Tema 3
Variables Aleatorias y Distribuciones de Probabilidad

- Variable aleatoria: Definición
- Variable aleatoria Discreta (VAD) y Continua (VAC)
- Propiedades de una Función de Distribución de Probabilidad de VAD f(x)
- Función de distribución acumulada F(x)
- Función de Densidad de Probabilidad de una VAC
- Propiedades de una f.d.p
- Relación entre F(x) y f(x)
- Distribución de Probabilidad Distribución Conjunta (f.d.c)
- Propiedades de una f.d.c
- Distribuciones Marginales
- Distribución de Probabilidad Condicional de una f.d.c
- Independencia Estadística

Variable Aleatoria: Definición

Es una función que asocia un número real a cada elemento o subconjunto (evento) del Espacio Muestral.

Este valor es el resultado de realizar un experimento y tiene por tanto una connotación aleatoria ya que no están controladas todas las variables (eventos fortuitos) ó se realiza ex profeso al azar.

Evento → subconjunto del Espacio Muestral ⇒ Variable Aleatoria

E	Asigno	X
P(E)	Asigno	$P(X = x)$

Toda la terna de conjuntos es válida para la variable aleatoria.

Variable Aleatoria Discreta (VAD)

Se puede contar conjunto de resultados posibles (Escala discreta) aunque el número de elementos sea infinito (Ej: Arrojar un dado muchas veces)

Variable Aleatoria Continua (VAC)

No es posible contar el número de posibilidades que puede tomar un valor de la VAC. (Escala Continua).
Ej: medición de temperaturas.

VAD: la VAD toma c/u de sus valores con cierta probabilidad según el experimento y el evento asociado.

Ej: experimento: seleccionar 3 artículos de una cadena de fabricación.

Observación: clasificar como Def (D) o No Def. (N)
N° de elementos = $2^3 = 8$
S = {DDD, DDN, DND, DNN, NDD, NDN, NND, NNN}
Variable aleatoria X = N° de artículos defectuosos.

Eventos:
A:{N° de artículos defectuosos igual a 3} $X = 3$

$P(A) \rightarrow P(X = 3)$

B:{N° de artículos defectuosos mayor o igual a 2} $x \geq 2$

$P(B) = P(X \geq 2)$

.-etc.

SI X es una VAD que representa el N° de artículos defectuosos:

X = 0 con P(X) = $\frac{1}{8}$ $\Rightarrow$ P(X = 0) = $\frac{1}{8}$
X = 1 con P(X) = $\frac{3}{8}$ $\Rightarrow$ P(X = 1) = $\frac{3}{8}$
X = 2 con P(X) = $\frac{3}{8}$ $\Rightarrow$ P(X = 2) = $\frac{3}{8}$
X = 3 con P(X) = $\frac{1}{8}$ $\Rightarrow$ P(X = 3) = $\frac{1}{8}$

$$\sum_{\forall x} P(X = x) = 1$$

Se puede también utilizar una expresión matemática ligada a la probabilidad de la VA donde ***f(x) = P(X = x)***

El conjunto ordenado *(x, f(x))* se llama Función de Probabilidad o Distribución de Probabilidad (fdp)

Propiedades de una fdp de VAD

1. $f(x) \geq 0$
2. $\sum_{\forall x} f(x) = 1$
3. $P(X = x) = f(x)$

<u>Distribución de probabilidad:</u> Tabla con los pares (x, fx)

