$Z = \frac{X_1 + X_2}{\lambda}$; λ is the scale, free for now

$$\boxed{\hat{p}_Z(k) = \langle e^{ikZ} \rangle = \langle e^{ikX/\lambda} \rangle^2 = \hat{p}_X^2\left(\frac{k}{\lambda}\right)}$$

We assume p_X narrowly distributed $\Leftrightarrow$ $\hat{p}_X(k)$ is analytic around $k = 0$

$$\mathcal{K}_X(k) = \log \hat{p}_X(k) = ik\langle X\rangle - \frac{k^2}{2} V(X) + \frac{(ik)^3}{3!} c_3 + \cdots$$

$$\mathcal{K}_Z(k) = \log \hat{p}_Z(k) = 2\, \mathcal{K}_X\left(\frac{k}{\lambda}\right)$$

Hence the functional equation:

$\mathcal{K}_Z(k) = 2\, \mathcal{K}_X(k/\lambda)$. When do we have $\mathcal{K}_X = \mathcal{K}_Z = \mathcal{K}^*$?

$$\mathcal{K}^*(k) = 2\, \mathcal{K}^*(k/\lambda)$$

$$\Rightarrow ik\langle Z\rangle - \frac{k^2}{2} V(Z) + \frac{(ik)^3}{3!} c_3 = 2\left[ik \underbrace{\left\langle \frac{Z}{\lambda}\right\rangle}_{\frac{k}{\lambda}\langle Z\rangle} - \frac{1}{2}\frac{k^2}{\lambda^2} V(Z) + \cdots\right]$$

$$\Rightarrow \begin{cases} \langle Z\rangle = \frac{2}{\lambda}\langle Z\rangle \\ V(Z) = \frac{2}{\lambda^2} V(Z) \\ c_k = \frac{2}{\lambda^k} c_k \quad \text{for cumulants of order } k \end{cases}$$

$\longrightarrow$ Discussion $\langle Z\rangle = (2/\lambda)\langle Z\rangle$

⊛ if $\langle Z\rangle \neq 0$, then $\boxed{\lambda = 2}$ and $V(Z) = \frac{1}{2} V(Z) \Rightarrow V(Z) = 0$

Also $c_k = 0$, $k \geq 3$

Thus $\mathcal{K}^*(k) = -ik\langle Z\rangle + 0$, which is the charact. function of $\delta(Z - \langle Z\rangle)$. This is the law of large numbers, stating that $\frac{1}{n}\sum_{i=1}^{n} X_i$ where $\langle X\rangle = m$, converges to m for $n \to \infty$.

The choice $\lambda = 2$ indeed ensures that we compute the (empirical) mean.

⊛ if $\langle Z\rangle = 0$, then $V(Z) = (2/\lambda^2) V(Z) \Rightarrow \boxed{\lambda = \sqrt{2}}$ to have non trivial result

Then $c_k = 2^{2-k/2} c_k \Rightarrow c_k = 0$ for all cumulant but the scd

$\hookrightarrow$ Gaussian behaviour ; central limit theorem

Indeed, by considering successively $\frac{1}{\sqrt{2}}(X_1 + X_2)$; $\frac{1}{2}(X_1 + X_2 + X_3 + X_4)$ $\frac{1}{\sqrt{8}}(X_1 + X_2 + \cdots + X_8)$, we deal with $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$, which is the proper combination to get CLT result emerge

$\rightarrow$ the gaussian is a fixed point of "renormalization" procedure

Is it a stable fixed point ?

→ Stability of the fixed point.

We introduce the operator

$$G_\lambda[\mathcal{K}](k) \equiv 2\mathcal{K}(k/\lambda)$$

operator G_λ, acting on a function $\mathcal{K}$ → gives a function of k.

The eigenfunctions are k^p, with eigenvalue $(2/\lambda)^p$

(*) $\lambda = 2$, eigenvalues are $2^{1-p} < 1$ for $p \geq 2$

So if we take $\mathcal{K}(k) = ik\langle z\rangle + \delta\mathcal{K}(k)$ and iterate, the terms of $G(k^p)$, $p \geq 2$ of $\delta\mathcal{K} \to 0$ (irrelevant perturbation).

For $p = 1$, eigenvalue is 1 : marginal perturbation in RG langage ;
the term in ik from $\delta\mathcal{K}$ shifts the mean.

↳ origin of universality of law of large numbers ... provided all moments exist (this latter condition is unnecessary, but we used it in the argument).

(*) $\lambda = \sqrt{2}$, with $\langle z\rangle = 0$: $\mathcal{K}^*(k) = -k^2 V(z)/2$

Take $\mathcal{K} = \mathcal{K}^* + \delta\mathcal{K}$, eigenvalues of $G_{\sqrt{2}}$ are $2^{1-p/2}$

$p=1$ eigenvalue $\sqrt{2} > 1$: unstable direction, relevant perturbation
eigenfunction is k → term $k\langle z\rangle$. We cannot afford $\langle z\rangle \neq 0$ in the perturbation

$p=2$ eigenvalue 1 : marginal, corresponds to rescaling of the variable
↳ innocuous.

$p>2$ eigenvalue $2^{1-p/2} < 1$ → irrelevant

↳ origin of universality of CLT (provided all moments exist).

