A few notes on the Central Limit Theorem

One form of the Central Limit Theorem states that if random variables $X_1, X_2, \ldots X_n$ are **independent and identically distributed**, as $n \to \infty$, the **sample mean** $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ approaches the normal distribution $N(\mu, \sigma^2/n)$. Note that μ, σ^2 are the mean/variance of each individual X_i.

Why is the variance σ^2/n? By the properties of variance and independence,

$$\begin{aligned} Var(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)) &= \tfrac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) \\ &= \tfrac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n)) \\ &= \tfrac{1}{n^2}(n \cdot \sigma^2) = \frac{\sigma^2}{n} \end{aligned}$$

However, in many applications, you will be asked to approximate the **sum** of the random variables $X_1 + X_2 + \cdots + X_n$. By the CLT, if these variables are independent and identically distributed, this sum is approximately normal for large n. Since we're using the normal distribution to approximate the sum, we want the mean and variance of the Normal to be the same as the sum's, so the sum is approximated by $N(n\mu, n\sigma^2)$.

If you standardize this (by subtracting the mean and then dividing by the standard deviation of the overall sum), you'll get the formula that was shown in class:

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \to N(0, 1)$$

There's one important catch. The normal distribution can take on any real value. However, if you're using the normal distribution to approximate anything where the random variable can only be an integer (such as the binomial), you'll have to apply the **continuity correction**. For example, let's say we want to find $P(X > 5)$, and X is discrete. Then, what we really want is for X to be 6, 7, 8, etc. So when using the normal distribution to approximate this, we want to find the probability that the value returned by the normal distribution **rounds to** 6, 7, 8, … so we should calculate $P(X > 5.5)$.

To summarize, here are the steps you should take for Central Limit Theorem problems.

1. Check that all random variables are independent and identically distributed.
2. Find the mean and variance of the normal distribution that should be used. Set the mean/variance of this normal distribution to be the same mean and variance as the sum or sample mean.
3. Apply the continuity correction, if we're approximating a discrete distribution.
4. Standardize the random variable: subtract the mean from both sides and divide both sides by the standard deviation of the normal distribution.
5. Look up the value in the Φ table.