

Estudios de sesgo estadístico en cosmología

L. Herrera-Zamorano*, A. Hernández-Almada**

Abstract

Based on Lambda-Cold-Dark Matter (Λ CDM) model, and using the Observational Hubble Data (OHD) and Supernovae type Ia (SNIa) sample, we study a possible bias due their limited statistics in the free model parameters. We develop an algorithm with the purpose of having the parameter best fits of the Λ CDM model for several data size by selecting with Monte Carlo Methods subsamples that contains 50%, 75%, and 90% of the OHD and SNIa dataset. We estimate a bias on the model parameters up to 1%, that is negligible with respect to other sources of systematics.

Keywords: *computational physics; statistics; cosmology*

Facultad de Ingeniería
Universidad Autónoma de Querétaro
Centro Universitario: Cerro de las Campanas 76010
Querétaro, México
*herrera31@alumnos.uaq.mx
**ahalmada@uaq.mx

Fecha de recepción: 4 de abril de 2018 Fecha de aceptación: 28 de junio de 2018

Resumen

Con base en el modelo cosmológico Λ CDM y utilizando los datos observacionales del parámetro de Hubble (OHD), así como la muestra de Supernova tipo Ia (SNIa), estudiamos un posible sesgo debido a la cantidad de datos estadísticos en los parámetros libres del modelo Ω_{m0}, h . Asimismo, desarrollamos un algoritmo con el propósito de estimar los valores del modelo Λ CDM que mejor se ajuste tomando en cuenta diferentes tamaños de submuestras, con 50, 75 y 90% de los datos de OHD y SNIa. Como resultado principal, se estimó un sesgo inferior a 1% en los parámetros del modelo, lo cual es despreciable si se consideran otras posibles fuentes de incertidumbre sistemática de estos datos.

Palabras clave: *física computacional; estadística; cosmología*

Introducción

La cosmología es la rama de la física que estudia el origen y la evolución del universo. Actualmente, las observaciones provenientes de las Supernovas tipo Ia (SNIa), la radiación de fondo de microondas (CMB), las oscilaciones acústicas de bariones (BAO) y las mediciones directas del parámetro de Hubble (OHD) [1-3] indican que el universo se está expandiendo de manera acelerada y su entendimiento es uno de los grandes retos de la cosmología; además de ello, las mediciones de las velocidades de rotación de las galaxias indican un exceso de materia aún no observado. Dichos fenómenos están catalogados como energía oscura y materia oscura, respectivamente.

Por ejemplo, una prueba de la expansión del universo —es decir, que las galaxias y cúmulos de galaxias a gran escala se están alejando unos de otros— es el corrimiento al rojo de las líneas espectrales del hidrógeno y otros elementos químicos de las estrellas, como se ejemplifica en la figura 1, donde las longitudes de ondas de la radiación se vuelven más largas por dicha expansión.

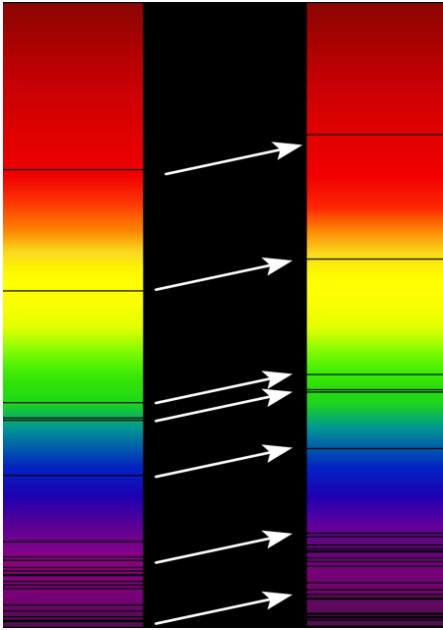


Figura 1. Líneas espectrales del hidrógeno en el laboratorio (izquierda) y las emitidas por las estrellas (derecha) mostrando un desplazamiento al rojo.

