Distribución Normal Multivariada

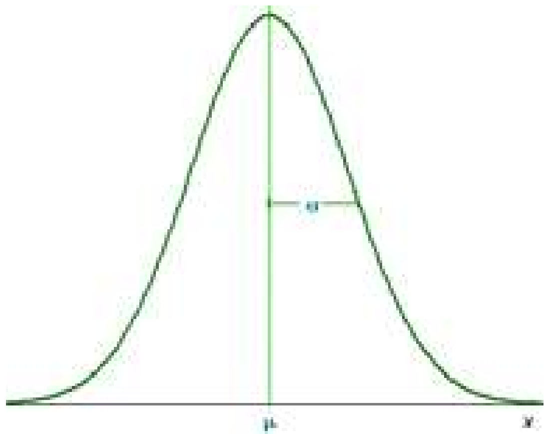
- La generalización de la densidad normal a varias dimensiones juega un rol fundamental en el análisis multivariado.
- La mayoría de las técnicas multivariadas asumen que los datos son generados de una distribución normal multivariada.
- En la realidad los datos casi nunca siguen exactamente una distribución normal multivariada, la densidad normal se utiliza frecuentemente como una aproximación a la verdadera distribución de los datos.
- Es muy atractiva por tres razones:

1. Es muy manipulable en un sentido matemático y tiene propiedades que producen resultados interesantes.
2. Resulta un modelo adecuado para la población bajo estudio, en algunas situaciones.
3. Las distribuciones muestrales de muchos estadísticos multivariados son aproximadamente normales, sin importar la distribución de la población raíz (teorema central del límite)

Revisión de la distribución normal Univariada

- La distribución normal univariada, con media μ y varianza σ^2, tiene la función de densidad de probabilidad

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



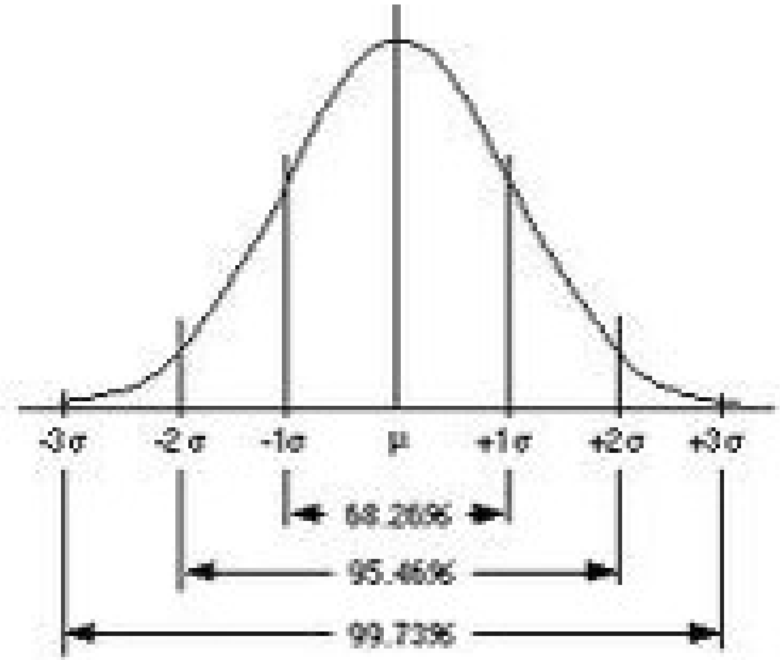
- La función de densidad normal univariada se denota por $N(\mu, \sigma^2)$

Características de la distribución normal

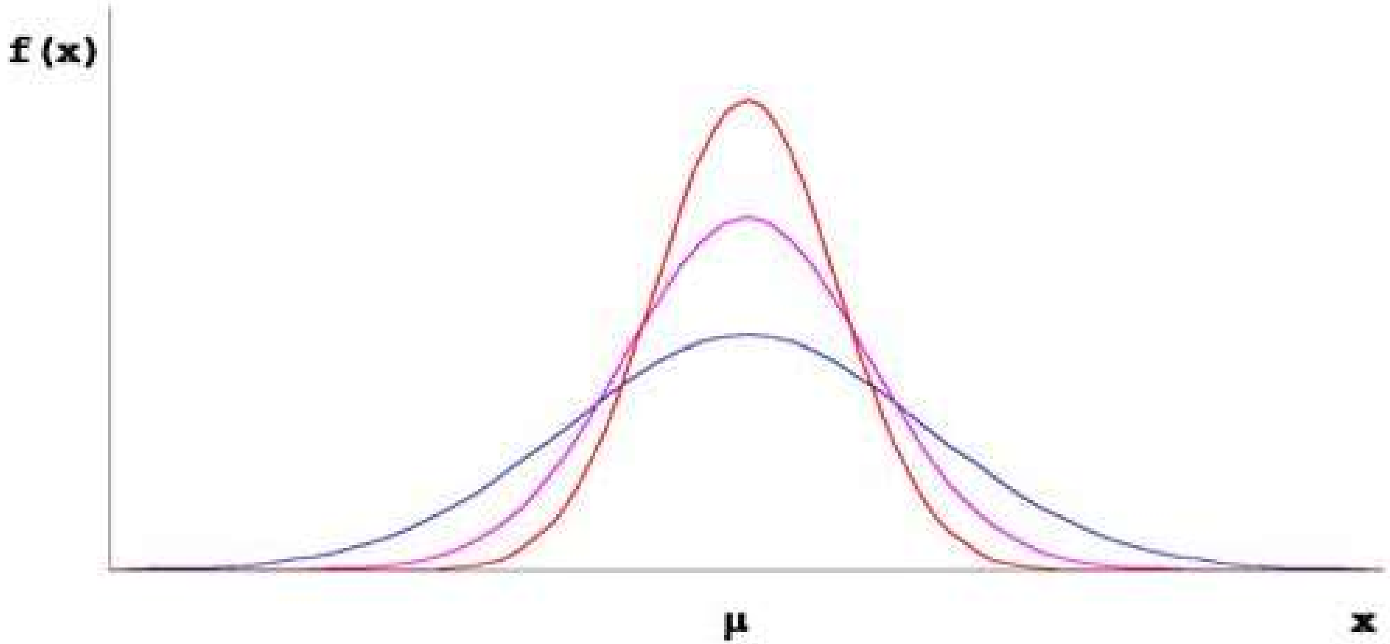
- Existe un número infinito de distribuciones normales, cada una definida por su combinación única de la media μ y la desviación estándar σ.
- μ determina la localización central y σ determina su anchura
- La distribución es simétrica alrededor de μ
- Es unimodal
- $\mu = M_d$
- Es asintótica con respecto al eje horizontal
- El área bajo la curva es 1.0
- Cumple la siguiente regla:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.99$$

