The University of Sydney
School of Mathematics and Statistics

Solutions to Tutorial Week 6

MATH1905: Statistics (Advanced) Semester 2, 2017

Web Page: `http://sydney.edu.au/science/maths/MATH1905`
Lecturer: Michael Stewart

Recall that if X and Y are independent random variables, then for all functions $g(\cdot)$ and $h(\cdot)$,

$$E\left[g(X)h(Y)\right] = E\left[g(X)\right]E\left[h(Y)\right] .$$

1. (**Multiple Choice**) Suppose that $X_i \sim B(50, 0.02)$. The distribution of sample mean $\bar{X}$ based on a random sample of size $n = 100$ is approximately:

(a) $N(50, 0.02)$

(b) $N(50, 1)$

(c) $N(1, 0.98)$

(d) $N(1, 0.0098)$

(e) $N(0.01, 0.098)$

Solution: There are various ways to attack this. One is to consider firstly the sum

$$T = X_1 + X_2 + \cdots + X_{100}$$

where $X_1, X_2, \ldots, X_{100}$ are independent $B(50, 0.02)$. Now since the sum of independent binomials (with the same p) is also binomial (with that same p), we have that

$$T \sim B(5000, 0.02) .$$

We have that

$$E(T) = 5000 \times 0.02 = 100 \quad \text{and} \quad Var(T) = 5000 \times 0.02 \times 0.98 = 98 .$$

Since this is a binomial with a large n, it is approximately normal with the same mean (expectation) and variance. So

$$T \overset{\text{approx}}{\sim} N(100, 98) .$$

Now the average $\bar{X} = T/100$ has

$$E(\bar{X}) = E\left(\frac{T}{100}\right) = \frac{E(T)}{100} = 1$$

and

$$Var(\bar{X}) = Var\left(\frac{T}{100}\right) = \left(\frac{1}{100}\right)^2 Var(T) = \frac{98}{10000} = 0.0098 .$$

Finally, since $\bar{X} = T/100$ and T is approximately normal, then so too is $\bar{X}$, so

$$\bar{X} \overset{\text{approx}}{\sim} N(1, 0.0098) .$$

Thus the solution is (d).

2. (**Multiple Choice**) Suppose that $X_1, X_2, \cdots, X_{16}$ is a random sample of size 16 from the distribution $N(100, 25)$. The distribution of $\bar{X}$ (the sample mean) is:

(a) $N(100, 25)$

(b) $N\left(100, \frac{5}{4}\right)$

(c) $N(0, 25)$

(d) $N\left(100, \frac{25}{16}\right)$

(e) $N(0, 1)$

Solution: Use $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ where $\mu = 100$, $\sigma^2 = 25$, and $n = 16$, i.e. $\bar{X} \sim N(100, \frac{25}{16})$. Thus the solution is (d).

3. Suppose that random variables X_1 and X_2 have joint probability distribution $P(X_1 = x_1, X_2 = x_2)$ given by

		x_1 = -1	x_1 = 0	x_1 = +1
x_2	-1	1/16	3/16	1/16
	0	3/16	0	3/16
	+1	1/16	3/16	1/16

(a) Find the marginal distributions of X_1 and X_2.

Solution:

		x_1 = -1	x_1 = 0	x_1 = $+1$
x_2	-1	1/16	3/16	1/16
	0	3/16	0	3/16
	$+1$	1/16	3/16	1/16
marginal $P(X_1 = x_1)$		5/16	6/16	5/16

The marginal distribution for X_1 is:

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) = \begin{cases} 5/16 & \text{for } x_1 = -1 \\ 6/16 & \text{for } x_1 = 0 \\ 5/16 & \text{for } x_1 = 1 \end{cases}$$

Note also that X_2 has the same marginal distribution as X_1.

(b) Show that X_1 and X_2 are **not** independent.

Solution: X_1 and X_2 are independent *if and only if*

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$$

for all pairs (x_1, x_2). To show that X_1 and X_2 are **not** independent, it's enough to show this doesn't hold for one (x_1, x_2) combination. Consider $(x_1, x_2) = (0, 0)$:

$$0 = P(X_1 = 0, X_2 = 0) \neq P(X_1 = 0)P(X_2 = 0) = \left(\frac{6}{16}\right)^2.$$

Hence, X_1 and X_2 are not independent.

