- HH0104 -

How to Quantify Uncertainty in a Statistical System?

Hsiu-Hau Lin
hsiuhau@phys.nthu.edu.tw
(Dec 31, 2013)

When the degrees of freedom for the interested system are small, we are often capable to write down the equations of motion and solve for the corresponding trajectories. However, when degrees of freedom are huge, finding the exact trajectories is not very illuminating and is literally impossible for realistic statistical systems. Take ideal gas as an example. Knowing the probability to find a particle with speed v,

$$P(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 \exp\left[-\frac{mv^2}{2kT}\right], \tag{1}$$

is more insightful than finding the zigzag paths for each molecule in the ideal gas. That is to say, we no longer wish to know the exact evolution of the velocity $\boldsymbol{v}(t)$. Instead, we treat them as random variables described by some definite distribution.

• random variables

Let us start with the simplest Bernoulli distribution. Consider a random variable $X = 0, 1$ with binary values described by the following probability distribution,

$$P(X) = \begin{cases} p, & X = 1; \\ 1 - p, & X = 0. \end{cases} \tag{2}$$

According to the classical probability theory, $0 \leq P(X) \leq 1$ and satisfies the sum rule $\sum_X P(X) = 1$. The average of the random variable X to the n-th power is defined as

$$\langle X^n \rangle \equiv \sum_X P(X) X^n. \tag{3}$$

It is straightforward to compute the average $\langle X \rangle$ for Bernoulli distribution,

$$\langle X \rangle = p \cdot 1 + (1 - p) \cdot 0 = p. \tag{4}$$

The variance ΔX for Bernoulli distribution can also be calculated easily,

$$\Delta X \equiv \sqrt{\langle X^2 \rangle - \langle X \rangle^2} = \sqrt{p(1-p)}. \tag{5}$$

The variance is usually a good indicator for fluctuations deviating from the average. You can see that ΔX vanishes for $p = 0, 1$ because the random variable is no longer random.

In many cases, the random variable is continuous and the summation is replaced by the integral,

$$\langle X^n \rangle \equiv \int dX \; P(X) X^n. \tag{6}$$

Here $P(X)$ is probability density, satisfying the sum rule (now an integral),

$$\int dX \; P(X) = 1. \tag{7}$$

The interpretation of the probability density is subtle: $P(X)dX$ denotes the probability to find the random variable falling into the regime between X and $X + dX$. Thus, $P(X)$ is not dimensionless but carries the reciprocal dimension of the random variable X.

• Shannon entropy

For Bernoulli distribution, it is clear that randomness disappears at $p = 0, 1$. However, for general $0 < p < 1$, how can we quantify the uncertainty for the random variable? It turns out that the uncertainty of a statistical system can be characterised by the Shannon entropy,

$$\sigma \equiv \langle -\ln P \rangle = \sum_s -P_s \ln P_s. \tag{8}$$

Let's compute the Shannon entropy for the Bernoulli distribution,

$$\sigma = -p \ln p - (1-p) \ln(1-p). \tag{9}$$

Making use of $\lim_{x \to 0^+}(x \ln x) = 0$, one can show that the Shannon entropy vanishes at $p = 0, 1$ as expected. The maximum value of the Shannon entropy occurs at $p = 1/2$ for Bernoulli distribution,

$$\sigma_{\max} = \ln 2. \tag{10}$$

This also agrees with our common sense. The definition presented here leads to additive property of the Shannon entropy for independent systems. Consider two independent random variables X, Y with probability distribution

$$P(X,Y) = f(X)g(Y), \tag{11}$$

where $f(X)$ and $g(Y)$ are the probability distributions for the corresponding random variables. According to the definition, the Shannon entropy is

$$\sigma = \langle -\ln P\rangle = \langle -\ln f - \ln g\rangle = \sigma_x + \sigma_y. \tag{12}$$

In consequence, the total entropy for two independent statistical systems is the sum of the individual entropies.

• dog-flea model

It is interesting to explore the dynamics of Shannon entropy in the so-called dog-flea model. Consider two dogs named $X = 0$ and $X = 1$. For simplicity, we treat time as discrete integers, $t = 0, 1, 2, ...$ and the probability to find the flea in dog $X = 1$ at time t is the time-dependent probability function $P(t)$. At each time step, the flea hops from one dog to the other with a transition probability $1/\tau$. Suppose the flea rests in dog $X = 0$ initially. We may guess the time evolution of the probability function follows the trend:

$$P(0) = 0 \quad \text{gradually evolves into} \quad P(t \to \infty) \equiv P_\infty = \frac{1}{2}. \tag{13}$$

When the dynamical equilibrium is reached at the end, we expect the probabilities to find the flea on either dog should be the same, i.e. $P_\infty = 1/2$.

Let us try to formulate the dynamical process with mathematical rigour. The probability at later time $P(t + 1)$ can be expressed in terms of $P(t)$ as long as the flea hops without referring to its previous memory,

$$P(t+1) = P(t)\left(\frac{\tau - 1}{\tau}\right) + [1 - P(t)]\,\frac{1}{\tau}. \tag{14}$$

The above equation for $P(t)$ is called the master equation for the stochastic processes. The first term accounts for the event when the flea was in dog $X = 1$ at time t already and remains in the same spot without hopping. On the other hand, the second term is attributed to the event when the flea was in dog $X = 0$ at time t and hops to dog $X = 1$.