X	f(x)
0	1/8
1	3/8
2	3/8
3	1/8

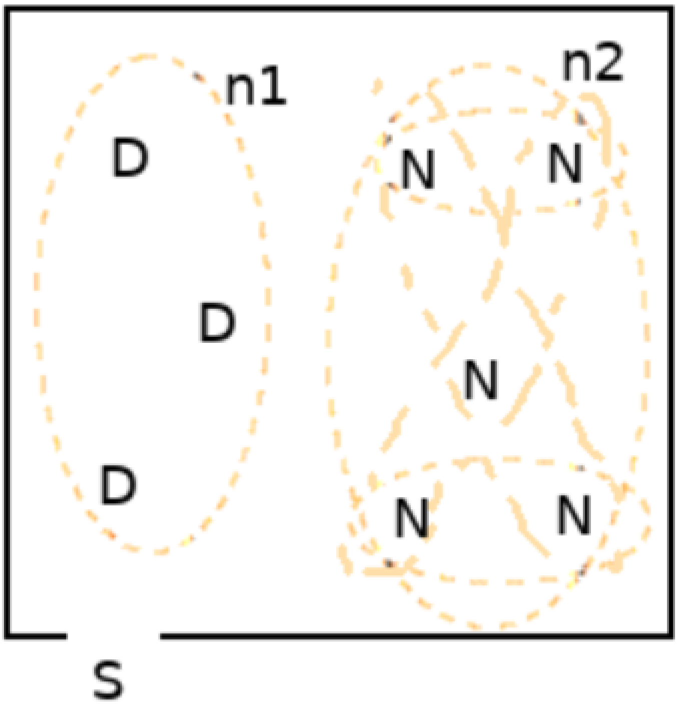
Ejemplo de tomar 3 artículos y clasificarlos como defectuosos o no defectuosos.

Definir *X* = N° de artículos defectuosos.

Ej. 2: un embarque de 8 computadoras similares para una tienda contiene 3 defectuosos. Una escuela compra 2 computadoras en esa tienda.
Encuentre la Distribución de Probabilidad para el N° de computadoras defectuosas que compra la escuela. (Comparar con el caso de fabricación en serie. No es lo mismo)

P(X) = $\frac{n}{N}$

X: N° de computadoras defectuosas en la compra de la escuela.



a) No compra ninguna defectuosa.

$$P(x) = \frac{n}{N} = \frac{n_1 \, . n_2}{N}$$

$X = 0 \Rightarrow f(x = 0) = ?$
$n_1 : 0 \; de \; 3, \quad n_2 : \; 2 \; de \; 5$
$N : 2 \; de \; 8$

$$f(0) = \frac{n_1 \, . n_2}{N} = \frac{\binom{3}{0} \cdot \binom{5}{2}}{\binom{8}{2}} = \frac{\frac{3!}{0!\,(3!)} \cdot \frac{5!}{2!\,(3!)}}{\frac{8!}{2!\,(6!)}}$$

$$f(0) = \frac{1 \, . 10}{28}$$

⇒ Hay una sola forma de no tomar defectos. Hay 10 formas de tomar 2 no defectuosas de 5. Hay 28 formas de tomar 2 computadoras de 8 totales.

$$f(1) = \frac{\binom{3}{1} \cdot \binom{5}{1}}{\binom{8}{2}} = \frac{3 \, . 5}{28}$$

b)

$$f(2) = \frac{\binom{3}{2} \cdot \binom{5}{0}}{\binom{8}{2}} = \frac{3 \, . 1}{28}$$

Distribución de Probabilidad:

X	f(x)
0	10/28 = 0,357
1	15/28 = 0,536
2	3/28 = 0,107

Otra forma de plantear el problema es sacar primero un artículo y luego otro.

$$P(x) = 1° \, extracción \, . \, 2° \, extracción = \frac{n_1}{N_1} . \frac{n_2}{N_2}$$

n_1: *opciones de la primera extracción.*
N_1: *n° total de artículos primera extracción.*
n_2: *opciones de la segunda extracción.*
N_2: *n° total de artículos en la segunda extracción.*

$$P(X = 0): No \; sacar \; ninguna \; defectuosa$$

$$P(X = 0) = \frac{5}{8} . \frac{4}{7} = \frac{20}{56} = \frac{10}{28}$$

$$P(X = 1) = \frac{3}{8} . \frac{5}{7} + \frac{5}{8} . \frac{3}{7} = \frac{15}{56} + \frac{15}{56} = \frac{30}{56} = \frac{15}{28}$$

$$P(X = 2) = \frac{3}{8} . \frac{2}{7} = \frac{6}{56} = \frac{3}{28}$$

Este método da resultados iguales al anterior.

