



JUN 2019

UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE CIENCIAS NATURALES Y
MATEMÁTICA
UNIDAD DE INVESTIGACION



INFORME FINAL DEL TEXTO


**“TEXTO: UNA INTRODUCCIÓN A LOS PROCESOS
ESTOCÁSTICOS”**

AUTOR: HERMINIA BERTHA TELLO BEDRIÑANA

(PERIODO DE EJECUCIÓN: Del 01/05/2017 al 30/04/2019)

(Resolución de aprobación N° 487-2017-R.- CALLAO)

Callao, 2019



I. ÍNDICE

I. ÍNDICE.....	1
TABLA DE CONTENIDO.....	4
II. PRÓLOGO	7
III. INTRODUCCIÓN.....	8
IV. CUERPO DEL TEXTO O CONTENIDO.....	9
CAPÍTULO I.....	10
PROBABILIDAD - VARIABLE ALEATORIA - VECTOR ALEATORIO.10	
1.1. Probabilidad.....	10
1.1.1. Causalidad y Aleatoriedad.....	10
1.1.2. Experimento, resultado, espacio muestral y suceso	10
1.1.3. Probabilidad y sus propiedades	13
1.1.4. Probabilidad condicionada. Teorema de Bayes	16
1.1.5. Independencia	20
1.2. Variable Aleatoria	23
1.2.1. Definición.....	23
1.2.2. Probabilidad inducida.....	24
1.2.3. Función de distribución de probabilidad.....	25
1.2.4. Función de cuantía o probabilidad: variable aleatoria discreta.....	26
1.3. Vector aleatorio	27
1.3.1. Probabilidad inducida	28
1.3.2. Funciones de distribución conjunta y marginales	28
CAPÍTULO II	29
ESPERANZA. DESIGUALDADES - FUNCIÓN CARACTERÍSTICA.....	29
2.1. Esperanza de una variable aleatoria.	29
2.1.1. Momentos de una variable aleatoria	30
2.1.2. Momentos de algunas variables aleatorias conocidas.....	32
2.2. Esperanza de un vector aleatorio.....	34
2.2.1. Momentos de un vector aleatorio.....	35
2.2.2. Covarianza. Aplicaciones.....	36
2.3. Esperanza condicionada	42
2.4. Desigualdades.....	47



2.4.1.	La distribución Normal multivariante	50
2.5.	Función característica	52
2.5.1.	Función característica e independiente	54
2.5.2.	Funciones características de algunas distribuciones conocidas	54
2.5.3.	Teorema de inversión. Unicidad	56
2.5.4.	Teorema de continuidad de Lévy	58
CAPÍTULO III.....		59
SUCESIONES DE VARIABLES ALEATORIAS - TEOREMAS DE		
CONVERGENCIA		
3.1.	Introducción.....	59
3.2.	Tipos de convergencia.....	60
3.3.	Leyes de los Grandes Números	63
3.4.	Teorema Central del Límite.....	64
3.4.1.	Aplicación del TCL: Estimación del valor de π	67
CAPÍTULO IV.....		68
PROCESOS ESTOCÁSTICOS.....		
4.1.	Introducción.....	68
4.2.	Definiciones básicas y descripción de un proceso estocástico.....	69
4.2.1.	Trayectoria de un proceso	71
4.2.2.	Distribuciones Finito-Dimensionales.....	72
4.2.3.	Funciones de Momento	73
CAPÍTULO V.....		76
ALGUNOS PROCESOS ESTOCÁSTICOS DE INTERÉS		
5.1.	Procesos IID	76
5.2.	Ruido Blanco	80
5.3.	Proceso Gaussiano.....	80
5.4.	Proceso De Poisson	81
5.5.	Señal telegráfica aleatoria (RTS).....	88
5.6.	Modulación por desplazamiento de fase (PSK)	91
5.7.	Proceso de Wiener. Movimiento Browniano	93
5.8.	Cadenas de Markov	96
CAPÍTULO VI.....		99
PROCESOS ESTACIONARIOS Y TRANSFORMACIÓN LINEAL DE UN		
PROCESO ESTACIONARIO		
6.1.	Procesos estacionarios.....	99
6.1.1.	Estacionariedad en sentido amplio (WSS).....	101

A

6.2. Procesos Cicloestacionarios	106
6.3. Densidad espectral de potencia (PSD) de un proceso WSS	108
6.3.1. PSD para procesos estocásticos WSS discretos en el tiempo	109
6.3.2. PSD para procesos estocásticos WSS continuos en el tiempo	118
6.4. Estimación de la densidad espectral de potencia.....	125
6.5. Ergodicidad.....	125
V. REFERENCIALES.....	129
VI. APÉNDICES	130
VII. ANEXOS	134



TABLA DE CONTENIDO

INDICE DE FIGURAS	Página
Figura N° 1.1“Teorema De Bayes Y Sistemas De Comunicación”	22
Figura N° 4.1“Ejemplos De Los Diferentes Tipos De Procesos Estocasticos”....	70
Figura N° 4.2“Trayectorias De Un Proceso De Poisson Y Realizaciones De Las Variables N_{20} Y N_{25} ”	71
Figura N° 5.1“Gráfico La Relación Entre Las S_n Y Las X_n De Un Proceso De Poisson”.	83
Figura N° 5.2: Relación Entre Los Tiempos De Espera Y Un Intervalo Arbitrario En Un Proceso De Poisson	84
Figura N° 5.3“Realización De Un Proceso Rts Con $a=1$ Y $\lambda=1,5$ Y Su Correspondiente Función De Autocorrelación”.....	90
Figura N° 5.4"Realización De Un Proceso Psk Y Su Correspondiente Función De Autocorrelación”	92
Figura N° 6.1“Una Trayectoria Del Proceso De Pulsos Modulados”.....	107
Figura N° 6.2 “La Función De Autocovarianza Del Proceso De Pulsos Modulados”	108
Figura N° 6.3 “Espectro Y Trayectoria Del Proceso Ar(1) Para $\alpha = -0.5$ (Superior), $\alpha = 0.5$ (Central) Y $\alpha = 0.9$ (Inferior)”	112
Figura N° 6.4 “Espectros Y Funcion De Transferencia De Los Procesos Ma(3) Y Ar(1)”	115
Figura N° 6.5 “El Fenomeno De Aliasing”.....	120

Figura N° 6.6 “Densidad Espectral De Potencia Del Proceso Rts Para $\lambda = 1(\dots)$ Y $\lambda = 4(\dots)$ ”	121
Figura N° 6.7 “P(W) Con Frecuencias Acotadas Para Un Ruido Blanco (Izquierda) Y Su Correspondiente $R(\tau)$ (Derecha)”	122
Figura N° 6.8 “Efecto Del Crecimiento w_0 Sobre $R(\tau)$ ”	123
Figura N° 7.1: “Procesos Estocástico”	135
Figura N° 7.2: “Esquema De Los Procesos Estocásticos.....”	136

INDICE DE TABLAS	Página
TABLA N° 2.1 "MOMENTO DE UNA VARIABLE ALEATORIA"	30
TABLA N° 4.1 "ESTADO Y TIEMPO (DISCRETO Y CONTINUO)"	69
TABLA N° 5.1 "EJEMPLO DE CADENA DE MARKOV"	98
TABLA N° 6.1 "TIPOS DE PROCESOS ESTACIONARIOS"	105



II. PRÓLOGO

El presente texto ha sido escrito para ser una guía en el tema: “Una Introducción a los Procesos Estocásticos”, que se desarrolla en el transcurso de la formación profesional de la Escuela de Matemática de la Facultad de Ciencias Naturales y Matemática de la Universidad Nacional del Callao.

Los cálculos que se presentan son una ayuda para entender los procesos estocásticos, para que los lectores entiendan sus cómo y por qué y asimismo la interpretación de los resultados obtenidos. Deja bien claro que en ningún momento se pretende adiestrar a los lectores en cálculos, sino en que aprendan los conocimientos teóricos básicos sobre los procesos estocásticos (saber), es por eso que, en el presente informe aborda sobre temas introductorios para entender mejor sobre el este tema, aplicando los procesos y desarrollar una actitud positiva hacia el tema. Esto es, que los procesos estocásticos no solamente es cálculo o el simple uso de fórmulas o expresiones que aparecen en este y en diversos libros relacionados, si no razonamiento crítico basado en evidencias objetivas. Una vez que el lector haya asimilado los conocimientos introductorios de los procesos estocásticos y sus aplicaciones que brinda el presente texto, estará en la capacidad de abordar a mayor profundidad sobre el tema. Sin embargo, el estudiante no debe conformarse con solo conocer sobre los procesos estocásticos, también tiene que saber sus aplicaciones en la teoría de la probabilidad de la mano a su carrera profesional.

R

III. INTRODUCCIÓN

Este es un texto, como su nombre lo indica, “Una Introducción a los Procesos Estocásticos”, que abarca temas introductorios sobre los procesos estocásticos, referentes a la teoría de la probabilidad y que refleja los avances en la enseñanza donde se han incorporado características tales como: ejercicios, uso de datos reales y texto claro. Se descartan los métodos de análisis de datos, en general se tiene preferencia a los ejemplos que requieren razonamientos críticos e interpretación de resultados. Cabe indicar que este libro también puede usarse sin ninguna tecnología específica, puesto que cada uno de los temas se ha ilustrado con ejemplos resueltos detalladamente. Por fortuna los avances tecnológicos con las computadoras han hecho que los cálculos requeridos sean relativamente fáciles, lo que ha permitido privilegiar la comprensión e interpretación de los resultados. Además, tengo el privilegio como autora de abordar este tema muy útil no solo para los estudiantes de la Facultad de Ciencias Naturales y Matemática, sino, para todos los interesados en los procesos estocásticos, ya que este material introductorio a dicho tema les será muy útil.

El contenido del libro se ha dividido en 6 capítulos. En el capítulo I se presentan conceptos fundamentales de probabilidades, variable aleatoria y vector aleatorio. En el capítulo II se aborda sobre la esperanza de una variable aleatoria, vector aleatorio y condicionada; también se tocará las desigualdades y la función característica. En el capítulo III, trata sobre las sucesiones de variables aleatorias y los teoremas de convergencia. El capítulo IV aborda la introducción, definiciones básicas y descripciones de un proceso estocástico. Algunos procesos estocásticos de interés se abordarán en el capítulo V mediante el ruido blanco, proceso gaussiano y Poisson, señal telefónica aleatoria, modulación por desplazamiento de fase, proceso Wiener, movimiento browniano y cadenas de Markov. Finalmente se cerrará el presente trabajo con el capítulo VI, donde se explicará brevemente sobre los procesos estacionarios y transformación lineal de un proceso estocástico.

La importancia del texto radica en el hecho de que constituye un instrumento para facilitar el proceso enseñanza-aprendizaje, de acuerdo con los objetivos y contenidos de los programas oficiales, de los procesos estocásticos en la teoría de la probabilidad que se imparte en las carreras de ciencias. Asimismo, debido a su enfoque predominantemente analítico, prepara al alumno para el estudio de éste a nivel más avanzado.

IV. CUERPO DEL TEXTO O CONTENIDO

CAPÍTULO I

PROBABILIDAD - VARIABLE ALEATORIA - VECTOR ALEATORIO.

1.1. Probabilidad

Es una ciencia exacta que se desarrolla en forma axiomática. Esto requiere primero conocer los conceptos básicos, es decir aprender el “lenguaje de la teoría de probabilidad”, luego se dan los axiomas de la probabilidad, para después desarrollar teoremas, corolarios, ejercicios y aplicaciones.

1.1.1. Causalidad y Aleatoriedad

Se tuvo en cuenta, cuánto tiempo se tardaría en recorrer de Huancayo a Lima, si nos desplazamos con velocidad constante de 100 km/hora, es entre 4 a 5 horas. Es muy distinta si, en un lanzamiento, que cara nos mostrará un dado. Se trata de dos fenómenos de naturaleza distinta:

- *El primero* se denomina deterministas, aquellos en los que la relación causa-efecto aparece perfectamente determinada. La que se conoce por la ecuación $e = v \cdot t$.
- *El segundo* se denomina aleatorios, que se caracterizan porque aun repitiendo en las mismas condiciones el experimento que lo produce, el resultado variará de una repetición a otra dentro de un conjunto de posibles resultados.

La Teoría de la Probabilidad pretendió oponerse el trabajo que los físicos y, en general, los científicos experimentales han llevado a cabo. Para lograr entenderlo, se observó que la ecuación anterior, $e = v \cdot t$, es un resultado experimental que se verificó como un modelo matemático que, haciendo abstracción del móvil concreto y del medio en el que se desplaza, describiendo la relación existente entre el espacio, el tiempo y la velocidad. La Teoría de la Probabilidad permitió la obtención de modelos aleatorios o estocásticos mediante los cuales se conoció, en términos de probabilidad, el comportamiento de los fenómenos aleatorios.

1.1.2. Experimento, resultado, espacio muestral y suceso

El lector puede verificar sobre el segundo ejemplo mencionado, el de los dados. se sabe que la extracción al azar de una carta de una baraja pertenece a uno de los cuatro: corazón, espada, rombo o trébol. El experimento a nuestro fenómeno aleatorio dio a lugar a un resultado, ω , de entre un conjunto de posibles resultados.

Este conjunto de posibles resultados se dio el nombre de espacio muestral: Ω . Subconjuntos de resultados con una característica común, se dio el nombre de sucesos aleatorios o sucesos. Cuando el resultado del experimento pertenece al suceso A, decimos que ha ocurrido o se ha realizado A.

A continuación, mostramos ejemplos de experimentos aleatorios, los espacios muestrales asociados y algunos sucesos relacionados.

- **Lanzamiento de dos monedas.** - Al lanzar dos monedas el espacio muestral viene definido por $\Omega = \{CC, C+, +C, ++\}$. En estos, dos ejemplos de sucesos en este espacio pueden ser:

$$A = \{\text{Ha salido una cara}\} = \{C+, +C\},$$

$$B = \{\text{Ha salido más de una cruz}\} = \{+, +\}$$

- **Elegir un punto al azar en el círculo unidad.**- El espacio muestral es $\Omega = \{\text{Los puntos del círculo}\}$ Ejemplos de sucesos:

$$A = \{w; d(w, \text{centro}) < 0,5\},$$

$$B = \{w; 0,3 < d(w, \text{centro}) < 0,75\}$$

➤ **Sucesos, conjuntos y σ -álgebra de sucesos.**

Los sucesos no son más que subconjuntos de Ω , se pudo operar con ellos de acuerdo con las reglas de la teoría de conjuntos. Las operaciones entre conjuntos fueron aplicables a los sucesos y el resultado de las mismas dio lugar a nuevos sucesos cuyo significado debemos conocer. Existen, por otra parte, sucesos cuya peculiaridad e importancia que se asignó nombre propio. De estos y de aquellas nos ocupamos a continuación:

- **Suceso cierto o seguro:** Cuando se lleva a cabo cualquier experimento aleatorio es seguro que el resultado pertenecerá al espacio muestral, por lo que Ω , al darse este suceso, recibe el nombre de suceso cierto o seguro.
- **Suceso imposible:** En el extremo opuesto aparece aquel suceso que no contiene ningún resultado que designamos mediante \emptyset y que, lógicamente, no ocurre nunca, por la cual se denomina suceso imposible.

- **Sucesos complementarios:** La ocurrencia de un suceso, A , supone la no ocurrencia del suceso que contiene a los resultados que no están en A (A^c). En estos casos, los sucesos reciben el nombre de complementarios.
- **Unión de sucesos:** La unión de dos sucesos, $A \cup B$, da lugar a un nuevo suceso que no es más que el conjunto resultante de dicha unión. Donde el lector puede verificar que, $A \cup B$ ocurre cuando el resultado del experimento pertenece a A , a B o ambos a la vez.
- **Intersección de sucesos:** La intersección de dos sucesos, $A \cap B$, es un nuevo suceso cuya realización tiene lugar si el resultado pertenece a ambos a la vez, lo que supuso que ambos ocurren simultáneamente.
- **Sucesos incompatibles:** Existen sucesos que al no compartir ningún resultado su intersección es el suceso imposible, $A \cap B = \emptyset$. Se les denomina, por ello, sucesos incompatibles. Un suceso A y su complementario A^c , son un buen ejemplo de sucesos incompatibles. Al verificarse el desarrollo, se concluyó que todo subconjunto de Ω será un suceso. La noción de suceso es un concepto que surge con naturalidad en el contexto de la experimentación aleatoria pero, aunque no hubiera sido así, la necesidad del concepto nos hubiera obligado a inventarlo. De la misma forma, es necesario que los sucesos posean una mínima estructura que garantice la estabilidad de las operaciones naturales que con ellos realicemos, entendiendo por naturales la complementación, la unión y la intersección. Estas dos últimas, para precisar que no se trata de uniones e intersecciones en número cualquiera, ya que más allá de la numerabilidad nos movemos con dificultad. Pues que se nos garantice que uniones e intersecciones numerables de sucesos son estables y dan lugar a otro suceso. Existe una estructura algebraica que verifica las condiciones de estabilidad que acabamos de enumerar.

Definición 1.1 (σ -álgebra de conjuntos): Una familia de conjuntos A definida sobre Ω decimos que es una σ -álgebra si:

1. $\Omega \in A$
2. $A \in A \Rightarrow A^c \in A$
3. $\{A_n\}_{n \geq 1} \subset A \Rightarrow \bigcup_{n \geq 1} A_n \in A$



La familia de las partes de Ω , $\mathcal{P}(\Omega)$, se verifica a la definición y es por tanto una σ -álgebra de sucesos, de hecho, la más grande de las existentes. En muchas ocasiones excesivamente grande para nuestras necesidades, que vienen determinadas por el núcleo inicial de sucesos objeto de interés. Para su mejor comprensión, se elaboró los siguientes ejemplos:

Ejemplo 1.1: Si se supone, un experimento consiste en elegir al azar un número en el intervalo $[0,1]$, el interés se centrará en conocer si la elección pertenece a cualquiera de los posibles su intervalos de $[0,1]$. La σ -álgebra de sucesos generada a partir de ellos, que es la menor que los contiene, se la conoce con el nombre de σ -álgebra de Borel en $[0,1]$, $\beta_{[0,1]}$, y es estrictamente menor que $\mathcal{P}([0, 1])$.

Resumiendo, el espacio muestral se comprobó que viene acompañado de la correspondiente σ -álgebra de sucesos, la más conveniente al experimento. La pareja que ambos constituyen, (Ω, \mathcal{A}) , recibe el nombre de espacio probabilizable.

Señalemos por último que en ocasiones no es posible economizar esfuerzos y \mathcal{A} coincide con $\mathcal{P}(\Omega)$. Por ejemplo, cuando el espacio muestral es numerable.

1.1.3. Probabilidad y sus propiedades

Se conoce que la naturaleza aleatoria del experimento impide predecir de antemano el resultado que se obtuvo al llevarse a cabo. Se quiso conocer si cada suceso de la σ -álgebra se realiza o no. Responder de una forma categórica fue un poco difícil, ya que fue imposible predecir en cada realización del experimento si el resultado va a estar o no en cada suceso. En Probabilidad la pregunta se formula del siguiente modo: ¿qué posibilidad hay de que tenga lugar cada uno de los sucesos? La respuesta exige un tercer elemento que proporcione esa información: Una función de conjunto P , es decir, una función definida sobre la σ -álgebra de sucesos, que a cada uno de ellos le asocie un valor numérico que exprese la mayor o menor probabilidad o posibilidad de producirse cuando se realiza el experimento. Esta función de conjunto se conoce como medida de probabilidad o simplemente probabilidad. El concepto de probabilidad aparece ligado en sus orígenes a los juegos de azar, razón por la cual se tiene constancia del mismo desde tiempos remotos. A lo largo de la historia se han hecho muchos y muy diversos intentos para formalizarlo, dando lugar a otras tantas definiciones de probabilidad donde todas ellas de haber sido confeccionadas, careciendo de la generalidad suficiente que permitió utilizar en cualquier contexto. Por ello, supusieron sucesivos avances que

permitieron a Kolmogorov enunciar su conocida y definitiva axiomática en 1933. De entre las distintas aproximaciones, dos son las más relevantes:

- **Método Frecuencialista.** - Cuando el experimento es susceptible de ser repetido en las mismas condiciones una infinidad de veces, la probabilidad de un suceso A , $P(A)$, se define como el límite al que tiende la frecuencia relativa de ocurrencias del suceso A .
- **Método Clásico (Fórmula de Laplace).**- Si el experimento conduce a un espacio muestral finito con n resultados posibles, $\Omega = \{w_1, w_2, \dots, w_n\}$, todos ellos igualmente probables, la probabilidad de un suceso A que contiene m de estos resultados se obtiene mediante la fórmula:

$$P(A) = \frac{m}{n}.$$

Conocida como fórmula de Laplace, propuesta a finales del siglo XVIII. La fórmula se enunció como el cociente entre el número de casos favorables y el número de casos posibles. Se observó la incorrección formal de esta aproximación en la medida que exige equiprobabilidad en los resultados para poder definir, precisamente, la probabilidad, lo cual implica un conocimiento previo de aquello que se quiere definir.

Las anteriores definiciones son aplicables cuando las condiciones exigidas al experimento son satisfechas y dejan un gran número de fenómenos aleatorios fuera de su alcance. Estos problemas se soslayan con la definición axiomática propuesta por A.N. Kolmogorov en 1933:

Definición. 2 (Probabilidad) *Una función de conjunto, P , definida sobre la σ -álgebra A es una probabilidad si:*

1. $P(A) \geq 0$ para todo $A \in A$.
2. $P(\Omega) = 1$.
3. P es numerablemente aditiva, es decir, si $\{A_n\}_{n \geq 1}$ es una sucesión de sucesos disjuntos de A , entonces:

$$P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} P(A_n).$$

A la terna (Ω, A, P) se denominó: espacio de probabilidad.



Se pudo verificar que las propiedades que se derivan de esta definición, los axiomas propuestos por Kolmogorov no son más que la generalización de las propiedades que posee la frecuencia relativa. La definición se apoya en las aproximaciones previas existentes y al mismo tiempo las incluye como situaciones particulares que son.

Ejemplo 1.2 (Espacio de probabilidad discreto. La fórmula de Laplace): En un espacio muestral, Ω , numerable y como σ -álgebra la familia formada por todos los posibles subconjuntos de Ω . Sea p una función no negativa definida sobre Ω verificando: $\sum_{w \in \Omega} p(w) = 1$. Si definimos $P(A) = \sum_{w \in A} p(w)$, se pudo comprobar que P es una probabilidad.

Hay un caso particular de especial interés, el llamado espacio de probabilidad discreto uniforme, en el que Ω es finito, $\Omega = \{w_1, w_2, \dots, w_n\}$ y $p(w_i) = \frac{1}{n}, \forall w_i \in \Omega$. Entonces, para

$A = \{w_{i_1}, w_{i_2}, \dots, w_{i_m}\}$ se tiene $P(A) = \frac{m}{n}$, que es la fórmula de

Laplace. El nombre de uniforme se justifica porque la masa de probabilidad está uniformemente repartida al ser constante en cada punto.

Un ejemplo de espacio de probabilidad discreto uniforme es el que resulta de lanzar dos dados. El espacio muestral, $\Omega = \{(1,1), (1,2), \dots, (6,5), (6,6)\}$, está formado por las 6x6 posibles parejas de caras. Si los dados son correctos, cualquiera de estos resultados tiene la misma probabilidad, 1/36. Sea ahora $A = \{\text{ambas caras son pares}\}$, el número de puntos que contiene A son 9, por lo que, aplicando la fórmula de Laplace, $P(A) = 9/36 = 1/4$.

➤ Propiedades de la probabilidad

De la definición de probabilidad se dedujo algunas propiedades muy útiles.

- La probabilidad del vacío es cero. - $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \dots$ y por la aditividad numerable.

- **Aditividad finita.** - Si A_1, A_2, \dots, A_n son elementos disjuntos de \mathcal{A} , aplicando la σ -aditividad, la propiedad anterior y haciendo $A_i = \emptyset, i > n$ se tuvo:

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i \geq 1}^n A_i\right) = \sum_{i=1}^n P(A_i),$$

se dedujo que: $\forall A \in \mathcal{A}, P(A^c) = 1 - P(A)$.

- **Monotonía.** - Si $A, B \in \mathcal{A}, A \subset B$, entonces de $P(B) = P(A) + P(B - A)$ se deduce que $P(A) \leq P(B)$.
- **Probabilidad de una unión cualquiera de sucesos (fórmula de inclusión - exclusión).** - Si $A_1, A_2, \dots, A_n \in \mathcal{A}$ entonces:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n).$$

- **Subaditividad.** - Dados los sucesos A_1, A_2, \dots, A_n , la relación existente entre la probabilidad de la unión de los A_i y la probabilidad de cada uno de ellos es la siguiente:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

- **Continuidad de la probabilidad.** - Sea $\{A_n\}_{n \geq 1}$ una sucesión monótona de sucesos y sea A su límite. Se demostró que:

$$P(A) = P\left(\lim_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

1.1.4. Probabilidad condicionada. Teorema de Bayes

Si se compra un número para una rifa que se celebra anualmente durante las fiestas de verano en nuestro pueblo y que está compuesta por 100 boletos numerados del 1 al 100, se sabe que nuestra probabilidad ganar el premio, suceso que se designa por A , vale:

$$P(A) = \frac{1}{100}.$$

A la mañana siguiente de celebrarse el sorteo alguien nos informa que el boleto premiado termina en 5. Con esta información, ¿se continúa pensando que nuestra probabilidad de ganar vale 10^{-2} ? Desde luego

sería absurdo continuar pensando si el número termina en 7, porque evidentemente la nueva probabilidad valdría $P(A) = 0$, pero aunque terminara en 5 también la probabilidad de ganar habría cambiado, porque los números que terminan en 5 entre los 100 son 10 y entonces 10 veces mayor que la inicial:

$$P(A) = \frac{1}{10},$$

se supuso que nuestro número es el 35 y que este pasara a los elementos que han intervenido en la nueva situación. Se tiene, un suceso original $A = \{\text{ganar el premio con el número 35}\}$, y, un suceso $B = \{\text{el boleto premiado termina en 5}\}$, de cuya ocurrencia se nos informa *a priori*. Se observa que la nueva probabilidad encontrada verifica:

$$P(A) = \frac{1}{10} = \frac{1/100}{10/100} = \frac{P(A \cap B)}{P(B)},$$

esto puso en evidencia algo que cabía esperar, que la nueva probabilidad depende de $P(B)$. Estas propiedades observadas justificaron la definición que damos a continuación.

Definición 1.3 (Probabilidad Condicionada): Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sean A y B dos sucesos, con $P(B) > 0$, se define la probabilidad de A condicionada a B mediante la expresión:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

➤ Teorema de factorización

Al considerar la definición de probabilidad condicionada, la probabilidad de la intersección de dos sucesos puede expresarse de la forma $P(A \cap B) = P(A|B)P(B)$. El teorema de factorización extiende este resultado para cualquier intersección finita de sucesos.

Se consideró los sucesos A_1, A_2, \dots, A_n , tales que $P(\bigcap_{i=1}^{n-2} A_i) > 0$, por inducción se comprobó que:

$$P(\bigcap_{i=1}^n A_i) = P(A_n | \bigcap_{i=1}^{n-1} A_i) P(A_{n-1} | \bigcap_{i=1}^{n-2} A_i) \dots P(A_2 | A_1) P(A_1).$$



Ejemplo 1.3 En una urna que contiene 5 bolas blancas y 4 negras, llevamos a cabo 3 extracciones consecutivas sin reemplazamiento. ¿Cuál es la probabilidad de que las dos primeras sean blancas y la tercera negra?

Cada extracción altera la composición de la urna y el total de bolas que contiene. De acuerdo con ello se tiene:

$$P(B_1 \cap B_2 \cap N_3) = P(N_3 | B_1 \cap B_2)P(B_2 | B_1)P(B_1) = \frac{4}{7} \cdot \frac{4}{8} \cdot \frac{5}{9}$$

➤ **Teorema de la probabilidad total**

Si los sucesos A_1, A_2, \dots, A_n constituyen una partición del Ω , tal que $P(A_i) > 0, \forall i \forall i$, tendremos que cualquier suceso B podrá particionarse de la forma, $B = \bigcup_{i=1}^n B \cap A_i$ y tratándose de una unión disjunta se puede escribir:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B | A_i)P(A_i).$$

Este resultado se conoce con el nombre de *Teorema de la Probabilidad Total*.

➤ **Teorema de Bayes**

Puede tener interés, y de hecho así ocurre en muchas ocasiones, conocer la probabilidad asociada a cada elemento de la partición dado que ha ocurrido B , es decir, $P(A_i | B)$. Para ello, recordemos la definición de probabilidad condicionada y apliquemos el resultado anterior.

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^N P(B | A_j)P(A_j)}.$$

Este resultado, conocido como el teorema de Bayes, permitió conocer el cambio que experimenta la probabilidad de A_i como consecuencia de haber ocurrido B . En el lenguaje habitual del Cálculo de Probabilidades a $P(A_i)$ se la denomina probabilidad a priori y a $P(A_i | B)$ probabilidad a posteriori, siendo la ocurrencia de B la que establece la frontera entre el antes y el después. ¿Cuál es, a efectos prácticos, el interés de este resultado? Veámoslo con un ejemplo.



Ejemplo 1.4 Tres urnas contienen bolas blancas y negras. La composición de cada una de ellas es la siguiente:

$$U_1 = \{3B, 1N\}, U_2 = \{2B, 2N\}, U_3 = \{1B, 3N\}.$$

Se elige al azar una de las urnas, se extrae de ella una bola al azar y resulta ser blanca. ¿Cuál es la urna con mayor probabilidad de haber sido elegida? Mediante U_1, U_2, U_3 , representaremos también la urna elegida. Estos sucesos constituyen una partición de Ω y se verificó, puesto que la elección de la urna es al azar,

$$P(U_1), P(U_2), P(U_3) = \frac{1}{3}.$$

Si: $B = \{la\ bola\ extraida\ es\ blanca\}$, tendremos:

$$P(B|U_1) = \frac{3}{4}, P(B|U_2) = \frac{2}{4}, P(B|U_3) = \frac{1}{4}.$$

Lo que nos piden es obtener $P(U_i|B)$, para conocer cuál de las urnas ha originado, más probablemente, la extracción de la bola blanca.

Aplicando el teorema de Bayes a la primera de las urnas:

$$P(U_1|B) = \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{1}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{3} \cdot \frac{1}{4}} = \frac{3}{6},$$

para las otras dos:

$$P(U_2|B) = 2/6 \wedge P(U_3|B) = 1/6.$$

Luego la primera de las urnas es la que con mayor probabilidad dio lugar a una extracción de bola blanca.

El teorema de Bayes es uno de aquellos resultados que indujo a pensar que la cosa no era para tanto. Se tiene ante él la sensación que produce lo trivial, hasta el punto de pensar que se pudo deducir, aunque afortunadamente el Reverendo Thomas Bayes se ocupó de ello en un trabajo titulado *An Essay towards solving a Problem in the Doctrine of Chances*, publicado en 1763. Conviene precisar que Bayes no planteó el teorema en su forma actual, que es debida a Laplace.

1.1.5. Independencia

La información previa que se nos proporcionó sobre el resultado del experimento modificó la probabilidad inicial del suceso. ¿Ocurre esto siempre?

Se supuso que, en lugar de comprar un único boleto, el que lleva el número 35, se compró todos aquellos que terminan en 5. Ahora $P(A) = 1/10$ puesto que se compró 10 boletos, pero al calcular la probabilidad condicionada a la información que se nos ha facilitado, $B = \{\text{el boleto premiado termina en } 5\}$, se observó que $P(A \cap B) = 1/100$, porque la intersección de ambos sucesos es justamente el boleto que está premiado, en definitiva la misma que originalmente tenía A. Parecen existir situaciones en las que la información previa no modifica la probabilidad inicial del suceso. Se observó, que este resultado tiene una consecuencia inmediata,

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B).$$

Esta es una situación de gran importancia en probabilidad que recibe el nombre de independencia de suceso, para el cual, se generalizó mediante la siguiente definición.

Definición 1.4 (Sucesos independientes): Sean A y B dos sucesos. Se dijo que A y B son independientes si $P(A \cap B) = P(A)P(B)$. De esta definición se obtiene como propiedad, su simétrica $P(B|A) = P(B)$.

