

1 Introducción

En la vida real es muy frecuente enfrentarse a problemas en los que interesa analizar varias características simultáneamente, como por ejemplo la velocidad de transmisión de un mensaje y la proporción de errores. De esta forma seremos capaces de estudiar no sólo el comportamiento de cada variable por separado, sino las relaciones que pudieran existir entre ellas. El objetivo principal de este tema es elaborar un modelo matemático que permita analizar experimentos aleatorios en los que cada resultado experimental tiene asociados varios valores, en general numéricos. Estos modelos serán la base para la construcción de los modelos inferenciales necesarios para extraer los resultados de una muestra a la población.

2 Vectores aleatorios

El concepto de vector aleatorio nace como una generalización natural de la noción de variable aleatoria, al considerar simultáneamente el comportamiento de varias características asociadas a un experimento aleatorio.

Definición 2.1 (Vector aleatorio.). *Dado un espacio probabilístico (E, \mathcal{A}, P) y el espacio probabilizable (\mathbb{R}^p, β) con β σ -álgebra de Borel en \mathbb{R}^p , se dice que una aplicación $\mathbf{x} = (x_1, \dots, x_p)^t$*

$$\mathbf{x} : E \longrightarrow \mathbb{R}^p$$

es un vector aleatorio si es medible, es decir $\mathbf{x}^{-1}(B) \in \mathcal{A}$ $\forall B \in \beta$, por tanto cada una de sus componentes x_i , $i = 1, \dots, p$ es una variable aleatoria.

Veamos algunos ejemplos sencillos de vectores aleatorios.

El vector $(X, Y) = (x_1, x_2)$ representa la temperatura máxima que puede alcanzar una resistencia y el tiempo que tarda en alcanzarla.

Otro ejemplo consistiría en analizar la edad, peso, estatura, sexo, colesterol y triglicéridos de una persona como consecuencia se obtendría un vector aleatorio de dimensión seis, con componentes $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)^t$.

Definición 2.2 (Probabilidad inducida.). *Dado un vector aleatorio \mathbf{x} definido sobre (E, \mathcal{A}, P) , se denomina probabilidad inducida por el vector a la aplicación $P_{\mathbf{x}} : \beta \longrightarrow \mathbb{R}^p$ definida por:*

$$P_{\mathbf{x}}(B) = P[\mathbf{x}^{-1}(B)] \quad \forall B \in \beta$$

$(\mathbb{R}^p, \beta, P_{\mathbf{x}})$ es el espacio probabilístico inducido por \mathbf{x} .

La distribución de probabilidad inducida por un vector aleatorio se puede caracterizar mediante la función de distribución.

Definición 2.3 (Función de distribución.). *Dado el vector aleatorio \mathbf{x} se denomina función de distribución asociada, a la función $F : \mathbb{R}^p \rightarrow \mathbb{R}$ definida por:*

$$F(\mathbf{a}) = P(x_1 \leq a_1, \dots, x_p \leq a_p)$$

Las propiedades más importantes de las funciones de distribución son:

1. $F(a_1, \dots, a_{i-1} - \infty, a_{i+1}, \dots, a_p) = 0$.
2. $\lim_{x_i \rightarrow \infty, \forall i} F(\mathbf{x}) = 1$.
3. F es continua por la derecha respecto de cada variable.
4. En el caso bidimensional si $a_1 \leq a_2 ; b_1 \leq b_2$.

$$P[(a_1 < x_1 \leq a_2) \cap (b_1 < x_2 \leq b_2)] = F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1) \geq 0$$

5. F es no decreciente respecto a cada variable.

Los vectores aleatorios se pueden clasificar teniendo en cuenta el tipo de las variables aleatorias que lo componen.

3 Vectores aleatorios discretos

La idea intuitiva asociada a un vector aleatorio discreto es que cada una de sus componentes sean v. a. discretas.

Definición 3.1 (Vector aleatorio discreto). *Sea \mathbf{x} un vector aleatorio definido sobre (E, \mathcal{A}, P) y $S = \{\mathbf{a} \in \mathbb{R}^p / P(\mathbf{a}) > 0\}$, se dice que \mathbf{x} es discreto si se cumple que:*

1. $S \neq \emptyset$
2. El cardinal de S es a lo sumo infinito numerable.
3. $\sum_{\mathbf{a} \in S} P(\mathbf{a}) = 1$.
4. $P[\mathbf{x} \in B] = \sum_{\mathbf{a} \in B \cap S} P(\mathbf{a})$.

Al conjunto S se le llama conjunto soporte del vector \mathbf{x} .

En este caso la función de distribución es discontinua y viene dada por la expresión:

$$F(\mathbf{x}) = \sum_{\mathbf{a} \in S_x} P(\mathbf{a})$$

con $S_x = \{\mathbf{y} \in \mathbb{R}^p / \mathbf{y} \leq \mathbf{x}\}$.

4 Vectores aleatorios continuos

En los vectores aleatorios continuos cada componente es una v. a. de tipo continuo y para caracterizar su distribución es usual emplear la función de densidad.

Definición 4.1 (Función de densidad). *Una función $f : \mathbb{R}^p \rightarrow \mathbb{R}$ se dice que es función de densidad si verifica las condiciones:*

1. $f(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^p$
2. $\int_{\mathbb{R}^p} f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{x}) d\mathbf{x} = 1$

Definición 4.2 (Vector aleatorio continuo). *Un vector aleatorio \mathbf{x} se dice que es continuo si su función de distribución se puede expresar como:*

$$F(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f(\mathbf{y}) d\mathbf{y} \quad \forall \mathbf{x} \in \mathbb{R}^p$$

donde f es una función de densidad en \mathbb{R}^p .

