Statistics for Computing

MA4413 Lecture 3A

Kevin O'Brien

Kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn Semester 2013

- Binomial Coefficients / The Choose Operator
- Definition: The Probability Mass Functions (pmf)
- Binomial Distribution : Example

Binomial Coefficients

In the last class, we came across binomial coefficients. Informally, binomial coefficients are the number of ways k items can be selected from a group of n items. The binomial coefficient indexed by n and k is usually written as ${}^{n}C_{k}$ or

$$\binom{n}{k}$$

. C is colloqially known as the "choose operator".

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

(We call the operator the choose operator. We will use both notations interchangeably.)

Binomial Coefficients

- $n!$ and $k!$ are the coefficients of n and k respectively.
- $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$
- For example $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
- $n! = n \times (n-1)!$
- Importantly $0! = 1$ not 0.

$$\binom{6}{2} = \frac{6!}{2! \times (6-2)!} = \frac{6!}{2! \times 4!}$$

$$= \frac{6 \times 5 \times 4!}{2! \times 4!} = 30/2 = 15$$

More examples of Binomial coefficients on blackboard.

Probability Mass Function

(Formally defining something mentioned previously)

- a probability mass function (pmf) is a ***function*** that gives the probability that a discrete random variable is exactly equal to some value.

$$P(X = k)$$

- The probability mass function is often the primary means of defining a discrete probability distribution
- It is conventional to present the probability mass function in the form of a table.
- The p.m.f of a value k is often denoted $f(k)$.

Binomial Example 1

(Revision from Last Class)
Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?
Solution: This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167.

Therefore, the binomial probability is:

$$P(X=2) = {}^5C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

Binomial Example 2

Suppose there is a container that contains 6 items. The probability that any one of these items is defective is 0.3. Suppose all six items are inspected.

- What is the probability of 3 defective components?
- What is the probability of 4 defective components?

$$P(3 \text{ defects}) = f(3) = P(X=3) = \binom{6}{3} 0.3^3 (1-0.3)^{6-3} = 0.1852$$

$$P(4 \text{ defects}) = f(4) = P(X=4) = \binom{6}{4} 0.3^4 (1-0.3)^{6-4} = 0.0595$$

Probability Tables

In the **Sulis** workspace there are two important tables used for this part of the course.
This class will feature a demonstration on how to read those tables.

- The Cumulative Binomial Tables (Murdoch Barnes Tables 1)
- The Cumulative Poisson Tables (Murdoch Barnes Tables 2)

Please get a copy of each as soon as possible.

Probability Tables

- For some value r the tables record the probability of $P(X \geq r)$.
- The Student is required to locate the appropriate column based on the parameter values for the distribution in question.
- A copy of the Murdoch Barnes Tables will be furnished to the student in the End of Year Exam. The Tables are not required for the first mid-term exam.
- Knowledge of the sample space, partitioning of the sample points, and the complement rule are advised.

Binomial Distribution : Using Tables

It is estimated by a particular bank that 25% of credit card customers pay only the minimum amount due on their monthly credit card bill and do not pay the total amount due. 50 credit card customers are randomly selected.

1. (3 marks) What is the probability that 9 or more of the selected customers pay only the minimum amount due?
2. (3 marks) What is the probability that less than 6 of the selected customers pay only the minimum amount due?
3. (3 marks) What is the probability that more than 5 but less than 10 of the selected customers pay only the minimum amount due?

Binomial Distribution : Using Tables

Demonstration on Blackboard re: how to use tables in class.

