Week 3 Probability and Statistics

- Statistical mechanics is an inherently probabilistic description of the macroscopic system

- **General Definitions**
  - Random variables
  - Probability density functiona

- **Important Probability Distributions**

- **Many Random Variables**

- **Central Limit Theorem**

- **Rules for Large Numbers**

- **References**
  - M. Kardar, Ch.2.1-2.3

General Definitions

The entity under investigation is a *random variable* x, which has a set of possible *outcomes* $\mathcal{S} \equiv \{x_1, x_2, \cdots\}$. The outcomes may be *discrete* as in the case of a coin toss, $S_{\text{coin}} = \{\text{head}, \text{tail}\}$, or a dice throw, $\mathcal{S}_{\text{dice}} = \{1, 2, 3, 4, 5, 6\}$, or *continuous* as for the velocity of a particle in a gas, $\mathcal{S}_{\vec{v}} = \left\{-\infty < v_x, v_y, v_z < \infty\right\}$, or the energy of an electron in a metal at zero temperature, $\mathcal{S}_\epsilon = \{0 \leq \epsilon \leq \epsilon_F\}$. An *event* is any subset of outcomes $E \subset \mathcal{S}$, and is assigned a *probability* $p(E)$, for example, $p_{\text{dice}}(\{1\}) = 1/6$, or $p_{\text{dice}}(\{1, 3\}) = 1/3$. From an axiomatic point of view, the probabilities must satisfy the following conditions:

(i) *Positivity*. $p(E) \geq 0$, that is, all probabilities must be real and non-negative.

(ii) *Additivity*. $p(A \text{ or } B) = p(A) + p(B)$, if A and B are disconnected events.

(iii) *Normalization*. $p(\mathcal{S}) = 1$, that is, the random variable must have some outcome in $\mathcal{S}$.

Objective and Subjective Probabilities

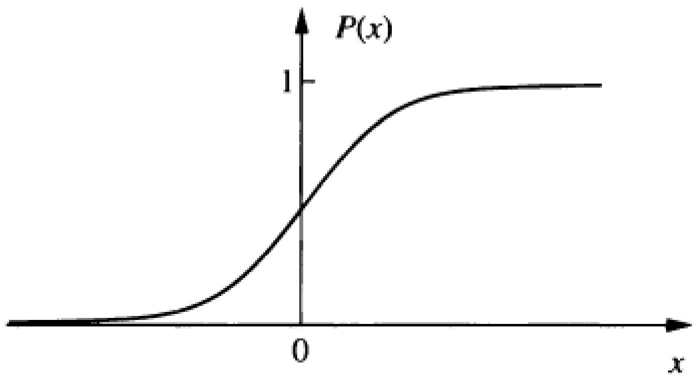
1. *Objective* probabilities are obtained *experimentally* from the relative frequency of the occurrence of an outcome in many tests of the random variable. If the random process is repeated N times, and the event A occurs N_A times, then

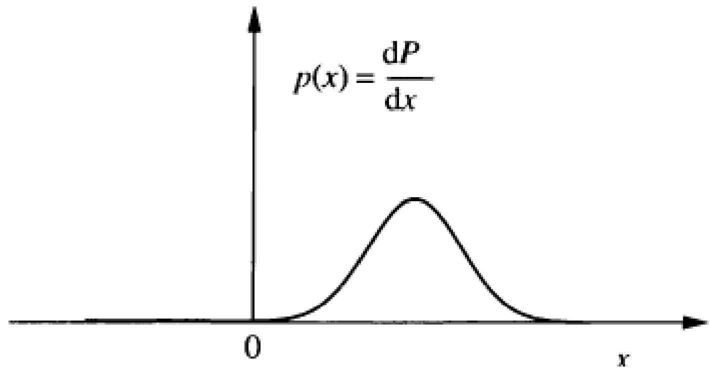
$$p(A) = \lim_{N \to \infty} \frac{N_A}{N}.$$

2. *Subjective* probabilities provide a *theoretical* estimate based on the uncertainties related to lack of precise knowledge of outcomes. For example, the assessment $p_{\text{dice}}(\{1\}) = 1/6$ is based on the knowledge that there are six possible outcomes to a dice throw, and that in the absence of any prior reason to believe that the dice is biased, all six are equally likely. All assignments of probability in statistical mechanics are subjectively based. The consequences of such subjective assignments of probability have to be checked against measurements, and they may need to be modified as more information about the outcomes becomes available.

