

Markov Processes

- Reference: 1. Ramon van Handel, Probability in High Dimension, Lecture notes, Princeton University, 2016 (RuH book)
2. Solution sets to problems in Prob. in High Dim., Available online at: tylertkf.github.io/prob-reading/
3. Sinho Chewi, Log-concave sampling, Online notes, Nov. 2024

(I closely follow Chaps. 2, 3 in ①)

The goal is to show that Langevin Dynamics converges exponentially fast in time independent of dimension, if the distribution is strongly log-concave. To do so, we develop the basic theory of Markov Processes, the Poincaré inequality, the Bakry-Emery criterion and the Log Sobolev Inequality. Some of these ideas will be useful again later when we study convergence analysis of discretized Langevin and DDPM Sampling. (we will revisit M.P.s again when we study Ito calculus).

$$\{X_t, t \geq 0\} \quad X_t \in \mathbb{R}^d$$

is a continuous time stochastic process.

(thing hold more generally, but I will stick to $\mathbb{R}^d$, or a discrete state space)

Define the semi-group $\{P_t, t \geq 0\}$ associated with it by: $P_0 = I$, $P_s \circ P_t = P_{s+t}$, where for any bounded measurable function $f \in \mathcal{F}$, $f: \mathbb{R}^d \mapsto \mathbb{R}$,

$$P_t: \mathcal{F} \mapsto \mathcal{F} \quad \text{s.t.}$$

$$P_t f(x) = E[f(X_t) | X_0 = x]$$

Let the filtration associated with $\{X_t, t \geq 0\}$ be denoted by $\{\mathcal{G}_t, t \geq 0\}$, where

$$\mathcal{G}_t = \sigma\{X_s, s \leq t\}.$$

$\{X_t, t \geq 0\}$ is a time homogeneous Markov Process if

$$E[f(X_t) | \mathcal{G}_s] = P_{t-s} f(X_s)$$

informally, $E[f(X_t) | \{X_\sigma, \sigma \leq s\}]$

A probability measure $\mu : \mathcal{F} \mapsto \mathbb{R}$ is defined as:

$$\mu(f) = E_{z \sim \mu} [f(z)]$$

A probability measure associated with $\{P_t\}$ is said to be stationary if

$$\mu_*(P_t f) = \mu_*(f)$$

Note: We will abuse notation to use μ for two different but related things. (i) As above, μ maps a function to its expectation (integral). (ii) μ is also used to denote the density / measure itself, i.e.,

$$\begin{aligned} \underbrace{\mu(f)}_{f \in \mathbb{R}} &= E_{z \sim \mu} [f(z)] = \int \underbrace{f(z)}_{f \in \mathbb{R}} \underbrace{d\mu(z)}_{\int d\mu(z) = 1} \\ &= \int f(z) \mu(z) dz \end{aligned}$$

From context, it will usually be clear which one we mean.

$L_p(\mu_*)$ norm: $\|f\|_{L_p(\mu_*)} = \left(\int f(z)^p \mu_*(z) dz \right)^{1/p}$

equivalently,

$$\|f\|_{L_p(\mu_*)}^p = E_{x_0 \sim \mu_*} [f(x_0)^p].$$

Properties 1 (see Lemma 2.7 in RVH book for proofs).

1. $\|P_t f\|_{L_p(\mu_*)} \leq \|f\|_{L_p(\mu_*)}$ (i.e., P_t is non-expansive)

2. $P_t(\alpha f + \beta g) = \alpha P_t f + \beta P_t g$ (linearity)

3. $P_{t+s} f = P_t \circ P_s f = P_t(P_s f) = P_s(P_t f)$.
(semi group)

4. $P_t 1 = 1$ μ_* -a.s.

Variance $\text{Var}_{\mu_*}(f) = \mu_*(f^2) - \mu_*(f)^2$

Lemma: μ_* stationary, $f \in \mathcal{F}$. Then $\text{Var}_{\mu_*}(P_t f)$ non-increasing in t .

