Lecture Notes on in Stochastic Processes

Book · March 2018

5 authors, including:

Yousif Hussein
Saint Petersburg State University

Hussein Yousif Eledum
University of Tabuk

UNIVERSITY of Tabuk

Faculty of Science

Department of Stat.

جامعة تبوك

كلية العلوم

قسم الاحصاء

محاضرات في العمليات التصادفية

Lecture Notes on Stochastic Processes

STAT 321

Dr.Hussein Yousif Eledum

Associate Professor in Applied Statistics at University of Tabuk, Department of Statistics

2018

Table of Contents

lesson1: Review of Probability

Random variable

A random variable is a real valued function whose numerical value is determined by the outcome of a random experiment. In other words the random variable X is a function that associated each element in the sample space Ω from with the real numbers (i.e. $X : \Omega \rightarrow \mathcal{R}$)

Notation:

X **(**capital letter**):** denotes the random variable .

x **(**small letter**):** denotes a value of the random variable X.

Discrete random variable

A random variable X is called a discrete random variable if its set of possible values is countable (integer).

Continuous random variable

A random variable X is called a continuous random variable if it can take values on a continuous scales.

Discrete probability distribution

If X is a discrete random variable with distinct values $x_1, x_2, \dots \dots .. x_t$, then the function

$$f(x) = \begin{cases} f(X = x_i), & if \quad x = x_1, x_2, \dots \dots .. x_t \\ 0, & Otherwise \end{cases}$$

Is defined to be the probability mass function pmf of X .

This means that a discrete random variable is a listing of all possible distinct (elementary) events and their probabilities of occurring for a random variable.

x_1	x_2	…………	x_t
$f(x_1)$	$f(x_2)$	…………	$f(x_t)$

The pmf $f(x)$ is a real valued function and satisfies the following properties:

3. $0 \leq f(x) \leq 1$ 2. $\sum_{\forall x} f(x) = 1$ 3. $P(X = A) = \sum_{x \epsilon A} f(x)$ where $A \subset x$ is

Continuous probability distribution

The probability density function pdf of a continuous random variable X is a mathematical function $f(x)$ which is defined as follows:

$$f(x) = \begin{cases} f(X = x_i), & if - \infty \leq x \leq \infty \\ 0, & Otherwise \end{cases}$$

The pdf $f(x)$ is satisfies the following properties:

1. $f(x) \geq 0 \; \forall \; x \, \epsilon \, \mathcal{R}$ 2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ 3. $P(X \epsilon A) = \int_A f(x)\, dx$, $A \subset x$

Cumulative distribution function (CDF)

For any random variable we define the cumulative distribution function cdf , $F(x)$ by:

$$f(x) = P(X \leq x)$$

Where, x is any real value.

$$F(x) = \begin{cases} \sum_{u=-\infty}^{x} f(u), & if\ X\ is\ a\ discrete \\ \int_{-\infty}^{x} f(u)du, & if\ X\ is\ a\ continuous \end{cases}$$

$F(x)$ is monotonic increasing i.e.

$F(a) \leq F(b) \quad whenever\ a \leq b$

And the limit of $F(x)$ to the left is 0 and to the right is 1:

$$\lim_{x \to -\infty} F(x) = 0 \qquad and \qquad \lim_{x \to \infty} F(x) = 1$$

For a continuous case :

1. $P(a < X < b) = P(X < b) - P(X < a) = F(b) - F(a)$
2. $f(x) = \dfrac{dF(x)}{dx}$

Mathematical Expectation

Let X be a random variable with a probability distribution $f(x)$ the expected value (mean) of X is denoted by E(X) or μ_x and is defined by:

$$E(X) = \mu_X = \begin{cases} \sum_{all\ x} x\ f(x); & if\ \ X\ \ is\ discrete \\ \int_{-\infty}^{\infty} x\ f(x)dx\ ; & if\ \ X\ \ is\ continuous \end{cases}$$

Linear property:

Let X be a random variable with the pdf *f(x)*, and let *a* and *b* are a constants, then:

$$E(a + bX) = a + bE(X)$$

The variance

Let X be a random variable with a probability distribution $f(x)$ the variance of X is denoted by $Var(X)$ or σ_X^2 and is defined by:

$$Var(X) = \sigma_X^2 = E[(X-\mu)^2] = \begin{cases} \sum_{all\ x} (x-\mu)^2\ f(x); & if\ \ X\ \ is\ discrete \\ \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)dx\ ; & if\ \ X\ \ is\ continuous \end{cases}$$

and it's also can be written as:

$$Var(X) = E(X^2) - \mu_X^2$$

Linear property:

Let X be a random variable with the pdf *f(x)*, and let *a* and *b* are a constants, then:

$$Var(a+bX) = b^2 Var(X)$$

The moments:

Let X be a random variable with the pdf *f(x)*, the r^{th} moment about the origin of X, is given by:

$$\acute{\mu}_r = E(X^r) = \begin{cases} \sum_{all\ x} x^r f(x); & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} x^r\ f(x)\,dx; & if\ X\ is\ continuous \end{cases}$$

if the expectation exists

As special case:

$$\acute{\mu}_1 = EX \quad = \text{mean of } X = \mu$$

Let X be a random variable with the pdf *f(x)*, the r^{th} central moment of X about μ , is defined as:

$$\mu_r = \mathrm{E}(X-\mu)^r = \begin{cases} \sum_{all\ x} (x-\mu)^r f(x); & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} (x-\mu)^r\ f(x)\,dx; & if\ X\ is\ continuous \end{cases}$$

As special case:

$$\mu_2 = \mathrm{E}(X-\mu)^2 = \sigma^2 \text{ the variance of X.}$$

Moment- Generating Function MGF:

Let X be a random variable with the pdf *f(x)*, the moment - generating function of X, is given by $E(e^{tx})$ and is denoted by $M_X(t)$. Hence :

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum_{all\ x} e^{tx} f(x); & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} e^{tx}\ f(x)\,dx; & if\ X\ is\ continuous \end{cases}$$

Moment-generating functions will exist only if the sum or integral of the above definition converges. If a moment-generating function of a random variable X does exist, it can be used to generate all the moments of that variable.

Definition:

Let X be a random variable with moment - generating function of X, is given by $M_X(t)$. then :

$\frac{d^r M_X(t)}{dt^2}\Big|_{t=0} = \acute{\mu}_r$ Therefore, $\frac{d\ M_X(t)}{dt}\Big|_{t=0} = \acute{\mu}_1 = \mu$

$\frac{d^2 M_X(t)}{dt^2}\Big|_{t=0} = \acute{\mu}_2$ $\sigma^2 = \acute{\mu}_2 - \acute{\mu}_1^2$

Example. Find the moment-generating function of the binomial random variable X and then use it to verify that $\mu = np$ and $\sigma^2 = npq.$

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}; & x = 0,1,\ldots\ldots,n \\ 0; & otherwise \end{cases}$$

$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}$

Recognizing this last sum as the binomial expansion of $(pe^t + q)^n$, we obtain:

$$M_X(t) = (pe^t + q)^n$$

now: $\frac{dM_X(t)}{dt} = npe^t(pe^t + q)^{n-1}$

and: $\frac{d^2M_X(t)}{dt^2} = npe^t[pe^t(n-1)(pe^t + q)^{n-2} + (pe^t + q)^{n-1}]$

Setting $t = 0$, we get: $\acute{\mu}_1 = np$ and : $\acute{\mu}_2 = np[p(n-1) + 1]$

Therefore, $\mu = \acute{\mu}_1 = np$

$$\sigma^2 = \acute{\mu}_2 - \acute{\mu}_1^2 = np[p(n-1) + 1] - (np)^2 = npq$$

Probability Generating Function PGF:

Let X be a random variable defined over the non-negative intergers. The probability generating function PGF is given by the polynomial

$$G_X(s) = E(s^X) = p_0 + p_1 s + p2s^2 + \cdots = \sum_{x=0}^{\infty} s^x P(X = x)$$

Example. Let X have a binomial distribution function such that $X \sim B(n,p)$. The PGF is given by

$$G_X(s) = E(s^X) = \sum_{x=0}^{n} \binom{n}{x} (sp)^x q^{n-x} = (q + sp)^n$$

An important property of a PGF is that it converges for $|s| \leq 1$ since

$G_X(1) = \sum_x P(X = x) = 1.$

The PGF can be used to directly derive the probability function of the random variable, as well as its moments. Single probabilities can be calculated as

$$P(X = j) = p_j = (j!)^{-1} \frac{d^j G_X(s)}{ds^j}\Big|_{s = 0}$$

Example: A binomial distributed random variable has PGF $G_X(s) = (q + sp)^n$. Thus,

$P(X = 0) = G_X(0) = q^n$

$P(X = 1) = G_X{}'(0) = nq^{n-1}p^1$

$P(X = 2) = (2!)^{-1} G_X{}''(0) = (2!)^{-1} n(n-1) q^{n-2} p^2$

$\vdots \qquad \qquad \vdots$

The expectation $E(X)$ satisfies the relation

$$E(X) = \sum_{x=0}^{\infty} x\, G_X(s) = G_X{}'(1)$$

Example: A binomial distributed random variable has mean

$$G_X{}'(1) = np(p+q)^{n-1} = np$$

Calculating first

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\, G_X(s) = G_X{}''(1)$$

the variance is obtained as

$$Var(X) = E[X(X-1)] + E(X) - [E(X)]^2 = G_X^{\prime\prime(1)} + G_X^{\prime(1)} - \left[\, G_X^{\prime(1)}\right]^2$$

Example: A binomial distributed random variable has variance

$$Var(X) = n(n-1)p^2 + np + [np]^2 = np(1-p)$$

Joint and marginal probability Distributions

Joint probability distribution (Discrete case)

If X and Y are two discrete random variables, then $f(x,y) = P(X = x, Y = y)$ is called joint probability mass function jpmf of X and Y, and $f(x,y)$ has the following properties:

1. $0 \leq f(x,y) \leq 1$ for all x and y. 2. $\sum_x \sum_y f(x,y) = 1$
3. $P[(X,Y) \in A] = \sum \sum_A f(x,y)$ for any region A in the X, Y plane.

Marginal probability distribution (Discrete case)

If X and Y are jointly discrete random variables with the jpmf $f(x,y)$, then $\mathrm{g}(x)$ and $\mathrm{h}(y)$ are called marginal probability mass functions of X and Y respectively which can be calculated as

- $\mathrm{g}(x) = \sum_{\forall y} f(x,y)$ - $\mathrm{h}(y) = \sum_{\forall x} f(x,y)$

Joint probability distribution (Continuous case)

If X and Y are two continuous random variables, then $f(x,y) = P(X = x, Y = y)$ is called joint probability density function jpdf of X and Y, and $f(x,y)$ has the following properties:

1. $f(x,y) \geq 0$ for all x and y.
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$
3. $P[(X,Y) \in A] = \int_A f(x,y)\, dx\, dy$ for any region A in the X, Y plane.

Marginal probability distribution (Continuous case)

If X and Y are jointly continuous random variables with the j.p.d.f $f(x,y)$, then $\mathrm{g}(x)$ and $\mathrm{h}(y)$ are called marginal probability density function of X and Y respectively which can be calculated as

- $\mathrm{g}(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$
- $\mathrm{h}(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

Conditional Distributions and conditional Expectation

Conditional distribution

If X and Y are jointly random variables discrete or continuous with the jpf $f(x,y)$, $\mathrm{g}(x)$ and $\mathrm{h}(y)$ are marginal probability distributions of X and Y respectively, then the conditional distribution of the random variable Y given that $X = x$ is

$$f(y \mid x) = \frac{f(x,y)}{\mathrm{g}(x)}, \quad \mathrm{g}(x) > 0$$

Similarly the conditional distribution of the random variable X given that $Y = y$ is

$$f(x \mid y) = \frac{f(x,y)}{\mathrm{h}(y)}, \quad \mathrm{h}(y) > 0$$

Statistical independence

If X and Y be two random variables discrete or continuous with the jpf $f(x,y)$, $\mathrm{g}(x)$ and $\mathrm{h}(y)$ are marginal probability distributions of X and Y respectively. The random variables X and Y are said to be statistically independent if and only if:

$$f(x,y) = \mathrm{g}(x)\mathrm{h}(y)$$

for all (x,y) within their ranges.

Conditional Expectation

If X and Y are jointly random variables discrete or continuous with the jpf $f(x,y)$, $\mathrm{g}(x)$ and $\mathrm{h}(y)$ are marginal probability distributions of X and Y respectively, then the conditional expectation of the random variable X given that $Y = y$ for all values of Y such that $\mathrm{h}(y) > 0$ is

$$E(X \mid Y = y) = \begin{cases} \sum_{all\ x} x f(x \mid y); & \text{for discrete} \\ \int_{\mathrm{R}} x f(x \mid y)\,dx; & \text{for continuous} \end{cases}$$

Note that $E(X \mid Y = y)$ is a function of Y.

Covariance

Let X and Y be a random variables with joint probability distribution $f(x,y)$ the covariance of X and Y which denoted by $\mathrm{Cov}(X,Y)$) or σ_{XY} is :

$$E(X-\mu_X)(Y-\mu_Y)=\begin{cases}\sum_{all\,x}\sum_{all\,y}(X-\mu_X)(Y-\mu_Y)\,f(x,y)\,; & \text{for discrete}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_X)(Y-\mu_Y)\,f(x,y)\,dx\,dy; & \text{for continuous}\end{cases}$$

The alternative and preferred formula for σ_{XY} is:

$$E(XY)-\mu_X\mu_Y$$

Linear combination

Let X and Y be a random variables with joint probability distribution $f(x,y)$, a and b are constants, then

$$\mathrm{Var}(aX\pm \mathrm{b}Y)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)\pm 2ab\mathrm{Cov}(X,Y)$$

If X and Y are independent random variables, then

$$\mathrm{Var}(aX\pm \mathrm{b}Y)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)$$

Correlation coefficient

Let X and Y be two random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y respectively. The correlation coefficient of X and Y is

$$\rho_{\mathrm{XY}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

Lesson2: Definition of Stochastic Process

Definition

A stochastic process *(random process)* is a family of random variables, $\{X(t), t \in T\}$ or $\{X_t, t \in T\}$ That is, for each t in the index set T, $X(t)$ is a random variable.

Random process also defined as a random variable which a function of time t, that means , $X(t)$ is a random variable for every time instant t or it's a random variable indexed by time.

We know that a random variable is a function defined on the sample space Ω. Thus a random process $\{X(t), t \in T\}$ is a real function of two arguments $\{X(t, \omega), t \in T,\ \omega \in \Omega\}$. For fixed $t(= t_k)$, $X(t_k, \omega) = X_k(\omega)$ is a random variable denoted by $X(t_k)$, as ω varies over the sample space Ω. On the other hand for fixed sample space $\omega_h \in \Omega$, $X(t, \omega_h) = X_h(t)$ is a single function of time t, called a sample function or a *realization* of the process. The totality of all sample functions is called an *ensemble*.

If both ω and t are fixed, $X(t_k, \omega_h)$ is a real number. We used the notation $X(t)$ to represent$X(t, \omega)$.

Description of a Random Process

In a random process $\{X(t), t \in T\}$ the index t called the *time-parameter* (or simply the time) *and* $T \in \mathbb{R}$ called the parameter set of the random process. Each $X(t)$ takes values in some set $S \in \mathbb{R}$ called the *state space*; then $X(t)$ is the state of the process at time *t*, and if $X(t) = i$ we said the process in state i at time t .

Definition:-

$\{X(t), t \in T\}$ is a discrete - time (discrete parameter) process if the index set T of the random process is discrete . A discrete-parameter process is also called a random sequence and is denoted by $\{X(n), n = 1,2, \dots\}$ or $\{X_n, n = 1,2, \dots\}$.

In practical this generally means T={1,2,3,…….}.

Thus a discrete-time process is $\{X(0), X(1),\ X(2), \dots\}$: a new random number recorded at every time 0, 1, 2, 3, . . .

Definition:-

$\{X(t), t \in T\}$ is continuous - time (continuous parameter) process if the index set T is continuous .

In practical this generally means T = [0, ∞), or T = [0,K] for some K.

Thus a continuous-time process $\{X(t), t \in T\}$ has a random number $X(t)$ recorded at every instant in time.

(Note that $X(t)$ needs not change at every instant in time, but it is allowed to change at any time; i.e. not just at t = 0, 1, 2, . . . , like a discrete-time process.)

Definition:-

The state space, S: is the set of real values that $X(t)$ can take.

Every $X(t)$ takes a value in $\mathbb{R}$, but S will often be a smaller set: $S \subset \mathbb{R}$. For example, if $X(t)$ is the outcome of a coin tossed at time *t*, then the state space is S = {0, 1}.

Definition:-

The state space S is called a discrete-state process if it is discrete , often referred to as a *chain*. In this case, the state space S is often assumed to be $\{0,1,2,\ldots\}$If the state space S is continuous then we have a continuous-state process.

Examples:

Discrete-time, discrete-state processes

Example 1: Tossing a balanced die more than once, if we interest on the number on the uppermost face at toss n , say $X(1)$ the number appears on the first toss , $X(2)$ number appears in the second one ,…………….. ect, then $\{X(n), n \in T\}$ is the random process, and the random variable $X(n)$ denotes the number appears at toss n . where n is the parameter. $T = \{1,2,3,\ldots\}$ and $S = \{1,2,3,4,5,6\}$.

*Example*2:The number of emails in your inbox at time t .$T = \{1,2,3,\ldots\}$ and $S = \{0,1,2,\ldots\}$.

Example 3: your bank balance on day *t* .

Example 4: the number of occupied channels in a telephone link at the arrival time of the n^{th} customer, n = 1,2,...