One Random Variable

- *The cumulative probability function* (CPF) $P(x)$ is the probability of an outcome with *any value* less than x, that is, $P(x) = \text{prob}(E \subset [-\infty, x])$. $P(x)$ must be a monotonically increasing function of x, with $P(-\infty) = 0$ and $P(+\infty) = 1$.





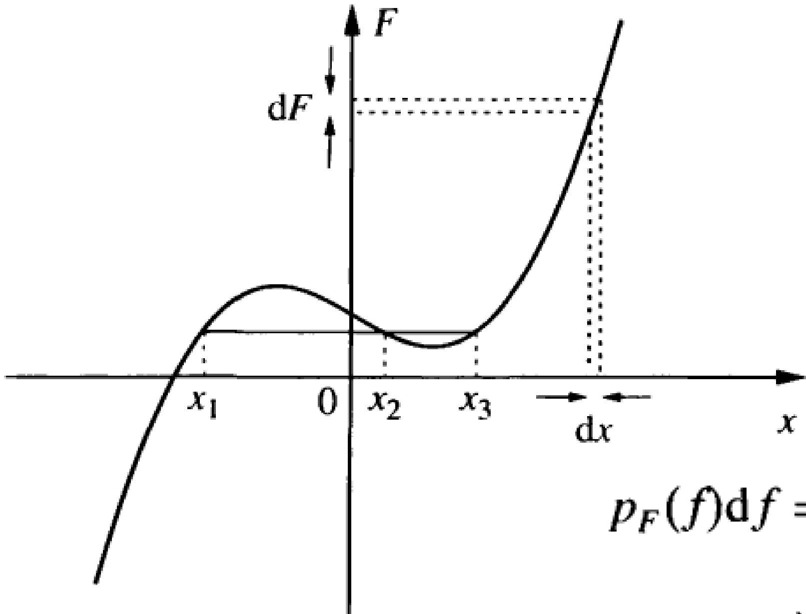
- *The probability density function* (PDF) is defined by $p(x) \equiv \mathrm{d}P(x)/\mathrm{d}x$. Hence, $p(x)\mathrm{d}x = \text{prob}(E \in [x, x+\mathrm{d}x])$. As a probability density, it is *positive*, and normalized such that

$$\text{prob}(\mathcal{S}) = \int_{-\infty}^{\infty} \mathrm{d}x \; p(x) = 1.$$

- *The expectation value* of any function, $F(x)$, of the random variable is

$$\langle F(x) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x \; p(x) F(x).$$

Example: PDF of the Function F(x)



$$p_F(f)\mathrm{d}f = \mathrm{prob}(F(x) \in [f, f+\mathrm{d}f]).$$

$|\mathrm{d}x/\mathrm{d}F|$ are the *Jacobians*

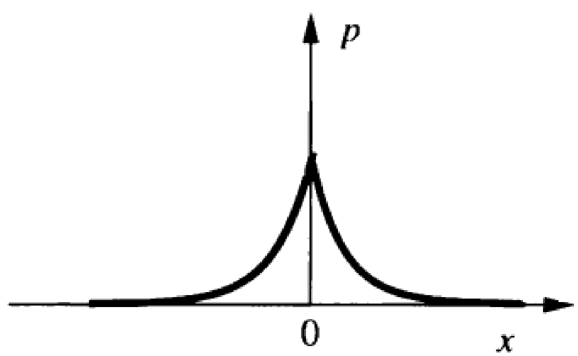
$$p_F(f)\mathrm{d}f = \sum_i p(x_i)\mathrm{d}x_i\,, \quad \Longrightarrow \quad p_F(f) = \sum_i p(x_i)\left|\frac{\mathrm{d}x}{\mathrm{d}F}\right|_{x=x_i}$$

