

Tarea 1

Muestreo de Variables Aleatorias

Integrantes: Catalina Zapata
Constanza Santori
Karina Hernández
Teresa Peralta
Profesor: Francisco Ortega Culaciati
Fecha de entrega: 25 de abril de 2025
Santiago de Chile

1. Introducción

La teoría de problemas inversos, que se expone en el trabajo de Tarantola (2005) [1], se sustenta en una concepción probabilística del conocimiento científico. En este enfoque, un problema inverso no se resuelve obteniendo una única solución determinista, sino caracterizando la incertidumbre asociada a las posibles soluciones mediante distribuciones de probabilidad.

Todo conocimiento acerca de los parámetros de un modelo m se representa a través de distribuciones de probabilidad que evolucionan desde un estado previo (a priori) hacia uno posterior (a posteriori), condicionado por los datos observados d .

Esta relación está regida por la **regla de Bayes**, cuya formulación general, tal como se presenta es:

$$\sigma(m) = \frac{\rho(d|m) \mu(m)}{\int \rho(d|m') \mu(m') dm'}$$

donde:

- $\mu(m)$: distribución de probabilidad a priori sobre el espacio de modelos,
- $\rho(d|m)$: verosimilitud del dato dado el modelo,
- $\sigma(m)$: distribución de probabilidad a posteriori,
- $\int \rho(d|m') \mu(m') dm'$: constante de normalización.

Este marco general permite incorporar el conocimiento previo sobre los modelos y la información contenida en los datos, y es fundamental para resolver problemas inversos donde las soluciones no son únicas ni exactas. En contextos donde la resolución analítica de estas distribuciones no es factible, se recurre al muestreo estadístico. La capacidad de generar muestras aleatorias a partir de una distribución de probabilidad dada es esencial para resolver problemas inversos por métodos Monte Carlo. Estos métodos permiten estimar estadísticas de interés, evaluar funciones objetivo, y explorar regiones del espacio de modelos con alta verosimilitud.

Además Tarantola [1] destaca que, los métodos de muestreo deben implementarse cuidadosamente, considerando la estructura de dependencia de las variables y utilizando transformaciones adecuadas para pasar de distribuciones estándar a distribuciones arbitrarias. Los tres métodos de muestreo considerados en esta tarea —el método análogo, el método inverso y el método relacional— responden a distintas necesidades según la naturaleza discreta o continua de las variables, y la forma funcional de la distribución.

En este trabajo se implementan códigos desarrollados en Python que permiten evaluar funciones de probabilidad, verosimilitud, y generar muestras de variables aleatorias que sigan distribuciones de probabilidad clásicas.

2. Metodología

Siguiendo lo visto en clases (Apuntes del curso de Métodos Inversos Avanzados [2]), se completan las funciones en el paquete **GF7013** y ejemplos que permitan verificar el buen funcionamiento de los códigos programados.

2.1. Distribución Multinomial (P1)

Se pide abordar la implementación de algoritmos para generar muestras de una variable aleatoria discreta siguiendo una distribución multinomial, utilizando tanto el método proporcionalizado por la librería *Numpy* como un *Método Análogo*, en el script `pmf_multinomial.py`.

El método análogo se basa en transformar un valor aleatorio uniforme en un valor discreto siguiendo una distribución de probabilidad. Para ello, primero se construye una función de distribución acumulada (CDF por sus siglas en inglés) sumando las probabilidades individuales de cada estado. Así, esta última define efectivamente una partición del intervalo $[0,1]$ en subintervalos cuyas longitudes son proporcionales a las probabilidades de cada estado discreto posible.

Luego, para generar una muestra, se configura un número aleatorio uniforme $u \sim \mathcal{U}([0, 1])$ y se determina en qué subintervalo de la CDF cae este valor. Esto se logra utilizando una búsqueda binaria (implementada mediante la función `np.searchsorted`) que identifica el índice del primer valor en la CDF que es mayor o igual que u . De manera que el estado correspondiente al índice es, entonces, seleccionado como el de la variable aleatoria multinomial.

Con esto, para generar múltiples muestras, simplemente se repite este proceso el número deseado de veces, notando que en cada paso, el algoritmo reduce a la mitad el número de elementos que necesita examinar. Así, si se tienen N estados posibles en nuestra distribución multinomial, el número máximo de comparaciones necesarias es aproximadamente $\log_2(N)$.

Posteriormente, en el script **P1.py** se implementa un análisis comparativo entre los métodos de *Numpy* y *analog* para la generación de muestras multinomiales. Para ello, se crea la función `freq`, la que encapsula el proceso recibiendo como parámetros el método de muestreo a utilizar, los valores discretos posibles, sus importancias (probabilidades no normalizadas), el generador de números aleatorios y la cantidad de muestras a generar. Además, se crea una instancia de la clase `pmf_multinomial` para generar las muestras solicitadas, se cuenta la frecuencia de cada valor discreto y se normaliza los conteos para obtener frecuencias relativas comparables con las probabilidades teóricas.

