

# Lecture notes

## Contents

<b>1</b>	<b>Introducción</b>	<b>2</b>
1.1	Motivación: Desafíos de la inferencia cosmológica moderna . . . . .	2
<b>2</b>	<b>Statistics in cosmology</b>	<b>2</b>
2.1	The Likelihood function . . . . .	3
2.2	From raw data to parameter constraints . . . . .	4
<b>3</b>	<b>Inferencia basada en simulaciones</b>	<b>5</b>
<b>4</b>	<b>Estimación secuencial de densidad neuronal</b>	<b>6</b>
<b>5</b>	<b>Introducciónnn</b>	<b>7</b>
5.1	Motivación . . . . .	7
5.2	Inferencia Bayesiana . . . . .	7
5.3	<i>Likelihood-free inference</i> . . . . .	7
<b>6</b>	<b>Simulador</b>	<b>8</b>
6.1	Generador de APS . . . . .	8
6.2	Ruido instrumental y cobertura parcial del cielo . . . . .	9
<b>7</b>	<b>Inference</b>	<b>9</b>
<b>8</b>	<b>Métodos bayesianos</b>	<b>10</b>
8.1	Función de verosimilitud . . . . .	10
<b>9</b>	<b>Inferencia Basada en Simulación con Simuladores Diferenciables</b>	<b>10</b>

## 1 Introducción

### 1.1 Motivación: Desafíos de la inferencia cosmológica moderna

La cosmología observacional enfrenta una coyuntura crítica impulsada por la explosión en la cantidad y complejidad de los datos disponibles. Experimentos recientes como *Planck* [1] y Simons Observatory [2], junto con proyectos futuros como CMB-S4 [3], generan conjuntos de datos cuyo análisis exhaustivo mediante métodos tradicionales basados en cadenas de Markov (MCMC) requiere recursos computacionales prohibitivos. Un caso emblemático es el análisis de los datos de *Planck* 2018, que necesitó más de 10 millones de muestras MCMC para alcanzar convergencia en espacios de parámetros de más de 30 dimensiones [1], con un coste escalando como  $\mathcal{O}(N^3)$  para  $N$  modos observados [4]. Esta barrera computacional se agudizará en la próxima generación de experimentos con resoluciones angulares sub-arcminuto.

Paralelamente, la reducción de errores estadísticos por debajo del 1% en mediciones de polarización [5] ha expuesto limitaciones fundamentales en el paradigma de inferencia actual. Errores sistemáticos anteriormente despreciables, tanto instrumentales (ruido correlacionado, efectos de haz asimétrico) como teóricos (no-Gaussianidad primordial, efectos de segundo orden en transferencia radiativa [6]), dominan ahora las incertidumbres. Particularmente problemática resulta la creciente evidencia de desviaciones de Gaussianidad en pequeños escalones angulares ( $\ell > 2000$ ), donde los efectos de lente gravitacional distorsionan significativamente las estadísticas del fondo cósmico de microondas [7]. Estas complejidades se amplifican al considerar análisis conjuntos de múltiples estadísticos como  $C_\ell^{TT}$ ,  $C_\ell^{TE}$  y funciones de correlación  $\xi_+(\theta)$ , cuyas covarianzas cruzadas introducen no-linealidades no triviales [8].

Este panorama exige desarrollar frameworks de inferencia alternativos que superen tres limitaciones clave: primero, la incapacidad de los métodos basados en likelihood exacto para escalar a los volúmenes de datos futuros; segundo, la dependencia crítica de aproximaciones Gaussianas cada vez más cuestionables; y tercero, la dificultad para incorporar sistemas complejos en modelos analíticos. La inferencia basada en simulaciones (SBI) emerge como solución prometedora al integrar emuladores de alta dimensión [9], técnicas de compresión óptima de datos [10] y algoritmos de muestreo eficiente [11], permitiendo una caracterización rigurosa de parámetros cosmológicos directamente desde observaciones sin necesidad de likelihoods explícitos.

## 2 Statistics in cosmology

Los cosmólogos enfrentan nuevos desafíos en el análisis de datos debido al crecimiento exponencial en la cantidad y calidad de información disponible. Experimentos recientes generan conjuntos de datos masivos que requieren métodos avanzados, ya que algoritmos tradicionales se vuelven inefficientes. Además, al reducirse los errores estadísticos, emergen errores sistemáticos previamente enmascarados, relacionados tanto con artefactos instrumentales como con limitaciones teóricas.

