ESQUEMA

ACERCA DE ESTE MÓDULO

- **Los Objetivos que nos proponemos**
- **Los Núcleos Temáticos que abordaremos?**
- **La Bibliografía sugerida**

UNIDAD V : SUMA DE VARIABLES ALEATORIAS

- **Introducción**
- **Variables aleatorias bidimensionales**
- **Función de probabilidad conjunta**
- **Función densidad conjunta**
- **Función de distribución acumulada**
- **Distribuciones marginales**
- **Distribuciones condicionales**
- **Variables aleatorias independientes**
- **Funciones de dos o más variables aleatorias**
- **Esperanza matemática y variancia para la suma de variables aleatorias**
- **Algunos resultados importantes**
  - **Propiedad reproductiva de la distribución de Poisson**
  - **Propiedad reproductiva de la distribución normal**
  - **La distribución chi - cuadrado**
  - **Teorema del límite central**
- **Covariancia y coeficiente de correlación**
- **Actividades y ejercicios de aplicación**

UNIDAD V

SUMA DE VARIABLES ALEATORIAS

ACERCA DE ESTA UNIDAD

En el estudio de las variables aleatorias hemos considerado hasta ahora, sólo el caso unidimensional. En estas condiciones, el resultado de un experimento se registra mediante un único número x.

En muchas situaciones interesa observar dos o más variables aleatorias simultáneamente. Por ejemplo, podría seleccionarse de un proceso de producción, una lámina de acero y observarse dos características: la resistencia al corte y los diámetros de los puntos de soldadura. Entonces, tanto la resistencia al corte como los diámetros de las soldaduras son las ***variables aleatorias*** de interés. En este caso el resultado de cada experiencia se registra a través de un ***par ordenado de números reales (x,y)***.

A lo largo de esta UNIDAD nos proponemos:

- estudiar la ***distribución conjunta de dos o más variables aleatorias*** ;

y a partir de estas distribuciones conjuntas,

- obtener las distribuciones ***marginales*** de cada variable, como así también las distribuciones ***condicionales***,

- proporcionar algunos resultados que permitan trabajar con problemas de la ingeniería, que involucran situaciones en las que una variable aleatoria es función de dos o más variables aleatorias.

A los fines de favorecer el logro de estos objetivos usted encontrará distintos tipos de ayuda gráfica. En particular, para un reconocimiento rápido de las diferentes propuestas de esta UNIDAD, hemos adoptado las siguientes **SEÑALES**:

Información:

Introduce los comentarios y los contenidos que el profesor a cargo del curso ha elaborado especialmente.

Lectura de bibliografía:

Introduce la lectura de los libros para la consulta de los contenidos de la asignatura

Actividades:

Presenta las preguntas, ejercitaciones o actividades para realizar.

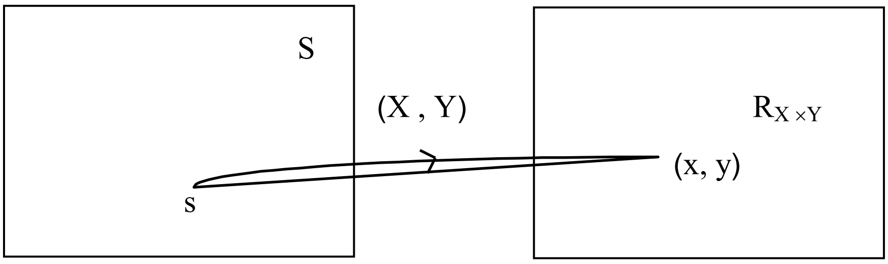
Recomendación:

Es una parada de alerta. Introduce alguna recomendación, observación o sugerencia para el aprendizaje.

VARIABLES ALEATORIAS BIDIMENSIONALES

En lo que sigue introducimos las **DEFINICIONES** y **CONCEPTOS** para describir el ***comportamiento simultáneo de dos o más variables aleatorias***.

⇒ Sea S un ***espacio muestral asociado a una experiencia aleatoria***; X e Y dos variables aleatorias definidas sobre S. Al par (X, Y) lo llamaremos ***variable aleatoria*** o ***vector aleatorio bidimensional***.



Siendo:

$x = X(s) \in R_X$ el ***rango o recorrido*** de X

$y = Y(s) \in R_Y$ el ***rango o recorrido*** de Y

Entonces:

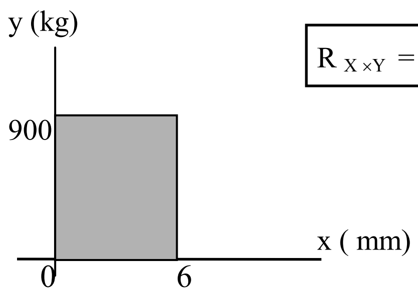
$R_{X \times Y} = \{(x, y) : x \in R_X \, , \, y \in R_Y\}$ es el ***rango o recorrido del vector (X ,Y)***

Veamos algunos EJEMPLOS

1. Se observa el número de repuestos que se demandan en un día en cierta ferretería y el número de repuestos en existencia en dicho día. Si X representa la variable aleatoria *demanda diaria* e Y la variable aleatoria *número de repuestos en existencia*, y se conoce que $R_X = \{0, 1, 2, 3\}$ y $R_Y = \{1, 2, 3\}$, entonces el rango o recorrido del vector es:

$$\boxed{R_{X\times Y} = \{(0, 1), (0, 2), (0, 3), (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}}$$

2. Consideremos nuevamente el caso en que se observa el *diámetro de una soldadura* y la *resistencia al corte de una lámina de acero*.
   Si X representa el diámetro expresado en mm e Y la resistencia expresada en Kg, y se conoce que: $0 \le X \le 6$ mm, mientras que $0 \le Y \le 900$ Kg, entonces el ***rango o recorrido*** para el vector (X, Y) es:



$$\boxed{R_{X\times Y} = \{(x, y) : 0 \le x \le 6 \text{ mm},\ 0 \le y \le 900 \text{ Kg}\}}$$

El análisis de estos ejemplos nos permite identificar:

$\Rightarrow$ En relación al **Ejemplo 1** diremos que el vector (X, Y) es un vector aleatorio bidimensional ***discreto***. En este caso el vector asume solamente un número finito de pares ordenados. (Podría tratarse también de un vector que asuma un número infinito numerable de pares ordenados).

