Normal Integrals

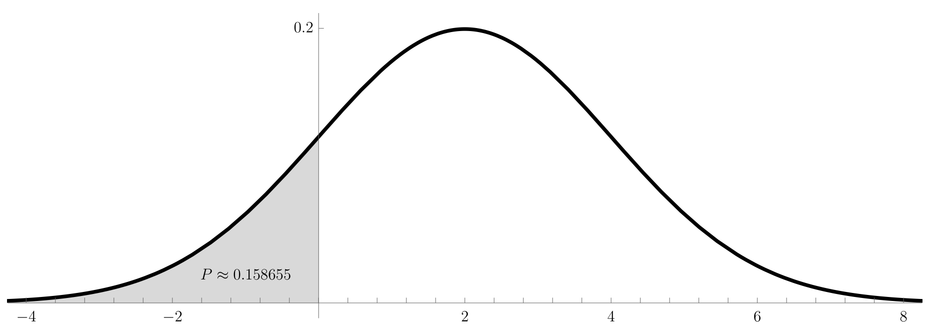

Figure 1: A normal integral

In statistics, we learn that data can often be analyzed using a *normal model*. For example, we can quantify how strong a score of 700 on the SAT or how tall the average NBA player is relative to the population since the data is normally distributed. In that context, we often compute a so-called *Z-score* using a formula like

$$Z = \frac{X - \mu}{\sigma}.$$

The point behind that formula is that it relates a normally distributed random variable X with mean μ and standard deviation σ to the *standard* normal with mean 0 and standard deviation 1. Ultimately, this formula arises from a change of the bounds of integration that arise when we perform u-substitution on a normal integral.

Normal random variables

In statistics, we learn that a lot of data is normally distributed and that probabilities associated with that data can be looked up in a normal table, like this or this. If Z represents a random variable with a standard normal distribution, we can compute the probability that Z lies between 0 and a (written as $P(0 < Z < a)$) just by looking the value up in the table. For example, if $a = 2$, then according to the tables,

$$P(0 < Z < 2) \approx 0.4772.$$

We also learn in statistics that there is a mysterious curve, the so-called *normal curve*, that allows these types of computations to be visualized in terms of area under that curve. The curve for this particular example is shown in Figure 2.

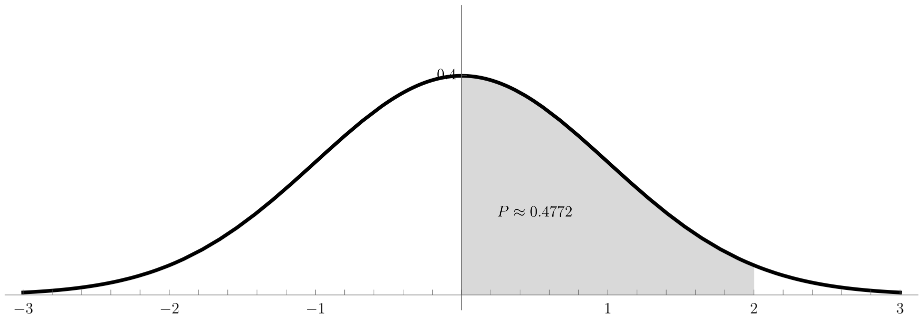

Figure 2: The normal integral for $P(0 < Z < 2)$

The symmetry of this curve allows us to compute normal probabilities that don't have zero as the lower bound. For example,

$$P(-1 < Z < 2) = P(0 < Z < 1) + P(0 < Z < 2) = 0.3413 + 0.4772 = 0.8185.$$

This is illustrated in figure 3.

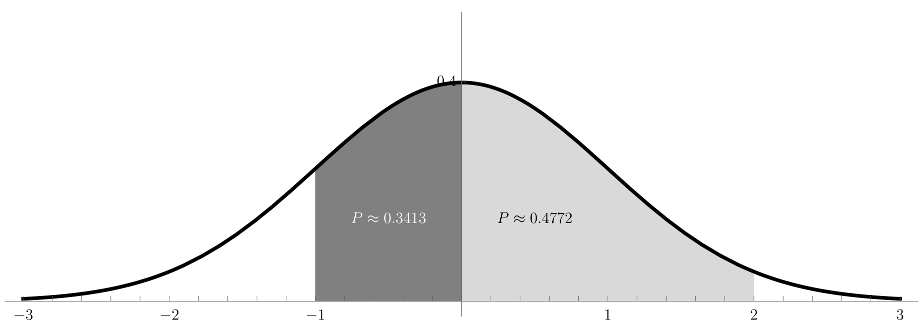

Figure 3: The normal integral for $P(-1 < Z < 2)$

The difficulty is that those tables only work for the *standard* normal, which assumes mean $\mu = 0$ and standard deviation $\sigma = 1$. What do we do when the mean or standard deviation are different? That's where the Z-score comes in. Assuming that X is normally distributed with mean μ and standard deviation σ, we learn in statistics that

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right).$$

Example: By design, SAT scores are normally distributed with mean 500 and standard deviation 100. If we pick a student at random, what is the probability that their SAT score is less than 700?

