The central limit theorem

Here is a proof of the central limit theorem, in a reasonably strong form. This is Lindeberg's proof, as presented by Terrence Tao in his notes (and made more concrete by specifying $G(x)$). Recall that the cumulative distribution function of the standard normal distribution is denoted by $\Phi(x)$.

Theorem 1 *Let $X_1, \dots, X_n$ be independent standard random variables, where there is also a constant c with $\mathbf{E}|X_i|^3 < c$, let $S_n = \sum_i X_i$, then*

$$\lim_{n\to\infty} \mathbf{P}\{S_n/\sqrt{n} < a\} = \Phi(a).$$

Proof. 1. Let Y_1, Y_2 be independent normal variables: then $Y_1 + Y_2$ is also normal. The proof of this is exercise. Let $Y_1, \dots, Y_n$ are independent standard normal variables, $T_n = \sum_i Y_i$. It follows from the above that $T_n/\sqrt{n}$ is standard normal.

2. Let $G(x)$ be any function defined on $(-\infty, \infty)$ that has continuous and bounded first, second and third derivatives: the third derivatives are bound by some parameter B. We will show that

$$\mathbf{E}G\big(S_n/\sqrt{n}\big) - \mathbf{E}G\big(T_n/\sqrt{n}\big) = O(Bn^{-1/2}). \tag{1}$$

We can assume that all the variables X_i, Y_j are over a common probability space, and the group of Y_j is also independent of the group of X_i. (One can always construct such a probability space.) We will show that exchanging one-by-one the X_i's in the sum for the Y_i's changes the expected value only little: namely

$$\mathbf{E}G(Z + X_i/\sqrt{n}) - \mathbf{E}G(Z + Y_i/\sqrt{n}) = O(Bn^{-3/2}),$$

where Z, X_i, Y_i are independent. Let us expand G into a Taylor series, using the fact that $|G'''(z)| \le B$:

$$G(Z + \delta) = G(Z) + G'(Z)\delta + G''(Z)\delta^2/2 + B\delta^3/6.$$

Substituting, $\delta = X_i/\sqrt{n}$ and $\delta = Y_i/\sqrt{n}$ we get

$$\begin{aligned} &G(Z + X_i/\sqrt{n}) - G(Z + Y_i/\sqrt{n}) \\ &\quad = G'(Z)(X_i - Y_i)n^{-1/2} + G''(Z)(X_i^2 - Y_i^2)n^{-1} + B \cdot O(|X_i|^3 + |Y_i|^3)n^{-3/2}. \end{aligned}$$

By definition, $\mathbf{E}X_i = \mathbf{E}Y_i$ and $\mathbf{E}X_i^2 = \mathbf{E}Y_i^2$. Hence, using the independence of Z from X_i, Y_i:

$$\begin{aligned}
\mathbf{E}G'(Z)(X_i - Y_i) &= \mathbf{E}G'(Z)\mathbf{E}(X_i - Y_i) = 0,\\
\mathbf{E}G''(Z)(X_i^2 - Y_i^2) &= \mathbf{E}G''(Z)\mathbf{E}(X_i^2 - Y_i^2) = 0,\\
\mathbf{E}G(Z + X_i/\sqrt{n}) &- \mathbf{E}G(Z + Y_i/\sqrt{n})\\
&= Bn^{-3/2}O(\mathbf{E}|X_i|^3 + \mathbf{E}|Y_i|^3) = O(Bn^{-3/2}),
\end{aligned}$$

since the third moments of X_i and Y_i are finite. This allows us to replace X_1 with Y_1, X_2 with Y_2 and so on one-by-one until S_n is replaced with T_n. The total price for the replacement is still $O(Bn^{-1/2})$, proving (1).

3. Below we will show a function $G(x)$ with the property that $0 \le G(x) \le 1$, $G(x) = 1$ for $x \le 0$, $G(x) = 0$ for $x \ge 1$, further it has three continous derivatives. Then $G_{a,\varepsilon}(x) = G(x/\varepsilon - a)$ sinks from 1 to 0 between a and $a + \varepsilon$, and $G'''_{a,\varepsilon}(x) = O(\varepsilon^{-3})$, so ε^{-3} plays the role of B above. For any random variable X we have then

$$\begin{aligned}
\mathbf{P}\{X < a\} \le \mathbf{E}G_{a,\varepsilon}(X) \le \mathbf{P}\{X < a + \varepsilon\},\\
\mathbf{E}G_{a-\varepsilon,\varepsilon}(X) \le \mathbf{P}\{X < a\} \le \mathbf{E}G_{a,\varepsilon}(X).
\end{aligned} \tag{2}$$

We found, since the bound B is now $O(\varepsilon^{-3})$, that

$$|\mathbf{E}G_{a,\varepsilon}(S_n/\sqrt{n}) - \mathbf{E}G_{a,\varepsilon}(T_n/\sqrt{n})| = O(\varepsilon^{-3}n^{-1/2}).$$

Applying (2) to both $S_n/\sqrt{n}$ and $T_n/\sqrt{n}$ and recalling that $\mathbf{P}\{T_n/\sqrt{n} < a\} = \Phi(a)$:

$$\Phi(a - \varepsilon) - O(\varepsilon^{-3}n^{-1/2}) \le \mathbf{P}\{S_n/\sqrt{n} < a\} \le \Phi(a + \varepsilon) + O(\varepsilon^{-3}n^{-1/2}).$$

Since the derivative of $\Phi(a)$ is bounded (by $1/\sqrt{2\pi}$):

$$|\,\mathbf{P}\{S_n/\sqrt{n} < a\} - \Phi(a)| = O(\varepsilon + \varepsilon^{-3}n^{-1/2}).$$

Choosing $\varepsilon = n^{-1/8}$ this becomes $O(n^{-1/8})$.

4. The function $F(x) = \cos(x^2(\pi/2 - x)^2)$ sinks from 1 to 0 between 0 and $\pi/2$, its three derivatives are continuous and are zero at $x = 0$ and $x = \pi/2$ (as can be checked by direct computation). Define the function $G(x)$ as $F(\frac{\pi}{2}x)$ between 0 and 1, further 1 for $x < 0$ and 0 for $x > 1$. Then it has the properties desired in point 3 above.

□