Assignment 3

Joshua Hwang (44302650)

30 March

1 White Walkers

A Poisson distribution is given by,

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

A neat property of the Poisson distribution is the mean and variance are equal. In this particular case, the second moment is $E(X^2) = 0.002a$, where a is the area in square kilometers.

We use the equation for variance to determine our λ,

$$\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
\lambda &= E(X^2) - \lambda^2 && \text{Both the variance and mean are } \lambda \\
E(X^2) &= \lambda + \lambda^2 && \text{The second moment is } 0.002a \\
0.002a &= \lambda + \lambda^2 \\
0 &= \lambda^2 + \lambda - 0.002a && \text{Solve the quadratic equation} \\
\lambda &= \frac{-1 \pm \sqrt{1^2 - 4(1)(-0.002a)}}{2(1)} \\
&= \frac{-1 \pm \sqrt{1 + 0.008a}}{2}
\end{aligned}$$

Since variance can never be negative (since it's the squared deviation from the mean and the standard deviation would not make sense), we produce the following probability function with regards to area a.

$$\lambda = \frac{-1 + \sqrt{1 + 0.008a}}{2}$$

1.1 Expected number of White Walkers

We're asked for expected number of White Walkers given $a = 20 \times 30 = 600$. Since we're using a Poisson distribution our lambda is our expected value.

$$\begin{aligned}
\lambda &= \frac{-1 + \sqrt{1 + 0.008a}}{2} \\
&= \frac{-1 + \sqrt{1 + 0.008(600)}}{2} \\
&= \frac{-1 + \sqrt{1 + 4.8}}{2} \\
&\approx 0.7042 && \text{To 4 decimal places}
\end{aligned}$$

There is expected to be 0.7042 White Walkers in a 20 by 30 km region.

1.2 10 White Walkers

Our new a will be 1 million hectares which is 10000 square kilometres.

$$\begin{aligned}
\lambda &= \frac{-1 + \sqrt{1 + 0.008a}}{2} \\
&= \frac{-1 + \sqrt{1 + 0.008(10000)}}{2} \\
&= \frac{-1 + \sqrt{1 + 80}}{2} \\
&= 4
\end{aligned}$$

Thus our PMF is now,

$$\begin{aligned}
P(X = k) &= \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \frac{4^k e^{-4}}{k!}
\end{aligned}$$

Now we find the probability of more than 10.

$$\begin{aligned}
P(X > 10) &= 1 - P(X \leq 10) \\
&= 1 - (P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = 9) + P(X = 10)) \\
&= 1 - (0.9972) && \text{To 4 decimal places} \\
&= 0.0028
\end{aligned}$$

Thus, the probability of having more than 10 White Walkers in a million hectares is 0.0028.

2 Fair dice

This situation can be modelled as a geometric distribution.

$$P(X = k) = (1 - p)^{k-1} p$$

Also note,

$$\begin{aligned}
P(X \geq k) &= \sum_{i=k}^{\infty} (1-p)^{i-1} p \\
&= p \sum_{i=k}^{\infty} (1-p)^{i-1} && \text{Infinite geometric series} \\
&= p \frac{(1-p)^{k-1}}{1-(1-p)} && \text{Probabilities are less than 1 as long as } p \text{ is not impossible} \\
&= p \frac{(1-p)^{k-1}}{p} \\
&= (1-p)^{k-1}
\end{aligned}$$

In our case the probability that any of the dice give a result of 5 or higher is,

$$\begin{aligned}
P(\text{any greater or equal to 5}) &= 1 - P(\text{all less than 5}) \\
&= 1 - \left(\frac{4}{6} \times \frac{4}{6} \times \frac{4}{6} \right) \\
&= \frac{19}{27}
\end{aligned}$$

Thus our PMF is,

$$\begin{aligned}
P(X = k) &= \left(1 - \frac{19}{27} \right)^{k-1} \frac{19}{27} \\
&= \left(\frac{8}{27} \right)^{k-1} \frac{19}{27}
\end{aligned}$$

