

Probabilidad Gaussiana multivariada

Diego Pedraza Barajas 2105940X
Abigail Sampedro Gutierrez 2105944A
Ana Berenice Garcia Hernandez 2105931K

Universidad Michoacana de San Nicolas de Hidalgo
Lic. En Actuaría y Ciencia de Datos



Probabilidad Gaussiana multivariada: Forma de cálculo, exposición de sus parámetros y el concepto de elipsidad relacionado con la correlación Pearson.

Administración de riesgos

2105940x@umich.mx

2105944a@umich.mx

2105931K@umich.mx

Introducción

La distribución normal univariada ($N(\mu, \sigma^2)$) es una distribución de probabilidad continua que se caracteriza principalmente por su forma de campana, esta se define usando principalmente dos parámetros:

1. μ (mu): La media o valor esperado de la distribución.
2. σ (sigma): La desviación típica, que mide la dispersión de la distribución.

La función de densidad de probabilidad (PDF) se da por:

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

Donde:

- \mathbf{X} : es la variable
- \mathbf{e} : es el número de Euler

La naturaleza de las distribuciones gaussianas se puede explicar por el teorema del límite central, el cual establece que la distribución de las medias muestrales se aproxima a una distribución normal a medida que la muestra aumenta ($n \geq 30$) incluso si la población está sesgada.

Aunque dentro de las distribuciones gaussianas hay un caso especial, la distribución gaussiana estándar, o distribución normal estándar, básicamente es:

1. La media (μ) es exactamente 0.
2. La desviación típica (σ) es exactamente 1.

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

En el análisis de riesgos, las variables aisladas son insuficientes, ya que existe una interconexión entre factores como precios de los activos, tasas de interés, tipos de cambio y precios de materias primas, etc., estos generan dependencias complejas, por lo que se vuelve imperativo evaluar de manera completa las variables implicadas.

Para comprender mejor esto, se debe conocer el concepto de vector aleatorio, el cual permite analizar un conjunto de variables aleatorias de manera simultánea, este vector puede representar rendimientos diarios de los activos en una cartera de inversión, lo cual resulta útil en un contexto financiero.

La distribución normal multivariada es la generalización de la distribución normal en un plano multidimensional, analizando la forma, orientación y densidad del conjunto de puntos en un

espacio de p dimensiones, captando la interrelación entre cada variable, cada conjunto de variables aleatorias multivariadas se caracteriza por un vector de medias y una matriz de covarianza, las cuales describen la tendencia central y la relación entre las variables. $X \sim N_p(\mu, \Sigma)$, donde p es el número de variables, μ es el vector de medias y Σ es la matriz de varianzas-covarianzas.

El vector de medias (μ) y la Matriz de Covarianzas (Σ)

El vector de medias (μ) se define como el vector de esperanzas matemáticas de cada componente del vector aleatorio X :

$$\mu = E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_p] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

(μ) Corresponde al vector de los rendimientos esperados para cada uno de los p activos que componen la cartera.

Por otro lado, la matriz de varianzas-covarianzas, Σ , define su forma, tamaño y orientación: $\Sigma = \Sigma^T$

Diagonal principal: $\sigma_{ii} = \sigma_i^2$ son las varianzas de cada variable aleatoria individual X_i , miden la dispersión de cada activo individualmente, sin considerarse los demás, es decir, la covarianza entre una variable y ella misma.

Fuera de la diagonal: σ_{ij} para $i \neq j$ son las covarianzas entre los pares de variables X_i y X_j . $Cov(X_i, X_j)$ mide la dirección de la relación lineal entre dos activos.

La matriz tiene dos propiedades matemáticas:

1. **Simetría:** esto significa que $\Sigma = \Sigma^T$ esto pasa porque la covarianza entre X_i y X_j es la misma que entre X_j y X_i ($\sigma_{ij} = \sigma_{ji}$).
2. **Semidefinición positiva:** para cualquier vector no nulo v , se cumple que $v^T \Sigma v \geq 0$ nos indica que la varianza de cualquier combinación lineal nunca puede ser negativa.

Si la matriz es definida positiva, entonces es invertible.

Covarianza y correlación

La covarianza es una medida cuantitativa en que la desviación de una variable X de su media está relacionada con la desviación de otra variable Y de su media, midiendo la variabilidad conjunta de dos variables aleatorias.

La covarianza se utiliza para medir la relación lineal entre dos variables aleatorias, pero no es una medida estandarizada y su magnitud depende de la escala de las variables, por lo que se vuelve un tanto complicada su interpretación.