→ For $\lambda \neq 2$ and $\neq \sqrt{2}$, other fixed points can be found, of the form $\mathcal{K}^*(k) \propto |k|^\mu$ with $0 < \mu < 2$ (divergent second moment) → Lévy-stable laws

$$|k|^\mu = 2\left(\frac{k}{\lambda}\right)^\mu \Rightarrow \lambda^\mu = 2 \quad ; \quad \lambda = 2^{1/\mu}$$

The groups considered are of type $\frac{1}{2^{1/\mu}}(X_1+X_2)$; $\frac{1}{4^{1/\mu}}(X_1+X_2+X_3+X_4)$ ...
ie $\frac{1}{N^{1/\mu}}(X_1 + \dots X_N)$ as already met.

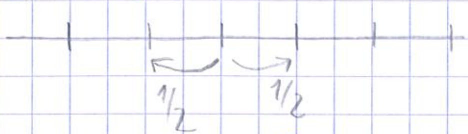
II LARGE DEVIATIONS

Goal : introduce the main tools (Sanov thm, Gärtner-Ellis) ; shed new light on and reinterpret thermodynamic potentials (free energy) ; discuss connection with mean-field approaches

1°) Introduction

What happens outside the central region where CLT applies (of extent $N^{2/3}$ or $N^{3/4}$ depending on symmetry), in the tails of $p(S_n)$ where $S_n = \sum_{i=1}^{n} X_i$?

Start by an explicit example : random walk, jumps ± 1 proba $1/2$, on lattice $\mathbb{Z}$.



$$x_n = x_{n-1} + \eta_n$$

η_n → jump step n : $+1$ proba $1/2$, -1 " $1/2$

x_n ↳ position after n steps

$$\Pr[x_n = m] \; ?$$

i.e $p(\eta) = \frac{1}{2}\left[\delta(\eta-1) + \delta(\eta+1)\right]$; η_i IID

For n large, CLT applies , $\langle x_n \rangle = 0 \equiv x_0$; $V(x_n) = n$

$$\Pr[x_n = m] \sim \frac{\text{Beware prefactor}^{(*)}}{\sqrt{2\pi n}} \exp\left(-\frac{m^2}{2n}\right)$$

Clearly wrong in the tail

$$\Pr[x_n = n] = \frac{1}{2^n} \text{ , exact while CLT gives } \doteq e^{-\frac{n}{2}} \gg 2^{-n} = e^{-n\ln 2}$$

Large deviation theory suitable to describe both small and large fluctuations, thus including the CLT, and going beyond. Large deviations are important in a number of context : risk management, climate, ...

Exact calculation possible here ; $n_+ \equiv$ number of steps to right ; $n_- \equiv$ to the left

$$\begin{aligned} n &= n_+ + n_- \\ m &= n_+ - n_- \end{aligned} \quad \Rightarrow \quad \begin{aligned} n_+ &= \frac{1}{2}(n+m) \\ n_- &= \frac{1}{2}(n-m) \end{aligned}$$

$$\Pr[x_n = m] = \frac{1}{2^n}\binom{n}{n_+} = \frac{1}{2^n}\binom{n}{\frac{n+m}{2}}$$

We are interested in the fluctuations, including those of $O(n)$: → $\boxed{\zeta \equiv \frac{m}{n} = \frac{x_n}{n}}$

Use Stirling : $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n = \sqrt{2\pi}\, e^{n\log n - n + \frac{1}{2}\log n}$

$$\Pr[x_n = m = zn] \approx \frac{1}{\sqrt{2\pi}} \frac{e^{n\log n - \cancel{n} + \frac{1}{2}\log n}}{\left(e^{n_+\log n_+ - \cancel{n_+} + \frac{1}{2}\log n_+}\right)\left(e^{n_-\log n_- - \cancel{n_-} + \frac{1}{2}\log n_-}\right)} \frac{1}{2^n}$$

$n_+ = \frac{n+m}{2} = n\frac{1+z}{2}$

$n_- = \frac{n-m}{2} = n\frac{1-z}{2}$

$$\sim \frac{1}{\sqrt{2\pi}} \exp\left\{(n_+ + n_-)\log n - n_+\log n_+ - n_-\log n_- + \frac{1}{2}\log\frac{n}{n_+ n_-} - n\ln 2\right\}$$

$$\sim \frac{1}{\sqrt{2\pi}} \exp\left\{-n\left[\frac{1+z}{2}\log\frac{1+z}{2} + \frac{1-z}{2}\log\frac{1-z}{2} + \ln 2\right]\right\}\sqrt{\frac{4}{n(1-z^2)}}$$

$$\Pr[x_n = nz] \doteq e^{-n\phi(z)}$$

$$\phi(z) = \frac{1+z}{2}\log(1+z) + \frac{1-z}{2}\log(1-z)$$ IS THE RATE FUNCTION

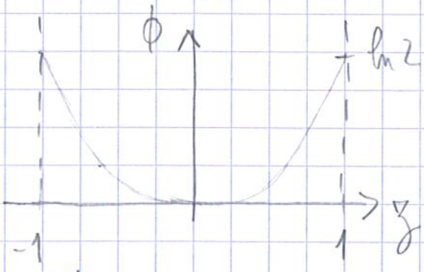
$\phi(z)$ characterizes all possible fluctuations, for n large.

When $\Pr[x_n] \doteq e^{-n\phi(x_n/n)}$ holds, ie ϕ exists, the proba is said to satisfy a large deviation principle

$\phi(z=1) = \ln 2 \Rightarrow \Pr[x_n = n] \doteq e^{-n\ln 2} = 2^{-n}$, true (even with $=$, not only $\doteq$)

$\phi(z) \underset{0}{\sim} \frac{1}{2}\left[(1+z)\left(z - \frac{z^2}{2}\right) + (1-z)\left(-z - \frac{z^2}{2}\right)\right] \sim \frac{1}{2}\left(z + \frac{z^2}{2} - z + \frac{z^2}{2}\right)$

$\underset{0}{\sim} z^2/2$



$$\phi(z) \xrightarrow{z \to 0} z^2/2$$
$$\phi(z) \xrightarrow{z \to 1} \ln 2$$

The quadratic behavior for $z \to 0$ is what is needed to recover CLT:

$$\Pr[x_n = m] \doteq e^{-n(m/n)^2/2} = \exp\left(-\frac{m^2}{2n}\right)$$

2°) Revisiting the validity of the CLT: where do "tails" start?

