2.1 General definitions

(i) Positivity: $p(E) \geq 0$

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

(iii) Normalization: $P(S) = 1$

Objective and Subjective probabilities

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

2.2 One random variable: (focus on continuous random variable)

$S_x = \{-\infty < x < \infty\}$

CDF: the cumulative prob. func.

1. $P(x) = \text{prob}(E \subset [-\infty, x])$
2. $P(x)$ must be a monotonically increasing func. of x
3. $P(-\infty) = 0$ and $P(+\infty) = 1$

PDF: the prob. density func.

1. $p(x) = \dfrac{dP(x)}{dx}$
2. $p(x)dx = \text{prob}(E \in [x, x+dx])$
3. $p(x)$ is positive
4. $\text{prob}(S) = \int_{-\infty}^{+\infty} dx\, p(x) = 1$ (integrable)
5. unit / dimension of $p(x)$ is $[x]^{-1}$
6. no upper bound ($0 < p(x) < \infty$) but contain divergences

The expectation value of $F(x)$

1. $\langle F(x) \rangle = \int_{-\infty}^{+\infty} dx\, p(x) F(x)$
2. $p_F(f)df = \text{prob}(F(x) \in [f, f+df])$
3. $p_F(f)df = \sum_i p(x_i) dx_i \Rightarrow p_F(f) = \sum_i p(x_i) \left|\dfrac{dx}{dF}\right|_{x=x_i}$

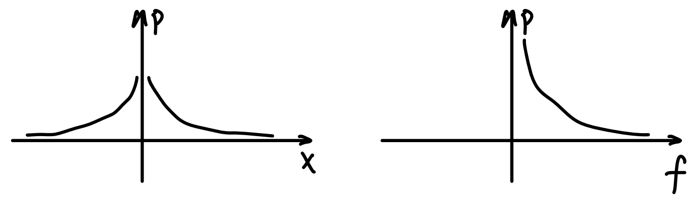
$\left|\dfrac{dx}{dF}\right|$ are Jacobians associated with the change of variables from x to F.

Ex: $p(x) = \frac{\lambda}{2} \exp(-\lambda|x|)$ & $F(x) = x^2$

① $F(x) = f \Rightarrow x_{\pm} = \pm\sqrt{f} \Rightarrow \text{Jacobians} = \left|\pm \dfrac{f^{-\frac{1}{2}}}{2}\right|$

② $p_F(f) = \sum_i p(x_i) \left|\dfrac{dx}{dF}\right|_{x=x_i} = \dfrac{\lambda}{2}\exp(-\lambda\sqrt{f})\left(\left|\dfrac{1}{2\sqrt{f}}\right| + \left|-\dfrac{1}{2\sqrt{f}}\right|\right) = \dfrac{\lambda \exp(-\lambda\sqrt{f})}{2\sqrt{f}}$

Prob. density func



• Moments of PDF:

expectation values for powers of the random variable $\quad m_n = \langle x^n \rangle = \int dx\, p(x)\, x^n$

• characteristic func:

generator of moments of the distribution

$$\tilde{p}(k) = \langle e^{-ikx} \rangle = \int dx\, p(x) e^{-ikx} \quad \text{(Fourier transform of PDF)}$$

$$p(x) = \frac{1}{2\pi} \int dk\, \tilde{p}(k) e^{ikx} \quad \text{(Inverse Fourier transform)}$$

$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n \right\rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \quad \text{(obtain moments of the distribution)}$$

$$e^{ikx_0} \tilde{p}(k) = \langle e^{-ik(x-x_0)} \rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle (x-x_0)^n \rangle \quad \text{(Moments of PDF at any } x_0\text{)}$$

• The cumulant generating func:

logarithm of the characteristic func

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

The relation of moments and cumulant:

As $\ln(1+\epsilon) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\epsilon^n}{n}$ and $\tilde{p}(k) = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle = 1 + \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle$