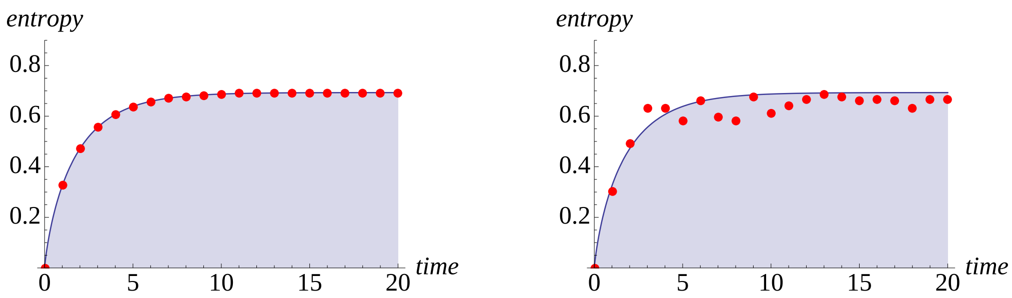

Figure 1: Entropy evolution $\sigma(t)$ for the dog-flea model with $\tau = 10$. The solid line is the prediction from the master equation. The red dots are simulated by $N = 1000$ fleas by computer program (left panel) and by $N = 68$ students in the classroom (right panel). Take $N = 1000$ fleas in the computer program and make them hop between $X = 0$ and $X = 1$ with the transition probability $1/\tau = 1/10$. The probability function $P(t)$ can be constructed by counting the fraction of the fleas in dog $X = 1$. The numerical results agree exceedingly well with the prediction from the master equation. We can also perform the simulation by the game demonstrated in the classroom. For $N = 68$ students, the outcome is still captured by the master equation but with visible fluctuations. Note that, for $t > \tau/2$, the Shannon entropy is very close to its maximum value $\sigma_{\max} = \ln 2$.

Let us try to find the solution in equilibrium in the infinite-time limit first. Because the probability function ceases to change,

$$\lim_{t\to\infty} P(t+1) = \lim_{t\to\infty} P(t) = P_\infty. \tag{15}$$

The time dependence in the master equation drops out and it turns into an algebraic equation for the stationary probability,

$$P_\infty = P_\infty \left(\frac{\tau - 1}{\tau}\right) + [1 - P_\infty]\,\frac{1}{\tau}, \qquad \to \qquad P_\infty = \frac{1}{2}. \tag{16}$$

The master equation indeed hosts the equilibrium solutions as we guessed beforehand. Now we are ready to solve the time evolution of $P(t)$. Let us massage the master equation into simpler form by changing variable,

$$Q(t) \equiv P(t) - P_\infty = P(t) - \frac{1}{2}. \tag{17}$$

After some algebra, the master equation takes a much simpler form,

$$Q(t+1) = \left(\frac{\tau - 2}{\tau}\right) Q(t). \tag{18}$$

The time evolution of $Q(t)$ can be computed by iterations and the solution is nothing but the simple geometric sequence,

$$Q(t) = \left(\frac{\tau - 2}{\tau}\right) Q(t-1) = \cdots = \left(\frac{\tau - 2}{\tau}\right)^t Q(0). \tag{19}$$

Switch back to the original variable $P(t) = 1/2 + Q(t)$ and the solution is

$$P(t) = \frac{1}{2} - \frac{1}{2}\left(\frac{\tau - 2}{\tau}\right)^t, \tag{20}$$

where $Q(0) = -1/2$ from the initial condition $P(0) = 0$ is used. It is interesting to observe that $P(t) \to P_\infty$ in exponential fashion. As shown in Figure 1, the Shannon entropy for the dog-flea model can be computed from the probability function $P(t)$,

$$\sigma(t) = -P(t) \ln P(t) - [1 - P(t)] \ln[1 - P(t)]. \tag{21}$$

It is peculiar that the second law of thermodynamics is at work here and the Shannon entropy is a monotonically increasing function to its maximum value $\sigma_{\max} = \ln 2 \approx 0.693$.

• continuous-time limit

Now we would like to derive the master equation in continuous-time limit. Rewrite the master equation into the following form,

$$\frac{Q(t+1) - Q(t)}{\Delta t} = -\frac{2}{\tau}\, Q(t), \tag{22}$$

where $\Delta t = 1$ for the discrete time steps. If the parameter $\tau \gg \Delta t = 1$, the change in each time step is small and the difference can be approximated by the differential. The resultant differential equation looks very familiar,

$$\frac{dQ}{dt} = -\frac{1}{\tau_{eq}}\, Q, \quad \to \quad Q(t) = Q(0) e^{-t/\tau_{eq}} \tag{23}$$

where $\tau_{eq} = \tau/2$ is the relaxation time to reach thermal equilibrium. Following similar steps, the probability function is

$$P(t) = \frac{1}{2} - \frac{1}{2}\, e^{-t/\tau_{eq}}. \tag{24}$$

The exponential form emerges naturally for stochastic processes without long-term memory. This turns out to be a universal feature and is independent of the model details.