PF:

$$\text{Var}_{\mu_*}(P_t f) = \|P_t f - \mu_* f\|_{L^2(\mu_*)}^2 = \|P_t (f - \mu_* f)\|_{L^2(\mu_*)}^2$$

$0 \leq s \leq t$

$$= \|P_{t-s} P_s (f - \mu_* f)\|_{L^2(\mu_*)}^2 \leq \|P_s (f - \mu_* f)\|_{L^2(\mu_*)}^2$$

$$= \text{Var}_{\mu_*}(P_s f).$$

□

Semigroup Generator $\mathcal{A}$: The Generator of the M.P. $\{P_t, t \geq 0\}$ is: $\mathcal{A}: \mathcal{F} \mapsto \mathbb{R}$ s.t.

$$\mathcal{A}f \triangleq \lim_{t \searrow 0} \frac{P_t f - f}{t} \quad (\text{limit in } L^2(\mu_*)),$$

for $f \in \mathcal{L}_2(\mu_*)$. Please read RVH for a discussion on the domain of $\mathcal{A}$. In general, $\text{Domain}(\mathcal{A}) \neq \mathcal{L}^2(\mu_*)$.

Remarks: ① Informally, $P_t = e^{t\mathcal{A}}$.

② The generator characterizes the time evolution of the Markov Process. In particular,

$$\frac{d}{dt} P_t f = \lim_{h \rightarrow 0} \left(\frac{P_{t+h} f - P_t f}{h} \right)$$

$$\Rightarrow = \lim_{h \rightarrow 0} \frac{P_h \overset{g}{P_t f} - \overset{g}{P_t f}}{h} = \mathcal{L} g = \mathcal{L} P_t f.$$

$\nearrow (P_t, \mathcal{L})$
commute

$$\Rightarrow = \lim_{h \rightarrow 0} P_t \left(\frac{P_h f - f}{h} \right) = P_t \mathcal{L} f.$$

Notation: $\partial_t \equiv \frac{\partial}{\partial t}$

$\therefore$ If $u_t = P_t f$, then

$\partial_t u_t = \mathcal{L} u_t$

Kolmogorov Backward Equation
(describes time evolution of expected function value)

③ The dual for this is the Fokker-Planck Equation, or the **Kolmogorov Forward Equation**, developed below.

Before we discuss, we digress to recall adjoint operators over Banach spaces.

Aside: Adjoint on Banach Spaces.

Let $f \in \mathcal{F}_b(\mathbb{R}^d)$, the space of continuous bounded functions over $\mathbb{R}^d$.

Then, its dual space (space of linear functionals over $\mathcal{F}_b(\mathbb{R}^d)$) is $\mathcal{M}$, the space of finite Borel measures, i.e.; if $\mu: \mathcal{F}_b \rightarrow \mathbb{R}$ is a linear functional, then it can be represented (isometric isomorphism) by $\mu \in \mathcal{M}$, i.e.,

$$\mu(f) = \int f d\mu.$$

↳ linear functional

Similar to Hilbert space adjoints, we can now consider adjoint operators here (but cannot compose an adjoint and its dual; see C. Heil, Functional Analysis Lecture Notes: Adjoint in Banach Spaces, online notes: heil.math.gatech.edu/handouts/adjoint-banach.pdf)

The linear functionals are represented in sesquilinear form, i.e., $\mu(f) \doteq \langle f, \mu \rangle$, $f \in \mathcal{S}_b(\mathbb{R}^d)$, $\mu \in \mathcal{M}$.
linear in f antilinear in μ .

$$\text{and } \langle Af, \mu \rangle = \langle f, A^* \mu \rangle$$

operator adjoint operator

(The notation of $\langle \cdot, \cdot \rangle$ is overloaded here; recall that a Banach space, in general, is NOT an inner product space. For a "true" inner product, we need a Hilbert space.)

Kolmogorov Forward Equation

We have P_t^* as the adjoint of P_t , i.e., P_t acts on f to push the function value forward,

i.e., $P_t f(x) = E[F(X_t) | X_0 = x],$

and $E_{X_0 \sim \mu_0}[f(X_t)] = \int P_t f(x) d\mu_0(x)$

$\hookrightarrow \mu_0 = \text{dist. of } X_0 \text{ (need not be the stat. dist.)}$

P_t^* acts on the measure μ_0 to push the measure forward, i.e., the dist. of $X_t \sim \mu_t$, with $\mu_t = P_t^* \mu_0$.

$$\begin{aligned} \text{i.e., } E_{X_0 \sim \mu_0}[f(X_t)] &= \int f(x) d\mu_t(x) \\ &= \int f(x) dP_t^* \mu_0(x). \end{aligned}$$

