LEGENDRE TRANSFORMATIONS

[CALLEN, p. 85 ff.]

Thermodynamics can be formulated and reformulated in many different guises. For instance, in the ENTROPY formulation the equilibrium states are characterized by the maximum of the entropy, as a function of the EXTENSIVE parameters subject to the constraints in place (e.g. constant ENERGY).

This is equivalent to the ENERGY representation, in which the equilibrium states have minimum energy, under the constraint of constant entropy.

IN ALL CASES, A CONSTRAINT IS FIXED AND ANOTHER QUANTITY IS EXTREMIZED

The equivalence is due to the fact that the (fundamental) surface of states, (FOR INSTANCE) $S = S(U, V, N_1, \ldots, N_t)$, is convex, is a homogeneous function of extensive parameters:

$$S(\lambda U, \lambda V, \ldots \lambda N_t) = \lambda S(U, V, \ldots, N_t)$$

and that

$$\left.\frac{\partial S}{\partial U}\right|_{V, N_1, \ldots, N_t} > 0$$

(monotonicity in U) so that the equation of the surface can be (INVERTED AND) written (ALTERNATIVELY AS) $U = U(S, V, N_1, \ldots, N_t)$.

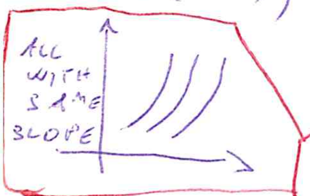
In both representations, the extensive parameters play the role of independent variables, whereas intensive parameters are derived concepts. This is in contrast to experimental practice, which finds often easier to control intensive variables.

In particular, there is no tool to measure the entropy, while thermometers and thermostats are common laboratory equipment.
Let us derive one further formulation of thermodynamics, in terms of intensive parameters.

Formally, let the fundamental relation take the form $Y = Y(X_0, X_1, \dots, X_t)$ FOR THE EXTENSIVE VARIABLES Y, X_i

and introduce $P_k = \frac{\partial Y}{\partial X_k}$ (IF Y, X_k ARE EXTENSIVE, THEN P_k ARE INTENSIVE

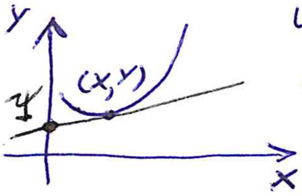
Begin with $t=1$. P is the slope of the curve $Y=Y(X)$, hence does not suffice by itself TO EXPRESS THE SET OF STATES, but knowledge of the tg lines suffices.



So, let Ψ be the intercept of the tg. line as a function of $P \Rightarrow \Psi = \Psi(P)$ is totally equivalent to $Y=Y(X)$.
Then P is the only needed independent variable IN THE NEW, EQUIVALENT REPRESENTATION.
In other words, $Y=Y(X)$ is a fundamental relation in the "Y-representation", whereas $\Psi = \Psi(P)$ is an EQUIVALENT

fundamental relation in the "Ψ-representation".

Ψ is obtained from Y as a "Legendre transformation":



We look for the intercept Ψ of the line of slope P tangent to $Y=Y(X)$, i.e. we minimize the distance between Y and $PX+q$: $\inf|Y-(PX+q)|=0 \Rightarrow q^*=\Psi=Y-PX$.

Indeed, $P=\dfrac{Y-\Psi}{X-0} \Rightarrow \boxed{\Psi=Y-PX}$ (*)

Ψ is called Legendre transform of Y; it exists if Y is convex, i.e. if Y' changes by changing X everywhere. This allows us to invert the relation $P=P(X)$ and obtain $X=X(P)$: e.g. $\dfrac{d^2Y}{dX^2}>0$ implies strictly increasing $\boxed{P(X)=\dfrac{dY}{dX}}$ and a different P for each different X, or a different X for a different P. In other words we can write $\boxed{\Psi=\Psi(P)}$.