En ocasiones se define la independencia de dos sucesos a partir de este resultado, obteniéndose entonces como propiedad la que nosotros hemos dado como definición. Existe equivalencia entre ambas definiciones, aunque a fuerza de ser rigurosos, hay que matizar que definir el concepto a partir de la probabilidad condicional exige añadir la condición de que el suceso condicionante tenga probabilidad distinta de cero. Hay además otra ventaja a favor de la definición basada en la factorización de $P(A \cap B)$, pone de inmediato en evidencia la simetría del concepto. El concepto de independencia puede extenderse a una familia finita de sucesos de la siguiente forma.

Definición 1.5 (Independencia mutua): Se dice que los sucesos de la familia $\{A_1, A_2, \dots, A_n\}$ son mutuamente independientes cuando:

$$P(A_{k_1} \cap \dots \cap A_{k_m}) = \prod_{i=1}^m P(A_{k_i}).$$

Siendo $\{k_1, \dots, k_m\} \subset \{1, \dots, n\}$ y los k_i distintos.

El lector podrá verificar que, conviene señalar que la independencia mutua de los n sucesos supone que han de verificarse ecuaciones si solamente se verificase: $\binom{n}{n} + \binom{n}{n-1} + \dots + \binom{n}{2} = 2^n - n - 1$, aquellas

igualdades que implican a dos elementos diríamos que los sucesos son independientes dos a dos, que es un tipo de independencia menos restrictivo que el anterior como pone de manifiesto el siguiente ejemplo. Solo cuando $n=2$ ambos conceptos son equivalentes.

La definición puede generalizarse a familias o clases de sucesos.

Definición 1.6 (Clases independientes de sucesos) Las clases de sucesos $A_1, A_2, \dots, A_n \subset A$ se dicen independientes, si al tomar A_i en cada $A_i, i=1, \dots, n$, los sucesos de la familia $\{A_1, A_2, \dots, A_n\}$ son independientes. Notemos que en la definición no se exige que los elementos de cada clase A_i sean independientes entre si. De hecho A y A^c solo lo son si $P(A)=0$ ó $P(A)=1$.

Para una colección infinita de clases de sucesos, se dedujo de la anterior definición, que se extiende con facilidad. Por ello, se puede comprobar que: $\{A_n\}_{n \geq 1} \subset A$ son independientes si cualquier su colección finita lo es.

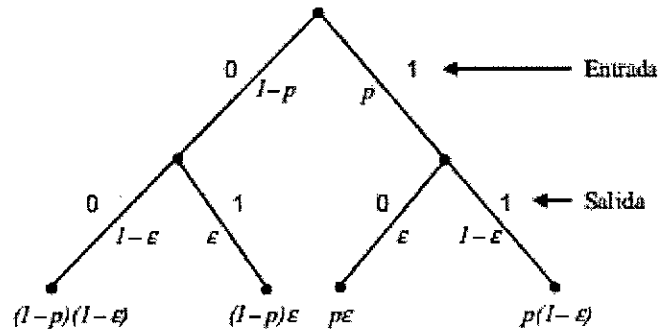
Ejemplo 1.5 (Teorema de Bayes y sistemas de comunicación): Muchos sistemas de comunicación pueden modelizarse de la siguiente forma. El usuario entra un 0 o un 1 en el sistema y la correspondiente señal es transmitida. El receptor toma una decisión acerca del cual fue la entrada del sistema en función de la señal recibida. Supongamos que el usuario envía un 0 con probabilidad $1-p$ y un 1 con probabilidad p , mientras que el receptor toma una decisión errónea con probabilidad ε . Para $i=0,1$, sea:

$$A_i = \{\text{el emisor envía } i\},$$

$$A_i = \{\text{el receptor decidió } i\}.$$

FIGURA N° 1.1

“TEOREMA DE BAYES Y SISTEMAS DE COMUNICACIÓN”



De acuerdo con el esquema de la figura, las probabilidades del tipo $P(A_i \cap B)$, supuesta independencia entre las acciones del emisor y del receptor, valen:

$$P(A_0 \cap B_0) = (1-p)(1-\varepsilon), \quad P(A_0 | B_1) = (1-p)\varepsilon$$

$$P(A_1 \cap B_0) = p\varepsilon, \quad P(A_1 | B_1) = p(1-\varepsilon).$$

Es interesante, en este contexto, conocer las probabilidades del tipo $P(A_i | B_j)$, con $i, j = 0, 1$. Es decir, la probabilidad de que habiendo el receptor interpretado la señal como j la que realmente se transmitió fuera i . De particular interés son, obviamente, $P(A_0 | B_0)$, y $P(A_1 | B_1)$. Para obtener estas probabilidades podemos hacer uso del Teorema de Bayes:

$$P(A_i | B_j) = \frac{P(B_j | A_i)P(A_i)}{\sum_{i=1}^n P(B_j | A_i)P(A_i)}$$

El denominador no es más que $P(B_j)$, que se calcula fácilmente a partir de las anteriores probabilidades. Así, para $j = 0$:

$$P(B_0) = P(B_0 | A_0)P(A_0) + P(B_0 | A_1)P(A_1) = (1-\varepsilon)(1-p) + \varepsilon p,$$

y para $j = 1$:

$$P(B_1) = P(B_1 | A_0)P(A_0) + P(B_1 | A_1)P(A_1) = \varepsilon(1-p) + (1-\varepsilon)p.$$

Handwritten signature

1.2. Variable Aleatoria

1.2.1. Definición

El interés al examinar el resultado de un experimento aleatorio no es tanto el espacio de probabilidad resultante, como la o las características numéricas asociadas, lo que supone cambiar nuestro objetivo de Ω a \mathbb{R} ó \mathbb{R}^k . Hay dos razones que justifican este cambio:

1. El espacio de probabilidad es un espacio abstracto, mientras que \mathbb{R} ó \mathbb{R}^k conocidos en los que resulta mucho más cómodo trabajar, son espacios bien conocidos lo que resultó mucho más cómodo trabajar.
2. Fijar nuestra atención en las características numéricas asociadas a cada resultado implica un proceso de abstracción que, al extraer los rasgos esenciales del espacio muestral, permitió construir un modelo probabilístico aplicable a todos los espacios muestrales que comparten dichos rasgos.

Las características numéricas ligadas, se trata de un experimento aleatorio, son ellas mismas cantidades aleatorias. Esto supuso que para su estudio y conocimiento no bastar 'a con saber que valores toman, para ello de conoció la probabilidad con que lo hacen. Todo ello exige trasladar la información desde el espacio de probabilidad a \mathbb{R} ó \mathbb{R}^k y la única forma que conocemos de trasladar información de un espacio a otro es mediante una aplicación. En este caso la aplicación tuvo que trasladar el concepto de suceso, lo que exige una mínima infraestructura en el espacio receptor de la información semejante a la σ -álgebra que contiene a los sucesos. Posteriormente, se ocupó del caso unidimensional, una sola característica numérica asociada a los puntos del espacio muestral, el espacio imagen fue \mathbb{R} . En \mathbb{R} , los intervalos son el lugar habitual de trabajo, por lo que más conveniente será exigir a esta infraestructura que los contenga. Existe en \mathbb{R} la llamada σ -álgebra de Borel, β , que tiene la propiedad de ser la menor de las que contienen a los intervalos, lo que la hace la más adecuada para convertir a \mathbb{R} en espacio probabilizable: (\mathbb{R}, β) . Estamos ahora en condiciones de definir la variable aleatoria.

Definición 1.7 (Variable aleatoria): Se consideró los dos espacios probabilizables $((\Omega, A)$ y (\mathbb{R}, β) . Una variable aleatoria es una aplicación, $X : \Omega \rightarrow \mathbb{R}$, que verifica:

$$X^{-1}(B) \in A, \quad \forall B \in \beta.$$

Cuando se hizo intervenir una variable aleatoria en nuestro proceso es porque ya estamos en presencia de un espacio de probabilidad, (Ω, \mathcal{A}, P) . La variable aleatoria traslada la información probabilística relevante de Ω a \mathbb{R} mediante una probabilidad inducida que se conoce como ley de probabilidad de X o distribución de probabilidad de X .

➤ **El concepto de σ -álgebra inducida**

Dada la definición de variable aleatoria, se comprobó el siguiente resultado.

Lema 1.1: La familia de sucesos:

$$\sigma(X) = \{X^{-1}(B), \forall B \in \beta\} = \{X^{-1}(\beta)\},$$

es una σ -álgebra, denominada σ -álgebra inducida por X , que verifica $\sigma(X) \subset \mathcal{A}$.

A los efectos de conocer el comportamiento probabilístico de una variable aleatoria X , tres funciones que proporcionaron la información necesaria:

- La probabilidad inducida, P_X ,
- La función de distribución de probabilidad F_X , y
- La función de cuantía o probabilidad, si la variable es discreta, o la función de densidad de probabilidad, si la variable es continua, en ambos casos se denotó por f_X .

Estas definiciones y propiedades se describen a continuación. Se puede demostrar, aunque está fuera del objetivo y alcance de estas notas, la equivalencia entre las tres funciones,

$$P_X \Leftrightarrow F_X \Leftrightarrow f_X.$$

Es decir, el conocimiento de una cualquiera de ellas permite obtener las otras dos.

1.2.2. Probabilidad inducida

X induce sobre (\mathbb{R}, β) una probabilidad P_X , de la siguiente forma,

$$P_X(A) = P(X^{-1}(A)), \forall A \in \beta.$$

Al comprobar P_X , es una probabilidad sobre la σ -álgebra de Borel, de manera que (\mathbb{R}, β, P_X) es un espacio de probabilidad al que podemos aplicar todo cuanto se dijo en el capítulo anterior. Se observó que P_X , hereda las características que tenía P pero lo hace a través de X . ¿Que quiere esto decir? Para poder entender mejor, se consideró el siguiente ejemplo.

Ejemplo 1.6: Sobre el espacio de probabilidad resultante de lanzar dos dados, definimos las variables aleatorias, $X = \text{suma de las caras}$ e $Y = \text{el valor absoluto de las diferencia de las caras}$. Aun cuando el espacio de probabilidad sobre el que ambas están definidas es el mismo, P_X y P_Y son distintas porque viene inducidas por variables distintas. En efecto,

$$P_X(\{0\}) = P(X^{-1}(\{0\})) = P(\emptyset) = 0,$$

sin embargo,

$$P_Y(\{0\}) = P(Y^{-1}(\{0\})) = P(\{1,1\}, \{2,2\}, \{3,3\}, \{4,4\}, \{5,5\}, \{6,6\}) = \frac{1}{6}.$$

La distribución de probabilidad de X, P_X , proporcionó cuanta información se necesita para conocer el comportamiento probabilístico de X , pero se trata de un objeto matemático complejo de incomodo manejo, al que no es ajena su condición de función de conjunto. Esta es la razón por la que recurrimos a funciones de punto para describir la aleatoriedad de X .

1.2.3. Función de distribución de probabilidad

Al tener en cuenta la probabilidad inducida, se pudo definir sobre \mathbb{R} la siguiente función,

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}\{(-\infty, x]\}) = P(X \leq x), \forall x \in \mathbb{R}.$$

Así definida esta función tiene las siguientes propiedades:

- **No negatividad.** - Consecuencia inmediata de su definición.
- **Monotonía.** - De la monotonía de la probabilidad se dedujo que $F_X(x_1) \leq F_X(x_2)$ si $x_1 \leq x_2$.
- **Continuidad por la derecha.** - Se consideró una sucesión decreciente de números reales $x_n \downarrow x$. La correspondiente

sucesión de intervalos verifica $\bigcap_n (-\infty, x_n] = (-\infty, x]$, y por la continuidad desde arriba de la probabilidad respecto del paso al límite tendremos $\lim_{x_n \downarrow x} F_x(x_n) = F(x)$. Por otra parte, se observó que $(-\infty, x] = \{x\} \cup \lim_{n \rightarrow +\infty} \left(-\infty, x - \frac{1}{n}\right]$, para el cual se tomó en cuenta las probabilidades para obtener:

$$F_x(x) = P(X = x) + \lim_{n \rightarrow +\infty} F_x\left(x - \frac{1}{n}\right) = P(X = x) + F(x-),$$

A partir de 1.8 se sigue que $F_x(x)$ es continua en x si y solo si:

$$P(X = x) = 0.$$

- **Valores límites.** – Si $x_n \uparrow +\infty$ ó $x_n \downarrow -\infty$ entonces $(-\infty, x_n] \uparrow \mathbb{R}$ y $(-\infty, x_n] \downarrow \emptyset$ y por tanto

$$F(+\infty) = \lim_{x_n \uparrow +\infty} F_x(x_n) = 1, \quad F(-\infty) = \lim_{x_n \downarrow -\infty} F(x_n) = 0.$$

A la función F_x , se la conoce como función de distribución de probabilidad de X (de distribución). En ocasiones se la denomina función de distribución acumulada, porque tal y como ha sido definida nos informa de la probabilidad acumulada por la variable X hasta el punto x . Esto permitió obtener probabilidades del tipo $P(a < X \leq b)$ a partir de la expresión:

$$P(a < X \leq b) = F_x(b) - F_x(a).$$

1.2.4. Función de cuantía o probabilidad: variable aleatoria discreta

Existe una segunda función de punto que permite describir el comportamiento de X , pero para introducirla hemos de referirnos primero a las características del soporte de X , entendiéndose por tal un conjunto $D_x \in \beta$ que verifica, $P_x(D_x) = P(X \in D_x) = 1$.

Cuando D_x es numerable, P_x es discreta y decimos que X es una variable aleatoria discreta. Como ya se vió en un ejemplo del capítulo anterior, $P_x(\{x_i\}) = P(X = x_i) > 0, \forall x_i \in D_x$, y siendo además $P(X \in D_x) = 1$, se deduce $P(X = x) = 0, \forall x \in D_x^c$. En este caso se comprobó que la F_x asociada viene dada por:



$$F_X(x) = \sum_{x_i \leq x} P(X = x_i).$$

De acuerdo con esto, si $x_{(i)}$ y $x_{(i+1)}$ son dos puntos consecutivos del soporte tse tuvo que $\forall x \in [x_{(i)}, x_{(i+1)}[$, $F_X(x) = F_X(x_{(i)})$. Como además $P_X(x) = 0, \forall x \in D_X$, la función será también continua. Por otra parte $P(X = x_i) > 0$, para $x_i \in D_X$, con lo que los únicos puntos de discontinuidad serán lo del soporte, discontinuidad de salto finito cuyo valor es $F_X(x_{(i)}) - F_X(x_{(i-1)}) = P(X = x_i)$. Se trata por tanto de una función escalonada, cuyos saltos se producen en los puntos de D_X .

A la variable aleatoria discreta podemos asociarle una nueva función puntual que nos será de gran utilidad. La definimos para cada $x \in \mathbb{R}$ mediante $f_X(x_{(i)}) = P_X(\{x\}) = P(X = x)$, lo que supone que

$$f_X(x) = \begin{cases} P(X = x), & \text{si } x \in D_X \\ 0, & \text{en el resto.} \end{cases}$$

Esta función es conocida como función de cuantía o de probabilidad de X y posee las dos propiedades siguientes:

- Al tratarse de una probabilidad, $f_X(x) \geq 0, x \in \mathbb{R}$,
- Como $P(X \in D_X) = 1$:

$$\sum_{x_i \in D_X} f_X(x_i) = 1.$$

La relación entre f_X y F_X viene recogida en las dos expresiones que siguen. La primera de ellas permite obtener F_X a partir de f_X ,

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i).$$

La segunda proporciona f_X en función de F_X ,

$$f_X(x) = F_X(x) - F_X(x^-).$$

1.3. Vector aleatorio

Cuando se estudió k características numéricas ligadas al resultado del k experimento, por ejemplo, la altura y el peso de las personas, se mueve en \mathbb{R}^k



imagen de la aplicación. La σ -álgebra de sucesos con la que se denota a \mathbb{R}^k para hacerlo probabilizables es la correspondiente σ -álgebra de Borel β^k , que tiene la propiedad de ser la menor que contiene a los rectángulos $(a, b] = \prod_{i=1}^k (a_i, b_i]$, con $a = (a_1, \dots, a_k)$, $b = (b_1, \dots, b_k)$, y $-\infty \leq a_i \leq b_i < +\infty$. De entre ellos merecen especial mención aquellos que tiene el extremo inferior en $+\infty$, $S_x = \prod_{i=1}^k (-\infty, x_i]$, a los que denota región suroeste de x , porque sus puntos están situados al suroeste de x .

Definición 1.8 (Vector aleatorio) Un vector aleatorio, $X = (X_1, X_2, \dots, X_k)$, $X = (X_1, X_2, \dots, X_k)$, es una aplicación de Ω en \mathbb{R}^k , que verifica $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \beta^k$.

La presencia de una probabilidad sobre el espacio (Ω, \mathcal{A}) permite al vector inducir una probabilidad sobre (\mathbb{R}^k, β^k) .

1.3.1. Probabilidad inducida

X induce sobre (\mathbb{R}^k, β^k) una probabilidad, P_X , de la siguiente forma:

$$P_X(B) = P(X^{-1}(B)), \forall B \in \beta^k.$$

Es sencillo comprobar que verifica los tres axiomas que definen una probabilidad, por lo que la terna $(\mathbb{R}^k, \beta^k, P_X)$ constituye un espacio de probabilidad con las características de (Ω, \mathcal{A}, P) heredadas a través de X .

1.3.2. Funciones de distribución conjunta y marginales

La función de distribución asociada a P_X se define para cada punto $x = (x_1, \dots, x_k)$ de \mathbb{R}^k mediante:

$$F_X(x) = F_X(x_1, \dots, x_k) = P_X(S_x) = P(X \in S_x) = P\left(\bigcap_{i=1}^k \{X_i \leq x_i\}\right).$$

De la definición se derivan las siguientes propiedades:

- **No negatividad.** - Consecuencia inmediata de ser una probabilidad.
- **Monotonía en cada componente.** - Si $x \leq y$, es decir, $x_i \leq y_i$, $i = 1, \dots, k$, $S_x \subset S_y$ y $F_X(x) \leq F_X(y)$.

- **Continuidad conjunta por la derecha.-** Si $x^{(n)} \downarrow x$, entonces $F_X(x^{(n)}) \downarrow F_X(x)$.
- **Valores límites.-** Al tender a $a \pm \infty$ las componentes del punto, se tiene:

$$\lim_{\forall x_i \uparrow +\infty} F_X(x_1, \dots, x_k) = 1,$$

O bien:

$$\lim_{\forall \exists x_i \downarrow +\infty} F_X(x_1, \dots, x_k) = 0,$$

que no son más que la versión multidimensional de las propiedades ya conocidas para el caso unidimensional. Existe una quinta propiedad sin la cual no sería posible recorrer el camino inverso, obtener P_X a partir de F_X , y establecer la deseada y conveniente equivalencia entre ambos conceptos.

- Se supuso que $k = 2$, considerando un rectángulo $(a, b] = (a_1, b_1] \times (a_2, b_2]$. Indudablemente $P_X([a, b]) \geq 0$, se pudo escribir:

$$P_X([a, b]) = F_X(b_1, b_2) - F_X(b_1, a_2) - F_X(a_1, b_2) + F_X(a_1, a_2) \geq 0.$$

CAPÍTULO II

ESPERANZA. DESIGUALDADES - FUNCIÓN CARACTERÍSTICA

2.1. Esperanza de una variable aleatoria.

La esperanza matemática juega el papel de centro de gravedad de la distribución y nos indica alrededor de qué valor se sitúa nuestra variable o vector. La varianza indica cuan dispersos o agrupados se presentan los valores alrededor de aquella. Existen también otras constantes que proporcionan información acerca de la distribución de probabilidad, son los llamados momentos, de los cuales la esperanza y varianza son casos particulares. Los momentos pueden llegar a aportarnos un conocimiento exhaustivo de la variable aleatoria.

Definición 2.1 (Esperanza de una variable aleatoria discreta) : Sea X aleatoria discreta, f_x su función de cuantía y D_x su soporte. Si g es una función medible definida de (\mathcal{R}, β) en (\mathcal{R}, β) , tal que

$\sum_{x_i \in D_x} |g(x_i) f_x(x_i)| < +\infty$, se deduce que existe la esperanza de $g(X)$, $E[g(X)]$, cuyo valor es:

$$E[g(X)] = \sum_{x_i \in D_x} g(x_i) f_x(x_i).$$

En particular, si $g(X) = X$, el valor es:

$$E(X) = \sum_{x_i \in D_x} x_i f_x(x_i).$$

Definición 2.2 (Esperanza de una variable aleatoria continua) :Sea X aleatoria continua, f_x su función de densidad. Si g es una función medible definida de (R, β) en (R, β) , tal que $\int_R |g(x) f_x(x) dx| < +\infty$, se dice que existe la esperanza de $g(X)$, $E[g(X)]$, cuya valor es:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

En particular, si $g(X) = X$, el valor es:

$$E(X) = \int_{-\infty}^{+\infty} x f_x(x)$$

2.1.1. Momentos de una variable aleatoria

Formas particulares de $g(X)$ dan lugar lo que se denomina momentos de X . En la tabla se resume los distintos tipos de momentos y la correspondiente función que los origina, siempre que ésta sea integrable pues de lo contrario la esperanza no existe.

TABLA N° 2.1

MOMENTO DE UNA VARIABLE ALEATORIA

	De orden k	Absoluto de orden k
Respecto del origen	X^k	$ X ^k$
Respecto de a	$(X-a)^k$	$ X-a ^k$
Factoriales	$X(X-1)\dots(X-k+1)$	$ X(X-1)\dots(X-k+1) $

FUENTE: VARIABLE ALEATORIA (2019); elaboración propia.

Tabla 2.1.- Forma de $g(X)$ para los distintos momentos de X

Respecto de la existencia de los momentos se verifica el siguiente resultado.

Proposición 2.1: Si $E(X^k)$ existe, existen todos los momentos de orden inferior.

La comprobación es inmediata a partir de la desigualdad $|X|^j \leq 1 + |X|^k$, $j \leq k$.

Los momentos de una variable aleatoria son características numéricas que resumen su comportamiento probabilístico. Bajo ciertas condiciones el conocimiento de todos los momentos permite conocer completamente la distribución de probabilidad de la variable.

En el caso $k=1$, cuyo correspondiente momento coincide con $E(X)$ y recibe también el nombre de media. Suele designarse mediante la letra griega μ (μ_x , si existe riesgo de confusión). Puesto que μ es una constante, en la tabla anterior podemos hacer $a = \mu$, obteniendo así una familia de momentos respecto de μ que tienen nombre propio: *Los momentos centrales de orden k* , $E[(X - \mu)^k]$. De entre todos ellos cabe destacar la *varianza*,

$$\text{Var}(X) = \sigma_x^2 = E[(X - \mu)^2].$$

➤ **Propiedades de $E(X)$ y $V(X)$**

Un primer grupo de propiedades no merecen demostración dada su sencillez. Conviene señalar que todas ellas derivan de las propiedades de la integral.

1. Propiedades de $E(X)$:

- La esperanza es un operador lineal:

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)],$$

en particular:

$$E[aX + b] = aE(X) + b.$$

- $P(a \leq X \leq b) = 1 \Rightarrow a \leq E(X) \leq b$.
- $P(g(X) \leq h(X)) = 1 \Rightarrow E[g(X)] \leq E[h(X)]$.
- $|E[g(X)]| \leq E[|g(X)|]$.

2. Propiedades de $V(X)$:

- $V(X) \geq 0$.
- $V(aX + b) = a^2V(X)$.
- $V(X) = E(X^2) - [E(X)]^2$.
- $V(X)$ hace mínima $E[(X - a)^2]$.

En efecto:

$$\begin{aligned} E[(X - a)^2] &= E[(X - E(X) + E(X) - a)^2] \\ &= \\ E[(X - E(X))^2] &+ E[(E(X) - a)^2] + 2E[(X - E(X))(E(X) - a)] \\ &= V(X) + (E(X) - a)^2 \end{aligned}$$

El siguiente resultado fue una alternativa de obtener la $E(X)$ cuando X es no negativa.

Proposición 2.2 Si para $X \geq 0$, existe $E(X)$, entonces:

$$E(X) = \int_0^{+\infty} P(X \geq x) dx = \int_0^{+\infty} P(X > x) dx = \int_0^{+\infty} (1 - F_x(x)) dx.$$

2.1.2. Momentos de algunas variables aleatorias conocidas

➤ Binomial

Si $X \sim B(n, p)$,

$$\begin{aligned} E(X) &= \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^n x \frac{n(n-1)\dots(n-x+1)}{x!} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)\dots(n-x+1)}{(x-1)!} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-y-1} = np \end{aligned}$$

Para obtener $V(X)$, se observó que $E[X(X-1)] = E(X^2) - E(X)$, y de aquí $V(X) = E[X(X-1)] + E(X) - [E(X)]^2$. Aplicando un

desarrollo análogo al anterior se obtuvo $E[X(X-1)] = n(n-1)p^2$ y finalmente:

$$V(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

➤ **Poisson**

Si $X \sim P(\lambda)$,

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

Por otra parte:

$$E[X(X-1)] = \sum_{x \geq 0} x(x-1) e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda} \sum_{x-2 \geq 0} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2.$$

De aquí:

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

➤ **Uniforme**

Si $X \sim U(0,1)$,

$$E(X) = \int_{-\infty}^{+\infty} x f_x dx = \int_0^1 x dx = \frac{1}{2}$$

Para obtener $V(X)$ se utilizó la expresión alternativa:

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3},$$

y de aquí:

$$V(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

➤ **Normal tipificada**

Si $X \sim N(0,1)$, como función de densidad es simétrica respecto del origen:



$$E(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \begin{cases} 0, & \text{Si } k = 2n+1 \\ m_{2n}, & \text{Si } k = 2n \end{cases}$$

Ello supone que $E(X) = 0$ y $V(X) = E(X^2)$. Para obtener los momentos de orden par:

$$m_{2n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}} dx.$$

Integrando por partes:

$$\begin{aligned} \int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}} dx &= \\ -x^{2n-1} e^{-\frac{x^2}{2}} \Big|_0^{+\infty} + (2n-1) \int_0^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}} dx &= (2n-1) \int_0^{+\infty} x^{2n-2} e^{-\frac{x^2}{2}} dx, \end{aligned}$$

Lo que conduce a la fórmula de recurrencia $m_{2n} = (2n-1)m_{2n-2}$ y recorriendo sobre n ,

$$m_{2n} = (2n-1)(2n-3)\dots 1 = \frac{2n(2n-1)(2n-2)\dots 2 \cdot 1}{2n(2n-2)\dots 2} = \frac{(2n)!}{2^n n!}.$$

La varianza es:

$$V(X) = E(X^2) = \frac{2!}{2 \cdot 1!} = 1$$

➤ Normal con parámetros μ y σ^2

Si $Z \sim N(0,1)$ es posible comprobar que la variable definida mediante la expresión $X = \sigma Z + \mu$ es $N(\mu, \sigma^2)$. Teniendo en cuenta las propiedades de la esperanza y de la varianza,

$$E(X) = \sigma E(Z) + \mu = \mu, \quad \text{var}(X) = \sigma^2 \text{var}(Z) = \sigma^2,$$

que son precisamente los parámetros de la distribución.

2.2. Esperanza de un vector aleatorio

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio y sea g una función medible de R^k en R , la expresión de su esperanza depende, como en el caso unidimensional, del carácter del vector.

- **Vector aleatorio discreto.**- Si D_x es el soporte del vector y f_x su función de cuantía conjunta, la esperanza a partir de:

$$E(g(X)) = \sum_{(x_1, \dots, x_k) \in D_x} g(x_1, \dots, x_k) f_x(x_1, \dots, x_k).$$

- **Vector aleatorio continuo.**- Si f_x es la función de densidad conjunta:

$$E(g(X)) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_k) f_x(x_1, \dots, x_k) dx_1, \dots, dx_k.$$

En ambos casos la existencia de la esperanza está supeditada a que $|g(x)f(x)|$ sea absolutamente sumable o integrable, respectivamente.

2.2.1. Momentos de un vector aleatorio

Se comprobó en el caso de que una variable aleatoria, tiene determinadas formas de la función g dando a lugar a los llamados *momentos* que se definen de forma análoga se hizo anteriormente.

Las situaciones de mayor interés son ahora:

- **Momento conjunto.**- El momento conjunto de orden (n_1, \dots, n_k) se obtiene, siempre que la esperanza exista, para:

$$g(X_1, \dots, X_k) = X_1^{n_1} \dots X_k^{n_k}, n_i \geq 0,$$

lo que da lugar a $E(X_1^{n_1} \dots X_k^{n_k})$. Se observó que los momentos de orden k respecto del origen para cada componente pueden obtenerse como casos particulares de la fórmula mostrada, haciendo $n_i = k$, $n_j = 0$ y $j \neq i$, pues entonces:

$$E(X_1^{n_1} \dots X_k^{n_k}) = E(X_i^k).$$

- **Momento conjunto central.**- En el momento conjunto central, se obtuvo de orden (n_1, \dots, n_k) , siempre que la esperanza exista, para:

$$g(X_1, \dots, X_k) = (X_1 - E(X_1))^{n_1} \dots (X_k - E(X_k))^{n_k}, n_i \geq 0.$$



2.2.2. Covarianza. Aplicaciones

De especial interés es el momento conjunto central obtenido para $n_i = 1, n_j = 1$ y $n_l = 0, l \neq (i, j)$. Recibe el nombre de covarianza de X_i y X_j y su expresión es:

$$\text{cov}(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))] = E(X_i X_j) - E(X_i)E(X_j).$$

La covarianza nos informa acerca del grado y tipo de dependencia existente entre ambas variables mediante su magnitud y signo, porque a diferencia de lo que ocurría con la varianza, la covarianza puede tener signo negativo.

➤ Covarianza en la Normal bivalente

Si (X, Y) tiene una distribución Normal bivalente de parámetros $\mu_x, \mu_y, \sigma_x, \sigma_y$ y ρ . Se sabe que $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Para simplificar la obtención de $\text{cov}(X, Y)$ se hizo el uso de la invarianza por traslación de la covarianza (se trata de una propiedad de demostración que se comprobó para la varianza). De acuerdo con ello, $\text{cov}(X, Y) = \text{cov}(X - \mu_x, Y - \mu_y)$, se puede suponer que X e Y tienen media 0, lo que permite expresar la covarianza como $\text{cov}(X, Y) = E(XY)$.

Procedamos a su cálculo:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{xy}(x, y) dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xye^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right) + \left(\frac{y}{\sigma_y}\right)^2\right\}} dx dy \\ &= \int_{-\infty}^{+\infty} \frac{y}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2} \left[\int_{-\infty}^{+\infty} \frac{x}{\sigma_x\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x}{\sigma_x} - \rho\frac{y}{\sigma_y}\right)^2} dx \right] dy \dots (i) \end{aligned}$$

La integral interior en (i) es la esperanza de una $N(\rho\sigma_x y / \sigma_y, \sigma_x^2(1-\rho^2))$ y su valor será por tanto $\rho\sigma_x y / \sigma_y$. Sustituyendo en (i):

$$E(XY) = \frac{\rho\sigma_x}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(\frac{y}{\sigma_y}\right)^2 e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2} dy = \frac{\rho\sigma_x\sigma_y}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{1}{2}z^2} dz$$

$$= \rho\sigma_x\sigma_y.$$

EL coeficiente de correlación entre ambas variables es:

$$\rho_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x\sigma_y} = \rho.$$

Todos los parámetros de la Normal bivalente adquieren ahora significado.

La covarianza nos permite también obtener la varianza asociada a la suma de variables aleatorias. Como se aprecia a continuación.

➤ **Esperanza y varianza de una suma**

La linealidad de la esperanza aplicada cuando $g(X) = X_1 + \dots + X_n$, permite escribir:

$$E(X_1 + \dots + X_k) = \sum_{i=1}^k E(X_i).$$

Si las varianzas de las variables X_1, \dots, X_k existen, la varianza de $S = \sum_{i=1}^k a_i X_i$, donde los a_i son reales cualesquiera, existe y viene dada por

$$\begin{aligned} V(S) &= E[(S - E(S))^2] = \\ &= E\left[\left(\sum_{i=1}^k a_i (X_i - E(X_i))\right)^2\right] \\ &= \sum_{i=1}^k a_i^2 V(X_i) + 2a_i a_j \sum_{i=1}^{k-1} \sum_{j=i+1}^k \text{cov}(X_i, X_j). \quad \dots(ii) \end{aligned}$$

➤ **Independencia y momentos de un vector aleatorio**

Un resultado acerca de los momentos conjuntos se recoge en la proposición que sigue.

