

Cosmo Estadística

J. Alberto Vazquez^{1,2} and Luis E. Padilla²

¹*ICF-UNAM*

²*CINVESTAV**

(Dated: February 13, 2018)

En este artículo se presenta una breve introducción sobre las propiedades que describen el Universo observable. Tanto la teoría como las observaciones, y la conexión existente entre ellas dada por la estadística Bayesiana a través de códigos numéricos.

I. INTRODUCCIÓN

Algo sobre el universo

II. EL UNIVERSO OBSERVABLE

La descripción estándar de las propiedades dinámicas del Universo se basa en la teoría de la Relatividad General de Einstein, la cual construye una conexión entre la geometría del espacio-tiempo y su contenido de materia a través de cantidades fundamentales, como son el tensor de Einstein $G_{\mu\nu}$ y el tensor de Energía-momento $T_{\mu\nu}$

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}. \quad (1)$$

La parte izquierda de la igualdad describe la geometría del espacio-tiempo englobada en las componentes del tensor de Ricci $R_{\mu\nu}$, el escalar de Ricci R y la métrica del espacio-tiempo $g_{\mu\nu}$. Éstas en conjunto dirigen el comportamiento de los constituyentes de universo $T_{\mu\nu}$. Una modificación aceptable a las ecuaciones de Einstein es la introducción de una invariante bajo transformaciones de Loretz $\Lambda g_{\mu\nu}$ en las ecuaciones de campo

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (2)$$

donde Λ es la denominada constante cosmológica y su valor, de acuerdo con observaciones astrofísicas, es $\Lambda \sim 10^{-52}m^{-2}$; veremos un poco más acerca de esta componente en la siguiente sección. Las ecuaciones (2) son, en general, un conjunto de ecuaciones diferenciales parciales acopladas cuasi-lineares y de segundo orden para los diez elementos del tensor métrico. Sin embargo, éstas pueden presentar soluciones analíticas simples en la presencia de simetrías genéricas, como veremos mas adelante.

III. GEOMETRÍA DEL ESPACIO-TIEMPO

Para poder especificar la geometría del Universo, una suposición esencial es el *Principio Cosmológico*: cuando se promedia sobre escalas suficientemente grandes, a un tiempo particular, el Universo observable puede considerarse como *Homogéneo e Isotrópico* con gran precisión. Por ejemplo, en escalas superiores a 100 Mega-parsecs[?] la distribución de galaxias observadas sobre la esfera celeste justifica la suposición de isotropía. Además de esto, la uniformidad observada en la distribución de temperaturas vista a través de la radiación Cósmica en el Fondo de Microondas (CMB), una parte en 10^5 es la mejor evidencia observacional que tenemos en favor de un universo isotrópico. Por tanto, sí la isotropía se da por sentado y teniendo en cuenta que nuestra posición ocupada en el Universo no tiene preferencia sobre ninguna otra -conocido como el *Principio Copernicano*- por tanto homogeneidad se sigue de considerar isotropía en cada punto.

Homogeneidad establece que el Universo se observa por igual en cada punto del espacio.

Isotropía establece que el Universo se observa por igual en todas las direcciones.

Métrica Friedmann-Robertson-Walker

Vamos a incorporar los postulados de Homogeneidad e Isotropía en la descripción geométrica del Universo. El primero demanda que todos los puntos sobre una hiper-superficie espacial son equivalentes, mientras que el segundo asegura que todas las direcciones sobre dicha hiper-superficie son equivalentes para observadores fundamentales. Esto es, isotropía requiere que la distribución de galaxias a dos tiempos distintos sea similar, mientras que la homogeneidad requiere que el factor de magnificación deba ser independiente de la posición de la distribución. Por lo tanto, se sigue que la métrica deba tener la siguiente forma.

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = dt^2 - a^2(t)\gamma_{ij}dx^i dx^j, \quad (3)$$

donde $a(t)$ es una función - comúnmente denominada como *factor de escala normalizado* - cuya dependencia solo es temporal. A través de los años, observaciones cosmológicas han dado evidencia contundente que el Universo actual se encuentra en un período de expansión y por tanto el factor de escala satisface la condición $\dot{a}(t) > 0$. La métrica que describe este Universo homogéneo e isotrópico sobre las escalas más grandes se le conoce como métrica de Friedmann-Robertson-Walker (FRW). Por otro lado, el tensor γ_{ij} contiene solo funciones dependientes de las coordenadas. Además, el espacio-tiempo que cumple con las propiedades de homogeneidad e isotropía es aquel cuyas propiedades quedan especificadas con tan solo una constante - la *curvatura* K . Este tipo de espacios, denominados *maximalmente simétricos*, definen la parte espacial de la métrica como

$$\gamma_{ij}dx^i dx^j = \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2, \quad (4)$$

donde r es la coordenada radial y $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ es la métrica de la 2-esfera. Por tanto, sustituyendo (4) en (3), la métrica FRW se reescribe de la siguiente forma

$$ds^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right]. \quad (5)$$

La constante K describe la geometría de las secciones espaciales y dependiendo del valor $K > 0$, $K = 0$ o $K < 0$ representa Universos tipo cerrado \mathcal{S}^3 , plano \mathcal{R}^3 o abierto \mathcal{H}^3 respectivamente. Considerando la métrica $g_{\mu\nu}$ de la forma (5), los únicos términos de curvatura distintos de cero son

$$R_0^0 = -\frac{3\ddot{a}}{a}, \quad (6)$$

$$R_j^i = \left(\frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2} + \frac{2\kappa}{a^2} \right) \delta_j^i, \quad (7)$$

$$R = 6 \left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{\kappa}{a^2} \right). \quad (8)$$

Por tanto, la dinámica del espacio-tiempo en un universo homogéneo e isotrópico se reduce a determinar el factor de escala normalizado $a(t)$, el cual se calcula una vez que el contenido de materia queda especificado.

