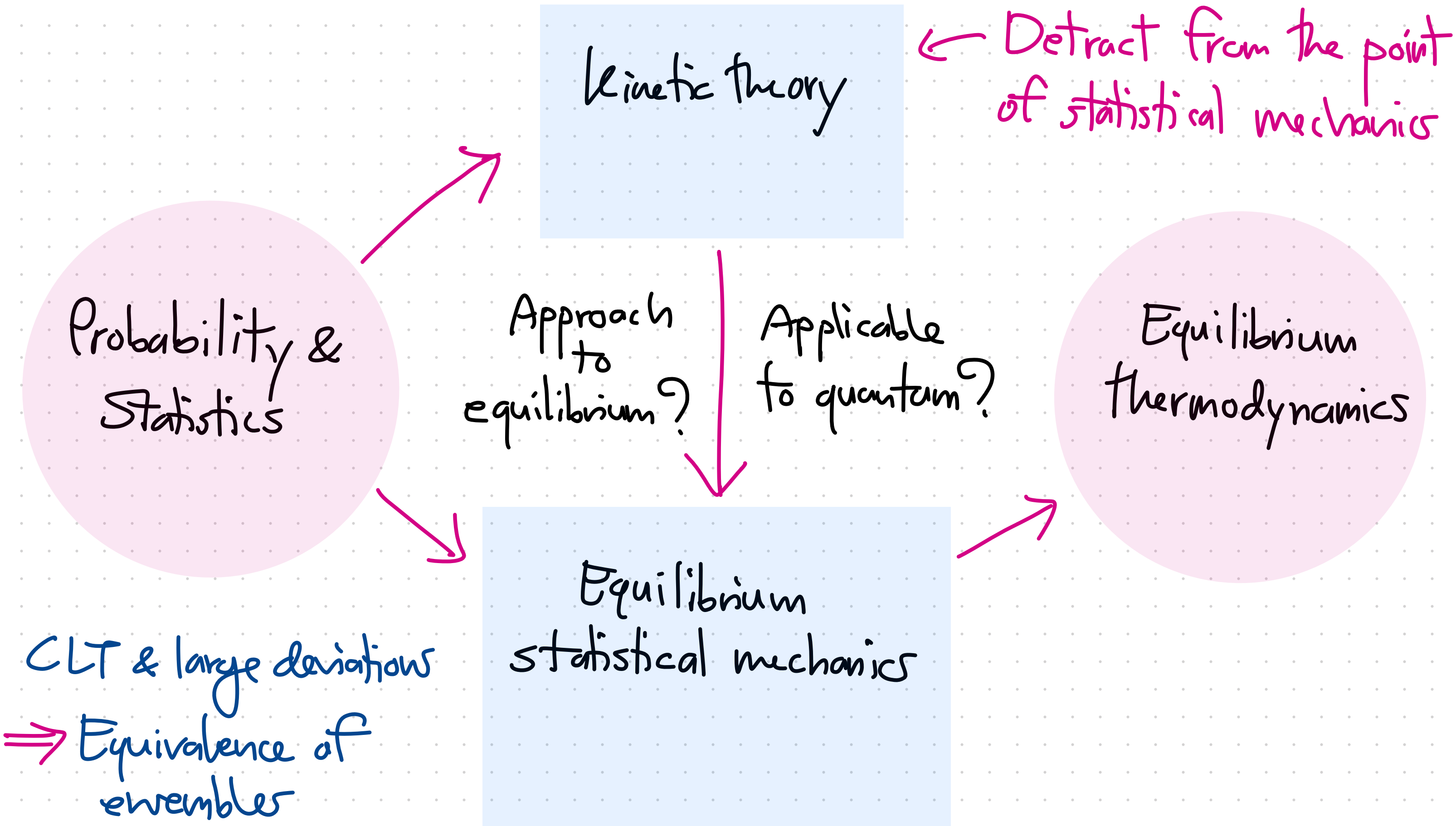
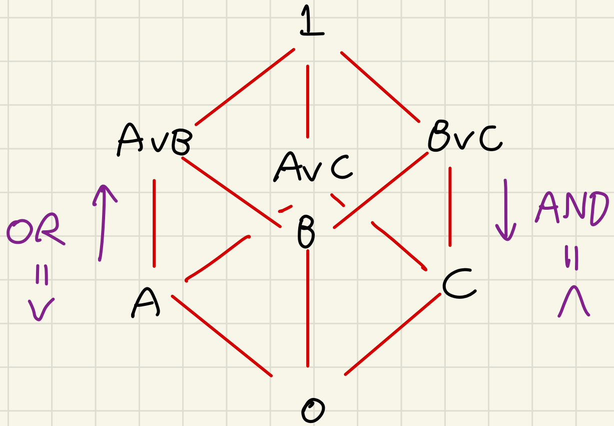

