# 9 Normal Distribution

An important and ubiquitous continuous distribution is the Normal distribution (also called the Gaussian). Normal distributions occur frequently in statistics, economics, natural sciences, and social sciences. For example, IQs approximately follow a Normal distribution. Men's heights and weights are approximately Normally distributed, as are women's heights and weights. Part of what makes the Normal distribution so relevant is the Central Limit Theorem (CLT; Section 9.4), which says that the average of a large number of independent and identically distributed (i.i.d.) quantities converges to a Normal. This explains, for example, why the Binomial random variable (r.v.) has a Normal shape when the number of coin flips is high. It also explains why noise (which is the mixture of many independent factors) is typically Normally distributed.

## 9.1 Definition

**Definition 9.1** *A continuous r.v. $X$ follows a* **Normal** *or* **Gaussian** *distribution, written $X \sim \text{Normal}(\mu, \sigma^2)$, if $X$ has probability density function (p.d.f.) $f_X(x)$ of the form*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty,$$

*where $\sigma > 0$. The parameter $\mu$ is called the* **mean***, and the parameter $\sigma$ is called the* **standard deviation***.*

**Definition 9.2** *$X$ follows a* **standard Normal** *distribution if $X \sim \text{Normal}(0, 1)$, that is,*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty.$$

The $\text{Normal}(\mu, \sigma^2)$ p.d.f. has a "bell" shape and is symmetric around $\mu$, as shown in Figure 9.1. The fact that $f_X(x)$ in Definition 9.1 is actually a density function can be seen by proving that it integrates to 1. This integration involves a change into polar coordinates (trust me, you do not want to see the gory details [71]).

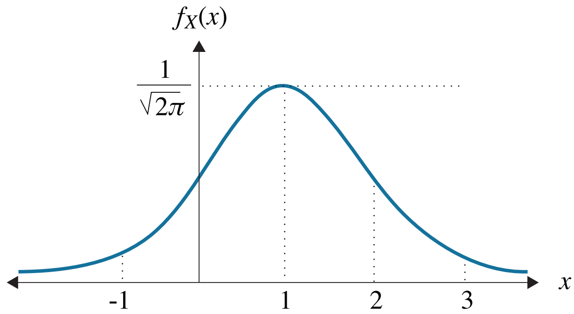


**Figure 9.1** *Normal*$(1, 1)$ *p.d.f.*

Theorem 9.3 shows that the parameters of the Normal distribution in fact represent its mean and variance.

**Theorem 9.3** *Let* $X \sim Normal(\mu, \sigma^2)$*, then* $\mathbf{E}[X] = \mu$ *and* $\mathbf{Var}(X) = \sigma^2$.

**Proof**: Because $f_X(x)$ is symmetric around $\mu$, it is obvious that $\mathbf{E}[X] = \mu$.

$$\begin{aligned}
\mathbf{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}((x-\mu)/\sigma)^2} dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \qquad \text{(let } y = (x-\mu)/\sigma \text{ and } dx = \sigma dy) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y \cdot \left(y e^{-y^2/2}\right) dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(-y e^{-y^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \qquad \text{(integration by parts)} \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\
&= \sigma^2.
\end{aligned}$$

The last line was obtained by using the fact that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1,$$

because the integrand is the density function of the standard Normal. ■

One of the things that makes the Normal distribution challenging is that its cumulative distribution function (c.d.f.) is not known in closed form. For the standard Normal, it is common to use the function $\Phi(\cdot)$ to represent the c.d.f., but the value of $\Phi(x)$ must be computed numerically. We will return to this point in Section 9.3.

**Definition 9.4** *If $X \sim Normal(0,1)$, then the c.d.f. of $X$ is denoted by*

$$\Phi(x) = F_X(x) = \mathbf{P}\{X \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

## 9.2 Linear Transformation Property

The Normal distribution has a very particular property known as the "Linear Transformation Property," which says that if $X$ is a Normal r.v., and you take a linear function of $X$, then that new r.v. will also be distributed as a Normal. Note that this property is *not* true for other distributions that we have seen, such as the Exponential.

**Theorem 9.5 (Linear Transformation Property)** *Let $X \sim Normal(\mu, \sigma^2)$. Let*

$$Y = aX + b,$$

*where $a > 0$ and $b \in \mathbb{R}$. Then, $Y \sim Normal(a\mu + b, a^2\sigma^2)$.*

**Proof**: Clearly $\mathbf{E}[Y] = a\mathbf{E}[X] + b = a\mu + b$ and $\mathbf{Var}(Y) = a^2\mathbf{Var}(X) = a^2\sigma^2$. All that remains is to show that $f_Y(y)$ is Normally distributed.

