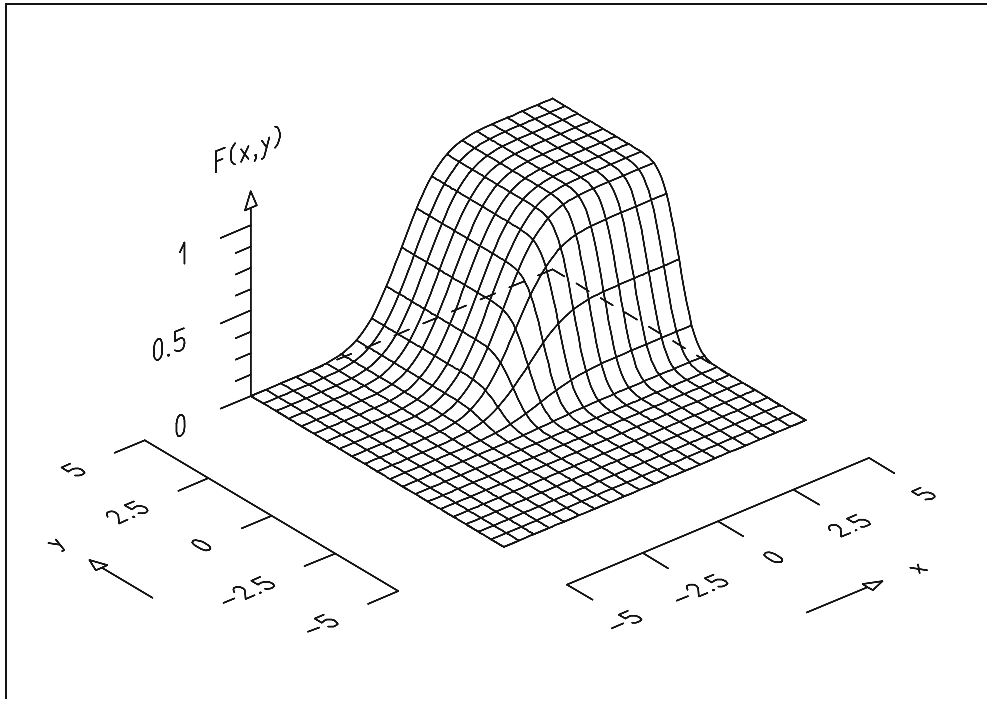
FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Sea x e y dos variables aleatorias

$$F(x, y) = P(\mathsf{x} < x, \mathsf{y} < y)$$



Ejemplo:

a) x e y pueden ser la Latitud y Longitud en la posición de un objeto.
b) x e y pueden ser la Presión y Temperatura de un gas

Es decir, x e y son dos variables aleatorias cualesquiera que las utilizaré de manera conjunta y lo que deseo conocer es sus propiedades estadísticas de manera de poder trabajar criteriosamente con ambas variables.

Haciendo una extensión de una variable, para dos variables definiré una serie de funciones útiles para su descripción estadística.

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Función de densidad de probabilidad conjunta

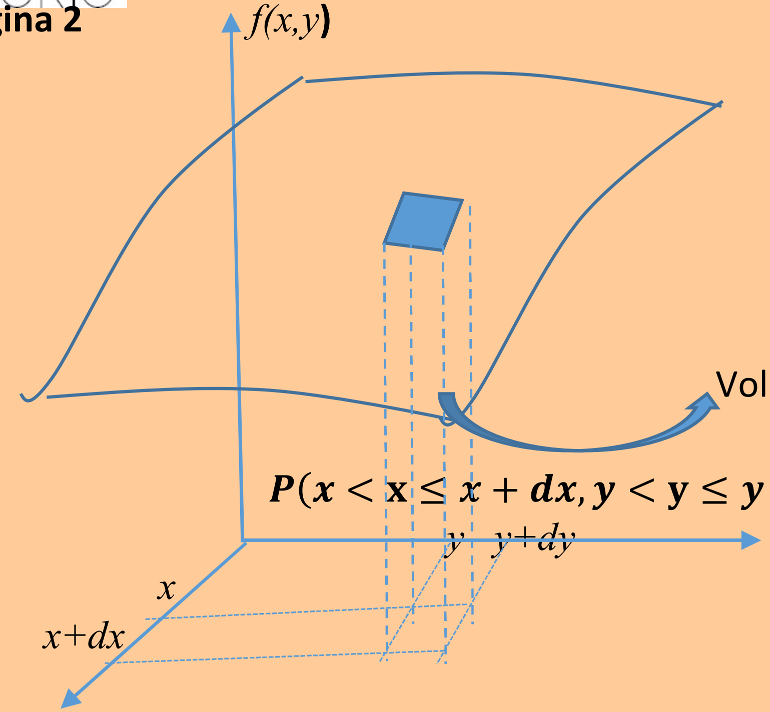
$$f(x, y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x, y)$$

$$P(a \leq x < b, c \leq y < d) = \int_a^b \left[\int_c^d f(x, y)\,\mathrm{d}y\right]\mathrm{d}x \quad .$$

