Lecture 17. The Normal Distribution

Fact:

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

Proof:

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right) e^{-y^2/2}\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2/2} e^{-y^2/2}\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy \\
&= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2}\, r\, dr\, d\theta \\
&= 2\pi \int_{0}^{\infty} e^{-r^2/2}\, r\, dr \\
&= 2\pi \int_{0}^{-\infty} -e^{u}\, du \quad \text{(letting } u = -r^2/2) \\
&= 2\pi \qquad\qquad /////
\end{aligned}$$

- From the fact above, we can conclude that

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

  is a pdf, since this is a non-negative, continuous function whose integral over the line is 1. A random variable having this as its pdf is called "standard normal." Often the letter Z is used to denote such a variable.

- If Z is standard normal, $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

- Suppose Z is standard normal. Let σ be a positive real number and let μ be any real number. Then

$$X = \sigma Z + \mu$$

  is said to be normal (μ, σ).

- If X normal (μ, σ), then $E(X) = \mu$ and $\mathrm{Var}(Z) = \sigma^2$.

$$\begin{aligned} P(X \le a) &= P(\sigma Z + \mu \le a) \\ &= P\left(Z \le \frac{a-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-z^2/2}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-(x-\mu)^2/2\sigma^2}\, dx \qquad \text{(letting } z = \tfrac{x-\mu}{\sigma}.\text{)} \end{aligned}$$

This shows that the pdf of a normal (μ, σ) random variable is:

$$\frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The method of Z-scores. This is simply the observation that if X normal (μ, σ), then the probability that X is less than $z_0\, \sigma$ above its mean is the same as the probability that the standard normal is less than z_0 above its mean:

$$P(X \le \mu + z_0\, \sigma) = P(Z \le z_0).$$

Thus, one can use a table of values of $P(Z \le z_0)$ to find $P(X \le x_0)$

Example. If moths of a certain species average 2.64 inches in length, with a standard deviation of 0.23 inches, what proportion of the moths are more than 3 inches long. *Solution.* Three inches is 36/23 standard deviations above the mean. By Mathematica, $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{36/22} e^{-z^2/2}\, dz$ is approximately 0.94