Teaching Poisson Distribution

Koh Liang Cheung Daniel
Department of Mathematics, Raffles Institution

Abstract

The Poisson Distribution is introduced in the Statistics syllabus within Advanced "A" level H2 Mathematics. In this article, some approaches to teaching the topic are introduced and discussed.

1 Introduction and Background

The Binomial and Poisson random variables are example of ***discrete*** random variables (DRVs), used in applications as "counting models", that is, to count the (random) number of occurrence of a "phenomenon" of interest. The emphasis in the H2 syllabus is on using the Binomial and Poisson model in different word problem scenarios to calculate probabilities with the aid of a Graphics Calculator. The expression for Binomial probabilities,

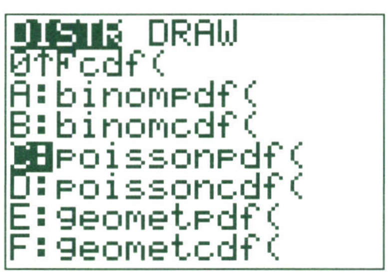
$$\binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0,1,2,\ldots,n$$

is easy for students to understand and can be interpreted simply, for example, as the probability of obtaining x number of "heads" when a dice is tossed n times. Here, p corresponds to the probability that a "head" turns up when a dice is thrown once. On the other hand, the expression for Poisson probabilities,

$$\frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0,1,2,\ldots$$

is not intuitive.

The general concept of a DRV is not emphasized in the H2 syllabus, hence many students may not know what a *probability mass function* (pmf) is. Interestingly the Graphics Calculator (GC) syntax is poisson***pdf***( ) instead of pmf for the computation of probabilities.



Screen Shot from TI-84 Plus Family of Graphing Calculators

Figure 1

Definition

A random variable X is said to be *Discrete* if it takes a countable set of possible values x_1, x_2, x_3, K . We define the pmf of the DRV X as the function p where

$$p(x_i) = P(X = x_i) \text{ for } i = 1, 2, 3, \ldots$$

The reader can easily check that the pmf satisfies the following properties:

(i) $p(x_i) \geq 0$ for $i = 1, 2, 3, \ldots$; and

(ii) $\sum_{i=1}^{\infty} p(x_i) = 1.$

It can be shown in a more advanced course in probability that these 2 properties in fact characterize a pmf, and hence the probability distribution of a random variable. We now define a Poisson random variable. The random variable X that takes values on the non-negative integers with pmf

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \text{ with parameter } \lambda > 0,$$

is said to have the Poisson distribution with parameter λ, and we write $X \sim Po(\lambda)$.

We leave it as an exercise for the reader to show that the requirements to be a pmf are indeed satisfied. In particular, students are reminded about the standard series expansion $e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \ldots$. We can define a Binomial distribution similarly.

JC students are typically not taught to view Poisson probabilities in this way but will interpret Poisson probabilities as a limit of Binomial probabilities or as an approximation to Binomial probabilities.

Starting with Binomial probabilities $\binom{n}{x} p^x (1-p)^{n-x}$ with $\lambda = np$, we can use Poisson probabilities $\frac{e^{-\lambda}\lambda^x}{x!}$ as a good approximation, in the situation when n is large and p is small.

Teachers attempt to justify this observation by comparing the probabilities obtained from the 2 models. For example when $n = 100$, $p = 0{,}04$ we observe that the probabilities produced using a Binomial model, $\binom{100}{x}(0.04)^x(0.96)^{100-x}$ (the Y_1 values in the table below) and those obtained through a Poisson model, $\frac{e^{-4}4^x}{x!}$ (the Y_2 values) have corresponding values that are very close. The two probability mass functions appear to differ only in the 3rd decimal place.

