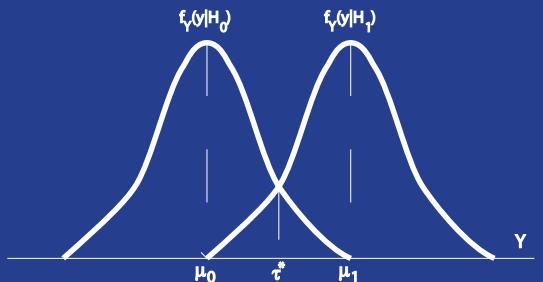


PROBABILIDAD, VARIABLES ALEATORIAS Y PROCESOS ESTOCÁSTICOS

Una introducción orientada a las Telecomunicaciones



PROBABILIDAD, VARIABLES ALEATORIAS Y PROCESOS ESTOCÁSTICOS

**Una introducción orientada a
las Telecomunicaciones**

Serie: INGENIERÍA, nº 12

ALBEROLA LOPÉZ, Carlos

Probabilidad, variables aleatorias y procesos estocásticos : una introducción orientada a las Telecomunicaciones / Carlos Alberola López. – Valladolid : Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial, 2004

260 p.; 24 cm. – (Ingeniería ; 12)
ISBN 978-84-8448-307-6

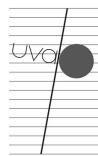
1. Telecomunicaciones – Procesos estocásticos – Variables aleatorias – Probabilidades
I. Alberola López, Carlos, aut. II. Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial, ed. III. Serie

519.21:621.39

CARLOS ALBEROLA LÓPEZ

PROBABILIDAD, VARIABLES ALEATORIAS Y PROCESOS ESTOCÁSTICOS

**Una introducción orientada a
las Telecomunicaciones**



**Universidad de Valladolid
Secretariado de Publicaciones
e Intercambio Editorial**

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros métodos, ni su préstamo, alquiler o cualquier otra forma de cesión de uso del ejemplar, sin el permiso previo y por escrito de los titulares del Copyright.

© CARLOS ALBEROLA LÓPEZ, VALLADOLID, 2004
SECRETARIADO DE PUBLICACIONES E INTERCAMBIO EDITORIAL
UNIVERSIDAD DE VALLADOLID

Primera edición: 2004
Segunda reimpresión: 2011

Diseño de cubierta: Miguel Ángel de la Iglesia
Álvaro Tapias García

ISBN: 978-84-8448-307-6
Dep. Legal: SE-3851-2004

Composición del autor

Imprime: Publidisa

*A quien fue,
a quien es,
a quien no pudo ser
y a quienes serán.*

Prólogo

Este libro es el resultado de los diez años de docencia en la materia de Teoría de la Probabilidad y Señales Aleatorias que, en el momento de redacción de este prólogo, el autor del mismo ha impartido en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad de Valladolid. Pretende ser, como objetivo prioritario, un manual que incluya los conocimientos probabilísticos básicos que se precisan para abordar el análisis y diseño de los Sistemas de Comunicación.

En el Plan de Estudios de la titulación de Ingeniero de Telecomunicación, tanto el que tuvo esta Escuela en su origen, como en el que actualmente está en vigor, la materia arriba mencionada se concentra fundamentalmente en una asignatura, a saber, *Señales y Sistemas de Transmisión I* en el plan antiguo, y *Señales Aleatorias en Ruido* (SAR, en adelante), en el plan actual. El resto de asignaturas que hacen uso de estos temas (fundamentalmente Teoría de la Comunicación, Teleráfico, Tratamiento Digital de la Señal y optativas afines así como alguna Telemática) construyen sobre los pilares de ésta. Por tanto, es SAR el soporte teórico sobre el que otras asignaturas se basan para impartir sus conocimientos específicos.

Este libro está diseñado para constituir el material docente, a impartir en un único cuatrimestre, de la única asignatura que proporciona formación en temas de probabilidad para los alumnos de este Plan de Estudios. Para el autor tal afirmación supone un reto; no es fácil sintetizar una materia tan amplia en un único cuatrimestre, partiendo además de un desconocimiento por parte de los alumnos prácticamente total de conceptos probabilísticos (la experiencia dice que es el caso). Se debe por tanto hacer una selección minuciosa de los temas a tratar, así como del grado de profundidad en su tratamiento, para conseguir transmitir esta materia persiguiendo, asimismo, que el alumno pueda llegar a verla como una poderosa herramienta que le sea útil para abordar problemas futuros.

Dado que la materia es clásica, es evidente que existe una enorme bibliografía al respecto. Sin embargo el autor (que es ingeniero del ramo) ha pretendido escribir un libro para ingenieros y para aspirantes a ello. Por ello las fuentes bibliográficas que se han tenido presentes (véase la sección de Bibliografía, epígrafe “Básica”) responden a este perfil de *probabilidad aplicada a las señales y comunicaciones*. De tales fuentes bibliográficas destaca, en el campo de la Ingeniería de Telecomunicación y, en concreto, en el área de conocimiento de Teoría de la Señal, la obra de Athanasios Papoulis, recientemente fallecido. Sin exageración ninguna, esta obra es referente mundial y cubre, con creces, los contenidos de la asignatura a que va dirigida.

do este libro. Sin embargo, como se dijo antes, el autor dispone de un único cuatrimestre para realizar su función. En tan poco tiempo esta obra resulta exhaustiva en exceso de forma que, si bien se tendrá presente a lo largo de todo el tratado, se debe realizar un trabajo de síntesis importante, síntesis que, a juicio de su autor, consigue el libro que ahora comienza. Se podría argumentar que existen otras obras también sintéticas; tal es el caso, por ejemplo, de la obra de Peebles, también referenciada en el epígrafe anteriormente indicado. No obstante, esta obra tal vez adolece de simplificación excesiva en determinados momentos, carencia que aquí hemos tratado de evitar.

Este libro no es un libro de estadística—hemos incluido, no obstante alguna referencia bibliográfica al respecto—, si entendemos por tal el conjunto de procedimientos a llevar a cabo para extraer conclusiones acerca de una población a partir de observaciones de una muestra. Por tanto, en el libro no se encontrará una sola mención, por ejemplo, del concepto de *intervalo de confianza*, ni de la estimación de *máxima verosimilitud*, ni se hablará de funciones como la *t de Student* o la *F de Snedecor*. Sin embargo, sí que se proporciona una base probabilística que permita al alumno, si resultase interesado en estos temas, avanzar en esa dirección. Excepción a lo dicho, esto es, descripción de un procedimiento estadístico, es el caso de la estimación de mínimo error cuadrático medio que se incluye al final del capítulo tercero. El objetivo de la impartición de tal concepto es proporcionar al alumno una pequeña base sobre la que construir, en otras asignaturas, procedimientos óptimos de filtrado de señales. Estaríamos hablando, concretamente, de proporcionar las bases para la comprensión del filtro de Wiener y, adicionalmente, del filtro de Kalman. Tampoco se describe formalmente el concepto de test de hipótesis (se hace una mención indirecta en el primer capítulo para ilustrar la trascendencia del Teorema de Bayes) ni se deducen las soluciones óptimas en sentido bayesiano ni en sentido Neyman-Pearson. Sin embargo, a lo largo del texto se plantean ejemplos de éstos y el alumno puede observar que, en un momento dado, dispone de las herramientas que le permiten calcular, pongamos, umbrales para conseguir un determinado requisito. Por tanto, este libro tiene como objetivo adicional proporcionar una base para poder profundizar, si es deseo del alumno, en temas de tratamiento estadístico de señales (estimación y detección) y reconocimiento de patrones, para lo cual hemos incluido en la bibliografía varias referencias enormemente conocidas.

El libro se ha estructurado en cinco capítulos, de los cuales los cuatro primeros formarían una unidad conceptual (probabilidad y variables aleato-

rias) y el quinto, segunda unidad conceptual, constituiría una introducción a los procesos estocásticos. El primer capítulo es una introducción a la Probabilidad, y el segundo describe el concepto de variable aleatoria. El tercero amplía el marco a dos variables aleatorias (haciendo especial hincapié en el concepto de comportamiento conjunto) y el cuarto extiende (con respecto a dos) el número de variables a considerar de forma conjunta. Los tres primeros capítulos contienen varios ejemplos de aplicación de los conceptos expuestos, y los capítulos cuatro y cinco finalizan con una sección de ejercicios ilustrativos, recopilatorios de los conceptos y procedimientos impartidos en las dos unidades conceptuales consideradas en el libro. Los ejercicios, que proceden fundamentalmente de exámenes de antiguas convocatorias de las asignaturas mencionadas arriba, se han resuelto en su totalidad y, algo que el autor considera de utilidad, contienen numerosos punteros a las secciones y ecuaciones que se requieren en cada momento. El autor entiende pues que, a pesar del número reducido de ejercicios incluidos, éstos abarcan un porcentaje muy elevado de los conceptos expuestos en el libro. Asimismo, los problemas del quinto capítulo introducen al lector en el mundo de las comunicaciones, tanto analógicas como digitales, con ejemplos de algunas modulaciones. De forma complementaria, algunos de los ejercicios del capítulo cuatro analizan el comportamiento de detectores, tanto de sistemas radar como de sistemas de comunicación.

Esto último hace que este libro no esté dirigido sólo a estudiantes de la titulación de Ingeniero de Telecomunicación, sino a todos los alumnos de las diversas titulaciones que se imparten actualmente en la Escuela de Telecomunicación de la Universidad de Valladolid. Todos ellos (ingenieros técnicos e ingenieros, de Telecomunicación o Electrónica) debe conocer los conceptos fundamentales de los Sistemas de Comunicación, y este libro proporciona una recopilación de los fundamentos probabilísticos en los que se basan. Para estos otros alumnos, igualmente, los problemas incluidos al final constituyen —a juicio del autor— una herramienta de interés teórico-práctico.

Finalmente, se debe hacer mención expresa de que este libro ha sido escrito bajo el supuesto de que el lector dispone de unos conocimientos básicos de Teoría de la Señal, en particular, de Sistemas Lineales. Se asume pues que el lector conoce fundamentos de señales, sistemas lineales e invariantes, la operación de convolución, señales básicas empleadas como soporte (nos referimos a funciones escalón y delta) y Transformada de Fourier. En el texto, por lo tanto, no se hace hincapié en tales cuestiones, sino que, simplemente, se emplean. No obstante, la necesidad de tales concep-

tos se produce de forma paulatina, de forma que si el lector de este prólogo no conoce los conceptos que se acaban de mencionar, pero planea leer este libro conjuntamente con otro de Sistemas Lineales, no debería tener mayor problema en tal simultaneidad.

Capítulo 1

Teoría de la Probabilidad

1.1 Introducción

La mayor parte de los fenómenos en la naturaleza tienen un carácter aleatorio si bien, en numerosos casos, se simplifican las situaciones con el fin de hacerlas abarcables y así se tratan de cerrar los experimentos en fórmulas determinísticas. Por ejemplo, si dejamos caer un objeto desde una cierta altura, asumimos que conocemos con qué aceleración se produce la caída y, asimismo, damos por supuesto que la posición de parada del objeto se producirá en la proyección vertical del mismo sobre el suelo. Tal aproximación deja de ser válida si se introducen otros factores que hacen del problema algo más complejo como, por ejemplo, la existencia de fuerzas descontroladas (digamos, la acción del viento). Asimismo, si medimos el tiempo de caída del objeto en varias ocasiones raramente conseguiremos dos medidas exactamente iguales. No es difícil por tanto aceptar que la realidad se puede aproximar mediante un modelo matemático, más un término de ambigüedad (al que típicamente se denomina ruido).

Podemos pensar en experimentos implícitamente aleatorios. Piénsese, por ejemplo, el tiempo que se invierte en llegar de un punto a otro en una ciudad mediante tráfico rodado. Claramente, este tiempo será uno u otro en función de las condiciones del tráfico, las cuales no son completamente predecibles.

Los sistemas de comunicación no son ajenos a estos escenarios de incertidumbre. En primer lugar, las señales que se manejan transportan información, de forma que no son caracterizables con exactitud mediante un conjunto finito de parámetros. Es decir, si una señal transporta información, no podremos predecir sin error —en general— el valor que toma

dicha señal en un cierto instante conocido el valor que ha tomado en instantes pasados. Tan sólo podremos hacer una caracterización parcial de la misma y así poder extraer conclusiones del tipo, por ejemplo, *qué valores típicamente toma la señal, con qué velocidad varía etc ...*, esto es, características promedio o características globales, ante nuestra incapacidad de hacer una caracterización más pormenorizada.

Asimismo, inherentes a los sistemas de comunicación son las señales aleatorias perturbadoras, superpuestas a las señales de interés. Las primeras, completamente inevitables en los sistemas de comunicación, deben ser estudiadas, con el fin de ser caracterizadas y así poder diseñar sistemas que sean inmunes, o que, cuanto menos, consigan minimizar el efecto de tales perturbaciones sobre las señales deseadas y, por ende, sobre la información que se trata de enviar de un punto a otro.

También en el diseño de sistemas de comunicación nos encontramos casos en los que se deben tomar decisiones de dimensionamiento en ambientes de incertidumbre. Por ejemplo, si pretendemos proporcionar un servicio telefónico en un edificio, una posibilidad de diseño sería colocar tantas líneas como individuos estén presentes en dicho edificio. De esta manera, al descolgar un teléfono, el usuario tendría seguridad de disponer de línea telefónica para hacer la llamada. No obstante, esta solución es claramente ineficiente, como dicta el sentido común. Parece mucho más razonable estudiar cómo es el tráfico que típicamente se generará en dicho edificio (en base, por ejemplo, a mediciones de tráfico solicitado en otros edificios de características similares) y diseñar el sistema de modo que la *probabilidad* de disponer de línea telefónica al descolgar el teléfono sea lo suficientemente elevada como para que el usuario, en general, disponga de la posibilidad de comunicarse, si bien en algunos—y poco frecuentes—casos esta persona tendrá que colgar y esperar unos instantes. De esta manera se consigue un uso de los recursos mucho más ajustado a la necesidad real.

Todos estos ejemplos ponen de manifiesto que debe haber un cuerpo de doctrina que, en base a datos recogidos de situaciones similares, nos permita caracterizar, aunque sea parcialmente, el comportamiento de un sistema que se vea sujeto a incertidumbre. Tal cuerpo doctrinal es la Teoría de la Probabilidad, objeto de este primer tema.

1.2 Álgebra de conjuntos

Como se pondrá de manifiesto en posteriores secciones, la teoría de la probabilidad¹ se asienta sobre el álgebra de conjuntos, de forma que es conveniente repasar, aunque sea de manera sucinta, los conceptos básicos en relación con la teoría de conjuntos. El lector interesado en un tratamiento más riguroso del que aquí se va a hacer puede consultar [2].

1.2.1 Definiciones, terminología y notación

Un conjunto se define genéricamente como una colección de elementos. Típicamente los elementos que forman parte de un conjunto tienen algún tipo de propiedad en común que les haga susceptibles de pertenecer al conjunto, pero tal requisito es meramente anecdótico. El conjunto se suele representar con una letra mayúscula, tipo A, B, C , etc., y los elementos del mismo se representan con una letra minúscula (a, b, c , etc.).

Sobre los conjuntos se define una relación de pertenencia, la cual se denota con el símbolo \in . Así pues, si el elemento a pertenece al conjunto A , este hecho se formaliza mediante la expresión $a \in A$. En el caso en que b no pertenezca a A se escribe $b \notin A$.

Respecto a la forma de descripción del conjunto, ésta se puede realizar de manera enumerativa, $A = \{a_1, a_2, \dots, a_n\}$, o bien, mediante la ley de formación a la que antes hacíamos referencia $A = \text{'los diez primeros números naturales'}$. Tal definición, como puede imaginarse, es equivalente a escribir de forma enumerativa $A = \{1, 2, 3, \dots, 10\}$.

Se define el cardinal de un conjunto como el número de elementos que forman parte de dicho conjunto. Si dicho cardinal es un número finito, el conjunto se denominará finito. Caso contrario será infinito. Dentro de estos últimos, debemos distinguir los de cardinal numerable, que serán aquellos cuyos elementos se pueden poner en relación 1:1 con los números enteros (por ejemplo, el conjunto de los números pares); por otra parte nos encontraremos conjuntos de cardinal no numerable, como, por ejemplo, el conjunto de los números reales comprendidos entre dos números a y b .

La relación de inclusión se deriva de la relación de pertenencia; un conjunto B se dice que está incluido dentro de un conjunto A si todos los elementos de B están en A . Si tal es el caso, podemos expresar de forma abreviada que $B \subset A$, o bien que $A \supset B$. Si se verifica que $B \subset A$ y que $A \subset B$ de forma simultánea, entonces es que los dos conjuntos son iguales.

¹Veremos en la sección 1.3.3 que esta teoría se define a partir de tres axiomas, de forma que, en sentido estricto, deberíamos denominarla *teoría axiomática de la probabilidad*.

Dos conjuntos se dice que son disjuntos si no tienen ningún elemento en común. A estos conjuntos se les denomina también mutuamente excluyentes.

Dado un problema, el conjunto universal, denotado por S , será el conjunto formado por todos los elementos del problema. De forma complementaria, el conjunto vacío, denotado por \emptyset será un conjunto sin ningún elemento. Como es natural, los conjuntos S y \emptyset son mutuamente excluyentes.

1.2.2 Operaciones con conjuntos

Para operar con conjuntos, una representación muy útil e intuitiva son los diagramas de Venn. Tales diagramas consisten, simplemente, en contornos cerrados en los que los elementos se representan como puntos. En base a tales diagramas la pertenencia de un elemento a un conjunto, la inclusión de un conjunto en otro, así como las operaciones que vamos a ver a continuación, se representan de forma natural y muy gráfica. Un ejemplo de diagrama de Venn puede verse en la figura 1.1.

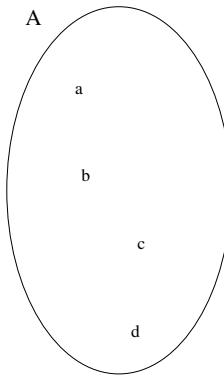


Figura 1.1: Ejemplo de diagrama de Venn.

Igualdad de conjuntos y conjunto diferencia

Como se ha indicado, dos conjuntos son iguales si el primero incluye al segundo, y viceversa. En tal caso se puede expresar que $A = B$.

Por otra parte, el conjunto diferencia de dos conjuntos A y B será el conjunto formado por aquellos elementos del primero que no están en el segundo, es decir si $C = A - B$ entonces $C = \{\forall a \in A, t.q. a \notin B\}$. La figura 1.2 ilustra mediante diagramas de Venn este concepto.

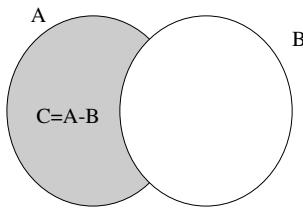


Figura 1.2: Conjunto diferencia.

Nótese que esta operación no es commutativa, es decir $A - B \neq B - A$.

Unión e intersección de conjuntos

Dados dos conjuntos A y B , el conjunto unión de éstos, denotado como $A \cup B$, es el conjunto formado por los elementos que pertenecen a cualquiera de los dos. De forma complementaria, el conjunto intersección, denotado por $A \cap B$, es el conjunto formado por los elementos que pertenecen a ambos conjuntos simultáneamente. Nótese pues que $(A \cup B) \supset A$ y $(A \cup B) \supset B$ mientras que $(A \cap B) \subset A$ y $(A \cap B) \subset B$.

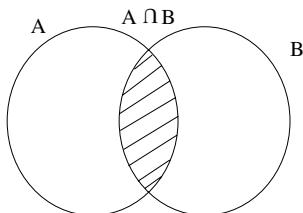
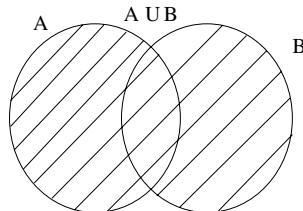


Figura 1.3: Unión e intersección de conjuntos.

En el caso en que se aplique esta operación de forma múltiple, la manera de escribirlo de forma compacta es la siguiente

$$A_1 \cup A_2 \cup \dots \cup A_N = \bigcup_{i=1}^N A_i \quad (1.1)$$

$$A_1 \cap A_2 \cap \dots \cap A_N = \bigcap_{i=1}^N A_i \quad (1.2)$$

Por razones que se verán posteriormente, es habitual en cálculo probabilístico escribir

$$A \cup B = A + B \quad (1.3)$$

$$A \cap B = AB \quad (1.4)$$

y, extendido al caso de operaciones múltiples,

$$A_1 \cup A_2 \cup \dots \cup A_N = \sum_{i=1}^N A_i \quad (1.5)$$

$$A_1 \cap A_2 \cap \dots \cap A_N = \prod_{i=1}^N A_i \quad (1.6)$$

Debe repararse que estos signos no tienen el significado algebraico de suma y multiplicación sino, como hemos dicho, de unión e intersección respectivamente.

Es interesante, por su utilidad, recordar que estas dos operaciones cumplen las propiedades conmutativa, asociativa y distributiva. Esta última hace referencia al uso combinado de las mismas, mientras que las dos primeras se refieren a cada una de las operaciones por separado. Concretamente, la conmutatividad implica que

$$A \cup B = B \cup A \quad (1.7)$$

$$A \cap B = B \cap A \quad (1.8)$$

Con respecto a la propiedad asociativa

$$A \cup (B \cup C) = (A \cup B) \cup C \quad (1.9)$$

$$A \cap (B \cap C) = (A \cap B) \cap C \quad (1.10)$$

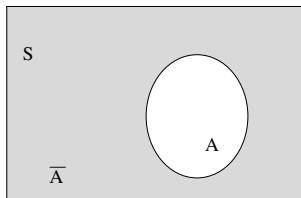
Finalmente la propiedad distributiva permite escribir

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (1.11)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (1.12)$$

Complemento de un conjunto

El conjunto complemento (o complementario) de un conjunto A (véase figura 1.4) es el conjunto diferencia entre el conjunto universal S y el conjunto A . Formalmente podemos escribirlo como $\overline{A} = S - A$. Es fácil ver que $A \cup \overline{A} = S$ así como que $A \cap \overline{A} = \emptyset$.

Figura 1.4: Conjunto A y su complementario \bar{A} .**Leyes de Morgan. Principio de dualidad**

Las leyes de Morgan constituyen una relación entre las operaciones de unión e intersección y la operación de complementación. Concretamente son:

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \quad (1.13)$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B} \quad (1.14)$$

Ambas expresiones pueden comprobarse con facilidad mediante diagramas de Venn.

Las expresiones anteriores, aplicadas de forma repetida, llevan a la afirmación: si en una expresión de igualdad de conjuntos se reemplazan todas las uniones por intersecciones, las intersecciones por uniones y los conjuntos por sus complementarios, la igualdad se mantiene.

Por ejemplo, el principio anterior llevaría a afirmar que, en vista de la igualdad (1.12), podemos escribir

$$\bar{A} \cup (\bar{B} \cap \bar{C}) = (\bar{A} \cup \bar{B}) \cap (\bar{A} \cup \bar{C}) \quad (1.15)$$

Finalmente, el principio de dualidad dice que si en una igualdad de conjuntos se reemplazan todas las uniones por intersecciones, todas las intersecciones por uniones, el conjunto S por \emptyset y viceversa, la igualdad se mantiene. Un ejemplo de ello serían las expresiones (1.11) y (1.12).

1.3 Definición de probabilidad

Se trata ahora de definir de forma precisa el concepto de probabilidad que vamos a manejar a lo largo de la asignatura. Dicho concepto, si bien lo tenemos todos asumido de alguna manera (piénsese, por ejemplo, en los apreciaciones que se hacen cuando se juega a las quinielas o a la ruleta) está basado en la intuición y no en un cuerpo matemático riguroso. Trataremos, por tanto, de conectar sentido común con rigor para crear un todo único.

Para ello, propondremos en primer lugar las definiciones más o menos intuitivas de la probabilidad, para luego integrarlas en un marco formal más ambicioso.

En lo que sigue haremos uso del concepto *experimento aleatorio*. Si bien éste puede definirse de forma rigurosa, entenderemos por tal *aquel experimento cuyo resultado no es completamente predecible*. Encajan en esta definición el lanzamiento de una moneda, el lanzamiento de un dado, el juego de la ruleta o, más cercano a las telecomunicaciones, la detección de un símbolo dentro de un alfabeto finito en un entorno ruidoso.

1.3.1 Probabilidad como frecuencia relativa

La probabilidad entendida como una frecuencia relativa se basa en la experimentación, y asocia a un determinado resultado de un experimento aleatorio una medida de *posibilidad de aparición* de dicho resultado mediante:

$$f_r(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N} \quad (1.16)$$

donde $f_r(A)$ representa la frecuencia relativa de aparición del resultado A , N_A el número de veces que se ha obtenido como resultado precisamente el A y N representa el número de ejecuciones del experimento.

Si bien es cierto que este planteamiento de la probabilidad encaja perfectamente con la intuición (y, de hecho, como se verá, la Ley de los Grandes Números—sección 4.5.3—asegura la convergencia de este parámetro a la probabilidad axiomática del resultado) emplearlo como base para la creación de una teoría matemática que nos permita calcular una probabilidades a partir de otras es bastante pobre, habida cuenta que estamos obligados a experimentar para conseguir obtener el valor deseado. No hay pues posibilidad de anticipar la probabilidad de resultados de un experimento en base al conocimiento de las probabilidades de otros experimentos relacionados.

1.3.2 Definición clásica

La definición clásica de la probabilidad se basa, a diferencia de la anterior, en un análisis del problema en vez de en la experimentación. Concretamente, dado un experimento aleatorio, la probabilidad asociada a un resultado A se define como

$$P_c(A) = \frac{N_A}{N} \quad (1.17)$$

donde, en este caso, $P_c(A)$ representa la probabilidad clásica asociada a dicho resultado, N_A representa el número de casos favorables a dicho resul-

tado y N el número total de casos posibles. Repárese que se han empleado la misma notación para ambas definiciones, pero los conceptos subyacentes son completamente distintos. Un ejemplo aclarará todavía más esta idea: si se pretende hallar la probabilidad de que salga el resultado *uno* ante el lanzamiento de un dado, para la definición clásica $N_A = 1$ y $N = 6$. Sin embargo, para la definición basada en frecuencia relativa es necesario experimentar. Si suponemos que se lanza el dado mil veces y que sale el resultado 1 doscientas veces, entonces $N_A = 200$ y $N = 1000$. Nótese por tanto que la definición clásica proporciona un valor único para la probabilidad del resultado deseado, mientras que el planteamiento basado en frecuencia relativa proporciona un resultado *distinto* (en general) cada vez que se realice el experimento.

A priori, pues, parece más interesante la definición de la probabilidad en su versión clásica, básicamente porque los resultados son siempre los mismos y, a su vez, porque no estamos obligados a experimentar. Sin embargo, debemos notar que esta definición no está exenta de inconvenientes. Concretamente:

1. Esta definición lleva implícito un concepto de equiprobabilidad en los resultados. ¿Qué sucedería en el ejemplo del dado si éste fuese un dado sesgado, es decir, si algún resultado tendiese a observarse de forma más habitual que otros? Nótese que el planteamiento clásico de la probabilidad no sería capaz de modelar correctamente esta situación, si bien a través del planteamiento basado en frecuencia relativa este sesgo sí que se haría patente.
2. ¿Qué sucedería si el número de casos posibles fuese infinito? En esta situación todas las probabilidades serían cero, si bien la realidad puede ser muy otra.

A pesar de sus inconvenientes, el planteamiento clásico es algo intuitivo y razonable, y funciona en un elevado número de casos. No obstante, tiene que plantearse algo más amplio tal que, englobando a los anteriores, no se quede atrapado en deficiencias como las comentadas.

1.3.3 Definición axiomática de la probabilidad

El planteamiento axiomático de la probabilidad hace uso del álgebra de conjuntos vista anteriormente; necesitamos, no obstante, alguna precisión terminológica antes de introducir la definición axiomática:

1. El conjunto universal que hemos denotado por S , recibirá ahora el nombre de espacio muestral o, alternativamente, y en base al segundo de los axiomas que definimos a continuación, suceso seguro. Complementariamente, el conjunto vacío recibe el nombre de suceso imposible.
2. Cada subconjunto del espacio muestral recibirá ahora el nombre de suceso. En particular, cada elemento del espacio muestral se denominará suceso elemental.

Con estos términos, el planteamiento de la probabilidad es el siguiente: dado un experimento con un espacio muestral S asociado, la probabilidad es una función que asocia a cada suceso del espacio muestral un número real, y que cumple los siguientes tres axiomas:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. $P(A \cup B) = P(A) + P(B) \quad \text{si} \quad A \cap B = \emptyset$

Nótese que de estos axiomas se pueden extraer conclusiones muy útiles, sin más que aplicar operaciones sobre conjuntos. Entre ellas

- $P(\overline{A}) = 1 - P(A)$. Esto es debido a que $S = A \cup \overline{A}$. Aplicando los axiomas 3 y 2 el resultado es directo.
- $P(\emptyset) = 0$. Nótese que esto es un caso particular del anterior, para cuando $A = S$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, si $A \cup B \neq \emptyset$. Esta consecuencia generaliza el axioma 3 y se deduciría de la forma siguiente:

$$\begin{aligned} A \cup B &= A \cup (B - A) = A \cup C \\ B &= C \cup (A \cap B) \\ P(A \cup B) &= P(A \cup C) = P(A) + P(C) \\ P(B) &= P(C) + P(A \cap B) \end{aligned}$$

Uniendo las dos últimas ecuaciones obtenemos el resultado anticipado.

- Si $A \subset B$ entonces $P(A) \leq P(B)$. Esta consecuencia se debe a que si A está incluida en B entonces podemos escribir $B = A \cup (B - A)$ y esta unión está formada por conjuntos disjuntos. Entonces, según el axioma tercero, $P(B) = P(A) + P(B - A)$, y acorde con el axioma 1, $P(B - A) \geq 0$.

Debe hacerse notar, por lo tanto, que la probabilidad de todo suceso es un número entre 0 y 1 (repárese que $\emptyset \subset A \subset S$).

Finalmente, demostremos que la probabilidad clásica está dentro de este planteamiento axiomático. Para ello, veamos que la definición proporcionada anteriormente cumple los axiomas. Así pues:

- $P_c(A) = \frac{N_A}{N} \geq 0$, dado que es el cociente entre dos magnitudes no negativas.
- $P_c(S) = \frac{N}{N} = 1$, pues todos los casos son favorables al espacio muestral.
- Suponiendo que los resultados A y B no puedan darse de forma simultánea, es decir, que $N_{A \cup B} = N_A + N_B$, entonces

$$P_c(A \cup B) = \frac{N_{A \cup B}}{N} = \frac{N_A + N_B}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P_c(A) + P_c(B)$$

Por ello, el planteamiento clásico forma parte del planteamiento axiomático.

Finalmente, podríamos hacer algo similar con respecto a la frecuencia relativa. No obstante, debemos esperar hasta el estudio de los teoremas asintóticos para demostrar que esto es así (en concreto, la ya indicada Ley de los Grandes Números, descrita en la sección 4.5.3). Por ahora, pidamos al lector un poco de fe.

Ejercicio: Se desea validar un sistema automático de medición de alturas de edificaciones. Con el objetivo de reducir costes en el proceso de validación se descarta la medición directa de los edificios para contrastar con la automática, de forma que se recurre a la medición por parte de n expertos, los cuales emplean sus propios sistemas de medida—y que, naturalmente, no están exentos de errores—. El sistema a validar se considerará apto para su uso si la medición que proporciona no se encuentra en ninguno de los dos extremos, esto es, no es ni la mayor ni la menor de las $n + 1$ mediciones. Bajo la hipótesis de

que los $n + 1$ equipos de medida proporcionen medidas similares (es decir, que todos funcionen correctamente, afectados por errores similares), obtenga la probabilidad de que el sistema sea validado.

Solución: Si las medidas son similares, el hecho de que la medida de un determinado equipo sea la mayor, o la menor, o caiga en cualquier otra posición con respecto a las demás, se debe exclusivamente al azar y no hay razón para pensar que exista tendencia por parte de ninguna medida a situarse en una determinada posición. Por ello, si se ordenan las medidas de manera creciente (o decreciente) cualquiera de las $n + 1$ medidas puede caer en cualquiera de las $n + 1$ posiciones de forma equiprobable. Consideremos la posición 1 como la asociada a la medida más pequeña y la posición $n + 1$ a la más grande. Sean los sucesos $A_i =$ “La medida automática cae en la posición i ”, $i = \{1, \dots, n + 1\}$ y $B =$ “el sistema automático es validado”. Entonces podemos escribir

$$\begin{aligned} P(B) &= 1 - P(\overline{B}) \\ &= 1 - P(A_1 \bigcup A_{n+1}) = 1 - (P(A_1) + P(A_{n+1})) \\ &= 1 - \frac{1}{n+1} - \frac{1}{n+1} = 1 - \frac{2}{n+1} = \frac{n-1}{n+1} \quad (1.18) \end{aligned}$$



Espacio de Probabilidad

Un experimento aleatorio se formalizará mediante la terna $\epsilon < S, \mathcal{F}, P >$. El primer elemento de la terna es el espacio muestral; el segundo es la clase de sucesos que podemos extraer del espacio muestral y el tercero es la ley de asignación de probabilidades.

Al respecto de la clase de sucesos debemos mencionar que si el espacio muestral tiene cardinal finito igual a N , la clase de sucesos tendrá un número de subconjuntos igual a 2^N . No obstante, si el cardinal del espacio muestral es infinito, el número de subconjuntos también lo será, de forma que podríamos pensar en uniones e intersecciones de infinitos sucesos. Para que el espacio de probabilidad esté bien definido la clase de sucesos a considerar debe tener estructura de *campo de Borel*², esto es, la clase debe ser cerrada frente a una cantidad numerable de aplicaciones de las operaciones

²También denominado σ -álgebra [3].

de complemento, unión e intersección entre los sucesos de la clase [2]. No obstante, no nos preocuparemos de estos detalles, y consideraremos que, a todos los efectos prácticos, esto será siempre así.

Finalmente, debemos repetir que el objetivo de la axiomática de la probabilidad es permitir calcular las probabilidades de unos sucesos en función de las probabilidades de otros que se asumen conocidas o que serán proporcionadas como datos del problema. De ahí que sea necesario, para definir completamente un experimento aleatorio, una ley de asignación de probabilidades. Éste es el significado del tercer elemento de la terna.

Respecto a dicho elemento, si el espacio muestral es finito, o si es infinito numerable, nos deben proporcionar la probabilidades de cada elemento del espacio muestral³. A tales probabilidades se les denominará probabilidades *elementales*. Si, por contra, el espacio muestral es infinito no numerable, deberemos ser informados de una ley continua de asignación de probabilidades. Esta ley será, según veremos en capítulos posteriores, una función de densidad de probabilidad o, dualmente, una función de distribución.

Ejemplo: Supongamos que un experimento aleatorio consiste en lanzar dos dados. Nos dicen que ambos dados son simétricos y que no existe interrelación entre los dados, esto es, que el lanzamiento del primero no afecta en nada al segundo, y viceversa.

Con esta información, el experimento aleatorio queda completamente definido. Veamos que es así:

$1^{\text{o}} \setminus 2^{\text{o}}$	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

1. El espacio muestral S viene dado por el conjunto de todos los posibles resultados, los cuales se muestran en la tabla adjunta. Como puede verse, consta de 36 elementos.
2. La clase de sucesos consta de todos los posibles subconjuntos que puedan extraerse de S . Sabemos de antemano que el número total de éstos es de $2^N = 2^{36}$.

³Estrictamente hablando, bastaría con que nos proporcionasen todas las probabilidades menos una, ya que la suma de todas ellas debe ser igual a la unidad.

3. Finalmente, no nos dicen de forma explícita la ley de asignación de probabilidades. Sin embargo, nos dicen que los datos son simétricos (luego no hay por qué pensar que los resultados no son equiprobables) y que además no existe influencia entre ellos (luego los resultados conjuntos mantienen esa equiprobabilidad). Así pues, la probabilidad de cada resultado elemental, digamos a_{ij} , con $1 \leq i, j \leq 6$ será igual a $\frac{1}{36}$ (consideremos que el índice i direcciona las filas de la tabla, y el j las columnas). En la sección 1.7 se discute una posible asignación de probabilidades en la que hay equiprobabilidad de cada dado por separado, pero no de ambos de manera conjunta.

Con ello podemos calcular la probabilidad de cualquier suceso que sea de interés. Por ejemplo, podemos calcular la probabilidad del suceso B =“la suma de las componentes es igual a 7”. Para ello, si denominamos $A_{ij} = \{a_{ij}\}$ entonces

$$\begin{aligned}
 P(B) &= P(A_{61} \cup A_{52} \cup A_{43} \cup A_{34} \cup A_{25} \cup A_{16}) \\
 &= P(A_{61} \cup (A_{52} \cup A_{43} \cup A_{34} \cup A_{25} \cup A_{16})) \\
 &= P(A_{61}) + P(A_{52} \cup A_{43} \cup A_{34} \cup A_{25} \cup A_{16}) \\
 &\quad \vdots \\
 &= P(A_{61}) + P(A_{52}) + P(A_{43}) + P(A_{34}) + \\
 &\quad P(A_{25}) + P(A_{16}) \\
 &= 6 \frac{1}{36} = \frac{1}{6}
 \end{aligned}$$

El paso de la primera línea a la segunda se ha realizado aplicando la propiedad asociativa de la unión (expresión 1.9). Ello, junto con el tercer axioma de la probabilidad, permite pasar a la tercera línea. La repetición de este proceso lleva a la penúltima igualdad y, finalmente, a su valor numérico.

1.4 Probabilidad condicionada

La probabilidad condicionada es una definición que se une a los axiomas para formar los pilares de la teoría de la probabilidad. Esta modalidad de probabilidad, la cual se denota por $P(A|B)$, mide cómo cambia nuestro conocimiento probabilístico de un suceso A cuando conocemos que otro suceso (digamos, B) se ha verificado.

El hecho de conocer que el suceso B se ha verificado implica que el experimento aleatorio con el que trabajamos ha cambiado. La situación puede expresarse como

$$\epsilon < S, \mathcal{F}, P > \xrightarrow{B} \epsilon_B < S_B, \mathcal{F}_B, P_B > \quad (1.19)$$

esto es, partimos del experimento ϵ y la verificación de B nos hace pasar a un experimento ϵ_B en el que las tres entidades involucradas (espacio muestral, clase de sucesos y ley de asignación de probabilidades) han cambiado. Es precisamente esta última la probabilidad condicionada.

La probabilidad condicionada puede calcularse en el experimento ϵ_B pero es práctica habitual calcularse directamente en el experimento original. Para ello se establece, por definición, que:

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)} \quad (1.20)$$

definición en la que se asume que el suceso condicionante B tienen probabilidad no nula, esto es, $P(B) \neq 0$. En lo que sigue, aunque no hagamos mención expresa de esta asunción, debemos tenerla en mente siempre que manejemos probabilidades condicionadas (de forma inmediata en las secciones 1.5.1 y 1.6, y a lo largo de los diferentes capítulos).

Esta nueva definición debe ser coherente con los axiomas de la probabilidad. Veamos que, en efecto, así es:

1. $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$, pues es el cociente de dos probabilidades axiomáticas y, por ello, el cociente de dos magnitudes no negativas.
2. $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$
3. Si $A \cap C = \emptyset$ entonces $P(A \cup C|B)$ debe ser igual a $P(A|B) + P(C|B)$.
En efecto,

$$\begin{aligned} P(A \cup C|B) &= \frac{P((A \cup C) \cap B)}{P(B)} \\ &= \frac{P((A \cap B) \cup (C \cap B))}{P(B)} \\ &= \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} = P(A|B) + P(C|B) \end{aligned}$$

puesto que si A y C son sucesos disjuntos, lo son también cualesquiera otros sucesos extraídos de éstos.

Ejemplo: Retomemos el ejemplo de los dos dados y pensemos que nos informan de que se ha verificado un suceso, llamémoslo B , consistente en que “ha salido al menos un seis en las tiradas”. Se pide que calculemos, en estas nuevas condiciones, la probabilidad del suceso A = “la suma de las componentes es igual a 7”.

Podemos calcular dicha probabilidad en ambos experimentos:

1. En experimento ϵ_B : el espacio muestral S_B ahora está formado por la última fila y la última columna de la tabla de resultados vista anteriormente, es decir $a_{ij} \in S_B$ si $i = 6$ ó $j = 6$. Por tanto $Card(S_B) = 11$, y habida cuenta de que los dados son físicamente los mismos, todos los resultados de este espacio muestral son equiprobables. Por tanto $P(a_{ij}|B) = \frac{1}{11} \forall a_{ij} \in S_B$. En estas condiciones:

$$\begin{aligned} P(A|B) &= P(A_{16} \cup A_{61}|B) = P(A_{16}|B) + P(A_{61}|B) \\ &= \frac{1}{11} + \frac{1}{11} = \frac{2}{11} \end{aligned}$$

2. En experimento ϵ : aplicando la definición de probabilidad condicionada (ecuación 1.20)

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A_{16} \cup A_{61})}{P(B)} \\ &= \frac{P(A_{16} \cup A_{61})}{P(\bigcup_{i=6 \text{ ó } j=6} A_{ij})} = \frac{\frac{2}{36}}{\frac{11}{36}} = \frac{2}{11} \end{aligned}$$

1.5 Independencia de sucesos

1.5.1 Independencia de pares de sucesos

Sean dos sucesos A y B con probabilidades distintas de 0. Se dice que los sucesos A y B son independientes entre sí si se verifica que

$$P(A|B) = P(A) \tag{1.21}$$

es decir, si la probabilidad de A condicionada a la verificación del suceso B coincide con la probabilidad incondicional de A . Esto significa que el hecho de que se haya verificado B no afecta, en absoluto, al suceso A , es decir,

A no es ni más probable ni más improbable cuando sabemos que B se ha verificado con respecto a cuando no lo sabíamos.

Nótese que si un primer suceso es independiente de un segundo, el segundo, como es natural, lo es del primero. Esto puede observarse de forma inmediata sin más que operar con la expresión anterior y con la definición de la probabilidad condicionada:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (1.22)$$

Si la condición anterior se cumple, entonces

$$P(A|B) = P(A) = \frac{P(B|A)P(A)}{P(B)} \rightarrow 1 = \frac{P(B|A)}{P(B)} \quad (1.23)$$

por ello, obtenemos la expresión de la condición de independencia con el suceso condicionante cambiado.

Una conclusión operativa de la independencia de sucesos es el hecho de que la probabilidad de la intersección de sucesos independientes se puede escribir como el producto de las probabilidades de cada uno de ellos. Esto se obtiene de forma directa dado que, para el caso que nos ocupa,

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

entonces, a partir de la segunda igualdad, despejamos la probabilidad de la intersección de los sucesos, y se deduce que

$$P(A \cap B) = P(A)P(B) \quad (1.24)$$

Esta expresión podemos tomarla como definición alternativa a la indicada en la expresión (1.21), la cual, asimismo, nos permite abordar de forma cómoda el caso de sucesos de probabilidad nula. En particular, a la vista de ésta podemos afirmar que si al menos uno de los dos sucesos involucrados tienen probabilidad nula entonces los sucesos son independientes. En efecto, supuesto que $P(A) = 0$, dado que $A \cap B \subset A$ entonces, recordando la cuarta consecuencia de los axiomas (página 23), podemos afirmar que $P(A \cap B) \leq P(A) = 0$. Por tanto $P(A \cap B) = 0 = P(A)P(B)$ y, por ello, los sucesos son independientes. Por otra parte, nótese que si dos sucesos de probabilidad no nula son disjuntos no pueden ser independientes; esto viene dado porque en la expresión anterior la probabilidad de la intersección sería cero (dado que los sucesos son disjuntos), lo cual sería distinto del producto

de la probabilidades de cada uno de ellos. Un razonamiento alternativo se basa en la intuición: si dos sucesos son disjuntos, no pueden verificarse de forma simultánea. Si es así, y se verifica uno de ellos, la verificación de éste impide la verificación del otro, lo cual está muy alejado del concepto de la independencia. Es por este motivo por el que los sucesos disjuntos reciben también el nombre de *mutuamente excluyentes*.

Ejemplo: Supongamos una baraja española de la que extraemos una carta. Calculemos las probabilidades siguientes:
a) $P(\text{"as"} | \text{"rey"})$ y b) $P(\text{"as"} | \text{"oros"})$.

a) $P(\text{"as"} | \text{"rey"}) = \frac{P(\text{"as"} \cap \text{"rey"})}{P(\text{"rey"})}$. Es obvio que si la carta que tenemos en la mano es un rey (como dice el condicionante) esa carta no puede ser un as. Por tanto, el suceso del numerador es el conjunto vacío y por tanto, la probabilidad que nos piden es nula. Estamos hablando pues de sucesos disjuntos y, por ello, no independientes.

b) $P(\text{"as"} | \text{"oros"}) = \frac{P(\text{"as"} \cap \text{"oros"})}{P(\text{"oros"})}$. Si tenemos una carta en la mano que es un oro, la probabilidad que dicha carta sea un as es 1 de 10 (1/10). En efecto, la probabilidad del numerador es la probabilidad de extraer *el as de oros*, esto es, 1 de 40, y la probabilidad del denominador es la de extraer un oro, la cual, habiendo diez oros en la baraja, es de 10 de 40.

Por otra parte, nótese que la probabilidad de extraer un as, dado que hay cuatro ases en la baraja, es 4 de 40, es decir, 1 de 10. Por tanto, $P(\text{"as"} | \text{"oros"}) = P(\text{"as"})$, y podemos concluir que ambos sucesos son independientes.

La interpretación parece clara: dado que en cada palo hay un as, el hecho de que salga un palo en particular no hace variar la probabilidad de que salga un as con respecto al hecho de no saber qué palo ha salido. Nótese que si en una hipotética baraja hubiese asimetría en algún palo, por ejemplo, hubiese un palo con más de un as, o hubiese un palo con más de diez cartas, tales sucesos ya no serían independientes.

1.5.2 Independencia de múltiple sucesos

La extensión del concepto de independencia al caso de N sucesos, digamos, A_1, A_2, \dots, A_N , es más o menos inmediata. Sin embargo, conviene hacer la precisión de que la independencia de N sucesos requiere el cumplimiento **de forma simultánea** de la independencia entre todos los subconjuntos (de al menos dos sucesos) que podamos extraer de tales sucesos. Por tanto, para que N sucesos sean independientes se requiere que se verifiquen las siguientes condiciones:

- Independencia de pares: $P(A_i \cap A_j) = P(A_i)P(A_j) \forall i \neq j$
- Independencia de tríos: $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \forall i \neq j, k; \forall j \neq k$
- ⋮
- Independencia global: $P\left(\bigcap_{i=1}^N A_i\right) = \prod_{i=1}^N p(A_i)$

Caso de que los sucesos no sean independientes, podemos hacer uso de la definición de probabilidad condicionada para expresar la probabilidad de la intersección múltiple de formas alternativas. Por ejemplo

$$\begin{aligned}
 P\left(\bigcap_{i=1}^N A_i\right) &= P\left(A_N \mid \bigcap_{i=1}^{N-1} A_i\right) P\left(\bigcap_{i=1}^{N-1} A_i\right) \\
 &= P\left(A_N \mid \bigcap_{i=1}^{N-1} A_i\right) P\left(A_{N-1} \mid \bigcap_{i=1}^{N-2} A_i\right) P\left(\bigcap_{i=1}^{N-2} A_i\right) \\
 &= P\left(A_N \mid \bigcap_{i=1}^{N-1} A_i\right) P\left(A_{N-1} \mid \bigcap_{i=1}^{N-2} A_i\right) \times \\
 &\quad P\left(A_{N-2} \mid \bigcap_{i=1}^{N-3} A_i\right) \cdots P(A_2|A_1)P(A_1)
 \end{aligned} \tag{1.25}$$

Esta forma de plantear la probabilidad global parece ser, a priori, más compleja que la inicial. Sin embargo, en muchas ocasiones se pueden hacer simplificaciones. Por ejemplo, en el caso en que los sucesos A_i estuviesen ordenados temporalmente, parece razonable pensar que la dependencia entre sucesos sea tanto menos intensa cuanto más alejados estén temporalmente.

De hecho, existirán muchas situaciones en que la expresión de arriba pueda simplificarse para dar lugar a:

$$P\left(\bigcap_{i=1}^N A_i\right) \sim P(A_N|A_{N-1})P(A_{N-1}|A_{N-2}) \times \\ P(A_{N-2}|A_{N-3}) \cdots P(A_2|A_1)P(A_1) \quad (1.26)$$

En tal caso, es decir, si lo anterior es razonablemente válido, el cálculo de la probabilidad de la intersección múltiple se simplifica notablemente.

1.6 Teoremas de la Probabilidad total y de Bayes

En esta sección presentaremos dos teoremas de enorme importancia en la Teoría de la Probabilidad. Para tal fin, introduciremos en primer lugar el concepto de *partición del espacio muestral*.

1.6.1 Concepto de partición

Una partición del espacio muestral consiste en un conjunto de sucesos A_i , $i = 1, \dots, N$, que cumplen las dos condiciones siguientes:

1. $A_i \cap A_j = \emptyset, \forall i \neq j;$

2. $S = \bigcup_{i=1}^N A_i$

Por tanto, una partición es un conjunto de sucesos disjuntos, tales que todos ellos unidos dan lugar al espacio muestral. Nótese que los elementos de la partición serían como las teselas de un mosaico.

Repárese en que hemos hablado de *una* partición del espacio muestral tratando así de enfatizar el hecho de que sobre un espacio muestral se puede definir, en general, un número arbitrario de particiones. Un caso particular de tal sería el conjunto de sucesos $A_1 = A$ y $A_2 = \overline{A}$, esto es, un suceso cualquiera y su complementario.

1.6.2 Teorema de la Probabilidad Total

Este teorema consiste en escribir la probabilidad de un suceso como una superposición de probabilidades de dicho suceso, condicionadas a cada uno de los elementos de una partición del espacio muestral. Concretamente, si B es un suceso del cual queremos hallar su probabilidad y $A_i, i = 1, \dots, N$

es una partición del espacio muestral, podemos escribir dicho suceso como una superposición de las partes del suceso que caen sobre cada elemento de la partición, es decir,

$$B = \bigcup_{i=1}^N (B \cap A_i) \quad (1.27)$$

Por tanto, podemos escribir $P(B)$ haciendo uso de la anterior igualdad

$$P(B) = P\left(\bigcup_{i=1}^N (B \cap A_i)\right) = \sum_{i=1}^N P(B \cap A_i) \quad (1.28)$$

dado que si los elementos de la partición son disjuntos, también lo serán los sucesos $B \cap A_i$, $i = 1, \dots, N$. Así pues, recordando la definición de probabilidad condicionada, podemos escribir

$$P(B) = \sum_{i=1}^N P(B|A_i)P(A_i) \quad (1.29)$$

que es la expresión del teorema.

Ejemplo: Podemos plantear el ejemplo propuesto en el apartado 1.4 mediante el teorema que acabamos de ver. Concretamente:

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) \\ &= \frac{2}{11} \frac{11}{36} + \frac{4}{25} \frac{25}{36} \\ &= \frac{2}{36} + \frac{4}{36} = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

1.6.3 Teorema de Bayes

Consecuencia de lo anterior, pero con un enorme trasfondo probabilístico que en este momento es difícil vislumbrar, es el teorema de Bayes. En este caso, tratamos de calcular la probabilidad de cada elemento de la partición supuesto que el suceso B se ha verificado. Así pues, podemos escribir

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} \quad (1.30)$$

y en virtud del teorema de la probabilidad total

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^N P(B|A_j)P(A_j)} \quad (1.31)$$

1.6.4 Sobre el teorema de Bayes y los sistemas de comunicaciones

El teorema de Bayes tiene un enorme interés para el planteamiento y resolución de muchos problemas de sistemas de comunicaciones y, como es natural, de otras muchas disciplinas. En lo que sigue, trataremos de hacer ver la importancia conceptual del mismo.

Pensemos que un sistema de comunicaciones puede enviar, en cada instante, uno de dos posibles símbolos. Al envío del primero, digamos, el símbolo lógico 0, le asociaremos el suceso H_0 . Al envío del segundo símbolo, léase, el símbolo lógico 1, le asociaremos el suceso H_1 . Tales sucesos⁴ llevan unas probabilidades asociadas, digamos, $P(H_0) = p_0$ y $P(H_1) = p_1$. Naturalmente $p_0 + p_1 = 1$. Estas probabilidades miden nuestro conocimiento previo de cómo de probable es el envío de cada uno de los dos símbolos. Por esta razón, es costumbre denominar a estas probabilidades *probabilidades a priori*, pues miden, en ausencia de observación ninguna, nuestro estado de conocimiento del problema.

Supongamos que a cada símbolo le asociamos una determinada señal física; por ejemplo, la hipótesis H_0 llevará asociada una señal de amplitud A_0 voltios, con una duración de T segundos. La hipótesis H_1 tendrá asociada una señal de amplitud A_1 voltios, con una duración de T segundos.

En un momento dado se envía uno de los dos símbolos, es decir, el transmisor envía una de las dos señales que se acaban de proponer. Como es natural, el receptor desconoce qué símbolo se ha enviado (si lo conociese no habría comunicación) y su misión es, en base a la observación obtenida, tomar una decisión sobre cuál de las dos señales se ha recibido y, por tanto, cuál de los dos símbolos se ha enviado.

Todo esto no sería problema si las señales recibidas fuesen exactamente iguales a las señales enviadas. En tal caso, dado que las señales en origen son distintas, las señales en destino también lo serían. Lamentablemente, la realidad es muy otra y las señales, conforme atraviesan el medio de transmisión, sufren distorsiones varias (recortes en banda, superposición de interferencias, adición de ruido etc...). Supongamos que la señal recibida tiene una amplitud de X voltios; el problema procede del hecho de que la amplitud X puede recibirse tanto si se envía un símbolo como si se envía el otro debido a los efectos de canal comentados. Ahora bien, debe haber *algo* que nos permita pronunciarnos—de forma no arbitraria—por un símbolo u otro a pesar de que los dos símbolos puedan, en un momento dado, dar

⁴La letra hache corresponde a la inicial de *hipótesis*.

lugar a la misma amplitud X .

Supongamos que tenemos la capacidad de calcular $P(X|H_0)$ y $P(X|H_1)$. Esto podría hacerse siempre que conociésemos qué tipo de distorsión sufre la señal. Para los modelos de canal típicamente empleados en la práctica, esto es perfectamente posible (de hecho, en buena medida el contenido de este tratado está encaminado a tal fin). A estas probabilidades se les denomina, generalmente, probabilidades de transición.

La cuestión es pues: recibidos X voltios, ¿a qué símbolo corresponde?

Podemos proporcionar varias respuestas al problema, cada una haciendo uso de diferentes grados de información:

1. Solución independiente de la observación: con independencia de la observación X recibida, el detector decide siempre que el símbolo enviado es el más probable a priori. Por tanto, el detector escoge el símbolo i de la forma siguiente:

$$\text{Escoger } H_i \text{ si } P(H_i) > P(H_j), i \neq j; i, j = \{0, 1\} \quad (1.32)$$

Como el lector puede imaginar, este tipo de decisión es muy pobre. El despreciar el valor numérico de la información recibida es un lujo que no es sensato permitirse. No obstante, si las probabilidades a priori son muy asimétricas (es decir, si una hipótesis es mucho más probable que otra) este detector no se equivocará con demasiada frecuencia.

2. Solución basada en maximizar la probabilidad de la observación: en este caso, el detector escogería como correcta aquella hipótesis que tiene asociada una mayor probabilidad de observar la tensión recibida; es decir, si se verifica

$$P(X|H_0) > P(X|H_1) \quad (1.33)$$

escogería la hipótesis H_0 y, si por el contrario, se verifica que

$$P(X|H_0) < P(X|H_1) \quad (1.34)$$

el detector escogería como cierta la hipótesis H_1 .

Obviamente, este criterio parece más razonable que el anterior puesto que la observación recibida aglutina mucha información y, además, no siempre tomaremos la misma decisión (como sucede en el caso anterior) sino que ahora la decisión de qué símbolo se ha enviado puede cambiar símbolo a símbolo en función de los datos concretos recibidos.

No obstante, si bien este modo de decisión parece más acertado, repárese que estamos despreciando una información muy valiosa, a saber, cómo de probable (o improbable) es, a priori, cada símbolo.

Cabe pues pensar en una tercera filosofía de decisión:

3. Escoger la hipótesis más probable dada la observación: en este caso el detector escogería la hipótesis H_0 si se verificase que

$$P(H_0|X) > P(H_1|X) \quad (1.35)$$

y escogería H_1 si, por el contrario,

$$P(H_0|X) < P(H_1|X) \quad (1.36)$$

Estas probabilidades reciben el nombre de probabilidades *a posteriori*, es decir, probabilidades una vez obtenida la observación.

No obstante, estas probabilidades no son conocidas. ¿O sí? Pues sí que lo son, dado que el teorema de Bayes nos permite escribir

$$P(H_0|X) = \frac{P(X|H_0)P(H_0)}{P(X)} < P(H_1|X) = \frac{P(X|H_1)P(H_1)}{P(X)} \quad (1.37)$$

Podemos ver que como nos interesa la comparación entre ambas, la probabilidad $P(X)$ no hay por qué calcularla. No obstante, si se desea conocer su valor, podemos hacerlo sin más que hacer uso del teorema de la probabilidad total.

Esta tercera filosofía es una filosofía bayesiana. Repárese que estamos *actualizando* la probabilidad a priori de cada hipótesis sobre la base de la información proporcionada por la observación recibida. Es decir, tenemos un conocimiento inicial del problema— $P(H_0)$ y $P(H_1)$ —pero la observación nos hace cambiar dicho conocimiento— $P(H_0|X)$ y $P(H_1|X)$ —a partir del cual tomamos una decisión.

Por todo ello, y a modo de resumen, el teorema de Bayes nos permite pasar de probabilidades a priori a probabilidades a posteriori, en base al conocimiento de las probabilidades de transición. En las secciones 2.4.1 y 2.4.2 incluimos un ejemplo numérico basado en esta filosofía.

1.7 Experimentos compuestos. Composición de ensayos de Bernoulli

Un experimento compuesto, como indica su nombre, consiste en la composición de experimentos simples, a los que, en este entorno, podemos llamar subexperimentos. Así pues, dados (al menos) dos experimentos aleatorios $\epsilon_1 < S_1, \mathcal{F}_1, P_1 >$ y $\epsilon_2 < S_2, \mathcal{F}_2, P_2 >$, podemos definir un experimento compuesto en base a la composición de éstos, el cual, asimismo, tendrá las tres componentes necesarias $\epsilon_c < S_c, \mathcal{F}_c, P_c >$. Veamos cada una de ellas:

1. S_c : El espacio muestral S_c vendrá dado por el producto cartesiano de los espacios muestrales S_1 y S_2 , es decir $S_c = S_1 \times S_2$. Consecuentemente, los sucesos elementales del espacio S_c consisten en los pares ordenados (a, b) , donde $a \in \mathcal{F}_1$ y $b \in \mathcal{F}_2$. En el ejemplo del lanzamiento de los dos dados que veíamos en la sección 1.3 dicho producto cartesiano dio lugar a una tabla de 36 elementos, la cual escribimos en su momento de forma natural. Otro ejemplo consistiría en que los espacios muestrales S_1 y S_2 fuesen la recta real. En tal caso $S_1 \times S_2$ sería el plano \mathbb{R}^2 .
2. \mathcal{F}_c : Denotemos por $A \times B$ el suceso formado por los pares ordenados (a, b) , donde $a \in A \in \mathcal{F}_1$ y $b \in B \in \mathcal{F}_2$. La clase de sucesos estará formada por sucesos de la forma $A \times B$, así como por sus uniones e intersecciones. Ejemplos de tales sucesos, de nuevo con el experimento de los dos dados, son *que la suma de las componentes sea igual a 4*, o *que aparezca al menos un seis en las tiradas*. Si los sucesos A y B fuesen sucesos definidos sobre la recta real, por ejemplo, $A = \{x_1 < x \leq x_2\}$ y $B = \{y_1 < y \leq y_2\}$, entonces el suceso $A \times B$ sería el rectángulo $\{x_1 < x \leq x_2, y_1 < y \leq y_2\}$.

Repárese que los sucesos $A \times S_2$ y $B \times S_1$ pertenecen a la clase \mathcal{F}_c , y que el suceso $A \times B$ puede escribirse como $A \times B = (A \times S_2) \cap (B \times S_1)$. Por otra parte, nótese que el suceso $A \times S_2$ se verifica siempre que en el primer experimento se verifique A , con independencia de lo que suceda en el segundo, ya que S_2 es el suceso seguro en el segundo experimento. De forma complementaria el suceso $B \times S_1$ se verifica siempre que en el experimento segundo se verifique B .

3. Ley de generación de probabilidades P_c : al contrario que los anteriores conceptos, la ley de asignación de probabilidades compuestas, en general, no viene únicamente caracterizada por el único conocimiento de las probabilidades P_1 y P_2 de cada subexperimento; esto

trae como consecuencia pensar que la composición de experimentos es una entidad formada por tres subentidades: el primer subexperimento, el segundo, y la forma de llevar a cabo la propia composición. Si no nos informan de alguno de los tres, no es posible, en general, conocer la ley de asignación de probabilidades del experimento compuesto. Un caso particular de esto es la composición de experimentos independientes. Dos experimentos serán independientes si el resultado de uno no afecta al resultado del otro. Formalmente los experimentos serán independientes si se verifica que los sucesos $A \times S_2$ y $B \times S_1$ lo son. Repárese que ambos sucesos pertenecen a la clase \mathcal{F}_c y, por tanto, podemos definir su independencia sobre la base de la expresión (1.24). Por ello, en caso de independencia de tales sucesos, podríamos escribir

$$\begin{aligned} P_c(A \times B) &= P_c((A \times S_2) \cap (B \times S_1)) \\ &= P_c(A \times S_2)P_c(B \times S_1) \\ &= P_1(A)P_2(B) \end{aligned} \quad (1.38)$$

donde la última igualdad procede del hecho de que $P_c(A \times S_2) = P(A)$ y $P_c(B \times S_1) = P(B)$. Nótese que en caso de independencia, por lo tanto, sí que es posible conocer P_c a partir del conocimiento de P_1 y P_2 .

La igualdad que acabamos de obtener justifica la notación que habitualmente emplearemos en la composición de experimentos. Dado que en el caso de experimentos independientes se verifica (1.38), expresión similar a la igualdad (1.24), escribiremos, aun resultando un abuso en la notación, $A \cap B$ en vez de $A \times B$. Esto debe entenderse como una simplificación operativa, puesto que la operación de intersección debe realizarse entre sucesos pertenecientes a la misma clase \mathcal{F} , lo cual, naturalmente, no es el caso. Esta notación, sin embargo, siempre que se tenga en mente la relación de orden en la composición de los experimentos, resulta muy cómoda para manejar casos más complejos (como es el propuesto en la sección 1.7.1).

Finalmente, ilustremos con un ejemplo que, en función de cómo sea la composición de experimentos, la ley de asignación de probabilidades puede ser, o no, determinada a partir de las leyes individuales de cada subexperimento:

- Consideremos el ejemplo del lanzamiento de dos dados. Asumamos que los dados son simétricos e independientes. Definamos

el suceso A_i como “ha salido la cara i -ésima en el primer dado”, y el suceso B_j como “ha salido la cara j -ésima en el segundo dado”, donde, naturalmente, tanto i como j varían en el rango $1, \dots, 6$.

Con esta información, si definimos el suceso C_{ij} como “ha salido la cara i -ésima en el primer dado, y la cara j -ésima en el segundo”, podemos escribir que

$$P(C_{ij}) = P(A_i \cap B_j) = P(A_i)P(B_j) = \frac{1}{6} \frac{1}{6} = \frac{1}{36} \quad (1.39)$$

En la expresión anterior la probabilidad $P(C_{ij})$ sería una probabilidad calculada en el experimento compuesto, mientras que las probabilidades $P(A_i)$ y $P(B_j)$ (con los índices i y j variando en los intervalos oportunos) procederían de cada uno de los subexperimentos.

- Imaginemos ahora que simulamos mediante ordenador el experimento de lanzamiento de dos dados, ambos simétricos. Esto sería sencillo, sin más que generar aleatoriamente dos números naturales entre 1 y 6. Supongamos, no obstante, que en nuestro programa incluimos una sentencia que diga *si en el primer dado sale un cinco, en el segundo no puede salir un tres y si en el primer dado sale un tres, en el segundo no puede salir un cinco*. Con este programa, escogiendo oportunamente las probabilidades de los sucesos elementales compuestos, sería perfectamente posible que las caras de cada dado (por separado) fuesen equiprobables.

En estas condiciones, el único conocimiento de la equiprobabilidad de las caras de cada dado no sería suficiente para conocer la ley de asignación de probabilidades del experimento compuesto.

Ejercicio: Haciendo uso del teorema de la probabilidad total, podemos escribir la probabilidad de cada cara de cada dado de la forma siguiente:

$$P(A_i) = \sum_{j=1}^6 P(A_i|B_j)P(B_j) = \sum_{j=1}^6 P(A_i \cap B_j) \quad (1.40)$$

$$P(B_j) = \sum_{i=1}^6 P(B_j|A_i)P(A_i) = \sum_{i=1}^6 P(A_i \cap B_j) \quad (1.41)$$

Con la restricción $P(A_i) = P(B_j) = \frac{1}{6}$ $i, j = \{1, \dots, 6\}$ construya una tabla con los valores $P(A_i \cap B_j)$ que cumplan, además, el condicionante de que las caras 3 y 5 no se pueden dar conjuntamente.

Solución: Las expresiones anteriores—ecuaciones (1.40) y (1.41)—significan que la probabilidad de cada cara se obtendría sumando por filas las probabilidades de cada resultado elemental del experimento compuesto para el primer dado y por columnas para el segundo. Así pues, una posible tabla sería

$A_i \setminus B_j$	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	2x	x	0	x
4	x	x	x	x	x	x
5	x	x	0	x	2x	x
6	x	x	x	x	x	x

donde $x = \frac{1}{36}$. Nótese que nos hemos limitado a paliar el efecto de la probabilidad cero en los resultados (3, 5) y (5, 3) forzando probabilidad doble en los resultados (3, 3) y (5, 5), sin necesidad de afectar a los demás elementos de la tabla.

1.7.1 Composición de ensayos de Bernoulli

Un caso particular de enorme interés en la práctica es la composición de ensayos de Bernoulli independientes. Veremos que tal tipo de composición es enormemente frecuente y, adicionalmente, es la base para la introducción de la ley de los grandes números.

En primer lugar definamos y propongamos la notación a emplear en los ensayos de Bernoulli. Un experimento aleatorio es un ensayo de Bernoulli si los resultados se engloban en dos (y sólo dos) sucesos. Tales sucesos los denotaremos por A y \bar{A} , y les asignaremos las probabilidades p y $q = 1 - p$ respectivamente. Nótese que, como es natural, tales sucesos forman una partición del espacio muestral, de forma que son disjuntos y sus probabilidades suman la unidad. Nótese, asimismo, que no estamos hablando de

que el número de resultados del experimento sea sólo dos, sino que estamos dividiendo las soluciones en dos subconjuntos. Así pues, el lanzamiento de un dado y la observación del resultado obtenido puede considerarse un ensayo de Bernoulli sin más que dividir los resultados en, por ejemplo, A =“ha salido un valor superior a tres”, y su complementario.

En este apartado nos interesa analizar la composición de tales ensayos. Supongamos que llevamos a cabo varias ejecuciones de un ensayo de Bernoulli de forma independiente entre ellas (recuérdese que esta asunción permite emplear el resultado de la expresión 1.38). Pensemos que el número de veces que realizamos el experimento es N . En estas condiciones, la cuestión a resolver es ¿cuál es la probabilidad de que el resultado A haya salido k veces (de N posibles)?

Si denominamos B_k al suceso cuya probabilidad buscamos conocer, el proceso para el cálculo de la misma será, básicamente, escribir dicho suceso B_k como función de sucesos cuya probabilidad sea conocida. Como, por ahora, sólo conocemos la probabilidad de cada resultado elemental, tendremos que escribir dicho suceso como función de los resultados elementales en cada uno de los ensayos.

Denominando B_k^j a cada una de las posibles formas de darse el suceso B_k , y denotando por A_i al suceso “ha salido el resultado A en el ensayo i -ésimo”, podremos escribir⁵:

$$B_k^1 = A_1 \cap A_2 \cap \dots \cap A_k \cap \overline{A}_{k+1} \dots \overline{A}_N \quad (1.42)$$

$$B_k^2 = \overline{A}_1 \cap A_2 \cap \dots \cap A_k \cap \overline{A}_{k+1} \cap \overline{A}_{k+2} \cap \dots \cap \overline{A}_{N-1} \cap A_N \quad (1.43)$$

...

$$B_k^M = \overline{A}_1 \cap \overline{A}_2 \cap \dots \cap \overline{A}_{N-k} \cap A_{N-k+1} \cap \dots \cap A_N \quad (1.44)$$

En este momento ya se encuentra los sucesos escritos en función de la intersección de sucesos cuyas probabilidades conocemos. Obsérvese que, si suponemos conocido el número de formas M en que puede darse el suceso B_k entonces podríamos escribir

$$B_k = \bigcup_{j=1}^M B_k^j \quad (1.45)$$

No obstante, quedan cuestiones importantes por resolver:

⁵En lo que sigue emplearemos la notación simplificada que acabamos de proponer para la composición de experimentos. En sentido estricto, y para el caso del suceso B_k^1 —por ejemplo—el suceso A_1 debería escribirse $A_1 \times S_2 \times \dots \times S_N$.

1. ¿Son los sucesos B_k^j disjuntos?
2. ¿Podemos calcular las probabilidades de cada B_k^j en función de p , q , N y k ?
3. ¿Tenemos que ser exhaustivos en la tabla anterior, es decir, debemos escribir todas las formas en que puede darse el suceso B_k ?
4. ¿Cuánto vale M ?

Tendremos que ir resolviendo cada pregunta:

1. Los sucesos son disjuntos. Obsérvese que cada suceso representa una forma de darse el resultado B_k , por ello, o se da de una forma, o se da de otra, pero no puede hacerlo de dos maneras a la vez. Esto puede verse de forma rigurosa escogiendo dos sucesos cualesquiera. Por ejemplo, escogiendo B_k^1 y B_k^2 , si realizamos la operación

$$B_k^1 \cap B_k^2 = (A_1 \cap \overline{A}_1) \cap \cdots \cap (A_N \cap \overline{A}_N)$$

Nótese que, con independencia del resto de las intersecciones, la primera es el conjunto vacío, de forma que la intersección global también lo es y por ello los sucesos son disjuntos.

Así pues podremos escribir

$$P(B_k) = P\left(\bigcup_{j=1}^M B_k^j\right) = \sum_{j=1}^M P(B_k^j) \quad (1.46)$$

2. Para calcular la probabilidad de cada B_k^j escojamos de forma genérica uno de ellos, por ejemplo, el suceso B_k^1 . Nótese que hablamos de una probabilidad de una intersección múltiple, concretamente de

$$P\left(\bigcap_{i=1}^k A_i \bigcap_{j=k+1}^N \overline{A}_j\right).$$

Dado que la composición de experimentos se ha hecho de forma independiente, los sucesos A_i son independientes entre sí, de forma que la probabilidad de la intersección se puede escribir como el producto de las probabilidades, es decir,

$$P(B_k^1) = P\left(\bigcap_{i=1}^k A_i \bigcap_{j=k+1}^N \overline{A}_j\right) = \prod_{i=1}^k P(A_i) \prod_{j=k}^N P(\overline{A}_j) = p^k q^{N-k} \quad (1.47)$$

3. La respuesta a la segunda pregunta proporciona también respuesta a la tercera. La ecuación anterior indica que la probabilidad de cada suceso B_k^j sólo es función de cuántos resultados A aparecen y no de su posición dentro de los N posibles posiciones. Por tanto, la probabilidad anterior es común para $B_k^j \forall j = 1, \dots, M$ de forma que la probabilidad (1.46) podremos escribirla

$$P(B_k) = Mp^k q^{N-k} \quad (1.48)$$

y el problema queda reducido a la determinación del parámetro M .

4. ¿Cuánto vale M ? Para responder a esta pregunta podemos acudir a varios planteamientos. Uno de ellos es el siguiente:

El número buscado M se puede encontrar de la forma siguiente: disponemos de N posiciones donde introducir k veces el resultado A . En las $N - k$ restantes tendremos que introducir el resultado \bar{A} . Tomemos como referencia las ejecuciones del experimento de la ecuación (1.42) y siguientes. En el primer caso, introducimos el resultado A en las k primeras posiciones. Así pues, podemos extraer el subconjunto de índices $\{1, 2, \dots, k\}$ donde hemos introducido el resultado A . En la segunda, dicho subconjunto de índices será $\{2, 3, \dots, k, k + 1\}$. Por tanto, si extraemos todos los subconjuntos de k índices donde podemos introducir el resultado A habremos extraído todas las formas de presentarse el resultado A k veces. Y, ¿cuántos subconjuntos de k elementos podemos extraer de un conjunto de N elementos? La respuesta a esta pregunta son las combinaciones de N elementos tomados de k en k y dicho número resulta ser $M = \binom{N}{k}$.

Así pues, la expresión de la probabilidad que estamos buscando es

$$P(B_k) = \binom{N}{k} p^k q^{N-k} \quad (1.49)$$

Ejercicio: Un sistema de control de calidad de productos químicos decide, para cada producto que inspecciona, si éste cumple los requisitos normativos para su aceptación. Supongamos que el protocolo de inspección rechaza los productos con una probabilidad p_r y el proceso de fabricación es tal que permite aceptar independencia entre productos. Se pide:

- a) Si se dispone de N productos, probabilidad de que una fracción no superior al 10% de ellos sea rechazada.

b) Probabilidad de que k productos sean rechazados antes de aceptar un número s de éstos.

Solución:

a) Denominemos k_r al primer número entero menor que $0.1N$ ($k_r = \text{floor}(0.1N)$). Dado que los experimentos son independientes entre sí y el resultado de cada experimento es rechazar el producto (suceso A , con probabilidad $p(A) = p_r$) o aceptarlo (suceso \bar{A} , con probabilidad $q_r = 1 - p_r$), estamos ante una composición de ensayos de Bernoulli independientes. En este caso el suceso (digamos) B cuya probabilidad se pide se puede escribir, siguiendo la definición de los sucesos B_k dada por la ecuación (1.45), de la forma

$$B = \bigcup_{k=0}^{k_r} B_k$$

Por otra parte los sucesos B_k , ($0 \leq k \leq N$), son disjuntos (si se verifica B_k es que se rechazan exactamente k productos y se aceptan $N - k$, luego no puede verificarse ningún otro B_j con $j \neq k$). Por ello:

$$P(B) = P\left(\bigcup_{k=0}^{k_r} B_k\right) = \sum_{k=0}^{k_r} P(B_k) = \sum_{k=0}^{k_r} \binom{N}{k} p_r^k q_r^{N-k}$$

donde la última igualdad procede de la expresión (1.49).

b) La situación que se plantea ahora consiste en tener, por una parte, $(s - 1) + k$ productos, de los cuales se deben aceptar $s - 1$ y rechazar k y, a continuación, aceptar el siguiente producto. Con ello se habrán rechazado k productos antes de aceptar s . Definiendo el suceso \bar{A}_{s+k} como “el producto es aceptado en la inspección $s + k$ -ésima” y denominando C al suceso cuya probabilidad se pide en este apartado, entonces podemos escribir (téngase en mente que ahora $N = s + k - 1$ para usar la terminología de la sección 1.7.1):

$$\begin{aligned} C &= B_k \cap \bar{A}_{s+k} \\ P(C) &= P(B_k \cap \bar{A}_{s+k}) = P(B_k) P(\bar{A}_{s+k}) \\ &= \left(\binom{s+k-1}{k} p_r^k q_r^{(s+k-1)-k} \right) q_r = \binom{s+k-1}{k} p_r^k q_r^s \end{aligned}$$



Aproximaciones de la probabilidad $P(B_k)$

La expresión de la probabilidad deducida en la ecuación (1.49) se puede aproximar por expresiones alternativas, las cuales, en algunos casos, pueden suponer una mayor comodidad en el cálculo de las probabilidades asociadas a composiciones de ensayos de Bernoulli. Dos de estas aproximaciones son las siguientes:

1. Aproximación de Poisson: esta primera aproximación es de validez cuando $N \gg 1$, $p \ll 1$ y el producto $Np = a < 5$. En estas condiciones se verifica que

$$P(B_k) = \binom{N}{k} p^k q^{N-k} \approx \frac{a^k}{k!} e^{-a} \quad (1.50)$$

Nótese que esta aproximación hace que sólo tengamos que calcular un factorial (en vez de los tres que corresponden al número combinatorio de la ecuación (1.49)). Asimismo, cuando las probabilidades son pequeñas, el elevar un número pequeño (p ó q) a una potencia grande puede traer consigo una pérdida de precisión en la operación. Con esta aproximación tales operaciones son asimismo obviadas.