<u>Distribución acumulada</u>

La $F(X)$ de una VAD con distribución de probabilidad $f(x)$ es:

$$F(X) = P(X \leq x) = \sum_{t \leq x} f(t) \quad para \quad -\infty < x < +\infty$$

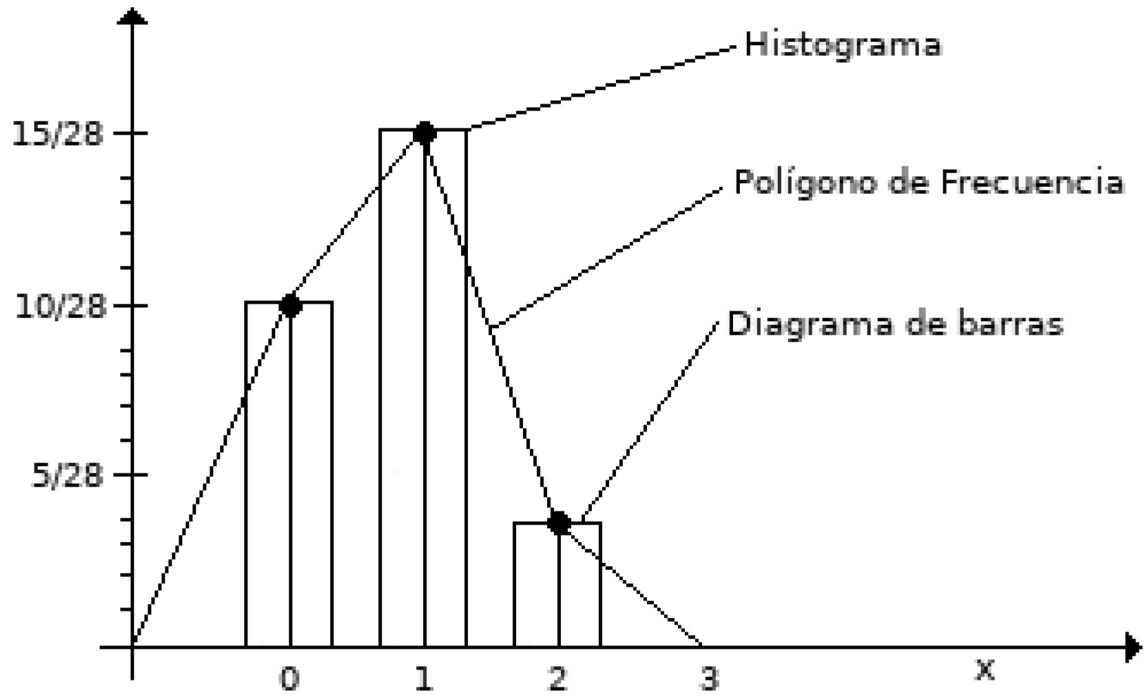
Para el ejemplo anterior: ¿Qué probabilidad existe de comprar <u>a lo sumo</u> 1 defectuosa?

$$F(X \leq 1) = \sum_{t \leq 1} f(t) = f(0) + f(1) = \frac{10}{28} + \frac{15}{28} = \frac{25}{28}$$

¿Qué probabilidad existe de comprar <u>al menos</u> 1 defectuosa?

$$F(X \geq 1) = 1 - \sum_{t \leq 1} f(t) = 1 - [\, f(0) + f(1)] = \frac{28}{28} - \frac{25}{28} = \frac{3}{28}$$

