1.4: General Aspects of Probability Distributions

Discrete and Continuous Distributions

Consider a system whose possible configurations $|\,n\,\rangle$ can be labeled by a discrete variable $n\in\mathcal{C}$, where $\mathcal{C}$ is the set of possible configurations. The total number of possible configurations, which is to say the *order* of the set $\mathcal{C}$, may be finite or infinite. Next, consider an ensemble of such systems, and let P_n denote the probability that a given random element from that ensemble is in the state (configuration) $|\,n\,\rangle$. The collection $\{P_n\}$ forms a *discrete probability distribution*. We assume that the distribution is *normalized*, meaning

$$\sum_{n\in\mathcal{C}} P_n = 1\ . \tag{1.4.1}$$

Now let A_n be a quantity which takes values depending on n. The average of A is given by

$$\langle A\rangle = \sum_{n\in\mathcal{C}} P_n\, A_n\ . \tag{1.4.2}$$

Typically, $\mathcal{C}$ is the set of integers ($\mathbb{Z}$) or some subset thereof, but it could be any countable set. As an example, consider the throw of a single six-sided die. Then $P_n=\frac{1}{6}$ for each $n\in\{1,\ldots,6\}$. Let $A_n=0$ if n is even and 1 if n is odd. Then find $\langle A\rangle=\frac{1}{2}$, on average half the throws of the die will result in an even number.

It may be that the system's configurations are described by several discrete variables $\{n_1,n_2,n_3,\ldots\}$. We can combine these into a vector $\mathbf{n}$ and then we write $P_\mathbf{n}$ for the discrete distribution, with $\sum_\mathbf{n} P_\mathbf{n}=1$.

Another possibility is that the system's configurations are parameterized by a collection of continuous variables, $\varphi=\{\varphi_1,\ldots,\varphi_n\}$. We write $\varphi\in\Omega$, where Ω is the phase space (or configuration space) of the system. Let $d\mu$ be a *measure* on this space. In general, we can write

$$d\mu = W(\varphi_1,\ldots,\varphi_n)\, d\varphi_1\, d\varphi_2\cdots d\varphi_n\ . \tag{1.4.3}$$

The phase space measure used in classical statistical mechanics gives equal weight W to equal phase space volumes:

$$d\mu=\mathcal{C}\prod_{\sigma=1}^r dq_\sigma\, dp_\sigma\ , \tag{1.4.4}$$

where $\mathcal{C}$ is a constant we shall discuss later on below[8].

Any continuous probability distribution $P(\varphi)$ is normalized according to

$$\int_\Omega d\mu\, P(\varphi) = 1\ . \tag{1.4.5}$$

The average of a function $A(\varphi)$ on configuration space is then

$$\langle A\rangle = \int_\Omega d\mu\, P(\varphi)\, A(\varphi)\ . \tag{1.4.6}$$

For example, consider the Gaussian distribution

$$P(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}\ . \tag{1.4.7}$$

From the result[9]

$$\int_{-\infty}^{\infty} dx\, e^{-\alpha x^2}\, e^{-\beta x} = \sqrt{\frac{\pi}{\alpha}}\, e^{\beta^2/4\alpha}\ , \tag{1.4.8}$$

we see that $P(x)$ is normalized. One can then compute

$$\begin{aligned}\langle x\rangle &= \mu\\ \langle x^2\rangle - \langle x\rangle^2 &= \sigma^2\ .\end{aligned}$$

We call μ the *mean* and σ the *standard deviation* of the distribution, Equation [pgauss].

The quantity $P(\varphi)$ is called the *distribution* or *probability density*. One has

$$P(\varphi)\, d\mu = \text{probability that configuration lies within volume } d\mu \text{ centered at } \varphi \tag{1.4.9}$$

For example, consider the probability density $P=1$ normalized on the interval $x\in\big[0,1\big]$. The probability that some x chosen at random will be *exactly* $\frac{1}{2}$, say, is infinitesimal – one would have to specify each of the infinitely many digits of x. However, we can say that $x\in\big[0.45\ ,\ 0.55\big]$with probability $\frac{1}{10}$.

