IMPORTANT: The answers are provided with the view to deepen your understanding of the topic and to stimulate more in depth discussions. They should not be shared with others or put on the web. The questions and answers may be revised in the next course delivery.

Answers to Tutorial 4

Qn 1

a) Let X_1 and X_2 be the number obtained for the first and second die respectively. Linearity of expectation holds no matter what is the relationship between the two random variables. Hence

$$E[X_1 + X_2] = E[X_1] + E[X_2] = 3.5 + 3.5 = 7$$

b) Since we may assume that the number obtained for the first and the second die are not independent,

$$Var[X_1 + X_2] = Var[X_1] + Var[X_2] + Cov(X_1, X_2) + Cov(X_2, X_1)$$
$$= \frac{35}{12} + \frac{35}{12} + 1.2 + 1.2 = 8.233333333$$

Standard deviation = $\sqrt{8.233333333} = 2.869378562 \approx 2.87$

Qn 2

a)

Let X denote the number of different Pokémon in 100 successful catches.

$$X = X_1 + \cdots + X_{151}$$

$$X_i = \begin{cases} 1 & if\ at\ least\ one\ type\ i\ \text{Pokémon}\ is\ in\ the\ 100\ successful\ catches \\ 0 & otherwise \end{cases}$$

$$\begin{aligned} E[X_i] &= 1P\{X_i = 1\} + 0P\{X_i = 0\} = P\{X_i = 1\} \\ &= P\{at\ least\ one\ type\ i\ Pokémon\ is\ in\ 100\ successful\ catches\} \\ &= 1 - P\{no\ type\ i\ Pokémon\ are\ in\ 100\ successful\ catches\} \\ &= 1 - \left(\frac{150}{151}\right)^{100} = 0.485445742 \end{aligned}$$

$$E[X] = E[X_1] + \cdots + E[X_{151}] = 151\left[1 - \left(\frac{150}{151}\right)^{100}\right] = 73.30230709 \approx 73$$

b)

One cannot simply put $E[X] = 151$

$151\left[1 - \left(\frac{150}{151}\right)^n\right] = 151 \Rightarrow n = \infty$

This is correct as to have expected value of 151, we need to have an infinite number of catches, as there is a very small chance we fail to catch a Pokémon after a large number of attempts.

We can find the expected number of attempts by the celebrated Coupon Collector's Theorem
The key to solving the problem is based on the observation that it takes very little time to collect the first few Pokémon, but it takes a long time to collect the last few Pokémon.

Let n be the number of different types of Pokémon. Let X be the number of catches until at least one of every type of Pokémon is caught.

Let X_i be the number of catches until a new Pokémon is caught while you have already collected $i-1$ different types of Pokémon. When exactly $i-1$ different types of Pokémon have been collected, the probability of obtaining a new type of Pokémon is

$$p_i = 1 - \frac{i-1}{n}$$

Hence X_i is a geometric random variable with parameter p_i, and

$$E[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

Using linearity of expectations,

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} E[X_i] \\ &= \sum_{i=1}^{n} \frac{n}{n-i+1} \\ &= n\sum_{i=1}^{n} \frac{1}{i} \end{aligned}$$

$\sum_{i=1}^{n} \frac{1}{i}$ is known as the harmonic number $H(n)$.

In our question, $n = 151$.

Using the harmonic series calculator
(https://www.math.utah.edu/~carlson/teaching/calculus/harmonic.html)

$E[X] = (151)(5.597803105200172) \approx 846$

References

1. M. Mitzenmacher and E. Upfal, Probability and Computing, Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, 2005.

2. https://en.wikipedia.org/wiki/Coupon_collector%27s_problem

c) It is known that some Pokémon can only be caught at certain districts in the world. For example, the Kangaskhan Pokémon can only be caught in areas near Australia. Thus the chance of catching a Pokémon is not equally likely.

Also, certain Pokémon only become available at a specific place at a specified period of time. Moreover, certain Pokémon only becomes catchable by invitation.

Qn 3

a)

From the web page, the expectation of life at birth by sex is

Expectation of life at birth by sex

Sex	*Years*		
	2014	2018	2019
Male	81.2	82.3	82.2
Female	86.9	87.7	88.1

Last revision date : 9 September 2020

Using 2019 male data, $E[X] = 82.2$. Using Markov's inequality,

$$P\{X \geq 100\} \leq \frac{82.2}{100} = 0.822 = 82.2\%$$

Using 2019 female data,

$$P\{X \geq 100\} \leq \frac{88.1}{100} = 0.881 = 88.1\%$$

b)

$$P\{|X - E[X]| \geq 20\} \leq \frac{5^2}{20^2} \approx 6.25\%$$

c)

$$P\{X - E[X] \geq \lambda\} \begin{cases} \leq \dfrac{\sigma^2}{\sigma^2 + \lambda^2} & if\ \lambda > 0 \\ \geq 1 - \dfrac{\sigma^2}{\sigma^2 + \lambda^2} & if\ \lambda < 0 \end{cases}$$

For male, $E[X] = 82.2$, $\lambda = 100 - 82.2 = 17.8$, $\sigma = 5$

$P\{X \geq 100\} \leq 0.073133629 \approx 7.3\%$

For female, $E[X] = 88.1$, $\lambda = 100 - 88.1 = 11.9$, $\sigma = 5$

$P\{X \geq 100\} \leq 0.150051017 \approx 15.0\%$

d)

The Markov's inequality can only provide very rough estimates. When the variance is also available, the quality of the estimate is better.

One can get the statistics of the mean death age at different years. Then use regression to estimate the expected life at birth. We shall study the topic of regression later on in the course.

e)

On average, female lives longer than male. As the mean age of Hong Kong population is increasing, the weighting of older people gets higher and higher. Thus, the sex ratio is dropping.

Or any other reasonable explanation. Can you identify other meaningful variables?