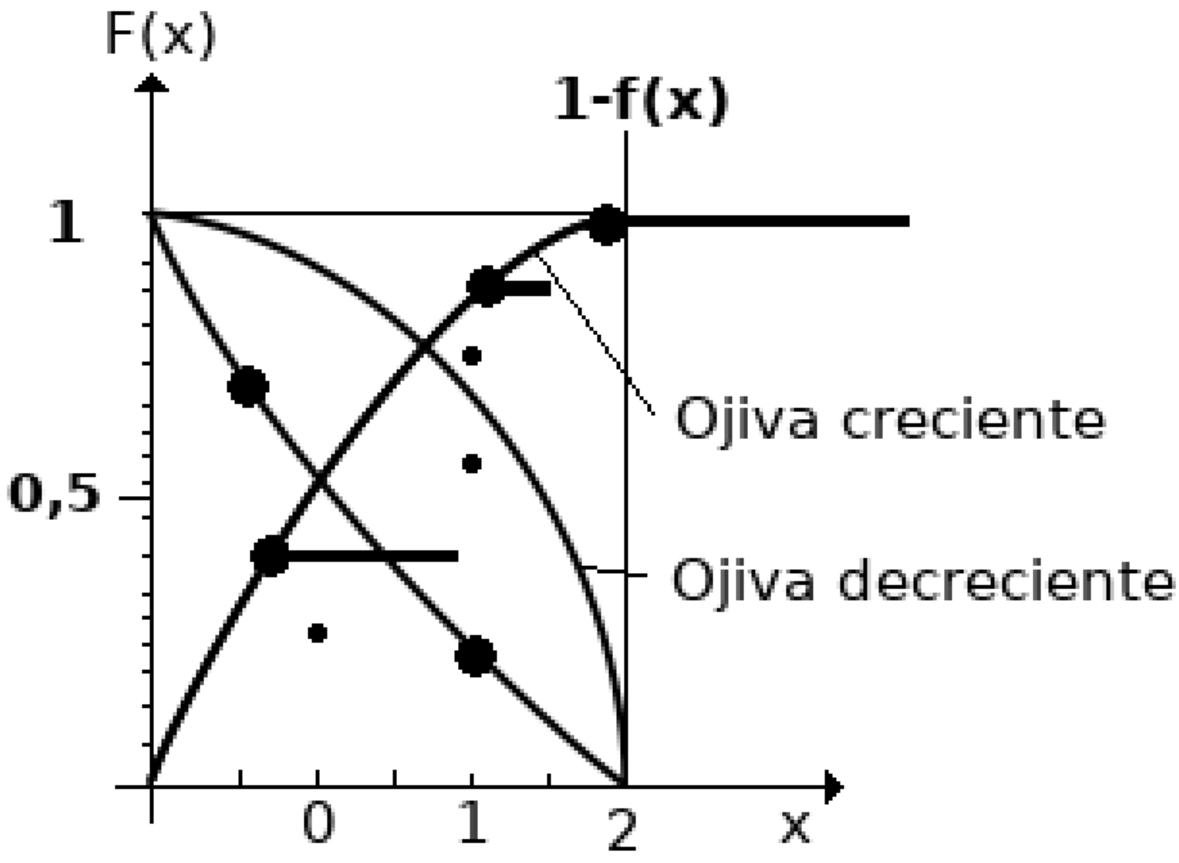
Gráfico de Barras → Polígono de Frecuencias



(x, f(x))

x	$f(x)$
0	0,357
1	0,536
2	0,107

Distribución Acumulada Discreta



(x, F(x))

x	f(x)
0	0,357
1	0,893
2	1

x	1 - f(x)
0	0,643
1	0,107
2	0

Distribuciones Continuas

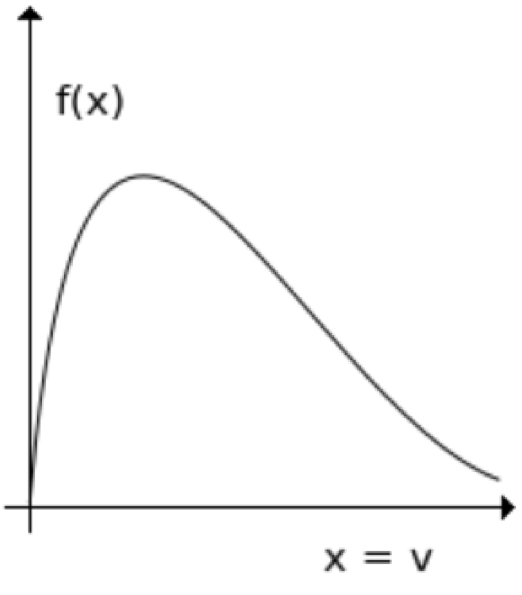
Una Variable Aleatoria Continua tiene una probabilidad P(X=x) = 0 de tomar exactamente cualquiera de sus valores.

