CHAPTER FIVE

Continuous Random Variables

- Discrete:

How many shoes do you own?

- Continuous:

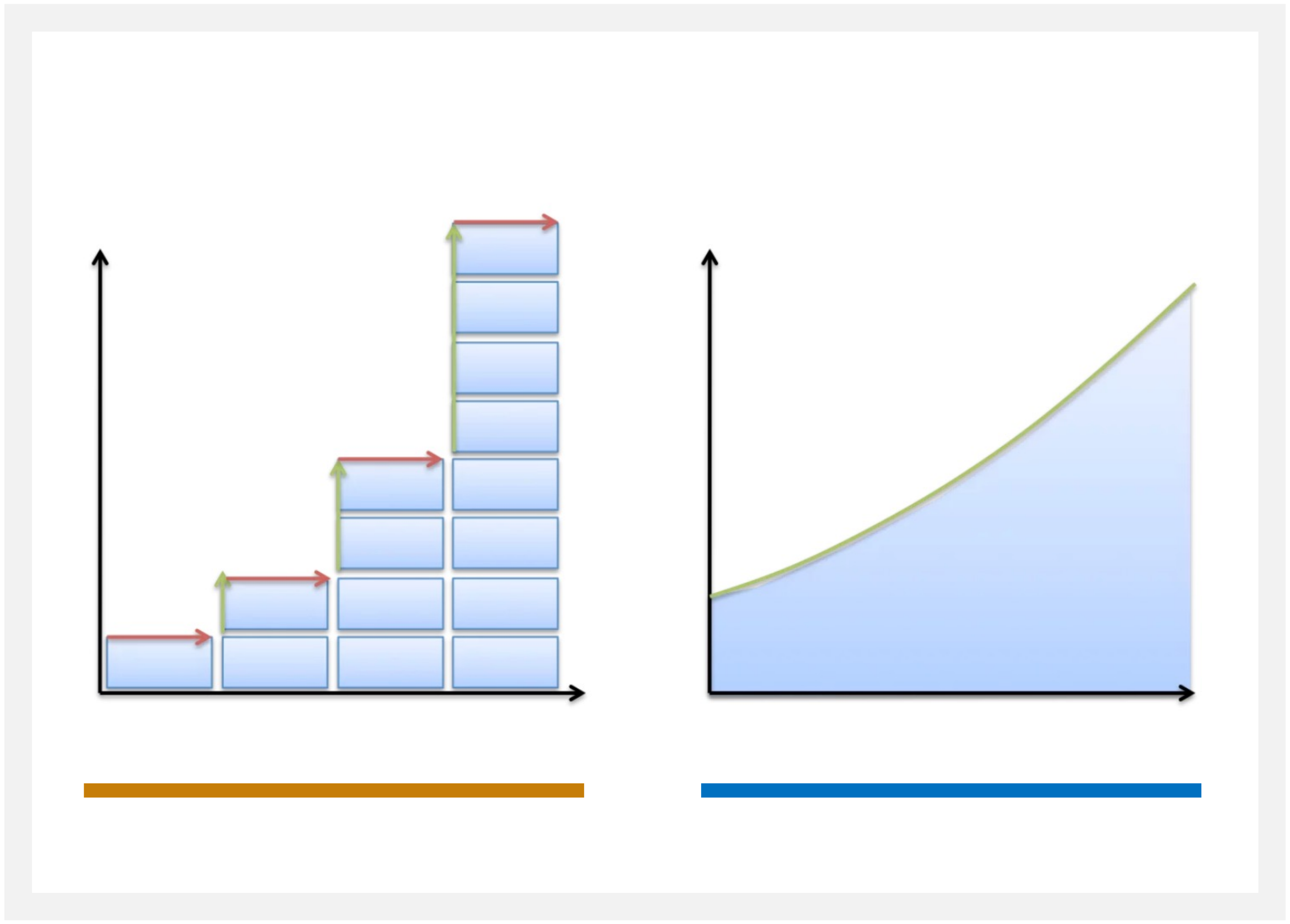
How tall are you?

- Discrete:

coin flip, dice, marbles, people, letter grade, …

- Continuous:

Pressure, height, mass, weight, density, volume, temperature, distance,…

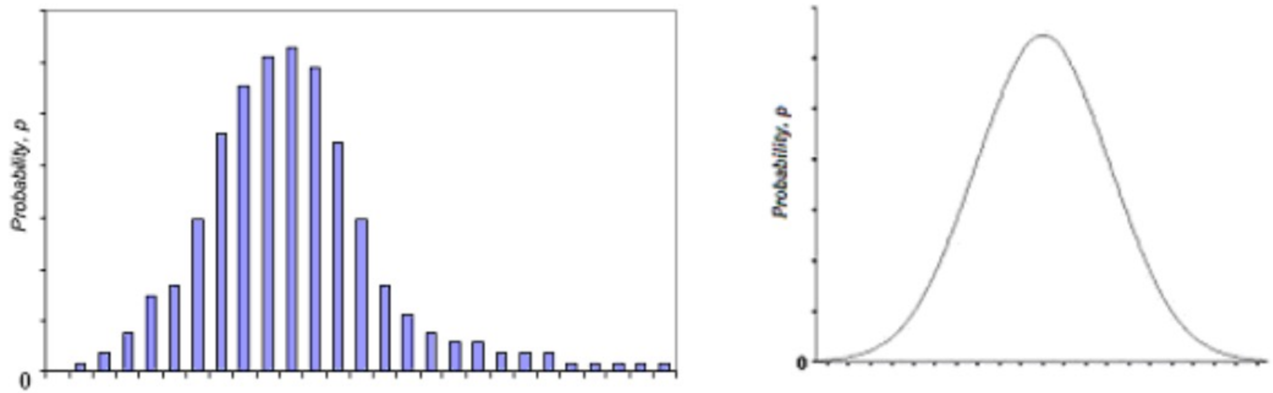


Rolling a die and considering the outcomes can be studied with a DISCRETE random Variable.

- i.e., There are a discrete quantity of outcomes: {1, 2, 3, 4, 5, 6}

- Loading a model bridge with weight to investigate structural integrity can be studied with a CONTINUOUS random variable.

- i.e., There are a continuous quantity of outcomes: $(0, \infty)$



(a) PMF for a discrete variable.   (b) PDF for a continuous variable

A random variable X is continuous if for some function $f: \mathbb{R} \to \mathbb{R}$ and for any numbers a and b with $a \leq b$,

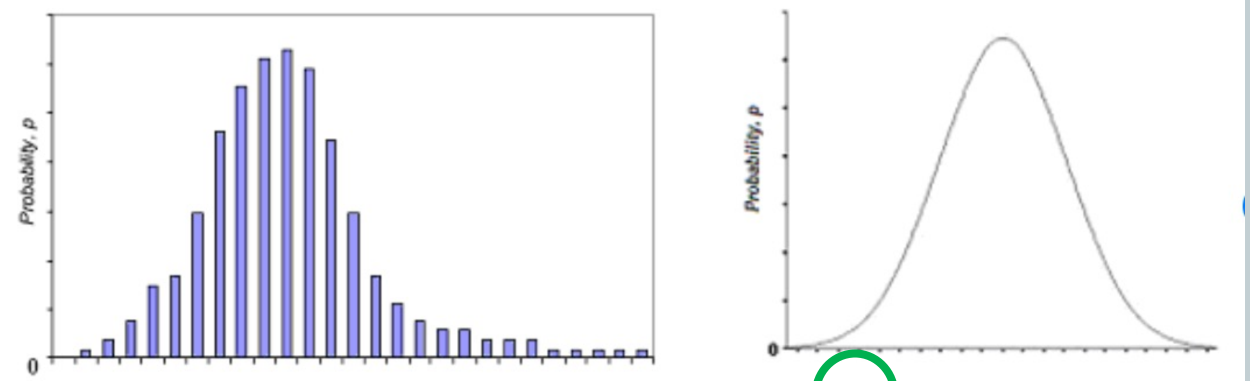
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

The function f has to satisfy:

[1] $f(x) \geq 0$ for all x

[2] $\int_{-\infty}^{\infty} f(x)dx = 1.$

We call f the probability density function, or probability density of X.



(a) PMF for a discrete variable.

(b) PDF for a continuous variable

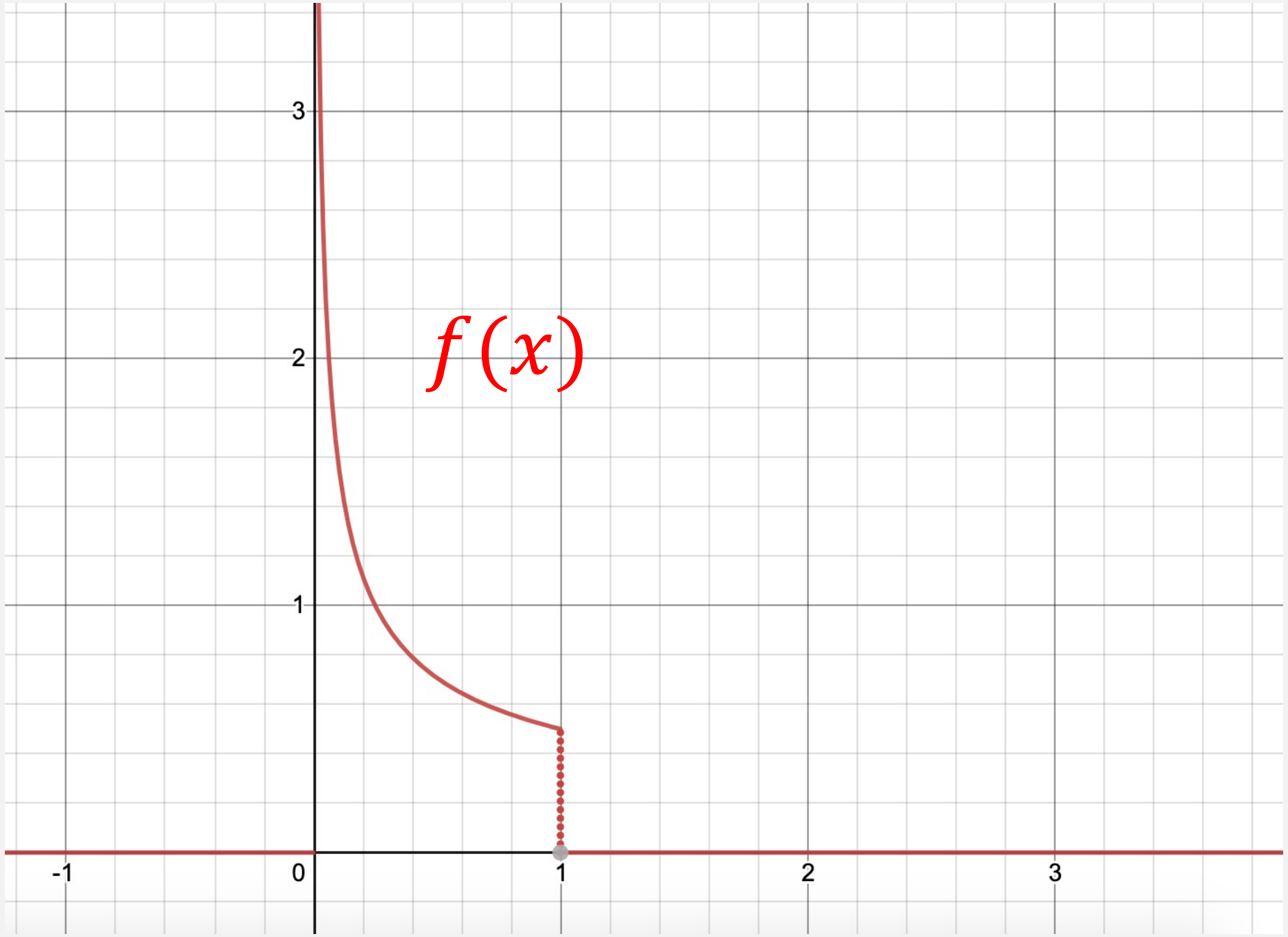
- Suppose $f(x)$

$$= \begin{cases} 0 \quad , & x \le 0 \text{ or } x \ge 1 \\ \dfrac{1}{2\sqrt{x}} \quad , & 0 < x < 1 \end{cases}$$

