MODULE 4 EXAMPLES

TAs: Nihal Reddy
Email: nireddy@ucsd.edu
OH: Thursdays 6-7pm

Slide Credits: Kira Fleischer

PROBLEM #: KEY TOPICS FROM PROBLEM

Problem setup and description.

Question

Key notes from readings/lectures needed to answer the question

Solution: written with as much detail as we expect you to give on your homework sets

PROBLEM 1: PROBABILITY MEASURES

Discuss whether the following is a valid assignment of probabilities to a discrete random variable which takes values 1, 3, and 5.

(a) P (X = 1) = 0.1; P (X = 3) = 1.3; P (X = 5) = 0.2

For a sample space Ω with events $A_i \epsilon \Omega$:

$P(A_i) \epsilon [0,1]$: Probability of each event in the sample space must be in the range [0, 1].

$P(\Omega) = 1$: Probabilities of all events in the sample space must sum to 1.

Solution: This is not valid since not all individual probabilities are in [0, 1].

(b) P (X = 1) = 0.3; P (X = 3) = 0.3; P (X = 5) = 0.4

Solution: This is valid since the probabilities do add up to 1 while every individual one is in [0, 1].

PROBLEM 2: DISCRETE AND CONTINUOUS RVS

Discuss whether the following are discrete or continuous random variables.

(a) The number of points that a player scores in a basketball game.

Discrete RV: RV has a countable number of distinct values

Continuous RV: RV can reflect an infinite number of potential values within a range

Solution: This will always be an integer and hence is a discrete random variable.

(b) The number of Chemistry courses that a UCSD student has taken in their sophomore year.

Solution: This will always be an integer and hence is a discrete random variable.

(c) The distance you walk on a given day.

Solution: This is on a continuous scale. Therefore it is a continuous random variable

PROBLEM 3: BINOMIAL RANDOM VARIABLES

Luke takes a True or False exam with 30 questions in total. However, he didn't study for the exam at all, so he flips a fair coin to randomly answer each question.

(a) What's the probability that he gets exactly 13 questions right?

Binomial Random Variable: a RV that counts the number of "successes" in a fixed number of trials

For $X \sim Bin(n, p)$, we have $P(X = k) = \binom{n}{k} * p^k * (1-p)^{n-k}$

Solution: Let X denote the number of questions he guesses correctly.

Since the number of questions on the exam is 30 and he gets questions correct with a probability of 0.5, we know $X \sim Bin(30, 0.5)$.

Since we are looking for the probability that he gets exactly 13 questions correct, we calculate $P(X = 13) = \binom{30}{13} * 0.5^{13} * (1-0.5)^{30-13} = \binom{30}{13} * 0.5^{30} = 0.11154$

PROBLEM 3: BINOMIAL RANDOM VARIABLES

Luke takes a True or False exam with 30 questions in total. However, he didn't study for the exam at all, so he flips a fair coin to randomly answer each question.

Binomial Random Variable: a RV that counts the number of "successes" in a fixed number of trials

For $X \sim Bin(n, p)$, we have $P(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

(b) What's the probability that he will get a grade of 95% or higher on the exam?

Solution: We use the same set up as in part (a). Now we know that to get a grade of at least 95%, he must answer at least $30 * 0.95 = 28.5 \rightarrow 29$ questions correctly.

Thus we calculate $P(X \geq 29) = P(X = 29) + P(X = 30) = \binom{30}{29} * 0.5^{29} * (0.5)^1 + \binom{30}{30} * 0.5^{30} = 0.5^{30}\left[\binom{30}{29} + \binom{30}{30}\right] = 0.5^{30}(30 + 1) = 2.89 * 10^{-8}$

PROBLEM 3: BINOMIAL RANDOM VARIABLES

Luke takes a True or False exam with 30 questions in total. However, he didn't study for the exam at all, so he flips a fair coin to randomly answer each question.

Binomial Random Variable: a RV that counts the number of "successes" in a fixed number of trials

For $X \sim Bin(n, p)$, we have $P(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

(c) Is it likely he will get 100% on the exam?

Solution: The probability he gets 100% is $P(X = 30) = \binom{30}{30} * 0.5^{30} = 9.31 * 10^{-10}$, thus it is very unlikely he will get 100% on the exam.

PROBLEM 4: PROPERTIES OF CONTINUOUS RVS

Given that X is a continuous random variable, discuss whether the following statements are true.

(a) The expected value of X must be 1.

Solution: False. The expected value of a continuous random variable can be any real number.

(b) P(X = 4) must be positive.

Solution: False. If X is a continuous random variable, then the probability that it takes on any one particular value must be zero.

PROBLEM 5: SYMMETRIC DISTRIBUTIONS

Suppose Z has a standard normal distribution. We know that Z is symmetric.

If given $P(Z < -a) = 0.25$, what is $P(Z > a)$? Explain.

Solution: Since Z is symmetric and $P(Z < -a) = 0.25$, we know that $P(Z > a) = 0.25$.

PROBLEM 6: BINOMIAL RVS

Suppose that in a factory, the percentage of broken light bulbs is about 4%. We take a look at a batch of 100 light bulbs.

(a) What is the expected number of broken light bulbs?

