1 The Normal Distribution

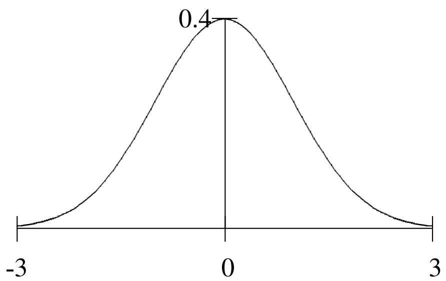
The Normal Distribution
Consider making the follow measurements.

Measure the height of everyone in DCU.

Measure the volume of coke in 200 1.5L bottles.

Measure the amount of salt in 100 samples of seawater.

Many observations like these which are obtained from measurements may be considered as random variables which follow a normal distribution.
The probability density function (pdf) for a normal distribution is called a normal curve:



The normal curve has these properties:

The curve is symmetrical about the center, which is the expectation (or mean) of X.

The total area under the curve is 1 (or 100%). Probability is equated with area.

The horizontal axis represents a continuous random variable (e.g. height, volume).

The area under the curve between two points on the horizontal axis gives the probability that the value of the variable lies between these two points.

The position and shape of the curve depend only on the expectation and variance of the random variable.

The normal curve is the graph of

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

Definition
We say that a random variable X is normally distributed if it has pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

where

$$\mu = E(X)$$

is the mean (or expectation) of X and

$$\sigma = \sqrt{V(X)}$$

is the square root of the variance of X. (σ is called the *standard deviation*.) We write

$$X \sim N(\mu, \sigma^2).$$

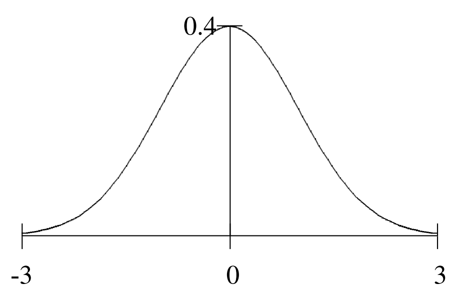


Figure 1: Normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.

So all normal distribution curves can be obtained by taking the basic normal curve $N(0, 1)$ and (i) squashing it down flatter or pulling it up sharper and (ii) shifting it up or down the horizontal axis.

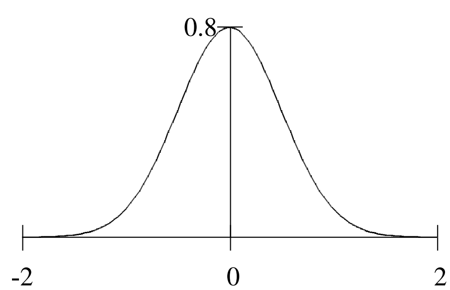


Figure 2: Normal curve with mean $\mu = 0$ and standard deviation $\sigma = 0.5$.

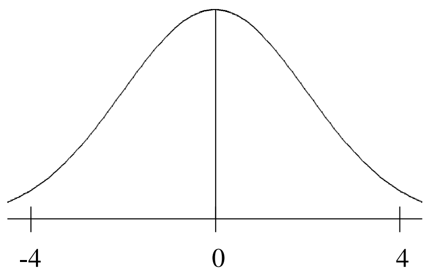


Figure 3: Normal curve with mean $\mu = 0$ and standard deviation $\sigma = 2$.

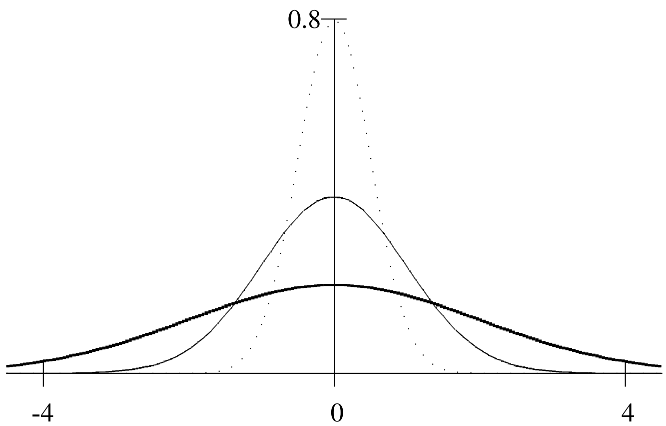


Figure 4: Figures 2, 3 (dashed curve) and 4 (bold curve) compared.

The proportion of the graph within one standard deviation of the mean is the same for every normal distribution.

The proportion of the graph within two standard deviations of the mean is the same for every normal distribution.

The proportion of the graph within three standard deviations of the mean is the same for every normal distribution.

...

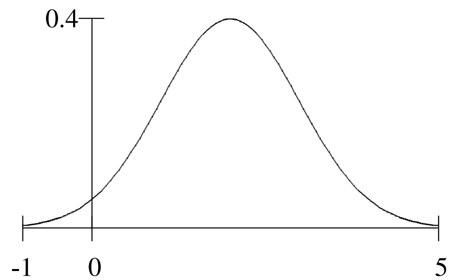


Figure 5: Normal curve with mean $\mu = 2$ and standard deviation $\sigma = 1$. Compare with Figure 2.

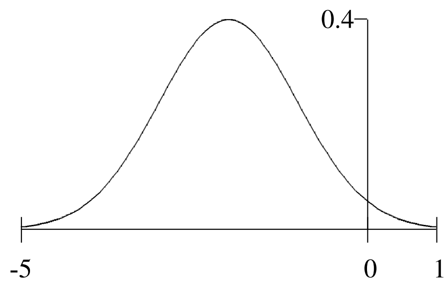


Figure 6: Normal curve with mean $\mu = -2$ and standard deviation $\sigma = 1$. Compare with Figures 2 and 6.