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

$$= \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle - \frac{1}{2} \left[\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right]^2 + \frac{1}{3} \left[\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right]^3 + \cdots\cdots$$

$$= (-ik)\langle x \rangle + \frac{(-ik)^2}{2} \langle x^2 \rangle - \frac{1}{2} (-ik)^2 \langle x \rangle^2 + \cdots\cdots$$

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

$$+ \cdots\cdots$$

✱ Mathematically get moments and cumulants

$$\langle x^n \rangle = \left. \frac{d^n M_x(t)}{dt^n} \right|_{t=0} \quad \& \quad \langle x^n \rangle_c = \left. \frac{d^n \ln M_x(t)}{dt^n} \right|_{t=0} \quad M_x(t) = \langle \exp(tX) \rangle$$

$(t = -ik)$

first four cumulants:

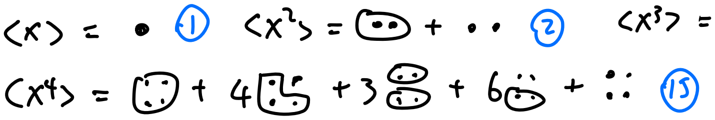
mean $\langle x \rangle_c = \langle x \rangle$

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

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

kurtosis / curtosis $\langle x^4 \rangle_c = \langle x^4 \rangle - 4\langle x^3 \rangle \langle x \rangle - 3\langle x^2 \rangle^2 + 12\langle x^2 \rangle \langle x \rangle^2 - 6\langle x \rangle^4$

✱ have some memorical tips with graphical computations.



use cumulants to calculate moments.

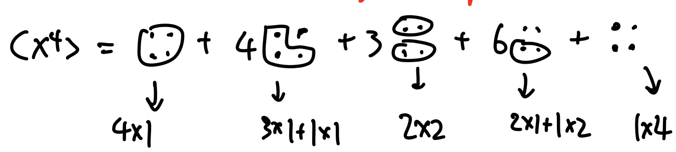
start point for various diagrammatic computations in stats. mech. and field theory.

Proof: $$\sum_{m=0}^{\infty} \frac{(-ik)^m}{m!}\langle x^m\rangle = \exp\left(\sum_{n=1}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle_c\right) = \prod_{n=1}^{\infty}\sum_{p_n}\left[\frac{(-ik)^{np_n}}{(p_n)!}\left(\frac{\langle x^n\rangle_c}{n!}\right)^{p_n}\right]$$

$$\langle x^m\rangle = \sum_{\{p_n\}} \overset{\star}{\boxed{m!\prod_{n=1}^{\infty}\frac{1}{p_n!\,(n!)^{p_n}}}}\langle x^n\rangle_c^{p_n} \quad \text{w.r.t. } \sum n p_n = m$$

☆ numerical factor is the number of ways of breaking m points into $\{p_n\}$ clusters of n points. ☆(total num is Bell number)

Ex:



2.3 Some important prob. distributions.

Normal (Gaussian) distribution

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

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

3. $\ln \tilde{p}(k) = -ik\lambda - \frac{k^2\sigma^2}{2}$

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

5. $\langle x\rangle = \lambda$; $\langle x^2\rangle = \sigma^2 + \lambda^2$; $\langle x^3\rangle = 3\sigma^2\lambda + \lambda^3$
   $\langle x^4\rangle = 3\sigma^4 + 6\sigma^2\lambda^2 + \lambda^4$

Summary:
① most perturbative computations in field theory
② vanishing of higher cumulants implies (only one)
all graphical computations involves only product of one/two-point clusters
(two point clusters $\equiv$ propagators)

Binomial distribution:

random variable A and B with prob. p_A and $p_B = 1 - p_A$
N trials the event A occurs N_A times

1. $p_N(N_A) = \binom{N}{N_A} p_A^{N_A} p_B^{N-N_A}$ $\equiv$ expansion of $(p_A + p_B)^N$

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

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

4. $\langle N_A\rangle_c = Np_A$; $\langle N_A^2\rangle_c = N(p_A - p_A^2) = Np_A p_B$

5. mean of binom. dist. scales as N, but standard devi. grows as $\sqrt{N}$;
   so relative uncertainty becomes smaller for large N.

