1 Review: Watson's Lemma and Laplace's method

1.1 Quick recap: Exponential integrals:

Assume for the moment that we want to calculate

$$F(x) = \int_0^T f(x,t)\,dt = \int_0^T e^{xh(x,t)} g(x,t)\,dt$$

for some large $x \gg 1$ (in general, both h, g are complex valued). Intuitively, we can guess that the behavior of this integral for $x \gg 1$ will depend only on the t for which $h(x,t)$ has the most positive real part (assuming that the imaginary part of h near that point doesn't fluctuate too wildly). On the other hand, if we insist on using a numerical quadrature rule, we find that the error of our approximation will be proportional to some (high-order) derivative of the integrand. For example, if we insist on using simpson's rule for quadrature we'll find that the error scales like x^4/N^4, where N is the number of discretization intervals we use. Now when $x \to \infty$, we will be forced to take $N \sim x/\sqrt[4]{\varepsilon}$ to obtain an error of ε. This is a supreme waste of time, given that the integral $F(x)$ typically has a simple form as $x \to \infty$.

1.2 Sketch of Watson's Lemma:

Assume $g(t)$ is smooth near $t = 0$, and that

$$\int_0^T |g(t)|\,dt < G < \infty.$$

Now let's consider

$$F(x) = \int_0^T f(x,t)\,dt = \int_0^T e^{-xt} g(t)\,dt \text{ as } x \to \infty.$$

Note that, because of the e^{-xt} term in the integrand, we don't have to worry much about values of $g(t)$ for any t that are significantly greater that 0. Thus, we can write:

$$F(x) \sim \int_0^s f dt \text{ for any } s > 0.$$

Now using a taylor's series for g in $t \in (0, s)$, we see:

$$g(t) = \sum_{n=0}^{N} g^{(n)}(0)/n! \cdot t^n + \cdots$$

which lets us write:

$$\begin{aligned}
\int_0^s e^{-xt} g(t)\,dt &= \sum_{n=0}^{N} \frac{g^{(n)}(0)}{n!} \int_0^s e^{-xt} t^n dt + \cdots \\
&\sim \sum_{n=0}^{N} \frac{g^{(n)}(0)}{n!} \int_0^\infty e^{-xt} t^n dt + \cdots \\
&= \sum_{n=0}^{N} \frac{g^{(n)}(0)}{n!} \frac{(n+1)!}{x^{n+1}} + \cdots
\end{aligned}$$

1.3 Official version of Watson's Lemma:

Lemma: Assume $g(t)$ has infinitely many derivatives in some neighborhood of $t = 0$, and that

$$\int_0^T |t^\sigma g(t)|\,dt < G^\sigma < \infty \text{ for } \sigma > -1.$$

Then

$$\begin{aligned} F(x) &= \int_0^T f(x,t)\,dt = \int_0^T e^{-xt}t^{\sigma}g(t)\,dt \\ &\sim \sum_{n=0}^{\infty} \frac{g^{(n)}(0)\,\Gamma(\sigma+n+1)}{n!x^{\sigma+n+1}} \text{ as } x\to\infty \end{aligned}$$

is an asymptotic series (note that we are defining the asymptotic series with respect to the functions $\left\{1/x^{\sigma+n+1}\right\}_{n\geq 0}$).

Proof: By assumption, $\exists s>0$ s.t. $g(t)$ has infinitely many derivatives within $t\in(0,s)$.

$$F(x) = \int_0^s f dt + \int_s^T f dt.$$

Now the latter term

$$\begin{aligned} \left|\int_s^T f dt\right| &\leq \int_s^T e^{-xt}\left|t^{\sigma}g(t)\right| dt \\ &\leq e^{-xs}\int_s^T \left|t^{\sigma}g(t)\right| dt \leq e^{-xs}G^{\sigma} \\ &= o\left(x^{-(\sigma+n+1)}\right) \text{ as } x\to\infty\ \forall n\geq 0 \end{aligned}$$

and is thus 'beyond all orders' (i.e., we don't have to consider it when determining our asymptotic series). On the other hand, $\int_0^s f dt$ is not beyond all orders.

$$\begin{aligned} g(t) &= \sum_{n=0}^{N} g^{(n)}(0)/n!\cdot t^n + \gamma_N(t) \\ |\gamma_N(t)| &\leq \sup_{0<\tau<s}\left|g^{(N+1)}(\tau)\right| \frac{t^{N+1}}{(N+1)!}, \text{ for } t\in[0,s]. \end{aligned}$$

Using this taylor-expansion for g, we can write

$$\begin{aligned} \int_0^s e^{-xt}t^{\sigma}g(t)\,dt &= \sum_{n=0}^{N}\frac{g^{(n)}(0)}{n!}\int_0^s e^{-xt}t^{\sigma+n}dt + \int_0^s e^{-xt}t^{\sigma}\gamma_N(t)\,dt \\ \left|\int_0^s e^{-xt}t^{\sigma}\gamma_N(t)\,dt\right| &\leq \int_0^s e^{-xt}t^{\sigma}\left|\gamma_N(t)\right| dt \\ &\leq \sup_{0<\tau<s}\left|g^{(N+1)}(\tau)\right|\frac{1}{(N+1)!}\int_0^s e^{-xt}t^{\sigma+N+1}dt. \end{aligned}$$

Now we need to determine the behavior of

$$\begin{aligned} F_p(x,s) &= \int_0^s e^{-xt}t^{\sigma+p}dt \\ &= \int_0^{\infty} e^{-xt}t^{\sigma+p}dt - \int_s^{\infty} e^{-xt}t^{\sigma+p}dt \\ &\quad \text{(and letting } \tau \text{ equal } xt) \\ &= \frac{1}{x^{\sigma+p+1}}\int_0^{\infty} e^{-\tau}\tau^{\sigma+p}d\tau - \int_s^{\infty} e^{-xt}t^{\sigma+p}dt \\ &= \frac{\Gamma(\sigma+p+1)}{x^{\sigma+p+1}} - \int_s^{\infty} e^{-xt}t^{\sigma+p}dt \end{aligned}$$

Noting that the very last term is beyond-all-orders:

$$\begin{aligned}\int_s^\infty e^{-xt}t^{\sigma+p}dt &\leq \left[\int_s^\infty e^{-xt}dt\right]^{1/2}\left[\int_s^\infty e^{-xt}t^{2\sigma+2p}dt\right]^{1/2} \\ &\quad (\text{using } \vec{a}\cdot\vec{c} \text{ less than } |\vec{a}|\,|\vec{c}|\,) \\ &\leq \frac{1}{x}e^{-xs}\left[\int_s^\infty e^{-t}t^{2\sigma+2p}dt\right]^{1/2} \\ &\quad (\text{using } x \text{ greater than } 1) \\ &= o\left(1/x^{\sigma+n+1}\ \forall n\geq 0\right) \text{ as } x\to\infty\end{aligned}$$

we see that

$$F_p(x) = \frac{\Gamma(\sigma+p+1)}{x^{\sigma+p+1}} + o\left(1/x^{\sigma+n+1}\ \forall n\geq 0\right) \text{ as } x\to\infty,$$

and that $F_p(x)$ does not depend on $s>0$. Now we can immediately write

$$\int_0^s e^{-xt}t^\sigma g(t)\,dt = \sum_{n=0}^N \frac{g^{(n)}(0)}{n!}\frac{\Gamma(\sigma+n+1)}{x^{\sigma+n+1}} + o\left(1/x^{\sigma+N+1}\right).$$

Note that the upper limit T is not important for this asymptotic series, since

$$\begin{aligned}\left|\int_T^{T'} e^{-xt}t^\sigma g(t)\,dt\right| &\leq e^{-xT}\int_T^{T'}|t^\sigma g(t)|\,dt \\ &= o\left(1/x^{\sigma+n+1}\ \forall n\geq 0\right).\end{aligned}$$

Note that this lemma can be generalized: $x>0$ can be replaced with $z\in\mathbb{C}$, $\mathrm{Re}(z)>0$, and $\sigma>-1$ can be replaced with $\sigma\in\mathbb{C}$, $\mathrm{Re}(\sigma)>-1$. Moreover, if g only has M derivatives near zero, the lemma holds up to order $\sigma+M+1$.

