Capítulo 3

Vector Aleatorio

El estudio de procesos complejos suele involucrar más de una variable y entonces las herramientas estadísticas hasta aquí desarrolladas resultan insuficientes. Por ejemplo, si tenemos Precio y Demanda de un bien, que por razones fortuitas o de azar, se alejan de sus valores teóricos o de equilibrio y se convierten en variables aleatorias, digamos X e Y, con valores esperados μ_X y μ_Y y desviaciones estándar σ_X y σ_Y respectivamente, podemos pronosticar que lo más probable es un precio entre $\mu_X - \sigma_X$ y $\mu_X + \sigma_X$, y una demanda entre $\mu_Y - \sigma_Y$ y $\mu_Y + \sigma_Y$, pero no podremos pronosticar con qué probabilidad habrá un exceso de demanda ni dar forma analítica a la conocida relación teórica (inversa) entre Precio X y Demanda Y, ni usar ésta para afinar nuestros pronósticos. El problema ocurre por no tener un sistema de análisis conjunto de variables aleatorias. En este capítulo desarrollaremos ese sistema y para ello, consideraremos el análisis simultáneo de variables aleatorias definidas sobre un mismo espacio muestral de modo que conformen un vector de variables aleatorias o sea un vector aleatorio. Para simplificar, empezaremos con el caso bidimensional y luego generalizaremos al caso n dimensional.

3.1 Definición y Clasificación

Definición de vector aleatorio bidimensional
Un vector aleatorio (X,Y) es un vector cuyas componentes son variables aleatorias X e Y definidas conjuntamente sobre el mismo espacio muestral S. El conjunto de posibles parejas (X,Y) se denota R_{XY} y se llama Rango del vector (X,Y)

Clasificación
Diremos que (X,Y) es discreto si sus componentes son variables aleatorias discretas
Diremos que (X,Y) es continuo si sus componentes son variables aleatorias continuas

Nota:
Hay vectores (X,Y) "mixtos", con una componente discreta y la otra continua.

Extensión al caso n-dimensional
En general, un vector aleatorio <u>fila</u> n-dimensional $X_{1xn} = (X_1, X_2,..., X_n)$ es un vector cuyas componentes son variables aleatorias definidas conjuntamente. Análogamente podemos hablar de un vector aleatorio <u>columna</u> n-dimensional $X_{nx1} = (X_1, X_2,..., X_n)^t$. Filas o columnas, los vectores se clasifican como continuos, discretos o mixtos.

Ejemplo 3.1
Una empresa hace sucesivas perforaciones en busca de pozos petroleros y la probabilidad de tener éxito en una perforación cualquiera es $p > 0$. El espacio muestral S lo podemos representar como un conjunto de sucesiones $\{a_1, a_2, a_3, \cdots\}$ donde a_i=1 si en la perforación i-ésima se encuentra petróleo o a_i=0 si en la perforación i-ésima no se encuentra petróleo. En este contexto definamos las variables aleatorias X = Número de la perforación donde se encuentra petróleo por primera vez

$Y =$ Número de la perforación donde se encuentra petróleo por segunda vez

En este caso, el vector (X,Y) es vector aleatorio discreto (en adelante v.a.) donde $X = 1,2,3,\cdots$ e $Y = 2,3,4,\cdots$ y además se debe cumplir $X < Y$. Se tiene entonces
$R_{XY} = \{(x,y) \mid x = 1,2,3,\cdots; \quad y = (x+1),(x+2),\cdots\}$ o también
$R_{XY} = \{(x,y) \mid x = 1,2,3,\cdots,y-1; \quad y = 2,3,4,\cdots\}$

Ejemplo 3.2
La producción de una empresa ocurre de modo que con 40% de probabilidad un producto se exporta a Europa, con 50% se exporta al Asia y con 10% se vende en el mercado nacional. Si Ud. toma una muestra al azar de 5 productos recién terminados y debe decidir el destino de cada uno.
Si $X = \text{Número de productos exportados a Europa}$ e $Y = \text{Número de productos exportados a Asia}$, entonces (X,Y) es v.a. con rango $R_{XY} = \{(x,y) \mid x = 0,1,2,\cdots,5; \quad y = 0,1,2,\cdots,5; \quad x + y \le 5\}$

Ejemplo 3.3
El gobierno convoca a un concurso para la construcción de un puente y se presentan dos compañías A y B que ofrecen costos de a y b millones de unidades monetarias respectivamente, donde a y b pueden tomar cualquier valor entre 0 y 1 indistintamente. El espacio muestral S lo podemos representar como el conjunto de todas las ofertas posibles, esto es $S = \{(a,b) \mid 0 < a < 1; \quad 0 < b < 1\}$. Sean las variables aleatorias X = Precio ganador e Y = Diferencia entre las ofertas presentadas. En este caso (X,Y) es v.a. aleatorio continuo con rango $R_{XY} = \{(x,y) \mid x = Mín\{a,b\}; \quad y = |a-b|; \quad (a,b) \in S\}$

Ejemplo 3.4
Una consultora recibe proyectos cuyo costo de ejecución c puede tomar cualquier valor en]0,1[y fija un precio de consultoría $p > c$ para el proyecto, donde p que puede tomar cualquier valor en $]c,1[$. Sean las v.a. X = Costo de ejecución e Y = Precio de consultoría. Entonces (X,Y) es v.a. aleatorio continuo con rango $R_{XY} = \{(x,y) \mid 0 < x < 1; \quad x < y < 1\}$

Ejemplo 3.5
Una acción puede subir o bajar su cotización en una rueda de bolsa. Si sube su cotización, la subida puede ser de 100r% donde $0 < r < 1/2$ y si baja, ésta puede ser de 100r% donde $0 < r < 3/4$. La probabilidad de que suba la cotización es p. Si el valor de la acción al inicio de la rueda de bolsa es 1 y definimos las v.a. $X = \{\begin{matrix} 1 & si & hay & subida \\ 0 & si & hay & baja \end{matrix}$ e Y = Valor de la acción al final de la rueda. Entonces (X,Y) es v.a. aleatorio *mixto* con rango $R_{XY} = \{(x,y) \mid x = 0,1; \quad y =]0.25,1.5[\}$

Ejemplo 3.6
El número X de trabajadores contratados en una empresa es aleatorio con distribución geométrica $X \sim G(x;p)$ y el tiempo T_i que el trabajador contratado i permanece en la empresa también es aleatorio con distribución exponencial $T_i \sim Exp(t_i,\beta)$. Sea la v.a. Y = Tiempo de permanencia del trabajador contratado más antiguo. Entonces (X,Y) es v.a. *mixto* con rango $R_{XY} = \{(x,y) \mid x = 1,2,3,\cdots; \quad y =]0,\infty[\}$

Nota:
En algunos ejemplos hemos definido explícitamente el espacio muestral para resaltar que la definición de las componentes del v.a. debe ser conjunta, es decir X e Y se miden o registran simultáneamente, sobre los mismos casos. Esto es así para tener información que luego permita hacer pronósticos de una de las componentes dando valores a la otra, esto es, pretendemos "explicar" el comportamiento de una de las componentes a partir de la otra.

3.2 Distribuciones Conjunta, Marginales y Condicionales

3.2.1 Caso discreto

Función de Probabilidad Conjunta
Si (X,Y) es vector aleatorio discreto, la función de probabilidad conjunta de (X,Y), denotada $P_{XY}(x,y)$, se define mediante:

$$P_{XY}(x,y) = P\left[(X = x) \cap (Y = y)\right]$$

Nota:
$P_{XY}(x,y)$ proporciona la probabilidad de que ocurra la pareja (x,y), o sea la probabilidad de que simultáneamente la v.a. X tome el valor particular x y la v.a. Y tome el valor particular y. Naturalmente si $(x,y) \notin R_{XY}$ entonces $P_{XY}(x,y) = 0$

Propiedades
(a) $P_{XY}(x,y) \geq 0$
(b) $\sum_x \sum_y P_{XY}(x,y) = 1$
(c) $P[(X,Y) \in A] = \sum\sum_{(x,y)\in A} P_{XY}(x,y)$
La verificación de estas propiedades es consecuencia directa de ser $P_{XY}(x,y)$ una función que distribuye probabilidades sobre el rango R_{XY} del vector aleatorio (X,Y).

Nota:
La sumatoria que figura en las propiedades (b) y (c) es una "sumatoria doble", con dos subíndices x e y que debemos manejar, y aunque podemos sumar de manera libre, es mejor hacerlo ordenadamente, esto es
$\sum\sum_{(x,y)} P_{XY}(x,y) = \sum_x \left[\sum_y P_{XY}(x,y) \right]$ donde primero se suma dentro del corchete y en este caso x es fija e y subíndice variable de modo que el resultado depende de x y luego se suma sobre x y el resultado final es ya es un número. Como la suma no depende del orden de los sumandos, tenemos también que

$$\sum\sum_{(x,y)} P_{XY}(x,y) = \sum_x \left[\sum_y P_{XY}(x,y) \right] = \sum_y \left[\sum_x P_{XY}(x,y) \right]$$

Ejemplo 3.7
Sea (X,Y) v.a. discreto del ejemplo 3.1 de la empresa petrolera, donde X = Número de la perforación donde se encuentra petróleo por primera vez e Y = Número de la perforación donde se encuentra petróleo por segunda vez. Hallar la función de probabilidad conjunta $P_{XY}(x,y)$ de (X,Y) y la probabilidad de que X e Y resulten números consecutivos.