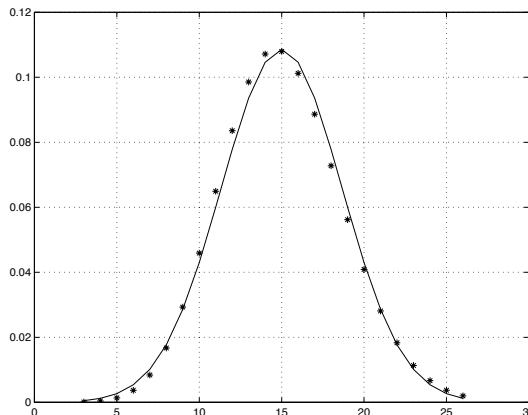


Figura 1.5: Ajuste de $P(B_k)$ mediante una curva dada por la expresión (1.51).

2. Aproximación de DeMoivre-Laplace: La figura 1.5 muestra en asteriscos el aspecto que presenta la expresión de la ecuación (1.49) para

$p = 0.1$ y $N = 150$. Asimismo, hemos superpuesto una aproximación de tales valores mediante la expresión

$$\frac{1}{b\sqrt{2\pi}}e^{\frac{(x-a)^2}{2b^2}} \quad (1.51)$$

donde el parámetro $a = Np$ es el parámetro de simetría de la curva anterior y el parámetro $b = \sqrt{Npq}$ es tal que cuanto mayor sea este valor más ancho es el lóbulo de la curva. Como puede verse, las probabilidades $P(B_k)$ se ajustan de manera bastante aproximada a la curva anterior para valores de k próximos al centro de simetría de la curva. De hecho, esta aproximación es válida cuando $Npq \gg 1$ y para valores de k en el intervalo $Np - 3\sqrt{Npq} \leq k \leq Np + 3\sqrt{Npq}$.

Cuando el objetivo sea calcular la probabilidad de la unión de varios B_k (es el caso del apartado a) del ejercicio anterior) debemos acudir al área bajo esta curva. Concretamente:

$$P\left(\bigcup_{k=0}^{k_0} B_k\right) \cong \int_{-\infty}^{k_0} \frac{1}{b\sqrt{2\pi}}e^{\frac{(x-a)^2}{2b^2}} dx = G\left(\frac{k_0 - Np}{\sqrt{Npq}}\right) \quad (1.52)$$

donde la función $G(\cdot)$ representa el área bajo la curva (1.51), y se encuentra tabulada (ella, u otras similares, véase pie de la página 62) en cualquier texto de Teoría de la Probabilidad (por ejemplo, [1, 2]).

En el caso en que el sumatorio no comience en cero, la expresión anterior debe transformarse de la manera que se indica:

$$P\left(\bigcup_{k=k_0}^{k_1} B_k\right) \cong G\left(\frac{k_1 - Np}{\sqrt{Npq}}\right) - G\left(\frac{(k_0 - 1) - Np}{\sqrt{Npq}}\right) \quad (1.53)$$

El motivo por el que esto es así, si bien en este momento no es evidente, lo será en capítulos posteriores (véase sección 4.5.2).

Capítulo 2

Variable aleatoria unidimensional

2.1 Concepto de variable aleatoria

En numerosos experimentos prácticos en que existan fenómenos aleatorios no disponemos de acceso directo a los resultados del experimento aleatorio, sino que, simplemente, observamos números que fluctúan. Por ejemplo, si medimos la tensión de salida de una antena—sobre una determinada carga—en diferentes situaciones, observaremos niveles de tensión fluctuantes. Tales fluctuaciones dependerán de numerosos factores, muchos de ellos completamente incontrolables por nosotros (condiciones atmosféricas, ambiente radioeléctrico local etc ...). En tales situaciones es necesario disponer de una herramienta que permita caracterizar probabilísticamente las mediciones que pueden obtenerse, de forma que, por ejemplo, podamos obtener un valor en torno al cual fluctuarán las medidas, un indicativo de la mayor o menor fluctuación esperable, intervalos de valores de mayor o menor probabilidad de aparición etc ... Esta herramienta está constituida por las variables aleatorias.

Visto el concepto intuitivo, tratemos de formalizar el concepto de variable aleatoria (VA, en adelante). Una VA se puede definir como una aplicación del espacio muestral en el cuerpo de los números complejos de orden N , que asocia a cada elemento del espacio muestral $a \in S$ (por ende, para cada posible resultado del experimento aleatorio) un vector de N números complejos $\bar{\mathbf{Z}} = [\mathbf{Z}_1 \dots \mathbf{Z}_N]$

$$\bar{\mathbf{Z}} : S \rightarrow \mathcal{C}^N$$

$$a \in S \rightarrow \bar{\mathbf{Z}}(a) = [\mathbf{Z}_1(a) \dots \mathbf{Z}_N(a)] \quad (2.1)$$

Un posible ejemplo de tal variable (compleja de orden N) sería la medición simultánea de N impedancias complejas de un circuito con el objetivo de analizar el comportamiento de las mismas con el paso del tiempo. Los elementos $a \in S$ serían los resultados de un experimento aleatorio de características desconocidas (envejecimiento de las impedancias) y lo único que se observaría del experimento sería el valor de las VAs (impedancias). Otro ejemplo lo constituiría la medición de la tensión a la salida de una antena. Imaginemos que se envía repetidamente una señal patrón; en tal caso, la medida de tensión a la salida de la antena debería reflejar fielmente esa repetición y así lo hará, pero con fluctuaciones superpuestas respecto del valor que se espera. En tal caso no sabríamos cuál es el resultado del experimento aleatorio que se está produciendo (de hecho, no sabemos qué experimento aleatorio es) sino que simplemente tendríamos acceso a las mediciones de tensión, las cuales son números reales.

Para poder analizar las cosas en un orden natural, consideremos en este tema el caso más sencillo, léase, el caso en que $N = 1$, y que el cuerpo sobre el que se define la aplicación sea el cuerpo de los números reales \mathcal{R} en vez del cuerpo de los números complejos \mathcal{C} . El resultado de esta simplificación será una VA unidimensional (real), a la cual nos referiremos con letras mayúsculas del tipo $(\mathbf{X}, \mathbf{Y}, \mathbf{Z} \dots)$. En sentido estricto las VAs deberían reflejar la dependencia con respecto al resultado del experimento aleatorio $a \in S$, es decir, deberíamos escribir $\mathbf{X}(a)$. Sin embargo, tal dependencia será dada por supuesto para no sobrecargar la notación.

Para terminar la formalización del concepto de VA debemos indicar que para que ésta sea tal, la aplicación anterior debe cumplir dos condiciones. Concretamente¹:

- $\{\mathbf{X} \leq x\} \in \mathcal{F}$, es decir, debe ser un suceso, de forma que sobre él puedan definirse probabilidades.
- $P(\mathbf{X} = -\infty) = P(\mathbf{X} = \infty) = 0$, de forma que aunque las variables puedan tomar valores arbitrariamente elevados, la probabilidad en el límite será nula.

Es interesante hacer una clasificación de las variables aleatorias en base al conjunto de valores que éstas pueden tomar. Esta clasificación será útil

¹En lo que sigue, y acorde con numerosos tratados [1, 2, 3] igualaremos los valores de las VAs a $\pm\infty$ y evaluaremos las funciones que definen su comportamiento probabilístico en esos mismos valores. Tal igualdad debe entenderse, naturalmente, como comportamientos en el límite.

durante el resto de la asignatura. Concretamente, si la VA puede tomar valores dentro de un conjunto discreto, esto es, dentro de un conjunto —a lo sumo— numerable, diremos que la VA es discreta. Un ejemplo típico es una variable que sólo pueda tomar los valores 0 y 1, o los valores $[a, b, c]$, o cualquier valor entero etc... Repárese que no se está exigiendo que el conjunto de valores que la VA puede tomar sea *finito* sino —a lo sumo— *numerable*, es decir, que se pueda hacer una correspondencia uno a uno entre el conjunto de valores que puede tomar la VA y el conjunto de los números enteros. Por contra, una variable *continua* tomará valores dentro de un conjunto no numerable, por ejemplo, cualquier valor del intervalo $[-1, 1]$ o, incluso cualquier valor real. Finalmente, hablaremos de variables mixtas, que son las que presentan un comportamiento intermedio entre el comportamiento de las continuas y el de las discretas. Para hablar más en detalle de estas últimas necesitamos, no obstante, introducir conceptos de caracterización probabilística de las variables, lo cual es nuestro siguiente paso.

2.2 Caracterización de las variables aleatorias: funciones de distribución y densidad

Caracterizar una VA consiste, simplemente, en asociar un comportamiento probabilístico a los valores que puede tomar una VA. De nada nos valdría decir que una variable puede tomar cualquier valor entre $-\infty$ y $+\infty$. Si así fuese, no estaríamos haciendo más que expresar nuestra ignorancia. Otra cosa sería decir que una variable puede tomar valores en ese rango, pero que toma con mucha mayor probabilidad los valores cercanos al cero, que los alejados de éste. Este tipo de información es, precisamente, extraíble de las funciones que caracterizan a las VAs.

La caracterización de las VAs puede verse (por comodidad) de forma distinta para variables continuas y para discretas. Decimos por comodidad pues, como veremos posteriormente, existen herramientas matemáticas que posibilitan que la caracterización de ambos tipos de variables se haga de una forma única. No obstante, como indicamos a continuación, es notablemente más sencillo considerarlas como casos distintos.

Empecemos por el caso de las variables discretas: una variable discreta \mathbf{X} queda completamente caracterizada si conocemos los valores que puede tomar (digamos, si conocemos los valores x_i , $i \in \mathcal{Z}$) y las probabilidades $p_i = P(\mathbf{X} = x_i)$, $i \in \mathcal{Z}$, con las que la variable toma cada uno de tales valores. Conociendo esto conocemos todo lo conocible con respecto a esta

variable \mathbf{X} . Pongamos un ejemplo para ver que es así: imaginemos que queremos conocer la probabilidad de que una variable \mathbf{X} , de la cual nos informan que puede tomar cualquier valor entero $i \in \mathcal{Z}$, tome valores entre -2 y 2 . Para poder hacer este cálculo, supongamos conocidos los p_i , con $i \in \mathcal{Z}$. La probabilidad deseada puede calcularse si conseguimos hacer uso del álgebra de sucesos que conocemos hasta ahora. En efecto

$$\begin{aligned}
 P(-2 \leq \mathbf{X} \leq 2) &= P(\{\mathbf{X} = -2\} \cup \{\mathbf{X} = -1\} \cup \{\mathbf{X} = 0\} \cup \\
 &\quad \{\mathbf{X} = 1\} \cup \{\mathbf{X} = 2\}) \\
 &= P(\mathbf{X} = -2) + P(\mathbf{X} = -1) + P(\mathbf{X} = 0) + \\
 &\quad P(\mathbf{X} = 1) + P(\mathbf{X} = 2) \\
 &= p_{-2} + p_{-1} + p_0 + p_1 + p_2 \\
 &= \sum_{i=-2}^2 p_i
 \end{aligned} \tag{2.2}$$

Repárese que el paso de la primera línea a la segunda se ha hecho en base a la extensión a N sucesos del tercer axioma de la probabilidad ya que, simultáneamente, la variable no puede tomar dos valores distintos, de forma que todos los sucesos involucrados son disjuntos.

El caso de las VAs continuas es un poco más complejo. El motivo no es otro que, para tales VAs, debemos proporcionar una función, también de variable continua, que caracterice probabilísticamente a todos y cada uno de los valores de éstas. Tales funciones son fundamentalmente dos, a saber, la función de distribución y la función de densidad de probabilidad. Empecemos por la primera, para pasar a continuación a describir la segunda.

2.2.1 Función de distribución

La función de distribución de una VA \mathbf{X} se denota por $F_{\mathbf{X}}(x)$ y se define mediante

$$F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x) \tag{2.3}$$

Repárese que \mathbf{X} representa la VA y $x \in \mathcal{R}$ es la abscisa donde se evalúa la función $F_{\mathbf{X}}(x)$. Esta función, por tanto, es una función real de variable real y , como puede verse, esta función indica cuál es la probabilidad de que la VA \mathbf{X} tome valores menores o iguales que un determinado punto de abscisa x . Repárese por ello que es imprescindible que $\{\mathbf{X} \leq x\} \in \mathcal{F}$, como antes hemos exigido, de forma que tenga sentido definir probabilidades sobre esta entidad. Pues bien, esta función nos permite hacer cualquier

cálculo probabilístico en relación con los valores que puede tomar la VA \mathbf{X} . Asimismo, esta función, dado que mide probabilidades, no puede ser una función arbitraria, sino que debe cumplir un conjunto de condiciones. Veamos esto paso por paso (véase figura 2.1 y recuérdese la nota en el pie de la página 48):

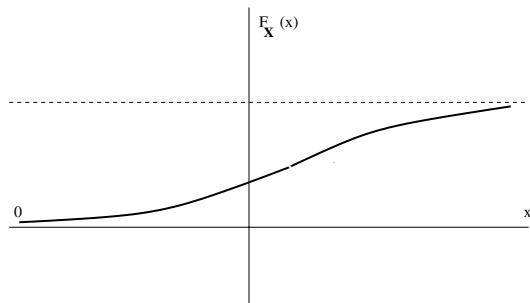


Figura 2.1: Un ejemplo de función de distribución.

- $F_{\mathbf{X}}(-\infty) = 0$, esto es, esta función presenta una asíntota horizontal de cota nula. En efecto, hemos dicho que para que una VA sea tal, $P(\mathbf{X} = -\infty) = 0$. Esto trae consigo que

$$F_{\mathbf{X}}(-\infty) = P(\mathbf{X} \leq -\infty) = P(\mathbf{X} = -\infty) = 0$$

- $F_{\mathbf{X}}(\infty) = 1$, esto es, esta función presenta una segunda asíntota horizontal, en este caso, de cota unidad. En efecto, el suceso $\{\mathbf{X} \leq \infty\} = S$, es decir, es el suceso seguro.
- $0 \leq F_{\mathbf{X}}(x) \leq 1$, lo cual es consecuencia inmediata de los axiomas de la probabilidad (repárese que los valores de una función de distribución son probabilidades).
- La función es no decreciente, es decir, si $x_1 < x_2$ entonces se verifica que $F_{\mathbf{X}}(x_1) \leq F_{\mathbf{X}}(x_2)$. Esto se debe a que estamos trabajando con sucesos, y, como es obvio, se verifica que $\{\mathbf{X} \leq x_1\} \subset \{\mathbf{X} \leq x_2\}$ si $x_1 \leq x_2$. Por ello, también es cierto que $P(\mathbf{X} \leq x_1) \leq P(\mathbf{X} \leq x_2)$.

Por esta razón a esta función de le denomina también función de probabilidad acumulada, ya que, conforme crece la abscisa donde evaluamos la función, crece el intervalo de la recta real considerado en el cálculo de la probabilidad, y crece, de forma acorde, la probabilidad asociada.

- Lo anterior nos permite realizar un cálculo probabilístico con la función de distribución. Pensemos que necesitamos conocer $P(x_1 < \mathbf{X} \leq x_2)$. Esto se puede realizar de manera sencilla haciendo uso del álgebra de sucesos que conocemos hasta ahora. Concretamente,

$$\begin{aligned}\{\mathbf{X} \leq x_2\} &= \{\mathbf{X} \leq x_1\} \cup \{x_1 < \mathbf{X} \leq x_2\} \rightarrow \\ F_{\mathbf{X}}(x_2) = P(\mathbf{X} \leq x_2) &= P((\mathbf{X} \leq x_1) \cup (x_1 < \mathbf{X} \leq x_2)) = \\ &= P(\mathbf{X} \leq x_1) + P(x_1 < \mathbf{X} \leq x_2) \\ &= F_{\mathbf{X}}(x_1) + P(x_1 < \mathbf{X} \leq x_2)\end{aligned}\quad (2.4)$$

Despejando de arriba es inmediato ver que

$$P(x_1 < \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1), \quad (2.5)$$

es decir, la probabilidad de que la VA tome valores en un intervalo es igual a la diferencia de los valores de la función de distribución en dicho intervalo.

- Podemos calcular la probabilidad del suceso complementario $\{\mathbf{X} > x\}$. En efecto, podemos escribir el suceso seguro como

$$S = \{\mathbf{X} \leq x\} \cup \{\mathbf{X} > x\}$$

Tomando probabilidades y aplicando el hecho de que los sucesos del segundo miembro de la igualdad anterior son disjuntos, concluimos que $P(\mathbf{X} > x) = 1 - F_{\mathbf{X}}(x)$.

Caso de las variables discretas

Al principio de esta sección hemos hablado de la función de distribución de una VA, sin especificar si ésta tenía que ser continua o discreta. Y, en efecto, es así: la función de distribución se define para VAs con independencia del tipo de variables que se esté manejando. No obstante, sólo es informativa en el caso de variables continuas, pues, para el caso de variables discretas, como hemos dicho antes, basta conocer los valores que puede tomar la variable, y las probabilidades con las que toma dichos valores.

Veamos que esto efectivamente es así mediante un ejemplo. Consideremos una VA discreta \mathbf{X} que pueda tomar los valores (a, b, c) con probabilidades respectivas (p_a, p_b, p_c) . Veremos que con estos datos podemos construir la función de distribución de la variable \mathbf{X} y así queda patente que la función de distribución no añade nada que no sepamos a partir de los

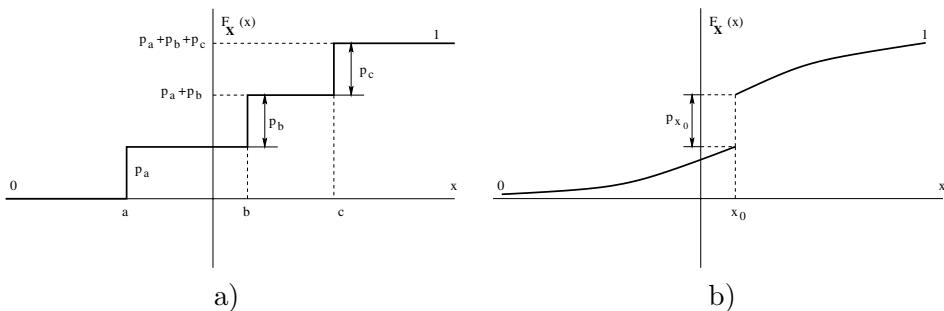


Figura 2.2: a) $F_{\mathbf{X}}(x)$ para una VA discreta que toma tres valores. b) $F_{\mathbf{X}}(x)$ para una VA mixta que toma el valor x_0 con probabilidad p_{x_0} .

datos iniciales. Construir la función de distribución $F_{\mathbf{X}}(x)$ no es otra cosa que calcular la probabilidad del suceso involucrado para todos los valores de x (repárese que x es la variable independiente de la función $F_{\mathbf{X}}(x)$ de forma que nada tiene que ver con los valores a, b y c que toma la variable \mathbf{X}). Para tal fin, distingamos diversos intervalos en la recta real en función de los valores (a, b, c). Concretamente:

1. $x < a$; $F_{\mathbf{X}}(x) = 0$, dado que \mathbf{X} no puede tomar ningún valor en esta zona de la recta real.
2. $a \leq x < b$; $F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x) = P(\mathbf{X} = a) = p_a$, pues para valores de x en el intervalo indicado, la variable aleatoria \mathbf{X} sólo puede ser menor o igual que el valor a , lo cual sucede con probabilidad p_a .
3. $b \leq x < c$; $F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x) = P((\mathbf{X} = a) \cup (\mathbf{X} = b)) = p_a + p_b$
4. $x \geq c$; $F_{\mathbf{X}}(x) = P((\mathbf{X} = a) \cup (\mathbf{X} = b) \cup (\mathbf{X} = c)) = 1$, dado que es el suceso seguro (repare que se está calculando la probabilidad de que la VA \mathbf{X} sea menor igual que un valor de x que es, a su vez, mayor que c . Dado que el valor máximo que puede tomar \mathbf{X} es c , el suceso cuya probabilidad buscamos es el suceso seguro, de forma que su probabilidad es la unidad).

La forma de la función de distribución puede verse en la figura 2.2(a). Como se aprecia de la figura, la función es constante a intervalos, con comportamiento asintótico por la izquierda igual a cero y por la derecha igual a uno. Los puntos de discontinuidad de la función se encuentran en los puntos donde la variable discreta toma valores, y la altura de cada discontinuidad coincide con la probabilidad de que la variable \mathbf{X} tome el

valor de que se trate. Por ello, la expresión de la función de distribución se puede escribir en base a los datos del ejemplo, de la forma

$$F_{\mathbf{X}}(x) = p_a u(x - a) + p_b u(x - b) + p_c u(x - c)$$

con $u(x)$ la función escalón. Generalizando a una variable que tome los valores x_i , $i = \{1, \dots, N\}$, con probabilidades respectivas p_i , podemos escribir

$$F_{\mathbf{X}}(x) = \sum_{i=1}^N p_i u(x - x_i) \quad (2.6)$$

Comentario adicional: Observe que, para que el comportamiento asintótico por la derecha de esta función sea el esperado, es condición necesaria que $\sum_i p_i = 1$.

Variables mixtas

Las funciones de distribución correspondientes a VAs continuas son funciones suaves, continuas, y derivables —con continuidad— salvo en una cantidad, a lo sumo, numerable de puntos. No obstante, las funciones de distribución correspondientes a variables discretas hemos visto que consisten en superposiciones de funciones escalón desplazadas a los puntos donde la VA toma valores. El tercer tipo de variable aleatoria, del cual hasta ahora sólo hemos dado el nombre, es un caso intermedio entre ambas. Tales variables —las variables mixtas— presentan funciones de distribución suaves y continuas pero presentan al menos un punto donde la función de distribución es discontinua. Un ejemplo de función de distribución correspondiente a una variable mixta puede verse en la figura 2.2(b)

La función de distribución en discontinuidades

Las discontinuidades en las funciones de distribución sólo aparecen en las variables discretas o mixtas. En tal caso, el manejo de probabilidades es más sencillo de realizar en términos de álgebra de sucesos que en términos de la función de distribución. No obstante, por completitud, maticemos el comportamiento de esta función, lo cual reviste sólo interés en el caso de discontinuidades²:

²En lo que sigue interpretaremos $F_{\mathbf{X}}(x^+) = \lim_{\epsilon \rightarrow 0} F_{\mathbf{X}}(x + \epsilon)$ y $F_{\mathbf{X}}(x^-) = \lim_{\epsilon \rightarrow 0} F_{\mathbf{X}}(x - \epsilon)$, con $\epsilon \in \mathcal{R}^+$.

- $F_{\mathbf{X}}(x) = F_{\mathbf{X}}(x^+)$, es decir, el valor que la función de distribución toma en un punto es el que esta función toma inmediatamente a la derecha de ese punto.
- $P(\mathbf{X} = x) = F_{\mathbf{X}}(x^+) - F_{\mathbf{X}}(x^-) = F_{\mathbf{X}}(x) - F_{\mathbf{X}}(x^-)$, según hemos visto en el ejemplo. Repárese que esto indica que la probabilidad de que una variable continua tome exactamente el valor x es nula. Para variables continuas la probabilidad debe considerarse en términos de intervalos, aunque éstos sean, como veremos a continuación, de longitud infinitesimal.
- $P(x_1 \leq \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1^-)$, dado que si se incluye el valor x_1 en el suceso cuya probabilidad se desea calcular tenemos que evaluar la función de distribución a la izquierda de dicho punto. Esta propiedad se extiende a los otros dos casos que nos faltan de la forma

$$\begin{aligned} P(x_1 < \mathbf{X} < x_2) &= F_{\mathbf{X}}(x_2^-) - F_{\mathbf{X}}(x_1) \\ P(x_1 \leq \mathbf{X} < x_2) &= F_{\mathbf{X}}(x_2^-) - F_{\mathbf{X}}(x_1^-) \end{aligned}$$

De nuevo es interesante apuntar que para variables continuas estas propiedades no revisten ningún interés, ya que la función de distribución es continua, de forma que los comportamientos por la izquierda y por la derecha de cada punto que se considere coinciden.

2.2.2 Función de densidad de probabilidad

La función de densidad de probabilidad de una VA \mathbf{X} se denota por $f_{\mathbf{X}}(x)$ y se define a partir de la función de distribución mediante

$$f_{\mathbf{X}}(x) = \frac{dF_{\mathbf{X}}(x)}{dx} \quad (2.7)$$

de forma que también es una función real de variable real. Esta función tiene exactamente la misma información que la función de distribución. De hecho, si, como hemos visto, de la primera podemos pasar a la segunda, podemos hacer el paso inverso, esto es, de la segunda pasar a la primera. Basta resolver la ecuación diferencial resultante de la definición integral anterior:

$$\begin{aligned} dF_{\mathbf{X}}(x) &= f_{\mathbf{X}}(x)dx \\ \int_{-\infty}^x dF_{\mathbf{X}}(\tau) &= \int_{-\infty}^x f_{\mathbf{X}}(\tau)d\tau \end{aligned}$$

$$\begin{aligned} F_{\mathbf{X}}(x) - F_{\mathbf{X}}(-\infty) &= \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau \\ F_{\mathbf{X}}(x) &= \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau \end{aligned} \quad (2.8)$$

donde la última igualdad procede del comportamiento asintótico por el lazo izquierdo de la función de distribución.

Dado que la función de distribución tiene un comportamiento marcado por el conjunto de características analizadas en la sección 2.2.1, la función de densidad de probabilidad no puede tener formas arbitrarias, sino que debe ser coherente con las propiedades de la función de distribución. En particular:

- $f_{\mathbf{X}}(x) \geq 0$, pues la función de distribución es una función no decreciente.
- El área bajo una función de densidad de probabilidad es siempre unitaria. Esto procede de la expresión (2.8), si evaluamos en $x = \infty$ (recuérdese que la función de distribución en dicho punto vale uno).
- $P(x_1 < \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1) = \int_{x_1}^{x_2} f_{\mathbf{X}}(x) dx$, lo cual puede obtenerse mediante el mismo procedimiento empleado para obtener la expresión (2.8), integrando entre x_1 y x_2 .

Por lo tanto, repárese que si bien la función de distribución es igual a una probabilidad, la función de densidad es una función que, integrada en un cierto intervalo, proporciona el valor de que la VA tome valores dentro de dicho intervalo. De aquí podemos deducir el motivo por el que dicha función recibe el nombre *densidad de probabilidad*. Concretamente, si consideramos el intervalo $(x, x + \Delta x)$, la probabilidad de que la variable tome valores en dicho intervalo es, según conocemos

$$P(x < \mathbf{X} \leq x + \Delta x) = F_{\mathbf{X}}(x + \Delta x) - F_{\mathbf{X}}(x)$$

Dado que la función de densidad es una derivada, podemos plantearla como el límite de un cociente incremental

$$f_{\mathbf{X}}(x) = \lim_{\Delta x \rightarrow 0} \frac{F_{\mathbf{X}}(x + \Delta x) - F_{\mathbf{X}}(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x < \mathbf{X} \leq x + \Delta x)}{\Delta x}$$

lo cual nos permite ver que la función de densidad mide la probabilidad de un intervalo infinitesimal alrededor del punto donde ésta se evalúa, dividido por la longitud de dicho intervalo. Es pues un probabilidad por unidad de longitud, de forma que es, en efecto, una densidad de probabilidad.

VARIABLES DISCRETAS Y MIXTAS

Según hemos visto, la definición de la función de densidad se lleva a cabo mediante una derivada de la función de distribución. Cuando la variable bajo estudio sea continua, la función de distribución lo es también, y será en general derivable salvo, tal vez, en puntos aislados. En esos puntos, la función de densidad es discontinua. Tal es el caso, por ejemplo, de las variables uniformes que veremos a continuación (sección 2.3.1).

Sin embargo, cuando la función de distribución presenta el aspecto dado por la ecuación (2.6), las cosas se complican ya que la función de distribución es discontinua en varios puntos (tantos como valores tome la variable discreta en cuestión). Para tales casos debemos acudir a las funciones *Delta de Dirac*, pues se verifica³

$$\delta(x) = \frac{du(x)}{dx} \quad (2.9)$$

Esta función es nula para todo punto de abscisa distinto de cero y, en ese punto, presenta una singularidad (su valor es infinito) de forma tal que el área encerrada bajo esta función es unitaria, es decir, $\int_{-\infty}^{\infty} \delta(x)dx = 1$.

Haciendo uso de esta función, así como de la linealidad del operador derivación, es obvio que la función de densidad de una variable discreta (y de una mixta en los puntos de discontinuidad de la función de distribución) puede escribirse

$$f_{\mathbf{X}}(x) = \sum_{i=1}^N p_i \delta(x - x_i) \quad (2.10)$$

es decir, consiste en un tren de deltas, colocadas éstas en cada uno de los puntos donde la variable toma valores y con peso igual a la probabilidad con que la variable toma el valor en cuestión.

Recuerde, no obstante, que raramente se acudirá al empleo de la función (2.10) pues todo cálculo probabilístico se puede hacer en términos de las probabilidades p_i . Tan sólo cuando se necesite explícitamente calcular la expresión de la función de densidad de tales variables será necesario acudir a este tipo de formulación.

³La expresión (2.9) debe tomarse con las oportunas reservas. La función escalón no es derivable en el cero, de forma que la función $\delta(x)$ no se podría definir de esta manera, sino a través del límite de la derivada de una función continua por intervalos, cuyo valor en el intervalo $(0, \Delta)$ sería $\frac{1}{\Delta}$. El límite se calcularía con Δ tendiendo a cero, esto es, la pendiente tendiendo a infinito con $x \rightarrow 0$. Los detalles puede consultarlos en [15]. En cualquier caso, lo que sí es estrictamente cierto es que $u(x) = \int_{-\infty}^x \delta(t)dt$ de forma que el paso de (2.10) a (2.6) a través de una integración es claro.

2.3 Algunas variables aleatorias frecuentes

En este apartado describiremos algunos tipos de VAs que aparecen con frecuencia en la práctica. Para el caso de las variables continuas plantearemos la función de densidad, obtendremos la función de distribución y comentaremos aspectos característicos de las variables. Para el caso de las discretas indicaremos, como ya hemos visto que es necesario, qué valores pueden tomar las diferentes variables y con qué probabilidad toman dichos valores.

Existen otros muchos tipos de VAs que los que aquí expondremos, pero el manejo de las funciones de densidad y distribución será similar metodológicamente al que aquí hagamos, si bien, como es natural, los cálculos analíticos precisos dependerán de cada caso.

2.3.1 Continuas

Uniforme

Una VA \mathbf{X} es una variable uniforme en el intervalo definido por los puntos⁴ a y b si su función de densidad es constante (y no nula) en el interior de dicho intervalo, y es nula fuera del mismo. Denotaremos tal tipo de variable de la forma $\mathbf{X} \sim U(a, b)$.

El valor de la constante asociada a esta función de densidad no puede ser arbitrario dado que, como hemos visto, el área bajo cualquier función de densidad es siempre unitaria. Aplicando esta restricción tendremos

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\mathbf{X}}(x)dx &= \int_a^b f_{\mathbf{X}}(x)dx \\ &= \int_a^b kdx \\ &= k \int_a^b dx = k(b-a) = 1 \end{aligned} \tag{2.11}$$

Por tanto los parámetros a y b determinan el valor de la constante k y éste resulta ser igual al inverso de la longitud del intervalo, es decir, igual a $1/(b-a)$.

Con respecto a la función de distribución tendremos que calcular

$$F_{\mathbf{X}}(x) = \int_{-\infty}^x f_{\mathbf{X}}(\tau)d\tau$$

⁴Todos los parámetros involucrados en las funciones de densidad y distribución serán parámetros reales, dado que estas funciones lo son.

integral que, en función de donde se encuentre x , tomará una forma u otra. Así pues, distinguimos

$$\begin{aligned} x < a \quad F_{\mathbf{X}}(x) &= \int_{-\infty}^x 0 d\tau = 0 \\ a \leq x < b \quad F_{\mathbf{X}}(x) &= \int_a^x \frac{1}{b-a} d\tau = \frac{x-a}{b-a} \\ x \geq b \quad F_{\mathbf{X}}(x) &= \int_a^x f_{\mathbf{X}}(\tau) d\tau = \int_a^b \frac{1}{b-a} d\tau = 1 \end{aligned} \tag{2.12}$$

Puede observarse, por tanto, que $F_{\mathbf{X}}(x)$ tiene los comportamientos asintóticos esperados (cero, por lado izquierdo, y uno por lado derecho) y que crece linealmente en el interior del intervalo (a, b) . Asimismo, dado que \mathbf{X} es una VA continua, $F_{\mathbf{X}}(x)$ es una función continua.

Es interesante realizar unos comentarios adicionales:

- La variable uniforme sólo puede tomar valores en el interior del intervalo (a, b) . En efecto, si calculamos la probabilidad

$$P(x_1 < \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1) = \int_{x_1}^{x_2} f_{\mathbf{X}}(x) dx \tag{2.13}$$

es fácil ver que esta integral sólo es distinta de cero si los intervalos (x_1, x_2) y (a, b) tienen intersección no vacía, dado que la función subintegral es no nula en el interior del segundo intervalo.

- Lo anterior se generaliza a cualquier tipo de variable: aquellos valores de abscisa x para los que la función de densidad presenta un valor nulo, son valores que la VA \mathbf{X} no puede tomar.
- Retomando la ecuación (2.13) es directo comprobar que un intervalo tiene probabilidad nula cuando los valores de la función de distribución en los extremos del intervalo coinciden. Por lo tanto, intervalos de constancia en la función de distribución son intervalos de probabilidad nula para la VA. Este resultado, que es general, se concreta en el caso de una VA $\mathbf{X} \sim U(a, b)$ viendo que la función de distribución para valores $x < a$ y $x > b$ es constante. Como hemos dicho, en estos intervalos la VA no puede tomar valores.
- Centrémonos ahora en los valores de probabilidad. Para ello, consideremos cuatro puntos de abscisa x_i ($i=\{1, \dots, 4\}$), tales que cumplan

$a \leq x_1 < x_2 < x_3 < x_4 \leq b$, y, asimismo, que se verifique que $\Delta = x_2 - x_1 = x_4 - x_3$, esto es, que los intervalos (x_1, x_2) y (x_3, x_4) tengan la misma longitud. Calculemos las probabilidades respectivas de que la VA \mathbf{X} tome valores en cada uno de los intervalos

$$\begin{aligned} P(x_1 < \mathbf{X} \leq x_2) &= \int_{x_1}^{x_2} f_{\mathbf{X}}(x) dx = \int_{x_1}^{x_2} \frac{1}{b-a} dx \\ &= \frac{x_2 - x_1}{b-a} = \frac{\Delta}{b-a} \\ P(x_3 < \mathbf{X} \leq x_4) &= \int_{x_3}^{x_4} f_{\mathbf{X}}(x) dx = \int_{x_3}^{x_4} \frac{1}{b-a} dx \\ &= \frac{x_4 - x_3}{b-a} = \frac{\Delta}{b-a} \end{aligned}$$

Como puede apreciarse, las dos probabilidades coinciden. Por tanto, ambos intervalos son equiprobables. Ésta es una característica general de las variables uniformes: en el interior del intervalo donde la variable toma valores no existe ningún intervalo privilegiado, sino que todos los intervalos son equiprobables. Si hacemos que el intervalo en cuestión sea de anchura infinitesimal podemos concluir que todos los valores que la variable uniforme puede tomar son equiprobables.

- Repárese que no se ha dicho de forma explícita si el intervalo (a,b) es abierto o cerrado. De hecho, a efectos de cálculo de probabilidades es irrelevante. Como hemos visto ya, la contribución a la probabilidad total de puntos aislados es nula, de forma que sólo toman sentido bajo el signo integral, esto es, dentro de intervalos no puntuales. Cuando se trabaje con variables discretas, sin embargo, puntos aislados sí son relevantes, pues éstos llevan asociadas probabilidades no nulas.

Gaussiana

Una VA \mathbf{X} se dice que es gaussiana, de parámetros a y b si su función de densidad tiene la expresión

$$f_{\mathbf{X}}(x) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} \quad (2.14)$$

donde a es cualquier número real, $b > 0$ y la expresión es válida $\forall x$. Denotaremos a este tipo de variable de la forma $\mathbf{X} \sim N(a, b)$. El motivo de emplear la letra N es que a este tipo de variables se les denomina también

variables normales puesto que es el tipo de variables normalmente asumido para el modelado y resolución de numerosos problemas prácticos (hay razones para ello, como veremos en la sección 4.5.1).

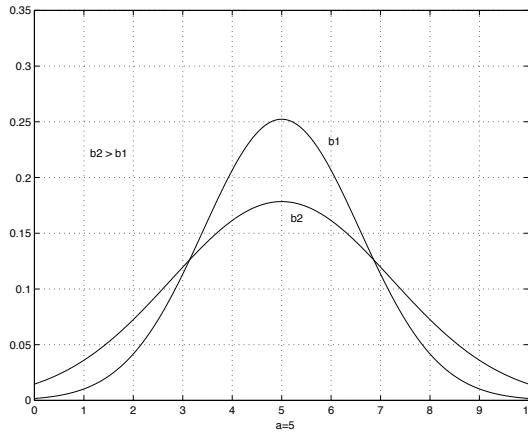


Figura 2.3: Ejemplos de función de densidad gaussiana para dos valores de b .

La figura 2.3 muestra la forma de la expresión (2.14) para dos valores distintos de b ; puede observarse que los parámetros a y b tienen un claro significado gráfico: el parámetro a es el punto de simetría de la curva (así como el punto de abscisa del máximo) y el parámetro b controla la anchura y altura de la curva. Repárese que cuanto más grande sea b más tarda en decrecer la exponencial cuando nos separamos del punto de abscisa $x = a$, de forma que más ancha es la curva. Si esto es así, dado que el área bajo esta función ha de ser unitaria, una *campana* más ancha implica una menor altura con el fin de que se cumpla este requisito.

Estos parámetros reciben nombre propios. Posponemos la presentación de tales nombres hasta la sección 2.5.4.

Respecto de la función de distribución, ésta se calcularía mediante la expresión

$$F_{\mathbf{X}}(x) = \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau = \int_{-\infty}^x \frac{1}{b\sqrt{2\pi}} e^{-\frac{(\tau-a)^2}{2b^2}} d\tau \quad (2.15)$$

Sin embargo, no existe primitiva conocida para esta integral, de forma que no existe una expresión cerrada para la función de distribución de una variable gaussiana.

Si no existe expresión para la función de distribución cabe preguntarse si es posible calcular probabilidades asociadas a este tipo de variables. La

respuesta, naturalmente, es afirmativa. Para tal fin, la integral anterior se calcula numéricamente, y existen tablas sobre los valores de esta función para numerosos valores de abscisa x . Algún comentario en relación con esto, no obstante:

- Repárese que esta función es biparamétrica, de forma que, aparentemente, se requeriría tabular la función para cada par (a, b) . Tales parámetros son números reales, de forma que existe una cantidad no numerable de combinaciones (a, b) ; debe existir, naturalmente, alguna manera de evitar semejante complicación.
- Aunque existen tablas publicadas de esta función para valores de abscisa tanto positivos como negativos [11], es más frecuente [1] que sólo se proporcione el valor de esta función en abscisa no negativas ($x \geq 0$).

Los dos problemas mencionados pueden resolverse oportunamente. Veamos cómo:

- En primer lugar, denominemos variable *normal estándar* a una VA $X \sim N(a = 0, b = 1)$. La función que típicamente se encuentra tabulada⁵

$$G(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau \quad (2.16)$$

Sin embargo, esta función puede amoldarse al caso de variables no estándar. Para ello, basta que hagamos el cambio de variable $\alpha = \frac{\tau-a}{b}$ en la expresión (2.15). Por tanto, para una VA $\mathbf{X} \sim N(a, b)$ tendremos

$$\begin{aligned} F_{\mathbf{X}}(x) &= \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau = \\ &= \int_{-\infty}^x \frac{1}{b\sqrt{2\pi}} e^{-\frac{(\tau-a)^2}{2b^2}} d\tau \\ &= \int_{-\infty}^{\frac{x-a}{b}} \frac{1}{b\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} b d\alpha \end{aligned}$$

⁵En [2] la función tabulada es la función $erf(x)$, la cual se define como $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\tau^2} d\tau$. En Teoría de la Comunicación [13, 14] es habitual hablar de la función $Q(x)$, la cual es el complemento a uno de $G(x)$, y de la función $erfc(x)$, la cual se define como $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\tau^2} d\tau$. Las funciones $erf(x)$ y $erfc(x)$ se relacionan fácilmente con las funciones $G(x)$ y $Q(x)$.

$$\begin{aligned}
 &= \int_{-\infty}^{\frac{x-a}{b}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} d\alpha \\
 &= G\left(\frac{x-a}{b}\right)
 \end{aligned} \tag{2.17}$$

Por tanto, basta con que conozcamos la función de distribución para una normal estándar (función $G(x)$, ecuación (2.16)) para poder encontrar la probabilidad de cualquier VA gaussiana, estándar o no.

- Centrémonos ahora en la función $G(x)$. Esta función es igual al área bajo la función de densidad de una VA normal estándar. Hemos visto que, para tal VA, el parámetro $a = 0$, de forma que la función de densidad es simétrica respecto a $x = 0$, es decir, es una función par. Por tanto, se ha de verificar que

$$\int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau$$

dado que la función subintegral es par, y los intervalos de integración son simétricos respecto a $x = 0$. Demos un significado probabilístico a estas integrales, lo cual nos va a permitir relacionarlas con la función $G(x)$. La integral de la izquierda no es otra cosa que $P(\mathbf{X} \leq -x) = F_{\mathbf{X}}(-x) = G(-x)$. La segunda integral, por su parte, es igual a $P(\mathbf{X} > x)$. Sabemos que $P(\mathbf{X} > x) = 1 - P(\mathbf{X} \leq x) = 1 - F_{\mathbf{X}}(x) = 1 - G(x)$. Como ambas cosas son iguales, concluimos que $G(-x) = 1 - G(x)$ y, por tanto, es claro que es suficiente conocer el área bajo la curva sólo en la parte derecha de la recta real.

Finalmente, debemos comentar una diferencia de importancia de esta variable, y en general, de todas las variables, con respecto a la uniforme: la no equiprobabilidad de intervalos de anchura constante de valores de la variable. En el caso de la VA uniforme cualquier intervalo de valores de anchura Δ (que se encuentre dentro del intervalo (a, b)) es equiprobable. Sin embargo, en el caso de las variables gaussianas, son mucho más probables los intervalos de la misma cercanos al punto de abscisa $x = a$ que los alejados de este punto. En efecto, según muestra la figura 2.4 y aceptando que la base de los dos rectángulos dibujados mide Δx , el área encerrada por el rectángulo situado en torno a $x = a$ es muy superior al área del segundo rectángulo. Como el área encerrada bajo la curva en el intervalo de que se trate es igual a la probabilidad de que la variable tome valores en dicho intervalo, es obvio qué valores de la variable gaussiana son más probables.

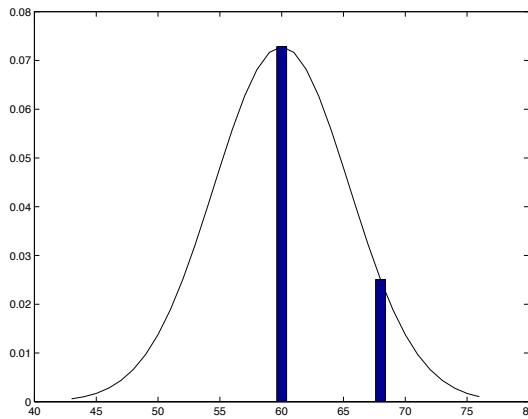


Figura 2.4: Comparación de probabilidades de valores en una gaussiana.

Exponencial

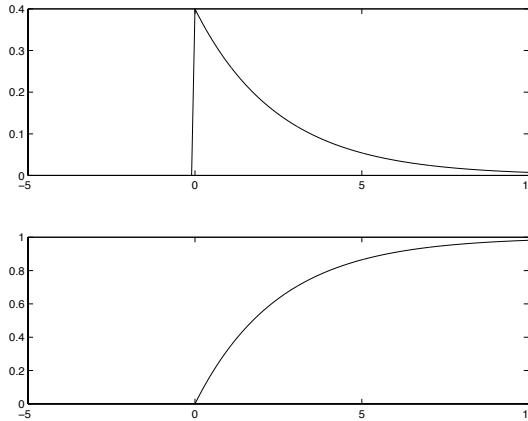


Figura 2.5: $f_{\mathbf{X}}(x)$ (superior) y $F_{\mathbf{X}}(x)$ (inferior) de una VA exponencial de parámetro $\lambda = 0.4$.

Una VA \mathbf{X} se dice que es exponencial de parámetro λ , y se denota por $\mathbf{X} \sim \exp(\lambda)$, si su función de densidad tiene la expresión

$$f_{\mathbf{X}}(x) = \lambda e^{-\lambda x} u(x) \quad (2.18)$$

donde $\lambda > 0$ y $u(x)$ es la función escalón. En base a ello puede observarse que una VA exponencial puede tomar sólo valores no negativos y que

la probabilidad de intervalos de anchura constante es tanto más pequeña cuanto más alejados se encuentren del cero (véase figura 2.5).

Asimismo:

- $f_{\mathbf{X}}(x) \geq 0 \quad \forall x$
- $\int_{-\infty}^{\infty} \lambda e^{-\lambda x} u(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = (-e^{-\lambda x})_0^{\infty} = 1.$

Finalmente, la función de distribución se obtiene fácilmente; por una parte es claro que $F_{\mathbf{X}}(x) = 0$ si $x < 0$ pues la función de densidad es nula en esta zona. Integrando en la zona en que la función de densidad es no nula, podemos escribir

$$\begin{aligned} F_{\mathbf{X}}(x) &= \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau \\ &= \int_0^x e^{-\lambda \tau} d\tau \\ &= \left[-e^{-\lambda \tau} \right]_0^x \\ &= \left[1 - e^{-\lambda x} \right], \quad x \geq 0 \end{aligned}$$

por lo que

$$F_{\mathbf{X}}(x) = \left[1 - e^{-\lambda x} \right] u(x) \quad (2.19)$$

2.3.2 Discretas

Bernoulli

Una VA \mathbf{X} discreta se dice que es de Bernoulli, de parámetro p ($p \geq 0$), si sólo puede tomar los valores $x = 0$ y $x = 1$ con probabilidades respectivas $P(\mathbf{X} = 0) = 1 - p = q$ y $P(\mathbf{X} = 1) = p$.

Tales variables pueden asociarse de manera inmediata a un ensayo de Bernoulli. Basta para ello llevar a cabo la igualdad

$$\begin{aligned} A &= \{X = 1\} \\ \overline{A} &= \{X = 0\} \end{aligned}$$

Binomial

Una VA \mathbf{X} discreta se dice que es Binomial, de parámetros N y p ($N \in \mathbb{Z}^+$ y $p \geq 0$), si puede tomar los valores $x = k$ con $k = \{0, 1, 2, \dots, N\}$, con

probabilidades

$$P(\mathbf{X} = k) = \binom{N}{k} p^k q^{N-k} \quad (2.20)$$

donde $q = 1 - p$. En tal caso la variable se denota $\mathbf{X} \sim B(N, k)$.

Puede observarse que esta variable se asocia fácilmente a la composición de ensayos (independientes) de Bernoulli, sin más que hacer la igualdad $B_k = \{\mathbf{X} = k\}$, con $k = \{0, \dots, N\}$ y B_k el suceso definido en la sección 1.7.1.

Poisson

Una VA \mathbf{X} discreta se dice que es de Poisson, de parámetro a ($a > 0$), si puede tomar los valores enteros $x = k$ con $k \geq 0$, con probabilidades

$$P(\mathbf{X} = k) = e^{-a} \frac{a^k}{k!} \quad (2.21)$$

Obsérvese que los valores que puede tomar esta variable no están acotados superiormente.

Geométrica

Una VA \mathbf{X} discreta se dice que es geométrica, de parámetro p ($p \geq 0$), si puede tomar los valores enteros $x = k$ con $k \geq 0$, con probabilidades

$$P(\mathbf{X} = k) = q^k p \quad (2.22)$$

donde $q = 1 - p$. Esta variable se emplea, por ejemplo, para calcular la probabilidad del número de **repeticiones** (independientes) de un ensayo de Bernoulli que se deben llevar a cabo hasta obtener el resultado A (supuesto que la probabilidad de obtener A sea igual a p). En efecto, si el número de repeticiones⁶ es k es porque en los $k - 1$ intentos anteriores se ha verificado el resultado \overline{A} y en el $k -$ ésimo experimento el resultado A . Denominando A_i al resultado del ensayo $i -$ ésimo, procedamos con k creciente

$$\begin{aligned} P(\mathbf{X} = 0) &= P(A_1 = A) = p \\ P(\mathbf{X} = 1) &= P((A_1 = \overline{A}) \cap (A_2 = A)) = qp \end{aligned}$$

⁶La palabra *repetición* está aquí empleada en el más puro sentido académico: si se repite un experimento (por ejemplo, cursar una asignatura) una vez es porque se ejecuta dos veces (el alumno se ha matriculado dos veces) y en la primera ejecución el resultado no ha sido el deseado (esto es, no se ha aprobado en primera matrícula). En este sentido repetir *cero* veces implica aprobar en primera matrícula.

$$\begin{aligned}
 P(\mathbf{X} = 2) &= P((A_1 = \bar{A}) \cap (A_2 = \bar{A}) \cap (A_3 = A)) = q^2 p \\
 &\vdots \quad \vdots \\
 P(\mathbf{X} = k) &= P\left(\bigcap_{i=1}^k (A_i = \bar{A}) \bigcap (A_{k+1} = A)\right) \\
 &= \left[\prod_{i=1}^k P(A_i = \bar{A}) \right] P(A_{k+1} = A) \\
 &= q^k p
 \end{aligned} \tag{2.23}$$

Ejercicio: Compruébese que, para las variables discretas definidas, se verifica la igualdad $\sum_i p_i = 1$, donde el recorrido del índice i depende de la definición de cada VA.

2.4 Funciones condicionadas

De igual manera que se definieron probabilidades condicionadas en el tema anterior, podemos hablar de funciones de distribución y densidad de probabilidad condicionadas. El condicionante será un suceso relacionado con la VA \mathbf{X} y tal suceso nos proporcionará más información sobre el comportamiento probabilístico de la variable.

Considerando pues que el suceso B es el suceso condicionante ($P(B) > 0$) la función de distribución condicionada se define como

$$F_{\mathbf{X}}(x|B) = P(\mathbf{X} \leq x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} \tag{2.24}$$

El suceso intersección estará formado, como es sabido, por todos los $a \in S$ para los que, simultáneamente se verifica que $a \in B$ y $a \in \{\mathbf{X} \leq x\}$.

A partir de la función de distribución condicionada se define la función de densidad condicionada mediante el operador derivada, de forma similar a como se hizo en la expresión (2.7)

$$f_{\mathbf{X}}(x|B) = \frac{dF_{\mathbf{X}}(x|B)}{dx} \tag{2.25}$$

Consideremos dos casos particulares de suceso condicionante B , y, para caso caso, obtengamos las funciones de distribución y densidad condicionadas como función de las funciones incondicionales:

- $B = \{\mathbf{X} \leq a\}$, con $a \in \mathcal{R}$. Procediendo según la definición:

$$F_{\mathbf{X}}(x|B) = \begin{cases} x \leq a & F_{\mathbf{X}}(x|B) = P(\mathbf{X} \leq x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} = \\ & \frac{P(\mathbf{X} \leq x)}{P(B)} = \frac{F_{\mathbf{X}}(x)}{F_{\mathbf{X}}(a)} \\ x > a & F_{\mathbf{X}}(x|B) = P(\mathbf{X} \leq x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} = \\ & \frac{P(B)}{P(B)} = 1 \end{cases} \quad (2.26)$$

por lo que la función de densidad condicionada sería:

$$f_{\mathbf{X}}(x|B) = \begin{cases} x \leq a & f_{\mathbf{X}}(x|B) = \frac{f_{\mathbf{X}}(x)}{F_{\mathbf{X}}(x)} = \frac{f_{\mathbf{X}}(x)}{\int_{-\infty}^a f_{\mathbf{X}}(x)dx} \\ x > a & f_{\mathbf{X}}(x|B) = 0 \end{cases} \quad (2.27)$$

Como puede observarse, dado que el condicionante informa de que la variable \mathbf{X} ha tomado valores menores o iguales que el valor de abscisa a , es obvio que la variable no puede tomar valores a la izquierda de dicho punto, de forma que la función de distribución condicionada es constante para valores $x > a$, o, de forma equivalente, la función de densidad condicionada es nula en dicho intervalo. Por lo tanto, las funciones condicionadas, como ya hemos dicho, actualizan nuestro conocimiento probabilístico sobre la variable en cuestión a partir de la información proporcionada por el condicionante.

- $B = \{a < \mathbf{X} \leq b\}$. Distingamos ahora tres zonas:

$$F_{\mathbf{X}}(x|B) = \begin{cases} x < a & F_{\mathbf{X}}(x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0 \\ a \leq x \leq b & F_{\mathbf{X}}(x|B) = P(\mathbf{X} \leq x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} = \\ & \frac{P(a < \mathbf{X} \leq x)}{P(B)} = \frac{F_{\mathbf{X}}(x) - F_{\mathbf{X}}(a)}{F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a)} \\ x > b & F_{\mathbf{X}}(x|B) = \frac{P(\mathbf{X} \leq x, B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \end{cases} \quad (2.28)$$

Derivando la expresión anterior obtenemos la función de densidad condicionada:

$$f_{\mathbf{X}}(x|B) = \begin{cases} x < a & f_{\mathbf{X}}(x|B) = 0 \\ a \leq x \leq b & f_{\mathbf{X}}(x|B) = \frac{f_{\mathbf{X}}(x)}{F_{\mathbf{X}}(b) - F_{\mathbf{X}}(a)} = \frac{f_{\mathbf{X}}(x)}{\int_a^b f_{\mathbf{X}}(x)dx} \\ x > b & f_{\mathbf{X}}(x|B) = 0 \end{cases} \quad (2.29)$$

- Supóngase ahora que \mathbf{X} es discreta y que puede tomar valores en el conjunto $\{x_1, x_2, x_3\}$ con probabilidades respectivas $P(\mathbf{X} = x_i) = p_i$,

$i = \{1, 2, 3\}$. Denominemos $B = \{\mathbf{X} \leq a\}$, donde $x_2 < a < x_3$. Entonces se verifica que:

$$F_{\mathbf{X}}(x|B) = \sum_i p_{i|B} u(x - x_i)$$

con $p_{i|B} = P(\mathbf{X} = x_i|B)$, $i = \{1, 2, 3\}$. En particular

$$\begin{aligned} p_{1|B} &= \frac{P(\mathbf{X} = x_1, \mathbf{X} \leq a)}{P(\mathbf{X} \leq a)} = \frac{P(\mathbf{X} = x_1)}{P(\mathbf{X} = x_1 \cup \mathbf{X} = x_2)} = \frac{p_1}{p_1 + p_2} \\ p_{2|B} &= \frac{P(\mathbf{X} = x_2, \mathbf{X} \leq a)}{P(\mathbf{X} \leq a)} = \frac{P(\mathbf{X} = x_2)}{P(\mathbf{X} = x_1 \cup \mathbf{X} = x_2)} = \frac{p_2}{p_1 + p_2} \\ p_{3|B} &= \frac{P(\mathbf{X} = x_3, \mathbf{X} \leq a)}{P(\mathbf{X} \leq a)} = \frac{0}{P(\mathbf{X} = x_1 \cup \mathbf{X} = x_2)} = 0 \end{aligned}$$

2.4.1 Teoremas de la Probabilidad Total y de Bayes

Los Teoremas de la Probabilidad Total y de Bayes vistos en el tema anterior (véase sección 1.6) se pueden aplicar de forma inmediata a las funciones de distribución y densidad condicionadas. En efecto, si consideramos que los sucesos A_i , $i = \{1, 2, \dots, N\}$, constituyen una partición del espacio muestral S , podemos identificar el suceso B empleado en la sección 1.6 con el suceso $\{\mathbf{X} \leq x\}$ e, identificando término a término,

$$\begin{aligned} P(\mathbf{X} \leq x) &= \sum_{i=1}^N P(\mathbf{X} \leq x|A_i)P(A_i) \implies \\ F_{\mathbf{X}}(x) &= \sum_{i=1}^N F_{\mathbf{X}}(x|A_i)P(A_i) \end{aligned} \tag{2.30}$$

y, derivando con respecto a la variable x ,

$$f_{\mathbf{X}}(x) = \sum_{i=1}^N f_{\mathbf{X}}(x|A_i)P(A_i) \tag{2.31}$$

Estas expresiones indican que se puede construir una función de densidad en base a la superposición de funciones de densidad condicionadas, cada una de ellas multiplicada por un peso relativo igual a la probabilidad del elemento de la partición a que está condicionada. A una función de densidad así constituida se le denomina *mezcla* de funciones de densidad.

Ejemplo: Continuando con lo expuesto en la sección 1.6.4 en relación con los sistemas de comunicación, pensemos que las observaciones asociadas al envío del símbolo lógico 0 son muestras de una VA $\mathbf{X} \sim N(a_0, b_0)$, mientras que las observaciones asociadas al símbolo lógico 1 son muestras de $\mathbf{X} \sim N(a_1, b_1)$. Manteniendo la terminología de dicha sección, la función de densidad $f_{\mathbf{X}}(x)$ de la observación podría escribirse

$$f_{\mathbf{X}}(x) = \frac{1}{b_0 \sqrt{2\pi}} e^{-\frac{(x-a_0)^2}{2b_0^2}} P(H_0) + \frac{1}{b_1 \sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2b_1^2}} P(H_1)$$

La figura 2.6 muestra el aspecto de dicha gráfica para los valores ($a_0 = -3, b_0 = 1.5$), ($a_1 = 3, b_1 = 1.5$), $P(H_0) = 1/3$, $P(H_1) = 1 - 1/3$. Como puede observar, los máximos de las curvas son distintos, dado que cada una de ellas se encuentra multiplicada por un peso $P(H_i)$ ($i = 0, 1$) diferente. Por otra parte existe un grado de solapamiento importante entre las dos curvas, lo cual puede llevar a error al decisor del símbolo enviado instalado en el receptor. En efecto, si consideramos un determinado valor dentro de la zona de solapamiento, digamos, $x = 0.5$, si bien es más probable que haya sido generado por la distribución $f_{\mathbf{X}}(x|H_1)$, también puede haber sido generado por la distribución $f_{\mathbf{X}}(x|H_0)$ y con una probabilidad razonablemente elevada (esa probabilidad será tanto mayor cuanto mayor sea la función de densidad condicionada en dicho punto). Ello, por lo tanto, induce a confusión en el receptor. Cuantificaremos estas afirmaciones en el ejemplo de la sección 2.4.2.

El Teorema de Bayes se obtiene sin más que intercambiar los condicionantes en las expresiones anteriores

$$P(A_i | \mathbf{X} \leq x) = \frac{F_{\mathbf{X}}(x|A_i)P(A_i)}{F_{\mathbf{X}}(x)} = \frac{F_{\mathbf{X}}(x|A_i)P(A_i)}{\sum_{j=1}^N F_{\mathbf{X}}(x|A_j)P(A_j)} \quad (2.32)$$

2.4.2 Teoremas de la Probabilidad Total y de Bayes mediante partición continua

Nada impide considerar como partición *cada uno* de los elementos de la recta real \mathcal{R} . En tal caso, no podremos indexar los elementos de la partición mediante un índice entero, pues la partición será no numerable, de

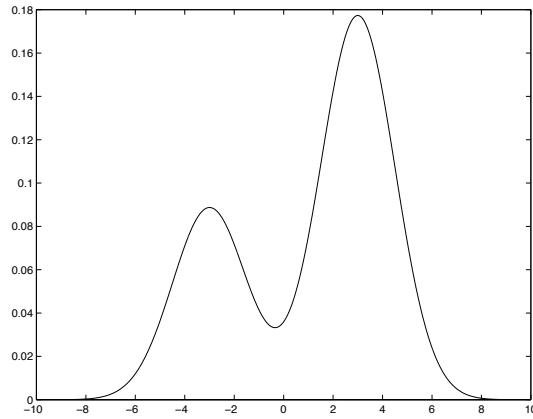


Figura 2.6: Función de densidad mezcla de dos gaussianas.

forma que los operadores de suma pasarán a ser sumas continuas, esto es, integrales. Sin embargo, conceptualmente, no existe diferencia alguna.

Para ver cómo podemos hacer esto, consideremos el suceso condicionante $B = \{x_1 < \mathbf{X} \leq x_2\}$. Acorde con él, podemos escribir

$$\begin{aligned} P(A|B) &= P(A|x_1 < \mathbf{X} \leq x_2) = \frac{P(x_1 < \mathbf{X} \leq x_2|A)}{P(x_1 < \mathbf{X} \leq x_2)} \\ &= \frac{F_{\mathbf{X}}(x_2|A) - F_{\mathbf{X}}(x_1|A)}{F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1)} P(A) \end{aligned} \quad (2.33)$$

El objetivo es hacer que el suceso B tienda a contener sólo un elemento. Para ello, hagamos que $x_1 = x$ y $x_2 = x + \Delta x$. Cuanto más pequeño sea Δx más estrecho será el intervalo que manejemos y, en el límite, contendrá sólo a $\mathbf{X} = x$. Por ello:

$$\begin{aligned} \lim_{\Delta x \rightarrow 0} P(A|x < \mathbf{X} \leq x + \Delta x) &= \frac{F_{\mathbf{X}}(x + \Delta x|A) - F_{\mathbf{X}}(x|A)}{F_{\mathbf{X}}(x + \Delta x) - F_{\mathbf{X}}(x)} P(A) \\ &= \frac{\frac{F_{\mathbf{X}}(x + \Delta x|A) - F_{\mathbf{X}}(x|A)}{\Delta x}}{\frac{F_{\mathbf{X}}(x + \Delta x) - F_{\mathbf{X}}(x)}{\Delta x}} P(A) \\ &= \frac{f_{\mathbf{X}}(x|A)}{f_{\mathbf{X}}(x)} P(A) = P(A|\mathbf{X} = x) \end{aligned} \quad (2.34)$$

La probabilidad $P(A|\mathbf{X} = x)$ debe entenderse como una forma compacta de escribir la expresión del primer miembro de la igualdad en la ecuación

anterior (primera línea de la ecuación 2.34). Véase que, de no entenderlo de esta manera sino en término estrictos, estaríamos hablando de una probabilidad condicionada en la que el condicionante sería un suceso de probabilidad nula. Recuérdese (sección 1.4) que la asunción de partida es la no nulidad en la probabilidad del suceso condicionante.

Quedándonos con la igualdad de la última línea de (2.34) podemos escribir

$$P(A)f_{\mathbf{X}}(x|A) = P(A|\mathbf{X} = x)f_{\mathbf{X}}(x) \quad (2.35)$$

Integrando ambos miembros en toda la recta real, y recordando que el área bajo una función de densidad es unitaria, obtenemos

$$\begin{aligned} P(A) &= \int_{-\infty}^{\infty} P(A)f_{\mathbf{X}}(x|A)dx \\ &= \int_{-\infty}^{\infty} P(A|\mathbf{X} = x)f_{\mathbf{X}}(x)dx \end{aligned} \quad (2.36)$$

lo cual no es otra cosa que el Teorema de la Probabilidad Total, empleando como partición cada uno de los elementos de la recta real.

Por otra parte, despejando de (2.34) la función de densidad condicionada, podemos escribir

$$\begin{aligned} f_{\mathbf{X}}(x|A) &= \frac{P(A|\mathbf{X} = x)f_{\mathbf{X}}(x)}{P(A)} \\ &= \frac{P(A|\mathbf{X} = x)f_{\mathbf{X}}(x)}{\int_{-\infty}^{\infty} P(A|\mathbf{X} = x)f_{\mathbf{X}}(x)dx} \end{aligned} \quad (2.37)$$

expresión que constituye el Teorema de Bayes en forma continua, es decir, la probabilidad de cada elemento de la partición ($\mathbf{X} = x$) condicionada al suceso A que se ha verificado. Estrictamente hablando, en este caso tendríamos densidades de probabilidad en vez de probabilidades. No obstante, como es sabido, bastaría multiplicar a ambos lados de la igualdad por una longitud elemental de intervalo (digamos, dx) para convertir esta igualdad en igualdad de probabilidades, de ahí que manejemos indistintamente uno u otro término.

Ejemplo: Podemos reescribir la ecuación (2.35) de la forma

$$P(A|\mathbf{X} = x) = \frac{f_{\mathbf{X}}(x|A)P(A)}{f_{\mathbf{X}}(x)} \quad (2.38)$$

y, volviendo al ejemplo del sistema de comunicaciones digitales, calcular esa probabilidad para cada uno de los símbolos cuando

el valor de señal observado sea $x = 0.5$. Si realizamos los cálculos para la selección de valores indicada anteriormente (valores para la construcción de la figura 2.6), tendríamos

$$\begin{aligned} P(H_0|\mathbf{X} = 0.5) &= \frac{f_{\mathbf{X}}(x = 0.5|H_0)P(H_0)}{f_{\mathbf{X}}(x = 0.5)} = 0.1165 \\ P(H_1|\mathbf{X} = 0.5) &= \frac{f_{\mathbf{X}}(x = 0.5|H_1)P(H_1)}{f_{\mathbf{X}}(x = 0.5)} = 0.8835 \end{aligned}$$

En efecto, el sistema actualiza su grado de conocimiento probabilístico sobre cada una de las hipótesis en base a la observación realizada. Anteriormente, $P(H_1) = 2/3 \approx 0.67$. Ahora, la creencia en que el símbolo enviado haya sido el correspondiente a H_1 se refuerza (pasa a ser 0.8835) en base a la observación obtenida $\{\mathbf{X} = 0.5\}$, de forma que el sistema decidirá que el símbolo enviado ha sido, efectivamente, el símbolo lógico 1 ya que $P(H_1|\mathbf{X} = 0.5) > P(H_0|\mathbf{X} = 0.5)$.

Sin embargo, el hecho de que $P(H_0|\mathbf{X} = 0.5) \neq 0$ hace que, si bien esta hipótesis es más improbable, que no sea imposible. De hecho, de cada 100 observaciones del valor $\{\mathbf{X} = 0.5\}$, aproximadamente 12 procederán del envío del símbolo lógico 0. Esto hace que el sistema no sea infalible.



2.5 Caracterización parcial de una variable aleatoria.

La caracterización probabilística de una VA viene dada, hasta donde conocemos, por dos funciones, bien la función de distribución, bien la de densidad de probabilidad. Sin embargo, en numerosos problemas basta conocer algunos parámetros que caracterizan— aunque sea de forma parcial —a la variable. En esta sección nos centraremos en los parámetros más habituales de caracterización parcial de las variables, y analizaremos cómo podemos obtener éstos para algunos de los tipos de variables vistos en la sección 2.3.

2.5.1 Media

Consideremos que disponemos de M observaciones de una variable (por ejemplo, medimos la altura de M personas). Aceptemos que medimos

con una precisión de 5 centímetros, de forma que nuestras medidas son $1.60m$, $1.65m$ etc.... Denotemos por h_j los valores de estas medidas, con $j = 1, \dots, N$. Pensemos, finalmente, que k_j personas, de las M posibles, han resultado tener una altura igual a h_j . Es obvio pues que $\sum_j k_j = M$. Con estas premisas, el valor medio de las alturas medidas se obtendría, de una forma natural, de la forma

$$\bar{h} = \sum_{j=1}^N \frac{k_j h_j}{M} = \sum_{j=1}^N \frac{k_j}{\sum_{i=1}^N k_i} h_j = \sum_{j=1}^N r_j h_j \quad (2.39)$$

es decir, el valor medio sería la media ponderada de todas las alturas h_j , donde el coeficiente de ponderación r_j indica la *importancia relativa* de la altura h_j con respecto a las otras alturas, y esa importancia relativa no es otra cosa que la fracción de observaciones, con respecto al total, cuya altura es h_j . Para el caso particular en que todas las alturas aparezcan por igual en el registro de datos ($r_j = 1/N$), entonces la media ponderada pasa a ser la media aritmética tradicional

$$\bar{h} = \sum_{j=1}^N r_j h_j = \frac{1}{N} \sum_{j=1}^N h_j \quad (2.40)$$

Consideremos ahora una VA \mathbf{X} y definamos su valor medio (al cual denotaremos por $\eta_{\mathbf{X}}$). Centrémonos, inicialmente, en las VAs discretas. El valor medio de una VA \mathbf{X} discreta se define como la *media ponderada* de los valores que puede tomar la variable. Si esta variable puede tomar los valores x_i con probabilidades respectivas $p_i = P(\mathbf{X} = x_i)$, $i = \{1, 2, \dots, N\}$, la media se obtiene mediante

$$\eta_{\mathbf{X}} = \sum_{i=1}^N x_i P(\mathbf{X} = x_i) = \sum_{i=1}^N x_i p_i \quad (2.41)$$

Como puede comprobarse, esta definición es una generalización de la ecuación (2.39), donde los coeficientes de ponderación r_i (los cuales serían función de los datos particulares de que se dispusiese) pasan a ser aquí las probabilidades p_i de cada uno de los valores de la variable. Huelga decir que si la VA discreta tiene cardinal infinito, el sumatorio se escribiría en ese mismo rango.

Para el caso general de variables continuas, la definición (2.41) se generaliza de la forma siguiente:

$$\eta_{\mathbf{X}} = E\{\mathbf{X}\} = \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x) dx \quad (2.42)$$

donde $E\{\cdot\}$ denota el operador *esperanza matemática*, operador que consiste en multiplicar su argumento por la función de densidad de la VA involucrada e integrar ese producto en todo el recorrido de la variable. Obsérvese que esta definición es la extensión continua de la ecuación (2.41), pues el operador sumatorio se ha sustituido por su extensión continua (integral), y las probabilidades p_i se han sustituido por su equivalente continuo, $P(x < \mathbf{X} \leq x + dx) = f_{\mathbf{X}}(x)dx$.

Asimismo, podemos comprobar de manera sencilla que la definición continua se convierte en la discreta cuando la función de densidad que se emplee venga dada por la ecuación (2.10). En efecto, si la VA \mathbf{X} es discreta, su función de densidad vendrá dada por la expresión

$$f_{\mathbf{X}}(x) = \sum_i p_i \delta(x - x_i) \quad (2.43)$$

de forma que el valor medio

$$\begin{aligned} \eta_{\mathbf{X}} = E\{\mathbf{X}\} &= \int_{-\infty}^{\infty} x \left[\sum_{i=1}^{\infty} p_i \delta(x - x_i) \right] dx \\ &= \sum_i p_i \int_{-\infty}^{\infty} x \delta(x - x_i) dx \\ &= \sum_i p_i x_i \int_{-\infty}^{\infty} \delta(x - x_i) dx \\ &= \sum_i p_i x_i \end{aligned} \quad (2.44)$$

como se indicó previamente.

2.5.2 Otras medidas de tendencia central

La media de una VA \mathbf{X} es un parámetro de tendencia central pues, como hemos dicho, indica un valor en torno al cual se distribuyen los valores de la variable. Este valor, sin duda alguna, es el más empleado, si bien no es el único. Otras medidas de tendencia central son las siguientes:

- Moda: La moda de una VA \mathbf{X} es el valor de dicha variable para el cual la función de densidad es máxima. Formalmente, si denotamos a la moda por x_{MODA} , podemos escribir

$$x_{MODA} = \arg \max_x f_{\mathbf{X}}(x) \quad (2.45)$$

El significado de la moda es el valor de la VA de mayor probabilidad. Por tanto, para el caso discreto, la moda se define como aquel valor x_i cuya probabilidad p_i es máxima.

- Mediana: el valor mediano, o mediana de una distribución, es el valor de abscisa para el cual la probabilidad de que la VA tome valores menores o iguales que dicho valor de abscisa coincida con la probabilidad de que la VA tome valores superiores a éste. Puede también definirse como el valor de la VA para el cual la probabilidad acumulada a su izquierda es igual a 0.5. En cualquier caso, denotando a la mediana por x_{ME} , entonces

$$\begin{aligned} P(\mathbf{X} \leq x_{ME}) &= P(\mathbf{X} > x_{ME}) \\ \int_{-\infty}^{x_{ME}} f_{\mathbf{X}}(x) dx &= \int_{x_{ME}}^{\infty} f_{\mathbf{X}}(x) dx \end{aligned} \quad (2.46)$$

Nótese que para el caso de VAs discretas este valor en general no coincidirá con ningún valor, digamos x_i , que la VA \mathbf{X} discreta pueda tomar. En tal caso si el valor x_p de la misma es tal que $P(\mathbf{X} \leq x_p) < 0.5$ y $P(\mathbf{X} \leq x_{p+1}) > 0.5$ entonces podemos definir el valor mediano como $x_{ME} = \frac{x_p + x_{p+1}}{2}$.

Tanto la moda como el valor mediano tiene interesantes aplicaciones prácticas dentro del ámbito de Teoría de la Señal; algunos ejemplos son el análisis de imagen mediante estimación bayesiana [4] o el filtrado de ruido impulsivo (en este último caso, el filtro de mediana [16, 17]). A pesar de ello no les prestaremos mayor atención en el resto de este tratado.

2.5.3 Varianza

La varianza de una VA \mathbf{X} es un valor de dispersión con respecto al valor medio $\eta_{\mathbf{X}}$. Concretamente, se denota por $\sigma_{\mathbf{X}}^2$, y se define

$$\sigma_{\mathbf{X}}^2 = E\{(\mathbf{X} - \eta_{\mathbf{X}})^2\} = \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})^2 f_{\mathbf{X}}(x) dx \quad (2.47)$$

A partir de la definición dada en la ecuación (2.42) se puede deducir que la varianza de una VA \mathbf{X} es la media de las desviaciones cuadráticas de los valores de la VA con respecto a su valor medio.

El parámetro $\sigma_{\mathbf{X}}$ ($\sigma_{\mathbf{X}} > 0$) se denomina *desviación típica* de la variable \mathbf{X} . El significado de este parámetro es el mismo que el de la varianza, sólo que en la misma unidad de medida que la media (repare que la varianza es una magnitud cuadrática con respecto a la de la media).

Si desarrollamos el cuadrado dentro de la función subintegral anterior

$$\begin{aligned}\sigma_{\mathbf{X}}^2 &= E\{(\mathbf{X} - \eta_{\mathbf{X}})^2\} = \int_{-\infty}^{\infty} (x^2 + \eta_{\mathbf{X}}^2 - 2x\eta_{\mathbf{X}}) f_{\mathbf{X}}(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{X}}(x) dx + \eta_{\mathbf{X}}^2 \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) dx - 2\eta_{\mathbf{X}} \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x) dx \\ &= E\{\mathbf{X}^2\} + \eta_{\mathbf{X}}^2 - 2\eta_{\mathbf{X}}^2 = E\{\mathbf{X}^2\} - \eta_{\mathbf{X}}^2\end{aligned}\quad (2.48)$$

donde al parámetro $E\{\mathbf{X}^2\}$ se le denomina *valor cuadrático medio* (VCM, en adelante) de la VA \mathbf{X} puesto que es igual a la media de los valores de la variable elevados al cuadrado.

Este resultado muestra varios conceptos de interés:

1. El operador *esperanza matemática*, dado que está definido mediante un operador integral, es un operador lineal. Podríamos haber obtenido el resultado de la ecuación (2.48) operando directamente sobre el operador esperanza, de la forma

$$\begin{aligned}\sigma_{\mathbf{X}}^2 &= E\{(\mathbf{X} - \eta_{\mathbf{X}})^2\} = E\{\mathbf{X}^2 - \eta_{\mathbf{X}}^2 - 2\eta_{\mathbf{X}}\mathbf{X}\} \\ &= E\{\mathbf{X}^2\} + \eta_{\mathbf{X}}^2 - 2E\{\eta_{\mathbf{X}}\mathbf{X}\} \\ &= E\{\mathbf{X}^2\} + \eta_{\mathbf{X}}^2 - 2\eta_{\mathbf{X}}E\{\mathbf{X}\} \\ &= E\{\mathbf{X}^2\} - \eta_{\mathbf{X}}^2\end{aligned}\quad (2.49)$$

dado que la esperanza de una constante es esa misma constante, y la esperanza de una constante multiplicada por una VA es igual a la constante multiplicada por la esperanza de la VA.

2. La varianza, según se desprende de su definición en la ecuación (2.47), es igual al valor medio de las desviaciones cuadráticas de los valores de la VA con respecto a su media $\eta_{\mathbf{X}}$. Por tanto, una VA cuya función de densidad esté muy concentrada alrededor de la media tendrá un valor pequeño de varianza, y sucederá lo contrario con funciones de densidad muy poco concentradas respecto a dicho valor medio.
3. La varianza puede calcularse de, al menos, dos maneras. La primera sería aplicando la definición dada en (2.47). La segunda sería calculando previamente $E\{\mathbf{X}^2\}$ y $\eta_{\mathbf{X}}$, para luego combinarlos según indica la ecuación (2.48). Son dos procedimientos distintos, hasta el punto de que, en función de las circunstancias, uno de ellos puede ser notablemente más sencillo que el otro.

