- UNIVERSIDAD NACIONAL DEL LITORAL
- FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

•ESTADÍSTICA

- INGENIERÍA EN INFORMÁTICA
- *MG. SUSANA VANLESBERG*

CARACTERÍSTICAS DE V. A.

- El comportamiento de una variable se describe a través de sus funciones de probabilidad. También puede hacerse por considerar un conjunto de medidas (**números**) que resumen aspectos salientes de esas funciones de probabilidad.
- Esos números se denominan

CARACTERÍSTICAS

- ESOS NÚMEROS SE AGRUPAN

SEGÚN LO QUE DESTACAN

- Medidas del **COMPORTAMIENTO CENTRAL**

- Medidas de **VARIABILIDAD**

- Medidas de **FORMA**
  - **ASIMETRÍA - ELEVACIÓN**

MEDIDAS DEL COMPORTAMIENTO CENTRAL

PROMEDIOS

- ESPERANZA MATEMÁTICA
- Media geométrica
- Media armónica

ESPERANZA MATEMÁTICA O VALOR MEDIO

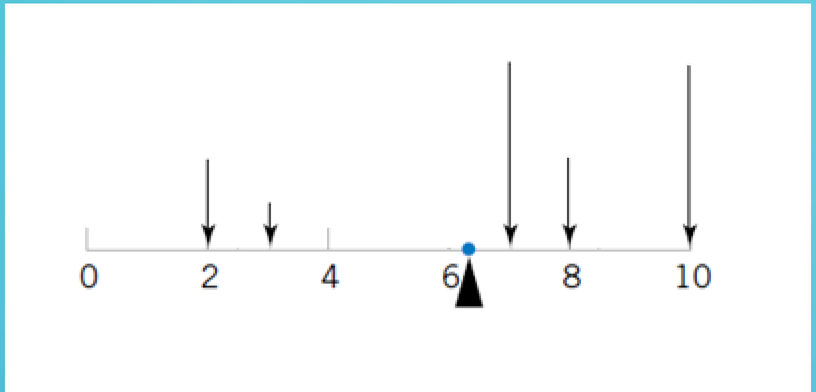
- Es un promedio muy importante. Físicamente ubica el centro de gravedad de la función de probabilidad que describe a la variable aleatoria. Representa a todos los valores, es un punto de equilibrio.
- Se obtiene de acuerdo al tipo de variable que se analice:

$$E(X)=\sum_{i=1}^{n} x_i\, f(x_i)$$

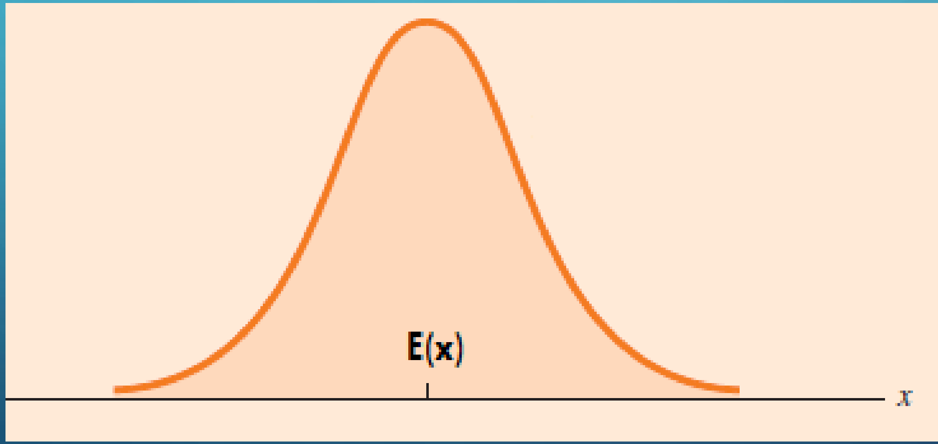
Variable Discreta

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

Variable continua



Variable Discreta



Variable continua

- **Propiedades:**

$$E(C) = C$$

$$E(C) = \int_{-\infty}^{\infty} C \,.\, f(x)dx = C. \int_{-\infty}^{\infty} f(x)dx = C$$

$$E(C\ X) = C\ E(X)$$

$$E(CX) = \int_{-\infty}^{\infty} C\ X\ \ f(X)dX = C \int_{-\infty}^{\infty} X\ \ f(X)dX = C\ E(X)$$