$\Rightarrow$ En relación al **Ejemplo 2** diremos que (X, Y) es un vector aleatorio bidimensional ***continuo*** . El vector aleatorio (X, Y) asume en este caso, cualquier par ordenado de valores (x, y) de un rectángulo. (Podría ser otra región del plano)

FUNCIÓN DE PROBABILIDAD CONJUNTA

Para describir la distribución de probabilidad conjunta de (X, Y) procedemos de un modo análogo al caso unidimensional.

DEFINIMOS:

Sea (X, Y) un ***vector aleatorio bidimensional discreto*** y $R_{X\times Y}$ su ***rango o recorrido***. Entonces, la función

$$p : R_{X\times Y} \longrightarrow \mathbb{R}$$
$$(x, y) \longrightarrow p(x, y)$$

se denomina ***función de probabilidad conjunta del vector $(X,\ Y)$ cuando satisface las siguientes condiciones***:

- $p\,(x, y) \ \geq\ 0$
- $\displaystyle\sum_{(x,\,y)\,\in\,R_{X\times Y}} p\,(x, y) = 1$
- Si $(x, y) \in A \subset R_{X\times Y}$, entonces: $\displaystyle\sum_{(x,\,y)\in A} p\,(x, y) = P\,\big((X, Y)\in A\big)$

Ésta se lee: ***probabilidad de que el vector asuma valores en A***

En relación con esto, consideramos conveniente **OBSERVAR** que:

- $p\,(x, y) = P\,(\,X = x\ ,\ Y = y)$

Asimismo:

- $\{(x,\ y,\ p\ (x,\ y))\ \text{con}\ (x,\ y)\ \in\ R_{X\times Y}\}$ se denomina ***distribución de probabilidad conjunta del vector (X,Y)***

Un poco más....

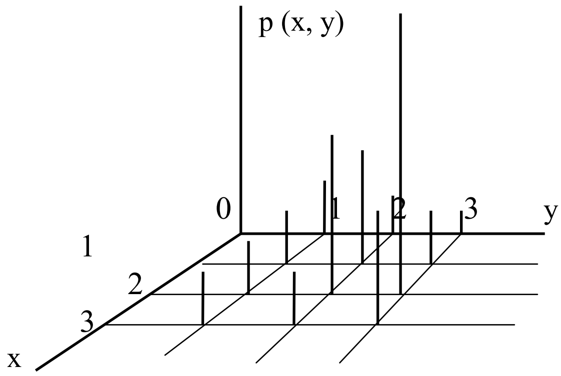
La distribución de probabilidad conjunta de un vector aleatorio bidimensional discreto (X, Y), suele presentarse mediante una ***tabla***.

Si retomamos el ejemplo ya visto, la siguiente tabla corresponde a la **distribución de probabilidad conjunta del vector (X,Y)**, donde X es la variable aleatoria *número de repuestos que se demandan en cierta ferretería*, e Y es la variable aleatoria *número de repuestos en existencia por día, en dicha ferretería*.

Y \ X	0	1	2	3
1	0.05	0.05	0.05	0.05
2	0.03	0.10	0.15	0.05
3	0.02	0.05	0.30	0.10

$$p\,(2, 3) = P\,(X = 2\,,\, Y = 3) = 0.30$$

La distribución de probabilidad conjunta del vector (X,Y) se visualiza en la siguiente **GRÁFICA**:



Si consideramos el suceso A: *en un día no se satisface la demanda*, entonces:

$$A = \{(2, 1), (3, 1), (3, 2)\}$$

y $P\,((X, Y) \in A) = p\,(2, 1) + p\,(3, 1) + p\,(3, 2) = 0.05 + 0.05 + 0.05 = 0.1$

En relación con este ejemplo que venimos trabajando le proponemos como **ACTIVIDAD**:

1. Verifique que $\displaystyle\sum_{(x, y) \in R_{X \times Y}} p\,(x, y) = 1$

2. Calcule la probabilidad de que un día se pierda de vender dos repuestos (demanda insatisfecha).

FUNCIÓN DENSIDAD CONJUNTA

Se requiere introducir más **DEFINICIONES** y **CONCEPTOS**.

Sea (X, Y) un vector aleatorio bidimensional y $R_{X\times Y}$ su recorrido.

Diremos que la función

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$(x, y) \longrightarrow f\,(x, y)$$

es la función ***densidad de probabilidad conjunta*** del vector (X, Y) cuando satisface las siguientes condiciones:

- $f\,(x, y) \geq 0$
- $\displaystyle\iint_{R^2} f\,(x, y)\ dx\ dy = 1$
- $\displaystyle P\left((X, Y) \in A \right) = \iint_{A} f\,(x, y)\ dx\ dy$; para cada $A \subset R_{X\times Y}$

Asimismo

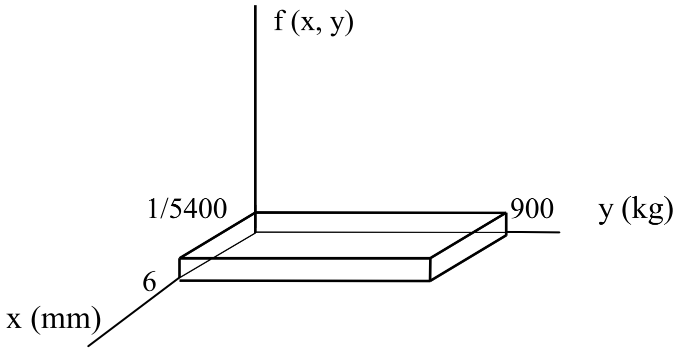
- $\{(x, y, f\,(x, y)\,) \text{ con } (x, y) \in R_{X \times Y}\}$ se denomina ***distribución de probabilidad conjunta del vector (X, Y).***