X	Y1	Y2
0	.01687	.01832
1	.07029	.07326
2	.14498	.14653
3	.19733	.19537
4	.19939	.19537
5	.15951	.15629
6	.10523	.1042
X=0		

Screen Shot from TI-84 Plus Family of Graphing Calculators

Figure 2

The above suggests that the approximation is valid in this instance but a student might still wonder,

Question 1 "How can we justify the claim in a mathematically more general way (as opposed to working out a few specific cases on a GC) that $\frac{e^{-\lambda}\lambda^x}{x!} \approx \binom{n}{x} p^x (1-p)^{n-x}$ when n is large and p is small, such that $\lambda = np$?"

Within H2 Statistics syllabus, students are also told that before a situation can be modeled with the Poisson model, events must occur:

C1 **randomly** in a given interval of time or space,

C2 **independently** of each other, i.e. the fact that an event has occurred (or not occurred) does not change the probability of another event occurring,

C3 **uniformly**, i.e. the **average rate of occurrences is a constant**, and

C4 **singly**, i.e. the probability of two or more events occurring within a very short interval is negligible.

But these conditions or requirements still appear rather imprecise. A student might wonder,

Question 2 "How exactly do the Poisson conditions (described so qualitatively) above give rise to Poisson probabilities?"

These 4 conditions when first "taught" are difficult to understand for the JC student, who could be asked in the A-level examination to ascertain whether any of these conditions are violated in a given context.

We illustrate with an example to discuss informally and intuitively what the 4 conditions mean, and how as a model it is an idealization to reality. An examination question might be posed as follows.

Example

Explain why a Poisson model would be valid if applied to count the number of car accidents taking place on an expressway (for example, PIE) during the morning from 7 am to 9 am (rush hour).

C1 "randomly" is sensible and says that accidents happen by chance and are not deterministic.
C2 "independently" says that given that an accident(s) has occurred does not change the probability of a subsequent accident occurring.
C3 "uniformly" suggests that the *propensity* for an accident or rate at which accidents occur is constant, which assumes all factors affecting an accident from 7 am to 9 am on the expressway remain the same throughout the entire interval. These factors could be a combination which includes weather, traffic volume, road condition etc.
C4 "singly" in this context means that simultaneous (cluster of) accidents (or at multiple locations) do not occur for a particular instance of time.

Later, we will see how to introduce precise Mathematical statements that indeed give rise to a Poisson distribution. The objective of this article is to present some possible approaches teachers might want to consider to address the 2 questions raised above. Many of these ideas were motivated by classroom discussions with students from the Raffles Academy (RA) Mathematics Programme. The RA Mathematics programme, started in 2009, caters to high ability students with an interest in Mathematics. We were strongly influenced by the introductory book by Ross [2] which we referred to heavily. We encourage JC students who are interested in learning Probability (and Statistics) to read this well written text.

2 Poisson distribution as an approximation to Binomial distribution

Result 1

We begin first by showing that if $X \sim B\left(n, \frac{\lambda}{n}\right)$ then $X \sim Po(\lambda)$ approximately when n is large, p is small, such that $\lambda = np$. The approximation is in fact exact when $n \to \infty$.

Recursive nature of Binomial and Poisson Probabilities

Solution 1

We consider the sequence $p_0, p_1, p_2, \ldots$ defined as follows:

$$\text{Let } p_k \equiv P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,2,\ldots,n.$$

The topic of recurrence relation is introduced as a H2 Pure Mathematics topic under the Chapter on Sequences and Series. We observe here that Binomial probabilities can be computed recursively, by the recurrence relation,

$$\frac{p_k}{p_{k-1}}=\frac{(n-k+1)p}{k(1-p)} \text{ with an initial value } p_0=(1-p)^n.$$

Now, if $X \sim B\left(n,\frac{\lambda}{n}\right)$ then $\dfrac{p_k}{p_{k-1}}=\dfrac{(n-k+1)\frac{\lambda}{n}}{k\left(1-\frac{\lambda}{n}\right)}=\dfrac{\lambda-\frac{(k-1)\lambda}{n}}{k-\frac{k\lambda}{n}}$.

