



Universidad Nacional del Litoral  
Facultad de Ingeniería y Ciencias Hídricas

# ESTADÍSTICA

## Ingeniería Informática

---

### TEORÍA

*Mg.Ing. Susana Vanlesberg*  
Profesor Titular

## UNIDAD 3

### Características de variables aleatorias

## **CARACTERÍSTICAS**

### **VARIABLE ALEATORIA UNIDIMENSIONAL**

Como ya se ha expresado, no es posible predecir con total seguridad el valor futuro de una variable aleatoria, pero sí es posible realizar una descripción completa de su comportamiento a través de sus leyes de probabilidad (función de cuantía, función de densidad, función de distribución).

Existen una serie de números que resumen las características dominantes del comportamiento de una variable aleatoria. Estos números son una forma conveniente de cuantificar la ubicación y la forma de una distribución de probabilidades. Se los puede clasificar de la siguiente manera:

#### **a - Medidas de la tendencia central**

#### **b - Medidas de variabilidad**

#### **c - Medidas de asimetría**

#### **d - Medidas de curtosis**

### **3.1 - Medidas de tendencia central**

Se distinguen, dentro de esta clase, las que son *promedios* y las que son *medidas de ubicación*.

#### **Promedios**

a- Esperanza matemática

b- Media geométrica

c- Media armónica

#### **Medidas de ubicación**

a- Mediana

b- Modo

c- Cuantiles

## Promedios

### a- Esperanza Matemática

Suele ser denominada también *valor medio* ó *valor esperado*. Es uno de los conceptos más importantes y es muy utilizado en la teoría de decisiones, en el análisis de sistemas y en muchos otros campos.

Se la define de la siguiente forma:

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i) \quad \text{variable discreta}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{variable continua} \quad (1)$$

En ella se condensa la información que hay en la función de probabilidad en un solo número. Geométricamente define el centro de gravedad de la masa de la distribución de probabilidades de la variable aleatoria.

### Propiedades

1- La esperanza de una constante es igual a la misma constante:

$$E(C) = C$$

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

2- La esperanza de una constante por una función es igual a la constante por la esperanza de la función:

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

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

3- Esperanza de una función de la variable X:

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

4- Esperanza de una suma de funciones de X:

$$E(g_1(x) \pm g_2(x)) = E(g_1(x)) \pm E(g_2(x))$$

5- Esperanza de un producto de variables aleatorias independientes:

$$E(X.Y) = E(X) \cdot E(Y)$$

## b - Media geométrica

$$M_g = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (2)$$

El logaritmo de la media geométrica es igual al valor esperado de los logaritmos de los  $x_i$ :

$$\log M_g = \sum_{\forall i} f(x_i) \log x_i \quad \text{para variable discreta} \quad (3)$$

$$\log M_g = \int_{-\infty}^{\infty} \log x f(x) dx \quad \text{para variable continua}$$

## c- Media armónica

$$M_H = \frac{1}{\sum_{\forall i} \frac{1}{x_i} f(x_i)} \quad \text{para variable discreta} \quad (4)$$

$$M_H = \frac{1}{\int_{-\infty}^{\infty} \frac{1}{x} f(x) dx} \quad \text{para variable continua}$$

## Medidas de ubicación

### a- Mediana

Al igual que los cuantiles y el modo es una medida de ubicación o posición de la función de probabilidades.

Es el valor de la variable que cumple con la siguiente condición:

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

$$P(X \leq \text{Mediana}) = 0.5$$

si  $X$  es variable continua

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

$$P(X > \text{Mediana}) = 0.5$$

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

si  $X$  es variable discreta

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

(5)

### b – Modo

El modo es el valor de la variable que ocurre más frecuentemente. De esta manera, si la variable aleatoria es continua será el valor de  $X$  que maximice a la función de densidad:

$$\frac{df(x)}{dx} = 0 \quad \text{y} \quad \frac{d^2 f(x)}{dx^2} < 0 \quad (6)$$

y, si la variable es discreta, será el valor de la variable asociado a la máxima probabilidad:

$$M_x = \underset{i=1}{\overset{n}{f(x_i)}} \quad (7)$$

### c - Cuantiles

Se denomina *cuantil de orden  $p$*  (siendo  $p$  un número perteneciente al intervalo  $[0,1]$ ) al valor de la variable  $X_p$  que cumple con la siguiente condición:

$$P(X \leq x_p) = p \text{ o bien } P(X > x_p) = 1 - p \quad (8)$$

**Cuartiles.** Existen tres cuartiles, y dividen a la distribución de probabilidades en cuatro partes; de allí su nombre. Su expresión es la siguiente:

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

El cuartil de orden 2 corresponde a la mediana.

