DISTRIBUCIONES MULTIDIMENSIONALES DE PROBABILIDAD

FUNCIÓN DE DISTRIBUCIÓN CONJUNTA DE UN VECTOR ALEATORIO

* En vectores aleatorios discretos y continuos:

$$F(x,y) = P(X \leq x , Y \leq y) = P\left[(X,Y) \in \left]-\infty ,x\right] \times \left]-\infty ,y\right] \right]$$

$$F(x_1, x_2 ,...,x_n) = P (X_1 \leq x_1, X_2 \leq x_2,..., X_n \leq x_n) =$$

$$= P\left[(X_1 ,X_2 ,...,X_n) \in \left]-\infty , x_1\right] \times \left]-\infty , x_2\right] \times ...\times \left]-\infty , x_n\right]\right]$$

propiedades

1.- $0 \leq F(x,y) \leq 1$

$$0 \leq F(x_1, x_2 ,...,x_n) \leq 1$$

2.- $F(\infty ,\infty)= 1$

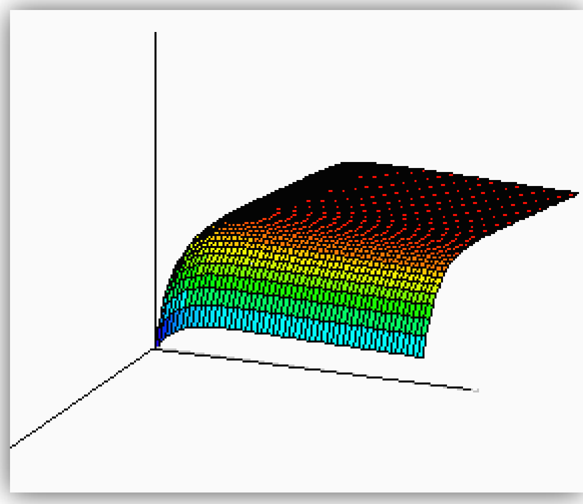
$$F(\infty ,\infty ,...,\infty)=1$$

3.- $F(-\infty ,y)= 0$; $F(x,-\infty)=0$

$F(-\infty, x_2, \ldots, x_n) = 0$;; $F(x_1, -\infty, \ldots, x_n) = 0$;; $F(x_1, x_2, \ldots, -\infty) = 0$

4.- Es siempre **No decreciente** para todas y cada una de las componentes

5.- Es siempre continua por la derecha: gráficamente



FUNCIÓN DE CUANTÍA CONJUNTA de vectores aletorios discretos

$P(x,y) = P(X=x, Y=y)$

$P(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

Propiedades

1.-

$$F(x,y) = \sum_{X \le x} \sum_{Y \le y} P(X,Y)$$

$$F(x_1, x_2, \ldots, x_n) = \sum_{X_1 \le x_1} \sum_{X_2 \le x_2} \cdots \sum_{X_n \le x} P(X_1, X_2, \ldots, X_n)$$

2.-

$P(x,y) \ge 0$

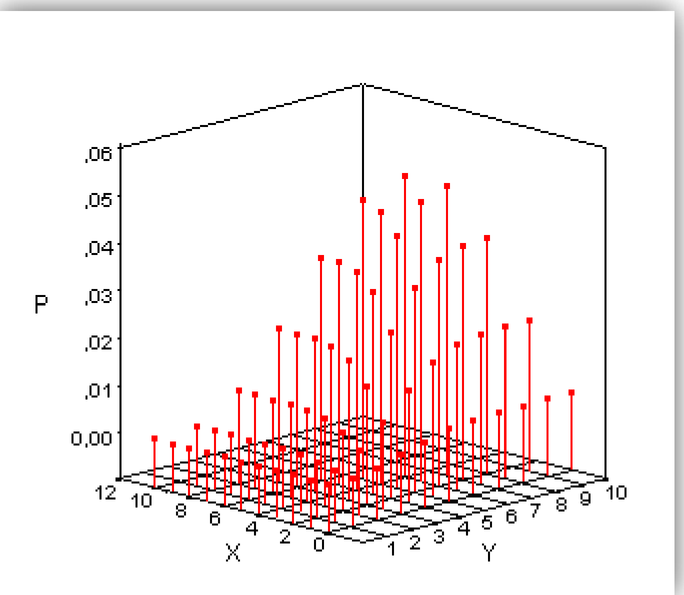
$P(x_1, x_2, \ldots, x_n) \ge 0$

3. -

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

$$\sum_{\forall x_1} \sum_{\forall x_2} \cdots \sum_{\forall x_n} P(x_1, x_2, \ldots, x_n) = 1$$

Así su representación sería:



FUNCIÓN DE DENSIDAD CONJUNTA de vectores aleatorios continuos

$$f(x,y) = \frac{\partial^2}{\partial x \, \partial y}\left[F(x,y)\right]$$

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \, \partial x_2 \ldots \partial x_n}\left[F(x_1, x_2, \ldots, x_n)\right]$$

propiedades

1.-

$f(x,y) \geq 0$

$f(x_1, x_2, \ldots, x_n) \geq 0$

2.-

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

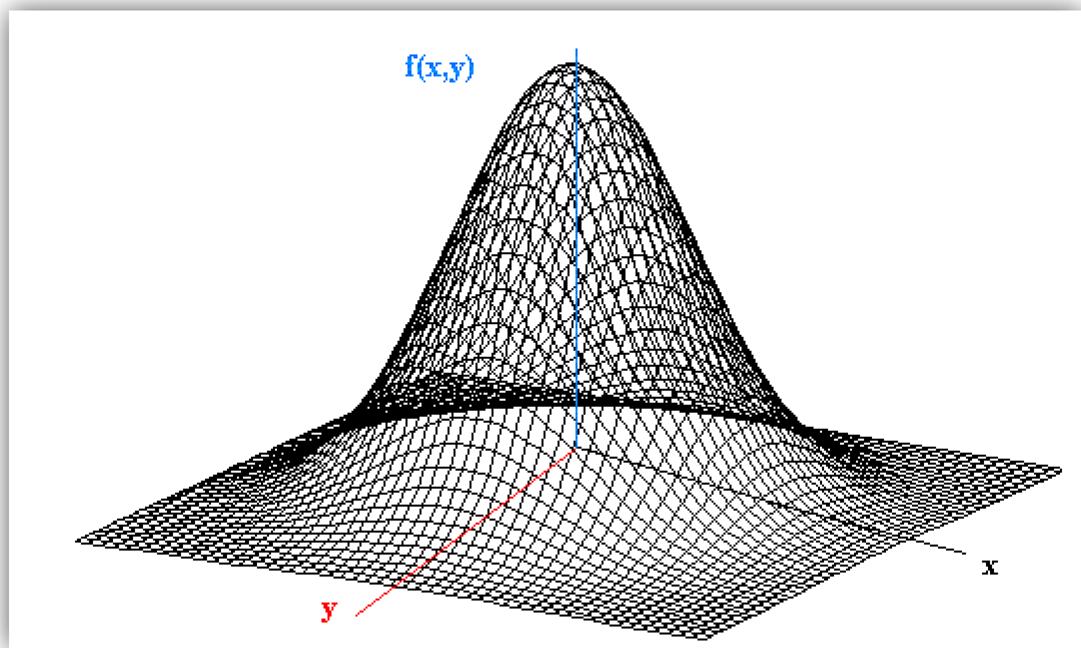
$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2 \ldots dx_n$$

3.-

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

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2 \ldots dx_n = 1$$

su representación sería



DISTRIBUCIONES MARGINALES DE PROBABILIDAD

CASO DISCRETO:

BIDIMENSIONAL MULTIDIMENSIONAL

$$P_1(x) = \sum_{\forall y} P(x,y) \qquad P_1(x_1) = \sum_{\forall x_2}\sum_{\forall x_3}\cdots\sum_{\forall x_n} P(x_1,x_2\ldots x_n)$$

$$P_2(y) = \sum_{\forall x} P(x,y) \qquad P_2(x_2) = \sum_{\forall x_1}\sum_{\forall x_3}\cdots\sum_{\forall x_n} P(x_1,x_2\ldots x_n)$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$P_n(x_n) = \sum_{\forall x_1}\sum_{\forall x_3}\cdots\sum_{\forall x_{n-1}} P(x_1,x_2\ldots x_n)$$

CASO CONTINUO:

BIDIMENSIONAL MULTIDIMENSIONAL

$$f_1(x) = \int_{\forall y} f(x,y)\,dy \qquad f_1(x_1) = \int_{\forall x_2}\int_{\forall x_3}\cdots\int_{\forall x_n} f(x_1,x_2,\ldots,x_n)dx_2,\ldots,dx_n$$

$$f_2(y) = \int_{\forall x} f(x,y)\,dx \qquad f_2(x_2) = \int_{\forall x_1}\int_{\forall x_3}\cdots\int_{\forall x_n} f(x_1,x_2,\ldots,x_n)dx_1,\ldots,dx_n$$

$$\cdots\cdots\cdots$$

$$f_n(x_n) = \int_{\forall x_1}\int_{\forall x_2}\cdots\int_{\forall x_{n-1}} f(x_1,x_2,\ldots,x_n)dx_1,\ldots,dx_{n-1}$$

