L05: Random Variables – Continuous

Yiming QIN
Assistant Professor
School of Energy and Environment
City University of Hong Kong

Continuous Random Variable

Discrete Random Variable: A quantitative random variable that can assume a countable number of values.

Intuitively, a discrete random variable can assume values corresponding to isolated points along a line interval. That is, there is a gap between any two values.

Note: Usually associated with counting.

Continuous Random Variable: A quantitative random variable that can assume an uncountable number of values.

Intuitively, a continuous random variable can assume any value along a line interval, including every possible value between any two values.

Note: Usually associated with a measurement.

Continuous Random Variable

- **Probability Density Function (PDF) and Distribution Function**

  The probability DENSITY function of a CONTINUOUS rv X, denoted by $f(x)$, is a **function that gives us a value for the measure of how likely it is that X is near to x**. It is valid for all possible values x of X.

  ✓ **Conditions for a pdf:**
  - $0 < f(x)$, for all x in the range of X.
  - $\int_{-\infty}^{+\infty} f(x) dx = 1$

$$P\{a < x \leq b\} = \int_a^b f(x) dx$$

Continuous Random Variable

- **Probability Density Function (PDF) and Distribution Function**

The probability DENSITY function of a CONTINUOUS rv *X*, denoted by $f(x)$, is **a function that gives us a value for the measure of how likely it is that *X* is near to x**. It is valid for all possible values x of *X*.

✓ **Conditions for a pdf:**

- $0 < f(x)$, for all x in the range of X.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\mathrm{P}\{a < x \le b\} = \int_a^b f(x)dx$$

The probabilities are given by integrating $f(x)$ over a particular interval.

- Probability that *X* takes values belonging to A:

$$P(x \in A) = \int_{x \in A} f(x)dx$$

Cumulative Distribution Function

- **CDF: the cumulative distribution function *F* of a continuous random variable can be expressed in terms of the PDF *f*(*x*) by**

$$F(a) = \mathrm{P}\{-\infty < X \leq a\} = \int_{-\infty}^{a} f(x)\,dx$$

- **Differentiate the CDF**

$$\frac{d}{da} F(a) = \mathrm{f(a)}$$

Rules for Integrals

Power Rule

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C \quad n \neq -1$$

$$\int x^{-1} dx = \ln|x| + C$$

Exponential

$$\int e^x dx = e^x + C$$

$$\int a^x dx = \frac{a^x}{\ln a} + C$$

Constant Multiples

$$\int kf(x)dx = k\int f(x)dx$$

Absolute Value

$$\int |x| dx = \frac{x|x|}{2} + C$$

Sums and Differences

$$\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$$

$$\int [f(x) - g(x)]dx = \int f(x)dx - \int g(x)dx$$

Example

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the value of C? What is the CDF?

✓ **<u>Conditions for a pdf:</u>**

- $0 < f(x)$, for all x in the range of X.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

Example

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the value of C? What is the CDF?

- By the property $\int_{-\infty}^{\infty} f(x) = 1$, we have

✓ **Conditions for a pdf:**
- $0 < f(x)$, for all x in the range of X.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$C\int_0^2 (4x - 2x^2)dx = 1 \Longrightarrow C\left[2x^2 - \frac{2x^3}{3}\right]\Big|_{x=0}^{x=2} = 1$$

- So $C = \frac{3}{8}$.

Example

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the value of C? What is the CDF?

- By the property $\int_{-\infty}^{\infty} f(x) = 1$, we have

✓ **Conditions for a pdf:**
  - $0 < f(x)$, for all x in the range of X.
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$C\int_0^2 (4x - 2x^2)dx = 1 \Longrightarrow C\left[2x^2 - \frac{2x^3}{3}\right]\Big|_{x=0}^{x=2} = 1$$

- So $C = \frac{3}{8}$.

$$F(a) = P\{X \leq a\} = \int_{-\infty}^{a} f(x)dx = \int_0^a \frac{3}{8}(4x - 2x^2)dx = \frac{3}{4}a^2 - \frac{1}{4}a^3$$

Example : Continuous Random Variable

Example: Let c > 0 and the probability density function of a random variable X is

$$f(x) = c(1 - x^2) \qquad for \ -1 \leq x \leq 1,$$

And is 0 otherwise. Compute c as well as the cumulative distribution function of f.

