Statistics for Computing

MA4413 Lecture 4B

Kevin O'Brien

Kevin.obrien@ul.ie

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

Autumn Semester 2012

Today's Class

- Review of Discrete Probability Distributions
- R Implementation
- A Few Examples
- Introduction to Continuous Probability Distributions
- The Uniform Distribution
- The Exponential Distribution

Discrete Probability Distributions

Three main distributions

- Binomial Distribution
- Poisson Distribution
- Geometric Distribution (mentioned, but not as important as the other two.)

Binomial Probability Distribution

Important Points:

- The experiment is a series of n independent trials.
- Two possible outcomes from each trial: a success and a failure.
- The probability of success (i.e. p) is constant.
- A binomial random variable can be written as

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

Poisson Probability Distribution

Important Points:

- This distribution is concerned with the number of occurrences per unit space.
- Unit space can mean a unit length, a unit area, a unit volume or a unit period of time.
- We will concern ourselves with unit time periods mostly.
- A Poisson random variable can be written as

$$X \sim Pois(m)$$

- The Poisson distribution can be used to approximate the binomial distribution under certain conditions.

PDFs and CDFs

Important Points:

- The probability density function (PDF) is the probability of a random variable taking a specific value i.e. $P(X = k)$
- The appropriate `R` functions are `dbinom` and `dpois`
- The cumulative distribution function (CDF) is the probability of a random variable not exceeding a specific value i.e. $P(X \leq k)$
- The appropriate `R` functions are `pbinom` and `ppois`

Geometric Probability Distribution

Important Points:

- This distribution is closely related to the binomial distribution.
- This distribution described the number of failures that occur before the first success, when the probability of success is p.
- The relevant `R` functions are `dgeom` and `pgeom`.

Geometric Probability Distribution : Example

If the probability of inserting a USB correctly is 0.40, what is the probability of successfully doing so on the second attempt. . In essence we have one

failure, then one success, and these are independent events. So the probability the second attempt will be successful is 0.6×0.4. The probability that we are successful on the first attempt (i.e. no failures beforehand) is 0.4

```
> dgeom(0,prob=0.4)
[1] 0.4
> dgeom(1,prob=0.4)
[1] 0.24
> dgeom(2,prob=0.4)
[1] 0.144
```

Cumulative Binomial Probability Distribution Tables

- Consider a binomial experiment with $n = 20$ and $p = 0.50$.
- First we have to find which section of the tables which tabulates the correspond binomial probabilities for $n = 20$ and $p = 0.50$
- Lets use the tables to compute $P(X \geq 10)$
- Remark
  $P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12) + \ldots + P(X = 20)$

Cumulative Binomial Probability Distribution Tables

Section of tables for $n = 20$ and $p = (0.10, 0.15, \ldots, 0.50)$

p=	...	0.45	**0.50**
r=0	...	1.0000	1.0000
1	...	1.0000	1.0000
2	...	0.9999	1.0000
...	...	...	...
9	...	0.5857	0.7483
10	...	0.4086	**0.5881**
11	...	0.2493	0.4119
...	...	...	...

Cumulative Binomial Probability Distribution Tables

- Consider a binomial experiment with $n = 25$ and $p = 0.50$.
- Suppose we wish to compute $P(X \geq 10)$
- Is there a section of the tables which tabulates the correspond binomial probabilities for $n = 25$ and $p = 0.50$
- No! We would not be able to use the Murdoch Barnes tables to compute this.
- Of course, we could more detailed set of tables, but that will not part of this module. For all problems using **cumulative** probabilities will be restricted to those that can be solved using the Murdoch Barnes tables.

Cumulative Binomial Probability Distribution Tables

- Consider a binomial experiment with $n = 20$ and $p = 0.50$.
- Suppose we wish to compute the probability of 9 successes or less, i.e. $P(X \leq 9)$
- The tables does not tabulate probabilities in such a way. Recall that it tabulates probabilities for r successes **or more**.
- However, we can still use the table to compute the desired probability.
- Consider the sample space for the number of successes. There can be between 0 and 20 successes.

$$S = \{0, 1, 2, 3, \ldots, 8, 9, 10, 11, 12 \ldots, 19, 20\}$$

Cumulative Binomial Probability Distribution Tables

- What are the sample points for the event where the number of success is less than or equal to 9? (Lets call this event A.)

$$A = \{0, 1, 2, 3, \ldots, 8, 9\}$$

- What are the sample points for the ***complement event***. (Lets call this event A^c).

$$A^c = \{10, 11, 12 \ldots, 19, 20\}$$

- Can we compute the probability of event A^c?
  Yes, it is $P(X \geq 10)$
- The solution is therefore

$$P(X \leq 9) = 1 - P(X \geq 10) = 1 - 0.5881 = 0.4119$$

Cumulative Binomial Probability Distribution Tables

- What is the probability that the number of successes is exactly 10 (with $n = 20, p = 0.50$)?
- i.e. $P(X = 10) = ?$
- Again, the tables does not tabulate probabilities in such a way.
- It would also be quite cumbersome to compute $P(X = 10)$ using the binomial probability formula.
- In the previous question, we found out the probability of 10 or more successes.

