THE UNIVERSITY OF SYDNEY
MATH1005 Statistics
maths.usyd.edu.au/u/UG/WS/WS1005/

Summer/Winter/Semester2 **Tutorial Solutions 5** 2015

This tutorial explores probability.
Most of the questions involve hand calculations, not R.
There are lots of questions, so work carefully through some questions in your tutorial and complete the rest at home.
The 1st 20 minutes of class will be Online Quiz 1.

Probability
For 2 events A and B on the same sample space Ω

A and B are mutually exclusive	$P(A \cap B) = 0$
A and B are independent	$P(A \cap B) = P(A)P(B)$
the union of A and B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
probability of A conditional on B	$P(A\|B) = \dfrac{P(A \cap B)}{P(B)}$

Combinatorial coefficients

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \qquad \text{where } N! = N(N-1)(N-2)\ldots 1$$

Generalised Hypergeometric Distribution (Sampling without replacement)
Given an urn of size N with N_1 balls of type 1, N_2 balls of type 2, ... N_k balls of type k.
Draw a sample of size n without replacement.
The probability of selecting n_1 balls of type 1, n_2 balls of type 2, ... n_k balls of type k is

$$\frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\ldots\binom{N_k}{n_k}}{\binom{N}{n}}$$

1. Combinatorial coefficients

(a) By hand, show that $6! = 720$, $\binom{6}{0} = 1$ and $\binom{6}{3} = 20$.

(b) Confirm your answers by finding the correct button on your calculator, and by using R:

```
> factorial(6)
> choose(6,0)
```

Solution

(a) $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$

$$\binom{6}{0} = \frac{6!}{0!(6-0!)} = 1$$

$$\binom{6}{3} = \frac{6!}{3!(6-3!)} = \frac{720}{6 \times 6} = 20.$$

(b) Output from R

```
> factorial(6)
[1] 720
> choose(6,0)
[1] 1
> choose(6,3)
[1] 20
```

2. Hypergeometric Distribution

A fish tank has 6 tropical fish, of which 3 are angel fish. Assume that each fish has an equal chance of being caught and that 3 fish are sampled without replacement.

(a) What is the probability of catching no angel fish?

(b) What is the probability of catching two angel fish?

Solution

(a) $P(\text{sample of 0 angel fish and 3 other fish}) = \dfrac{\binom{3}{0}\binom{3}{3}}{\binom{6}{3}} = \dfrac{1}{20}$

(b) $P(\text{sample of 2 angel fish and 1 other fish}) = \dfrac{\binom{3}{2}\binom{3}{1}}{\binom{6}{3}} = \dfrac{9}{20}$

3. Hypergeometric Distribution

There are 10 marbles in a jar: 7 marbles are red, 2 are blue and 1 is white. John picks out 2 marbles without replacement. What is the probability that they are the same colour?

Solution

The jar can be modelled by an hypergeometric distribution, where $N = 10$, $N_1 = 7$ (red marbles), $N_2 = 2$ (blue marbles) and $N_3 = 1$ (white marbles).

$P(\text{sample of same colour}) = P(\text{2 reds}) + P(\text{2 blues}) = \dfrac{\binom{7}{2}\binom{2}{0}\binom{1}{0}}{\binom{10}{2}} + \dfrac{\binom{7}{0}\binom{2}{2}\binom{1}{0}}{\binom{10}{2}} = \dfrac{21}{45} + \dfrac{1}{45} = \dfrac{22}{45}$

4. Two dependent events

A smoke-detector system consists of two parts A and B. If smoke occurs then A detects it with probability 0.95, B detects it with probability 0.98 and both of them detect it with probability 0.94.

(a) Write down $P(A), P(B)$ and $P(A \cap B)$.

(b) Show that A and B are not independent.

(c) What is the probability that the smoke will not be detected?

(d) What is the probability that A will not detect the smoke, given that B did detect the smoke.

