Computer Science

CSC196: Analyzing Data

Some Continuous Probability Distributions

Jason Pacheco and Cesim Erten

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
  - Area under the normal curve
  - Normal approximation of the binomial
- Lognormal Distribution

Continuous Uniform Distribution

Example: Weather forecast.

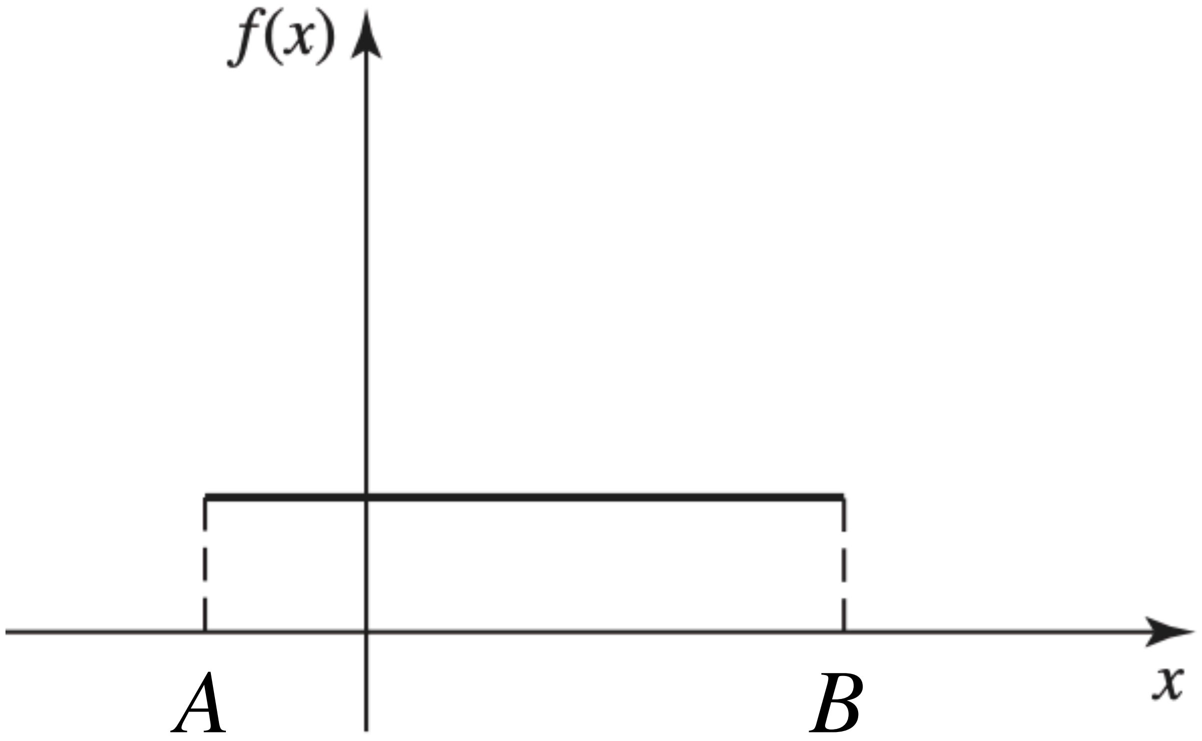
Rounded to the nearest integer.

We can assume

- The predicted temperature is equally likely any number in the interval from 55.5 and 56.5
- The temperature has continuous uniform distribution on $[55.5, 56.5]$.

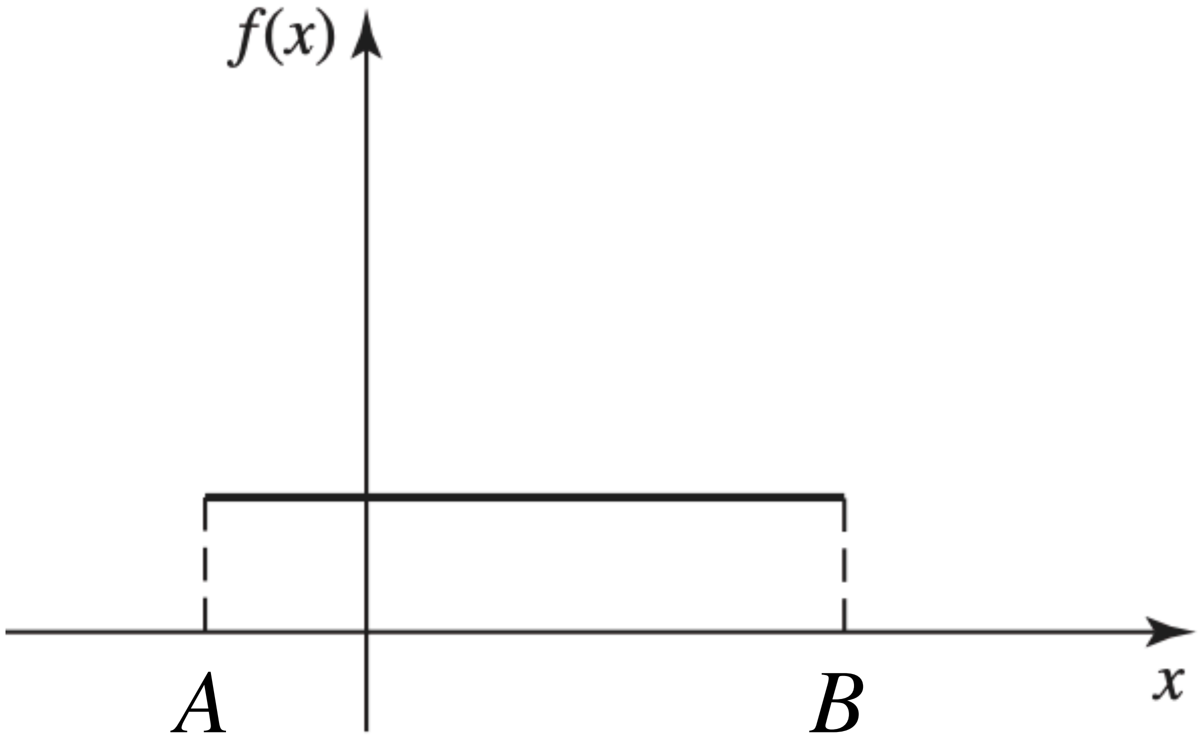
Continuous Uniform Distribution

Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



Continuous Uniform Distribution

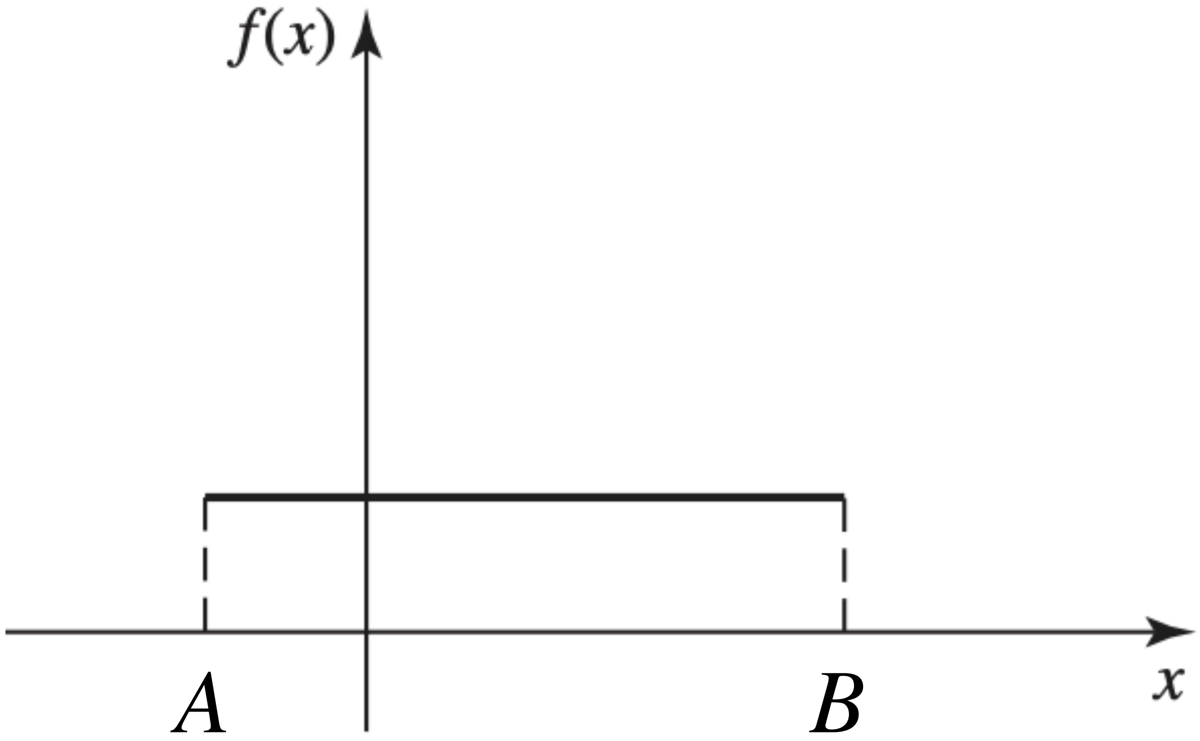
Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



$$f(x) = \begin{cases} \dfrac{1}{B - A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Definition: If $A \le X \le B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.

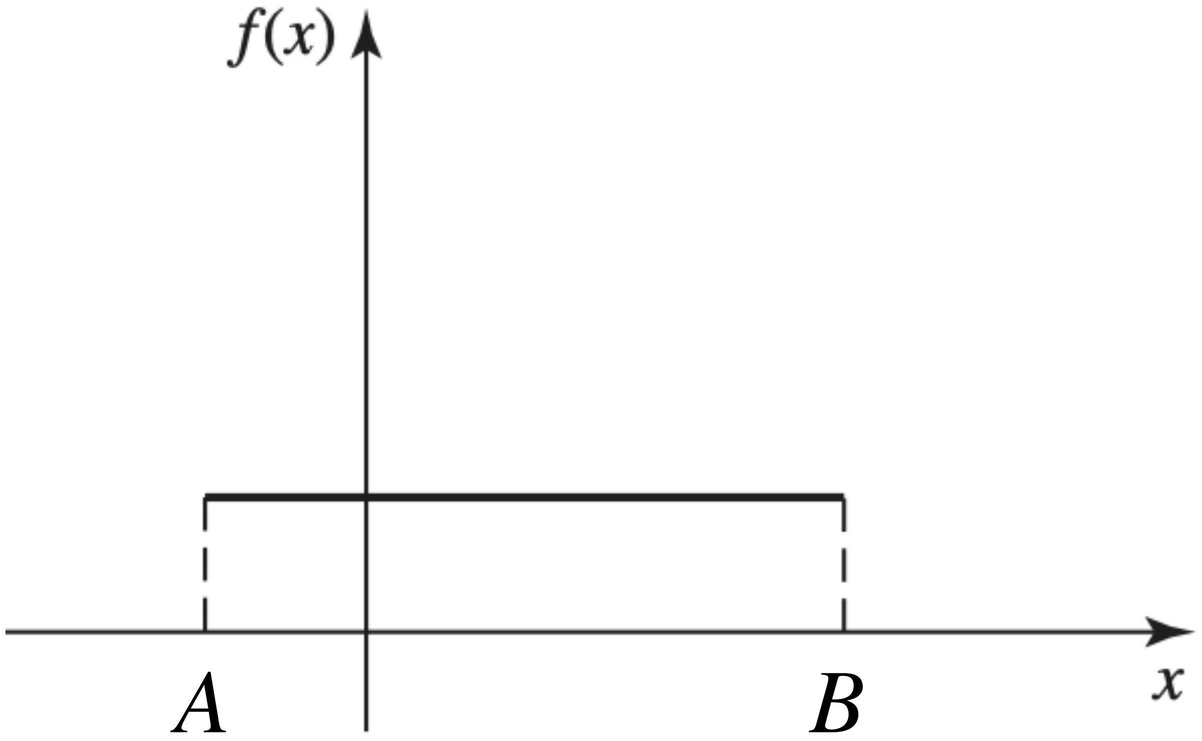


Can it be larger than 1?

$$f(x) = \begin{cases} \dfrac{1}{B-A}, & A \le x \le B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Definition: If $A \le X \le B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



Is it a pdf: Always nonnegative?
Sum to 1?

$$f(x) = \begin{cases} \dfrac{1}{B-A}, & A \le x \le B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Example: A conference room can be reserved for no more than 5 hours. Both long and short conferences occur often and it is assumed that the length X of a conference has a uniform distribution on the interval $[0, 5]$.

- The pdf?
- The probability that any given conference lasts at least 2 hours?

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Let's verify the mean:

Outline

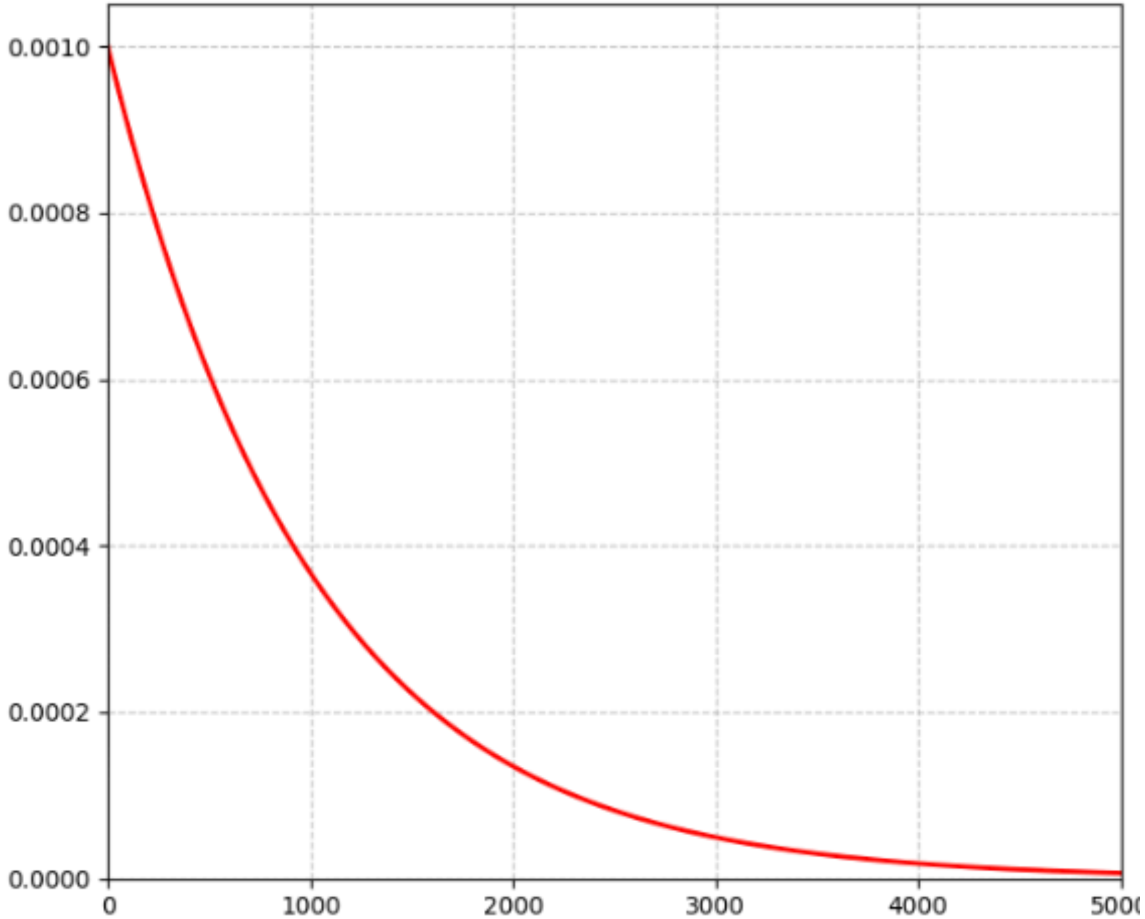
- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
  - Area under the normal curve
  - Normal approximation of the binomial
- Lognormal Distribution

Exponential Distribution

Example Applications:

Time between arrivals at service facilities

Time to failure of component parts



Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

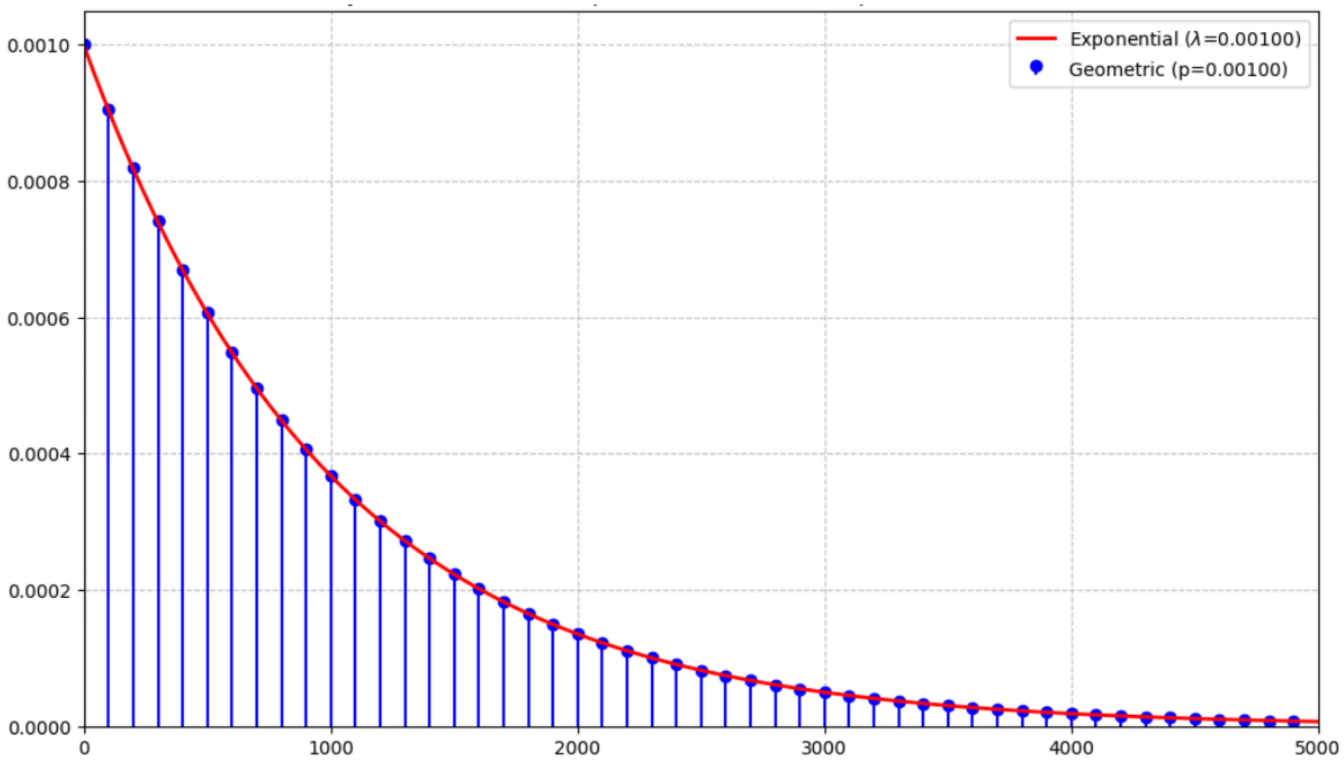
Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

Split time into minuscule units, say seconds: Each second 0 or 1 students.

Probability of x seconds until next arrival: $0.999^{x-1}0.001$



Exponential Distribution

Definition: X has exponential distribution with parameter λ, if

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

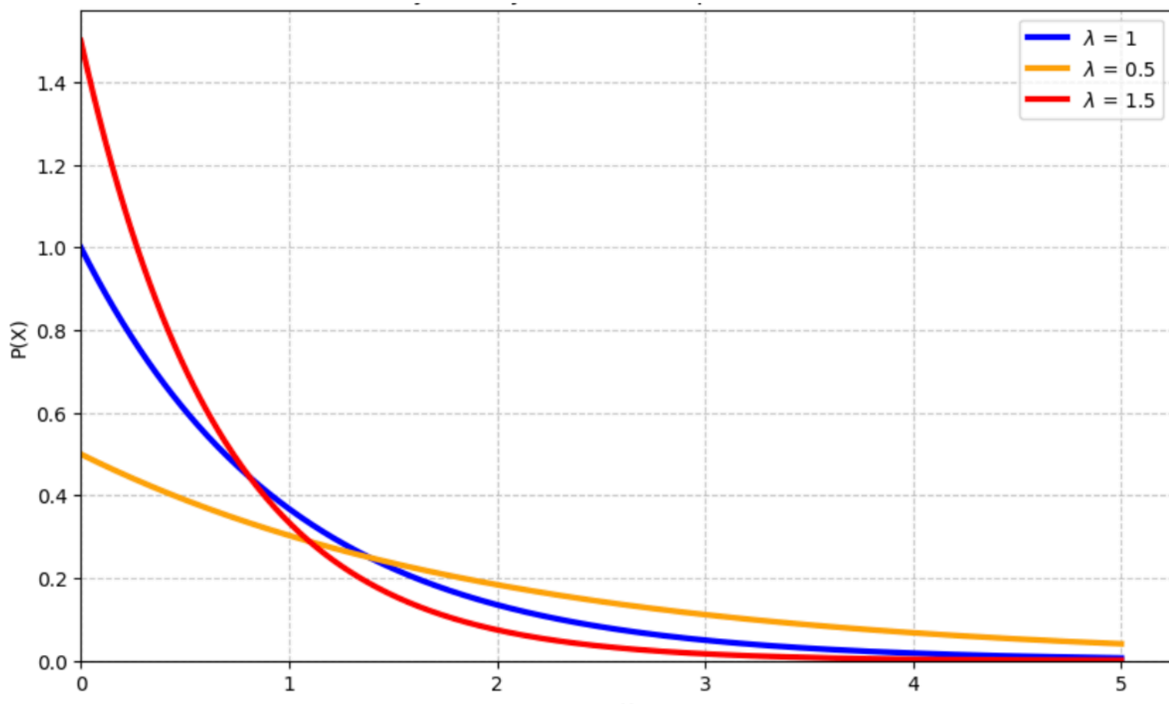
Time until next event nonnegative

Note: Textbook uses $1/\beta$ instead of λ.

Exponential Distribution

Definition: X has exponential distribution with parameter λ, if

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



Note: Textbook uses $1/\beta$ instead of λ.

Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✓

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

$$\Rightarrow P(X \leq x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$

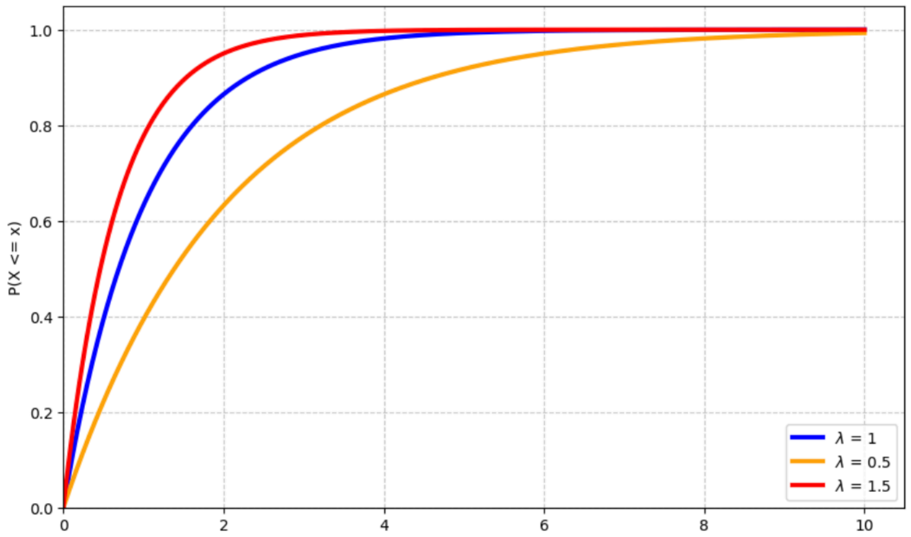
Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

$$\Rightarrow P(X \le x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$



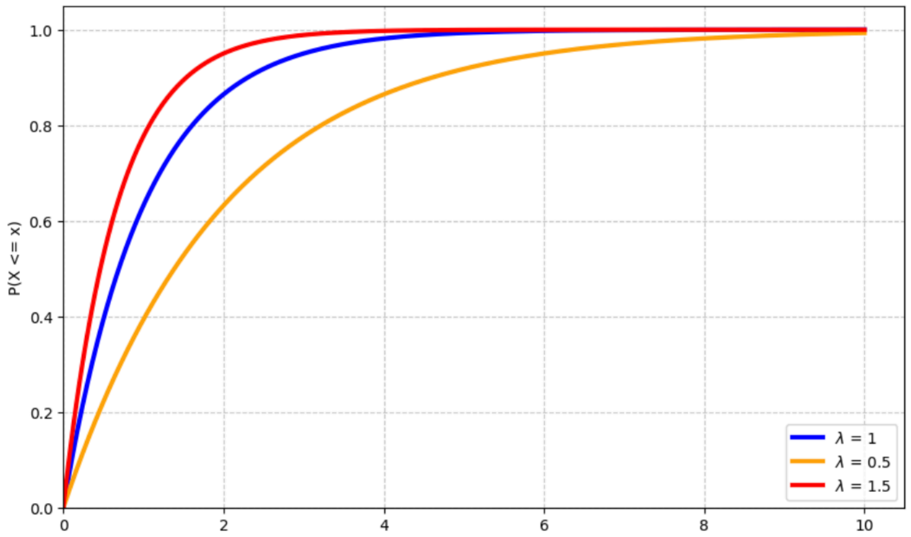
Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t}dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

$$\Rightarrow P(X \le x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$



Mean and Variance: $\mu = \frac{1}{\lambda}, \ \sigma^2 = \frac{1}{\lambda^2}$

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Number of random arrivals during a period of time $\longrightarrow$ Poisson distribution

Time until occurrence of next Poisson event $\longrightarrow$ Exponential distribution

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Recall with Poisson random variable X: $$Pois(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$\Rightarrow \dfrac{(\lambda t)^x e^{-\lambda t}}{x!}$ is the probability that x events occur in t units of time.

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Recall with Poisson random variable X: $$Pois(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$\Rightarrow \quad \dfrac{(\lambda t)^x e^{-\lambda t}}{x!}$ is the probability that x events occur in t units of time.

Y: time to first occurrence.

$P(Y > y) = e^{-\lambda y}$ (probability of 0 events up to time y)

$P(0 \le Y \le y) = 1 - e^{-\lambda y}$ is cdf of Y. Then $f(y) = \lambda e^{-\lambda y}$

Compare with pdf of Exponential.

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Note that λ is the mean of Poisson and $\dfrac{1}{\lambda}$ is the mean of Exponential.

If on average 4 events occur in 1 hour, time for first occurrence of the event on average is $1/4$ hours.

Exponential Distribution is Memoryless

$P(X \geq i + a \,|\, X \geq i)$ is the same as $P(X \geq a)$.

Example:

Arrival of your bus due to a Poisson process, rather than a fixed schedule.

Say you already waited for $i = 10$ mins. Probability that you wait ≥ 5 more mins is the same as the probability of you originally waiting for ≥ 5 mins (Consider a new passenger who comes after you waited for 10 mins.)

Note: This doesn't apply to cases where wear and tear is a factor.

Exponential Distribution is Memoryless

$P(X \geq i + a \,|\, X \geq i)$ is the same as $P(X \geq a)$.

Proof:

Some Exercises on Exponential Distribution

Example 6.17: Suppose that a system contains a certain type of component whose time, in years, to failure is given by X. The random variable X is modeled nicely by the exponential distribution with mean time to failure $= 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

Use $e^{-8/5} \approx 0.2$ and

```
print (1-binom.cdf(1, 5, 0.2))
```

```
0.26272000000000006
```

Solution:

Some Exercises on Exponential Distribution

Example: An exam is taken by 5 students independently. The number of minutes required by any particular student to complete it has exponential distribution with mean 80. The exam starts at 9:00 a.m. Probability that at least one of the students will complete the exam before 9:40 a.m.?

Solution:

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Gamma Distribution

A generalization of the exponential distribution:

Time until the next Poisson event $\longrightarrow$ Exponential

Time until a specified number of Poisson events $\longrightarrow$ Gamma

Gamma Distribution

A generalization of the exponential distribution:

Time until the next Poisson event $\longrightarrow$ Exponential

Time until a specified number of Poisson events $\longrightarrow$ Gamma

Definition: X has gamma distribution with parameters λ and α if

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

λ : Poisson mean

α : Specified number of Poisson events

Note: Textbook uses $1/\beta$ instead of λ.

Gamma Distribution

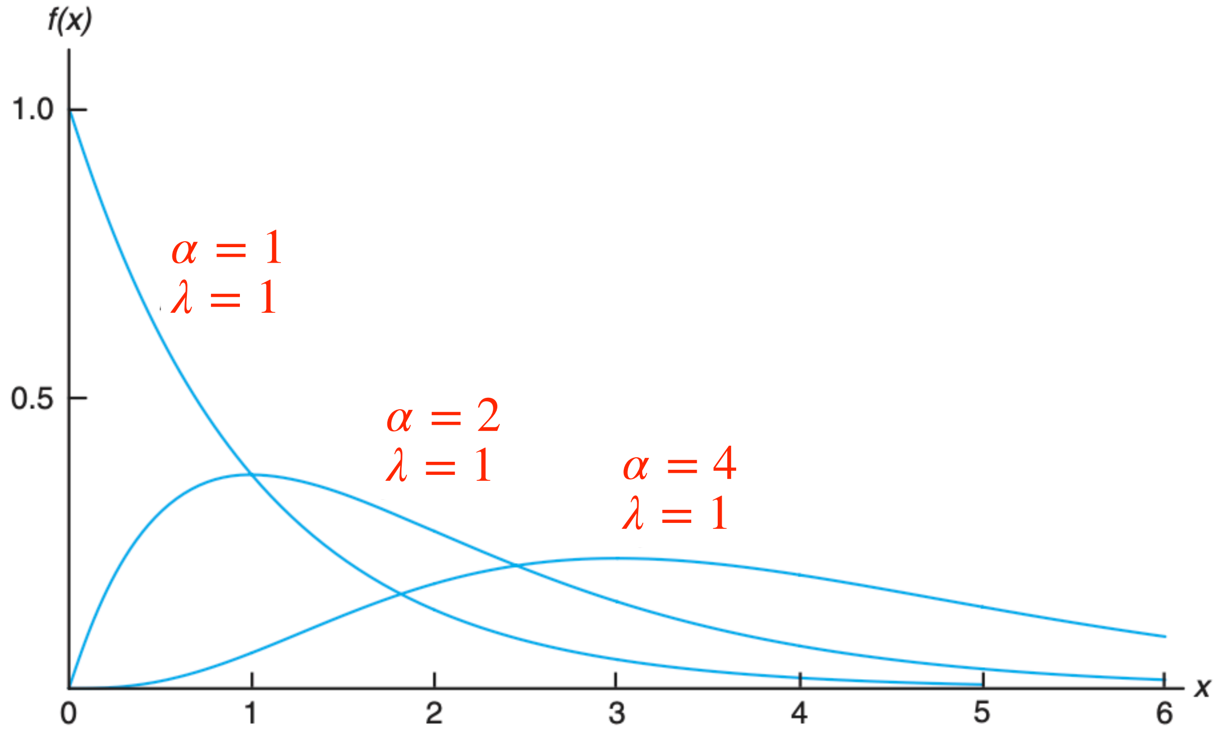
A generalization of the exponential distribution:

Time until the next Poisson event $\longrightarrow$ Exponential

Time until a specified number of Poisson events $\longrightarrow$ Gamma

Definition: X has gamma distribution with parameters λ and α if

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



Note: Textbook uses $1/\beta$ instead of λ.

Gamma Distribution

$\Gamma(\alpha)$ in the pmf is the Gamma function defined as:

Recursive formulation:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha)$$

which implies factorial for integers: $\Gamma(1) = 1,\ \ \Gamma(n) = (n-1)!$

Exercise on Gamma Distribution

Example 6.18: Suppose that telephone calls arriving at a particular switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will elapse by the time 2 calls have come in to the switchboard?

Use $\int xe^{-5x}dx = -\frac{e^{-5x}}{25}(5x+1)$

Exercise on Gamma Distribution

Example 6.18: Suppose that telephone calls arriving at a particular switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will elapse by the time 2 calls have come in to the switchboard?