1.4 Example: $g(t) = \log(1+t^2)/t^2$:

Consider

$$F(x) = \int_0^\infty e^{-xt}\log\left(1+t^2\right)dt.$$

Note that

$$\lim_{t^2\to 0^+}\log\left(1+t^2\right) \sim t^2 - \frac{\left(t^2\right)^2}{2} + \frac{\left(t^2\right)^3}{3} - \cdots$$

So we can choose

$$\sigma = 2,\ \ g(t) = 1 - \frac{1}{2}t^2 + \frac{1}{3}\left(t^2\right)^2 - \cdots$$

$$g^{(2n)}(0)/(2n)! = (-1)^n/n$$

and thus write

$$F(x) = \int_0^\infty e^{-xt}\log\left(1+t^2\right)dt \sim \sum_{n=0}^N \frac{(-1)^n}{n}\frac{(2+2n)!}{x^{2+2n+1}}.$$

1.5 Laplace's method:

To motivate Laplace's method, first consider a gaussian convolution

$$F(x) = \int_{-\alpha}^\beta e^{-xt^2}g(t)\,dt.$$

When x is large, the behavior is completely dominated by the point $t = 0$, at which point the function g can be approximated by a low-order polynomial. Because of symmetry considerations, only the even powers of g's taylor-expansion should matter (i.e., constant, parabolic term, 4th order term, etc). We can formalize this intuition:

$$\begin{aligned} F(x) &= \int_{-\alpha}^{\beta} e^{-xt^2} g(t)\, dt \\ &= \int_{-\alpha}^{0} e^{-xt^2} g(t)\, dt + \int_{0}^{\beta} e^{-xt^2} g(t)\, dt \\ &\quad \text{(using } \pm\sqrt{\tau} \text{ equals } t\text{)} \\ &= \frac{1}{2}\int_0^{\alpha^2} e^{-x\tau} g\left(-\sqrt{\tau}\right)\tau^{-1/2} d\tau + \frac{1}{2}\int_0^{\beta^2} e^{-x\tau} g\left(+\sqrt{\tau}\right)\tau^{-1/2} dt \end{aligned}$$

Now if

$$g(t) = \sum_{n=0}^{N} a_n t^n \text{ for } |t| < s,$$

we can approximate

$$\begin{aligned} \frac{1}{2}\int_0^{\alpha^2} e^{-x\tau} g\left(-\sqrt{\tau}\right)\tau^{-1/2} d\tau &\sim \frac{1}{2}\int_0^{T} e^{-x\tau} \sum_{n=0}^{N} a_n (-1)^n \tau^{n/2-1/2} d\tau. \\ \frac{1}{2}\int_0^{\beta^2} e^{-x\tau} g\left(+\sqrt{\tau}\right)\tau^{-1/2} d\tau &\sim \frac{1}{2}\int_0^{T} e^{-x\tau} \sum_{n=0}^{N} a_n (+1)^n \tau^{n/2-1/2} d\tau. \\ F(x) &\sim \int_0^{T} e^{-x\tau} \sum_{n=0}^{N} a_{2n} \tau^{n-1/2} d\tau. \end{aligned}$$

Note that the $a_{n\,=\,\text{odd}}$ terms have canceled out. Also note that:

$$\begin{aligned} \int_T^{\infty} e^{-xt} t^{n-1/2} dt &= O\left(e^{-xt}/x\right) \text{ as } x \to \infty \\ \int_0^{\infty} e^{-xt} t^{n-1/2} dt &= \frac{1}{x}\int_0^{\infty} e^{-s}\left(\frac{s}{x}\right)^{n-1/2} ds = \frac{\Gamma(n+1/2)}{x^{n+1/2}} \\ \Gamma(n+1/2) &= \frac{2n-1}{2}\Gamma(n-1/2) = \frac{(2n)!}{2^{2n} n!}\Gamma(1/2) \\ \Gamma(1/2) &= \sqrt{\pi}, \end{aligned}$$

thus

$$\int_0^{T} e^{-x\tau} t^{n-1/2} dt = \sqrt{\frac{\pi}{x}}\frac{(2n)!}{2^{2n} n!}\frac{1}{x^n}$$

allowing us to write

$$F(x) \sim \sqrt{\frac{\pi}{x}} \sum_{n=0}^{\infty} a_{2n} \frac{(2n)!}{2^{2n} n!}\frac{1}{x^n} \text{ as } x \to \infty.$$

Note that the limits α, β, do not matter, since the 'exterior' integrals

$$\int_{\alpha^2}^{\infty} e^{-x\tau} g\left(-\sqrt{\tau}\right)\tau^{-1/2} d\tau \text{ and } \int_{\beta^2}^{\infty} e^{-x\tau} g\left(+\sqrt{\tau}\right)\tau^{-1/2} d\tau \text{ are both beyond-all-orders in } \{1/x^n\} \text{ as } x \to \infty.$$

1.6 Leading-order behavior of Laplace's method

Consider

$$F(x) = \int_{\alpha}^{\beta} e^{xh(t)} g(t)\, dt, \text{ for } x > 0.$$

This type of integral can occur when using green's functions to evaluate solutions of PDEs (see below). The behavior of $F(x)$ is determined by the behavior of $h(t)$ near its internal maximae. If, say, $h(t)$ has a single internal maximum

$a \in [\alpha, \beta]$, and h is smooth across a, then the integral basically behaves like a gaussian convolution (see example above). If, on the other hand, h is not smooth across a (but has some sort of cusp), or if the maximum of h occurs at an endpoint (either α or β) then things change.

For the first case, without loss of generality, consider

$$F(x) = \int_0^T e^{xh(t)} g(t)\, dt,$$

assuming that $h(0)$ is the only maximum of h in $t \in [0, T]$, and that $h'(0) = 0$, $h''(0) < 0$. Now

$$\begin{aligned} h(t) - h(0) &= \frac{1}{2} t^2 h''(0) + O\left(t^3\right) \\ e^{xh(t)} &= e^{xh(0)} e^{-s^2} \\ t &= s\sqrt{-2/h''(0)} + O\left(s^2\right) \end{aligned}$$

Near 0 we also have

$$\begin{aligned} g(t) &= g(0) + t g'(0) + \frac{1}{2} t^2 g''(0) + \cdots \\ g(t) &= g(0) + s\sqrt{\frac{-2}{h''(0)}} g'(0) + O\left(s^2\right), \end{aligned}$$

thus

$$\begin{aligned} F(x) &= \int_0^T e^{xh(t)} g(t)\, dt \\ &\sim e^{xh(0)} \sqrt{\frac{-2}{h''(0)}} \int_0^{T(s)} e^{-xs^2} \left[g(0) + O(s)\right] ds \\ &\sim e^{xh(0)} g(0) \sqrt{\frac{-2}{h''(0)}} \int_0^{\infty} e^{-xs^2} ds + e^{xh(0)} O\left(\int_0^{\infty} s e^{-xs^2} ds\right) \\ &\sim e^{xh(0)} g(0) \sqrt{\frac{-\pi}{2xh''(0)}} + e^{xh(0)} O(1/x), \end{aligned}$$

since $\int_0^{\infty} e^{-xs^2} ds = \sqrt{\pi/4x}$. Note that the leading-order behavior is like $1/\sqrt{x}$.

Similar reasoning can be used to tackle the case where $h(0)$ is the maximum with $h'(0) < 0$, in which case the leading-order behavior is like $1/x$:

$$F(x) \sim \frac{-g(0)}{h'(0)} \frac{e^{xh(0)}}{x} + e^{xh(0)} O\left(\frac{1}{x^2}\right) \text{ as } x \to \infty.$$

1.7 Example: Application to Gamma-function:

Consider the following form for the gamma-function:

$$\begin{aligned} \Gamma(x+1) &= \int_0^{\infty} t^x e^{-t} dt \text{ as } x \to \infty. \\ &= \int_0^{\infty} e^{x \log t} e^{-t} dt \end{aligned}$$

Now letting $g = e^{-t}$ and $h = \log t$ won't work, since h doesn't have a maximum in $(0, \infty)$. So insted we can let $\tau = t/x$ and consider

$$\begin{aligned} \Gamma(x+1) &= \int_0^{\infty} e^{x(\log x\tau)} e^{-x\tau} x d\tau = \int_0^{\infty} e^{x \log x} e^{x \log \tau} e^{-x\tau} x d\tau \\ &= x^{x+1} \int_0^{\infty} e^{x(\log \tau - \tau)} d\tau. \end{aligned}$$

Now letting $g = 1$ and $h = \log\tau - \tau$, we get a maximum of h at 1, with $h(1) = -1$, $h'(1) = 0$ and $h''(1) = -1/\tau^2|_{\tau=1} = -1$. Thus

$$\begin{aligned}\Gamma(x+1) &\sim x^{x+1}e^{-x}\sqrt{2\pi/x} \\ x! &\sim x^x e^{-x}\sqrt{2\pi x}.\end{aligned}$$

Higher-order terms can be obtained using laplace's method. For example:

$$\Gamma(x) \sim \sqrt{2\pi}x^{x-1/2}e^{-x}\left(1+\frac{1}{12x}\right).$$

2 Method of Steepest Descent (Saddle-Point Method)

2.1 Quick recap: Cauchy Riemann conditions:

Assume we have a complex differentiable function

$$h(z = x+iy) = R(x,y) + iI(x,y),$$

with $R = \mathrm{Re}(h)$ and $I = \mathrm{Im}(h)$ the real and imaginary parts of $h(z)$, respectively.