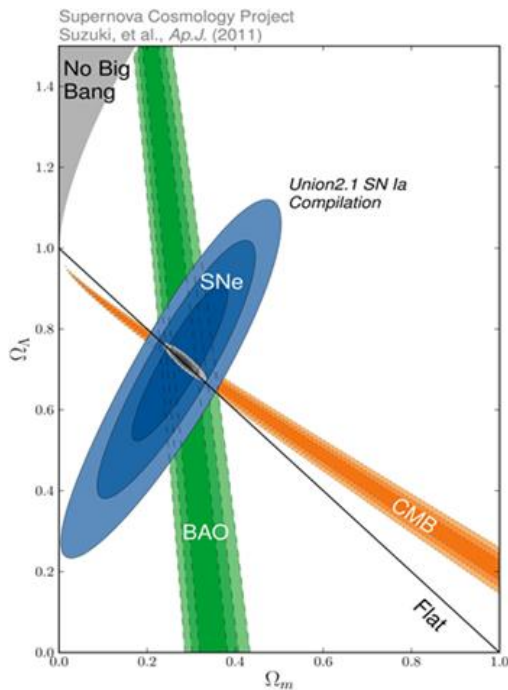


Figura 2. Contorno de los parámetros Ω_{m0} y Ω_{Λ} obtenidos por mediciones de SNe, CMB y BAO. En cada uno de los contornos, las regiones sombreadas de mayor a menor intensidad corresponden al 68, 95 y 99.7%, respectivamente, de nivel de confianza. La línea sólida representa la región donde el universo tiene curvatura plana. [5]

Hasta ahora, el modelo cosmológico más sencillo para entender la materia oscura y la energía oscura se conoce como Lambda-Cold Dark Matter (Λ CDM). En este modelo, se estima que 68.5 y 26.6% corresponden a estos dos componentes, respectivamente; la materia formada por protones y electrones (e.g. estrellas, seres humanos) corresponde al 5% restante y comúnmente es llamada materia bariónica. Además de estos parámetros, se estima un valor del parámetro de Hubble de alrededor de 67.31 km/s/Mpc a épocas actuales [4]. En la siguiente sección, se darán más detalles de este modelo y se explicará su relación con los datos observacionales.

En este trabajo, estudiamos un posible sesgo en los parámetros libres del modelo Λ CDM, debido al tamaño de los datos provenientes de las SNIa y de OHD, las cuales constan de 580 mediciones colectadas por el proyecto Union2.1 [5] y 51 puntos de OHD [6].

Modelo Λ CDM

La cosmología estudia el origen y evolución del universo. La ecuación que rige la dinámica es conocida como *la ecuación de Friedmann* y usualmente se expresa en términos del corrimiento al rojo z . Para el modelo Λ CDM, la ecuación de Friedmann está dada por:

$$E^2(z) = \frac{H^2(z)}{H_0^2} = \Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{\Lambda} \quad (1)$$

donde $H(z) = \frac{\dot{a}}{a}$ con a el factor de escala y $\dot{a} = \frac{da}{dt}$. La cantidad de H_0 está relacionada con el parámetro de Hubble h por la expresión $H_0 = 100h$. El parámetro Ω_{r0} corresponde a la cantidad de radiación en la época actual del universo y debido a que es del orden de 10^{-4} se considerará despreciable su contribución. Las expresiones Ω_{m0} y Ω_{Λ} representan la cantidad de materia (bariones y materia oscura) y energía oscura, respectivamente; en conjunto con h , forman los parámetros libres del modelo. En este trabajo, asumimos que el universo tiene una curvatura plana, la cual está soportada por diferentes observaciones cosmológicas (figura 2) provenientes de SNIa, CMB y BAO. De la ecuación

(1), en la época actual ($z = 0$ y $H(0) = H_0$) se cumple la siguiente relación:

$$\Omega_{m0} + \Omega_{\Lambda} = 1 \quad (2)$$

la cual nos reduce el número de parámetros libres en una unidad.

Una prueba latente de que el universo se encuentra en expansión acelerada está dada por las mediciones provenientes de SNIa del módulo de la distancia, el cual es

$$\mu(z) = 5 \log \left[\frac{d_L(z)}{\text{Mpc}} \right] + \mu_0 \quad (3)$$

donde μ_0 es un parámetro libre, y:

$$d_L(z) = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \quad (4)$$

es la distancia lumínica d_L , medida en Mpc.

En esta ecuación, $E(z)$ está dada por la ecuación 1 y c es la velocidad de la luz en el vacío.