IV. INVENTARIO CÓSMICO

La geometría del espacio-tiempo queda determinada por las ecuaciones (6)-(8) y por tanto para resolver las ecuaciones de Einstein solo falta especificar el contenido de materia. Consideraremos a los *fluidos perfectos* ideales como la principal fuente del tensor de energía-momento

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu - pg^{\mu\nu}, \quad (9)$$

donde ρ es la densidad de energía y p es la presión isotrópica del fluido, ambas medidas por un observador en el sistema de referencia inercial local, en el cual el fluido se encuentra en reposo. En este sistema, donde la 4-velocidad del fluido es $u^\mu = (1, 0, 0, 0)$, el tensor de energía-momento se reduce a $T^\mu_\nu = \text{diag}(-\rho, g^{ii}p)$. A partir de éste, se puede obtener la conservación de la energía $\nabla_\mu T^\mu_\nu = 0$, la cual guía a la ecuación de continuidad

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \quad (10)$$

Para poder entender las propiedades dinámicas del Universo, se necesita primero tener en cuenta el contenido total de éste. Nuestro enfoque se basa en los fluidos perfectos barotrópicos que satisfacen una *ecuación de estado* dependiente del tiempo $w(a)$ de la forma

$$p = w(a)\rho, \quad (11)$$

para cualquier componente. Si además se considera w igual a una constante, la ecuación de continuidad puede ser integrada fácilmente para obtener

$$\rho \propto a^{-3(1+w)}. \quad (12)$$

Además, en un Universo dominado por la densidad de energía ρ la ecuación de Friedmann nos lleva a la evolución temporal del factor de escala

$$a \propto t^{2/3(1+w)}, \quad w \neq -1. \quad (13)$$

Esto es, la evolución del Universo cuya única componente es un fluido perfecto puede calcularse una vez especificada su ecuación de estado. El modelo de materia oscura fría con constante cosmológica (Λ CDM) se basa en cuatro ingredientes principales descritos por: la radiación (fotones y neutrinos sin masa), materia en forma de bariones y una contribuyente de materia oscura fría (DM) y la energía del vacío Λ . El comportamiento de cada uno de estos constituyentes se resume de la siguiente manera:

Radiación

Esta componente relativista domina durante las etapas tempranas del Universo y se caracteriza por tener una presión asociada $p_r = \rho_r/3$, y por tanto una ecuación de estado $w_r = 1/3$. Además la evolución de la densidad de energía y el factor de escala están dados por

$$\rho_r(t) \propto a^{-4}, \quad \text{and} \quad a(t) \propto t^{1/2}. \quad (14)$$

La densidad de energía de la radiación ρ_r total en el Universo puede describirse en términos de dos contribuyentes principales: fotones (γ) y neutrinos sin masa (ν):

$$\rho_r = \rho_\gamma(t) + \rho_\nu(t). \quad (15)$$

Fotones. Los fotones primordiales juegan un papel muy importante en la cosmología observacional ya que ellos constituyen la radiación cósmica en el fondo de microondas que observamos hoy en día?

Neutrinos sin masa. Los neutrinos son clasificados como leptones débilmente interactuantes los cuales existen en tres sabores: electrón ν_e , muon ν_μ y tau ν_τ ; todos ellos tienen asociado una

antipartícula. La densidad de energía de los neutrinos sin masa en el fondo cósmico (estimada a partir de argumentos teóricos) esta dada por

$$\rho_\nu = N_{\text{eff}} \times \frac{7}{8} \times \left(\frac{4}{11} \right)^{4/3} \rho_\gamma, \quad (16)$$

donde N_{eff} es el número efectivo de especies de neutrinos; nótese que considerando el modelo estándar $N_{\text{eff}} = 3.046$. Recientemente varios experimentos sugieren que los neutrinos efectivamente si poseen una masa aunque ésta muy pequeña, por ejemplo experimentos que detectan neutrinos atmosféricos o neutrinos solares. Aunado a esto, observaciones cosmológicas también pueden imponer límites en las masas de los neutrinos; algunos artículos de revisión sobre el tema [1].

Materia

Cualquier tipo de material con presión despreciable es usualmente considerada como *polvo*. Esta propiedad se representa a través de una ecuación de estado $w_m = 0$ y densidad de energía dada por

$$\rho_m(t) \propto a^{-3}, \quad a(t) \propto t^{2/3}. \quad (17)$$

El contenido total de materia del universo viene en diferentes formas, por ejemplo, además de la materia bariónica que comúnmente conocemos y observamos en el laboratorio, observaciones en la estructura del Universo a gran escala sugieren que la mayor cantidad del contenido galáctico es de la forma de algún tipo no-bariónico de materia, denominada como *materia oscura*. La densidad de materia total puede expresarse como la suma de las contribuciones barionicas (b) y de materia oscura (DM)

$$\rho_m(t) = \rho_b(t) + \rho_{DM}(t). \quad (18)$$

Bariones. Son las bases de la materia que fácilmente localizamos alrededor de nuestro universo (protones p^+ y neutrones n^0). Además si consideramos que el Universo como un todo tiene una carga total neutra, por tanto debe de haber un número igual de protones que de electrones (e^- - leptones cargados). Un artículo de revisión sobre la formación de elementos en el big bang puede encontrarse en [2].

Materia oscura. La existencia de materia oscura no-bariónica puede ser inferida a partir de manifestaciones gravitacionales a través de, por ejemplo, el comportamiento aplanado detectado en las curvas de rotación de diversas galaxias espirales, la razón masa-luz observada en cúmulos de galaxias y la distorsión de la luz proveniente de fuentes lejanas medida por lentes gravitacionales. Una discusión extensa sobre el estado actual de la materia oscura, incluyendo evidencia experimental y motivaciones teóricas es presentado por [3].

