Varianza

- **V.a. continuas:** integrar

$$\sigma^2 = \int_{\Omega} (x - \mu)^2 f(x)dx$$

- **Propiedades:**

$\sigma^2 = E(X^2) - E(X)^2$

$V(aX+b) = a^2 \sigma^2 (X)$

Distribuciones Truncadas

- Ej: *"La longitud de las varillas fabricadas por una máquina es una variable aleatoria X distribuida según f_X. Si nos quedamos solamente con las varillas que miden más de 2cm, ¿cómo se distribuye la longitud de las varillas que quedan?"*

- La nueva distribución excluye la probabilidad de la condición, por lo que la nueva distribución no suma/integra 1. Es necesario corregir la distribución dividiendo por la probabilidad de la condición y excluyendo los valores descartados.

Distribuciones Truncadas(2)

- Ej (cont):

$$f_X(x) = \begin{cases} \frac{1}{4} & 1 < x < 5 \\ 0 & \forall \; otro \; x \end{cases}$$

- Condición: X>2

$$\mathbb{P}(X > 2) = \int_2^{\infty} f(x)dx = \int_2^5 \frac{1}{4} dx = \frac{3}{4}$$

- Se excluyen los valores descartados:

$$\begin{cases} \frac{1}{4} & 2 < x < 5 \\ 0 & \forall \; otro \; x \end{cases}$$

Distribuciones Truncadas(3)

- Ej (cont): Se corrige la distribución dividiendo por la probabilidad del suceso:

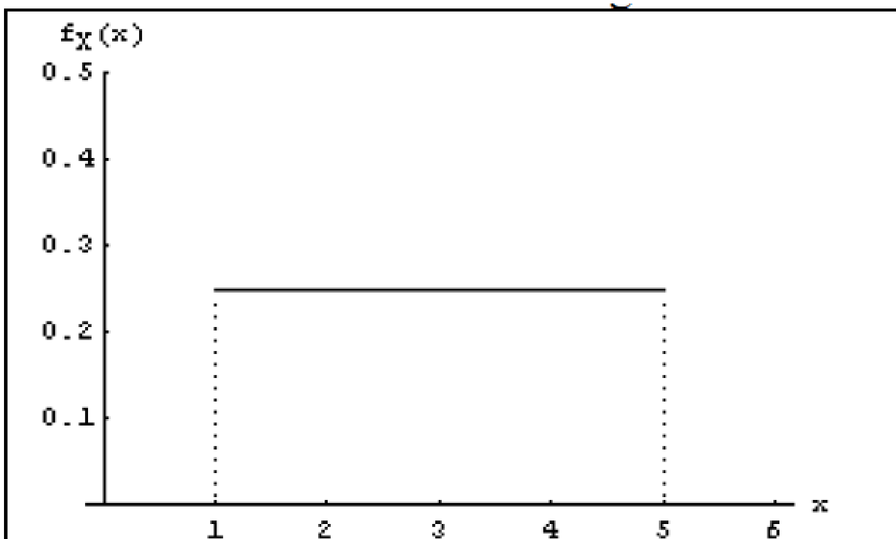
$$f_{x|x>2} = \begin{cases} \dfrac{1/4}{a} & 2 < x < 5 \\ 0 & \forall\ otro\ x \end{cases} = \begin{cases} \dfrac{1}{3} & 2 < x < 5 \\ 0 & \forall\ otro\ x \end{cases}$$

- Verificamos que es efectivamente una distribución:

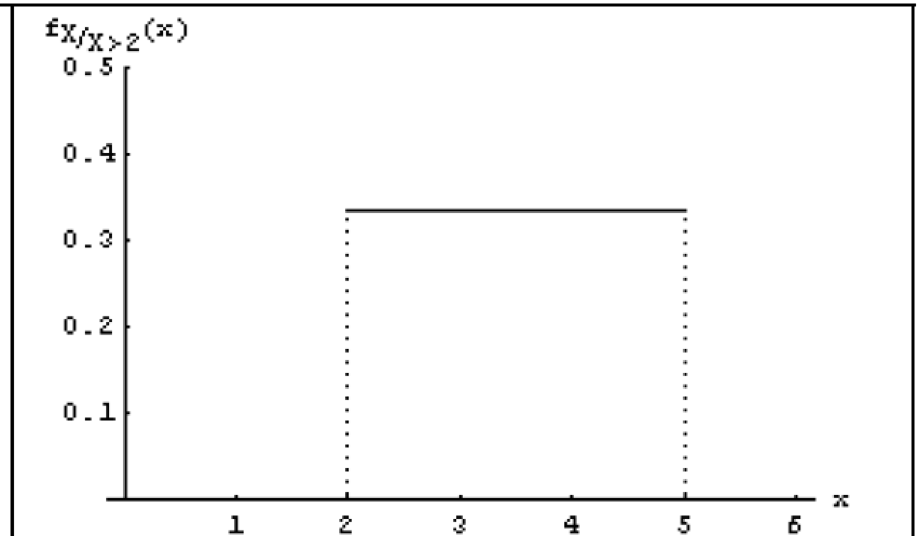
$$\int_{\Omega} f_{x|x>2}(x)dx = \int_{2}^{5} \frac{1}{3} dx = 1$$

Distribuciones Truncadas(4)

- Gráficamente:



$f_X(x)$



$f_{X/X>2}(x)$

Distribuciones Truncadas(5)

- Ej 2(v.a. discretas): Se tienen piezas de tipo 1, 2, 3 y 4, ubicadas, mezcladas, en una caja. El experimento consiste en tomar una pieza al azar de la caja. Hay un 20% de piezas tipo 1, 40% de tipo 2, 3% de tipo 3, y 10% de tipo 4. Luego alguien se toma el trabajo de quitar todas las piezas tipo 3 de la caja. ¿Cómo se distribuye X ahora?

- Se tiene la siguiente distribución:

$$P_X(x) = \begin{cases} 2/10 & x = 1 \\ 4/10 & x = 2 \\ 3/10 & x = 3 \\ 1/10 & x = 4 \\ 0 & \forall \ otro \ x \end{cases}$$

Distribuciones Truncadas(5)

- Ej 2: condición X ≠ 3 $\mathbb{P}(X \neq 3) = \frac{7}{10}$

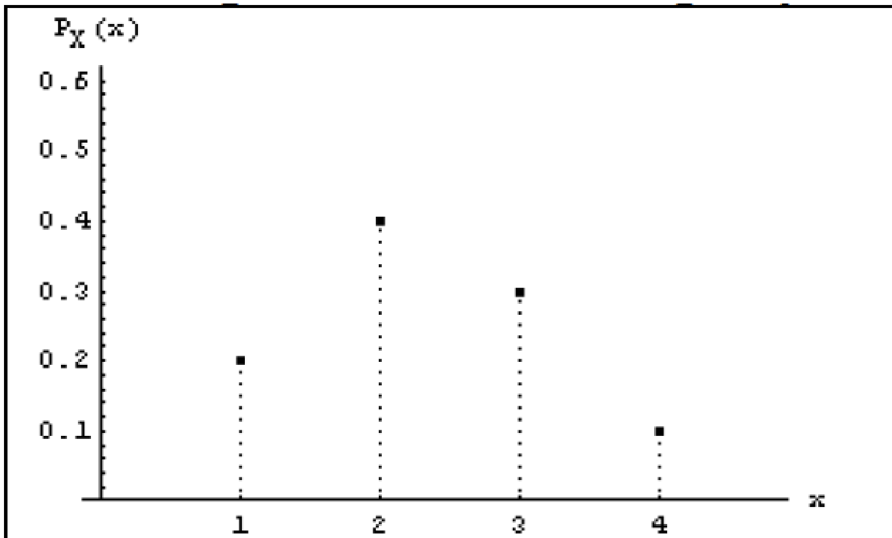
- Función con dominio restringido

$$P_X(x) = \begin{cases} 2/10 & x = 1 \\ 4/10 & x = 2 \\ 1/10 & x = 4 \\ 0 & \forall \; otro \; x \end{cases}$$