- Luego:

$$E(a+bX)=a+b\,E(X)$$

MEDIDAS DE UBICACIÓN

- Mediana
- Modo
- Cuantiles

- **La Mediana no tiene una fórmula, pero sí cumple la condición de ser el valor de la variable aleatoria que divide en dos partes iguales a la distribución de probabilidad:**

$$\int_{-\infty}^{Mediana} f(x)dx = 0.5$$

$$P(X \leq Mediana) = 0.5$$

$$si \quad X \ es \ variable \quad continua$$

$$o \ bien \quad \int_{Mediana}^{\infty} f(x)dx = 0.5$$

$$P(X > Mediana) = 0.5$$

$$Prob(X \leq Mediana) = \sum_{i=1}^{Mediana} f(x_i) = 0.5$$

$$si\ X\ es\ variable\ discreta$$

$$o\ bien\ Prob(X > Mediana) = \sum_{Mediana}^{\infty} f(x_i) = 0.5$$

MODO

- Es el valor de la variable aleatoria que se corresponde con el máximo de la función de probabilidad:

$$\frac{df(x)}{dx} = 0 \quad y \quad \frac{d^2 f(x)}{dx} < 0 \qquad Máx\ f(x_i)$$

CUANTILES

- Corresponden a valores de la variable aleatoria que dividen a la distribución de probabilidad en distintas partes:
- Si son cuatros son **CUARTILES**
- Si son diez **DECILES**
- Si son cien **PORCENTILES**

- Se obtienen de forma similar a la Mediana:

$$P(X \leq x_p) = p \quad o\, bien \quad P(X > x_p) = 1 - p$$

$$P(X \leq x_{p_i}) = \frac{i}{4}, para\, i = 1,2,3$$

$$P(X \leq x_{p_i}) = \frac{i}{10}, para\, i = 1,2,\ldots,9$$

$$P(X \leq x_{p_i}) = \frac{i}{100}, para\, i = 1,2,\ldots,99$$

MEDIDAS DE VARIABILIDAD

VARIANZA

- Es la medida de dispersión más usada.
- Obtiene el valor promedio de las distancias o apartamientos de los valores de la variable aleatoria al centro de gravedad, sin tener en cuenta el signo es decir la ubicación de cada valor.