Función de densidad de probabilidad marginal de x, $g(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y$

$$P(a \leq x < b, -\infty < y < \infty) = \int_a^b \left[\int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y\right]\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x$$

Apunte: Pagina 2



Volumen bajo la superficie:

$$\boldsymbol{P(x < \mathbf{x} \leq x + dx, y < \mathbf{y} \leq y + dy) = f(x,y)dx\, dy}$$

Nota: $\iint_{-\infty}^{+\infty} f(x,y)dx\, dy = 1$

Si la variable x e y son Independientes

a) ¿Cómo podemos escribir la función de densidad de probabilidad conjunta? Demuestre.

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Independencia entre las variables x e y

$$f(x, y) = g(x)h(y)$$

Probabilidad condicional

$$P(y \leq \mathsf{y} < y + \mathrm{d}y \,|\, x \leq \mathsf{x} \leq x + \mathrm{d}x)$$

La función de densidad de probabilidad de y, si x:

$$f(y|x) = \frac{f(x, y)}{g(x)}$$

$$f(y|x)\,\mathrm{d}y$$

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

La regla de probabilidad total puede escribirse como

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x = \int_{-\infty}^{\infty} f(y|x) g(x)\, \mathrm{d}x$$

Si existe independencia entre las variables x e y

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{g(x)h(y)}{g(x)} = h(y)$$

¿Cuáles son los parámetros para el caso de trabajar con dos variables x e y?

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Parámetros que caracterizan el comportamiento de **dos variables** aleatorias, resultado de un experimento aleatorio.

- Nuevo parámetro para cuando trabajo con dos ……

- Valor esperado cuando trabajo con dos variables:

$$E\{H(\mathrm{x},\mathrm{y})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)\,f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$\sigma^2\{H(\mathrm{x},\mathrm{y})\} = E\{[H(\mathrm{x},\mathrm{y}) - E(H(\mathrm{x},\mathrm{y}))]^2\}$$

Cuando trabajo con dos variables, se tiene que los parámetros fundamentales son los mismos que los presentados para una variable (la media y la varianza) pero para cada variable y a estos se suma un parámetro que me describa la dependencia estadística entre los mismos, el cual definimos como covarianza.

$$E[\mathrm{x}] = \int_{-\infty}^{+\infty} x\, g(x)dx = \mu_{\mathrm{X}}$$

$$\sigma_x^2 = E[(\mathrm{x} - \mu_{\mathrm{X}})^2] = \int_{-\infty}^{+\infty} (x - \mu_{\mathrm{X}})^2\, g(x)dx$$

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Tenemos media en x, media en y, varianza en x y varianza en y

$$E[\mathrm{x}] = \int_{-\infty}^{+\infty} x\, g(x)dx = \mu_{\mathrm{X}}$$

Sería una extensión de lo que ya vimos para una variable, donde la f.d.p es la función de densidad de probabilidad marginal de la variable

$$\sigma_x^2 = E[(\mathrm{x}-\mu_{\mathrm{X}})^2] = \int_{-\infty}^{+\infty} (x-\mu_{\mathrm{X}})^2\, g(x)dx$$

Y un par equivalente para la variable y ……

Una función en x e y lineal se tiene que:

$$H(\mathsf{x},\mathsf{y}) = a\mathsf{x} + b\mathsf{y}$$

$$E(a\mathsf{x}+b\mathsf{y}) = aE(\mathsf{x}) + bE(\mathsf{y})$$

$\sigma^2(a\mathsf{x}+b\mathsf{y})$??

Ejemplo a resolver:

Se dispone de la información del consumo de gas natural (variable x), expresada en 100 metros cúbicos, además del consumo de energía eléctrica (variable y) expresada en 100 kW, de una serie de viviendas ubicadas en el sector suroeste de la capital en el mes de abril pasado. Se sabe que la función de densidad conjunta de dichas variables es la siguiente:

$$f(x,y) = \begin{cases} \frac{x+y}{24} & si\ 0 < x < 2;\ 0 < y < 4 \\ 0 & resto \end{cases}$$

Se tiene la intención de revisar los medidores de aquellos hogares donde el consumo de gas y energía eléctrica no sobrepasa las respectivas cantidades esperadas.

a) ¿Qué porcentaje de los hogares de este sector se deberían revisar?
b) De los hogares con un consumo de 100 m^3mensuales en gas ¿Qué proporción consume menos de 100 KW en energía eléctrica?