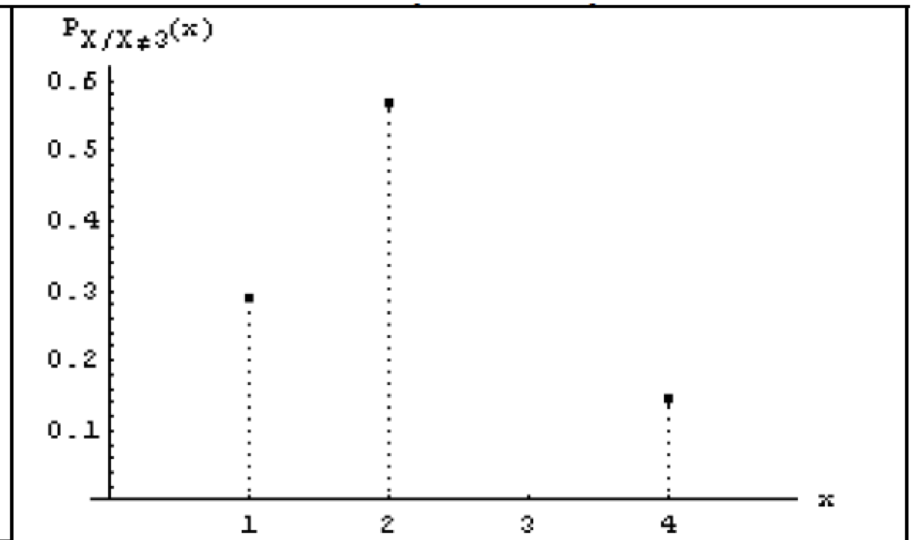
- Corrección de la distribución: $P_{X / X\neq 3}(x) = \begin{cases} 2/7 & x = 1 \\ 4/7 & x = 2 \\ 1/7 & x = 4 \\ 0 & \forall \; otro \; x \end{cases}$

Distribuciones Truncadas(6)

- Gráficamente:



$P_X(x)$



$P_{X/X\neq 3}(x)$

Distribución conjunta

- Se deben cumplir los mismos supuestos que para variables unidimensionales:

$$P(x_i, y_j) \geq 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) = 1 \qquad \int_{\Omega} \int_{\Omega'} f(x,y)dxdy = 1$$

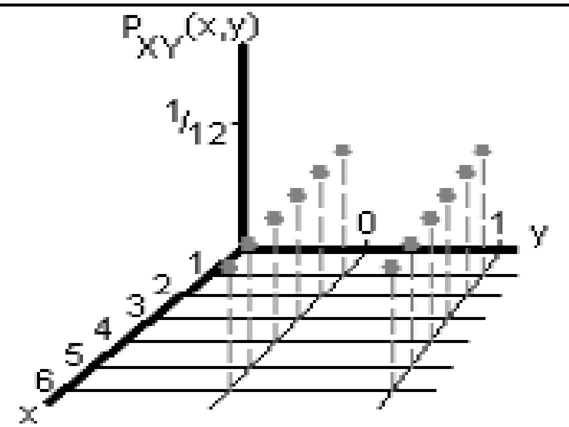
- La interpretación de la función de densidad conjunta es una función bidimensional.

Distribución conjunta(2)

- Ej: X: el número que sale al tirar un dado.

  Y: la cantidad de caras que salen al tirar una moneda.