Finalmente, visualizando comparativamente mediante gráficos de barras que muestran las probabilidades teóricas junto a las frecuencias obtenidas con cada método, se verifica que ambos métodos convergen a la distribución teórica.

2.2. Distribución Uniforme en N dimensiones (P2)

Distribución uniforme en una dimensión (1D)

Si se considera X una variable aleatoria continua que sigue una distribución uniforme en el intervalo $]a, b]$, es decir, $X \sim \mathcal{U}(a, b)$. Entonces, su función de densidad de probabilidad (fdp) asociada esta dada por:

$$f(x) = \frac{1}{(b-a)} \text{ para } x \in]a, b], \text{ y } 0 \text{ en otro caso.}$$

De este modo, la constante de normalización es $1/(b-a)$, con lo que se asegura que la integral de la fdp en todo el dominio sea igual a 1.

Distribución uniforme en n dimensiones (espacio \mathbb{R}^n)

Ahora, si se consideran n variables aleatorias continuas independientes X_1, X_2, \dots, X_n , donde cada $X_i \sim \mathcal{U}(a_i, b_i)$, es decir, cada componente está uniformemente distribuido en su respectivo intervalo $[a_i, b_i]$. Entonces, se tendrá que la fdp conjunta, está dada por el producto de las densidades marginales, debido a la independencia de las variables:

$$f(x_1, x_2, \dots, x_n) = f(x_1) \times f(x_2) \times \dots \times f(x_n) = \frac{1}{[(b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n)]}$$

Para cada $x_i \in [a_i, b_i]$, y 0 si algún x_i está fuera de su respectivo intervalo, para todo $i = 1, \dots, n$. Así, se construye la constante de normalización para la fdp conjunta que sería $1/[(b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n)]$, que puede entenderse como $1/V$, donde V es el volumen n -dimensional sobre el dominio definido por $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$ donde se asegura que la integral de la fdp sea igual a 1.

Muestreo en \mathbb{R}^n con Método Inverso

En la función `_draw` del módulo `pdf_uniform_nD.py` se utiliza el método del inverso para generar muestras aleatorias desde una distribución uniforme definida en \mathbb{R}^n . Así, sea $x_i \sim \mathcal{U}(a_i, b_i)$, la función de densidad de probabilidad (fdp) está dada por la ecuación antes mencionada. En cuanto a la función de distribución acumulada, esta se define como:

$$F(x_i) = \begin{cases} 0 & \text{si } x_i < a_i \\ \frac{x_i - a_i}{b_i - a_i} & \text{si } x_i \in [a_i, b_i] \\ 1 & \text{si } x_i > b_i \end{cases}$$

El Método Inverso, consiste en generar una variable auxiliar $u_i \sim \mathcal{U}(0, 1)$ y aplicar la inversa de la $F(x_i)$ para obtener una transformación que garantiza que las muestras generadas siempre tengan verosimilitud positiva de x_i :

$$x_i = a_i + (b_i - a_i) \cdot u_i$$

Dado que las variables x_i son independientes entre sí, la generación de muestras multivariadas consiste simplemente en aplicar esta transformación para cada componente del vector.

Además, es importante hacer notar que en el script `pdf_uniform_nD.py`, se implementa calculando directamente el logaritmo de la normalización como la suma negativa de los logaritmos de las diferencias entre ciertos límites, y en la función `_log_pdf` que suma el logaritmo de la verosimilitud con el logaritmo de normalización, así, evitan multiplicaciones que podrían causar overflow numérico con distribuciones de múltiples dimensiones.

Funciones de Densidad de Probabilidad Conjunta y Marginal

Sea un conjunto de variables aleatorias $d = [d_1, d_2, \dots, d_N]^\top$. La función de densidad de probabilidad conjunta (FDP conjunta) $p(d)$ describe la probabilidad de que estas variables tomen simultáneamente valores dentro de un cierto rango:

$$p(d) = p(d_1, d_2, \dots, d_N)$$

La FDP conjunta permite analizar relaciones, correlaciones y dependencias entre variables aleatorias. Si las variables son independientes, se cumple:

$$p(d_1, d_2) = p(d_1) \cdot p(d_2)$$

La FDP marginal de una variable aleatoria d_i , denotada como $p(d_i)$, se obtiene integrando la FDP conjunta respecto a las demás variables. Esta operación, conocida como *marginalización*, permite obtener el comportamiento individual de una variable sin considerar el de las otras. Por ejemplo, si $p(d_1, d_2)$ es la FDP conjunta de dos variables, se tiene:

$$p(d_1) = \int p(d_1, d_2) dd_2, \quad p(d_2) = \int p(d_1, d_2) dd_1$$

Así, el código implementa la distribución uniforme multidimensional aplicando la teoría con independencia entre variables, donde la función de verosimilitud verifica si un punto está dentro del dominio y el muestreo utiliza el método inverso transformando variables uniformes $\mathcal{U}(0, 1)$ mediante $x_i = a_i + (b_i - a_i) \cdot u_i$ para cada dimensión.

