A few useful mathematical tools

These notes provide a brief, informal, and non-rigorous summary of some mathematical concepts and tools that will be useful in this course. The main topics covered are: the moments and cumulants of a probability distribution, the law of large numbers, the central limit theorem, large deviation theory, and Poisson processes.

Throughout these notes, $f(x)$ denotes a probability distribution of a real variable x. We assume $f(x)$ is normalized to unity, $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$, and all of its cumulants (defined below) have well-defined, non-infinite values. We refer to x as a *random variable*, which is sampled from the distribution $f(x)$. Angular brackets $\langle \cdots \rangle$ denote averages with respect to values of x sampled from this distribution.

The *moments* of the distribution are averages of integer powers of x:

$$\mu_n = \langle x^n \rangle = \int \mathrm{d}x\, f(x) x^n \quad , \quad n = 0, 1, 2, \cdots \tag{1}$$

μ (without a subscript) and σ^2 will denote the *mean* and *variance* of the distribution:

$$\mu = \mu_1 \qquad , \qquad \sigma^2 = \mu_2 - \mu_1^2 \tag{2}$$

The *cumulant-generating function*, or simply *generating function*, is given by

$$g(\lambda) = \ln \left\langle e^{\lambda x} \right\rangle = \ln \int \mathrm{d}x\, f(x) e^{\lambda x} \tag{3}$$

where λ is a real number. The *cumulants* κ_m are defined through the expansion

$$g(\lambda) = \sum_{m=1}^{\infty} \kappa_m \frac{\lambda^m}{m!} = \kappa_1 \lambda + \frac{1}{2} \kappa_2 \lambda^2 + \cdots \tag{4}$$

(Note that $g(0) = 0$.) Equivalently,

$$\kappa_m = \left(\frac{\mathrm{d}^m g}{\mathrm{d}\lambda^m} \right)_{\lambda=0} \quad , \quad m \geq 1 \tag{5}$$

Computing the first two derivatives of $g(\lambda)$ (Eq. 3),

$$g'(\lambda) \;=\; \frac{\mathrm{d}g}{\mathrm{d}\lambda} = \frac{\int f(x)\, x\, e^{\lambda x}}{\int f(x)\, e^{\lambda x}} \tag{6}$$

$$g''(\lambda) \;=\; \frac{\mathrm{d}^2 g}{\mathrm{d}\lambda^2} = \frac{\int f(x)\, x^2\, e^{\lambda x}}{\int f(x)\, e^{\lambda x}} - \left[\frac{\int f(x)\, x\, e^{\lambda x}}{\int f(x)\, e^{\lambda x}} \right]^2 \tag{7}$$

and evaluating them at $\lambda = 0$ gives us

$$\kappa_1 = g'(0) = \mu_1 \qquad , \qquad \kappa_2 = g''(0) = \mu_2 - \mu_1^2 \tag{8}$$

Thus the first and second cumulants are the mean, $\kappa_1 = \mu$, and variance, $\kappa_2 = \sigma^2$. Higher cumulants such as the *skewness*, κ_3 and the *kurtosis*, κ_4, can similarly be evaluated.

Scaling property: if $y = ax$, where $a > 0$ is a constant, then

$$\mu_m[y] = a^m \, \mu_m[x] \quad , \quad \kappa_m[y] = a^m \, \kappa_m[x] \tag{9}$$

The notation $\mu_m[y]$ denotes the m'th moment of the random variable y, etc.

The cumulants (but not the moments) satisfy the useful property

$$\kappa_m[X] = N \, \kappa_m[x] \tag{10}$$

where $X = \sum_{n=0}^{N} x_n$ is the sum of N independent samples from the distribution $f(x)$. This can be seen by comparing the corresponding generating functions:

$$G(\lambda) = \ln\langle e^{\lambda X}\rangle = \ln\langle e^{\lambda x_1} e^{\lambda x_2} \cdots e^{\lambda x_N}\rangle = N \ln\langle e^{\lambda x}\rangle = N g(\lambda) \tag{11}$$

Eq. 10 then follows from Eq. 4.

Now consider the *sample mean*, $\bar{x} = N^{-1} \sum_{n=1}^{N} x_n = X/N$, which is the average over N independent samples from $f(x)$. Let $f_N(\bar{x})$ denote the probability distribution of the sample mean. Combining Eqs. 9 and 10 we get

$$\kappa_m[\bar{x}] = \frac{1}{N^m} \kappa_m[X] = \frac{1}{N^{m-1}} \kappa_m[x] \tag{12}$$

For the mean and variance, this gives us the familiar (hopefully!) results

$$\mu_{\bar{x}} = \mu_x \equiv \mu \qquad , \qquad \sigma_{\bar{x}}^2 = \frac{1}{N} \sigma_x^2 \tag{13}$$

Thus as $N \to \infty$, $f_N(\bar{x})$ becomes infinitely narrow:

$$\lim_{N \to \infty} f_N(\bar{x}) = \delta(\bar{x} - \mu) \tag{14}$$

This is the *law of large numbers* (LLN), which states that the sample mean $\bar{x}$ converges to the ensemble mean μ in the limit of infinitely many samples.