(c) Evaluate $E(X_1)$, $E(X_2)$ and $E(X_1X_2)$.

Solution:

$$E(X_1) = \sum_{x_1} x_1 P(X_1 = x_1) = -1 \times \frac{5}{16} + 0 \times \frac{6}{16} + 1 \times \frac{5}{16} = 0.$$

$$E(X_2) = \sum_{x_2} x_2 P(X_2 = x_2) = -1 \times \frac{5}{16} + 0 \times \frac{6}{16} + 1 \times \frac{5}{16} = 0.$$

$$\begin{aligned} E(X_1X_2) &= \sum_{x_1}\sum_{x_2} x_1 x_2 P(X_1 = x_1, X_2 = x_2) \\ &= (-1)(-1)\frac{1}{16} + (-1)(1)\frac{1}{16} + (1)(-1)\frac{1}{16} + (1)(1)\frac{1}{16} + \cdots \\ &= 0. \end{aligned}$$

On the second line of working for $E(X_1X_2)$ there should be 9 terms (9 pairs) in the sum, but all the terms not shown explicitly are equal to zero as either x_1 or x_2 equal zero.

(d) Determine whether the variables are uncorrelated. That is, check whether $Cov(X_1, X_2) = 0$. Comment on this result comparing with part (b).

Solution: If $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$ then $Cov(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$. However we have shown that $\mu_1 = \mu_2 = 0$ in this case so

$$Cov(X_1, X_2) = E(X_1 X_2) = 0\,.$$

Therefore, X_1 and X_2 are uncorrelated (which simply means their covariance is zero). Note that this means $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$ i.e. the variance of the sum is the sum of the variances (even though they are **not** independent!) Recall that X, Y independent **implies** $Cov(X, Y) = 0$. However this example shows that the reverse implication does **not** hold, hence being independent is a **stronger** condition than having zero covariance.

4. How many possible different words can be made by rearranging the letters of the word STATISTICS?

Solution: There are 3 S's, 3 T's, 2 I's, 1 A and 1 C giving 10 letters in all. The answer is simply the multinomial coefficient

$$\frac{10!}{3!3!2!1!1!} = 10 \times 9 \times 8 \times 7 \times 5 \times 2 = 50400\,.$$

Alternatively, using R we have

```
factorial(10)/((factorial(3)^2)*factorial(2))
```

```
[1] 50400
```

5. Suppose that an office receives telephone calls as a Poisson distribution with mean $\lambda = 0.5$ per min. What is the probability of receiving exactly 1 call during a 1 minute interval? What is the probability of receiving no call during a 1 minute interval? The number of calls in a 5 minute interval (also) follows a Poisson distribution with $\lambda = 5 \times 0.5$. What is the probability of receiving no call during a 5 minute interval?

Solution:

Let X denote the number of calls during a 1 minute interval: $X \sim \text{Pois}(0.5)$, the probability of no calls is $\text{P}(X = 0) = 0.6065$ and the probability of one call is $\text{P}(X = 1) = 0.3033$.

Let Y denote the number of calls during a 5 minute interval: $Y \sim \text{Pois}(2.5)$, the probability of no calls during a five minute interval is $\text{P}(Y = 0) = \exp(-2.5) = 0.082$.

6. Let $Z \sim N(0, 1)$. Consider the following R commands and output:

```
z=c(0.3,0.5,0.72,0.75,1,1.4,1.96)
Phi.z=pnorm(z)
cbind(z,Phi.z)
```

```
        z     Phi.z
[1,] 0.30 0.6179114
[2,] 0.50 0.6914625
[3,] 0.72 0.7642375
[4,] 0.75 0.7733726
[5,] 1.00 0.8413447
[6,] 1.40 0.9192433
[7,] 1.96 0.9750021
```

```
p=c(0.9,0.95)
Phi.inv.p=qnorm(p)
cbind(p,Phi.inv.p)
```

```
        p Phi.inv.p
[1,] 0.90  1.281552
[2,] 0.95  1.644854
```

(a) Use the information above to find (to 4 decimal places)