Use $\int xe^{-5x}dx = -\frac{e^{-5x}}{25}(5x+1)$

$$P(X \le 1) = \int_0^1 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx.$$ Since $\lambda = 5$, $\alpha = 2$, we want:

$$P(X \le 1) = 25 \int_0^1 xe^{-5x}\, dx = -e^{-5x}(5x+1)\Big|_0^1 = -6e^{-5} + 1.$$

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Normal Distribution

Definition: X has Normal (Gaussian) distribution with parameters $\mu,\ \sigma^2$, if

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

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

or equivalently with the notation of the textbook $n(x; \mu, \sigma)$

Normal Distribution

Definition: X has Normal (Gaussian) distribution with parameters μ, σ^2, if

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

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

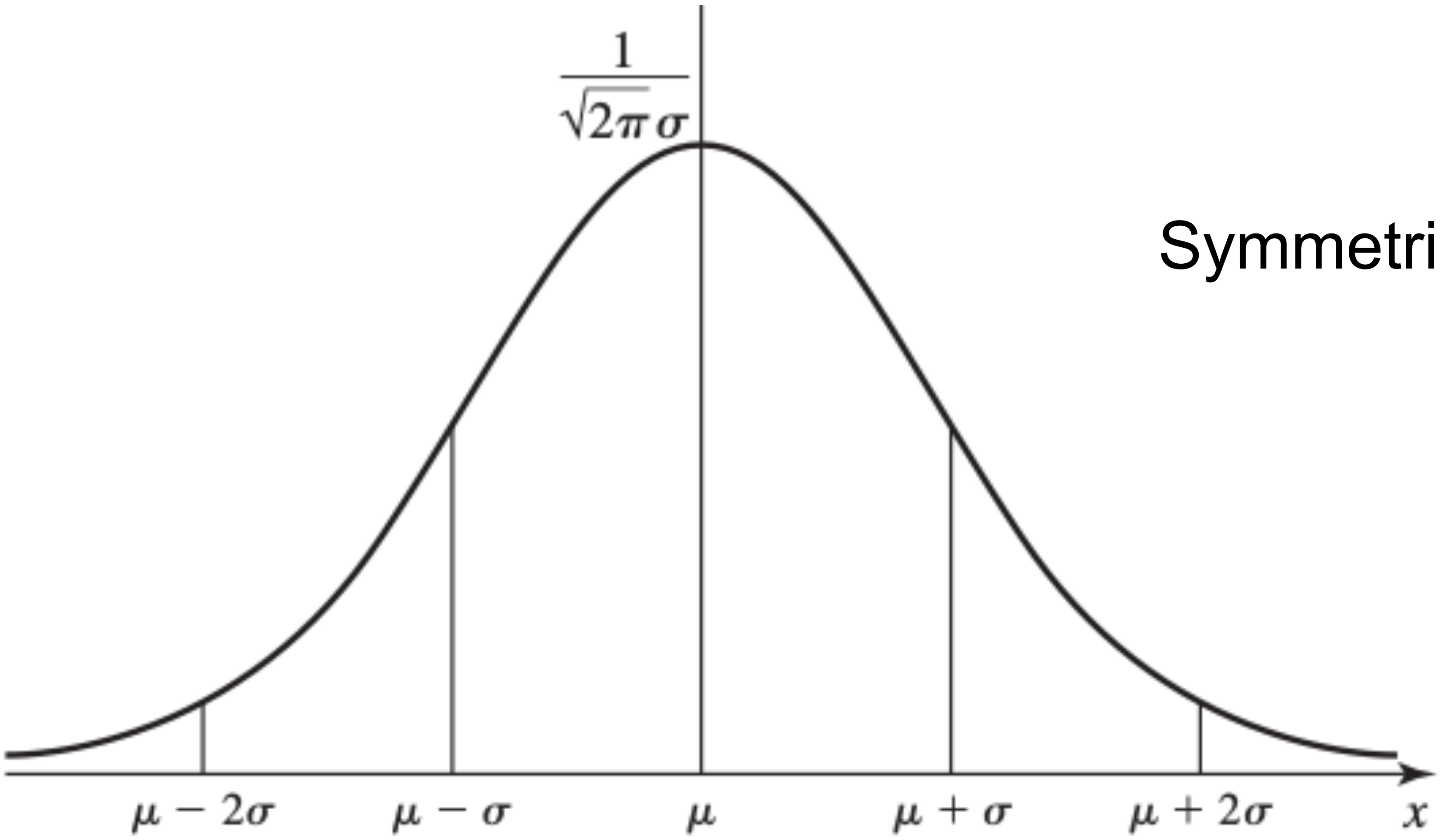
or equivalently with the notation of the textbook $n(x; \mu, \sigma)$

A very common distribution:

Many physical phenomena are approximately normal

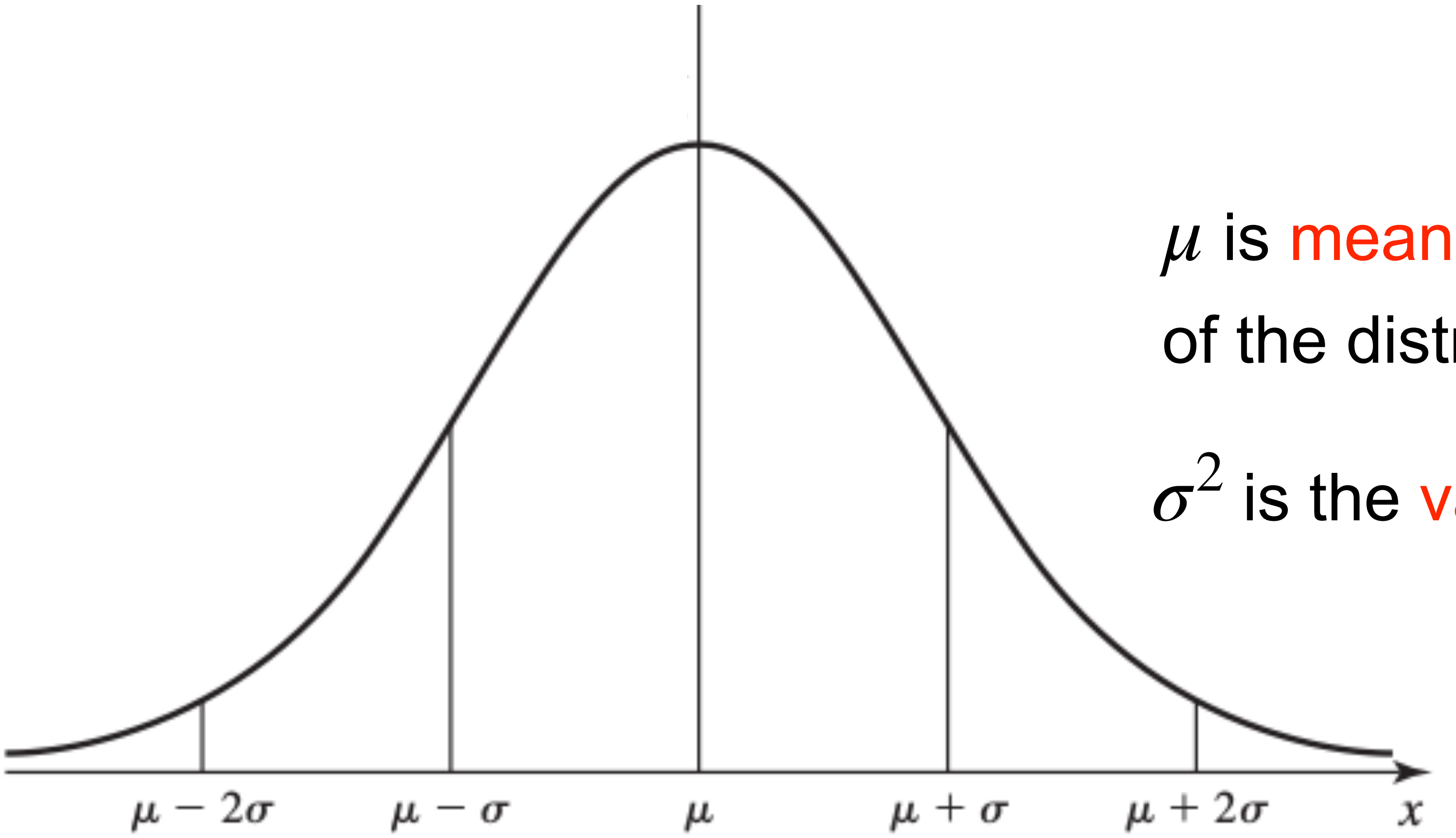
Central Limit Theorem: Mean of samples from any distribution is approximately normal (We'll see the details later.)

Normal Distribution



Symmetric around μ.

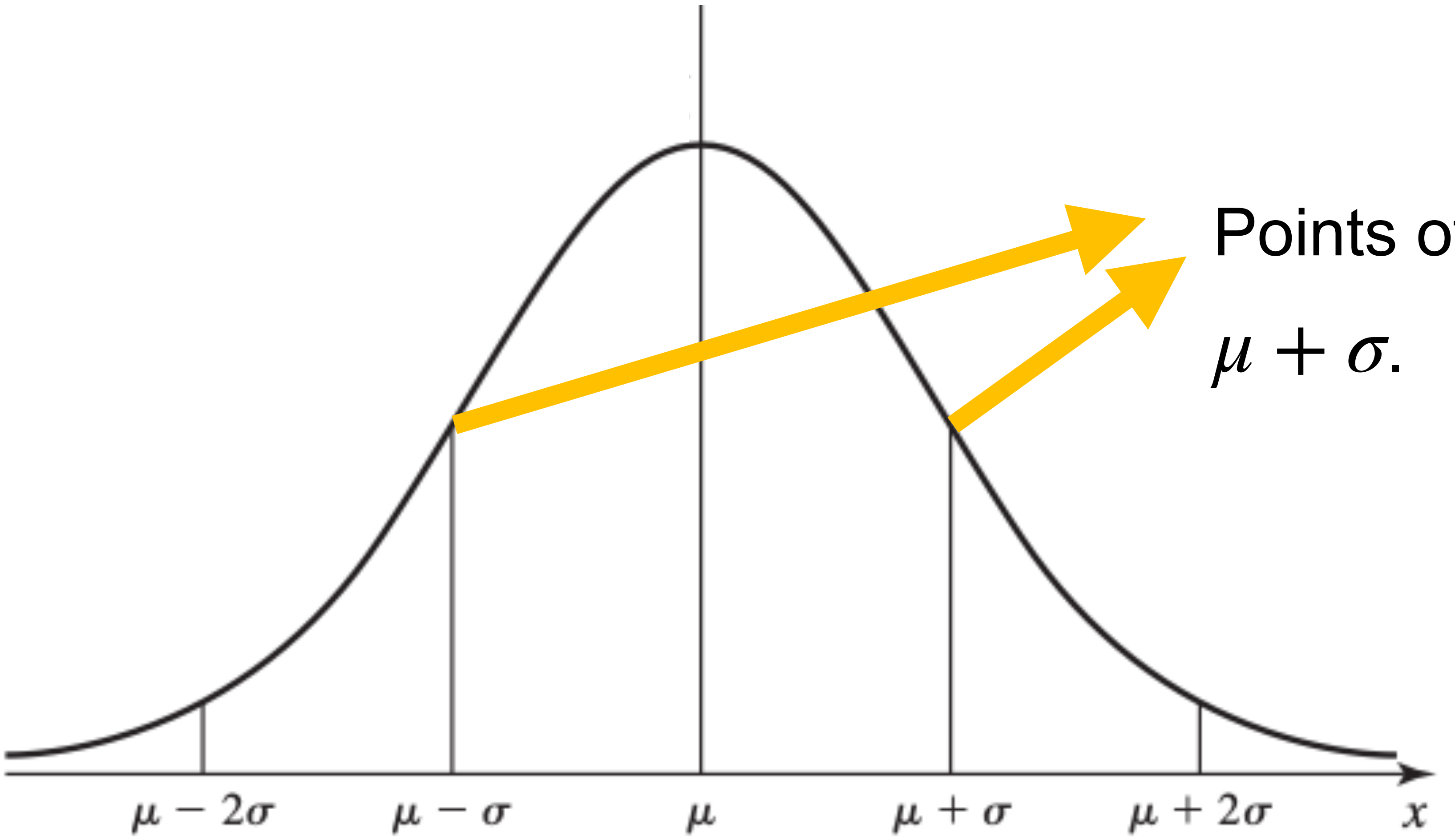
Normal Distribution



μ is mean, median, and the mode of the distribution.

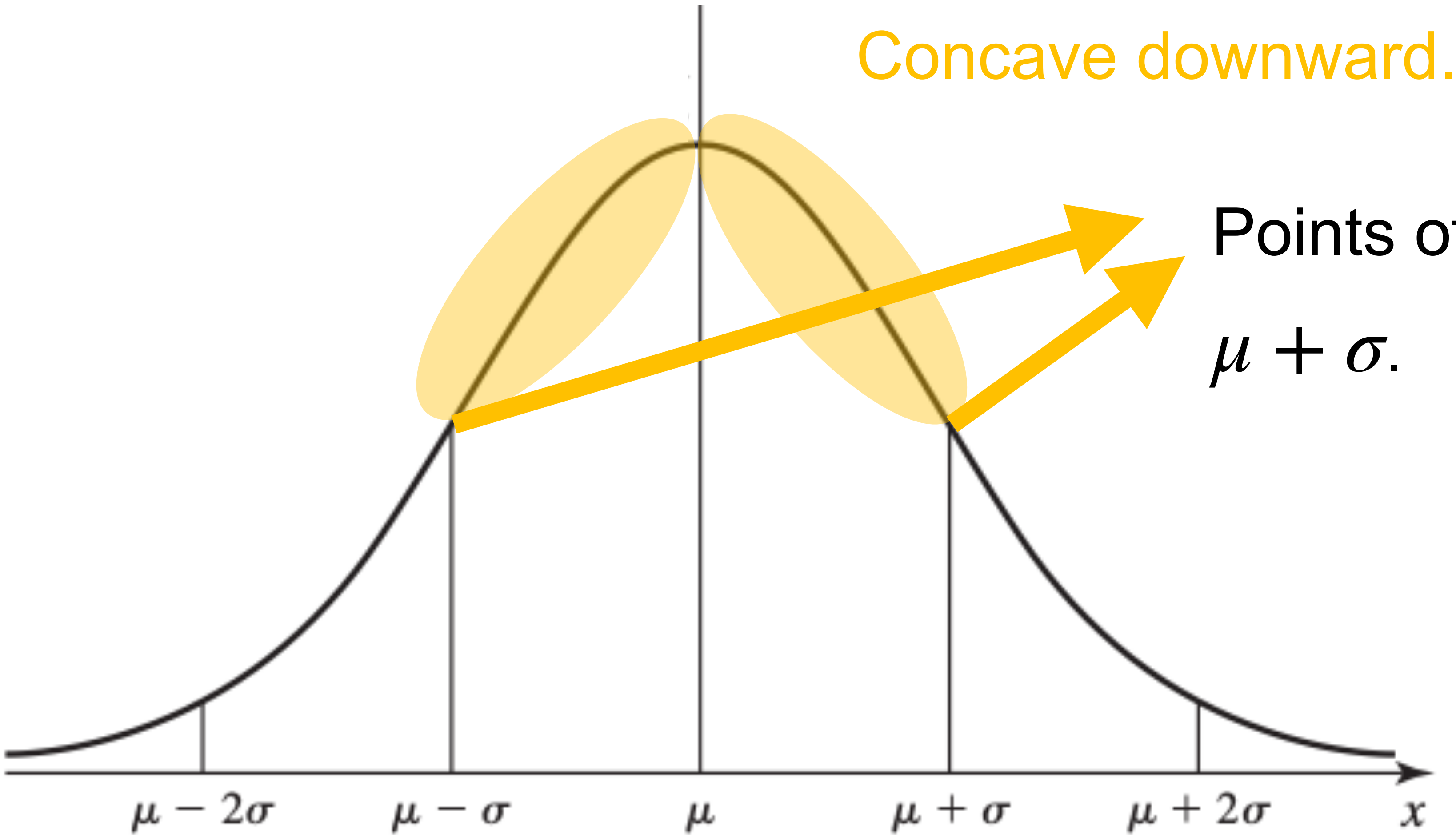
σ^2 is the variance of the distribution.

Normal Distribution



Points of inflection at $\mu - \sigma$ and $\mu + \sigma$.