Continuous-time, discrete-state processes

Example 6: The number of occupied channels in a telephone link at time $t > 0$

Example 7: The number of packets in the buffer of a statistical multiplexer at time $t > 0$

Lesson 3: Characterization of Stochastic Process

Distribution function CDF and Probability distribution PDF for (t) :

Consider the stochastic process $\{X(t), t \in T\}$, for any $t_0 \in T$, $X(t_0) = X$ is a random variable, and it's a CDF $F_{X(t_0)}(x)$ or $F_X(x;t_0)$ is defined as:

$$F_X(x;t_0) \quad = P(X(t_0) \le x)$$

$F_X(x;t_0)$is known as a *first - order distribution function* of the random process $X(t)$.

Similarly, Given t_1 and t_2 , $X(t_1) = X_1$ and $X(t_2) = X_2$ represent two random variables their joint CDF $F_{X(t_1)X(t_2)}(x_1, x_2)$ or $F_X(x_1, x_2; t_1, t_2)$ is given by

$$F_X(x_1, x_2; t_1, t_2) \quad = P(X(t_1) \le x_1, X(t_2) \le x_2)$$

$F_X(x_1, x_2; t_1, t_2)$ is known as the *second - order distribution* of $X(t)$.

In general we define the *nth-order distribution function* of $X(t)$ by

$$F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \quad = P(X(t_1) \le x_1, X(t_2) \le x_2, \dots, X(t_n) \le x_n)$$

Similarly, we can write joint PDFs or PMFs depending on whether $X(t)$ is continuous-valued (the $X(t_i)$'s are continuous random variables) or discrete-valued (the $X(t_i)$'s are discrete random variables). For example the *second – order* PDF and PMF given respectively by

$$f_X(x_1, x_2; t_1, t_2) \;= P(X(t_1) = x_1, X(t_2) = x_2) = \frac{\partial^2 F_{X(t_1)X(t_2)}(x_1, x_2)}{\partial x_1 \partial x_2}$$

$$P_X(x_1, x_2; t_1, t_2) \;= P(X(t_1) = x_1, X(t_2) = x_2)$$

Mean and Variance functions of random process:

As in the case of r.v.'s, random processes are often described by using statistical averages.

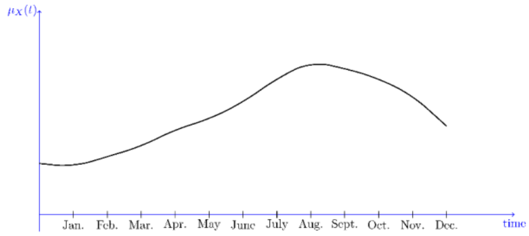
For the random process $\{X(t), t \in T\}$, the *mean function* $\mu_X(t): T \to \mathbb{R}$ is defined as

$$\mu_X(t) = E[X(t)] = \int x f_X(t)\, dx$$

The above definition is valid for both continuous-time and discrete-time random processes. In particular, if $\{X(n), n \in T\}$ is a discrete-time random process, then

$$\mu_X(n) = E[X(n)] \quad \forall \; n \in \mathbb{R}$$

The mean function gives us an idea about how the random process behaves on average as time evolves (a function of time). For example, if $X(t)$ is the temperature in a certain city, the mean function $\mu_X(t)$ might look like the function shown in Figure below. As we see, the expected value of $X(t)$ is lowest in the winter and highest in summer.



The variance of a random process $X(t)$, also a function of time, given by:

$$\sigma_X^2(t) = Var[X(t)] = E[X(t) - \mu_X(t)]^2 = E[X_t]^2 - [\mu_X(t)]^2$$

Autocorrelation, and Covariance Functions:

The mean function $\mu_X(t)$ gives us the expected value of $X(t)$ at time t, but it does not give us any information about how $X(t_1)$ and $X(t_2)$ are related. To get some insight on the relation between $X(t_1)$ and $X(t_2)$, we define correlation and covariance functions.

Given two random variables $X(t_1)$, $X(t_2)$ the *autocorrelation function* or simply *correlation function* $R_{XX}(t_1, t_2)$, defined by:

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1\, x_2\, f_X(x_1, x_2; t_1, t_2)\, dx_1 dx_2$$

Where $f_X(x_1, x_2; t_1, t_2)$ is a joint probability function for t_1 and t_2.

For a random process, t_1 and t_2 go through all possible values, and therefore, $E[X(t_1)X(t_2)]$ can change and is a function of t_1 and t_2.

Note that:

$$R_{XX}(t_1, t_2) = R_{XX}(t_2, t_1)$$

The *autocovariance function* of $X(t)$ is defined by:

$$C_{XX}(t_1, t_2) = \mathrm{Cov}[X(t_1), X(t_2)] = E\big[\big(X(t_1) - \mu_X(t_1)\big)\big(X(t_2) - \mu_X(t_2)\big)\big]$$
$$= R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \quad \text{for} \quad \mathrm{t}_1, \mathrm{t}_2 \in \mathrm{T}$$

It is clear that if the mean of $X(t)$ is zero, then $C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2)$.

If $t_1 = t_2 = t$ we obtain

$$R_{XX}(t_1, t_2) = R_{XX}(t, t) = E[X(t)X(t)] = E[X(t)]^2$$

$$C_{XX}(t_1, t_2) = C_{XX}(t, t) = \mathrm{Cov}[X(t), X(t)]$$
$$= Var\big(X(t)\big) \quad \text{for } \mathrm{t} \in \mathrm{T}$$

The normalized autocovariance function is defined by:

$$\rho(t_1, t_2) = \frac{C_{XX}(t_1, t_2)}{\sqrt{C_{XX}(t_1, t_1) C_{XX}(t_2, t_2)}}$$

Example: wireless signal model

Consider RP $X(t) = \alpha\cos(2\pi F_c(t) + \theta)$

Where α : amplitude (capacity) $F_c(t)$: carrier frequency θ: phase

$\theta \sim U(-\pi, \pi)$ that is $f_\theta(\theta) = \frac{1}{2\pi}$ if $-\pi \le \theta \le \pi$

α, F_c: are constant $F_c(t)$: function of a time

Find

i. Mean function of $X(t)$
ii. Autocorrelation function of $X(t)$

Solution

i. Mean function of $\boldsymbol{X(t)}$

$$\mu_X(t) = E[X(t)] = E\{\alpha\cos(2\pi F_c(t) + \theta)\} = \int \alpha\cos(2\pi F_c(t) + \theta)\, f_\theta(\theta) d\theta$$

$$= \int_{-\pi}^{\pi} \alpha\cos(2\pi F_c(t) + \theta)\ 1/2\pi\ d\theta = \frac{a}{2\pi}\int_{-\pi}^{\pi} \cos(2\pi F_c(t) + \theta)\ \ d\theta$$

$$= \frac{a}{2\pi}\ \sin(2\pi F_c(t) + \theta)\Big|_{-\pi}^{\pi} = \frac{a}{2\pi}\ \{\sin(2\pi F_c(t) + \pi) - \sin(2\pi F_c(t) - \pi)\}$$

$$= \frac{a}{2\pi} \times \{0\} = 0$$

$$\Longrightarrow \mu_X(t) = 0$$

ii. Autocorrelation function of $X(t)$

$$R_{XX}(t_1, t_2) = E\{X(t_1)X(t_2)\}$$

Let $t_1 = t$ $t_2 = t + \tau$ and $\tau \in \mathbb{R}$ $\tau = t_2 - t_1$ time shift

$$R_{XX}(t, t+\tau) = E\{[\alpha\cos(2\pi F_c(t) + \theta)][\alpha\cos(2\pi F_c(t+\tau) + \theta)]\}$$

$$= \alpha^2 E\{[\cos(2\pi F_c(t) + \theta)][\cos(2\pi F_c(t+\tau) + \theta)]\}$$

$$\boxed{\cos(\alpha)\cos(\beta) = \tfrac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)]}$$

Let $\alpha = 2\pi F_c(t) + \theta$ $\beta = 2\pi F_c(t+\tau) + \theta$ then

$\alpha + \beta = 2\pi F_c(2t+\tau) + 2\theta$ $\alpha - \beta = 2\pi F_c(\tau)$

$$R_{XX}(t, t+\tau) = \frac{a^2}{2} E\{\cos(2\pi F_c(2t+\tau) + 2\theta) + \cos\big(2\pi F_c(\tau)\big)\}$$

$$= \frac{a^2}{2}\Big(E\{\cos(2\pi F_c(2t+\tau) + 2\theta)\} + E\big\{\cos\big(2\pi F_c(\tau)\big)\big\}\Big)$$

The first term is 0 , and $E\{\cos(2\pi F_c(\tau))\} = \cos(2\pi F_c(\tau))$ is the constant (no θ)

$$R_{XX}(t, t+\tau) = \frac{a^2}{2}\cos\big(2\pi F_c(\tau)\big)$$

Example:

A random process $\{X(t), t \in T\}$ with $\mu_X(t) = 5$ and $R_{XX}(t_1, t_2) = 25 + 3e^{-0.6|t_1 - t_2|}$ Determine the mean, the variance and the covariance of the random variables $U = X(6)$ and $V = X(9)$.

Solution:

$E(U) = E[X(6)] = \mu_X(6) = 5,\ E(V) = E[X(9)] = \mu_X(9) = 5$

$Var(U) = E\{[X(6)]^2\} - \{E[X(6)]\}^2$

since $R_{XX}(t_1, t_1) = E\{[X(t_1)]^2\}$

$$Var(U) = R_{XX}(t_1, t_1) - \{\mu_X(6)\}^2 = R_{XX}(6,6) - 25$$

$$= 25 + 3e^{-0.6|6-6|} - 25 = 28 - 25 = 3$$

Similarly,

$$Var(V) = R_{XX}(t_1, t_1) - \{\mu_X(9)\}^2 = R_{XX}(9,9) - 25$$

$$= 25 + 3e^{-0.6|9-9|} - 25 = 28 - 25 = 3$$

$$\mathrm{Cov}[X(t_1), X(t_2)] = C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

$\mathrm{Cov}(U, V) = C_{XX}(6,9) = R_{XX}(6,9) - \mu_X(6)\mu_X(9)$

Since, $R_{XX}(6,9) = 25 + 3e^{-0.6|6-9|} = 25 + 3e^{-1.8} = 25.496$

$$\mathrm{Cov}(U, V) = 25.496 - 25 = 0.496$$

Lesson 4: Classification of Stochastic Processes

We can classify random processes based on many different criteria.

Stationary and Wide-Sense Stationary Random Processes

A. *Stationary Processes:*

A random process $\{X(t), t \in T\}$ is *stationary* or *strict-sense stationary* SSS if its statistical properties do not change by time. For example, for stationary process, $X(t)$ and $X(t+\Delta)$ have the same probability distributions. In particular, we have

$$F_X(x,t) = F_X(x; t+\Delta) \qquad \forall \quad t, t+\Delta \in T$$

More generally, for stationary process a random $\{X(t), t \in T\}$, the joint distributions of the two random variables $X(t_1)$, $X(t_2)$ is the same as the joint distribution of $X(t_1+\Delta)$, $X(t_2+\Delta)$, for example, if you have stationary process $X(t)$, then

$$P\left[\left(X(t_1), X(t_2)\right) \in A\right] = P\left[\left(X(t_1+\Delta), X(t_2+\Delta)\right) \in A\right]$$

For any set of $A \in \mathbb{R}^2$.

In short, a random process is stationary if a time shift does not change its statistical properties.

<u>*Definition*</u>. A discrete-time random process $\{X(n), n \in \mathbb{Z}\}$ is strict-sense stationary or simply stationary if, for all $n_1, n_2, \dots, n_r \in \mathbb{N}$ and all $D \in \mathbb{Z}$, the joint CDF of

$$X(n_1), X(n_2), \dots\dots, X(n_r)$$

Is the same CDF as

$$X(n_1+D), X(n_2+D), \dots, X(n_r+D)$$

That is, for real numbers $x_1, x_2, \dots, x_r$ we have

$$F_X(x_1, x_2, \dots, x_r; t_1, t_2, \dots, t_r) = F_X(x_1, x_2, \dots, x_r; t_1+D, t_2+D, \dots, t_r+D)$$

This can be written as

$$F_{X(t_1),X(t_2),\dots,X(t_r)}(x_1, x_2, \dots, x_r) = F_{X(t_1+D),X(t_2+D),\dots,X(t_r+D)}(x_1, x_2, \dots, x_r)$$

<u>*Definition.*</u> A continuous-time random process $\{X(t), t \in \mathbb{R}\}$ is strict-sense stationary or simply stationary if, for all $t_1, t_2, \dots, t_r \in \mathbb{R}$ and all $\Delta \in \mathbb{R}$, the joint CDF of

$$X(t_1), X(t_2), \dots\dots, X(t_r)$$

Is the same CDF as

$$X(t_1+\Delta), X(t_2+\Delta), \dots, X(t_r+\Delta)$$

That is, for real numbers $x_1, x_2, \dots, x_r$ we have

$F_X(x_1, x_2, \dots, x_r; t_1, t_2, \dots, t_r) = F_X(x_1, x_2, \dots, x_r; t_1 + \Delta, t_2 + \Delta, \dots, t_r + \Delta)$

This can be written as

$$F_{X(t_1),X(t_2),\dots,X(t_r)}(x_1, x_2, \dots, x_r) = F_{X(t_1+\Delta),X(t_2+\Delta),\dots,X(t_r+\Delta)}(x_1, x_2, \dots, x_r)$$

B. *Wide-Sense Stationary Processes :*

A random process is called *weak-sense stationary* or *wide-sense stationary* (WSS) if its mean function and its autocorrelation function do not change by shifts in time. More precisely, $X(t)$ is WSS if, for all $t_1, t_2 \in \mathbb{R}$,

1. $E[X(t_1)] = E[X(t_2)] = \mu_X$ constant (stationary mean in time)

For $t_1 = t$ $t_2 = t + \tau$ and $\tau \in \mathbb{R}$ $\tau = t_2 - t_1$ time shift

2. $R_{XX}(t, t+\tau) = E[X(t)X(t+\tau)] = R_{XX}(\tau)$

Note that the first condition states that the mean function $\mu_X(t)$ is not a function of time t, thus we can write $\mu_X(t) = \mu_X$. The second condition states that the correlation function $R_{XX}(t, t+\tau)$ is only a function of time shift τ and not on specific times t_1, t_2.

Definition

A continuous-time random process $\{X(t), t \in \mathbb{R}\}$ *weak-sense stationary* or *wide-sense stationary* (WSS) if

1. $\mu_X(t) = \mu_X \quad \forall\ t \in \mathbb{R}$
2. $R_{XX}(t, t+\tau) = R_{XX}(t_1 - t_2) = R_{XX}(\tau) \quad \forall\ t_1, t_2 \in \mathbb{R}$

Definition

A discrete-time random process $\{X(n), n \in \mathbb{Z}\}$ *weak-sense stationary* or *wide-sense stationary* (WSS) if

1. $\mu_X(n) = \mu_X \quad \forall\ n \in \mathbb{Z}$
2. $R_{XX}(n_1, n_2) = R_X(n_1 - n_2) \quad \forall\ n_1, n_2 \in \mathbb{Z}$

Note that a strict-sense stationary process is also a *WSS* process, but in general, the converse is not true.

Example: wireless signal model

Consider RP $X(t) = \alpha \cos(2\pi F_c(t) + \theta)$

Where α : amplitude (capacity) $F_c(t)$: carrier frequency θ: phase

$\theta \sim U(-\pi, \pi)$ that is $f_\theta(\theta) = \frac{1}{2\pi}$ if $-\pi \le \theta \le \pi$

α, F_c: are constant $F_c(t)$: function of a time

Show that $X(t)$ is WSS.

Solution

The Mean function of $\boldsymbol{X(t)}$ is $\mu_X(t) = 0$ constant

The autocorrelation function $R_{XX}(t, t+\tau) = \frac{a^2}{2}\cos\big(2\pi F_c(\tau)\big)$ function of τ

Since $\mu_X(t)$ is a constant doesn't depend on time and the $R_{XX}(t, t+\tau)$ depends only on time shift τ, therefore, $X(t)$ is WSS.