Metodología

El objetivo principal de este trabajo es buscar un posible sesgo en los componentes de materia y energía oscura del modelo Λ CDM, debido a las limitadas mediciones cosmológicas proporcionadas por OHD y SNIa.

La muestra de datos de OHD consiste en 51 puntos del parámetro de Hubble como función del corrimiento al rojo, la cual es compilada en [6]. Para dichas mediciones, se toman en cuenta galaxias elípticas viejas que están formadas principalmente de estrellas con masas que comprenden entre 1 y 2 veces la masa del Sol. Cabe mencionar que en estas mediciones se utilizan galaxias elípticas que contienen una cantidad despreciable de materia

interestelar, como polvo o gas, lo cual no favorece a la formación de nuevas estrellas y hace las hace ideales para la medición directa del parámetro de Hubble.

La colección de las mediciones del módulo de las distancias proviene de 580 SNIa. Las SNIa son explosiones de estrellas que suceden en sistemas formados por dos estrellas, donde una de ellas es una enana blanca. Estas explosiones se conocen como *candelas estándar*.

Los valores de los parámetros libres del modelo (Ω_{m0}, h) que mejor se ajustan a estas mediciones, son los que minimizan la función χ^2 , dada por:

$$\chi^2 = \sum_i^N \left(\frac{y_i - M(z_i | \Omega_{m0}, h)}{e_i} \right)^2 \quad (5)$$

donde $y_i \pm e_i$ son las N observaciones y sus incertidumbres medidas al corrimiento al rojo z_i . Cuando se consideran los datos OHD, $M(z_i | \Omega_{m0}, h)$ es la estimación de $H(z)$ (ver ecuación 1), según el modelo Λ CDM. Por otro lado, cuando se toman en cuenta los datos de SNIa, $M(z_i | \Omega_{m0}, h, \mu_0)$ es el valor teórico del módulo de la distancia calculado por la ecuación 3.

El sesgo en los parámetros del modelo debido al tamaño de los datos lo definimos como:

$$b = \text{Valor correcto} - \text{Valor estimado}$$

donde el valor correcto representa los valores que mejor se ajustan a los datos cuando se toman en cuenta las muestras completas, es decir, 51 puntos para OHD y 580 para SNIa. El valor estimado corresponde a los valores de los parámetros cuando reducimos la muestra a cierto tamaño, como se detalla a continuación.

Caso I.

- Seleccionamos 10,000 subconjuntos, aleatoriamente, que contenga 90, 75 y 50% de los datos de OHD y SNIa, es decir, muestras de 46, 38 y 25 puntos para OHD y 522, 435 y 290 para SNIa. Para cada una de las muestras, el promedio de los valores de los parámetros obtenidos en los 10,000 ajustes

corresponderá al valor estimado de la definición de sesgo.

Caso II.

- Se ordenan los datos en forma ascendente según el corrimiento al rojo, z , y dividimos los datos de OHD y SNIa en dos subconjuntos seleccionando datos pares e impares, respectivamente. La motivación de esta situación consiste en que pueden faltar datos entre las mediciones actuales, por lo que nos da una idea de cómo puede fluctuar la estimación de los parámetros del modelo. Por ejemplo, en la muestra de los datos de OHD, son evidentes las regiones con déficit de puntos; esto podemos interpretarlo como falta de puntos entre los puntos existentes (ver figura 3).

Nótese que el sesgo de los parámetros también está relacionado con el error que se obtiene entre la estimación como referencia (cuando tomamos en cuenta todos los datos) y el valor obtenido en cada caso (cuando tomamos las submuestras), esto es:

$$\epsilon = \frac{b}{\text{Valor correcto}} \times 100 \%$$

En otras palabras, en este caso, el sesgo lo podemos interpretar como el error porcentual obtenido entre estas estimaciones.

Finalmente cabe mencionar que el ajuste se obtuvo desarrollando un algoritmo usando Minuit bajo el ambiente de Python.

Resultados

Después de realizar el procedimiento establecido en la sección anterior, los resultados obtenidos se muestran en las tablas I y II con base en los datos de OHD y SNIa, respectivamente. Los valores de referencia son los valores que contienen el 100% de los datos. En las figuras 3 y 4, se muestran los datos

OHD y SNIa, así como los correspondientes ajustes, respectivamente.

