The Entropic Central Limit Theorem for Discrete Random Variables

Lampros Gavalakis  Ioannis Kontoyiannis

University of Cambridge

ISIT, 27 June 2022
Outline

1. Information Theoretic CLTs
   - The Entropic CLT

2. Discrete Random Variables

3. On Monotonicity

4. Proof
   - Bernoulli Smoothing
Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean 0 and finite variance $\sigma^2$ and let $\hat{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ denote their standardised sum.

Information theory and the central limit theorem (CLT) have a long history starting with [Linnik '59], [Shimizu '75] and [Brown '82].
Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean 0 and finite variance $\sigma^2$ and let $\hat{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ denote their standardised sum.

Information theory and the central limit theorem (CLT) have a long history starting with [Linnik '59], [Shimizu '75] and [Brown '82].

The *Fisher information* of $X$ with density $f$ is $I(X) = \int f \left( \frac{f'}{f} \right)^2$. For $X, Y$ independent

$$I(X + Y) \leq I(X) \quad \text{(convolution inequality)}$$
The differential entropy of $X$ is $h(X) = - \int f \log f$
and the relative entropy between $X$ and $Y$ with densities $f$ and $g$ is
$D(X||Y) = D(f||g) = \int f \log \frac{f}{g}$

Let $Z$ be a 0 mean Gaussian with variance $\sigma^2$. Then

$$D(X) := D(X||Z) = h(Z) - h(X)$$
$$= \frac{1}{2} \log (2\pi e\sigma^2) - h(X)$$

and by positivity of relative entropy

$$h(X) \leq h(Z) \quad \text{(Gaussian maximum entropy)}$$
Theorem (Entropic CLT [Barron, ’86])

Let \( \hat{S}_n \) denote the standardised sum of i.i.d. \( X_1, \ldots, X_n \)

\[
D(\hat{S}_n) \to 0 \quad \text{if and only if} \quad D(\hat{S}_n) \text{ is finite for some } n
\]

Equivalently, writing \( S_n = \sum_{i=1}^{n} X_i \),

\[
h(\hat{S}_n) = h(S_n) - \log \sqrt{n} \to \frac{1}{2} \log (2\pi e \sigma^2)
\]

By Pinsker’s inequality

\[
\|\hat{S}_n - Z\|_{TV} \to 0
\]
Theorem (Entropic CLT [Barron, ’86])

Let $\hat{S}_n$ denote the standardised sum of i.i.d. $X_1, \ldots, X_n$

$$D(\hat{S}_n) \to 0 \text{ if and only if } D(\hat{S}_n) \text{ is finite for some } n$$

Equivalently, writing $S_n = \sum_{i=1}^n X_i$,

$$h(\hat{S}_n) = h(S_n) - \log \sqrt{n} \to \frac{1}{2} \log (2\pi e\sigma^2)$$

By Pinsker’s inequality

$$\|\hat{S}_n - Z\|_{TV} \to 0$$

What about discrete?
The \textit{entropy} of a discrete random variable $Y$ with PMF $p$ on $A$ is

$$H(Y) = - \sum_{y \in A} p(y) \log p(y)$$

$Y$ has a lattice distribution with span $h > 0$ if its support is a subset of

$$\{ a + kh : k \in \mathbb{Z} \}$$

for some $a \in \mathbb{R}$.

$h$ is \textit{maximal} if it is the largest such $h$.

Let $\{X_n\}$ be i.i.d. lattice with variance $\sigma^2$ and maximal span $h$.

Let $S_n = \sum_{i=1}^{n} X_i$

Unlike the differential entropy

$$H\left(\frac{1}{\sqrt{n}} S_n\right) = H(S_n) \to \infty$$
Discrete Random Variables

Theorem (Entropy convergence)

\[
\lim_{n \to \infty} \left[ H(S_n) - \log \frac{\sqrt{n}}{h} \right] = \frac{1}{2} \log(2\pi e \sigma^2)
\]

This expansion has been derived in [Takano '87], [Verdú & Han '97], but using strong forms of the CLT.

