Statistics for Computing

MA4413 Lecture 4A

Kevin O'Brien

Kevin.obrien@ul.ie

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

Autumn Semester 2012

The Binomial Probability Distribution

A Quick Review of the Binomial Distribution

- The number of independent trials is denoted n.
- The outcome of interest is known as a "Success".
- The other outcome is known as a "failure".
- Often the applications of these names is counter-intuitive, i.e. defective components being the "success".
- The probability of a 'success' is p
- The expected number of 'successes' from n trials is $E(X) = np$
- The `binom` family of commands in `R` are what we use to compute necessary values.

Characteristics of a 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.
- The `pois` family of functions are used to compute probabilities and quantiles.

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.

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.
- This distribution describes the number of occurrences in a unit period (or space)
- The expected number of occurrences is m.
- R refers to the mean number of occurrences as `lambda` rather than `m`.

Poisson Formulae

The probability that there will be k occurrences in a unit time period is denoted $P(X = k)$, and is computed as below. Remark: This is known as the probability density function. The corresponding `R` command is `dpois()`.

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

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

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

Poisson Distribution (Example)

- Suppose that electricity power failures occur according to a Poisson distribution with an average of 2 outages every twenty weeks.
- Calculate the probability that there will not be more than one power outage during a particular week.

Solution:

- The average number of failures per week is: $m = 2/20 = 0.10$
- "Not more than one power outage" means we need to compute and add the probabilities for "0 outages" plus "1 outage".

Poisson Distribution (Example)

Recall:

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

- $P(X = 0)$

$$P(X = 0) = e^{-0.10} \frac{0.10^0}{0!} = e^{-0.10} = 0.9048$$

- $P(X = 1)$

$$P(X = 1) = e^{-0.10} \frac{0.10^1}{1!} = e^{-0.10} \times 0.1 = 0.0905$$

- $P(X \leq 1)$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.9048 + 0.0905 = 0.995$$

Implementation using R

- Probability Density Function $P(X = k)$
  - For a given poisson mean m, which in R is specified as `lambda`
  - `dpois(k,lambda = ...)`
- Cumulative Density Function $P(X \leq k)$
  - `ppois(k,lambda = ...)`

Implementation using R

From before: $P(X = 0)$ given than the mean number of occurrences is 2.

```
> dpois(0,lambda=2)
[1] 0.1353353
> dpois(1,lambda=2)
[1] 0.2706706
> dpois(2,lambda=2)
[1] 0.2706706
```

Implementation using R

Compute the cumulative distribution functions for the values $k = \{0, 1, 2\}$, given that the mean number of occurrences is 2

```
> ppois(0,lambda=2)
[1] 0.1353353
> ppois(1,lambda=2)
[1] 0.4060058
> ppois(2,lambda=2)
[1] 0.6766764
```

Poisson Approximation of the Binomial

- The Poisson distribution can sometimes be used to approximate the binomial distribution
- When the number of observations n is large, and the success probability p is small, the $\text{Bin}(n,p)$ distribution approaches the Poisson distribution with the parameter given by $m = np$.
- This is useful since the computations involved in calculating binomial probabilities are greatly reduced.
- As a rule of thumb, n should be greater than 50 with p very small, such that np should be less than 5.
- If the value of p is very high, the definition of what constitutes a "success" or "failure" can be switched.

Poisson Approximation: Example

Suppose we sample 1000 items from a production line that is producing, on average, 0.1% defective components.

Using the binomial distribution, the probability of exactly 3 defective items in our sample is

$$P(X = 3) = ^{1000}C_3 \times (0.001)^3 \times 0.999^{997}$$

Poisson Approximation: Example

Lets compute each of the component terms individually.

- ${}^{1000}C_3$

$$ {}^{1000}C_3 = \frac{1000 \times 999 \times 998}{3 \times 2 \times 1} = 166,167,000 $$

- 0.001^3

$$ 0.001^3 = 0.000000001 $$

- 0.999^{997}

$$ 0.999^{997} = 0.36880 $$

Multiply these three values to compute the binomial probability

$$ P(X = 3) = 0.06128 $$

Poisson Approximation: Example

- Lets use the Poisson distribution to approximate a solution.
- First check that $n \geq 50$ and $np < 5$ (Yes to both).
- We choose as our parameter value $m = np = 0.001 \times 1000 = 1$

$$P(X=3) = e^{-1}\frac{1^3}{3!} = \frac{e^{-1}}{6} = \frac{0.36787}{6} = 0.06131$$

- Compare this answer with the Binomial probability $P(X=3) = 0.06128$.
- Very good approximation, with much less computation effort.

Implementation using R

```
> # Poisson Mean m = 1000 * 0.001 = 1
> dbinom(3,size=1000,prob=0.001)
[1] 0.06128251
>
> dpois(3,lambda=1)
[1] 0.06131324
```