So if n is large, and $\dfrac{k\lambda}{n}\approx 0$ we have $\dfrac{p_k}{p_{k-1}}\approx\dfrac{\lambda}{k}$.

If we can justify that $p_0=\left(1-\dfrac{\lambda}{n}\right)^n\approx \mathrm{e}^{-\lambda}$ then we can conclude that we have satisfied the recurrence relation that defines Poisson probabilities, that is, $X \sim Po(\lambda)$approximately!

Why? We observe that if $X\sim Po(\lambda)$ then $p_k \equiv P(X=k)=\dfrac{\mathrm{e}^{-\lambda}\lambda^k}{k!}$ which yields the recurrence relation, $\dfrac{p_k}{p_{k-1}}=\dfrac{\lambda}{k}$ and $p_0=\mathrm{e}^{-\lambda}$.

We still have to show that $p_0=\left(1-\dfrac{\lambda}{n}\right)^n\approx \mathrm{e}^{-\lambda}$!

The following is a very useful analytic result which we consider here and can also be found in most undergraduate Calculus books.

Lemma : If n is large then $\left(1+\dfrac{a}{n}\right)^n\approx \mathrm{e}^a$.

To show $\left(1-\dfrac{\lambda}{n}\right)^n\approx \mathrm{e}^{-\lambda}$ we simply apply the lemma with $a=-\lambda$ noting also that for "n large", $0<p=\dfrac{\lambda}{n}<1$ so $-1<-\dfrac{\lambda}{n}<0$.

Proof

Recall that in secondary school we could define $\mathrm{e} \equiv \lim_{n \to +\infty} \left(1 + \frac{1}{n}\right)^n$, and we could investigate the sequence $u_n = \left(1 + \frac{1}{n}\right)^n$ and its limiting value through the use of a calculator.

n	u(n)	
1	2	
11	2.6042	
21	2.6563	
31	2.6757	
41	2.6859	
51	2.6921	
61	2.6963	

n=1

Screen Shot from TI-84 Plus Family of Graphing Calculators

Figure 3

Informally, we can fix value of a (> 0) and large n, and investigate similarly with the use of a calculator that $\left(1 + \frac{a}{n}\right)^{\frac{n}{a}} \approx \mathrm{e}$. Therefore $\left(1 + \frac{a}{n}\right)^n = \left[\left(1 + \frac{a}{n}\right)^{\frac{n}{a}}\right]^a \approx \mathrm{e}^a$.

The graphics calculator is useful for exploratory purposes. We now incorporate ideas from H2 Mathematics to proceed further.

We use a Binomial series expansion with large n and argue informally,

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= 1 + n\left(\frac{1}{n}\right) + \frac{n(n-1)}{2}\left(\frac{1}{n}\right)^2 + \frac{n(n-1)(n-2)}{3!}\left(\frac{1}{n}\right)^3 + \ldots \\ &= 1 + 1 + \frac{1}{2}\left(1-\frac{1}{n}\right) + \frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) + \ldots \\ &\approx 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \ldots\end{aligned}$$

We recognize this last line as the series definition of e. A more careful analysis here would not be palatable to a JC student and is reserved for an introductory undergraduate Real Analysis course.

Similarly, for large n we also have,

$$\begin{aligned}\left(1-\frac{1}{n^2}\right)^n &= 1 + n\left(-\frac{1}{n^2}\right) + \frac{n(n-1)}{2}\left(-\frac{1}{n^2}\right)^2 + \ldots \\ &= 1 - \frac{1}{n} + \frac{1}{2n^2}\left(1-\frac{1}{n}\right) + \ldots \\ &\approx 1.\end{aligned}$$

