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	$Var(X) = E(X^2) - E(X)^2$

Combinatorial coefficients

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

Eg1: 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.

$$P(\text{Select } n_i \text{ balls of each type } i\ ) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\ldots\binom{N_k}{n_k}}{\binom{N}{n}}$$

Eg2: Binomial Distribution (Sampling with replacement)
Consider a sequence of n independent, Binary trials with probability of success p, and let X count the number of successes.

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

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

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)
```

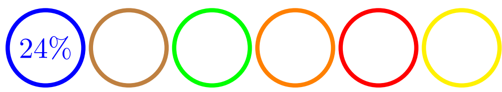
(c) What does $\binom{6}{x}$ represent?

Solution

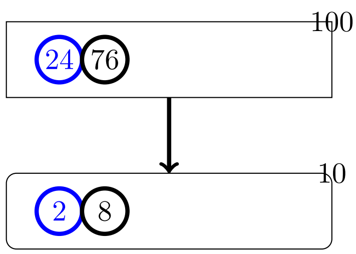
(c) $\binom{6}{x}$ represents the number of ways of choosing x items from 6, where $x = 0, 1, 2, 3, 4, 5, 6$.

2. Modelling M&Ms

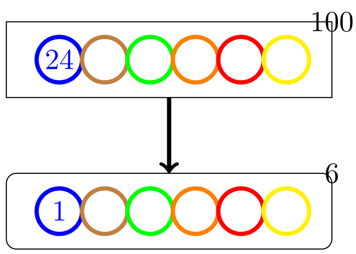
(a) According to the following article, what are the proportions of colours of M&Ms in each standard pack? http://joshmadison.com/mms-color-distribution-analysis/



(b) If I randomly pick 10 M&Ms from a certain packet containing 100 M&Ms, what is the chance that I get 2 blues.



(c) If I randomly pick 6 M&Ms from a certain packet containing 100 M&Ms, what is the chance that I get 1 of each colour?

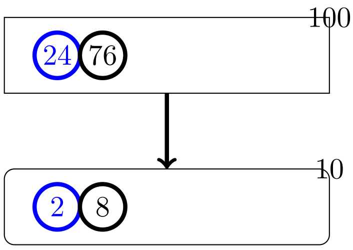


(d) M&Ms runs a competition with a gift voucher in 10% of packets. Suppose I buy 20 packets. What is the chance that I don't get any vouchers.

Solution

(a) Blue: 24%, Brown 14%, Green 16%, Orange 20%, Red 13%, Yellow 14%.

(b) Model the packet by a hypergeometric.
$N = 100$, $N_1 =$ Number of Blues $= 24$ and $N_2 =$ Number of non-blues $= 76$; $n = 10$.



$P(2 \text{ blue, } 8 \text{ non-blues}) = \dfrac{\binom{24}{2}\binom{76}{8}}{\binom{100}{10}} \approx 0.3.$

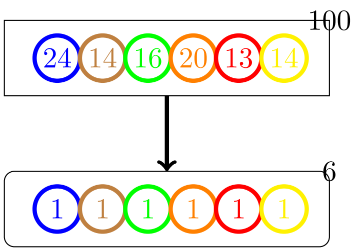
```
choose(24,2)*choose(76,8)/choose(100,10)

## [1] 0.300643

dhyper(2,24,76,10,log=FALSE)

## [1] 0.300643
```

(c) Model the packet by a hypergeometric.
$N = 100$, $N_1 =$ Number of Blues $= 24$; $N_2 =$ Number of Browns $= 14, \ldots, N_6 =$ Number of Yellows $=$ 14; $n = 6$.



$P(1 \text{ of each colour}) = \dfrac{\binom{24}{1}\binom{14}{1}\binom{16}{1}\binom{20}{1}\binom{13}{1}\binom{14}{1}}{\binom{100}{6}} \approx 0.02.$

```
choose(24,1)*choose(14,1)*choose(16,1)*choose(20,1)*choose(13,1)*choose(14,1)/choose(100,6)

## [1] 0.01641592
```

(d) Model the competition by a Binomial.
$X =$ The total number of vouchers in 20 packets $\sim Bin(n = 20, p = 0.1)$.
$P(X = 0) = \dbinom{20}{0}(0.1)^0(0.9)^{20} \approx 0.12.$