Then, $\Psi=Y-PX \Rightarrow d\Psi=\underbrace{dY-dP\cdot X-PdX}_{P=\frac{dY}{dX}}=-X\,dP$

$\Rightarrow \boxed{-X=\dfrac{d\Psi}{dP}}$. If $\dfrac{d^2\Psi}{dX^2}\neq 0$, verified by convex Ψ, $X=X(P)$ can be inverted to yield $P=P(X)$ and knowledge of $\Psi=\Psi(P)$ can be used to obtain $Y=Y(X)$, where now

$$Y=\Psi+PX, \quad \text{with } \Psi=\Psi(P(X)),\ P=P(X).$$

Let us compare the relation between the representation Ψ and the representation Y (the first in terms of intensive parameter P, the second in terms of the extensive X):

$Y = Y(X)$	$\Psi = \Psi(P)$
$P = \dfrac{dY}{dX}$	$-X = \dfrac{d\Psi}{dP}$
$\Psi = -PX + \cancel{Y}$	$Y = PX + \Psi$
EXTENSIVE Y - REPRESENTATION	INTENSIVE Ψ - REPRESENTATION

In general, $Y = Y(X_0, X_1, \dots, X_t)$ = FUNDAMENTAL RELATION

$$\boxed{P_k = \frac{\partial Y}{\partial X_k}} = \text{PARTIAL SLOPE OF TANGENT HYPERPLANES}$$

$$\Psi = \Psi(P_0, \dots, P_t) = \text{INTERCEPT OF HYPERPLANES}$$

$$= Y - \sum_{k=0}^{t} P_k X_k$$

RULE: SUBTRACT PRODUCT OF OLD & NEW INDEPENDENT VARIABLES FROM GIVEN FUNCT WHERE THE NEW VARIABLE I THE DERIVATIVE OF THE FUNCTI W.R.T. THE OLD VARIABLE.

$$d\Psi = -\sum_{k=0}^{t} X_k \, dP_k \Rightarrow \boxed{-X_k = \frac{\partial \Psi}{\partial P_k}}.$$

RULE: SUBTRACT PRODUCT OF OLD & NEW VARIABLES, WITH NEW VARIABLE = DERIVATIVE OF FUNCTION W.R.T. OLD VARIABLE

NOTE: a Legendre transformation may EVEN be performed on a subset of coordinates.

In classical mechanics, the fundamental relation is the Lagrangian

$$L = L(v_1, \dots, v_r, q_1, \dots, q_r)$$ OLD RELATION IN OLD VARIABL v_k

$$P_k = \frac{\partial L}{\partial v_k} = \text{generalized momenta}$$ NEW VARIABLE

$$(-H) = L - \sum_{k=1}^{r} P_k v_k = \text{PARTIAL LEGENDRE TRANSFORM W.r.t. VELOCITIES}$$

$$= -H(P_1, \dots, P_r, q_1, \dots, q_r) = \text{HAMILTONIAN}$$ NEW RELATI IN NEW VARIABLES P_k

In thermodynamics, the fundamental relation in energy-representation is $U = U(S, V, N_1, \dots, N_r)$. Its derivatives are the intensive parameters

$$T = \frac{\partial U}{\partial S}\ ; \quad -P = \frac{\partial U}{\partial V}\ ; \quad \mu_k = \frac{\partial U}{\partial N_k}$$

The Legendre transform functions (OF U) are called THERMODYNAMIC POTENTIALS (e.g.):

$$\overset{\text{NEW RELATION}}{F} = \underset{\text{OLD RELATION}}{U} - \overset{\text{NEW VARIABL}}{T}\ \underset{\text{ASSOCIA OLD VARIABL}}{S}$$

where elimination of U and S yields

THIS REPLACES S, ALL OTHERS REMAIN

$F = F(T, V, N_1, \dots, N_r)$. Then

$$-S = \frac{\partial F}{\partial T}\ ; \quad U = F + TS$$

where elimination of F and T yields $U = U(S, V, N_k$

The Helmholtz potential (free energy) is the partial Legendre transform of U which replaces S by T as independent variable.

Its elementary variation gives:

$$dF = -S\,dT - P\,dV + \sum_{k=1}^{r} \mu_k\, dN_k$$

THERE IS NO TdS TERM HERE, BECAUSE S IS NOT INDEPENDENT ANYMORE. P WAS NOT IND TO BEGIN WITH.

The ENTHALPY is the partial Legendre transfor of U that replaces the V by P as independent variable: $H = U + PV$ (BECAUSE $-P = \frac{\partial U}{\partial V}$)

$$dH = T\,dS + V\,dP + \sum_{k=1}^{r} \mu_k\, dN_k$$

The sign inversion in P is because P is the intensive parameter associated with V.

$$U = H - PV \; ; \; V = \frac{\partial H}{\partial P} .$$

The Gibbs function (free energy) is the parti[al] Legendre transform of U which replaces S by T and V by P as independent variable[s]

$$U = U(S, V, N_k)$$

$$T = \frac{\partial U}{\partial S}$$

$$-P = \frac{\partial U}{\partial V}$$

$G \equiv U - TS + PV \Rightarrow$ eliminating $U, S, V \Rightarrow$

$$G = G(T, P, N_k)$$

$$dG = -S\,dT + V\,dP + \sum_{k=1}^{r} \mu_k \, dN_k$$

If one starts from the entropy formulation with fundamental relation $S = S(U, V, N_1, \ldots, N_r)$ Legendre-transforming one obtains the so-called MASSIEU FUNCTIONS which are particularly useful in irreversible thermodynamics and in the theory of fluctuations.