2.3. Distribución Normal Multivariada (P3)

Distribución normal en una dimensión (1D)

Una variable aleatoria x que sigue una distribución normal $x \sim \mathcal{N}(\mu, \sigma^2)$ tiene una función

de densidad de probabilidad (fdp) dada por:

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

donde en esta expresión, el término $\frac{1}{\sqrt{2\pi\sigma^2}}$ corresponde a la constante de normalización, la cual garantiza que la integral de la fdp sea igual a 1.

Distribución normal multivariada

Para el caso de una variable aleatoria $x \in \mathbb{R}^n$ que sigue una distribución normal multivariada $x \sim \mathcal{N}(\boldsymbol{\mu}, C)$, su fdp conjunta está dada por:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |C|}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu})^\top C^{-1}(x - \boldsymbol{\mu})\right)$$

donde $\boldsymbol{\mu}$ es el vector de medias, C la matriz de covarianza y n la dimensión de la variable aleatoria. Cuya constante de normalización es $\frac{1}{\sqrt{(2\pi)^n |C|}}$.

Método relacional

Si se conoce que una variable aleatoria X tiene una relación funcional con otra variable aleatoria Y , se pueden obtener muestras de X a través de esta relación. Considerando una distribución normal estándar, supongamos:

$$Y \sim \mathcal{N}(0, 1)$$

Entonces se puede generar una muestra de X usando la función inversa acumulada F^{-1} (método inverso). Alternativamente, es posible generar muestras de $Y \sim \mathcal{N}(0, 1)$ utilizando el código *NumPy*. Luego, existe una función f tal que $X = f(Y)$. Generando Q muestras de Y , se obtienen Q muestras de X evaluando $X_i = f(Y_i)$.

Sea $X \sim \mathcal{N}(\boldsymbol{\mu}, C)$, con $\boldsymbol{\mu} \in \mathbb{R}^n$ y $C \in \mathbb{R}^{n \times n}$, podemos generar $Y \sim \mathcal{N}(0, I)$. A continuación, se tiene una matriz W tal que $C = WW^\top$ y la relación $X = \boldsymbol{\mu} + WY$ permite generar muestras de X a partir de muestras estándar de Y .

Para determinar W , se utiliza la fórmula de propagación del error. Si,

$$X = MY + V, \quad Y \sim \mathcal{N}(\boldsymbol{\mu}_Y, C_Y)$$

entonces $C_X = MC_YM^\top$; como $C_Y = I$, resulta que $C = C_X = WW^\top$, por lo que W corresponde a la descomposición de Cholesky de la matriz de covarianza C .

Este procedimiento es la base para el desarrollo de las tareas propuestas. En este trabajo

se requiere evaluar el comportamiento del muestreo de una variable aleatoria que sigue una distribución normal multivariada en \mathbb{R}^2 , repitiendo este experimento con diferentes cantidades de muestras. Este análisis tiene como objetivo comparar visual y cuantitativamente la convergencia de las muestras generadas hacia la distribución teórica esperada, tanto en su densidad conjunta como en las densidades marginales de cada componente. Para cada tamaño de muestra, se generaron realizaciones a partir de la distribución $\mathcal{N}(\boldsymbol{\mu}, C)$, donde:

$$\boldsymbol{\mu} = \begin{bmatrix} 0.5 \\ 3.0 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

Las muestras fueron visualizadas mediante un histograma bidimensional (2D), que aproxima la función de densidad de probabilidad conjunta, y mediante histogramas unidimensionales marginales para cada componente, sobre los cuales se superpusieron las curvas teóricas de densidad correspondientes.

Se comenzó trabajando en el *script* `pdf_normal.py`, durante la inicialización, se extraen y validan los parámetros de entrada, que incluyen el vector de medias y la matriz de covarianza, asegurando que esta última sea simétrica, definida positiva y no singular. Se calcula también la matriz inversa de la covarianza, necesaria para la evaluación de la verosimilitud, así como su descomposición de Cholesky, utilizada en el proceso de generación de muestras. En este *script* se trabaja con la función `__draw`, para generar muestras aleatorias a partir de la distribución, utilizando el método relacional. Se contemplaron dos modos de operación: generación de una única muestra o de múltiples muestras, organizadas en forma matricial.

A partir de esta clase, se trabajó en el *script* `P3.py`, donde se generaron conjuntos de muestras con diferentes tamaños (1E4, 1E5 y 1E6) para observar el efecto del tamaño muestral en la aproximación a la distribución teórica. Se utilizó una función auxiliar para representar gráficamente las distribuciones obtenidas. Esta función permite generar una figura compuesta por subgráficos que incluyen un histograma 2D de las muestras (como aproximación a la densidad conjunta), así como histogramas marginales individuales con sus respectivas curvas teóricas superpuestas, facilitando la evaluación visual de la calidad del muestreo.