Solución:
Ya vimos que $R_{XY} = \{(x,y) \mid x = 1,2,3,\cdots; \quad y = (x+1),(x+2),\cdots\}$. Tomemos una pareja específica (x,y) en R_{XY}. En este contexto sea el evento A_i = "En la perforación número i se encuentra petróleo", entonces $\left[(X = x) \cap (Y = y)\right] = A_1^C \cap A_2^C \cap \cdots \cap A_{x-1}^C \cap A_x \cap A_{x+1}^C \cap A_{x+2}^C \cap \cdots \cap A_{y-1}^C \cap A_y = A \cap B$, donde
$A = A_1^C \cap A_2^C \cap \cdots \cap A_{x-1}^C \cap A_x$ = "En las (x-1) primeras perforaciones no se encuentra petróleo y en la x^a sí se encuentra petróleo" y $B = A_{x+1}^C \cap A_{x+2}^C \cap \cdots \cap A_{y-1}^C \cap A_y$ = "En las perforaciones (x+1) hasta la (y-1) no se

encuentra petróleo y en la y[a] sí se encuentra petróleo". Por dato $P(A_i)=p$, luego $P(A_i^C)=1-p$ y podemos escribir

$$P[(X=x)\cap(Y=y)]=P\left[A_1^C\cap A_2^C\cap\cdots\cap A_{x-1}^C\cap A_x\cap A_{x+1}^C\cap A_{x+2}^C\cap\cdots\cap A_{y-1}^C\cap A_y\right]=P[A\cap B]=$$

$$P[A]\times P[B\mid A]=\overbrace{(1-p)(1-p)\cdots(1-p)}^{(x-1)\ veces}p\times\overbrace{(1-p)(1-p)\cdots(1-p)}^{(y-x-1)\ veces}p=p^2(1-p)^{y-2}$$

Es decir $P_{XY}(x,y)=p^2(1-p)^{y-2}\quad x=1,2,3,\cdots\quad y=(x+1),(x+2),\cdots$ y fuera del rango de (X,Y), naturalmente tenemos $P_{XY}(x,y)=0$ o equivalentemente:

$$P_{XY}(x,y)=\{\begin{matrix}p^2(1-p)^{y-2} & x=1,2,3,\cdots\quad y=(x+1),(x+2),\cdots\\ 0\quad en\quad otro\quad caso\end{matrix}$$

Finalmente sea el evento D = "X e Y resultan números consecutivos" entonces:
$D=\{(x,y)\in R_{XY}\mid x=1,2,,3\cdots\quad y=x+1\}$ y

$$P[(X,Y)\in D]=\sum_{x=1}^{\infty}\left[\sum_{y=x+1}^{x+1}p^2(1-p)^{y-2}\right]=\sum_{x=1}^{\infty}\left[p^2(1-p)^{(x+1)-2}\right]=\sum_{x=1}^{\infty}p^2(1-p)^{x-1}=p^2(1-p)^{-1}\sum_{x=1}^{\infty}(1-p)^x=$$

$$p^2(1-p)^{-1}\frac{(1-p)}{1-(1-p)}=p$$ (recordemos la "serie geométrica": $\sum_{j=1}^{\infty}r^j=\frac{r}{1-r}\quad si\,|r|<1$)

En este ejemplo hemos podido escribir fórmulas explícitas para $P_{XY}(x,y)$ y $P[(X,Y)\in D]$. Estas fórmulas explícitas ("fórmulas cerradas" es su nombre técnico) son más bien la excepción y no la regla. Muchas veces no queda sino hacer una tabla, y en casos extremos, usar la computadora para los cálculos, haciendo lo que se llama "Estadística computacional", que es la tendencia actual.

Ejemplo 3.8
Sea (X,Y) v.a. discreto, donde X = Número créditos concedidos en una agencia bancaria e Y = Número de créditos impagos dentro de los concedidos, con función de probabilidad conjunta $P_{XY}(x,y)$ dada por

$$P_{XY}(x,y)=cx\qquad x=1,2,3,4.\qquad y=0,1,2,\ldots,x$$

Hallar la constante c que hace a $P_{XY}(x,y)$ función de probabilidad ("constante normalizadora") y la probabilidad de que todos los créditos sean pagados.

Solución:
Apliquemos $\sum_x\sum_y P_{XY}(x,y)=1$. En este caso, es más sencillo escribir $P_{XY}(x,y)$ en formato de una tabla de doble entrada y sumar:

Y	X			
	1	2	3	4
0	c	2c	3c	4c
1	c	2c	3c	4c
2	0	2c	3c	4c
3	0	0	3c	4c
4	0	0	0	4c

$$\sum_x\sum_y P_{XY}(x,y)=40c=1\Rightarrow c=1/40\text{ y }P_{XY}(x,y)=\frac{x}{40}\quad x=1,2,3,4;\quad y=0,1,2,\ldots,x$$

Finalmente si A denota el evento "Todos los créditos son pagados", entonces

$$A = \{Y = 0\} = \{(x,0) \in R_{XY}\} \text{ y } P(A) = \sum_{x=1}^{4} P_{XY}(x,0) = c + 2c + 3c + 4c = 10c = 10/40 = 0.25$$

Ejemplo 3.9
En el ejemplo anterior ¿Cuál es la probabilidad de que se hayan concedido 3 créditos? ¿De que se hayan concedido 3 créditos si se sabe que hubo 2 créditos impagos?

Solución:
"Se han concedido 3 créditos" equivale a $X = 3$ y esto último equivale a

$(X = 3) \cap (Y = 0)$ o $(X = 3) \cap (Y = 1)$ o $(X = 3) \cap (Y = 2)$ o $(X = 3) \cap (Y = 3)$, así que:

$$P(X = 3) = P[(X = 3) \cap (Y = 0)] + P[(X = 3) \cap (Y = 1)] + P[(X = 3) \cap (Y = 2)] + P[(X = 3) \cap (Y = 3)]$$
$$= P_{XY}(3,0) + P_{XY}(3,1) + P_{XY}(3,2) + P_{XY}(3,3) = 3c + 3c + 3c + 3c = 12c = 12/40 = 0.3$$

Análogamente:

$$P(\textit{De que se hayan concedido 3 créditos si se sabe que hubo 2 créditos impagos}) = P(X = 3 \mid Y = 2)$$

y aplicando la definición de probabilidad condicional

$$P(X = 3 \mid Y = 2) = \frac{P[(X = 3) \cap (Y = 2)]}{P(Y = 2)} = \frac{P_{XY}(3,2)}{P(Y = 2)} = \frac{3c}{P(Y = 2)} = \frac{3c}{9c} = 0.33$$

pues

$$P(Y = 2) = P[(X = 2) \cap (Y = 2)] + P[(X = 3) \cap (Y = 2)] + P[(X = 4) \cap (Y = 2)]$$
$$= P_{XY}(2,2) + P_{XY}(3,2) + P_{XY}(4,2) = 2c + 3c + 4c = 9c$$

Función de Probabilidad Marginal
Sea (X,Y) v.a.d, con función de probabilidad conjunta $P_{XY}(x,y)$, se define la Función de Probabilidad Marginal de X, denotada $P_X(x)$, mediante:

$$P_X(x) = \sum_y P_{XY}(x,y)$$

Análogamente, la Función de Probabilidad Marginal de Y se denota $P_Y(y)$ y se define:

$$P_Y(y) = \sum_x P_{XY}(x,y)$$

Nota
Si deseamos hallar la probabilidad de que la componente X del v.a. (X,Y) tome el valor particular x podemos escribir
$P[X = x] = P[(X = x) \cap (Y \in R)] = \sum_y P[(X = x) \cap (Y = y)] = \sum_y P_{XY}(x,y) = P_X(x)$. Similarmente se comprueba que $P_Y(y) = P[Y = y] = \sum_x P_{XY}(x,y)$. Es decir, las distribuciones marginales proporcionan las distribuciones individuales de X y de Y, que son las mismas que obtendríamos si definimos estas v.a. por separado. Lo anterior muestra que no perdemos ninguna información si trabajamos desde el inicio con la distribución conjunta. Es más, en realidad con esta entrada ganamos información, como lo demuestran las definiciones dadas en el siguiente parágrafo

Función de Probabilidad Condicional

En el contexto anterior, se define la Función de Probabilidad Condicional de Y dado que $X=x$, denotada $P_{Y|X}(y|x)$, mediante:

$$P_{Y|X}(y|x)=\frac{P_{XY}(x,y)}{P_X(x)} \quad \text{con } x \text{ \underline{valor dado} tal que } P_X(x)>0$$

Análogamente se define la Función de Probabilidad Condicional de X dado que $Y=y$, denotada $P_{X|Y}(x|y)$, mediante:

$$P_{X|Y}(x|y)=\frac{P_{XY}(x,y)}{P_Y(y)} \text{ con } y \text{ \underline{valor dado} tal que } P_Y(y)>0$$

Nota

En realidad se trata de una aplicación directa de la definición de probabilidad condicional: $P[Y=y\mid X=x]=\dfrac{P[(X=x)\cap(Y=y)]}{P(X=x)}=\dfrac{P_{XY}(x,y)}{P_X(x)}=P_{Y|X}(y|x)$. Similarmente se obtiene e interpreta $P_{X|Y}(x|y)=P[X=x|Y=y]$. ¿Cuál es la utilidad de las distribuciones condicionales? Pues es inmensa, ya que pone al analista en posición de hacer pronósticos de una de las componentes del v.a (X,Y), dando valores a la otra componente. Esta es la base de toda explicación, podemos pronosticar y saber las tendencias de una variable de interés a partir de otra, que usamos como "variable explicativa". En Economía rara vez se estudia una variable aislada; al contrario, se trata siempre de asociarla a otra u otras que permitan hacer pronósticos confiables y tomar decisiones. Los modelos económicos en el fondo son formulaciones que muestran cómo una o unas variables económicas responden a otras que podemos usar como base para políticas. Las distribuciones condicionales están en la base de los modelos económicos cuando en éstos hay variables aleatorias.

Ejemplo 3.10

En el ejemplo 3.7, hallar $P_Y(y)$ y $P_{X|Y}(x|y)$. En promedio ¿Cómo varía X según va cambiando Y?