Central limit theorem (CLT)

The CLT states, roughly, that $f_N(\bar{x})$ looks like a Gaussian for sufficiently large N. For a Gaussian distribution, $\kappa_m = 0$ for all $m \geq 3$. (Proof assigned as exercise!) Moreover, any distribution that is *not* a Gaussian has infinitely many non-zero cumulants (Marcinkiewicz, 1939). To establish the CLT, consider a distribution $f(x)$ with mean $\mu = 0$ and variance $\sigma^2 > 0$. [1] Now define a "rescaled" sample mean,

$$y = \sqrt{\frac{N}{\sigma^2}}\,\bar{x} \tag{15}$$

where $\bar{x} = N^{-1}\sum_{n=1}^{N} x_n$ as before. Using Eqs. 9 and 10 we get

$$\kappa_m[y] = \frac{1}{N}\left(\frac{N}{\sigma^2}\right)^{m/2} \kappa_m[x] \tag{16}$$

hence as $N \to \infty$ we get $\kappa_2[y] = 1$, and $\kappa_m[y] \to 0$ for all $m \geq 3$. Thus the distribution of the variable y converges toward a Gaussian whose variance is unity. From Eq. 15 we conclude that the distribution of the sample mean $\bar{x}$ tends toward a Gaussian of variance σ^2/N.

Dropping the assumption $\mu = 0$, we summarize LLN and CLT as follows. As $N \to \infty$, $f_N(\bar{x})$ becomes ever sharper and ever more Gaussian:

$$f_N(\bar{x}) \approx \sqrt{\frac{N}{2\pi\sigma^2}}\exp\left[-\frac{N(\bar{x}-\mu)^2}{2\sigma^2}\right] \tag{17}$$

Large deviation theory (LDT)

Both LLN and CLT follow naturally from large deviation theory. For our purposes, LDT is a set of tools built around the following statement. For a well-behaved distribution $f(x)$, the distribution of the sample mean $\bar{x}$ obeys

$$f_N(\bar{x}) \sim e^{-NI(\bar{x})} \qquad \text{as } N \to \infty \tag{18}$$

The notation $\sim$ indicates that Eq. 18 captures the dominant dependence on N. That is, if $f_N(\bar{x}) = \exp -N[I(\bar{x}) + \text{other terms}]$, then the "other terms" are *subdominant*, i.e. they vanish as $N \to \infty$.[2] Somewhat more precisely, LDT states that the limit

$$\lim_{N\to\infty} -\frac{1}{N}\ln f_N(\bar{x}) \equiv I(\bar{x}) \tag{19}$$

[1] There is no loss of generality in assuming $\mu = 0$. If $\mu \neq 0$ then we can define a new random variable $x' = x - \mu$ whose mean is zero, and which is trivially related to x by a simple shift.

[2] Subdominant terms might arise from an N-dependent normalization factor, such as the one in Eq. 17.

exists. The definition of "well-behaved" is a bit logically circular: if the above limit exists, then we say that $f(x)$ "satisfies the large deviation principle". In practice, most of the distributions we will encounter in this course will satisfy this principle.

The *large deviation function* (or *Cramér function*) $I(\bar{x})$ has these properties:

1. It is *convex*, that is $I''(\bar{x}) \geq 0$ for all $\bar{x}$.

2. The unique minimum of $I(\bar{x})$ occurs at $\bar{x} = \mu$, where $I(\mu) = 0$.

3. $I''(\mu) = 1/\sigma^2$.

As $N \to \infty$, the distribution $f_N(\bar{x}) \sim e^{-NI(\bar{x})}$ becomes negligible except in the immediate vicinity of the minimum at $\bar{x} = \mu$. Expanding around this minimum, we have

$$I(\bar{x}) \approx \frac{(\bar{x} - \mu)^2}{2\sigma^2} \tag{20}$$

using properties 2 and 3. Combining with Eq. 18 we see that at large N, $f_N(\bar{x})$ becomes approximately a Gaussian of mean μ and variance σ^2/N, in agreement with Eq. 17.

Note that Eq. 20 is valid only near $\bar{x} = \mu$, in other words $f_N(\bar{x})$ is Gaussian only in this "central" region. The tails of the distribution are described by the shape of $I(\bar{x})$ away from this minimum. In other words, the central limit theorem does not describe these tails, but only the central region near $\bar{x} = \mu$, where most of the probability resides.

The function $I(\bar{x})$ can be computed either directly from Eq. 19, or else via the Legendre transform of the generating function:

$$I(\bar{x}) = \max_\lambda \{\lambda\bar{x} - g(\lambda)\} = \lambda^*\bar{x} - g(\lambda^*) = \text{Legendre transform of } g(\lambda) \tag{21}$$

where $\lambda^* = \lambda^*(\bar{x})$ is defined by the condition $g'(\lambda^*) = \bar{x}$. I won't derive Eq. 21 here. For a pedagogical introduction to Legendre transforms, see Zia, Redish and McKay, "Making sense of the Legendre transform", *Am. J. Phys.* **77**, 614 (2009).

