

where X_i denotes the error on the i^{th} assignment. Estimate the probability that the grade is within 2 points of the correct grade. Estimate the probability that the grade is within one half a point of the correct grade.

Note. You may assume that the X_i are mutually independent. We also assume for purposes of this problem that the average of these 20 errors is very well approximated by a normal distribution. In fact, this is something that you can verify directly.

3 Hypothesis-testing

Alice claims that the coin she is using in her coin-flipping contest with Bob is fair. Bob claims that it is unfair, and has been doctored to come up heads at least 60% of the time.

To test these hypotheses, they will flip the coin a large number N of times. The goal of this problem is to find values m and N so that if the coin comes up heads more than m times, we will reject Alice's claim (in Stats talk, the *null hypothesis*) with high confidence, and if the coin comes up less than m times, we will reject Bob's hypothesis (the *alternate hypothesis*) with high confidence

(a) Assuming that Alice is correct, write an expression for the *exact* probability that the coin comes up heads at least m times. (The expression will involve m and N, of course, as well as binomial coefficients.)

(b) Suppose $N = 100$. Using the normal approximation to the binomial distribution, find the smallest value m_1 so that

$$P(X > m_1) < 0.02,$$

where X is the number of heads.

(c) Now redo (a) and (b) above, this time assuming Bob's hypothesis that the coin comes up heads 60% of the time: that is, use the normal approximation to find the *largest* value m_2 so that

$$P(X < m_2) < 0.02,$$

assuming $N = 100$.

(*d) You should find that with $N = 100$, $m_2 < m_1$. This means that there is a gray area–if the number of heads is between m_2 and m_1, then we cannot rule out either hypothesis. However, you should notice that if we increase N, the m_2 will grow larger and m_1 smaller. Find as small a value of N as you can that makes $m_1 < m_2$. (We can then use any value m between m_1 and m_2 as the *critical value* for this experiment: If the number of heads is less than m it is highly unlikely that Bob 's claim is correct; if it is greater than m then it is highly unlikely that the coin is fair.)

4 Central Limit Theorem for a Continuous Distribution

On each turn of a game, you spin two spinners and you receive as a score the the maximum of these two spinners. We let Y denote the random variable that gives the score. Note that Y can take on any value between 0 and 1.

(a) Compute and plot the density of Y.

(b) Compute the mean and the variance of Y.

(c) What is the probability that the average score on 100 spins is within $\frac{1}{10}$ of the mean?

(d) Now write a simulation of this game: You should have a function `spingame(N)` that spins the pair of spinners N times and returns the average score. Another function should call `spingame` M times, and return an array of the resulting scores. Try this for $N = 100$ and $M = 1000$ and draw a probability plot. The data should be pretty close to a straight line, indicating that the distribution is close to normal, and you should be able to recover the mean and standard deviation from the y-intercept and slope of the line.