1 Intro to Probability Theory

Kolmogorov's axioms
The probability measures $\mathbb{P}(.)$ satisfies

1. $0 \leq \mathbb{P}(E) \leq 1$ for any event E
2. $\mathbb{P}(S) = 1$
3. For any chain of mutually exclusive events

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

Frequentists definition of probability
If the experiment is repeated independently over and over again, the proportion of time that the event E occurs is $\mathbb{P}(E)$

Conditional probability
The conditional probability of E_1, conditional on E_2 is defined as

$$\mathbb{P}(E_1|E_2) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)}$$

Bayes first rule:
If $\mathbb{P}(E_1), \mathbb{P}(E_2) > 0$ then

$$\mathbb{P}(E_1|E_2) = \mathbb{P}(E_2|E_1)\frac{E_1}{E_2}$$

Independence of two events
An event is defined to be **independent** iff

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \times \mathbb{P}(E_2)$$

Independence of several events
Several events are defined to be independent iff for every subset $\{i_1, i_2 ... i_n\}$ of $1, 2...n$

$$\mathbb{P}(\bigcap_{j=1}^{\infty} E_{ij}) = \prod_{j=1}^{\infty}(E_{ij})$$

Partition
A sequence of events $E_1, E_2....E_n$ such that

1. $S = \cup_{i=1}^{n} E_i$
2. $E_i \cap E_j = \theta$ for all $i \neq j$

Is called a partition of S.

Laws of total probability
Given a partition $\{E_1, E_2...E_n\}$ of S such that $\mathbb{P}(E_i) > 0$ for all i, the probability

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|E_i) \times \mathbb{P}(E_i)$$

Bayes second rule:
Given a partition $\{E_1, E_2, E_3...E_n\}$ of S such that $\mathbb{P}(E_i) > 0$, for all i we have for any event A such that $\mathbb{P}(A) > 0$

$$\mathbb{P}(E_i|A) = \frac{\mathbb{P}(A|E_i)\mathbb{P}(E_i)}{\sum_{j=1}^{n} \mathbb{P}(A|E_j)\mathbb{P}(E_j)}$$

2 Random Variables

Random Variables
A random variables is a real valued function defined on the sample space.

$$
\begin{aligned}
X : S &\to \mathbb{R} \\
\omega &\to X(\omega)
\end{aligned}
$$

Cumulative Distribution function
The cdf of a function.

$$F_X(x) = \mathbb{P}(X \leq x)$$

Probability mass function
The probability mass function of a discrete random variable is defined by

$$p_X(x) = \mathbb{P}(X = x)$$

Binomial distribution

Given

- Success with probability p
- Numbers of independent repetitions n

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Bernouli distribution

$$X \sim \text{Bern}$$

With a pmf

$$p_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Geometric distribution

$$P_X(x) = (1-p)^{x-1} p \text{ for } S_x$$

Poisson distribution

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Poisson distribution is good for modelling assuming continuity, stationarity, independence and non-simultaneity.

λ represents the intensity of the phenomenon.

Uniform Distribution Uniform distribution is defined as

$$X \sim U_{[\alpha,\beta]}$$

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

Exponential distribution

$$X \sim Exp(\lambda)$$

The probability density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Gamma Distribution

$$X = \Gamma(\alpha, \lambda)$$

Its probability funciton is defined by

$$f_X(x) = \begin{cases} \frac{\lambda e^{\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Where

$$\Gamma(\alpha) = \int_0^{+\inf} e^{-x} x^{\alpha-1} dx$$

Normal Distribution
A random variable is said to have a normal distribution which has parameters μ and σ with the following probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parameters of a distribution The two most important quantities of a distribution are

- Expectation
- Variance

Expectation is defined as:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x dF_X(x)$$

Cauchy Distribution A random variable is said to have a Cauchy distribution if its probability density function is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

The Cauchy distribution does not have an expectation value.

Moments of a random variable

Lets define the **kth moment of a random variable** as

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k dF_X(x)$$

Variance
The variance of a random variable is defined as

$$Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

Explicitly

$$Var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 dF_X(x)$$

Properties of variance
Property 1: $Var(X) = E(X^2) - (E(X))^2$
Property 2: $Var(aX + b) = a^2 Var(X)$

The standard deviation is defined as $\sqrt{Var(X)}$.