Example:

Consider RP $X(t) = A\sin(\omega_c(t) + \theta)$ A is a r.v. with mean μ_A and variance σ_A^2, $\theta \sim U(-\pi, \pi)$. A and θ are independent. Find

i. Mean function of $X(t)$
ii. Autocorrelation function of $X(t)$ and
iii. Show that $X(t)$ is WSS

Solution:

i. Mean function:

$$\mu_X(t) = E[X(t)] = E\{A\sin(\omega_c(t)+\theta)\} = E\{A\}\,E\{\sin(\omega_c(t)+\theta)\} \quad A,\theta \text{ Independent}$$

$$= \mu_A \int \sin(\omega_c(t)+\theta)\, f(\theta) d\theta = \mu_A \int_{-\pi}^{\pi} \sin(\omega_c(t)+\theta)\ 1/2\pi\ d\theta$$

$$= -\frac{\mu_A}{2\pi}\cos(\omega_c(t)+\theta)\Big|_{-\pi}^{\pi}$$

$$= -\frac{\mu_A}{2\pi}[\cos(\omega_c(t)+\pi) - \cos(\omega_c(t)-\pi)]$$

$$\cos(u) - \cos(v) = -2\sin\left(\frac{u+v}{2}\right)\sin\left(\frac{u-v}{2}\right)$$

Let $u = \omega_c(t) + \pi$ $\quad v = u = \omega_c(t) - \pi$ $\quad u + v = 2\omega_c(t)$ $\quad u - v = 2\pi$

$$\cos(\omega_c(t)+\pi) - \cos(\omega_c(t)-\pi) = -2\sin\big(\omega_c(t)\big)\sin(\pi)$$

$$-\frac{\mu_A}{2\pi}[\cos(\omega_c(t)+\pi) - \cos(\omega_c(t)-\pi)] = \frac{2\mu_A}{2\pi}\sin\big(\omega_c(t)\big)\sin(\pi)$$

Since $\sin(\pi) = 0$ therefore

$$\frac{\mu_A}{\pi}\sin\big(\omega_c(t)\big)\sin(\pi) = 0$$

$$\mu_X(t) = 0$$

ii. Correlation function:

$$R_{XX}(t_1, t_2) = E\{X(t_1)X(t_2)\} = E\{A^2\sin(\omega_c(t_1)+\theta)\sin(\omega_c(t_2)+\theta)\}$$

Because A, θ Independent

$$E\{A^2\sin(\omega_c(t_1)+\theta)\sin(\omega_c(t_2)+\theta)\} = E\{A^2\}E\{\sin(\omega_c(t_1)+\theta)\sin(\omega_c(t_2)+\theta)\}$$

$$sin(\alpha)\,sin(\beta) = \frac{1}{2}[cos(\alpha-\beta) - cos(\alpha+\beta)]$$

Let $\quad \alpha = \omega_c(t_1) + \theta \qquad \beta = \omega_c(t_2) + \theta \quad$ then

$$\alpha - \beta = \omega_c(t_1 - t_2) \qquad \alpha + \beta = \omega_c(t_1 + t_2) + 2\theta$$

Therefore

$$E\{A^2\}E\{\sin(\omega_c(t_1) + \theta)\sin(\omega_c(t_2) + \theta)\}$$

$$= E\{A^2\}\,E\left[\frac{1}{2}\cos\big(\omega_c(t_1 - t_2)\big) - \frac{1}{2}\cos(\omega_c(t_1 + t_2) + 2\theta)\right]$$

$$= E\{A^2\}\,\left[\frac{1}{2}E\big\{\cos\big(\omega_c(t_1 - t_2)\big)\big\} - \frac{1}{2}E\{\cos(\omega_c(t_1 + t_2) + 2\theta)\}\right]$$

The second term is zero.

$$R_{XX}(t_1, t_2) = \frac{1}{2}\,E\{A^2\}\cos\big(\omega_c(t_1 - t_2)\big) = \frac{\mu_A}{2}\,\cos\big(\omega_c(\tau)\big)$$

$X(t)$ is WSS random process because the mean function is a constant (=0) and the autocorrelation function is only a function of a time difference $t_1 - t_2$.

Independent and independent identically distributed iid Random Processes

A. *Independent Processes:*

In a random process $X(t)$, if $X(t_i)$ for $i = 1,2, \dots, n$ are independent r.v.'s, so that for $n = 1,2, \dots$,

$$f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \prod_{i=1}^{n} f_X(x_i; t_i)$$

and $\qquad F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \prod_{i=1}^{n} F_X(x_i; t_i)$

Or $\qquad P(X(t_1) \le x_1, X(t_2) \le x_2, \dots, X(t_n) \le x_n)$

$$= P(X(t_1) \le x_1).\,P(X(t_2) \le x_2), \dots P(X(t_n) \le x_n)$$

then we call $X(t)$ an independent random process. Thus, a first-order distribution is sufficient to characterize an independent random process $X(t)$.

B. *Independent and identically distributed iid random process*

A Random process $\{X(t), t \in T\}$ is said to be independent and identically distributed (iid) if any finite number, say *k*, of random variables $X(t_1)$, $X(t_2)$, ……, $X(t_k)$ are mutually independent and have a common cumulative distribution function $F_X(.)$. The joint cdf and pdf for $X(t_1)$, $X(t_2)$, ……, $X(t_k)$ are given respectively by:

$$F_X(x_1, x_2, \dots, x_k; t_1, t_2, \dots, t_k) = \prod_{i=1}^{k} F_X(x_i; t_i)$$

$$f_X(x_1, x_2, \dots, x_k; t_1, t_2, \dots, t_k) = \prod_{i=1}^{k} f_X(x_i; t_i)$$

Example.

Consider the random process $\{X_n,\ n = 0,1,2, \dots\}$ in which $X_i's$ are iid standard normal random variables.

(a) Write down $f_{x_n}(x)$ for $n = 0,1,2, \dots$

(b) Write down $f_{x_n,x_m}(x_1, x_2)$ for $m \neq n$

Solution.

(a) Since $X_n \sim N(0,1)$, we have

$$f_{x_n}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2} \qquad \forall\ x \in \mathbb{R}$$

(b) If $m \neq n$, then x_n and x_m are independent (because of the i.i.d. assumption) so,

$$\begin{aligned} f_{x_n,x_m}(x_1, x_2) &= f_{x_n}(x_1) f_{x_m}(x_2) \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x_1^2} \ \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x_2^2} \\ &= \frac{1}{2\pi} e^{\frac{1}{2}(x_1^2 + x_2^2)} \qquad \forall\ x_1, x_2 \in \mathbb{R} \end{aligned}$$

Solved Problems (1)

Problem 1

Let $Y_1, Y_2, ...$ be a sequence of iid random variables with mean $E[Y_i] = 0$ and $\text{var}[Y_i] = 4$. Define the discrete time random process $\{X_n,\ n \in N\}$ as

$$X_n = Y_1 + \ Y_2, \ldots . + Y_n \qquad\qquad \forall\, n \in N$$

(a) Find the mean function X_n

(b) Find the autocorrelation function and covariance function for X_n

Problem 2

Consider the random process $X_n = 1000(1 + R)^n$, for $n = 0,1,2, \ldots$ $R \sim U(0.04, 0.05)$.

(a) Find the mean function X_n

(b) Find the autocorrelation function and covariance function for X_n

Problem 3

Consider the random process $\{X(t), t \in \mathbb{R}\}$ defined as

$$X(t) = Cos(t + U)\}$$

where $U \sim Uniform(0,2\pi)$. Show that $X(t)$ is a WSS process.

Problem 4

Given a random process $\{X(t), t \in T\}$ with $\mu_X(t) = 4$ and $R_X(t_1, t_2) = 20 + 2e^{-0.4|t_1 - t_2|}$. Suppose $Y_1 = 2X(3)$ and $Y_2 = X(6)$. Find:

(a) $\text{E}(Y_1)$.

(b) $\text{Var}(Y_2)$.

(c) $\text{Cov}(Y_1, Y_2)$

Solutions (1)

Problem1 (Solution)

(a)
$$\mu_n = E[X_n]$$
$$= E[Y_1 + Y_2, \dots. + Y_n]$$
$$= E[Y_1] + E[Y_2] + \cdots + E[Y_n]$$
$$= 0$$

(b) Let $m \le n$

$$R_{XX}(m,n) = E[X_m X_n]$$
$$= E[(Y_1 + Y_2, \dots. + Y_m)(Y_1 + Y_2, \dots. + Y_n)]$$
$$= E[Y_1^2] + E[Y_2^2] + \cdots + E[Y_m^2]$$

since $E\left[Y_i Y_j\right] = E[Y_i]E\left[Y_j\right] = 0$ then $E[Y_1^2] = \mathrm{var}[Y]$

$R_{XX}(m,n) = \mathrm{var}[Y_1] + \mathrm{var}[Y_2] + \cdots + \mathrm{var}[Y_m]$

$= 4 + 4 + \cdots + 4 = 4m$

Similarly for $m \ge n$

$$R_{XX}(m,n) = E[X_m X_n]$$
$$= E[(Y_1 + Y_2, \dots. + Y_m)(Y_1 + Y_2, \dots. + Y_n)]$$
$$= E[Y_1^2] + E[Y_2^2] + \cdots + E[Y_n^2]$$
$$= \mathrm{var}[Y_1] + \mathrm{var}[Y_2] + \cdots + \mathrm{var}[Y_n]$$
$$= 4\mathrm{n}$$

Problem2 (Solution)

(a) Let $Y = 1 + R$ so, $Y \sim U(1.04, 1.05)$.

$$\mu_n = E[X_n]$$
$$= 1000\, E[Y^n]$$
$$= 1000 \int_{1.04}^{1.05} y^n \frac{1}{0.01}\, dy$$
$$= 100000 \int_{1.04}^{1.05} y^n\, dy$$
$$= \frac{10^5}{n+1}[y^{n+1}]_{1.04}^{1.05}$$
$$= \frac{10^5}{n+1}[(1.05)^{n+1} - (1.04)^{n+1}] \qquad \forall\, n \in \{0,1,2,\dots\}$$

(b) $R_{XX}(m,n) = E[X_m X_n]$

$$= 10^6\, E[Y^m Y^n]$$

$$= 10^6\, E[Y^{m+n}]$$

$$= 10^8 \int_{1.04}^{1.05} y^{m+n}\, dy$$

$$= \frac{10^8}{n+m+1}\left[y^{m+n+1}\right]_{1.04}^{1.05}$$

$$= \frac{10^8}{n+m+1}\left[(1.05)^{m+n+1} - (1.04)^{m+n+1}\right] \qquad \forall \;\; m,n \in \{0,1,2,\dots\}$$

To find covariance function

$$C_{XX}(m,n) = R_{XX}(m,n) - E[X_m]E[X_n]$$

$$= \frac{10^8}{n+m+1}\left[(1.05)^{m+n+1} - (1.04)^{m+n+1}\right]$$

$$-\frac{10^{10}}{(m+1)(n+1)}\left[(1.05)^{m+1} - (1.04)^{m+1}\right]\left[(1.05)^{n+1} - (1.04)^{n+1}\right]$$

Problem 3 (Solution)

We need to check two conditions

1. $\mu_X(t) = \mu_X \qquad \forall \;\; t \in \mathbb{R}$ and
2. $R_{XX}(t_1,t_2) = R_{XX}(t_1 - t_2) \quad \forall \;\; t_1, t_2 \in \mathbb{R}$

We have

$$\mu_n = E[X_n]$$

$$= E[\cos(t+U)]$$

$$= \int_0^{2\pi} \cos(t+u)\frac{1}{2\pi}\, du$$

$$= \frac{1}{2\pi}\int_0^{2\pi} \cos(t+u)\; du$$

$$= 0 \qquad \forall\, t \in \mathbb{R}$$

we can also find $R_{XX}(t_1,t_2)$

$$R_{XX}(t_1,t_2) = E[X_1 X_2]$$

$$= \; E[\cos(t_1+U)\cos(t_2+U)]$$

$$\cos(\alpha)\cos(\beta) = \tfrac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)]$$

Let $\qquad \alpha = t_1 + U \qquad\qquad \beta = t_2 + U \qquad$ then

$$\alpha+\beta = t_1 + t_2 + 2U \qquad\qquad \alpha - \beta = t_1 - t_2$$

$$R_{XX}(t_1,t_2) = \frac{1}{2}E[\cos(t_1+t_2+2U)+\cos(t_1-t_2)]$$
$$= \frac{1}{2}\cos(t_1-t_2)+\frac{1}{2}E[\cos(t_1+t_2+2U)]$$
$$= \frac{1}{2}\cos(t_1-t_2)+0$$
$$= \frac{1}{2}\cos(t_1-t_2) \qquad \forall\, t_1,t_2 \in \mathbb{R}$$

As we see, both conditions are satisfied, thus $X(t)$ is WSS.

Problem 4 (Solution)

(a) $\mathrm{E}(Y_1) = 8$.

(b) $\mathrm{Var}(Y_2) = E\{[X(6)]^2\} - \{E[X(6)]\}^2$

Since $R_X(t_1,t_1) = E\{[X(t_1)]^2\}$

$$Var(Y_2) = R_X(t_1,t_1) - \{\mu_X(6)\}^2$$
$$= R_X(6,6) - 16$$
$$= 20 + 2e^{-0.4|6-6|} - 16$$
$$= 22 - 16 = 6$$

(c) $\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}[X(t_1),X(t_2)]$

$$= C_{XX}(t_1,t_2) = R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X(t_2)$$

We know that $\mathrm{Cov}(aX_1,bX_2) = ab\mathrm{Cov}(X_1,X_2)$. By using this,

$\mathrm{Cov}(Y_1,Y_2) = \mathrm{Cov}\big(2X(3),X(6)\big) = 2\mathrm{Cov}\big(X(3),X(6)\big)$. Therefore,

$\mathrm{Cov}(Y_1,Y_2) = 2C_X(3,6) = 2(R_X(3,6) - \mu_X(3)\mu_X(6))$

Since, $R_X(3,6) = 20 + 2e^{-0.4|3-6|} = 20 + 2e^{-1.2} = 20.604$ and $\mu_X(3) = \mu_X(6) = 4$

Therefore, $\mathrm{Cov}(Y_1,Y_2) = 2(20.604 \quad - 16) = 9.208$

Markov Chains

Lesson 5: Discrete -Time Markov Chains

Basic Definitions

Let $\{X_n, n \geq 0\}$ be a stochastic process taking values in a state space S that has N states. To understand the behavior of this process we will need to calculate probabilities like:

$$P\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots., X_n = i_n\} \ldots\ldots\ldots\ldots..(1)$$

This can be computed by multiplying conditional probabilities as follows:

$$P\{X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0)P(X_2 = i_2 \mid X_1 = i_1, X_0 = i_0)$$
$$\times \ldots . \times P(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots\ldots.., X_1 = i_1, X_0 = i_0)$$

Example A .

We randomly select playing cards from an ordinary deck.

The state space is $S = \{Red, Black\}$. Let's calculate the chance of observing the sequence RRB using two different sampling methods.

(a) Without replacement:

$$P\{X_0 = R, X_1 = R, X_2 = B\}$$
$$=\{P\{X_0 = R)P(X_1 = R \mid X_0 = R)P(X_2 = B \mid X_1 = R, X_0 = R)$$
$$= \frac{26}{52} \times \frac{25}{51} \times \frac{26}{50} = 0.12745$$

(b) With replacement:

$$P\{X_0 = R, X_1 = R, X_2 = B\}$$
$$=\{P\{X_0 = R)P(X_1 = R \mid X_0 = R)P(X_2 = B \mid X_1 = R, X_0 = R)$$
$$= \frac{26}{52} \times \frac{26}{52} \times \frac{26}{52} = 0.125$$

Definition.

The process $\{X_n, n \geq 0\}$ is called a Markov chain if for any n and any collection of states $i_0, i_1, \ldots, i_{n+1}$ we have:

$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

For a Markov chain, the future depends only on the current state and not on history.

Exercise.

In example A, calculate $P(X_2 = B \mid X_1 = R)$ and confirm that only "with replacement" do we get a Markov chain.

Discrete-Time Markov chains

A discrete time Markov chain $\{X_n, n \geq 0\}$ with discrete state space $S = \{0,1,2, \dots \dots\}$ where this set may be finite or infinite. If $X_n = i$ then the Markov chain is said to be in state i at time n (or the n^{th} step). A discrete-time Markov chain $\{X_n, n \geq 0\}$ is characterized by:

$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

where $P(X_{n+1} = j \mid X_n = i)$ are known as one-step transition probabilities. If $P(X_{n+1} = j \mid X_n = i)$ is independent of n, then the Markov chain is said to possess stationary transition probabilities and the process is referred to as a homogeneous Markov chain. Otherwise the process is known as a nonhomogeneous Markov chain. Note that the concepts of a Markov chain's having stationary transition probabilities and being a stationary random process should not be confused. The Markov process, in general, is not stationary. we will assume that all our Markov chains are time homogeneous.

Definition.

A Markov chain $\{X_n, n \geq 0\}$ is called time homogeneous if, for any $i, j \in S$; we have:

$$P(X_{n+1} = j \mid X_n = i) = p_{ij}$$

For some function $p: S \times S \to [0,1]$.

Transition probability matrix

Often transition probabilities listed in matrix. The matrix called the *State transition matrix* or *transition probability matrix* and is usually shown by P.

Let $\{X_n, n \geq 0\}$ be a homogeneous Markov chain with a discrete finite state space S={0,1,2,……,m} then,

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) \quad i \geq 0, j \geq 0$$

Regardless of the value of *n*. A transition probability matrix of $\{X_n, n \geq 0\}$ is defined by:

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \qquad \text{where,}$$

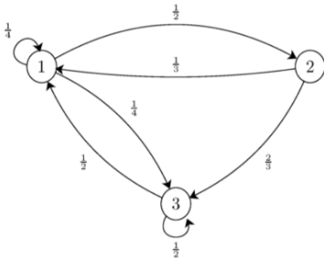
$$p_{ij} \geq 0 \qquad \sum_{j=1}^{m} p_{ij} = 1 \qquad i = 1,2, \dots, m$$

State Transition Diagram

A Markov Chain is usually shown by *a state transition diagram*. Consider a Markov Chain with three possible states 1,2 and 3, and the following transition probabilities.

$$P = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/3 & 0 & 2/3 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

Figure below shows the state transition diagram for this Markov Chain. In this diagram there are three possible states 1,2 and 3, and the arrows from each state to other states show the transition probabilities p_{ij}.When there is no arrow from stat i to state j, it means that $p_{ij} = 0$



Example

Consider the Markov chain shown in Figure above. Find

(a) $P(X_4 = 3 \mid X_3 = 2)$

(b) $P(X_3 = 1 \mid X_2 = 1)$

(c) If we know $P(X_0 = 1) = 1/3$, find $P(X_0 = 1, X_1 = 2)$

(d) If we know $P(X_0 = 1) = 1/3$, find $P(X_0 = 1, X_1 = 2, X_2 = 3)$

Solution

(a) By definition $P(X_4 = 3 \mid X_3 = 2) = p_{23} = \frac{2}{3}$

(b) By definition $P(X_3 = 1 \mid X_2 = 1) = p_{11} = \frac{1}{4}$

(c) $P(X_0 = 1, X_1 = 2) = P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1)$

$$= (1/3)p_{12} = (1/3)(1/2) = \frac{1}{6}$$

(d) $P(X_0 = 1, X_1 = 2, X_2 = 3)$

$= P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1)P(X_2 = 3 \mid X_0 = 1, X_1 = 2)$

$= P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1)P(X_2 = 3 \mid X_1 = 2)$ By Markov properties

$= (1/3)p_{12}p_{23} = (1/3)(1/2)(2/3) = \frac{2}{18} = \frac{1}{9}$

Example.

