II. Probability

The heart of statistical mechanics

II.A General Definitions

The laws of thermodynamics are based on observations of *macroscopic bodies*, and encapsulate their thermal properties. On the other hand, matter is composed of atoms and molecules whose motions are governed by fundamental laws of classical or quantum mechanics. It should be possible, in principle, to derive the behavior of a macroscopic body from the knowledge of its components. This is the problem addressed by kinetic theory in the following lectures. Actually, describing the full dynamics of the enormous number of particles involved is quite a daunting task. As we shall demonstrate, for discussing equilibrium properties of a macroscopic system, full knowledge of the behavior of its constituent particles is not necessary. All that is required is the *likelihood* that the particles are in a particular microscopic state. Statistical mechanics is thus an inherently *probabilities* description of the system. Familiarity with manipulations of probabilities is therefore an important prerequisite to statistical mechanics. Our purpose here is to review some important results in the theory of probability, and to introduce the notations that will be used in the rest of the course.

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, $S_{\rm coin} = \{\text{head, tail}\}$, or a dice throw, $\mathcal{S}_{\rm dice} = \{1, 2, 3, 4, 5, 6\}$, or *continuous* as for the velocity of a particle in a gas, $\mathcal{S}_{\vec{v}} = \{-\infty < v_x, v_y, v_z < \infty\}$, or the energy of an electron in a metal at zero temperature, $\mathcal{S}_\epsilon = \{0 \le \epsilon \le \epsilon_F\}$. An *event* is any subset of outcomes $E \subset \mathcal{S}$, and is assigned a *probability* $p(E)$, e.g. $p_{\rm dice}(\{1\}) = 1/6$, or $p_{\rm dice}(\{1, 3\}) = 1/3$. From an axiomatic point of view, the probabilities must satisfy the following conditions:

(i) *Positivity*: $p(E) \ge 0$, i.e. all probabilities must be non-zero.

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

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

From a practical point of view, we would like to know how to assign probability values to various outcomes. There are two possible approaches:

(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

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

For example, a series of $N = 100,\ 200,\ 300$ throws of a dice may result in $N_1 = 19,\ 30,\ 48$ occurrences of 1. The ratios .19, .15, .16 provide an increasingly more reliable estimate of the probability $p_{\text{dice}}(\{1\})$.

(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.

II.B One Random Variable

As the properties of a discrete random variable are rather well known, here we focus on continuous random variables, which are more relevant to our purposes. Consider a random variable x, whose outcomes are real numbers, i.e. $\mathcal{S}_x = \{-\infty < x < \infty\}$.

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

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

$$\text{prob.}(\mathcal{S}) = \int_{-\infty}^{\infty} dx\ p(x) = 1\ . \tag{II.1}$$

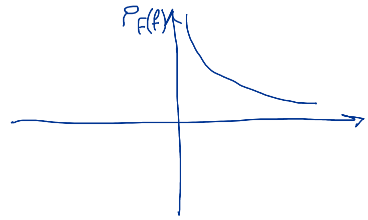
Note that since $p(x)$ is a *probability density*, it has no upper bound, i.e. $0 < p(x) < \infty$.

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

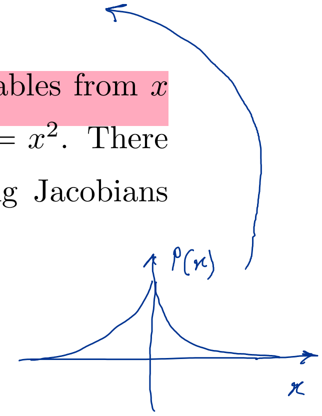
$$\langle F(x) \rangle = \int_{-\infty}^{\infty} dx\ p(x) F(x)\ . \tag{II.2}$$

The function $F(x)$ is itself a random variable, with an associated PDF of $p_F(f)df = \text{prob.}(F(x) \subset [f, f + df])$. There may be multiple solutions x_i, to the equation $F(x) = f$, and

$$p_F(f)df = \sum_i p(x_i)dx_i\,, \quad \Longrightarrow \quad p_F(f) = \sum_i p(x_i) \left|\frac{dx}{dF}\right|_{x=x_i}\ . \tag{II.3}$$