- Q: Is f a PDF?
- A: f has to satisfy:

  [1] $f(x) \ge 0$ for all x

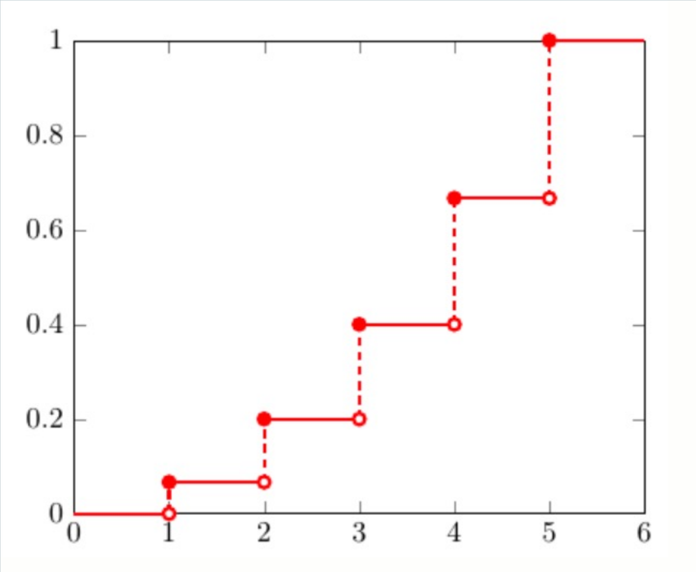
  [2] $\int_{-\infty}^{\infty} f(x)dx = 1.$

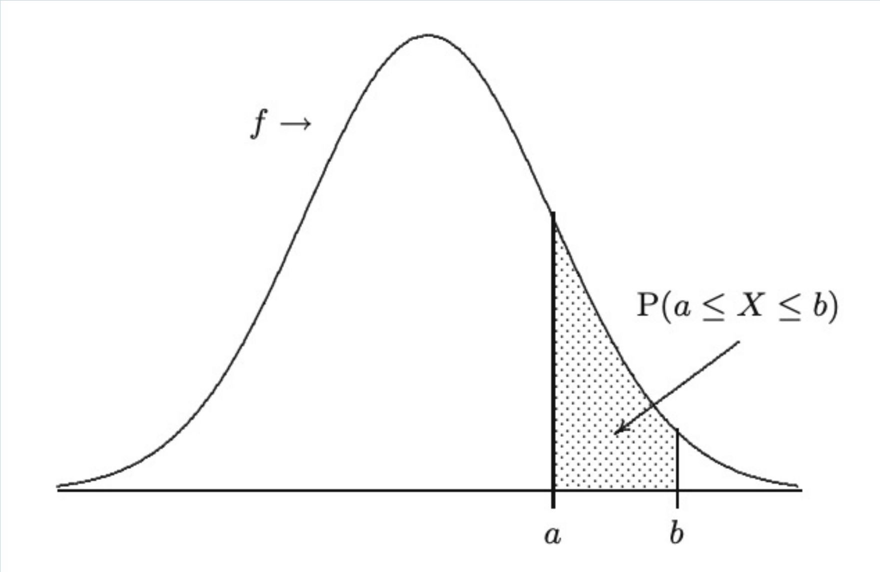


A discrete random variable does not have a probability density function, f.
A continuous random variable does not have a probability mass function, p.
But they both have a cumulative distribution function $F(a) = P(X \leq a)$.

In both cases (discrete and continuous), we can express the probability that X lies in an interval $(a, b]$ directly in terms of F:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

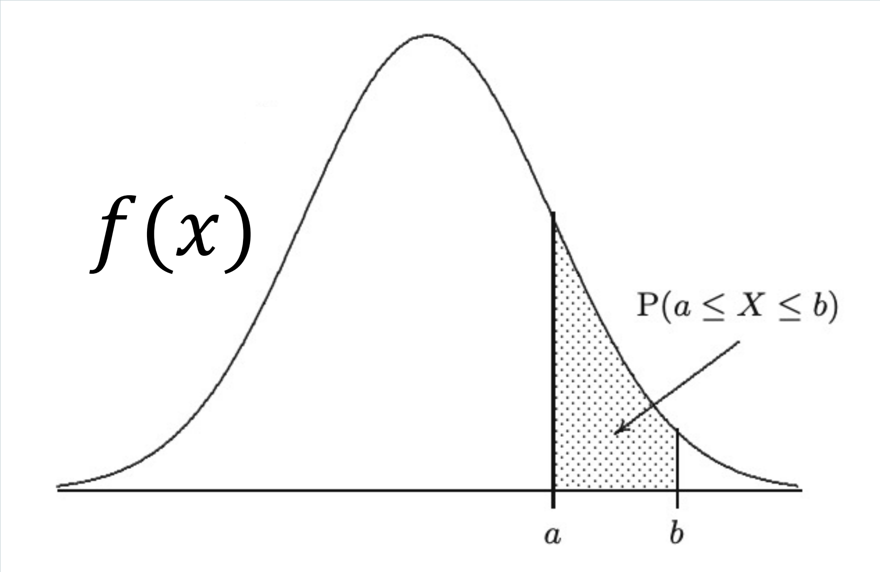




Q: What is the relation between the cumulative distribution function F and the probability density function f of a continuous random variable?

A:

$$F(b) = \int_{-\infty}^{b} f(x)dx \quad \text{and} \quad f(x) = \frac{d}{dx}F(x)$$



The probability that X lies in an interval $[a, b]$ is equal to the area under the probability density function f of X over the interval $[a, b]$.

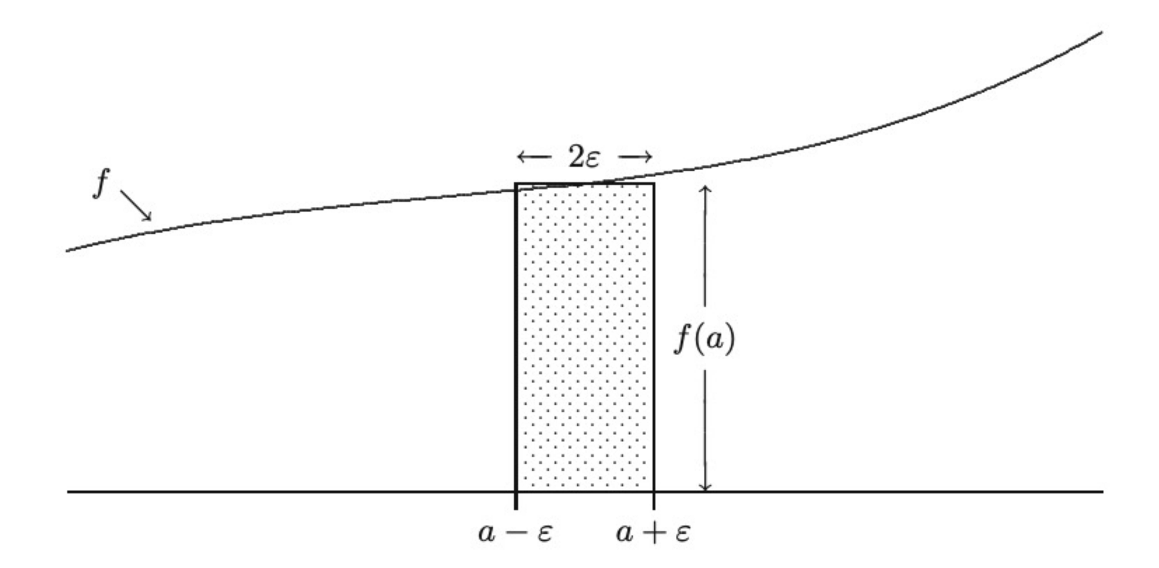
$$P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

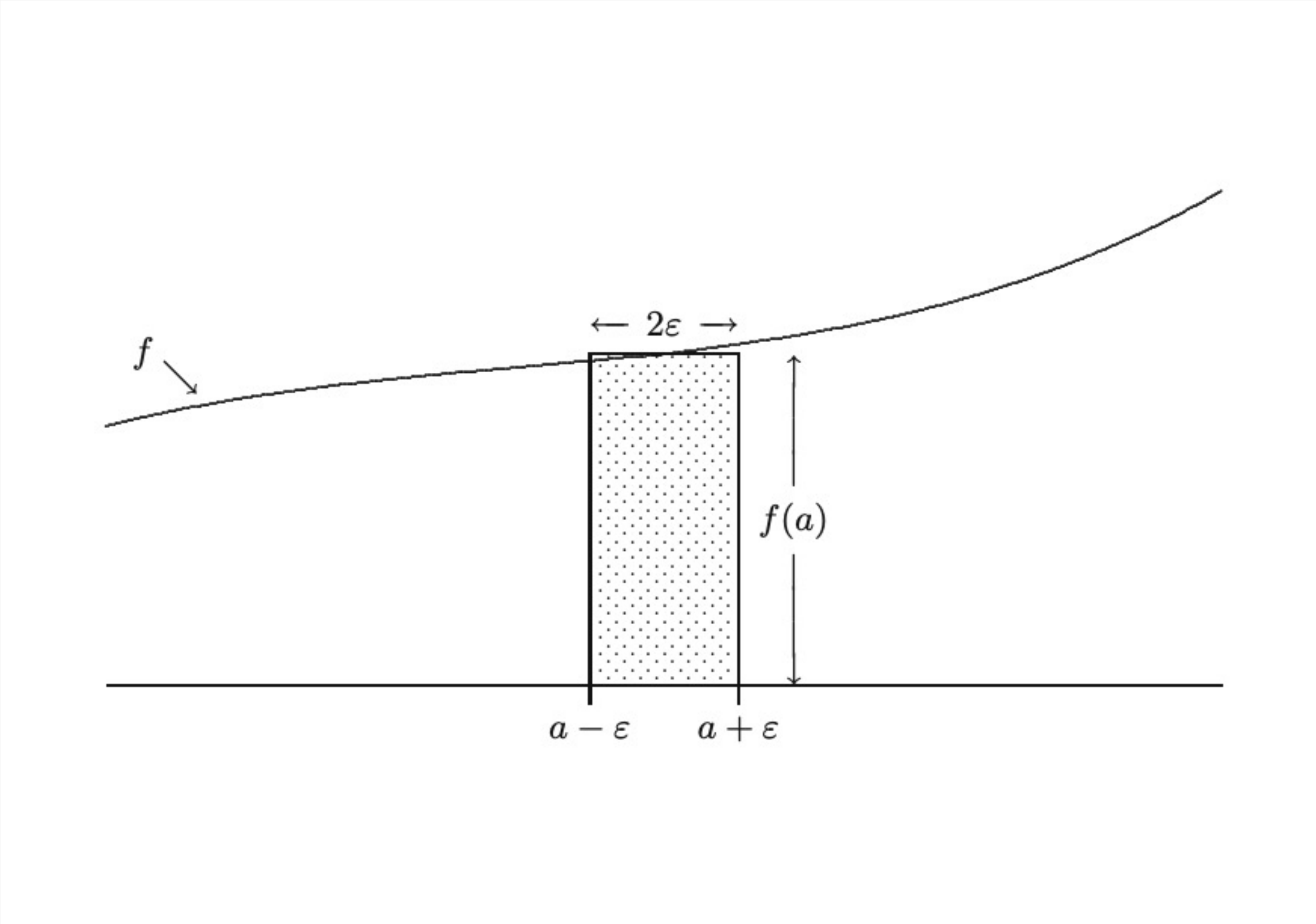
What happens as the interval gets smaller and smaller?

$$P(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x)dx$$

Consider $\varepsilon \to 0$, then $P(X = a) = 0$.

