Solutions to Assignment 7

1. There are about 8 Colloquiums at IISER Mohali during each semester (which is 16 weeks long). What is the probability that there is a gap of at least two weeks when there is no colloquium?

Solution: The frequency of colloquium per week is 0.5. Let W be the random variable that signifies the waiting time (in weeks) for a colloquium. The distribution is given by $P(W \leq t) = \int_0^t ce^{-cs}ds$ where $c = 0.5$. We calculate the answer as

$$1 - P(W \leq 2) = \int_2^\infty ce^{-cs}ds = e^{-2 \cdot 0.5} = e^{-1}$$

2. In a class of 200 students, the average in an examination was 10.5 marks with a standard deviation of 3.1. Estimate the number of students with marks between 8 and 11. Estimate the highest and lowest marks.

Solution: Let X be the random variable that gives the marks obtained by a randomly chosen student. We assume that X follows the normal distribution with mean $\mu = 10.5$ and standard deviation $\sigma = 3.1$.

The probability that a student has marks between 8 and 11 is given by

$$p = \frac{1}{\sigma\sqrt{2\pi}} \int_8^{11} e^{\frac{(s-\mu)^2}{2\sigma^2}} ds$$

Thus, the estimated number of students is $200p$.

If 199 students obtain less than a certain mark, then that mark will be the highest. This means we want M so that

$$199/200 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-infty}^{M} e^{\frac{(s-\mu)^2}{2\sigma^2}} ds$$

Equivalently, we want $t = (M - \mu)/\sigma$ so that

$$199/200 = \frac{1}{\sqrt{2\pi}} \int_{-infty}^{t} e^{\frac{s^2}{2}} ds$$

It turns out that this value of t is approximately 2.5758. Thus $M = 18.5$ approximately. The same constant (why?) can be used for lowest marks as well to get 2.5 as the approximately the lowest marks.

3. On the average 2.5 students enter the mess every second during the lunch session. If you wait for 10 seconds, what is the probability that you will see exactly 20 students? at least 20 students? at most 20 students?

Solution: An *average* of 25 students enter every 10 seconds. Let W be the number of students that we see in 10 seconds. This is a random variable following the the Poisson distribution with mean $c = 25$. In other words, $P(W = k) = (c^k/k!)e^{-c}$. We can use this to calculate the answers:

$$(c^{20}/20!)e^{-25} \; ; \; \sum_{k=20}^{\infty}(c^k/k!)e^{-c} \; ; \; \sum_{k=0}^{20}(c^k/k!)e^{-c}$$

to the above questions.

4. The average height of people in Punjab is 170 centimetres with a standard deviation of 5. Estimate the percentage of people who have a height above 183 centimetres? Estimate the percentage of people who have height between 160 and 180?

Solution: The height H of a randomly chosen person from Punjab can be taken as normally distributed with mean $\mu = 170$ and standard deviation $\sigma = 5$. The percentage of people with a height above 183 is estimated as

$$100 \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{183}^{\infty} e^{\frac{(s-\mu)^2}{\sigma^2}} ds$$

Similarly, the percentage of people with height between 160 and 180 is given by

$$100 \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{160}^{180} e^{\frac{(s-\mu)^2}{\sigma^2}} ds$$

5. There are 71 faculty members in IISER Mohali. Estimate the probability that at least 5 of them have birthday on the same day? (Assume that there are exactly 365 birthdays possible.)

Solution: Let $W = W_d$ be the random variable that counts the number of people with birthday on a fixed given day d. The expected value for this is $c = 71/365$. Since the number 71 is "large", we can use the Poisson distribution as an approximation and so

$$P(W = k) = \frac{c^k}{k!}e^{-c}$$

Thus, the probability that at most 4 have birthday on a given day is (approximately) given by

$$P(W \leq 4) = e^{-c} \sum_{k=0}^{4} \frac{c^k}{k!}$$

Let us assume (incorrectly!) that the variables W_d for different days are independent. In that case, probability that *all* 365 days have 4 or less birthdays, is given by $p = P(W \leq 4)^{365}$. Hence, the probability that there is at least one day when there are at least 5 birthdays is given by $1 - p$. We can calculate this to be about 0.02. This is is actually drastically over-estimated!

A different way of arguing is as follows.

Let $V = V_k$ be the random variable that counts the number of k-tuples of people with the same birthday. In our case $k = 5$.

For a given k-tuple the probability is $(1/365)^{k-1}$ that they all have the same birthday. The number of k-tuples among 71 faculty members is $\binom{71}{k}$.

We thus estimate that the expected value of W is $c = \binom{71}{k}/365^{k-1}$. In our case $k = 5$. We can use the Poisson distribution as an approximation and so

$$P(V \geq 1) = 1 - e^{-c}$$

We can approximate this to do a rough calculation

$$c = \frac{\binom{71}{5}}{365^4} \simeq 1/1363$$

Since c is small, we see that $e^{-c} \simeq 1 - c$. This gives us $P(W \geq 1) \simeq 1/1363$ which is quite low. However, it seems to be a bit of an over-estimate.

Yet another way of arguing is as follows. Let W denote the vector (365-dimensional!) random variable representing the number of birthdays on a given day of the year; in other words, $W = (a_1, \ldots, a_{365})$ represents that case that there are a_i birthdays on the i-th day of the year. We have the usual multinomial formula

$$P(W = (a_1, \ldots, a_{365})) = \binom{71}{a_1, \ldots, a_{365}} (365)^{-71}$$

here we use the notation

$$\binom{71}{a_1, \ldots, a_{365}} = \frac{71!}{a_1! \cdots a_{365}!}$$

Now we put V as the random variable that denotes the maximum of the entries of W. We then have

$$P(V \leq m) = \sum_{\vec{a} \in [0,m]^{365}; |\vec{a}| = 71} P(W = \vec{a})$$

where $|\vec{a}|$ denotes the sum of the entries of $\vec{a}$. The probability we want to compute is $1 - P(V \leq 4)$.

We can re-write the above as

$$P(V \leq m) = \frac{71!}{365^{71}} \sum_{\vec{a} \in [0,m]^{365}; |\vec{a}|=71} \frac{1}{a_1! \cdots a_{365}!}$$

Now we can make a clever sum (due to Yashonidhi Pandey in part and Shruti Paranjape for a correction) as follows

$$\sum_{\vec{a} \in [0,m]^{365}} \frac{1}{a_1! \cdots a_{365}!} x^{|\vec{a}|} = \left(\sum_{a=0}^{4} \frac{x^a}{a!} \right)^{365}$$

Thus, we want the ceofficient of x^{71} in the latter expression. This looks a bit difficult to do by hand but Sage can do it!

```
sage: e4=taylor(exp(x),x,0,4)
sage: f=taylor((e4)^365,x,0,71)
sage: 1 - N(f.coefficient(x,71)*factorial(71)/365^71)
0.000630906277814430
```

So we get roughly 1/1584.

6. A certain sample of cells has 1 cell division every 10 minutes. A student needs to observe 5 cell divisions in order to get a complete picture of the process. What is the probability that the student will take at most one hour to complete the experiment?

Solution: Let T denote the expected amount of time (in minutes) that the student will take to complete the experiment. Then T is a random variable following the waiting time distribution for 5 events. We are given that the frequency is $c = 0.1$ events per minute. Thus, the probability of waiting for at most 60 minutes is

$$\int_0^{60} \frac{c^5 s^4}{4!} e^{-cs} ds$$

We can show this to be the same as

$$1 - e^{-c} \sum_{k=0}^{4} \frac{c^k}{k!}$$