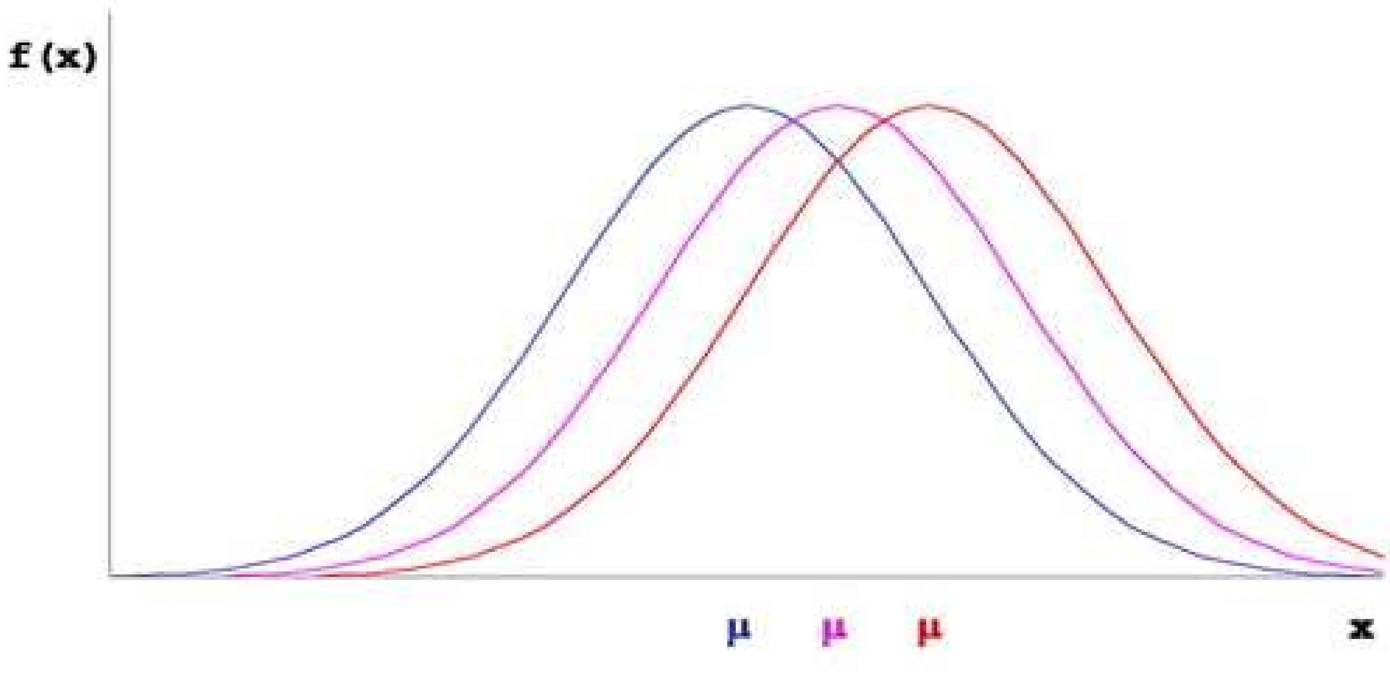
Regla Empirica



Distribuciones Normales con la misma media pero con diferentes desviaciones estándar:



Distribuciones normales con la misma desviación estándar pero diferentes medias:



Probabilidad en un intervalo

La probabilidad de que x tome valores en un intervalo [a,b] está dada por

$$\begin{aligned} P(a \leq x \leq b) &= \int_a^b f(x)dx \\ &= \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

- Estas integrales no se pueden calcular facilmente, de hecho no se pueden obtener directamente por metodos ordinarios, se usan metodos de integracion numerica para calcularlas y obtener la probabilidades
- ¿Debo integrar esta función cada vez que quiera calcular las probabilidades para una variable aleatoria normal?

Distribución normal estándar

- La variable normal x la podemos transformar a una variable que llamamos z y que se define como $z = (x - \mu)/\sigma$
- z **sigue una distribucion normal estandar** con media 0 y varianza 1, $x \sim N(0,1)$, es decir su funcion densidad esta dada por

$$f(z) = \frac{1}{2\pi} e^{-\frac{1}{2}z^2}$$

- Al transformar la distribucion normal x a una distribucion normal estandar z se tiene la ventaja de que podemos comparar todas la distribuciones normales que tienen diferentes valores de μ y σ con esta **distribucion normal común o estándar**.