Example : Continuous Random Variable

Example: Let $c > 0$ and the probability density function of a random variable X is

$$f(x) = c(1 - x^2) \qquad for \ -1 \leq x \leq 1,$$

And is 0 otherwise. Compute c as well as the cumulative distribution function of f.

Solution: since f is the density of a continuous random variable,

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-1}^{1} c(1 - x^2)dx = \frac{4}{3}c.$$

Hence $c = \frac{3}{4}$.

Example : Continuous Random Variable

Example: Let *c* > 0 and the probability density function of a random variable X is

$$f(x) = c(1 - x^2) \qquad for \quad -1 \le x \le 1,$$

And is 0 otherwise. Compute *c* as well as the cumulative distribution function of *f*.

Solution: since f is the density of a continuous random variable,

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-1}^{1} c(1 - x^2)dx = \frac{4}{3}c.$$

Hence $c = \frac{3}{4}$.

We denote the cumulative distribution function

$$\begin{array}{ll} F(x) = P(X \le x) = 0 & \text{for } x \le -1; \\ F(x) = \int_{-\infty}^{x} f(t)dt = \frac{3}{4}\int_{-1}^{x}(1 - t^2)dt = \frac{3}{4}[t - \frac{1}{3}t^3 + \frac{2}{3}] & \text{for } -1 \le x \le 1; \\ F(x) = P(X \le x) = 1 & \text{for } x \ge 1. \end{array}$$

Mean (Expectation)

Mean

If X is a continuous rv with its pdf $f(x)$, then the mean (expectation, expected value) of X is defined as

$$\boxed{E(X) = \int_{-\infty}^{\infty} [x\, f(x)]\, dx, \qquad \text{if it exists.}}$$

- The population mean is usually denoted by μ or μ_X.
- It is a common measure of central location of the random variable X.

Mean (Expectation)

Mean

If X is a continuous rv with its pdf $f(x)$, then the mean (expectation, expected value) of X is defined as

$$\boxed{E(X) = \int_{-\infty}^{\infty} [x\, f(x)]\, dx, \qquad if\ it\ exists.}$$

- The population mean is usually denoted by μ or μ_X.
- It is a common measure of central location of the random variable X.
- It is different from the sample mean, the mean of data.
  - Population mean is determined by the distribution of the random variable, while sample mean is determined by the collection of the actual observations of the random variable.
  - Thus, mean is **fixed** (even it is often unknown in practice) but sample mean is **different** when different data are used.

Example

- Suppose X has PDF $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, compute the expectation.

Example

- Suppose X has PDF $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, compute the expectation.

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$
$$= \int_{0}^{\infty} x e^{-x} dx$$

Example

- Suppose X has PDF $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, compute the expectation.

$$\int u dv = uv - \int v du$$ (Integration by parts)

$$u = x; v = -e^{-x}$$
$$dv = e^{-x}dx; du = dx$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$
$$= \int_0^{\infty} x e^{-x} dx$$

Example

- Suppose X has PDF $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, compute the expectation.

$$\int u dv = uv - \int v du$$ (Integration by parts)

$$u = x; v = -e^{-x}$$
$$dv = e^{-x}dx; du = dx$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$
$$= \int_0^{\infty} x e^{-x} dx$$
$$= \cdots$$
$$= 1$$

$$\int_0^{\infty} x e^{-x} dx$$
$$= [-xe^{-x}]_0^{\infty} + \int_0^{\infty} e^{-x} dx$$
$$= [-xe^{-x}]_0^{\infty} - [e^{-x}]_0^{\infty}$$
$$= 1$$

Law of the Unconscious Statistician (LOTUS)

LOTUS

- *If X is continuous*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

It provides a method for calculating the **expected value of a function of a random variable without directly knowing the probability distribution of the random variable itself**.