Para el caso de VAs discretas, es fácil comprobar, sin más que sustituir la expresión genérica de la función de densidad por la particularización de la ecuación (2.10), que la expresión de la varianza pasa a ser

$$\begin{aligned}\sigma_{\mathbf{X}}^2 &= \sum_i (x_i - \eta_{\mathbf{X}})^2 p_i \\ &= E\{\mathbf{X}^2\} - \eta_{\mathbf{X}}^2 \\ &= \sum_i x_i^2 p_i - \left(\sum_i x_i p_i \right)^2\end{aligned}\quad (2.50)$$

La primera línea de la ecuación mostraría el cálculo directo, equivalente discreto de la ecuación (2.47), mientras que la tercera línea sería el equivalente discreto de (2.48).

Finalmente, es interesante mencionar que la varianza de una constante es igual a cero. Insistimos que la varianza mide *dispersión* con respecto a la media. Por ello, si hablamos de una constante, hablamos de una magnitud que no tiene variabilidad ninguna, lo cual debe ser coherente con varianza nula. En efecto, una constante es modelable como una VA discreta que sólo puede tomar un valor, digamos, valor a , con probabilidad $p_a = 1$. Por ello, su función de densidad sería (llamémosla \mathbf{X}) $f_{\mathbf{X}}(x) = \delta(x - a)$. Haciendo uso de las definiciones vistas anteriormente:

- $E\{\mathbf{X}\} = \sum_i p_i x_i = a \cdot 1 = a$
- $E\{\mathbf{X}^2\} = \sum_i p_i x_i^2 = a^2 \cdot 1 = a^2$
- $\sigma_{\mathbf{X}}^2 = E\{\mathbf{X}^2\} - E^2\{\mathbf{X}\} = a^2 - a^2 = 0$

2.5.4 Ejemplos para varias distribuciones

Uniforme

Consideremos $\mathbf{X} \sim U(a, b)$. Los cálculos en este caso son particularmente sencillos dada la simplicidad de la función de densidad

- Media:

$$\begin{aligned}E\{\mathbf{X}\} &= \int_{\infty}^{\infty} x f_{\mathbf{X}}(x) dx \\ &= \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}\end{aligned}\quad (2.51)$$

- VCM

$$\begin{aligned} E\{\mathbf{X}^2\} &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{X}}(x) dx \\ &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3} \left(\frac{b^3 - a^3}{b-a} \right) \end{aligned} \quad (2.52)$$

- Varianza: la integral es igualmente sencilla, pero requiere de un poco más de álgebra. Se puede comprobar que el resultado es:

$$\begin{aligned} \sigma_{\mathbf{X}}^2 &= \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})^2 f_{\mathbf{X}}(x) dx \\ &= \int_a^b \left(x - \frac{a+b}{2} \right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12} \end{aligned} \quad (2.53)$$

Puede verse por lo tanto que la media de una VA uniforme es igual al punto medio del intervalo donde la variable toma valores, y la varianza es función cuadrática de la longitud del intervalo. En efecto, cuanto mayor sea el intervalo, dado que todos los valores de la variable son equiprobables, mayor es la dispersión de los valores con respecto de la media.

Gaussiana

Consideremos ahora que $\mathbf{X} \sim N(a, b)$. El cálculo de media y varianza de la distribución se haría como sigue:

- Media: debemos resolver la integral

$$\begin{aligned} E\{\mathbf{X}\} = \eta_{\mathbf{X}} &= \int_{-\infty}^{\infty} x f_{\mathbf{X}} dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx \end{aligned} \quad (2.54)$$

Para resolver esta integral parece razonable, dado que la exponencial presenta como punto de simetría el punto $x = a$, bien desplazar la exponencial a $x = 0$, bien desplazar la función $f(x) = x$ al punto $x = a$. Lo segundo se hace de manera sencilla sin más que sumar y restar la constante a , obteniendo

$$E\{\mathbf{X}\} = \eta_{\mathbf{X}} = \int_{-\infty}^{\infty} (x - a + a) \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} (x-a) \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx + \\
 &\quad \int_{-\infty}^{\infty} a \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx
 \end{aligned} \tag{2.55}$$

La segunda integral es igual a a , dado que es el cálculo del área bajo una función de densidad, multiplicada por a . Al respecto de la primera, repárese que la función subintegral es el producto de la función $f(x) = x-a$, la cual es impar con respecto al punto $x=a$, por una exponencial con exponente al cuadrado, la cual es una función par con respecto al punto $x=a$. El producto por tanto es impar, e integrando en toda la recta real, dado que la integral existe, el resultado debe ser nulo. Por ello, podemos afirmar que

$$E\{\mathbf{X}\} = \eta_{\mathbf{X}} = a \tag{2.56}$$

- Varianza: debemos resolver la integral

$$\begin{aligned}
 \sigma_{\mathbf{X}}^2 &= \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})^2 f_{\mathbf{X}} dx \\
 &= \int_{-\infty}^{\infty} (x-a)^2 \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx
 \end{aligned} \tag{2.57}$$

En este caso no podemos explotar la paridad/imparidad de las funciones, puesto que todas las funciones involucradas son cuadráticas. El cálculo directo de la integral parece también complicado.

Sin embargo, sí que existen soluciones sencillas para resolver este problema. Una forma de resolverlo sería explotar el conocimiento previo de que disponemos. Concretamente, sabemos que

$$\begin{aligned}
 1 &= \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx \Rightarrow \\
 b &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} dx
 \end{aligned} \tag{2.58}$$

La primera línea procede del hecho de que el área bajo una función de densidad es igual a uno. La segunda igualdad se ha puesto por conveniencia y se obtiene de multiplicar a ambos miembros de la igualdad por el parámetro b . Si consideramos ahora esta igualdad como una función del parámetro b (por lo tanto, el parámetro pasa a ser una variable), nada impide que hagamos transformaciones sobre

esa función, en particular, que derivemos. ¿Qué ganamos con esto? Si comparamos las ecuaciones (2.57) y (2.58) veremos que para pasar de la segunda a la primera necesitamos que el término $(x - a)^2$ aparezca multiplicando a la exponencial. Esto podemos conseguirlo derivando la igualdad (2.58), considerada ésta como función de b , ya que la regla de la cadena dice que

$$\frac{de^{f(b)}}{db} = e^{f(b)} f'(b) \quad (2.59)$$

y es obvio que al derivar la función del exponente con respecto a b , el término $(x - a)^2$, el cual actúa como constante a efectos de b , pasará a multiplicar a la exponencial, como necesitamos.

Procediendo de esta manera

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} \left(-\frac{(x-a)^2}{2} \right) (-2) \frac{1}{b^3} dx \\ b^2 &= \int_{-\infty}^{\infty} (x-a)^2 \frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}} \end{aligned} \quad (2.60)$$

de forma que, según dice la expresión (2.57), $\sigma_{\mathbf{X}}^2 = b^2$.

Según los resultados obtenidos, el primer parámetro de la gaussiana (a) es la media y el segundo parámetro (b) es la desviación típica. Por ello, es práctica habitual denotar las variables gaussianas mediante $\mathbf{X} \sim N(\eta_{\mathbf{X}}, \sigma_{\mathbf{X}})$.

Ejercicio: Obtenga los valores de x_{MODA} y x_{ME} para $\mathbf{X} \sim N(\eta_{\mathbf{X}}, \sigma_{\mathbf{X}})$.

Solución: El primer parámetro de la curva es el punto de simetría, que coincide con el máximo. Por ello, $x_{MODA} = \eta_{\mathbf{X}}$. Asimismo y, de nuevo, debido a la simetría de la curva con respecto a $x = \eta_{\mathbf{X}}$, el área bajo la curva en el intervalo $(-\infty, \eta_{\mathbf{X}})$ coincide con el área en el intervalo $(\eta_{\mathbf{X}}, \infty)$. Por ello, $x_{ME} = \eta_{\mathbf{X}}$.



Exponencial

Consideremos ahora que $\mathbf{X} \sim exp(\lambda)$. El cálculo de media y varianza de la distribución se haría como sigue:

- Media: debemos resolver la integral

$$\begin{aligned} E\{\mathbf{X}\} = \eta_{\mathbf{X}} &= \int_0^\infty x f_{\mathbf{X}} dx \\ &= \int_0^\infty x \lambda e^{\lambda x} dx \end{aligned} \quad (2.61)$$

- VCM: en este caso, debemos resolver

$$\begin{aligned} E\{\mathbf{X}^2\} = \eta_{\mathbf{X}} &= \int_0^\infty x^2 f_{\mathbf{X}} dx \\ &= \int_0^\infty x^2 \lambda e^{\lambda x} dx \end{aligned} \quad (2.62)$$

Ambas integrales son de corte similar, y pueden resolverse sin más que aplicar integración *por partes*, con la consiguiente dosis de paciencia. No obstante, existen resultados conocidos, los cuales pueden servir de base para obtener los parámetros que nos piden de forma más rápida y elegante. Concretamente, hablamos de la función *gamma*, o factorial generalizado, la cual se define

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx = k^p \int_0^\infty x^{p-1} e^{-kx} dx \quad (2.63)$$

donde se ha de verificar que $\Re k > 0$ y $\Re p > 0$. Esta función verifica la ley de recurrencia $\Gamma(p+1) = p\Gamma(p)$, lo cual particulariza al caso de $p \in \mathbb{Z}^+$ en $\Gamma(p+1) = p!$.

Así pues no tenemos más que identificar términos en media y VCM con respecto a la función Γ . Por tanto

- $E\{\mathbf{X}\} = \int_0^\infty x \lambda e^{\lambda x} dx = \frac{1}{\lambda} \Gamma(2) = \frac{1}{\lambda}$
- $E\{\mathbf{X}^2\} = \int_0^\infty x^2 \lambda e^{\lambda x} dx = \frac{1}{\lambda^2} \Gamma(3) = \frac{2}{\lambda^2}$
- $\sigma_{\mathbf{X}}^2 = E\{\mathbf{X}^2\} - E^2\{\mathbf{X}\} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Bernoulli

Este caso es particularmente sencillo, pues los sumatorios se extienden a sólo dos valores. Por tanto

- Media: $E\{\mathbf{X}\} = \eta_{\mathbf{X}} = \sum_{i=1}^N x_i p_i = 0 \cdot q + 1 \cdot p = p$
- VCM: $E\{\mathbf{X}^2\} = \sum_{i=1}^N x_i^2 p_i = 0^2 \cdot q + 1^2 \cdot p = p$
- Varianza: $\sigma_{\mathbf{X}}^2 = E\{\mathbf{X}^2\} - E^2\{\mathbf{X}\} = p - p^2 = p(1 - p) = pq$

Binomial

Para el caso en que $\mathbf{X} \sim B(N, p)$ tenemos:

- Media:

$$E\{\mathbf{X}\} = \eta_{\mathbf{X}} = \sum_{i=0}^N x_i p_i = \sum_{i=0}^N \binom{N}{i} i p^i q^{N-i} \quad (2.64)$$

- VCM:

$$E\{\mathbf{X}^2\} = \sum_{i=0}^N x_i^2 p_i = \sum_{i=0}^N \binom{N}{i} i^2 p^i q^{N-i} \quad (2.65)$$

El cómo llevar a cabo estos sumatorios de forma directa no parece obvio, pero sí que es abordable en base a un resultado conocido, que no es otro que el desarrollo de un binomio de Newton:

$$(p + q)^N = \sum_{i=0}^N \binom{N}{i} p^i q^{N-i}$$

Necesitamos que el índice del sumatorio i multiplique a las probabilidades, tanto linealmente — para el caso de la media — como cuadráticamente, para el caso del VCM. Esto lo podemos conseguir sin más que considerar la expresión anterior como una función f de la variable p , y derivar con respecto a ella. En efecto:

$$\frac{df(p)}{dp} = N(p + q)^{N-1} = \sum_{i=1}^N \binom{N}{i} ip^{i-1} q^{N-i} = \sum_{i=0}^N \binom{N}{i} ip^{i-1} q^{N-i}$$

donde la última igualdad procede de que el término en $i = 0$ es nulo. Basta ahora que multipliquemos ambos miembros por p para que la última expresión coincida con la expresión (2.64). Por ello,

$$\begin{aligned} E\{\mathbf{X}\} = \eta_{\mathbf{X}} &= p \sum_{i=0}^N \binom{N}{i} ip^{i-1} q^{N-i} |_{p+q=1} \\ &= pN(p + q)^{N-1} |_{p+q=1} \\ &= Np \end{aligned} \quad (2.66)$$

Con respecto al VCM, partimos de la expresión

$$pN(p + q)^{N-1} = \sum_{i=0}^N \binom{N}{i} ip^i q^{N-i}$$

y derivando con respecto a p , obtenemos:

$$\begin{aligned} N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2} &= \sum_{i=1}^N \binom{N}{i} i^2 p^{i-1} q^{N-i} \\ &= \sum_{i=0}^N \binom{N}{i} i^2 p^{i-1} q^{N-i} \end{aligned}$$

de forma que multiplicando por p e igualando $p+q = 1$ obtenemos

$$Np + p^2 N(N-1) = \sum_{i=0}^N \binom{N}{i} i^2 p^i q^{N-i}$$

por lo que $E\{\mathbf{X}^2\} = Np(1 + p(N-1))$.

Estos resultados nos permiten obtener la varianza

$$\begin{aligned} \sigma_{\mathbf{X}}^2 &= E\{\mathbf{X}^2\} - E^2\{\mathbf{X}\} = Np + p^2 N(N-1) - (Np)^2 = Np - Np^2 \\ &= Np(1-p) = Npq. \end{aligned}$$

En vista de estos resultados, repárese que la aproximación de DeMoivre-Laplace (introducida en la sección 1.7.1 y que veremos más detalladamente en la sección 4.5) no es otra cosa que aproximar una variable $B(N, p)$ por una gaussiana, donde esta última tiene media y varianza iguales a las de la variable binomial.

Poisson

Si \mathbf{X} es una VA de Poisson de parámetro a , es fácil obtener que:

- Media: $\eta_{\mathbf{X}} = a$
- VCM: $E\{\mathbf{X}^2\} = a^2 + a$
- Varianza: $\sigma_{\mathbf{X}}^2 = a$

Ejercicio: Sabiendo que el desarrollo en serie de Taylor de la función exponencial es:

$$S(a) = e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}$$

obténganse los valores de la media, VCM y varianza de una variable exponencial de parámetro a .

Geométrica

Si \mathbf{X} es una VA geométrica de parámetro p ($q = 1 - p$), es fácil obtener que:

- Media: $\eta_{\mathbf{X}} = \frac{q}{p}$
- VCM: $E\{\mathbf{X}^2\} = \frac{q}{p^2}(1 + q)$
- Varianza: $\sigma_{\mathbf{X}}^2 = \frac{q}{p^2}$

Ejercicio: Sabiendo que la suma de los términos de una progresión geométrica es:

$$S(p) = \frac{p}{1 - q} = \sum_{k=0}^{\infty} q^k p$$

obténganse los valores de la media, VCM y varianza de una variable geométrica de parámetro p .

2.5.5 Desigualdad de Tchebycheff

La desigualdad de Tchebycheff proporciona una cota a la probabilidad de un determinado suceso de la VA \mathbf{X} , cota que es de utilidad en aquellas situaciones en las que se desconozca la función de densidad de la variable en cuestión. En efecto, como es sabido, el cálculo de cualquier probabilidad en relación con los valores de una VA requiere del conocimiento de, bien la función de distribución, bien la función de densidad. Sin embargo, cuando ésta no sea conocida pero se conozcan la media y la varianza de la VA, si bien no es posible calcular de forma exacta probabilidad alguna, sí que se puede proporcionar una cota.

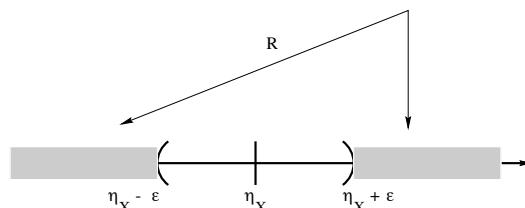


Figura 2.7: Ilustración de la región R para la cual $P(|\mathbf{X} - \eta_{\mathbf{X}}| \geq \epsilon)$.

Repárese que por *cota*, entendemos *cota no trivial*. Es obvio que cualquier probabilidad se encuentra entre cero y uno, de forma que está intrínsecamente acotada. Esta desigualdad pretende proporcionar un valor que

sirva de guía en aquellas situaciones en las que se disponga sólo de conocimiento parcial de la situación.

Dicho esto, indiquemos cuál es la cota. Se puede escribir como sigue:

$$P(|\mathbf{X} - \eta_{\mathbf{X}}| \geq \epsilon) \leq \frac{\sigma_{\mathbf{X}}^2}{\epsilon^2} \quad (2.67)$$

El significado físico de la desigualdad parece inmediato: la probabilidad de que una VA tome valores alejados de su media un determinado radio ϵ es tanto mayor cuanto mayor sea la dispersión de los valores de la misma con respecto al valor medio, es decir, es tanto mayor cuanto mayor sea la varianza (y tanto menor cuanto mayor sea el radio ϵ).

Es ilustrativo ver por qué esto es así. Si planteamos el cálculo estricto de la probabilidad contemplada en la desigualdad, debemos escribir

$$\begin{aligned} P(|\mathbf{X} - \eta_{\mathbf{X}}| \geq \epsilon) &= \int_{-\infty}^{\eta_{\mathbf{X}} - \epsilon} f_{\mathbf{X}}(x)dx + \int_{\eta_{\mathbf{X}} + \epsilon}^{\infty} f_{\mathbf{X}}(x)dx \\ &= \int_R f_{\mathbf{X}}(x)dx \end{aligned} \quad (2.68)$$

donde, por brevedad, se ha denominado región R a la región $(-\infty, \eta_{\mathbf{X}} - \epsilon) \cup (\eta_{\mathbf{X}} + \epsilon, \infty)$ (véase la ilustración de la figura 2.7).

Si recordamos ahora la definición de la varianza y explotamos el hecho de que en la región R se verifica que $(x - \eta_{\mathbf{X}})^2 \geq \epsilon^2$, podemos escribir

$$\begin{aligned} \sigma_{\mathbf{X}}^2 &= \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})^2 f_{\mathbf{X}}(x)dx \\ &\geq \int_R (x - \eta_{\mathbf{X}})^2 f_{\mathbf{X}}(x)dx \\ &\geq \int_R \epsilon^2 f_{\mathbf{X}}(x)dx \\ &= \epsilon^2 \int_R f_{\mathbf{X}}(x)dx = \epsilon^2 P(|\mathbf{X} - \eta_{\mathbf{X}}| \geq \epsilon) \end{aligned} \quad (2.69)$$

entonces, si más que pasar ϵ^2 al primer miembro de la desigualdad, obtenemos el resultado deseado.

Esta desigualdad, si bien es una herramienta de enorme aplicabilidad pues sólo necesita del conocimiento de la media y la varianza de la VA, paga un precio debido a su generalidad: el valor de la cota es bastante conservador. Esto lo podemos comprobar escogiendo un caso en el que conozcamos la caracterización completa de la VA, de tal forma que podamos comparar el valor exacto de la probabilidad del intervalo pedido con el valor

de la cota previsto en la desigualdad. Consideremos, por ejemplo, que $\mathbf{X} \sim N(\eta, \sigma)$ y pensemos que estamos interesados en hallar la probabilidad indicada en la ecuación (2.67) para el caso en que $\epsilon = 3\sigma$. Escogiendo un valor del radio del intervalo proporcional a la desviación típica de la variable podremos, como se verá, obtener un valor numérico de la probabilidad sin necesidad de seleccionar valores numéricos para los parámetros η y σ . En estas condiciones comparemos valor exacto con valor de la cota:

- Valor exacto:

$$\begin{aligned}
 P(|\mathbf{X} - \eta| \geq \epsilon) &= P(\mathbf{X} \leq \eta - \epsilon) + P(\mathbf{X} \geq \eta + \epsilon) \\
 &= G\left(\frac{\eta - \epsilon - \eta}{\sigma}\right) + (1 - G\left(\frac{\eta + \epsilon - \eta}{\sigma}\right)) \\
 &= G\left(\frac{-3\sigma}{\sigma}\right) + \left(1 - G\left(\frac{3\sigma}{\sigma}\right)\right) \\
 &= 2\left(1 - G\left(\frac{3\sigma}{\sigma}\right)\right) = 0.0027
 \end{aligned} \tag{2.70}$$

- Valor previsto por la cota

$$P(|\mathbf{X} - \eta| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} = \frac{\sigma^2}{9\sigma^2} = \frac{1}{9} = 0.11 \tag{2.71}$$

Nótese pues que, si bien la cota no es trivial, la diferencia entre el valor real y el previsto por la cota es de dos órdenes de magnitud. Es por tanto muy conservadora, como hemos indicado.

Ejercicio: Existen versiones derivadas del teorema. Una de ellas es la siguiente: considere que \mathbf{X} es una VA que sólo puede tomar valores no negativos. Se pide que demuestre que

$$P(\mathbf{X} \geq \delta) \leq \frac{E\{\mathbf{X}\}}{\delta}$$

con $\delta > 0$.

2.6 Transformación de variable aleatoria

Supongamos que a partir de un experimento aleatorio $\epsilon < S, \mathcal{F}, P >$ se define la VA \mathbf{X} . Aceptemos que conocemos la función de densidad de probabilidad $f_{\mathbf{X}}(x)$ de esta variable.

En esta sección consideraremos que definimos una función $y = g(x)$ sobre los valores de la VA \mathbf{X} de forma que para todo valor x de ésta existe una (y sólo una) imagen en la función $g(x)$.

La situación puede por tanto formalizarse de la manera

$$\begin{aligned} \mathbf{X} : S &\rightarrow \mathcal{R} \\ a \in S &\rightarrow \mathbf{X}(a) \end{aligned} \quad (2.72)$$

y

$$\begin{aligned} g : \mathcal{R} &\rightarrow \mathcal{R} \\ x \in \mathcal{R} &\rightarrow y = g(x) \in \mathcal{R} \end{aligned} \quad (2.73)$$

Nada impide componer ambas operaciones para dar lugar a

$$S \xrightarrow{\mathbf{X}} \mathcal{R} \xrightarrow{g} \mathcal{R} \quad (2.74)$$

de forma que, considerando los extremos de la composición, podemos resumir

$$\begin{aligned} g(\mathbf{X}) : S &\rightarrow \mathcal{R} \\ a \in S &\rightarrow \mathbf{Y}(\mathbf{X}(a)) \in \mathcal{R} \end{aligned} \quad (2.75)$$

Vemos por lo tanto que los valores reales que toma la función g son función del resultado obtenido en el experimento aleatorio $\epsilon < S, \mathcal{F}, P >$. Cabe pues considerar que la colección de resultados, \mathbf{Y} , constituye una VA, generada a partir de la transformación de \mathbf{X} mediante la función $g(x)$. Pues, en efecto, así será siempre que, como hemos visto en la sección 2.1, se cumplan unas determinadas condiciones, a saber⁷:

1. Que el dominio en que $g(x)$ está definida contenga a todos los valores de \mathbf{X} . Caso contrario, podrían existir valores $a \in S$ sin imagen en la variable \mathbf{Y} .
2. $\{\mathbf{Y} \leq y\} = \{\forall x \in \mathcal{R} \text{ t.q. } g(x) \leq y\} \in \mathcal{F}$, es decir, debe ser un suceso, de forma que sobre él puedan definirse probabilidades.
3. $P(\mathbf{Y} = -\infty) = P(\mathbf{Y} = \infty) = 0$, de forma que aunque las variables puedan tomar ambos valores, que lo hagan con probabilidad nula.

⁷Recuérdese el tratamiento que hacemos de los valores infinitos (véase sección 2.1).

Aceptando que se cumplen estas condiciones \mathbf{Y} sería una variable aleatoria y , por ello, susceptible de ser caracterizada probabilísticamente. Éste es el objetivo a partir de este momento: conociendo la caracterización de la variable \mathbf{X} , origen de la transformación, y conociendo la propia función de transformación $g(x)$, hallar la caracterización probabilística de la VA $\mathbf{Y} = g(\mathbf{X})$.

2.6.1 Caso de \mathbf{X} discreta

Supongamos que la VA \mathbf{X} puede tomar los valores x_i , con probabilidades $P(\mathbf{X} = x_i) = p_i$, con i índice entero, variando en un cierto rango. Supongamos que se define una función de transformación $\mathbf{Y} = g(\mathbf{X})$. En estas condiciones es obvio que la VA \mathbf{Y} también será discreta, de forma que su caracterización se hará hallando los valores que puede tomar, así como las probabilidades con las que los toma. Veamos:

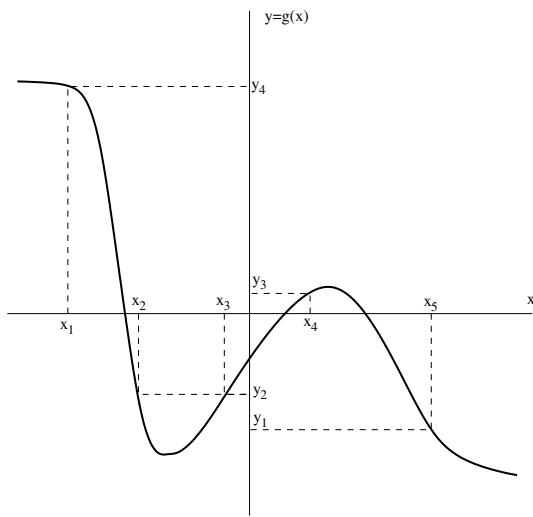


Figura 2.8: Transformación de una VA discreta

1. Valores y_j : Los valores que puede tomar \mathbf{Y} son los valores transformados directamente de $g(x_i)$. Si la transformación es inyectiva, es decir, si $\forall x_i \neq x_j \rightarrow g(x_i) \neq g(x_j)$ la variable \mathbf{Y} tomará tantos valores distintos como la variable \mathbf{X} y estos valores serán $y_j = g(x_i)$.

Si la función no es inyectiva hay que ver qué valores origin x_i se transforman en la misma imagen y . En este caso, la variable \mathbf{Y} tomará un número menor de valores distintos que la VA \mathbf{X} .

2. Probabilidades $P(\mathbf{Y} = y_i)$:

- Si $g(x)$ inyectiva e $y_j = g(x_i)$ entonces $p_i = P(\mathbf{X} = x_i) = P(\mathbf{Y} = y_j)$, con (i, j) variando en los rangos oportunos..
- Si $g(x)$ no es inyectiva: considerando la situación representada en la figura 2.8, en la cual la VA \mathbf{X} toma cinco posibles valores (x_1, \dots, x_5) , de ellos tres (x_1, x_4, x_5) se transforman en distintas imágenes, pero existen dos puntos $(x_2$ y $x_3)$ cuyas imágenes coinciden. Así pues, analizando cada caso individual, podemos escribir

$$\begin{aligned}
 P(\mathbf{Y} = y_1) &= P(\mathbf{X} = x_5) = p_5 \\
 P(\mathbf{Y} = y_2) &= P((\mathbf{X} = x_2) \cup (\mathbf{X} = x_3)) \\
 &= P(\mathbf{X} = x_2) + P(\mathbf{X} = x_3) = p_2 + p_3 \\
 P(\mathbf{Y} = y_3) &= P(\mathbf{X} = x_4) = p_4 \\
 P(\mathbf{Y} = y_4) &= P(\mathbf{X} = x_1) = p_1
 \end{aligned} \tag{2.76}$$

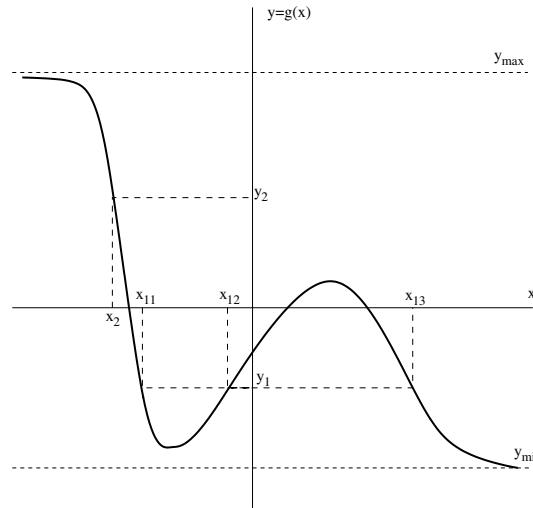
Esta metodología es común y de aplicabilidad general para cualquier VA discreta y cualquier transformación.

2.6.2 Caso general. Obtención de la función de distribución

El cálculo de la función de distribución se lleva a cabo de manera constructiva. Recordando que la función de distribución se define $F_{\mathbf{Y}}(y) = P(\mathbf{Y} \leq y)$, el procedimiento consiste en ir barriendo el eje de ordenadas en los diferentes valores de y e ir escribiendo el suceso $\{\mathbf{Y} \leq y\}$ como función de los sucesos correspondientes de la VA \mathbf{X} , variable cuya caracterización probabilística tenemos.

Para ilustrar este procedimiento, haciendo uso de la figura 2.9, analicemos los diferentes casos que pueden presentarse:

1. Si consideramos valores $y < y_{min}$, se verifica que $P(\mathbf{Y} \leq y) = 0$. Esto es debido a que ningún valor de la VA \mathbf{X} se transforma en valores de la VA \mathbf{Y} menores que y_{min} . Por ello, la función de distribución es nula en el intervalo $(-\infty, y_{min})$.
2. Centrémonos ahora en la parte superior de la transformación, es decir, en $y > y_{max}$. La probabilidad $P(\mathbf{Y} \leq y) = 1$ pues, como se deduce de la gráfica, todos los valores de \mathbf{X} se transforman en valores

Figura 2.9: Cálculo de la función $F_{\mathbf{Y}}(y)$.

de \mathbf{Y} menores o iguales que y_{max} . De esta manera, la función de distribución toma el valor 1 en el intervalo (y_{max}, ∞) .

3. Debemos ahora ir recorriendo el rango no explorado todavía, es decir, el intervalo $y \in (y_{min}, y_{max})$. Un punto perteneciente a este intervalo es y_2 . Para este punto, podemos escribir

$$F_{\mathbf{Y}}(y_2) = P(\mathbf{Y} \leq y_2) = P(\mathbf{X} > x_2) = 1 - P(\mathbf{X} \leq x_2) = 1 - F_{\mathbf{X}}(x_2) \quad (2.77)$$

4. Consideremos ahora el punto $\mathbf{Y} = y_1$. Para este caso podemos escribir

$$\begin{aligned} F_{\mathbf{Y}}(y_1) &= P(\mathbf{Y} \leq y_1) = P((x_{11} \leq \mathbf{X} \leq x_{12}) \cup (\mathbf{X} \geq x_{13})) \\ &= P(x_{11} \leq \mathbf{X} \leq x_{12}) + P(\mathbf{X} \geq x_{13}) \\ &= (F_{\mathbf{X}}(x_{12}) - F_{\mathbf{X}}(x_{11}^-)) + (1 - F_{\mathbf{X}}(x_{13}^-)) \end{aligned} \quad (2.78)$$

Si la VA \mathbf{X} es continua, naturalmente, la expresión (2.78) se puede reescribir $F_{\mathbf{Y}}(y_1) = F_{\mathbf{X}}(x_{12}) - F_{\mathbf{X}}(x_{11}) + (1 - F_{\mathbf{X}}(x_{13}))$ (véase sección 2.2.1).

Ejercicio: Considerando que $\mathbf{X} \sim U(-1/2, 1/2)$ y que $\mathbf{Y} = g(\mathbf{X}) = \mathbf{X}^2$, obtenga mediante el método expuesto la función de distribución $F_{\mathbf{Y}}(y)$.

Solución:

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ 2\sqrt{y} & 0 < y \leq \frac{1}{4} \\ 1 & y > \frac{1}{4} \end{cases}$$



Algunos casos particulares a tener en cuenta

El procedimiento anterior es válido para cualquier situación. No obstante, es interesante resaltar algunas situaciones particulares en las que, a primera vista, pueden surgir dificultades de interpretación.

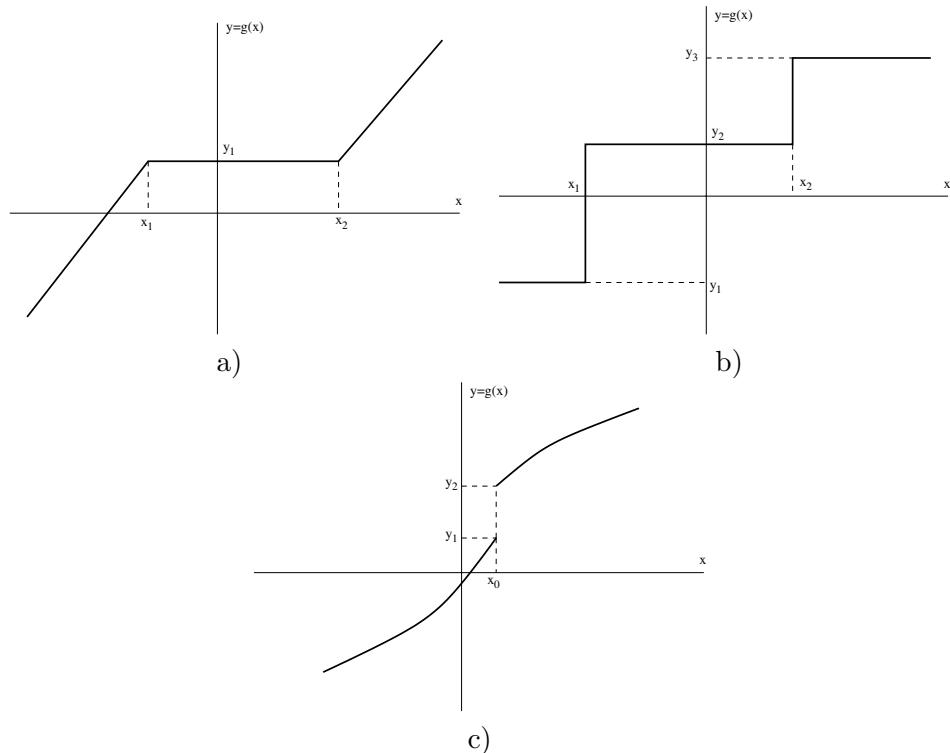


Figura 2.10: a) Transformación con intervalo de constancia. b) Transformación en escalera. c) Transformación con discontinuidad.

1. Si la transformación es constante en un intervalo de valores de la VA \mathbf{X} y esta VA es continua, entonces la VA resultante \mathbf{Y} será mixta

(figura 2.10a). Repárese que en la figura existe un rango continuo de valores de la VA \mathbf{X} que se transforma en el punto $\mathbf{Y} = y_1$, en vez de, como sucedía en el caso general, un número discreto de puntos transformándose en cualquier imagen. Por ello, el suceso $\{\mathbf{Y} = y_1\}$ presenta una probabilidad

$$P(\mathbf{Y} = y_1) = P(x_1 < \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1) = p_{y_1} \quad (2.79)$$

y eso trae consigo que la función $F_{\mathbf{Y}}(y)$ presente una discontinuidad de altura p_{y_1} en el punto de abscisa $y = y_1$. Una posible forma de función de distribución resultante para este caso es la que se muestra en la figura 2.11a).

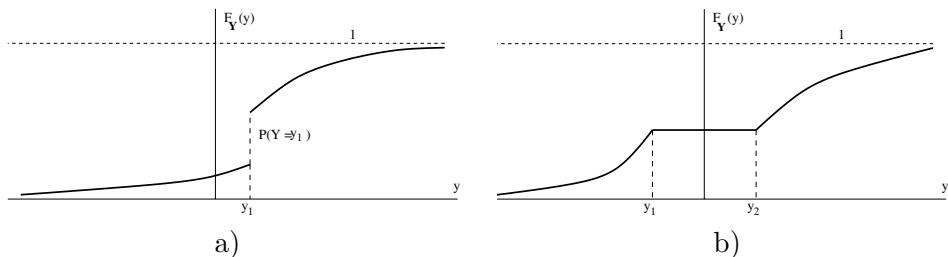


Figura 2.11: a) $F_{\mathbf{Y}}(y)$ resultante de transformación con intervalo de constancia. b) $F_{\mathbf{Y}}(y)$ resultante de transformación con discontinuidad.

2. Lo anterior generaliza a una transformación *en escalera* (véase figura 2.10b). Para una transformación como la de esta figura sabemos de antemano que la VA \mathbf{Y} será discreta y que sólo tomará los valores (y_1, y_2, y_3) con probabilidades respectivas:

$$\begin{aligned} P(\mathbf{Y} = y_1) &= P(\mathbf{X} \leq x_1) = F_{\mathbf{X}}(x_1) \\ P(\mathbf{Y} = y_2) &= P(x_1 < \mathbf{X} \leq x_2) = F_{\mathbf{X}}(x_2) - F_{\mathbf{X}}(x_1) \\ P(\mathbf{Y} = y_3) &= P(\mathbf{X} > x_2) = 1 - F_{\mathbf{X}}(x_2) \end{aligned} \quad (2.80)$$

3. La función $g(x)$ presenta una discontinuidad en el punto x_0 , donde $g(x_0^-) = y_1$ y $g(x_0^+) = y_2$ (figura 2.10c). En tal caso, dado que la VA \mathbf{Y} no puede tomar valores en el intervalo (y_1, y_2) (dado que no existen valores de la VA \mathbf{X} que se transformen en dicho intervalo) la función de distribución presentará un intervalo de constancia en el intervalo (y_1, y_2) , es decir, $F_{\mathbf{Y}}(y_2) = F_{\mathbf{Y}}(y_1)$. Un posible forma para esta función se muestra en la figura 2.11b).

2.6.3 Caso general. Obtención de la función de densidad: Teorema Fundamental

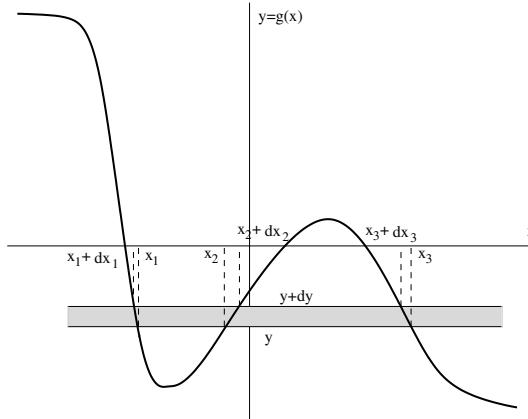


Figura 2.12: Cálculo de la función $f_Y(y)$.

Hemos visto que para el cálculo de la función de distribución $F_Y(y)$ no existe un procedimiento cerrado sino que el proceso es enteramente constructivo. Sin embargo, para el cálculo de la función de densidad sí que podemos hacer uso de una expresión cerrada que nos dé solución al problema. Tal expresión cerrada se conoce bajo el nombre *Teorema fundamental* y éste será nuestro foco de atención en el resto de la sección.

Consideremos la situación representada en la figura 2.12. Para hacer el cálculo de la función de densidad, consideremos un punto de la VA \mathbf{Y} y analicemos la probabilidad de que esta variable tome valores en un intervalo infinitesimal en torno a dicho punto. El objetivo es expresar esa probabilidad en función de sucesos equivalentes de la VA \mathbf{X} . En virtud de la gráfica, podemos escribir

$$\begin{aligned} P(y < \mathbf{Y} \leq y + dy) &= P(x_1 + dx_1 < \mathbf{X} \leq x_1) + P(x_2 < \mathbf{X} \leq x_2 + dx_2) + \\ &\quad P(x_3 + dx_3 < \mathbf{X} \leq x_3) \\ f_{\mathbf{Y}}(y)dy &= f_{\mathbf{X}}(x_1)|dx_1| + f_{\mathbf{X}}(x_2)|dx_2| + f_{\mathbf{X}}(x_3)|dx_3| \end{aligned} \tag{2.81}$$

La primera igualdad de la expresión anterior procede de la lectura directa de la gráfica. Como puede verse, en la gráfica hay zonas donde la pendiente es negativa (zonas de puntos x_1 y x_3) de forma que los diferenciales dx_1 y dx_3 se han tomado con signos negativos. En la segunda línea se ha igualado

la probabilidad de un intervalo diferencial al producto de la función de densidad por la longitud del intervalo elemental. Los diferenciales se han tomado en valor absoluto para que las contribuciones de cada punto de la VA **X** a la probabilidad completa sean todas ellas positivas.

Suponiendo un caso general en el que N abscisas se transformen en la ordenada y , podríamos escribir

$$\begin{aligned} f_{\mathbf{Y}}(y)dy &= \sum_{i=1}^{N(y)} f_{\mathbf{X}}(x_i)|dx_i| \\ f_{\mathbf{Y}}(y) &= \sum_{i=1}^{N(y)} f_{\mathbf{X}}(x_i) \frac{|dx_i|}{dy} = \sum_{i=1}^{N(y)} \frac{f_{\mathbf{X}}(x_i)}{\left|\frac{dy}{dx_i}\right|} \\ f_{\mathbf{Y}}(y) &= \sum_{i=1}^{N(y)} \frac{f_{\mathbf{X}}(x)}{\left|g'(x)\right|} \Big|_{x=x_i} \end{aligned} \quad (2.82)$$

constituyendo la última línea la expresión del Teorema Fundamental.

Repárese que este teorema plantea una expresión cerrada para encontrar la función de densidad. Sin embargo, el análisis que debe hacerse para obtener tal función es paralelo al realizado en el caso de la sección 2.6.2. El motivo es que, según dice la ecuación (2.82), el valor de la función $f_{\mathbf{Y}}(y)$, para cada punto y , procede de la superposición de $N(y)$ contribuciones de la VA **X**, siendo cada una de esas contribuciones las procedentes de los $N(y)$ puntos de solución de la ecuación $y=g(x)$, o si se prefiere, de los $N(y)$ puntos de corte de la recta horizontal de ordenada y con la función $y = g(x)$. No obstante, es cierto que este teorema proporciona una expresión muy útil y cómoda, la cual, de hecho, se empleará enormemente en la práctica. Debemos, no obstante, recalcar que debe aplicarse pensando *en qué punto* se está aplicando.

Adviértase, por último, que este teorema proporciona la manera de llegar a una expresión analítica de la función de densidad. No obstante, la aplicación de la misma requiere dos pasos adicionales:

1. Escribir la expresión directamente en función de la variable y . Por tanto, debe invertirse la función $y = g(x)$, teniendo que obtenerse una expresión— función de y —para cada x_i involucrado.
2. Deben indicarse explícitamente los rangos de validez, en los valores de y , de la expresión/expresiones resultantes de la aplicación de este teorema.

Los ejemplos siguientes ilustran cómo llevar a cabo la operativa completa.

Ejemplo 1: Considerando que $\mathbf{X} \sim U(-1/2, 1/2)$ y que $\mathbf{Y} = g(\mathbf{X}) = \mathbf{X}^2$, obtenga mediante el Teorema fundamental la expresión de $f_{\mathbf{Y}}(y)$.

Solución: La función de transformación $y = g(x) = x^2$ es una función par, tal que $\forall y > 0$ existen dos puntos del dominio origen que se transforman en un mismo punto destino. Para el caso $y = 0$ esto no es cierto, pero, como se sabe, la contribución de un punto aislado de una VA continua a la probabilidad es irrelevante, de forma que no prestaremos atención a este caso. Por otra parte, el valor máximo que puede tomar la VA \mathbf{Y} es igual a $1/4$, valor que procede de la transformación de los extremos del intervalo $-1/2 \leq x \leq 1/2$ en que la VA \mathbf{X} toma valores. Por tanto, las conclusiones que obtengamos en relación con la VA \mathbf{Y} serán válidas dentro del intervalo $(0, 1/4)$. Fuera de este intervalo sabemos, por simple inspección, que la función de densidad $f_{\mathbf{Y}}(y)$ es nula.

Por ello, podemos escribir

$$f_{\mathbf{Y}}(y) = \frac{f_{\mathbf{X}}(x)}{|g'(x)|} \Big|_{x=x_1} + \frac{f_{\mathbf{X}}(x)}{|g'(x)|} \Big|_{x=x_2}$$

La función de densidad $f_{\mathbf{X}}(x)$ es constante y de valor igual a la unidad en el intervalo $(-1/2, 1/2)$. Asimismo, dado que $g(x) = x^2$ entonces $g'(x) = 2x$. Entonces

$$f_{\mathbf{Y}}(y) = \frac{1}{|2x|} \Big|_{x=x_1} + \frac{1}{|2x|} \Big|_{x=x_2}$$

Para escribir las expresión anterior en función de la variable y , como dicta el primer miembro de la igualdad, basta ver que $y = x^2 \rightarrow x = \pm\sqrt{y}$. Las dos soluciones de las que hablábamos antes, son precisamente, las correspondientes a cada uno de los signos de la raíz. Como, según vemos en la expresión anterior, la función del denominador va acompañada de un valor absoluto, podemos escribir

$$f_{\mathbf{Y}}(y) = \frac{1}{|-2\sqrt{y}|} + \frac{1}{|2\sqrt{y}|} = \frac{2}{2\sqrt{y}} = \frac{1}{\sqrt{y}}$$

Así pues, concluimos que

$$f_{\mathbf{Y}}(y) = \begin{cases} \frac{1}{\sqrt{y}} & 0 < y \leq \frac{1}{4} \\ 0 & \text{resto} \end{cases}$$



Ejemplo 2: Considerando que $\mathbf{X} \sim U(-\pi, \pi)$ y que $\mathbf{Y} = g(\mathbf{X}) = a \sin(\mathbf{X} + \theta)$, con $a, \theta \in \mathcal{R}$, obtenga mediante el Teorema fundamental la expresión de $f_{\mathbf{Y}}(y)$.

Solución: En este caso, la función de transformación $y = g(x) = a \sin(x + \theta)$ es una sinusoide con fase inicial θ . El número de puntos de corte entre una recta horizontal de ordenada y y esta función sería infinito si la expresión se analizase $\forall x$. Sin embargo, como sólo nos interesa el intervalo donde la VA \mathbf{X} toma valores, y este intervalo es de anchura 2π , el número de puntos de corte queda reducido a 2 (salvo en algún punto aislado, por ejemplo, en los puntos de mínimo y máximo de esta curva). Por otra parte, los valores máximo y mínimo de la transformación son $\pm a$, de forma que sabemos de antemano que la función de densidad fuera del intervalo $(-a, a)$ es nula. Centrándonos en el interior del intervalo, podemos escribir

$$f_{\mathbf{Y}}(y) = \left. \frac{f_{\mathbf{X}}(x)}{|g'(x)|} \right|_{x=x_1} + \left. \frac{f_{\mathbf{X}}(x)}{|g'(x)|} \right|_{x=x_2}$$

La derivada de la función de transformación es $g'(x) = a \cos(x + \theta)$. Por ello, la expresión anterior se particulariza

$$f_{\mathbf{Y}}(y) = \frac{\frac{1}{2\pi}}{|a \cos(x_1 + \theta)|} + \frac{\frac{1}{2\pi}}{|a \cos(x_2 + \theta)|}$$

Para escribir la expresión en función de y tenemos que hacer uso de la función inversa. En este caso

$$\begin{aligned} y &= a \sin(x + \theta) \\ \frac{y}{a} &= \sin(x + \theta) \\ x &= \arcsin\left(\frac{y}{a}\right) - \theta \end{aligned} \tag{2.83}$$

de forma que

$$f_{\mathbf{Y}}(y) = \frac{\frac{1}{2\pi}}{|a \cos(\arcsin(\frac{y}{a}))|} \Big|_{x=x_1} + \frac{\frac{1}{2\pi}}{|a \cos(\arcsin(\frac{y}{a}))|} \Big|_{x=x_2}$$

Repárese, no obstante, que lo que realmente necesitamos saber es cómo escribir $a \cos(x + \theta)$ en función de y , para cada una de las dos soluciones x_1 y x_2 . Por el camino que hemos empleado hasta ahora las cosas no son obvias y no parece obvio tampoco cómo obtener de forma explícita las dos soluciones que hemos anticipado que existen. Debe haber otros caminos, los cuales, en general, suelen encontrarse haciendo uso combinado de la función $g(x)$ y de su derivada. En efecto,

$$\begin{aligned} y = g(x) &= a \sin(x + \theta) \\ g'(x) &= a \cos(x + \theta) = \pm a \sqrt{1 - \sin^2(x + \theta)} \\ &= \pm \sqrt{a^2 - (a \sin(x + \theta))^2} \\ g'(x) &= \pm \sqrt{a^2 - y^2} \end{aligned} \tag{2.84}$$

por lo que,

$$f_{\mathbf{Y}}(y) = \frac{\frac{1}{2\pi}}{|-\sqrt{a^2 - y^2}|} + \frac{\frac{1}{2\pi}}{|\sqrt{a^2 - y^2}|} = \frac{1}{\pi \sqrt{a^2 - y^2}}$$

de modo que, finalmente,

$$f_{\mathbf{Y}}(y) = \begin{cases} \frac{1}{\pi \sqrt{a^2 - y^2}} & |y| \leq a \\ 0 & \text{resto} \end{cases}$$



Sobre los casos particulares

En la sección 2.6.2 hemos visto varios casos particulares a tener en cuenta. Tales casos particulares se reflejan de manera directa en las funciones de densidad, de forma que aquí también tenemos que prestar atención a tales situaciones. Estos casos son:

- Si la transformación es la mostrada en la figura 2.10a), se aplicará el teorema fundamental en las regiones $x \leq x_1$ y $x > x_2$. A continuación se añadirá a la función $f_{\mathbf{Y}}(y)$ el término $p_{y_1} \delta(y - y_1)$ (véase sección 2.6.2).

- Si la transformación es la mostrada en la figura 2.10b) no se aplicará el teorema, pues sabemos, por simple inspección, que

$$f_{\mathbf{Y}}(y) = p_1\delta(y - y_1) + p_2\delta(y - y_2) + p_3\delta(y - y_3)$$

con los valores p_i que se calcularon en la sección 2.6.2.

- Si la transformación es la mostrada en la figura 2.10c) se aplicará el teorema fundamental, y se obtendrán resultados para $y \leq y_1$ e $y > y_2$. En el interior del intervalo (y_1, y_2) la función $f_{\mathbf{Y}}(y)$ es nula.
- Si \mathbf{X} es discreta, \mathbf{Y} también lo será, luego no ha lugar aplicar el teorema fundamental, sino que se seguirá el método propuesto en la sección 2.6.1.
- Si la transformación es no derivable en un número discreto de puntos, tales puntos serán ignorados. Como es sabido, puntos aislados en una VA continua no tienen relevancia alguna en cuanto a cálculo de probabilidades. Tal sería el caso, por ejemplo, de la transformación $g(x) = |x|$. Esta función es no derivable en $x = 0$, pero puede aplicarse el teorema $\forall x \neq 0$.

2.7 Caracterización parcial de una función de variable aleatoria

En la sección anterior se ha visto el procedimiento mediante el cual se puede obtener la caracterización probabilística de una función de VA, conocida la función de transformación así como la función de densidad de la variable origen de la transformación.

Sin embargo, en múltiples ocasiones puede que no sea necesario conocer la caracterización probabilística (digamos, total) de la VA destino de la transformación, sino que baste con conocer algunos parámetros de la distribución de dicha variable. Nos referimos, típicamente, a la media, la varianza y/o el VCM. Tales parámetros, y otros similares, se conocen de forma genérica como *caracterización parcial de la VA*, en contraposición a la caracterización total, o conocimiento de la función de densidad.

En resumen, el objetivo de esta sección se puede enunciar como sigue: dada \mathbf{X} , con función de densidad conocida $f_{\mathbf{X}}(x)$, y dada una función de transformación $\mathbf{Y} = g(\mathbf{X})$, obténganse *parámetros* de caracterización parcial de la VA \mathbf{Y} sin llevar a cabo el cálculo explícito de la función de densidad $f_{\mathbf{Y}}(y)$.

Planteemos la operativa para el caso de la media. Si varios valores x_i de la VA \mathbf{X} se transforman en el valor y de la \mathbf{Y} , podemos escribir:

$$\begin{aligned} P(y < \mathbf{Y} \leq y + dy) &= \sum_i P(x_i < \mathbf{X} \leq x_i + dx_i) \\ f_{\mathbf{Y}}(y)dy &= \sum_i f_{\mathbf{X}}(x_i)dx_i \end{aligned} \quad (2.85)$$

donde x_i denota las diferentes particularizaciones de la abscisa x . En este caso, en contraposición a la operativa propuesta en la sección 2.6.3, tomamos los diferenciales directamente en sentido creciente de ambas variables, dado que no vamos a hacer uso de la derivada de la función $g(x)$, la cual relaciona ambos diferenciales.

Para obtener la expresión de la media de \mathbf{Y} debemos premultiplicar en (2.85) por los valores de y :

$$yf_{\mathbf{Y}}(y)dy = \sum_i g(x_i)f_{\mathbf{X}}(x_i)dx_i \quad (2.86)$$

Si ahora integramos ambos miembros, cada uno en su respectivas variables, vamos recorriendo el eje de los valores de la variable y y, como consecuencia, vamos recorriendo, por grupos de valores, el eje de la variable x . El resultado neto finalmente es:

$$E\{\mathbf{Y}\} = \int_{-\infty}^{\infty} yf_{\mathbf{Y}}(y)dy = \int_{-\infty}^{\infty} g(x)f_{\mathbf{X}}(x)dx = E\{g(\mathbf{X})\} \quad (2.87)$$

Nótese pues que si bien para realizar la primera de las integrales necesitamos conocer la función de densidad de la VA \mathbf{Y} , para realizar la segunda basta con conocer la función de transformación y la función de densidad de la VA \mathbf{X} , información conocida según hemos asumido desde el principio.

Este resultado generaliza a cualquier función de la VA \mathbf{Y} sin más que escribirla como una función de la variable \mathbf{X} . Por ejemplo, si se desea conocer la esperanza de una variable $\mathbf{Z} = h(\mathbf{Y})$, y sólo se dispone de información probabilística de la variable \mathbf{X} , variable origen de la transformación $\mathbf{Y} = g(\mathbf{X})$, escribiremos:

$$\begin{aligned} E\{\mathbf{Z}\} &= E\{h(\mathbf{Y})\} = \int_{-\infty}^{\infty} h(y)f_{\mathbf{Y}}(y)dy \\ &= E\{h(g(\mathbf{X}))\} = \int_{-\infty}^{\infty} h(g(x))f_{\mathbf{X}}(x)dx \end{aligned} \quad (2.88)$$

Algunas consecuencias inmediatas de lo dicho, son las siguientes:

- VCM de \mathbf{Y} :

$$\begin{aligned} E\{\mathbf{Y}^2\} &= \int_{-\infty}^{\infty} y^2 f_{\mathbf{Y}}(y) dy \\ &= E\{g^2(\mathbf{X})\} \\ &= \int_{-\infty}^{\infty} g^2(x) f_{\mathbf{X}}(x) dx \end{aligned} \quad (2.89)$$

- Varianza

$$\begin{aligned} \sigma_{\mathbf{Y}}^2 &= \int_{-\infty}^{\infty} (y - \eta_{\mathbf{Y}})^2 f_{\mathbf{Y}}(y) dy \\ &= E\{(g(\mathbf{X}) - E\{g(\mathbf{X})\})^2\} \\ &= \int_{-\infty}^{\infty} (g(x) - E\{g(\mathbf{X})\})^2 f_{\mathbf{X}}(x) dx \end{aligned} \quad (2.90)$$

- Linealidad del operador esperanza: Si $\mathbf{Y} = a\mathbf{X} + b$, entonces

$$\begin{aligned} E\{\mathbf{Y}\} &= E\{a\mathbf{X} + b\} = \int_{-\infty}^{\infty} (ax + b) f_{\mathbf{X}}(x) dx \\ &= a \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x) dx + b \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) dx \\ &= aE\{\mathbf{X}\} + b \end{aligned} \quad (2.91)$$

- La varianza de una VA no aumenta si se añade a ésta una constante, y varía con el cuadrado del coeficiente que multiplica a la variable. En efecto, si consideramos que $\mathbf{Y} = a\mathbf{X} + b$, entonces, haciendo uso de la media calculada anteriormente, podemos escribir

$$\begin{aligned} \sigma_{\mathbf{Y}}^2 &= \int_{-\infty}^{\infty} (g(x) - E\{g(\mathbf{X})\})^2 f_{\mathbf{X}}(x) dx \\ &= \int_{-\infty}^{\infty} (ax + b - (aE\{\mathbf{X}\} + b))^2 f_{\mathbf{X}}(x) dx \\ &= a^2 \int_{-\infty}^{\infty} (x - E\{\mathbf{X}\})^2 f_{\mathbf{X}}(x) dx \\ &= a^2 \sigma_{\mathbf{X}}^2 \end{aligned} \quad (2.92)$$

2.7.1 Momentos de una variable aleatoria

Los momentos de una VA son una generalización de algunos de los parámetros que hemos visto hasta ahora. Concretamente, consideremos una VA \mathbf{X} con función de densidad $f_{\mathbf{X}}(x)$. Se definen:

- Momento no central de orden $n \in \mathcal{Z}^+$:

$$m_n = E\{\mathbf{X}^n\} = \int_{-\infty}^{\infty} x^n f_{\mathbf{X}}(x) dx$$

Naturalmente $m_1 = E\{\mathbf{X}\}$ y $m_2 = E\{\mathbf{X}^2\}$. Los momentos m_n son por tanto una generalización de la media y del VCM. El caso trivial es $m_0 = 1$, pues coincide con el área bajo una función de densidad.

- Momento central de orden $n \in \mathcal{Z}^+$:

$$\mu_n = E\{(\mathbf{X} - \eta_{\mathbf{X}})^n\} = \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})^n f_{\mathbf{X}}(x) dx$$

En este caso es fácil ver que $\mu_2 = \sigma_{\mathbf{X}}^2$. Tales momentos son por tanto una generalización de la varianza. Asimismo, puede comprobarse que $\mu_0 = 1$ y $\mu_1 = 0$.

Para el caso de las VAs discretas las expresiones anteriores pasarían a ser sumatorios como ya sucedió en el caso de las expresiones (2.41) y (2.50). En particular, por ejemplo, el momento central anteriormente definido se escribiría

$$\mu_n = E\{(\mathbf{X} - \eta_{\mathbf{X}})^n\} = \sum_i (x_i - \eta_{\mathbf{X}})^n P(\mathbf{X} = x_i) \quad (2.93)$$

Es interesante mencionar dos ideas adicionales:

1. Existen distribuciones que se caracterizan por determinadas propiedades en los momentos. Hablamos, en particular, de la VA gaussiana. Puede comprobarse que los momentos centrales de orden impar $n = 2m + 1, m \geq 0$, para esta distribución son todos nulos. Piénsese en ello.
2. Existen relaciones entre momentos centrales y no centrales. Concretamente, haciendo uso del desarrollo de un binomio de Newton, podemos escribir:

$$\begin{aligned} \mu_n &= f(m_0, m_1, \dots, m_n) = E\{(\mathbf{X} - m_1)^n\} \\ &= E \left\{ \sum_{k=0}^n \binom{n}{k} \mathbf{X}^k (-m_1)^{(n-k)} \right\} \\ &= \sum_{k=0}^n \binom{n}{k} E \left\{ \mathbf{X}^k (-m_1)^{(n-k)} \right\} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k=0}^n \binom{n}{k} (-m_1)^{(n-k)} E\{\mathbf{X}^k\} \\
 &= \sum_{k=0}^n \binom{n}{k} (-m_1)^{(n-k)} m_k
 \end{aligned} \tag{2.94}$$

2.8 Funciones que generan momentos

Bajo el término *funciones que generan momentos* se agrupan un conjunto de funciones a partir de las cuales, y mediante operaciones relativamente sencillas, se pueden obtener los momentos de la VA bajo análisis. Comenzamos nuestra exposición con la *función característica*, para pasar luego a funciones más generales.

2.8.1 Función característica

La función característica de una VA \mathbf{X} se define como la transformada de Fourier⁸ de la función de densidad $f_{\mathbf{X}}(x)$ de la variable. Así pues, podemos escribir

$$\phi_{\mathbf{X}}(\omega) = E\{e^{j\omega\mathbf{X}}\} = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) e^{j\omega x} dx \tag{2.95}$$

de forma que esta función puede interpretarse en los términos indicados, o bien como la esperanza de una función exponencial compleja definida sobre la VA \mathbf{X} .

Naturalmente, a partir de la función característica $\phi_{\mathbf{X}}(\omega)$ podemos obtener la función de densidad sin más que llevar a cabo la transformación inversa

$$f_{\mathbf{X}}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{\mathbf{X}}(\omega) e^{-j\omega x} d\omega \tag{2.96}$$

Como hemos dicho antes, esta función se engloba dentro de las funciones que generan momentos, pues éstos se obtienen a partir de la función de manera sencilla. Concretamente si derivamos de forma n-ésima la función característica

$$\begin{aligned}
 \phi_{\mathbf{X}}(\omega) &= \int_{-\infty}^{\infty} e^{j\omega x} f_{\mathbf{X}}(x) dx \\
 \frac{d\phi_{\mathbf{X}}(\omega)}{d\omega} &= \int_{-\infty}^{\infty} (jx) e^{j\omega x} f_{\mathbf{X}}(x) dx
 \end{aligned}$$

⁸La proyección sobre la exponencial compleja $e^{j\omega x}$ ó $e^{-j\omega x}$ no altera el sentido de la transformada. Sólo produce un cambio de signo.

$$\begin{aligned} \frac{d^2\phi_{\mathbf{X}}(\omega)}{d\omega^2} &= \int_{-\infty}^{\infty} (jx)^2 e^{j\omega x} f_{\mathbf{X}}(x) dx \\ &\vdots \\ \frac{d^n\phi_{\mathbf{X}}(\omega)}{d\omega^n} &= \int_{-\infty}^{\infty} (jx)^n e^{j\omega x} f_{\mathbf{X}}(x) dx \end{aligned} \quad (2.97)$$

Evaluando las anteriores expresiones en $\omega = 0$, tenemos

$$\begin{aligned} \phi_{\mathbf{X}}(\omega)|_{\omega=0} &= \phi_{\mathbf{X}}^{(0)}(\omega)|_{\omega=0} = 1 = j^0 m_0 \\ \left. \frac{d\phi_{\mathbf{X}}(\omega)}{d\omega} \right|_{\omega=0} &= \phi_{\mathbf{X}}^{(1)}(\omega)|_{\omega=0} = jm_1 \\ \left. \frac{d^2\phi_{\mathbf{X}}(\omega)}{d\omega^2} \right|_{\omega=0} &= \phi_{\mathbf{X}}^{(1)}(\omega)|_{\omega=0} = j^2 m_2 \\ &\vdots \\ \left. \frac{d^n\phi_{\mathbf{X}}(\omega)}{d\omega^n} \right|_{\omega=0} &= \phi_{\mathbf{X}}^{(n)}(\omega)|_{\omega=0} = j^n m_n \end{aligned} \quad (2.98)$$

La última línea indica pues la relación genérica que existen entre la derivada n -ésima de la función característica y el momento no central de orden n de la VA en cuestión. Por ello, en efecto, la función característica $\phi_{\mathbf{X}}(\omega)$ es una función que genera momentos.

Por otra parte, recordando el desarrollo en serie de Taylor de una función analítica, vemos que los momentos nos permiten sintetizar la función característica $\phi_{\mathbf{X}}(\omega)$ en cualquier punto ω , dado que podemos escribir

$$\phi_{\mathbf{X}}(\omega) = \sum_{n=0}^{\infty} \left. \frac{\phi_{\mathbf{X}}^{(n)}(\omega)}{n!} \right|_{\omega=0} \omega^n = \sum_{n=0}^{\infty} \frac{(j\omega)^n}{n!} m_n \quad (2.99)$$

Finalmente, la función característica puede ser un útil instrumento para hacer el cálculo de la función de densidad de una VA $\mathbf{Y} = g(\mathbf{X})$ a partir de la función de densidad $f_{\mathbf{X}}(x)$ sin hacer uso expreso del Teorema fundamental. El siguiente ejercicio ilustra el procedimiento.

Ejercicio 1: Considere que \mathbf{X} es una VA con función de densidad $f_{\mathbf{X}}(x)$ conocida. A partir de esta VA se genera una segunda VA mediante la transformación $\mathbf{Y} = g(\mathbf{X})$. Se pide que, mediante funciones características, desarrolle una metodología para la obtención de la función de densidad $f_{\mathbf{Y}}(y)$. Particularice para el caso en que $\mathbf{X} \sim N(0, \sigma)$ y $g(x) = ax^2$.

Solución: La función característica de la VA \mathbf{Y} se define:

$$\begin{aligned}\phi_{\mathbf{Y}}(\omega) &= E\{e^{j\omega \mathbf{Y}}\} = \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y) e^{j\omega y} dy \\ &= E\{e^{j\omega g(\mathbf{X})}\} = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) e^{j\omega g(x)} dx\end{aligned}$$

Por tanto, a partir de la segunda expresión, la cual es conocida, si podemos hacer cambios sobre ella para dar lugar a la primera, es decir, para llegar a una expresión del tipo

$$\phi_{\mathbf{Y}}(\omega) = \int_{-\infty}^{\infty} h(y) e^{j\omega y} dy \quad (2.100)$$

entonces podremos concluir que $h(y) = f_{\mathbf{Y}}(y)$.

Si concretamos para el caso que se plantea, es decir $\mathbf{Y} = g(\mathbf{X})$, es obvio que \mathbf{Y} va a tomar sólo valores no negativos, de forma que su función de densidad será nula $\forall y < 0$. Por ello, podemos escribir

$$\begin{aligned}\phi_{\mathbf{Y}}(\omega) &= \int_0^{\infty} f_{\mathbf{Y}}(y) e^{j\omega y} dy \\ &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) e^{j\omega g(x)} dx \\ &= \int_{-\infty}^{\infty} e^{j\omega a x^2} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx\end{aligned} \quad (2.101)$$

Para poder transformar la última ecuación en la primera debemos, en primer lugar, hacer que los intervalos de integración coincidan. Para ello podemos explotar el hecho de que la función subintegral es par, y el intervalo de integración es simétrico. Por ello, podemos escribir

$$\phi_{\mathbf{Y}}(\omega) = \frac{2}{\sigma \sqrt{2\pi}} \int_0^{\infty} e^{j\omega a x^2} e^{-\frac{x^2}{2\sigma^2}} dx \quad (2.102)$$

Si ahora hacemos el cambio

$$\begin{aligned}y &= ax^2 \\ dy &= 2axdx = 2\sqrt{ay}dx\end{aligned}$$

entonces

$$\begin{aligned}\phi_{\mathbf{Y}}(\omega) &= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty e^{j\omega y} e^{-\frac{y}{2a\sigma^2}} \frac{dy}{2\sqrt{ay}} \\ &= \frac{1}{\sigma\sqrt{2\pi ay}} \int_0^\infty e^{j\omega y} e^{-\frac{y}{2a\sigma^2}} dy\end{aligned}\quad (2.103)$$

de forma que podemos concluir que

$$f_{\mathbf{Y}}(y) = h(y) = \frac{1}{\sigma\sqrt{2\pi ay}} e^{-\frac{y}{2a\sigma^2}} u(y) \quad (2.104)$$



Ejercicio 2: Repítase el ejercicio anterior para el caso en que $\mathbf{X} \sim U(-\pi, \pi)$ y $\mathbf{Y} = g(\mathbf{X}) = a \sin(\mathbf{X})$.

Solución: En este caso \mathbf{Y} tomará valores en el intervalo $(-a, a)$, pues ése es el recorrido de la función de transformación $g(x)$. Por tanto, podemos escribir

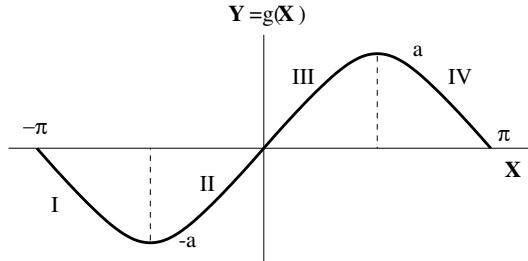
$$\begin{aligned}\phi_{\mathbf{Y}}(\omega) &= E\{e^{j\omega \mathbf{Y}}\} = \int_{-a}^a f_{\mathbf{Y}}(y) e^{j\omega y} dy \\ &= E\{e^{j\omega g(\mathbf{X})}\} = \int_{-\pi}^{\pi} f_{\mathbf{X}}(x) e^{j\omega g(x)} dx \\ &= \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{j\omega a \sin(x)} dx\end{aligned}$$

Dado que

$$\begin{aligned}y = g(x) &= a \sin(x) \\ dy &= a \cos(x) dx = \pm \sqrt{a^2 - (a \sin(x))^2} dx \\ dy &= \pm \sqrt{a^2 - y^2} dx\end{aligned}\quad (2.105)$$

La cuestión es que, en función de la zona de la curva donde nos encontremos, tendremos que emplear el signo de dy adecuado. La figura 2.13 ilustra el proceso: conforme avanzamos por el eje x desde $-\pi$ hasta π atravesamos:

- Una primera zona (zona I) donde la variable y varía en el intervalo $(0, -a)$, en este sentido, y donde los diferenciales

Figura 2.13: Función $\mathbf{Y} = \sin(\mathbf{X})$ con $\mathbf{X} \sim U(-\pi, \pi)$.

tienen signos contrarios (cuando x crece, y decrece). En esta zona, por tanto, la integral anterior pasa a ser

$$\int_0^{-a} \frac{1}{2\pi} e^{j\omega y} \left(-\frac{1}{\sqrt{a^2 - y^2}} \right) dy$$

- La zona II, donde y varía en el intervalo $(-a, 0)$ y los diferenciales tienen el mismo signo. Por ello

$$\int_{-a}^0 \frac{1}{2\pi} e^{j\omega y} \left(\frac{1}{\sqrt{a^2 - y^2}} \right) dy$$

- La zona III, donde podemos escribir, siguiendo la gráfica

$$\int_0^a \frac{1}{2\pi} e^{j\omega y} \left(\frac{1}{\sqrt{a^2 - y^2}} \right) dy$$

- Y, por último, la zona IV, donde, de nuevo, los diferenciales tienen signos contrarios.

$$\int_a^0 \frac{1}{2\pi} e^{j\omega y} \left(-\frac{1}{\sqrt{a^2 - y^2}} \right) dy$$

La solución que buscamos es por tanto la suma de las cuatro contribuciones, lo cual podemos escribir

$$\begin{aligned} \phi_{\mathbf{Y}}(\omega) &= \frac{1}{2\pi} \left[\int_0^{-a} e^{j\omega y} \left(-\frac{1}{\sqrt{a^2 - y^2}} \right) dy + \int_{-a}^0 e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy + \right. \\ &\quad \left. \int_0^a e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy + \int_a^0 e^{j\omega y} \left(-\frac{1}{\sqrt{a^2 - y^2}} \right) dy \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi} \left[\int_{-a}^0 e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy + \int_{-a}^0 e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy + \right. \\
&\quad \left. \int_0^a e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy + \int_0^a e^{j\omega y} \frac{1}{\sqrt{a^2 - y^2}} dy \right] \\
&= \frac{1}{2\pi} \left[2 \int_{-a}^0 \frac{1}{\sqrt{a^2 - y^2}} e^{j\omega y} dy + 2 \int_0^a \frac{1}{\sqrt{a^2 - y^2}} e^{j\omega y} dy \right] \\
&= \int_{-a}^a \frac{1}{\pi \sqrt{a^2 - y^2}} e^{j\omega y} dy
\end{aligned} \tag{2.106}$$

de forma que, identificando esta última expresión con la ecuación (2.95) podemos concluir que

$$f_{\mathbf{Y}}(y) = \begin{cases} \frac{1}{\pi \sqrt{a^2 - y^2}} & |y| \leq a \\ 0 & \text{resto} \end{cases}$$

resultado que coincide con el obtenido en el ejemplo 2 de la sección 2.6.3. Como puede verse, en el caso en que la transformación no sea inyectiva en el intervalo de integración empleado en la variable x para el cálculo de la función característica, el empleo del método expuesto para el cálculo de $f_{\mathbf{Y}}(y)$ resulta más laborioso.



Como comentario final debemos decir que si la VA \mathbf{X} es una VA discreta, la función característica se define de forma similar al caso continuo, es decir, mediante la esperanza de una función de VA, de la forma

$$\phi_{\mathbf{Y}}(\Omega) = E\{e^{j\Omega \mathbf{X}}\} = \sum_i p_i e^{j\Omega x_i} \tag{2.107}$$

En el caso particular en que $x_i = i$, $i \in \mathcal{Z}$ entonces esta función se obtendrá mediante la transformada de Fourier de la secuencia $y[i] = p_i$

$$\phi_{\mathbf{Y}}(\Omega) = E\{e^{j\Omega \mathbf{X}}\} = \sum_i p_i e^{j\Omega i} \tag{2.108}$$

transformación invertible, como es conocido, mediante la expresión

$$p_i = \frac{1}{2\pi} \int_{<2\pi>} \phi_{\mathbf{Y}}(\Omega) e^{-j\Omega i} d\Omega \tag{2.109}$$

donde el intervalo de integración en la expresión anterior es un intervalo cualquiera de longitud 2π , dado que la expresión (2.108) es periódica, de periodo 2π .

2.8.2 Función generadora de momentos

La función más general es la así llamada *función generadora de momentos*, la cual se define como la Transformada de Laplace de la función de densidad, es decir,

$$M_{\mathbf{X}}(s) = E\{e^{s\mathbf{X}}\} = \int_{-\infty}^{\infty} e^{sx} f_{\mathbf{X}}(x) dx \quad (2.110)$$

A partir de esta función es fácil comprobar que

$$m_n = \left. \frac{d^n M_{\mathbf{X}}(s)}{ds^n} \right|_{s=0} \quad (2.111)$$

A partir de ésta se define una segunda función $\Psi_{\mathbf{X}}(s) = \ln M_{\mathbf{X}}(s)$, cuyas derivadas dan lugar a los cumulantes de orden n

$$\lambda_n = \left. \frac{d^n \Psi_{\mathbf{X}}(s)}{ds^n} \right|_{s=0} \quad (2.112)$$

Es fácil comprobar que

$$\begin{cases} \lambda_1 = \eta_{\mathbf{X}} \\ \lambda_2 = \sigma_{\mathbf{X}}^2 \end{cases} \quad (2.113)$$

Sobre los cumulantes de orden n se construye la teoría de *estadísticos de orden superior* la cual ha tenido un moderado éxito en la comunidad científica para el manejo de señales no gaussianas.

Capítulo 3

Variable aleatoria bidimensional

3.1 Introducción

Para introducir el concepto de VA bidimensional, acudiremos a un ejemplo consistente en el juego del lanzamiento de dardos sobre una diana (véase figura 3.1). En este juego, como es bien sabido, se pretende que los dardos caigan tan cerca como sea posible del centro y que, además, éstos caigan en los sectores circulares marcados con una puntuación más alta. Si consideramos el origen de coordenadas localizado en el centro de la diana, las desviaciones con respecto a este punto pueden considerarse, en coordenadas cartesianas, mediante el par (x, y) , con x la desviación horizontal e y la desviación vertical. Cada lanzamiento puede considerarse la ejecución de un experimento aleatorio pues, naturalmente, el lanzador no sabe en qué posición va a caer el dado (por eso el juego tiene interés). Si consideramos un único lanzador y aceptamos que todos los lanzamientos son probabilísticamente idénticos, los diferentes lanzamientos pueden considerarse los diferentes valores que puede tomar el par (\mathbf{X}, \mathbf{Y}) . A priori, y cara a este ejemplo, posiblemente sea razonable pensar que los errores en el eje horizontal no tienen relación con los errores en el eje vertical. Pero esto no necesariamente tiene por qué ser así: los movimientos de hombro, codo y muñeca de un determinado individuo pueden relacionar los errores en una coordenada y en la otra. Asimismo, y considerando la estrategia en el juego, un jugador puede deliberadamente alejarse del centro de la diana en sentido horizontal si desea tener mayor tolerancia vertical para caer en el sector angular que deseé. En tal caso estaría acoplando deliberadamente

los errores en ambas dimensiones.

En este ejemplo, un jugador se consideraría mejor que otro si el primero, *habitualmente*, alcanzase mejores puntuaciones que el segundo. Esto es equivalente a afirmar, por ejemplo, que la *probabilidad* de que el primer jugador consiga mayor puntuación que el segundo es elevada. Tal afirmación se podrá cuantificar si se dispone de las herramientas y de la información probabilística oportunas. Ambas entidades son, precisamente, el cometido de este capítulo.

