Reciprocity laws and covariances

We can easily derive relationships between partial derivatives of the constraints F_k and Lagrange multipliers λ_k. By changing the order of partial differentiations we get

$$\left.\frac{\partial F_k}{\partial \lambda_j}\right|_{\{\lambda\}} = \left.\frac{\partial^2 - \log Z}{\partial \lambda_j \partial \lambda_k}\right|_{\{\lambda\}} = \left.\frac{\partial^2 - \log Z}{\partial \lambda_k \partial \lambda_j}\right|_{\{\lambda\}} = \left.\frac{\partial F_j}{\partial \lambda_k}\right|_{\{\lambda\}} \tag{13}$$

Similarly

$$\left.\frac{\partial \lambda_k}{\partial F_j}\right|_{\{F\}} = \left.\frac{\partial^2 S}{\partial F_j \partial F_k}\right|_{\{F\}} = \left.\frac{\partial^2 S}{\partial F_k \partial F_j}\right|_{\{F\}} = \left.\frac{\partial \lambda_j}{\partial F_k}\right|_{\{F\}} \tag{14}$$

By cursory observation one might say the second equation is just the inverse of the first one, so it is not telling anything new. This is wrong, as the quantities that are kept fixed at differentiation are not the same. However, the naive notion of inverse holds in a more intricate way: the matrices with elements $A_{jk} = \partial F_j / \partial \lambda_k$ and $B_{jk} = \partial \lambda_j / \partial F_k$ are inverses of each other: $A = B^{-1}$.

When we set $\langle f_k(X) \rangle = F_k$, we required that *on average* $f_k(X)$ is what is prescribed, but still it varies from observation to observation. Now we look at how large these fluctuations are.

The *covariance* of two random variables is defined as

$$\mathrm{Cov}(X,Y) \stackrel{\mathrm{def}}{=} \langle [X - \langle X \rangle][Y - \langle Y \rangle] \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$$

which is a measure of "how much Y is above its average at the same time when X is above its average". A covariance of a random variable with itself is called *variance*:

$$\mathrm{Var}(X) \stackrel{\mathrm{def}}{=} \mathrm{Cov}(X,X) = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$$

with the convenient meaning that its square root (the *standard deviation* σ) measures how much a random variable differs from its average, suitably weighted. (The variance is always non-negative, as it is the average of a non-negative quantity: a square.)

So we can calculate the covariance of $f_k(X)$ and $f_j(X)$:

$$\mathrm{Cov}(f_j(X), f_k(X)) = \langle f_j(X) f_k(X) \rangle - \langle f_j(X) \rangle \langle f_k(X) \rangle$$

The first term using (9) is

$$\langle f_j(X) f_k(X) \rangle = \frac{1}{Z} \sum_{i=1}^{n} f_j(x_i) f_k(x_i) \exp\left(- \sum_{\ell=1}^{m} \lambda_\ell f_\ell(x_i) \right) = \frac{1}{Z} \frac{\partial^2 Z(\lambda_1, \ldots, \lambda_m)}{\partial \lambda_j \partial \lambda_k}$$

As a side remark, the above calculation easily generalises to averages of arbitrary products of f_ks:

$$\langle f_j^{m_j}(X) f_k^{m_k}(X) \ \cdots \rangle = \frac{1}{Z} \left(\frac{\partial^{m_j}}{\partial \lambda_j^{m_j}} \frac{\partial^{m_k}}{\partial \lambda_k^{m_k}} \ \cdots \right) Z$$

Coming back to the covariance

$$\begin{aligned} \mathrm{Cov}(f_j(X), f_k(X)) &= \frac{1}{Z} \frac{\partial^2 Z}{\partial \lambda_j \partial \lambda_k} - \frac{1}{Z^2} \frac{\partial Z}{\partial \lambda_j} \frac{\partial Z}{\partial \lambda_k} = \frac{\partial^2 \log Z}{\partial \lambda_j \partial \lambda_k} \\ &= -\frac{\partial F_k}{\partial \lambda_j} = -\frac{\partial F_j}{\partial \lambda_k} \end{aligned} \tag{15}$$

where we have seen the last steps already in (13). Similarly for variance

$$0 \le \mathrm{Var}(f_k(X)) = \frac{\partial^2 \log Z}{\partial \lambda_k^2} = -\frac{\partial F_k}{\partial \lambda_k} \tag{16}$$

This confirms that the second derivative of $\log Z$ is non-negative, ie. $\log Z$ is a convex function, which we implicitly assumed when mentioned that $-\log Z$ and S are Legendre transforms of each other.

