

Teoría de la información, mecánica estadística y teoría de matrices aleatorias

Fernando de Juan Sanz

1. Introducción

He hecho estas notas como introducción a la teoría de matrices aleatorias, para que el que quiera pueda ver con un poco más de detalle de donde salen las ideas de las que hablaré en el seminario. En realidad las notas son sobre la parte más matemática, sobre la definición de la teoría. Cómo se aplica y qué resultados da es de lo que hablaré en el propio seminario, pero aquí introduzco un poco la idea física para saber a donde vamos con las ideas de las que hablo en las notas.

La teoría de matrices aleatorias intenta responder una pregunta relativamente simple. Supongamos que intentamos modelizar un sistema cuántico muy complicado, tanto que ni siquiera sabemos por donde empezar. Nos resulta imposible en términos prácticos elaborar un modelo. ¿Existen aun así propiedades genéricas que compartan todos los sistemas suficientemente complejos, que nos permitan hacer alguna predicción física sin saber realmente nada sobre el sistema?

En términos simples, dado que los sistemas cuánticos están descritos por Hamiltonianos, la teoría de matrices aleatorias sirve para decidir si un Hamiltoniano aleatorio tiene o no propiedades genéricas. La idea es una analogía con mecánica estadística. En esta teoría las propiedades del sistema se obtienen como media sobre todas las configuraciones posibles. Sin embargo, en el límite termodinámico, por ejemplo en el colectivo canónico, hay muchísimos más estados con la energía media que con cualquier otra energía. En realidad no necesitamos hacer la media, pues todos los estados van a contribuir con la misma energía, basta con mirar cual es la energía de un microestado cualquiera. Pese a que hemos considerado microestados de todas las energías, en realidad en el límite termodinámico los estados con energías distintas de la media son un número despreciable. La energía media es una propiedad genérica de (casi) todos los microestados.

Del mismo modo, vamos a definir ahora en vez de todos los estados posibles, todos los Hamiltonianos posibles. Nuestro límite termodinámico es el límite de matrices infinitamente grandes. Lo que vamos a intentar es ver si, en este límite, una matriz cualquiera del colectivo de todas las posibles tiene propiedades genéricas. Si esto fuera así, podríamos adjudicar esta propiedad a nuestro Hamiltoniano complejo sin saber nada sobre el.

Por supuesto, esto requiere ser más cuidadoso. Para empezar, tenemos que definir cual es el colectivo de todos los Hamiltonianos posibles. En la discusión anterior hemos obviado que cada microestado tiene un peso estadístico (lo que no cambia la conclusión, aun así). ¿Cual sería el peso estadístico a asignar a cada matriz dentro del colectivo? ¿Dependen las propiedades genéricas de ésta elección? De esta cuestión se ocupa la primera parte de las notas.

Por otra parte, por muy complicado que sea un sistema, hay cosas que podemos saber sobre él. Hay simetrías generales como paridad, simetría rotacional, inversión temporal... que quizá podemos postular para nuestro sistema. ¿Cómo influyen pues las simetrías en la determinación del colectivo de matrices? De esto se ocupa la segunda parte de las notas.

Una vez definidos estos dos puntos, tendremos un colectivo de matrices sobre el que hacer estadística. A partir de ahí podemos empezar a hacer predicciones. De esto es de lo que me ocuparé en el seminario.

2. Información y entropía de Shannon

La entropía de Shannon fue introducida en 1949 como una medida de la incertidumbre que tenemos sobre el resultado de una variable aleatoria, en el contexto del problema de ingeniería de transmisión de datos. Como veremos, también puede entenderse como la cantidad de información que tiene el resultado de un proceso aleatorio conocida su distribución de probabilidad. Trabajaremos midiendo la información en bits, una medida intuitiva que hace los cálculos fácilmente interpretables.

