Statistics for Computing

MA4413 Lecture 4A

Kevin O'Brien

Kevin.obrien@ul.ie

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

Autumn Semester 2011

The Binomial Probability Distribution

- The number of independent trials is denoted n.
- The probability of a 'success' is p
- The expected number of 'successes' from n trials is $E(X) = np$

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.

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 Distribution

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

Poisson Formulae

The probability that there will be k occurences 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 occurences 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

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

Continuous Random Variables

- Probability Density Function
- Cumulative Density Function

If X is a continuous random variable then we can say that the probability of obtaining a **precise** value x is infinitely small, i.e. close to zero.

$$P(X = x) \approx 0$$

Consequently, for continuous random variables (only), $P(X \leq x)$ and $P(X < x)$ can be used interchangeably.

$$P(X \leq x) \approx P(X < x)$$

Continuous Uniform Distribution

A random variable X is called a continuous uniform random variable over the interval (a,b) if it's probability density function is given by

$$f_X(x) = \frac{1}{b-a} \qquad \text{when } a \leq x \leq b$$

The corresponding cumulative density function is

$$F_x(x) = \frac{x-a}{b-a} \qquad \text{when } a \leq x \leq b$$

The mean of the continuous uniform distribution is

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

The Memoryless property

The most interesting property of the exponential distribution is the ***memoryless*** property. By this , we mean that if the lifetime of a component is exponentially distributed, then an item which has been in use for some time is a good as a brand new item with regards to the likelihood of failure.
The exponential distribution is the only distribution that has this property.

Random Variables

A pair of dice is thrown. Let X denote the minimum of the two numbers which occur. Find the distributions and expected value of X.

Random Variables

A fair coin is tossed four times. Let X denote the longest string of heads. Find the distribution and expectation of X.

Random Variables

A fair coin is tossed until a head or five tails occurs. Find the expected number E of tosses of the coin.

Random Variables

A coin is weighted so that $P(H) = 0.75$ and $P(T) = 0.25$
The coin is tossed three times. Let X denote the number of heads that appear.

- (a) Find the distribution f of X.
- (b) Find the expectation $E(X)$.

- Now consider an experiment with only two outcomes. Independent repeated trials of such an experiment are called Bernoulli trials, named after the Swiss mathematician Jacob Bernoulli (16541705).
- The term ***independent trials*** means that the outcome of any trial does not depend on the previous outcomes (such as tossing a coin).
- We will call one of the outcomes the "success" and the other outcome the "failure".

- Let p denote the probability of success in a Bernoulli trial, and so $q = 1 - p$ is the probability of failure. A binomial experiment consists of a fixed number of Bernoulli trials.
- A binomial experiment with n trials and probability p of success will be denoted by

$$B(n,p)$$

Probability Mass Function

- a probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value.
- The probability mass function is often the primary means of defining a discrete probability distribution

Thirty-eight students took the test. The X-axis shows various intervals of scores (the interval labeled 35 includes any score from 32.5 to 37.5). The Y-axis shows the number of students scoring in the interval or below the interval.

cumulative frequency distribution A can show either the actual frequencies at or below each interval (as shown here) or the percentage of the scores at or below each interval. The plot can be a histogram as shown here or a polygon.