

III.1. Variable Aleatoria Conjunta

May 13, 2019

1 Variable Aleatoria Conjunta

1.1 Introducción

Consideremos el **espacio de probabilidad** $\mathcal{E}(\Omega, \mathcal{F}, P)$. Sobre él podemos definir no solo una, como ya hemos hecho, sino varias **variables aleatorias** X_1, X_2, \dots, X_N :

$$\begin{aligned} X_n &: \longmapsto \mathbb{R} \\ \alpha &\longmapsto X_n(\alpha) \end{aligned}$$

Cada una de estas variables aleatorias puede ser discreta, continua o mixta. Al estar definidas sobre el mismo espacio de probabilidad, los distintos sucesos descritos por cada una de las variables aleatorias tendrán, en general una dependencia probabilística entre sí, y los valores de estas variables tendrán entre sí relaciones no deterministas. Caben dos casos extremos:

- **Dependencia funcional:** en cuyo caso las variables aleatorias satisfacen una ecuación sin componente aleatoria alguna
- **Independencia:** si por cómo están construidas las variables aleatorias los sucesos que se representan con cada una de ellas son independientes dos a dos con los de todas las demás. En tal caso, **los valores de unas variables aleatorias no ofrecen información alguna sobre las restantes.**

El espacio aleatorio sobre el que se definen las variables aleatorias puede corresponder a un experimento compuesto (recordar lo visto en el Bloque I), de forma que cada una de las variables aleatorias se define sobre cada uno de los espacios que lo componen.

$$\mathcal{E}(\Omega, \mathcal{F}, P) = \mathcal{E}_1(\Omega_1, \mathcal{F}_1, P_1) \times \dots \times \mathcal{E}_N(\Omega_N, \mathcal{F}_N, P_N)$$

$$\begin{aligned} X_n &: \longmapsto \mathbb{R} \\ \alpha &\longmapsto X_n(\alpha) \end{aligned}$$

Por supuesto, si los experimentos marginales (componentes del experimento compuesto) son independientes, también lo serán las variables aleatorias definidas sobre cada uno de ellos.

Adviértase, no obstante, que no es necesario definir varias variables aleatorias sobre experimentos compuestos.

Consideremos algunos ejemplos:

- En el lanzamiento de un dado, número de cara (X_1) y su doble (X_2)
- Lanzamiento simultáneo de una moneda X_1 y un dado X_2
- Lanzamiento sucesivo de una moneda cinco veces: X_1, X_2, \dots, X_5
- Lanzamiento de un dado (X_1) e instante de llamada (X_2) de mis padres entre 22h y 23h (con certeza de que lo harán en esa franja)
- Dos muestras sucesivas, espaciadas 1 mseg: X_1 y X_2 , de la señal de voz del profesor dando la clase
- El valor de cierre del índice Ibex cinco días seguidos: X_1, X_2, \dots, X_5
- Estatura (X_1), m peso X_2 y edad en años X_3 de una persona cogida al azar
- En el lanzamiento de un dado, cara (X_1) y acierto si es par (X_2)

Reflexionar sobre los siguientes aspectos:

- Carácter continuo, discreto o mixto de las variables aleatorias
- Dependencia probabilística, funcional e independencia

1.1.1 Vectores aleatorios

Es habitual reunir todas las variables aleatorias como componentes de un vector aleatorio n -dimensional.

$$\mathbf{X} = (X_1, \dots, X_N)^T$$

Cada una de las variables aleatorias X_n , componentes del vector aleatorio \mathbf{X} , tiene una **caracterización probabilística marginal**, es decir, una a una por separado, descrita como se ha visto en los temas de variable aleatoria:

- Funciones de distribución de probabilidad $F_{X_n}(x)$, que pueden expresarse como tablas con las funciones acumulada de probabilidad en el caso discreto, $F_{X_n}(x_i)$
- Funciones de densidad de probabilidad $f_{X_n}(x)$, que pueden expresarse mediante funciones de masa de probabilidad $f_{X_n}(x_i)$ en el caso discreto
- Momentos de orden k no centrados, $E(X_n^k)$, y centrados en la media, $E((X_n - \eta_{X_n})^k)$. En particular:
 - Media: $\eta_{X_n} = E(X_n)$
 - Valor cuadrático medio: $E(X_n^2)$
 - Varianza: $\sigma_{X_n}^2 = \text{Var}(X_n) = E((X_n - \eta_{X_n})^2) = E(X_n^2) - \eta_{X_n}^2$

El objetivo de este tema es, como veremos, considerar la **caracterización probabilística conjunta** del vector aleatorio \mathbf{X} o, dicho de otro modo, de las variables aleatorias X_n .

1.1.2 Secuencias aleatorias y procesos estocásticos

Un caso interesante es cuando el número de variables aleatorias crece de forma indefinida y ya no es posible representarlo mediante un vector de dimensión finito, dando lugar a una **secuencia aleatoria**, $X[n]$. Se trata de una secuencia indefinida, en general sin comienzo y sin final, de muestras aleatorias, cada una de ellas representable mediante una variable aleatoria.

Si en vez de una secuencia de muestras aleatorias consideramos una función que toma valores aleatorios, no sobre un índice discreto sino sobre todo el eje real (por ejemplo, un eje de tiempo), estamos ante un **proceso estocástico**, $X(t)$, en el que el valor que toma la función en cada instante

es una variable aleatoria. Estrictamente, un proceso estocástico asigna a cada resultado de un experimento aleatorio una función real de variable real del conjunto $\mathcal{F}(\mathbb{R}, \mathbb{R})$:

$$\begin{aligned} X &: \mapsto \mathcal{F}(\mathbb{R}, \mathbb{R}) \\ \alpha &\mapsto X(t, \alpha) \end{aligned}$$

Adviértase que, si se fija el resultado α estamos ante una función determinista mientras que, si se fija el instante t estamos ante una variable aleatoria.

Para cada índice de la muestra o instante de tiempo puede hacerse una **caracterización marginal de la secuencia o proceso**:

$$F_X(x; n) \quad f_X(x; n) \quad F_X(x; t) \quad f_X(x; t)$$

Igualmente con los **momentos**. Por ejemplo:

$$\eta_X[n] = E(X[n]) \quad \sigma_X^2[n] = \text{Var}(X[n]) = E(X^2[n]) - \eta_X^2[n]$$

$$\eta_X(t) = E(X(t)) \quad \sigma_X^2(t) = \text{Var}(X(t)) = E(X^2(t)) - \eta_X^2(t)$$

Secuencias y procesos estacionarios Si la caracterización probabilística de la secuencia o proceso es **temporalmente invariante**, esto es, si tal caracterización no depende de cuándo se establece el origen de tiempos o, alternativamente, de desplazar temporalmente la secuencia o el proceso, estamos ante una **secuencia o proceso estacionario en sentido estricto**. Dicho de otra manera, las caracterizaciones probabilísticas de $X(t)$ y de $X(t + t_0)$ son idénticas para cualquier $t_0 \in \mathbb{R}$. Igualmente pasa con $X[n]$ y $X[n + n_0]$ para cualquier $n_0 \in \mathbb{Z}$.

Lógicamente la estacionariedad en sentido estricto conlleva que las funciones de distribución de probabilidad, las funciones de densidad de probabilidad y, en el caso discreto, las funciones de masa de probabilidad, así como todos los momentos, no varían a lo largo del tiempo o el índice de la muestra

$$F_X(x; n) \equiv F_X(x) \quad f_X(x; n) \equiv f_X(x) \quad F_X(x; t) \equiv F_X(x) \quad f_X(x; t) \equiv f_X(x)$$

$$\eta_X[n] \equiv \eta_X \quad \sigma_X^2[n] \equiv \sigma_X^2 \quad \eta_X(t) \equiv \eta_X \quad \sigma_X^2(t) \equiv \sigma_X^2$$

Secuencias y procesos ergódicos En principio, la estimación de parámetros estadísticos de secuencias y procesos aleatorios requiere promediar en los resultados del espacio probabilístico suya-
cente. Por ejemplo, si N es el número de realizaciones, la media de un proceso $X(t)$ será:

$$\eta_X(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\alpha_i \in \Omega} X(t, \alpha_i)$$

Frecuentemente esto no resulta práctico, pues no es posible repetir muchas veces el experimento para obtener cada una de las funciones y promediarlas.

Sin embargo, **para muchas secuencias o procesos estacionarios** es posible estimar los parámetros estadísticos a partir de una única realización. En tal caso la secuencia o proceso es **ergódica**.