Una covarianza positiva indica que los rendimientos de los dos activos tienden a moverse en la misma dirección, por lo que comparten un comportamiento similar, mientras que una covarianza negativa nos dice que tienden a moverse en direcciones contrarias, lo que implica que si una variable aumenta, la otra disminuye, finalmente, la correlación cero o nula se produce cuando no hay relación lineal entre las variables, expresándose como un valor cercano a 0, esto nos indica que no hay relación entre las variables.

La correlación se calcula utilizando un método conocido como Correlación producto-momento de Pearson o como Coeficiente de correlación, este coeficiente varía entre -1 y 1. Un valor cercano a 1 indica correlación positiva fuerte, mientras que un valor cercano a -1 indica correlación negativa fuerte, un valor cercano a 0 indica correlación nula entre las variables.

Finalmente, la covarianza indica el grado en que dos variables varían entre sí, mientras que la correlación determina la fuerza y dirección de esta relación.

Coeficiente de correlación de Pearson

Este coeficiente es denotado por ρ , se calcula como:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

El valor ρ_{ij} está acotado por -1 y 1, lo que nos da una medida estandarizada y de fácil interpretación.

Sabiendo todo esto, tenemos que a partir de la matriz de covarianzas Σ , se puede derivar la matriz de correlación R, la cual tiene unos en su diagonal principal y los coeficientes de correlación ρ_{ij} en los elementos fuera de la diagonal.

Fórmula de la Densidad Gaussiana Multivariada

La función de densidad de probabilidad (PDF) de la distribución normal multivariada asigna una densidad de probabilidad a cada punto x en el espacio p -dimensional.

Para un vector aleatorio $X \sim N_p(\mu, \Sigma)$ con una matriz invertible, su fórmula es:

$$f(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

El propósito del término de la normalización es asegurarse de que el volumen total bajo la hipersuperficie de densidad sea igual a 1. El determinante de la matriz de covarianzas o varianza generalizada, mide el volumen del hiperparalelepípedo formados por los vectores columna de la matriz. Un valor grande nos dice que hay una gran dispersión general de los datos, mientras que un determinante cercano a 0 nos dice que las variables son aleatoriamente colineales, lo que nos indica que los datos pueden chocar en un subespacio menor.

El exponente y distancia de Mahalanobis con fórmula $(x - \mu)^T \Sigma^{-1}(x - \mu)$, es quien nos define la forma de la campana gaussiana en más de una dimensión, este término se conoce formalmente como distancia de Mahalanobis al cuadrado, este es una medida de distancia estadística que generaliza el z-score (mide distancia en unidades de desviación estándar) a un entorno de múltiples variables, teniendo en cuenta la estructura de covarianza de los datos, midiendo la distancia de un punto x al centro de masa μ en unidades de desviación estándar correlacionadas.

Por lo tanto, la densidad de probabilidad es una función que disminuye exponencialmente a medida que aumenta la distancia de Mahalanobis al cuadrado desde el centro, por lo que todos los puntos que se encuentran a la misma distancia de Mahalanobis del centro tienen la misma densidad de probabilidad.

Para visualizar la función de densidad de más de dos dimensiones, se trazan sus curvas de nivel o contornos de hipodensidad, lo que se reflejaría como cortar la superficie en planos horizontales paralelos a la base, su fórmula es:

$$(x - \mu)^T \Sigma^{-1}(x - \mu) = c^2$$

La forma y orientación de estas elipses de contorno se determinan por la matriz de covarianzas y por el coeficiente de correlación de Pearson.

1. **Correlación 0** ($\rho = 0$) Cuando no hay correlación entre dos variables, la matriz de covarianzas es diagonal, por lo que los ejes de las elipses estarán alineados con los ejes de coordenadas x_1 y x_2 , mientras que su longitud va a depender de las varianzas individuales. Si las covarianzas son iguales, la matriz de covarianzas se convierte en un múltiplo de la matriz identidad, por lo que las elipses ahora serán círculos, por lo que la dispersión será isotropa e igual en todas las direcciones.

2. **Correlación positiva** ($0 < \rho < 1$) Las elipses se alargan y se inclinan con una pendiente positiva, habrá valores elevados de X_1 y de X_2 . En el límite teórico de $\rho = 1$, la distribución colapsaría sobre una línea recta y la matriz sería singular.
3. **Correlación negativa** ($-1 < \rho < 0$) Hace que las elipses se alarguen y se inclinen con pendiente negativa, asociado a la tendencia de que valores altos de una variable se asocien con valores bajos de la otra. Si ρ se acerca a -1 las elipses se vuelven más delgadas y alargadas en esa dirección.

Eigenvectores y Eigenvalores

Los eigenvectores de la matriz de covarianzas Σ definen las direcciones de los ejes principales de las elipses de isodensidad, representando las direcciones de máxima y mínima varianza en los datos. Este debe ser un vector no nulo que cuando se le aplica su transformación, solo cambia en su magnitud o sentido, pero no en su dirección.