En esta introducción se presentan técnicas para el análisis cosmológico moderno, comenzando con conceptos estadísticos básicos (verosimilitud, previos y posteriores) y su aplicación a espectros de potencia. Introduce herramientas como la *Matriz de Fisher* (para estimación rápida de errores) y métodos *Markov Chain Monte Carlo* (MCMC), que permiten manejar verosimilitudes complejas. Estas metodologías, aunque desarrolladas para cosmología, tienen aplicaciones transversales en diversas áreas científicas que trabajan con grandes volúmenes de datos.

## 2.1 The Likelihood function

La función de verosimilitud  $L(\{d_i\}|\theta)$  representa el núcleo conceptual de los análisis estadísticos modernos en cosmología. Se define formalmente como la probabilidad condicional de observar un conjunto de datos experimentales  $\{d_i\}_{i=1}^m$  dado un modelo teórico parametrizado por  $\theta = (w, \sigma_w)$ . Para mediciones independientes, esta función se expresa como el producto de probabilidades individuales:

$$L(\{d_i\}|\theta) = \prod_{i=1}^m P(d_i|\theta), \quad (1)$$

donde cada término  $P(d_i|\theta)$  corresponde a la distribución de probabilidad del modelo para cada dato observado. La potencia analítica de este constructo matemático reside en su capacidad para invertir la relación lógica entre datos y teoría mediante la aplicación del teorema de Bayes:

$$P(\theta|\{d_i\}) \propto L(\{d_i\}|\theta)P(\theta), \quad (2)$$

permitiendo inferir los parámetros del modelo a partir de las observaciones. En el contexto del ejemplo pedagógico: -la medición del peso  $w$  de un individuo utilizando  $m = 100$  básculas independientes con error gaussiano-, cada observación  $d_i$  sigue una distribución normal  $d_i \sim \mathcal{N}(w, \sigma_w^2)$ . La verosimilitud para una sola medición adopta la forma gaussiana clásica:

$$L(d_i|w, \sigma_w) = (2\pi\sigma_w^2)^{-1/2} \exp\left(-\frac{(d_i - w)^2}{2\sigma_w^2}\right). \quad (3)$$

Al extenderse a múltiples observaciones independientes, la verosimilitud conjunta se factoriza en un producto de términos gaussianos, lo que en escala logarítmica se traduce en una suma cuadrática:

$$\ln L(\{d_i\}|w, \sigma_w) = -\frac{m}{2} \ln(2\pi\sigma_w^2) - \frac{1}{2\sigma_w^2} \sum_{i=1}^m (d_i - w)^2. \quad (4)$$

Los estimadores de máxima verosimilitud (MLE) para los parámetros se obtienen mediante la optimización de esta función. Para el peso  $w$ , la condición  $\partial \ln L / \partial w = 0$  conduce al estimador:

$$\hat{w} = \frac{1}{m} \sum_{i=1}^m d_i, \quad (5)$$

que corresponde a la media muestral. Análogamente, al resolver  $\partial \ln L / \partial \sigma_w^2 = 0$  se obtiene el estimador para la varianza

$$\hat{\sigma}_w^2 = \frac{1}{m} \sum_{i=1}^m (d_i - \hat{w})^2. \quad (6)$$

Estos resultados ejemplifican cómo los principios de máxima verosimilitud recuperan estimadores intuitivos en casos simples. La transición desde la verosimilitud hacia la distribución posterior de parámetros se realiza mediante el teorema de Bayes:

$$P(w, \sigma_w | \{d_i\}) \propto L(\{d_i\} | w, \sigma_w) P_{\text{prior}}(w, \sigma_w), \quad (7)$$

donde  $P_{\text{prior}}$  incorpora cualquier conocimiento a priori sobre los parámetros. Esta distribución posterior contiene toda la información probabilística sobre los parámetros condicionada a los datos observados. La constante de normalización, denominada evidencia, asegura que la posterior integre a unidad sobre el espacio de parámetros. El análisis de incertidumbres en los estimadores se fundamenta en el examen de la curvatura del logaritmo de la verosimilitud alrededor de su máximo. Para el parámetro  $w$ , la varianza del estimador resulta:

$$\text{Var}(\hat{w}) = \frac{\sigma_w^2}{m}, \quad (8)$$

reflejando la reducción clásica de ruido conforme  $\sqrt{m}$ . En contraste, la estimación de  $\sigma_w^2$  presenta comportamientos no-gaussianos, con varianza:

$$\text{Var}(\hat{\sigma}_w^2) = \frac{2}{m} \sigma_w^4. \quad (9)$$

Este error en  $\sigma_w$  puede parecer un detalle técnico, pero en cosmología, gran parte de lo que medimos (como fluctuaciones en la densidad de galaxias o la temperatura del CMB) es análogo a  $\sigma_w^2$ . Estas fluctuaciones siguen distribuciones (a menudo gaussianas) cuyos parámetros dependen del modelo cosmológico. Por ello, esta ecuación es fundamental: al estimar la varianza de una distribución, hay una incertidumbre intrínseca proporcional a  $\sigma_w^2/\sqrt{m}$ , llamada varianza muestral o varianza cósmica.