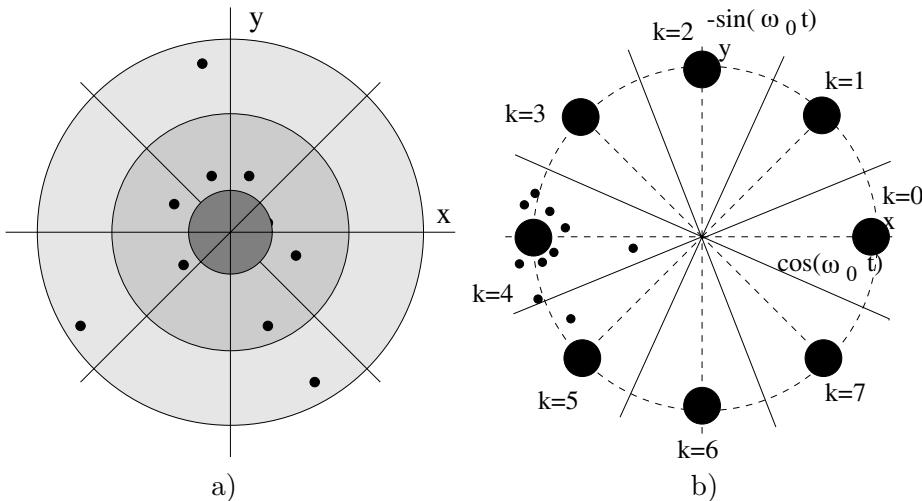


Figura 3.1: a) Lanzamientos de dardos sobre una diana. Pueden considerarse realizaciones de la VA bidimensional “errores con respecto a $(0, 0)$ ”. b) Modulación digital en fase, con nubes de puntos alrededor de uno de los valores esperados.

Este ejemplo puede resultar ingenuo pero no está lejos de las Tecnologías de la Información. Concretamente, una determinada modulación digital consiste en el envío, recepción y discriminación de señales de la forma

$$s_k(t) = A \cos(\omega_0 t + \phi_k(t)) \quad (3.1)$$

donde la información de qué símbolo digital se está enviando recae, precisamente, en la fase $\phi_k(t)$. Si consideramos que durante un determinado tiempo de símbolo $0 \leq t \leq T$ se envía una de las M posibles fases

$$\phi_k(t) = \frac{2\pi k}{M} \quad 0 \leq k \leq M - 1$$

los posibles símbolos a recibir serían los indicados en la figura 3.1b)¹. Un modelo más real de señal recibida sería $x(t) = s_k(t) + n(t)$, con $n(t)$ una señal de ruido superpuesta que distorsiona las fases de los símbolos recibidos, de forma que, según se indica en la figura 3.1b) para el símbolo de índice $k = 4$, los símbolos recibidos fluctuarían respecto a los símbolos esperados. Así pues se hace necesario establecer protocolos de decisión que minimicen la probabilidad de error, definida ésta como la probabilidad de decidir que se ha enviado un símbolo cuando en realidad se ha enviado otro. Tales protocolos se traducen en regiones de decisión asociadas a cada símbolo esperado y, para el modelo de ruido típicamente empleado en comunicaciones, tales regiones coinciden con el modelo de sectores circulares del juego de la diana (véase en la figura 3.1b que a cada símbolo esperado se le ha asociado una región, marcada por líneas continuas, de forma que se decide como símbolo enviado aquel símbolo esperado en cuya región asociada ha caído la observación). Además, debe disponerse de una herramienta que permita el cálculo exacto de la probabilidad de error o, dicho de otra manera, el cálculo de la separación necesaria entre los símbolos (lo cual será controlado por la amplitud A de la sinusoida de la ecuación 3.1) de forma que la calidad de las comunicaciones pueda ser garantizada por el diseñador. Tal cálculo, como el lector puede imaginar, requiere del análisis conjunto de dos VAs (los errores en cada una de las direcciones de la figura), metodología que se analizará en este capítulo.

Dicho esto, debemos decir pues qué es una VA bidimensional. Pues bien, tal VA no es otra cosa que el par (\mathbf{X}, \mathbf{Y}) , lo cual constituye una entidad formada por **tres** subentidades, a saber, la VA \mathbf{X} , la VA \mathbf{Y} , y su comportamiento conjunto, esto es, su interacción. La primera variable procede del experimento aleatorio $\epsilon_1 < S_1, \mathcal{F}_1, P_1 >$; la VA \mathbf{Y} procederá de un segundo experimento $\epsilon_2 < S_2, \mathcal{F}_2, P_2 >$; el comportamiento conjunto constituye el experimento compuesto ϵ_c (véase sección 1.7), con espacio muestral $S_c = S_1 \times S_2$, clase de sucesos \mathcal{F}_c , función de \mathcal{F}_1 y \mathcal{F}_2 , pero con ley de asignación de probabilidades P_c no anticipable en base al conocimiento exclusivo de P_1 y P_2 . Esto refuerza la idea de la tercera entidad presente en una VA bidimensional: la propia composición de las dos VAs unidimensionales. Habrá casos en los que el comportamiento conjunto sea

¹La señal $s(t) = A \cos(\omega_0 t + \phi(t))$ puede expresarse de forma equivalente mediante $s(t) = A \cos(\omega_0 t) \cos(\phi(t)) - A \sin(\omega_0 t) \sin(\phi(t))$. Escogiendo oportunamente el valor de ω_0 para un T dado, las funciones $\cos(\omega_0 t)$ y $-\sin(\omega_0 t)$ son ortogonales, es decir, $\int_0^T \cos(\omega_0 t) \sin(\omega_0 t) dt = 0$, por lo que tales funciones pueden constituir un par de ejes coordenados cartesianos sobre los que representar los símbolos $(A \cos(\phi(t)), A \sin(\phi(t)))$ a enviar.

inexistente, de forma que la VA bidimensional se reducirá a dos VA unidimensionales. Pero en general no será así, de forma que, como hemos dicho, serán tres las entidades a considerar en nuestro análisis.

3.2 Caracterización de variable aleatoria bidimensional

3.2.1 Función de distribución conjunta

De forma análoga al caso de la VA unidimensional, la VA bidimensional debe caracterizarse completamente para poder hacer cualquier cálculo probabilístico sobre ella. Tal caracterización viene dada por la extensión bidimensional de las funciones que conocemos hasta ahora. Comenzaremos, en primer lugar, con la función de distribución conjunta o bidimensional. Tal función se define, como en el capítulo anterior, a través de una probabilidad. Lo que ahora cambia es el suceso sobre el que ésta se define. Concretamente:

$$F_{\mathbf{X}\mathbf{Y}}(x, y) = P(\mathbf{X} \leq x \cap \mathbf{Y} \leq y) \quad (3.2)$$

es decir, esta función mide la probabilidad de que, simultáneamente, la

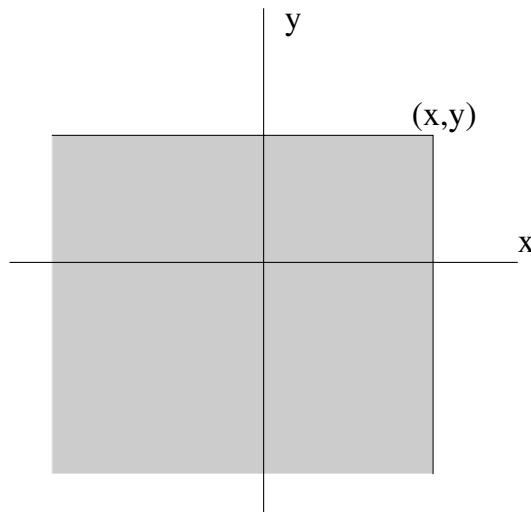


Figura 3.2: Suceso $\{\mathbf{X} \leq x\} \cap \{\mathbf{Y} \leq y\}$.

primera componente de la VA bidimensional sea menor o igual que un

punto de abscisa x , y la segunda lo sea de un punto de ordenada y . Tal suceso² se ha representado en la figura 3.2.

Esta función, por medir una probabilidad, no puede tener una forma de variación arbitraria. Mostramos a continuación algunas propiedades y formas de uso de la misma; antes de ello indicaremos, a nivel de notación, que por simplicidad en la escritura la definición (3.2) se escribirá habitualmente de la forma $F_{\mathbf{XY}}(x, y) = P(\mathbf{X} \leq x, \mathbf{Y} \leq y)$.

Propiedades y uso de la función de distribución

1. Comportamientos en $x = -\infty$ e $y = -\infty$ Podemos escribir que³:

$$F_{\mathbf{XY}}(-\infty, y) = F_{\mathbf{XY}}(x, -\infty) = F_{\mathbf{XY}}(-\infty, -\infty) = 0$$

Esto es debido a que los sucesos $\{\mathbf{X} = -\infty\}$ e $\{\mathbf{Y} = -\infty\}$ son sucesos de probabilidad nula, de forma que la probabilidad de la intersección de ellos con cualquier otro suceso es, asimismo, nula.

2. Comportamiento en $(x, y) = (\infty, \infty)$. En este caso:

$$F_{\mathbf{XY}}(\infty, \infty) = 1$$

puesto que se está calculando la probabilidad del suceso seguro.

3. La función de distribución conjunta es una función de probabilidad acumulada y, por ello, es no decreciente. En efecto, si consideramos que $A = \{\mathbf{X} \leq x_0, \mathbf{Y} \leq y_0\}$, $B = \{\mathbf{X} \leq x_1, \mathbf{Y} \leq y_1\}$, y que $x_0 \leq x_1$ e $y_0 \leq y_1$, podemos escribir acorde con la figura 3.3a)

$$\begin{aligned} B &= A \cup C, A \cap C = \emptyset \rightarrow \\ P(B) &= P(A \cup C) = P(A) + P(C) \\ F_{\mathbf{XY}}(x_1, y_1) &= F_{\mathbf{XY}}(x_0, y_0) + P(C) \\ F_{\mathbf{XY}}(x_1, y_1) &\geq F_{\mathbf{XY}}(x_0, y_0) \end{aligned} \tag{3.3}$$

pues, según el axioma primero de la probabilidad (página 22), $P(C) \geq 0$.

²Recuérdese que, en sentido estricto, el suceso mencionado debería escribirse $(\{\mathbf{X} \leq x\} \times S_2) \cap (S_1 \times \{\mathbf{Y} \leq y\})$, con S_1 y S_2 los espacios muestrales de los respectivos subexperimentos. Haremos uso, no obstante, de la notación simplificada indicada en la expresión 3.2.

³Una vez más, recuérdese el tratamiento que hacemos de los valores infinitos (véase sección 2.1).

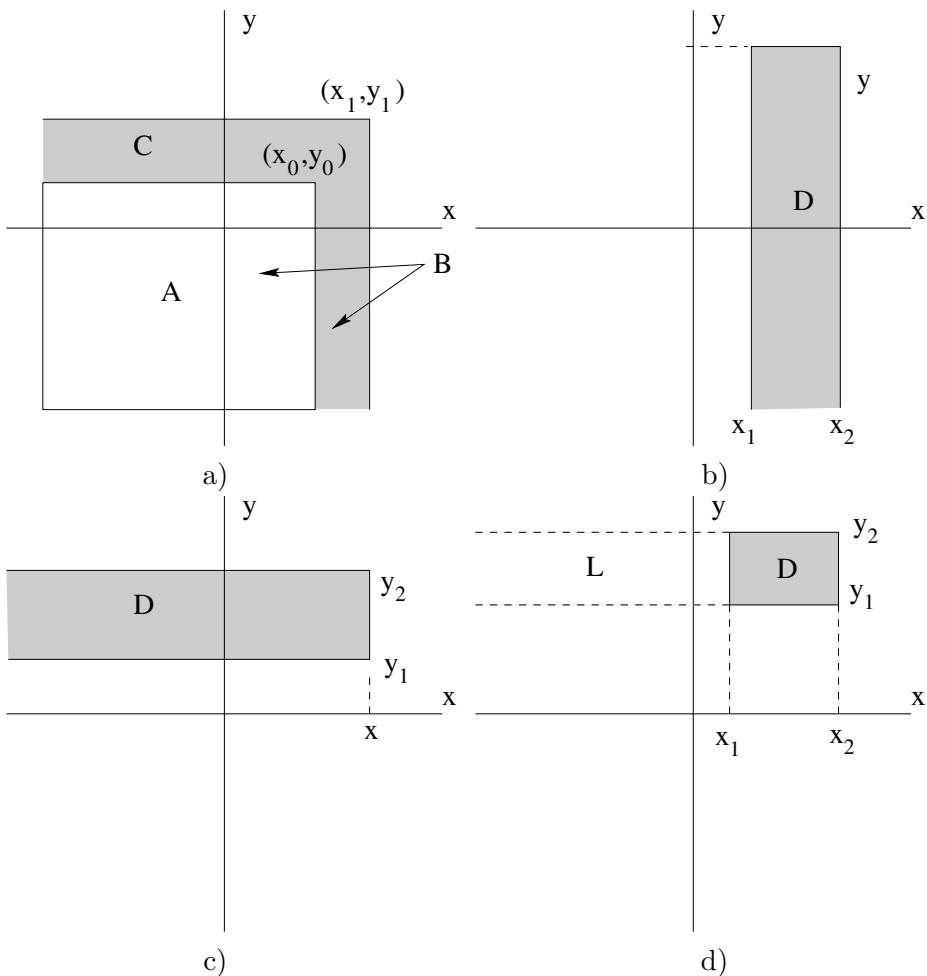


Figura 3.3: Casos de uso de la función $F_{\mathbf{X}\mathbf{Y}}(x, y)$.

4. La función de distribución permite calcular probabilidades de regiones con topología sencilla, concretamente, regiones con fronteras paralelas a los ejes coordinados. Si consideramos, en primer lugar, la situación representada en la figura 3.3b) podemos escribir

$$\begin{aligned}
 \{\mathbf{X} \leq x_2, \mathbf{Y} \leq y\} &= D \cup \{\mathbf{X} \leq x_1, \mathbf{Y} \leq y\} \\
 P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y) &= P(D \cup \{\mathbf{X} \leq x_1, \mathbf{Y} \leq y\}) \\
 &= P(D) + P(\mathbf{X} \leq x_1, \mathbf{Y} \leq y) \\
 P(D) &= P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y) - P(\mathbf{X} \leq x_1, \mathbf{Y} \leq y)
 \end{aligned}$$

$$= F_{\mathbf{XY}}(x_2, y) - F_{\mathbf{XY}}(x_1, y) \quad (3.4)$$

Por ello, la probabilidad asociada a una franja vertical limitada por los puntos (x_1, y) y (x_2, y) es igual a la diferencia de la función de distribución en dichos puntos.

5. Se puede repetir el razonamiento anterior para una franja horizontal. A partir de la situación representada en la figura 3.3c) podemos escribir

$$\begin{aligned} \{\mathbf{X} \leq x, \mathbf{Y} \leq y_2\} &= D \cup \{\mathbf{X} \leq x, \mathbf{Y} \leq y_1\} \\ P(\mathbf{X} \leq x, \mathbf{Y} \leq y_2) &= P(D \cup \{\mathbf{X} \leq x, \mathbf{Y} \leq y_1\}) \\ &= P(D) + P(\mathbf{X} \leq x, \mathbf{Y} \leq y_1) \\ P(D) &= P(\mathbf{X} \leq x, \mathbf{Y} \leq y_2) - P(\mathbf{X} \leq x, \mathbf{Y} \leq y_1) \\ &= F_{\mathbf{XY}}(x, y_2) - F_{\mathbf{XY}}(x, y_1) \end{aligned} \quad (3.5)$$

resultado análogo al obtenido anteriormente.

6. Esta función también permite hallar la probabilidad almacenada en el rectángulo representado en la figura 3.3d). Para ello, podemos escribir

$$\begin{aligned} \{\mathbf{X} \leq x_2, \mathbf{Y} \leq y_2\} &= D \cup \{\mathbf{X} \leq x_2, \mathbf{Y} \leq y_1\} \cup L \\ P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y_2) &= P(D \cup \{\mathbf{X} \leq x_2, \mathbf{Y} \leq y_1\} \cup L) \\ &= P(D) + P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y_1) + P(L) \\ P(D) &= P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y_2) - \\ &\quad P(\mathbf{X} \leq x_2, \mathbf{Y} \leq y_1) - P(L) \\ &= F_{\mathbf{XY}}(x_2, y_2) - F_{\mathbf{XY}}(x_2, y_1) - \\ &\quad (F_{\mathbf{XY}}(x_1, y_2) - F_{\mathbf{XY}}(x_1, y_1)) \\ &= F_{\mathbf{XY}}(x_2, y_2) - F_{\mathbf{XY}}(x_2, y_1) - \\ &\quad F_{\mathbf{XY}}(x_1, y_2) + F_{\mathbf{XY}}(x_1, y_1) \end{aligned} \quad (3.6)$$

Así pues, la probabilidad asociada a un rectángulo es igual a la suma algebraica de los valores de la función de distribución en las cuatro esquinas del mismo, con los signos dependiendo de la esquina de que se trate.

3.2.2 Función de densidad de probabilidad conjunta

La función de densidad de probabilidad conjunta $f_{\mathbf{XY}}(x, y)$ se define, de manera análoga al caso unidimensional, a partir de la derivada de la función

de distribución $F_{\mathbf{XY}}(x, y)$. En este caso la derivada debe ser con respecto a las dos variables, de forma que

$$f_{\mathbf{XY}}(x, y) = \frac{\partial^2 F_{\mathbf{XY}}(x, y)}{\partial x \partial y} \quad (3.7)$$

La relación inversa, naturalmente, es integral, con la expresión

$$F_{\mathbf{XY}}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{\mathbf{XY}}(\alpha, \beta) d\alpha d\beta \quad (3.8)$$

La ecuación (3.8) es equivalente al límite de un cociente incremental. Concretamente:

$$\begin{aligned} f_{\mathbf{XY}}(x, y) &= \lim_{\Delta x \rightarrow 0 \Delta y \rightarrow 0} \frac{F_{\mathbf{XY}}(x + \Delta x, y + \Delta y) - F_{\mathbf{XY}}(x, y + \Delta y)}{\Delta x \Delta y} \\ &\quad + \frac{-F_{\mathbf{XY}}(x + \Delta x, y) + F_{\mathbf{XY}}(x, y)}{\Delta x \Delta y} \\ &= \lim_{\Delta x \rightarrow 0 \Delta y \rightarrow 0} \frac{P(x < \mathbf{X} \leq x + \Delta x, y < \mathbf{Y} \leq y + \Delta y)}{\Delta x \Delta y} \quad (3.9) \end{aligned}$$

donde la segunda igualdad se ha escrito haciendo uso de la expresión (3.6). Puede verse, por tanto, que la función de densidad es equivalente a la probabilidad de que la VA bidimensional tome valores dentro de un rectángulo de lados diferenciales, situado en el punto (x, y) , dividido por el área de dicho rectángulo elemental. Es, consecuentemente, una probabilidad por unidad de área, de ahí su nombre de *densidad de probabilidad*.

Asimismo, dado que cualquier región D en el plano \mathcal{R}^2 puede aproximarse mediante una concatenación de rectángulos diferenciales, sea cual sea la forma de la frontera de dicha región, la probabilidad de que la VA (\mathbf{X}, \mathbf{Y}) tome valores en el interior de D es igual al volumen encerrado por la función de densidad en esta región, esto es,

$$P(D) = \int \int_D f_{\mathbf{XY}}(x, y) dx dy \quad (3.10)$$

Puede por ello apreciarse que la función de densidad es notablemente más versátil que la función de distribución, de forma que es ésta la función habitualmente empleada en la práctica.

Finalmente, debe hacerse dos comentarios adicionales:

1. La función de densidad $f_{\mathbf{XY}}(x, y)$ es la derivada de una función de distribución, la cual, por ser una función de probabilidad acumulada, es *no decreciente* (entendiéndose esto en los términos de la expresión 3.3); así pues, la función de densidad es no negativa, es decir, $f_{\mathbf{XY}}(x, y) \geq 0 \forall (x, y) \in R^2$.

2. El volumen encerrado bajo una función de densidad $f_{\mathbf{XY}}(x, y)$ es igual a uno. Esto es debido a que si evaluamos la expresión (3.8) en el punto (∞, ∞) , lo que evaluamos es la función de distribución en dicho punto, y sabemos que el valor de esta función en tal punto es unitario. Con ello se extiende a dos dimensiones el requisito equivalente que vimos para las funciones de densidad unidimensionales.

3.2.3 Funciones marginales

Las funciones vistas hasta ahora ($F_{\mathbf{XY}}(x, y)$ y $f_{\mathbf{XY}}(x, y)$) son funciones de caracterización probabilística total de la VA bidimensional (\mathbf{X}, \mathbf{Y}) . Por ello, estas funciones tienen toda la información necesaria para poder realizar cualquier cálculo probabilístico en relación con sucesos definidos sobre dicha variable. Ello implica que tales funciones tengan también toda la información probabilística sobre cada una de las dos componentes de la VA bidimensional, pues éstas, junto con la interacción entre ellas, configuran las tres entidades que componen la VA bidimensional. Por ello, a partir de las funciones conjuntas podremos hallar las funciones de caracterización de cada una de las VAs unidimensionales aisladas. A estas funciones, que son las vistas en el capítulo anterior, se les denomina en este ámbito *funciones de distribución/densidad marginales*.

Nótese que el recíproco no es, en general, cierto. Es decir, salvo excepciones, a partir del comportamiento individual (marginal) de las VAs unidimensionales no podremos conocer el comportamiento conjunto. La razón no es otra que lo dicho en el párrafo anterior: dado que la VA conjunta está formada por tres subentidades, y dado que las funciones marginales codifican el comportamiento de cada una de las VAs unidimensionales por separado, el comportamiento mutuo entre las VAs se ha perdido, y por ello no es posible restablecerlo a partir exclusivamente de la información marginal. Hablaremos de las excepciones a esta regla general en apartados posteriores.

El objetivo es, pues, obtener las funciones marginales a partir de las funciones conjuntas. Empecemos con las funciones de distribución. Para construir $F_{\mathbf{X}}(x)$ tenemos que hallar

$$F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x)$$

probabilidad de un suceso que podemos escribir, de forma equivalente,

$$\{\mathbf{X} \leq x\} = \{\mathbf{X} \leq x\} \cap \{\mathbf{Y} \leq \infty\}$$

pues el segundo suceso del segundo miembro es el suceso seguro. Por ello, podemos concluir que

$$F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x, \mathbf{Y} \leq \infty) = F_{\mathbf{XY}}(x, \infty)$$

De la misma manera

$$F_{\mathbf{Y}}(y) = P(\mathbf{X} \leq \infty, \mathbf{Y} \leq y) = F_{\mathbf{XY}}(\infty, y)$$

Repárese, por tanto, que la función de distribución es igual a la unidad cuando *ambas* VAs están evaluadas en el infinito. Si sólo evaluamos una de ellas en tal punto lo que hacemos es eliminar la restricción que impone dicha VA, de forma que estamos calculando la función de distribución marginal de la otra VA.

Con respecto a las funciones de densidad, hagamos el cálculo empleando el resultado anterior como puente entre $F_{\mathbf{XY}}(x, y)$ y $f_{\mathbf{XY}}(x, y)$. Para ello

$$\begin{aligned} f_{\mathbf{X}}(x) &= \frac{dF_{\mathbf{X}}(x)}{dx} \\ &= \frac{dF_{\mathbf{XY}}(x, \infty)}{dx} \\ &= \frac{d}{dx} \int_{-\infty}^x d\alpha \int_{-\infty}^{\infty} f_{\mathbf{XY}}(\alpha, y) dy \end{aligned} \quad (3.11)$$

Para realizar esta operación debemos derivar bajo el signo integral. Esto se puede realizar en base a la Regla de Leibnitz, la cual se formula

$$\begin{aligned} \frac{d\varphi(u)}{du} &= \frac{d}{du} \int_{\mu_i(u)}^{\mu_s(u)} G(\tau, u) d\tau \\ &= G(\mu_s(u), u) \frac{d\mu_s(u)}{du} - G(\mu_i(u), u) \frac{d\mu_i(u)}{du} + \int_{\mu_i(u)}^{\mu_s(u)} \frac{\partial G(\tau, u)}{\partial u} d\tau \end{aligned} \quad (3.12)$$

En nuestro caso la dependencia con la variable exterior de derivación es sólo a través del extremo superior de la integral en variable α . Por ello, si reescribimos la expresión como

$$f_{\mathbf{X}}(x) = \frac{d}{dx} \int_{-\infty}^x \phi(\alpha) d\alpha$$

con $\phi(\alpha) = \int_{-\infty}^{\infty} f_{\mathbf{XY}}(\alpha, y) dy$ entonces

$$f_{\mathbf{X}}(x) = \phi(x) \frac{dx}{dx} = \phi(x)$$

y por ello

$$f_{\mathbf{X}}(x) = \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x, y) dy \quad (3.13)$$

De forma similar

$$f_{\mathbf{Y}}(y) = \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x, y) dx \quad (3.14)$$

Por tanto, como hemos visto, la obtención de las funciones marginales de una VA a partir de la conjunta se reduce a eliminar la dependencia con respecto a la otra VA. Para tal fin, si hablamos de la función de distribución, ésta debe ser evaluada en el infinito de la otra VA. Si es la función de densidad, la conjunta debe ser integrada con respecto a la VA a eliminar.

Ejercicio: Supóngase una función de densidad $f_{\mathbf{XY}}(x, y)$ no nula en el interior de un recinto como el sombreado de la figura 3.8a). Se pide que obtenga el valor de la función $f_{\mathbf{X}}(x)$ para los puntos x_0 y x_1 .

Solución: Aplicamos la expresión (3.13), particularizada a esta situación. Por ello, los intervalos de integración se extienden sólo a la zona en la que la función $f_{\mathbf{XY}}(x, y)$ es no nula. Por ello

$$\begin{aligned} f_{\mathbf{X}}(x_0) &= \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x_0, y) dy = \int_{y_1}^{y_2} f_{\mathbf{XY}}(x_0, y) dy \\ f_{\mathbf{X}}(x_1) &= \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x_1, y) dy = \int_{y_3}^{y_4} f_{\mathbf{XY}}(x_1, y) dy \end{aligned}$$

Nótese por ello que, si bien genéricamente las expresiones (3.13) y (3.14) presentan intervalos de integración independientes de la otra VA (de x e y , respectivamente) en el caso frecuente en que la función de densidad sea no nula sólo en una cierta región del plano R^2 los límites de integración serán función del punto x e y donde se evalúan, respectivamente, $f_{\mathbf{X}}(x)$ y $f_{\mathbf{Y}}(y)$.



Ejercicio: Se desea conocer si $P(\mathbf{X} > x, \mathbf{Y} > y) = 1 - P(\mathbf{X} \leq x, \mathbf{Y} \leq y) = 1 - F_{\mathbf{XY}}(x, y)$ por analogía a su equivalente unidimensional $P(\mathbf{X} > x) = 1 - P(\mathbf{X} \leq x) = 1 - F_{\mathbf{X}}(x)$.

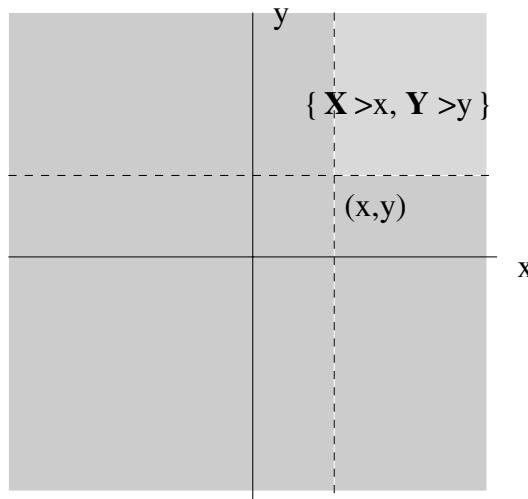


Figura 3.4: Suceso $\{\mathbf{X} > x\} \cap \{\mathbf{Y} > y\}$.

Solución: La igualdad anterior no es correcta. Según puede verse en la figura 3.4 el suceso complementario del suceso $\{\mathbf{X} > x\} \cap \{\mathbf{Y} > y\}$ es el suceso $\{\mathbf{X} \leq x\} \cup \{\mathbf{Y} \leq y\}$. Por ello.

$$\begin{aligned}
 P(\mathbf{X} > x, \mathbf{Y} > y) &= 1 - P(\{\mathbf{X} \leq x\} \cup \{\mathbf{Y} \leq y\}) \\
 &= 1 - (P(\mathbf{X} \leq x) + P(\mathbf{Y} \leq y)) - \\
 &\quad P(\mathbf{X} \leq x, \mathbf{Y} \leq y)) \\
 &= 1 - (F_{\mathbf{X}}(x) + F_{\mathbf{Y}}(y) - F_{\mathbf{XY}}(x, y)) \\
 &= 1 - (F_{\mathbf{XY}}(x, \infty) + F_{\mathbf{XY}}(\infty, y) - F_{\mathbf{XY}}(x, y))
 \end{aligned} \tag{3.15}$$



3.2.4 Casos particulares

Si bien lo dicho hasta ahora ha sido completamente general, es interesante tener en cuenta un conjunto de casos particulares de aparición relativamente frecuente en la práctica. Éstos son los siguientes:

X e Y discretas

En tal caso, la VA bidimensional (\mathbf{X}, \mathbf{Y}) no toma valores en todo el plano, o en partes del plano de forma continua, sino que sólo toma valores en un

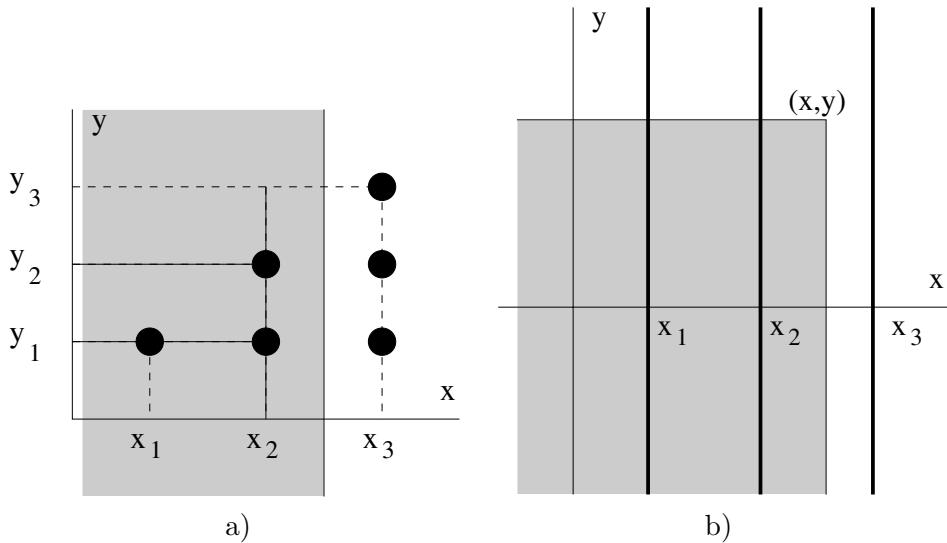


Figura 3.5: a) \mathbf{X} e \mathbf{Y} discretas b) \mathbf{X} discreta e \mathbf{Y} continua. En ambos casos \mathbf{X} toma sólo tres posibles valores.

conjunto discreto de puntos, esto es, en una rejilla de puntos. Un ejemplo de tal VA puede verse en la figura 3.5a).

En tal caso, de igual manera a su equivalente unidimensional, la VA viene caracterizada a partir del conjunto de valores (x_i, y_j) que puede tomar, junto con las probabilidades $p_{ij} = P(\mathbf{X} = x_i, \mathbf{Y} = y_j)$ con las que los toma. Los índices i y j variarán, en cada caso, en los intervalos adecuados.

A partir de estas probabilidades, podemos hacer cualquier cálculo probabilístico que nos interese. Por ejemplo, escogiendo un valor de x en el intervalo $x_2 < x < x_3$ (figura 3.5a), podemos escribir

$$P(\mathbf{X} \leq x) = P(\{\mathbf{X} = x_1\} \cup \{\mathbf{X} = x_2\}) = P(\mathbf{X} = x_1) + P(\mathbf{X} = x_2)$$

y dado que

$$\begin{aligned}\{\mathbf{X} = x_1\} &= \{\mathbf{X} = x_1, \mathbf{Y} = y_1\} \\ \{\mathbf{X} = x_2\} &= \{\mathbf{X} = x_2, \mathbf{Y} = y_1\} \cup \{\mathbf{X} = x_2, \mathbf{Y} = y_2\}\end{aligned}$$

entonces

$$P(\mathbf{X} = x_1) = P(\mathbf{X} = x_1, \mathbf{Y} = y_1) = p_{11}$$

$$P(\mathbf{X} = x_2) = P(\mathbf{X} = x_2, \mathbf{Y} = y_1) + P(\mathbf{X} = x_2, \mathbf{Y} = y_2) = p_{21} + p_{22}$$

de forma que

$$P(\mathbf{X} \leq x) = p_{11} + p_{21} + p_{22}$$

si $x_2 < x < x_3$.

X discreta e Y continua

En el caso en que las VAs sean discreta y continua, la VA bidimensional tomará valores sólo sobre rectas horizontales o verticales en función de quién de las VAs sea discreta. Para la selección hecha en el título del apartado, las rectas serán verticales, situadas en los valores de abscisa que tome la VA \mathbf{X} (véase figura 3.5b).

Para poder analizar una situación como ésta basta con que recordemos algunas expresiones introducidas en el capítulo anterior (véase sección 2.4). Tomando como base la figura 3.5b), calculemos $P(\mathbf{X} \leq x, \mathbf{Y} \leq y)$ para los valores de x e y seleccionados en la figura. Así pues

$$\begin{aligned} P(\mathbf{X} \leq x, \mathbf{Y} \leq y) &= P((\{\mathbf{X} = x_1\} \cup \{\mathbf{X} = x_2\}) \cap \{\mathbf{Y} \leq y\}) \\ &= P((\{\mathbf{X} = x_1\} \cap \{\mathbf{Y} \leq y\}) \cup (\{\mathbf{X} = x_2\} \cap \{\mathbf{Y} \leq y\})) \\ &= P(\mathbf{X} = x_1, \mathbf{Y} \leq y) + P(\mathbf{X} = x_2, \mathbf{Y} \leq y) \end{aligned}$$

Las expresiones anteriores son probabilidades de intersección de sucesos, de forma que nada nos impide escribirlas en términos de probabilidades condicionadas. Por ello

$$\begin{aligned} P(\mathbf{X} \leq x, \mathbf{Y} \leq y) &= P(\mathbf{X} = x_1, \mathbf{Y} \leq y) + P(\mathbf{X} = x_2, \mathbf{Y} \leq y) \\ &= P(\mathbf{Y} \leq y | \mathbf{X} = x_1)P(\mathbf{X} = x_1) + \\ &\quad P(\mathbf{Y} \leq y | \mathbf{X} = x_2)P(\mathbf{X} = x_2) \end{aligned}$$

Recordando ahora que (expresión 2.31)

$$f_{\mathbf{Y}}(y) = \sum_{i=1}^N f_{\mathbf{Y}}(y|A_i)P(A_i)$$

podemos considerar como elementos de la partición A_i cada uno de los valores x_i que puede tomar la VA \mathbf{X} . Por ello, y denominando, por brevedad $P(\mathbf{X} = x_i) = p_i$ y $f_{\mathbf{Y}}(y|\mathbf{X} = x_i) = f_{\mathbf{Y}}(y|x_i)$, podemos escribir

$$\begin{aligned} P(\mathbf{X} \leq x, \mathbf{Y} \leq y) &= P(\mathbf{Y} \leq y | \mathbf{X} = x_1)P(\mathbf{X} = x_1) + \\ &\quad P(\mathbf{Y} \leq y | \mathbf{X} = x_2)P(\mathbf{X} = x_2) \\ &= p_1 \int_{-\infty}^y f_{\mathbf{Y}}(\tau|x_1)d\tau + p_2 \int_{-\infty}^y f_{\mathbf{Y}}(\tau|x_2)d\tau \end{aligned} \tag{3.16}$$

Como puede verse, por lo tanto, para abordar el caso que nos ocupa, la información que debemos tener de la VA bidimensional se reduce a

1. Los valores $\mathbf{X} = x_i$ y las probabilidades $P(\mathbf{X} = x_i) = p_i \forall i$.
2. Las funciones de densidad $f_{\mathbf{Y}}(y|x_i)$ para los valores de x_i de la VA \mathbf{X} .

ya que esta información nos ha permitido obtener (3.16), función de distribución de la VA (\mathbf{X}, \mathbf{Y}) , es decir, la caracterización probabilística total de la VA bidimensional.

Ejercicio: Para la situación descrita en la figura 3.5b) y suponiendo conocida la información involucrada en la expresión (3.16), calcule $P(\mathbf{X} = x_i | \mathbf{Y} = y)$.

Solución: Recordando que

$$P(A|\mathbf{X} = x) = \frac{f_{\mathbf{X}}(x|A)P(A)}{f_{\mathbf{X}}(x)}$$

(expresiones 2.34 y 2.38), y haciendo que $A = \{\mathbf{X} = x_i\}$ y que la VA sea ahora \mathbf{Y} , podemos escribir

$$P(\mathbf{X} = x_i | \mathbf{Y} = y) = P(\mathbf{X} = x_i | y) = \frac{f_{\mathbf{Y}}(y|x_i)p_i}{\sum_{j=1}^3 f_{\mathbf{Y}}(y|x_j)p_j}$$



Componentes relacionadas mediante la función $\mathbf{Y} = g(\mathbf{X})$

Supongamos que las componentes \mathbf{X} e \mathbf{Y} de una VA bidimensional están relacionadas mediante la función $\mathbf{Y} = g(\mathbf{X})$. En tal caso la VA bidimensional sólo toma valores sobre la curva de transformación, y no en todo el plano; éste es un caso particular en el que se puede construir la función de distribución conjunta $F_{\mathbf{XY}}(x, y)$ en base a las marginales. En efecto, tomando como base la función dada en la figura 3.6a) podemos escribir

$$\begin{aligned} y > g(x) &\quad F_{\mathbf{XY}}(x, y) = P(\mathbf{X} \leq x, \mathbf{Y} \leq y) = P(\mathbf{X} \leq x) = F_{\mathbf{X}}(x) \\ y \leq g(x) &\quad F_{\mathbf{XY}}(x, y) = P(\mathbf{X} \leq x, \mathbf{Y} \leq y) = P(\mathbf{X} \leq g^{-1}(y)) = F_{\mathbf{X}}(g^{-1}(y)) \end{aligned}$$

Ejemplo: Supóngase que $\mathbf{Y} = g(\mathbf{X}) = 2\mathbf{X}$. Se pide que calcule:

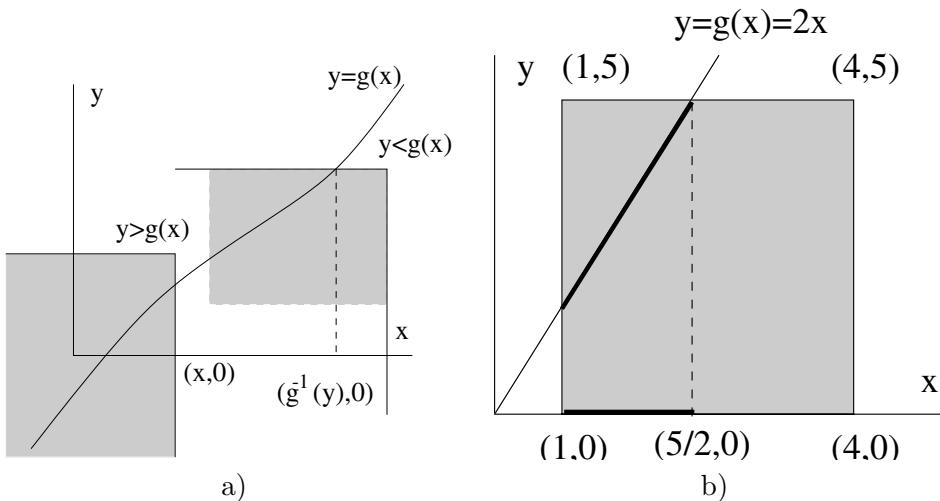


Figura 3.6: a) Componentes relacionadas mediante la función $\mathbf{Y} = g(\mathbf{X})$
b) $\mathbf{Y} = g(\mathbf{X}) = 2\mathbf{X}$.

1. $F_{\mathbf{XY}}(x, y)$
2. $P(1 < \mathbf{X} \leq 4, 0 < \mathbf{Y} \leq 5)$

Solución: A partir de la figura 3.6b) es fácil ver que:

1.

$$\begin{aligned} y > 2x &\quad F_{\mathbf{XY}}(x, y) = F_{\mathbf{X}}(x) \\ y \leq 2x &\quad F_{\mathbf{XY}}(x, y) = F_{\mathbf{X}}(g^{-1}(y)) = F_{\mathbf{X}}(y/2) \end{aligned}$$

2.

$$\begin{aligned} P(1 < \mathbf{X} \leq 4, 0 < \mathbf{Y} \leq 5) &= F_{\mathbf{XY}}(4, 5) - F_{\mathbf{XY}}(4, 0) - \\ &\quad F_{\mathbf{XY}}(1, 5) + F_{\mathbf{XY}}(1, 0) \\ &= F_{\mathbf{X}}(5/2) - F_{\mathbf{X}}(0/2) - \\ &\quad F_{\mathbf{X}}(1) + F_{\mathbf{X}}(0/2) \\ &= F_{\mathbf{X}}(5/2) - F_{\mathbf{X}}(1) \quad (3.17) \end{aligned}$$

resultado que se podría haber anticipado viendo los límites de variación de la VA \mathbf{X} dentro de la región $\{1 < \mathbf{X} \leq 4, 0 < \mathbf{Y} \leq 5\}$.



3.3 Funciones condicionadas

De igual manera que se vio en la sección 2.4 para el caso de la VA unidimensional, podemos considerar funciones de caracterización probabilística de una VA bidimensional cuando se sabe que un determinado suceso, digamos M , se ha verificado. Tales funciones, las cuales serán denominadas funciones condicionadas, indican cómo se actualiza nuestro conocimiento probabilístico sobre los valores que puede tomar la VA bidimensional, en las condiciones indicadas, esto es, cuando el suceso M se ha verificado.

Las funciones se definen

$$F_{\mathbf{XY}}(x, y|M) = \frac{P(\mathbf{X} \leq x, \mathbf{Y} \leq y, M)}{P(M)} \quad (3.18)$$

$$f_{\mathbf{XY}}(x, y|M) = \frac{\partial^2 F_{\mathbf{XY}}(x, y|M)}{\partial x \partial y} \quad (3.19)$$

La figura 3.7 muestra la situación planteada en la ecuación (3.18). El nu-

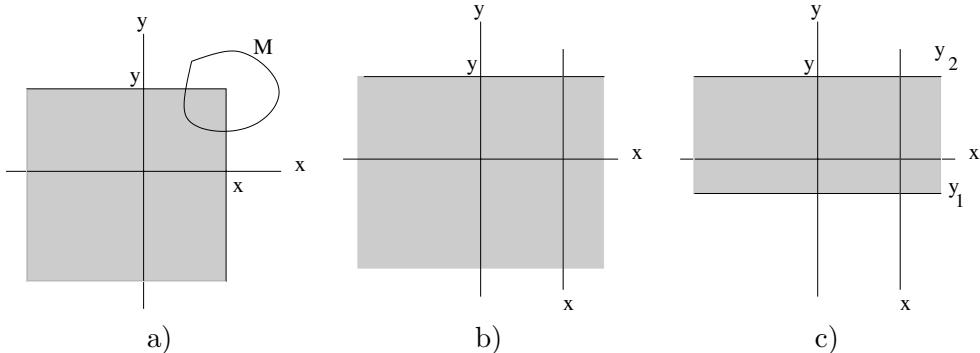


Figura 3.7: a) Esquema de definición de la función de distribución condicionada. b) Suceso $\{\mathbf{Y} \leq y\}$. c) Suceso $\{y_1 < \mathbf{Y} \leq y_2\}$.

merador de dicha expresión mide la probabilidad de que, simultáneamente, se verifique el suceso a partir del cual se define la función de distribución conjunta y el suceso M . El denominador es la normalización con respecto a la probabilidad del suceso que condiciona, por definición de la probabilidad condicionada (ecuación 1.20).

La segunda ecuación indica que la función de densidad condicionada se obtiene, como ya conocemos, derivando la correspondiente función de distribución.

3.3.1 Funciones marginales, condicionadas y conjuntas

La definición de funciones condicionadas nos permite encontrar interesantes relaciones entre funciones marginales condicionadas, funciones conjuntas, y funciones marginales incondicionales. Para ello, simplemente, debemos elegir oportunamente el suceso condicionante. Retomemos la funciones unidimensionales condicionadas y elijamos un condicionante función de la segunda variable. Concretamente:

1. $M = \{\mathbf{Y} \leq y\}$: El suceso corresponde a la zona sombreada de la figura 3.7b). A partir de él podemos escribir:

$$F_{\mathbf{X}}(x|M) = F_{\mathbf{X}}(x|\mathbf{Y} \leq y) = \frac{P(\mathbf{X} \leq x, \mathbf{Y} \leq y)}{P(\mathbf{Y} \leq y)} = \frac{F_{\mathbf{XY}}(x, y)}{F_{\mathbf{Y}}(y)}$$

de forma que despejando a favor de la función conjunta obtenemos

$$F_{\mathbf{XY}}(x, y) = F_{\mathbf{X}}(x|\mathbf{Y} \leq y)F_{\mathbf{Y}}(y) \quad (3.20)$$

Naturalmente, si hubiésemos hecho el razonamiento condicionando con la VA \mathbf{X} habríamos obtenido

$$F_{\mathbf{XY}}(x, y) = F_{\mathbf{Y}}(y|\mathbf{X} \leq x)F_{\mathbf{X}}(x) \quad (3.21)$$

Estas expresiones indican qué información se requiere para poder construir una función de distribución conjunta, esto es, para poder caracterizar probabilísticamente de forma total a una VA bidimensional. Tal información es el comportamiento aislado (marginal) de una de las dos variables, lo cual se encuentra codificado en la función de distribución marginal, así como el comportamiento de la segunda VA en relación con la primera, lo cual se encuentra en la función condicionada. Una lectura alternativa a esta expresión es que, en general, el conocimiento exclusivo de las dos funciones marginales no es suficiente para construir una función conjunta ya que tales funciones no tienen información alguna sobre la relación entre ambas. Tal información se encuentra en la función condicionada.

2. Podemos encontrar una segunda relación, en este caso entre funciones de densidad, escogiendo como suceso condicionante el suceso $M = \{y_1 < \mathbf{Y} \leq y_2\}$ (véase la figura 3.7c). Con esta elección podemos escribir

$$\begin{aligned} F_{\mathbf{X}}(x|M) &= F_{\mathbf{X}}(x|y_1 < \mathbf{Y} \leq y_2) = \frac{P(\mathbf{X} \leq x, y_1 < \mathbf{Y} \leq y_2)}{P(y_1 < \mathbf{Y} \leq y_2)} = \\ &= \frac{F_{\mathbf{XY}}(x, y_2) - F_{\mathbf{XY}}(x, y_1)}{F_{\mathbf{Y}}(y_2) - F_{\mathbf{Y}}(y_1)} \end{aligned}$$

El numerador de la anterior expresión procede de la probabilidad recogida en una lámina horizontal (expresión 3.5). El objetivo ahora es obtener una relación entre funciones de densidad, de forma que necesitaremos derivadas tanto con respecto a la variable x como con respecto a la variable y . Para derivar con respecto a y , repárese que en la expresión anterior tenemos el cociente entre dos incrementos de una función de y . Conseguiríamos derivadas construyendo un cociente de cocientes incrementales (esto es, dividiendo a ambos incrementos de función entre los correspondientes incrementos de la variable independiente). Para hacer esto renombremos los puntos $y_1 = y$ e $y_2 = y + \Delta y$ y hagamos que Δy tienda a ser cada vez más pequeño, esto es, que la lámina de probabilidad con la que estamos trabajando tienda a contener sólo al punto y . Segundo lo dicho

$$\begin{aligned} F_{\mathbf{X}}(x|y < \mathbf{Y} \leq y + \Delta y) &= \frac{F_{\mathbf{XY}}(x, y + \Delta y) - F_{\mathbf{XY}}(x, y)}{F_{\mathbf{Y}}(y + \Delta y) - F_{\mathbf{Y}}(y)} \\ &= \frac{\frac{F_{\mathbf{XY}}(x, y + \Delta y) - F_{\mathbf{XY}}(x, y)}{\Delta y}}{\frac{F_{\mathbf{Y}}(y + \Delta y) - F_{\mathbf{Y}}(y)}{\Delta y}} \\ \lim_{\Delta y \rightarrow 0} F_{\mathbf{X}}(x|y < \mathbf{Y} \leq y + \Delta y) &= F_{\mathbf{X}}(x|\mathbf{Y} = y) = \frac{\frac{\partial F_{\mathbf{XY}}(x, y)}{\partial y}}{\frac{dF_{\mathbf{Y}}(y)}{dy}} \end{aligned} \quad (3.22)$$

Con respecto a la última línea de la ecuación anterior debemos hacer el mismo comentario que hicimos al respecto de la última línea de la ecuación (2.34): $F_{\mathbf{X}}(x|\mathbf{Y} = y)$ debe entenderse como una forma compacta de escribir una expresión definida mediante un límite.

Derivando ahora con respecto a x llegamos a la expresión

$$f_{\mathbf{X}}(x|\mathbf{Y} = y) = \frac{\frac{\partial^2 F_{\mathbf{XY}}(x, y)}{\partial x \partial y}}{\frac{dF_{\mathbf{Y}}(y)}{dy}} = \frac{f_{\mathbf{XY}}(x, y)}{f_{\mathbf{Y}}(y)} \quad (3.23)$$

Razonando con respecto a la variable x se obtendría un resultado equivalente con los papeles de las variables intercambiados. Así pues, si despejamos a favor de la función de distribución conjunta (y escribiendo, por brevedad, $f_{\mathbf{X}}(x|\mathbf{Y} = y) = f_{\mathbf{X}}(x|y)$ y su equivalente en la otra variable), obtenemos

$$f_{\mathbf{XY}}(x, y) = f_{\mathbf{X}}(x|y)f_{\mathbf{Y}}(y) \quad (3.24)$$

$$= f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x) \quad (3.25)$$

expresiones muy similares a las obtenidas para el caso de funciones de distribución, y cuya interpretación es exactamente la misma que para tales funciones.

Comentarios adicionales:

- La función $f_{\mathbf{Y}}(y|x)$ es una función de densidad unidimensional como las manejadas en el capítulo anterior. La única novedad es que en ella existe la influencia de la otra variable, bien en la expresión funcional de la misma, bien en el intervalo de valores donde está definida. Por ejemplo, una posible concreción de esta función sería

$$f_{\mathbf{Y}}(y|x) = \frac{1}{\sigma(x)\sqrt{2\pi}} e^{-\frac{(y-\eta(x))^2}{2\sigma_x^2}} \quad (3.26)$$

lo cual se interpretaría como que la VA \mathbf{Y} , cuando la VA \mathbf{X} ha tomado el valor x , se comporta como una gaussiana de parámetros $\eta(x)$ y $\sigma(x)$. Repárese que con esta afirmación *sólo* conocemos cómo se distribuye la VA \mathbf{Y} cuando \mathbf{X} ha tomado una determinado valor, es decir, no conocemos cómo se comporta \mathbf{Y} aisladamente de \mathbf{X} . Para conocer esto, como veremos a continuación, necesitamos conocer la función de densidad de \mathbf{X} (expresión 3.27).

- La expresión (3.25) nos permite construir, mediante ordenador, muestras de VAs bidimensionales que tengan una determinada función de densidad conjunta. En efecto, esta expresión dice qué distribución tiene que tener la VA \mathbf{X} y qué distribución tiene que tener la VA \mathbf{Y} para cada uno de los valores que pueda tomar la VA \mathbf{X} , si queremos que la VA bidimensional (\mathbf{X}, \mathbf{Y}) tenga la función de densidad $f_{\mathbf{XY}}(x, y)$. Por ello, podemos construir muestras de tal VA bidimensional con el siguiente procedimiento:

1. Construir muestras de la VA \mathbf{X} distribuidas acorde con $f_{\mathbf{X}}(x)$.
2. Para cada valor x generado de la VA \mathbf{X} , constrúyanse muestras de la VA \mathbf{Y} con función de densidad condicionada a \mathbf{X} $f_{\mathbf{Y}}(y|x)$. Si, por ejemplo, tal función fuese la planteada anteriormente (ecuación 3.26), bastaría con que construyésemos VAs con distribución $N(\eta(x), \sigma(x))$.

3.3.2 Teoremas de la Probabilidad Total y de Bayes para partición continua

Las expresiones (3.24) y (3.25) nos permiten hacer una interpretación de las funciones marginales en términos del Teorema de la Probabilidad Total. En efecto, igualando (3.24) y (3.25) podemos escribir

$$f_{\mathbf{X}}(x|y)f_{\mathbf{Y}}(y) = f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)$$

Si ahora integramos con respecto, por ejemplo, a la variable x , obtendremos

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\mathbf{X}}(x|y)f_{\mathbf{Y}}(y)dx &= \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)dx \\ f_{\mathbf{Y}}(y) \int_{-\infty}^{\infty} f_{\mathbf{X}}(x|y)dx &= \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)dx \\ f_{\mathbf{Y}}(y) &= \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)dx \end{aligned} \quad (3.27)$$

pues el área bajo toda función de densidad (condicionada o incondicional) es igual a la unidad. Si tomamos esta expresión y multiplicamos por dy en ambos miembros de la igualdad (con el único objetivo de convertir las funciones densidad de probabilidad en probabilidades), tenemos

$$\begin{aligned} f_{\mathbf{Y}}(y)dy &= \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|x)dyf_{\mathbf{X}}(x)dx \\ P(y < \mathbf{Y} \leq y + dy) &= \int_{-\infty}^{\infty} P(y < \mathbf{Y} \leq y + dy|\mathbf{X} = x)P(x < \mathbf{X} \leq x + dx) \end{aligned} \quad (3.28)$$

lo cual es el equivalente continuo de la expresión

$$P(B) = \sum_{i=1}^N P(B|A_i)P(A_i)$$

donde se está empleando como elemento de la partición cada uno de los valores que puede tomar la VA \mathbf{X} .

Si hablamos del Teorema de la Probabilidad Total, podemos hablar del Teorema de Bayes sin más que intercambiar los condicionantes. Para ello, recordemos que

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

Haciendo uso, una vez más, de las igualdades (3.24) y (3.25) llegamos a la conclusión de que:

$$f_{\mathbf{X}}(x|y) = \frac{f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)}{f_{\mathbf{Y}}(y)} = \frac{f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)}{\int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|\alpha)f_{\mathbf{X}}(\alpha)d\alpha}$$

Finalmente, recordando las igualdades (3.13) y (3.14), y haciendo uso de las identidades (3.24) y (3.25) podemos escribir

$$\begin{aligned} f_{\mathbf{X}}(x) &= \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x,y)dy = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x|y)f_{\mathbf{Y}}(y)dy \\ f_{\mathbf{Y}}(y) &= \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x,y)dx = \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y|x)f_{\mathbf{X}}(x)dx \end{aligned}$$

de forma que es claro que el Teorema de la Probabilidad Total para dos VAs es una forma alternativa de obtención de las funciones de densidad marginales a partir, en su caso, de una función marginal condicionada y una marginal incondicional.

3.4 Independencia

En la introducción a este capítulo dijimos que la VA bidimensional (\mathbf{X}, \mathbf{Y}) se construye a partir de un experimento compuesto, $\epsilon_c < S_c, \mathcal{F}_c, P_c >$, el cual es el resultado de la composición de dos subexperimentos, a saber, $\epsilon_1 < S_1, \mathcal{F}_1, P_1 >$ y $\epsilon_2 < S_2, \mathcal{F}_2, P_2 >$. Allí se dijo que la ley para asignación de probabilidades no es, en general, determinable a partir de las leyes P_1 y P_2 de cada uno de los subexperimentos, sino que interviene una tercera entidad para la creación de tal ley, la cual es, precisamente, la propia composición de los experimentos.

Sin embargo, esto no es siempre así. Consideremos que los sucesos $A_{\mathbf{X}} \in \mathcal{F}_1$ y $B_{\mathbf{Y}} \in \mathcal{F}_2$. Si se verifica que⁴

$$P(A_{\mathbf{X}} \cap B_{\mathbf{Y}}) = P(A_{\mathbf{X}})P(B_{\mathbf{Y}})$$

entonces las VAs \mathbf{X} e \mathbf{Y} se dice que son independientes. En tal caso, la ley P_c sí puede obtenerse como función exclusiva de las leyes P_1 y P_2 . Repárese que, de hecho, cuando existe independencia entre variables la composición de los experimentos es meramente nominal, esto es, los experimentos se

⁴De forma coherente con la notación de la sección 1.7, la escritura estricta del suceso compuesto sería $(A_{\mathbf{X}} \times S_2) \cap (S_1 \times B_{\mathbf{Y}})$.

están considerando simultáneamente, pero en realidad no existe influencia alguna de uno sobre otro.

A nivel operativo, podemos comprobar la independencia entre variables verificando el comportamiento de las funciones de distribución o densidad conjunta y marginales. Para ello basta con elegir los sucesos $A_{\mathbf{X}}$ y $B_{\mathbf{Y}}$ de forma oportuna. Escogiendo $A_{\mathbf{X}} = \{\mathbf{X} \leq x\}$ y $B_{\mathbf{Y}} = \{\mathbf{Y} \leq y\}$ podemos afirmar que 2 VAs son independientes si

$$\begin{aligned} P(\mathbf{X} \leq x, \mathbf{Y} \leq y) &= P(\mathbf{X} \leq x)P(\mathbf{Y} \leq y) \\ F_{\mathbf{XY}}(x, y) &= F_{\mathbf{X}}(x)F_{\mathbf{Y}}(y) \end{aligned} \quad (3.29)$$

pues, al ser la función de distribución una función de caracterización probabilística total de la VA (\mathbf{X}, \mathbf{Y}) , la definición de independencia dada antes para sucesos genéricos $A_{\mathbf{X}}$ y $B_{\mathbf{Y}}$ se concreta en la expresión (3.29) con validez general.

Derivando esta expresión con respecto a ambas variables, tendremos un criterio equivalente de comprobación de independencia en términos de funciones de densidad. En efecto, dos VAs serán independientes si

$$f_{\mathbf{XY}}(x, y) = f_{\mathbf{X}}(x)f_{\mathbf{Y}}(y) \quad (3.30)$$

Es interesante contrastar la expresión obtenida con la vista en el caso general (ecuaciones 3.24 y 3.25). Para que las dos expresiones coincidan es necesario que las funciones marginales condicionadas coincidan con las incondicionales, es decir, $f_{\mathbf{X}}(x) = f_{\mathbf{X}}(x|y)$ y lo propio con respecto a la VA \mathbf{Y} . Pero esto, como vimos ya en el capítulo primero, es algo asociado a la independencia: los condicionantes no condicionan. Si es así, es decir, si las variables son independientes, el hecho de que la VA \mathbf{Y} haya tomado algún valor en particular no afecta en absoluto al comportamiento probabilístico de la VA \mathbf{X} .

Comentarios adicionales

- Lo dicho tiene validez general. No obstante, si la VA bidimensional es discreta, esto es, si cada componente de la misma es discreta, el criterio de independencia se particulariza en la factorización de la probabilidad de cada uno de los puntos del plano donde la variable toma valores. Concretamente, si la variable (\mathbf{X}, \mathbf{Y}) puede tomar los valores (x_i, y_j) con i y j variando en los recorridos oportunos, y con probabilidades $P(\mathbf{X} = x_i, \mathbf{Y} = y_j) = p_{ij}$, la VA

está formada por componentes independientes si se verifica que

$$P(\mathbf{X} = x_i, \mathbf{Y} = y_j) = p_{ij} = p_i^x p_j^y = P(\mathbf{X} = x_i)P(\mathbf{Y} = y_j) \quad (3.31)$$

- La comprobación de la independencia entre dos VAs exige comprobar cualquiera de las dos condiciones vistas (expresiones 3.29 y 3.30) para todos los valores de x e y . Sin embargo, la comprobación de la no independencia es más sencilla e intuitiva. Por ejemplo, si la función de densidad conjunta es no nula en el interior de la región sombreada de la figura 3.8a), el simple hecho de que la región R no presente lados paralelos a los ejes coordenados indica que las variables no son independientes. En efecto, si la variable \mathbf{X} toma el valor x_0 la figura indica que la VA \mathbf{Y} puede tomar valores en el intervalo (y_1, y_2) . Sin embargo, si \mathbf{X} toma el valor x_1 entonces \mathbf{Y} puede tomar valores en el intervalo (y_3, y_4) , intervalo distinto al anterior. Por tanto, el hecho de que \mathbf{X} tome un valor u otro afecta notablemente a los valores que puede tomar \mathbf{Y} , de forma que las variables no son independientes.

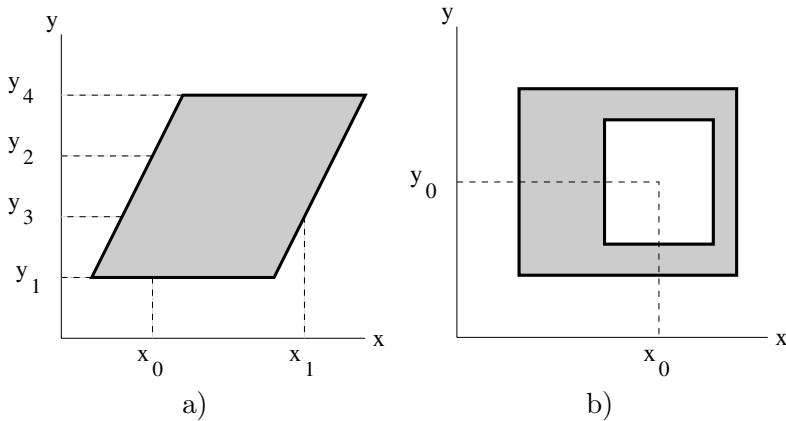


Figura 3.8: Dos casos de VAs no independientes.

- Similares comentarios pueden hacerse en relación con la figura 3.8b (supóngase no nula en el interior de la zona sombreada). En este caso, a pesar de que el dominio donde la VA bidimensional está definida presenta lado paralelos

a los ejes coordenados, es claro que la VA (\mathbf{X}, \mathbf{Y}) no puede tomar el valor (x_0, y_0) , pero que cada una de las componentes, de forma aislada, sí que pueden tomarlo; decimos esto porque la figura muestra que $f_{\mathbf{XY}}(x_0, y_0) = 0$, pero, sin embargo, $f_{\mathbf{X}}(x_0) \neq 0$ y $f_{\mathbf{Y}}(y_0) \neq 0$.

Ejercicio: Justifique por qué $f_{\mathbf{X}}(x_0) \neq 0$ y $f_{\mathbf{Y}}(y_0) \neq 0$ en el caso representado en la figura 3.8b).



3.5 Transformación de variable aleatoria bidimensional

Abordamos en este punto la extensión bidimensional de la transformación de VA vista en el capítulo anterior (sección 2.6). En este caso abordaremos en primer lugar el problema de una única transformación de dos VAs, digamos, $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$. Un segundo paso será abordar, mediante el mismo método que en el primero, el problema de dos funciones de dos VAs, es decir, $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$ y $\mathbf{W} = h(\mathbf{X}, \mathbf{Y})$. Sin embargo, para este segundo caso es más sencillo el uso del Teorema Fundamental, extendido a dos variables, el cual ya introdujimos, para el caso unidimensional, en el capítulo anterior. Las ventajas de tal teorema serán obvias tan pronto como se haga uso del mismo. Por ello, en el último punto, veremos cómo poder emplear el Teorema Fundamental para el caso de un única función de dos VAs. Dispondremos, consecuentemente, de caminos alternativos para abordar el mismo problema.

3.5.1 Una función de dos variables aleatorias

Consideremos la situación en la que la VA bidimensional (\mathbf{X}, \mathbf{Y}) , cuya caracterización probabilística asumiremos conocida, se transforma mediante la transformación $g(x, y)$ en una VA $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$. Asumiremos que la transformación es tal que se cumplen las condiciones para que \mathbf{Z} sea una VA (véase sección 2.1).

Con estas premisas, el objetivo es encontrar la caracterización probabilística de la VA \mathbf{Z} en base a la caracterización de la VA bidimensional (\mathbf{X}, \mathbf{Y}) y de la transformación $g(x, y)$. Para ello tendremos que calcular

$$F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) = P(D_z) = \int \int_{D_z} f_{\mathbf{X}, \mathbf{Y}}(x, y) dx dy \quad (3.32)$$

donde la región D_z se define como el lugar geométrico de los puntos del plano (x, y) que transformados mediante $g(x, y)$ dan lugar a valores menores o iguales que z . Formalmente:

$$D_z = \{(x, y) \in R^2 \text{ t.q. } g(x, y) \leq z\}$$

Por ello, el procedimiento para encontrar la caracterización probabilística de \mathbf{Z} consiste en dos fases:

1. Identificar la región D_z .
2. Calcular $P(D_z) = \int \int_{D_z} f_{\mathbf{X}\mathbf{Y}}(x, y) dx dy$

Ilustraremos este procedimiento mediante varios ejemplos:

1. Considérese que $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \mathbf{Y}$. El objetivo es encontrar, para esta transformación, la función de distribución de \mathbf{Z} como función de $f_{\mathbf{X}, \mathbf{Y}}(x, y)$. Para ello

$$F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) = P(\mathbf{X} + \mathbf{Y} \leq z) = P(D_z)$$

En este caso la región D_z es el lugar geométrico de los puntos del

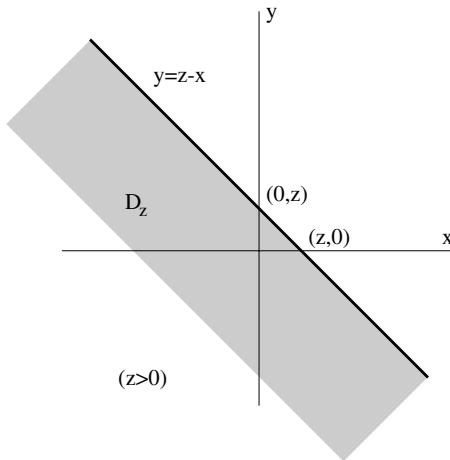


Figura 3.9: Representación de D_z para $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \mathbf{Y}$.

plano que verifican que $x + y \leq z$. Esta región tiene como frontera la ecuación $x + y = z$, la cual puede escribirse como $y = z - x$, es decir, una recta de pendiente -1 y ordenada en el origen igual a z

(véase figura 3.9 donde se indica la región D_z). A partir de la figura podemos escribir

$$F_{\mathbf{Z}}(z) = P(\mathbf{X} + \mathbf{Y} \leq z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f_{\mathbf{X}, \mathbf{Y}}(x, y) dy$$

Para hallar la función de densidad $f_{\mathbf{Z}}(z)$ tenemos que derivar con respecto a z la expresión anterior. Eso puede conseguirse fácilmente sin más que realizar un cambio de variable en la integral en y , con el objetivo de que el extremo superior de esta integral no dependa de x . Si hacemos $t = x + y$, podemos escribir

$$F_{\mathbf{Z}}(z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^z f_{\mathbf{X}, \mathbf{Y}}(x, t-x) dt = \int_{-\infty}^z \varphi(t) dt$$

con $\varphi(t) = \int_{-\infty}^{\infty} f_{\mathbf{X}, \mathbf{Y}}(x, t-x) dx$. Si derivamos la anterior expresión con respecto a z (recuérdese la regla de Leibnitz, expresión 3.12), tenemos

$$f_{\mathbf{Z}}(z) = \frac{dF_{\mathbf{Z}}(z)}{dz} = \varphi(z) \frac{dz}{dz} = \varphi(z) = \int_{-\infty}^{\infty} f_{\mathbf{X}, \mathbf{Y}}(x, z-x) dx$$

El resultado hasta ahora obtenido es completamente general. No obstante, si asumimos que las variables \mathbf{X} e \mathbf{Y} son independientes, entonces $f_{\mathbf{X}, \mathbf{Y}}(x, z-x) = f_{\mathbf{X}}(x)f_{\mathbf{Y}}(z-x)$, de forma que obtenemos

$$f_{\mathbf{Z}}(z) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x)f_{\mathbf{Y}}(z-x) dx = f_{\mathbf{X}}(z) * f_{\mathbf{Y}}(z)$$

con $*$ el operador convolución. A este resultado se le conoce como el *Teorema de la Convolución*, que dice, como hemos obtenido, que la *función de densidad de la suma de dos VAs independientes es igual a la convolución de sus funciones de densidad*.

Ejercicio: Sean \mathbf{X} e \mathbf{Y} dos VAs independientes, uniformes entre $(0, 1)$ y $(0, 2)$ respectivamente. Se pide la función de densidad de la variable $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$.

Solución: Según acabamos de ver la función de densidad es la convolución de las funciones de densidad. Por tanto, tenemos que realizar la operación

$$f_{\mathbf{Z}}(z) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\tau)f_{\mathbf{Y}}(z-\tau) d\tau = \int_{-\infty}^{\infty} f_{\mathbf{Y}}(\tau)f_{\mathbf{X}}(z-\tau) d\tau$$

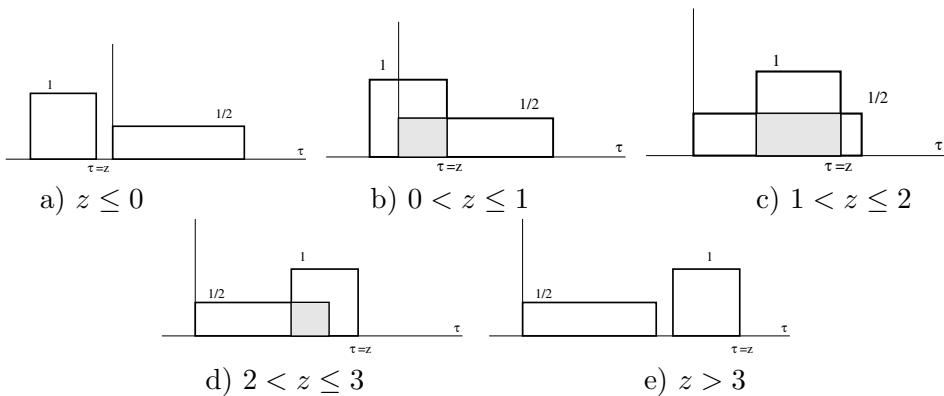


Figura 3.10: Las cinco secciones de la convolución de dos rectángulos.

en cualquiera de sus dos versiones. Escojamos la segunda. Dejamos entonces inalterada la función de densidad de la VA \mathbf{Y} , reflejamos la función de densidad de la VA \mathbf{X} alrededor del eje $\tau = 0$ y desplazamos esta función al punto z donde evaluamos la convolución. Podemos pues considerar cinco casos:

- (a) $z \leq 0$: En este caso (véase figura 3.10a) el área bajo el producto de ambas funciones es nula (las funciones no tienen puntos en común) de forma que la convolución vale cero.
- (b) $0 < z \leq 1$ (figura 3.10b): El cuadrado en que consiste $f_{\mathbf{X}}$ empieza a entrar en el rectángulo de $f_{\mathbf{Y}}$, pero no lo hace completamente. Así pues

$$f_{\mathbf{Z}}(z) = \int_0^z f_{\mathbf{X}}(\tau) f_{\mathbf{Y}}(z - \tau) d\tau = \int_0^z 1 \frac{1}{2} d\tau = \frac{z}{2}$$

- (c) $1 < z \leq 2$ (figura 3.10c): El cuadrado ha entrado completamente en el rectángulo de $f_{\mathbf{Y}}$, y se mantiene en el interior en todo este intervalo. Por ello

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \int_{z-1}^z f_{\mathbf{X}}(\tau) f_{\mathbf{Y}}(z - \tau) d\tau = \int_{z-1}^z 1 \frac{1}{2} d\tau \\ &= \frac{z - (z - 1)}{2} = \frac{1}{2} \end{aligned}$$

- (d) $2 < z \leq 3$ (figura 3.10d): En este caso el cuadrado

comienza a salir del rectángulo. Por ello

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \int_{z-1}^2 f_{\mathbf{X}}(\tau) f_{\mathbf{Y}}(z-\tau) d\tau = \int_{z-1}^2 1 \frac{1}{2} d\tau \\ &= \frac{2 - (z-1)}{2} = \frac{3-z}{2} \end{aligned}$$

- (e) $z > 3$ (figura 3.10e): El cuadrado ha abandonado el interior del rectángulo de forma que el área bajo el producto de ambas funciones vuelve a ser nulo.

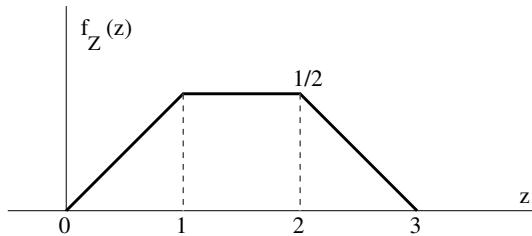


Figura 3.11: Función de densidad $f_{\mathbf{Z}}(z)$ para el ejercicio planteado.

La figura 3.11 muestra la función de densidad resultante. Como puede verse, la convolución de dos rectángulos es un trapecio, en general, y si la longitud de ambos rectángulos coincide, la parte de la convolución que hemos clasificado en tercer lugar desaparece, de forma que, en tal caso, el trapecio se convierte en un triángulo.



2. Considérese que $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \max(\mathbf{X}, \mathbf{Y})$. El objetivo es, como antes, encontrar la función de distribución de \mathbf{Z} . Para ello

$$F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) = P(\max(\mathbf{X}, \mathbf{Y}) \leq z) = P(D_z)$$

Para que el máximo de dos variables sea menor que una cierta cantidad z debe verificarse que ambas variables sean menores o iguales que z . Por ello, D_z es la región $\{\mathbf{X} \leq z, \mathbf{Y} \leq z\}$ (zona sombreada en figura 3.12a), y la función que nos piden es:

$$F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) = P(\mathbf{X} \leq z, \mathbf{Y} \leq z) = F_{\mathbf{XY}}(z, z)$$

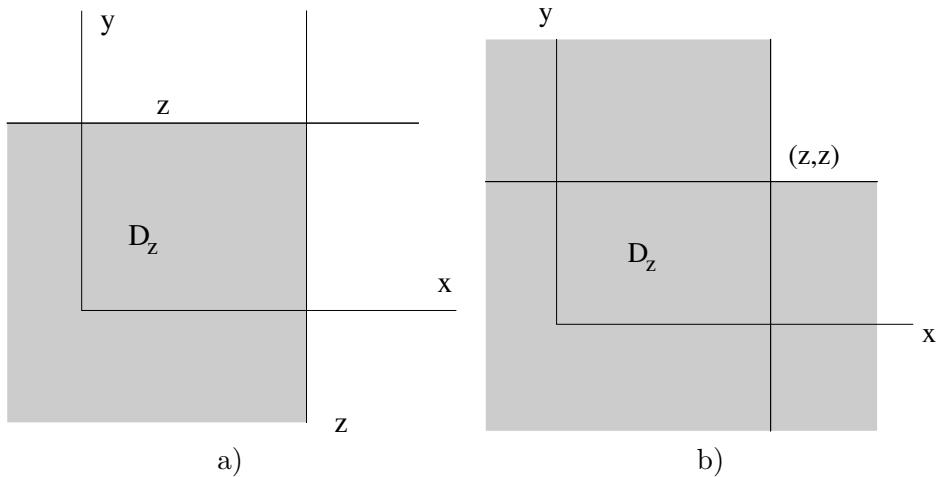


Figura 3.12: a) D_z para $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \max(\mathbf{X}, \mathbf{Y})$ b) D_z para $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \min(\mathbf{X}, \mathbf{Y})$.