Kinetic theory
← Detract from the point of statistical mechanics
Probability & Statistics
Approach to equilibrium?
Applicable to quantum?
Equilibrium thermodynamics
Equilibrium statistical mechanics
CLT & large deviations
⟹ Equivalence of ensembles

Probability & statistics



Boolean algebra	Set-theoretic
$\vee$	$\cup$
$\wedge$	$\cap$
$\neg$	Complement

Hesse diagram of Boolean algebra with three elementary events A, B, C

Probability axioms

Notations $p \leftrightarrow P \leftrightarrow Pr$

① $p(A) \geqslant 0$

② A is certain $\Longleftrightarrow$ $p(A) = 1$

③ $p(A \vee B) = p(A) + p(B)$ if $A \wedge B = \emptyset$ ← Mutually exclusive

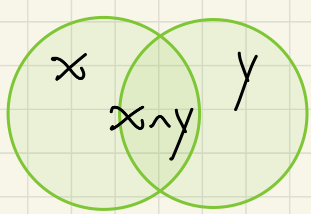
Random variable X has values $\underbrace{x_1, x_2, \dots, x_N}_{\text{Elementary events/alternatives}}$

$Pr(X=x)$ or $Pr_X(x)$

Joint $p_{X,Y}(x,y)$ **Marginal** $\begin{cases} p_X(x) = \sum_y p_{X,Y}(x,y) \\ p_Y(y) = \sum_x p_{X,Y}(x,y) \end{cases}$

Conditional

$\underbrace{p_{X|Y}(x|y)}_{\text{Prob. of } x \text{ given that } Y=y} = p_{X,Y}(x,y) / p_Y(y)$



Statistical experiment

From IFxNAS Summer School

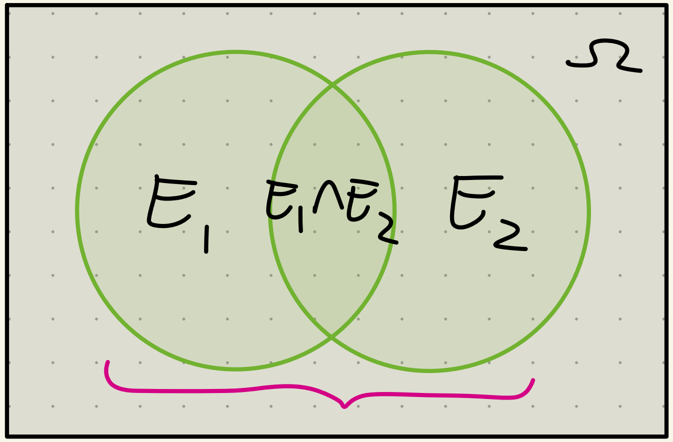
Denote by Ω the set of all possible outcomes/events. (Ω is called a sample space.) A model (possibly unknown) of a statistical experiment assigns probabilities to all events $E \subset \Omega$.

Random variable

$Pr(X = E)$

Logical AND $\wedge$, OR $\vee$, NOT $\neg$

Probability axioms



$E_1 \vee E_2$

① $Pr(E) \geq 0$

② $Pr(\Omega) = 1$

③ $Pr(E_1 \vee E_2) = Pr(E_1) + Pr(E_2)$

if $E_1 \wedge E_2 = 0$

Everything else can be derived from these three axioms, for example that $Pr(\neg E) = 1 - Pr(E)$ or $Pr(E_1 \vee E_2)$ $= Pr(E_1) + Pr(E_2) - Pr(E_1 \wedge E_2)$ for an arbitrary pair of events.