Solución:

Para hallar $P_Y(y)$, fijemos un valor y para aplicar la definición, recordando que fijado el valor y, entonces X sólo puede tomar valores desde 1 hasta $(y-1)$. Veamos:

$P_Y(y)=\sum_x P_{XY}(x,y)=\sum_{x=1}^{y-1} p^2(1-p)^{y-2}=(y-1)p^2(1-p)^{y-2} \qquad y=2,3,\cdots$. Ahora vayamos por $P_{X|Y}(x|y)$:

$$P_{X|Y}(x|y)=\frac{P_{XY}(x,y)}{P_Y(y)}=\frac{p^2(1-p)^{y-2}}{(y-1)p^2(1-p)^{y-2}}=\frac{1}{(y-1)} \qquad x=1,2,\cdots,(y-1) \quad (y=\text{valor dado})$$

Examinando $P_{X|Y}(x|y)$, podemos concluir que dado un valor $Y=y$, la componente X puede tomar cualquier valor entre 1 e (y-1) con la misma probabilidad, pero conforme aumentamos el valor y, el rango de posibilidades para X aumenta. Podemos representar esta tendencia usando la media de X en esta distribución condicional:

$\textit{Media de } X=\sum_{x=1}^{(y-1)} xP_{X|Y}(x|y)=\sum_{x=1}^{(y-1)} x\frac{1}{(y-1)}=\frac{1}{(y-1)}\sum_{x=1}^{(y-1)} x=\frac{1}{(y-1)}\frac{(y-1)(y)}{2}=\frac{y}{2}$. Como pensamos, la *media de X aumenta conforme aumenta y*. Para resaltar que no es la media o valor esperado individual de *X* sino el valor esperado *condicionado* a *Y=y*, usaremos la notación $E[X\mid Y=y]$ y así escribiremos $E[X\mid Y=y]=\dfrac{y}{2}$ para mostrar cómo *en promedio la componente X del vector aumenta conforme* *aumenta la componente Y.*

Ejemplo 3.11
En el ejemplo 3.8, hallar $P_X(x)$ y $P_{Y|X}(y|4)$

Solución:
Sumando en el cuadro de $P_{XY}(x,y)$ tenemos en cada margen:

Y	X				
	1	2	3	4	$P_Y(y)$
0	c	2c	3c	4c	10c
1	c	2c	3c	4c	10c
2		2c	3c	4c	9c
3			3c	4c	7c
4				4c	4c
$P_X(x)$	2c	6c	12c	20c	1

Y como $c=1/40$, reemplazando obtenemos

Y	X				
	1	2	3	4	$P_Y(y)$
0	1/40	2/40	3/40	4/40	10/40
1	1/40	2/40	3/40	4/40	10/40
2		2/40	3/40	4/40	9/40
3			3/40	4/40	7/40
4				4/40	4/40
$P_X(x)$	2/40	6/40	12/40	20/40	1

Las funciones de probabilidad marginales se encuentran en los márgenes derecho e inferior de la Tabla de Probabilidad conjunta, esto es:

x	1	2	3	4
$P_X(x)$	2/40	6/40	12/40	20/40

y	0	1	2	3	4
$P_Y(y)$	10/40	10/40	9/40	7/40	4/40

Finalmente, cuando $X=4$ tenemos $P_{Y|X}(y|4)=\dfrac{P_{XY}(4,y)}{P_X(4)}=\dfrac{4/40}{20/40}=0.20 \quad y=0,1,...,4$

Ejemplo 3.12
La producción de una empresa ocurre de modo que con 40% de probabilidad un producto se exporta a Europa, con 50% se exporta al Asia y con 10% se vende en el mercado nacional. Si Ud. toma una muestra al azar de 5 productos recién terminados y debe decidir el destino de cada uno.
Si $X = \textit{Número de productos exportados a Europa}$ e $Y = \textit{Número de productos exportados a Asia}$

a) ¿Con qué probabilidad enviará $X=3$ productos a Europa?
b) ¿Con qué probabilidad enviará $X=2$ productos a Europa e $Y=2$ productos al Asia?
c) Halle $P_{XY}(x,y)$
d) Halle $P_{Y|X}(1|3)$

Solución:

a) $X=3 \Leftrightarrow$ "de los cinco productos 3 van Europa y 2 no van a Europa", luego
$P(X=3)=C_3^5(0.4)^3(0.6)^2$ y en general $P(X=k)=P_X(k)=C_k^5(0.4)^k(0.6)^{5-k} \quad k=0,1,....5$

b) $(X=2\cap Y=2) \Leftrightarrow$ "de los cinco productos 2 van Europa y 2 van a Asia y 1 queda aquí", luego
$P(X=2\cap Y=2)=C_2^5(0.4)^2\times C_2^3(0.5)^2\times C_2^1(0.1)^1=\frac{5!}{2!2!1!}(0.4)^2(0.5)^2(0.1)^1$

c) $P_{XY}(x,y)=P(X=x\cap Y=y)=C_x^5(0.4)^x C_y^{5-x}(0.5)^y C_{5-x-y}^{5-x-y}(0.1)^{5-x-y}$
$=\frac{5!}{x!\,y!(5-x-y)!}(0.4)^x(0.5)^y(0.1)^{5-x-y}\ x=0,1,... \quad y=0,1,.. \quad x+y\le 5$

d) $P_{Y|X}(1|3)=\frac{P_{XY}(3,1)}{P_X(3)}=\frac{\frac{5!}{3!1!1}(0.4)^3(0.5)^1(0.1)^1}{C_3^5(0.4)^3(0.6)^2}=C_1^2\,\frac{(0.5)^1(0.1)^1}{(0.6)^2}$.

Y en general

$$P_{Y|X}(y|x)=\frac{P_{XY}(x,y)}{P_X(x)}=\frac{\frac{5!}{x!\,y!(5-x-y)!}(0.4)^x(0.5)^y(0.1)^{5-x-y}}{C_x^5(0.4)^x(0.6)^{5-x}}$$
$$=C_y^{5-x}\,\frac{(0.5)^y(0.1)^{5-x-y}}{(0.6)^{5-x}}=C_y^{5-x}\left(\frac{5}{6}\right)^y\left(\frac{1}{6}\right)^{5-x-y}\ y=0,1,...,5-x \quad x=0,1,...,5$$

3.2.2 Caso continuo

Función de Densidad Conjunta

Si (X,Y) es vector aleatorio continuo, la función de densidad conjunta de (X,Y), denotada $f_{XY}(x,y)$ es una función continua tal que:

- $f_{XY}(x,y)\ge 0$
- $\iint_{R^2} f_{XY}(x,y)=1$
- $P[(X,Y)\in A)]=\iint_A f_{XY}(x,y)$ para cualquier región A contenida en el rango R_{XY} de (X,Y)

Nota (Integrales dobles)

1) Recordemos que si una función $f_{XY}(x,y)$ es continua sobre una región $A\subset R^2$ dada por $A=\{(x,y)\,|\,a\le x\le b, \quad c\le y\le d\}$, entonces la integral doble $\iint_A f_{XY}(x,y)$ se calcula como una <u>integral iterada</u>, primero sobre y luego sobre x, o también en el orden inverso (Teorema de Fubini). Esto es:
$$\iint_A f_{XY}(x,y)=\int_a^b\left[\int_c^d f_{XY}(x,y)dy\right]dx=\int_c^d\left[\int_a^b f_{XY}(x,y)dx\right]dy\,.$$

2) También, si $A \subset R^2$ se puede escribir como una región de fronteras definidas en términos de funciones, como $A = \{(x, y) \mid a \le x \le b, \quad h_1(x) \le y \le h_2(x)\}$, entonces:
$$\iint_A f_{XY}(x, y) = \int_a^b \left[\int_{h_1(x)}^{h_2(x)} f_{XY}(x, y) dy \right] dx.$$
3) Análogamente, si $A \subset R^2$ es $A = \{(x, y) \mid g_1(y) \le x \le g_2(y), \quad c \le y \le d\}$, entonces:
$$\iint_A f_{XY}(x, y) = \int_c^d \left[\int_{g_1(y)}^{g_2(y)} f_{XY}(x, y) dx \right] dy$$
4) Geométricamente, en todos los casos, $\iint_A f_{XY}(x, y) =$ Volumen debajo de la gráfica de la superficie $z = f_{XY}(x, y)$ sobre la región $A \subset R^2$.

Por lo anterior y partiendo de la condición $\iint_{R^2} f_{XY}(x, y) = 1$ (o equivalentemente de $\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_{XY}(x, y) dy \right] dx = 1$), vemos que $f_{XY}(x, y)$ distribuye probabilidades de modo continuo sobre R_{XY}, asignando <u>volúmenes</u> sobre los distintos eventos A contenidos en R_{XY}, de modo que el volumen total (o probabilidad total) es 1.

Nótese finalmente que aunque en la definición se integra sobre todo R^2, en la práctica la integral es sólo sobre R_{XY}, <u>pues fuera</u> de R_{XY}, $f_{XY}(x, y) = 0$.

Ejemplo 3.13
Sea (X, Y) vector aleatorio donde X = Ingreso e Y = Consumo, con f. de densidad conjunta
$$f_{XY}(x, y) = ce^{-x} \quad 0 < y < x < \infty$$
Halle la constante normalizadora c y la probabilidad de que el consumo sea menor al 75% del ingreso.

Solución:
Apliquemos $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1$ para hallar c. Como $R_{XY} = \{(x, y) \mid 0 < y < x < \infty\}$ que se puede reescribir $R_{XY} = \{(x, y) \mid 0 < x < \infty \quad 0 < y < x\}$, entonces:
$$\iint_{R^2} f_{XY}(x, y) = \iint_{R_{XY}} f_{XY}(x, y) = \int_0^\infty \left[\int_0^x ce^{-x} dy \right] dx = \int_0^\infty ce^{-x} \left[\int_0^x dy \right] dx = \int_0^\infty ce^{-x} [y]_0^x dx = \int_0^\infty ce^{-x} x dx = c \int_0^\infty xe^{-x} dx = c\Gamma(2) = c$$

Igualando a 1, se obtiene $c = 1$ y finalmente $f_{XY}(x, y) = e^{-x} \quad 0 < y < x < \infty$

Finalmente, $A =$ "Consumo menor al 75% del ingreso" equivale al evento
$A = \{(x, y) \in R_{XY} \mid 0 < x < \infty, \quad 0 < y < x, \quad y < 0.75x\} = \{(x, y) \in R_{XY} \mid 0 < x < \infty, \quad 0 < y < 0.75x\}$ y
$$P((X, Y) \in A) = \iint_A f_{XY}(x, y) = \int_0^\infty \left[\int_0^{0.75x} e^{-x} dy \right] dx = \int_0^\infty e^{-x} \left[\int_0^{0.75x} dy \right] dx = \int_0^\infty 0.75xe^{-x} dx = 0.75$$

Función de Densidad Marginal
Sea (X, Y) v.a.c, con función de probabilidad conjunta $f_{XY}(x, y)$, se define la Función de Densidad Marginal de X, denotada $f_X(x)$, mediante:

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

Análogamente, la Función de Densidad de Y se denota $f_Y(y)$ y se define:

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

Función de Densidad Condicional

En el contexto anterior, definimos la Función de Densidad Condicional de Y dado que $X = x$, denotada $f_{Y|X}(y|x)$, mediante:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \text{ donde } x \text{ es valor dado tal que } f_X(x) > 0$$

Análogamente se define la Función de Densidad Condicional de X dado que $Y = y$, denotada $f_{X|Y}(x|y)$, mediante:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ donde } y \text{ es valor dado tal que } f_Y(y) > 0$$

Nota:

- $P(a < X \leq b) = P((a < X \leq b) \cap (-\infty < Y \leq \infty)) = \int_a^b \left[\int_{-\infty}^{\infty} f_{XY}(x,y)dy \right] dx$. Luego, $\int_{-\infty}^{\infty} f_{XY}(x,y)dy$ es una función de X que integrada sobre]a,b] proporciona la probabilidad de X en dicho intervalo, es decir, $\int_{-\infty}^{\infty} f_{XY}(x,y)dy$ cumple los requisitos de una función de densidad para X y así podemos escribir $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$. Similarmente ocurre con $f_Y(y)$
- Se define la probabilidad condicional $P(c < Y \leq d \mid X = x) := \int_c^d f_{Y|X}(y|x)dy$ y también la probabilidad condicional $P(a < X \leq b \mid Y = y) := \int_a^b f_{X|Y}(x|y)dx$

