



UNIVERSIDAD
DE GRANADA

Teoría de la Comunicación

Autor:
Ignacio Alvarez Illán

Abril 2022



ETSIIT

Escuela Técnica Superior
de Ingenierías Informática
y de Telecomunicación

Variable Aleatoria Unidimensional

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons “Reconocimiento-NoCommercial-CompartirIgual 3.0 España”](#).



Una variable aleatoria X asigna un valor numérico real a la salida de un experimento aleatorio.

Ejemplo 1 El experimento aleatorio constituido por lanzar una moneda al aire y registrar el resultado puede caracterizarse por una variable aleatoria X tal que $X(\oplus) = 1$ y $X(\odot) = 0$, donde \oplus simboliza el resultado 'cara' y \odot simboliza el resultado 'cruz'.

Ejemplo 2 El experimento aleatorio constituido por medir el volumen de un líquido dentro de un termómetro a diferentes temperaturas puede caracterizarse por una variable aleatoria X tal que $X(30^\circ) = 25, 3^\circ C$ constituye un resultado entre infinitos posibles.

Si el conjunto de resultados posibles del experimento aleatorio es discreto hablamos de *variable aleatoria discreta*. Si por el contrario el conjunto de resultados posibles del experimento aleatorio es continuo, la variable aleatoria será una *variable aleatoria continua*.

1. Variable aleatoria discreta

Una variable aleatoria discreta X puede tomar n diferentes valores x_i , con $i = 1, 2, \dots, n$.

Ejemplo 1 La variable X puede tomar $n = 2$ valores, siendo $x_1 = 1$ y $x_2 = 0$.

Si repetimos el experimento aleatorio N veces y la variable aleatoria toma el valor x en m de las repeticiones, diremos que x ha aparecido en una proporción m/N . Si imaginamos que repetimos el experimento infinitas veces, a ese cociente lo llamamos *probabilidad* de x y lo denotamos por $P[X = x]$

1.1. Función masa de probabilidad

Una variable aleatoria discreta quedará completamente caracterizada por la función $p_X(X = x_i)$, es decir, la probabilidad $P[X = x_i]$ de cada una de los resultados de la variable aleatoria en todo el espacio muestral Ω , donde $x_i \in \Omega, i = 1, 2, \dots, n$ simboliza los n posibles valores que puede tomar la variable aleatoria X . A esta función se conoce por el nombre de función masa de probabilidad y se abrevia *pmf* por sus siglas en inglés. Cuando no hay lugar a confusión, la notación se suele simplificar y se omite la referencia explícita a la variable aleatoria X y se hace referencia solo al valor que puede tomar ésta. Así, se denota la pmf como $p_X(x_i)$.

Ejemplo 1 Suponiendo que la moneda no está trucada, sabemos que $P[X = 1] = 0,5$ y $P[X = 0] = 0,5$, donde $\{1, 0\}$ constituye todo el espacio muestral y $p_X = \{0,5, 0,5\}$ es la pmf.

La pmf debe cumplir que $P(\Omega) = 1$, es decir que la probabilidad de que la variable aleatoria tome *alguna* de los valores dentro del espacio muestral es la unidad. Esta condición se suele escribir como:

$$\sum_{i=1}^n p_X(x_i) = 1 \quad (1)$$

1.2. Valores esperados y momentos

Si repetimos un experimento aleatorio N veces, podremos determinar la media aritmética de la variable aleatoria X sumando los valores que aparece el resultado x_i multiplicado por la proporción en la que aparece $\frac{m_i}{N}$:

$$\bar{X} = \sum_{i=1}^n x_i \frac{m_i}{N} \quad (2)$$

El valor esperado (media o esperanza matemática) de una variable aleatoria X es una generalización del concepto de media aritmética para el caso imaginario en el que repitiéramos el experimento infinitas veces. Se denota por $E[X]$ y se define como:

$$\mu_X = E[X] = \sum_{i=1}^n x_i p_X(x_i) \quad (3)$$

El valor esperado es el momento de orden 1 de la variable aleatoria X y es una función lineal, esto es $E[a \cdot X + b] = a \cdot E[X] + b$, donde a y b son constantes.

Ejemplo 1 Suponiendo que repetimos el lanzamiento de la moneda 100 veces y obtenemos 47 caras y 53 cruces, $\bar{X} = (1) \cdot \frac{47}{100} + (0) \cdot \frac{53}{100} = 0,47$ y $E[X] = (1) \cdot 0,5 + (0) \cdot 0,5 = 0,5$

Conociendo $\mu_X = E[X]$ se puede definir una nueva variable aleatoria $Y = X - \mu_X$ cuyo valor esperado es cero, y se dice que la variable Y está centrada.

Se define el momento de orden k de una variable aleatoria X como:

$$\mu_X^k = E[X^k] = \sum_{i=1}^n x_i^k p_X(x_i) \quad (4)$$

Un caso de particular interés es el momento de orden 2 de la variable aleatoria X cuando la centramos. Se conoce con el nombre de varianza:

$$\sigma_X^2 = VAR[X] = E[(X - E[X])^2] = \sum_{i=1}^n (x_i - \mu_X)^2 p_X(x_i) \quad (5)$$

Ejemplo 1 $\sigma_X^2 = (1 - 0,5)^2 \cdot 0,5 + (0 - 0,5)^2 \cdot 0,5 = 0,25$

La varianza mide la dispersión de los datos alrededor de la media, mientras otros momentos de orden superior dan información complementaria: $k = 3$ se denomina skewness y mide la asimetría en la distribución de los datos alrededor de la media, $k = 4$ se denomina kurtosis y mide la presencia de outliers, etc...

2. Variable aleatoria continua

Una variable aleatoria continua X puede tomar infinitos valores $x \in \mathbb{R}$ en el intervalo $x \in (a,b) \subset \mathbb{R}$, donde el intervalo puede ser finito o infinito, abierto o cerrado.

Ejemplo 2 La variable X puede tomar infinitos valores x comprendidos en el rango de temperaturas disponible en el termómetro, que podrían ser $x \in (-20,60)$.

La probabilidad de que X tome un valor concreto es cero¹, así que para el caso continuo hablaremos de la probabilidad de que la variable aleatoria X pertenezca a un subconjunto del espacio muestral A , $P[X \in A]$. Puesto que el espacio muestral en el caso más general será $\Omega \subset \mathbb{R}$, un subconjunto de Ω será un intervalo de \mathbb{R} , y escribiremos $P[a < x < b]$.

2.1. Función densidad de probabilidad

Una variable aleatoria continua quedará completamente caracterizada por $f_X(x)$, es decir, la densidad de probabilidad de la variable aleatoria en todo el espacio muestral Ω , donde $x \in \Omega$ simboliza los infinitos posibles valores que puede tomar la variable aleatoria X . A esta función se conoce por el nombre de función densidad de probabilidad y se abrevia **pdf** por sus siglas en inglés.

Ejemplo 2 Suponiendo que las temperaturas que medimos están distribuidas uniformemente en el intervalo del termómetro, $f_X(x) = 1/80$, con $-20 < x < 60$ donde $\Omega = (-20, 60) \subset \mathbb{R}$ constituye todo el espacio muestral y $f_X(x)$ es la pdf.

La pdf no representa la probabilidad de que ocurra un valor x , ya que ésta es cero para un valor concreto y la pdf no. La pdf representa la posibilidad de que ocurra este valor y el área bajo la curva $f_X(x)$ representa la probabilidad.² Así:

$$P[a < x < b] = \int_a^b f_X(x) dx \quad (6)$$

La pdf debe cumplir que $P(\Omega) = 1$, es decir que la probabilidad de que la variable aleatoria tome *alguna* de los valores dentro del espacio muestral es la unidad. Esta condición se suele escribir como:

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (7)$$

2.2. Valores esperados y momentos

La receta general para pasar del caso discreto al continuo es sustituir variables discretas x_i por variables continuas x , sustituir pmfs por pdfs y:

$$\sum_{i=1}^n \iff \int_{-\infty}^{\infty} dx \quad (8)$$

El valor esperado para el caso continuo se define como:

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (9)$$

El valor esperado es el momento de orden 1 de la variable aleatoria X y es una función lineal, esto es $E[a \cdot X + b] = a \cdot E[X] + b$, donde a y b son constantes.

Ejemplo 2 $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-20}^{60} \frac{x}{80} dx = \frac{x^2}{160} \Big|_{-20}^{60} = 20$

Conociendo $\mu_X = E[X]$ se puede definir una nueva variable aleatoria $Y = X - \mu_X$ cuyo valor esperado es cero, y se dice que la variable Y está centrada. Se define el momento de orden k de una variable aleatoria X como:

¹La extrapolación de las ideas de probabilidad para el caso continuo es sutil, y requiere el desarrollo de herramientas matemáticas adecuadas. Véase el video ilustrativo al respecto: [Por qué 'probabilidad igual a 0' no significa 'imposible'?](#)

²En términos más rigurosos hablaríamos de la integral de Lebesgue con respecto a cierta medida[1]

$$\mu_X^k = E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx \quad (10)$$

Un caso de particular interés es el momento de orden 2 de la variable aleatoria X cuando la centramos. Se conoce con el nombre de varianza:

$$\sigma_X^2 = VAR[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \quad (11)$$

Ejemplo 2 $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx = \int_{-20}^{60} \frac{(x-20)^2}{80} dx = \int_{-40}^{40} \frac{y^2}{80} dy = \frac{y^3}{240} \Big|_{-40}^{40} = 533,3$

En general, el valor esperado de una función g de la variable aleatoria X se obtendrá como:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad (12)$$

3. Ejercicios

1. Muestre que, en el ejemplo 1, si resto 0.5 a los valores que puede tomar X , entonces el valor esperado de esta nueva variable aleatoria es cero.
2. Usando la definición de valor esperado de la ecuación 3, muestre que el valor esperado de una variable aleatoria $Y = X + b$, donde b es una constante, es igual a $E[Y] = E[X] + b$.
3. Usando la propiedad demostrada en el ejercicio anterior, muestre que el valor esperado de una variable aleatoria centrada $Y = X - E[X]$ es cero.
4. Muestre que el valor esperado de una variable aleatoria $Y = a \cdot X$, donde a es una constante, es igual a $E[Y] = a \cdot E[X]$.
5. Usando las propiedades de linealidad demostradas en los ejercicios anteriores, demuestre que $E[(X - E[X])^2] = E[X^2] - (E[X])^2$.

Referencias

- [1] Terence Tao (2010) *An introduction to measure theory*, <http://www.stat.rice.edu/~dobelman/courses/texts/qualify/Measure.Theory.Tao.pdf>.

Variable Aleatoria Multidimensional

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons](#) “Reconocimiento-NoCommercial-CompartirIgual 3.0 España”.



1. Variable aleatoria bidimensional

Una variable aleatoria bidimensional esta formada por un conjunto de dos variables aleatorias X e Y .

Ejemplo 1 El experimento aleatorio constituido por tirar dos dados simultáneamente puede caracterizarse por una variable aleatoria bidimensional $\mathbf{X} = (X, Y)$ tal que X asigna valores numéricos $\{1, 2, 3, 4, 5, 6\}$ a los posibles resultados $\{\square, \square, \square, \square, \square, \square\}$ al igual que Y .