Now, (we will suppress explicitly showing x below).

$$\partial_t E_{X_0 \sim \mu_0}[f(X_t)] = \partial_t \int f d \underbrace{P_t^* \mu_0}_{\mu_t}$$

(defn. of P_t, P_t^*)

$$= \partial_t \int P_t f d\mu_0 = \int \partial_t P_t f d\mu_0$$

$$= \int P_t \partial_t f d\mu_0 = \int \partial_t f dP_t^* \mu_0$$

Similar to P_t, P_t^* , we can define the adjoint of α , denoted by α^* , that acts on measures, i.e., $\alpha f = g$ (i.e., α maps functions to functions), and $\alpha^* \nu = \tilde{\nu}$ (i.e., α^* maps measures to measures). Specifically,

$$\int \alpha f d\nu = \int f d(\alpha^* \nu)$$

$$\partial_t \int f d(\underbrace{P_t^* \mu_0}_{\mu_t}) = \int \alpha f d(P_t^* \mu_0) = \int f d(\alpha^* P_t^* \mu_0)$$

This is true $\forall f \in \mathcal{F}$, thus we have

$$\partial_t \mu_t = \underbrace{\alpha^* P_t^* \mu_0}_{= \mu_t} = P_t^* \alpha^* \mu_0$$

i.e., α^* acts on measures, and describes density evolution:

$$\partial_t \mu_t = \alpha^* \mu_t$$

$\equiv$ Kolmogorov Forward Equation
Fokker-Planck Equation

This is the same F-P Equation we derived earlier for an SDE from first principles!

④ Bringing things together, we now have

$$\mu_* \text{ is stationary} \iff \mathcal{L}^* \mu_* = 0$$

$$\iff E_{z \sim \mu_*} [\mathcal{L} f(z)] = 0. \quad \forall f \in \mathcal{F}$$

Remark: Duals are used in two different ways, when discussing M.P., which sometimes can be a source of confusion.

(a) Banach space duality: When we discussed (P_t, P_t^*) and $(\mathcal{L}, \mathcal{L}^*)$, earlier, we are concerned with P_t and $\mathcal{L}$ acting on $f \in \mathcal{F}$, and P_t^* and $\mathcal{L}^*$ acting on Borel measure $\mu \in \mathcal{P}(\mathbb{R}^d)$. (see discussion earlier in these notes).

(b) Hilbert space duality with respect to μ -weighted inner product:

In reversibility, we are concerned with a different dual: P_t, α and $\tilde{P}_t^*, \tilde{\alpha}^*$ where all of these act on $f \in \mathcal{F}$:
 $\underbrace{\tilde{P}_t^*, \tilde{\alpha}^*}_{\rightarrow \text{Hilbert space adjoint operators}}$

$$\langle P_t f, g \rangle_{\mu_*} = \int P_t f(x) \cdot g(x) d\mu_*(x)$$

$$= \int f(x) \tilde{P}_t^* g(x) d\mu_*(x) = \langle f, \tilde{P}_t^* g \rangle_{\mu_*}$$

and similarly, $\langle \alpha f, g \rangle_{\mu_*} = \langle f, \tilde{\alpha}^* g \rangle_{\mu_*}$

A M.P. with (α, μ) is reversible if $P_t = \tilde{P}_t^*$, and $\alpha = \tilde{\alpha}^*$. In these notes, we will not be using $\tilde{P}_t^*, \tilde{\alpha}^*$, as for the most part, we consider only reversible M.P., which are self-dual.

Reversibility: A key idea in M.P. is reversibility, which informally says that the evolution of density when time is reversed, is the same as the evolution when time moves forward.