3. Considérese que $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \min(\mathbf{X}, \mathbf{Y})$. Buscamos

$$F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) = P(\min(\mathbf{X}, \mathbf{Y}) \leq z) = P(D_z)$$

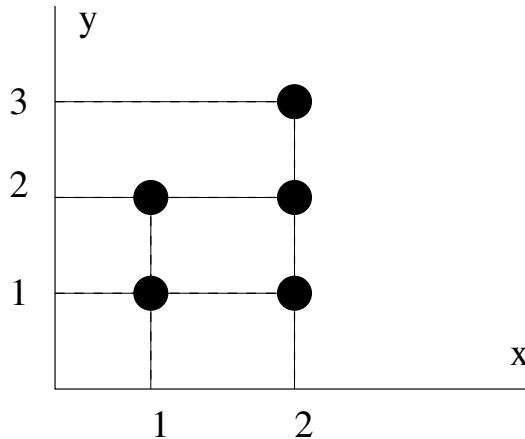
En este caso D_z varía ligeramente respecto del anterior, puesto que para que el mínimo entre dos variables sea menor que una cierta cantidad z basta con que cualquiera de ellas lo sea. Por ello, D_z es la región $\{\mathbf{X} \leq z\} \cup \{\mathbf{Y} \leq z\}$ (zona sombreada en figura 3.12b) y la función que nos piden es:

$$\begin{aligned}
F_{\mathbf{Z}}(z) = P(\mathbf{Z} \leq z) &= P(\{\mathbf{X} \leq z\} \cup \{\mathbf{Y} \leq z\}) \\
&= P(\mathbf{X} \leq z) + P(\mathbf{Y} \leq z) - P(\mathbf{X} \leq z, \mathbf{Y} \leq z) \\
&= F_{\mathbf{X}}(z) + F_{\mathbf{Y}}(z) - F_{\mathbf{XY}}(z, z) \\
&= F_{\mathbf{XY}}(z, \infty) + F_{\mathbf{XY}}(\infty, z) - F_{\mathbf{XY}}(z, z)
\end{aligned} \tag{3.33}$$

Evidentemente, un procedimiento alternativo al empleado sería

$$F_{\mathbf{Z}}(z) = 1 - P(\mathbf{Z} > z) = 1 - P(\mathbf{X} > z, \mathbf{Y} > z)$$

4. Supongamos ahora que \mathbf{X} e \mathbf{Y} son discretas, y que toman los valores que se representan en la figura 3.13, con probabilidades p_{ij} . Se plantea la función $\mathbf{Z} = q(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \cdot \mathbf{Y}$. La VA resultante \mathbf{Z} también

Figura 3.13: Puntos de probabilidad no nula de la VA (\mathbf{X}, \mathbf{Y}) .

será discreta, de forma que su caracterización viene dada por los valores z_k que tomará la variable y por las probabilidades $p_k = P(\mathbf{Z} = z_k)$. Por tanto, podemos construir la siguiente tabla:

z_k	p_k
1	$p_1 = p_{11}$
2	$p_2 = p_{12} + p_{21}$
4	$p_4 = p_{22}$
6	$p_6 = p_{23}$

puesto que

$$\begin{aligned} P(\mathbf{Z} = 1) = p_1 &= P(\mathbf{X} = 1, \mathbf{Y} = 1) = p_{11} \\ P(\mathbf{Z} = 2) = p_2 &= P(\mathbf{X} = 1, \mathbf{Y} = 2) + P(\mathbf{X} = 2, \mathbf{Y} = 1) = p_{12} + p_{21} \\ P(\mathbf{Z} = 4) = p_4 &= P(\mathbf{X} = 2, \mathbf{Y} = 2) = p_{22} \\ P(\mathbf{Z} = 6) = p_6 &= P(\mathbf{X} = 2, \mathbf{Y} = 3) = p_{23} \end{aligned}$$

3.5.2 Dos funciones de dos variables aleatorias

Consideremos ahora la situación en la que (\mathbf{X}, \mathbf{Y}) se transforma en otra VA bidimensional por medio del par de transformaciones:

$$\begin{aligned} \mathbf{Z} &= g(\mathbf{X}, \mathbf{Y}) \\ \mathbf{W} &= h(\mathbf{X}, \mathbf{Y}) \end{aligned}$$

En este caso, como en el apartado anterior, conocida la caracterización probabilística de la VA (\mathbf{X}, \mathbf{Y}) origen de la transformación, y conocidas las expresiones analíticas de la transformación, se pretende obtener la caracterización probabilística de la VA bidimensional resultante (\mathbf{Z}, \mathbf{W}).

El procedimiento de resolución en este caso sería similar al que acabamos de ver. Concretamente pretendemos hallar

$$F_{\mathbf{ZW}}(z, w) = P(\mathbf{Z} \leq z, \mathbf{W} \leq w) = P(D_{zw}) = \int \int_{D_{zw}} f_{\mathbf{XY}}(x, y) dx dy$$

con D_{zw} el lugar geométrico de los puntos del plano (x, y) que se transforman en puntos del plano (z, w) pertenecientes al suceso cuya probabilidad buscamos, esto es

$$D_{zw} = \{(x, y) \in R^2 \text{ t.q. } g(x, y) \leq z, h(x, y) \leq w\}$$

Para tal fin, debemos identificar la región D_{zw} para, a continuación, hallar la probabilidad asociada a dicha región.

Ejemplo: Para el par de transformaciones

$$\begin{aligned} \mathbf{Z} &= g(\mathbf{X}, \mathbf{Y}) = (+)\sqrt{\mathbf{X}^2 + \mathbf{Y}^2} \\ \mathbf{W} &= h(\mathbf{X}, \mathbf{Y}) = \frac{\mathbf{Y}}{\mathbf{X}} \end{aligned}$$

identifíquese la región D_{zw} para un par (z, w) genérico, con $z > 0$ y $w > 0$.

Solución: La región $\{\mathbf{Z} \leq z\} = \{\sqrt{\mathbf{X}^2 + \mathbf{Y}^2} \leq z\}$ es, en el plano (x, y) , el interior de una circunferencia de radio z . Por lo que respecta a la segunda transformación,

$$\mathbf{W} = \frac{\mathbf{Y}}{\mathbf{X}} \leq w \rightarrow \mathbf{Y} \leq w\mathbf{X} \quad (3.34)$$

región cuya frontera es una recta dada por el signo igual de la anterior desigualdad, es decir, $y = wx$. Esta frontera es una recta de pendiente w y ordenada en el origen igual a 0. Los valores que pertenecen a la región $y \leq wx$ son los valores que están por debajo de la recta, como puede comprobarse tomando cualquier punto del primer cuadrante por debajo de la recta de división (véase figura 3.14a).

Sin embargo, hay que hacer una llamada de precaución: la implicación escrita en la ecuación (3.34) no es completamente correcta, puesto que la desigualdad en su estado original es

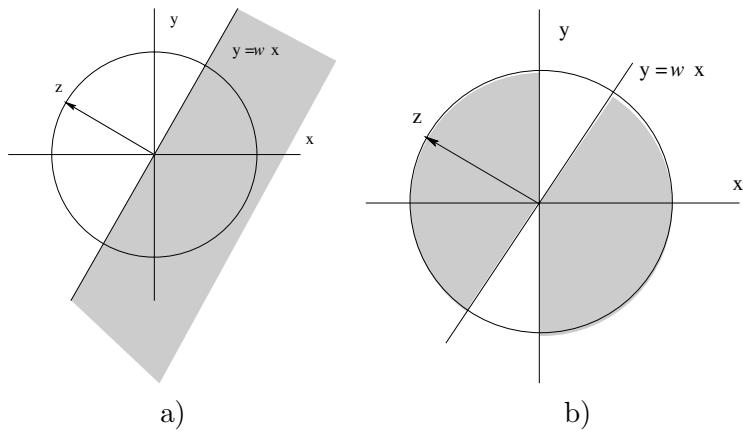


Figura 3.14: a) D_{zw} obtenida a partir de $y \leq wx$ b) D_{zw} obtenida a partir de $\frac{y}{x} \leq w$.

sensible a los signos positivos/negativos de las variables involucradas. Por ello, hay que hacer un análisis más pormenorizado:

1. Primer cuadrante: en este caso no hay duda, pues ambas magnitudes son positivas. Los puntos pertenecientes a la región bajo estudio son los que se encuentran por debajo de la recta.
2. Segundo cuadrante: x es negativa, y es positiva, el cociente es negativo, luego es menor que w si $w > 0$.
3. Tercer cuadrante: los puntos de este cuadrante son puntos de ambas coordenadas negativas, luego el cociente es positivo. De éstos, los que se encuentran por debajo de la recta tienen una ordenada que puede tomar valores arbitrariamente grandes, incluso con $x = 0$. Por ello, para estos puntos $y/x > w$. Por contra, los puntos por encima de la recta tienen signo positivo, y es la x la coordenada que puede crecer arbitrariamente, mientras que y crece de forma moderada. Estos puntos pertenecen a la región.
4. Cuarto cuadrante: el cociente es negativo, luego todos los puntos pertenecen a la región.

La región resultante es pues la que se encuentra sombreada en la figura 3.14b).



En cualquier caso, el método aquí descrito no es empleado con frecuencia pues, como veremos a continuación, existe una expresión cerrada para la obtención de la función de densidad conjunta de dos variables transformadas, la cual constituye el Teorema Fundamental bidimensional.

3.5.3 Teorema Fundamental

Buscamos una extensión bidimensional de la metodología descrita en la sección 2.6.3 del capítulo anterior. Tal metodología, recuérdese, proporcionaba una expresión cerrada para la obtención de la función de densidad de probabilidad de la VA transformada a partir de la VA origen de la transformación. Por tanto, dado que trabajaremos con funciones de densidad de probabilidad, tendremos que hacer uso de cálculo diferencial, empleando en nuestros razonamientos sucesos asociados a rectángulos elementales en cada uno de los dominios (z, w) y (x, y) .

Consideremos una situación como la descrita en la figura 3.15. Podemos

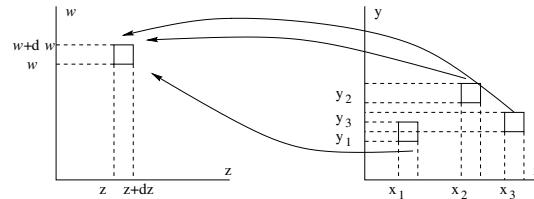


Figura 3.15: Puntos que se corresponden en planos destino y origen.

ver que el punto (z, w) procede de la transformación de varios puntos del dominio (x, y) . Por ello, dado que queremos hallar la función de densidad de la VA bidimensional (\mathbf{Z}, \mathbf{W}) podemos escribir

$$\begin{aligned}
 f_{\mathbf{ZW}}(z, w)dzdw &= P(z < \mathbf{Z} \leq z + dz, w < \mathbf{W} \leq w + dw) \\
 &= P((x_1 < \mathbf{X} \leq x_1 + dx_1, y_1 < \mathbf{Y} \leq y_1 + dy_1) \cup \\
 &\quad (x_2 < \mathbf{X} \leq x_2 + dx_2, y_2 < \mathbf{Y} \leq y_2 + dy_2) \cup \\
 &\quad (x_3 < \mathbf{X} \leq x_3 + dx_3, y_3 < \mathbf{Y} \leq y_3 + dy_3))
 \end{aligned} \tag{3.35}$$

Generalizando, podemos escribir

$$f_{\mathbf{ZW}}(z, w)dzdw = P(z < \mathbf{Z} \leq z + dz, w < \mathbf{W} \leq w + dw)$$

$$\begin{aligned}
&= P \left(\bigcup_{i=1}^N [x_i < \mathbf{X} \leq x_i + dx_i, y_i < \mathbf{Y} \leq y_i + dy_i] \right) \\
&= \sum_{i=1}^N P(x_i < \mathbf{X} \leq x_i + dx_i, y_i < \mathbf{Y} \leq y_i + dy_i) \\
&= \sum_{i=1}^N f_{\mathbf{XY}}(x_i, y_i) |dx_i| |dy_i|
\end{aligned} \tag{3.36}$$

donde los valores absolutos se han incluido para evitar problemas de signos entre los diferenciales, dado que vamos a hacer uso de una relación explícita entre ambos. En efecto, la relación que liga los cuatro diferenciales es

$$dzdw = |J(x, y)|dxdy$$

con $J(x, y)$ el jacobiano de la transformación, el cual es igual a

$$J(x, y) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} \tag{3.37}$$

Despejando la expresión anterior a favor de la función de densidad $f_{\mathbf{ZW}}(z, w)$ podemos escribir

$$\begin{aligned}
f_{\mathbf{ZW}}(z, w) &= \sum_{i=1}^N \frac{f_{\mathbf{XY}}(x_i, y_i)}{|dzdw|} \\
&= \sum_{i=1}^N \frac{f_{\mathbf{XY}}(x, y)}{|J|} \Big|_{(x_i, y_i)}
\end{aligned} \tag{3.38}$$

expresión que constituye el teorema fundamental extendido a dos dimensiones. Repárese que esta expresión permite construir *punto a punto* la función de densidad de la VA (\mathbf{Z}, \mathbf{W}) lo cual significa que para cada punto (z, w) el número N de raíces (x_i, y_i) , esto es, el número de puntos del plano origen que se transforman en el punto (z, w) , puede variar. Por ello, estrictamente, el parámetro N debe escribirse como $N(z, w)$.

Al respecto de este teorema debe recordarse que:

- La solución debe proporcionarse en función de las variables (z, w) . Por ello, debe acudirse a las funciones inversas de $g(x, y)$ y $h(x, y)$ o a cualquier otro procedimiento que se estime conveniente, para escribir la expresión resultante del empleo del teorema en función de las variables correctas.

- Asimismo, debe indicarse cuál es la zona de validez de la expresión o expresiones resultantes de la aplicación del teorema.

Finalmente debe recordarse que la expresión (3.38) puede escribirse, de forma alternativa, a través del jacobiano de la función inversa. Denominando a dicho jacobiano $J_{-1}(z, w)$ podemos escribir

$$dxdy = |J_{-1}(z, w)|dzdw$$

con

$$J_{-1}(z, w) = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix}$$

por lo que

$$f_{\mathbf{ZW}}(z, w) = \sum_{i=1}^N f_{\mathbf{XY}}(x_i, y_i) |J_{-1}| \Big|_{(x_i, y_i)} \quad (3.39)$$

igualdad que puede resultar de interés en función de las expresiones concretas de las operaciones a realizar.

Ejemplo 1: Sean \mathbf{X} e \mathbf{Y} VAs uniformes en el intervalo $(0, 1)$ e independientes. Se generan las VAs \mathbf{Z} y \mathbf{W} mediante:

$$\begin{aligned} \mathbf{Z} &= g(\mathbf{X}, \mathbf{Y}) = \mathbf{XY} \\ \mathbf{W} &= h(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \end{aligned}$$

Obténgase mediante el Teorema Fundamental la función de densidad $f_{\mathbf{ZW}}(z, w)$.

Solución: Puesto que las VAs \mathbf{X} e \mathbf{Y} son independientes, la función de densidad conjunta puede escribirse $f_{\mathbf{XY}}(x, y) = f_{\mathbf{X}}(x)f_{\mathbf{Y}}(y)$, por lo que ésta es constante de valor unidad en el interior del cuadrado $0 \leq x \leq 1, 0 \leq y \leq 1$, y nula fuera de dicha región.

La transformación es inyectiva (salvo para el punto $(z, w) = (0, 0)$, de contribución irrelevante a la probabilidad de sucesos definidos sobre la VA), de forma que podemos escribir:

$$f_{\mathbf{ZW}}(z, w) = \frac{f_{\mathbf{XY}}(x, y)}{|J|}$$

Sustituyendo términos

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{1}{\begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix}} \\ &= \frac{1}{|-x|} = \frac{1}{|x|} \end{aligned}$$

Llegado a este punto resta indicar las dos cuestiones anteriormente apuntadas. Así pues

1. Para escribir la expresión en función de (z, w) en este ejemplo las cosas son particularmente sencillas pues existe una identidad entre \mathbf{X} y \mathbf{W} . Ello nos permite escribir:

$$f_{\mathbf{ZW}}(z, w) = \frac{1}{w}$$

2. La expresión anterior es válida en la región del plano (z, w) sobre la que se transforma el cuadrado de lado 1 donde es no nula la función de densidad $f_{\mathbf{XY}}(x, y)$. Para ver qué región es ésta podemos razonar componente a componente:

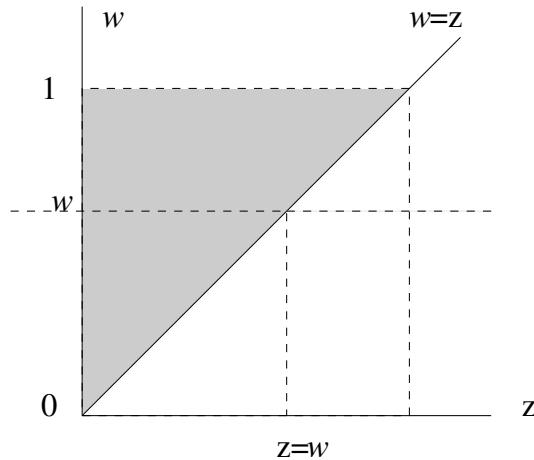


Figura 3.16: Región de soporte de la función $f_{\mathbf{ZW}}(z, w)$.

- Dado que $\mathbf{W} = \mathbf{X}$, si \mathbf{X} varía en el intervalo $(0, 1)$, \mathbf{W} hará lo propio.

- Para cada valor w de la VA \mathbf{W} (es decir, cortando por una horizontal de ordenada arbitraria w , véase la figura 3.16), y dado que $\mathbf{W} = \mathbf{X}$, entonces la VA \mathbf{Z} sería

$$\mathbf{Z} = w\mathbf{Y}$$

lo cual indica que, sobre esa horizontal, el valor mínimo que puede tomar \mathbf{Z} es 0 (correspondiente al valor mínimo de \mathbf{Y}) y el valor máximo es w (correspondiente al valor máximo de \mathbf{Y}). Por ello, para cada valor de \mathbf{W} la VA \mathbf{Z} varía desde 0 hasta \mathbf{W} , es decir, desde 0 hasta la bisectriz del primer cuadrante (zona sombreada de la figura 3.16).

Así pues, la función pedida es

$$f_{\mathbf{ZW}}(z, w) = \begin{cases} \frac{1}{w} & 0 \leq z \leq w \leq 1 \\ 0 & \text{resto del plano} \end{cases}$$

Ejercicio: Compruébese que el volumen encerrado bajo esta función de densidad es unitario.



Ejemplo 2: Sean \mathbf{R} y Θ dos VAs independientes donde \mathbf{R} tiene una función de densidad Rayleigh, a saber,

$$f_{\mathbf{R}}(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r)$$

y $\Theta \sim U(-\pi, \pi)$. Se generan las VAs \mathbf{Z} y \mathbf{W} mediante:

$$\begin{aligned} \mathbf{Z} &= g(\mathbf{R}, \Theta) = \mathbf{R} \cos(\Theta) \\ \mathbf{W} &= h(\mathbf{R}, \Theta) = \mathbf{R} \sin(\Theta) \end{aligned}$$

Obténgase mediante el Teorema Fundamental la función de densidad $f_{\mathbf{ZW}}(z, w)$.

Solución: Es evidente que esta transformación es un cambio de coordenadas polares a coordenadas cartesianas, de forma que la relación es 1:1. Asimismo, los índices de variación de las variables origen son tales que las VAs cartesianas \mathbf{Z} y \mathbf{W} podrán

tomar cualquier valor del plano R^2 . Por ello las conclusiones que extraigamos serán válidas para todo el plano.

Aplicando el Teorema Fundamental, tenemos

$$f_{\mathbf{ZW}}(z, w) = \frac{f_{\mathbf{R}\Theta}(r, \theta)}{|J|} = \frac{f_{\mathbf{R}}(r)f_{\Theta}(\theta)}{|J|}$$

donde el jacobiano es igual a

$$J(r, \theta) = \begin{vmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{vmatrix} = r(\cos^2(\theta) + \sin^2(\theta)) = r$$

Sustituyendo términos

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{\frac{1}{2\pi} \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}}{r} \\ &= \frac{1}{\sigma^2 2\pi} e^{-\frac{r^2}{2\sigma^2}} \\ &= \frac{1}{\sigma^2 2\pi} e^{-\frac{z^2+w^2}{2\sigma^2}} \end{aligned}$$

ya que $r^2 = z^2 + w^2$.

Repárese que la expresión anterior puede factorizarse en

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{w^2}{2\sigma^2}} \\ &= f_{\mathbf{Z}}(z)f_{\mathbf{W}}(w) \end{aligned}$$

de donde se deduce que ambas variables son independientes, y marginalmente gaussianas, de media nula y desviación típica σ .



3.5.4 Método de la variable auxiliar

Los ejemplos anteriores han puesto de manifiesto que el Teorema Fundamental es una herramienta poderosa para realizar el cálculo de la función de densidad conjunta de dos VAs. Sería pues deseable poder emplear este teorema para la situación que analizamos en el apartado 3.5.1, a saber, la caracterización probabilística de una única función de dos VAs $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$.

La manera mediante la cual esto es posible no es otra que convirtiendo artificialmente el problema $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$ en un problema del tipo

$$\begin{aligned}\mathbf{Z} &= g(\mathbf{X}, \mathbf{Y}) = \mathbf{XY} \\ \mathbf{W} &= h(\mathbf{X}, \mathbf{Y})\end{aligned}$$

definiendo una VA \mathbf{W} arbitraria. A esta VA se le denomina *variable auxiliar*, de la que este método toma el nombre. Realizado esto, bastaría ejecutar las dos acciones siguientes:

1. Obtener la función de densidad conjunta $f_{\mathbf{ZW}}(z, w)$ mediante el Teorema Fundamental.
2. Obtener la función de densidad marginal de la VA realmente involucrada en el problema. Para ello.

$$f_{\mathbf{Z}}(z) = \int_{-\infty}^{\infty} f_{\mathbf{ZW}}(z, w) dw$$

Es práctica habitual, por sensatez básica, escoger una VA auxiliar que coincida con una de las dos VAs origen de la transformación. El motivo es obvio: dado que las operaciones a realizar pueden ser complicadas, interesa que la VA auxiliar sea lo más sencilla posible. El hecho de que coincida con una de las dos VAs origen garantiza que un elemento de la matriz de derivadas para el cálculo del jacobiano sea nulo. Asimismo, a la hora de escribir $f_{\mathbf{ZW}}(z, w)$ como función de las variables (z, w) la inversión es más sencilla si una de las dos variables (x, y) pasa directamente al dominio (u, w) .

Ejemplo: Obténgase la función de densidad de probabilidad de la suma de dos VAs \mathbf{X} e \mathbf{Y} , cuya función de densidad $f_{\mathbf{XY}}(x, y)$ se asumirá conocida. Particularícese para el caso en que éstas sean independientes.

Solución: En el enunciado se plantea, exclusivamente, la función $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \mathbf{Y}$. Para emplear el método de la variable auxiliar hacemos uso de una segunda transformación. Para ello, definimos

$$\begin{aligned}\mathbf{Z} &= g(\mathbf{X}, \mathbf{Y}) = \mathbf{X} + \mathbf{Y} \\ \mathbf{W} &= h(\mathbf{X}, \mathbf{Y}) = \mathbf{X}\end{aligned}$$

Con estas dos VAs podemos aplicar el Teorema Fundamental; dado que la transformación es lineal de determinante no nulo, la relación es 1:1, por lo que podemos escribir

$$f_{\mathbf{ZW}}(z, w) = \frac{f_{\mathbf{XY}}(x, y)}{|J|}$$

Sustituyendo términos

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{f_{\mathbf{XY}}(x, y)}{\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}} \\ &= \frac{f_{\mathbf{XY}}(x, y)}{|-1|} = \frac{f_{\mathbf{XY}}(x, y)}{1} = f_{\mathbf{XY}}(x, y) \\ &= f_{\mathbf{XY}}(w, z - w) \end{aligned}$$

Por tanto

$$f_{\mathbf{Z}}(z) = \int_{-\infty}^{\infty} f_{\mathbf{ZW}}(z, w) dw = \int_{-\infty}^{\infty} f_{\mathbf{XY}}(w, z - w) dw$$

Si particularizamos para el caso en que las VAs \mathbf{X} e \mathbf{Y} sean independientes entonces

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \int_{-\infty}^{\infty} f_{\mathbf{XY}}(w, z - w) dw \int_{-\infty}^{\infty} f_{\mathbf{X}}(w) f_{\mathbf{Y}}(z - w) dw \\ &= f_{\mathbf{X}}(z) * f_{\mathbf{Y}}(z) \end{aligned}$$

es decir, como ya conocemos de la sección 3.5.1, la convolución de las funciones de densidad de cada una de las VAs. En dicha sección llegamos a este mismo resultado con algo más de esfuerzo.



3.6 Caracterización parcial de una función de dos variables aleatorias

En el capítulo anterior (sección 2.7) vimos que si nos enfrentábamos a la situación $\mathbf{Y} = g(\mathbf{X})$, pero sólo estábamos interesados en calcular ciertos parámetros de caracterización parcial de la VA \mathbf{Y} , entonces no había razón

para calcular la función de densidad $f_{\mathbf{Y}}(y)$ de esta VA, sino que bastaba con la transformación $g(\mathbf{X})$ y la función de densidad de la VA \mathbf{X} para hallar cualquier momento de la VA \mathbf{Y} .

Estas conclusiones aplican, por idéntico motivo, al caso $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$. Si necesitamos conocer, en relación con la VA \mathbf{Z} , el valor de $E\{h(\mathbf{Z})\}$, podemos reflejar el problema sobre el dominio (\mathbf{X}, \mathbf{Y}) (al cual, naturalmente, supondremos caracterizado) y escribir

$$E\{h(\mathbf{Z})\} = E\{h(g(\mathbf{X}, \mathbf{Y}))\} = \int \int h(g(x, y)) f_{\mathbf{XY}}(x, y) dx dy \quad (3.40)$$

En particular, si $h(\mathbf{Z}) = \mathbf{Z}$ entonces

$$E\{h(\mathbf{Z})\} = E\{\mathbf{Z}\} = E\{g(\mathbf{X}, \mathbf{Y})\} = \int \int g(x, y) f_{\mathbf{XY}}(x, y) dx dy \quad (3.41)$$

Si $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = a\mathbf{X} + b\mathbf{Y} + c$, con a, b, c constantes reales, tendremos

$$\begin{aligned} E\{\mathbf{Z}\} &= E\{g(\mathbf{X}, \mathbf{Y})\} \\ &= \int \int (ax + by + c) f_{\mathbf{XY}}(x, y) dx dy \\ &= \int \int ax f_{\mathbf{XY}}(x, y) dx dy + \int \int by f_{\mathbf{XY}}(x, y) dx dy + \\ &\quad \int \int c f_{\mathbf{XY}}(x, y) dx dy \\ &= a \int \int x f_{\mathbf{XY}}(x, y) dx dy + b \int \int y f_{\mathbf{XY}}(x, y) dx dy + \\ &\quad c \int \int f_{\mathbf{XY}}(x, y) dx dy \\ &= a \int x dx \int f_{\mathbf{XY}}(x, y) dy + b \int y dy \int f_{\mathbf{XY}}(x, y) dx + \\ &\quad c \int \int f_{\mathbf{XY}}(x, y) dx dy \\ &= a E\{\mathbf{X}\} + b E\{\mathbf{Y}\} + c \end{aligned} \quad (3.42)$$

es decir, el operador *esperanza matemática* es, también para funciones de dos variables, un operador lineal.

Finalmente, indiquemos que si las VAs son discretas podremos escribir

$$\begin{aligned} E\{\mathbf{Z}\} &= E\{g(\mathbf{X}, \mathbf{Y})\} = \sum_i \sum_j g(x_i, y_j) P(\mathbf{X} = x_i, \mathbf{Y} = y_j) \\ &= \sum_i \sum_j g(x_i, y_j) p_{ij} \end{aligned} \quad (3.43)$$

3.6.1 Esperanzas condicionadas

El cálculo de una esperanza condicionada es similar al de una esperanza incondicional. Lo único que varía es el empleo de una función de densidad condicionada en lugar de una incondicional. Concretamente, si $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y})$ y $M \in \mathcal{F}_c$, con \mathcal{F}_c la clase de sucesos del experimento compuesto ϵ_c , podremos escribir

$$E\{\mathbf{Z}|M\} = E\{g(\mathbf{X}, \mathbf{Y})|M\} = \int \int g(x, y) f_{\mathbf{XY}}(x, y|M) dx dy \quad (3.44)$$

Especial interés tiene el caso particular $\mathbf{Z} = g(\mathbf{X}, \mathbf{Y}) = \mathbf{Y}$ con el condicionante $M = \{\mathbf{X} = x\}$. En tal caso calcularíamos

$$E\{\mathbf{Y}|\mathbf{X} = x\} = E\{\mathbf{Y}|x\} = \int y f_{\mathbf{Y}}(y|x) dy = \int y \frac{f_{\mathbf{XY}}(x, y)}{f_{\mathbf{X}}(x)} dy = \varphi(x) \quad (3.45)$$

Haremos uso extensivo de la misma en la sección 3.7.6.

3.6.2 Momentos conjuntos

Los momentos vistos en el capítulo anterior se extienden al caso bidimensional de la forma que sigue:

- Momento no central de órdenes (r, s) :

$$m_{rs} = E\{\mathbf{X}^r \mathbf{Y}^s\} = \int \int x^r y^s f_{\mathbf{XY}}(x, y) dx dy \quad (3.46)$$

- Momento central de órdenes (r, s) :

$$\mu_{rs} = E\{(\mathbf{X} - \eta_{\mathbf{X}})^r (\mathbf{Y} - \eta_{\mathbf{Y}})^s\} = \int \int (x - \eta_{\mathbf{X}})^r (y - \eta_{\mathbf{Y}})^s f_{\mathbf{XY}}(x, y) dx dy \quad (3.47)$$

donde, como es conocido, los órdenes $r, s \in \mathbb{Z}^+$ típicamente. De nuevo, si las VAs involucradas fuesen discretas, los operadores anteriores se reducirían a sumatorios. Por ejemplo, el momento central anteriormente definido se escribiría

$$\mu_{rs} = E\{(\mathbf{X} - \eta_{\mathbf{X}})^r (\mathbf{Y} - \eta_{\mathbf{Y}})^s\} = \sum_i \sum_j (x_i - \eta_{\mathbf{X}})^r (y_j - \eta_{\mathbf{Y}})^s P(\mathbf{X} = x_i, \mathbf{Y} = y_j) \quad (3.48)$$

Entre todos estos momentos son de particular interés los introducidos en el capítulo anterior, léase, media, varianza y VCM. Asimismo, y de nueva introducción, son los siguientes:

- Correlación: $R_{\mathbf{XY}} = m_{11} = E\{\mathbf{XY}\}$
- Covarianza: $C_{\mathbf{XY}} = \mu_{11} = E\{(\mathbf{X} - \eta_{\mathbf{X}})(\mathbf{Y} - \eta_{\mathbf{Y}})\}$

Estas dos magnitudes están relacionadas debido a la propiedad de linealidad del operador esperanza. En efecto:

$$\begin{aligned} C_{\mathbf{XY}} &= E\{(\mathbf{X} - \eta_{\mathbf{X}})(\mathbf{Y} - \eta_{\mathbf{Y}})\} \\ &= E\{\mathbf{XY} - \eta_{\mathbf{Y}}\mathbf{X} - \eta_{\mathbf{X}}\mathbf{Y} + \eta_{\mathbf{X}}\eta_{\mathbf{Y}}\} \\ &= R_{\mathbf{XY}} - \eta_{\mathbf{Y}}\eta_{\mathbf{X}} - \eta_{\mathbf{X}}\eta_{\mathbf{Y}} + \eta_{\mathbf{X}}\eta_{\mathbf{Y}} \\ &= R_{\mathbf{XY}} - \eta_{\mathbf{X}}\eta_{\mathbf{Y}} \end{aligned} \quad (3.49)$$

Nótese que esta igualdad es análoga a la igualdad que relaciona varianza con VCM y media al cuadrado. De hecho, si en la expresión anterior la VA \mathbf{Y} se sustituye por la VA \mathbf{X} , entonces obtenemos la igualdad (2.48). Por tanto, la covarianza entre dos variables es la generalización del concepto de varianza, y el correlación es la generalización del valor cuadrático medio.

Un coeficiente que se emplea con la práctica es el llamado *coeficiente de correlación* o covarianza normalizada. Se define:

$$\rho_{\mathbf{XY}} = \frac{C_{\mathbf{XY}}}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} \quad (3.50)$$

y es un valor real que verifica que $|\rho_{\mathbf{XY}}| \leq 1$. Mide el grado de relación lineal que existe entre las VAs, esto es, el grado en el que una VA, digamos \mathbf{Y} , se puede predecir mediante una función lineal de la otra (VA \mathbf{X} , en este caso). La predicción perfecta se da cuando $\rho_{\mathbf{XY}} = \pm 1$ y tal relación no existe cuando $\rho_{\mathbf{XY}} = 0$. Posponemos los detalles sobre esta cuestión hasta la sección 3.7.4 (téngase en cuenta, no obstante, el comentario que se hace en el ejercicio de la página 155).

Las definiciones anteriores se complementan con las siguientes:

- Incorelación: dos VAs se dice que son (o que están) incorreladas si $C_{\mathbf{XY}} = 0$. En tal caso se verifica también que $\rho_{\mathbf{XY}} = 0$ y, por tanto, incorrelación implica ausencia de relación lineal entre las VAs.
- Ortogonalidad: dos VAs se dice que son ortogonales si $R_{\mathbf{XY}} = 0$.

Tales definiciones traen consigo las siguientes consecuencias:

1. Si las VAs \mathbf{X} e \mathbf{Y} son independientes, entonces son incorreladas. En efecto,

$$C_{\mathbf{XY}} = E\{(\mathbf{X} - \eta_{\mathbf{X}})(\mathbf{Y} - \eta_{\mathbf{Y}})\}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})(y - \eta_{\mathbf{Y}}) f_{\mathbf{XY}}(x, y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}})(y - \eta_{\mathbf{Y}}) f_{\mathbf{X}}(x) f_{\mathbf{Y}}(y) dx dy \\
&= \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}}) f_{\mathbf{X}}(x) dx \int_{-\infty}^{\infty} (y - \eta_{\mathbf{Y}}) f_{\mathbf{Y}}(y) dy \\
&= (E\{\mathbf{X}\} - \eta_{\mathbf{X}})(E\{\mathbf{Y}\} - \eta_{\mathbf{Y}}) = 0
\end{aligned}$$

Conceptualmente esto es debido a que la independencia entre VAs implica *ausencia de relación* entre las variables, mientras que la incorrelación implica *ausencia de relación lineal* entre las VAs. Por ello, la segunda está incluida en la primera, de forma que si se verifica la primera, debe verificarse la segunda.

Por esta misma razón el recíproco, en general, no es cierto.

Ejercicio: Considere $\mathbf{X} \sim N(0, \sigma)$. Se define $\mathbf{Y} = \mathbf{X}^2$. Indique si las VAs son incorreladas e independientes.

Solución: Por simple inspección es obvio que las VAs no pueden ser independientes, dado que una es función de la otra (sabido lo que vale \mathbf{X} es conocido el valor de \mathbf{Y}). Respecto a la incorrelación, calculemos su covarianza:

$$\begin{aligned}
C_{\mathbf{XY}} &= E\{(\mathbf{X} - \eta_{\mathbf{X}})(\mathbf{Y} - \eta_{\mathbf{Y}})\} \\
&= R_{\mathbf{XY}} - \eta_{\mathbf{X}}\eta_{\mathbf{Y}} \\
&= R_{\mathbf{XY}} = E\{\mathbf{XY}\} = E\{\mathbf{X}^3\} \\
&= \int_{-\infty}^{\infty} x^3 f_{\mathbf{X}}(x) dx = 0
\end{aligned}$$

La integral anterior es nula dado que la función de densidad de la VA \mathbf{X} es una función par, mientras que la función $f(x) = x^3$ es impar, y el intervalo de integración es simétrico con respecto al cero.

Por tanto, las VAs son incorreladas, pero no independientes. Repárese que, como diremos en el apartado 3.7.4, la incorrelación implica que la variable \mathbf{Y} no se puede predecir con una función lineal de la VA \mathbf{X} *mejor* que mediante una simple constante. Sin embargo no debe entenderse que esto es debido a que no existe un término lineal en la relación entre \mathbf{X} e \mathbf{Y} . En particular, si se repite el ejercicio para $\mathbf{Y} = \mathbf{X}^3$ se verá que, si bien no existe una relación analítica

lineal entre las VAs, sí que existe una covarianza no nula. Con ello se está afirmando que una función lineal de la VA \mathbf{X} predice *mejor* a la VA \mathbf{Y} (apartado 3.7.4) que una recta horizontal de valor igual a la media de la VA \mathbf{Y} (apartado 3.7.3).



2. Si \mathbf{X} e \mathbf{Y} son incorreladas:

- (a) $C_{\mathbf{XY}} = 0 \rightarrow R_{\mathbf{XY}} = E\{\mathbf{XY}\} = E\{\mathbf{X}\}E\{\mathbf{Y}\}$, es decir, la esperanza del producto es igual al producto de las esperanzas
- (b) La varianza de la suma es igual a la suma de las varianzas. En efecto:

$$\begin{aligned}\sigma_{\mathbf{X}+\mathbf{Y}}^2 &= E\{((\mathbf{X} + \mathbf{Y}) - (\eta_{\mathbf{X}} + \eta_{\mathbf{Y}}))^2\} \\ &= E\{((\mathbf{X} - \eta_{\mathbf{X}}) + (\mathbf{Y} - \eta_{\mathbf{Y}}))^2\} \\ &= E\{(\mathbf{X} - \eta_{\mathbf{X}})^2 + (\mathbf{Y} - \eta_{\mathbf{Y}})^2 + 2(\mathbf{X} - \eta_{\mathbf{X}})(\mathbf{Y} - \eta_{\mathbf{Y}})\} \\ &= \sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2 + 2C_{\mathbf{XY}}\end{aligned}\quad (3.51)$$

Si las VAs son incorreladas, el tercer sumando se anula, y se verifica el resultado indicado.

- 3. Incorrelación y ortogonalidad son una misma propiedad si al menos una de las dos VAs involucradas tiene media nula. Tal afirmación es obvia a partir de la relación (3.49).
- 4. Sin las VAs son ortogonales, entonces el VCM de la suma es igual a la suma de los VCMs. Esto se deduce haciendo un razonamiento paralelo al hecho para la varianza de la suma, de la forma:

$$\begin{aligned}E\{(\mathbf{X} + \mathbf{Y})^2\} &= E\{\mathbf{X}^2 + \mathbf{Y}^2 + 2\mathbf{XY}\} \\ &= E\{\mathbf{X}^2\} + E\{\mathbf{Y}^2\} + 2E\{\mathbf{XY}\} \\ &= E\{\mathbf{X}^2\} + E\{\mathbf{Y}^2\} + 2R_{\mathbf{XY}}\end{aligned}\quad (3.52)$$

Interpretación geométrica

Es fácil comprobar que la operación $E\{\mathbf{XY}\}$ es un producto escalar, de forma que se puede considerar el espacio de las VAs como un espacio vectorial con una métrica asociada. En base a esto, los resultados expuestos anteriormente tienen una interpretación geométrica relativamente directa.

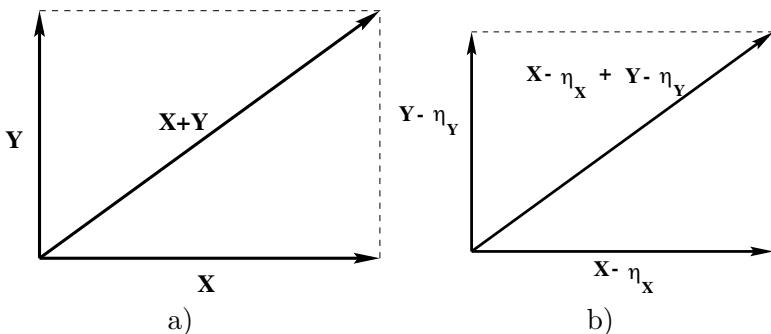


Figura 3.17: a) Ilustración del concepto de ortogonalidad. b) Ilustración del concepto de incorrelación.

Si consideramos el caso de dos VAs ortogonales, podemos interpretarlas como dos vectores cuyo producto escalar es nulo, esto es, como dos vectores perpendiculares. El VCM de cada una de las VAs sería igual a la norma del vector, esto es, a su módulo al cuadrado. Por ello, $E\{(\mathbf{X} + \mathbf{Y})^2\}$ sería igual a la norma del vector suma $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, la cual, como muestra la figura 3.17a) sería igual a la suma de las normas de cada uno de los vectores. Éste es el resultado obtenido en la expresión (3.52) para el caso $R_{\mathbf{XY}} = 0$.

La incorrelación puede ilustrarse mediante la figura 3.17b). En tal caso los vectores $\mathbf{X} - \eta_{\mathbf{X}}$ e $\mathbf{Y} - \eta_{\mathbf{Y}}$ son ortogonales por lo que la norma del vector suma, léase, $\mathbf{X} + \mathbf{Y} - (\eta_{\mathbf{X}} + \eta_{\mathbf{Y}})$, es igual a la suma de las normas de cada uno de ellos. Como el VCM (la norma) calculado sobre una VA menos su media es igual a la varianza de la misma, la varianza de la suma es igual a suma de varianzas si las VAs son incorreladas (resultado indicado en la expresión 3.51 para el caso $C_{\mathbf{XY}} = 0$).

3.7 Estimación de mínimo error cuadrático medio

3.7.1 Introducción

En este apartado indicaremos el procedimiento mediante el cual el valor y de una VA \mathbf{Y} puede predecirse en base a la observación x de que podamos disponer de otra VA \mathbf{X} . Sabemos que si la relación entre las variables fuese del tipo $\mathbf{X} = h(\mathbf{Y})$ con $h()$ una función determinística y monovaluada, es decir, con $h()$ una función no aleatoria y tal que cada origen tenga una única imagen, entonces, conocido el valor x que ha tomado la VA \mathbf{X} conoceríamos sin error ninguno el valor y de la VA \mathbf{Y} sin más que invertir la función en

el valor concreto observado x de la variable \mathbf{X} , es decir, $y = h^{-1}(x)$.

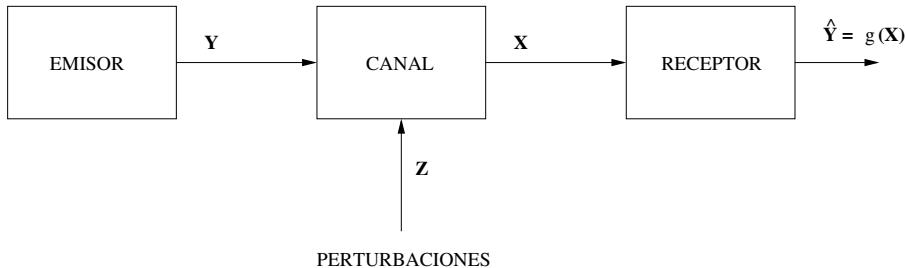


Figura 3.18: Diagrama de bloques de un esquema de comunicaciones.

No obstante, existen numerosos casos en los que la relación entre las variables \mathbf{X} e \mathbf{Y} no es determinística, sino aleatoria. De hecho, éste es el caso más frecuente en los sistemas de comunicaciones. Según se representa en la figura 3.18, un emisor envía un determinada señal; considerando un esquema muy simplificado del problema —con el objetivo de hacerlo encajar en un esquema de variables bidimensionales— pensemos que analizamos la situación en un estado estático, esto es, en un único instante temporal. Aceptando esto como premisa, la señal enviada por el emisor, en el instante temporal de análisis, se reduce⁵ a un escalar y . Si el mensaje enviado cambia, el número a enviar sería otro, de forma que el conjunto de números, cada uno asociado a un mensaje, constituiría el conjunto de valores de una VA \mathbf{Y} . Esta variable es transformada por el canal y, además, sobre ésta se superponen un conjunto de perturbaciones aleatorias, de forma que la señal recibida se podría modelar como una VA \mathbf{X} cuya relación con la \mathbf{Y} sería del tipo

$$\mathbf{X} = h(\mathbf{Y}) + \mathbf{Z} \quad (3.53)$$

con \mathbf{Z} una VA que agruparía todas las perturbaciones que se habrían superpuesto sobre la transformación de \mathbf{Y} mediante el canal. En este caso, el receptor tratará de averiguar el valor y que ha tomado la variable \mathbf{Y} , pero, como se ve, en base a una observación x de la VA \mathbf{X} , versión transformada y ruidosa de la variable de origen. Para ello construirá una función $\hat{\mathbf{Y}} = g(\mathbf{X})$, la cual pretenderá acercarse lo máximo posible al verdadero valor y que haya tomado la variable \mathbf{Y} . No obstante, dado que la relación

⁵Si la señal fuese $y(t) \in \mathcal{R}$ y analizamos el instante temporal t_0 entonces el valor que manejaríamos sería $y(t_0) = y$. Planteamos en este ejemplo un problema de *filtrado de señales* pero, para evitar emplear conceptos que analizaremos en el capítulo 5, hemos considerado la simplificación indicada.

es una relación aleatoria ya no es posible invertir la función de transformación, de modo que nos tendremos que conformar simplemente con tratar de *adivinar* (en lenguaje técnico, *estimar*) el valor que ha tomado la variable original. Por esta razón, a la función $\hat{\mathbf{Y}} = g(\mathbf{X})$ se le denomina *estimador* de la VA \mathbf{Y} , y es una función de la VA \mathbf{X} , esto es, una función que a cada observación concreta x de la VA \mathbf{X} le asocia un valor y que podría haber tomado la VA \mathbf{Y} .

Dado que $g(\mathbf{X})$ es una función de VA, es también una VA, de forma que la expresión *tratar de adivinar el valor que ha tomado la variable original* tiene significado sólo en términos probabilísticos. ¿Qué criterio puedes emplear para construir esta función?

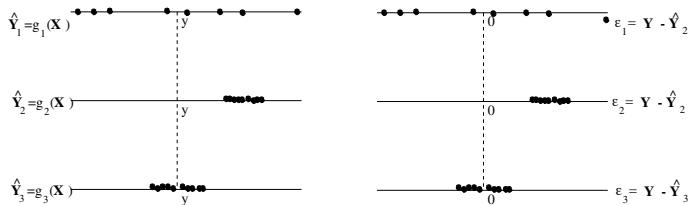


Figura 3.19: Ejemplo de prestaciones de diferentes selecciones de la función de estimación $g_i(\mathbf{X})$, $\{1, 2, 3\}$. Cada punto procede de una realización de la VA \mathbf{X} para el caso particular en que $\mathbf{Y} = y$.

Para ilustrar tal criterio, consideremos la situaciones que se indican en la figura 3.19. En ella se muestran los posibles resultados de múltiples experimentaciones con tres estimadores $\hat{\mathbf{Y}}_i = g_i(\mathbf{X})$ ($i = 1, 2, 3$) cuando la cantidad correcta es $\mathbf{Y} = y$. La parte izquierda de la figura muestra dónde han caído cada uno de los valores $g_i(x)$, tomando como referencia el valor verdadero y . La derecha de la figura muestra cómo se distribuyen los errores en la estimación, entendiendo por tales los diferentes valores que ha tomado la VA *error en la estimación*, definida genéricamente como $\epsilon = \mathbf{Y} - \hat{\mathbf{Y}}$. Es claro, a partir de la figura que:

- En el caso del estimador $\hat{\mathbf{Y}}_1 = g_1(\mathbf{X})$ se cometen errores centrados en el error cero (los valores del estimador oscilan en torno a y), pero la dispersión de los mismos es elevada.
- Para $\hat{\mathbf{Y}}_2 = g_2(\mathbf{X})$ la dispersión de los valores es pequeña ya que todos se concentran en una parte reducida de los ejes dibujados. Sin embargo, los valores del error no están centrados sobre el valor 0 de error, sino que existe un término sistemático de error (sesgo, en adelante).

- En el caso de $\hat{\mathbf{Y}}_3 = g_3(\mathbf{X})$ la variabilidad en el error es pequeña y además no existe el error sistemático anterior sino que los valores de error oscilan en torno al cero.

El tercer estimador parece el más adecuado y los estimadores que construiremos se diseñarán con el objetivo de que se comporten como éste. En la terminología que conocemos, es decir, recordando que $E\{\epsilon^2\} = \sigma_\epsilon^2 + E^2\{\epsilon\}$ podríamos decir que:

- El estimador $\hat{\mathbf{Y}}_1 = g_1(\mathbf{X})$ posiblemente se ha construido tratando de minimizar el término $E^2\{\epsilon\}$ de la expresión anterior, pues, en efecto, el error tiene media nula, pero no se ha prestado atención a la varianza σ_ϵ^2 del mismo.
- El estimador $\hat{\mathbf{Y}}_2 = g_2(\mathbf{X})$ posiblemente se ha construido tratando de minimizar el término σ_ϵ^2 de la expresión anterior, pero no se ha considerado el término $E^2\{\epsilon\}$, de forma que el error tiene una media distinta de cero.
- En el tercer caso se ha atendido a ambos criterios a la vez, esto es, se ha diseñado con el criterio de minimizar $E\{\epsilon^2\}$, por lo que se ha logrado una variabilidad pequeña y ausencia de sesgo⁶.

Los estimadores que se construyen acorde con la filosofía de minimizar el valor de $E\{\epsilon^2\}$ se denominan *estimadores de mínimo error cuadrático medio* y se denotan de forma abreviada por las siglas de su equivalente en inglés MMSEE (*minimum mean square error estimator*). En lo que sigue éste será el criterio que emplearemos para diseñar estimadores.

La estructura de la exposición es como sigue: en primer lugar haremos una recopilación de los tres resultados básicos de esta sección 3.7. Con ello se pretende disponer de una guía de referencia rápida, ajena a los detalles de obtención de los mismos. A continuación deduciremos tales resultados: en primer lugar estimaremos mediante una constante y obtendremos

⁶Lo dicho es cierto, únicamente, para el caso en que la magnitud a estimar sea una VA ya que, como veremos, los estimadores presentarán— felizmente —ausencia de sesgo (a los cuales se denomina *insesgados*). Si la magnitud a estimar fuese una constante determinística desconocida, y estimada a partir de observaciones de VAs, la estimación que se construyese en base a minimizar el VCM del error no tendría por qué dar lugar a estimadores insesgados, pudiéndose perfectamente llevar a cabo un intercambio entre sesgo y varianza para minimizar tal VCM. Los estimadores de constante determinística que sí se ajustarían a este tercer caso serían los conocidos como *estimadores insesgados de mínima varianza* [4]. Estas cuestiones, no obstante, caen fuera de los objetivos de este tratado, mencionándose aquí únicamente por completitud.

qué constante es la óptima en el sentido MMSE. Seguidamente haremos lo propio, pero permitiendo que el estimador sea una función lineal de la observación. Encontraremos qué recta es la óptima en el sentido MMSE. Por último, eliminaremos todo tipo de restricción a la expresión funcional del estimador, de forma que deduciremos cuál es el estimador óptimo sin restricciones.

3.7.2 Resumen de estimadores

- Estimador mediante una constante: $\hat{\mathbf{Y}} = g(\mathbf{X}) = a$.

$$a = E\{\mathbf{Y}\} \quad (3.54)$$

- Estimador lineal $\hat{\mathbf{Y}} = g(\mathbf{X}) = a\mathbf{X} + b$

$$a = \frac{C_{XY}}{\sigma_X^2} \quad (3.55)$$

$$b = E\{\mathbf{Y}\} - aE\{\mathbf{X}\} \quad (3.56)$$

- Estimador óptimo sin restricciones $\hat{\mathbf{Y}} = g(\mathbf{X})$

$$g(\mathbf{X}) = E\{\mathbf{Y}|\mathbf{X}\} = \int_{-\infty}^{\infty} y f_Y(y|x) dy \quad (3.57)$$

3.7.3 Estimación mediante una constante

Se va a construir un estimador del tipo $\hat{\mathbf{Y}} = g(\mathbf{X}) = a$, es decir, estimar mediante una constante, con independencia de la observación recibida. Como es natural, este estimador será de una calidad cuestionable, pero es interesante conocerlo con el fin de poder ver que existen condiciones en las que un estimador tan simple es el óptimo.

Para hallar el valor de la constante que minimiza el error cuadrático medio procedemos de la forma siguiente:

$$\begin{aligned} \epsilon &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - a \\ E\{\epsilon^2\} &= E\left\{\left(\mathbf{Y} - \hat{\mathbf{Y}}\right)^2\right\} = E\left\{\left(\mathbf{Y} - a\right)^2\right\} \end{aligned}$$

Entonces el valor de a óptimo será (llamemos al óptimo a^*)

$$a^* = \arg \min_a E\left\{\left(\mathbf{Y} - a\right)^2\right\}$$

y por ello

$$\frac{dE\{(\mathbf{Y} - a)^2\}}{da} = E\{2(\mathbf{Y} - a)(-1)\}$$

con lo que igualando a cero y despejando a favor de a obtenemos el resultado $a^* = E\{\mathbf{Y}\}$.

En cuanto al error

$$\begin{aligned} E\{\epsilon\} &= E\{\mathbf{Y} - a^*\} = E\{\mathbf{Y}\} - a^* = 0 \\ E\{\epsilon^2\} &= E\{(\mathbf{Y} - a^*)^2\} = E\{(\mathbf{Y} - E\{\mathbf{Y}\})^2\} = \sigma_{\mathbf{Y}}^2 \end{aligned}$$

es decir, estamos cometiendo un error de media nula, y con una varianza tan grande como la de la varianza de la variable que pretendemos observar.

3.7.4 Estimación mediante una función lineal de la observación

En este caso, el estimador será del tipo $\hat{\mathbf{Y}} = a\mathbf{X} + b$, es decir, una función lineal del dato observado. Para hallar el valor de las constantes que minimizan el error cuadrático medio procedemos de una forma similar al caso anterior:

$$\begin{aligned} \epsilon &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - (a\mathbf{X} + b) \\ E\{\epsilon^2\} &= E\{(\mathbf{Y} - \hat{\mathbf{Y}})^2\} = E\{[\mathbf{Y} - (a\mathbf{X} + b)]^2\} \end{aligned} \tag{3.58}$$

Derivando con respecto a cada uno de los parámetros e igualando a cero

$$\frac{\partial E\{[\mathbf{Y} - (a\mathbf{X} + b)]^2\}}{\partial a} = E\{2(\mathbf{Y} - (a\mathbf{X} + b))(-\mathbf{X})\} = 0 \tag{3.59}$$

$$\frac{\partial E\{[\mathbf{Y} - (a\mathbf{X} + b)]^2\}}{\partial b} = E\{2(\mathbf{Y} - (a\mathbf{X} + b))(-1)\} = 0 \tag{3.60}$$

A partir de la ecuación (3.60) despejamos para obtener

$$E\{\mathbf{Y}\} - aE\{\mathbf{X}\} - b = 0 \rightarrow b^* = E\{\mathbf{Y}\} - aE\{\mathbf{X}\}$$

y sustituyendo en (3.59)

$$E\{\mathbf{XY}\} - aE\{\mathbf{X}^2\} - bE\{\mathbf{X}\} = 0$$

$$\begin{aligned} E\{\mathbf{XY}\} - aE\{\mathbf{X}^2\} - (E\{\mathbf{Y}\} - aE\{\mathbf{X}\})E\{\mathbf{X}\} &= 0 \\ E\{\mathbf{XY}\} - E\{\mathbf{X}\}E\{\mathbf{Y}\} - a(E\{\mathbf{X}^2\} - E^2\{\mathbf{X}\}) &= 0 \\ C_{\mathbf{XY}} - a\sigma_{\mathbf{X}}^2 = 0 \rightarrow a^* &= \frac{C_{\mathbf{XY}}}{\sigma_{\mathbf{X}}^2} \end{aligned}$$

Por lo tanto

$$a^* = \frac{C_{\mathbf{XY}}}{\sigma_{\mathbf{X}}^2} \quad (3.61)$$

$$b^* = E\{\mathbf{Y}\} - a^*E\{\mathbf{X}\} \quad (3.62)$$

Al respecto del error podemos hacer un análisis similar al hecho en la sección anterior. Concretamente:

$$\begin{aligned} E\{\boldsymbol{\epsilon}\} &= E\{\mathbf{Y} - (a^*\mathbf{X} + b^*)\} = 0 \quad (\text{compruébese}) \\ E\{\boldsymbol{\epsilon}^2\} = \sigma_{\boldsymbol{\epsilon}}^2 &= E\{[\mathbf{Y} - (a^*\mathbf{X} + b^*)]^2\} \end{aligned} \quad (3.63)$$

$$\begin{aligned} &= E\{[\mathbf{Y} - (a^*\mathbf{X} + E\{\mathbf{Y}\} - a^*E\{\mathbf{X}\})]^2\} \\ &= E\{[\mathbf{Y} - E\{\mathbf{Y}\} - a^*(\mathbf{X} - E\{\mathbf{X}\})]^2\} \\ &= \sigma_{\mathbf{Y}}^2 + (a^*)^2\sigma_{\mathbf{X}}^2 - 2a^*E\{(\mathbf{Y} - E\{\mathbf{Y}\})(\mathbf{X} - E\{\mathbf{X}\})\} \\ &= \sigma_{\mathbf{Y}}^2 + (a^*)^2\sigma_{\mathbf{X}}^2 - 2a^*C_{\mathbf{XY}} \\ &= \sigma_{\mathbf{Y}}^2 + \left(\frac{C_{\mathbf{XY}}}{\sigma_{\mathbf{X}}^2}\right)^2\sigma_{\mathbf{X}}^2 - 2\frac{C_{\mathbf{XY}}}{\sigma_{\mathbf{X}}^2}C_{\mathbf{XY}} \\ &= \sigma_{\mathbf{Y}}^2 - \frac{C_{\mathbf{XY}}^2}{\sigma_{\mathbf{X}}^2} = \sigma_{\mathbf{Y}}^2 \left(1 - \frac{C_{\mathbf{XY}}^2}{\sigma_{\mathbf{Y}}^2\sigma_{\mathbf{X}}^2}\right) \\ &= \sigma_{\mathbf{Y}}^2(1 - \rho_{\mathbf{XY}}^2) \end{aligned} \quad (3.64)$$

con $\rho_{\mathbf{XY}}$ el coeficiente de correlación entre las variables involucradas.

Nótese por lo tanto:

- Si $\rho_{\mathbf{XY}} = 0$ entonces $\sigma_{\boldsymbol{\epsilon}}^2 = \sigma_{\mathbf{Y}}^2$, y coincidiría con el caso de estimación mediante una constante. Nótese que, en efecto, $a^* = 0$ dado que la covarianza $C_{\mathbf{XY}}$ sería nula (ecuación 3.61), y así pues estimaríamos sólo con $b^* = E\{\mathbf{Y}\}$ (ecuación 3.62). Éste sería el caso de *ausencia de relación lineal entre las VAs*, esto es, no poder estimar *mejor* (en sentido MMSE) mediante una función lineal que mediante una constante (según indicamos ya en el ejercicio de la página 155).

- Si $|\rho_{\mathbf{XY}}| = 1$ entonces $\sigma_\epsilon^2 = 0$. La estimación en este caso no sería una VA, sino una constante, y el error sería nulo. Estaríamos estimando con precisión perfecta. Por tanto, nótese que cuando se produce esta situación, el conocimiento de \mathbf{X} hace que se conozca sin error ninguno la variable \mathbf{Y} . Estaríamos en el caso de ambas variables relacionadas mediante una función determinística, concretamente, mediante una función lineal.

3.7.5 Principio de ortogonalidad

La estimación lineal cumple el llamado *principio de ortogonalidad* que consiste en lo siguiente: considerando, como ya hemos dicho, la operación $E\{\mathbf{XY}\}$ como un producto escalar de las variables \mathbf{X} e \mathbf{Y} , aplicamos esta operación al error

$$\begin{aligned} E\{\epsilon \mathbf{X}\} &= E\{(\mathbf{Y} - (a^* \mathbf{X} + b^*)) \mathbf{X}\} \\ &= E\{\mathbf{XY}\} - (a^*) E\{\mathbf{X}^2\} - b^* E\{\mathbf{X}\} \\ &= E\{\mathbf{XY}\} - (a^*) E\{\mathbf{X}^2\} - E\{\mathbf{Y}\} E\{\mathbf{X}\} + a^* E^2\{\mathbf{X}\} \\ &= C_{\mathbf{XY}} - a^* \sigma_{\mathbf{X}}^2 = C_{\mathbf{XY}} - C_{\mathbf{XY}} = 0 \end{aligned} \quad (3.65)$$

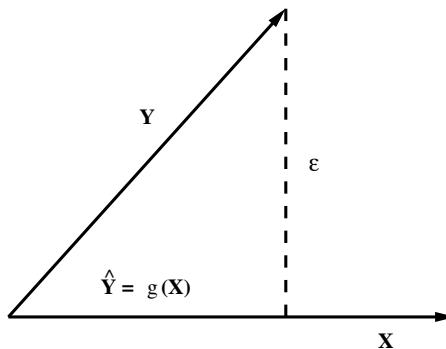


Figura 3.20: Ilustración del principio de ortogonalidad.

Puede por tanto concluirse que el estimador lineal es tal que el error cometido es ortogonal a la observación empleada. Podemos interpretar esto gráficamente como que estaríamos proyectando la variable a estimar sobre la observación, de modo que el error cometido sería la componente de \mathbf{Y} que se pierde al proyectar (véase figura 3.20).

Este principio es una forma equivalente de obtener las ecuaciones de los parámetros de la estimación. Repárese que, de hecho, la expresión (3.65) es equivalente a la ecuación (3.59).

3.7.6 Estimación óptima sin restricciones

Se pretende ahora construir el estimador óptimo sin ninguna restricción funcional sobre el mismo. Por ello, tendremos que acudir a su definición genérica mediante la función $g(\mathbf{X})$ y optimizar sobre dicha función

$$\begin{aligned}\epsilon &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - g(\mathbf{X}) \\ E\{\epsilon^2\} &= E\{(Y - g(X))^2\} = \int \int (y - g(x))^2 f_{\mathbf{XY}}(x, y) dx dy \\ &= \int \int (y - g(x))^2 f_{\mathbf{Y}}(y|x) f_{\mathbf{X}}(x) dx dy\end{aligned}\quad (3.66)$$

Entonces el estimador óptimo será (llamémosle $g^*(\mathbf{X})$)

$$g^*(\mathbf{X}) = \arg \min_{g(\mathbf{X})} E\{\epsilon^2\} \quad (3.67)$$

Acorde con la ecuación (3.66) podemos escribir

$$\int f_{\mathbf{X}}(x) dx \int (y - g(x))^2 f_{\mathbf{Y}}(y|x) dy = \int \varphi(g(x)) f_{\mathbf{X}}(x) dx \quad (3.68)$$

Minimizar esta ecuación con respecto a $g(x)$ equivale a minimizar $\varphi(g(x), x)$ con respecto a $g(x)$. Por ello, podemos escribir

$$\begin{aligned}\frac{d\varphi(g(x), x)}{dg(x)} &= \frac{\partial}{\partial g(x)} \int (y - g(x))^2 f_{\mathbf{Y}}(y|x) dy \\ &= (-2) \int (y - g(x)) f_{\mathbf{Y}}(y|x) dy \\ \frac{d\varphi(g(x))}{dg(x)} &= 0 \rightarrow \int (y - g(x)) f_{\mathbf{Y}}(y|x) dy = 0 \\ \int y f_{\mathbf{Y}}(y|x) dy &= g(x) \int f_{\mathbf{Y}}(y|x) dy = g^*(x) \\ g^*(x) &= \int (y f_{\mathbf{Y}}(y|x) dy) = E\{\mathbf{Y}|x\}\end{aligned}\quad (3.69)$$

Por lo tanto, podemos ver que el estimador óptimo sin restricciones es $g^*(\mathbf{X}) = E\{\mathbf{Y}|\mathbf{X}\}$, es decir, la esperanza de la variable \mathbf{Y} condicionada al valor que haya tomado la variable \mathbf{X} . Nótese que en el caso en que las variables sean independientes, dado que $f_{\mathbf{Y}}(y|x) = f_{\mathbf{Y}}(y)$, entonces $E\{\mathbf{Y}|\mathbf{X}\} = E\{\mathbf{Y}\}$ de forma que el estimador óptimo, en este caso, coincidiría con el estimador mediante una constante.

Como resumen, podemos decir que

1. En general $\sigma_{\epsilon \text{ constante}}^2 \geq \sigma_{\epsilon \text{ lineal}}^2 \geq \sigma_{\epsilon \text{ no lineal}}^2$
2. Si las VAs son incorreladas $\sigma_{\epsilon \text{ constante}}^2 = \sigma_{\epsilon \text{ lineal}}^2$
3. Si las VAs son independientes $\sigma_{\epsilon \text{ constante}}^2 = \sigma_{\epsilon \text{ no lineal}}^2$

Ejercicio: Obtenga la media $E\{\epsilon\}$ y la varianza σ_{ϵ}^2 del error de estimación $\epsilon = \mathbf{Y} - E\{\mathbf{Y}|\mathbf{X}\}$.

Solución:

Al respecto del valor medio:

$$\begin{aligned}
 E\{\epsilon\} &= E\{\mathbf{Y} - E\{\mathbf{Y}|\mathbf{X}\}\} = E\{\mathbf{Y}\} - E\{E\{\mathbf{Y}|\mathbf{X}\}\} \\
 &= E\{\mathbf{Y}\} - E\{g(\mathbf{X})\} \\
 &= E\{\mathbf{Y}\} - \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{\mathbf{Y}}(y|x) dy \right) f_{\mathbf{X}}(x) dx \\
 &= E\{\mathbf{Y}\} - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{\mathbf{XY}}(x,y) dx dy \\
 &= E\{\mathbf{Y}\} - \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} f_{\mathbf{XY}}(x,y) dx \\
 &= E\{\mathbf{Y}\} - \int_{-\infty}^{\infty} y f_{\mathbf{Y}}(y) dy \\
 &= E\{\mathbf{Y}\} - E\{\mathbf{Y}\} = 0
 \end{aligned}$$

Respecto de la varianza:

$$\begin{aligned}
 \sigma_{\epsilon}^2 = E\{\epsilon^2\} &= E\{(\mathbf{Y} - E\{\mathbf{Y}|\mathbf{X}\})^2\} \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - E\{\mathbf{Y}|x\})^2 f_{\mathbf{XY}}(x,y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - E\{\mathbf{Y}|x\})^2 f_{\mathbf{Y}}(y|x) f_{\mathbf{X}}(x) dx dy \\
 &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(x) dx \int_{-\infty}^{\infty} (y - E\{\mathbf{Y}|x\})^2 f_{\mathbf{Y}}(y|x) dy \\
 &= \int_{-\infty}^{\infty} \sigma_{\mathbf{Y}|x}^2 f_{\mathbf{X}}(x) dx \\
 &= E\{\sigma_{\mathbf{Y}|\mathbf{X}}^2\}
 \end{aligned}$$



Capítulo 4

Variable N-dimensional

4.1 Conceptos básicos

Se pretende ahora generalizar al caso N-dimensional los conceptos vistos para $N = 2$ en el capítulo anterior. En este capítulo consideraremos que las VAs vienen dadas mediante un vector $\bar{\mathbf{X}} = [\mathbf{X}_1, \dots, \mathbf{X}_N]^T$ de N componentes. Esta notación vectorial nos permitirá hacer, en algunos casos, operaciones con vectores y matrices de forma sencilla.

El hecho de trabajar con un número genérico N de variables en vez de con $N = 2$ no supone cambio conceptual alguno; simplemente añade complejidad a las operaciones, pero los conceptos básicos y, en particular, el concepto de comportamiento conjunto, se extienden directamente del caso de $N = 2$ a un caso de dimensión arbitraria. Repárese, no obstante, que los momentos definidos en las expresiones (3.46) y (3.47) de la sección 3.6.2, así como el coeficiente de correlación, son conceptos implícitamente bidimensionales, de los cuales aquí también haremos uso de ellos.

Así pues podemos recordar y extender conceptos generales de las VAs de la forma que sigue:

1.

$$\begin{aligned} F_{\bar{\mathbf{X}}}(\bar{x}) &= F_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N) \\ &= P(\mathbf{X}_1 \leq x_1, \mathbf{X}_2 \leq x_2, \dots, \mathbf{X}_N \leq x_N) \end{aligned}$$

2. $f_{\bar{\mathbf{X}}}(\bar{x}) = f_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N) = \frac{\partial^N F_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N)}{\partial x_1 \partial x_2 \dots \partial x_N}$

3. $P(\bar{\mathbf{X}} \in D) = \int_D f_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N$

con D una región arbitraria en el hiperespacio N -dimensional.

4. Reducción del número de variables de las funciones. Si consideramos, por ejemplo, que $N = 4$, podemos escribir

$$F_{\mathbf{X}_1 \mathbf{X}_3}(x_1, x_3) = F_{\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3 \mathbf{X}_4}(x_1, \infty, x_3, \infty) \quad (4.1)$$

$$f_{\mathbf{X}_1 \mathbf{X}_4}(x_1, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3 \mathbf{X}_4}(x_1, x_2, x_3, x_4) dx_2 dx_3 \quad (4.2)$$

5. Transformaciones: si ahora se verifica que $\bar{\mathbf{Y}} = \bar{g}(\bar{\mathbf{X}})$, donde por $\bar{g}()$ representamos en modo compacto un vector $\bar{g}() = [g_1(), g_2(), \dots, g_N()]^T$ de N transformaciones de $R^N \rightarrow R$, entonces el teorema fundamental extendido a N dimensiones se puede escribir

$$f_{\mathbf{Y}_1 \dots \mathbf{Y}_N}(y_1, \dots, y_N) = \sum_{i=1}^M \frac{f_{\mathbf{X}_1 \dots \mathbf{X}_N}(x_1, \dots, x_N)}{|J|} \Big|_{(x_1^i, \dots, x_N^i)} \quad (4.3)$$

donde el jacobiano de la transformación es ahora

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_N}{\partial x_1} & \dots & \frac{\partial g_N}{\partial x_N} \end{vmatrix} \quad (4.4)$$

y los puntos (x_1^i, \dots, x_N^i) , $i = (1, \dots, M)$ son los M puntos del dominio origen que se convierten en el punto (y_1, \dots, y_N) donde estamos construyendo la función de densidad de la VA $\bar{\mathbf{Y}}$.

Si el número de transformaciones g_i fuese $k < N$, entonces podríamos hacer uso de $N - k$ variables auxiliares para poder aplicar el teorema fundamental. Con ello, extenderíamos a N VAs el método de la VA auxiliar visto en la sección 3.5.4.

6. Independencia: también este concepto se extiende de una forma natural al caso de N variables. Si éstas son independientes se verifica que existe independencia entre los sucesos definidos a partir de las mismas, en particular, entre los sucesos $\{\mathbf{X}_i \leq x_i\}$ y $\{\mathbf{X}_j \leq x_j\} \forall i \neq j$. Por ello podremos escribir¹

$$F_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N) = \prod_{i=1}^N F_{\mathbf{X}_i}(x_i) \quad (4.5)$$

¹En términos estrictos y según la notación de la sección 1.7, el suceso en el experimento compuesto asociado al suceso $\{\mathbf{X}_i \leq x_i\}$ debería escribirse $S_1 \times \dots \times S_{i-1} \times \{\mathbf{X}_i \leq x_i\} \times S_{i+1} \times \dots \times S_N$. Emplearemos la notación simplificada, no obstante.

$$f_{\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_N}(x_1, x_2, \dots, x_N) = \prod_{i=1}^N f_{\mathbf{X}_i}(x_i) \quad (4.6)$$

Nótese que esta expresión implica independencia entre cada dos variables, entre cada tres etc ..., hasta la independencia conjunta de las N variables. El motivo no es otro que a partir de las expresiones anteriores podemos reducir a nuestro gusto el número de VAs involucradas sin más que aplicar la operativa indicada en las expresiones (4.1) y (4.2) y la igualdad se seguiría manteniendo.

7. Distribuciones condicionadas: en el capítulo anterior obtuvimos que

$$f_{\mathbf{XY}}(x, y) = f_{\mathbf{X}}(x|y)f_{\mathbf{Y}}(y) \quad (4.7)$$

(expresiones 3.24 y 3.25). Para el caso de N variables las relaciones anteriores se generalizan de una manera similar a como se escribía la probabilidad de la intersección de N sucesos en el primer capítulo (ecuación 1.25). Allí escribíamos

$$\begin{aligned} P(A_1, A_2, \dots, A_N) &= P(A_1|A_2, \dots, A_N)P(A_2, \dots, A_N) = \\ &= P(A_1|A_2, \dots, A_N)P(A_2|A_3 \dots, A_N) \cdot \\ &\quad P(A_3|A_4 \dots A_N) \cdots P(A_{N-1}|A_N)P(A_N) \end{aligned} \quad (4.8)$$

o bien, de forma equivalente

$$\begin{aligned} P(A_1, A_2, \dots, A_N) &= P(A_1, A_2|A_3 \dots, A_N)P(A_3, \dots, A_N) = \\ &= P(A_1, A_2|A_3 \dots, A_N) \cdot \\ &\quad P(A_3|A_4 \dots, A_N) \cdots P(A_{N-1}|A_N)P(A_N) \end{aligned} \quad (4.9)$$

Siguiendo este planteamiento, podemos escribir de forma similar²,

$$\begin{aligned} f(x_1, x_2, \dots, x_N) &= f(x_1|x_2, \dots, x_N)f(x_2, \dots, x_N) = \\ &= f(x_1|x_2, \dots, x_N)f(x_2|x_3 \dots, x_N) \times \\ &\quad f(x_3|x_4 \dots x_N) \cdots f(x_{N-1}|x_N)f(x_N) \end{aligned} \quad (4.10)$$

Como norma general:

²Para no sobrecargar la notación quitamos los subíndices a las funciones de densidad por ser obvio a qué variables nos referimos.

- (a) Para eliminar una variable a la izquierda de la línea del condicionante, integramos con respecto a ella. Por ejemplo, para pasar de $f(x_1, x_2|x_3)$ a $f(x_1|x_3)$, integramos con respecto a x_2 . En efecto, véase que

$$f(x_1, x_2|x_3) = \frac{f(x_1, x_2, x_3)}{f(x_3)}$$

y, como sabemos, para eliminar la dependencia de x_2 en la variable conjunta, integramos respecto a ella. Así pues

$$\int f(x_1, x_2|x_3)dx_2 = \frac{\int f(x_1, x_2, x_3)dx_2}{f(x_3)} = \frac{f(x_1, x_3)}{f(x_3)} = f(x_1|x_3)$$

- (b) Para eliminar la influencia de una variable que se encuentra a la derecha del condicionante, multiplicaríamos por la función de densidad de esa variable condicionada al resto de las VAs condicionantes, e integraríamos con respecto a ella. Por ejemplo, para eliminar la influencia de x_2 en $f(x_1|x_2, x_3)$ dado que

$$f(x_1|x_2, x_3) = \frac{f(x_1, x_2, x_3)}{f(x_2, x_3)}$$

si ahora multiplicamos por $f(x_2|x_3)$, tendríamos

$$f(x_1|x_2, x_3)f(x_2|x_3) = \frac{f(x_1, x_2, x_3)}{f(x_2, x_3)} \frac{f(x_2, x_3)}{f(x_3)}$$

de forma que ya aparece la dependencia de x_2 sólo en la función conjunta. Por tanto, integrando respecto a ella tendríamos

$$\begin{aligned} \int f(x_1|x_2, x_3)f(x_2|x_3)dx_2 &= \int \frac{f(x_1, x_2, x_3)}{f(x_2, x_3)} \frac{f(x_2, x_3)}{f(x_3)} dx_2 \\ &= \int \frac{f(x_1, x_2, x_3)}{f(x_3)} dx_2 \\ &= \frac{f(x_1, x_3)}{f(x_3)} = f(x_1|x_3) \end{aligned} \tag{4.11}$$

8. Respecto de la estimación MMSE: si la estimación es lineal, en el caso de N variables se desearía estimar la VA \mathbf{Y} como función lineal de $\mathbf{X}_1, \dots, \mathbf{X}_N$, esto es,

$$\hat{\mathbf{Y}} = g(\mathbf{X}_1, \dots, \mathbf{X}_N) = \sum_{i=1}^N a_i \mathbf{X}_i + b \tag{4.12}$$

Los coeficientes a_i , $i=\{1,\dots,N\}$ y b se obtienen siguiendo un procedimiento similar al empleado para obtener las ecuaciones (3.59) y (3.60) de la sección 3.7.4. En este caso el número de ecuaciones resultante será de $N+1$, coincidente con el número de incógnitas en la expresión (4.12).

Si el estimador es el óptimo sin restricciones, la extensión al caso de N variables de lo visto en la sección 3.7.6 es inmediata: debe calcularse la esperanza de \mathbf{Y} condicionada a todas las variables observadas. Concretamente, la expresión (3.69) ahora se generaliza de la forma

$$\begin{aligned}\hat{\mathbf{Y}} &= g(\mathbf{X}_1, \dots, \mathbf{X}_N) = \int y f_{\mathbf{Y}}(y | \mathbf{X}_1, \dots, \mathbf{X}_N) dy \\ &= E\{\mathbf{Y} | \mathbf{X}_1, \dots, \mathbf{X}_N\}\end{aligned}\quad (4.13)$$

4.2 Esperanzas matemáticas

El operador esperanza lleva a cabo la misma función que para el caso de $N = 2$. No obstante, para un N genérico, es cómodo ordenar las esperanzas de cada variable por separado, y las esperanzas del producto de cada dos variables, respectivamente, en vectores y matrices. Con esta ordenación, aparte de poder expresar la información de forma más compacta, se podrá hacer operaciones de forma más rápida como veremos en la sección siguiente.

