Solutions to Assignment 4

1. Calculate the following expressions and the error in comparison with the target expression.

   (a) $\sum_{k=0}^{5} \frac{1}{3^k}$ for the target $1/(1-\frac{1}{3})$.

   Solution: We know that the error is given by

   $$\sum_{k=6}^{\infty} \frac{1}{3^k} = \frac{1}{3^6} \sum_{k=0}^{\infty} \frac{1}{3^k} = \frac{1}{3^6(1-\frac{1}{3})} = \frac{1}{3^5 \cdot 2}$$

   This is 1/483 or approximately 0.002.

   (b) $\sum_{k=0}^{3} \frac{1}{k!}$ for the target $\exp(1)$.

   Solution: We know that the error is given by

   $$\sum_{k=4}^{\infty} \frac{1}{k!} \leq \frac{1}{4!} \sum_{k=0}^{\infty} 14^k = \frac{1}{24(1-\frac{1}{4})} = 1/18$$

   This is approximately 0.06.

   (c) $\sum_{k=0}^{4} \frac{1}{4^{k+1}(k+1)}$ for the target $-\log(1-\frac{1}{4})$.

   Solution: We know that the error is given by

   $$\sum_{k=5}^{\infty} \frac{1}{4^{k+1}(k+1)} = \frac{1}{4^6 6} \sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{1}{4^6 6(1-\frac{1}{4})} = \frac{1}{2^{11}3^2}$$

   This is $1/(1024 \cdot 18)$ or approximately 0.0006.

2. Suppose we want to calculate the value of log(2) using the formula

   $$\log(2) = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}(k+1)}$$

   How many terms will we use to calculate it correctly to 2 places of decimal? (Note that the error should be at most 1/200; justify this statement!) Use a calculator to calculate this value and compare it with the built-in value.

Solution: Since we want the second place of decimal to be correct as per rounding off rules, the error can be at most 0.005 or 1/200. Using the series given above we see that if we take $n+1$ terms the error is

$$\sum_{k=n+1}^{\infty} \frac{1}{2^{k+1}(k+1)} \leq \frac{1}{2^{n+2}(n+2)} \sum_{r=0}^{\infty} \frac{1}{2^r} = \frac{1}{2^{n+2}(n+2)(1-1/2)} = \frac{1}{2^{n+1}(n+2)}$$

In order for this to be at most 1/200, we need $n = 5$. Hence the sum $\sum_{k=0}^{5} 1/(2^{n+1}(n+1))$ gives an accurate value of $\log(2)$ upto 2 places of decimal.

3. Using the above value of $\log(2)$ and the following formula

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

to calculate $\log(3)$, $\log(5)$ accurate upto two places of decimal. (note that $\log(4) = 2\log(2)$!) Compare it with the built-in value on your calculator.

Solution: We can do a calculation similar to the previous exercise to see that the error due to summing only the first $n+1$ terms of the series for $\log(1-1/3)$ is $1/(3^{n+1}2(n+2))$. This is less than 1/200 if $n = 3$.

Similarly, the error due to summing only the first $n+1$ terms of the series for $\log(1-1/5)$ is $1/(5^{n+1}4(n+2))$ which is less than 1/200 if $n = 2$.

Now, if we are going to further add these values 2 at a time, then errors can add up. So it is probably better to get all three series upto an accuracy of 1/1000. After that, adding 2 or even 4 of the values will still not loose the accuracy of the digits in the 2nd place of decimal.

4. A die is tossed 500 times and S is the random variable that counts the number of 6's that occur.

   (a) Write a formula for $p_k = P(S = k)$, for $m = E(S)$ and $s = \sigma(S)$.

Solution: The random variable S follows the Binomial distribution with $p = 1/6$ and $n = 500$. So

$$p_k = P(S = k) = \binom{500}{k} \frac{5^{500-k}}{6^{500}}$$

Moreover

$$m = E(S) = 500 \cdot (1/6) \text{ and } s^2 = \sigma^2(S) = 500(1/6)(5/6)$$

So $m = 250/3$ and $s = 25/3$.

(b) Write a formula for the probability of that the number of 6's that we throw is betweem $m - 3s$ and $m + 3s$.

Solution: We note that $m-3s = 250/3-25 = 175/3$ and $m+3s = 250/3+25 = 325/3$. Since S takes integer values we see that we want $P(59 \leq S \leq 108)$. So the expression is

$$\sum_{k=59}^{108} p_k$$

(c) Assuming that we can use the normal distribution with mean m and standard deviation s to approximate this probability write an expression for it.