1. $P(X \geq 9) = 0.9084$
2. $P(X < 6) = 1 - P(X \geq 6) = 1 - 0.9930 = 0.0070$
3. $P(6 \leq X \leq 9) = P(X \geq 6) - P(X \geq 10) = 0.9930 - 0.8363 = 0.1567$

Binomial Distribution: Expected Value and Variance

If the random variable X has a binomial distribution with parameters n and p, we write

$$X \sim B(n,p)$$

Expectation and Variance If $X \sim B(n,p)$, then:

- Expected Value of X : $E(X) = np$
- Variance of X : $Var(X) = np(1-p)$

Suppose n=3 and p=0.5 Then $E(X) = 1.5$ and $V(X) = 0.75$.
Remark: Referring to the expected value and variance may be used to validate the assumption of a binomial distribution.

The Geometric Distribution

- The Geometric distribution is related to the Binomial distribution in that both are based on independent trials in which the probability of success is constant and equal to p.
- However, a Geometric random variable is the number of trials until the first failure, whereas a Binomial random variable is the number of successes in n trials.
- The Geometric distributions is often used in IT security applications.

The Geometric Distribution

Suppose that a random experiment has two possible outcomes, success with probability p and failure with probability 1-p .
The experiment is repeated until a success happens. The number of trials before the success is a random variable X computed as follows

$$P(X = k) = (1 - p)^{(k-1)} \times p$$

(i.e. The probability that first success is on the k-th trial)

The Geometric Distribution: Notation

If X has a geometric distribution with parameter p, we write

$$X \sim Geo(p)$$

Expectation and Variance If $X \sim Geo(p)$, then:

- Expected Value of X : E(X) = 1/p
- Variance of X : Var(X) = $(1-p)/p^2$.

Poisson Experiment

A Poisson experiment is a statistical experiment that has the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (m) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the region.
- The probability that a success will occur in an extremely small region is virtually zero.

Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

Poisson Probability Distribution

- A Poisson random variable is the number of successes that result from a Poisson experiment.
- The probability distribution of a Poisson random variable is called a Poisson distribution.

The Poisson Probability Distribution

- The number of occurrences in a unit period (or space)
- The expected number of occurrences is m
- Given the mean number of successes (m) that occur in a specified region, we can compute the Poisson probability based on the following formula (next slide).

Poisson Formulae

The probability that there will be k occurrences in a unit time period is denoted $P(X=k)$, and is computed as follows.

$$P(X=k) = \frac{m^k e^{-m}}{k!}$$

Poisson Formulae

Given that there is on average 2 occurrences per hour, what is the probability of no occurrences in the next hour?

i.e. Compute $P(X=0)$ given that $m=2$

$$P(X=0) = \frac{2^0 e^{-2}}{0!}$$

- $2^0 = 1$
- $0! = 1$

The equation reduces to

$$P(X=0) = e^{-2} = 0.1353$$

Poisson Formulae

What is the probability of one occurrences in the next hour?

i.e. Compute $P(X=1)$ given that $m=2$

$$P(X=1) = \frac{2^1 e^{-2}}{1!}$$

- $2^1 = 2$
- $1! = 1$

The equation reduces to

$$P(X=1) = 2 \times e^{-2} = 0.2706$$

The Cumulative Distribution Function

- The Cumulative Distribution Function, denoted $F(x)$, is a common way that the probabilities of a random variable (both discrete and continuous) can be summarized.
- The Cumulative Distribution Function, which also can be described by a formula or summarized in a table, is defined as:

$$F(x) = P(X \leq x)$$

- The notation for a cumulative distribution function, F(x), entails using a capital "F". (The notation for a probability mass or density function, f(x), i.e. using a lowercase "f". The notation is not interchangeable.

Useful Results

(Demonstration on the blackboard re: partitioning of the sample space, using examples on next slide)

- $P(X \leq 1) = P(X = 0) + P(X = 1)$
- $P(X \leq r) = P(X = 0) + P(X = 1) + \ldots P(X = r)$
- $P(X \leq 0) = P(X = 0)$
- $P(X = r) = P(X \geq r) - P(X \geq r + 1)$
- **Complement Rule**: $P(X \leq r - 1) = P(X < r) = 1 - P(X \geq r)$
- **Interval Rule**:$P(a \leq X \leq b) = P(X \geq a) - P(X \geq b + 1)$.

