Probability and Applied Statistics

Final Report

Ivan Wang

Chapter 2: Probability

2.3 A Review of Set Notation

From a group of students in the 'StudentsPerformance.csv' dataset, it was found that 150 students have completed a test preparation course, 200 students scored 70 or above in math, and 50 of these students both have completed a test preparation course and scored 70 or above in math. Find the number of students who either have completed a test preparation course, scored 70 or above in math, or both.

The number of students who either have completed a test preparation course, scored 70 or above in math, or both, is 300.

2.4: A Probabilistic Model for an Experiment: The Discrete Case

A student's performance in a test can be categorized as 'Pass' or 'Fail' in each of the three subjects: math, reading, and writing. The experiment consists of observing the performance of a single student across these three subjects.

Sample space S:
(Pass in Math, Pass in Reading, Pass in Writing)
(Pass in Math, Pass in Reading, Fail in Writing)
(Pass in Math, Fail in Reading, Pass in Writing)
(Pass in Math, Fail in Reading, Fail in Writing)
(Fail in Math, Pass in Reading, Pass in Writing)
(Fail in Math, Pass in Reading, Fail in Writing)
(Fail in Math, Fail in Reading, Pass in Writing)
(Fail in Math, Fail in Reading, Fail in Writing)

2.5: Calculating the Probability of an Event: The Sample Point Method

There are students who scored above 90 in math (considered 'high scorers') and those who didn't. A study group is being formed by randomly selecting two students from a group of ten students, which includes four high scorers in math and six who aren't. What is the probability that both students selected for the study group are high scorers in math?

$$\text{Comb}(10,2) = \frac{10!}{2!\,(10-2)!}$$

$$\text{Probability} = \frac{\text{Ways to select high scorers}}{\text{Total ways to select}} = \frac{6}{45}$$

2.6: Tools for Counting Sample Points

Consider two groups of students based on their scores in math and reading. Group A consists of students who scored above 80 in math, and Group B consists of students who scored above 80 in reading. If a student can be a member of both groups, how many different combinations of students can be formed from these two groups?

$$|A| + |B| - |A \cap B| = 100 + 120 - 30 = 190$$

2.7: Conditional Probability and the Independence of Events

Consider that students are randomly selected one at a time. If the first two students selected both scored above 90 in math, what is the probability that the next three students selected will also score above 90 in math?

$$P = \frac{98}{998} \times \frac{97}{997} \times \frac{96}{996} \approx 0.00092$$

2.8: Two Laws of Probability

Suppose there is a 1 in 50 chance of a student scoring below 50 in math on a single test. If we assume that the outcomes of different tests are independent, what is the probability that a student scores below 50 in math if they take two tests?

$$P(\text{below 50 on both tests}) = \left(\frac{1}{50}\right) \times \left(\frac{1}{50}\right) = 0.04\%.$$

2.10: The Law of Total Probability and Bayes' Rule

In a certain school, 40% of students prefer subject A and 60% prefer subject B. It is known that 30% of the students who prefer subject A and 70% of those who prefer subject B excel in a particular extracurricular activity. A student is chosen at random and is found to excel in this activity. What is the conditional probability that this student prefers subject B?

$$P(B|\text{Excel}) = \frac{P(\text{Excel}|B) \cdot P(B)}{P(\text{Excel}|A) \cdot P(A) + P(\text{Excel}|B) \cdot P(B)}$$

$$P(B|\text{Excel}) = \frac{0.70 \cdot 0.60}{0.30 \cdot 0.40 + 0.70 \cdot 0.60} \approx 0.778$$

Chapter 3: Discrete Random Variables and Their Probability Distributions

3.2: The Probability Distribution for a Discrete Random Variable

In a certain school, it's found that 20% of students don't participate in any extracurricular activities, 40% participate in Activity A, and 50% participate in Activity B. (Some students participate in both activities.) If a student is randomly chosen from this school, find the probability distribution for Y, the number of extracurricular activities participated in by the student.

$$P(Y=1) = P(\text{participating in A only}) + P(\text{participating in B only})$$

$$P(Y=2) = P(\text{participating in A}) + P(\text{participating in B}) - P(Y=1) - P(Y=0)$$

$$P(Y = 0) = 0.20$$

$$P(Y = 1) = 0.70$$

$$P(Y = 2) = 0.10$$

3.4: The Binomial Probability Distribution

Experience shows that 30% of students in a certain course pass the final exam. A new study method is introduced, and ten students randomly selected use this method; nine of them pass the exam. Suppose the study method is absolutely ineffective. What is the probability that at least nine out of the ten students using the method will pass the exam?

$$P(\text{at least 9 out of 10 pass}) = P(X=9) + P(X=10)$$

$$P(X=9) = \binom{10}{9}(0.30)^9(1-0.30)^{10-9}$$

$$P(X=10) = \binom{10}{10}(0.30)^{10}(1-0.30)^{10-10}$$

$$P(\text{at least 9 out of 10 pass}) \approx 0.000144$$

3.5: The Geometric Probability Distribution

If the probability of a student getting a perfect score on any one test is 30%, and Y denotes the number of tests taken until the first perfect score, find the probability that the first student with a perfect score is the fifth student to turn in their test.

$$P(Y=5) = (1-0.30)^{5-1} \times 0.30$$

$$P(Y=5) = (1-0.30)^{5-1} \times 0.30 \approx 0.07203$$

3.6: The Negative Binomial Probability Distribution

Assuming a 20% chance that any given student in a class answers a question correctly, find the probability that the third correct answer comes from the fifth student asked.