*** multinomial distribution:**

outcomes $\{A, B, \cdots, M\}$ with probs. $\{p_A, p_B, \cdots, p_M\}$
$N = N_A + N_B + \cdots + N_M$

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

Poisson distribution:

Ex. radioactive decay

(a) prob. of one or only one event (decay) in the interval $[t, t+dt]$ is proportional to dt as $dt \to 0$

(b) prob. of events (decay) at different intervals are independent of each other.

$\to$ observing exactly M decays in the interval T

$N = T/dt \gg 1$

In each segment dt, an event (decay) occurs with $p = \alpha dt$

doesn't occur with $q = 1 - \alpha dt$

As the prob. of more than one event in dt is too small, so the process can be regarded as binomial distribution.

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

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

as $\int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{ik(x-M)} = \delta(x-M)$;

3. $$P_{\alpha T}(x) = \sum_{M=0}^{\infty} e^{-\alpha T} \frac{(\alpha T)^M}{M!} \delta(x-M) \overset{M=x}{\Longrightarrow} P_{\alpha T}(M) = e^{-\alpha T} \frac{(\alpha T)^M}{M!}$$

4. $$\ln \tilde{p}(k) = \alpha\left(e^{-ik} - 1\right)T = \alpha T \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \Rightarrow \langle M^n \rangle_c = \alpha T$$

5. $$\langle M \rangle = \alpha T \ ; \ \langle M^2 \rangle = (\alpha T) + \alpha^2 T^2 \ ; \ \langle M^3 \rangle = (\alpha T)^3 + 3(\alpha T)^2 + (\alpha T)$$

Ex. star distribution:

nearest star at distance R

density $= n$; $\alpha = n$; $T = V$ or dV

$$p(R)\,dR = P_{nV}(0)\, P_{ndV}(1) = e^{-\frac{4\pi}{3}R^3 n} \cdot \frac{\left(\frac{4\pi}{3}R^3 n\right)^0}{0!} \cdot e^{-4\pi R^2 n\, dR} \cdot \frac{\left(4\pi R^2 n\, dR\right)^1}{1!}$$

$$\Rightarrow p(R) \approx 4\pi R^2 n \exp\left(-\frac{4\pi}{3}R^3 n\right) \quad \left(\text{as } dR \to 0,\ e^{-4\pi R^2 n\, dR} \to 1\right)$$

2.4 Many random variables

$S_x = \{-\infty < x_1, x_2, \cdots, x_N < \infty\}$, e.x. location and velocity of gas (6 coordinates)

- Joint PDF $p(x)$

$d^N x = \prod_{i=1}^{N} dx_i$

$$P_x(S) = \int d^N x \, p(x) = 1$$

$$p(x) = \prod_{i=1}^{N} p_i(x_i) \quad \text{(all } N \text{ random variables are independent)}$$

- Unconditional PDF : the behaviour of a subset of random variables, independent of the values of others

$$p(x_1, \cdots, x_m) = \int \prod_{i=m+1}^{N} dx_i \, p(x_1, \cdots, x_N)$$

(All random variables are independent, and unconditional PDF = conditional PDF)

- Conditional PDF : the behaviour of a subset of random variables

e.x. $p(\vec{v}|\vec{x}) = \frac{1}{N} p(\vec{v}, \vec{x})$ and $N = \int d^3\vec{v} \, p(\vec{x}, \vec{v}) = p(\vec{x})$

comes from Bayes' theorem:

$$p(x_1, \cdots, x_m | x_{m+1}, \cdots, x_N) = \frac{p(x_1, \cdots, x_N)}{p(x_{m+1}, \cdots, x_N)}$$

