STAT2001 Tutorial 3 Solutions

Problem 1

(a) Let: D = "Person has disease"
H = "Person doesn't have disease" (is healthy)
Y = "Person tests positive" (positive = yes)
N = "Person tests negative" (negative = no).

Then: $P(D) = 0.01$
$P(Y|D) = 0.9 = P(N|H)$.

So: $P(H) = 1 - P(D) = 0.99$
$P(Y|H) = 1 - P(N|H) = 0.1$.

Hence: $P(Y) = P(D)P(Y|D) + P(H)P(Y|H)$ (by LTP)
$= 0.01(0.9) + 0.99(0.1) = 0.108$.

(b) $P(YD) = P(D)P(Y|D) = 0.01(0.9) = 0.009$.

(c) $P(D \mid Y) = \dfrac{P(DY)}{P(Y)} = \dfrac{0.009}{0.108} = \dfrac{1}{12} = 0.0833$.

The posterior probability of the person having the disease (8%) seems rather small, considering the high accuracy of the test (90%). However, 8% is very large when compared to 1%, the prior probability. Ie, the positive test result has in fact greatly increased the chance of the person having the disease.

Another way to solve the problem is to consider giving the test to 1000 people in the population.

About 10 (1%) will have the disease,
of which 9 (90%) will test positive and 1 (10%) negative.

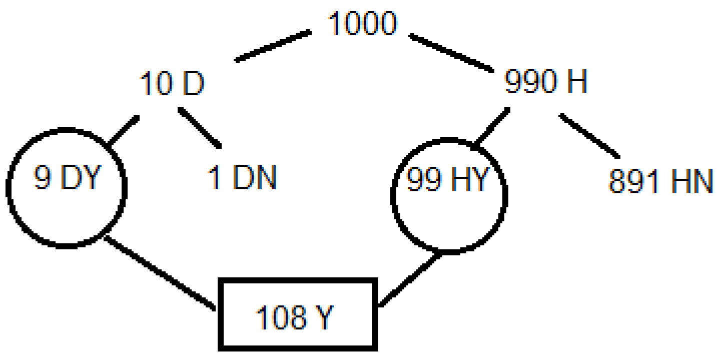
Conversely, about 990 (99%) will not have the disease,
of which 891 (90%) will test negative and 99 (10%) positive.

Thus a total of about 9 + 99 = 108 will test positive.

Of these 108, however, only 9 will actually have the disease.

Therefore the probability that a person will actually have the disease,
given that they test positive is 9/108 = 1/12.

This probability is small because the number of false positives (99) is high relative to the number of true positives (9).



(d) Repeat with $P(D) = 0.3$.
We can do the calculations in one step using Baye's rule:

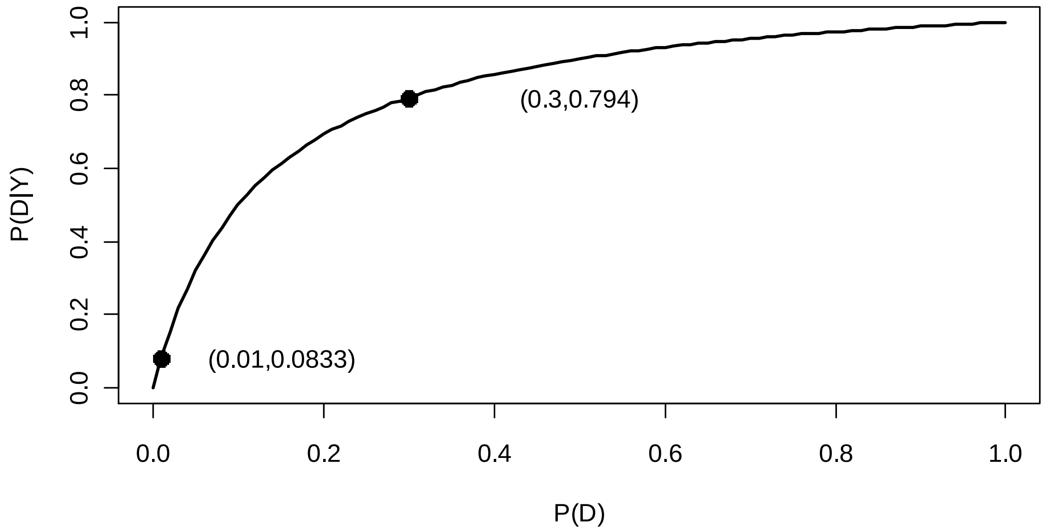
$$P(D\,|\,Y) = \frac{P(D)P(Y\,|\,D)}{P(D)P(Y\,|\,D) + P(H)P(Y\,|\,H)} = \frac{0.3(0.9)}{0.3(0.9) + 0.7(0.1)} = \frac{0.27}{0.34} = 0.794\,.$$

In this case the posterior probability of the person having the disease is high (79%). This is because the prior probability (30%) is already quite high.

Out of 1000 people with the stated symptoms, the number of true positives will now be about 300(0.9) = 270. This is no longer small compared to the number of false positives, namely 700(0.1) = 70. Therefore the required probability is high at
270/(270 + 70) = 0.794.