Los eigenvalores son proporcionales al cuadrado de las longitudes de estos ejes principales. Un eigenvalor grande indica una gran dispersión a lo largo de la dirección de su eigenvector asociado, mientras que un eigenvalor pequeño indica poca dispersión. Este valor es el escalar asociado al eigenvector y representa el factor por el cual este es escalado o “estirado” cuando se le aplica la transformación lineal.

- Si $\lambda > 1$, el eigenvector se alarga.
- Si $0 < \lambda < 1$, el eigenvector se acorta.
- Si $\lambda < 0$, el eigenvector se invierte y se escala.
- Si $\lambda = 1$, el eigenvector no cambia de magnitud.
- Si $\lambda = 0$, el eigenvector se mapea al vector cero.

Heavy Tails

El supuesto de normalidad en modelos financieros tradicionales, son un supuesto poco probable, generalmente las distribuciones de rendimientos reales exhiben leptocurtosis (exceso de curtosis, lo que significa que los datos están muy concentrados alrededor de la media, dando paso a un pico más alto y estrecho en comparación a una distribución normal, asimismo tienen colas más largas, lo que aumenta la probabilidad de encontrar valores extremos o fuera de lo común, ya sean positivos o negativos en comparación a lo que se esperaría en una distribución normal de igual forma).

Este fenómeno de colas pesadas implica que los eventos extremos son mucho más probables de lo que el modelo gaussiano propone, por lo que algunos modelos como el VaR subestiman el tail risk, que es el riesgo de eventos catastróficos, que son los que más se deben cuantificar, un ejemplo de ellos fue la crisis financiera del 2008, la cual sucedió por confiar en modelos que no visualizaban adecuadamente la probabilidad de estos eventos extremos.

Alternativas

Se han desarrollado modelos que buscan calcular mejor las características de los datos financieros reales, una de las más populares es la distribución t-Student multivariada, la cual es una generalización de la normal, pero incluye un parámetro adicional: los grados de libertad (v), este parámetro controla el grosor de las colas de la distribución, cuando v disminuye, las colas se vuelven más pesadas, teniendo una mayor probabilidad a los eventos extremos, pero cuando v tiende a infinito, la distribución t-Student converge a una distribución normal, la cual la hace un caso límite.

Bisualizacion geometrica bivariada

```
library(mvtnorm)
library(ggplot2)
library(clusterGeneration)

mu <- c(x = runif(1, -5, 5), y = runif(1, -5, 5))

sigma_result <- genPositiveDefMat(dim = 2, covMethod = "unifcor", rangeVar = c(1, 2.5))
sigma <- sigma_result$Sigma

datos <- rmvnorm(n = 1500, mean = mu, sigma = sigma)
datos_df <- as.data.frame(datos)
colnames(datos_df) <- c("X", "Y")

ggplot(datos_df, aes(x = X, y = Y)) +
  geom_point(alpha = 0.5, color = "darkgray") +
  geom_density_2d(aes(color = ..level..), linewidth = 1) +
  scale_color_viridis_c(name = "Densidad") +
  labs(
    title = "Distribución Normal Multivariada Aleatoria",
    subtitle = paste0(
      "Centro en (", round(mu[1], 2), ", ", round(mu[2], 2), ")",
      " | Correlación: ", round(cov2cor(sigma)[1, 2], 2)
    ),
    x = "Variable X",
    y = "Variable Y"
  ) +
  theme_minimal() +
  coord_equal()
```