(i) $P(Z \leq 1.4)$

Solution: Reading directly from the output, since `pnorm(z)` is precisely $P(Z \leq \texttt{z})$,

$$P(Z \leq 1.4) \approx 0.9192\,.$$

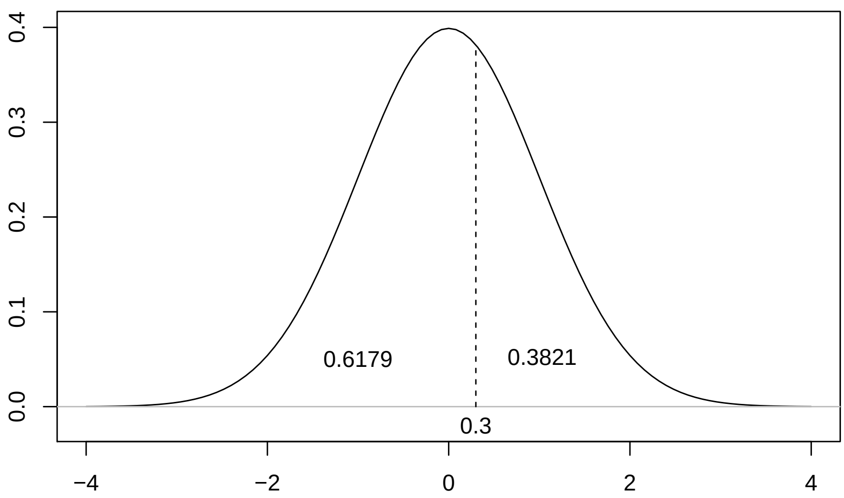
(ii) $P(Z > 0.3)$

Solution: From the output we have that

$$P(Z \leq 0.3) \approx 0.6179$$

and so

$$P(Z > 0.3) = 1 - P(Z \leq 0.3) \approx 0.3821\,.$$



(iii) $P(-0.72 < Z < 0.72)$

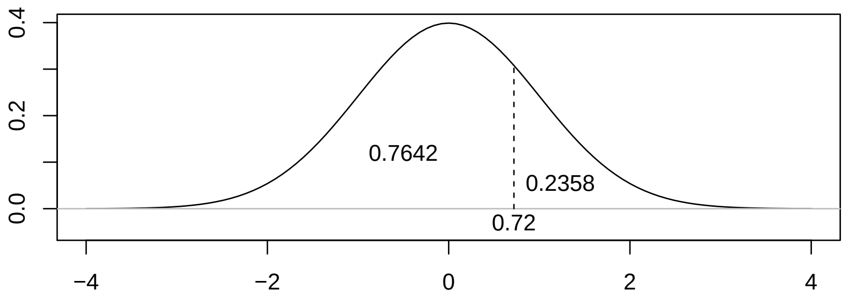
Solution: Note that the desired probability can be expressed as

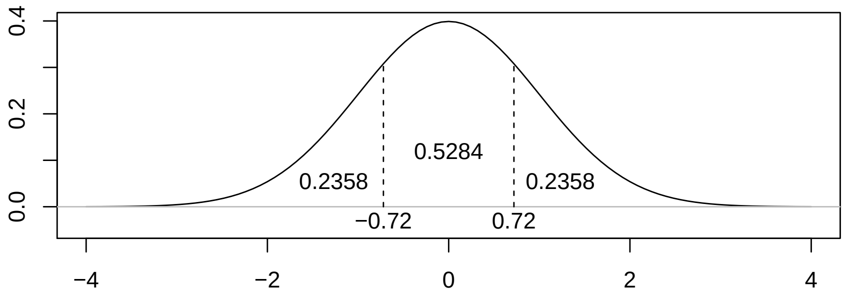
$$\begin{aligned} 1 - P(Z \leq -0.72) - P(Z \geq 0.72) &= 1 - 2P(Z \geq 0.72) \text{ (by symmetry)} \\ &= 1 - 2[1 - P(Z < 0.72)] \\ &= 1 - 2[1 - P(Z \leq 0.72)] \end{aligned}$$

since Z has a continuous distribution, and thus $P(Z = 0.72) = 0$.
Therefore the answer is (approximately, after rounding)

$$1 - 2(1 - 0.7642) = 1 - (2 \times 0.2358) = 1 - 0.4716 = 0.5284\,.$$

See also the graphs below:





(iv) $P(|Z| > 1.96)$.