El conjunto $S_{\mathbf{x}} = \{\mathbf{a} \in \mathbb{R}^p / f(\mathbf{a}) > 0\}$ se denomina conjunto soporte de \mathbf{x} . Las propiedades más importantes de los vectores continuos son:

1. La función de distribución es continua.
2. En los puntos de continuidad de la función de densidad se cumple que:

$$f(\mathbf{x}) = \frac{\partial^p F(\mathbf{x})}{\partial x_1 \dots \partial x_p}$$
3. $S = \{\mathbf{a} \in \mathbb{R}^p / P_{\mathbf{x}}(\mathbf{a}) > 0\} = \emptyset$, es decir, $P(\mathbf{a}) = 0 \quad \forall \mathbf{a} \in \mathbb{R}^p$
4. $P[\mathbf{x} \in B] = \int_B f(\mathbf{x}) d\mathbf{x}$

5 Distribuciones marginales

Las componentes de los vectores aleatorios son v. a. y resulta en muchos casos de interés de estudiar el comportamiento de cada una de esas variables por separado o de un subconjunto de las mismas. Para abordar este problemas se definen las distribuciones marginales.

Definición 5.1 (Distribuciones marginales). *Sea \mathbf{x} un vector aleatorio con función de distribución $F(\mathbf{x})$, entonces cada una de sus componentes, x_i es una variable aleatoria unidimensional, con la siguiente función de distribución, llamada distribución marginal de la componente x_i*

$$F_i(x_i) = F(+\infty, \dots + \infty, x_i, +\infty, \dots + \infty)$$

También se puede hablar de distribución marginal de un subvector, así si $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ se define la función de distribución marginal de \mathbf{x}_1 como $F_1(\mathbf{x}_1) = F(\mathbf{x}_1, +\infty, \dots + \infty)$

Si el vector es continuo se tiene la función de densidad marginal de la componente x_i , $f_i(x_i) = \int_{\mathbb{R}^{p-1}} f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) da_1 \dots da_{i-1} da_{i+1} \dots da_n$

Mientras que si \mathbf{x} es discreto la expresión para la distribución de probabilidad marginal de la componente x_i es:

$$P_i(x_i) = \sum_{\mathbf{y} \in S, y_i = x_i} P(\mathbf{y})$$

Análogamente se tendrían las distribuciones de los subvectores.

6 Distribuciones condicionadas

En muchos problemas reales puede interesar estudiar el comportamiento de una variable teniendo cierta información complementaria sobre el experimento. Por ejemplo sea (X, Y) un vector aleatorio cuyas componentes representan, respectivamente, las intensidades de las señales enviadas y recibidas a través de un canal de información. Un aspecto fundamental es este tipo de problemas es conocer el comportamiento de la señal recibida Y condicionada a que se ha enviado una señal $(X=x)$, es decir, analizar la distribución de $(Y / X=x)$. Otro problema muy interesante es el complementario del anterior, o sea, estudiar la señal que se envió cuando se conoce la señal recibida $(X / Y=y)$.

Cuando el suceso que condiciona tiene una probabilidad estrictamente positiva el problema se reduce a utilizar la probabilidad condicionada para calcular la distribución de $(\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a})$.

$$F_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}(\mathbf{b}) = \frac{P(\mathbf{x}_1 = \mathbf{a}, \mathbf{x}_2 \leq \mathbf{b})}{P(\mathbf{x}_1 = \mathbf{a})}$$

y si la variable es continua, la función de distribución es continua, función de densidad será:

$$f_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}(\mathbf{b}) = \left. \frac{\partial^{p_2} (F_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}(\mathbf{x}_2))}{\partial \mathbf{x}_2} \right|_{\mathbf{b}}$$

Sin embargo cuando el vector \mathbf{x}_1 es de tipo continuo, el suceso $(\mathbf{x}_1 = \mathbf{a})$ tiene siempre una probabilidad igual a cero, y no se puede emplear el método anterior. En este caso se define la función de densidad condicionada.

Definición 6.1 (Función de densidad condicionada). *Dado el vector aleatorio $(\mathbf{x}_1, \mathbf{x}_2)$ la función de densidad de \mathbf{x}_2 condicionada a $(\mathbf{x}_1 = \mathbf{a})$, se define como una función no negativa $f_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}$, que satisface:*

$$F_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}(\mathbf{b}) = \int_{\mathbf{t} \in \mathbb{R}^{p_2} / \mathbf{t} \leq \mathbf{b}} f_{\mathbf{x}_2 / \mathbf{x}_1 = \mathbf{a}}(\mathbf{t}) d\mathbf{t} \quad \forall \mathbf{x}_2 \in \mathbb{R}^{p_2}$$

El siguiente resultado permite calcular la función de densidad condicionada en el caso de trabajar con vectores aleatorios continuos.

Proposición .1. *Dado un vector aleatorio $(\mathbf{x}_1, \mathbf{x}_2)$ cuya función de densidad es $f(\mathbf{x}_1, \mathbf{x}_2)$ se verifica que en todo punto de continuidad de f en el que $f_{\mathbf{x}_2}(\mathbf{b}) > 0$, la función de densidad de $(\mathbf{x}_1 | \mathbf{x}_2 = \mathbf{b})$ existe y vale:*

$$f_{\mathbf{x}_1 | \mathbf{x}_2 = \mathbf{b}}(\mathbf{a}) = \frac{f(\mathbf{a}, \mathbf{b})}{f_{\mathbf{x}_2}(\mathbf{b})}$$

Definición 6.2 (Independencia de vectores aleatorios). *Dado el vector aleatorio $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ se dice que \mathbf{x}_1 y \mathbf{x}_2 son independientes cuando se verifica que:*

$$F_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) = F_{\mathbf{x}_1}(\mathbf{a}) F_{\mathbf{x}_2}(\mathbf{b}) \quad \forall (\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$$

Si trabajamos con la función de densidad esta condición es equivalente a:

$$f_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) = f_{\mathbf{x}_1}(\mathbf{a}) f_{\mathbf{x}_2}(\mathbf{b}) \quad \forall (\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$$

Cuando las variables son discretas la independencia equivale a:

$$P_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) = P_{\mathbf{x}_1}(\mathbf{a}) P_{\mathbf{x}_2}(\mathbf{b}) \quad \forall (\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$$

Una propiedad muy interesante es que si \mathbf{x}_1 y \mathbf{x}_2 son independientes entonces nuevas variables aleatorias $g(\mathbf{x}_1)$ y $h(\mathbf{x}_2)$, obtenidas como transformaciones de las anteriores, también los son.