Solution: Let X follow the Normal distribution with mean 250/3 and standard deviation 25/3. The probability that X lies in the range 175/3 and 225/3 is

$$\frac{1}{(25/3)\sqrt{2\pi}} \int_{175/3}^{225/3} \exp\left(-\frac{(t-250/3)^2}{2\cdot(25/3)^2}\right) dt$$

We also see that $Y = (X-m)/s$ is a random variable with mean 0 and standard deviation. The given condition is the same as the condition that Y lies between -3 and 3. So the probability is

$$\frac{1}{\sqrt{2\pi}} \int_{-3}^{3} e^{-t^2/2} dt$$

5. A fair coin is flipped 1000 times and the number of heads is recorded. Give an integral formula for the probability that the number of heads lies in the range $[480, 520]$. Also write the estimate for this probability using the Chebychev formula.

   Suppose that we want to estimate the probability that the number of heads lies in the range $[490, 510]$. What is the problem with the estimate using Chebychev's inequality?

Solution: We will use the normal approximation to the binomial distribution.

We note that if S is the random variable that counts the number of heads, then $m = E(S) = 500$ and $s^2 = \sigma^2(S) = 500/4$. So $s = 5\sqrt{5}$.

If we take $X = (S-m)/s$, then X lies in the range $[-a, a]$ where $a = 4/\sqrt{5}$; note that $a \cdot s = 20$. So the normal approximation gives the probability of X in the range $[-a, a]$ as

$$\frac{1}{\sqrt{2\pi}} \int_{-4/\sqrt{5}}^{4/\sqrt{5}} e^{-t^2/2} dt$$

The Chebychev inequality can also be applied. The probability that S differs from its mean m by at least as is bounded above by $1/a^2 = 5/16$. Hence the probability that the number of heads lies in the given range is *at least* $11/16$ by Chebychev's inequality.

If we replace a by $a/2$, then the integral still gives some estimate. However, in Chebychev's inequality we have $4/a^2 = 5/4$ which is *bigger* than 1. Thus, it says that the probability that we are outside the range is bounded by a number bigger than 1. In other words, it says that the probability that the number of heads lies in the range $[490, 510]$ is more than a negative number which is not useful!

6. Out of a collection of 500 samples of DNA, 10 are known to have a mutation. An experimentalist is checking these samples *one-by-one* for the presence of the mutation. Give an estimate for the smallest number of samples needed to be checked so that the experimentalist can have a probability of at least 4/5 (or 80%) of seeing the mutation.

Solution: The frequency of a mutation is $c = 10/500 = 1/50$. The waiting time approximates the geometric distribution (as above). If S denotes the random variable denoting the number of samples to be checked before seeing a mutation, then $P(S \leq t)$ is *approximated* by $1-\exp(-ct)$. So we want the smallest t such that $1-\exp(-ct) \geq 4/5$ or equivalently $\exp(-ct) \leq 1/5$. Such a t is approximated by

$$-(1/c)\log(1/5) = \log(5) \cdot 50 \simeq 80$$

In the last calculation we used the value of log we computed in the previous assignment!

7. Suppose that the probability of success in a trial is 1/50. Write a formula for the probability of 2 successes in 20 trials. Also apply the discrete Poisson distribution to get an estimate of the probability. Note how the two values differ. What will happen to the difference if the probability of success is in a single trial is 1/100)?

Solution: The probability of 2 successes in 20 trials is given by

$$\binom{20}{2}(1/50)^2(1-1/50)^{20-2}$$

We then write this as

$$(1/2)\cdot(20)^2\cdot(1-1/20)\cdot(1/50)^2\cdot(1-(20/50)(1/20)^{20}(1-1/50)^{-2}$$

Combining all the terms with the power 2, we have

$$(1/2) \cdot (20/50)^2 \cdot \left((1-1/20)(1-1/50)^{-2}\right) \cdot \left(1 - \frac{20/50}{20}\right)^{20}$$

We now put $a = 20/50 = 2/5$ and get

$$(1/2) \cdot a^2 \cdot \left(1 - \frac{a}{20}\right)^{20} \cdot \left((1-1/20)(1-1/50)^{-2}\right)$$

Comparing with $(a^2/2)\exp(-a)$ we see that the error comes from the approximation of $\exp(-a)$ by $(1-a/20)^{20}$ and from treating the last two terms as 1 (since they are close to 1!). Both of these errors will become *somewhat* smaller when we replace 50 by 100. However, to increase accuracy, we must also increase 20 to a larger number.