The factors of $|dx/dF|$ are the *Jacobians* associated with the change of variables from x to F. For example, consider $p(x) = \lambda \exp(-\lambda|x|)/2$, and the function $F(x) = x^2$. There are two solutions to $F(x) = f$, located at $x_\pm = \pm\sqrt{f}$, with corresponding Jacobians $|\pm f^{-1/2}/2|$. Hence,



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

for $f > 0$, and $p_F(f) = 0$ for $f < 0$. Note that $p_F(f)$ has an (integrable) divergence at $f = 0$.

• *Moments* of the PDF are expectation values for powers of the random variable. The n^{th} moment is

$$m_n \equiv \langle x^n \rangle = \int dx p(x)\; x^n. \tag{II.4}$$

• *The characteristic function*, is the generator of moments of the distribution. It is simply the Fourier transform of the PDF, defined by

! why does the Fourier transform produce the PDF moments?

$$\tilde{p}(k) = \left\langle e^{-ikx} \right\rangle = \int dx p(x)\; e^{-ikx}. \tag{II.5}$$

The PDF can be recovered from the characteristic function through the inverse Fourier transform

expanded

$$p(x) = \frac{1}{2\pi}\int dk \tilde{p}(k)\; e^{+ikx}. \tag{II.6}$$

Moments of the distribution are obtained by expanding $\tilde{p}(k)$ in powers of k,

$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n \right\rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \quad . \tag{II.7}$$

$\tilde{p}(k) = 1 + \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle$

Moments of the PDF around any point x_0 can also be generated by expanding

$$e^{ikx_0}\tilde{p}(k) = \left\langle e^{-ik(x-x_0)} \right\rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle (x-x_0)^n \rangle \quad . \tag{II.8}$$

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

$$\ln \tilde{p}(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c \quad . \tag{II.9}$$

$\ln\tilde{p}(k) = \overbrace{\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle}^{\varepsilon} - \frac{1}{2}\left(\overbrace{\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle}^{\varepsilon}\right)^2 + \frac{1}{3}\varepsilon^3$

$\rightarrow (-ik)\langle x_c\rangle + \frac{(-ik)^2}{2}\langle x_c^2\rangle + \frac{(-ik)^3}{6}\langle x_c^3\rangle + \cdots = (-ik)\langle x\rangle + \frac{(-ik)^2}{2}\left(\langle x^2\rangle - \langle x\rangle^2\right) + \frac{(-ik)^3}{6}\left(\langle x^3\rangle - 3\langle x\rangle\langle x^2\rangle + 2\langle x\rangle^3\right) + \cdots$

$\varepsilon = (-ik)\langle x\rangle + \frac{(-ik)^2}{2}\langle x^2\rangle + \frac{(-ik)^3}{6}\langle x^3\rangle + \cdots$

$\varepsilon^2 = (-ik)^2\langle x\rangle^2 + \frac{(-ik)^4}{4}\langle x^2\rangle^2 + (-ik)^3\langle x\rangle\langle x^2\rangle + \cdots$

$\varepsilon^3 = (-ik)^3\langle x\rangle^3 + \cdots$

Relations between moments and cumulants can be obtained by expanding the logarithm of $\tilde{p}(k)$ in eq.(II.7), and using

$\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle_c = \ln\tilde{p}(k) = \ln\left(1 + \underbrace{\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle}\right)$

$$\ln(1+\epsilon) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\epsilon^n}{n} \approx \varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3} - \frac{\varepsilon^4}{4}\cdots \qquad \text{(II.10)}$$