Joint distribution function

The joint cumulative distribution function is defined by

$$F_{XY}(x,y) = \mathbb{P}(X \geq x, Y \geq y) \forall (x,y) \in \mathbb{R} \times \mathbb{R}$$

Marginal cdf's are obtained by

$$F_X(x) = F_{XY}(x, +\infty) \text{ and } F_Y(y) = F_{XY}(+\infty, y)$$

Continuous case
X and Y are said to be jointly continuous if there exists a function $f_{XY}(x,y) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ such that for any sets A and B of real numbers.

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{XY}(x,y) dy dx$$

Independent random variables
The random variables X and Y are said to be independent if for all $(x,y) \in \mathbb{R}$

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y)$$

Their expected value will behave like:

$$\mathbb{E}(h(x)g(x) = \mathbb{E}(h(x)) \times \mathbb{E}g(x)$$

Covariance of two random variables

$$Cov(X,Y) = \mathbb{E}(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$$

If $Cov(X,Y) > 0$, X tends to increase as Y tends to increase.

If $Cov(X,Y) < 0$, X tends to decrease as Y tends to increase and vica versa.

Variance of a sum of random variables

$$\begin{aligned}
Var(\sum_{i=1}^{n} X_i) &= Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i X_j) \\
&= \sum_{i=1}^{n} Cov(X_i, X_i) + \sum_{i=1}^{n} \sum_{j \neq 1} Cov(X_i, X_j) \\
&= \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i<j} Cov(X_i, X_j)
\end{aligned}$$

Sum of independent variables
Let X and Y be 2 random variables. We know

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

If X and Y are independent then

$$f_{X+Y} = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

Important slides to look at: Slides 107-110

Moment Generating function
The moment generating function of a random variable is defined as

$$\varphi_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} dF_x(x)$$

If $\varphi_X(t)$ exists around zero, then for any positive integer n, $\varphi_X^{(n)} = \frac{d^n\varphi_X(t)}{dt^n}|_{t=0}$
Characteristic equation is defined as:

$$\phi_X(t) = \mathbb{E}(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} dF_X(x)$$

To find the expectation and variance from the mgf, see slide 114-116

Properties

- Moment generating function uniquely determines the distribution of a random variable
- For any real numbers $\varphi_{aX+b}(t) = e^{bt}\varphi_X(at)$
- $\varphi_{X+Y}(t) = \varphi_X(t) * \varphi_Y(t)$ if X and Y are two independent variables.

Joint Moment Generating function

$$\varphi_{X_1X_2...X_n}(t_1, t_2...t_n) = \mathbb{E}(e^{t_1X_1+t_2X_2+...+t_nX_n})$$

Multivariate normal distribution
The random variables $X_1, X_2...X_n$ are said to have a multivariate normal distribution if there exists m independent standard normal variables $Z_1...Z_m$ such that

$$X_1 = \mu_1 + a_{11}Z_1 + a_{12}Z_2 + ... + a_{1m}Z_m$$

and so on.

Cramer-Wold theorem
The random variables $X_1, X_2, ..., X_n$ are said to have a multivariate normal distribution iff for any vector $(b_1, b_2...b_n) \in \mathbb{R}$ the linear combination $(b_1X_1 + b_2X_2 + ... + b_nX_n)$ has a normal distribution.

Copula
There always exists a function C_{XY} such that $\forall X, Y$

$$F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y))$$

Where C_{XY} is a copula of the distribution.

Independence Copula
If $C_{XY}(uv) = uv$ then $F_{XY}(x, y) = F_X(x) + F_Y(y)$

Gaussian Copula
Two marginal normal distributions will interact so as to produce a bivariate normal distribution with correlation ρ iff

$$C_{X,Y}(u, v) = \int_0^u \int_0^v \frac{exp(\frac{-\rho^2(\theta^{-1}(\zeta)-\theta^{-1}(\mu))^2}{2(1-\rho^2)})}{\sqrt{1-\rho^2}} d\zeta d\mu$$

But any other copula function gives us a joint distribution with normal marginals which is not bivariate normal.

Normally distributed functions

If X and Y are normally distributed, X and Y are independent iff $Cov(X, Y) = 0$.