If indeed

$$h'(z) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[h(z+\varepsilon) - h(z)\right]$$

exists for all sequences $\varepsilon \to 0$, then in particular

$$\begin{aligned}h'(z) &= \lim_{\varepsilon\in\mathbb{R}\to 0}\frac{1}{\varepsilon}\left[h(z+\varepsilon) - h(z)\right] \\ &= \lim_{\varepsilon\in\mathbb{R}\to 0}\frac{1}{\varepsilon}\left[R(x+\varepsilon,y) + iI(x+\varepsilon,y) - R(x,y) - iI(x,y)\right] \\ &= \partial_x R(x,y) + i\partial_x I(x,y) \\ &= \lim_{\varepsilon\in\mathbb{R}\to 0}\frac{1}{i\varepsilon}\left[h(z+i\varepsilon) - h(z)\right] \\ &= \lim_{\varepsilon\in\mathbb{R}\to 0}\frac{1}{i\varepsilon}\left[R(x,y+\varepsilon) + iI(x,y+\varepsilon) - R(x,y) - iI(x,y)\right] \\ &= \partial_y I(x,y) - i\partial_y R(x,y).\end{aligned}$$

Thus

$$\partial_x R(x,y) = \partial_y I(x,y), \text{ and } \partial_x I(x,y) = -\partial_y R(x,y).$$

Note also that

$$\begin{aligned}\Delta R &= \partial_{xx}R + \partial_{yy}R = \partial_{xy}I - \partial_{yx}I = 0 \\ \Delta I &= \partial_{xx}I + \partial_{yy}I = -\partial_{xy}R + \partial_{yx}R = 0,\end{aligned}$$

and so both I and R are 'harmonic' if h is differentiable. Furthermore:

$$\nabla I(x,y) = \partial_x I\ \hat{x} + \partial_y I\ \hat{y}, \text{ and } \nabla R(x,y) = \partial_x R\ \hat{x} + \partial_y R\ \hat{y}.$$

Because $\partial_x R = \partial_y I$ and $\partial_y R = -\partial_x I$, this implies $\nabla R(x,y) = \partial_y I\ \hat{x} - \partial_x I\ \hat{y}$, and so

$$\nabla R(x,y)\cdot\nabla I(x,y) = 0.$$

In other words, contours of R are typically perpendicular to contours of I (except perhaps at points where the gradients are both 0).

2.2 Quick recap: maximum principle:

Given any function $u(\vec{x})$, and letting B_r be the ball of radius r around 0 (with surface area $S_r = S_1 r^2$) we have that the spherical average of $\vec{u}$ about 0 is given by

$$\frac{1}{S_r}\int_{B_r} u(\vec{x})\, dS = \frac{1}{S_1}\int_{B_1} u(r\vec{x})\, dS.$$

How does this change as a function of r?

$$\begin{aligned}\frac{d}{dr}\left[\frac{1}{S_1}\int_{B_1} u(r\vec{x})\, dS\right] &= \frac{1}{S_1}\int_{B_1} \partial_r u(r\vec{x})\, dS \\ &= \frac{1}{S_1}\int_{B_1} \nabla u(r\vec{x})\cdot \vec{n} dS \\ &= \frac{1}{S_1}\int_{B_1} \Delta u(r\vec{x})\, d\vec{x}.\end{aligned}$$

Thus, the average of u over a sphere changes (as the sphere inflates) at a rate proportional to the integral of Δu inside the sphere. This immediately implies that, if $\Delta u = 0$, then the surface-average of u over a sphere is equal to u at the center (i.e., mean-value-theorem). Consequently, the maximum of u within a sphere cannot occur in the interior, and must occur on the boundary (i.e., maximum principle).

2.3 Residue theorem:

Note that, if $k \neq -1$ and C is a closed contour enclosing 0, then

$$\oint_C z^k dz = 0, \text{ and } \oint_C \frac{1}{z} dz = 2\pi i \times [\text{number of times } C \text{ encircles 0 counter-clockwise}].$$

since every power of z^k has a smooth anti-derivative with the exception of z^{-1}, for which $\int z^{-1} dz = \ln(z)$ has an unavoidable branch cut (i.e., $\ln\left(re^{i\theta}\right) = \ln(r) + i\theta$). Now let's assume C encircles 0 counter-clockwise exactly once. If we have a function

$$f = \cdots + a_{-2}\frac{1}{z^2} + a_{-1}\frac{1}{z} + a_0 + a_1 z + a_2 z^2 + \cdots$$

then

$$\begin{aligned}\oint_C f dz &= 2\pi i a_{-1} \\ \oint_C \frac{f}{z^{k+1}} dz &= 2\pi i a_k.\end{aligned}$$

Note that we can interpret a_k as $f^{(k)}(0)/k!$.

2.4 Method of steepest descent: constant I in exponent

Consider first the evaluation of the complex integral

$$F(\lambda) = \int_C g(z)\exp(\lambda h(z))\, dz, \text{ for } \lambda \in R^+ \text{ as } \lambda \to \infty$$

where C is some contour in the complex plane. We would like to use Laplace's method to approximate this sort of integral. If $\operatorname{Im} h(z) = I_0$ is constant along C, then evaluating this integral is easy:

$$\begin{aligned}h(z = x + iy) &= R(x,y) + iI_0 \\ C &\to z(t) = x(t) + iy(t), \text{ for } t \in (a,b) \\ F(\lambda) &= \int_a^b g(x(t), y(t))\exp\lambda(R(x(t), y(t)) + iI_0)\frac{dz}{dt} dt \\ &= e^{i\lambda I_0}\left[\int_a^b e^{\lambda R}\operatorname{Re}(gz')\, dt + i\int_a^b e^{\lambda R}\operatorname{Im}(gz')\, dt\right]\end{aligned}$$

And laplace's method can be used on each part of this last equation. In other words, if the maximum of $R(x(t), y(t))$ occurs at t_c within the interior of (a, b), then

$$\int_a^b e^{\lambda R} g z' dt \sim e^{\lambda R(t_c)} g(t_c) z'(t_c) \sqrt{\frac{-2\pi}{\lambda R''(t_c)}} + e^{\lambda R(t_c)} O(1/\lambda).$$

On the other hand, if the maximum of R occurs at the endpoint a then

$$\int_a^b e^{\lambda R} g z' dt \sim \frac{-g(a) z'(a)}{R'(a)} \frac{e^{\lambda R(a)}}{\lambda} + e^{\lambda R(a)} O\left(\frac{1}{\lambda^2}\right).$$

2.5 Method of steepest descent: nonconstant I in exponent

Now we turn to the more general case in which $\mathrm{Im}(h(z))$ is not constant along C. In this case we can't directly use laplace's method, since the complex-phase of the integrand will oscillate wildly along the contour (with more and more rapid oscillations as $M \to \infty$). One conceptual step which avoids this problem is to shift the contour from C to some C' along which $\mathrm{Im}(h)$ is constant. How would such a contour C' behave? If $z(x, y)$ is a point on the contour C' along which $\mathrm{Im}(h(x, y)) = I(x, y)$ is constant, then we must have that $\nabla I(x, y)$ is normal to C' at $z(x, y)$. Now, by the powerful structure associated with complex differentiation, we know that if $h(x, y) = R(x, y) + iI(x, y)$ is complex-differentiable, then $\nabla R \cdot \nabla I = 0$. In other words, the contour C' which is a level set of I must in fact be a 'gradient ascent' (or descent) for the function R. Equivalently, curves of constant R and curves of constant I intersect perpendicularly. (all of these statements simply recapitulate the Cauchy-Riemann conditions). Note that, at a saddle point (where $h' = 0$), the contours for I and R intersect. This degenerate case will become important momentarily. Note that, due to maximum principle, the critical points where $h' = 0$ will never be a local maximum or minimum.

Ignoring saddle-points for the moment, an arbitrary contour C from z_1 to z_2 can typically be shifted to a new contour $C' = C_a + C_b + C_c$, where $\mathrm{Im}(h)$ is constant on $C_a : z_1 \to \hat{z}_1$ and $C_c : \hat{z}_2 \to z_2$, and $R(h)$ is minimized (and constant) along $C_b : \hat{z}_1 \to \hat{z}_2$. Thus, along C_a, the integral is dominated by the behavior of $R(z_1)$ (by construction the maximum value of R along this contour), and along C_b the integral is dominated by the behavior of $R(z_2)$. Typically either $R(z_1) > R(z_2)$ or the other way around, so only one maximum contributes to the overall integral. Along C_b on the other hand, R is fixed (and held at $R_{\min} < R(z_1), R(z_2)$). Thus, this contribution to the integral is beyond all orders in λ as $\lambda \to \infty$:

$$\left| \int_{C_b} e^{\lambda h(z)} g(z) dz \right| \leq e^{\lambda R_{\min}} \int_{C_b} |g| dz = o\left(e^{\lambda R(z_{1,2})} / \lambda^N \ \forall N \geq 0\right).$$

We may find that shifting the original contour C in the manner described above is impossible (see, e.g., the Boltzmann distribution below). In this case we may wish to cross a saddle point (by following an $\mathrm{Im}(h)$ constant contour over the saddle point). In this case the maximum of R along this stretch of the contour does not necessarily occur at the endpoint, and usually occurs in the middle (i.e, at the saddle point). Now $R(\text{saddle})$ will compete with $R(z_1)$ and $R(z_2)$ for dominance of the integral. In some situations (e.g., if the contour is closed or extends to infinity) the saddle-point might dominate the calculation.