Veamos un EJEMPLO

Supongamos que la función de densidad de probabilidad conjunta para el vector (X, Y) , - donde X es el *diámetro de la soldadura* e Y *la resistencia al corte* -, viene dada de la forma que se expresa a continuación:

$$f(x, y) = \begin{cases} 1/5400 & \text{si} \quad 0 \le x < 6 \ ; \quad 0 \le y < 900 \\ 0 & \text{caso contrario} \end{cases}$$

La GRÁFICA correspondiente es:



Entonces resulta:

$$P(1 \le X \le 3 \ ; \quad 400 \le Y \le 500) = \int_{1}^{3} \left[\int_{400}^{500} 1/5400 \, dy \right] dx = \frac{200}{5400} \cong 0.037$$

Interpretamos una vez más el *SIGNIFICADO DE UNA PROBABILIDAD*.

El resultado hallado en el ejemplo anterior nos indica que: si observamos el diámetro de una soldadura y la resistencia al corte en cada lámina de acero de una muestra de tamaño n, con n convenientemente grande, alrededor del 3,7 % de las láminas presentan el diámetro entre 1 y 3 mm y la resistencia al corte entre 400 y 500 kg.

FUNCIÓN DE DISTRIBUCIÓN ACUMULADA

Al estudiar variables aleatorias unidimensionales encontramos que la función de distribución acumulada F, juega un papel importante. En el caso bidimensional también podemos definir una función de distribución acumulada como se indica a continuación:

- **Para el caso discreto**:

$$F(u, v) = P\,(X \leq u, Y \leq v) = \sum_{x \leq u} \sum_{y \leq v} p\,(x, y)$$

- **Para el caso continuo:**

$$F(u, v) = P\,(X \leq u, Y \leq v) = \int_{-\infty}^{u} \left[\int_{-\infty}^{v} f\,(x, y)\; dy \right] dx$$

Los conceptos introducidos para un vector aleatorio bidimensional se extienden de manera natural al caso n-dimensional.

DISTRIBUCIONES MARGINALES

Dada $p(x, y) = P(X = x, Y = y)$ la función de probabilidad conjunta del vector (X,Y), se puede calcular con facilidad :

$$p_X(x) = P(X = x) \quad \forall\, x \in R_x \qquad ; \qquad p_Y(y) = P(Y = y) \quad \forall\, y \in R_y$$

En otras palabras, conocida la probabilidad conjunta del vector (X,Y) puede determinarse la función de probabilidad de X e Y respectivamente.

Veamos..

Si consideramos de nuevo **la tabla de la distribución conjunta** del vector (X,Y), donde X es *el número de repuestos que se demandan diariamente en una ferretería* e Y es el *número de repuestos en existencia por día* resulta, por ejemplo, que:

- $p_X(0) = P(X = 0) = P(X = 0, Y = 1) + P(X = 0, Y = 2) + P(X = 0, Y = 3) =$
  $= p(0, 1) + p(0, 2) + p(0, 3) = \sum_{y \in R_Y} p(0, y)$

- $p_Y(1) = P(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) + P(X = 2, Y = 1) +$
  $+ P(X = 3, Y = 1) =$
  $= p(0, 1) + p(1, 1) + p(2, 1) + p(3, 1) = \sum_{x \in R_X} p(x, 1)$

En general:

$$p_X(x) = P(X = x) = \sum p(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_{x \in R_X} p(x, y)$$

Las distribuciones $\{(x, p_X(x))$ con $x \in R_X\}$ e $\{(y, p_Y(y))$ con $y \in R_Y\}$, se denominan ***distribuciones marginales*** de X e Y respectivamente.

En forma análoga:

Dada $f(x, y)$, ***la función densidad de probabilidad conjunta del vector (X, Y)***, se puede determinar :

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

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

Las funciones $f_X(x)$ y $f_Y(y)$ se denominan ***densidades marginales*** de X e Y respectivamente. Asimismo, $\{(x, f_X(x))$ con $x \in R_X\}$ e $\{(y, f_Y(y))$ con $y \in R_Y\}$ son las ***distribuciones marginales*** de X e Y respectivamente.

Volvamos a nuestro ejemplo:

Si consideramos la ***función densidad conjunta*** del vector (X, Y) donde X es el *diámetro de la soldadura* e Y la *resistencia al corte*, se obtiene que:

$$f_X(x) = \int_{-\infty}^{\infty} 1/5400 \, dy = \int_{0}^{900} 1/5400 \, dy = 900/5400 = 1/6 \qquad \text{para } 0 \le x < 6$$

Luego:

$$f_X(x) = \begin{cases} 1/6 & 0 \le x < 6 \\ 0 & \text{caso contrario} \end{cases}$$

es la función de densidad de la variable aleatoria continua X o ***función marginal X.***

Con el propósito de afianzar estos conceptos, en el espacio que sigue le proponemos como **ejercicio** que determine $f_Y(y)$.

....

....

A partir de lo visto surge en forma natural la siguiente pregunta:

¿Puede determinarse la función de probabilidad conjunta o la función de densidad conjunta de un vector (X, Y) a partir de las distribuciones marginales?

Antes de dar respuesta a esta pregunta abordaremos brevemente el siguiente tema.

DISTRIBUCIONES CONDICIONALES

En la UNIDAD II , hemos definido la probabilidad condicional de que ocurra un evento, cuando se sabe que ha ocurrido otro evento.

El concepto de la probabilidad condicional para eventos se puede extender al concepto de distribución condicional de una variable aleatoria X, dado que una segunda variable aleatoria Y ha asumido un valor específico, por ejemplo, y.

Supongamos que:

* X e Y son variables aleatorias discretas con distribución de probabilidad conjunta conocida,
* A es el evento $X = x$, en tanto
* B es el evento $Y = y$.

