Poisson Processes and Applications in Hockey

by
Zachary Viljo Calvin DeJardine

A project submitted to the Department of
Mathematical Sciences in conformity with the requirements
for Math 4301 (Honours Seminar)

Lakehead University
Thunder Bay, Ontario, Canada

Abstract

My honours project is on Poisson processes, which are named after the French mathematician Siméon-Denis Poisson. The Poisson process is a stochastic counting process that arises naturally in a large variety of daily-life situations. My main focus is to understand Poisson processes. In this project, a brief historical review of this topic is provided, some basic concepts, tools, several properties, and theoretical approaches for studying Poisson processes are introduced and discussed. We will further briefly discuss how to use Poisson processes to model scoring in a hockey game.

Acknowledgements

I would like to thank my supervisor Dr. Deli Li. His knowledge and assistance have been invaluable in the successful completion of this project. I would also like to thank Dr. Adam Van Tuyl for the feedback, as well as the LaTeXtemplates and tutorials that he provided.

Contents

Abstract i

Acknowledgements ii

Chapter 1. Introduction 1
1. Probability Review 1
2. Is Scoring in Hockey Poisson? 2

Chapter 2. Poisson Distribution 4
1. History and Definitions 4
2. Basic Properties of Poisson Distributions 6
3. Applications 8

Chapter 3. Poisson Process 10
1. Defining Poisson Processes 10
2. Some Properties of Poisson Processes 13

Chapter 4. Applying Poisson Process to Hockey 17
1. A Basic Model 17
2. How will the Game End? 18
3. Overtime 18
4. Additional Notes 20

Bibliography 22

CHAPTER 1

Introduction

The purpose of this project is to introduce Poisson processes, study some of the basic properties of Poisson processes, and make an application of the properties covered. We will show that we can use Poisson processes to model the number of goals scored in a hockey game and determine the likelihood of a given team winning. Poisson processes are also useful to model radioactive decay, telephone calls, and many other phenomena.

1. Probability Review

This section reviews some basic concepts and definitions of probability theory which will used throughout this project.

DEFINITION 1.1. A *binomial experiment* (Bernoulli trial) is a statistical experiment in which outcomes are categorized as either a success with probability p or failure with probability $q = (1 - p)$.

EXAMPLE 1.2. Consider flipping a coin n times and counting the number of heads that occur. This is a binomial experiment with the probability of success (Heads) has probability $p = 0.5$.

DEFINITION 1.3. A *random variable* is a set of of possible outcomes, where the probability of each outcome is given by the probability density function.

There are two types of random variables, discrete random variables and continuous random variables. This project will only be working with discrete random variables.

DEFINITION 1.4. A *discrete random variable* is random variable that has a finite or countable number of outcomes.

EXAMPLE 1.5. Let X represent the outcome when rolling a fair D6, the random variable X is a discrete random variable with the following outcomes $X = \{1, 2, 3, 4, 5, 6\}$

DEFINITION 1.6. A *probability density function* (pdf) is a function that describes the likelihood for a random variable to take on a given value.

$$P\{X = x\} = f(x)$$

where $f(x) \leq 1$ for all x and $\sum f(x) = 1$.

DEFINITION 1.7. Given two random variables X and Y the *joint probability density function* defines the probability of events in terms of both X and Y.

Definition 1.8. The *conditional probability* is the probability that an event A will occur, when a given event B is known to occur. This is said to be the "probability of A given B", denoted $P\{A|B\}$.

Definition 1.9. Two random variables, X and Y are *independent* if and only if they have a joint density function such that

$$f(x, y) = f(x)f(y).$$

In other words $P\{X = x \text{ and } Y = y\} = P\{X = x\}P\{Y = y\}$.

Definition 1.10. Let X be a discrete random variable with density function f. The *expected value* or mean of X denoted $E[X]$, is given by

$$E[X] = \sum_{\textit{all } x} xf(x)$$

provided $E[X] = \sum_{\textit{all } x} xf(x)$ is finite. The summation is over all values of X with nonzero probability.

Definition 1.11. Let X be a discrete random variable with mean μ. The *variance* of X, denoted by Var$X$ is given by

$$\text{Var } X = E[X^2] - E[X]^2.$$

Definition 1.12. A *stochastic process* $\{N_t : t \geq 0\}$ is a collection of random variables indexed by a totally ordered set.

Definition 1.13. A stochastic process $\{N_t : t \geq 0\}$ is said to be a *counting process* if N_t represents the total number of events or arrivals that have occurred up to time $t > 0$, such that the following properties hold

(1) $N_t \geq 0$
(2) N_t is non negative integer valued.
(3) If $s \leq t$ then $N_s \leq N_t$
(4) If $s < t$ then $N_t - N_s$ is the number of events that have occurred in interval $(s, t]$

Example 1.14. If N_t represents the number of goals scored in a hockey game by time t, then N_t is a counting process. On the contrary, if X_t represents the goal differential in a hockey game at time t, then it is not a counting process.

2. Is Scoring in Hockey Poisson?

For events to follow a Poisson process they must have three properties; events must be random, rare, and memoryless. Being random means that events occur with no apparent pattern. For hockey this is mostly the case, except when teams pull their goalies. If an event is rare it means that an event could occur many times but only occurs a few times. As a rule an event can be considered rare for the number of trials $n \geq 100$ and the probability of success $p \leq 0.10$. In ice hockey scoring a goal takes only seconds, so a game that lasts at least 60 minutes could potentially have hundreds of goals scored. If events are memoryless then the probability of the next event occurring does not depends on previous events. If scoring in hockey is not memoryless then goals would occur in bunches similar to baseball.