2.1 Third toss

The probability we stop at or before 3 rolls is determined by,

$$\begin{aligned}
P(X \le 3) &= 1 - P(X > 3) \\
&= 1 - P(X \ge 4) \\
&= 1 - \left(\left(\frac{8}{27}\right)^{4-1} \right) \\
&= 0.9740 && \text{To 4 decimal places} \\
&= P(X = 1) + P(X = 2) + P(X = 3) && \text{An alternative approach...} \\
&= 0.9740 && \text{To 4 decimal places}
\end{aligned}$$

2.2 More than 10

We're looking for,

$$P(X > 10|\text{No 6 in 6 throws}) = \frac{P(X > 10 \cap \text{No 6 in 6 throws})}{P(\text{No 6 in 6 throws})}$$

Note: achieving no 6s in 6 throws is guaranteed to occur if more than 10 throws occur. Thus 6 throws is a subset of $X > 10$.

$$= \frac{P(X > 10)}{P(\text{No 6 in 6 throws})}$$

Now we need to find the probability of not rolling a 6 in 6 rolls. We use the same techniques as before. The chance that three dice fail to roll a 6 is,

$$\begin{aligned}
P(\text{3 dice no 6}) &= \left(\frac{5}{6}\right)^3 \\
&= 0.5787
\end{aligned}$$

Now we find how it fails after 6 throws.

$$\begin{aligned}
P(\text{No 6 in 6 rolls}) &= P(\text{3 dice no 6})^6 \\
&= 0.0376
\end{aligned}$$

$$
\begin{aligned}
&= \frac{P(X > 10)}{P(\text{No 6 in 6 throws})} \\
&= \frac{P(X \geq 11)}{0.0376} \\
&= \frac{(1-p)^{11-1}}{0.0376} \\
&= \frac{\left(1 - \frac{19}{27}\right)^{11-1}}{0.0376} \\
&= 0.0001
\end{aligned}
$$

Thus the probability of taking more than 10 throws given that the firt 6 throws did not produce a single 6 is 0.0001 (to 4 decimal places).

3 Fish

This problem can be modelled with a hypergeometric distribution. N is the total number of fish, r is the number of desired fish and n are then number of fish we catch.

$$
\begin{aligned}
P(X = k) &= \frac{{}^{r}C_k \times {}^{N-r}C_{n-k}}{{}^{N}C_n} \\
&= \frac{{}^{r}C_k \times {}^{21-r}C_{n-k}}{{}^{21}C_n} \qquad \text{We leave } n \text{ and } r \text{ for each specific question}
\end{aligned}
$$

3.1 Remaining gold fish

Instead of looking at it like the cat chose 17 fish, we instead "choose" those 4 lucky fish that were spared. We will use the 15 goldfish as our focus. Our PMF is now,

$$
\begin{aligned}
P(X = k) &= \frac{{}^{r}C_k \times {}^{21-r}C_{n-k}}{{}^{21}C_n} \\
&= \frac{{}^{15}C_k \times {}^{21-15}C_{4-k}}{{}^{21}C_4} \\
&= \frac{{}^{15}C_k \times {}^{6}C_{4-k}}{5985}
\end{aligned}
$$

From basic rules of the choice function, we know neither of the numbers in a choice can be negative and the top number must be larger (can't choose 5 items from a set of 4). From these we know, $0 \leq k \leq 4$ this gives us the domain. From here we evaluate every possible k and determine the range, all taken to 4 decimal places.

$$\begin{aligned}
P(X = 0) &= 0.0025 \\
P(X = 1) &= 0.0501 \\
P(X = 2) &= 0.2632 \\
P(X = 3) &= 0.4561 \\
P(X = 4) &= 0.2281
\end{aligned}$$

The range is $[0.0025, 0.4561]$.

3.2 Expected silver carp

For this question we will now focus on the 6 silver carps and the number of fish caught, $r = 6$ and $n = 17$. Luckily we have a formula for expectation,

$$\begin{aligned}
E(X) &= n\frac{r}{N} \\
&= 17\frac{6}{21} \\
&= 4.8571
\end{aligned}$$

We can expect the cat to catch 4.2857 of the 6 silver carps.

