Guía 2: Maximización de entropía y ejercicio 18b

Facundo Rost

Física Teórica 3 - 1er cuatrimestre de 2020

1. Ensambles genéricos

Consideremos un conjunto M, llamado espacio muestral, formado por posibles estados $r \in M$ de nuestro sistema físico, donde cada estado r posee una probabilidad p_r de ser el estado del sistema. Por simplicidad, consideremos que M es discreto ($\#(M) = \Omega \in \mathbb{N}$ es la multiplicidad del sistema y es un número natural), por lo que:

$$\sum_{r \in M} p_r = 1$$

Definimos la entropía estadística o de Shannon del sistema como:

$$S_{\text{est}} = -\sum_{r \in M} p_r \ln(p_r) = \langle -\ln(p_r) \rangle = \langle \ln\left(\frac{1}{p_r}\right) \rangle \tag{1}$$

Ver https://en.wikipedia.org/wiki/Entropy_(information_theory) para más detalle.

Notar que en general vale: $S_{\text{est}} \geq 0$, y ocurre que $S_{\text{est}} = 0 \iff \exists\, r \in M / p_r = 1 \iff p_{r'} = \delta_{r'r}$.

Si maximizamos $S_{\text{est}}(p_1, \cdots, p_r, \cdots, p_\Omega)$ sujetos al vínculo $\sum_{r \in M} p_r = 1$, obtenemos (usando el método de los multiplicadores de Lagrange): [1]

$$W(p_r) = S_{\text{est}}(p_r) - \nu \left(\sum_{r \in M} p_r - 1 \right)$$

Luego, si maximizamos esta función respecto de las variables p_r, obtenemos (para todo $r \in M$):

[1] Probablemente hayan visto en Matemática 1 que para maximizar una función $f(x_1, \cdots, x_\Omega)$ sujeta a los n vínculos $g_k(x_1, \cdots, x_\Omega) = 0$ con $1 \leq k \leq n$, el método de los multiplicadores de Lagrange consiste en plantear que $\nabla f(\vec{x}) = \sum_{k=1}^{n} \lambda_k \nabla g_k(\vec{x})$, donde λ_k con $1 \leq k \leq n$ son los multiplicadores de Lagrange. Es decir, que el gradiente de f es perpendicular a la superficie formada por los vínculos $S = \{g_k(\vec{x}) = 0 \ \forall\ 1 \leq k \leq n\} \subseteq \mathbb{R}^\Omega$. No es difícil ver que plantear ésto es equivalente a plantear: $\nabla W(\vec{x}) \equiv \nabla \left(f(\vec{x}) - \sum_{k=1}^{n} \lambda_k g_k(\vec{x}) \right) = \vec{0}$ o equivalentemente que $\frac{\partial W(\vec{x})}{\partial x_r} = \frac{\partial}{\partial x_r} \left(f(\vec{x}) - \sum_{k=1}^{n} \lambda_k g_k(\vec{x}) \right) = 0$ para todo $1 \leq r \leq \Omega$, donde se definió la función auxiliar $W(\vec{x}) \equiv f(\vec{x}) - \sum_{k=1}^{n} \lambda_k g_k(\vec{x})$.

$$0 = \frac{\partial W}{\partial p_r} = -\ln(p_r) - 1 - \nu \iff p_r = e^{-1-\nu} \equiv A$$

Y como se debe verificar el vínculo:

$$\sum_{r\in M} p_r = A\Omega = p_r\Omega = 1 \implies p_r = \frac{1}{\Omega}$$

Vemos que el máximo de entropía se da cuando todos los estados son equiprobables, que es la distribución de probabilidad en la cual la incertidumbre antes de conocer el estado es máxima, y la información que se gana al conocer el estado es máxima.

En general, cuanto mayor sea S_{est}, mayor es la incertidumbre antes de conocer el estado, y mayor información se gana al conocer el estado.

Maximizar la entropía se corresponde luego con maximizar la incertidumbre antes de conocer el estado.

