6 Physics, entropy, and large deviations

6a Entropy in physics and the rate function

Each macrostate is compatible with, and hence is a summary of, many microstates. The entropy of a macrostate is a measure of this multiplicity.
Ellis [1, p. 5]

Here are two basic physical models and the corresponding formulas for the (physical) entropy.

The first (discrete) model consists of a large number n of classical[1] spin-$1/2$ particles; each such particle has two states: spin up, and spin down (and no other degrees of freedom). The value $r = (n_+ - n_-)/n$ is treated as 'the order parameter' (whatever it could meam); here n_+, n_- are the number of spins up and down, respectively. The entropy is

$$S = -k_{\mathrm{B}} n \biggl(\frac{1+r}{2} \ln \frac{1+r}{2} + \frac{1-r}{2} \ln \frac{1-r}{2} \biggr) + \mathrm{const} \, ;$$

k_{B} ($= 1.38 \cdot 10^{-23} \mathrm{J}/K$) is the Boltzmann constant (recall 1a), and 'const' is an arbitrary constant (it does not depend on r, but may depend on n).

The second (continuous) model consists of a large number n of classical particles moving in a container of volume V (with no other degrees of freedom). Let E be the total kinetic energy of the particles. The entropy is

$$S = k_{\mathrm{B}} n \Bigl(\frac{3}{2} \ln E + \ln V \Bigr) + \mathrm{const} \, ;$$

'const' is an arbitrary constant (it does not depend on E, V, but may depend on n and the mass m of each particle).

The question is, can we translate these physical statements into the language of large deviations?

[1] Spin is beyond the reach of classical mechanics, but within the reach of classical thermodynamics and classical statistical mechanics.

Amasingly, physical entropy emerged in thermodynamics, on the macroscopic level, without any idea of microscopic particles at all![1] This is the origin of the coefficient k_{B} in the entropy. We put the coefficient aside, and consider the entropy per particle,

$$\frac{S}{k_{\mathrm{B}} n} \, .$$

In classical physics, only the increments of entropy are well-defined; the additive constant remains arbitrary. (It is determined in quantum physics, but this is another story.)

In both models we have a sequence of measure spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and measurable maps $X_n : \Omega_n \to \mathbb{R}$. The measures P_n are positive but not just probability measures; moreover, they may be not finite (like Lebesgue measure on $\mathbb{R}$). However,

$$P_n\bigl(\{\omega \in \Omega_n : X_n(\omega) \in [a,b]\}\bigr) < \infty$$

whenever $-\infty < a < b < \infty$. We treat each P_n up to an arbitrary coefficient; that is, P_n may be replaced with $c_n P_n$ for any $c_1, c_2, \dots \in (0,\infty)$. (Ultimately, the choice of the coefficients is dictated by quantum physics...)

In the first (discrete) model, $\Omega_n = \{-1,+1\}^n$, P_n is the counting measure on Ω_n (or, if you prefer, the uniform probability distribution on Ω_n), and

$$X(s_1, \dots, s_n) = \frac{s_1 + \dots + s_n}{n} \, . \tag{6a1}$$

In the second (continuous) model we have a container, — an open set $G \subset \mathbb{R}^3$, $\operatorname{mes}_3 G = V \in (0,\infty)$. A particle is described by coordinates (a point of G) and momenta (a vector of $\mathbb{R}^3$). Accordingly, $\Omega_n = (G \times \mathbb{R}^3)^n$, with Lebesgue measure.[2] Thus, a point of Ω_n is a sequence of n points of $G \times \mathbb{R}^3$. Alternatively, you may define a point of Ω_n as an n-point subset of $G \times \mathbb{R}^3$ (which reflects indistinguishability of the particles[3]); it introduces the factor $1/n!$, which does not matter in our framework. The function X_n is the kinetic energy,

$$X_n\bigl((q_1,p_1), \dots, (q_n,p_n)\bigr) = \frac{1}{2m}\bigl(|p_1|^2 + \dots + |p_n|^2\bigr) \tag{6a2}$$

for $q_1, \dots, q_n \in G$ and $p_1, \dots, p_n \in \mathbb{R}^3$; here m is the mass of each particle.