Markov's Inequality

If X is a random variable that takes only non negative values, then for any value $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Chevyshev's inequality

Let X be a random variable with mean μ and variance σ^2. Then for any value $k > 0$,

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Strong law of large numbers

Let $X_1, X_2 ... X_n$ be a sequence of independent random variables having a common distribution, and let $\mathbb{E}(X_i) = \mu < \infty$. Then with probability 1

$$\bar{X_n} = \frac{X_1 + X_2 + ... + X_n}{n} = \mu \text{ as } n \to \infty$$

Central Limit Theorem

Let $X_1, X_2 ... X_n$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(X_i) = \mu < \infty$ and $Var(X_i) = \sigma^2 < \infty$. Then the distribution of $\frac{S_n - n_\mu}{\sigma n}$ tends to the standard normal as $n \to \infty$. That is

$$\mathbb{P}(\frac{S_n - n_\mu}{\sigma \sqrt{n}} \leq z) \to \theta(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}dx}$$

Where $S_n = n\bar{X_n}$.

3 Conditional Probability

Conditional Probability discrete case
The conditional expectation of X given by $Y = y$ is defined by

$$\mathbb{E}(X|Y = y) = \sum_{x \in S_x} x p_{X|Y}(x|y)$$

Conditional probability density function

The conditional probability density function of X given $Y = y$ is defined for all values y such that $f_Y(y) > 0$ by

$$f_{X|Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Marginal density is defined as

$$\mathbb{P}(y - \delta/2 \leq Y \leq y + \delta/2) \approx \delta f_Y(y)$$

Law of iterred expectations
Let X and Y be two random variables. Then:

$$\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$$

$\mathbb{E}(X)$ is the weighted average of the conditional expected values of X given $Y = y$, each of the terms $\mathbb{E}(X|Y = y)$ being weighted by the probability of the event on which it is conditioned.

This also satisfies:

$$\mathbb{E}[\mathbb{E}(X|Y)g(Y)] = \mathbb{E}(Xg(Y))$$

for any function g where both expectations exist.

Conditional Variance
The conditional variance of a random variable X given that $Y = y$ is given by

$$Var(X|Y = y) = \mathbb{E}[(X - \mathbb{E}(X|Y = y))^2|Y = y]$$

All properties of Variance hold here.

Conditional variance formula
Let X and Y be two random variables. Then

$$Var(X) = \mathbb{E}[Var(X|Y) + Var[\mathbb{E}(X|Y)]]$$

Iterred Conditonal expectation

$$\mathbb{E}(X|Y) = \mathbb{E}[\mathbb{E}(X|W, Y)|Y]$$

where $\mathbb{E}(X|W, Y)$ is a random variable $\mathbb{E}(X|W = w, Y = y)$

(Example of this being used slide 190-194)

4 Markov Chains

textbfDefinition of Markov chains

A discrete Markov Chain is

- A discrete-time stochastic process $\{X_n, n = 0, 1, 2...\}$
- That takes on a finite or countable number of possible value (discrete state space S_x) called "states" and usually denoted by integers without loss of generality
- Satisfying the Markov property. That is, for any $i, j, i_0, i_1...i_{n-1} \in S_x = \{0, 1, ...\}$

$$\mathbb{P}$$

Champan-Kolmogorov Equations
For all $n, m \geq 0, i, j \geq 0$ we have

$$P_{ij}^{[n+m]} = \sum_{k \in S_X} P_{ik}^{[n]} P_{kj}^{[m]}$$

Note: $P^{[0]} = I$
Transformed transition matrix
Assume that we've entered some states $\mathcal{A}$ by time m

Then we reset the matrix to:

$$Q_{ij} = \mathbb{P}*(X_{n+1} = j | X_n = i) = \begin{cases} 1 & \text{if } i \in \mathcal{A}, j = i \\ 0 & \text{if } i \in \mathcal{A}, j \neq i \\ P_{ij} & \text{if } i \notin \mathcal{A} \end{cases}$$

Accessible states

State j is said to be accessible from state i denoted by $i \to j$ if

$$[P^n]_{ij} = P_{ij}^[n] > 0$$

for some $n \geq 0$.

Communicating states
Two states i and j communicate, denoted by $i \leftrightarrow j$ if i is accessible form j and j is accessible from i.