Imaginemos el conjunto de todos los mensajes posibles de M bits, que son en total $N = 2^M$. La entropía de Shannon para una distribución de probabilidad sobre el conjunto de los mensajes es¹:

$$I = - \sum_i p_i \log_2 p_i \quad (1)$$

Donde p_i es la probabilidad de que ocurra el mensaje i . Un par de ejemplos ayudan a interpretar esta definición. Si uno de los mensajes tiene probabilidad 1 y el resto 0, la entropía es cero, es decir, la incertidumbre que tenemos sobre el mensaje es cero (el mensaje solo puede ser uno dado con probabilidad uno). Por el contrario, si todos los mensajes tienen una probabilidad constante $1/N = 2^{-M}$ de ocurrir, la entropía es:

$$I = - \sum_i^{2^M} 2^{-M} \log_2 2^{-M} = M \quad (2)$$

Es decir, la incertidumbre que tenemos es total, M bits, todo el mensaje. No es difícil demostrar que este es el valor máximo que puede alcanzar la entropía². Si normalizamos al número de bits, podemos entender la entropía como el “porcentaje de incertidumbre” sobre el mensaje. Alternativamente, podemos interpretar la entropía como la cantidad de información que obtenemos cuando tomamos un mensaje del conjunto de los mensajes, de acuerdo a la probabilidad p_i . En el primer caso, la información que obtenemos es cero, pues ya sabemos cual es el único mensaje que podemos obtener. En el segundo caso, la información que obtenemos es exactamente M bits, pues no sabíamos nada de antemano sobre el posible mensaje a obtener. Un último ejemplo: Si dos mensajes distintos tienen probabilidad $1/2$ y el resto 0, la entropía es 1 bit. La incertidumbre es sólo un bit, u obtenemos un mensaje o el otro. La información que obtenemos es 1 bit, el dato de saber cual de los dos hemos obtenido. (Por supuesto, la cantidad de bits no es necesariamente entera³, estos son ejemplos para facilitar la comprensión)

¿Qué es pues la información exactamente? Mi conclusión es que depende de como lo pensemos, lo único que está claramente definido es la entropía mediante 1. Podemos pensar que la información es:

- Conocida por nosotros una distribución de probabilidad en el espacio de mensajes, la información I_1 que obtenemos al recibir un mensaje concreto. Ésta coincide con la entropía de Shannon I .
- Si interpretamos la propia distribución de probabilidad como un “mensaje”, podríamos considerar la información I_2 como aquella que nos aporta conocer la distribución. Esto sería $(M - I)$, sería máxima para una distribución picada y cero para una distribución uniforme. Sería la información “contenida” en la distribución. Notemos que las dos informaciones suman M bits, conocida la información I_2 de una distribución, cualquier mensaje de M bits nos transmitirá $M - I_2 = I_1$ bits.

¹La función de entropía más general lleva una constante multiplicativa, que ha sido elegida de manera que el resultado se obtenga en bits.

²Nótese que $M = \log_2 2^M = \log_2 N$, y en este caso la entropía resulta ser el número de mensajes posibles. Esta “interpretación física” de Boltzmann de la entropía como el logaritmo del número de microestados sólo es válida si todos son igual de probables.

³Entendido el bit como unidad de entropía

De esta discusión nos quedaremos con la idea de que *tendremos la mayor incertidumbre sobre el mensaje (menor información I_2 sobre la distribución) cuando la entropía sea máxima*. En ausencia de otras condiciones sobre p_i , la entropía máxima se obtiene para $p_i = 1/N$

Más adelante será conveniente considerar una definición para una distribución continua, en cuyo caso:

$$I = - \int dx P(x) \ln P(x) \quad (3)$$

Un comentario para matemáticos, de acuerdo a [2], pag. 93, hay algunos problemas con este paso al continuo, de los que no nos vamos a preocupar aquí. Hasta aquí lo que necesitamos de teoría de la información.

3. Entropía de Gibbs y mecánica estadística

En mecánica estadística estamos acostumbrados a tratar con la entropía, $S = k_B \ln \Omega$, el logaritmo del número de microestados que corresponden a un macroestado dado. Existe otra definición de entropía más general dada por Gibbs, que mostramos a continuación, para una distribución de probabilidad general, que podemos entender como un determinado macroestado. Esta entropía de Gibbs es:

$$S_G = -k_B H_G = -k_B \int dx P(x) \ln P(x) \quad (4)$$

Donde x debe entenderse como un símbolo para todas las variables continuas de las que depende la distribución $P(x)$ y dx como el producto de todos los diferenciales correspondientes. (Nótese la correspondencia total con la entropía de Shannon). La entropía $S = k_B \ln \Omega$ se recupera de esta expresión en el caso de microestados igualmente probables, como vimos en la sección anterior, y solo en este caso. Las entropías de los estados de equilibrio de otros colectivos se obtienen de esta fórmula, que incluso define la entropía para ser empleada en termodinámica fuera del equilibrio.