Recall that $g(0) = 0$ and $g'(0) = \mu$, and note that $g''(\lambda) \geq 0$ for all λ (this follows from Eq. 7). <u>Exercise</u>: using these properties of the generating function $g(\lambda)$, show that its Legendre transform $I(\bar{x})$ satisfies properties 1-3 listed above.

Large deviation theory plays an important role in both equilibrium and nonequilibrium statistical physics, as well as in statistics, finance and other fields. A review of LDT can be found in Touchette, "The large deviation approach to statistical mechanics", *Phys. Reports* **478**, 1 (2009).

Poisson processes

In a Poisson process, some event occurs repeatedly, with a fixed *probability rate*, r. That is, during every infinitesimal time interval δt, the probability that an event occurs is $r \cdot \delta t$. An example is the emission of α-particles from a radioactive sample with a long half-life. Each emission of an α-particle is an event. If our period of observation is much shorter than the half-life, we can model these events as a Poisson process.

For a Poisson process with rate r, let's compute $p(n;\tau)$, the probability to observe exactly n events during a given time interval of duration τ. We divide the interval into $K \gg 1$ sub-intervals of duration $\delta t = \tau/K$. By considering all the ways that n events can be distributed among K bins, we have

$$\begin{aligned} p(n;\tau) &= \lim_{K\to\infty} \frac{K!}{n!(K-n)!} (r\cdot\delta t)^n (1 - r\cdot\delta t)^{K-n} && (22)\\ &= \frac{1}{n!} \lim_{K\to\infty} \left[\frac{K(K-1)\cdots(K-n+1)}{K^n}\right] \alpha^n \left(1-\frac{\alpha}{K}\right)^{K-n} && (23)\\ &= \frac{\alpha^n}{n!} \lim_{K\to\infty} \left[\left(1-\frac{\alpha}{K}\right)^K\right]^{1-(n/K)} && (24)\end{aligned}$$

where $\alpha \equiv r\tau$. If we now use $\lim_{N\to\infty}[1-(x/N)]^N = e^{-x}$,[3] we arrive at the result

$$p(n;\tau) = \frac{\alpha^n}{n!} e^{-\alpha} \qquad , \qquad \alpha = r\tau \tag{25}$$

It is easy to verify that normalization is satisfied: $\sum_{n=0}^{\infty} p(n;\tau) = 1$.

To compute the generating function $g(\lambda)$, we first evaluate

$$\begin{aligned} \langle e^{\lambda n}\rangle &= \sum_{n=0}^{\infty} e^{\lambda n} \frac{\alpha^n}{n!} e^{-\alpha} && (26)\\ &= \sum_{n=0}^{\infty} \frac{\left(\alpha e^{\lambda}\right)^n}{n!} \cdot e^{-\alpha e^{\lambda}} \cdot e^{\alpha\left(e^{\lambda}-1\right)} && (27)\\ &= e^{\alpha\left(e^{\lambda}-1\right)} \sum_{n=0}^{\infty} \frac{\beta^n}{n!} e^{-\beta} \qquad , \qquad \beta = \alpha e^{\lambda} && (28)\end{aligned}$$

Since the sum on the last line converges to unity, we arrive at

$$g(\lambda) = \ln\left\langle e^{\lambda n}\right\rangle = \alpha\left(e^{\lambda}-1\right) \tag{29}$$

Eq. 5 then gives us

$$\kappa_m[n] = \alpha \qquad , \qquad m = 1, 2, \cdots \tag{30}$$

[3] We can motivate this by noting that $\ln[1-(x/N)]^N = N\ln[1-(x/N)] \to -x$ as $N \to \infty$.

Thus all of the cumulants have the same value, $\alpha = r\tau$. In particular the mean and variance of the number of events occurring in the interval τ are

$$\mu = \langle n \rangle = r\tau \qquad , \qquad \sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = r\tau \tag{31}$$

The result $\langle n \rangle = r\tau$ makes sense; we could have guessed this without doing any calculations. The quantity $\sigma = \sqrt{r\tau}$ characterizes the width of the distribution $p(n;\tau)$, thus it provides a measure of the size of typical fluctuations in n. For very long time intervals, $r\tau \gg 1$, both $\langle n \rangle$ and σ become large, but the *relative* size of the fluctuations $\sigma/\langle n \rangle = 1/\sqrt{r\tau}$, goes to zero, in agreement with the law of large numbers.

Here's how to simulate a Poisson process using a standard random number generator:

1. Set $t_0 = 0$ and $k = 0$.

2. Generate a random number ξ between 0 and 1.

3. Set
$$t_{k+1} = t_k - \frac{\ln \xi}{r} \tag{32}$$

4. Update $k \to k+1$ and go back to step 2.

The sequence $t_1, t_2, t_3, \cdots$ will represent the times at which the events occur.

Exercise: Convince yourself that this algorithm generates a Poisson process.