$$Pr(E_1 | E_2) := \frac{Pr(E_1 \wedge E_2)}{Pr(E_2)}$$ (Conditional probability)

$$Pr(E_1 | E_2) = \frac{Pr(E_2 | E_1)\, Pr(E_1)}{Pr(E_2)}$$ (Bayes' rule)

$$\text{Mutually exclusive} \iff E_1 \wedge E_2 = \emptyset$$

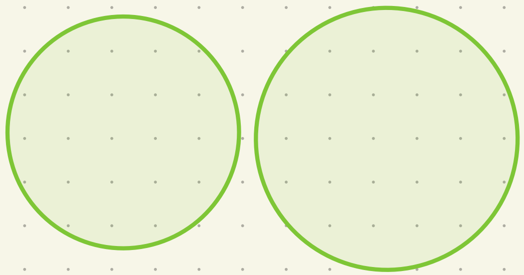
$$\text{Independent} \iff \Pr(E_1 \wedge E_2) = \Pr(E_1)\Pr(E_2)$$

Remarks

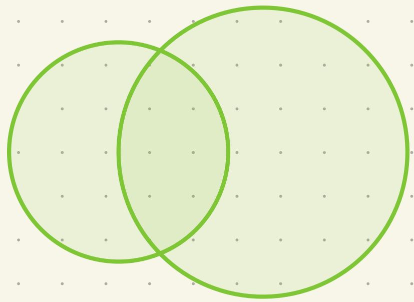
① Mutually exclusive events can't be independent; if a coin flip comes up head, I know that it did not come up tail.

- $E_1 \wedge E_2 = \emptyset \Rightarrow \Pr(E_1 \wedge E_2) = 0$. But $\Pr(E_1 \wedge E_2) = \Pr(E_1)\Pr(E_2)$ can't be 0 if $\Pr(E_1), \Pr(E_2) > 0$. $\square$

② Independence can't be easily visualized on a Venn diagram, in contrast to mutual exclusivity, since an area in a Venn diagram doesn't conventionally represent probability



Mutually exclusive



Merely overlapping doesn't gaurantee independence

Bayes $\quad p_{X|Y}(x|y)\,p_Y(y) = p_{Y|X}(y|x)\,p_X(x) = p_{X,Y}(x,y)$

Usual form $\quad p(y|x) = p(x|y)\,p(y) / \underbrace{p(x)}$

("Law of total probabilities") $\displaystyle\sum_y p(x,y) = \sum_y p(x|y)\,p(y)$

We almost always omit the subscripts that indicate which random variables are assigning the probabilities.

How do we assign probabilities in the real world?

1. Objective (Frequentist) $\leftarrow \; p(x) = \lim_{N\to\infty} \frac{N_x}{N}$
2. Subjective (Bayesian) $\leftarrow$ Laplace (1818)
3. "I just use probability theory, man"

We will adopt a pragmatic stance that we assign subjective probabilities based on the best available information (this is our "model" of the system) and then experimentally check the model by comparing our probability assignments to the measured frequencies.

$$p(\text{Model}\,|\,\text{Data}) \propto p(\text{Data}\,|\,\text{Model})\;p(\text{Model})$$

Equality if we divide by p(Data) but we usually don't know that.

Laws of large number (LOL)

We will derive Markov $\rightarrow$ Chebyshev $\rightarrow$ Weak LOL

Moments $\langle x^n \rangle = \int dx\, x^n p(x)$

Mean $\langle x \rangle = \int dx\, x\, p(x)$

Variance $\sigma_X^2 := \langle (x - \langle x \rangle)^2 \rangle$
$$= \langle x^2 \rangle - \langle x \rangle^2$$

Deviation
$$\Delta X = X - \langle X \rangle$$
$$\langle \Delta X \rangle = 0$$

For a non-negative random variable X

$$\boxed{p(X \geq a) \leq \frac{\langle x \rangle}{a}} \quad \text{(Markov)}$$

- $\langle x \rangle = \int_0^\infty dx\, x\, p(x) \geq \int_a^\infty dx\, x\, p(x)$
$$\geq \int_a^\infty dx\, a\, p(x) = a\, p(X \geq a) \qquad \square$$

Ineq. saturated if $\mathrm{Supp}(p) \in \{0, a\}$ (Delta functions at $x=0$ and a)

Applying this to the ΔX^2 to obtain

$$p(\Delta X^2 \geq a^2) \leq \frac{\overbrace{\langle \Delta X^2 \rangle}^{=\sigma_X^2}}{a^2} \iff \boxed{p(|x - \langle x \rangle| \geq a) \leq \frac{\sigma_X^2}{a^2}}$$

(Chebyshev)

Let $X_1, X_2, \dots, X_N$ be independent, identically distributed (i.i.d.) random variables with finite mean and variance.