La entropía de Gibbs y la información de Shannon son pues la misma cosa. La entropía de una distribución en mecánica estadística (no necesariamente la de equilibrio) es exactamente la incertidumbre sobre la distribución, o alternativamente de ella se puede obtener el “contenido de información” de la misma. Los estados de mayor entropía son los que menos información tienen. (Ver la discusión sobre I_1 e I_2 en la sección anterior.)

De hecho, no es difícil demostrar con la definición 4 la distribución $P(x)$ que minimiza la información (maximiza la entropía) satisfaciendo las ligaduras que definen un colectivo estadístico es la distribución de equilibrio de ese colectivo.

Esto simplemente significa que la condición de igual probabilidad a priori de los microestados es equivalente a la condición de tener incertidumbre máxima acerca del sistema. Es de hecho aún mejor, porque nos permite calcular la distribución para cualquier colectivo prácticamente sin cálculo, comparando con la manera usual de obtener las distribuciones para los colectivos no microcanónicos, separando el sistema en una reserva de calor y una parte pequeña en la que medimos. En algunos libros se asume el principio de máxima entropía en sustitución al postulado, y se considera el punto de partida de la mecánica estadística, ver [6],[2]. Este principio será también nuestra guía para la discusión sobre matrices aleatorias.

El principio también implica que, en general, en termodinámica fuera del equilibrio, donde el número de microestados no es un concepto útil, los sistemas tienden a establecerse en la configuración con menos información.

Veamos como podemos obtener las distribuciones de los colectivos habituales. Supongamos el espacio de fases de N partículas. La probabilidad de encontrarnos en el microestado (x, p) es simplemente $P(x, p) dx dp$.

Para maximizar la entropía siempre tendremos que incluir la ligadura de la normalización de P , que entra como un multiplicador de Lagrange constante. Así, si por ejemplo nos restringimos a la región del espacio de fases con una energía determinada, esto impone los límites de integración en la ligadura. La maximización nos da, como esperamos:

$$\delta I = \int dx dp \delta P(x, p) (1 + \ln P(x, p)) = 0 \implies P(x, p) = cte \quad (5)$$

Esta es la distribución para el colectivo microcanónico, en la que todos los microestados son igualmente probables. En vez de esto, podríamos considerar los microestados de todas las energías, pero imponer que la energía media sea una constante E_0 , lo que representamos por la ligadura:

$$\int E(x, p) P(x, p) dx dp = E_0 \quad (6)$$

El multiplicador de Lagrange entra en la minimización de esta forma:

$$\delta I = \int dx dp \delta P(x, p) (1 + \ln P(x, p) - \lambda E(x, p)) = 0 \implies P(x, p) = Z \exp(-\lambda E) \quad (7)$$

Y vemos que hemos obtenido la distribución del colectivo canónico (en una línea!), donde λ es el inverso de la temperatura, y está determinado por E_0 . Si consideramos un espacio de fases de infinitas partículas pero imponemos la ligadura de un número medio de partículas dado, obtendremos análogamente la distribución del colectivo gran canónico, donde el potencial químico entrará como el multiplicador de Lagrange asociado a la segunda ligadura.

Ésta es básicamente toda la discusión que necesitamos para el siguiente punto. Sólo nos queda mencionar un punto de importancia fundamental, el de la equivalencia de los colectivos. En física estadística sabemos que, en el llamado límite termodinámico, todos los colectivos resultan equivalentes. Esto significa, por ejemplo, que aunque el colectivo canónico tiene miembros con todas las energías, en el límite termodinámico los miembros con energías iguales a la energía media E_0 forman la inmensa mayoría del colectivo, por lo que éste resulta equivalente al microcanónico de energía E_0 . Una propiedad física en el límite termodinámico es indiferente del colectivo sobre el que se calcula como media. Dado que en la interpretación de máxima entropía un colectivo está definido por las ligaduras, podríamos expresar este último punto (abusando quizá un poco de la intuición...) como que las propiedades físicas en el límite termodinámico son independientes de las ligaduras escogidas.

4. Los colectivos de matrices aleatorias de información mínima

La teoría de matrices aleatorias puede considerarse como una “mecánica estadística” de todos los sistemas posibles (en vez de todos los estados posibles de un sistema). Las propiedades del sistema a describir se obtienen como la media sobre el colectivo de todos los Hamiltonianos posibles.