3 Microcanonical Ensemble (Darwin-Fowler derivation)

Here we will provide a discussion which illuminates the origin of the boltzmann distribution. Consider first an ensemble of M noninteracting particles (each particle may be a 'system' in its own right). Let's assume that these particles don't interact with each other, and are distinguishable from one another. Let's also assume that each particle has some value associated with it. This value, say an 'energy' (purely notational for now), takes one of countably many values

$$E_k, \text{ for } k = 0, 1, 2, \ldots.$$

Let's assume for now that each particle is equally likely to take on any of the possible energy values. In other words, knowing nothing about the ensemble, our 'prior-distribution' for particle energies is uniform (this assumption can be relaxed later).

Now let's make a further (and very important) assumption — assume that we have measured the ensemble and find that the average energy is some U. Not every ensemble will have an average energy U. What is the distribution

of particle energies within the set of admissable ensembles? I.e., if I measure the average energy to be U, and then I measure the energy of a particular particle from the ensemble, what value am I likely to see for the individual particle energy?

3.1 Quick aside: motivating example:

Let's say we have an ensemble of $M = 4$ particles, each which can take on an energy value within the set $\{-3, -1, +1, +3\}$. If we measure our ensemble and find the average energy is $U = +0.5$, then the possible energy distributions are:

$$\begin{aligned}
[-1,+1,+1,+1]\text{, and permutations (4} &= \text{4!/1!3! total)} \\
[-1,+1,+3,-1]\text{, and permutations (12} &= \text{4!/2!1!1! total)} \\
[-3,+3,+1,+1]\text{, and permutations (12} &= \text{4!/2!1!1! total)} \\
[-3,+3,+3,-1]\text{, and permutations (12} &= \text{4!/2!1!1! total).}
\end{aligned}$$

If we presume that the prior distribution of energy values is uniform (i.e., each particle has a 25% chance of possessing any particular energy value), then the distribution of individual particle energy values *GIVEN* that the average energy is $+0.5$ is:

$$\begin{aligned}
P\left(E=-3|U=+0.5\right) &= \left(0\times 4+0\times 12+1\times 12+1\times 12\right)/\left(4\times\left(4+12+12+12\right)\right) = 24/160 \\
P\left(E=-1|U=+0.5\right) &= \left(1\times 4+2\times 12+0\times 12+1\times 12\right)/\left(4\times\left(4+12+12+12\right)\right) = 40/160 \\
P\left(E=+1|U=+0.5\right) &= \left(3\times 4+1\times 12+2\times 12+0\times 12\right)/\left(4\times\left(4+12+12+12\right)\right) = 48/160 \\
P\left(E=+3|U=+0.5\right) &= \left(0\times 4+1\times 12+1\times 12+2\times 12\right)/\left(4\times\left(4+12+12+12\right)\right) = 48/160
\end{aligned}$$

which is not uniform. The probability that an individual particle possesses an energy of -3 is less than the probability that a particle has an energy of $+3$ (given that $U = +0.5$). As you may suspect, if -5 was also a valid energy, then $P\left(E=-5|U=+0.5\right)$ would be even less than $P\left(E=-1|U=+0.5\right)$. Systematically applying this sort of reasoning in the limit as the number of particles (and number of valid energies) increases is at the heart of the boltzmann distribution.

3.2 Back to microcanonical ensemble:

Now suppose that we have m_k particles with energy E_k. We say that m_k is the 'energy count' associated with the energy E_k. Given average energy U, we have two constraints:

$$\sum_{k=0}^{\infty} m_k = M, \text{ and } \sum_{k=0}^{\infty} m_k E_k = UM.$$

Let's define $W_{\{m_k\}}$ as the number of distinct ways in which the various energy values can be assigned with their respective energy counts (i.e., with m_k particles having energy E_k).

$$W_{\{m_k\}} = M!/\left(m_0!m_1!m_2!\cdots\right).$$

Eventually, we would like to derive an expression for the mean energy-counts $\langle m_l \rangle$ for each energy E_l. We also want to show that (in the limit as $M \to \infty$) the actual energy-counts converge to these mean energy-counts. That is to say, the variance in energy-counts tends to 0:

$$\frac{\left\langle m_l^2 \right\rangle - \langle m_l \rangle^2}{M^2} \to 0 \text{ as } M \to \infty,$$

implying that the most likely l-energy-count of the ensemble is $\langle m_l \rangle$, and the probability of observing any particle at energy E_l converges to $\langle m_l \rangle / M$.

3.3 Define a generating-function Γ:

A useful mathematical device is the 'generating function'

$$\Gamma_{M,U}\left(g_0, g_1, g_2, \ldots\right) = \sum_{\substack{\{m_k\} \text{ such that} \\ \Sigma_k m_k = M \text{ and} \\ \Sigma_k m_k E_k = UM}} W_{\{m_k\}} g_0^{m_0} g_1^{m_1} g_2^{m_2} \cdots = \widetilde{\sum} W_{\{m_k\}} \Pi_k g_k^{m_k}$$

where the sum $\tilde{\Sigma}$ encapsulates the constraints. This generating function Γ is useful since it allows us to elegantly represent the expected value for each energy count:

$$\begin{aligned} g_l \partial_{g_l} \log(\Gamma)|_{g_k=1\ \forall k} &= g_l \frac{\partial_{g_l}\Gamma}{\Gamma}\Big|_{g_k=1\ \forall k} \\ &= g_l \frac{\widetilde{\sum} W_{\{m_k\}} m_l g_l^{m_l-1} \Pi_{k\neq l} g_k^{m_k}}{\widetilde{\sum} W_{\{m_k\}} g_l^{m_l} \Pi_{k\neq l} g_k^{m_k}}\Big|_{g_k=1\ \forall k} \\ &= \frac{\widetilde{\sum} W_{\{m_k\}} m_l}{\widetilde{\sum} W_{\{m_k\}}} \\ &= \langle m_l \rangle . \end{aligned}$$

(I think the prefactor of g_l is only kept for convention — as we'll see in a moment, restoring the power of g_l in the numerator allows for further application of $g_l\partial_{g_l}$ to calculate higher-order moments).

Note that, with appropriate relabelling of indices, the simple example above has the following expected energy counts:

$$\begin{aligned} \langle m_{-3} \rangle &= \frac{\widetilde{\sum} W_{\{m_k\}} m_{-3}}{\widetilde{\sum} W_{\{m_k\}}} = \frac{0\times 4 + 0\times 12 + 1\times 12 + 1\times 12}{4+12+12+12} = \frac{24}{40} \\ \langle m_{-1} \rangle &= \frac{40}{40}, \quad \langle m_{+1} \rangle = \frac{48}{40}, \quad \langle m_{+3} \rangle = \frac{48}{40}, \end{aligned}$$

where the condition $U = +0.5$ is implicitly woven into the sum $\tilde{\Sigma}$. Note that the probability of measuring a single particle and obtaining energy E_l is given by $\langle m_l \rangle / M$.

The generating function Γ also allows for a convenient representation of the variance in energy counts:

$$\begin{aligned} \left\langle m_l^2 \right\rangle &= \frac{1}{\Gamma} g_l \partial_{g_l} \left(g_l \partial_{g_l} \Gamma \right)\Big|_{g_k=1\ \forall k} \\ &= g_l \partial_{g_l} \left(\frac{1}{\Gamma} g_l \partial_{g_l} \Gamma \right) - g_l \left(\partial_{g_l} \frac{1}{\Gamma} \right) \left(g_l \partial_{g_l} \Gamma \right)\Big|_{g_k=1\ \forall k} \\ &= g_l \partial_{g_l} \left(g_l \partial_{g_l} \log\Gamma \right) + 2 g_l \frac{1}{\Gamma^2} \left(\partial_{g_l} \Gamma \right) \left(g_l \partial_{g_l} \Gamma \right)\Big|_{g_k=1\ \forall k} \\ &= g_l \partial_{g_l} \left(g_l \partial_{g_l} \log\Gamma \right) + \left(g_l \partial_{g_l} \log\Gamma \right)^2\Big|_{g_k=1\ \forall k} \\ &= g_l \partial_{g_l} \left(g_l \partial_{g_l} \log\Gamma \right)|_{g_k=1\ \forall k} + \langle m_l \rangle^2 , \end{aligned}$$

Thus:

$$\left\langle m_l^2 \right\rangle - \langle m_l \rangle^2 = g_l \partial_{g_l} \left(g_l \partial_{g_l} \log\Gamma \right)|_{g_k=1\ \forall k} .$$

Recall that we would (ultimately) like to show that

$$\frac{\left\langle m_l^2 \right\rangle - \langle m_l \rangle^2}{M^2} \to 0 \text{ as } M \to \infty,$$

implying that the most likely l-energy-count of the ensemble is $\langle m_l \rangle / M$. To obtain such an estimate, we need to get a better understanding of $\log\Gamma$. The tricky part is dealing with the constraints (which, right now, are implicitly accounted for in $\tilde{\Sigma}$).