Entonces

$$P(A / B) = \frac{P(A \cap B)}{P(B)},$$ o sea que:

$$P(X = x / Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}$$

El cociente $\frac{p(x, y)}{p_Y(y)}$ se conoce como la ***función de probabilidad condicional*** para X, dado que $Y = y$.

Este cociente es apropiado para evaluar cualquier declaración de probabilidad en relación a X, cuando se ha observado que la variable Y asumió un valor y.

Continuamos con otras DEFINICIONES importantes.

A continuación nos ocupamos tanto de las ***funciones de probabilidad condicional***, - para el caso discreto -, como de las ***funciones de densidad condicional***, - para el caso continuo -.

Sea (X, Y) un vector aleatorio con función de probabilidad conjunta $p(x, y)$.

La función de probabilidad condicional para Y, dado que $X = x$, se define como:

$$p\,(y / x) = \frac{p\,(x, y)}{p_X\,(x)} \quad \text{para} \quad p_X\,(x) > 0$$

En forma análoga:

$$p\,(x / y) = \frac{p\,(x, y)}{p_Y\,(y)} \quad \text{para} \quad p_Y\,(y) > 0$$

Otra DEFINICIÓN...

Sea (X, Y) un vector aleatorio con función de densidad conjunta $f(x, y)$. La función de densidad condicional para Y, dado que $X = x$, se define como:

$$f\,(y / x) = \frac{f\,(x, y)}{f_X\,(x)} \quad , \text{ para} \quad f_X\,(x) > 0$$

En forma análoga:

$$f(x/y) = \frac{f(x, y)}{f_Y(y)} \quad , \text{para} \quad f_Y(y) > 0$$

Veamos algunos EJEMPLOS de aplicación :

Consideramos la ***distribución conjunta*** del vector (X, Y), donde X es la *demanda de un repuesto* e Y es la *existencia del repuesto*, entonces:

$$p(3/2) = P(X = 3 / Y = 2) = \frac{P(X = 3, Y = 2)}{P(Y = 2)} = \frac{p(3, 2)}{p_Y(2)} = \frac{0{,}05}{0{,}33} = 0{,}15$$

Esto significa: ***cuando hay dos repuestos en existencia alrededor del 15% de los días no se satisface la demanda***.

Si $f(x, y)$ es la función densidad conjunta del vector (X, Y) donde X e Y son respectivamente las variables aleatorias *peso* y *altura* de una persona, entonces:

$$\int_{60}^{80} f(x / 1{,}70)\, dx = \int_{60}^{80} \frac{f(x, 1{,}70)}{f_Y(1{,}70)}\, dx$$

es la probabilidad de que el peso de una persona que mide $1{,}70$ m se encuentre entre 60 y 80 kg. $(P(60 \leq X \leq 80 \;/\; Y = 1{,}70))$

VARIABLES ALEATORIAS INDEPENDIENTES

Recordemos de la Unidad II que las condiciones

i) $P(A/B) = P(A)$; ii) $P(B/A) = P(B)$; iii) $P(A \cap B) = P(A)\,P(B)$

son equivalentes y caracterizan la independencia de los sucesos A y B.

En forma análoga puede suceder que la densidad condicional (o la función de probabilidad) para Y sea igual a la densidad marginal (o función de probabilidad) para Y, sin importar el valor dado de X; es decir:

$f(y/x) = f_Y(y)$ en el caso continuo y

$p(y/x) = p_Y(y)$ en el caso discreto.

Esta situación motiva las siguientes **DEFINICIONES *de variables aleatorias independientes*.**

Para las **variables aleatorias discretas** decimos:

1.- Dos **variables aleatorias discretas** X e Y son **independientes** si se verifica cualquiera de las siguientes condiciones equivalentes:

i) $p(x, y) = p_X(x)\,.\,p_Y(y) \quad \forall (x, y)$

ii) $p(x/y) = p_X(x) \quad \forall (x, y)$ con $p_Y(y) > 0$

iii) $p(y/x) = p_Y(y) \quad \forall (x, y)$ con $p_X(x) > 0$

Del mismo modo, para las **continuas decimos:**

2.- Dos **variables aleatorias continuas** X e Y son **independientes** si se verifica cualquiera de las siguientes condiciones equivalentes:

i) $f(x, y) = f_X(x) \, . \, f_Y(y) \qquad \forall (x, y)$

ii) $f(x / y) = f_X(x) \qquad \forall (x, y)$ con $f_Y(y) > 0$

iii) $f(y / x) = f_Y(y) \qquad \forall (x, y)$ con $f_X(x) > 0$

ACTIVIDADES PROPUESTAS

1. En relación a la distribución conjunta del vector (X, Y) donde X es la variable aleatoria *número de repuestos que se demandan diariamente en cierta ferretería* e Y es la variable aleatoria *número de repuestos en existencia*, le pedimos que determine:

   a) las distribuciones marginales de X e Y ;

   ....
   ....
   ....
   ....

   b) la probabilidad de que, habiendo un repuesto en existencia , no se satisfaga la demanda ;

   ....
   ....

....

....

c) si X e Y son o no son variables aleatorias independientes.

....

....

....

2. Sean:

$$g(v) = \begin{cases} e^{-v} & v \geq 0 \\ 0 & \text{caso contrario} \end{cases} \qquad \text{y}$$

$$h(i) = \begin{cases} 3\,e^{-3i} & i \geq 0 \\ 0 & \text{caso contrario} \end{cases}$$

las densidades marginales de las variables aleatorias independientes v (voltaje) e i (corriente), respectivamente, ¿cuál es la función densidad conjunta del vector (v,i)?

....

....

....

....