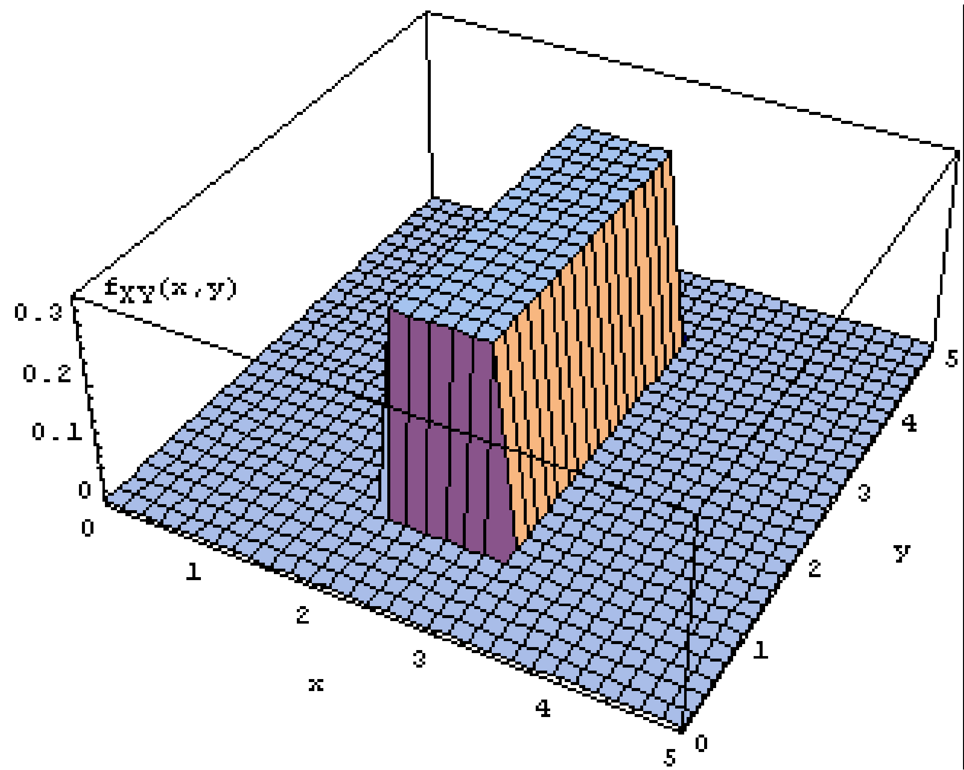
		Y	
	P_{XY}	0	1
X	1	$^1/_{12}$	$^1/_{12}$
	2	$^1/_{12}$	$^1/_{12}$
	3	$^1/_{12}$	$^1/_{12}$
	4	$^1/_{12}$	$^1/_{12}$
	5	$^1/_{12}$	$^1/_{12}$
	6	$^1/_{12}$	$^1/_{12}$



Distribución conjunta(2)

- Ej 2: Caso continuo:

$$f_{XY}(x,y)=\begin{cases}\frac{1}{3} & 2<x<3,1<y<4\\ 0 & \forall \text{ otro } x\end{cases}$$



Distribución Marginal

- Cada componente de una variable aleatoria bidimensional es una variable aleatoria unidimensional en sí misma.

- Nos puede interesar conocer la distribución de una componente por separado, sin tener en cuenta a la otra componente.

- Eso se denomina "marginar", y la distribución de la variable unidimensional por separado se llama "distribución marginal".

Distribución Marginal(2)

- (v.a. discretas) Sea la variable aleatoria bidimensional XY distribuida según $P_{XY}(x,y)$, la distribución marginal de X es:

$$P_X(x_i) = \sum_{j=1}^{m} P(x_i, y_j)$$

- Análogamente, la distribución marginal de Y es:

$$P_Y(y_j) = \sum_{i=1}^{n} P(x_i, y_j)$$

Distribución Marginal(3)

- (v.a. continuas) Sea la variable aleatoria bidimensional XY distribuida según $f_{XY}(x,y)$, la distribución marginal de X es:

$$f_X(x) = \int_{\Omega'} f(x,y)\,dy$$

- Análogamente, la distribución marginal de Y es:

$$f_Y(y) = \int_{\Omega} f(x,y)\,dx$$

Distribución Marginal(4)

- **Ej 1:**

Si la distribución conjunta es:

P_{XY}		**Y**	
		20	**30**
X	**1**	0,1	0,3
	2	0,4	0,2

Vamos a hallar la distribución de X.

Primero enumeramos los valores posibles de X: 1; 2.
Y ahora para cada valor posible de X, aplicamos la fórmula.

$$P_X(1) = \sum_{y=-\infty} P_{XY}(1, y) = P_{XY}(1{,}20) + P_{XY}(1{,}30) = 0{,}1 + 0{,}3 = 0{,}4$$

$$P_X(2) = \sum_{y=-\infty} P_{XY}(2, y) = P_{XY}(2{,}20) + P_{XY}(2{,}30) = 0{,}4 + 0{,}2 = 0{,}6$$

Entonces obtuvimos:

$$P_X(x) = \begin{cases} 0{,}4 & x = 1 \\ 0{,}6 & x = 2 \\ 0 & \forall \text{ otro } x \end{cases}$$

Distribución Marginal(5)

Ahora hallemos la distribución de Y:
Primero enumeramos los valores posibles de Y: 20; 30.
Y ahora para cada valor posible de X, aplicamos la fórmula.