In our above example, $\phi(z) = \phi(-z)$ by symmetry ie $\phi(z) \underset{0}{\sim} z^2/2 + (\cdot\cdot)z^4 + \cdots$

Hence the central CLT region is such that

$$n z^4 \ll 1 \Leftrightarrow n\left(\frac{x_n}{n}\right)^4 \ll 1 \Leftrightarrow \boxed{|x_n| \ll n^{3/4}}$$

This argument is generic, as soon as a large deviation principle holds. If the pdf of jumps is non symmetric $p(\eta) \neq p(-\eta)$, then $\phi(z) \neq \phi(-z)$ and we have $\phi(z) \underset{0}{\sim} (\cdot\cdot)z^2 + (\cdot\cdot)z^3$

The CLT region is then for $n|z|^3 \ll 1 \Leftrightarrow n\left(\frac{|x_n|}{n}\right)^3 \ll 1$; $\boxed{|x_n| \ll n^{2/3}}$

3c) An entropy appears

Wait... $\phi(z)$ looks like the entropy for the mean-field Ising model solution.

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} S_i S_j \quad ; \quad z \equiv \langle S_i \rangle \text{ magnetization per spin (homogeneous)}$$

$$F_{mf}(z) = -J \frac{nq}{2} z^2 + kTn \left[\frac{1+z}{2} \log \frac{1+z}{2} + \frac{1-z}{2} \log \frac{1-z}{2} \right]$$

n: # spins ; q: number nearest neighbors, eg 4 on square lattice

$$= \underbrace{\quad}_{U} - T S n \quad ; \quad S \equiv \text{entropy per spin}$$

Hence here $\boxed{-S(z)/k = \phi(z) - \log 2}$. Is this a coincidence ?

Not quite, although in general, ϕ is not directly to an entropy, but to a relative entropy (langage of information theory). Here, S arises from pure combinatorics :

$$\Pr[x_n = m] = \frac{1}{2^n} \mathcal{N}\left(n_+ = \frac{n+m}{2}\right)$$

↳ total # of walks from 0 to $x_n = \binom{n}{n_+}$ here

Remember Boltzmann's grave : " $S = k \log W$ " $\rightarrow$ $nS = k \log \mathcal{N}$ ie $\mathcal{N} = e^{nS/k}$

$$\Rightarrow e^{-n\phi(z)} = \frac{1}{2^n} e^{nS/k} \quad \Rightarrow \quad S/k = -\phi(z) + \log 2$$

↳ this is the good old entropy of statistical physics ! It takes here a noticeable form

$$S/k = -\frac{1+z}{2} \log \frac{1+z}{2} - \frac{1-z}{2} \log \frac{1-z}{2} = -p_+ \log p_+ - p_- \log p_-$$

↳ Recover Shannon entropy of information theory $\quad p_\pm = \frac{1}{2}(1 \pm z)$, proba of + and − spin resp.

$$S = -k \sum_i p_i \log p_i$$

for an ensemble of events where $\sum_i p_i = 1$.

S measures : uncertainty ; surprise ; missing information

We also recover that an entropy is a rate function. Take syst with short range interact's, denote $\mathcal{N}(E, \delta E)$ the # of microstates of energy between E and $E + \delta E$ (count eigenstates, or integrate volume in phase space) : with N molecules

$$\mathcal{N}(E, \delta E) \doteq \exp\left[N s \left(\frac{E}{N e_0}\right) / k \right] \quad e_0 \text{ some micro energy}$$

Translates extensivity. Related to free energy by a Legendre transform

$$Z(\beta) = \sum_E \mathcal{N}(E) e^{-\beta E}$$

$$= \sum_E e^{-\beta E + \log \mathcal{N}} \qquad E = Ne$$

$$= \sum_E e^{-\beta N [e - T k s(e)]}$$

$$\doteq \exp\left[-\beta N \min_e \left(\boxed{e - T s(e)} \right) \right] \rightarrow f(T)$$

4°) Sanov's theorem

We need one more tool of information theory to express a large dev principle, and understand the emergence of an entropy / free energy : introduce a "distance" between 2 proba distrib ; the Kullback-Leibler divergence.

$$D(q\|p) = \sum_{x\in\mathcal{X}} q(x) \log \frac{q(x)}{p(x)}$$; q and p are 2 proba distribution on the same ensemble $\mathcal{X}$ ("alphabet")

$$D(q\|p) \geqslant 0 \quad \text{since} \quad -D(q\|p) = \sum q(x) \log \frac{p(x)}{q(x)}$$
$$\leqslant \sum_x q(x) \left(\frac{p(x)}{q(x)} - 1 \right) = 0 \quad \text{since } \ln x \leqslant x-1$$

$D(q\|p)$ is called the relative entropy, a key tool in statistics, computer science. It is not a real distance, since not symmetric : $D(p\|q) \neq D(q\|p)$; and does not satisfy the triangle inequality.