Ejemplo 3.14

En el ejemplo anterior, halle $f_X(x)$, $f_{Y|X}(x|y)$ y $P(Y \leq 0.8 \mid X = 1.5)$

Solución:

Aplicando la definición para hallar la densidad marginal de X, tenemos:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_0^x e^{-x}dy = e^{-x}[y]_0^x = xe^{-x} \text{ donde } 0 < x < \infty$$

En cuanto a la condicional de Y dado X:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{e^{-x}}{xe^{-x}} = \frac{1}{x} \text{ donde } 0 < y < x \text{ siendo } x \text{ un valor dado.}$$

En particular, para $X = 1.5$, $f_{Y|X}(y|1.5) = \dfrac{1}{1.5}$ donde $0 < y < 1.5$ y por tanto**:**

$$P(Y \leq 0.8 \mid X = 1.5) = \int_0^{0.8} \frac{1}{1.5} dy = \frac{0.8}{1.5} = 0.53$$

Nota:
Obsérvese que $X \sim \Gamma(\alpha = 2, \beta = 1)$ y **dado** $X = x,\ Y$ tiene distribución uniforme en el intervalo $[0, x]$, o sea $Y \mid X \sim U(y; 0, X)$ y $E[Y \mid X = x] = \frac{x}{2}$: *Por cada unidad adicional de ingreso, la mitad se destina al consumo adicional.*

También: $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_y^{\infty} e^{-x} dx = \left[-e^{-x}\right]_y^{\infty} = e^{-y}$. Esto es $Y \sim Exp(\beta = 1)$

Ejemplo 3.15
Resuelva el ejemplo anterior si la f. de densidad conjunta de (X, Y) es $f_{XY}(x, y) = ye^{-x} \quad 0 < y < x < \infty$

Solución:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^x ye^{-x} dy = e^{-x}\left[\frac{y^2}{2}\right]_0^x = \frac{x^2}{2} e^{-x} \quad 0 < x < \infty$$

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{ye^{-x}}{\frac{x^2}{2} e^{-x}} = \frac{2y}{x^2} \quad 0 < y < x \quad (x = \text{valor dado})$$

y

$$f_{Y|X}(y \mid 1.5) == \frac{2y}{1.5^2} \quad 0 < y < 1.5$$, así que

$$P(Y \leq 0.8 \mid X = 1.5) = \int_0^{0.8} \frac{2y}{1.5^2} dy = \frac{1}{1.5^2}\left[y^2\right]_0^{0.8} = \frac{0.64}{1.5^2} = 0.28$$

Nota:

En este ejemplo $E[Y \mid X = x] = \int_0^x y f_{Y|X}(y \mid x) dy = \int_0^x y \frac{2y}{x^2} dy = \frac{2}{x^2} \int_0^x y^2 dy = \frac{2}{x^2}\left[\frac{y^3}{3}\right]_0^x = \frac{2}{3} x$, o sea que *conforme aumenta el ingreso, el consumo también aumenta y lo hace a una tasa de constante de 2/3 por cada unidad adicional de ingreso.*

3.2.3 Independencia

- Sean X e Y v.a. discretas, diremos que X e Y son independientes si:

$$P_{XY}(x, y) = P_X(x) P_Y(y) \quad \forall (x, y)$$

- Sean X e Y v.a. continuas, diremos que X e Y son independientes si:

$$f_{XY}(x, y) = f_X(x) f_Y(y) \quad \forall (x, y)$$

Ejemplo 3.15
Las variables X e Y del ejemplo donde X = Número créditos concedidos en una agencia bancaria e Y = Número de créditos impagos dentro de los concedidos, no son independientes.

En cambio, X e Y con función de probabilidad conjunta

$P_{XY}(x,y)=\left(\frac{1}{2}\right)^{x+y}$ $x=1,2,3,...$ $y=1,2,3,...$ sí son independientes.

Ejemplo 3.16
El FMI diseña un rescate financiero de dos países A y B, cubriendo totalmente sus déficits en cuenta corriente con X e Y en miles de millones de dólares respectivamente. Si (X,Y) tiene función de densidad conjunta

$$f_{XY}=\begin{cases} x^2+xy/3 & si \quad 0<x<1, \quad 0<y<2 \\ 0 & en \quad otro \quad caso \end{cases}$$

a) Halle la distribución marginal del monto de ayuda a A
b) El FMI sólo tiene mil millones ¿Con qué probabilidad no podrá ayudar a los dos países simultáneamente?
c) Si A recibe 500 millones ¿Con qué probabilidad B recibirá más de esa cantidad? ¿Cuánto esperaría recibir B en este caso?
d) Use propiedades de valor esperado para estimar la cantidad de dinero que gastaría el FMI en A y B
e) ¿El Fondo financia a cada país sin tomar en cuenta al otro?. Justifique.

Solución:

a) $f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy=\int_0^2 (x^2+xy/3)dy=\left[x^2y+xy^2/6\right]_0^2=2x^2+2x/3 \quad 0<x<1$

b) Sea C el evento "El FMI no podrá ayudar a los dos países", entonces esto equivale a $C=\{(x,y)\in R_{XY} \mid x+y>1\}$ y graficando vemos que es mejor calcular $P(C)$ mediante $P(C)=1-P(C^C)$:

$$P(C^C)=P((X,Y)\in C^C)=P(X+Y\le 1)=\iint_{C^C} f_{XY}(x,y)=\int_0^1\left[\int_0^{1-x}(x^2+xy/3)dy\right]dx=\int_0^1\left[x^2y+xy^2/6\right]_0^{1-x}dx=$$

$$=\int_0^1[x^2(1-x)+x(1-x)^2/6]dx=\frac{1}{6}\int_0^1[x+4x^2-5x^3]dx=\frac{1}{6}\left[\frac{x^2}{2}+\frac{4x^3}{3}-\frac{5x^4}{4}\right]_0^1=\frac{7}{72} \text{ y } P(C)=\frac{65}{72}$$

c) Si A recibe 500 millones, entonces $X=0.5$ y necesitamos calcular $P(Y>0.5\mid X=0.5)=\int_{0.5}^2 f_{Y|X}(y\mid 0.5)dy$

Como en general $f_{Y|X}(y\mid x)=\dfrac{f_{XY}(x,y)}{f_X(x)}=\dfrac{x^2+xy/3}{2x^2+2x/3}$ $0<y<2$; $0<x<1, x=valor\,dado$

en particular, $f_{Y|X}(y\mid 0.5)=\dfrac{0.25+0.5y/3}{2\times 0.25+2\times 0.5/3}=\dfrac{4+y}{10}$ $0<y<2$ así que

$$P(Y>0.5\mid X=0.5)=\int_{0.5}^2 f_{Y|X}(y\mid 0.5)dy=\int_{0.5}^2\left(\frac{4+y}{10}\right)dy=\left[\frac{8y+y^2}{20}\right]_{0.5}^2=1-\frac{17}{80}=\frac{63}{80}$$

Lo que "esperaría recibir B dado que A ya recibió 0.5" es

$$E[Y\mid X=0.5]=\int_0^2 yf_{Y|X}(y\mid x)dy=\int_0^2 y\left(\frac{4+y}{10}\right)dy=\left[\frac{6y^2+y^3}{30}\right]_0^2=\frac{32}{30}=1.07\,mil\,millones$$

3.3 Valor Esperado

3.3.1 Definición

Si (X,Y) es vector aleatorio y $H(X,Y)$ es una función de (X,Y), se define el Valor Esperado de $H(X,Y)$, denotado $E[H(X,Y)]$, mediante

$$E[H(X,Y)] = \begin{cases} \sum\limits_{x}\sum\limits_{y} H(x,y)P_{XY}(x,y) & si \quad (X,Y) \quad es \quad discreto \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} H(x,y) f_{XY}(x,y)dxdy & si \quad (X,Y) \quad es \quad continuo \end{cases}$$

Nota

$E[H(X,Y)]$ es una constante obtenida como un promedio ponderado de los distintos valores que puede tomar $H(X,Y)$.

3.3.2 Casos Especiales

- $\mu_X := E[X]$ y $\mu_Y := E[Y]$, son las medias poblacionales de X e Y respectivamente
- $\sigma_X^2 \equiv V(X) := E[(X-\mu_X)^2]$ y $\sigma_Y^2 \equiv V(Y) := E[(Y-\mu_Y)^2]$, son las varianzas poblacionales de X e Y respectivamente
- $\sigma_{XY} \equiv Cov(X,Y) := E[(X-\mu_X)(Y-\mu_Y)]$, llamada Covarianza entre X e Y. Este indicador mide asociación (**lineal**) entre X e Y, o sea la propensión a variar conjuntamente que tienen X e Y

Interpretación de σ_{XY}:

Como σ_{XY} es el promedio de los valores del producto $(X-\mu_X)(Y-\mu_Y)$ entonces:

- $\sigma_{XY} > 0$ quiere decir que 'en promedio' $(X-\mu_X)$ y $(Y-\mu_Y)$ tienen el mismo signo, es decir, por lo general ocurre que $X > \mu_X$ y a la vez $Y > \mu_Y$ ó ocurre que $X < \mu_X$ y a la vez $Y < \mu_Y$. Podemos concluir que X e Y tienden a "crecer o disminuir a la vez", i.e. están asociadas directamente.
- $\sigma_{XY} < 0$ quiere decir que 'en promedio' $(X-\mu_X)$ y $(Y-\mu_Y)$ tienen signo opuesto, es decir, por lo general ocurre que $X > \mu_X$ y a la vez $Y < \mu_Y$ ó ocurre que $X < \mu_X$ y a la vez $Y > \mu_Y$. Podemos concluir que X e Y tienden a "moverse en dirección opuesta", están asociadas inversamente.
- Si $\sigma_{XY} = 0$, no hay relación lineal entre X e Y, aunque puede haber una relación no lineal