Communication classes
A class $\mathcal{C}$ is a non-empty set of states such that for each state $i \in \mathcal{C}$ i communicates with all $j \in \mathcal{C}$ and does not communicate with any $j \notin \mathcal{C}$

Irreducible chain
A markov chain is said to be irreducible if there is only one class, that is if all states communicate with each other.

Classification of states
State i is said to be

1. **recurrent** if $p_i = 1$
2. **transient** if $p_i < 1$

State i is recurrent if $\sum_{n=0}^{\infty} P_{ii}^[n] = +\infty$
and transient if the same sum $< \infty$

Period
The period of a state i is denoted by $d(i)$ is the greatest common divisor of the values of n for which $P_{ii}^{[n]} > 0$.

If the period of the state is 1, the state is aperiodic.

In a Markov Chain, all state in a given class has the same period. (Proof slide 235)

Positive and null recurrent
A state i of a Markov Chain is **positive recurrent** if its recurrent and $\mathbb{E}(T_{ii}) < +\infty$. It is **null recurrent** if it is recurrent and $\mathbb{E}(T_{ii}) = +\infty$.

Ergodic A state is said to be **ergodic** if it is positive-recurrent and aperiodic. A class of ergodic states is known as an ergodic chain.

Theorem - Ergodic markov Chain

For an ergodic markov chain

$$lim_{n\to\infty} P_{ij}^{[n]} \text{ exists for all } j$$

and is independent of i. Furthermore, letting

$$\pi_j = \lim_{n\to\infty} P_{ij}^{[n]}$$

then $\pi = (\pi_0, \pi_1, \pi_2, ...)$ is the unique nonnegative solution of the system of equations

$$\begin{cases} \pi_j = \sum_{[i\in S_x]} \pi_i P_{ij} \\ 1 = \sum_{[i\in S_X]} \pi_j \end{cases}$$

Stationary distribution

Take the law of total probability:

$$\mathbb{P}(X_{n+1}) = \sum_{i\in S_X} \mathbb{P}(X_{n+1} = j|X_n = i)\mathbb{P}(X_n = i)$$

Suppose that the distribution of X_n is π. Then

$$\mathbb{P}(X_{n+1} = j) = \sum_{i\in S_X} P_{ij}\pi_j = \pi_j$$

The vector π is called the stationary distribution. It is also the long run probability of time that the process will be in the state j.
Limiting probabilities

For examples, see slides 249-253
Non Ergodic cases
If the chain is ergodic then P^n converges as $n \to +\infty$ and there is one single stationary distribution π, which is the unique solution of $\pi = \pi P$

Non ergodic means

- not reducible
- periodic
- null recurrent

A non-irreducible chain means with one recurrent class and some transient classes meansthat it experiences a:

- Stationary distribution = long term behaviour of chain
- transient classes play a role only for a finite number of transitions

- After a certain number of transitions the chain enters a recurrent class and never leaves it.

Ergodic theorem

Let $\{X_n, n \geq 0\}$ be an ergodic Markov Chain with stationary probabilities $\pi_j \geq 0$. Let r be a bounded function: $S_X \to R \subset \mathbb{R}$. Then with probability 1

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} r(X_n) = \sum_{j \in S_X} r(j)\pi_j$$

for any starting value X_0.
Loosely time average = space average

In fact most of the limit theorems for sums of independent random variables essentially holds true for sums of dependent random variables, provided that the dependance isn't too strong.

Canonical form
You can always rearrange states of the chain to have this form

$$\mathbf{P}^* = \left(\begin{array}{c|c} \mathbf{P}_t & \mathbf{P}_{tr} \\ \hline \mathbf{0} & \mathbf{P}_r \end{array} \right)$$

If R is the number of recurrent states then

- P_t is the $(T \times T)$ matrix of transitions from a transient state to a transient state.

- P_{tr} is the matrix of transitions from a transient state to a recurrent state
- P_r is the matrix of transitions from a recurrent state to a recurrent state.

$P*$ is called the canonical form of P.

How long will the chain be in transient states before its absorbed?

Conditioning on first transition from i we get:

$$S_{ij} = \delta_{ij} + \sum_{k=1}^{T+R} P_{ik} * S_{kj} = \delta_{ij} + \sum_{k=1}^{T} [P_t]_{ik} + S_{ij}$$

where δ_{ij} is the Kronecker delta. $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

In Matrix notation

$$S = I + P_t S$$

so:

$$S = (I - P_T)^{-1}$$

It follows that S can be referred to as the **fundamental matrix**.