Vacío

Sí el término de la constante cosmológica se mueve hacia el lado derecho de las ecuaciones de Einstein, éste puede ser asociado a la densidad de energía del vacío, dado por

$$\rho_\Lambda \equiv \frac{\Lambda}{8\pi G}. \quad (19)$$

Extrapolando nuestras teorías hacia el futuro, donde la materia y radiación son diluidas debido a la expansión del Universo, la densidad de energía del vacío permanece siempre con el mismo valor constante ρ_Λ . Este tipo de energía puede ser modelada como un fluido perfecto con presión negativa igual a $p_\Lambda = -\rho_\Lambda$, la cual tiene una ecuación de estado asociada $w_\Lambda = -1$. La constante cosmológica se puede ver como la forma más simple de una componente de *energía oscura*, comúnmente considerada como el principal candidato para dar explicación de la aceleración actual del cosmos. En [1] encuentran que una ecuación de estado dependiente del tiempo $w_{DE}(a)$ provee una mejor descripción de las observaciones actuales.

Curvatura

Una componente importante de la densidad total del Universo es la contribución de la curvatura espacial, la cual puede ser considerada como cualquier otra componente si definimos una densidad de energía ficticia como

$$\rho_k \equiv -\frac{3k}{8\pi G}a^{-2}. \quad (20)$$

Por tanto esta densidad de energía queda descrita por una ecuación de estado $w_k = -1/3$ para la cual el factor de escala evoluciona proporcionalmente al tiempo cósmico $a \propto t$. El comportamiento general del término de curvatura se entiende fácilmente si la ecuación de Friedmann se reescribe (sin considerar la constante cosmológica) de la siguiente manera

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho + \rho_k). \quad (21)$$

Considerando que la densidad total del universo ρ es positiva, la expansión universal solamente puede detenerse si la geometría de este es cerrada $k > 0$ ($\rho_k < 0$) o de cualquier otra manera el universo se expandirá por siempre. Notamos de la ecuación (21) que para un parámetro de Hubble muy particular existe una densidad para la cual el Universo posee una curvatura plana ($k = 0$), a este valor de la densidad se le conoce como la *densidad crítica*, y esta dada por

$$\rho_c(t) = \frac{3H^2}{8\pi G}. \quad (22)$$

V. ECUACIONES COSMOLÓGICAS

Una vez que se ha especificado el contenido de materia toda la información sobre la dinámica del cosmos queda englobada en el factor de escalar. Por ejemplo, consideremos una expansión en serie de Taylor del factor de escala alrededor del tiempo actual t_0

$$a(t) = a[t_0 - (t_0 - t)] \quad (23)$$

$$= a(t_0) - (t_0 - t)\dot{a}(t_0) + \frac{1}{2}(t_0 - t)^2\ddot{a}(t_0) - \dots \quad (24)$$

$$= a(t_0)[1 - (t_0 - t)H(t_0) - \frac{1}{2}(t_0 - t)^2q(t_0)H^2(t_0) - \dots] \quad (25)$$

donde se ha definido la función de Hubble $H(t)$ y la función de desaceleración $q(t)$ como

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}, \quad q(t) \equiv -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)} \quad (26)$$

Los valores de estas funciones al día de hoy usualmente son denotados como $H_0 \equiv H(t_0)$ y $q_0 \equiv q(t_0)$. También podemos incluir la razón de la densidad de energía relativa a la densidad crítica del universo como el parámetro de densidad adimensional

$$\Omega_i(t) \equiv \frac{\rho_i(t)}{\rho_c(t)}, \quad (27)$$

donde el índice i etiqueta un tipo particular de componente como lo es la materia, radiación, etc..

Por lo tanto, las ecuaciones de Friedmann para un universo lleno de un conjunto de fluidos perfectos se pueden escribir de la siguiente manera

$$\left(\frac{H}{H_0}\right)^2 = \sum_i \Omega_{i,0} a^{-3(1+w_i)} + \Omega_{k,0} a^{-2}, \quad (28)$$

$$q = \frac{1}{2} \sum_i \Omega_i (1 + 3w_i). \quad (29)$$

VI. PARÁMETROS COSMOLÓGICOS

En la sección anterior hemos desarrollado las principales ecuaciones que describen la evolución del universo a gran escala. Notamos que la estructura total del CMB, espectro de potencias en la materia y distancias luminosas dependen fuertemente de las condiciones esenciales que surgen a partir de la época infacionaria, de la densidad de materia y de la razón de expansión del universo H_0 . En esta sección se da una breve introducción a las cantidad utilizadas en la descripción de las propiedades del cosmos.

A. Parámetros base

Este tipo de parámetros, comúnmente denominados como *parámetros estándar*, son considerados como las principales cantidades utilizadas en la descripción del universo. Estos, sin embargo, no son predichos por una teoría fundamental sino que tienen que ser modulados a mano de tal manera que la mejor combinación provee la mejor descripción de las observaciones astrofísicas y cosmológicas actuales. Variaciones de estos parámetros afectan la amplitud y forma del espectro así como las distancias luminosas medidas y por tanto guiarnos a diferentes realizaciones de nuestro universo. Existe una clasificación de estos parámetros dependiendo de las escalas donde se ven involucrados.

Parámetros de Fondo

La descripción actual del universo homogéneo puede darse simplemente en términos de los parámetros de densidad $\Omega_{0,i}$ y el parámetro de Hubble H_0 a través de la ecuación de Friedmann

$$H^2 = H_0^2 [(\Omega_{\gamma,0} + \Omega_{\nu}, 0) a^{-4} + (\Omega_{b,0} + \Omega_{DM,0}) a^{-3} + \Omega_{k,0} a^{-2} + \Omega_{\Lambda_0}] \quad (30)$$

Los términos que contienen contribuciones de radiación son medidos con gran precisión, por ejemplo, utilizando mediciones del satélite WMAP encuentran que $\Omega_{\gamma,0} = 2.469 \times 10^{-5} h^{-2}$ correspondiente a una temperatura $T_{cmb} = 2.725 K$. De manera muy similar para los neutrinos ν mientras estos mantengan un comportamiento relativista, ya que pueden relacionarse con la

densidad de los fotones a través de (16). Sin embargo, variaciones del resto de los parámetros deja distintos tipos de huellas en la historia de fondo y la evolución de las perturbaciones observadas a través de las distancias luminosas. La figura () muestra los valores teóricos del modulo de las distancias para tres modelos diferentes tomando en cuenta combinaciones de $\Omega_{m,0}$ y $\Omega_{\Lambda,0}$. Nótese que los objetos parecen estar más alejados (se observan más tenues) en un universo con constante cosmológica que en uno dominado solo por materia.