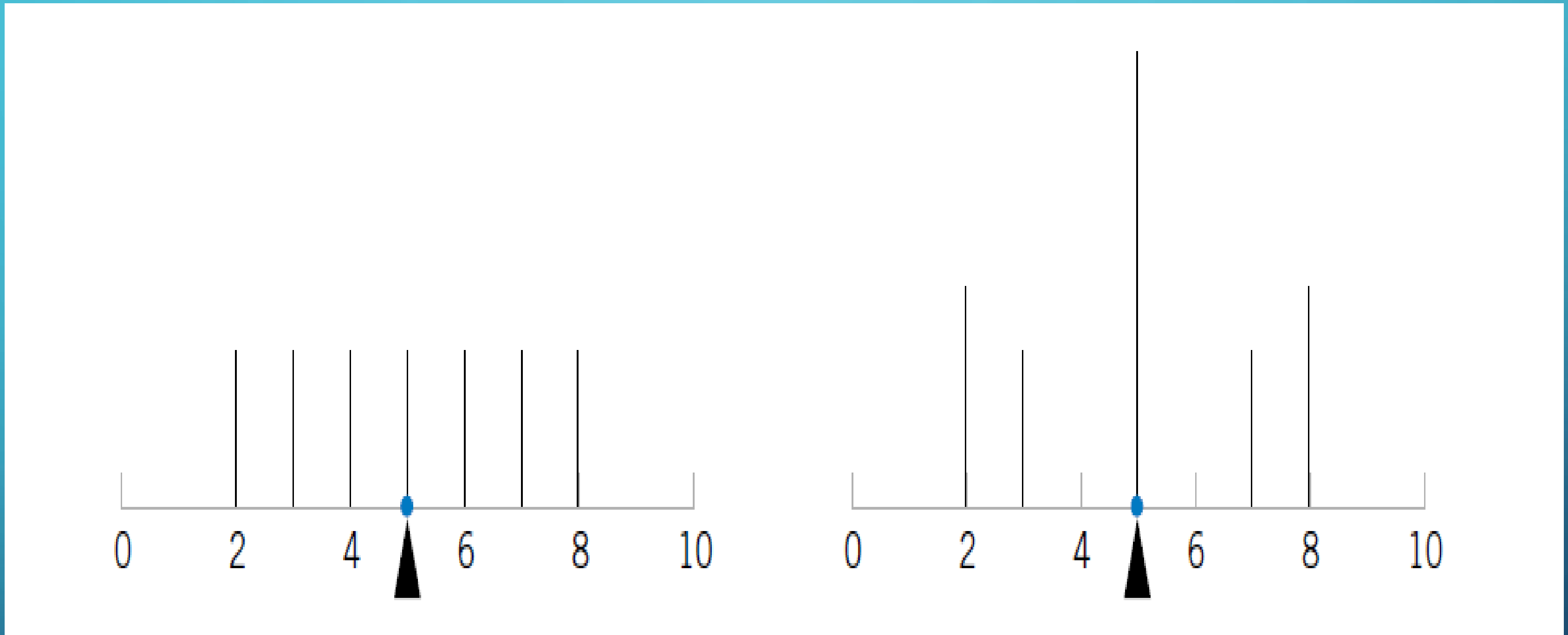
$$Var[X] = \sigma_x^2 = E[X - E[X]]^2$$

$$Var[X] = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x)dx, \text{ si } X \text{ es continua}$$

$$Var[X] = \sum_{\forall i} (x_i - E[X])^2 f(x_i), \text{ si } X \text{ es discreta}$$

- La raíz cuadrada positiva de la varianza es el **DESVÍO** y la ventaja frente a la **VARIANZA** es que tiene las mismas unidades de la variable que se considera.

$$\sigma_x = +\sqrt{Var(X)}$$


0
2
4
6
8
10
0
2
4
6
8
10

PROPIEDADES

$$Var(C) = 0$$

$$E[C - E(C)]^2 = E[C - C]^2 = E(0) = 0$$

$$Var(CX) = C^2 Var(X)$$

$$E[CX - E(CX)]^2 = E[CX - CE(X)]^2 =$$

$$= E\{[C.(X - E(X))]^2\} = E\{C^2.[X - E(X)]^2\} =$$

$$= C^2 E[X - E(X)]^2 = C^2 Var(X)$$

MOMENTOS

- Son promedios de distintas potencias de la variable.
- **Respecto al origen:**

$$\alpha_k = E[X^k] = \sum_{\forall i} x_i^k . f(x_i) \quad ,para\ variable\ discreta$$

$$\alpha_k = E[x^k] = \int_{-\infty}^{+\infty} x^k f(x)dx \quad ,para\ variable\ continua$$

- Momentos respecto al valor medio E(x)=μ:

$$\mu_k = E[(X - E[X])^k] = \int_{-\infty}^{+\infty} (x - E[X])^k f(x)dx, si\ X\ es\ continua$$

$$\mu_k = E[(X - E[X])^k] = \sum_{\forall i} (x_i - E[X])^k . f(x_i), si\ X\ es\ discreta$$

PROPIEDADES

$$\mu_1 = E[x - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$$

- La varianza es el momento centrado de orden 2:

$$Var[X] = \sigma_x^2 = E[X - E[X]]^2$$

- Una medida muy adecuada, ya que no posee unidades es el **COEFICIENTE DE VARIABILIDAD:**