Therefore, $P(a \leq X \leq b) =$
$P(a < X \leq b) =$
$P(a < X < b) =$
$P(a \leq X < b)$.





$f(a)$ can be interpreted as a relative measure of how likely it is that X will be near a.

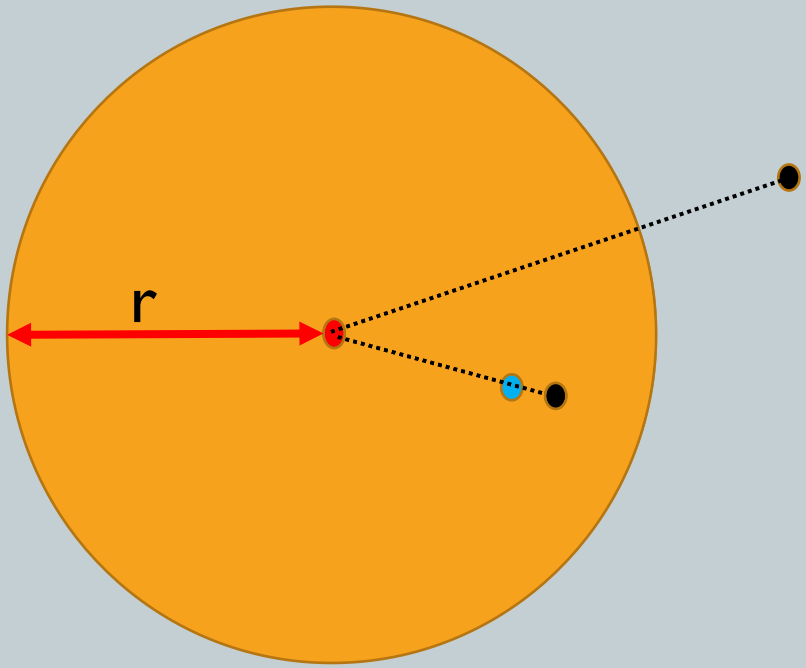
- Do not think of $f(a)$ as a probability though:

$f(a)$ can be arbitrarily large, it is not restricted to values inside of $[0, 1]$.

Experiment: We are looking at darts **that hit** a circular target of radius r. Consider distances, b, from the center of the target.

We are interested in the distance X between the hitting point and the center of the target. Distance cannot be negative, so for $b < 0$ we have $F(b) = P(X \leq b) = 0$.

Furthermore, since we are only looking at darts that actually hit the target, we have $F(b) = 1$ when $b > r$.



We can now discuss the probability that a dart is inside a particular region:

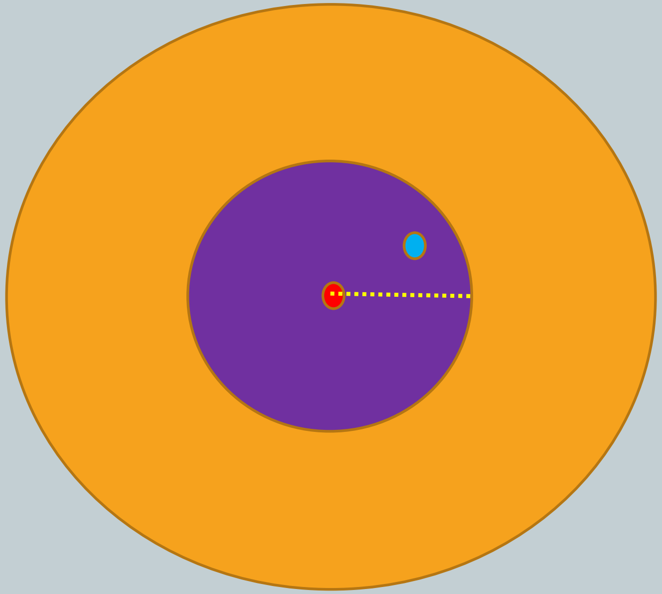
$$F(b) = P(X \leq b) = \frac{\pi b^2}{\pi r^2} = \frac{b^2}{r^2} \quad \text{for } \; 0 \leq b \leq r.$$

What is the probability that $0 < X \leq \frac{r}{2}$ and the probability that $\frac{r}{2} < X \leq r$?

ANS: $F\left(\frac{r}{2}\right) - F(0) = {}^{1}/_{4}$

$F(r) - F\left(\frac{r}{2}\right) = {}^{3}/_{4}$

regardless of radius.



Also, if $F(b) = P(X \leq b) = \frac{\pi b^2}{\pi r^2} = \frac{b^2}{r^2}$ for $0 \leq b \leq r$,

then the probability density function f of X is equal to 0 outside the interval $[0, r]$ and $f(x) = \frac{d}{dx}F(x) = \frac{1}{r^2}\frac{d}{dx}x^2 = \frac{2x}{r^2}$ for $0 \leq x \leq r$.

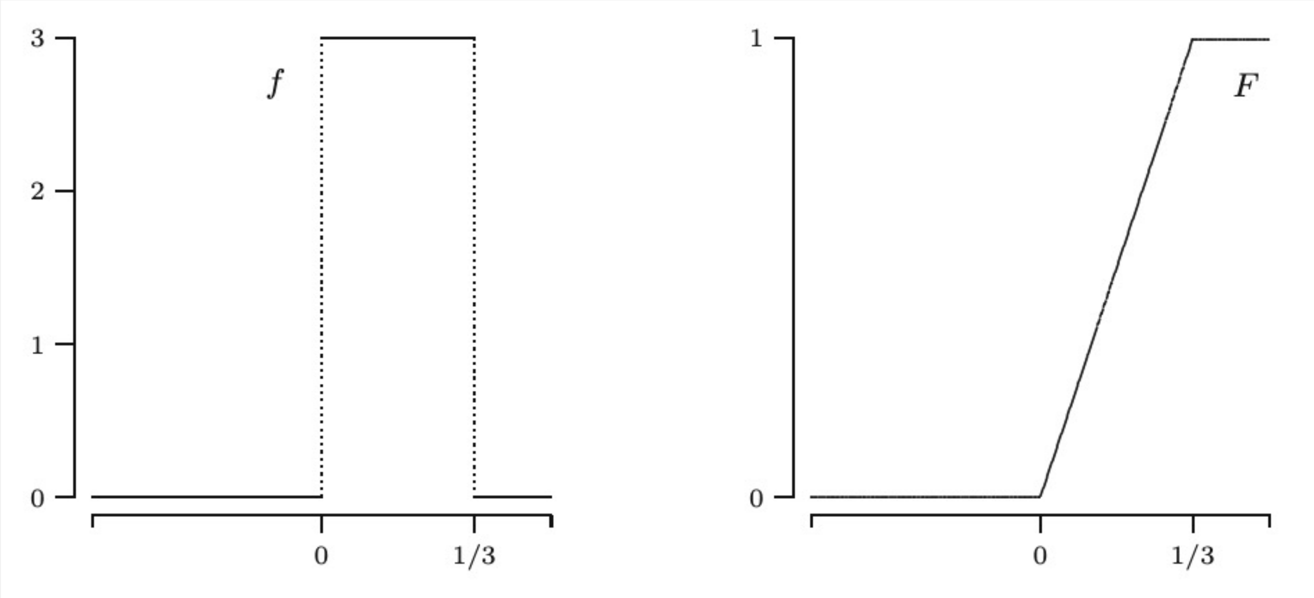
The pdf, $f(a)$, gives you the "relative likelihood of a continuous random variable taking that value", i.e., the relative measure of how likely it is that X will be near a.

But do not think of $f(a)$ as a probability because $f(a)$ can be arbitrarily large.

A continuous random variable has a **uniform distribution** on the interval $[\alpha, \beta]$ if its probability density function f is given by:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{for } \alpha \leq x \leq \beta. \\ 0 & \text{otherwise} \end{cases}$$

We denote this by $U(\alpha, \beta)$.



$$U\left(0, \frac{1}{3}\right)$$

Experiment:

You arrive into a building and are about to take an elevator to your floor.
Once you call the elevator, it will take between 0 and 40 seconds to arrive to you
We will assume that the elevator arrives uniformly between 0 and 40 seconds after you press the button.

Q: What is the probability that the elevator takes less than 15 seconds to arrive ?