Linear of Expectation

The expectation of linear combination equals to the linear combination of the expectation

Linearity

For any constants a, b

$$E[aX + b] = aE[X] + b$$

Linearity – Multi

For any constants $a_1, a_2, \ldots, a_K$ and b

$$E[a_1X_1 + a_2X_2 + \cdots a_KX_K + b] = a_1E[X_1] + a_2E[X_2] + \cdots + a_KE[X_K] + b$$

Independence of Random Variables

INDEPENDENCE

Random variables X and Y are independent means for all x, y

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

EXPECTATION OF PRODUCT GIVEN INDEPENDENCE

If X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Example

Let X denote a continuous random variable with pdf $f(x)$:

$$f(x) = \begin{cases} 3x^2 & if\ 0 < x < 1, \\ 0 & otherwise. \end{cases}$$

Compute the expected value of $\boldsymbol{X^2}$.

Law of the Unconscious Statistician (LOTUS)

Example

Let *X* denote a continuous random variable with pdf *f(x)*:

$$f(x) = \begin{cases} 3x^2 & if\ 0 < x < 1\,, \\ 0 & otherwise. \end{cases}$$

Compute the expected value of ***X*²**.

Law of the Unconscious Statistician (LOTUS)

Solution

$$E\left(g(x)\right) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Let *g(x)* be *X*²

$$E(X^2) = \int_0^1 x^2 \cdot 3x^2 dx$$

$$E(X^2) = \frac{3}{5} \times \left[x^5\right]_0^1 = \frac{3}{5}$$

Variance

Variance

If X is a continuous rv with its pdf $f(x)$, then the population variance of X is defined as

$$\boxed{Var(X) = \int_{-\infty}^{\infty} [(x-\mu)^2 f(x)]\, dx, \qquad \textit{if it exists.}}$$

- The population variance is usually denoted by σ^2 or σ_X^2.
- The positive square root of σ^2, denoted by σ, is called the population standard deviation (sd) of X.

Variance

Variance

If X is a continuous rv with its pdf $f(x)$, then the population variance of X is defined as

$$\boxed{Var(X) = \int_{-\infty}^{\infty} [(x-\mu)^2 f(x)]\, dx, \qquad if\ it\ exists.}$$

- The population variance is usually denoted by σ^2 or σ_X^2.
- The positive square root of σ^2, denoted by σ, is called the population standard deviation (sd) of X.
- Both σ^2 and σ are common measures of spread of the random variable X.
- Population variance (sd) is **different** from the sample variance (sd), the variance (sd) of data.
  - Population variance (sd) is determined by the distribution of the random variable, while sample variance (sd) is determined by the collection of the actual observations of the random variable.
  - Thus, population variance (sd) is **fixed** (even it is often unknown in practice) but sample variance (sd) is **different** when different data are used.

It can be showed that:

$$\boxed{Var(X) = E(X^2) - [E(X)]^2}$$

Property of Variance

If X and Y are $\boxed{\text{independent,}}$ then

$$Var(X+Y) = Var(X) + Var(Y)$$

Example

Let X denote a continuous random variable with pdf *f(x)*: $f(x) = \begin{cases} x^3 & if\ 0 < x < 1\ , \\ 0 & otherwise. \end{cases}$

Compute the variance value of *X*.

Example

Let X denote a continuous random variable with pdf *f(x)*: $f(x) = \begin{cases} x^3 & if\ 0 < x < 1, \\ 0 & otherwise. \end{cases}$

Compute the variance value of *X*.

Solution:

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X) = \int_{-\infty}^{\infty} [x\, f(x)]\, dx,$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Example

Let X denote a continuous random variable with pdf *f(x)*: $f(x) = \begin{cases} x^3 & if\ 0 < x < 1, \\ 0 & otherwise. \end{cases}$

Compute the variance value of *X*.