También es interesante hacer notar que si dos vectores \mathbf{x}_1 y \mathbf{x}_2 son independientes no quiere decir que las componentes de \mathbf{x}_1 lo sean. Es decir en el caso multidimensional pueden aparecer muchos tipos de independencia como la independencia condicional

Definición 6.3 (Independencia condicional). *Dado el vector aleatorio $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ se dice que \mathbf{x}_1 y \mathbf{x}_2 son condicionalmente independientes dado \mathbf{x}_3 cuando se verifica que $\mathbf{x}_1 | \mathbf{x}_3$ y $\mathbf{x}_2 | \mathbf{x}_3$ son independientes.*

7 Momentos: vector de medias y matriz de varianzas-covarianzas

En este apartado se da la definición de momento para poder describir de manera resumida el comportamiento de los vectores aleatorios.

Definición 7.1 (Momentos respecto al origen). *Dado una vector aleatorio p -dimensional continuo \mathbf{x} se denomina momento respecto al origen de orden (r_1, \dots, r_p) , si existe, al valor dado por la expresión:*

$$a_{r_1, \dots, r_p}(\mathbf{x}) = E(x_1^{r_1} \dots x_p^{r_p}) = \int_{\mathbb{R}^p} x_1^{r_1} \dots x_p^{r_p} f(\mathbf{x}) d\mathbf{x}$$

donde $f(\mathbf{x})$ es la función de densidad.

Análogamente se definirá para el caso discreto.

Los momentos respecto al origen más importantes son los orden uno que permiten definir el vector esperanza del vector aleatorio que será muy importante en la caracterización de una v.a. p-dimensional.

Definición 7.2 (Vector esperanza). *Dado un vector aleatorio p-dimensional \mathbf{x} se denomina vector esperanza, $E(\mathbf{x})$, al vector μ de componentes $(\mu_1, \dots, \mu_p)^t$ definidas por :*

$$\mu_i = \int_{-\infty}^{\infty} x f_{x_i}(x) dx$$

Es inmediato comprobar que:

$$\mu_i = E(x_i) = a_{0, \dots, 0, \underline{1}_i, 0, \dots, 0}(\mathbf{x}).$$

Cuando se trabaja con una función $g(\mathbf{x})$ la esperanza de la nueva variable aleatoria se puede calcular mediante la expresión:

$$E[g(\mathbf{x})] = \int_{\mathbb{R}^p} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

Entre las propiedades más importantes del vector esperanza, si existe, se encuentran:

1 Linealidad para transformaciones $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, siendo A una matriz (n,p):

$$E[\mathbf{y}] = A E[\mathbf{x}] + \mathbf{b}.$$

$$2 E[g_1(\mathbf{x}_1)] = \int_{\mathbb{R}^p} g_1(\mathbf{x}_1) f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^{p_1}} g_1(\mathbf{x}_1) f_1(\mathbf{x}_1) d\mathbf{x}_1.$$

$$3 \text{ Linealidad } E[a_1 g_1(\mathbf{x}) + a_2 g_2(\mathbf{x})] = a_1 E[g_1(\mathbf{x})] + a_2 E[g_2(\mathbf{x})], \text{ siendo } a_i \in \mathbb{R}$$

4 Si \mathbf{x} es un v.a. con media μ se llama vector centrado a $\mathbf{x} - \mu$

Otros momentos muy importantes son los centrados respecto al vector de medias. La expresiones que aparecen a continuación se refieren al caso continuo, ya que el caso discreto se desarrolla de manera análoga a la realizada hasta este momento.

Definición 7.3 (Momentos centrados). *Dado una vector aleatorio p-dimensional continuo \mathbf{x} se denomina momento centrado de orden (r_1, \dots, r_p) , si existe, al valor dado por la expresión:*

$$m_{r_1, \dots, r_p}(\mathbf{x}) = \int_{\mathbb{R}^p} (x_1 - \mu_1)^{r_1} \dots (x_p - \mu_p)^{r_p} f(\mathbf{x}) d\mathbf{x}$$

donde $f(\mathbf{x})$ es la función de densidad.

Los momentos centrados de orden dos para una de las componentes $r_i = 2$, cero para el resto, $r_j = 0$ si $j \neq i$), dan lugar a las *varianzas* de cada componente x_i , que se va a denotar por σ_{ii} .

$$Var(x_i) = \sigma_{ii} = E(x_i - \mu_i)^2 = m_{0, \dots, 0, 2_i, 0, \dots, 0}(\mathbf{x})$$

El momento centrado de orden $(0, \dots, 0, \underline{1}_i, 0, \dots, 0, \underline{1}_j, 0, \dots, 0)$ se denomina *covarianza* entre la componentes x_i, x_j , y se va a denotar por σ_{ij} , tiene una gran importancia porque permitirá estudiar si las componentes están relacionadas linealmente.

Definición 7.4 (Matriz de Varianzas-Covarianzas). *Dado un vector aleatorio \mathbf{x} se define su matriz de varianzas-covarianzas, si existe, como:*

$$\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^t] = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix}$$

Entre sus propiedades destacan:

- Las matrices de varianzas-covarianzas son siempre matrices simétricas y semidefinidas positivas
- $\Sigma = E[\mathbf{x} \mathbf{x}^t] - \mu \mu^t$
- $\text{Var}(\mathbf{a}^t \mathbf{x}) = \mathbf{a}^t \Sigma \mathbf{a}$
- $\text{Var}(A\mathbf{x} + \mathbf{b}) = A \Sigma A^t$
- Si una matriz de varianzas covarianzas tiene determinante cero significa que hay una combinación lineal de las componentes del vector aleatorio que tiene varianza cero, por tanto una componente, al menos, se puede eliminar ya que es función lineal de las otras y no aportaría información nueva.
- Si \mathbf{x} es un v.a. con media μ y matriz de varianzas $\Sigma > 0$. La descomposición de Σ mediante el teorema de Jordan: $\Sigma = B \Lambda B^t$ (B matriz de vectores propios normalizados $B B^t = B^t B = I$ y Λ matriz de valores propios, que por ser simétrica y definida positiva son todos reales y positivos) conduce a la matriz inversible $C = B(\Lambda)^{1/2}$ de modo que $\Sigma = C * C^t$ que es usada para la tipificación del vector \mathbf{x} , $C^{-1}(\mathbf{x} - \mu) = \Lambda^{-1/2} B^t(\mathbf{x} - \mu)$.