A: $P(0 \leq x \leq 15) = \int_0^{15} \frac{1}{40-0} dx = \frac{x}{40}\Big|_0^{15} = \frac{15}{40} - \frac{0}{40} = \frac{15}{40} = \frac{3}{8} = 0.375$

Because the PDF is $f(x) = \frac{1}{\beta - \alpha} = \frac{1}{40-0}$

A continuous random variable has an

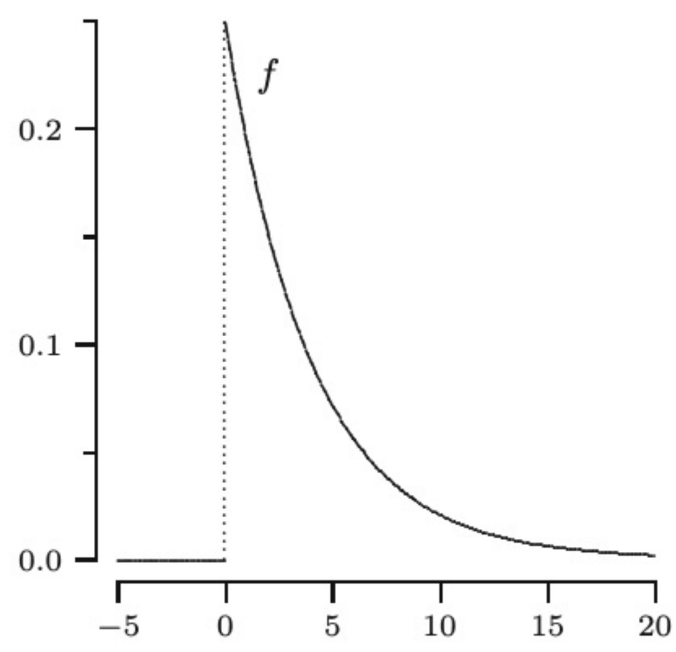
exponential distribution

with parameter λ if its probability density function f is given by:

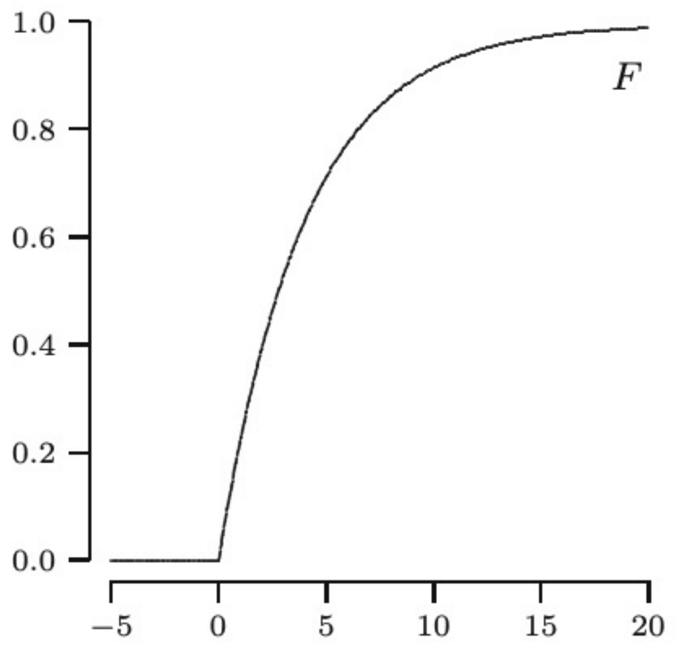
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

This is denoted $Exp(\lambda)$.

Probability Density Function



Cumulative Distribution Function



$Exp(0.25)$

Notice that when the probability density function f is given we can get the cumulative distribution function F via integration:

$$f(x) = \lambda e^{-\lambda x} \text{ implies } F(x) = \int \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda x}$$

So, $F(a) = 1 - e^{-\lambda a}$ for $a \geq 0$.

The exponential distribution: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$

The PDF is $f(x)$ and the CDF is $F(x)$.

$F(a) = \int_{-\infty}^{a} f(x)dx$ and $f(x) = \frac{d}{dx}F(x)$

Therefore,

$F(a) = \int_{-\infty}^{a} f(x)dx = \int_{0}^{a} \lambda e^{-\lambda x}\, dx$

$$u = -\lambda x \implies du = -\lambda\, dx \implies F(a) = 1 - e^{-\lambda a}$$

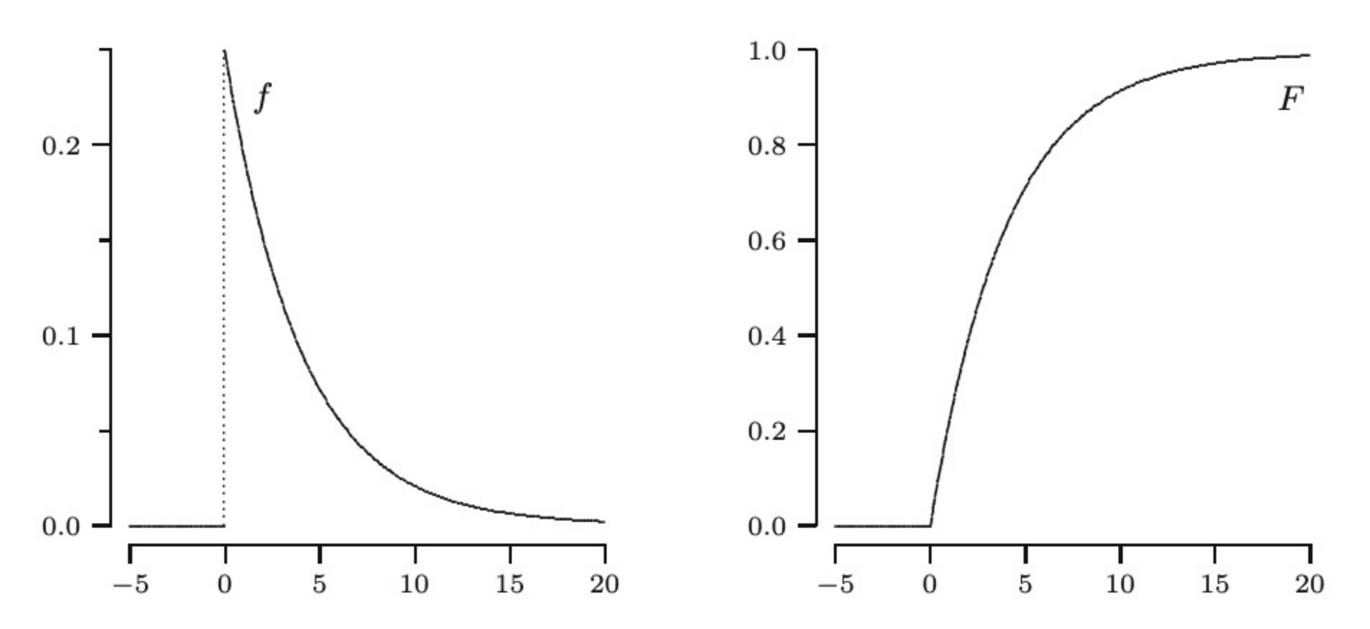
An investigation of the response time of a certain computer system yields that the response time in seconds has an exponentially distributed time with parameter 0.25.

Q: What is the probability that the response time exceeds 5 seconds?

A: If X is the response time, we ask for $P(X > 5)$.
$F(a) = 1 - e^{-\lambda a}$ is used to determine $P(X \leq a)$.

$$P(X > 5) = 1 - P(X \leq 5) = 1 - F(5) = 1 - (1 - e^{-0.25 \cdot 5})$$

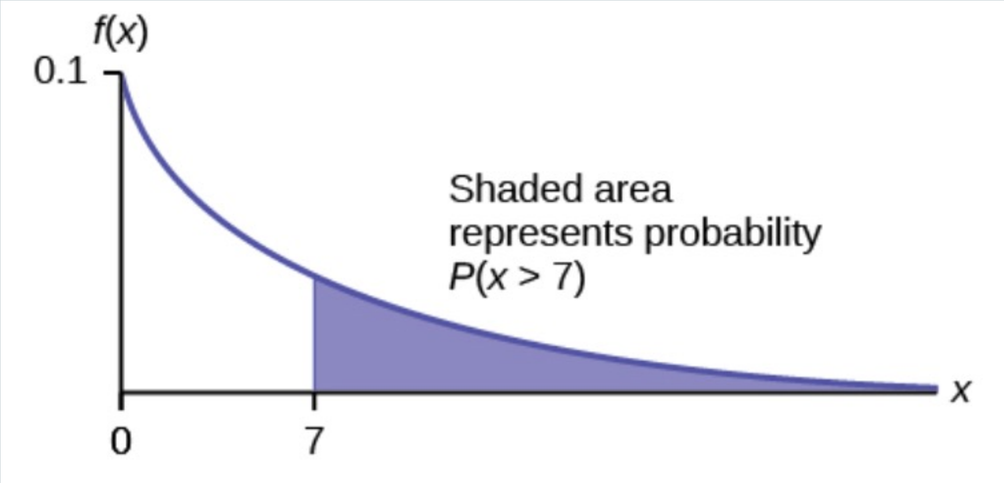
$$\begin{aligned} P(X > 5) &= e^{-0.25 \cdot 5} \\ &= e^{-1.25} \\ &= 0.2865 \ldots \end{aligned}$$



On the average, a certain computer part lasts ten years. The length of time the computer part lasts is exponentially distributed, with parameter $\lambda = 0.1$.

Q: What is the probability that a computer part lasts more than 7 years?

A: X is the amount of time (in years) a computer part lasts.
We need $P(X > 7) = 1 - P(X < 7) = e^{-0.1 \cdot 7} = 0.4966$



Let X = amount of time (in minutes) a postal clerk spends with a customer.
The time is known from historical data to have an average amount of time equal to four minutes, with parameter $\lambda = 0.25$.

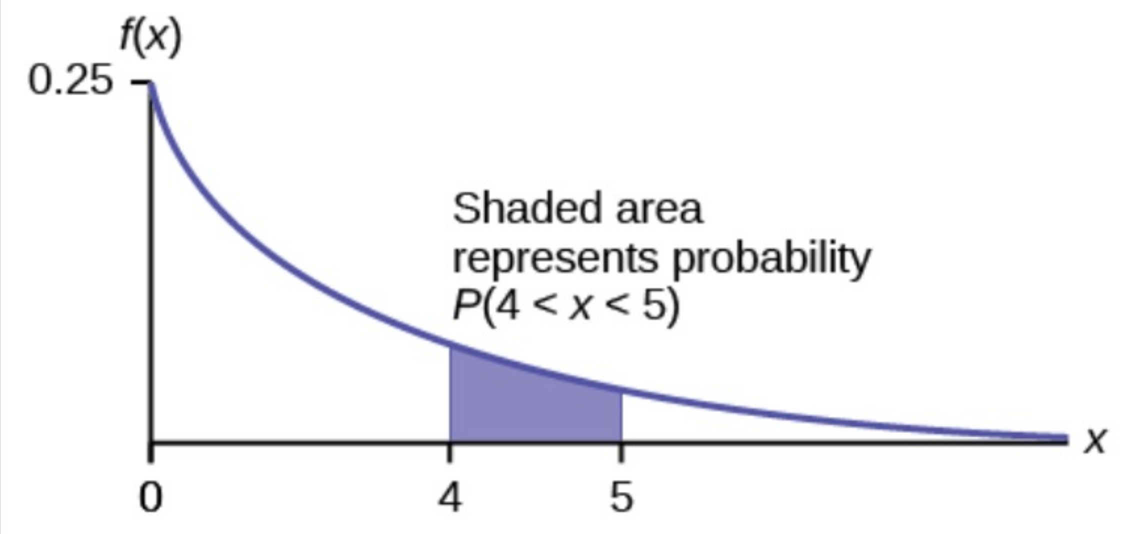
Q: What is the probability that a clerk spends four to five minutes with a randomly selected customer?

