Probability Distributions

Random variables
Probability Distributions
Bernoulli Trials
The Common Distributions:

- Bernoulli
- Binomial
- Geometric
- Poisson

Random Variables

A *Random Variable* is the numerical outcome of a random process. Technically it is the mapping from the set out outcomes to subsets of the real line.

Many outcomes are numerical anyway. If the outcome is not numerical we can code it to be so:

Ex Coin = {H,T} we can code H as 1, T as 0.

{Male, Female} Male =0 ,Female =1.

{Bad, Average, Good} as 1,2,3 etc.

The purpose of doing this is so that we can define quantities such as:

X is a random variable.

P(X<x), Average value of X etc.

Further the advantage of doing this is that if the coding is such that these quantities are sensible, (such as the 1 and 0 above) we can regard them as properties of (0,1) = *binary* variables. These have the same meaning irrespective of the underlying process.

Sometimes the coding is meaningless, then so are some of the quantities but that doesn't matter as long as we take this into account.

Probability Distributions

The *probability distribution of X* is a list of all possible values of X and their associated probabilities – (at least for now)

p(x)=P(X=x); the probability that X=x

The list might be a simple list :

Example: No of PCs sold in a retail outlet on a given day:

PCs **X**	**Probability** **p(x)**
2	.2
3	.3
4	.25
5	.15
6	.1

For completeness we should add:

0 or 1	0
more than 6	0

More frequently we shall be dealing with a situation where the probabilities are given as a formula.

X is the number of goals scored by a team during its matches.

$$p(x) = e^{-2}\frac{2^x}{x!} \quad x=0,1,2\ldots$$

This is the Poisson distribution with parameter 2.

For discrete distributions we have:

P(X=x) = p(x)

For the distribution to be valid we need:

$0 \le p(x) \le 1$ as these are probabilities,

and

$\sum_{all\ x} p(x) = 1$ As all possible events X=x are included; the total probability must be 1.

Otherwise there are no restrictions any such list defines a probability model.

Standard distributions

These correspond to models possessing particular properties that are frequently encountered in practice.

<u>The Bernoulli Trial</u>

Any experiment with only two outcomes

win/lose a game,
succeed/fail an exam,
Bit 1 or 0
head/tail of a coin

is called a *Bernoulli Trial*.

For convenience, take outcomes to be 1 (= win, succeed, head) or 0 (= lose, fail, tail).

Let p = P(1) (called "Probability of Success")

Probability distribution of a Bernoulli trial is

$$P(x) = \begin{cases} p, & \text{if } x = 1 \\ 1 - p, & \text{if x} = 0 \end{cases}$$

The Bernoulli trial is a model for a *single* outcome of any binary process:

As a single outcome the B-trial is not very informative; however sequences and sets of B-trials are interesting.

Binomial Distribution

Example: Conduct a sequence of 3 B-trials (0,1) with P(X=1)=p

The are 8 possible outcomes:

	000	001 010,100	011 101 110	111
Y=	0	1	2	3

Y is the number of 1s in the 3 trials.

Also note $Y = \sum_{i=1}^{3} X_i$

Assume that the trials are independent

So that:

$$P(111) = p \times p \times p = p^3$$

$$P(011) = (1 - p) \times p \times p = (1 - p) p^2 = P(101) = P(110)$$

$$P(001) = (1 - p)^2 p = P(010) = P(100)$$

and finally:

$$P(000) = (1 - p)^3$$

Thus:

$$P(Y = 3) = p^3$$

$$P(Y = 2) = P(110) + P(101) + P(011) = 3p^2 (1 - p)$$

$$P(Y = 1) = 3p(1 - p)^2$$

$$P(Y = 0) = (1 - p)^3$$

Y follows the *Binomial Distribution*

Binomial gives *the prob of number of successes from a set of Bernoulli trials*

It has two parameters:

n – the number of trials,

p – the probability of success in each trial.

The trials must be independent.