$$P_Y(20) = \sum_{x=-\infty}^{+\infty} P_{XY}(x,20) = P_{XY}(1,20) + P_{XY}(2,20) = 0{,}1 + 0{,}4 = 0{,}5$$

$$P_Y(30) = \sum_{x=-\infty}^{+\infty} P_{XY}(x,30) = P_{XY}(1,30) + P_{XY}(2,30) = 0{,}3 + 0{,}2 = 0{,}5$$

Entonces obtuvimos:

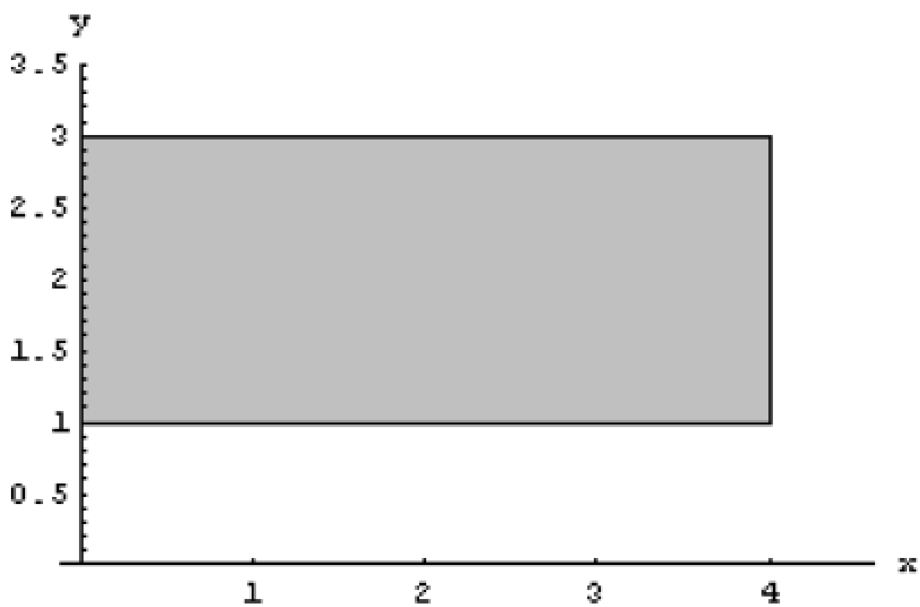
$$P_Y(y) = \begin{cases} 0{,}5 & y = 20 \\ 0{,}5 & y = 30 \\ 0 & \forall \text{ otro } y \end{cases}$$

P_{XY}		Y		
		20	**30**	
X	**1**	0,1	0,3	*0,4*
	2	0,4	0,2	*0,6*
		0,5	*0,5*	

Distribución Marginal(5)

- Ej 2

$$f_{XY}(x,y)=\begin{cases}\dfrac{1}{8} & 0<x<4\,,1<y<3\\ 0 & \forall\, otro\ x,y\end{cases}$$



Distribución Marginal(6)

- Ej 2: Distribución marginal de X: se aplica la fórmula al único intervalo relevante (0 < x < 4):

$$\int_{-\infty}^{+\infty} f_{XY}(x, y)\, dy = \int_{1}^{3} \frac{1}{8}\, dy = \frac{1}{4}$$

$$f_X(x) = \begin{cases} 1/4 & 0 < x < 4 \\ 0 & \forall\ otro\ x \end{cases}$$

Distribución Marginal(7)