Ahora planteemos otra situación: Maximicemos la entropía dado el vínculo $\sum_{r\in M} p_r = 1$, y también dados los vínculos $\langle A_k\rangle = a_k$ con $1 \leq k \leq n$ (hay n valores medios fijos), donde a_k es un número real fijo y A_k es una cierta magnitud del sistema cuyo valor depende del estado del sistema: Si el estado es r entonces A_k vale $A_{k,r}$ (O sea, $A_k : M \to \mathbb{R}$). Cada uno de los n vínculos $\langle A_k\rangle = a_k$ los podemos reescribir como: $-a_k + \langle A_k\rangle = -a_k + \displaystyle\sum_{r\in M} p_r A_{k,r} = 0$.

La distribución de probabilidad sobre el espacio muestral M que se obtenga a partir de maximizar la entropía dados los anteriores vínculos de valores medios $\langle A_k\rangle = a_k$ con $1 \leq k \leq n$ (además del vínculo $\sum_{r\in M} p_r = 1$), nos determina lo que denominamos un ensamble estadístico con vínculos $\langle A_k\rangle = a_k$ con $1 \leq k \leq n$.

Entonces, si maximizamos S_{est} sujeto a los anteriores vínculos usando multiplicadores de Lagrange, debemos maximizar la función:

$$W(p_r) = S_{\text{est}}(p_r) - \nu\left(\sum_{r\in M} p_r - 1\right) - \sum_{k=1}^{n} \lambda_k \left(-a_k + \sum_{r\in M} p_r A_{k,r}\right)$$

Si maximizamos esta función respecto de las variables p_r, obtenemos (para todo $r \in M$):

$$0 = \frac{\partial W}{\partial p_r} = -\ln(p_r) - 1 - \nu - \sum_{k=1}^{n} \lambda_k A_{k,r} \iff p_r = A\exp\left(-\sum_{k=1}^{n} \lambda_k A_{k,r}\right)$$

con $A \equiv e^{-1-\nu}$. Y como se debe verificar el vínculo:

$$1 = \sum_{r\in M} p_r = A\sum_{r\in M} \exp\left(-\sum_{k=1}^{n} \lambda_k A_{k,r}\right) \equiv AZ = 1 \implies A = \frac{1}{Z}$$

donde:

$$Z \equiv \sum_{r\in M} \exp\left(-\sum_{k=1}^{n} \lambda_k A_{k,r}\right) \tag{2}$$

se define como la función de partición del sistema. Luego, las probabilidades del sistema que maximizan la entropía dados los vínculos es:

$$p_r = \frac{1}{Z}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right) = \frac{\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right)}{\sum_{r'\in M}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r'}\right)} \tag{3}$$

Si planteamos los otros vínculos:

$$a_k = \langle A_k\rangle = \sum_{r\in M} p_r A_{k,r} = \frac{1}{Z}\sum_{r\in M} A_{k,r}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right) = -\frac{1}{Z}\left.\frac{\partial Z}{\partial \lambda_k}\right|_{\lambda_j, j\neq k} = -\left.\frac{\partial(\ln Z)}{\partial\lambda_k}\right|_{\lambda_j, j\neq k}$$

En definitiva obtenemos (recordar que por definición: $d(f(x_1,\cdots,x_n)) \equiv \sum_{k=1}^{n}\left.\frac{\partial f}{\partial x_k}\right|_{x_j, j\neq k} dx_k$):

$$a_k = -\left.\frac{\partial(\ln Z)}{\partial\lambda_k}\right|_{\lambda_j, j\neq k} \iff d(\ln Z) = -\sum_{k=1}^{n} a_k d\lambda_k \tag{4}$$

entendiendo a $\ln Z(\vec{\lambda})$ como una función de los multiplicadores de Lagrange $\vec{\lambda} = (\lambda_1,\cdots,\lambda_n)$.