Formally, $\{P_t, t \geq 0\}$, associated with (d, μ_*) is reversible if

generator ↗ ↘ stat. measure

$$\langle P_t f, g \rangle_{\mu_*} = \langle f, P_t g \rangle_{\mu_*},$$

equivalently

$$E_{x_0 \sim \mu_*} [P_t f(x_0) \cdot g(x_0)] = E_{x_0 \sim \mu_*} [f(x_0) P_t g(x_0)]$$

$$\text{i.e., } E_{x_0 \sim \mu_*} [E[f(x_t) | x_0] g(x_0)] =$$

$$E_{x_0 \sim \mu_*} [f(x_0) E[g(x_t) | x_0]]$$

$$\text{i.e., } E_{x_0 \sim \mu_*} [f(x_t) g(x_0)] = E_{x_0 \sim \mu_*} [f(x_0) g(x_t)]$$

Ergodicity: A M.P. is ergodic if $\forall f \in \mathcal{L}^2(\mu)$,

$$P_t f \xrightarrow{t \rightarrow \infty} \mu(f) \quad \text{in } \mathcal{L}^2(\mu).$$

$$\downarrow \quad \quad \quad \downarrow$$
$$P_t f(x) = E[f(X_t) | X_0 = x] \quad \quad \mu(f) = E_{\mu}[f(z)]$$

i.e., irrespective of initial state $X_0 = x$, the M.P. time-varying density converges to μ .

Dirichlet form: This is a notion of "average energy" of the M.P. (as will become clearer later), and is defined as:

$$\begin{aligned} \mathcal{E}(f, g) &= - \langle f, \mathcal{L}g \rangle_{\mu} \\ &= - \int f(x) \mathcal{L}g(x) d\mu(x). \end{aligned}$$

Claim 2: For the SDE (Markov Process):

→ OU process.

$$dX_t = -X_t dt + \sqrt{2} dB_t, \quad X_t \in \mathbb{R}$$

we have:

(a) $\mu(x) = N(0, 1)$

(b) $P_t f(x) = E_{Z \sim \mu} \left[f(e^{-t}x + \sqrt{1-e^{-2t}} \cdot Z) \right]$

(c) $\mathcal{L} f(x) = -x f'(x) + f''(x)$

(For future ref: $\mathcal{L}^* p_n(x) = \frac{\partial}{\partial x} (x p_n(x)) + \frac{\partial^2}{\partial x^2} p_n(x)$)
density f_n associated with μ .

(d) $\mathcal{E}(f, g) = \langle f', g' \rangle_\mu$

$$\mathcal{E}(f, f) = \|f'\|_\mu^2 = \int \|\nabla f\|^2 d\mu.$$

Remark: We that the SDE above can be explicitly written as: $X_t = e^{-t} X_0 + e^{-t} B_{e^{2t}-1}$, $X_0 \perp W$

Further, if $X_0 \sim N(0,1)$, then $X_t \sim N(0,1)$
 $\forall t \geq 0$.

$$\begin{aligned}\therefore \text{Var}(X_t) &= (e^{-t})^2 + (e^{-t})^2 \cdot (e^{2t} - 1) \\ &= e^{-2t} + 1 - e^{-2t} = 1\end{aligned}$$

Fact:

$$Z \sim N(0,1) \iff E[Z f(Z)] = E[f'(Z)]$$

Stein's
Lemma
(see Note 4a).

$\forall$ cont. bounded f

Remark: We see that by writing $f = g'$, and $U \sim N(0,1)$, the above condition states

$$Z \sim N(0,1) \iff \begin{aligned}E_{Z \sim \mu} [Z g'(Z) - g''(Z)] &= 0, \text{ i.e.,} \\ E_{Z \sim \mu} [\mathcal{L}g(Z)] &= 0, \text{ where}\end{aligned}$$

$\mathcal{L}g(x) = x g'(x) - g''(x)$, which
(as we will show below) is the generator of the
OU process.

This "certification" of a r.v. have dist. μ

via expectations of functions operated on by a generator of a M.P. $(\mathcal{A}, \mathcal{U})$ is very useful for many other applications, and is popularly known as **Stein's method**. (See notes on Stein's method).

Proof: (of Claim 2. to calculate $\mathcal{A}$, $\mathcal{E}(\cdot, \cdot)$ for the OU process)

$$X_t = e^{-t} X_0 + e^{-t} B_{e^{2t}-1}, \quad \forall t \geq 0.$$

$$\text{I.I.} \quad X_s = e^{-s} X_0 + e^{-s} B_{e^{2s}-1}$$

$$\Rightarrow X_t = e^{-(t-s)} X_s + e^{-t} \underbrace{(B_{e^{2t}-1} - B_{e^{2s}-1})}_{\substack{= W \sim N(0, e^{2t}-e^{2s}) \\ \text{and indep. of } X_s}}$$

$$\text{i.e., } X_t = e^{-(t-s)} X_s + e^{-t} \cdot \sqrt{e^{2t}-e^{2s}} \cdot Z \quad \rightarrow N(0, 1)$$

$$\text{i.e., } X_t = e^{-(t-s)} X_s + \sqrt{1 - e^{-2(t-s)}} \cdot Z \quad \rightarrow \textcircled{*}$$

