Teoría Avanzada de la Termodinámica – 1er. cuatrimestre de 2018

Guía 4: *Procesos estocásticos, entropía e infomación*

1. Para un proceso de Markov, demostrar que todas las densidades de probabilidad conjunta de n tiempos, $p_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)$, pueden construirse a partir de $p(x,t)$ y $p(x,t|x',t')$, con $t \geq t'$. [Para fijar la notación, $p(x,t|x',t')$ significa "la probabilidad (o densidad, según el caso) de que el estado del sistema sea x a tiempo t dado que era x' en el tiempo anterior t'.]

2. Por definición, en un proceso de Markov la densidad de probabilidad condicional satisface

$$p(x_n,t_n|x_0,t_0;\ldots;x_{n-2},t_{n-2};x_{n-1},t_{n-1}) = p(x_n,t_n|x_{n-1},t_{n-1}), \qquad \text{donde } t_n \geq t_{n-1} \geq \ldots \geq t_0.$$

Esto significa que inferencias sobre el estado actual del proceso dependen sólo del pasado inmediato.

a) Aunque esto parece distinguir una dirección del tiempo privilegiada, demostrar que vale la propiedad recíproca: inferencias sobre el estado actual del proceso dependen sólo del conocimiento de su futuro inmediato; es decir:

$$p(x_n,t_n|x_{n+1},t_{n+1};x_{n+2},t_{n+2};\ldots;x_{n+m},t_{n+m}) = p(x_n,t_n|x_{n+1},t_{n+1}),$$

donde $t_n \leq t_{n+1} \leq \ldots \leq t_{n+m}$.

b) Suponga que el sistema es observado en t_0 en el estado x_0, y en t_n en el estado x_n. Los estados intermedios no son conocidos. Interesa hacer inferencias acerca del camino que siguió el sistema entre los dos estados observados. Para eso, dé explícitamente $p(x_k,t_k|x_0,t_0;x_n,t_n)$, donde $t_n \geq t_k \geq t_0$, en términos de las probabilidades de transición directas, es decir de pasado a futuro. Generalice a $n-1$ estados intermedios entre t_0 y t_n.

3. *a*) Demostrar que las densidades de probabilidad condicional de tipo gaussiano y de Cauchy,

$$p_{\mathrm{G}}(x,t|x',t') = \frac{1}{\sqrt{2\pi(t-t')}} \exp\left[-\frac{(x-x')^2}{2(t-t')}\right],$$

$$p_{\mathrm{C}}(x,t|x',t') = \frac{(t-t')/\pi}{(t-t')^2+(x-x')^2}, \tag{1}$$

satisfacen la ecuación de Chapman−Kolmogorov,

$$p(x_2,t_2|x_0,t_0) = \int dx_1\; p(x_2,t_2|x_1,t_1)p(x_1,t_1|x_0,t_0), \qquad (t_0 < t_1 < t_2),$$

y que para $t' \to t$ ambas tienden a $\delta(x-x')$.

b) Demostrar que las densidades condicionales anteriores son compatibles con las siguientes densidades de probabilidad de un tiempo, según el caso,

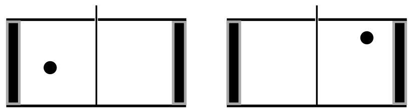
$$p_{\mathrm{G}}(x,t) = \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}, \qquad p_{\mathrm{C}}(x,t) = \frac{t/\pi}{t^2+x^2}.$$

Compatibles significa que

$$p(x_2,t_2) = \int dx_1\; p(x_2,t_2|x_1,t_1)p(x_1,t_1).$$

c) Para hacer en la computadora. Bastan unas pocas líneas para generar caminatas al azar de acuerdo a las dos distribuciones anteriores. Suponga que una partícula parte del origen en $t = 0$. A partir de un generador de números aleatorios obtenga la posición en tiempos sucesivos, Δt, $2\Delta t$, etc. (Muchos programas ya tienen definido el generador gaussiano. En todo caso, cualquier distribución puede obtenerse a partir de la distribución uniforme en el intervalo $[0, 1]$. Ver la Guía 1.) El paso temporal no necesita ser pequeño, porque se usan las densidades condicionales exactas, Ec. (1). Variando el paso, lo que uno consigue es construir trayectorias con mayor o menor definición temporal, lo que equivale a filmar a la partícula con mayor o menor cantidad de cuadros por segundo. Note que no importa qué tan pequeño sea el paso temporal, las trayectorias nunca se suavizan. ¿Qué característica muy clara distingue la caminada según Gauss de la caminata según Cauchy?

