Seminario 1: Teoría de la Información.

- Información, incertidumbre e ignorancia.
- Reglas para asignar probabilidades.
- Información y Física estadística.
- Otras aplicaciones.

La FE no involucra (ni necesita de) la **teoría de la información**, desarrollada en sus orígenes por Shanon (1948) y Jaynes (1957), pero tiene cierto interés pedagógico en relación con la FE, puesto que las ideas de *información* e *incertidumbre* se aceptan intuitivamente, y conduce rápidamente a la idea y caracterización de una colectividad, por lo que es una alternativa (*no rigurosa*) a la teoría de colectividades (con su problema ergódico incluido). Además es susceptible de aplicaciones interesantes en Física.

Bibliografía:

- L. Brillouin, *Science and Information Theory,* Academic.
- H.G. Harris, *Introduction to Modern Theoretical Physics*, vol. 2; y libro de Gopal.
- A. Katz, *Principles of Statistical Mechanics,* W.H. Freeman, San Francisco.
- E.T. Jaynes, *Information Theory and Statistical Mechanics,* Brandeis Summer Institute, Benjamin, NY.

Conceptos relevantes.

El concepto científico de información no coincide exactamente con el usual, y aquí interesan más los conceptos de *ausencia de información* o *incertidumbre* (por su relación con el de entropía), en cierto modo complemetarios del de *información.*

El concepto de *incertidumbre* se sigue de consideraciones puramente estadísticas. Por ejemplo, sea una situación en la que pueden presentarse n sucesos distintos de un experimento aleatorio, todos igualmente probables 'a priori', es decir, de acuerdo con nuestro conocimiento inicial. Podemos pensar en

- n cajas, una (y sólo una) de las cuales contiene un objeto, y nos piden encontrar éste sin otra información y sin poder realizar el experimento;
- un alfabeto con n letras igualmente probables, y queremos conocer la primera letra de una palabra oculta escrita por otra persona.
- lanzar un dado ($n = 6$), donde cada cara tiene la misma probabilidad de salir y queremos saber cual va a ser el resultado.

Nuestra incertidumbre o *ignorancia* es máxima en esta condición inicial. Se tiene la primera fila de la tabla:

estado	posibilidades	incertidumbre, U	información, I
inicial	n_0, máx.	U_0, máx.	$I_0 = 0$, mín.
intermedio	$n_i < n_0$	$U_i < U_0$	$I_i > I_0$
final ideal	$n_f = 1$	$U_f = 0$	I_f, máx.

Pero puede ocurrir que consigamos información sobre el sistema, reduciendo así el número n de posibilidades que pueden presentarse o, incluso, llegando a conocer con certeza cuál se presentará. Se tienen las otras dos filas de la tabla.

Los casos anteriores son ejemplos de *experimentos aleatorios*, esto es, toda experiencia en la que no se sabe con certeza cual va ser el resultado que va a salir.
Ante un experimento aleatorio uno tiene cierta información (por ejemplo, 6 caras de un dado, dado trucado, etc) pero también hay cierta incertidumbre (no sabemos con certeza el resultado del experimento).
En FE al estudiar los sistemas macroscópicos no podemos conocer posiciones y momentos de todas las partículas. Solo tenemos acceso a los macroestados y no a los microestados (información parcial).

- I. Postulado

$$A \longrightarrow \hat{A}$$
$$\text{valores observados} \longrightarrow \langle \hat{A} \rangle$$

- II. Postulado

$$\langle \hat{A} \rangle = \mathrm{Tr}(\hat{\rho}\hat{A}) \qquad \hat{\rho} = \sum_{\alpha} |\alpha\rangle p_\alpha \langle\alpha|$$