También se pueden definir matrices de varianzas-covarianzas entre dos vectores aleatorios definidos sobre el mismo espacio probabilístico,

Definición 7.5. *Dado un vector aleatorio p -dimensional, \mathbf{x} , y otro q -dimensional, \mathbf{y} , se define su matriz de varianzas-covarianzas, si existe, como:*

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = E[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)^t] = \begin{pmatrix} \sigma_{x_1, y_1} & \dots & \sigma_{x_1, y_q} \\ \vdots & \ddots & \vdots \\ \sigma_{x_p, y_1} & \dots & \sigma_{x_p, y_q} \end{pmatrix}$$

Entre sus propiedades destacan:

- $\Sigma = \text{Cov}(\mathbf{x}, \mathbf{x})$
- $\text{Cov}(\mathbf{x}, \mathbf{y}) = (\text{Cov}(\mathbf{y}, \mathbf{x}))^t$

- $\text{Cov}(\mathbf{Ax}, \mathbf{By}) = \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}^t$
- $\text{Cov}(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = \text{Cov}(\mathbf{x}_1, \mathbf{y}) + \text{Cov}(\mathbf{x}_2, \mathbf{y})$
- Si los vectores tienen la misma dimensión
 $\text{Var}(\mathbf{x} + \mathbf{y}) = \Sigma_{\mathbf{x} + \mathbf{y}} = \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y}) + \text{Cov}(\mathbf{y}, \mathbf{x}) + \text{Var}(\mathbf{y})$
- Si \mathbf{x} e \mathbf{y} son dos vectores aleatorios independientes entonces $\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, pero el recíproco no tiene porque ser cierto.

Lema 7.1 (Desigualdad de Cauchy-Schwartz). *Dado un vector aleatorio bidimensional (X, Y) en el que existen los momentos centrados de orden dos, se verifica que:*

$$E^2[X Y] \leq E(X^2) E(Y^2)$$

además la igualdad se alcanza si y solo si existen dos números reales α y β , con al menos uno de ellos distinto de cero, tales que $P[\alpha X + \beta Y = 0] = 1$.

Definición 7.6 (Coeficiente de correlación de Pearson). *Dadas dos componentes x_i, x_j de un vector aleatorio \mathbf{x} se denomina coeficiente de correlación de Pearson entre esas componentes, ρ_{ij} a la expresión:*

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}$$

Las propiedades más importantes del Coeficiente de correlación de Pearson son:

1. $-1 \leq \rho_{ij} \leq 1$.
2. ρ_{ij} alcanza los extremos si y solo si x_j es una función lineal de x_i , o al revés.
3. Si $\rho_{ij}=0$, se dice que x_i e x_j son linealmente independientes, pero en general la independencia lineal no implica la independencia estadística.
4. Si x_i e x_j son independientes entonces son linealmente independientes.

Definición 7.7 (Coeficiente de correlación multiple). *Dadas la componente x_1 y el vector $\mathbf{y} = (x_2, \dots, x_p)$ de un vector aleatorio \mathbf{x} se denomina coeficiente de correlación multiple entre entre x_1 y el vector \mathbf{y} esas componentes, $\rho_{1,2,\dots,p}^2$ a la máxima correlación entre x_1 y una combinación lineal de las componentes de \mathbf{y}*

$$\rho_{1,2,\dots,p}^2 = \max_b \rho_{x_1, b^t y}^2 = \max_b \frac{b^t \Sigma_{y, x_1} \Sigma_{x_1, y} b}{b^t \Sigma_y b \sigma_{11}}$$

Definición 7.8 (Matriz de correlación de Pearson). *Dado un vector aleatorio \mathbf{x} se denomina matriz de correlación entre sus componentes a:*

$$R = \begin{pmatrix} 1 & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \dots & 1 \end{pmatrix}$$

Definición 7.9 (Función generatriz de momentos). *Dado un vector aleatorio \mathbf{x} p -dimensional la función generatriz de momentos, si existe, se define como:*

$$g_{\mathbf{x}}(\mathbf{t}) = E(e^{t^t \mathbf{x}}) = \int_{\mathbb{R}^p} e^{t^t \mathbf{x}} f(\mathbf{x}) d\mathbf{x}$$

Entre las propiedades más interesantes de la función generatriz de momentos están las siguientes:

1. $g(\mathbf{0})=1$
2. g cuando existe caracteriza la distribución.
3. Los vectores aleatorios \mathbf{x}, \mathbf{y} son independientes si y solo si la función generatriz de momentos conjunta es igual al producto de las funciones generatrices de las marginales.

$$g_{\mathbf{x},\mathbf{y}}(\mathbf{t}_1, \mathbf{t}_2) = g_{\mathbf{x}}(\mathbf{t}_1) g_{\mathbf{y}}(\mathbf{t}_2) \quad \forall (\mathbf{t}_1, \mathbf{t}_2) \in \text{entorno de } \mathbf{0} \in \mathbb{R}^{p+q}$$

4. Si \mathbf{x}, \mathbf{y} son vectores aleatorios independientes entonces la función generatriz de $\mathbf{x} + \mathbf{y}$ verifica:

$$g_{\mathbf{x}+\mathbf{y}}(\mathbf{t}) = g_{\mathbf{x}}(\mathbf{t}) g_{\mathbf{y}}(\mathbf{t})$$