**Deciles.** Dividen la distribución de probabilidades en diez partes, por lo tanto hay 9 deciles:

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

**Porcentiles:** Dividen la distribución de probabilidades en 100 partes; por lo tanto, hay 99 porcentiles. Son de interés cuando se desea analizar detalladamente la distribución de probabilidades:

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

### 3.2 - Medidas de variabilidad

Indican la dispersión o aleatoriedad en el comportamiento de la variable aleatoria.

- a - Rango.
- b - Varianza - Desvío.
- c - Coeficiente de variabilidad.

#### a – Rango

Es simplemente la diferencia entre el mayor y el menor valor de la variable; es una cantidad que no aporta mucha información, ya que solamente se considera un par de números. Suele variar desde  $-\infty$  a  $+\infty$  ó de 0 a  $+\infty$ , o bien entre dos números.

#### b - Varianza

Es la medida de dispersión más usada. Previo a definirla, se hace necesario definir una herramienta muy importante: los *momentos*.

**Momentos:** se los define como los promedios de distintas potencias de la variable aleatoria:

$$\alpha_k = E[X^k] = \sum_{\forall i} x_i^k \cdot f(x_i) \text{ ó}$$

$$\alpha_k = E[X^k] = \int_{-\infty}^{+\infty} x^k f(x) dx$$

(9)

Estas expresiones corresponden a los momentos k-ésimos del área de la función de probabilidades con respecto al origen. Si k=1, se obtiene la esperanza.

Es posible definir también los momentos de áreas con respecto a otro punto que no sea el origen. En particular, momentos con respecto a la media se denominan **momentos centrados**.

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

(10)

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

El momento de orden 1 es igual a 0:

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

El momento centrado de orden 2 define a la **varianza de X**:

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

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

(11)

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

Esta expresión puede simplificarse y ser expresada en función de los momentos con respecto al origen, por desarrollar la potencia del binomio:

$$\begin{aligned}
E[X - E[X]]^2 &= E[X^2 - 2XE[X] + E^2[X]] = \\
&= E[X^2] - 2E[X]E[X] + E[E^2[X]] = \\
&= E[X^2] - 2E^2[X] + E^2[X] = \\
&= E[X^2] - E^2[X] = \alpha_2 - \alpha_1^2
\end{aligned}
\tag{12}$$

### Propiedades

1- La varianza de una constante es igual a cero:

$$Var(C) = 0$$

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

2- La varianza de una constante por X es igual a la constante al cuadrado por la varianza de la variable X:

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

$$\begin{aligned}
E[CX - E(CX)]^2 &= E[CX - CE(X)]^2 = \\
&= E\{[C \cdot (X - E(X))]^2\} = E\{C^2 \cdot [X - E(X)]^2\} = \\
&= C^2 E[X - E(X)]^2 = C^2 Var(X)
\end{aligned}$$

3- Varianza de una función de X:

$$Var(a + bX) = Var(a) + Var(bX) = 0 + b^2 Var(X)$$

### Desvío Estándar

Para algunas aplicaciones suele ser más conveniente utilizar una medida de variabilidad en las mismas unidades de la variable; para eso se define el *desvío estándar* como la raíz cuadrada positiva de la varianza:

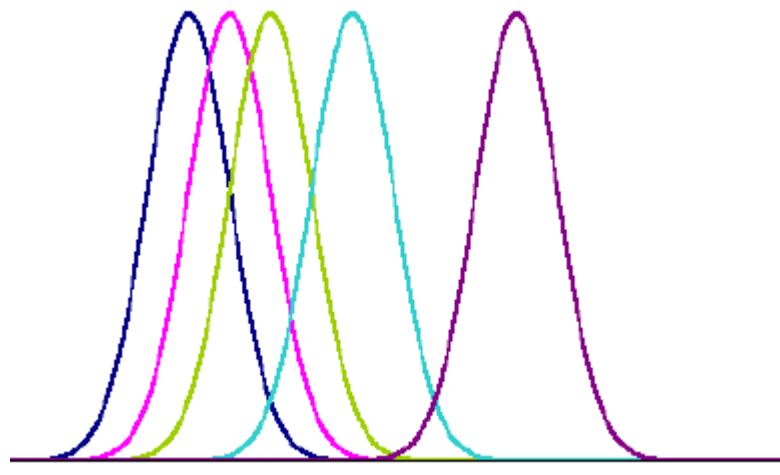
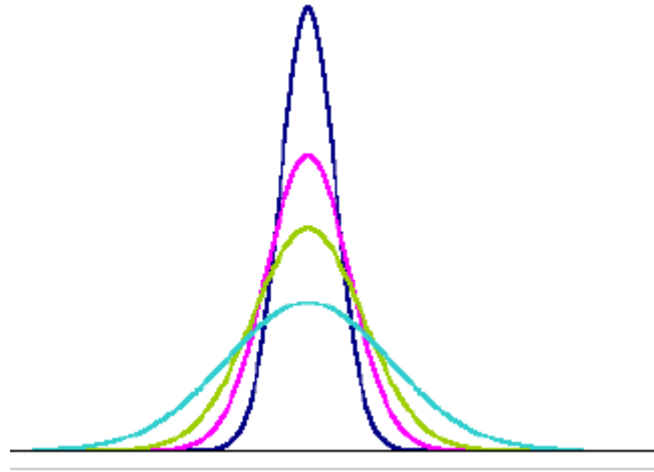


$$\sigma_x = +\sqrt{\text{Var}(X)}$$

(13)

$$\sigma_x = +\sqrt{\alpha_2 - \alpha_1^2}$$

En las siguientes figuras se observa el efecto de variar  $\mu$  y  $\sigma$ .



### **Coeficiente de Variabilidad**

Es un coeficiente adimensional que se obtiene de dividir el desvío por el valor esperado. Se lo utiliza para comparar las dispersiones de poblaciones que correspondan a variables con diferentes unidades de medida o bien en aquellos casos en que si bien las unidades son iguales los valores medios sean muy diferentes:

$$C_v = \frac{\sigma(X)}{E(X)} \quad (14)$$

### 3.3 - Medidas de asimetría

De igual forma que la media y la varianza miden la ubicación y dispersión de una distribución, los momentos más altos miden otras propiedades de la misma.

El tercer momento respecto a la media es usado para determinar si una distribución es simétrica o asimétrica.

Si la *distribución es simétrica*, como las desviaciones están elevadas al cubo, las positivas y negativas tienden a anularse y, por lo tanto,  $\mu_3 = 0$ .

Si la *distribución es asimétrica a la derecha*,  $\mu_3 < 0$ .

Si la *distribución es asimétrica a la izquierda*,  $\mu_3 > 0$ .

$\mu_3$ , tomado como valor aislado, no es una buena medida de la asimetría, ya que tiene las mismas unidades que la variable; por eso es que se define una medida relativa denominada *coeficiente de asimetría*:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (15)$$

Esto puede expresarse en función de los momentos respecto al origen, desarrollando  $\mu_3$ :

$$\begin{aligned} \mu_3 &= E(X - E[X])^3 = E[X^3 - 3X^2 E[X] + 3X E^2[X] - E^3[X]] = \\ &= E[X^3] - 3E[X^2]E[X] + 3E^3[X] - E^3[X] = \\ &= \alpha_3 - 3\alpha_2\alpha_1 + 3\alpha_1^3 - \alpha_1^3 \end{aligned}$$

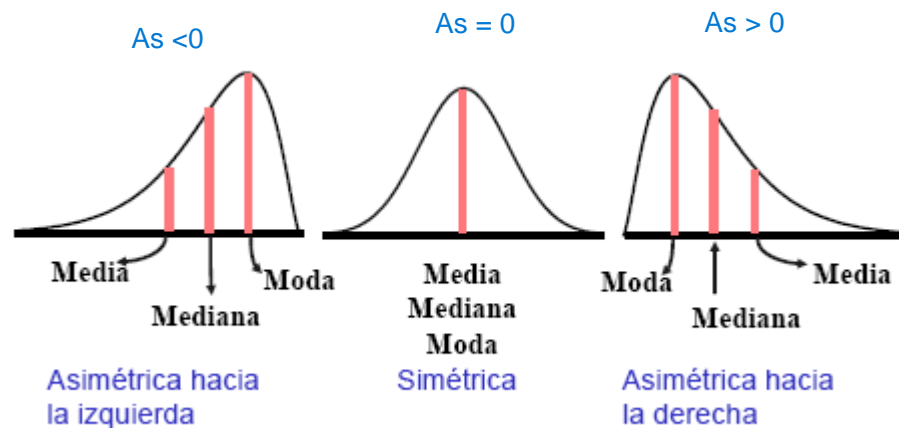
$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3$$

$$\gamma_1 = \frac{\alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3}{(\alpha_2 - \alpha_1^2)^{3/2}}$$

Se cumplen también las siguientes relaciones aproximadas:

$$As = \frac{E(X) - \text{Modo}}{\sigma(X)} \quad As = \frac{3 [E(X) - \text{Mediana}]}{\sigma(X)}$$

La distribución será simétrica si  $As=0$ ; asimétrica a la derecha si  $As>0$ ; y asimétrica a la izquierda si  $As<0$ .