Además, se implementó la función `test_draw()`, la cual permite validar cuantitativamente el correcto funcionamiento del muestreo. Esta función genera un conjunto de muestras utilizando una semilla fija, calcula la media y matriz de covarianza empíricas, y las compara con los parámetros originales del modelo. Si ambas comparaciones se encuentran dentro de tolerancias predefinidas, el *test* devuelve True (y False si no). Esto proporciona una verificación robusta de que las muestras generadas convergen adecuadamente a los parámetros estadísticos esperados de la distribución multivariada.

3. Resultados y Discusión

3.1. Distribución Multinomial (P1)

En este trabajo, se comparan dos métodos de generación de muestras para una variable aleatoria discreta con distribución multinomial: el método basado en la función `rng.multinomial()` de *NumPy*, y el *Método Análogo*. Ambos métodos fueron evaluados frente a la distribución teórica de probabilidad. Como resultado de la comparación, se obtuvo un histograma con tres barras por cada uno de los estados posibles.

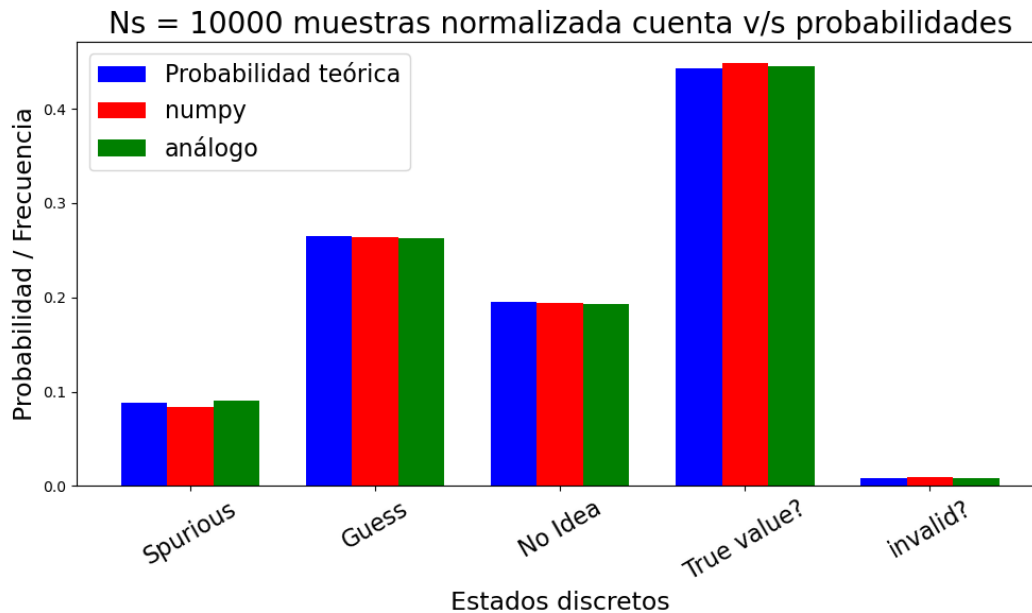


Figura 1: Comparación entre la función de masa de probabilidad (teórica, en azul) y los histogramas obtenidos por muestreo utilizando el Método Análogo y el proporcionado por *NumPy* para una distribución discreta.

Las muestras generadas mediante ambos métodos reproducen adecuadamente la distribución teórica, lo cual sugiere que la implementación del Método Análogo, es correcta desde el punto de vista estadístico. Esto se valida visualmente mediante la superposición de las barras (Fig. 1). El *Método Análogo* se basa en la técnica de *inversión de la función de distribución acumulada*, en la cual un número aleatorio $u \sim \mathcal{U}[0, 1]$ es asignado a un estado según los intervalos definidos por la FDP de la distribución. Este procedimiento, detallado en Tarantola (2005) [1], refleja un enfoque bayesiano en el que las realizaciones de una variable aleatoria representan elecciones posibles del "mundo real" bajo un conocimiento probabilístico.

Por otro lado, si se hace un análisis desde el punto de vista de la eficiencia computacional, el *Numpy* presenta una implementación optimizada y vectorizada internamente, mientras que el *Método Análogo*, aunque también puede vectorizarse (como se hizo en este trabajo), conlleva mayor esfuerzo de programación. No obstante, la correcta implementación del método análogo refuerza la comprensión del proceso de generación de muestras desde una perspec-

tiva probabilística y computacional.

Se observan pequeñas discrepancias entre las frecuencias relativas empíricas y las probabilidades teóricas. Esto se explica con la naturaleza aleatoria de la simulación, y tienden a desaparecer a medida que el número de muestras N_s crece.