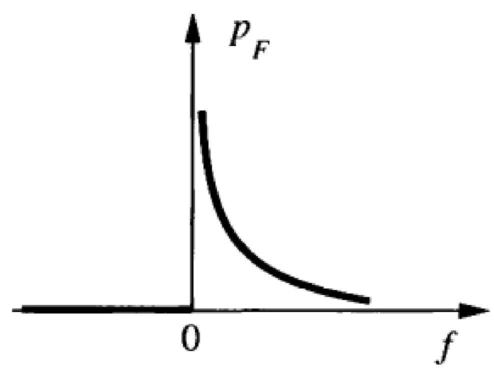
$$p(x) = \lambda\exp(-\lambda|x|)/2,$$

$$F(x) = x^2.$$

$$P_F(f) = \frac{\lambda}{2}\exp\left(-\lambda\sqrt{f}\right)\left(\left|\frac{1}{2\sqrt{f}}\right| + \left|\frac{-1}{2\sqrt{f}}\right|\right) = \frac{\lambda\exp\left(-\lambda\sqrt{f}\right)}{2\sqrt{f}},$$

for $f > 0$, and $p_F(f) = 0$ for $f < 0$.





- *Moments* of the PDF are expectation values for powers of the random variable. The nth moment is
$$m_n \equiv \langle x^n \rangle = \int \mathrm{d}x p(x) x^n .$$
- *The characteristic function* is the generator of moments of the distribution. It is simply the Fourier transform of the PDF, defined by
$$p(x) = \frac{1}{2\pi} \int \mathrm{d}k \tilde{p}(k)\, \mathrm{e}^{+\mathrm{i}kx}.$$
$$p(x) = \frac{1}{2\pi} \int \mathrm{d}k \tilde{p}(k)\, \mathrm{e}^{+\mathrm{i}kx}.$$

Moments of the distribution are obtained by expanding $\tilde{p}(k)$ in powers of k,
$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(-\mathrm{i}k)^n}{n!} x^n \right\rangle = \sum_{n=0}^{\infty} \frac{(-\mathrm{i}k)^n}{n!} \langle x^n \rangle .$$

Moments of the PDF around any point x_0 can also be generated by expanding
$$\mathrm{e}^{\mathrm{i}kx_0} \tilde{p}(k) = \left\langle \mathrm{e}^{-\mathrm{i}k(x-x_0)} \right\rangle = \sum_{n=0}^{\infty} \frac{(-\mathrm{i}k)^n}{n!} \langle (x - x_0)^n \rangle .$$

Cumulant Generating Functions

- *The cumulant generating function* is the logarithm of the characteristic function. Its expansion generates the *cumulants* of the distribution defined through

$$\ln \tilde{p}(k) = \sum_{n=1}^{\infty} \frac{(-\mathrm{i}k)^n}{n!} \langle x^n \rangle_c \, .$$

$$\ln(1+\epsilon) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\epsilon^n}{n}.$$

The first four cumulants are called the *mean*, *variance*, *skewness*, and *curtosis*

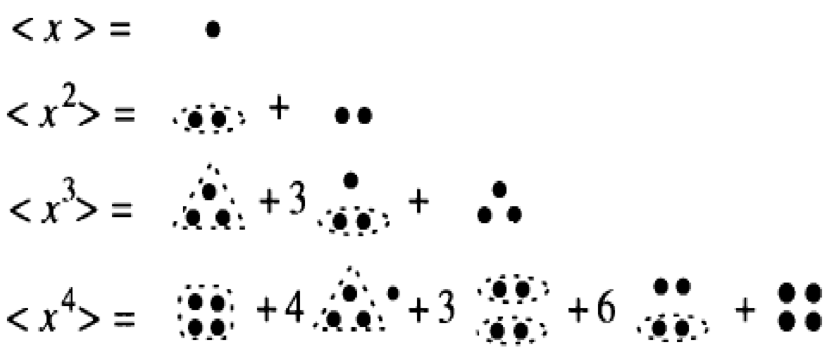
$$\langle x \rangle_c = \langle x \rangle \, ,$$

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 \, ,$$

$$\langle x^3 \rangle_c = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \, ,$$

$$\langle x^4 \rangle_c = \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4 \, .$$