**Question:** What do we want $f_Y(y)$ to look like?

**Answer:** We want to show that

$$f_Y(y) = \frac{1}{\sqrt{2\pi}(a\sigma)} e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2}.$$

**Question:** Can we relate the p.d.f. of $Y$ to the p.d.f. of $X$ as follows:

$$f_Y(y) = \mathbf{P}\{Y = y\} = \mathbf{P}\{aX + b = y\} = \mathbf{P}\left\{X = \frac{y-b}{a}\right\} = f_X\left(\frac{y-b}{a}\right)?$$

**Answer:** The above is WRONG, because we can't say that $f_Y(y) = \mathbf{P}\{Y = y\}$. To make this argument correctly, we need to go through the c.d.f., which represents a valid probability.

We relate the c.d.f. of $Y$ to the c.d.f. of $X$ as follows:

$$F_Y(y) = \mathbf{P}\{Y \le y\} = \mathbf{P}\{aX + b \le y\} = \mathbf{P}\left\{X \le \frac{y-b}{a}\right\} = F_X\left(\frac{y-b}{a}\right).$$

We now differentiate both sides with respect to $y$:

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}\int_{-\infty}^{y} f_Y(t)dt \overset{\text{FTC}}{=} f_Y(y)$$

$$\frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = \frac{d}{dy}\int_{-\infty}^{\frac{y-b}{a}} f_X(t)dt \overset{\text{FTC}}{=} f_X\left(\frac{y-b}{a}\right)\cdot\frac{d}{dy}\left(\frac{y-b}{a}\right) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right),$$

where FTC denotes the Fundamental Theorem of Calculus (Section 1.3).

Thus we have shown that

$$f_Y(y) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right).$$

Evaluating this, we have

$$\begin{aligned}
f_Y(y) &= \frac{1}{a}f_X\left(\frac{y-b}{a}\right)\\
&= \frac{1}{a\sqrt{2\pi}\sigma}e^{-(\frac{y-b}{a}-\mu)^2/2\sigma^2}\\
&= \frac{1}{\sqrt{2\pi}(a\sigma)}e^{-(y-b-a\mu)^2/2a^2\sigma^2}\\
&= \frac{1}{\sqrt{2\pi}(a\sigma)}e^{-(y-(b+a\mu))^2/2a^2\sigma^2}.
\end{aligned}$$

So $f_Y(y)$ is a Normal p.d.f. with mean $a\mu + b$ and variance $a^2\sigma^2$. ■

## 9.3 The Cumulative Distribution Function

As stated earlier, unfortunately we do not know how to compute the c.d.f. of a Normal distribution. We must therefore use a table of numerically integrated results for $\Phi(y)$, such as that given in [82].[1]

Here is a snippet of the numerical table for $\Phi(y)$:

[1] In practice no one ever goes to the table anymore, because there are approximations online that allow you to compute the $\Phi(\cdot)$ values to within seven decimal places; see, for example, [75].

| $y$ | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|
| $\Phi(y)$ | 0.6915 | 0.8413 | 0.9332 | 0.9772 | 0.9938 | 0.9987 |

**Question:** Looking at the table you see, for example, that $\Phi(1) = 0.8413$. What does this tell us about the probability that the standard Normal is within one standard deviation of its mean?

**Answer:** Let $Y \sim \text{Normal}(0, 1)$. Since $\Phi(1) \doteq 0.84$, we know that $\mathbf{P}\{Y < 1\} = 0.84$. We want to know $\mathbf{P}\{-1 < Y < 1\}$.

$$\begin{aligned}
\mathbf{P}\{-1 < Y < 1\} &= \mathbf{P}\{Y < 1\} - \mathbf{P}\{Y < -1\} \\
&= \mathbf{P}\{Y < 1\} - \mathbf{P}\{Y > 1\} \quad \text{(by symmetry)} \\
&= \mathbf{P}\{Y < 1\} - (1 - \mathbf{P}\{Y < 1\}) \\
&= 2\mathbf{P}\{Y < 1\} - 1 \\
&= 2\Phi(1) - 1 \\
&\doteq 2 \cdot 0.84 - 1 \\
&= 0.68.
\end{aligned}$$

So with probability approximately 68%, we are within one standard deviation of the mean.