  Teorema: sistema en equilibrio $\Longleftrightarrow [\hat{\rho}, \hat{H}] = 0$ con $\hat{\rho} = \sum_r |r\rangle p_r \langle r|$ donde $|r\rangle$ estado propio de $\hat{H}$ con autovalor E_r. Sistema aislado en equilibrio

$$E < H < E + \Delta E \qquad p_r = \begin{cases} \text{cte} & |r\rangle / E \leq E_r \leq E + \Delta E \\ 0 & \text{resto} \end{cases}$$

aqui p_r son las probabilidades de que suceda algo. Son probabilidades de que al elegir al azar un estado del sistema y medir sobre él H, el valor obtenido fuera el autovalor E_r. El experimento aleatorio en este caso es elegir un estado del sistema. La única información que tengo sobre microestados es si son o no compatibles con el sistema. Ante la falta de información (I) le asignamos a todos los microestados compatibles con el sistema, la misma probabilidad ("Postulado de igual probabilidad a priori."). Sin embargo tenemos una información adicional que es la información del macroestado del sistema.

Vamos a formalizar todo lo que acabamos de decir:

- Exp. Aleatorio: Experimento en el que, a priori, no sabemos el resultado. Cada uno de los posibles resultados se llama **suceso** del experimento aleatorio. Los sucesos han de verificar:
  - Incompatibles entre sí (no se den a la vez)
  - El conjunto de sucesos ha de ser completo: siempre que realicemos el experimento debe salir un de ellos.
  - El número posible de sucesos ha de ser finito.

- Dado un experimento aleatorio se dice que se puede dar un esquema cuando se da los sucesos y sus probabilidades:

  Esquema A $\begin{pmatrix} A_1 & A_2 \dots A_n \\ p_1 & p_2 \dots p_n \end{pmatrix}$ $\begin{matrix} A_i & \text{sucesos del experimento A} \\ p_i & \text{probabilidades o frequencias del suceso } A_i \end{matrix}$

  Las p_i describen el mecanismo de aleatoriedad del experimento. Al dar un esquema tenemos una cierta información pero también hay incertidumbre.

Función Incertidumbre o Ignorancia

Vamos a tratar de cuantificar la incertidumbre definiendo una cierta función $U(n)$. Que propiedades debe tener?

1 Debe ser función de las n probabilidades $U(n) = U(p_1, \dots, p_n)$ y definida positiva $U(n) > 0$ si $n > 1$ (supondremos que $n \geq 1$) además $U(1) = 0$ (si solo hay un posible suceso no hay incertidumbre).

2 $U(n)$ debe ser máxima cuando las p_i sean iguales entre sí.

$$U_{\max} = U(p, \dots, p) = U(\tfrac{1}{n} \dots \tfrac{1}{n}) \text{ por normalización de las } p_i$$

3 El valor de la incertidumbre no cambia si le añadimos un suceso imposible (suceso que nunca ocurre)

$$U(p_1, \dots, p_n) = U(p_1, \dots, p_n, 0)$$

4 Si tenemos dos esquemas A, B y estos son mutuamente idependientes (no hay correlaciones entre los dos esquemas $\Rightarrow A_iB_j \longrightarrow p_iq_j$)

$$U(AB) = U(A) + U(B)$$

$$\text{A=}\begin{pmatrix} A_1 \dots A_n \\ p_1 \dots p_n \end{pmatrix} \qquad \text{B=}\begin{pmatrix} B_1 \dots B_m \\ q_1 \dots q_m \end{pmatrix}$$

$$\text{AB=}\begin{pmatrix} A_1B_1 & A_1B_2 \dots A_nB_m \\ p_1q_1 & p_1q_2 \ \ \dots p_nq_m \end{pmatrix}$$

Intuimos que

$$U(n) = K \ln n, \qquad K = \text{ cte. positiva,}$$

Vamos a verlo:

- Consideremos primero el caso de n sucesos equiprobables
  - $U(n)$ es una función creciente del argumento

    $$U(n) = U(\tfrac{1}{n} \dots \tfrac{1}{n})$$
    $$U(n) = U(\tfrac{1}{n} \dots \tfrac{1}{n}, 0) \leq U(\tfrac{1}{n+1} \dots \tfrac{1}{n+1}) = U(n+1)$$