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Una función en x e y lineal se tiene que:

$H(x, y) = ax + by,$ $\qquad E(ax + by) = aE(x) + bE(y)$

$\sigma^2(ax + by)$

$$E\left\{\left[(ax+by) - E[ax+by]\right]^2\right\} = E\left\{\left[ax + by - a\mu_x - b\mu_y\right]^2\right\}$$

$$= E\left\{\left[a(x-\mu_x) + b(y-\mu_y)\right]^2\right\}$$

$$= E\left\{a^2(x-\mu_x)^2 + b^2(y-\mu_y)^2 + 2ab(x-\mu_x)(y-\mu_y)\right\}$$

$$= a^2 E[(x-\mu_x)^2] + b^2 E[(y-\mu_y)^2] + 2ab\, E[(x-\mu_x)(y-\mu_y)]$$

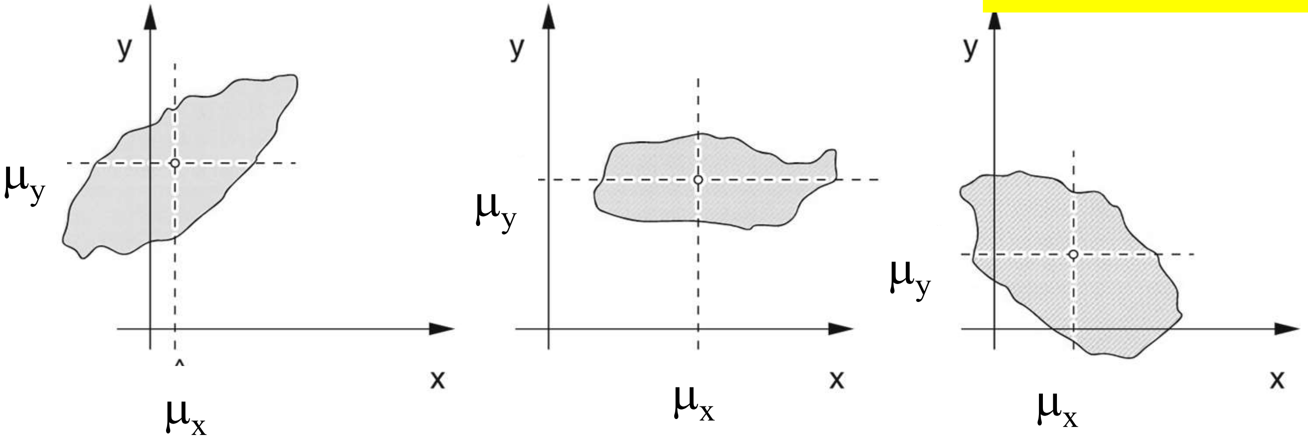
$cov(x,y) = \sigma_{xy}$

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

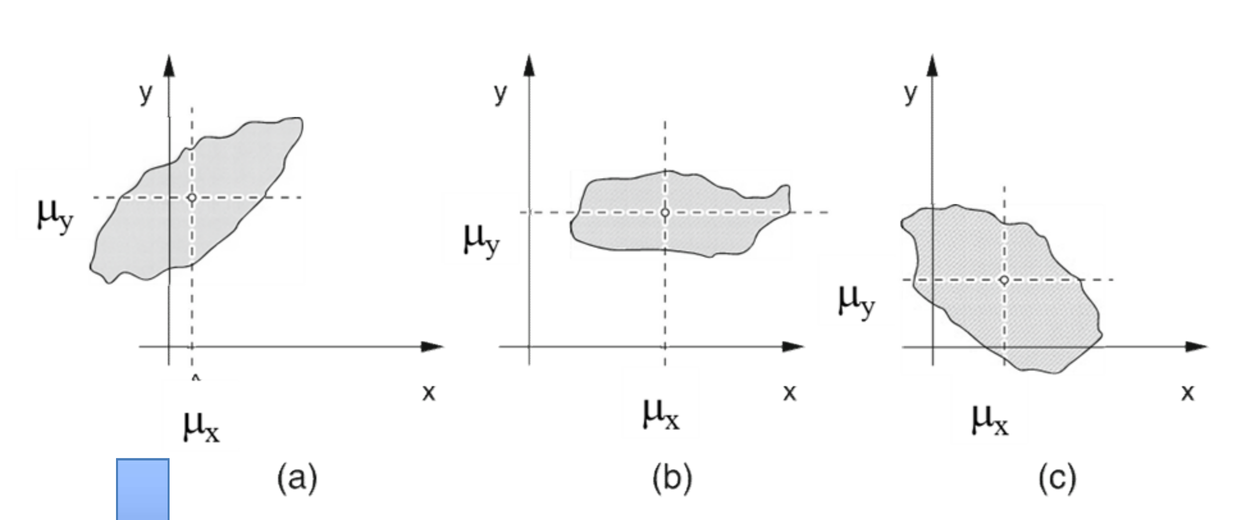
$$\sigma^2(a\mathsf{x}+b\mathsf{y}) \;=\; a^2\sigma^2(\mathsf{x})+b^2\sigma^2(\mathsf{y})+2ab\,\boxed{\mathrm{cov}(\mathsf{x},\mathsf{y})}$$

