The University of Sydney
MATH1005 Statistics
maths.usyd.edu.au/u/UG/WS/WS1005/

Summer/Winter/Semester2 **Tutorial Solutions 6** 2015

This tutorial explores random variables and discrete distributions, with special focus on the Binomial.

Discrete Random Variables
For a discrete random variable X

probability distribution	$\{x, P(X=x)\}$ (often represented in table)
expected value or mean	$E(X) = \sum_x xP(X=x)$
expected value of function	$E(f(X)) = \sum_x f(x)P(X=x)$
Eg	$E(X^2) = \sum_x x^2 P(X=x)$
variance	$Var(X) = E(X^2) - (E(X))^2$

Binomial Distribution (cf Sampling with replacement)
Consider a sequence of n independent, Binary trials with probability of success p.

The probability of getting x successes is

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \dots, n$$

Moments: $E(X) = np$ and $Var(X) = np(1-p)$

1. Expected values using a table and formula

(a) Given the following probability distribution table

x	1	2	3	4	Total
$P(X=x)$	0.1	0.2	0.3	0.4	1

Find $E(X), E(X^2)$, $E(\frac{1}{X})$ and $Var(X)$?

(b) What is the formula for the probability distribution function?

Solution

(a) $E(X) = \sum_{i=1}^{n} xP(X=x) = 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4 = 3$

$E(X^2) = \sum_{i=1}^{n} x^2 P(X=x) = 1^2 \times 0.1 + 2^2 \times 0.2 + 3^2 \times 0.3 + 4^2 \times 0.4 = 10$

$E(1/X) = \sum_{i=1}^{n} 1/xP(X=x) = 1/1 \times 0.1 + 1/2 \times 0.2 + 1/3 \times 0.3 + 1/4 \times 0.4 = 0.4$

$Var(X) = E(X^2) + E^2(X) = 10 - (3)^2 = 1$

Check in R:
```
> x=c(1,2,3,4)
> p=c(0.1,0.2,0.3,0.4)
> sum(x*p)
> sum(x^2*p)-sum(x*p)^2
```

(b) $P(X = x) = \dfrac{x}{10}$ for $x = 1, 2, 3, 4$.

2. Using Binomial tables

If $X \sim B(9, 0.7)$, find
(a) $P(X \leq 3)$
(b) $P(2 < X < 7)$
(c) $P(3 \leq X < 7)$
(d) $P(X \geq 1.5)$.

Solution

Given: $X \sim B(9, 0.7)$

(a) $P(X \leq 3) = 0.0253$ (Binomial tables).

Check in R:
```
> pbinom(3,9,0.7)
[1] 0.02529484
```

(b) $P(2 < X < 7) = P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2) = 0.5372 - 0.0043 = 0.5329$.

Check in R:
```
> pbinom(6,9,0.7)-pbinom(2,9,0.7)
[1] 0.5328779
OR
> x=c(3,4,5,6)
> sum(dbinom(x,9,0.7))
```

(c) $P(3 \leq X < 7) = P(3 \leq X \leq 6) = 0.5329$. (previous answer)

(d) As the Binomial is a discrete integer valued distribution,
$P(X \geq 1.5) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.0004 = 0.9996$.

Check in R:
```
> 1-pbinom(1,9,0.7)
[1] 0.999567
```

3. Binomial Probabilities

(a) Given $X \sim Bin(n = 10, p = 0.4)$. Find $P(X = 2)$ and $P(X \leq 2)$ by both using the formula and looking up the Binomial tables.

(b) Check your answers using R.

```
> dbinom(2,10,0.4)    # This calculates probability P(X=2)
> pbinom(2,10,0.4)     # This calculates cumulative probability P(X <= 2)
```

Solution

(a) Formula:

$P(X = 2) = \binom{10}{2}(0.4)^2(0.6)^8 = 0.1209324$

$P(X \leq 2) = \binom{10}{2}(0.4)^2(0.6)^8 + \binom{10}{1}(0.4)^1(0.6)^9 + \binom{10}{0}(0.4)^0(0.6)^{10} = 0.1673$

Binomial Tables:
$P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0.1673 - 0.0464 = 0.1209$
$P(X \leq 2) = 0.1673$

(b) Check your answers using R.

```
> dbinom(2,10,0.4)
[1] 0.1209324
> pbinom(2,10,0.4)
[1] 0.1672898
```

4. Mean and Variance of Binomial

(a) Suppose that $X \sim B(4, 0.2)$. State the mean, variance and standard deviation of X.

(b) Using hand calculations, construct the following table

x	0	1	2	3	4	Total
$P(X = x)$						

(c) Using the table, calculate $E(X)$. Check that this agrees with (a).