8. Suppose that a shooting star (meteorite) passes overhead once per hour on average. What is the probability that you watch for one hour and see no shooting star? How often should you look at the sky so that your probability is within 1% of this probability?

Solution: This is just the same as the second problem above written with an application in mind! If we look at the sky n times in an hour, the possibility that we see a meteor at that $1/n$-th of an hour is $1/n$ on average. Thus, if we divide the hour in n parts, the probability that we do not see a meteor is $(1-1/n)^n$. As n goes to infinity, this approximates $\exp(-1)$ which is the probability that if you watch continuously, then you will see no shooting star. If one calculates this one finds that n about 30 gives an answer with an error of at most 1%. (We can also use the estimates given in part 1 to calculate a suitable value of n.

9. Suppose that the frequency of occurrence of an certain scintillation is f times per second. What is the probability that you will have to wait for t seconds before you see a scintillation? What is the expected waiting time for a scintillation?

Solution: This is the direct application of the waiting time distribution. If W denotes the time waited before the scintillation, then (for $t \geq 0$)!)

$$P(W < t) = 1 - f\exp(-ft) = \int_0^t f\exp(-fx)dx$$

The expectation is $1/f$ (which we can also see intuitively since 1 scintillation occurs, on average, every $1/f$ seconds!).

10. Two hundred students eat in the mess in a period of two hours on average. Assume that the students enter the mess at random times and that the mess runs continuously for all time! What is the probability that no student will enter the mess over a period of ten minutes? What is the amount of time that one should wait to be 90% certain that a student will enter the mess during that time?

> **Solution:** The average amount of time (in seconds) that one needs to wait for a student to enter the mess is $c = 7200/200 = 36$. Let W be the random variable that measures the amount of time that one has to wait before a student enters the mess. We can use the waiting time distribution (with $\lambda = 1/c = 1/36$)
>
> $$P(W < t) = \int_0^t \lambda \exp(-\lambda x) dx = 1 - \exp(-\lambda t)$$
>
> The probability $P(W \geq 600)$ (that we have to wait at least 10 minutes to see a student) is $\exp(-600/36) = \exp(-50/3)$.
>
> In order to have $P(W < t) \geq 9/10$ we need $1 - \exp(-t/36) \geq 9/10$ so $-t/36 \leq \log(1/10)$ (natural log!) or $t \geq 36 \log(10)$. (Note that the answer is in seconds).

11. An buzzer goes with a frequency of f times per hour. We observe the buzzer for a period of one hour. What is the probability that the buzzer will go off k times? What is the expected number of times the buzzer will go off?

> **Solution:** This is a case where the Poisson distribution is applicable. Let X denote the number of times the buzzer goes off. We have
>
> $$P(X = k) = \frac{f^k}{k!} \exp(-f)$$
>
> The expectation of this random variable is f which is what we see intuitively as well!

12. About 10 planes pass low overhead in IISER Mohali campus very day from 9:00 Hours to 19:00 hours on average. We have a sound sensitive experiment that takes 30 minutes to carry out. Assume that the experiment is completely ruined if 3 (or more) planes pass overhead during the experiment. What is the probability that an experiment is completely ruined?

> **Solution:** The frequency with which the planes fly low overhead is $f = 1/2$ every 30 minutes. Let L be the random variable denoting the number of planes that fly

overhead in the 30 minutes of the experiment. It is distributed according to the Poisson distribution

$$P(L = k) = \frac{1}{(2^k)k!} \exp(-1/2)$$

We want to calculate the probability $P(L \geq 3)$ which is

$$P(L \geq 3) = \sum_{k=3}^{\infty} \frac{1}{(2^k)k!} \exp(-1/2)$$

We note that this is

$$P(L \geq 3) = 1 - \sum_{k=0}^{2} \frac{1}{(2^k)k!} \exp(-1/2) = 1 - 2\exp(-1/2)$$

13. A chef adds 200 cashew pieces to 2 kilogrammes of rawa idly *maavu*. Each idly weighs 100 grammes before cooking. Assuming that kaju pieces weigh nothing (they are so light!), what is the probability of finding at most 1 kaju piece in an idly? What is the expected number of kaju pieces in an idly?

Solution: The density of kaju pieces is $20/20 = 1$ kaju piece per 100 grammes of maavu. Let X be the number of kaju pieces per idly. This follows the Poisson distribution:

$$P(X = k) = \frac{1}{k!} \exp(-1)$$

The probability that there is at most 1 kaju piece is

$$P(X = 0) + P(X = 1) = 2\exp(-1)$$

(Note that this is quite close to 1!)

The expected number of Kaju pieces is 1 which is easy to see intuitively as well.