Solution

Let A = detector A detects the smoke and B = detector B detects the smoke.

(a) $P(A) = 0.95$, $P(B) = 0.98$, $P(A \cap B) = 0.94$.

(b) Check whether $P(A \cap B) = P(A)P(B)$.
LHS = 0.94.
RHS = $0.95 \times 0.98 = 0.931$.
Hence, A and B are not independent.

(c) $P(\text{smoke will not be detected}) = P(\text{neither A nor B works}) = P(A' \cap B') = 1 - P(A \cup B)$.
We have $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.95 + 0.98 - 0.94 = 0.99$.
Hence, $P(\text{neither A nor B works}) = 1 - 0.99 = 0.01$.

(d) This is conditional probability.
$P(\text{A will not detect smoke}|\text{B did detect smoke}) = P(A'|B) = \dfrac{P(A' \cap B)}{P(B)} = \dfrac{0.04}{0.98} \approx 0.0408$ where $P(A' \cap B) = P(B) - P(A \cap B) = 0.98 - 0.94 = 0.04$.

Note: Drawing a Venn diagram can assist in the working.

5. Three independent events

Three football players will attempt to kick a field goal. Let A_1, A_2, A_3 denote the events that the field goal is made by player $1, 2, 3$, respectively. Assume that A_1, A_2, A_3 are independent with $P(A_1) = 0.5$, $P(A_2) = 0.7$ and $P(A_3) = 0.6$. Compute the probability that exactly one player is successful.

Solution

$P(\text{Exactly one person kicks a goal}) = P(A_1 \cap A_2^c \cap A_3^c) + P(A_1^c \cap A_2 \cap A_3^c) + P(A_1^c \cap A_2^c \cap A_3)$

Using independence,
$P(\text{Exactly one person kicks a goal})$
$= (0.5)(0.3)(0.4) + (0.5)(0.7)(0.4) + (0.5)(0.3)(0.6)$
$= 0.06 + 0.14 + 0.09$
$= 0.29$

6. Probability in R

Hospital data of discharged patients contains the following columns:

```
Column  Label
1       ID no.            2       Duration of hospital stay
3       Age               4       Sex  1=male  2=female
5       First temperature following admission
6       First WBC(x1000) following admission
7       Received antibiotic 1=yes  2=no
8       Received bacterial culture 1=yes  2=no
9       Service 1=med 2=surg.
```

(a) Read in the data and have a look at it.

```
> data=read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/hospital.txt"),skip=1)
```

(b) How many patients were discharged?

(c) Isolate the data on treatments A (antibiotic) and B (bacterial culture).

```
> a = data[,7]
> b = data[,8]
```

(d) Count the number of discharged patients receiving each treatment and both treatments.

```
> number.a = length(a[a==1])
> number.b = length(b[b==1])
> number.both=length(a[a==1 & b==1])
```

(e) Use the data to estimate the following probabilities.

Event	A	B	$A \cap B$
Probability			

(f) Are the events A and B mutually exclusive? Are the events A and B independent?

Solution

(a)

```
> data=read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/hospital.txt"),skip=1)

> data
   V1 V2 V3 V4   V5 V6 V7 V8 V9
1   1  5 30  2 99.0  8  2  2  1
2   2 10 73  2 98.0  5  2  1  1
...
25 25  4 41  2 98.0  5  2  2  1
```

(b) The number of patients is 25.

(c)

```
> a = data[,7]
> a
 [1] 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 1 2 1 2 2 1 2 2 2
> b = data[,8]
> b
 [1] 2 1 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2
```

(d)

```
> number.a = length(a[a==1])
> number.a
[1] 7      # number of patients receiving A
> number.b = length(b[b==1])
> number.b
[1] 6      # number of patients receiving B
> number.both=length(a[a==1 & b==1])
> n.both
[1] 2       # number of patients receiving both A and B
```

(e)

Event	A	B	$A \cap B$
Probability	$6/25 = 0.24$	$7/25 = 0.28$	$2/25 = 0.08$

(f)
As $P(A \cap B) \neq 0$, events A and B are not mutually exclusive.
As $P(A \cap B) \neq P(A)P(B)$, events A and B are not independent.