Suppose now that the constraint functions f_k depend on an external parameter: $f_k(X;\alpha)$. Everything, including Z and S become dependent on α. To see its effect we calculate partial derivatives:

$$\begin{aligned} -\frac{\partial \log Z}{\partial \alpha}\bigg|_{\{\lambda\}} &= -\frac{1}{Z}\sum_{i=1}^{n} \exp\left(-\sum_{k=1}^{m}\lambda_k f_k(x_i;\alpha)\right)\sum_{k=1}^{m} -\lambda_k \frac{\partial f_k(x_i;\alpha)}{\partial \alpha} \\ &= \sum_{k=1}^{m}\lambda_k \left\langle \frac{\partial f_k}{\partial \alpha}\right\rangle \end{aligned} \tag{17}$$

Similarly, using $S = \log Z + \sum_k \lambda_k F_k$

$$\begin{aligned} \frac{\partial S(F_1,\ldots,F_n;\alpha)}{\partial \alpha}\bigg|_{\{F\}} &= \sum_{k=1}^{m} \underbrace{\frac{\partial \log Z}{\partial \lambda_k}\bigg|_{\{\lambda\}}}_{-F_k} \frac{\partial \lambda_k}{\partial \alpha}\bigg|_{\{F\}} + \frac{\partial \log Z}{\partial \alpha}\bigg|_{\{\lambda\}} + \sum_{k=1}^{m}\frac{\partial \lambda_k}{\partial \alpha}\bigg|_{\{F\}} F_k \\ &= \frac{\partial \log Z}{\partial \alpha}\bigg|_{\{\lambda\}} \end{aligned}$$

So the partial derivatives of $\log Z$ and S with respect to α are equal, though one should note that the variables kept fixed are the natural variables in each case.

Applications of the maximum entropy framework

The microcanonical ensemble

The simplest system to consider is the isolated one, with no interaction with its environment. A physical example can be a thermally and mechanically isolated box containing some gas, traditionally these are called *microcanonical ensembles*. With no way to communicate, we have no information about the current state of the system. To put it in the maximum entropy framework, we do not have any constraint to apply.

The maximum entropy solution for such a system is

$$Z = \sum_{i=1}^{n} 1\,, \qquad p_i = \frac{1}{Z}\,, \qquad S = \log Z$$

Using the conventions of statistical physics the number of states is denoted by Ω, and the unit of entropy is k_B: recall this sets the prefactor and/or the base of the logarithm in (2)-(3). Using this notation (the MC subscript denotes microcanonical):

$$Z = \Omega\,, \qquad p_i = \frac{1}{Z} = \frac{1}{\Omega}\,, \qquad S_{\text{MC}} = k_B \ln \Omega$$

In this most simple system all internal states have equal probability.

The canonical ensemble

In the next level of increasing complexity, we allow the exchange of one conserved quantity with the external environment. The physical example is a system which is thermally coupled (allowing energy exchange) with its environment; traditionally these are called *canonical ensembles*. Using this terminology we label the internal states with their energy. By having the ability to interact with the system, we can control eg. the average energy of the system by changing the condition of the environment, corresponding to having one constraint in the maximum entropy formalism.