La idea es que es posible (no siempre) obtener tales estimaciones promediando a lo largo del tiempo o índice, en vez de hacerlo a lo largo de las distintas realizaciones. La ergodicidad es una

propiedad muy útil que vincula la estadística con las secuencias y procesos estacionarios. Pero es también matemáticamente compleja, y no vamos a profundizar en ella.

Por ejemplo, la media de un proceso estacionario y ergódico podría calcularse:

$$\eta_X = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t) dt \quad \eta_X = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{-N}^N X[n]$$

En términos prácticos, si la secuencia o proceso es ergódico, podemos muestrearlo temporalmente para estimar sus parámetros estadísticos.

Señales aleatorias y ruido Cuando todas las variables aleatorias correspondientes tanto a las muestras de la secuencia aleatoria como a los instantes del proceso estocástico son **independientes** entre sí, decimos que la secuencia o el proceso es **blanco**. Una señal o secuencia aleatoria e indeseada suele denominarse **ruido**. El **ruido blanco** es una secuencia aleatoria, $N[n]$, o proceso estocástico, $N(t)$, en el que todas las variables aleatorias de las muestras o instantes son independientes entre sí.

En ingeniería de telecomunicación y electrónica, así como en otras áreas de la ciencia y la técnica, es muy habitual considerar procesos estocásticos que modelan **señales en ruido**, donde $S(t)$ puede ser determinista desconocida o aleatoria, según sea el modelo que se aplique, respectivamente, clásico (basado en la *verosimilitud*) o bayesiano. El ruido es, por definición, siempre aleatorio:

$$X(t) = S(t) + N(t)$$

Al muestrear tales procesos, con $X[n] = S(n\Delta T_s)$, siendo T_s el periodo de muestreo, obtenemos secuencias aleatorias de señal en ruido:

$$X[n] = S[n] + N[n]$$

Es también habitual trabajar sólo con las muestras correspondientes a un intervalo, no con toda la secuencia. En tales casos, se trabaja con un número finito y predeterminado de muestras que se ordenan en vectores aleatorios de dimensión N (no confundir con el ruido), resultando el siguiente modelo:

$$\mathbf{X} = \mathbf{S} + \mathbf{N}$$

Los problemas más habituales ante los que nos encontramos son los siguientes:

- **Filtrado** consiste en estimar la señal S a partir de las medidas ruidosas X .
- **Clasificación** consiste en identificar la señal S_i entre un conjunto finito de señales posibles a partir de medidas ruidosas.
- **Predicción** consiste en estimar la señal S en un instante o muestra futuro a partir de medidas ruidosas anteriores.
- **Interpolación** consiste en estimar la señal S en instantes en los que no disponemos de medida a partir de medidas ruidosas tanto anteriores como posteriores

En todos los casos, la señal S puede considerarse determinista o aleatoria, dando lugar a métodos clásicos o bayesianos para resolver los problemas mencionados.

1.2 Dos variables aleatorias o vector aleatorio bidimensional

Ya hemos visto la **caracterización marginal** de vectores y secuencias aleatorias y de procesos estocásticos. Sin embargo esto solo muestra vistas parciales de los mismos, una por cada variable aleatoria que podemos extraer. Hemos de añadir el conocimiento de la dependencia probabilística entre las variables aleatorias que los componen, lo que nos lleva a la **caracterización conjunta de vectores y secuencias aleatorias y de procesos estocásticos**. Es ilustrativo para ello estudiar en detalle el caso de dos variables aleatorias o vector aleatorio bidimensional, $\mathbf{V} = (X, Y)^t$, que tiene una fácil representación gráfica en el plano mediante un **diagrama de dispersión** de sus realizaciones.

Ya sabemos que las variables aleatorias, X e Y las definimos sobre un espacio de probabilidad subyacente $\mathcal{E}(\Omega, \mathcal{F}, P)$. Podemos definir sobre el plano $X \times Y$ conjuntos que serán sucesos de \mathcal{F} . Por ejemplo:

- $\{X \leq x\}$ es el semiplano que se extiende a la izquierda de la línea vertical $X = x$, y es un suceso de \mathcal{F} .
- $\{Y \leq y\}$ es el semiplano que se extiende por debajo de la línea horizontal $Y = y$, y también un suceso de \mathcal{F} .
- $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$ es, por tanto, el cuadrante que queda a la izquierda de la línea vertical $X = x$ y por debajo de la línea horizontal $Y = y$, y es también un suceso de \mathcal{F} .

Adviértase que una variable aleatoria proporciona, por lo general, una **descripción parcial del espacio de probabilidad** subyacente, que, cuando tenemos definidas varias, llamamos **caracterización marginal**. Esto resulta evidente sin más que pensar, por ejemplo, en un experimento compuesto, al que asignamos una variable aleatoria diferente a cada experimento componente.

1.2.1 Función de Distribución Conjunta de Probabilidad

Considérese un vector aleatorio (X, Y) o, equivalentemente, dos variables aleatorias, X e Y . Sus **funciones de distribución marginales** son:

- $F_X(x) = P(\{X \leq x\}) \quad -\infty \leq x \leq \infty$
- $F_Y(y) = P(\{Y \leq y\}) \quad -\infty \leq y \leq \infty$

El conocimiento de F_X y F_Y no es, en general, suficiente para poder obtener las probabilidades de sucesos del tipo $\{X \leq x, Y \leq y\}$. Necesitamos para ello conocer una función diferente, la **función de distribución conjunta de probabilidad**, $F_{XY}(x, y)$ del vector aleatorio (X, Y) es:

$$F_{XY}(x, y) = P(\{X \leq x, Y \leq y\})$$

Podemos considerar que $F_{XY}(x, y)$ proporciona la **masa de probabilidad acumulada en el cuadrante** que queda a la izquierda de la línea vertical $X = x$ y por debajo de la línea horizontal $Y = y$

Las principales propiedades de $F_{XY}(x, y)$ son las siguientes, y se entienden intuitivamente sin más que considerar cómo se distribuye la masa de probabilidad en el plano:

- $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0 \quad \forall x, y \in \mathbb{R}$, pues no hay masa de probabilidad acumulada
- $F_{XY}(\infty, \infty) = 1$, pues corresponde a la totalidad de la masa de probabilidad
- $F_{XY}(x, y)$ es **no decreciente** tanto en x como y . EN los puntos donde se localicen masas discretas de probabilidad F_{XY} muestra discontinuidades de salto. En el resto, es continua.

- El suceso $\{x_1 < X \leq x_2, Y \leq y\}$ es una franja vertical limitada por $X = x_1$ (sin incluir) y $X = x_2$ (incluido) cuya masa de probabilidad es $F_{XY}(x_2, y) - F_{XY}(x_1, y)$
- El suceso $\{X \leq x, y_1 < Y \leq y_2\}$ es una franja horizontal limitada por $Y = y_1$ (sin incluir) e $Y = y_2$ (incluido) cuya masa de probabilidad es $F_{XY}(x, y_2) - F_{XY}(x, y_1)$
- El suceso $\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$ es un rectángulo limitado por $X = x_1$ (sin incluir), $X = x_2$ (incluido), $Y = y_1$ (sin incluir) e $Y = y_2$ (incluido), cuya masa de probabilidad es $F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$

Si X e Y son variables aleatorias discretas, puede trabajarse con una tabla con la función de probabilidad acumulada: $F_{XY}(x_i, y_j) \quad i, j \in \mathbb{Z}$

Relación entre las funciones de distribución marginales y la conjunta Es inmediato obtener las funciones de distribución marginales a partir de la conjunta, sin más que entender que la ausencia de una variable aleatoria significa que puede darse cualquiera de ellos y, por tanto, debe cubrirse por completo el eje correspondiente:

- $F_X(x) = F_{XY}(x, \infty)$
- $F_Y(y) = F_{XY}(\infty, y)$

Sin embargo, como ya se ha dicho, **no es posible, en general, obtener la distribución conjunta a partir de las marginales**. La excepción es cuando **ambas variables aleatorias son independientes** pues, en tal caso:

$$F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\})P(\{Y \leq y\}) = F_X(x)F_Y(y)$$

Por tanto, si X e Y son **variables aleatorias independientes**: $F_{XY}(x, y) = F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R}$.