In general we have n-trials , with y 1s (successes) the probability of observing such a sequence is:

$$p^y(1-p)^{n-y}$$

Thus

$$P(Y=y)=C(y,n)p^y(1-p)^{n-y}$$

Where C(y,n) is the number of ways a sequence of y 1s and (n-y) 0s can be ordered.

$C(y,n)=\binom{n}{y}=\dfrac{n!}{y!(n-y)!}$ ie choosing y from n.

What we are choosing here are the positions for the 1s. Each 1 can be in any position (in the sequence) from 1 to n. Choose the y positions for the 1s fill the rest of the sequence with 0s and the sequence is defined.

$$P(Y=y)=\binom{n}{y}p^y(1-p)^{n-y} \qquad 0 \le \mathrm{y} \le \mathrm{n}, 0 \le \mathrm{p} \le 1$$

The binomial is abbreviated to B(n,p).

Examples:

Ex1.

Circuit boards produced by an assembly line have a probability of 0.04 being defective. The probability that 2 are defective in a batch of 10 CBs is

P(X=2)=

$$\binom{10}{2} * 0.04^2 * 0.96^8 = 45 * 0.04^2 * 0.96^8 = 0.05194$$

The Excel function :

BINOMDIST(x,n,p,0) returns P(X=x).

The function BINOMDIST(x,n,p,1) returns $P(X \le x)$.

Actually the last variable is FALSE =(0) ,TRUE =(1)

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) =$$
$$BINOMDIST(2,10,0.04,1) = 0.993786.$$

Binomial probabilities are computed by hand in only the simplest cases .

Ex 2

The example considered in lecture 1 – the number that PCs that crash during the term.

If we assume that each PC (with a given operating system) has the same chance of crashing during the term.

And

That each PC crashes or not independently of other PCs

Then the number of PCs that crash is Binomially distributed.

For a comparison of reliability we assume that
X_w the number of Windows PCs that crash is Binomial with N=742 and p = p_w

X_m the number of Mac OS PCs that crash is Binomial with N=733 and p = p_m

Comparing the reliabilities of the two reduces to answering the question:

Is $p_w = p_m$?

Later we shall see how to do this.

The Binomial model also could serve for the bugs in software data.

At stage i there are N_i bugs the probability that a particular bug is uncovered during stage i is p , independently of other bugs.

So the number of bugs discovered during stage i is
$X_i \quad B(N_i, p)$

Also we have $N_{i+1} = N_i - X_i$

More sophisticated models would change p from stage to stage, introduce non-independence of bugs, possibility of introducing new bugs – but the above is a good starting point.

Ex 3

A bridge hand consist of 13 cards. 1/4th of the cards in the pack are spades (p=0.25). When dealt the cards are randomly selected. The probability of the hand containing exactly 6 spades can be computed from the Bin ….

NO!

We have 13 binary trials, the absolute probability on each trial is in fact 1/4. But the trials are NOT independent.

Consider just the first two trials: 1 = spade, 0=other suit.

$$P(X_1 = 1) = \frac{13}{52} = \frac{1}{4}$$

but

$$P(X_2 = 1 \mid X_1 = 1) = \frac{12}{51} \text{ and } P(X_2 = 1 \mid X_1 = 0) = \frac{13}{51}$$

So the trials are not independent !

Note by Partition law:

$$P(X_2 = 1) = \frac{1}{4} \times \frac{12}{51} + \frac{3}{4} \times \frac{13}{51} = \frac{51}{4 \times 51} = \frac{1}{4}.$$

However if we deal a hand of 13 from 10 packs shuffled together (520 cards 130 spades) the probability is nearly BINOMDIST(6,13,0.25,0). With 100 packs (5200 cards 1300 spades) for practical purposes we may use P(X=6)=BINOMDIST(6,13,0.25,0)

Although there is still dependence its effects are so small that they can be ignored.

A rule of thumb is that if the selection size is less than 1% of the population size we may use the Binomial.