Proposición 2.3 Si las variables aleatorias X_1, \dots, X_k son independientes, entonces:



$$E(X_1^{n_1} \dots X_k^{n_k}) = \prod_{i=1}^k E(X_i^{n_i}).$$

Demstración. – Se supuso el vector continuo, la densidad conjunta puede factorizarse como producto de las marginales y

$$\begin{aligned} E(X_1^{n_1} \dots X_k^{n_k}) &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1^{n_1} \dots x_k^{n_k} f_x(x_1, \dots, x_k) dx_1 \dots dx_k \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left[\prod_{i=1}^k x_i^{n_i} f_i(x_i) \right] dx_1 \dots dx_k \\ &= \prod_{i=1}^k \left[\int_{-\infty}^{+\infty} x_i^{n_i} f_i(x_i) dx_i \right] = \prod_{i=1}^k E(X_i^{n_i}). \end{aligned}$$

El caso discreto se demuestra análogamente.

Observación 2.1: El anterior resultado admite una formulación más general. Si las funciones $g_i, i=1, \dots, k$ son medibles, las $g_i(X_i)$ también son variables independientes y podemos escribir

$$E \left[\prod_{i=1}^k g_i(X_i) \right] = \prod_{i=1}^k E[g_i(X_i)].$$

Corolario 2.1: Si las variables X_1, \dots, X_k son independientes, entonces $\text{cov}(X_i, X_j) = 0, \forall i, j$.

Corolario 2.2: Si las variables X_1, \dots, X_k son independientes y sus varianzas existen, la varianza de $S = \sum_{i=1}^k a_i X_i$, donde los a_i son reales cualesquiera, existe y viene dada por $V(S) = \sum_{i=1}^k a_i^2 V(X_i)$.

Una aplicación de los anteriores resultados permite la obtención de la esperanza y la varianza de algunas conocidas variables de manera mucho más sencilla a como lo hicimos anteriormente. A continuación, mostraremos algunos ejemplos.

Ejemplo 2.1 (La Binomial y la Hipergeométrica como suma de Bernoullis): Las características de una Binomial fácilmente se comprueba que si $X \sim B(n, p)$ entonces $X = \sum_{i=1}^n X_i$ con $X_i \sim B(1, p)$ e independientes. Como $\forall_i,$

$$E(X_i) = p, \quad \text{y} \quad \text{var}(X_i) = p(1-p),$$

tendremos que:

$$E(X) = \sum_{i=1}^n E(X_i) = nE(X_i) = np, \quad \text{y}$$

$$\text{var}(X) = \sum_{i=1}^n \text{var}(X_i) = np(1-p) \dots (iii)$$

Si $X \sim H(n, N, r)$, se deduce que también se puede escribir:

$X = \sum_{i=1}^n X_i$ con $X_i \sim B(1, r/N)$ pero a diferencia de la Binomial X_i son ahora dependientes. Se tuvo que:

$$E(X) = \sum_{i=1}^n E(X_i) = nE(X_i) = n \frac{r}{N},$$

y aplicando (iii):

$$\text{var}(X) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^k \sum_{j>1}^k \text{cov}(X_i, X_j) = n \frac{r}{N} \frac{N-r}{N} + n(n-1) \text{cov}(X_1, X_2) \dots (iv),$$

puesto que todas las covarianzas son iguales.

Para obtener $\text{cov}(X_1, X_2)$,

$$\begin{aligned} \text{cov}(X_1, X_2) &= E(X_1 X_2) - E(X_1)E(X_2) \\ &= P(X_1 = 1, X_2 = 1) - \left(\frac{r}{N}\right)^2 \\ &= P(X_2 = 1 | X_1 = 1)P(X_1 = 1) - \left(\frac{r}{N}\right)^2 \\ &= \frac{r-1}{N-1} \frac{r}{N} - \left(\frac{r}{N}\right)^2 \\ &= -\frac{r(N-r)}{N^2(N-1)}. \end{aligned}$$

Sustituyendo en (iv):

$$\text{var}(X) = n \frac{r}{N} \frac{N-r}{N} \left(1 - \frac{n-1}{N-1}\right) \dots iv)$$

Es interesante comparar (v) con (iii), para $p = r/N$. Se vio que difieren en el último factor, factor que será muy próximo a la unidad si $n \ll N$. Es decir, si la fracción de muestreo (así se la denomina en Teoría de Muestras) es muy pequeña. Conviene recordar que se dijo, que hay la relación entre ambas distribuciones.

Ejemplo 2.2 (La Binomial Negativa como suma de Geométricas):

Si $X \sim BN(r, p)$, recondando su definición, podemos expresarla como $X = \sum_{i=1}^r X_i$, donde cada $X_i \sim BN(1, p)$ e independiente de las demás y representa las pruebas Bernoulli necesarias después del $(i-1)$ -ésimo éxito para alcanzar el i -ésimo.

Primero, se obtiene la esperanza y la varianza de una variable Geométrica de parámetro p .

$$E(X_i) = \sum_{n \geq 0} np(1-p)^n = p \sum_{n \geq 1} n(1-p)^n = \frac{1-p}{p}, \forall i.$$

Para calcular la varianza se necesita conocer $E(X_i^2)$:

$$E(X_i^2) = \sum_{n \geq 0} n^2 p(1-p)^n = p \sum_{n \geq 1} n^2 (1-p)^n = \frac{1-p}{p^2} (2-p), \forall i$$

y de aquí:

$$\text{var}(X) = \frac{1-p}{p^2} (2-p) - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}.$$

La esperanza y la varianza de la Binomial Negativa de parámetros r y p , valdrán

$$E(X) = \sum_{i=1}^r E(X_i) = \frac{r(1-p)}{p},$$

$$\text{var}(X) = \sum_{i=1}^r \text{var}(X_i) = \frac{r(1-p)}{p^2}.$$



➤ **Covarianzas en una Multinomial**

Se supo, si $X \sim M(n; p_1, p_2, \dots, p_k)$, cada componente $X_i \sim B(n, p_i), i = 1, \dots, k$, puede expresarse como suma de n Bernoulli de parámetro p_i . La covarianza entre dos cualesquiera puede expresarse como:

$$\text{cov}(X_i, X_j) = \text{cov}\left(\sum_{k=1}^n X_{ik}, \sum_{l=1}^n X_{jl}\right).$$

Se muestra fácilmente que:

$$\text{cov}\left(\sum_{k=1}^n X_{ik}, \sum_{l=1}^n X_{jl}\right) = \sum_{k=1}^n \sum_{l=1}^n \text{cov}(X_{ik}, X_{jl}). \quad \dots (vi)$$

Para calcular $\text{cov}(X_{ik}, X_{jl})$ se debe considerar:

$$X_{ik} = \begin{cases} 1, & \text{si en la prueba } k \text{ ocurre } A_i; \\ 0, & \text{en cualquier otro caso,} \end{cases}$$

y

$$X_{jl} = \begin{cases} 1, & \text{si en la prueba } l \text{ ocurre } A_j; \\ 0, & \text{en cualquier otro caso,} \end{cases}$$

En consecuencia, $\text{Cov}(X_{ik}, X_{jl}) = 0$ si $k \neq l$ porque las pruebas de Bernoulli son independientes:

$$\text{Cov}(X_{ik}, X_{jl}) = E(X_{ik} X_{jl}) - E(X_{ik})E(X_{jl}) = 0 - p_i p_j,$$

donde $E(X_{ik} X_{jl}) = 0$ porque en una misma prueba, la k -ésima, no puede ocurrir simultáneamente A_i y A_j . En definitiva:

$$\text{Cov}(X_i, X_j) = \sum_{k=1}^n \text{cov}(X_{ik}, X_{jk}) = -np_i p_j.$$

El coeficiente de correlación entre ambas variables vale:

$$\rho_{ij} = -\frac{np_i p_j}{\sqrt{np_i(1-p_i)}\sqrt{np_j(1-p_j)}} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}.$$



El valor negativo de la covarianza y del coeficiente de correlación se explica por el hecho de que siendo el número total de pruebas fijo, n , a mayor número de ocurrencias de A_i , menor número de ocurrencias de A_j .

2.3. Esperanza condicionada

Sea (X, Y) un vector aleatorio definido sobre el espacio de probabilidad (Ω, \mathcal{A}, P) y se denota por $P_{X|Y=y}$ la distribución de probabilidad de X condicionada a $Y = y$. Si g es una función medible definida de (R, β) en (R, β) , tal que $E(g(X))$ existe, la *esperanza condicionada* de $g(X)$ dado Y , $E[g(X)|Y]$, es una variable aleatoria que para $Y = y$ toma el valor:

$$E[g(X)|y] = \int_{\Omega} g(X) dP_{X|Y=y} \dots (vii)$$

La forma de (vii) depende de las características de la distribución del vector (X, Y) .

(X, Y) es **discreto**.- Ello supone que el soporte de la distribución, D , es numerable :

$$E[g(X)|y] = \sum_{x \in D_y} g(x) P(X = x | Y = y) = \sum_{x \in D_y} g(x) f_{X|Y}(x|y),$$

donde $D_y = \{x; (x, y) \in D\}$ es la sección de D mediante y .

(X, Y) es continuo. Entonces:

$$E[g(X)|y] = \int_R g(x) f_{X|Y}(x|y) dx.$$

Una definición similar puede darse para $E[h(Y)|X]$ siempre que $E[h(Y)]$ exista.

Ejemplo 2.3 Sean X e Y variables aleatorias independientes ambas con distribución $B(n, p)$.

$$\begin{aligned} f_{X+Y}(m) &= P\left(\bigcup_{k=0}^m \{X = k, Y = m - k\}\right) = \sum_{k=0}^m P(X = k, Y = m - k) \\ &= \sum_{k=0}^m P(X = k)P(Y = m - k) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)} \\
&= p^m (1-p)^{2n-m} \sum_{k=0}^m \binom{n}{k} \binom{n}{m-k} \\
&= \binom{2n}{m} p^m (1-p)^{2n-m},
\end{aligned}$$

de donde $X + Y \sim B(2n, p)$.

La distribución condicionada de $Y | X + Y = m$ es:

$$\begin{aligned}
P(Y = k | X + Y = m) &= \frac{P(Y = k, X + Y = m)}{P(X + Y = m)} \\
&= \frac{P(Y = k, X = m - k)}{P(X + Y = m)} \\
&= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)}}{\binom{2n}{m} p^m (1-p)^{2n-m}} \\
&= \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}},
\end{aligned}$$

es decir, $Y | X + Y = m \sim H(m, 2n, n)$. La $E(Y | X + Y = m)$ valdrá:

$$E(Y | X + Y = m) = \frac{nm}{2n} = \frac{m}{2}.$$

La definición de la esperanza condicionada goza de todas las propiedades inherentes al concepto de esperanza anteriormente estudiadas. No se debe olvidar:

- La esperanza condicionada es un operador lineal:

$$E[(ag_1(X) + bg_2(X) | Y)] = aE[g_1(X) | Y] + bE[g_2(X) | Y].$$

En particular:

$$E[aX + b | Y] = aE[X | Y] + b.$$

- $P(a \leq X \leq b) = 1 \Rightarrow a \leq E(X | Y) \leq b.$
- $P(g_1(X) \leq g_2(X)) = 1 \Rightarrow E[g_1(X) | Y] \leq E[g_2(X) | Y].$
- $E(c | Y) = c,$ para C constante.

Momentos de todo tipo de la distribución condicionada se definen de forma análoga a como hicimos en el caso de la distribución absoluta y gozan de las mismas propiedades. Existen, no obstante, nuevas propiedades derivadas de las peculiares características de este tipo de distribuciones y de que $E[g(X)]$, en tanto que función de Y , es una variable aleatoria. A continuación se mostrará.

Proposición 2.4: Si $E(g(X))$ existe, entonces:

$$E(E[g(X) | Y]) = E(g(X)). \quad \dots(viii)$$

Demostración. - Supongamos el caso continuo.

$$\begin{aligned} E(E[g(X) | Y]) &= \int_{\mathcal{R}} E[g(X) | y] f_y(y) dy \\ &= \int_{\mathcal{R}} \left[\int_{\mathcal{R}} g(x) f_{X|Y}(x | y) dx \right] f_y(y) dy \\ &= \int_{\mathcal{R}} \left[\int_{\mathcal{R}} g(x) \frac{f_{XY}(x, y)}{f_y(y)} dx \right] f_y(y) dy \\ &= \int_{\mathcal{R}} g(x) \left[\int_{\mathcal{R}} f_{XY}(x, y) dy \right] dx \\ &= \int_{\mathcal{R}} g(x) f_x(x) dx = E(g(X)). \end{aligned}$$

Ejemplo 2.4: Considerando el vector (X, Y) con densidad conjunta

$$f_{XY}(x, y) = \begin{cases} \frac{1}{x}, & 0 < y \leq x \leq 1 \\ 0, & \text{en el resto.} \end{cases}$$

Fácilmente se obtendrá que $X \sim U(0, 1)$ y que $Y | X = x \sim U(0, x)$. Aplicando el resultado anterior podemos calcular $E(Y)$,

$$E(Y) = \int_0^1 E(Y | x) f_X(x) dx = \int_0^1 \frac{1}{2} x dx = \frac{1}{4}.$$

Ejemplo 2.5 :Un trabajador está encargado del correcto funcionamiento de n máquinas situadas en línea recta y distantes una de otra l metros. El trabajador debe repararlas cuando se averían, cosa que sucede con igual probabilidad para todas ellas e independiente de una a otra. El operario puede seguir dos estrategias:

Acudir a reparar la máquina estropeada y permanecer en ella hasta que otra máquina se avería, desplazándose entonces hacia ella, o

Situarse en el punto medio de la línea de máquinas y desde allí acudir a la averiada regresando nuevamente a dicho punto cuando la avería está resuelta.

Si X es la distancia que recorre el trabajador entre dos averías consecutivas, ¿Cuál de ambas estrategias le conviene más para andar menos?

Se trata, en cualquiera de las dos estrategias, de obtener la $E(X)$ y elegir aquella que la proporciona menor. Designamos por $E_i(X)$ la esperanza obtenida bajo la estrategia $i = 1, 2$.

Estrategia 1.- Sea A_k el suceso, el operario se encuentra en la máquina k . Para obtener $E_1(X)$ recurriremos a la propiedad anterior, pero utilizando como distribución condicionada la que se deriva de condicionar respecto del suceso A_k . Tendremos $E_1(X) = E(E(X | A_k))$.

Para obtener $E(X | A_k)$ tengamos en cuenta que si i es la próxima máquina averiada, $P(A_i) = 1/n, \forall i$ y el camino a recorrer será

$$X|A_k = \begin{cases} (k-i)l, & \text{si } i \leq k, \\ (i-k)l, & \text{si } i > k. \end{cases}$$

Así:

$$E(X | A_k) = \frac{1}{n} \left(\sum_{i=1}^k (k-i)l + \sum_{i=k+1}^n (i-k)l \right) = \frac{l}{2n} [2k^2 - (n+1)k + n(n+1)].$$

Utilizando:

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6},$$

obtenemos para $E_1(X)$:

$$E_1(X) = E(E(X | A_k)) = \frac{1}{n} \sum_k E(X | A_k) = \frac{l(n^2 - 1)}{3n}$$

Estrategia 2.- Para facilitar los cálculos supongamos que n es impar, lo que supone que hay una máquina situada en el punto medio de la línea, la $\frac{n+1}{2}$ -ésima. Si la próxima máquina averiada es la i la distancia a recorrer será:

$$X = \begin{cases} 2\left(\frac{n+1}{2} - i\right)l, & \text{si } i \leq \frac{n+1}{2}k, \\ 2\left(i - \frac{n+1}{2}\right)l, & \text{si } i > \frac{n+1}{2}k, \end{cases}$$

donde el 2 se justifica porque el operario en esta segunda estrategia regresa siempre al punto medio de la línea de máquinas. La esperanza viene dada por:

$$E_2(X) = \frac{2}{n} \sum_{i=1}^n \left| \frac{n+1}{2} - i \right| l = \frac{l(n-1)}{2}.$$

Como $E_1(X) \leq E_2(X) \rightarrow (n+1)/3n \leq 1/2 \rightarrow n \geq 2$, podemos deducir que la primera estrategia es mejor, salvo que hubiera una sola máquina.

También es posible relacionar la varianza absoluta con la varianza condicionada, aunque la expresión no es tan directa como la obtenida para la esperanza.

Proposición 2.5: Si $E(X^2)$ existe, entonces:

$$\text{var}(X) = E(\text{var}[X|Y]) + \text{var}(E[X|Y]). \dots (ix)$$

Demostración. – Del anterior resultado, se obtiene:

$$\begin{aligned} \text{var}(X) &= E[(X - E(X))^2] = E\{E[(X - E(X))^2 | Y]\} \\ &= E\{E[X^2 + (E(X))^2 - 2XE(X) | Y]\} \\ &= E\{E[X^2 | Y] + E(X)^2 - 2E(X)E[X | Y]\} \\ &= E\{E[X^2 | Y] - (E[X | Y])^2 + (E[X | Y])^2 + E(X)^2 - 2E(X)E[X | Y]\} \\ &= E\{\text{var}[X | Y] + (E[X | Y] - E(X))^2\} \\ &= E\{\text{var}[X | Y]\} + E\{(E[X | Y] - E(X))^2\} \\ &= E(\text{var}[X | Y]) + \text{var}(E[X | Y]). \end{aligned}$$

Corolario 2.3 Si $E(X^2)$ existe, por la propiedad anterior

$$\text{var}(X) \geq \text{var}(E[X|Y]).$$

Si se verifica la igualdad, de (2.14) deducimos que $E(\text{var}[X|Y])=0$, y como $\text{var}[X|Y] \geq 0$, tendremos que $P(\text{var}[X|Y]=0)=1$, es decir

$$P\left(E\left\{(X - E[X|Y])^2 | Y\right\} = 0\right) = 1,$$

y como $(X - E[X|Y])^2 \geq 0$, aplicando de nuevo el anterior razonamiento concluiremos que

$$P(X = E[X|Y]) = 1,$$

lo que supone que, con probabilidad 1, X es una función de Y puesto que $E[X|Y]$ lo es.

2.4. Desigualdades

Si para $X \geq 0$ existe su esperanza, sea $\varepsilon > 0$ y escribamos (2.3) de la forma,

$$E(X) = \int_0^{+\infty} P(X \geq x) dx = \int_0^{\varepsilon} P(X \geq x) dx + \int_{\varepsilon}^{+\infty} P(X \geq x) dx.$$

Como la segunda integral es no negativa y la función $P(X \geq x)$ es decreciente:

$$E(X) \geq \int_0^{\varepsilon} P(X \geq x) dx \geq \int_0^{\varepsilon} P(X \geq \varepsilon) dx = \varepsilon P(X \geq \varepsilon),$$

y de aquí:

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}. \quad \dots(x)$$

Este resultado da lugar a dos conocidas desigualdades generales que proporcionan cotas superiores para la probabilidad de ciertos conjuntos. Estas desigualdades son válidas independientemente de cuál sea la distribución de probabilidad de la variable involucrada.

- **Desigualdad de Markov.**- La primera de ellas se obtiene al sustituir en (2.15) X por $|X|^k$ y ε por ε^k ,

$$P(|X| \geq \varepsilon) = P(|X|^k \geq \varepsilon^k) \leq \frac{1}{\varepsilon^k} E(|X|^k), \quad \dots(x_i)$$

y es conocida como la desigualdad de Markov.

- **Desigualdad de Chebyshev.**- Un caso especial de (2.16) se conoce como la desigualdad Chebyshev y se obtiene para $k=2$ y $X = X - E(X)$,

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(X). \quad \dots(x_{ii})$$

Proposición 2.6 Si $V(X) = 0$, entonces X es constante con probabilidad 1.

Demostración.- Suponiendo $E(X) = \mu$ y considerando los conjuntos $A_n = \{|X - \mu| \geq 1/n\}$, se aplicó (xii):

$$P(A_n) = P\left(|X - \mu| \geq \frac{1}{n}\right) = 0, \forall n$$

y de aquí $P\left(\bigcup_n A_n\right) = 0$ y $P\left(\bigcap_n A_n^c\right) = 1$. Pero:

$$\bigcap_{n \geq 1} A_n^c = \bigcap_{n \geq 1} \left\{ |x - \mu| < \frac{1}{n} \right\} = \{X = \mu\},$$

luego $P(X = \mu) = 1$.

- **Desigualdad de Jensen.**- Si $g(X)$ es convexa sabemos que $\forall a, \exists \lambda_a$ tal que $g(x) \geq g(a) + \lambda_a(x - a), \forall x$. Si hacemos ahora $a = E(X)$,

$$g(X) \geq g(E(X)) + \lambda_a(X - E(X)),$$

y tomando esperanzas obtenemos la que se conoce como desigualdad de Jensen:

$$E(g(X)) \geq g(E(X)).$$

Teorema 2.1 (Desigualdad de Cauchy-Schwarz) Sean X e Y variables aleatorias con varianzas finitas. Entonces $\text{cov}(X, Y)$ existe y se verifica

$$[E(XY)]^2 \leq E(X^2)E(Y^2),$$

verificándose la igualdad si y solo si existe un real α tal que $P(\alpha X + Y = 0) = 1$.

Demostración.- Para cualesquiera números reales a y b se verifica

$$|ab| \leq \frac{a^2 + b^2}{2},$$

lo que significa que $E(XY) < \infty$ si $E(X^2) < \infty$ y $E(Y^2) < \infty$. Por otra parte, para cualquier real α , se tiene

$$E[(\alpha X + Y)^2] = \alpha^2 E(X^2) + 2\alpha E(XY) + E(Y^2) \geq 0,$$

lo que supone que la ecuación de segundo grado tiene a lo sumo una raíz y su discriminante será no positivo. Es decir,

$$[E(XY)]^2 \leq E(X^2)E(Y^2).$$

Si se diera la igualdad, la ecuación tendría una raíz doble, α_0 y $E[(\alpha_0 X + Y)^2] = 0$. Tratándose de una función no negativa, esto implica que $P(\alpha_0 X + Y = 0) = 1$.

➤ El coeficiente de correlación

El coeficiente de correlación entre dos componentes cualesquiera de un vector aleatorio, X y Y , se define como la covarianza de dichas variables tipificadas.

$$\rho_{XY} = \text{cov}(X, Y) = E\left[\left(\frac{X - E(X)}{\sigma_X}\right)\left(\frac{Y - E(Y)}{\sigma_Y}\right)\right] = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Desde la desigualdad de Cauchy-Schwarz se desprende que $\rho_{XY}^2 \leq 1$ y en particular que $-1 \leq \rho_{XY} \leq 1$. Recordamos que cuando en Cauchy-Schwarz se daba la igualdad X e Y estaban relacionadas linealmente con probabilidad 1, $P(\alpha_0 X + Y = 0) = 0$, pero por otra parte dicha igualdad implica $\rho_{XY}^2 = 1$. El valor 0 es otro valor de interés para ρ_{XY} . Si $\rho_{XY} = 0$ decimos que las variables están incorreladas y además $\text{cov}(X, Y) = 0$. Hay entonces una ausencia total de relación lineal entre ambas variables. Podemos decir que el valor absoluto del coeficiente de correlación es una medida del grado de linealidad entre las variables medido, de menor a mayor, en una escala de 0 a 1.

2.4.1. La distribución Normal multivariante

La expresión de la densidad de la Normal bivalente admite una forma alternativa más compacta a partir de lo que se conoce como la matriz de covarianzas de la distribución, Σ , una matriz 2×2 cuyos elementos son las varianzas y las covarianzas del vector (X_1, X_2) ,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

La nueva expresión de la densidad es:

$$f(x_1, x_2) = \frac{|\Sigma|^{-\frac{1}{2}}}{2\pi} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad (x_1, x_2) \in R^2,$$

donde $|\Sigma|$ es el determinante de Σ , cuyo valor es:

$$|\Sigma| = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2),$$

$x' = (x_1, x_2)$ es el vector de variables y $\mu' = (\mu_1, \mu_2)$ es el vector de medias.

Si el vector tiene n componentes, X_1, X_2, \dots, X_n , la extensión a n dimensiones de la expresión de la densidad es ahora inmediata con esta notación. La matriz de covarianzas es una matriz $n \times n$ con componentes:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1(n-1)} & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2(n-1)} & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{1(n-1)} & \sigma_{2(n-1)} & \dots & \sigma_{n-1}^2 & \sigma_{(n-1)n} \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_{(n-1)n} & \sigma_n^2 \end{pmatrix},$$

el vector de medias es $\mu' = (\mu_1, \mu_2, \dots, \mu_n)$ y la densidad tiene por extensión:

$$f(x_1, x_2, \dots, x_n) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad (x_1, x_2, \dots, x_n) \in R^n. \dots (xiii)$$

Cuando las componentes del vector son independientes, las covarianzas son todas nulas y Σ es una matriz diagonal cuyos elementos son las varianzas de cada componente, por tanto:

$$|\Sigma|^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 \dots \sigma_n^2}.$$

Además, la forma cuadrática que aparece en el exponente (2.18) se simplifica y la densidad adquiere forma

$$f(x_1, x_2, \dots, x_n) = \frac{|\Sigma|^{-\frac{1}{2}}}{\prod_{i=1}^n (2\pi\sigma_i^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} \right],$$

que es el producto de las densidades de cada una de las componentes.

Añadamos por último que la matriz Σ es siempre definida positiva y lo es también la forma cuadrática que aparece en el exponente de (2.18).

➤ **Transformación lineal de una Normal multivariante**

A partir del vector X definimos un nuevo vector Y mediante una transformación lineal cuya matriz A es invertible. Tendremos:

$$Y = AX, \quad \text{y} \quad X = A^{-1}Y,$$

y dada la linealidad de la esperanza, la misma relación se verifica para los vectores de medias:

$$\mu_Y = A\mu_X, \quad \text{y} \quad \mu_X = A^{-1}\mu_Y,$$

Para conocer la distribución del vector Y recurrimos a la fórmula del cambio de variable. El jacobiano de la transformación inversa es precisamente $|J^{-1}| = |A^{-1}|$, con lo que la densidad conjunta de $Y = (Y_1, Y_2, \dots, Y_n)$ valdrá:

$$\begin{aligned}
f_Y(y_1, y_2, \dots, y_n) &= \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(A^{-1}y - A^{-1}\mu_Y)' \Sigma^{-1} (A^{-1}y - A^{-1}\mu_Y)} |A^{-1}| \\
&= \frac{|A^{-1}| |\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} [A^{-1}(y - \mu_Y)]' \Sigma^{-1} [A^{-1}(y - \mu_Y)]} \\
&= \frac{|A^{-1}| |\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} (y - \mu_Y)' [A^{-1}]' \Sigma^{-1} A^{-1} (y - \mu_Y)} \\
&= \frac{|A \Sigma A'|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} (y - \mu_Y)' [A \Sigma A']^{-1} (y - \mu_Y)},
\end{aligned}$$

que es la densidad de una Normal multivariante con vector de medias $\mu_Y = A\mu_X$ y matriz de covarianzas $\Sigma_Y = A \Sigma A'$

2.5. Función característica

La función característica es una herramienta de gran utilidad en Teoría de la Probabilidad, una de sus mayores virtudes reside en facilitar la obtención de la distribución de probabilidad de la suma de variables aleatorias y la del límite de sucesiones de variables aleatorias, situaciones ambas que aparecen con frecuencia en Inferencia Estadística.

El concepto de función característica, introducido por Lyapunov en una de las primeras versiones del Teorema Central del Límite, procede del Análisis Matemático donde se le conoce con el nombre de transformada de Fourier.

Definición 2.3 Sea X una variable aleatoria y sea $t \in R$. La función característica de X , $\phi_X(t)$, se define como $E(e^{itX}) = E(\cos tX) + iE(\sin tX)$.

Como $|e^{itX}| \leq 1, \forall t, \phi_X(t)$ existe siempre y está definida $\forall t \in R$. Para su obtención recordemos que,

Caso discreto.- Si X es una v.a. discreta con soporte D_X y función de cuantía $f_X(x)$,

$$\phi_X(t) = \sum_{x \in D_X} e^{itx} f_X(x) \dots (xvi)$$

Caso continuo.- Si X es una v.a. continua con función de densidad de probabilidad $f_X(x)$,

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx.$$

De la definición se derivan, entre otras, las siguientes propiedades:

$P\phi 1) \phi_X(0) = 1$

$P\phi 2) |\phi_X(t)| \leq 1$

$P\phi 3) \phi_X(t)$ es uniformemente continua

En efecto:

$$\phi_X(t+h) - \phi_X(t) = \int_{\Omega} e^{itx} (e^{ihx} - 1) dP.$$

Al tomar módulos:

$$|\phi_X(t+h) - \phi_X(t)| \leq \int_{\Omega} |e^{ihx} - 1| dP \dots (xv)$$

pero $|e^{ihx} - 1| \leq 2$ y (2.21) será finito, lo que permite intercambiar integración y paso al límite, obteniendo:

$$\lim_{h \rightarrow 0} |\phi_X(t+h) - \phi_X(t)| \leq \int_{\Omega} \lim_{h \rightarrow 0} |e^{ihx} - 1| dP = 0.$$

$P\phi 4)$ Si definimos $Y = aX + b$,

$$\phi_Y(t) = E(e^{itY}) = E(e^{it(ax+b)}) = e^{itb} \phi_X(at)$$

$P\phi 5)$ Si $E(X^n)$ existe, la función característica es n veces diferenciable y $\forall k \leq n$ se verifica $\phi_X^{(k)}(0) = i^k E(X^k)$

La propiedad 5 establece un interesante relación entre las derivadas de $\phi_X(t)$ y los momentos de X cuando estos existen, relación que permite desarrollar $\phi_X(t)$ en serie de potencias. En efecto, si $E(X^n)$ existe $\forall n$, entonces,

$$\phi_X(t) = \sum_{k \geq 0} \frac{i^k E(X^k)}{k!} t^k. \dots (xvi)$$

2.5.1. Función característica e independiente

Si X_1, X_2, \dots, X_n son variables aleatorias independientes y definimos $Y = X_1 + X_2 + \dots + X_n$, considerando (ii):

$$\phi_Y(t) = E\left(e^{it(X_1+X_2+\dots+X_n)}\right) = E\left(\prod_{k=1}^n e^{itX_k}\right) = \prod_{k=1}^n E\left(e^{itX_k}\right) = \prod_{k=1}^n \phi_{X_k}(t) \dots (xvi),$$

expresión que permite obtener con facilidad la función característica de la suma de variables independientes y cuya utilidad pondremos de manifiesto de inmediato.

2.5.2. Funciones características de algunas distribuciones conocidas

➤ **Bernoulli.**- Si $X \sim B(1, p)$

$$\phi_X(t) = e^0 q + e^{it} p = q + pe^{it}.$$

➤ **Binomial.**- Si $X \sim B(n, p)$

$$\phi_X(t) = \prod_{k=1}^n (q + pe^{it}) = (q + pe^{it})^n.$$

➤ **Poisson.**- Si $X \sim P(\lambda)$

$$\phi_X(t) = \sum_{x \geq 0} e^{itx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x \geq 0} \frac{(\lambda e^{it})^x}{x!} = e^{\lambda(e^{it}-1)}.$$

➤ **Normal Tipificada.**- Si $Z \sim N(0,1)$, sabemos que existen los momentos de cualquier orden y en particular, $E(Z^{2n+1}) = 0, \forall n$ y

$$E(Z^{2n}) = \frac{(2n)!}{2^n n!}, \forall n. \text{ Aplicando (2.22),}$$

$$\phi_Z(t) = \sum_{n \geq 0} \frac{i^{2n} (2n)!}{2^n (2n)! n!} t^{2n} = \sum_{n \geq 0} \frac{\left(\frac{-t^2}{2}\right)^n}{n!} = e^{-\frac{t^2}{2}}.$$

Para obtener $\phi_Z(t)$ si $X \sim N(\mu, \sigma^2)$, podemos utilizar el resultado anterior y P4. En efecto, recordemos que X puede expresarse en función de Z mediante $X = \mu + \sigma Z$ y aplicando P4,

$$\phi_X(t) = e^{i\mu t} \phi_Z(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

Observación 2.2: Obsérvese que $\text{Im}(\phi_Z(t)) = 0$. El lector puede comprobar que no se trata de un resultado exclusivo de la Normal tipificada, sino de una propiedad que poseen todas las v.a. con distribución de probabilidad simétrica, es decir, aquellas que verifican $(P(X \geq x)) = P(X \leq -x)$.