Distribución normal estándar

- El problema de calcular la probabilidad de que la variable normal x tome valores en el intervalo $[a, b]$, es decir

$$P(a \leq x \leq b) = P\left(\frac{a-\mu}{\sigma} \leq z \leq \frac{b-\mu}{\sigma}\right)$$
$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} f(z)dz = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{2\pi} e^{-z^2/2} dz$$

- Estas integrales tampoco se pueden calcular por metodos ordinarios y es necesario utilizar métodos numericos
- Sin embargo existen tablas que se utilizan para obtener los valores de estas integrales para la variable normal estándar.

Tablas de distribución normal estándar

- La mayoría de las tablas trabajan a partir de la distribución de probabilidad acumulada de la normal estándar $\phi(z_0) = P(z \leq z_0)$
- Entonces

$$P(a \leq x \leq b) = P(\frac{a-\mu}{\sigma} \leq z \leq \frac{b-\mu}{\sigma}) = P(z_1 \leq z \leq z_2)$$

$$= P(z \leq z_2) - P(z \leq z_1) = \phi(z_2) - \phi(z_1) = \int_{-\infty}^{z_2} f(z)dz - \int_{-\infty}^{z_1} f(z)dz$$

- En cualquier libro existen tablas que dan los resultados de la integración

La distribución normal univariada tiene una forma generalizada en p dimensiones. La función de densidad normal p-dimensional es :

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} e^{\overbrace{-(\boldsymbol{x}-\mu)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu)/2}_{\text{Distancia cuadrada generalizada de x a } \mu}}$$

donde $-\infty \leq x_i \leq \infty$, $i = 1, \ldots, p$

La función de densidad normal p-dimensional se denota por $N_p(\mu, \boldsymbol{\Sigma})$ donde

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

Distribución Normal Multivariada como una generalizacion de la normal univariada

- El término del exponente en el caso univariado mide el cuadrado de la distancia de x a μ, medida en unidades de desviación estándar
- Esta distancia se puede generalizar para un vector p-dimensional, $\mathbf{x}$, de observaciones de varias variables mediante la expresión $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$, que representa la distancia cuadrada generalizada de $\mathbf{x}$ a μ.
- Para completar la generalización al caso multivariado es necesario reemplazar el término constante de la univariada por una constante mas general que haga que el volumen bajo la superficie de la función de densidad normal multivariada sea 1, para cualquier p.
- Se puede probar que la constante que generaliza la constante univariada está dada por el denominador de la función de normal multivariada

Distribución Normal Bivariada

La distribución normal multivariada mas simple es la distribución normal bivariada (2 dimensiones), su función de densidad está dada por :

$$f(\mathbf{x}) = \frac{1}{(2\pi)\,|\boldsymbol{\Sigma}|^{1/2}} e^{\overbrace{-(\boldsymbol{x}-\mu)'\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\mu)/2}_{\text{Distancia cuadrada generalizada de x a } \mu}}$$

donde $-\infty \leq x_i \leq \infty$, $i = 1, \ldots, 2$

Esta función de densidad normal bidimensional se denota por $N_2(\mu, \boldsymbol{\Sigma})$ donde

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Ejercicio. Evaluación de la densidad normal bivariada en términos de los parámetros individuales:

- $\mu_1 = E(x_1), \mu_2 = E(x_2), \sigma_{11} = var(x_1), \sigma_{22} = var(x_2), \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} = corr(x_1, x_2)$
- Primero encontramos la inversa de la matriz de covarianzas

$$\mathbf{\Sigma^{-1}} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix}$$

Ahora bien, de la definición de ρ_{12} se obtiene que

$$\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$$

entonces el determinante se puede reescribir como

$$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho_{12}^2)$$

Ejercicio. Continuación

Sustituyendo, podemos ahora escribir la distancia cuadrada generalizada como

$$A = (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$$

$$\begin{aligned}
A &= \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\
&= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \\
&= \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]
\end{aligned}$$

Ejercicio. Continuación

lo cual significa que podemos reescribir la función de densidad de probabilidad normal bivariada como:

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

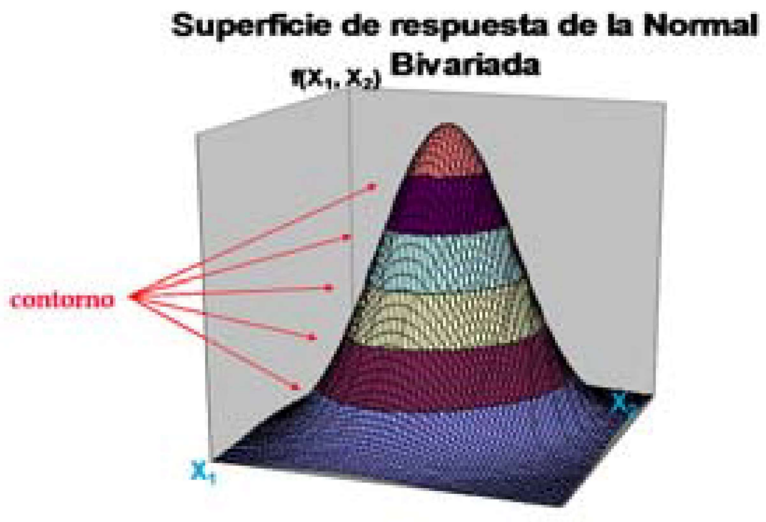
$$= \frac{1}{(2\pi)\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} e^{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]}$$