Por otro lado, la existencia de fuertes degeneraciones a partir de distintas combinaciones de parámetros tambien es notoria. En particular la *degeneración geométrica* involucrando Ω_m , Ω_Λ y el parámetro de curvatura $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$. Para reducir estas degeneraciones es común introducir una combinación de parámetros cosmológicos de tal manera que estos tengan efectos ortogonales en las mediciones, por ejemplo, una parametrización estándar se basa en las *densidades de energía físicas* de materia oscura $\Omega_{DM}h^2$ y bariones $\Omega_b h^2$.

Parámetros derivados

El conjunto estándar de parámetros, mencionado previamente, provee una descripción adecuada de los modelos cosmológicos comparandola con los datos observacionales. Sin embargo, esta parametrización no es unica y alguna otra pueden ser tan buena como ésta. Algunas parametrizaciones hacen uso del conocimiento de la física o de la sensibilidad de los detectores y por tanto pueden interpretarse de manera más natural. En general se pudo haber utilizado otro conjunto de parámetros para describir el universo, por ejemplo, algunos de ellos incluyen: la edad del universo, la temperatura actual del fondo de neutrinos, la época de igualdad materia-radiación, la época de reionización o alguna otra combinación. En el modelo LCDM con el fin de disminuir degeneraciones se utilizan como parámetros base las densidades de energía físicas $\Omega_{DM,0}h^2$ y $\Omega_{b,0}h^2$, y además se consideran como cantidades derivadas los parámetros de densidad $\Omega_{i,0}$ y la constante de Hubble H_0 .

VII. OBSERVACIONES

El rápido avance en el desarrollo de instrumentos observacionales altamente potentes ha guiado al surgimiento de la cosmología de alta precisión. En particular experimentos desarrollados para medir las anisotropías observadas en el CMB, distancias luminosas a partir de explosiones de Supernovas tipo IA y formación de estructura a gran escala. En este artículo nos enfocaremos en las supernovas.

A. Supernovas

Por más de casi dos décadas las observaciones en las supernovas han proporcionado evidencia decisiva que la expansión actual del universo se encuentra en una etapa de aceleramiento. En particular estudios de supernovas tipo IA consideradas como velas estándar: ellas tiene la propiedad de poseer la misma luminosidad intrínseca con alta precisión, hasta un factor de reescalamiento. Por tanto, la aceleración actual sugiere de la existencia de una componente exótica o la posibilidad de teorías alternativas las cuales producirían tal efecto. Algunos ejemplos de detección de supernovas tipo IA importante mencionar, incluyen:

- El Sloan Digital Sky Survey-II (SDSS) quien descubrió y midió curvas de luz de más de 327 supernovas tipo IA espectroscópicamente confirmadas, en el rango $0.05 < z < 0.35$.

- El programa The Equation of State: SupErNovae trace Cosmic Expansion (ESSENCE), encontró y analizó 60 supernovas en el intervalo de corrimiento al rojo $0.15 < z < 0.70$.
- La muestra del Supernova Legacy Survey (SNLS) tercer año de medición presenta 252 supernovas a altos corrimientos al rojo ($0.15 < z < 1.1$).
- El Telescopio Espacial Hubble (HST) descubrió 21 supernovas tipo IA a corrimientos al rojo $z \geq 1$.
- Recientemente la compilación de datos de todas las fuentes antes mencionadas es nombrada como Union 2.1

Las mediciones de supernovas pueden ser graficadas en el diagrama de Hubble con el modulo o luminosidad de la distancia en términos del corrimiento al rojo, para después localizar la mejor combinación de parámetros que describen las observaciones, como lo veremos en las siguientes secciones.

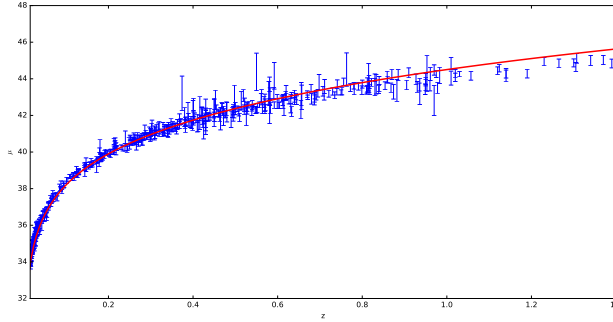


FIG. 1: A summary of possible geometries.

VIII. ESTADÍSTICA BAYESIANA

Durante la última década la información proveniente de un amplio rango de fuentes se ha incrementado sorprendentemente, por ejemplo de la radiación cósmica del fondo de microondas, de Supernovas y de la estructura a gran escala del universo. Este artículo nos enfocamos en traducir la información experimental/observacional en constricciones de nuestros modelos resumido en la estimación de los parámetros cosmológicos involucrados. Uno de los objetivos principales de la cosmología observacional es determinar la combinación de parámetros que mejor describen los datos observacionales. En esta sección introduciremos brevemente los conceptos matemáticos básicos necesarios para poder emplear el formalismo Bayesiano a la cosmología.

Comenzaremos definiendo el concepto más importante para cualquier procedimiento estadístico, esto es, el concepto de probabilidad. Si consideramos que x es una variable aleatoria relacionada con un evento en particular y $P(x)$ su probabilidad correspondiente, entonces para ambos casos se debe cumplir que

$$P(x) \geq 0, \quad (31a)$$

$$\int_{-\infty}^{\infty} dx P(x) = 1, \quad (31b)$$

y además para eventos mutuamente excluyentes,

$$P(x_1 \cup x_2) = P(x_1) + P(x_2), \quad \text{si } x_1 \cap x_2 = \emptyset. \quad (31c)$$

En general

$$P(x_1, x_2) = P(x_1)P(x_1|x_2) \quad (31d)$$

La última regla nos dice que la probabilidad de que x_1 y x_2 ocurran estará determinada por el producto de la probabilidad de que x_1 ocurra multiplicada por la probabilidad condicional de que x_2 ocurra dado que tal que x_1 ha ocurrido también.