If x is distributed according to $P_1(x)$, then the probability distribution on the product space $(x_1\ ,\ x_2)$ is simply the product of the distributions: $P_2(x_1,x_2)=P_1(x_1)\,P_1(x_2)$. Suppose we have a function $\phi(x_1,\ldots,x_N)$. How is it distributed? Let $P(\phi)$ be the distribution for ϕ. We then have

$$\begin{aligned}P(\phi) &= \int\limits_{-\infty}^{\infty}\!dx_1\cdots\int\limits_{-\infty}^{\infty}\!dx_N\, P_N(x_1,\ldots,x_N)\,\delta\Big(\phi(x_1,\ldots,x_N)-\phi\Big)\\ &= \int\limits_{-\infty}^{\infty}\!dx_1\cdots\int\limits_{-\infty}^{\infty}\!dx_N\, P_1(x_1)\cdots P_1(x_N)\,\delta\Big(\phi(x_1,\ldots,x_N)-\phi\Big)\ ,\end{aligned}$$

where the second line is appropriate if the $\{x_j\}$ are themselves distributed independently. Note that

$$\int\limits_{-\infty}^{\infty}\!d\phi\ P(\phi)=1\ , \tag{1.4.10}$$

so $P(\phi)$ is itself normalized.

Central limit theorem

In particular, consider the distribution function of the sum $X=\sum_{i=1}^N x_i$. We will be particularly interested in the case where N is large. For general N, though, we have

$$P_N(X)=\int\limits_{-\infty}^{\infty}\!dx_1\cdots\int\limits_{-\infty}^{\infty}\!dx_N\, P_1(x_1)\cdots P_1(x_N)\,\delta\big(x_1+x_2+\ldots+x_N-X\big)\ . \tag{1.4.11}$$

It is convenient to compute the Fourier transform[10] of $P(X)$:

$$\begin{aligned}\hat{P}_N(k) &= \int\limits_{-\infty}^{\infty}\!dX\,P_N(X)\,e^{-ikX}\\ &= \int\limits_{-\infty}^{\infty}\!dX\int\limits_{-\infty}^{\infty}\!dx_1\cdots\int\limits_{-\infty}^{\infty}\!dx_N\,P_1(x_1)\cdots P_1(x_N)\,\delta\big(x_1+\ldots+x_N-X\big)\,e^{-ikX}=\big[\hat{P}_1(k)\big]^N\ ,\end{aligned}$$

where

$$\hat{P}_1(k)=\int\limits_{-\infty}^{\infty}\!dx\,P_1(x)\,e^{-ikx} \tag{1.4.12}$$

is the Fourier transform of the single variable distribution $P_1(x)$. The distribution $P_N(X)$ is a *convolution* of the individual $P_1(x_i)$ distributions. We have therefore proven that *the Fourier transform of a convolution is the product of the Fourier transforms*.

OK, now we can write for $\hat{P}_1(k)$

$$\begin{aligned}\hat{P}_1(k) &= \int_{-\infty}^{\infty} dx\, P_1(x)\, \left(1 - ikx - \frac{1}{2}\, k^2 x^2 + \frac{1}{6}\, i\, k^3\, x^3 + \ldots\right) \\ &= 1 - ik\langle x\rangle - \frac{1}{2}\, k^2 \left\langle x^2\right\rangle + \frac{1}{6}\, i\, k^3 \left\langle x^3\right\rangle + \ldots\, .\end{aligned}$$

Thus,

$$\ln \hat{P}_1(k) = -i\mu k - \frac{1}{2}\sigma^2 k^2 + \frac{1}{6}\, i\, \gamma^3\, k^3 + \ldots\ , \tag{1.4.13}$$

where

$$\begin{aligned}\mu &= \langle x\rangle \\ \sigma^2 &= \left\langle x^2\right\rangle - \langle x\rangle^2 \\ \gamma^3 &= \left\langle x^3\right\rangle - 3\left\langle x^2\right\rangle\langle x\rangle + 2\,\langle x\rangle^3\end{aligned}$$

We can now write

$$\left[\hat{P}_1(k)\right]^N = e^{-iN\mu k}\, e^{-N\sigma^2 k^2/2}\, e^{iN\gamma^3 k^3/6} \cdots \tag{1.4.14}$$