3.3 Probability of more silver carps

Let the number of goldfish the cat caught be x and the number of caught silver carps is y. Assume the cat did catch more silver carps, $x < y$. The number of fish left in the pond is $6 - y$ carps and $15 - x$ goldfish, it is completely reasonable to conclude there can't be a negative number of fish left in the pond. Also the number of fish caught must be 17, $x + y = 17$. Now we perform basic algebra,

$$\begin{aligned}
x &< y && \text{add } y \text{ to both sides} \\
x + y &< 2y \\
17 &< 2y \\
8.5 &< y \\
9 &\leq y && \text{Swap them} \\
-y &\leq -9 && \text{Add 6 to both sides} \\
6 - y &\leq 6 - 9 \\
6 - y &\leq -3 && 6 - y \text{ is the silver carp left} \\
\text{Num. of silver carp left} &\leq -3
\end{aligned}$$

But now the number of silver carp left in the pond is negative which is directly against our previous declaration for a positive number of fish. We have arrived at a contradiction, the cat cannot catch more silver carp than goldfish.

4 Camera chips

Both chips use a normal distribution as models for their lifetimes. Since this is a normal distribution we can standardise our distributions and compare them. We standardise both to get their comparable Z scores for $X \geq 24000$.

$$\begin{aligned}
Z &= \frac{x - \mu}{\sigma} \\
Z &= \frac{24000 - 21000}{3000} && \text{Chip 1} \\
Z &= 1 \\
Z &= \frac{24000 - 22000}{1000} && \text{Chip 2} \\
Z &= 2
\end{aligned}$$

Using a standard deviation table we find the following, (we need to reverse these since they are $\leq$)

$$\begin{aligned}
P(Z \geq 1) &= 1 - .8643 && \text{Chip 1} \\
&= 0.1587 \\
P(Z \geq 2) &= 1 - .9772 && \text{Chip 2} \\
&= 0.0228
\end{aligned}$$

We can make this comparison now because both distribution are in standard form, they both follow the same PMF. Without out even looking at the actual probabilities we have an intuitive understanding that Chip 1 will do better. In Chip 2's case, achieving a 240000 hour lifetime is given a Z score much higher than Chip 1's. Higher Z score means less likely to achieve. Looking at the actual probabilities now, Chip 1 has a probability of 0.1587 compared to Chip 2's 0.0228 (much lower). Therefore, Chip 1 will have a higher probability of reaching the 24000 mark.

5 Malfunction time

We use the exponential distribution for the chance of malfunction. The PDF (f) and CDF (F) are,

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} && x \geq 0 \\ F(x) &= 1 - e^{-\lambda x} && x \geq 0 \end{aligned}$$

We expect a malfunction in 12 days. Luckily we have a simple formula for the expectation of a exponential function.

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ 12 &= \frac{1}{\lambda} \\ \lambda &= \frac{1}{12} \end{aligned}$$

So our complete model is now,

$$\begin{aligned} f(x) &= \frac{1}{12} e^{-\frac{1}{12}x} && x \geq 0 \\ F(x) &= 1 - e^{-\frac{1}{12}x} && x \geq 0 \end{aligned}$$

5.1 First day malfunction

We're looking for $P(X \leq 1) = F(1)$.

$$\begin{aligned} F(1) &= 1 - e^{-\frac{1}{12}1} \\ &= 0.0800 && \text{To 4 decimal places} \end{aligned}$$

The chance of a malfunction on the first day is 0.0800.

5.2 Another 12 days

We denote another feature of our model,

$$\begin{aligned} P(X > x) &= 1 - P(X \leq x) \\ &= 1 - (1 - e^{-\lambda x}) \\ &= e^{-\lambda x} \end{aligned}$$

Now we can properly evaluate our problem,

$$\begin{aligned} P(X > 24 | X > 12) &= \frac{P(X > 24 \cap X > 12)}{P(X > 12)} \\ &= \frac{P(X > 24)}{P(X > 12)} \\ &= \frac{e^{-\frac{1}{12}24}}{e^{-\frac{1}{12}12}} \\ &= e^{-\frac{1}{12}12} && \text{Note: memoryless} \\ &= e^{-1} \\ &= 0.3679 && \text{To 4 decimal places} \end{aligned}$$

Here we see the memoryless property pop up. It is trivial to generalise this example but we will just move on. The probability of malfunctioning after 12 days is 0.3679.