Ahora bien, cualquier procedimiento estadístico consiste de tres ingredientes básicos que deben ser entendidos: los datos, el modelo y un método de estimación.

- *Los datos.* Son una medida de nuestras observaciones, los cuales denotaremos como D .
- *El modelo.* En general, un *modelo* Q es una colección de mediciones de probabilidades P . Las distribuciones P_θ son llamadas distribuciones del modelo, donde θ son aquellos parámetros que contiene Q . En la estadística Bayesiana se considera que no existe diferencia matemática formal entre parámetros y datos, por lo que para ambos casos se debe asociar una distribución de probabilidad.
- Finalmente nos encontramos con el estimador de P_0 . Un estimador para P_0 es la representación de nuestra “mejor creencia” de P dados los datos D , i.e. $P_0 = P_{\text{mejor}}$.

A. Teorema de Bayes, priors, posteriores y esas cosas

El tratamiento Bayesiano de un problema estadístico se centra en el teorema de Bayes. Este teorema es una consecuencia directa de los axiomas de la probabilidad (31). Podemos observar de (31c), sin pérdida de generalidad, que siempre es posible reescribir $P(x_2, x_1) = P(x_2)P(x_2|x_1)$. Luego, como es de esperarse que la relación $P(x_1, x_2) = P(x_2, x_1)$ se cumpla, obtenemos que

$$P(x_2|x_1) = \frac{P(x_2)P(x_1|x_2)}{P(x_1)}. \quad (32)$$

Este resultado es conocido como el *teorema de Bayes*. Por otro lado, ya que en la estadística Bayesiana no existe distinción formal entre datos y parámetros, podemos reescribir la ecuación anterior haciendo los cambios $x_1 \rightarrow D$ y $x_2 \rightarrow H$, como

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)} \quad (33)$$

En esta expresión agregamos el término θ para especificar que H depende de dichos parámetros. Por otro lado, notemos que hemos agregado una nueva cantidad H , que llamaremos nuestra “hipótesis”. Esta última corresponde al modelo Q que según nuestra creencia mejor se ajustaría a los datos, i.e. $H = Q_{\text{mejor}}$.

Notemos, de la expresión anterior, que contamos con cuatro nuevos entes que deben ser entendidos a la perfección. En primera instancia $P(\theta, H|D)$ es conocida como la distribución posterior de probabilidad (o simplemente posterior). Esta cantidad es básicamente el resultado principal a obtener ya que nos habla sobre la probabilidad de nuestro modelo (o parámetros del modelo). Usualmente, dicha cantidad se utiliza para acotar los parámetros del modelo. Luego tenemos los denominados “priors”, $P(\theta, H)$, los cuales son una distribución de probabilidad para nuestros parámetros, definidos a partir de nuestro propio conocimiento acerca del modelo. Generalmente, en el límite donde tenemos muchos datos, estos priors no son estadísticamente importantes, por lo que un prior típico para cada parámetro del modelo es uno plano (una distribución uniforme). A continuación tenemos lo que es conocido como el Likelihood $L(D; \theta) \equiv P(D|\theta, H)$ y es básicamente el elemento más importante al momento de hacer un estudio estadístico para la inferencia de los parámetros de nuestro modelo. Nos concentraremos en este elemento más adelante. Finalmente nos encontramos con la evidencia Bayesiana (o simplemente evidencia). Podemos notar que ésta actúa como un factor de normalización

$$P(D) = \int d\theta P(\theta, H) P(D|\theta, H). \quad (34)$$

Esta cantidad es usualmente ignorada cuando el espacio de parámetros de un único modelo es probado. Sin embargo resulta ser fundamental si lo que se quiere es hacer una comparación entre modelos. Para propósitos de este artículo, omitiremos esta cantidad $P(D)$.

Podemos ver que el teorema de Bayes tiene una implicación enorme respecto a una inferencia (de parámetros, por ejemplo) desde el punto de vista Bayesiano. En un escenario típico podemos tomar un conjunto de datos y esperar interpretarlos en términos de algún modelo. Sin embargo, lo que usualmente podemos hacer es lo opuesto, es decir, podemos tener un conjunto de datos y posteriormente confrontar un modelo tomando en cuenta la probabilidad que nuestro modelo se ajuste a los datos. De esta manera, como se puede ver de (33), el teorema de Bayes nos permite relacionar ambos escenarios, dándonos la posibilidad de conocer cual es el modelo (o parámetros del modelo) que mejor se ajusta a los datos.

B. Likelihood

Tomando el teorema de Bayes (33), pero por ahora ignorando la evidencia Bayesiana y considerando un prior uniforme, el cálculo de la probabilidad aposteriori esta determinado únicamente por maximizar el Likelihood. Por supuesto, el ignorar la evidencia ocasiona que no sea posible el poder dar una probabilidad absoluta de los parámetros de nuestro modelo, sin embargo, lo que sí podemos hacer es dar una probabilidad relativa definida como el cociente entre probabilidades. De esta manera, el Likelihood en un punto particular en el espacio de parámetros puede ser comparado con el que mejor se ajuste a las observaciones, L_{max} . Podemos decir entonces que un modelo es aceptable si el cociente de los Likelihoods

$$\Lambda = -2 \ln \left[\frac{L(D; \theta, H)}{L_{max}} \right] \quad (35)$$

es mayor que algún número dado.