[1] See [2, Sect. 5.1].

[2] In fact, the Liouville measure on the phase space, see [2, Sect. 4.1].

[3] See [2, Sect. 3.5].

It appears that in both cases

$$\|f(X_n(\cdot))\|_{L_n(c_n P_n)} \to \max_{\mathbb{R}}\bigl(|f|\mathrm{e}^{-I}\bigr) \quad \text{as } n \to \infty \tag{6a3}$$

for every *compactly supported* continuous $f : \mathbb{R} \to \mathbb{R}$, and some $I : \mathbb{R} \to (-\infty, +\infty]$, provided that $c_n \in (0, \infty)$ are chosen appropriately. The freedom of choosing c_n leads to an arbitrary additive constant in I. In the first model,

$$I(x) = \frac{1+x}{2} \ln \frac{1+x}{2} + \frac{1-x}{2} \ln \frac{1-x}{2} + \text{const} \quad \text{for } x \in [-1, 1] \tag{6a4}$$

and $I(x) = +\infty$ for other x. In the second model,

$$I(x) = -\frac{3}{2} \ln x + \text{const} \quad \text{for } x \in [0, \infty) \tag{6a5}$$

and $I(x) = +\infty$ for other x. The similarity to LDP is evident, and

$$I = -\frac{S}{k_{\mathrm{B}} n} + \text{const}\,, \quad S = -k_{\mathrm{B}} n I + \text{const}\,. \tag{6a6}$$

The dependence of S on the volume V disappeared because V was fixed. We may restore it by considering models with different G (and V) together, thus introducing $\Omega_{n,G}$ and stipulating the same c_n for all G. Doing so we get

$$I_V(x) = -\Bigl(\frac{3}{2} \ln x + \ln V\Bigr) + \text{const}\,.$$

The relation (6a3) for the first (discrete) model is basically the binomial LDP of 3a (Sanov's theorem). For the second (continuous) model we have

$$\int_{\Omega_n} |f(X_n(\cdot))|^n \, \mathrm{d}P_n = V^n \int_0^\infty |f(x)|^n \, \mathrm{d}\bigl(a_{3n}(2mx)^{3n/2}\bigr)\,,$$

where a_n is the volume of the unit ball in $\mathbb{R}^n$;[1] indeed, $P_n\bigl(\{\omega \in \Omega_n : X_n(\omega) \le x\}\bigr)$ is the Lebesgue measure of G^n multiplied by the volume of the $3n$-dimensional ball $\{(p_1, \dots, p_n) \in (\mathbb{R}^3)^n : (|p_1|^2 + \dots + |p_n|^2)^{1/2} \le \sqrt{2mx}\}$. Thus,

$$\begin{aligned} \|f(X_n(\cdot))\|_{L_n(P_n)} \sim V a_{3n}^{1/n} (2m)^{3/2} \Bigl(\frac{3n}{2} \int |f(x) x^{3/2}|^n \, \frac{\mathrm{d}x}{x}\Bigr)^{1/n} \sim \\ \sim \text{const}(n) \cdot V \max_{x \ge 0}\bigl(|f(x)| x^{3/2}\bigr)\,. \end{aligned}$$

[1] In fact, $a_n^{1/n} \sim \sqrt{\frac{2\pi \mathrm{e}}{n}}$ as $n \to \infty$.

Is energy relevant? I am afraid, if a physicist will read this text, he/she will say: 'the continuous model describes a simple (monatomic) ideal gas, but the author did not say so, why? Probably these words sound too physical for mathematicians... But worse, the author forgot to say that the particles do not interact.'