Replacing U by $\frac{1}{T}$ leads to

BECAUSE
$$\left(F + TS = U \Rightarrow S = \frac{1}{T} U - \frac{F}{T} \right)$$
AND TO REMOVE U : $\frac{\partial S}{\partial U} = \frac{1}{T}$ ← NEW VARIABLE

$$\underset{\text{NEW S-FUNCTION}}{S^*\left[\tfrac{1}{T}\right]} = \underset{\text{OLD FUNCTION}}{S} - \underset{\text{NEW VARIABLE} = \frac{\partial S}{\partial U}}{\frac{1}{T}} \; \underset{\text{OLD VARIABLE}}{U} = -\frac{F}{T}$$

replacing V by $\frac{P}{T}$, because $S = -\frac{1}{T}G + \frac{1}{T}U + \frac{P}{T}V$; $\frac{\partial S}{\partial V} = \frac{P}{T}$ ← NEW VARIABLE (†)

OLD VARIABLE

YIELDS

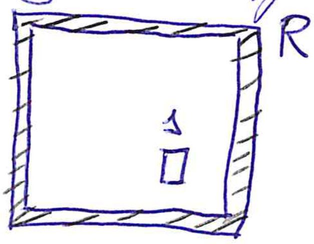
$$\boxed{S^*[P/T] = S - \frac{P}{T}V} \; ; \; \boxed{-V = \frac{\partial S^*}{\partial (P/T)}}$$

Replacing U by $\frac{1}{T}$ and V by $\frac{P}{T}$, because $S = -\frac{1}{T}G + \frac{1}{T}U + \frac{P}{T}V$ hence $\frac{\partial S}{\partial U} = \frac{1}{T}$; $\frac{\partial S}{\partial V} = \frac{P}{T}$, yields

$$\boxed{S^*\left[\frac{1}{T}, \frac{P}{T}\right] = S - \frac{1}{T}U - \frac{P}{T}V = -\frac{1}{T}G}$$

The above constitutes the phenomenological thermodynamic description. Statistical mechanics tries to reproduce thermodynamics from the molecular point of view, that is characterized by fluctuating microscopic quantities, for which a statistical description is required.

For instance, consider a large isolated system R, of which we observe a part s.



Let x be an extensive variable of s, i.e. a LOCAL variable. In general x fluctuates and its statistics can be quite generally given by a sort of CANONICAL DISTRIBUTION

$$f(x) = \frac{1}{e^{\Psi(h)}} e^{-hx}, \quad e^{\Psi(h)} = \int_{-\infty}^{\infty} e^{-hx}\, dN(x)$$

Where $N(x)$ can be selected according to the needs (quite "arbitrarily" or quite "generally") so that

one can write

$$P(x < x_0) = \int_{-\infty}^{x_0} e^{-hx}\, dN(x)\, e^{-\Psi(h)}$$

Introduce the moment generating function of X (8)

$$\langle e^{\alpha X}\rangle = e^{-\Psi(h)}\int e^{\alpha x} e^{-hx} dN(x) = e^{-\Psi(h)}\int e^{-(h-\alpha)x} dN(x)$$

$$= e^{-\Psi(h)} e^{\Psi(h-\alpha)}$$

the cumulant generating function is then:

$$\ln\langle e^{\alpha X}\rangle = -\Psi(h)+\Psi(h-\alpha) = -\Psi(h)+\sum_{n=0}^{\infty}\frac{(-\alpha)^n}{n!}\Psi^{(n)}(h)$$

$$= \sum_{n=1}^{\infty}\frac{(-\alpha)^n}{n!}\Psi^{(n)}(h)$$

BY DEFINITION OF CUMULANTS

$$= \sum_{n=1}^{\infty}\frac{\alpha^n}{n!}\langle X^n\rangle_c \Rightarrow \boxed{\langle X^n\rangle_c = (-1)^n \Psi^{(n)}(h)}$$

