

Reporte AS4103-1

Autor: Julián Proboste
Profesora guía: Claudia Scoccola
Facultad de Ciencias Físicas y Matemáticas
GitHub

July 4, 2025

Contents

1	Introducción	2
1.1	Contexto cosmológico y relevancia del CMB	2
1.2	Estadística tradicional en cosmología	2
1.3	Inferencia basada en simulaciones	3
1.4	Objetivos del proyecto	4

1 Introducción

1.1 Contexto cosmológico y relevancia del CMB

El descubrimiento del Fondo Cósmico de Microondas por Arno Penzias y Robert Wilson en 1965 [1] confirmó la existencia de una radiación isotrópica de microondas que permea el universo. Experimentos posteriores demostraron que este fondo sigue un espectro de cuerpo negro con una temperatura de $T = 2.7255$ K. La densidad de energía del CMB es baja, con aproximadamente 411 fotones por centímetro cúbico, y su longitud de onda característica (2 mm) lo sitúa en la región de microondas del espectro electromagnético [2].

La existencia del CMB fue una evidencia clave a favor del modelo del Big Bang frente al modelo del Estado Estacionario. En un universo en expansión inicialmente caliente y denso, la materia estaba ionizada, lo que hacía al universo opaco. A medida que el universo se expandió y enfrió, los electrones y protones se combinaron formando átomos neutros, permitiendo que los fotones se liberaran. Esta radiación, que en su origen tenía una temperatura de ~ 2970 K, se ha enfriado hasta los 2.7255 K actuales debido a la expansión cósmica.

Con la llegada de las sondas *WMAP* (Wilkinson Microwave Anisotropy Probe) y *Planck*, se lograron mediciones de alta precisión de las anisotropías del CMB, que son fundamentales para entender las inhomogeneidades primordiales que dieron lugar a la estructura a gran escala del universo. La combinación de estos resultados con observaciones independientes, como las de supernovas tipo Ia y la distribución de galaxias, ha permitido refinar la estimación de los parámetros cosmológicos [3], consolidando el modelo Λ CDM como el marco estándar de la cosmología moderna.

1.2 Estadística tradicional en cosmología

En la cosmología moderna, el análisis del CMB sigue una metodología bien establecida que enfrenta nuevos retos ante el volumen y complejidad de los datos actuales. El punto de partida son los mapas de anisotropías de temperatura, que representan fluctuaciones del orden de $\Delta T/T \sim 10^{-5}$ respecto a la temperatura media de 2.7255 K. Estos mapas, obtenidos por misiones como *Planck* [3] y próximamente por el Observatorio Simons [4], contienen información primordial sobre las condiciones iniciales del universo. El análisis tradicional procede mediante la extracción de espectros de potencia C_ℓ , que cuantifican las correlaciones angulares a diferentes escalas ($\ell \sim 180^\circ/\theta$). Esta transformación reduce la dimensionalidad de los datos preservando la información gaussiana:

$$C_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 \quad (1)$$

donde $a_{\ell m}$ son los coeficientes de la expansión en armónicos esféricos del mapa de temperatura. La inferencia de parámetros cosmológicos se realiza mediante funciones de verosimilitud usualmente gaussianas que comparan los espectros observados \hat{C}_ℓ con predicciones teóricas $C_\ell^{\text{teo}}(\lambda_\alpha)$, por ejemplo [5]:

$$\ln \mathcal{L}(\lambda_\alpha) = -\frac{1}{2} \sum_{\ell \ell'} \left(\hat{C}_\ell - C_\ell^{\text{teo}}(\lambda_\alpha) \right) \text{Cov}_{\ell \ell'}^{-1} \left(\hat{C}_{\ell'} - C_{\ell'}^{\text{teo}}(\lambda_\alpha) \right) + \text{cte.} \quad (2)$$

La inferencia bayesiana combina estas verosimilitudes con distribuciones previas $\pi(\lambda_\alpha)$ que incorporan conocimiento teórico y restricciones independientes. En la práctica, la exploración del espacio paramétrico se realiza típicamente mediante métodos de Monte Carlo mediante Cadenas

de Markov (MCMC), que muestrean eficientemente la distribución posterior incluso en espacios de alta dimensionalidad. Este enfoque ha permitido determinar con precisión sin precedentes los parámetros del modelo Λ CDM, pero enfrenta desafíos crecientes ante la complejidad de los nuevos datos. La Figura 1 muestra un esquema de la inferencia tradicional en cosmología.

