IS 606 Assignment 8

David Stern

April 8, 2015

1. A professor is constructing a multiple-choice history quiz. The quiz will have ten questions, each with four possible answers. The minimum passing score on the quiz is 60% (6 of 10 correct).

a. Suppose a student has not studied at all and decides to guess randomly on all ten questions. What is the probability that the student is able to pass the quiz?

For this problem, we will use a binomial probability distribution.

$$P(r) = \frac{n!}{r!(n-r)!}p^r q^{n-r}$$

$$n = trials$$

$$r = successes$$

$$p = p(success)$$

$$q = p(failure)$$

$$P(PassingtheQuiz) = P(r \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10)$$
$$= \frac{10!}{6!(4)!}0.25^6 0.75^4 + \frac{10!}{7!3!}0.25^7 0.75^3 + \frac{10!}{8!2!}0.25^8 0.75^2 + \frac{10!}{9!1!}0.25^9 0.75^1 + \frac{10!}{10!0!}0.25^{10} 0.75^0$$
$$= 0.0197$$

The student has about a 2% chance of passing the quiz.

b. For the student in the previous question, how much lower would his chance of passing be if each question had five possible answers instead of four?

If there were five possible answers, then we would calculate the probability of passing by setting p=0.2 and q=0.8

$$P(PassingTheQuizWith5MultipleChoiceOptions) =$$
$$\frac{10!}{6!(4)!}0.2^6 0.8^4 + \frac{10!}{7!3!}0.2^7 0.8^3 + \frac{10!}{8!2!}0.2^8 0.8^2 + \frac{10!}{9!1!}0.2^9 0.8^1 + \frac{10!}{10!0!}0.2^{10} 0.8^0$$
$$= 0.00637$$

If there were five possible answers instead of four, the probability of passing the quiz would drop to 0.00637, under 1%. This is one-third of the probability of passing the quiz with four possible multiple-choice answers.

c. Using your answer from part (a) as the probability a randomly selected unprepared student passes using this method of guessing, imagine you have 100 such students taking the quiz. How many would you typically expect to pass this quiz out of 100?

For a binomial distribution, $\mu = np$ If I had 100 randomly-selected unprepared students taking the test, I would expect $100 \times 0.0197 = 1.97 \approx 2$ students to pass.

d. Would you recommend that the professor go through extra work to ensure that each question has five possible answers instead of just four? Why or why not?

I would make the recommendation based on the preferences of the professor and the size of the course. The probability of passing the quiz is very low whether there are 4 multiple-choice options or 5. With a relatively small class,i.e. 20 students, it is unlikely that more unprepared students will pass the quiz if the probability is 0.0197 (with 4 options) or 0.00637 (with 5 options). However, if the professor is running a massive online open course with 1000 students, then she can expect about 20 unprepared students to pass the 4-option quiz as opposed to about 6 students who would be expected to pass the 5-option quiz. The professor would not be happy with 14 additional unprepared students passing a quiz.

2. Worksite injuries at a large manufacturing plant occur at an average rate of 1.5 per month. Accidents typically occur independent of one another.

a. What is the probability that exactly three accidents occur in a given month?

For this problem, we will use a poisson probability distribution.