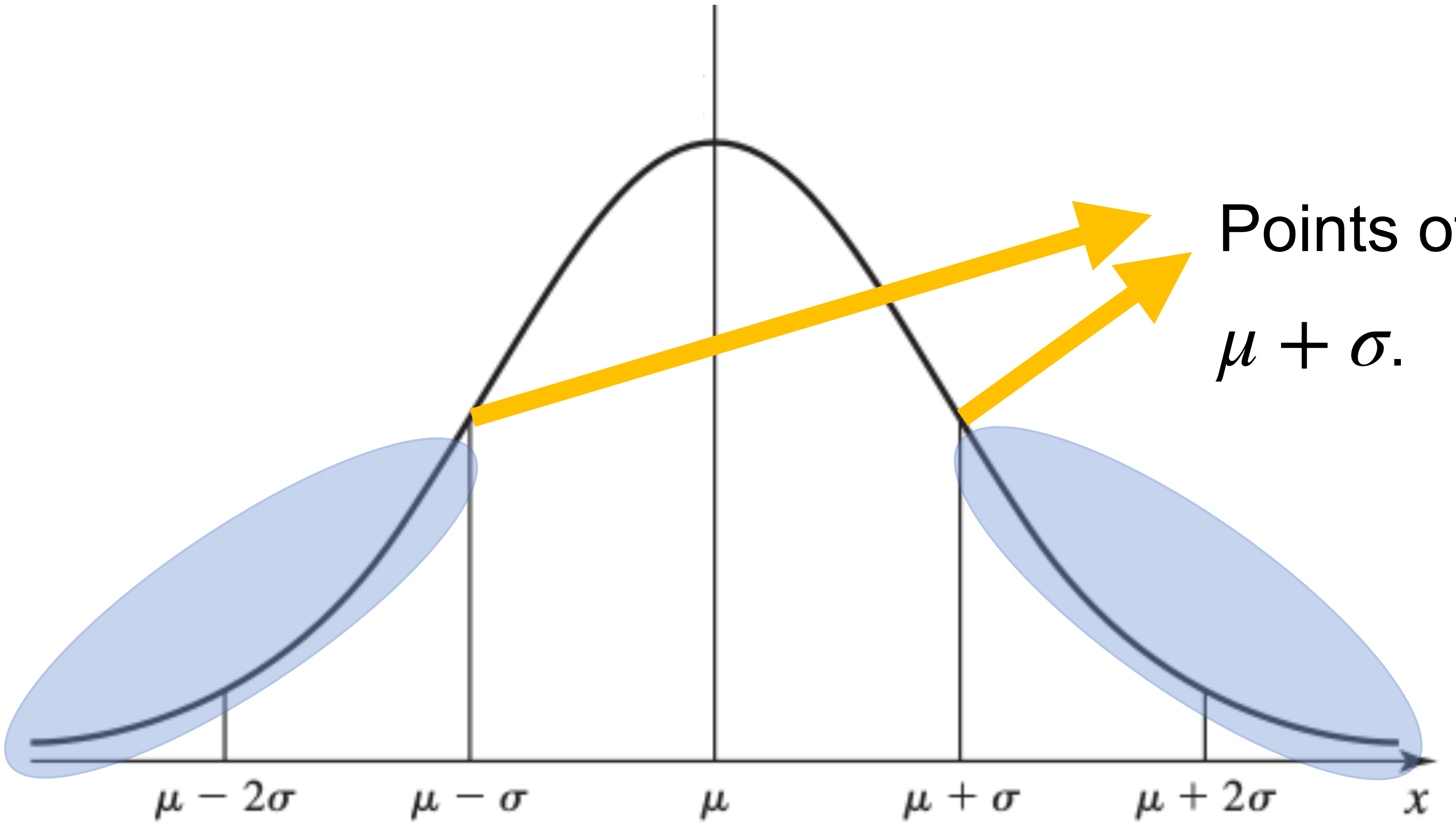
Normal Distribution



Concave downward.

Points of inflection at $\mu - \sigma$ and $\mu + \sigma$.

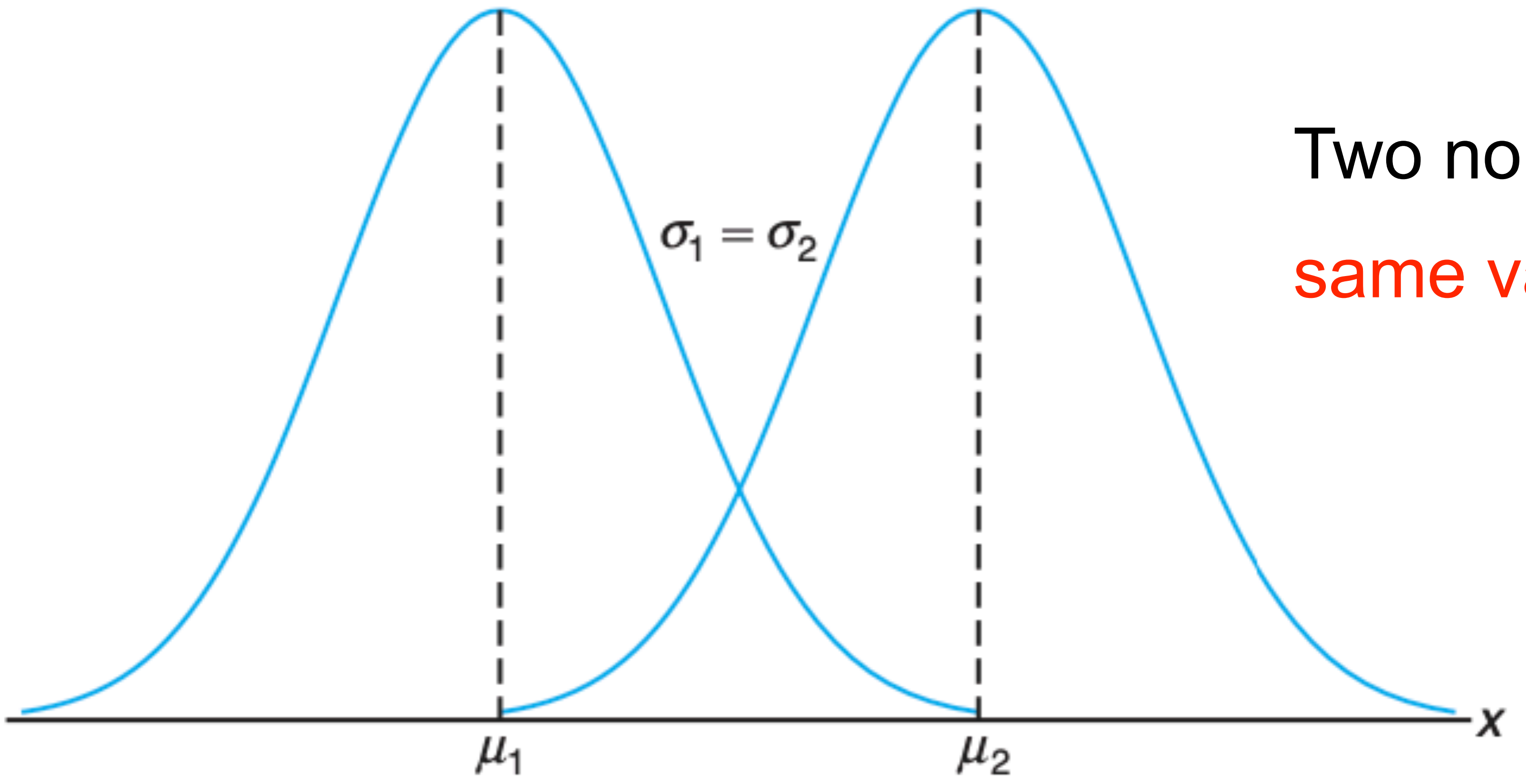
Normal Distribution



Points of inflection at $\mu - \sigma$ and $\mu + \sigma$.

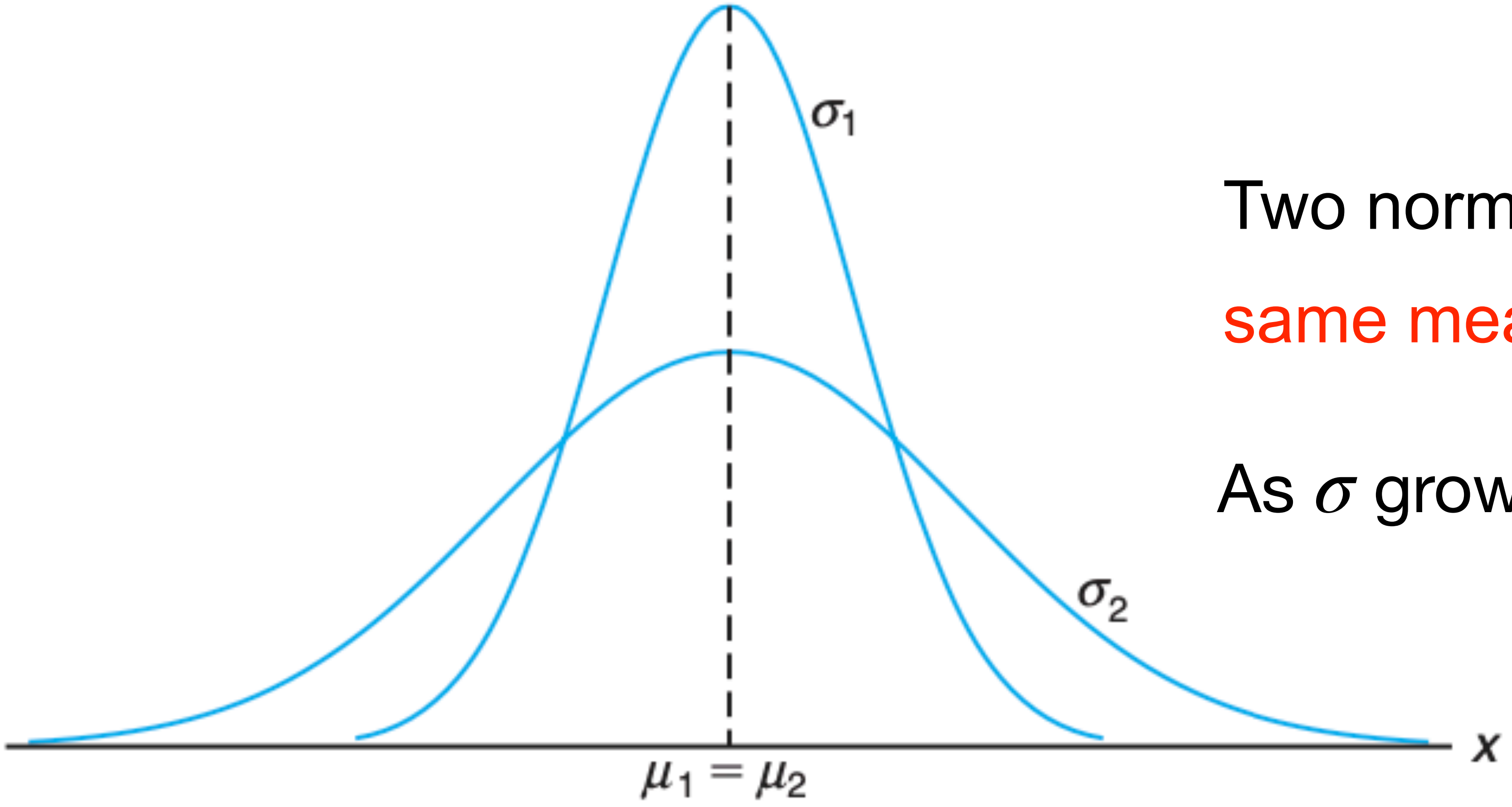
Concave upward.

Normal Distribution



Two normal distributions with the same variance and $\mu_1 < \mu_2$.

Normal Distribution



Two normal distributions with the same mean and $\sigma_1 < \sigma_2$.

As σ grows, pdf gets more flat.

Normal Distribution

Linear combinations/transformations of normal independent variables:

If X_1, X_2 are independent and each is normally distributed then

$Y = a_1X_1 + a_2X_2 + b$ has a normal distribution.

(Can be generalized to more than 2 random variables.)

Revisit: What is the mean and variance of Y?

Normal Distribution

Linear combinations/transformations of normal independent variables:

If X_1, X_2 are independent and each is normally distributed then

$Y = a_1X_1 + a_2X_2 + b$ has a normal distribution.

(Can be generalized to more than 2 random variables.)

Revisit: What is the mean and variance of Y?

Mean is $a_1\mu_1 + a_2\mu_2 + b$ and variance is $a_1^2\sigma_1^2 + a_2^2\sigma_2^2$.

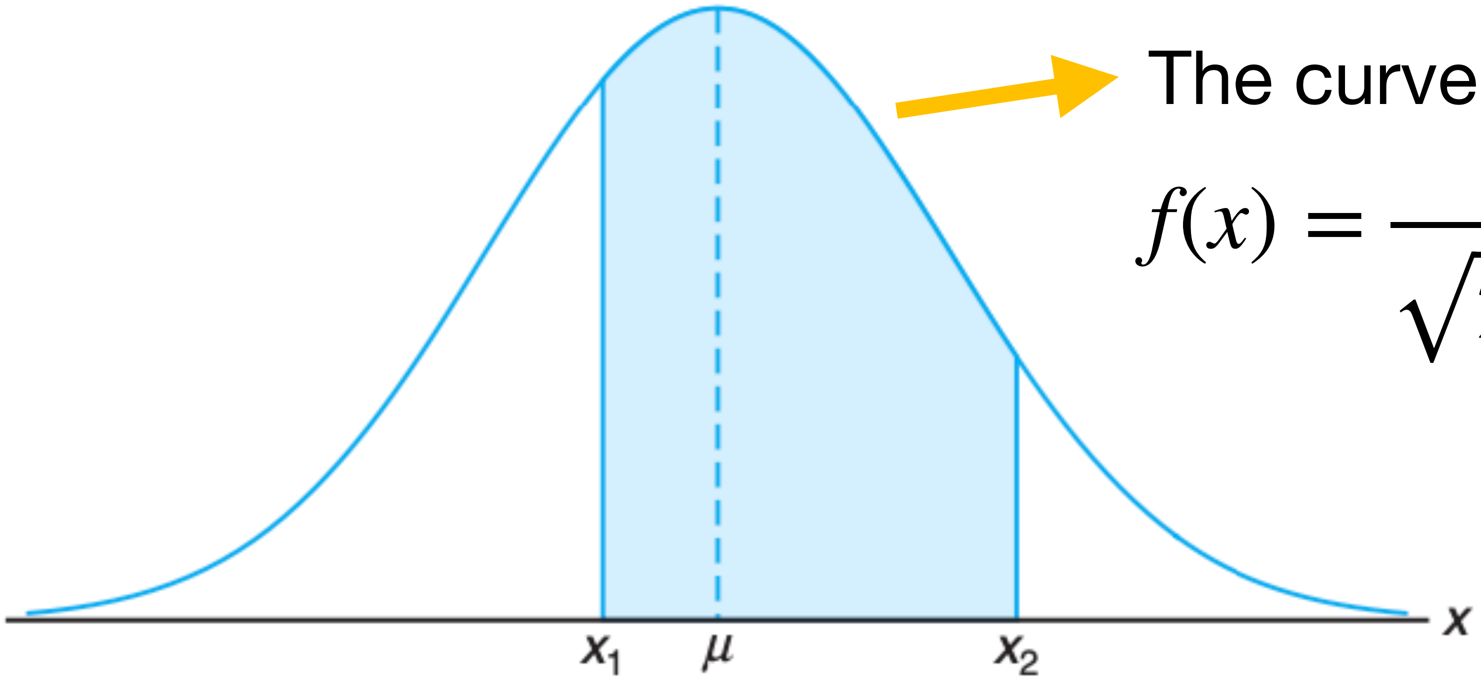
Normal Distribution

Generalizes to sum of independent random variables (regardless of their distributions)

More on this later when we talk about the central limit theorem.

