CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture 3B: The Normal Distribution

Prof. Oke

UMass Amherst

College of Engineering

September 25, 2025

Outline

Reading: OpenIntro Statistics 4.1

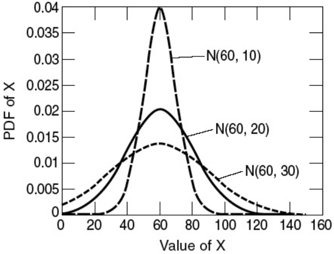
Normal (Gaussian) distribution

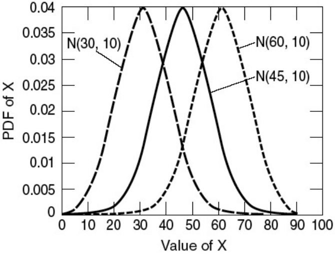
Definition

Denoted as $\mathcal{N}(\mu, \sigma^2)$ or $\mathcal{N}(\mu, \sigma)$, the normal distribution is continuous with PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \equiv \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad -\infty < x < \infty \quad (1)$$

where μ and σ^2 (or σ) are its parameters (mean and variance (or standard deviation)).





Example 1: Normal distribution parameters

(a) A random variable X is normally distributed as: $X \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$. What are the mean and standard deviation of this distribution? Mean: $\mu = 0$, standard deviation: $\sigma = 1$

(b) A random variable X is normally distributed as: $X \sim \mathcal{N}(\mu = 2, \sigma^2 = 4)$. What are the mean and variance of this distribution? Mean: $\mu = 2$, variance: $\sigma^2 = 4$

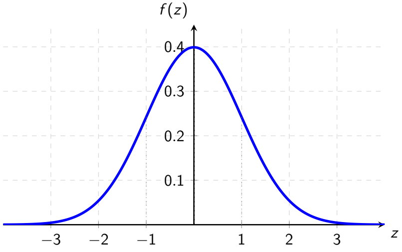


Figure: Standard normal distribution $\mathcal{N}(0, 1)$

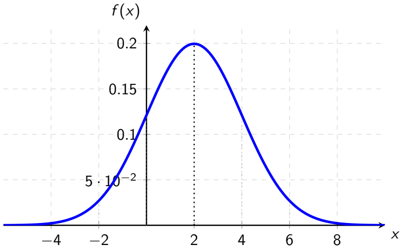


Figure: Normal distribution with $\mu = 2$, $\sigma = 2$

CDF of a normal distribution

- The CDF of a normal distribution is the integral of the PDF:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_x(x)dx = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \tag{2}$$

- There is no closed-form solution to this integral
- So, in texts/tables, we denote the **standard normal** CDF as $\Phi(z)$, where:

$$\Phi(z) = \int_{-\infty}^{z} f_Z(z) \frac{1}{1\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] dz = P(Z \leq z) \tag{3}$$

where

$$z = \frac{x-\mu}{\sigma} \tag{4}$$

- The standardized normal variable z is often referred to as the Z-score

Standard normal distribution

If a random variable X has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, then the r.v. Z has a standard normal distribution if

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1) \tag{5}$$

The **standardized normal** therefore has a mean of 0 and variance of 1. Its PDF is thus:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad -\infty < z < \infty \tag{6}$$

The CDF Φ of the standard normal variate Z is given by:

$$\Phi(z) = F_Z(z) \equiv P(Z \leq z) \tag{7}$$

Example 2: Computing the Z-score

If $X \sim \mathcal{N}(10, \sigma^2 = 4)$, what is the Z-score of a sample $x = 5$?

$$
\begin{aligned}
z &= \frac{x-\mu}{\sigma} = \frac{5-10}{\sqrt{4}} \\
&= -\frac{5}{2} = \boxed{-2.5}
\end{aligned}
$$

This means that the sample $x = 5$ is 2.5 standard deviations below the mean

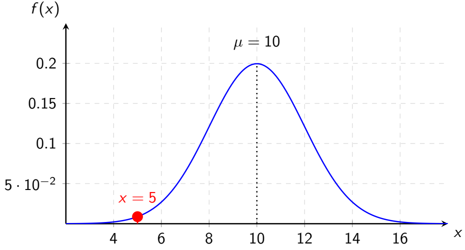


Figure: Normal distribution with $\mu = 10$, $\sigma^2 = 4$, and $x = 5$ indicated