A man either drives his car or takes a train to work each day. Suppose he never takes the train two days in a row, but if he drives to work, then the next day he is just as likely to drive again as he is to take the train.

The state space of the system is $\{t(train), d(drive)\}$. This stochastic process is a Markov chain since the outcome on any day depends only on what happened the preceding day. The transition matrix of the Markov chain is

$$\begin{array}{cc} & \begin{array}{cc} t & d \end{array} \\ \begin{array}{c} t \\ d \end{array} & \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

The first row of the matrix corresponds to the fact that he never takes the train two days in a row and so he definitely will drive the day after he takes the train. The second row of the matrix corresponds to the fact that the day after he drives he will drive or take the train with equal probability.

Example.

Card colour with replacement.

$$\begin{array}{cc} & \begin{array}{cc} B & R \end{array} \\ \begin{array}{c} B \\ R \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

Example.

Three boys *A*, *B* and *C* are throwing a ball to each other. *A* always throws the ball to *B* and *B* always throws the ball to *C*; but *C* is just as likely to throw the ball to *B* as to *A*. let X_n denote the n^{th} person to be thrown the ball. The state space of the system is *{A,B,C}*. This is a Markov chain since the person throwing the ball is not influenced by those who previously had the ball. The transition matrix of the Markov chain is

$$\begin{array}{cc} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \end{array}$$

The first row of the matrix corresponds to the fact that *A* always throws the ball to *B*. The second row of the matrix corresponds to the fact that *B* always throws the ball to *C*. the last row corresponds to the fact that *C* throws the ball to *A* or *B* with equal probability(and does not throw it to himself)

Lesson 6: Higher Transition probability matrix and Probability Distributions

n- Step Transition probability matrix

Consider a Markov chain $\{X_n, n = 0, 1, \dots\}$ if $X_0 = i$ then $X_1 = j$ with probability p_{ij} . That is, p_{ij} is the probability of going from state i to state j in one step: $i \to j$. Now suppose that we are interested in finding the probability of going from state i to state j in two steps, i.e.

$$p_{ij}^{(2)} = P(X_2 = j \mid X_0 = i)$$

We can find this probability by applying the law of total probability. In particular we argue that X_1 can take one of the possible values in S. Thus we can write

$$\begin{aligned} p_{ij}^{(2)} = P(X_2 = j \mid X_0 = i) &= \textstyle\sum_{k\in S} P(X_2 = j \mid X_1 = k, X_0 = i)P(X_1 = k \mid X_0 = i) \\ &= \textstyle\sum_{k\in S} P(X_2 = j \mid X_1 = k)P(X_1 = k \mid X_0 = i) \quad \text{(by Markov properties)} \\ &= \textstyle\sum_{k\in S} p_{kj}p_{ik} \end{aligned}$$

We conclude

$$p_{ij}^{(2)} = P(X_2 = j \mid X_0 = i) = \textstyle\sum_{k\in S} p_{kj}p_{ik}$$

That means, In order to get to state j, we need to pass through some intermediate state k.

$$i \to k \to j$$

Accordingly, we can define the two-step transition matrix as follows:

$$P^{(2)} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & \dots. & p_{1r}^{(2)} \\ p_{21}^{(2)} & p_{21}^{(2)} & \dots. & p_{2r}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1}^{(2)} & p_{r2}^{(2)} & \dots. & p_{rr}^{(2)} \end{bmatrix}$$

Thus, we conclude that the two-step transition matrix can be obtained by squaring the state transition matrix, i.e.,

$$P^{(2)} = PP = P^2$$

Similarly, $P^{(3)} = PP^2$

Generally, we can define the n-step transition probabilities $p_{ij}^{(n)}$ as

$$p_{ij}^{(n)} = P(X_n = j \mid X_0 = i) \qquad \text{for n} = 0,1,2, \dots$$

That means, In order to get to state j, we need to pass through some intermediate states $k_1, k_2, \dots, k_{n-1}$.

$$i \to k_1 \to k_2, \dots \to k_{n-1} \to j$$

and the n-step transition matrix, $P^{(n)}$, as

$$P^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1r}^{(n)} \\ p_{21}^{(n)} & p_{21}^{(n)} & \cdots & p_{2r}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1}^{(n)} & p_{r2}^{(n)} & \cdots & p_{rr}^{(n)} \end{bmatrix}$$

Similar to the case of two-step transition probabilities, we can show that

$$P^{(n)} = P^n \qquad \text{for n} = 1,2,3, \ldots$$

More generally, let m and n be two positive integers and assume $X_0 = i$. In order to get to state j in $(m+n)$ steps, the chain will be at some intermediate state k after m steps. To obtain $p_{ij}^{(m+n)}$, we sum over all possible intermediate states:

$$p_{ij}^{(m+n)} = P(X_{m+n} = j \mid X_0 = i) = \sum_{k\in S} p_{kj}^{(m)} p_{ik}^{(n)}$$

The above equation is called the **Chapman-Kolmogorov equation.**

The probability distribution $\{X_n, n \geq 0\}$

Consider a Markov chain $\{X_n, n = 0, 1, \ldots\}$ where $X_n \in S = \{1,2,\cdots,r\}$. Suppose that we know the probability distribution of X_0. More specifically, define the row vector π^0 as

$$\pi^{(0)} = [P(X_0 = 1) \quad P(X_0 = 2) \ldots P(X_0 = r)]$$

How can we obtain the probability distribution of X_1, X_2, $\cdots\cdots$? We can use the law of total probability. More specifically, for any $j \in S$, we can write

$$P(X_1 = j) = \sum_{k=1}^{r} P(X_1 = j \mid X_0 = k) P(X_0 = k)$$

$$= \sum_{k=1}^{r} p_{kj} P(X_0 = k)$$

If we generally define

$$\pi^{(n)} = [P(X_n = 1) \quad P(X_n = 2) \ldots P(X_n = r)]$$

we can rewrite the above result in the form of matrix multiplication

$$\pi^{(1)} = \pi^{(0)} P$$

where P is the state transition matrix. Similarly, we can write

$$\pi^{(2)} = \pi^{(1)} P$$

More generally, we can write

$$\pi^{(n+1)} = \pi^{(n)} P \qquad \text{for } n = 0,1,2, \ldots$$

$$\pi^{(n)} = \pi^{(0)} P^n \qquad \text{for } n = 0,1,2, \ldots$$

Example.

Consider the Markov chain for the example of the man who either drives his car or takes a train to work.

$$\begin{array}{cc} & \begin{array}{cc} t & d \end{array} \\ \begin{array}{c} t \\ d \end{array} & \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

Here *t* is the state of taking a train to work and *d* of driving to work

$$P^4 = P^2P^2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix} = \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix}$$

Thus the probability that the process changes from, say, state *t* to state *d* in exactly 4 steps is $\frac{5}{8}$, i.e. $p_{td}^{(4)} = \frac{5}{8}$. similarly, $p_{tt}^{(4)} = \frac{3}{8}$, $p_{dt}^{(4)} = \frac{5}{16}$ and $p_{dd}^{(4)} = \frac{11}{16}$.

Suppose that on the first day of work, the man toss affair die and drove to work if only if a 6 is appeared. In other words, $\pi^{(0)} = [5/6 \quad 1/6]$ is the initial probability distribution then,

$$\pi^{(4)} = \pi^{(0)}P^4 = [5/6 \quad 1/6] \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix} = [35/96 \quad 61/96]$$

Is the probability distribution after 4 days, i.e. $\pi_t^{(4)} = \frac{35}{96}$ and $\pi_d^{(4)} = \frac{61}{96}$.

Example.

Consider a system that can be in one of two possible states, $S = \{0,1\}$. In particular, suppose that the transition matrix is given by

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix}$$

Suppose that the system is in state 0 at time $n = 0$, i.e., $X_0 = 0$. Find the probability that the system is in state 1 at time n =3.

Solution:

Here we know $\pi^{(0)} = [P(X_0 = 0) \quad P(X_0 = 1)] = [1 \quad 0]$

Thus, the probability that the system is in state 1 at time n =3 is $43/72$.

$$\pi^{(3)} = \pi^{(0)}P^3 = [1 \quad 0]\begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix}^3 = \begin{bmatrix} 29/72 & 43/72 \end{bmatrix}$$

Thus, the probability that the system is in state 1 at time n =3 is $\frac{43}{72}$.

Example.

Consider the Markov chain for the example of the three boys *A*, *B* and *C* who are throwing a ball to each other .

$$\begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

Suppose C was the first person with ball, i.e. $\pi^{(0)} = [0 \quad 0 \quad 1]$ is the initial probability distribution then,

$$\pi^{(1)} = \pi^{(0)}P = [0 \quad 0 \quad 1]\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix}$$

$$\pi^{(2)} = \pi^{(1)}P = \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix}$$

$$\pi^{(3)} = \pi^{(2)}P = \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 1/2 \end{bmatrix}$$

Thus, after 3 throws, the probability that A has the ball is 1/4 , that B has the ball is 1/4 and that C has a ball is 1/2 . $\pi_A^{(3)} = \frac{1}{4}$, $\pi_B^{(3)} = \frac{1}{4}$ and $\pi_C^{(3)} = \frac{1}{2}$

Example. A school contains 200 boys and 150 girls. One student is selected after another to take an eye examination.

Explain whether this process is a Markov Chain or not and why?

Solution

The state space of the stochastic process is $\{m(male), f(female)\}$. However, this process is not a Markov chain since, for example the probability that the third person is a girl depends not only on the outcome of the second trial but on both the first and second trials.

Lesson 7: Stationary Distribution and Regular Markov Chain

Stationary Distribution

Let P be the transition probability matrix of a Markov chain $\{X_n, n \geq 0\}$. If there exists a probability vector $\hat{p}$ such that:

$$\hat{p}P = \hat{p} \qquad (1)$$

then $\hat{p}$ is called a stationary distribution for the Markov chain.

Example.

Find stationary distribution $\hat{p}$ for the transition matrix P of a Markov chain:

$$P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

We seek probability vector with two components, which we can denote by $\hat{p} = [x \quad 1-x]$ such that $\hat{p}P = \hat{p}$:

$$[x \quad 1-x]\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = [x \quad 1-x]$$

Multiplying the left side of the above matrix equation, we obtain

$$\left[\tfrac{1}{2} - \tfrac{1}{2}x \qquad \tfrac{1}{2} + \tfrac{1}{2}x\right] = [x \quad 1-x] \text{ or } \begin{cases} \tfrac{1}{2} - \tfrac{1}{2}x & = x \\ \tfrac{1}{2} + \tfrac{1}{2}x & = 1-x \end{cases} \quad \text{or} \quad x = \frac{1}{3}$$

Thus $\hat{p} = [x \quad 1-x] = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$.

Thus in the long run, the man will take the train to work $\frac{1}{3}$ of the time, and drive to work the other $\frac{2}{3}$ of the time.

Example

Find stationary distribution $\hat{p}$ for the transition matrix P of a Markov chain:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Suppose that the vector is $\hat{p} = [x \qquad y \qquad 1-x-y]$

Solution

$$\hat{p} = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{bmatrix}$$

Thus in the long run, *A* will be thrown the ball 20% of the time, and B and C 40% of the time.

Regular Markov chain

A Markov chain is called regular if there is a finite positive integer m such that after m time-steps, every state has a nonzero chance of being occupied, no matter what the initial state. Let $A > 0$ denote that every element a_{ij} of A satisfies the condition $a_{ij} > 0$. Then, for a regular Markov chain with transition probability matrix P, there exists an $m > 0$ such that $P^m > 0$.

Example.

The following transition matrix P of a Markov chain:

$$P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

Is regular matrix since,

$$P^2 = P.P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

Is positive in every entry

Example.

The following transition matrix P of a Markov chain:

$$P = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

Is irregular matrix since,

$$P^2 = \begin{bmatrix} 1 & 0 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}, \qquad P^3 = \begin{bmatrix} 1 & 0 \\ \frac{7}{8} & \frac{1}{8} \end{bmatrix}, \qquad P^4 = \begin{bmatrix} 1 & 0 \\ \frac{15}{16} & \frac{1}{16} \end{bmatrix}$$

every power m will have 1 and 0 in the first row.

Stationary distribution of regular Markov chain

Let $\{X_n, n \geq 0\}$ be a regular finite-state Markov chain with transition matrix P. Then

$$\lim_{n\to\infty} P^n = \hat{P}$$

Where $\hat{P}$ is a matrix whose rows are identical and equal to the stationary distribution $\hat{p}$ for the Markov chain defined by Eq. (1). In other words, P^n approaches $\hat{P}$ means that each entry of P^n approaches the corresponding entry of $\hat{P}$, and $\hat{p}P$ approaches $\hat{p}$ means that each component of $\hat{p}P$ approaches the corresponding responding component of $\hat{p}$.

Example. For the regular transition matrix P of a Markov chain below find the matrix $\hat{P}$:

$$P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

We found before that: $\hat{p} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$. Thus: $\hat{P} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0.33 & 0.67 \\ 0.33 & 0.67 \end{bmatrix}$

We exhibit some of the powers of P to indicate the above result:

$$P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 0.50 & 0.50 \\ 0.25 & 0.75 \end{bmatrix}, \qquad P^3 = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.75 \\ 0.37 & 0.63 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{bmatrix} = \begin{bmatrix} 0.37 & 0.63 \\ 0.31 & 0.69 \end{bmatrix}, \qquad P^5 = \begin{bmatrix} \frac{5}{16} & \frac{11}{16} \\ \frac{11}{32} & \frac{21}{32} \end{bmatrix} = \begin{bmatrix} 0.31 & 0.69 \\ 0.34 & 0.66 \end{bmatrix}$$

Theorem.

If a stochastic matrix P has 1 on the main diagonal, the P is not regular (unless P is 1×1)

Calculating probability

The probabilities for a Markov chain are computed using the initial probabilities $\pi_{i_0}^{(0)} = P(X_0 = i_0)$ and the transition probabilities p_{ij}

$$P(X_0 = i_0, X_1 = i_1, \ldots., X_n = i_n) = \pi_{i_0}^{(0)}\, p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$$

Example.

Consider a Markov chain with the following transition matrix:

$$P = \begin{array}{c|cc} & 0 & 1 \\ 0 & \frac{3}{4} & \frac{1}{4} \\ 1 & \frac{1}{6} & \frac{5}{6} \end{array}$$

1. To find the probability that the process follows a certain path, you multiply the initial probability with conditional probabilities. For example, what is the chance that the process begins with 01010?

$$P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0) = \pi_0^{(0)}\, p_{01} p_{10} p_{01} p_{10}$$

$$\pi_0^{(0)} \times \tfrac{1}{4} \times \tfrac{1}{6} \times \tfrac{1}{4} \times \tfrac{1}{6} = \pi_0^{(0)} \times \tfrac{1}{576}$$

2. Find the chance that the process begins with 00000.

$$P(X_0 = 0, X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0) = p_{(0)}p(0,0)p(0,0)p(0,0)p(0,0)$$

$$= \pi_0^{(0)} p_{00}p_{00}p_{00}p_{00} = \pi_0^{(0)} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \pi_0^{(0)} \times \frac{81}{256}$$

If (as in many situations) we were interested in conditional probabilities, given that $X_0 = 0$, we simply drop $= \pi_0^{(0)}$, that is,

$$P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0 \mid X_0 = 0) = \frac{1}{576} = 0.00174$$

$$P(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0 \mid X_0 = 0) = \frac{81}{256} = 0.31641$$

Example.

If we start in state zero, then $\pi^{(0)} = [0 \quad 1]$ and

$$\begin{aligned} \pi^{(4)} &= \pi^{(0)}P^4 \\ &= [0 \quad 1]\begin{bmatrix} \frac{3245}{6912} & \frac{3667}{6912} \\ \frac{3667}{10368} & \frac{6701}{10368} \end{bmatrix} \\ &= \begin{bmatrix} \frac{3245}{6912} & \frac{3667}{6912} \end{bmatrix} \\ &= [0.46947 \quad 0.53053] \end{aligned}$$

On the other hand, if we flip a coin to choose the starting position then,

$\pi^{(0)} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ and

$$\begin{aligned} \pi^{(4)} &= \pi^{(0)}P^4 \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{3245}{6912} & \frac{3667}{6912} \\ \frac{3667}{10368} & \frac{6701}{10368} \end{bmatrix} \\ &= \begin{bmatrix} \frac{17069}{41472} & \frac{24403}{41472} \end{bmatrix} \\ &= [0.41158 \quad 0.58840] \end{aligned}$$

Lesson 8: Classification of States

Accessible States:

State j is said to be accessible from state i if , $p_{ij}^{(n)} > 0$ for some n , written as $i \rightarrow j$.

Communicative states

Two states i and j are called communicative states written as $i \leftrightarrow j$, if they are accessible from each other. In other words $i \leftrightarrow j$ means $j \rightarrow i$ and $i \leftarrow j$

Irreducible Markov Chain

A Markov chain is said to be irreducible if all states communicate with each other.