A: We want $P(4 < X < 5)$.

$$F(5) = P(X < 5) = 1 - e^{-0.25 \cdot 5} = 0.7135$$

$$F(4) = P(X < 4) = 1 - e^{-0.25 \cdot 4} = 0.6321$$

$$P(4 < X < 5) = 0.7135 - 0.6321 = 0.0814$$

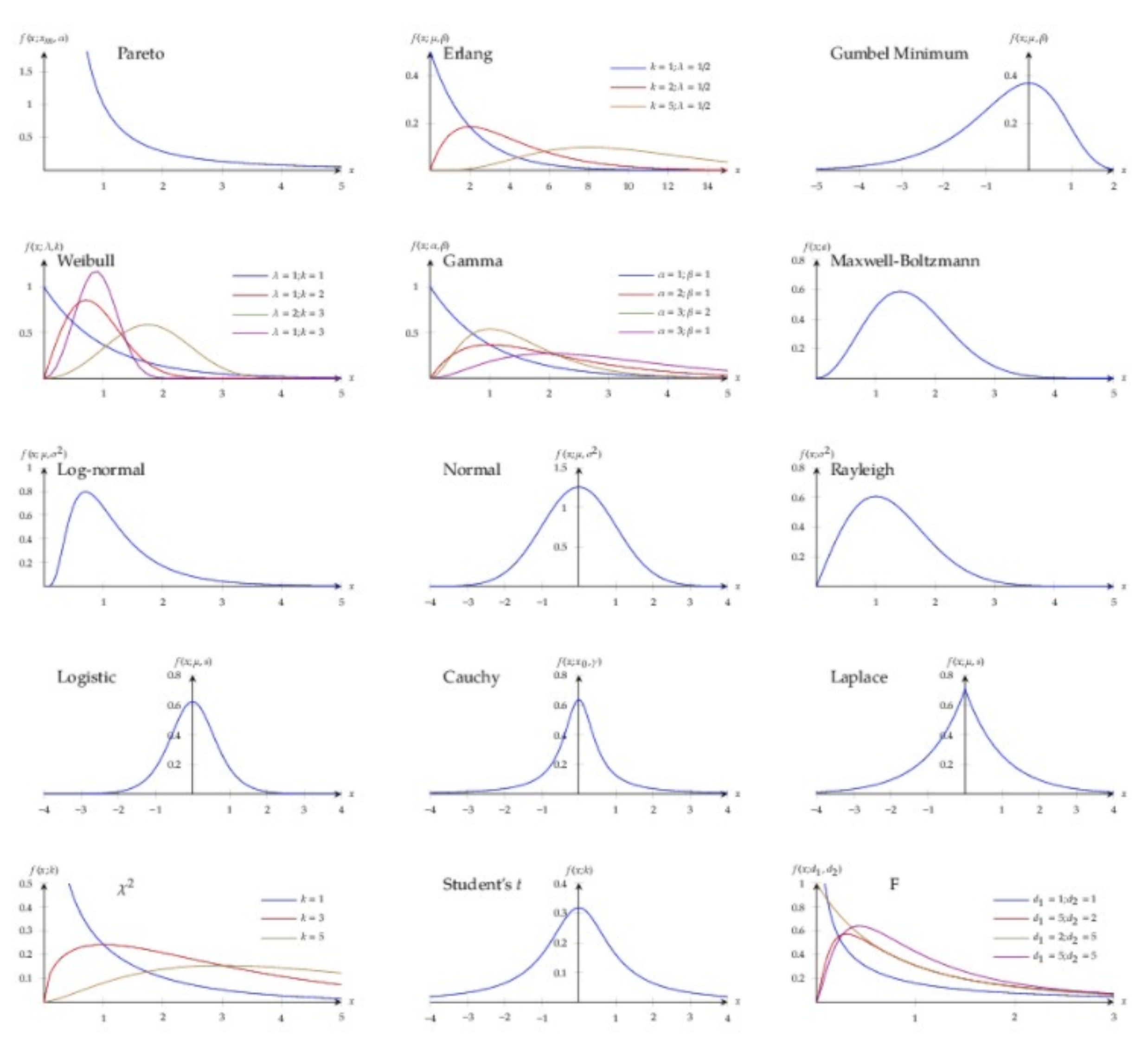


- We have seen a number of distributions, both discrete and continuous.

- Now, we are introduced to the normal distribution.

- The normal plays a central role in probability theory and statistics.

- The normal is an important tool we will use in approximating throughout the course.

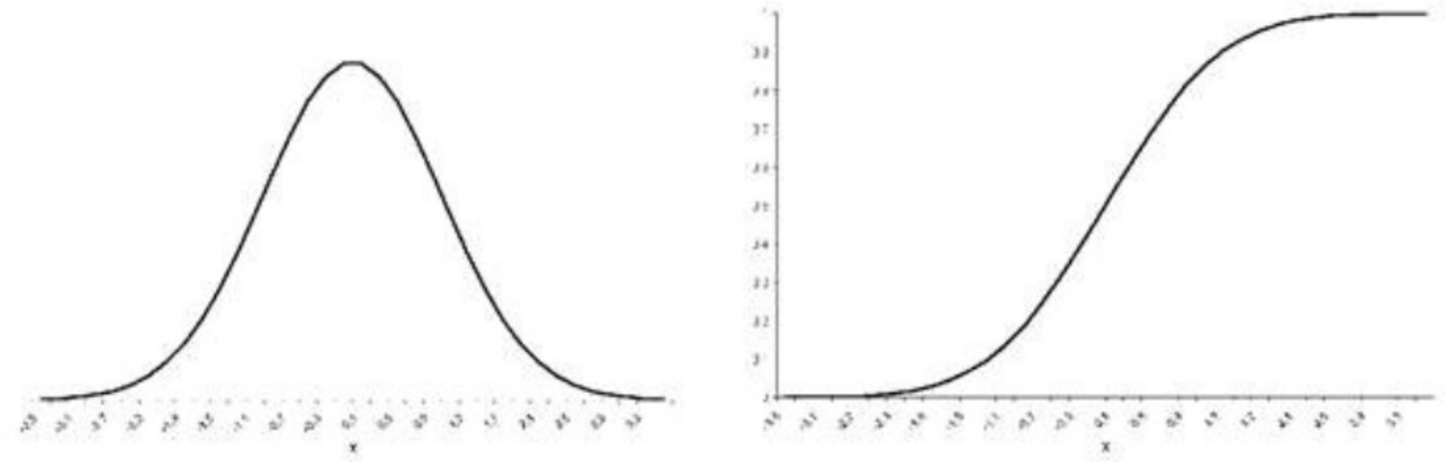


- A continuous random variable has a **normal distribution** with parameters μ and $\sigma^2 > 0$ if its probability density function f is given by:

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

for $-\infty < x < \infty$.

This is denoted $N(\mu, \sigma^2)$.



a) Normal Probability Density Function

b) Normal Cumulative Distribution Function

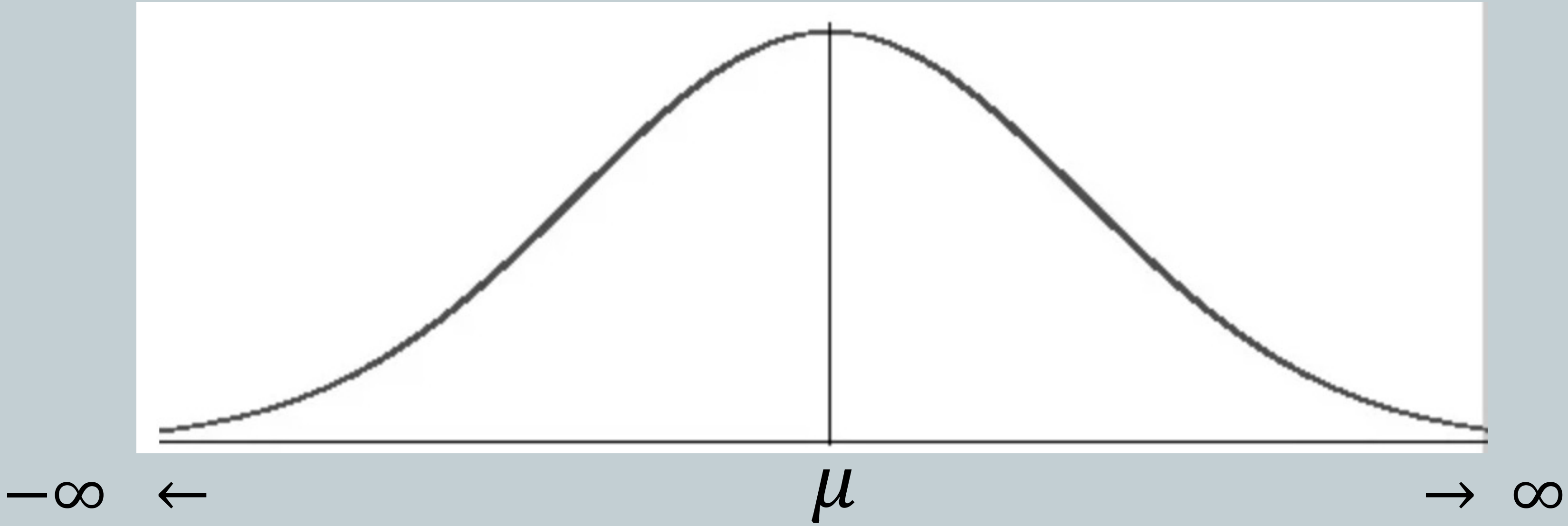
Suppose X has an $N(\mu, \sigma^2)$ distribution, then its CDF is given by:

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \text{ for } -\infty < a < \infty.$$

Unfortunately, f doesn't have an antiderivative!
Therefore, there is no explicit expression for F.

Luckily though, for any μ and σ^2, a $N(\mu, \sigma^2)$ distribution can easily be changed to a $N(0,1)$ distribution.

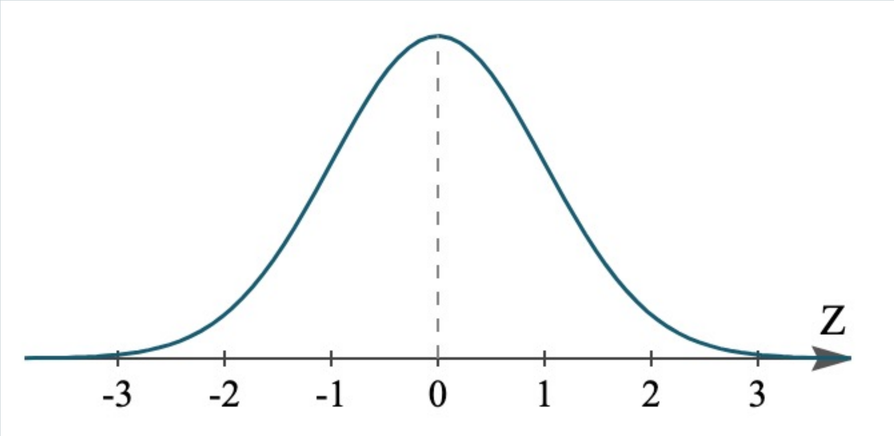
Area under the curve = 1



Consider a Normal Distribution with any given μ and σ: $N(\mu, \sigma^2)$

This distribution can be changed into a Standard Normal Distribution:

$z = \frac{x - \mu}{\sigma}$ or $x = \sigma z + \mu$.



$x - \mu$ centers the distribution at 0.

$\frac{x - \mu}{\sigma}$ scales to a standard deviation of 1.

The $N(\mu, \sigma^2)$ distribution has been changed to a $N(0, 1)$ distribution.

The $N(0,1)$ distribution is so important that it gets its own name and symbol!
The $N(0,1)$ distribution is called the **Standard Normal Distribution, $\phi(x)$**.

The Standard Normal Distribution, $\phi(x)$

Since $N(0,1)$ means $\mu = 0$ and $\sigma = 1$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ becomes:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

for $-\infty < x < \infty$.

Instead of PDF f, we have PDF ϕ.
Instead of CDF F, we have CDF Φ.

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

for $-\infty < a < \infty$.