Por otro lado, es conveniente definir intervalos de credibilidad para los parámetros. En el ejemplo de la verosimilitud gaussiana, los intervalos de confianza al 68% (1- $\sigma$ ) corresponden a la región donde  $\ln L$  decrece en 1/2 unidades desde su valor máximo, delimitando así el rango de parámetros consistentes con los datos. De manera general, se definen los valores  $\omega_-$  y  $\omega_+$  tales que:

$$\int_{\omega_-}^{\omega_+} d\omega P(\omega|\{d_i\}) = 0.68, \quad (10)$$

En aplicaciones reales, usualmente hay múltiples parámetros desconocidos. Si algunos son irrelevantes, debemos marginalizar sobre ellos, formalmente esto se implementa mediante la integración multidimensional:

$$P(\theta|\{d_i\}) = \int_{\Phi} P(\theta, \phi|\{d_i\}) d\phi, \quad (11)$$

sobre el espacio completo  $\Phi$  de parámetros *nuisance*.

## 2.2 From raw data to parameter constraints

El análisis cosmológico moderno sigue una metodología bien establecida que comienza con los mapas observacionales. Estos representan la materia prima del estudio, correspondiendo a representaciones espaciales de los fenómenos cósmicos investigados. En el caso de la radiación cósmica de fondo (CMB), se trabaja con mapas de anisotropías de temperatura; para sondeos galácticos, con distribuciones tridimensionales de densidad; y en estudios de lente gravitacional débil, con campos de elipticidad galáctica.

A partir de estos mapas, se extraen los estadísticos de dos puntos, que capturan las correlaciones fundamentales mediante funciones de dos puntos. Estas pueden expresarse tanto en el espacio espectral como en el espacio de configuraciones, proporcionando así una descripción estadística compacta de las propiedades de los campos observados. La transición de los mapas brutos a

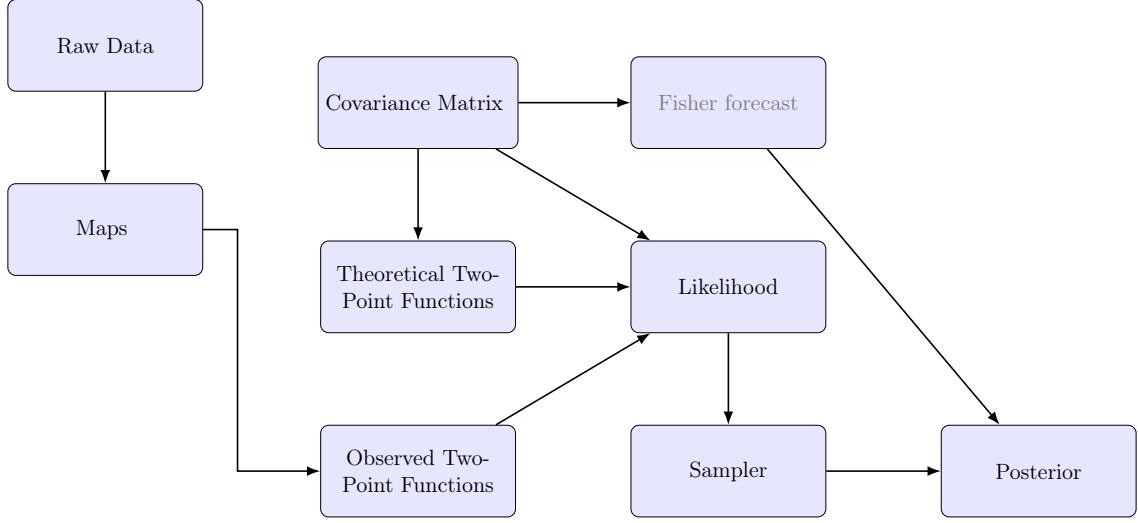


Figure 1: Pipeline desde los datos crudos hasta las restricciones cosmológicas.

estos estadísticos permite reducir considerablemente la dimensionalidad de los datos, reduciendo la complejidad computacional, pero perdiendo información no gaussiana.