➤ **Gamma.-** Si $X \sim G(\alpha, \lambda)$, su función de densidad de probabilidad viene dada por

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

por lo que aplicando (2.20),

$$\phi_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-\lambda x} dx,$$

que con el cambio $y = x(1 - it/\lambda)$ conduce a

$$\phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}.$$

Hay dos casos particulares que merecen ser mencionados:

➤ **Exponencial.-** La distribución exponencial puede ser considerada un caso particular de $G(\alpha, \lambda)$ cuando $\alpha = 1$. A partir de aquí,

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

➤ **Chi-cuadrado.-** Cuando $\alpha = n/2$ y $\lambda = 1/2$, decimos que X tiene una distribución X^2 con n grados de libertad, $X \sim X_n^2$. Su función característica será

$$\phi_X(t) = (1 - 2it)^{-\frac{n}{2}}.$$

2.5.3. Teorema de inversión. Unicidad

Se obtiene la función característica de una v.a. X a partir de su distribución de probabilidad, pero es posible proceder de manera inversa por cuanto el conocimiento de $\phi_x(t)$ permite obtener $F_x(x)$.

Teorema 2.2 (Fórmula de inversión de Lévy) Sean $\phi(t)$ y $F(x)$ las funciones característica y de distribución de la v.a. X y sean $a \leq b$ sendos puntos de continuidad de F , entonces :

$$F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt.$$

Que puede también expresarse:

$$F(x_0 + h) - F(x_0 - h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\operatorname{sen} ht}{t} e^{-itx_0} \phi(t) dt,$$

donde $h > 0$ y $x_0 + h$ y $x_0 - h$ son puntos de continuidad de F .

Este resultado permite obtener $F(x)$ en cualquier x que sea punto de continuidad de F . Basta para ello que en $F(x) - F(y)$ hagamos que $y \rightarrow \infty$ a través de puntos de continuidad. Como $F_x(x)$ es continua por la derecha, la tendremos también determinada en los puntos de discontinuidad sin más que descender hacia ellos a través de puntos de continuidad.

Si la variable es continua, un corolario del anterior teorema permite obtener la función de densidad directamente a partir de la función característica.

Corolario 2.4 Si $\phi(t)$ es absolutamente integrable en R , entonces la función de distribución es absolutamente continua, su derivada es uniformemente continua:

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{2\pi} \int_R e^{-itx} \phi(t) dt.$$

Este teorema tiene una trascendencia mayor por cuanto implica la unicidad de la función característica, que no por casualidad recibe este nombre, porque caracteriza, al igual que lo hacen otras funciones asociadas a X (la de distribución, la de probabilidad o densidad de probabilidad,...), su distribución de probabilidad. Podemos afirmar que si dos variables X e Y comparten la misma función característica tienen idéntica distribución de probabilidad. La combinación de este

resultado con las propiedades antes enunciadas da lugar a una potente herramienta que facilita el estudio y obtención de las distribuciones de probabilidad asociadas a la suma de variables independientes. Veámoslo con algunos ejemplos.

1) Suma de Binomiales Independientes.- Si las variables $X_k \sim B(n_k, p)$, $k = 1, 2, \dots, m$ son independientes, al definir $X = \sum_{k=1}^m X_k$, se sabe de (xvi) que:

$$\phi_X(t) = \prod_{k=1}^m \phi_{X_k}(t) = \prod_{k=1}^m (q + pe^{it})^n,$$

con $n = n_1 + n_2 + \dots + n_m$. Pero siendo esta la función característica de una variable $B(n, p)$, podemos afirmar que $X \sim B(n, p)$.

2) Suma de Poisson Independientes.- Si nuestra suma es ahora de variables Poisson independientes, $X_k \sim P(\lambda_k)$, entonces:

$$\phi_X(t) = \prod_{k=1}^m \phi_{X_k}(t) = \prod_{k=1}^m e^{\lambda_k(e^{it}-1)} = e^{\lambda(e^{it}-1)},$$

con $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_m$. Así pues, $X \sim P(\lambda)$.

3) Combinación lineal de Normales independientes.- Si $X = \sum_{k=1}^n c_k X_k$ con $X_k \sim N(\mu_k, \sigma_k^2)$ e independientes,

$$\begin{aligned} \phi_X(t) &= \phi_{\sum c_k X_k}(t) = \prod_{k=1}^n \phi_{X_k}(c_k t) \\ &= \prod_{k=1}^n e^{ic_k t \mu_k - \frac{\sigma_k^2 c_k^2 t^2}{2}} \\ &= e^{i\mu t - \frac{\sigma^2 t^2}{2}} \quad \dots(xvii) \end{aligned}$$

4) Suma de Exponenciales independientes.- En el caso de que la suma esté formada por n variables independientes todas ellas $Exp(\lambda)$,

$$\phi_X(t) = \left(\frac{\lambda}{\lambda - it} \right)^n = \left(1 - \frac{it}{\lambda} \right)^{-n},$$

y su distribución será la de una $G(n, 1/\lambda)$.

5) **Cuadrado de una $N(0,1)$** .- Sea ahora $Y = X^2$ con $X \sim N(0,1)$, su función característica viene dada por:

$$\phi_Y(t) = \int_{-\infty}^{+\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{x^2}{2}(1-2it)} dx = \frac{1}{(1-2it)^{\frac{1}{2}}},$$

lo que nos asegura que $Y \sim X^2$.

Algunos de estos resultados fueron obtenidos anteriormente, pero su obtención fue entonces mucho más laboriosa de lo que ahora ha resultado.

2.5.4. Teorema de continuidad de Lévy

Permite conocer la convergencia de una sucesión de v.a. a través de la convergencia puntual de la sucesión de sus funciones características.

Teorema 2.3 (Directo) Sea $\{X_n\}_{n \geq 1}$ una sucesión de v.a. y $\{F_n(x)\}_{n \geq 1}$ y $\{\phi_n(t)\}_{n \geq 1}$ las respectivas sucesiones de sus funciones de distribución y características. Sea X una v.a. y $F_X(x)$ y $\phi(t)$ sus funciones de distribución y característica, respectivamente. Si $F_n \xrightarrow{w} F$ (es decir, $X_n \xrightarrow{L} X$), entonces

$$\phi_n(t) \rightarrow \phi(t), \forall t \in \mathbb{R}.$$

El resultado se completa con el teorema inverso.

Teorema 2.4 (Inverso) Sea $\{\phi_n(t)\}_{n \geq 1}$ una sucesión de funciones características y $\{F_n(x)\}_{n \geq 1}$ la sucesión de funciones de distribución asociadas. Sea $\phi(t)$ una función continua, si $\forall t \in \mathbb{R}, \phi_n(t) \rightarrow \phi(t)$, entonces:

$$F_n \xrightarrow{w} F,$$

donde $F(x)$ es una función de distribución cuya función característica es $\phi(t)$.

Este resultado permite, como decíamos antes, estudiar el comportamiento límite de sucesiones de v.a. a través de sus funciones características, generalmente de mayor sencillez. Una de sus aplicaciones más relevantes ha sido el conjunto de resultados que se

R

conocen como Teorema Central del Limite (TCL), denotado por Lyapunov que pretendió así destacar el papel central de estos resultados en la Teoría de la Probabilidad.

CAPÍTULO III

SUCESIONES DE VARIABLES ALEATORIAS - TEOREMAS DE CONVERGENCIA

3.1. Introducción

Los capítulos anteriores están relacionados con el concepto de variable y vector aleatorio, el cual nos brinda las herramientas que nos permiten conocer su comportamiento probabilístico. En el caso de un vector aleatorio se estudia el comportamiento conjunto de un número finito de variables aleatorias. Los modelos probabilísticos asociados a los siguientes fenómenos:

Lanzamientos sucesivos de una moneda.

Tiempos transcurridos entre dos llamadas consecutivas a una misma centralita.

Sucesión de estimadores de un parámetro cuando se incrementa el tamaño de la muestra.

En todos los casos las variables aleatorias involucradas lo son en cantidad numerable y se podrá estudiar su comportamiento conjunto y, de conocer cuánto haga referencia al límite de la sucesión. Del comportamiento conjunto se ocupa una parte de la Teoría de la Probabilidad que dado su interés ha tomado entidad propia: La teoría de los procesos estocásticos. En este capítulo nos ocuparemos de estudiar la relación con el límite de las sucesiones de variables aleatorias. Este estudio requiere en primer lugar, introducir los tipos de convergencia apropiados a la naturaleza de las sucesiones que nos ocupan, para en segundo lugar obtener las condiciones bajo las que tienen lugar las dos convergencias que nos interesan: La convergencia de la sucesión de variables a una constante (Leyes de los Grandes Números) y la convergencia a otra variable (Teorema central del límite). El estudio de esta segunda situación se ve facilitado con el uso de una herramienta conocida como función característica.

Ejemplos relacionados con la distribución binomial.

Ejemplo 3.1 (Un resultado de J. Bernoulli) Si repetimos n veces un experimento cuyo resultado es la ocurrencia o no del suceso A , tal que $P(A) = p$, y si las repeticiones son independientes unas de otras, la variable

X_n = número de ocurrencias de A, tiene una distribución $B(n, p)$. La variable X_n/n representa la frecuencia relativa de A y sabemos que

$$E\left(\frac{X_n}{n}\right) = \frac{1}{n} E(X_n) = \frac{np}{n} = p,$$

y

$$\text{var}\left(\frac{X_n}{n}\right) = \frac{1}{n^2} \text{var}(X_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Si aplicamos la desigualdad de Chebyshev:

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{\text{var}(X_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Deducimos que la frecuencia relativa de ocurrencias de A converge, en algún sentido, a $P(A)$.

Ejemplo 3.2 (Binomial vs Poisson) El segundo ejemplo hacía referencia a la aproximación de la distribución Binomial mediante la distribución de Poisson. Vimos que cuando tenemos un gran número de pruebas Bernoulli con una probabilidad de éxito muy pequeña de manera que $\lim_n np_n = \lambda$, $0 < \lambda < +\infty$, la sucesión de funciones de cuantía de las variables aleatorias $X_n \sim B(n, p_n)$ converge a la función de cuantía de $X \sim Po(\lambda)$.

Ejemplos con sucesiones de variables Binomiales que han conducido a límites muy distintos. En el primero, el valor límite es la probabilidad de un suceso, y por tanto una constante; en el segundo la función de cuantía tiende a otra función de cuantía.

3.2. Tipos de convergencia.

Sobre un espacio de probabilidad (Ω, \mathcal{A}, P) consideremos la sucesión de variables aleatorias $\{X_n\}$ y la variable aleatoria X .

Definición 3.1 (Convergencia en probabilidad) Decimos que $\{X_n\}$ converge a X en probabilidad, $X_n \xrightarrow{P} X$, si para cada $\delta > 0$,

$$\lim_n P\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\} = 0.$$

Definición 3.2 (Convergencia casi segura o con probabilidad 1) Decimos que $\{X_n\}$ converge a X (o con probabilidad 1), $X_n \xrightarrow{a.s.} X$.

Definición 3.3 (Convergencia débil) Sean F_n , $n \geq 1$ y F , funciones de distribución de probabilidad. Decimos que la sucesión F_n converge débilmente a F , $F_n \xrightarrow{w} F$, si $\lim_n F_n(x) = F(x)$, $\forall x$ que sea punto de continuidad de F .

Definición 3.4 (Convergencia en ley) Decimos que $\{X_n\}$ converge en ley a X , $X_n \xrightarrow{L} X$, si $F_{X_n} \xrightarrow{w} F_X$. Teniendo en cuenta la definición de F_n y F , la convergencia en ley puede expresarse también:

$$X_n \xrightarrow{L} X \Leftrightarrow \lim_n P(X_n \leq x) = P(X \leq x).$$

Definición 3.5 (Convergencia en media cuadrática) Decimos que $\{X_n\}$ converge en media cuadrática a X , $X_n \xrightarrow{m.s} X$, si:

$$\lim_n E[(X_n - X)^2] = 0.$$

Las relaciones entre los tres tipos de convergencia se establecen en el siguiente teorema.

Teorema 3.1 (Relación entre convergencias) Sean X_n y X variables aleatorias definidas sobre un mismo espacio de probabilidad entonces:

$$\left. \begin{array}{l} X_n \xrightarrow{a.s.} X \\ X_n \xrightarrow{m.s.} X \end{array} \right\} \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{L} X$$

Las convergencias casi seguras y en probabilidad tienen distinta naturaleza, mientras aquella es de tipo puntual, esta última es de tipo conjuntista.

Ejemplo 3.3 (La convergencia en probabilidad \neq la convergencia casi segura) Como espacio de probabilidad consideraremos el intervalo unidad dotado de la σ - álgebra de Borel y de la medida de Lebesgue, es decir, un espacio de probabilidad uniforme en $]0,1[$. Definimos la sucesión $X_n = 1_{I_n}$,

$\forall n$, con $I_n = \left] \frac{p}{2^q}, \frac{p+1}{2^q} \right]$, siendo p y q los únicos enteros positivos que verifican, $p + 2^q$.

Obviamente $q = q(n)$ y $\lim_n q(n) = +\infty$. Los primeros términos de la sucesión son,

$$n=1 \quad q=0, p=0 \quad X_1 = 1_{]0,1]}$$

$$n=2 \quad q=1, p=0 \quad X_2 = 1_{]0, \frac{1}{2}]}$$

$$n=3 \quad q=1, p=1 \quad X_3 = 1_{] \frac{1}{2}, 1]}$$

$$n=4 \quad q=2, p=0 \quad X_4 = 1_{]0, \frac{1}{4}]}$$

$$n=5 \quad q=2, p=1 \quad X_5 = 1_{] \frac{1}{4}, \frac{1}{2}]}$$

$$n=6 \quad q=2, p=2 \quad X_6 = 1_{] \frac{1}{2}, \frac{3}{4}]}$$

$$n=7 \quad q=2, p=3 \quad X_7 = 1_{] \frac{3}{4}, 1]}$$

...

Se observa que si $X = 0$, $\forall \delta > 0$ se tiene $\lambda\{\omega: |X_n(\omega)| \geq \delta\} = \lambda\{\omega: |X_n(\omega)| = 1\} = \lambda(I_n) = 2^{-q}$, $2^q \leq n < 2^{q+1}$ y $X_n \xrightarrow{\lambda} 0$; pero dada la construcción de las X_n en ningún $\omega \in [0, 1]$ se verifica $\lim X_n(\omega) = 0$.

Las convergencias casi y en media cuadrática no se implican mutuamente.

Ejemplo 3.4 (La convergencia a.s. \Leftrightarrow la convergencia m.s) Con la misma sucesión del ejemplo anterior, como $X_n = 1_{I_n}$, con $I_n =]\frac{p}{2^{q(n)}, \frac{p+1}{2^{q(n)}}]$,

$$E(X_n^2) = \frac{1}{2^{q(n)}},$$

Con $\lim_n q(n) = \infty$. En definitiva, $E(X_n^2) \xrightarrow{n} 0$, lo que pone de manifiesto que la convergencia en media cuadrática \nrightarrow la convergencia casi segura.

Definamos ahora $X_n = \sqrt{n} 1_{]0, \frac{1}{n}]}$. Claramente $X_n \xrightarrow{a.s.} 0$ pero $E(X_n^2) = 1$,

$\forall n$ y $X_n \not\xrightarrow{a.s.} 0$.

La convergencia en ley (débil) no implica la convergencia de probabilidad.

BR

Ejemplo 3.5: Consideremos una variable Bernoulli con $p=1/2$, $X \sim B(1,1/2)$ y definamos una sucesión de variables, $X_n = X, \forall n$. La variable $Y=1-X$ tiene la misma distribución que X , es decir, $Y \sim B(1,1/2)$. Obviamente $X_n \xrightarrow{L} Y$, pero como $|X_n - Y| = |2X - 1| = 1$, no puede haber convergencia en probabilidad.

3.3. Leyes de los Grandes Números

El nombre de leyes de los grandes números hace referencia al estudio de un tipo especial de límites derivados de la sucesión de variables aleatorias $\{X_n\}$.

Concretamente los de la forma $\lim_n \frac{S_n - a_n}{b_n}$, con $S_n = \sum_{i=1}^n X_i$ y siendo $\{a_n\}$

y $\{b_n\}$ sucesiones de constantes tales que $\lim b_n = +\infty$. En esta sección fijaremos las condiciones para saber cuándo existe convergencia a.s. y como nos ocuparemos también de la convergencia en probabilidad, las leyes se denominaran fuerte y débil, respectivamente.

Teorema 3.2. (Ley débil) Sea $\{X_k\}$ una sucesión de variables aleatorias independientes tales que $E(X_k^2) < +\infty, \forall k$, y $\lim_n \frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k) = 0$, entonces:

$$\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{P} 0.$$

Demostración.- Para $S_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k))$, $E(S_n) = 0$ y $\text{var}(S_n) = \frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k)$. Por la desigualdad de chebyshev, $\forall \varepsilon > 0$:

$$P(|S_n| \geq \varepsilon) \leq \frac{\text{var}(S_n)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^n \text{var}(X_k),$$

Que al pasar al límite nos asegura la convergencia en probabilidad de S_n a 0.

Corolario 3.1 Si las X_n son *i.i.d.* con varianza finita y esperanza común $E(X_1)$, entonces $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} E(X_1)$.



Demostración.- Si $\text{var}(X_k) = \sigma^2$, $\forall k$, tendremos $\frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k) = \frac{\sigma^2}{n}$ que tiende a cero con n . Es por tanto de aplicación la ley débil que conduce al resultado enunciado.

Este resultado fue demostrado por primera vez por J. Bernoulli para variables con distribución Binomial (véase el ejemplo 3.1), versión que se conoce como la ley de los grandes números de Bernoulli.

El siguiente paso será fijar las condiciones para que el resultado sea válido bajo convergencia a.s.

Teorema 3.3 (Ley fuerte): Si $\{X_k\}$ es una sucesión de variables aleatorias *i.i.d.* con media finita, entonces

$$\sum_{k=1}^n \frac{X_k}{n} \xrightarrow{\text{a.s.}} E(X_1).$$

Corolario 3.2 Si $\{X_k\}$ es una sucesión de variables aleatorias *i.i.d.* con

$$E(X_1^-) < +\infty \text{ y } E(X_1^+) < +\infty, \text{ entonces } \frac{S_n}{n} \xrightarrow{\text{a.s.}} \infty.$$

La demostración de la ley fuerte es de una complejidad, aun en su versión más sencilla debida a Etemadi. La demostración más compleja, se debe a Kolmogorov y es el resultado final de una cadena de propiedades previas de gran interés y utilidad en sí mismas.

3.4. Teorema Central del Límite

Una aplicación inmediata es el Teorema de De Moivre-Laplace, una versión temprana del TCL, que estudia el comportamiento asintótico de una $B(n, p)$.

Teorema 3.4 (De Moivre-Laplace) Sea $X_n \sim B(n, p)$ y definamos

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}. \text{ Entonces:}$$

$$Z_n \xrightarrow{L} N(0,1).$$

Demostración. - Aplicando los resultados anteriores, se obtiene:

$$\phi_{Z_n}(t) = \left((1-p)e^{-it\sqrt{\frac{p}{n(1-p)}}} + pe^{it\sqrt{\frac{(1-p)}{np}}} \right)^n,$$

que admite un desarrollo en serie de potencias de la forma:

$$\phi_{Z_n}(t) = \left[1 - \frac{t^2}{2n}(1 + R_n) \right]^n,$$

Con $R_n \rightarrow 0$, si $n \rightarrow \infty$. En consecuencia:

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}}.$$

La unicidad y el teorema de continuidad culmina con la demostración.

Observación 3.1 El teorema afirma que si $X \sim B(n, p)$, para n suficientemente grande, tenemos

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) \approx \Phi(x).$$

Donde $\Phi(x)$ es la función de distribución de la $N(0,1)$.

¿De qué forma puede generalizarse este resultado? Se sabe que $X_n \sim B(n, p)$ es la suma de n v.a. *i.i.d.*, todas ellas Bernoulli ($Y_k \sim B(1, p)$), cuya varianza común, $\text{var}(Y_1) = p(1-p)$, es finita. Nos lleva a generalizar variables independientes, con igual distribución y con varianza finita.

Teorema 3.5 (Lindeberg) Sean X_1, X_2, \dots, X_n , v.a. *i.i.d.* con media y varianza finitas, μ y σ^2 , respectivamente. Sea $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ su media muestral, entonces:

$$Y_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{var}(\bar{X}_n)}} \xrightarrow{L} N(0,1).$$

Demostración. – Teniendo en cuenta la definición de \bar{X}_n podemos escribir

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k,$$

Con $Z_k = (X_k - \mu) / \sigma$, variables aleatorias *i.i.d.* con $E(Z_1) = 0$ y $\text{var}(Z_1) = 1$:

$$\phi_{Y_n}(t) = \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right) \right]^n$$

Existiendo los dos primeros momentos de Z_1 también expresarse de la forma:

$$\phi_{Z_1}(t) = 1 - \frac{t^2}{2n}(1 + R_n),$$

Con $R_n \rightarrow 0$, si $n \rightarrow \infty$. En consecuencia:

$$\phi_{Y_n}(t) = \left[1 - \frac{t^2}{2n}(1 + R_n) \right]^n.$$

Así pues: $\lim_{n \rightarrow \infty} \phi_{Y_n}(t) = e^{-\frac{t^2}{2}}$.

Que es la función característica de una $N(0,1)$.

El teorema de De Moivre-Laplace es un caso particular del Teorema de Lindeberg, cuya importancia dice, sea cual sea la distribución común a las X_i , su medida muestral \bar{X}_n , adecuadamente tipificada, converge a una $N(0,1)$ cuando $n \rightarrow \infty$.

El teorema de Lindeberg, que puede considerarse el teorema central del límite básico, admite una generalización a las variables. Las llamadas condiciones de Lindeberg y Lyapunov muestran sendos resultados que permiten eliminar aquella condición.

Ejemplo 3.6 (La fórmula de Stirling para aproximar n!) Consideremos una sucesión de variables aleatorias X_1, X_2, \dots , independientes e idénticamente distribuidas, Poisson de parámetro $\lambda = 1$. La variable $S_n = \sum_{i=1}^n X_i$ es también Poisson con parámetro $\lambda_n = n$. Si $Z \sim N(0,1)$, para n suficientemente grande el TCL nos permite escribir:

$$\begin{aligned} P(S_n = n) &= P(n-1 < S_n \leq n) \\ &= P\left(-\frac{1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leq 0\right) \\ &\approx P\left(-\frac{1}{\sqrt{n}} < Z \leq 0\right) \\ &= \frac{1}{2\pi} \int_{-1/\sqrt{n}}^0 e^{-x^2/2} dx \approx \frac{1}{2\pi} \frac{1}{\sqrt{n}}, \end{aligned}$$

En donde la última expresión surge de aproximar la integral entre $[-1/\sqrt{n}, 0]$ de $f(x) = e^{-x^2/2}$ mediante el área del rectángulo que tiene por base el intervalo de integración y por altura el $f(0) = 1$.

Por otra parte,

$$P(S_n = n) = e^{-n} \frac{n^n}{n!}.$$

Igualando ambos resultados y despejando $n!$ se obtiene la llamada fórmula de Stirling:

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}.$$

3.4.1. Aplicación del TCL: Estimación del valor de π

De Moivre y Laplace dieron en primer lugar una versión local del TCL al demostrar que si $X \sim B(n, p)$,

$$P(X = m) \sqrt{np(1-p)} \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad (3.1)$$

Para n suficientemente grande y $x = \frac{m - np}{\sqrt{np(1-p)}}$. Esta aproximación

nos va a servir para estudiar la credibilidad de algunas aproximaciones al número π obtenidas a partir del problema de la aguja de Buffon.

El problema planteado por Buffon se pretende calcular la probabilidad de que una aguja de longitud $2l$, lanzada al azar sobre una trama de paralelas separadas entre sí una distancia $2a$, con $a > l$, corte a alguna de las paralelas.

La respuesta es: $P(\text{corte}) = \frac{2l}{a\pi}$.

Resultado que permite obtener una aproximación de π si, conocidos a y l , sustituimos en $\pi = \frac{2l}{aP(\text{corte})}$ la probabilidad de corte por su

estimador natural la frecuencia relativa de corte, p , a lo largo de n lanzamientos. Podremos escribir, si en lugar de trabajar con π lo hacemos con su inverso:

$$\frac{1}{\pi} = \frac{am}{2 \ln}$$



Donde m es el número de cortes en los n lanzamientos.

El año 1901 Lazzarini realizó 3408 lanzamientos obteniendo para π el valor de 3,1415929 con ¡6 cifras decimales exactas!. La aproximación es tan buena que merece como mínimo alguna pequeña reflexión. Se supuso que el número de cortes aumenta en una unidad, las aproximaciones de los inversos de π correspondientes a los m y $m+1$ cortes diferirían en:

$$\frac{a(m+1)}{2l.n} - \frac{am}{2l.n} = \frac{a}{2l.n} \geq \frac{1}{2n}.$$

Que si $n \approx 5000$, da lugar a $\frac{1}{2n} \approx 10^{-4}$. Es decir, un corte más produce

una diferencia mayor que la precisión de 10^{-6} alcanzada. Lazzarini tuvo la suerte de obtener exactamente el número de cortes, m , que conducía a tan excelente aproximación. La pregunta inmediata es, cual es la probabilidad de que ello ocurriera:

$$P(X = m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}} \leq \frac{1}{\sqrt{2\pi np(1-p)}},$$

que suponiendo $a = 2l$ y $p = 1/\pi$ nos da para $P(X = m)$ la siguiente cota:

$$P(X = m) \leq \frac{\pi}{\sqrt{2\pi np(\pi-1)}}.$$

Para el caso de Lazzarini $n = 3408$ y $P(X = m) \leq 0,0146, \forall m$.

CAPÍTULO IV

PROCESOS ESTOCÁSTICOS.

4.1. Introducción

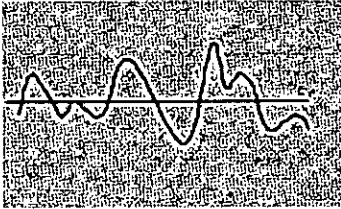
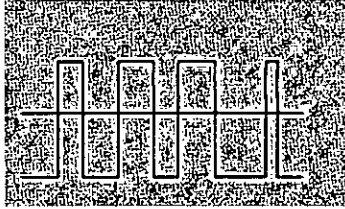
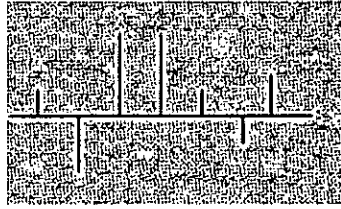
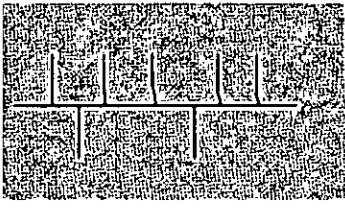
Los temas precedentes nos han dotado de las herramientas necesarias para conocer el comportamiento de una variable aleatoria o de un conjunto finito de ellas, un vector aleatorio. En cualquier caso, se trataba de un número finito de variables aleatorias. El capítulo 3 nos ha permitido, además, estudiar el comportamiento asintótico de una sucesión de variables aleatorias a través de las llamadas Leyes de los Grandes Números y del Teorema Central del Límite.



Antes de introducir las definiciones básicas, conviene puntualizar que el concepto de proceso estocástico abarca situaciones más generales que las sucesiones de variables aleatorias. Se trata de estudiar el comportamiento de un conjunto de variables aleatorias que puede estar indexado por un conjunto no numerable.

Un ejemplo sería el estudio de la ocurrencia de un mismo suceso a lo largo del tiempo, si dichas ocurrencias son independientes y ocupados en el intervalo de tiempo que transcurre entre una y otra, nuestro interés se centra en la sucesión $\{X_n\}_{n \geq 1}$, donde $X_i = \{\text{Tiempo transcurrido entre las ocurrencias } i\text{-ésima y la } i-1\text{-ésima}\}$. Si lo que nos interesa es el número de veces que el suceso ha ocurrido en el intervalo $[0, t]$, la familia a considerar es \mathbb{N}_t con $t > 0$, que desde luego no es numerable.

TABLA N° 4.1
 “ESTADO Y TIEMPO(DISCRETO Y CONTINUO)”

		PROCESO ESTOCÁSTICO	
		ESTADO	
		CONTINUO	DISCRETO
TIEMPO	CONTINUO		
	DISCRETO		

FUENTE: Elaboración propia.

4.2. Definiciones básicas y descripción de un proceso estocástico

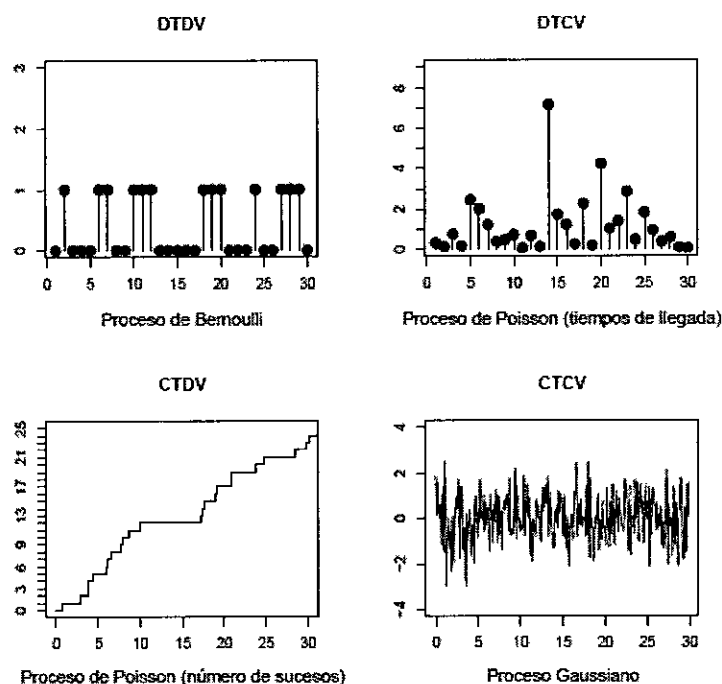
Definición.- Un proceso estocástico es una colección de variables aleatorias $\{X_t, t \in T\}$ definidas sobre un espacio de probabilidad (Ω, A, P) .

Obsérvese que el índice y el conjunto de índices se denotan mediante las letras t y T , respectivamente. La razón para ello es que su origen los procesos estocásticos surgen del estudio de la evolución temporal de fenómenos aleatorios. Ello no presupone nada respecto a la numerabilidad de T .

Una primera clasificación de los procesos estocásticos toma como criterios el tipo de variables involucradas y la dimensión del índice. Para esto se estableció cuatro tipos:

FIGURA N° 4.1

“EJEMPLOS DE LOS DIFERENTES TIPOS DE PROCESOS ESTOCASTICOS”.



- DTDV, acrónimo del inglés *Discret Time / Discret Values*, procesos con índice numerable y variables discretas.
- DTCV, acrónimo del inglés *Discret Time / Continuous Values*, procesos con índice numerable y variables continuas.
- CTDV, acrónimo del inglés *Continuous Time / Discret Values*, procesos con índice no numerable y variables discretas.
- CTCV, acrónimo del inglés *Continuous Time / Continuous Values*, procesos con índice no numerables y variables continuas discretas.

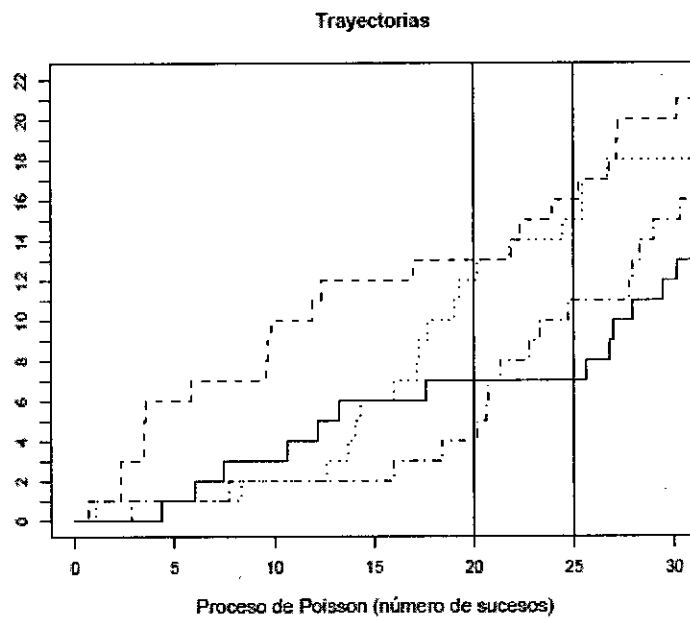
La figura 4.1 muestra la gráfica de un proceso de cada uno de los cuatro tipos.