7. Unfair die

A six-sided die is loaded in such a way that every even number is twice as likely to occur as every odd number. The die is tossed once.

(a) Fill out the following probability distribution.

x	1	2	3	4	5	6	Total
P(X=x)							1

(b) What is the probability that a number (strictly) less than 4 occurs?

(c) Let A be the event that an even number occurs and let B be the event that a number divisible by 3 occurs. Find $P(A \cup B)$ and $P(A \cap B)$.

Solution

(a)

x	1	2	3	4	5	6	Total
P(X=x)	1/9	2/9	1/9	2/9	1/9	2/9	1

Reasoning:
For some value p, the weights for each outcome are as follows

x	1	2	3	4	5	6	Total
P(X=x)	p	$2p$	p	$2p$	p	$2p$	1

Since the weights must add to 1, $9p = 1$, hence $p = 1/9$.

(b) $P(\text{number strictly less than 4 occurs}) = P(1) + P(2) + P(3) = 4/9$.

(c) $A = \{2, 4, 6\}$ and $B = \{3, 6\}$.
$P(A \cup B) = P(\{2, 3, 4, 6\}) = 7/9$
$P(A \cap B) = P(\{6\}) = 2/9$

8. Table of frequencies

In 1988 the Physicians' Health Study Research Group released the results of a five-year experiment based on 22,071 people between the ages of 40 and 84. The purpose was to determine whether taking aspirin reduces the risk of a heart attack. The people had been randomly assigned to one of two treatment groups: one group took aspirin, the other took the placebo (non drug). The following information was obtained.

	people with heart attack	people without heart attack
Aspirin	104	10,933
Placebo	189	10,845

One person is selected by chance.
(a) Let A be the event that the selected person had a heart attack. Find $P(A)$.
(b) Let B be the event that the selected person took aspirin. Find $P(B)$.
(c) Find the probability that the selected person had a heart attack given that they were in the Aspirin group.
(d) Are A and B independent?

Solution

(a)
$P(A) = (104 + 189)/22071 = 293/22071 \approx 0.013$

(b)
$P(B) = (104 + 10933)/22071 = 11037/22071 \approx 0.5$

(c)
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{104/22071}{11037/22071} = 104/11037 \approx 0.009$$

(d)
Check whether $P(A \cap B) = P(A)P(B)$.
$LHS = P(A \cap B) = 104/22071 \approx 0.005$
$RHS = P(A)P(B) = (293/22071)(11037/22071) \approx 0.007$
Hence A and B are not independent.

9. Total of three dice

Suppose three fair six-sided dice are rolled independently and the numbers on the 3 top faces are added together.

(a) What is the total number of possible sequences $\{1,1,1\} \ldots \{6,6,6\}$?

(b) How many different ways can the numbers $\{1,1,1\}$, $\{1,1,2\}$, $\{1,1,3\}$, $\{1,2,2\}$, $\{1,1,4\}$, $\{1,2,3\}$, and $\{2,2,2\}$ can be arranged in order?

(c) What is the probability that the total is at most 6?

Solution

(a) There are $6^3 = 216$ possible sequences, all equally likely.

(b)
$\{1,1,1\}$ 1 way
$\{1,1,2\}$ 3 ways (i.e. 112, 121, 211)
$\{1,1,3\}$ 3 ways
$\{1,2,2\}$ 3 ways
$\{1,1,4\}$ 3 ways
$\{1,2,3\}$ 6 ways (i.e. 123, 132, 213, 231, 312, 321)
$\{2,2,2\}$ 1 way
Hence there is a total of 20 ways of getting a total at most 6.

