1 Poisson process

1.1 Introduction

Text messages arrive on your cell phone at irregular times throughout the day. Accidents occur on the highway in a seemingly random distribution of time and place. Babies are born at chance moments on a maternity ward. All of these phenomena are well modeled by the Poisson process, a stochastic process used to model the occurrence, or arrival, of events over a continuous interval. Typically, the interval represents time.

A Poisson process is a special type of counting process. Given a stream of events that arrive at random times starting at $t = 0$, let N_t denote the number of arrivals that occur by time t, that is, the number of events in $[0, t]$. For instance, N_t might be the number of text messages received up to time t.

For each $t \geq 0$, N_t is a random variable. The collection of random variables $(N_t)_{t\geq 0}$ is a continuous-time, integer-valued stochastic process, called a counting process. Since N_t counts events in $[0, t]$, as t increases, the number of events N_t increases.

Definition 1.1. (Counting Process.) A counting process $(N_t)_{t\geq 0}$ is a collection of non-negative, integer-valued random variables such that if $0 \leq s \leq t$, then $N_s \leq N_t$.

Important note 1.1. There are several ways to characterize the Poisson process. One can focus on:

- the number of events that occur in fixed intervals
- when events occur, and the times between those events
- the probabilistic behavior of individual events on infinitesimal intervals.

This leads to three equivalent definitions of a Poisson process, each of which gives special insights into the stochastic model.

Definition 1.2. (Poisson Process - Definition 1). A Poisson process with parameter λ is a counting process $(N_t)_{t\geq 0}$ with the following properties:

1. $N_0 = 0$.
2. For all $t > 0$, N_t has a Poisson distribution with parameter λt.
3. (Stationary increments) For all $s, t > 0, N_{t+s} - N_s$ has the same distribution as N_t. That is, $P(N_{t+s} - N_s = k) = P(N_t = k) = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}$, for $k = 0, 1, \ldots$
4. (Independent increments) For $0 \leq q < r \leq s < t, N_t - N_s$ and $N_r - N_q$ are independent random variables.

Important note 1.2. The stationary increments property says that the distribution of the number of arrivals in an interval depends only on the length of the interval.

The independent increments property says that the number of arrivals on disjoint intervals are independent random variables.

Since N_t has a Poisson distribution, $E(N_t) = \lambda t$. That is, we expect about λt arrivals in t time units. Thus, the rate of arrivals is $\dfrac{E(Nt)}{t} = \lambda$.

Example 1.1. Starting at 6 a.m., customers arrive at Martha's bakery according to a Poisson process at the rate of 30 customers per hour (or equivalently one customer per two minutes). Find the probability that more than 65 customers arrive between 9 and 11 a.m.

Answer 1.1. Let $t = 0$ represent 6 a.m. Then 9 a.m. and 11 a.m. are represented as $t = 3$ and $t = 5$, respectively. The desired probability is thus $P(N_5 - N_3 > 65)$, with N_t is a Poisson distributed random variable, with mean $30t$. By stationary increments, we have:

$$P(N_5 - N_3 > 65) = P(N_2 > 65) = 1 - P(N_2 \leq 65) = 1 - \sum_{k=0}^{65} P(N_2 = k) = 1 - \sum_{k=0}^{65} \frac{e^{-2\lambda}(2\lambda)^k}{k!} = 0.2355$$

The result can be obtained by using R-software command: 1-ppois(65,2*30).

Example 1.2. Joe receives text messages starting at 10 a.m. at the rate of 10 texts per hour according to a Poisson process. Find the probability that he will receive exactly 18 texts by noon (12 a.m.) and 70 texts by 5 p.m.

Answer 1.2. Let $t = 0$ represent 10 a.m. Then 12 a.m. and 5 p.m. are represented as $t = 2$ and $t = 7$, respectively. The desired probability is thus $P(N_2 = 18, N_7 = 70)$, with N_t is a Poisson distributed random variable, with mean $10t$. By stationary increments, we have:

$$\begin{aligned} P(N_2 = 18, N_7 = 70) &= P(N_2 = 18, N_7 - N_2 = 52) = P(N_2 = 18)P(N_7 - N_2 = 52) \\ &= P(N_2 = 18)P(N_5 = 52) = \frac{e^{-10*2}(10*2)^{18}}{18!}\frac{e^{-10*5}(10*5)^{52}}{52!} = 0.0045 = 0.45\%, \end{aligned}$$

where the second equality is because of independent increments, and the third equality is because of stationary increments. The result can be obtained by using R-software command: dpois(18,2*10)*dpois(52,5*10).