Normal Distribution

Area under the Normal Curve



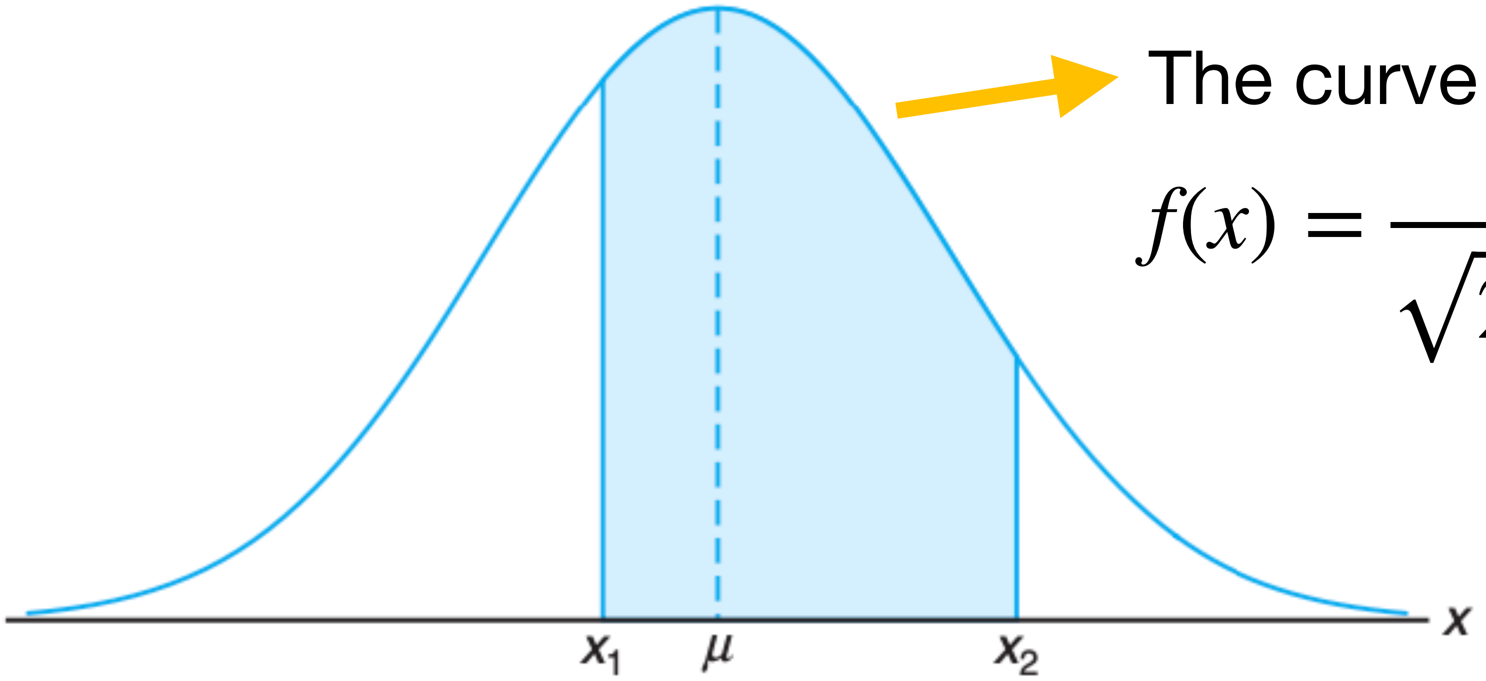
The curve

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

$\int_{x_1}^{x_2} f(x)\, dx$ gives the area of the shaded region, that is $P(x_1 < X < x_2)$.

Normal Distribution

Area under the Normal Curve



The curve

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

$\int_{x_1}^{x_2} f(x)\, dx$ gives the area of the shaded region, that is $P(x_1 < X < x_2)$.

This integral can't be solved. A standard transformation is applied, then looked up from a table or computed via software.

Normal Distribution

Standard Transformation:

If X is normally distributed with mean μ and variance σ^2 then

$Z = \dfrac{X - \mu}{\sigma}$ is normally distributed with mean 0 and variance 1.

(Follows from linear transformations of normal independent random variables)

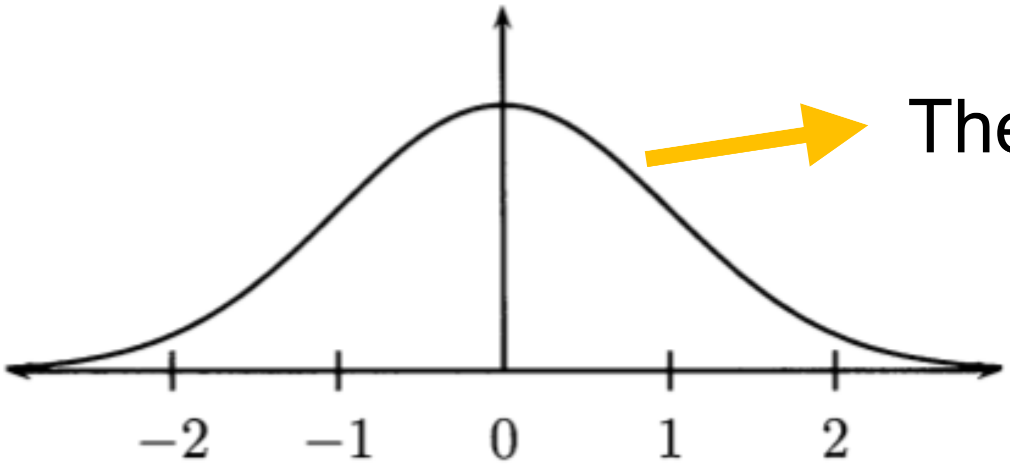
Normal Distribution

Standard Transformation:

If X is normally distributed with mean μ and variance σ^2 then

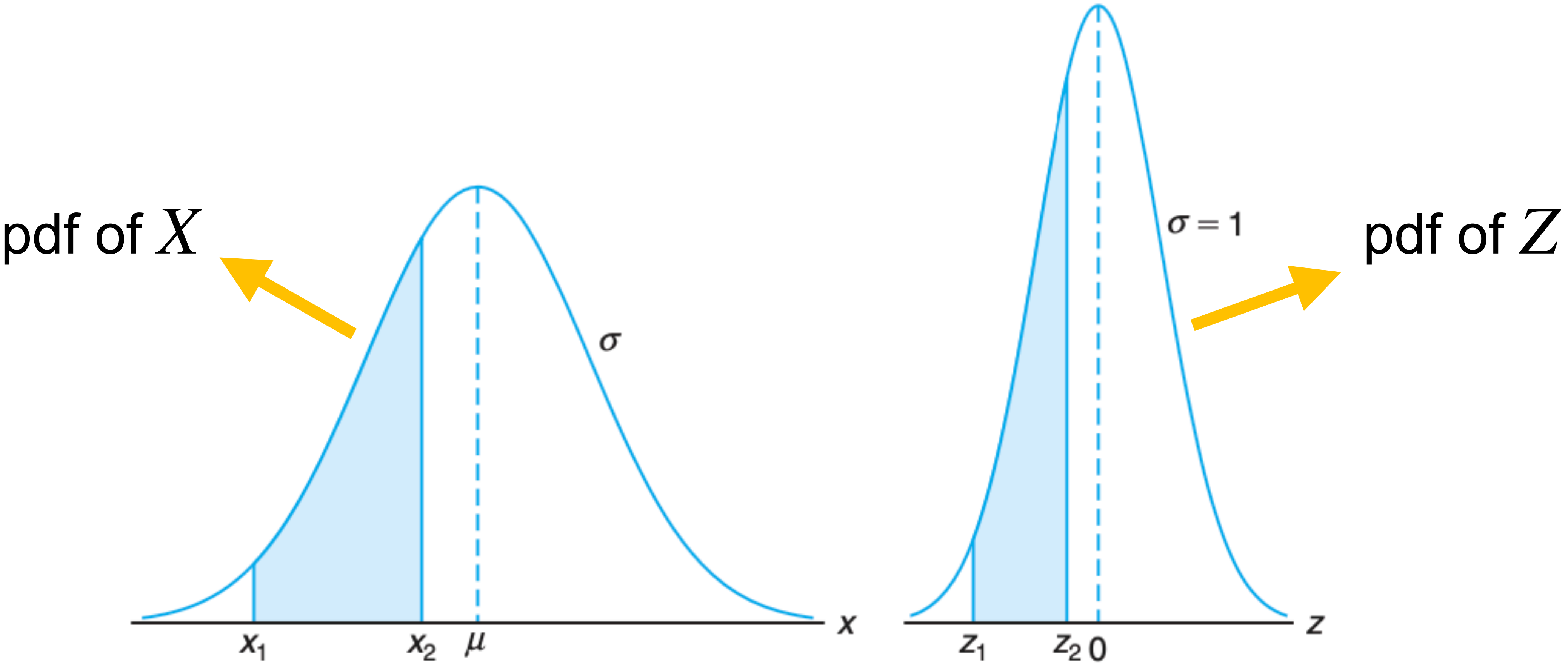
$Z = \dfrac{X - \mu}{\sigma}$ is normally distributed with mean 0 and variance 1.

(Follows from linear transformations of normal independent random variables)



The standard normal distribution Z.

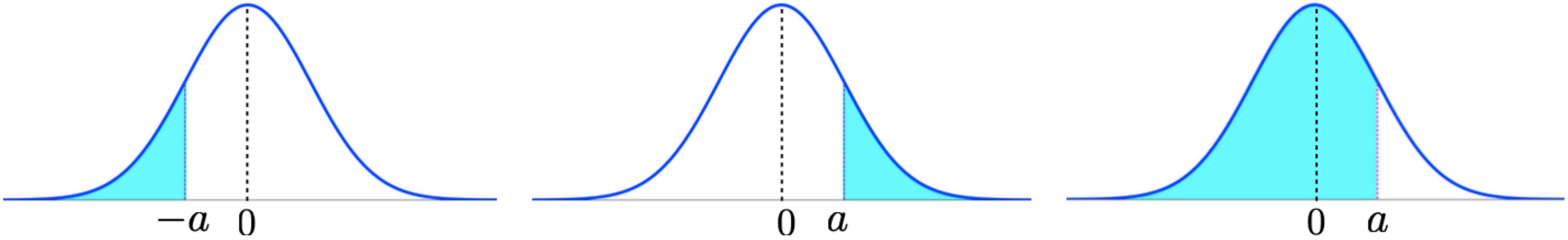
Normal Distribution



$$P(x_1 < X < x_2) \quad = \quad P\left(\frac{x_1 - \mu}{\sigma} < Z < \frac{x_2 - \mu}{\sigma}\right)$$

Normal Distribution

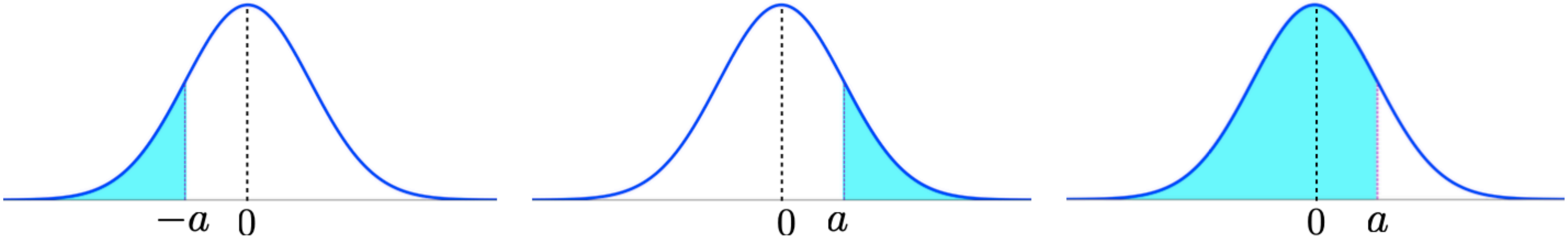
Standard Normal: Consequence of symmetry



$$P(Z \le -a) \quad = \quad P(Z \ge a) \quad = \quad 1 - P(Z \le a)$$

Normal Distribution

Standard Normal: Consequence of symmetry

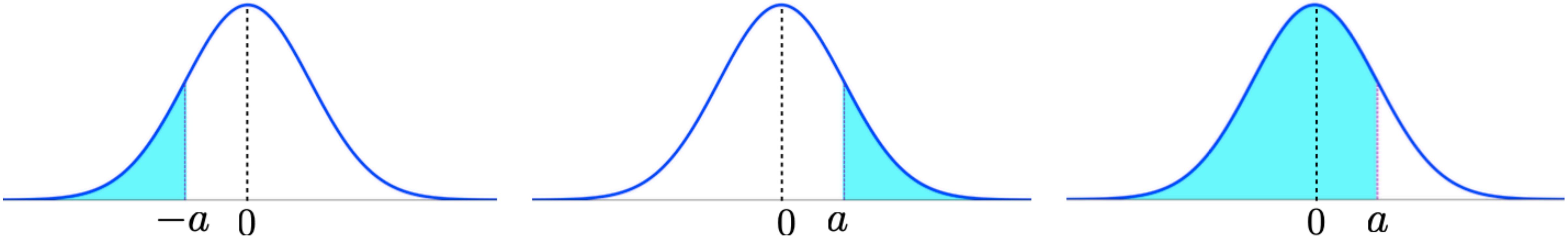


$P(Z \le -a) \quad = \quad P(Z \ge a) \quad = \quad 1 - P(Z \le a)$

The tables provide cdf, that is $\le$ not $\ge$.

Normal Distribution

Standard Normal: Consequence of symmetry

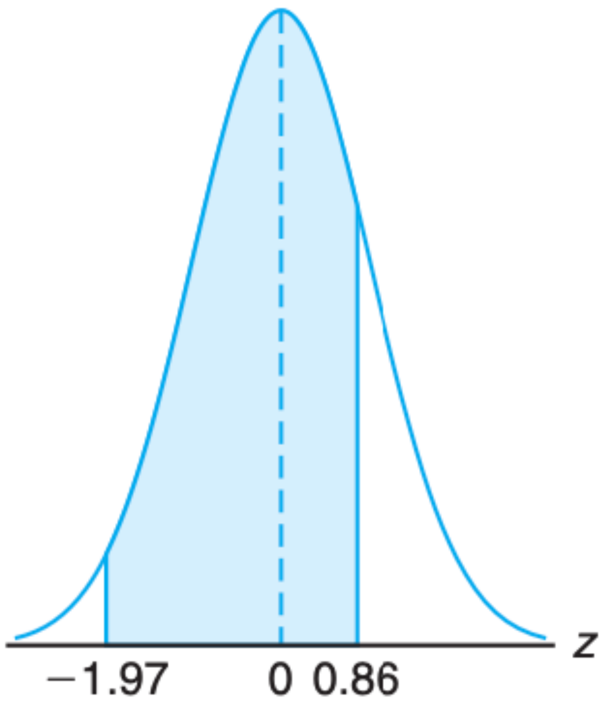


$$P(Z \leq -a) \quad = \quad P(Z \geq a) \quad = \quad 1 - P(Z \leq a)$$

Some tables provide the cdf just for positive values.

Normal Distribution

Example: What is $P(-1.97 < Z < 0.86)$?

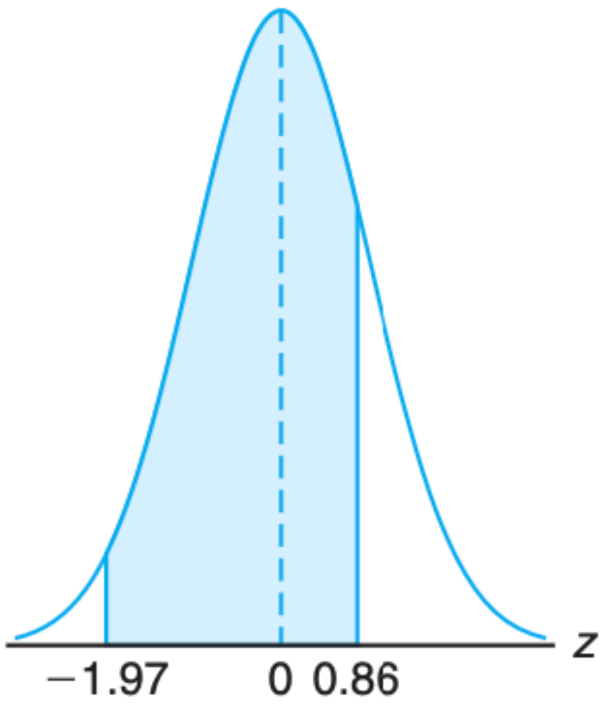


Normal Distribution

Example: What is $P(-1.97 < Z < 0.86)$?