### 3.4 - Medidas de curtosis

Una cuarta propiedad de las variables aleatorias se basa en el momento centrado de cuarto orden, que permite evaluar el empinamiento o aplastamiento de la distribución de probabilidades comparada con una curva tomada como modelo (la curva normal).

Puede introducirse una medida relativa para independizarse de las unidades:

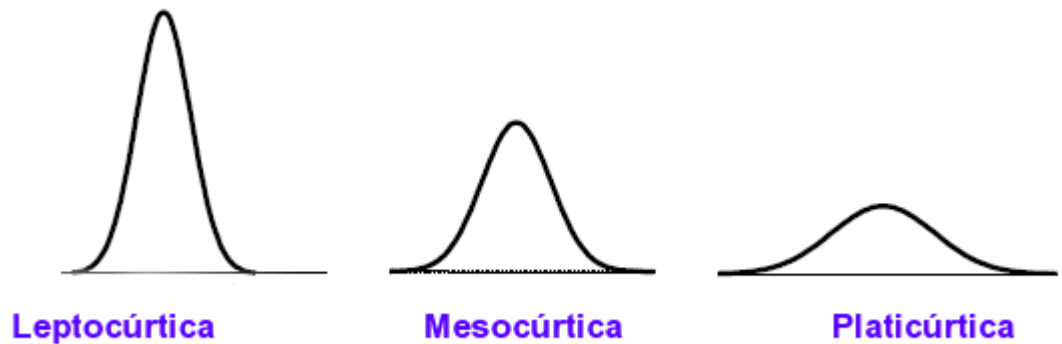
$$\gamma_2 = \frac{\mu_4}{\sigma^4} \quad (16)$$

Desarrollando  $\mu_4$  en función de los momentos respecto al origen:

$$\begin{aligned} \mu_4 &= E[X - E[X]]^4 = E[X^4 - 4X^3 E[X] + 6X^2 E^2[X] - 4X E^3[X] + E^4[X]] = \\ &= E[X^4] - 4E[X^3]E[X] + 6E[X^2]E^2[X] - 4E[X]E^3[X] + E^4[X] \end{aligned}$$

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

La Curtosis de la curva Normal es igual a 3, y se la denomina distribución *mesocúrtica*. Para distribuciones que presenten mayor concentración de probabilidad cerca de la media, mayor que en la Normal, la curtosis será mayor que 3 y se denominará *leptocúrtica*. En caso que la concentración alrededor de la media sea menor que en la Normal la curtosis será menor que 3 y la distribución se dice *platicúrtica*.



### 3.5 - Momentos de variables aleatorias distribuidas conjuntamente

Se parte de considerar una función de las variables X e Y de la siguiente forma:

$$g(X,Y) = X^l Y^n$$

Para obtener el momento conjunto de orden l,n, se obtiene la esperanza de la expresión anterior, desarrollada para el caso de una variable bidimensional continua:

$$E(X^l, Y^n) = \sum_{\forall x_i} \sum_{\forall y_j} x_i^l y_j^n f(x, y) \quad (17)$$

Los momentos más importantes son los de orden (1,0), (0,1), (2,0) y (0,2).

En el caso que  $g(X,Y) = X$ , se obtiene la esperanza de X:

$$E[g(X, Y)] = E(X) = \sum_{\forall x_i} \sum_{\forall y_j} x_i f(x_i, y_j)$$

$$E(X) = \alpha_{1,0} = \sum_{\forall x_i} x_i \sum_{\forall y_j} f(x_i, y_j)$$

$$E(X) = \alpha_{1,0} = \sum_{\forall x_i} x_i f(x_i) \quad (18)$$

Esta expresión es similar a la obtenida para el caso de una variable; de esta manera la esperanza de X es el valor esperado de X sin considerar a Y.