Important note 1.3. It would be incorrect to write $P(N_2 = 18, N_7 - N_2 = 52) = P(N_2 = 18, N_5 = 52)$. It is not true that $N_7 - N_2 = N_5$. The number of arrivals in $(2, 7]$ is not necessarily equal to the number of arrivals in $(0, 5]$.What is true is that the distribution of $N_7 - N_2$ is equal to the distribution of N_5. Note that while $N_7 - N_2$ is independent of N_2, the random variable N_5 is not independent of N_2. Indeed, $N_5 \geq N_2$.

Example 1.3. On election day, people arrive at a voting center according to a Poisson process. On average, 100 voters arrive every hour. If 150 people arrive during the first hour, what is the probability that at most 350 people arrive before the third hour?

Answer 1.3. Let N_t denote the number of arrivals in the first t hours. Then, $(N_t)_{t\geq 0}$ is a Poisson process with parameter $\lambda = 100$. Given $N_1 = 150$, the distribution of $N_3 - N_1 = N_3 - 150$ is equal to the distribution of N2. This gives

$$\begin{aligned} & P(N_3 \leq 350|N_1 = 150) = P(N_3 - N_1 + N_1 \leq 350|N_1 = 150) = P(N_3 - N_1 \leq 200|N_1 = 150) \\ = \; & P(N_3 - N_1 \leq 200) = P(N_2 \leq 200) = \sum_{k=0}^{200} P(N_2 = k) = \sum_{k=0}^{200} \frac{e^{-100*2}(100*2)^k}{k!} = 0.519. \end{aligned}$$

The result could be obtained by using R-software command ppois(200,200).

How do we model a Poisson process? Suppose we want to model a Poisson process $(X_t)_{t\geq 0}$ in a time interval $[0, T]$. Note that X_t is the number of arrivals in $[0, t]$. We divide the interval into n equal subintervals by discrete time points $0 = t_0 < t_1 < \cdots < t_n = T$, with the time step $h = t_i - t_{i-1}$. The value of X_{t_i} is determined by a recursive formula: $X_{t_i} = X_{t_{i-1}} + U$, where U is a Poisson random variable with mean $h\lambda$, i.e., $U \sim P(h\lambda)$. All we need to simulate the Poisson process is to generate a Poisson random variable, which can be done by using command "rpois(1,hλ" in software R. The recursive (iteration) process occurs as follows:

- $X_0 = 0$;
- $X_1 = X_0 + U$

- $X_2 = X_1 + U$
- $\cdots$
- $X_n = X_{n-1} + U$

For each iteration, as U is a random variable, the value of U may change.

```
Pois=function(lambda,T,N) {
### we need to specify the input of the Poisson process
# T=1  expiry time
# N=100  number of simulation points
h=T/N  # the timestep of the simulation
X=rep(0, (N+1)) # generate Poisson vector, #with length N+1
X[1]=0
for(i in 1:N) { X[i+1]=X[i] +rpois(1,lambda*h)}
return(X)
}
```

We now try to vectorise the above code using the fact that $X_t - X_s$ is a Poisson random variable with mean and variance (t-s) and

$$X_{t_i} = (X_{t_i} - X_{t_{i-1}}) + (X_{t_{i-1}} - X_{t_{i-2}}) + \cdots + (X_1 - X_0).$$

In other words, each X_{t_i} is a sum of i i.i.d Poisson random variables with mean and variance $h * \lambda$, where h is the time step and λ is the intensity of the Poisson process. This idea is the key for producing a vectorized code. We first generate n i.i.d Poisson random variables by using command in R, "rpois(n,$\lambda * h$)", then we use "cumsum" command to cumulatively sum up the resulted vector. The code is provided as follows:

```
### an alternative code for Poisson process
## this is R code for simulating Poisson process.
Pois=function(lambda,T,N) {
### we need to specify the input of the Poisson process
# T=1  expiry time
# N=100  number of simulation points
ptm <- proc.time() #start clock
X<-c(0,cumsum(rpois(N,lambda*t/N))) ## a method of vetorized code
## we use the fact that X_t - X_s
##is a Poisson random variable with mean and variance (t-s)
return(X)
#stop clock
proc.time() - ptm}
```

```
## code used to plot a sample Poisson process if we want to
steps <- seq(0,T,length=n+1)
plot(steps,X,type="l")
```

It is not easy as it looks to run many sample paths of Poisson process. The following code will allow us to do it.

