Michaelmas Exercises 2004/5
Submit by 4PM Monday 24th Jan 2005 (Stats Office)

1. A survey of web users as to the type activities they engaged in.

1. Research — Yes = R , No = $\bar{R}$
2. Entertainment — Yes = E, No = $\bar{E}$
3. Shopping — Yes = S, No = $\bar{S}$

		S	$\bar{S}$
	E	0.08	0.20
R	$\bar{E}$	0.15	0.08
$\bar{R}$	E	0.15	0.16
	$\bar{E}$	0.06	0.12

(a) Compute the following probabilities and give a one sentence description of their meaning.

P(R), P(E), $P(R \cap E)$ are R and E independent?

$P(S \mid E)$ and $P(S \mid R)$.

$P(E \mid R)$ and $P(E \mid \bar{R})$ What inference do you draw from these about the relationship of E and R?

(b) Suppose three individuals are randomly selected from this population (probabilities of R, E and S are as above)

What is the probability that none of them shop via the Web?

If none of them shop what is the probability that at least one of them uses the Web for research?

2. Two tests are applied to assembled circuit boards . Test V is a visual inspection for missing components and soldering quality. Test R is a resistivity test.
The circuit board is either defective or not defective. (D or $\bar{D}$)

Suppose that:
P(Fails V| D) =0.9 , P(Fails V | $\bar{D}$) =0.1, P(Fails R|D)=0.9,
P(Fails R|$\bar{D}$)=0.03. and P(D)=0.04.

Compute:

P(D|fails V) (only V is applied)

P(D| passes R) (only R applied)

If V and R are independent (for both D boards and $\bar{D}$ boards)

P(fails both tests)
P($\bar{D}$ | passes both tests)

If V and R are independent for $\bar{D}$ boards so that $P(V \cap R \mid \bar{D}) = P(V \mid \bar{D}) * P(R \mid \bar{D})$ but not independent for D boards $P(V \cap R \mid D) = 0.8$, reconstruct the three way table (as in Q1)

		D	$\bar{D}$
	R		
V	$\bar{R}$		
$\bar{V}$	R		
	$\bar{R}$		

Hence compute the probability that a defective board fails at least one test and the probability that a board that has passed both is actually defective.

3. An automatic system is developed for visual grading of timber planks. The grading is based on the number of knots in the planks. A digital image of the plank is analysed by "knot recognition software". Based on the count of knots the plank is graded as:

Grade A not more than 2 knots
Grade B not more than 5 knots
Grade C more than 5 knots.

An extensive study of planks determined that the number of knots in a random plank is Poisson distributed with a mean of 2 (planks are independent)

(a) Explain what the Poisson model is and why it may or may not be reasonable in this context .

What proportion of planks are actually Grades A, B and C? Give excel expressions and evaluate these.

In a batch of 15 (independent) planks what is the appropriate model for the number of grade A planks.

What is the probability (use excel) that at least 8 of them are grade A.

What is the model for the number N of planks that I have to select (randomly) until I get a grade A plank? Compute P(N>5)

(b) The knot recognition software doesn't always get it right.
It recognises a genuine knot with probability 0.8 (independently for each knot)

Suppose that a plank actually has 8 knots, give an excel expression for the probability that at least 4 of these will be recognised as knots (evaluate the probability)

The software also "invents" knots –classifies other blemishes as knots.
The number of invented knots is Poisson distributed with mean 0.2. (per plank independent of the actual number of knots in the plank).

What is the mean of the number of knots recorded by the software? (taking into account Poisson 1.8 genuine knots, some of which are missed, and the invented ones)

It can be shown algebraically that with the scheme above the number of <u>genuine knots not missed</u> by the software is Poisson(λp) i.e. Poisson(1.8*0.8).
You are not required to show this but you will get 10 bonus marks by demonstrating this either algebraically or numerically.

Taking the above as true (not forgetting the invented ones) compute the proportion of planks classified by the software as grade A.