Para obtener la esperanza de Y se procede de forma análoga:

$$E(Y) = \sum_{\forall x_i} \sum_{\forall y_j} y_j f(x_i, y_j) = \sum_{\forall y_j} y_j \sum_{\forall x_i} f(x_i, y_j)$$

$$E(Y) = \alpha_{0,1} = \sum_{\forall y_j} y_j f(y_j) \quad (19)$$

$\alpha_{1,0}$  y  $\alpha_{0,1}$  establecen el centro de masa de la distribución de probabilidades.

### 3.5.1 - Momentos centrados

Como en el caso de una variable, y como en mecánica, el momento centrado más usado es el de segundo orden. Tomando ahora una función de ambas variables igual a:

$$g(X, Y) = [X - E(X)]^l \cdot [Y - E(Y)]^n$$

y hallando la esperanza de esta función, se obtiene :

$$E[g(x, y)] = E\{[x - E(X)]^l \cdot [y - E(Y)]^n\} \quad (20)$$

De la misma manera que para el caso unidimensional los momentos centrados de orden 1,0 o 0,1 son iguales a cero.

Los momentos más usados son los de orden 2,0, 0,2 y 1,1. Los momentos de orden 2,0 y 0,2 llevan a obtener las varianzas marginales.

Expresadas en función de los momentos respecto al origen,

$$\sigma^2(X) = \alpha_{2,0} - \alpha_{1,0}^2 \quad (21)$$

Un resultado similar se obtiene para la varianza de Y:

$$\sigma^2(Y) = \alpha_{0,2} - \alpha_{0,1}^2 \quad (22)$$

### Covarianza

Un momento muy importante se obtiene de considerar  $l=1$  y  $n=1$ . Este momento se denomina *covarianza*.

$$Cov(X,Y) = \sigma_{x,y} = E \{ [X - E(X)] [y - E(Y)] \} = \quad (23)$$

$$= \sum_{\forall x_i} \sum_{\forall y_j} [x_i - E(X)] [y_j - E(Y)] f(x_i, y_j)$$

Puede ser expresada en función de los momentos respecto al origen por desarrollar la ecuación anterior:

$$\sigma_{x,y} = \alpha_{1,1} - \alpha_{1,0} \cdot \alpha_{0,1}$$

### Coefficiente de correlación

Una versión normalizada de la covarianza es el denominado *coeficiente de correlación*  $\rho(x,y)$ , y se obtiene dividiendo la covarianza por el producto de sus desviaciones:

$$\rho_{x,y} = \frac{cov_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{\mu_{1,1}}{\sqrt{\mu_{2,0} \cdot \mu_{0,2}}} \quad (24)$$

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

Este coeficiente merece algunos comentarios:

-de la misma manera que la media y la varianza, el coeficiente de correlación es útil cuando la ley de probabilidades no se conoce en forma completa; en este caso pares de observaciones de ambas variables pueden ayudar a estimar el coeficiente de correlación y poder obtener alguna conclusión acerca de su comportamiento a través de su valor.

-de la definición de covarianza se ve que valores positivos de ella resultarán de pares de valores altos de x con valores altos de y o valores pequeños de x y pequeños de y; mientras que valores negativos de la covarianza se darán con la asociación de valores pequeños de una con valores grandes de la otra. En ambos casos es posible decir que existe al menos alguna dependencia estocástica entre ambas.

-si dos variables aleatorias son independientes, su covarianza y por lo tanto su coeficiente de correlación serán iguales a cero. La independencia implica que  $f(x,y)$  puede descomponerse en el producto de las marginales  $f_1(x) \cdot f_2(y)$  por lo tanto la covarianza se puede obtener como:

$$\begin{aligned} & \sum_{\forall x_i} \sum_{\forall y_j} [x_i - E(X)] \cdot [y_j - E(Y)] \cdot f(x) \cdot f(y) = \\ & = \sum_{\forall x_i} [x_i - E(X)] \cdot f(x_i) \sum_{\forall y_j} [y_j - E(Y)] \cdot f(y_j) = \\ & = E[x_i - E(X)] \cdot E[y_j - E(Y)] = \mu_{1,0} \cdot \mu_{0,1} \end{aligned}$$

$$\text{siendo } \mu_{1,0} = 0 \text{ y } \mu_{0,1} = 0$$

-lo inverso no es cierto, o sea si el coeficiente de correlación o la covarianza son cero esto no implica que las variables sean independientes, puede haber alta dependencia estocástica y ser  $\rho=0$ . **Más específicamente, el coeficiente de correlación es una medida de dependencia lineal entre dos variables aleatorias.**