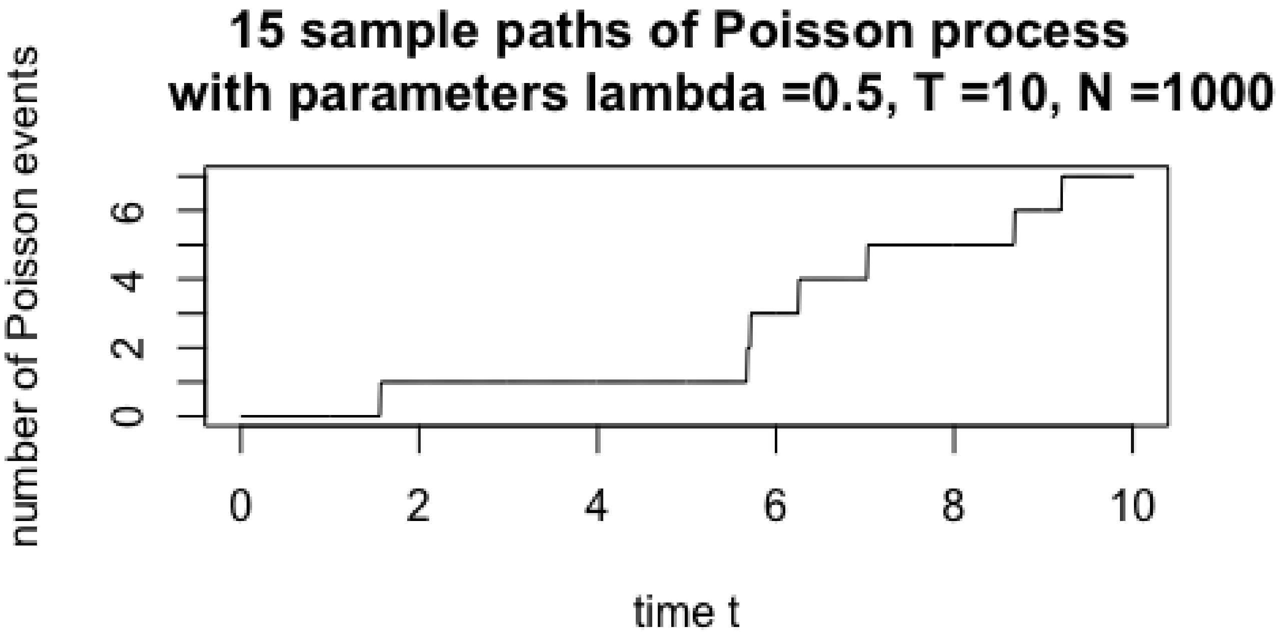

Figure 1:

```
PoisSamplepaths <- function(lambda,T,N,nt)
{
proc.time() - ptm
h=T/N
t=seq(0,T,by=h) # a way to produce a sequence of time points
X=matrix(rep(0,length(t)*nt), nrow=nt)
# #return(X)
for (i in 1:nt) {X[i,]= Pois(lambda,T=T,N=N)}
# # ##Plot
ymax=max(X); ymin=min(X) #bounds for simulated prices
plot(t,X[1,],t='l',main='50 sample paths of Poisson process
with parameters lambda =0.5, T =10, N =1000',ylim=c(ymin, ymax), col=1,
ylab="number of Poisson events",xlab="time t")
for(i in 2:nt){lines(t,X[i,], t='l',ylim=c(ymin, ymax),col=i)}
proc.time() - ptm
}
```

1.2 Arrival, inter-arrival time

For a Poisson process with parameter λ, let X denote the time of the first arrival. Then, $X > t$ if and only if there are no arrivals in $[0, t]$. Thus, $P(X > t) = P(N_t = 0) = e^{-\lambda t}$, for $t > 0$. Hence, X has an exponential distribution with parameter λ.

The exponential distribution plays a central role in the Poisson process. What is true for the time of the first arrival is also true for the time between the first and second arrival, and for all inter-arrival times. **A Poisson process is a counting process for which inter-arrival times are independent and identically distributed exponential random variables.**

Definition 1.3. (Poisson Process - Definition 2). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. exponential random variables with parameter λ. For $t > 0$, let $N_t = \max\{n : X_1 + ... + X_n \leq t\}$, with $N_0 = 0$. Then, $(N_t)_{t\geq 0}$ defines a Poisson process with parameter λ.