Example

> A bakery produces 3,000 loaves of bread each day.
>
> The mean weight of a loaf is $800g$ and the standard deviation for the loaves is $8g$.
>
> What proportion of the loaves will weigh more than $815g$?
>
> Assuming that the weight of loaves follows a normal distribution, we can draw the normal distribution $N(800, 64)$.

We want to calculate the shaded area (see board).
In one sense, this is very straightforward. We want to calculate the area under the curve

$$f(x) = \frac{1}{8\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-800}{8}\right)^2\right]$$

to the right of $x = 815$. We can write down the answer:

$$\text{Area} = \int_{815}^{\infty} \frac{1}{8\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-800}{8}\right)^2\right] dx.$$

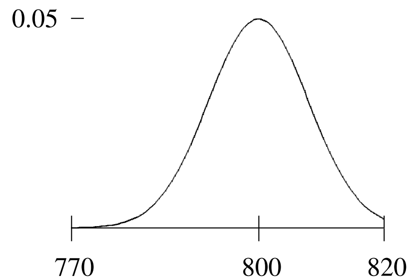


Figure 7: Normal curve with mean $\mu = 800$ and standard deviation $\sigma = 8$.

The problem is that this integral is very difficult to evaluate.

The Standard Normal Distribution
To see how to deal with this problem, we will focus on the simplest normal distribution. This is the one with $\mu = 0$ and $\sigma = 1$. We call it the *standard normal distribution*. So X has the standard normal distribution if X has pdf

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

We write $X \sim N(0, 1)$. This is important for the following reasons:

> By shifting and squashing, every normal distribution can be reduced to this one.
>
> The probabilities we are interested in calculating are integrals of the form
>
> $$\int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$
>
> These can be evaluated using tables of values given on page 36 of the log tables.

Important Fact
If $X \sim N(\mu, \sigma^2)$, then $Z \sim N(0, 1)$ where

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

In other words, if X has mean μ and standard deviation σ, then Z has mean $\mu = 0$ and standard deviation $\sigma = 1$.
For a particular value x of the random variable X, the quantity

$$z = \frac{x - \mu}{\sigma}$$

tells us the number of standard deviations by which x differs from μ.

Using the tables.
The cumulative distribution function (cdf) for a standard normal random variable is

$$\begin{aligned}\Phi(z_1) &= P(Z \leq z_1) = \int_{-\infty}^{z_1} f(z)\,dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}}\,dz.\end{aligned}$$

$$\Phi(z_1) = P(Z \leq z_1) = \text{Area to left of } z_1.$$

See Page 36 of log tables.
The following rules allow us to calculate probabilities for standard normal r.v.'s.

Probability of being above a positive number z_1:

$$P(z > z_1) = 1 - P(z \leq z_1) = 1 - \Phi(z_1).$$

Probability of being below a negative number $-z_1$:

$$\begin{aligned}P(z \leq -z_1) &= P(z > z_1) && \text{by symmetry} \\ &= 1 - P(z \leq z_1) && \text{by part 1} \\ &= 1 - \Phi(z_1).\end{aligned}$$

Probability of being above a negative number $-z_1$:

$$P(z \geq -z_1) = P(z \leq z_1) = \Phi(z_1).$$

Probability of being between two numbers a, b:

$$P(a \leq z \leq b) = P(z \leq b) - P(z \leq a).$$

Examples

Verify that 95.44% of the area under the standard normal curve lies between -2 and $+2$ (i.e. is within 2 standard deviations of the mean). In other words, verify that

$$P(-2 \leq Z \leq 2) = 0.9544.$$

Calculate $\Phi(1.96)$.

Assuming that $X \sim N(\mu, \sigma^2)$, use the tables to find $P(X > \mu + 2.2\sigma)$.

Let $Z \sim N(0, 1)$. Calculate $P(-1.35 \leq Z \leq 2.44)$.

Let $Z \sim N(0, 1)$. Calculate $P(Z \geq 1.05)$.

Individual weights of a batch of screws are normally distributed with mean $\mu = 2.10g$ and standard deviation $\sigma = 0.15g$.

What proportion of screws weigh more than $2.55g$?

The strength of individual bars made by a certain manufacturing process is normally distributed with mean $\mu = 24$ units and standard deviation $\sigma = 3$ units.

A customer requires at least 95% of the bars to have strength at least 20 units. Do the bars meet this specification?

Proof of the important fact

Let X be a normally distributed r.v. with mean μ and standard deviation σ; i.e. $X \sim N(\mu, \sigma)$. Then X has pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

Let

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

Our aim is to show that the r.v. Z has pdf

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

i.e. that $Z \sim N(0, 1)$. We do this by looking at the cdf's for X and Z.

Note A random variable which follows a normal distribution is sometimes called a *normal variate*.

Independent normal variates

If X, Y are independent normal variates with mean μ_X, μ_Y respectively and standard deviation σ_X, σ_Y respectively, then

the random variable $X + Y$ follows a normal distribution with mean $\mu_X + \mu_Y$ and variance $\sigma^2 = \sigma_X^2 + \sigma_Y^2$;

the random variable $X - Y$ follows a normal distribution with mean $\mu_X - \mu_Y$ and variance $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.

Example

The diameters of a shaft A and a bearing B are independent normal variates with means $\mu_A = 2.000$ cm, $\mu_B = 2.004$ cm and standard deviations $\sigma_A = 0.001$ cm, $\sigma_B = 0.0018$ cm. Find the probability of interference (i.e. diameter of shaft $\geq$ diameter of bearing).