② From $\textcircled{*}$, and $s=0$, it is clear that if $X_0 \sim N(0, 1)$,
 $X_t \sim N(0, 1) \quad \forall t$, i.e., $\mu = N(0, 1)$

$$(b) E[f(x_t) | \mathcal{F}_s]$$

(substitute e^*)

$$= E\left[f\left(e^{-(t-s)} x_s + \sqrt{1-e^{-2(t-s)}} z\right) \mid x_s\right]$$

↪ indep. of x_s .

$$\therefore P_{t-s} f(x) = E\left[f(x_t) \mid x_s = x\right]$$

$$= E_{z \sim N}\left[f\left(e^{-(t-s)} x + \sqrt{1-e^{-2(t-s)}} z\right)\right]$$

(we get (b) by setting $s=0$).

$$(c) \frac{d}{dt} P_t f(x) = E\left[f'\left(e^{-t} x + \sqrt{1-e^{-2t}} z\right) \cdot \left(-e^{-t} x + \frac{(-2)e^{-2t}}{(2)\sqrt{1-e^{-2t}}} z\right)\right]$$

Aside: $g(z) = f'(e^{-t} x + \sqrt{1-e^{-2t}} z)$. Then,

$$E[zg(z)] = E[g'(z)] = E\left[\sqrt{1-e^{-2t}} \cdot f''(e^{-t} x + \sqrt{1-e^{-2t}} z)\right]$$

$$\Rightarrow = E\left[-x e^{-t} f'(e^{-t} x + \sqrt{1-e^{-2t}} z) + e^{-2t} f''(e^{-t} x + \sqrt{1-e^{-2t}} z)\right]$$

(i.e.)

$$\frac{d}{dt} P_t f(x) = \left(-x \frac{d}{dx} + \frac{d^2}{dx^2} \right) P_t f(x).$$

$$\Rightarrow \mathcal{L} f(x) = -x f'(x) + f''(x).$$

$$\textcircled{d} \quad \varepsilon(f, g) = -\langle f, \mathcal{L} g \rangle_\mu$$

$$= E \left[f(z) (z g'(z) - g''(z)) \right]$$

(Aside: $E \left[\underbrace{z f(z)}_{h(z)} \underbrace{g'(z)}_{h'(z)} \right] = E \left[\underbrace{(f(z) g'(z))'}_{h'(z)} \right]$)

$$= E \left[f'(z) g'(z) + f(z) g''(z) \right]$$

$$\hookrightarrow = E \left[f'(z) g'(z) + \cancel{f(z) g''(z)} - \cancel{f(z) g''(z)} \right]$$

$$= E_{\mu \nu z} [f'(z) g'(z)] = \langle f', g' \rangle_\mu$$



Remark: In the vector case, $\mu = N(0, I_d)$
 $X_t \in \mathbb{R}^d$ ↙
d x d matrix

$$\mathcal{L} f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle$$

$$\left(\begin{array}{l} \text{↙ Laplacian;} \\ = \text{Tr}(\nabla^2 f(x)). \end{array} \quad \Delta f(x) = \sum_{i=1}^d \frac{\partial^2 f(x)}{\partial x_i^2} \right)$$

$$\mathcal{E}(f, f) = \|\nabla f\|_{L^2(\mu)}^2.$$

Finally, for future reference, the following are useful:

$$\text{Suppose } \pi^*(x) = \frac{e^{-V(x)}}{Z}, \quad x \in \mathbb{R}^d$$

$$\text{Observe that } \nabla \ln \pi^*(x) = -\nabla V(x).$$

Thus, the **Langevin Dynamics** in this case becomes:

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t, \quad X_t \in \mathbb{R}^d$$

Then, $\mu(x) = \pi^*(x)$, and $f \in \mathcal{F}$,
i.e., $f: \mathbb{R}^d \rightarrow \mathbb{R}$.