Los datos que se representan con el 90, 75 y 50% corresponden al caso I mencionado en la sección anterior se generan aleatoriamente.

En la figura 5, se muestran las distribuciones de los mejores ajustes de Ω_{m0} para los datos OHD y SNIa. La flecha que aparece del lado izquierdo representa el valor de Ω_{m0} obtenido cuando se toman en cuenta los datos pares de OHD, y la flecha del lado derecho denota el valor obtenido para los datos impares de SNIa. Estos valores corresponden a fluctuaciones con respecto del valor correcto del 5 y el 9%, respectivamente (ver tablas 1 y 2). La probabilidad de ocurrencia de obtener estos valores de Ω_{m0} , o menores (mayores) para OHD (SNIa), es el área bajo la distribución en dicha región, las cuales son de 0.12 (0.10). En otras palabras, se obtienen fluctuaciones de hasta 5 y 9%, con un nivel de confianza de 88 y 90%, respectivamente.

Conclusiones

En este trabajo, se estudió un sesgo en los parámetros Ω_{m0} y h contenidos en el modelo Λ CDM, debido al tamaño de las observaciones cosmológicas proporcionadas por las muestras de datos de OHD y SNIa. Este análisis reporta un sesgo menor a 1% y 0.1% cuando se redujeron las muestra de los datos de OHD y SNIa al 50%, respectivamente, las cuales son incertidumbres aceptables en el área de cosmología.

Con la finalidad de entender lo que sucede cuando se dejan huecos entre las mediciones existentes, se removieron puntos que correspondían a posiciones pares o impares de la lista de los datos. Los sesgos registrados fluctuaron hasta un 10% en el parámetro Ω_{m0} a un 90% de nivel de confianza. Para el parámetro h , se obtuvieron sesgos despreciables (<0.001%) a un 99.98% de nivel de confianza.

Datos OHD	Ω_{m0}	h	$\epsilon_{\Omega_{m0}}\%$	$\epsilon_h\%$
100%	0.251 ± 0.011	0.712 ± 0.007	-	-
90%	0.250 ± 0.003	0.713 ± 0.002	0.398	0.140
75%	0.250 ± 0.007	0.713 ± 0.004	0.398	0.140
50%	0.249 ± 0.011	0.715 ± 0.007	0.797	0.421
<i>Impar</i>	0.253 ± 0.014	0.712 ± 0.010	0.797	<0.001
<i>Par</i>	0.238 ± 0.014	0.719 ± 0.010	5.179	0.983

TABLA I. Valores de los parámetros libres usando los datos OHD. De izquierda a derecha: tamaño de la muestra, cantidad de materia (bariones + materia oscura), parámetro de Hubble, error porcentual de cantidad de materia y error porcentual del parámetro de Hubble. Los valores de los parámetros obtenidos con el 100% corresponden al valor de referencia.

Datos SNIa	Ω_{m0}	h	μ_0	$\epsilon_{\Omega_{m0}}\%$	$\epsilon_h\%$	$\epsilon_\mu\%$
100%	0.277 ± 0.019	0.732 ± 0.017	25.096 ± 0.052	-	-	-
90%	0.277 ± 0.006	0.732 ± 0.004	25.097 ± 0.003	<0.001	<0.001	0.003
75%	0.278 ± 0.011	0.732 ± 0.019	25.097 ± 0.008	<0.001	<0.001	0.003
50%	0.277 ± 0.019	0.732 ± 0.019	25.097 ± 0.012	<0.001	<0.001	0.003
<i>Impar</i>	0.302 ± 0.029	0.732 ± 0.017	25.106 ± 0.054	9.025	<0.001	0.040
<i>Par</i>	0.256 ± 0.026	0.732 ± 0.017	25.088 ± 0.054	7.581	<0.001	0.032

TABLA II. Valores de los parámetros libres usando los datos SNIa. De izquierda a derecha: tamaño de la muestra, cantidad de materia (bariones + materia oscura), parámetro de Hubble, error porcentual de cantidad de materia y error porcentual del parámetro de Hubble.