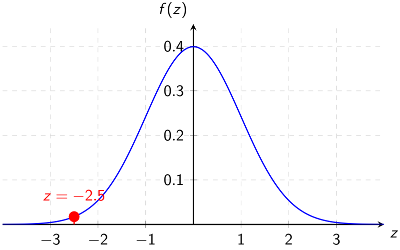


Figure: Standard normal distribution with $z = -2.5$ indicated

Example 3: Computing the Z-score

If $X \sim \mathcal{N}(\mu = 6, \sigma^2 = 9)$, what is the Z-score of a sample $x = 8$?

$$\begin{aligned} z &= \frac{x-\mu}{\sigma} = \frac{8-6}{\sqrt{9}} \\ &= \frac{2}{3} = \boxed{0.67} \end{aligned}$$

This means that the sample $x = 8$ is 0.67 standard deviations above the mean

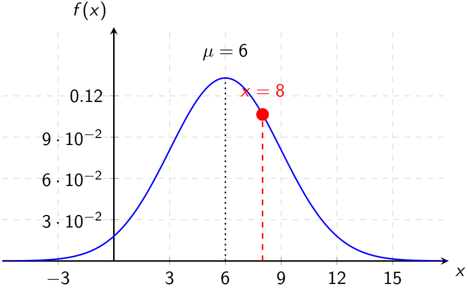


Figure: Normal distribution with $\mu = 6$, $\sigma^2 = 9$, and $x = 8$ indicated

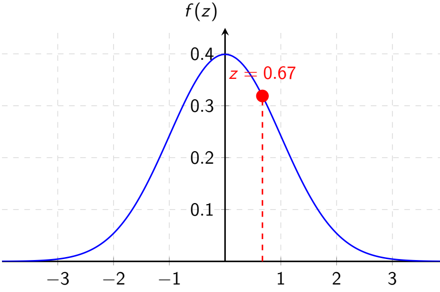


Figure: Standard normal distribution with $z = 0.67$ indicated

Example 4: Finding the normal variate from a Z-score

A normal r.v. $X \sim \mathcal{N}(\mu = 2, \sigma = 3)$ has a Z-score of $z = 1.25$. What is the corresponding x value?

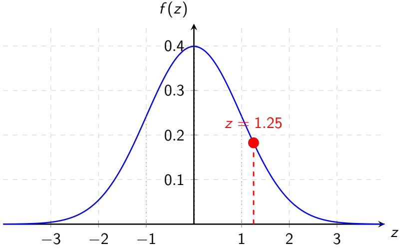


Figure: Standard normal distribution with $z = 1.25$ indicated

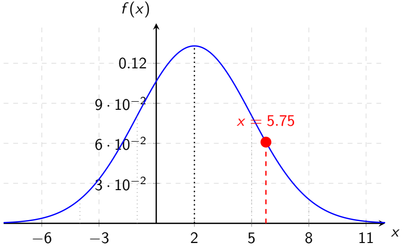


Figure: Normal distribution with $\mu = 2$, $\sigma = 3$, and $x = 5.75$ indicated

$$
\begin{aligned}
z &= \frac{x - \mu}{\sigma} \\
x &= \mu + z\sigma = 2 + 1.25(3) = \boxed{5.75}
\end{aligned}
$$

68-95-99.7 rule

The probabilities of a normal r.v. within ± 1, ± 2 and ± 3 standard deviations are 68.3%, 95.4% and 99.7%, respectively. This is known as the **68-95-99.7 rule** or the **empirical rule**.

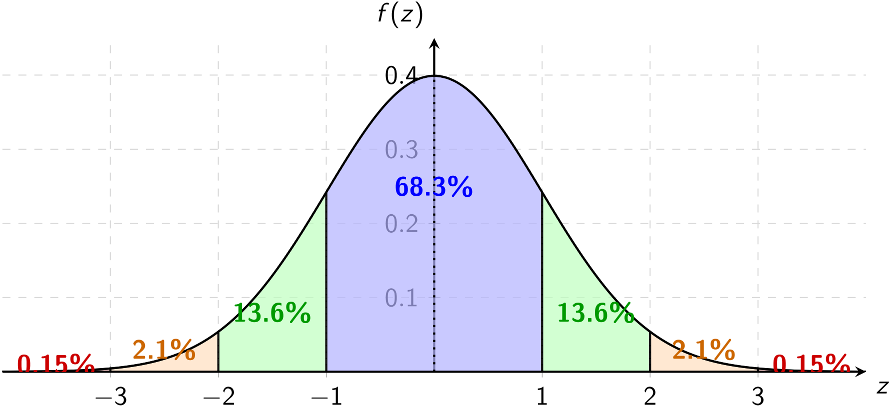


Figure: Standard normal distribution with color-coded regions showing probability percentages

Probability of a normal random variable

The probability that a normal r.v. lies within a certain interval is given by the *area* under the PDF in that interval.

