# 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 $h(\cdot)$,

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

**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) $N(50, 0.02)$ (c) $N(1, 0.98)$ (e) $N(0.01, 0.098)$
(b) $N(50, 1)$ (d) $N(1, 0.0098)$

**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) $N(100, 25)$ (c) $N(0, 25)$ (e) $N(0, 1)$
(b) $N\left(100, \frac{5}{4}\right)$ (d) $N\left(100, \frac{25}{16}\right)$

**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 | 0 | +1 |
| | -1 | 1/16 | 3/16 | 1/16 |
| $x_2$ | 0 | 3/16 | 0 | 3/16 |
| | +1 | 1/16 | 3/16 | 1/16 |

(a) Find the marginal distributions of $X_1$ and $X_2$.
(b) Show that $X_1$ and $X_2$ are **not** independent.
(c) Evaluate $E(X_1)$, $E(X_2)$ and $E(X_1X_2)$.
(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).

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

**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?

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

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

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

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

(i) $P(X > 12)$

**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?

(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.

**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 $|s| < 1/(1-p)$.

**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 $x = 0, 1, 2, \ldots$,

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

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 $P(Y = 3)$.

(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 $P(Y = 3)$.

**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, $P(X \leq 10) = P(X < 11)$).