Exercises 2
Relationships between events illustrated.

Two screening test are available for identifying loan defaulters. These tests are used to evaluate requests for new loans.

Test A – Credit rating result: good (G) or poor (P)

Test B - Employment Status result : satisfactory (S) or not satisfactory (NS)

Statistical analysis has provided the following estimates of the test performances produced the following tables:

Sample size = 10158 loans

Default = D, non default = $\overline{D}$

Test A		
	D	$\overline{D}$
P	135	1008
G	255	8760

Test B		
	D	$\overline{D}$
NS	120	444
S	270	9324

We shall estimate the probabilities from this large sample.

Dividing the cells of A by the sample size we get the following probabilities:

Test A		
	D	$\overline{D}$
P	0.01329	0.099232
G	0.025103	0.862374

What we have in the table is that:

$P((A = P) \cap D) = 0.01329$ etc

$P(D) = 0.01329 + 0.025103 = 0.038393$ or nearly 4% of loans default.

Accuracy of the test at picking up potential defaulters is described by:

$$P(D \mid P) = \frac{P(D \cap P)}{P(P)} = \frac{0.01329}{0.01329 + 0.099232} = 0.11811$$

Thus 12% of people with poor credit rating default, in 88% of these cases the loan is OK.

Suppose the evaluating software applies this criterion rejecting those with the result P.

$$P(D \mid G) = \frac{P(D \cap G)}{P(G)} = \frac{0.025103}{0.025103 + 0.862374} = 0.028286$$

A reduction in the bad credit rate from 3.8% to 2.8%. However customers will be lost:

$P(P) = 0.01329 + 0.099232 = 0.112522$, 11.3% of customers are rejected.

$P(\overline{D} \mid P) = \dfrac{0.099232}{0.112522} = 0.88199$, 88.2% of these would have been OK.

Suppose an OK loan nets the company \$2000, the loss on a defaulted loan is \$10000.

If they don't use the criterion they will make on (large) N loan applications:

N * (0.038393* (-10000) + 0.961607*2000)=1539.279*N

If they do then

N*0.887478*(0.028286*(-10000)+0.971714*2000)=1473.715*N

So they should leave well alone or get better info they are rejecting too many.

Exercises:

As exercise evaluate Test B in the same way.

If you could only use one test which would you use?

Using both the tests.

We now have 3 events (P,G), (S,NS) and (D, $\overline{D}$)

The two tables above do not (in general) contain enough information for the 3 event model. What is the relationship between the tests?

$P(P \cap S) = ?$
How do the tests jointly impact on P(D) ?

The full model involves 7 parameters; suppose the disaggregated table is :

		Defaulting	
Test A	Test B	D	$\overline{D}$
P	NS	108	432
	S	27	576
G	NS	12	12
	S	243	8748

Note: the data in this table is the disaggregated version of the data in the tables above.

So that $P(D) = 0.038393$ is the same as before.

We can get the probabilities as:
$P(P \cap NS \cap D) = \dfrac{108}{10158}$ etc.

Exercises:
Compute the probabilities of defaulting and of rejecting a potentially good client with the following strategies:

1. Only grant a loan if pass both tests.
2. Only reject if fail both tests .

Are the tests independent?