Let's use Table A.3 of textbook.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
					⋮					
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
					⋮					
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133

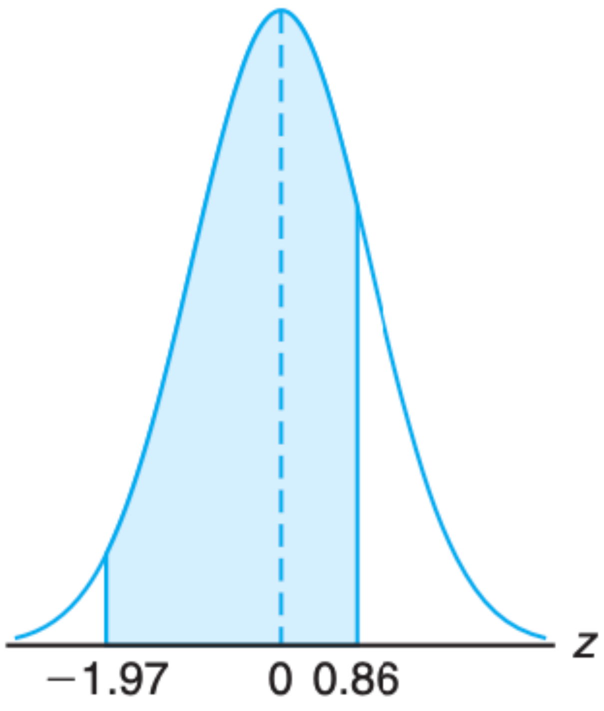


Normal Distribution

Example: What is $P(-1.97 < Z < 0.86)$?

Let's use Table A.3 of textbook.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
					⋮					
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
					⋮					
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133



$$P(-1.97 < Z < 0.86) = P(Z \leq 0.86) - P(Z \leq -1.97)$$

$$= 0.8051 - 0.0244 = 0.7807$$

Normal Distribution

Example: X is normally distributed with mean 5 and variance 4.

What is $P(1 < X < 8) = ?$

Let's use scipy.stats.

```
from scipy.stats import norm
print(norm.cdf(1.5)-(1-norm.cdf(2)))
```

```
0.9104426667829627
```

Normal Distribution

Example: X is normally distributed with mean 5 and variance 4.

What is $P(1 < X < 8) = ?$

Let's use scipy.stats.

Note that we could also directly use:

```
norm.cdf(8, loc=5, scale=2) - norm.cdf(1, loc=5, scale=2)
```

Normal Distribution

The Quantile Function:

Sometimes called the inverse cdf or percent point function.

Given a desired probability p, find k such that $P(Z \leq k) = p$.

Normal Distribution

Example: Find k such that $P(Z \leq k) = 0.9$

Using a table like Table A.3 of textbook:

Locate the table entry closest to p to find k.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
					⋮					
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

$$P(Z \leq 1.28) \approx 0.9$$

Normal Distribution

Example: Find k such that $P(Z \leq k) = 0.9$

Using spicy.stats:

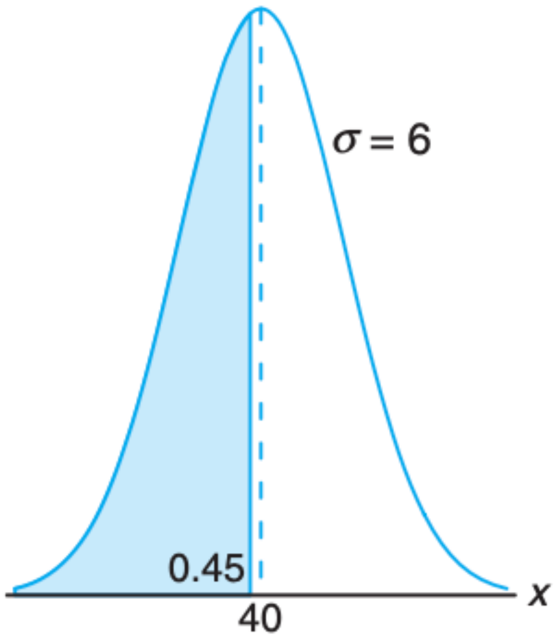
```
from scipy.stats import norm
norm.ppf(0.9)
```

```
1.2815515655446004
```

$$P(Z \leq 1.28) \approx 0.9$$

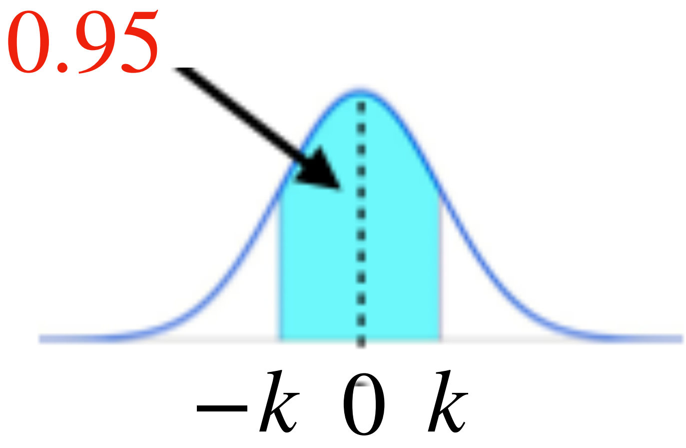
Normal Distribution

Example: Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has $45\ \%$ of the area to the left. Use norm.ppf$(0.45) \approx -0.13$



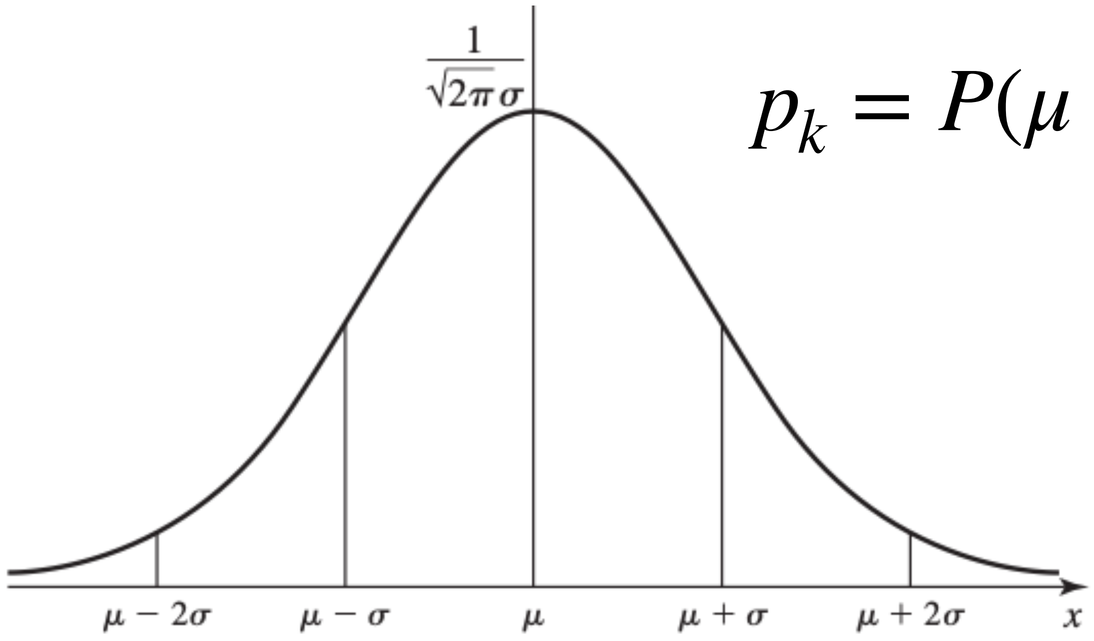
Normal Distribution

Example: Find k s.t. $P(-k \leq Z \leq k) = 0.95$ Use norm.ppf$(0.975) \approx 1.96$



Normal Distribution

Probability that normal X within k standard deviation of the mean:

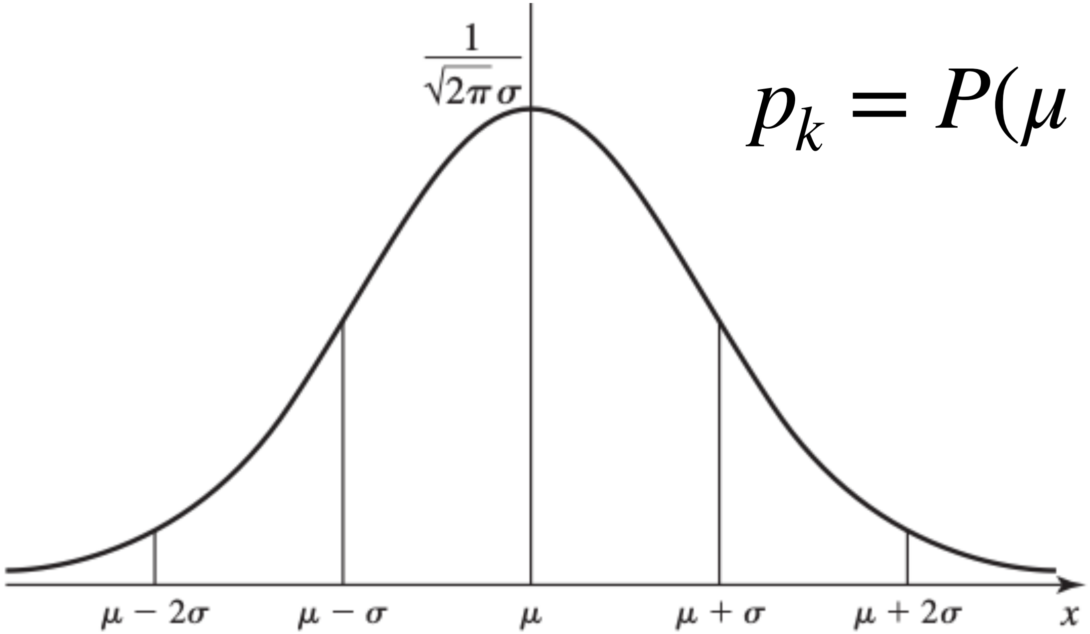


$$p_k = P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(-k \leq Z \leq k)$$

$$= 2P(Z \leq k) - 1$$

Normal Distribution

Probability that normal X within k standard deviation of the mean:



$$p_k = P(\mu - k\sigma \le X \le \mu + k\sigma) = P(-k \le Z \le k)$$

$$= 2P(Z \le k) - 1$$

68, 95, 99.7 rule.

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
$\vdots$	$\vdots$

Normal Distribution

Comparing Probability Bounds from Chebyshev's and Normal RV:

Normal RV:

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
$\vdots$	$\vdots$

Normal Distribution

Comparing Probability Bounds from Chebyshev's and Normal RV:

Chebyshev's Theorem's Bound:

k	$p_k \geq 1 - \frac{1}{k^2}$
1	0
2	0.75
3	0.8889
4	0.9375
$\vdots$	$\vdots$

Normal RV:

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
$\vdots$	$\vdots$

Normal Distribution

Comparing Probability Bounds from Chebyshev's and Normal RV:

Lower bound. Applies to any distribution.

Stronger bounds.

Chebyshev's Theorem's Bound:

k	$p_k \geq 1 - \frac{1}{k^2}$
1	0
2	0.75
3	0.8889
4	0.9375
$\vdots$	$\vdots$

Normal RV:

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
$\vdots$	$\vdots$

Some Exercises on Normal Random Variables

6.17 The average life of a certain type of small motor is 10 years with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If she is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.

Use norm.ppf$(0.03) \approx -1.88$

Some Exercises on Normal Random Variables

Example: Heights of women in a population is normally distributed with $\mu_W = 65$ and $\sigma_W = 1$, heights of men have normal distribution with $\mu_M = 68$ and $\sigma_M = 3$. One woman and one man are selected independently and randomly. Probability that the woman is taller than the man?

Use norm.cdf$(0.949) \approx 0.829$

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
  - Area under the normal curve
  - Normal approximation of the binomial
- Lognormal Distribution

Approximating Binomial with Normal

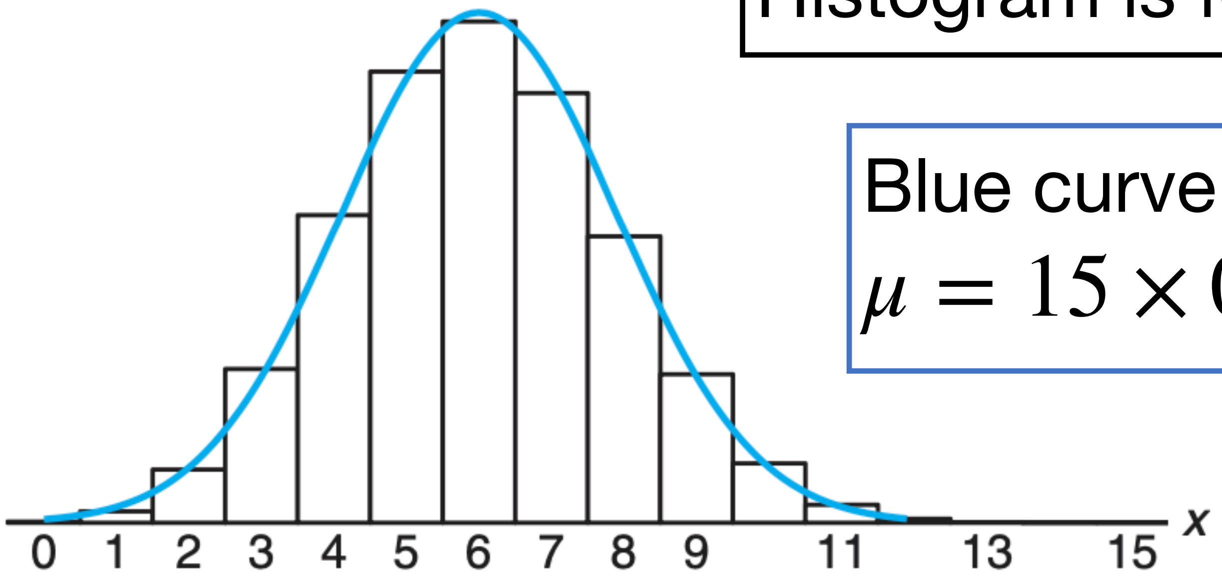
Let X be a binomial RV with mean np and variance $np(1-p)$.

Normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$ is a good approximation of X, especially for large n and p close to 0.5.

Approximating Binomial with Normal

Let X be a binomial RV with mean np and variance $np(1-p)$.

Normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$ is a good approximation of X, especially for large n and p close to 0.5.

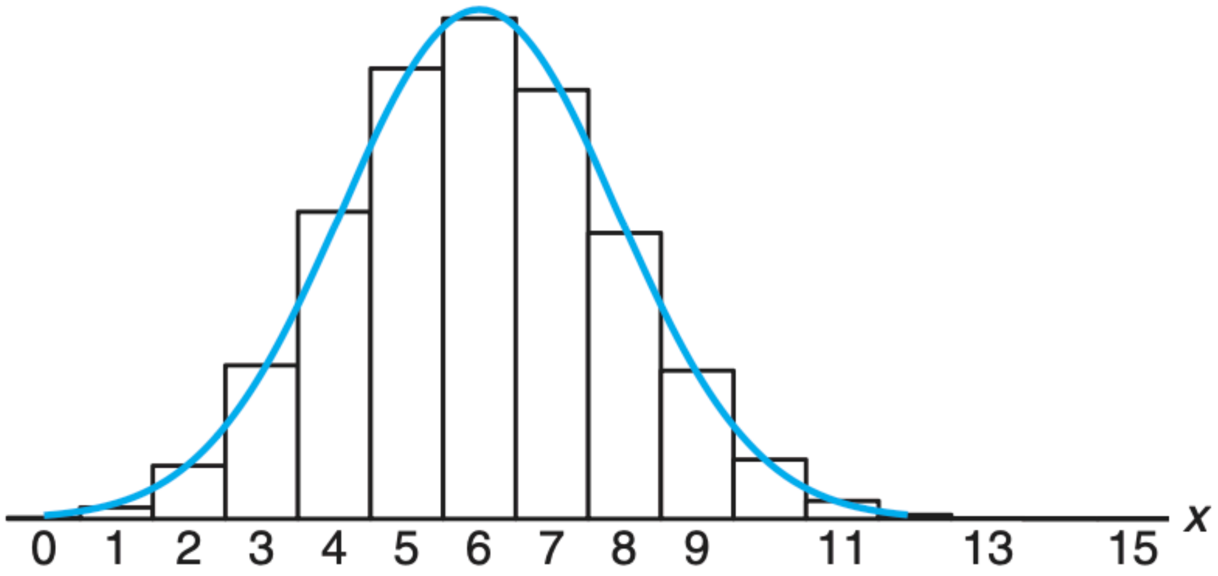


Histogram is for binomial RV with $n = 15, p = 0.4$.