Communication is an equivalence relation. That means that

1. Every state communicates with itself $i \leftrightarrow i$
2. $i \leftrightarrow j$ implies $j \leftrightarrow i$ and
3. $i \leftrightarrow j$ and $j \leftrightarrow k$ together imply $i \leftrightarrow k$

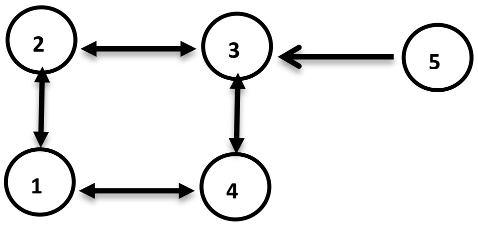
- If the transition matrix is not irreducible, then it is not regular
- If the transition matrix is irreducible and at least one entry of the main diagonal is nonzero, then it is regular.

Class structure

The accessibility relation divides states into classes. Within each class, all states communicate to each other, but no pair of states in different classes communicates. The chain is irreducible if there is only one class.

Example:

Consider a Markov Chain shown in figure below



Any state 1, 2, 3, 4 is accessible from any of the five states, but 5 is not accessible from 1, 2, 3, 4. So we have two classes: $\{1, 2, 3, 4\}$, and $\{5\}$. The chain is not irreducible.

Example

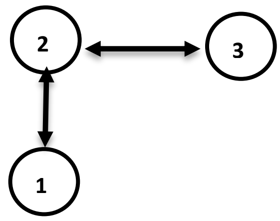
Consider the chain on states 1, 2, 3, determine whether it is irreducible or not?

$$P=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Solution

There is only one class because $1 \leftrightarrow 2$ and $2 \leftrightarrow 3$, this is an irreducible Markov chain.

Other solution is that



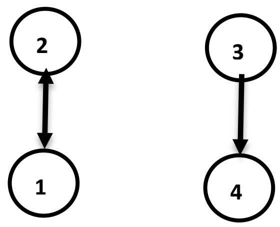
$$P^2=\begin{bmatrix} \frac{4}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{18}{48} & \frac{19}{48} & \frac{11}{48} \\ \frac{18}{108} & \frac{33}{108} & \frac{57}{108} \end{bmatrix}$$

Since all $p_{ij}^{(2)} > 0$, then this chain is irreducible.

Example

Consider the chain on states 1, 2, 3, 4, and

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

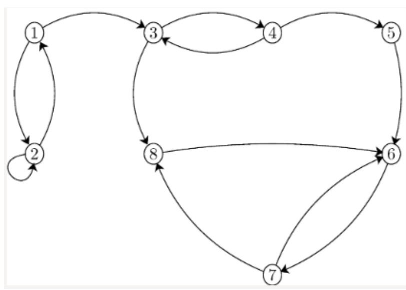


This chain has three classes {1, 2}, {3} and {4} and is hence not irreducible.

and also some $p_{ij}^{(m)} = 0$.

Example

Consider the Markov chain shown in Figure below. It is assumed that when there is an arrow from state i to state j, then $p_{ij} > 0$. Find the classes for this Markov chain.



Solution

There are 4 communicating classes in this Markov Chain. States 1, 2 and states 3, 4 communicate with each other, but they do not communicate with any other notes. State 5 does not communicate with any other states, so it by itself is a class. States 6,7 and 8 construct another class. Thus, the class are

Class 1= {state 1, state2}

Class 2= {state 3, state4}

Class 3= {state 5}

Class 4= {state 6, state7, state8}

Closed Set (class)

A set of states S in a matrix chain is closed set if no state outside of S is reachable from any state in S. In the above example the set $\{1,2\}$ is closed.

Absorbing States:

State j is said to be an absorbing state if $p_{jj} = 1$; that is, once state j is reached, it is never left. Thus state j is absorbing if and only if the j^{th} row of transition matrix P has a 1 on the main diagonal and zeros everywhere else. (The main diagonal of an n-square matrix $P = (p_{ij}) = (p_{11} \quad p_{22} \quad \cdots \quad p_{nn})$). An absorbing state is a **closed set** containing only one state.

Absorption Probabilities:

Consider a Markov chain $\{X_n, n \geq 0\}$ with finite state space $S = (1, 2, \ldots, N)$ and transition probability matrix P. Let $A = (1, \ldots, m)$ be the set of absorbing states and $B = \{m + 1, \ldots, N)$ be a set of non-absorbing states. Then P can be expressed as

$$P = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 & \cdots & 0 \\ p_{m+1,1} & \cdots & \cdots & p_{m+1,m} & p_{m+1,m+1} & \cdots & p_{m+1,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & \cdots & \cdots & p_{N,m} & \cdots & \cdots & p_{N,N} \end{pmatrix} = \begin{pmatrix} I & O \\ R & Q \end{pmatrix}$$

where I is an $m \times m$ identity matrix, O is an $m \times N - m$ zero matrix, and:

$$R = \begin{bmatrix} p_{m+1,1} & \cdots & p_{m+1,m} \\ \vdots & \ddots & \vdots \\ p_{N,1} & \cdots & p_{N,m} \end{bmatrix} \qquad Q = \begin{bmatrix} p_{m+1,m+1} & \cdots & p_{m+1,N} \\ \vdots & \ddots & \vdots \\ p_{N,m+1} & \cdots & p_{N,N} \end{bmatrix}$$

Note that the elements of R are the one-step transition probabilities from non-absorbing to absorbing states, and the elements of Q are the one-step transition probabilities among the non-absorbing states.

Example.

Suppose the following matrix P is a transition matrix of a Markov chain.

The state a_2 and a_5 are each absorbing, since each of the second and fifth rows has a 1 on the main diagonal.

$$\begin{array}{c|c} & \begin{array}{ccccc} a_1 & a_2 & a_3 & a_4 & a_5 \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} & \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Example. (**Random walk with absorbing barriers**)

A man is at in integral point on the x-axis between the original O and, say, the 4. He takes a unit step to the right with probability p or to the left with probability $q = 1 - p$, we assume that the man remains at either endpoint whenever he reaches there.

We call this process a random walk with absorbing barriers, since the state a_0 and a_4 are each absorbing. In this case $\pi_0^{(n)}$ denotes the probability that the man reaches the state a_0 on or before the nth step. Similarly, $\pi_4^{(n)}$ denotes the probability that he reaches the state a_4 on or before the nth step.

$$\begin{array}{c|c} & \begin{array}{ccccc} a_0 & a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{array}{c} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Example.

A man tosses a fair coin until 3 heads occur. Let $X_n = k$ if, at the nth trail, the last tail occurred at the $(n-k)$-th trial, i.e. X_n denotes the longest string of heads ending at the nth trial. This is a Markov chain process with state space $S = (a_0, a_1, a_2, a_3)$, where a_i means the string of heads has length i. the transition matrix is

$$\begin{array}{c|c} & \begin{array}{cccc} a_0 & a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} a_0 \\ a_1 \\ a_2 \\ a_3 \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Each row, except the last, corresponds to the fact that a string of heads is either broken if a tail occurs or is extended by one if a head occurs. The last line corresponds to the fact that the game ends if 3 heads are tossed in a row. Note that a_3 is an absorbing state.

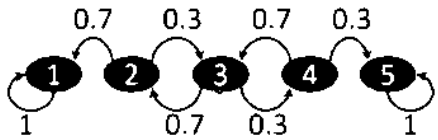
Recurrent and Transient States

A state is said to be **recurrent** if, any time that we leave that state, we will return to that state in the future with probability one. On the other hand, if the probability of returning is less than one, the state is called **transient**.

Definition.

State i is **transient** state if there exists state j that is reachable from i, but the state i is not reachable from state j. If state is not transient, it is **recurrent** state.

Example. Consider the following Markov chain



Transient state: 2, 3,4 Recurrent state:1,5

For example we can go from 3 to 2 then 2 to 1 then we get trap in state 1 and we will never come back to state 3 a gain.

Note that:

1. If two states are in the same class, either both of them are recurrent, or both of them are transient.
2. A class is said to be recurrent if the states in that class are recurrent. If, on the other hand, the states are transient, the class is called transient.
3. A Markov chain might consist of several transient classes as well as several recurrent classes.

Definition (Alternative)

A set (class) **C** is called irreducible if whenever $i, j \in C, i$ communicates with j.

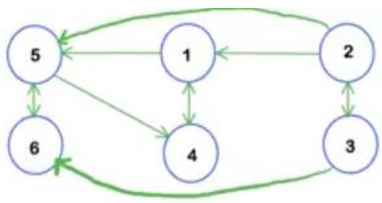
A set (class) **C** is closed if it is impossible to get out.

If **C** is a finite closed and irreducible set, then all states in C are recurrent.

If state A goes to state B but state B doesn't go to state A then state A is transient.

Example

Consider the Markov chain shown in Figure below. Identify the transient and recurrent states, and the irreducible closed classes.



$2 \rightarrow 1$ $1 \nrightarrow 2$ $\Rightarrow$ 2 is transient $3 \rightarrow 6$ $6 \nrightarrow 3$ $\Rightarrow$ 3 is transient

Class=$\{1,4,6,5\}$ is closed and irreducible 1,4,6,5 are recurrent

Example.

Consider the chain on states 1, 2, 3, 4, and

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Obviously, state 4 is recurrent, as it is an absorbing state $f_{44} = 1$. The only possibility to return to 3 is to do so in one step, so we have $f_{33} = \frac{1}{4}$ and 3 is transient.($3 \rightarrow 4$ $3 \nrightarrow 4$).

Class=$\{1,2\}$ is closed and irreducible 1,2 are recurrent.

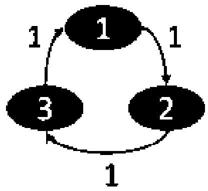
Periodic and Aperiodic States

A state i is **periodic** with period $k > 1$, if k is the smallest number such that all paths leading from state i back to state i have a length that is a multiple of k.

- Absorbing states are aperiodic
- If we can return to a recurrent state at irregular times, it is aperiodic

Example

Consider a Markov chain with the following state diagram

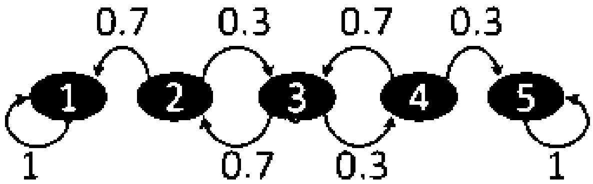


All state are periodic Period k =3

For example starting from state 1 we need 3, 6, or 9 steps to come back to state 1the multiple is 3.

Example

Consider a Markov chain with the following state diagram



States 1 and 5 are **aperiodic** because they are absorbing states

The transient states 2, 3 and 4 are **aperiodic** we may not come back to this state again.

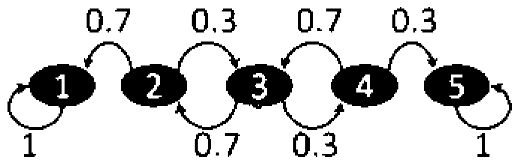
Note that the class {2,3,4} **is not closed.**

Ergodic Markov Chain

If all states in a Markov chain are **recurrent** (not transient), **aperiodic** (not periodic) and **communicate** with each other, the chain is said to be **ergodic**.

Example

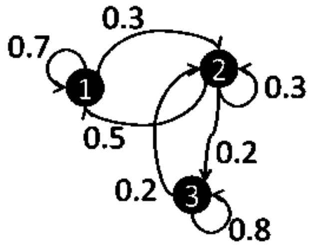
Consider a Markov chain with the following state diagram



Not ergodic

Example

Consider a Markov chain with the following state diagram



S1

Is ergodic because all state are recurrent and communicate with other and it is aperiodic because (k=0) there is no period every state can communicate with other state.

Solved Problems (2)

Problem 1: (Bernoulli Process)

Let $X_1, X_2, \ldots$ be independent Bernoulli r.v.'s with $P(X_n = 1) = p$ and$P(X_n = 0) = q = 1 - p \; \forall \; n$ The collection of r.v.'s $\{X_n, n \geq 1\}$ is a random process, and it called a Bernoulli process.

(a) Describe the Bernoulli process.

(b) Construct a typical sample sequence of the Bernoulli process.

(c) Find the transition matrix of this process considering a sequence of coin flips, where each flip has probability of p of having the same outcome as the previous coin flip, regardless of all previous flips

Problem 2: (Binomial Process)

Let $X_1, X_2, \ldots$ be independent Bernoulli r.v.'s with $P(X_n = 1) = p$ and $P(X_n = 0) = q = 1 - p$ for all n . Let S_n be number of success in n trial of Bernoulli, then the stochastic process $\{S_n, n \in N\}$ is called Binomial process and it denoted by

$$S_n = \begin{cases} 0, & n = 0, \\ X_1 + X_2 + \ldots + X_n, & n \geq 1. \end{cases}$$

(a) Describe the Binomial process.

(b) Find the transition matrix of this process

Problem 3:(Simple Random Walk Process)

Let $Z_1, Z_2, \ldots$ be independent identically distributed r.v.'s with $P(Z_n = 1) = p$ and $P(Z_n = -1) = q = 1 - p$ for all n . let $X_n = \sum_{i=1}^{n} Z_i \quad n = 1,2, \ldots$ and $X_0 = 0$. The collection of r.v.'s $\{X_n, n \geq 0\}$ is a random process, and it is called the simple random walk process in one dimension.

(a) Describe the simple random walk process.

(b) Construct a typical sample sequence of the process.

(c) Find the transition matrix of this process

Problem 4:

Student's study habits are as follows. If he studies one night, he is 70% sure not to study the next night. On the other hand, if he does not study one night, he is 60% sure not to study the next night as well. Find

(a) The transition matrix of this process. (b) The transition matrix after 4 nights.

(c) Suppose that on the first night, the student toss affair die and studied if only if a 2 or 3 are appeared. What is the probability that he didn't study in the fourth night.

Problem 5:

Consider a Markov Chain with three possible states $S = \{1, 2, 3\}$, that has the following transition matrix.

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 0 & 2/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

(a) Draw the state transition diagram for this chain.

(b) $P(X_5 = 2 \mid X_4 = 1)$

(c) $P(X_3 = 2 \mid X_2 = 2)$

(d) If we know $P(X_0 = 1) = 1/4$, find $P(X_0 = 1, X_1 = 2)$

(e) If we know $P(X_0 = 1) = 1/4$, find $P(X_0 = 1, X_1 = 2, X_2 = 3)$

(f) If we know $P(X_1 = 1) = P(X_1 = 2) = 1/4$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$

Problem 6:

A psychologist makes the following assumptions concerning the behavior of mice subjected to a particular feeding schedule. For any particular trail 80% of the mice that went right on the previous experiment will go right on this trial, and 60% of those mice that went left on the previous experiment will go right on this trial. If 50% went right on the first trial, what would he predict for:

(a) The second trial. (b) The third trial. (c) The thousandth trial.

Problem 7:

Consider a Markov chain which has the transition matrix

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

Determine

(a) $P(X_1 = 1, X_2 = 1 \mid X_0 = 0)$

(b) 2. $P(X_3 = 1 \mid X_0 = 0)$

(c) $P(X_2 = 1, X_3 = 1 \mid X_0 = 0)$

Problem 8

Determine whether each of the following is stochastic matrix or not and why given that $s = \{1,2\}$?

(a) $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 \end{bmatrix}$ (b) $B = \begin{bmatrix} \frac{15}{16} & \frac{1}{16} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$ (c) $D = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$

Problem 9

Given the transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad \text{and} \qquad \pi^{(0)} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

Find:

(a) $\pi^{(3)}$ (b) $p_{21}^{(3)}$ (c) $\pi_2^{(3)}$

Problem 10

Given the transition matrix

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \pi^{(0)} = \begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}.$$

Find:

(a) $p_{32}^{(2)}$ (b) $p_{31}^{(2)}$ (c) $\pi^{(4)}$ (d) $\pi_3^{(4)}$

Problem 11

A salesman's area consists of three cities, A, B and C. He never sells in the same city on successive days. If he sells in city A, then the next day he sells in city B. however, if he sells in either B or C, then the next day he is twice as likely to sell in city A as in the other city. Find

(a) The transition matrix of this process.

(b) In the long run, how often does he sell in each of the cities.

Problem 12

There are 2 white balls in urn A and 3 red balls in urn B. at each step of the process a ball is selected from each urn and the two balls selected are interchanged. Let the state a_i of the process be the number i of red balls in urn A. find:

(a) The transition matrix of this process.

(b) What is the probability that there are 2 red balls in urn A after 3 steps.

(c) In the long run, what is the probability that there are 2 red balls in urn A.

Problem 13

Consider the transition matrix P of a Markov chain of S={0,1} .

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

Given that $P(X_0 = 0) = P(X_0 = 1) = 0.5$.

(a) Find the distribution of X_n

(b) Find the distribution of X_n, when $n \to \infty$.

Problem 14

Consider a Markov chain of the transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \quad \text{with } S = \{0 \quad 1\}.$$

Compute,

(a) $P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 1)$ (b) $P(X_1 = 1, X_2 = 0, X_3 = 1 \mid X_0 = 0)$

(c) $P(X_4 = 1, X_5 = 1 \mid X_0 = 0)$

Problem 15

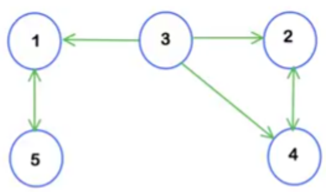
Determine whether each of the given matrices is recurrence or not

$$P = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

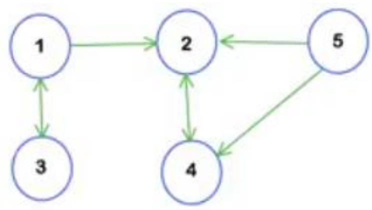
Problem 16

Consider the two Markov chains shown in Figure below. Identify the transient and recurrent states, and the irreducible closed classes in each one.