Independently on the meaning of x and of N, let us call ENTROPY the following quantity:

$$S = -k_B\int f(x)\ln f(x)\, dN(x) = -k_B\int f(x)\left[-\Psi(h)-hx\right]dN(x)$$

$$= k_B\Psi(h)+k_B h\langle X\rangle \Rightarrow \boxed{k_B\Psi(h) = S - k_B h\langle X\rangle}$$

Then $k_B\Psi$ looks like the "Masieu function" of S in which the variable $\langle X\rangle$ is replaced by the variable $k_B h$

$k_B h = \frac{\partial S}{\partial\langle X\rangle}$ = intensive field conjugated to extensive variable X; indeed, we also have

$$\langle X\rangle = e^{-\Psi(h)}\int x e^{-hx} dN = e^{-\Psi(h)}\left(-\frac{\partial}{\partial h}\int e^{-hx} dN\right) = e^{-\Psi(h)}\frac{\partial e^{\Psi(h)}}{\partial h} =$$

$$= e^{-\Psi}\Psi' e^{\Psi} = \frac{\partial\Psi}{\partial h} = \frac{\partial(k_B\Psi)}{\partial(k_B h)}$$

AND $kh = \frac{1}{T}$ SO THAT $h = \beta$

If we identify x with mechanical energy E and $\langle X\rangle$ with internal energy U, then whatever choice for $N(x)$ leads to:

$$k_B\Psi = S - \beta U k_B = S - \frac{1}{T}U = -\frac{1}{T}F$$

i.e. $\boxed{\Psi(\beta) = -\beta F(\beta)}$

and the cumulants reproduce all known thermodynamic relations.

Then $\langle E\rangle_c = -\frac{d}{d\beta}\left(-\beta F(\beta)\right) = \frac{d}{d(1/\cancel{k_B}T)}\left(\frac{1}{\cancel{k_B}T}F\right) =$

$$= F + \frac{1}{T}\frac{\partial F}{\partial T}\frac{d(1/T)^{-1}}{d(1/T)} = U - TS + \frac{1}{T}(-S)(-1)\left(\frac{1}{T}\right)^{-2} = U$$

$$\langle E^2\rangle_c = \frac{d^2}{d\beta^2}\left[-\beta F(\beta)\right] = -\frac{d}{d\beta}U = -k_B\frac{dU}{d(1/T)}$$

$$= -k_B\frac{dU}{dT}\frac{\overbrace{d(1/T)^{-1}}^{T}}{d(1/T)} = k_B T^2 C_V$$

NOTE: C_V IS EXTENSIVE, i.e. $O(N)$ ~~like~~ $\langle E\rangle \Rightarrow$

$$\frac{\sqrt{\langle (E-\langle E\rangle)^2\rangle}}{\langle E\rangle} = O\left(\frac{1}{\sqrt{N}}\right) \longrightarrow 0 \text{ AS } N\to\infty$$

$$\langle E^3\rangle_c = -\frac{d^3}{d\beta^3}\left(\beta F(\beta)\right) = -\frac{d}{d\beta}k_B T^2 C_V =$$

$$= -k_B^2\frac{d}{d(1/T)}\left[T^2 C_V\right] = k_B^2 T^3\left[2C_V + T\frac{dC_V}{dT}\right]$$

These cumulants represent the connection between the fluctuations of $X = E$ and the derivatives of the associated thermodynamic potential Ψ.

THE ABOVE RELATIONS ALLOW US TO COMPUTE MICROSCOPIC QUANTITIES, SUCH AS THE MECHANICAL ENERGY AND ITS FLUCTUATIONS, IN TE[RMS] OF EQUILIBRIUM MACROSCOPIC QUANTITIES: THE PROCESS OPPOSITE TO THE O[NE] USUALLY THOUGHT TO BE FOLLOW[ED]. AND IN THE LARGE N LIMIT, ANY OBSERVABLE FLUCTUATION WILL BE A LARGE DEVIATION.

GENERAL THEORY [ARNOLD § 14]

The Legendre transform is a mathematical algori[thm] which allows us to pass from functions define[d] on a linear space to functions defined on its dual space (the space of all linear functionals on the original space).

DEF.: Let $y = f(x)$ be convex ($f''(x) > 0$). Take $p \in \mathbb{R}$ and consider the line $y = px$. Introd[uce] $F(p,x) = px - f(x)$ and look for the point $x(p)$ which maximizes F at constant $p \Rightarrow$

$$\boxed{g(p) = F(p, x(p)) = \max_x\left(px - f(x)\right)}$$

is called LEGENDRE TRANSFORM of f.

