Review of Probability and Statistics

Probability Distribution

• *Mathematically:* A probability is a normalized measure on a sample space X of possible events or outcomes; that is, for some collection $\mathcal{A}$ of subsets of X among which we can take set differences and countable unions, a function $P : \mathcal{A} \to \mathbb{R}$ taking $\mathcal{A} \ni A \mapsto P(A) \geq 0$, which is countably additive and satisfies $P(X) = 1$.

• *In physics:* A probability distribution assigns a likelihood to each statement about a system S, identified with the state of the system belonging to some subset of the set of all states. For example, for a classical theory the sample space X or set of possible states is the classical phase space Γ of S. We are often interested in finding the mean value $\langle f\rangle$ and uncertainty $\sigma(f)$ or Δf of functions $f(x)$ of random variables $x \in X$.

• *Discrete sample space:* The set X is finite or countable. Examples are the possible outcomes for a coin toss or a die ($P(i) = \frac{1}{6}$), the states $X = \mathbf{Z}_2^N$ for N spin-$\frac{1}{2}$ particles, or the locations on an infinite lattice.

• *Continuous sample space:* The set X is parametrized as a subset of $\mathbb{R}^n$, for some n. For example, the phase space $\Gamma = \{(\vec{x}_{(i)}, \vec{p}_{(i)})\} \subseteq \mathbb{R}^{6N}$ for N pointlike particles in $\mathbb{R}^3$. Here, what we often work with are not the probabilities themselves, but probability densities; for example, if $P(x)$ is a probability density on $\mathbb{R}$, for any $A \subseteq \mathbb{R}$ the probability is $P(A) := \int_A P(x)\, \mathrm{d}x$.

• *Interpretation:* The stated interpretation of probability in physics is usually frequentist ($P(i)$ is the fraction of times outcome i occurs as the number of trials becomes very large), but in practice the working meaning of probability physicists use is Bayesian (an expectation or "degree of belief" in each possible event).

Characterizing a Probability Distribution

• *Mean:* The mean value of a function $f(x)$ of a discrete or continuous random variable x is, respectively,

$$\langle f\rangle = \textstyle\sum_i f(x_i)\, P(x_i) \qquad \text{or} \qquad \langle f\rangle = \displaystyle\int_X \mathrm{d}x\, f(x)\, P(x)\ .$$

• *Variance:* The mean-square deviation from the mean (squared uncertainty or variation),

$$\sigma^{(2)}(f) = (\Delta f)^2 := \langle (f - \langle f\rangle)^2\rangle = \langle f^2\rangle - \langle f\rangle^2\ .$$

• *Higher-order moments:* Defined using higher powers of the deviation; for all k, $\sigma^{(k)}(f) := \langle (f - \langle f\rangle)^k\rangle$. For example, $\sigma^{(3)}(f)$ is the skewness and $\sigma^{(4)}(f)$ the kurtosis. A probability distribution $P(f)$ is characterized by the set of all moments $\sigma^{(k)}(f)$, but in practice we will only use the first two, $\langle f\rangle$ and $\sigma^{(2)}(f)$.

Distributions with Several Variables: Independence and Correlations

• *Joint probabilities:* Given two subsets A, $B \in \mathcal{A}$, the joint probability for both statements on the system is $P(A \cap B) = P(A|B)\, P(B)$, where $P(A|B)$ is the conditional probability for A given B. If the two subsets are identified by giving the value of two variables x and y, we can write this is $P(x, y) = P(x|y)\, P(y)$.

• *Marginal probabilities:* The unconditional probability $P(x)$ for a set of variables x, regardless of the value of some other variables y, is calculated as $P(x) = \int \mathrm{d}y\, P(x, y)$.

• *Independent variables:* Two variables x and y are statistically independent if the conditional probability $P(x|y)$ does not depend on the value of y, so for all values of x and y

$$P(x, y) = P(x)\, P(y)\ .$$

• *Correlations:* Given two observables f and g, we can define the correlation function by

$$\sigma^{(2)}(f, g) := \langle fg\rangle - \langle f\rangle\langle g\rangle\ .$$

• *Special cases:* If x and y are independent variables, the probabilities of obtaining any pair of values for them factorize, and the correlation vanishes. Notice however that the correlation between random variables is not enough to define their dependence structure; from $\sigma^{(2)}(f, g) = 0$ we cannot conclude that f and g are independent. The maximal correlation is obtained for $f = g$, for which $\sigma^{(2)}(f, f) = \sigma^{(2)}(f)$, so the correlation of an observable with itself is its variance. For this reason, one sometimes defines

$$\mathrm{corr}(f, g) := \frac{\langle fg\rangle - \langle f\rangle\langle g\rangle}{\sqrt{(\langle f^2\rangle - \langle f\rangle^2)(\langle g^2\rangle - \langle g\rangle^2)}}\ .$$