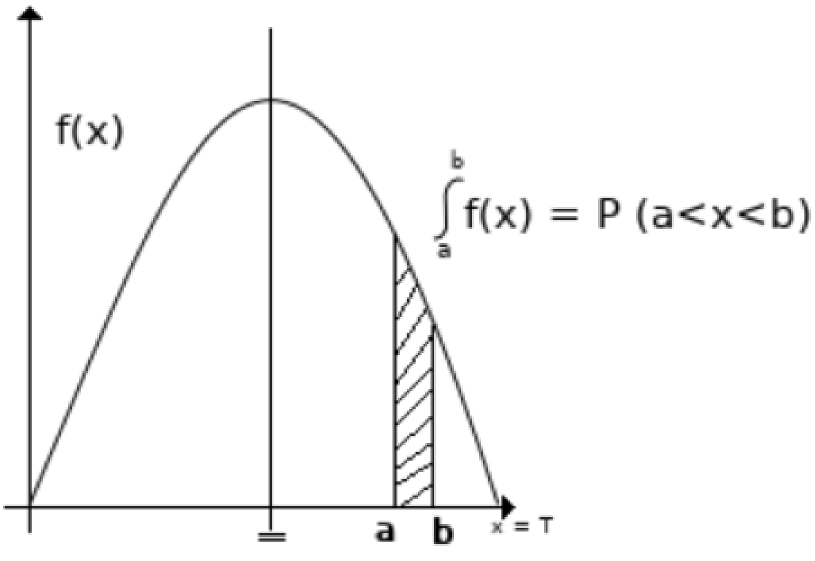
$$P_{(n)} = \frac{n}{N} \qquad si\ N \rightarrow \infty \qquad P_{(n)} \rightarrow \infty$$

Si en cambio se trata de un intervalo $(a < x < b)$

- Una Variable Aleatoria Continua no se puede presentar de forma tabular ⇒ se utiliza una función f(x)

f(x) en una Variable Aleatoria Continua se denomina: **Función de Densidad de Probabilidad.**





$$\boldsymbol{P(a < x < b) = \int_a^b f(x)dx}$$

Una función $f(x)$ es una <u>función de densidad de probabilidad</u> para una $VAC \in \mathbb{R}$ si:

1- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
2- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$
3- $P(a < x < b) = \int_a^b f(x)\,dx$

Ejemplo: sea $f(x)$

$$f(x) = \begin{cases} \frac{x^2}{3} & ; \quad -1 < x < 2 \\ 0 & ; \quad \text{para cualquier otro caso} \end{cases}$$

a) Comprobar que $f(x)$ $es\ una\ f.d.p.$

1) $f(x) \geq 0$ si $\frac{x^2}{3} \geq 0 \quad \forall\, x \in \mathbb{R}$ se cumple

2) $\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{+\infty} \frac{x^2}{3}\,dx = \frac{x^3}{9}$; $\big(para\ (x = 2) - (x = -1)\big) = \frac{8}{9} + \frac{1}{9} = 1$ ✔

b) Obtener $P(-1 < x < 1)$

$$P(-1 < x < 1) = \int_{-1}^{1} \frac{x^2}{3}\,dx = \frac{x^3}{9}\ ;\ (para\ (x = 1) - (x = -1)) = \frac{2}{9}$$

c) Obtener $P(x < 1{,}5)$

$$P(x < 1{,}5) = P(x < \frac{3}{2}) = \int_{-\infty}^{3/2} f(x)\,dx = \int_{-1}^{3/2} \frac{x^2}{3}\,dx$$

$$P(x < 1{,}5) = \frac{x^3}{9}\ ;\ (para\ (x = 3/2) - (x = -1)) = \frac{27}{9.8} + \frac{1}{9} = \frac{27}{72} + \frac{8}{72} = \frac{35}{72}$$

Función de distribución Acumulada de una VAC

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

Relación entre $F(x)$ y $f(x)$

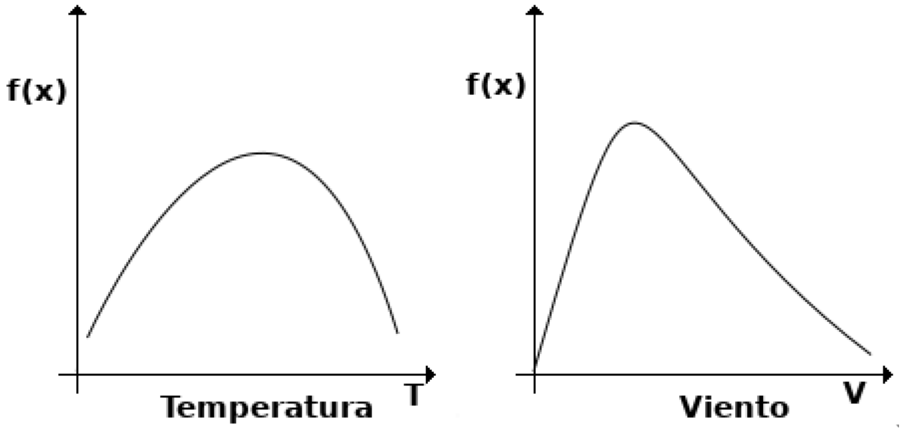
$$Si \quad F(x) = \int_{-\infty}^{x} f(t)\, dt \Rightarrow f(x) = \frac{d\,F(x)}{dx}$$