Con los estadísticos medidos en mano, el siguiente paso es la construcción de la verosimilitud, que compara estas mediciones con sus contrapartes teóricas. Esta comparación se realiza ponderando las diferencias mediante la matriz de covarianza inversa. Para el caso gaussiano, la función de verosimilitud para las anisotropías del CMB toma la forma:

$$\ln \mathcal{L}(\lambda_\alpha) = -\frac{1}{2} \sum_{ll'} \left( \hat{C}(l) - C_{\text{theory}}(l, \lambda_\alpha) \right) \text{Cov}_{ll'}^{-1} \left( \hat{C}(l') - C_{\text{theory}}(l', \lambda_\alpha) \right). \quad (12)$$

Para explorar el espacio de parámetros cosmológicos  $\lambda_\alpha$  se emplean técnicas de muestreo de parámetros, particularmente algoritmos MCMC (Monte Carlo Markov Chain). Estos algoritmos evalúan sistemáticamente millones de combinaciones paramétricas, permitiendo caracterizar completamente la distribución posterior. Este proceso debe manejar adecuadamente dos desafíos clave: las degeneraciones entre parámetros cosmológicos y la marginalización sobre parámetros molestos no cosmológicos, como el parámetro de bias  $b_1$  en estudios de distribución galáctica.

Como complemento al muestreo exhaustivo, las estimaciones analíticas mediante pronósticos de Fisher ofrecen una alternativa eficiente para proyectar barras de error. Estas aproximaciones analíticas son especialmente valiosas en etapas preliminares de diseño experimental, cuando se requiere evaluar rápidamente el potencial científico de diferentes configuraciones instrumentales sin recurrir a costosos análisis completos. El proceso completo se ilustra en la Figura 1.

### 3 Inferencia basada en simulaciones

La inferencia sin verosimilitud (*Likelihood-Free Inference*, LFI) busca estimar con precisión la distribución posterior  $p(\theta|x_o)$  sin acceso a la verosimilitud  $p(x|\theta)$ , utilizando un número limitado de simulaciones. En este contexto, la tarea puede formularse como un problema de estimación de densidad condicional, donde se entrena una red neuronal  $F$  para mapear los datos simulados  $x$  a una estimación de la posterior  $p(\theta|x)$ . Una vez entrenada, la red puede aplicarse directamente sobre los datos observados  $x_o$ .

Para el entrenamiento, se simulan pares de parámetros y datos  $(\theta_j, x_j)$  muestreados del prior, y

se minimiza la pérdida

$$L(\phi) = - \sum_{j=1}^N \log q_{F(x_j, \phi)}(\theta_j),$$

donde  $q_{F(x, \phi)}(\theta)$  es la estimación neuronal de la densidad posterior. Si el modelo es suficientemente expresivo, el mapeo aprendido aproxima correctamente la distribución posterior real.

## 4 Estimación secuencial de densidad neuronal

El enfoque secuencial busca mejorar la eficiencia del entrenamiento al priorizar simulaciones en regiones del espacio de parámetros donde la densidad posterior  $p(\theta|x_o)$  es mayor. Esto se logra mediante una propuesta  $\tilde{p}(\theta)$  que se actualiza iterativamente, un proceso conocido como *Sequential Neural Posterior Estimation* (SNPE). Sin embargo, al usar muestras de una propuesta distinta del prior, el entrenamiento optimiza una *posterior de la propuesta*  $\tilde{p}(\theta|x)$ , que difiere de la verdadera posterior por un factor de corrección:

$$\tilde{p}(\theta|x) = p(\theta|x) \frac{\tilde{p}(\theta)}{p(\theta)} \frac{p(x)}{\tilde{p}(x)}.$$

Se han desarrollado tres variantes principales para abordar este problema. **SNPE-A** corrige post-hoc la distribución mediante una solución analítica bajo suposiciones gaussianas, aunque no maneja propuestas multimodales y puede generar matrices de covarianza no definidas positivas. **SNPE-B** incorpora pesos de importancia en la función de pérdida, eliminando la necesidad de corrección posterior, pero a costa de una alta varianza en las actualizaciones de parámetros. Finalmente, **SNL** estima la verosimilitud  $p(x|\theta)$  directamente, permitiendo el uso de propuestas arbitrarias o esquemas de aprendizaje activo, aunque requiere un muestreo MCMC adicional para recuperar la posterior, lo cual puede resultar costoso.

APT supera estas limitaciones al combinar las ventajas de los enfoques anteriores sin requerir pesos de importancia ni correcciones post-hoc, logrando un equilibrio entre flexibilidad, estabilidad y eficiencia computacional.