$$Cv = \frac{\sigma}{\mu} \quad (\%)$$

MEDIDAS DE FORMA

MEDIDAS DE ASIMETRÍA

$$As = \frac{E(X) - Modo}{\sigma(X)}$$

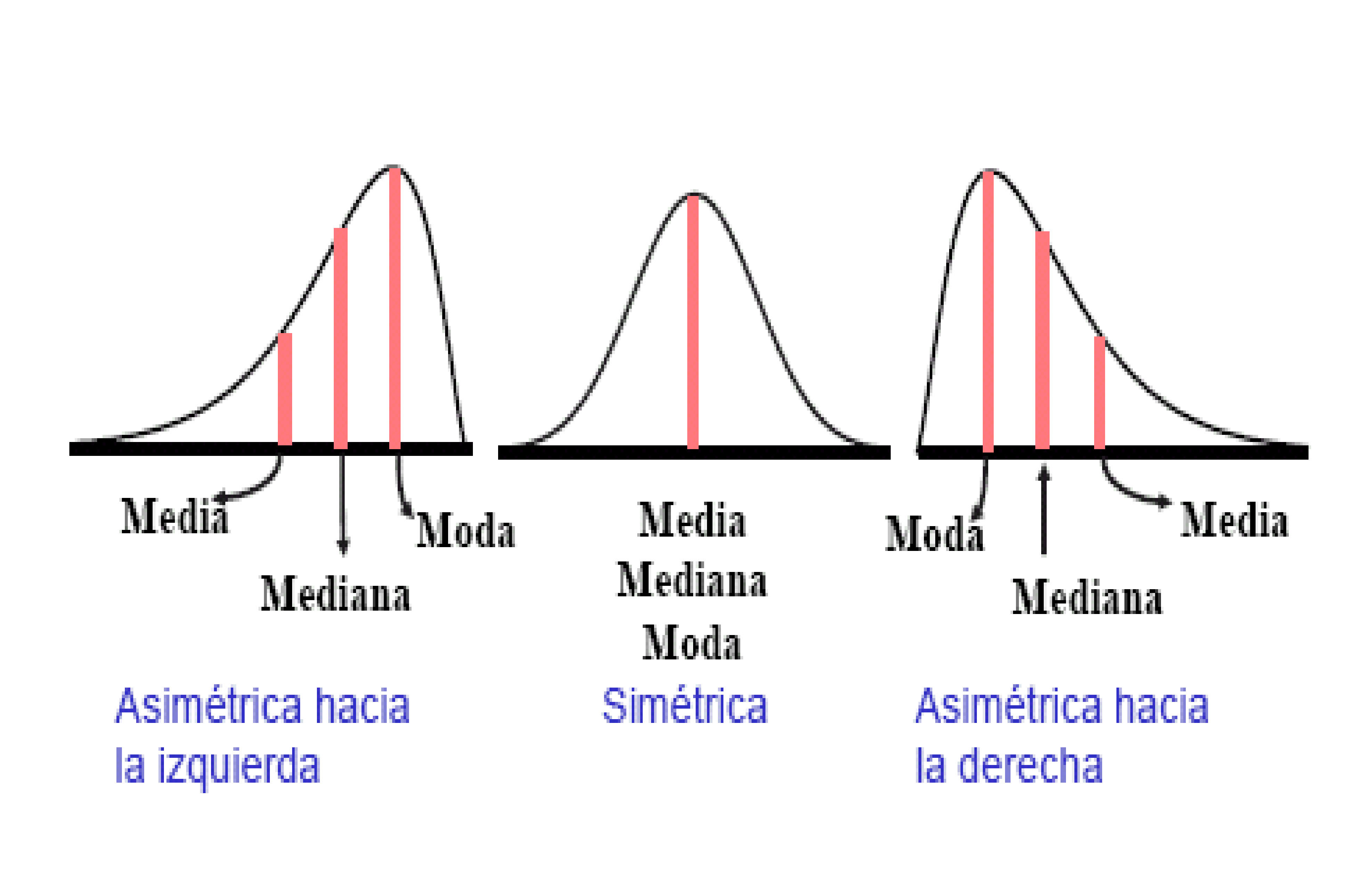
$$o \ \ bien \qquad As = \frac{3\,[E(X) - Mediana]}{\sigma(X)}$$

- LAS ANTERIORES SON MEDIDAS APROXIMADAS, PERO TODAS COINCIDEN EN EL SIGNO.

COEFICIENTE DE ASIMETRÍA

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

- Es un coeficiente que al considerar el momento centrado de orden tres es una medida más exacta.