$$\Delta f(x) = \Delta f(x) - \langle \nabla V(x), \nabla f(x) \rangle$$

(Laplacian)

$$\text{i.e., } \Delta(f)(x) = \sum_{i=1}^d \frac{\partial^2 f(x)}{\partial x_i^2} - \sum_{i=1}^d \frac{\partial V(x)}{\partial x_i} \cdot \frac{\partial f(x)}{\partial x_i}$$

π a prob. density with domain $\mathbb{R}^d$, i.e., $\pi^*: \mathbb{R}^d \rightarrow \mathbb{R}$

$$\Delta^* \pi(y) = \Delta \pi(y) + \nabla \cdot (\pi(y) \nabla V(y))$$

$$\text{Finally, } \mathcal{E}(f, f) = \int \|\nabla f(x)\|_2^2 d\mu(x).$$

For completeness, recall: $\Delta f(x) = \sum_{i=1}^d \frac{\partial^2 f(x)}{\partial x_i^2}$

$$\nabla \cdot g(y) = \sum_{i=1}^d \frac{\partial}{\partial y_i} g(y).$$

Remark: de Bruijn's Identity, Fisher Information, and KL Divergence

For any Langevin Dynamic, we see above that

$$\mathcal{E}(f, f) = \int \|\nabla f(x)\|_2^2 d\mu(x).$$

We note that the Fisher Divergence (or Fisher Information Distance) is defined as:

$$FI(\mu \parallel \nu) = \int \left\| \nabla_x \ln \frac{d\mu}{d\nu}(x) \right\|^2 d\mu(x).$$

Observe that if $f(x) = \frac{d\mu}{d\nu}$ (the likelihood

ratio, or the Radon-Nikodym derivative), then

$$\mathcal{E}(f, f) = FI(\mu \parallel \nu).$$

Next, we have de Bruijn's Identity:

Let γ_1, γ_2 be r.v.s (with diff. distributions),

and $Z_\theta \sim N(0, \theta^2)$ be indep. of γ_1, γ_2 .

Let $\mu_\theta = \text{dist.}(\gamma_1 + Z_\theta)$, $\nu_\theta = \text{dist.}(\gamma_2 + Z_\theta)$

(i.e., θ^2 -Normal perturbed versions of the dist. of γ_1, γ_2 respectively).

Then,
$$\frac{\partial}{\partial \theta} KL(\mu_\theta \parallel \nu_\theta) = -\frac{1}{2} FI(\mu_\theta \parallel \nu_\theta).$$

(de Bruijn's identity for divergence)

There is also a version that relates entropy and Fisher information:

$$FI(\mu) = E_{x \sim \mu} [\|\nabla_x \ln p_\mu(x)\|^2],$$

where $p_\mu(x)$ is the density associated with measure $\mu(x)$. ($d\mu(x) = p_\mu(x)dx$).

We also have entropy: $E_{x \sim \mu} [-\ln p_\mu(x)]$.

As before, let $\gamma \sim \mu$, and $\gamma + Z_\theta \sim \mu_\theta$,

where $Z_\theta \perp \gamma$, $Z_\theta \sim N(0, \sigma^2)$.

Then,

$$\frac{\partial}{\partial \theta} H(\mu_\theta) = \frac{1}{2} \text{FI}(\mu_\theta)$$

(de Bruijn's identity for Entropy)

Starting from above, we can show the following result
(see [4] below):

Let $X \sim p$ with $E[X] = \mu$, $\text{Var}(X) = \sigma^2$

$Z, W \sim N(\mu, \sigma^2)$ $Z \perp W \perp X$.
 $\sigma_{(\mu, \sigma^2)}$

Further, let $X_t = \sqrt{t} X + \sqrt{1-t} Z$ (Gaussian smoothed X), and $X_t \sim P_t$.

$$KL(p \parallel \sigma_{(\mu, \sigma^2)}) = \sigma^2 \int_0^1 \frac{\text{FI}(P_t \parallel \sigma_{(\mu, \sigma^2)})}{2t} dt$$

We will revisit KL and FI when we discuss the problem of learning score functions through Tweedie's Formula as well as when deriving discretization bounds for DDPM sampling.

References:

4. Ben Chugg, de Bruijn, Relative Entropy and the CLT, online blog, October 2023.

benchugg.com/research-notes/entropy-and-clt.