Well, I did not forget it. I insist that my models are kinematical, not dynamical. Energy is just irrelevant. Maybe the particles interact. Maybe, an external field is present. The entropy is $k_{\mathrm{B}} n \big(\frac{3}{2} \ln E + \ln V \big)$ anyway, provided that the macrostate is specified by the kinetic energy (only). In this sense I am right. However, the physicist is also right! We have no feasible way to prepare the macrostate mentioned above, if the corresponding microstates are of different (total) energy.

See also Sect. 6d below, and the book [3]: Sect. 2.1 about the notion of entropy, Sect. 4.3 about the continuous model, and Sect. 7.2.1 about the discrete model.

6b Entropy in physics and entropy in mathematics

A discrete probability measure, consisting of a finite or countable set of atoms,[1] has the entropy

$$S = - \sum_n p_n \ln p_n$$

(just by definition); here p_n are the probabilities of the atoms.

The question is, can we relate this 'mathematical' entropy to the 'physical' entropy of 6a?

Recall the general scheme of 6a,

$$X_n : \Omega_n \to \mathbb{R} \, , \quad \| f(X_n(\cdot)) \|_{L_n(c_n P_n)} \to \max_{\mathbb{R}} \big(|f| \mathrm{e}^{-I} \big)$$

for some lower semicontinuous $I : \mathbb{R} \to (-\infty, +\infty]$. Assume in addition that for every n the measure P_n on Ω_n is discrete, consisting of *atoms of mass* 1 *each.*

Given an interval $(a, b) \subset \mathbb{R}$ such that $\inf_{(a,b)} I = \min_{[a,b]} I \ne +\infty$ we introduce *probability* measures Q_n on Ω_n by conditioning P_n on $a < X_n(\cdot) < b$. That is,

$$Q_n(A) = \frac{P_n \big(A \cap X_n^{-1}((a,b)) \big)}{P_n \big(X_n^{-1}((a,b)) \big)} \quad \text{for measurable } A \subset \Omega_n \, .$$

[1]Not to be confused with atoms in physics...

Clearly, Q_n consists of a finite number $P_n\big(X_n^{-1}((a,b))\big)$ of *equiprobable* atoms. Thus its entropy is

$$S(Q_n) = \ln P_n\big(X_n^{-1}((a,b))\big) .$$

Similarly to 4b6, $\frac{1}{n}\ln\big(c_n P_n\big(X_n^{-1}((a,b))\big) \to -\min_{[a,b]} I$, therefore

$$S(Q_n) = -n \min_{[a,b]} I - \ln c_n + o(n) .$$

Given $x \in \mathbb{R}$ such that $I(x) \ne +\infty$, and $\varepsilon > 0$, we may condition on $x - \varepsilon < X_n(\cdot) < x + \varepsilon$ getting probability measures $Q_{n,\varepsilon}$. Similarly to 4b12,

$$S(Q_{n,\varepsilon}) = -nI(x) - \ln c_n + o(n) \tag{6b1}$$

as $\varepsilon \to 0+$, uniformly in $n \ge n_\varepsilon$.

For a small ε and large $n \ge n_\varepsilon$ the measure $Q_{n,\varepsilon}$ may be thought of as a macrostate corresponding to $X \approx x$. According to 6a (especially, (6a6)), its 'physical' entropy is

$$S = -k_{\mathrm{B}} n I(x) + \mathrm{const}$$

where 'const' does not depend on x, but may depend on n. The similarity to (6b1) is evident.

Why does the measure consist of atoms of mass 1? These appear naturally in the discrete model, bot not in the continuous model! Well, in fact they do appear also in the continuous model... after quantization. Physicists call them 'phase cells'. A large region in the phase space (smoothed a bit) corresponds to an operator in the Hilbert space, close to a projection onto a subspace whose dimension is close to the volume of the domain (if the units are chosen so that the Planck constant is equal to 1). The trace of this operator is the quantal counterpart of the classical Liouville measure of the region. See also [2, Sect. 3.5 and Exercise (7.3)] and [3, Sect. 2.1]. Quantization is simpler for spins (the discrete model); each point of $\Omega_n = \{0,1\}^n$ corresponds to a one-dimensional subspace of the 2^n-dimensional Hilbert space, and these subspaces are orthogonal.