The maximum entropy solution for one constraint reads

$$Z(\lambda) = \sum_{i=1}^{n} e^{-\lambda f(x_i)}\,, \qquad p_i = \frac{1}{Z} e^{-\lambda f(x_i)}\,, \qquad S(F) = \log Z(\lambda) + \lambda F$$

The conventional units for entropy is k_B for canonical ensembles as well, and as we mentioned the states are labelled with energy: $f(x_i) = E_i$ with average energy (the value of the constraint) $F = E$. Finally, by convention the Lagrange multiplier λ is called $\beta = 1/(k_B T)$ in statistical physics, where T is temperature (measured in Kelvins), and k_B is the Boltzmann constant. Thus we have

$$Z(\beta) = \sum_{i=1}^{n} e^{-\beta E_i}\,, \qquad p_i = \frac{1}{Z} e^{-\beta E_i} = \frac{1}{Z} \exp\left(-\frac{E_i}{k_B T}\right)\,, \qquad S_C(\langle E\rangle) = k_B \ln Z + \frac{\langle E\rangle}{T}$$

In p_i the exponential factor $e^{-\beta E_i}$ is called Boltzmann factor, while Z provides the normalisation.

Having established this connection, we can easily translate the results of the maximum entropy formalism. Eqs. (8) and (11) become

$$\langle E\rangle = -\frac{\partial \ln Z}{\partial \beta} \qquad \text{and} \qquad \frac{1}{T} = \frac{\partial S_C}{\partial \langle E\rangle}$$

Eq. (16) gives the energy fluctuation:

$$\sigma_E^2 = \mathrm{Var}(E) = \frac{\partial^2 \ln Z}{\partial \beta^2} = -\frac{\partial \langle E\rangle}{\partial \beta} = \underbrace{\frac{\partial \langle E\rangle}{\partial T}}_{C_V} k_B T^2$$

where C_V is the *heat capacity* of the system. This is an interesting relation, connected microscopic fluctuations with macroscopic thermodynamic quantities.

In practice it is useful to define the following quantity, called *Helmholtz free energy*:

$$A \stackrel{\text{def}}{=} -k_B T \ln Z = \langle E\rangle - T S_C\,,$$

If we consider it as a function of temperature, $A(T)$, its derivative is

$$\frac{\partial A}{\partial T} = -k_B \ln Z - k_B T \frac{\partial \ln Z}{\partial \beta} \frac{1}{k_B T^2} = -S_C$$

This leads to a relation with the energy. Our approach so far determined the entropy S_C as a function of average energy $\langle E\rangle$. Considering its inverse function $\langle E\rangle(S_C)$, we see that its Legendre transform is $-A(T)$.

It is interesting to note that

$$\sum_i \exp\left(-\frac{E_i}{k_B T}\right) = Z = \exp\left(-\frac{A}{k_B T}\right)\,,$$

so the sum of Boltzmann factors equals to a single Boltzmann factor with energy replaced with the Helmholtz free energy. We will see its implications later in the grand canonical ensemble.

Next we consider a system made of two subsystems, which are sufficiently uncoupled. The joint partition function can be written as [labeling the left and right subsystem with (L) and (R)]:

$$Z = \sum_i \sum_j \exp -\beta\left(E_i^{(L)} + E_j^{(R)}\right) = \left(\sum_i \exp -\beta E_i^{(L)}\right)\left(\sum_j \exp -\beta E_j^{(R)}\right) = Z^{(L)} Z^{(R)}$$

This means that $\ln Z$ is additive: $A = -k_B T \log Z = A^{(L)} + A^{(R)}$. Other quantities, like the entropy or the energy have the similar additive property, and we call these *extensive* quantities.

Physical examples for canonical ensembles

We have seen that the way to calculate any statistical mechanics quantity for a given system is to calculate first the partition function, and then any other quantity is easily expressed. Consider, for example, a particle with position x and momentum $p = mv$, and energy $E = p^2/(2m) + U(x)$, where U is the potential. In systems made of discrete states the formula involves a sum over the states. For continuous systems, however, the sum needs to be replaced by integration:

$$\sum_i (\cdot) \quad \leftrightarrow \quad \frac{1}{h} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp \, (\cdot) \tag{18}$$

This is a *semiclassical* formula: not quantum mechanical, as x and p are independent variables and not non-commuting operators; but not purely classical either as the Plack constant h is involved. Instead of fully understanding, we just rationalise this formula as (i) a constant needs to appear in front of the integrals to make the full expression dimensionless, as Z should be, and (ii) in quantities involving $\log Z$ the prefactor $1/h$ becomes an additive constant, and in particular for the entropy it sets its zero level.