Take sequence $\vec{s} = (x_1, \dots x_n)$ of IID from $p(x)$ ~~(for n large)~~
discrete distribution : the possible values of each x_i belong to some alphabet $\mathcal{X}$
The type of the sequence $\vec{s}$ is the frequency of occurence of the symbols, for all symbols.
Eg for a symbol x from the alphabet

$$q_s(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i = x} = \frac{\text{number of } x \text{ in sequence}}{n = \text{number of elements in sequence}}$$

$\delta_{x_i;x}$, Kronecker

↳ this is the empirical probability, from measured frequencies

NB $\displaystyle\sum_{x\in\mathcal{X}} q_s(x) = 1$ and $\langle q_s(x) \rangle = p(x)$

Sanov thm gives the probability that the type of $\vec{s}$, ie the empirical probability distribution, differs from $p(x)$, for n large.

Let K be an ensemble of probability distributions over $\mathcal{X}$, with some chosen properties

$$\Pr[q \in K] \doteq \exp\left[-n\, D(q^*\|p)\right] \; ; \quad q^* = \operatorname*{argmin}_{q\in K} D(q\|p)$$

We thus have to find in K, the distribution that is closest to p, in terms of relative entropy. K can be "anything". This tells us that the typical sequence have type $p(x)$, while the others are exponentially rare.

↳ Sanov yields a fine and powerful statement : it gives the probability of a (empirical) probability.

Proof : x takes values $x_\alpha, x_\beta, \dots$ in $\mathcal{X}$ with proba $p_\alpha, p_\beta, \dots$

What is the proba to observe a given type q for the sequence $x_1, \dots, x_n$

(do not confuse n with the size of the alphabet → sample size)

We take $n \ggg |\mathcal{X}|$ = size of " " .

x_α appears n_α times ... then $q_\alpha = \frac{n_\alpha}{n}$; $q_\beta = \frac{n_\beta}{n}$; ...

$$\Pr(q) = \frac{n!}{n_\alpha! \, n_\beta! \dots} p_\alpha^{n_\alpha} p_\beta^{n_\beta} \dots \quad ; \quad n_\alpha + n_\beta + \dots + n_{|\mathcal{X}|} = n$$

Multinomial

$$\log \Pr(q) \simeq n \log n - \cancel{n} - n_\alpha \log n_\alpha + \cancel{n_\alpha} - n_\beta \log n_\beta + \cancel{n_\beta} - \dots + n_\alpha \log p_\alpha + n_\beta \log p_\beta$$

$$\frac{1}{n} \log \Pr(q) = -q_\alpha \log q_\alpha - q_\beta \log q_\beta \dots + q_\alpha \log p_\alpha + q_\beta \log p_\beta + \dots$$

$$= -q_\alpha \log \frac{q_\alpha}{p_\alpha} - q_\beta \log \frac{q_\beta}{p_\beta} - \dots$$

$$= -D[q \| p]$$

and we then have to minimize over all eligible q in K. Because of the fast decay with n, only the q with smallest $D(q\|p)$ matters → noted q^*.

All but the terms corresponding to the minimum are exponentially suppressed.

Sanov encodes detailed info on the large deviations of the sequence, and in particular, on its empirical mean $\overline{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$, but also on quantities like $\overline{f} \equiv \frac{1}{n} \sum_{i=1}^n f(x_i)$

$f(x)$ arbitrary function

Indeed, this is a corollary of Sanov, taking

$$K = \left\{ q, \text{ proba on } \mathcal{X} \text{ s.t } \sum_{x \in \mathcal{X}} q(x) f(x) \in A, \text{ some interval} \right\}$$

$$\Pr[\overline{f} \in A] \doteq \exp(-n \, D(q^* \| p)) \quad ; \quad q^* = \underset{q}{\operatorname{argmin}} \, D[q\|p] \ \text{ s.t } \sum_x q(x) f(x) \in A$$

Ex 1 : back to r.w in 1d, $x_n = \sum_{i=1}^n \eta_i$; $\eta_i = \pm 1$ proba $1/2$, IID

$$z \equiv \frac{x_n}{n} = \overline{\eta}_n \quad \text{, ie here } f(\eta) = \eta$$

$$\Pr[x_n = nz] = \Pr[\overline{\eta} = z] \doteq \exp[-n \, \phi(z)]$$

$$\phi(z) = D[q^* \| p] \ , \ q^* = \underset{q}{\operatorname{argmin}} \, D[q\|p] \ \text{ where } \begin{cases} q_1 + q_{-1} = 1 \\ q_1 + - q_{-1} = z \end{cases}$$

ie A is here a point, z

Here, there is nothing to minimize over, since the constraint fully determines the probability $q(z)$: $q(1)$ and $q(-1)$

$$D[q\|p] = q_1 \log \frac{q_1}{1/2} + q_{-1} \log \frac{q_{-1}}{1/2} = \log 2 + \frac{1+z}{2} \log \frac{1+z}{2} + \frac{1-z}{2} \log \frac{1-z}{2}$$

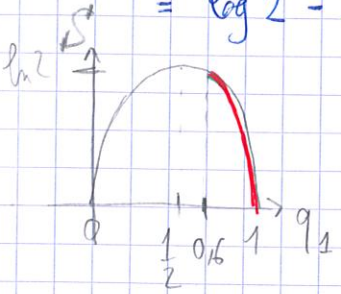
$$\phi(z) = \frac{1}{2}\left[(1+z)\log(1+z) + (1-z)\log(1-z)\right]$$ as already found

Ex2 Take a fair coin : what is proba of observing more than 600 heads in 1000 toss?

$X_i = 1$ for H ; $X_i = 0$ for T ; $S_n = \sum_{i=1}^{1000} X_i$; $\Pr[S_n \geqslant 600]$?