1.2.2 Función de masa de probabilidad conjunta

Concentrémonos ahora en el caso en que tanto X como Y son discretas. Ya hemos visto que la función de distribución conjunta podemos expresarla mediante una tabla de probabilidades acumuladas indexada por las posiciones donde se localizan las masa de probabilidad. Es conveniente repasar la forma de hacerlo con una sola variable aleatoria.

$$F_{XY}(x_i, y_j) \quad i, j \in \mathbb{Z}$$

Podemos fácilmente obtener las masa de probabilidad en cada localización, e introducir con ello la **función de masa de probabilidad conjunta** del vector aleatorio (X, Y) o de ambas variables aleatorias X e Y :

$$p_{XY}(x_i, y_j) \equiv P(X = x, Y = y) = F_{XY}(x_i, y_j) - F_{XY}(x_{i-1}, y_j) - F_{XY}(x_i, y_{j-1}) + F_{XY}(x_{i-1}, y_{j-1})$$

La **tabla de probabilidades acumuladas** correspondiente a la función de distribución conjunta se obtiene a partir de la función de masa de probabilidad conjunta:

$$F_{XY}(x_i, y_j) = \sum_{m=-\infty}^i \sum_{n=-\infty}^j p_{XY}(x_m, y_n)$$

Es posible obtener una expresión recursiva para $F_{XY}(x_i, y_j)$ sin más que despejar:

$$F_{XY}(x_i, y_j) = (F_{XY}(x_{i-1}, y_j) + F_{XY}(x_i, y_{j-1}) - F_{XY}(x_{i-1}, y_{j-1})) + p_{XY}(x_i, y_j)$$

Adviértase que, por tratarse de probabilidades, $p_{XY}(x_i, y_j) \geq 0$ y, también:

$$F_{XY}(\infty, \infty) = 1 \implies \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} p_{XY}(x_i, y_j) = 1$$

Relación entre las funciones de masa de probabilidad marginales y la conjunta Considerando que $F_X(x_i) = F_{XY}(x_i, \infty)$ y que $p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$, pueden obtenerse las funciones de masa de probabilidad marginales a partir de la conjunta:

$$p_X(x_i) = \sum_{m=-\infty}^i \sum_{n=-\infty}^{\infty} p_{XY}(x_m, y_n) - \sum_{m=-\infty}^{i-1} \sum_{n=-\infty}^{\infty} p_{XY}(x_m, y_n) = \sum_{j=-\infty}^{\infty} p_{XY}(x_i, y_j)$$

Igualmente,

$$p_Y(y_j) = \sum_{i=-\infty}^{\infty} p_{XY}(x_i, y_j)$$

En general, **la función de masa de probabilidad conjunta no puede obtenerse a partir de las funciones marginales**. Sin embargo, **si ambas variables son independientes**:

$$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j) = p_X(x_i)p_Y(y_j)$$

Ejemplo: caracterización conjunta de dos variables aleatorias discretas independientes Considérese el lanzamiento simultáneo de un dado trucado y de una moneda trucada. Representamos con la variable aleatoria X_1 los resultados del dado y con X_2 los de la moneda (0 es cara y 1 es cruz). Las funciones de masa de probabilidad marginales son como sigue:

X_1	x_{1i}	1	2	3	4	5	6	
	$p_{X_1}(x_i)$	$\frac{1}{24}$	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{6}{24}$	$\sum_i p_X(x_i) = 1$

X_2	x_{2i}	0	1	
	$p_{X_2}(x_{2i})$	$\frac{11}{24}$	$\frac{13}{24}$	$\sum_i p_X(x_i) = 1$

Considerando que ambos lanzamientos son independientes, resulta la siguiente función de masa de probabilidad conjunta. Téngase en cuenta que, para cualesquiera otros valores de las variables aleatorias no presentes en la tabla, la masa de probabilidad es cero.

$p_{X_1 X_2}(x_{1i}, x_{2j})$		0	1	$p_{X_1}(x_{1i}) = \sum_j p_{X_1 X_2}(x_{1i}, x_{2j})$
$X_2 X_1$	1	$\frac{1}{24} \frac{11}{24} = \frac{11}{24^2}$	$\frac{1}{24} \frac{13}{24} = \frac{13}{24^2}$	$\frac{1}{24}$
	2	$\frac{5}{24} \frac{11}{24} = \frac{55}{24^2}$	$\frac{5}{24} \frac{13}{24} = \frac{65}{24^2}$	$\frac{5}{24}$
	3	$\frac{5}{24} \frac{11}{24} = \frac{55}{24^2}$	$\frac{5}{24} \frac{13}{24} = \frac{65}{24^2}$	$\frac{5}{24}$
	4	$\frac{3}{24} \frac{11}{24} = \frac{33}{24^2}$	$\frac{3}{24} \frac{13}{24} = \frac{39}{24^2}$	$\frac{3}{24}$
	5	$\frac{4}{24} \frac{11}{24} = \frac{44}{24^2}$	$\frac{4}{24} \frac{13}{24} = \frac{52}{24^2}$	$\frac{4}{24}$
	6	$\frac{6}{24} \frac{11}{24} = \frac{66}{24^2}$	$\frac{6}{24} \frac{13}{24} = \frac{78}{24^2}$	$\frac{6}{24}$
$p_{X_2}(x_{2j}) = \sum_i p_{X_1 X_2}(x_{1i}, x_{2j})$		$\frac{11}{24}$	$\frac{13}{24}$	1

La tabla con la función acumulada de probabilidad se obtiene de forma inmediata con la expresión recursiva ya vista:

$$F_{X_1 X_2}(x_{1_i}, x_{2_j}) = \left(F_{X_1 X_2}(x_{1_{i-1}}, x_{2_j}) + F_{X_1 X_2}(x_{1_i}, x_{2_{j-1}}) - F_{X_1 X_2}(x_{1_{i-1}}, x_{2_{j-1}}) \right) + p_{X_1 X_2}(x_{1_i}, x_{2_j})$$

Aunque, por supuesto, también podemos simplemente acumular probabilidades:

$$F_{X_1 X_2}(x_{1_i}, x_{2_j}) = \sum_{m=-\infty}^i \sum_{n=-\infty}^j p_{X_1 X_2}(x_{1_m}, x_{2_n})$$

Adviértase también que las funciones de distribución marginales se obtienen también de la tabla:

- $F_{X_1}(x_{1_i}) = F_{X_1 X_2}(x_{1_i}, \infty) = F_{X_1 X_2}(x_{1_i}, 1) = F_{X_1 X_2}(x_{1_i}, 2) \dots$
- $F_{X_2}(x_{2_j}) = F_{X_1 X_2}(\infty, x_{2_j}) = F_{X_1 X_2}(6, x_{2_j}) = F_{X_1 X_2}(7, x_{2_j}) \dots$

$F_{X_1 X_2}(x_{1_i}, x_{2_j})$		-1	0	1	2
$X_2 X_1$	0	0	0	0	0
	1	0	$\frac{11}{24^2}$	$\frac{11}{24^2} + \frac{13}{24^2} = \frac{24}{24^2}$	$\frac{24}{24^2} = \frac{1}{24}$
	2	0	$\frac{11}{24^2} + \frac{55}{24^2} = \frac{66}{24^2}$	$\left(\frac{24}{24^2} + \frac{66}{24^2} - \frac{11}{24^2} \right) + \frac{65}{24^2} = \frac{144}{24^2}$	$\frac{144}{24^2} = \frac{6}{24}$
	3	0	$\frac{66}{24^2} + \frac{55}{24^2} = \frac{121}{24^2}$	$\left(\frac{144}{24^2} + \frac{121}{24^2} - \frac{66}{24^2} \right) + \frac{65}{24^2} = \frac{264}{24^2}$	$\frac{264}{24^2} = \frac{11}{24}$
	4	0	$\frac{121}{24^2} + \frac{33}{24^2} = \frac{154}{24^2}$	$\left(\frac{264}{24^2} + \frac{154}{24^2} - \frac{121}{24^2} \right) + \frac{39}{24^2} = \frac{336}{24^2}$	$\frac{336}{24^2} = \frac{14}{24}$
	5	0	$\frac{154}{24^2} + \frac{44}{24^2} = \frac{198}{24^2}$	$\left(\frac{336}{24^2} + \frac{198}{24^2} - \frac{154}{24^2} \right) + \frac{52}{24^2} = \frac{432}{24^2}$	$\frac{432}{24^2} = \frac{18}{24}$
	6	0	$\frac{198}{24^2} + \frac{66}{24^2} = \frac{264}{24^2}$	$\left(\frac{432}{24^2} + \frac{264}{24^2} - \frac{198}{24^2} \right) + \frac{78}{24^2} = \frac{576}{24^2}$	$\frac{576}{24^2} = \frac{24}{24} = 1$
	7	0	$\frac{264}{24^2} = \frac{11}{24}$	$\frac{576}{24^2} = \frac{24}{24} = 1$	1

Caracterización conjunta de dos variables aleatorias discretas (caso general) Consideremos ahora que lanzamos un dado y una moneda, trucados de tal manera que hay dependencia entre sus resultados. En este caso nos tienen que dar tal dependencia en forma de una tabla con la función de masa de probabilidad conjunta, a partir de la cual podemos obtener las marginales (pero no al revés):

$p_{X_1 X_2}(x_{1_i}, x_{2_j})$		0	1	$p_{X_1}(x_{1_i}) = \sum_j p_{X_1 X_2}(x_{1_i}, x_{2_j})$
$X_2 X_1$	1	$\frac{1}{24}$	0	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{5}{24}$
	3	$\frac{1}{24}$	$\frac{4}{24}$	$\frac{5}{24}$
	4	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{3}{24}$
	5	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{4}{24}$
	6	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{6}{24}$
$p_{X_2}(x_{2_j}) = \sum_i p_{X_1 X_2}(x_{1_i}, x_{2_j})$		$\frac{11}{24}$	$\frac{13}{24}$	1

Adviértase que se ha mantenido para el ejemplo las mismas funciones de masa de probabilidad marginales que en el ejemplo anterior, si bien la función de masa de probabilidad conjunta es diferente.