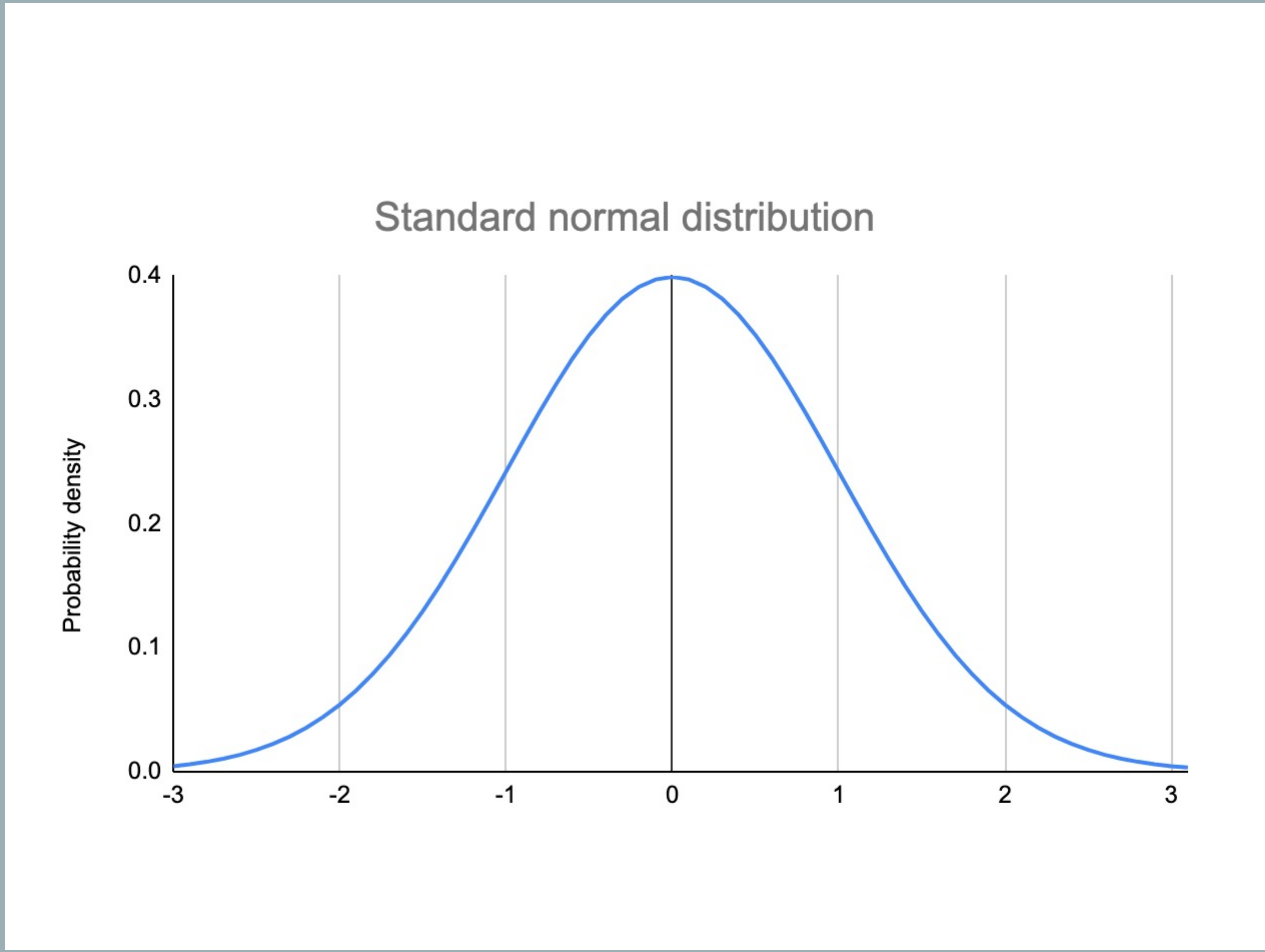
NOTICE

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

IS SYMMETRIC ABOUT 0.

$$\phi(-x) = \phi(x)$$

FOR EACH x.



Since $\phi(x)$ is symmetric about 0, we know 50% of the mass is on the negative axis. Therefore, if the random variable Z has a standard normal distribution, then we know right away $P(Z \leq 0.75) > 0.5$.

When coding, Python provide answers to standard normal distribution questions.

```
import scipy.stats as stats
stats.norm.cdf()
stats.norm.pdf()
stats.norm.ppf()
```

However, if you use a chart to find answers by hand, then be aware of how the chart lists area. Different books/charts list area differently.

For instance, your book uses a chart that does not contain the values of $\Phi(a)$ but rather $1 - \Phi(a)$, aka right-tail probabilities.

Suppose the random variable Z has a standard normal distribution.

- $P(Z \geq -0.75) = ?$

- A: Table B1 shows the number 2266.

The table lists a right-tail distribution.

Therefore,

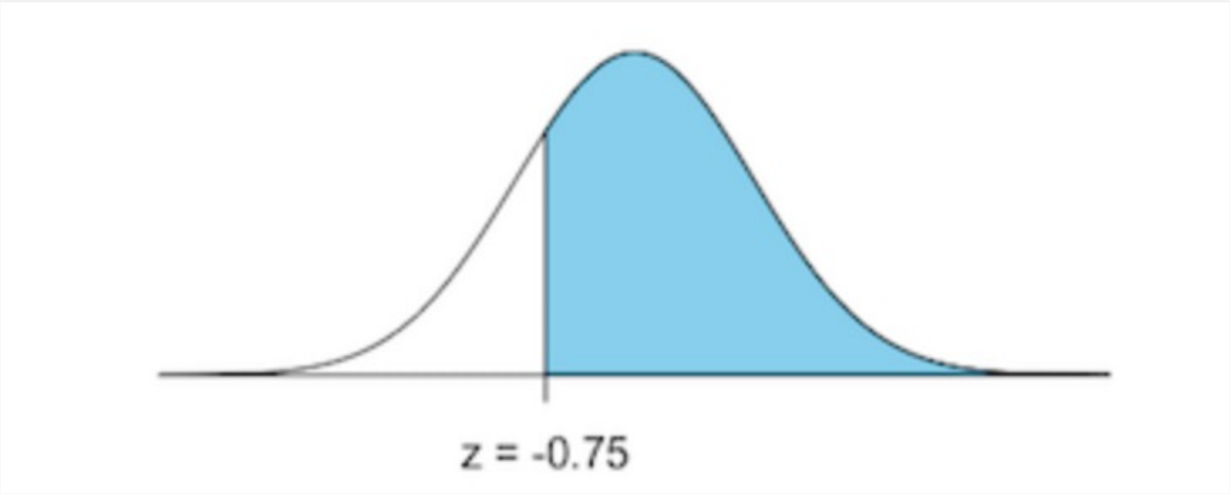
$$P(Z \geq -0.75) = P(Z \leq 0.75) = 1 - 0.2266 = 0.7734$$

aka $\Phi(0.75) = 0.7734$

```
scipy.stats.norm.cdf(0.75)
```


Table B.1. Right tail probabilities $1 - \Phi(a) = P(Z \geq a)$ for an $N(0,1)$ distributed random variable Z.

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611

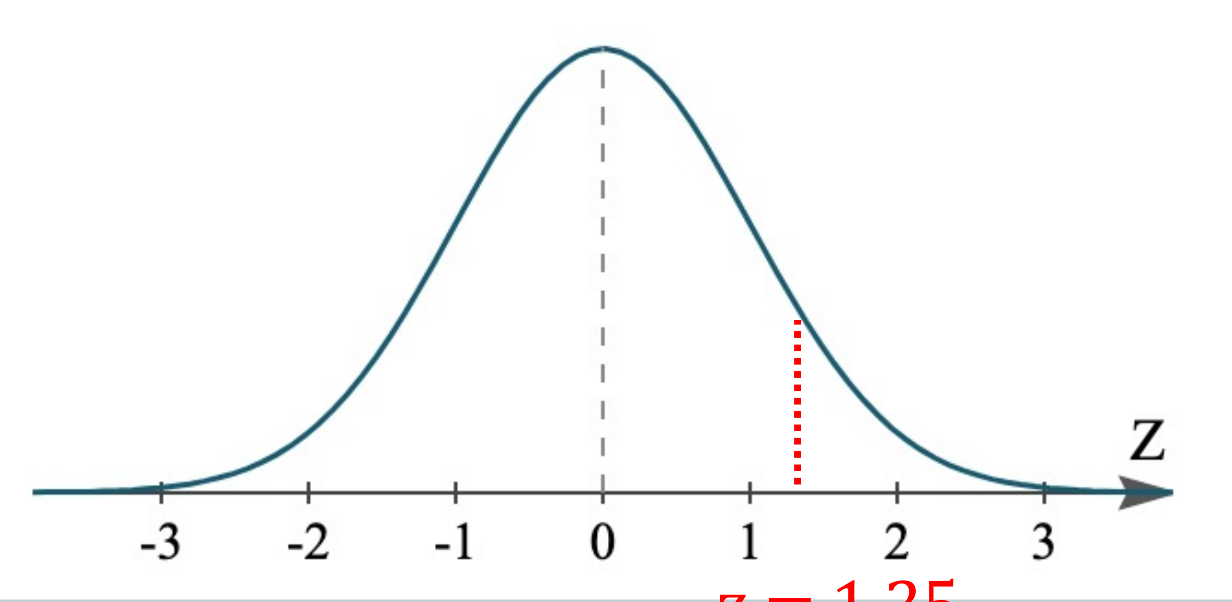


What is $P(Z \leq 1.25)$ on the Standard Normal Distribution?