## 5 Introducción

### 5.1 Motivación

La inferencia sin verosimilitud (*likelihood-free inference*) permite estimar parámetros cosmológicos sin asumir una forma analítica para la función de verosimilitud, evitando así sesgos por supuestos incorrectos como Gaussianidad. Esto es especialmente relevante en cosmología, donde procesos no lineales y efectos sistemáticos complican el modelado estadístico. En el contexto del CMB, aunque las fluctuaciones primordiales son aproximadamente Gaussianas, efectos secundarios, como lentes gravitacionales, polarización inducida y *foregrounds* introducen no-Gaussianidades que desafían los análisis tradicionales basados en espectros de potencia.

El CMB es uno de los observables más poderosos para estudiar los parámetros cosmológicos, pero su buena explotación requiere ir más allá de las estadísticas de dos puntos. En este trabajo, comenzaré con el análisis de estadísticas resumidas tradicionales, partiendo de los *angular power spectrum* (APS) de temperatura, que capturan la información gaussiana primordial. Posteriormente, escalaré hacia estadísticas más complejas que incluyan modos de polarización, para finalmente explorar métodos no lineales (de ser posible).

Métodos de aprendizaje profundo, como la compresión neuronal, pueden extraer información no lineal de mapas del CMB, mejorando las restricciones sobre parámetros como  $\Omega_m$  o  $\sigma_8$ . Además, la inclusión de efectos no ideales como ruido instrumental, cortes en el cielo y *foregrounds* en simulaciones realistas es importante para un análisis robusto. Esta aproximación gradual—desde espectros de potencia hasta estadísticas de alto orden—permitirá validar resultados intermedios y cuantificar ganancias al incorporar información no gaussiana.

### 5.2 Inferencia Bayesiana

El marco bayesiano permite determinar la distribución posterior de parámetros cosmológicos  $\theta$  mediante la expresión

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \quad (13)$$

donde  $x$  representa los datos observacionales. La determinación precisa de la función de verosimilitud  $p(x|\theta)$  presenta varios desafíos en el análisis del CMB. Aunque las fluctuaciones primordiales son aproximadamente gaussianas, efectos secundarios como las lentes gravitacionales, la contaminación por *foregrounds* y las no-linealidades instrumentales introducen complejidades. Estas complicaciones se acentúan al considerar estadísticas más allá del espectro de potencia, como los estudios de no-gaussianidad primordial o los análisis de lentes gravitacionales a pequeña escala.

Los métodos tradicionales basados en supuestos gaussianos para la verosimilitud muestran limitaciones crecientes frente a la precisión de los nuevos experimentos. Particularmente, para análisis que involucran reconstrucción de lentes, estudios de no-gaussianidad o el tratamiento de regiones con máscaras complejas, la distribución exacta de los estadísticos se desvía significativamente de la aproximación gaussiana. Esta discrepancia puede llevar a estimaciones sesgadas de parámetros cosmológicos.

### 5.3 *Likelihood-free inference*

La inferencia sin verosimilitud (también llamada *Simulation-based inference*) busca realizar inferencia estadística en situaciones donde la función de verosimilitud es intratable o desconocida. Métodos tradicionales como el *Approximate Bayesian Computation* (ABC) 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  $\theta$  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 1 se muestra la pipeline general utilizada para estimar posteriores con el método ABC.

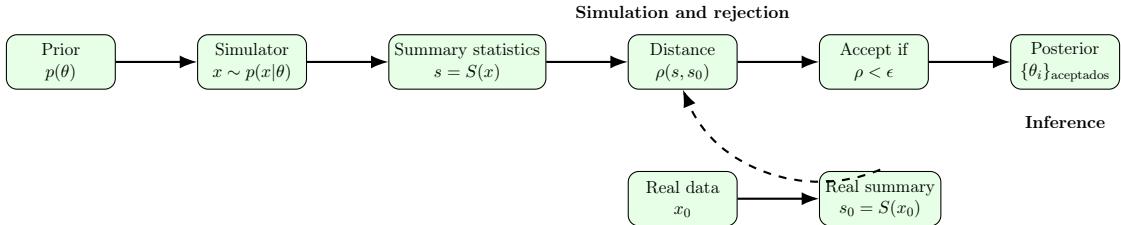


Figure 2: Pipeline general de ABC.

Más recientemente, técnicas modernas basadas en aprendizaje automático 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  $(\theta, x)$  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 2 muestra la pipeline general para técnicas de sbi basadas en aprendizaje automático.