Ejemplo 2 El experimento aleatorio constituido por sacar dos cartas de una baraja puede caracterizarse por una variable aleatoria bidimensional $\mathbf{X} = (X, Y)$ cuyos valores dependerán de si se vuelve a introducir la carta en la baraja tras sacar la primera carta o no.

Ejemplo 3 El experimento aleatorio constituido por medir el volumen de un líquido dentro de dos termómetros a diferentes temperaturas, uno en el interior de un coche y otro en el exterior, puede caracterizarse por una variable aleatoria bidimensional $\mathbf{X} = (X, Y)$ tal que $(X = 25, 3^\circ C, Y = 32, 7^\circ C)$ constituye un resultado entre infinitos posibles.

Si el conjunto de resultados posibles del experimento aleatorio es discreto hablamos de *variable aleatoria discreta*. Si por el contrario el conjunto de resultados posibles del experimento aleatorio es continuo, la variable aleatoria será una *variable aleatoria continua*.

1.1. Variable aleatoria bidimensional discreta

Una variable aleatoria bidimensional discreta $\mathbf{X} = (X, Y)$ puede tomar un conjunto numerable de valores (x_i, y_j) , con $i = 1, 2, \dots, j = 1, 2, \dots$, pudiendo ser numerable finito o infinito.

Ejemplo 1 La variable $\mathbf{X} = (X, Y)$ puede tomar $N = 36$ valores. Un valor posible entre ellos es $\mathbf{X} = (X(\square), Y(\square)) = (5, 3)$.

Si repetimos el experimento aleatorio N veces y la variable aleatoria toma el valor (x, y) en m de las repeticiones, diremos que (x, y) ha aparecido en una proporción m/N . Si imaginamos que repetimos el experimento infinitas veces, a ese cociente lo llamamos *probabilidad* de (x, y) y lo denotamos por $P[X = x, Y = y]$ ¹

¹En algunas ocasiones, cuando se utiliza la teoría de conjuntos para formalizar los conceptos estadísticos, se utiliza la nomenclatura $P[\{X = x\} \cap \{Y = y\}]$.

1.1.1. Función masa de probabilidad conjunta

Una variable aleatoria bidimensional discreta quedará completamente caracterizada por la función $p_{X,Y}(x_i, y_j)$, es decir, la probabilidad $P[X = x_i, Y = y_j]$ de cada una de los resultados de la variable aleatoria en todo el espacio muestral Ω , donde $(x_i, y_j) \in \Omega, i = 1, 2, \dots, j = 1, 2, \dots$ simboliza los posibles valores que puede tomar la variable aleatoria $\mathbf{X} = (X, Y)$. A esta función se conoce por el nombre de función masa de probabilidad conjunta.

Ejemplo 1 El espacio muestral de este ejemplo estará constituido por todas las parejas posibles de valores $\{(1, 1), (1, 2), (1, 3), \dots, (5, 6), (6, 6)\}$, habiendo 36 posibilidades. Suponiendo que los dados no están trucados, sabemos que $P[X = 1, Y = 1] = 1/36$, al igual que para cualquier otro par de valores, ya que cualquier par de valores será igualmente probable. La pmf será $p_{\mathbf{X}}(x_i, y_j) = 1/36 \quad i, j = 1, 2, \dots, 6$.

La pmf debe cumplir que $P(\Omega) = 1$, es decir que la probabilidad de que la variable aleatoria tome *alguno* de los valores dentro del espacio muestral es la unidad. Esta condición se suele escribir como:

$$\sum_i \sum_j p_{X,Y}(x_i, y_j) = 1 \quad (1)$$

1.1.2. Probabilidades Marginales

Cuando tratamos con una variable aleatoria bidimensional $\mathbf{X} = (X, Y)$, cabe preguntarse cuál es la probabilidad de que una de las dos variables aleatorias (por ejemplo Y) tome un valor concreto con independencia de qué le ocurra a la variable X , es decir, $P[Y = y]$. Esta probabilidad se conoce como *probabilidad marginal* de la variable aleatoria Y . Conociendo las probabilidades marginales de todos los valores del espacio muestral que puede tomar Y , se define la función masa de probabilidad marginal p_Y . Conociendo la pmf $p_{X,Y}(x_i, y_j)$, las funciones masa de probabilidad marginales se determinan sumando todos los valores posibles de la pmf en la variable que no es de interés, es decir como:

$$p_Y(y_j) = P[Y = y_j] = P[X = \text{cualquiera}, Y = y_j] = \sum_i p_{X,Y}(x_i, y_j) \quad (2)$$

De la misma forma:

$$p_X(x_i) = \sum_j p_{X,Y}(x_i, y_j) \quad (3)$$

Ejemplo 1 La probabilidad de obtener un 5 en el primero de los dados, con independencia de lo que saque en el otro dado (la probabilidad marginal) se obtiene a través de:

$$P[X = 5] = \sum_{i=1}^6 P[X = 5, Y = y_i] = P[5, 1] + P[5, 2] + P[5, 3] + P[5, 4] + P[5, 5] + P[5, 6] = 1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36 = 1/6.$$

Conocida la función masa de probabilidad conjunta se pueden determinar las marginales pero, en general, a partir de las marginales no es posible determinar la pmf conjunta.

1.1.3. Probabilidad condicional

A menudo será de gran interés determinar cuál es la probabilidad de obtener un valor en una de las variables, por ejemplo Y , una vez sabemos que la otra variable tiene un valor fijo x_i . A esta probabilidad se le denomina probabilidad condicional, se denota por $P[Y|X = x_i]$, leído como probabilidad de Y dado X .

La probabilidad condicional se comporta como una probabilidad, en el sentido de que cumple que, fijada una condición, la suma de todas las probabilidades condicionales es uno ($P[\Omega|X] = 1$) y $0 < P[Y|X] < 1$.

Ejemplo 2 Fijada la condición de que la primera carta que se saca Y es un as $Y = \diamond$, quedan las dos posibilidades: que la siguiente carta sacada X sea un as $X = \clubsuit$ o que no lo sea $X \neq \clubsuit$. Como una o la otra se tienen que dar $P[X = \clubsuit|Y = \diamond] + P[X \neq \clubsuit|Y = \diamond] = 1$, y ambas tienen una probabilidad entre 0 y 1.

Conocida la pmf, la pmf condicional, *por definición* es igual a:

$$p_Y(y_j|x_i) = P[Y|X = x_i] = \frac{P[X = x_i, Y = y_j]}{P[X = x_i]} = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} \quad (4)$$

De esta definición se deduce el teorema de la probabilidad total: pasando la probabilidad marginal multiplicando al otro miembro y sumando para la condición:

$$p_Y(y_j) = \sum_i p_Y(y_j|x_i)p_X(x_i) \quad (5)$$

También se puede deducir de esta definición el teorema de Bayes.

1.1.4. Independencia estadística

Dadas dos variables aleatorias X e Y , se dice que son independientes cuando la probabilidad de que ocurra un evento $X = x$ no se ve afectada por los valores que tome Y .

$$P[X = x|Y] = P[X = x] \quad (6)$$

Teniendo en cuenta la definición de probabilidad condicional, esta condición equivale a:

$$p_{X,Y}(x_i, y_j) = P[X = x_i, Y = y_j] = P[X = x_i] \cdot P[Y = y_j] = p_X(x_i) \cdot p_Y(y_j) \quad (7)$$

es decir, cuando dos variables aleatorias son independientes su pmf conjunta se factoriza como el producto de pmf marginales para cada variable.

Ejemplo 2 Consideramos la probabilidad condicional de sacar un as de una baraja de 40 cartas, dado que la anterior carta sacada fue un as $P[X = \clubsuit|Y = \diamond]$. En el caso de que volvamos a meter la carta en el mazo tras sacar la primera, los dos sucesos son independientes y podemos ver que $P[X = \clubsuit|Y = \diamond] = P[X = \clubsuit] \cdot P[Y = \diamond] = 4/40 \cdot 4/40 = 0,01$, pudiéndose sacar el mismo as del mismo palo las dos veces. Si no volvemos a meter la carta, los sucesos ya no son independientes.

1.1.5. Valores esperados y momentos conjuntos

Definimos el momento conjunto $k\ell$ -ésimo de la variable aleatoria $\mathbf{X} = (X, Y)$ como:

$$E[X^k Y^\ell] = \sum_i \sum_j x_i^k y_j^\ell p_{X,Y}(x_i, y_j) \quad (8)$$

Cuando $k = 0$ se obtienen los momentos de la variable Y , y los de X cuando $\ell = 0$. El momento conjunto con $k = 1, \ell = 1$ recibe un nombre especial: *correlación*:

$$CORR(X, Y) = E[XY] \quad (9)$$

Los momentos conjuntos centrales de X e Y , obtenidos centrando las variables $X - E[X]$ y $Y - E[Y]$ se definen como $E[(X - E[X])^k (Y - E[Y])^\ell]$. Nótese que $k = 2, \ell = 0$ corresponde con $VAR(X)$, mientras que $k = 0, \ell = 2$ corresponde con $VAR(Y)$. Otro caso de particular interés es el momento conjunto central con $k = 1, \ell = 1$, que se conoce como *covarianza*:

$$COV(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \quad (10)$$

La covarianza mide la *dependencia lineal* entre las variables X e Y .

Si dos variables aleatorias X e Y son independientes entonces $E[XY] = E[X]E[Y]$, y por lo tanto $COV(X, Y) = 0$. Sin embargo, $COV(X, Y) = 0$ no garantiza que X e Y sean independientes.

Se define el coeficiente de correlación (de Pearson):

$$\rho_{X,Y} = \frac{COV(X, Y)}{\sigma_X \sigma_Y} \quad -1 < \rho_{X,Y} < 1 \quad (11)$$

donde $\sigma_X = \sqrt{VAR(X)}$ y $\sigma_Y = \sqrt{VAR(Y)}$. Dos variables aleatorias se dirá que están correladas si su coeficiente de correlación es distinto de cero. En el caso contrario se dirá que no existe correlación o están incorreladas.

La ventaja de usar el coeficiente de correlación para comparar el grado de dependencia lineal entre diferentes pares de variables con respecto a la covarianza es que el coeficiente de correlación está normalizado, y por lo tanto los valores son comparables. Es posible definir otros coeficientes de correlación diferentes al de Pearson.

2. Múltiples variables aleatorias continuas

Siguiendo los principios establecidos en el caso de una única variable aleatoria continua, podemos extender la exposición anterior al caso de múltiples variables aleatorias continuas X_1, X_2, \dots, X_n . Como en el caso univariado, la densidad de probabilidad cobra una relevancia especial en el caso de variables aleatorias continuas.

2.1. Función densidad de probabilidad conjunta

Un conjunto de n variables aleatorias continuas $\mathbf{X} = (X_1, X_2, \dots, X_n)$ quedará completamente caracterizada por su función densidad de probabilidad conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$, es decir, la densidad de probabilidad de la variable aleatoria en todo el espacio muestral Ω , donde $\mathbf{x} \in \Omega$ simboliza los infinitos posibles valores que puede tomar la variable aleatoria \mathbf{X} .