Blue curve is normal RV with
$\mu = 15 \times 0.4 = 6, \sigma^2 = 15 \times 0.4 \times 0.6 = 3.6$

Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.



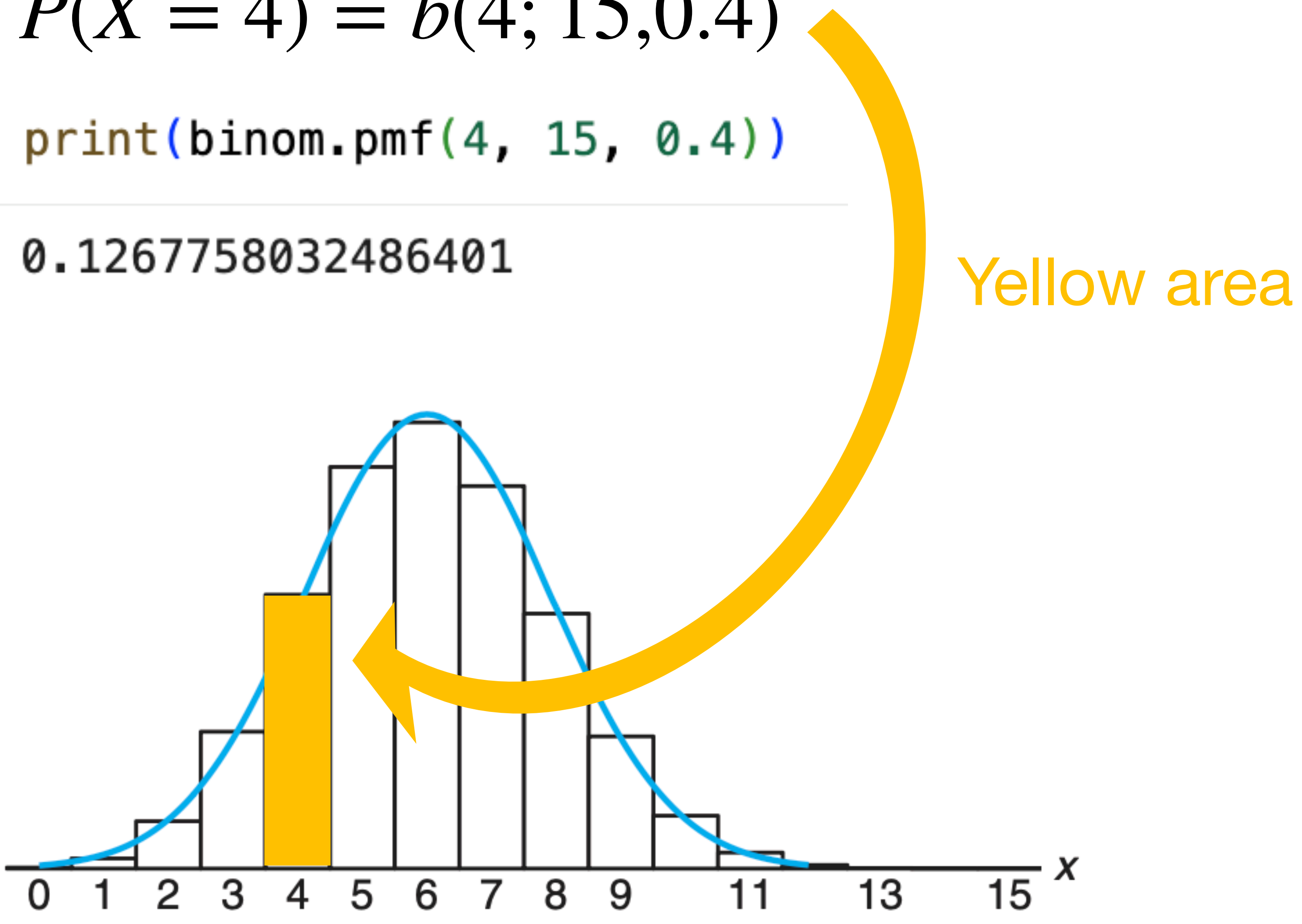
Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$$P(X = 4) = b(4; 15, 0.4)$$

```
print(binom.pmf(4, 15, 0.4))
```

```
0.1267758032486401
```



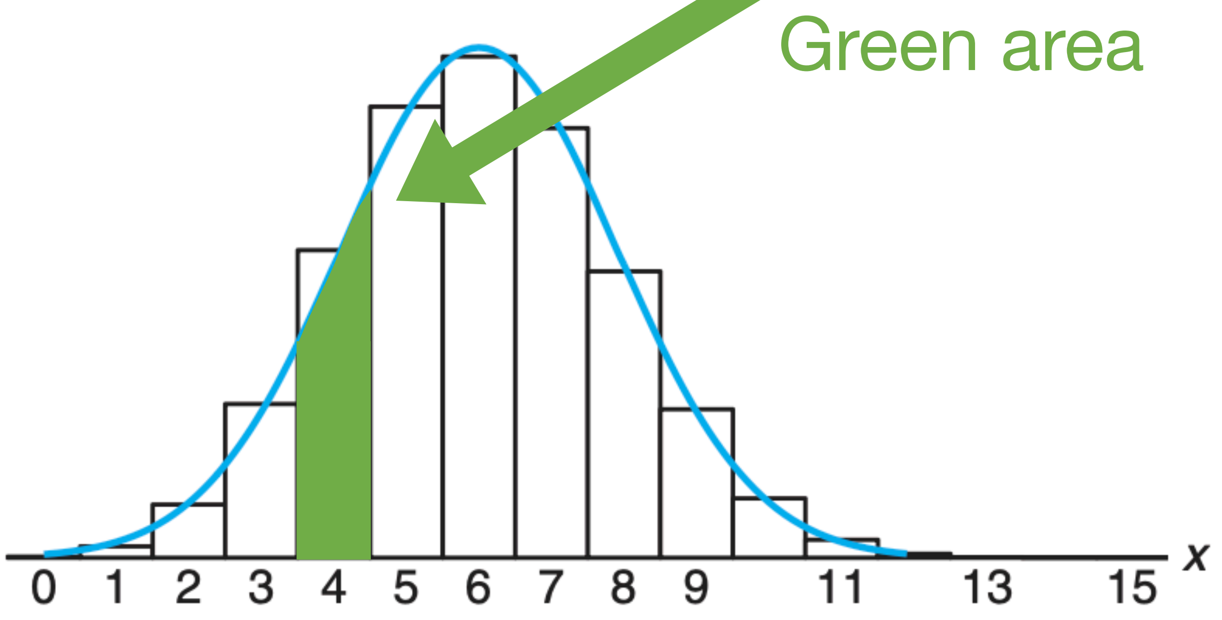
Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$P(X = 4) = b(4; 15, 0.4)$ $\approx$ Area under the corresponding normal curve between 3.5 and 4.5

```
print(binom.pmf(4, 15, 0.4))
```

```
0.1267758032486401
```



Approximating Binomial with Normal

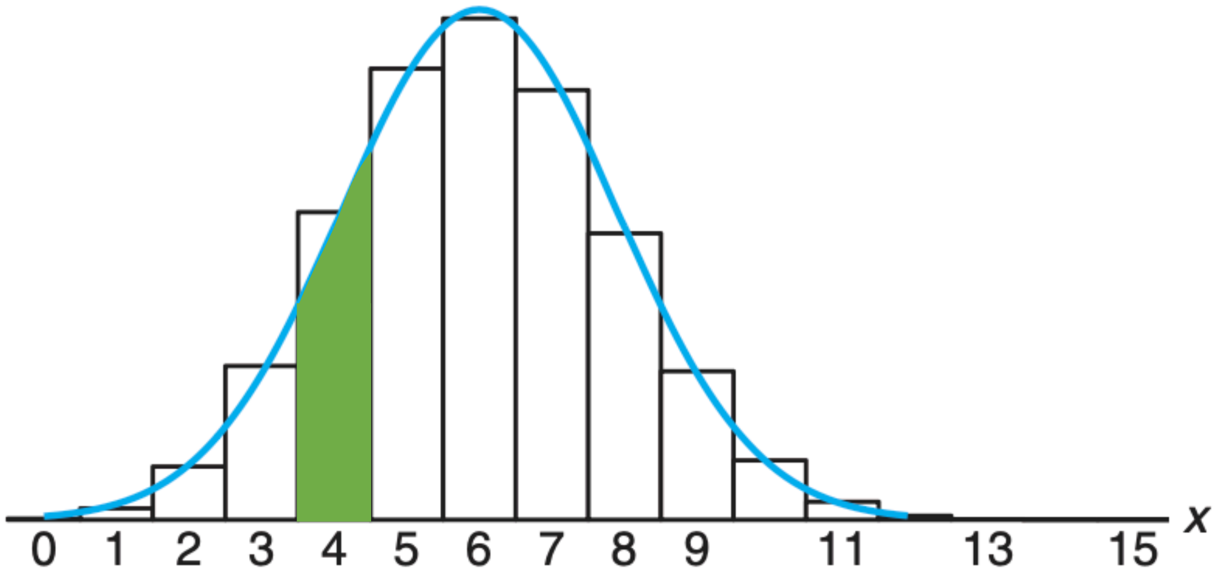
Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$P(X = 4) = b(4; 15,0.4)$ $\approx$ Area under the corresponding normal curve between 3.5 and 4.5

```
print(binom.pmf(4, 15, 0.4))
```

```
0.1267758032486401
```

Continuity correction



Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$P(X = 4) = b(4; 15, 0.4)$ $\approx$ Area under the corresponding normal curve between 3.5 and 4.5

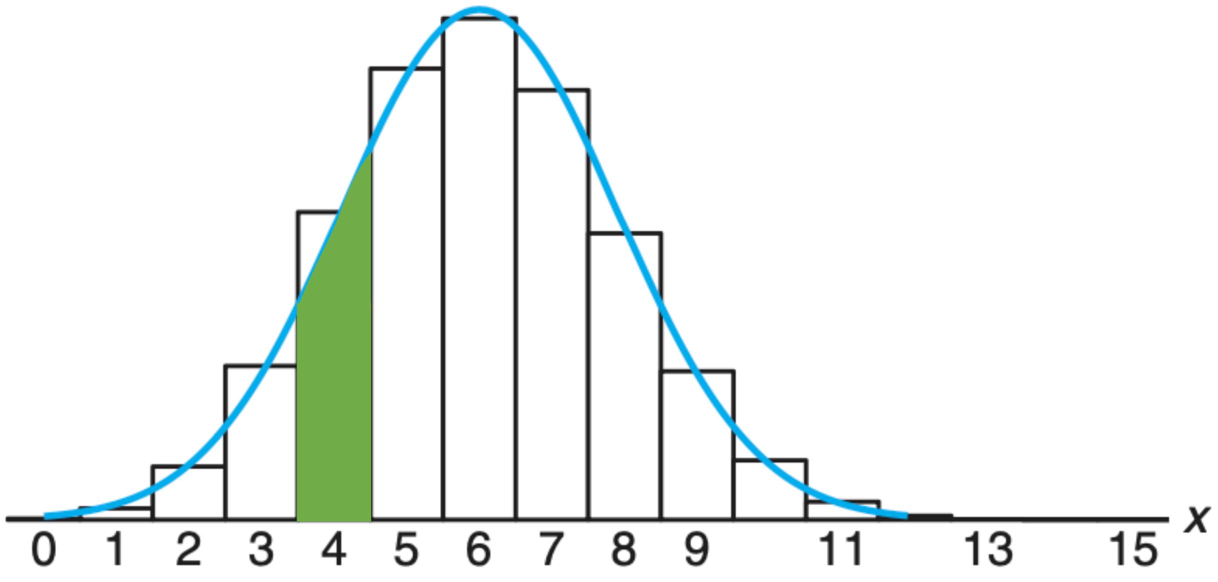
```
print(binom.pmf(4, 15, 0.4))
```

```
0.1267758032486401
```

$$P\left(\frac{3.5 - 6}{1.897} \leq Z \leq \frac{4.5 - 6}{1.897}\right)$$

```
print(norm.cdf(-0.79)-norm.cdf(-1.32))
```