The first four cumulants are called the *mean*, *variance*, *skewness*, and *curtosis* of the distribution respectively, and are obtained from the moments as

$$\begin{aligned}
\langle x\rangle_c &= \langle x\rangle\,, && \ln(1+\varepsilon)\approx\varepsilon \\
\langle x^2\rangle_c &= \langle x^2\rangle - \langle x\rangle^2\,, && \ln(1+\varepsilon)\approx\varepsilon-\varepsilon^2/2 \\
\langle x^3\rangle_c &= \langle x^3\rangle - 3\langle x^2\rangle\langle x\rangle + 2\langle x\rangle^3\,, && \ln(1+\varepsilon)\approx\varepsilon-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3} \\
\langle x^4\rangle_c &= \langle x^4\rangle - 4\langle x^3\rangle\langle x\rangle - 3\langle x^2\rangle^2 + 12\langle x^2\rangle\langle x\rangle^2 - 6\langle x\rangle^4\,.
\end{aligned} \qquad \text{(II.11)}$$

The cumulants provide a useful and compact way of describing a PDF. → The most important info about a PDF

An important theorem allows easy computation of moments in terms of the cumulants: Represent the n^{th} cumulant graphically as a *connected cluster* of n points. The m^{th} moment is then obtained by summing all possible subdivisions of m points into groupings of smaller (connected or disconnected) clusters. The contribution of each subdivision to the sum is the product of the connected cumulants that it represents. Using this result the first four moments are easily computed as

Graphical representations of moments in terms of cumulants

$$\begin{aligned}
\langle x\rangle &= \langle x\rangle_c\,, \\
\langle x^2\rangle &= \langle x^2\rangle_c + \langle x\rangle_c^2\,, \\
\langle x^3\rangle &= \langle x^3\rangle_c + 3\langle x^2\rangle_c\langle x\rangle_c + \langle x\rangle_c^3\,, \\
\langle x^4\rangle &= \langle x^4\rangle_c + 4\langle x^3\rangle_c\langle x\rangle_c + 3\langle x^2\rangle_c^2 + 6\langle x^2\rangle_c\langle x\rangle_c^2 + \langle x\rangle_c^4\,.
\end{aligned} \qquad \text{(II.12)}$$

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

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

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

This theorem, which is the starting point for various diagrammatic computations in statistical mechanics and field theory, is easily proved by equating the expression in eqs. (II.7) and (II.9) for $\tilde{p}(k)$

$$\sum_{m=0}^{\infty}\frac{(-ik)^m}{m!}\langle x^m\rangle = \exp\left[\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle_c\right] = \prod_n\sum_{p_n}\left[\frac{(-ik)^{np_n}}{p_n!}\left(\frac{\langle x^n\rangle_c}{n!}\right)^{p_n}\right]. \qquad \text{(II.13)}$$

Equating the powers of $(-ik)^m$ on the two sides of the above expression leads to

$$\langle x^m\rangle = \sum_{\{p_n\}}{}' \prod_n \frac{m!}{p_n!(n!)^{p_n}}\langle x^n\rangle_c^{p_n}\,. \qquad \text{(II.14)}$$

The sum is restricted such that $\sum np_n = m$, and leads to the graphical interpretation given above, as the numerical factor is simply the number of ways of breaking m points into $\{p_n\}$ clusters of n points.

II.C Some Important Probability Distributions

The properties of three commonly encountered probability distributions are examined in this section.

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

$$\boxed{(x-\lambda)=y \qquad dx=dy}$$

$$\text{Gaussian} \xrightarrow{\text{Fourier}} \text{Gaussian} \qquad p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\lambda)^2}{2\sigma^2}\right] \quad . \tag{II.15}$$

$$\boxed{z=y+ik\sigma^2 \to dz=dy \to z^2=y^2-k^2\sigma^4+2yik\sigma^2}$$

The corresponding characteristic function also has a Gaussian form,

$$\to \exp^{-ik\lambda}\int_{-\infty}^{+\infty} dy\,\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{y^2}{2\sigma^2}-iky+\frac{k^2\sigma^2}{2}-\frac{k^2\sigma^2}{2}\right]$$

$$\tilde{p}(k)=\int_{-\infty}^{\infty}dx\,\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\lambda)^2}{2\sigma^2}-ikx\right]=\exp\left[-ik\lambda-\frac{k^2\sigma^2}{2}\right] \quad . \tag{II.16}$$

$$\tilde{p}(k)=e^{-ik\lambda-\frac{k^2\sigma^2}{2}}\int dz\,\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{z^2}{2\sigma^2}}$$