    Luego es una función creciente del argumento. Esto viene a decir que si el numero de sucesos posibles se reduce en uno (por ejemplo porque adquiero cierta información) la incertidumbre decrece.
  - Consideremos r esquemas mútuamente independientes

    $$U(A^1 A^2 \dots A^r) = \sum_i^r U(A^i)$$

    que en el caso de que todos sean esquemas de n sucesos equiprobables

    $$U(A^1 A^2 \dots A^r) = \sum_i^r U(n) = rU(n) = U(n^r)$$

    para la última igualdad hemos utilizado que por definición r esquemas independientes de n sucesos equiprobables cada uno es un esquema de n^r sucesos equiprobables. Luego

    $$U(n^r) = rU(n)$$
  - Sean $s, m, n \in \mathbb{N}^*$, arbitrarios, siempre $\exists\ r$ tal que

    $$n^r \leq m^s \leq n^{r+1}$$

    $$r \ln n \leq s \ln m \leq (r+1) \ln n$$

    y dividiendo por $s \ln n$

    $$\frac{r}{s} \leq \frac{\ln m}{\ln n} \leq \frac{r}{s} + \frac{1}{s}$$

Las incertidumbres para ese tipo de esquemas verifican (por ser creciente)

$$U(n^r) \leq U(m^s) \leq U(n^{r+1}) \Rightarrow rU(n) \leq sU(m) \leq (r+1)U(n)$$

y dividiendo por $U(n) \neq 0$ se tiene

$$\frac{r}{s} \leq \frac{U(m)}{U(n)} \leq \frac{r}{s} + \frac{1}{s}$$

luego

$$\left|\frac{\ln m}{\ln n} - \frac{U(m)}{U(n)}\right| \leq \frac{1}{s}$$

que tiende a cero para s suficientemente grande luego concluimos

$$U(m)/\ln m = U(n)/\ln n = k = \text{cte}$$

y por tanto $U(m) = k \ln m$.[62]

- Pero esta definición no es completa pues se refiere a sucesos equiprobables. En la práctica interesan esquemas más generales con sucesos con probabilidades 'a priori':

$$p_i \geq 0, \quad i = 1, 2, ..., n; \quad \sum_{i=1}^{n} p_i = 1.$$

[62]Otra demostración alternativa que se deriva tambien de las condiciones solicitadas a U. En efecto, supongamos que $U(n)$ es continua con buen comportamiento (derivable e integrable) y que $U > 1$. Entonces, $4 \Longrightarrow U(x^n) = nU(x)$. Derivando, y usando la notación $U'(\alpha) \equiv \mathrm{d}U(\alpha)/\mathrm{d}\alpha$,, tenemos

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[U(x^n) = nU(x)\right] \quad \rightarrow \quad U'(x^n)\, nx^{n-1} = nU'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}n}\left[U(x^n) = nU(x)\right] \quad \rightarrow \quad U'(x^n)\, x^n \ln x = U(x).$$

Combinando estas dos ecuaciones:

$$U(x) = U'(x)\, x \ln x \quad \rightarrow \quad \frac{U'(x)}{U(x)} = \frac{1}{x \ln x} \quad \rightarrow \quad \frac{\partial}{\partial n} \ln U(x) = \frac{1}{x \ln x}.$$

Integrando entre x y a :

$$\ln U(x)|_x^a = \ln(\ln x)|_x^a \quad \rightarrow \quad U(x)|_x^a = \ln x|_x^a \quad \rightarrow$$

$U(x) - U(a) = \ln x - \ln a \quad \rightarrow \quad U(x) = \ln x + U(a) - \ln a$, pero $U(1) = 0 \quad \rightarrow$ $U(a) - \ln a = 0$, luego $U(x) = \ln x$. Falta cte. de proporcionalidad pues la prueba involucra derivaciones; esta cte. es postiva por propiedades 1 y 3.

