Elementary Proof of the CLT
Sketch
L. Brown, 5/24/12 rev 11/15/12

I first presented a version of this proof in math 470 at Cornell University in the late 60s. I've occasionally presented versions in probability courses since then. The most recently that I can remember doing this is the last time I taught stat 430 here at UPenn, which was about 10 years ago. The following is a very condensed account, but with added material in **4**, below, and with minor improvements to other parts of the argument. It is meant as a reminder of the ideas rather than a full classroom presentation. I have transcribed this material now at the urging of G. Dumais who heard of it via George Casella who was my treasured friend and colleague over decades together at Rutgers and Cornell.

Let X denote a random variable with mean and variance $\mu(X), \sigma^2(X)$, resp. Write $X \in \mathbb{N}$ if X satisfies the CLT; i.e., for iid copies $X_1,..,X_n$

$$\frac{\sum_1^n \left(X_i - \mu(X)\right)}{\sigma(X)\sqrt{n}} \to \Phi \text{ in distribution.}$$

As a convention, if X is a constant then also write $X \in \mathbb{N}$.

1. If $X \in \mathbb{N}$ so is $aX + b$.

2. If X, Y are indep normal variables then $X + Y \in \mathbb{N}$.

3. If $X, Y \in \mathbb{N}$ and X, Y are indep then $X + Y \in \mathbb{N}$.
Proof: Rearrange sums and use **2**.||

4. $B = \text{Bern}(1/2)$ satisfies $B \in \mathbb{N}$.
Proof: This is the simplest Bernoulli CLT. It can be proved using Stirling's formula. See e.g., *Feller*, *v.1*.||
Note: With some extra computations one can also use Stirling's formula to prove that $B \in \mathbb{N}$ for any Bernoulli B. But this isn't needed at this stage.
Alternate Proof:
Step 1. (CLT for the median of uniforms.) Let $U_1,..,U_n$ be iid Unif[0,1] random variables. Assume n is an odd number (for convenience); modification of the argument for even n is straightforward. Let M_n denote their median. Let $B_n = \#\{i : U_i \le 1/2\}$. Note that $B_n \sim Bin(n,1/2)$.

We know (or can derive) that M_n has a beta distribution. What matters is that its density is of the form

$$f_{M_n}(m) = C_n m^{[n/2]} (1-m)^{[n/2]}.$$

Let $W_n = (M_n - 1/2)\sqrt{n}$. Then

(4.1)
$$
\begin{aligned}
f_{W_n}(w) &= \frac{C_n}{\sqrt{n}}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)^{[n/2]}\left(\frac{1}{2}-\frac{w}{\sqrt{n}}\right)^{[n/2]} = \frac{C_n}{2^{n-1}\sqrt{n}}\left(1+\frac{2w}{\sqrt{n}}\right)^{[n/2]}\left(1-\frac{2w}{\sqrt{n}}\right)^{[n/2]} \\
&= \frac{C_n}{2^{n-1}\sqrt{n}}\left(1-\frac{4w^2}{n}\right)^{[n/2]} \sim \frac{C_n}{2^{n-1}\sqrt{n}}e^{-2w^2}
\end{aligned}
$$

This shows that the limiting distribution of W_n is $N(0,1/4)$ since this has the form of a normal density with variance $1/4$. [It follows that $\frac{C_n}{2^{n-1}\sqrt{n}} \to \frac{2}{\sqrt{2\pi}}$.]

Step 2. (Limiting distribution of the median is the same as that of Bin(n, ½).)

For this paragraph suppose $M_n > 1/2$ and condition on M_n. The conditional distribution of U_i given $U_i < M_n$ is uniform on $(0, M_n)$ with all such U_i being (conditionally) independent. We then have $P(U_i > 1/2 | U_i < M_n) = (M_n - 1/2)/M_n \sim 2W_n/\sqrt{n}$. There are $[n/2]$ values of U_i satisfying $U_i < M_n$. Hence

(4.2)
$$\Delta_n @ \left(\left[\frac{n}{2}\right] - B_n\right) \sim \text{Bin}\left(\left[\frac{n}{2}\right], (M_n - 1/2)/M_n\right).$$

It's easily computed that $E(\Delta_n) \sim W_n\sqrt{n}$ and also $\text{var}(\Delta_n) \sim W_n\sqrt{n}$. Invoking Chebyshev's inequality yields

$$\frac{\Delta_n}{\sqrt{n}} - W_n \to 0 \text{ (in probability)}.$$

A similar argument holds when $M_n < 1/2$. It follows that

(4.3)
$$\frac{B_n - [n/2]}{\sqrt{n}} - W_n \to 0 \text{ (in probability)}$$

and consequently $(B_n - [n/2])/\sqrt{n} \to N(0, 1/4)$ in distribution, by (4.1).

This shows that $\text{Bern}(1/2) \in \mathbb{N}$ and hence completes a proof of the DeMoivre Central Limit Theorem.

5. Let $X, Y, B \in \mathbb{N}$, all independent, with $B = \text{Bern}(p)$. Let $W = BX + (1-B)Y$. Then $Z \in \mathbb{N}$.

Proof: Assume $0 = \mu(W) = p\mu(X) + (1-p)\mu(Y)$. Because of #1 there is no loss of generality.