Y si calculamos la entropía estadística que corresponde al estado que la maximiza (dados los vínculos):

$$\begin{aligned} S_{\text{est}}^{\max} &= S_{\text{est}}\left(p_r = \frac{1}{Z}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right)\right) = -\sum_{r\in M}\frac{1}{Z}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right)\ln\left(\frac{1}{Z}\exp\left(-\sum_{k=1}^{n}\lambda_k A_{k,r}\right)\right) \\ &= \sum_{r\in M} p_r \ln Z + \sum_{r\in M} p_r \sum_{k=1}^{n}\lambda_k A_{k,r} = \left(\sum_{r\in M} p_r\right)\ln Z + \sum_{k=1}^{n}\lambda_k\left(\sum_{r\in M} p_r A_{k,r}\right) = \ln Z + \sum_{k=1}^{n}\lambda_k a_k \end{aligned}$$

En definitiva:

$$S_{\text{est}}^{\max} = \ln Z + \sum_{k=1}^{n}\lambda_k a_k \tag{5}$$

y en consecuencia:

$$dS_{\text{est}}^{\max} = d(\ln Z) + \sum_{k=1}^{n}(a_k d\lambda_k + \lambda_k da_k) = \sum_{k=1}^{n}(-a_k d\lambda_k + a_k d\lambda_k + \lambda_k da_k) = \sum_{k=1}^{n}\lambda_k da_k$$

o equivalentemente:

$$\lambda_k = \left.\frac{\partial S_{\text{est}}^{\max}}{\partial a_k}\right|_{a_j, j\neq k} \iff dS_{\text{est}}^{\max} = \sum_{k=1}^{n}\lambda_k da_k \tag{6}$$

Notemos que $S_{\text{est}}^{\max} = S_{\text{est}}^{\max}(\vec{a})$ es la transformada total de Legendre de $\ln Z$, y viceversa.

1.1. Ensambles en continuo

En caso de que M no sea discreto, la fórmula de la entropía estadística se obtiene reemplazando la sumatoria $\sum_{r\in M}$ de la expresión 1 por una integral $\int_M d^L R$, si especificar el estado $r \in M$ consiste en especificar el valor de L variables aleatorias $\vec{R} = (R_1, \cdots, R_L)$ que toman valores continuos:

$$S_{\text{est}} = -\int_M f(\vec{R}) \ln(f(\vec{R})) d^L R$$

y se denomina entropía diferencial o continua (para más detalles ver https://en.wikipedia.org/wiki/Differential_entropy).

Un pequeño comentario al respecto: Como se debe verificar $\int_M f(\vec{R}) d^L R = 1$ entonces las unidades de la distribución de probabilidad son trivialmente: $[f] = [d^L R]^{-1} = [R_1]^{-1}(\cdots)[R_L]^{-1}$. Ahora bien, puesto que en el cálculo de la entropía se está aplicando un logaritmo a f, entonces es necesario que $f(\vec{R})$ sea adimensional (ver http://math.ucr.edu/home/baez/physics/General/logs.html), para lo cual el elemento de volumen $d^L R$ debe ser adimensional. Si partimos de variables $r_1, (\cdots), r_L$ tales que $d^L r$ no es adimensional, entonces podemos normalizar $d\omega \equiv d^L r \to d^L R \equiv \dfrac{d\omega}{\omega_0} = \dfrac{d^L r}{\omega_0}$ donde ω_0 es una cantidad con unidades de volumen $[\omega_0] = [d^L r]$ que va a servir para adimensionalizar el elemento de volumen $d^L r$.