Por otra parte, notemos que si nuestro posterior posee un único máximo global en θ_0 , siempre es posible hacer una expansión en serie de la forma

$$\ln L(D; H) = \ln L(D; H_0) + \frac{1}{2}(\theta_\alpha - \theta_{0\alpha}) \frac{\partial^2 \ln L}{\partial \theta_\alpha \partial \theta_\beta} (\theta_\beta - \theta_{0\beta}) + \dots \quad (36)$$

donde H_0 corresponde al modelo con los parámetros que mejor se ajustan a los datos y $\theta_{0\alpha}$ son los componentes del vector de parámetros θ_0 . De esta manera podemos reescribir nuestro Likelihood como

$$L(D; H) = L(D; H_0) \exp \left[-\frac{1}{2}(\theta_\alpha - \theta_{0\alpha}) H_{\alpha\beta} (\theta_\beta - \theta_{0\beta}) \right] \quad (37)$$

donde $H_{\alpha\beta} = \partial^2 \ln L / \partial \theta_\alpha \partial \theta_\beta$ es llamada la matriz Hessiana y controla si la estimación de θ_α y θ_β se encuentran correlacionadas. Si esta es diagonal, se dice que las estimaciones no están correlacionadas. Aquí es preciso remarcar que la aproximación es buena siempre que contemos con un único máximo global en nuestro posterior. Si por el contrario éste posee múltiples máximos locales, entonces la aproximación no serviría.

C. Chi-cuadrada

Como se comentó anteriormente, la forma precisa para encontrar el posterior de algún modelo en particular estará dado por la maximización del Likelihood. Notemos que en la aproximación gaussiana, dicho tarea es equivalente a minimizar la expresión

$$\chi^2 \equiv (\theta_\alpha - \theta_{0\alpha}) H_{\alpha\beta} (\theta_\beta - \theta_{0\beta}). \quad (38)$$

La cantidad χ^2 , usualmente llamada *chi-cuadrada*, se relaciona con el Likelihood Gaussiano vía $L = L_0^{-\chi^2/2}$. Así pues, podemos decir que el proceso de maximizar un Likelihood Gaussiano es equivalente con el de minimizar un chi-cuadrada.

D. Curvas de contorno y regiones de confianza

Una vez que se han obtenido los parámetros que mejor se ajustan a los datos, lo que nos gustaría saber es si hay regiones de confianza donde otros parámetros pueden ser considerados como buenos candidatos para nuestro modelo. Una elección natural sería considerar regiones donde χ^2 es menor que algún cierto valor dado. Para el caso de distribuciones Gaussianas, como es el caso de (37), dichas regiones corresponden con elipsoides.

La forma en la que estas regiones de confianza pueden ser calculadas es un poco técnica para los fines de este capítulo, sin embargo, para los lectores interesados se les recomienda revisar las referencias [ref].

E. Marginalización

Se puede dar el caso en el que al realizar un experimento sólo estemos interesados en analizar ciertos parámetros de nuestro modelo o existan parámetros extras considerados como ruido. Un ejemplo de estos parámetros de ruido podría ser aquellos correspondientes a efectos de calibración del experimento. Si este es el caso, lo que usualmente se hace es marginalizar sobre los parámetros que no son de interés mediante la expresión

$$P(\theta_1, \dots, \theta_j, H|D) = \int d\theta_{j+1} \dots d\theta_m P(\theta, H|D) \quad (39)$$

donde m es el número total de parámetros en nuestro modelo y $\theta_1, \dots, \theta_j$ corresponden a los parámetros en los que estamos interesados.

IX. MÉTODOS NUMÉRICOS

En un escenario típico casi nunca es posible hacer el cálculo del posterior de manera analítica. Para estas circunstancias existe toda una caja de herramientas computacionales que pueden ayudarnos a realizar esta tarea de forma numérica. En esta sección nos enfocaremos en revisar las herramientas más simples (pero no por eso menos eficientes) que están a nuestra disposición.

A. MCMC para la inferencia de parámetros

Una secuencia X_1, X_2, \dots de elementos aleatorios se dice que es una cadena de Markov si la distribución condicional de X_{n+1} dados X_1, \dots, X_n depende únicamente de X_n . Lo que es importante de este proceso es que se puede demostrar que éstas convergen a un estado estacionario donde varios elementos sucesivos de la cadena recaen. De esta manera, es posible estimar todas las cantidades de interés, por ejemplo promedios, varianza, etc..

El uso de este método en nuestro análisis Bayesiano es el de tratar de muestrear nuestro posterior con ayuda de estas cadenas. Para esto $P(\theta, H|D)$ es aproximada por un conjunto de funciones delta

$$p(\theta, H|D) \simeq \frac{1}{N} \sum_{i=1}^N \delta(\theta - \theta_i) \quad (40)$$

Así, el promedio de nuestro posterior puede ser calculado como

$$\langle \theta \rangle = \int d\theta \theta P(\theta, H|D) \simeq \frac{1}{N} \sum_{i=1}^N \theta_i \quad (41)$$

De esta manera, podemos promediar sobre funciones de nuestros parámetros mediante la expresión

$$\langle f(\theta) \rangle \simeq \frac{1}{N} \sum_{i=1}^N f(\theta_i) \quad (42)$$

1. Algoritmo “Metropolis Hasting”

Como se explicó anteriormente, en un proceso MCMC es necesario proponer un nuevo paso en nuestra cadena tomando sólo como información el paso presente. Sin embargo, como es de esperarse, necesitamos un criterio de aceptación (o rechazo) de este nuevo paso, dependiendo de si éste resulta ser mejor para nuestro modelo o no. Por otro lado, si resulta que el nuevo paso es mejor, nos gustaría no siempre aceptarlo, ya que de hacerlo, se podría dar el caso de que no estemos mapeando completamente el espacio de parámetros e ir convergiendo a lo que podría ser un máximo local de probabilidad para nuestro posterior. El algoritmo más simple que contiene toda esta información en su metodología es conocido como el *algoritmo de Metropolis Hasting* (MHA). Éste básicamente consiste en lo siguiente:

Comenzando con un punto inicial aleatorio θ_i , con probabilidad posterior asociada $p_i = P(\theta_i, H|D)$, se necesita proponer un candidato θ_c tomándolo de alguna *distribución propuesta* $q(\theta_i, \theta_c)$ que sea

simétrica, i.e. $q(\theta_i, \theta_c) = q(\theta_c, \theta_i)$. De esta manera, la probabilidad de aceptar un nuevo punto estaría dada por

$$p(\text{aceptar}) = \min \left[1, \frac{p_c}{p_i} \right] \quad (43)$$

De esta manera, el algoritmo completo puede ser descrito por los siguientes pasos:

1. Elegir un valor inicial aleatorio θ_i en el espacio de parámetros y calcular su correspondiente posterior.
2. Generar un nuevo candidato en el espacio de parámetros considerando alguna distribución propuesta y calcular su correspondiente posterior.
3. Aceptar (o no) el nuevo punto con ayuda del MHA.
4. Si el punto no es aceptado, repetir el punto anterior en la cadena.
5. Repetir los pasos del 2 al 4 hasta tener una cadena lo suficientemente larga.