$x(p)$ is found by imposing $\frac{\partial F}{\partial x} = p - f'(x) = 0$ (10)
i.e. $f'(x) = p$, hence the condition $f'' > 0$ makes it univocally determined, if it exists. W.r.t. our thermodynamic notation, here we have $g = -\Psi$; $y = f$; $P = p$; $X = x$.

EX.1: $f(x) = x^2 \Rightarrow F(p,x) = px - x^2$; $\frac{\partial F}{\partial x} = p - 2x \Rightarrow$

$x(p) = \frac{1}{2}p \Rightarrow g(p) = p\overbrace{\frac{1}{2}p}^{x} - \overbrace{\frac{1}{4}p^2}^{f(x)} = \frac{1}{4}p^2 = F(p, x(p))$

$G(p) = px - g(p)$; $\frac{\partial G}{\partial p} = x - g'(p) = 0 \Rightarrow p = 2x \Rightarrow f(x) = p(x)x - g(p(x)) = 2x^2 - \frac{1}{4}(2x)^2 = x^2$

EX.2: $f(v) = \frac{m}{2}v^2 \Rightarrow F(p,v) = pv - \frac{1}{2}mv^2 \Rightarrow \frac{\partial F}{\partial v} = p - m$

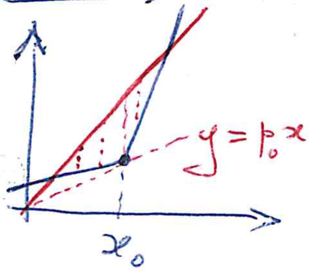
$v(p) = \frac{1}{m}p \Rightarrow g(p) = p\frac{1}{m}p - \frac{1}{2}\cancel{m}\frac{1}{m^{\cancel{2}}}p^2 = \frac{p^2}{2m} = F(p, v(p))$

EX.3: $f(x) = \frac{1}{\alpha}x^\alpha$; $F(p,x) = px - \frac{1}{\alpha}x^\alpha$; $\frac{\partial F}{\partial x} = p - x^{\alpha-1}$

$\Rightarrow x(p) = p^{\frac{1}{\alpha-1}} \Rightarrow g(p) = p\,p^{\frac{1}{\alpha-1}} - \frac{1}{\alpha}p^{\frac{\alpha}{1-\alpha}} = p^{\frac{\alpha}{1-\alpha}}\left(1 - \frac{1}{\alpha}\right)$

i.e. $g(p) = \frac{1}{\beta}p^\beta$ with $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ $(\alpha, \beta > 1)$

EX.4: consider the angle $f(x) = \begin{cases} p_1 x + q_1, & x \le x_0 \\ p_2 x + q_2, & x \ge x_0 \end{cases}$



$p_1 x_0 + q_1 = p_2 x_0 + q_2$

$\Rightarrow F(p,x) = \begin{cases} px - p_1 x - q_1, & x \le x_0 \\ px - p_2 x - q_2, & x \ge x_0 \end{cases}$

Here, we cannot find $x(p)$ by differentiation but the point of maximum difference (BETWEEN px ~~and~~ $f(x)$) is clearly x_0 in the convex part of the angle (distances grow linearly from 0 to x_0, decrease linearly from x_0 to second intersection) $\Rightarrow$
$g(p) = x_0 p - p_1 x_0 - q_1 = x_0 p - p_2 x_0 - q_2$

For $p<p_1$ and $p>p_2$, $y=px$ does not cut the convex part of the angle $\Rightarrow$ $D(g)=[p_1, p_2]$
DOMAIN OF g

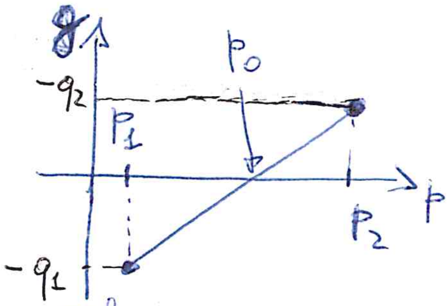
Also, $g(p_1) = p_1 x_0 - p_1 x_0 - q_1 = -q_1$

$g(p_2) = p_2 x_0 - p_2 x_0 - q_2 = -q_2$

$p_0 x_0 = p_1 x_0 + q_1 = p_2 x_0 + q_2 \Rightarrow g(p_0) = 0 \Rightarrow$

$\parallel$

$x_0 p_0 - p_i x_0 - q_i = 0$, $i=1,$



This example leads to the conclusion that, convex polygonal f produce a convex polygon g, such that vertices of f correspond to segments of g and segments of f to vertices of g.