Now for the inverse transform. In computing $P_N(X)$, we will expand the term $e^{iN\gamma^3 k^3/6}$ and all subsequent terms in the above product as a power series in k. We then have

$$\begin{aligned}P_N(X) &= \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ik(X-N\mu)}\, e^{-N\sigma^2 k^2/2} \left\{1 + \frac{1}{6}\, i\, N\gamma^3 k^3 + \ldots\right\} \\ &= \left(1 - \frac{\gamma^3}{6} N\, \frac{\partial^3}{\partial X^3} + \ldots\right) \frac{1}{\sqrt{2\pi N\sigma^2}}\, e^{-(X-N\mu)^2/2N\sigma^2} \\ &= \left(1 - \frac{\gamma^3}{6}\, N^{-1/2}\, \frac{\partial^3}{\partial \xi^3} + \ldots\right) \frac{1}{\sqrt{2\pi N\sigma^2}}\, e^{-\xi^2/2\sigma^2}\ .\end{aligned}$$

In going from the second line to the third, we have written $X = N\mu + \sqrt{N}\,\xi$, in which case $\partial_X = N^{-1/2}\,\partial_\xi$, and the non-Gaussian terms give a subleading contribution which vanishes in the $N \to \infty$ limit. We have just proven the *central limit theorem*: in the limit $N \to \infty$, the distribution of a sum of N independent random variables x_i is a Gaussian with mean $N\mu$ and standard deviation $\sqrt{N}\,\sigma$. Our only assumptions are that the mean μ and standard deviation σ exist for the distribution $P_1(x)$. Note that $P_1(x)$ itself need not be a Gaussian – it could be a very peculiar distribution indeed, but so long as its first and second moment exist, where the k^\ssr{th} moment is simply $\langle x^k\rangle$, the distribution of the sum $X = \sum_{i=1}^N x_i$ is a Gaussian.

Moments and cumulants

Consider a general multivariate distribution $P(x_1, \ldots, x_N)$ and define the multivariate Fourier transform

$$\hat{P}(k_1, \ldots, k_N) = \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N\, P(x_1, \ldots, x_N)\, \exp\left(-i\sum_{j=1}^N k_j x_j\right)\, . \tag{1.4.15}$$

The inverse relation is

$$P(x_1, \ldots, x_N) = \int_{-\infty}^{\infty} \frac{dk_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{dk_N}{2\pi}\, \hat{P}(k_1, \ldots, k_N)\, \exp\left(+i\sum_{j=1}^N k_j x_j\right)\, . \tag{1.4.16}$$

Acting on $\hat{P}(\mathbf{k})$, the differential operator $i\,\frac{\partial}{\partial k_i}$ brings down from the exponential a factor of x_i inside the integral. Thus,

$$\left[\left(i\,\frac{\partial}{\partial k_1}\right)^{m_1} \cdots \left(i\,\frac{\partial}{\partial k_N}\right)^{m_N} \hat{P}(\mathbf{k})\right]_{\mathbf{k}=0} = \left\langle x_1^{m_1} \cdots x_N^{m_N}\right\rangle\, . \tag{1.4.17}$$

Similarly, we can reconstruct the distribution from its moments, *viz*.

$$\hat{P}(\mathbf{k})=\sum_{m_1=0}^{\infty}\cdots\sum_{m_N=0}^{\infty}\frac{(-ik_1)^{m_1}}{m_1!}\cdots\frac{(-ik_N)^{m_N}}{m_N!}\left\langle x_1^{m_1}\cdots x_N^{m_N}\right\rangle . \tag{1.4.18}$$

The *cumulants* $\langle\langle x_1^{m_1}\cdots x_N^{m_N}\rangle\rangle$ are defined by the Taylor expansion of $\ln\hat{P}(\mathbf{k})$:

$$\ln\hat{P}(\mathbf{k})=\sum_{m_1=0}^{\infty}\cdots\sum_{m_N=0}^{\infty}\frac{(-ik_1)^{m_1}}{m_1!}\cdots\frac{(-ik_N)^{m_N}}{m_N!}\left\langle\left\langle x_1^{m_1}\cdots x_N^{m_N}\right\rangle\right\rangle . \tag{1.4.19}$$