B. Pruebas de convergencia

Es claro la necesidad de algún criterio que nos permita saber cuando nuestras cadenas han convergido al estado estacionario, de lo contrario, uno podría tener una gran pérdida de tiempo computacionalmente hablando. La forma más simple de hacer dicha prueba consiste en correr multiples cadenas y detectar a simple vista que todas converjan a la misma región de parámetros. Sin embargo, existen métodos más formales que nos permiten hacer dicha verificación. La prueba clásica de convergencia utilizada en Cosmología es el llamado *criterio de convergencia de Gelman-Rubin* (1992) [ref]. En éste se considera que una cadena ha convergido siempre que la cantidad

$$\hat{R} = \frac{\frac{N-1}{N}W + B(1 + \frac{1}{M})}{W}, \quad (44)$$

se aproxime a la unidad. Un criterio típico de convergencia es cuando $R < 1.03$. En esta expresión M es el número total de cadenas, N el número de puntos por cadena, B es la varianza relacionada entre cadenas y W es la varianza de cada cadena.

$$B = \frac{1}{M-1} \sum_{j=1}^M (\langle \theta^j \rangle - \langle \theta \rangle)^2, \quad W = \frac{1}{M(N-1)} \sum_{i=1}^N \sum_{j=1}^M (\theta_i^j - \langle \theta^j \rangle)^2 \quad (45)$$

C. Algunos detalles importantes

Sobre la distribución propuesta.- La elección de una función propuesta para generar el nuevo paso en nuestra cadena es de vital importancia para la eficiencia al momento de hacer la inferencia en cuestión. Se puede dar el caso de que si el paso propuesto en la función es muy pequeño, entonces nuestro código muestreará muy lentamente el espacio de parámetros, lo que ocasionaría que nuestro método se hiciera muy ineficiente. Por otro lado, si el paso de nuestra función propuesta es muy grande ocasionaría que este pudiera mapear de forma muy ineficiente el espacio de parámetros, haciendo saltos grandes y dejando el valor de nuestros parámetros que mejor se ajustan a los datos entre las regiones intermedias entre pasos.

X. AJUSTANDO UNA LÍNEA RECTA

En esta sección pondremos en uso las herramientas aprendidas anteriormente con uno de los ejemplos más simples que hay: ajustar una línea recta. Para esto, primero generaremos un conjunto de 25 datos a lo largo de la recta $y = 1 + x$. Digamos que estos datos fueron obtenidos a partir de una teoría en que la línea recta es el modelo que puede describir al sistema. Para dificultar un poco más las cosas, vamos a agregar a cada dato un ruido gaussiano con desviación estandar $\sigma = 0.3$, que puede ser debido a problemas en la medición. Con esto, nuestros datos lucirán como se pueden ver en la figura ??.

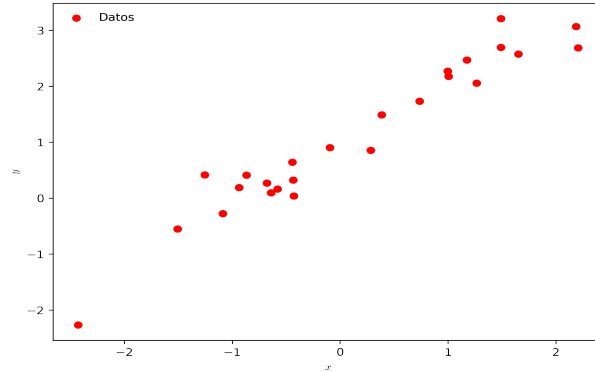


FIG. 2: (jav: Graficar barras de error). Datos a lo largo de la recta $y = 1 + x$ con un ruido gaussiano con desviación estandar $\sigma = 0.3$.

Ahora bien, supongamos que estamos interesados en conocer el valor de los parámetros a y m de nuestra recta $y = a + mx$, por lo que nos gustaría estimar dichos parámetros mediante un proceso de inferencia Bayesiana. También es posible estimar cual es la desviación estandar de nuestro conjunto de datos a lo largo de la recta, lo que nos agregaría un parámetro σ extra por estimar. Si consideramos que en principio no conocemos nada sobre el valor de nuestros parámetros libres, salvo las cotas límite en que éstos deberían estar, un buen prior para comenzar nuestra inferencia Bayesiana sería el considerar distribuciones planas. Para esto comenzaremos considerando los priors

$$a \propto U[0, 1.5], \quad m \propto U[0, 1.5] \quad \text{y} \quad \sigma \propto U[0, 0.5] \quad (46)$$

Considerando que existen valores de a , m y σ para los cuales los datos se fijan mejor (tal que el posterior es máximo global), entonces de (37) podemos escribir el Likelihood de nuestro sistema como

$$L(D; \text{recta}) \propto \exp \left[-\frac{(y_d - y)^2}{2\sigma} \right]. \quad (47)$$

Con esto ya podemos generar nuestras MCMC utilizando el MHA. Para ésto hicimos uso del módulo PyMC3 [ref] ya implementado en Python que nos permite emplear este método de forma más sencilla. Para el lector interesado, el código puede encontrarse en [ref]. En nuestro análisis corrimos un total de 6 cadenas con 10,000 pasos cada una. Nuestro resultado obtenido puede verse en la tabla I y la figura ??. Como podemos ver en el lado izquierdo de la figura, existen regiones para las cuales la frecuencia de eventos en nuestro muestreo se ve incrementada. De esta manera, podemos decir que dichas regiones poseen un posterior más probable para ajustarse a nuestros datos. De hecho, de la tabla I podemos observar que los valores reales para a y b parecen