Minimize $q_0 \log \frac{q_0}{1/2} + q_1 \log \frac{q_1}{1/2}$ with $q_1 \geqslant 0{,}6$

$$= \log 2 - S(q_0, q_1)$$



Maximize $S(q_1)$ for $q_1 \geqslant 0{,}6 \Rightarrow q_1 = 0{,}6$

and let us take $\log_2$, log in base 2

$$D[q\|p] = 1 + 0{,}6 \log_2 0{,}6 + 0{,}4 \log_2 0{,}4$$
$$\simeq 0{,}029$$

$$\Rightarrow \Pr(S_n \geqslant 600] \doteq 2^{-29}$$
$$= 10^{-0{,}3 \times 29} \simeq 10^{-9}$$

Note that $\Pr(S_n = 600)$ would be given by the same result, at our level of approximation

↳ due to asymptotic nature of our tools, valid for $n \to \infty$. Is $n = 1000$ large enough? Yes for having, say, $\Pr(S_n \geqslant 700) \ll \Pr(S_n \geqslant 600)$

since $\Pr(S_n \geqslant 700) \doteq 2^{-119} \simeq 10^{-35}$ but not for $\Pr(S_n \geqslant 610) \doteq 2^{-35} \simeq 10^{-10}$

This is not so small compared to $\Pr(S_n \geqslant 600)$ found above (10^{-9})

We come back below to the pb of giving a rigorous upper bound to $\Pr(S_n \geqslant \text{sthg})$.

Ex3 We wish to find $\Pr\left[\frac{1}{n}\sum_{i=1}^{n} g_j(x_i) \geqslant \alpha_j \ , j = 1, 2, \dots k\right]$, ie we have k constraints

$$\sum_{x \in \mathcal{X}} g_j(x)\, q(x) \geqslant \alpha_j \quad , \quad j = 1, \dots k$$

For a given $p(x)$, minimize $D[q\|p]$ under the k constraints $\Rightarrow$ k Lagrange multipliers

Minimize $\sum_x q(x) \log \frac{q(x)}{p(x)} + \sum_{j=1}^{k} \sum_x g_j(x) q(x) \lambda_j + \mu \sum q(x)$

↳ for normalization

$$\frac{\partial}{\partial q(x)} = 0 = \log \frac{q(x)}{p(x)} + 1 + \sum_{j=1}^{k} g_j(x) \lambda_j + \mu$$

$$\Rightarrow q(x) = \frac{1}{Z} p(x) \exp\left(-\sum_{j=1}^{k} \lambda_j g_j(x)\right).$$

The constraints determine the variation range for the λ_j ; then feed into $D[q\|p]$

q^*, the optimal

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} g_j(x_i) \geqslant \alpha_j\right] \doteq \exp\left[-n D(q^*\|p)\right]$$

Ex 4 Sanov can be used to count the # of sequences with a given type (such as a given empirical average). Idea: take $p(x)$ uniform; all sequences are equally probable. A sequence is $(x_1, \dots x_n)$ where each x_i can take $|\mathcal{X}|$ values. There are thus $|\mathcal{X}|^n$ sequences

$$\Pr[q \in K] = \frac{\#(\text{sequences of type in } K)}{|\mathcal{X}|^n} \doteq e^{-n D[q^* \| p_{\text{unif}}]}$$

$$\#(\text{seq type } K) = |\mathcal{X}|^n e^{-D[q^*\|p]} = e^{n \log|\mathcal{X}| - n D[q^*\|p]} = e^{n S(q \in K)}$$

we recover the entropy

5°) The case of correlated variables (Gärtner-Ellis theorem): simple-minded exposition.

Take a set of $\boxed{\text{correlated r.v}}$ $\vec{x} = (x_1, \dots, x_n)$, real, each x_i in $\mathcal{X}$,

joint proba $P_n(\vec{x}) = P_n(x_1, \dots, x_n)$. Take arbitrary function $f: \mathbb{R} \to \mathbb{R}$ and it empirical av

$$\bar{f}_n(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

We assume that a large deviation principle holds, i.e.

$$P_n(\bar{f}) \doteq \exp(-n\, \phi(\bar{f}))$$

Define the cumulant generating function

$$K_n(t) = \frac{1}{n} \log \left\langle e^{t n \bar{f}_n} \right\rangle$$

$$= \frac{1}{n} \log \int d\bar{f}\; e^{t n \bar{f}}\, e^{-n \phi(\bar{f})}$$

$$\xrightarrow[n\to\infty]{} \sup_{\bar{f}\in\mathbb{R}} \left[t\bar{f} - \phi(\bar{f}) \right] \equiv K(t)$$

and the large n limit can be evaluated with saddle-point method

↳ this is a LEGENDRE TRANSFORM up to sign.

Such a definition of the LT (that of the mathematicians) does not require ϕ to be differentiable

Here, $K_n(t)$ is convex-up, and it is natural to take its Legendre Transform, for "inversion"

$$\frac{d^2 K_n}{dt^2} = \frac{d}{dt}\left(\langle n\bar{f} e^{nt\bar{f}} \rangle / \langle e^{nt\bar{f}} \rangle \right) = \frac{1}{\langle e^{nt\bar{f}} \rangle} \langle n^2 \bar{f}^2 e^{nt\bar{f}} \rangle - \frac{\langle n\bar{f} e^{nt\bar{f}} \rangle \langle n\bar{f} e^{nt\bar{f}} \rangle}{\langle e^{nt\bar{f}} \rangle^2}$$