- Esta expresión no es muy manejable y su forma compacta general es mas informativa
- Sin embargo es muy útil para ilustrar ciertas propiedades de la distribución normal

Que sucede cuando las dos variables aleatorias x_1 y x_2 no están correlacionadas (i.e., $\rho_{12} = 0$):

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{1}{(2\pi)\,|\boldsymbol{\Sigma}|^{1/2}} e^{-(x-\mu)'\boldsymbol{\Sigma}^{-1}(x-\mu)/2} \\
&= \frac{1}{(2\pi)\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} e^{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]} \\
&= \frac{1}{(2\pi)\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} e^{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]} \\
&= \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-(x_1-\mu_1)^2/2\sigma_{11}}\right)}_{f(x_1)} \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma_{22}}} e^{-(x_2-\mu_2)^2/2\sigma_{22}}\right)}_{f(x_2)}
\end{aligned}
$$

Forma Gráfica de la Distribución Normal Bivariada

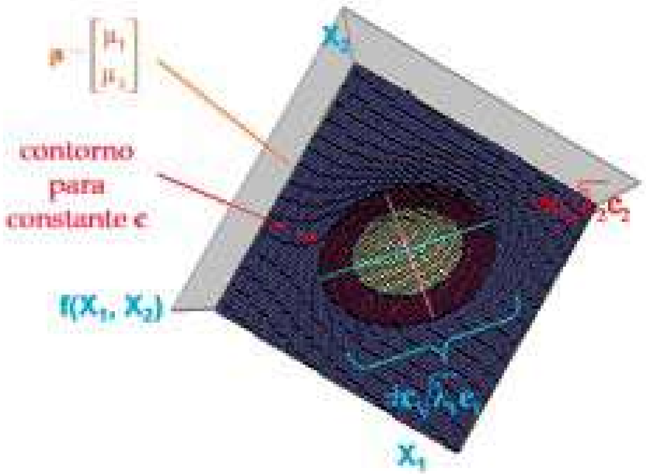


- Las probabilidades son representadas por el volumen bajo la superficie sobre las regiones definidas por los intervalos de los valores de las x_i
- Todos los puntos de igual densidad son llamados un **contorno**. Un *contorno* para p-dimensiones se define como todo **x** tal que

$$(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = \mathbf{c}^2$$

Contornos de distribución Normal Bivariada

- Los contornos: $(\boldsymbol{x}-\mu)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu)=\mathbf{c}^2$ forman elipsoides concéntricos centrados en μ
- Sus ejes están definidos por $\pm c\sqrt{\lambda_i}\boldsymbol{e}_i$, donde $\sum_i \boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i$ para $i = 1, \ldots, p$
- Los vectores propios determinan la dirección de los ejes y los valores propios su longitud.



Ejercicio: Contornos de una normal bivariada

Obtener la forma general de los ejes de los contornos para una densidad normal bivariada, asumiendo que las variables tienen igual varianza ($\sigma_{11} = \sigma_{22}$):
Primero obtenemos los valores propios de $\mathbf{\Sigma}$,

$$|\mathbf{\Sigma} - \lambda I| = 0$$

$$0 = \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} = (\sigma_{11} - \lambda)^2 - \sigma_{12}^2$$
$$= (\lambda - \sigma_{11} - \sigma_{12})(\lambda - \sigma_{11} + \sigma_{12})$$

Así $\lambda_1 = \sigma_{11} + \sigma_{12}$, $\lambda_2 = \sigma_{11} - \sigma_{12}$

El siguiente paso es obtener los vectores propios $\mathbf{e}_1$ y $\mathbf{e}_2$ correspondientes a λ_1 y λ_2
Para λ_1, $\boldsymbol{\Sigma e_1} = \lambda_1 \mathbf{e_1}$

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \lambda_1 \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Equivalente a

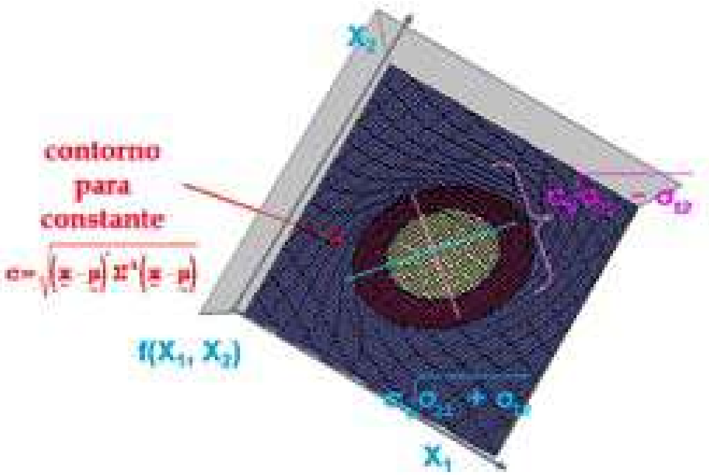
$$\begin{aligned} \sigma_{11} e_1 + \sigma_{12} e_2 &= (\sigma_{11} + \sigma_{12}) e_1 \\ \sigma_{12} e_1 + \sigma_{11} e_2 &= (\sigma_{11} + \sigma_{12}) e_2 \end{aligned}$$

Lo cual implica $e_1 = e_2$, por tanto $\mathbf{e}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