We can see that goals in hockey are indeed rare, memoryless and for the most part they are random, the exception is during the final minutes of the third period when teams are trailing by one or two goals, they pull the goalie for a extra attacker in the attempt to score late goals to send the game to overtime. In the 2011-2012 NHL season there were 238 empty net goals scored. It is a reasonable assumption that all the empty net goals were scored during the final minutes of the third period since it is extremely rare for a team to pull the goalie at any other time in the game. In [**3**] the author discusses this phenomena in more detail and calls it the "end game effect".

In the 2011-2012 NHL season there were a total of 6545 goals scored, 6426 in regulation and 119 in overtime. Breaking down goals scored by period we see that number of goals scored in the second and third periods are nearly identical (2^{nd} 2254, 3^{rd} 2248), while there were 1924 goals scored in the first period. There are many factors that can lead to teams scoring less in the first period, but for the most part we can see that goals a evenly distributed throughout the game.

In Chapter 2 this project will introduce the Poisson distribution, give a extremely brief history about the Poisson distribution and prove some of its important properties. Chapter 3 will focus on defining Poisson processes and studying some of the properties of independent Poisson processes. Chapter 4 we use Poisson processes to model hockey game, and provide a method to determine what team will win a given game.

CHAPTER 2

Poisson Distribution

Before we can begin to study the properties of Poisson processes we need to review some properties of the Poisson distribution. This section will have a brief historical review of the Poisson distribution, cover some of its properties, and give two examples that follow Poisson distribution.

1. History and Definitions

DEFINITION 2.1. Let X be a binomial random variable. Then the probability of getting exactly x successes in n trials is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

where p is the probability of a success. We say X has *binomial distribution*.

The Poisson distribution was derived by Siméon Denis Poisson (1781-1840) a French mathematician, physicist, and geometer. It was discovered as a result of Poisson's discussion of The Law of Large Numbers in his work "Recherches sur la probability des jugements en matiére criminelle et en matiére civile" (Research on the Probability of Judgments in Criminal and Civil Matters) published in 1837. In the chapter "the law of large numbers", Poisson said the probability of the number of occurrences of an event denoted E will fall within certain limits. Using the binomial distribution, he found that for large values of n that it was simpler to let n tend to infinity and p tend to zero, to avoid the growing factorial. This concept is now known as the Law of Small Numbers.

Poisson's observation was not noticed by the public until 1898 when Ladislaus von Bortkiewicz (1868-1931), a professor at University of Berlin, illustrated the importance of Poisson's formula in a monograph entitled "The Law of Small Numbers". Poisson's distribution was used by Ladislaus von Bortkiewicz to model the number of deaths by accidental horse kick in the Prussian army.

The Poisson distribution is one of the most important probability distributions in the field of probability. A Poisson distribution is the probability distribution that results from a Poisson experiment.

DEFINITION 2.2. A *Poisson experiment* is a statistical experiment that has the following properties:

(1) The experiment results in outcomes that can be classified as successes or failures.
(2) The average number of successes, denoted λ, that occurs in a specified region is known.
(3) The probability that a success will occur is proportional to the size of the region.
(4) The probability that a success will occur in an extremely small region is virtually zero.

(Note: The specified region may take on many forms, for example length, area, volume, time)

DEFINITION 2.3. A *Poisson random variable* is the number of successes that result from a Poisson experiment.

DEFINITION 2.4. Suppose we conduct a Poisson experiment, in which λ is the average number of successes within a given region. Let X be the number of successes that result from the experiment. Then

$$P\{X = x\} = f(x) = \frac{\lambda^x}{x!}e^{-\lambda}\; x \in \mathbb{N}\,,\; \lambda > 0$$

where x is the number of successes and e is the natural logarithm. Then X has *Poisson distribution* denoted $X \sim \mathrm{P}_\lambda$.

The Poisson distribution is useful because many random events follow it. We give some examples below,

EXAMPLE 2.5. The occurrence of a major earthquake could be considered to be a random event. If there are an average of 5 major earthquakes each year, then the number X of major earthquakes in any given year will have a Poisson distribution with parameter $\lambda = 5$

The Poisson distribution can be used as an approximation to the binomial distribution. For those situations in which n is large and p is very small, the Poisson distribution can be used to approximate the binomial distribution. Let X be a binomial random variable with parameters n and p. Then

$$P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x}, x \in \{1, 2, 3, \ldots, n\}.$$

If n is large and p is very small then,

$$P\{X = x\} \approx \frac{(np)^x}{x!}e^{-np}\; x \in \{1, 2, \ldots, n\}$$

The larger the n and smaller the p, the closer the approximation becomes.