In fact, this theorem implies the CLT!
Theorem (Entropy convergence)

\[
\lim_{n \to \infty} \left[ H(S_n) - \log \frac{\sqrt{n}}{h} \right] = \frac{1}{2} \log(2\pi e\sigma^2)
\]

This expansion has been derived in [Takano ’87], [Verdú & Han ’97], ... But using strong forms of the CLT.
**Theorem (Entropy convergence)**

\[
\lim_{n \to \infty} \left[ H(S_n) - \log \frac{\sqrt{n}}{h} \right] = \frac{1}{2} \log(2\pi e\sigma^2)
\]

This expansion has been derived in [Takano ’87], [Verdú & Han ’97], … But using strong forms of the CLT.

In fact, this theorem implies the CLT!
Suppose $Y$ is lattice with PMF $p$, maximal span $h$, values in $A = \{a + kh : k \in \mathbb{Z}\}$, mean $\mu$, and variance $\sigma^2$.

**Definition (Discrete Gaussianity)**

Define

$$D(Y) := D(p \| q) = \sum_{k \in \mathbb{Z}} p(a + kh) \log \frac{p(a + kh)}{q(a + kh)}$$

where $q$ is the PMF of a Gaussian $Z \sim N(\mu, \sigma^2)$ quantised on $A$ as,

$$q(a + kh) = \int_{a + kh}^{a + (k+1)h} \phi(x) \, dx, \quad k \in \mathbb{Z},$$

where $\phi$ is the $N(\mu, \sigma^2)$ density.

By definition, $D(Y + c) = D(Y)$ for any constant $c$. 
Theorem (Discrete entropic CLT)

If \( X_1, X_2, \ldots \) are i.i.d. lattice and \( \hat{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \), then

\[
D(\hat{S}_n) \to 0, \quad \text{as } n \to \infty.
\]

Take WLOG \( \mu = 0 \), let \( Z \sim \mathcal{N}(0, \sigma^2) \) and let \( Z_n \) be the quantised Gaussian. Then, from Pinsker’s and the triangle inequality for the total variation norm

\[
\|\hat{S}_n - Z\|_{TV} \leq \sqrt{\frac{1}{2} D(\hat{S}_n)} + \|Z_n - Z\|_{TV} \to 0 \quad \text{(strong version of CLT)}
\]

Alternatively, \( \|\hat{S}_n - Z_n\|_{TV} \to 0 \)
Theorem (Entropy-relative entropy solidarity)

\[ D(\hat{S}_n) = \frac{1}{2} \log (2\pi e\sigma^2) - \left[ H(S_n) - \log \frac{\sqrt{n}}{h} \right] \\
+ O\left(\frac{1}{\sqrt{n}}\right) \]

By positivity of the relative entropy,

\[ H(S_n) - \log \frac{\sqrt{n}}{h} \leq \frac{1}{2} \log (2\pi e\sigma^2) + O\left(\frac{1}{\sqrt{n}}\right), \]

so the standardised entropy converges to its maximum limit!
On Monotonicity

- **Continuous**
  
  Entropy Power Inequality (EPI):

  \[
  h(X_1 + X_2) \geq h(X_1) + \frac{1}{2} \log 2
  \]

  \[\Rightarrow h(\hat{S}_{2n}) \geq h(\hat{S}_n) \text{ for all } n\]

  In fact, \(h(\hat{S}_n) \uparrow \frac{1}{2} \log(2\pi e\sigma^2)\)
On Monotonicity

- **Continuous**
  Entropy Power Inequality (EPI):

  \[ h(X_1 + X_2) \geq h(X_1) + \frac{1}{2} \log 2 \]

  \[ \Rightarrow h(\hat{S}_{2n}) \geq h(\hat{S}_n) \text{ for all } n \]

  In fact, \( h(\hat{S}_n) \uparrow \frac{1}{2} \log(2\pi e\sigma^2) \)

- **Discrete**
  \( H(X_1 + X_2) \geq H(X_1) + \frac{1}{2} \log 2 \) fails in general.