For the binomial distribution, if the probability of success is greater than 0.5, instead of considering the number of successes, to use the table we consider the number of failures.

Binomial Example 1

Suppose a signal of 100 bits is transmitted and the probability of sending a bit correctly is 0.9. What is the probability of

1. at least 10 errors
2. exactly 7 errors
3. Between 5 and 15 errors (inclusively).

Binomial Example 1

- Since the probability of success is 0.9. We consider the distribution of the number of failures (errors).
- We reverse the definition of 'success' and 'failure'. Success is now defined as an error.
- The probability that a bit is sent incorrectly is 0.1.
- Let X be the total number of errors. $X \sim B(100, 0.1)$.
- Answer : $P(X \geq 10) = 0.5487$.
- $P(X = 7) = P(X \geq 7) - P(X \geq 8) = 0.8828 - 0.7939 = 0.0889$.
- $P(5 \leq X \leq 15) = P(X \geq 5) - P(X \geq 16) = 0.9763 - 0.0399 = 0.9364$

The Poisson Probability Distribution

- A Poisson random variable is the number of successes that result from a Poisson experiment.
- The probability distribution of a Poisson random variable is called a Poisson distribution.
- Very Important: This distribution describes the number of occurrences in a ***unit period (or space)***
- Very Important: The expected number of occurrences is m

The Poisson Probability Distribution

We use the following notation.

$$X \sim Poisson(m)$$

Note the expected number of occurrences per unit time is conventionally denoted λ rather than m. As the Murdoch Barnes cumulative Poisson Tables

(Table 2) use m, so shall we. Recall that Tables 2 gives values of the probability $P(X \geq r)$, when X has a Poisson distribution with parameter m.

The Poisson Probability Distribution

Consider cars passing a point on a rarely used country road. Is this a Poisson Random Variable? Suppose

1. Arrivals occur at an average rate of m cars per unit time.
2. The probability of an arrival in an interval of length k is constant.
3. The number of arrivals in two non-overlapping intervals of time are independent.

This would be an appropriate use of the Poisson Distribution.

Changing the unit time.

- The number of arrivals, X, in an interval of length t has a Poisson distribution with parameter $\mu = mt$.
- m is the expected number of arrivals in a unit time period.
- μ is the expected number of arrivals in a time period t, that is different from the unit time period.
- Put simply : if we change the time period in question, we adjust the Poisson mean accordingly.
- If 10 occurrences are expected in 1 hour, then 5 are expected in 30 minutes. Likewise, 20 occurrences are expected in 2 hours, and so on.
- (Remark : we will not use μ in this context anymore).

Poisson Example

A motor dealership which specializes in agricultural machinery sells one vehicle every 2 days, on average. Answer the following questions.

1. What is the probability that the dealership sells at least one vehicle in one particular day?
2. What is the probability that the dealership will sell exactly one vehicle in one particular day?
3. What is the probability that the dealership will sell 4 vehicles or more in a six day working week?

Poisson Example

1. Expected Occurrences per Day: m = 0.5
2. Probability that the dealership sells at least one vehicle in one particular day?

$$P(X \geq 1) = 0.3935$$

3. Probability that the dealership will sell exactly one vehicle in one particular day?

$$P(X = 1) = P(X \geq 1) - P(X \geq 2) = 0.3935 - 0.0902 = 0.3031$$

4. Probability that the dealership will sell 4 vehicles or more in a six day working week?
    - For a 6 day week, m=3
    - $P(X \geq 4) = 0.3528$

Knowing which distribution to use

- For the end of semester examination, you will be required to know when it is appropriate to use the Poisson distribution, and when to use the binomial distribution.
- Recall the key parameters of each distribution.
- Binomial : number of ***successes*** in n ***independent trials***.
- Poisson : number of ***occurrences*** in a ***unit space***.