(a)

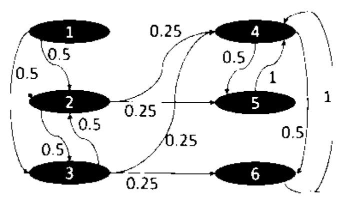


(b)



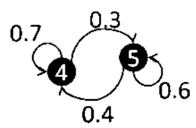
Problem 17

Consider the Markov chains shown in Figure below. Identify the transient, recurrent, periodic and aperiodic states.



Problem 18

Determine whether the following matrix is ergodic or not and Why?



Problem 19. Determine whether each of the following matrix is regular or not and why?

(a) $P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$ (b) $P = \begin{bmatrix} 0.5 & 0.5 \\ 1 & 0 \end{bmatrix}$ (c) $P = \begin{bmatrix} 1 & 0 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 1 & 0 \end{bmatrix}$

Solutions (2)

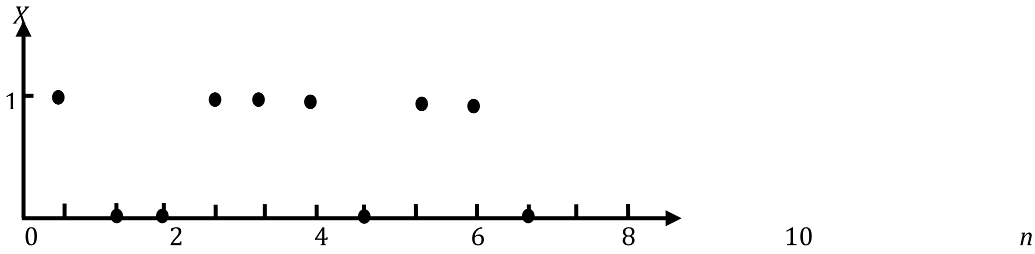
Problem 1: (Solution)

(a) The Bernoulli process $\{X_n, n \geq 1\}$ is a discrete-parameter, discrete-state process. The state space is $S=\{0,1\}$, and the parameter set is $N=\{1,2,3,\ldots\ldots\}$.

(b) A sample sequence of the Bernoulli process can be obtained by tossing a coin consecutively. If a head appears, we assign 1, and if a tail appears, we assign 0.

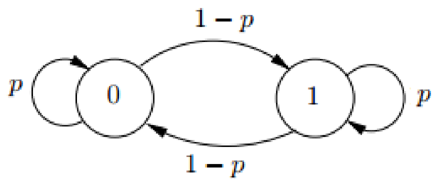
n	1	2	3	4	5	6	7	8	9	10	……
Coin tossing	H	T	T	H	H	H	T	H	H	T	……
X_n	1	0	0	1	1	1	0	1	1	0	……

The sample sequence $\{X_n\}$ obtained above is plotted in



(c) Transition matrix

$$p_{ij} = \begin{cases} p & j = i \\ 1-p & j \neq i \end{cases} \qquad P = \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \end{array}$$



Problem 2: (Solution)

(a) The Binomial process $\{S_n, n \geq 1\}$ is a discrete-parameter, discrete-state process. The state space is $S = \{0,1,2,\ldots\}$, and the parameter set is $N = \{1,2,3,\ldots\ldots\}$.

(b) Transition matrix

$$P(S_{n+1} = j \mid S_n = i, S_{n-1} = i_{n-1}, \ldots, S_1 = i_1) = p_{ij} = \begin{cases} p, & j = i+1 \quad (i = 0,1,\ldots) \\ 1-p, & j = i \quad (i = 0,1,\ldots) \\ 0, & \text{otherwise} \end{cases}$$

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & \ddots \end{array} \\ \begin{pmatrix} 1-p & p & 0 & 0 & \ddots \\ 0 & 1-p & p & 0 & \ddots \\ 0 & 0 & 1-p & p & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \end{array}$$

The first row of the matrix corresponds to the fact that in the next trial still have 1 success that means failed with probability $1-p$, get 2 successes means succeeded with probability p. Moving from 1 to 3,4,…. That impossible.

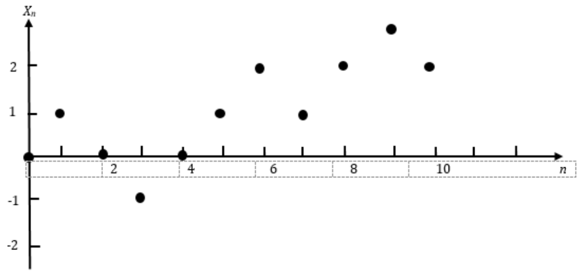
Problem 3: (Solution)

(a) The simple random walk process $\{X_n, n \geq 0\}$ is a discrete-parameter, discrete-state process. The state space is *S*={…..,-2,-1,0,1,2,…..}, and the parameter set is *T*={0,1,2,3,…….}.

(b) A sample sequence of the simple random walk process can be obtained by tossing a coin every second and letting X_n increase by unity if a head appears, and decrease by unity if a tail appears. Thus, for instance,

n	0	1	2	3	4	5	6	7	8	9	10	…
Coin tossing		H	T	T	H	H	H	T	H	H	T	…
X_n	0	1	0	-1	0	1	2	1	2	3	2	…

The sample sequence $\{X_n\}$ obtained above is plotted in



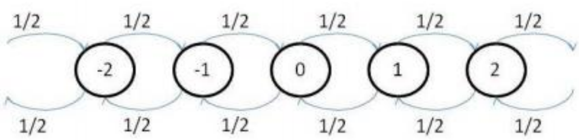
(c) Find the transition matrix of this process

(d) $P(Z_{n+1} = j - i) = \begin{cases} p & \text{if } j-i=1 \\ 1-p & \text{if } j-i=-1 \\ 0 & \text{Otherwise} \end{cases}$

$$P = \begin{array}{c} \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} \dots & -2 & -1 & 0 & 1 & 2 & \dots \end{array} \\ \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \dots & 0 & p & 0 & 0 & 0 & \dots \\ \dots & 1-p & 0 & p & 0 & 0 & \dots \\ \dots & 0 & 1-p & 0 & p & 0 & \dots \\ \dots & 0 & 0 & 1-p & 0 & p & \dots \\ \dots & 0 & 0 & 0 & 1-p & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

Suppose $P(Z_n = 1) = 1/2$ and $P(Z_n = -1) = 1/2$

$$P(Z_{n+1} = j - i) = \begin{cases} \frac{1}{2} & \text{if } |j-i| = 1 \\ 0 & \text{Otherwise} \end{cases}$$



Problem 4: (Solution)

(a) The transition matrix of this process

The state space is S={S"study",T "Not study"}

$$P = \begin{array}{c} \\ S \\ T \end{array}\begin{array}{c} \begin{array}{cc} S & T \end{array} \\ \begin{pmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{pmatrix} \end{array}$$

(b) The transition matrix after 4 nights.

$$P^4 = \begin{array}{c} \\ S \\ T \end{array}\begin{array}{c} \begin{array}{cc} S & T \end{array} \\ \begin{pmatrix} 0.36 & 0.64 \\ 0.36 & 0.64 \end{pmatrix} \end{array}$$

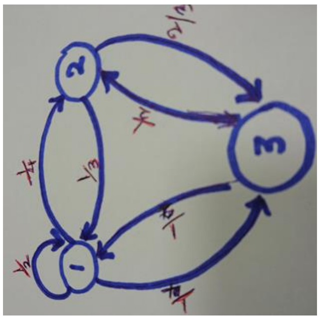
(c) The probability that he didn't study in the second night

$\pi^{(0)} = (\frac{1}{3}, \frac{2}{3})$ is the initial probability distribution then,

$$\pi^{(2)} = \pi^{(0)}P^2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}\begin{bmatrix} 0.37 & 0.63 \\ 0.36 & 0.64 \end{bmatrix} = [0.36 \ , 0.64]$$

The probability that he didn't study in the second night is $\pi_T^{(2)} = 0.64$.

Problem 5: (Solution)



(b) By definition$P(X_5 = 2 \mid X_4 = 1) = p_{12} = \frac{1}{4}$

(c) By definition $P(X_3 = 2 \mid X_2 = 2) = p_{22} = 0$

(d) $P(X_0 = 1, X_1 = 2) = P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1) = (1/4)p_{12} = (1/4)(1/4) = \frac{1}{16}$

(e) $P(X_0 = 1, X_1 = 2, X_2 = 3) = P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1)P(X_2 = 3 \mid X_0 = 1, X_1 = 2)$
$= P(X_0 = 1)P(X_1 = 2 \mid X_0 = 1)P(X_2 = 3 \mid X_1 = 2)$ By Markov properties
$= (1/4)p_{12}p_{23} = (1/4)(1/4)(2/3) = \frac{2}{48} = \frac{1}{24}$

(f) $P(X_1 = 3, X_2 = 2, X_3 = 1) = P(X_1 = 3)P(X_2 = 2 \mid X_1 = 3)P(X_3 = 1 \mid X_1 = 3, X_2 = 2)$

$$= P(X_1 = 3)P(X_2 = 2 \mid X_1 = 3)P(X_3 = 1 \mid X_2 = 2)$$

$$P(X_1 = 3) = 1 - P(X_1 = 1) = P(X_1 = 2)$$
$$= 1 - \frac{1}{4} - \frac{1}{4} = 1/2$$
$$P(X_1 = 3, X_2 = 2, X_3 = 1) = \frac{1}{2}p_{32}p_{21}$$
$$= (1/2)(1/2)(1/3) = 1/12$$

Problem 6: (Solution)

The state space is S={R"Right",L "Left"} , then transition matrix of this process is,

$$P = \begin{array}{c c} & \begin{array}{cc} R & L \end{array} \\ \begin{array}{c} R \\ L \end{array} & \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \end{array}$$

The probability distribution for the first trial $\pi^{(1)} = [0.5 \quad 0.5]$.

(a) To compute the probability distribution for the next step, i.e. the second trial, multiply p by the transition matrix P.

$$[0.5 \quad 0.5] \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = [0.7 \quad 0.3]$$

Thus, on the second trial he predicts that 70% of the mice will go right and 30% will go left.

(b) To compute the probability distribution for the third trial, multiply that of the second trial by the transition matrix P. $[0.7 \quad 0.3] \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = [0.74 \quad 0.26]$

Thus, on the third trial he predicts that 74% of the mice will go right and 26% will go left.

(c) We assume that the probability distribution for the thousandth trial is essentially the stationary probability distribution of the Markov chain, and we compute it by the following

$$[x \quad 1-x] \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = (x \quad 1-x) = [0.75 \quad 0.25]$$

Thus, on the thousandth trial he predicts that 75% of the mice will go right and 25% will go left.

Problem 7: (Solution)

(a) $P(X_1 = 1, X_2 = 1 \mid X_0 = 0) = p_{01}p_{11} = 0.2 \times 0.6 = 0.12$

(b) $P(X_3 = 1 \mid X_0 = 0) = p^{(3)}{}_{01} = 0.478$

$$P^3 = \begin{pmatrix} 0.478 & 0.264 & 0.250 \\ 0.36 & 0.256 & 0.304 \\ 0.57 & 0.18 & 0.25 \end{pmatrix}$$

(c) $P(X_2 = 1, X_3 = 1 \mid X_0 = 0) = p^{(2)}{}_{01}p^{(2)}{}_{11} = 0.26 \times 0.36 = 0.0936$

$$P^2 = \begin{pmatrix} 0.54 & 0.26 & 0.2 \\ 0.44 & 0.36 & 0.44 \\ 0.6 & 0.1 & 0.3 \end{pmatrix}$$

Problem 8: (Solution)

(a) Stochastic because $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1 \quad \forall i$

(b) not stochastic because $\sum_j p_{ij} > 1$ for $i = 1$

(c) not stochastic because $p_{12} < 0$

Problem 9: (Solution)

$P^3 = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} \end{bmatrix} \quad \pi^{(0)} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$

(a) $\pi^{(3)} = \pi^{(0)} P^3 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$ (b) $p_{21}^{(3)} = 3/8$ (c) $\pi_2^{(3)} = 2/3$

Problem 10: (Solution)

(a) $p_{32}^{(2)} = \frac{1}{2}$ (b) $p_{31}^{(2)} = 0$ (c) $\pi^{(4)} = \begin{bmatrix} \frac{3}{12} & \frac{7}{12} & \frac{2}{12} \end{bmatrix}$ (d) $\pi_3^{(4)} = \frac{1}{6}$

Problem 11: (Solution)

(a) The state space is S={A,B,C} $P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}$

(b) $\hat{p} = \begin{bmatrix} \frac{2}{5} & \frac{9}{20} & \frac{3}{20} \end{bmatrix} = [40\% \quad 45\% \quad 15\%]$

Thus in the long run he sell 40% of the time in city *A*, 45% of the time in city *B*, and 15% of the time in city *C*.

Problem 12: (Solution)

(a) There are 3 states a_0, a_1, a_2 describe by the following diagrams:

2 W	3 R	1 W 1 R	1 W 2 R	2 R	2 W 1 R
A	B	A	B	A	B
a_0		a_1		a_2	

The transition matrix is:

$$\begin{array}{c|c} & \begin{matrix} a_0 & a_1 & a_2 \end{matrix} \\ \begin{matrix} a_0 \\ a_1 \\ a_2 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix} \end{array}$$

For example, if the process is in state a_0 then a white ball must be selected from urn A and a red ball from urn B, so the process must move to state a_1 . Accordingly, the first row of the transition matrix is $[0 \quad 1 \quad 0]$. To move from a_1 to a_0 a red ball must be selected from urn A and a white ball from urn B, with prob. $\frac{1}{6}$ ($P_A(R).P_B(W) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$).

Problem 13: (Solution)

Since the initial distribution $\pi^{(0)} = [0.5 \quad 0.5]$

(a) the distribution of X_n is given by: $\pi^{(n)} = \pi^{(0)} P^n = [0.5 \quad 0.5] \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^n$

(b) since P is regular so

$$\lim_{n\to\infty} P^n = \hat{P} \text{ and each row in } \hat{P} \text{ is } \hat{p} \quad \text{so} \quad \hat{p}P = \hat{p}$$

$$[x \quad 1-x]\begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = (x \quad 1-x) = [2/3 \quad 1/3]$$

distribution of X_n, when $n \to \infty$ is $[2/3 \quad 1/3]$

Problem 14: (Solution)

(a) $P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 1) = \pi_0^{(0)} p_{10} p_{10} p_{01} = \pi_0^{(0)} \times 1 \times \frac{1}{3} \times 1 = \frac{1}{3}\pi_0^{(0)}$

(b) $P(X_1 = 1, X_2 = 0, X_3 = 1 \mid X_0 = 0) = p_{10} p_{10} p_{01} = \frac{1}{3}$

(c) $P(X_4 = 1, X_5 = 1 \mid X_0 = 0) = p_{01}^4 p_{11}^4 = 0.740741 \times 0.753086 = 0.557842$

$$P^4 = \begin{bmatrix} 0.259259 & 0.740741 \\ 0.246914 & 0.753086 \end{bmatrix}$$

Problem 15: (Solution)

$$P^2 = \begin{bmatrix} 0.16 & 0.49 & 0.35 \\ 0.07 & 0.33 & 0.60 \\ 0.02 & 0.24 & 0.74 \end{bmatrix}$$

All $p_{ij} > 0$ hence, P is irreducible .

Problem 16: (Solution)

(a)

$3 \to 1 \quad 1 \nrightarrow 3 \quad \Rightarrow \quad 3$ is transient

Class=$\{1,5\}$ is closed and irreducible

1,5 are recurrent.

Class=$\{2,4\}$ is closed and irreducible

2,4 are recurrent.

(b)

$1 \to 2 \quad 2 \nrightarrow 1 \quad \Rightarrow \quad 1$ is transient

$3 \to 2 \quad 2 \nrightarrow 3 \quad \Rightarrow \quad 3$ is transient

$5 \to \begin{matrix} 2 \\ 4 \end{matrix} \quad \begin{matrix} 2 \\ 4 \end{matrix} \nrightarrow 2 \quad \Rightarrow \quad 5$ is transient

Class=$\{2,4\}$ is closed and irreducible

2,4 are recurrent.

Class=$\{2,4\}$ is closed and irreducible

2,4 are recurrent.

Problem 17: (Solution)

Transient: 1, 2 and 3

Recurrent: 4, 5 and 6

Aperiodic : transient states 1,2 and 3

Periodic : 4, 5 and 6 with a period $k = 2$.

Problem 18: (Solution)

A Markov chain is ergodic because state 4 and 5 are recurrent and with aperiodic period $k = 0$. Transient: 1, 2 and 3

Problem 19: (Solution)

(a) yes, all entries are positive

(b) yes because has only positive entries.

$$P^2 = \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix}$$

(c) No, because it is not irreducible (not connectable). Also, if you multiply it by itself over and over it will still contain zeros

Poisson Processes

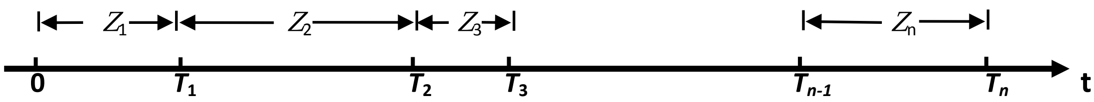
Lesson 9: Counting Process

Definition: (Points of Occurrence)

Let t represent a time variable. Suppose an experiment begins at $t = 0$. Events of a particular kind occur randomly, the first at T_I, the second at T_2, and so on. The random variable T_i denotes the time at which the ith event occurs, and the values t_i of T_i $(i = 1,2,\ldots)$ are called points of occurrence.