3.4 Define generating-function G:

One useful device is (yet another) generating function G:

$$G_M(z; g_0, g_1, g_2, \ldots) = \sum_{U=0}^{\infty} z^{MU} \cdot \Gamma_{M,U}(g_0, g_1, g_2, \ldots),$$

which can be 'unpacked' into the following expression:

$$\begin{aligned} G_M(z; g_0, g_1, g_2, \ldots) &= \sum_{U=0}^{\infty} z^{MU} \cdot \widetilde{\sum} W_{\{m_k\}} g_0^{m_0} g_1^{m_1} g_2^{m_2} \cdots \\ &= \sum_{U=0}^{\infty} \sum_{\substack{\{m_k\} \text{ such that} \\ \Sigma_k m_k = M \text{ and} \\ \Sigma_k m_k E_k = UM}} z^{\Sigma_k m_k E_k} \cdot W_{\{m_k\}} g_0^{m_0} g_1^{m_1} g_2^{m_2} \cdots . \end{aligned}$$

Note that the two sums in the final expression actually account for every combination $\{m_k\}$ of energy-counts such that $\Sigma_k m_k = M$ and $\Sigma_k m_k E_k = UM$ exactly once. Thus, this double-sum can be replaced by a simpler sum:

$$G_M(z; g_0, g_1, g_2, \ldots) = \sum_{\substack{\{m_k\} \text{ such that} \\ \Sigma_k m_k = M}} W_{\{m_k\}} \left(g_0 z^{E_0}\right)^{m_0} \left(g_1 z^{E_1}\right)^{m_1} \left(g_2 z^{E_2}\right)^{m_2} \cdots .$$

Now note that this sum is simply the multinomial-formula:

$$\begin{aligned} G_M(z; g_0, g_1, g_2, \ldots) &= \left(g_0 z^{E_0} + g_1 z^{E_1} + g_2 z^{E_2} + \cdots\right)^M \\ &= \left[f(z; g_0, g_1, \ldots)\right]^M, \text{ with} \\ f(z; g_0, g_1, \ldots) &= \left(g_0 z^{E_0} + g_1 z^{E_1} + g_2 z^{E_2} + \cdots\right). \end{aligned}$$

3.5 Use Residue-Theorem:

Note that $\Gamma_{M,U}$ is the coefficient of z^{MU} in the expansion of G_M about $z = 0$. By using the residue-theorem we can write:

$$\begin{aligned} \Gamma_{M,U} &= \frac{1}{2\pi i} \oint_C \frac{f(z)^M}{z^{MU+1}} dz \\ &= \frac{1}{2\pi i} \oint_C \frac{1}{z} \exp M\left(\ln\left(f(z)\right) - U \log(z)\right) dz \\ &= \frac{1}{2\pi i} \oint_C \frac{1}{z} e^{Mh(z)} dz, \text{ with} \\ h(z) &= \ln\left(f(z)\right) - U \log(z), \text{ or } \exp\left(h(z)\right) = f(z)/z^U. \end{aligned}$$

3.6 With Laplace's-method in mind, find critical point of $h(z)$:

Note that M is a large parameter, and so in the limit $M \to \infty$ the behavior of the integral is dominated by the saddle-point (i.e., the x_c where $h'(x_c) = 0$). If we try and solve for the x_c such that $h'(x_c) = 0$, we find:

$$\begin{aligned} h'(z) &= 0 \\ f'(z)/f(z) &= U/z \\ \frac{\sum_k g_k E_k z^{E_k - 1}}{\sum_k g_k z^{E_k}} &= \frac{U}{z}. \end{aligned}$$

Assuming that $g_k = 1\ \forall k$, we have the last equality reducing to:

$$\frac{\sum_k g_k E_k z^{E_k}}{\sum_k g_k z^{E_k}} = U \implies \frac{\sum_k E_k z^{E_k}}{\sum_k z^{E_k}} = U.$$

Let's assume that $E_0 = 0$. Note that, for g_k sufficiently close to 1, the function $f(0) = 1$, and $f(x)$ is monotonically increasing faster than x^U (as a function of x) for $x \in \mathbb{R}^+$. Also, the function $1/x^U \to \infty$ as $x \to 0^+$, $1/x^U \to 0$ as $x \to \infty$, and $1/x^U$ is monotonically decreasing for $x \in \mathbb{R}^+$. Therefore, $\left[f(x)/x^U\right]^M = f(x)^M/x^{MU}$ has a minimum for some $x_c \in \mathbb{R}^+$ along the positive-real-axis. Since

$$\partial_z \left[f(x_c)^M / x_c^{MU}\right] = M h'(x_c) \exp\left(M h(x_c)\right) = 0,$$

we must have that $h'(x_c) = 0$. Now at this critical point x_c the second derivative of h takes the form:

$$\begin{aligned} h''(x_c) &= \frac{f''(x_c)}{f(x_c)} - \left(\frac{f'(x_c)}{f(x_c)}\right)^2 + \frac{U}{x_c^2} \\ &= \frac{f''(x_c)}{f(x_c)} - \left(\frac{U}{x_c}\right)^2 + \frac{U}{x_c^2}, \end{aligned}$$

and is real. Thus for $z \in \mathbb{C}$ near x_c, the function $h(z) - h(x_c)$ looks like a quadratic function of $z - x_c$:

$$\begin{aligned} h(z = x + iy) &= R(x,y) + iI(x,y) \\ z - x_c &= r\exp(i\theta) \\ R(x,y) - R(x_c,0) &= \frac{1}{2}|h''(x_c)|\, r^2\cos(2\theta) + O\left(r^3\right) \\ I(x,y) - I(x_c,0) &= \frac{1}{2}|h''(x_c)|\, r^2\sin(2\theta) + O\left(r^3\right). \end{aligned}$$

Note that the orientation of the contours of h near x_c is fixed by the behavior of h along the positive-real-axis: since the real axis is an I-contour the 'vertical' line $\theta = \pm\pi/2$ must be an I-contour as well. Since the saddle point along this vertical line occurs at $(x_c, 0)$, the maximum value of R (along this vertical line) occurs at the saddle-point $(x_c, 0)$.

3.7 Deform contour C to run through critical point of $h(z)$ (& use Laplace's-method):

Now if we try to approximate

$$\Gamma_{M,U} = \frac{1}{2\pi i}\oint_C \frac{1}{z}\exp(Mh(z))\,dz \text{ as } M \to \infty$$

we can deform C so that C_{x_c} is a circle of radius x_c, which passes 'vertically' through $(x_c, 0)$. How do we know that all other local maximae of $R(x,y)$ (along C_{x_c}) are smaller than $R(x_c, 0)$? Well, if we consider all $z = x_c\exp(i\theta)$, then

$$|\exp(h(z))| \le \frac{1}{x_c^U}\left|1 + \left(x_c e^{i\theta}\right)^{E_1} + \left(x_c e^{i\theta}\right)^{E_2} + \cdots\right|$$

which has a maximum when all the vectors $\left(x_c e^{i\theta}\right)^{E_k}$ are parallel, which happens when $\theta = 0$. We can ensure that these vectors are never parallel for any other $\theta \neq 0$ by choosing the E_k so that they don't share a common divisor (just choose units of energy appropriately so that $E_0 = 0$ and the next largest E_k, say E_1, is equal to 1). Thus, no maximum of $R(x,y)$ along C_{x_c} is comparable to the maximum which exists at $(x_c, 0)$.

Therefore, we can evaluate Γ as:

$$\begin{aligned} \Gamma_{M,U} &= \frac{1}{2\pi i}\oint_{C_{x_c}} \frac{1}{z}\exp(Mh(z))\,dz \\ &\quad \text{since } |dz/z| \text{ is } |dz|/x_c \text{ and } dz/z \text{ points in the } i \text{ direction} \\ &= \frac{1}{x_c 2\pi i}\oint_{C_{x_c}} \exp M\left(R(x_c,0) - \frac{1}{2}|h''(x_c)|\,(z-x_c)^2\right)|dz| \\ &\sim \frac{\exp(MR(x_c,0))}{x_c}\frac{1}{2\pi i}\int_{x_c-\infty i}^{x_c+\infty i}\exp\left(\frac{-M}{2}|h''(x_c)|\,(z-x_c)^2\right)|dz| \\ &\quad \text{letting } z - x_c \text{ be } iy \\ &\sim \frac{\exp(MR(x_c,0))}{x_c}\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-\frac{M}{2}|h''(x_c)|\,y^2\right)dy \\ &= \frac{\exp(MR(x_c,0))}{x_c}\sqrt{\frac{1}{2\pi M h''(x_c)}} + \frac{\exp(MR(x_c,0))}{x_c}\cdot O(1/M). \end{aligned}$$

Thus

$$\frac{1}{M}\log\Gamma_{M,U} = h(x_c) - \frac{1}{M}\log x_c - \frac{1}{2M}\log(2\pi M h''(x_c)) \;+ O\left(1/M^{3/2}\right) \text{ as } M \to \infty.$$

The important thing to note is that – to leading order in $1/M$ – the value of $1/M \cdot \log\Gamma_{M,U}$ is given by the value of $h(x_c)$.