Definición 7.10 (Función característica). *Dado un vector aleatorio \mathbf{x} p -dimensional la función característica se define como:*

$$\phi_{\mathbf{x}}(\mathbf{t}) = E(e^{it^t \mathbf{x}}) = \int_{\mathbb{R}^p} e^{it^t \mathbf{x}} f(\mathbf{x}) d\mathbf{x}$$

Esta función existe siempre y caracteriza totalmente la distribución del vector aleatorio \mathbf{x} . Presenta unas propiedades análogas a la función generatriz de momentos. y un resultado importante basado en ella es el siguiente teorema

Teorema 7.2 (Cramer -Wald). *La distribución de cualquier vector aleatorio p -dimensional \mathbf{x} está perfectamente determinada si se conoce la de todas las combinaciones lineales de sus componentes.*

8 Función de un vector aleatorio

En este apartado se estudiaran algunos métodos que permiten calcular la distribución de probabilidad de funciones de un vector aleatorio.

El método más general consiste en obtener directamente la función de distribución del nuevo vector aleatorio. Dado $\mathbf{y} = g(\mathbf{x})$, entonces

$$F_{\mathbf{y}}(\mathbf{y}) = P(\mathbf{y} \leq \mathbf{y}) = P[\mathbf{x} \in \mathbb{R}^p / g(\mathbf{x}) \leq \mathbf{y}]$$

Veamos algunos ejemplos sencillos en los que g es una función de \mathbb{R}^2 en \mathbb{R} , es decir Z es una variable aleatoria.

Ejemplo

Sea (X, Y) un vector aleatorio continuo cuya función de densidad vale:

$$f(x, y) = \begin{cases} 1 & \text{si } 0 < x < 1; 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Calcular la distribución de la variable $Z = \frac{Y}{X}$.

El conjunto soporte de Z es $S_Z = (0, \infty)$.

$$P(Z \leq z) = P\left(\frac{Y}{X} \leq z\right) = P(Y \leq zX)$$

Cuando $0 < z < 1$ la función de distribución vale:

$$F(z) = \int_0^1 \int_0^{zx} f(x, y) dx dy = \int_0^1 \int_0^{zx} dx dy = \frac{z}{2}$$

Cuando $z > 1$ la función de distribución vale:

$$F(z) = \int_0^{1/z} \int_0^{zx} dx dy + \int_{1/z}^1 \int_0^1 dx dy = \frac{1}{2z} + (1 - \frac{1}{z}) = 1 - \frac{1}{2z}$$

Por lo tanto la función de distribución es:

$$F(z) = \begin{cases} 0 & \text{si } z \leq 0 \\ \frac{z}{2} & \text{si } 0 < z \leq 1 \\ 1 - \frac{1}{2z} & \text{si } z > 1 \end{cases},$$

Cuando la función g verifique ciertas condiciones de regularidad es posible calcular directamente la función de densidad del vector aleatorio transformado.

Proposición .2. *Dado un vector aleatorio continuo \mathbf{x} con función de densidad $f(\mathbf{x})$, sea $\mathbf{z} = g(\mathbf{x})$ una transformación que verifica en el conjunto soporte S_x las siguientes condiciones :*

1. *La función $g: \mathbb{R}^p \rightarrow \mathbb{R}^p$ es biyectiva, salvo un conjunto de Lebesgue con medida nula, es decir existen las transformaciones*

$$z_i = g_i(\mathbf{x}), i = 1, \dots, p$$

$$x_i = h_i(\mathbf{z}), i = 1, \dots, p$$

2. *Todas las transformaciones son continuas.*

3. *Existen y son continuas las derivadas parciales*

$$\frac{\partial \mathbf{x}}{\partial \mathbf{z}}, \quad \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

$$J = \left| \frac{\partial(\mathbf{x})}{\partial(\mathbf{z})} \right|$$

es distinto de cero en el soporte de la transformación.

Entonces se cumple que

$$f_z(z) = \begin{cases} f_x[h(z)] \times |J| & \text{si } z \in S_z \\ 0 & \text{en el resto} \end{cases}$$

Corolario .3. Si el conjunto soporte S se puede expresar como una unión finita de intervalos disjuntos S_i $i = 1 \dots n$, en los que la función g verifica las condiciones del teorema anterior, entonces

$$f_z(z) = \begin{cases} \sum_{i=1}^n f_x[h_z \times |J_i|] & \text{si } z \in S_z \\ 0 & \text{en el resto} \end{cases}$$

Ejemplo 3.

Sea (X, Y) un vector aleatorio continuo cuya función de densidad vale:

$$f(x, y) = \begin{cases} 2x & \text{si } 0 < x < 1 ; 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Calcular la distribución de la variable $(U, V) = (X+Y, X-Y)$.

Las funciones g y h están definidas de la siguiente formas:

$$\begin{aligned} u &= g_1(x, y) = x + y & v &= g_2(x, y) = x - y \\ x &= h_1(u, v) = \frac{u+v}{2} & y &= h_2(u, v) = \frac{u-v}{2} \end{aligned}$$

El conjunto soporte de (U, V) se calcula teniendo en cuenta el soporte de (X, Y) y la transformación realizada.

Como $u=x+y$, es evidente que $0 < u < 2$; por otra parte se tiene:

$$\begin{aligned} 0 < x < 1 &\iff 0 < \frac{u+v}{2} < 1 \iff 0 < u+v < 2 \iff -u < v < 2-u \\ 0 < y < 1 &\iff 0 < \frac{u-v}{2} < 1 \iff -2 < -u+v < 0 \iff u-2 < v < 2-u \end{aligned}$$

es decir,

$$0 < u < 2 ; \quad \text{Max}(u-2, -u) < v < \min(u, 2-u)$$

Si $u \in (0, 1]$ $\implies \text{Max}(u-2, -u) = -u$; $\min(u, 2-u) = u$.