La función de masa de probabilidad acumulada es ahora:

$F_{X_1 X_2}(x_{1_i}, x_{2_j})$		0	1	$F_{X_1}(x_{1_i}) = F_{X_1 X_2}(x_{1_i}, \infty)$
$X_2 X_1$	1	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
	2	$\frac{3}{24}$	$\frac{6}{24}$	$\frac{6}{24}$
	3	$\frac{4}{24}$	$\frac{11}{24}$	$\frac{11}{24}$
	4	$\frac{6}{24}$	$\frac{14}{24}$	$\frac{14}{24}$
	5	$\frac{7}{24}$	$\frac{18}{24}$	$\frac{18}{24}$
	6	$\frac{11}{24}$	$\frac{24}{24} = 1$	1
$F_{X_2}(x_{2_j}) = F_{X_1 X_2}(\infty, x_{2_j})$		$\frac{11}{24}$	1	

1.2.3 Función de densidad de probabilidad conjunta

Consideremos ahora que las variables aleatorias X e Y son continuas. Por igual razón que en el caso de una única variable aleatoria, en general, la masa de probabilidad en un punto es nula, $P(X = x, Y = y) = 0$. Ello hace necesario recurrir a densidades de probabilidad, generalizando lo visto en el caso de una variable aleatoria. La función de densidad de probabilidad conjunta del vector aleatorio (X, Y) , o de ambas variables aleatorias X e Y , se define como sigue a partir de la función de distribución de probabilidad conjunta $F_{XY}(x, y)$:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \geq 0$$

que es positiva o nula por ser $F_{XY}(x, y)$ no decreciente. Además:

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(u, v) du dv$$

Si se ponen la derivadas parciales como límites, de forma semejante a como se hizo con una sola variable aleatoria:

$$\begin{aligned} f_{XY}(x, y) &= \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = \\ &= \lim_{\Delta x \Delta y \rightarrow 0} \frac{F_{XY}(x + \Delta x/2, y + \Delta y/2) - F_{XY}(x - \Delta x/2, y + \Delta y/2) - F_{XY}(x + \Delta x/2, y - \Delta y/2) + F_{XY}(x - \Delta x/2, y - \Delta y/2)}{\Delta x \Delta y} \\ &= \lim_{\Delta x \Delta y \rightarrow 0} \frac{P(\{(x - \Delta x/2) < x \leq (x + \Delta x/2)\} \cap \{(y - \Delta y/2) < y \leq (y + \Delta y/2)\})}{\Delta x \Delta y} \end{aligned}$$

Donde se hace uso de un suceso rectangular infinitesimal centrado en $X = x, Y = y$. Por tanto:

$$P(\{(x - \Delta x/2) < x \leq (x + \Delta x/2)\} \cap \{(y - \Delta y/2) < y \leq (y + \Delta y/2)\}) \approx f_{XY}(x, y) \Delta x \Delta y$$

Puede asimismo calcularse la masa de probabilidad continua en un dominio D arbitrario:

$$P\{(x, y) \in D\} = \int \int_D f_{XY}(x, y) dx dy$$

De la definición, se obtienen fácilmente las **funciones de densidad marginales**:

$$f_X(x) = \frac{\partial F_{XY}(x, \infty)}{\partial x} = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

$$f_Y(y) = \frac{\partial F_{XY}(\infty, y)}{\partial y} = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

Además, como siempre, la masa total de probabilidad es la unidad:

$$F_{XY}(\infty, \infty) = 1 \implies \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$$

1.2.4 Distribuciones condicionadas

La función de distribución conjunta puede resultar condicionada por un suceso M , sin más que extender lo visto para una variable aleatoria:

$$F_{XY}(x, y|M) = P(\{X \leq x, Y \leq y\} | M) = \frac{P(\{X \leq x, Y \leq y\} \cap M)}{P(M)}$$

Si las variables aleatorias X e Y son discretas, la función de masa de probabilidad conjunta condicionada por M es:

$$p_{XY}(x_i, y_j|M) = \frac{P(\{X = x_i, Y = y_j\} \cap M)}{P(M)}$$

Adviértase que $\sum_i \sum_j p_{XY}(x_i, y_j|M) = 1$

Y en caso de que las variables aleatorias fueran continuas, la función de densidad de probabilidad conjunta condicionada es:

$$f_{XY}(x, y|M) = \frac{\partial^2 F_{XY}(x, y|M)}{\partial x \partial y}$$

Podemos también definir la probabilidad de un suceso M condicionada por un valor arbitrario (x_i, y_j) del vector aleatorio (X, Y) . Para ello no hay más que advertir que $\{(X = x_i, Y = y_j)\}$ es un suceso del espacio muestral:

$$P(M | \{X = x_i, Y = y_j\}) = \frac{P(M \cap \{X = x_i, Y = y_j\})}{P(\{X = x_i, Y = y_j\})} = \frac{P(M \cap \{X = x_i, Y = y_j\})}{p_{XY}(x_i, y_j)}$$

Mantendremos el habitual abuso de notación $P(M|x_i, y_j) \equiv P(M|X = x_i, Y = y_j)$. Adviértase que $P(M|x_i, y_j) \neq p_{XY}(x_i, y_j|M)$, como resulta del Teorema de Bayes.

En el caso de variables aleatorias continuas, cuando el suceso condicionante es del tipo $\{X = x, Y = y\}$ es necesario considerar un suceso rectangular infinitesimal centrado en dicho punto, pues la masa de probabilidad puntual es nula. Las expresiones resultan semejantes a las del caso discreto, pero utilizando funciones de densidad en vez de funciones de masa de probabilidad.

En particular, es posible obtener las masas de probabilidad de una variable aleatoria discreta condicionada por la otra:

$$p_{X|Y}(x_i|y_j) = \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)}$$

$$p_{Y|X}(y_j|x_i) = \frac{p_{XY}(x_i, y_j)}{p_X(x_i)}$$

Adviértase que las probabilidades condicionadas proporcionan verosimilitudes dadas las observaciones. En particular:

- Si la variable cuyos valores se observan es X la verosimilitud de los valores de Y es $L(y_j) = p_{X|Y}(x_i|y_j)$
- Si la variable cuyos valores se observan es Y la verosimilitud de los valores de X es $L(x_i) = p_{Y|X}(y_j|x_i)$

Por supuesto, en general $p_{X|Y}(x_i|y_j) \neq p_{Y|X}(y_j|x_i)$, estando relacionados por el Teorema de Bayes.

De forma semejante, si las variables aleatorias son continuas, las funciones de densidad de probabilidad condcionadas resultan:

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

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

Igual que antes, las funciones de probabilidad condicionadas proporcionan verosimilitudes dadas las observaciones. En particular:

- Si la variable cuyos valores se observan es X la verosimilitud de los valores de Y es $L(y) = p_{X|Y}(x|y)$
- Si la variable cuyos valores se observan es Y la verosimilitud de los valores de X es $L(x) = p_{Y|X}(y|x)$

De nuevo, en general $f_{X|Y}(x/y) \neq f_{Y|X}(y/x)$, estando ambas relacionadas por el Teorema de Bayes.