Let $S_n = X_1 + ... + X_n$, for $n = 1, 2, ...$ We call $S_1, S_2, ...$ the arrival times of the process, where S_k is the time of the kth arrival. Furthermore, $X_k = S_k - S_{k-1}$, for $k = 1, 2, \ldots$ is the inter-arrival time between the $(k-1)$-th and k-th arrival, with $S_0 = 0$.

Important note 1.4. The arrival time $S_n = X_1 + X_2 + ... + X_n$, where X_k is the time of the k-th arrival, has a gamma distribution with parameters n and λ. The density of S_n is $f_{S_n}(t) = \lambda \dfrac{e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$, for $t > 0$ (notice the similarity of the formula with that of the probability mass function of Poisson random variables). The mean and variance of S_n are $E(S_n) = \dfrac{n}{\lambda}, Var(S_n) = \dfrac{n}{\lambda^2}$.

For a general gamma distribution, the parameter n does not have to be an integer. When it is, the distribution is sometimes called an Erlang distribution. Observe that if $n = 1$, the gamma distribution reduces to the exponential distribution with parameter λ.

A direct method for constructing, and simulating, a Poisson process:

1. $S_0 = 0$.
2. Generate i.i.d. exponential random variables $X_1, X_2, \ldots$
3. Let $S_n = X_1 + ... + X_n$, for $n = 1, 2, ...$
4. For each $k = 0, 1, ...$, let $N_t = k$, for $S_k \leq t < S_{k+1}$.

```
###Code the Poisson process using definition 2
# Start the clock!
ptm <- proc.time()
#set.seed(123)  #keep the random state fixed
lambda <- 0.8  # average occurrence per unit time
T <- 10 #10 units of time
avg <- lambda*T  # average occurence per 10 unit times
avg
t <- 0  # assign t to the value 0 # t will a collection of exponential random variable
N <- 0 # assign N to the value 0
k <- 0 # assign k to the value 0
continue <- TRUE # assign a dummy variable \continue" to the value TRUE
# continue the following loop while the \continue" variable stills receive a value
# TRUE
while(continue){
event <- rexp(1, lambda)  # assign the variable \event" as a exponential random variable
# we will model the Poisson process during the interval [0,T]: N_T=max{n: X_1+X_2+...+X_n<=T}
if(sum(t) + event < T){
k <- k +1 # k counts for the number of iterations
```

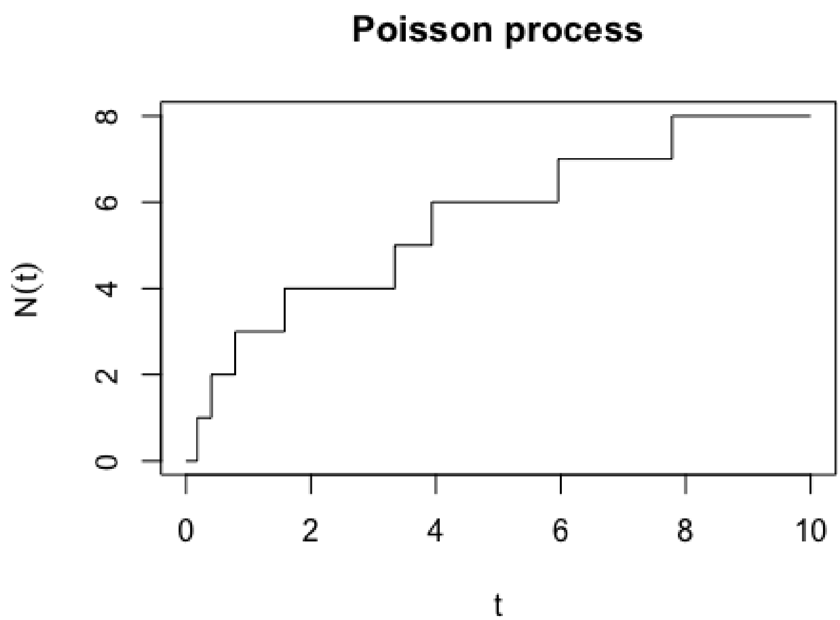

Figure 2:

```
N <- c(N,k) # variable N lists the number of occurrence
t <- c(t, event) # variable t lists the occurrence time after the previous occurrence
} else {
continue <- FALSE  # assign a dummy variable \continue" to the value FALSE to stop the loop
t <- cumsum(t) # variable t lists the occurrence time of the whole Poisson process
N <- c(N,k) # adding the final occurrence
t <- c(t,T) # adding the expiry time
}
}
N  # list the number of occurrences
t # list the time corresponding to each occurrence
plot(t, N,type="s", main="Poisson process",ylab=expression(N(t)),xlim=c(0,T))

# Stop the clock
proc.time() - ptm
```

Definition 1.4. (Memorylessness). A random variable X is memoryless if, for all $s, t > 0$, $P(X > s + t|X > s) = P(X > t)$ or equivalently $P(X > s + t) = P(X > s)P(X > t)$.