- Expectation value of $F(x)$

$$\langle F(x) \rangle = \int d^N x \, p(x) F(x)$$

- Joint characteristic func: N-dimensional Fourier transform.

$$\tilde{p}(k) = \left\langle \exp\left(-i \sum_{j=1}^{N} k_j x_j\right) \right\rangle$$

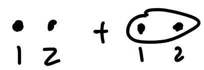
Joint moments

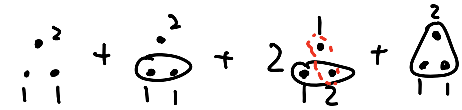
$$\langle x_1^{n_1} x_2^{n_2} \cdots x_N^{n_N} \rangle = \left[\frac{\partial}{\partial(-ik_1)}\right]^{n_1} \left[\frac{\partial}{\partial(-ik_2)}\right]^{n_2} \cdots \left[\frac{\partial}{\partial(-ik_N)}\right]^{n_N} \tilde{p}(k=0)$$

Joint cumulants

$$\langle x_1^{n_1} x_2^{n_2} \cdots x_N^{n_N} \rangle_c = \left[\frac{\partial}{\partial(-ik_1)}\right]^{n_1} \left[\frac{\partial}{\partial(-ik_2)}\right]^{n_2} \cdots \left[\frac{\partial}{\partial(-ik_N)}\right]^{n_N} \ln \tilde{p}(k=0)$$

Graphical relation

$\langle x_1 x_2 \rangle =$  $= \langle x_1 \rangle_c \langle x_2 \rangle_c + \langle x_1 x_2 \rangle_c$

$\langle x_1^2 x_2 \rangle =$  $= \langle x_1 \rangle_c^2 \langle x_2 \rangle_c + \langle x_1^2 \rangle_c \langle x_2 \rangle_c + 2 \langle x_1 x_2 \rangle_c \langle x_1 \rangle + \langle x_1^2 x_2 \rangle_c$

* connected correlation $\langle x_\alpha x_\beta \rangle_c = 0 \iff x_\alpha$ and x_β are independent

* covariance

- Joint Gaussian distribution:

$$p(x) = \frac{1}{\sqrt{(2\pi)^N \det(C)}} \exp\left[-\frac{1}{2}\sum_{mn}(C^{-1})_{mn}(x_m-\lambda_m)(x_n-\lambda_n)\right]$$

C is symmetric and positive-definite matrix

λ is the eigenvalues

$$\tilde{p}(k) = \exp\left[-ik_m\lambda_m - \frac{1}{2}C_{mn}k_mk_n\right]$$

with summation convention (Einstein Notation)

$\ln\tilde{p}(k) \rightarrow \langle x_m\rangle_c = \lambda_m \quad \langle x_m\rangle_c\langle x_n\rangle_c = C_{mn}$

with all higher cumulants $= 0$

Special case: $\{\lambda_n\} = 0$

all odd moments of the distribution is zero.

any even moment is obtained by summing over all ways of grouping the involved random variables into pairs

$$\langle x_a x_b x_c x_d\rangle = C_{ab}C_{cd} + C_{ac}C_{bd} + C_{ad}C_{bc}$$ (Wick's theorem)

2.5 Sum of random variables and the central limit theorem

$$X = \sum_{i=1}^{N} x_i$$

as $x = \sum x_i$, $x - x_1 - \cdots - x_{N-1} = x_N$

PDF: $$P_X(x) = \int d^N x \, p(x)\, \delta(x - \sum x_i) = \int \prod_{i=1}^{N-1} dx_i \, p(x_1, \cdots, x_{N-1}, x - x_1 - x_2 \cdots - x_{N-1})$$