Definition: (An Interarrival process)

Let $Z_n = T_n - T_{n-1}$ and $T_o = 0$. Then Z_n denotes the time between the $(n-1)th$ and the nth events. The sequence of ordered random variables $\{Z_n, n \geq 1\}$ called ***an interarrival process*****.** Figure below shows a possible realization and the corresponding sample function of interarrival process.



Definition: (Renewal process)

If all random variables Z_n are independent and identically distributed, then $\{Z_n, n \geq 1\}$ is called ***a renewal process or a recurrent process***.

Definition: (Arrival process)

If $T_n = Z_1 + Z_2 + \cdots + Z_n$ where T_n denotes the time from the beginning until the occurrence of the nth event. Thus, $\{T_n, n \geq 0\}$ is called ***an arrival process***.

Counting process

In some problems, we count the occurrences of some types of events. In such scenarios, we are dealing with a counting process. For example, you might have a random process $N(t)$ that shows the number of customers who arrive at a supermarket by time t starting from time 0. For such a processes, we usually assume $N(0) = 0$, so as time passes and customers arrive, $N(t)$ takes positive integer values.

Definition

A ransom process $\{N(t): t \in [0, \infty)\}$ is a counting process if $N(t)$ represents the total number of "events" that have occurred in the interval $(0, t)$.

Examples: - Number of persons entering a store before time t

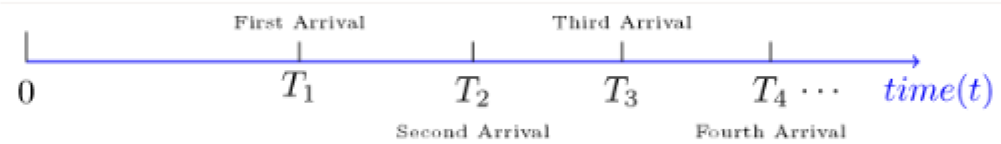
- Number of people who were born by time t

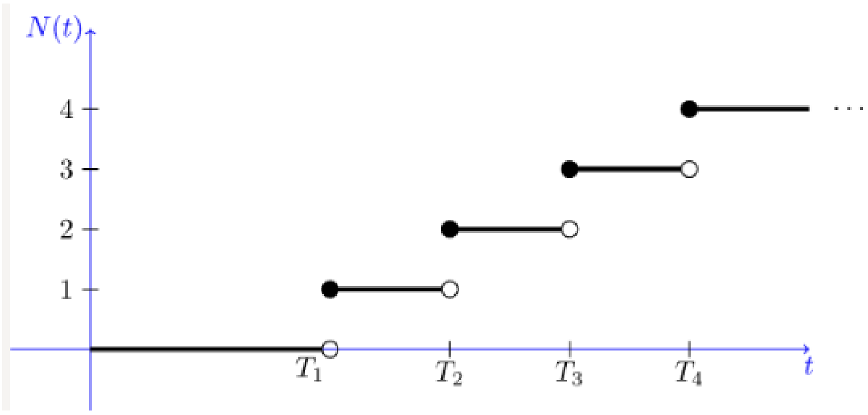
- Number of goals a soccer player scores by time t.

$N(t)$ should satisfy:

1. $N(t) \geq 0$ and $N(0) = 0$
2. $N(t)$ is integer valued that is, $N(t) \in \{0,1,2,\dots\}$, for all $t \in [0,\infty)$
3. If $s < t$, then $N(s) \leq N(t)$
4. For $0 \leq s < t$, $N(t) - N(s)$ equals the number of events that have occurred on the interval (s,t).

Since counting processes have been used to model arrivals (such as the supermarket example above), we usually refer to the occurrence of each event as an "arrival". For example, if $N(t)$is the number of accidents in a city up to time t, we still refer to each accident as an arrival. Figure below shows a possible realization and the corresponding sample function of a counting process $N(t)$.





By the above definition, the only sources of randomness are the arrival times T_i.

Definition: (Independent Increment)

Let $\{X(t): t \in [0,\infty)\}$ be a continuous-time random process, we say that $X(t)$ has **independent increment** if, for all $0 \leq t_1 < t_2 < t_3 < \cdots < t_n$ the random variables

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$$

are independent.

Note that for a **counting process**, $N(t_i) - N(t_{i-1})$ is the number of arrivals in the interval $(t_{i-1}, t_i]$.Thus, a counting process has independent increments if the numbers of arrivals in non-overlapping (disjoint) intervals

$$(t_1, t_2], (t_2, t_3], \dots, (t_{n-1}, t_n].$$

are independent.

A counting process has independent increments if the numbers of arrivals in non-overlapping (disjoint) intervals are independent.

Having independent increments simplifies analysis of a counting process. For example, suppose that we would like to find the probability of having 2 arrivals in the interval (1,2], and 3 arrivals in the interval (3,5]. Since the two intervals (1,2] and (3,5] are disjoint, we can write

$$P(2 \text{ arrivals in } (1,2], \text{and } 3 \text{ arrivals in } (3,5])$$
$$= P(2 \text{ arrivals in } (1,2]) \times P(3 \text{ arrivals in } (3,5])$$

Definition: (Stationary Increment)

Let $\{X(t): t \in [0, \infty)\}$ be a continuous-time random process, we say that $X(t)$ has stationary increment if, for all $t_2 > t_1 \geq 0$ and all $r > 0$ the two random variables $X(t_2) - X(t_1)$ and $X(t_2 + r) - X(t_1 + r)$, have the same distributions. In other words, the distribution of the difference depends only on the length of the interval $(t_1, t_2]$, and not on the exact location of the interval on the real line.

A counting process has stationary increments if, for all $t_2 > t_1 \geq 0$, $N(t_2) - N(t_1)$ has the same distribution as $N(t_2 - t_1)$.

Lesson 10: Poisson Process

One of the most important types of counting processes is the Poisson process (or Poisson counting process), it is usually used in scenarios where we are counting the occurrences of a certain events that appear to happen at a certain rate, but completely at random (without a certain structure). For example, suppose that from historical data, we know that earthquakes occur in a certain area with a rate of 2 per month. Other than this information, the timings of earthquakes seem to be completely random.

Further examples	- The number of car accidents at a site or in an area. - The location of users in a wireless network. - The requests for individual documents on a web server. - The outbreak of wars.

Poisson Random Variable

A discrete random variable X is said to be a Poisson random variable with parameter λ, shown as $X \sim Poisson(\lambda)$, if its range is $R_X = \{0,1,2,3,\dots\}$, and its pmf is given by

$$P_X(k) = \begin{cases} \frac{e^{-\lambda}\mu^{\lambda}}{k!} & \text{for } k \in R_X \\ 0 & \text{otherwise} \end{cases}$$

- if $X \sim Poisson(\lambda)$, then $E[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$
- if $X_i \sim Poisson(\lambda_i)$, for $i = 1,2,\dots,n$ and the $\mathrm{X_i}'s$ are independent, then

$$\mathrm{X}_1 + X_2 + \cdots + X_n \sim Poisson(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$

Definition (Poisson Process)

The counting process $\{N(t): t \in [0,\infty)\}$ is said to be a Poisson process having rate (intensity) $\lambda(> 0)$ if:

1. $N(t) = 0$
2. $N(t)$ has independent increments.
3. The number of arrivals in any interval of length $\tau > 0$ has $Poisson(\lambda\tau)$ distribution with mean $\lambda\tau$.

It follows from condition 3 that a Poisson process has stationary increments and that

$$E[X_t] = \lambda\tau \quad \text{and} \quad Var[X_t] = \lambda\tau$$

We conclude that in a Poisson process, the distribution of the number of arrivals in any interval depends only on the length of the interval and not on the exact location of the interval on the real line.

Result. Let $P_n(\tau)$ be the probability that exactly n events occur in an interval of length τ, namely,

$P_n(\tau) = P(N(\tau) = n)$. We have, for each $n \in \mathbb{N}, \tau \in [0, \infty)$

$$P_n(\tau) = \frac{e^{-\lambda\tau}\,(\lambda\tau)^n}{n!} \qquad n = 0,1,\dots$$

Example.

The number of customers arriving at a grocery store can be modelled by a Poisson process with intensity λ=10 customers per hour.

1. Find the probability that there are 2 customers between 10:00 and 10:20.
2. Find the probability that there are 3 customers between 10:00 and 10:20 and 7 customers between 10:20 and 11.

Solution

1. Here λ=10 and the interval between 10:00 and 10:20 has length $\tau = \frac{1}{3}$ hours. Thus, if X is the number of arrivals in that interval, we can write $X \sim Poisson(\frac{10}{3})$. Therefore,

$$P(X = 2) = \frac{e^{-\frac{10}{3}}\left(\frac{10}{3}\right)^2}{2!}$$
$$= 0.198$$

2. Here we have non-overlapping interval I_1=(10:00 , 10:20] and I_2=(10:20 , 11:00]. Thus, we can write

$$P(3 \text{ arrivals in } \mathrm{I}_1, \text{and } 7 \text{ arrivals in } \mathrm{I}_2)$$
$$= P(3 \text{ arrivals in } \mathrm{I}_1) \times P(7 \text{ arrivals in } \mathrm{I}_2)$$

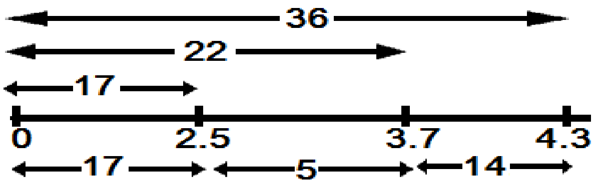
Since the lengths of the two intervals $\tau_1 = \frac{1}{3}$ and $\tau_2 = \frac{2}{3}$ respectively, we obtain $\lambda\tau_1 = \frac{10}{3}$ and $\tau_2 = \frac{20}{3}$. Thus, we have

$$P(3 \text{ arrivals in } \mathrm{I}_1, \text{and } 7 \text{ arrivals in } \mathrm{I}_2) = \frac{e^{-\frac{10}{3}}\left(\frac{10}{3}\right)^3}{3!}\;\frac{e^{-\frac{20}{3}}\left(\frac{20}{3}\right)^7}{7!}$$
$$= (0.220)(0.148)$$
$$= 0.0325$$

Example.

Suppose the process $\{N(t): t \in [0,\infty)\}$ be a Poisson process having rate $\lambda = 8$. Find $\text{P}\{\, N(2.5) = 17, N(3.7) = 22, N(4.3) = 36\}$.

Solution:



We have, $P\{\, N(2.5) = 17, N(3.7) - N(2.5) = 5, N(4.3) - N(3.7) = 14\}$.

From independent increments properties we notice that the r.v.'s $N(2.5), N(3.7) - N(2.5), N(4.3) - N(3.7)$ are independents, according to stationary properties the r.v.'s follow Poisson distribution with the parameters $8 \times 2.5 = 20, 8 \times (3.7 - 2.5) = 9.6, and\ 8 \times (4.3 - 3.7) = 4.8$ respectively. Therefore,

$$\text{P}\{\, N(2.5) = 17, N(3.7) = 22, N(4.3) = 36\}$$

$$= e^{-20}\frac{(20)^{17}}{17!} \times e^{-9.6}\frac{(9.6)^5}{5!} \times e^{-4.8}\frac{(4.8)^{14}}{14!}$$

$$= 0.076 \times 0.046 \times 0.0003$$

$$= 0.00000114$$

Second definition for Poisson process:

Let $N(t)$ be a Poisson process with rate λ. Consider a very short interval of length Δ. Then, the number of arrivals in this interval has the same distribution as $N(\Delta)$. In particular, we can write

$$P(N(\Delta) = 0) = \frac{e^{-\lambda\Delta}\,(\lambda\Delta)^0}{0!} = e^{-\lambda\Delta}$$

$$= 1 - \lambda\Delta + \frac{\lambda^2}{2}\,\Delta^2 \ldots \text{(Taylor Series)}$$

Note that if Δ is small, the terms that include second or higher powers of Δ are negligible compared to Δ. We write this as

$$P(N(\Delta) = 0) = 1 - \lambda\Delta + \text{o}(\Delta)$$

$P(N(\Delta) = 0)$ is the probability that no event occurs in the interval Δ.

Where $\text{o}(\Delta)$ is a function of Δ which goes to zero faster than does Δ; that is,

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

The letter of Omicron (o) was originally used in mathematics as a symbol for Big O notation, representing the asymptotic rate of growth of a function.

Now, let us look at the probability of having one arrival in an interval of length Δ.

$$P(N(\Delta)=1)=\frac{e^{-\lambda\Delta}\,(\lambda\Delta)^1}{1!}=\lambda\Delta\,e^{-\lambda\Delta}$$

$$=\lambda\Delta(1-\lambda\Delta+\frac{\lambda^2}{2}\,\Delta^2\,\dots)\text{ (Taylor Series)}$$

$$=\lambda\Delta+\left(-\lambda^2\Delta^2+\frac{\lambda^3}{2}\,\Delta^3\,\dots\right)$$

$$=\lambda\Delta+\mathrm{o}(\Delta)$$

We conclude that

$$P(N(\Delta)=1)=\lambda\Delta+\mathrm{o}(\Delta)$$

Similarly,

$$P(N(\Delta)\geq 2)=\mathrm{o}(\Delta)$$

Definition

The counting process $\{N(t): t\in[0,\infty)\}$ is said to be a Poisson process having rate (intensity) $\lambda(>0)$ if:

1. $N(0)=0$
2. $N(t)$ has independent and stationary increments.
3. We have

$$P(N(\Delta)=0)=1-\lambda\Delta+\mathrm{o}(\Delta)$$
$$P(N(\Delta)=1)=\lambda\Delta+\mathrm{o}(\Delta)$$
$$P(N(\Delta)\geq 2)=\mathrm{o}(\Delta)$$

Distribution of Interarrival times

Exponential Distribution

It is often used to model the time elapsed between events. A continuous random variable X is said to have an exponential distribution with parameter λ, shown as $X\sim Exp(\lambda)$, if its probability density function is of the form

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{for } x>0\\ 0 & \text{otherwise}\end{cases}$$

An CDF is given as

$F_X(t)=P(X\leq t)=\begin{cases}1-e^{-\lambda t}, & t>0\\ 0, & \text{otherwise}\end{cases}$ implies that $P(X>t)=e^{-\lambda t}$

- if $X\sim Exp(\lambda)$, then $E[X]=1/\lambda$ and $\mathrm{Var}[X]=1/\lambda^2$

Connection between a Poisson process and the exponential distributions

There is actually a strong connection between a Poisson process and the exponential distribution.

Let $N(\mathrm{t})$ be a Poisson process with rate . Let Z_1 be the time of the first arrival. Then,

$$P(Z_1 > t) = P(\text{no arrival in } (0, t]) = \frac{e^{-\lambda t}\,(\lambda t)^0}{0!} = e^{-\lambda \mathrm{t}}$$

We conclude

$$F_{Z_1}(t) = P(Z_1 > t) = 1 - e^{-\lambda \mathrm{t}}$$

Therefore, $Z_1 \sim \text{Exponential}(\lambda)$. Let Z_2 be the time elapsed between the first and the second arrival.

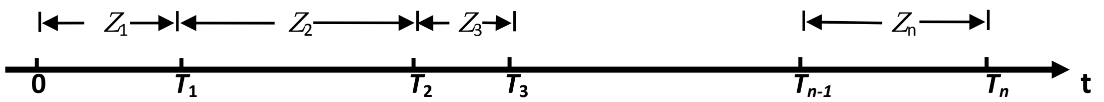


Figure: The random variables $Z_1, Z_2, \ldots$ are called the interarrival times of the counting process $N(\mathrm{t})$.

Let $s > 0$ and $t > 0$. Note that the two intervals $(0, s]$ and $[s, s+t]$ are independent. We can write

$$P(Z_2 > t \mid Z_1 = s) = P(\text{no arrival in } [s, s+t] \mid Z_1 = s)$$

$$= P(\text{no arrival in } [s, s+t]) \text{ (independent increments)} = e^{-\lambda \mathrm{t}}$$

We conclude that $Z_2 \sim \text{Exponential}(\lambda)$, and that Z_1 and Z_2 are independent. The random variables $Z_1, Z_2, \ldots$ are called the interarrival times of the counting process $N(\mathrm{t})$. Similarly, we can argue that all $Z_i's$ are independent and $Z_i \sim \text{Exponential}(\lambda)$ for $i = 1,2,3, \ldots$

> *Interarrival time for Poisson process:*
>
> Let $N(\mathrm{t})$ be a Poisson process with rate λ. Then the interarrival time $Z_1, Z_2, \ldots$ are independent and
>
> $$Z_i \sim \text{Exponential}(\lambda), \qquad \text{for } i = 1,2,3, \ldots$$

Remember that if X is exponential with parameter $\lambda > 0$, then X is a memoryless random variable, that is

$$P(X > x + a \mid X > a) = P(X > x) \qquad \text{for } a, x \geq 0$$

Thinking of the Poisson process, the memoryless property of the interarrival times is consistent with the independent increment property of the Poisson distribution. In some sense, both are implying that the number of arrivals in non-overlapping intervals are independent.

Example

Let $N(\mathrm{t})$ be a Poisson process with intensity $\lambda = 2$, and let $Z_1, Z_2, \ldots$ be the corresponding interarrival times.

1. Find the probability that the first arrival occurs after $t = 0.5$, i.e., $P(Z_1 > 0.5)$.
2. Given that we have had no arrivals before $t = 1$, find $P(Z_1 > 3)$.
3. Given that the third arrival occurred at time $t = 2$, find the probability that the fourth arrival occurs after $t = 2$.