Ejemplo 3 Suponiendo que las temperaturas que medimos están distribuidas uniformemente en el intervalo del termómetro, $f_{X,Y}(x, y) = 1/1600$, con $-20 < x < 60, -20 < y < 60$

La pdf no representa la probabilidad de que ocurra un valor (x_1, x_2, \dots, x_n) , ya que ésta es cero para un valor concreto y la pdf no. La pdf representa la posibilidad de que ocurra este valor y el área bajo la curva $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ representa la probabilidad. En concreto, la probabilidad de que la variable aleatoria multidimensional \mathbf{X} tome un valor restringido a la región A del espacio muestral es:

$$P[\mathbf{x} \in A] = \int \int \dots \int_{\mathbf{x} \in A} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (12)$$

La pdf debe cumplir que $P(\Omega) = 1$, es decir que la probabilidad de que la variable aleatoria tome *alguno* de los valores dentro del espacio muestral es la unidad. Esta condición se suele escribir como:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1 \quad (13)$$

2.2. Densidades de Probabilidad Marginales

A la pdf conjunta se le asocia una familia de pdf marginales. La pdf marginal de X_1 será:

$$f_{X_1}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_2 \dots dx_n \quad (14)$$

es decir, se integra en todas las variables salvo la variable marginal X_1 . De la misma forma, la pdf marginal de X_2, X_3, \dots, X_n :

$$f_{X_2, X_3, \dots, X_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 \quad (15)$$

Conocida la pdf conjunta se pueden determinar las marginales pero, en general, a partir de las marginales no es posible determinar la pdf conjunta.

2.3. Densidades de Probabilidad condicional

A su vez, se pueden definir una familia de pdf condicionales a partir de la pdf conjunta:

$$f_{X_n}(x_n | x_1, x_2, \dots, x_{n-1}) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{f_{X_1, X_2, \dots, X_{n-1}}(x_1, x_2, \dots, x_{n-1})} \quad (16)$$

en general, fijada una condición, la pdf condicional será el conciente entre la pdf conjunta y la pdf marginal de la condición. De esta definición se puede deducir el teorema de la probabilidad total y el teorema de Bayes.

2.4. Independencia estadística

En el caso de multiples variables aleatorias continuas, la condición de independencia se expresa como:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \dots \cdot f_{X_n}(x_n) \quad (17)$$

es decir, cuando dos variables aleatorias son independientes su pdf conjunta se factoriza como el producto de pdf marginales para cada variable.

3. Vectores de variables aleatorias

Un conjunto de variables aleatorias puede ser agrupada en un vector para aprovechar la nomenclatura y agilidad del cálculo vectorial.

Un vector de variables aleatorias es:

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \quad (18)$$

donde X_1, X_2, \dots, X_n son variables aleatorias.

3.1. Valores esperados y momentos conjuntos para vectores de variables aleatorias

El valor esperado para un vector de variables aleatorias es:

$$E[\vec{X}] = E \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix} = \vec{\mu}_X \quad (19)$$

Se puede también construir la *matriz de correlación* \mathbf{R}_X , considerando las correlaciones dos a dos para cada una de las variables aleatorias X_1, X_2, \dots, X_n que componen el vector de variables aleatorias:

$$\mathbf{R}_X = \begin{pmatrix} E[X_1^2] & E[X_1 X_2] & \cdots & E[X_1 X_n] \\ E[X_2 X_1] & E[X_2^2] & \cdots & E[X_2 X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_n X_1] & E[X_n X_2] & \cdots & E[X_n^2] \end{pmatrix} \quad (20)$$

Esta matriz es simétrica y se puede escribir de forma resumida usando notación vectorial como:

$$\mathbf{R}_X = E[\vec{X} \cdot \vec{X}^t] \quad (21)$$

El elemento ij -ésimo de la matriz de correlación viene determinado por:

$$(\mathbf{R}_X)_{ij} = E[X_i X_j] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X_i, X_j}(x, y) dx dy, & \text{para } X_i, X_j \text{ continuas} \\ \sum_{x_i} \sum_{x_j} x_i x_j p_{X_i, X_j}(x_i, x_j), & \text{para } X_i, X_j \text{ discretas} \end{cases} \quad (22)$$

La *matriz de covarianza* \mathbf{C}_X se define como la matriz de correlación para vectores de variables aleatorias centradas, es decir:

$$\mathbf{C}_X = E[(\vec{X} - E[\vec{X}]) \cdot (\vec{X} - E[\vec{X}])^t] = E[\vec{X} \cdot \vec{X}^t] - E[\vec{X}](E[\vec{X}])^t = \mathbf{R}_X - \vec{\mu}_X \cdot \vec{\mu}_X^t \quad (23)$$

3.2. Transformaciones lineales de vectores de variables aleatorias

Las variables aleatorias \vec{X} se pueden combinar para formar nuevas variables aleatorias \vec{Z} , como $\vec{Z} = g(\vec{X})$, donde g es una función.

Ejemplo 1 Se puede construir nueva la variable $Z = X + Y$, es decir, una variable aleatoria cuyo valor sea la suma de los dos valores obtenidos por cada dado. El valor $Z = 5$ podrá obtenerse de varias maneras $(X = 1, Y = 4), (X = 2, Y = 2), \dots$ y será, por tanto, más probable que el valor $Z = 2$.

Entre todas las funciones g posibles, las más sencillas son las funciones lineales. En ese caso, cada una de las nuevas variables definidas Z_i será una combinación lineal de las anteriores X_i , de forma que podemos expresar $Z_i = a_{i1} \cdot X_1 + a_{i2} \cdot X_2 + \cdots + a_{in} X_n$. Expresado en forma matricial:

$$\vec{Z} = \mathbf{A} \cdot \vec{X} \quad (24)$$

donde \mathbf{A} es una matriz $n \times n$ con coeficientes constantes. Los valores esperados del nuevo vector de variables aleatorias \vec{Z} se pueden expresar en función de \vec{X} como:

$$\vec{\mu}_Z = E[\vec{Z}] = E[\mathbf{A} \cdot \vec{X}] = \mathbf{A} \cdot E[\vec{X}] = \mathbf{A} \cdot \vec{\mu}_X \quad (25)$$

donde hemos utilizado la propiedad de linealidad del valor esperado. Igualmente los momentos se obtienen como:

$$\mathbf{R}_Z = E[\vec{Z} \cdot \vec{Z}^t] = E[(\mathbf{A} \cdot \vec{X})(\mathbf{A} \cdot \vec{X})^t] = E[\mathbf{A} \cdot \vec{X} \vec{X}^t \cdot \mathbf{A}^t] = \mathbf{A} \cdot E[\vec{X} \vec{X}^t] \cdot \mathbf{A}^t = \mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^t \quad (26)$$

donde de nuevo hemos usado la linealidad del momento esperado y $(\mathbf{A} \cdot \vec{X})^t = \vec{X}^t \cdot \mathbf{A}^t$. De igual forma $\mathbf{C}_Z = \mathbf{A} \cdot \mathbf{C}_X \cdot \mathbf{A}^t$. Puesto que la matriz de covarianza es simétrica y se transforma de la anterior forma bajo una transformación lineal, podemos afirmar que, dado un conjunto de variables aleatorias \vec{X} , siempre existe un nuevo conjunto de variables aleatorias \vec{Z} combinación lineal de las anteriores, cuya matriz de covarianza es diagonal. Dicho de otro modo, la matriz de covarianza es diagonalizable mediante una transformación de congruencia.

Ejemplo Considere un conjunto de variables aleatorias incorreladas cuya matriz de covarianza es diagonal e igual a la unidad $\mathbf{C}_X = \mathbf{1}$. Cualquier transformación lineal de estas variables producirá una nueva matriz de covarianza $\mathbf{C}_Z = \mathbf{A} \mathbf{A}^t$, que en general no será diagonal, y por lo tanto las nuevas variables estarán correladas.

3.3. Variables Gaussianas

Una variable unidimensional gaussiana se caracteriza por tener una pdf con la forma:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (27)$$

donde μ es el valor esperado y σ^2 la varianza.

Esta distribución se generaliza para el caso multivariado como:

$$f_{\vec{X}}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{C}}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^t \mathbf{C}^{-1}(\vec{X}-\vec{\mu})} \quad (28)$$

donde \mathbf{C} es la matriz de covarianza, y $\vec{\mu} = E[\vec{X}]$. Las variables aleatorias descritas por la anterior pdf se dice que son conjuntamente gaussianas.

3.3.1. Dos variables aleatorias conjuntamente gaussianas

En el caso de dos variables aleatorias X, Y conjuntamente gaussianas la matriz de covarianza se simplifica a:

$$\mathbf{C} = \begin{pmatrix} \sigma_x^2 & \rho_{XY}\sigma_x\sigma_y \\ \rho_{XY}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \quad (29)$$

donde σ_x^2, σ_y^2 son las varianzas y ρ_{XY} el coeficiente de correlación. Podemos calcular su inversa como:

$$\mathbf{C}^{-1} = \kappa \begin{pmatrix} \sigma_y^2 & -\rho_{XY}\sigma_x\sigma_y \\ -\rho_{XY}\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} \quad (30)$$

donde $\kappa = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho_{XY}^2)} = (\det \mathbf{C})^{-1}$. Por lo tanto la pdf de dos variables conjuntamente gaussianas es:

$$f_{XY}(x, y) = \frac{\sqrt{\kappa}}{(2\pi)} \exp \left[-\frac{\kappa}{2} (x - \mu_x, y - \mu_y) \begin{pmatrix} \sigma_y^2 & -\rho_{XY}\sigma_x\sigma_y \\ -\rho_{XY}\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \right] \quad (31)$$

desarrollando el producto matricial contenido en el exponente se llega a:

$$f_{XY}(x, y) = \frac{\sqrt{\kappa}}{(2\pi)} \exp \left[-\frac{1}{2(1-\rho_{XY}^2)} \left(\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + 2\rho_{XY} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right) \right] \quad (32)$$

Es interesante considerar el caso particular en el que $\rho_{XY} = 0$, es decir, cuando no existe correlación entre las dos variables aleatorias conjuntamente gaussianas. En ese caso, la ecuación anterior se simplifica y podemos escribir que:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right) \right] \quad (33)$$

lo que permite afirmar que:

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} = f_X(x) \cdot f_Y(y) \quad (34)$$

En otras palabras, dos variables aleatorias conjuntamente gaussianas e incorreladas son también *independientes*.

3.3.2. Correlación e independencia de multiples variables aleatorias gaussianas

En el caso general de multiples variables aleatorias conjuntamente gaussianas podemos comprobar que la condición de no correlación también implica independencia estadística entre las variables.

Un conjunto de n variables aleatorias conjuntamente gaussianas $\vec{X} = (X_1, X_2, \dots, X_n)^t$ que no están correladas entre sí, tiene una matriz de covarianza diagonal $\mathbf{C} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Por lo tanto, su pdf conjunta es factorizable:

$$f_{\vec{X}}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{C}}} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) \quad (35)$$

En general, para variables conjuntamente gaussianas siempre es posible encontrar una transformación lineal $\vec{Z} = \mathbf{A} \cdot \vec{X}$ que defina un nuevo conjunto de variables aleatorias conjuntamente gaussianas \vec{Z} , cuya matriz de covarianza sea diagonal $C_Z = \mathbf{A} \cdot \mathbf{C}_X \cdot \mathbf{A}^t$ y conformen un conjunto de variables aleatorias independientes.