distintas. Si el experimento se repite muchas, N, esperamos que el suceso i ocurra Np_i veces. La única incertidumbre adicional que tenemos es el orden que aparecen los Np_i resultados del suceso A_i. Los sucesos de este experimento son cadenas del tipo por ejemplo

$$A_1 A_1 A_r \ldots A_n$$

El número de cadenas distintas (teniendo en cuenta que no vale permutar sucesos iguales pues me dan la misma cadena) es

$$W = \frac{N!}{\prod_{r=1}^{n}(Np_r)!}$$

Podemos asociar a cada una de estas cadenas un suceso de hacer N experimentos y todos ellos tienen asociada la misma probabilidad pues en todas ellas el suceso A_1 aparece Np_1 veces, el A_2 aparece Np_2, etc. Luego aplicamos nuestro resultado anterior

$$U_N\left(\frac{N!}{\prod_{r=1}^{n}(Np_r)!}\right) = k \ln \frac{N!}{\prod_{r=1}^{n}(Np_r)!},$$

En consecuencia,

$$U_N\left(W\right) = k\left[\ln N! - \sum_{r=1}^{n} \ln\left(Np_r\right)!\right]$$

y, suponiendo N suficientemente grande

$$\begin{aligned} \ln N! \quad &\simeq N \ln N - N \\ \ln\left(Np_r\right)! \; &\simeq Np_r \ln\left(Np_r\right) - Np_r, \end{aligned}$$

se sigue

$$U_N(W) = -Nk \sum_{r=1}^{n} p_r \ln p_r$$

y la **incertidumbre media**, ee, la asociada con uno de los experimentos, es

$$U(n) \equiv \frac{U_N(W)}{N} = -k\sum_{r=1}^{n} p_r \ln p_r. \qquad (74)$$

Si todos los sucesos fueran igualmente probables, $p_i = 1/n$, y recobraríamos la definición de arriba.

(74) es la **medida de Shanon** para la incertidumbre (de un suceso que puede prsentarse de n formas, cada una con probabilidad p_i) o para la información que ganamos cuando una de esos sucesos ocurre. Algunos la llaman "entropía de información".[63]

Si la variable (discreta) i puede considerarse continua, x, escribimos

$$U = -K \int \mathrm{d}x\ P(x)\ \ln P(x),$$

no vamos a demostrarlo (Jaynes 1963, Katz 1967).[64]

En algunas aplicaciones (comunicaciones, ordenadores digitales,...) juega un papel importante el sistema binario.

Sea un problema con m sucesos posibles, cada uno con dos posibilidades, 0 ó 1. El número total de posibilidades es $n = 2^m$ y, en consecuencia, la incertidumbre asociada es

$$U = k \ln n = km \ln 2.$$

Si ahora queremos identificar U con el n$^o m$ de sucesos (dígitos binarios):

[63] Tal denominación es consecuencia de la analogía que puede establecerse entre incertidumbre y entropía cuando la teoría de Shanon se aplica a un sistema físico en el que los casos posibles se corresponden con microestados accesibles (pero no en otros casos).

[64] La cte. aditiva k fija las unidades de medida de la incertidumbre, luego queda determinada por las unidades del problema/sistema en cuestión. Cuando tenga sentido comparar U con entropía, es conveniente que ambas tengan las mismas unidades y se sigue $k = k_{\mathrm{Boltzmann}} = 8'617 \times 10^{-5}$ eV/^{o}K. Otras veces conviene que U no tenga dimensiones; basta con que no las tenga k, por ejemplo midiendo la temperatura en unidades de energía.

$$U = m \quad \Longrightarrow \quad K = 1/\ln 2 = \log_2 \mathrm{e},$$

luego se sigue:

$$U = \log_2 n$$

en lugar de la definición de Shanon arriba; se dice entonces que la incertidumbre está medida en *bits* (ejemplo, en Brillouin 1962, p.2).[65]

Asignar probabilidades en ausencia de información.

La teoría proporciona método sistemático para, de acuerdo con nuestros conocimientos, asignar probabilidades a distintos sucesos. Se basa en un **principio de máxima incertidumbre** que la teoría de la información enuncia en analogía con el principio de máxima entropía de la Física: (*grosso modo*) perturbaciones aleatorias —inevitables en el mundo real— tienden a degenerar la información y, por tanto, a incrementar la ignorancia y la incertidumbre.