En consecuencia se pasa de tener $1 = \int_M \rho_N(\vec{r}) d^L r = \int_M \rho_N(\vec{r}) \omega_0 \dfrac{d^L r}{\omega_0} = \displaystyle\int_M f(\vec{R}) d^L R$, donde ρ_N es una densidad normalizada. También se puede tener una densidad ρ no normalizada ($V_0 \equiv \int_M \rho(\vec{r}) d^L r \neq 1$), por lo que podemos considerar que la densidad normalizada es $\rho_N(\vec{r}) = \dfrac{\rho(\vec{r})}{V_0}$ y luego se tiene: $f(\vec{R}) = \rho_N(\vec{r})\omega_0 = \frac{1}{V_0}\rho(\vec{r})\omega_0 = \dfrac{\rho(\vec{r})}{\int_M \rho(\vec{r}) d^L r}\omega_0 = \dfrac{\rho(\vec{r})}{\int_M \rho(\vec{r}) d^L r/\omega_0}$.

Si interpretamos a ω_0 como el volumen de un microestado, entonces la cantidad de microestados en un espacio muestral M es igual a $\Omega = \dfrac{\omega}{\omega_0} \equiv \dfrac{\text{Vol}(M)}{\omega_0} = \dfrac{1}{\omega_0}\displaystyle\int_M d^L R$.

No es difícil ver, al maximizar $S_{\text{est}}[f(\vec{R})]$ (que ahora es un funcional), que todas las ecuaciones 2, 3, 4, 5, 6 de la anterior sección son válidas reemplazando

$$\sum_{r\in M} \to \int_M d^L R = \int_M \frac{d^L r}{\omega_0} \tag{7}$$

1.2. Distribuciones de máxima entropía, ejemplo: Ejercicio 18b, guía 2

La distribución de probabilidad que maximice la entropía va a depender fuertemente de los vínculos $\langle A_k(\vec{R})\rangle = a_k$ en cuestión, y del dominio de las variables aleatorias. De hecho, podemos ver en la tabla de la página https://en.wikipedia.org/wiki/Maximum_entropy_probability_distribution#Other_examples que se pueden obtener distribuciones de probabilidad que maximicen la entropía muy diversas, que dependen en esencia de los vínculos elegidos y del dominio de las variables aleatorias.

En particular, mostraremos como es la distribución de probabilidad de máxima entropía en un sistema con una variable aleatoria x que pueden tomar cualquier valor real (continua), con los vínculos $\langle x\rangle = \mu$ (con multiplicador de Lagrange asociado λ) y $\langle \Delta x^2\rangle = \langle (x-\mu)^2\rangle = \sigma^2$ (con multiplicador de Lagrange asociado ξ). Entonces la función de partición del sistema queda de la forma:

$$\begin{aligned} Z &= \int_{\mathbb{R}} dx\ e^{-\lambda x-\xi(x-\mu)^2} = \int_{\mathbb{R}} dx\ e^{-\lambda(x-\mu+\mu)-\xi(x-\mu)^2} = e^{-\lambda\mu}\int_{\mathbb{R}} dx\ e^{-\lambda(x-\mu)-\xi(x-\mu)^2} \\ &= e^{-\lambda\mu}\int_{\mathbb{R}} dx\ e^{-\lambda\Delta x-\xi\Delta x^2} \end{aligned}$$

donde definimos $\Delta x \equiv x-\mu$. Luego, para poder integrar, completamos cuadrados en la expresión dentro de la exponencial del integrando:

$$\begin{aligned} &-\lambda\Delta x-\xi\Delta x^2 = -\xi\left(\Delta x^2+\frac{\lambda}{\xi}\Delta x\right) = -\xi\left(\Delta x^2+2\left(\frac{\lambda}{2\xi}\right)\Delta x\right) \\ &= -\xi\underbrace{\left(\Delta x^2+2\left(\frac{\lambda}{2\xi}\right)\Delta x+\frac{\lambda^2}{4\xi^2}\right)}_{\left(\Delta x+\frac{\lambda}{2\xi}\right)^2}+\frac{\lambda^2}{4\xi} = -\xi\left(\Delta x+\frac{\lambda}{2\xi}\right)^2+\frac{\lambda^2}{4\xi} \end{aligned} \tag{8}$$

donde se asumió que $\xi \neq 0$. Entonces la cuenta que hay que hacer es la siguiente:

$$\begin{aligned} Z(\lambda,\xi) &= e^{-\lambda\mu}\int_{\mathbb{R}} dx\ e^{-\lambda\Delta x-\xi\Delta x^2} \\ &= e^{-\lambda\mu}\int_{\mathbb{R}} dx\ e^{-\xi\left(\Delta x+\frac{\lambda}{2\xi}\right)^2+\frac{\lambda^2}{4\xi}} \\ &= e^{-\lambda\mu+\frac{\lambda^2}{4\xi}}\int_{\mathbb{R}} dx\ e^{-\xi\left(\Delta x+\frac{\lambda}{2\xi}\right)^2} \\ &= e^{-\lambda\mu+\frac{\lambda^2}{4\xi}}\int_{\mathbb{R}} dy\ e^{-\xi y^2} \\ &= e^{-\lambda\mu+\frac{\lambda^2}{4\xi}}\sqrt{\frac{\pi}{\xi}} \end{aligned} \tag{9}$$

donde en el segundo renglón se utilizó la ecuación 8, en el tercer renglón se extrajo de la integral el factor constante $e^{\frac{\lambda^2}{4\xi}}$, en el cuarto renglón se realizó el cambio de variable $x \to y \equiv x-\mu+\frac{\lambda}{2\xi} = \Delta x+\frac{\lambda}{2\xi}$ con jacobiano igual a 1, y en el quinto y último renglón se utilizó que la integral de una gaussiana en los reales es precisamente $\int_{\mathbb{R}} dy\ e^{-\xi y^2} = \sqrt{\frac{\pi}{\xi}}$. [2]

[2] Podemos demostrar ésto si lo elevamos al cuadrado, lo vemos como una integral en dos dimensiones y pasamos a coordenadas polares: $\left(\int_{\mathbb{R}} dy\ e^{-\xi y^2}\right)^2 = \left(\int_{\mathbb{R}} dx\ e^{-\xi x^2}\right)\left(\int_{\mathbb{R}} dy\ e^{-\xi y^2}\right) = \iint_{\mathbb{R}^2} dxdy\ e^{-\xi(x^2+y^2)} = \int_0^{2\pi} d\theta \int_0^{+\infty} dr\ r\ e^{-\xi r^2} = 2\pi\ \frac{1}{2\xi}\int_0^{+\infty} dr\ e^{-u} = \frac{\pi}{\xi}(-0+1) = \frac{\pi}{\xi}$, donde se utilizó que el jacobiano al pasar de coordenadas cartesianas a polares en $\mathbb{R}^2$ es igual a r, y se realizó el cambio de variable $r \to u \equiv \xi r^2$ (luego $rdr = \frac{1}{2\xi}du$). Por lo tanto, tomando raíz cuadrada de lo anterior: $\int_{\mathbb{R}} dy\ e^{-\xi y^2} = \sqrt{\frac{\pi}{\xi}}$.

Y luego:

$$\ln Z(\lambda, \xi) = \frac{1}{2}\ln\pi - \lambda\mu + \frac{\lambda^2}{4\xi} - \frac{1}{2}\ln\xi \tag{10}$$

Entonces, si se plantean los vínculos se obtiene:

$$\mu = -\frac{\partial}{\partial\lambda}\ln Z = \mu - \frac{\lambda}{2\xi} \iff \frac{\lambda}{2\xi} = 0 \iff \lambda = 0$$

Donde se utilizó que $\xi \neq 0$.