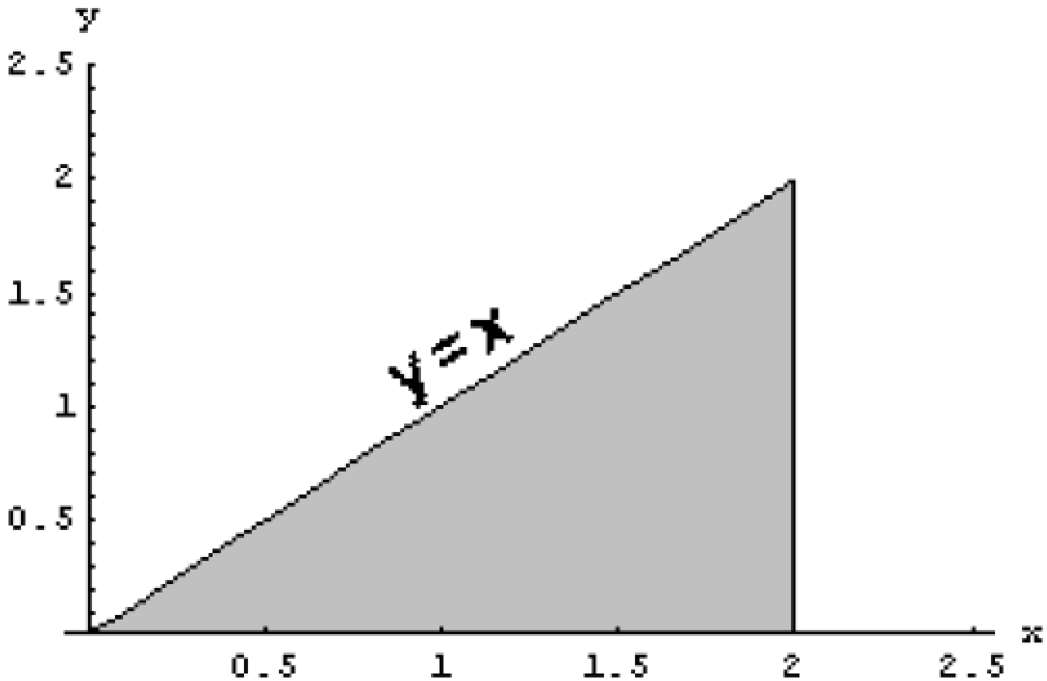
- Ej 2: Distribución marginal de Y: se aplica la fórmula al único intervalo relevante (1 < y < 3):

$$\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx = \int_{0}^{4} \frac{1}{8}\,dx = \frac{1}{2}$$

$$f_Y(y) = \begin{cases} 1/2 & 1 < y < 3 \\ 0 & \forall\ otro\ y \end{cases}$$

Distribución Marginal(8)

- Ej 3:



$$f_{XY}(x,y) = \begin{cases} \dfrac{x+y}{4} & 0 < x < 2\,,\, 0 < y < x \\ 0 & \forall\ otro\ x, y \end{cases}$$

Distribución Marginal(9)

- Ej 3: Distribución marginal de X: se aplica la fórmula al único intervalo relevante (0 <x< 2):

$$\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy = \int_{0}^{x} \frac{x+y}{4}\,dy = \frac{3x^2}{8} \qquad f_X(x) = \begin{cases} \dfrac{3x^2}{8} & 0 < x < 2 \\ 0 & \forall\ otro\ x \end{cases}$$

- Distribución marginal de Y: se aplica la fórmula al único intervalo relevante (0 < y < 2):

$$\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx = \int_{y}^{2} \frac{x+y}{4}\,dx = \frac{4+4y-3y^2}{8} \qquad f_Y(y) = \begin{cases} \dfrac{4+4y-3y^2}{8} & 0 < y < 2 \\ 0 & \forall\ otro\ y \end{cases}$$

Covarianza

- Introducimos el concepto de probabilidad conjunta, cuando dos v.a. pueden tomar simultáneamente ciertos valores concretos (caso discreto):

$$P(x_i, y_i) = [\mathbb{P}(X = x_i) \cap \mathbb{P}(Y = y_i)]$$

- La covarianza refleja la relación lineal de 2 v.a. X e Y:

$$\sigma\,(X,Y) = E(\mathrm{XY}) - E(\mathrm{X})\,E(\mathrm{Y})$$

Covarianza(2)

Donde:

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_i P(x_i, y_i)$$ caso v.a. discretas

$$E(XY) = \int_{\Omega}\int_{\Omega'} xy\, f(x,y)\,dx\,dy$$ caso v.a. continuas

- Correlación de Pearson:

$$\rho(X,Y) = \frac{\sigma(X,Y)}{\sigma(X)\sigma(Y)}$$

Distribuciones Condicionales.