If the population is finite (small) the Hypergeometric distribution is used:
N=population size
R – number in population with the characteristic.
n- sample (selection) size
x – number with characteristic in sample.

W/o proof

$$P(X = x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$$

We can compute the probability that we get 6 spades:

$$\frac{\binom{13}{6}\binom{39}{7}}{\binom{52}{13}} = HYPERGEOMDIST(x,n,R,N) = 0.041564$$

For comparison
BINOMDIST(6,13.0.25,0)=0.055922

This might not seem much but the binomial predicts 1/18 hands with 6 spades whereas the actual value is 1/24.

Geometric Distribution

Examines a different facet of a sequence of independent B-trials

X is the number of trials up to and including the first success .

Sequence =000001 => X=6

$$P(1st\ \text{success at}\ x^{th}\ \text{trial}) = p(1-p)^{x-1}$$

for x=1,2,3,.....

Examples:

1. My computer dials my ISP at 7pm, the probability that it establishes a connection on any attempt is 0.6.

The number of attempts required is Geometric(0.6).

$P(X = 3) = 0.6 \times 0.4^2 = 0.096$.

I am playing Monopoly and find myself in Jail – to get out I need to throw a double. The probability of throwing a double is 6/36 = 1/6=p
If I don't get out in 3 goes I have to pay a fine – so $P(X>3)$ is important.

$$P(X>3)=1-P(X\leq 3)=1-\sum_{x=1}^{3}P(X=x)$$

But simple manipulation of the Geometric series:

$$\sum_{x=1}^{r}p(1-p)^{x-1}=p\sum_{x=0}^{r-1}(1-p)^{x}=p\frac{1-(1-p)^{r}}{1-(1-p)}=1-(1-p)^{r}$$

So

$$P(X\leq x)=1-(1-p)^{r}\quad P(X>x)=(1-p)^{r}$$

P(X>3)=$\left(\frac{5}{6}\right)^{3}=\frac{125}{216}$

Poisson Distribution

This distribution is not obviously related to Bernoulli trials.
It is a distribution of counts in many random processes.

- # calls to a telephone exchange in a time period
- # emissions from a radioactive source in a time period

- Random events in a time period.

Distribution is defined to be

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

for x=0,1,2,3,.....

λ is called the Mean Parameter. $\lambda > 0$.

In Excel:
P(X=x)=POISSON(x,λ,0)
$P(X \leq x) = POISSON(x, \lambda, 1)$.

Example:

Analysis of football results suggests that the number of goals scored by a team in a match maybe modelled by a Poisson distribution. λ is a function of the offensive ability of the team and the defensive ability of the opposition.

Suppose that when Man-Utd play Chelsea λ=2.

P(Man-Utd score 3 goals)=$e^{-2}\frac{2^3}{3!}$=0.180447.

In the same match Chelsea score according to and *independent* Poisson with λ=1.3 (smaller λ because they are not as good as Man-Utd).

P(Chelsea score 3 goals)

=POISSON(3,1.2,0)=0.086744

P(3-3 result)=0.015653

More ambitiously P(Man-Utd Win)=?

$$P(MUnotlose) =$$

$$P(CH \leq MU) \approx \sum_{x=0}^{x=\infty} P(C \leq x \mid MU = x) \times P(MU = x)$$

By independence assumption the conditional probabilities are just $P(C \leq x)$. As scores in excess of 10 goals are extremely unlikely, we can replace the infinity in the summation by 10

So

$$P(CH \leq MU) \approx \sum_{x=0}^{10} POISSON(x,1.2,1) \times POISSON(x,2,0)$$

=0.775

$$P(Draw) = P(CH = MU) =$$

$$\sum_{x=0}^{10} POISSON(x,1.2,0) * POISSON(x,2,0)$$

$$= 0.216$$

P(MU win)=0.775-0.216 = 0.559

The real trick is to get good estimates for the λs.