	Promedio	Desv. Est.	Gelman-Rubin
a	0.957023	0.075513	0.99999
b	1.059392	0.062649	1.00029
σ	0.374450	0.060069	1.00027

TABLE I: Promedios obtenidos tras nuestra inferencia Bayesiana. Se calcula también el criterio de Gelman-Rubin para la convergencia de las cadenas.

encuentrarse dentro de una desviación estándar del valor promedio estimado para estos parámetros, mientras que el valor real de σ parece encontrarse ligeramente por arriba. Adicionalmente se obtuvo el valor del criterio de convergencia de Gelman-Rubin para cada variable con la intención de verificar que nuestros resultados obtenidos han convergido. Como podemos ver este número es muy cercano a la unidad, por lo que nuestro criterio de convergencia se cumple.

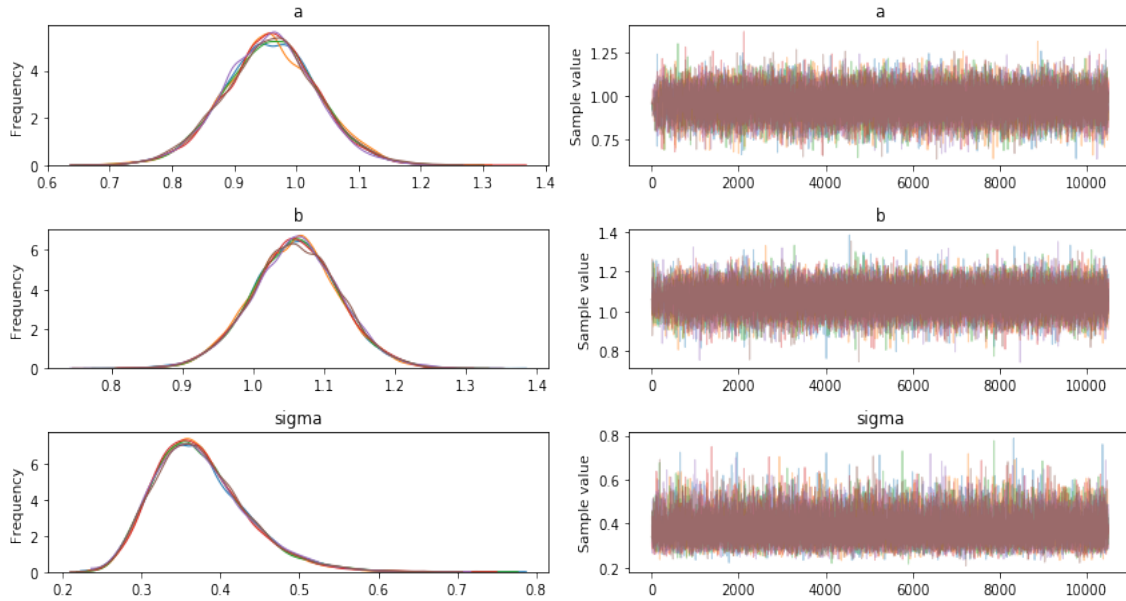


FIG. 3: Resultados obtenidos de nuestro muestreo de pasos en nuestras cadenas de Markov

Finalmente sólo resta mostrar las típicas regiones de confianza para nuestros parámetros. Las regiones que son usuales mostrar son aquellas que corresponden a un múltiplo de desviaciones estándar de los parámetros de interés. En la figura ?? mostramos las típicas regiones de confianza a 1-4 σ . Además hemos graficado en color rojo los valores reales de nuestros parámetros. Lo que podemos observar rápidamente es que ninguno de estos se encuentra alejado a más de 2 σ s del máximo estimado. Por supuesto, la razón de esta discrepancia es debido a que sólo contamos con 25 datos para hacer nuestra inferencia Bayesiana; de contar con más, nuestros resultados deberían ser más precisos y podríamos ser capaces de acotar en una región muy pequeña los posibles valores que nuestros parámetros podrían tener.

(jav: tomando los parametros que mejor ajustan, graficar la linea teoria en la misma figura 2, tambien falta decir que valor en el promedio \pm su barra de error para a,b, m)

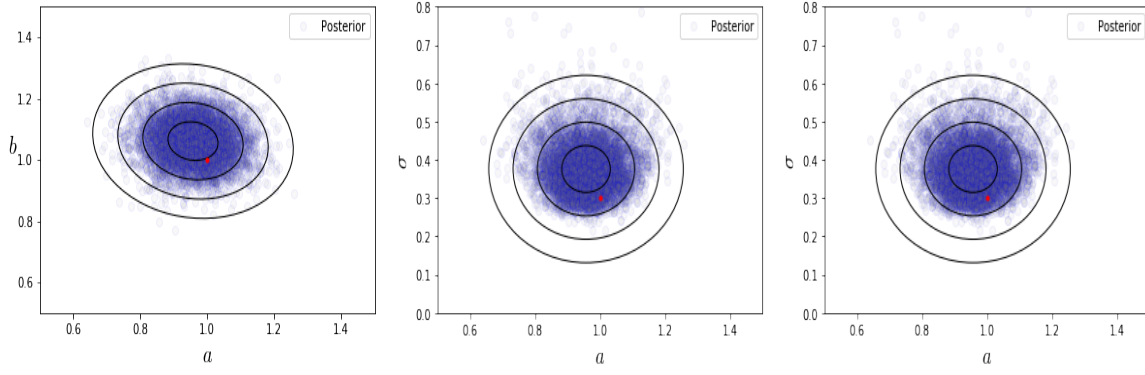


FIG. 4: Regiones de confianza en 2D para nuestros parámetros a , b y σ . Graficamos los contornos a 1-4 σ .