Binomial Random Variable: a RV that counts the number of "successes" in a fixed number of trials

For $X \sim Bin(n, p)$, we have $\mathbb{E}[X] = np$

Solution: Let X be the number of broken light bulbs in a batch. Then X has binomial distribution with parameters n = 100; p = 0.04. Then the expected number of broken light bulbs you should see in a batch is:

$$\mathbb{E}[X] = np = 100 * 0.04 = 4$$

PROBLEM 6: BINOMIAL RVS

Suppose that in a factory, the percentage of broken light bulbs is about 4%. We take a look at a batch of 100 light bulbs.

(b) What is the standard deviation?

For $X \sim Bin(n, p)$, we have $\sigma_X = \sqrt{np(1-p)}$

Solution: Following the same set up as in part (a), we have

$$\sigma_X = \sqrt{100 * 0.04 * 0.96} \approx 1.96$$

PROBLEM 7: APPROXIMATING BINOMIAL DIST WITH NORMAL

At UCSD, 36% of the undergraduate student body in Fall 2023 identify as first-generation college students.

(a) Can we use the Central Limit Theorem to approximate the probability of getting 110 or more first-generation students out of 300 randomly selected undergraduate students in Fall 2023?

From the Central Limit Theorem, we get that a binomial distribution $Bin(n, p)$ can be approximated by a normal distribution $N(\mu, \sigma)$ with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$ when the sample size n is sufficiently large so that $np \geq 10$ and $n(1-p) \geq 10$.

Solution: Since $np = 300 * 0.36 = 108 \geq 10$ and $n(1-p) = 300 * 0.64 = 192 \geq 10$, we can approximate $Bin(300, 0.36)$ with $N(\mu, \sigma)$ where $\mu = 108$ and $\sigma = \sqrt{108 * 0.64} \approx 8.314$

PROBLEM 7: APPROXIMATING BINOMIAL DIST WITH NORMAL

At UCSD, 36% of the undergraduate student body in Fall 2023 identify as first-generation college students.

(b) If we can apply CLT, approximate the probability that more than 110 students are first-generation in the student body of Fall 2023 out of 300 students.

Since we can approximate this Binomial distribution with a Normal distribution, follow the typical procedure of computing a z-score to find the probability in question:

$$Z = \frac{X-\mu}{\sigma};\ P(X > x) = P\left(\frac{X-\mu}{\sigma} > \frac{x-\mu}{\sigma}\right) = P\left(Z > \frac{x-\mu}{\sigma}\right)$$

Solution: Let X be the number of first-generation students in the student body of Fall 2023 out of those 300 students. Then

$$P(X > 110) = P\left(\frac{X-108}{8.314} > \frac{110-108}{8.314}\right) = P(Z > 0.2406) = 1 - 0.5948 = .4052$$

PROBLEM 8: USING THE STANDARD NORMAL DIST.

A certain test has its scores approximately normally distributed with a mean of 18 and a standard deviation of 6. **What is the IQR of the test scores?**

Ultimately we want to find the IQR of the test scores, which is the 75^{th} percentile of test scores minus the 25^{th} percentile of test scores.

Suppose we call our test scores random variables X_i, with $X_i \sim N(18, 6^2)$.

What are the 25^{th} and 75^{th} percentiles of this random variable Xi?

If we normalize X, we know that this new random variable (Z) will follow a standard normal distribution (due to the CLT), and we can easily find the 25^{th} and 75^{th} percentiles of this distribution by using the z-score table (or technology).

PROBLEM 8: USING THE STANDARD NORMAL DIST.

A certain test has its scores approximately normally distributed with a mean of 18 and a standard deviation of 6. **What is the IQR of the test scores?**

Solution: We have $X_i \sim N(18, 6^2)$

Normalizing X, we get $Z = \frac{X-\mu}{\sigma} = \frac{X-18}{6} \sim N(0,1)$

To find the 25th and 75th percentiles of Z, you can use the z-scores table, Python, R, Excel, etc.

Thus we have $P(Z < -0.6745) = 0.25$, and

$$P(Z < 0.6745) = 0.75$$

Using these lower and upper quartiles, we have

$\frac{Q3-18}{6} = 0.6745$, and $\frac{Q1-18}{6} = -0.6745$

Thus IQR $= Q3 - Q1 = 8.094$

Example (Python code):

```
[2]  import scipy.stats as stats

     # 25th percentile
     stats.norm.ppf(0.25)

     -0.6744897501960817

[4]  # 75th percentile
     stats.norm.ppf(0.75)

     0.6744897501960817
```

PROBLEM 9: Z AND T STATISTICS

Suppose that for a random variable X we know the population standard deviation σ with sample size 100, but we do not know the population mean μ.

(a) What distribution does the random variable $\frac{\bar{X}-\mu}{\sigma/\sqrt{100}}$ follow?

Solution: If we know the population standard deviation σ, then $\frac{\bar{X}-\mu}{\sigma/\sqrt{100}}$ follows a standard normal distribution N(0,1) due to the Central Limit Theorem.

(b) If we don't know the population standard deviation, but use s instead, what distribution would the random variable $\frac{\bar{X}-\mu}{s/\sqrt{100}}$ follow? With what degrees of freedom?

Solution: If we don't know σ and use s instead, $\frac{\bar{X}-\mu}{s/\sqrt{100}}$ follows a t-distribution with df = 99.