4. Aproximación axiomática a la probabilidad

Basándose en la teoría de conjuntos y teoría de la medida, los siguientes tres axiomas definen el concepto de espacio de probabilidad, constituido por (Ω, F, P) , donde el espacio muestral Ω es un conjunto, F es una generalización de subconjunto², y P es una medida tal que:

Amioma 1: Dado un evento E dentro del espacio de eventos F , la probabilidad es siempre positiva $P(E) > 0$

Amioma 2: La probabilidad de que ocurra al menos un evento es 1: $P(\Omega) = 1$

Amioma 3: La probabilidad de un conjunto de eventos *mutuamente exclusivos* es la suma de sus probabilidades: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

De estas definiciones se pueden hacer varias deducciones, haciendo uso de la teoría de conjuntos para describir los eventos, como que $P(\emptyset) = 0$ ó $P(A^c) = 1 - P(A)$. También se puede deducir la probabilidad de un conjunto de eventos a $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$, para casos no excluyentes y se puede definir la probabilidad condicional como $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Para más detalles consultar la bibliografía.

5. Bibliografía recomendada

1. Capítulos 5 y 6 de A. León-García, *Probability and Random Processes for Electrical Engineering*, (2^a edición). Addison Wesley. 1994.
2. Jr. Peebles, Z. Peyton, *Probability, Random Variables and Random Signal Principles*, (4^a edición), Mc Graw-Hill. 2001.
3. S. Kay, *Intuitive probability and random processes using matlab*. 2006. New York: Springer, 2006.

² F es una σ -álgebra

Tema 2: Procesos Aleatorios

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons](#) “Reconocimiento-NoCommercial-CompartirIgual 3.0 España”.



1. Definición

A una familia de variables aleatorias $X_i(\xi)$ indexadas por un parámetro t se denomina proceso aleatorio o estocástico $X(\xi, t)$. El parámetro que indexa la familia de variables aleatorias puede ser discreto n o continuo t , asimilándose habitualmente al tiempo.

Ejemplo 1 Continuando con el ejemplo 1 de variable aleatoria unidimensional: El experimento aleatorio constituido por lanzar una moneda al aire y registrar el resultado puede caracterizarse por una variable aleatoria X tal que $X(\oplus) = 1$ y $X(\odot) = 0$, donde \oplus simboliza el resultado ‘cara’ y \odot simboliza el resultado ‘cruz’. Podemos genéricamente llamar ξ a cualquiera de los resultados \odot, \oplus . Si se registra el resultado de este experimento aleatorio obtenido en cada instante de tiempo con un índice n , se construye un proceso aleatorio que se conoce con el nombre de proceso de Bernuilli $X(\xi, n)$. La notación se suele simplificar a X_n , omitiendo la mención al evento ξ y enfatizando el carácter discreto del tiempo n . Así, para un valor concreto de tiempo i , X_i toma valores $\{0, 1\}$ con probabilidad p y $1 - p$. Una realización del proceso tendrá forma de cadena de caracteres binarios, como 000110100101...

Ejemplo 2 Se puede construir una nueva variable aleatoria D_i a partir de una variable aleatoria de Bernuilli X_i como $D_i = 2 \cdot X_i - 1$, de tal manera que ahora D_i toma valores en $\{-1, 1\}$. Podemos ahora definir un proceso aleatorio S_n tal que $S_n = D_1 + D_2 + \dots + D_n$. Este proceso aleatorio recibe el nombre de camino aleatorio.

Ejemplo 3 El proceso aleatorio definido por $X(\xi, t) = A(\xi) \cos \omega t$, donde A es una variable aleatoria uniformemente distribuida en el intervalo $[-1, 1]$.

Ejemplo 4 El proceso aleatorio definido por $X(\xi, t) = \cos(\omega t + \Theta(\xi))$, donde Θ es una variable aleatoria uniformemente distribuida en el intervalo $[-\pi, \pi]$.

En un proceso aleatorio $X(\xi, t)$, fijado el evento ξ para cada instante de tiempo, se genera una *función del tiempo* $x(t)$ que se conoce como *realización* del proceso aleatorio. Si por el contrario se fija el instante del tiempo t , se obtiene una variable aleatoria $X(\xi)$. A menudo se omite la referencia al evento ξ , y se habla de un proceso aleatorio $X(t)$, donde se mantiene el uso de mayúsculas para referirse a variables aleatorias.

Un proceso aleatorio queda completamente especificado por su pdf (o CDF) conjunta:

$$f_{X(t_1)X(t_2)\dots X(t_n)}(x_1, x_2, \dots, x_n) \quad \forall t_1, t_2, \dots, t_n \quad \forall n \quad (1)$$

En general, la pdf conjunta cambiará en función de los t_1, t_2, \dots, t_n que elijamos, haciéndola una función también del tiempo. En el caso discreto queda especificada por su pmf conjunta:

$$p_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) \quad \forall n \quad (2)$$

El valor esperado de un proceso aleatorio en el caso continuo es:

$$E[X(t)] = \int_{-\infty}^{\infty} x(t)f_{X(t)}(x)dx = \mu(t) \quad (3)$$

y en el caso discreto:

$$E[X[n]] = \sum_i x_i[n]p_{X[n]}(x_i[n]) = \mu[n] \quad (4)$$

Cuando no existe lugar a confusión, se denota como X_i a la variable aleatoria obtenida en el instante i del proceso aleatorio ($X(t_i)$ ó $X[i]$).

Se pueden también definir los momentos de orden superior de un proceso aleatorio, como la *autocorrelación*:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{X(t_1)X(t_2)}(x, y)dxdy \quad (5)$$

para tiempo continuo, y:

$$R_X[n_1, n_2] = E[X[n_1]X[n_2]] = \sum_i \sum_j x_i y_j p_{X[n_1]X[n_2]}(x_i, y_j) \quad (6)$$

Si centramos las variables, se define la *autocovarianza*. Para tiempo continuo:

$$C_X(t_1, t_2) = E[(X_1 - E[X_1])(X_2 - E[X_2])] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu(t_1))(y - \mu(t_2))f_{X_1 X_2}(x, y)dxdy \quad (7)$$

la autocovarianza a tiempos iguales $t_1 = t_2$, es igual a la varianza $C_{XX} = VAR(X) = \sigma^2(t)$. Para tiempo discreto:

$$C_X[n_1, n_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = \sum_i \sum_j (x_i - \mu[n_1])(y_j - \mu[n_2])p_{X_1 X_2}(x_i, y_j) \quad (8)$$

donde de nuevo la autocovarianza a tiempos iguales $n_1 = n_2$, es igual a la varianza $C_{XX} = VAR(X) = \sigma^2[n]$

1.1. Procesos i.i.d

Los procesos aleatorios formados por variables aleatorias independientes cumplen que su pdf conjunta se puede factorizar:

$$f_{X(t_1)X(t_2)\dots X(t_n)}(x_1, x_2, \dots, x_n) = f_{X(t_1)}(x_1) \cdot f_{X(t_2)}(x_2) \dots f_{X(t_n)}(x_n) \quad \forall t_1, t_2, \dots, t_n \quad \forall n \quad (9)$$

si además se cumple que las variables aleatorias tienen la misma pdf para cualquier instante de tiempo: $f_{X_i} = f_{X_j} \quad \forall i, j$ diremos que las variables están idénticamente distribuidas. Un proceso i.i.d. es aquel formado por un conjunto de variables aleatorias independientes e idénticamente distribuidas.

Ejemplo 1 El proceso de Bernuilli es un proceso i.i.d. Se puede determinar la pmf conjunta del proceso, ya que $P[\xi_1, \xi_2, \dots, \xi_n] = p^k(1-p)^{n-k}$ donde k es el número de veces que $\xi = 1$ en la secuencia de n caracteres binarios.

1.2. Procesos Gaussianos

Un proceso Gaussiano $X(t)$ es aquel cuya pdf conjunta esta formada por variables conjuntamente gaussianas. Si denominamos \vec{X} al vector de variables aleatorias formadas por:

$$\vec{X} = (X(t_1), X(t_2), \dots, X(t_n)), \quad \forall n \quad \forall t \quad (10)$$

entonces la pdf conjunta viene dada por:

$$f_{\vec{X}}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{C}}} e^{-\frac{1}{2}(\vec{X} - \vec{\mu})^t \mathbf{C}^{-1}(\vec{X} - \vec{\mu})} \quad (11)$$

donde \mathbf{C} es la matriz de correlación y $\vec{\mu}$ el vector de valores esperados.

Los procesos gaussianos tienen varias propiedades:

- Un proceso gaussiano queda completamente especificado por 2 parametros: $\vec{\mu} = E[\vec{X}]$ y $\mathbf{C} = E[(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^t]$
- La combinación lineal de procesos gaussianos produce procesos gaussianos¹
- En un proceso gaussiano de media cero², si la autocorrelación es cero $R_X(t, t+\tau) = 0, \forall \tau \neq 0$ entonces el proceso es i.i.d., formado por variables aleatorias gaussianas *independientes* (no correlación implica independencia)
- Un proceso gaussiano estacionario en sentido amplio, es también estacionario en sentido estricto.

2. Procesos estacionarios

Un proceso estacionario es aquel cuyas propiedades estadísticas no dependen de dónde se fije el origen de tiempos $t = 0$. En términos matemáticos, se define un proceso estacionario (en sentido estricto) como aquel cuya densidad de probabilidad (pdf o CDF) conjunta no cambia al desplazarse una cantidad τ en el tiempo:

$$f_{X(t_1+\tau)X(t_2+\tau)\dots X(t_n+\tau)}(x_1, x_2, \dots, x_n) = f_{X(t_1)X(t_2)\dots X(t_n)}(x_1, x_2, \dots, x_n) \quad (12)$$

esto debe cumplirse para todo n y para cualquier elección de tiempos. En concreto, la pdf de primer orden ($n = 1$) debe cumplirlo, cumpliéndose en consecuencia que:

$$E[X(t + \tau)] = \int_{-\infty}^{\infty} x f_{X(t+\tau)}(x) dx = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx = E[X(t)] = \text{cte} \quad (13)$$

Es decir, el valor esperado de un proceso aleatorio estacionario no cambia con el tiempo. Por otro lado, la pdf conjunta de segundo orden ($n = 2$) también debe cumplir la condición de estacionaridad, y como consecuencia:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(t_1)X(t_2)}(x, y) dx dy = \quad (14)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(0)X(t_2-t_1)}(x, y) dx dy = R_X(t_2 - t_1) \quad (15)$$

$\forall t_1, t_2$, donde se hace uso de la libertad para fijar $\tau = -t_1$.

Tanto $E[X(t)]$ como R_X solo describen el proceso aleatorio $X(t)$ de forma parcial, y las condiciones dadas en las eqs. 14 y 13 no garantizan que el proceso sea estacionario. En general será difícil determinar si un proceso es estacionario en sentido estricto o no.

¹Esta propiedad se utiliza para demostrar que la salida de un sistema LTI es un proceso gaussiano si la entrada es un proceso gaussiano, *en tiempo discreto*. Cuando se generaliza esta idea a sistemas en tiempo continuo existen numerosas sutilezas matemáticas que deben tratarse con rigor, dando lugar a definiciones más generales de procesos gaussianos que la presentada aquí (véase por ejemplo Haykin *Digital communications*, [o este debate](#))

²también valido si la media no es cero pero $\mathbf{C}_X(t_1, t_2) = 0, \forall t_1 \neq t_2$

2.1. Procesos estacionarios en sentido amplio (WSS)

Un proceso estacionario en sentido amplio (Wide sense stationary o WSS) es aquel que cumple las condiciones dadas en las eqs. 14 y 13. A menudo, la condición 14 se escribirá como:

$$R_X(t, t + \tau) = E[X(t)X(t + \tau)] = R_X(\tau) \quad (16)$$

donde $\tau = t_2 - t_1$.