Si $u \in (1, 2)$ $\implies \text{Max}(u-2, -u) = u-2$; $\min(u, 2-u) = 2-u$.

Por lo tanto

$$S_{UV} = \{(u, v) \in \mathbb{R}^2 / 0 < u < 1 \quad -u < v < u \quad 1 < u < 2 \quad u-2 < v < 2-u\}$$

El jacobiano de la transformación inversa vale $J = \frac{d(x, y)}{d(u, v)} = -\frac{1}{2}$

y la función de densidad de (U, V) se calcula a partir de la expresión

$$f(u, v) = f\left[\frac{u+v}{2}, \frac{u-v}{2}\right] \times |J|$$

es decir,

$$f(u, v) = \begin{cases} \frac{u+v}{2} & \text{si } 0 < u \leq 1; -u < v < u \\ \frac{u+v}{2} & \text{si } 1 < u < 2; u-2 < v < 2-u \\ 0 & \text{en el resto} \end{cases}$$

El vector aleatorio (U, V) del ejemplo anterior con $U=X+Y$, $V=X-Y$, es un caso particular de las transformaciones lineales, que se pueden representar en forma matricial como

$$(U, V)^t = \mathbf{A} (X, Y)^t.$$

Las transformaciones lineales $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$ donde \mathbf{A} es una matriz no singular, es decir, $\det(\mathbf{A}) \neq 0$, son transformaciones biyectivas, su inversa es

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})^t$$

y el jacobiano de la transformación inversa vale $J = \det(\mathbf{A}^{-1})$.

Aplicando el teorema anterior la función de densidad de \mathbf{y} es:

$$f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{x}}[\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})^t] \times |\det(\mathbf{A}^{-1})|$$

8.1 Propiedades de las esperanzas condicionadas

En el caso bidimensional se considera la variable (X, Y) puesto que $E(X/Y = y)$ es una función de y se puede considerar la variable $h(Y) = E(X/Y)$ que tiene las siguientes propiedades interesantes

- $E(X) = E(E(X/Y))$
- $Var(X) = Var(E(X/Y)) + E(Var(X/Y))$

La $E(Y/X)$ también se llama la función de regresión de Y sobre X y sirve para aproximar el valor de Y cuando se conoce el valor de X por lo que el error cometido sería $\mathcal{E} = Y - E(Y/X)$

Los resultados anteriores se pueden generalizar al caso multidimensional.

Teorema 8.1. *Dado el vector aleatorio $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{x}_1 \in \mathbb{R}^k$, $\mathbf{x}_2 \in \mathbb{R}^{p-k}$, el vector $(p-k)$ -dimensional $\mathcal{E} = \mathbf{x}_2 - E(\mathbf{x}_2/\mathbf{x}_1)$ verifica que*

- $E(\mathcal{E}) = 0$
- $E(\mathbf{x}_2/\mathbf{x}_1)$ es la mejor aproximación de \mathbf{x}_2 por una función $h(\mathbf{x}_1)$ en el sentido de minimizar el error cuadrático medio (ECM) siendo

$$ECM(h) = E((\mathbf{x}_2 - h(\mathbf{x}_1))^t(\mathbf{x}_2 - h(\mathbf{x}_1)))$$

9 Formas cuadráticas

En el trabajo con vectores aleatorios una de las trasformaciones más sencillas, después de las transformaciones lineales, que pueden aparecer son las del tipo $\vec{x}' A \vec{x}$, con A una matriz simétrica. Este tipo de variables aleatorias unidimensionales reciben el nombre de **formas cuadráticas** sobre \vec{x}

Las formas cuadráticas juegan un papel muy importante en la inferencia estadística bajo hipótesis de normalidad, ya que son la base de los procedimientos inferenciales sobre la varianza de una variable, el ANOVA y la Regresión.

En general si se trabaja con una v.a. unidimensional \mathbf{x} y se tiene una m.a.s. de tamaño n de la misma, \vec{x} , ya se ha trabajado con la forma cuadrática $nS = \vec{x}' H \vec{x} / n$, donde H es la matriz de centrado $H_n = I_n - \vec{1}_n * \vec{1}_n' / n$, es decir, con la varianza muestral; estadístico que es muy importante para las inferencias sobre la varianza de la variable X.

Proposición .4. *Sea \vec{x} un vector aleatorio p -dimensional con vector de medias $\vec{\mu}$ y matriz de varianzas-covarianza Σ . La variable aleatoria $\vec{x}' A \vec{x}$ tiene por media:*

$$E(\vec{x}' A \vec{x}) = \text{traza}(A\Sigma) + \vec{\mu}' A \vec{\mu}$$

Corolario .5.

$$E((\vec{x} - \vec{b})' A (\vec{x} - \vec{b})) = \text{traza}(A\Sigma) + (\vec{\mu} - \vec{b})' A (\vec{\mu} - \vec{b})$$

Corolario .6. *Si \vec{x} un vector aleatorio n -dimensional proveniente del m.a.s de la variable unidimensional X con media μ y varianza σ^2 entonces*

$$E(nS) = E(\vec{x}' H \vec{x}) = \sigma^2 \text{traza}(H) + \mu \vec{1}_n' H \mu \vec{1}_n = (n-1)\sigma^2$$

10 Estadísticos importantes

Para realizar inferencias sobre el comportamiento de los vectores aleatorios es usual recoger datos y manipularlos en forma matricial, así al examinar los datos de n individuos se dispondrán los datos en una matriz X de n filas, una para cada individuo y p columnas (una para cada variable)

Es importante distinguir cuando se está hablado de un variable (column de la matriz) es decir los datos de una am.a.s de la variable $X = (x_{(1)}, \dots, x_{(p)})$ o cuando se está hablado de los valores del vector aleatorio para cada individuo (fila de la matriz) $X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$.