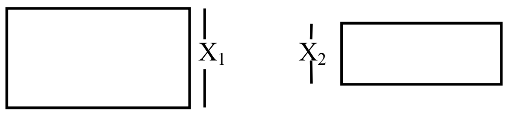
FUNCIONES DE DOS O MÁS VARIABLES ALEATORIAS

Con frecuencia se presentan problemas en los cuales aparece una variable aleatoria en función de dos o más variables aleatorias. Veamos algunos EJEMPLOS:

1. Cuatro resistores están conectados en serie como se muestra en la figura. Si la resistencia del resistor i es una ***variable aleatoria*** que notamos con R_i, entonces la resistencia del ensamble es $R = R_1 + R_2 + R_3 + R_4$

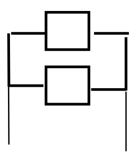
2. Dos partes deben ensamblarse como se muestra en la figura. El espacio libre puede expresarse como $Y = X_1 - X_2$ (combinación lineal de X_1 y X_2 con escalares 1 y -1)

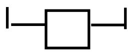
   Un espacio libre negativo significa interferencia.



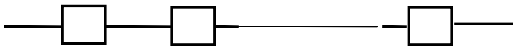
3. Un sistema está formado por n componentes que funcionan independientemente. Si el tiempo transcurrido hasta la falla de la componente i es una variable aleatoria T_i, entonces el tiempo transcurrido hasta la falla del sistema es una variable aleatoria T donde:

   a) $T = \text{máx}\ (T_1, T_2, \ldots T_n)$ si las componentes están conectadas en paralelo





b) $T = \text{mín}\ (T_1,\ T_2,\ .....T_n)$ si las componentes están conectadas en serie



Existen diferentes **métodos** para encontrar ***la distribución de probabilidad de una variable aleatoria, que es función de otras variables aleatorias.***

Nos limitaremos a enunciar algunos resultados útiles que nos permitirán resolver numerosos problemas de la ingeniería y abordar algunos temas de la inferencia estadística, tales como la estimación puntual y por intervalos de confianza de ciertos parámetros.

En lo que sigue obtendremos la **distribución de la variable aleatoria T del ejemplo 3.a.- , bajo el supuesto de que las variables aleatorias T_1, T_2, ...T_n son independientes e idénticamente distribuidas** (es decir, tienen la misma curva de densidad, que notaremos $f_{Ti}\ (t)$).

Si las componentes del sistema provienen de un mismo proceso de producción es razonable suponer que las variables aleatorias T_i están idénticamente distribuidas.

$$F_T(t) = P(T \leq t) = P(T_1 \leq t, T_2 \leq t, \ldots. T_n \leq t) =$$

$$= P(T_1 \leq t)\ P(T_2 \leq t) \ldots..P(T_n \leq t) = \left(\int_0^t f_{Ti}(x)\,dx\right)^n$$

Derivando miembro a miembro se obtiene la función densidad de la variable aleatoria T:

$$f_T(t) = n \left(\int_0^t f_{Ti}(x)\,dx\right)^{n-1} .\ f_{Ti}(t)$$

Para el ejemplo 3.b. conviene plantear:

$$P(T > t) = P(T_1 > t, T_2 > t, \ldots. T_n > t).$$

ESPERANZA MATEMÁTICA Y VARIANCIA PARA LA SUMA DE VARIABLES ALEATORIAS

Una variable aleatoria de interés es la que aparece cuando se considera una **combinación lineal de n variables aleatorias** $X_1, X_2, \ldots\ldots..X_n$, es decir:

$$Z = H(X_1, X_2, \ldots\ldots..X_n) = a_1 X_1 + \ldots\ldots...+ a_n X_n$$

Nos interesa estudiar en primer término la ***esperanza matemática y la variancia*** de $Z = H(X_1, X_2, \ldots\ldots X_n) = a_1 X_1 + \ldots\ldots\ldots + a_n X_n$ bajo diferentes hipótesis.

Comencemos por analizar un **EJEMPLO** simple.

Sean X: nº de piezas producidas por una máquina por la mañana

Y: nº de piezas producidas por la misma máquina por la tarde

La siguiente TABLA corresponde a la **distribución de probabilidad conjunta del vector (X, Y)**

Y \ X	1	2	3
1	0,05	0,10	0,15
2	0,05	0,35	0,30

La variable aleatoria $Z = X + Y$ es el *número total de piezas producidas por la máquina en un día.*

A partir de la distribución conjunta del vector (X, Y) se puede determinar la ***distribución de probabilidad de la variable $Z = X + Y$.***

En efecto, $R_z = \{2, 3, 4, 5\}$ y

$$P(Z = 2) = P(X = 1, Y = 1) = 0{,}05$$

$$P(Z = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0{,}15$$

Del mismo modo usted puede completar las líneas de puntos

$$P(Z = 4) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots = 0{,}50$$

$$P\,(Z = 5) \;=\; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \;= 0{,}30$$

Conocida la distribución de probabilidad de la variable aleatoria Z podemos calcular $E\,(Z)$

$$E\,(Z) = 2\,P\,(Z = 2) + 3\,P\,(Z = 3) + 4\,P\,(Z = 4) + 5\,P\,(Z = 5) = 4.05$$

Como **ACTIVIDAD le** proponemos calcular $E\,(X)$ y $E\,(Y)$ y verificar que:

$$E\,(Z) = E\,(X + Y) = E\,(X) + E\,(Y)$$

Este resultado es un caso particular de la siguiente **PROPIEDAD** (1):

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i\, E\,(X_i)$$

cualesquiera sean las variables aleatorias $X_1, X_2, \dots X_n$ y los escalares $a_1 \dots a_n$

Le proponemos asimismo como **ACTIVIDAD** calcular $V(X)$, $V(Y)$, $V(Z)$ donde $Z = X + Y$ y verificar que $V\,(X+Y) \neq V(X) + V(Y)$

Si recordamos que $V\,(X) = E\left[(X - E\,(X))^2\right]$, y aplicamos la **PROPIEDAD (1),** después de desarrollar el cuadrado obtenemos:

$$V(X) = E(X^2) - (E(X))^2 \qquad (2)$$

La relación (2) nos facilitará el cálculo de la **variancia para la suma de dos variables**.