Hence, from

$$\left(1-\frac{1}{n}\right)^n\left(1+\frac{1}{n}\right)^n=\left(1-\frac{1}{n^2}\right)^n,$$

$$\left(1-\frac{1}{n}\right)^n\approx\frac{1}{\left(1+\frac{1}{n}\right)^n}.$$

We have the following result for $a<0$ and large n,

$$\left(1+\frac{a}{n}\right)^n=\left(1-\frac{-a}{n}\right)^n\approx\frac{1}{\left(1+\frac{-a}{n}\right)^n}\approx\frac{1}{\mathrm{e}^{-a}}=\mathrm{e}^a.$$

We consider 2 other approaches which teachers might want to consider. In both approaches, we show that $\ln\left(1+\frac{a}{n}\right)^n\approx a$ for n large and leave it to the reader to explain why this is an equivalent result to $\left(1+\frac{a}{n}\right)^n\approx\mathrm{e}^a$ for n large.

Firstly, we consider instead the standard series for $\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\ldots$ for $-1<x\le 1$.

We have $\ln\left(1+\frac{a}{n}\right)^n=n\ln\left(1+\frac{a}{n}\right)$

$$=n\left(\frac{a}{n}-\frac{1}{2}\left(\frac{a}{n}\right)^2+\frac{1}{3}\left(\frac{a}{n}\right)^3\ldots\right)$$

$$=a-\frac{a^2}{2n}+\frac{a^3}{3n^2}\ldots$$

$$\approx a \quad \text{for } n \text{ large.}$$

Secondly, and perhaps the most *basic approach* for JC students, we consider differentiation from 1st principles,

$$\lim_{n\to\infty}\ln\left(1+\frac{a}{n}\right)^n=a\lim_{n\to\infty}\frac{\ln\left(1+\frac{a}{n}\right)-\ln 1}{\frac{a}{n}}=a\lim_{h\to 0}\frac{\ln(1+h)-\ln 1}{h}$$

$$=a\frac{d}{dx}\ln(1+x)|_{x=0}$$

$$=a.$$

Solution 2

We consider a second analytic approach which will be useful in the next section.

$$\text{If } X \sim B\left(n, \frac{\lambda}{n}\right) \text{ then } P(X=k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

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

$$\approx \frac{e^{-\lambda}\lambda^k}{k!} \text{ for fixed } k \text{ and large } n.$$

The student is in a position now to work out the details to explain how the 3rd line follows from the 2nd or she can refer to [2] Page 144.

3 Poisson Postulates

To see how the Poisson conditions give rise to the Poisson distribution as a limit of the Binomial distribution, these 4 conditions need to be ***re-cast*** into more precise Mathematical statements which are commonly referred to as Poisson postulates.

As a limit of the Binomial Distribution

Suppose we are interested in the number of occurrences X, of a certain phenomenon in a time period which we take to be $[0,1]$ for convenience. Let λ be a positive number, indicating the rate of occurrence. The Poisson postulates are:

P1 The number of events happening in disjoint (that is, non-overlapping) time intervals are mutually independent,

P2 The probability of exactly one event occurring in a small interval Δt is approximately $\lambda \Delta t$,

P3 The probability of more than one event occurring in a small interval Δt is approximately zero.

Result 2

If the Poisson postulates stated above are satisfied then X follows a Poisson distribution.

Solution

To obtain an expression for $P(X=k)$ we start by breaking the interval $[0,1]$ into n non-overlapping subintervals each of length $\frac{1}{n}$ and analyze what happens in each interval, before letting n get large.

We see immediately that **P1** is a more precise statement about the condition of **C2** "independence". From **P2** and **P3** we see that for n large, the probability of zero events occurring in the interval $\left[\frac{k-1}{n},\frac{k}{n}\right]$, $k=1,2,\ldots,n$ is approximately $1-\frac{\lambda}{n}$ while the probability of exactly one event occurring is approximately $\frac{\lambda}{n}$. In other words, we can treat each of the n sub-intervals of length $\frac{1}{n}$ as a Bernoulli trial ! Each instance we have at most one occurrence. This is what is meant by **C4** "singly". Observe also that the number of events occurring in any time interval of fixed length is not dependent on its starting time but only on the length of the interval.