Corresponding characteristic func: $$\tilde{P}_X(k) = \left\langle \exp\left(-ik \sum_{j=1}^{N} x_j\right) \right\rangle = \tilde{P}_X(k_1 = k_2 = \cdots = k_N = k)$$

Cumulants of sum: $$\ln \tilde{P}_X(k = k_1 = k_2 \cdots = k_N) = -ik \underbrace{\sum_{i_1=1}^{N} \langle x_{i_1} \rangle_c}_{\langle X \rangle_c} + \frac{(-ik)^2}{2} \underbrace{\sum_{i_1, i_2}^{N} \langle x_{i_1} x_{i_2} \rangle_c}_{\langle X^2 \rangle_c} + \cdots \cdots$$

$\Rightarrow$ from single random variable

Independent random variables $\quad p(x) = \prod p_i(x_i)$; $\tilde{p}(k) = \prod \tilde{p}_i(k)$

cross cumulants vanish and $\langle X^n \rangle_c = \sum_{i=1}^{N} \langle x_i^n \rangle_c$

⊕ N random variables are taken from the same dist. $p(x)$

$\langle X^n \rangle_c = N \langle x^n \rangle_c$ Hence, for large N, the average value of sum is proportional to N; $SD = \sqrt{\langle X^n \rangle_c} \propto \sqrt{N}$

The random variable $y = (X - N\langle x \rangle_c) / \sqrt{N}$ has zero mean

$\Rightarrow \langle y \rangle_c = \langle y \rangle = 0$

and cumulants that scales as $\langle y^n \rangle_c \propto N / (\sqrt{N})^n = N^{1-\frac{n}{2}}$ (Indep. r.v.)

$\langle y^n \rangle_c = N^{1-\frac{n}{2}} \cdot \langle x^n \rangle_c$

As $N \to \infty$, $\lim_{N\to\infty} N^{1-\frac{n}{2}} = \lim_{N\to\infty} e^{(1-\frac{n}{2})\ln N}$ involves w.r.t. $1 - \frac{n}{2} = 0$ $\quad (n = 2)$

$\Rightarrow \langle y^2 \rangle_c = \langle x^2 \rangle_c$ and $\langle y^n \rangle_c = 0 \ (n \neq 2)$

PDF for y converges to $$\lim_{N\to\infty} p\left(y = \frac{\sum_{i=1}^{N} x_i - N\langle x_i \rangle_c}{\sqrt{N}}\right) = \frac{1}{\sqrt{2\pi \langle x^2 \rangle_c}} \exp\left(-\frac{y^2}{2\langle x^2 \rangle_c}\right)$$

(Gaussian dist. is the only dist. with only first and second cumulants)

Central limit theorem : the coverage of PDF for the same of many random variables to a normal distribution (not necessary to be independent, as the condition $\sum_{i_1, \cdots i_m}^{N} \langle x_{i_1}, \cdots, x_{i_m} \rangle \ll O(N^{m/2})$)

[Assumes the cumulants of individual r.v. are finite]

The sum converges to Levy distribution if variables are taken from wide PDF (or the cumulants of individual r.v. are infinite)

The variance doesn't exist if individual PDF falls off slowly at large values

e.g. $p_i(x) = p_1(x) \propto \frac{1}{|x|^{1+\alpha}}$ $(0<\alpha\leq 2)$ (Guess with $|k|^\alpha$)

$$\ln \tilde{p}_x(k) = N \ln \tilde{p}_1(k) = N \ln\left[\sqrt{\frac{2}{\pi}} \cos\left(\frac{\pi\alpha}{2}\right) \Gamma(-\alpha)|k|^\alpha\right]$$ by mathematica

$$= N(-a|k|^\alpha + \text{higher order terms}) \boxed{\Rightarrow ?}$$ P47

let $y = X/N^{1/\alpha}$ to get rid of N dependence of leading term

$$\lim_{N\to\infty} \tilde{p}_y(k) = -a|k|^\alpha$$ (higher order terms with negative powers of N) $\boxed{\Rightarrow ?}$