There is no general form for the cumulants. It is straightforward to derive the following low order results:

$$\begin{aligned}\langle\langle x_i\rangle\rangle &=\langle x_i\rangle\\ \langle\langle x_ix_j\rangle\rangle &=\langle x_ix_j\rangle-\langle x_i\rangle\langle x_j\rangle\\ \langle\langle x_ix_jx_k\rangle\rangle &=\langle x_ix_jx_k\rangle-\langle x_ix_j\rangle\langle x_k\rangle-\langle x_jx_k\rangle\langle x_i\rangle-\langle x_kx_i\rangle\langle x_j\rangle+2\langle x_i\rangle\langle x_j\rangle\langle x_k\rangle .\end{aligned}$$

Multidimensional Gaussian integral

Consider the multivariable Gaussian distribution,

$$P(\mathbf{x})\equiv\left(\frac{\det A}{(2\pi)^n}\right)^{1/2}\exp\left(-\frac{1}{2}\,x_i\,A_{ij}\,x_j\right) , \tag{1.4.20}$$

where A is a positive definite matrix of rank n. A mathematical result which is extremely important throughout physics is the following:

$$Z(\mathbf{b})=\left(\frac{\det A}{(2\pi)^n}\right)^{1/2}\int_{-\infty}^{\infty}\!dx_1\cdots\int_{-\infty}^{\infty}\!dx_n\,\exp\left(-\frac{1}{2}\,x_i\,A_{ij}\,x_j+b_i\,x_i\right)=\exp\left(\frac{1}{2}\,b_i\,A_{ij}^{-1}\,b_j\right) . \tag{1.4.21}$$

Here, the vector $\mathbf{b}=(b_1\,,\,\ldots\,,\,b_n)$ is identified as a *source*. Since $Z(0)=1$, we have that the distribution $P(\mathbf{x})$ is normalized. Now consider averages of the form

$$\begin{aligned}\langle\,x_{j_1}\cdots\,x_{j_{2k}}\,\rangle &=\int\! d^nx\;P(\mathbf{x})\;x_{j_1}\cdots\,x_{j_{2k}}=\frac{\partial^nZ(\mathbf{b})}{\partial b_{j_1}\cdots\,\partial b_{j_{2k}}}\bigg|_{\mathbf{b}=0}\\ &=\sum_{contractions}A^{-1}_{j_{\sigma(1)}j_{\sigma(2)}}\cdots A^{-1}_{j_{\sigma(2k-1)}j_{\sigma(2k)}} .\end{aligned}$$

The sum in the last term is over all *contractions* of the indices $\{j_1\,,\,\ldots\,,\,j_{2k}\}$. A contraction is an arrangement of the $2k$ indices into k pairs. There are $C_{2k}=(2k)!/2^kk!$ possible such contractions. To obtain this result for C_k, we start with the first index and then find a mate among the remaining $2k-1$ indices. Then we choose the next unpaired index and find a mate among the remaining $2k-3$ indices. Proceeding in this manner, we have

$$C_{2k}=(2k-1)\cdot(2k-3)\cdots3\cdot1=\frac{(2k)!}{2^kk!} . \tag{1.4.22}$$

Equivalently, we can take all possible permutations of the $2k$ indices, and then divide by $2^kk!$ since permutation within a given pair results in the same contraction and permutation among the k pairs results in the same contraction. For example, for $k=2$, we have $C_4=3$, and

$$\langle\,x_{j_1}x_{j_2}x_{j_3}x_{j_4}\,\rangle=A^{-1}_{j_1j_2}A^{-1}_{j_3j_4}+A^{-1}_{j_1j_3}A^{-1}_{j_2j_4}+A^{-1}_{j_1j_4}A^{-1}_{j_2j_3} . \tag{1.4.23}$$

If we define $b_i=ik_i$, we have

$$\hat{P}(\mathbf{k})=\exp\left(-\frac{1}{2}\,k_i\,A_{ij}^{-1}\,k_j\right) , \tag{1.4.24}$$

from which we read off the cumulants $\langle\langle x_ix_j\rangle\rangle=A_{ij}^{-1}$, with all higher order cumulants vanishing.