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

$$r = successes$$

$$\lambda = mean$$

$$\mu = \lambda, \sigma = \sqrt{\lambda}$$

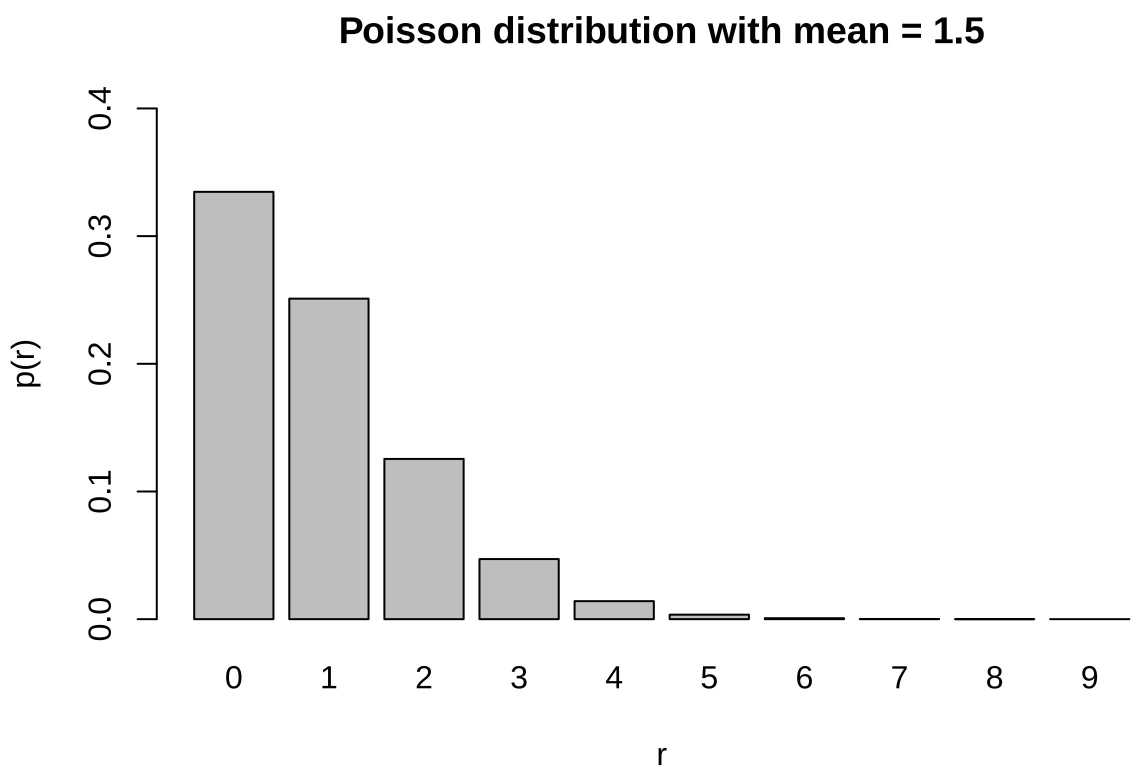
The probability of 3 accidents occurring in a given month is:

$$p(3) = \frac{e^{-1.5}1.5^3}{3!} = 0.126$$

b. Suppose you are looking out for months with unusually high numbers of accidents. How many accidents would have to occur in a month for you to take special notice?

If I was looking for outliers in the data, I would expect that 95% of accident totals by month would fall within two standard deviations of the mean. For the poisson distribution $\sigma = \sqrt{\lambda} = \sqrt{1.5} = 1.225$ Therefore, 95% of values should fall between $1.5 \pm 2(1.225)$, or between $0 \leq$ r ≤ 3.94. As we can see in this plot, the probability of accidents drops off considerably at r = 4. If I was looking at accident totals for 100 months of work history, I would be surprised if there were more than 5 months with more than 4 accidents.

```
ourDist <- c()
for (i in 0:10){ourDist[i] <- dpois(i,1.5)}
barplot(ourDist,ylim=c(0,0.4),names.arg=c(0:9),xlab="r",ylab="p(r)",
        main="Poisson distribution with mean = 1.5")
```



c. What is the probability that there are at most two accidents in a particular month?

The probability that $P(r \leq 2) = P(0) + P(1) + P(2) = 0.251 + 0.167 + 0.112 = 0.530$

3. For a particular population of river otters, the age (in months) is uniformly distributed between 0 months and 250 months. Imagine an otter is chosen at random from this population.

For this problem, the probability that an otter is x months old is $f(x) = \frac{1}{250-0} = 0.004$ for all $0 \leq x \leq 250$.

a. Is there any age you would find surprising? Explain.

I would be suprised if the otter was more than 250 months old. Any age less than or equal to 250 months has a uniform probability, but the probability of any value over 250 is equal to zero.

b. What is the probability that the otter is less than 24 months old?