Ejemplo Tomando las $p_{X_1X_2}(x_i, y_j)$ y $p_{X_1}(x_{1_i})$ del ejemplo anterior obtenemos $p_{X_2|X_1}(x_{2_j}|x_{1_i})$:

$p_{X_2 X_1}(x_{2_j} x_{1_i})$		0	1	$\sum_j p_{X_1X_2}(x_{1_i}, x_{2_j})$
$X_2 X_1$	1	$\frac{1}{24} = 1$	$\frac{0}{1} = 0$	1
	2	$\frac{2}{24} = \frac{2}{5}$	$\frac{3}{24} = \frac{3}{5}$	1
	3	$\frac{1}{24} = \frac{1}{5}$	$\frac{4}{24} = \frac{4}{5}$	1
	4	$\frac{2}{24} = \frac{2}{3}$	$\frac{1}{24} = \frac{1}{3}$	1
	5	$\frac{1}{24} = \frac{1}{4}$	$\frac{3}{24} = \frac{3}{4}$	1
	6	$\frac{4}{24} = \frac{4}{6}$	$\frac{2}{24} = \frac{2}{6}$	1

Igualmente obtenemos $p_{X_1|X_2}(x_{1_i}|x_{2_j})$ a partir de $p_{X_1X_2}(x_i, y_j)$ y $p_{X_2}(x_{2_j})$:

$p_{X_1 X_2}(x_{1_i} x_{2_j})$	0	1
1	$\frac{\frac{1}{24}}{\frac{11}{24}} = \frac{1}{11}$	$\frac{\frac{0}{13}}{\frac{24}{24}} = 0$
2	$\frac{\frac{2}{24}}{\frac{11}{24}} = \frac{2}{11}$	$\frac{\frac{3}{24}}{\frac{13}{24}} = \frac{3}{13}$
3	$\frac{\frac{1}{24}}{\frac{11}{24}} = \frac{1}{11}$	$\frac{\frac{4}{24}}{\frac{13}{24}} = \frac{4}{13}$
4	$\frac{\frac{2}{24}}{\frac{11}{24}} = \frac{2}{11}$	$\frac{\frac{1}{24}}{\frac{13}{24}} = \frac{1}{13}$
5	$\frac{\frac{1}{24}}{\frac{11}{24}} = \frac{1}{11}$	$\frac{\frac{3}{24}}{\frac{13}{24}} = \frac{3}{13}$
6	$\frac{\frac{4}{24}}{\frac{11}{24}} = \frac{4}{11}$	$\frac{\frac{2}{24}}{\frac{13}{24}} = \frac{2}{13}$
$\sum_i p_{X_1 X_2}(x_{1_i} x_{2_j})$	1	1

1.2.5 Teorema de la Probabilidad Total

Es una extensión inmediata del visto para una única variable aleatoria. Consideremos una partición del espacio muestral y que conocemos las probabilidades *a priori* de los sucesos de la misma, $P(M_p)$:

$$M_p \in \mathcal{P}(\Omega) \quad p = 1 \dots N \iff \begin{aligned} &\bigcup_{p=1}^N M_p = \Omega \\ &M_p \cap M_q = \emptyset \quad p \neq q \end{aligned}$$

Si conocemos las funciones de de distribución de probabilidad condicionada conjunta de un vector aleatorio (X, Y) condicionadas por todos los sucesos de dicha partición $F_{XY}(x, y|M_p)$, es fácil obtener la función de distribución total (sin condicionar):

$$F_{XY}(x, y) = \sum_{p=1}^N F_{XY}(x, y|M_p)P(M_p)$$

Para el **caso discreto**, el Teorema de la Probabilidad Total puede expresarse a partir de las funciones de masa de probabilidad $p_{XY}(x_i, y_j|M_p)$, de modo que la función de masa de probabilidad total (sin condicionar) conjunta del vector aleatorio (X, Y) resulta:

$$p_{XY}(x_i, y_j) = \sum_{p=1}^N p_{XY}(x_i, y_j|M_p)P(M_p)$$

En el **caso continuo**, si conocemos las funciones de densidad de probabilidad condicionadas, $f_{XY}(x, y|M_p)$, la función de densidad de probabilidad total (sin condicionar) conjunta del vector aleatorio (X, Y) es

$$f_{XY}(x, y) = \sum_{p=1}^N f_{XY}(x, y|M_p)P(M_p)$$

El Teorema de la Probabilidad Total permite obtener **mezclas (mixture) de distribuciones**, consistentes en la superposición ponderada de varias distribuciones, algo muy útil en la práctica.

La probabilidad total también puede aplicarse de forma alternativa, si lo que se conocen son las probabilidades un suceso arbitrario M condicionadas por cada uno de los valores posibles del vector aleatorio (X, Y) , esto es, si se conocen $P(M|x_i, y_j) \equiv P(M| \{X = x_i, Y = y_j\})$, en el caso discreto, y $P(M|x, y) \equiv P(M| \{X = x, Y = y\})$, en el caso continuo.

En tal caso, la probabilidad total del suceso M (sin condicionar) es, para el caso discreto,

$$P(M) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} P(M|x_i, y_j) p_{XY}(x_i, y_j)$$

Y, para el caso continuo,

$$P(M) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(M|x, y) f_{XY}(x, y) dx dy$$

En el caso del vector aleatorio discreto, si disponemos de la verosimilitud o función de masa de probabilidad condicionada de una variable (observación) respecto de la otra (condicionante) y de la función de masa de probabilidad marginal *a priori* de la variable condicionante, resulta la función de masa de probabilidad total marginal de las observaciones:

$$p_X(x_i) = \sum_j p_{X|Y}(x_i|y_j) p_Y(y_j)$$

$$p_Y(y_i) = \sum_i p_{Y|X}(y_i|x_i) p_X(x_i)$$

Y si el vector aleatorio es continuo:

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx$$

1.2.6 Teorema de Bayes

El Teorema de Bayes puede obtenerse en distinta formas, para lo cual es necesario recordar su definición y, si el suceso condicionante es un valor de una variable aleatoria continua y, por ello, tiene una masa de probabilidad infinitesimal, hay que considerar un rectángulo infinitesimal.

Veamos una primera forma, que hace uso de la función e distribución conjunta y de un suceso arbitrario M :

$$P(M | \{X \leq x, Y \leq y\}) = \frac{F_{XY}(x, y | M) P(M)}{F_{XY}(x, y)}$$

$$F_{XY}(x, y | M) = \frac{P(M | \{X \leq x, Y \leq y\}) F_{XY}(x, y)}{P(M)}$$

En el **caso discreto**, es inmediato obtener el **Teorema de Bayes** a partir de la definición de función de masa de probabilidad conjunta condicionada (recordar que $P(M|x_i, y_j) \equiv P(M | \{X = x_i, Y = y_j\})$):

$$p_{XY}(x_i, y_j | M) = \frac{P(M|x_i, y_j) p_{XY}(x_i, y_j)}{P(M)} \iff P(M|x_i, y_j) = \frac{p_{XY}(x_i, y_j | M) P(M)}{p_{XY}(x_i, y_j)}$$

Con frecuencia, los denominadores de las expresiones anteriores se obtienen mediante el *Teorema de la Probabilidad Total*, pues tan sólo se conocen las probabilidades condicionadas:

$$p_{XY}(x_i, y_j|M) = \frac{P(M|x_i, y_j)p_{XY}(x_i, y_j)}{\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} P(M|x_i, y_j)p_{XY}(x_i, y_j)}$$

$$P(M|x_i, y_j) = \frac{p_{XY}(x_i, y_j|M)P(M)}{p_{XY}(x_i, y_j|M)P(M) + p_{XY}(x_i, y_j|\bar{M})(1 - P(M))}$$

En el caso del vector aleatorio, si disponemos de la verosimilitud o función de masa de probabilidad condicionada de una variable (observación) respecto de la otra (condicionante) y de la función de masa de probabilidad marginal *a priori* de la variable condicionante, el Teorema de Bayes permite obtener la probabilidad *a posteriori* de la variable condicionante, dadas las observaciones:

$$p_{X|Y}(x_i|y_j) = \frac{p_{Y|X}(y_j|x_i)p_X(x_i)}{p_Y(y_j)} = \frac{p_{Y|X}(y_j|x_i)p_X(x_i)}{\sum_i p_{Y|X}(y_j|x_i)p_X(x_i)}$$

$$p_{Y|X}(y_j|x_i) = \frac{p_{X|Y}(x_i|y_j)p_Y(y_j)}{p_X(x_i)} = \frac{p_{X|Y}(x_i|y_j)p_Y(y_j)}{\sum_j p_{X|Y}(x_i|y_j)p_Y(y_j)}$$

Para el **caso continuo** las expresiones anteriores se adaptan sustituyendo funciones de masa de probabilidad por funciones de densidad y sumatorios por integrales:

$$f_{XY}(x, y|M) = \frac{P(M|x, y)p_{XY}(x, y)}{P(M)} \iff P(M|x, y) = \frac{f_{XY}(x, y|M)P(M)}{f_{XY}(x, y)}$$

$$f_{XY}(x, y|M) = \frac{P(M|x, y)f_{XY}(x, y)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(M|x, y)f_{XY}(x, y)dx dy}$$

$$P(M|x, y) = \frac{f_{XY}(x, y|M)P(M)}{f_{XY}(x, y|M)P(M) + f_{XY}(x, y|\bar{M})(1 - P(M))}$$