Ejemplo 3 Mientras que $E[X(t)] = 0$ para este proceso, cumpliéndose la primera condición, la autocorrelación no cumple la condición de ser dependiente únicamente del desplazamiento.

Ejemplo 4 Para este proceso $E[X(t)] = 0$ y $R_X(t, t + \tau) = \frac{1}{2} \cos(w\tau)$, siendo por lo tanto WSS.

De la definición de la función de autocorrelación de un proceso estacionario en sentido amplio se obtienen las siguientes propiedades:

- La autocorrelación a diferencia de tiempo $\tau = 0$ mide la **potencia promedio** del proceso:³

$$R_X(0) = E[X^2(t)] \quad (17)$$

- La autocorrelación es una función par:

$$R_X(\tau) = R_X(-\tau) \quad (18)$$

- La autocorrelación a diferencia de tiempo $\tau = 0$ toma su valor máximo:⁴

$$R_X(0) \geq |R_X(\tau)| \quad (19)$$

2.2. Procesos cicloestacionarios

La periodicidad de un proceso aleatorio estacionario se estudia desde el punto de vista estadístico. En este sentido, se define un proceso *cicloestacionario* como aquel que cuya pdf conjunta se repite cuando el desplazamiento temporal es un múltiplo entero n de un periodo T : $f_{X(t_1+nT)X(t_2+nT)\dots X(t_n+nT)}(x_1, \dots, x_n) = f_{X(t_1)X(t_2)\dots X(t_n)}(x_1, \dots, x_n)$.

Se puede definir también una versión de cicloestacionariedad en sentido amplio relajando esta restricción a los momentos de primer orden:

$$\mu_X(t + nT) = \mu_X(t), \quad \forall t, \quad n = 0, 1, 2, \dots \quad (20)$$

$$R_X(t_1 + nT, t_2 + nT) = R_X(t_1, t_2), \quad \forall t_1, t_2, \quad n = 0, 1, 2, \dots \quad (21)$$

En el caso de procesos WSS, la ecuación 21 implica que la autocorrelación es una función periódica, con periodo T . Además, se puede afirmar que el valor máximo de la autocorrelación, la potencia media $R(0)$, se repite periódicamente y, por el mismo argumento utilizado para deducir que $R(0)$ debe ser el valor máximo, se deduce que el proceso WSS cicloestacionario es periódico en sentido cuadrático medio, es decir:

$$E[(X(t + nT) - X(t))^2] = 0 \quad n = 0, 1, \dots \quad (22)$$

³Si las variables del proceso están centradas (tienen media cero) entonces coincide también con la varianza, es decir mide la dispersión de los datos.

⁴A partir de la cantidad positiva $(X(t + \tau) - X(t))^2$, se puede comprobar que $E[(X(t + \tau) - X(t))^2] \geq 0$. Desarrollando el cuadrado y haciendo uso de que el proceso es WSS se llega a $2R_X(0) - 2R_X(\tau) \geq 0$, deduciéndose la desigualdad.

3. Promedios temporales

Se define el promedio temporal de un proceso aleatorio como:

$$\bar{\mu}_X = \frac{1}{2T} \int_{-T}^T x(t) dt \quad (23)$$

donde $x(t)$ es una *realización* del proceso aleatorio $X(t)$. El promedio temporal $\bar{\mu}_X$ es una variable aleatoria que toma un valor para cada realización del proceso que, en general, depende de t y T . En el caso de procesos WSS, se cumple que $E[\bar{\mu}_X] = \mu_X$, como se deduce de la definición.

De la misma forma, se define la autocorrelación promedio como:

$$\bar{\mathcal{R}}_X = \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) dt \quad (24)$$

donde, de nuevo, $x(t)$ es una *realización* del proceso aleatorio $X(t)$. Para procesos estacionarios $E[\bar{\mathcal{R}}_X(\tau)] = R_X(\tau)$.

3.1. Procesos ergódicos

Un proceso aleatorio $X(t)$ es ergódico si se puede estimar los parámetros estadísticos del proceso a partir de los promedios temporales. Concretamente, se define un proceso WSS⁵ ergódico en la media como aquel que cumple que:

$$\lim_{T \rightarrow \infty} \bar{\mu}_X = \mu_X \quad (25)$$

siempre que la varianza $\text{VAR}(\bar{\mu}_X)$ tienda también a cero en el mismo límite.

Se denominan procesos ergódicos en la autocorrelación a aquellos procesos aleatorios WSS $X(t)$ que cumplen que:

$$\lim_{T \rightarrow \infty} \bar{\mathcal{R}}_X(\tau) = R_X(\tau) \quad (26)$$

La varianza del promedio de la autocorrelación temporal debe también tender a cero (en sentido cuadrático medio) para ser ergódico en la autocorrelación.

4. Código

Los ejemplos 1 y 2 vienen acompañados de un archivo en MATLAB para simularlos numéricamente:

`ProcesosAleatoriosDiscretos.m`

Los ejemplos 3 y 4 vienen acompañados de un archivo en MATLAB para simularlos numéricamente:

`ProcesosAleatoriosContinuos.m`

5. Bibliografía recomendada

1. Capítulos 5 y 6 de A. León-García, *Probability and Random Processes for Electrical Engineering*, (2ª edición). Addison Wesley. 1994.
2. Capítulo 1, S. Haykin *Communication systems*, (4ª edición), John Wiley and Sons. 2001.
3. Capítulos 15 a 22 [Una aproximación con énfasis en procesos en tiempo discreto] de S. Kay, *Intuitive probability and random processes using matlab*. 2006. New York: Springer, 2006.

⁵La condición de estacionariedad es necesaria porque el promedio temporal del proceso $\bar{\mu}_X$ es un valor constante, y por tanto solo tiene sentido aproximar el promedio estadístico con el temporal para procesos en los que el promedio estadístico μ_X es también constante, es decir, WSS.

Tema 3: Caracterización espectral de procesos aleatorios

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons](#) “Atribución-NoComercial-CompartirIgual 3.0 No portada”.



1. Densidad espectral de potencia

Conocer el contenido espectral de un proceso aleatorio $X(t)$ es de especial relevancia en telecomunicaciones. Sin embargo, la forma más inmediata de obtenerlo, a través del proceso transformada de fourier $\hat{X}(f)$, tiene varias dificultades técnicas. La principal es que no está bien definida para la mayorá de procesos de interés.¹

El atajo mas habitual es restringir el estudio del proceso aleatorio $X(t)$ a un intervalo temporal finito $[-T, T]$. Llamaremos $X_T(t) = u(t)X(t)$ a esta restricción, donde $u(t)$ es una función de tipo escalón que vale 1 si $|t| < T$ y 0 en otro caso. Para una realización $x_T(t)$ del proceso $X_T(t)$ se puede obtener su transformada de fourier:

$$\hat{x}_T(f) = \int_{-\infty}^{\infty} x_T(t)e^{-j2\pi ft}dt = \int_{-T}^T x_T(t)e^{-j2\pi ft}dt \quad (1)$$

donde ahora no hay problemas de convergencia por estar limitados al intervalo $[-T, T]$. Existirán diferentes transformadas $\hat{x}_T(f)$ para las diferentes realizaciones del proceso, dando lugar al proceso transformado $\hat{X}_T(f)$, del que será de interés estudiar sus propiedades estadísticas tales como el valor esperado o la autocorrelación. Mientras que el valor esperado no suele ser de gran interés, sí lo es la autocorrelación, por estar relacionado con la potencia de la señal como veremos a continuación.

La potencia de una señal estadística se asocia con $|x(t)|^2$, por analogía con los casos en los que $x(t)$ representa un voltaje. Podemos definir el promedio temporal de la potencia de una realización del proceso aleatorio como:

$$\bar{p}_T = \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt = \frac{1}{2T} \int_{-\infty}^{\infty} |x_T(t)|^2 dt \quad (2)$$

donde se ha introducido por conveniencia la función restringida $x_T(t)$. Gracias al teorema de Parseval, se expresa de forma equivalentemente que:

$$\bar{p}_T = \frac{1}{2T} \int_{-\infty}^{\infty} |\hat{x}_T(f)|^2 df \quad (3)$$

donde ahora la variable de integración es la frecuencia y, por tanto, la potencia promedio para ésta realización se puede obtener a partir de la integral de cierta densidad espectral de potencia. El

¹Para que exista la transformada de fourier de una función $x(t)$, a menudo se exige que sea de cuadrado integrable $\int |x(t)|^2 dt < \infty$ o integrable en valor absoluto $\int |x(t)| dt < \infty$, condición que es suficiente (pero no necesaria). Muchos procesos aleatorios de interés (ruido blanco) no son de cuadrado integrable.

integrando de esta ecuación $\frac{1}{2T}|\hat{x}_T(f)|^2$ se conoce con el nombre de estimación de *periodograma*. En la práctica, puesto que las observaciones de cualquier señal están limitadas a un intervalo temporal, el periodograma será la estimación de la densidad espectral de potencia.

El valor de \bar{p}_T será distinto para cada realización del proceso, haciendo que estemos interesados en conocer su valor esperado $P_T = E[\bar{p}_T]$. Por otro lado, también estaremos interesados en conocer la potencia promedio para todo tiempo, y no solo para la restricción entre $[-T, T]$. Por ello, definimos la potencia promedio del proceso como:

$$P = \lim_{T \rightarrow \infty} E[\bar{p}_T] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} E[|\hat{X}_T(f)|^2] df \quad (4)$$

donde ahora se hace referencia a todo el proceso \hat{X}_T , y no a una única realización. Puesto que la integral esta definida en el rango de frecuencias, podemos identificar el integrando como una densidad de potencia:

$$S_X(f) = \lim_{T \rightarrow \infty} \frac{E[|\hat{X}_T(f)|^2]}{2T} \quad (5)$$

de forma que:

$$P = \int_{-\infty}^{\infty} S_X(f) df \quad (6)$$

A $S_X(f)$ se conoce con el nombre de *densidad espectral de potencia* o PSD, por sus siglas en inglés. La densidad espectral de potencia se obtiene como el promedio estadístico de las diferentes estimaciones de periodograma tomando el límite en el que observamos el proceso durante un tiempo indefinidamente largo. Para el caso de procesos WSS (ver Tema 2), la potencia promedio viene dada por:

$$P = R_X(0) = E[X^2(t)] \quad (7)$$

lo que sugiere una conexión entre la autocorrelación (en el dominio del tiempo) y la densidad espectral de potencia (en el dominio de la frecuencia).