**Question:** If $Y \sim \text{Normal}(0, 1)$, what's the probability that $Y$ is within $k$ standard deviations of its mean?

**Answer:**

$$\mathbf{P}\{-k < Y < k\} = 2\Phi(k) - 1. \tag{9.1}$$

Equation (9.1) tells us the following useful facts:

- With probability $\approx$ 68%, the Normal is within 1 standard deviation of its mean.
- With probability $\approx$ 95%, the Normal is within 2 standard deviations of its mean.
- With probability $\approx$ 99.7%, the Normal is within 3 standard deviations of its mean.

**Question:** The "useful facts" were expressed for a standard Normal. What if we do not have a standard Normal?

**Answer:** We can convert a non-standard Normal into a standard Normal using the Linear Transformation Property. That is:

$$X \sim \text{Normal}(\mu, \sigma^2) \iff Y = \frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1).$$

Thus, if $Y \sim \text{Normal}(0, 1)$, and $X \sim \text{Normal}(\mu, \sigma^2)$, then the probability that $X$ deviates from its mean by less than $k$ standard deviations is:

$$\mathbf{P}\{-k\sigma < X - \mu < k\sigma\} = \mathbf{P}\left\{-k < \frac{X - \mu}{\sigma} < k\right\} = \mathbf{P}\{-k < Y < k\}.$$

This point is summarized in Theorem 9.6.

**Theorem 9.6** *If $X \sim$ Normal$(\mu, \sigma^2)$, then the probability that $X$ deviates from its mean by less than $k$ standard deviations is the same as the probability that the standard Normal deviates from its mean by less than $k$.*

Theorem 9.6 illustrates why it is often easier to think in terms of standard deviations than absolute values.

**Question:** Proponents of IQ testing will tell you that human intelligence (IQ) has been shown to be Normally distributed with mean 100 and standard deviation 15. What fraction of people have an IQ greater than 130 ("the gifted cutoff")?

**Answer:** We are looking for the fraction of people whose IQ is more than two standard deviations *above* the mean. This is the same as the probability that the standard Normal exceeds its mean by more than two standard deviations, which is $1 - \Phi(2) = 0.023$. Thus only about 2.3% of people have an IQ above 130.

Other properties of the Normal distribution will be proven later in the book. A particularly useful property is that the sum of two independent Normal distributions is Normally distributed.

**Theorem 9.7 (Sum of two independent Normals)** *Let $X \sim$ Normal$(\mu_x, \sigma_x^2)$. Let $Y \sim$ Normal$(\mu_y, \sigma_y^2)$. Assume $X \perp Y$. Let $W = X + Y$. Then*

$$W \sim \textit{Normal}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$

**Proof**: This will be proven in Exercise 11.10 via Laplace transforms. ■

## 9.4 Central Limit Theorem

Consider sampling the heights of 1000 individuals within the country and taking that average. The CLT, which we define soon, says that this average will tend to be Normally distributed. This would be true even if the distribution of individual heights were not Normal. Likewise, the CLT would apply if we took the average of a large number of Uniform random variables. It is this property that makes the Normal distribution so important! We now state this more formally.

Let $X_1, X_2, X_3, \ldots, X_n$ be independent and identically distributed random variables with some mean $\mu$ and variance $\sigma^2$. Note: We are *not* assuming that these are Normally distributed random variables. In fact we are not even assuming that they are necessarily continuous random variables – they may be discrete.

Let

$$S_n = X_1 + X_2 + \cdots + X_n. \tag{9.2}$$

**Question:** What are the mean and standard deviation of $S_n$?

**Answer:** $\mathbf{E}[S_n] = n\mu$ and $\mathbf{Var}(S_n) = n\sigma^2$. Thus $\mathbf{std}(S_n) = \sigma\sqrt{n}$.

Let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

**Question:** What are the mean and standard deviation of $Z_n$?

**Answer:** $Z_n$ has mean 0 and standard deviation 1.

**Theorem 9.8 (Central Limit Theorem (CLT))** *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with common mean $\mu$ and finite variance $\sigma^2$, and define*

$$S_n = \sum_{i=1}^{n} X_i \qquad \textit{and} \qquad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

*Then the distribution of $Z_n$ converges to the standard normal, Normal$(0, 1)$, as $n \to \infty$. That is,*

$$\lim_{n\to\infty} \mathbf{P}\{Z_n \le z\} = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

*for every $z$.*

**Proof**: Our proof uses Laplace transforms, so we defer it to Exercise 11.12. ■

It should seem counter-intuitive to you that $Z_n$ converges to a Normal in distribution, especially when the $X_i$'s might be very skewed and not-at-all Normal themselves.

