Handout 15: Normal Distribution

We write $X \sim N(\mu, \sigma^2)$ to indicate that X is a random variable with a normal distribution having mean μ and variance σ^2.[1] The pdf is given by:

$$f_X(x) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] \times \left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]$$

[1] I have split the density into two parts. The one on the left is the normalizing constant. This is fixed for any choice of the parameters μ and σ^2 with respect to x. The shape of the distribution is set by the part on the right.

In the special case that the mean is 0 and the variance is 1, we say that this is a *standard normal* distribution. The standard normal is traditionally denoted with the letter Z; it has special notation for its pmf and cdf:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(z)^2}{2}}, \quad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x)^2}{2}} dx$$

All of the results that we want to establish regarding the normal can be derived from the mgf, which is given by the following:[2]

$$m_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$$

[2] The notation indicates that we raise the values inside of the *exp* to the power of e.

We will show several results regarding the normal distribution on today's worksheet, including a derivation of the mgf as a starred problem. Here, let's state and prove the most important result that we will show all semester.

Theorem 15.1 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables from a distribution with a mean of 0 and a variance of 1. Define the sample average $\bar{X}_n$ for an n as:*

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

Then, as $n \to \infty$, we have:

$$\sqrt{n} \cdot \bar{X}_n = Z_n \underset{d}{\to} N(0, 1)$$

Proof. We prove the theorem assuming that the mgf of X_1, which we will call $M(t)$, exists. We know that this is the mfg for all of the X_i's because the X_i all come from the same distribution. Because X_1 has a mean of 0 and a variance of 1, we know that $M(0) = 1$, $M'(0) = 0$, and $M''(0) = 1$. Now, notice that the moment generating function of Z_n is given by:

$$\begin{aligned}
M_Z(t) &= \mathbb{E}\left[e^{t\cdot Z}\right] \\
&= \mathbb{E}\left[e^{t\cdot\sqrt{n}\cdot\bar{X}_n}\right] \\
&= \mathbb{E}\left[e^{t\cdot\frac{1}{\sqrt{n}}\cdot\sum_i X_i}\right] \\
&= \prod_i \mathbb{E}\left[e^{t\cdot\frac{1}{\sqrt{n}}\cdot X_i}\right] \\
&= \left[\mathbb{E}\left[e^{t\cdot\frac{1}{\sqrt{n}}\cdot X_1}\right]\right]^n \\
&= \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

We want to show that the limit of this quantity is $e^{t^2/2}$, the mgf of a standard normal distribution. A helpful trick is that we can instead show that the logarithm of this function converges to the logarithm of the standard normal mgf.[3] Then, we have:

$$\begin{aligned}
\lim_{n\to\infty} \log\left(\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n\right) &= n\log M\left(\frac{t}{\sqrt{n}}\right) \\
&= \lim_{y\to 0}\frac{log(M(yt))}{y^2} && \text{with } y = 1/\sqrt{n} \\
&= \lim_{y\to 0}\frac{tM'(yt)}{2yM(yt)} && \text{L'Hôpital's rule} \\
&= \frac{t}{2}\lim_{y\to 0}\frac{M'(yt)}{y} && M(yt)\to 1 \\
&= \frac{t^2}{2}\lim_{y\to 0} M''(yt) && \text{L'Hôpital's rule} \\
&= \frac{t^2}{2}.
\end{aligned}$$

Therefore, the limit of the log of the mgf is the log of the standard normal distribution. By definition, then, the value $Z_n = \sqrt{n}\bar{X}_n$ limits in distribution to a standard normal in distribution ■.

[3] The logarithm of the mgf is a frequently used tool called the cumulant-generating function.

Technically the CLT only says that the scaled sample mean will eventually approach the distribution of a standard normal. We will use the following theorem to justify this on today's worksheet in terms of the approximate distribution of a random variable.[4]

[4] You can define this concept in several ways, the most straightforward being that the maximum distance between the CDF of the true distribution and the approximate one can be bounded away by ϵ for some large enough n.

Theorem 15.2 (Central Limit Theorem Approximation) *Define $X_1, X_2, \ldots$ to be a sequence of i.i.d. random variables with finite mean μ and finite variance σ^2. Define the sample average $\bar{X}_n$ as in the central limit theorem. Then:*

$$\bar{X}_n \dot{\sim} N(\mu, \sigma^2/n).$$

Where $\dot{\sim}$ indicates the concept of approximately distributed as.