Solution:

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X) = \int_0^\infty x x^3 dx = \frac{1}{5} x^5 \Big|_0^1 = \frac{1}{5}$$

$$E(X^2) = \int_0^\infty x^2 x^3 dx = \frac{1}{6} x^6 \Big|_0^1 = \frac{1}{6}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{6} - (\frac{1}{5})^2 = \frac{19}{150}$$

$$E(X) = \int_{-\infty}^{\infty} [x\, f(x)]\, dx,$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Discrete and Continuous

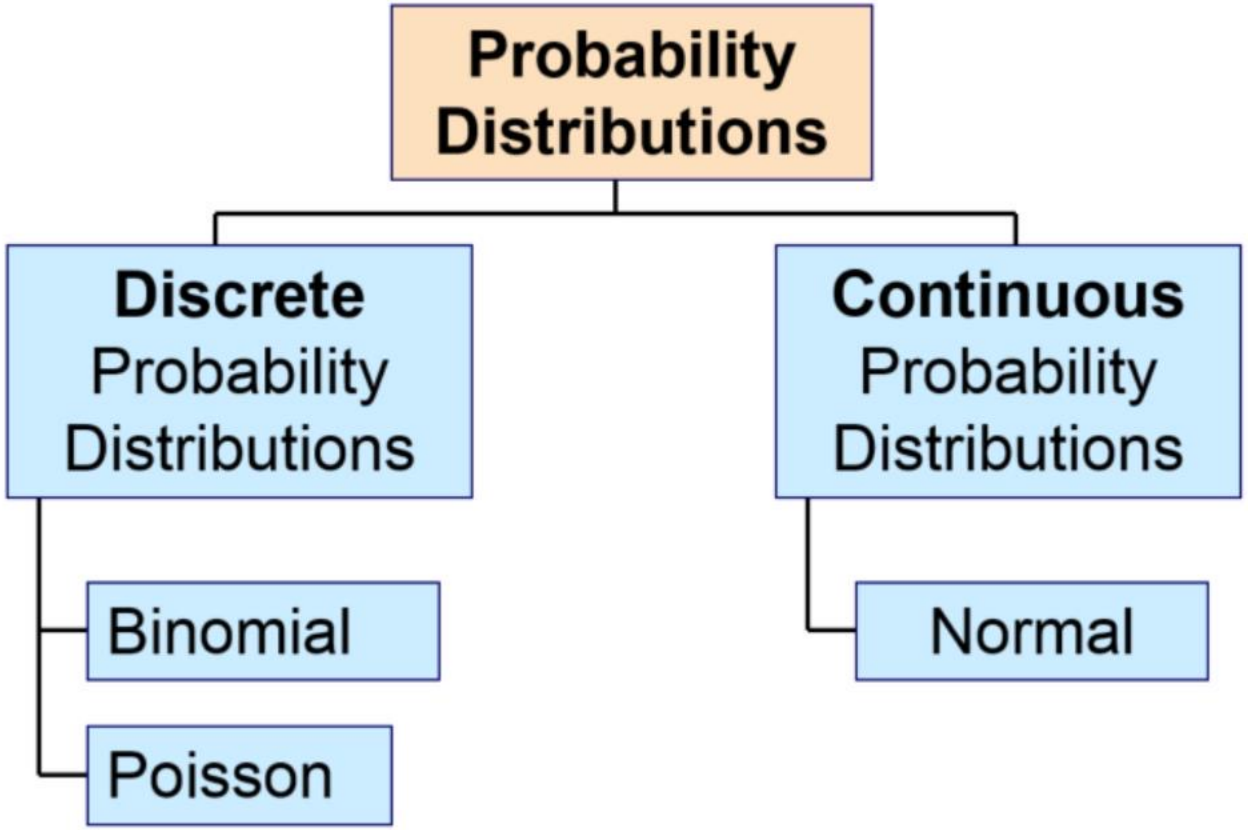
- Discrete Distribution
  - The distribution of a random variable X is discrete in the sense that all the possible values of X are isolated points.
  - Its cdf only increases at jump points.