Cumulative Binomial Probability Distribution Tables

- Recall that
  $P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12) + \ldots + P(X = 20)$.
- Equivalently $P(X \geq 11) = P(X = 11) + P(X = 12) + \ldots + P(X = 20)$.
- Therefore $P(X \geq 10) = P(X = 10) + P(X \geq 11)$.
- Re-arranging this expression we get
  $P(X = 10) = P(X \geq 10) - P(X \geq 11)$.
- From the tables $P(X \geq 11) = 0.4119$.
- Therefore $P(X = 10) = 0.5881 - 0.4119) = 0.1762$.

Cumulative Binomial Probability Distribution Tables

Section of tables for $n = 20$ and $p = (0.10, 0.15, \ldots, 0.50)$

p=	...	0.45	**0.50**
r=0	...	1.0000	1.0000
1	...	1.0000	1.0000
2	...	0.9999	1.0000
...	...	...	...
10	...	0.4086	0.5881
11	...	0.2493	**0.4119**
12	...	0.1308	0.2517
...	...	...	...

Cumulative Binomial Probability Distribution Tables

The vice-president of a computer firm has reviewed the records of the firm's personnel and has found that 70% of the employees read a well known industry magazine "The IT Journal".

If the vice-president was to choose 10 employees at random, what is the probability that the number of these employees who do not read the "IT Journal" is the following?

1. At least five.
2. Between four and eight, inclusive.
3. No more than seven.

Solution :

- Firstly, identify the probability distribution to be used?
  - Answer: the binomial distribution
- We are given the number of trials (“ choose 10 employees”)
- We are given a definition of a “success”, which is finding an employee that did NOT reads the journal.
- We are given the probability of such a success : 30% or 0.30
- So our binomial parameters are n= 10 and p = 0.30
- Let’s use the Murdoch Barnes Table 1 and find the relevant columns.

Cumulative Binomial Probability Distribution Tables

Section of tables for $n = 10$ and $p = (0.10, 0.15, \ldots, 0.30, \ldots, 0.50)$

Part 1: compute $P(X \geq 5)$
Answer: $P(X \geq 5) = 0.1503$

p=	...	0.25	**0.30**	...
r=0	...	1.0000	1.0000	...
1	...	0.9437	0.9718	...
2	...	0.7560	0.8507	...
...	...	...	...	...
5	...	0.0781	**0.1503**	...
...	...	...	...	...

Cumulative Binomial Probability Distribution Tables

Part 2: Compute $P(4 \leq X \leq 8)$

- $P(4 \leq X \leq 8) = P(X=4) + P(X=5) + P(X=6) + P(X=7) + P(X=8)$
- $P(X \geq 4) = P(X=4) + \ldots + P(X=8) + P(X=9) + P(X=10)$
- $P(X \geq 9) = P(X=9) + P(X=10)$
- $P(X \geq 4) = P(4 \leq X \leq 8) + P(X \geq 9)$
- Re-arranging $P(4 \leq X \leq 8) = P(X \geq 4) - P(X \geq 9)$

Cumulative Binomial Probability Distribution Tables

Part 2: compute $P(4 \leq X \leq 8)$
Answer: $P(4 \leq X \leq 8) = 0.3504 - 0.0001 = 0.3503$

p=	...	0.25	**0.30**	...
r=0	...	1.0000	1.0000	...
1	...	0.9437	0.9718	...
...	...	...	...	...
4	...	0.2241	**0.3504**	...
...	...	...	...	...
9	...		**0.0001**	...
...	...	...	...	...

Cumulative Binomial Probability Distribution Tables

Part 3: Compute $P(X \leq 7)$

- From before, this is the complement of 8 or more successes
- i.e. $P(X \leq 7) = 1 - P(X \geq 8)$
- Determine $P(X \geq 8)$ and subtract that value from 1.

Cumulative Binomial Probability Distribution Tables

Part 3: compute $P(X \leq 7)$
Answer: $P(X \leq 7) = 1 - 0.0016 = 0.9984$

p=	...	0.25	**0.30**	...
r=0	...	1.0000	1.0000	...
1	...	0.9437	0.9718	...
...	...	...	...	...
7	...	0.0.0035	0.0106	...
8	...	0.0004	**0.0016**	...
...	...	...	...	...

Cumulative Poisson Probability Distribution Tables

- The Murdoch Barnes Table set 2 ("cumulative Poisson probabilities") tabulates cumulative values for Poisson probabilities.
- For some value r, these tables give the probability for r or more occurrences (i.e $P(X \geq r)$) for some value m, the expected number of occurences per unit period.
- To use the tables, the correct column for m must be found.
- The cumulative Poisson tables is easier to use, compared to the cumulative binomial tables.
- All of the problem solving approaches we learned for the cumulative binomial tables, also apply for the cumulative Poisson tables.

Cumulative Poisson Probability Distribution Tables

m=	...	0.40	0.50
r=0	...	1.0000	1.0000
1	...	0.3297	0.3935
2	...	0.0616	0.0902
3	...	0.0079	0.0144
...	...	...	...

Cumulative Poisson Probability Distribution Tables

- Suppose that for a Poisson random variable that mean number of occurrences per minute was 0.5.
- Compute the probability that there will more than one occurrence in a given minute.
- Solution : $P(X \geq 1) = 0.3935$

Cumulative Poisson Probability Distribution Tables

$P(X \geq 1) = 0.3935$

m=	...	0.40	**0.50**
r=0	...	1.0000	1.0000
1	...	0.3297	**0.3935**
2	...	0.0616	0.0902
3	...	0.0079	0.0144
...	...	...	...