De momento sólo vamos con las matemáticas de las matrices aleatorias. Sabemos (o sabremos en el seminario) que la física (en particular, las simetrías) impone ciertos requisitos sobre las matrices a considerar (matrices reales y simétricas, hermíticas, o de cuaterniones reales, dependiendo de si hay o no inversión temporal o simetría de rotaciones, ver [4]). En adelante consideraremos el espacio de las matrices que tienen las características de simetría apropiadas, que no serán relevantes para la discusión. Ahora nos interesa definir la distribución de probabilidad para estas matrices en su colectivo. Comparando con mecánica estadística, las matrices son los microestados, y como no sabemos nada sobre el sistema más que su simetría, pues no sabemos que distribución de probabilidad asignar a las matrices dentro del colectivo. (¿equiprobables, quizá, como en microcanónico?, pero, ¿cual es la “energía”?)

Recordemos que la motivación de RMT es describir un Hamiltoniano complicado del que se espera que no tenga propiedades específicas, que pueda ser cualquiera dentro de los límites marcados por la simetría. Como después de tanta introducción no puede ser de otra forma, tomaremos la distribución de probabilidad para las matrices en el colectivo como aquella que maximiza la entropía de Shannon de la distribución, de modo que nuestra incertidumbre sobre el Hamiltoniano (sobre el mensaje en cuestión, volviendo al principio) sea máxima. Esta es la mejor manera de no asumir inadvertidamente un conocimiento sobre el Hamiltoniano que no tenemos.

Si denotamos el colectivo de matrices por \mathcal{H} , y la probabilidad de encontrar un elemento $H \in \mathcal{H}$ por $P(H)d[H]$, donde $P(H)$ es la distribución de probabilidad y $d[H]$ la medida de intergración (producto de los diferenciales necesarios), podemos definir la entropía I de una distribución de acuerdo a 3, y maximizándola sujeta a ligaduras obtenemos:

$$\delta I = \int d[H] \delta P(x) \left(1 + \ln P(H) - \sum_i \lambda_i f_i(H) \right) = 0 \quad (8)$$

Y la distribución de probabilidad queda⁴:

$$P(H) = A \exp \left(\sum_i \lambda_i f_i(H) \right) \quad (9)$$

Lo primero que observamos aquí es que la razón de que tengamos una exponencial en $P(H)$ aparece naturalmente. En cualquier colectivo, tendremos una exponencial que viene del logaritmo de la I de Shannon. (En el seminario veremos que se suele tomar la exponencial sin muchas explicaciones) Debido a la propia definición de los colectivos con los que tratamos, éstos deben ser invariantes bajo cambios de base (esto es consecuencia de la simetría), por lo que la probabilidad $P(H)$ debe ser también invariante. Esto se puede conseguir imponiendo solamente ligaduras que dependan de la traza de una función cualquiera de H^5 .

Lo único que nos queda es definir cuales son las ligaduras que imponemos, y tendremos un colectivo con el que trabajar. Por diversas razones que comentaré, habitualmente se toma la llamada distribución Gaussiana:

$$P(H) = A \exp(-\text{tr} H^2) \quad (10)$$

Que da nombre a los colectivos Gaussianos GOE, GUE y GSE que veremos. Es fácil comprobar que estos colectivos Gaussianos, obtenidos de una forma un tanto arbitraria imponiendo la independencia estadística de los elementos de cada matriz, se recuperan con la ligadura $\langle \text{tr} H^2 \rangle = k$. Pero este método de información mínima permite definir muchos otros colectivos sujetos a distintas ligaduras. También tenemos ya respuesta a una pregunta del principio: Matrices equiprobables es de hecho una buena distribución, la que no tiene ligaduras, sólo que habría que imponer una región finita del espacio de matrices sobre la que integrar, equivalente a la región de energía constante. Podríamos quizá tomar todas las matrices de una determinada traza.