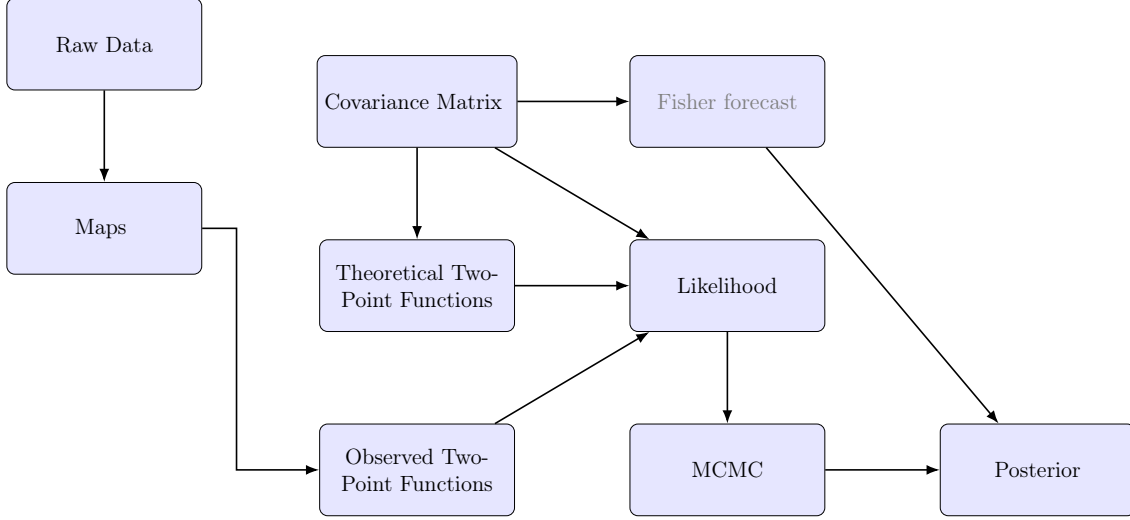


Figure 1: Inferencia tradicional con métodos de MCMC.

1.3 Inferencia basada en simulaciones

El enfoque tradicional presenta dos limitaciones fundamentales: (1) la pérdida de información no-gaussiana durante la compresión a espectros de potencia, y (2) la dependencia de supuestos gaussianos en la verosimilitud. Ambos problemas están estrechamente relacionados, pues si se desea hacer inferencia con resúmenes estadísticos no gaussianos, asumir una verosimilitud gaussiana podría llevar a resultados sesgados. La inferencia basada en simulaciones (SBI) busca realizar inferencia estadística en situaciones donde la función de verosimilitud es intratable o desconocida.

Métodos híbridos como ABC (*Approximate Bayesian Computation*) abordan este problema comparando directamente los datos simulados con los datos observados mediante una métrica de distancia. En este enfoque, se generan múltiples simulaciones a partir de diferentes valores del parámetro θ y se conservan aquellos para los cuales los datos simulados x se parecen lo suficiente a los datos observados, según un umbral de tolerancia. Este método depende fuertemente de la elección de estadísticas resumidas y puede requerir un número muy elevado de simulaciones para obtener una aproximación razonable de la distribución posterior. La Figura 2 se muestra el pipeline general utilizado para estimar posteriores con el método ABC.