3.3.3 Propiedades del Valor Esperado

Propiedad 1

$E[C] = C$ para toda constante o variable no aleatoria C

Propiedad 2

Si α es constante o variable no aleatorias y $H(X,Y)$ es función de (X,Y), entonces:

$$E[\alpha H(X,Y)] = \alpha E[H(X,Y)]$$

Propiedad 3

Si $\alpha_1, \alpha_2, \ldots, \alpha_n$ son constantes (o variables no aleatorias) y $H_1(X,Y)$, $H_2(X,Y)$, ... $H_n(X,Y)$ son funciones de (X,Y), entonces:

$$E[\sum_{j=1}^{n}\alpha_j H_j(X,Y)]=\sum_{j=1}^{n}\alpha_j E[H_j(X,Y)]$$

Nota:
Lo anterior se resume diciendo que el valor esperado $E[\bullet]$ es un "operador lineal". Pero obviamente $E[H(X,Y)]\neq H(E[X],E[Y])$

Propiedad 3
Si X e Y **son independientes** y $H(X)$, $G(Y)$ son funciones de X e Y, entonces:

$$E[H(X)G(Y)]=E[H(X)]E[G(Y)]$$

Consecuencias de las propiedades
(a) $Cov(X,Y)=E[XY]-E[X]E[Y]$
(b) Si X e Y **son independientes**, entonces $Cov(X,Y)=0$
(c) Si $W=\alpha X+\beta Y$, donde α y β son constantes, entonces:

$$\mu_W=E[W]=\alpha E[X]+\beta E[Y] \text{ y } \sigma_W^2=V[W]=\alpha^2 V[X]+\beta^2 V[Y]+2\alpha\beta Cov(X,Y)$$

Definición: Coeficiente de Correlación de Pearson
Denotado ρ_{XY} y definido mediante $\rho_{XY}=\dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$, este coeficiente mide lo mismo que la covarianza, pero carece de unidades y por tanto es más preciso en la medición de la asociación

Propiedades de ρ_{XY}
- $0\leq|\rho_{XY}|\leq 1$
- $\rho_{XY}>0$ indica asociación directa o positiva entre X e Y
- $\rho_{XY}<0$ indica asociación inversa o negativa entre X e Y
- $|\rho_{XY}|=1 \Leftrightarrow$ Existen constantes α y β tales que $Y=\alpha+\beta X$
- $|\rho_{XY}|\cong 1$ indica que entre X e Y hay asociación (lineal) 'fuerte'
- $|\rho_{XY}|\cong 0$ indica que entre X e Y hay asociación (lineal) 'débil'

Propiedad:
Sean $T=(X+Y)$ y $D=(X-Y)$ entonces:
- $\mu_T=\mu_X+\mu_Y$ y $\sigma_T^2=\sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_X$
- $\mu_D=\mu_X-\mu_Y$ y $\sigma_D^2=\sigma_X^2+\sigma_Y^2-2\rho_{XY}\sigma_X\sigma_X$

Ejemplo 3.17
En el ejemplo 3.13 de (X,Y) donde X =Ingreso e Y =Consumo, hallar σ_{XY} y ρ_{XY}

Solución:
Aplicando la propiedad $\sigma_{XY}=Cov(X,Y)=E[XY]-E[X]E[Y]$.
Como $X\sim\Gamma(\alpha=2,\beta=1)$, entonces $E(X)=\mu_X=\alpha\beta=2$ y $V(X)=\sigma_X^2=\alpha\beta^2=2$ y como $Y\sim Exp(\beta=1)$, entonces $E(Y)=\mu_Y=1/\beta=1$ y $V(Y)=\sigma_Y^2=1/\beta^2=1$. De modo que sólo necesitamos hallar $E[XY]$.

$$E[XY]=\iint_{R^2} xyf_{XY}(x,y)=$$

$$\int_0^\infty\left[\int_0^x xye^{-x}dy\right]dx=\int_0^\infty xe^{-x}\left[\int_0^x ydy\right]dx=\int_0^\infty xe^{-x}\left[\frac{y^2}{2}\right]_0^x dx=\int_0^\infty \frac{x^3e^{-x}}{2}dx=\frac{1}{2}\int_0^\infty x^3e^{-x}dx=\frac{1}{2}\Gamma(4)=3$$

Por tanto, $\sigma_{XY}=E[XY]-E[X]E[Y]=3-2x1=1$ y $\rho_{XY}=\dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}=\dfrac{1}{\sqrt{2}x1}=0.71$

Viendo la correlación ρ_{XY} se puede deducir que entre X e Y hay relación directa

3.3.4 Esperanza Condicional E[Y|X]

Denotada $E[Y\,|\,X]$, es el valor esperado de Y en la distribución condicional de Y dado $X=x$. También se llama Función de Regresión de Y sobre X, pues suele ser una función de X que se usa para pronosticar el valor de Y cuando se conoce el valor de X.

En efecto, si por ejemplo (X,Y) es continuo, entonces $E[Y\,|\,X=x]=\int_{-\infty}^{+\infty}yf_{Y|X}(y\,|\,x)dy$ y como x está fijo, el resultado es una función del valor dado x que fijemos, i.e. $E[Y\,|\,X=x]=\int_{-\infty}^{+\infty}yf_{Y|X}(y\,|\,x)dy\equiv\varphi(x)$; si ahora dejamos variar libremente a x, ésta recupera su condición aleatoria y podemos escribir $E[Y\,|\,X]=\varphi(X)$ que es una función de la v.a. X

Análogamente se define la Regresión de X sobre Y como $E[X\,|\,Y]$, que es una función $\phi(Y)$

Propiedad 1

$E(E[Y\,|\,X])=E[\varphi(X)]=E[Y]$

Demostración (caso continuo):

$$E(E[Y\,|\,X])=\int_{-\infty}^{+\infty}\varphi(x)f_X(x)dx=\int_{-\infty}^{+\infty}[\int_{-\infty}^{+\infty}yf_{Y|X}(y\,|\,x)dy]f_X(x)dx=\int_{-\infty}^{+\infty}[\int_{-\infty}^{+\infty}y\frac{f_{XY}(x,y)}{f_X(x)}dy]f_X(x)dx=$$

$$\int_{-\infty}^{+\infty}[\int_{-\infty}^{+\infty}yf_{XY}(x,y)dy]dx=E[Y]$$

Observaciones:

- Se define la varianza condicional de $Y\,|\,X$, denotada $V[Y\,|\,X]$ o también $\sigma^2_{Y|X}$, mediante $V[Y\,|\,X]=E[Y^2\,|\,X]-(E[Y\,|\,X])^2$
- $E[Y\,|\,X]=\varphi(X)$ nos proporciona el valor esperado de Y conociendo (o en función de) el valor de X y es el mejor pronóstico de Y a partir del conocimiento de X. Se puede decir que es la parte de Y "explicada" por X.
- $\sigma^2_{Y|X}$ mide la variabilidad de Y alrededor del pronóstico $E[Y\,|\,X]=\varphi(X)$ y $\sigma_{Y|X}$ puede verse como el "margen de error" en el pronóstico de Y a partir de X vía $E[Y\,|\,X]=\varphi(X)$ y en este sentido se escribe $Y=E[Y\,|\,X]\pm\sigma_{Y|X}$
- Muchos modelos económicos se formulan en términos de valores esperados, esto es, se escribe algo como $E[Y\,|\,X]=\varphi(X)$ para explicar los valores de la variable Y como "dependiente" de la variable X, llamada por esto variable independiente.

Propiedad 2
$V(Y) = E(V[Y \mid X]) + V(E[Y \mid X])$ **y** $E(V[Y \mid X]) \leq V(Y)$

Ejemplo 3.18
En el ejemplo 3.13 de (X,Y) donde X = Ingreso e Y = Consumo, calcule $E[Y \mid X]$. ¿Cómo interpretaría $E[Y \mid X]$ y $\frac{\partial E[Y \mid X]}{\partial X}$?

Solución:
Ya vimos que
$f_{Y|X}(y \mid x) = \frac{1}{x}$ donde $0 < y < x$ siendo x un valor dado, por tanto:
$E[Y \mid X] = \int_{\infty}^{\infty} y f_{Y|X}(y \mid x)dy = \int_0^x y\frac{1}{x}dy = \frac{1}{x}\int_0^x ydy = \frac{1}{x}\left[\frac{y^2}{2}\right]_0^x = \frac{x}{2}$, esto es, dado un ingreso, en promedio la mitad del mismo se va en consumo.

$\frac{\partial E[Y \mid X]}{\partial X} = \frac{\partial}{\partial X}\frac{X}{2} = \frac{1}{2}$, que se interpreta como una propensión marginal a consumir, o sea, por cada unidad monetaria adicional de ingreso, la mitad se dedica al consumo.

3.3.5 Vector de Medias y Matriz de Varianza Covarianza

Definición (Vector de Medias)

Sea $(X,Y)^T$ vector aleatorio columna, el vector de medias de $(X,Y)^T$, denotado μ, se define mediante
$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$

Definición (Matriz de Varianza-Covarianza)

Sea $(X,Y)^T$ vector aleatorio columna, la matriz de varianza covarianza de $(X,Y)^T$, denotada Σ, se define mediante $\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$

Observaciones:

- μ resume la "tendencia central" de $(X,Y)^T$
- Σ resume la variabilidad de $(X,Y)^T$. En este sentido hay dos maneras de ver la "variabilidad de $(X,Y)^T$: Mediante la traza de Σ y mediante el determinante de Σ
- La Varianza Total de $(X,Y)^T$, que denotaremos VT, es $VT = Tr(\Sigma) = \sigma_X^2 + \sigma_Y^2$ y mide la variabilidad presente en $(X,Y)^T$ como la suma de las variabilidades de cada una de sus componentes.
- La Varianza Generalizada de $(X,Y)^T$, que denotaremos VG, es $VG = \det(\Sigma) = = \sigma_X^2\sigma_Y^2 - (\sigma_{XY})^2 = \sigma_X^2\sigma_Y^2 - (\rho_{XY}\sigma_X\sigma_Y)^2 = \sigma_X^2\sigma_Y^2(1-\rho_{XY}{}^2)$. VG mide variabilidad en $(X,Y)^T$, pero toma en cuenta la correlación entre X e Y: cuanto más variabilidad haya en cada componente, mayor será VG, pero esto es atenuado si hay correlación entre X e Y; y en el caso extremo de correlación perfecta (o sea $\rho_{XY} = \pm 1$), VG será nula indicando que aunque hay variabilidad en cada componente,

las parejas (X,Y) tienen un aspecto constante en su comportamiento: siempre caen siguiendo una recta en el plano XY, lo que permite predicciones exactas de una componente, digamos Y, si llegamos a conocer el valor de la otra, digamos X.

Propiedad:
Sea $U_{2x1}=(X,Y)^T$ vector aleatorio con vector de medias μ_U y matriz de varianza covarianza Σ_U; sean A_{2x2} y B_{2x1} matrices dadas. Si definimos el vector $V_{2x1}:=AU+B$ entonces el vector de medias y la matriz de varianza covarianza de V son $\mu_V=A\mu_U+B$ y $\Sigma_V=A\Sigma_U A^T$ respectivamente.