A partir de esta matriz se calcularán dos estadísticos que son fundamentales:

1. El vector de medias muestrales $\vec{x}^t = 1/n \vec{1}^t X$, que es una transformación lineal de la matriz de datos y permite obtener la estimación del vector de medias $\vec{\mu}$

2. La matriz de varianzas covarianzas muestrales $S = \frac{X^t H X}{n}$, que es una transformación matricial de la matriz de datos y permite obtener la estimación de la matriz de varianzas -covarianzas del v. a. Σ .

En su cálculo será muy útil la *matriz de centrado*, $H_n = I_n - 1/n \vec{1}_n \vec{1}_n^t = I_n - \vec{1}_n (\vec{1}_n^t \vec{1}_n)^{-1} \vec{1}_n^t$ que actuando sobre cualquier matriz de datos X , centra los datos por columnas, es decir a cada valor de la variable le resta la media muestral de esa variable. Esta matriz de centrado tiene dos propiedades importantes

- (a) H_n es idempotente de rango $n-1$.
- (b) $H \vec{1} = \vec{0}$

Distribución normal p-dimensional

11 Introducción

La distribución normal p-dimensional tiene en el análisis multivariante un papel similar al de la normal en el caso unidimensional, siendo la base de muchas técnicas estadísticas. Por todo ello merece un estudio detallado, junto con las propiedades y distribuciones asociadas a su muestreo

Entre sus ventajas se pueden destacar las siguientes:

1. Es una generalización sencilla del caso univariante
2. Tiene propiedades matemáticas que la hacen muy manejable
3. Depende de un número relativamente pequeño de parámetros: p para la media y $\frac{p(p+1)}{2}$ para la varianza
4. Independencia lineal es equivalente en este caso a independencia estadística
5. Es el límite de la suma de vectores aleatorios independiente y con la misma distribución (TCL multivariante)

Definición 11.1. *Un vector aleatorio p -dimensional se dice que tiene distribución normal p -dimensional, N_p si cualquier combinación lineal de sus componentes es una normal o una distribución degenerada*

Si una combinación lineal de sus componentes fuese degenerada querría decir que una componente se podría eliminar ya que es función lineal de otras.

Propiedad: Si \mathbf{x} tiene una distribución N_p entonces existen el vector de medias, μ , y la matriz de varianzas-covarianzas, Σ y caracterizan la distribución que se denotará $\mathbf{x} \equiv N_p(\mu, \Sigma)$

Propiedad: Si \mathbf{x} tiene una distribución $N_p(\mu, \Sigma)$ entonces su función generatriz de momentos es

$$g(t) = \exp\{t'\mu + \frac{1}{2}t'\Sigma t\}, t \in \mathbb{R}^p$$

Propiedad: Si \mathbf{x} tiene una distribución $N_p(\mu, \Sigma)$ cualquier transformación lineal $A\mathbf{x} + \mathbf{b}$, $A_{q,p}$, $\mathbf{b} \in \mathbb{R}^q$ tiene distribución normal, $N_q(A\mu + \mathbf{b}, A\Sigma A')$. En particular cualquier componente es normal y cualquier subconjunto de componentes es normal.

En general se trabajará con transformaciones lineales tales que $A\Sigma A'$ sea definida positiva

Propiedad: Dado $\mu \in \mathbb{R}^p$ y una matriz simétrica de dimensión p semidefinida positiva, $\Sigma \geq 0$ existe un vector aleatorio $\mathbf{x} \equiv N_p(\mu, \Sigma)$.

Propiedad: La función de densidad de $\mathbf{x} \equiv N_p(\mu, \Sigma)$ $\Sigma > 0$ es

$$f(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \mu)' \Sigma^{-1}(\mathbf{x} - \mu)\}}{|2\pi\Sigma|^{1/2}}$$

Los puntos de la variable que tienen igual función de densidad se sitúan sobre elipsoides llamados líneas o superficies de equiprobabilidad.

Propiedad: Sea $\mathbf{x} \equiv N_p(\mu, \Sigma)$ que partimos en dos subvectores $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$, lo que conlleva las particiones $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ y $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, \mathbf{x}_1 y \mathbf{x}_2 son independientes si y sólo si $\Sigma_{12} = 0$

Por tanto en el caso normal la independencia entre componentes es equivalente a que las variables tengan covarianza cero o sean incorreladas.

Teorema 11.1. Sea $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \equiv N_p(\mu, \Sigma)$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

Sea $\Sigma_{1.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Se verifica:

1. $\mathbf{x}_{1.2} = \mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$ es un vector aleatorio con distribución normal e independiente de \mathbf{x}_2 , $\mathbf{x}_{1.2} \equiv N_r(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{1.2})$.
2. La distribución condicionada de $\mathbf{x}_1/\mathbf{x}_2$ es $N_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{1.2})$, siendo por tanto la varianza la varianza independiente de la condición.

Propiedad: Si $\mathbf{x}_i, i = 1, \dots, n$ son una colección de vectores aleatorios p -dimensionales normales independientes entonces cualquier combinación lineal de ellos es un vector aleatorio normal p -dimensional.

En particular la media muestral de un v.a. $N_p(\mu, \Sigma)$, $\bar{\mathbf{x}}_n$ sigue una distribución $N_p(\mu, \Sigma/n)$, donde n es el tamaño de la muestra aleatoria simple.

Propiedad: Si $\mathbf{x}_i, i = 1, \dots, n, \dots$ son una colección de vectores aleatorios p-dimensionalas con la misma distribución de media μ y matriz de varianzas-covarianzas $\Sigma > 0$, e independientes, entonces $\sqrt{n}(\bar{\mathbf{x}}_n - \mu)$ converge asintóticamente en ley, cuando $n \rightarrow \infty$, a $N_p(0, \Sigma)$

Propiedad: Si \mathbf{t}_n es un estadístico p-dimensional verificando $\sqrt{n}(\mathbf{t}_n - \mu) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(0, \Sigma)$ y f es una función $f : \mathbb{R}^p \rightarrow \mathbb{R}^p$ diferenciable en μ y 'regular' entonces $\sqrt{n}(f(\mathbf{t}_n) - f(\mu)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(0, D\Sigma D')$ donde $D = (\frac{\partial f}{\partial t})|_{\mu}$

12 Formas cuadráticas

Las formas cuadráticas sobre distribuciones normales dan lugar, bajo determinadas condiciones a distribuciones de tipo χ^2 y χ^2 -descentralizada, así como las funciones de estas que dan lugar a las distribuciones t y F centradas y descentradas.