2. Teorema de Wiener-Khinchin

La relación que existe entre la autocorrelación y la densidad espectral de potencia viene motivada por el teorema de convolución, que establece que la transformada de fourier de una convolución (como la autocorrelación) es el producto de sus transformadas. Si tomamos la definición de densidad espectral de potencia de la ecuación 5, y lo expresamos en el dominio del tiempo:

$$S_X(f) = \lim_{T \rightarrow \infty} \frac{E[|\hat{X}_T(f)|^2]}{2T} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T dt \int_{-T}^T dt' E[X_T(t)X_T(t')]e^{-j2\pi f(t-t')} \quad (8)$$

haciendo el cambio de variables $\tau = t - t'$ y reconociendo que en el integrando tenemos el promedio temporal de la autocorrelación $\bar{\mathcal{R}}_X$, se llega a:

$$S_X(f) = \int_{-\infty}^{\infty} d\tau \left\{ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T dt E[X_T(t)X_T(t+\tau)] \right\} e^{-j2\pi f\tau} = \int_{-\infty}^{\infty} d\tau E[\bar{\mathcal{R}}_X(t, t+\tau)]e^{-j2\pi f\tau} \quad (9)$$

Para procesos WSS se cumple que el valor esperado del promedio temporal de la autocorrelación coincide con la autocorrelación $E[\bar{\mathcal{R}}_X] = R_X$, dando lugar al teorema de Wiener-Khinchin:

$$S_X(f) = \int_{-\infty}^{\infty} d\tau R_X(\tau)e^{-j2\pi f\tau} \quad (10)$$

es decir, la densidad espectral de potencia y la autocorrelación son pares de transformadas de fourier. El hecho de que la autocorrelación sea una función par $R_X(\tau) = R_X(-\tau)$ hace que la PSD sea una función real y también par $S_X(f) = S_X(-f)$.

3. Ruido

3.1. Rudio blanco

El ruido blanco es un proceso aleatorio que se caracteriza por tener una densidad espectral de potencia constante. En analogía con el espectro electromagnético, en el que la luz blanca esta compuesta por todos los colores, el ruido blanco está compuesto por todas las frecuencias posibles. El ruido blanco se modela por un conjunto de variables aleatorias $X[n]$ i.i.d. en tiempo discreto de media cero y varianza σ^2 , de forma que:

$$R_X[n, n+k] = E[X[n]X[n+k]] = \begin{cases} 0 & \text{si } n \neq k \\ \sigma^2 & \text{si } n = k \end{cases} \quad (11)$$

donde se ha hecho uso de que el valor esperado del producto es el producto de valores esperados cuando las variables son independientes. Podemos reescribir este resultado usando la delta de Kronecker como $R_X[k] = \sigma^2 \delta_{nk} = \sigma^2 \delta[k]$. Puesto que el proceso es WSS, podemos usar el teorema de Wiener-Khinchin para determinar la densidad espectral de potencia como:

$$S_X[f] = \sum_{k=-\infty}^{\infty} R_X[k] e^{-j2\pi f k} = \sum_{k=-\infty}^{\infty} \sigma^2 \delta[k] e^{-j2\pi f k} = \sigma^2 \quad (12)$$

de forma que la densidad espectral de potencia es constante en toda frecuencia.

Este modelo se transfiere al caso de tiempo continuo simplemente sustituyendo la delta de Kronecker por la delta de Dirac, de forma que $R_X(\tau) = N_0/2\delta(\tau)$, y por lo tanto $S_X(f) = N_0/2$. Este modelo idealizado de ruido tiene la propiedad de que la potencia necesaria para generar este tipo de ruido es infinita:

$$P = \int_{-\infty}^{\infty} S_X(f) df = \int_{-\infty}^{\infty} \frac{N_0}{2} df \rightarrow \infty \quad (13)$$

por lo tanto no representa un modelo realista de ruido. Sin embargo, es un modelo sencillo que permite describir procesos observados empíricamente siempre y cuando se den las condiciones adecuadas y el proceso atraviese un sistema lineal.

Si las variables $X[n]$ i.i.d. en tiempo discreto de media cero y varianza σ^2 son además gaussianas hablaremos de ruido blanco gaussiano.²

3.2. Otras formas de ruido

En general, cuando el ruido no tenga un espectro constante hablaremos de ruido coloreado. En concreto, se habla de ruido rosa cuando el espectro de frecuencias sigue una forma $1/f^2$. En ese caso, las frecuencias bajas tendrán mayor presencia que las altas, y de ahí su nombre (de nuevo por analogía con el espectro visible, donde el color rojo pertenece a las frecuencias bajas).

El ruido puede afectar a la transmisión de señales de muchas formas diferentes. El modo más sencillo de modelar los efectos del ruido es considerar que el ruido se añade a la señal de interés. Se denomina *ruido aditivo* cuando se puede modelar la presencia de ruido $W(t)$ en una señal $X(t)$ como:

$$Y(t) = X(t) + W(t) \quad (14)$$

donde $Y(t)$ es la señal contaminada con ruido.

Un ejemplo de ruido muy presente en sistemas electrónicos es el ruido de Jitter. Éste ocurre en sistemas electrónicos que contienen un reloj interno de sincronización. El Jitter es una fluctuación indeseada en la sincronización de algún evento. Se modela como una señal a la que se añade un desplazamiento temporal estocástico $v_j(t) = v(t + j(t))$. El ejemplo 3 del tema anterior sería un caso de ruido de Jitter.

²De nuevo, al trasladar el modelo a tiempo continuo tendremos dificultades para que represente un modelo realista, ya que las variables de ruido blanco gaussiano deberán tener una varianza que tiende a ser infinita $VAR(X(t)) = R_X(0) = N_0/2\delta(0)$.

En señales continuas, se suelen definir eventos cuando se traspasan umbrales predeterminados. Si a la señal se le añade ruido, el traspaso del umbral queda modificado ligeramente, de forma que un umbral convierte ruido aditivo en jitter. Esta es la forma habitual de formación de ruido de jitter. El ruido aditivo y el jitter se pueden relacionar expandiendo $v_j(t)$ en serie de Taylor, y truncando la expansión en primer orden.

4. Código

La última sección viene acompañada de código para generar un ruido aditivo gaussiano en una secuencia digital y ruido de jitter en una imagen:

`Ruido.m`

5. Bibliografía recomendada

1. Capítulo 10 de A. León-García, *Probability and Random Processes for Electrical Engineering*, (2ª edición). Addison Wesley. 1994.
2. Capítulo 1, S. Haykin *Communication systems*, (4ª edición), John Wiley and Sons. 2001.
3. Capítulo 6, PZ. Peebles, *Probability, random variables and random signal principles* (2ª edición). McGraw-Hill
4. Capítulos 17.6 a 18 [Una aproximación con énfasis en procesos en tiempo discreto] de S. Kay, *Intuitive probability and random processes using matlab*. 2006. New York: Springer, 2006.

Tema 4: Respuesta de un sistema LTI a un proceso aleatorio

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons](#) “Atribución-NoComercial-CompartirIgual 3.0 No portada”.



1. Sistemas lineales e invariantes en el tiempo

Considerando una señal $x(t)$ como input de un sistema, obtendremos un output $y(t)$, que podemos representar como $y(t) = T[x(t)]$. Un sistema lineal es aquel que cumple que:

$$T[\alpha_1 x_1(t) + \alpha_2 x_2(t)] = \alpha_1 T[x_1(t)] + \alpha_2 T[x_2(t)] \quad (1)$$

donde $\alpha_1, \alpha_2 \in \mathbb{R}$. Si además se cumple que:

$$T[x(t + \tau)] = y(t + \tau) \quad (2)$$

entonces se dice que el sistema es lineal e invariante en el tiempo (LTI).

Un sistema LTI se caracteriza por su respuesta al impulso $h(t) = T[\delta(t)]$. La respuesta al impulso permite describir la salida del sistema $y(t)$ a una entrada arbitraria $x(t)$ a través del producto de convolución:

$$y(t) = h(t) \star x(t) = \int_{-\infty}^{\infty} h(s)x(t-s)ds \quad (3)$$

El comportamiento de un sistema LTI se puede estudiar también en el dominio de la frecuencia. Se conoce como respuesta en frecuencia $H(f)$ a la transformada de Fourier de la respuesta al impulso:

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt \quad (4)$$

Además, se dice que el sistema es causal si $h(t) = 0$ cuando $t < 0$, y se dice que el sistema es estable si $\int |h(t)| < \infty$.

2. LTI en procesos aleatorios

Considerando que una realización $x(t)$ de un proceso aleatorio $X(t)$ se puede mostrar que, bajo ciertas restricciones suaves¹, podemos escribir el proceso salida $Y(t)$ de un sistema LTI como:

$$Y(t) = h(t) \star X(t) = \int_{-\infty}^{\infty} h(s)X(t-s)ds \quad (5)$$

¹La integral debe entenderse que converge en sentido cuadrático medio, mientras que no debe haber ninguna realización del proceso que no sea integrable y el sistema debe ser estable.

Es posible relacionar el valor esperado del proceso salida con el de entrada; tomando el valor esperado a ambos lados:

$$\mu_Y = E[Y(t)] = E \left[\int_{-\infty}^{\infty} h(s)X(t-s)ds \right] \quad (6)$$

si el sistema es estable y $E[X(t)]$ es finito para todo t , podemos intercambiar el orden entre valor esperado e integración:

$$\mu_Y = E[Y(t)] = \int_{-\infty}^{\infty} h(s)E[X(t-s)]ds = \mu_X \int_{-\infty}^{\infty} h(s)ds = \mu_X H(0) \quad (7)$$

Como consecuencia de este resultado podemos ver que si el proceso es de media nula a la entrada, a la salida también lo será. Además, si el sistema no tiene componente DC $H(0) = 0$, también producirá un proceso a la salida de media nula.

Igualmente podemos relacionar la autocorrelación a la salida del sistema con la autocorrelación a la entrada como:

$$R_Y(t, t+\tau) = E[Y(t)Y(t+\tau)] = E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s_1)h(s_2)X(t-s_1)X(t+\tau-s_2)ds_1ds_2 \right] \quad (8)$$

donde de nuevo podemos intercambiar el orden de las operaciones valor esperado e integral si se cumplen las condiciones anteriores para dar:

$$R_Y(t, t+\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s_1)h(s_2)E[X(t-s_1)X(t+\tau-s_2)]ds_1ds_2 = \quad (9)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s_1)h(s_2)R_X(t, t+\tau+s_1-s_2)ds_1ds_2 \quad (10)$$

Suponiendo que el proceso $X(t)$ es WSS, podemos deducir que $Y(t)$ es también WSS, y la autocorrelación en la salida viene dada por:

$$R_Y(\tau) = R_X \star R_h = R_X \star (h \star \hat{h}) = \int_{-\infty}^{\infty} R_X(\tau-t)R_h(t)dt \quad (11)$$

donde $\hat{h}(t) = h(-t)$ es la respuesta al impulso invertida temporalmente y cuya transformada de fourier es $\hat{H}(f) = H^*(f)$, el complejo conjugado de la respuesta en frecuencia, y $R_h(t) = h \star \hat{h}$ es la función de autocorrelación en sentido determinista (donde ahora no se hace promedio estadístico) $\int_{-\infty}^{\infty} h(\tau+t)h(t)dt$ de la respuesta al impulso $h(t)$. Según la ecuación 11, la autocorrelación a la salida es la 'función de correlación' de las autocorrelaciones R_X y R_h .

