1. Let $S_n := \sum_{i=1}^n X_i$ be the empirical sum of iid random variables $X_1, X_2, \ldots$ in $\mathbb{R}$. In the proof of the large-deviation principle for the empirical average $\frac{1}{n} S_n$, use Cramér's Theorem to show the alternative lower bound for the case where $x < \mathbb{E}X_1$ and $\epsilon > 0$ is sufficiently small, that is:

$$\text{(LB')} \qquad \liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big(\tfrac{1}{n} S_n \in B_\epsilon(x)\big) \geq -I(x).$$

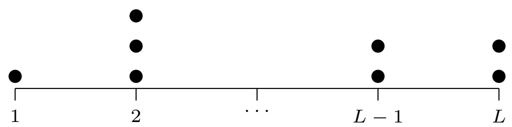
2. For any two finite measures $\rho, \nu \in \mathcal{M}(\mathcal{X})$ on a measurable space, the *relative entropy* is defined as

$$\mathcal{H}(\rho \mid \nu) := \begin{cases} \int_{\mathcal{X}} \big(\rho(dx) \log \frac{d\rho}{d\nu}(x) - \rho(dx) + \nu(dx)\big), & \rho \ll \nu, \\ \infty, & \text{otherwise}, \end{cases}$$

Show that $\inf_\rho \mathcal{H}(\rho \mid \nu) = 0$.

The next exercise does not require much preknowledge, but it shows a deep connection between large deviations, thermodynamics and information theory.

3. Consider n 'particles' $x_1, \ldots, x_n$ on a finite space $\{1, \ldots, L\}$. We first study information theoretic aspects and later introduce randomness. In thermodynamics/statistical mechanics one often distinguishes between *microscopic states* and *macroscopic states*. In this setting, one may think of all particle coordinates $x = (x_1, \ldots, x_n)$ as a micro state (where all particles are distinct), and its corresponding empirical measure $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i}$ as the macro state (where only the number of particles at each site matters), see picture:



(If you are having trouble with this exercise just take $L = 2$.)

(a) Write down the Boltzmann entropy of a given macro state $\rho^n \in (\frac{1}{n}\mathbb{N})^L \cap \mathcal{P}(\{1, \ldots, L\})$, defined as:

$$\mathrm{Ent}_n(\rho^n) := k_B \log \#\Omega_n(\rho^n), \qquad \text{where}$$
$$\Omega_n(\rho^n) := \Big\{\text{micro states } x \in \{1, \ldots, L\}^n : \text{macro state } \tfrac{1}{n} \textstyle\sum_{i=1}^n \mathbb{1}_{x_i} = \rho^n\Big\},$$

and k_B is the Boltzmann constant.

(b) The Boltzmann entropy blows up as $n \to \infty$, so for large particle numbers one usually takes the average entropy per particle. Calculate this limit (formally):

$$\mathrm{Ent}(\rho) := \lim_{n\to\infty} \tfrac{1}{n} \mathrm{Ent}_n(\rho^n),$$

assuming each $\rho^n \in (\frac{1}{n}\mathbb{N})^L \cap \mathcal{P}(\{1, \ldots, L\})$ and $\rho^n \to \rho \in \mathcal{P}(\{1, \ldots, L\})$.

Remark: the resulting formula is what physicists usually use as the entropy, whereas mathematicians usually drop the constant and flip the minus sign.

The Boltzmann entropy is purely combinatoric/information theoretic, so it is only a useful concept if all sites $1, \ldots, L$ are equally likely. In general one needs something more advanced. Let $X_1, X_2, \ldots$ be iid random variables with values in the finite set $\{1, \ldots, L\}$, with probabilities $\mathbb{P}(X_1 = l) =: \nu_l$. Now the *empirical measure* $L^n := \dfrac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i}$ is a random variable in $\mathcal{P}(\{1, \ldots, L\})$.

(c) For an abitrary test 'function' $f \in \mathbb{R}^L$, what is the weak limit of $f \cdot L_n$ as $n \to \infty$? No proof needed, but where does L^n weakly converge to?

(d) Write down the probability that $L^n = \rho^n$ for a given $\rho^n \in (\frac{1}{n}\mathbb{N})^L \cap \mathcal{P}(\{1, \dots, L\})$.

(e) Again assuming $\rho^n \to \rho \in \mathcal{P}(\{1, \dots, L\})$, (formally) show that:

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(L^n = \rho^n) = \mathcal{H}(\rho \mid \nu).$$

(f) In physics, one often has probabilities of the form

$$\nu_l = \frac{1}{Z_T} e^{-V_l/(k_B T)},$$

where V_l is some energy function, T is the temperature, and Z_T is a normalisation constant. Show that

$$\lim_{n\to\infty} -\frac{k_B T}{n} \log \mathbb{P}(L^n = \rho^n) = U(\rho) - T\mathrm{Ent}(\rho) + \mathrm{const}$$

for some function $U(\rho)$ and a constant not depending on ρ.

Remark: physicists call $U(\rho)$ the internal energy, $U(\rho) - T\mathrm{Ent}(\rho)$ the Helmholtz free energy, and they don't care about the constant.