Solution: We first compute the Z-score of 700 relative to the mean of $\mu = 500$ and standard deviation of $\sigma = 100$. That is:

$$Z = \frac{700 - 500}{100} = 2.$$

Refferring to the standard normal table, we again find that $P(0 < Z < 2) \approx 0.4772$. By symmetry, $P(-\infty < Z < 0) = 1/2$. Thus,

$$P(-\infty < Z < 2) \approx 0.9772.$$

This is illustrated in figure 4. We could interpret this to mean that a score of 700 should put us at nearly the 98^{th} percentile.

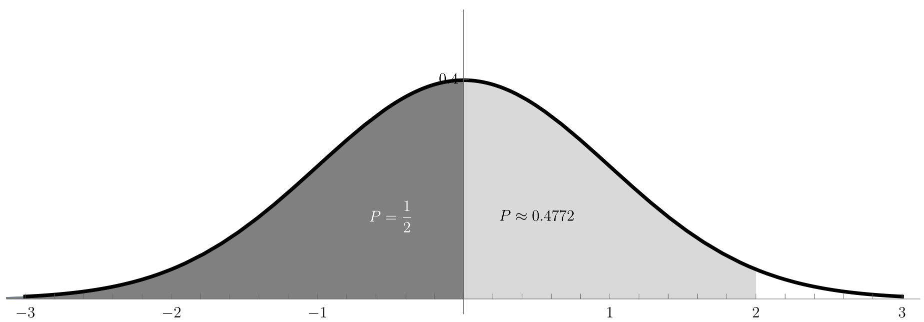

Figure 4: The normal integral for $P(-\infty < Z < 2)$

The normal curve

It turns out that there's a very specific formula for the normal curve, namely:

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

You can plot these things on Desmos.

The standard normal distribution arises when $\mu = 0$ and $\sigma = 1$; thus, it has formula

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

When a statistician tells us that a normal probability can be visualized in terms of area under a normal curve, they mean that

$$P(a < Z < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Unfortunately, this integral cannot be evaluated in closed form; it must be estimated using numerical techniques. That's exactly the point behind the normal table.

When a statistician tell us that

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right),$$

what they really mean is that

$$\frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}}\int_{(a-\mu)/\alpha}^{(b-\mu)/\sigma} e^{-x^2/2}dx.$$

We can use u-substitution to verify this. By writing the exponent of the integrand as

$$-\frac{(x-\mu)^2}{2\sigma^2} = -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2,$$

we are led to the substitution $u = (x-\mu)/\sigma$. Then, $du = \frac{1}{\sigma}dx$ so

$$\begin{aligned}\frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,\frac{1}{\sigma}dx\\ &= \frac{1}{\sqrt{2\pi}}\int_{(a-\mu)/\alpha}^{(b-\mu)/\sigma} e^{-x^2/2}dx.\end{aligned}$$

Example: Express the normal integral

$$\frac{1}{\sqrt{2\pi}3}\int_{-2}^{1} e^{-(x+2)^2/18}\,dx$$

as a *standard* normal integral.

Solution: Writing the exponent of the integrand as

$$-\frac{(x+2)^2}{18} = -\frac{1}{2}\left(\frac{x+2}{3}\right)^2,$$

we are led to the substitution $u = (x+2)/3$. Then, $du = \frac{1}{3}dx$ so

$$\begin{aligned}\frac{1}{\sqrt{2\pi}3}\int_{-2}^{1} e^{-\frac{(x+2)^2}{2\times 3^2}}\,dx &= \frac{1}{\sqrt{2\pi}}\int_{-2}^{1} e^{-\frac{1}{2}\left(\frac{x+2}{3}\right)^2}\,\frac{1}{3}dx\\ &= \frac{1}{\sqrt{2\pi}}\int_{(-2+2)/3}^{(1+2)/3} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_0^1 e^{-x^2/2}dx.\end{aligned}$$

Problems

1. Referring to a table of standard normal integrals, compute the following:

a) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{0}^{1.3} e^{-x^2/2}\, dx$

b) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-0.4}^{1.3} e^{-x^2/2}\, dx$

c) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{0.4}^{1.3} e^{-x^2/2}\, dx$

2. Using u-substitution, convert the following normal integrals into standard normal integrals. Then evaluate the integral using the table on the last page or your favorite numerical integrator.

a) $\dfrac{1}{\sqrt{2\pi}2} \displaystyle\int_{0}^{1} e^{-(x-1)^2/8}\, dx$

b) $\dfrac{1}{\sqrt{2\pi}4} \displaystyle\int_{12}^{18} e^{-(x-10)^2/32}\, dx$

3. Given that

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-x^2/2}\, dx = \frac{1}{2},$$

show that

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mu}^{\infty} e^{-(x-\mu)^2/\left(2\sigma^2\right)}\, dx = \frac{1}{2},$$

for all $\mu \in \mathbb{R}$ and $\sigma > 0$.