Aplicando el teorema de la convolución² a la ecuación 11 obtenemos que:

$$S_Y(f) = S_X(f)|H(f)|^2 \quad (12)$$

3. Filtrado lineal de ruido blanco

El resultado de tener ruido a la entrada de un amplificador (que puede ser modelado como un sistema LTI, un filtro con ganancia) es un output variable cuya densidad espectral de potencia es proporcional a la respuesta en magnitud del filtro, resultado al que se llega empíricamente. A la constante de proporcionalidad se le suele denominar $N_0/2$. De la misma forma, por razones experimentales, resulta razonable modelar este proceso de salida como un proceso $Y(t)$ WSS de media cero, de modo que:

$$S_Y(f) = \frac{N_0}{2}|H(f)|^2 \quad \sigma_Y^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} |h(t)|^2 dt = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df \quad (13)$$

²Dadas dos funciones f y g y su convolución $f \star g$, el teorema de la convolución demuestra que su transformada de fourier es igual al producto de las transformadas $\mathcal{F}(f \star g) = \mathcal{F}(f)\mathcal{F}(g)$

donde la varianza se obtiene a través de la potencia promedio por ser un proceso de media cero. La constante $N_0/2$ se elige de esta forma para que N_0 represente la potencia del ruido por unidad de ancho de banda, en watts/Hz, en sentido que la potencia de ruido a la salida de un filtro paso-baja, o paso-banda de B Hz de ancho de banda y ganancia unidad es N_0B watts.

Según lo desarrollado en la sección anterior, podemos identificar la constante de proporcionalidad con la densidad espectral del ruido de entrada $X(t)$, de forma que:

$$S_X(f) = \frac{N_0}{2} \quad (14)$$

y por lo tanto nos lleva a concluir que $R_X(\tau) = N_0/2\delta(\tau)$, donde $\delta(t)$ es la delta de Dirac. Haciendo uso de la ecuación 11:

$$R_Y(\tau) = \int_{-\infty}^{\infty} \frac{N_0}{2} \delta(\tau - t) R_h(t) dt = \frac{N_0}{2} R_h(\tau) \quad (15)$$

de forma que las autocorrelación a la salida también es proporcional a la autocorrelación del filtro con la misma constante de proporcionalidad. Igualmente, la varianza:

$$\sigma_Y^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} R_X(t) R_h(t) dt = \frac{N_0}{2} \int_{-\infty}^{\infty} \delta(t) R_h(t) dt = \frac{N_0}{2} R_h(0) = \frac{N_0}{2} \int_{-\infty}^{\infty} |h(t)|^2 dt \quad (16)$$

Sin embargo, es fácil ver que la potencia promedio generada por el proceso $X(t)$ debería ser infinita, ya que $P = \int_{-\infty}^{\infty} S_X(f) df$. Desde el punto de vista práctico, el ruido blanco aquí descrito es irrealizable. Sin embargo, en la mayoría de los casos, la densidad espectral de ruido blanco puede aproximarse por $S_X(f) = N_0/2 \text{rect}(f/f_0)$, donde f_0 es del orden de 10^{13} Hz. Para todos efectos prácticos, $|H(f)|^2$ es muy pequeño para valores grandes de $|f|$, y como consecuencia hay una diferencia despreciable entre el valor teórico $\frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)| df$ y el valor práctico dados por la expresión $\frac{N_0}{2} \int_{-f_0}^{f_0} |H(f)| df$. Por esta razón, el modelo de ruido blanco es una aproximación aceptable en la mayoría de los casos de interés. Además, el uso de señales de tipo impulso como $\delta(t)$ simplifica notablemente el análisis de muchos problemas en sistemas de telecomunicaciones.

En la mayoría de los casos, el ruido térmico es omnipresente en los circuitos electrónicos, produciendo ruido estacionario gaussiano. El ruido blanco gaussiano es igual al descrito anteriormente, pero con la condición adicional de estar formado por variables aleatorias conjuntamente gaussianas. Sin embargo, este proceso a la entrada de un filtro cuya salida mediríamos en las condiciones anteriormente expuestas es ficticio, ya que técnicamente las variables aleatorias gaussianas deberían tener varianza infinita para encajar en el molde que se ha descrito.

4. Bibliografía recomendada

1. Capítulo 10 de A. León-García, *Probability and Random Processes for Electrical Engineering*, (2ª edición). Addison Wesley. 1994.
2. Capítulo 1, S. Haykin *Communication systems*, (4ª edición), John Wiley and Sons. 2001.
3. Capítulo 8, PZ. Peebles, *Probability, random variables and random signal principles* (2ª edición). McGraw-Hill
4. Capítulo 18 [Una aproximación con énfasis en procesos en tiempo discreto] de S. Kay, *Intuitive probability and random processes using matlab*. 2006. New York: Springer, 2006.

Tema 5: Teoría de la información

Ignacio Alvarez Illan

4 de abril de 2022

Esta obra está bajo una licencia [Creative Commons](#) “Atribución-NoComercial-CompartirIgual 3.0 No portada”.



1. Información y Entropía

Dada una variable aleatoria discreta $X(\xi)$, cuya pmf $p_X(x_i)$, $i = 1, 2, 3, \dots$ determina las probabilidades $P(X(\xi_i) = x_i)$ para todos los elementos del espacio muestral, se define la *información* de la variable X asociada al evento ξ como:

$$I(\xi) = \log \frac{1}{P[X(\xi)]} \quad (1)$$

donde la base del logaritmo no está especificada, pero frecuentemente se usará la base 2.

De esta definición se deduce que la información asociada a un evento de probabilidad 1 es nula, y es siempre una cantidad positiva debido a que las probabilidades están comprendidas entre 0 y 1. Si un evento ξ_1 es mas probable que otro ξ_2 , $P(x_1) > P(x_2)$, entonces la información asociada al primero es menor que la segunda $I(\xi_1) < I(\xi_2)$. Podemos interpretar la información como una cuantificación de la sorpresa asociada al evento ξ , o a la incertidumbre.

Conocida la pmf de la variable X , se puede determinar el valor esperado de la información para todos los posibles eventos dentro del espacio muestral. Esta cantidad $H(X)$ se denomina *entropía*:

$$H(X) = E[I(\xi)] = \sum_i p_X(x_i) \log \frac{1}{p_X(x_i)} = - \sum_i p_X(x_i) \log p_X(x_i) \quad (2)$$

Si la base del logaritmo utilizado es 2, entonces la entropía se mide en unidades de *bits*. Si logaritmo es natural mediremos en *natts*, y si la base es 10 usaremos *hartleys*. Haciendo uso de $\log_a(x) = \log_a(b) \log_b(x)$ se puede cambiar de una unidad a otra. Podríamos también pensar que la entropía es el valor esperado del logaritmo de la pmf $H(X) = E[\log \frac{1}{p_X}]$

1.1. Cota en la entropía

Consideramos una variable aleatoria discreta $X(\xi)$ que puede tomar N valores diferentes y conocemos su pmf $p_X(x_i)$, $i = 1, 2, \dots, N$. La entropía de la variable aleatoria X está acotada. Para demostrarlo usaremos el hecho¹ de que $\log(x) \leq (x - 1)$, donde la igualdad se da para $x = 1$.

¹No demostraremos explícitamente esta desigualdad, pero es fácil convencerse de que es así sin más que representar ambas funciones

Investigaremos la cantidad $\log N - H(X)$:

$$\log N - H(X) = \sum_i p_X(x_i) \log N + \sum_i p_X(x_i) \log p_X(x_i) = \quad (3)$$

$$= \sum_i p_X(x_i) \log (N p_X(x_i)) \geq \quad (4)$$

$$\geq \sum_i p_X(x_i) \left(1 - \frac{1}{N p_X(x_i)}\right) = 0 \quad (5)$$

donde hemos hecho uso de que² $\log(1/x) \geq (1-x)$ identificando x con $N p_X(x_i)$. De esta forma deducimos que la entropía de la variable X está acotada:

$$H(X) \leq \log N \quad (6)$$

la entropía es siempre menor a un valor fijado por el número de posibilidades dentro del espacio muestral, y alcanza su máximo cuando³ $p_X(x_i) = \frac{1}{N}$, es decir, cuando los resultados de los experimentos aleatorios son equiprobables.

1.2. Entropía de dos variables aleatorias

La generalización de la definición de entropía a dos variables aleatorias discretas X e Y , cuya pmf conjunta es p_{XY} , es:

$$H(X, Y) = \sum_i \sum_j p_{XY}(x_i, y_j) \log \frac{1}{p_{XY}(x_i, y_j)} \quad (7)$$

ecuación que define la *entropía conjunta* entre X e Y , donde de nuevo las unidades de medida dependerán de la base del logaritmo usada. Podemos entender la entropía conjunta de dos variables como el valor esperado del logaritmo de su pmf conjunta $H(X, Y) = -E[\log(p_{XY})]$. Es fácil ver que si X e Y son independientes:

$$H(X, Y) = H(X) + H(Y) \quad \text{si } X, Y \text{ independientes} \quad (8)$$

En el caso de independencia podemos ver que la entropía conjunta es la suma de la información promedio de cada una de las variables. En casos mas generales, cuando exista dependencia entre las variables, la información asociada a la variable X contendrá información sobre la variable Y . Una forma de cuantificar esa dependencia es a través de la entropía condicional.

Supongamos que la variable X toma un valor concreto $X = x$. Puesto que la probabilidad condicional $P(Y|X = x)$ se comporta como una probabilidad⁴, puedo calcular la información promedio $E[I(\xi)]$ asociada a la variable Y fijada la condición $X = x$:

$$H|_{X=x} = \sum_j p(y_j|X=x) \log \frac{1}{p(y_j|X=x)} \quad (9)$$

Si promediamos también para todos los posibles valores x_i que puede tomar $X = x$ obtengo la *entropía condicional* (o *equivocación*):

$$H(Y|X) = \sum_i p_X(x_i) H|_{X=x_i} = \sum_i \sum_j p_X(x_i) p(y_j|X=x_i) \log \frac{1}{p(y_j|X=x_i)} = \quad (10)$$

$$= \sum_i \sum_j p_{XY}(x_i, y_j) \log \frac{1}{p(y_j|X=x_i)} = -E[\log(p(Y|X))] \quad (11)$$

De esta definición se deduce que la entropía condicional cumple que $H(Y|X) = 0$ cuando el valor de Y queda completamente determinado por el valor de X . En cambio, $H(Y|X) = H(Y)$ cuando X e Y son independientes.

²esta desigualdad se deduce de la anterior multiplicando por -1 a ambos lados.

³en $\log(1/x) \geq (1-x)$ la igualdad se da cuando $x = 1$, que corresponde con $N p_X(x_i) = 1$

⁴suma 1, esta comprendida entre 0 y 1

1.3. Regla de la cadena para la entropía

Con las definiciones anteriormente descritas se demuestra que existe una relación entre la entropía conjunta, la entropía condicional y la entropía individual, que se conoce con el nombre de regla de la cadena para la entropía:

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (12)$$

Esta ecuación nos permite ver que la entropía conjunta de X e Y esta formada por dos terminos: la entropía de X mas la entropía condicional de $Y|X$ (o la de Y mas la de $X|Y$). La entropía condicional de $Y|X$ es, por tanto, la entropía de Y que no depende de lo que valga X .