- Continuous Distribution
  - Describes events over a continuous range, where the probability of a specific outcome is zero.
  - Its cdf is continuous.

	Discrete	Continuous
P*F	PMF $p(x_i) = P(X = x_i)$	PDF $P(a \le X \le b) = \int_a^b f(x)dx$
CDF	$F(a) = P(X \le a) = \sum_{x_i \le a} p(x_i)$	$F(a) = P(X \le a) = \int_{-\infty}^a f(x)dx$
LOTUS	$E[g(X)] = \sum g(x_i)\, P(x_i)$	$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)dx$

Well-Known Distributions

The following well-known distributions are important in statistics as many results are derived for them leading to quick analyses

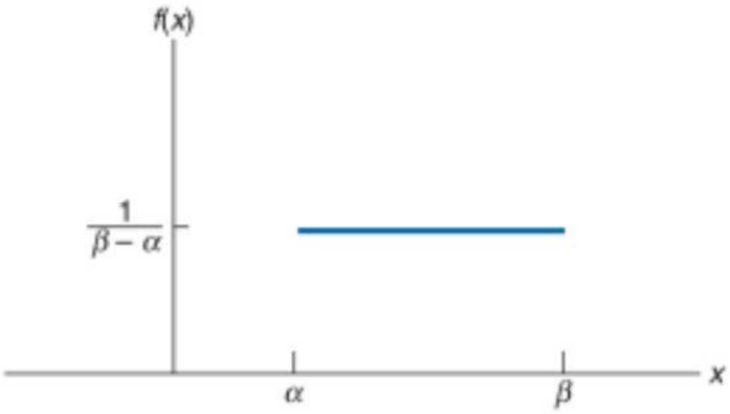


Uniform Random Variables

A random variable X is uniformly distributed over the interval $[\alpha, \beta]$ if its probability density function is given by

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

It is clear that $\int_{-\infty}^{\infty} f(x)dx = 1$



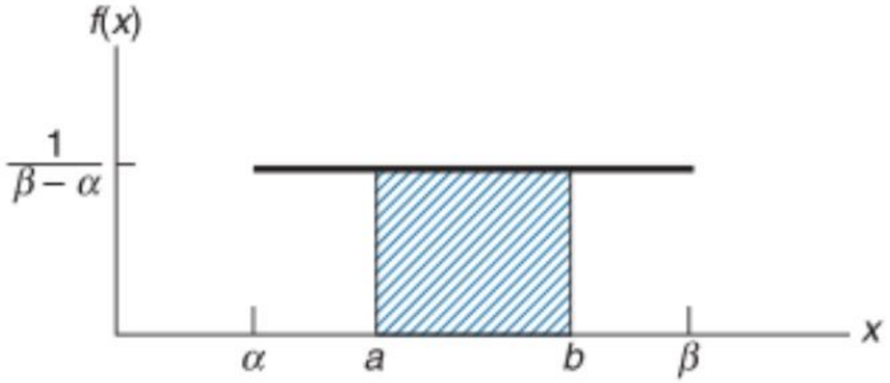
Graph of $f(x)$ for a uniform $[\alpha, \beta]$.

The uniform distribution can be either continuous or discrete, depending on whether the random variable is continuous or discrete.

Uniform Random Variables

How to calculate probabilities:

$$P\{a < X < b\} = \int_a^b f(x)dx = \frac{1}{\beta - \alpha}\int_a^b 1dx$$
$$= \frac{b-a}{\beta-\alpha}$$



Probabilities of a uniform random variable.

Uniform Random Variables

Mean (middle point)

$$E[X] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{x^2}{2(\beta - \alpha)} \Big|_{\alpha}^{\beta}$$
$$= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{(\beta + \alpha)(\beta - \alpha)}{2(\beta - \alpha)}$$
$$= \frac{\beta + \alpha}{2}$$

The center or balance point of the distribution.