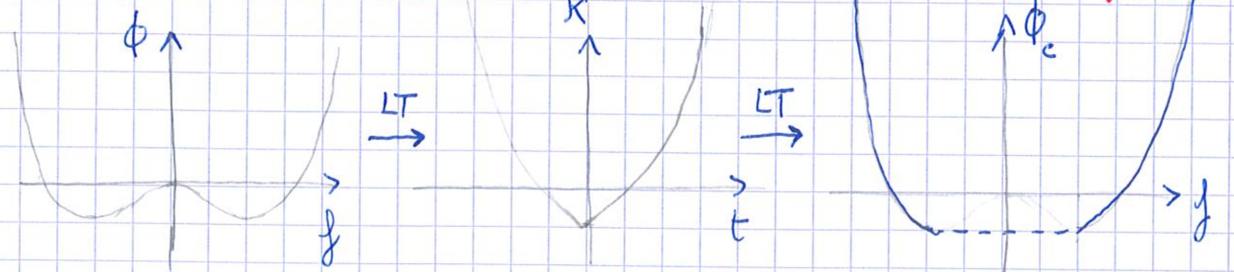
≥ 0 from Cauchy-Schwarz

t and $\bar{f}$ are conjugate variables

$$\phi_c(\bar{f}) = \sup_{t\in\mathbb{R}} \left[t\bar{f} - K(t) \right]$$

; it is the convex envelope of $\phi(\bar{f})$

Indeed, take



This is the G-E theorem, apart from subtleties on convexity: the LT of $\kappa(t)$ is the rate function.

5°) a) Application to independent variables

$z \Leftrightarrow \bar{\eta}$ above

Go back to our simple ± 1 random walk ; $x_n = \sum_{i=1}^{n} \eta_i$; $x_n = nz$

Gärtner-Ellis : $\phi(z) \xrightarrow{LT} \kappa_n(t) = \frac{1}{n} \log \langle e^{ntz} \rangle = \frac{1}{n} \log \langle e^{t\sum_i \eta_i} \rangle$

$$= \frac{1}{n} \log \langle e^{t\eta} \rangle^n$$

$$= \log \langle e^{t\eta} \rangle \quad \text{, does not depend on } n.$$

$$\kappa(t) = \log\left(\frac{1}{2}e^{t} + \frac{1}{2}e^{-t}\right) = \log(\cosh t)$$

Minimize $tz - \kappa(t)$ wrt $t \Rightarrow z - \kappa'(t) = 0$

$$z = \text{th}(t) \quad ; \quad t^* = \text{argth}\, z = \frac{1}{2} \log \frac{1+z}{1-z}$$

$$\phi(z) = z t^* - \log \cosh t^*$$

$$= \frac{z}{2} \log \frac{1+z}{1-z} - \frac{1}{2} \log \frac{1}{1-\text{th}^2 t^*}$$

$\log \cosh = \frac{1}{2} \log \cosh^2$
$= \frac{1}{2} \log \frac{\cosh^2}{\cosh^2 - \sinh^2}$

$$= \frac{z}{2} \log(1+z) - \frac{z}{2} \log(1-z) + \frac{1}{2} \log(1-z^2)$$

$$= \frac{1}{2}\left[(1+z)\log(1+z) + (1-z)\log(1-z)\right]$$

We recover a known result, derived by 2 other means: explicit calc & Sanov theorem. Gärtner-Ellis provides a 3rd route.

5°) b) Coupled variables

Take Ising 1D, N spins ± 1 ; $P(\sigma_1, \dots, \sigma_N) = \frac{1}{Z} e^{-\beta \mathcal{H}(\sigma_1, \dots, \sigma_N)}$

$$\mathcal{H}(\sigma_1, \dots, \sigma_N) = -J \sum_{i=1}^{N-1} \sigma_i \sigma_{i+1} \quad , \quad B = 0 \quad , \text{ periodic boundary conditions}$$

What are the large deviation properties of the magnetization?

$$m(\sigma_1, \dots, \sigma_N) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \quad ; \quad \Pr[Nm] = \frac{1}{Z} \sum_{\mathcal{C}} e^{-\beta \mathcal{H}} \delta_{mN, \sum_i \sigma_i}$$

Kronecker

$$\frac{1}{N} \log \langle e^{tNm} \rangle = \kappa_N(t) = \frac{1}{N} \log \sum_{\mathcal{C} \equiv (\sigma_1, \sigma_2, \dots \sigma_N)} \sum_m e^{Ntm} \frac{1}{Z} e^{-\beta \mathcal{H}} \delta_{mN, \sum \sigma_i}$$

$$= \frac{1}{N} \log \frac{1}{Z} \sum_{\mathcal{C}} e^{-\beta \mathcal{H}} e^{t \sum_i \sigma_i}$$

$$= \frac{1}{N}\left[-\log Z(\beta, B=0) + \log Z(\beta, B = t/\beta)\right]$$

$$\xrightarrow{N \to \infty} \beta f(\beta, B=0) - \beta f(\beta, B = t/\beta)$$

Hence the connexion between the free energy, and the cumulant generating function of magnetizat°

$f(\beta, B)$ can be computed with the transfer matrix technique : (27)

$$\mathcal{K}(t) = \log\left[\frac{\cosh t + \sqrt{\sinh^2 t + e^{-4K}}}{1+e^{-2K}}\right] \qquad K = \beta J$$