EXAMPLE 2.6. There are 50 misprints in a book which has 250 pages. We will find the probability that page 100 has no misprints. Let X be the number of misprints on page 100. Then X follows a binomial distribution with $n = 50$ and $p = \frac{1}{250}$. Since n

is large and p is small, by using the Poisson distribution to approximate the binomial distribution, we find:

$$P\{X=0\} \approx \frac{(50\frac{1}{250})^0}{0!}e^{-50\frac{1}{250}} = e^{-0.2} \approx .0819$$

2. Basic Properties of Poisson Distributions

Recall that $X \sim \mathrm{P}_\lambda$ if

$$f(x) = P\{X=x\} = \frac{\lambda^x}{x!}e^{-\lambda} \; x \in \mathbb{N},\ \lambda > 0$$

Let us verify that this is indeed a probability density function (pdf) by showing that the sum of $f(x)$ over all $x \geq 0$ is 1. We have

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}.$$

This last sum is the power series formula for, e^λ, so

$$e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

Theorem 2.7. *If $X \sim P_\lambda$ then $E[X]=\lambda$ and $Var\ X = \lambda$.*

Proof. First we will prove that $E[X]=\lambda$. This is a simple exercise in series manipulation. We begin by noticing that

$$E[X] = \sum_{x=0}^{\infty} x f(x) = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}.$$

Noting that

$$n! = n \cdot (n-1)! \cdot (n-2)! \cdots 2 \cdot 1,$$

we see that $\frac{x}{x!} = \frac{1}{(x-1)!}$. So, pulling out $e^{-\lambda}$ and canceling x, we see that

$$E[X] = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}.$$

(Note that the term $x=1$ has $0!$ in the denominator.)
Setting $m = x-1$, we find

$$E[X] = \lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!} = \lambda e^{-\lambda}e^{\lambda},$$

and we are done.

Now we will prove Var$X = \lambda$. Again we will make use of some simple series manipulation. By definition the variance is

$$\text{Var}X = E[X^2] - (E[X]^2) = \sum_{x=0}^{\infty} x^2 f(x) - (\sum_{x=0}^{\infty} x f(x))^2.$$

Since we have just proved that $E[X] = \lambda$ we know that $(E[X]^2) = \lambda^2$, which just leaves us with finding the value of $E[X^2]$.

$$E[X^2] = \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}.$$

As above we can see that

$$E[X^2] = e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \left(\sum_{x=1}^{\infty} (x-1) \frac{\lambda^{x-1}}{(x-1)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right).$$

$$= \lambda e^{-\lambda} \left(\lambda \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right)$$

Letting $i = x - 1$ and $j = x - 2$, we find that

$$E[X^2] = \lambda e^{-\lambda} \lambda \left(\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right) = \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda.$$

When we substitute the values of $E[X]$ and $E[X^2]$ into the definition of Var$X$ we get

$$\text{Var } X = E[X^2] - (E[X]^2) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

and we are done. □

THEOREM 2.8. *Let $X \sim P_{\lambda_1}$ and $Y \sim P_{\lambda_2}$. If X and Y are independent, then $X + Y \sim P_{\lambda_1+\lambda_2}$.*

PROOF. By definition we say X and Y are independent if and only if, $P\{X = x \text{ and } Y = y\} = P\{X = x\}P\{Y = y\}$. Now letting

$$Z = X + Y$$

we have that

$$\begin{aligned} P\{Z=z\}=P\{X+Y=z\} &= \sum_{\substack{x,y\geq 0\\ x+y=z}} P\{X=x \text{ and } Y=y\} \\ &= \sum_{\substack{x,y\geq 0\\ x+y=z}} P\{X=x\}P\{Y=y\} \\ &= \sum_{\substack{x,y\geq 0\\ x+y=z}} \frac{e^{-\lambda_1}\lambda_1^x}{x!}\cdot\frac{e^{-\lambda_2}\lambda_2^y}{y!} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{\substack{x,y\geq 0\\ x+y=z}} \frac{\lambda_1^x\lambda_2^y}{y!x!} \end{aligned}$$

To finish the proof we require a little trick. We begin by recognizing that $x!y!$ is the denominator of

$$\frac{z!}{x!y!}=\binom{z}{x}=\binom{z}{y}.$$

So, we multiply by a factor of $\frac{z!}{z!}$ we have

$$P\{Z=z\}=\frac{e^{-(\lambda_1+\lambda_2)}}{z!}\sum_{\substack{x,y\geq 0\\ x+y=z}}\binom{z}{x}\lambda_1^x\lambda_2^y.$$

Notice that from the binomial theorem we deduce that

$$\sum_{\substack{x,y\geq 0\\ x+y=z}}\binom{z}{x}\lambda_1^x\lambda_2^y=(\lambda_1+\lambda_2)^z.$$

So

$$P\{Z=z\}=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^z}{z!}.$$

and therefore Z has a Poisson distribution with parameter $\lambda_1+\lambda_2$ as claimed. □

3. Applications

This section will briefly discuss two of the many possible applications of the Poisson distribution. The first example is to demonstrate that scoring in hockey does indeed follow a Poisson distribution. The second example is a interesting example regarding prime numbers from [**8**], more information on the distribution of prime numbers can be found in [**7**].

3.1. Scoring in Hockey. If we are to assume that scoring in hockey follows a Poisson process then there are some things that we should be able to predict. In the 2011-2012 NHL season there were 6545 goals scored giving an average $\lambda = 2.66$ goals/game. To find the probability of a shut-out we simply need to compute $P\{X = 0\}$.

$$P\{X = x\} = \frac{e^{-\lambda}\lambda^x}{x!} \Rightarrow P\{X = 0\} = \frac{e^{-2.66}2.66^0}{0!} = .06995.$$

This is the probability of a team scoring zero goals in a game. Because there are 2460 games in a season, so the expected number of shutouts in a season is $2460 \times .06995 \approx 172$ (actual 177).