DISTRIBUCIONES CONDICIONADAS

CASO DISCRETO: BIDIMENSIONAL

$$X/Y \quad (X/Y = y_0) \qquad P(x/Y = y_0) = \frac{P(x,y_0)}{P(y_0)}$$

$$Y/X \quad (Y/X = x_0) \qquad P(y/X = x_0) = \frac{P(x_0,y)}{P(x_0)}$$

CASO CONTINUO: BIDIMENSIONAL

$$X/Y \qquad f(x/y) = \frac{f(x,y)}{f(y)}$$

$$Y/X \qquad f(y/x) = \frac{f(x,y)}{f(x)}$$

INDEPENDENCIA ESTOCÁSTICA

Dos variables aleatorias X e Y se dice que son estocásticamente independientes cuando las distribuciones condicionadas y las marginales coinciden:

La independencia estocástica puede caracterizarse de una forma más operativa como:

X e Y son independientes $\Leftrightarrow$ P(x,y)= P(x) P(y), en el caso discreto,

o bien, X e Y son independientes $\Leftrightarrow$ f(x,y)= f(x) f(y) , en el caso continuo

ESPERANZA MATEMÁTICA EN DISTRIBUCIONES n-DIMENSIONALES.

Dado un vector aleatorio n-dimensional (v.a. n-dimensional), $X_1,X_2,...,X_n$, y dada una función cualquiera $g(X_1,X_2,...,X_n)$ llamaremos esperanza matemática de $g(X_1,X_2,...,X_n)$ a la expresión:

$$E\left[g(x_1,x_2,...,x_n)\right]=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,x_2,\dots,x_n)f(x_1,x_2,\dots,x_n)\quad dx_1dx_2\dots dx_n$$

para caso continuo

$$=\sum\sum\dots\sum g(x_1,x_2,...,x_n).P(x_1,x_2,...,x_n)$$

para caso discreto

Si la función g( ) no es completa (no afecta a todas las variables) la expresión del operador sólo dependerá de las variables afectadas. Así por ejemplo:

$E\left[x_1\right]$ quedaría tan sólo como la esperanza tomada sobre la distribución marginal:

$$E[x_1] = \int_{-\infty}^{\infty} x_1 f_1(x_1) dx_1$$

caso continuo

$$E[x_1] = \sum_{\forall i} x_1 \cdot P_1(x_1)$$

caso discreto

INDICADORES FUNDAMENTALES DE UNA DISTRIBUCIÓN

n-DIMENSIONAL

Al igual que ocurría en el caso de distribuciones de frecuencias los indicadores fundamentales de la distribución son únicamente los que se pueden obtener a partir de las distribuciones marginales unidimensionales (media, varianza, etc.) y de cada par de variables (covarianza y coeficiente de correlación)

COVARIANZA Y CORRELACIÓN ENTRE VARIABLES ALEATORIAS

La **covarianza** entre dos v.a. X e Y se define como:

$$\sigma_{xy} = E[(x-\mu_x)(y-\mu_y)] = E[x.y] - \mu_x \mu_y$$

siendo E[x.y] el momento ordinario mixto de orden 1,1 ($\alpha_{1,1}$):

$$E[x,y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(xy)dxdy$$

caso continuo

$$E[x,y] = \sum_{\forall i}\sum_{\forall j} x_i y_i P(x_i y_j)$$

caso discreto

La covarianza nos informa de la covariación conjunta de las dos variables, análogamente a como ocurría en el caso de distribuciones de frecuencias. Sólo hay que considerar aquí que la covariación hay que entenderla en términos de probabilidad y no de frecuencia: una covariación positiva sería el que a valores altos de una de la variables le corresponden "**con mayor probabilidad**" valores altos de la otra y a valores bajos valores bajos.(Una covariación negativa , inversamente).

Al igual también que en el caso de las distribuciones de frecuencias puede obtenerse un indicador de la **correlación**
(covariación estandarizada, relativizada y que permite la comparación al estar acotada):

Se define, entonces el **coeficiente de correlación** como:

$$\rho = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Indicador que viene a tener al mismo sentido que en las distribuciones de frecuencias y que igualmente está acotado entre -1 y 1.

Si dos variables aleatorias son estocásticamente independientes su coeficiente de correlación (y su covarianza es cero).

Sin embargo el resultado recíproco no es necesariamente cierto (dos variables incorrelacionadas no tienen porqué ser independientes, por lo general)

VECTOR DE MEDIAS (centro de gravedad) de una distribución de probabilidad

vector de medias**:** Es el vector columna formado por las medias de las distribuciones marginales univariantes:

$$M = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdot\cdot \\ \mu_n \end{pmatrix}$$

MATRIZ DE VARIANZA (DE VARIANZAS Y COVARIANZAS; DE COVARIANZAS O DE MOMENTOS) Y DE CORRELACIÓN

matriz de varianzas: Es la matriz n× n formada por las varianzas y covarianzas:

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdot & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{n1} & \sigma_{n2} & \cdot & \sigma_n^2 \end{pmatrix}$$

matriz de correlación: Es la matriz n× n formada por los coeficientes de correlación:

$$R = \begin{pmatrix} 1 & \rho_{12} & . & \rho_{1n} \\ \rho_{21} & 1 & . & . \\ . & . & . & . \\ \rho_{n1} & \rho_{n2} & . & 1 \end{pmatrix}$$

REGRESION DE VARIABLES ALEATORIAS

Al igual que en estadística descriptiva se puede concebir la regresión como un procedimiento de expresar una variable como función de otra u otras.