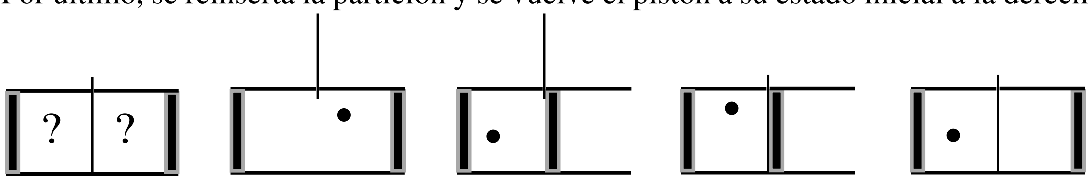
4. Una memoria elemental consiste en un conjunto de cajas o “bits”, cada una de volumen V y conteniendo una sola partícula. A cada lado de la caja hay un pistón y en el medio una partición removible. Si la partícula está en el lado izquierdo, el estado es **0** y si está en el lado derecho es **1**.



estado **0** estado **1**

Las cajas están en contacto con un foco térmico a temperatura T. La partícula puede tratarse como un gas ideal, con ecuación de estado $PV = kT$ y entropía $S = S_\mathrm{r} + k \log(V/V_\mathrm{r}) + ck \log(T/T_\mathrm{r})$, donde S_r, V_r y T_r corresponden a un estado de referencia y c es constante. Un bit es borrado cuando, independientemente de su estado inicial, y sin conocimiento de este estado, se lo fuerza al estado **0** mediante las siguientes operaciones:

- Se quita la partición, permitiendo que el gas se expanda libremente de $V/2$ a V.
- Isotérmica y cuasiestáticamente se comprime el gas hasta un volumen $V/2$ mediante el pistón de la derecha, de manera que, sea cuál sea el estado inicial, la partícula termina en el lado izquierdo.
- Por último, se reinserta la partición y se vuelve el pistón a su estado inicial a la derecha de la caja.



a) Encuentre el cambio total de entropía en el sistema formado por la memoria y el reservorio térmico durante este proceso de borrado de un bit.

b) El segundo pistón, que hasta aquí no se usó, tiene una función: proponga uno o más procedimientos para pasar del estado **0** al **1**, de manera reversible, sin transferencia de calor ni realización de trabajo.

5. En Teoría de la Información, la entropía contenida en un mensaje se define como

$$S = -\sum_{r=1}^{M} p(r) \log p(r),$$

donde $p(r)$ es la probabilidad del mensaje r, entre M mensajes posibles. Notar que la entropía no está referida a un mensaje en particular, sino a un ensamble de mensajes posibles. A mayor entropía mayor es la incertidumbre antes de recibir un mensaje, y entonces mayor es la información transmitida por cada mensaje. La base en la que se calcula el logaritmo define la unidad de medida de la información. En este contexto suele usarse el logaritmo en base 2 y la entropía se mide en *bits*.

a) Graficar $x \log x$ en el intervalo $[0,1]$. ¿Qué pasa con S cuando uno de los mensajes ocurre con probabilidad 1?

b) Demostrar que S es no negativa y que es máxima cuando todos los mensajes tienen igual probabilidad, $p(r) = 1/M$. *Sugerencia:* que hay un extremo es fácil; demostrar que es un máximo es más difícil. Las dos cosas pueden hacerse en un solo paso usando la llamada desigualdad de Jensen, de la teoría de funciones convexas. Si Φ es continua y convexa, entonces

$$\Phi\left(\frac{1}{M}\sum_{i=1}^{M} a_k\right) \leq \frac{1}{M}\sum_{i=1}^{M}\Phi(a_k).$$

(Es fácil entender esta desigualdad si se piensa a $\Phi(x)$ como la función que define la coordenada y de cierto arco convexo, con una distribución de masas unidad sobre el arco y cuyas coordenadas x son las cantidades a_k. La desigualdad dice que el centro de masa está *dentro* del arco, lo que es bastante intuitivo; Callen, §17-1.)

También puede hablarse de la entropía de una fuente de mensajes. Supongamos que la fuente emite mensajes con un número arbitrario de caracteres. La entropía asociada a los mensajes de longitud N es

$$S_N = -\sum_{r=1}^{M_N} p_N(r) \log p_N(r),$$

donde la suma sólo se extiende a los M_N mensajes de longitud N. Si existe el límite

$$s = \lim_{N\to\infty} \frac{S_N}{N},$$

entonces s es la entropía por caracter asociada a la fuente. Los siguientes ítems analizan lo que sucede cuando la probabilidad de cada caracter depende de un número finito de caracteres inmediatamente anteriores. El caso más simple es una fuente sin correlaciones, donde cada caracter es emitido con independencia de los anteriores. El siguiente caso en complejidad es cuando la probabilidad de emisión de un caracter depende sólo del inmediatamente anterior. Y así siguiendo con fuentes más complicadas, donde la probabilidad de cada caracter depende de los dos, tres o más caracteres anteriores. Por ejemplo, en español, $p(\mathrm{o}|\mathrm{e,n,t,r,o,p,í}) \approx 0$, $p(\mathrm{K,o,l,m,o,g,o,r,o,v}|\mathrm{C,h,a,p,m,a,n,\text{-}}) \approx 1$.

