Solutions to Assignment 4

1. In a population of people in a city, the distribution of heights of individuals (crudely measured using a scaled with least count 2 cm) is given as per the following table:

Height (in cm)	160	162	164	166	168	170	172	174	176	178	180
People (in 1000's)	5	10	25	50	100	120	140	100	50	10	5

X denotes the random variable "the height of a randomly chosen person from the city where each person is equally likely to be chosen". Calculate the following:

1. The mathematical expectation $E(X)$ (also called $\mu(X)$).
2. The most likely height.
3. The smallest number h so that at least 50% of the population has neight less than or equal to h.
4. The variance $\sigma^2(X)$ and the standard deviation $\sigma(X)$.

Use a calculator if necessary. Also, try to understand the way in which X is defined.

Solution: The idea is to just use the formulas for $E(X)$ and $\sigma^2(X)$ to get (approximately) 170.699 and (approximately) 13.072. (Note that the latter calculation can be simplified by calculating $\sigma^2(X - 170)$ which is the same!).

We see immediately that the most likely height is 172 cm since 140 is the largest frequency. The total population size is 615 Thousand. A rolling sum shows that the population of people with height less than or equal to 170 is 310 Thousand and that those with height less than or equal to 168 is only 190 Thousand.

2. A multiple choice paper has 10 questions. Each question has 4 choices. The correct answer gets 3 marks and a wrong answer gets -1 mark. A student throws a 4-sided unbiased die to answer each question (the different throws are independent). Let X be the random variable that denotes the score of the student in the examination.

1. Calculate the expected score $E(X)$.
2. Calculate the value of X for which the probability is the highest.
3. What is the smallest s so that $P(X \leq s) \geq 1/2$?
4. Calculate the variance $\sigma^2(X)$.

Use a calculator if necessary.

Solution: If there are r correct answers, there are $10 - r$ wrong answers and the score is $3r - (10 - r) = 4r - 10$. The probability of a correct answer is $1/4$ and the probability of a wrong answer is $3/4$. Hence, by the binomial distribution, the probability of r correct answers is

$$\binom{10}{r} 3^{10-r}/4^{10} = \binom{10}{r} (3/4)^{10}(1/3)^r$$

Hence, no calculator is actually required for the first half of the problem!

By the Binomial theorem,

$$(1+t)^{10} = \sum_{r=0}^{10} \binom{10}{r} t^r$$

Differentiating both sides with respect to t and then multiplying by t, we have

$$10t(1+t)^9 = \sum_{r=0}^{10} r\binom{10}{r} t^r$$

Doing the same operation once more we have

$$10t(1+t)^9 + 90t^2(1+t)^8 = \sum_{r=0}^{10} r^2\binom{10}{r} t^r$$

The second formula gives us, on putting $t = 1/3$

$$\sum_{r=0}^{10} r\binom{10}{r}(1/3)^r = (10/3)(4/3)^9$$

or equivalently,

$$E(X+10) = \sum_{r=0}^{10} 4r\binom{10}{r}(3/4)^{10}(1/3)^r = 4(3/4)(10/3) = 10$$

In other words, as we might expect(!) we get $E(X) = 0$. Similarly, we can use the third formula with $t = 1/3$ to get

$$\sum_{r=0}^{10} r^2\binom{10}{r}(1/3)^r = (10/3)(4/3)^9 + 10(4/3)^8$$

or equivalently,

$$E((X+10)^2) = \sum_{r=0}^{10} 16r^2\binom{10}{r}(3/4)^{10}(1/3)^r = 16\left((3/4)(10/3) + 10(3/4)^2\right) = 130$$

This gives the variance $\sigma^2(X) = \sigma^2(X+10) = 30$. Hence, the standard deviation is $\sigma(X) = \sqrt{30} = 5.477$.

The table of probabilities is (this may require a calculator!):

r	0	1	2	3	4	5	6	7	8	9	10
$4^{10} * p$	59049	196830	295245	262440	153090	61236	17010	3240	405	30	1

We note that to get at least 50% of the population we can take $r \leq 2$. Since the score is then $4r - 10 \leq -2$ it means that more than 50% of the time you will get a negative score!

3. A die is rolled repeated until we get a 6. The number of rolls is recorded. Let X denote the random variable that denotes the number of rolls.

    1. Calculate the expectation $E(X)$.
    2. Calculate the variance $\sigma^2(X)$.
    3. Calculate the value of X for which the probability is the highest.
    4. What is the smallest s so that $P(X \leq s) \geq 1/2$?