Por ello, podemos definir:

1. Vector de medias:

$$\bar{\eta}_{\bar{\mathbf{X}}} = E\{\bar{\mathbf{X}}\} = \begin{bmatrix} \eta_{\mathbf{X}_1} \\ \eta_{\mathbf{X}_2} \\ \vdots \\ \eta_{\mathbf{X}_N} \end{bmatrix} \quad (4.14)$$

2. Matriz de correlación:

$$\bar{R}_{\bar{\mathbf{X}}} = E\{\bar{\mathbf{X}}\bar{\mathbf{X}}^T\} = \begin{bmatrix} E\{\mathbf{X}_1\mathbf{X}_1\} & E\{\mathbf{X}_1\mathbf{X}_2\} & \cdots & E\{\mathbf{X}_1\mathbf{X}_N\} \\ E\{\mathbf{X}_2\mathbf{X}_1\} & E\{\mathbf{X}_2\mathbf{X}_2\} & \cdots & E\{\mathbf{X}_2\mathbf{X}_N\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{\mathbf{X}_N\mathbf{X}_1\} & E\{\mathbf{X}_N\mathbf{X}_2\} & \cdots & E\{\mathbf{X}_N\mathbf{X}_N\} \end{bmatrix}$$

$$= \begin{bmatrix} E\{\mathbf{X}_1^2\} & R_{\mathbf{X}_1 \mathbf{X}_2} & \cdots & R_{\mathbf{X}_1 \mathbf{X}_N} \\ R_{\mathbf{X}_2 \mathbf{X}_1} & E\{\mathbf{X}_2^2\} & \cdots & R_{\mathbf{X}_2 \mathbf{X}_N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\mathbf{X}_N \mathbf{X}_1} & R_{\mathbf{X}_N \mathbf{X}_2} & \cdots & E\{\mathbf{X}_N^2\} \end{bmatrix} \quad (4.15)$$

Nótese, pues, que esta matriz no es otra cosa que una tabla de N^2 elementos, donde los términos de la diagonal principal son los VCMs de cada variable y el resto de los elementos son las correlaciones cruzadas entre cada dos variables distintas. Obviamente, para variables reales esta matriz es simétrica.

3. Matriz de covarianza: matriz similar a la anterior, pero ahora el término c_{ij} es la covarianza entre \mathbf{X}_i y \mathbf{X}_j , es decir, $c_{ij} = E\{(\mathbf{X}_i - \eta_{\mathbf{X}_i})(\mathbf{X}_j - \eta_{\mathbf{X}_j})\}$. Ello trae consigo que, en particular, los elementos de la diagonal principal sean las varianzas de cada una de las variables. Así pues³

$$\bar{C}_{\bar{\mathbf{X}}} = \begin{bmatrix} \sigma_{\mathbf{X}_1}^2 & C_{\mathbf{X}_1 \mathbf{X}_2} & \cdots & C_{\mathbf{X}_1 \mathbf{X}_N} \\ C_{\mathbf{X}_2 \mathbf{X}_1} & \sigma_{\mathbf{X}_2}^2 & \cdots & C_{\mathbf{X}_2 \mathbf{X}_N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{\mathbf{X}_N \mathbf{X}_1} & C_{\mathbf{X}_N \mathbf{X}_2} & \cdots & \sigma_{\mathbf{X}_N}^2 \end{bmatrix} \quad (4.16)$$

En este caso la matriz es una tabla de N^2 números en los que los elementos de la diagonal principal son las varianzas de cada variable y el resto son las covarianzas entre cada dos variables distintas. Esta matriz también es simétrica para VAs reales.

De manera equivalente al caso bidimensional:

- El vector $\bar{\mathbf{X}}$ está formado por componentes ortogonales si $R_{\mathbf{X}_i \mathbf{X}_j} = 0 \quad \forall i \neq j$.
- El vector $\bar{\mathbf{X}}$ está formado por componentes incorreladas si $C_{\mathbf{X}_i \mathbf{X}_j} = 0 \quad \forall i \neq j$.

³Según la expresión (4.16) deberíamos llamar a esta matriz *de varianzas y covarianzas*, así como a la matriz de la expresión (4.15) matriz *de valores cuadráticos medios y correlaciones*. Aquí se ha empleado, sin embargo, la denominación más común en el área de Teoría de la Señal, aun a riesgo de pecar de anglicista.

El interés de que las componentes sean ortogonales y/o incorreladas viene dado por la nulidad de las esperanzas cruzadas en un problema de múltiples variables. Concretamente:

- Si las variables \mathbf{X}_i son ortogonales, entonces se verifica que el VCM de la variable $\mathbf{Z} = \sum_{i=1}^N \mathbf{X}_i$ es igual a la suma de los VCMs. En efecto

$$\begin{aligned}
E\{\mathbf{Z}^2\} &= E\left\{\left(\sum_{i=1}^N \mathbf{X}_i\right)^2\right\} = E\left\{\sum_{i=1}^N \sum_{j=1}^N \mathbf{X}_i \mathbf{X}_j\right\} \\
&= E\left\{\sum_{i=1}^N \mathbf{X}_i^2\right\} + E\left\{\sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{X}_i \mathbf{X}_j\right\} \\
&= \sum_{i=1}^N E\{\mathbf{X}_i^2\} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N E\{\mathbf{X}_i \mathbf{X}_j\} \\
&= \sum_{i=1}^N E\{\mathbf{X}_i^2\} + \sum_{i=1}^N \sum_{j=1}^N R_{\mathbf{X}_i \mathbf{X}_j} \\
&\stackrel{\text{Ortog.}}{=} \sum_{i=1}^N E\{\mathbf{X}_i^2\}
\end{aligned} \tag{4.17}$$

- Si las variables fuesen incorreladas, se podrían obtener conclusiones similares, pero en este caso en relación con la varianza. Concretamente, la varianza de la suma sería igual a la suma de las varianzas. En efecto:

$$\begin{aligned}
E\{(\mathbf{Z} - \eta_Z)^2\} &= E\left\{\left(\sum_{i=1}^N \mathbf{X}_i - \sum_{i=1}^N \eta_{\mathbf{X}_i}\right)^2\right\} \\
&= E\left\{\left(\sum_{i=1}^N (\mathbf{X}_i - \eta_{\mathbf{X}_i})\right)^2\right\} \\
&= E\left\{\sum_{i=1}^N \sum_{j=1}^N (\mathbf{X}_i - \eta_{\mathbf{X}_i})(\mathbf{X}_j - \eta_{\mathbf{X}_j})\right\} \\
&= E\left\{\sum_{i=1}^N (\mathbf{X}_i - \eta_{\mathbf{X}_i})^2\right\} +
\end{aligned}$$

$$\begin{aligned}
& E \left\{ \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N (\mathbf{X}_i - \eta_{\mathbf{X}_i})(\mathbf{X}_j - \eta_{\mathbf{X}_j}) \right\} \\
&= \sum_{i=1}^N E \left\{ (\mathbf{X}_i - \eta_{\mathbf{X}_i})^2 \right\} + \\
&\quad \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N E \left\{ (\mathbf{X}_i - \eta_{\mathbf{X}_i})(\mathbf{X}_j - \eta_{\mathbf{X}_j}) \right\} \\
&= \sum_{i=1}^N E \left\{ (\mathbf{X}_i - \eta_{\mathbf{X}_i})^2 \right\} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N C_{\mathbf{X}_i \mathbf{X}_j} \\
&\stackrel{\text{Incorr.}}{=} \sum_{i=1}^N \sigma_{\mathbf{X}_i}^2
\end{aligned} \tag{4.18}$$

4.3 Variables conjuntamente gaussianas

N variables aleatorias son conjuntamente gaussianas si la función de densidad conjunta de las mismas se puede escribir de la forma que sigue:

$$\begin{aligned}
f_{\mathbf{X}_1 \dots \mathbf{X}_N}(x_1, \dots, x_N) &= f_{\bar{\mathbf{X}}}(\bar{x}) \\
&= \frac{1}{|\bar{C}_{\bar{\mathbf{X}}}|^{\frac{1}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\bar{x} - \bar{\eta}_{\bar{\mathbf{X}}})^T \bar{C}_{\bar{\mathbf{X}}}^{-1} (\bar{x} - \bar{\eta}_{\bar{\mathbf{X}}})} \tag{4.19}
\end{aligned}$$

donde $\bar{x} = [x_1, x_2, \dots, x_N]^T$ y $\bar{\eta}_{\bar{\mathbf{X}}}$ y $\bar{C}_{\bar{\mathbf{X}}}$ son, respectivamente, el vector de medias (expresión 4.14) y la matriz de covarianzas (expresión 4.16) de la variable N-dimensional.

Si N VAs son conjuntamente gaussianas se verifica [12]:

1. que $\mathbf{X}_i \sim N(\eta_{\mathbf{X}_i}, \sigma_{\mathbf{X}_i}^2)$, $\forall i = 1, \dots, N$;
2. que cualquier subconjunto de $k < N$ VAs extraído del vector $\bar{\mathbf{X}}$ también constituye un vector de VAs conjuntamente gaussianas;
3. que la función de densidad de cualquier conjunto de $k < N$ VAs del vector $\bar{\mathbf{X}}$, condicionada al conjunto de $N - k$ VAs restantes, es también conjuntamente gaussiana.

El estudio de variables gaussianas y conjuntamente gaussianas es particularmente interesante por diversas razones. Entre ellas:

1. Como se acaba de ver, la función de densidad conjunta de la variable gaussiana N -dimensional viene completamente caracterizada por el vector de medias y por la matriz de covarianzas. Por tanto, la función de densidad conjunta (a partir de la cual, recuérdese, se pueden calcular momentos centrales y no centrales de órdenes arbitrarios y que involucren a tantas variables como deseemos) viene dada, exclusivamente, por el momento no central m_1 de cada VA (véase la sección 2.7.1), y por el momento central de órdenes $r = 1$ y $s = 1$ para cada par de VAs (véase sección 3.6.2).
2. Según se verá (sección 4.5.1), el Teorema del Límite Central hace que en la naturaleza muchos problemas de índole estocástico tiendan a comportarse de forma gaussiana.
3. Los estimadores lineal y no lineal óptimos coinciden, es decir, para el caso de variables conjuntamente gaussianas las expresiones resultantes de (4.12) y (4.13) son idénticas. Esto es debido a que la función de densidad conjunta (4.19) se construye haciendo uso, únicamente, de medias y covarianzas, que es, a su vez, la información de que hace uso el estimador lineal (4.12).
4. En la gaussiana N -dimensional incorrelación también implica independencia. Nótese que si las variables \mathbf{X}_i están incorreladas entre sí entonces la matriz $\overline{\overline{C_X}}$ (ecuación 4.16) es una matriz diagonal. Esto hace en la expresión (4.19)
 - que el determinante de la matriz $\overline{\overline{C_X}}$ pueda escribirse como el producto de las varianzas $\sigma_{\mathbf{X}_i}^2$ de las N variables involucradas;
 - que la forma cuadrática que acompaña la exponencial no tenga términos cruzados del tipo $(x_i - \eta_{\mathbf{X}_i})(x_j - \eta_{\mathbf{X}_j})$, $i \neq j$. Por ello, la exponencial se puede escribir como productos de exponentiales con exponente de la forma

$$\exp \left(-\frac{(x_i - \eta_{\mathbf{X}_i})^2}{2\sigma_{\mathbf{X}_i}^2} \right)$$

Aplicando estas simplificaciones es inmediato comprobar que $f_{\overline{\mathbf{X}}}(\overline{x}) = \prod_{i=1}^N f_{\mathbf{X}_i}(x_i)$, de forma que, como hemos dicho, incorrelación implica, para este tipo de variables, independencia.

5. La transformación lineal $\bar{\mathbf{Y}} = \bar{\mathbf{A}}\bar{\mathbf{X}}$ del vector $\bar{\mathbf{X}}$ de VAs conjuntamente gaussianas da lugar a un vector $\bar{\mathbf{Y}}$ de VAs también conjuntamente gaussianas si la matriz $\bar{\mathbf{Y}}$ es cuadrada y de rango completo N .

Ejercicio 1: Demuéstrese por medio del Teorema Fundamental N-dimensional (expresión 4.3) que si $\bar{\mathbf{A}}$ es una matriz cuadrada de $N \times N$ y rango completo, entonces la VA generada mediante $\bar{\mathbf{Y}} = \bar{\mathbf{A}}\bar{\mathbf{X}}$ es un vector de VAs conjuntamente gaussianas si el vector $\bar{\mathbf{X}}$ lo es. Obtenga la expresión de $\bar{\eta}_{\bar{\mathbf{Y}}}$ y $\bar{\mathbf{C}}_{\bar{\mathbf{Y}}}$ en función de $\bar{\eta}_{\bar{\mathbf{X}}}$ y $\bar{\mathbf{C}}_{\bar{\mathbf{X}}}$.

Solución: Si $\bar{\mathbf{A}}$ es de rango completo entonces la combinación lineal asociada a ella es biyectiva. Por ello, podemos escribir:

$$\begin{aligned} f_{\mathbf{Y}_1 \dots \mathbf{Y}_N}(y_1, \dots, y_N) &= \frac{f_{\mathbf{X}_1 \dots \mathbf{X}_N}(x_1, \dots, x_N)}{|J|} \Big|_{\bar{x}=\bar{\mathbf{A}}^{-1}\bar{y}} = \\ &= \frac{\frac{1}{|\bar{\mathbf{C}}_{\bar{\mathbf{X}}}|^{\frac{1}{2}}(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\bar{\mathbf{A}}^{-1}\bar{y}-\bar{\eta}_{\bar{\mathbf{X}}})^T \bar{\mathbf{C}}_{\bar{\mathbf{X}}}^{-1} (\bar{\mathbf{A}}^{-1}\bar{y}-\bar{\eta}_{\bar{\mathbf{X}}})}}{|\bar{\mathbf{A}}|} \\ &= f_{\bar{\mathbf{Y}}}(\bar{y}) \end{aligned}$$

Denominando por conveniencia

$$\bar{\eta} = \bar{\mathbf{A}}\bar{\eta}_{\bar{\mathbf{X}}} \quad (4.20)$$

entonces podemos escribir

$$\begin{aligned} f_{\bar{\mathbf{Y}}}(\bar{y}) &= \frac{1}{|\bar{\mathbf{C}}_{\bar{\mathbf{X}}}|^{\frac{1}{2}}|\bar{\mathbf{A}}|(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\bar{\mathbf{A}}^{-1}\bar{y}-\bar{\mathbf{A}}^{-1}\bar{\eta})^T \bar{\mathbf{C}}_{\bar{\mathbf{X}}}^{-1} (\bar{\mathbf{A}}^{-1}\bar{y}-\bar{\mathbf{A}}^{-1}\bar{\eta})} \\ &= \frac{1}{|\bar{\mathbf{C}}_{\bar{\mathbf{X}}}|^{\frac{1}{2}}|\bar{\mathbf{A}}|(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\bar{\mathbf{A}}^{-1}(\bar{y}-\bar{\eta}))^T \bar{\mathbf{C}}_{\bar{\mathbf{X}}}^{-1} (\bar{\mathbf{A}}^{-1}(\bar{y}-\bar{\eta}))} \\ &= \frac{1}{|\bar{\mathbf{C}}_{\bar{\mathbf{X}}}|^{\frac{1}{2}}|\bar{\mathbf{A}}|(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\bar{y}-\bar{\eta})^T (\bar{\mathbf{A}}^{-1})^T \bar{\mathbf{C}}_{\bar{\mathbf{X}}}^{-1} \bar{\mathbf{A}}^{-1} (\bar{y}-\bar{\eta})} \end{aligned} \quad (4.21)$$

Si denominamos ahora

$$\bar{\mathbf{C}} = \bar{\mathbf{A}} \bar{\mathbf{C}}_{\bar{\mathbf{X}}} \bar{\mathbf{A}}^T \quad (4.22)$$

y recordando que

$$\begin{aligned} (\overline{\overline{A}}\overline{\overline{B}})^T &= \overline{\overline{B}}^T\overline{\overline{A}}^T \\ (\overline{\overline{B}}^{-1})^T &= (\overline{\overline{B}}^T)^{-1} \\ |\overline{\overline{A}}\overline{\overline{B}}| &= |\overline{\overline{A}}||\overline{\overline{B}}| \end{aligned}$$

(suponiendo en el último caso que ambas matrices son cuadradas y de la misma dimensión), entonces

$$\overline{\overline{C}}^{-1} = (\overline{\overline{A}}^T)^{-1}\overline{\overline{C}}_{\overline{\mathbf{X}}}^{-1}\overline{\overline{A}}^{-1} = (\overline{\overline{A}}^{-1})^T\overline{\overline{C}}_{\overline{\mathbf{X}}}^{-1}\overline{\overline{A}}^{-1}$$

y también se verifica que

$$|\overline{\overline{C}}| = |\overline{\overline{A}}\overline{\overline{C}}_{\overline{\mathbf{X}}}\overline{\overline{A}}^T| = |\overline{\overline{A}}||\overline{\overline{C}}_{\overline{\mathbf{X}}}||\overline{\overline{A}}^T| = |\overline{\overline{C}}_{\overline{\mathbf{X}}}||\overline{\overline{A}}|^2$$

de forma que volviendo a la ecuación (4.21) podemos escribir

$$f_{\mathbf{Y}_1 \dots \mathbf{Y}_N}(y_1, \dots, y_N) = \frac{1}{|\overline{\overline{C}}|^{\frac{1}{2}}(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\overline{\overline{y}}-\overline{\overline{\eta}})^T\overline{\overline{C}}^{-1}(\overline{\overline{y}}-\overline{\overline{\eta}})} \quad (4.23)$$

por lo que concluimos que, en efecto, el vector $\overline{\mathbf{Y}}$ es un vector de variables conjuntamente gaussianas.

Asimismo, comparando la expresión obtenida con la indicada por definición en la ecuación (4.19) vemos que los parámetros $\overline{\eta}_{\overline{\mathbf{Y}}}$ y $\overline{\overline{C}}_{\overline{\mathbf{Y}}}$, coinciden con los definidos por comodidad en las expresiones (4.20) y (4.22), por tanto

$$\begin{aligned} \overline{\eta}_{\overline{\mathbf{Y}}} &= \overline{\overline{A}}\overline{\eta}_{\overline{\mathbf{X}}} \\ \overline{\overline{C}}_{\overline{\mathbf{Y}}} &= \overline{\overline{A}}\overline{\overline{C}}_{\overline{\mathbf{X}}}\overline{\overline{A}}^T \end{aligned}$$



Ejercicio 2: Razónese que si $\overline{\overline{A}}$ es una matriz de dimensiones $M \times N$, con $M < N$, y de rango M , entonces la VA generada mediante $\overline{\mathbf{Y}} = \overline{\overline{A}}\overline{\mathbf{X}}$ es un vector de VAs conjuntamente gaussianas. Obtenga la expresión de $\overline{\eta}_{\overline{\mathbf{Y}}}$ y $\overline{\overline{C}}_{\overline{\mathbf{Y}}}$ en función de $\overline{\eta}_{\overline{\mathbf{X}}}$ y $\overline{\overline{C}}_{\overline{\mathbf{X}}}$.

Solución: La respuesta es afirmativa y para ello podemos razonar:

- Si $M < N$ entonces tenemos un problema de M VAs en dominio destino y N VAs en dominio origen.
- Podemos construir $N - M$ VAs $\bar{\mathbf{Z}} = [\mathbf{Z}_1 \dots \mathbf{Z}_{N-M}]^T$ auxiliares, y construir la transformación lineal

$$\begin{bmatrix} \bar{\mathbf{Y}} \\ \bar{\mathbf{Z}} \end{bmatrix} = \bar{\bar{B}}\bar{\mathbf{X}}$$

con $\bar{\bar{B}}$ una matriz cuadrada de rango completo N . Una elección razonable es que el vector de VAs coincida, por ejemplo, con las $N - M$ últimas VAs, es decir que

$$\bar{\mathbf{Z}} = [\mathbf{Z}_1 \dots \mathbf{Z}_{M-N}]^T = [\mathbf{X}_{M+1} \dots \mathbf{X}_N]^T$$

lo cual puede conseguirse fácilmente haciendo que

$$\bar{\bar{B}} = \begin{pmatrix} \bar{\bar{A}} \\ \bar{\bar{0}} \quad \bar{\bar{I}}_{N-M} \end{pmatrix}$$

con $\bar{\bar{0}}$ la matriz nula de dimensiones $(N - M) \times M$ e $\bar{\bar{I}}_{N-M}$ la matriz identidad de dimensión $N - M$.

- Segundo se ha visto en el ejercicio anterior, el vector $[\bar{\mathbf{Y}} \quad \bar{\mathbf{Z}}]^T$ es un vector de VAs conjuntamente gaussianas.
- Por definición, si el vector completo está formado por VAs conjuntamente gaussianas, cualquier subconjunto de VAs extraído del mismo también es conjuntamente gaussiano, en particular, un subconjunto formado por M ($M < N$) VAs.

Al respecto de $\bar{\eta}_{\bar{\mathbf{Y}}}$ y $\bar{\bar{C}}_{\bar{\mathbf{Y}}}$, aplicando la linealidad del operador esperanza:

$$\begin{aligned} \bar{\eta}_{\bar{\mathbf{Y}}} &= E\{\bar{\mathbf{Y}}\} &= E\{\bar{\bar{A}}\bar{\mathbf{X}}\} = \bar{\bar{A}}\bar{\eta}_{\bar{\mathbf{X}}} \\ \bar{\bar{C}}_{\bar{\mathbf{Y}}} &= E\{(\bar{\mathbf{Y}} - \bar{\eta}_{\bar{\mathbf{Y}}})(\bar{\mathbf{Y}} - \bar{\eta}_{\bar{\mathbf{Y}}})^T\} \\ &= E\{(\bar{\bar{A}}\bar{\mathbf{X}} - \bar{\bar{A}}\bar{\eta}_{\bar{\mathbf{X}}})(\bar{\bar{A}}\bar{\mathbf{X}} - \bar{\bar{A}}\bar{\eta}_{\bar{\mathbf{X}}})^T\} \\ &= E\{\bar{\bar{A}}(\bar{\mathbf{X}} - \bar{\eta}_{\bar{\mathbf{X}}})(\bar{\mathbf{X}} - \bar{\eta}_{\bar{\mathbf{X}}})^T\bar{\bar{A}}^T\} \\ &= \bar{\bar{A}}E\{(\bar{\mathbf{X}} - \bar{\eta}_{\bar{\mathbf{X}}})(\bar{\mathbf{X}} - \bar{\eta}_{\bar{\mathbf{X}}})^T\}\bar{\bar{A}}^T = \bar{\bar{A}}\bar{\bar{C}}_{\bar{\mathbf{X}}}\bar{\bar{A}}^T \end{aligned} \tag{4.24}$$

expresiones que coinciden, respectivamente, con (4.20) y (4.22)



Ejercicio 3: Particularícese la expresión (4.19) para el caso $N = 2$ (gaussiana bivariante).

Solución: Para el caso que se pide empleemos la notación (\mathbf{X}, \mathbf{Y}) de forma similar a como hacíamos en el capítulo 3. La matriz de covarianza puede escribirse:

$$\overline{\overline{C}}_{\mathbf{XY}} = \begin{bmatrix} \sigma_{\mathbf{X}}^2 & C_{\mathbf{XY}} \\ C_{\mathbf{XY}} & \sigma_{\mathbf{Y}}^2 \end{bmatrix}$$

cuyo determinante es

$$|\overline{\overline{C}}_{\mathbf{XY}}| = \sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 - C_{\mathbf{XY}}^2 = \sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 (1 - \rho_{\mathbf{XY}}^2)$$

y su inversa

$$\overline{\overline{C}}_{\mathbf{XY}}^{-1} = \frac{1}{\sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 (1 - \rho_{\mathbf{XY}}^2)} \begin{bmatrix} \sigma_{\mathbf{Y}}^2 & -C_{\mathbf{XY}} \\ -C_{\mathbf{XY}} & \sigma_{\mathbf{X}}^2 \end{bmatrix}$$

La expresión (4.19) se particulariza para este caso de la forma

$$f_{\mathbf{XY}}(x, y) = \frac{1}{2\pi \sqrt{\sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 (1 - \rho_{\mathbf{XY}}^2)}} e^{-\frac{1}{2}\alpha}$$

con α la forma cuadrática presente en (4.19), la cual pasa a ser:

$$\begin{aligned} \alpha &= \begin{bmatrix} x - \eta_{\mathbf{X}} & y - \eta_{\mathbf{Y}} \end{bmatrix} \overline{\overline{C}}_{\mathbf{XY}}^{-1} \begin{bmatrix} x - \eta_{\mathbf{X}} \\ y - \eta_{\mathbf{Y}} \end{bmatrix} \\ &= \frac{1}{\sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 (1 - \rho_{\mathbf{XY}}^2)} \times \\ &\quad \begin{bmatrix} x - \eta_{\mathbf{X}} & y - \eta_{\mathbf{Y}} \end{bmatrix} \begin{bmatrix} \sigma_{\mathbf{Y}}^2 & -C_{\mathbf{XY}} \\ -C_{\mathbf{XY}} & \sigma_{\mathbf{X}}^2 \end{bmatrix} \begin{bmatrix} x - \eta_{\mathbf{X}} \\ y - \eta_{\mathbf{Y}} \end{bmatrix} \\ &= \frac{\sigma_{\mathbf{Y}}^2 (x - \eta_{\mathbf{X}})^2 - 2C_{\mathbf{XY}}(x - \eta_{\mathbf{X}})(y - \eta_{\mathbf{Y}}) + \sigma_{\mathbf{X}}^2 (y - \eta_{\mathbf{Y}})^2}{\sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2 (1 - \rho_{\mathbf{XY}}^2)} \\ &= \frac{1}{(1 - \rho_{\mathbf{XY}}^2)} \times \\ &\quad \left(\frac{(x - \eta_{\mathbf{X}})^2}{\sigma_{\mathbf{X}}^2} - 2\rho_{\mathbf{XY}} \frac{(x - \eta_{\mathbf{X}})(y - \eta_{\mathbf{Y}})}{\sigma_{\mathbf{X}} \sigma_{\mathbf{Y}}} + \frac{(y - \eta_{\mathbf{Y}})^2}{\sigma_{\mathbf{Y}}^2} \right) \end{aligned}$$

Por tanto, la expresión de la gaussiana bivariante es

$$\begin{aligned} f_{\mathbf{XY}}(x, y) &= \frac{1}{\sigma_X \sigma_Y 2\pi \sqrt{1 - \rho_{XY}^2}} \times \\ &\quad e^{-\frac{1}{2} \frac{1}{(1-\rho_{XY}^2)} \left(\frac{(x-\eta_X)^2}{\sigma_X^2} - 2\rho_{XY} \frac{(x-\eta_X)(y-\eta_Y)}{\sigma_X \sigma_Y} + \frac{(y-\eta_Y)^2}{\sigma_Y^2} \right)} \end{aligned} \quad (4.25)$$



Ejercicio 4: Compruebe que, para el caso de una VA gaussiana bivariante, el estimador lineal óptimo coindice con el estimador óptimo sin restricciones.

Solución: A partir de la expresión (4.25) denominemos, por simplicidad en la escritura $\tau_1 = x - \eta_X$ y $\tau_2 = y - \eta_Y$; podemos entonces escribir

$$\begin{aligned} f_{\mathbf{XY}}(\tau_1, \tau_2) &= \frac{1}{\sigma_X \sigma_Y 2\pi \sqrt{1 - \rho_{XY}^2}} \times \\ &\quad e^{-\frac{1}{2} \frac{1}{(1-\rho_{XY}^2)} \left(\frac{\tau_1^2}{\sigma_X^2} - 2\rho_{XY} \frac{\tau_1 \tau_2}{\sigma_X \sigma_Y} + \frac{\tau_2^2}{\sigma_Y^2} \right)} \\ &= \frac{1}{\sigma_X \sigma_Y 2\pi \sqrt{1 - \rho_{XY}^2}} e^{-\frac{1}{2} \frac{1}{(1-\rho_{XY}^2)} \Psi(\tau_1, \tau_2)} \end{aligned} \quad (4.26)$$

Entonces

$$\Psi(\tau_1, \tau_2) = \frac{\sigma_Y^2 \tau_1^2 - 2\rho_{XY} \sigma_X \sigma_Y \tau_1 \tau_2 + \sigma_X^2 \tau_2^2}{\sigma_X^2 \sigma_Y^2}$$

Sumando y restando $\rho_{XY}^2 \sigma_Y^2 \tau_1^2$ podemos conseguir un cuadrado perfecto de la forma

$$\begin{aligned} \Psi(\tau_1, \tau_2) &= \frac{\sigma_Y^2 \tau_1^2 - 2\rho_{XY} \sigma_X \sigma_Y \tau_1 \tau_2 + \sigma_X^2 \tau_2^2}{\sigma_X^2 \sigma_Y^2} + \\ &\quad \frac{\rho_{XY}^2 \sigma_Y^2 \tau_1^2 - \rho_{XY}^2 \sigma_Y^2 \tau_1^2}{\sigma_X^2 \sigma_Y^2} \\ &= \frac{(\sigma_X \tau_2 - \rho_{XY} \sigma_Y \tau_1)^2 + \sigma_Y^2 \tau_1^2 (1 - \rho_{XY}^2)}{\sigma_X^2 \sigma_Y^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma_{\mathbf{X}}^2 \left(\tau_2 - \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} \tau_1 \right)^2 + \sigma_{\mathbf{Y}}^2 \tau_1^2 (1 - \rho_{\mathbf{XY}}^2)}{\sigma_{\mathbf{X}}^2 \sigma_{\mathbf{Y}}^2} \\
&= \frac{\left(\tau_2 - \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} \tau_1 \right)^2}{\sigma_{\mathbf{Y}}^2} + \frac{\tau_1^2 (1 - \rho_{\mathbf{XY}}^2)}{\sigma_{\mathbf{X}}^2}
\end{aligned}$$

Por ello, retomando la expresión (4.26)

$$\begin{aligned}
f_{\mathbf{XY}}(\tau_1, \tau_2) &= \frac{1}{\sigma_{\mathbf{X}} \sqrt{2\pi}} e^{-\frac{1}{2(1-\rho_{\mathbf{XY}}^2)} \frac{\tau_1^2 (1 - \rho_{\mathbf{XY}}^2)}{\sigma_{\mathbf{X}}^2}} \cdot \\
&\quad \frac{1}{\sigma_{\mathbf{Y}} \sqrt{2\pi(1 - \rho_{\mathbf{XY}}^2)}} e^{-\frac{\left(\tau_2 - \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} \tau_1 \right)^2}{\sigma_{\mathbf{Y}}^2}}
\end{aligned}$$

con lo que si devolvemos la expresión a las variables originales x e y tendremos

$$\begin{aligned}
f_{\mathbf{XY}}(x, y) &= \frac{1}{\sigma_{\mathbf{X}} \sqrt{2\pi}} e^{-\frac{(x - \eta_{\mathbf{X}})^2}{\sigma_{\mathbf{X}}^2}} \cdot \\
&\quad \frac{1}{\sigma_{\mathbf{Y}} \sqrt{2\pi(1 - \rho_{\mathbf{XY}}^2)}} e^{-\frac{\left((y - \eta_{\mathbf{Y}}) - \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} (x - \eta_{\mathbf{X}}) \right)^2}{2(1 - \rho_{\mathbf{XY}}^2) \sigma_{\mathbf{Y}}^2}} \\
&= \frac{1}{\sigma_{\mathbf{X}} \sqrt{2\pi}} e^{-\frac{(x - \eta_{\mathbf{X}})^2}{\sigma_{\mathbf{X}}^2}} \cdot \\
&\quad \frac{1}{\sigma_{\mathbf{Y}} \sqrt{2\pi(1 - \rho_{\mathbf{XY}}^2)}} e^{-\frac{\left(y - \left(\eta_{\mathbf{Y}} + \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} (x - \eta_{\mathbf{X}}) \right) \right)^2}{2(1 - \rho_{\mathbf{XY}}^2) \sigma_{\mathbf{Y}}^2}} \\
&= f_{\mathbf{X}}(x) f_{\mathbf{Y}}(y|x)
\end{aligned}$$

donde la última igualdad procede de las identidades (3.24) y (3.25). Entonces podemos afirmar que

$$\begin{aligned}
\mathbf{X} &\sim N(\eta_{\mathbf{X}}, \sigma_{\mathbf{X}}) \\
\mathbf{Y}|x &\sim N\left(\eta_{\mathbf{Y}} + \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} (x - \eta_{\mathbf{X}}), \sigma_{\mathbf{Y}} \sqrt{1 - \rho_{\mathbf{XY}}^2}\right)
\end{aligned} \tag{4.27}$$

por lo que, según se desprende de (4.27)

$$E\{\mathbf{Y}|\mathbf{X}\} = \eta_{\mathbf{Y}} + \rho_{\mathbf{XY}} \frac{\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}} (\mathbf{X} - \eta_{\mathbf{X}})$$

que resulta ser una función lineal de la VA \mathbf{X} .



Ejercicio 5: Considere que la secuencia $x[n] = (x[0], x[1], x[2], x[3])$ es filtrada a través de un sistema lineal e invariante (LTI) con respuesta al impulso $h[n] = (h[0], h[1], h[2])$. Se pide que escriba la secuencia de salida en forma matricial.

Solución: Como es bien sabido, la secuencia de salida de un sistema LTI con respuesta al impulso $h[n]$ cuando la entrada es $x[n]$ se obtiene mediante la expresión de la convolución de señales discretas

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k]$$

Para el caso que se plantea en el ejercicio, el sumatorio tiene un recorrido finito. La secuencia $x[n]$ tiene un tamaño de $L = 4$ puntos y la respuesta al impulso de $P = 3$ puntos. Por ello, la secuencia de salida $y[n]$ tiene un tamaño de $L + P - 1 = 4 + 3 - 1 = 6$ puntos, cuyos valores son:

$$\begin{aligned} y[0] &= x[0]h[0] \\ y[1] &= x[0]h[1] + x[1]h[0] \\ y[2] &= x[0]h[2] + x[1]h[1] + x[2]h[0] \\ y[3] &= x[1]h[2] + x[2]h[1] + x[3]h[0] \\ y[4] &= x[2]h[2] + x[3]h[1] \\ y[5] &= x[3]h[2] \end{aligned}$$

lo cual puede escribirse matricialmente como

$$\begin{bmatrix} y[0] \\ y[1] \\ y[2] \\ y[3] \\ y[4] \\ y[5] \end{bmatrix} = \begin{pmatrix} h[0] & 0 & 0 & 0 \\ h[1] & h[0] & 0 & 0 \\ h[2] & h[1] & h[0] & 0 \\ 0 & h[2] & h[1] & h[0] \\ 0 & 0 & h[2] & h[1] \\ 0 & 0 & 0 & h[2] \end{pmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \end{bmatrix}$$

Por tanto dado que la expresión de la convolución no es otra que la de una combinación lineal, es obvio que si la entrada a

un sistema LTI es un vector de VAs conjuntamente gaussianas, la salida lo será también⁴.



4.4 Variables complejas

En el caso en que las variables involucradas sean complejas, los resultados que hemos visto hasta ahora son aplicables con pequeñas variaciones. Concretamente, una variable compleja $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ tiene dos componentes reales, de forma que la función de densidad de la variable compleja coincide con la función de densidad de la variable bidimensional (\mathbf{X}, \mathbf{Y}) . Para el caso de dos variables complejas tendríamos cuatro variables reales, de forma que la función conjunta sería una función de densidad tetradimensional. Estos resultados se extenderían al caso de N variables.

Con respecto a las esperanzas, $E\{\mathbf{Z}\} = E\{\mathbf{X} + j\mathbf{Y}\} = E\{\mathbf{X}\} + jE\{\mathbf{Y}\}$. Como puede verse, una esperanza es una medida de tendencia central, de forma que la media localiza a la variable compleja en torno a un punto del plano complejo.

Por otra parte, la varianza es una medida de dispersión, y mediría en este caso cómo se dispersa la variable compleja con respecto a su valor medio. Así pues, se plantea como un radio de dispersión de la forma

$$\begin{aligned}\sigma_{\mathbf{Z}}^2 &= E\{|\mathbf{Z} - \eta_{\mathbf{Z}}|^2\} = E\{(\mathbf{Z} - \eta_{\mathbf{Z}})(\mathbf{Z} - \eta_{\mathbf{Z}})^*\} \\ &= E\{\mathbf{Z}\mathbf{Z}^*\} - \eta_{\mathbf{Z}}\eta_{\mathbf{Z}}^* \\ &= E\{|\mathbf{Z}|^2\} - |\eta_{\mathbf{Z}}|^2\end{aligned}\tag{4.28}$$

donde $E\{|\mathbf{Z}|^2\}$ es el VCM de la variable \mathbf{Z} .

Si $g(\mathbf{Z})$ es una función de la variable compleja \mathbf{Z} el cálculo de $E\{g(\mathbf{Z})\}$ se hace definiendo esta función en base a las componentes real e imaginaria de \mathbf{Z} . Así pues, considerando que $g(\mathbf{Z})$ se puede expresar en base a estas componentes como $\varphi(\mathbf{X}, \mathbf{Y})$ entonces

$$E\{g(\mathbf{Z})\} = E\{\varphi(\mathbf{X}, \mathbf{Y})\} = \int \int \varphi(x, y) f_{\mathbf{XY}}(x, y) dx dy\tag{4.29}$$

⁴Estrictamente hablando, sólo las L primeras VAs serían conjuntamente gaussianas pues las $P - 1$ restantes son calculables a partir de las primeras. No obstante, en casos prácticos el valor de L suele ser mucho mayor de P , por lo que los efectos de bordes de los transitorios del filtro son despreciables.

Si $\mathbf{Z}_1 = \mathbf{X}_1 + j\mathbf{Y}_1$ y $\mathbf{Z}_2 = \mathbf{X}_2 + j\mathbf{Y}_2$ son dos variables complejas independientes, entonces podremos escribir

$$f_{\mathbf{X}_1 \mathbf{Y}_1 \mathbf{X}_2 \mathbf{Y}_2}(x_1, y_1, x_2, y_2) = f_{\mathbf{X}_1 \mathbf{Y}_1}(x_1, y_1) f_{\mathbf{X}_2 \mathbf{Y}_2}(x_2, y_2) \quad (4.30)$$

Finalmente, la correlación y la covarianza se generalizan al caso complejo, simplemente, conjugando la segunda variable. Concretamente

$$R_{\mathbf{Z}_1 \mathbf{Z}_2} = E\{\mathbf{Z}_1 \mathbf{Z}_2^*\} \quad (4.31)$$

$$\begin{aligned} C_{\mathbf{Z}_1 \mathbf{Z}_2} &= E\{(\mathbf{Z}_1 - \eta_{\mathbf{Z}_1})(\mathbf{Z}_2 - \eta_{\mathbf{Z}_2})^*\} \\ &= R_{\mathbf{Z}_1 \mathbf{Z}_2} - \eta_{\mathbf{Z}_1} \eta_{\mathbf{Z}_2}^* \end{aligned} \quad (4.32)$$

Nótese pues que

$$R_{\mathbf{Z}_2 \mathbf{Z}_1} = E\{\mathbf{Z}_2 \mathbf{Z}_1^*\} = E^*\{(\mathbf{Z}_1 \mathbf{Z}_2^*)\} = R_{\mathbf{Z}_1 \mathbf{Z}_2}^* \quad (4.33)$$

$$\begin{aligned} C_{\mathbf{Z}_2 \mathbf{Z}_1} &= E\{(\mathbf{Z}_2 - \eta_{\mathbf{Z}_2})(\mathbf{Z}_1 - \eta_{\mathbf{Z}_1})^*\} \\ &= E^*\{(\mathbf{Z}_1 - \eta_{\mathbf{Z}_1})(\mathbf{Z}_2 - \eta_{\mathbf{Z}_2})^*\} = C_{\mathbf{Z}_1 \mathbf{Z}_2}^* \end{aligned} \quad (4.34)$$

de modo que las expresiones dadas para las matrices de correlación y covarianza (ecuaciones 4.15 y 4.16) eran específicas del caso real. Para VAs complejas las matrices tienen que incorporar los complejos conjugados, de modo que las matrices no serán simétricas sino hermíticas (la matriz coincide con su conjugada traspuesta).

4.5 Teoremas asintóticos

4.5.1 Teorema del Límite Central

En términos muy amplios, el *Teorema del Límite Central* establece que la distribución de la suma de un número elevado de variables tiende a comportarse como la distribución de una variable gaussiana, tanto más cuantas más variables intervengan en la suma. Este teorema aplica cuando las variables involucradas son independientes, aunque también puede ser cierto en algunos casos de dependencia.

Las versiones más exigentes del teorema requieren que las variables que intervienen en la suma sean *independientes e idénticamente distribuidas* (en adelante IID). No obstante esta segunda condición se puede relajar siempre que ninguna variable domine sobre las demás. ¿Cómo puede cuantificarse dicho dominio? Una posible forma es la siguiente [1]: denominando a las variables \mathbf{X}_i , $i = \{1, \dots, N\}$, entonces

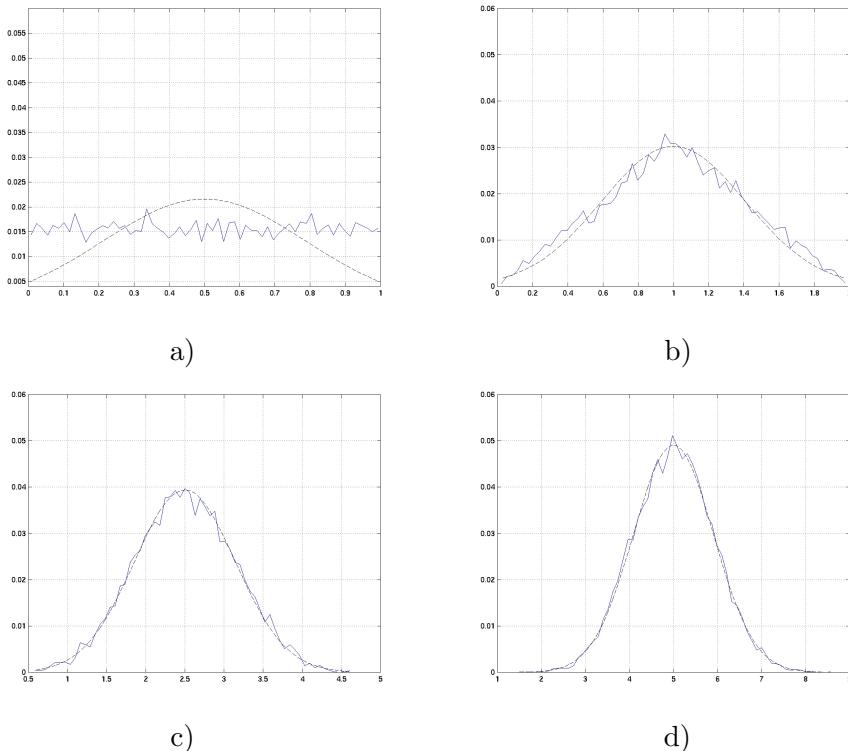


Figura 4.1: Ejemplo de convergencia hacia una distribución gaussiana a partir de VAs $U(0, 1)$. a) Histograma de una única VA. b) Suma de dos VAs. c) Suma de cinco VAs d) Suma de 10 VAs. En todos los casos se ha superpuesto en línea discontinua la función de densidad gaussiana con la media y varianza de la VA resultante de la suma de las VAs uniformes.

1. $\sigma_{\mathbf{X}_i}^2 > B_1 > 0 \quad \forall i = 1, \dots, N$
2. $E\{| \mathbf{X}_i - \eta_{\mathbf{X}_i} |^3\} > B_2 \quad \forall i = 1, \dots, N$

con B_1 y B_2 dos constantes.

Una expresión, por tanto, del teorema es: dadas \mathbf{X}_i , $i = \{1, \dots, N\}$, variables independientes y que satisfacen las condiciones de suficiencia indicadas, entonces la distribución de la variable

$$\mathbf{Z} = \sum_{i=1}^N \mathbf{X}_i$$

tiene el comportamiento

$$F_{\mathbf{Z}}(z) \longrightarrow_{N \rightarrow \infty} G\left(\frac{z - \eta_{\mathbf{Z}}}{\sigma_{\mathbf{Z}}}\right) \quad (4.35)$$

con

$$\begin{aligned} \eta_{\mathbf{Z}} &= \sum_{i=1}^N \eta_{\mathbf{X}_i} \\ \sigma_{\mathbf{Z}}^2 &= \sum_{i=1}^N \sigma_{\mathbf{X}_i}^2 \end{aligned}$$

Nótese que se ha hablado de *convergencia en distribución*, y no se ha dicho nada del comportamiento de la función de densidad de probabilidad. En general, en el caso de las variables continuas una convergencia suele traer consigo también la otra, de forma que si se cumple (4.35) se verifica también que

$$f_{\mathbf{Z}}(z) \longrightarrow_{N \rightarrow \infty} \frac{1}{\sigma_{\mathbf{Z}} \sqrt{2\pi}} e^{-\frac{(z - \eta_{\mathbf{Z}})^2}{2\sigma_{\mathbf{Z}}^2}} \quad (4.36)$$

No obstante, si las variables \mathbf{X}_i fuesen discretas, la función de densidad de la suma sería la correspondiente a una VA también discreta (es decir, formada por impulsos) de forma que una función de densidad gaussiana no tendría por qué poder ajustarse a este esquema. Un caso particular de esto, como veremos a continuación, es el caso en que los valores que tomen las VAs \mathbf{X}_i sean equiespaciados, es decir, $\mathbf{X}_i = a \cdot i$.

Intuitivamente puede ilustrarse el funcionamiento del teorema en base al teorema de la convolución (véase sección 3.5.1). Como es sabido, la función de densidad de la suma de variables independientes es igual a la convolución de sus funciones de densidad. Considerando variables IID uniformes entre cero y uno, la figura 4.1a) muestra el histograma normalizado de la VA $\mathbf{Z}_1 = \mathbf{X}_1$; la figura 4.1b) lo hace para la VA $\mathbf{Z}_2 = \mathbf{X}_1 + \mathbf{X}_2$ y las figuras 4.1c) y 4.1d) muestran los histogramas respectivos de las VAs $\mathbf{Z}_5 = \sum_{i=1}^5 \mathbf{X}_i$ y $\mathbf{Z}_{10} = \sum_{i=1}^{10} \mathbf{X}_i$. Como puede verse, la función de densidad resultante va adquiriendo progresivamente una envolvente más parecida a la curva gaussiana, tanto más cuantas más funciones de densidad intervienen en la convolución.

4.5.2 Teorema de DeMoivre-Laplace

Es el caso particular del anterior teorema para VAs discretas con valores equiespaciados, en concreto, VAs \mathbf{X}_i de Bernoulli (de parámetro p). En

este caso, asumiendo independencia entre las VAs, la distribución de la VA

$$\mathbf{Z} = \sum_{i=1}^N \mathbf{X}_i$$

que, como sabemos, en términos exactos es una binomial, puede aproximarse (con $N \gg 1$) mediante

$$F_{\mathbf{Z}}(z) \approx G\left(\frac{z - \eta_{\mathbf{Z}}}{\sigma_{\mathbf{Z}}}\right)$$

con

$$\begin{aligned}\eta_{\mathbf{Z}} &= Np \\ \sigma_{\mathbf{Z}} &= Npq\end{aligned}$$

Asimismo, por ser una VA de valores equiespaciados $z_i = i \cdot a = i$, $i = (0, \dots, N)$ la envolvente de la función de densidad de la misma también es gaussiana (véase ecuación (1.51) y figura 1.5 del primer capítulo). Esta aproximación es tanto más válida cuanto más cerca estén los valores de la VA de su valor medio. Por este motivo, se plantearon en el capítulo primero condiciones adicionales de validez de la expresión (1.51).

4.5.3 Ley de los Grandes Números

La Ley de los Grandes Números es la conexión entre la probabilidad definida en términos de frecuencia relativa y la probabilidad axiomática. Constituye pues la base sólida de argumentación de que la frecuencia relativa tiende a la probabilidad axiomática conforme el número de experimentos sobre los que se calcula esta frecuencia es cada vez mayor.

Como se verá, la argumentación a emplear hace uso de conceptos vistos hasta ahora. Consideremos un determinado experimento aleatorio $\epsilon < S, \mathcal{F}, P >$, y aceptemos que dado un $A \in \mathcal{F}$ se verifica que $P(A) = p$. Acéptese que realizamos este experimento N veces, naturalmente, ejecuciones independientes entre sí, y que llevamos cuenta de cuántas veces (de las N posibles) se ha verificado el suceso A . Denominemos a esta cantidad \mathbf{N}_A . En estas condiciones, es fácil demostrar⁵ que la frecuencia relativa

⁵Podría argumentarse que lo que haremos a continuación no es una demostración pues nos basaremos en una aproximación para conseguir la expresión (4.38); no obstante, dado que la conclusión se obtiene en el límite (en el cual se verifica 4.35), la validez del resultado (4.39) está fuera de duda.

$\mathbf{f}_r(A)$ (véase sección 1.3.1) de aparición de este suceso tiende a la probabilidad $P(A) = p$, o, formalmente

$$\lim_{N \rightarrow \infty} P(|\mathbf{f}_r(A) - p| < \delta) = 1 \quad (4.37)$$

$$\forall \delta > 0, \delta \in R.$$

Para ver que esto es así basta caracterizar la VA \mathbf{N}_A . Esta VA es igual, como hemos dicho antes, al número de veces que se verifica el suceso deseado, de N veces posibles, cuando se ejecutan experimentos independientes. Esta VA, por definición, es una VA Binomial, concretamente, $B(N, p)$. Así pues, aplicando la aproximación de DeMoivre-Laplace, podemos escribir:

$$\begin{aligned} P(|\mathbf{f}_r(A) - p| < \delta) &= P(-\delta < \mathbf{f}_r(A) - p < \delta) \\ &= P(-\delta < \frac{\mathbf{N}_A}{N} - p < \delta) \\ &= P(N(p - \delta) < \mathbf{N}_A < N(p + \delta)) \\ &\approx G\left(\frac{N(p + \delta) - Np}{\sqrt{Npq}}\right) - G\left(\frac{N(p - \delta) - Np}{\sqrt{Npq}}\right) \\ &= G\left(\frac{N\delta}{\sqrt{Npq}}\right) - G\left(-\frac{N\delta}{\sqrt{Npq}}\right) \\ &= G\left(\frac{N\delta}{\sqrt{Npq}}\right) - \left(1 - G\left(\frac{N\delta}{\sqrt{Npq}}\right)\right) \\ &= 2G\left(\frac{N\delta}{\sqrt{Npq}}\right) - 1 \end{aligned} \quad (4.38)$$

Por lo tanto

$$\begin{aligned} \lim_{N \rightarrow \infty} P(|\mathbf{f}_r(A) - p| < \delta) &= \lim_{N \rightarrow \infty} 2G\left(\frac{N\delta}{\sqrt{Npq}}\right) - 1 \\ &= 2G(\infty) - 1 = 2 - 1 = 1 \end{aligned} \quad (4.39)$$

Naturalmente, en un caso real el número de ejecuciones del experimento no será infinito, sino limitado. En tal caso la expresión (4.38) sirve, para cada valor de p , para encontrar un intervalo de confianza donde se encontrará la frecuencia relativa con una determinada probabilidad.

Ejercicio: Supóngase que se sospecha que para un cierto problema $p = 0.6$. Se pide que halle el número de veces N que debe realizarse un experimento para que la frecuencia relativa $\mathbf{f}_r(A)$ de aparición del suceso esté en un entorno de p de radio $\delta = 0.05$ con probabilidad 0.98.

Solución: Aplicando la expresión (4.38) tenemos

$$\begin{aligned} P(|\mathbf{f}_r(A) - p| < \delta) &= 2G\left(\frac{N\delta}{\sqrt{Npq}}\right) - 1 = 0.98 \\ &= G\left(\frac{N\delta}{\sqrt{Npq}}\right) = 0.99 \end{aligned}$$

Por tanto

$$\frac{N\delta}{\sqrt{Npq}} = G^{-1}(0.99) \approx 2.33$$

por lo que

$$\begin{aligned} \sqrt{N} &= 2.33 \frac{\sqrt{pq}}{\delta} \\ N &= \left(2.33 \frac{\sqrt{pq}}{\delta}\right)^2 = (2.33)^2 \frac{0.6 \cdot 0.4}{0.05^2} = 521.17 \end{aligned}$$

de forma que necesitamos hacer al menos $N = 522$ experimentos.



Ejercicio: Demuestre la Ley de los Grandes Números mediante la desigualdad de Tchebycheff.

Solución: En este caso la desigualdad de Tchebycheff (expresión 2.67) se escribiría

$$P(|\mathbf{X} - \eta_{\mathbf{X}}| < \epsilon) \geq 1 - \frac{\sigma_{\mathbf{X}}^2}{\epsilon^2} \quad (4.40)$$

La VA en juego es \mathbf{N}_A . Esta VA se define $\mathbf{N}_A = \sum_{i=1}^N \mathbf{X}_i$ con \mathbf{X}_i VA de Bernoulli de parámetro p , la cual toma el valor 1 cuando el suceso A se ha verificado. Las VAs \mathbf{X}_i y \mathbf{X}_j son independientes (luego, en particular, incorreladas) $\forall i \neq j$. Así pues, dado que $\mathbf{f}_r(A) = \frac{\mathbf{N}_A}{N}$, se verifica que

$$\begin{aligned} E\{\mathbf{f}_r(A)\} &= E\left\{\frac{\mathbf{N}_A}{N}\right\} = \frac{1}{N} \sum_{i=1}^N E\{\mathbf{X}_i\} = p \\ \sigma_{\mathbf{f}_r(A)}^2 &= \frac{1}{N^2} \sigma_{\mathbf{N}_A}^2 = \frac{1}{N^2} \sum_{i=1}^N \sigma_{\mathbf{X}_i}^2 = \frac{Npq}{N^2} = \frac{pq}{N} \end{aligned}$$

Por ello, sustituyendo en (4.40), tenemos

$$\begin{aligned} P(|\mathbf{f}_r(A) - p| < \epsilon) &\geq 1 - \frac{\frac{pq}{N}}{\epsilon^2} \\ \lim_{N \rightarrow \infty} P(|\mathbf{f}_r(A) - p| < \epsilon) &\geq \lim_{N \rightarrow \infty} 1 - \frac{pq}{N\epsilon^2} \\ &= 1 - \lim_{N \rightarrow \infty} \frac{pq}{N\epsilon^2} = 1 \quad (4.41) \end{aligned}$$

$\forall \epsilon > 0$, como se quería demostrar.



4.6 Algunos ejercicios ilustrativos

1.- Un sistema de vigilancia aérea está constituido por N subsistemas y un control central; cada subsistema toma, en un determinado instante, una medida, tal que en **condiciones de ausencia de aeronave** puede modelarse como una variable aleatoria \mathbf{X}_i $i = \{1, \dots, N\}$ uniforme entre 0 y 1. Cada variable \mathbf{X}_i está sometida a un proceso de acondicionamiento mediante la función $g(\mathbf{X}_i) = \frac{1}{\alpha} [-\ln(\mathbf{X}_i)]^{\frac{1}{\beta}}$, (con $\alpha, \beta > 0$), originando nuevas variables \mathbf{Y}_i , más apropiadas para basar en ellas las decisiones. El funcionamiento de cada subsistema puede considerarse independiente del resto de ellos. El control central combinará la información que recibe de los subsistemas para gestionar la presencia o ausencia de aeronave. La forma en que el sistema central combina la información dependerá de la estrategia de detección a seguir. En estas condiciones:

- a) Considerando que la función de transformación $g(\mathbf{X}_i)$ $i = \{1, \dots, N\}$ es biunívoca, obtenga la función de densidad de probabilidad y la media de cada variable \mathbf{Y}_i .
- b) Cada subsistema envía al control central una señal de alarma (indicando que el sistema piensa que hay presencia de aeronave) si la amplitud de la \mathbf{Y}_i recibida supera un cierto umbral μ . Calcule el valor de éste para que la probabilidad de falsa alarma (en adelante Pfa) en cada subsistema sea igual a λ . Particularice para $\lambda = 0.6$, $\alpha = 4$ y $\beta = 3$.
- c) Para aumentar la fiabilidad el control central sigue una estrategia de decidir presencia de blanco si al menos M subsistemas (de los N posibles) han enviado señal de alarma. Calcule la expresión de la Pfa del sistema, así como su valor numérico si $M = 13$, $N = 20$ ($N \gg 1$) y el umbral μ es el del apartado anterior.
- d) Con el fin de reducir el número de operaciones a realizar por el

sistema de control para tomar una decisión se propone otra estrategia; ésta consiste en los subsistemas en orden creciente de índice (de 1 a N), y dar alarma en cuanto alguno de ellos haya enviado señal de alarma. Obtenga el umbral μ para cada subsistema de modo que la Pfa de esta estrategia coincida con la del apartado c), así como el número medio de operaciones realizadas por el sistema de control para tomar una decisión, suponiendo que la consulta a cada subsistema requiere P operaciones.

Solución:

a) La función de densidad de la VA cada \mathbf{Y}_i puede obtenerse mediante el Teorema Fundamental para variables unidimensionales (véase sección 2.6.3) cuya expresión es

$$f_{\mathbf{Y}_i}(y) = \sum_{i=1}^{N(y)} \frac{f_{\mathbf{X}_i}(x)}{|g'(x)|} \Big|_{x=x_i}.$$

En nuestro caso $N(y) = 1 \forall y$ que resulte ser imagen de la transformación, puesto que ésta es biunívoca. Por otra parte podemos calcular la derivada de la transformación de la forma

$$\begin{aligned} g(x) &= \frac{1}{\alpha} [-\ln(x)]^{\frac{1}{\beta}} \\ g'(x) &= \frac{1}{\alpha\beta} [-\ln(x)]^{\frac{1}{\beta}-1} \left(-\frac{1}{x}\right) \\ |g'(x)| &= \frac{1}{\alpha\beta} [-\ln(x)]^{\frac{1-\beta}{\beta}} \left(\frac{1}{x}\right) \end{aligned}$$

Asimismo, a partir de la transformación es claro que

$$\begin{aligned} -(\alpha y)^\beta &= \ln(x) \\ x &= e^{-(\alpha y)^\beta} \end{aligned}$$

por lo que la expresión del valor absoluto de la derivada podemos escribirla, de forma alternativa

$$|g'(x)| = \frac{1}{\alpha\beta} [(\alpha y)^\beta]^{\frac{1-\beta}{\beta}} e^{(\alpha y)^\beta}$$

Ello trae consigo que la función de densidad resulte ser

$$\begin{aligned} f_{\mathbf{Y}_i}(y) &= \frac{1}{|g'(x)|} = \frac{1}{\frac{1}{\alpha\beta} [(\alpha y)^{1-\beta}] e^{(\alpha y)^\beta}} = \alpha\beta(\alpha y)^{\beta-1} e^{-(\alpha y)^\beta} \\ &= \alpha^\beta \beta y^{\beta-1} e^{-(\alpha y)^\beta} \end{aligned}$$

siendo esta expresión válida $y > 0$ y cero en el resto de la recta real.

Al respecto del cálculo de $E\{\mathbf{Y}_i\}$ podemos escribir, haciendo uso de la igualdad (2.87)

$$E\{\mathbf{Y}_i\} = E\{g(\mathbf{X}_i)\} = \int_0^1 \frac{1}{\alpha} (-\ln(x))^{\frac{1}{\beta}} dx$$

Efectuando el cambio de variable $\ln(x) = \tau$ y recordando la definición de la función $\Gamma(\cdot)$ (expresión (2.63)), podemos escribir

$$\begin{aligned} E\{\mathbf{Y}_i\} = E\{g(\mathbf{X}_i)\} &= \int_{-\infty}^0 \frac{1}{\alpha} (-\tau)^{\frac{1}{\beta}} e^\tau d\tau \\ &= \int_0^\infty \frac{1}{\alpha} \tau^{\frac{1}{\beta}} e^{-\tau} d\tau = \frac{1}{\alpha} \int_0^\infty \tau^{[(\frac{1}{\beta}+1)-1]} e^{-\tau} d\tau \\ &= \frac{1}{\alpha} \Gamma\left(1 + \frac{1}{\beta}\right) = \frac{1}{\alpha \beta} \Gamma\left(\frac{1}{\beta}\right) \end{aligned}$$

b) El valor μ debe ser tal que

$$\begin{aligned} \lambda = Pfa = P(\mathbf{Y}_i > \mu) &= \int_\mu^\infty f_{\mathbf{Y}_i}(y) dy \\ &= \int_\mu^\infty \alpha^\beta \beta y^{\beta-1} e^{-(\alpha y)^\beta} dy \\ &= \left[-e^{-(\alpha y)^\beta}\right]_\mu^\infty = e^{-(\alpha \mu)^\beta} \end{aligned}$$

Por ello

$$\begin{aligned} (\alpha \mu)^\beta &= -\ln(\lambda) \\ \alpha \mu &= [-\ln(\lambda)]^{\frac{1}{\beta}} \\ \mu &= \frac{1}{\alpha} [\ln(\lambda)]^{\frac{1}{\beta}} \end{aligned} \tag{4.42}$$

Para los valores planteados en el enunciado $\mu = 0.1998$.

c) Se plantea una composición de ensayos de Bernoulli (decisión en cada subsistema de presencia/ausencia de aeronave) mediante subsistemas independientes. Por ello, podemos emplear las conclusiones obtenidas en la sección 1.7.1. En particular, definiendo el suceso B_k como “ha habido k alarmas de N posibles”, la probabilidad pedida se puede escribir

$$Pfa = P\left(\bigcup_{k=M}^N B_k\right) = 1 - P\left(\bigcup_{k=0}^{M-1} B_k\right) = 1 - \sum_{k=0}^{M-1} P(B_k)$$

lo cual, en virtud de la expresión (1.49), podemos escribir

$$Pfa = 1 - \sum_{k=0}^{M-1} \binom{N}{k} p^k q^{N-k}$$

con p la probabilidad de que cada subsistema dé alarma; como el umbral que se emplea en cada subsistema es μ entonces $p = \lambda$. Por ello

$$Pfa = 1 - \sum_{k=0}^{M-1} \binom{N}{k} \lambda^k (1-\lambda)^{N-k} \quad (4.43)$$

Al respecto del valor numérico y con el objetivo de facilitar el cálculo, podría plantearse el uso de la aproximación de DeMoivre-Laplace (expresión (1.52)); para ello debemos comprobar que se cumplen las condiciones de aproximación de forma satisfactoria (véase página 46). Por una parte repárese que $\lambda(1-\lambda)N = 0.6 \cdot 0.4 \cdot 20 = 4.8$ valor que, si bien es mayor que uno, no es excesivamente mayor. El valor $M - 1 = 13 - 1 = 12$, por otra parte, sí se encuentra suficientemente próximo al valor $N\lambda = 20 \cdot 0.6 = 12$; de hecho $N = M - 1$. Por ello, si bien no se cumplen enteramente las condiciones de la aproximación, ésta nos podría proporcionar una primera idea del valor de la Pfa . Aplicando dicha aproximación, con las reservas indicadas, obtenemos

$$Pfa \cong 1 - G\left(\frac{M-1-N\lambda}{\sqrt{N\lambda(1-\lambda)}}\right) = 1 - G(0) = 1 - 0.5 = 0.5$$

Si se calcula de forma numérica el resultado de la expresión (4.43) para los valores propuestos resulta una probabilidad $Pfa = 0.41$, por lo que, en efecto, la aproximación de DeMoivre-Laplace se aleja del valor real al no cumplirse, por completo, las condiciones para su empleo.

d) Definamos el suceso A_i ($i = \{1, \dots, N\}$) como “el subsistema i -ésimo envía alarma”, y aceptemos que $P(A_i) = p$. En este caso el sistema decide falsa alarma, digamos, suceso FA si se verifica

$$FA = A_1 \cup (\overline{A}_1 \cap A_2) \cup (\overline{A}_1 \cap \overline{A}_2 \cap A_3) \cup \dots \cup \left(\left[\bigcap_{i=1}^{N-1} \overline{A}_i \right] \cap A_N \right)$$

Sin embargo, a efectos del cálculo de Pfa es más sencillo emplear el suceso complementario

$$\overline{FA} = \bigcap_{i=1}^N \overline{A}_i$$

por lo que

$$\begin{aligned} Pfa = P(FA) &= 1 - P(\overline{FA}) = 1 - P\left(\bigcap_{i=1}^N \overline{A}_i\right) \\ &= 1 - \prod_{i=1}^N P(\overline{A}_i) = 1 - (1-p)^N \end{aligned}$$

Para $Pfa = 0.5$ el valor de p resultante es $p = 0.0341$. El umbral μ se calcularía ahora haciendo uso de la expresión (4.42), escribiendo $p = P(A_i)$ en lugar de λ . Sustituyendo los valores de α y β indicados en el apartado b), obtenemos $\mu = 0.3751$.

Al respecto del número medio de operaciones, definimos la VA **Z** como el “número de operaciones que realiza el control central para tomar una decisión”. A partir de esta definición podemos escribir

$$\begin{aligned} \mathbf{Z} = P &\rightarrow P(\mathbf{Z} = P) = P(A_1) = p \\ \mathbf{Z} = 2P &\rightarrow P(\mathbf{Z} = 2P) = P(\overline{A}_1 \cap A_2) = qp \\ \mathbf{Z} = 3P &\rightarrow P(\mathbf{Z} = 3P) = P(\overline{A}_1 \cap \overline{A}_2 \cap A_3) = q^2p \\ \vdots & \\ \mathbf{Z} = (N-1)P &\rightarrow P(\mathbf{Z} = (N-1)P) = P\left(\left[\bigcap_{i=1}^{N-2} \overline{A}_i\right] \cap A_{N-1}\right) = q^{N-2}p \\ \mathbf{Z} = NP &\rightarrow P(\mathbf{Z} = NP) = P\left(\left[\bigcap_{i=1}^{N-1} \overline{A}_i\right] \cap [A_N \cup \overline{A}_N]\right) = \\ &= q^{N-1}(p+q) = q^{N-1} \end{aligned}$$

A partir de ello el cálculo del valor medio $E\{\mathbf{Z}\}$ se llevaría a cabo en base a la expresión (2.41) y operando de forma acorde. Concretamente

$$E\{\mathbf{Z}\} = \sum_{i=1}^N z_i P(\mathbf{Z} = z_i) = \sum_{i=1}^{N-1} (iP)q^{i-1}p + (NP)q^{N-1}$$

Esta expresión puede reformularse para extender el sumatorio al último valor de la VA empleando el hecho de que $q^{N-1} = q^{N-1}(p+q)$, de modo que

$$E\{\mathbf{Z}\} = pP \sum_{i=1}^N iq^{i-1} + NPq^N$$

Para encontrar el valor de la serie anterior podemos partir de la igualdad

$$S(q) = \sum_{i=1}^N q^i = \frac{q(1-q^N)}{1-q} = \frac{q - q^{N+1}}{1-q}$$

y derivando con respecto al parámetro q tenemos

$$\frac{dS(q)}{dq} = \sum_{i=1}^N iq^{i-1} = \frac{[1 - (N+1)q^N][1-q] + (q - q^{N+1})}{(1-q)^2}$$

Por tanto

$$\begin{aligned} E\{\mathbf{Z}\} &= pP \frac{[1 - (N+1)q^N][1-q] + (q - q^{N+1})}{(1-q)^2} + q^N NP \\ &= \frac{P \left[(1 - (N+1)q^N)(1-q) + (q - q^{N+1}) \right] + pq^N NP}{p} \\ &= \frac{P(1 - q^N)p - PNq^N p + Pq(1 - q^N) + PNq^N p}{p} \\ &= \frac{P(1 - q^N)(p + q)}{p} = P \frac{1 - q^N}{p} \end{aligned}$$



2.- Un detector de un sistema de comunicaciones digitales binario tiene dos canales (superior e inferior). La salida del canal superior será la variable aleatoria \mathbf{X}_1 y la del canal inferior la variable \mathbf{X}_2 . Cuando el emisor transmite el símbolo 1, las funciones de densidad de las variables mencionadas se sabe que son

$$f_{\mathbf{X}_i}(x) = \frac{\beta^{-\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i-1} e^{-x/\beta} \quad x \geq 0$$

con $i = \{1, 2\}$, $\alpha_i > 0$, $\beta > 0$, y en las condiciones indicadas, $\alpha_1 > \alpha_2$. Se sabe además que las variables son independientes. El detector construye la variable $\mathbf{Z} = \frac{\mathbf{X}_1}{\mathbf{X}_2}$, y decide que el símbolo enviado es un 1 si $Z \geq 1$, y caso contrario considera que el símbolo enviado es un 0. Suponiendo que el emisor transmite el símbolo 1, se pide:

- a) $f_{\mathbf{Z}}(z)$
- b) $\eta_{\mathbf{Z}}$

c) Sabiendo que la función de distribución de las variables arriba indicadas puede escribirse (si $\alpha_i \in Z$):

$$F_{\mathbf{X}_i}(x) = 1 - e^{-x/\beta} \sum_{j=0}^{\alpha_i-1} \frac{(x/\beta)^j}{j!} \quad x \geq 0,$$

con $i = \{1, 2\}$, calcule, para $\alpha_1 = 3$ y $\alpha_2 = 2$, la probabilidad de detectar el símbolo 1 correctamente.

Solución:

a) Las VAs \mathbf{X}_1 y \mathbf{X}_2 toman valores no negativos, luego su cociente también será no negativo. Podemos emplear el método de la variable auxiliar (sección 3.5.4) de la forma

$$\begin{aligned} \mathbf{Z} &= \frac{\mathbf{X}_1}{\mathbf{X}_2} \\ \mathbf{W} &= \mathbf{X}_2 \end{aligned}$$

Entonces

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{f_{\mathbf{X}_1 \mathbf{X}_2}(x_1, x_2)}{\left| \begin{array}{cc} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \\ 0 & 1 \end{array} \right|} = x_2 f_{\mathbf{X}_1 \mathbf{X}_2}(x_1, x_2) \\ &= x_2 f_{\mathbf{X}_1}(x_1) f_{\mathbf{X}_2}(x_2) = w f_{\mathbf{X}_1}(zw) f_{\mathbf{X}_2}(w), w \geq 0, z \geq 0 \end{aligned}$$

Marginalizando, obtenemos

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \int_0^\infty w f_{\mathbf{X}_1}(zw) f_{\mathbf{X}_2}(w) dw \\ &= \int_0^\infty w \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} (zw)^{\alpha_1-1} e^{-zw/\beta} \frac{\beta^{-\alpha_2}}{\Gamma(\alpha_2)} (w)^{\alpha_2-1} e^{-w/\beta} dw \\ &= \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} \frac{\beta^{-\alpha_2}}{\Gamma(\alpha_2)} z^{\alpha_1-1} \int_0^\infty w^{\alpha_1+\alpha_2-1} e^{-\frac{z+1}{\beta} w} dw \end{aligned}$$

Denominando $k = \frac{z+1}{\beta}$ y $p = \alpha_1 + \alpha_2$, y recordando la definición de la función $\Gamma(\cdot)$ (expresión (2.63)), podemos escribir

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \frac{\beta^{-(\alpha_1+\alpha_2)}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1-1} \left(\frac{1+z}{\beta} \right)^{-(\alpha_1+\alpha_2)} \Gamma(\alpha_1 + \alpha_2) \\ &= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{z^{\alpha_1-1}}{(1+z)^{\alpha_1+\alpha_2}} \end{aligned}$$

expresión válida para $z \geq 0$.

b)

$$\begin{aligned}
 \eta_{\mathbf{Z}} = E\{\mathbf{Z}\} &= \int_0^\infty z f_{\mathbf{Z}}(z) dz \\
 &= E\left\{\frac{\mathbf{X}_1}{\mathbf{X}_2}\right\} = \int_0^\infty \int_0^\infty x_1 x_2^{-1} f_{\mathbf{X}_1}(x_1) f_{\mathbf{X}_2}(x_2) dx_1 dx_2 \\
 &= \int_0^\infty \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1} e^{-x_1/\beta} dx_1 \int_0^\infty \frac{\beta^{-\alpha_2}}{\Gamma(\alpha_2)} x_2^{\alpha_2-2} e^{-x_2/\beta} dx_2 \\
 &= \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{1}{\beta}\right)^{-(\alpha_1+1)} \Gamma(\alpha_1+1) \frac{\beta^{-\alpha_2}}{\Gamma(\alpha_2)} \left(\frac{1}{\beta}\right)^{-(\alpha_2-1)} \Gamma(\alpha_2-1) \\
 &= \beta^{-(\alpha_1+\alpha_2)} \frac{1}{\beta^{-(\alpha_1+\alpha_2)}} \frac{\Gamma(\alpha_1+1)\Gamma(\alpha_2-1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \\
 &= \frac{\alpha_1\Gamma(\alpha_1)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_2-1)}{(\alpha_2-1)\Gamma(\alpha_2-1)} = \frac{\alpha_1}{\alpha_2-1}
 \end{aligned}$$

c) Sea suceso $A = \text{"detección correcta de símbolo 1"}$; entonces

$$\begin{aligned}
 P(A) &= P(\mathbf{Z} \geq 1) = P\left(\frac{\mathbf{X}_1}{\mathbf{X}_2} \geq 1\right) = P(\mathbf{X}_1 \geq \mathbf{X}_2) \\
 &= \int_0^\infty dx_1 \int_0^{x_1} f_{\mathbf{X}_1 \mathbf{X}_2}(x_1, x_2) dx_1 dx_2 = \\
 &= \int_0^\infty f_{\mathbf{X}_1}(x_1) dx_1 \int_0^{x_1} f_{\mathbf{X}_2}(x_2) dx_2 = \int_0^\infty F_{\mathbf{X}_2}(x_1) f_{\mathbf{X}_1}(x_1) dx_1 \\
 &= \int_0^\infty \left(1 - e^{-x_1/\beta} \left[1 + \frac{x_1}{\beta}\right]\right) f_{\mathbf{X}_1}(x_1) dx_1 \\
 &= \int_0^\infty f_{\mathbf{X}_1}(x_1) dx_1 - \int_0^\infty e^{-x_1/\beta} \left[1 + \frac{x_1}{\beta}\right] f_{\mathbf{X}_1}(x_1) dx_1 \\
 &= 1 - \int_0^\infty e^{-x_1/\beta} \left[1 + \frac{x_1}{\beta}\right] \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1-1} e^{-x_1/\beta} dx_1 \\
 &= 1 - \int_0^\infty \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1-1} e^{-2x_1/\beta} dx_1 - \int_0^\infty \frac{\beta^{-(\alpha_1+1)}}{\Gamma(\alpha_1)} x_1^{\alpha_1} e^{-2x_1/\beta} dx_1 \\
 &= 1 - \frac{\beta^{-\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{2}{\beta}\right)^{-\alpha_1} \Gamma(\alpha_1) - \frac{\beta^{-(\alpha_1+1)}}{\Gamma(\alpha_1)} \left(\frac{2}{\beta}\right)^{-(\alpha_1+1)} \Gamma(\alpha_1+1) \\
 &= 1 - 2^{-\alpha_1} - \alpha_1 2^{-(\alpha_1+1)} = 1 - \frac{1}{8} - \frac{3}{16} = \frac{11}{16}
 \end{aligned}$$



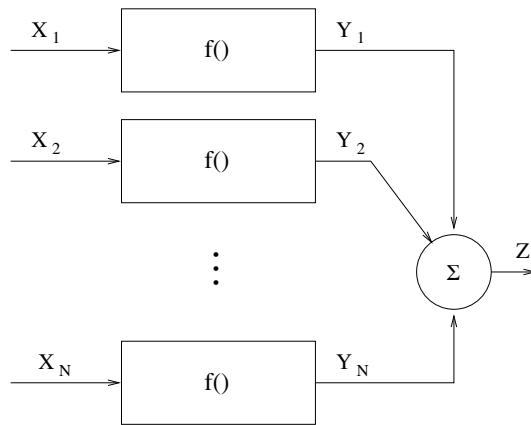


Figura 4.2: Figura problema 3.

3.- La figura adjunta representa el diagrama de bloques de una parte de un detector radar. Las variables \mathbf{X}_i son independientes e idénticamente distribuidas, con una función de densidad de probabilidad para $x \geq 0$ igual a $f_{\mathbf{X}_i}(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp(-(\frac{x}{\beta})^\alpha)$, y nula en el resto de la recta real, $\forall i = \{1, 2, \dots, N\}$. Se sabe asimismo que la media de cada variable $E\{\mathbf{X}_i\} = \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$. En estas condiciones, se pide:

a) Considerando que $f(\mathbf{X}_i) = \mathbf{X}_i$, proporcione, en función de N , una cota a la probabilidad de que la variable $\mathbf{Z} = \sum_{i=1}^N \mathbf{Y}_i$ presente valores alejados de su media una cantidad superior a tres veces dicha media (particularice para $\alpha = 1/2$).

b) Obtenga de forma aproximada la probabilidad de que la variable \mathbf{Z} presente valores alejados de su media una cantidad superior a dos veces su desviación típica, considerando que el número de variables que intervienen es elevado.

c) Obtenga la función característica de una variable \mathbf{U} cuya función de densidad de probabilidad es la que se indica (a y ϕ son parámetros positivos):

$$f_{\mathbf{U}}(u) = \begin{cases} \frac{\phi^{-a}}{\Gamma(a)} u^{a-1} \exp(-\frac{u}{\phi}) & u \geq 0 \\ 0 & u < 0 \end{cases}$$

d) Obtenga, mediante funciones características, la función de densidad de probabilidad de la variable \mathbf{Z} , considerando ahora que $f(\mathbf{X}_i) = (\mathbf{X}_i)^\alpha$.