4. Two makes of computer monitor are available.

(a) Type A have lifetimes that are exponentially distributed. Thus the CDF is

$$F_X(x) = 1 - e^{-\lambda t}$$

$\lambda = 0.3$ (1/years)

What is the mean lifetime?
What is the probability that a monitor will last at least 3 years?
A particular monitor has been in use for 1 year, what is the probability that it will last at least another year?
What is the probability that not more than 2 monitors (original one plus one other) will be required for 3 years continuous operation.

(b) Type B monitors have lifetimes that are Gamma distributed with shape parameter a=2 and scale parameter $\alpha = 0.6$.

Thus the PDF is: $f_X(x) = \alpha^2 x e^{-\alpha x}$ x>0

You may assume that the cdf is
$F_X(x) = 1 - e^{-\alpha y} - \alpha y e^{-\alpha y}$

What is the mean lifetime?

What is the probability that a monitor will last at least 2 years?

A particular monitor has been in use for 1 year, what is the probability that it will last at least another year?

5. The actual resistance $1000\,\Omega$ resistors produced by a certain process is Normally distributed with a mean of $1000\,\Omega$ and a variance of 900.

(a) For a resistor produced by this process, compute the following probabilities:

The resistance is less than $975\,\Omega$.

The resistance is within 10% of the nominal value.

(b) What would the standard deviation of a process have to be so that only 1 in a 1000 resistors would be outside the range $\pm 10\%$ of the nominal value? (mean=1000)

(c) Two resistors are connected in series so that the total resistance = $R_1 + R_2$ what are the probabilities that the total resistance is:

within $50\,\Omega$ of the expected value.
within $\pm 10\%$ of the expected value.

(d) A voltage of 5V is applied to the ends of the resistor. Using Ohm's law, V=IR, and assuming that V is a known constant. What is the probability that the current, I, exceeds 5.5mA (=0.0055A) ?

(e) Suppose further that the process is such that for a resistor of nominal value $X\,\Omega$ the mean is X and the standard deviation is 0.02X.

A $1000\,\Omega$ and a $5000\,\Omega$ resistor are connected in series, what are the mean and variance of the total resistance.

5V is applied across this circuit – compute 95% a range for the current.

6. (a) A single machine in a network of 10 machines is infected by a virus. At random time intervals T each machine contacts one other machine (randomly chosen from the other 9). With probability p the contacted machine becomes infected and on its next contact may infect other machines.

T is exponentially distributed with a mean of 30 minutes, p=0.8.

Using the random numbers below simulate the process until 3 machines are infected. Explain what you are doing and report the time the 3rd machine becomes infected.

0.882527	0.689	0.217338	0.278995	0.380009
0.679072	0.200944	0.563485	0.88406	0.665581
0.249003	0.618067	0.696305	0.835708	0.947014
0.553888	0.544888	0.992601	0.417145	0.169657
0.491494	0.53307	0.261717	0.260366	0.619494
0.019286	0.543713	0.659946	0.464153	0.930175
0.617627	0.059412	0.043565	0.250976	0.161189
0.068364	0.894099	0.564778	0.129249	0.804696
0.336509	0.650402	0.779395	0.98621	0.676385
0.294789	0.869918	0.662932	0.104109	0.290693
0.284603	0.184103	0.981771	0.426104	0.299177
0.29733	0.860716	0.887247	0.873096	0.309201
0.446297	0.898289	0.781355	0.878278	0.482707
0.700971	0.072424	0.732294	0.603352	0.661683

Note: You need only simulate infected machines. Because of lack of memory of the exponential distribution the of the delay until a newly infected machine makes contact has the same exponential distribution as above. With a different distribution for T you would have to simulate all contacts.
(*Alternatively, if you wish, you may submit the listing of a program that does this (with different random numbers if you like) but you are not required to do so).*

(b) A simulation model is being built to investigate changes/improvements to the running of a PC maintenance/repair department of a large organisation.

As part of the model the type of fault needs to be simulated for each malfunctioning PC.
After a cursory inspection the faults are classified as follows:

Type	Description	Frequency
A	Physical damage to HD,CD or Floppy etc	26
B	No Power	37
C	Complete OS failure	64
D	Networking Fault	52

Frequency = number of the type of fault during the past year.

Illustrate the modified rejection method by simulating two PCs using the random numbers in (a) above (start from the bottom right corner for a change)