The probability that an otter is less than 24 months old is:

$$P(x < 24) = \frac{23}{250} = 0.092$$

c. What is the expected age of the otter?

$$E[X] = \frac{b+a}{2} = \frac{250+0}{2} = 125 months$$

d. Which is more likely for the above population of river otters – that you select an otter less than one year old or that you select an otter more than 20 years old? Explain, giving the probability for each possibility.

$$P(LessThanOneYear) = P(x < 12) = \frac{11}{250} = 0.044$$

$$P(GreaterThan20Years) = P(x > 240) = \frac{250 - 240}{250} = 0.04$$

There is a slightly larger probability that the otter is less than one year old. This make sense intuitively if each age (by month) has an equal probability. There would be a greater likelihood of selecting an otter than is less than one year old (11 possible months) than one that is older than 20 years (10 possible months).

4. Consider a normal distribution with mean 80 and standard deviation 10 as shown below.

To find area and probability for this normal distribution, we have to convert the x-values to z-scores that correspond to a standard normal curve.

$$z = \frac{x - \mu}{\sigma}$$

a. What is the probability that an observation falls between the cutoffs of 65 and 75 (as shown in the image)?

$$z_{65} = \frac{65 - 80}{10} = -1.5$$

$$z_{85} = \frac{75 - 80}{10} = -0.5$$

$$P(65 \leq x \leq 75) = P(-1.5 \leq z \leq -0.5) = P(z \leq -0.5) - P(z \leq -1.5)$$
$$= 0.3085 - 0.0668 = 0.2417$$

b. What is the probability that an observation falls above the value of 92?

$$z_{92} = \frac{92 - 80}{10} = 1.2$$
$$P(x > 92) = P(z > 1.2) = 0.1151$$

c. What is the probability that an observation falls below the value of 68? (How does this compare to your previous answer?)

$$z_{68} = \frac{68 - 80}{10} = -1.2$$
$$P(x < 68) = P(z < -1.2) = 0.1151$$

The probability that an observation is greater than 92 is the same as the probability that it is less than 68.

d. What is the cutoff that separates the bottom 30 percent from the top 70 percent? (This is the same as asking what the 30th percentile is.)

The z-score that represenents the 30th percentile is z = -0.52. If we plug this back into x = μ + zσ we find the 80th percentile at x = 74.8.

e. What is the 80th percentile?

The z-score that represents the 80th percentile is z = 0.84. Using the same method, x = 88.4.

f. What are the cutoffs that contain the middle 60%? (That would be the 20th and 80th percentiles.)

The middle 60% would be found between $-0.84 \leq z \leq 0.84$

We know the upper bound is x = 88.4. The lower bound for z = - 0.84 is x = 71.6. The middle 60% is found between $71.6 \leq x \leq 88.4$

5. The arrival time (measured in minutes late) for a particular flight each day is normally distributed with a mean of 10 minutes late and a standard deviation of 20 minutes late. Answer the following questions:

a. What is the probability that a flight is between ten and twenty minutes late?

The probability that the flight is between 10 and twenty minutes late is $P(10 \leq x \leq 20)$.

$$z_{10} = \frac{10-10}{20} = 0$$
$$z_{20} = \frac{20-10}{20} = 0.5$$
$$P(10 \leq x \leq 20) = P(0 \leq z \leq 0.5) = P(z \leq 0.5) - P(z \leq 0)$$
$$= 0.6915 - 0.5 = 0.1915$$

b. What is the probability that the flight is more than an hour late?

The probability that the flight is more than an hour late is $P(x \geq 60)$

$$z_{60} = \frac{60-10}{20} = 2.5$$
$$P(x \geq 60) = P(z \geq 2.5) = 0.0062$$

c. What percentage of the time would you expect the arrival time to be negative? What does this mean?

I would expect the arrival time to be negative anytime x is less than 0. Strictly speaking, a negative arrival time would mean the plane arrived early, not late.

$$z_0 = \frac{0-10}{20} = -0.5$$

The percentage of time that flights arrive early will be the area under the standard normal curve to the left of z = -0.5. This is 0.3085. I would expect flights to arrive early about 31% of the time.

d. What is the 25th percentile for arrival times? What is the 75th percentile?

Using a table of areas under the standard normalcurve for z-values, we know that the 25th and 75h percentile are found at z = -0.68,0.68. If we plug those values into x = μ + zσ we find the x-values for each percentile.

25th percentile:

$$x = 10 - 0.68(20) = -3.6$$

75th percentile:

$$x = 10 + 0.68(20) = 23.6$$