Análogamente a como ocurría allí:

La <u>regresión Y/X</u> en sentido estricto se define como: E[Y/X]

La <u>regresión X/Y</u> en sentido estricto se define como: E[X/Y]

y la regresión en sentido estricto puede ajustarse a alguna función analítica. En el caso lineal (y por el método de mínimos cuadrados) resulta que:

regresión lineal Y/X: $Y^{*}=\mu_y+ (\rho/\sigma^2_x(x-\mu_x))$

regresión lineal X/Y : $X^{*}=\mu_x+ (\rho/\sigma^2_y(y-\mu_y))$

ESPERANZA DE UN VECTOR .Linealidad

El operador esperanza puede extenderse para que actúe sobre un vector aleatorio, sin más que considerar que la esperanza de un vector es el vector formado por las esperanzas.

De esta manera: dado el vector aleatorio

$$X = \begin{pmatrix} x_1 \\ x_2 \\ . \\ x_n \end{pmatrix}$$

la esperanza del vector vendrá dada por:

$$E(X) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ . \\ E(x_n) \end{pmatrix}$$

PROPIEDADES:

Si se realiza una transformación lineal del vector X de manera que el nuevo vector aleatorio k-dimensional Z sea:

$$Z = \mathbf{A}X + \mathbf{b}$$

Donde

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1} & a_{k2} & \cdot & a_{kn} \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_k \end{pmatrix}$$

Entonces $E[Z] = \mathbf{A}\, E[X] + \mathbf{b}$:

LA ESPERANZA DE UNA TRANSFORMACIÓN LINEAL DE UN VECTOR ALEATORIO ES LA TRANSFORMACIÓN LINEAL DE LA ESPERANZA DEL VECTOR

VARIANZA DE UN VECTOR. OPERADOR VARIANZA

El operador varianza puede generalizarse para aplicarse sobre un vector aleatorio de forma que se podrá definir como:

$$\mathbf{V(X) = E\Big[(X\text{-}E(X))\ (X\text{-}E(X))\ '\Big]}$$

El resultado de aplicar el operador a un vector aleatorio es, obviamente la matriz de varianzas-covarianzas.

PROPIEDADES:

Si se realiza una transformación lineal del vector X de manera que el nuevo vector aleatorio k-dimensional Z sea:

$$\mathbf{Z = AX + b}$$

donde:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1} & a_{k2} & \cdot & a_{kn} \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_k \end{pmatrix}$$

RESULTADOS COMPLEMENTARIOS IMPORTANTES:

1) Dada dos o más variables aleatorias independientes si las sumamos obtendremos una nueva variable aleatoria cuya F.G.M.(F.C.) será el producto de las F.G.M. (F.C) de las variables originales.

2) Distribuciones reproductivas por adición.

Un modelo de probabilidad (distribución-tipo de probabilidad) *univariante* se dice que es reproductivo por adición o que verifica el teorema de adición cuando se cumple la propiedad siguiente:

Dadas dos o más variables aleatorias que tengan por distribución ese modelo de probabilidad y que sean (todas) independientes, la suma de ellas es una nueva variable aleatoria que tiene ese mismo modelo de probabilidad y cuyos parámetros son la suma de los parámetros

Modelos que cumplen la reproductividad aditiva (Teorema de adición):

BINOMIAL (para el parámetro n): la suma de dos o más variables binomiales independientes es una binomial de parámetro n la suma de los parámetros n.

POISSON (para el parámetro λ): la suma de dos o más variables de poisson independientes es una variable de poisson con parámetro λ , la suma de los parámetros λ .

NORMAL (para la media y la **VARIANZA**): la suma de dos o más variables normales independientes es una variable normal con media la suma de las medias y con varianza la suma de las varianzas (con D.Típica la raíz cuadrada de la suma de los cuadrados de las desviaciones típicas).

Todas estas reproductividades pueden probarse fácilmente utilizando el resultado 1) y teniendo en cuenta el efecto caracterizador de la F.G.M (o de la F.C.)

3) PROPIEDAD FUNDAMENTAL (TEOREMA FUNDAMENTAL) DE LAS DISTRIBUCIONES NORMALES."Una combinación lineal cualquiera de variables normales independientes es también normal y tendrá por media la misma combinación lineal de las medias y por varianza la combinación, con los coeficientes al cuadrado, de las varianzas.