E.X. $\alpha=1$ $\quad p_1(x) = \frac{a}{\pi(a^2+x^2)}$ (Cauchy distribution) $\Rightarrow$ HW

$0<\alpha\leq 2$ $\quad p_\alpha(x) = \frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n+1}\sin\left(\frac{n\pi}{2}a\right)\frac{\Gamma(1+n\alpha)\cdot a^n}{n!\cdot x^{1+n\alpha}}$

$\Rightarrow$ use for large rare events

the tail $p_\alpha(x\to+\infty) \sim x^{-1-\alpha}$

2.6 Rules for large numbers:

Deal with large N of microscopic degrees of freedom.

thermodynamic limit $N \to \infty$

(a) Intensive quantities, e.g. $T, f(P, \vec{B})$ are independent of N $\Rightarrow O(N^0)$

(b) Extensive quantities, e.g. $E, S, \vec{x}(V, \vec{M})$ are proportional to N $\Rightarrow O(N^1)$

(c) Exponential dependence, encounted in enumerating discrete micro-states; or computing available volumes in phase space $\Rightarrow O(\exp(N\phi))$

(d) other asymptotic dependence, e.g. Coulomb energy of N ions at fixed density scales as $Q^2/R \sim N^{5/3}$

Simplification:

(1) Summation of exponential quantities

$S = \sum_{i=1}^{\mathcal{N}} \mathcal{E}_i$ & $0 \leq \mathcal{E}_i \sim O(\exp(N\phi_i))$

As $0 \leq \mathcal{E}_i \leq \mathcal{E}_{max}$, and the number of $\mathcal{N}$ is proportional to some power of N, $\mathcal{E}_{max} \leq S \leq \mathcal{N}\mathcal{E}_{max}$

An intensive quantity can be constructed from $\frac{\ln S}{N}$

$$\frac{\ln \mathcal{E}_{max}}{N} \leq \frac{\ln S}{N} \leq \frac{\ln \mathcal{N} + \ln \mathcal{E}_{max}}{N}$$

For $\mathcal{N} \propto N^P$, $\frac{\ln \mathcal{N}}{N}$ vanishes in large N limit ($\lim_{N\to\infty} \frac{\ln \mathcal{N}}{N} = 0$)

$$\lim_{N\to\infty} \frac{\ln S}{N} = \frac{\ln \mathcal{E}_{max}}{N} = \phi_{max}$$

(2) Saddle point integration:

$\mathcal{I} = \int dx \exp(N\phi(x)) \xrightarrow{\text{approx.}}$ max. value of the integrand, obtained at a point x_{max} maximizing the exponent $\phi(x)$

$$\mathcal{I} = \int dx \exp\left\{N\left[\phi(x_{max}) - \frac{1}{2}|\phi''(x_{max})|(x - x_{max})^2 + \cdots\right]\right\} \quad (\text{expand around } x_{max})$$

$x \to x_{max}$

at the max., $\phi'(x_{max}) = 0$, $\phi''(x_{max}) < 0$

$$\mathcal{I} \approx e^{N\phi(x_{max})} \int_{-\infty}^{+\infty} dx \exp\left[-\frac{N}{2}|\phi''(x_{max})|(x - x_{max})^2\right] \approx \sqrt{\frac{2\pi}{N|\phi''(x_{max})|}}\, e^{N\phi(x_{max})}$$

* The integrand is negligibly small outside the neighborhood of x_{max}

Corrections:

① higher-order terms in the expansion of $\phi(x)$ around x_{max} (treat it perturbatively and lead to a series in powers of $1/N$ $\Leftarrow \left(\frac{1}{N}\right)^m$)

② Additional local maxima for the function. At x'_{max}, do similar Gaussian integrals.