Note that as $P(D)$ approaches 1 or 0, so also does $P(D|Y)$. This behaviour can be expressed by the following formula:

$$P(D\,|\,Y) = \frac{P(D)\times 0.9}{P(D)\times 0.9 + (1 - P(D))\times 0.1} = \frac{9P(D)}{1 + 8P(D)}$$



R Code (non-assessable)

```
PD = seq(0,1,0.01); PDY = 9*PD/(1+8*PD)
plot(PD,PDY,type="l",lwd=2,xlab="P(D)",ylab="P(D|Y)")
points(0.01,0.0833,pch=16,cex=1.5); text(0.15,0.0833,"(0.01,0.0833)")
points(0.3,0.794,pch=16,cex=1.5); text(0.5,0.794,"(0.3,0.794)")
```

Problem 2

(a) $1 \geq P(A \cup B)$ since no event can have a probability greater than 1

$= P(A) + P(B) - P(AB)$ by the additive law of probability.

Therefore $P(AB) \geq P(A) + P(B) - 1$.

(b) Consider the following Venn diagram, where $a = A\bar{B}\bar{C}$, $g = ABC$, etc.

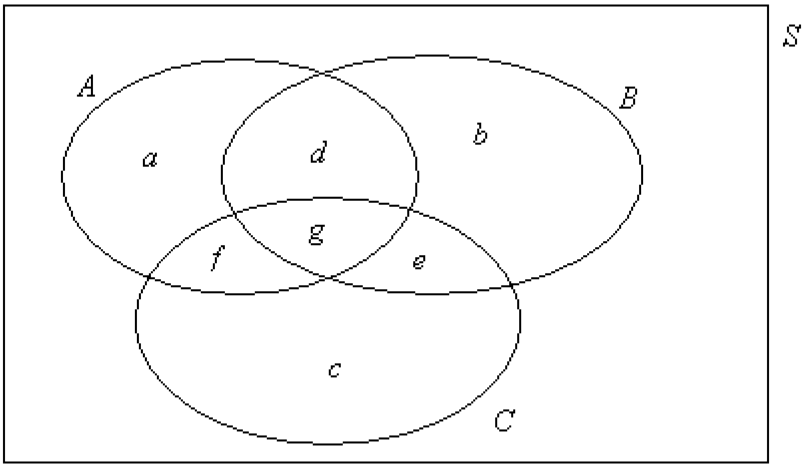
In this Venn diagram we see that

$$P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$= (a + d + g + f) + (b + e + g + d) + (c + f + g + e) - (d + g) - (f + g) - (g + e) + g$$

$$= a + b + c + d + e + f + g$$

$$= P(A \cup B \cup C).$$



The above does not actually prove the result. The following is a formal proof:

$P(A \cup B \cup C) = P(A \cup (B \cup C))$ by the associative laws

$= P(A) + P(B \cup C) - P(A(B \cup C))$

by the additive law of probability for two events

$= P(A) + \{P(B) + P(C) - P(BC)\} - P((AB) \cup (AC))$

by the additive law again and the distributive laws.

But $P((AB) \cup (AC)) =$ $P(AB) + P(AC) - P((AB)(AC))$ by the additive law yet again, where $P((AB)(AC)) = P(ABC)$ by the associative laws.

The result follows.

A similar argument can be used to prove, by induction, the additive law of probability for any number of events.

Problem 3

(a) There are $n = \binom{52}{5} = 2598960$ different possible hands.

The number of different hands with 3 cards of any one suit and 2 of another is

$a = P_2^4 \binom{13}{3}\binom{13}{2}$ = 267696. There are P_2^4 choices for the two suits.

So the required probability is a/n = 0.103.

(b) The number of different hands with no aces is $b = \binom{4}{0}\binom{48}{5}$ = 1712304.

So the number of different hands with at least one ace is $c = n - b = 886656$.

Hence the required probability is $c/n = 0.341$.

Alternatively, the probability of no aces is $\frac{48}{52}\frac{47}{51}\frac{46}{50}\frac{45}{49}\frac{44}{48}=0.659$,

where 48/52 is the probability of not getting an ace on the first draw, etc.
Hence the probability of at least one ace is 1 – 0.659 = 0.341.

Yet another solution is as follows (after defining A = number of aces):

$$P(A\geq 1)=P(A=1)+P(A=2)+P(A=3)+P(A=4)$$

$$=\frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}}+\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}+\frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}}+\frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} = 0.341.$$

Problem 4

(a) $\binom{7}{4}\binom{5}{3}7!=1764000$.

(b) $\binom{5}{1}+\binom{5}{2}+\binom{5}{3}+\binom{5}{4}+\binom{5}{5}=31$.

Alternatively, each ingredient can either be in the salad or not in it (two possibilities). The total number of possibilities is therefore 2^5 = 32. But this includes the case where none of the ingredients is in the salad. Hence the number of different salads is
32 – 1 = 31.

Note: These two solutions illustrate the following result:

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n.$$

This result follows from the binomial theorem, which says that:

$$\binom{n}{0}a^0b^n+\binom{n}{1}a^1b^{n-1}+\binom{n}{2}a^2b^{n-2}+\ldots+\binom{n}{n}a^nb^0=(a+b)^n$$

(put a = b = 1).