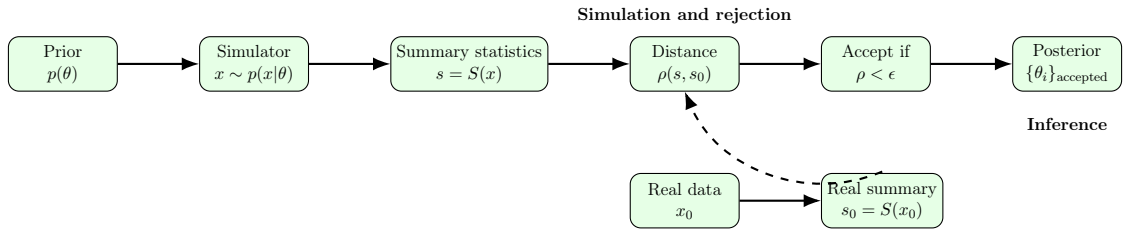


Figure 2: Pipeline general de ABC.

Más recientemente, técnicas modernas basadas en aprendizaje automático como SNPE (*Sequential Neural Posterior Estimation*) [6] o NPSE (*Neural Posterior Score Estimation*) [7] [8] han

revolucionado este enfoque. En lugar de comparar datos de manera directa, estos métodos reformulan el problema como uno de estimación de densidad: se modela la distribución conjunta de pares (θ, x) simulados, lo que permite aproximar la distribución posterior $p(\theta|x)$. Herramientas como redes neuronales profundas y flujos normalizantes permiten entrenar modelos expresivos que aprenden directamente la relación entre datos y parámetros a partir de muestras sintéticas, evitando el cálculo explícito de la verosimilitud y haciendo la inferencia mucho más eficiente. La Figura 3 muestra la pipeline general para técnicas de SBI basadas en aprendizaje automático.

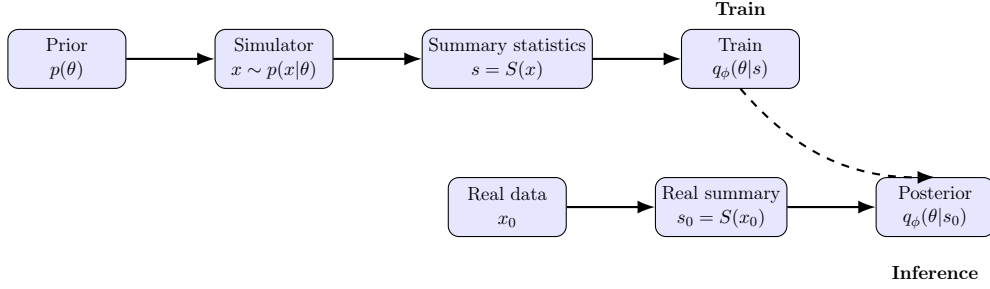


Figure 3: Pipeline general para métodos con aprendizaje automático.

1.4 Objetivos del proyecto

El objetivo principal de este proyecto es la implementación de técnicas de inferencia basadas en simulaciones para estimar parámetros cosmológicos a partir de espectros de potencia angulares del CMB

References

- [1] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 1965. doi: 10.1086/148307.
- [2] Barbara Ryden. *Introduction to Cosmology*. Cambridge University Press, 2 edition, 2017.
- [3] Planck Collaboration. Planck 2018 results. VI. cosmological parameters. *arXiv preprint arXiv:1807.06209*, Jun 2018.
- [4] Simons Observatory Collaboration. The simons observatory: Science goals and forecasts. *arXiv preprint arXiv:1808.07445*, Mar 2019.
- [5] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. Academic Press, 2 edition, 2020.
- [6] David S Greenberg, Marcel Nonnenmacher, and Jakob H Macke. Automatic posterior transformation for likelihood-free inference. *arXiv preprint arXiv:1905.07488*, May 2019.
- [7] Louis Sharrock, Jack Simons, Song Liu, and Mark Beaumont. Sequential neural score estimation: Likelihood-free inference with conditional score based diffusion models. *arXiv preprint arXiv:2210.04872*, Jun 2024.
- [8] Tomas Geffner, George Papamakarios, and Andriy Mnih. Compositional score modeling for simulation-based inference. *arXiv preprint arXiv:2209.14249*, Jul 2023.