1.4. Información mutua

Las relaciones anteriormente descritas pueden reformularse introduciendo el concepto de *información mutua*:

$$I(X, Y) = \sum_i \sum_j p_{XY}(x_i, y_j) \log \frac{p_X(x_i)p_Y(y_j)}{p_{XY}(x_i, y_j)} \quad (13)$$

La información mutua mide la distancia⁵ entre una distribución conjunta p_{XY} y el producto de sus marginales p_X y p_Y , de forma que es cero en el caso de independencia estadística. La información mutua mide la información que comparten X e Y . De la definición se deduce que:

$$H(X, Y) = H(X) + H(Y) - I(X, Y) \quad (14)$$

2. Fuentes y canales discretos

Dentro de un sistema de comunicaciones, las fuentes son los emisores de los mensajes. Una fuente S se describe como un proceso aleatorio, de forma que la fuente genera símbolos s_1, s_2, s_3, \dots en el tiempo. El conjunto de símbolos que puede generar la fuente se denomina alfabeto $\mathcal{A} = \{s_1, s_2, \dots, s_q\}$, y cada símbolo del alfabeto tiene una probabilidad asociada $P(S = s_i)$. Se pueden construir palabras o mensajes combinando varios símbolos del alfabeto.

Se llama *extensión* de orden 2 de una fuente al nuevo conjunto de símbolos generado haciendo todas las posibles combinaciones de dos elementos de la fuente original. Así, el alfabeto de la extensión de orden 2 de la fuente S sería $\mathcal{A}^2 = \{\sigma_1, \sigma_2, \dots, \sigma_{q^2}\} = \{s_1s_1, s_1s_2, \dots, s_2s_1, s_2s_2, \dots, s_{q-1}s_q, s_qs_q\}$. En general, la extensión de orden n de una fuente S representa todos los posibles mensajes (o palabras) que es posible generar con n símbolos de la fuente, y tiene un alfabeto $\mathcal{A}^n = \{\sigma_1, \sigma_2, \dots, \sigma_{q^n}\}$ con q^n elementos.

Si una fuente S emite a razón de r símbolos por unidad de tiempo, se llama *razón de transmisión* R a la cantidad de información promedio que emite la fuente por unidad de tiempo, y se define como:

$$R = r \cdot H(S) \quad \text{bits/s} \quad (15)$$

suponiendo que la entropía se mide usando el logaritmo en base 2 y el tiempo en segundos. Si se usa otra base para el logaritmo y otra unidad para el tiempo, las unidades cambiarán correspondientemente.

2.1. Fuentes Discretas y sin memoria

Una fuente discreta y sin memoria (DMS por sus siglas en inglés) es una fuente que produce símbolos que son independientes entre sí e igualmente distribuidos (i.i.d). Esta restricción permite caracterizar las extensiones de orden n de una fuente DMS. Concretamente, una extensión de orden 2 de una fuente DMS cumple que la probabilidad de emitir un símbolo σ_i formado por dos símbolos cualquiera de la fuente original s_js_k se puede determinar a partir de las probabilidades

⁵la distancia de Kullback-Leibler

de éstos como: $P[S^2 = \sigma_i] = P[S = s_j] \cdot P[S = s_k]$. Usando las propiedades de los logaritmos, se demuestra que:

$$H(S^n) = nH(S) \quad (16)$$

para una fuente DMS.

2.2. Canales Discretos

Dado un alfabeto de entrada $\mathcal{X} = \{x_1, x_2, \dots, x_q\}$ y uno de salida $\mathcal{Y} = \{y_1, y_2, \dots, y_q\}$, se modela un canal de transmisión discreto de memoria nula (DMC channel) a través del conjunto de probabilidades $P(y_i|x_j)$. Se dice de memoria nula porque la probabilidad de que llegue al receptor el símbolo y_j dado que se envió el símbolo a_i solo depende del símbolo enviado en este instante y no de los anteriores símbolos enviados ni de los símbolos anteriormente recibidos. Si los alfabetos de entrada y salida son binarios, se habla de canales binarios. Si además la probabilidad $P(y_i|x_j) = P(y_j|x_i)$, el canal es simétrico. El ejemplo más simple de canal discreto es el canal BSC, o canal discreto binario simétrico, que queda completamente especificado por la probabilidad p de emitir 0s (o 1s).

3. Códigos

Un código permite expresar un conjunto de símbolos de un alfabeto $\mathcal{A} = \{s_1, s_2, \dots, s_q\}$ en un nuevo conjunto de símbolos $\mathcal{X} = \{x_1, x_2, \dots, x_r\}$ y sus extensiones (en general $q \neq r$), formando X_1, X_2, \dots, X_q palabras. A menudo estaremos interesados en códigos binarios, $x_1 = 0$ y $x_2 = 1$, por sus aplicaciones en comunicaciones digitales. Los códigos de interés serán aquellos códigos unívocamente decodificables. Un código es unívocamente decodificable si y solo si su extensión de orden n es no singular (todas sus palabras son distintas) para todo n . Dentro de los códigos unívocos son también de interés aquellos códigos instantáneos; aquellos códigos que nunca forman palabras con secuencias prefijo. Dada una palabra $X_i = x_{i_1}x_{i_2}\dots x_{i_m}$, un prefijo es cualquier secuencia $x_{i_1}x_{i_2}\dots x_{i_j}$, con $j \leq m$.

3.1. Codificación de fuentes

Dado una fuente S que emite símbolos de un alfabeto $\mathcal{A} = \{s_1, s_2, \dots, s_q\}$ con probabilidades $\{p(s_1), p(s_2), \dots, p(s_q)\}$ y un código unívoco que lo codifica mediante palabras X_i , $i = 1, 2, \dots, q$, se define la longitud media del código como:

$$\bar{L} = \sum_{i=1}^q p(s_i) \cdot \ell_i \quad (17)$$

donde ℓ_i son las longitudes de las palabras código X_i .

Dada una fuente S , cuya entropía es $H(S)$, se define la eficiencia de un código como:

$$\eta = \frac{H(S)}{\bar{L}} \quad (18)$$

En la siguiente sección se demuestra que la eficiencia debe ser un número entre 0 y 1.

4. Primer Teorema de Shannon: Coding theorem

El primer teorema de Shannon establece los límites posibles en la compresión de datos, y establece una cota en la longitud media de las palabras código para codificación de fuentes. Concretamente, la longitud media de las palabras de un código para una fuente dada debe ser siempre mayor o igual a la entropía de esa fuente.

4.1. Demostración

Para demostrar el primer teorema de Shannon se hace uso de la cota en la entropía anteriormente expuesta y la desigualdad de Kraft.

La inecuación de Kraft afirma que: Dado una fuente S que emite símbolos de un alfabeto $\mathcal{A} = \{s_1, s_2, \dots, s_q\}$ y un código instantáneo $\{x_1, x_2, \dots, x_r\}$ que lo codifica mediante palabras X_i de longitud ℓ_i , entonces

$$k = \sum_{i=1}^q r^{-\ell_i} \leq 1$$

inecuación que se deduce de las restricciones impuestas por ser un código instantáneo. McMillan demostró que esta desigualdad también se cumple para cualquier código unívoco.

Si definimos unas cantidades q_i tal que $q_i \geq 0$ y $\sum q_i = 1$ entonces:

$$H(X) = - \sum_{i=1}^n p_i \log p_i \quad (19)$$

$$\leq - \sum_{i=1}^n p_i \log q_i \quad (20)$$

$$= - \sum_{i=1}^n p_i \log r^{-\ell_i} + \sum_{i=1}^n p_i \log \sum q_i \quad (21)$$

$$= - \sum_{i=1}^n p_i \log r^{-\ell_i} + \log \sum q_i \quad (22)$$

$$\leq - \sum_{i=1}^n -\ell_i p_i \log r \quad (23)$$

$$\leq \bar{L} \log r \quad (24)$$

$$(25)$$

donde se ha hecho uso de que $q_i = r^{-\ell_i} / \sum r^{-\ell_i}$. Si asumimos que la entropía está expresada en unidades en las que $\log r = 1$, entonces llegamos a conclusión de que la longitud media debe ser mayor o igual a la entropía tal y como establece el teorema de Shannon.

4.2. Códigos compactos

Alcanzar el límite impuesto por el primer teorema de Shannon es el objetivo de las técnicas de compactación sin pérdidas. El código óptimo para una codificación de símbolos uno a uno es el código de Huffmann, que constituye un ejemplo de código compacto. Un código compacto es aquel cuya longitud media es mínima comparada con todos los códigos unívocos posibles. Para construir un código de Huffmann se siguen los siguientes pasos:

1. Ordenar los símbolos de la fuente en orden decreciente de probabilidad
2. Combinar los dos símbolos de menor probabilidad y calcular su probabilidad asociada
3. Repetir los pasos 1 y 2 hasta que solo queden dos símbolos cuya probabilidad combinada sea igual a 1
4. etiquetar cada rama del árbol creado con 0s y 1s. Leer el código generado para cada símbolo de derecha a izquierda.

5. Segundo Teorema de Shannon: Noisy channel theorem

Con la caracterización de un canal de comunicación descrita en la sección 2.2 es posible formular el primer teorema de Shannon de una forma alternativa que nos prepara para formular el segundo teorema de Shannon.

Se define la *capacidad* de un canal como:

$$C = \sup_{p(x)} I(X; Y) \quad (26)$$

donde sup se toma como aquella distribución de probabilidad de los datos de entrada que hace máxima la información mutua entre la entrada y la salida. La capacidad se mide en bits por input de entrada. De esta forma, si transmito un input por segundo a través de un canal de capacidad C , podré transmitir información a una razón de C bits/s.

Por otro lado, la capacidad operacional se define como la tasa en bits/s que se puede transmitir por un canal con una probabilidad de error arbitrariamente pequeña. El segundo teorema de Shannon establece que la capacidad de un canal es igual a la capacidad operacional de este.

5.1. Canal sin ruido

Un canal sin ruido es aquel que solo tiene un elemento por columna distinto de cero en la matriz de canal $p(y_j|x_i)$, de forma que si el número de símbolos de entrada es igual al número de símbolos de salida entonces la matriz de canal es diagonal. Es fácil calcular la equivocación para un canal sin ruido $H(Y|X)$, que resulta ser igual a 0. En otras palabras, no hay incertidumbre asociada a la variable recibida Y una vez enviada X .

Según la definición de capacidad, en el caso de canal sin ruido se reduce a:

$$C = \sup_{p(x)} I(X; Y) = \sup_{p(x)} H(Y) - H(Y|X) = \sup_{p(x)} H(Y) \quad (27)$$

es decir, la capacidad de un canal sin ruido viene determinada por la $p(x)$ de los datos de entrada que hace máxima la entropía a la salida. Se puede ver que en este caso la entropía de entrada coincide con la entropía de salida. En este caso, la tasa de información máxima que se puede transmitir por un canal sin ruido con una probabilidad arbitrariamente pequeña de error viene determinada por la entropía de la fuente.

5.2. Canal con ruido

En un canal con ruido, la tasa de información que se puede transmitir a través de ese canal viene de nuevo determinada por la capacidad del canal, que en este caso depende de la información mutua entre la entrada y la salida.

6. Bibliografía recomendada

1. Norman Abrahamson, *Teoría de la información y la codificación*, (6ª edición). Paraninfo 1986.
2. Capítulos 9 y 10, S. Haykin *Communication systems*, (4ª edición), John Wiley and Sons. 2001.
3. James V. Stone, *Information theory: a tutorial introduction* Sebtel Press 2018.