Esta manera de ver las cosas puede arrojar luz sobre la conjetura de que las llamadas propiedades de fluctuación de las matrices aleatorias, calculadas habitualmente para los colectivos Gaussianos (en los que los cálculos son más fáciles), son de hecho independientes de la distribución de probabilidad en el límite $N \rightarrow \infty$. Hay una explicación para esto en términos de la analogía del gas de Coulomb (que daré en el

⁴Estrictamente, la constante de normalización A provendría de otro multiplicador de Lagrange que no incluimos asociado a la propia condición de normalización

⁵Pues si desarrollamos esta función en serie de Taylor para H obtendremos una serie potencias cuya traza será invariante. De hecho se le puede dar la vuelta a esta idea, pues, según Metha, cualquier invariante de una matriz $N \times N$ puede expresarse en términos de las trazas de las N primeras potencias de la matriz.

seminario), pero si entendemos este límite como el análogo al límite termodinámico, es inmediato conjeturar que en este límite las propiedades físicas no dependen de las ligaduras impuestas, exactamente igual que en el caso de mecánica estadística, los colectivos se vuelven equivalentes. No existe una demostración de esta conjetura, pero existe amplia evidencia numérica [1],[4] de que, en efecto, las propiedades de fluctuación son las mismas. La condición sobre la traza o la independencia estadística aparecerían así como una elección conveniente para los cálculos, sin tener un significado especial.

5. Bibliografía

Aquí incluyo algunos libros y artículos que se citan durante la exposición, o bien que resultan interesantes para la discusión, y comento por encima porque me han sido útiles o mi opinión acerca de ellos. No es para nada una bibliografía exhaustiva sobre el tema. ¿Se puede citar a Wikipedia en los seminarios de formación? Aunque está floja en algunas cosas, otros artículos de introducción están bien.

Referencias

- [1] M.L. Mehta, *Random matrices*. Academic Press (1991) ;
Suele ser el libro de referencia para las matemáticas de matrices aleatorias, es bastante extenso. La teoría de la información la menciona de pasada.
- [2] H.P. Yockey, *Information theory and molecular biology*. Cambridge University Press (1992) ;
Yo lo he usado para la parte de teoría de la información, tiene una buena discusión sobre la entropía de Shannon, la entropía física y el principio de entropía máxima, la información...
- [3] C.W.J. Beenakker, *Random matrix theory of quantum transport*. Rev. Mod. Phys. Vol 69 No. 3 (1997) ;
Pues eso, la parte de la que no voy a hablar, la aplicación de RMT a transporte en mesoscópicos, fluctuaciones de la conductancia...
- [4] T. Guhr, A. Müller-Groeling, H.A. Weidenmüller. *Random Matrix Theories in Quantum Physics: Common concepts*. Phys.Rept. 299 (1998) 189-425;
Un review muy extenso sobre RMT en todos sus aspectos, las matemáticas, los distintos campos de aplicación... muchas referencias. Es con el que primero trabajé, tiene bastante introducción, y explica muchas cosas. La página web de Guhr (<http://www.matfys.lth.se/Thomas.Guhr/research.html>) tiene un resumen interesante de cosas con RMT.
- [5] R. Blümel, W. P. Riehnardt. *Chaos in atomic physics*. Cambridge University Press, 1997. ;
Un librito sobre caos cuántico relativamente fácil en el que introducen las matrices aleatorias sin muchas matemáticas, también interesante. Tiene ejemplos de caos en física atómica.
- [6] E.T. Jaynes, *Gibbs vs. Boltzmann entropies*. American Journal of Physics, Vol. 33, 5, 391-398 (1965). ;
Un artículo cortito de mecánica estadística que defiende la entropía de Gibbs como cantidad fundamental, de un señor muy enfadado porque todo el mundo está confundido en el tema, muy interesante. El autor tiene un libro de mecánica estadística citado en el artículo. El artículo está gratis en http://en.wikipedia.org/wiki/Gibbs_entropy
- [7] O. Bohigas, M. J. Giannoni, C. Schmit, Phys. Rev. Lett. 52 (1984) ;
El artículo de la conjetura BGS.
- [8] P. Leboeuf, A. G. Monastera, O. Bohigas. *The Riemannium*. Regular and Chaotic Dynamics Vol. 6 (2001) N. 2, 205-210. ;
Una aplicación muy interesante de RTM que probablemente no mencione, las fluctuaciones de los ceros de la función de Riemann son las del GUE.
- [9] M.-J. Giannoni, A. Voros, J. Zinn-Justin. Les Houches. *Chaos and Quantum Physics*. North-Holland (1991) ;

Tiene un par de capítulos de matrices aleatorias que recuerdo que estan bien, y tiene muchas otras cosas de caos cuántico, no relacionadas con RMT que también son interesantes.

- [10] Thomas Guhr. *Random matrix theory in physics*
<http://www.matfys.lth.se/Thomas.Guhr/encrmt.pdf> ;

Un resumen muy bueno sobre el tema con mucha información, de uno de los autores que más ha trabajado en el tema.