The University of Sydney
School of Mathematics and Statistics

Solutions to Tutorial Week 5

MATH1905: Statistics (Advanced) — Semester 2, 2017

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

For a discrete random variable X we have

$$E(X) = \sum_x xP(X = x)\,,\ E(X^2) = \sum_x x^2 P(X = x)\,,\ Var(X) = E(X^2) - [E(X)]^2\,.$$

1. (**Multiple Choice**) For $X \sim B(8, 0.1)$, $P(X \leq 2$ is closest to

(a) 0.9950

(b) 0.8131

(c) 0.7969

(d) 0.9619

(e) 0.6259

Solution: There are various ways to compute this using R. Perhaps the quickest is

```
pbinom(2,8,0.1)
```

```
[1] 0.9619082
```

but a slightly more instructive way is

```
x=0:2
x
```

```
[1] 0 1 2
```

```
dbinom(x,8,0.1)
```

```
[1] 0.4304672 0.3826375 0.1488035
```

```
sum(dbinom(x,8,0.1))
```

```
[1] 0.9619082
```

Thus the correct answer is (d).

2. The following table shows the probability distribution of a random variable X:

x	1	2	3	4	Total
$P(X = x)$	0.35	0.30	0.25	0.10	1.00

Find $E(X)$, $E(1/X)$, $E(X^2)$, $Var(X)$. Verify that $E(1/X) \neq 1/E(X)$.

Solution:

$$\begin{aligned}
E(X) &= 1 \times .35 + 2 \times .30 + 3 \times .25 + 4 \times .10 = 2.1 \\
E(1/X) &= 1 \times .35 + \frac{1}{2} \times .30 + \frac{1}{3} \times .25 + \frac{1}{4} \times .10 = 0.6083 \neq 1/E(X) = 1/2.1 \approx 0.476 \\
E(X^2) &= 1^2 \times .35 + (2)^2 \times .30 + (3)^2 \times .25 + (4)^2 \times .10 = 5.4 \\
Var(X) &= E(X^2) - [E(X)]^2 = 5.4 - (2.1)^2 = 0.99
\end{aligned}$$

3. Let X be a discrete random variable with the following incomplete probability distribution table:

x	0	1	2	3	4
$P(X = x)$	0.17	0.36	0.31		0.03

Find

(a) $P(X = 3)$

Solution: Since the probabilities add to 1, $P(X = 3) = 1 - 0.17 - 0.36 - 0.31 - 0.03 = 0.13$.

(b) $E(X)$

Solution: $E(X) = 0 \times 0.17 + 1 \times 0.36 + 2 \times 0.31 + 3 \times 0.13 + 4 \times 0.03 = 1.49$.

(c) $E(X^2)$

Solution: $E(X^2) = 0^2 \times 0.17 + 1^2 \times 0.36 + 2^2 \times 0.31 + 3^2 \times 0.13 + 4^2 \times 0.03 = 3.25$.

(d) $Var(X)$

Solution: $Var(X) = E(X^2) - [E(X)]^2 = 3.25 - 1.49^2 = 1.03$.

4. When Mendel crossed a tall strain of pea with a dwarf strain of pea, he found that $\frac{3}{4}$ of the offspring were tall and $\frac{1}{4}$ were dwarf. Suppose five such offspring were selected at random. Let X be the number of 'tall offspring' in this random sample. Find the probability distribution of X by completing the following table:

x	
$P(X = x)$	

Verify that $\sum_x P(X = x) = 1$.

Solution: Assuming that the tall/dwarf attribute of each plant is determined *independently from each other* and that each is tall(resp. dwarf) with probability $\frac{3}{4}$(resp. $\frac{1}{4}$), then the number of tall offspring in a random sample of 5 has a binomial$(5, \frac{3}{4})$ distribution, so for $x = 0, 1, \ldots, 5$,

$$P(X = x) = \binom{5}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{5-x} = \frac{\binom{5}{x} 3^x}{4^5}.$$

The binomial coefficients $\binom{5}{x}$ for $x = 0, 1, \ldots, 5$ are (respectively) $1, 5, 10, 10, 5, 1$. This yield the following table:

x	0	1	2	3	4	5
$P(X = x)$	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$

To check that these add to 1 it suffices to add the numerators:

$$1 + 15 + 90 + 270 + 405 + 243 = 1024.$$

5. Use R to verify the answers to the previous question

Solution:

```
x=0:5
x
```

```
[1] 0 1 2 3 4 5
```

```
dbinom(x,5,0.75)
```

```
[1] 0.0009765625 0.0146484375 0.0878906250 0.2636718750 0.3955078125
[6] 0.2373046875
```

```
1024*dbinom(x,5,0.75)
```

```
[1]   1  15  90 270 405 243
```

6. In a small pond there are 50 fish, 20 of which have been tagged. Seven fish are caught and X represents the number of tagged fish in the catch.