3.2. Prime Numbers. Prime numbers are very important numbers to many areas of mathematics. A prime number is an integer $n \in \mathbb{N}$ that is only divisible by ± 1 and $\pm n$ itself. It turns out that prime numbers in short intervals obey a "Poisson distribution". To discuss this, we need to look at the Prime Number Theorem:

THEOREM 2.9. *(Prime Number Theorem): The number of primes $p \leq x$ has size about $x/\log x$, where here the log is to the base-e. That is, if $\pi(x)$ denotes this number of primes less then or equal to x, then*

$$\pi(x) \sim x/\log x.$$

This means that

$$\lim_{n\to\infty} \frac{\pi(x)}{x/\log x} = 1.$$

In what follows, we will use the notation $\pi(I)$ to denote the number of primes in a given interval I.

Now suppose that x is a "really large" number , say $x = 10^{100}$. Then, if you pick a number $n \leq x$ at random, there is about a $1/100\log(10)$ (about 0.43 percent) chance that it will be prime; and we expect that a typical interval $[n, n + \log n] \subset [1, x]$ will contain about one prime.

In fact, much more is true. If we pick $n \leq x$ at random, and choose $\lambda > 0$ and j not too large (say $\lambda, j \leq 20$), then the number of primes in $[n, n + \lambda \log n]$ roughly obeys a Poisson distribution:

$$P\{\pi([n, n + \lambda \log n]\} = j \approx \frac{e^{-\lambda}\lambda^j}{j!}.$$

Notice that we do not have an equality; in order to get an equality we would have to let $x \to \infty$ in some way. Certainly the larger we take x to be, the close the above probability comes to

$$\frac{e^{-\lambda}\lambda^j}{j!}$$

CHAPTER 3

Poisson Process

In this Chapter we will define Poisson process in two different ways, and discuss how Poisson processes can be combined or broken down.

1. Defining Poisson Processes

There are three different ways to define Poisson processes: using arrivals, using the inter-arrival times, or using differential equations. This section will define a Poisson process by the arrivals and determine the distributions of the arrivals and the inter-arrival times. Constructing Poisson processes with the inter-arrival times and differential equations is left out due to the length of this project. More information on the other methods of defining Poisson processes can be found in [**6**].

1.1. Arrivals. Consider N_t the number of arrivals by time $t \geq 0$. The counting process N_t is said to be a Poisson process if it satisfies the following properties:

(1) The number of arrivals that occur during one interval of time does not depend on the number of arrivals that occur over a different time interval.
(2) The "average" rate at which arrivals occur remains constant.
(3) Arrivals occur one at a time.

Though these properties help us understand the Poisson process, in order to make use of them we need to make them more mathematically precise.

(1) For the first property let $t_1 \leq t_2 \leq t_3 \leq t_4$, then the random variables $N_{t_2} - N_{t_1}$, and $N_{t_4} - N_{t_3}$ are independent.
(2) The second property says that the number of arrivals after time t should be λt.
(3) The third property indicated that the probability of more then one arrival occurring in a small time interval is extremely unlikely. This can be written as:

$$P\{N_{t+\Delta t} - N_t = 0\} = 1 - \lambda \Delta t + o(\Delta t), \tag{1.1}$$

$$P\{N_{t+\Delta t} - N_t = +1\} = \lambda \Delta t + o(\Delta t), \tag{1.2}$$

$$P\{N_{t+\Delta t} - N_t \geq 2\} = o(\Delta t). \tag{1.3}$$

where Δt is a small time interval and $o(\Delta t)$ is a function such that

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

Now that we have a definition of the Poisson process that we can use, we will determine the distribution of N_t. We divide the time interval $[0, t]$ into n intervals, and write

$$N_t = \sum_{j=1}^{n} [N_{\frac{j}{n}t} - N_{\frac{(j-1)}{n}t}].$$

We have rewritten N_t as the sum of n independent, identically distributed random variables. For a large number n, $P[N_{t/n} \geq 2] = o(t/n)$ by (1.3). So we can approximate N_t by a sum of independent random variables that equal:

- 1 with probability $\lambda(t/n)$.
- 0 with probability $1 - \lambda(t/n)$.

Using the formula for the binomial distribution,

$$P\{N_t = x\} \approx \binom{n}{x} (\lambda t/n)^x (1 - (\lambda t/n))^{n-x}.$$

We can then show:

$$P\{N_t = x\} = \lim_{n\to\infty} \binom{n}{x} (\lambda t/n)^x (1 - (\lambda t/n))^{n-x}.$$

To take this limit, note that

$$\lim_{n\to\infty} \binom{n}{x} n^{-x} = \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!n^x} = \frac{1}{x!}$$