$$V(X+Y) = E\left[(X+Y)^2\right] - \left(E(X+Y)\right)^2 =$$

$$= E(X^2) + 2\,E(XY) + E(Y^2) - (E(X))^2 - 2\,E(X)\,E(Y) - (E(Y))^2 =$$

$$= E(X^2) - (E(X))^2 + E(Y^2) - E(Y^2) + 2\,(E(XY) - E(X)\,E(Y)) =$$

$$= V(X) + V(Y) + 2\,(E(XY) - E(X)\,E(Y))$$

Si X e Y son variables aleatorias independientes, entonces (se prueba fácilmente) :

$$E(XY) = E(X)\,E(Y)$$

En este caso es: $V(X+Y) = V(X) + V(Y)$

Este resultado es un caso particular de la siguiente **PROPIEDAD:**

Si $X_1, X_2, \ldots\ldots X_n$ son n variables aleatorias independientes y $a_1, a_2, \ldots a_n$, n escalares, entonces:

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2\, V(X_i)$$

Ⓟ Es importante detenerse a **OBSERVAR**:

Si X_1 y X_2 son **variables aleatorias independientes** entonces:

1.- $V(X_1 + X_2) = V(X_1 - X_2) = V(X_1) + V(X_2)$

2.- $\sigma(X_1 + X_2) = \sigma(X_1 - X_2) = \sqrt{V(X_1) + V(X_2)} \neq \sigma(X_1) + \sigma(X_2$

ALGUNOS RESULTADOS IMPORTANTES

1. Propiedad reproductiva de la distribución de Poisson

- Si $X_1, X_2, ...X_n$ son n variables aleatorias independientes y cada X_i tiene una distribución de Poisson con parámetros λ_i, entonces la variable aleatoria :

$$X = \sum_{i=1}^{n} X_i ,$$

también tiene distribución de Poisson de parámetro $\lambda = \sum_{i=1}^{n} \lambda_i$

Veamos un **EJEMPLO:**

Si la *demanda diaria de repuesto en una ferretería* es **una X con distribución de Poisson de parámetro λ** . entonces la *demanda semanal* (6 días) **es una variable aleatoria D** , siendo

$$D = \sum_{i=1}^{6} X_i$$

con **distribución de Poisson, de parámetro 6λ.** (se supone independencia entre las demandas diarias).

Es conveniente detenerse y **OBSERVAR** la siguiente desigualdad:

$$D = \sum_{i=1}^{6} X_i \neq 6 X_i \qquad \text{¿ Por qué es así?}$$

Veamos más PROPIEDADES

2. **Propiedad reproductiva de la distribución normal**

- Si $X_1, X_2, ...X_n$ son $\mathbf{n}$ **variables aleatorias independientes** y cada X_i tiene una **distribución normal** con $E(X_i) = \mu_i$ y $V(X_i) = \sigma_i^2$,

entonces: $Y = c_1 X_1 + c_2 X_2 + ... c_n X_n$

es una **variable aleatoria normal**, con:

$$E(Y) = c_1 \mu_1 + c_2 \mu_2 + ... + c_n \mu_n \quad \text{y}$$

$$V(Y) = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + ... + c_n^2 \sigma_n^2$$

Veamos un **EJEMPLO**:

La envoltura de plástico para un disco magnético está formada por dos hojas. Si **el espesor de una hoja es una variable aleatoria**, siendo:

$$X \sim N(\mu_X = 1{,}5 \text{ mm}, \sigma_x^2 = 0{,}01 \text{ mm}^2) \quad ,$$

entonces **el espesor de las dos hojas es una variable aleatoria** Y, resultando:

$$Y = X_1 + X_2 \sim N(\mu_Y = 3 \text{ mm}, \sigma_Y^2 = 0{,}02 \text{ mm}^2)$$

(X_i simboliza la variable aleatoria espesor de la hoja i). Hemos supuesto además que las variables aleatorias X_1 y X_2 son independientes.

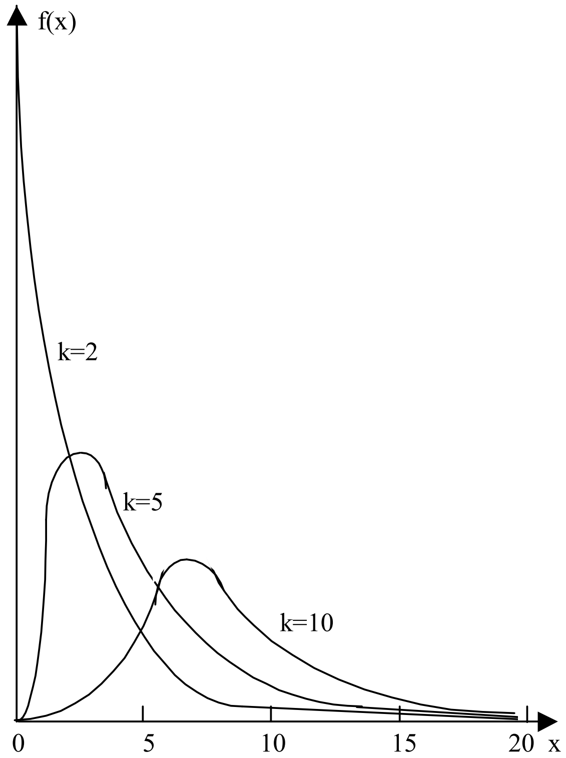
Un poco más....

3. La distribución chi- cuadrado

- Si $X_1, X_2, \ldots\ldots X_n$ son ***n* variables aleatorias independientes donde cada una tiene una distribución** $\mathbf{N(0, 1)}$**,**
  entonces $T = X_1^2 + X_2^2 + \ldots\ldots X_k^2$ tiene ***una distribución que llamaremos chi- cuadrado con k grados de libertad*** y notaremos $T \sim \chi_k^2$.