NOTA:

$$G(2) = 0.9773$$

$$\Gamma(p) = k^p \int_0^\infty \tau^{p-1} \exp(-k\tau) d\tau; p, k > 0$$

$$\Gamma(p+1) = p\Gamma(p), p \in \mathcal{R}; \Gamma(\frac{1}{2}) = \sqrt{\pi}; \Gamma(z+1) = z!, z \in \mathcal{N}$$

Solución:

- a) Recordando que la desigualdad de Tchebycheff (véase sección 2.5.5)

$$P(|\mathbf{Z} - \eta_{\mathbf{Z}}| \geq \epsilon) \leq \frac{\sigma_{\mathbf{Z}}^2}{\epsilon^2}$$

donde, según dice el enunciado, $\epsilon = 3\eta_{\mathbf{Z}}$, entonces el problema se reduce a calcular la media y la varianza de la VA \mathbf{Z} . En las condiciones del apartado se verifica que $\mathbf{Z} = \sum_{i=1}^N \mathbf{Y}_i = \sum_{i=1}^N \mathbf{X}_i$ de forma que

$$\begin{aligned} E\{\mathbf{Z}\} &= \sum_{i=1}^N E\{\mathbf{X}_i\} = N \frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) \\ \sigma_{\mathbf{Z}}^2 &= \sum_{i=1}^N \sigma_{\mathbf{X}_i}^2 = \sum_{i=1}^N \left(E\{\mathbf{X}_i^2\} - E^2\{\mathbf{X}_i\} \right) \end{aligned}$$

donde se ha hecho uso de la expresión (4.18). Respecto del VCM de cada VA \mathbf{X}_i

$$\begin{aligned} E\{\mathbf{X}_i^2\} &= \int_0^\infty x^2 \alpha \beta^{-\alpha} x^{\alpha-1} \exp(-(\frac{x}{\beta})^\alpha) dx \\ &= \int_0^\infty \alpha \beta^{-\alpha} x^{\alpha+1} \exp(-(\frac{x}{\beta})^\alpha) dx \end{aligned}$$

Según el enunciado $\Gamma(p) = k^p \int_0^\infty \tau^{p-1} \exp(-k\tau) d\tau$, de forma que haciendo $x^\alpha = \tau$ la ecuación anterior se puede escribir

$$\begin{aligned} E\{\mathbf{X}_i^2\} &= \int_0^\infty \beta^{-\alpha} \tau^{\frac{1+\alpha}{\alpha}} e^{-\frac{\tau}{\beta^\alpha}} \frac{d\tau}{\tau^{\frac{\alpha-1}{\alpha}}} \\ &= \int_0^\infty \beta^{-\alpha} \tau^{\frac{1+\alpha-\alpha+1}{\alpha}} e^{-\frac{\tau}{\beta^\alpha}} d\tau \\ &= \int_0^\infty \beta^{-\alpha} \tau^{\frac{2}{\alpha}} e^{-\frac{\tau}{\beta^\alpha}} d\tau \end{aligned}$$

Denominando $k = 1/\beta^\alpha$ y $p-1 = 2/\alpha \rightarrow p = 2/\alpha + 1$ podemos escribir

$$\begin{aligned} E\{\mathbf{X}_i^2\} &= k^{-p} \int_0^\infty \beta^{-\alpha} k^p \tau^{p-1} e^{-k\tau} d\tau = k^{-p} \beta^{-\alpha} \Gamma(p) \\ &= (\beta^\alpha)^{\frac{2+\alpha}{\alpha}} \beta^{-\alpha} \Gamma\left(\frac{2}{\alpha} + 1\right) \\ &= \beta^2 \frac{2}{\alpha} \Gamma\left(\frac{2}{\alpha}\right) \end{aligned}$$

Por ello

$$\begin{aligned}\sigma_{\mathbf{x}_i}^2 &= \beta^2 \frac{2}{\alpha} \Gamma\left(\frac{2}{\alpha}\right) - \frac{\beta^2}{\alpha^2} \Gamma^2\left(\frac{1}{\alpha}\right) = \frac{\beta^2}{\alpha} \left[2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \Gamma^2\left(\frac{1}{\alpha}\right)\right] \\ \sigma_{\mathbf{Z}}^2 &= N \sigma_{\mathbf{x}_i}^2\end{aligned}$$

siendo estas expresiones válidas $\forall i = 1, \dots, N$. Así pues la cota pedida es

$$\begin{aligned}\frac{\sigma_{\mathbf{Z}}^2}{\epsilon^2} &= \frac{\sigma_{\mathbf{Z}}^2}{(3\eta_{\mathbf{Z}})^2} = \frac{N \frac{\beta^2}{\alpha} \left[2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \Gamma^2\left(\frac{1}{\alpha}\right)\right]}{9 \frac{N^2 \beta^2}{\alpha^2} \Gamma^2\left(\frac{1}{\alpha}\right)} \\ &= \frac{\alpha}{9N} \frac{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \Gamma^2\left(\frac{1}{\alpha}\right)}{\Gamma^2\left(\frac{1}{\alpha}\right)}\end{aligned}$$

Particularizando para $\alpha = 1/2$ resulta

$$\frac{\sigma_{\mathbf{Z}}^2}{\epsilon^2} = \frac{1}{2} \frac{1}{9N} \frac{2\Gamma(4) - 2\Gamma^2(2)}{\Gamma^2(2)} = \frac{1}{2} \frac{1}{9N} \frac{10}{1} = \frac{5}{9N}$$

b)

$$\begin{aligned}P(|\mathbf{Z} - \eta_{\mathbf{Z}}| > 2\sigma_{\mathbf{Z}}) &= 1 - P(|\mathbf{Z} - \eta_{\mathbf{Z}}| \leq 2\sigma_{\mathbf{Z}}) \\ &= 1 - P(-2\sigma_{\mathbf{Z}} < \mathbf{Z} - \eta_{\mathbf{Z}} \leq 2\sigma_{\mathbf{Z}}) \\ &= 1 - P(\eta_{\mathbf{Z}} - 2\sigma_{\mathbf{Z}} < \mathbf{Z} \leq \eta_{\mathbf{Z}} + 2\sigma_{\mathbf{Z}}) \\ &\approx 1 - \left[G\left(\frac{\eta_{\mathbf{Z}} + 2\sigma_{\mathbf{Z}} - \eta_{\mathbf{Z}}}{\sigma_{\mathbf{Z}}}\right) - G\left(\frac{\eta_{\mathbf{Z}} - 2\sigma_{\mathbf{Z}} - \eta_{\mathbf{Z}}}{\sigma_{\mathbf{Z}}}\right)\right] \\ &= 1 - [G(2) - G(-2)] = 1 - [G(2) - (1 - G(2))] \\ &= 1 - [2G(2) - 1] = 1 - 0.9546 = 0.0454\end{aligned}$$

donde se ha empleado el Teorema del Límite Central (véase sección 4.5.1), así como la simetría de la función $G(\cdot)$ (véase página 62).

c) Si escribimos la definición de función característica (ecuación 2.95) adaptada al caso que se plantea

$$\phi_{\mathbf{U}}(\omega) = E\{e^{j\omega \mathbf{U}}\} = \int_0^\infty f_{\mathbf{U}}(u) e^{j\omega u} du = \frac{\phi^{-a}}{\Gamma(a)} \int_0^\infty u^{a-1} e^{-(\frac{1}{\phi} - j\omega)u} du$$

parece claro que podemos explotar el parecido entre esta expresión y la definición de la función $\Gamma(\cdot)$ (expresión (2.63)). En concreto haciendo $k =$

$\frac{1}{\phi} - j\omega$ y $a = p$ entonces

$$\begin{aligned}\phi_{\mathbf{U}}(\omega) &= \frac{\phi^{-a}}{\Gamma(a)} k^{-a} \int_0^\infty k^a u^{a-1} e^{-ku} du \\ &= \frac{\phi^{-a}}{\Gamma(a)} k^{-a} \Gamma(a) \\ &= \phi^{-a} \left(\frac{1 - j\omega\phi}{\phi} \right)^{-a} = \left(\frac{1}{1 - j\omega\phi} \right)^a\end{aligned}$$

d) En este caso $\mathbf{Z} = \sum_{i=1}^N \mathbf{Y}_i = \sum_{i=1}^N \mathbf{X}_i^\alpha$. Por ello

$$\phi_{\mathbf{Z}}(\omega) = E\{e^{j\omega\mathbf{Z}}\} = E\{e^{j\omega \sum_{i=1}^N \mathbf{X}_i^\alpha}\} = \prod_{i=1}^N E\{e^{j\omega \mathbf{X}_i^\alpha}\} = \prod_{i=1}^N \phi_{\mathbf{X}_i^\alpha}(\omega)$$

donde se ha hecho uso de la independencia de las variables \mathbf{X}_i para convertir la esperanza del producto en el producto de las esperanzas (véase página 156 y ecuación 4.6). Con respecto a la función característica de las VAs \mathbf{X}_i^α

$$\begin{aligned}\phi_{\mathbf{X}_i^\alpha}(\omega) &= E\{e^{j\omega \mathbf{X}_i^\alpha}\} = \int_0^\infty \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(\frac{x}{\beta})^\alpha} e^{j\omega x^\alpha} dx \\ &= \int_0^\infty \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(\frac{1}{\beta^\alpha} - j\omega)x^\alpha} dx \\ &= \beta^{-\alpha} \left[\frac{-1}{\frac{1}{\beta^\alpha} - j\omega} e^{-(\frac{1}{\beta^\alpha} - j\omega)x^\alpha} \right]_0^\infty \\ &= \beta^{-\alpha} \frac{1}{\frac{1}{\beta^\alpha} - j\omega} = \frac{1}{1 - j\omega\beta^\alpha}\end{aligned}$$

Por lo que

$$\phi_{\mathbf{Z}}(\omega) = \left(\frac{1}{1 - j\omega\beta^\alpha} \right)^N$$

Según el resultado del apartado anterior, podemos escribir

$$f_{\mathbf{Z}}(z) = \frac{\phi^{-a}}{\Gamma(a)} z^{a-1} \exp\left(-\frac{z}{\phi}\right), z > 0$$

con $a = N$ y $\phi = \beta^\alpha$.



4.- Sean $\mathbf{Y}_i, i = \{1, \dots, n\}$, n variables aleatorias independientes e idénticamente distribuidas (IID), con función de distribución $F(y)$ (común para todas), tal que $F(y) = 0 \forall y < 0$. Se pide:

a) Definiendo $\mathbf{Z} = \max(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ y $\mathbf{W} = \min(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$, obtenga la función $F_{\mathbf{ZW}}(z, w)$ para $z \geq w \geq 0$.

b) Suponiendo que \mathbf{Y}_0 es una variable IID con respecto a las \mathbf{Y}_i anteriormente definidas, obtenga $P(\mathbf{W} < \mathbf{Y}_0 \leq \mathbf{Z})$ como función exclusivamente de n . Para resolver las expresiones que obtenga se recomienda que analice el parecido formal de la función $f_{\mathbf{ZW}}(z, w)$ calculada a partir de n y $n + 1$ variables IID.

c) Suponga que \mathbf{X}_j es una variable que toma el valor 1 cuando $\{\mathbf{W} < \mathbf{Y}_0 \leq \mathbf{Z}\}$ y 0 en cualquier otro caso, con $j = \{1, \dots, N\}$ y N el número de veces que se ejecutan experimentos independientes del tipo del indicado ($N \gg 1$). Se pide que obtenga el valor de ϵ que hace que

$$P\left(\left|\frac{1}{N} \sum_{j=1}^N \mathbf{X}_j - p\right| < \epsilon\right) = 1 - \alpha$$

con α un valor de probabilidad, y p la probabilidad calculada en el apartado anterior (si no contestó a éste emplee un p genérico). Particularice para $\alpha = 0.05$.

Solución:

a) Para hallar la función de distribución conjunta $F_{\mathbf{ZW}}(z, w)$ en el punto que se indica definamos el suceso B_k como “hay k variables a la izquierda de w y $n - k$ entre w y z ”. Asimismo, dado que las VAs \mathbf{Y}_i son IID empleemos el símbolo \mathbf{Y} para referirnos a una cualquiera de ellas. En estas condiciones, y siguiendo un razonamiento similar al expuesto en la sección 1.7.1, podemos escribir

$$\begin{aligned} P(B_k) &= \binom{n}{k} P(\mathbf{Y} < w)^k P(w < \mathbf{Y} \leq z)^{n-k} \\ &= \binom{n}{k} F_{\mathbf{Y}}^k(w)(F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w))^{n-k} \end{aligned}$$

Entonces

$$\begin{aligned} F_{\mathbf{ZW}}(z, w) &= P\left(\bigcup_{k=1}^n B_k\right) = \sum_{k=1}^n P(B_k) \\ &= \sum_{k=1}^n \binom{n}{k} F_{\mathbf{Y}}^k(w)(F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w))^{n-k} \\ &= F_{\mathbf{Y}}(z) - P(B_0) = F_{\mathbf{Y}}(z) - [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^n \end{aligned}$$

con $0 \leq w \leq z$.

b) Recordando la definición de función de densidad conjunta (ecuación 3.7)

$$\begin{aligned} f_{\mathbf{ZW}}(z, w) &= \frac{\partial^2 F_{\mathbf{ZW}}(z, w)}{\partial z \partial w} \\ &= \frac{\partial^2}{\partial z \partial w} [F_{\mathbf{Y}}(z) - [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^n] \\ &= \frac{\partial}{\partial z} \left[n [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^{n-1} f_{\mathbf{Y}}(w) \right] \\ &= n(n-1) [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^{n-2} f_{\mathbf{Y}}(z) f_{\mathbf{Y}}(w) \end{aligned}$$

Para calcular la probabilidad que se pide, empleemos el Teorema de la Probabilidad Total (definición original en ecuación 1.29 y un caso particular del mismo para una única VA en la ecuación 2.36) adaptado a este caso. Concretamente

$$\begin{aligned} P(\mathbf{W} < \mathbf{Y}_0 \leq \mathbf{Z}) &= \int_{zw} P(w < \mathbf{Y}_0 \leq z | \mathbf{W} = w, \mathbf{Z} = z) f_{\mathbf{ZW}}(z, w) dz dw \\ &= \int_{zw} P(w < \mathbf{Y}_0 \leq z) f_{\mathbf{ZW}}(z, w) dz dw \\ &= \int_{zw} [F_{\mathbf{Y}_0}(z) - F_{\mathbf{Y}_0}(w)] f_{\mathbf{ZW}}(z, w) dz dw \\ &= \int_{zw} n(n-1) [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^{n-1} f_{\mathbf{Y}}(z) f_{\mathbf{Y}}(w) dz dw \end{aligned}$$

donde se ha hecho uso del hecho de que la VA \mathbf{Y}_0 es IID con respecto a las VAs \mathbf{Y}_i , por lo que también es independiente de una función de las mismas, en particular, de las funciones máximo y mínimo. Definiendo ahora $f_{\mathbf{ZW}}^{n+1}(z, w)$ como la función de densidad conjunta del máximo y mínimo obtenido a partir de $n+1$ VAs \mathbf{Y}_i , y comparando con la expresión que se acaba de obtener, es claro que

$$f_{\mathbf{ZW}}^{n+1}(z, w) = (n+1)n [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^{n-1} f_{\mathbf{Y}}(z) f_{\mathbf{Y}}(w)$$

de modo que podemos continuar el razonamiento anterior de la forma

$$\begin{aligned} P(\mathbf{W} < \mathbf{Y}_0 \leq \mathbf{Z}) &= \frac{n(n-1)}{n(n+1)} \times \\ &\quad \int_{zw} n(n+1) [F_{\mathbf{Y}}(z) - F_{\mathbf{Y}}(w)]^{n-1} f_{\mathbf{Y}}(z) f_{\mathbf{Y}}(w) dz dw \end{aligned}$$

$$\begin{aligned}
&= \frac{n(n-1)}{n(n+1)} \int_{zw} f_{\mathbf{ZW}}^{n+1}(z, w) dz dw \\
&= \frac{n(n-1)}{n(n+1)} = \frac{(n-1)}{(n+1)}
\end{aligned}$$

resultado que coincide con el obtenido en el ejercicio de la página 23.

c) Las VAs \mathbf{X}_i son variables de Bernoulli con probabilidades

$$\mathbf{X}_i \left\{ \begin{array}{l} p(\mathbf{X}_i = 1) = p = \frac{n-1}{n+1} \\ p(\mathbf{X}_i = 0) = 1 - p = q = 1 - \frac{n-1}{n+1} = \frac{2}{n+1} \end{array} \right.$$

Definiendo

$$\mathbf{U} = \frac{\mathbf{V}}{N} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i$$

y siendo $N \gg 1$ podemos hacer uso de la Ley de los Grandes Números (sección 4.5.3) y escribir

$$\begin{aligned}
P(|\mathbf{U} - p| < \epsilon) &= P(-\epsilon < \mathbf{U} - p \leq \epsilon) = P(p - \epsilon < \mathbf{U} \leq p + \epsilon) \\
&= P(p - \epsilon < \frac{\mathbf{V}}{N} \leq p + \epsilon) = P(N(p - \epsilon) < \mathbf{V} \leq N(p + \epsilon)) \\
&\approx G\left(\frac{N(p + \epsilon) - Np}{\sqrt{Npq}}\right) - G\left(\frac{N(p - \epsilon) - Np}{\sqrt{Npq}}\right) \\
&= G\left(\frac{N\epsilon}{\sqrt{Npq}}\right) - G\left(-\frac{N\epsilon}{\sqrt{Npq}}\right) = 2G\left(\frac{N\epsilon}{\sqrt{Npq}}\right) - 1 \\
&= 1 - \alpha
\end{aligned}$$

donde se ha hecho uso de la simetría de la función $G(\cdot)$ (véase página 62). Entonces

$$\begin{aligned}
2G\left(\frac{N\epsilon}{\sqrt{Npq}}\right) - 1 &= 1 - \alpha \\
G\left(\frac{N\epsilon}{\sqrt{Npq}}\right) &= 1 - \frac{\alpha}{2} \\
G^{-1}\left(1 - \frac{\alpha}{2}\right) = G^{-1}(0.9750) &= \epsilon \sqrt{\frac{N}{pq}} = 1.96
\end{aligned}$$

por lo que

$$\epsilon = 1.96 \sqrt{\frac{pq}{N}}$$



5.-Dos especialistas médicos diagnostican un número N de radiografías. En cada una de ellas el juicio de cada especialista puede ser *presencia de patología* (PP, al que asociaremos el valor numérico 0) o (AP, al que asociaremos el valor numérico 1). Se asumirá que no existen dependencias entre juicios de dos radiografías distintas. La tabla adjunta refleja los resultados obtenidos para esas radiografías:

Esp1 \ Esp2	PP	AP
PP	a	b
AP	c	d

con $a, b, c, d \geq 0$ y $a + b + c + d = N$. Se define la variable aleatoria binaria \mathbf{X}_i como el *valor numérico del diagnóstico* del primer especialista para la radiografía i ($i = \{1, \dots, N\}$), y, de forma equivalente, \mathbf{Y}_i para el segundo especialista. Denominemos \mathbf{I} a la variable *número de veces que coinciden los juicios de los especialistas*, y, para una radiografía cualquiera, se definen p_j^x como la probabilidad de que el primer especialista diagnostique resultado j , ($j=(0,1)$), p_k^y como la probabilidad de que el segundo diagnostique k ($k = (0, 1)$), y p_{jk} como la probabilidad de que, conjuntamente, el primero diagnostique j y el segundo k , con $j, k = (0, 1)$. Se pide:

a) Denominando \mathbf{I} a la variable *número de veces que coinciden los juicios de los especialistas*, escriba \mathbf{I} como función de \mathbf{X}_i e \mathbf{Y}_i , con $i = \{1, \dots, N\}$.

b) Se pretende analizar probabilísticamente si los especialistas tienen el mismo criterio. Para ello, denominemos hipótesis H_0 a la hipótesis *los especialistas comparten el mismo criterio* (entendemos por tal que los juicios respectivos sobre una misma radiografía sean dependientes) y H_1 a la hipótesis contraria (es decir, que los juicios respectivos sean independientes). En estas condiciones obtenga $f_{\mathbf{I}}(i|H_0)$ empleando sólo las p_{jk} , y $f_{\mathbf{I}}(i|H_1)$ empleando sólo las p_j^x y las p_k^y .

c) Se considerará que los especialistas comparten el mismo criterio si $E\{\mathbf{I}|H_0\} > E\{\mathbf{I}|H_1\}$. Suponiendo que las probabilidades p_j^x , p_k^y y p_{jk} son desconocidas, se pide que proporcione una regla, función de a, b, c y d , que permita materializar el criterio anterior. Para introducir estos parámetros en el problema identificaremos probabilidad con frecuencia relativa, calculada esta última a partir de los datos de la tabla adjunta.

Solución:

a) $\mathbf{I} = \sum_{i=1}^N \delta[\mathbf{X}_i - \mathbf{Y}_i]$ donde, $\mathbf{X}_i = 0$ si el primer especialista dice PP y $\mathbf{X}_i = 1$ en caso contrario, y lo propio para el segundo especialista

con la VA \mathbf{Y}_i .

b) Por definición la VA \mathbf{I} es binomial en ambos supuestos (esto es, en supuestos H_0 y H_1), ya que esta variable se construye en base a una composición de ensayos de Bernoulli independientes (véase secciones 1.7.1 y 2.3.2). Lo que cambia entre H_0 y H_1 es la probabilidad de cada valor de la VA \mathbf{I} .

Así pues

$$\begin{aligned}\mathbf{I}_{H_0} &\sim B(N, p_{H_0}) \\ \mathbf{I}_{H_1} &\sim B(N, p_{H_1})\end{aligned}$$

donde

$$\begin{aligned}p_{H_j} &= p(\mathbf{X}_i = \mathbf{Y}_i | H_j) = p((\mathbf{X}_i = 1, \mathbf{Y}_i = 1) \cup (\mathbf{X}_i = 0, \mathbf{Y}_i = 0) | H_j) \\ &= p(\mathbf{X}_i = 1, \mathbf{Y}_i = 1 | H_j) + p(\mathbf{X}_i = 0, \mathbf{Y}_i = 0 | H_j) \\ p_{H_0} &= p_{11} + p_{00} \\ p_{H_1} &= p(\mathbf{X}_i = 1 | H_1)p(\mathbf{Y}_i = 1 | H_1) + p(\mathbf{X}_i = 0 | H_1)p(\mathbf{Y}_i = 0 | H_1) \\ &= p_1^x p_1^y + p_0^x p_0^y\end{aligned}$$

c) La media de las VAs binomiales anteriores, acorde con la expresión (2.64), es

$$\begin{aligned}E\{\mathbf{I}_{H_0}\} &= Np_{H_0} \\ E\{\mathbf{I}_{H_1}\} &= Np_{H_1}\end{aligned}$$

Para calcular las probabilidades p_{H_0} y p_{H_1} empleamos los datos siguiendo las expresiones escritas en el apartado anterior, e identificando — según indica el enunciado — probabilidad con frecuencia relativa. Así pues

$$\begin{aligned}p_{H_j} &= p(\mathbf{X}_i = \mathbf{Y}_i | H_j) = p((\mathbf{X}_i = 1, \mathbf{Y}_i = 1) \cup (\mathbf{X}_i = 0, \mathbf{Y}_i = 0) | H_j) \\ p_{H_0} &= p(\mathbf{X}_i = 1, \mathbf{Y}_i = 1 | H_j) + p(\mathbf{X}_i = 0, \mathbf{Y}_i = 0 | H_j) \\ &= \frac{d}{N} + \frac{a}{N} = \frac{a+d}{N} \\ p_{H_1} &= p(\mathbf{X}_i = 1 | H_1)p(\mathbf{Y}_i = 1 | H_1) + p(\mathbf{X}_i = 0 | H_1)p(\mathbf{Y}_i = 0 | H_1) \\ &= \frac{c+d}{N} \frac{b+d}{N} + \frac{a+b}{N} \frac{a+c}{N} = \frac{1}{N^2} ((a+b)(a+c) + (b+d)(c+d))\end{aligned}$$

La regla por tanto sería decidir H_0 si

$$\begin{aligned} E\{\mathbf{I}_{H_0}\} &> E\{\mathbf{I}_{H_1}\} \\ a+d &> \frac{(a+b)(a+c)}{N} + \frac{(b+d)(c+d)}{N} \end{aligned}$$



Capítulo 5

Procesos estocásticos

Hasta ahora, hemos abordado el estudio de los conceptos básicos de Teoría de la Probabilidad, así como de las VAs a través de un planteamiento incremental de una, dos y N variables.

En este tema encontraremos ya la relación que liga a las señales con las VAs. Toda señal que transporta información tendrá algún grado de aleatoriedad, de forma que en general no podremos predecir sin error el valor que tomará una señal en el futuro, conocidos los valores que ha tomado en el pasado. Pero, no obstante, hemos visto formas de predecir de manera óptima haciendo uso de la información probabilística de que dispongamos. Por otra parte, las señales de comunicaciones siempre se mueven en entornos ruidosos, siendo el ruido también una señal aleatoria, en este caso indeseada. Por ello, será necesario disponer de herramientas que nos permitan minimizar el efecto del ruido.

En el tema que comenzamos, por lo tanto, conectaremos los conceptos vistos hasta ahora con el tratamiento de la señal.

5.1 Concepto de proceso estocástico. Clasificación

Un proceso estocástico es una regla que asigna a cada resultado de un cierto experimento aleatorio $\epsilon < S, \mathcal{F}, P >$ una *función*, la cual depende de un determinado número de variables, y del propio resultado $a \in S$ obtenido del experimento (figura 5.1). El caso más sencillo sería una función de una única variable, por ejemplo, el tiempo. En tal caso, un proceso estocástico sería una colección de funciones del tiempo, cada una de ellas asociada a cada uno de los resultados $a \in S$. Pero, también, la variables podrían

ser 2 y con significado de coordenadas espaciales. En tal caso, un proceso estocástico podría ser una colección de imágenes estáticas, cada una de ellas asociada a cada resultado del espacio muestral. Si fuesen tres variables, por ejemplo, dos espaciales y una temporal, podríamos considerar un proceso estocástico como una colección de imágenes en movimiento.

En esta asignatura nos centraremos en el caso de señales unidimensionales funciones del tiempo.

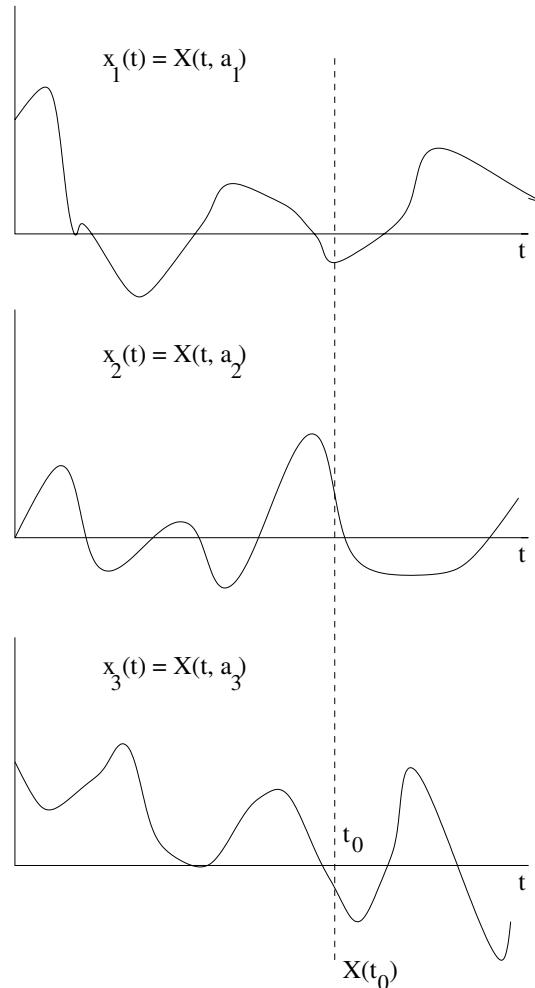


Figura 5.1: Concepto de proceso estocástico: colección de funciones del tiempo o de VAs.

El concepto de proceso estocástico es muy similar al concepto de una

VA. Recuérdese que una variable era una función del espacio muestral en el cuerpo de los números reales, de forma que para cada $a \in S$ obteníamos un número real $\mathbf{X}(a)$. No obstante, por simplicidad en la notación, nos hemos referido siempre a las VAs sin explicitar la dependencia con el resultado a del experimento aleatorio. En el caso de los procesos estocásticos (particularizados a funciones unidimensionales del tiempo) la dependencia es similar. A cada $a \in S$ le asociamos una función del tiempo $\mathbf{X}(t, a)$. De forma similar al caso de las VAs, denotaremos al proceso estocástico como $\mathbf{X}(t)$ dando por sentado la dependencia con el resultado aleatorio.

No obstante, antes de eliminar de la notación esta dependencia, hagamos una última aclaración. Para ello, tomaremos como base tanto la figura 5.1 como el proceso estocástico $\mathbf{Y}(t, a) = A \cos(\omega_0 t + \Theta(a))$, con $\Theta(a)$ una VA uniforme en un rango de 2π radianes, y el resto de los parámetros involucrados determinísticos.

- $\mathbf{X}(t, a)$ con los dos parámetros libres, es un proceso estocástico.
- $\mathbf{X}(t, a_i)$ donde a_i es un resultado concreto del experimento aleatorio, es decir, con la primera variable libre y fija la segunda, es una función *determinística* del tiempo. En este caso estamos ante una *realización* del proceso. Correspondría en la figura 5.1 a una cualquiera de las funciones $x_i(t)$. Por su parte, en el proceso estocástico $\mathbf{Y}(t, a)$ tal particularización correspondería a la función del tiempo observada para el valor particular de la variable $\Theta(a_i)$, por ejemplo, $\mathbf{Y}(t, a_i) = A \cos(\omega_0 t + \frac{\pi}{3})$. El equivalente en el caso de las VAs sería el haber observado un valor concreto de la variable \mathbf{X} , es decir, haber observado un número real.
- $\mathbf{X}(t_0, a)$, donde t_0 indica un instante de tiempo concreto, es decir, con la primera variable fija y libre la segunda, sería una VA: en efecto, el resultado para cada $a \in S$ sería un número real, lo cual corresponde en la figura 5.1 al conjunto de números que podríamos observar sobre la vertical de abscisa $t = t_0$. Respecto del proceso $\mathbf{Y}(t, a)$, fijado el instante temporal en $t = t_0$ tendríamos $\mathbf{Y}(t_0, a) = A \cos(\omega_0 t_0 + \Theta(a))$, función de la VA $\Theta(a)$ y, por tanto, tendríamos una VA. Por todo ello, una visión alternativa de un proceso estocástico es la de una colección de VAs indexadas por un índice continuo¹ t .

¹El índice no tiene que ser, necesariamente, continuo. Veremos a continuación que éste puede ser discreto.

- $\mathbf{X}(t_0, a_i)$, esto es, ambas variables fijas, sería un número real. En la figura 5.1 seleccionaríamos una función concreta $x_i(t)$ y observaríamos un instante temporal concreto $t = t_0$. Para el caso del proceso $\mathbf{Y}(t, a)$, fijados $t = t_0$ y $a = a_i$ tendríamos, por ejemplo, el escalar $\mathbf{Y}(t_0, a_i) = A \cos(\omega_0 t_0 + \frac{\pi}{3})$.

5.1.1 Clasificación de procesos estocásticos

Una posible clasificación de los procesos estocásticos consistiría en analizar tanto la variable temporal como las características de cada una de las VAs involucradas. Concretamente:

1. Proceso estocástico continuo: en este caso, la variable t es continua, y cada una de las variables de $\mathbf{X}(t)$ toman valores en un rango continuo.
2. Proceso estocástico discreto: en este caso, la variable t es continua, pero las variables de $\mathbf{X}(t)$ son VAs discretas.
3. Secuencia aleatoria continua: la variable de indexación temporal es discreta, pero las VAs involucradas toman valores en un rango continuo. Típicamente lo denotaremos por $\mathbf{X}[n]$.
4. Secuencia aleatoria discreta: secuencia (como la anterior) de VAs discretas. La denotaremos como la anterior.

Se podrían hacer otra clasificaciones. Por ejemplo, si un proceso estocástico es predecible sin error a partir de la observación de los valores del mismo en el pasado, el proceso se denomina **predecible**. Un ejemplo típico sería el proceso ya empleado $\mathbf{Y}(t) = A \cos(\omega_0 t + \Theta)$, con Θ una VA. En este caso, hablaríamos de una sinusoides cuya fase inicial es desconocida. No obstante, si se observa lo que vale el proceso en un instante se puede calcular el valor que ha tomado la variable Θ , de forma que a partir de ahí el proceso sería perfectamente predecible.

Por contra, si tal predicción no se puede llevar a cabo, el proceso sería impredecible, y, por tanto el comportamiento del mismo no se podría resumir en un conjunto finito de parámetros.

Otras clasificaciones de los procesos requieren de más conocimiento del que hemos expuesto hasta el momento.

5.2 Funciones de distribución y densidad

Dado que un proceso estocástico no es otra cosa que una colección de VAs, indexadas éstas mediante un índice continuo (procesos estocásticos propiamente dichos) o discreto (secuencias aleatorias), podemos hablar también de funciones de densidad y distribución del proceso, bien entendido que tales funciones se definen sobre *cada una* de las VAs del proceso. Por tanto, podemos definir

- Función de densidad (de primer orden) del proceso

$$F_{\mathbf{X}}(x; t) = P(\mathbf{X}(t) \leq x) \quad (5.1)$$

Como puede verse, la función de distribución es función de dos variables. La primera (variable x) tiene el significado tradicional: el punto de abscisa donde estamos evaluando la probabilidad del suceso indicado. La segunda variable es, simplemente, el índice que indica sobre qué VA del proceso estamos haciendo el cálculo de la función de distribución.

Esta función se denomina *de primer orden* dado que sólo involucra, para cada t , a una variable del proceso.

- Función de densidad de probabilidad de primer orden. De forma coherente con lo que conocemos hasta ahora, la función de densidad se obtiene a partir de la función de distribución de la forma:

$$f_{\mathbf{X}}(x; t) = \frac{dF_{\mathbf{X}}(x; t)}{dx} \quad (5.2)$$

Nada nos impide extraer más de una variable del proceso. Si extraemos dos, léase, $(\mathbf{X}(t_1), \mathbf{X}(t_2))$, habríamos creado una variable bidimensional a partir del proceso. Si extraemos N podemos crear un vector aleatorio a partir de variables del proceso de la forma $[\mathbf{X}(t_1), \mathbf{X}(t_2), \dots, \mathbf{X}(t_N)]^T$, por lo que los conceptos ya conocidos de temas anteriores aplican directamente. A modo de ejemplo:

- Funciones de distribución y densidad de segundo orden

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2; t_1, t_2) &= P(\mathbf{X}(t_1) \leq x_1, \mathbf{X}(t_2) \leq x_2) \\ f_{\mathbf{X}}(x_1, x_2; t_1, t_2) &= \frac{\partial^2 F_{\mathbf{X}}(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2} \end{aligned}$$

- Obtención de las marginales a partir de las anteriores

$$\begin{aligned} F_{\mathbf{X}}(x_1, t_1) &= F_{\mathbf{X}}(x_1, \infty; t_1, t_2) \\ f_{\mathbf{X}}(x_1, t_1) &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2; t_1, t_2) dx_2 \end{aligned}$$

- Funciones de distribución y densidad de orden N

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) &= P\left(\bigcap_{i=1}^N \mathbf{X}(t_i) \leq x_i\right) \\ f_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) &= \frac{\partial^N F_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N)}{\partial x_1 \partial x_2 \dots \partial x_N} \end{aligned}$$

- En el caso en que tengamos dos procesos, la caracterización probabilística total del mismo vendría dada por la función de densidad conjunta de órdenes N y M ($\forall N$ y $\forall M$), es decir, a partir de la función:

$$f_{\mathbf{XY}}(x_1, \dots, x_N, y_1, \dots, y_M; t_1, \dots, t_N, t'_1, \dots, t'_M) \quad (5.3)$$

En la práctica, salvo para procesos gaussianos, es impensable poder disponer de toda esta información probabilística, tanto para uno como para dos o más procesos, de forma que es habitual trabajar con parámetros de caracterización parcial del proceso, léase, medias, varianzas, covarianzas etc Éste es nuestro siguiente punto de atención.

5.3 Caracterización parcial de procesos estocásticos. Ruido blanco

- Media:

$$\eta_{\mathbf{X}}(t) = E\{\mathbf{X}(t)\} = \int_{-\infty}^{\infty} x f_{\mathbf{X}}(x; t) dx \quad (5.4)$$

Como puede verse, la media de un proceso estocástico será en general función del tiempo, ya que esta función contiene la media de cada una de las variables que podemos extraer del proceso. Como las variables en general serán de diferente distribución, las medias en general no coincidirán.

- VCM:

$$E\{\mathbf{X}^2(t)\} = \int_{-\infty}^{\infty} x^2 f_{\mathbf{X}}(x; t) dx \quad (5.5)$$

- Varianza:

$$\begin{aligned}\sigma_{\mathbf{X}}^2(t) &= E\{(\mathbf{X}(t) - \eta_{\mathbf{X}}(t))^2\} = \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}}(t))^2 f_{\mathbf{X}}(x; t) dx \\ &= E\{\mathbf{X}^2(t)\} - \eta_{\mathbf{X}}^2(t)\end{aligned}\quad (5.6)$$

- Autocorrelación²: como es sabido, este parámetro involucra a dos VAs, de forma que, en relación con los procesos estocásticos, la función de correlación será función de dos variables temporales (los índices temporales de las dos variables involucradas).

$$R_{\mathbf{X}}(t_1, t_2) = E\{\mathbf{X}(t_1)\mathbf{X}(t_2)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{\mathbf{X}}(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (5.7)$$

Nótese que $R_{\mathbf{X}}(t, t) = E\{\mathbf{X}^2(t)\}$, es decir, es igual al VCM de la VA $\mathbf{X}(t)$.

- Autocovarianza

$$\begin{aligned}C_{\mathbf{X}}(t_1, t_2) &= E\{(\mathbf{X}(t_1) - \eta_{\mathbf{X}}(t_1))(\mathbf{X}(t_2) - \eta_{\mathbf{X}}(t_2))\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \eta_{\mathbf{X}}(t_1))(x_2 - \eta_{\mathbf{X}}(t_2)) f_{\mathbf{X}}(x_1, x_2; t_1, t_2) dx_1 dx_2 \\ &= R_{\mathbf{X}}(t_1, t_2) - \eta_{\mathbf{X}}(t_1)\eta_{\mathbf{X}}(t_2)\end{aligned}\quad (5.8)$$

Nótese que $C_{\mathbf{X}}(t, t) = E\{(\mathbf{X}(t) - \eta_{\mathbf{X}}(t))^2\} = \sigma_{\mathbf{X}}^2(t)$, es decir, es igual a la varianza de la variable aleatoria $\mathbf{X}(t)$.

- Coeficiente de correlación entre las variables $\mathbf{X}(t_1)$ y $\mathbf{X}(t_2)$

$$r_{\mathbf{X}}(t_1, t_2) = \frac{C_{\mathbf{X}}(t_1, t_2)}{\sigma_{\mathbf{X}}(t_1)\sigma_{\mathbf{X}}(t_2)} \quad (5.9)$$

- Para el caso de procesos complejos, las cosas funcionan como conocemos. Concretamente:

$$\begin{aligned}- \sigma_{\mathbf{X}}^2(t) &= E\{|\mathbf{X}(t) - \eta_{\mathbf{X}}(t)|^2\} = E\{|\mathbf{X}(t)|^2\} - |\eta_{\mathbf{X}}(t)|^2 \\ - R_{\mathbf{X}}(t_1, t_2) &= E\{\mathbf{X}(t_1)\mathbf{X}^*(t_2)\}\end{aligned}$$

²El prefijo *auto* enfatiza el hecho de que calculamos la correlación entre dos variables extraídas del mismo proceso. Cuando los procesos involucrados sean dos, la correlación se denominará correlación *cruzada*.

$$- C_{\mathbf{X}}(t_1, t_2) = E\{(\mathbf{X}(t_1) - \eta_{\mathbf{X}}(t_1))(\mathbf{X}(t_2) - \eta_{\mathbf{X}}(t_2))^*\} = R_{\mathbf{X}}(t_1, t_2) - \eta_{\mathbf{X}}(t_1)\eta_{\mathbf{X}}^*(t_2)$$

- Si los índices son discretos, esto es, si estamos tratando con una secuencia aleatoria, los operadores son los mismos, dado que en este caso el índice temporal simplemente es un seleccionador de la variable o variables empleadas para el cálculo. Por ejemplo, si la secuencia (real, por simplicidad) es $\mathbf{X}[n]$ podremos escribir
 - $\eta_{\mathbf{X}}[n] = E\{\mathbf{X}[n]\} = \int_{-\infty}^{\infty} xf_{\mathbf{X}}(x; n)dx$
 - $\sigma_{\mathbf{X}}^2[n] = E\{(\mathbf{X}[n] - \eta_{\mathbf{X}}[n])^2\} = E\{\mathbf{X}^2[n]\} - \eta_{\mathbf{X}}^2[n]$, sustituyendo t por n en la expresión (5.6).
 - $R_{\mathbf{X}}[n_1, n_2] = E\{\mathbf{X}[n_1]\mathbf{X}[n_2]\}$, sustituyendo t_1 y t_2 por n_1 y n_2 , respectivamente, en la expresión (5.7).
 - $C_{\mathbf{X}}[n_1, n_2] = E\{(\mathbf{X}[n_1] - \eta_{\mathbf{X}}[n_1])(\mathbf{X}[n_2] - \eta_{\mathbf{X}}[n_2])\} = R_{\mathbf{X}}[n_1, n_2] - \eta_{\mathbf{X}}[n_1]\eta_{\mathbf{X}}[n_2]$, haciendo la misma sustitución que en el caso anterior, ahora en la expresión (5.8).
- Finalmente, si el proceso o la secuencia consiste en variables discretas, los anteriores operadores probabilísticos continuos se pueden expresar como sumatorios en vez de como integrales. Así pues, tendríamos
 - $\eta_{\mathbf{X}}(t) = E\{\mathbf{X}(t)\} = \sum_i x_i P(\mathbf{X}(t) = x_i)$
 - $\sigma_{\mathbf{X}}^2(t) = \sum_i (x_i - \eta_{\mathbf{X}}(t))^2 P(\mathbf{X}(t) = x_i)$
 - $R_{\mathbf{X}}(t_1, t_2) = \sum_i \sum_j x_{1i} x_{2j} P(\mathbf{X}(t_1) = x_{1i}, \mathbf{X}(t_2) = x_{2j})$
 - $C_{\mathbf{X}}(t_1, t_2) = \sum_i \sum_j (x_{1i} - \eta_{\mathbf{X}}(t_1))(x_{2j} - \eta_{\mathbf{X}}(t_2)) \times P(\mathbf{X}(t_1) = x_{1i}, \mathbf{X}(t_2) = x_{2j})$

5.3.1 Concepto de ruido blanco

Un proceso se denomina ruido blanco si cumple unas determinadas condiciones. Dentro del ruido blanco se distinguen típicamente dos casos, aunque en la práctica el empleado es el primero. Éstos son:

- Ruido blanco en sentido amplio: un proceso se dice que es un proceso de ruido blanco en sentido amplio si presenta incorrelación total, es decir, si cada dos variables del mismo están incorreladas. Por ello, para el caso de procesos estocásticos (esto es, para índices temporales continuos) se puede escribir

$$C_{\mathbf{X}}(t_1, t_2) = q(t_1)\delta(t_1 - t_2) \quad (5.10)$$

si el proceso $\mathbf{X}(t)$ es ruido blanco en sentido amplio. Si lo es la secuencia $\mathbf{X}[n]$ podremos escribir

$$C_{\mathbf{X}}[n_1, n_2] = q[n_1]\delta[n_1 - n_2] \quad (5.11)$$

Repárese que la función δ correspondiente al proceso estocástico es una delta de Dirac, es decir, presenta una discontinuidad en 0, mientras que la segunda es una delta de Kronecker, es decir, es una función nula en todos los puntos de su argumento salvo en el 0, donde es unitaria.

- Ruido blanco en sentido estricto: concepción más restrictiva que la anterior, en la que se exige no sólo incorrelación a las variables del proceso, sino también independencia. Por tanto, si un proceso $\mathbf{X}(t)$ es un ruido blanco en sentido estricto, podremos escribir la función de densidad N -dimensional ($\forall N$) como sigue:

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = \prod_{i=1}^N f_{\mathbf{X}}(x_i; t_i) \quad (5.12)$$

En la práctica, como hemos dicho, el modelo comúnmente empleado es el de ruido blanco en sentido amplio. Asimismo, salvo mención expresa, se asumirá que el ruido blanco tiene media nula.

5.3.2 Caso de dos procesos

Cuando los procesos involucrados sean dos, por ejemplo, $\mathbf{X}(t)$ e $\mathbf{Y}(t)$, la caracterización conjunta de ambos vendría dada por la función de densidad conjunta de órdenes N y M , con estos parámetros arbitrariamente grandes. No obstante, no suele ser posible poder calcular esta función, de forma que el recurso es limitarse a caracterizaciones parciales del tipo medias, varianzas y covarianzas. Las covarianzas, en este caso, involucran a variables de dos procesos y por ello se las denomina covarianzas (o correlaciones, en su caso) *cruzadas*. Así pues, se definen:

$$\begin{aligned} R_{\mathbf{XY}}(t_1, t_2) &= E\{\mathbf{X}(t_1)\mathbf{Y}(t_2)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{\mathbf{XY}}(x, y; t_1, t_2) dx dy \\ C_{\mathbf{XY}}(t_1, t_2) &= E\{(\mathbf{X}(t_1) - \eta_{\mathbf{X}}(t_1))(\mathbf{Y}(t_2) - \eta_{\mathbf{Y}}(t_2))\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \eta_{\mathbf{X}}(t_1))(y - \eta_{\mathbf{Y}}(t_2)) f_{\mathbf{XY}}(x, y; t_1, t_2) dx dy \\ &= R_{\mathbf{XY}}(t_1, t_2) - \eta_{\mathbf{X}}(t_1)\eta_{\mathbf{Y}}(t_2) \end{aligned} \quad (5.13)$$

De manera similar a lo ya conocido, podemos hablar de procesos inco-
rrelados, ortogonales e independiente. Formalmente:

- Incorrelación: dos procesos son incorrelados si $C_{\mathbf{XY}}(t_1, t_2) = 0, \forall t_1, t_2$.
Nótese que esto implica que $R_{\mathbf{XY}}(t_1, t_2) = \eta_{\mathbf{X}}(t_1)\eta_{\mathbf{Y}}(t_2), \forall t_1, t_2$
- Ortogonalidad: dos procesos son ortogonales si $R_{\mathbf{XY}}(t_1, t_2) = 0, \forall t_1, t_2$
- Independencia: dos procesos son independientes si $\forall N$ y $\forall M$ se veri-
fica

$$\begin{aligned} f_{\mathbf{XY}}(x_1, \dots, x_N, y_1, \dots, y_M; t_1, \dots, t_N, t'_1, \dots, t'_M) &= \\ f_{\mathbf{X}}(x_1, \dots, x_N; t_1, \dots, t_N) f_{\mathbf{Y}}(y_1, \dots, y_M; t'_1, \dots, t'_M) \end{aligned} \tag{5.14}$$

Todo lo conocido para el caso de VAs, como es natural, aplica en este momento. En particular, independencia implica incorrelación, y si dos procesos son gaussianos (esto es, si las variables que podamos extraer de éstos son conjuntamente gaussianas) incorrelación implicaría independencia. Piénsese en ello.

5.4 Estacionariedad

Hasta ahora, como se ha observado, apenas se ha hecho uso de la dimensión temporal del proceso estocástico $\mathbf{X}(t)$. Básicamente, no hemos centrado en la visión de un proceso estocástico como una colección de VAs y las relaciones que hemos manejado entre ellas han sido las ya conocidas (correlaciones, covarianzas, etc ...). En este punto, y en los que siguen, explotaremos el hecho de que un proceso estocástico es una función del tiempo.

El concepto de estacionariedad está ligado a las variaciones de las propiedades estadísticas del proceso a lo largo del tiempo. En este caso, también distinguiremos entre estacionariedad en sentido estricto y estacionariedad en sentido amplio. El primer sentido hará referencia a las propiedades que debe cumplir la función de densidad, mientras que el segundo impondrá condiciones sólo sobre la media y la correlación del proceso.

- Sentido estricto³: un proceso estocástico $\mathbf{X}(t)$ es estacionario en sen-
tido estricto si su función de densidad de orden N ($\forall N$) es invariante

³Se denota por SSS, de *strict sense stationary*

a un desplazamiento en el origen de tiempos, esto es, si se verifica que

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = f_{\mathbf{X}}(x_1, x_2, \dots, x_N; t_1 + c, t_2 + c, \dots, t_N + c) \quad (5.15)$$

$\forall c$. Nótese por tanto que un proceso es estacionario en sentido estricto si se verifica que la distribución conjunta de variables igualmente separadas coincide. Dos casos particulares importantes de lo anterior son

$$\begin{aligned} f_{\mathbf{X}}(x; t) &= f_{\mathbf{X}}(x) \\ f_{\mathbf{X}}(x_1, x_2; t_1, t_2) &= f_{\mathbf{X}}(x_1, x_2; t_1 - t_2) \end{aligned} \quad (5.16)$$

esto es, las variables de un proceso estacionario en sentido estricto son marginalmente equidistribuidas, y la función de densidad bidimensional depende de la separación entre las dos variables y no de sus posiciones absolutas en el eje de tiempos.

Si la condición prevista en la ecuación (5.15) se verifica sólo hasta un cierto valor de N entonces el proceso se denomina estacionario de orden N .

Las definiciones anteriores se extienden al caso de dos procesos. Se dice que $\mathbf{X}(t)$ e $\mathbf{Y}(t)$ son conjuntamente estacionarios si

$$\begin{aligned} f_{\mathbf{XY}}(x_1, \dots, x_N, y_1, \dots, y_M; t_1, \dots, t_N, t'_1, \dots, t'_M) &= \\ f_{\mathbf{XY}}(x_1, \dots, x_N, y_1, \dots, y_M; t_1 + c, \dots, t_N + c, t'_1 + c, \dots, t'_M + c) & \end{aligned} \quad (5.17)$$

Nótese que de lo anterior se desprende que si dos procesos son conjuntamente estacionarios en sentido estricto, lo son individualmente. Bastaría calcular las funciones de densidad marginales de cada proceso a partir de la expresión anterior.

Finalmente, el proceso $\mathbf{Z}(t) = \mathbf{X}(t) + j\mathbf{Y}(t)$ es estacionario si sus componentes real e imaginaria son conjuntamente estacionarias.

- Sentido amplio⁴: un proceso $\mathbf{X}(t)$ es estacionario en sentido amplio si de verifican las dos condiciones siguientes:

$$E\{\mathbf{X}(t)\} = \eta_{\mathbf{X}}(t) = \eta_{\mathbf{X}} \quad (5.18)$$

$$R_{\mathbf{X}}(t_1, t_2) = R_{\mathbf{X}}(t_1 - t_2) = R_{\mathbf{X}}(\tau) \quad (5.19)$$

⁴Se denota por WSS, de *wide sense stationary*

es decir, si la media no varía con el tiempo y la autocorrelación es función de la separación temporal de las variables, no de la posición absoluta de cada una en el eje temporal.

Nótese que estas condiciones implican que varianza y VCM deben ser constantes, así como que la covarianza es función sólo de la separación entre las variables.

Para el caso de dos procesos se dice que son conjuntamente estacionarios si lo son marginalmente, y si, además, la correlación cruzada $R_{\mathbf{XY}}(t_1, t_2) = R_{\mathbf{XY}}(t_1 - t_2) = R_{\mathbf{XY}}(\tau)$.

Puede también apreciarse que si el proceso es SSS, o bien estacionario hasta orden 2, también es WSS. No obstante, el recíproco no es cierto en general, salvo en el caso de procesos gaussianos.

Finalmente, y a nivel de convenio, estableceremos la siguiente notación para procesos WSS (para el caso complejo, por generalidad)

$$\begin{aligned} R_{\mathbf{X}}(\tau) &= E\{\mathbf{X}(t + \tau)\mathbf{X}^*(t)\} \\ C_{\mathbf{X}}(\tau) &= E\{(\mathbf{X}(t + \tau) - \eta_{\mathbf{X}})(\mathbf{X}(t) - \eta_{\mathbf{X}})^*\} = R_{\mathbf{X}}(\tau) - |\eta_{\mathbf{X}}|^2 \\ R_{\mathbf{XY}}(\tau) &= E\{\mathbf{X}(t + \tau)\mathbf{Y}^*(t)\} \\ C_{\mathbf{XY}}(\tau) &= E\{(\mathbf{X}(t + \tau) - \eta_{\mathbf{X}})(\mathbf{Y}(t) - \eta_{\mathbf{Y}})^*\} = R_{\mathbf{XY}}(\tau) - \eta_{\mathbf{X}}\eta_{\mathbf{Y}}^* \end{aligned} \quad (5.20)$$

Para el caso de secuencias aleatorias el convenio será el mismo salvo sustituir los índices continuos por índices discretos. Por ejemplo, $R_{\mathbf{X}}[m] = E\{\mathbf{X}(n + m)\mathbf{X}^*[n]\}$.

En lo que sigue, y hasta el final del tema, con el objetivo de no duplicar resultados consideraremos procesos de índices continuos, bien entendido que los resultados a los que llegaremos aplican también a secuencias aleatorias sin más que hacer las oportunas adaptaciones de operadores continuos a operadores discretos (el ejercicio 3 de la sección 5.8 es un ejemplo de manejo de una secuencia aleatoria).

5.4.1 Propiedades de la correlación de procesos estacionarios

Autocorrelación

Si $\mathbf{X}(t)$ es un proceso WSS se verifica:

1. La autocorrelación del proceso en $\tau = 0$ es igual al VCM del proceso, el cual coincide con la potencia del mismo (según indica la ecuación (5.55) de la sección 5.6.3). Formalmente

$$R_{\mathbf{X}}(0) = E\{|\mathbf{X}(t)|^2\} = P_{\mathbf{X}} \quad (5.21)$$

2. $|R_{\mathbf{X}}(\tau)| \leq R_{\mathbf{X}}(0)$. La demostración es sencilla mediante el concepto de densidad espectral de potencia (véase sección 5.6) y el por qué de esta propiedad se puede razonar en términos muy intuitivos haciendo uso del concepto de ergodicidad (véase sección 5.5.3).
3. $R_{\mathbf{X}}(-\tau) = R_{\mathbf{X}}^*(\tau)$. En general la correlación es una función hermítica. Para el caso real, la correlación es una función par.
4. Si $\mathbf{X}(t)$ es periódico, $R_{\mathbf{X}}(\tau)$ también es periódico y del mismo periodo.
5. La función de autocorrelación no puede tener una forma arbitraria. Concretamente, escogiendo N variables del proceso, la matriz de correlación debe ser, al menos, semidefinida positiva. Esto es debido a que

$$E \left\{ \left| \sum_{i=1}^N a_i \mathbf{X}(t_i) \right|^2 \right\} = \sum_{i=1}^N \sum_{j=1}^N a_i a_j^* R_{\mathbf{X}}(t_i, t_j) \geq 0 \quad (5.22)$$

6. Si el proceso $\mathbf{X}(t)$ es un proceso de ruido blanco entonces $C_{\mathbf{X}}(\tau) = q\delta(\tau)$.

Ejercicio: El proceso estocástico $\mathbf{X}(t)$ se define de la forma

$$\mathbf{X}(t) = \sum_{k=1}^K \mathbf{A}_k e^{j(k\omega_o t + \Theta_k)}$$

con \mathbf{A}_k VAs complejas de VCM conocido y Θ_k VAs uniformes en un rango de 2π radianes ($k=\{1, \dots, K\}$). Supóngase que todas las VAs involucradas son independientes entre sí. Se pide que obtenga la media y la autocorrelación del proceso $\mathbf{X}(t)$ e indique si el proceso es estacionario en sentido amplio.

Solución: Para calcular la media, recordando la linealidad del operador esperanza (sección 3.6) así como que independencia

implica incorrelación y lo que ello conlleva en relación a la esperanza del producto (sección 3.6.2), podemos escribir

$$\begin{aligned}
 \eta_{\mathbf{X}}(t) &= E\{\mathbf{X}(t)\} = E\left\{\sum_{k=1}^K \mathbf{A}_k e^{j(k\omega_o t + \Theta_k)}\right\} \\
 &= \sum_{k=1}^K e^{jk\omega_o t} E\left\{\mathbf{A}_k e^{j\Theta_k}\right\} \\
 &= \sum_{k=1}^K e^{jk\omega_o t} E\{\mathbf{A}_k\} E\left\{e^{j\Theta_k}\right\} \\
 &= \sum_{k=1}^K e^{jk\omega_o t} E\{\mathbf{A}_k\} \left(\frac{1}{2\pi} \int_{<2\pi>} (\cos(\theta_k) + j\sin(\theta_k)) d\theta_k\right) \\
 &= 0
 \end{aligned}$$

pues la integral de una sinusoide en un periodo de la misma es nula.

Respecto de la autocorrelación:

$$\begin{aligned}
 R_{\mathbf{X}}(t_1, t_2) &= E\{\mathbf{X}(t_1)\mathbf{X}^*(t_2)\} \\
 &= E\left\{\sum_{p=1}^K \sum_{q=1}^K \mathbf{A}_p \mathbf{A}_q^* e^{j(p\omega_o t_1 + \Theta_p)} e^{-j(q\omega_o t_2 + \Theta_q)}\right\} \\
 &= \sum_{p=1}^K \sum_{q=1}^K e^{jp\omega_o t_1} e^{-jq\omega_o t_2} E\left\{\mathbf{A}_p \mathbf{A}_q^*\right\} E\left\{e^{j\Theta_p} e^{-j\Theta_q}\right\}
 \end{aligned}$$

Centrándonos en la última esperanza repárese que para $p \neq q$ tendríamos la esperanza del producto de funciones de VAs independientes. La independencia implica incorrelación, luego la esperanza del producto sería igual al producto de las esperanzas. La esperanza de cada exponencial compleja, según acabamos de ver, es nula. Por ello, los únicos términos que permanecen en la expresión anterior son los términos de índice $p = q$, para los cuales la exponencial compleja es constante y de valor unidad (el exponente es nulo). Por ello, resulta

$$\begin{aligned}
 R_{\mathbf{X}}(t_1, t_2) &= \sum_{p=1}^K e^{jp\omega_o(t_1 - t_2)} E\left\{|\mathbf{A}_p|^2\right\} \\
 &= \sum_{p=1}^K E\left\{|\mathbf{A}_p|^2\right\} e^{jp\omega_o \tau}
 \end{aligned}$$

$$= R_{\mathbf{X}}(\tau)$$

En virtud del resultado, el proceso es WSS. Definiendo el proceso $\mathbf{X}(t)$ como $\mathbf{X}(t) = \sum_{k=1}^K \mathbf{X}_k(t)$, con $\mathbf{X}_k(t) = \mathbf{A}_k e^{j(k\omega_0 t + \Theta_k)}$, puede comprobarse que la autocorrelación de $\mathbf{X}(t)$ es igual a la suma de las autocorrelaciones de los procesos $\mathbf{X}_k(t)$. Asimismo, repárese en que la autocorrelación es periódica de periodo $T = \frac{2\pi}{\omega_0}$, de la misma manera que cada proceso lo es; finalmente, nótese que también presenta simetría conjugada así como un máximo en $\tau = 0$.



Correlación cruzada

Si los procesos $\mathbf{X}(t)$ e $\mathbf{Y}(t)$ son procesos conjuntamente WSS entonces se verifica que:

1. $R_{\mathbf{XY}}(-\tau) = R_{\mathbf{YX}}^*(\tau)$
2. $|R_{\mathbf{XY}}(\tau)| \leq \sqrt{R_{\mathbf{X}}(0)R_{\mathbf{Y}}(0)}$
3. Si los procesos son ortogonales entonces $R_{\mathbf{XY}}(\tau) = 0 \ \forall \tau$. Esto trae consigo que si $\mathbf{Z}(t) = \mathbf{X}(t) + \mathbf{Y}(t)$ entonces $R_{\mathbf{Z}}(\tau) = R_{\mathbf{X}}(\tau) + R_{\mathbf{Y}}(\tau)$.
4. Si los procesos son incorrelados entonces $C_{\mathbf{XY}}(\tau) = 0 \ \forall \tau$. Esto trae consigo que si $\mathbf{Z}(t) = \mathbf{X}(t) + \mathbf{Y}(t)$ entonces $C_{\mathbf{Z}}(\tau) = C_{\mathbf{X}}(\tau) + C_{\mathbf{Y}}(\tau)$.
5. Si los procesos son independientes, definiendo $\mathbf{Z}(t) = \mathbf{X}(t)\mathbf{Y}(t)$, se verifica que $R_{\mathbf{Z}}(\tau) = R_{\mathbf{X}}(\tau)R_{\mathbf{Y}}(\tau)$.

5.5 Ergodicidad

5.5.1 Ergodicidad con respecto de la media

Según conocemos de los temas anteriores, si observamos el valor que han tomado N variables \mathbf{X}_i idénticamente distribuidas, cada una con valor medio η y varianza σ^2 , entonces si creamos otra variable

$$\mathbf{Z} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \quad (5.23)$$

podemos afirmar que $E\{\mathbf{Z}\} = \eta$ y, si estas variables son, al menos, incorreladas, entonces $\sigma_{\mathbf{Z}}^2 = \sigma^2/N$. Por tanto, conforme N aumenta, la variable \mathbf{Z} tiene cada vez una varianza menor y, en el límite ($N \rightarrow \infty$), \mathbf{Z} pasaría a ser una constante de valor igual a η . En términos prácticos, dado que nunca podremos llegar al límite, diremos que \mathbf{Z} es un buen estimador del valor de la media de cada una de las variables \mathbf{X}_i .

Consideremos ahora el caso de un proceso estocástico WSS $\mathbf{X}(t)$. Por ser estacionario, su media es constante, esto es, no es función del tiempo, tampoco lo es su varianza y la correlación es función de la separación entre las variables escogidas. Pensemos que queremos calcular la media de la variable $\mathbf{X}(t_0)$, con t_0 un valor concreto del tiempo. Para ello, según acabamos de ver, necesitaríamos múltiples realizaciones del proceso $\mathbf{X}(t, a_i)$, desecharíamos todos los valores de las realizaciones en instantes diferentes de t_0 y nos limitaríamos a promediar en el índice i , es decir

$$\mathbf{Z} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}(t_0, a_i) \quad (5.24)$$

Procediendo de esta manera, la variable \mathbf{Z} tendría un comportamiento similar al de la VA de la expresión (5.23): su media coincidiría con la del proceso y su varianza sería tanto más pequeña cuanto mayor fuese el número de realizaciones del proceso, de forma que en el límite la varianza sería nula.

Sin embargo, repárese en que hemos desperdiciado mucha información: nos hacen falta numerosas realizaciones del proceso pero, de cada una de ellas, sólo escogemos un valor (el correspondiente al instante temporal t_0) y renunciamos a hacer uso del resto. No obstante, el proceso es WSS, de forma que la media del resto de las variables coincide con la media de la variable $\mathbf{X}(t_0)$. ¡Y disponemos de una realización de cada una de las infinitas variables de que consta el proceso! Por ello, ¿no sería posible calcular el valor medio a partir de una única realización, sin necesidad de acudir a infinitas realizaciones? Pues bien, si esto es posible, el proceso estocástico $\mathbf{X}(t)$ se dice que es ergódico, en este caso, con respecto a la media.

Veamos qué condiciones son necesarias para que un proceso pueda ser ergódico con respecto a la media. La primera condición es obvia: la media del proceso no puede variar con el tiempo, de forma que el proceso, al menos, ha de ser WSS. Por otra parte, definamos el operador *media temporal* de la forma siguiente:

$$\mathbf{M}_T = \mathbf{M}_T[\mathbf{X}] = \frac{1}{2T} \int_{-T}^T \mathbf{X}(t) dt \quad (5.25)$$

Este operador, como es natural, resulta en una VA puesto que es una función de las infinitas variables del proceso estocástico. ¿Cuáles son la media y la varianza de dicho operador? Al respecto de la media:

$$\begin{aligned} E\{\mathbf{M}_T\} &= E\left\{\frac{1}{2T} \int_{-T}^T \mathbf{X}(t) dt\right\} \\ &= \frac{1}{2T} \int_{-T}^T E\{\mathbf{X}(t)\} dt = \frac{1}{2T} \int_{-T}^T \eta_{\mathbf{X}} dt = \eta_{\mathbf{X}} \end{aligned} \quad (5.26)$$

la cual, como puede verse, coincide con la media del proceso. Al respecto de la varianza el cálculo es un poco más complejo, pero metodológicamente interesante:

$$\begin{aligned} \sigma_{\mathbf{M}_T}^2 &= E\{(\mathbf{M}_T - \eta_{\mathbf{X}})(\mathbf{M}_T - \eta_{\mathbf{X}})^*\} \\ &= E\left\{\frac{1}{2T} \int_{-T}^T (\mathbf{X}(t_1) - \eta_{\mathbf{X}}) dt_1 \frac{1}{2T} \int_{-T}^T (\mathbf{X}(t_2) - \eta_{\mathbf{X}})^* dt_2\right\} \\ &= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T E\{(\mathbf{X}(t_1) - \eta_{\mathbf{X}})(\mathbf{X}(t_2) - \eta_{\mathbf{X}})^*\} dt_1 dt_2 \\ &= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_{\mathbf{X}}(t_1, t_2) dt_1 dt_2 \\ &= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_{\mathbf{X}}(t_1 - t_2) dt_1 dt_2 \end{aligned} \quad (5.27)$$

Para calcular esta integral es conveniente hacer el cambio de variable

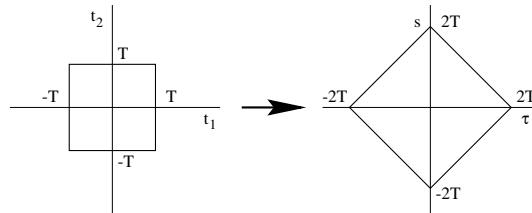


Figura 5.2: Dominios original (t_1, t_2) y transformado (τ, s) de integración.

$$\begin{aligned} s &= t_1 + t_2 \\ \tau &= t_1 - t_2 \end{aligned} \quad (5.28)$$

(la figura 5.2 indica los dominios de integración original y transformado) y dado que el jacobiano de la transformación es

$$\left| \begin{array}{cc} \frac{\partial s}{\partial t_1} & \frac{\partial s}{\partial t_2} \\ \frac{\partial \tau}{\partial t_1} & \frac{\partial \tau}{\partial t_2} \end{array} \right| = \left| \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right| = -2 \quad (5.29)$$

entonces

$$\begin{aligned} dsd\tau &= | - 2|dt_1 dt_2 = 2dt_1 dt_2 \\ dt_1 dt_2 &= \frac{1}{2}dsd\tau \end{aligned} \quad (5.30)$$

de forma que la integral pasa a ser

$$\begin{aligned} \sigma_{\mathbf{M}_T}^2 &= \left(\frac{1}{2T}\right)^2 \int_{-2T}^{2T} d\tau \int_{-(2T-|\tau|)}^{(2T-|\tau|)} C_{\mathbf{X}}(\tau) \frac{1}{2} ds \\ &= \left(\frac{1}{2T}\right)^2 \int_{-2T}^{2T} \frac{1}{2} C_{\mathbf{X}}(\tau) 2(2T - |\tau|) d\tau \\ &= \left(\frac{1}{2T}\right)^2 \int_{-2T}^{2T} C_{\mathbf{X}}(\tau) 2T \left(1 - \frac{|\tau|}{2T}\right) d\tau \\ &= \frac{1}{2T} \int_{-2T}^{2T} C_{\mathbf{X}}(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau \end{aligned} \quad (5.31)$$

Según hemos visto en la ecuación (5.26), la media de la variable *media temporal* \mathbf{M}_T coincide con la media (estadística) $\eta_{\mathbf{X}}$ del proceso. Pues bien, para que el proceso sea ergódico con respecto de la media hace falta que la varianza de esta variable tienda a cero cuando el tiempo de integración tienda a infinito, de forma que, en el límite, la variable media temporal sea igual, sin error ninguno, a $\eta_{\mathbf{X}}$. Formalmente el proceso es ergódico con respecto de la media si

$$\lim_{T \rightarrow \infty} \sigma_{\mathbf{M}_T}^2 = 0 \quad (5.32)$$

es decir, si se verifica que

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-2T}^{2T} C_{\mathbf{X}}(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau = 0 \quad (5.33)$$

5.5.2 Ergodicidad con respecto de la autocorrelación

Podríamos hacer las mismas consideraciones acerca de la ergodicidad de un proceso respecto de la autocorrelación. En este caso, a través de un operador temporal aplicado sobre una única realización del proceso pre-tenderíamos que el resultado de esta operación coincidiese con la autocorrelación $R_{\mathbf{X}}(\tau)$ del proceso. Pues bien, en este caso el operador es la autocorrelación temporal del proceso, definido de la forma:

$$\mathbf{A}_T(\tau) = \mathbf{M}_T[\mathbf{X}(t + \tau)\mathbf{X}^*(t)] = \frac{1}{2T} \int_{-T}^T \mathbf{X}(t + \tau)\mathbf{X}^*(t) dt \quad (5.34)$$

El proceso será ergódico con respecto de la autocorrelación si el límite de la anterior expresión coincide (sin error alguno) con la autocorrelación del proceso para todo τ , es decir, si se verifica que

$$\lim_{T \rightarrow \infty} \mathbf{A}_T(\tau) = R_{\mathbf{X}}(\tau), \forall \tau \quad (5.35)$$

Para que esto sea así se pueden hacer las mismas consideraciones que en el caso de ergodicidad con respecto de la media. Es sencillo comprobar que la esperanza de la ecuación (5.34) coincide con la autocorrelación $R_{\mathbf{X}}(\tau)$ del proceso. Por la tanto, se cumple la primera condición para que el anterior límite se verifique. La segunda condición tiene que ver con la reducción progresiva de la varianza del anterior operador conforme el intervalo de integración crece. Se puede verificar que la varianza de $\mathbf{A}_T(\tau)$ es:

$$\sigma_{\mathbf{A}_{T(\tau)}}^2 = \frac{1}{2T} \int_{-2T}^{2T} C_{\mathbf{Y}_\tau}(\lambda) \left(1 - \frac{|\lambda|}{2T}\right) d\lambda \quad (5.36)$$

con $C_{\mathbf{Y}_\tau}(\lambda)$ es la covarianza del proceso $\mathbf{Y}_\tau(t) = \mathbf{X}(t + \tau)\mathbf{X}^*(t)$ evaluada en el punto λ . Así pues, para que el proceso sea ergódico respecto de la autocorrelación, se debe verificar que

$$\lim_{T \rightarrow \infty} \sigma_{\mathbf{A}_{T(\tau)}}^2 = 0, \forall \tau \quad (5.37)$$

5.5.3 Algunas consecuencias de la ergodicidad de procesos estocásticos

La ergodicidad es una propiedad que nos permite conectar conceptos probabilísticos con ideas intuitivas que podemos obtener haciendo razonamientos temporales. Por ejemplo, imaginemos que un proceso estocástico $\mathbf{X}(t)$ es igual a una VA \mathbf{Y} , es decir

$$\mathbf{X}(t) = \mathbf{Y}, \forall t \quad (5.38)$$

Supongamos, por concreción, que la media de la variable \mathbf{Y} es igual a $\eta_{\mathbf{Y}}$. En estas condiciones es obvio que una realización cualquiera del proceso será una función constante del tiempo de valor igual al del valor que haya tomado la variable \mathbf{Y} en el experimento aleatorio, digamos, valor y . Por tanto, el operador media temporal

$$\mathbf{M}_T = \frac{1}{2T} \int_{-T}^T y dt = y \quad (5.39)$$

de forma que, también en el límite ($T \rightarrow \infty$), la media temporal es igual al valor y de la realización. Como es natural, este valor en general será distinto de la media (salvo que dé la casualidad que coincidan) por lo que el operador media temporal no coincidirá con la media η_Y del proceso, de modo que un proceso así definido no puede ser ergódico.

Un segundo ejemplo ilustrativo es el de un proceso estocástico ergódico de media nula. Si el proceso es ergódico, como ya hemos dicho, la integración temporal del mismo, en el límite, ha de ser igual a la media (estadística) del proceso. Por ello, si la media de éste es nula, la media temporal del proceso también lo ha de ser, de forma que cualquier realización del mismo será una señal con fluctuaciones en torno a la amplitud cero, con *más o menos simetría* de los valores de la señal en torno a dicho valor nulo. Si, por contra, el proceso tuviese media igual a 5, las fluctuaciones de las realizaciones del proceso se producirían sobre un nivel constante de valor 5. La figura 5.3 muestra un ejemplo de cada uno de los dos procesos.

El concepto de ergodicidad también nos permite justificar, de forma intuitiva, por qué la correlación de un proceso WSS ergódico ha de ser máxima en el cero, así como por qué debe ser periódica si el proceso es periódico⁵. Al respecto de la primera, tomemos como referencia la figura 5.4. En esta figura observamos en línea continua una realización del proceso $\mathbf{X}(t)$ y en línea discontinua una realización de este proceso desplazado, es decir, del proceso $\mathbf{X}(t + \tau)$ para $\tau = -1$ (figura inferior) y para $\tau = -4$ (figura superior). Recuérdese que el operador autocorrelación temporal calcula el área encerrada bajo el producto de estas dos funciones. Como es natural, cuando el proceso no está desplazado con respecto a él mismo, el área encerrada bajo el producto punto a punto de las dos funciones temporales (la realización del proceso original, y la realización del proceso desplazado) es máxima, dado que todos los productos tendrán signo positivo, de forma que todos los puntos de la realización contribuyen positivamente. Por contra, cuando desplazamos la señal con respecto a ella misma, los productos punto a punto de una con otra tendrán, dependiendo de cada punto, signos positivos o negativos, de forma que la contribución al área total será constructiva o destructiva según el signo y, por ello, el área total será menor.

Un argumento similar se puede emplear para un proceso periódico. En la figura 5.5 se muestra una realización de un proceso periódico junto a

⁵Debe hacerse una llamada de precaución. La ergodicidad no es condición necesaria para que se cumplan las propiedades de la autocorrelación vistas en la sección 5.4.1. Tales propiedades se cumplen para procesos ergódicos y no ergódicos. Lo que aquí se pretende es proporcionar una visión gráfica de por qué ha de ser así, no una demostración.

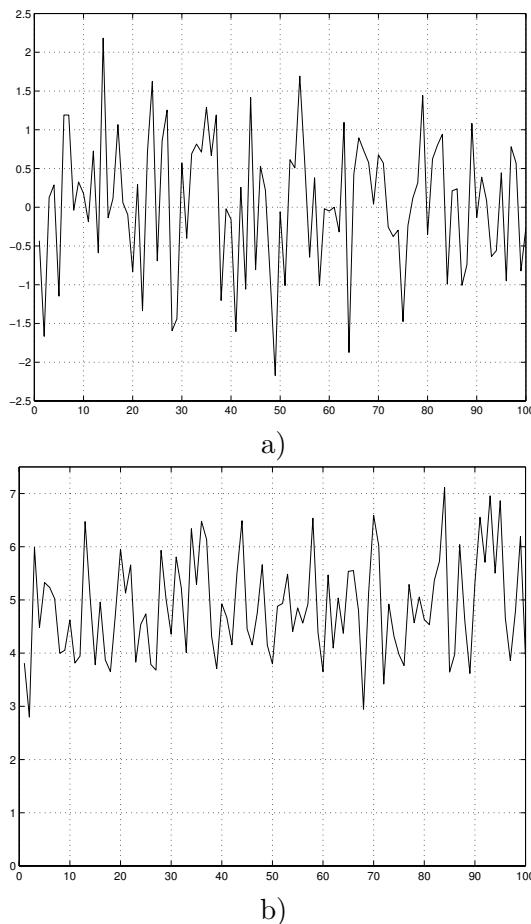


Figura 5.3: a) Realización de un proceso estocástico ergódico de media nula. b) Realización de un proceso estocástico ergódico de media cinco.

otra realización del mismo desplazada 2 unidades temporales (uts). Según se observa, el periodo del proceso es de 10 uts. La realización en línea discontinua ha desplazado su máximo al instante $t = 8$, mientras que el máximo central de la realización en línea continua se encuentra en $t = 10$. Asimismo, en la realización en línea discontinua observamos que un segundo máximo aparece por la derecha de la ventana. Por tanto, si seguimos desplazando el proceso en línea discontinua hasta 10 uts, el máximo que ha entrado por la derecha se habrá desplazado hacia la izquierda hasta encajar perfectamente con el máximo de la realización del proceso en línea continua situado en $t = 10$. En este momento ($\tau = 10$) la autocorrelación

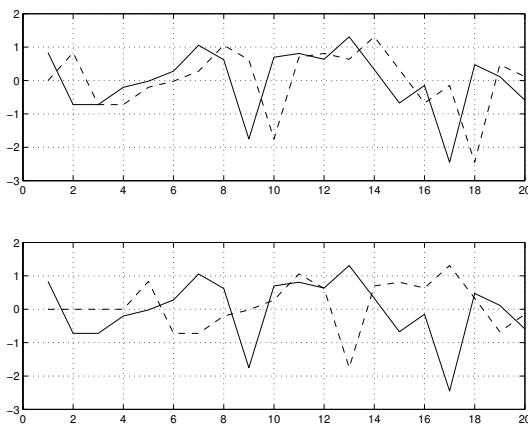


Figura 5.4: Ejemplo de ajuste entre una función y ella misma desplazada en el tiempo. Superior: retraso de $\tau = 1$ unidades temporales (ut). Inferior: retraso de $\tau = 4$ ut.

temporal (y, por ende, la autocorrelación estadística del proceso) volverá a tener el mismo valor que para $\tau = 0$. Así pues, la autocorrelación debe ser periódica, y del mismo periodo que el proceso.