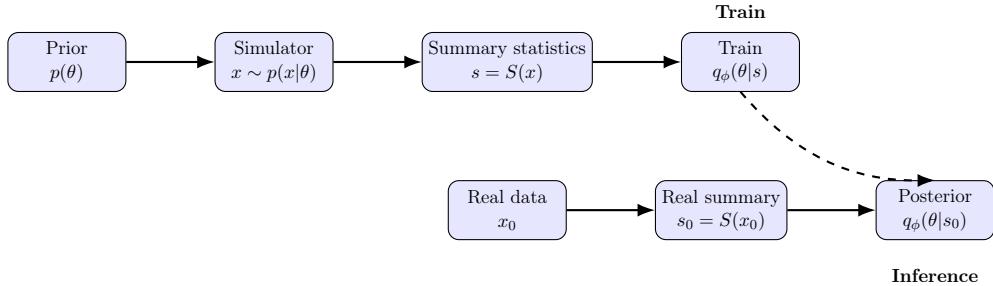


Figure 3: Pipeline general para métodos con aprendizaje automático.

El paquete `sbi`, desarrollado en Python sobre PyTorch, implementa este paradigma de manera flexible. Ofrece herramientas para: (1) simular datos bajo distintos parámetros  $\theta$ , (2) entrenar estimadores de densidad condicional  $p(\theta|x)$  usando arquitecturas modernas, y (3) realizar inferencia posterior dado un conjunto de datos observados. La calidad de las simulaciones y de las estadísticas utilizadas continúa siendo muy importante, ya que el modelo entrenado sólo puede aprender lo que está presente en los datos simulados. Sin embargo, al entrenar modelos que generalizan bien, se puede reducir de forma significativa el número de simulaciones necesarias, logrando una inferencia más precisa y escalable que con métodos como ABC.

## 6 Simulador

### 6.1 Generador de APS

El generador implementa un cálculo teórico completo de las anisotropías del fondo cósmico de microondas mediante la resolución numérica de las ecuaciones de Boltzmann a través del código

CAMB. Parte de un conjunto de parámetros cosmológicos que definen un modelo  $\Lambda$ CDM estándar y devuelve su APS correspondiente. En la actual implementación, el cálculo se restringe al espectro de temperatura, aunque la arquitectura del código permite una extensión futura para incluir los modos de polarización. El resultado final consiste en el APS completo expresado en microkelvin, que cuantifica las fluctuaciones de temperatura en función del multipolo  $\ell$ .

## 6.2 Ruido instrumental y cobertura parcial del cielo

Para generar simulaciones realistas de los APS, se consideran dos fuentes principales de ruido: el ruido instrumental del experimento y la cobertura parcial del cielo, que introduce varianza cósmica adicional. Primero, se añade el ruido instrumental a los espectros teóricos mediante un modelo basado en la resolución angular del experimento ( $\theta_{\text{fwhm}}$ ) y la sensibilidad por píxel ( $\sigma_T$ ). El término de ruido instrumental  $N_\ell^{\text{TT}}$  que se suma al espectro de potencia teórico  $C_\ell$  es:

$$N_\ell^{\text{TT}} = (\theta_{\text{fwhm}} \cdot \sigma_T)^2 \exp \left[ \ell(\ell + 1) \frac{\theta_{\text{fwhm}}^2}{8 \ln 2} \right], \quad (14)$$

donde  $\theta_{\text{fwhm}}$  se expresa en radianes. Esta fórmula modela el suavizado del cielo debido a la resolución finita del instrumento. Posteriormente, se simula la observación del cielo parcial considerando que solo se mide una fracción  $f_{\text{sky}} < 1$  del cielo completo. Como consecuencia, el estimador observable de  $C_\ell$ , denotado  $\hat{C}_\ell$ , no es determinístico, sino una variable aleatoria cuya dispersión depende de  $f_{\text{sky}}$ . Para multipolos bajos ( $\ell < \ell_{\text{transition}}$ ), se modela esta varianza como una distribución chi-cuadrado escalada:

$$\hat{C}_\ell = \frac{1}{\nu_\ell} \sum_{i=1}^{\nu_\ell} X_i^2, \quad X_i \sim \mathcal{N}(0, \sqrt{C_\ell}), \quad (15)$$

donde  $\nu_\ell = \text{round}(f_{\text{sky}} \cdot (2\ell + 1))$  representa el número efectivo de grados de libertad. Esta formulación captura correctamente la dispersión estadística del estimador cuando el número de modos disponibles es pequeño. Para multipolos altos ( $\ell \geq \ell_{\text{transition}}$ ), se asume que el estimador puede aproximarse mediante una distribución normal centrada en  $C_\ell$  con varianza:

$$\hat{C}_\ell \sim \mathcal{N} \left( C_\ell, \frac{2C_\ell^2}{f_{\text{sky}}(2\ell + 1)} \right). \quad (16)$$

Este enfoque es válido debido al teorema central del límite, ya que en esta región el número de modos es suficientemente grande como para justificar una aproximación gaussiana.