  Esta distribución se aplicará para estimar la **variancia de una variable aleatoria con distribución normal**.

Observemos la gráfica correspondiente para distintos valores de k



Funciones de densidad de probabilidad de varias distribuciones χ^2

4. Teorema del límite central

Si $X_1, X_2, \ldots\ldots X_n$ son n variables aleatorias independientes cualesquiera, con $E(X_i) = \mu_i$ y $V(X_i) = \sigma^2_i$,

entonces, bajo ciertas condiciones que son válidas en la mayor parte de las aplicaciones, la variable aleatoria $X = X_1 + X_2 + \ldots\ldots + X_n$, con n convenientemente grande, tiene una **distribución aproximadamente normal** con parámetros

$$E(X) = \sum_{i=1}^{n} \mu_i \quad \text{y} \quad V(X) = \sum_{i=1}^{n} \sigma^2_i$$

Es importante realizar las siguientes **OBSERVACIONES:**

- Por lo general se considera que n es convenientemente grande cuando es mayor que 30.

- En muchos problemas, la variable aleatoria que se considera puede ser representada como la suma de n variables aleatorias independientes, y por lo tanto, su distribución puede aproximarse por la distribución normal.

  Por ejemplo: El consumo diario de electricidad en una ciudad es igual a la suma de un número considerable de consumos individuales.

Las condiciones generales citadas anteriormente pueden resumirse como sigue:

- Los términos individuales en la suma contribuyen con una cantidad despreciable a la variación de la suma y es poco probable que cualquier término haga una gran contribución a la suma.

- El Teorema del Límite Central establece que los sumandos no necesitan estar distribuidos normalmente a fin de que la suma sea aproximada a una distribución normal.

Veamos un **PROBLEMA DE APLICACIÓN:**

- Al sumar números, un computador aproxima cada número al entero más próximo. Supongamos que todos los errores de aproximación son independientes y distribuidos uniformemente en (-0,5 ; 0,5). Si se suman 1500 números, ¿cuál es la probabilidad de que la magnitud del error total exceda 15?

Sea X_i :error de aproximación en el i-ésimo número y X: error en la suma de los 1500 números

$$X = \sum_{i=1}^{1500} X_i \quad ; \quad E(X) = 1500.E(X_i) \quad ; \quad V(X) = 1500.V(X_i)$$

donde $E(X_i) = 0$ y $V(X_i) = \dfrac{(0{,}5 - (-0{,}5))^2}{12} = 1/12$

Por el teorema del límite central X tiene distribución aproximadamente normal con $E(X) = 0$ y $V(X) = 1500/12$

Por lo tanto:

$$P(|X| > 15) = 1 - P(|X| \leq 15) =$$

$$= 1 - P\left(\frac{|X - 0|}{\sqrt{1500/12}} \leq \frac{15}{\sqrt{1500/12}}\right) = 1 - P(|Z| \leq 1{,}34) =$$

$$= 1 - (2 \times 0{,}9099 - 1) = 1 - 0{,}8198 = 0{,}1802$$

ACTIVIDADES PROPUESTAS

1.- El espesor de una lámina metálica en (mm) es una variable aleatoria,

$$X \sim N\,(\mu_X = 0.5;\ \sigma_X = 0.05)$$

El espesor de una lámina de papel aislante en (mm) es también una variable aleatoria,

$$Y \sim N\,(\mu_Y = 0.05;\ \sigma_Y = 0.02)$$

Obtenga la distribución del espesor del núcleo de un transformador que consta de 50 capas de láminas metálicas y 49 láminas de papel aislante.

....
....
....
....

2.- La demanda diaria de agua por habitante en una ciudad de 10000 habitantes es una variable aleatoria X con

$$E\,(X) = 0{,}4\ m^3 \quad \text{y} \quad \sigma_X = 0{,}09\ m^3.$$

La disponibilidad de agua para el consumo almacenado diariamente en una represa es una variable aleatoria

$$Y \sim N\,(\mu_Y = 4500\ m^3;\ \sigma_Y = 450\ m^3).$$

Calcule la probabilidad de que un día cualquiera no sea satisfecha la demanda.

....
....
....
....
....

3. Un supermercado tiene dos puertas de ingreso que indicamos con A y B. El número de clientes que entran por la puerta A en un período de 1 minuto es una variable aleatoria con distribución de Poisson de parámetro $\lambda_A = 2{,}5$, y el número de clientes que entran por la puerta B en un período de 1 minuto es una variable aleatoria independiente de la anterior, con distribución de Poisson de parámetro $\lambda_B = 2$.

   Calcule la probabilidad de que en un período de 2 minutos entren por lo menos 5 clientes.

   ....
   ....
   ....
   ....
   ....

COVARIANCIA Y COEFICIENTE DE CORRELACIÓN

Cuando vemos un niño de cinco años y decidimos que está muy alto, lo comparamos con la altura media de otros niños de la misma edad; es decir, relacionamos dos variables, la edad y la altura.

Cuando se tienen dos variables aleatorias X e Y suele ser útil describir cómo varían conjuntamente, o disponer de una medida de la relación que existe entre las variables.

Una **medida de la relación que existe entre dos variables aleatorias** es la **covariancia.**

DEFINIMOS:

> Se llama covariancia entre dos variables aleatorias X e Y, y se nota **cov (X.Y),** al número
>
> $$\mathbf{cov\,(X, Y) = E\,[(X - \mu_X)\,(Y - \mu_Y)]}$$

De la definición surge de inmediato que:

- Si X e Y son variables aleatorias discretas con función de probabilidad conjunta $p\,(x, y)$, entonces:

$$\text{cov}\,(X, Y) = \sum_{(x, y)\, \in\, R_{X \times Y}} (x - \mu_X)\,(y - \mu_Y)\,p\,(x, y) \qquad (1)$$

- Si X e Y son variables aleatorias continuas con función densidad conjunta $f\,(x, y)$, entonces:

$$\text{cov}\,(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)\,(y - \mu_Y)\,f\,(x, y)\,dx\,dy \qquad (2)$$

Ⓟ

Las expresiones (1) y (2) nos permiten una interpretación rápida de la covariancia. Veamos algunas cuestiones importantes:

- Si los valores grandes de X están asociados con valores grandes de Y, con una probabilidad alta, entonces la covariancia es positiva. En cambio si existe una probabilidad alta de que valores grandes de X se encuentren asociados con valores pequeños de Y o viceversa entonces la covariancia será negativa.