Solution

1. $Z_1 \sim \mathrm{Exp}(2)$, we can write

$$P(Z_1 > 0.5) = e^{-(2\times 0.5)} \approx 0.37$$

another to solve this is to note that

$$P(Z_1 > 0.5) = P(\text{no arrival in } (0{,}0.5]) = e^{-(2\times 0.5)} \approx 0.37$$

2. we can write

$$P(\,Z_1 > 3 \mid Z_1 > 1\,) = P(Z_1 > 2).. \qquad \text{(memoryless property)}$$

$$e^{-(2\times 2)} \approx 0.0183$$

another to solve this is to note that the number of arrivals in (1,3] is independent of the arrivals before $t = 1$. Thus

$$P(\,Z_1 > 3 \mid Z_1 > 1\,) = P(\text{no arrival in } (1{,}3] \mid \text{no arrival in } (0{,}1])$$

$$= P(\text{no arrival in } (1{,}3]) = e^{-(2\times 2)} \approx 0.0183$$

3. the time between the third and the fourth arrival is $Z_4 \sim Exp(2)$ we can write

$$P(\,Z_4 > 2 \mid Z_1 + Z_2 + Z_3 = 2\,) = P(Z_1 > 2) \qquad \text{(independent of the Z's)}$$

$$= e^{-(2\times 2)} \approx 0.0183$$

Distribution of arrival times

Gamma Distribution

A continuous random variable X is said to have a gamma distribution with parameters $\alpha > 0$, and $\lambda > 0$ shown as $X \sim \text{gamma}(\alpha, \lambda)$, if its probability density function is of the form

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma\alpha} & \text{for} \quad x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- if $X \sim \text{gamma}(\alpha, \lambda)$, then $E[X] = \alpha/\lambda$ and $\text{Var}[X] = \alpha/\lambda^2$
- if we let $\alpha = 1$ we get $X \sim Exp(\lambda)$
- $\Gamma n = (n-1)!$

We know that If $T_n = Z_1 + Z_2 + \cdots + Z_n$ where T_n denotes the time from the beginning until the occurrence of the nth event. Thus, $\{T_n, n \geq 0\}$ is called ***an arrival process***. So T_n is the sum of n independent $Exp(\lambda)$ random variables.

Theorem. If $T_n = Z_1 + Z_2 + \cdots + Z_n$, where the Z_i's are independent $Exp(\lambda)$ random variables, then $T_n \sim \text{gamma}(n, \lambda)$.

The gamma distribution also called Erlang distribution, i.e, we can write

$$T_n \sim \text{Erlang}(n, \lambda) = T_n \sim \text{gamma}(n, \lambda) \quad \text{for } n = 1,2,3, \ldots$$

The PDF of T_n for $n = 1,2,3, \ldots$ is given by

$$f_{T_n}(t) = \begin{cases} \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} & \text{for} \quad t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Arrival time for Poisson process:

Let $N(\mathrm{t})$ be a Poisson process with rate λ. Then the interarrival time $T_1, T_2, \ldots$ have $\text{gamma}(n, \lambda)$ distribution. In particular for $n = 1,2,3, \ldots$, we have

$$E(T_n) = \frac{\mathrm{n}}{\lambda}, \quad \text{and} \quad Var(T_n) = \frac{\mathrm{n}}{\lambda^2}$$

Solved Problems (3)

Problem 1

Suppose we know that a receptionist receives an average of 15 phone calls per hour.

a) What is the probability that he will receive at least two calls between 8 and 8:12 am.
b) If the receptionist absents for 10 minutes what is the probability that no call has been lost.

Problem 2

Consider the failures of a link in a communication network. Failures occur according to a Poisson process with rate 2.4 per day. Find:

(i) Probability of time between failure greater than t days
(ii) Probability of time between failure less than t days
(iii) Probability of k failures in t days
(iv) Probability of 0 failures in next day.

Problem 3

Damages occur in a connection wire under the ground follow Poisson process at rate of $\lambda = 0.1$ per mile.

a) What is the probability that no damages in the first 2 miles.
b) In condition of no damages in the first 2 miles, what is the probability that no damages between the 2^{nd} and 3^{rd} miles.

Problem 4

Suppose the process $\{X_t : t \geq 0\}$ be a Poisson process having rate $\lambda = 2$. Find:

1. $\mathrm{P}\{\mathrm{X}_2 = 4, \mathrm{X}_5 = 12, \mathrm{X}_9 = 16\}$.
2. $\mathrm{P}\{\mathrm{X}_{1.5} = 10, \mathrm{X}_{3.5} = 18, \mathrm{X}_5 = 20\}$.

Solutions (3)

Problem 1: (Solution)

(a) Suppose X is a random variable associated to the number of calls received between 8 and 8:12 am, hence, X is follows Poisson distribution with mean $\frac{12\times 15}{60} = 3$ then,

$$P\{X \geq 2\} = 1 - (P\{X = 0\} + P\{X = 1\})$$
$$= 1 - (e^{-3} + 3e^{-3}) \quad = 1 - 4e^{-3}$$

(b) Suppose Y is a random variable associated to the number of calls received within 10 minutes, hence, Y is follows Poisson distribution with mean $\frac{10\times 15}{60} = 2.5$ then the probability that no calls have been received within 10 minutes is,

$$P\{Y = 0\} = e^{-2.5}\frac{(2.5)^0}{0!} = e^{-2.5}$$

Problem 2: (Solution)

(i) $P(\text{time between failures} > \text{ t days}) = e^{-2.4t}$

(ii) $P(\text{time between failures} < \text{ t days}) = 1 - e^{-2.4t}$

(iii) $P(k \text{ failures in } t \text{ days}) = \frac{(2.4t)^k}{k!}e^{-2.4t}$

(iv) $P(0 \text{ failures in next day}) = e^{-2.4}$

Problem 3: (Solution)

Suppose $N(t)$ a number of damages occur until mile t . then,

a) The random variable $N(2)$ follows Poisson distribution with rate parameter $0.1 \times 2 = 0.2$,hence

$$P\{N(2) = 0\} = e^{-0.2} = 0.8187$$

b) Since the two random variables $N(3) - N(2)$ and $N(2) - N(0)$ are independent, therefore, the conditional probability and unconditional probability are equivalent,then

$$P\{N(3) - N(2) = 0\} = e^{-0.1} = 0.9048$$

Problem 4: (Solution)

1. $P\{X_2 = 4, X_5 = 12, X_9 = 16\} = P\{X_2 = 4, X_5 - X_2 = 8, X_9 - X_5 = 4\}$

equivalent to

$$P[N(2) = 4, N(5) = 12, N(9) = 16] = P[N(2) = 4, N(5) - N(2) = 8, N(9) - N(5) = 4]$$

$$= e^{-4}\frac{(4)^4}{4!} \times e^{-6}\frac{(6)^8}{8!} \times e^{-8}\frac{(8)^4}{4!}$$

$$= \frac{(4)^4}{4!}\frac{(6)^8}{8!}\frac{(8)^4}{4!}e^{-18}$$

$$= 0.001155$$

2. $P\{X_{1.5} = 10, X_{3.5} = 18, X_5 = 20\}$

$$= P\{X_{1.5} = 10, X_{3.5} - X_{1.5} = 8, X_5 - X_{3.5} = 2\}$$

$$= P\{X_{1.5} = 10, X_2 = 8, X_{1.5} = 2\}$$

$$= e^{-3}\frac{(3)^{10}}{10!} \times e^{-4}\frac{(4)^8}{8!} \times e^{-3}\frac{(3)^2}{2!}$$

$$= \frac{(3)^{10}}{10!}\frac{(4)^8}{8!}\frac{(3)^2}{2!}e^{-10}$$

$$= 0.00000540352$$

Branching Processes

Lesson 11: Simple Branching Process

Lesson 11: Branching Processes

Consider some sort of population consisting of reproducing individuals.

Examples:

Living things (animals, plants, bacteria, royal families); diseases; computer viruses; rumours, gossip, lies (one lie always leads to another!)

Start conditions: start at time $n = 0$, with a single individual.

Each individual: lives for 1 unit of time. At time $n = 1$, it produces a family of Offspring, and immediately dies.

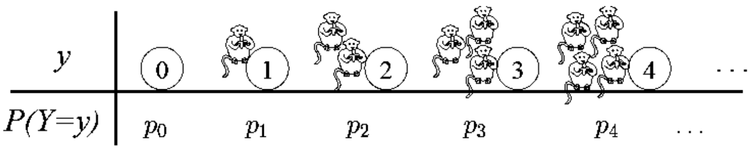
How many offspring? Could be 0, 1, 2, ... This is the family size, . (Y stands for number of Young).

Each offspring: lives for 1 unit of time. At time n = 2, it produces its own family of offspring, and immediately dies. and so on. . .

Assumptions

1. All individuals reproduce independently of each other.
2. The family sizes of different individuals are independent, identically distributed random variables. Denote the family size by Y (number of Young).

<u>*Family size distribution*, Y $P(Y = k) = p_k$</u>



Definition:

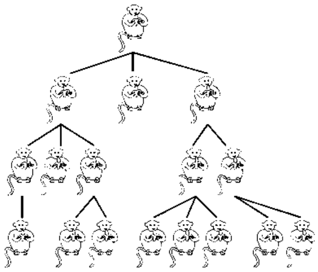
A branching process is defined as follows.

- Single individual at time $n = 0$.
- Every individual lives exactly one unit of time, then produces Y offspring and dies.
- The number of offspring Y takes values 0, 1, 2, . . . , and the probability of producing k offspring is $P(Y = k) = p_k$.
- All individuals reproduce independently. Individuals $1, 2, \dots, n$ have family sizes $Y_1, Y_2, \dots Y_n$, where each Y_i has the same distribution as Y .
- Let Z_n be the number of individuals born at time n, for $n = 0, 1, 2, \ldots$ Interpret Z_n as the size of generation n.
- Then the branching process is $\{Z_0, Z_1, Z_2, Z_3, \dots\} = \{Z_n , n \in \mathbb{N}\}$.

Definition:

The **state** of the branching process at time n is z_n, where each z_n can take values $0, 1, 2, 3, \ldots$ Note that $z_0 = 1$ always. z_n represents the size of the population at time n.

Branching Process



Analysing the Branching Process

Z_n as a randomly stopped sum:

Consider the following

The population size at time $n-1$ is given by Z_{n-1}.

Label the individuals at time $n-1$ as $1, 2, 3, \ldots, Z_{n-1}$.

Each individual $1, 2, 3, \ldots, Z_{n-1}$ starts a new branching process. Let $Y_1, Y_2, \ldots, Y_{Z_{n-1}}$ be the random family sizes of the individuals $1, 2, 3, \ldots, Z_{n-1}$.

The number of individuals at time n, Z_n, is equal to the total number of offspring of the individuals $1, 2, 3, \ldots, Z_{n-1}$. That is,

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_i$$

Thus Z_n is a randomly stopped sum: a sum of $Y_1, Y_2, \ldots$, randomly stopped by the random variable Z_{n-1}.

Note:

1. Each $Y_i \sim Y$: that is, each individual $i = 1, \ldots, Z_{n-1}$ has the same family size distribution.

2. $Y_1, Y_2, \ldots, Y_{Z_{n-1}}$ are independent.

Probability Generating Function of Z_n

Theorem. Let $G_Y(s) = E(s^Y) = \sum_{y=0}^{\infty} p_y s^y$ be the PGF of the family size distribution Y. Let $Z_0 = 1$ (start from a single individual at time 0), and let Z_n be the population size at time n $(n = 0, 1, 2, \ldots)$. Let $G_n(s)$ be the PGF of the random variable Z_n. Then

$$G_n(s) = \underbrace{\left(G\left(G\left(G\left(G(\ldots\ldots G(s)\ldots)\right)\right)\right)\right)}_{n\ times}$$

Proof. Let $G_Y(s) = E(s^Y)$ be the probability generating function of Y .

(Recall that Y is the number of Young of an individual: the family size.)

Now Z_n is a randomly stopped sum: it is the sum of $Y_1, Y_2, \ldots$ stopped by the random variable Z_{n-1}. So we can use to Theorem below to express the PGF of Z_n directly in terms of the PGFs of Y and Z_{n-1}.

Theorem: if $Z_n = Y_1 + Y_2 + \ldots + Y_{Zn-1}$, and $Zn - 1$ is itself random, then the PGF of Z_n is given by:

$$G_{Z_n}(s) = G_{Z_{n-1}}\big(G_Y(s)\big) \qquad (*)$$

where $G_{Z_{n-1}}$ is the PGF of the random variable Z_{n-1}.

For ease of notation, we can write:

$$G_{Z_n}(s) = G_n(s)\ ,\ G_{Z_{n-1}}(s) = G_{n-1}(s), \text{ and so on.}$$

Note that $Z_1 = Y$ (the number of individuals born at time $n = 1$), so we can also write: $G_Y(s) = G_1(s) = G(s)$ (for simplicity). Thus, from (*),

$$G_n(s) = G_{n-1}\big(G(s)\big)$$

Note:

1. $G_n(s) = E(s^{Z_n})$ the PGF of the population size at time n, Z_n.
2. $G_{n-1}(s) = E(s^{Z_{n-1}})$ the PGF of the population size at time $n-1, Z_{n-1}$.
3. $G(s) = E(s^Y) = E(s^{Z_1})$ the PGF of the family size Y.

We are trying to find the PGF of Z_n, the population size at time n.

So far, we have:

$$G_n(s) = G_{n-1}\big(G(s)\big) \qquad (**)$$

But by the same argument,

$$G_{n-1}(r) = G_{n-2}\big(G(r)\big)$$

(use r instead of s to avoid confusion in the next line.)

Substituting in (**),

$$\begin{aligned} G_n(s) &= G_{n-1}\big(G(s)\big) \\ &= G_{n-1}(r) \quad \text{where } r = G(s) \\ &= G_{n-2}\big(G(r)\big) \\ &= G_{n-2}(G(G(s))) \ \text{ replacing } r = G(s) \end{aligned}$$

By the same reasoning, we will obtain:

$$G_n(s) = \underbrace{G_{n-3}}_{3}\underbrace{\left(G\left(G\big(G(s)\big)\right)\right)}_{3\text{ times}}$$

and so on, until we finally get

$$G_n(s) = \underbrace{G_{n-(n-1)}}_{n-1}\underbrace{\left(G\left(G\left(G(\ldots\ldots G(s)\ldots)\right)\right)\right)}_{\text{n−1 times}}$$

$$= \underbrace{G_1}_{G}\underbrace{\left(G\left(G\left(G(\ldots\ldots G(s)\ldots)\right)\right)\right)}_{\text{n−1 times}}$$

$$= \underbrace{\left(G\left(G\left(G\left(G(\ldots\ldots G(s)\ldots)\right)\right)\right)\right)}_{\text{n times}}$$

Mean of Z_n

Theorem. $\{Z_0, Z_1, Z_2, Z_3, \ldots\}$ be a branching process with $Z_0 = 1$ (start with a single individual). Let Y denote the family size distribution, and suppose that $\mathrm{E}(Y) = \mu$. Then

$$E(Z_n) = \mu^n$$

Proof. We know that $Z_n = Y_1 + Y_2 + \ldots + Y_{Zn-1}$ is a randomly stopped sum:

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_i$$

$$E(Z_n) = E\left(\sum_{i=1}^{Z_{n-1}} Y_i\right) = E(Y_i) \times E(Z_{n-1})$$

$$= \mu\, E(Z_{n-1}) = \mu\left(\mu\, E(Z_{n-2})\right)$$

$$= \mu^2\, E(Z_{n-2})$$

$$\vdots$$

$$= \mu^{n-1}\, E(Z)$$

$$= \mu^{n-1}\,\mu = \mu^n$$

Example. Let family size ~Geometric$(p = 0.3)$. So $\mu = E(Y) = \frac{q}{p} = \frac{0.7}{0.3} = 2.33$.

Expected population size by generation $n = 10$ is:

$$E(Z_{10}) = \mu^{10} = (2.33)^{10} = 4784$$

Example. Let family size ~Geometric$(p = 0.5)$. So $\mu = E(Y) = \frac{q}{p} = \frac{0.5}{0.5} = 1$.

$$E(Z_{10}) = \mu^{10} = (1)^{10} = 1$$

Variance of Z_n

Theorem. $\{Z_0, Z_1, Z_2, Z_3, \ldots\}$ be a branching process with $Z_0 = 1$ (start with a single individual). Let Y denote the family size distribution, and suppose that $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$ Then

$$\mathrm{Var}(Z_n) = \begin{cases} \sigma^2 n & \text{if } \mu = 1 \\ \sigma^2\mu^{n-1}\left(\frac{1-\mu^n}{1-\mu}\right) & \text{if } \mu \neq 1 \text{ or } 1 < \mu < 1 \end{cases}$$

Example. Let family size ~Geometric$(p = 0.3)$. So $\mu = E(Y) = \frac{q}{p} = \frac{0.7}{0.3} = 2.33$.

$$\sigma^2 = \text{Var}(Y) = \frac{q}{p^2} = \frac{0.7}{(0.3)^2} = 7.78$$

Since $\mu \neq 1$ is:

$$Var(Z_{10}) = \sigma^2\mu^9\left(\frac{1-\mu^{10}}{1-\mu}\right) = (7.78)^2(2.33)^9\left(\frac{1-(2.33)^{10}}{1-2.33}\right) = 5.72 \times 10^7$$

Example. Let family size ~Geometric$(p = 0.5)$. So $\mu = 1, \sigma^2 = \frac{q}{p^2} = \frac{0.5}{(0.5)^2} = 2$

Since $\mu = 1$ is:

$$Var(Z_{10}) = \sigma^2 n = 2 \times 10 = 20$$