```
choose(20,0)*(0.1^0)*(0.9^20)

## [1] 0.1215767

dbinom(0,20,0.1)

## [1] 0.1215767
```

Note: Any data drawn from a blog must be scrutinised in terms of reliability. A more reliable source of information about M&Ms is found here:
http://www.maths.usyd.edu.au/u/UG/JM/MATH1005/r/loc/Mars.pdf

3. Simulating a Binomial Distribution

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

(b) Using the Binomial distribution formula, complete the following table

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

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

```
#Check answers in R
x=c(0:4)
p=dbinom(x,4,0.2)
sum(x*p)
sum(x^2*p)
sum(x^2*p)-sum(x*p)^2
```

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

```
pbinom(3,4,0.2)

## [1] 0.9984
```

(e) Simulate the distribution by generating 100 random variables from $X \sim Bin(n = 4, p = 0.2)$.

```
set.seed(1234)        #Chooses the random number generator
x=rbinom(100,4,0.2)   #Simulates 100 numbers from a Bin(4,0.2)
plot(table(x))        #Produces a frequency table of simulations
hist(x)               #Produces a histogram of simulations
mean(x)               #Produces mean of simulations
var(x)                #Produces the variance of simulations
```

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).

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

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

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

```
pbinom(3,4,0.2)

## [1] 0.9984
```

Extra Questions

4. Expected values and variance

Given the following probability distribution table

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

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

(b) By hand, find $E(X), E(X^2)$, $E(\frac{1}{X})$ and $Var(X)$?

```
#Check answers in R
x=c(1:4)
p=x/10
sum(x*p)
sum(x^2*p)
sum((1/x)*p)
sum(x^2*p) - sum(x*p)^2
```

(c) Estimate your answers by simulation.

```
y=sample(x, 10000, prob=p, replace=T)   #Simulates 10000 draws
mean(y)
mean(y^2)
mean(1/y)
var(y)
```

Solution

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

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

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

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

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

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

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

6. 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).

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

7. 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).

```
pbinom(3,9,0.7)

## [1] 0.02529484
```

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

```
pbinom(6,9,0.7)-pbinom(2,9,0.7)

## [1] 0.5328779

x=c(3,4,5,6)
sum(dbinom(x,9,0.7))

## [1] 0.5328779
```

(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,
$P(X \geq 1.5) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.0004 = 0.9996$.

```
1-pbinom(1,9,0.7)

## [1] 0.999567
```

8. 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$)$
$= \frac{\binom{5}{1}\binom{20}{4}}{\binom{25}{5}}$
$= \dfrac{5 \times 4845}{53130}$
$= 0.4559571$

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

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

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

```r
pbinom(8,10,0.02)-pbinom(4,10,0.02)

## [1] 7.41464e-07
```

10. 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:

$$\Omega = \{ \begin{array}{lll} (1,1),(1,2), & \cdots & ,(1,6), \\ (2,1),(2,2), & \cdots & ,(2,6), \\ \vdots & \ddots & \vdots \\ (6,1),(6,2), & \cdots & ,(6,6) \end{array} \}$$

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 $P(A) = P(B) = P(C) = \frac{6}{36} = \frac{1}{6}$

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

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

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

12. 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 $x = 0, 1, 2, \ldots$.

```
dgeom(0,0.5)

## [1] 0.5
```

13. 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)    #Calculates probability P(X=2)
pbinom(2,10,0.4)     #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:

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

$P(X \leq 2) = 0.1673$

14. Simulation of Binomial Distribution

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)      #Produces the outcomes 0,1,...9
p=dbinom(x,9,0.4)             #Produces the probabilities
rbind(x,p)                    #Combines x and p
sum(x*p)                      #Produces 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
```

15. (Extension: Poisson Distribution)

A Poisson Distribution models rare events: $X \sim P_0(\lambda)$, where $\mu = E(X)$ and $P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, for $x = 0, 1, 2, \ldots$.

The expected number of daily car accidents outside a busy shopping centre is 0.75. On a particular day, what is the probability of 1 accident?

Solution

Let $X =$ Number of daily car accidents $\sim P_0(\lambda = 0.75)$.

$$P(X = 1) = \frac{0.75^1 e^{-0.75}}{1!} = 0.35$$

```r
dpois(1,0.75)

## [1] 0.3542749
```