- Le proponemos probar que:

$$\operatorname{cov}(X, Y) = E(X . Y) - E(X) . E(Y)$$

- En el caso en que las variables aleatorias son independientes, ya hemos observado que

$$E(X . Y) = E(X) . E(Y) \qquad \text{y por lo tanto} \qquad \operatorname{cov}(X, Y) = 0$$

Sin embargo puede mostrarse mediante un ejemplo que la recíproca no se cumple.

- El inconveniente que presenta la covariancia es que depende de las unidades en que vienen dadas las variables aleatorias. Si en vez de X e Y se toman las variables aleatorias aX , bY, donde a y b son escalares, resulta:

$$\text{cov}(aX, bY) = a.b.\ \text{cov}(X, Y)$$

Para evitar este inconveniente se ha recurrido al siguiente concepto.

Introducimos la siguiente **DEFINICIÓN**

Se llama coeficiente de correlación entre dos variables aleatorias X e Y y se nota $\rho(X, Y)$, al cociente:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

De la definición se deduce en forma inmediata que

$$\rho(X, Y) = E\left[\frac{(X - \mu_X)}{\sigma_X} \frac{(Y - \mu_Y)}{\sigma_Y}\right]$$

es la ***covariancia entre dos variables estandarizadas***, es decir, variables que tienen esperanza matemática igual a cero y variancia igual a uno.

Enunciaremos, sin demostrar, algunas **PROPIEDADES DEL COEFICIENTE DE CORRELACIÓN.**

1. $-1 \leq \rho(X, Y) \leq 1$

2. $\rho(X, Y) = -1 \quad \Leftrightarrow \quad Y = AX + B$ con $A < 0$

3. $\rho(X, Y) = 1 \quad \Leftrightarrow \quad Y = AX + B$ con $A > 0$

- Cuando el coeficiente de correlación toma el valor -1 entonces las observaciones del vector (X, Y) se encuentran (con probabilidad uno) sobre una recta con pendiente negativa; en cambio cuando el coeficiente de correlación toma el valor 1, entonces las observaciones del vector (X, Y) se encuentran (con probabilidad uno) sobre una recta con pendiente positiva.

- El valor $\rho = 0$ indica la ausencia de cualquier relación lineal entre las variables, sin embargo puede existir otro tipo de relación.

- Si existe una relación estrecha entre dos variables, el conocimiento de una de ellas puede ayudar a predecir el comportamiento de la otra; en cambio cuando la relación es débil, tener información sobre una de las variables no ayuda demasiado a extraer conclusiones sobre la otra.

- Si se conoce la función de densidad conjunta $f(x, y)$ de un vector aleatorio continuo (X, Y), entonces, es posible calcular el coeficiente de correlación entre X e Y. ¿Por qué?

- Si sólo se tiene una muestra formada por n pares ordenados (x_i, y_i), con $i = 1,....n$, de observaciones del vector (X, Y), una medida adecuada del grado de asociación lineal entre las variables X e Y viene dada por el estadístico

n

$$r = \frac{1/n \sum_{i=1} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{1/n \sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{1/n \sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

El estadístico r se denomina ***coeficiente de correlación de la muestra.***

- Observemos que el numerador de r es la media aritmética de los productos de las desviaciones correspondientes de x_i e y_i , y sus medias muestrales. Este valor se denomina covariancia muestral entre X e Y, y se simboliza por S_{XY} . En el denominador de r aparece el producto del desvío estándar de los valores de x_i (S_X) con el desvío estándar de los valores y_i (S_Y).

Con la notación introducida, podemos escribir:

$$r = \frac{S_{XY}}{S_X \; S_Y}$$

Veamos las PROPIEDADES de r , las que pueden ser deducidas a partir de su definición:

i) $-1 \leq r \leq 1$
ii) Si $r = 1$ (ó -1), los puntos de la muestra están situados sobre una línea recta de pendiente positiva (negativa)
iii) Si r asume valores próximos a 1 ó -1, habrá una asociación lineal fuerte entre las variables.
iv) Si r es próximo a cero, entonces hay una correlación muy débil entre las variables. (Podrá existir una relación no lineal)

v) r es un número sin dimensión, es decir, no depende de las unidades empleadas para X e Y.

Es claro que **r es una medida muy útil para saber el grado en que la asociación entre las dos variables se aproxima a una relación funcional lineal**.

Por otra parte es fundamental que se comprenda que **r** por sí mismo no puede probar ni desmentir una relación causal entre **X** e **Y** aún cuando $\mathbf{r = \pm 1.}$ La manifestación de una relación causa - efecto es posible sólo a través de la comprensión de la relación natural que existe entre X e Y, y ésta no debe manifestarse sólo por la existencia de una fuerte correlación entre X e Y.

PROPUESTA DE ACTIVIDAD

Los datos que figuran en la siguiente TABLA fueron recopilados con el fin de estudiar la relación entre la vida útil de una herramienta y su velocidad de corte.

vida útil	41	18	22	43	6	20	21	35	35	11	15	13	32
velocidad de corte	86	104	100	85	107	104	106	99	90	110	108	105	91

Le proponemos que:

a) Construya un diagrama de dispersión (visto en la **Unidad I**) representando sobre el eje horizontal la *velocidad de corte*.

b) ¿Cuáles son sus observaciones?

....
....
....

c) Calcule el coeficiente de correlación muestral.

....
....
....
....
....