Media
Mediana
Moda
Media
Mediana
Moda
Moda
Mediana
Media
Asimétrica hacia
la izquierda
Simétrica
Asimétrica hacia
la derecha

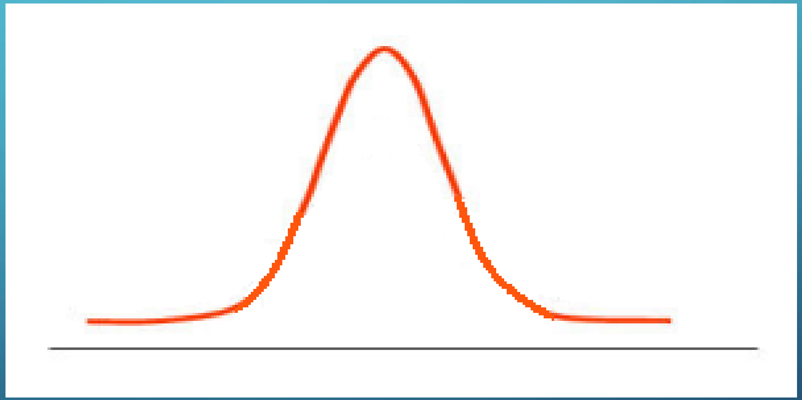
MEDIDAS DE ELEVACIÓN O CURTOSIS

- Esta medida se define en relación a una distribución que se toma como referencia y que tiene un valor constante igual a 3.

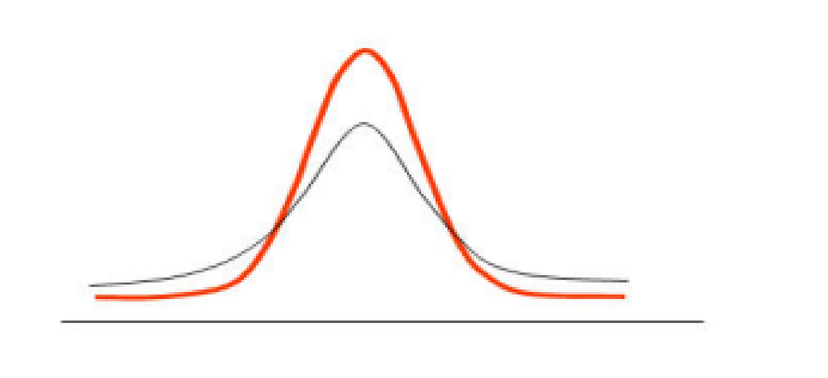
$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$$

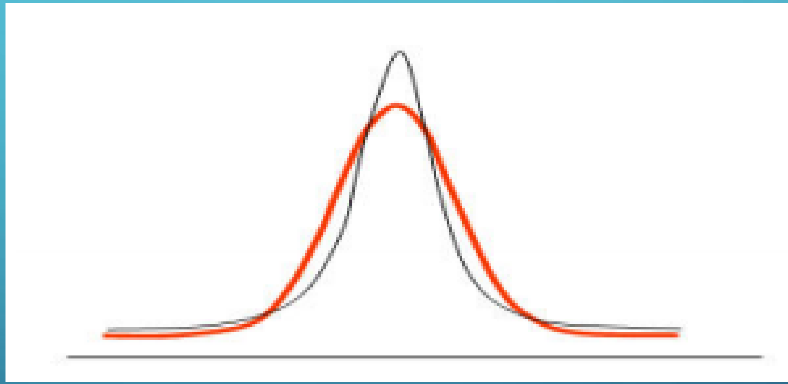
- Si el coeficiente es igual a 3 la curva coincide con la distribución tomada como patrón:



- Si el coeficiente es menor a 3 la curva es más baja que la distribución tomada como referencia:



- Si el coeficiente es mayor a 3 la distribución es más elevada que la distribución patrón:



CARACTERÍSTICAS DE VARIABLES ALEATORIAS DISTRIBUIDAS CONJUNTAMENTE

- De la misma forma que las características de variables unidimensionales permiten resaltar rasgos destacados de la distribución de probabilidad, también para variables bidimensionales es posible hacer lo mismo:

MOMENTOS PARA VARIABLES DISTRIBUIDAS CONJUNTAMENTE

- Si se toman potencias para cada variable se definen los promedios, es decir momentos respecto al origen bidimensionales:

$$E(X^l, Y^n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^l \; y^n \; f(x,y)\,dx\,dy$$