**Question:** Does the sum $S_n$ also converge to a Normal?

**Answer:** This is a little trickier, but, for practical purposes, yes. Since $S_n$ is a linear transformation of $Z_n$, then by the Linear Transformation Property, $S_n$ gets closer and closer to a Normal distribution too. However, $S_n$ is not well defined as $n \to \infty$, because $S_n$ is getting closer and closer to Normal$(n\mu, n\sigma^2)$, which has infinite mean and variance as $n \to \infty$. There's another problem with looking at $S_n$. Suppose all the $X_i$'s are integer-valued. Then $S_n$ will also be integer-valued and hence not exactly Normal (although it will behave close to Normal for high $n$ – see Exercise 9.6). For all these reasons, CLT involves $Z_n$ rather than $S_n$.

**Question:** Does the average $A_n = \frac{1}{n}S_n$ converge to a Normal?

**Answer:** Yes! Applying the Linear Transformation Property to $Z_n$, we see that $A_n$ gets closer and closer to a Normal with mean $\mu$ and variance $\frac{\sigma^2}{n}$.

The CLT is extremely general and explains many natural phenomena that result in Normal distributions. The fact that CLT applies to any sum of i.i.d. random variables allows us to prove that the Binomial$(n, p)$ distribution, which is a sum of i.i.d. Bernoulli$(p)$ random variables, can be approximated by a Normal distribution when $n$ is sufficiently high. In Exercise 9.7 you will use a similar argument to explain why the Poisson$(\lambda)$ distribution is well represented by a Normal distribution when $\lambda$ is high.

In the next example, we illustrate how the CLT is used in practice.

**Example 9.9 (Signal with noise)**

Imagine that we are trying to transmit a signal. During the transmission, there are 100 sources independently making low noise. Each source produces an amount of noise that is Uniformly distributed between $a = -1$ and $b = 1$. If the total amount of noise is greater than 10 or less than $-10$, then it corrupts the signal. However, if the absolute value of the total amount of noise is under 10, then it is not a problem.

**Question:** What is the approximate probability that the absolute value of the total amount of noise from the 100 signals is less than 10?

**Answer:** Let $X_i$ be the noise from source $i$. Observe that

$$\mu_{X_i} = 0$$
$$\sigma^2_{X_i} = \frac{(b-a)^2}{12} = \frac{1}{3}$$
$$\sigma_{X_i} = \frac{1}{\sqrt{3}}.$$

Let $S_{100} = X_1 + X_2 + \cdots + X_{100}$.

$$\begin{aligned}\mathbf{P}\{-10 < S_{100} < 10\} &= \mathbf{P}\left\{\frac{-10}{\sqrt{100/3}} < \frac{S_{100}-0}{\sqrt{100/3}} < \frac{10}{\sqrt{100/3}}\right\} \\ &\approx \mathbf{P}\left\{-\sqrt{3} < \text{Normal}(0,1) < \sqrt{3}\right\} \\ &= 2\Phi\left(\sqrt{3}\right) - 1 \\ &\approx 0.91.\end{aligned}$$

Hence the approximate probability of the signal getting corrupted is < 10%. In practice, this CLT approximation is excellent, as we'll see in Chapter 18.

## 9.5 Exercises

9.1 **Practice with the $\Phi(\cdot)$ table**
Let $X \sim \text{Normal}(0,1)$. Let $Y \sim \text{Normal}(10, 25)$. Using the table for $\Phi(\cdot)$ values given in the chapter, answer the following questions:
(a) What is $\mathbf{P}\{X > 0\}$?
(b) What is $\mathbf{P}\{-1 < X < 1.5\}$?
(c) What is $\mathbf{P}\{-2.5 < Y < 22.5\}$?

9.2 **Total work processed by server**
A server handles 300 jobs per day. Job sizes are i.i.d. and are Uniformly distributed between 1 second and 3 seconds. Let $S$ denote the sum of the sizes of jobs handled by the server in a day. Approximately, what is $\mathbf{P}\{590 < S < 610\}$?

9.3 **Bytes at a server**
A server receives 100 messages a day. Message sizes (in bytes) are i.i.d. from distribution $\text{Exp}(\mu)$. Let $S$ denote the total number of bytes received by the server. Approximately, what is $\mathbf{P}\left\{\frac{90}{\mu} < S < \frac{110}{\mu}\right\}$?