(d) Using the table, calculate $E(X^2)$. Calculate $Var(X)$. Check that this agrees with (a).

(e) Using the table, find the probability $P(X \leq 3)$. Check you can also get this answer from the Binomial tables and by using R.

```
> pbinom(3,4,0.2)
```

Solution

(a) Given $X \sim B(4, 0.2)$,
$E(X) = np = 4 \times 0.2 = 0.8.$
$Var(X) = np(1 - p) = 4 \times 0.2 \times 0.8 = 0.64.$
$SD(X) = \sqrt{0.64} = 0.8.$

(b)

x	0	1	2	3	4	Total
$P(X = x)$	0.4096	0.4096	0.1536	0.0256	0.0016	1

Working:

$P(X = 0) = \binom{4}{0}(0.2)^0(0.8)^4 = 0.4096$

$$P(X=1) = \binom{4}{1}(0.2)^1(0.8)^3 = 0.4096$$
$$P(X=2) = \binom{4}{2}(0.2)^2(0.8)^2 = 0.1536$$
$$P(X=3) = \binom{4}{3}(0.2)^3(0.8)^1 = 0.0256$$
$$P(X=4) = \binom{4}{4}(0.2)^4(0.8)^0 = 0.0016$$

(c) $E(X) = 0 \times 0.4096 + 1 \times 0.4096 + 2 \times 0.1536 + 3 \times 0.0256 + 4 \times 0.0016 = 0.8$
Agrees with (a).

(d) $E(X^2) = 0^2 \times 0.4096 + 1^2 \times 0.4096 + 2^2 \times 0.1536 + 3^2 \times 0.0256 + 4^2 \times 0.0016 = 1.28$

$Var(X) = 1.28 - (.8)^2 = 0.64$. Agrees with (a).

(e) $P(X \leq 3) = P(0) + P(1) + P(2) + P(3) = 0.9984$.
Or $P(X \leq 3) = 1 - P(X = 4) = 1 - 0.0016 = 0.9984$.

```
> pbinom(3,4,0.2)
[1] 0.9984
```

5. Identifying Binomial Distribution

(a) Does X have a binomial distribution in each of the following situations? Explain your reasoning.

(i) We observe the gender of the babies born in the next 15 births at a local hospital. X is the number of girls.

(ii) A couple decides to continue to have children until their first girl is born. X is the total number of children in the family.

(iii) Each child born to a particular set of parents has probability 0.25 of having blood type O. These parents have 6 children. X is the number of children with blood type O.

(b) Where a Binomial distribution is appropriate, write $X \sim Bin(n, p)$, identifying the parameters n and p, and calculate $P(X = 1)$.

Solution

(a) (i) Yes: We have $n = 15$ independent Binary trials, where $p = P(\text{baby is girl})$. We assume there are no identical twins.

(ii) No: The number of trials is random, not fixed.

(iii) Yes: We have $n = 6$ independent Binary trials, where $p = P(\text{Blood group O}) = 0.25$. We assume that each child's blood group is assigned independently of any other.

(b) (i) $X \sim Bin(15, 0.5)$. $P(X = 1) = \binom{15}{1}(0.5)^1(0.5)^{14} = 0.0004577637$.

(ii) N/A

(iii) $X \sim Bin(6, 0.25)$. $P(X = 1) = \binom{6}{1}(0.25)^1(0.75)^5 = 0.355957$

6. Comparing Binomial and Hypergeometric

Suppose that a hat contains 25 tickets of which 20% are blue and the rest are white.

(a) Suppose 20% of them are drawn out randomly without replacement. What is the probability that 20% of the tickets in the sample are blue?

(b) How does the answer change if the sampling is done with replacement?

Solution

(a) This is modelled by a hypergeometric, with $N = 25$, $N_1 = 20\% \times 25 = 4$ (Blue) and $N_2 = 20$ (White). Draw a sample of size $n = 20\% \times 25 = 5$ randomly without replacment.

P(20% of the tickets in the sample are blue)
$= P$(20% of the 5 tickets in the sample are blue)
$= P$(1 ticket in the sample is blue)
$= \frac{\binom{5}{1}\binom{20}{4}}{\binom{25}{5}}$
$= \frac{5 \times 4845}{53130}$
$= 0.4559571$

(b) This is now modelled by a Binomial: $X \sim Bin(n = 5, p = P(Blue) = 1/5)$.
P(1 ticket in the sample is blue) $= P(X = 1) = \binom{5}{1}(0.2)^1(0.8)^4 = 0.4096$

7. Geometric Distribution

A fair coin is tossed until a head occurs. X is the number of coin tosses (failures) before the first success (head) . Fill out this table:

x	0	1	2	3+	Total
$P(X = x)$					1

What is the probability distribution function?