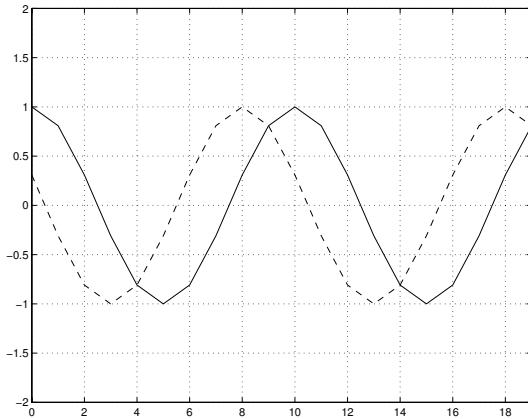


Figura 5.5: Ilustración del desplazamiento temporal de una realización de un proceso periódico.

5.6 Densidad espectral de potencia

5.6.1 Introducción

En este punto vamos a llevar a cabo una caracterización espectral de los procesos estocásticos en el dominio de Fourier, es decir, en el dominio de la frecuencia. Como es sabido [15] el dominio transformado hace que ciertas operaciones, en particular, la operación de filtrado, sean mucho más intuitivas que en el dominio temporal original. El motivo es que una convolución es una operación de cierta complejidad, mientras que su equivalente en el dominio de la frecuencia es un simple producto punto a punto de la transformada de Fourier de la señal con la transformada de Fourier de la respuesta en frecuencia del filtro. Por tanto, conocidas ambas, es evidente ver, por ejemplo, qué extensión espectral tendrá la señal de salida del filtro, si bien la respuesta a esta pregunta en el dominio temporal original no es, en absoluto, trivial.

Con el objetivo de poder hacer uso de esta operativa vamos a relacionar la teoría de los procesos estocásticos con la teoría de los sistemas lineales. Para tal fin, haremos un análisis paralelo a [1].

Una manera inmediata de hacerlo sería llevar a cabo el siguiente razonamiento: dado que los procesos estocásticos son una colección de funciones temporales, podemos limitarnos a calcular las transformadas de Fourier de las infinitas (o finitas, en su caso) realizaciones del mismo y así tendremos la caracterización del proceso en el dominio espectral. Este razonamiento es correcto, sin embargo, es poco práctico. El motivo es obvio: nos obliga a trabajar con una colección de transformadas de Fourier. Sería mucho más útil poder definir *una única función espectral* que caracterizase conjuntamente a *todas* las posibles realizaciones del proceso. En cierta manera, sería algo equivalente a un *espectro promedio* del proceso. Con esta función, a la cual denominaremos densidad espectral de potencia del proceso, podríamos ver, por ejemplo, si el discurso de un locutor podría pasar sin distorsión a través de un determinado sistema lineal (por ejemplo, un amplificador) *con independencia* de lo que el locutor diga en concreto, simplemente, haciendo uso de las características globales de la voz de dicho locutor.

5.6.2 Deducción de la densidad espectral de potencia

Como es sabido, una condición suficiente para que la transformada de Fourier de una señal $x(t)$ exista, es que ésta sea de cuadrado integrable. Sin embargo, las sinusoides, por ejemplo, no tienen cuadrado integrable, pero

sí que disponen de transformada de Fourier. En el caso de los procesos estocásticos, en general, tampoco serán de cuadrado integrable, de forma que tendremos que asegurarnos que, de alguna manera, la integral de Fourier sea convergente. Esto se consigue trabajando en términos de potencia de las señales, en vez de en términos de energía. Para tal fin, consideremos un proceso estocástico $\mathbf{X}(t)$, y definamos, a partir de éste, un proceso estocástico $\mathbf{X}_T(t)$ el cual coincidirá con el primero en una parte del eje temporal y será nulo en el resto. Concretamente:

$$\mathbf{X}_T(t) = \begin{cases} \mathbf{X}(t) & |t| \leq T \\ 0 & |t| > T \end{cases} \quad (5.40)$$

Un proceso así definido será de cuadrado integrable salvo que el proceso $\mathbf{X}(t)$ presente en el intervalo $|t| \leq T$ algún tipo de patología. Dado que asumiremos que esto no será así, partamos de la hipótesis de

$$\int_{-T}^T |\mathbf{X}_T(t)|^2 dt < \infty \quad (5.41)$$

Con esta asunción, podemos afirmar que la transformada de Fourier del proceso $\mathbf{X}_T(t)$ (es decir, la de cada una de las realizaciones de dicho proceso) existe. Además, es cierto que

$$\int_{-T}^T |\mathbf{X}_T(t)|^2 dt = \int_{-T}^T |\mathbf{X}(t)|^2 dt = E_{R=1\Omega} \quad (5.42)$$

dado que en esa ventana temporal los dos procesos coinciden. Asimismo, si aceptamos que los procesos son una magnitud eléctrica (tensión o corriente) la anterior expresión se puede considerar que es igual a la energía disipada por el proceso sobre una resistencia normalizada de valor igual a 1Ω . No se pierda de vista que, naturalmente, dado que el proceso estocástico es una colección de VAs, dicha energía es una función de VA y, por ello, una VA.

En virtud del teorema de Parseval para señales continuas podemos escribir

$$E_{R=1\Omega} = \int_{-T}^T |\mathbf{X}(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathbf{X}_T^f(\omega)|^2 d\omega \quad (5.43)$$

donde $\mathbf{X}_T^f(\omega)$ es la transformada de Fourier del proceso entranado $\mathbf{X}_T(t)$. Nótese que esta transformada de Fourier es también un proceso estocástico, dado que es una función aleatoria de una variable, en este caso, de la pulsación ω .

Podemos pasar de energías a potencias simplemente dividiendo por el tiempo empleado en el cálculo de la energía. Por ello

$$P_{R=1\Omega} = \frac{E_{R=1\Omega}}{2T} = \frac{1}{2T} \int_{-T}^T |\mathbf{X}(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\mathbf{X}_T^f(\omega)|^2}{2T} d\omega \quad (5.44)$$

Extendiendo el intervalo de integración para abarcar todo el proceso ($T \rightarrow \infty$) y tomando valores promedio (es decir, calculando la esperanza de las transformadas de Fourier de cada realización) tendríamos

$$\begin{aligned} \bar{P}_{\mathbf{X}} &= \lim_{T \rightarrow \infty} E\{P_{R=1\Omega}\} \\ &= \lim_{T \rightarrow \infty} E\left\{ \frac{1}{2T} \int_{-T}^T |\mathbf{X}(t)|^2 dt \right\} = \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E\{|\mathbf{X}(t)|^2\} dt = \\ &= \lim_{T \rightarrow \infty} E\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\mathbf{X}_T^f(\omega)|^2}{2T} d\omega \right\} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} E\left\{ \frac{|\mathbf{X}_T^f(\omega)|^2}{2T} \right\} d\omega \end{aligned} \quad (5.45)$$

Resultado de esta operativa es que a la función subintegral de la última expresión, es decir a

$$S_{\mathbf{X}}(\omega) = \lim_{T \rightarrow \infty} E\left\{ \frac{|\mathbf{X}_T^f(\omega)|^2}{2T} \right\}, \quad (5.46)$$

se le denomina *densidad espectral de potencia* del proceso $\mathbf{X}(t)$, ya que es una función que integrada en un eje de frecuencias (de pulsaciones, estrictamente) proporciona la potencia media desarrollada por el proceso (sobre una resistencia normalizada de 1Ω). Mide, por lo tanto, cómo se reparte la potencia media del proceso en cada una de las componentes espectrales que contribuyen a la formación del proceso.

5.6.3 Caso particular de procesos estacionarios en sentido amplio

En el caso particular, aunque prácticamente habitual, de procesos estocásticos estacionarios al menos en sentido amplio, las expresiones anteriores se transforman en otras de más fácil manejo y mucho más intuitivas, pues relacionan la potencia media y la densidad espectral de potencia con parámetros de los procesos que ya conocemos.

Concretamente, si consideramos la expresión (5.45), y nos centramos en la relación entre la potencia media y el proceso en el tiempo, es decir

$$\overline{P}_{\mathbf{X}} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \left\{ |\mathbf{X}(t)|^2 \right\} dt, \quad (5.47)$$

podemos escribir esta expresión haciendo uso del operador media temporal, de forma que

$$\overline{P}_{\mathbf{X}} = \lim_{T \rightarrow \infty} M_T \left[E \left\{ |\mathbf{X}(t)|^2 \right\} \right] \quad (5.48)$$

es decir, estamos calculando la media temporal de la función de VCMs del proceso. En el caso de procesos WSS, el VCM es constante, de forma que, dado que la media temporal de una constante coincide con la constante, podemos escribir

$$\overline{P}_{\mathbf{X}} = E \left\{ |\mathbf{X}(t)|^2 \right\} \quad (5.49)$$

Así pues, para un proceso WSS, el VCM tiene el significado de la potencia media del mismo.

Por otra parte, la expresión general de la densidad espectral de potencia $S_{\mathbf{X}}(\omega)$ se simplifica notablemente, resultando una expresión mucho más intuitiva. Concretamente, llevando a cabo una operativa muy similar a la desarrollada en la sección 5.5, podemos escribir

$$\begin{aligned} \mathbf{X}_T^f(\omega) &= \int_{-T}^T \mathbf{X}(t) e^{-j\omega t} dt \\ E \left\{ |\mathbf{X}_T^f(\omega)|^2 \right\} &= \int_{-T}^T \int_{-T}^T E \left\{ \mathbf{X}(t_1) \mathbf{X}^*(t_2) \right\} e^{-j\omega(t_1-t_2)} dt_1 dt_2 \end{aligned} \quad (5.50)$$

y haciendo el cambio de variable

$$\begin{aligned} s &= t_1 + t_2 \\ \tau &= t_1 - t_2 \end{aligned} \quad (5.51)$$

podemos escribir

$$\begin{aligned} \frac{E \left\{ |\mathbf{X}_T^f(\omega)|^2 \right\}}{2T} &= \frac{1}{2T} \int_{-2T}^{2T} d\tau \int_{-(2T-|\tau|)}^{(2T-|\tau|)} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} \frac{1}{2} ds \\ &= \frac{1}{2T} \int_{-2T}^{2T} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} \frac{1}{2} 2(2T - |\tau|) d\tau \\ &= \frac{1}{2T} \int_{-2T}^{2T} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} 2T \left(1 - \frac{|\tau|}{2T}\right) d\tau \\ &= \int_{-2T}^{2T} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} \left(1 - \frac{|\tau|}{2T}\right) d\tau \end{aligned} \quad (5.52)$$

Si ahora calculamos el límite de la anterior expresión cuando $T \rightarrow \infty$ para dar lugar a la función de densidad de potencia, obtendremos

$$S_{\mathbf{X}}(\omega) = \lim_{T \rightarrow \infty} E \left\{ \frac{|\mathbf{X}_T^f(\omega)|^2}{2T} \right\} = \int_{-\infty}^{\infty} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} d\tau \quad (5.53)$$

o lo que es igual, la transformada de Fourier de la función de correlación del proceso. La relación, por tanto, entre la densidad espectral de potencia de un proceso estocástico WSS y la función de autocorrelación del mismo es de transformadas directa e inversa, es decir

$$\begin{aligned} S_{\mathbf{X}}(\omega) &= \int_{-\infty}^{\infty} R_{\mathbf{X}}(\tau) e^{-j\omega\tau} d\tau \\ R_{\mathbf{X}}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\mathbf{X}}(\omega) e^{j\omega\tau} d\omega \end{aligned} \quad (5.54)$$

Estas expresiones reciben el nombre de relaciones de *Wiener-Khinchin*.

Algunas propiedades de $S_{\mathbf{X}}(\omega)$

La función de densidad de potencia de un proceso verifica las siguientes propiedades:

1. $S_{\mathbf{X}}(\omega)$ es una función real. Nótese, según indica la expresión (5.46), que se calcula mediante una esperanza de un módulo al cuadrado de un número complejo. Por lo tanto es una esperanza de una magnitud real, luego ha de ser real.
2. $S_{\mathbf{X}}(\omega) \geq 0 \forall \omega$. En efecto, según hemos dicho, es la esperanza de una magnitud no negativa (un módulo al cuadrado) luego no puede ser negativa.
3. Si el proceso $\mathbf{X}(t)$ es real y estacionario, $S_{\mathbf{X}}(\omega) = S_{\mathbf{X}}(-\omega)$, es decir, es una función par. Esto es debido a que es igual a la transformada de Fourier de $R_{\mathbf{X}}(\tau)$, función que, como hemos visto, es real y par si el proceso $\mathbf{X}(t)$ es real. Si el proceso $\mathbf{X}(t)$ fuese complejo, $S_{\mathbf{X}}(\omega)$ sería una función hermítica, puesto que también lo sería la correlación $R_{\mathbf{X}}(\tau)$.
4. Habida cuenta de la relación entre el VCM de un proceso estocástico WSS y su función de autocorrelación, se verifica que

$$\overline{P}_{\mathbf{X}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\mathbf{X}}(\omega) d\omega = R_{\mathbf{X}}(0) = E\{\mathbf{X}^2(t)\} \quad (5.55)$$

5.6.4 Espectros cruzados de procesos estacionarios en sentido amplio

La relaciones de Wiener-Khinchin se pueden extender al caso de dos procesos $\mathbf{X}(t)$ e $\mathbf{Y}(t)$ conjuntamente estacionarios en sentido amplio. Podemos, por tanto, definir el par de transformadas

$$\begin{aligned} S_{\mathbf{XY}}(\omega) &= \int_{-\infty}^{\infty} R_{\mathbf{XY}}(\tau) e^{-j\omega\tau} d\tau \\ R_{\mathbf{XY}}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\mathbf{XY}}(\omega) e^{j\omega\tau} d\omega \end{aligned} \quad (5.56)$$

Estos espectros cruzados son, en general, funciones complejas incluso aunque los procesos $\mathbf{X}(t)$ e $\mathbf{Y}(t)$ sean reales. Además, habida cuenta de las propiedades de la correlación cruzada $R_{\mathbf{XY}}(\tau)$ es fácil comprobar que $S_{\mathbf{XY}}(\omega) = S_{\mathbf{YX}}^*(\omega)$.

Por otra parte, los espectros cruzados surgen con frecuencia en la práctica. Por ejemplo, supongamos que necesitamos calcular la densidad espectral de un proceso $\mathbf{Z}(t) = \mathbf{X}(t) + \mathbf{Y}(t)$. Entonces, calculando su autocorrelación

$$\begin{aligned} R_{\mathbf{Z}}(\tau) &= E\{\mathbf{Z}(t + \tau)\mathbf{Z}^*(t)\} \\ &= E\{(\mathbf{X}(t + \tau) + \mathbf{Y}(t + \tau))(\mathbf{X}(t) + \mathbf{Y}(t))^*\} \\ &= E\{\mathbf{X}(t + \tau)\mathbf{X}^*(t)\} + E\{\mathbf{X}(t + \tau)\mathbf{Y}^*(t)\} + \\ &\quad E\{\mathbf{Y}(t + \tau)\mathbf{X}^*(t)\} + E\{\mathbf{Y}(t + \tau)\mathbf{Y}^*(t)\} \\ &= R_{\mathbf{X}}(\tau) + R_{\mathbf{Y}}(\tau) + R_{\mathbf{XY}}(\tau) + R_{\mathbf{YX}}(\tau) \end{aligned} \quad (5.57)$$

por lo que, tomando transformadas de Fourier, obtenemos

$$S_{\mathbf{Z}}(\omega) = S_{\mathbf{X}}(\omega) + S_{\mathbf{Y}}(\omega) + S_{\mathbf{XY}}(\omega) + S_{\mathbf{YX}}(\omega) \quad (5.58)$$

densidad espectral que, como se observa, es función de dos densidades espetrales y dos espectros cruzados. No obstante, para el caso particular en el que los procesos sean ortogonales se verifica que $R_{\mathbf{XY}}(\tau) = R_{\mathbf{YX}}(\tau) = 0 \forall \tau$, de forma que los espectros cruzados serían nulos $\forall \omega$ y la densidad espectral de la suma de dos procesos coincidiría con la suma de las densidades espetrales. Una vez más, la ortogonalidad de procesos hace que ciertas operaciones sean mucho más sencillas.

5.7 Sistemas lineales con entradas estocásticas

Consideremos que el proceso $\mathbf{X}(t)$ atraviesa un sistema lineal e invariante con respuesta al impulso $h(t)$. Denotemos al proceso a la salida del mismo

como $\mathbf{Y}(t)$. El objetivo de esta sección es obtener las principales propiedades del proceso de salida en función de las propiedades del proceso de entrada.

Como es sabido [15], la relación entrada-salida se obtiene mediante la operación de convolución:

$$\mathbf{Y}(t) = \int_{-\infty}^{\infty} \mathbf{X}(t - \tau) h(\tau) d\tau$$

5.7.1 Media y valor cuadrático medio

Con respecto a la media:

$$\begin{aligned} E\{\mathbf{Y}(t)\} &= E\left\{\int_{-\infty}^{\infty} \mathbf{X}(t - \tau) h(\tau) d\tau\right\} \\ &= \int_{-\infty}^{\infty} h(\tau) E\{\mathbf{X}(t - \tau)\} d\tau = \int_{-\infty}^{\infty} h(\tau) \eta_{\mathbf{X}}(t - \tau) d\tau \end{aligned} \quad (5.59)$$

y si el proceso es WSS

$$E\{\mathbf{Y}(t)\} = \eta_{\mathbf{X}} \int_{-\infty}^{\infty} h(\tau) d\tau = \eta_{\mathbf{X}} H(\omega = 0) \quad (5.60)$$

donde $H(\omega = 0)$ es la transformada de Fourier de la respuesta al impulso del filtro evaluada en el cero.

Con respecto al VCM:

$$\begin{aligned} E\{|\mathbf{Y}(t)|^2\} &= E\left\{\int_{-\infty}^{\infty} h(\tau_1) \mathbf{X}(t - \tau_1) d\tau_1 \int_{-\infty}^{\infty} h^*(\tau_2) \mathbf{X}^*(t - \tau_2) d\tau_2\right\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{\mathbf{X}(t - \tau_1) \mathbf{X}^*(t - \tau_2)\} h(\tau_1) h^*(\tau_2) d\tau_1 d\tau_2 \end{aligned} \quad (5.61)$$

En el caso en que $\mathbf{X}(t)$ sea real y WSS

$$E\{|\mathbf{Y}(t)|^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{\mathbf{X}}(\tau_2 - \tau_1) h(\tau_1) h^*(\tau_2) d\tau_1 d\tau_2 \quad (5.62)$$

Un caso particular de interés consiste en que el proceso de entrada $\mathbf{X}(t)$ sea un ruido blanco de media nula y función de autocorrelación $R_{\mathbf{X}}(\tau) = \frac{N_0}{2}\delta(\tau)$. En estas condiciones

$$E\{\mathbf{Y}(t)\} = \eta_{\mathbf{X}} H(\omega = 0) = 0 \quad (5.63)$$

$$\begin{aligned} E\{|\mathbf{Y}(t)|^2\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{N_0}{2} \delta(\tau_2 - \tau_1) h(\tau_1) h^*(\tau_2) d\tau_1 d\tau_2 \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} |h(\tau)|^2 d\tau \end{aligned} \quad (5.64)$$

5.7.2 Correlación cruzada y autocorrelación

Calculemos ahora la autocorrelación $R_{\mathbf{Y}}(t_1, t_2)$ del proceso $\mathbf{Y}(t)$ como función de $R_{\mathbf{X}}(t_1, t_2)$. Al respecto del proceso $\mathbf{X}(t)$, consideraremos que es WSS.

Para calcular la autocorrelación del proceso $\mathbf{Y}(t)$ emplearemos como paso intermedio la correlación cruzada $R_{\mathbf{XY}}(t_1, t_2)$:

$$\begin{aligned} R_{\mathbf{XY}}(t_1, t_2) &= E\{\mathbf{X}(t_1)\mathbf{Y}^*(t_2)\} = E\left\{\mathbf{X}(t_1) \int_{-\infty}^{\infty} \mathbf{X}^*(t_2 - \alpha) h^*(\alpha) d\alpha\right\} \\ &= \int_{-\infty}^{\infty} h^*(\alpha) E\{\mathbf{X}(t_1)\mathbf{X}^*(t_2 - \alpha)\} d\alpha \\ &= \int_{-\infty}^{\infty} h^*(\alpha) R_{\mathbf{X}}(t_1 - t_2 + \alpha) d\alpha = \int_{-\infty}^{\infty} h^*(\alpha) R_{\mathbf{X}}(\tau + \alpha) d\alpha \\ &= R_{\mathbf{XY}}(\tau) \end{aligned} \quad (5.65)$$

es decir, que la correlación cruzada $R_{\mathbf{XY}}(t_1, t_2)$ es función de la diferencia entre los índices $\tau = t_1 - t_2$, y se obtiene mediante la siguiente operación de convolución:

$$R_{\mathbf{XY}}(\tau) = R_{\mathbf{X}}(\tau) * h^*(-\tau) \quad (5.66)$$

Con esto, la función de autocorrelación $R_{\mathbf{Y}}(t_1, t_2)$ se puede calcular de la forma

$$\begin{aligned} R_{\mathbf{Y}}(t_1, t_2) &= E\{\mathbf{Y}(t_1)\mathbf{Y}^*(t_2)\} = E\left\{\left[\int_{-\infty}^{\infty} h(\alpha) \mathbf{X}(t_1 - \alpha) d\alpha\right] \mathbf{Y}^*(t_2)\right\} \\ &= \int_{-\infty}^{\infty} h(\alpha) E\{\mathbf{X}(t_1 - \alpha)\mathbf{Y}^*(t_2)\} d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha) R_{\mathbf{XY}}(t_1 - t_2 - \alpha) d\alpha = \int_{-\infty}^{\infty} h(\alpha) R_{\mathbf{XY}}(\tau - \alpha) d\alpha \\ &= R_{\mathbf{Y}}(\tau) = h(\tau) * R_{\mathbf{XY}}(\tau) \end{aligned} \quad (5.67)$$

es decir,

$$\begin{aligned} R_{\mathbf{Y}}(\tau) &= h(\tau) * R_{\mathbf{XY}}(\tau) \\ &= h(\tau) * h^*(-\tau) * R_{\mathbf{X}}(\tau) \end{aligned} \quad (5.68)$$

5.7.3 Densidad espectral de potencia

Consecuencia inmediata de lo anterior es que

$$\begin{aligned} S_{\mathbf{XY}}(\omega) &= \mathcal{F}\{R_{\mathbf{XY}}(\tau)\} = \mathcal{F}\{R_{\mathbf{X}}(\tau) * h^*(-\tau)\} \\ &= S_{\mathbf{X}}(\omega)H^*(\omega) \end{aligned} \quad (5.69)$$

así como que

$$S_{\mathbf{Y}}(\omega) = S_{\mathbf{X}}(\omega)H^*(\omega)H(\omega) = S_{\mathbf{X}}(\omega)|H(\omega)|^2 \quad (5.70)$$

5.8 Algunos ejercicios ilustrativos

1.- Sea $\mathbf{X}(t)$ un proceso estocástico definido como sigue:

$$\mathbf{X}(t) = A \cos(2\pi \mathbf{F}(n)t)$$

con $nT < t \leq (n+1)T$, $n \in \mathbb{Z}$, $T > 0$ y $\mathbf{F}[n]$ una secuencia de variables independientes de Bernoulli, las cuales toman los valores $f_i = \frac{i}{T}$, $i = 1, 2$, con igual probabilidad $\forall n \in \mathbb{Z}$. Se pide:

- a) Obtenga $\eta_{\mathbf{X}}(t)$ y $\sigma_{\mathbf{X}}^2(t)$. ¿Es el proceso WSS?
- b) Se define el proceso

$$\mathbf{Z}[n] = \int_{nT}^{(n+1)T} \mathbf{X}(t) \cos(2\pi f_1 t) dt$$

Obtenga la caracterización probabilística de primer orden de este proceso.

- c) Se define el proceso

$$\mathbf{W}[n] = \int_{nT}^{(n+1)T} \mathbf{X}(t) \cos(2\pi f_2 t) dt$$

Se pide que justifique si los procesos $\mathbf{Z}[n]$ y $\mathbf{W}[n]$ son ortogonales.

Solución:

- a) El proceso $\mathbf{X}(t)$ consiste en una colección de VAs discretas, pues para un t dado el proceso puede tomar los valores $A \cos(2\pi f_1 t)$ ó $A \cos(2\pi f_2 t)$ de forma equiprobable. Por ello, para el cálculo de los

parámetros que se piden emplearemos las expresiones (2.41) y (2.50). Así pues

$$\begin{aligned}\eta_{\mathbf{X}}(t) = E\{\mathbf{X}(t)\} &= \sum_{i=1}^2 x_i p_i \\ &= \frac{A}{2} [\cos(2\pi f_1 t) + \cos(2\pi f_2 t)] \\ &= \frac{A}{2} \left[\cos\left(2\pi \frac{1}{T} t\right) + \cos\left(2\pi \frac{2}{T} t\right) \right]\end{aligned}$$

Si particularizamos para dos valores temporales concretos

$$\begin{aligned}\eta_{\mathbf{X}}(t=0) &= \frac{A}{2} [1+1] = A \\ \eta_{\mathbf{X}}\left(t=\frac{T}{4}\right) &= \frac{A}{2} \left[\cos\left(\frac{\pi}{2}\right) + \cos(\pi) \right] = -\frac{A}{2}\end{aligned}$$

de forma que es evidente que $\eta_{\mathbf{X}}(t) \neq \eta_{\mathbf{X}}$, es decir, la media es función del tiempo, por lo que el proceso no puede ser WSS.

Para calcular la varianza, calculemos primero el VCM

$$\begin{aligned}E\{\mathbf{X}^2(t)\} &= \sum_{i=1}^2 x_i^2 P(X(t)=x_i) = \sum_{i=1}^2 x_i^2 p_i \\ &= \frac{A^2}{2} \left[\cos^2\left(2\pi \frac{1}{T} t\right) + \cos^2\left(2\pi \frac{2}{T} t\right) \right]\end{aligned}$$

para finalmente escribir

$$\begin{aligned}\sigma_{\mathbf{X}}^2(t) &= E\{\mathbf{X}^2(t)\} - \eta_{\mathbf{X}}^2(t) \\ &= \frac{A^2}{2} \left[\cos^2\left(2\pi \frac{1}{T} t\right) + \cos\left(2\pi \frac{2}{T} t\right) \right] - \\ &\quad \frac{A^2}{4} \left[\cos\left(2\pi \frac{1}{T} t\right) + \cos^2\left(2\pi \frac{2}{T} t\right) \right]^2 \\ &= \frac{A^2}{4} \left[\cos^2\left(2\pi \frac{1}{T} t\right) + \cos^2\left(2\pi \frac{2}{T} t\right) - 2\cos\left(2\pi \frac{1}{T} t\right) \cos\left(2\pi \frac{2}{T} t\right) \right]\end{aligned}$$

b) $\mathbf{Z}[n]$ es un proceso de VAs discretas dado que cada VA del proceso procede de una función de cada variable del proceso $\mathbf{F}[n]$. Por ello, su caracterización de primer orden consiste en hallar los valores que toma cada variable, así como las probabilidades con las que los toman (sección

2.2.2). Así pues si $\mathbf{F}[n] = f_1$ entonces

$$\begin{aligned}
 \mathbf{Z}[n] &= \int_{nT}^{(n+1)T} A \cos(2\pi f_1 t) \cos(2\pi f_1 t) dt \\
 &= A \int_{nT}^{(n+1)T} \cos^2(2\pi f_1 t) dt \\
 &= A \int_{nT}^{(n+1)T} \frac{1 + \cos(2\pi(2f_1)t)}{2} dt \\
 &= \frac{A}{2} T + \frac{A}{2} \int_{nT}^{(n+1)T} \cos\left(\frac{4\pi}{T}t\right) dt = \frac{A}{2} T + \frac{A}{2} \frac{T}{4\pi} \sin\left(\frac{4\pi}{T}t\right) \Big|_{nT}^{(n+1)T} \\
 &= \frac{A}{2} T
 \end{aligned}$$

Por otra parte si $\mathbf{F}[n] = f_2$

$$\begin{aligned}
 \mathbf{Z}[n] &= \int_{nT}^{(n+1)T} A \cos(2\pi f_2 t) \cos(2\pi f_1 t) dt \\
 &= \frac{A}{2} \int_{nT}^{(n+1)T} [\cos(2\pi(f_2 + f_1)t) + \cos(2\pi(f_2 - f_1)t)] dt \\
 &= \frac{A}{2} \int_{nT}^{(n+1)T} \left[\cos\left(2\pi \frac{3}{T}t\right) + \cos\left(2\pi \frac{1}{T}t\right) \right] dt \\
 &= \frac{A}{2} \left(\frac{T}{6\pi} \sin\left(\frac{6\pi}{T}t\right) \Big|_{nT}^{(n+1)T} + \frac{T}{2\pi} \sin\left(\frac{2\pi}{T}t\right) \Big|_{nT}^{(n+1)T} \right) \\
 &= 0
 \end{aligned}$$

Así pues $\mathbf{Z}[n]$ es un proceso de VAs discretas que toman los valores $\frac{A}{2}T$ y 0 de forma equiprobable.

c) Es evidente, a partir del procedimiento seguido en el apartado anterior, que el proceso $\mathbf{W}[n]$ es también un proceso de VAs discretas que toman los mismos valores que $\mathbf{Z}[n]$ y con las mismas probabilidades. Sin embargo, debido a la definición de estos procesos, cuando $\mathbf{Z}[n] = 0$ entonces $\mathbf{W}[n] = \frac{A}{2}T$ y viceversa. Por ello la VA bidimensional $(\mathbf{Z}[n], \mathbf{W}[n])$ toma los valores $(0, \frac{A}{2}T)$ y $(\frac{A}{2}T, 0)$ de forma equiprobable. Recordando la definición de correlación entre dos VAs (sección 3.6.2) podemos escribir

$$\begin{aligned}
 R_{\mathbf{ZW}}[n, n] &= E\{\mathbf{Z}[n]\mathbf{W}[n]\} = \sum_i z_i w_i P(\mathbf{Z}[n] = z_i, \mathbf{W}[n] = w_i) \\
 &= \frac{1}{2} \left(0 \cdot \frac{A}{2}T + \frac{A}{2}T \cdot 0 \right) = 0
 \end{aligned}$$

Por ello las VAs $\mathbf{Z}[n]$ y $\mathbf{W}[n]$ son ortogonales.

Sin embargo si escogemos dos índices distintos $m \neq n$ las VAs $\mathbf{Z}[m]$ y $\mathbf{W}[n]$ proceden de funciones de las VAs $\mathbf{F}[m]$ y $\mathbf{F}[n]$, las cuales son independientes y, por ello, también lo son $\mathbf{Z}[m]$ y $\mathbf{W}[n]$. Al ser independientes son incorreladas (sección 3.6.2) por lo que

$$\begin{aligned} R_{ZW}[m, n] &= E\{\mathbf{Z}[m]\mathbf{W}[n]\} = E\{\mathbf{Z}[m]\}E\{\mathbf{W}[n]\} = \left(\frac{AT}{2}\frac{1}{2} + 0\frac{1}{2}\right)^2 \\ &= \left(\frac{AT}{4}\right)^2 \neq 0 \end{aligned}$$

Por ello los procesos no son ortogonales.

Este problema describe un tipo de modulación digital conocido [13] como FSK (frequency shift keying), en la cual cada dígito binario se envía como un pulso sinusoidal de frecuencia f_i ($i=\{1,2\}$). Los procesos $\mathbf{Z}[m]$ y $\mathbf{W}[n]$ toman valores no nulos cuando se reciben los símbolos asociados a las frecuencias f_1 y f_2 , en los intervalos $(mT, (m+1)T)$ y $(nT, (n+1)T)$, respectivamente.



2.- La cabecera de una instalación de TV recibe la superposición de tres procesos estocásticos $\mathbf{X}(t)$, $\mathbf{Y}(t)$ y $\mathbf{N}(t)$. Los dos primeros soportan, respectivamente, la información de intensidad y color, mientras que el tercero es un proceso de ruido blanco, de media nula, estacionario en sentido amplio con función de autocorrelación $R_N(\tau) = \frac{N_0}{2}\delta(\tau)$, ergódico e independiente de los primeros, el cual se ha introducido durante la transmisión de la señal. Los procesos $\mathbf{X}(t)$ e $\mathbf{Y}(t)$ responden, respectivamente, a las expresiones $\mathbf{X}(t) = \sum_{i=1}^N \mathbf{A}_i \cos(\omega_i^x t + \Theta_i)$, $\mathbf{Y}(t) = \sum_{j=1}^M \mathbf{B}_j \cos(\omega_j^y t + \Phi_j)$, siendo \mathbf{A}_i exponenciales de parámetro $\lambda \forall i$, \mathbf{B}_j exponenciales de parámetro $\mu \forall j$ y las fases iniciales Θ_i y Φ_j uniformes $[0, 2\pi]$, $\forall i, j$. Considere, asimismo, independencia entre todas las variables $\mathbf{A}_i, \mathbf{B}_j, \Theta_p, \Phi_q, \forall i, j, p, q$. Se pide:

a) Denominando $\mathbf{Z}(t)$ al proceso suma de los tres arriba indicados, obtener $E\{\mathbf{Z}(t)\}$ y $R_Z(t_1, t_2)$. ¿Es el proceso $\mathbf{Z}(t)$ estacionario en sentido amplio?

b) ¿Es el proceso $\mathbf{Z}(t)$ ergódico respecto de la media? Sin necesidad de hacer las operaciones, razoné la ergodicidad respecto de la autocorrelación del proceso $\mathbf{X}(t) + \mathbf{Y}(t)$.

c) El proceso $\mathbf{Z}(t)$ es filtrado mediante un filtro cuya respuesta al impulso es $h(t) = \delta(t) - \delta(t - T_0)$ originando un proceso de salida $\mathbf{W}(t)$. Obtener la densidad espectral del proceso de salida $\mathbf{W}(t)$, sabiendo que $f_i^x = \frac{2i+1}{2}f_0$, $f_j^y = jf_0$, siendo $f_0 = 1/T_0$. Explique brevemente la misión

del filtro $h(t)$.

Solución:

a) Aplicando linealidad del operador esperanza (apartado 3.6) podemos escribir

$$\begin{aligned} E\{\mathbf{Z}(t)\} &= E\{\mathbf{X}(t) + \mathbf{Y}(t) + \mathbf{N}(t)\} \\ &= E\{\mathbf{X}(t)\} + E\{\mathbf{Y}(t)\} + E\{\mathbf{N}(t)\} \end{aligned}$$

Al respecto de la esperanza de $\mathbf{X}(t)$

$$\begin{aligned} E\{\mathbf{X}(t)\} &= E\left\{\sum_{i=1}^N \mathbf{A}_i \cos(\omega_i^x t + \Theta_i)\right\} \\ &= \sum_{i=1}^N E\{\mathbf{A}_i \cos(\omega_i^x t + \Theta_i)\} \\ &= \sum_{i=1}^N E\{\mathbf{A}_i\} E\{\cos(\omega_i^x t + \Theta_i)\} \\ &= \sum_{i=1}^N E\{\mathbf{A}_i\} \frac{1}{2\pi} \int_0^{2\pi} \cos(\omega_i^x t + \theta_i) d\theta_i \end{aligned}$$

donde el paso de la segunda a la tercera línea se ha hecho en base a la independencia de las VAs \mathbf{A}_i y Θ_i (véase apartado 3.6.2). Con ello, la integral resulta ser nula, dado que es la integración de un coseno en un periodo del mismo (recuérdese que se integra en la variable θ_i y ésta varía en el intervalo $[0, 2\pi]$). Por ello, $E\{\mathbf{X}(t)\} = 0$. Similares consideraciones podrían hacerse con el proceso $\mathbf{Y}(t)$; además el enunciado dice que el proceso $\mathbf{N}(t)$ tiene media nula. Ello hace finalmente que $E\{\mathbf{Z}(t)\} = 0$.

Respecto de la autocorrelación:

$$\begin{aligned} R_{\mathbf{Z}}(t_1, t_2) &= E\{\mathbf{Z}(t_1)\mathbf{Z}(t_2)\} \\ &= E\{(\mathbf{X}(t_1) + \mathbf{Y}(t_1) + \mathbf{N}(t_1))(\mathbf{X}(t_2) + \mathbf{Y}(t_2) + \mathbf{N}(t_2))\} \end{aligned}$$

Si se realiza el producto indicado tendremos numerosos factores del tipo $E\{\mathbf{AB}\}$ con \mathbf{A} y \mathbf{B} VAs independientes (las procedentes de cada uno de los procesos). Dado que $E\{\mathbf{AB}\} = E\{\mathbf{A}\}E\{\mathbf{B}\}$ para el caso de VAs independientes, y dado que la media de cada proceso es nula, del producto anterior basta con que nos quedemos con los términos consistentes en productos de variables que proceden del mismo proceso. Por ello

$$R_{\mathbf{Z}}(t_1, t_2) = E\{\mathbf{Z}(t_1)\mathbf{Z}(t_2)\}$$

$$\begin{aligned}
&= E\{\mathbf{X}(t_1)\mathbf{X}(t_2)\} + E\{\mathbf{Y}(t_1)\mathbf{Y}(t_2)\} + E\{\mathbf{N}(t_1)\mathbf{N}(t_2)\} \\
&= R_{\mathbf{X}}(t_1, t_2) + R_{\mathbf{Y}}(t_1, t_2) + R_{\mathbf{N}}(t_1, t_2) \\
&= R_{\mathbf{X}}(t_1, t_2) + R_{\mathbf{Y}}(t_1, t_2) + R_{\mathbf{N}}(t_1 - t_2)
\end{aligned}$$

Debemos calcular los dos primeros sumandos. Comenzando por $R_{\mathbf{X}}(t_1, t_2)$

$$\begin{aligned}
R_{\mathbf{X}}(t_1, t_2) &= \sum_{i=1}^N \sum_{j=1}^N E\{\mathbf{A}_i \mathbf{A}_j \cos(\omega_i^x t_1 + \Theta_i) \cos(\omega_j^x t_2 + \Theta_j)\} \\
&= \sum_{i=1}^N E\{\mathbf{A}_i^2 \cos(\omega_i^x t_1 + \Theta_i) \cos(\omega_i^x t_2 + \Theta_i)\} + \\
&\quad \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N E\{\mathbf{A}_i \mathbf{A}_j \cos(\omega_i^x t_1 + \Theta_i) \cos(\omega_j^x t_2 + \Theta_j)\} \\
&= \sum_{i=1}^N E\{\mathbf{A}_i^2\} E\{\cos(\omega_i^x t_1 + \Theta_i) \cos(\omega_i^x t_2 + \Theta_i)\} + \\
&\quad \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N E\{\mathbf{A}_i\} E\{\mathbf{A}_j\} \times \\
&\quad E\{\cos(\omega_i^x t_1 + \Theta_i)\} E\{\cos(\omega_j^x t_2 + \Theta_j)\} \\
&= \sum_{i=1}^N E\{\mathbf{A}_i^2\} E\{\cos(\omega_i^x t_1 + \Theta_i) \cos(\omega_i^x t_2 + \Theta_i)\} \\
&= \sum_{i=1}^N \frac{E\{\mathbf{A}_i^2\}}{2} E\{\cos(\omega_i^x(t_1 + t_2) + 2\Theta_i) + \cos(\omega_i^x(t_1 - t_2))\} + \\
&= \sum_{i=1}^N \frac{E\{\mathbf{A}_i^2\}}{2} \cos(\omega_i^x(t_1 - t_2)) = \sum_{i=1}^N \frac{2/\lambda^2}{2} \cos(\omega_i^x(t_1 - t_2)) \\
&= \sum_{i=1}^N \frac{1}{\lambda^2} \cos(\omega_i^x(t_1 - t_2)) = \sum_{i=1}^N \frac{1}{\lambda^2} \cos(\omega_i^x \tau)
\end{aligned}$$

con $\tau = t_1 - t_2$. Repárese que se ha hecho uso de la independencia de VAs, así como del VCM de una VA exponencial (apartado 2.5.4).

Respecto del proceso $\mathbf{Y}(t)$ el razonamiento sería similar, y el único cambio procedería de los VCMs de las VAs \mathbf{B}_i . Por ello, podemos concluir que

$$R_{\mathbf{Z}}(t_1, t_2) = \sum_{i=1}^N \frac{1}{\lambda^2} \cos(\omega_i^x \tau) + \sum_{j=1}^N \frac{1}{\mu^2} \cos(\omega_j^y \tau) + \frac{N_0}{2} \delta(\tau)$$

$$= R_{\mathbf{Z}}(\tau)$$

por lo que el proceso es WSS (apartado 5.4).

b) La ergodicidad respecto de la media se puede plantear analizando si el límite del operador media temporal (expresión 5.25) coincide con la esperanza matemática del proceso. Para tal fin

$$\mathbf{M}_T[\mathbf{Z}] = \mathbf{M}_T[\mathbf{X} + \mathbf{Y} + \mathbf{N}] = \mathbf{M}_T[\mathbf{X}] + \mathbf{M}_T[\mathbf{Y}] + \mathbf{M}_T[\mathbf{N}] \quad (5.71)$$

Respecto del proceso $\mathbf{N}(t)$, el enunciado dice que es ergódico. Queda pues comprobar qué sucede con los otros dos. Razonando, como anteriormente, con el proceso $\mathbf{X}(t)$

$$\begin{aligned} \mathbf{M}_T[\mathbf{X}] &= \frac{1}{2T} \int_{-T}^T \mathbf{X}(t) dt \\ &= \frac{1}{2T} \int_{-T}^T \sum_{i=1}^N \mathbf{A}_i \cos(\omega_i^x t + \Theta_i) dt \\ \lim_{T \rightarrow \infty} \mathbf{M}_T[\mathbf{X}] &= \sum_{i=1}^N \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mathbf{A}_i \cos(\omega_i^x t + \Theta_i) dt \end{aligned}$$

Cada una de las integrales corresponde a la integración de una sinusoides en toda la recta real. El valor de esta integral es nulo. Podemos ver esto redefiniendo el origen de tiempos para que el coseno se convierta en un seno. En tal caso, la integral sería la de una función impar en un intervalo simétrico.

No obstante, para evitar dudas, empleemos el procedimiento consistente en el cálculo de la varianza de la VA $\mathbf{M}_T[\mathbf{X}]$ (expresión 5.31):

$$\begin{aligned} \sigma_{\mathbf{M}_T}^2 &= \frac{1}{2T} \int_{-2T}^{2T} C_{\mathbf{X}}(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau \\ &= \frac{1}{2T} \int_{-2T}^{2T} R_{\mathbf{X}}(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau \\ &= \frac{1}{T} \int_0^{2T} \left(\sum_{i=1}^N \frac{1}{\lambda^2} \cos(\omega_i^x \tau) \right) \left(1 - \frac{\tau}{2T}\right) d\tau \\ &= \sum_{i=1}^N \frac{1}{T} \int_0^{2T} \frac{1}{\lambda^2} \cos(\omega_i^x \tau) \left(1 - \frac{\tau}{2T}\right) d\tau \\ &= \sum_{i=1}^N \sigma_i^2 \end{aligned}$$

Centrándonos, por simplicidad, en un único valor de i podemos escribir

$$\begin{aligned}\sigma_i^2 &= \frac{1}{T\lambda^2} \int_0^{2T} \cos(\omega_i^x \tau) d\tau - \frac{1}{T\lambda^2} \int_0^{2T} \cos(\omega_i^x \tau) \frac{\tau}{2T} d\tau \\ &= I_1 + I_2 \\ I_1 &= \frac{1}{T\lambda^2} \frac{1}{\omega_i^x} \left. \sin(\omega_i^x \tau) \right|_0^{2T} = \frac{1}{T\lambda^2 \omega_i^x} \sin(\omega_i^x 2T)\end{aligned}$$

Es evidente que $\lim_{T \rightarrow \infty} I_1 = 0$ pues el seno toma un valor entre 0 y 1.

Respecto de I_2 podemos integrar por partes de la forma:

$$\begin{aligned}I_2 &= \frac{1}{2T^2 \lambda^2} \left(\tau \frac{1}{\omega_i^x} \left. \sin(\omega_i^x \tau) \right|_0^{2T} - \int_0^{2T} \frac{1}{\omega_i^x} \sin(\omega_i^x \tau) d\tau \right) \\ &= \frac{1}{2T^2 \lambda^2} \left(\frac{2T}{\omega_i^x} \sin(\omega_i^x 2T) + \frac{1}{(\omega_i^x)^2} \left. \cos(\omega_i^x \tau) \right|_0^{2T} \right) \\ &= \frac{1}{2T^2 \lambda^2} \left(\frac{2T}{\omega_i^x} \sin(\omega_i^x 2T) + \frac{1}{(\omega_i^x)^2} (\cos(\omega_i^x 2T) - 1) \right)\end{aligned}$$

En este caso el factor T^2 que divide a la expresión hace que el límite sea nuevamente cero. Por ello $\lim_{T \rightarrow \infty} \sigma_i^2 = 0$, de forma que $\lim_{T \rightarrow \infty} \sigma_{\mathbf{M}_T}^2 = 0$ y el proceso es ergódico con respecto de la media.

Lo mismo sucede para el proceso $\mathbf{Y}(t)$ por lo que podemos concluir que el proceso $\mathbf{Z}(t)$ es ergódico respecto de la media.

En relación con la ergodicidad con respecto a la autocorrelación repárese que la autocorrelación temporal (expresión (5.34)) será función de los valores concretos observados de las VAs \mathbf{A}_i y \mathbf{B}_j . Por ello la autocorrelación temporal será una función aleatoria, de forma que no coincidirá, en general, con $R_{\mathbf{Z}}(\tau)$, por lo que el proceso no será ergódico respecto de la autocorrelación.

c) Haciendo uso de la expresión (5.70) y dado que $H(\omega) = 1 - e^{-j\omega T_0}$ [15] entonces

$$\begin{aligned}S_{\mathbf{W}}(\omega) &= |H(\omega)|^2 S_{\mathbf{Z}}(\omega) \\ &= 2(1 - \cos(\omega T_0)) S_{\mathbf{Z}}(\omega)\end{aligned}$$

con

$$\begin{aligned}S_{\mathbf{Z}}(\omega) &= \sum_{i=1}^N \frac{\pi}{\lambda^2} [\delta(\omega - \omega_i^x) + \delta(\omega + \omega_i^x)] + \\ &\quad \sum_{j=1}^N \frac{\pi}{\mu^2} [\delta(\omega - \omega_j^y) + \delta(\omega + \omega_j^y)] + \frac{N_0}{2}\end{aligned}$$

Repárese entonces que:

$$\begin{aligned}|H(\omega)|^2 = 0 \rightarrow \omega T_0 &= 2k\pi \rightarrow 2\pi f T_0 = 2k\pi \rightarrow f_k &= k \frac{1}{T_0} \\|H(\omega)|^2 = 4 \rightarrow \omega T_0 &= (2k+1)\pi \rightarrow 2\pi f T_0 = (2k+1)\pi \rightarrow f_k &= \frac{2k+1}{2T_0} \\&&= \frac{2k+1}{2} f_0\end{aligned}$$

Por ello, habida cuenta de los valores que toman f_i^x y f_j^y es claro que el filtro elimina la componente de color $\mathbf{Y}(t)$ de la señal $\mathbf{Z}(t)$, multiplica por 4 la componente de intensidad $\mathbf{X}(t)$ y colorea el ruido adoptando éste un espectro sinusoidal. Así pues

$$S_{\mathbf{Z}}(\omega) = 4S_{\mathbf{X}}(\omega) + 2(1 - \cos(\omega T_0))S_{\mathbf{N}}(\omega)$$



3.- Considere el proceso estocástico real $\mathbf{X}[n]$, ($n \in \mathcal{Z}$), estacionario en sentido amplio y de media nula, con secuencia de autocorrelación $R_{\mathbf{X}}[m]$ conocida ($R_{\mathbf{X}}[m] < R_{\mathbf{X}}[0] \forall m \neq 0$). Se pide:

a) Identifique el valor de a_{11} que hace que $\hat{\mathbf{X}}[n] = a_{11}\mathbf{X}[n-1]$ sea el estimador lineal de $\mathbf{X}[n]$ de mínimo error cuadrático medio. Obtenga la varianza σ^2 del error en la estimación.

b) Un filtro causal con $H(z) = \frac{1}{1-a_{11}z^{-1}}$ es excitado por un ruido blanco $\mathbf{W}[n]$, real y estacionario, de media nula y secuencia de autocorrelación $R_{\mathbf{W}}[m] = \sigma^2\delta[m]$, originando así un nuevo proceso $\mathbf{Y}[n]$. Obtenga la secuencia de autocorrelación $R_{\mathbf{Y}}[m]$.

c) Para mejorar la calidad de la estimación de $\mathbf{X}[n]$ se propone el estimador

$$\hat{\mathbf{X}}[n] = a_{21}\mathbf{X}[n-1] + a_{22}\mathbf{X}[n-2]$$

Obtenga los valores de los coeficientes a_{11} y a_{22} de modo que $\hat{\mathbf{X}}[n]$ sea el estimador lineal de mínimo error cuadrático medio, y deduzca la relación entre éstos y σ^2 .

Solución:

a) Dado que las medias de las VAs del proceso $\mathbf{X}[n]$ son nulas, la expresión del estimador lineal óptimo es, en efecto, la indicada en el enunciado (pues el valor de la ordenada en el origen b sería nulo según indica la expresión (3.62)). Por otra parte el coeficiente a_{11} se puede obtener directamente del empleo de la expresión (3.61). Asimismo y también debido

a la nulidad de la media, correlación y covarianza coinciden (expresión 3.49). Por ello

$$a_{11} = \frac{C_{\mathbf{X}}[1]}{\sigma_{\mathbf{X}}^2} = \frac{R_{\mathbf{X}}[1]}{R_{\mathbf{X}}[0]}$$

Con respecto a la varianza del error de estimación y según la expresión (3.64) podemos escribir

$$\sigma^2 = \sigma_{\mathbf{X}}^2(1 - \rho_{\mathbf{X}}^2[1])$$

Dado que

$$\rho_{\mathbf{X}}[1] = \frac{R_{\mathbf{X}}[1]}{\sqrt{R_{\mathbf{X}}[0]}\sqrt{R_{\mathbf{X}}[0]}} = \frac{R_{\mathbf{X}}[1]}{R_{\mathbf{X}}[0]} = a_{11}$$

entonces

$$\sigma^2 = R_{\mathbf{X}}[0](1 - a_{11}^2)$$

b) El filtro planteado tiene una respuesta al impulso [15] $h[n] = a^n u[n]$, con $u[n]$ la función escalón. La autocorrelación a la salida del filtro puede calcularse mediante el equivalente discreto de la expresión (5.68) o bien a través del dominio transformado, haciendo uso de la expresión (5.70) adaptada al dominio discreto. Empleando el dominio original, la autocorrelación pedida puede calcularse mediante

$$R_{\mathbf{Y}}[m] = h[m] * h^*[-m] * R_{\mathbf{W}}[m]$$

En este caso, dado que la autocorrelación a la entrada es la correspondiente a un ruido blanco, la expresión anterior se puede escribir

$$R_{\mathbf{Y}}[m] = \sigma^2 (h[m] * h^*[-m]) \quad (5.72)$$

Para calcular esta convolución podemos hacer uso de la propiedad de paridad de la autocorrelación para un proceso estocástico WSS real (apartado 5.4.1); de esta manera, podemos calcular la convolución para valores $m \geq 0$ y luego forzar a que ese resultado sea par. Procediendo del modo indicado:

$$\begin{aligned} h[m] * h^*[-m] &= \sum_{n=0}^{\infty} h[n]h[n-m] \\ &= \sum_{n=m}^{\infty} a^n a^{n-m} = a^{-m} \sum_{n=m}^{\infty} a^{2n}, \end{aligned} \quad (5.73)$$

expresión válida $\forall m \geq 0$. Denominando $a^2 = \tau$ debemos hacer la suma $S(\tau) = \sum_{n=m}^{\infty} \tau^n$ lo cual se obtiene de forma sencilla haciendo

$$\begin{aligned} S(\tau) &= \sum_{n=m}^{\infty} \tau^n = \tau^m + \tau^{m+1} + \dots \\ \tau S(\tau) &= \tau \sum_{n=m}^{\infty} \tau^n = \tau^{m+1} + \tau^{m+2} + \dots \end{aligned}$$

Restando ambas expresiones llegamos a

$$\begin{aligned} (1 - \tau)S(\tau) &= \tau^m \\ S(\tau) &= \frac{\tau^m}{1 - \tau} \end{aligned} \tag{5.74}$$

por lo que uniendo los resultados de (5.74) y (5.73) y sustituyendo en (5.72) llegamos a

$$R_{\mathbf{Y}}[m] = \sigma^2 \frac{a^m}{1 - a^2},$$

lo cual es válido $\forall m \geq 0$. Como dijimos, la correlación debe ser par. Por ello

$$R_{\mathbf{Y}}[m] = \sigma^2 \frac{a^{|m|}}{1 - a^2}, \forall m$$

c) En este caso nos plantean la extensión a dos variables del estimador lineal óptimo, esto es, la particularización de la expresión (4.12) para $N = 2$. Para tal fin, dado que $b = 0$ puesto que el proceso es de media nula, podemos, simplemente, hacer uso del principio de ortogonalidad (véase apartado 3.7.5). Concretamente, denominando $\boldsymbol{\epsilon}[n] = \mathbf{X}[n] - \hat{\mathbf{X}}[n]$ podemos escribir

$$\begin{aligned} E\{\boldsymbol{\epsilon}[n]\mathbf{X}[n-1]\} &= 0 \\ E\{\boldsymbol{\epsilon}[n]\mathbf{X}[n-2]\} &= 0 \end{aligned}$$

Por lo que

$$\begin{aligned} E\{(\mathbf{X}[n] - (a_{21}\mathbf{X}[n-1] + a_{22}\mathbf{X}[n-2]))\mathbf{X}[n-1]\} &= 0 \\ E\{(\mathbf{X}[n] - (a_{21}\mathbf{X}[n-1] + a_{22}\mathbf{X}[n-2]))\mathbf{X}[n-2]\} &= 0 \end{aligned}$$

$$\begin{aligned} a_{21}R_{\mathbf{X}}[0] + a_{22}R_{\mathbf{X}}[1] &= R_{\mathbf{X}}[1] \\ a_{21}R_{\mathbf{X}}[1] + a_{22}R_{\mathbf{X}}[0] &= R_{\mathbf{X}}[2], \end{aligned}$$

sistema de dos ecuaciones con dos incógnitas que podemos resolver, por ejemplo, mediante la regla de Cramer, esto es:

$$\begin{aligned}
 a_{21} &= \frac{\begin{vmatrix} R_{\mathbf{X}}[1] & R_{\mathbf{X}}[1] \\ R_{\mathbf{X}}[2] & R_{\mathbf{X}}[0] \end{vmatrix}}{\begin{vmatrix} R_{\mathbf{X}}[0] & R_{\mathbf{X}}[1] \\ R_{\mathbf{X}}[1] & R_{\mathbf{X}}[0] \end{vmatrix}} = \frac{R_{\mathbf{X}}[1]R_{\mathbf{X}}[0] - R_{\mathbf{X}}[1]R_{\mathbf{X}}[2]}{R_{\mathbf{X}}^2[0] - R_{\mathbf{X}}^2[1]} \\
 a_{22} &= \frac{\begin{vmatrix} R_{\mathbf{X}}[0] & R_{\mathbf{X}}[1] \\ R_{\mathbf{X}}[1] & R_{\mathbf{X}}[2] \end{vmatrix}}{\begin{vmatrix} R_{\mathbf{X}}[0] & R_{\mathbf{X}}[1] \\ R_{\mathbf{X}}[1] & R_{\mathbf{X}}[0] \end{vmatrix}} = \frac{R_{\mathbf{X}}[2]R_{\mathbf{X}}[0] - R_{\mathbf{X}}^2[1]}{R_{\mathbf{X}}^2[0] - R_{\mathbf{X}}^2[1]} \\
 &= \frac{R_{\mathbf{X}}[0] \left(R_{\mathbf{X}}[2] - \frac{R_{\mathbf{X}}^2[1]}{R_{\mathbf{X}}[0]} \right)}{R_{\mathbf{X}}[0] \left(1 - \frac{R_{\mathbf{X}}^2[1]}{R_{\mathbf{X}}^2[0]} \right) R_{\mathbf{X}}[0]} \\
 &= \frac{R_{\mathbf{X}}[2] - a_{11}R_{\mathbf{X}}[1]}{\sigma^2}
 \end{aligned}$$



4.- El proceso estocástico real $\mathbf{M}(t)$ es WSS, de media nula, con densidad espectral $S_{\mathbf{M}}(\omega)$, de la que se sabe que $S_{\mathbf{M}}(\omega) = 0$ para $|\omega| > \omega_0$. A partir de éste se construye el proceso $\mathbf{X}(t)$ de la forma

$$\mathbf{X}(t) = \mathbf{M}(t)\cos(\omega_c t) - \hat{\mathbf{M}}(t)\sin(\omega_c t)$$

donde el proceso $\hat{\mathbf{M}}(t)$ es la *transformada de Hilbert* del proceso $\mathbf{M}(t)$. Esta transformada se obtiene como salida del filtro lineal e invariante cuya respuesta al impulso es $h(t) = \frac{1}{\pi t}$ y, de forma complementaria, su respuesta en frecuencia es $H(\omega) = -j\text{sgn}(\omega)$ con sgn la función signo. Bajo el supuesto de que $\omega_c \gg \omega_0$ se pide que obtenga la densidad espectral de potencia $S_{\mathbf{X}}(\omega)$ del proceso $\mathbf{X}(t)$.

Solución:

Para obtener $S_{\mathbf{X}}(\omega)$ calcularemos, como paso previo, $R_{\mathbf{X}}(t_1, t_2)$ y, de resultar el proceso WSS, nos limitaremos a calcular la transformada de Fourier de la autocorrelación. Repárese que de no resultar WSS tendríamos que emplear la vía indicada en la expresión (5.46). Por otra parte, dado que el proceso $\hat{\mathbf{M}}(t)$ se obtiene a partir de $\mathbf{M}(t)$ mediante un sistema lineal e

invariante, al ser $\mathbf{M}(t)$ WSS, también lo es $\hat{\mathbf{M}}(t)$ y son, asimismo, conjuntamente estacionarios (véase sección 5.7). Así pues

$$\begin{aligned}
 R_{\mathbf{X}}(t_1, t_2) &= E\{\mathbf{X}(t_1)\mathbf{X}(t_2)\} \\
 &= E\left\{\left(\mathbf{M}(t_1)\cos(\omega_c t_1) - \hat{\mathbf{M}}(t_1)\sin(\omega_c t_1)\right) \times \right. \\
 &\quad \left.\left(\mathbf{M}(t_2)\cos(\omega_c t_2) - \hat{\mathbf{M}}(t_2)\sin(\omega_c t_2)\right)\right\} \\
 &= E\{\mathbf{M}(t_1)\mathbf{M}(t_2)\cos(\omega_c t_1)\cos(\omega_c t_2) - \\
 &\quad \mathbf{M}(t_1)\hat{\mathbf{M}}(t_2)\cos(\omega_c t_1)\sin(\omega_c t_2) - \\
 &\quad \hat{\mathbf{M}}(t_1)\mathbf{M}(t_2)\sin(\omega_c t_1)\cos(\omega_c t_2) + \\
 &\quad \left.\hat{\mathbf{M}}(t_1)\hat{\mathbf{M}}(t_2)\sin(\omega_c t_1)\sin(\omega_c t_2)\right\} \\
 &= R_{\mathbf{M}}(t_1 - t_2)\cos(\omega_c t_1)\cos(\omega_c t_2) - \\
 &\quad R_{\mathbf{M}\hat{\mathbf{M}}}^*(t_1 - t_2)\cos(\omega_c t_1)\sin(\omega_c t_2) - \\
 &\quad R_{\hat{\mathbf{M}}\mathbf{M}}^*(t_1 - t_2)\sin(\omega_c t_1)\cos(\omega_c t_2) + \\
 &\quad R_{\hat{\mathbf{M}}\hat{\mathbf{M}}}^*(t_1 - t_2)\sin(\omega_c t_1)\sin(\omega_c t_2)
 \end{aligned}$$

En virtud de las relaciones entrada-salida obtenidas en la sección mencionada y de las propiedades de las autocorrelaciones y correlaciones cruzadas vistas en la sección 5.4.1, podemos escribir

$$\begin{aligned}
 R_{\hat{\mathbf{M}}\hat{\mathbf{M}}}(\tau) &= R_{\mathbf{M}\hat{\mathbf{M}}}(-\tau) \\
 R_{\mathbf{M}\hat{\mathbf{M}}}(\tau) &= R_{\mathbf{M}}(\tau) * h^*(-\tau) = R_{\mathbf{M}}(\tau) * \frac{1}{\pi(-\tau)} = R_{\mathbf{M}}(\tau) * \left(-\frac{1}{\pi\tau}\right) \\
 R_{\hat{\mathbf{M}}\mathbf{M}}(\tau) &= R_{\mathbf{M}}(-\tau) * \frac{1}{\pi\tau} = R_{\mathbf{M}}(\tau) * \frac{1}{\pi\tau} = -R_{\mathbf{M}\hat{\mathbf{M}}}(\tau)
 \end{aligned}$$

Por otra parte recuérdese que la función $sgn(\omega)$ es igual a 1 para $\omega > 0$ y a -1 para $\omega < 0$. Para $\omega = 0$ la función vale cero. Sin embargo, dado que el proceso $\mathbf{M}(t)$ se ha supuesto de media nula, la densidad espectral $S_{\mathbf{M}}(\omega)$ no presenta una delta en el origen (no está presente un término proporcional a $\delta(\omega)$), de forma que el comportamiento de la función $sgn(\omega)$ en $\omega = 0$ es, a todos los efectos, irrelevante. Por ello, podemos escribir

$$\begin{aligned}
 S_{\hat{\mathbf{M}}}(\omega) &= |H(\omega)|^2 S_{\mathbf{M}}(\omega) \\
 &= (-j sgn(\omega)) (j sgn(\omega)) S_{\mathbf{M}}(\omega) \\
 &= S_{\mathbf{M}}(\omega)
 \end{aligned}$$

$\forall \omega \neq 0$. Por ello podemos afirmar que $R_{\hat{\mathbf{M}}}(\tau) = R_{\mathbf{M}}(\tau) \forall \tau$. Volviendo a la expresión de $R_{\mathbf{X}}(t_1, t_2)$, y aceptando que $\tau = t_1 - t_2$, podemos escribir

$$R_{\mathbf{X}}(t_1, t_2) = R_{\mathbf{M}}(\tau)\cos(\omega_c t_1)\cos(\omega_c t_2) -$$

$$\begin{aligned}
& R_{\mathbf{M}\hat{\mathbf{M}}}(\tau) \cos(\omega_c t_1) \sin(\omega_c t_2) + \\
& R_{\mathbf{M}\hat{\mathbf{M}}}(\tau) \sin(\omega_c t_1) \cos(\omega_c t_2) + \\
& R_{\mathbf{M}}(\tau) \sin(\omega_c t_1) \sin(\omega_c t_2) \\
= & R_{\mathbf{M}}(\tau) [\cos(\omega_c t_1) \cos(\omega_c t_2) + \sin(\omega_c t_1) \sin(\omega_c t_2)] + \\
& R_{\mathbf{M}\hat{\mathbf{M}}}(\tau) [\sin(\omega_c t_1) \cos(\omega_c t_2) - \cos(\omega_c t_1) \sin(\omega_c t_2)] \\
= & R_{\mathbf{M}}(\tau) \cos(\omega_c(t_1 - t_2)) + R_{\mathbf{M}\hat{\mathbf{M}}} \sin(\omega_c(t_1 - t_2)) \\
= & R_{\mathbf{M}}(\tau) \cos(\omega_c \tau) + R_{\mathbf{M}\hat{\mathbf{M}}} \sin(\omega_c \tau)
\end{aligned}$$

Finalmente, dado que

$$\begin{aligned}
E\{\mathbf{X}(t)\} &= E\left\{\mathbf{M}(t) \cos(\omega_c t) - \hat{\mathbf{M}}(t) \sin(\omega_c t)\right\} \\
&= E\{\mathbf{M}(t)\} \cos(\omega_c t) - E\{\hat{\mathbf{M}}(t)\} \sin(\omega_c t) = 0
\end{aligned}$$

(recuérdese la expresión (5.60)) podemos afirmar que el proceso $\mathbf{X}(t)$ es WSS.

Si recordamos ahora la expresión (5.69), dado que $H^*(\omega) = j \operatorname{sgn}(\omega)$ entonces

$$S_{\mathbf{M}\hat{\mathbf{M}}}(\omega) = H^*(\omega) S_{\mathbf{M}}(\omega) = j \operatorname{sgn}(\omega) S_{\mathbf{M}}(\omega)$$

Con todo ello la densidad espectral $S_{\mathbf{X}}(\omega)$ se puede expresar

$$\begin{aligned}
S_{\mathbf{X}}(\omega) &= S_{\mathbf{M}}(\omega) * (\pi [\delta(\omega - \omega_c) + \delta(\omega + \omega_c)]) + \\
&\quad S_{\mathbf{M}\hat{\mathbf{M}}}(\omega) * \left(\frac{\pi}{j} [\delta(\omega - \omega_c) - \delta(\omega + \omega_c)] \right) \\
&= \pi [S_{\mathbf{M}}(\omega - \omega_c) + S_{\mathbf{M}}(\omega + \omega_c)] + \\
&\quad \frac{\pi}{j} [S_{\mathbf{M}\hat{\mathbf{M}}}(\omega - \omega_c) - S_{\mathbf{M}\hat{\mathbf{M}}}(\omega + \omega_c)] \\
&= \pi [S_{\mathbf{M}}(\omega - \omega_c) + S_{\mathbf{M}}(\omega + \omega_c)] + \\
&\quad \pi [S_{\mathbf{M}}(\omega - \omega_c) \operatorname{sgn}(\omega - \omega_c) - S_{\mathbf{M}}(\omega + \omega_c) \operatorname{sgn}(\omega + \omega_c)]
\end{aligned}$$

La figura 5.6 muestra, para una forma escogida de $S_{\mathbf{M}}(\omega)$ que cumple la restricción del enunciado, la densidad espectral de potencia $S_{\mathbf{X}}(\omega)$ del proceso $\mathbf{X}(t)$. Como puede verse, se ha desplazado la densidad espectral $S_{\mathbf{M}}(\omega)$ del proceso $\mathbf{M}(t)$ al valor $\omega = \omega_c$ pero sólo se ha conservado la banda lateral superior de tal densidad espectral (esto es, los valores $S_{\mathbf{M}}(\omega)$ para $\omega > 0$). Al ser $\mathbf{M}(t)$ un proceso real su densidad espectral es par, de forma que al conservar sólo la banda lateral superior (o la inferior) no perdemos información de dicha función. El procedimiento descrito en este

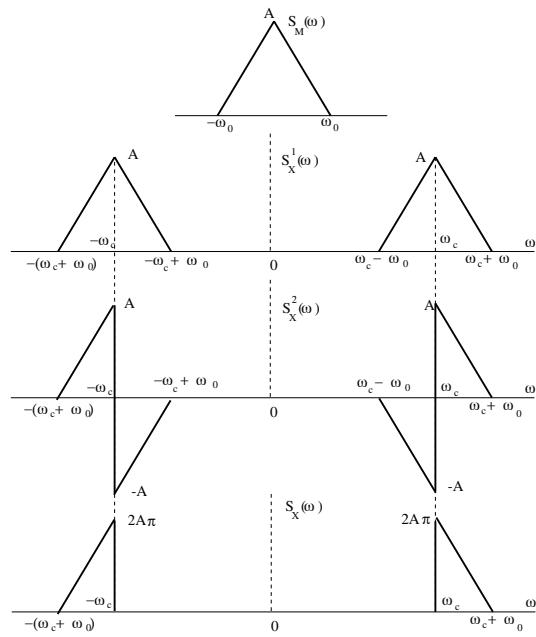


Figura 5.6: Figura problema 4. Se ha denominado $S_X^1(\omega) = S_M(\omega - \omega_c) + S_M(\omega + \omega_c)$ y $S_X^2(\omega) = S_M(\omega - \omega_c)sgn(\omega - \omega_c) - S_M(\omega + \omega_c)sgn(\omega + \omega_c)$.

ejercicio consiste en un tipo de modulación lineal conocida [13] como SSB (single sideband, esto es, modulación de banda lateral única) y permite modular señales analógicas ocupando la mitad de ancho de banda que se ocuparía si se transmitiesen las dos bandas laterales, sin pérdida alguna de información.



Bibliografía

Básica

- [1] P. Z. Peebles, Probability, Random Variables and Random Signal Principles, *Mc-Graw Hill Int. Ed.*, 3rd Ed., 1994.
- [2] A. Papoulis, Probability, Random Variables and Stochastic Processes, *Mc-Graw Hill Int. Ed.*, 3rd Ed., 1993.
- [3] H. Starks, J. W. Woods, Probability, Random Processes and Estimation Theory for Engineers, *Mc-Graw Hill Int. Ed.*, 2nd Ed., 1994.

Avanzada

- [4] S. M. Kay, Fundamentals of Statistical Signal Processing. Estimation Theory, *Prentice-Hall Int. Ed.*, 1st Ed., 1993.
- [5] S. M. Kay, Fundamentals of Statistical Signal Processing. Detection Theory, *Prentice-Hall Int. Ed.*, 1st Ed., 1998.
- [6] H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, *John Wiley and Sons Inc.*, 1st Ed., 1968.
- [7] H. V. Poor, An Introduction to Signal Detection and Estimation, *Springer-Verlag*, 2nd Ed., 1994.
- [8] L. L. Scharf, Statistical Signal Processing, *Addison Wesley*, 1st Ed., 1991.
- [9] R. O. Duda, P. E. Hart, Pattern Analysis and Scene Classification, *John Wiley and Sons Inc.*, 1st Ed., 1973.

- [10] S. Theodoridis, K. Koutroumbas, Pattern Recognition, *Academic Press*, 1st Ed. 1999.
- [11] B. W. Lindgren Statistical Theory, *Collier-Macmillan Int. Ed.*, 4th Ed., 1976.
- [12] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, *John Wiley and Sons Inc.*, 2nd Ed., 1984.

Sistemas de comunicación

- [13] S. Haykin, Communication Systems, *John Wiley and Sons Inc.*, 3rd Ed., 1994.
- [14] B. Sklar, Digital Communications: Fundamentals and Applications, *Prentice Hall Int. Ed.*, 1st Ed., 1988.

Procesado de Señal e Imagen

- [15] A. V. Oppenheim, A. S. Willsky, I. T Young, Signals and Systems, *Prentice Hall Int. Ed.*, 1st Ed., 1983.
 - [16] A. K. Jain, Fundamentals of Digital Image Processing, *Prentice-Hall Int. Ed.*, 1st Ed., 1989.
 - [17] J. S. Lim, Two dimensional Signal and Image Processing, *Prentice-Hall Int. Ed.*, 1st Ed., 1990.
-

Índice General

Prólogo	9
1 Teoría de la Probabilidad	13
1.1 Introducción	13
1.2 Álgebra de conjuntos	15
1.2.1 Definiciones, terminología y notación	15
1.2.2 Operaciones con conjuntos	16
1.3 Definición de probabilidad	19
1.3.1 Probabilidad como frecuencia relativa	20
1.3.2 Definición clásica	20
1.3.3 Definición axiomática de la probabilidad	21
1.4 Probabilidad condicionada	26
1.5 Independencia de sucesos	28
1.5.1 Independencia de pares de sucesos	28
1.5.2 Independencia de múltiple sucesos	31
1.6 Teoremas de la Probabilidad total y de Bayes	32
1.6.1 Concepto de partición	32
1.6.2 Teorema de la Probabilidad Total	32
1.6.3 Teorema de Bayes	33
1.6.4 Sobre el teorema de Bayes y los sistemas de comunicaciones	34
1.7 Experimentos compuestos. Composición de ensayos de Bernoulli	37
1.7.1 Composición de ensayos de Bernoulli	40
2 Variable aleatoria unidimensional	47
2.1 Concepto de variable aleatoria	47
2.2 Caracterización de las variables aleatorias: funciones de distribución y densidad	49

2.2.1	Función de distribución	50
2.2.2	Función de densidad de probabilidad	55
2.3	Algunas variables aleatorias frecuentes	58
2.3.1	Continuas	58
2.3.2	Discretas	65
2.4	Funciones condicionadas	67
2.4.1	Teoremas de la Probabilidad Total y de Bayes . . .	69
2.4.2	Teoremas de la Probabilidad Total y de Bayes mediante partición continua	70
2.5	Caracterización parcial de una variable aleatoria.	73
2.5.1	Media	73
2.5.2	Otras medidas de tendencia central	75
2.5.3	Varianza	76
2.5.4	Ejemplos para varias distribuciones	78
2.5.5	Desigualdad de Tchebycheff	85
2.6	Transformación de variable aleatoria	87
2.6.1	Caso de \mathbf{X} discreta	89
2.6.2	Caso general. Obtención de la función de distribución	90
2.6.3	Caso general. Obtención de la función de densidad: Teorema Fundamental	94
2.7	Caracterización parcial de una función de variable aleatoria	99
2.7.1	Momentos de una variable aleatoria	101
2.8	Funciones que generan momentos	103
2.8.1	Función característica	103
2.8.2	Función generadora de momentos	109
3	Variable aleatoria bidimensional	111
3.1	Introducción	111
3.2	Caracterización de variable aleatoria bidimensional	114
3.2.1	Función de distribución conjunta	114
3.2.2	Función de densidad de probabilidad conjunta . . .	117
3.2.3	Funciones marginales	119
3.2.4	Casos particulares	122
3.3	Funciones condicionadas	127
3.3.1	Funciones marginales, condicionadas y conjuntas .	128
3.3.2	Teoremas de la Probabilidad Total y de Bayes para partición continua	131
3.4	Independencia	132
3.5	Transformación de variable aleatoria bidimensional	135
3.5.1	Una función de dos variables aleatorias	135

3.5.2	Dos funciones de dos variables aleatorias	141
3.5.3	Teorema Fundamental	144
3.5.4	Método de la variable auxiliar	149
3.6	Caracterización parcial de una función de dos variables aleatorias	151
3.6.1	Esperanzas condicionadas	153
3.6.2	Momentos conjuntos	153
3.7	Estimación de mínimo error cuadrático medio	157
3.7.1	Introducción	157
3.7.2	Resumen de estimadores	161
3.7.3	Estimación mediante una constante	161
3.7.4	Estimación mediante una función lineal de la observación	162
3.7.5	Principio de ortogonalidad	164
3.7.6	Estimación óptima sin restricciones	165
4	Variable N-dimensional	167
4.1	Conceptos básicos	167
4.2	Esperanzas matemáticas	171
4.3	Variables conjuntamente gaussianas	174
4.4	Variables complejas	183
4.5	Teoremas asintóticos	184
4.5.1	Teorema del Límite Central	184
4.5.2	Teorema de DeMoivre-Laplace	186
4.5.3	Ley de los Grandes Números	187
4.6	Algunos ejercicios ilustrativos	190
5	Procesos estocásticos	209
5.1	Concepto de proceso estocástico. Clasificación	209
5.1.1	Clasificación de procesos estocásticos	212
5.2	Funciones de distribución y densidad	213
5.3	Caracterización parcial de procesos estocásticos. Ruido blanco	214
5.3.1	Concepto de ruido blanco	216
5.3.2	Caso de dos procesos	217
5.4	Estacionariedad	218
5.4.1	Propiedades de la correlación de procesos estacionarios	220
5.5	Ergodicidad	223
5.5.1	Ergodicidad con respecto de la media	223
5.5.2	Ergodicidad con respecto de la autocorrelación	226

5.5.3	Algunas consecuencias de la ergodicidad de procesos estocásticos	227
5.6	Densidad espectral de potencia	231
5.6.1	Introducción	231
5.6.2	Deducción de la densidad espectral de potencia	231
5.6.3	Caso particular de procesos estacionarios en sentido amplio	233
5.6.4	Espectros cruzados de procesos estacionarios en sentido amplio	236
5.7	Sistemas lineales con entradas estocásticas	236
5.7.1	Media y valor cuadrático medio	237
5.7.2	Correlación cruzada y autocorrelación	238
5.7.3	Densidad espectral de potencia	239
5.8	Algunos ejercicios ilustrativos	239
	Bibliografía	255