Let $N_n = \sum_1^n B_i$. Then (after some algebraic manipulation)

$$\frac{\sum_1^n W_i}{\sqrt{n}} =_{Dist} \frac{\sum_1^{N_n} X_i + \sum_1^{n-N_n} Y_i}{\sqrt{n}}$$

(5.1) $$= \sqrt{\frac{N_n}{n}} \frac{\sum_1^{N_n}\left(X_i - \mu(X)\right)}{\sqrt{N_n}} + \sqrt{\frac{n-N_n}{n}} \frac{\sum_1^{n-N_n}\left(Y_i - \mu(Y)\right)}{\sqrt{N_n}} + \left(\mu(X) - \mu(Y)\right)\frac{N_n - p_n}{\sqrt{n}}$$

$$\to_{Dist} \left[pZ_1 + (1-p)Z_2 + Z_3 \right]$$

where Z_1, Z_2, Z_3 are independent and $Z_{1[2]}:\ N\left(0, \sigma^2\left(X[Y]\right)\right),\ Z_3 \sim N\left(0, \left(\mu(X) - \mu(Y)\right)^2\right)$. ||

6. If D is a discrete random variable of the form $P\left(D = d_j\right) = k_j / 2^K$, $j = 1,..,J \le 2^K$ then $D \in \mathbb{N}$.

Proof: This can be established by induction on *K*, using **4**, **5**. (Note that in the above one can take $k_j \equiv 1$ and allow repeated values of d_j.) ||

In particular, this shows that $\text{Bern}(p) \in \mathbb{N}$ whenever *p* is a dyadic rational (i.e., of the form $p = k/2^K$ for any *k*, *K*).

7. (CLT for continuous r.v.[s] -- i.e., if *X* is any continuous r.v. with finite mean and variance then $X \in \mathbb{N}$.)

Proof: Partition the real line as $-\infty = a_0 < a_1 < ... < a_L = \infty$ with $L = 2^K$ and with

$$P\left(a_{j-1} < X < a_j\right) = 1/L.$$

Let $b_j = E\left(X \middle| a_{j-1} < X < a_j\right)$. Let D^L denote the discrete r.v. with $P\left(D^L = b_j\right) = 1/L$. Then $\mu(X) = \mu(D)$ and in a natural fashion,

$$X = D^L + \varepsilon^L$$

where $\mu\left(\varepsilon^L\right) = 0$, $\sigma^2\left(\varepsilon^L\right) \to 0$ as $L \to \infty$. Of course, D^L, ε^L are generally not independent. [In modern terminology, I've created a "coupling" representation for *X*.]

Then,

$$T_n^L \triangleq \frac{\sum\left(X_i - \mu(X)\right)}{\sqrt{n}} = \frac{\sum\left(D_i^L - \mu\left(D^L\right)\right)}{\sqrt{n}} + \frac{\sum \varepsilon_i^L}{\sqrt{n}} \triangleq Z_n^L + E_n^L.$$

Here Z_n^L, E_n^L are not necessarily independent. It follows from **6** that $Z_n^L \to N\left(0, \sigma^2\left(D^L\right)\right)$ in distribution as $n \to \infty$. Also $\sigma^2\left(E_n^L\right) = \sigma^2\left(\varepsilon^L\right) \to 0$ and $\sigma^2\left(D^L\right) \to \sigma^2(X)$ as $L \to \infty$. And, $\mu\left(E_n^L\right) = 0$.

Now let $\alpha > 0$ and write,

(8.1) $$P\left(Z_n^L \le c - \alpha\right) - P\left(E_n^L \ge \alpha\right) \le P\left(T_n^L \le c\right) \le P\left(Z_n^L \le c + \alpha\right) + P\left(E_n^L \le -\alpha\right).$$

Then apply Chebyshev's inequality to get that the left and right sides of (8.1) satisfy

$$P\left(Z_n^L \le c-\alpha\right) - P\left(E_n^L \ge \alpha\right) \ge P\left(Z_n^L \le c-\alpha\right) - \sigma\left(\varepsilon^L\right)\big/\alpha$$
$$P\left(Z_n^L \le c+\alpha\right) + P\left(E_n^L \le -\alpha\right) \le P\left(Z_n^L \le c+\alpha\right) + \sigma\left(\varepsilon^L\right)\big/\alpha \, .$$

Now let $n \to \infty$ and then $L \to \infty$ and then $\alpha \to 0$ to get the desired result that $X \in \bullet$. ||

9. (Extension of CLT to all distributions having finite mean and variance.)
Let X be any variable with finite variance. Let $\varepsilon > 0$. Let $Z : N(0,1)$, independent of X. Let $W = X + \varepsilon Z$. Then, $W \in \mathbb{N}$ by **8**. Hence $n^{-1/2}\sum_1^n (W_i - \mu) + n^{-1/2}\sum_1^n \varepsilon Z_i \to N\left(0, \sigma^2 + \varepsilon^2\right)$. Letting $n \to \infty$ and then $\varepsilon \to 0$ yields the desired result that $X \in \mathbb{N}$, since $\operatorname{var}\left(n^{-1/2}\sum_1^n \varepsilon Z_i\right) = \varepsilon^2 \to 0$ as $\varepsilon \to 0$.