Para fijar la notación, un mensaje de N caracteres es una $N-$upla ordenada $(x_1, x_2, \ldots, x_N)$. El índice de cada x_i es la posición que ocupa dentro del mensaje. Cada x_i se elige de entre un conjunto de M símbolos distintos. Se asume siempre que el proceso es homogéneo, de modo que la probabilidad de encontrar un dado grupo de n caracteres dentro del mensaje no depende de qué parte del mensaje se mire. Por ejemplo $p(x_1 = a) = p(x_2 = a)$, $p(x_1 = a, x_2 = b) = p(x_3 = a, x_4 = b)$, etc. (Esta hipótesis sólo es razonable para mensajes largos y no muy cerca de los extremos; en algún sentido, se desprecian los efectos de borde.) Así, el índice i puede usarse también para señalar la posición en que aparece un dado caracter dentro de un grupo de caracteres **consecutivos**. Notar que no es lo mismo $p(x_1 = a, x_2 = b)$ que $p(x_1 = a, x_3 = b)$.

c) Si los caracteres no están correlacionados, demostrar que S_N se escribe en términos de S_1. Encontrar s. A esta s, que se calcula usando sólo $p_1(x_1)$, y que es exacta en el caso en que no hay correlaciones, la llamaremos s_1.

d) Suponga que el proceso de emisión de caracteres liga la probabilidad de cada nuevo caracter sólo con el anterior. Es decir, es un proceso de Markov en el sentido habitual: basta con conocer las probabilidades $p_2(x_1, x_2)$ y $p_1(x_1)$. Escriba S_N en términos de S_2 y S_1. ¿Cuánto vale s? A esta s, que se calcula usando sólo $p_2(x_1, x_2)$ y $p_1(x_1)$, y que es exacta cuando las probabilidades condicionales sólo ligan 2 caracteres, la llamaremos s_2.

Más generalmente, cada caracter estará ligado a los n anteriores (notar que los caracteres en las posiciones $i \leq n$ están ligados en verdad con un número menor de caracteres, pues el mensaje empieza en $i = 1$). Esta es una extensión natural de los procesos de Markov usuales. Para $N > n$ vale

$$p(x_N | \underbrace{x_1, x_2, \ldots, x_{N-2}, x_{N-1}}_{N-1}) = p(x_N | \underbrace{x_{N-n}, \ldots, x_{N-2}, x_{N-1}}_{n}). \tag{2}$$

e) En analogía con los procesos de Markov ordinarios, escriba cualquier probabilidad conjunta de N caracteres consecutivos usando $p_1(x_1)$ y las condicionales

$$p(x_2|x_1), \quad p(x_3|x_1, x_2), \quad \ldots \quad p(x_N|x_{N-n}, \ldots, x_{N-2}, x_{N-1}).$$

f) Demostrar que es posible escribir S_N en términos de S_{n+1} y S_n. ¿Cuánto vale s? A esta s, que se calcula usando las probabilidades condicionales que ligan hasta $n + 1$ caracteres consecutivos, y que es exacta si vale la Ec. (2), la llamaremos s_{n+1}.

Puede darse el caso de que la probabilidad de cada nuevo caracter dependa siempre de todos los anteriores, sin importar lo extenso que sea el mensaje. Así, en principio, para calcular la entropía s sería necesario conocer las probabilidades conjuntas $p_N(x_1, \ldots, x_N)$ con N arbitrariamente grande. El cálculo de la entropía usando la estadística que incluye a lo sumo grupos de $n + 1$ caracteres, es decir, lo que hemos llamado s_{n+1}, será, como mucho, una aproximación a la s verdadera. El ítem anterior dice como calcular s_{n+1}. Puede demostrarse que bajo ciertas condiciones $s = \lim_{n\to\infty} s_{n+1}$. Así, para ciertas fuentes tiene sentido tratar a s_1, s_2, s_3, etc., como aproximaciones cada vez mejores de s. Además, puede demostrarse que $s \leq s_{n+1} \leq s_n$, de modo que cada s_n da una cota más precisa para s.