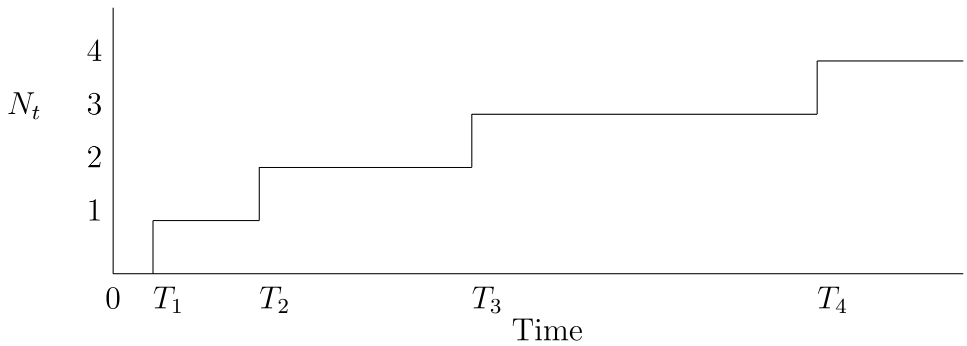
and

$$\lim_{n\to\infty} (1 - \frac{\lambda t}{n})^{n-x} = \lim_{n\to\infty} (1 - \frac{\lambda t}{n})^n \lim_{n\to\infty} (1 - \frac{\lambda t}{n})^{-x} = e^{-\lambda t}.$$

Hence,

$$P\{N_t = x\} = e^{-\lambda t} \frac{(\lambda t)^x}{x!},$$

so N_t has a Poisson distribution with parameter λt.



The above figure shows N_t over time.

1.2. Inter-arrival Times. The second way we can define the Poisson process is by considering the time between arrivals. Let $\tau_n, n \in \mathbb{N}$ be the time between the nth and $(n-1)$st arrivals. Let $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$ be the total time until the nth arrival. We can write

$$T_n = \inf\{t : N_t = n\},$$

$$\tau_n = T_n - T_{n-1}.$$

The τ_i should be independent, identically distributed random variables. The τ_i should also satisfy the memoryless property. That is if there is no arrival in time $[0, s]$, then the probability of an arrival occurring in the interval $[s, t]$ is the same as if arrivals had occurred in $[0, s]$. This can be written mathematically as

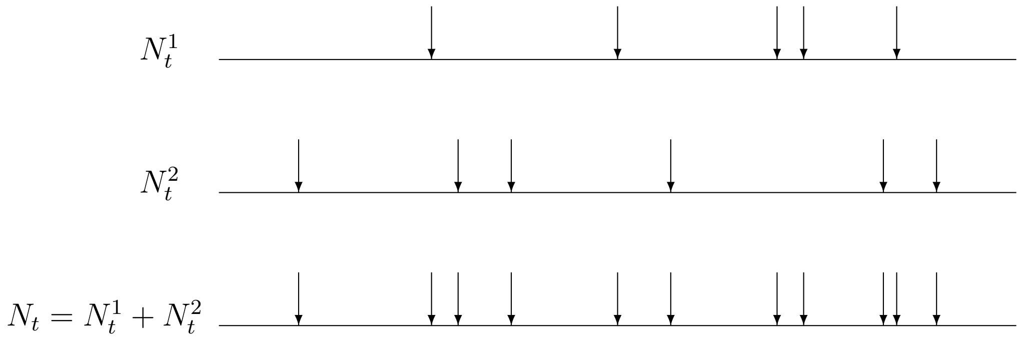
$$P\{\tau_i \geq s + t | \tau_i \geq s\} = P\{\tau_i \geq t\}.$$

The only real-valued functions that satisfy $f(t + s) = f(t)f(s)$ are $f(t) = 0$ for all t, or $f(t) = e^{-\lambda t}$. The function $f(t) = 0$ contradicts the third property of the Poisson process, since if there is a 0 probability of $N_t = 0$ no matter how small t is, then there must be a infinite number of arrivals in any interval. Thus the τ_i is exponentially distributed.

Definition 3.1. A random variable X has *exponential distribution* with rate parameter λ if it has probability density function given by

$$P\{X = x\} = f(x) = \lambda e^{-\lambda x}, 0 < x < \infty.$$

This gives us another method of constructing a Poisson process.



The above figure shows the superposition N_t of 2 independent Poisson processes N_t^1 and N_t^2

2. Some Properties of Poisson Processes

This section will go over some important properties of independent Poisson processes.

2.1. Superposition of Poisson Processes.

EXAMPLE 3.2. Consider a store that has two entrances, A and B. Let N_t^A count the number of customers that arrive at entrance A with rate $\lambda_A t$, and N_t^B count the number of customers that arrive at entrance B with rate $\lambda_B t$. Determine the total number of customers arriving to the store by time t.

To determine the number of customers that arrive by time t, we need to consider the counting process $N_t = N_t^A + N_t^B$.

THEOREM 3.3. *Let N_t^1 and N_t^2 be Poisson processes such that, $N_t = N_t^1 + N_t^2$. Since N_t, for $t \geq 0$, is the sum of two independent Poisson random variables, it is also a Poisson random variable with expected value $\lambda t = \lambda_1 t + \lambda_2 t$*

PROOF. To prove this we will use the same method as with proving the sum of two Poisson random variables is Poisson. We begin by recalling the definition of independence, that is, $P\{X = x \text{ and } Y = y\} = P\{X = x\}P\{Y = y\}$. Then we have

$$\begin{aligned} P\{N_t = n\} = P\{N_t^1 + N_t^2 = n\} &= \sum_{\substack{x,y \geq 0 \\ x+y=n}} P\{N_1^t = x \text{ and } N_2^t = y\} \\ &= \sum_{\substack{x,y \geq 0 \\ x+y=n}} P\{N_t^1 = x\}P\{N_t^2 = y\} \\ &= \sum_{\substack{x,y \geq 0 \\ x+y=n}} \frac{e^{-\lambda_1 t}(\lambda_1 t)^x}{x!} \cdot \frac{e^{-\lambda_2 t}(\lambda_2 t)^y}{y!} \\ &= e^{-t(\lambda_1 + \lambda_2)} \sum_{\substack{x,y \geq 0 \\ x+y=n}} \frac{(\lambda_1 t)^x (\lambda_2 t)^y}{y!x!} \end{aligned}$$