On average how long will it take for the chain to be absorbed?

Let σ_i be the amount of time before the chain is absorbed, given the chain is now in transient state i.

$$\sigma_i = \sum_{j=1}^{T} S_{ij}$$

If $\sigma = (\sigma_1 \sigma_2 .... \sigma_T)^t$ we have

$$\sigma = Se = (1 - P_t)^{-1} e$$

For absorption examples see slides 283-289
Probability generating function

The probability generating function of a non negative integer valued random value X iis defined by

$$G_X(s) = \mathbb{E}(s^X) = \sum_{j=0}^{+\infty} s^j \mathbb{P}(X = j)$$

The pgf $G_X(s)$ is such that

1. It exists for all values $s \in [0, 1]$ as $\sum_j \mathbb{P}(X = j) = 1$
2. It is continusous, non decreasing and convex over $[0, 1]$
3. $G_X(0) = \mathbb{P}(X = 0)$ and $G_X(1) = 1$
4. It is infinitely many times differentiable over $[0, 1]$
5. (See slide 292)
6. (See slide 293)

Branching process
The evolution of a total population as time proceeds is called a **branching process**

Galton Watson definition of the process

By assumption, $\{Y_{n,k}\}$ are i.i.d random variables with common probability distribution $\mathbb{P}(Y = j) = p_j$ for $j = 0, 1, 2...$ and

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

5 The exponential distribution and the Poisson process

Exponential distribution: moments

$$\varphi_X(t) = \mathbb{E}(e^{tX}) = \frac{\lambda}{\lambda - t}$$

$$\mathbb{E} = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

A non-negative random variable X is said to be without memory or memoryless if

$$P(X > s + t | X > t) = \mathbb{P}(X > s)$$

Another equivalent formulation

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > t) \times \mathbb{P}(X > s)$$

Exponential distribution is memoryless and is the only function that is memoryless.

Hazard rate function

For a continuous random variable X having cdf F_X and density f_X the hazard rate function is defined by

$$r_X(t) = \frac{f_X(t)}{1 - F_X(t)}$$

The Geometric distribution is the only discrete distribution that has the memoryless property.

Properties 1. Let $X_1, X_2 ... X_n$ be i.i.d Exponential random variables with parameter λ. Then

$$\sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda)$$

2. Let X_1 and X_2 be independent exponential random variables with respective rates λ_1 and λ_2. Then

$$\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

This can be generalised to

$$\mathbb{P}(X_i \text{ is the minimum}) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

3. Let $X_1, X_2 ... X_n$ be independent exponential random variables with respective rates $\lambda_1, \lambda_2, ... \lambda_n$. Then the distribution of $X_{(1)} = min_i X_i$ is exponentially distribted with rate equal to $\sum_{i=1}^{n} \lambda_i$, ie

$$X_{(1)} \sim \text{Exp}(\sum_{i=1}^{n} \lambda_i)$$

Sum of Exponential random variables

$$f_{X_1+X_2+...+X_n}(t) = \sum_{i=1}^{n} C_{i,n} \lambda_i e^{-\lambda_i t}$$

Where

$$C_{i,n} = \prod_{i \neq j} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

Geometric sum of exponential variables

$$X = \sum_{i=1}^{N} X_i \sim \text{Exp}(p\lambda)$$

Arrival process

An arrival process is a sequence of increasing random variables $0 < S_1 < S_2 < ...$ is called arrival times and representing the times at which some repeating phenomenon occurs.

Interarrival times

The interarrival times $X_1, X_2, ...$ are positive random variables defined in terms of the arrival times by $X_1 = S_1$ and $X_i = S_i - S_{i-1}$ for $i > 1$. Similarly we can write

$$S_n = \sum_{i=1}^{n} X_i$$

Counting Process

A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of occurances of a certain phenomenon by (and including) time t.

Some properties

1. $N(t) > 0$ for all $t \geq 0$
2. $N(t)$ is an integer valued for all $t \geq 0$
3. $N(0) = 0$
4. For all $t_2 > t_1, N(t_2) \geq N(t_1)$

Increment

The increment of the counting process between time t_1 and time t_2 is the number of occurances of the phenomenon in the interval $(t_1, t_2]$ ie $N(t_2) - N(t_1)$