$$cov(\mathrm{x,y}) = \iint_{-\infty}^{+\infty}(x-\mu_x)(y-\mu_y)f(x,y)\,dx\,dy$$

mide la dependencia estadística entre la variable x e y



(a) (b) (c)



$$\sum\sum(x\mathrm{i} - \mu_x)(y\mathrm{j} - \mu_y)P(xi, xj)$$

Mayoritariamente:

$(x\mathrm{i} - \mu_x) > 0 \; y \; (y\mathrm{j} - \mu_y) > 0$

$(x\mathrm{i} - \mu_x) < 0 \; y \; (y\mathrm{j} - \mu_y) < 0$

$(x\mathrm{i} - \mu_x)(y\mathrm{j} - \mu_y) > 0$

$\sigma_{\mathrm{xy}} = cov(\mathrm{x,y}) > 0$

MAGGIA LABORATORIO

FUNCIÓN DE DISTRIBUCION DE DOS VARIABLES

Coeficiente de correlación: mide la dependencia estadística pero independientemente de la desviación inherente de cada variable

$$\rho(\mathsf{x},\mathsf{y}) = \frac{\operatorname{cov}(\mathsf{x},\mathsf{y})}{\sigma(\mathsf{x})\sigma(\mathsf{y})}$$

Características del coeficiente de correlación:

a) $-1 \leq \rho(\mathsf{x},\mathsf{y}) \leq 1$.

b) Si x e y son variables independientes entonces $\rho(\mathsf{x},\mathsf{y})$ es igual a 0

c) Si y es una función determinista de x, H(x), entonces $\rho(\mathsf{x},\mathsf{y})$ =+/- 1

MAGGIA LABORATORIO

- Demostración del punto a)

Recurrimos de una variable intermediaria, sea t variable real, encontraremos las raíces de $q(t)$=E[(v+tu)^2] donde u=x-E[x] y v=y-E[y];

Siempre es mayor que 0, entonces q(t)≥ 0

$$q(t) = E[\mathrm{v}^2] + 2\ t\ E[\mathrm{vu}] + t^2\ E[\mathrm{u}^2]$$

$q(t) = 0$ solo presenta un punto igual a cero o ninguno, si no tiene ninguna raíz real el discriminante debe ser < 0.

$$4\ (E[\mathrm{uv}])^2 \text{- } 4\ \mathrm{E}[\mathrm{v}^2]\ \mathrm{E}[\mathrm{u}^2] \leq 0$$

$$\frac{(E[\mathrm{uv}])^2}{\mathrm{E}[\mathrm{v}^2]\ \mathrm{E}[\mathrm{u}^2]} \leq 1$$

$$\frac{E[(\mathrm{x\text{-}E[x]})(\mathrm{y\text{-}E[y]})]}{\sqrt{\mathrm{E}[(\mathrm{x}-\mathrm{E}[\mathrm{x}])^2]\ \mathrm{E}[(\mathrm{y}-\mathrm{E}[\mathrm{y}])^2]}} \leq \mp 1$$

Finalmente:

$$-1 \leq \rho(\mathrm{x},\mathrm{y}) \leq 1 \quad .$$

- Demostración del punto c)

Si y=H(x) , donde H es una función analítica. Realizo un desarrollo en Taylor entorno al valor más representativo.

$$\mathrm{y} = H(\mu_x) + H^{'}(\mu_x)(x - \mu_x) + H^{''}(\mu_x)(x - \mu_x)^2 + \cdots$$

Aproximamos a primer orden, entonces podemos escribir

$\mathrm{y} = \alpha + \beta\mathrm{x}$, teniendo en cuenta que $\sigma_{\mathrm{y}}^2 = \beta^2\sigma_{\mathrm{x}}^2$

y que $\sigma_{\mathrm{xy}} = E[(\mathrm{x}\text{-}\mathrm{E}[\mathrm{x}])(\mathrm{y}\text{-}\mathrm{E}[\mathrm{y}])] = \beta\sigma_{\mathrm{x}}^2$

$$\rho_{\mathrm{xy}} = \mp 1$$