## 7 Inference

## 8 Métodos bayesianos

En las últimas décadas, la cosmología ha experimentado una explosión en la cantidad y calidad de datos observacionales, desde mediciones de anisotropías en el Fondo Cósmico de Microondas (CMB) hasta sondeos de galaxias y lentes gravitacionales a gran escala. Este crecimiento ha impulsado la necesidad de técnicas de análisis más sofisticadas, capaces de manejar conjuntos de datos masivos y modelos teóricos cada vez más complejos.

En este contexto, los métodos bayesianos se han convertido en una herramienta fundamental para la inferencia cosmológica. Estos permiten combinar observaciones con predicciones teóricas mediante el uso de distribuciones de probabilidad, actualizando nuestro conocimiento a través del teorema de Bayes:

$$P(\theta | \mathcal{D}) = \frac{P(\mathcal{D} | \theta) P(\theta)}{P(\mathcal{D})}, \quad (17)$$

donde  $P(\theta | \mathcal{D})$  es la distribución posterior,  $P(\mathcal{D} | \theta)$  es la verosimilitud,  $P(\theta)$  es la distribución previa, y  $P(\mathcal{D})$  actúa como evidencia.

Sin embargo, en cosmología, el cálculo de la verosimilitud suele ser computacionalmente costoso o incluso intratable cuando se trabaja con simulaciones numéricas. Esto ha llevado al desarrollo de técnicas de *Simulation-Based Inference* (SBI), que evitan la evaluación explícita de la verosimilitud mediante aproximaciones basadas en simulaciones. En las siguientes secciones, exploraremos cómo estos métodos están revolucionando el análisis de datos cosmológicos.

### 8.1 Función de verosimilitud

La función de verosimilitud  $L$  es fundamental en análisis estadísticos modernos, definida como la probabilidad de observar los datos dado un modelo teórico. Esta herramienta permite estimar parámetros del modelo y sus incertidumbres. Por ejemplo, al medir repetidamente el peso de una persona con múltiples básculas, la verosimilitud modela cómo las mediciones  $d_i$  se distribuyen alrededor del peso real  $w$ , asumiendo un ruido gaussiano con desviación estándar  $\sigma_w$ .

$$L(w|d_i, \sigma_w) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{(d_i - w)^2}{2\sigma_w^2}\right) \quad (18)$$

Para  $m$  mediciones independientes, la verosimilitud conjunta es el producto de las verosimilitudes individuales, resultando en una expresión exponencial que depende de la suma de cuadrados de las diferencias entre datos y modelo.

$$L(w|d_i, \sigma_w) = \prod_i \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{(d_i - w)^2}{2\sigma_w^2}\right) \quad (19)$$

## 9 Inferencia Basada en Simulación con Simuladores Diferenciables

La inferencia bayesiana tiene como objetivo inferir los parámetros  $\theta_0$  que han generado una observación dada  $x_0$ . Del teorema de Bayes tenemos

$$p(\theta|x_0) = \frac{p(x_0|\theta)p(\theta)}{p(x_0)} \propto p(x_0|\theta)p(\theta), \quad (20)$$

y dado que  $x_0$  es el resultado de un gran número de transformaciones que involucran un gran número de variables latentes  $z$ , la verosimilitud marginal

$$p(x_0|\theta) = \int p(x_0|\theta, z)p(z)dz, \quad (21)$$

es intratable.

SBI es particularmente útil en este caso ya que proporciona un marco para aproximar la posterior sin usar una verosimilitud analítica. Para aproximar  $p(\theta|x_0)$  los algoritmos SBI requieren la observación  $x_0$ , una previa  $p(\theta)$  para los parámetros del modelo  $\theta$  y un simulador  $x \sim p(x|\theta)$  para muestrear de la verosimilitud intratable.

En este trabajo estamos interesados en hacer NPE que tiene como objetivo aprender directamente la siguiente distribución

$$p(\theta|x) \propto \int p(x|\theta, z)p(z|\theta)p(\theta)dz, \quad (22)$$

usando un Estimador de Densidad Neuronal (NDE) como un Flujo Normalizante (en adelante NF).

Se puede entrenar el NDE para aprender la distribución aproximada  $p_\phi(\theta|x)$  a partir de muestras  $(x, \theta)$  de la distribución conjunta minimizando

$$\mathcal{L}_{\text{NLL}} = \mathbb{E}_{p(\theta, x)} [-\log p_\phi(\theta|x)], \quad (23)$$

y luego evaluarla en la observación dada  $x_0$  para obtener la posterior aproximada  $p_\phi(\theta|x=x_0) \approx p(\theta|x_0)$ .