If $f''>0$, g is also convex, $g''>0$, therefore one may Legendre transform twice & give convex function.

THM: $\mathcal{L}(f)=g \Rightarrow \mathcal{L}(g)=\mathcal{L}^2(f)=f$, i.e. the Legendre transform is an involution (hence has unit norm).

COR: Consider the family of straight lines $y=px-g$ for some g. The envelope of these lines is the function $f(x)$ whose Legendre transform is g.

REMARK: by definition, $F(x,p)=px-f(x)\leq \underbrace{g(p)}_{\sup_x(px-f(x))}$. This yields

YOUNG'S INEQUALITY $\quad px \leq f(x)+(\mathcal{L}f)(p)$

LARGE DEVIATIONS: a generalization of the CLT (1c) and the law of large numbers.

Consider a sequence of N coin tosses, where $H=1$, $T=-1$, and let X_n be the random variable of the outcome at the n-th tossing, $n\in\{1,\dots,N\}$, such that α is the probability of H, and $(1-\alpha)$ is the probability of

Introduce the R.V.

$$Y_N=\frac{1}{N}\sum_{n=1}^{N}X_N \quad [\langle Y_N\rangle=\langle X\rangle \text{ if all } X_n\sim$$

which takes values in $[-1,1]$. For large N, the values of Y_N are closely spaced. If the tosses are independent CLT states that

$$\mathbb{P}(Y_N\le y)\simeq\int_{-\infty}^{y}\frac{e^{-\frac{N(y'-\langle Y\rangle)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dy'$$

where $\langle Y\rangle=\alpha(+1)+(1-\alpha)(-1)=2\alpha-1$; $\langle Y^2\rangle=\alpha(+1)+(1-\alpha)(+1)=1\Rightarrow$

$\sigma^2=\langle Y^2\rangle-\langle Y\rangle^2=1-(4\alpha^2-4\alpha+1)=4\alpha(1-\alpha)$. This approximation, however, is good only for small deviations from $\langle Y\rangle$ i.e. for $O(|Y_N-\langle Y\rangle|)<O(\frac{1}{\sqrt{N}})$. Therefore, let us introduce the exact calculation, based on enumerating the possible cases [CLASSICAL NOTION OF PROBABILITY MODULATED BY α and $(1-\alpha)$] $\Rightarrow$ for K heads and $N-K$ tails, we have:

$$Y_N=\frac{1}{N}\sum_{k=1}^{N}X_k=\frac{1}{N}\left[K(+1)+(N-K)(-1)\right]=\frac{2K-N}{N}\Rightarrow$$

$$\mathbb{P}\left(Y_N=\frac{2K}{N}-1\right)=\underbrace{\frac{N!}{K!\,(N-K)!}}_{\#\text{ WAYS IN WHICH } K \text{ HEADS ARE OBTAINED}}\alpha^K(1-\alpha)^{N-K}$$

For moderately large N, $N-K$, K, one may use Stirling's approximation, that yields

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}}\frac{e^n}{n^n}=1\Rightarrow n!\simeq\sqrt{2\pi n}\,\frac{n^n}{e^n}\Rightarrow \ln n!=n\ln n-n+O(\ln n).$$

One may further introduce a new variable p s.t. $K=pN$; $N-K=(1-p)N$ $(p\in[0,1])$ and express

$$\boxed{\mathbb{P}(Y_N=2p-1)=e^{-NI(p)}}$$, where $p=\alpha$ means $Y_N=\langle$

One finds that, for large N, K, $N-K$

$$I(p)=I(p;\alpha)=p\ln\frac{p}{\alpha}+(1-p)\ln\frac{1-p}{1-\alpha}$$

which is known as RELATIVE ENTROPY of p w.r.t. α or KULLBACK–LIEBLER divergence.

Clearly, $I(p;\alpha)>0$ for $\alpha\neq p$, and $I(p;\alpha)=0$ for $\alpha=p$.