Important note 1.5. Memorylessness means that regardless of how long you have waited, the distribution of the time you still have to wait is the same as the original waiting time.

Theorem 1. The exponential distribution is memoryless.

Proof. Let X be an exponential random variable with parameter λ. That means $P(X > t) = e^{-\lambda t}$.

For any $s, t > 0$, we have $P(X > s + t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t} = P(X > s)P(X > t)$. □

Important note 1.6. The exponential distribution is the only continuous distribution that is memoryless.

Theorem 2. (Minimum of Independent Exponential Random Variables). Let $X_1, ..., X_n$ be independent exponential random variables with respective parameters $\lambda_1, \lambda_2, ..., \lambda_n$. Let $M = \min(X_1, ..., X_n)$.

1. For $t > 0$, $P(M > t) = e^{-t(\lambda_1+...+\lambda_n)}$. That is, M has an exponential distribution with parameter $\lambda_1 + \lambda_2 + ... + \lambda_n$.
2. For $k = 1, ..., n$, $P(M = X_k) = \dfrac{\lambda_k}{\lambda_1 + ... + \lambda_n}$.

Proof. 1. For $t > 0$, we have

$$P(M > t) = P(\min(X_1, ..., X_n) > t) = P(X_1 > t, ..., X_n > t) = P(X_1 > t)...P(X_n > t) = e^{-t(\lambda_1+...+\lambda_n)}.$$

2. For $1 \le k \le n$, conditioning on X_k gives:

$$\begin{aligned}
&P(M = X_k) = P(\min(X_1, ..., X_n)) = P(X_1 \ge X_k, ..., X_{k-1} \ge X_k, X_{k+1} \ge X_k, ...X_n \ge X_k) \\
&= \int_0^\infty P(X_1 \ge X_k, ..., X_{k-1} \ge X_k, X_{k+1} \ge X_k, ...X_n \ge X_k | X_k = t) f_{X_k}(t)dt \\
&= \int_0^\infty P(X_1 \ge t, ..., X_{k-1} \ge t, X_{k+1} \ge t, ...X_n \ge t) f_{X_k}(t)dt \\
&= \int_0^\infty P(X_1 \ge t)...P(X_{k-1} \ge t)P(X_{k+1} \ge t)...P(X_n \ge t)\lambda_k e^{-\lambda_k t}dt \\
&= \int_0^\infty e^{-\lambda_1 t}...e^{-\lambda_{k-1} t}e^{-\lambda_{k+1} t}...e^{-\lambda_n t}\lambda_k e^{-\lambda_k t}dt = \lambda_k \int_0^\infty e^{-t(\lambda_1+...+\lambda_{k-1}+\lambda_k+\lambda_{k+1}+...+\lambda_n)}dt \\
&= \frac{\lambda_k}{\lambda_1 + ... + \lambda_n}.
\end{aligned}$$

□

Example 1.4. A Boston subway station services the red, green, and orange lines. Subways on each line arrive at the station according to three independent Poisson processes. On average, there is one red train every 10 minutes, one green train every 15 minutes, and one orange train every 20 minutes.

a) When you arrive at the station what is the probability that the first subway that arrives is for the green line?

b) How long will you wait, on average, before some train arrives?

Answer 1.4. a) Let X_G, X_R, and X_O denote, respectively, the times of the first green, red, and orange subways that arrive at the station. The event that the first subway is green is the event that X_G is the minimum of the three independent random variables. The desired probability is

$$P(\min(X_G, X_R, X_O) = X_G) = \frac{1/15}{1/10 + 1/15 + 1/20} = \frac{4}{13} = 0.31$$

b) The time of the first train arrival is the minimum of X_G, X_R, and X_O, which has an exponential distribution with parameter $\dfrac{1}{10} + \dfrac{1}{15} + \dfrac{1}{20} = \dfrac{13}{60}$.
Thus, you will wait, on average $60/13 = 4.615$ minutes.