$$P(Y=5) = \binom{5-1}{3-1}(0.20)^3(1-0.20)^{5-3} \approx 0.03072$$

3.7: The Hypergeometric Probability Distribution

A class has ten students, four of whom did not complete their homework. A teacher selects five of the students at random for a review, thinking all have completed their homework. What is the probability that all five of the selected students have completed their homework?

$$P(\text{all 5 completed homework}) = \frac{\text{Comb}(6,5)}{\text{Comb}(10,5)} \approx 0.02381$$

3.8: The Poisson Probability Distribution

In a large area, a certain type of question in a test bank is randomly dispersed, with an average density of five questions per topic. If an instructor randomly selects ten topics for a quiz, each representing one question, find the probability that none of the topics will contain these specific questions.

$$P(X = 0) = \frac{e^{-50} \times 50^0}{0!} \approx 1.93 \times 10^{-22}$$

3.11: Chebychev's Theorem

The number of correctly answered questions per test by students, Y, has been observed over a long period and found to have a mean of 20 and a standard deviation of 2. The probability distribution of Y is not known. What can be said about the probability that, on the next test, Y will be greater than 16 but less than 24?

$$P(16 < Y < 24) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 0.75$$

Chapter 4: Continuous Variables and Their Probability Distributions

4.2: The Probability Distribution for a Continuous Random Variable

A school cafeteria has a 150-student capacity and notices that the number of students using the cafeteria increases steadily up to 100 students and then levels off between 100 and 150 students. If Y denotes the weekly number of students in hundreds using the cafeteria, the relative frequency of students can be modeled by a piecewise function.

$$f(y)\begin{cases} y, & 0 \leq y \leq 1 \\ 1, & 1 < y \leq 1.5 \\ 0, & elsewhere \end{cases}$$

a. Find the cumulative distribution function F(y).

$$\text{F}(y)\begin{cases} \frac{1}{2}y^2, & 0 \leq y \leq 1 \\ \frac{1}{2} + (\text{y - 1}), & 1 < y \leq 1.5 \\ 0, & elsewhere \end{cases}$$

b. Find the probability that up to 50 students use the cafeteria in a week.

$$P(0 \leq Y \leq .5) = F(.5) = \tfrac{1}{2}(.5)^2 = \tfrac{1}{8}$$

c. Find the probability that between 50 and 120 students use the cafeteria in a week.

$$P(.5 \leq Y \leq 1.2) = F(1.2) - F(.5) = 0.7 - \frac{1}{8} = 0.575$$

4.3: Expected Values for Continuous Random Variables

The proportion of time Y that a student spends studying during a 40-hour week is a random variable with a probability density function (PDF) given by

$$f(y)\begin{cases} 2y, & 0 \leq y \leq 1 \\ 0, & elsewhere \end{cases}$$

Find the expected value E(Y) and the variance V(Y).

$$E(Y) = \int_0^1 2y^2\, dy = \left[\frac{2y^3}{3}\right]_0^1 = \frac{2}{3}$$

$$E(Y^2) = \int_0^1 2y^3\, dy = \left[\frac{2y^4}{4}\right]_0^1 = \frac{1}{2}$$

$$V(Y) = E(Y^2) - \left(E(Y)\right)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2$$

4.4: The Uniform Probability Distribution

After analyzing the scores for mid-term exams, a teacher finds that the students' scores are uniformly distributed between 20 and 25 points. Find the probability that the next student's score on the mid-term exam is below 22 points.

$$P(X < 22) = F(22) = \frac{22 - 20}{25 - 20} = \frac{2}{5}$$

Chapter 5: Multivariate Probability Distributions

5.2: Bivariate and Multivariate Probability Distributions

Consider a school canteen that prepares a large batch of lunches at the beginning of the week. Let Y1 represent the proportion of the batch that is prepared at the beginning of the week, and Y2 the proportion of the batch that is actually sold during the week. The joint density of Y1 and Y2, where Y2 is always less than or equal to Y1, is given by:

$$f(x, y)\begin{cases} 3x, & 0 \le y \le x \le 1, \\ 0, & elsewhere. \end{cases}$$

Find the probability that less than half of the batch is prepared and less than a third is sold, i.e., P(Y1 < 1/2, Y2 < 1/3)

$$P(X < 1/2\ , Y < 1/3) = \int_0^{1/3}\int_y^{1/2} 3x\ dx\, dy$$

$$\int_y^{1/2} 3x\ dx = \frac{3}{2}\left[\left(\frac{1}{2}\right)^2 - y^2\right]$$

$$\int_0^{1/3} \frac{3}{2}\left[\left(\frac{1}{2}\right)^2 - y^2\right] dy$$

5.3: Marginal and Conditional Probability Distributions

A classroom has a random number of Y students present at the beginning of a school day, and a random number of X homework assignments are turned in during the day (with measurements in numbers of assignments). No additional homework is turned in during the day, hence X ≤ Y. It has been observed that X and Y have a joint density given by:

$$f(x, y)\begin{cases} 1/2, & 0 \le y \le x \le 1, \\ 0, & elsewhere. \end{cases}$$

Which means the points (x, y) are uniformly distributed over the triangle with the given boundaries. Find the conditional density of X given Y = y. Evaluate the probability that less than half of the homework assignments will be turned in, given that there are 1.5 times as many students as assignments at the start of the day.

$$f(x|Y = y) = \frac{f(x, y)}{f_Y(y)} = \frac{\frac{1}{2}}{\frac{y}{2}} = \frac{1}{y}, \quad 0 \le x \le y$$

$$P(X < 1/2|Y = 1.5) = \int_0^{1/2} \frac{1}{1.5} dx = \frac{1}{1.5} \times \frac{1}{2} = \frac{1}{3}$$