Local CLT, Stirling's formula

Daniel Rui 10/12/25

I assume the reader already knows/[believes in] the proof of the standard CLT. See e.g. TT's 275A course notes using Lindeburg exchange. It is a substantial proof, but I think presentable in roughly 1-1.5 hours of lecture, so not too long. Not much prerequisite material is needed, besides the basic definitions of probability theory, the "weak* definition" of convergence in distribution (Def. 2 in TT, which is equivalent to the c.d.f. definition by Prop. 4), and Taylor expansion.

The local (discrete) CLT proof in TT Notes 5 is trickier, using Fourier/characteristic function methods. I would like to use this article to explain Burgess Davis' elementary proof of a local discrete CLT, in the context of a simple application of it towards a probabilistic proof of Stirling's formula for the factorial.

ex:stirling

Example 0.10: Stirling

Let $X_1, X_2, \dots$ be i.i.d. Poisson(1) r.v.'s (i.e. valued in $\mathbb{N}$, where $\mathbb{P}[X = k] = f_X(k) := \frac{e^{-1}}{k!}$).
Recall (or calculate for yourself) $\mathbb{E}[X_i] = 1, \mathrm{Var}[X_i] = 1$.
Let $S_n := X_1 + \ldots + X_n$. A probabilistic argument tells us that $S_n \sim \mathrm{Poisson}(n)$, so in ptclr., $\mathbb{E}[S_n] = n, \mathrm{Var}[S_n] = n$.
CLT tells us $\frac{S_n - n}{\sqrt{n}} \to_d \mathrm{Normal}(0,1)$, or equivalently (for all $t \in \mathbb{R}$)

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n}{\sqrt{n}} \leqslant t\right] = \Phi(t) := \int_{-\infty}^{t} \phi(x) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

For the sake of relating to the writeups of McDonald and/or Davis, I introduce the notation $a_n := n$, $b_n := \sqrt{n}$ so that $\frac{S_n - a_n}{b_n} := \frac{S_n - n}{\sqrt{n}}$.

Intuitively, this says that

$$\mathbb{P}[S_n \leqslant \underbrace{b_n t + a_n}_{:=k}] \approx \Phi(\underbrace{t}_{=\frac{k-a_n}{b_n}})$$

so taking the "derivative in k" (use FTC2 on Φ) heuristically gives $\mathbb{P}[S_n = k] \approx \frac{1}{b_n}\phi(\frac{k-a_n}{b_n})$.

Indeed, we aim to show the following "local discrete CLT"

$$\sup_{k\in\mathbb{Z}} \left| \underbrace{b_n}_{:=\sqrt{n}} \cdot \underbrace{\mathbb{P}[S_n = k]}_{=\frac{e^{-n}n^k}{k!}} - \phi\Big(\underbrace{\frac{k - a_n}{b_n}}_{:=\frac{k-n}{\sqrt{n}}} \Big) \right| \xrightarrow{n\to\infty} 0$$

In particular, taking $k = n$, this implies that

$$\left| \sqrt{n} \cdot \frac{e^{-n} n^n}{n!} - \frac{1}{\sqrt{2\pi}} \right| \xrightarrow{n\to\infty} 0,$$

or equivalently $\frac{\sqrt{2\pi n}\cdot(\frac{n}{e})^n}{n!} = 1 + o_{n\to\infty}(1) \iff n! = \sqrt{2\pi n} \cdot (\frac{n}{e})^n \cdot (1 + o_{n\to\infty}(1))$.

Now, we prove the above "local discrete CLT", following McDonald.
Define variation "norm" on functions $f : \mathbb{Z} \to [0, \infty)$ by $\|f\|_V := \sup_{k\in\mathbb{Z}} |f(k) - f(k+1)|$. Define $\underline{f}$

as the function $k \mapsto \min(f(k), f(k+1))$ — observe that $\underline{f} \leqslant f$ pointwise on $\mathbb{Z}$. The "f" we deal with will be p.m.f.'s of our random variables X_i (f_{X_i} doesn't depend on i because we assumed i.i.d.), S_n.

lem:varLemma

Lemma 0.11: Variation norm condition to discrete LCLT

[Lemma 3.1 in McDonald, Lemma 1 in Davis]
If $a_n, b_n \geqslant 0$ are as above ($\frac{S_n - a_n}{b_n}$ satisfy CLT, etc.) and $b_n \to \infty$ (as $n \to \infty$), and

$$\sup_n \left[b_n^2 \cdot \|f_{S_n}\|_V \right] < \infty, \quad \text{(e.g. } \sup_n \left[n \cdot \sup_k \left| \frac{e^{-n} n^k}{k!} - \frac{e^{-n} n^{k+1}}{(k+1)!} \right| \right] < \infty \text{)}$$

then the LCLT result holds:

$$\sup_{k \in \mathbb{Z}} \left| b_n \cdot \mathbb{P}[S_n = k] \; - \; \phi\Big(\frac{k - a_n}{b_n} \Big) \right| \xrightarrow{n \to \infty} 0$$