Chart: 1.25 corresponds to 1056, implies $\Phi(1.25) = 1 - 0.1056 = 0.8944$

Python: `phi = stats.norm.cdf(1.25) = 0.89435`

a	0	1	2	3	4	5	6
0.0	5000	4960	4920	4880	4840	4801	4761
0.1	4602	4562	4522	4483	4443	4404	4364
0.2	4207	4168	4129	4090	4052	4013	3974
0.3	3821	3783	3745	3707	3669	3632	3594
0.4	3446	3409	3372	3336	3300	3264	3228
0.5	3085	3050	3015	2981	2946	2912	2877
0.6	2743	2709	2676	2643	2611	2578	2546
0.7	2420	2389	2358	2327	2296	2266	2236
0.8	2119	2090	2061	2033	2005	1977	1949
0.9	1841	1814	1788	1762	1736	1711	1685
1.0	1587	1562	1539	1515	1492	1469	1446
1.1	1357	1335	1314	1292	1271	1251	1230
1.2	1151	1131	1112	1093	1075	$\boxed{1056}$	1038
1.3	0968	0951	0934	0918	0901	0885	0869



z = 1.25

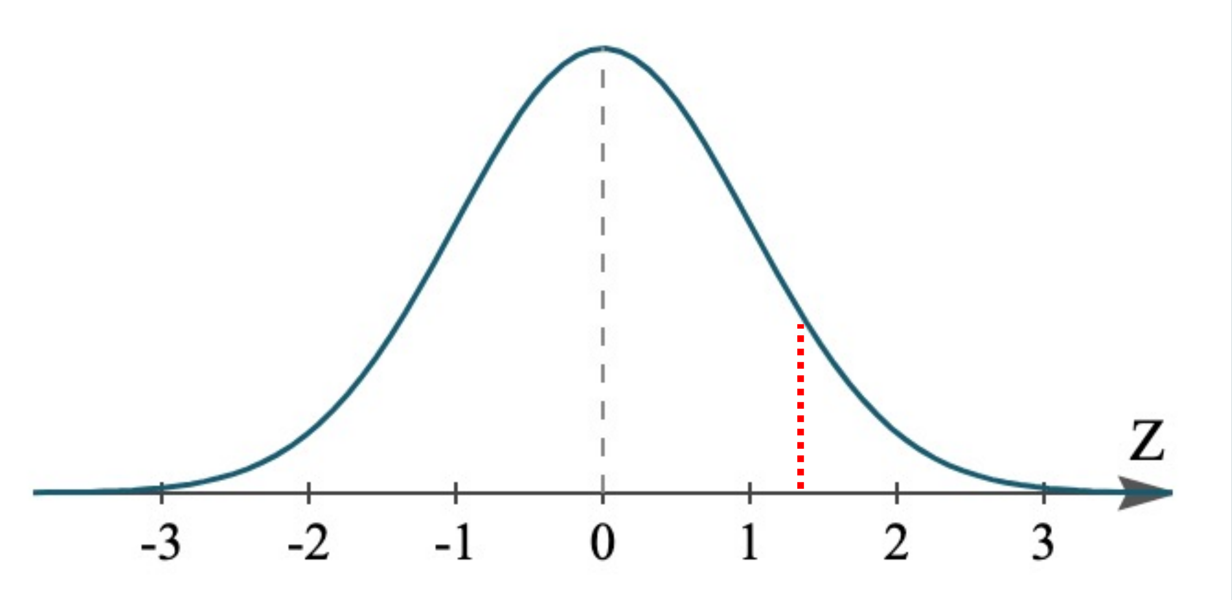
What is $P(Z \geq 1.25)$ on the Standard Normal Distribution?

Chart: 1.25 corresponds to 1056, implies $\mathrm{P}(Z \geq 1.25) = 0.1056$

$$P(Z \geq 1.25) = 1 - P(Z \leq 1.25) = P(Z < -1.25)$$

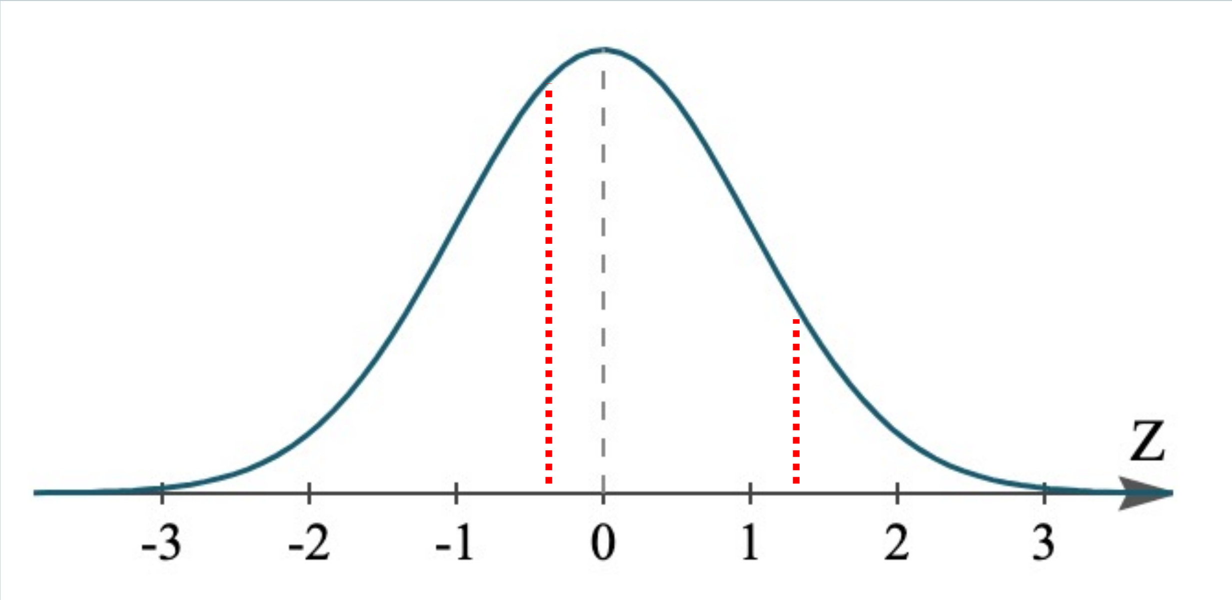
Python: `phi = stats.norm.cdf(-1.25) = 0.10564977`

a	0	1	2	3	4	5	6
0.0	5000	4960	4920	4880	4840	4801	4761
0.1	4602	4562	4522	4483	4443	4404	4364
0.2	4207	4168	4129	4090	4052	4013	3974
0.3	3821	3783	3745	3707	3669	3632	3594
0.4	3446	3409	3372	3336	3300	3264	3228
0.5	3085	3050	3015	2981	2946	2912	2877
0.6	2743	2709	2676	2643	2611	2578	2546
0.7	2420	2389	2358	2327	2296	2266	2236
0.8	2119	2090	2061	2033	2005	1977	1949
0.9	1841	1814	1788	1762	1736	1711	1685
1.0	1587	1562	1539	1515	1492	1469	1446
1.1	1357	1335	1314	1292	1271	1251	1230
1.2	1151	1131	1112	1093	1075	$\boxed{1056}$	1038
1.3	0968	0951	0934	0918	0901	0885	0869



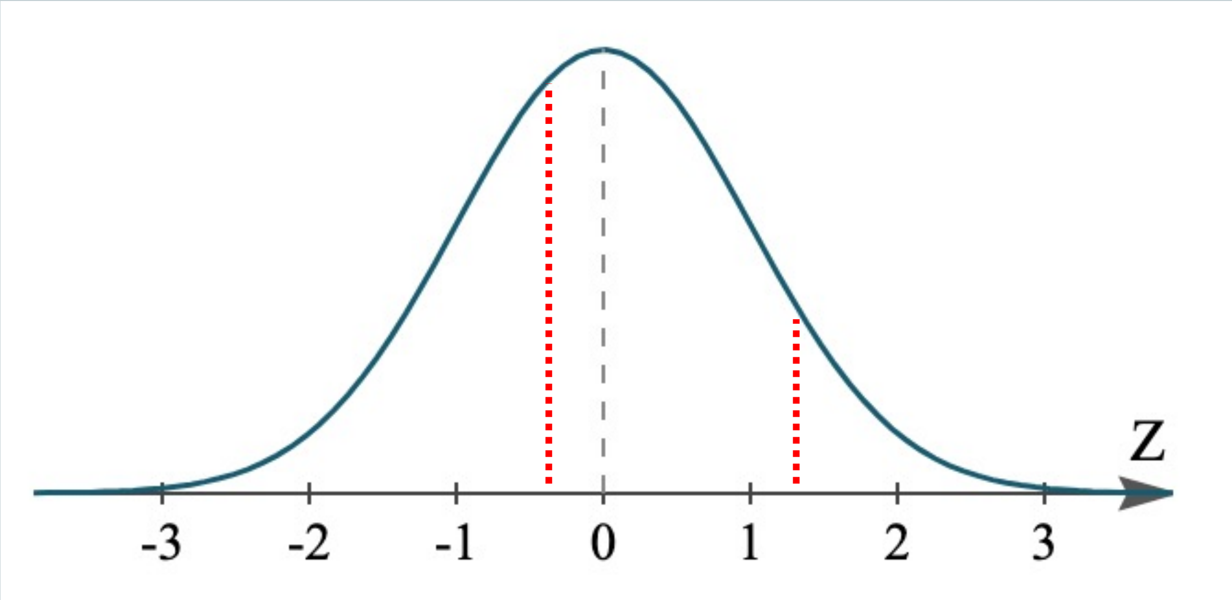
$z = 1.25$

$$P(-0.38 \leq Z \leq 1.25) = ?$$



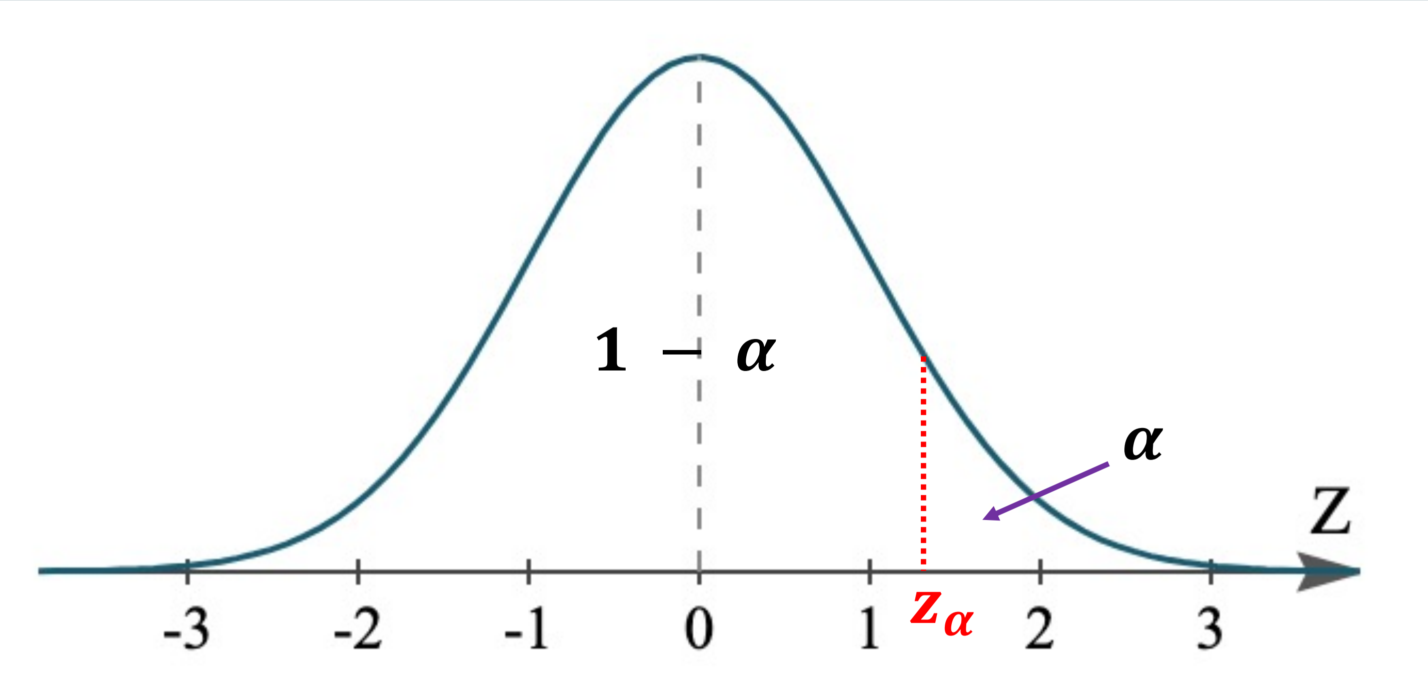
$P(-0.38 \ \leq Z \leq 1.25) = ?$

A: $P(Z \leq 1.25) - P(Z \leq -0.38) = 0.54237751875 \ldots$



$$\Phi(z_\alpha) = \int_{-\infty}^{z_\alpha} f(z)dz = 1 - \alpha = P(Z \leq z_\alpha)$$

$$P(Z \geq z_\alpha) = \alpha = 1 - P(Z \leq z_\alpha) = 1 - \Phi(z_\alpha)$$



z_α is the $100(1 - \alpha)^{\text{th}}$ percentile.