3.8 Now take derivatives of $\Gamma_{M,U}$ to finds energy-counts:

The expression $\log(x_c)$ will show up frequently, and we know that $x_c \in \mathbb{R}^+$, so we can define

$$x_c = \exp(-\beta).$$

This allows us to write:

$$f(x_c) = \sum_{k=0}^{\infty} g_k x_c^{E_k} = \sum_{k=0}^{\infty} g_k \exp(-\beta E_k)$$

$$\begin{aligned} h(x_c) &= \log\left(\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)\right) + \beta U \\ g_l \partial_{g_l} h(x_c) &= g_l \frac{\exp(-\beta E_l)}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)} \end{aligned}$$

So, as $M \to \infty$, the energy-count tends towards:

$$\begin{aligned} \frac{\langle m_l \rangle}{M} &= g_l \partial_{g_l} \frac{1}{M} \log(\Gamma)\Big|_{g_k=1\ \forall k} \\ &= \frac{\exp(-\beta E_l)}{\sum_{k=0}^{\infty} \exp(-\beta E_k)} + O(1/M), \end{aligned}$$

which is the boltzmann distribution.

3.9 Variance about the mean-energy-counts tends to 0:

Recall that

$$\frac{1}{M^2}\left[\langle m_l^2 \rangle - \langle m_l \rangle^2\right] = g_l \partial_{g_l} \frac{1}{M}\left(g_l \partial_{g_l} \frac{1}{M} \log \Gamma\right)\Big|_{g_k=1\ \forall k}$$

and that

$$\frac{1}{M} \log \Gamma_{M,U} = h(x_c) - \frac{1}{M} \log x_c - \frac{1}{2M} \log(2\pi M h''(x_c)) + O\left(1/M^{3/2}\right).$$

The derivatives of the $h(x_c)$ term can be approximated via:

$$g_l \partial_{g_l} \frac{1}{M} g_l \frac{\exp(-\beta E_l)}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)} = O(1/M).$$

Similarly, we can consider the h'' term (calculated above):

$$\begin{aligned} h''(x_c) &= \frac{f''(x_c)}{f(x_c)} - \frac{U(U-1)}{x_c^2} \\ &= \frac{\sum_{k=0}^{\infty} g_k E_k (E_k - 1) \exp(-\beta(E_k - 2))}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)} - U(U-1)\exp(2\beta) \\ &= \exp(2\beta)\left[\frac{\sum_{k=0}^{\infty} g_k E_k (E_k - 1) \exp(-\beta E_k)}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)} - U(U-1)\right] \\ &= \exp(2\beta)\left[\frac{\sum_{k=0}^{\infty} g_k \left[E_k^2 - E_k - U^2\right] \exp(-\beta E_k)}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)} + U\right] \end{aligned}$$

Recalling that, at x_c, we have:

$$\frac{\sum_k g_k E_k z^{E_k}}{\sum_k g_k z^{E_k}} = U,$$

we can rewrite the last equality as:

$$h''(x_c) == \exp(2\beta)\left[\frac{\sum_{k=0}^{\infty} g_k \left[E_k^2 - U^2\right] \exp(-\beta E_k)}{\sum_{k=0}^{\infty} g_k \exp(-\beta E_k)}\right].$$

Thus, after taking derivatives we find:

$$g_l \partial_{g_l} \log h''(x_c) = g_l \left[\frac{\left(E_l^2 - U^2\right) \exp\left(-\beta E_l\right)}{\left[\sum_{k=0}^{\infty} g_k \left[E_k^2 - U^2\right] \exp\left(-\beta E_k\right)\right]} - \frac{\exp\left(-\beta E_l\right)}{\left[\sum_{k=0}^{\infty} g_k \exp\left(-\beta E_k\right)\right]} \right]$$

and so

$$\frac{1}{M} g_l \partial_{g_l} g_l \partial_{g_l} \log h''(x_c) = O(1/M) \text{ as well.}$$

These results can be combined to indicate that

$$\frac{1}{M^2} \left[\left\langle m_l^2 \right\rangle - \left\langle m_l \right\rangle^2 \right] \to 0 \text{ as } M \to \infty.$$

3.10 Summarize:

In conclusion, we see that the mean-energy-counts are given by:

$$\frac{\langle m_l \rangle}{M} = \frac{\exp\left(-\beta E_l\right)}{\sum_{k=0}^{\infty} \exp\left(-\beta E_k\right)} + O(1/M),$$

where the dependence on the mean-energy U is embedded in β by the constraint

$$h'(x_c) = 0, \text{ or rather } \frac{\sum_k E_k e^{-\beta E_k}}{\sum_k e^{-\beta E_k}} = U = \langle E \rangle.$$

4 Relationship to entropy:

4.1 Using the previous derivation:

Recall that, given N energy values $E_0, \ldots, E_{N-1}$, then the probability of finding the ensemble with average energy U is:

$$P(U) = \frac{1}{N^M} \sum_{\substack{\{m_k\} \text{ such that} \\ \Sigma_k m_k = M \text{ and} \\ \Sigma_k m_k E_k = UM}} W_{\{m_k\}} = \frac{1}{N^M} \widetilde{\sum} W_{\{m_k\}} = \frac{1}{N^M} \Gamma_{M,U}|_{g_i=1,\ \forall i}.$$

If we rely on the analysis above, we can associate $\Gamma_{M,U}$ with $h(x_c)$:

$$\Gamma_{M,U} \sim \exp\left(M h(x_c)\right) \text{ as } M \to \infty,$$

therefore

$$\begin{aligned} \frac{1}{M} \log\left(P(U)\right) &= h(x_c)|_{g_i=1\ \forall i} - \ln N + O(1/M) \\ P(U) &\sim \exp M \left(h(x_c)|_{g_i=1\ \forall i} - \ln N\right) \\ &= \exp M \left(\log\left(\sum_{k=0}^{\infty} \exp\left(-\beta E_k\right) \right) + \beta U - \ln N \right). \end{aligned}$$

This expression can be interpreted as:

$$P(U) \sim \exp\left(-M \cdot I(U)\right) \text{ for some 'rate-function' } -I(U) \sim \log\left(\sum_{k=0}^{\infty} \exp\left(-\beta E_k\right) \right) + \beta U - \ln N.$$

Note that, in this expression, the constant β is defined implicitly through $h'(x_c) = 0$.

4.2 Without using the previous derivation:

Now let's see if we can derive a similar relationship more directly, without using our analysis above. Let's again assume that we have M particles, each with an energy drawn from the set $\{E_0, E_1, \ldots, E_{N-1}\}$. Let's further assume that the energy of each particle is drawn independently and randomly, with a prior-probability $P(E = E_k)$ of $\tilde{\rho}_k$. If we draw our M particles in this fashion, the probability of obtaining a specific collection of energy-counts $\{m_k\}$ is given by:

$$P[\{m_k\}] = \frac{M!}{m_0!m_1!\cdots m_{N-1}!} \cdot \tilde{\rho}_0^{m_0}\tilde{\rho}_1^{m_1}\cdots\tilde{\rho}_{N-1}^{m_{N-1}}.$$

and since $\sum_k m_k = M$, we have that

$$\frac{1}{M}\log P[\{m_k\}] \sim \frac{1}{M}\log\frac{M!}{m_0!m_1!\cdots m_{N-1}!} - \frac{1}{M}\sum_k m_k \log\tilde{\rho}_k.$$

Now, by stirling's formula

$$\begin{aligned}
\log(M!) &\sim M\log M - M \\
\frac{1}{M}\log\frac{M!}{m_0!m_1!\cdots m_{N-1}!} &\sim \frac{1}{M}\left[M\log M - M - \sum_k m_k\log m_k + \sum_k m_k\right] \\
&= \log M - \sum_k \frac{m_k}{M}\log m_k \\
&= -\sum_k \frac{m_k}{M}\log\left(\frac{m_k}{M}\right).
\end{aligned}$$

Thus,

$$\frac{1}{M}\log P[\{m_k\}] \sim -\sum_k \frac{m_k}{M}\log\left(\frac{m_k}{M}/\tilde{\rho}_k\right).$$

Now as $M \to \infty$, the ratio m_k/M tends towards ρ_k — the fraction of particles drawn with an energy of E_k (i.e., the 'empirical' probability of finding a particle with $E = E_k$). This means that:

$$\frac{1}{M}\log P[\{m_k\}] \sim -\sum_k \rho_k \log(\rho_k/\tilde{\rho}_k),$$

or in other words:

$$P[\{m_k\}] \sim \exp\left(-M\sum_k \rho_k\log(\rho_k/\tilde{\rho}_k)\right) = \prod_k \frac{\rho_k^{M\rho_k}}{\tilde{\rho}_k^{M\rho_k}} = \prod_k \frac{\rho_k^{m_k}}{\tilde{\rho}_k^{m_k}}.$$

The function $\frac{1}{M}\log P$ is often referred to as the 'Relative-Entropy' (denoted $-I_{\tilde{\rho}}[\rho]$) of the empirical-distribution ρ with-respect-to the prior-distribution $\tilde{\rho}$.

