Poisson Distribution Presentation Notes

A. Jeremy Mahoney

The Scenario

Consider the following scenario: there is a field with a mole in a hole. From time to time, the mole pokes its head out of the hole. You decide to make a daily routine of sitting down next to the hole for a fixed time t (say, four hours), counting the number of times the mole comes out of the hole during t. During your hours of contemplation, you ponder how to guess the number of days you should expect to see the mole once, twice, thrice, and so on. You have, in fact, been pondering the Poisson distribution! The Poisson Distribution PMF f is defined as follows, where k is the number of occurrences of the event in question:

$$f(k) = \begin{cases} e^{\lambda}\frac{\lambda^k}{k!} & \text{if } k \in \mathbb{N} \cup \{0\} \\ 0 & \text{otherwise} \end{cases}$$

This is all well and good, but how did this distribution come to be? This is the question we'll explore in this presentation.

The Derivation

Let λ be the average number of times that the mole emerges from the hole during t. After sitting by the hole for many days, you'd be able to produce a pretty good estimate of λ. And your estimate would probably get better over time. It so happens that λ is all we need to form an entire Poisson distribution.

We'll break t into n small intervals of size $\frac{t}{n}$. Assuming that the mole is equally likely to emerge at any time, the probability that the mole appears during an interval is approximately $\frac{\lambda}{n}$. (Note that this is only valid if $n >> t$, so we need to make n large and thus break t into a large number of intervals.) And the probability that the mole does not appear during an interval is approximately $1 - \frac{\lambda}{n}$. Feels like the Binomial Distribution may apply to this situation...each interval can be viewed as a trial that's a success if the mole appears, and a failure otherwise. Thus, we use the Binomial Distribution to define the probability $p(k)$ of seeing exactly k successful trials as follows:

$$p(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

To form the Poisson Distribution, we consider what happens to $p(k)$ as $n \to \infty$. We'll begin by examining what happens to $\binom{n}{k}$ as $n \to \infty$.

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{k} &= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \\
&= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \\
&= \frac{n^k}{k!}
\end{aligned}$$

Note that the final equality above results from the fact that when $n >> k$, subtracting the terms that are being multiplied by a constant less than k is insignificant, so the product of those terms approaches n^k as $n \to \infty$.

Observe that $\binom{n}{k}$ is multiplied by $\left(\frac{\lambda}{n}\right)^k = \frac{\lambda^k}{n^k}$. Thus, because $\lim_{n\to\infty} \binom{n}{k} = \frac{n^k}{k!}$, we conclude the following:

$$\lim_{n\to\infty} \left(\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \right) = \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} = \frac{\lambda^k}{n^k}$$

Finally, we turn our attention to the final term $\left(1 - \frac{\lambda}{n}\right)^{n-k}$. As $n \to \infty$, k becomes insignificant in the exponent, so the value of the term approaches $\left(1 - \frac{\lambda}{n}\right)^n$. Recall that

$$e = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n$$

and similarly,

$$e^a = \lim_{n\to\infty} \left(1 - \frac{a}{n}\right)^n$$

Thus, we conclude the following:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^\lambda$$

Putting everything together, we arrive at the formula for the PMF of the Poisson Distribution as follows:

$$\begin{aligned}
f(k) &= \lim_{n\to\infty} p(k) \\
&= \lim_{n\to\infty} \left(\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \right) \\
&= \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} \cdot e^\lambda \\
&= e^\lambda \frac{\lambda^k}{k!}
\end{aligned}$$

This concludes the derivation. Note that this function is only defined on the set of nonnegative integers, because k represents a discrete number of times that an event occurs: in this case, the number of times the mole emerges from its hole during t.

Exercises

1. Use a software tool (i.e. Python with Jupyter Notebook, R, Matlab) to plot the value of $p(k)$ for a reasonable range of k values and a very large value for n, using a value of your choice for λ. Overlay this plot with a plot of the Poisson Distribution $f(k)$, to show graphically that $p(k)$ is very close to $f(k)$ for a sufficiently large value of n.

2. Explain why we may assume that the probability that the mole appears during an interval is approximately $\frac{\lambda}{n}$ only if $n >> t$.

3. Prove the following:
$$e^a = \lim_{n \to \infty} \left(1 - \frac{a}{n}\right)^n$$

4. Prove that the average value of the Poisson Distribution is λ.

5. Prove the following:
$$\sum_{k=0}^{\infty} f(k) = 1$$