Example 1.5. (Marketing.) A commonly used model in marketing is the so-called NBD model, introduced in late 1950s by Andrew Ehrenberg (1959) and still popular today. (NBD stands for negative binomial distribution, but we will not discuss the role of that distribution in our example.) **Individual customers' purchasing**

occasions are modeled as a Poisson process. Different customers purchase at different rates λ and often the goal is to estimate such λ. Many empirical studies show a close fit to a Poisson process.

Suppose the time scale for such a study is in days. We assume that Ben's purchases form a Poisson process with parameter $\lambda = 0.5$ (on average one purchase per two days). Consider the following questions of interest.

1. What is the average rate of purchases?
2. What is the probability that Ben will make at least three purchases within the next 7 days?
3. What is the probability that his 10th purchase will take place within the next 20 days?
4. What is the expected number of purchases Ben will make next month?
5. What is the probability that Ben will not buy anything for the next 5 days given that he has not bought anything for 2 days?

Answer 1.5. Let N_t be the number of Ben's purchases in t days.

1. The average rate of purchases is 0.5 purchase per day or one purchase per two days.
2. The number of Ben's purchases N_t in t days is a Poisson random number with mean $0.5t$. The probability that Ben will make at least three purchases within the next 7 days is then

$$\begin{aligned} P(N_7 \geq 3) &= 1 - P(N_7 < 3) = 1 - P(N_7 = 0) - P(N_7 = 1) - P(N_7 = 2) \\ &= 1 - e^{-0.5*7}\frac{(0.5*7)^0}{0!} - e^{-0.5*7}\frac{(0.5*7)^1}{1!} - e^{-0.5*7}\frac{(0.5*7)^2}{2!} = 0.679. \end{aligned}$$

3. The probability that his 10th purchase will take place within the next 20 days is equal to the probability that the number of purchases in the next 20 days is at least 10. The desired probability is then

$$P(N_{20} \geq 10) = 1 - P(N_{20} \leq 9) = 0.542.$$

The result can be obtained using the R software command ppois(9,20*0.5).

4. The number of purchases Ben will make next month is a Poisson random variable with mean $30 * 0.5 = 15$.
5. The probability that Ben will not buy anything for the next 5 days given that he has not bought anything for 2 days is $P(N_5 = 0|N_2 = 0) = P(N_5 - N_2 = 0|N_2 = 0) = P(N_5 - N_2 = 0) = P(N_3 = 0) = e^{-3*0.5} = 22.3$

1.3 Exercise

1. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda = 1.5$. Find the following:
   a) $P(N_1 = 2, N_4 = 6)$
   b) $P(N_4 = 6|N_1 = 2)$
   c) $P(N_1 = 2|N_4 = 6)$

2. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda = 2$. Compute $E[N_3 N_4]$.
3. Calls are received at a company call center according to a Poisson process at the rate of five calls per minute.
   (a) Find the probability that no call occurs over a 30-second period.

(b) Find the probability that exactly four calls occur in the first minute, and six calls occur in the second minute.

(c) Find the probability that 25 calls are received in the first 5 minutes and six of those calls occur in the first minute.

4. Starting at 9 a.m., patients arrive at a doctor's office according to a Poisson process. On average, three patients arrive every hour.

   (a) Find the probability that at least two patients arrive by 9:30 a.m.

   (b) Find the probability that 10 patients arrive by noon and eight of them come to the office by 11 a.m.

   (c) If six patients arrive by 10 a.m., find the probability that only one patient arrives by 9:15 a.m.

5. Let $(N_t)_{t\geq 0}$ be a Poisson process. Explain what is wrong with the following proof that N_3 is a constant.

   $E((N_3)^2) = E(N_3N_3) = E(N_3(N_6-N_3)) = E(N_3)E(N_6-N_3) = E(N_3)E(N_3) = E(N_3)^2$. Thus, $Var(N_3) = E((N_3)^2) - E(N_3)^2 = 0$, which gives that N_3 is a constant with probability 1.

6. Occurrences of landfalling hurricanes during an El Nino event are modeled as a Poisson process. The research shows that "During an El Nino year, the probability of two or more hurricanes making landfall in the United States is 28%." Find the rate of the Poisson process.

7. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $\{N_t, t \geq 0\}$ be a Poisson process with intensity $\lambda > 0$. Compute the covariance and correlation coefficient of N_t and N_s, for $s < t$.