6c Increase of entropy

Any real engine will create net entropy during a cycle; no engine can reduce the net amount of entropy in the Universe.

Sethna [2, p. 80].

We consider the general scheme of 6a, but restrict ourselves to compact spaces and probability measures. Namely, $(\Omega_n, \mathcal{F}_n, P_n)$ are probability spaces, K is a compact metrizable space, $X_n : \Omega_n \to K$ are measurable maps, $\mu_n \in P(K)$ is the distribution of X_n.

We assume that $(\mu_n)_n$ satisfies LDP with a *continuous* rate function $I : K \to [0, \infty)$.

A new ingredient is dynamics: measure preserving maps $T_n : \Omega_n \to \Omega_n$.

Let $G \subset K$ be an open set (not empty). For each n we choose an initial state $\omega_n \in \Omega_n$ at random among the points satisfying $X_n(\omega_n) \in \overline{G}$. We know (recall 4c4) that $I(X_n(\omega_n))$ is close to $\min_{\overline{G}} I$ with high probability, if n is large. It appears that the same holds for $\max\big(I(X_n(\omega_n)), I(X_n(T\omega_n)), I(X_n(T^2\omega_n)), \dots, I(X_n(T^{t_n}\omega_n))\big)$, if t_n grows not too fast. (However, it does not mean that $I(X_n(T\omega_n))$ is close to $\min_{\overline{G}} I$.)

6c1 Proposition. Let $t_1, t_2, \dots \in \{1, 2, \dots\}$ satisfy $(t_n)^{1/n} \to 1$. Then

$$\frac{P_n\big(X_n^{-1}(\overline{G}) \cap \{\omega_n \in \Omega_n : \forall t \le t_n \quad I(X_n(T^t\omega_n)) \le \varepsilon + \min_{\overline{G}} I\}\big)}{P_n\big(X_n^{-1}(\overline{G})\big)} \to 1$$

as $n \to \infty$, for every $\varepsilon > 0$.

Proof. Denoting for convenience $c = \min_{\overline{G}} I$ we have

$$\begin{aligned} P_n\big(X_n^{-1}(\overline{G}) \cap \{\omega_n \in \Omega_n : I(X_n(T\omega_n)) > c + \varepsilon\}\big) &\le \\ \le P_n\big(\{\omega_n \in \Omega_n : I(X_n(T\omega_n)) > c + \varepsilon\}\big) &= \\ = P_n\big(\{\omega_n \in \Omega_n : I(X_n(\omega_n)) > c + \varepsilon\}\big) &= \\ = \mu_n\big(\{x \in K : I(x) > c + \varepsilon\}\big)& . \end{aligned}$$

By 4b6 (and continuity of I),

$$\lim_n \big(\mu_n(\{x \in K : I(x) > c + \varepsilon\})\big)^{1/n} \le \mathrm{e}^{-(c+\varepsilon)} \,,$$
$$\lim_n \big(\mu_n(\overline{G})\big)^{1/n} = \mathrm{e}^{-c} \,.$$

Therefore

$$\begin{aligned} \limsup_n \bigg(\frac{P_n\big(X_n^{-1}(\overline{G}) \cap \{\omega_n \in \Omega_n : I(X_n(T\omega_n)) > c + \varepsilon\}\big)}{P_n\big(X_n^{-1}(\overline{G})\big)} \bigg)^{1/n} &\le \\ \le \limsup_n \bigg(\frac{\mu_n\big(\{x \in K : I(x) > c + \varepsilon\}\big)}{\mu_n(\overline{G})} \bigg)^{1/n} \le \frac{\mathrm{e}^{-(c+\varepsilon)}}{\mathrm{e}^{-c}} = \mathrm{e}^{-\varepsilon} \,. \end{aligned}$$