Ejemplo 3.19
En el ejemplo 3.13 de (X,Y) donde X = Ingreso e Y = Consumo hallar μ, Σ y la media y la varianza del Ahorro.

Solución:

Ya hemos calculado todos los elementos necesarios para responder.

$$\mu=\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}=\begin{pmatrix}2\\ 1\end{pmatrix} \text{ y } \Sigma=\begin{pmatrix}\sigma_X^2 & \sigma_{XY}\\ \sigma_{XY} & \sigma_Y^2\end{pmatrix}=\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$$

Sea Z = Ahorro, entonces $Z=X-Y=\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}X\\ Y\end{pmatrix}$ y aplicando la propiedad:

$$\mu_Z=\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix}=2-1=1;\ \sigma_Z^2=\Sigma_Z=\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}=\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}=1$$

En este ejemplo, también es aplicable la propiedad:

- Si $W=\alpha X+\beta Y$, donde α y β son constantes, entonces $\mu_W=E[W]=\alpha E[X]+\beta E[Y]$ y $\sigma_W^2=V[W]=\alpha^2V[X]+\beta^2V[Y]+2\alpha\beta Cov(X,Y)$.

Basta notar que en este caso $\alpha=1$ y $\beta=-1$ de modo que:
$E[Z]=1\times E[X]+(-1)\times E[Y]=2-1=1$ y $V[Z]=1^2\times V[X]+(-1)^2\times V[Y]+2\times1\times(-1)\times Cov(X,Y)=$
$1^2\times2+(-1)^2\times1+2\times1\times(-1)\times1=2+1-2=1$

3.4 Vector Aleatorio n-dimensional

Como ya se dijo, un vector aleatorio es un vector cuyas componentes son n variables aleatorias definidas conjuntamente. En particular:

- Un vector aleatorio n-dimensional (fila) es de la forma $\vec{\mathbf{X}}_{1xn} = (X_1, X_2, ..., X_n)$
- Un vector aleatorio n-dimensional (columna) es de la forma $\vec{\mathbf{X}}_{nx1} = (X_1, X_2, ..., X_n)^T$.
- Cuando el contexto lo permite, se usa una notación unificada $\vec{\mathbf{X}}$ para representar el vector, sin hacer alusión a si es fila o columna.

La definición anterior puede ampliarse al caso de matrices aleatorias.

Matriz aleatoria
Una matriz aleatoria M de orden $n \times m$ es una matriz $M = \left(M_{ij}\right)_{n\times m}$ de nxm variables aleatorias $M_{ij} \quad i = 1,2,...,n \quad j = 1,2,...,m$ definidas conjuntamente sobre el mismo espacio muestral S.

Vector particionado
Cuando se tiene un vector aleatorio $X_{nx1} = (X_1, X_2, ..., X_n)^T$, a veces es útil separar partes de él, por ejemplo $X_{nx1} = (X_1, \cdots, X_p, X_{p+1}, \cdots, X_n)^T$ donde las primeras
p componentes forman un subvector $X_{px1} = (X_1, \cdots, X_p)^T$ de orden px1 y el resto de (n-p) componentes un subvector $X_{(n-p)x1} = (X_{p+1}, \cdots, X_n)^T$. En este contexto se escribe $X_{nx1} = (X_1 \vdots X_2)^T$ donde $X_1 = (X_1, \cdots, X_p)^T$ y $X_2 = (X_{p+1}, \cdots, X_n)^T$. O más explícitamente, en formato de vector columna:

$$X_{nx1} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ \cdots \\ X_{p+1} \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix} \text{ donde } X_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \text{ y } X_2 = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_n \end{pmatrix}.$$

3.4.1 Distribuciones

Para simplificar, tanto en el caso discreto como en el continuo usaremos la misma notación para las distribuciones.

Las nociones de distribución conjunta $f_{X_1X_2\ldots X_n}(x_1, x_2, \ldots, x_n)$; distribuciones marginales $f_{X_1}(x_1), f_{X_2}(x_2), ..., f_{X_n}(x_n)$; y condicionales $f_{X_i|X_j}(x_i \mid x_j)$, son extensiones directas del caso bivariado. En particular es importante:

Definición de Independencia.
$X_1, X_2, \ldots, X_n$ se dicen **independientes** si:

$$f_{X_1X_2\ldots X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n) \equiv \prod_{j=1}^{n} f_{X_j}(x_j)$$

Nota
En verdad, se necesita que esta regla se cumpla con todos los subconjuntos de componentes del vector, por ejemplo, debe cumplirse que
$f_{X_1X_2}(x_1,x_2)= f_{X_1}(x_1)f_{X_2}(x_2)$; $f_{X_1X_2X_3}(x_1,x_2,x_3)= f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3)$; etc.

En términos de vector particionado se escribe

- La función de distribución conjunta $f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots,x_n)\equiv f_{X_{nx1}}(x_{nx1})\equiv f_{X_1X_2}(x_1,x_2)$, donde en la última expresión $X_1\in R^p$ y $X_2\in R^{n-p}$ son subvectores
- La distribución marginal del subvector $X_1\in R^p$ como
  $f_{X_1}(x_1,\ldots,x_p)\equiv f_{X_{px1}}(x_{px1})\equiv f_{X_1}(x_1)$
- La distribución marginal del subvector $X_2\in R^{n-p}$ como
  $f_{X_2}(x_{p+1},\ldots,x_n)\equiv f_{X_{(n-p)x1}}(x_{(n-p)x1})\equiv f_{X_2}(x_2)$
- La distribución condicional del subvector $X_1\in R^p$, dado el subvector $X_2\in R^{n-p}$, como
  $$f_{X_1|X_2}(x_1\mid x_2)=\frac{f_{X_1X_2}(x_1,x_2)}{f_{X_2}(x_2)}$$
- $X_1\in R^p$ y $X_2\in R^{n-p}$ son independientes si $f_{X_1X_2}(x_1,x_2)= f_{X_1}(x_1)f_{X_2}(x_2)\quad \forall x_1\,\forall x_2$

Nota (Necesidad de vectores particionados)
Los modelos económicos suelen ser multiecuacionales, donde el economista trata de "explicar" el comportamiento de un conjunto de ***p*** variables de interés y que llama "endógenas", como resultado de su interacción con otro conjunto de **(*n-p*)** variables que llama "exógenas". Si las variables tienen elementos de aleatoriedad (porque es obtienen con encuestas por muestreo o porque en sí son aleatorias), el conjunto total de variables en el modelo completo es un vector aleatorio $X_{nx1}=(X_1,\cdots,X_p,X_{p+1},\cdots,X_n)^T$, donde por comodidad de lectura primero, y de análisis después, es preferible escribir $X_1=(X_1,\cdots,X_p)^T$ para representar al vector de variables "endógenas" y $X_2=(X_{p+1},\cdots,X_n)^T$ para representar al vector de variables "exógenas". En este contexto, entendemos por "explicación", el dar cuenta del comportamiento de $X_1\in R^p$ a partir del comportamiento de $X_2\in R^{n-p}$. En términos de probabilidades, esto se logra a través de $f_{X_1|X_2}(x_1\mid x_2)=\dfrac{f_{X_1X_2}(x_1,x_2)}{f_{X_2}(x_2)}$. Ciertamente, sería muy fatigoso hacer análisis económico al grado de explicitar totalmente $f_{X_1|X_2}(x_1\mid x_2)$. Lo que hace el analista económico es plantear las relaciones entre X_1 y X_2 primero y luego explicitarlas a través de ecuaciones simultáneas. Lo primero equivale a plantear la forma de la distribución condicional $f_{X_1|X_2}(x_1\mid x_2)$ y lo segundo se hace usando esperanzas condicionales. Para eso, mejor pasemos a tratar este punto en contexto de vectores aleatorios

3.4.2 Valor Esperado

Como en el caso bivariado, tenemos

- **Media Poblacional** de la componente X_j de $\vec{\mathbf{X}}$:
  $\mu_j\equiv\mu_{X_j}:=E[X_j]$
- **Varianza Poblacional** de la componente X_j de $\vec{\mathbf{X}}$:
  $\sigma_j^2\equiv\sigma_{X_j}^2:=V[X_j]=E[(X_j-\mu_j)^2]$

- **Covarianza Poblacional** entre las componentes X_i y X_i de $\vec{\mathbf{X}}$:
  $\sigma_{ij} \equiv \sigma_{X_i X_j} \equiv Cov(X_i, X_j) := E[(X_i - \mu_i)(X_j - \mu_j)]$
- **Coeficiente de Correlación Lineal de Pearson** entre las componentes X_i y X_i de $\vec{\mathbf{X}}$:
  $\rho_{ij} \equiv \rho_{X_i X_j} := \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$
- **Vector de Medias** del vector (columna) $\vec{\mathbf{X}} = (X_1, X_2, ..., X_n)^T$:

$$\boldsymbol{\mu}_{\vec{\mathbf{X}}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}_{nx1}$$

- **Matriz de Varianza-covarianza** del vector (columna) $\vec{\mathbf{X}} = (X_1, X_2, ..., X_n)^T$:

$$\Sigma_X = \left(\sigma_{ij}\right)_{n\times n} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}_{nxn}$$

- **Varianza Total** $VT := tr(\boldsymbol{\Sigma}_{\vec{\mathbf{X}}}) = \sum_{j=1}^{n} \sigma_j^2$
- **Varianza Generalizada** $VG := Det(\Sigma_{\vec{\mathbf{X}}}) = |\Sigma_{\vec{\mathbf{X}}}|$
- **Vector de Medias del vector particionado** $X_{nx1} = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix}$. Se define mediante

$$\boldsymbol{\mu}_{\vec{\mathbf{X}}} = E\left(X_{nx1}\right) = \begin{pmatrix} E(X_1) \\ \cdots \\ E(X_2) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix} \text{ donde } \mu_1 = E\left(X_1\right) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} \text{ y } \mu_2 = E\left(X_2\right) = \begin{pmatrix} E(X_{p+1}) \\ \vdots \\ E(X_{n-p}) \end{pmatrix}$$

- **Matriz de Varianza-covarianza del vector particionado** $X_{nx1} = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix}$ Se define

$$\Sigma_X = \left(\sigma_{ij}\right)_{n\times n} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{nxn} \text{ donde } \Sigma_{11} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}, \ \Sigma_{12} = \begin{pmatrix} \sigma_{1(p+1)} & \cdots & \sigma_{1n} \\ \sigma_{2(p+1)} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots \\ \sigma_{p(p+1)} & \cdots & \sigma_{pn} \end{pmatrix}$$

$$\Sigma_{21} = \begin{pmatrix} \sigma_{(p+1)1} & \cdots & \sigma_{(p+1)p} \\ \sigma_{(p+2)1} & \cdots & \sigma_{(p+2)p} \\ \vdots & \vdots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{np} \end{pmatrix} \text{ y } \Sigma_{22} = \begin{pmatrix} \sigma_{(p+1)(p+1)}^2 & \sigma_{(p+1)(p+2)} & \cdots & \sigma_{(p+1)n} \\ \sigma_{(p+2)(p+1)} & \sigma_{(p+2)(p+2)}^2 & \cdots & \sigma_{(p+2)n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n(p+1)} & \sigma_{n(p+2)} & \cdots & \sigma_{nn}^2 \end{pmatrix}$$

. Nótese que Σ_{11} es la matriz de varianza-covarianza de X_1, Σ_{22} la matriz de varianza-covarianza de X_2 y $\Sigma_{12} = \Sigma_{21}^T$ es la matriz de covarianzas entre X_1 y X_2.