Solution: The absolute values sometimes cause confusion. One way to mitigate this, in this example, is to consider the *complementary* event:

$$\begin{aligned} P(|Z| > 1.96) &= 1 - P(|Z| \leq 1.96) \\ &= 1 - P(-1.96 \leq Z \leq 1.96)\,. \end{aligned}$$

The complementary probability $P(-1.96 \leq Z \leq 1.96)$ can be determined using the same approach as the previous question:

$$P(-1.96 \leq Z \leq 1.96) = 1 - 2[1 - P(Z \leq 1.96)]$$

So we see that the desired answer is one minus this, that is

$$2[1 - P(Z \leq 1.96)] \approx 2(1 - 0.9750) = 0.05\,.$$

(b) Use the information above to find (to 3 decimal places) z such that

(i) $P(Z \leq z) = 0.90$

Solution: Reading directly from the output, since $z = $ `qnorm(p)` $\Leftrightarrow P(Z \leq z) = $ `p`, the desired $z \approx 1.282$.

(ii) $P(Z > z) = 0.95$

Solution: Note that the desired z also satisfies

$$P(Z \leq z) = 0.05\,.$$

Thus the desired z is negative, being in the *lower* tail of the $N(0, 1)$ distribution. By symmetry, *minus* that value will satisfy the property:

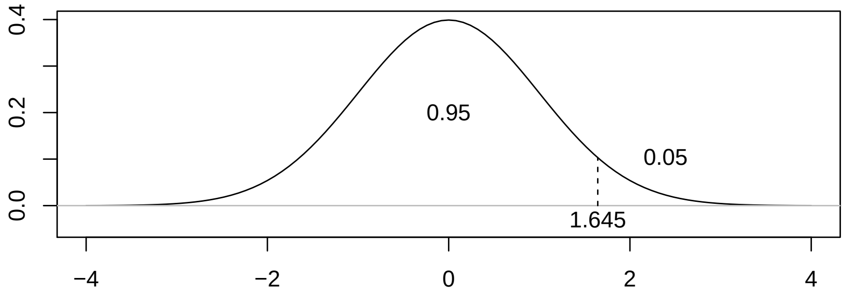
$$P(Z \geq -z) = 0.05\,,$$

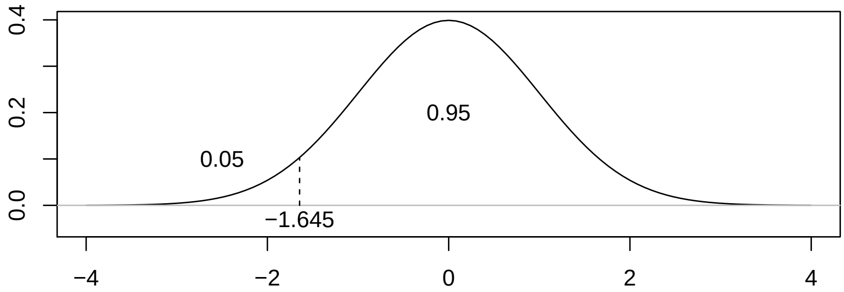
and thus also

$$P(Z < -z) = P(Z \le -z) = 0.95\,,$$

Thus we have

$$-z \approx 1.645\,, \quad \text{and thus} \quad z \approx -1.645\,.$$





(iii) $P(|Z| < z) = 0.90$.

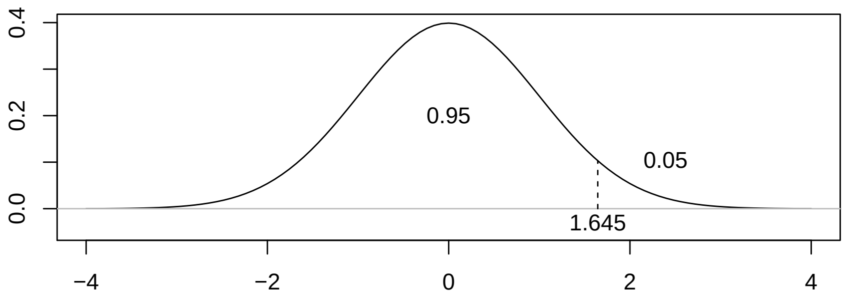
Solution: Using the same argument as in (a)(iii) above,