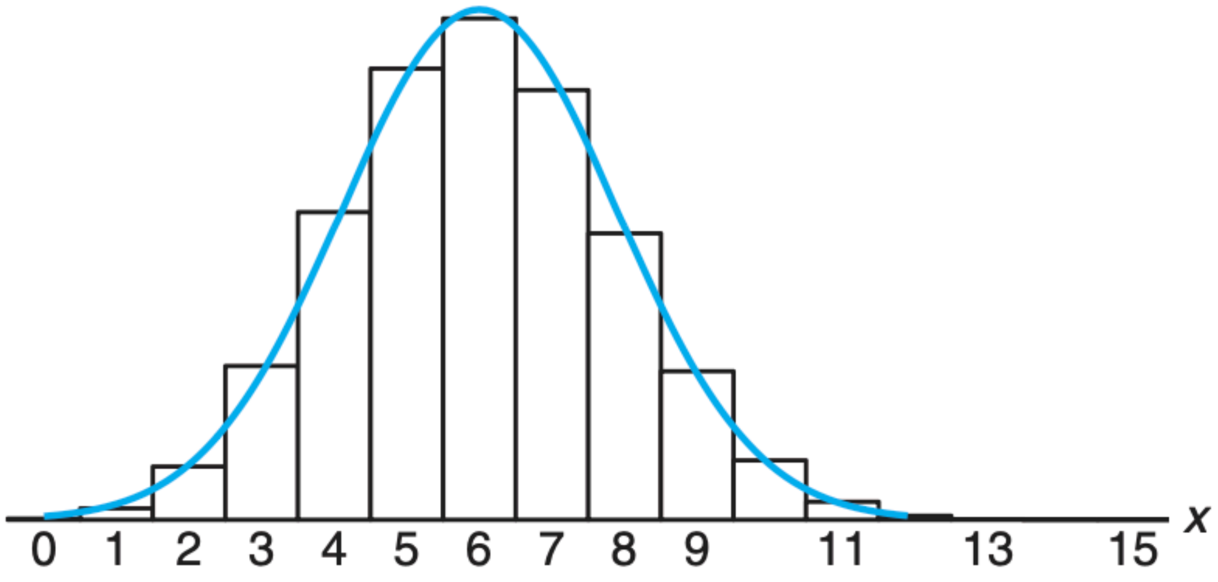
```
0.1213463751701653З
```



Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

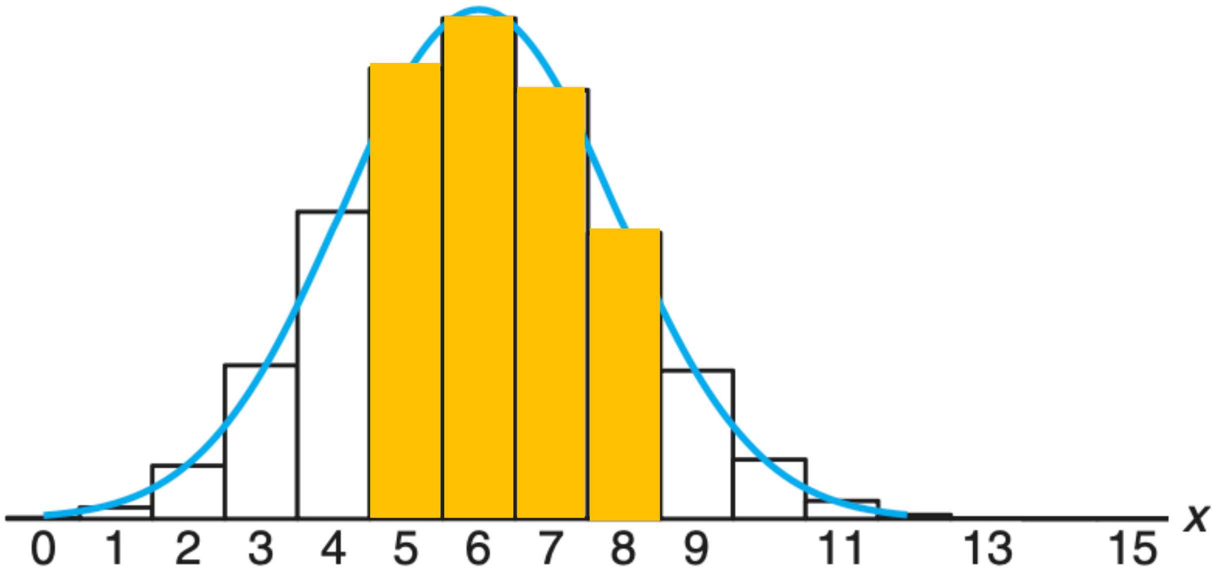
$P(5 \leq X \leq 8)$



Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$P(5 \leq X \leq 8)$

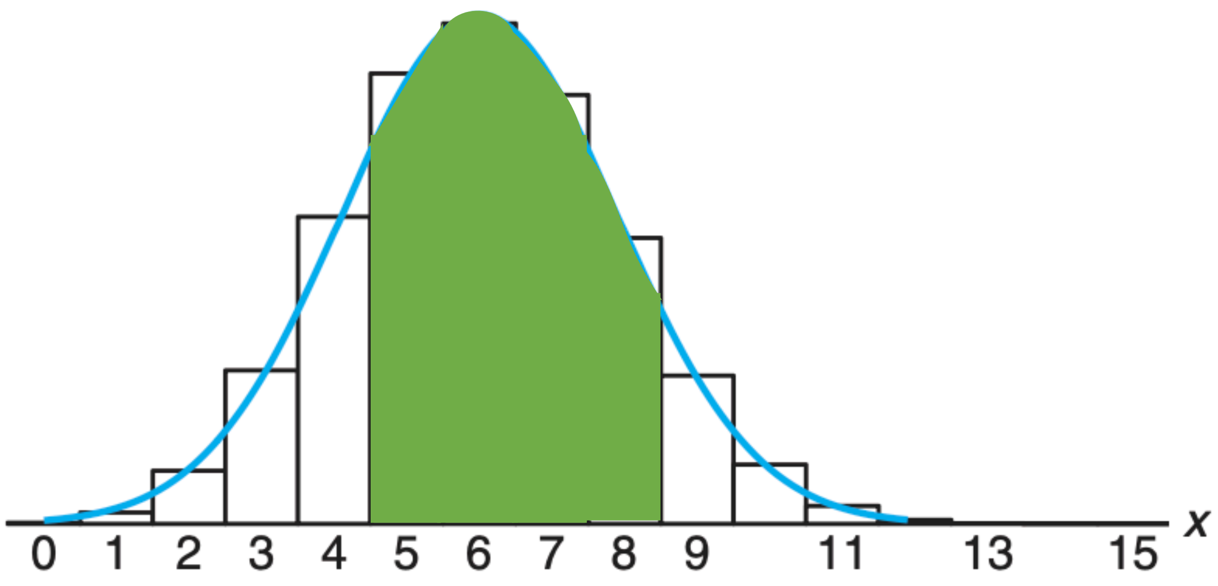


Approximating Binomial with Normal

Example: Consider the previous binomial RV X, where $n = 15$, $p = 0.4$.

$P(5 \leq X \leq 8)$ $\approx$ Approximate probability using Z:

$$P\left(\frac{4.5 - 6}{1.897} \leq Z \leq \frac{8.5 - 6}{1.897}\right)$$



Approximating Binomial with Normal

Example 6.16: A multiple-choice quiz has 200 questions, each with 4 possible answers of which only 1 is correct. What is the probability that sheer guesswork yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge? Use normal approximation. Leave your solution as $P(Z \leq k_1) - P(Z \leq k_2)$.

Outline

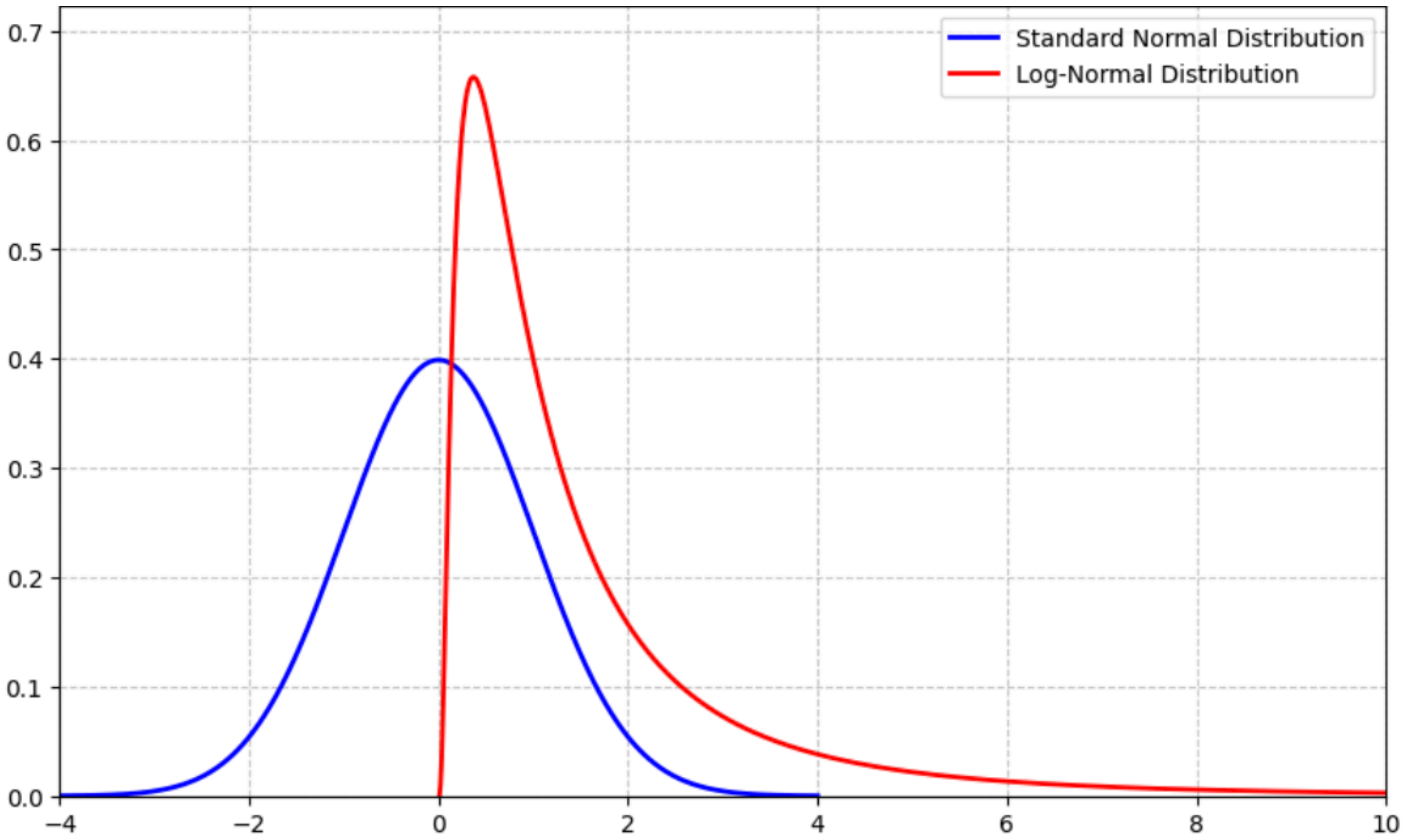
- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Lognormal Distribution

X is said to have lognormal distribution, if $ln(X)$ has normal distribution.

Lognormal Distribution

$\equiv$ e^X is said to have lognormal distribution, if X has normal distribution.



Horizontal axis:

X values

Horizontal axis:

e^X values

Lognormal Distribution

Previously we have seen that:

Linear combinations of normal random variables

$\Rightarrow$ normal random variable

Translates to:

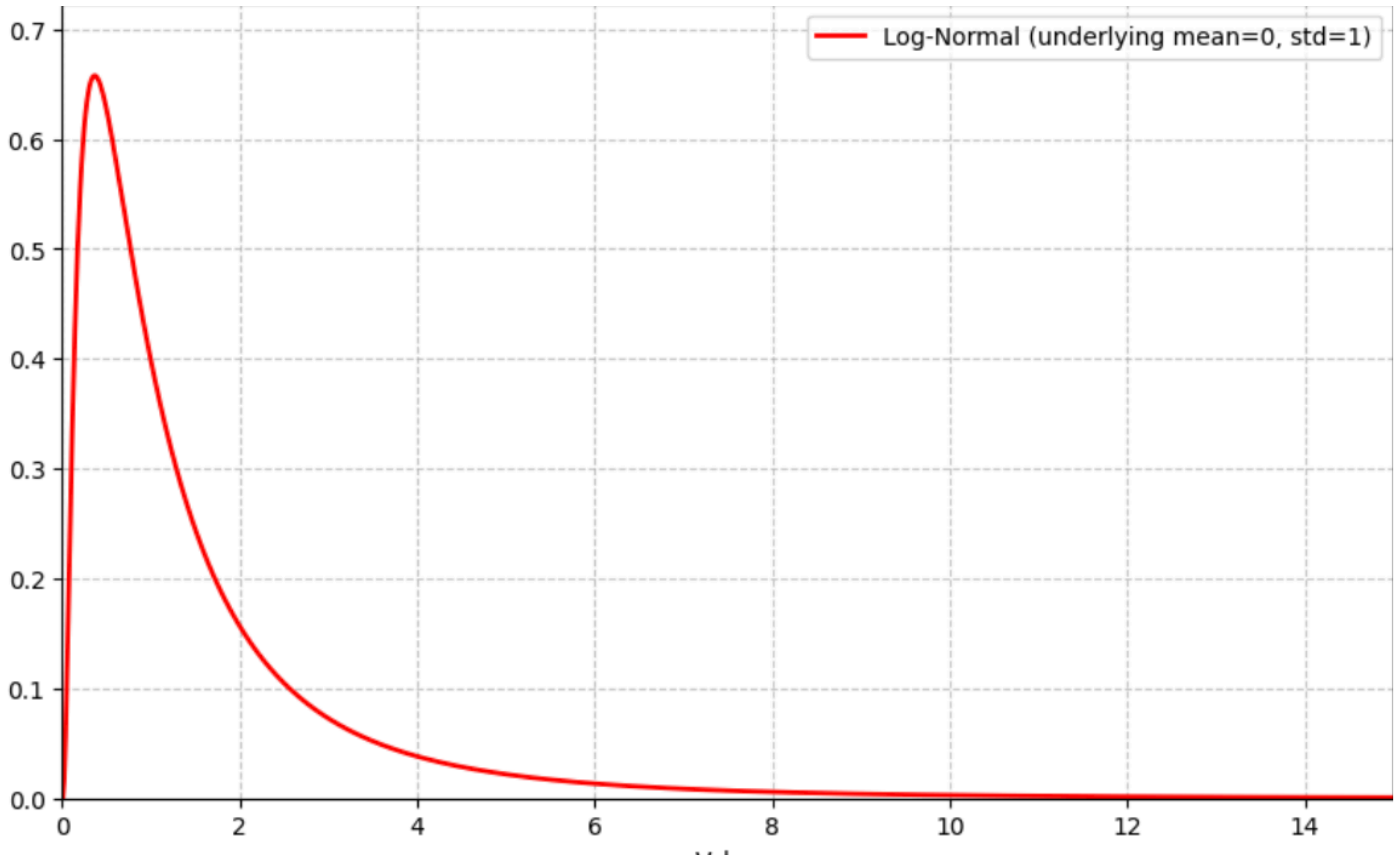
Product of lognormal random variables

$\Rightarrow$ lognormal random variable

(Consider $e^{a_1 X_1} \times e^{a_2 X_2} \times \cdots \times e^{a_n X_n}$ where $X_1, X_2, \cdots, X_n$ are normal independent random variables.)

Lognormal Distribution

If $ln(X)$ has normal distribution then X has lognormal distribution.



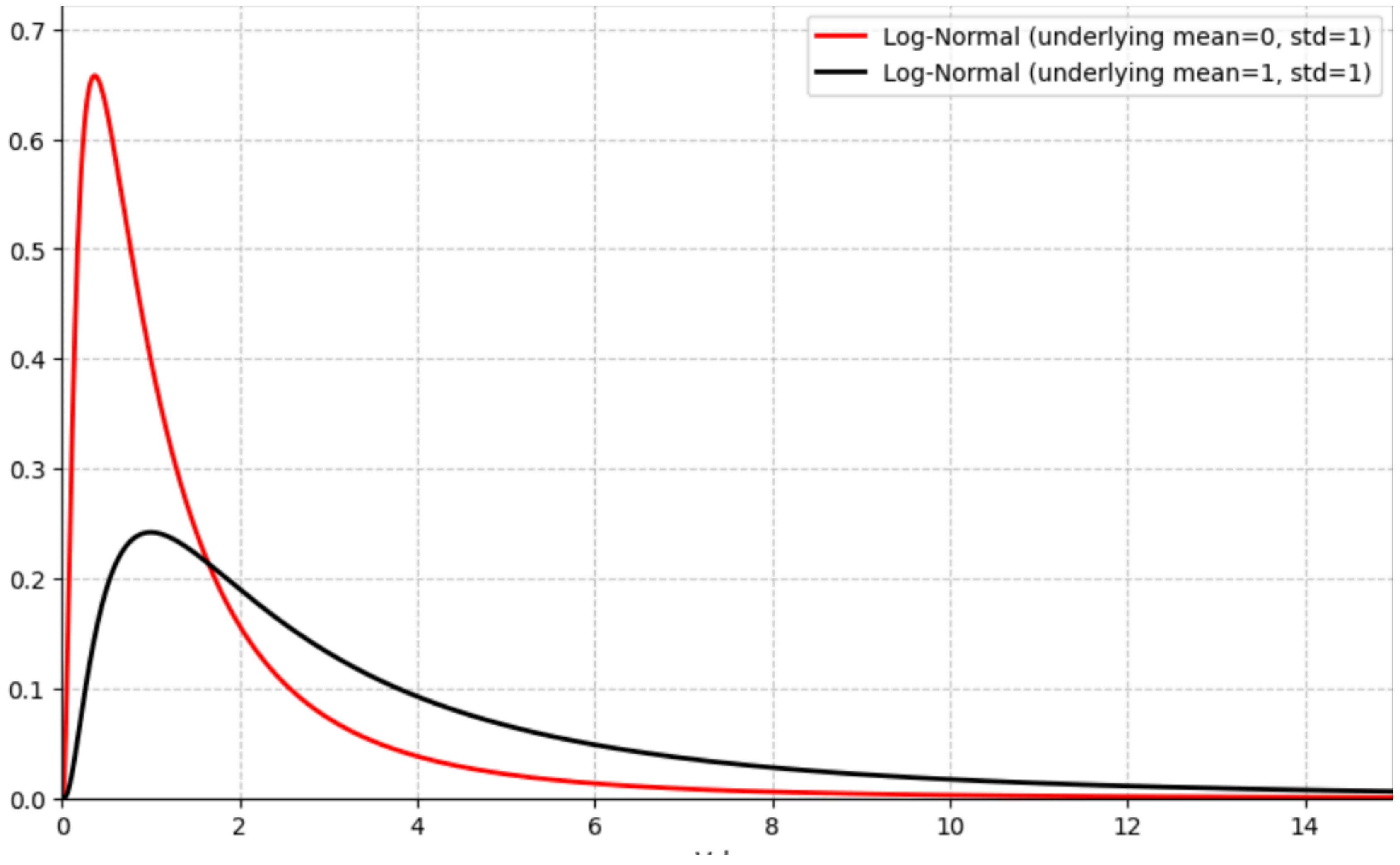
Mean of X: $e^{\mu+\sigma^2/2}$

Median of X: e^{μ}

Mean is larger than the median:
Skewed to the right.

Lognormal Distribution

What is modeled with lognormal?



Company size

Income: Median household income: $83,730, Mean: $121,000 (still doesn't capture extreme high income)

Human height (normal) vs weight (lognormal)

Some Exercises on Lognormal Distribution

Example 6.22: Concentrations of pollutants produced by chemical plants historically are known to exhibit behavior that resembles a lognormal distribution. This is important when one considers issues regarding compliance with government regulations. Suppose it is assumed that the concentration of a certain pollutant, in parts per million, has a lognormal distribution with parameters $\mu = 3.2$ and $\sigma = 1$. What is the probability that the concentration exceeds 8 parts per million?

Use norm.cdf$(-1.12) \approx 0.1314$

Some Exercises on Lognormal Distribution

Example 6.23: The life, in thousands of miles, of a certain type of electronic control for locomotives has an approximately lognormal distribution with $\mu = 5$ *and* $\sigma = 2$. Find the 5th percentile of the life of such an electronic control.

Use norm.ppf(0.05)$=-1.645$ and $e^{1.71} = 5.53$