Examples of Discrete Probability Distributions

• *Binomial:* An event consists of N repetitions of a basic choice between two outcomes, with probabilities p and $q = 1 - p$, respectively, where $p \in [0, 1]$ is a parameter. The probability that the first outcome occurs n times out of N is given by

$$P(N, n) = \binom{N}{n} p^n\, q^{N-n}\ .$$

The mean $\langle n \rangle$ and Δn can be found using the generating function $G_N(p, q) := \sum_n P(N, n) = (p + q)^N$:

$$\langle n \rangle = \Big(p\,\frac{\partial}{\partial p}\Big)\, G(p, q)\Big|_{q=1-p} = Np\ , \qquad \langle n^2 \rangle = \Big(p\,\frac{\partial}{\partial p}\Big)^2 G(p, q)\Big|_{q=1-p} = (Np)^2 + Npq\ , \quad \text{or} \quad \Delta n = \sqrt{Npq}\ .$$

In physics, a common application is to uncoupled spins, where the outcomes are the two possible values $s = \pm\hbar/2$ for each spin, with probability parameter p depending on the temperature and external magnetic field. A 1D spin chain of length N has a probability $P(N, n)$ that n of the spins have value $+1$.

Another application is to the random walk. In 1D there are two elementary outcomes (left and right), with probabilities p and $q = 1 - p$. A random walk of length N is a sequence of N independent steps, and the probability of n of them to be to the right (say) is $P(N, n)$.

• *Poisson:* Depends on a parameter μ. It is given by $P_\mu(n) = \mathrm{e}^{-\mu}\mu^n/n!$ and one finds that $\langle n \rangle = (\Delta n)^2 = \mu$.

One application is to Poisson point processes. When choosing a point at random in a manifold of volume V_0, the probability of it falling in a region R of volume V is $p = V/V_0$. For N random points, the probability of n of them falling inside R is $P(N, n)$; as $N \to \infty$ with Np = constant, this becomes a Poisson distribution.

Examples of Continuous Probability Distributions

• *Uniform probability density:* Defined for compact manifolds with a measure. On the standard 2D sphere, for example, $p(\theta, \phi) = (1/4\pi)\, \sin\theta$.

• *Gaussian:* The 1D distribution with mean $\bar{x}$ and standard deviation σ is $P_{\bar{x},\sigma}(x) = (2\pi\sigma^2)^{-1/2}\, \mathrm{e}^{-(x-\bar{x})^2/2\sigma^2}$.

• *Relationships:* In the $N \to \infty$, $p \to 0$ limit, with Np = constant =: μ, the binomial $P(N, n)$ becomes

$$P(N, n) \approx P_\mu(n) = \frac{\mathrm{e}^{-Np}\, (Np)^n}{n!}\ ,$$

while if $N \to \infty$ with p = constant, consistently with the general result of the central limit theorem we get,

$$P(N, n) \approx P_{\bar{x},\sigma}(x)\ , \qquad \text{with} \qquad \bar{x} = Np, \quad \sigma^2 = Npq\ .$$

• *Law of large numbers:* The well known and intuitive fact that if we make repeated measurements of the outcome of independent random events, as the number of measurements increases the calculated average of the values obtained will approach the mean of the quantity of interest. But we can make a more precise statement, which specifies the limiting probability distribution for the values of the sample average.

• *Central limit theorem:* If a random variable is a sum $s := x_1 + x_2 + ... + x_N$ of independent, identically distributed random variables x_i with mean μ and standard deviation σ, then in the limit $N \to \infty$ the probability distribution for s approaches a Gaussian with mean $\langle s \rangle = N\mu$ and standard deviation $\sigma_s = \sqrt{N}\,\sigma$. Similarly, for $x := N^{-1} \sum_i x_i$ the probability distribution approaches a Gaussian with mean $\langle s \rangle = \mu$ and standard deviation $\sigma_x = \sigma/\sqrt{N}$. This makes thermodynamical quantities meaningful.

On Applications in Physics

• *Examples of concepts used:* Distinction we will make between mean and average value of a physical quantity. Average over the time evolution of a system. Mean value of an observable in phase space.

Reading

• *Course textbook:* Kennett, § 1.2.
• *Other books:* Mattis & Swendsen, Ch 1; Reichl, App A; Reif, Ch 1 (very detailed discussion of random walk and binomial distributions); Schwabl, §§ 1.2 (central limit theorem) and 1.5.1; Pathria & Beale does not have a review of concepts of probability and statistics.