Similarmente para λ_2, se obtiene $\mathbf{e}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Ejercicio: Continuación

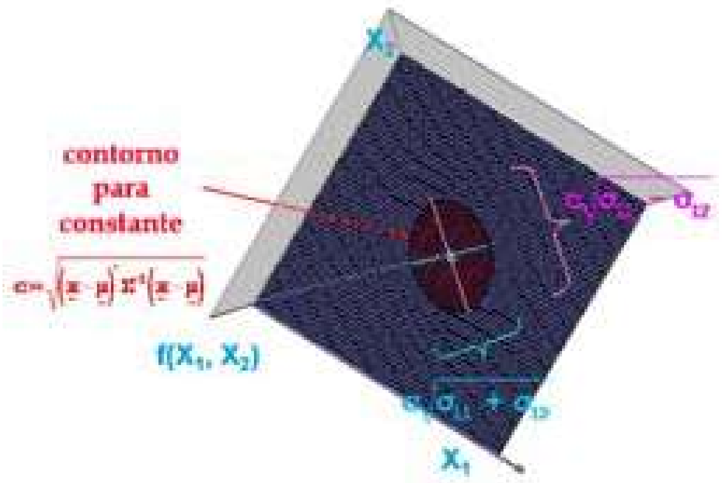
Para una covarianza positiva σ_{12}, λ_1 sera el mayor valor propio, entonces el eje mayor descansa a lo largo de la linea de 45^o que pasa a través del centroide μ:



¿Qué sucede cuando la covarianza es negativa?

Ejercicio: Continuación

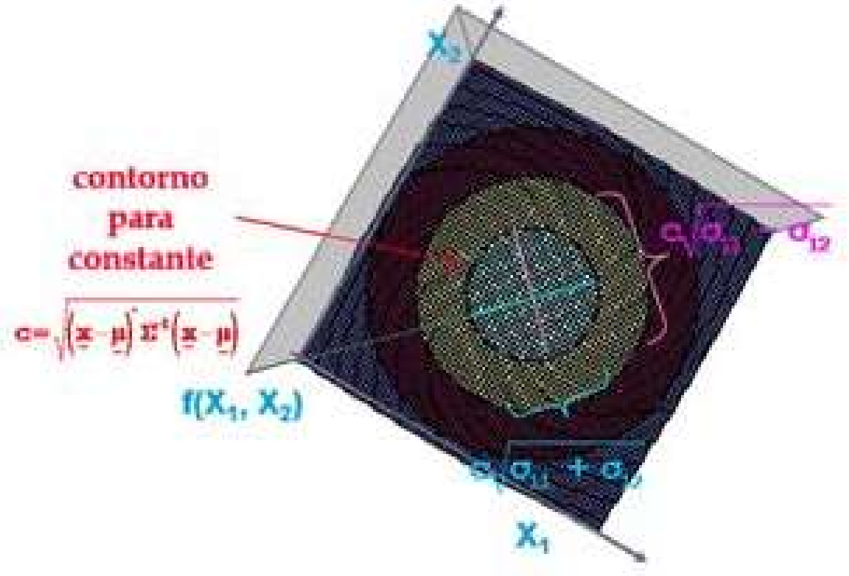
Para una covarianza negativa σ_{12}, λ_2 sera el mayor valor propio, entonces el eje mayor descansará a lo largo de una linea recta perpendicular a la linea de 45^o que pasa a través del centroide μ:



¿Qué sucede cuando la covarianza es cero?

Ejercicio: Continuación

Para una covarianza σ_{12} igual a cero, los dos valores propios y sus vectores propios son iguales (excepto por los signos) -uno corre a lo largo de la linea de 45^o que pasa a través del centroide μ y el otro es perpendicular:



Nota: Estos resultados son válidos únicamente para el caso en que las varianzas son iguales, es decir, $\sigma_{11} = \sigma_{22}$

Contornos de una normal bivariada

- Los contornos también tienen una interpretación probabilística. Se puede probar que $(\mathbf{x}-\mu)'\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\mu)$ se distribuye como una χ_p^2 con p grados de libertad.
- Entonces el elipsoide sólido de valores de $\mathbf{x}$ que satisface

$$(\mathbf{x}-\mu)'\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\mu) \leq \chi_p^2(\alpha)$$

tiene una probabilidad $1-\alpha$, donde $\chi_p^2(\alpha)$ denota el percentil superior $(100\alpha)\%$ de la distribución χ_p^2.

- Es decir,

$$Pr\left[(\mathbf{x}-\mu)'\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\mu) \leq \chi_p^2(\alpha)\right] = 1-\alpha,$$

que son contornos que contienen $(1-\alpha)$ x 100% de los valores de $\mathbf{x}$

Características y Propiedades de la Distribución Normal Multivariada

Para cualquier vector aleatorio normal multivariado $\mathbf{x}$

1. La densidad

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

tiene un valor máximo en

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

i.e., la media es igual a la moda

Características y Propiedades de la Distribución Normal Multivariada

2. La densidad

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

es simétrica a través de sus contornos constantes y está centrada en μ.