4.2.1. Trayectoria de un proceso

Un proceso estocástico puede también ser visto como una función aleatoria con un doble argumento, $\{X(t, \omega), t \in T, \omega \in \Omega\}$. Si con esta notación fijamos $\omega = \omega_0$, tendremos una realización del proceso, $X(\cdot, \omega_0)$, cuya representación gráfica constituye lo que denominamos la trayectoria del proceso (sample path). La figura 4.2 muestra cuatro trayectorias de un proceso de Poisson.

FIGURA N° 4.2

“TRAYECTORIAS DE UN PROCESO DE POISSON Y REALIZACIONES DE LAS VARIABLES N_{20} Y N_{25} ”



Por el contrario, si lo que fijamos es $t = t_0$, estamos refiriéndonos a la variable aleatoria $X_{t_0} = X(t_0, \cdot)$. Las líneas verticales que aparecen en la Figura 4.2 representan a las variables aleatorias N_{20} y N_{25} , su intersección con las cuatro trayectorias del proceso son los valores que dichas variables han tomado en cada realización.

AP

4.2.2. Distribuciones Finito-Dimensionales

Un proceso estocástico se describe a partir de las distribuciones de probabilidad que induce sobre R^n . Si $\{t_1, t_2, \dots, t_n\} \subset T$ es un subconjunto finito cualquiera de índices, la distribución conjunta del vector (t_1, t_2, \dots, t_n) vendrá dada por:

$$F_{t_1, t_2, \dots, t_n}(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = P(X_{t_1} \leq x_{t_1}, \dots, X_{t_n} \leq x_{t_n}), \quad \dots(4.1)$$

Que recibe el nombre de distribución finito-dimensional del proceso. Estas distribuciones pueden igualmente venir dadas en términos de las funciones de densidad o probabilidad conjunta del vector.

Un proceso estocástico se describe especificando sus distribuciones finito-dimensionales, que permiten obtener la probabilidad de cualquier suceso involucrado en el proceso. No obstante, que el conjunto de distribuciones finito-dimensionales no determina por completo las características del proceso, por lo que en ocasiones hay que definir ciertas propiedades adicionales.

Si el proceso se puede especificar completamente a partir de dichas distribuciones, decimos que el proceso es separable.

Dos de estas propiedades son las que conocen como *incrementos independientes* y de Markov.

Definición 4.2 (Incrementos Independientes) :Se dice que un proceso estocástico tiene sus incrementos independientes si para $t_1 < t_2 < \dots < t_n$, las variables:

$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}, \quad \dots(4.2)$$

Son independientes.

Definición 4.3 (Markov): Se dice que un proceso estocástico es un proceso de Markov si la evolución del proceso depende solo del pasado inmediato, es decir, datos $t_1 < t_2 < \dots < t_n$

$$P(X_{t_n} \in B | X_{t_1}, X_{t_2}, \dots, X_{t_{n-1}}) = P(X_{t_n} \in B | X_{t_{n-1}}) \quad \dots(4.3)$$

Donde B es cualquier conjunto de Borel (suceso) en R .

La condición (4.3) puede igualmente expresarse en términos de las funciones de densidad o de probabilidad, según sea la naturaleza de las variables del proceso,

$$f_{t_n|t_1, \dots, t_{n-1}}(x_n | x_{t_1}, \dots, x_{t_{n-1}}) = f_{t_n|t_{n-1}}(x_n | x_{t_{n-1}}).$$

En el caso discreto las probabilidades $P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1})$ se denominan probabilidades de transición, cuyo conocimiento caracteriza completamente el proceso de Markov.

Estas propiedades están relacionadas. Los incrementos independientes implican la propiedad de Markov, pero el recíproco no es cierto. En el caso discreto la implicación se demostró:

$$\begin{aligned} P(X_{t_n} = x_n | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) &= \\ &= \\ P(X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1} | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) & \quad (4.4) \end{aligned}$$

$$= P(X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1} | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) \quad (4.5)$$

$$= P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1}), \quad (4.6)$$

Del paso de (4.4) a (4.5) es consecuencia de la independencia de los incrementos.

4.2.3. Funciones de Momento

Las llamadas funciones de momento, obtenidas a partir de los momentos de variables involucradas en un proceso estocástico, juegan un papel muy importante a la hora de conocer su comportamiento y en las aplicaciones del mismo. Se tienen las siguientes:

➤ **Función media.** - Se define como:

$$\mu_X(t) = E[X_t], t \in T. \quad \dots(4.7)$$

Para su obtención tendremos en cuenta el tipo de variables que conforman el proceso:

En el caso discreto,

$$\mu_X(t) = \sum_{x \in D_{X_t}} xP(X_t = x)$$

Donde D_{X_t} es el soporte de X_t .

En el caso continuo:



$$\mu(t) = \int_{-\infty}^{\infty} xf_i(x)dx.$$

- **Función de autocorrelacion.** - Se define a partir del momento conjunto de dos variables asociadas a dos tiempos cualesquiera, t_1 y t_2 ,

$$R(t_1, t_2) = E[X_{t_1} X_{t_2}]. \quad \dots(4.8)$$

Para el caso discreto (4.8) se obtiene mediante:

$$R(t_1, t_2) = \sum_{x_1 \in D_{X_{t_1}}, x_2 \in D_{X_{t_2}}} x_1 x_2 P(X_{t_1} = x_1, X_{t_2} = x_2).$$

En el caso continuo:

$$R(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{t_1, t_2}(x_1, x_2) dx_1 dx_2.$$

- **Función de autocovarianza.** - Se define a partir del momento central conjunto de dos variables asociadas a dos tiempos cualquiera, t_1 y t_2 ,

$$C(t_1, t_2) = E[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))], \quad \dots(4.9)$$

Con sus correspondientes versiones discreta y continua. Se induce fácilmente la siguiente relación $R(t_1, t_2)$ y $C(t_1, t_2)$,

$$C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2).$$

El caso particular $t_1 = t_2 = t$ recibe el nombre de *función varianza*, $\sigma^2(t) = C(t, t)$. La función de correlación se obtiene mediante:

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{\sigma^2(t_1)\sigma^2(t_2)}},$$

Las definiciones anteriores tienen sentido si las correspondientes integrales o series son finitas.

Ejemplo 4.1.- Consideremos el siguiente proceso estocástico del tipo CTCV,

$$X_t = A \sin(\omega_0 t + \Theta), t \in R,$$

Donde A y Θ son variables aleatorias independientes, con $\Theta \sim U(-\pi, \pi)$. Obtengamos sus momentos.

$$\begin{aligned} \mu(t) &= E[A \operatorname{sen}(\omega_0 t + \Theta)] \\ &= E(A)E[\operatorname{sen}(\omega_0 t + \Theta)] \\ &= \mu_A \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{sen}(\omega_0 t + \theta) d\theta \\ &= \mu_A \cdot 0 = 0. \end{aligned}$$

La autocorrelación vale:

$$R(t_1, t_2) = E[X_{t_1} X_{t_2}]$$

$$\begin{aligned} &= E[A^2 \operatorname{sen}(\omega_0 t_1 + \Theta) \operatorname{sen}(\omega_0 t_2 + \Theta)] \\ &= E(A^2) E[\operatorname{sen}(\omega_0 t_1 + \Theta) \operatorname{sen}(\omega_0 t_2 + \Theta)] \quad \dots (4.10) \end{aligned}$$

$$\begin{aligned} &= E(A^2) \cdot \frac{1}{2} \left\{ E[\cos \omega_0 (t_1 - t_2)] - E[\cos(\omega_0 (t_1 + t_2) + 2\Theta)] \right\} \quad \dots (4.11) \\ &= \frac{1}{2} E(A^2) \cos \omega_0 (t_1 - t_2). \end{aligned}$$

Para pasar de (4.10) a (4.11) hemos aplicado:

$$\operatorname{sen} \alpha \operatorname{sen} \beta = \frac{\cos(\alpha - \beta) - \cos(\alpha + \beta)}{2}.$$

Observe que $\mu(t) = 0$ es constante y que $R(t_1, t_2)$ depende de $t_1 - t_2$ en realidad de $|t_1 - t_2|$ porque $\cos(\alpha) = \cos(-\alpha)$. Un proceso de estas características se dice que es estacionario en sentido amplio (wide-sense stationary, WSS).



CAPÍTULO V

ALGUNOS PROCESOS ESTOCÁSTICOS DE INTERÉS

5.1. Procesos IID

Se trata de procesos constituidos por variables aleatorias independientes e idénticamente distribuidas (IID). Como consecuencia de ello, las distribuciones finito – dimensionales se obtienen fácilmente a partir de la distribución común. Si $F(x)$ denota la función de distribución común a todas las X_t del proceso,

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = F(x_1)F(x_2) \dots F(x_n).$$

La función media es:

$$\mu(t) = E(X_t) = \mu,$$

Donde μ es la esperanza común a todas las X_t .

La función de autocovarianza vale 0 porque,

$$C(t_1, t_2) = E[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))] = E[(X_{t_1} - \mu(t_1))]E[(X_{t_2} - \mu(t_2))] = 0,$$

y para $t_1 = t_2 = t$ da lugar a la varianza común, $\sigma^2(t) = \sigma^2$. Es decir:

$$C(t_1, t_2) = \begin{cases} 0 & , \text{si } t_1 \neq t_2; \\ \sigma^2 & , \text{si } t_1 = t_2. \end{cases}$$

La función de autocorrelacion es:

$$R(t_1, t_2) = \begin{cases} \mu^2 & , \text{si } t_1 \neq t_2; \\ \sigma^2 + \mu^2 & , \text{si } t_1 = t_2. \end{cases}$$

Entre los procesos estocásticos IID existen algunos de especial interés.

- **Proceso de Bernoulli.**- Se trata de una sucesión ($T = N$) de variables Bernoulli independientes, $X_n \sim B(1, p)$. La sucesión de lanzamientos de una moneda (cara = 1, cruz = 0) da lugar a un proceso de Bernoulli. La media y la varianza del proceso valen

$$\mu(n) = p, \quad \sigma^2(n) = p(1-p).$$

La probabilidad que los cuatro primeros lanzamientos sean $\{C++C\} = \{1001\}$ es:

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) = P(X_1 = 0)P(X_2 = 1)P(X_3 = 1)P(X_4 = 0) \\ = p^2(1-p)^2.$$

- **Procesos suma IID**

La suma de procesos estocásticos es una forma de obtener nuevos procesos de interés.

$$S_n = X_1 + X_2 + X_3 + \dots + X_n, \quad n \geq 1,$$

O también:

$$S_n = S_{n-1} + X_n, \text{ Con } S_0 = 0.$$

Si el proceso original es un proceso IID, el proceso suma tiene incrementos independientes para intervalos de tiempo no solapados. El caso en que las X_k son discretas, si $n_0 < n \leq n_1$ y $n_2 < n \leq n_3$, con $n_1 \leq n_2$,

$$S_{n_1} - S_{n_0} = X_{n_0+1} + \dots + X_{n_1}$$

$$S_{n_3} - S_{n_2} = X_{n_2+1} + \dots + X_{n_3}.$$

Cada sumando está constituido por variables que son independientes de las del otro, lo que supone la independencia de los incrementos:

$$S_{n_2} - S_{n_1} = X_{n_1+1} + \dots + X_{n_2} = X_1 + X_2 + \dots + X_{n_2-n_1} = S_{n_2-n_1}.$$

Si las variables son discretas lo será la suma por la expresión anterior tendremos,

$$P(S_{n_2} - S_{n_1} = y) = P(S_{n_2-n_1} = y).$$

La probabilidad depende tan solo de la longitud del intervalo y no de su posición, por lo que el proceso suma tiene incrementos estacionarios.

Las distribuciones finito-dimensionales del proceso suma se obtienen haciendo uso de la propiedad de incrementos estacionarios independientes. Si $n_1 < n_2 < \dots < n_k$,

$$P(S_{n_1} = x_1, S_{n_2} = x_2, \dots, S_{n_k} = x_k) = \\ = P(S_{n_1} = x_1, S_{n_2} - S_{n_1} = x_2 - x_1, \dots, S_{n_k} - S_{n_{k-1}} = x_k - x_{k-1}) \\ = P(S_{n_1} = x_1)P(S_{n_2} - S_{n_1} = x_2 - x_1) \dots P(S_{n_k} - S_{n_{k-1}} = x_k - x_{k-1}) \\ = P(S_{n_1} = x_1)P(S_{n_2-n_1} = x_2 - x_1) \dots P(S_{n_k-n_{k-1}} = x_k - x_{k-1}).$$



Las variables del proceso original son continuas, es fácil obtener una expresión equivalente para la densidad conjunta:

$$f_{S_{n_1}, S_{n_2}, \dots, S_{n_k}}(x_1, x_2, \dots, x_k) = f_{S_{n_1}}(x_1) f_{S_{n_2-n_1}}(x_2 - x_1) \dots f_{S_{n_k-n_{k-1}}}(x_k - x_{k-1}).$$

$$\mu(n) = E(S_n) = n\mu, \quad \sigma^2(n) = \text{var}(S_n) = n\sigma^2,$$

Donde μ y σ^2 son la media y varianza comunes a las X_k . La función de autocovarianza es:

$$\begin{aligned} C(n_1, n_2) &= E[(S_{n_1} - n_1\mu)(S_{n_2} - n_2\mu)] \\ &= E\left[\left\{\sum_{i=1}^{n_1} (X_i - \mu)\right\}\left\{\sum_{j=1}^{n_2} (X_j - \mu)\right\}\right] \\ &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} E[(X_i - \mu)(X_j - \mu)] \\ &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \text{cov}(X_i, X_j). \end{aligned}$$

Como X_i y X_j son IID:

$$\text{cov}(X_i, X_j) = \begin{cases} \sigma^2 & , \text{si } i = j; \\ 0 & , \text{si } i \neq j. \end{cases}$$

En definitiva:

$$C(n_1, n_2) = \sum_{i=1}^{\min(n_1, n_2)} \sigma^2 = \min(n_1, n_2) \sigma^2.$$

Veamos tres ejemplos interesantes de procesos suma.

- **Camino aleatorio.-** Se trata de proceso estocástico que describe el desplazamiento aleatorio de un móvil en R^n . En el caso unidimensional, del que nos ocuparemos por más sencillo, el móvil se desplaza a saltos unitarios e independientes por la recta real, haciéndolo hacia la derecha con probabilidad p y hacia la izquierda con probabilidad $1-p$. Al cabo de n desplazamientos, la posición del móvil viene dada por $S_n = D_1 + D_2 + \dots + D_n$, donde D_k tiene por función de probabilidad:

$$f_k(x) = \begin{cases} 1-p, & \text{si } x = -1; \\ p, & \text{si } x = 1; \\ 0, & \text{en el resto.} \end{cases}$$

Si en los n desplazamientos k de ellos sido hacia la derecha y los restantes $n-k$ en sentido contrario, la posición final será $S_n = k - (n-k) = 2k - n$ por lo que su función de probabilidad valdrá:

$$f_{S_n}(2k - n) = P(S_n = 2k - n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

La media y la varianza del camino aleatorio unidimensional:

$$\mu(n) = E(S_n) = \sum_{k=1}^n E(D_k) = n(2p - 1),$$

Y

$$\sigma_n^2 = \text{var}(S_n) = \sum_{k=1}^n \text{var}(D_k) = 4np(1-p).$$

$$C(n_1, n_2) = \min(n_1, n_2)4p(1-p),$$

Y la autocorrelación:

$$R(n_1, n_2) = C(n_1, n_2) + \mu(n_1)\mu(n_2) = \min(n_1, n_2)4p(1-p) + n_1n_2(2p-1)^2.$$

- **Proceso Binomial.** - Si el proceso original es un proceso Bernoulli, el proceso suma resultante es el proceso de recuento o Binomial, que proporciona el número de éxitos entre las n primeras pruebas de Bernoulli. Se llama así porque, como bien sabemos, las variables del nuevo proceso son $S_n \sim B(n, p)$.

$$\begin{aligned} P(S_{n_1} = m_1, S_{n_2} = m_2, \dots, S_{n_k} = m_k) &= \\ &= \\ P(S_{n_1} = m_1)P(S_{n_2 - n_1} = m_2 - m_1) \dots P(S_{n_k - n_{k-1}} = m_k - m_{k-1}) &= \\ &= \binom{n_1}{m_1} \binom{n_2 - n_1}{m_2 - m_1} \dots \binom{n_k - n_{k-1}}{m_k - m_{k-1}} p^{m_k} (1-p)^{n_k - m_k}. \end{aligned}$$

La medida del proceso Binomial vale:

$$\mu(n) = E(S_n) = np,$$

Que no es constante y crece linealmente con n . Su varianza es:

$$\sigma^2(n) = \text{var}(S_n) = np(1-p),$$

Y las funciones de autocovarianza y autocorrelación:

$$C(n_1, n_2) = \min(n_1, n_2)p(1-p),$$

$$R(n_1, n_2) = \min(n_1, n_2)p(1-p) + n_1 n_2 p^2.$$

Proceso suma Gaussiano.- Se obtiene a partir de la suma de variable $X_k \sim N(0, \sigma^2)$, lo que implica que $S_n \sim N(0, n\sigma^2)$.

$$f_{S_{n_1}, S_{n_2}, \dots, S_{n_k}}(x_1, x_2, \dots, x_k) = \frac{1}{\sigma \sqrt{2\pi n_1}} e^{-\frac{x_1^2}{2n_1 \sigma^2}} \dots \frac{1}{\sqrt{2\pi(n_k - n_{k-1})}} e^{-\frac{(x_k - x_{k-1})^2}{2(x_k - x_{k-1})^2}}.$$

Las funciones de momento son:

$$\mu(n) = 0, \quad \sigma^2(n) = n\sigma^2, \quad C(n_1, n_2) = R(n_1, n_2) = \min(n_1, n_2)\sigma^2.$$

5.2 Ruido Blanco

Un ruido blanco es un proceso estocástico con media cero, $\mu(t) = 0$, varianza constante, $\sigma^2(t) = \sigma^2$ y con componentes incorreladas. Como consecuencia de ello, la función de autocovarianza y autocorrelación coinciden y valen:

$$R(t_1, t_2) = C(t_1, t_2) = \begin{cases} \sigma^2 & , t_1 = t_2; \\ 0 & , t_1 \neq t_2. \end{cases}$$

La definición es válida tanto si se trata de una sucesión, t discreto, como si t es continuo. Un proceso IID es, según esta definición, un ruido blanco, porque la independencia de las variables implica la incorrelación. En otros casos se da una definición más restrictiva de este tipo de procesos, para determinar que las variables sean independientes. A continuación, en el caso de que las variables sean normales o gaussianas ambas definiciones son equivalentes.

5.3. Proceso Gaussiano

Se trata de un proceso en el que sus variables $X_i \sim N(\mu(t), \sigma^2(t))$ y sus distribuciones finito-dimensionales son normales multivariantes:

$$f(x_1, x_2, \dots, x_n) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)},$$

Donde:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1(n-1)} & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2(n-1)} & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{1(n-1)} & \sigma_{2(n-1)} & \dots & \sigma_{n-1}^2 & \sigma_{(n-1)n} \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_{(n-1)n} & \sigma_n^2 \end{pmatrix},$$

Son el vector de medidas y la matriz de covarianzas del vector $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$. Esta definición es válida tanto para tiempos discretos como continuos.

Si el proceso Gaussiano es tal que $\mu_i = 0, \forall i, \sigma_{ij} = 0, i \neq j$ y $\sigma_i^2 = \sigma^2, \forall i$; es un caso particular de ruido blanco, llamado ruido blanco Gaussiano, algunas de las propiedades de la Normal multivariante, en la que icorrelación equivale a independencia, está constituido por variables independientes.

La importancia del proceso Gaussiano se debe en parte a sus propiedades, heredadas de las propiedades de la Normal. Entre las más destacadas,

Las distribuciones finito-dimensionales del proceso están completamente especificadas con los momentos de primer y segundo orden, es decir, μ y Σ ,

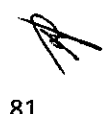
La transformación lineal del proceso da lugar a un nuevo Gaussiano.

Son muchos los fenómenos relacionados con señal y ruido que pueden ser modelizados con éxito utilizando un proceso Gaussiano.

5.4. Proceso De Poisson

Consideremos una sucesión de ocurrencias de un determinado suceso a lo largo del tiempo. Sea X_1 el tiempo de espera necesario para que el suceso ocurra por primera vez, X_2 el tiempo transcurrido entre la primera y la segunda ocurrencia, y en general, X_i el tiempo entre las ocurrencias consecutivas $i-1$ e i . El modelo formal que describe este fenómeno es una sucesión de variables aleatorias definidas sobre un determinado espacio de probabilidad.

Otra característica de interés ligada al fenómeno es el número de sucesos que han ocurrido en un determinado intervalo de tiempo $]t_1, t_2]$. Por ejemplo, para $t_1 = 0$ y $t_2 = t$, definimos la variable $N_t = \{\text{número de sucesos ocurridos hasta el tiempo } t\}$. Es del proceso estocástico que estas variables definen del



que nos vamos a ocupar. Si la sucesión de tiempos de espera verifica las siguientes condiciones iniciales:

C1) No pueden ocurrir dos sucesos simultáneamente.

C2) En cada intervalo finito de tiempo ocurren a lo sumo un número finito de sucesos.

C3) Los tiempos de espera son independientes e idénticamente distribuidos según una $\text{Exp}(\lambda)$.

El proceso recibe el nombre de proceso de Poisson. De tal manera que las variables N_t son variables discretas que toman valores en $\{0,1,2,\dots\}$.

Para estudiar el comportamiento probabilístico de las variables N_t es conveniente recurrir a una nueva sucesión de variables, S_n , que se define a partir de los tiempos de espera entre sucesos:

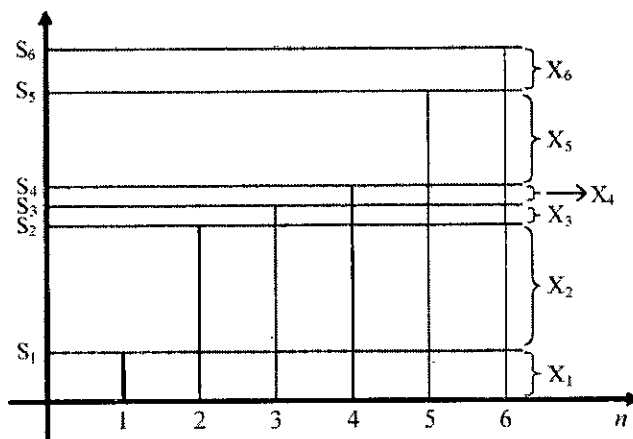
$$S_n = \sum_{i=1}^n X_i, \quad n \geq 0,$$

Con $S_0 = 0$. La variable S_n representa el tiempo transcurrido hasta la llegada del n -ésimo suceso. Como consecuencia de C1 y C2, la sucesión es estrictamente creciente:

$$0 = S_0 < S_1 < S_2 < \dots, \quad \sup_n S_n = \infty.$$

FIGURA N° 5.1

“GRÁFICO LA RELACIÓN ENTRE LAS S_n Y LAS X_n DE UN PROCESO DE POISSON”.



Tiempos entre sucesos, X_i , y tiempos acumulados, S_k , en proceso de Poisson.

La distribución de S_n , al ser suma de exponenciales con parámetro común λ , una $G(n, 1/\lambda)$, conocida como distribución de Erlang cuya densidad es:

$$g_n(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0,$$

Y cuya función de distribución es:

$$G_n(t) = \sum_{k \geq n} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!},$$

El número N_t de sucesos que ha ocurrido en el intervalo $]0, t]$ es el mayor n tal que $S_n \leq t$,

$$N_t = \text{máx}\{n; S_n \leq t\}.$$

Teniendo en cuenta esta relación es fácil deducir que:

$$\{N_t \geq n\} = \{S_n \leq t\},$$

$$P(N_t \geq n) = G_n(t) = \sum_{k \geq n} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Se obtiene:

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

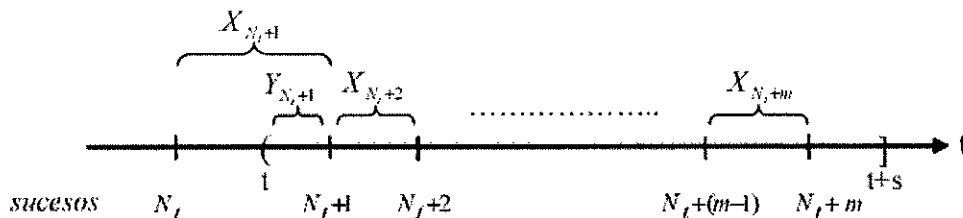
Es decir, $N_t \sim Po(\lambda t)$, lo que justifica el nombre dado al proceso. Este resultado nos permite además darle significado al parámetro λ porque, si recordamos que $E(N_t) = \lambda t$, se deduce de aquí que λ es el número de sucesos que ocurren por unidad de tiempo, una característica específica del fenómeno aleatorio que da lugar al proceso. Algunas propiedades del proceso de Poisson.

➤ **Incrementos independientes y estacionarios**

La sucesión de tiempos acumulados es un proceso suma obtenido a partir de variables exponenciales IID y como tal, tiene incrementos independientes y estacionarios. Dada la relación vista anteriormente entre las N_t y las S_n , es lógico pensar que también el proceso de Poisson goce de la misma propiedad. La implicación no es inmediata si tenemos en cuenta que al considerar los tiempos de espera sobre un intervalo cualquiera $[t, t+s]$, el primero de ellos puede estar contenido solo parcialmente en él, tal como se muestra en la figura 5.2. En ella se observa que $N_{t+s} - N_t = m$ y que parte del tiempo transcurrido entre el suceso N_t y el N_{t+1}, X_{N_t+1} , está en $(t, t+s]$. La comprobación de la probabilidad no es sencilla, razón por la cual no la presentamos.

FIGURA N° 5.2

“RELACIÓN ENTRE LOS TIEMPOS DE ESPERA Y UN INTERVALO ARBITRARIO $(t, t+s]$ EN UN PROCESO DE POISSON”



Resulta sencillo comprobar la estacionariedad de los incrementos. En efecto, dada la independencia entre las X_i , como Y_{N_t+1} depende exclusivamente de X_{N_t+1} y de S_{N_t} ,

$$X_{N_t+1} - Y_{N_t+1} = t - S_{N_t},$$

Y_{N_t+1} será también independiente de todos los tiempos de espera posteriores, X_{N_t+k} , $k \geq 2$. La consecuencia inmediata es que la variable $N_{t+s} - N_t$ está relacionada con la suma de variables exponenciales IID,

$$Y_{N_t+1} + \sum_{k=2}^m X_{N_t+k},$$

$$P(N_{t+s} - N_t = m) = P(N_s = m) = e^{-\lambda s} \frac{(\lambda s)^m}{m!},$$

En forma general, si $t_2 > t_1$, $N_{t_2} - N_{t_1} \sim Po(\lambda(t_2 - t_1))$, que depende sólo del incremento de tiempos y no de su posición. Son pues estacionarios.

➤ **Distribuciones finito-dimensionales y funciones de momento**

La independencia y estacionariedad de los incrementos permite obtener fácilmente las distribuciones finito-dimensionales. Para el caso de dos tiempos cualesquiera, $t_2 > t_1$,

$$\begin{aligned} P(N_{t_1} = n_1, N_{t_2} = n_2) &= P(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2 - n_1) \\ &= P(N_{t_1} = n_1)P(N_{t_2} - N_{t_1} = n_2 - n_1) \\ &= e^{-\lambda t_1} \frac{(\lambda t_1)^{n_1}}{n_1!} e^{-\lambda(t_2 - t_1)} \frac{(\lambda(t_2 - t_1))^{n_2 - n_1}}{(n_2 - n_1)!} \\ &= e^{-\lambda t_2} \frac{\lambda^{n_2} t_1^{n_1} (t_2 - t_1)^{n_2 - n_1}}{n_1! (n_2 - n_1)!}. \end{aligned}$$

Por inducción se obtiene para cualquier número finito de tiempos. Respecto a las funciones de momento:

- **Media del proceso.** -

$$\mu(t) = \lambda t.$$

- **Función de autocorrelación.**- Para dos tiempos cualesquiera $t_2 > t_1$, haciendo uso de la estacionariedad e independencia de los incrementos:

$$\begin{aligned}
E[N_{t_1}N_{t_2}] &= E[(N_{t_1} + (N_{t_2} - N_{t_1}))N_{t_1}] \\
&= E[N_{t_1}^2] + E[N_{t_2} - N_{t_1}]E[N_{t_1}] \\
&= \lambda t_1 + \lambda^2 t_1^2 + \lambda(t_2 - t_1)\lambda t_1 \\
&= \lambda t_1 + \lambda^2 t_1 t_2.
\end{aligned}$$

Si $t_1 > t_2$ intercambiaríamos t_1 y t_2 en la expresión anterior. En consecuencia, la función de autocorrelación vale:

$$R(t_1, t_2) = E[N_{t_1}N_{t_2}] = \lambda \min(t_1, t_2) + \lambda^2 t_1 t_2.$$

- **Función de autocovarianza.**- De la relación entre $R(t_1, t_2)$ y $C(t_1, t_2)$ se deduce:

$$C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2) = \lambda \min(t_1, t_2).$$

Para $t_1 = t_2 = t$ obtendremos la varianza del proceso que, como en caso de la variable Poisson, coincide con la media:

$$\sigma^2(t) = \lambda t.$$

➤ **Proceso de Poisson y llegadas al azar**

El proceso de Poisson surge de la necesidad de modelizar fenómenos aleatorios consistentes en las ocurrencias sucesivas a lo largo del tiempo de un determinado suceso. Las partículas radiactivas que llegan a un contador Geiger son un buen ejemplo. En este tipo de fenómenos es costumbre hablar de “llegadas al azar” del suceso. Para entender el significado de la expresión y al mismo tiempo justificarla, suponga que observamos el proceso en un intervalo $]0, t]$ a través de su variable asociada, N_t , y que el número de llegadas ha sido n , $N_t = n$. Consideremos un subintervalo cualquiera de $]0, t]$, sin pérdida de generalidad podemos hacer coincidir los orígenes de ambos intervalos, $]0, t] \subset [0, t]$, y vamos obtener la probabilidad de que k , de los n sucesos hayan ocurrido en $]0, t_1]$. Haciendo uso de las propiedades del proceso:



$$\begin{aligned}
P(N_{t_1} = k | N_t = n) &= \frac{P(N_{t_1} = k, N_t = n)}{P(N_t = n)} \\
&= \frac{P(N_{t_1} = k, N_{t-t_1} = n-k)}{P(N_t = n)} \\
&= \frac{[e^{-\lambda t} (\lambda t)^k / k!][e^{-\lambda(t-t_1)} (\lambda(t-t_1))^{n-k} / (n-k)!]}{e^{-\lambda t} (\lambda t)^n / n!} \\
&= \frac{n!}{k!(n-k)!} \times \frac{e^{-\lambda t} \lambda^n t_1^k (t-t_1)^{n-k}}{e^{-\lambda t} \lambda^n t^n} \\
&= \binom{n}{k} \left(\frac{t_1}{t}\right)^k \left(\frac{t-t_1}{t}\right)^{n-k}.
\end{aligned}$$

Es decir, $N_{t_1} | N_t = n \sim B(n, p)$ con $p = t_1 / t$, lo que significa que la probabilidad de que cualesquiera de los n sucesos ocurra en $]0, t_1]$ es proporcional a la longitud del subintervalo o, equivalentemente, los n sucesos se distribuyen uniformemente, al azar, en el intervalo $]0, t]$.

Este resultado tiene una aplicación inmediata en la simulación de un proceso de Poisson de parámetro λ . Un sencillo algoritmo para simular las llegadas en un intervalo $]0, t]$ es:

- Generar un valor de una variable $Po(\lambda t)$, lo que nos dará el número de llegadas.
- Si el valor anteriormente simulado es n_o , generar n_o valores de una $U(0, t)$, que determinarán los tiempos de llegada de los n_o sucesos.