Cumulants of the distribution can be identified from $\ln\tilde{p}(k)=-ik\lambda-k^2\sigma^2/2$, using eq.(II.9), as

$$\langle x\rangle_c=\lambda \quad , \quad \langle x^2\rangle_c=\sigma^2 \quad , \quad \langle x^3\rangle_c=\langle x^4\rangle_c=\cdots=0 \quad . \tag{II.17}$$

The normal distribution is thus completely specified by its two first cumulants. This makes the computation of moments using the cluster expansion (eqs.(II.12)) particularly simple, and

$$\begin{aligned}
\langle x\rangle &=\lambda\,,\\
\langle x^2\rangle &=\sigma^2+\lambda^2,\\
\langle x^3\rangle &=3\sigma^2\lambda+\lambda^3,\\
\langle x^4\rangle &=3\sigma^4+6\sigma^2\lambda^2+\lambda^4, \quad \cdots \quad .
\end{aligned} \tag{II.18}$$

field theory & propagators (?)

The normal distribution serves as the starting point for most perturbative computations in field theory. The vanishing of higher cumulants implies that all graphical computations involve only products of one point, and two point (known as propagators) clusters.

(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

$$p_N(N_A)=\binom{N}{N_A}p_A^{N_A}p_B^{N-N_A} \quad . \tag{II.19}$$

The prefactor,

$$\binom{N}{N_A}=\frac{N!}{N_A!(N-N_A)!}\,, \tag{II.20}$$

is just the coefficient obtained in the binomial expansion of $(p_A+p_B)^N$, and gives the number of possible orderings of N_A events A and $N_B = N - N_A$ events B. The characteristic function for this discrete distribution is given by

Binomial expansion theorem

$$\tilde{p}_N(k) = \left\langle e^{-ikN_A}\right\rangle = \sum_{N_A=0}^{N} \frac{N!}{N_A!(N-N_A)!} p_A^{N_A} p_B^{N-N_A} e^{-ikN_A} = \left(p_A e^{-ik} + p_B\right)^N \quad . \tag{II.21}$$

The resulting cumulant generating function is

$$\ln \tilde{p}_N(k) = N \ln\left(p_A e^{-ik} + p_B\right) = N \ln \tilde{p}_1(k), \tag{II.22}$$

where $\ln \tilde{p}_1(k)$ is the cumulant generating function for a single trial. Hence, the cumulants after N trials are simply N times the cumulants in a single trial. In each trial, the allowed values of N_A are 0 and 1 with respective probabilities p_B and p_A, leading to $\underbrace{\left\langle N_A^\ell \right\rangle = p_A}_{\text{for 1 trial}}$, $= 0 \times P_B + 1 \times P_A$ for all ℓ. After N trials the first two cumulants are

$\langle N_A^2\rangle c = \langle N_A^2\rangle - \langle N_A\rangle^2$

$$\langle N_A\rangle_c = Np_A \quad , \quad \left\langle N_A^2\right\rangle_c = N(p_A - p_A^2) = Np_Ap_B \quad . \tag{II.23}$$

Stand deviation $= \sqrt{N p_A p_B}$

A measure of fluctuations around the mean is provided by the *standard deviation*, which is the square root of the variance. While the mean of the binomial distribution scales as N, its standard deviation only grows as $\sqrt{N}$. Hence, the *relative* uncertainty becomes smaller for large N. → It has important implications for statistical mechanics. How?

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

$$p_N\left(\{N_A, N_B, \cdots, N_M\}\right) = \frac{N!}{N_A!N_B!\cdots N_M!} p_A^{N_A} p_B^{N_B} \cdots p_M^{N_M} \quad . \tag{II.24}$$

(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 $[t, t+dt]$ is proportional to dt as $dt \to 0$, $p = \alpha dt$ why is it important?