3.2. Distribución Uniforme en N dimensiones (P2)

Las figuras 2, 3 y 4 muestran, los resultados de la generación de muestras a partir de una distribución uniforme bidimensional $\mathcal{U}([-2, 5] \times [3.2, 7])$ para $N = 1E4$, $N = 1E5$ y $N = 1E6$. En cada caso, se presenta un histograma 2D que aproxima la función de densidad de probabilidad conjunta, acompañado por histogramas marginales unidimensionales que permiten observar el comportamiento individual de las variables x_1 y x_2 . Los resultados permiten apreciar con claridad cómo el número de muestras afecta la aproximación empírica a la distribución teórica.

Para $N = 1E4$ (Fig. 2), se observa una mayor variabilidad en las frecuencias relativas de los histogramas marginales, y la distribución conjunta aún presenta una granularidad notable. Esto refleja el hecho de que, con una muestra pequeña, las aproximaciones empíricas tienden a desviarse, lo cual es esperable dado el carácter estocástico del muestreo. Esto implica que la estimación de cualquier parámetro basada en este conjunto puede estar sujeta a mayor incertidumbre.

Por otro lado, en $N = 1E5$ (Fig. 3), la aproximación mejora con respecto a la anterior. Las frecuencias de los histogramas marginales tienden a estabilizarse, y la densidad conjunta adquiere una apariencia más homogénea. Esto se alinea con lo esperado, en la medida que aumenta el tamaño muestral, la frecuencia relativa de cada evento converge a su probabilidad teórica.

Por último, en el caso de $N = 1E6$ (Fig. 4), se alcanza una alta fidelidad en la aproximación empírica de la distribución teórica, con frecuencias notablemente uniformes tanto en los histogramas marginales como en el conjunto 2D. En este caso, la muestra puede considerarse representativa del conocimiento asumido, y el espacio de parámetros queda suficientemente explorado.

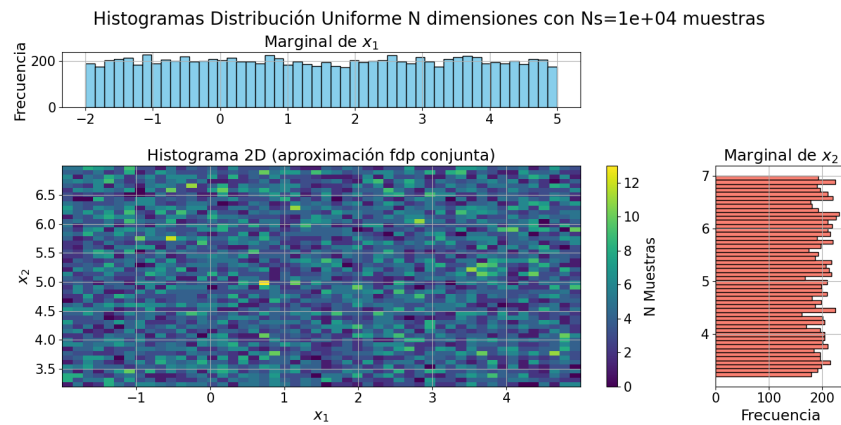


Figura 2: Aproximación de la función de densidad de probabilidad conjunta mediante histogramas 2D para $N_s = 1E4$ de muestras.

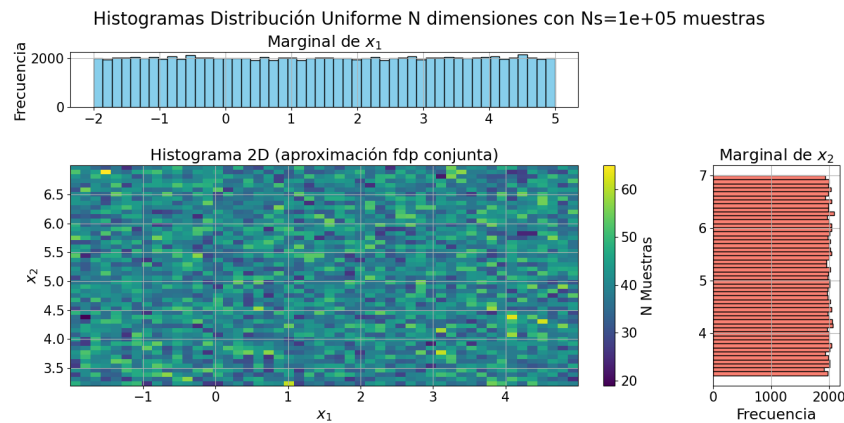


Figura 3: Aproximación de la función de densidad de probabilidad conjunta mediante histogramas 2D para $N_s = 1E5$ de muestras.