Las funciones para generar valores de variables Poisson y Uniformes están disponibles en cualquier operador.

➤ Derivación alternativa del proceso de Poisson

Existe una forma alternativa de obtener el proceso de Poisson basada en resultados elementos de la Teoría de la Probabilidad y estableciendo condiciones iniciales para el fenómeno. Con la notación habitual, estas condiciones son:

CA1) Si $t_1 < t_2 < t_3$, los sucesos $\{N_{t_2-t_1} = n\}$ y $\{N_{t_3-t_2} = m\}$ son independientes, para cualesquiera valores no negativos de n y m ,

CA2) Los sucesos $\{N_{t_2-t_1} = n\}$, $n=0,1,\dots$, constituyen una partición de espacio muestral y $P(N_{t_2-t_1} = n)$ depende sólo de la diferencia $t_2 - t_1$.

CA3) Si t es suficientemente pequeño, entonces $P(N_t \geq 2)$ es despreciablemente pequeña comparada con $P(N_t = 1)$, es decir:

$$\lim_{x \downarrow \infty} \frac{P(N_t \geq 2)}{P(N_t = 1)} = \lim_{x \downarrow \infty} \frac{1 - P(N_t = 0) - P(N_t = 1)}{P(N_t = 1)} = 0,$$

Equivale a:

$$\lim_{x \downarrow \infty} \frac{1 - P(N_t = 0)}{P(N_t = 1)} = 1.$$

➤ **Generalización del proceso de Poisson**

El proceso de Poisson admite generalizaciones en varios sentidos. La más inmediata es aquella que suprime la estacionariedad admitiendo que su intensidad λ , número de ocurrencias por unidad de tiempo, dependa del tiempo, $\lambda(t) > 0, \forall t \geq 0$. Tenemos entonces un proceso de Poisson no uniforme o no homogéneo, en el que su media vale:

$$\mu(t) = \int_0^t \lambda(x) dx, \quad \forall t \geq 0.$$

5.5. Señal telegráfica aleatoria (RTS)

El proceso conocido como RTS, Random Telegraph Signal en inglés, es proceso relacionado con el proceso de Poisson. Se trata de un proceso CTDV en el que la variable X_t toman solo dos valores simétricos, $\{-a, a\}$, de acuerdo con el siguiente esquema.

$$X_0 = \pm a \text{ con igual probabilidad } p = 1/2,$$

X_t cambia de signo con cada ocurrencia de un suceso en un proceso de Poisson de parámetro λ .

Un proceso de estas características surge cuando toda la información de interés en una señal aleatoria está contenida en los puntos donde se produce un cambio de signo, los puntos donde se cruza el eje de las X 's. Como las variaciones de amplitud carecen de interés.

La función de probabilidad de X_t depende del valor N_t , más concretamente de su paridad.

En efecto:

$$P(X_t = \pm a) = P(X_t = \pm a | X_0 = a)P(X_0 = a) + P(X_t = \pm a | X_0 = -a)P(X_0 = -a).$$

Se deduce que X_t y X_0 tomarán el mismo valor si $N_t = par$, en caso contrario tomarán valores opuestos. Así:

$$\begin{aligned} P(X_t = \pm a | X_0 = \pm a) &= P(N_t = par) \\ &= \sum_{n \geq 0} e^{-\lambda t} \frac{(\lambda t)^{2n}}{(2n)!}. \end{aligned}$$

Sumar la serie observemos que:

$$e^{\lambda t} + e^{-\lambda t} = \sum_{k \geq 0} \frac{(\lambda t)^k + (-\lambda t)^k}{k!} = 2 \sum_{n \geq 0} \frac{(\lambda t)^{2n}}{(2n)!}.$$

$$P(X_t = \pm a | X_0 = \pm a) = e^{-\lambda t} \frac{e^{\lambda t} + e^{-\lambda t}}{2} = \frac{1}{2}(1 + e^{-2\lambda t}).$$

$$P(X_t = \pm a | X_0 = \mp a) = e^{-\lambda t} \frac{e^{\lambda t} - e^{-\lambda t}}{2} = \frac{1}{2}(1 - e^{-2\lambda t}).$$

$$P(X_t = +a) = \frac{1}{2} \left[\frac{1}{2}(1 + e^{-2\lambda t}) + \frac{1}{2}(1 - e^{-2\lambda t}) \right] = \frac{1}{2},$$

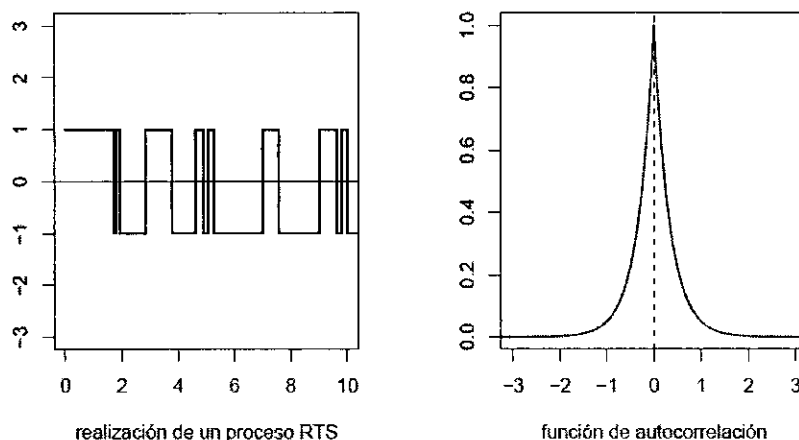
$$P(X_t = -a) = \frac{1}{2} \left[\frac{1}{2}(1 - e^{-2\lambda t}) + \frac{1}{2}(1 + e^{-2\lambda t}) \right] = \frac{1}{2},$$

Lo que no dice que el proceso RTS toma cualquiera de los dos valores con igual probabilidad en cualquier instante de tiempo.



FIGURA N° 5.3

“REALIZACIÓN DE UN PROCESO RTS CON $a=1$ Y $\lambda=1,5$ Y SU CORRESPONDIENTE FUNCIÓN DE AUTOCORRELACIÓN”



Distribuciones finito-dimensionales y funciones de momento

Para dos tiempos cualesquiera t y s .

Si $t > s$:

$$f_{st}(x, y) = P(X_s = x, X_t = y) = P(X_s = x | X_t = y)P(X_t = y),$$

Si $s > t$:

$$f_{st}(x, y) = P(X_s = x, X_t = y) = P(X_t = y | X_s = x)P(X_s = x).$$

En cualquiera de los dos casos lo que determina el valor es la paridad de $N_{|t-s|}$,

$$f_{st}(x, y) = \begin{cases} \frac{1}{2}P(N_{|t-s|} = \text{par}) = \frac{1}{2} \times \frac{1}{2}(1 + e^{-2\lambda|t-s|}), & \text{si } x = y; \\ \frac{1}{2}P(N_{|t-s|} = \text{impar}) = \frac{1}{2} \times \frac{1}{2}(1 - e^{-2\lambda|t-s|}), & \text{si } x \neq y. \end{cases}$$

La media del proceso RTS es nula dada la simetría de las variables X_t , $\mu(t) = 0$. La varianza vale:

$$\sigma^2(t) = E(X_t^2) = a^2 \times \frac{1}{2} + (-a)^2 \times \frac{1}{2} = a^2.$$

Las funciones de autocorrelación y autocovarianza coinciden por ser la media nula:

$$\begin{aligned}
 C(t_1, t_2) &= E(X_{t_1} X_{t_2}) \\
 &= a^2 P(X_{t_1} = X_{t_2}) - a^2 P(X_{t_1} \neq X_{t_2}), \\
 &= \frac{a^2}{2} (1 + e^{-2\lambda|t_1 - t_2|}) - \frac{a^2}{2} (1 - e^{-2\lambda|t_1 - t_2|}) \\
 &= a^2 e^{-2\lambda|t_1 - t_2|}.
 \end{aligned}$$

La función se amortigua a medida que aumenta la diferencia entre ambos tiempos, disminuyendo así la correlación entre ambas variables.

5.6. Modulación por desplazamiento de fase (PSK)

La transmisión de datos binarios a través de ciertos medios, la línea telefónica sería un ejemplo, exige una modulación de dichos datos. La modulación por desplazamiento de fase (PSK del inglés Phase-Shift Keying) es un método básico que consiste transformar los datos, una sucesión de variables aleatorias independientes todas ellas Bernoulli uniformes, $B_n \sim B(1, 1/2)$, en una sucesión de ángulo-fase Θ_n que utiliza para modular el $\cos(2\pi f_0 t)$ de la portada de la señal. Para ello definimos,

$$\Theta_n = \begin{cases} +\frac{\pi}{2} & \text{si } B_n = 1; \\ -\frac{\pi}{2} & \text{si } B_n = 0. \end{cases}$$

Y el ángulo del proceso mediante:

$$\Theta_t = \Theta_k, \quad \text{para } kT \leq t < (k+1)T,$$

Donde T es una constante que denota el tiempo de transmisión de un bit, que suele elegirse como múltiplo de $1/f_0$ para tener así un número entero de ciclos en el tiempo de transmisión del bit. El inverso de T se denomina la tasa de baudios.

El proceso de la señal modulada es el proceso PSK, cuya expresión es:

$$X_t = \cos(2\pi f_0 t + \Theta_t).$$

Para la obtención de las funciones de momentos del proceso es conveniente hacer uso de las funciones:

$$h^{(c)}(t) = \begin{cases} \cos(2\pi f_0 t), & 0 \leq t < T; \\ 0, & \text{en el resto.} \end{cases}$$

$$h^{(s)}(t) = \begin{cases} \sin(2\pi f_0 t), & 0 \leq t < T; \\ 0, & \text{en el resto.} \end{cases}$$

Aplicando la fórmula del coseno del ángulo suma, podemos escribir:

$$\cos(2\pi f_0 t + \Theta_k) = \cos(\Theta_k) \cos(2\pi f_0 t) - \sin(\Theta_k) \sin(2\pi f_0 t)$$

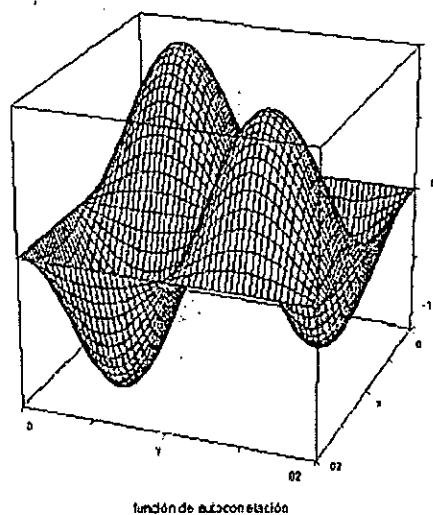
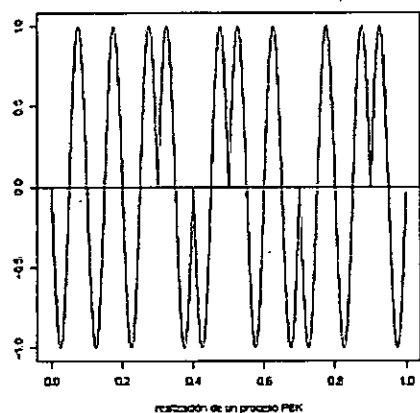
$$\doteq \sum_{k=-\infty}^{k=+\infty} \cos(\Theta_k) h^{(c)}(t - kT) - \sum_{k=-\infty}^{k=+\infty} \sin(\Theta_k) h^{(s)}(t - kT)$$

$$= - \sum_{k=-\infty}^{k=+\infty} \sin(\Theta_k) h^{(s)}(t - kT),$$

Donde el primer sumatorio se anula porque todos sus términos son 0, $\Theta_k = \pm\pi/2, \forall k$ y en consecuencia $\cos(\Theta_k) = 0, \forall k$. La expresión anterior permite obtener fácilmente $\mu(t) = 0$ dado que $\sin(\Theta_k) = \pm 1$ con igual probabilidad.

FIGURA N° 5.4

REALIZACIÓN DE UN PROCESO PSK Y SU
CORRESPONDIENTE FUNCIÓN DE AUTOCORRELACIÓN



Las funciones autocovarianza y autocorrelación coinciden:

$$R(t_1, t_2) = E(X_{t_1} X_{t_2}) = \sum_{k,l} E[\sin(\Theta_k) \sin(\Theta_l)] h^{(s)}(t_1 - kT) h^{(s)}(t_2 - lT),$$

Teniendo en cuenta la independencia de la sucesión original y que $\Theta_k = \pm\pi/2, \forall k$,

$$E[\sin(\Theta_k) \sin(\Theta_l)] = \begin{cases} 0, & \text{si } k \neq l; \\ 1, & \text{si } k = l, \end{cases}$$

Con lo que:

$$R(t_1, t_2) = \sum_{k=-\infty}^{k=+\infty} h^{(s)}(t_1 - kT) h^{(s)}(t_2 - kT).$$

Por la definición, el soporte de $h^{(s)}(t)$ es el intervalo $[0, T]$, por lo que el anterior sumatorio será siempre nulo a menos que $\exists k_0$ tal que t_1 y t_2 estén ambos en $[(k_0 - 1)T, k_0 T]$. Si denotamos por $t^* = t \bmod T$, la expresión final de $R(t_1, t_2)$ será:

$$R(t_1, t_2) = \begin{cases} h^{(s)}(t_1^*) h^{(s)}(t_2^*), & \text{si } [t_1/T] = [t_2/T]; \\ 0, & \text{en el resto,} \end{cases}$$

Donde $[.]$ representa la parte entera por defecto del argumento. La figura 5.4 muestra una realización de un proceso PSK y su función de autocorrelación.

5.7. Proceso de Wiener. Movimiento Browniano

➤ Límite de un camino aleatorio

El camino aleatorio que se describe anteriormente es la suma de desplazamientos independientes de derecha a izquierda, de forma que sólo saltos unitarios estaban permitidos en uno u otro sentido en cada intervalo unitario de tiempo. Imaginemos ahora desplazamientos pequeños de longitud Δ que se producen en intervalos de tiempo pequeños de longitud τ , cuando desplazamiento y tiempo tiendan a cero obtendremos un proceso cuyas realizaciones serán funciones continuas en el tiempo. Debemos precisar en qué condiciones tienden a cero ambas cantidades

Consideremos una partida situada en el origen. En cada intervalo de tiempo τ se desplaza una cantidad aleatoria Z de manera que:

$$P(Z = +\Delta) = p, \quad P(Z = -\Delta) = 1 - p = q,$$

Siendo los distintos desplazamientos independientes. Se trata de una variable dicotómica con media y varianza:

$$\mu_Z = (p - q)\Delta, \quad \sigma_Z^2 = 4pq\Delta^2,$$

Y cuya función característica vale:

$$\varphi_Z(u; \Delta) = E(e^{iuZ}) = pe^{iu\Delta} + qe^{-iu\Delta}.$$

En un tiempo t se habrán producido $n = \lfloor t/T \rfloor$ desplazamientos, siendo X_t la posición final de la partícula. Dicha posición es por tanto la suma $X_t = \sum_{i=1}^n Z_i$, con las Z_i i.i.d como la Z anterior. Así, la función característica de X_t valdrá:

$$\varphi_{X_t}(u, \tau, \Delta) = E(e^{iuX_t}) = (pe^{iu\Delta} + qe^{-iu\Delta})^n = (pe^{iu\Delta} + qe^{-iu\Delta})^{\lfloor t/\tau \rfloor}.$$

La media y la varianza de X_t se obtienen fácilmente a partir de μ_Z y σ_Z^2 ,

$$\mu_{X_t} = \lfloor t/\tau \rfloor (p - q)\Delta, \quad \sigma_{X_t}^2 = \lfloor t/\tau \rfloor 4pq\Delta^2.$$

Nuestro objetivo es $\Delta \rightarrow 0$ y $\tau \rightarrow 0$ de manera tal que obtengamos un resultado razonable. Por ejemplo, por razonable podríamos entender que media y varianza de X_t , para $t=1$, fueran finitas e iguales a μ y σ^2 , respectivamente. Ello supondría que Δ y τ deben tender a cero de forma tal que:

$$\frac{(p - q)\Delta}{\tau} \rightarrow \mu \quad \text{y} \quad \frac{4pq\Delta^2}{\tau} \rightarrow \sigma^2.$$

Para ello debe cumplirse:

$$\Delta = \sigma\sqrt{\tau}, \quad p = \frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma} \right), \quad q = \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma} \right).$$

Estas relaciones suponen que tanto p como q deben ser valores muy próximos a $1/2$ si queremos evitar degeneraciones en el proceso límite y que Δ es de un orden de magnitud mucho mayor que τ puesto que como infinitésimo $\Delta = O(\tau^{1/2})$.

La distribución de probabilidad límite de las X_t podemos conocerla a través comportamiento límite de su función característica:

$$\varphi_{X_t}(u, \tau) = \left[\frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma} \right) e^{iu\sigma\sqrt{\tau}} + \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma} \right) e^{-iu\sigma\sqrt{\tau}} \right]^{t/\tau},$$

Y desarrollaremos en potencias de τ la expresión entre corchetes. Hecho esto y haciendo $\tau \rightarrow 0$, tendremos:

$$\varphi_{X_t}(u) = \exp\left(\mu tu - \frac{1}{2}\sigma^2 tu\right),$$

Que es la función característica de una $N(\mu t, \sigma^2 t)$.

Observamos además que el proceso formado por las variables X_t hereda las propiedades del camino aleatorio que lo genera, al fin y al cabo lo único que hemos hecho es aplicar el teorema Central de Límite a la suma de variable dicotómicas que definen X_t . De entre ellas conviene destacar la propiedad de incrementos independientes y estacionarios. Así, si $t_1 < t_2$, $X_{t_2} - X_{t_1} \sim N(\mu(t_2 - t_1), \sigma^2(t_2 - t_1))$.

Para obtener ahora las distribuciones finito-dimensional del proceso límite podemos recurrir a la expresión anterior, que proporciona la densidad conjunta a partir de las densidades de los incrementos. Si $t_1 < t_2 < \dots < t_n$,

$$\begin{aligned} f_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) &= f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \dots f_{t_n - t_{n-1}}(x_n - x_{n-1}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2 t_1}} e^{-\frac{1}{2}\left(\frac{(x_1 - \mu t_1)^2}{\sigma^2 t_1}\right)} \dots \frac{1}{\sqrt{2\pi\sigma^2 (t_n - t_{n-1})}} e^{-\frac{1}{2}\left(\frac{[(x_n - x_{n-1}) - (\mu t_n - \mu t_{n-1})]^2}{\sigma^2 (t_n - t_{n-1})}\right)} \\ &= \frac{e^{-\frac{1}{2}\left\{\frac{(x_1 - \mu t_1)^2}{\sigma^2 t_1} + \dots + \frac{[(x_n - x_{n-1}) - (\mu t_n - \mu t_{n-1})]^2}{\sigma^2 (t_n - t_{n-1})}\right\}}}{\sqrt{(2\pi\sigma^2)^n t_1 \dots (t_n - t_{n-1})}}, \end{aligned}$$

Que es la densidad conjunta de una Normal multivariable. El proceso límite es un proceso Gaussiano con incrementos independientes.

➤ Proceso de Wiener

El proceso de Wiener es un proceso límite como el descrito en el párrafo anterior con $p=q=1/2$.

Definición (Proceso de Wiener) Un proceso de Wiener es un proceso estocástico W_t , que verifica,

Su valor inicial es 0, $P(W_0 = 0) = 1$.

Sus incrementos son independientes.

Para $0 \leq t_1 < t_2$, el incremento $W_{t_2} - W_{t_1} \sim N(0, \sigma^2(t_2 - t_1))$.

Los fundamentos matemáticos del proceso establecidos por Norbert Wiener, matemático americano especialista en inferencia estadística y teoría de la comunicación. El proceso es conocido también como movimiento browniano, expresión utilizada para describir el movimiento aleatorio de las moléculas de un gas al entrec chocar entre sí, que recibe este nombre en honor de Robert Brown, un botánico del siglo diecinueve que lo describió por primera vez.

La media y la varianza del proceso de Wiener son ya conocidas:

$$\mu(t) = 0, \quad \sigma^2(t) = \sigma^2 t.$$

Autocovarianza y autocorrelación coinciden. Para su obtención, como los incrementos son independientes, obteniendo:

$$C(t_1, t_2) = R(t_1, t_2) = \sigma^2 \min(t_1, t_2).$$

Las funciones de autocovarianza del proceso de Wiener y el proceso de Poisson, son iguales a pesar de tratarse de X_t dos procesos de naturaleza muy distinta. Las trayectorias del segundo son funciones escalonadas mientras que las del primero son continuas. Esta igualdad pone en evidencia que las funciones de momento son sólo descripciones parciales del proceso.

5.8. Cadenas de Markov

Un proceso estocástico X_t es un proceso de Markov si verifica la propiedad de Markov, enunciada anteriormente. Dicha propiedad supone que la evolución del proceso depende tan sólo de su pasado inmediato. En términos de probabilidad, si las variables del proceso son discretas, la propiedad se expresa mediante,

$$P(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}, \dots, X_{t_1} = x_{t_1}) = P(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}).$$

Si las variables del proceso son continuas su expresión es:

$$P(X_{t_{n+1}} \leq x_{t_{n+1}} | X_{t_n} \leq x_{t_n}, \dots, X_{t_1} \leq x_{t_1}) = P(X_{t_{n+1}} \leq x_{t_{n+1}} | X_{t_n} \leq x_{t_n}).$$

O sus equivalentes en términos de las funciones de probabilidad o de distribución, respectivamente.

Algunos de los procesos estudiados hasta ahora son procesos de Markov. De tal manera los procesos suma IID y el proceso de Poisson, tenían incrementos independientes, eso implica que poseen propiedad de Markov.

Para el proceso RTS, recordemos que el valor de $X_{t_{n+1}} = \pm a$ y que su valor depende exclusivamente del signo que tenga X_{t_n} y de la paridad de $N_{t_{n+1}-t_n}$, número de cambios (sucesos) del proceso de Poisson subyacente, y es por tanto un proceso de Markov.

Ejemplo (Un proceso que no es de Markov).- Se consideró en el siguiente proceso:

$$Y_n = \frac{1}{2}(X_n + X_{n-1}),$$

Donde las X_k son independientes, ambas Bernoulli con $p=1/2$. Un proceso de estas características recibe el nombre de proceso de medias móviles de orden 2, puesto que se define a partir de la media aritmética de las dos últimas realizaciones de otro proceso.

$$P(Y_n = 0) = P(X_n = 0, X_{n-1} = 0) = \frac{1}{4}.$$

Para los demás:

$$P(Y_n = 1/2) = \frac{1}{2}, \quad P(Y_n = 1) = \frac{1}{4}.$$

Obteniendo la probabilidad condicionada para dos valores consecutivos de Y_n ,

$$\begin{aligned} P(Y_n = 1 | Y_{n-1} = 1/2) &= \frac{P(Y_n = 1, Y_{n-1} = 1/2)}{P(Y_{n-1} = 1/2)} \\ &= \frac{P(X_n = 1, X_{n-1} = 1, X_{n-2} = 0)}{P(Y_{n-1} = 1/2)} \\ &= \frac{1/8}{1/2} = \frac{1}{4}. \end{aligned}$$

Si se tiene información de un instante anterior, por ejemplo, $Y_{n-2} = 1$, podemos calcular:

$$P(Y_n = 1 | Y_{n-1} = 1/2, Y_{n-2} = 1) = \frac{P(Y_n = 1, Y_{n-1} = 1/2, Y_{n-2} = 1)}{P(Y_{n-1} = 1/2, Y_{n-2} = 1)},$$

Pero el suceso $\{Y_n = 1, Y_{n-1} = 1/2\} = \{X_n = 1, X_{n-1} = 1, X_{n-2} = 0\}$ y el suceso con lo que ambos sucesos son incompatibles, por tanto:

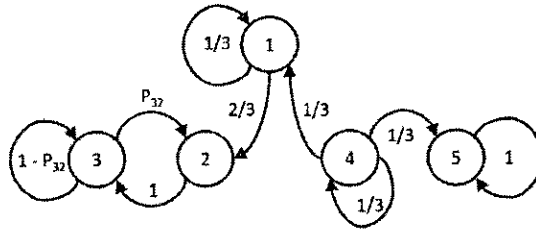
$$P(Y_n = 1 | Y_{n-1} = 1/2, Y_{n-2} = 1) = 0 \neq P(Y_n = 1 | Y_{n-1} = 1/2)$$

Y el proceso no puede ser Markov.

TABLA N° 5.1
“EJEMPLO DE CADENA DE MARKOV”

CADENA DE MARKOV

DESDE DIAGRAMA
(EJEMPLO)



DESDE MATRIZ
(EJEMPLO)

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 9/10 & 0 & 1/10 & 0 \\ 0 & 1/10 & 0 & 9/10 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

FUENTE: Elaboración propia.

CAPÍTULO 6

PROCESOS ESTACIONARIOS Y TRANSFORMACIÓN LINEAL DE UN PROCESO ESTACIONARIO

6.1. Procesos estacionarios

La estacionariedad al comprobar que la distribución ligada a los incrementos dependía tan solo de la diferencia de tiempos y no de su posición.

El concepto de estacionariedad recoge una propiedad común a muchos procesos consistentes en que su comportamiento probabilístico no cambia con el tiempo. La definición formal es la siguiente.

Definición (Proceso estacionario)

Un proceso estocástico X_t se dice que es estacionario si sus distribuciones finito-dimensionales son invariantes por traslación. Es decir:

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_1+\tau, t_2+\tau, \dots, t_n+\tau}(x_1, x_2, \dots, x_n), \forall (t_1, t_2, \dots, t_n), \forall \tau.$$

La primera consecuencia de esta definición es que las distribuciones individuales de las variables que componen el proceso no dependió de t :

$$F_t(x) = F_{t+\tau}(x) = F(x), \forall t, \tau,$$

Y como consecuencia: $\mu(t) = E(X_t) = \mu, \forall t$

Y $\sigma^2(t) = E[(X_t - \mu)^2] = \sigma^2.$

La distribución conjunta de (X_{t_1}, X_{t_2}) dependerá tan solo de la diferencia de tiempos, $t_2 - t_1$. Basta para ello hacer $\tau = t_1$:

$$F_{t_1, t_2}(x_1, x_2) = F_{0, t_2 - t_1}(x_1, x_2), \forall t_1, t_2.$$

La consecuencia es que los momentos de segundo orden y las funciones correspondientes dependen también, solamente, de dicha diferencia.

$$R(t_1, t_2) = R(t_1 - t_2), \quad C(t_1, t_2) = C(t_1 - t_2), \forall t_1, t_2.$$

Algunos de los procesos antes definidos gozan de esta propiedad. Así, el proceso IID es estacionario porque:

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F(x_1) \dots F(x_n)$$

$$= F_{t_1+\tau, t_2+\tau, \dots, t_n+\tau}(x_1, x_2, \dots, x_n), \forall \tau, t_1, t_2, \dots, t_n.$$

Sin embargo, el proceso suma IID no es aleatorio porque recordemos que su media y varianza valían $\mu(n) = n\mu$ y $\sigma^2(n) = n\sigma^2$, que crecen con n y no son constantes.

Ejemplo (Estacionariedad del proceso RTS)

El proceso RTS se estudió anteriormente, recordemos que el proceso toma solo dos valores simétricos, $\{-a, a\}$,

De acuerdo con el siguiente esquema.

$X_0 = \pm a$ con igual probabilidad $p = 1/2$,

X_t cambia de signo con cada ocurrencia de un suceso en un proceso de Poisson de parámetro λ .

Obtengamos la distribución finito-dimensional, para lo que haremos uso de la propiedad de incrementos independientes que el proceso RTS hereda del proceso de Poisson.

$$P(X_{t_1} = x_1, \dots, X_{t_n} = x_n) = P(X_{t_1} = x_1)P(X_{t_2} = x_2 | X_{t_1} = x_1) \dots P(X_{t_n} = x_n | X_{t_{n-1}} = x_{t_{n-1}}).$$

Las probabilidades condicionadas que aparecen en la expresión dependen del número de cambios que se han producido entre los dos tiempos implicados, concretamente de la paridad de $N_{|t_i - t_j|}$, número de sucesos ocurridos en el intervalo de tiempo.

$$P(X_{t_j} = x_j | X_{t_i} = x_i) = \begin{cases} \frac{1}{2} P(N_{|t_j - t_i|} = \text{par}) = \frac{1}{2} (1 + e^{-2\lambda|t_j - t_i|}), & \text{si } x_i = x_j \\ \frac{1}{2} P(N_{|t_j - t_i|} = \text{impar}) = \frac{1}{2} (1 - e^{-2\lambda|t_j - t_i|}), & \text{si } x_i \neq x_j. \end{cases}$$

Si efectuamos ahora una traslación τ de todos los tiempos:

$$|t_i - t_j| = |(t_i - \tau) - (t_j - \tau)|$$

$$Y \quad P(X_{t_j} = x_j | X_{t_i} = x_i) = P(X_{t_j+\tau} = x_j | X_{t_i+\tau} = x_i),$$

Y como $P(X_{t_i} = x_i) = 1/2, \forall t_i, x_i$, se deduce la estacionariedad del proceso RTS. Si generalizamos el proceso RTS de manera que $P(X_0 = +a) = p$ y

$P(X_0 = -a) = 1 - p$ la estacionariedad se pierde. Las probabilidades condicionadas no cambian al efectuar la traslación τ , pero si pueden hacerlo $P(X_t = x_1)$. Por ejemplo; si $x_1 = a$:

$$\begin{aligned} P(X_t = a) &= \\ &= P(X_t = a | X_0 = a)P(X_0 = a) + P(X_t = a | X_0 = -a)P(X_0 = -a) \\ &= P(N_t = \text{par}) + (1 - p)P(N_t = \text{impar}) \\ &= \frac{1}{2} \left(1 - e^{-2\lambda t} + 2pe^{-2\lambda t} \right). \end{aligned}$$

Al efectuar la traslación,

$$\begin{aligned} P(X_{t+\tau} = a) &= P(X_{t+\tau} = a | X_0 = a)P(X_0 = a) + P(X_{t+\tau} = a | X_0 = -a)P(X_0 = -a) \\ &= P(N_{t+\tau} = \text{par}) + (1 - p)P(N_{t+\tau} = \text{impar}) \\ &= \frac{1}{2} \left(1 - e^{-2\lambda(t+\tau)} + 2pe^{-2\lambda(t+\tau)} \right). \end{aligned}$$

Si desde el inicio hay equiprobabilidad, $p=1/2$, $P(X_t = \pm a) = 1/2, \forall t$

6.1.1. Estacionariedad en sentido amplio (WSS)

Definición. Decimos que un proceso estocástico es estacionario en sentido amplio (WSS, Wide-Sense Stationary) si su media es constante y su función de autocorrelación (autocovarianza) es invariante por traslación. Es decir,

$$\mu(t) = \mu, \forall t, \quad \text{y} \quad R(t_1, t_2) = R(t_1 + \tau, t_2 + \tau) = R(\tau), \forall t_1, t_2.$$

La WSS es una propiedad más débil que la estacionariedad. Esta última implica a aquella, pero no al revés, como pone de manifiesto el siguiente contraejemplo.

Ejemplo (WSS \nRightarrow la estacionariedad)

El proceso X_n está formado por dos procesos intercalados de variables independientes, si $n=2k$, $X_n = \pm 1$ con probabilidad $1/2$, si $n=2k+1$, X_n toma los valores $1/3$ y -3 con probabilidad $9/10$ y $1/10$, respectivamente. El proceso no puede ser estacionario porque su función de probabilidad varía con n .

Por otra parte, se comprueba fácilmente que $\mu(n) = 0, \forall n$

$$Y \quad C(n_1, n_2) = \begin{cases} E(X_{n_1})E(X_{n_2}) = 0, & \text{si } n_1 \neq n_2, \\ E(X^2), & \text{si } n_1 = n_2. \end{cases}$$

El proceso es WSS.

Propiedades de la función de autocorrelación de un proceso WSS

En un proceso WSS la función de autocorrelación tiene una serie de propiedades que por su interés posterior vamos a deducir.

PA1) Para $\tau = 0$:

$$R(0) = R(t, t) = R(X_t^2), \forall t,$$

que es la potencia media del proceso.