Solution: We have seen that $P(X = n) = (5/6)^{n-1}(1/6)$. To calculate $E(X)$ and $\sigma^2(X)$ we need to calculate sums of the form $\sum_{n=0}^{\infty} n^k(1-t)^n$ for $t = 1/6$. We have

$$\frac{1}{t} = \sum_{n=0}^{\infty}(1-t)^n$$

Differentiating both sides, changing sign and renumbering

$$\frac{1}{t^2} = \sum_{n=0}^{\infty}(n+1)(1-t)^n$$

Repeating this gives

$$\frac{2}{t^3} = \sum_{n=0}^{\infty}(n+2)(n+1)(1-t)^n$$

This gives the formulae

$$\sum_{n=0}^{\infty} n(1-t)^n = \frac{1}{t^2} - \frac{1}{t} = \frac{1-t}{t^2}$$

and

$$\sum_{n=0}^{\infty} n^2(1-t)^n = \frac{2}{t^3} - \frac{3}{t^2} + \frac{1}{t} = \frac{(2-t)(1-t)}{t^3}$$

We can now apply this to calculate

$$E(X) = \sum_{n=1}^{\infty} n(5/6)^{n-1}(1/6) = (1/5)\sum_{n=0}^{\infty} n(1-(1/6))^n = (1/5)\frac{5/6}{1/6^2} = 6$$

Another calculation gives

$$E(X^2) = \sum_{n=1}^{\infty} n^2(5/6)^{n-1}(1/6) = (1/5)\sum_{n=0}^{\infty} n^2(1-(1/6))^n = (1/5)\frac{(11/6)(5/6)}{1/6^3} = 66$$

This gives

$$\sigma^2(X) = E(X^2) - E(X)^2 = 66 - 36 - 30$$

We see easily that $P(X = 1)$ is the largest among all probabilities as the terms of the series are decreasing!

To get the median, we need to use the formula $\sum_{n=1}^{N} t(1-t)^{n-1} = 1-(1-t)^N$. This gives us

$$\sum_{n=1}^{N}(1/6)(5/6)^{n-1} = 1 - (5/6)^N$$

In order for this to be at least half, we need $(5/6)^N \leq 1/2$. Taking log of both sides gives $N\log(5/6) \leq -\log(2)$. Thus, we need $N \geq \log(2)/\log(6/5) = 3.80$ or $N \geq 4$.

4. Let X denote the random variable the denotes the sum of the numbers obtained on rolling two dice.

   1. Calculate the expectation $E(X)$.
   2. Calculate the value of X for which the probability is the highest.
   3. What is the smallest s so that $P(X \leq s) \geq 1/2$?
   4. Calculate the variance $\sigma^2(X)$.

Solution: During the previous assignment we had seen that the table of probabilities is as follows

Sum	2	3	4	5	6	7	8	9	10	11	12
$36p$	1	2	3	4	5	6	5	4	3	2	1

From this we see that the probability of 7 is the highest. The probability that the sum is less than 7 is at least 50% (whereas the probability that the sum is less than 6 is less than 50%).

The mathematical expectation and the variance can be easily calculated from the above table. We get $E(X) = 7$ and $\sigma^2(X) = 35/6$.

5. • Player A plays a game where he wins 2 rupees with each Head and loses 1 rupee with each Tail; X is the random variable measuring the money won by A in a single game.

- Player B plays a game where he wins one rupee with each Head and loses one rupee with each tail but the coin has $2/3$ chance of getting head; Y is the random variable measuring the money won by B in a single game.

Calculate and compare the mathematical expectation, median, mode, variance of the variables X and Y.

Solution: We get $E(X) = 2(1/2) + (-1)(1/2) = 1/2$. We have $E(X^2) = 4(1/2) + 1(1/2) = 5/2$, so $\sigma^2(X) = 9/4$. The probability of both Head and Tail is the same so there is no well-defined mode. The median of X is -1 since $P(X \leq -1) = P(X = -1) = 1/2$.

We get $E(Y) = (2/3) + (-1)(1/3) = 1/3$. We have $E(Y^2) = (2/3) + (1/3) = 1$, so $\sigma^2(Y) = 1 - 1/9 = 10/9$. The probability of Head is greater so it $Y = 1$ is the mode. The median of Y is 1 since $P(Y \leq -1) = P(X = -1) = 1/3$ and $P(Y \leq 1) = 1$.