As we can see, the relative-entropy serves as a 'rate-function' describing the probability of observing the empirical event-counts $m_k = M\rho_k$ (given prior-distribution $\tilde{\rho}$). Given a choice amongst different empirical-distributions ρ, the most probable empirical-distribution will be the one that minimizes this rate-function $I_{\tilde{\rho}}(\rho)$ (or, equivalently, maximizes the relative-entropy). With this in mind, let's see what happens if we try to minimize $I_{\tilde{\rho}}$ subject to the constraint that the average energy is U:

$$\min_{\sum_k \rho_k E_k = U} I[\{\rho_k\}] \implies I_{\min} = \sum_k \rho_k^c \log\frac{\rho_k^c}{\tilde{\rho}_k}, \text{ where } \rho_c \text{ is the optimal empirical-distribution,}$$

$$\text{subject to the constraint } K(\rho) = \sum_k E_k\rho_k = U.$$

The critical distribution ρ^c satisfies the following constrained optimization problem:

$$\nabla I = \beta\cdot\nabla K, \text{ for some lagrange-multiplier } \beta, \text{ where}$$

$$\begin{aligned}\partial_{\rho_j} I\left[\rho^c\right] &= \log\frac{\rho_j^c}{\tilde{\rho}_j} + \sum_k \tilde{\rho}_k = \log\frac{\rho_j^c}{\tilde{\rho}_j} + 1, \text{ and} \\ \partial_{\rho_j}\left[K\left(\rho\right)\right] &= \partial_{\rho_j}\left[\sum_k E_k\rho_k\right] = E_j.\end{aligned}$$

Putting these together we see:

$$\log\frac{\rho_k^c}{\tilde{\rho}_k} + 1 + \beta E_k = 0 \text{ , or } \rho_k^c = \frac{\tilde{\rho}_k}{e}\exp\left(-\beta E_k\right) \quad \forall k.$$

In addition, the critical distribution (and hence β) is subject to the constraint

$$K\left(\rho^c\right) = \sum_k \rho_k^c E_k = U.$$

If we now further stipulate that $\tilde{\rho}_k = 1/N$ (as we did when deriving the boltzmann-distribution), then we can see that

$$\rho_k^c = \frac{\exp\left(-\beta E_k\right)}{Z}, \text{ for } Z\left(\beta\right) = \sum_j \exp\left(-\beta E_j\right),$$

and that

$$\begin{aligned}\sum_k \rho_k^c \log\frac{\rho_k^c}{\tilde{\rho}_k} &= \sum_k \frac{\exp\left(-\beta E_k\right)}{Z}\log\frac{N\exp\left(-\beta E_k\right)}{Z} \\ &= \sum_k \frac{\exp\left(-\beta E_k\right)}{Z}\left[\log N - \beta E_k - \log Z\right] \\ &= \log N - \sum_k \frac{\beta E_k \exp\left(-\beta E_k\right)}{Z} - \log Z \\ &= \log N - \beta U - \log Z\end{aligned}$$

implying that, in fact,

$$P\left(U\right) \sim \exp M\left(\log Z + \beta U - \log N\right) = \exp M\left(\log\left(\sum_{k=0}^{\infty}\exp\left(-\beta E_k\right)\right) + \beta U - \ln N\right)$$

for the entropy maximizing distribution ρ_c. In other words, the lagrange mutliplier β (associated with the relative-entropy optimization problem contstrained via average energy U) is equal to the logarithm of the critical point of the exponent associated with the generating function Γ (also subject to the same mean-energy constraint).

5 One interpretation relating β to temperature

Let's revisit the system considered above, with M particles each with an 'energy' drawn from $\{E_0, E_1, \ldots, E_{N-1}\}$. Recall that the probability $P\left[\{m_k\}\right]$ of observing a given collection of energy-counts $\{m_k\}$ satisfies:

$$P\left[\{m_k\}\right] \sim \exp\left(-M\sum_k \rho_k \log\left(\rho_k/\tilde{\rho}_k\right)\right) = \prod_k \frac{\rho_k^{M\rho_k}}{\tilde{\rho}_k^{M\rho_k}} = \prod_k \frac{\rho_k^{m_k}}{\tilde{\rho}_k^{m_k}}, \text{ or}$$

$$\frac{1}{M}\log P\left[\{m_k\}\right] \sim -\sum_k \rho_k \log\left(\rho_k/\tilde{\rho}_k\right).$$

This means that, when M is large, we can expect $\log P$ to be a very steep function of $\{m_k\}$ (or, equivalently, of $\rho = \{m_k\}/M$).

Let's now assume that this system itself undergoes some kind of fast dynamics, which we'll abstract as follows: different particles in this system are allowed to rapidly exchange energy with one another, but the total energy in the system cannot change (i.e., energy is conserved[1]). If the mean-energy is known to be U, then even though the

[1] Up until now, the 'energy' has simply been a name we used, laking any inherent physical meaning; this is the first time that the energy takes on a physical interpretation – namely as a quantity that we expect to be conserved under some kind of natural dynamics.

system's energy-counts can change, the system's empirical-distribution ρ must always have mean-energy U; i.e., the system must live along the U-energy-shell. Because the dynamics of the system are fast, we expect the system to shift between various energy-counts rather quickly. Moreover, because we expect $\log P$ to be steep and for the contours of P to (in general) have nothing to do with the shape of the energy-shell, we expect the system to quickly adopt the single unique critical-empirical-distribution ρ^c that both (i) lies on the U-energy-shell and (ii) lies on on the I-contour with the smallest-possible value on the U-energy-shell. We'll later refer to the time it takes the system to reach ρ^c as the 'equilibration-time' of the system, and we'll assume that this equilibration-time is very fast compared to other time-scales.

If we assume that the prior-distribution for this system is uniform (i.e., $\tilde{\rho} = 1/N$), then the total number of system-configurations which correspond to the critical empirical-distribution ρ^c is given by:

$$\# = \frac{M!}{m_0! m_1! \cdots m_{N-1}!}, \text{ with } m_k = M\rho_k^c,$$

which can be approximated by:

$$\log(\#) = \log \frac{M!}{m_0! m_1! \cdots m_{N-1}!} \sim -\sum_k M\rho_k^c \log(\rho_k^c) = M\left[\beta U + \log Z\right],$$

$$\text{with } Z = \sum_j \exp(-\beta E_j) \text{ and } \beta \text{ implicitly defined via: } U = \frac{1}{Z}\sum_k E_k \exp(-\beta E_k).$$

If we were to now increase the mean-energy U by a small amount μ (say $U \to U + \mu$), then we shift the energy constraint from the U-energy-shell to the $(U+\mu)$-energy-shell. If we allow the system to evolve, it will rapidly redistribute this additional energy μ, quickly adopting a new critical-distribution $\rho^c(U+\mu)$ on the new $(U+\mu)$-energy-shell. Because β is the lagrange multiplier of the relative-entropy-optimization we performed earlier, we can see immediately that the rate-function I evaluated at this new $\rho^c(U+\mu)$ will change by $\beta\mu$ (indeed, β is defined so that $\nabla I = \beta \nabla K = \beta\mu$). This statement is the same as stating that $\log(\#)$ will increase by $M\beta\mu$, or that the total number of configurations associated with the critical empirical-distribution $\rho^c(U+\mu)$ will increase by a factor of $\exp(M\beta\mu)$. Conversely, if we were to remove μ energy from the system, the number of configurations would decrease by a factor of $\exp(M\beta\mu)$.

Now let's now consider two different systems, A and B, each with their own particles and energies. The first system 'A' will have M_A particles, each with an energy drawn from $\{E_0^A, E_1^A, \ldots\}$, a mean-energy of U_A, and a lagrange-multiplier of β_A. The second system 'B' will have M_B particles, each with an energy drawn from $\{E_0^B, E_1^B, \ldots\}$, a mean-energy of U_B, and a lagrange-multiplier of β_B. To start out let's assume that A and B are completely independent. With this independence assumption, we expect that each system will rapidly adopt its own critical empirical-distribution (i.e., ρ_A^c or ρ_B^c), each corresponding to its own number of admissible system-configurations (i.e., $\#_A$ or $\#_B$). Because of independence, the total number of system-configurations we expect when considering both systems simultaneously is

$$\#_{A\&B} = \#_A \cdot \#_B, \text{ or } \log(\#_{A\&B}) = \log(\#_A) + \log(\#_B).$$

Now we bring the two systems into 'contact', and allow the two systems to exchange energy very slowly (i.e., slowly with respect to the equilibration-time of each individual system). For example, we might assume that every once in a while any one system can donate some small amount of energy to a few randomly chosen particles from the other system. With such a slow energy-exchange, we expect both systems to remain *almost* entirely independent, with the exception that (i) their individual mean-energies change very slowly, and (ii) their total energy $M_A U_A + M_B U_B$ is constant. With these assumptions we expect that, after each occasional energy-exchange, each system rapidly reassumes the new critical empirical-distribution associated with its new energy. In other words, for most times each system can be described using the $\#$ and ρ^c associated with its U. Using this notation, we can describe the pair of systems together with a vector expressing their total energies: $[M_A U_A, M_B U_B]$, constrained such that $M_A U_A + M_B U_B = M U_T$ is fixed.