(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 $N = T/dt \gg 1$ segments of size dt. In each segment, an event occurs with probability $p = \alpha dt$, and there is no event with probability $q = 1 - \alpha dt$. As the probability

See the theorem on next page!

of more than one event in dt is too small to consider, the process is equivalent to a binomial one. Using eq.(II.21), the characteristic function is given by

$$\tilde{p}(k) = \left(pe^{-ik} + q\right)^n = \lim_{dt\to 0}\left[1 + \alpha dt\left(e^{-ik} - 1\right)\right]^{T/dt} = \exp\left[\alpha(e^{-ik} - 1)T\right] \quad . \tag{II.25}$$

The Poisson PDF is obtained from the inverse Fourier transform in eq.(II.6) as

$$p(x) = \int_{-\infty}^{\infty}\frac{dk}{2\pi}\exp\left[\alpha\left(e^{-ik} - 1\right)T + ikx\right] = e^{-\alpha T}\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ikx}\sum_{M=0}^{\infty}\frac{(\alpha T)^M}{M!}e^{-ikM} \ , \tag{II.26}$$

using the power series for the exponential. The integral over k is

$$\boxed{\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ik(x-M)} = \delta(x - M)}\ , \tag{II.27}$$

This relation was also used in the derivation of momentum eigenstates

leading to

$$p_{\alpha T}(x) = \sum_{m=0}^{\infty} e^{-\alpha T}\,\frac{(\alpha T)^M}{M!}\delta(x - M)\ . \tag{II.28}$$

→ Math was smart enough to say "no! x can only be integers." → It was derived from binomial distribution

The probability of M events is thus $p_{\alpha T}(M) = e^{-\alpha T}(\alpha T)^M/M!$. The cumulants of the distribution are obtained from the expansion

$$\ln\tilde{p}_{\alpha T}(k) = \alpha T(e^{-ik} - 1) = \alpha T\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}, \quad \Longrightarrow \quad \langle M^n\rangle_c = \alpha T \quad . \tag{II.29}$$

All cumulants have the same value, and the moments are obtained from eqs.(II.12) as

$$\langle M\rangle = (\alpha T), \quad \left\langle M^2\right\rangle = (\alpha T)^2 + (\alpha T), \quad \left\langle M^3\right\rangle = (\alpha T)^3 + 3(\alpha T)^2 + (\alpha T). \tag{II.30}$$

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 dV is ndV, and they are assumed to be independent, the number of stars in a volume V is described by a Poisson process as in eq.(II.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 $p_{ndV}(1)$, of finding one star in the shell of volume $dV = 4\pi R^2 dR$ at a distance R. Both $p_{nV}(0)$ and $p_{ndV}(1)$ can be calculated from eq.(II.28), and

$$\begin{aligned} p(R)dR = p_{nV}(0)\,p_{ndV}(1) =& e^{-4\pi R^3 n/3}\,e^{-4\pi R^2 ndR}\,4\pi R^2 ndR, \\ \Longrightarrow \qquad p(R) =& 4\pi R^2 n\exp\left(-\frac{4\pi}{3}R^3 n\right) \ . \end{aligned} \tag{II.31}$$

Included here for easy reference:

Theorem

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x \quad (= \exp(x))$$

Proof.
If $x = 0$ then the result clearly holds and if $x \neq 0$ then

$$\begin{aligned}
\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n &= \lim_{n\to\infty}\exp\left(n\ln\left(1+\frac{x}{n}\right)\right) = \lim_{n\to\infty}\exp\left(x\left(\frac{\ln(1+x/n)}{x/n}\right)\right)\\
&= \lim_{h\to 0}\exp\left(x\left(\frac{\ln(1+h)}{h}\right)\right)\\
&= \exp\left(x\left(\lim_{h\to 0}\frac{\ln(1+h)}{h}\right)\right)\\
&= \exp(x)
\end{aligned}$$

using the continuity of the exp( ) function and since $e^0 = 1$ so $\ln(1) = 0$ we have that

$$\lim_{h\to 0}\frac{\ln(1+h)}{h} = \lim_{h\to 0}\frac{\ln(1+h)-\ln(1)}{h} = \left.\frac{d(\ln(x))}{dx}\right|_{x=1} = \left.\left(\frac{1}{x}\right)\right|_{x=1} = 1\,.$$

□