$$\mathcal{I} = A e^{N\phi(x_{max})} + A' e^{N\phi(x'_{max})}$$
$$= A e^{N\phi(x_{max})}\left(1 + \frac{A'}{A} e^{-N(\phi(x_{max}) - \phi(x'_{max}))}\right) \overset{N\to\infty}{=} A e^{N\phi(x_{max})}$$

Therefore, all these correlations vanish: $\lim_{N\to\infty} \frac{\ln J}{N} = \lim_{N\to\infty} \left[\phi(x_{max}) - \frac{1}{2N}\ln\left(\frac{N|\phi''(x_{max})|}{2\pi}\right) + \mathcal{O}\left(\frac{1}{N^2}\right)\right] = \phi(x_{max})$

Saddle point method: use for general integrands and integration path in complex plane
(focus on extremum)

E.x.
Stirling's approx. for N!

$$\int_0^\infty dx\, e^{-\alpha x} = \frac{1}{\alpha} \xrightarrow[\text{for } N \text{ times}]{\text{differentiation } (\alpha)} \int_0^\infty dx\, x^N e^{-\alpha x} = \frac{N!}{\alpha^{N+1}} \quad (N \in \mathbb{N})$$

$$\Gamma(N+1) = N! = \int_0^\infty dx\, x^N e^{-x} \quad (\text{analytical one})$$

$\exp(N\phi(x))$ with $\phi(x) = \ln x - x/N$;

$$\phi(x_{max}) = \phi(N) = \ln N - 1 ; \; \phi''(x_{max}) = -\frac{1}{N^2}$$

$$N! \simeq \int_0^\infty dx \exp\left\{N\left[(\ln N - 1) - \frac{1}{2N^2}(x-N)^2\right]\right\} \simeq N^N e^{-N}\sqrt{2\pi N}$$

(* $\int_0^\infty \to \int_{-N}^\infty$ and $N \to +\infty$, so integral is evaluated from $-\infty$ to $+\infty$

Therefore, $\ln N! = N\ln N - N + \frac{1}{2}\ln(2\pi N) + \mathcal{O}\left(\frac{1}{N}\right)$

2.7 Information, entropy, and estimation

Information: A random variable with discrete outcome $S=\{x_i\}$, occuring with prob. $\{p(i)\}$

E.X. N independent outcomes; M prob. for each character in message.

info contents: $\boxed{N\log_2 M}$ bits

$N\to\infty$, "roughly" $\{N_i = Np_i\}$ occurrences of each symbol as $p(N_i)$ that is different from Np_i by more than $\mathcal{O}(\sqrt{N})$ becomes exponentially small in N. $(N\to\infty)$

The number of messages: number of ways of rearranging the $\{N_i\}$ occurrence of $\{x_i\}$

$$g = \frac{N!}{\prod_{i=1}^{M} N_i!} \ll \text{total number of messages } \boxed{M^N}$$

$$\ln g = \ln N! - \sum_{i=1}^{M} \ln N_i! = N\ln N - \sum_{i=1}^{M} N_i \ln N_i$$

$\downarrow N\to\infty$

$$= N\ln N - \sum_{i=1}^{M} Np_i \ln(Np_i)$$

$$= N\ln N - \left(\sum_{i=1}^{M} Np_i \ln p_i + \sum_{i=1}^{M} N_i \ln N\right)$$

$$= N\ln N - N\sum_{i=1}^{M} p_i \ln p_i - \cancel{\ln N \cdot \sum_{i=1}^{M} N_i} \rightarrow N\ln N$$

$$= -N\sum_{i=1}^{M} p_i \ln p_i$$

Result 1: $\log_2 g = \dfrac{\ln g}{\ln 2} = -N\sum_{i=1}^{M} p_i \dfrac{\ln p_i}{\ln 2} = -N\sum_{i=1}^{M} p_i \log_2 p_i$ bits of info.