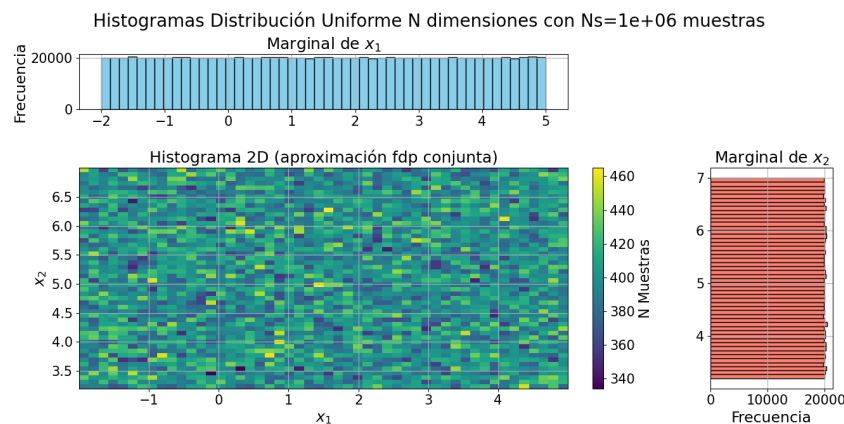


Figura 4: Aproximación de la función de densidad de probabilidad conjunta mediante histogramas 2D para $N_s = 1E6$ de muestras.

3.3. Distribución Normal Multivariada (P3)

En esta sección del trabajo, se analiza el comportamiento del muestreo de una variable aleatoria en \mathbb{R}^2 que sigue una distribución normal multivariada con media $\boldsymbol{\mu} = [0.5, 3.0]^\top$ y matriz de covarianza $C = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$. Las figuras que se presentan a continuación, muestran para tres tamaños de muestras diferentes ($N = 1E4$, $N = 1E5$, $N = 1E6$), el histograma bidimensional (como aproximación de la función de densidad de probabilidad conjunta) y los histogramas marginales de cada componente, junto con sus respectivas funciones de densidad teóricas superpuestas.

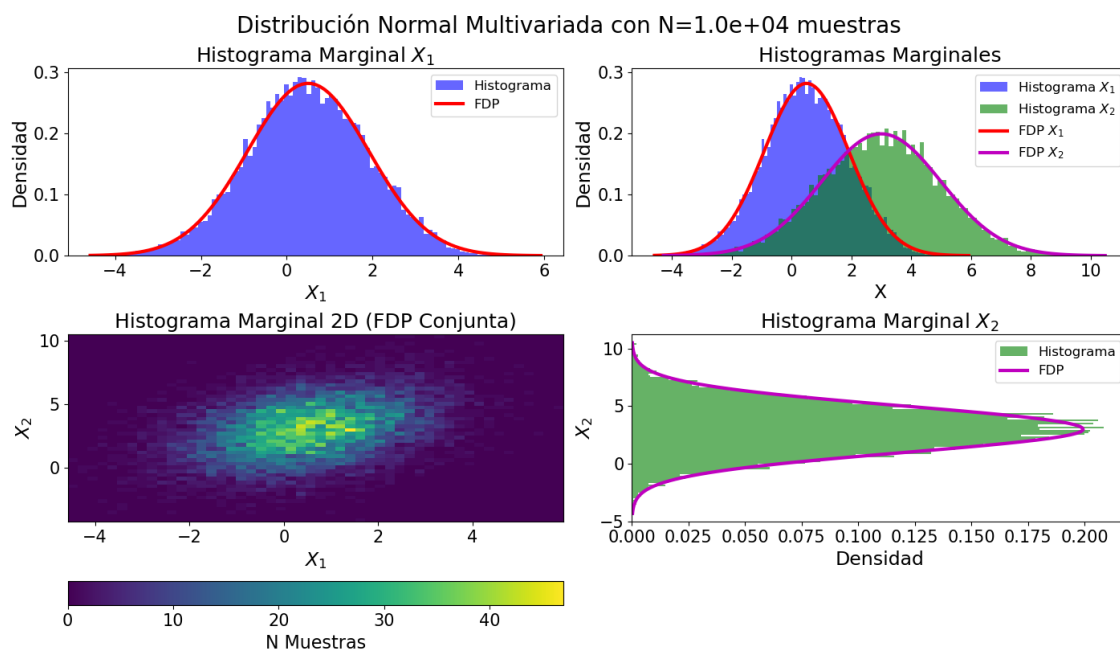


Figura 5: Histogramas de $N = 1E4$ muestras de la función de densidad de probabilidad conjunta de la distribución Normal Multivariada en 2D.

Para $N = 1E4$, se observa una aproximación general a la distribución teórica, pero con variaciones locales visibles tanto en la densidad conjunta como en los histogramas marginales. Las curvas teóricas no coinciden perfectamente con los histogramas, lo cual es esperable debido al bajo tamaño muestral, lo que puede derivar en estimaciones sesgadas y en una representación deficiente de la incertidumbre.