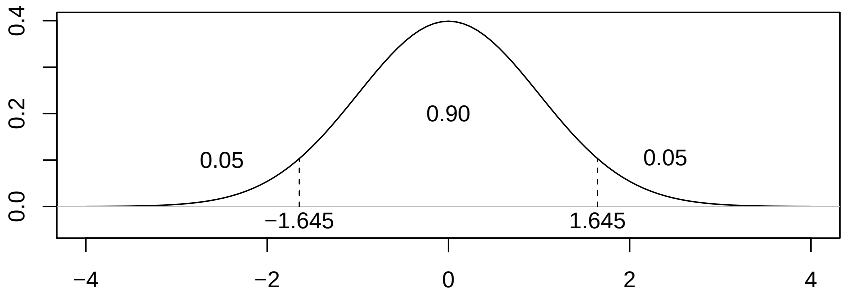
$$0.90 = P(|Z| < z) = P(-z < Z < z) = 1 - 2[1 - P(Z < z)]$$

and so

$$P(Z < z) = (1 - 0.90)/2 = 0.05\,,$$

and so $z \approx 1.645$; see also the graphs below:





(c) If $X \sim N(10, 16)$, use the information above to find

(i) $P(X > 12)$

Solution: The important thing to note is that

$$X \sim N(10, 16) \Leftrightarrow Z = \frac{X - 10}{\sqrt{16}} = \frac{X - 10}{4} \sim N(0, 1)\,;$$

(this follows from the *definition* of a $N(\mu, \sigma^2)$ random variable: $X \sim N(\mu, \sigma^2)$ *if and only if* $Z = (X - \mu)/\sigma \sim N(0, 1)$). So

$$\begin{aligned} P(X > 12) &= P\left(\frac{X - 10}{4} > \frac{12 - 10}{4}\right) \\ &= P\left(Z > \frac{1}{2}\right) \\ &= 1 - P(Z \le 0.5)\,. \end{aligned}$$

Reading directly from the output above part (a), $P(Z \le 0.5) \approx 0.6915$, so

$$P(X > 12) \approx 0.3085\,.$$

(ii) $P(X < 14)$

Solution:

$$P(X < 14) = P\left(\frac{X - 10}{4} < \frac{14 - 10}{4}\right) = P(Z < 1) \approx 0.8413\,,$$

reading directly from the R output above part (a).

(iii) $P(8 < X < 13)$

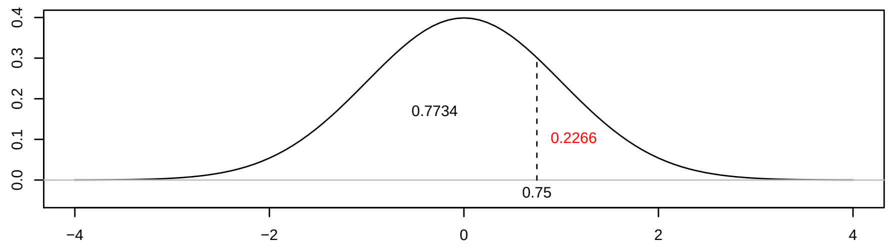
Solution:

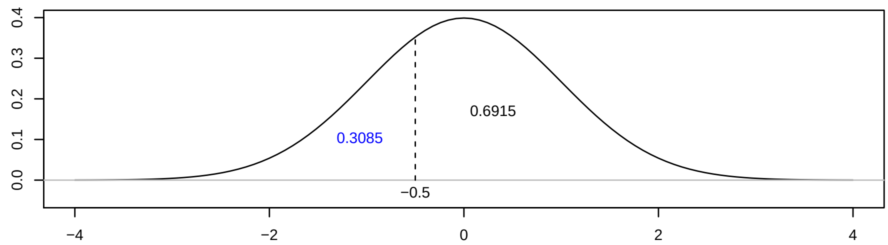
$$\begin{aligned}
P(8 < X < 13) &= P\left(\frac{8-10}{4} < \frac{X-10}{4} < \frac{13-10}{4}\right) \\
&= P(-0.5 < Z < 0.75) \\
&= P(Z < 0.75) - P(Z \leq -0.5) \\
&= P(Z < 0.75) - P(Z \geq 0.5) \quad \text{(by symmetry)} \\
&= P(Z < 0.75) - [1 - P(Z < 0.5)] \\
&= P(Z < 0.75) + P(Z < 0.5) - 1\,.
\end{aligned}$$