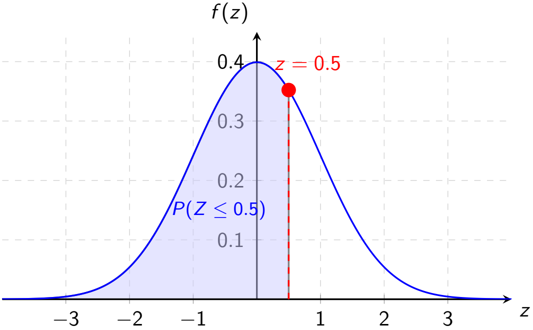


Figure: Standard normal distribution with $z = 0.5$ indicated and $P(Z \leq 0.5)$ shaded

- Recall that the area under the PDF within a given interval is the CDF evaluated in that range
- Thus, in the above figure: $p = P(Z \leq z_p) = \Phi(z_p)$

Probability of a normal random variable (cont.)

Given a normal r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$P(a < X \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \tag{8}$$

Substituting $z = \frac{x-\mu}{\sigma}$ and $dx = \sigma dz$, we obtain:

$$P(a < X \leq b) = \frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} e^{-\frac{1}{2}z^2} dz \tag{9}$$

Recognizing that the integrand is the PDF of a standard normal distribution, we have:

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{10}$$

Example 5a: Normal probabilities

SAT scores are normally distributed as $X \sim \mathcal{N}(1100, \sigma = 200)$.

(a) What is the probability that a randomly selected student has a score that is at least 1200? The Z-score is $z = \frac{1200-1100}{200} = 0.5$
Thus,

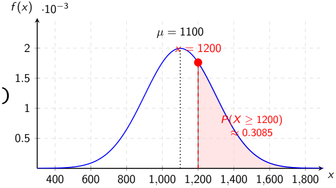
$$P(X \geq 1200) \quad = \quad 1 - \Phi(.5) = 1 - .695 = \boxed{.3085}$$

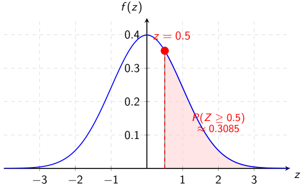
In Python, you can compute this probability using the scipy.stats library:

```
import scipy.stats as stats
p = 1 - stats.norm.cdf(1200, 1100, 200)
```

The first 3 arguments of stats.norm.cdf are the value, mean (texttloc), and standard deviation (texttscale), respectively. OR, you can use the Z-score (default mean=0, std=1):

```
import scipy.stats as stats
p = 1 - stats.norm.cdf(0.5)
```





Example 5b: Normal probabilities

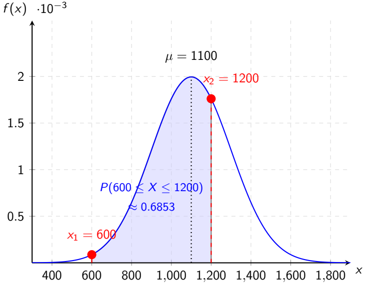
SAT scores are normally distributed as $X \sim \mathcal{N}(1100, \sigma = 200)$.

(b) What is the probability that another randomly selected student's score is greater than 600 but less than 1200?

$$\begin{aligned} P(600 \leq X < 1200) &= \Phi\left(\frac{1200 - 1100}{200}\right) - \Phi\left(\frac{600 - 1100}{200}\right) \\ &= \Phi(.5) - \Phi(-2.5) = \boxed{.6853} \end{aligned}$$

In Python:

```
from scipy.stats import norm
p = norm.cdf(1200, 1100, 200)
    - norm.cdf(600, 1100, 200)
```



Example 5b: Normal probabilities (cont)

(b) OR, you can use the Z-scores (default mean=0, std=1):

```
import scipy.stats as stats
p = norm.cdf(0.5) - norm.cdf(-2.5)
```

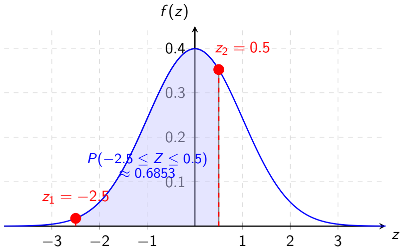


Figure: Standard normal PDF with probability area shaded

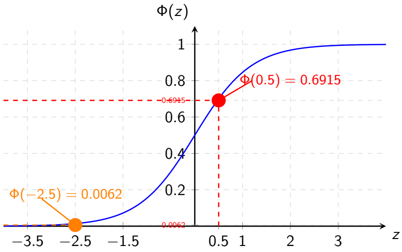


Figure: Standard normal CDF with corresponding probability values marked

Example 6: Inverse normal probabilities

SAT scores are normally distributed as $X \sim \mathcal{N}(1100, \sigma = 200)$.