So, we multiply by a factor of $\frac{n!}{n!}$ we have

$$P\{N_t = n\} = \frac{e^{-t(\lambda_1+\lambda_2)}}{n!} \sum_{\substack{x,y \geq 0 \\ x+y=n}} \binom{n}{x} (\lambda_1 t)^x (\lambda_2 t)^y.$$

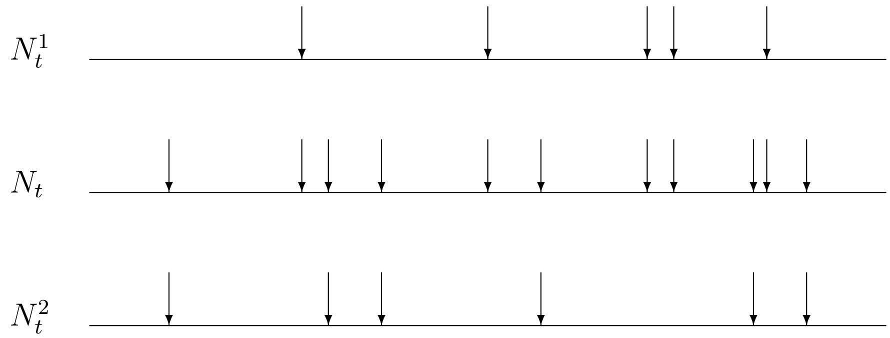
From the binomial theorem we deduce that

$$\sum_{\substack{x,y \geq 0 \\ x+y=n}} \binom{n}{x} (\lambda_1 t)^x (\lambda_2 t)^y = (\lambda_1 t + \lambda_2 t)^n.$$

So

$$P\{N_t = n\} = e^{-t(\lambda_1+\lambda_2)} \frac{(\lambda_1 t + \lambda_2 t)^n}{n!}$$

Therefore we have proven N_t also has a Poisson distribution with parameter $\lambda_1 t + \lambda_2 t$ as claimed. □



The above figure shows the thinning of the Poisson process N_t.

2.2. Thinning Poisson Processes. Now that we have shown that a Poisson process N_t can be constructed by the superposition of two independent Poisson process N_t^1 and N_t^2, we will now look at how to split a Poisson process N_t with rate λt in to two independent Poisson processes.

Suppose we have a Poisson process N_t with rate λt. for each arrival toss a unfair coin with the probability of heads p and the probability of tail $q = (1-p)$. Then the process N_t^1 counts the number of heads with rate $p\lambda t$ and N_t^2 counts the number of tails with rate $q\lambda t = (1-p)\lambda t$.

THEOREM 3.4. *A Poisson process N_t with rateλt can be split into independent Poisson processes N_t^1 and N_t^2 with rates $p\lambda t$ and $(1-p)\lambda t$ respectively.*

PROOF. To show that N_t^1 and N_t^2 are independent we first need to calculate the joint pdf of N_t^1 and N_t^2 for t arbitrary. Determining the conditional probability on a given number of arrivals N_t for the original process, we have

$$P\{N_t^1 = m, N_t^2 = n | N_t = m+n\} = \frac{(m+n)!}{m!n!} p^m (1-p)^n \tag{2.1}$$

This is simply the binomial distribution since the $m+n$ arrivals to the original process independently goes to N_t^1 with probability p and N_t^2 with probability $(p-1)$. Since the event $\{N_t^1 = m, N_t^2 = n\}$ is a subset of the conditioning event above.

$$P\{N_t^1 = m, N_t^2 = n | N_t = m+n\} = \frac{P\{N_t^1 = m, N_t^2 = n\}}{P\{N_t = m+n\}} \tag{2.2}$$

Combining (2.1) and (2.2), we have

$$\frac{(m+n)!}{m!n!} p^m (1-p)^n \frac{\lambda t^{m+n} e^{-\lambda t}}{(m+n)!} \tag{2.3}$$

Rearranging terms we find.

$$P[N_t^1 = m,\ N_t^2 = n] = \frac{(p\lambda t)^m e^{-\lambda p t}}{m!} \frac{((1-p)\lambda t)^n e^{-\lambda(1-p)t}}{n!} \tag{2.4}$$

This shows that N_t^1 and N_t^2 are independent. $\square$

We have now seen that a single Poisson process can be split into two independent Poisson processes. The most useful consequence is that, if we allow $N_t = N_t^1 + N_t^2$ with rate $\lambda t = \lambda_1 t + \lambda_2 t$ then each arrival in the combined process is labled as N_t^1 with probability $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and N_t^2 with probability $1 - p$.

CHAPTER 4

Applying Poisson Process to Hockey

In this section we will use the Poisson process to model scoring in a hockey game. This section will also briefly introduce a new distribution that is vital in determining the outcome of the game.