Sample mean $S = \frac{1}{N}\sum_{k=1}^{N} X_k$ from many trials

Mean of sample mean $\langle S\rangle = \frac{1}{N}\sum_{k=1}^{N}\underbrace{\langle X_k\rangle}_{\langle X\rangle \text{ i.i.d.}} = \boxed{\langle X\rangle}$

$$\langle S^2\rangle = \frac{1}{N^2}\sum_{k,l}\underbrace{\langle X_k X_l\rangle}$$

$$\langle(\langle X\rangle + \Delta X_k)(\langle X\rangle + \Delta X_l)\rangle$$

$$= \langle X\rangle^2 + 2\cancel{\langle \Delta X_k\rangle}^{0}\cancel{\langle \Delta X_l\rangle}^{0} + \langle \Delta X_k \Delta X_l\rangle$$

$$= \langle X\rangle^2 + \delta_{kl}\langle \Delta X^2\rangle \quad \text{i.i.d.}$$

$$\therefore \langle S^2\rangle = \langle X\rangle^2 + \frac{1}{N}\langle \Delta X^2\rangle = \langle X\rangle^2 + \frac{\sigma_X^2}{N}$$

Thus, $\langle \Delta S^2\rangle = \langle S^2\rangle - \langle S\rangle^2 = \boxed{\frac{\sigma_X^2}{N}}$

Now we are equipped to prove the weak LOL.

$$p(|S - \langle S\rangle| \geq \epsilon) \leq \frac{\langle \Delta S^2\rangle}{\epsilon^2}$$

$$\Downarrow$$

$$\boxed{p(|S - \langle X\rangle| \geq \epsilon) \leq \frac{\sigma_X^2}{N}}$$

Taking the limit $N \to \infty$, we see that

$$\lim_{N\to\infty} p(|S - \langle X \rangle| \leq \epsilon) \to 1 \quad \left(\begin{matrix}\text{Weak} \\ \text{LOL}\end{matrix}\right)$$

This provides a link between the theoretical mean and the sample mean. But be careful about what the weak law of large number doesn't say. It doesn't say that S becomes $\langle X \rangle$ identically. It only says that S takes the value $\langle X \rangle$ with overwhelming probability (convergence in probability)

How to obtain this relation at the level of frequencies?

$\Rightarrow$ Next

Binomial & multinomial distributions

Let X be a discrete random variable that can take on 2 values x_1, x_2 with probability p_1 and p_2 respectively.

x_1 — Head, x_2 — Tail, p_1 — p, p_2 — q

In N trials (N independent and identical coin tosses),

$$\begin{pmatrix}\text{Prob. for a}\\ \text{particular sequence}\\ \text{with } n \text{ heads}\end{pmatrix} = p^n q^{N-n}$$

$$\begin{pmatrix}\text{Prob. for any}\\ \text{sequence with}\\ n \text{ heads}\end{pmatrix} = \binom{N}{n} p^n q^{N-n} =: p(n)$$

Normalization $\displaystyle\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} \overset{\text{Binomial thm}}{=} (p+q)^N = 1$

$$\langle n \rangle = \sum_{n=0}^{N} n\, p(n) = \sum n \binom{N}{n} p^n q^{N-n}$$

$$= p \frac{\partial}{\partial p} \sum \binom{N}{n} p^n q^{N-n} = p \frac{\partial}{\partial p} (p+q)^N$$

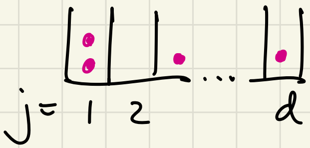
$$= Np\, (\cancel{p+q})^{N-1} = \boxed{Np} \qquad \boxed{\sigma_n^2 = p(1-p)N}$$

with $p+q = 1$

d outcomes

$$p(n_1,\dots,n_d) = \frac{N!}{n_1!\cdots n_d!}\, p_1^{n_1}\cdots p_d^{n_d}$$

$$\sum_{j=1}^{d} n_j = N$$



Multinomial theorem

$$\sum_{\substack{n_1,\dots,n_d \\ \sum n_j = N}} p(n_1,\dots,n_d) \overset{\downarrow}{=} (p_1+\cdots+p_d)^N = 1$$

$$\langle n_j\rangle = p_j \frac{\partial}{\partial p_j}\sum p(n_1,\dots,n_d)$$

$$= N_j p_j (p_1+\cdots+p_d)^{N-1} = N_j p_j$$

Correlation matrix

$$\langle n_j n_k\rangle = \sum_{n_1,\dots,n_d} n_j n_k\, p(n_1,\dots,n_d)$$

$$= p_j\frac{\partial}{\partial p_j}\left[p_k\frac{\partial}{\partial p_k}\left(\sum p(n_1,\dots,n_d)\right)\right]$$