Indeed, for large N, pN, $N-K$, writing

$$P(Y_N = 2p-1) = \frac{N!(\alpha)^{pN}(1-\alpha)^{(1-p)N}}{(pN)![(1-p)N]!} = e^{-NI(p;\alpha)} \text{ implies}$$

$$NI(p;\alpha) = -\ln\left[N!\alpha^{pN}(1-\alpha)^{(1-p)N}\right] + \ln\left[(pN)![(1-p)N]!\right] =$$

$$= -\ln N! - pN\ln\alpha - (1-p)N\ln(1-\alpha) +$$
$$+ \ln(pN)! + \ln[(1-p)N]! \simeq$$

$$\simeq -\{N\ln N - N + pN\ln\alpha + (1-p)N\ln(1-\alpha)\} +$$
$$+ pN\ln pN - \cancel{pN} + (1-p)N\ln(1-p)N - (\cancel{1-p})N =$$

$$= pN\ln p + \cancel{pN\ln N} + (1-p)N\ln(1-p) + (\cancel{1-p})N\ln N - \cancel{N} +$$
$$- N\cancel{\ln N} + \cancel{N} - pN\ln\alpha - (1-p)N\ln(1-\alpha) =$$

$$= pN\ln\frac{p}{\alpha} + (1-p)N\ln\frac{1-p}{1-\alpha} \quad \text{to}$$

$$I(p;\alpha) \simeq p\ln\frac{p}{\alpha} + (1-p)\ln\frac{1-p}{1-\alpha}$$

as anticipated.
The argument can be repeated in a higher dimensional case, leading to

$$I(\{p\};\{\pi\}) = \sum_{j=1}^{m} p_j \ln\frac{p_j}{\pi_j}$$

if π_j are the m probabilities of the R.V. taking the values $a_1, a_2, \ldots, a_m$ rather than only a and $1-a$.
Coming back to one dimension, replace p with $y = 2p-1 \Rightarrow p = \frac{1}{2}(y+1)$ and I can be replaced by

$$S(y;\alpha) = \frac{y+1}{2}\ln\frac{y+1}{2\alpha} + \frac{1-y}{2}\ln\frac{1-y}{2(1-\alpha)}$$

Which is known as CRAMER FUNCTION, and one obtains

$$P(Y_N = y) \simeq e^{-NS(y)} \quad (*)$$

For y close to $\langle Y \rangle$, i.e. p close to α, one (14)
realizes that the central limit approximation
is good. Indeed, take $p = \alpha + \varepsilon$. Then:

$$I(p;\alpha) = (\alpha+\varepsilon)\ln\left(\frac{\alpha+\varepsilon}{\alpha}\right) + (1-\alpha-\varepsilon)\ln\frac{1-\alpha-\varepsilon}{1-\alpha} =$$

$$= (\alpha+\varepsilon)\ln\left(1+\frac{\varepsilon}{\alpha}\right) + (1-\alpha-\varepsilon)\ln\left(1-\frac{\varepsilon}{1-\alpha}\right) =$$

$$= (\alpha+\varepsilon)\left[\frac{\varepsilon}{\alpha} - \frac{\varepsilon^2}{2\alpha^2} + O\left(\frac{\varepsilon^3}{\alpha^3}\right)\right] + (1-\alpha-\varepsilon)\left[-\frac{\varepsilon}{1-\alpha} - \frac{\varepsilon^2}{2(1-\alpha)^2} + O\left(\frac{\varepsilon^3}{(1-\alpha)^3}\right)\right]$$

$$= \cancel{\varepsilon} - \frac{\varepsilon^2}{2\alpha} + O\left(\frac{\varepsilon^3}{\alpha^2}\right) + \frac{\varepsilon^2}{\alpha} - \frac{\varepsilon^3}{2\alpha^2} + O\left(\frac{\varepsilon^4}{\alpha^3}\right) +$$

$$-\cancel{\varepsilon} - \frac{\varepsilon^2}{2(1-\alpha)} + O\left(\frac{\varepsilon^3}{(1-\alpha)^2}\right) + \frac{\varepsilon^2}{1-\alpha} - \frac{\varepsilon^3}{2(1-\alpha)^2} + O\left(\frac{\varepsilon^4}{(1-\alpha)^3}\right)$$

FOR ε small compared to α and $1-\alpha$

$$= \frac{\varepsilon^2}{2\alpha} + O\left(\frac{\varepsilon^3}{\alpha^2}\right) + \frac{\varepsilon^2}{2(1-\alpha)} + O\left(\frac{\varepsilon^3}{(1-\alpha)^2}\right) =$$

if $\alpha \in (0,1)$

$$= \frac{\varepsilon^2}{2}\left(\frac{1}{\alpha} + \frac{1}{1-\alpha}\right) + O(\varepsilon^3)$$

$$= \frac{\varepsilon^2}{2}\,\frac{1-\alpha+\alpha}{\alpha(1-\alpha)} + O(\varepsilon^3) = \frac{\varepsilon^2}{2\alpha(1-\alpha)} + O(\varepsilon^3)$$

$$= \frac{(p-\alpha)^2}{2\alpha(1-\alpha)} + O(\varepsilon^3).$$ At the same time

$$\frac{(y-\langle Y\rangle)^2}{2\sigma^2} = \frac{(2p-1-2\alpha+1)^2}{8\alpha(1-\alpha)} = \frac{4(p-\alpha)^2}{8\alpha(1-\alpha)} = \frac{(p-\alpha)^2}{2\alpha(1-\alpha)}$$