(c) $P(\text{total is at most } 6) = 20/6^3 \approx 0.09$

10. Deck of cards

A standard deck of cards consists of 52 different cards, with 13 cards of each of the 4 suits (hearts, diamonds, spades and clubs). 5 cards are taken out without replacement and placed face-down on a table.

(a) What is the probability that all 5 cards belong to the suit of hearts?

(b) What is the probability that all 5 cards belong to the same suit?

(c) What is the probability that exactly 4 cards belong to the suit of hearts?

(d) If 4 of the cards are turned face-up on the table, and are all hearts, what is the probability that the 5th card also belongs to the suit of hearts?

Solution

(a) This can be modelled by a hypergeometric distribution, with $N = 52$, $N_1 = 13$ (hearts) and $N_2 = 39$ (non hearts). Draw 5 cards without replacement.
Hence $P(\text{5 cards are all hearts}) = \dfrac{\binom{13}{5}\binom{39}{0}}{\binom{52}{5}} = \dfrac{\binom{13}{5}}{\binom{52}{5}} \doteq 0.0005.$

(b) $P(\text{5 cards same suit}) = P(\text{5 cards hearts}) + P(\text{5 cards clubs}) + P(\text{5 cards diamonds}) + P(\text{5 cards spades}) = 4 \times \dfrac{\binom{13}{5}}{\binom{52}{5}} = 0.0019$

(c) $P(\text{4 cards are hearts, 1 card is non heart}) = \dfrac{\binom{13}{4}\binom{39}{1}}{\binom{52}{5}} \doteq 0.0107$

(d) If we know that 4 hearts are on the table, then of the 48 remaining cards, only 9 of them are hearts. Thus $P(\text{5th card is hearts, given first four are hearts}) = \dfrac{\binom{9}{1}\binom{39}{0}}{\binom{48}{1}} = \dfrac{9}{48} = 0.1875$.

Note: This can also be calculated using the conditional probability formula. Let A be the event for which the 5th card is hearts, and B the event in which 4 first cards are hearts. Thus

$$\begin{aligned} P(A|B) &= \frac{P(A\cap B)}{P(B)} \\ &= \frac{P(\text{first 5 cards are all hearts})}{P(\text{first 4 cards are all hearts})} \\ &= \frac{(a)}{\frac{\binom{13}{4}\binom{39}{0}}{\binom{52}{4}}} \\ &= \frac{\frac{\binom{13}{5}}{\binom{52}{5}}}{\frac{\binom{13}{4}}{\binom{52}{4}}} \\ &= 0.1875 \end{aligned}$$

11. Events on two dice

Two 6-sided dice (one red, one green) are rolled, the red first and then the green. The set of all possible outcomes may be represented as follows:

$$\begin{aligned} \Omega = \{ \quad & (1,1),(1,2), \quad \cdots \quad ,(1,6), \\ & (2,1),(2,2), \quad \cdots \quad ,(2,6), \\ & \quad \vdots \qquad\qquad \ddots \qquad \vdots \\ & (6,1),(6,2), \quad \cdots \quad ,(6,6) \quad \} \end{aligned}$$

Let A = 'first die shows 1', B = 'sum of rolls is 7', C = 'both rolls have the same number'.

(a) The event A can be written as $A = \{(1,1),(1,2),\ldots,(1,6)\}$. In a similar way, write out B, C, $A\cap B$, $A\cap C$ and $B\cap C$.

(b) Assuming all 36 possible outcomes are equally likely, determine the probabilities of A, B and C, and of the three pairwise intersections.

(c) Are any of the pairs (A and B, A and C or B and C) independent? Explain.

Solution

(a) $A = \{(1,1),(1,2),\ldots,(1,6)\}$
$B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$
$C = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$
$A\cap B = \{(1,6)\}$
$B\cap C = \emptyset$
$A\cap C = \{(1,1)\}$

(b) Since all the outcomes in Ω are equally likely, and since A, B and C all have 6 outcomes, it follows that $P(A) = P(B) = P(C) = \dfrac{6}{36} = \dfrac{1}{6}$

In addtion, $P(A\cap B) = P(A\cap C) = \dfrac{1}{36}$ and $P(B\cap C) = 0$.