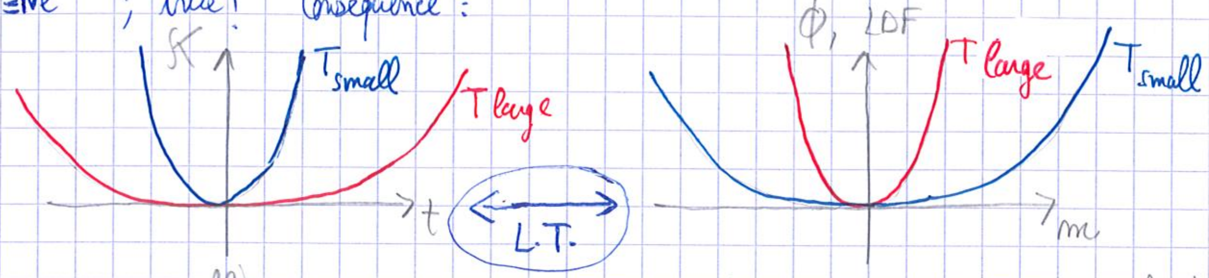
$$\underset{t=0}{\sim} \log\left(\frac{1}{1+e^{-2K}}\left(1+t^2/2 + e^{-2K}\sqrt{1+e^{+4K}\sinh^2 t}\right)\right)$$

$$\underset{t=0}{\sim} \log\left[\frac{1}{1+e^{-2K}}\left(1+t^2/2 + e^{-2K}\left\{1+t^2 e^{4K}/2\right\}\right)\right]$$

$$\sim \log\left[1+\frac{t^2}{2}\,\frac{1+e^{2K}}{1+e^{-2K}}\right] \rightarrow \frac{e^{2K}\left(e^{-2K}+1\right)}{\left(1+e^{-2K}\right)} = e^{2K}$$

$$\underset{t\to 0}{\sim} \frac{t^2}{2}\, e^{2K} + \mathcal{O}(t^4)$$

By the fluctuation-response connexion $\langle m^2\rangle - \langle m\rangle^2 = \chi k T$, this tells us that $\chi k T = N e^{2K}$; true! Consequence :



small / large curvature $\longleftrightarrow$ L.T. large / small curvature $\qquad$ g^{al} p^{ty} of L.T.

Thus, at small T, the large dev function is broader $\rightarrow$ larger fluctuations !
This may seem counterintuitive, but is already present in the susceptibility $\chi \propto T e^{+2J/T}$
This is because the correlation length is larger at small T : larger "coherent" clusters

Rk 1 More on the fluctuation-response connexion

$$\langle m^2\rangle - \langle m\rangle^2 = \frac{d^2}{dt^2}\mathcal{K}(t)\Big|_{t=0} = -\beta\,\frac{\partial^2 f}{\partial t^2}(\beta, B = t/\beta)$$

and $\langle m\rangle = -\frac{\partial f}{\partial B}\Big|_{T \Leftrightarrow \beta}$

per spin $\leftarrow$ $$\frac{\chi}{N} \equiv \frac{\partial \langle m\rangle}{\partial B}\Big|_T = -\frac{\partial^2 f}{\partial B^2}\Big|_T$$

$$\langle m^2\rangle - \langle m\rangle^2 = -\frac{\beta}{\beta^2}\,\frac{\partial^2 f}{\partial B^2}\Big|_T = \chi k T$$ here proven at $B=0$

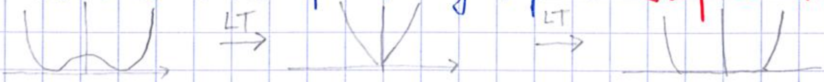
We could add an external field $B_{ext} \neq 0$: $\mathcal{K}(t) = \beta f(\beta, B_{ext}) - \beta f(\beta, B_{ext} + t/\beta)$

Rk 2 We stated Gärtner-Ellis in a mellow physicist sense. The thm essentially says that

$$\lim_{n\to\infty} \frac{1}{n} \log P_n\left(\bar{J} \in F\right) = -\inf_{j\in F} \phi(j)$$

Thm does not start by assuming that a large deviation principle holds.

It allows to treat cases where ϕ "has flat parts" $\Leftrightarrow$ phase transitions



Rk 3 The Legendre transform view allows to have more rigorous statements than before.

We will need Markov inequality: $\Pr[X \geq x] \leq \dfrac{\langle X \rangle}{x}$ for a non negative variable

$$\Pr[X \geq x] = \int_x^\infty p(x')dx' \leq \int_x^\infty \frac{x'}{x} p(x')dx' \leq \int_0^\infty \frac{x'}{x} p(x')dx' = \frac{\langle X \rangle}{x}$$

$$\Pr\left[\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n) \geq x\right] = \Pr[X_1 + \dots X_n \geq n x]$$

$$= \Pr[t(X_1 + \dots + X_n) \geq t n x] \quad , \quad \forall t > 0$$

$$= \Pr\left[e^{t(X_1+\dots+X_n)} \geq e^{tnx}\right]$$

$$\leq e^{-tnx} \left\langle e^{t(X_1+\dots+X_n)} \right\rangle \qquad \text{by Markov inequality}$$

$$\leq e^{-tnx} \langle e^{tX} \rangle^n \qquad X_i \text{ are independent (IID)}$$

$$\leq \exp\left\{-n\left(tx - \log\langle e^{tX}\rangle\right)\right\}$$

$$\leq \exp\{-n\,\phi(x)\} \quad ; \quad \phi(x) = \sup_{t>0}\left[tx - \log\langle e^{tX}\rangle\right]$$

$\phi(x)$ is the LT of the cumulant gen. function (up to the condition $t>0$ in the maximization) $\to$ the above bound is rigorous.