PA2) La función de autocorrelación es una función par. En efecto,

$$R(\tau) = E(X_{t+\tau} X_t).$$

y haciendo $s = t + \tau$,

$$R(\tau) = E(X_{t+\tau} X_t) = E(X_s X_{s-\tau}) = R(-\tau).$$

PA3) La función de autocorrelación mide la tasa de cambio del proceso en términos de probabilidad. Si consideramos el cambio que ha sufrido el proceso entre t y $t + \tau$,

$$\begin{aligned} P(|X_{t+\tau} - X_t| > \varepsilon) &= P(|X_{t+\tau} - X_t|^2 > \varepsilon^2) \\ &\leq \frac{E[(X_{t+\tau} - X_t)^2]}{\varepsilon^2} \\ &= \frac{E[X^2 + X^2 - 2X_{t+\tau} X_t]}{\varepsilon^2} \\ &= \frac{2[R(0) - R(\tau)]}{\varepsilon^2}. \end{aligned}$$

PA4) La función de autocorrelación alcanza su máximo en 0. Aplicando la desigualdad de Schwartz, $[E(XY)]^2 \leq E(X^2)E(Y^2)$, a $X_{t+\tau}$ y X_t ,



$$R(\tau)^2 = [E(X_{t+\tau} X_t)]^2 \leq E(X_{t+\tau}^2) E(X_t^2) = R(0)^2.$$

y por tanto $|R(\tau)| \leq R(0)$.

PA5) Si $R(d)=R(0)$, la función de autocorrelación es periódica con periodo d . Aplicando de nuevo la desigualdad de Schwartz a $X_{t+\tau+d} - X_{t+\tau}$ y X_t ,

$$\{E[(X_{t+\tau+d} - X_{t+\tau})]\}^2 \leq E[(X_{t+\tau+d} - X_{t+\tau})]^2 E[X_t]^2,$$

Que en términos de $R(\tau)$ podemos escribir:

$$[R(\tau + d) - R(\tau)]^2 \leq 2[R(0) - R(d)]R(0).$$

Como $R(d)=R(0)$ entonces $R(\tau + d) = R(\tau), \forall \tau$ con lo que la función es periódica de periodo d .

Por otra parte, de:

$$E[(X_{t+d} - X_t)^2] = 2[R(0) - R(d)] = 0,$$

Se deduce que el proceso es periódico en media cuadrática.

PA6) Si $X_t = m + N_t$ donde N_t es un proceso de media cero con $R_N(\tau) \xrightarrow{\tau \rightarrow \infty} 0$, entonces:

$$R_X(\tau) = E[(m + N_{t+\tau})(m + N_t)] = m^2 + 2mE(N_t) + R_N(\tau) = m^2 + R_N(\tau) \xrightarrow{\tau \rightarrow \infty} m^2.$$

Deducimos de estas propiedades que la función de autocorrelación puede tener tres tipos de componentes:

- Una componente que se aproxima a 0 cuando $t \rightarrow \infty$
- Una componente periódica, y
- Una componente con media no nula.
- Procesos Gaussianos y estacionariedad

Se ha visto que la WSS \rightarrow estacionariedad, pero este resultado tiene una excepción en el caso del proceso Gaussiano. Como vimos anteriormente, las distribuciones finito-dimensionales del proceso están completamente especificadas mediante un vector de medias y la matriz de covarianzas.



Si el proceso Gaussiano es WSS, sabemos que $\mu(t) = \mu, \forall t$ y $C(t_1, t_2) = g(|t_1 - t_2|)$. La consecuencia inmediata es que la distribución conjunta de $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ será invariante por traslación y el proceso será estacionario.

Ejemplo (Un proceso Gaussiano de medias móviles)

Definamos $Y_n = \frac{X_n + X_{n-1}}{2}$.

Donde X_n es un proceso Gaussiano IID con media 0 y varianza σ^2 .

La media de Y_n es también 0 y su covarianza:

$$C_Y(n_1, n_2) = \frac{1}{4} E[(X_{n_1} + X_{n_1-1})(X_{n_2} + X_{n_2-1})]$$

$$= \frac{1}{4} E[X_{n_1} X_{n_2} + X_{n_1} X_{n_2-1} + X_{n_1-1} X_{n_2} + X_{n_1-1} X_{n_2-1}]$$

$$= \begin{cases} \frac{\sigma^2}{2}, & \text{si } n_1 - n_2 = 0, \\ \frac{\sigma^2}{4}, & |n_1 - n_2| = 1, \\ 0, & \text{en el resto.} \end{cases}$$

En el proceso Y_n es por tanto WSS y además es Gaussiano por ser combinación lineal de variables Gaussianas. Las distribuciones finito dimensionales del proceso están especificados con el vector de medias nulo y matriz de covarianzas que define $C_Y(n_1, n_2)$.

En este capítulo, para facilitar la comprensión del lector, se ilustrará con los tipos de los procesos estacionarios, en el cual se mencionó: los procesos estocásticos estacionarios, no estacionarios y estacionarios en media. Esto permitirá al lector una comprensión más fácil sobre el tema.



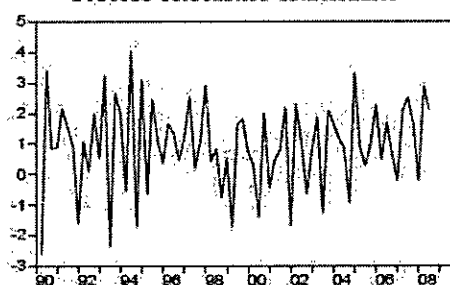
TABLA N° 6.1

“TIPOS DE PROCESOS ESTACIONARIOS”

TIPOS DE PROCESOS ESTACIONARIOS

ESTACIONARIO

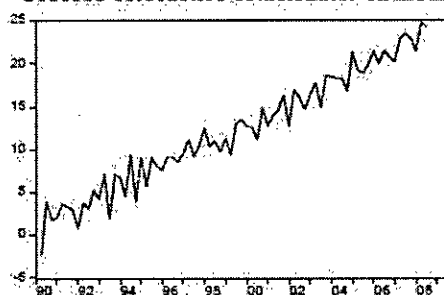
Proceso estocástico estacionario



ESTACIONARIO EN MEDIA

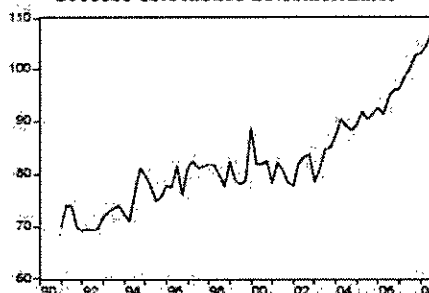
EN

Proceso estocástico estacionario en media



NO ESTACIONARIO

Proceso estocástico no estacionario



FUENTE: Elaboración propia.

6.2. Procesos Cicloestacionarios

Definición (Proceso cicloestacionario (CE)) Decimos que el proceso X_t es cicloestacionario, CE si sus distribuciones finito-dimensionales son invariantes por traslación mediante múltiplos enteros de un cierto periodo T. Es decir, $\forall(t_1, t_2, \dots, t_n)$ y $\forall k \in \mathbb{Z}$,

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_1+kT, t_2+kT, \dots, t_n+kT}(x_1, x_2, \dots, x_n).$$

Definición (Proceso cicloestacionario en sentido amplio (WSC))

Decimos que el proceso X_t es cicloestacionario en sentido amplio, WSC si su media y su función de autocovarianza son invariantes por traslación mediante múltiplos enteros de un cierto periodo T:

$$\mu(t+kT) = \mu(t), \quad C(t_1+kT, t_2+kT) = C(t_1, t_2).$$

Ejemplo: Sea el proceso $X_t = A \cos\left(\frac{2\pi t}{T}\right)$.

La distribución conjunta de $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ es

$$\begin{aligned} F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) &= P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) \\ &= P\{A \cos(2\pi t_1 / T) \leq x_1, \dots, A \cos(2\pi t_n / T) \leq x_n\} \\ &= \\ &P\{A \cos(2\pi(t_1+kT) / T) \leq x_1, \dots, A \cos(2\pi(t_n+kT) / T) \leq x_n\} \\ &= F_{t_1+kT, t_2+kT, \dots, t_n+kT}(x_1, x_2, \dots, x_n), \end{aligned}$$

Y el proceso es cicloestacionario.

El proceso del ejemplo anterior es periódico en el sentido que todas sus trayectorias lo son. Pero hay que tener en cuenta que el proceso puede ser cicloestacionario. Hay procesos con un comportamiento cíclico que no tienen este tipo de trayectorias y que no pudiendo ser cicloestacionarios, son WSC.

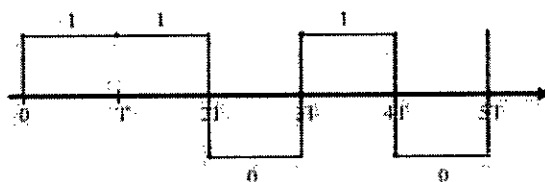
Ejemplo: Un modem transmite señales binarias 0 y 1 IID de la siguiente forma,

Para transmitir un 1, emite una señal rectangular de amplitud 1 y duración T,

Para transmitir un 0, emite una señal rectangular de amplitud -1 y duración T.

FIGURA N° 6.1

“UNA TRAYECTORIA DEL PROCESO DE PULSOS MODULADOS”



La FIGURA 6.1 muestra la trayectoria correspondiente a la secuencia de datos 11010...

El proceso puede ser descrito mediante.

$$X_n = \sum_{n=-\infty}^{+\infty} A_n 1_{[0,T]}(t-nT),$$

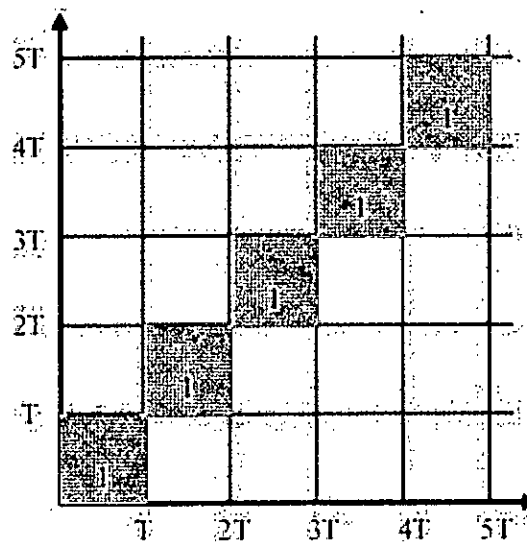
Donde $A_n = \pm 1$ según el valor a transmitir y $1_{[0,T]}(\cdot)$ es la función indicatriz del intervalo $[0, T]$. La media del proceso es 0 por que $E(A_n) = 0, \forall n$. La función de autocovarianza vale:

$$\begin{aligned} C(t_1, t_2) &= E \left[\left(\sum_{n=-\infty}^{+\infty} A_n 1_{[0,T]}(t_1 - nT) \right) \left(\sum_{m=-\infty}^{+\infty} A_m 1_{[0,T]}(t_2 - mT) \right) \right] \\ &= E \left[\left(\sum_{n=-\infty}^{+\infty} A_n^2 1_{[0,T]}(t_1 - nT) 1_{[0,T]}(t_2 - nT) \right) \right] \\ &= \begin{cases} 1, & \text{si } (t_1, t_2) \in [nT, (n-1)T] \times [nT, (n-1)T], \\ 0, & \text{en el resto} \end{cases} \end{aligned}$$

La figura muestra la región del plano (celdas en gris) en la que $C(t_1, t_2)$ vale la unidad, de donde se deduce que $C(t_1 + kT, t_2 + kT) = C(t_1, t_2)$ y el proceso es WSC.

FIGURA N° 6.2

“LA FUNCIÓN DE AUTOCOVARIANZA DEL PROCESO DE PULSOS MODULADOS”



6.3. Densidad espectral de potencia (PSD) de un proceso WSS

El desarrollo en serie de Fourier de una función determinista permite expresar como una suma ponderada de funciones periódicas. Si la función no presenta cambios bruscos y varía suavemente, los coeficientes (pesos) de la serie correspondientes a las sinusoides de baja frecuencia presentan valores elevados frente a los de alta frecuencia. Las funciones muy variables muestran una distribución de coeficientes opuesta. Se deduce que la tasa de variación de la función está relacionada con el valor de los distintos coeficientes, que son considerados como una función y recibe el nombre de espectro de la función original.

Este tipo de descomposición es representar la función en el dominio de las frecuencias y es trasladable a los procesos estocásticos, transformándose en una herramienta de gran utilidad.

Pero la traslación no puede hacerse de forma directa debido a la diferencia existente entre los fenómenos deterministas y aleatorios.

Como ya sabemos, las distintas realizaciones de un proceso dan lugar a diferentes funciones (trayectorias) y por esta razón el espectro de un proceso estocástico expresa la variación media de todas aquellas trayectorias a lo largo del tiempo.

Por otra parte, usando las propiedades de la función de autorrelación de un proceso WSS mide la tasa de cambio del proceso en términos de probabilidad.

$$P(|X_{t+\tau} - X_t| > \varepsilon) \leq \frac{2[R(0) - R(\tau)]}{\varepsilon^2}.$$

De la descomposición en serie de Fourier surge el concepto de densidad espectral de potencia (PSD de sus siglas en ingles)

6.3.1. PSD para procesos estocásticos WSS discretos en el tiempo

Definición.- Sea X_t un proceso estocástico WSS discreto en el tiempo, su espectro o densidad espectral de potencia, $P(w)$, es la transformada de Fourier de su función de autocorrelación $R(k)$.

$$P(w) = F[R(k)] = \sum_{k=-\infty}^{k=+\infty} R(k) \exp(-i2\pi wk), -\infty < w < \infty.$$

Si se tiene en cuenta que $R(k) = R(-k)$ y $e^{-ix} + e^{ix} = 2 \cos x$, la expresión anterior adopta la forma:

$$P(w) = R(0) + 2 \sum_{k \geq 1} R(k) \cos 2\pi wk,$$

Lo que implica que $P(w)$ es una función real par, $P(w) = P(-w)$. De esta expresión para $P(w)$ se deduce que es periódica con periodo 1, por lo que solo consideramos el rango de frecuencias $-1/2 < w \leq 1/2$, pero dada su paridad bastara con definir $P(w)$ para $w \in [0, 1/2]$.

Se obtendrá:

$$P^*(w) = 1 + 2 \sum_{k \geq 1} \frac{R(k)}{R(0)} \cos 2\pi wk.$$

En caso de tratarse de un proceso centrado con $\mu(n) = 0, \forall n$, el cociente:

$$\frac{R(k)}{R(0)} = \frac{C(k)}{C(0)} = \rho(k),$$

es la función de correlación y se escribe:

$$P^*(w) = 1 + 2 \sum_{k \geq 1} \rho(k) \cos 2\pi wk.$$



Ejemplo (Espectro de un ruido blanco)

Si el proceso $\{X_t\}$ es un ruido blanco, $\mu(t) = 0, \sigma^2(t) = \sigma^2, \forall t$ y las variables son incorreladas (algunos autores exigen independencia). Se trata, evidentemente, de un proceso WSS en el que:

$$C(k) = R(k) = \begin{cases} \sigma^2, & \text{si } k=0, \\ 0, & \text{si } k \neq 0. \end{cases}$$

Sustituyendo, tendremos:

$$P(w) = \sigma^2,$$

Su versión normalizada:

$$P^*(w) = 1.$$

En cualquiera de sus formas, el espectro de un ruido blanco es constante para cualquier frecuencia w , a semejanza de lo que ocurre con el espectro óptico plano de la luz blanca, lo que justifica el nombre que reciben este tipo de procesos.

➤ PSD versus R(k)

La definición de $P(w)$ como $F[R(k)]$ permite recuperar la función de autocorrelación si $P(w)$ es conocida. Basta para ello recurrir a la transformada inversa de Fourier,

$$R(k) = F^{-1}[P(w)] = \int_{-1/2}^{1/2} P(w) \exp(i2\pi kw) dw = 2 \int_0^{1/2} P(w) \cos(2\pi kw) dw.$$

Un resultado conocido como el Teorema de Wold, afirma que cualquier función $P(w)$ no negativa definida sobre $[0, 1/2]$ define un espectro y, en consecuencia, $R(k)$ es una función de autocorrelación si y solo si puede expresarse a partir de $P(w)$. Se deriva de ello, que las condiciones a cumplir para ser función de autocorrelación son mucho más restrictivas que las exigidas para el espectro.

Las herramientas matemáticas $P(w)$ y $R(k)$ equivalentes.

Ejemplo (Espectro de un proceso autoregresivo de orden uno, AR(1))

Un proceso estocástico AR(1) se define de esta manera,

$$X_t = \alpha X_{t-1} + Z_t.$$

AP

Donde $\{Z_t\}$ es una sucesión de ruido blanco y $|\alpha| < 1$. La restricción para el valor de α es necesaria para que el proceso WSS.

Para calcular $R(k)$ multiplicamos ambas partes por X_{t-k} y al tomar esperanzas se obtiene:

$$R(k) = \alpha R(k-1),$$

$$R(k) = \alpha^k R(0) = \alpha^k \gamma^2.$$

Con $R(k) = R(-k)$, $k = 1, 2$, siendo γ^2 la varianza del proceso. Su espectro vale:

$$\begin{aligned} P(w) &= \gamma^2 \sum_{-\infty}^{+\infty} \alpha^k \exp(-i2\pi wk) \\ &= \gamma^2 + \gamma^2 \sum_{k \geq 1} \alpha^k \exp(-i2\pi wk) + \gamma^2 \sum_{k \geq 1} \alpha^k \exp(i2\pi wk) \\ &= \gamma^2 \left(1 + \frac{\alpha \exp(-i2\pi w)}{1 - \alpha \exp(-i2\pi w)} + \frac{\alpha \exp(i2\pi w)}{1 - \alpha \exp(i2\pi w)} \right). \end{aligned}$$

Donde:

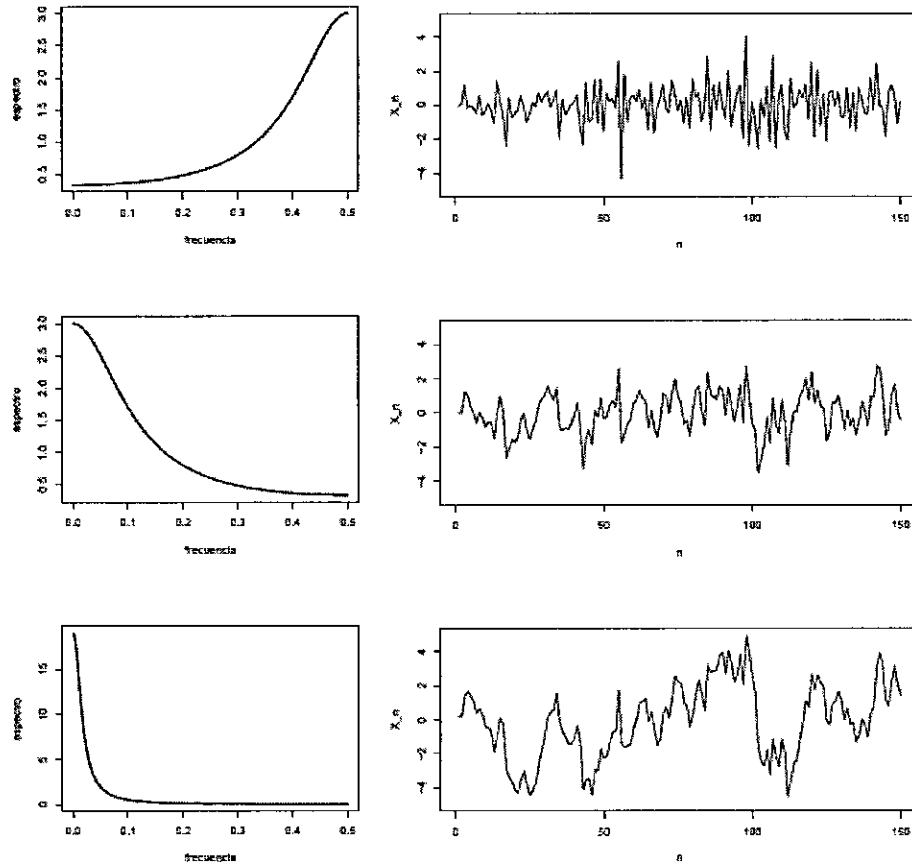
$$P(w) = \frac{\gamma^2 (1 - \alpha^2)}{1 - 2\alpha \cos 2\pi w + \alpha^2}.$$

En la figura hemos representado para $\sigma^2 = 1$ y $\alpha = -0.5, 0.5, 0.9$ y una trayectoria de cada uno de estos procesos.

El espectro para $\alpha = -0,5$ (superior) es creciente lo que significa un mayor peso de las altas frecuencias, la potencia se concentra en ellas, y ello se corresponde con una trayectoria que alterna valores positivos y negativos con rápidas oscilaciones entre unos y otros. Para valores positivos de α los espectros son decrecientes, indicando predominio de bajas frecuencias, más acusado para $\alpha = 0,9$ (inferior). Las trayectorias correspondientes muestran oscilaciones menos frecuentes con largas permanencias de valores de un mismo signo.

FIGURA N° 6.3

“ESPECTRO Y TRAYECTORIA DEL PROCESO AR(1) PARA $\alpha = -0.5$ (SUPERIOR), $\alpha = 0.5$ (CENTRAL) Y $\alpha = 0.9$ (INFERIOR)”



Ejemplo (Un espectro divergente)

Consideremos el proceso:

$$X_t = A \cos \theta t + B \sin \theta t + Z_t,$$

Con $\{Z_t\}$ una sucesión de ruido blanco con varianza σ^2 , y A y B sendas variables aleatorias IID con media 0 y varianza γ^2 e independientes de Z_t . La media del proceso será por tanto $E(X_t) = 0$ y su varianza:

$$\text{var}(X_t) = E(A^2) \cos^2 \theta t + E(B^2) \sin^2 \theta t + E(Z_t^2) = \gamma^2 + \sigma^2.$$

Handwritten signature

La función de autocorrelación, para $t \neq s$,

$$\begin{aligned} R(t,s) &= E\{[A \cos \theta t + B \sin \theta t + Z_t][A \cos \theta s + B \sin \theta s + Z_s]\} \\ &= E(A^2) \cos \theta t \cos \theta s + E(B^2) \sin \theta t \sin \theta s \\ &= \gamma^2 \cos \theta(t-s). \end{aligned}$$

Por tanto:

$$R(x) = \begin{cases} \gamma^2 + \sigma^2, & \text{si } k = 0, \\ \gamma^2 \cos k\theta, & k \neq 0. \end{cases}$$

Y el proceso es WSS. Su espectro:

$$\begin{aligned} P(w) &= \gamma^2 + \sigma^2 + 2\gamma^2 \sum_{k \geq 1} \cos(k\theta) \cos(2\pi w k) \\ &= \gamma^2 + \sigma^2 + 2\gamma^2 \sum_{k \geq 1} [\cos k(\theta + 2\pi w) + \cos k(\theta - 2\pi w)], \end{aligned}$$

Que para $w = \theta / 2\pi$ diverge y no existe.

Filtros lineales

Un filtro es simplemente una transformación de un proceso X_t para dar lugar a otro proceso Y_t . Si la transformación es de la forma.

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j},$$

Hablamos de un filtro lineal como una combinación lineal, eventualmente infinita, de elementos del proceso original.

Si hay un número finito de a_j distintos de cero y X_t es WSS, también lo será Y_t , pero en el caso de que la combinación lineal sea estrictamente infinita, la WSS de Y_t depende de los a_j .

Si ambos procesos son WSS, es interesante establecer la relación entre $R_X(k)$ y $R_Y(k)$ y $P_X(w)$ y $P_Y(w)$. Comenzamos por la función de autocorrelación,

R

$$\begin{aligned}
R_Y(k) &= E(Y_t Y_{t-k}) \\
&= E\left[\left(\sum_{l=-\infty}^{\infty} a_l X_{t-l}\right)\left(\sum_{j=-\infty}^{\infty} a_j X_{t-k-j}\right)\right] \\
&= \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-l).
\end{aligned}$$

Para obtener la relación entre las funciones de densidad espectral bastara calcular la transformada de Fourier.

$$\begin{aligned}
P_Y(w) &= \sum_{k=-\infty}^{k=+\infty} R_Y(k) \exp(-i2\pi wk) \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-l) e^{-i2\pi wk} \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-l) e^{-i2\pi wl} e^{-i2\pi wj} e^{-i2\pi w(k+j-l)} \\
&= \sum_{k=-\infty}^{\infty} a_l e^{-i2\pi wl} \sum_{l=-\infty}^{\infty} a_j e^{-i2\pi wj} \sum_{j=-\infty}^{\infty} R_X(k+j-l) e^{-i2\pi w(k+j-l)}.
\end{aligned}$$

Si $h(w) = \sum_{l=-\infty}^{\infty} a_l e^{-i2\pi wl}$ mediante un cambio adecuado de índices. se puede escribir, $P_Y(w) = |h(w)|^2 P_X(w)$.

La función $h(w)$ recibe el nombre de función de transferencia del filtro lineal.

Ejemplo (Un proceso MA(3) de un proceso AR(1))

A partir del proceso AR(1):

$$X_t = \alpha X_{t-1} + Z_t,$$

Con Z_t un ruido blanco de varianza σ^2 , definimos un nuevo proceso:

$$Y_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1}),$$

Un proceso de medias móviles de orden 3, MA(3), que no es más que un filtro lineal con $a_j = 1/3, j = -1, 0, 1$ y $a_j = 0$ para cualquier otro j .



Para obtener $P_Y(w)$, recordemos el espectro del proceso AR(1) que obtuvimos en el ejemplo anterior.

$$P_X(w) = \frac{\sigma^2}{1 - 2\alpha \cos 2\pi w + \alpha^2}.$$

Por otra parte, la función de transferencia vale:

$$h(w) = \frac{1}{3}(e^{i2\pi w} + 1 + e^{-i2\pi w}) = \frac{1}{3}(1 + 2 \cos 2\pi w), \quad y$$

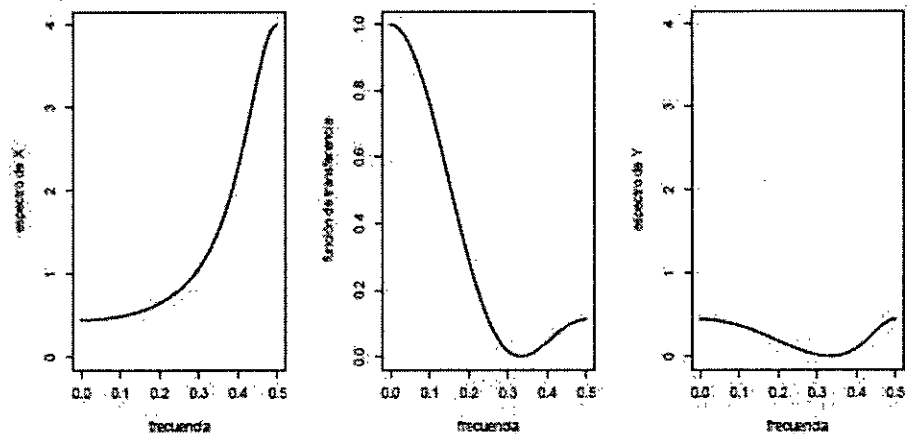
$$|h(w)|^2 = \frac{1}{9}(1 + 4 \cos 2\pi w + 4 \cos^2 2\pi w) = \frac{1}{9}(3 + 4 \cos 2\pi w + 2 \cos 4\pi w).$$

Finalmente:

$$P_Y(w) = \frac{\sigma^2(3 + 4 \cos 2\pi w + 2 \cos 4\pi w)}{9(1 - 2\alpha \cos 2\pi w + \alpha^2)}.$$

La figura muestra los espectros de ambos procesos y la función de transferencia que los relaciona, para $\sigma^2 = 1$ y $\alpha = -0,5$. El efecto del filtrado del proceso original es disminuir la varianza, menor área bajo el espectro, como resultado del suavizado que las medias móviles implican. El suavizado se evidencia de bajas frecuencias en el espectro de Y_t .

FIGURA N° 6.4
 “ESPECTROS Y FUNCION DE TRANSFERENCIA DE
 LOS PROCESOS MA(3) Y AR(1)”



[Handwritten signature]

La figura 6.4 muestra los espectros de ambos procesos y la función de transferencia que los relaciona para $\sigma^2 = 1$ y $\alpha = -0.5$

➤ **Filtrado lineal de Ruido Blanco. Proceso lineal general**

Resulta de filtrar una sucesión de ruido blanco, Z_t , con varianza σ^2 :

$$X_t = \sum_{j=0}^{\infty} a_j Z_{t-j}.$$

Se recuerda $P_Z(w) = \sigma^2$ y $h(w) = \sum_{j=0}^{\infty} a_j e^{-i2\pi wj}$.

El espectro del nuevo proceso valdrá:

$$\begin{aligned} P_Z(w) &= \sigma^2 |h(w)|^2 \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} a_j e^{-i2\pi wj} \sum_{k=0}^{\infty} a_k e^{i2\pi wk} \right) \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + \sum_{j \neq k} a_j a_k e^{-i2\pi w(k-j)} \right) \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + 2 \sum_{k \geq 1} \sum_{j < k} a_j a_k \cos 2\pi w(k-j) \right) \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + 2 \sum_{m \geq 1} \sum_{k > m} a_{k-m} a_k \cos 2\pi wm \right) \\ &= b_0 + \sum_{m \geq 1} b_m \cos 2\pi wm, \end{aligned}$$

$$\text{Con } b_0 = \sigma^2 \sum_{j \geq 0} a_j^2 \text{ y } b_m = \sum_{k > m} a_{k-m} a_k.$$

Este proceso se denomina proceso lineal general, y el caso de mayor interés en la práctica es el que se obtiene cuando solo un número finito de coeficientes a_j son distintos de 0. Los procesos ARMA, autoregresivos de medias móviles, son sin duda los más interesantes de entre los que cumplen esta condición.

La WSS de X_t no está garantizada aun cuando Z_t lo sea. Solo estaría en el caso de una combinación lineal finita. El caso más general requiere una condición sobre a_j .



Para obtenerla:

$$\begin{aligned}
 R(t,s) &= E(X_t X_s) \\
 &= E \left[\sum_{j=0}^{\infty} a_j Z_{t-j} \sum_{l=0}^{\infty} a_l Z_{s-l} \right] \\
 &= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} a_j a_l E(Z_{t-j} Z_{s-l}).
 \end{aligned}$$

Pero siendo Z_t un ruido blanco, $E(Z_{t-j} Z_{s-l}) \neq 0$ solo si $t-j = s-l$, por lo que se escribirá, $t \leq s$,

$$R(t,s) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+s-t}.$$

$E(X_t) = 0, \forall t$, y la expresión anterior evidencia que $R(t,s) = R(s-t) = R(k)$, con lo que el proceso será WSS si la $\sum_{j=0}^{\infty} a_j a_{j+k}$ es convergente $\forall k$. Ello supone que para $k=0$:

$$\sigma_{X_t}^2 = R(t,t) = R(0) = \sum_{j=0}^{\infty} a_j^2 < \infty.$$

Por otra parte, si $R(0)$ es finito, como $|\rho_{X_t, X_{t-k}}| \leq 1$, tendremos:

$$|\rho_{X_t, X_{t-k}}| = \left| \frac{R(k)}{\sqrt{\sigma_{X_t}^2 \sigma_{X_{t-k}}^2}} \right| = \left| \frac{R(k)}{R(0)} \right| \leq 1,$$

Despejando:

$$|R(k)| \leq R(0) < \infty.$$

En definitiva, la condición necesaria y suficiente para que el proceso lineal general sea WSS es que $\sum_{j=0}^{\infty} a_j^2 < \infty$.

6.3.2. PSD para procesos estocásticos WSS continuos en el tiempo

Definición. - Sea X_t un proceso estocástico WSS continuo en el tiempo, su densidad espectral de potencia, $P(w)$, es la transformada de Fourier de su función de autocorrelación $R(\tau)$,

$$P(w) = F[R(\tau)] = \int_{-\infty}^{+\infty} R(\tau) \exp(-i2\pi w\tau) d\tau, -\infty < w < \infty.$$

$P(w)$ se define como:

$$\lim_{T \rightarrow \infty} \frac{1}{T} E \left[\left| \int_{-T/2}^{T/2} X_t \exp(-i2\pi w\tau) d\tau \right|^2 \right],$$

Pero un resultado conocido como el teorema de Einstein-Wiener-Khinchin establece la igualdad anterior. De ambas definiciones y de las propiedades de $R(\tau)$ se derivan las siguientes propiedades para $P(w)$, análogas, como no podía ser de otra forma, a las que hemos visto en el caso discreto.