(a) Under what additional conditions is X well-modelled by a binomial random variable? Write a formula for $P(X = x)$ and indicate all possible values x.

Solution: If

- each fish in the pond has the same chance of being caught at a the first attempt;
- each fish caught is then returned to the pond and any subsequent fish are caught "independently", i.e. enough time is allowed to pass so that (again) all fish are equally likely to be caught at the next attempt,

then X would be well-modelled as a $B(7, 0.4)$ random variable, with distribution given by

$$P(X = x) = \binom{7}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{7-x} = \frac{\binom{7}{x}2^x 3^{7-x}}{5^7},$$

for $x = 0, 1, \ldots, 7$.

(b) Under what additional conditions is X well-modelled by a hypergeometric random variable? Write a formula for $P(X = x)$ and indicate all possible values x.

Solution: If each possible sample (taken **without replacement**) of 7 fish is equally likely then X is well-modelled by a hypergeometric random variable with distribution given by

$$P(X = x) = \frac{\binom{20}{x}\binom{30}{7-x}}{\binom{50}{7}},$$

for $x = 0, 1, \ldots, 7$ (no extra "boundary constraints" come into play since the number of both tagged and untagged in the "population" exceeds the sample size).

Conditions "implying" this might be that *at each attempt* all fish actually in the pond have the same chance of being caught. This would result in any possible *ordered* sample having probability

$$\frac{1}{50} \times \frac{1}{49} \times \cdots \times \frac{1}{44} = \frac{1}{{}^{50}P_7}$$

which is equivalent to each possible *unordered* sample having probability

$$\frac{7!}{{}^{50}P_7} = \frac{1}{\binom{50}{7}}.$$

(c) Find the probability of exactly one tagged fish being in the sample of 7 in each case.

Solution: In the binomial case:

```
(7*2*(3^6))/(5^7)
```

```
[1] 0.1306368
```

or just use

```
dbinom(1,7,2/5)
```

```
[1] 0.1306368
```

In the hypergeometric case:

```
(20*choose(30,6)/choose(50,7))
```

```
[1] 0.1188924
```

or use

```
dhyper(1,20,30,7)
```

```
[1] 0.1188924
```

7. (a) Suppose that for some fixed positive integer n, for $x = 1, 2, \ldots, n$, the random variable X has distribution given by $P(X = x) = 1/n$. Find

(i) $E(X)$;

Solution: $E(X) = \sum_{i=x}^{n} xP(X = x) = \frac{1}{n}\sum_{i=1}^{n} x = \frac{n+1}{2}$.

(ii) $Var(X)$.

Hint: you may find the following results useful:

$$\sum_{x=1}^{n} x = \frac{n(n+1)}{2}, \quad \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}.$$

Solution: $E(X^2) = \sum_{x=1}^{n} x^2 P(X = x) = \frac{1}{n}\sum_{i=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6}$. Thus $\mathrm{V}ar(X) = \frac{(n+1)(n-1)}{12}$.

(b) If the distribution of X is given by $P(X = x) = C_n x$, $x = 1, 2, 3, \ldots, n$, find C_n.

Solution: Using $\sum_{x=1}^{n} P(X = x) = 1$, we have $1 = C_n \sum_{x=1}^{n} x = \frac{C_n n(n+1)}{2}$. Thus $C_n = \frac{2}{n(n+1)}$.

8. Recall that the expectation of a geometric random variable X with

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

is given by

$$p\left\{1 + 2(1-p) + 3(1-p)^2 + \cdots\right\}. \tag{†}$$

We shall show that in fact $E(X) = 1/p$ by proving that the curly-bracketed expression is $1/p^2$.

(a) For $0 < q < 1$ we know that

$$1 + q + q^2 + \cdots = \frac{1}{1-q}.$$

By differentiating both sides evaluate the infinite sum $1 + 2q + 3q^2 + \cdots$.

Solution: Differentiating both sides we get

$$1 + 2q + 3q^2 + \cdots = \frac{1}{(1-q)^2}. \tag{$*$}$$

(b) Hence show that $E(X) = 1/p$.

Solution: Substitute $q = 1 - p$ in $(*)$ above to get

$$1 + 2(1-p) + 3(1-p)^2 + \cdots = \frac{1}{p^2}.$$

Substituting this into the expression (†) above for $E(X)$ gives $1/p$.