3. Si $\mathbf{x} \sim \mathbf{N_p}(\mu, \Sigma)$, entonces cualquier combinación lineal

$$\mathbf{a}'\mathbf{x} = \sum_{i=1}^{p} a_i x_i \sim N(\boldsymbol{a}'\mu, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$$

Además, si $\mathbf{a}'\mathbf{x} \sim N(\boldsymbol{a}'\mu, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$ para cada $\mathbf{a}$, entonces $\mathbf{x} \sim N_p(\mu, \boldsymbol{\Sigma})$

4. Si $\mathbf{A}$ es una matriz $q \times p$, $\mathbf{x} \sim N_p(\mu, \mathbf{\Sigma})$, entonces cualquier conjunto de q combinaciones lineales

$$\mathbf{Ax} = \begin{bmatrix} \sum_{i=1}^{p} a_{1i}x_i \\ \sum_{i=1}^{p} a_{2i}x_i \\ \vdots \\ \sum_{i=1}^{p} a_{qi}x_i \end{bmatrix} \sim N_q(\mathbf{A}\mu, \mathbf{A\Sigma A}^{'})$$

Además, si $\mathbf{d}$ es un vector conformado de constantes, entonces $\mathbf{x} + \mathbf{d} \sim N_p(\mu + \boldsymbol{d}, \mathbf{\Sigma})$

5. Si $\mathbf{x} \sim N_p(\mu, \mathbf{\Sigma})$, entonces todos los subconjuntos de $\mathbf{x}$ siguen una distribucion normal multivariada, i.e., para cualquier partición

$$\underset{(px1)}{\mathbf{X}} = \left(\begin{array}{c} \underset{(qx1)}{\mathbf{X}_1} \\ \hline \underset{((p-q)x1)}{\mathbf{X}_2} \end{array} \right), \underset{(px1)}{\mu} = \left(\begin{array}{c} \underset{(qx1)}{\mu_1} \\ \hline \underset{((p-q)x1)}{\mu_2} \end{array} \right), \underset{(pxp)}{\mathbf{\Sigma}} = \left(\begin{array}{c|c} \underset{(qxq)}{\mathbf{\Sigma}_{11}} & \underset{(qx(p-q))}{\mathbf{\Sigma}_{12}} \\ \hline \underset{((p-q)xq)}{\mathbf{\Sigma}_{21}} & \underset{((p-q)x(p-q))}{\mathbf{\Sigma}_{22}} \end{array} \right)$$

entonces $\mathbf{x_1} \sim N_q(\mu_1, \mathbf{\Sigma_{11}})$,$\mathbf{x_2} \sim N_{p-q}(\mu_2, \mathbf{\Sigma_{22}})$

Características y Propiedades de la Distribución Normal Multivariada

6. a) Si $\mathbf{x_1} \sim N_{q1}(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_{11}})$ y $\mathbf{x_2} \sim N_{q2}(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_{22}})$ son independientes, entonces $Cov(\mathbf{x_1}, \mathbf{x_2}) = \boldsymbol{\Sigma_{12}} = 0$

b) Si

$$\begin{pmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \boldsymbol{\mu_1} \\ \hline \boldsymbol{\mu}_2 \end{pmatrix}, \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right)\right)$$

entonces $\mathbf{x_1}$y $\mathbf{x_2}$ son independientes si y solo si $\boldsymbol{\Sigma}_{12} = 0$

c) Si $\mathbf{x_1} \sim N_{q1}(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_{11}})$ y $\mathbf{x_2} \sim N_{q2}(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_{22}})$ son independientes, entonces

$$\begin{pmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{pmatrix}, \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & 0 \\ \hline 0 & \boldsymbol{\Sigma}_{22} \end{array}\right)\right)$$

7.
$$\mathbf{X} = \left[\begin{array}{c}\mathbf{X}_1 \\ \hline \mathbf{X}_2\end{array}\right] \sim N_p(\mu, \boldsymbol{\Sigma}), \text{con } \mu = \left[\begin{array}{c}\mu_1 \\ \hline \mu_2\end{array}\right], \boldsymbol{\Sigma} = \left[\begin{array}{c|c}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{array}\right], |\boldsymbol{\Sigma}_{22}| > 0$$

Entonces la distribución condicional de $\mathbf{X}_1$ dado que $\mathbf{X}_2 = x_2$, $f(\mathbf{X}_1|\mathbf{X}_2)$, es normal multivariada con

$$Media = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$Covarianza = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

- Todas las distribuciones condicionales de los componentes son normales multivariadas
- La covarianza condicional no depende de los valores de las variables condicionantes

Características y Propiedades de la Distribución Normal Multivariada

8. Si $\mathbf{x} \sim N_p(\mu, \mathbf{\Sigma})$ y $|\Sigma| > 0$, entonces los elipsoides de densidad constante satisfacen:

1. $(\mathbf{x} - \mu)'\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{x} - \mu) \sim \chi^2_p$
2. La distribución $N_p(\mu, \mathbf{\Sigma})$ asigna probabilidad $1 - \alpha$ al elipsoide sólido definido por los valores de $\mathbf{x}$ tal que

$$\left\{\mathbf{x} : (\mathbf{x} - \mu)'\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{x} - \mu) \leq \chi^2_p(\alpha)\right\}$$

Donde $\chi^2_p(\alpha)$ denota el percentil superior $(100\alpha)\%$ de la distribución $\chi^2_p(\alpha)$.