(c) Since $P(A\cap B) = 1/36 = \dfrac{1}{6}\times\dfrac{1}{6} = P(A)\times P(B)$, it follows that A and B are independent.
Since $P(A\cap C) = 1/36 = \dfrac{1}{6}\times\dfrac{1}{6} = P(A)\times P(C)$, it follows that A and C are independent.
However, since $P(A\cap B) = 0 \neq \dfrac{1}{6}\times\dfrac{1}{6} = P(A)\times P(B)$, it follows that B and C are not independent.

12. Probability with 3 tosses of a coloured die

A six-sided die has two of its faces white, one red and three green. It is thrown three times in such a way that each face is equally likely to land facing upwards and that each throw is independent. Compute the probability that

(a) a white face is uppermost at each throw

(b) the same colour is uppermost at each throw.

Solution

(a)
Let W_i be the event that White is uppermost on throw i. $i = 1, 2, 3$.
Let R_i be the event that Red is uppermost on throw i. $i = 1, 2, 3$.
Let G_i be the event that Green is uppermost on throw i. $i = 1, 2, 3$.

If each of the 6 faces is equally likely, then $P(\text{White uppermost on throw i}) = P(W_i) = 2/6 = 1/3$, for $i = 1, 2, 3$.
Similarly, $P(\text{Red uppermost on throw i}) = P(R_i) = 1/6$
and $P(\text{Green uppermost on throw i}) = P(G_i) = 3/6 = 1/2$.

Using independence,
$P(\text{a white face is uppermost at each throw}) = P(W_1)P(W_2)P(W_3) = (1/3)^3 = 1/27$.

(b) $P(\text{the same colour is uppermost at each throw}) = (1/3)^3 + (1/6)^3 + (1/2)^3 = 1/6$.

Reasoning:
$P(\text{the same colour is uppermost at each throw}) = P(\text{all red OR all white OR all green})$
$= P\{(W_1 \cap W_2 \cap W_3) \cup (R_1 \cap R_2 \cap R_3) \cup (G_1 \cap G_2 \cap G_3)\}$

As we have decomposed the event into three mutually exclusive events, it follows that
$P(\text{all the same colour}) = P(W_1 \cap W_2 \cap W_3) + P(R_1 \cap R_2 \cap R_3) + P(G_1 \cap G_2 \cap G_3)$.

As W_i, R_i and G_i are all independent, it follows that
$P(\text{all the same colour}) = P(W_1)P(W_2)P(W_3) + P(R_1)P(R_2)P(R_3) + P(G_1)P(G_2)P(G_3) = (1/3)^3 + (1/6)^3 + (1/2)^3$.

13. Extension ('Simulation' in R of Q9)

The following commands are not examinable,

```
> rolls = sample(3000,x=c(1,2,3,4,5,6),replace=T)     # 3000 random die rolls
> mat=matrix(rolls,ncol=3)                     # puts rolls into a 1000-by-3 matrix
> getsums=apply(mat,1,sum)                              # gets the sums of each row
> table(getsums)                          # displays a frequency table of the sums
> sum(getsums<=6)                              # counts how many sums are at most 6
> mean(getsums<=6)          # same as above but divides by length(getsums)=1000
```

Note: You can copy and paste the above block of commands into the R command line, as the `#`'s "comment out" the trailing text, so R ignores the rest of the line.

Solution

```
> rolls = sample(3000,x=c(1,2,3,4,5,6),replace=T)
> mat=matrix(rolls,ncol=3)
```

```
> getsums=apply(mat,1,sum)
> table(getsums)
> table(getsums)
> sum(getsums<=6)
[1] 92
> mean(getsums<=6)
[1] 0.092
```