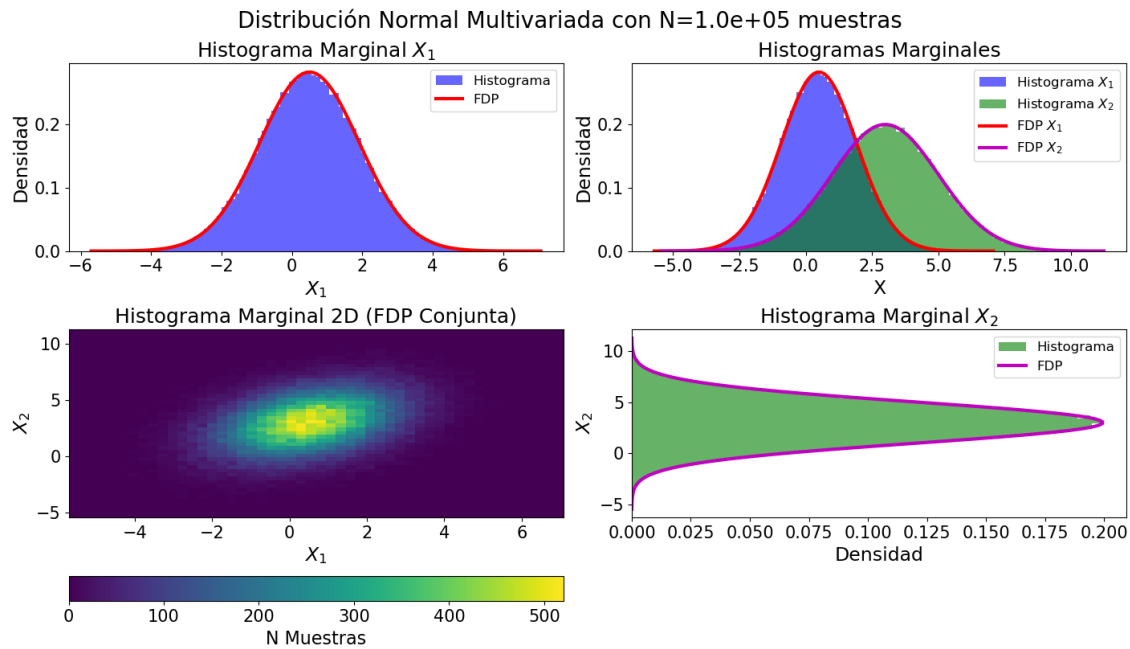


Figura 6: Histogramas de $N = 1E5$ muestras de la función de densidad de probabilidad conjunta de la distribución Normal Multivariada en 2D.

Para $N = 1E5$, la concordancia entre la densidad teórica y los histogramas mejora notablemente. La densidad conjunta presenta una estructura elíptica bien definida, y las curvas teóricas en los marginales se ajustan adecuadamente a los histogramas empíricos. Este comportamiento es esperado ya que al aumentar el número de muestras, las estadísticas muestrales convergen hacia los valores teóricos.

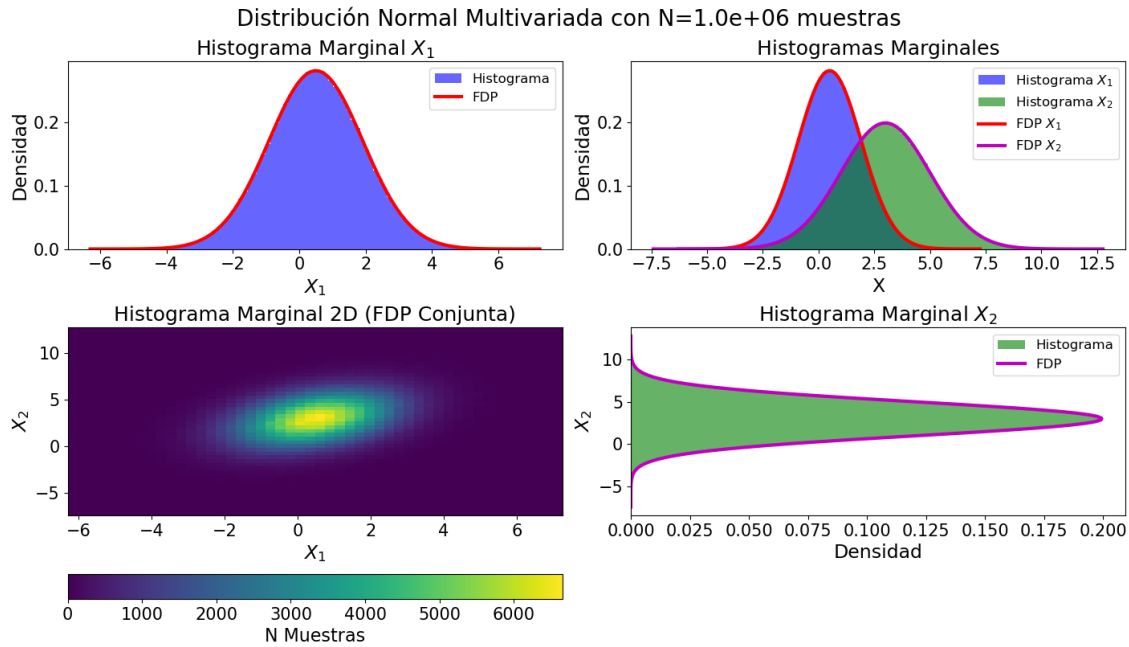


Figura 7: Histogramas de $N = 1E6$ muestras de la función de densidad de probabilidad conjunta de la distribución Normal Multivariada en 2D.