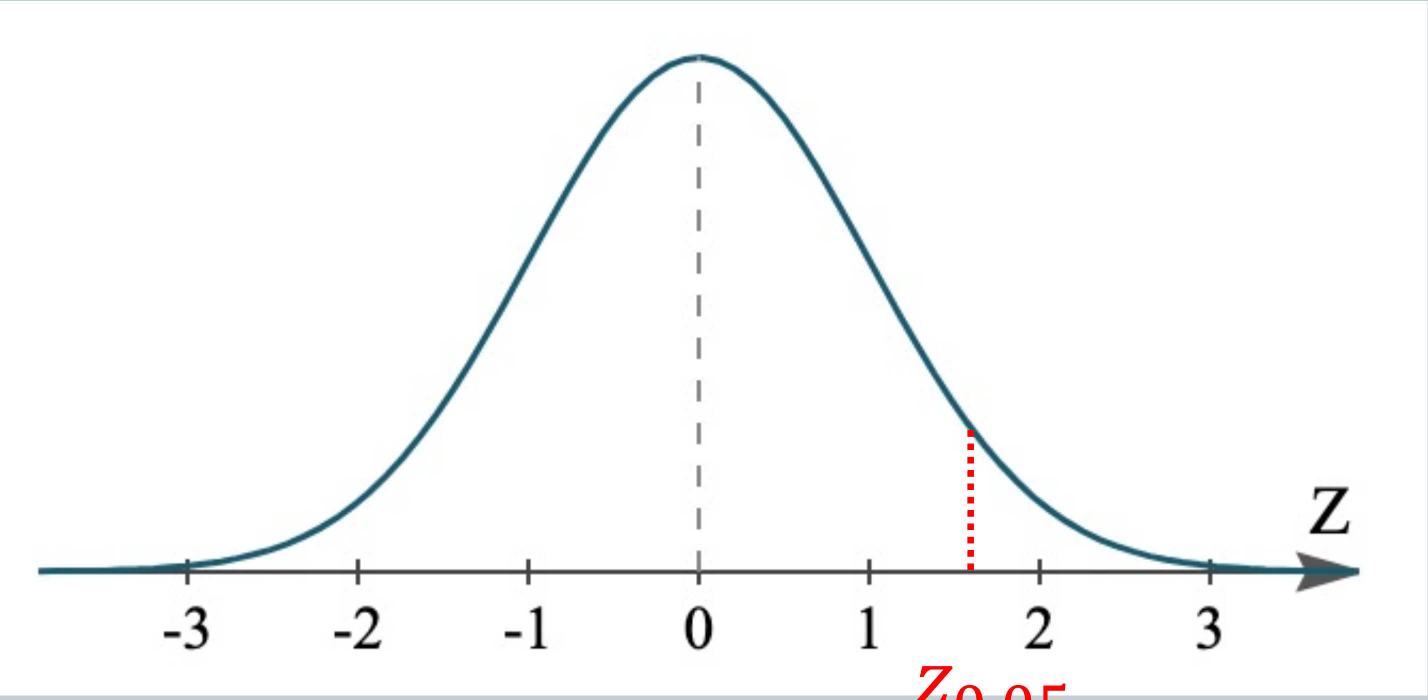
On the standard normal distribution, $z_{0.05} = 1.645$ (interpolate)

$z_{0.05}$ is the 95^{th} percentile.

$\Phi(1.645) = 0.95$

```
stats.norm.ppf(0.95) = 1.6448536269514722…
```

a	0	1	2	3	4	5	6	7
0.0	5000	4960	4920	4880	4840	4801	4761	4721
0.1	4602	4562	4522	4483	4443	4404	4364	4325
0.2	4207	4168	4129	4090	4052	4013	3974	3936
0.3	3821	3783	3745	3707	3669	3632	3594	3557
0.4	3446	3409	3372	3336	3300	3264	3228	3192
0.5	3085	3050	3015	2981	2946	2912	2877	2843
0.6	2743	2709	2676	2643	2611	2578	2546	2514
0.7	2420	2389	2358	2327	2296	2266	2236	2206
0.8	2119	2090	2061	2033	2005	1977	1949	1922
0.9	1841	1814	1788	1762	1736	1711	1685	1660
1.0	1587	1562	1539	1515	1492	1469	1446	1423
1.1	1357	1335	1314	1292	1271	1251	1230	1210
1.2	1151	1131	1112	1093	1075	1056	1038	1020
1.3	0968	0951	0934	0918	0901	0885	0869	0853
1.4	0808	0793	0778	0764	0749	0735	0721	0708
1.5	0668	0655	0643	0630	0618	0606	0594	0582
1.6	0548	0537	0526	0516	0505	0495	0485	0475
1.7	0446	0436	0427	0418	0409	0401	0392	0384
1.8	0359	0351	0344	0336	0329	0322	0314	0307



$z_{0.05}$

Q: What is another name for the $z_{0.5}$?

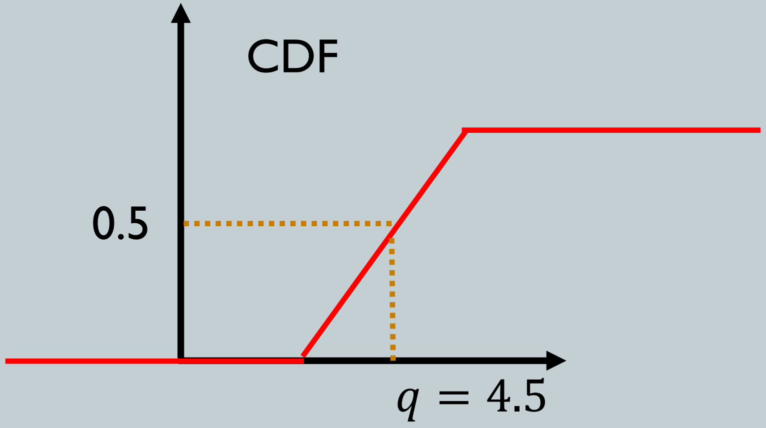
Q: What is another name for the $z_{0.5}$?

A: The 50^{th} percentile, or the median.

In general, for any distribution:

Let X be a continuous random variable and let p be a number between 0 and 1. The $100p^{th}$ percentile (aka p^{th} quantile) of the distribution of X is the smallest number q_p such that $F(q_p) = P(X \leq q_p) = p$.

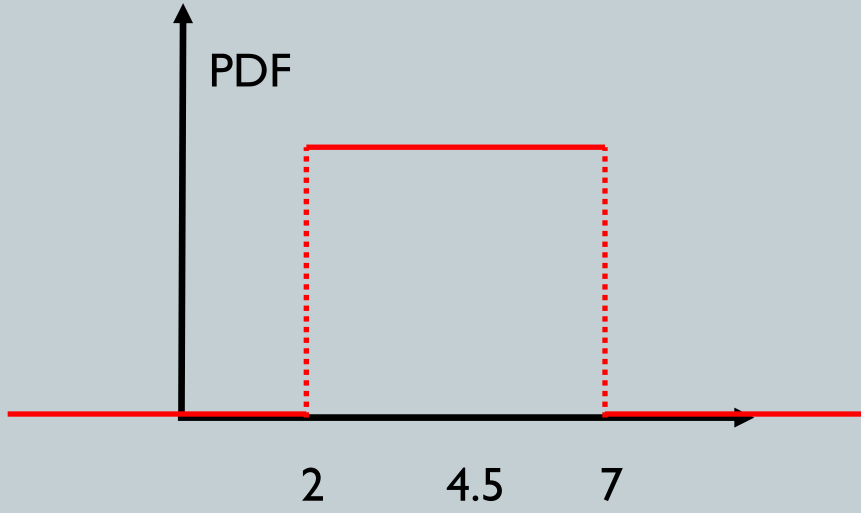
i.e., The median of a distribution is its 50^{th} percentile.



Example: What is the median of a $U(2, 7)$ distribution?

A: We know $F(q_{0.5}) = 0.5$ or $q_{0.5} = F^{inv}(0.5)$

Therefore, $F(q) = \frac{q-2}{7-2} = \frac{1}{2} \implies q = 4.5$



Research demonstrates that the average time it takes a driver to depress car brakes after seeing the brake lights of the car in front of them light up is 1.25 seconds with a standard deviation of 0.46 seconds.
Across all subjects this data is normally distributed.

What is the probability that a given driver has a reaction time between 1 second and 1.75 seconds?

Mean value, $\mu = 1.25$
Std. Dev., $\sigma = 0.46$

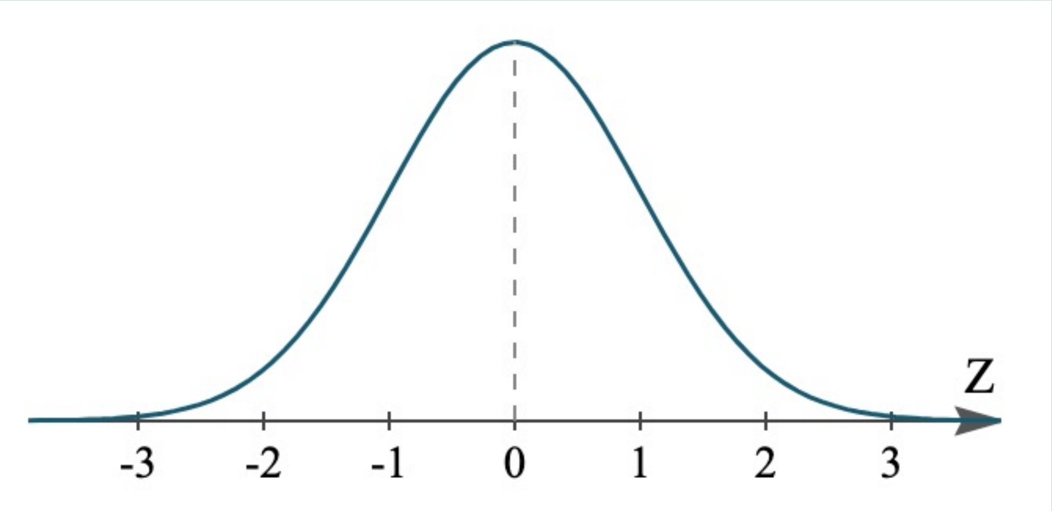
Change the normal distribution into a standard normal distribution:

$$P(1.0 \le x \le 1.75) = P\left(\frac{1.0 - 1.25}{0.46} \le Z \le \frac{1.75 - 1.25}{0.46}\right)$$

$$= P(-0.54 \le Z \le 1.09)$$

$$= \Phi(1.09) - \Phi(-0.54)$$

$$\approx 0.5675449$$



Next time, Chapter 7: Expectation and Variance.