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx}$$

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy}$$

Ejemplo Consideremos en el ejemplo del lanzamiento del dado y moneda trucados y dependientes entre sí que las observaciones son las caras del dado X_1 y queremos hacer inferencia sobre el lado de la moneda X_2

- Si sólo tenemos las probabilidades de las caras condicionadas por el lado de la moneda, $p_{X_1|X_2}(x_{1_i}|x_{2_j})$ la inferencia se basa en el principio de máxima verosimilitud
- Si se cuenta con las probabilidades *a priori* de los lados de la moneda, $p_{X_2}(x_{2_j})$, es posible obtener las probabilidades *a posteriori* de las mismas dadas las observaciones $p_{X_2/X_1}(x_{2_j}|x_{1_i})$. En este caso utilizamos el **principio de Máximo a Posteriori (MAP)** y podemos obtener las probabilidades de acierto y de error de cada elección. Adviértase que la búsqueda del **MAP** suele conllevar la división por la probabilidad total, que es la misma en todas las posibilidades y, por tanto, puede evitarse

- Si se dispone de la probabilidad conjunta, se dispone de la máxima información probabilística para realizar inferencias

Verosimilitudes: si sale 1, 4 ó 6 seleccionamos cara (0), y cruz (1) en los restantes casos:

$L(x_{2j}) = p_{X_1 X_2}(x_{1i} x_{2j})$		0	1
$X_2 X_1$	1	$\frac{1}{11}$	0
	2	$\frac{2}{11}$	$\frac{3}{13}$
	3	$\frac{1}{11}$	$\frac{4}{13}$
	4	$\frac{2}{11}$	$\frac{1}{13}$
	5	$\frac{1}{11}$	$\frac{3}{13}$
	6	$\frac{4}{11}$	$\frac{2}{13}$
$\sum_i p_{X_1 X_2}(x_{1i} x_{2j})$		1	1

Probabilidades *a priori* de los lados de la moneda: $p_{X_2}(0) = \frac{11}{24}$ $p_{X_2}(1) = \frac{13}{24}$.

Probabilidades totales de las observaciones del dado: $p_{X_1}(x_{1i}) = \sum_j p_{X_1|X_2}(x_{1i}|x_{2j})p_{X_2}(x_{2j})$

x_{1i}	$p_{X_1}(x_{1i})$
1	$\frac{1}{11} \frac{11}{24} + 0 \frac{13}{24} = \frac{1}{24}$
2	$\frac{2}{11} \frac{11}{24} + \frac{3}{13} \frac{13}{24} = \frac{5}{24}$
3	$\frac{1}{11} \frac{11}{24} + \frac{4}{13} \frac{13}{24} = \frac{5}{24}$
4	$\frac{2}{11} \frac{11}{24} + \frac{1}{13} \frac{13}{24} = \frac{3}{24}$
5	$\frac{1}{11} \frac{11}{24} + \frac{3}{13} \frac{13}{24} = \frac{4}{24}$
6	$\frac{4}{11} \frac{11}{24} + \frac{2}{13} \frac{13}{24} = \frac{6}{24}$
$\sum_i p_{X_1}(x_{1i})$	1

Probabilidades *a posteriori* de los lados de la moneda, dadas las observaciones del dado:

De nuevo seleccionamos cara (0) si sale 1, 4 ó 6 y cruz (1) en los restantes casos. Aunque en este ejemplo coinciden las elecciones utilizando los criterios **ML** (máxima verosimilitud) y **MAP** (máximo a posteriori), en general no tiene por qué ser así:

$p_{X_2 X_1}(x_{2j} x_{1i})$		0	1	$\sum_j p_{X_1 X_2}(x_{1i} x_{2j})$
$X_2 X_1$	1	1	0	1
	2	$\frac{2}{5}$	$\frac{3}{5}$	1
	3	$\frac{1}{5}$	$\frac{4}{5}$	1
	4	$\frac{2}{3}$	$\frac{1}{3}$	1
	5	$\frac{1}{4}$	$\frac{3}{4}$	1
	6	$\frac{4}{6}$	$\frac{2}{6}$	1

Probabilidades de acierto y de error: para cada observación, el criterio **MAP** da la probabilidad de acertar, que es máxima. Por tanto, su complementario a 1 da la probabilidad de error, que es mínima:

x_{1i}	1	2	3	4	5	6	
$P(A x_{1i})$	1	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{4}{6}$	$\sum_i p_X(x_i) = 1$

La probabilidad total de acierto es:

$$P(A) = \sum_i P(A|x_{1i})p_{x_{1i}} = 1 \frac{1}{24} + \frac{3}{5} \frac{5}{24} + \frac{4}{5} \frac{5}{24} + \frac{2}{3} \frac{3}{24} + \frac{3}{4} \frac{4}{24} + \frac{4}{6} \frac{6}{24} = \frac{17}{24}$$

$Error$	x_{1_i}	1	2	3	4	5	6	
	$P(E x_{1_i})$	0	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{2}{6}$	$\sum_i p_X(x_i) = 1$

La probabilidad total de error es:

$$P(E) = \sum_i P(E|x_{1_i})p_{x_{1_i}} = 0\frac{1}{24} + \frac{2}{5}\frac{5}{24} + \frac{1}{5}\frac{5}{24} + \frac{1}{3}\frac{3}{24} + \frac{1}{4}\frac{4}{24} + \frac{2}{6}\frac{6}{24} = \frac{7}{24}$$

La probabilidad conjunta nos proporciona toda la información precedente muy fácilmente:

- Regla de decisión (ver tabla): (1, cara), (4, cara), (6, cara) (2, cruz), (3, cruz), (5, cruz)
- $P(A) = p_{X_1X_2}(1,0) + p_{X_1X_2}(4,0) + p_{X_1X_2}(6,0) + p_{X_1X_2}(2,1) + p_{X_1X_2}(3,1) + p_{X_1X_2}(5,1) = \frac{17}{24}$:

$p_{X_1X_2}(x_{1_i}, x_{2_j})$		0	1	$p_{X_1}(x_{1_i}) = \sum_j p_{X_1X_2}(x_{1_i}, x_{2_j})$
X_2X_1	1	$\frac{1}{24}$	0	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{5}{24}$
	3	$\frac{1}{24}$	$\frac{4}{24}$	$\frac{5}{24}$
	4	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{3}{24}$
	5	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{4}{24}$
	6	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{6}{24}$
$p_{X_2}(x_{2_j}) = \sum_i p_{X_1X_2}(x_{1_i}, x_{2_j})$		$\frac{11}{24}$	$\frac{13}{24}$	1

1.2.7 Representación matricial

El manejo de las tablas anteriores puede facilitarse haciendo uso de vectores y matrices de probabilidades. Considerando que la variable aleatoria X puede tomar M valores y que la Y puede tomar N :

- Vectores fila de probabilidades:
 - $\pi_X = [p_X(x_1), p_X(x_2) \dots p_X(x_M)]$
 - $\pi_Y = [p_Y(y_1), p_Y(y_2) \dots p_Y(y_N)]$
- Matrices de probabilidades condicionadas. Cada fila corresponde a un valor de la variable condicionante y, por tanto, todas y cada una de las filas suman 1:

$$\begin{aligned}
- \Pi_{X|Y} &= \begin{bmatrix} p_{X|Y}(x_1|y_1) & \dots & p_{X|Y}(x_M|y_1) \\ \vdots & \ddots & \vdots \\ p_{X|Y}(x_1|y_N) & \dots & p_{X|Y}(x_M|y_N) \end{bmatrix} \\
- \Pi_{Y|X} &= \begin{bmatrix} p_{Y|X}(y_1|x_1) & \dots & p_{Y|X}(y_N|x_1) \\ \vdots & \ddots & \vdots \\ p_{Y|X}(y_1|x_N) & \dots & p_{Y|X}(y_N|x_N) \end{bmatrix}
\end{aligned}$$

- Teorema de la Probabilidad Total: se hace uso del producto de matrices para multiplicar un vector fila por una matriz de probabilidades condicionadas y obtener otro vector fila:

$$\begin{aligned}
- \pi_Y &= \pi_X \Pi_{Y|X} \\
- \pi_X &= \pi_Y \Pi_{X|Y}
\end{aligned}$$

- Matriz de probabilidades conjuntas. Sus elementos suman 1

$$- \Pi_{XY} = \begin{bmatrix} p_{XY}(x_1, y_1) & \dots & p_{XY}(x_M, y_1) \\ \vdots & \ddots & \vdots \\ p_{XY}(x_1, y_N) & \dots & p_{XY}(x_M, y_N) \end{bmatrix}$$

Recurriendo a operaciones puntuales (elemento a elemento) es posible obtener algoritmos sencillos para aplicar el Teorema de Bayes o para obtener las matrices de probabilidades conjuntas a partir de las condicionadas y los vectores de probabilidades a priori.