(c) If the probability of an SAT score lower than x is 0.4, find x.

$$\begin{aligned} z &= \Phi^{-1}(.4) = -.2533 = \frac{x-\mu}{\sigma} \\ \therefore x &= z\sigma + \mu = -.2533(200) + 1100 = \boxed{1049} \end{aligned}$$

```python
from scipy.stats import norm
p = norm.ppf(0.4, 1100, 200)
```

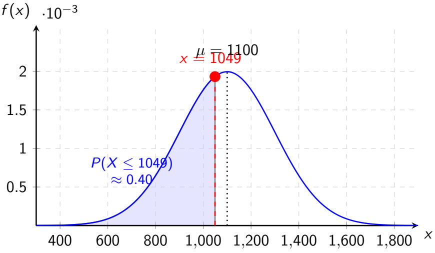


Figure: Normal distribution with $\mu = 1100$, $\sigma = 200$, and $P(X \leq 1049)$ shaded

Example 6: Inverse normal probabilities (cont.)

(c) You can think of the inverse CDF as finding the x value that corresponds to a given percentile.

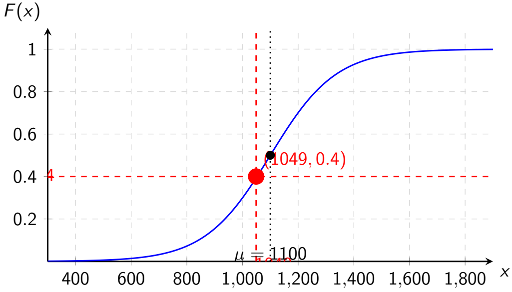


Figure: Normal distribution CDF showing the 40th percentile at $x = 1049$

More on the normal CDF

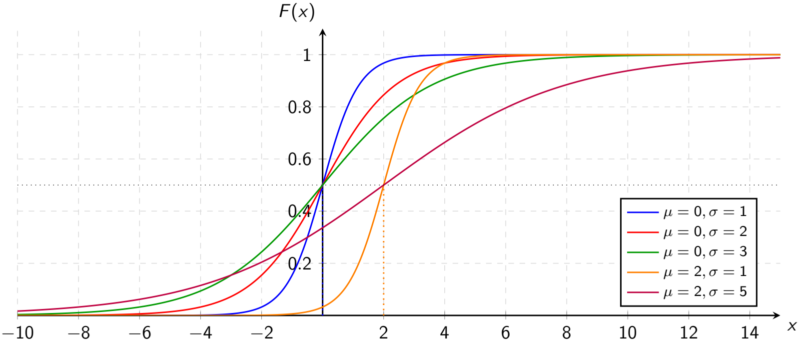


Figure: Comparison of normal distribution CDFs with different parameters

- The standard normal CDF is the blue curve in the above figure
- Quantiles can be read off the plot (e.g. the median is the value of X corresponding to the y value of 0.5)
- $\Phi(-z) = 1 - \Phi(z)$
- $z = \Phi^{-1}(p) = -\Phi^{-1}(1-p)$

Example 7: Probability of flooding

The drainage from a community during a storm is a normal random variable estimated to have a mean of 1.2 million gallons per day (mgd) and an SD of 0.4 mgd. If the storm drain system is designed with a maximum drainage capacity of 1.5 mgd:

(a) What is the underlying probability of flooding during a storm that is assumed in the design of the drainage system?

(b) Find $P(1.0 < X \leq 1.6)$.

(c) Find the 90th-percentile drainage load from the community during a storm.

Example 7: Probability of flooding (cont.)

Given $\mu = 1.2$ and $\sigma = 0.4$.

(a) What is the underlying probability of flooding during a storm that is assumed in the design of the drainage system?

Solution

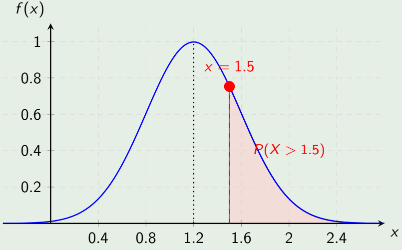


Figure: Normal distribution with $\mu = 1.2$, $\sigma = 0.4$, and $P(X > 1.5)$ shaded

$$
\begin{aligned}
P(X > 1.5) &= 1 - P(X \leq 1.5) \\
&= 1 - \Phi\left(\frac{1.5 - 1.2}{0.4}\right) \\
&= 1 - \Phi(0.75) \\
&= 1 - 0.7734 = \boxed{0.227}
\end{aligned}
$$

In Python: `1 - norm.cdf(1.5, 1.2, 0.4)`

Example 7: Probability of flooding (cont.)