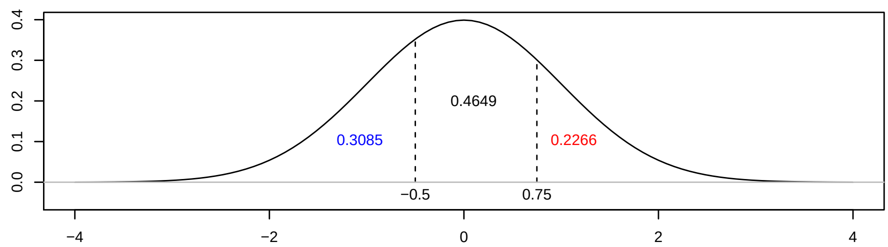
Reading directly from the R output, this is (approximately, after rounding),

$$0.7734 + 0.6915 - 1 = 0.4649\,.$$

See also the graphs below:







7. Glaucoma is a disease of the eye that is manifested by high intraocular pressure. The distribution of intraocular pressure in unaffected adults is approximately normal with mean 16 mm Hg and standard deviation 4 mm Hg.

(a) If the normal range for intraocular pressure (in mm Hg) is considered to be 12 − 20, what percentage of unaffected adults would fall within this range?

Solution: Let X denote intraocular pressure, $X \sim N(16, 4^2)$. $P(12 \leq X \leq 20) = 0.6826$ i.e. about 68% of unaffected adults are in this range.

(b) An adult is considered to have *abnormally high* intraocular pressure if the pressure reading is in the top percentile (1 percent) for unaffected adults. Determine pressures considered to be abnormally high.

Solution: We want c such that $P(X > c) = 0.01$: In R we would obtain this using

```
c=qnorm(1-0.01,16,4)
c
```

```
[1] 25.30539
```

Pressures above 25.31mm Hg are considered abnormally high.

8. Suppose the random variable X has probability distribution given by

$$P(X = x) = p(1-p)^x\,, \quad \text{for } x = 0, 1, 2, \ldots$$

for some $0 < p < 1$. Then X has a geometric distribution, but this is the version describing the *number of failures before the first success* in a sequence of independent success/failure trials, where the success probability at each trial is p.

Show that the probability generating function $\pi_X(s) = E\left(s^X\right)$ is given by

$$\pi_X(s) = \frac{p}{1 - s(1-p)}$$

so long as $|s| < 1/(1-p)$.

Solution:

$$\begin{aligned}\pi_X(s) &= \sum_{x=0}^{\infty} s^x P(X = x)\\ &= \sum_{x=0}^{\infty} s^x p(1-p)^x\\ &= p\sum_{x=0}^{\infty}[s(1-p)]^x\,.\end{aligned}$$

The infinite sum here is a geometric series of the form $1 + q + q^2 + \cdots$ which equals $\frac{1}{1-q}$ for all $|q| < 1$. Thus here, if $|s(1-p)| = |s|(1-p) < 1$ the infinite sum is $\frac{1}{1-s(1-p)}$, giving the desired result.

9. Suppose that X_1, X_2 and X_3 are independent random variables all of which have the same distribution as X in the previous question, i.e for $i = 1, 2, 3$ and each $x = 0, 1, 2,$

$$P(X_i = x) = p(1-p)^x\,.$$

Define the sum $Y = X_1 + X_2 + X_3$. We are going to derive $P(Y = 3)$ in two ways:

- directly;
- using probability generating functions.

(a) Enumerate all possible triples (x_1, x_2, x_3) where

- each x_i is a non-negative integer;
- $x_1 + x_2 + x_3 = 3$.

Hence compute $P(Y = 3)$.

Solution: We may list all possible triples as follows:

- (3,0,0), (0,3,0), (0,0,3);
- (2,1,0), (2,0,1), (0,2,1), (1,2,0), (1,0,2), (0,1,2);
- (1,1,1).