Características y Propiedades de la Distribución Normal Multivariada

9. Si $\mathbf{x_j} \sim N_p(\boldsymbol{\mu_j}, \mathbf{\Sigma})$, $j = 1, \ldots, n$ mutuamente independientes. Entonces:

$$\mathbf{V}_1 = \sum_{j=1}^{n} c_j \mathbf{x}_j \sim N_p\left(\sum_{j=1}^{n} c_j \boldsymbol{\mu}_j, \left(\sum_{j=1}^{n} c_j^2\right)\mathbf{\Sigma}\right)$$

Además, si consideramos otro vector $\mathbf{V}_2$ tal que

$$\mathbf{V}_2 = \sum_{j=1}^{n} b_j \mathbf{x}_j \sim N_p\left(\sum_{j=1}^{n} b_j \boldsymbol{\mu}_j, \left(\sum_{j=1}^{n} b_j^2\right)\mathbf{\Sigma}\right)$$

entonces $\mathbf{V}_1$ y $\mathbf{V}_2$ son normales conjuntamente, con matriz de covarianza

$$\begin{bmatrix} \left(\sum_{j=1}^{n} c_j^2\right)\mathbf{\Sigma} & (b\,'c)\mathbf{\Sigma} \\ (b\,'c)\mathbf{\Sigma} & \left(\sum_{j=1}^{n} b_j^2\right)\mathbf{\Sigma} \end{bmatrix}$$

Así $\mathbf{V}_1$ y $\mathbf{V}_2$ son independientes si $b\,'c = 0$, Es decir, si b y c son vectores perpendiculares

Muestreo a partir de una Distribución Normal Multivariada y estimación de Máxima Verosimilitud

- Sea $\mathbf{X} = (\mathbf{x_1}, ..., \mathbf{x_n})$ una muestra aleatoria de n vectores obtenida de una distribución normal multivariada, es decir, $\mathbf{x_j} \sim N_p(\mu, \mathbf{\Sigma}), j = 1, \ldots, n$.
- Debido a que la muestra de observaciones es aleatoria, entonces las $\mathbf{x}_j$ son mutuamente independientes y su densidad conjunta es el producto de sus densidades marginales, i.e.,

$$\{\text{densidad de conjunta de } x_1, x_2, \ldots, x_n\}$$

$$= \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} e^{-(\mathbf{x_j}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x_j}-\mu)/2}$$

$$= \frac{1}{(2\pi)^{np/2} |\mathbf{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n}(\mathbf{x_j}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x_j}-\mu)/2}$$

Para observaciones fijas $\boldsymbol{x_j}, j = 1, \ldots, n$, la función depende únicamente de $\mu, \mathbf{\Sigma}$, se conoce como **función de verosimilitud**

Estimación de Máxima verosimilitud

- El objetivo es obtener los estimadores de los parametros $(\mu, \mathbf{\Sigma})$ de la distribución normal multivariada, dada la muestra recoletada $\mathbf{X}$.
- El enfoque de estimación de **máxima verosimilitud** consiste en encontrar los valores de los parámetros que **mejor** expliquen la muestra de datos recolectada, i.e., valores de los parámetros que produzcan la mayor probabilidad de obtener esa muestra.
- En otras palabras, estimar los valores de parámetros que maximicen la distribución conjunta evaluada en la muestra de observaciones (función de verosimilitud).
- Este enfoque de estimación se conoce estimación de **máxima verosimilitud**, y los valores de los parámetros encontrados son llamados **estimadores de máxima verosimilitud.**

Estimación de Máxima verosimilitud

- Para simplificar los cálculos, es necesario utilizar algunas propiedades de la traza para reescribir la función de verosimilitud de otra forma.
- **Algunas propiedades de la traza**:
  Para una matriz simétrica $\mathbf{A}$ $(k \times k)$ y un vector $\mathbf{x}$ $(k \times 1)$:
  1. $\mathbf{x}^{'}\mathbf{A}\mathbf{x} = tr(\mathbf{x}^{'}\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{x}\mathbf{x}^{'})$
  2. $tr(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$ donde $\lambda_i, i = 1, \ldots, k$ son los valores propios de $\mathbf{A}$

Entonces, primero, podemos reescribir

$$
\begin{aligned}
(\boldsymbol{x_j} - \mu)^{\prime}\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x_j} - \mu) &= tr\left[(\boldsymbol{x_j} - \mu)^{\prime}\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x_j} - \mu)\right] \\
&= tr\left[\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x_j} - \mu)(\boldsymbol{x_j} - \mu)^{\prime}\right]
\end{aligned}
$$

Entonces

$$\begin{aligned}\sum_{j=1}^{n}(\mathbf{x_j}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x_j}-\mu) &= \sum_{j=1}^{n} tr\left[(\mathbf{x_j}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x_j}-\mu)\right]\\ &= \sum_{j=1}^{n} tr\left[\mathbf{\Sigma^{-1}}(\mathbf{x_j}-\mu)(\mathbf{x_j}-\mu)'\right]\\ &= tr\left[\mathbf{\Sigma^{-1}}\left(\sum_{j=1}^{n}(\mathbf{x_j}-\mu)(\mathbf{x_j}-\mu)'\right)\right]\end{aligned}$$

Lo anterior debido a que la traza de la suma de las matrices es igual a la suma de sus trazas individuales.

Estimación de Máxima verosimilitud

La sumatoria de la expresión anterior también se puede reescribir como

$$\sum_{j=1}^{n}(\boldsymbol{x_j}-\mu)(\boldsymbol{x_j}-\mu)' = \sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)'$$
$$= \sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})' + \sum_{j=1}^{n}(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'$$
$$= \sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})' + n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'$$

La simplificación en el segundo término se debe a que los términos del producto cruzado

$$\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\mu)' \; y \; \sum_{j=1}^{n}(\bar{\mathbf{x}}-\mu)(\mathbf{x}_j-\bar{\mathbf{x}})'$$

son matrices de ceros.