How do we interpret **C3** "uniformly"?

Recall that if $T \sim B(1,p)$, $E(T)=p$, so for large n, the expected number of occurrences in $\left[\frac{k-1}{n},\frac{k}{n}\right]$, $k=1,2,\ldots,n$ (which is of length $\frac{1}{n}$) is approximately $\lambda\left(\frac{1}{n}\right)$.

By dividing through by $\frac{1}{n}$, we have

$$\frac{\text{Expected number of occurrence in sub-interval of length 1/n}}{1/n} \approx \lambda.$$

In other words the "average" intensity (or instantaneous rate) of occurrence is a constant λ.

By **P1**, since the n intervals $\left[\frac{k-1}{n},\frac{k}{n}\right]$, $k=1,2,\ldots,n$ are non-overlapping, the number of occurrences in each sub-interval, are mutually independent.

Hence, the number of occurrence in $[0,1]$ is obtained by summing the outcomes of n ***independent*** identically distributed Bernoulli random variables which results in an (approximate) Binomial model.

That is, $P(X=k)\approx\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$ with the approximation becoming exact as $n\to\infty$.

Following the argument of Solution 2 in Section 2 we have $P(X=k)\approx\frac{e^{-\lambda}\lambda^k}{k!}$ for large n.

A brief technical note

The idea conveyed by the Poisson condition of "singly" can be made much more explicit and illuminating in **P2** and **P3**. We now make what we mean by "approximately" and "negligible" more precise by introducing a "little o" notation.

We re-write the Poisson Postulate

P2 The probability of exactly one event occurring in a small interval Δt is $\lambda\Delta t + o(\Delta t)$,

P3 The probability of more than one event occurring in a small interval Δt is $o(\Delta t)$.

Informally, $o(\Delta t)$ is a general expression for a term or collection of terms that "goes to zero" faster than Δt "goes to zero". More formally, the notation $o(h)$ stands for any function $\mathrm{f}(h)$ such that $\lim_{h\to 0}\frac{\mathrm{f}(h)}{h} = 0$. For example, the function $\mathrm{f}(h) = h^2$ is $o(h)$ and so is $\mathrm{g}(h) = \ln(1+h)$ whereas $\mathrm{f}(h) = h$ is not.

Some further words about "limits" and how **P3** is applied to show that there are no accumulation of simultaneous events taking place at a particular instance. Recall we took the interval $[0,1]$ and divided it into n non-overlapping subintervals each of length $\frac{1}{n}$.

If A is the event that 2 or more events occur in ***some*** subinterval (with k events in $[0,1]$) then,

$$P(A) \le P\left(\bigcup_{i=1}^{n}\left\{2 \text{ or more events in } \left[\frac{i-1}{n},\frac{i}{n}\right]\right\}\right) \to 0 \ as\ n \to \infty.$$

The student can work out the details for herself of refer to [2] Page 153.

4 Poisson Process

In the previous section we considered a time interval [0,1] and assumed that the average number of occurrences is λ. We now consider how to generalize for an interval of length t. We can think of it here as applying the Poisson condition **C3** of "uniformly" so that for the interval $[0,t]$ the average number of occurrences is scaled accordingly to give us an average of λt occurrences.

Alternatively, for positive integer t, we break up the interval $[0,t]$ into sub-intervals $[0,1],[1,2],\ldots,[t-1,t]$ and use **P1** which informs us that the number of events that occur in each sub-interval are mutually independent. We apply **result 2** to each interval of length 1, and use the additive property of (independent) Poisson random variables to give us a Poisson distribution with mean λt for the number of events occurring over $[0,t]$.

We consider another approach here. This last section may appeal to students with a stronger Mathematics background. The students in the RA Mathematics programme are introduced to solutions of 1^{st} order differential equations at the level of H3 Mathematics (or the old Further Mathematics (FM)) which is required here.