Cumulant and Moments



$$\langle x \rangle = \langle x \rangle_c \,,$$

$$\langle x^2 \rangle = \langle x^2 \rangle_c + \langle x \rangle_c^2 \,,$$

$$\langle x^3 \rangle = \langle x^3 \rangle_c + 3\langle x^2 \rangle_c \langle x \rangle_c + \langle x \rangle_c^3 \,,$$

$$\langle x^4 \rangle = \langle x^4 \rangle_c + 4\langle x^3 \rangle_c \langle x \rangle_c + 3\langle x^2 \rangle_c^2 + 6\langle x^2 \rangle_c \langle x \rangle_c^2 + \langle x \rangle_c^4 \,.$$

$$\sum_{m=0}^{\infty} \frac{(-\mathrm{i}k)^m}{m!} \langle x^m \rangle = \exp\left[\sum_{n=1}^{\infty} \frac{(-\mathrm{i}k)^n}{n!} \langle x^n \rangle_c \right] = \prod_n \sum_{p_n} \left[\frac{(-\mathrm{i}k)^{np_n}}{p_n!} \left(\frac{\langle x^n \rangle_c}{n!} \right)^{p_n} \right]$$

$$\langle x^m \rangle = \sum_{\{p_n\}}{}' \, m! \prod_n \frac{1}{p_n!(n!)^{p_n}} \langle x^n \rangle_c^{p_n} \,.$$

$$\sum n p_n = m$$

Important PDFs: Normal Distribution

(1) *The normal (Gaussian) distribution* describes a continuous real random variable x, with

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\lambda)^2}{2\sigma^2}\right].$$

$$\tilde{p}(k) = \int_{-\infty}^{\infty} \mathrm{d}x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\lambda)^2}{2\sigma^2} - \mathrm{i}kx\right] = \exp\left[-\mathrm{i}k\lambda - \frac{k^2\sigma^2}{2}\right].$$

$$\ln \tilde{p}(k) = -\mathrm{i}k\lambda - k^2\sigma^2/2,$$

$$\langle x \rangle_c = \lambda, \quad \langle x^2 \rangle_c = \sigma^2, \quad \langle x^3 \rangle_c = \langle x^4 \rangle_c = \cdots = 0.$$

$$\langle x \rangle = \lambda,$$

$$\langle x^2 \rangle = \sigma^2 + \lambda^2,$$

$$\langle x^3 \rangle = 3\sigma^2\lambda + \lambda^3,$$

$$\langle x^4 \rangle = 3\sigma^4 + 6\sigma^2\lambda^2 + \lambda^4,$$

Important PDFs: Binomial Distribution

(2) *The binomial distribution*: consider a random variable with two outcomes A and B (e.g., a coin toss) of relative probabilities p_A and $p_B = 1 - p_A$. The probability that in N trials the event A occurs exactly N_A times (e.g., 5 heads in 12 coin tosses) is given by the binomial distribution

$$p_N(N_A) = \binom{N}{N_A} p_A^{N_A} p_B^{N-N_A}. \qquad \binom{N}{N_A} = \frac{N!}{N_A!(N-N_A)!},$$

$$\tilde{p}_N(k) = \left\langle e^{-ikN_A} \right\rangle = \sum_{N_A=0}^{N} \frac{N!}{N_A!(N-N_A)!} p_A^{N_A} p_B^{N-N_A} e^{-ikN_A} = \left(p_A e^{-ik} + p_B\right)^N .$$

$$\ln \tilde{p}_N(k) = N \ln \left(p_A e^{-ik} + p_B\right) = N \ln \tilde{p}_1(k),$$

$$\langle N_A \rangle_c = N p_A, \quad \left\langle N_A^2 \right\rangle_c = N\left(p_A - p_A^2\right) = N p_A p_B.$$

The binomial distribution is straightforwardly generalized to a *multinomial* distribution, when the several outcomes $\{A, B, \cdots, M\}$ occur with probabilities $\{p_A, p_B, \cdots, p_M\}$. The probability of finding outcomes $\{N_A, N_B, \cdots, N_M\}$ in a total of $N = N_A + N_B + \cdots + N_M$ trials is

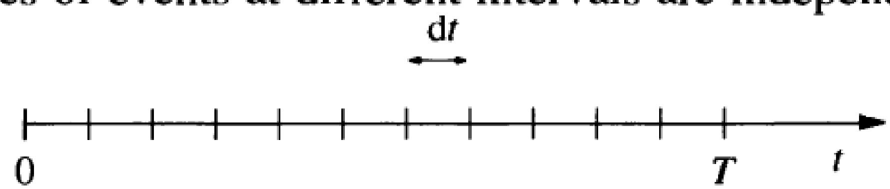
$$p_N(\{N_A, N_B, \cdots, N_M\}) = \frac{N!}{N_A! N_B! \cdots N_M!} p_A^{N_A} p_B^{N_B} \cdots p_M^{N_M} .$$