> Es decir, una distribución picuda tipo gaussiano asociada con un observable determinado tiende a ensancharse debido a causas naturales.

Evitamos discutir este principio, que requiere entrar en terreno filosófico. Lo aceptamos provisionalmente, deducimos una regla concreta de él, aplicamos la regla a casos concretos, y analizamos si las consecuencias son asumibles.

La regla en cuestión puede enunciarse:

al tratar de asignar probababilidades a un suceso,
hemos de utilizar toda la información disponible
y nada más que la información cierta disponible.

[65] Nótese que la relación entre los valores de la constante k en estas aplicaciones, $k = 1/\ln 2$, y en aplicaciones de la física, $k_{16} = k_{\mathrm{Boltzmann}} = 8'617 \times 10^{-5}$ eV/^{o}K, es próxima a 10^{16}, lo que evidencia la gran cantidad de información —medida en bits— necesaria para modificar significantemente la entropía.

En teoría de la información, esta última condición equivale a
maximizar nuestra falta de información (o incertidumbre) acerca del suceso o del conjunto de sucesos en cuestión.

Aplicación de Teoría de la información a la FE

Ejemplos de aplicación del principio de máxima ignorancia o incertidumbre:
Antes de aplicarlo a las colectividades veamos un ejemplo sencillo de un dado para ver el formalismo: Disponemos de un dado con caras numeradas del 1 al 6 que, al ser tirado sobre superficie plana, puede mostrar cualquiera de sus caras. ¿Qué probabilidades, p_i, hemos de asignar a cada uno de los 6 sucesos posibles?

Suponemos primero que no disponemos de otra información. La incertidumbre asociada es

$$U(6) = -k \sum_{i=1}^{6} p_i \ln p_i$$

y la regla indica que maximicemos esta magnitud, luego:

$$\delta U = -k \sum_{i=1}^{6} \delta p_i \left(\ln p_i + 1 \right) = 0. \tag{75}$$

Pero los δp_i no son independientes, pues tenemos la ligadura $\sum p_i = 1$ que:

$$\sum_{i=1}^{6} \delta p_i = 0. \tag{76}$$

Usando multiplicadores de Lagarange, multiplicamos (76) por $k\alpha$ y sumamos a (75):

$$-k \sum_{i=1}^{6} \delta p_i \left(\ln p_i + 1 - \alpha \right) = 0.$$

Esta ecuación ya no está sometida a restricción alguna luego, al ser los δp_i independientes y arbitrarios, hemos de exigir que se anulen sus coeficientes,

$$\ln p_i + 1 - \alpha = 0 \quad \forall i,$$

de donde

$$p_i = \mathrm{e}^{\alpha - 1} = \text{ const.}$$

El parámetro α puede determinarse usando la condición de normalización, $\sum \mathrm{e}^{\alpha-1} = 6\mathrm{e}^{\alpha-1} = 1 \Longrightarrow \mathrm{e}^{\alpha-1} = 1/6$, luego

$$p_i = \mathrm{e}^{\alpha-1} = \frac{1}{6}.$$

El resultado es razonable, aunque trivial: en ausencia de otra info, lo mejor que podemos hacer es asignar probabilidades $1/6$ a todas las caras.

Supongamos que arrojamos el dado muchas veces y obtenemos i_0 como media de los números que aparecen el la cara superior, es decir,

$$\langle i \rangle = \sum_{i=1}^{6} i p_i = i_0.$$

En el caso anterior, ha de ser

$$i_0 = \frac{1}{6}(1+2+3+4+5+6) = 3'5.$$

Supongamos, sin embargo, que el experimento arroja $i_0 = 4'5$, lo que delata cierta inhomogeneidad en el peso de las caras que favorece los números altos (dado trucado).
¿Cómo escoger p_i si contamos con esta nueva información?
Usando multiplicadores de nuevo, la situación actual requiere que:

$$\begin{array}{ll} \delta U & = -k \sum_{i=1}^{6} \delta p_i (\ln p_i + 1) = 0 \\ \delta i_0 & = \sum_{i=1}^{6} i \; \delta p_i = 0 \\ \delta 1 & = \sum_{i=1}^{6} \delta p_i = 0, \end{array}$$

de donde, multiplicando por 1, $k\beta$ y $k\alpha$, respectivamente, y sumando:

$$-k \sum_{i=1}^{6} \delta p_i (\ln p_i + 1 - \beta i - \alpha) = 0.$$

Puesto que ha de satisfacerse para δp_i independientes y arbitrarios, han de anularse los coeficientes de donde se sigue que:

$$p_i = \mathrm{e}^{\alpha-1}\mathrm{e}^{\beta i} \equiv \frac{1}{Z}\mathrm{e}^{\beta i}, \tag{77}$$

donde los parámetros α (equivalentemente, $Z \equiv \mathrm{e}^{-\alpha+1}$) y β se siguen de normalización:

$$\sum_i p_i = 1 \quad \Longrightarrow \quad Z = \sum_i \mathrm{e}^{\beta i}$$

(que supone tener una relación entre α y β) y

$$i_0 = \sum_i i p_i \quad \Longrightarrow \quad i_0 = \frac{1}{Z}\sum_i i\mathrm{e}^{\beta i} = \frac{\partial}{\partial\beta}\ln Z.$$

A partir de estas expresiones es sencillo relacionar i_0 y β :[66]

$$i_0 = \frac{7}{2} + 3\coth(3\beta) - \frac{1}{2}\coth\frac{\beta}{2}. \tag{78}$$

En definitiva, dado i_0, la ec. (78) puede resolverse para β y utilizar el resultado en (77) para determinar las probabilidades buscadas p_i.

Notemos que

- $\beta = 0$ corresponde a $i_0 = 7/2$, el caso trivial para el que $p_i = 1/6\ \forall i$;
- si $3'5 < i_0 < 6$, se obtienen valores positivos de β (lo que equivale a "primar" las P_i con mayor i); mientras que
- si $0 < i_0 < 3'5$, se obtienen valores negativos de β.

[66]En efecto, tenemos sucesivamente:

$$\begin{aligned} Z &= \textstyle\sum_{i=1}^{6} \mathrm{e}^{\beta i} = \mathrm{e}^{\beta}\frac{1-\mathrm{e}^{6\beta}}{1-\mathrm{e}^{\beta}} && \text{pues } s_n = a_1\frac{1-r^n}{1-r} \\ &= \mathrm{e}^{\beta}\frac{\mathrm{e}^{-3\beta}-\mathrm{e}^{3\beta}}{\mathrm{e}^{-\beta/2}-\mathrm{e}^{\beta/2}}\cdot\frac{\mathrm{e}^{3\beta}}{\mathrm{e}^{\beta/2}} = \frac{\mathrm{e}^{3\beta}-\mathrm{e}^{-3\beta}}{\mathrm{e}^{\beta/2}-\mathrm{e}^{-\beta/2}}\mathrm{e}^{\frac{7}{2}\beta} \end{aligned}$$

luego

$$\begin{aligned} i_0 &= \tfrac{\partial}{\partial\beta}\ln Z = \tfrac{\partial}{\partial\beta}\left[\tfrac{7}{2}\beta + \ln\left(\mathrm{e}^{3\beta}-\mathrm{e}^{-3\beta}\right) - \ln\left(\mathrm{e}^{\beta/2}-\mathrm{e}^{-\beta/2}\right)\right] \\ &= \tfrac{7}{2} + \frac{3\mathrm{e}^{3\beta}+3\mathrm{e}^{-3\beta}}{\mathrm{e}^{3\beta}-\mathrm{e}^{-3\beta}} - \frac{\frac{1}{2}\mathrm{e}^{\beta/2}+\frac{1}{2}\mathrm{e}^{-\beta/2}}{\mathrm{e}^{\beta/2}-\mathrm{e}^{-\beta/2}} \\ &= \tfrac{7}{2} + 3\coth(3\beta) - \tfrac{1}{2}\coth\tfrac{\beta}{2} && \text{pues } \coth x = \frac{\mathrm{e}^{x}+\mathrm{e}^{-x}}{\mathrm{e}^{x}-\mathrm{e}^{-x}}. \end{aligned}$$

Física estadística: colectividades.