Here we will regard the Poisson model as an example of a Counting process. It counts the number of times, N something happens in a given period of time. More formally we write $N([s,t])$ to count the number of occurrence of an event of interest in the time interval $[s,t]$.

Clearly, $N((s,t]) = N([0,t]) - N([0,s])$.

To simplify notation we define $N(t) = N([0,t])$ and $p_n(t) = P(N[0,t]=n)$ that is, $p_n(t)$ is the probability that there are n events in the interval $[0,t]$

Definition

A counting process $N([0,t])$ is said to be a Poisson process with rate $\lambda > 0$ if

PP1 $N(0)=0$,

PP2 $P(N[0,t]=k) = P(N[s,s+t]=k)$,

This means that the distribution of the number of occurrences in any time interval depends only on the "length" of the interval and not the starting time which communicates the idea of "stationary increments".

PP3 $a<b<c<d$ we have $N[a,b]$ and $N[c,d]$ are independent,

This means that the number of occurrences in disjoint time intervals are independent and convey the idea of "independent increments".

PP4 We use the "little o" notation introduced in the last section,

$$p_0(h) = P(N[0,h]=0) = 1-\lambda h + o(h)$$
$$p_1(h) = P(N[0,h]=1) = \lambda h + o(h)$$
$$P(N[0,h]>1) = o(h).$$

These 4 conditions are enough to produce a Poisson distribution, $p_n(t) = \dfrac{e^{-\lambda t}(\lambda t)^n}{n!}$!

Solution

Using the above 4 conditions we will form a "**Differential-Difference equation**" and show by Mathematical induction that the Poisson distribution $p_n(t)=\frac{e^{-\lambda t}(\lambda t)^n}{n!}$ is a solution.

A Differential Equation

Formulating Differential Equations of the form $\frac{dy}{dx}=f(y)$ is taught in the H2 Syllabus.

All RA Mathematics students take Physics and have encountered questions (for example dilution problems) which required them to form a differential equation (DE) by considering "what happens in a small time interval Δt". We illustrate.

Example

Consider a tank which initially holds V_0 litres of brine that contains a grams of salt. Another brine solution, containing b grams of salt per litre, is poured into the tank at the rate of e litres per minute, while simultaneously, the well stirred solution leaves the tank at the rate of f litres per minute. Let $Q(t)$ denote the amount (in grams) of salt in the tank at any time t.

Amount of salt "flowing in" in an interval of time $(t,t+\Delta t)$ is $be\Delta t$

Amount of salt "flowing out" in an interval of time $(t,t+\Delta t)$ is approximately $f\frac{Q(t)}{V_0+(e-f)t}\Delta t$

$$Q(t+\Delta t)-Q(t)\approx be\Delta t-f\frac{Q(t)}{V_0+(e-f)t}\Delta t$$

$$\frac{Q(t+\Delta t)-Q(t)}{\Delta t}\approx be-f\frac{Q(t)}{V_0+(e-f)t}$$

As $\Delta t\to 0$, we obtain the following differential equation,

$$\frac{dQ}{dt}=be-f\frac{Q}{V_0+(e-f)t},$$

which we can solve in H2 Mathematics because it is of the familiar form $\frac{dQ}{dt}=f(Q)$.

A Differential-Difference Equation

The idea here will be to condition on what happen in the 2 intervals $[0,t]$ and $[t,t+h]$ then use independent and stationary properties. "No arrivals in $t+h$" is same as "No arrivals in t and no arrivals in next h".

$$\begin{aligned}
p_0(t+h) &= P(N[0,t+h]=0) \\
&= P(N[0,t]=0 \text{ and } N[t,t+h]=0) \\
&= P(N[0,t]=0)P(N[0,h]=0) \\
&= p_0(t)p_0(h).
\end{aligned}$$

Now we want to form a differential equation and also use the initial condition that $p_0(0)=1$,

$$p'_0(t)=\lim_{h\to 0}\frac{p_0(t+h)-p_0(t)}{h}=\lim_{h\to 0}\frac{(p_0(h)-p_0(0))p_0(t)}{h}=p'_0(0)p_0(t).$$

For $p'_0(0)$ we have,

$$p'_0(0)=\lim_{h\to 0}\frac{p_0(h)-p_0(0)}{h}=\lim_{h\to 0}\frac{1-\lambda h+o(h)-1}{h}=-\lambda.$$

So

$$p'_0(t)=-\lambda p_0(t).$$

This simple differential equation which can be solved in H2 Mathematics has solution $p_0(t)=e^{-\lambda t}$.