So

$$e^{-NI(p;\alpha)} \simeq e^{-N\frac{\varepsilon^2}{2\alpha(1-\alpha)}} = e^{-N\frac{(y-\langle Y\rangle)^2}{2\sigma^2}}$$

Obviously, however, (*) of p. 13 is more general.

Consider, indeed, the random variable $NY_N = \sum_{n=1}^{N} x_n$, where the x_n are i.i.d. random variables. Let x be any of them; its cumulant generating function is $L(q) = \ln\langle e^{qx}\rangle$. Then, by independence we have:

$$\langle e^{qNY_N}\rangle = \langle e^{q\sum_{n=1}^{N} x_n}\rangle = \langle e^{qx_1}\cdots e^{qx_N}\rangle = \langle e^{qx_1}\rangle\cdots\langle e^{qx_N}\rangle$$

$$= \langle e^{qx}\rangle^N = e^{L(q)N}$$

Moreover, $\langle e^{qNY_N}\rangle = \sum_{k=0}^{N} e^{qNy_k}\, P(y_k)$ where

$Ny_k \in \{-N, -N+2, \dots, N-2, N\}$ which are $N+1$ values including 0 if N is even; including $-1, 1$ if N is odd:

$-N$ if 0 heads; $-N+2$ if 1 head etc. Furthermore,

$$\langle e^{qNY_N}\rangle \simeq \sum_{k=0}^{N} e^{qNy_k}\, e^{-NS(y_k)} = \sum_{k=0}^{N} e^{N[qy_k - S(y_k)]}$$

$$\boxed{e^{N\sup}/e^{N(\sup-\varepsilon)} = e^{N\varepsilon}}$$

For large N, this sum is dominated by the cases for which $N[qy - S(y)]$ is largest

$$\Rightarrow \langle e^{qNY_N}\rangle = O\left(e^{N\sup_y[qy - S(y)]}\right)$$

$$\| $$

$$L(q)^N$$

More precisely, given a range of values around $\sup_y[qy - S(y)]$, the number of terms in it only grows linearly with N so

$$\langle e^{qNY_N}\rangle \simeq cN\, e^{N\sup_y[qy - S(y)]}$$

$$\Rightarrow \quad e^{NL(q)} \simeq O(N)\, e^{N \sup_y [qy - S(y)]}$$

$$NL(q) \simeq \ln O(N) + N \sup_y [qy - S(y)] \quad \text{i.e.}$$

$$L(q) = \sup_y [qy - S(y)]$$

which says that L is the Legendre transform of S, and viceversa:

$$S(y) = \sup_q [qy - L(q)] \qquad (*)$$

For dependent variables x_n, one defines

$$L(q) = \lim_{N\to\infty} \frac{1}{N} \ln \langle e^{q \sum_{n=1}^{N} x_n} \rangle$$

and $(*)$ is exact if S is convex, otherwise $(*)$ gives the convex envelope of the correct S.

This way we see how to identify macroscopic thermodynamic quantities and microscopic fluctuating quantities. In particular we may continue the analogy and proceed with the cumulants.

INDEPENDENT

U, like S, is extensive, so let $X = \sum_{i=1}^{N} Z_i = N Y_N = N\left(\frac{1}{N}\sum_i^N Z_i\right)$

$$\Rightarrow \langle e^{\alpha X}\rangle = \langle e^{\alpha N Y}\rangle = \langle e^{\alpha Z}\rangle^N = e^{N L(\alpha)}$$

$$\langle e^{\alpha X}\rangle = \int e^{\alpha N Y} P_N(Y)\, dY$$

What could we take for P_N?

If we take $P_N \sim e^{-S/k_B} = e^{-Ns/k_B}$

Where S = entropy and $s = \frac{S}{N}$ = entropy per particle

as Einstein did, then for large N we repeat the argument of large deviations and we have

$$L(\alpha) = \sup_Y \left[\alpha Y - \frac{1}{k_B} s\right] \text{ etc.}$$

Indeed, Einstein inverted Boltzmann entropy $S = +k_B \ln W$