Estimación de Máxima verosimilitud

La sustitución de estos dos resultados nos da una expresión alternativa para la densidad conjunta de una muestra aleatoria obtenida de una población p-dimensional:

$$f(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}) = \frac{1}{(2\pi)^{np/2} |\mathbf{\Sigma}|^{n/2}} e^{-tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \mu)(\bar{\mathbf{x}} - \mu)'\right)\right]/2}$$

$$= ((2\pi)^{-np/2} |\mathbf{\Sigma}|^{-n/2}) \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \mu)(\bar{\mathbf{x}} - \mu)'\right)\right]\right\}$$

Sustituyendo los valores observados $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ en la densidad conjunta se obtiene la función de verosimilitud para la muestra correspondiente $\mathbf{X}$, la cual es denotada frecuentemente como $\boldsymbol{L}(\mu, \mathbf{\Sigma})$.

Estimación de Máxima verosimilitud

Así, para los valores observados $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ que conforman la muestra aleatoria $\mathbf{X}$ elegidos de una población p-dimensional normalmente distribuida, la función de verosimilitud es

$$\boldsymbol{L}(\mu, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} \, |\boldsymbol{\Sigma}|^{n/2}} e^{-tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})' + n(\overline{\mathbf{x}} - \mu)(\overline{\mathbf{x}} - \mu)'\right)\right]/2}$$

Finalmente, notemos que el exponente de la función de verosimilitud se puede expresar de muchas maneras – una expresión alterna particularmente útil es:

$$
\begin{aligned}
& tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})' + n(\overline{\mathbf{x}}-\mu)(\overline{\mathbf{x}}-\mu)'\right)\right] \\
&= tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})'\right)\right] + n\left[tr\left(\Sigma^{-1}(\overline{\mathbf{x}}-\mu)(\overline{\mathbf{x}}-\mu)'\right)\right] \\
&= tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})'\right)\right] + n(\overline{\mathbf{x}}-\mu)'\Sigma^{-1}(\overline{\mathbf{x}}-\mu)
\end{aligned}
$$

Estimación de Máxima verosimilitud

Entonces sustituyendo esta expresión, se obtiene la función de verosimilitud

$$\boldsymbol{L}(\mu,\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}\,|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\left[tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})'\right)\right]+n(\overline{\mathbf{x}}-\mu)'\Sigma^{-1}(\overline{\mathbf{x}}-\mu)\right]}$$

De nuevo, teniendo en mente que el objetivo es obtener las estimaciones de μ y $\boldsymbol{\Sigma}$ que maximicen la función de verosimilitud $\boldsymbol{L}(\mu,\boldsymbol{\Sigma})$ para una muestra aleatoria dada $\mathbf{X}$.

El siguiente resultado será también útil en la derivación de los estimadores de máxima verosimilitud de μ y $\boldsymbol{\Sigma}$.

Resultado: Para una matriz $\mathbf{B}$ $(p \times p)$, simétrica y positiva definida y un escalar $b > 0$, se sigue que

$$\frac{1}{|\boldsymbol{\Sigma}|^b} e^{-tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{B})/2} \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-pb}$$

para toda $\boldsymbol{\Sigma}$ positiva definida de dimensión $p \times p$, donde la igualdad se sostiene únicamente para

$$\boldsymbol{\Sigma} = \frac{1}{2b}\mathrm{B}$$

Resultado

Para una muestra aleatoria $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ de una población normal con media μ y covarianza $\boldsymbol{\Sigma}$, los *estimadores de maxima verosimilitud* $\hat{\mu}$ y $\hat{\boldsymbol{\Sigma}}$ de μ y $\boldsymbol{\Sigma}$ son

$$\hat{\mu} = \overline{\mathbf{x}}, \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x_j} - \overline{\mathbf{x}})(\mathbf{x_j} - \overline{\mathbf{x}})' = \frac{n-1}{n} \mathbf{S}.$$

Así, dada la muestra $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$, sus valores observados

$$\overline{\mathbf{x}} \quad \text{y} \quad \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x_j} - \overline{\mathbf{x}})(\mathbf{x_j} - \overline{\mathbf{x}})'$$

son llamados los estimadores de máxima verosimilitud para μ y $\boldsymbol{\Sigma}$.

Estimación de Máxima verosimilitud

Se observa que el máximo de la verosimilitud se alcanza en

$$\boldsymbol{L}(\hat{\mu}, \hat{\boldsymbol{\Sigma}}) = \left(\frac{1}{(2\pi)^{np/2}} e^{-\frac{np}{2}} \right) \left(\frac{1}{\left|\hat{\boldsymbol{\Sigma}}\right|^{n/2}} \right)$$

y como

$$\left|\hat{\boldsymbol{\Sigma}}\right| = \left(\frac{n-1}{n} \right)^{p} |\boldsymbol{S}|$$

tenemos que

$$\boldsymbol{L}(\hat{\mu}, \hat{\boldsymbol{\Sigma}}) = \left(\frac{1}{(2\pi)^{np/2}} e^{-\frac{np}{2}} \right) \left(\frac{1}{\left(\left(\frac{n-1}{n} \right)^{p} |\mathbf{S}| \right)^{n/2}} \right) = \underbrace{\left(\left(\frac{n-1}{n} 2\pi e \right)^{-\frac{np}{2}} \right)}_{\textbf{constante}} \underbrace{|\mathbf{S}|^{-\frac{n}{2}}}_{\textit{Varianza generalizada}}$$

La varianza generalizada determina lo **puntiagudo** de la funcion de verosimilitud, es una medida de global de variabilidad