Variance (square of width)

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= \int_{\alpha}^{\beta} \frac{x^2}{\beta - \alpha} dx - (\frac{\beta + \alpha}{2})^2$$
$$= \frac{(\beta^3 - \alpha^3)}{3(\beta - \alpha)} - \left(\frac{\beta + \alpha}{2}\right)^2$$
$$= \frac{(\beta - \alpha)^2}{12}$$

The spread or dispersion of the values around the mean. The wider range, the more spread out.

Example

Given that the lifetimes of a certain brand of lightbulbs follow a **uniform distribution between 1000 hours and 2000 hours**. What is the probability of lasting **more than 1500 hours?**

Solution:

The pdf is 1/1000 within the range [1000, 2000].

P = Integral of pdf from 1500 to 2000

P= (1/1000) * (2000 - 1500)= (1/1000) * 500 = 0.5 or 50%

Example

Let's say we have a uniform distribution on the interval [0, 10].

Calculate the following:

1. What is the probability of randomly selecting a value between 2 and 6 from this uniform distribution?
2. Calculate the mean of the uniform distribution on the interval [0, 10].
3. Calculate the variance of the uniform distribution on the interval [0, 10].

Example

Let's say we have a uniform distribution on the interval [0, 10].
Calculate the following:

1. What is the probability of randomly selecting a value between 2 and 6 from this uniform distribution?
2. Calculate the mean of the uniform distribution on the interval [0, 10].
3. Calculate the variance of the uniform distribution on the interval [0, 10].

Solution:

Probability = (Length of [2, 6]) / (Length of [0, 10]) = 4 / 10 = 0.4

Mean = (Minimum + Maximum) / 2 = (0 + 10) / 2 = 5

Variance = ((Maximum - Minimum)^2) / 12 =((10 - 0)^2) / 12 = 100 / 12 = 8.33

Exponential Random Variables

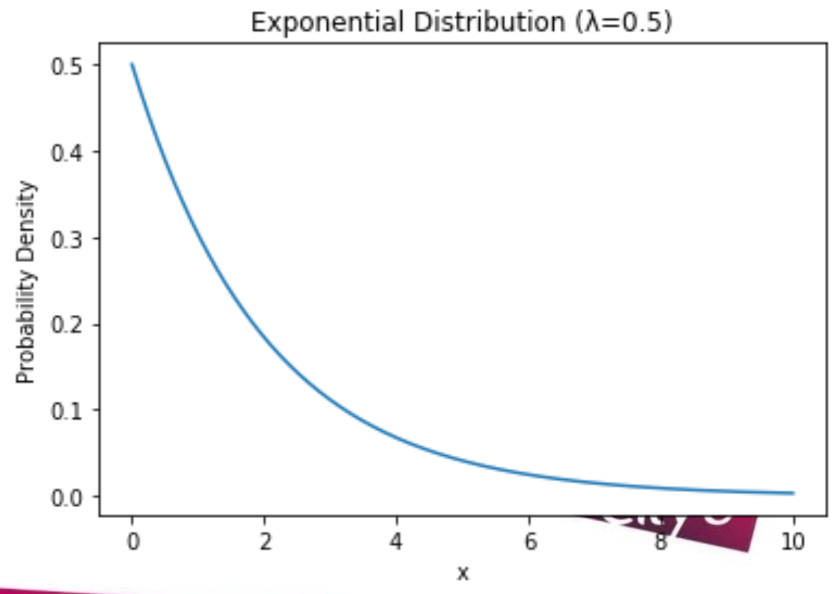
Exponential random variables are a type of continuous probability distribution that models the time between events occurring in a Poisson process.

Exponential RVs mean the time until the next event occurs (for each process) follows an exponential distribution.

An exponential random variable has the following density function:

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

Note that the possible values it can take: [0, ∞)



Exponential Random Variables

Exponential random variables are a type of continuous probability distribution that models the time between events occurring in a Poisson process.