$$= p_j\frac{\partial}{\partial p_j}\left[N p_k (p_1+\cdots+p_d)^{N-1}\right]$$

$$= N p_j \delta_{jk}(\overbrace{p_1+\cdots+p_d}^{1})^{N-1} + N(N-1)p_j p_k(\overbrace{\qquad\qquad}^{1})^{N-2}$$

$$= N^2 p_j p_k + N p_j(\delta_{jk} - p_k)$$

$$\langle \Delta n_j \Delta n_k \rangle = \langle (n_j - \langle n_j \rangle)(n_k - \langle n_k \rangle) \rangle$$
$$= \langle n_j n_k \rangle - \langle n_j \rangle \langle n_k \rangle$$
$$= N p_j (\delta_{jk} - p_k)$$

Variances $\langle \Delta n_j^2 \rangle = N p_j (1 - p_j)$

Frequencies

$$f_j = \frac{n_j}{N} \Longrightarrow \begin{cases} \langle f_j \rangle = \dfrac{\langle n_j \rangle}{N} = p_j \\ \langle f_j f_k \rangle = \dfrac{\langle n_j n_k \rangle}{N^2} = p_j p_k + \dfrac{p_j}{N}(\delta_{jk} - p_k) \\ \langle \Delta f_j \Delta f_k \rangle = \dfrac{p_j}{N}(\delta_{jk} - p_k) \quad \boxed{\text{Important}} \\ \langle \Delta f_j^2 \rangle = \dfrac{p_j(1-p_j)}{N} \leftarrow \text{Variance goes as } 1/N \end{cases}$$

Therefore, the weak LOL implies that

$$p(\underbrace{|f_j - p_j| \geq \epsilon_j, \forall j}_{\vec{f} \text{ is within a hypercube}}) \geq p\Big[\underbrace{\textstyle\sum (f_j - p_j)^2}_{\vec{f} \text{ within the sphere inside the hypercube}} \geq \epsilon_j \ \forall j\Big]$$

$$= \frac{\sum_j \langle \Delta f_j^2 \rangle}{\epsilon^2} = \frac{\cancelto{1}{\sum_j p_j} - \sum p_j^2}{N\epsilon^2}$$

$$p(|x - \langle x \rangle| \geq a) \leq \frac{\sigma_x^2}{a^2}$$

(Chebyshev)

Moments $\langle x^n \rangle = \int dx\, x^n p(x)$

Not clear yet that knowing all the moments is equivalent to knowing the PDF

Characteristic function is just the Fourier transform of the PDF

$$\varphi_X(k) = \langle e^{-ikx} \rangle = \int dx\, e^{-ikx} p(x)$$

Inverse FT gives the PDF back

$$p(x) = \frac{1}{2\pi} \int dk\, e^{ikx} \varphi_X(k)$$

The characteristic function "generates" all the moments of $p(x)$ (up to the factor $(-i)^n$), meaning that it is a polynomial (here in k) whose coefficients are the moments:

$$\varphi_X(k) = \left\langle 1 - ikx + \cdots + \frac{(-ikx)^n}{n!} + \cdots \right\rangle$$

$$= \underset{\substack{\uparrow \\ \text{Normalization}}}{1} - i\underset{\substack{\uparrow \\ \text{Mean}}}{\langle x \rangle} k - \frac{\langle x^2 \rangle}{2} k^2 + \cdots + (-i)^n \frac{\langle x^n \rangle}{n!} k^n + \cdots$$

Moments of $p(x)$ can be computed by differentiating w.r.t. k and then setting $k=0$.

$$\frac{d^n}{dk^n} \varphi_X(k) = \int dx\, p(x) \frac{d^n}{dk^n} e^{-ikx} \Big|_{k=0}$$

$$= (-i)^n \int dx\, x^n p(x) = (-i)^n \langle x^n \rangle$$

Cumulants

The cumulant generating function is the natural log of the characteristic function.