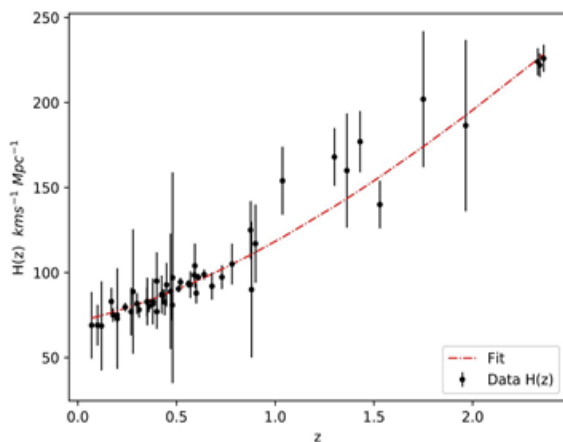


Figura 3. Ajuste de los datos OHD en el caso I.

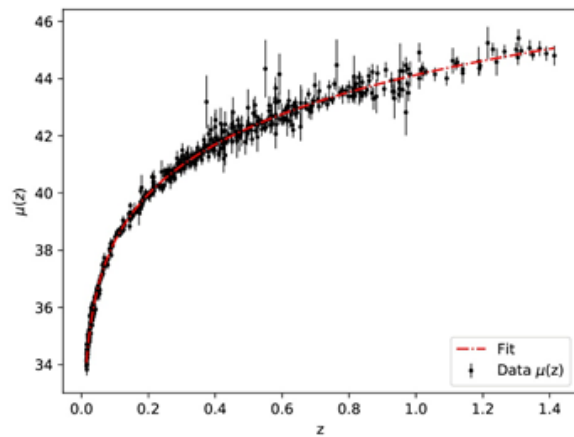


Figura 4. Ajuste de los datos SNIa en el caso I.

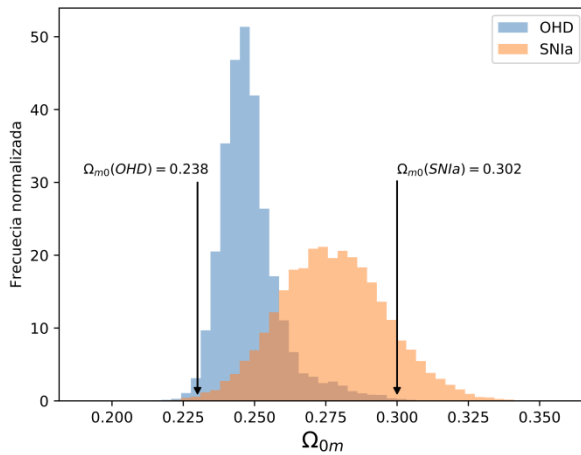


Figura 5. Distribución del parámetro Ω_{m0} usando datos OHD (azul) y SNIa (naranja). La fecha del lado izquierdo representa el valor obtenido cuando se toman en cuenta los datos pares de OHD con un error del 5%. La fecha a la derecha corresponde al valor medido de Ω_{m0} cuando se toman en cuenta los datos impares de SNIa con una desviación del 9% con respecto del valor correcto.

Referencias

- [1] Ryden B (2017). *Introduction to cosmology*, 2ª ed. EUA: Cambridge University Press.
- [2] Liddle A (2015). *An introduction to modern cosmology*, 3ª ed. Escocia: Blackford Hill.
- [3] Jain P (2015). *An introduction to astronomy and astrophysics*. India: CRC Press.
- [4] Ade P, Aghanim N *et al.* (2016). Planck 2015 results, XIII. Cosmological parameters, *A&A* 594, A(13).
- [5] Suzuki N *et al.* The Supernova Cosmology Project: the hubble space telescope cluster supernova survey: V. improving the dark-energy constraints above $z > 1$ and building an early-type-hosted supernova sample. *The Astrophysical Journal*, vol. 746(1):.
- [6] Magaña J, MH Amante, Garcia A y Motta V. The Cardassian expansion revisited: constraints from updated Hubble parameter measurements and type Ia supernova data. *Monthly Notices of the Royal Astronomical Society*, vol. 1, núm. 476, p. 103.

Agradecimientos

Los autores reconocen los recursos de computo, conocimientos técnicos y la asistencia proporcionados por “ABACUS” Laboratorio de Matemática Aplicada y Cómputo de Alto Rendimiento del CINVESTAV-IPN, Proyecto CONACYT-EDOMEX-2011-C01-165873.