- Es posible obtener las esperanzas marginales a partir de los momentos anteriores, por ejemplo para x y de manera similar para y:

$$E(X^1, Y^0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^1\, y^0\, f(x,y)\,dx\,dy =$$

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\,dy\,dx = \int_{-\infty}^{\infty} x \underbrace{\left(\int_{-\infty}^{\infty} f(x,y)\,dy \right)}_{f(x)} dx =$$

Momentos centrados

- El concepto para una variable se extiende, se definen para diferentes potencias de las variables las distancias respecto a cada esperanza o centro de gravedad:

$$g(X,Y) = [X - E(X)]^{l} . [Y - E(Y)]^{n}$$

- Para la potencia dos de cada variable se obtiene su **varianza marginal**:

$$E\{[X - E(X)]^2 . [Y - E(Y)]^0\}$$

$$\sigma_x^2 = \alpha_{2,0} - \alpha_{1,0}^2$$

Esperanza de una suma algebraica de variables aleatorias

- Aplicando el concepto de esperanza a la suma algebraica de variables:

$$E(X \pm Y) = E(X) \pm E(Y)$$

Covarianza

- Se obtiene de considerar el concepto de momento centrado para ambas variables con potencia uno:

$$Cov(X,Y) = \sigma_{x,y} = E\{\ [X - E(X)]\ [Y - E(Y)]\ \} =$$

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

- más sencillamente:

$$Cov(XY) = E\{\ [X - E(X)].[Y - E(Y)]\ \} =$$

$$= E\{XY - X.E(Y) - E(X).y + E(X).E(Y)\ \} =$$

$$= E(XY) - E(X).E(Y) - \cancel{E(X).E(Y)} + \cancel{E(X).E(Y)} =$$

$$Cov = E(X, Y) - E(X).E(Y)$$

- **Si las variables aleatorias X e Y son independientes la función conjunta se puede expresar como el producto de sus marginales:**

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x - E(X)].[y - E(Y)].\, f\ \ (x).\, f\ \ (y)\, dx\, dy =$$

$$= \int_{-\infty}^{\infty}[x - E(X)].\, f\ \ (x)\, dx \int_{-\infty}^{\infty}[y - E(Y)].\, f\ \ (y)\, dy =$$

$$= E\ [x - E(X)].E\ [y - E(Y)] = \underbrace{\mu_{1,0}}_{=0}.\underbrace{\mu_{0,1}}_{=0}$$

$$luego\ la\ Cov = 0$$

Varianza de una suma algebraica de variables aleatorias

- Se aplica el concepto de varianza a la suma algebraica de las variables.

$$Var(X \pm Y) = E\left\{\left[(X - E(X)) \pm (Y - E(Y))\right]^2\right\}$$

$$Var(X + Y) = Var(X) + Var(Y) + 2.Cov(XY)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2.Cov(XY)$$

COEFICIENTE DE CORRELACIÓN

- Este coeficiente mide el **grado de dependencia lineal de una variable** bidimensional:

$$\rho_{x,y} = \frac{cov_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{\mu_{1,1}}{\sqrt{\mu_{2,0} \cdot \mu_{0,2}}}$$

$$-1 \leq \rho \leq 1$$

- **Si la Cov=0 entonces la dependencia lineal entre las variables no existe. Pero podría existir otro tipo de dependencia.**

- **Si las variables son independientes se deduce que la Cov=0 ya que no habrá ningún tipo de dependencia entre ellas.**

APLICAMOS

Un sistema electrónico tiene dos componentes diferentes en operación conjunta. Las duraciones de cada componente pueden considerarse v.a. con función conjunta dada por:

$$f(x,y) = \frac{1}{8} x\, e^{-(x+y)/2} \qquad x > 0;\, y > 0$$

PUEDE DECIR SI ESTAS VARIABLES SON INDEPENDIENTES?

Una empresa dedicada a la construcción de sistemas, subcontrata a dos empresas especializadas en programación.

Los contratos asignados a estas empresas se pueden considerar variables aleatorias que poseen la distribución conjunta que se presenta.

Interesa saber si estas variables pueden considerarse independientes; utilizar todos los conceptos desarrollados para tal fin.

	y		
x	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0