9.4 **Estimating failure probability**
Suppose that 10% of cars have engine light problems at some point in their lifetime. If a dealer sells 200 cars, what is the (approximate) probability that fewer than 5% of the cars she sells will eventually have engine light problems? Use the appropriate Normal distribution table. Express your answer as a decimal.

9.5 **Linear Transformation of Exponential**
Recall that the Normal distribution has a pretty Linear Transformation property. Does the Exponential distribution have this as well? Let $X \sim \text{Exp}(\mu)$. Let $Y = aX + b$, where $a$ and $b$ are positive constants. Is $Y$ Exponentially distributed? Prove your answer.

9.6 **Accuracy of the Central Limit Theorem**
Bill Gater invites 1,000 friends to a dinner. Each is asked to make a contribution. The contributions are i.i.d. Poisson-distributed random variables with mean \$1,000 each. Bill hopes to raise \$1,000,000. Your job is to compute the probability that Bill raises < \$999,000.
(a) Compute this using the Normal approximation from this chapter.
(b) Now write an *exact* expression for this probability, and then use your calculator or small program to evaluate the expression.

9.7 **Why a Poisson looks like a Normal**
You may have noticed that the Poisson($\lambda$) distribution looks very similar in shape to a Normal with mean $\lambda$ and variance $\lambda$. This is particularly true for high $\lambda$. Use the CLT approximation to explain why this is, in the case where $\lambda$ is a high integer. [Hint: The exercises on the Poisson distribution from Chapter 6 are useful here.]

9.8 **Heuristic proof of Stirling's approximation**
[Contributed by Ishani Santurkar] Stirling's approximation, Theorem 1.14, says that $n!$ grows in accordance with (9.3) for large $n$:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{9.3}$$

In this problem you will come up with a heuristic proof for this fact.
(a) Let $X \sim \text{Poisson}(n)$. What is $p_X(n)$?
(b) Now assume that $n$ is large, and use the Normal approximation from Exercise 9.7 to write an alternative approximate expression for $p_X(n)$. Note that for a continuous r.v. $Y$ we can't talk about $\mathbf{P}\{Y = i\}$, but we can write: $\mathbf{P}\{i < Y < i+1\} \approx f_Y(i) \cdot 1$.
(c) Equate (a) and (b) to get (9.3).

9.9 **Fractional moments**
Given the ugliness of the Normal distribution, I am happy to say that it never comes up in my research . . . until a few days ago! Here is the story: I had a r.v. $X \sim \text{Exp}(1)$ and I needed to compute $\mathbf{E}\left[X^{\frac{1}{2}}\right]$. Figure out why I needed a Normal distribution to do this and what answer I finally got. [Hint: Start by applying integration by parts. Then make the right change of variables. If you do it right, the standard Normal should pop out. Remember that the Exponential ranges from 0 to $\infty$, whereas the Normal ranges from $-\infty$ to $\infty$.]

9.10 **Sampling from an unknown distribution**
We want to understand some statistics (e.g., mean and variance) of the webpage load time distribution, $X$. To do that, we randomly choose $n$ websites and measure their load times, $X_1, X_2, \ldots, X_n$. We assume that the $X_i$'s are i.i.d. samples of $X$, where $X_i \sim X$. Our goal is to use these samples to estimate $X$'s mean, $\mu = \mathbf{E}[X]$, and $X$'s variance, $\sigma^2 = \mathbf{Var}(X)$.
Our *sample mean* $\overline{X}$ is defined as

$$\overline{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Our *sample variance* is defined as

$$S^2 \equiv \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

(a) If the expectation of the sample mean is the same as the actual mean, that is if $\mathbf{E}\left[\overline{X}\right] = \mu$, then $\overline{X}$ is called an *unbiased estimator of the mean* of the sampling distribution. Prove that $\overline{X}$ is an unbiased estimator of the mean.
(b) If the expectation of the sample variance is the same as the actual variance, that is, if $\mathbf{E}\left[S^2\right] = \sigma^2$, then $S^2$ is called an *unbiased estimator of the variance* of the sampling distribution. Prove that $S^2$ is an unbiased estimator of the variance.
It will help to follow these steps:
(i) Start by expressing $(n-1)S^2 = \sum_{i=1}^{n}\Big((X_i - \mu) + (\mu - \overline{X})\Big)^2$.
(ii) From (i), show that: $(n-1)S^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\mu - \overline{X})^2$.
(iii) Take expectations of both sides of (ii).