An exponential random variable has the following density function:

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

Note that the possible values it can take: [0, ∞)

The cumulative distribution function is

$$\boxed{\begin{aligned} F(a) &= \int_0^a \lambda e^{-\lambda x} dx \\ &= 1 - e^{-\lambda a}, \quad a \geq 0 \end{aligned}}$$

Exponential Random Variables

Exponential random variables are a type of continuous probability distribution that models the time between events occurring in a Poisson process.

An exponential random variable has the following density function:

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

Expectation: $E[X] = \frac{1}{\lambda}$

Variance: $Var(X) = \frac{1}{\lambda^2}$

Note that the possible values it can take: [0, ∞)

The cumulative distribution function is

$$\begin{aligned} F(a) &= \int_0^a \lambda e^{-\lambda x} dx \\ &= 1 - e^{-\lambda a}, \quad a \geq 0 \end{aligned}$$

CityU

Exponential Random Variables

Expectation: $E[X] = \frac{1}{\lambda}$

An exponential random variable has the following density function:

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

Note that the possible values it can take: [0, ∞)

$$E(X) = \int_{-\infty}^{\infty} [x\, f(x)]\, dx$$

$$\int_0^{\infty} \lambda x e^{-\lambda x}\, dx$$

$$u = \lambda x;\ v = -e^{-\lambda x}$$

$$du = \lambda dx;\ dv = \lambda e^{-\lambda x} dx$$

$$\int_0^{\infty} \lambda x e^{-\lambda x}\, dx$$

$$= \frac{1}{\lambda}\int_0^{\infty} u dv = \frac{1}{\lambda}\left[uv - \int_0^{\infty} v du\right]$$

$$= [-xe^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx$$

$$= 0 - \frac{1}{\lambda}\left[e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}$$

Exponential Random Variables

$$Var(X) = \frac{1}{\lambda^2}$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^\infty \lambda x^2 e^{-\lambda x}\,dx$$

$$u = \lambda x^2; v = -e^{-\lambda x}$$

$$du = 2\lambda x dx; dv = \lambda e^{-\lambda x} dx$$

$$\int_0^\infty \lambda x^2 e^{-\lambda x}\,dx$$

$$= \frac{1}{\lambda}\left[uv - \int_0^\infty v du\right]$$

$$= [-x^2 e^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} 2x dx$$

$$= 0 + 2\int_0^\infty x e^{-\lambda x} dx$$

$$= \frac{2}{\lambda}[-xe^{-\lambda x}]_0^\infty + \frac{2}{\lambda}\int_0^\infty e^{-\lambda x} dx$$

$$= \left(\frac{2}{\lambda}\right)\left(-\frac{1}{\lambda}\right)[e^{-\lambda x}]_0^\infty = \frac{2}{\lambda^2}$$

Expectation (proved at the previous slide): $E(X) = \frac{1}{\lambda}$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Exponential Random Variables

Competition of Two Exponential RV's

If

$$X \sim \text{Exponential}(\lambda_1), \quad Y \sim \text{Exponential}(\lambda_2)$$

And *X*, *Y* are independent, then

$$\boxed{P\{X < Y\} = \frac{\lambda_1}{\lambda_1 + \lambda_2}}$$

Exponential RVs mean the time until the next event occurs (for each process) follows an exponential distribution.

Which process will experience the event first?

Exponential Random Variables

Competition of Two Exponential RV's

If

$$X \sim \text{Exponential}(\lambda_1), \quad Y \sim \text{Exponential}(\lambda_2)$$

Exponential RVs mean the time until the next event occurs (for each process) follows an exponential distribution.

And *X*, *Y* are independent, then

$$\boxed{P\{X < Y\} = \frac{\lambda_1}{\lambda_1 + \lambda_2}}$$

Which process will experience the event first?