PSDc1) Del resultado anterior se dedujo que:

$$P(w) \geq 0, \forall w.$$

PSDc2) En la propiedad vista anteriormente, vimos que la función de autocorrelación de un proceso WSS es una función par, $R(\tau) = R(-\tau)$, en consecuencia, la PSD es una función real:

$$\begin{aligned} P(w) &= \int_{-\infty}^{+\infty} R(\tau) \exp(-i2\pi w\tau) d\tau \\ &= \int_{-\infty}^{+\infty} R(\tau) \exp(\cos 2\pi w\tau - i \operatorname{sen} 2\pi w\tau) d\tau \\ &= 2 \int_0^{+\infty} R(\tau) \cos 2\pi w\tau d\tau, \end{aligned}$$

Justificado por que la función seno es una función impar.

PSDc3) De la expresión anterior y la paridad de $R(\tau)$ y de la función coseno implican la paridad de PSD, $P(w) = P(-w)$.

Recordemos la relación entre las funciones de autovarianza y autocorrelación en un proceso WSS,

$$R(\tau) = C(\tau) + \mu^2.$$

Al calcular la PSD:



$$\begin{aligned}
P(w) &= F[R(\tau)] = F[C(\tau) + \mu^2] \\
&= F[C(\tau)] + F[\mu^2] \\
&= F[C(\tau)] + \mu^2 \delta(w),
\end{aligned}$$

Donde $\delta(w)$ es la función delta de Dirac, que es la transformada de Fourier de una función constante e igual a 1 en todo \mathbb{R} . El comentario que hicimos al final del ejemplo anterior es válido también ahora.

Observacion (El fenómeno de aliasing)

Al enumerar las propiedades de la PSD para el caso continuo hemos mencionado su analogía con las del caso discreto, pero se observará que ahora la periodicidad está ausente.

En la expresión anterior los límites de la integral son $-\infty$ y $+\infty$. La razón para ello es que cuando observamos un proceso en tiempo continuo es posible identificar cualquier componente por elevada que sea su frecuencia, por el contrario, si la observación se lleva a cabo a intervalos de tiempo, los efectos debidos a frecuencias superiores a $1/2$ o inferiores a $-1/2$ son indistinguibles de los efectos de baja frecuencia.

Este fenómeno se conoce con el nombre de "aliasing". Para ilustrarlo consideremos la función:

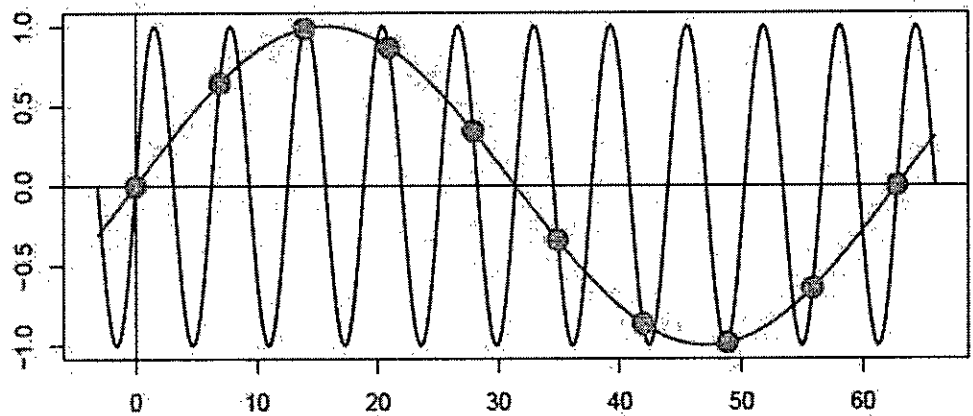
$$x(t) = \cos 2\pi t w,$$

Con $w = w_0 + 1/2$ y $0 < w_0 < 1/2$. La función es observada solo para tiempos enteros. Si tenemos en cuenta que $\cos x = \cos(x + 2\pi t)$ para t entero, y que $\cos x = \cos(-x)$, las distintas observaciones de $x(t)$ cumplen,

$$\begin{aligned}
x(t) &= \cos 2\pi t w \\
&= \cos(2\pi t w - 2\pi t) \\
&= \cos[2\pi t(w_0 + 1/2) - 2\pi t] \\
&= \cos(2\pi t w_0 - \pi t) \\
&= \cos(\pi t w - 2\pi t w_0) \\
&= \cos(2\pi t w'),
\end{aligned}$$

Con $w' = 1/2 - w_0$ que verifica $0 < w' < 1/2$. En definitiva, las frecuencias $w > 1/2$ no se distinguen de las frecuencias $w' < 1/2$. La figura muestra el fenómeno, se observa en ella que al muestrear s veces por segundo, las muestras de una y otra senoide se confunden. En procesamiento de señales se dice que cada una de las sinusoides se convierte en un alias para la otra, de ahí el nombre que este fenómeno recibe.

FIGURA N° 6.5
“EL FENOMENO DE ALIASING”



➤ **PSD versus R(k)**

Si la PSD es conocida, la transformada inversa de Fourier permite recuperar la función de autocorrelación:

$$R(\tau) = F^{-1}[P(w)] = \int_{-\infty}^{+\infty} P(w) \exp(i2\pi w\tau) dw.$$

De la definición de $R(\tau)$ se tiene que $R(0) = E(X_t^2)$. Aplicando lo anterior podemos calcular esta esperanza, conocida como potencia media de X_t , mediante:

$$E(X_t^2) = R(0) = \int_{-\infty}^{+\infty} P(w) dw.$$

R

Ejemplo (PSD del proceso RTS)

Recordemos que el proceso RTS es estacionario, y por tanto WSS, y que su función de autocorrelación viene dada por:

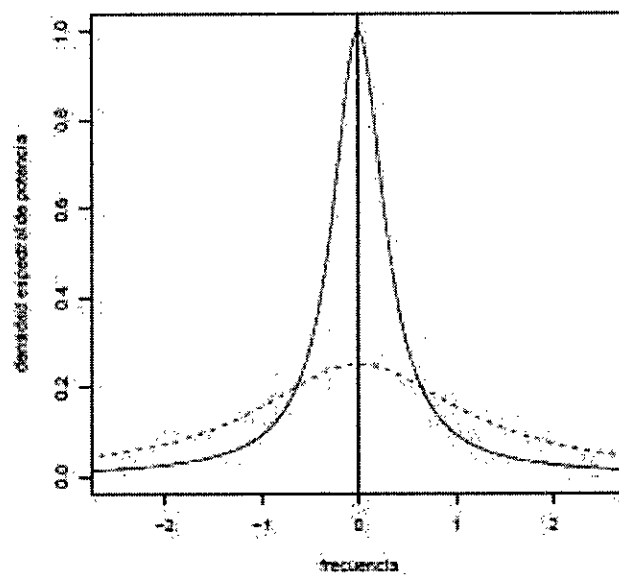
$$R(\tau) = a^2 e^{-2\lambda|\tau|},$$

Donde λ es el parámetro del proceso de Poisson subyacente. Para obtener $P(w)$:

$$\begin{aligned} P(w) &= \int_{-\infty}^{+\infty} a^2 e^{-2\lambda|\tau|} e^{-i2\pi w\tau} d\tau \\ &= a^2 \int_{-\infty}^0 e^{\tau(2\lambda - i2\pi w)} d\tau + a^2 \int_0^{+\infty} e^{-\tau(2\lambda + i2\pi w)} d\tau \\ &= \frac{a^2}{2\lambda - i2\pi w} + \frac{a^2}{2\lambda + i2\pi w} \\ &= \frac{\lambda a^2}{\lambda^2 + \pi^2 w^2}. \end{aligned}$$

FIGURA N° 6.6

“DENSIDAD ESPECTRAL DE POTENCIA DEL PROCESO RTS PARA $\lambda=1(\dots)$ Y $\lambda=4(---)$ ”



Esta figura muestra la gráfica de la densidad espectral de potencia para sendos procesos RTS con $\lambda = 1$ y $\lambda = 4$.

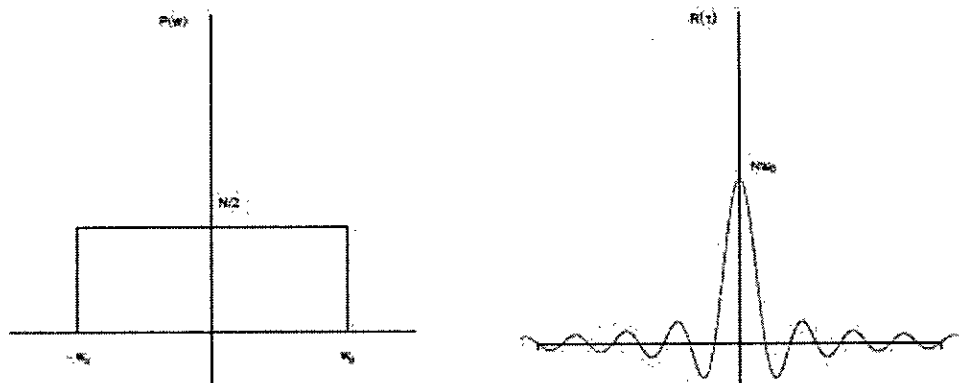
Ejemplo (PSD de un ruido blanco continuo)

La grafica de la izquierda de la figura 6.7 muestra la PSD de un ruido blanco recortada a un rango de frecuencias $-w_0 \leq w \leq w_0$. Se observa que $P(w)$ es constante en todo intervalo.

Ya comentábamos en el ejemplo anterior que el ruido blanco recibe este nombre por similitud con la luz blanca, cuyo espectro es constante para todas las frecuencias.

FIGURA N° 6.7

“ $P(W)$ CON FRECUENCIAS ACOTADAS PARA UN RUIDO BLANCO (IZQUIERDA) Y SU CORRESPONDIENTE $R(\tau)$ (DERECHA)”



Para obtener la función de autocorrelación recurrimos a:

$$\begin{aligned}
 R(\tau) &= \frac{N}{2} \int_{-w_0}^{w_0} \exp(i2\pi w\tau) dw \\
 &= \frac{N}{2} \times \frac{e^{-i2\pi w_0\tau} - e^{i2\pi w_0\tau}}{-i2\pi\tau} \\
 &= \frac{N \operatorname{sen} 2\pi w_0\tau}{2\pi\tau}
 \end{aligned}$$

A

La grafica de la derecha en la figura corresponde a $R(\tau)$, que anula en los puntos $\tau = \pm k / 2w_0, k = 1, 2, \dots$, lo que implica que las variables X_t y $X_{t+\tau}$ son incorreladas para dichos valores de τ .

La potencia del proceso vale:

$$E(X_t^2) = R(0) = \frac{N}{2} \int_{-w_0}^{w_0} dw = Nw_0,$$

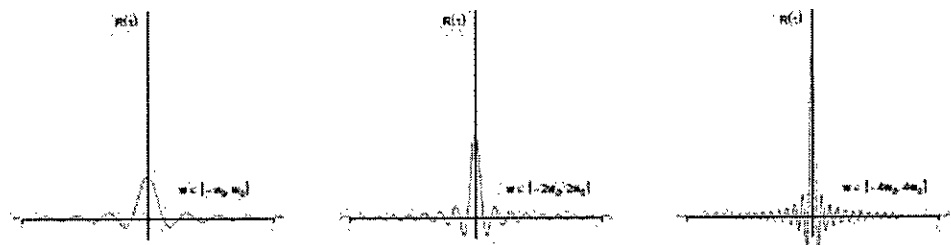
Que coincide con el máximo de la función $R(\tau)$, tal como se observa en la gráfica anterior.

A medida que $w_0 \rightarrow \infty$, el efecto sobre $R(\tau)$ es el que se muestra en las gráficas de la figura. En el límite, $P(w)$ será constante para cualquier frecuencia, hecho que caracteriza a un ruido blanco, y su función de autocorrelacion acumulará toda la potencia en 0 y valdrá por tanto infinito. Utilizando la función δ lo expresariamos:

$$R(\tau) = \frac{N}{2} \delta(\tau).$$

FIGURA N° 6.8

“EFECTO DEL CRECIMIENTO w_0 SOBRE $R(\tau)$ ”



Discretización temporal de un ruido blanco continuo

Anteriormente se ilustra $R(\tau)$ para un ruido blanco como límite, cuando $w_0 \rightarrow \infty$, de las $R_{w_0}(\tau)$ correspondientes a ruidos blancos filtrados del original mediante una banda de frecuencias limitadas por $|w| \leq w_0$. Es un modelo utilizado habitualmente en el tratamiento de señales.



Esta limitación real simplifica la descripción matemática de los procesos sin sacrificar su realismo.

Pero las simplificaciones no acaban con un filtrado paso bajo. Una segunda e inevitable simplificación es la que supone muestrear un proceso continuo para obtener un proceso discreto en el tiempo con el que realmente trabajaremos. Hemos pasado del proceso X_t a la sucesión X_n .

El nuevo proceso X_n , antes filtrado del original, puede representarse de la forma,

$$X_n = X_t |_{t=nT}, -\infty < n < +\infty,$$

Donde $T = f_0^{-1}$ es el periodo de muestreo y f_0 la frecuencia de muestreo. Si se quiere reconstruir la señal original a partir de la muestra, mediante un teorema se exige que f_0 sea mayor o igual que $2w_0$, la frecuencia de Nyquist.

Es inmediato comprobar que al igual que el proceso original, $E(X_n) = 0$. Calculemos ahora $R(k)$ para comprobar si la sucesión es también WSS.

$$\begin{aligned} R_{X_n}(k) &= E(X_n X_{n+k}) \\ &= E(X_{nT} X_{(n+k)T}) \\ &= E(X_{nT} X_{nT+kT}) \\ &= R_{X_t}(kT), \end{aligned}$$

Como X_n no depende de n es WSS. Además, se deduce que $R_{X_n}(k)$ es, a su vez, el resultado de muestrear la autocorrelación del proceso original con el mismo periodo T con el que se obtuvo X_n . En concreto, si $f_0 = 2w_0$, el periodo valdrá $T = (2w_0)^{-1}$ y $R_{X_n}(k) = R_{X_t}[k(2w_0)^{-1}]$,

Densidad espectral de potencia cruzada

Cuando tenemos dos procesos estocásticos, X_t e Y_t , ambos WSS, se define la función de autocorrelación cruzada mediante:

$$R_{XY}(\tau) = E(X_{t+\tau}Y_t).$$

Análogamente, se puede definir su densidad espectral cruzada, $P_{XY}(w)$, $F[R_{XY}(\tau)]$, que es una función compleja.

6.4. Estimación de la densidad espectral de potencia

La inferencia estadística nos enseña que la estimación de cualquier característica ligada a un fenómeno aleatorio exige la observación repetida del mismo, que proporciona lo que denominamos una muestra aleatoria. Así la media poblacional, μ , de una variable X podemos estimarla mediante la media muestral, $\bar{X}_n = \sum_{i=1}^n X_i$, obtenida a partir de una muestra aleatoria de tamaño n , X_1, \dots, X_n .

Lo esencial de este procedimiento es la repetibilidad del experimento, es la forma de conseguir las n observaciones que componen la muestra. Nuestro objetivo es ahora estimar las características ligadas a un proceso estocástico, en concreto su PSD, y el problema que se plantea es la imposibilidad de obtener una muestra del proceso. Se efectúa una única observación del proceso a lo largo del tiempo, también durante un periodo largo de tiempo. La interrogante que uno se plantea ¿Cómo obtener estimaciones en tales condiciones? Para ello es necesario el concepto de ergodicidad.

6.5. Ergodicidad

La ergodicidad es una propiedad por la cual las medias a lo largo del tiempo convergen a los valores esperados poblacionales. Existen diferentes tipos de ergodicidad, según la característica poblacional involucrada.

➤ Definición. (Ergodicidad en media)

Decimos que un proceso WSS es ergódico en media si la media temporal converge en media cuadrática a la media del proceso, μ ,

$$X_T = \frac{1}{2T} \int_{-T}^{+T} X_t dt \xrightarrow{m.s.} \mu, \text{ cuando } T \rightarrow \infty.$$

El estimador X_T es un estimador insesgado de μ , y su varianza depende de la función de autocovarianza.

➤ Definición (Ergodicidad en autocorrelación)

Decimos que un proceso WSS es ergódico en autocorrelación si

$$\frac{1}{2T} \int_{-T}^{+T} X_{t+\tau} X_t dt \xrightarrow{m.s.} R(\tau), \forall \tau.$$

Si definimos un nuevo proceso para cada k mediante:

$$Y_\tau(t) = X_{t+\tau} X_t,$$

La ergodicidad en autocorrelacion admite una caracterización alternativa, como nos muestra el siguiente teorema.

Teorema.- Un proceso estocástico WSS, X_t , es ergodico en autocorrelacion si y solo si:

$$\frac{1}{2T} \int_{-T}^{+T} \left(1 - \frac{|u|}{2T}\right) C_{Y_\tau}(u) du \xrightarrow{m.s.} 0, \forall \tau.$$

➤ **Periodograma: Definición y propiedades**

La estimación de la densidad espectral de potencia puede llevarse a cabo de dos formas distintas. La primera se basa en la definición de la PSD como $F[R(\tau)]$, supone estimar en primer lugar la función de autocorrelación y estimar a continuación la PSD mediante la transformada de Fourier de dicha estimación.

La segunda obtiene la PSD directamente de los datos a partir de lo que se conoce como el periodograma.

Definición (Periodograma)

Sean X_0, X_1, \dots, X_{n-1} , n observaciones de un proceso WSS, X_t , discreto en el tiempo. Sea $x_n(w)$ la transformada de Fourier discreta de dichas observaciones:

$$x_n(w) = \sum_{j=0}^{n-1} X_j \exp(-i2\pi wj).$$

Se define el periodograma, como el cuadrado medio de dicha transformada:

$$I_n(w) = \frac{1}{n} |x_n(w)|^2.$$

Si tenemos en cuenta que $|x_n(w)|^2$ es una medida de la energía del proceso en la frecuencia w , $I_n(w)$, la media temporal de dicha energía, es una estimación de la potencia en w . La justificación de porque el periodograma es un estimador adecuado para la PSD la encontramos al calcular la esperanza.

Tengamos primero en cuenta que:

$$|x_n(w)|^2 = \left[\sum_{j=0}^{n-1} X_j \exp(-i2\pi wj) \right] \left[\sum_{k=0}^{n-1} X_k \exp(i2\pi wk) \right],$$

Por tratarse de una variable aleatoria compleja. Calculemos ahora la esperanza.

$$\begin{aligned} E[I_n(w)] &= \frac{1}{n} E \left\{ \left[\sum_{j=0}^{n-1} X_j e^{-i2\pi wj} \right] \left[\sum_{k=0}^{n-1} X_k e^{i2\pi wk} \right] \right\} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} E(X_j X_k) e^{-i2\pi w(j-k)} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} R_X(j-k) e^{-i2\pi w(j-k)}. \end{aligned}$$

Esta expresión puede simplificarse si tenemos en cuenta que el argumento $j-k$ toma valores entre $-(n-1)$ y $n-1$, tantos como diagonales tiene el retículo $[0, 1, \dots, n-1] \times [0, 1, \dots, n-1]$, puesto que en cada una de esas diagonales $j-k$ toma el mismo valor. Así que, haciendo $j-k=m$,

$$\begin{aligned} E[I_n(w)] &= \frac{1}{n} \sum_{m=-(n-1)}^{n-1} (n-|m|) R_X(m) e^{-i2\pi wm} \\ &= \sum_{m=-(n-1)}^{n-1} \left(n - \frac{|m|}{n} \right) R_X(m) e^{-i2\pi wm}. \end{aligned}$$

Por tanto $I_n(w)$ es un estimador asintóticamente insesgado de $P_X(w)$, lo que justifica su elección. Es necesario que la varianza del estimador sea pequeña, en particular que sea asintóticamente 0. Pero no es el caso del periodograma, puede demostrarse que para un proceso WSS:

$$\text{var}[I_n(w)] \approx P^2(w).$$

Se trata de un estimador inconsistente, lo que puede conducir a estimaciones con gran variabilidad. En los textos de Diggle (1990) y Stark y Woods (2002) puede el lector encontrar un desarrollo de estos aspectos.

Si el proceso X_t es un proceso continuo en el tiempo, cuanto hemos dicho para el caso discreto puede extenderse de forma inmediata. Si observamos el proceso en el intervalo $[0, T]$, definimos el periodograma mediante

$$I_T(w) = \frac{1}{T} |x_T(w)|^2,$$

$$x_n(w) = \int_0^T X_t e^{-i2\pi w t} dt.$$

Desarrollando, nos dá:

$$E[I_T(w)] = \int_{-T}^{+T} \left(1 - \frac{\tau}{T}\right) R_x(\tau) e^{-i2\pi w \tau} d\tau,$$

Para $T \rightarrow \infty$ implica:

$$E[I_T(w)] \rightarrow P_x(w).$$



V. REFERENCIALES

- Barragués Fuentes, J. I., & Guisasola Aranzabal, J. (2006). La introducción de los conceptos relativos. *Enseñanza de las Ciencias* (2006), pp. 241-256.
- Brzézniak, Z., & Zastawniak, T. (1999). *Basic stochastic processes*. Londres: Springer.
- CMAPS, I. P. (s.f.). Obtenido de *Procesos estocásticos* [Figura]: https://cmapspublic.ihmc.us/rid=1304736021908_233554253_508/Procesos%20estocasticos.cmap
- Einstein, A. (1956). *Investigations on the theory of the brownian movement*. United States of America: dover.
- Feller, W. (1973). *Introducción de la teoría de probabilidades y sus aplicaciones*. Mexico: Limusa.
- Jones, P., & Smith, P. (2001). *Stochastic processes: An Introduction*. Londres: Hodder Education Publishers.
- Jurado, S. (20 de JUNIO de 2014). *Semana 15 procesos estocásticos series de tiempo*
Obtenido de https://cmapspublic.ihmc.us/rid=1304736021908_233554253_508/Procesos%20estocasticos.cmap
- Kunita, H. (1990). *Stochastic Flows and Stochastic Differential Equations*. Cambridge: Cambridge University Press.
- Lawler, G. (2006). *Introduction to stochastic processes: Whit applications to signal processing*. Boca Raton: Champan & Hall.
- Montes Suay, F. (2007). *Introducción a la Probabilidad*. Recuperado el 04 de Mayo de 2017, de <https://www.uv.es/montes/probabilitat/manual.pdf>
- Montes Suay, F. (2007). *Procesos Estocásticos para Ingenieros*. Recuperado el 04 de Mayo de 2017, de https://www.uv.es/montes/SPE/manual_material.pdf
- Stark, h., & Woods, J. (2002). *Probability and random processes: with applications to signal processing*. Prentice Hall N.J.
- Taylor, H., & Karlin, S. (1994). *An Introduction to Stochastic Modeling*. San Diego: Academic Press.
- Van Kampen, N. (1992). *Stochastic Processes in Physics and Chemistry*. Utrech: North- Holland Personal Library.

VI. APÉNDICES

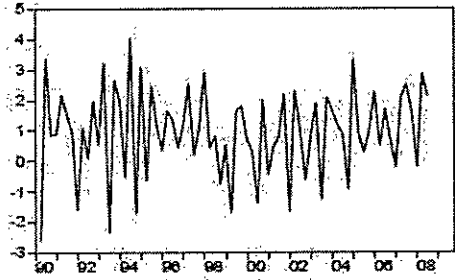

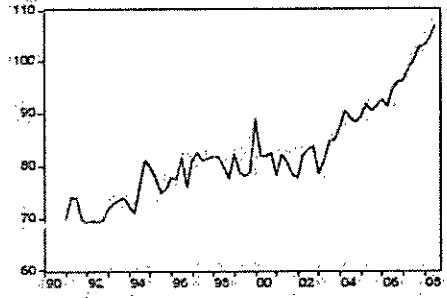
TABLA N° A-1: "TIPOS DE PROCESOS ESTACIONARIOS"

TABLA N° A-2: "ESTADO Y TIEMPO (DISCRETO Y CONTINUO)"

TABLA N° A-3: "EJEMPLO DE CADENA DE MARKOV"

TABLA N° 6.1

“TIPOS DE PROCESOS ESTACIONARIOS”


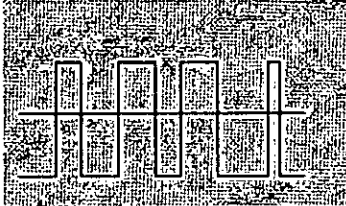
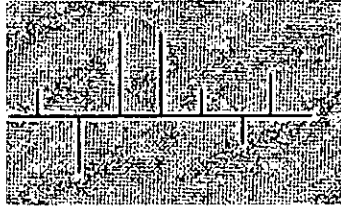
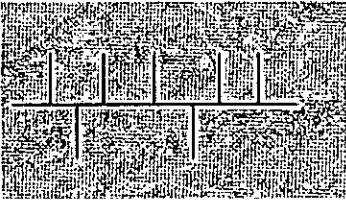
TIPOS DE PROCESOS ESTACIONARIOS	
ESTACIONARIO	<p>Proceso estocástico estacionario</p> 
ESTACIONARIO EN MEDIA	<p>Proceso estocástico estacionario en media</p> 
NO ESTACIONARIO	<p>Proceso estocástico no estacionario</p> 

FUENTE: Elaboración propia.



TABLA N° 4.1

“ESTADO Y TIEMPO(DISCRETO Y CONTINUO)”

		PROCESO ESTOCÁSTICO	
		CONTINUO	DISCRETO
TIEMPO	CONTINUO		
	DISCRETO		

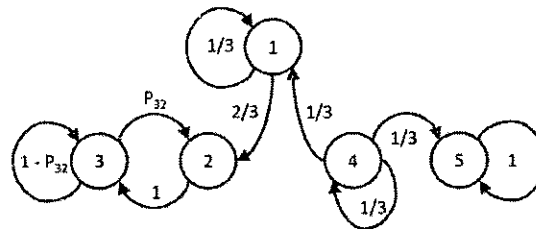
FUENTE: Elaboración propia.

Handwritten mark

TABLA N° 5.1
 “EJEMPLO DE CADENA DE MARKOV”

CADENA DE MARKOV

DESDE DIAGRAMA
 (EJEMPLO)



DESDE MATRIZ
 (EJEMPLO)

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 9/10 & 0 & 1/10 & 0 \\ 0 & 1/10 & 0 & 9/10 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

FUENTE: Elaboración propia.

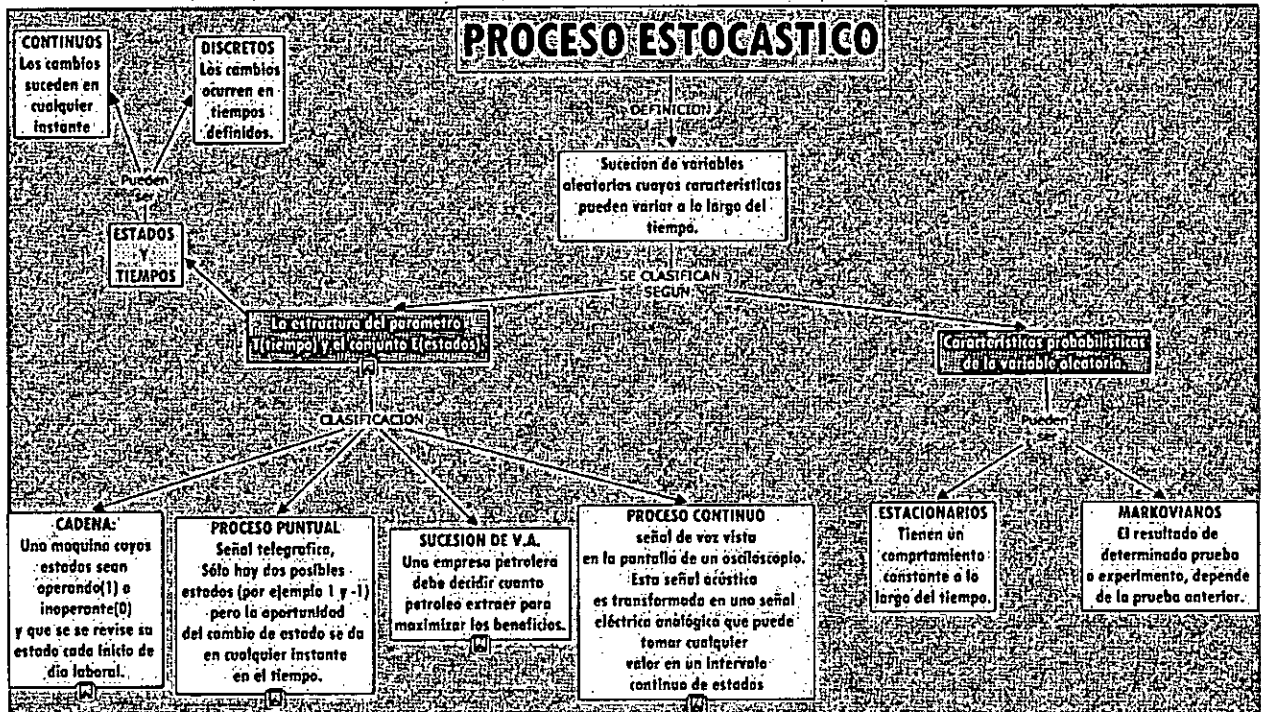
VII. ANEXOS

FIGURA N° 7.1: "PROCESOS ESTOCÁSTICO"

FIGURA N° 7.2: "ESQUEMA DE LOS PROCESOS ESTOCÁSTICOS"



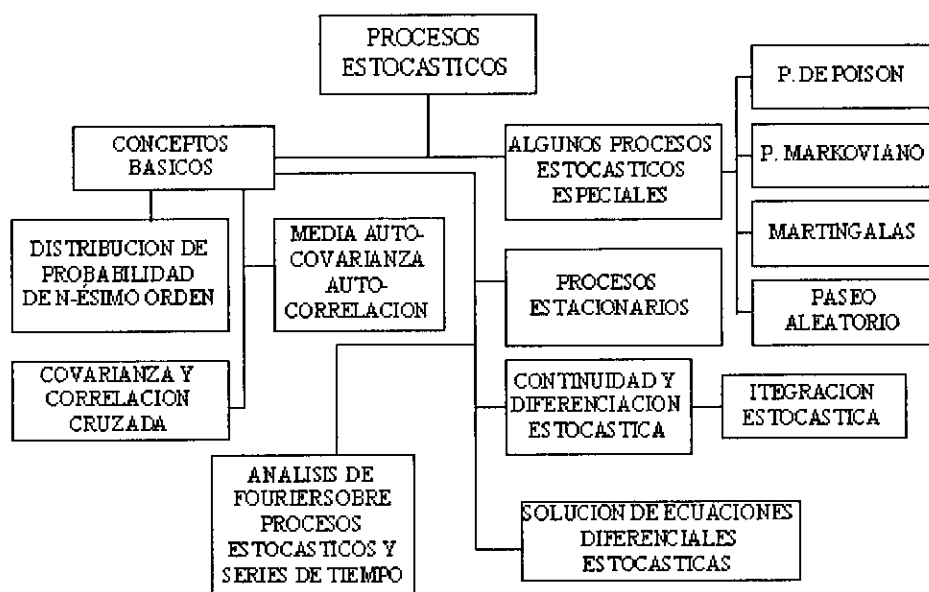
FIGURA N° 7.1
 "PROCESOS ESTOCÁSTICO"



FUENTE: *Procesos estocásticos*[Figura]. Recuperado de:
https://cmapspublic.ihmc.us/rid=1304736021908_233554253_508/Procesos%20estocasticos.cmap

FIGURA N° 7.2

“ESQUEMA DE LOS PROCESOS ESTOCÁSTICOS”



FUENTE: Jurado, S. (2014.) *Semana 15 procesos estocásticos series de tiempo* [Figura]. Recuperado de: [“https://cmapspublic.ihmc.us/rid=1304736021908_233554253_508/Procesos%20estocasticos.cmap”](https://cmapspublic.ihmc.us/rid=1304736021908_233554253_508/Procesos%20estocasticos.cmap)