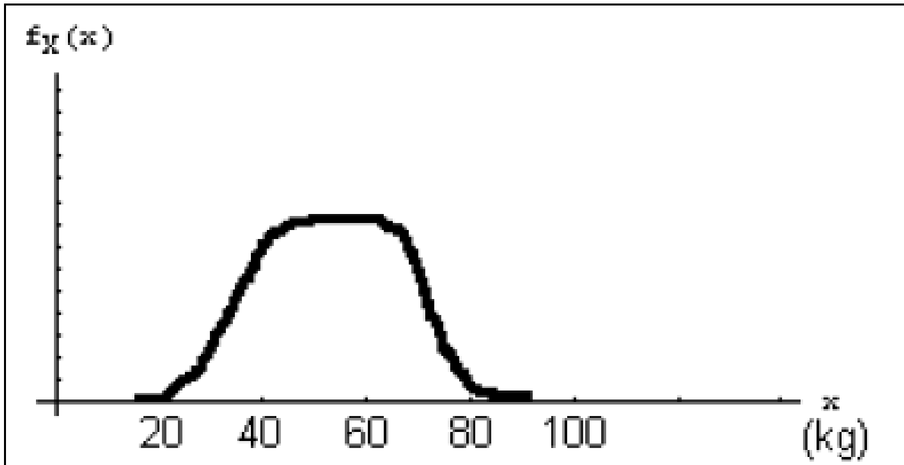
- Supongamos que tenemos las v.a. *peso* y *altura*.

- Intuitivamente, si conocemos el valor que tomó una de las v.a. al hacer el experimento, eso nos modificará la distribución de probabilidad de la otra variable aleatoria.

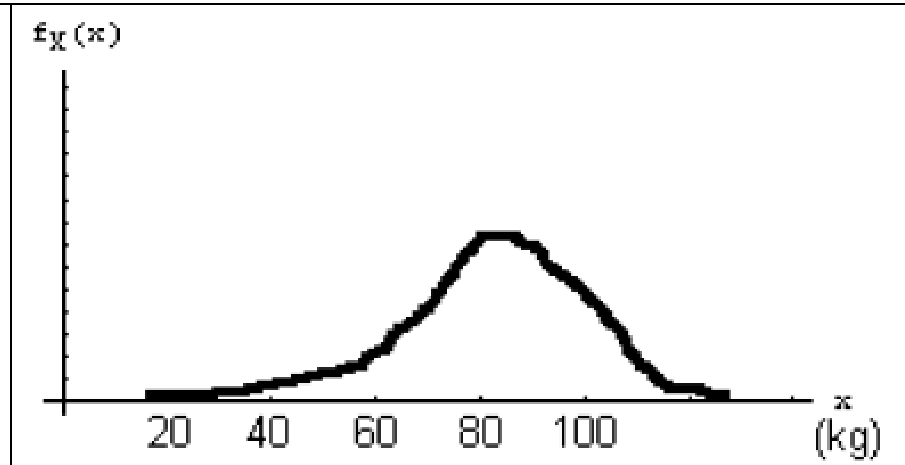
- Conociendo función de densidad conjunta de las dos variables aleatorias, podemos obtener la distribución marginal del peso. Pero si conociéramos que la variable altura tomó el valor 1.90m, ¿la distribución marginal del peso que teníamos sigue siendo válida?

Distribuciones Condicionales(2).

- No, Seguramente, la masa de probabilidad del peso tenderá a distribuirse más hacia los valores más altos:



Distribución marginal de la variable peso, es decir, sin saber nada de la altura.



Distribución de la variable peso, sabiendo que la variable altura tomó el valor 1.90

Distribuciones Condicionales(3).

- Se define entonces la función de densidad condicional:

- Caso 2 v.a. discretas X e Y:

$$P_{X/Y}(x,y)=\frac{P_{XY}(x,y)}{P_Y(y)}$$

- Caso 2 v.a. continuas X e Y:

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

Independencia de v.a.

- Intuitivamente, X e Y son independientes si $f_{X/Y}(x,y)$ es idéntica para todos los posibles valores de y.