Para $N = 1E6$, se alcanza una mejor representación de la distribución teórica. Las curvas marginales casi coinciden con los histogramas y la densidad conjunta muestra una forma suave y simétrica. Desde el punto de vista de la inferencia probabilística, esto indica que las muestras generadas capturan de forma robusta las características estadísticas del modelo, permitiendo utilizarlas como base para realizar inferencias confiables en el contexto de problemas inversos.

Por otro lado, en la figura 5, correspondiente al caso de $N = 1E4$, se puede observar una orientación elíptica en la distribución conjunta, aunque con ruido visible producto del bajo número de muestras. A medida que el tamaño muestral aumenta (Fig. 6 y 7), la estructura elíptica se vuelve más definida, reflejando con mayor fidelidad la geometría de la distribución normal multivariada subyacente. Esta forma elíptica no es arbitraria, sino que responde directamente a la estructura de la matriz de covarianza C .

El hecho de que los términos fuera de la diagonal sean distintos de cero indica que existe correlación lineal entre las variables aleatorias X_1 y X_2 . Esta correlación se traduce visualmente en la inclinación del contorno de densidad: si las variables fueran independientes (i.e., matriz diagonal), los contornos serían círculos (en caso de varianzas iguales) o elipses alineadas con los ejes (si las varianzas fueran distintas). En cambio, la presencia de correlación hace que los contornos de igual densidad tomen la forma de *elipses rotadas*, cuya orientación depende de los valores propios (autovalores) y vectores propios de C . Este fenómeno es una manifestación directa de la *transformación lineal* que se realiza al aplicar la matriz de Cholesky W , en el método relacional, sobre muestras $Y \sim \mathcal{N}(0, I)$, generando las muestras finales $X = \mu + WY$. La reproducción de esta estructura elíptica en las figuras 6 y 7 no solo valida la implementación del método relacional, sino que evidencia que la matriz de

covarianza empírica de las muestras está convergiendo apropiadamente hacia C .

Este conjunto de resultados ilustra una idea fundamental de los métodos inversos probabilísticos: el conocimiento no se representa con un valor puntual, sino con una distribución que expresa incertidumbre. Y la calidad de la inferencia depende de la capacidad de capturar y representar correctamente esa distribución.

4. Conclusión

El desarrollo de este trabajo permitió implementar y validar distintos métodos de generación de muestras aleatorias, fundamentales en el marco de los problemas inversos probabilísticos. A través de la aplicación del método análogo, el método inverso y el método relacional, se constató que es posible representar adecuadamente el conocimiento y su incertidumbre mediante muestras que aproximan distribuciones teóricas, como se observa en las figuras 1 a 7.

En la figura 1, se evidencia que tanto el método de NumPy como el método análogo reproducen adecuadamente la distribución multinomial teórica, validando estadísticamente su implementación. Esto permite comprender en profundidad la relación entre la función de distribución acumulada y el espacio de estados discretos. Tal como señalan Kaipio y Somersalo [3], la fidelidad del muestreo es crucial para la propagación de la incertidumbre en esquemas bayesianos.

Los resultados de la P2 (Figs. 2–4) destacan cómo el tamaño muestral impacta directamente en la calidad de la aproximación empírica. Mientras que con $N = 1E4$ muestras la distribución conjunta presenta irregularidades notorias, con $N = 1E6$ se alcanza una homogeneidad destacable tanto en los histogramas 2D como en los marginales.

En la P3 (Figs. 5–7), se observa que el uso del método relacional y la descomposición de Cholesky permiten generar muestras de una distribución normal multivariada que reproducen sus propiedades estadísticas.

En los problemas inversos no se busca una única “*mejor solución*”, sino una representación robusta del conjunto de soluciones plausibles, dadas las restricciones impuestas por los datos y el conocimiento previo [1]. En este sentido, el muestreo aleatorio no se limita a ser una herramienta numérica, sino que constituye una vía para explorar el espacio de modelos, estimar parámetros y caracterizar la incertidumbre. Así, puede concluirse que su valor no reside en obtener un único valor “*correcto*”, sino en su capacidad para capturar adecuadamente la incertidumbre inherente al problema. Como afirma Tarantola [1], “*la solución a un problema inverso no es un modelo único, sino una distribución de probabilidad sobre el espacio de modelos*”.

Referencias

- [1] Tarantola, A., Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, 2005.
- [2] Ortega, F., Apuntes del curso de Métodos Inversos Avanzados. Depto, Geofísica, Universidad de Chile, 2025.
- [3] Kaipio, J. y Somersalo, E., Statistical and Computational Inverse Problems. Springer, 2005.