1.2.8 Función de dos variables aleatorias

Consideremos un vector aleatorio discreto (X, Y) . Podemos aplicar sobre sus componentes una función g que devolverá valores aleatorios, al serlo también sus variables independientes. Lo representamos como sigue:

$$Z = g(X, Y)$$

Cada par (x_i, y_j) de (X, Y) se transformará por la función en un nuevo valor z_k de Z . Adviértase que valores distintos $(x_i, y_j) \neq (x_m, y_n)$ pueden dar lugar a valores iguales $z_k = y_j$ con tal que $g(x_i, y_j) = g(x_m, y_n)$. Por tanto, definimos la función de masa de probabilidad de Z como sigue:

$$p_Z(z_k) = \sum_{\substack{i,j \\ g(x_i, y_j) = z_k}} p_{XY}(x_i, y_j)$$

Adviértase que tal y como está expuesto no podemos invertir la transformación, pues al tratarse de una función de dos variables necesitaríamos una segunda función. No vamos a profundizar en la formulación matemática, y vamos a desarrollar la intuición con un ejemplo.

Ejemplo

Supongamos el lanzamiento de sucesivo e independiente de dos monedas, cuyos resultados modelamos con las variables aleatorias X e Y . En cada caso, cero significa cara y uno significa cruz. Las funciones de masa de probabilidad marginales son las siguientes:

X	0	1	
$p_X(x_i)$	$\frac{1}{3}$	$\frac{2}{3}$	$\sum_i p_X(x_i) = 1$

Y	0	1	
$p_Y(y_j)$	$\frac{5}{6}$	$\frac{1}{6}$	$\sum_i p_Y(y_j) = 1$

Al ser ambas variables aleatorias independientes puede obtenerse la distribución conjunta a partir de las marginales: $p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$

	$p_{XY}(x_i, y_j)$	0	1	$p_X(x_i) = \sum_j p_{XY}(x_i, y_j)$
YX	0	$\frac{1}{3} \cdot \frac{5}{6} = \frac{5}{18}$	$\frac{1}{3} \cdot \frac{1}{6} = \frac{1}{18}$	$\frac{1}{3}$
	1	$\frac{2}{3} \cdot \frac{5}{6} = \frac{10}{18}$	$\frac{2}{3} \cdot \frac{1}{6} = \frac{2}{18}$	$\frac{2}{3}$
	$p_Y(y_j) = \sum_i p_{XY}(x_i, y_j)$	$\frac{5}{6}$	$\frac{1}{6}$	1

La función de masa de probabilidad de la suma de ambas variables aleatorias es:

$Z = X + Y$	0	1	2	
$p_Z(z_k)$	$\frac{5}{8}$	$\frac{1}{18} + \frac{10}{18} = \frac{11}{18}$	$\frac{2}{18}$	$\sum_i p_Y(y_j) = 1$

Una propiedad muy importante de la **suma de dos variables aleatorias discretas independientes** permite obtener la función de masa de probabilidad de la variable aleatoria suma a partir de las de los dos sumandos:

La función de masa de probabilidad de la suma de dos variables aleatorias discretas **independientes**, $Z = X + Y$, se obtiene mediante la **convolución** discreta de las funciones de masa de probabilidad de cada uno de los sumandos:

$$p_Z(z_i) = p_X(x_i) * p_Y(y_i) = \sum_{k=-\infty}^{\infty} p_X(x_k) p_Y(y_{i-k}) = \sum_{k=-\infty}^{\infty} p_Y(y_k) p_X(x_{i-k})$$

La convolución es una operación matemática que se realiza entre dos secuencias o señales y que se verá en la asignatura de Señales y Sistemas. No vamos a profundizar aquí en su cálculo.

1.2.9 Esperanza Matemática y Valor Medio. Momentos

Se define la esperanza matemática de una función $g(X, Y)$ de dos variables aleatorias discretas como sigue:

$$E(g(X, Y)) = \sum_i \sum_j g(x_i, y_j) p_{XY}(x_i, y_j)$$

Si las variables aleatorias son continuas:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

Como ya se vio en el caso unidimensional, es fácil advertir que el operador $E()$ es **lineal**:

$$E(a_1 g_1(X, Y) + a_2 g_2(X, Y)) = a_1 E(g_1(X, Y)) + a_2 E(g_2(X, Y))$$

donde a_1 y a_2 son coeficiente constantes arbitrarios y g_1 y g_2 son dos funciones. También es evidente que, si k es una constante arbitraria

$$E(k) = k$$

De la linealidad se concluye inmediatamente:

$$E(X + Y) = E(X) + E(Y)$$

$$E(kX) = kE(X)$$

Veamos como obtener las esperanzas marginales en el caso discreto. En el caso continuo sería igual, sin más que cambiar funciones de masa de probabilidad por densidades y sumas por integrales.

Si $g(X, Y) = X$ se obtiene el valor medio de X :

$$\eta_X = E(X) = \sum_i \sum_j x_i p_{XY}(x_i, y_j) = \sum_i x_i \sum_j p_{XY}(x_i, y_j) = \sum_i x_i p_X(x_i)$$

Y si $g(X, Y) = Y$ se obtiene el valor medio de Y :

$$\eta_Y = E(Y) = \sum_i \sum_j y_j p_{XY}(x_i, y_j) = \sum_j y_j \sum_i p_{XY}(x_i, y_j) = \sum_j y_j p_Y(y_j)$$

Valor Cuadrático Medio y Correlación Considérese ahora que $g(X, Y) = (X + Y)^2$, esto es, calculamos el valor cuadrático medio de la suma de X e Y . Desarrollando el cuadrado y aplicando linealidad:

$$E((X + Y)^2) = E(X^2) + E(Y^2) + 2E(XY)$$

El término $R_{XY} = E(XY)$ se denomina **correlación**, y es un indicador de la dependencia probabilística de ambas variables aleatorias, X e Y .

- X e Y son **ortogonales** si $R_{XY} = E(XY) = 0$. Ello indica que no podemos pensar que ambas variables guarden una relación de proporcionalidad $Y \approx kX$ pues ello supondría que $E(X^2) \approx 0$.
- En general, $E(XY) \neq E(X)E(Y)$. Como veremos, si $E(XY) = E(X)E(Y)$ ambas variables aleatorias están **incorreladas**.
- Si X e Y son independientes es fácil advertir que $E(XY) = E(X)E(Y)$, de modo que la independencia implica incorrelación, si bien el contrario no es cierto en general.

Es habitual utilizar una formulación matricial, haciendo uso de la **matriz de correlación**:

$$\mathbf{R}_{XY} = E\left(\begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix}\right) = \begin{pmatrix} E(X^2) & E(XY) \\ E(XY) & E(Y^2) \end{pmatrix}$$

Como se ve, la matriz de correlación es una matriz simétrica, cuyos elementos sobre la diagonal corresponden a los valores cuadrático medio, $E(X^2)$ y $E(Y^2)$ y fuera de la misma está la correlación entre X e Y , $E(XY)$.

Si ambas variables fueran ortogonales, la matriz de correlación sería diagonal.

Covarianza Considérese ahora que centramos el vector aleatorio (X, Y) en el centro de masa probabilístico (η_X, η_Y) y que $g(X, Y) = ((X - \eta_X) + (Y - \eta_Y))^2$

$$E(((X - \eta_X) + (Y - \eta_Y))^2) = \sigma_X^2 + \sigma_Y^2 + 2E((X - \eta_X)(Y - \eta_Y))$$

El término $C_{XY} = E((X - \eta_X)(Y - \eta_Y))$ se denomina **covarianza**, y es también un indicador de la dependencia probabilística de ambas variables aleatorias, X e Y .

- La covarianza se relaciona con la correlación sin más que operar y aplicar linealidad: $C_{XY} = R_{XY} - \eta_X \eta_Y$
- X e Y son incorreladas si $C_{XY} = 0$. Decir que X e Y están incorreladas es equivalente a afirmar que $X - \eta_X$ e $Y - \eta_Y$ son ortogonales.
- Es fácil demostrar que $C_{XY} = 0 \iff R_{XY} = \eta_X \eta_Y$.
- La incorrelación supone ausencia de relación lineal entre X e Y , esto es, $Y \not\approx aX + b$.