Solution

$X =$ number of tosses until a head occurs $\sim Geo(p = P(Head) = 0.5)$.

x	0	1	2	3+	Total
$P(X = x)$	0.5	0.25	0.125	0.125	1

Working:
$P(X = 0) = P(\text{Head on 1st toss}) = 0.5$.
$P(X = 1) = P(\text{1st Head on 2nd toss}) = P(\text{Tail, then Head}) = 0.5^2 = 0.25$.
$P(X = 2) = P(\text{1st Head on 3rd toss}) = P(\text{2 Tails, then Head}) = 0.5^3 = 0.125$.
$P(X = 3+) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 0.125$.

Probability distribution function: $P(X = x) = (0.5)^{x+1}$, for $x = 0, 1, 2, \ldots$.

Check in R:
```
> dgeom(0,0.5)
[1] 0.5
```

8. Binomial Distribution

Of a large number of mass-produced articles, it is known that 2% of them are defective. Write X for the number of defective items in a random sample of 10 of these articles.

(a) Explain why the distribution of X is binomial.

(b) Suppose that for quality control purposes, the line is shut down for repairs, if the number of faulty items is two or more. Find the probability that the line is shut down. Check your answer in R.

(c) What is the probability that the number of defectives is between 5 and 8 inclusive? Use R.

Solution

$X =$ the number of defective items in a random sample of 10 $\sim Bin(n = 10, p = 0.02)$

(a) X is a Binomial because it counts the number of successes (defectives) in a series of independent Bernoulli trials.

(b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0}(0.02)^0(0.98)^{10} - \binom{10}{1}(0.02)^1(0.98)^9 = 0.01617764$

Hence there is approximately a 2% chance that the line is shut down.

Check in R:

```
> 1-pbinom(1,10,0.02)
[1] 0.01617764
```

(c) We want $P(5 \leq X \leq 8) = P(X \leq 8) - P(X \leq 4)$

Using R:

```
> pbinom(8,10,0.02)-pbinom(4,10,0.02)
[1] 7.41464e-07
```

9. Extension (Simulation of Binomial Distribution in R)

It is known from previous testing that 40% of mice used in an experiment become agressive within one minute of being administered an experimental drug.

(a) Produce a frequency table for $Bin(9, 0.4)$.

```
> x=c(0,1,2,3,4,5,6,7,8,9)      # This produces the outcomes 0,1,...9
> p=dbinom(x,9,0.4)             # This produces the probabilities
> rbind(x,p)                    # This combines x and p
> sum(x*p)                      # This produce the mean of x [Check:E(X)= np]
```

(b) Conduct a simulation experiment to generate 100 random variables $X \sim Bin(n = 9, p = 0.4)$.

```
> set.seed(1234)        # This chooses the random number generator
> x=rbinom(100,9,0.4)   # This simulates 100 numbers from a Bin(9,0.4)
> plot(table(x))        # This produces a frequency table of simulations
> hist(x)               # This produces a histogram of simulations
> mean(x)               # This produces mean of simulations
> var(x)                # This produces variance of simulations
```

Solution

(a) Produce a frequency table for $Bin(9, 0.4)$.

```
> x=c(0,1,2,3,4,5,6,7,8,9)
> p=dbinom(x,9,0.4)
> rbind(x,p)
       [,1]       [,2]      [,3]      [,4]      [,5]
x 0.0000000 1.00000000 2.0000000 3.0000000 4.0000000
p 0.0100777 0.06046618 0.1612431 0.2508227 0.2508227
       [,6]       [,7]       [,8]        [,9]       [,10]
x 5.0000000 6.00000000 7.00000000 8.000000000 9.000000000
p 0.1672151 0.07431782 0.02123366 0.003538944 0.000262144
> sum(x*p)
[1] 3.6         # Check: E(X)= np = 9*0.4=3.6
> sum(x^2*p)-sum(x*p)^2
[1] 2.16        # Check: Var(X)= np(1-p) = 9*0.4*0.6=2.16
```

(b) Conduct a simulation experiment to generate 100 random variables $X \sim Bin(n = 9, p = 0.4)$.

```
> set.seed(1234)
> x=rbinom(100,9,0.4)
> plot(table(x))
> hist(x)
> mean(x)
[1] 3.35        # Note: Compare to exact E(X)=3.6
> var(x)
[1] 1.825758    # Note: Compare to exact Var(X)=2.16
```