Y si calculamos ahora la derivada segunda de $\ln Z$ respecto a λ, deberíamos obtener la matriz de covarianza de x (o sea, como hay una sola variable, deberíamos obtener la varianza σ^2):

$$\sigma^2 = \langle (x-\mu)^2 \rangle = \frac{\partial^2(\ln Z)}{\partial\lambda^2} = -\frac{\partial}{\partial\lambda}\left(-\frac{\partial}{\partial\lambda}\ln Z\right) = -\frac{\partial}{\partial\vec{\lambda}}\left(\mu - \frac{\lambda}{2\xi}\right) = \frac{1}{2}\,\xi^{-1} \iff \xi = \frac{1}{2\sigma^2}$$

También podemos obtener la varianza σ^2 derivando a $\ln Z$ respecto al multiplicador de Lagrange ξ asociado al vínculo que fija la varianza:

$$\sigma^2 = -\frac{\partial}{\partial\xi}\ln Z = -\frac{\partial}{\partial\xi}\left(\frac{\lambda^2}{4\xi} - \frac{1}{2}\ln\xi\right) = -\frac{\lambda^2}{4}\frac{\partial}{\partial\xi}\left(\xi^{-1}\right) + \frac{1}{2}\frac{\partial}{\partial\xi}\left(\ln\xi\right) = \frac{\lambda^2}{4\xi^2} + \frac{1}{2\xi} = \frac{1}{2}\xi^{-1}$$

donde se utilizó en el último paso que $\lambda = 0$ por el anterior vínculo. De ésta forma se reobtuvo que $\sigma^2 = (2\xi)^{-1}$.

Ésto nos permite escribir la distribución de probabilidades de la forma:

$$\begin{aligned} f(x) &= \frac{1}{Z}e^{-\lambda x - \xi(x-\mu)^2} = \frac{1}{Z}e^{-\lambda\mu}e^{-\lambda\Delta x - \xi\Delta x^2} = \frac{1}{Z}e^{-\lambda\mu + \frac{\lambda^2}{4\xi}}\ e^{-\xi\left(\Delta x + \frac{\lambda}{2\xi}\right)^2} \\ &= \sqrt{\frac{\xi}{\pi}}\ e^{-\xi\left(x - \mu + \frac{\lambda}{2\xi}\right)^2} = \sqrt{\frac{1}{2\pi\sigma^2}}\ e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \end{aligned}$$

donde se utilizó la ecuación 8 en el tercer paso, se reemplazó $Z(\lambda, \xi)$ en el cuarto paso, y en el quinto y último paso se utilizó lo obtenido a partir de los vínculos: $\lambda = 0$ y $\xi = (2\sigma^2)^{-1}$. En definitiva, la función de distribución de probabilidad de una variable aleatoria x con dominio en $\mathbb{R}$, que maximiza la entropía sujeto a los vínculos $\langle x \rangle = \mu$ y $\langle (x-\mu)^2 \rangle = \sigma^2$, es una distribución gaussiana de una variable.

También se puede calcular la entropía estadística de ésta distribución simplemente haciendo:

$$\begin{aligned} S_{\text{est}}(f(\vec{R})) = S_{\text{est}}^{\max} &= \ln Z + \lambda\mu + \xi\sigma^2 = \frac{1}{2}\ln\pi + \frac{\lambda^2}{4\xi} - \frac{1}{2}\ln\xi + \xi\sigma^2 \\ &= \frac{1}{2}\ln\pi - \frac{1}{2}\ln(2\sigma^2)^{-1} + (2\sigma^2)^{-1}\sigma^2 = \frac{1}{2}\left(\ln\pi + \ln(2\sigma^2) + 1\right) = \frac{1}{2}\ln(2\pi e\ \sigma^2) \end{aligned}$$

donde en el tercer paso se reemplazó la expresión de $\ln Z$, en el cuarto paso se reemplazaron las expresiones $\lambda = 0$ y $\xi = (2\sigma^2)^{-1}$ obtenidas con los vínculos, y luego se hicieron cuentas.

Nótese de hecho que derivando a $S_{\text{est}}^{\max}(\mu, \sigma^2)$ podemos reobtener: $\lambda = \dfrac{\partial S_{\text{est}}^{\max}}{\partial\mu} = 0$ y también $\xi = \dfrac{\partial S_{\text{est}}^{\max}}{\partial(\sigma^2)} = \dfrac{1}{2\sigma^2}$, es decir: Recuperamos lo obtenido con los vínculos.