The same holds for T^t (in place of T) for each t. Taking into account that $(t_n)^{1/n} \to 1$ we get

$$\limsup_n \left(\frac{P_n\big(X_n^{-1}(\overline{G}) \cap \{\omega_n \in \Omega_n : \exists t \le t_n \quad I(X_n(T^t \omega_n)) > c + \varepsilon\}\big)}{P_n\big(X_n^{-1}(\overline{G})\big)} \right)^{1/n} \\ \le \limsup_n \left(\frac{(t_n + 1)\mu_n\big(\{x \in K : I(x) > c + \varepsilon\}\big)}{\mu_n(\overline{G})} \right)^{1/n} \le \mathrm{e}^{-\varepsilon} < 1 \, .$$

□

In the light of (6a6), decrease of the rate function means increase of the entropy. We see that dynamics does not reduce the entropy as compared with the initial state. However, it does not mean that the entropy increases (nearly) monotonically, since, say, $I(X_n(T^2\omega_n))$ may be much greater than $I(X_n(T\omega_n))$.

It is possible to get a (nearly) monotone increase of entropy by introducing a small random perturbation as follows. We start at some $\omega_n(0)$, choose $\omega'_n(0)$ at random among all points satisfying $\operatorname{dist}\big(X_n(\omega'_n(0)), X_n(\omega_n(0))\big) < \varepsilon$, and jump to $\omega_n(1) = T\omega'_n(0)$. Then we choose $\omega'_n(1)$ at random among all points satisfying $\operatorname{dist}\big(X_n(\omega'_n(1)), X_n(\omega_n(1))\big) < \varepsilon$ and jump to $\omega_n(2) = T\omega'_n(1)$. And so on. The random choices are mutually independent and governed by the measure P_n conditioned as prescribed. A metric on K should be chosen.

Of course, a physicist would prefer an autonomous deterministic dynamics; alas, we are unable to deduce monotone increase of entropy without additional hypotheses.

6d The energy and the temperature

The scheme of $X_n : \Omega_n \to K$ is too general for physics. In order to be a bit more specific we should consider the energy $E_n : \Omega_n \to \mathbb{R}$. The corresponding macrostates appear naturally. About the two models of 6a: (6a2) is the energy of the simple ideal gas (particles do not interact), and (6a1) is the energy of noninteracting spins in an external magnetic field.

We may combine both models into a single physical system: the gas *and* the spins.[1] It means the product of measure spaces,

$$(\Omega_n, \mathcal{F}_n, P_n) = (\Omega'_n, \mathcal{F}'_n, P'_n) \times (\Omega''_n, \mathcal{F}''_n, P''_n) \, ,$$

[1] The particles with spins and the particles of the gas may be the same particles, or not the same. For simplicity I assume that the number n of particles is the same for both subsustems.

and nearly additive energy,

$$E_n(\omega_n) = E_n(\omega_n', \omega_n'') = E_n'(\omega_n') + E_n''(\omega_n'') + E_n^{\text{int}}(\omega_n', \omega_n'') \,,$$

the interaction energy E_n^{int} being small,[1]

$$\sup_{\omega_n', \omega_n''} |\tfrac{1}{n} E_n^{\text{int}}(\omega_n', \omega_n'')| \to 0 \quad \text{as } n \to \infty \,.$$

We know the rate function for each subsystem; denoting $r = \frac{E'}{nE_1'}$,

$$I'(\tfrac{1}{n}E') = \frac{1+r}{2} \ln \frac{1+r}{2} + \frac{1-r}{2} \ln \frac{1-r}{2} + \text{const} \,,$$
$$I''(\tfrac{1}{n}E'') = -\frac{3}{2} \ln \tfrac{E''}{n} + \text{const}_n$$

(recall (6a4), (6a5)). Similarly to 4d we get the rate function $(\frac{1}{n}E', \frac{1}{n}E'') \mapsto I'(\frac{1}{n}E') + I''(\frac{1}{n}E'')$ for the combined system. Similarly to the contraction principle 2b1, the rate function I for the total energy E is

$$I(\tfrac{1}{n}E) = \min\{ I'(\tfrac{1}{n}E') + I''(\tfrac{1}{n}E'') : \tfrac{1}{n}E' + \tfrac{1}{n}E'' = \tfrac{1}{n}E \} \,.$$

Due to 5d (slightly generalized), the small interaction energy does not invalidate this relation.