Como se reconoce en Brehmer et al. (2020), cuando los simuladores son diferenciables se puede extraer para cada simulación los gradientes con respecto a los parámetros de simulación, lo que proporciona significativamente más información que las muestras del simulador, y puede usarse para ayudar a constreñir las estimaciones de densidad posterior obtenidas por SBI. Por lo tanto, en nuestro caso, extraemos para cada simulación  $(x_i, \theta_i)$  el gradiente de la log-probabilidad conjunta del simulador con respecto a los parámetros de entrada. En lo siguiente, notaremos este gradiente como

$$\nabla_\theta \log p(\theta|x, z) = \nabla_\theta \log p(x|\theta, z) \quad (24)$$

$$+ \nabla_\theta \log p(z|\theta) \quad (25)$$

$$+ \nabla_\theta \log p(\theta), \quad (26)$$

donde  $z$  son variables estocásticas latentes del simulador. En un ligero abuso de lenguaje, nos referiremos a  $\nabla_\theta \log p(\theta|x, z)$  como el *score* conjunto, pero debe notarse que la definición convencional de la función *score* en estadística (también adoptada en Brehmer et al. (2020)) es el gradiente de la log-verosimilitud  $\nabla_\theta \log p(x|\theta, z)$ .

Podemos entonces definir una pérdida de emparejamiento de *score* directa

$$\mathcal{L}_{\text{SM}} = \mathbb{E}_{p(x, z, \theta)} [\|\nabla_\theta \log p(\theta|x, z) - \nabla_\theta \log p_\phi(\theta|x)\|_2^2]. \quad (27)$$

Inspirados por Brehmer et al. (2020), esta cantidad se minimiza por

$$\mathbb{E}_{p(z|x, \theta)} [\nabla_\theta \log p(\theta|x, z)] = \mathbb{E}_{p(z|x, \theta)} \left[ \nabla_\theta \log \frac{p(\theta, z|x)p(x)}{p(z, x)} \right] \quad (28)$$

$$= \mathbb{E}_{p(z|x, \theta)} [\nabla_\theta \log p(\theta, z|x)] \quad (29)$$

$$= \nabla_\theta \log p(\theta|x) \quad (30)$$

lo que significa que al minimizar  $\mathcal{L}_{\text{SM}}$  aproximamos el *score* marginal intratable.

Finalmente, entrenamos nuestro NDE usando la siguiente pérdida combinada:

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + \lambda \mathcal{L}_{\text{SM}}, \quad (31)$$

donde  $\lambda$  es un hiper-parámetro que puede usarse para sintonizar la contribución del *score* a la función de pérdida. El valor óptimo para  $\lambda$  típicamente depende del problema considerado.

Nótese que para entrenar el NDE constriñendo su *score*  $\nabla_{\theta} \log p_{\phi}(\theta|x)$ , necesita ser suficientemente suave con respecto a  $\theta$ , motivando el desarrollo de arquitecturas de NF dedicadas descritas en la siguiente sección.

Mientras que en este trabajo nos enfocamos en NPE, que permite inferencia amortizada, como se demuestra en Brehmer et al. (2020) ecuaciones similares pueden formularse para mejorar la estimación de verosimilitud y razón de verosimilitud con información de gradientes.

## References

- [1] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, 2020. doi: 10.1051/0004-6361/201833910.
- [2] Simons Observatory Collaboration. The Simons Observatory: Science goals and forecasts. *JCAP*, 2019:056, 2019.
- [3] K. Abazajian et al. CMB-S4 Science Book. *arXiv e-prints*, 2016.
- [4] A. Schneider and R. Teyssier. A fast and accurate large-scale structure matter power spectrum calculator. *JCAP*, 2020:019, 2020.
- [5] BICEP/Keck Collaboration. Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season. *Phys. Rev. Lett.*, 127:151301, 2021.
- [6] A. Challinor and H. Peiris. The CMB bispectrum. *NewAR*, 55:223–239, 2011.
- [7] Planck Collaboration. Planck 2018 results. VIII. Gravitational lensing. *A&A*, 641:A8, 2020.
- [8] G. Efstathiou, C. Frenk, and N. Sugiyama. On the covariance matrix of CMB power spectrum estimates. *MNRAS*, 354:L72–L76, 2004.
- [9] J. Alsing and B. Wandelt. Fast likelihood-free cosmology with neural density estimators. *MNRAS*, 488:4443–4460, 2019.
- [10] N. Jeffrey, J. Alsing, and F. Lanusse. Dimensionality reduction for CMB data. *MNRAS*, 505:4626–4640, 2021.
- [11] G. Papamakarios, D. Sterratt, and I. Murray. Sequential Neural Likelihood: Fast Likelihood-Free Inference with Autoregressive Flows. *AISTATS*, 2019.