Under these assumptions let's examine what happens if system B were to donate u-total-energy to system A. I.e., the system-energies change from $[M_A U_A, M_B U_B]$ to $[M_A U_A + u, M_B U_B - u]$. Because the mean-energy of B decreases by $\mu = u/M_B$, we expect the total number of configurations $\#_B$ to decrease by a factor of $\exp(M_B \beta_B \mu) = \exp(\beta_B u)$. On the other hand, because the mean-energy of A increased by $\mu M_B/M_A$, we expect the total number of configurations $\#_A$ to increase by a factor of $\exp(M_A \beta_A \mu M_B/M_A) = \exp(\beta_A u)$. When considered together, the total number of configurations to the pair of systems has changed by:

$$\#_{A\&B} \to \#_{A\&B} \cdot \exp\left[(\beta_A - \beta_B)\, u\right].$$

One way to interpret this is as follows: If the lagrange-multiplier β_B associated with system B is smaller than the lagrange-multiplier β_A associated with system A, then the redistribution of u-total-energy from B to A drastically increases the total number of configurations available to the pair of systems. In other words, even though removing energy from B reduces $\#_B$, the increase in $\#_A$ more than makes up for it. Thus, if we were to consider all possible system configurations with total energy MU_T, there will be many more configurations corresponding to $[M_A U_A + u, M_B U_B - u]$ than to $[M_A U_A, M_B U_B]$. Conversely, if β_A is smaller than β_B, then the redistribution of u-total-energy from A to B drastically increases the total number of configurations available by a factor of $\exp\left[(\beta_B - \beta_A)\, u\right]$. Consequently, if we were to look at all possible system configurations with total energy MU_T, then we would see many more corresponding to $[M_A U_A - u, M_B U_B + u]$ than to $[M_A U_A, M_B U_B]$.

If we now allow the systems to exchange energy freely (but slowly), we can carry this logic forward. The most probable configuration of energies that we expect to see must have $\beta_A = \beta_B$ equals some equilibrium β_{eq}. Moreover, if the system naturally adopts such a configuration (i.e., if $\beta_A = \beta_B = \beta_{eq}$ is stable), then we expect that, nearby this equilibrium, a probabilistically-favorable redistribution of energy from A to B should increase β_A and decrease β_B (and vice-versa). With this interpretation it is natural to associate β_A and β_B with something like an 'inverse-temperature' for each system.

6 Example: Curie-Weiss Model

Now let's consider a system with N particles, each with spin $s_i = \pm 1$, for which the system's energy is:

$$E\left\{s_i\right\} = -J/2N \cdot \sum_{i,j} s_i s_j, \text{ for some constant } J > 0.$$

The energy will depend on the configuration of the s_i. Note however, that any permutation of the set $\{s_i\}$ has the same energy (i.e., the total energy only depends on the total numbers $m_\pm$ of $+1$ and -1 spins, respectively). This motivates the definition of the order parameter U:

$$U = \frac{1}{N} \sum_i s_i, \qquad U \in [-1, +1]\,.$$

Now, given U, we may ask how the spins are distributed (i.e., how many positive and negative spins — m_+ and m_- — are there?). Given U, we know that

$$\begin{aligned} m_+ - m_- &= NU \\ m_+ + m_- &= N \\ m_- &= N/2 - NU/2. \end{aligned}$$

Thus, out of the 2^N possible spin combinations, only

$$\binom{N}{N/2 - NU/2} = \frac{N!}{m_+! m_-!}$$

are admissible (i.e., can give rise to an average spin of U). Given an admissible collection of spins, the energy E is given by

$$\begin{aligned} -2N/J \cdot E\left\{s_i\right\} &= \sum_{i,j} s_i s_j \\ &= m_+ m_+ + m_- m_- - 2 m_+ m_- \\ &= \left(m_+ - m_-\right)^2 \\ &= N^2 U^2 \\ E\left\{s_i\right\} &= -\left(\frac{NU}{2}\right)^2 \cdot \frac{2J}{N}. \end{aligned}$$

Why am I grouping the terms $NU/2$? This is because $NU/2 = \left(m_+ - m_-\right)/2$, and if N is a large even integer, then $NU/2$ can take on any integer value from $-N/2$ (all negative spin) to $N/2$ (all positive spin). Now let's define

the partition function Z_N:

$$\begin{aligned} Z_N(z) &= \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} z^{E\{s_i\}} \\ &= \sum_{\substack{\frac{NU}{2}=\frac{m_+-m_-}{2} \\ =-N/2,\ldots,N/2}} \sum_{\substack{\{s_i\} \text{ such that} \\ m_++m_-=N \\ m_+-m_-=NU}} z^{E\{s_i\}} \\ &= \sum_{\substack{\frac{NU}{2}=\frac{m_+-m_-}{2} \\ =-N/2,\ldots,N/2}} \binom{N}{N/2-NU/2} z^{-\left(\frac{NU}{2}\right)^2 \cdot \frac{2J}{N}} \end{aligned}$$

In the limit as $N \to \infty$, we have that

$$\begin{aligned} Z_N(z) &\sim \int_{-\infty}^{\infty} \binom{N}{N/2-NU/2} z^{-\left(\frac{NU}{2}\right)^2 \frac{2J}{N}} d\left(\frac{NU}{2}\right) \\ &\sim \int_{-\infty}^{\infty} \binom{N}{N\left(\frac{1}{2}-\frac{U}{2}\right)} z^{-\left(\frac{U}{2}\right)^2 2JN} N d\left(\frac{U}{2}\right). \end{aligned}$$

Now using stirling's approximation:

$$\begin{aligned} \log(x!) &\sim x\log x - x \\ \log\binom{N}{N\left(\frac{1}{2}-\frac{U}{2}\right)} &\sim N\log N - N \\ &\quad -N\left(\frac{1}{2}-\frac{U}{2}\right)\log N\left(\frac{1}{2}-\frac{U}{2}\right) + N\left(\frac{1}{2}-\frac{U}{2}\right) \\ &\quad -N\left(\frac{1}{2}+\frac{U}{2}\right)\log N\left(\frac{1}{2}+\frac{U}{2}\right) + N\left(\frac{1}{2}+\frac{U}{2}\right) \\ &= -N\left[\log\left(\frac{1}{2}\right) + \frac{1}{2}(1-U)\log(1-U) + \frac{1}{2}(1+U)\log(1+U)\right]. \end{aligned}$$

Writing

$$z = \exp(-\beta),$$

we have:

$$Z_N(\beta) \sim \frac{N}{2} 2^N \int_{-\infty}^{\infty} \exp N\left(\left(\frac{U}{2}\right)^2 2J\beta - \frac{1}{2}(1-U)\log(1-U) - \frac{1}{2}(1+U)\log(1+U)\right) dU.$$

Using laplace's method, we know the integral is dominated by the sumpremum (critical point) of the exponent — namely,

$$\begin{aligned} \text{given } F(x) &= \int_{-\infty}^{\infty} e^{xh(t)} g(t)\, dt \\ \frac{1}{x}\log F(x) &= h(t_c) + O(1/x) \end{aligned}$$

and thus

$$\frac{1}{N}\log(Z_N(\beta)) = \sup_U \left\{\left(\frac{U}{2}\right)^2 2J\beta - \frac{1}{2}(1-U)\log(1-U) - \frac{1}{2}(1+U)\log(1+U)\right\}.$$

Now how can we characterize the critical point?

$$\begin{aligned} h(U) &= \left(\frac{U}{2}\right)^2 2J\beta - \frac{1}{2}(1-U)\log(1-U) - \frac{1}{2}(1+U)\log(1+U) \\ h'(U) &= U2J\beta + \frac{1}{2}\log(1-U) - \frac{1}{2}\log(1+U) \\ h''(U) &= 2J\beta - \frac{1}{(1-U^2)}. \end{aligned}$$

$$
\begin{aligned}
h\left(U_c\right) &= \text{maximum} \implies h'\left(U_c\right) = 0 \\
U_c 4J\beta &= \log\left(\frac{1-U_c}{1+U_c}\right) \\
U_c &= \frac{\exp\left(U_c 4J\beta\right) - 1}{\exp\left(U_c 4J\beta\right) + 1} \\
&= \tanh\left(U_c 2J\beta\right).
\end{aligned}
$$