Ejemplo: Obtener $F(x)$ del ejemplo anterior

$$F(x) = \int_{-\infty}^{x} \frac{t^2}{3}\, dt = \int_{-1}^{x} \frac{t^2}{3}\, dt = \frac{x^3}{9} + \frac{1}{9} = \frac{x^3+1}{9}$$

<u>Distribuciones Empíricas</u>: A menudo no sabemos la forma de la distribución de probabilidad (caso discreto) o la función de distribución de probabilidad (caso continuo).

Ejemplo:



¿Cómo sabemos que tienen esta forma?

Se han realizado mediciones → se han tabulado y graficado este conjunto de mediciones (muestra) → se infiere según estas gráficas que el comportamiento de la variable (población) tiene esta forma.

<u>Distribuciones Conjuntas</u>

Espacios muestrales multidimensionales.

Propiedades

Discreta	Continua
$F(x,y) \geq 0\ ;\ \forall (x,y)$	$f(x,y) \geq 0\ ;\ f(x,y)$
$\sum_x \sum_y f(x,y) = 1;\ \forall (x,y)$	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\, dx\, dy = 1$
$P(X = x, Y = y) = f(x,y)$	$P(a < x < b, c < y < d) = \int_a^b \int_c^d f(x,y) dx\, dy$ $si\ \{a \leq x \leq b\ ;\ c \leq y \leq d\}$

$P[(x,y) \in \mathbb{R}] = \sum_{x}\sum_{y} f(x,y)$	

Distribuciones Marginales

Distribución de una variable cuando la otra toma todos los valores posibles.

$g(x) = \sum_y f(x,y)$ y $h(y) = \sum_x f(x,y)$ caso Discreto

$g(x) = \int_{y=-\infty}^{y=+\infty} f(x,y)\, dy$ y $h(x) = \int_{x=-\infty}^{x=+\infty} f(x,y)\, dx$ caso Continuo

Distribución de Probabilidad Condicional

De la definición: $P(B/A) = \frac{P(A \cap B)}{P(A)}$; $P(A) > 0$

A y B son dos eventos definidos por X = x; Y = y (Variables aleatorias asociadas a los eventos)

$P(Y = y \,/\, X = x) = \frac{P\,(X = x\,;\, Y = y)}{P(X = x)} = \frac{f(x\,,y)}{g(x)}$ si $g(x) > 0$;

Siendo $g(x)$ todos los valores posibles de y

$f\,(y/x) = \frac{f(x,y)}{g(x)}$ y a la inversa $f\,(x/y) = \frac{f(x,y)}{h(y)}$ con $g(x), h(y) > 0$

Siendo $h(y)$ todos los valores posibles de x

Independencia Estadística

Al igual que con la teoría de conjuntos:

$$P\left({}^{A}\!/_{B}\right) = P(A)$$

$$P\left({}^{B}\!/_{A}\right) = P(B)$$

Se dice que los eventos A y B son independientes.

Como consecuencia:

$$P\left({}^{A}\!/_{B}\right) = \frac{P(A \cap B)}{P(B)} --\!\!\rightarrow P(A \cap B) = P(A).P(B)$$

$$P\left({}^{B}\!/_{A}\right) = \frac{P(A \cap B)}{P(B)} -\!-\!\!\rightarrow P(A \cap B) = P(B).P(A)$$

De la misma forma entonces para la variable aleatoria de distribución conjunta (*x*,*y*):

$$f\left({}^{x}\!/_{y}\right) = g(x)$$

$$f\left({}^{y}\!/_{x}\right) = h(y)$$

Se dice que *x* e *y* son independientes.

Entonces:

$$f\left({}^{x}\!/_{y}\right) = g(x) = \frac{f(x,y)}{h(y)} \;-\!-\!\!\rightarrow f(x,y) = f(x).h(y)$$

$$f\left({}^{y}\!/_{x}\right) = h(y) = \frac{f(x,y)}{g(x)} \;-\!-\!\!\rightarrow f(x,y) = h(y).f(x)$$

Que son las consecuencias de la independencia estadística entre las variables *x*,*y*