On the other hand, Sanov would say

$$\Pr[\overline{X}_n \geq x] \doteq e^{-nD[q^*\|p]} \quad ; \quad q^* = \operatorname{argmin}\left[D[q\|p] \, ; \int q(x')x'dx' \geq x\right]$$

6°) Mean-field recovered

Our discussion hides a strong connexion to stat. phys & phase transitions, that we would like to make explicit. Let $P(\mathcal{C}) = Z^{-1} e^{-\beta \mathcal{H}(\mathcal{C})}$ be the equil distrib for some interacting p model, with $F(\beta) = -\beta^{-1} \ln Z$. $\mathcal{C}$ denotes again a microscopic configuration.

Take any other distrib $P_0(\mathcal{C})$

$$D[P_0\|P] \equiv \sum_{\mathcal{C}} P_0 \log\left(\frac{P_0}{P}\right) = \sum_{\mathcal{C}} P_0\left[\log P_0 + \beta \mathcal{H}(\mathcal{C}) + \log Z\right]$$

$$= \underbrace{\underbrace{\beta \sum_{\mathcal{C}} P_0(\mathcal{C}) \mathcal{H}(\mathcal{C})}_{\beta\langle \mathcal{H}\rangle_0} + \underbrace{\sum_{\mathcal{C}} P_0 \log P_0}_{-S_0/k}}_{\beta\mathcal{F}[P_0]} - \beta F$$

Boltzmann weight

$$D[P_0\|P] = \beta\,\mathcal{F}[P_0] - \beta F(\beta) \geq 0$$

↳ convex-up functional of P_0, that achieves its unique minimum for $P_0 = P$; then $\mathcal{F}[P] = F(\beta)$, the "true" free energy

In computer science, $\mathcal{F}$ is sometimes called the "Gibbs free energy".
Misleading! The true Gibbs free energy is $G = U + PV - TS$ ... Also called "availability"

Opens a probabilistic interpretation. Take collection of $\mathcal{N} \ggg 1$ of independent copies of the same system, with the same $\mathcal{H}(\mathcal{C})$, given once for all. Then, compute $P_0(\mathcal{C})$, out of these $\mathcal{N}$ copies (empirical distribution). $P_0(\mathcal{C})$ will most probably be close to $P(\mathcal{C})$, the Boltzmann's distrib. For typical and atypical distributions, we can use Sanov

$$\Pr[P_0] \doteq \exp\left[-\mathcal{N} D[P_0 \| P]\right] \doteq \exp\left[-\mathcal{N}\beta\left(\mathcal{F}[P_0] - F\right)\right] \quad F \to \mathcal{F}[P]$$

When $F(\beta) \equiv \mathcal{F}[P]$ is not known (too complicated to compute, as usual), this provides a variational estimation for it.

Introduce the trial proba $P_0(\mathcal{C})$, depending on some parameters $(\lambda_1, \dots, \lambda_p)$ and minimize $\mathcal{F}[P_0]$ wrt $\lambda_1 \dots \lambda_p$ → gives an upper bound for $F(\beta)$

↳ essence of mean-field approximation, that often takes

$$P_0(\mathcal{C}) = \prod_i \underbrace{q_0(\sigma_i)}_{\text{for spins}} \quad ; \text{ factorized} \to \text{reduces to a 1-body problem.}$$

Particular situation: take $P_0 = \frac{1}{Z_0} e^{-\beta \mathcal{H}_0(\mathcal{C})}$ (and often $\mathcal{H}_0 = \sum_i \dots$)

$$F \leq \mathcal{F}[P_0] = \langle \mathcal{H} \rangle_0 + kT \sum_{\mathcal{C}} P_0 \ln P_0 = \langle \mathcal{H} \rangle_0 + kT\left[-\beta \langle \mathcal{H}_0 \rangle_0 + \beta F_0\right]$$

$$F \leq \underbrace{F_0 + \langle \mathcal{H} - \mathcal{H}_0 \rangle_0}_{\text{mean-field free energy, to be minimized wrt } \lambda_1 \dots \lambda_p} \qquad \text{Gibbs-Bogoliubov inequality}$$

Proper m.f. is always variational, as here. ⚠ naive mean-field, that would amount to minimizing F_0 wrt $\lambda_1, \dots \lambda_p$ → can lead to wrong results.

Ex Take Ising on some lattice $\mathbb{L}$, not necessarily 1d, $\mathcal{C} \equiv (\sigma_1, \dots, \sigma_N)$

and $P_0(\mathcal{C}) = \prod_{i \in \mathbb{L}} q_m(\sigma_i)$ $\qquad q_m(\sigma) = \frac{1+m}{2} \delta_{\sigma,1} + \frac{1-m}{2} \delta_{\sigma,-1}$

With P_0, the σ_i are IID with mean m: $\langle \sigma \rangle_0 = m$.

With P, " " " iD, but of course not independent

$$\mathcal{H}(\mathcal{C}) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

$$\mathcal{F}[P_0] = \langle \mathcal{H} \rangle_0 + kT \sum_{\mathcal{C}} P_0 \log P_0 = -J \frac{Nz}{2} m^2 + kT \overbrace{\sum_{\sigma_1 \dots \sigma_N} \sum_{i \in \mathbb{L}} P_0 \log q_m(\sigma_i)}^{\sum_{i \in \mathbb{L}} \sum_{\sigma_1} q_m(\sigma_1) \log q_m(\sigma_1)}$$

$$= -J \frac{Nz}{2} m^2 + kT N \left[\frac{1+m}{2} \log \frac{1+m}{2} + \frac{1-m}{2} \log \frac{1-m}{2}\right] \quad \text{already met!}$$

$$\underset{m=0}{\sim} N \frac{m^2}{2} \left(-zJ + kT\right)$$

Hence the critical temperature, where $m=0$ ceases to be optimal: $kT_c = zJ$