Definición 12.1. Sea $\vec{x} \equiv N_p(\vec{\mu}, I_p)$. La variable $z = \vec{x}' \vec{x}$ tiene una distribución, $\chi^2_{p,\delta}$, ji-cuadrado descentrada con p grados de libertad y parámetro de descentralización $\delta = \vec{\mu}' \vec{\mu}$, $\chi^2_{p,\delta}$.

Ya es conocido que si $\vec{x} \equiv N_p(\vec{0}, I_p)$ la variable $z = \vec{x}' \vec{x}$ tiene una distribución, $\chi^2_{p,\delta}$, ji-cuadrado con p grados de libertad

Definición 12.2. Sea $x \equiv N(\mu, 1)$ y $\equiv \chi^2_p$, independientes. La variable $z = \frac{x}{\sqrt{y/p}}$ tiene una distribución, $t_{p,\delta}$, t descentrada con p grados de libertad y parámetro de descentralización $\delta = \mu$.

Definición 12.3. Sea $x \equiv \chi^2_{p,\delta}$ y $y \equiv \chi^2_q$, independientes. La variable $z = \frac{x/p}{y/q}$ tiene una distribución, $F_{p,q,\delta}$, F descentrada con p y q grados de libertad y parámetro de descentralización δ .

En cuanto a formas cuadráticas sobre distribuciones normales se tiene que:

Teorema 12.1. Sea $\vec{x} \equiv N_p(\vec{\mu}, \Sigma)$, $\Sigma > 0$ se verifica que:

1. $(\vec{x} - \vec{\mu})' \Sigma^{-1} (\vec{x} - \vec{\mu}) \equiv \chi^2_p$
2. $\vec{x}' \Sigma^{-1} \vec{x} \equiv \chi^2_{p,\delta}$, $\delta = \vec{\mu}' \Sigma^{-1} \vec{\mu}$

Teorema 12.2. Sea $\vec{x} \equiv N_p(\vec{\mu}, I_p)$ A una matriz simétrica y semidefinida positiva, se verifica que:

$$\vec{x}' A \vec{x} \equiv \chi^2_{k, \vec{\mu}' A \vec{\mu}} \Leftrightarrow A \text{ es idempotente de rango } k.$$

Corolario 7. Sea $\vec{x} \equiv N_p(\vec{\mu}, \Sigma)$, $\Sigma > 0$ y A una matriz simétrica y semidefinida positiva, se verifica que: $\vec{x}' A \vec{x} \equiv \chi^2_{k, \vec{\mu}' A \vec{\mu}} \Leftrightarrow A\Sigma$ es idempotente de rango k .

Teorema 12.3. Sea $\vec{x} \equiv N_p(\vec{\mu}, I_p)$, $Q_1 = \vec{x}' A \vec{x} \equiv \chi^2_{r_1, \delta_1}$, $Q_2 = \vec{x}' B \vec{x} \equiv \chi^2_{r_2, \delta_2}$.

Q_1, Q_2 son independientes si y sólo si $AB=0$

(Existe un resultado más general (*Graybill*) que no exige que las formas cuadráticas tengan distribución χ^2)

Corolario .8. Sea $\vec{x} \equiv N_p(\vec{\mu}, \Sigma)$, $\Sigma > 0$, $Q_1 = \vec{x}' A \vec{x} \equiv \chi^2_{r_1, \delta_1}$, $Q_2 = \vec{x}' B \vec{x} \equiv \chi^2_{r_2, \delta_2}$.

Q_1, Q_2 son independientes si y sólo si $A\Sigma B = 0$

Lema (de Loynes): Sean A una matriz simétrica e idempotente, $B \geq 0$ una matriz simétrica. Si $I - A - B \geq 0$ entonces $AB = 0$

Lema : Sean A_1, \dots, A_k matrices simétricas con $\text{rango}(A_i) = r_i$, $A = A_1 + \dots + A_k$ se verifica:

A es idempotente con $\text{rango}(A) = r = \sum_{i=1}^k r_i \Leftrightarrow A_i$ es idempotente y $A_i A_j = 0$ $i \neq j$

Teorema 12.4 (Teorema (de Cochran):). Sea $\vec{x} \equiv N_p(\vec{\mu}, I_p)$, A_i matrices simétricas con $\text{rango}(A_i) = r_i$, $A = A_1 + \dots + A_k$. Son equivalentes

1. $\vec{x}' A_i \vec{x} \equiv \chi^2_{r_i, \vec{\mu}' A_i \vec{\mu}}$; $\vec{x}' A_i \vec{x}, \vec{x}' A_j \vec{x}$ son independientes
2. $\vec{x}' A \vec{x} \equiv \chi^2_{r, \vec{\mu}' A \vec{\mu}}$, $r = \sum_{i=1}^k r_i$

Lema : Sean A y B matrices simétricas idempotentes con $A-B \geq 0$, entonces:

1. $AB = BA = B$
2. $A-B$ es idempotente

Teorema 12.5. Sea $\vec{x} \equiv N_p(\vec{\mu}, I_p)$, $Q_1 = \vec{x}' A \vec{x} \equiv \chi^2_{r_1, \vec{\mu}' A \vec{\mu}}$, $Q_2 = \vec{x}' B \vec{x} \equiv \chi^2_{r_2, \vec{\mu}' B \vec{\mu}}$.

Si $A-B \geq 0$ entonces $Q_1 - Q_2 \equiv \chi^2_{r_1 - r_2, \vec{\mu}' A \vec{\mu}}$ y además $Q_1 - Q_2$ y Q_2 son independientes.