Las ideas anteriores pueden aplicarse de manera natural a la Física, por ejemplo, en FE para establecer distintas colectividades.

En este caso, definimos la ignorancia asociada con la función densidad clásica, es decir, $\rho(\alpha)$, donde $\alpha \equiv (q_1, ..., q_N; p_1, ..., p_N)$ es el conjunto de coord. generalizadas, como

$$S = -k \int \mathrm{d}\alpha \; \rho(\alpha) \; \ln \rho(\alpha),$$

con la integral extendida al espacio Γ.[67]

Se trata ahora de maximizar S sujeta a toda la información disponible.

Supongamos primero que ésta consiste en el conocimiento de la energía media de los miembros de la colectividad,

$$\langle E \rangle = \int \mathrm{d}\alpha \; H \; \rho \tag{79}$$

y queremos determinar la distribución de probabilidad, es decir, ρ, asociada con tal colectividad.

Se sigue que hemos de maximizar S respetando (79) y la normalización:

$$1 = \int \mathrm{d}\alpha \; \rho(\alpha).$$

En definitiva:

$$\delta S - k\beta \; \delta \langle E \rangle + k\alpha \; \delta 1 = -k \int \mathrm{d}\alpha \; \delta\rho \; (\ln \rho + 1 + \beta H - \alpha) = 0,$$

con α y β multiplicadores de Lagrange, luego

[67]Usamos aquí directamente S, en lugar de U, pues coincide con la entropía, de donde, $k =$ cte. de Boltzmann.

$$\ln \rho + 1 + \beta H - \alpha = 0$$

de donde

$$\rho = \mathrm{e}^{\alpha-1}\mathrm{e}^{-\beta H} = \frac{1}{Z}\mathrm{e}^{-\beta H} \tag{80}$$

con $Z = e^{1-\alpha} = \int d\alpha\ \mathrm{e}^{-\beta H}$, que es la **distribución canónica**.

Notas:

- De la condición de normalización:

$$Z = \int \mathrm{d}\alpha\ \mathrm{e}^{-\beta H}. \tag{81}$$

- Usando (80) en la definición de S :

$$\begin{aligned} S \quad &= -k \textstyle\int \mathrm{d}\alpha\ \frac{1}{Z}\mathrm{e}^{-\beta H}\left(-\ln Z - \beta H\right) \\ &= k\tfrac{\ln Z}{Z} \underbrace{\int \mathrm{d}\alpha\ \mathrm{e}^{-\beta H}}_{Z} + \tfrac{k\beta}{Z}\underbrace{\int \mathrm{d}\alpha\ \mathrm{e}^{-\beta H} H}_{Z\,\langle E\rangle} \\ &= k \ln Z + k\beta \langle E\rangle \end{aligned}$$

  Pero (79) y (81) implican $\langle E\rangle = -\frac{\partial}{\partial\beta}\ln Z$, luego

$$S = k\ln Z - k\beta\frac{\partial \ln Z}{\partial\beta} = -k\beta^2\frac{\partial}{\partial\beta}\frac{1}{\beta}\ln Z,$$

  que coincide con la entropía canónica si interpretamos $\beta \equiv 1/kT$ y $A = -kT\ln Z$.[68]

- Puede concluirse lo mismo en el caso de un **sistema cuántico**, donde hay que obtener el operador densidad a partir de la definición

$$S = -k\mathrm{Tr}\left(\hat{\rho}\ln\hat{\rho}\right)$$

[68]De hecho, puede verse que esta interpretación es correcta siguiendo el método de la FE. Por ejemplo, para que tenga sentido la definición

$$\langle E\rangle = \int \mathrm{d}\alpha\ H\ \mathrm{e}^{-\beta H},$$

he de ser $\beta = 1/kT$.