This is rather interesting. To have no arrivals by time t means that the 1st arrival takes place after time t. We introduce the random variable T as the time until the 1st arrival.

$$P(N[0,t]=0)=e^{-\lambda t} \Leftrightarrow P(T>t)=e^{-\lambda t},$$

We obtain the cumulative distribution function,

$$P(T\le t)=1-e^{-\lambda t}.$$

Teachers who studied under the old Further Mathematics syllabus would recognize this as the distribution function of an exponential random variable, commonly used as a model for failure times and which has the memory-less property. [2] Page 418 shows us that in a Poisson model, the time between arrivals are independent and follow an exponential distribution!

We proceed in a similar way for the general case. Using the law of total probability,

$$
\begin{aligned}
p_n(t+h) &= P\left(N[0,t+h]=n\right)\\
&= P\left(N[0,t]=n \text{ and } N[t,t+h]=0\right)+P\left(N[0,t]=n-1 \text{ and } N[t,t+h]=1\right)\\
&\quad +\sum_{k=2}^{n} P\left(N[0,t]=n-k \text{ and } N[t,t+h]=k\right)\\
&= p_n(t)\,p_0(h)+p_{n-1}(t)\,p_1(h)+o(h).
\end{aligned}
$$

The little $o(h)$ arises because

$$\bigcup_{k=2}^{n}\left\{N[0,t]=n-k \text{ and } N[t,t+h]=k\right\}\subseteq\bigcup_{k=2}^{n}\left\{N[t,t+h]=k\right\}\subseteq\left\{N[t,t+h]>1\right\}.$$

We can now form our differential equation as follows,

$$
\begin{aligned}
p'_n(t) &= \lim_{h\to 0}\frac{p_n(t+h)-p_n(t)}{h}\\
&= p_n(t)\lim_{h\to 0}\left(\frac{p_0(h)-1}{h}\right)+p_{n-1}(t)\lim_{h\to 0}\frac{p_1(h)}{h}+\lim_{h\to 0}\frac{o(h)}{h}\\
&= -\lambda p_n(t)+\lambda p_{n-1}(t).
\end{aligned}
$$

Now rearranging this a little we have $p'_n(t)+\lambda p_n(t)=\lambda p_{n-1}(t)$.

This equation looks a little daunting at first glance, but we consider $n = 1$ and use $p_0(t)=e^{-\lambda t}$,

$$p'_1(t)+\lambda p_1(t)=\lambda e^{-\lambda t}.$$

Students taking H3 Mathematics (or in the past, students taking Further Mathematics) would learn how to solve a 1st order differential equation with constant coefficient,

$$\frac{dy}{dt}+\lambda y=\lambda \mathrm{e}^{-\lambda t}.$$

We multiply through using the integrating factor $e^{\lambda t}$ to obtain an exact differential equation,

$$\mathrm{e}^{\lambda t}\left(\frac{dy}{dt}+\lambda y\right)=\lambda$$

$$\frac{d}{dt}\left(y\mathrm{e}^{\lambda t}\right)=\lambda$$

In this setting we have,

$$\frac{d}{dt}\left(\mathrm{e}^{\lambda t}p_1(t)\right)=\lambda.$$

We can solve by integrating to obtain,

$$e^{\lambda t} p_1(t) = \lambda t$$
$$p_1(t) = e^{-\lambda t} \lambda t.$$

We can now repeat the process recursively and establish by induction that $p_n(t) = \dfrac{e^{-\lambda t}(\lambda t)^n}{n!}$.