  However, for i.i.d. \( X_1, X_2 \)

  \[ H(X_1 + X_2) \geq H(X_1) + \frac{1}{2} \log 2 - o_{H(X_1)}(1), \quad [\text{Tao, '10}] \]
Proof

WLOG $h = 1$. Three steps:

1. **Binomial entropy**
   
   If $S_n \sim \text{Bin}(n, 1/2)$, then
   
   $$H(S_n) - \log \sqrt{n} \to 1/2 \log \frac{1}{2} \pi e$$

2. **Bernoulli smoothing**
   
   If $\{V_n\}$ are i.i.d. lattice and $\{B_n\}$ i.i.d. $\text{Bern}(1/2)$ independent,

   $$H_n = \sum_{i=1}^n (V_i + B_i) - \log \sqrt{n} \to 1/2 \log \frac{2}{\pi e} \sigma^2 + 1/4$$

3. **Bernoulli part decomposition**

   $S_n = \sum_{i=1}^n (V_i + W_i)$ for some lattice $V_i$ and $W_i \sim \text{Bern}(q(n))$ with $q(n) \to 1$
Proof

WLOG $h = 1$. Three steps:

1. Binomial entropy
   If $S_n \sim \text{Bin}(n, 1/2)$,
   
   $$H(S_n) - \log \sqrt{n} \to \frac{1}{2} \log \left( \frac{1}{2} \pi e \right)$$
Proof

WLOG $h = 1$. Three steps:

1. **Binomial entropy**
   If $S_n \sim \text{Bin}(n, 1/2)$,
   \[
   H(S_n) - \log \sqrt{n} \to \frac{1}{2} \log \left(\frac{1}{2} \pi e\right)
   \]

2. **“Bernoulli smoothing”**
   If $\{V_n\}$ are i.i.d. lattice and $\{B_n\}$ i.i.d. $\text{Bern}(1/2)$ independent,
   \[
   H\left(\sum_{i=1}^{n} [V_i + B_i]\right) - \log \sqrt{n} \to \frac{1}{2} \log \left(2\pi e \left(\sigma_V^2 + \frac{1}{4}\right)\right)
   \]
Proof

WLOG \( h = 1 \). Three steps:

1. **Binomial entropy**
   If \( S_n \sim \text{Bin}(n, 1/2) \),
   \[
   H(S_n) - \log \sqrt{n} \to \frac{1}{2} \log \left( \frac{1}{2} \pi e \right)
   \]

2. **“Bernoulli smoothing”**
   If \( \{V_n\} \) are i.i.d. lattice and \( \{B_n\} \) i.i.d. \( \text{Bern}(1/2) \) independent,
   \[
   H\left( \sum_{i=1}^{n} [V_i + B_i] \right) - \log \sqrt{n} \to \frac{1}{2} \log \left( 2\pi e \left( \sigma_V^2 + \frac{1}{4} \right) \right)
   \]

3. **Bernoulli part decomposition**
   \[
   S_n \xrightarrow{D} V^{(n)} + W^{(n)} B,
   \]
   for some lattice \( V^{(n)} \), \( W^{(n)} \sim \text{Bern}(q^{(n)}) \) with \( q^{(n)} \to 1 \).
Proof: Bernoulli Smoothing

Lemma

Let $U$ be an independent uniform on $(-1/2, 1/2)$. Then

$$D(\hat{S}_n) = D\left(\hat{S}_n + \frac{1}{\sqrt{n}} U\right) + O\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$. 
Proof: Bernoulli Smoothing

Standardised Fisher information: \( J(X) := \text{Var}(X) I(X) - 1 \)

de Bruijn’s identity: \( D(X) = \int_0^1 J(\sqrt{1 - tX} + \sqrt{tZ}) \frac{dt}{2(1-t)} \)

\( \hat{S}_n = \frac{1}{\sqrt{n}} [\sum_{i=1}^n V_i + B_i] \)
Proof: Bernoulli Smoothing

Standardised Fisher information: $J(X) := \text{Var}(X) I(X) - 1$

de Bruijn’s identity: $D(X) = \int_0^1 J(\sqrt{1-tX} + \sqrt{tZ}) \frac{dt}{2(1-t)}$

$\hat{S}_n = \frac{1}{\sqrt{n}} [\sum_{i=1}^n V_i + B_i]$

$$D \left( \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^n V_i + B_i \right] + \frac{1}{\sqrt{n}} U \right) = D \left( \frac{1}{\sqrt{2n}} [\hat{S}_n + U] + \frac{1}{\sqrt{2}} Z \right)$$

$$+ \int_0^{1/2} J \left( \sqrt{\frac{1-t}{n}} \left[ \sum_{i=1}^n V_i + B_i \right] + \sqrt{\frac{1-t}{n}} U + \sqrt{tZ} \right) \frac{dt}{2(1-t)},$$
Proof: Bernoulli Smoothing