y la información, por ejemplo, de que:

$$\langle E \rangle = \mathrm{Tr}\left(H\hat{\rho}\right), \quad 1 = \mathrm{Tr}\left(\hat{\rho}\right).$$

Maximizando S con estas condiciones,

$$\delta S - k\beta\,\delta\left\langle E\right\rangle + k\alpha\,\delta 1 = -k\mathrm{Tr}\left[\delta\hat{\rho}\left(\ln\hat{\rho} + 1 + \beta\hat{H} - \alpha\right)\right] = 0$$

de donde

$$\hat{\rho} = \mathrm{e}^{\alpha-1}\mathrm{e}^{-\beta\hat{H}} = \frac{1}{Z}\mathrm{e}^{-\beta\hat{H}}$$

Supongamos ahora que nuestra información consiste en saber que todos los miembros de la colectividad tienen energías en el rango infinitesimal: $E_0 < E < E_0 + \mathrm{d}E$. Escribimos el elemento de volumen en esa región de Γ:

$$\mathrm{d}\alpha = \mathrm{d}E\ \mathrm{d}\sigma\left(E_0\right),$$

donde $\mathrm{d}\sigma\left(E_0\right)$ es un elemento de área en la superficie de energía constante E_0 en el espacio Γ, y definimos la incertidumbre:

$$S = -k\int_{E_0<E<E_0+\mathrm{d}E} d\alpha\ \rho\ln\rho = -k\int \mathrm{d}E\ \mathrm{d}\sigma\left(E_0\right)\ \rho\ \ln\rho.$$

aqui la primera integral está extendida a la región de Γ tal que

$E_0 < E < E_0 + \mathrm{d}E$ pues fuera $\rho = 0$ no hay microestados accesibles al sistema. Maximizando con la condición

$$1 = \int_{E_0<E<E_0+\mathrm{d}E} d\alpha\ \rho = \int \mathrm{d}E\ \mathrm{d}\sigma\left(E_0\right)\ \rho, \tag{82}$$

se tiene:

$$\delta S + k\alpha\,\delta 1 = -k\int \mathrm{d}E\,\mathrm{d}\sigma\,\delta\rho\ \left(\ln\rho + 1 - \alpha\right) = 0 \quad \Longrightarrow \quad \rho = \mathrm{e}^{\alpha-1} = \text{ cte.}$$

y la cte. se sigue de (82):

$$\rho = \begin{cases} 1/\Omega, & E_0 < E < E_0 + \mathrm{d}E, \quad \Omega = \int \mathrm{d}E \; \mathrm{d}\sigma\,(E_0) \\ 0, & \text{en otro caso} \end{cases}$$

que es la **distribución microcanónica.**

Hay que decir que la FE no es sólo teoría de la información pues

- Necesitamos la mecánica: Tenemos que resolver el H, y en cuántica necesitamos conocer el espectro de $\hat{H}$.
- Necesitamos saber cual es el significado de las p_i para lo cual necesitamos conocer los postulados de la ME IA y IB, si no conocemos esto no sabemos cual es el experimento aleatorio.
- El postulado II (de equiprobabilidad a priori, no mecánico), es algo más general que el Principio de Maxima Ignorancia. Además hacemos uso de la hipótesis ergódica, y para esto hay razones físicas.

Problema: Se van lanzando monedas de modo que siempre caen en una recta (eje x). Sabiendo que:

$$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{\infty} \mathrm{d}x \; x \; P\,(x) = x_0 \\ \left\langle x^2 \right\rangle &= \int_{-\infty}^{\infty} \mathrm{d}x \; x^2 \; P\,(x) = a^2, \end{aligned}$$

donde $P\,(x)\,\mathrm{d}x$ es la probabilidad de que la moneda caiga en $\mathrm{d}x$, mostrad que

$$P\,(x) = \sqrt{\frac{2\,(a^2 - x_0^2)}{\pi}} \exp\left[-\frac{(x - x_0)^2}{a^2 - x_0^2}\right].$$