(Inductive step) Assume now that $p_k(t) = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}$ for some positive integer k.

$$\frac{d}{dt}\left(e^{\lambda t} p_{k+1}(t)\right) = e^{\lambda t} \lambda p_k(t)$$
$$\frac{d}{dt}\left(e^{\lambda t} p_{k+1}(t)\right) = \frac{\lambda(\lambda t)^k}{k!}$$
$$e^{\lambda t} p_{k+1}(t) = \int_0^t \frac{\lambda(\lambda t)^k}{k!} dt = \frac{(\lambda t)^{k+1}}{(k+1)!}$$
$$p_{k+1}(t) = \frac{e^{-\lambda t}(\lambda t)^{k+1}}{(k+1)!}.$$

We are done !

4 Further Discussion

We end this article by a further discussion of some other aspects of the 4 Poisson conditions.

Firstly, we might ask if the 1st Poisson condition that "events must occur **randomly** in a given interval of time or space" is truly necessary. Upon reflection, if this condition is removed, then "deterministic", "independently", "uniformly" and "singly" might refer instead to a phenomenon or process such as the manufacturing (or production) of items on an assembly line. The term "independently" might convey a different, more common everyday meaning.

However, beyond understanding the term randomly to mean that events occurring are not *pre-determined* in nature, the word "**randomly**" is itself not a precise term. A well-known problem in (Geometric) Probability known as Bertrand's Paradox (see [2] Page 197) gives fair warning to this concern. It states: "Consider a random chord of a circle. What is the probability that the length of the chord will be greater than the side of the equilateral triangle inscribed in that circle?" This problem is incapable of a solution because it is not clear what is meant to "choose a chord at random".

In general the 2nd condition of "**independence**" is hard to verify. In our earlier traffic accident example, the impact of how information introduced by a given accident will affect future accidents is hard to ascertain. For example, a student might argue that radio stations broadcast early morning accident reports. Hence, drivers may by chance tune in to the radio channel and this may change their driving behavior reducing the likelihood of a subsequent accident. The famous Monte Hall problem (see [1] Page 32 Exercise 1.4.30) or Prisoner's Dilemma (see [2] Page 105 Problem 3.44) shows how subtly information can be introduced which can affect decision making and change probabilities.

In examination questions, it is common to ask students to answer in context why the Poisson model, say for our traffic accident model, may fail when applied "for a long period of time". The most common explanation is that the 3rd Poisson condition of "**uniformly**" may fail to apply due to seasonal variations or factors. Interestingly if this seasonal variation is well known and regular it is possible to incorporate a time dependent rate $\lambda(t)$. Another generalization is to treat the intensity rate as a random variable itself. These are best left discussed in a more advanced stochastic modeling course.

Finally, in a traffic accident model, it is quite conceivable for an accident to involve multiple vehicles and to observe the condition of "**singly**" fail. Such models would also be worth studying in a more advanced Probability course.

5 Conclusion

The topic of Binomial and Poisson Distributions can be made more attractive and appealing to Mathematics students by drawing links with other topics within the H2 (and H3) Mathematics syllabus for explanatory purposes. In particular, conveying the importance of Mathematical definitions to describe Statistical concepts.

6 Acknowledgement

The author would like to express his thanks to his colleagues Lim Hway Kiong and Mok Hoi Nam who taught from the material presented here. A special word of thanks goes out to Professor Yap Von Bing, National University of Singapore, for his invaluable suggestions and comments, and Professor Victor Tan, National University of Singapore, for his encouragement.

References

[1] R. V. Hogg, J. W. McKean, and A. T. Craig, *Introduction to Mathematics Statistics* (6th ed.), New Jersey: Pearson Prentice Hall, 2005.

[2] S. Ross, *A first course in probability* (8th ed.), New Jersey: Pearson Prentice Hall, 2010.