The first 3 all have the same probability, which is

$$p(1-p)^3 \times p \times p = p^3(1-p)^3\,;$$

the next 6 all have the same probability, which is also

$$p(1-p)^2 \times p(1-p) \times p = p^3(1-p)^3\,;$$

The final one also has probability

$$p(1-p) \times p(1-p) \times p(1-p) = p^3(1-p)^3 .$$

Thus the answer is

$$10p^3(1-p)^3 .$$

(b) Writing $\pi_X(s)$ for the probability generating function of X in question 8 above, the probability generating function of $Y = X_1 + X_2 + X_3$ is given by

$$\pi_Y(s) = E\left(s^Y\right) = E\left(s^{X_1+X_2+X_3}\right) = E\left(s^{X_1}\right) E\left(s^{X_2}\right) E\left(s^{X_3}\right) = [\pi_X(s)]^3 = \left[\frac{p}{1-s(1-p)}\right]^3 .$$

Differentiate this three times, and hence determine $P(Y = 3)$.

Solution: Write the PGF as $\pi_Y(s) = p^3[1-s(1-p)]^{-3}$. The first derivative is

$$\begin{aligned}\pi_Y'(s) &= p^3(-3)[1-s(1-p)]^{-4}[-(1-p)] \\ &= 3p^3(1-p)[1-s(1-p)]^{-4} .\end{aligned}$$

The second derivative is

$$\begin{aligned}\pi_Y''(s) &= 3p^3(1-p)(-4)[1-s(1-p)]^{-5}[-(1-p)] \\ &= 12p^3(1-p)^2[1-s(1-p)]^{-5} .\end{aligned}$$

The third derivative is

$$\begin{aligned}\pi_Y'''(s) &= 12p^3(1-p)^2(-5)[1-s(1-p)]^{-6}[-(1-p)] \\ &= 60p^3(1-p)^3[1-s(1-p)]^{-6} .\end{aligned}$$

Since, in general

$$\pi_Y^{(k)}(0) = k!P(Y = k) ,$$

we have that

$$P(Y = 3) = \frac{\pi'''(0)}{3!} = \frac{60p^3(1-p)^3}{6} = 10p^3(1-p)^3 ,$$

agreeing with part (a).

10. Using R, find the exact probability $P(X \leq 10)$ for $X \sim B(20, 0.6)$. Find the corresponding normal approximation with continuity correction (**hint**: if you are unsure whether to "add $\frac{1}{2}$" or "subtract $\frac{1}{2}$", note that since X is integer-valued, $P(X \leq 10) = P(X < 11)$).

Solution: The exact binomial probability can be found using the following commands:

```
x=0:10
bin.probs=dbinom(x,20,0.6)
cbind(x,bin.probs)
```

```
       x    bin.probs
 [1,]  0 1.099512e-08
 [2,]  1 3.298535e-07
 [3,]  2 4.700412e-06
 [4,]  3 4.230371e-05
 [5,]  4 2.696862e-04
 [6,]  5 1.294494e-03
 [7,]  6 4.854351e-03
 [8,]  7 1.456305e-02
 [9,]  8 3.549744e-02
[10,]  9 7.099488e-02
[11,] 10 1.171416e-01
```

```
sum(bin.probs)
```

```
[1] 0.2446628
```

The approximating normal random variable is $Y \sim N(12, 4.8)$ (having the same expectation and variance as X). Note that because X only takes integer values, we have

$$P(X \leq 10) = P(X < 11) = P(X \leq 10.5) = P(X < 10.5)$$

These last two are equal because $P(X = 10.5) = 0$. The two "naïve" normal approximations are $P(Y \leq 10)$ which is given by

```
pnorm(10,12,sqrt(4.8))
```

```
[1] 0.1806552
```

which is clearly too small and

```
pnorm(11,12,sqrt(4.8))
```

```
[1] 0.3240384
```

which is clearly too big. The approximation with the appropriate correction for continuity is $P(Y \leq 10.5)$ which is

```
pnorm(10.5,12,sqrt(4.8))
```

```
[1] 0.2467814
```

which is very close to the true value.