Result 2: $1 = \left(\sum_i p_i\right)^N = \sum_{\{N_i\}} N! \prod_{i=1}^{M} \dfrac{p_i^{N_i}}{N_i!} \simeq g_{max} \prod_{i=1}^{M} p_i^{Np_i} \quad (N\to\infty)$

multinormial coeffi. — sum replaced by largest term

Shannon's theorem: minimum number of bits ensure that the percentage of errors in N trials vanishes in the $N\to\infty$ limit is $\log_2 g \ll \log_2 M^N = N\log_2 M$ (for any non-uniform distibution without any info. on relative probabilities)

*Difference per trial → info. content of the prob. distribution

$$I[\{p_i\}] = \log_2 M + \sum_{i=1}^{M} p_i \log_2 p_i$$

Entropy:

a measure of dispersity of the distri. (does not depends on the values of random variables $\{x_i\}$)

$$g = \frac{N!}{\prod_{i=1}^{M} N_i!} \Leftarrow \text{mixing of } M \text{ distinct components} \Leftarrow \text{entropy of mixing}$$

General definition: $S = -\sum_{i=1}^{M} p(i) \ln p(i) = -\langle \ln p(i) \rangle$

Minimum: $S_{min} = 0$ w.r.t. $p(i) = \delta_{i,j}$ (Delta func. distri. 0/1)

Maximum: $S_{max} = \ln M$ w.r.t. $p(i) = 1/M$ (uniform distri.)

one-to-one mapping to $f_i = F(x_i)$ $\Rightarrow$ entropy unchange

many-to-one mapping $\Rightarrow$ more order $\Rightarrow$ entropy decreases

E.x.

$$\Delta S(x_1, x_2 \to f) = -(p_1+p_2)\ln(p_1+p_2) - \left[-(p_1 \ln p_1 + p_2 \ln p_2)\right] = p_1 \ln \frac{p_1}{p_1+p_2} + p_2 \ln \frac{p_2}{p_1+p_2} < 0$$

Estimation: use S to quantify subjective estimates of probabilities.

In absence of any info, the best unbiased estimate :

all M outcomes are equally likely (maximum entropy)

With additional info, the unbiased estimate :

maximizing entropy subject to the constraints imposed by info.

E.x. know $\langle F(x) \rangle = f$

$$S(\alpha, \beta, \{p_i\}) = -\sum_i p(i) \ln p(i) - \underbrace{\alpha \Big(\overbrace{\sum_i p(i) - 1}^{\text{normalization}}\Big) - \beta \Big(\overbrace{\sum_i p(i) F(x_i) - f}^{\text{mean average}}\Big)}_{\text{Lagrange multipliers}}$$

$\Rightarrow$ optimization $p_i \propto \exp(-\beta F(x_i))$ (Here β is fixed by constraints.)

Proof: $\frac{\partial S}{\partial p_i} = -(\ln p(i) + 1) - \alpha - \beta F(x_i) = 0 \Rightarrow p(i) = e^{-\alpha-1} e^{-\beta F(x_i)}$

* Constraints: $n = 2k$ moments (or n cumulants), $\ln p(i) = P(p(i))$

Analytical method: Lambert W function

Numerical method: bisection / Newton-Raphson method

$\Rightarrow$ the exponential of an n-th order polynomial.

Continous random variable

$S_x = \{-\infty < x < +\infty\}$

Entropy calculation:

$$S = -\int dx\, p(x) \ln p(x) = -\langle \ln p(x) \rangle$$

Involves some problems under this defintion.

P1: S is not invariant under a one-to-one map

After change $f = F(x)$, the entropy is changed by $\langle |F'(x)| \rangle$

↑ sol.

Canonically cojugate pairs offer a suitable choice of coordinates in classical statistical mechanics, since Jacobian of a canonical transformation is unity. Ambiguities are removed if continuous variables is discretized.

↳ similar to quantum: a discrete ladder of states

volume of discretization of phase space is set by $\hbar$.