Math 318 – homework 6 – due 2023-03-10

Problem 1. Assume X_i are iid samples with $N(\mu, \sigma^2)$ distribution, with unknown μ, σ. We would like to test the null hypothesis $\mu = 200$. A sample of size 9 has sample mean $X = 205$.

(a) Find the p-value (probability of deviating at least that much from the mean)if the standard deviation is known to be
  (i) $\sigma = 5$
  (ii) $\sigma = 10$
  (iii) $\sigma = 15$

(b) In which of the three cases would the null hypothesis be rejected at the 5% level of significance?

(c) In which of the three cases would the null hypothesis be rejected at the 1% level of significance?

Problem 2. ESP cards have one of five shapes on them (cross, circle, square, star and waves) with equal probability. A psychic claims he can guess a card correctly with probability 0.5. James Randy designs an experiment, where the psychic tries to guess N cards, and is considered successful if he made at least $N/3$ correct guesses.

(a) How large does N need to be so that the probability of success is at most 1/1000, if the psychic has no special ability (probability 1/5 at each guess)?

(b) For that N, what is the probability of failure if the psychic's claim is true?

Problem 3. A bakery claims their bread weighs on average 1Kg.

(a) An investigator weighed 1000 loaves over a year, and summarized his results in a file `bread_data` (available on the course website). Based on this data, find a 95% confidence interval for μ.

(b) Based on this data, can we rule out the bakery's claim at a 5% level?

(c) If the mean weight actually is 1Kg, what is the likelyhood of us reaching that decision?

(d) If 50 investigators repeated this procedure, how many of them on average would reject the bakery's claim?

Problem 4. Let (X, Y) be a point in the rectangle $0 \le x \le 2$ and $0 \le y \le 1$, with joint pdf $f(x, y) = \frac{xy+1}{3}$.

(a) Find the conditional pmf $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

(b) Find the conditional expectations $E[X|Y]$ and $E[Y|X]$.

Problem 5. Let X be uniform from the set $\{4, 6, 8, 12\}$. Given X, we take an X-sided die, and toss it X times. Let Y be the sum of these dice.

(a) Find $E[Y|X]$. (Hint: the expected result on an n-sided die is $\frac{n+1}{2}$.)

(b) Use that to fine $E[Y]$.

Problem 6. Consider the standard normal probability density function $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. It is known that there is no closed form for the antiderivative of this function, i.e., for the c.d.f. Φ of the standard normal. However, the c.d.f. can be approximated accurately, and the tables give

$$\int_0^1 f(x)\,dx = \Phi(1) - \Phi(0) \approx 0.3413.$$

(a) To demonstrate the method of Monte Carlo integration, approximate the integral $\int_0^1 f(x)\,dx$ by generating 10000 i.i.d. uniform random numbers on $[0, 1]$ and computing the approximation

$$I_{10000} = \frac{f(U_1) + f(U_2) + \cdots + f(U_{10000})}{10000}.$$

Do this 100 times and record the results from each run.
Compute the average result (sample mean) and the sample variance of the 100 results, and compare them to their expectation. Submit your code and your output.

(b) Another method for approximating this integral is to recall that for $f \geq 0$, the integral $\int_0^1 f(x)\,dx$ represents the area underneath the graph of f from $x = 0$ to $x = 1$. To estimate this area, one could simulate a large number of uniform points in the *square* with corners at $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$; then, find the proportion of points that lie underneath the curve $y = f(x)$. Give an argument (a formal proof is not required, just a motivation) as to why this is a reasonable way to approximate this area.

(c) Perform the approximation in (b) by writing code to simulate 10000 i.i.d. uniform random numbers in the square $[0,1] \times [0,1]$ and determine the proportion of them falling in the region $y \leq f(x)$. Do this 100 times and record the sample mean and sample variance of the results. Submit your code and your output.

(d) Is one of the two methods better? Which and why?

Extra practice problems

(a)

(b) Chapter 2: 78 (read about Markov's inequality), *80.

(c) 25 measurements are made of the splitting tensile stress (lb/in^2) of concrete cylinders. The following table shows the frequency of each measured value, with the strength on the first line and the frequency on the second line. Assuming a normal distribution, determine a 99% confidence interval for the mean splitting tensile stress of the population from which the sample was drawn.

320	330	340	350	360	370	380	390
1	1	3	3	8	3	5	1

(d) 1000 people play roulette at a casino, each betting \$100 on either even or odd. Each wins \$100 with probability $\frac{18}{38}$ and loses their bet otherwise. Estimate the probability that the casino loses money on that game at the end of the day.

(e) Repeat the same if the people bet on a single number, so they win \$3600 with probability $\frac{1}{38}$ and lose their bet otherwise.