Cumulant (Defined implicitly)

$$K_X(k) = \ln \varphi(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c$$

Beware! Log and integration can't be interchanged

$$\ln\left[\int_0^\infty dx\, e^{-x}\right] = \ln 1 = 0 \neq \int_0^\infty dx \ln e^{-x} = -\int_0^\infty dx\, x = -\infty$$

$$\ln(1+\epsilon) = -\sum_{n=1}^{\infty} (-1)^n \frac{\epsilon^n}{n} = \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} + \cdots$$

$$\Rightarrow \ln\left(1 - i\langle x \rangle k - \frac{\langle x^2 \rangle}{2} k^2 + \cdots\right)$$

$$= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(-i\langle x \rangle k - \frac{\langle x^2 \rangle}{2} k^2 + \cdots\right)^n$$

First-order term: $-i\langle x \rangle k \Longrightarrow \boxed{\langle x \rangle_c = \langle x \rangle}$

Second-order term: $\underbrace{-\frac{\langle x^2 \rangle}{2} k^2}_{\text{From } n=1} + \underbrace{\frac{1}{2} \langle x \rangle^2 k^2}_{\text{From } n=2}$

Compare with $-\langle x^2 \rangle k^2/2 \Longrightarrow \boxed{\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \sigma_X^2}$

This is the part of the 2nd moment that doesn't come from the 1st moment.

Central limit theorem (CLT) is stronger

$$p(S) = \frac{1}{\sqrt{2\pi(\Delta S)^2}} \exp\left[-\frac{(S-\langle S\rangle)^2}{(\Delta S)^2}\right]$$

To show convergence to a Gaussian distribution (not rigorous), we only need to show that all the cumulants higher than the 3rd cumulants vanish in the $N \to \infty$ limit.

Joint cumulant

$$\langle x_1^{n_1} * \cdots * x_N^{n_N} \rangle = (-i)^{n_1} \frac{\partial^{n_1}}{\partial k_1^{n_1}} \cdots (-i)^{n} \frac{\partial^{n_N}}{\partial k_N^{n_N}} \ln \tilde{\varphi} \Big|_{\vec{k}=0}$$

$\langle x_1 * x_2 \rangle = 0$ if x_1 and x_2 are independent random variables

Independent random variables

$$\tilde{\varphi}_{\Sigma x/\sqrt{N}}(k) = \langle e^{-i\Sigma k_j x_j/\sqrt{N}} \rangle = \prod_{j=1}^{N} \tilde{\varphi}_{x_j}\left(\frac{k_j}{\sqrt{N}}\right)$$

$$\Rightarrow K_{\Sigma x/\sqrt{N}}(k) = \ln \tilde{\varphi}_{\Sigma x/\sqrt{N}}(k) = \sum_{j=1}^{N} K_{x_j}\left(\frac{k_j}{\sqrt{N}}\right)$$

So the nth cumulant will be attached to the $\frac{k^n}{N^{n/2}}$ term in the cumulant generating function.

Identical and cumulant bounded by a const. C

$$\Rightarrow \text{nth cumulant} = \frac{\sum \langle x^n \rangle_c}{N^{n/2}} \leq \frac{NC}{N^{n/2}} = N^{1-\frac{n}{2}} C$$

In the limit $N \to \infty$, the number of times x_j appears in a sequence is $\langle n_j \rangle = N p_j$ ("typical" sequence)

$$\left(\begin{array}{c}\text{Prob. of any} \\ \text{given typical} \\ \text{sequence}\end{array}\right) = p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d} = p_1^{N p_1} \cdots p_d^{N p_d}$$

$$= 2^{N p_1 \log p_1} \cdots 2^{N p_d \log p_d}$$

$$= 2^{N(p_1 \log p_1 + \cdots + p_d \log p_d)}$$

$$= 2^{-N H(X)}$$

$$\boxed{d^N = 2^{N \log d}}$$

Count the number of typical sequences

$$\ln\left(\overbrace{\frac{N!}{n_1! \cdots n_d!}}\right) = \ln N! - \sum_j \ln n_j!$$

Stirling

$\ln N! \sim N \ln N - N$

$$= N \ln N - \cancel{N} - \sum_j (n_j \ln n_j - \cancel{n_j})$$

$$= N \ln N - \underbrace{\sum_j N p_j \ln (N p_j)}$$

$$\sum_j N p_j (\ln N + \ln p_j)$$

$$\parallel$$

$$N \ln N + N \sum_j p_j \ln p_j$$

$$= N\left(-\sum_j p_j \ln p_j\right) = N H(X)$$

Each typical sequence appears with prob. $2^{-NH(X)}$ and there are $2^{NH(X)}$ of them, so they take all the probability.