Proof:
define $\tilde{\phi}(\frac{k-a_n}{b_n}) \coloneqq b_n \cdot \big(\Phi(\frac{(k+1)-a_n}{b_n}) - \Phi(\frac{k-a_n}{b_n}) \big)$ (named $\tilde{\phi}$ because the RHS is a Newton quotient of Φ, hence close to $\Phi' = \phi$).
In Davis' notation, $e_{k,n} \coloneqq \frac{k-a_n}{b_n}, \tilde{\phi}(\frac{k-a_n}{b_n}) = b_n \cdot h_{k,n}$.
Davis (1) notes

$$\sup_{k \in \mathbb{Z}} |\tilde{\phi}(\tfrac{k-a_n}{b_n}) - \phi(\tfrac{k-a_n}{b_n})| \xrightarrow{n \to \infty} 0.$$

(this only needs $b_n \to \infty$; proof by mean value theorem or FTC).
Consider deviation $\Delta_n(k) \coloneqq b_n \cdot \mathbb{P}[S_n = k] - \tilde{\phi}(\frac{k-a_n}{b_n})$. Recall we WTS $\|\Delta_n(k)\|_{\ell^\infty_k(\mathbb{Z})} \xrightarrow{n \to \infty} 0$.

We consider successive difference $|\Delta_n(k+1) - \Delta_n(k)|$. Triangle inequality splits into 2 terms, and $\| \bullet \|_V$ assumption bounds the $\mathbb{P}[\bullet]$ term, but we need something for the $\tilde{\phi}$ term.
For any fixed $k \in \mathbb{Z}$, Davis' (6) claims

$$\sup_{n \in \mathbb{Z}} \left[b_n \cdot |\tilde{\phi}(\tfrac{(k+1)-a_n}{b_n}) - \tilde{\phi}(\tfrac{k-a_n}{b_n})| \right] < \infty.$$

(intuition is that it is like Newton quotient for ϕ). More rigorously, it is 2nd order Newton quotient for Φ, so 2nd order MVT (or FTC) will give a bound in terms of Φ'', which is uniformly bounded on $\mathbb{R}$.
Putting all the bounds of this textblock together, we get

$$b_n \cdot \; |\Delta_n(k+1) - \Delta_n(k)| \leqslant b_n^2 \|f_{S_n}\|_V + b_n |\tilde{\phi}(\tfrac{(k+1)-a_n}{b_n}) - \tilde{\phi}(\tfrac{k-a_n}{b_n})|,$$

which is uniformly (in $n \in \mathbb{Z}$) bounded $\leqslant C < \infty$.
Thus, $|\Delta_n(j) - \Delta_n(k)| \leqslant C|j-k| \cdot \frac{1}{b_n}$, for all $n, j, k \in \mathbb{Z}$ (cf. Davis' (7)).

Fix $\epsilon > 0$. Fix $k \in \mathbb{Z}$.
We show that $|\Delta_n(k)| > \epsilon$ can happen for only finitely many n.
Well, suppose we do have an $n^\star$ s.t. $\Delta_{n^\star}(k) > \epsilon$ (the $< -\epsilon$ case is handled similarly). Then by Davis' (7), $\Delta_{n^\star}(j) > \frac{\epsilon}{2}$ for all j s.t. $C|j-k| \leqslant \frac{\epsilon}{2} \cdot b_{n^\star}$. Define J to be this range of j. The number of such j, i.e. $|J|$, is $\geqslant \frac{\epsilon}{2C} \cdot b_{n^\star}$.
Summing $\sum_{j \in J} \Delta_{n^\star}(j)$ over this (size $\sim b_n$ 🤯) range of j yields a value $> (\frac{\epsilon}{2C} b_{n^\star}) \cdot (\frac{\epsilon}{2})$.

But

$$\sum_{j\in J}\Delta_n(j) = b_n\left(\sum_{j\in J}\mathbb{P}[S_n = j] - \sum_{j\in J}\tfrac{1}{b_n}\tilde{\phi}\big(\tfrac{j-a_n}{b_n}\big)\right)$$

J is long enough ($|J| \sim b_n$ 🤯) that the rightmost sum is essentially a Riemann sum converging to $\int_{\text{an interval}} \phi(x)\,dx$, so within the regime of CLT.

So, if there is an $n^\star$ as above, then (dividing by $b_{n^\star}$ everywhere)

$$\left(\sum_{j\in J}\mathbb{P}[S_{n^\star} = j] - \sum_{j\in J}\tfrac{1}{b_n}\tilde{\phi}\big(\tfrac{j-a_n}{b_n}\big)\right) > \frac{\epsilon^2}{4C} > 0$$

which CLT tells us can only be true for finitely many $n^\star$. ■

Ok, now just need some way to check the weird $\|\bullet\|_V$ assumption to apply the Lemma to our Stirling example.

McDonald pg. 10-11 takes it from here, using the idea of Bernoulli decomposition, and one simple Lemma on successive difference of normalized binomial coefficients $\binom{n}{k}/2^n$ (on the same row of Pascal's triangle).