Along the line $\frac{1}{n}E' + \frac{1}{n}E'' = \frac{1}{n}E$ the sum $I'(\frac{1}{n}E') + I''(\frac{1}{n}E'')$ has a single minimum (since both functions are strictly convex). Similarly to 4c4, the macrostate corresponding to E is concentrated near the point of minimum. That is, the given total energy E decomposes into $E' + E''$ according to the point of minimum. The latter can be found by differentiation,

$$\text{(6d1)} \qquad \frac{\mathrm{d} I'(\frac{1}{n}E')}{\mathrm{d}\frac{1}{n}E'} = \frac{\mathrm{d} I''(\frac{1}{n}E'')}{\mathrm{d}\frac{1}{n}E''} \,.$$

Specifically,

$$-\frac{1}{k_{\mathrm{B}} T'} = \frac{1}{E_1'} \cdot \frac{1}{2} \ln \frac{1+r}{1-r} = -\frac{3}{2} \frac{n}{E''} = -\frac{1}{k_{\mathrm{B}} T''} \,.$$

In physics,

$$\frac{\mathrm{d}(\text{entropy})}{\mathrm{d}(\text{energy})} = \frac{1}{(\text{temperature})}$$

[1] The mathematics is easier with $E_n^{\text{int}} = 0$, however, physically it means that the two subsustems do not interact, which is uninteresting.

(see [2, p. 46] and [3, Sect. 2.1 and 2.3.1]). Thus, (6d1) means that the two subsystems are at the same temperature! This is well-known as the equilibrium. Starting at equilibrium, dynamical evolution does not lead to different temperatures (unless you wait an exponentially long time).

Specifically, for the ideal gas the temperature is $T'' = \frac{2E''}{3k_{\mathrm{B}}n}$, which looks quite believable. For the spin system, $T' = \frac{2E'_1}{k_{\mathrm{B}}n \ln \frac{1-r}{1+r}}$, which is rather counterintuitive for a non-physicist.[1]

Thermal and mechanical energy. We may also combine one of the two multiparticle systems (gas, or spins) with a few-particle system, say, just one particle (maybe, quite massive).[2] It means

$$(\Omega_n, \mathcal{F}_n, P_n) = (\Omega'_n, \mathcal{F}'_n, P'_n) \times (\Omega'', \mathcal{F}'', P'') \, .$$

Thus, $I''(\cdot) = \text{const}$ (on the support of the measure; other points are irrelevant).[3] The minimum of $I'(E') + I''(E'')$ on the line $E' + E'' = E$ is reached when E'' is minimal. This is the equilibrium: all the energy is thermal (heat), not mechanical. Starting at equilibrium, dynamical evolution does not convert any heat into mechanical energy (unless you wait an exponentially long time). See also [2, p. 81] about the heat death of the Universe.

References

[1] R.S. Ellis, *Entropy, large deviations, and statistical mechanics,* Springer, 1985.

[2] J.P. Sethna, *Statistical mechanics: entropy, order parameters, and complexity,* Oxford, 2006.

[3] D. Yoshioka, *Statistical physics: an introduction,* Springer, 2007.

[1]Negative temperature is really possible in such systems, especially in lasers! See [2, Exercise (6.3)].

[2]It may be also a collective degree of freedom of the multiparticle system.

[3]Which could be the definition of a mechanical system in our framework.