Dada una M una matriz aleatoria de orden $n \times m$, podemos ensayar una definición general que será de utilidad a futuro, a saber:

Valor esperado de una matriz aleatoria
Sea M una matriz aleatoria de orden $n \times m$, su valor esperado, denotado $E(M)$, se define como la matriz de valores esperados de las respectivas componentes. Esto es $E(M) = \left(E(M_{ij})\right)_{n\times m}$

Propiedad
Si $X_{n\times 1}$ es vector columna con vector de medias $\boldsymbol{\mu}_X$ y matriz de varianza covarianza Σ_X, entonces se cumple que $\Sigma_X = E\left[(X-\mu_X)(X-\mu_X)^T\right]$

Demostración
En efecto, según la definición de valor esperado de una matriz aleatoria, tenemos

$$E\left[(X-\mu_X)(X-\mu_X)^T\right] = E\left[\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \ \vdots \ \vdots \\ X_n-\mu_n \end{pmatrix}(X_1-\mu_1, X_2-\mu_2, ..., X_n-\mu_n)\right] =$$

$$E\left[\begin{pmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_n-\mu_n) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \dots & (X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \cdots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & (X_n-\mu_n)(X_2-\mu_2) & \cdots & (X_n-\mu_n)^2 \end{pmatrix}\right] =$$

$$\left[\begin{pmatrix} E\left((X_1-\mu_1)^2\right) & E\left((X_1-\mu_1)(X_2-\mu_2)\right) & \dots & E\left((X_1-\mu_1)(X_n-\mu_n)\right) \\ E\left((X_2-\mu_2)(X_1-\mu_1)\right) & E\left((X_2-\mu_2)^2\right) & \dots & E\left((X_2-\mu_2)(X_n-\mu_n)\right) \\ \vdots & \vdots & \cdots & \vdots \\ E\left((X_n-\mu_n)(X_1-\mu_1)\right) & E\left((X_n-\mu_n)(X_2-\mu_2)\right) & \cdots & E\left((X_n-\mu_n)^2\right) \end{pmatrix}\right] =$$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}_{nxn} = \Sigma_X .$$

A veces es más simple usar la notación $V(X)$ para la matriz Σ

Propiedad
Si $X_1 \in R^p$ y $X_2 \in R^{n-p}$ son vectores aleatorios columna con vectores de medias μ_1 y μ_2 respectivamente, entonces la matriz de covarianza Σ_{12} satisface $\Sigma_{12} = E\left[(X_1-\mu_1)(X_2-\mu_2)^T\right]$.
Análogamente $\Sigma_{21} = E\left[(X_2-\mu_2)(X_1-\mu_1)^T\right]$

Demostración
Es similar a la de la propiedad anterior.

Propiedad
$\Sigma_{\bar{\mathbf{X}}}$ es simétrica y semidefinida positiva, esto es $Y^T\Sigma_{\bar{\mathbf{X}}}Y \geq 0$ para todo vector no nulo $Y \in R^n$

Proposición

Sea $\vec{X}$ vector aleatorio y sean $H_1(\vec{X}), H_2(\vec{X}),..., H_k(\vec{X})$ k funciones reales de $\vec{X}$ (esto es $H_j : R^n \to R$). Sean $\alpha_1, \alpha_2, \cdots, \alpha_k$ constantes (o también variables pero no aleatorias), entonces se cumple

$$E[\sum_{j=1}^{k} \alpha_j H_j(\vec{X})] = \sum_{j=1}^{k} \alpha_j E[H_j(\vec{X})]$$

Proposición

Sea $\vec{X}$ vector aleatorio y sean $H_1(\vec{X}), H_2(\vec{X}),..., H_k(\vec{X})$ k funciones vectoriales de $\vec{X}$ (esto es $H_j : R^n \to R^m$). Sean $\alpha_1, \alpha_2, \cdots, \alpha_k$ constantes (o también variables, pero no aleatorias), entonces se cumple $E[\sum_{j=1}^{k} \alpha_j H_j(\vec{X})] = \sum_{j=1}^{k} \alpha_j E[H_j(\vec{X})]$

Observación:

El operador Valor Esperado $E[\bullet]$ es operador lineal.

Proposición

Sea $\vec{X}$ vector aleatorio n-dimensional $\vec{X}_{1xn} = (X_1, X_2,..., X_n)$ tal que sus componentes **son independientes** y sean $H_1(X_1), H_2(X_2), \cdots, H_n(X_n)$ funciones reales de $X_1, X_2,..., X_n$ respectivamente.

Entonces se cumple:

$$E[\prod_{j=1}^{n} H_j(X_j)] = \prod_{j=1}^{n} E[H_j(X_j)]$$

Corolario

Sea $\vec{X}$ vector aleatorio n-dimensional $\vec{X}_{1xn} = (X_1, X_2,..., X_n)$ tal que sus componentes **son independientes,** entonces se cumple $\sigma_{ij} = Cov(X_i, X_j) = 0$

Proposición

Sea $\mathbf{X}$ vector aleatorio n-dimensional $\mathbf{X}_{1xn} = (X_1, X_2,..., X_n)$ y sean $\alpha_1, \alpha_2, \cdots, \alpha_k$ constantes (o también variables no aleatorias). Entonces:

- $E[\sum_{j=1}^{n} \alpha_j X_j] = \sum_{j=1}^{n} \alpha_j E[X_j]$ y
- $V[\sum_{j=1}^{n} \alpha_j X_j] = \sum_{j=1}^{n} \alpha_j^2 V[X_j] + 2\sum_{i<j} \alpha_i \alpha_j Cov(X_i, X_j)$

Corolario

Si $X_1, X_2,..., X_n$ s**on independientes**, entonces $E[\sum_{j=1}^{n} \alpha_j X_j] = \sum_{j=1}^{n} \alpha_j E[X_j]$ y $V[\sum_{j=1}^{n} \alpha_j X_j] = \sum_{j=1}^{n} \alpha_j^2 V[X_j]$.

En particular $E[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} E[X_j]$ y $V[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} V[X_j]$

Nota:

Si $X_1, X_2,..., X_n$ s**on independientes**, entonces $E(\prod_{j=1}^{n} X_j) = \prod_{j=1}^{n} E(X_j)$ pero no ocurre lo mismo con las varianzas, esto es: $V(\prod_{j=1}^{n} X_j) \neq \prod_{j=1}^{n} V(X_j)$

Proposición
Si A_{nxm} es matriz constante (o no aleatoria) entonces $E(A_{nxm}) = A_{nxm}$
Demostración
Basta aplicar la definición
$A_{nxm} = (a_{ij})$ donde cada componente a_{ij} es no aleatoria así que $E(a_{ij}) = a_{ij}$ y por tanto
$E(A_{nxm}) = (E(a_{ij})) = (a_{ij})_{nxm} = A_{nxm}$

Proposición
Si A_{nxq} es matriz de constante (o no aleatoria) y M_{qxm} es matriz aleatoria entonces
$E(A_{nxq}M_{qxm}) = A_{nxq}E(M_{qxm})$. O en notación más simple $E(AM) = AE(M)$.
Demostración
Especificando componentes tenemos $A_{nxq} = (a_{ik})_{nxq}$ y $M_{qxm} = (M_{kj})_{qxm}$ de modo que recordando la definición de producto matricial, es $A_{nxq}M_{qxm} = \left(\sum_{k=1}^{q} a_{ik}M_{kj}\right)_{nxm}$.
Tomando valor esperado:
$E(A_{nxq}M_{qxm}) = \left(E\left(\sum_{k=1}^{q} a_{ik}M_{kj}\right)\right)_{nxm} = \left(\sum_{k=1}^{q} a_{ik}E(M_{kj})\right)_{nxm} = (a_{ik})_{nxq}(E(M))_{qxm} = A_{nxq}E(M_{qxm})$. Análogamente se prueba la siguiente propiedad

Proposición
Si C_{pxm} es matriz de constante (o no aleatoria) y M_{nxp} es matriz aleatoria entonces
$E(M_{nxp}C_{pxm}) = E(M_{nxp})C_{pxm}$. O en notación más simple $E(MC) = E(M)C$.
<u>En general</u> podemos escribir $E(AMC + B) = AE(M)C + B$ donde A, B y C son matrices no aleatorias y M es matriz aleatoria.

Proposición
Sea $\vec{X}_{nx1}$ vector columna con vector de medias $\mu_{\vec{X}}$ y matriz de varianza covarianza $\Sigma_{\vec{X}}$. Sea $\vec{Y}_{mx1}$ otro vector tal que $\vec{Y}_{mx1} = A_{mxn}\vec{X}_{nx1} + B_{mx1}$ donde A_{mxn} y B_{mx1} son no aleatorias.