1. A Basic Model

One of the most important things to sports analysts, fans, and the gambling community is having the ability to predict the outcome of a sports match between two teams. Now that have we covered some of the basic properties of Poisson processes and also shown that scoring in hockey is basically a Poisson process, we are now able to use properties of Poisson processes to predict outcomes of a hockey game. Let N_t represent the total number of goals scored in time t with rate λt. If we let N_t^1 and N_t^2 represent the number of goals scored by Team 1 and Team 2 with rates $\lambda_1 t$ and $\lambda_2 t$ respectively. We can see that $N_t = N_t^1 + N_t^2$ and $\lambda t = (\lambda_1 + \lambda_2)t$. From here there are several things that we can easily see, including the expected number of goals scored the expected time till the first goal is scored, and which team is expected team to score first. However we are still unable to determine the probability of Team 1 winning ($N_t^1 > N_t^2$), the probability of a tie ($N_t^1 = N_t^2$), and the probability of Team 2 winning ($N_t^2 > N_t^1$). In order to determine the probability of a given team winning we need to consider the goal differential to do this we need to introduce the Skellam distribution.

Definition 4.1. If X and Y are independent random variables with rates λ_1 and λ_2 receptively. The difference $Z = X + Y$ follows a *Skellam distribution* with probability density function.

$$P\{Z = X - Y = z\} = S(z) = e^{-(\lambda_1+\lambda_2)}(\frac{\lambda_1}{\lambda_2})^{\frac{z}{2}} I_{|z|}(2\sqrt{\lambda_1\lambda_2})$$

Where $I_{|x|}()$ is the modified Bessel function of the first kind of order x.

Definition 4.2. The *modified Bessel function of the first kind of order* n denoted $I_n(x)$ where n is a integer can be written as.

$$I_n(x) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(n\theta)d\theta$$

More information on the Skellam distribution and Bessel functions can be found in [**9**] and [**10**]

We set $X_t = x$ to be the goal differential at time t (i.e $X_t = N_t^1 - N_t^2$). If $x > 0$ then Team 1 is winning by x goals at time t, if $x < 0$ then Team 2 is winning by x goals at time t, and if $x = 0$ the game is tied at time t.

2. How will the Game End?

This section will cover the probability of each possible outcome of a given hockey game.

2.1. End of the 3rd Period. If we are to assume that the game ends when time expires in the third period, then to compute the probability of a team winning we simply need to determining the probability of the Skellam distribution for values of x. That is:

(1) Team 1 will win after the third period with probability

$$P\{N_t^1 - N_t^2 = x > 0\} = \sum_{x>0} S(x) = \sum_{x>0} e^{-(\lambda_1+\lambda_2)} (\frac{\lambda_1}{\lambda_2})^{\frac{x}{2}} I_{|x|}(2\sqrt{\lambda_1 \lambda_2})$$

(2) Team 2 will win after the third period with probability

$$P\{N_t^1 - N_t^2 = x < 0\} = \sum_{x<0} S(x) = \sum_{x>0} e^{-(\lambda_1+\lambda_2)} (\frac{\lambda_2}{\lambda_1})^{\frac{x}{2}} I_{|x|}(2\sqrt{\lambda_1 \lambda_2})$$

(where $S(x)$ is the Skellam distribution at x, and N_t^1 and N_t^2 have rates λ_1 and λ_2 respectively.)

Current rules in NHL forbid the outcome of a game to be a tie. If after 60 minutes of play the score is tied, the teams play additional time to determine a winner. This added time is called "overtime".

3. Overtime

When we are trying to predict the outcome of a game, it is important to know whether the game in question is a playoff game or is a regular-season game as the overtime rules differ. To avoid any confusion when explaining the rules of overtime, I have copied the relevant information from Rule 84 "Overtime" in the NHL rule book [**5**].

3.1. Playoffs. In the Stanley Cup Playoffs, when a game is tied after three (3) twenty (20) minute regular periods of play, the teams resume playing twenty (20) minute periods, changing ends for the start of each overtime period. The team scoring the first goal in overtime shall be declared the winner of the game.

We start by determining how a team would win in playoff overtime, since this is the easiest case. We have shown that we can model goals scored as independent Poisson processes with rates λ_1 and λ_2 per game (60 min), respectively. Team 1 wins if the first goal produced in overtime is from N_1^t, this event has probability $\lambda_1/(\lambda_1 + \lambda_2)$.

(1) The probability that Team 1 will win a playoff game in OT is:

$$S(0)\frac{\lambda_1}{(\lambda_1+\lambda_2)}$$

(2) The probability that Team 2 will win a playoff game in OT is:

$$S(0)\frac{\lambda_2}{(\lambda_1+\lambda_2)}$$

(where $S(0)$ is the probability the teams will be tied at the end of regulation, $P\{X = 0\} = S(0)$)

3.2. Regular-season. During regular-season games, if at the end of the three (3) regular twenty (20) minute periods, the score shall be tied. The teams will then play an additional overtime period of not more than five (5) minutes with the team scoring first declared the winner. The overtime period shall be played with each team at a numerical strength of four (4) skaters and one (1) goalkeeper.

For the regular-season the probability of a team scoring first in overtime remains the same, however the fact that regular-season over time lasts only 5 minutes adds an extra condition. The game winning goal must be scored within the first 5 minutes ($t = \frac{1}{12}$ since λ_1 and λ_2 are goals per 60 minute).