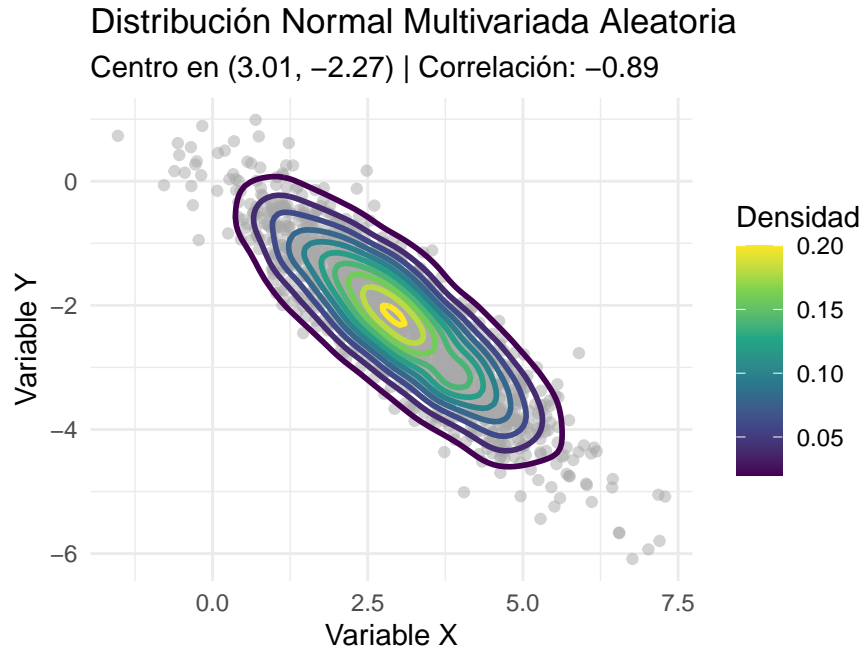


Figura 1: Distribución Normal Multivariada con Parámetros Aleatorios

Conclusión

La distribución gaussiana multivariada es básicamente una extensión de la curva de campana a varias dimensiones. En lugar de usar una media y una varianza, usamos un vector de medias (μ) y una matriz de covarianzas (Σ).

El vector μ : nos dice dónde está el centro o pico de la distribución en ese espacio multidimensional.

La matriz Σ define la forma, el tamaño y la orientación de la distribución. Aquí es donde se codifica cómo se relacionan las diferentes dimensiones o variables.

Hay un vínculo directo y esencial entre la matriz Σ y la geometría de la distribución. La estructura de dependencia que Σ captura, especialmente la correlación, se refleja en las formas de los elipsoides de isodensidad (que son como los contornos tridimensionales de la distribución).

- Si las variables no están correlacionadas, estos elipsoides se alinean perfectamente con los ejes.
- Si hay correlación (positiva o negativa), la distribución se “inclina” o rota en el espacio en función a su signo.

Lo que realmente une la fórmula matemática de la probabilidad con esta forma geométrica es la distancia de Mahalanobis. Esta medida asegura que cualquier punto que se encuentre en el mismo elipsoide de contorno (es decir, a la misma distancia de Mahalanobis del centro μ) tendrá la misma densidad de probabilidad.

A pesar de las limitaciones, especialmente prediciendo las heavy tails de las distribuciones, la distribución normal multivariada sigue siendo una parte fundamental en la teoría financiera y para la gestión de riesgos; sin embargo, en la actualidad ya solo se considera como un punto de partida o como modelo de referencia, con el cual se comparan algunos modelos más sofisticados, el analista ya entra en acción al momento de comprender y analizar los supuestos e identificar las condiciones del mercado en las que esos supuestos puedan fallar, actualmente la distribución t-Student es una herramienta mucho más completa para la estimación de estos parámetros y predicciones, por lo que también es de las más usadas en el presente.

Referencias

National Library of Medicine. (2023). *Explainable AI models for healthcare applications*. PubMed Central. Recuperado el 1 de octubre de 2025, de [LINK](#)

Stanford University. (s. f.). *Multivariate Gaussians*. Stanford CS229. Recuperado el 1 de octubre de 2025, de [LINK](#)

Peng, H. (s. f.). *Lecture note 3: Multivariate normal distribution*. Hong Kong Baptist University. Recuperado el 1 de octubre de 2025, de [LINK](#)

Ayala, J. (2018). *Distribución Normal Multivariada*. RPubS. Recuperado el 1 de octubre de 2025, de [LINK](#)

Scribd. (s. f.). *Gaussiana multivariable*. Recuperado el 1 de octubre de 2025, de [LINK](#)

Baíllo, A. (s. f.). *Normal multivariante*. Universidad Autónoma de Madrid. Recuperado el 1 de octubre de 2025, de [LINK](#)

FasterCapital. (s. f.). *Covarianza: comprensión de la covarianza a través del coeficiente de Pearson*. Recuperado el 1 de octubre de 2025, de [LINK](#)

Denapo, P. (2021). *Clase 13: Normal multivariada*. Universidad de Buenos Aires. Recuperado el 1 de octubre de 2025, de [LINK](#)

DataCamp. (s. f.). *Distribución gaussiana*. DataCamp Tutorial. Recuperado el 1 de octubre de 2025, de [LINK](#)

Ayala, J. (2018). *Índice de Masa Corporal con distribución normal*. RPubS. Recuperado el 1 de octubre de 2025, de [LINK](#)

Wikipedia. (s. f.). *Distribución normal multivariada*. En *Wikipedia*. Recuperado el 1 de octubre de 2025, de [LINK](#)

Wikipedia. (s. f.). *Multivariate normal distribution*. En *Wikipedia*. Recuperado el 1 de octubre de 2025, de [LINK](#)