Entonces:

- El vector de medias de $\vec{Y}_{mx1}$ es $\mu_{\vec{Y}} = A\mu_{\vec{X}} + B$
- La matriz de varianza covarianza de $\vec{Y}_{mx1}$ es $\Sigma_{\vec{Y}} = A\Sigma_{\vec{X}}A^T$

Demostración

- En el caso de $E(\vec{Y}_{mx1})$, o equivalentemente de $\mu_{\vec{Y}}$: $E(\vec{Y}_{mx1}) = E(A_{mxn}\vec{X}_{nx1} + B_{mx1}) = A_{mxn}E(\vec{X}_{nx1}) + B_{mx1}$
  $= A_{mxn}\mu_{\vec{X}} + B_{mx1}$

- Para $\Sigma_{\vec{Y}}$:
  $$\Sigma_Y = E\left[(Y_{mx1} - \mu_Y)(Y_{mx1} - \mu_Y)^T\right]$$
  $$= E\left[(A_{mxn}\vec{X}_{nx1} + B_{mx1} - A_{mxn}\mu_{\vec{X}} - B_{mx1})(A_{mxn}\vec{X}_{nx1} + B_{mx1} - A_{mxn}\mu_{\vec{X}} - B_{mx1})^T\right]$$
  $$= E\left[(A_{mxn}\vec{X}_{nx1} - A_{mxn}\mu_{\vec{X}})(A_{mxn}\vec{X}_{nx1} - A_{mxn}\mu_{\vec{X}})^T\right]$$
  $$= E\left[(A_{mxn}(\vec{X}_{nx1} - \mu_{\vec{X}}))(A_{mxn}(\vec{X}_{nx1} - \mu_{\vec{X}}))^T\right] = E\left[A_{mxn}(\vec{X}_{nx1} - \mu_{\vec{X}})(\vec{X}_{nx1} - \mu_{\vec{X}})^T A_{nxm}^T\right]$$
  $$= A_{mxn}E\left[(\vec{X}_{nx1} - \mu_{\vec{X}})(\vec{X}_{nx1} - \mu_{\vec{X}})^T\right]A_{nxm}^T = A_{mxn}\Sigma_{\vec{X}}A_{nxm}^T$$

Ejemplo 3.20
Se tiene tres instrumentos financieros con rentabilidades anuales porcentuales: X ~ N(6,9); Y ~ N(10,25) y Z ~ N(10,16), de modo que la matriz de varianza covarianza del vector $(X,Y,Z)^T$ es $\begin{pmatrix} 9 & -1 & 3 \\ -1 & 25 & -3 \\ 3 & -3 & 16 \end{pmatrix}$. Si Ud. invierte 100 u.m. de modo que en X y en Y pone 30 u.m. y el resto lo pone en Z. Halle la ganancia esperada de su inversión así como su varianza.

Solución:

La ganancia U es $U = 0.3X + 0.3Y + 0.4Z = (0.3 \quad 0.3 \quad 0.4)\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ y aplicando la proposición:

$$E[U] = (0.3 \quad 0.3 \quad 0.4)\begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix} = (0.3 \quad 0.3 \quad 0.4)\begin{pmatrix} 6 \\ 10 \\ 10 \end{pmatrix} = 8.8 \text{ y}$$

$$V[U] = (0.3 \quad 0.3 \quad 0.4)\begin{pmatrix} 9 & -1 & 3 \\ -1 & 25 & -3 \\ 3 & -3 & 16 \end{pmatrix}(0.3 \quad 0.3 \quad 0.4)^T = (3.6 \quad 6 \quad 6.4)\begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix} = 5.44$$

Propiedad
Si $\vec{X}_{nx1}$ tiene componentes con distribución normal con vector de medias $\mu_{\vec{X}}$ y matriz de varianza covarianza $\Sigma_{\vec{X}}$ y $\vec{Y}_{mx1} = A_{mxn}\vec{X} + B_{mx1}$ entonces $\vec{Y}_{mx1}$ tiene componentes con distribución normal con $\mu_{\vec{Y}} = A\mu_{\vec{X}} + B$ y $\Sigma_{\vec{Y}} = A\Sigma_{\vec{X}}A^T$

3.4.3 Valor Esperado Condicional

La noción de vector de medias y matriz de varianza covarianza se puede aplicar en las distribuciones condicionales. En este contexto, es mejor usar la notación de vector particionado

Sea $X_{nx1} = (X_1, \cdots, X_p, X_{p+1}, \cdots, X_n)^T$ vector aleatorio, $X_{nx1} \in R^n$, donde las primeras p componentes forman un subvector $X_{px1} = (X_1, \cdots, X_p)^T$, $X_{px1} \in R^p$ y las demás (n-p) componentes forman un subvector $X_{(n-p)x1} = (X_{p+1}, \cdots, X_n)^T$, $X_{(n-p)x1} \in R^{n-p}$. En este contexto escribimos $X_{nx1} = (X_1 \vdots X_2)^T$ donde $X_1 = (X_1, \cdots, X_p)^T$ y $X_2 = (X_{p+1}, \cdots, X_n)^T$
Sea $f_{X_1|X_2}(x_1 | x_2)$ la distribución condicional del subvector $X_1 \in R^p$, dado el subvector $X_2 \in R^{n-p}$, esto es $f_{X_1|X_2}(x_1 | x_2) = \dfrac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}$.

Definición
La esperanza condicional de X_1 dado X_2, denotada $E(X_1 | X_2)$, es el vector de medias de X_1 en la distribución condicional $f_{X_1|X_2}(x_1 | x_2)$.
También usaremos las notaciones $E(X_1 | X_2) = \mu_{X_1|X_2}$ y en cualquier caso se trata de un vector $E(X_1 | X_2) \in R^p$, que por extensión del caso bidimensional, llamamos "Regresión de X_1 sobre X_2".

$E(X_1 | X_2) = \mu_{X_1|X_2}$ se puede usar para pronosticar X_1 en función de X_2

Definición

La matriz de varianza covarianza condicional de X_1 dado X_2, denotada $\Sigma_{X_1|X_2}$, es la matriz de varianza covarianza de X_1 en la distribución condicional $f_{X_1|X_2}(x_1 \mid x_2)$. La calculamos entonces, mediante

$$\Sigma_{X_1|X_2} = E\left[\left(X_1 - \mu_{X_1|X_2}\right)\left(X_1 - \mu_{X_1|X_2}\right)^T \mid X_2\right]$$

Esta matriz mide las variabilidades de las p componentes de X_1 cuando las predecimos y explicamos a partir de las componentes de X_2. También explica cómo se relacionan las p componentes de X_1 a causa de su relación con X_2 y de las interacciones entre las componentes de X_2.

En Economía, $E\left(X_1 \mid X_2\right) = \mu_{X_1|X_2}$ vendría a ser un modelo de ecuaciones simultáneas que formaliza una teoría explicativa de las variables económicas X_1, llamadas endógenas, partiendo de otro conjunto de variables X_2 llamadas exógenas. En las componentes de $\Sigma_{X_1|X_2}$ están reflejadas las interacciones entre las variables endógenas X_1 generadas o explicadas por las variables exógenas X_2.

Ejemplo 3.21

Sea $X \in R^4$ v.a. con vector de medias $\mu = \begin{pmatrix} 6 \\ 2 \\ 6 \\ 5 \end{pmatrix}$ y $\Sigma = \begin{pmatrix} 9 & 5 & 4 & -2 \\ 5 & 8 & 3 & 0 \\ 4 & 3 & 4 & 0 \\ -2 & 0 & 0 & 5 \end{pmatrix}$ su matriz de varianza covarianza. Sean $X_1 = (X_1, X_2)^T \in R^2$ el vector de las dos primeras componentes de X y $X_2 = (X_3, X_4)^T \in R^2$ el vector de las dos siguientes componentes de X

Entonces por identificación de componentes tenemos que:

a) El vector de medias y la matriz de varianza covarianza de $X_1 \in R^2$ son

$$\mu_{X_1} = \begin{pmatrix} 6 \\ 2 \end{pmatrix} \text{ y } \Sigma_{X_1} = \begin{pmatrix} 9 & 5 \\ 5 & 8 \end{pmatrix}$$

b) El vector de medias y la matriz de varianza covarianza de $X_2 \in R^2$ son $\mu_{X_2} = \begin{pmatrix} 6 \\ 5 \end{pmatrix}$ y $\Sigma_{X_2} = \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix}$

c) La matriz de covarianzas $\Sigma_{X_1X_2}$ de $X_1 \in R^2$ y $X_2 \in R^2$ es $\Sigma_{X_1X_2} = \begin{pmatrix} 4 & -2 \\ 3 & 0 \end{pmatrix}$

d) Sean las variables aleatorias $Y_1 = X_1 + 2X_2$ e $Y_2 = 2X_1 - X_2$, donde X_1 y X_2 son las componentes de $X_1 \in R^2$. Si definimos $Y = (Y_1, Y_2)^T \in R^2$, entonces es fácil hallar el vector de medias y la matriz de varianza covarianza de Y, si escribimos matricialmente la relación entre $Y \in R^2$ y $X_1 \in R^2$. En efecto, como $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 + 2X_2 \\ 2X_1 - X_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, llamando $A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$ aplicamos la linealidad del valor esperado matricial y obtenemos

$\mu_Y = E(Y) = E\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} E\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$ y la matriz de varianza covarianza

$$\Sigma_Y = E\left((Y - \mu_Y)(Y - \mu_Y)^T\right) = E\left((AX_1 - A\mu_{X_1})(AX_1 - A\mu_{X_1})^T\right) = E\left(A(X_1 - \mu_{X_1})(X_1 - \mu_{X_1})^T A^T\right)$$

$$AE\left((X_1-\mu_{X_1})(X_1-\mu_{X_1})^T\right)A^T = A\Sigma_{X_1}A^T = \begin{pmatrix}1 & 2\\ 2 & -1\end{pmatrix}\begin{pmatrix}9 & 5\\ 5 & 8\end{pmatrix}\begin{pmatrix}1 & 2\\ 2 & -1\end{pmatrix} = \begin{pmatrix}19 & 21\\ 13 & 2\end{pmatrix}\begin{pmatrix}1 & 2\\ 2 & -1\end{pmatrix}$$
$$= \begin{pmatrix}840 & 17\\ 17 & 24\end{pmatrix}$$

e) Sea ahora $Y_3 = X_3 + X_4$ y definamos $Z = (Y_1, Y_2, Y_3)^T \in R^3$, con Y_1 e Y_2 escritas como antes, a las que se agrega $Y_3 = X_3 + X_4$. Podemos escribir matricialmente que $Z = \begin{pmatrix}Y_1\\ Y_2\\ Y_3\end{pmatrix} = \begin{pmatrix}1 & 2 & 0 & 0\\ 2 & -1 & 0 & 0\\ 0 & 0 & 1 & 1\end{pmatrix}\begin{pmatrix}X_1\\ X_2\\ X_3\\ X_4\end{pmatrix}$, así si $D = \begin{pmatrix}1 & 2 & 0 & 0\\ 2 & -1 & 0 & 0\\ 0 & 0 & 1 & 1\end{pmatrix}$ entonces $Z = DX$ y por propiedades del valor esperado matricial el vector de medias de Z es $E(Z) = DE(X) = D\mu_X$, con matriz de varianza covarianza $V(Z) \equiv \Sigma_z = D\Sigma_X D^T$