(1) The probability of Team 1 winning a regular-season game in OT is:

$$S(0)\frac{\lambda_1}{(\lambda_1+\lambda_2)}(1-e^{-\frac{1}{12}(\lambda_1+\lambda_2)})$$

(2) The probability of Team 2 winning a regular-season game in OT is:

$$S(0)\frac{\lambda_2}{(\lambda_1+\lambda_2)}(1-e^{-\frac{1}{12}(\lambda_1+\lambda_2)})$$

(where $S(0)$ is the probability the teams will be tied at the end of regulation, $P\{X = 0\} = S(0)$)

3.2.1. *Shootout.* During regular-season games, if the game remains tied at the end of the five (5) minute overtime period, the teams will proceed to a shootout. The rules governing the shootout shall be the same as the rules for Penalty Shots. The home team shall have the choice of shooting first or second. The teams shall alternate shots. Three (3) players from each team shall participate in the shootout and they shall proceed in such order as the Coach selects. If after each teams players have gone and the score remains tied, the shootout will continue until a winner is decided.

When determining the probability of winning in shootout we assume that each team has a equal chance to win. So the probability of a game going to shootout is $e^{-\frac{1}{12}(\lambda_1+\lambda_2)}$ which is the probability $P\{N_{\frac{1}{12}} = 0\}$ that no goal is scored in overtime. The total probability of winning in shootout is $0.5e^{-\frac{1}{12}(\lambda_1+\lambda_2)}$.

3.3. Who will Win? Now we have covered the probabilities of each possible outcome of a NHL game, to determine the probability of a team to win a game we need only to sum the probabilities of a team winning in each scenario.

(1) If the game in question is a regular season then the probability that Team 1 will win is,

$$\sum_{x>0} S(x) + S(0)\left(\frac{\lambda_1}{(\lambda_1+\lambda_2)}(1-e^{-\frac{1}{12}(\lambda_1+\lambda_2)}) + 0.5e^{-\frac{1}{12}(\lambda_1+\lambda_2)}\right)$$

(2) If the game in question is a playoff game then the probability that Team 1 will win is,

$$\sum_{x>0} S(x) + S(0)\frac{\lambda_1}{(\lambda_1+\lambda_2)}$$

4. Additional Notes

This simple model assumes that the scoring rate λt is constant. Though this is good to establish the basic idea, assuming that scoring is constant could hurt the accuracy of the model. One of the ways that we can improve the accuracy of the model would be to let $\lambda(t)$ be a rate that is dependent on time. This is called a non-homogeneous Poisson process.

DEFINITION 4.3. A *non-homogeneous Poisson process* with arrival rate $\lambda(t)$ a function of time, is defined as a counting process $\{N_t \mid t > 0\}$ which has the independent increment property and, satisfies;

$$\begin{aligned} P\{N_{t+dt} - N_t = 0\} &= 1 - \lambda(t)dt + o(dt), \\ P\{N_{t+dt} - N_t = 1\} &= \lambda(t)dt + o(dt), \\ P\{N_{t+dt} - N_t \geq 2\} &= o(dt) \end{aligned}$$

for all $t \geq 0, dt > 0$ the number of arrivals in time t satisfies

$$P\{N_t = x\} = \frac{m(t)^x e^{-m(t)}}{x!} \quad where \;\; m(t) = \int_0^t \lambda(u)du$$

Using a non-homogeneous Poisson process with rate $\lambda(t)$ would improve the accuracy of the model since we are able to adjust the scoring rate for different values of t, to compensate for the lack of scoring in the first period and the influx of goals in the last minutes.

What we now have is a method of determining what team will win a hockey game. We have assumed that Team 1 and Team 2 have scoring rates λ_1 and λ_2. But how do we determine what the values of λ? This topic had to be left out do to size restrictions, for more information on estimating λ, as well as a good method for collecting and sorting data can be found in [**4**].

Poisson processes are a very powerful tool for determining the probability of random events occurring over time. There is a wide range of phenomena that can be modeled using Poisson processes including but not limited to: scoring in some team sports, radioactive decay, and queuing theory. I have found that hockey, and soccer are Poisson processes and it is relatively easy to create models for since goals are always one point, I believe that football may also be Poisson and will be looking more into this in the future. Overall while a simple model like this may have little practical value, they are extremely useful for is providing insight into the problem at hand.

Bibliography

[1] J. F. C. Kingman, "Poisson Processes" Oxford University Press.

[2] W. Feller, (1950 1966) "An introduction to probability theory and its applications", 12, Wiley .

[3] Alan Ryder (2004). "Poisson Toolbox". 3

[4] Samuel E. Buttrey, Alan R Washburn, Wilson L. Price, (2011). "Estimating NHL Scoring Rates" Journal of Quantitative Analysis in Sports" Volume 7, Issue 3. 20

[5] National Hockey League (2012). "Official Rules 2012-2013'. 18

[6] R.G Gallager (2009). "Discrete Stochastic Processes" . 10

[7] P.X., Gallagher (1976). "On the distribution of primes in short intervals". Mathematika 23: 49. 8

[8] Ernie Croot (2010). "Some notes on the Poisson distribution". 8

[9] J. G. Skellam, The frequency distribution of the difference between two Poisson variates belonging to different populations, J. R. Statist. Soc., vol. 109, pp. 296, 1946. 17

[10] Abramowitz, M. and Stegun, I. A. (Eds.). "Modified Bessel Functions and ." 9.6 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 374-377, 1972. 17