Math 302.102 Fall 2010
The Central Limit Theorem

As you saw in preparation for Midterm #2, if we know that $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, then the distribution of

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

can be determined using moment generating functions. In fact,

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Moreover, by *normalizing* we conclude that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

As we will now see, the special case of normal random variables is an idealized version of the Central Limit Theorem. Suppose, therefore, that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with common mean μ and common variance σ^2, and let

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Without knowing the common distribution of $X_1, X_2, \ldots, X_n$, it is not possible to determine the exact distribution of $\overline{X}$. However, we can conclude that

$$\begin{aligned}\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)}{n} &= \frac{\mu + \mu + \cdots + \mu}{n} \\ &= \frac{n\mu}{n} = \mu,\end{aligned}$$

and using the fact that $X_1, X_2, \ldots, X_n$ are independent,

$$\begin{aligned}\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) &= \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)}{n^2} \\ &= \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.\end{aligned}$$

Therefore,

$$\mathbb{E}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = 0 \quad \text{and} \quad \mathrm{Var}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = 1.$$

Now, even though we cannot determine the *exact* distribution of

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

without knowledge of the common distribution of $X_1, X_2, \ldots, X_n$, we will see that it is possible to determine its *approximate* distribution.

Observe that

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{n\overline{X}-n\mu}{\sigma\sqrt{n}} = \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sigma\sqrt{n}}$$
$$= \frac{X_1-\mu}{\sigma\sqrt{n}} + \frac{X_2-\mu}{\sigma\sqrt{n}} + \cdots + \frac{X_n-\mu}{\sigma\sqrt{n}}. \qquad (*)$$

We now write

$$Y = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \quad \text{and} \quad Y_j = \frac{X_j-\mu}{\sigma\sqrt{n}}$$

so that $Y_1, Y_2, \ldots, Y_n$ are iid and $(*)$ is equivalent to

$$Y = Y_1 + \cdots + Y_n.$$

We can determine the moment generating function of Y in terms of the moment generating functions of Y_j. That is,

$$m_Y(t) = \mathbb{E}(e^{tY}) = \mathbb{E}[e^{t(Y_1+\cdots+Y_n)}] = \mathbb{E}(e^{tY_1}\cdots e^{tY_n}) = \mathbb{E}(e^{tY_1})\cdots\mathbb{E}(e^{tY_n}) = [\mathbb{E}(e^{tY_1})]^n \qquad (**)$$

using the fact that $Y_1, \ldots, Y_n$ are iid. The next step is to approximate $\mathbb{E}(e^{tY_1})$. The basic idea is to write out the power series expansion for e^a and take expectations. That is,

$$e^a = 1 + a + \frac{a^2}{2!} + \cdots$$

and so

$$\mathbb{E}(e^{tY_1}) = \mathbb{E}\left[1 + tY_1 + \frac{t^2Y_1^2}{2!} + \cdots\right] = 1 + t\mathbb{E}(Y_1) + \frac{t^2\mathbb{E}(Y_1^2)}{2!} + \cdots$$
$$= 1 + \frac{t^2}{2n} + \cdots$$

since $\mathbb{E}(Y_1) = 0$ and $\text{Var}(Y_1) = \mathbb{E}(Y_1^2) = 1/n$. Hence, we see from $(**)$ that

$$m_Y(t) = [\mathbb{E}(e^{tY_1})]^n = \left[1 + \frac{t^2}{2n} + \cdots\right]^n$$

But the quantity on the right side above just happens to look like the limit definition of e. That is,

$$\lim_{n\to\infty} m_Y(t) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \cdots\right]^n = e^{t^2/2}$$

which just so happens to be the moment generating function of a $\mathcal{N}(0,1)$ random variable.

Theorem (Central Limit Theorem)**.** *If $X_1, X_2, \ldots, X_n$ are independent and identically distributed with common mean μ and common variance σ^2, then the limiting distribution of*

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

is $\mathcal{N}(0,1)$. That is,

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in A\right\} = \frac{1}{\sqrt{2\pi}}\int_A e^{-x^2/2}\,\mathrm{d}x.$$

Surprise! Wikipedia has an article titled *Illustration of the Central Limit Theorem.*
`http://en.wikipedia.org/wiki/Illustration_of_the_central_limit_theorem`