Important PDFs: Poisson Distribution

(3) *The Poisson distribution*: the classical example of a Poisson process is radioactive decay. Observing a piece of radioactive material over a time interval T shows that:

(a) The probability of one and only one event (decay) in the interval $[t, t+\mathrm{d}t]$ is proportional to $\mathrm{d}t$ as $\mathrm{d}t \to 0$.

(b) The probabilities of events at different intervals are independent of each other.



The probability of observing exactly M decays in the interval T is given by the Poisson distribution. It is obtained as a limit of the binomial distribution by subdividing the interval into $N = T/\mathrm{d}t \gg 1$

$$\tilde{p}(k) = \left(p\mathrm{e}^{-\mathrm{i}k} + q\right)^n = \lim_{\mathrm{d}t\to 0}\left[1 + \alpha\mathrm{d}t\left(\mathrm{e}^{-\mathrm{i}k} - 1\right)\right]^{T/\mathrm{d}t} = \exp\left[\alpha(\mathrm{e}^{-\mathrm{i}k} - 1)T\right].$$

$$p(x) = \int_{-\infty}^{\infty}\frac{\mathrm{d}k}{2\pi}\exp\left[\alpha\left(\mathrm{e}^{-\mathrm{i}k} - 1\right)T + \mathrm{i}kx\right] = \mathrm{e}^{-\alpha T}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{2\pi}\mathrm{e}^{\mathrm{i}kx}\sum_{M=0}^{\infty}\frac{(\alpha T)^M}{M!}\mathrm{e}^{-\mathrm{i}kM},$$

$$p_{\alpha T}(x) = \sum_{M=0}^{\infty}\mathrm{e}^{-\alpha T}\frac{(\alpha T)^M}{M!}\delta(x - M).$$

$$\ln\tilde{p}_{\alpha T}(k) = \alpha T(\mathrm{e}^{-\mathrm{i}k} - 1) = \alpha T\sum_{n=1}^{\infty}\frac{(-\mathrm{i}k)^n}{n!}, \quad \Longrightarrow \quad \langle M^n\rangle_c = \alpha T.$$

$$\langle M\rangle = (\alpha T), \quad \langle M^2\rangle = (\alpha T)^2 + (\alpha T), \quad \langle M^3\rangle = (\alpha T)^3 + 3(\alpha T)^2 + (\alpha T).$$

Example

Example. Assuming that stars are randomly distributed in the Galaxy (clearly unjustified) with a density n, what is the probability that the nearest star is at a distance R? Since the probability of finding a star in a small volume $\mathrm{d}V$ is $n\mathrm{d}V$, and they are assumed to be independent, the number of stars in a volume V is described by a Poisson process as in Eq. (2.28), with $\alpha = n$. The probability $p(R)$ of encountering the first star at a distance R is the product of the probabilities $p_{nV}(0)$ of finding zero stars in the volume $V = 4\pi R^3/3$ around the origin, and $p_{n\mathrm{d}V}(1)$ of finding one star in the shell of volume $\mathrm{d}V = 4\pi R^2 \mathrm{d}R$ at a distance R. Both $p_{nV}(0)$ and $p_{n\mathrm{d}V}(1)$ can be calculated from Eq. (2.28),

$$p_{\alpha T}(x) = \sum_{M=0}^{\infty} \mathrm{e}^{-\alpha T} \frac{(\alpha T)^M}{M!} \delta(x - M). \qquad (2.28)$$

$$p(R)\mathrm{d}R = p_{nV}(0)\, p_{n\mathrm{d}V}(1) = \mathrm{e}^{-4\pi R^3 n/3}\, \mathrm{e}^{-4\pi R^2 n \mathrm{d}R}\, 4\pi R^2 n \mathrm{d}R,$$

$$\Longrightarrow \qquad p(R) = 4\pi R^2 n \exp\left(-\frac{4\pi}{3} R^3 n\right).$$