Standardised Fisher information: \( J(X) := \text{Var}(X) I(X) - 1 \)

de Bruijn’s identity: \( D(X) = \int_0^1 J(\sqrt{1 - tX} + \sqrt{tZ}) \frac{dt}{2(1-t)} \)

\( \hat{S}_n = \frac{1}{\sqrt{n}} [\sum_{i=1}^n V_i + B_i] \)

\[
D \left( \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^n V_i + B_i \right] + \frac{1}{\sqrt{n}} U \right) =
D \left( \frac{1}{\sqrt{2n}} \left[ \hat{S}_n + U \right] + \frac{1}{\sqrt{2}} Z \right)
+ \int_0^{1/2} J \left( \sqrt{\frac{1-t}{n}} \left[ \sum_{i=1}^n V_i + B_i \right] + \sqrt{\frac{1-t}{n}} U + \sqrt{t}Z \right) \frac{dt}{2(1-t)},
\]

- **First term** vanishes by the continuous entropic CLT
Proof: Bernoulli Smoothing

Standardised Fisher information: \( J(X) := \text{Var}(X) I(X) - 1 \)

de Bruijn’s identity: \( D(X) = \int_0^1 J(\sqrt{1-tX} + \sqrt{tZ}) \frac{dt}{2(1-t)} \)

\( \hat{S}_n = \frac{1}{\sqrt{n}} [\sum_{i=1}^n V_i + B_i] \)

\[
D \left( \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^n \sqrt{n} V_i + B_i \right] + \frac{1}{\sqrt{n}} U \right) = D \left( \frac{1}{\sqrt{2n}} \left[ \hat{S}_n + U \right] + \frac{1}{\sqrt{2}} Z \right) + \int_0^{1/2} J \left( \sqrt{\frac{1-t}{n}} \left[ \sum_{i=1}^n V_i + B_i \right] + \sqrt{\frac{1-t}{n}} U + \sqrt{tZ} \right) \frac{dt}{2(1-t)} ,
\]

- First term vanishes by the continuous entropic CLT
- The integrand vanishes for each fixed \( t \in (0, 1) \) by the results of [Barron, ’86] and, by the convolution inequality, is

\[
\leq \left( 1 + \frac{\sigma^2}{\sigma^2} \right) J \left( \sqrt{\frac{1-t}{n}} \sum_{i=1}^n B_i + \sqrt{\frac{1-t}{n}} U + \sqrt{tZ'} \right) + \frac{\sigma^2}{\sigma^2} ,
\]

whose integral vanishes by the binomial case (Step 1)!
Proof: Bernoulli Smoothing

Standardised Fisher information: \( J(X) := \text{Var}(X)I(X) - 1 \)

de Bruijn’s identity: \( D(X) = \int_0^1 J(\sqrt{1-t}X + \sqrt{t}Z) \frac{dt}{2(1-t)} \)

\( \hat{S}_n = \frac{1}{\sqrt{n}} [\sum_{i=1}^{n} V_i + B_i] \)

\[
D \left( \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^{n} V_i + B_i \right] + \frac{1}{\sqrt{n}} U \right) = D \left( \frac{1}{\sqrt{2n}} [\hat{S}_n + U] + \frac{1}{\sqrt{2}} Z \right) \\
+ \int_0^{1/2} J \left( \sqrt{\frac{1-t}{n}} \left[ \sum_{i=1}^{n} V_i + B_i \right] + \sqrt{\frac{1-t}{n}} U + \sqrt{t}Z \right) \frac{dt}{2(1-t)},
\]

- **First term** vanishes by the continuous entropic CLT
- **The integrand** vanishes for each fixed \( t \in (0, 1) \) by the results of [Barron, ’86] and, by the convolution inequality, is

\[
\leq \left( 1 + \frac{\sigma^2}{\sigma^2} \right) J \left( \sqrt{\frac{1-t}{n}} \sum_{i=1}^{n} B_i + \sqrt{\frac{1-t}{n}} U + \sqrt{t}Z' \right) + \frac{\sigma^2}{\sigma^2}, \text{ whose integral vanishes by the binomial case (Step 1)!}
\]

\[ \Rightarrow \text{Uniform integrability} \]
Further Work

- Non-lattice
- Rates of convergence under additional moment assumptions
- (Approximate) Monotonicity (of any of the quantities appearing in the proof)
- Dependent random variables
- Random vectors
Thank you!