Es habitual utilizar una formulación matricial, haciendo uso de la **matriz de covarianza**:

$$\mathbf{C}_{XY} = E\left(\begin{pmatrix} X - \eta_X \\ Y - \eta_Y \end{pmatrix} \begin{pmatrix} X - \eta_X & Y - \eta_Y \end{pmatrix}\right) = \begin{pmatrix} E((X - \eta_X)^2) & E((X - \eta_X)(Y - \eta_Y)) \\ E((X - \eta_X)(Y - \eta_Y)) & E((Y - \eta_Y)^2) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & C_{XY} \\ C_{XY} & \sigma_Y^2 \end{pmatrix}$$

Como se ve, la matriz de covarianza es una matriz simétrica, cuyos elementos sobre la diagonal corresponden a las varianzas, σ_X^2 y σ_Y^2 y fuera de la misma está la covarianza entre X e Y , C_{XY} .

Si ambas variables fueran incorreladas, la matriz de covarianza sería diagonal.

Coefficiente de correlación Si ahora trabajamos con las variables aleatorias tipificadas, $\hat{X} = \frac{X - \eta_X}{\sigma_X}$ e $\hat{Y} = \frac{Y - \eta_Y}{\sigma_Y}$ podemos definir el **coeficiente de correlación** (recuérdese que las variables tipificadas tienen, por construcción, valor medio nulo y desviación típica unidad):

$$r_{XY} = E\left(\frac{X - \eta_X}{\sigma_X} \frac{Y - \eta_Y}{\sigma_Y}\right) = \frac{1}{\sigma_X \sigma_Y} C_{XY} \quad -1 \leq r_{XY} \leq 1$$

El coeficiente de correlación r_{XY} :

- Tiene valor nulo si hay ausencia de relación lineal entre X e Y
- Tiene valor 1 si $Y = aX + b$ con $a > 0$
- Tiene valor -1 si $Y = aX + b$ con $a < 0$
- Si $0 < |r_{XY}| < 1$ el grado de dependencia lineal (positiva o negativa) será tanto mayor cuanto más cerca de 1.

Esperanza matemática condicionada Podemos definir la esperanza matemática de una función g de las variables aleatorias (X, Y) condicionada por el suceso M como sigue, si son discretas:

$$E(g(X, Y)|M) = \sum_i g(x_i, y_j) p_{XY}(x_i, y_j|M)$$

Si las variables aleatorias son continuas:

$$E(g(X, Y)|M) = \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y|M)$$

Podemos extender con la esperanza condicionada las propiedades y definiciones vistas anteriormente. Lo ilustraremos para el caso de variables aleatorias discretas, si bien es igual para continuas, sin más que utilizar funciones de densidad e integrales en vez de funciones de masa de probabilidad y sumatorios. Por ejemplo:

$$\begin{aligned} \eta_{X|M} &= E(X|M) = \sum_i \sum_j x_i p_{XY}(x_i, y_j|M) = \sum_i x_i p_X(x_i|M) \\ E(X^2|M) &= \sum_i \sum_j x_i^2 p_{XY}(x_i, y_j|M) = \sum_i x_i^2 p_X(x_i|M) \\ \sigma_{X|M}^2 &= \text{Var}(X|M) = E(X^2|M) - \eta_{X|M}^2 \end{aligned}$$

En particular, podemos considerar los momentos de X condicionados por $Y = y_j$, donde utilizamos el abuso de notación $E(\bullet | \{Y = y_j\}) \equiv E(\bullet | y_j)$:

$$\begin{aligned} \eta_{X/+|y_j} &= E(X|y_j) = \sum_i x_i p_{X|Y}(x_i|y_j) \\ E(X^2|y_j) &= \sum_i x_i^2 p_{X|Y}(x_i|y_j) \\ \sigma_{X|y_j}^2 &= \text{Var}(X|y_j) = E(X^2|y_j) - \eta_{X|y_j}^2 \end{aligned}$$

Adviértase que $E(X) = \sum_j E(X|y_j) p_Y(y_j)$, sin más que sustituir $E(X|y_j)$ por su definición y utilizar el Teorema de la Probabilidad Total.

Ejemplo Consideramos el ejemplo ya visto de lanzamiento mutuamente dependiente de un dado y una moneda trucados, cuya función de masa de probabilidad se indica de nuevo:

$p_{X_1 X_2}(x_{1i}, x_{2j})$		0	1	$p_{X_1}(x_{1i}) = \sum_j p_{X_1 X_2}(x_{1i}, x_{2j})$
$X_2 \backslash X_1$	1	$\frac{1}{24}$	0	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{5}{24}$
	3	$\frac{1}{24}$	$\frac{4}{24}$	$\frac{5}{24}$
	4	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{3}{24}$
	5	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{4}{24}$
	6	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{6}{24}$
$p_{X_2}(x_{2j}) = \sum_i p_{X_1 X_2}(x_{1i}, x_{2j})$		$\frac{11}{24}$	$\frac{13}{24}$	1

- Variable aleatoria X_1 (dado)

- $\eta_{X_1} = E(X_1) = 1 \frac{1}{24} + 2 \frac{5}{24} + 3 \frac{5}{24} + 4 \frac{3}{24} + 5 \frac{4}{24} + 6 \frac{6}{24} = \frac{94}{24} \approx 3.917$
- $E(X_1^2) = 1^2 \frac{1}{24} + 2^2 \frac{5}{24} + 3^2 \frac{5}{24} + 4^2 \frac{3}{24} + 5^2 \frac{4}{24} + 6^2 \frac{6}{24} = \frac{430}{24} \approx 17.917$
- $Var(X_1) = E(X_1^2) - \eta_{X_1}^2 = \frac{430}{24} - (\frac{94}{24})^2 \approx 2.576$
- $\sigma_{X_1} \approx 1.605$

- Variable aleatoria X_2 (moneda)

- $\eta_{X_2} = E(X_2) = 0 \frac{11}{24} + 1 \frac{13}{24} = \frac{13}{24} \approx 0.542$
- $E(X_2^2) = 0^2 \frac{11}{24} + 1^2 \frac{13}{24} = \frac{13}{24} \approx 0.542$
- $Var(X_2) = E(X_2^2) - \eta_{X_2}^2 = \frac{13}{24} - (\frac{13}{24})^2 = \frac{13}{24} \frac{11}{24} \approx 0.248$
- $\sigma_{X_2} \approx 0.498$
- Momentos conjuntos
 - * Correlación: $R_{XY} = 2 \frac{3}{24} + 3 \frac{4}{24} + 4 \frac{1}{24} + 5 \frac{3}{24} + 6 \frac{2}{24} = \frac{49}{24} \approx 2.042$
 - * Covarianza: $C_{XY} = R_{XY} - \eta_X \eta_Y = \frac{49}{24} - \frac{94}{24} \frac{13}{24} \approx -0.80$
 - * Coeficiente de correlación: $r_{\{XY\}} = \frac{1}{\sigma_X \sigma_Y} C_{\{XY\}} \approx -0.1$

Con $p_{X_1|X_2}(x_{1i}, x_{2j})$ pueden calcularse los momentos de los resultados del dado para cada resultado del lanzamiento de la moneda. Por ejemplo, la media:

$p_{X_1 X_2}(x_{1i} x_{2j})$		0	1
$X_2 \backslash X_1$	1	$\frac{1}{11}$	0
	2	$\frac{2}{11}$	$\frac{3}{13}$
	3	$\frac{1}{11}$	$\frac{4}{13}$
	4	$\frac{2}{11}$	$\frac{1}{13}$
	5	$\frac{1}{11}$	$\frac{3}{13}$
	6	$\frac{4}{11}$	$\frac{2}{13}$
$E(X_1 x_{2j})$		$1 \frac{1}{11} + 2 \frac{2}{11} + 3 \frac{1}{11} + 4 \frac{2}{11} + 5 \frac{1}{11} + 6 \frac{4}{11} = \frac{45}{11}$	$0 + 2 \frac{3}{13} + 3 \frac{4}{13} + 4 \frac{1}{13} + 5 \frac{3}{13} + 6 \frac{2}{13} = \frac{49}{13}$

Por supuesto, $E(X_1) = E(X_1/0)p_{X_2}(0) + E(X_1/1)p_{X_2}(1) = \frac{45}{11} \frac{11}{24} + \frac{49}{13} \frac{13}{24} = \frac{94}{24}$.