6 Derive Z

We know $Y \sim N(1,9)$ so we can represent it as an affine transformation of the standard Gaussian random variable, G.

$$
\begin{aligned}
Y &= \mu + \sigma G \\
Y &= 1 + 3G
\end{aligned}
$$

Now we look at Z as an affine transformation of Y then G.

$$
\begin{aligned}
Z &= 2Y - 5 \\
Z &= 2(1 + 3G) - 5 \\
Z &= 2 + 6G - 5 \\
Z &= 6G - 3 \\
Z &\sim N(-3, 36)
\end{aligned}
$$

Our random variable will be positive if $G > \frac{1}{2}$. Luckily we can find the probability of this since G is the normal Gaussian random variable. Looking through the standard normal distribution table we find the following,

$$
\begin{aligned}
P(G > \frac{1}{2}) &= 1 - P(G \leq \frac{1}{2}) \\
&= 1 - 0.6915 \\
&= 0.3085
\end{aligned}
$$

The probability of having a positive Z is 0.3085.

We now consider $Y \sim U(1,9)$. We again represent our random variable as an affine tranformation of a more standard, $U(0,1)$.

$$
\begin{aligned}
Y &= 1 + (9 - 1) \times U(0,1) \\
Y &= 1 + 8 \times U(0,1)
\end{aligned}
$$

We look at Z as an affine transformation of $U(0,1)$ then the equivalent distribution in standard notation.

$$
\begin{aligned}
Z &= 2Y - 5 \\
Z &= 2(1 + 8 \times U(0,1)) - 5 \\
Z &= 2 + 16U(0,1) - 5 \\
Z &= 16U(0,1) - 3 \\
Z &\sim U(-3, 13)
\end{aligned}
$$

In this case to get a positive Z we need $U(0,1) > \frac{3}{16}$. Thankfully it is quite easy to calculate the CDF for a uniform random number.

The height of $U(0,1)$ is also its PDF which is determined by,

$$
\begin{aligned}
f(x) &= \frac{1}{b-a} && a < x < b \\
f(x) &= \frac{1}{1-0} && 0 < x < 1 \\
f(x) &= 1 && 0 < x < 1
\end{aligned}
$$

Now we find $P\left(U(0,1) > \frac{3}{16}\right) = 1 - P\left(U(0,1) < \frac{3}{16}\right)$.

$$
\begin{aligned}
P(U(0,1) < x) = F(x) &= \frac{x}{1} && 0 < x < 1 \\
P\left(U(0,1) < \frac{3}{16}\right) &= \frac{3}{16} \\
&= 0.1875
\end{aligned}
$$

Thus $P(U(0,1) > \frac{3}{16}) = 1 - 0.1875 = 0.8125$

7 Countable discontinuities

To prove a countable number of points a we prove there exists an injective function from those points a to the rational numbers. Injective functions prove that a domain is less than or equal to the range. Consider the function the following function.

For every point a a discontinuity exists from $F(a-)$ to $F(a)$. Since $F(a-) \neq F(a)$ then we have an interval $(F(a-), F(a)]$. Since this is an interval on the real numbers, there exists a rational number in there. Our function will map a to one of the rational numbers in $(F(a-), F(a)]$.

Now we just need to prove uniqueness and we've created an injective function. All CDFs are monotone increasing thus $(F(a_1-), F(a_1)] \cap (F(a_2-), F(a_2)] = \emptyset$. It is not possible for any as to share a rational number. Thus we have successfully constructed an injective function from a to the rationals so,

$$\text{Num. of discontinuities} \leq \mathbb{Q}$$

Since the rationals are countably infinite, the number of a must be at least countable.