(b) Find $p = P(1.0 < X \leq 1.6)$:

Solution

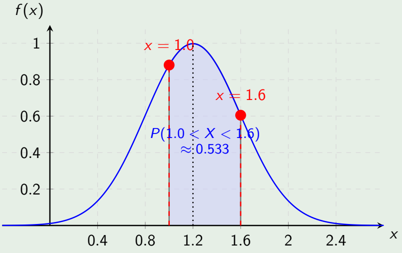


Figure: Normal distribution with $\mu = 1.2$, $\sigma = 0.4$, and $P(1.0 < X < 1.6)$ shaded

$$
\begin{aligned}
p &= \Phi\left(\frac{1.6 - 1.2}{0.4}\right) \\
&\quad -\Phi\left(\frac{1.0 - 1.2}{0.4}\right) \\
&= \Phi(1.0) - \Phi(-0.5) \\
&= 0.8413 - [1 - \Phi(0.5)] \\
&= 0.8413 - (1 - 0.6915) \\
&= 0.8413 - 0.3085 \\
&= 0.5328 \approx \boxed{0.533}
\end{aligned}
$$

In Python: `norm.cdf(1.6, 1.2, 0.4) - norm.cdf(1.0, 1.2, 0.4)`

Example 7: Probability of flooding (cont.)

(c) Find the 90th-percentile drainage load from the community during a storm.

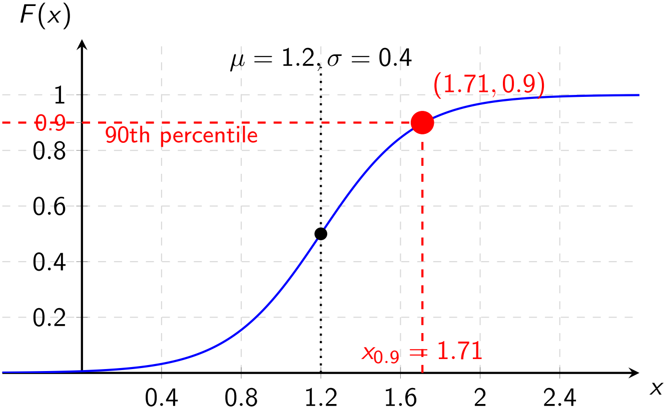


Figure: Normal CDF showing the 90th percentile at $x = 1.71$

Example 7: Probability of flooding (cont.)

(c) Find the 90th-percentile drainage load from the community during a storm.

Solution

$$
\begin{aligned}
P(X \le x_{0.90}) &= \Phi\left(\frac{x_{0.90} - 1.2}{0.40}\right) = 0.90 \\
\Longrightarrow \frac{x_{0.90} - 1.2}{0.40} &= \Phi^{-1}(0.90) = 1.28 \\
\therefore x_{0.90} &= 1.28(0.40) + 1.2 = 1.71 \text{ mgd}
\end{aligned}
$$

In Python:

```python
from scipy.stats import norm
p90 = norm.ppf(0.9, 1.2, 0.4)
```

gives 1.7095 mgd

Example 8: Steel beam reliability

Assume the variability E in the lengths of steel beams is normally distributed. What is the precision (in terms of σ) required for a reliability of 99.7%, given that the specified tolerance for a construction project is ± 5 mm?

Definitions:

- **Precision**: in physical terms is the inverse of the variance (i.e. higher precision means lower variance). In this context, all you need to do is find the standard deviation σ.
- **Reliability**: probability that the deviation in the length of a beam meets (falls within) the specified tolerance

Recap of normal distribution

- The PDF of the normal distribution (parameters μ and σ^2) is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{11}$$

- The parameters of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ correspond to its mean and variance, respectively.
- There is no closed-form solution to the integral of the normal CDF
- Instead, it is customary to standardize a normal variable to its "Z-score":

$$Z = \frac{X-\mu}{\sigma} \tag{12}$$

- The mean and variance of the standard normal distribution are 0 and 1, respectively.
- The symbol Φ ("phi") is used to represent the CDF of the *standard* normal distribution, whose values can be looked up in a table.
- In Python, the `scipy.stats.norm.cdf(x, mu, sigma)` and `scipy.stats.norm.ppf(p, mu, sigma)` can be used to compute probabilities and inverse CDFs of the normal distribution, respectively.