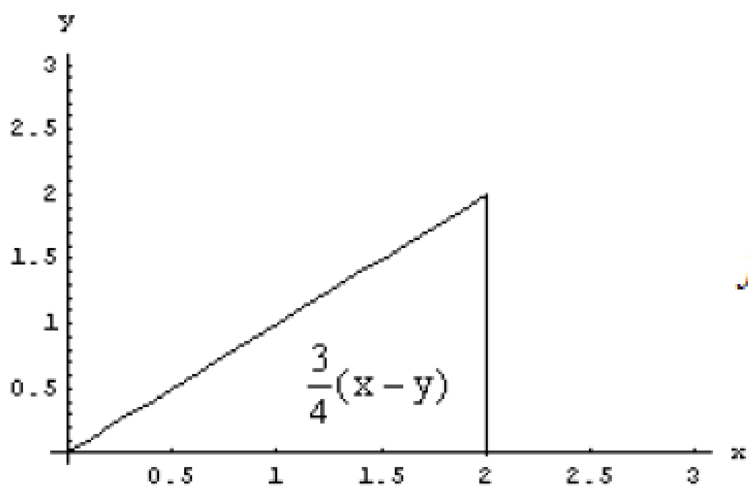
- Por ejemplo, que la distribución del peso es idéntica para los distintos valores de altura (claramente no se cumple).

- Si $f_{X/Y}(x,y)$ no depende de y, entonces es en realidad $f_X(x)$, es decir, la distribución marginal de X.

Independencia de v.a.(2)

Para X, Y variables aleatorias continuas:	Para X, Y variables aleatorias discretas:
X e Y son estadísticamente independientes <=> $f_{X/Y}(x,y) = f_X(x)$ <=> $f_{Y/X}(x,y) = f_Y(y)$ <=> $f_{XY}(x,y) = f_X(x) \,.\, f_Y(y)$	X e Y son estadísticamente independientes <=> $P_{X/Y}(x,y) = P_X(x)$ <=> $P_{Y/X}(x,y) = P_Y(y)$ <=> $P_{XY}(x,y) = P_X(x) \,.\, P_Y(y)$

- Ej 1:



$$f_{XY}(x,y) = \begin{cases} \frac{3}{4}(x-y) & 0 < x < 2,\, 0 < y < x \\ 0 & \forall\ otro\ x, y \end{cases}$$

Independencia de v.a.(3)

Marginamos:

$$f_X(x,y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^{x} \frac{3}{4}(x-y)\,dy = \frac{3}{8}x^2$$

lo cual vale para 0 < x < 2.

$$f_Y(x,y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_y^{2} \frac{3}{4}(x-y)\,dx = \frac{3}{2}(\frac{1}{4}y^2 - y + 1)$$

lo cual vale para 0 < y < 2. Tenemos entonces:

$$f_X(x) = \begin{cases} \frac{3}{8}x^2 & 0 < x < 2 \\ 0 & \forall\, otro\ x \end{cases} \qquad f_Y(y) = \begin{cases} \frac{3}{2}\left(\frac{1}{4}y^2 - y + 1\right) & 0 < y < 2 \\ 0 & \forall\, otro\ y \end{cases}$$

Independencia de v.a.(4)

Multiplicándolas se obtiene que el valor es:

$$\frac{3}{8}x^2\,\frac{3}{2}\left(\frac{1}{4}y^2-y+1\right)=\frac{9}{16}x^2\left(\frac{1}{4}y^2-y+1\right)$$

Y el dominio es $0 < x < 2 \cap 0 < y < 2$.

Se ve claramente que ni los valores ni el dominio coinciden con los de la función conjunta original. Luego, **X e Y no son independientes**.

Ej 2:

Tenemos las variables aleatorias discretas X e Y, cuya distribución conjunta es:

		Y		
X	**P_{XY}**	**1**	**2**	**3**
	1	0.08	0.12	0.2
	2	0.12	0.18	0.3

Independencia de v.a.(5)

Hallamos las distribuciones marginales:

X		Y			P_X
	P_{XY}	**1**	**2**	**3**	
	1	0.08	0.12	0.2	0.4
	2	0.12	0.18	0.3	0.6
P_Y		0.2	0.3	0.5	---

Si multiplicamos las distribuciones marginales obtenemos:

X		Y		
	$P_X P_Y$	**1**	**2**	**3**
	1	0.08	0.12	0.2
	2	0.12	0.18	0.3

Vemos que $P_X P_Y = P_{XY}$ $\forall$ x, y. Por lo tanto X e Y **son independientes**.