If

$$X_1 \sim \text{Exponential}(\lambda_1), \cdots, X_n \sim \text{Exponential}(\lambda_n)$$

And $X_1, \ldots, X_n$ are independent, then

$$\boxed{P\{X_i \text{ is the smallest}\} = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}}$$

CityU

Example

Two ATM's at a bank. Service time at ATM A is Exp(2), service time at ATM B is Exp(1). When you arrive, you find that Tony is using A and Paul is using B. How likely Tony will leave first?

$\lambda_A = 2; \lambda_B = 1$

$$\boxed{P\{X < Y\} = \frac{\lambda_1}{\lambda_1 + \lambda_2}}$$

Example

Two ATM's at a bank. Service time at ATM A is Exp(2), service time at ATM B is Exp(1). When you arrive, you find that Tony is using A and Paul is using B. How likely Tony will leave first?

$\lambda_A = 2; \lambda_B = 1$

$$\boxed{P\{X < Y\} = \frac{\lambda_1}{\lambda_1 + \lambda_2}}$$

Solution:

- Tony's remaining service time (call it *X*) is still Exp(2)
- Paul's remaining service time (call it *Y*) is also Exp(1).

So, by the "competition property" the probability is

$$P\{X < Y\} = \frac{2}{2+1} = \frac{2}{3}$$

Normal Distribution

- The **normal probability distribution** is the most important distribution in all of statistics.
- Many continuous random variables have normal or approximately normal distributions.

It is sometimes called the Guassion distribution because it was given by Johann Carl Friedrich Gauss---the Prince of Mathematics.

Normal Distribution

Johann Carl Friedrich Gauss (30 April 1777 – 23 February 1855) was a German mathematician whose contributions included too many fields, such as number theory, algebra, statistics, analysis, differential geometry, geophysics, mechanics, mechanics, electrostatics, astronomy and optics.

From 1989 through 2001, Gauss's portrait and a normal density curve were featured on the German ten-mark banknote.

More about Gauss can be found on http://www.storyofmathematics.com/19th_gauss.html

Normal Distribution

Normal Probability Distribution:

Normal probability distribution function:

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

μ: mean
σ: standard deviation

The probability that *x* lies in some interval is the area under the curve.

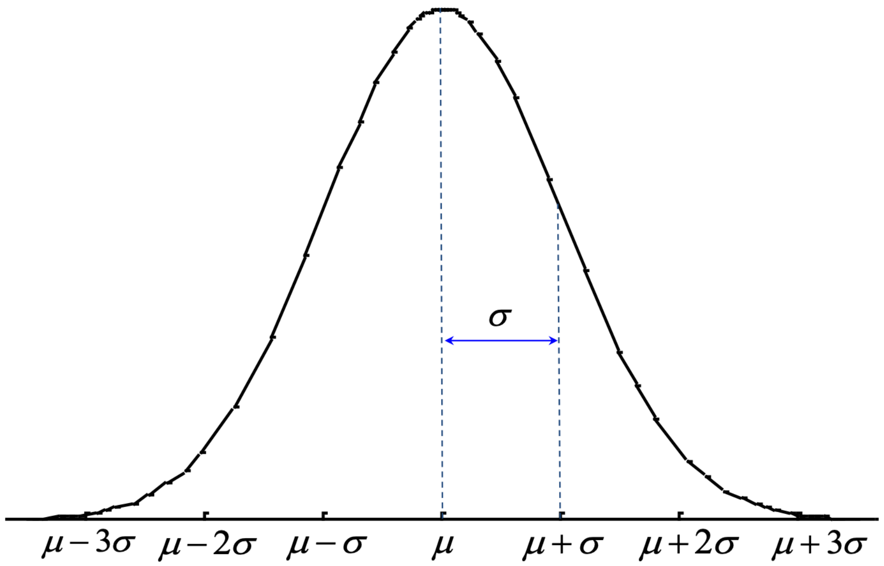
- Notation:

$$X \sim N(\underset{a}{\mu}, \underset{b}{\sigma^2}), \text{ where } \mu \in (-\infty, \infty) \text{ and } \sigma \in (0, \infty).$$

- Population mean and population variance:

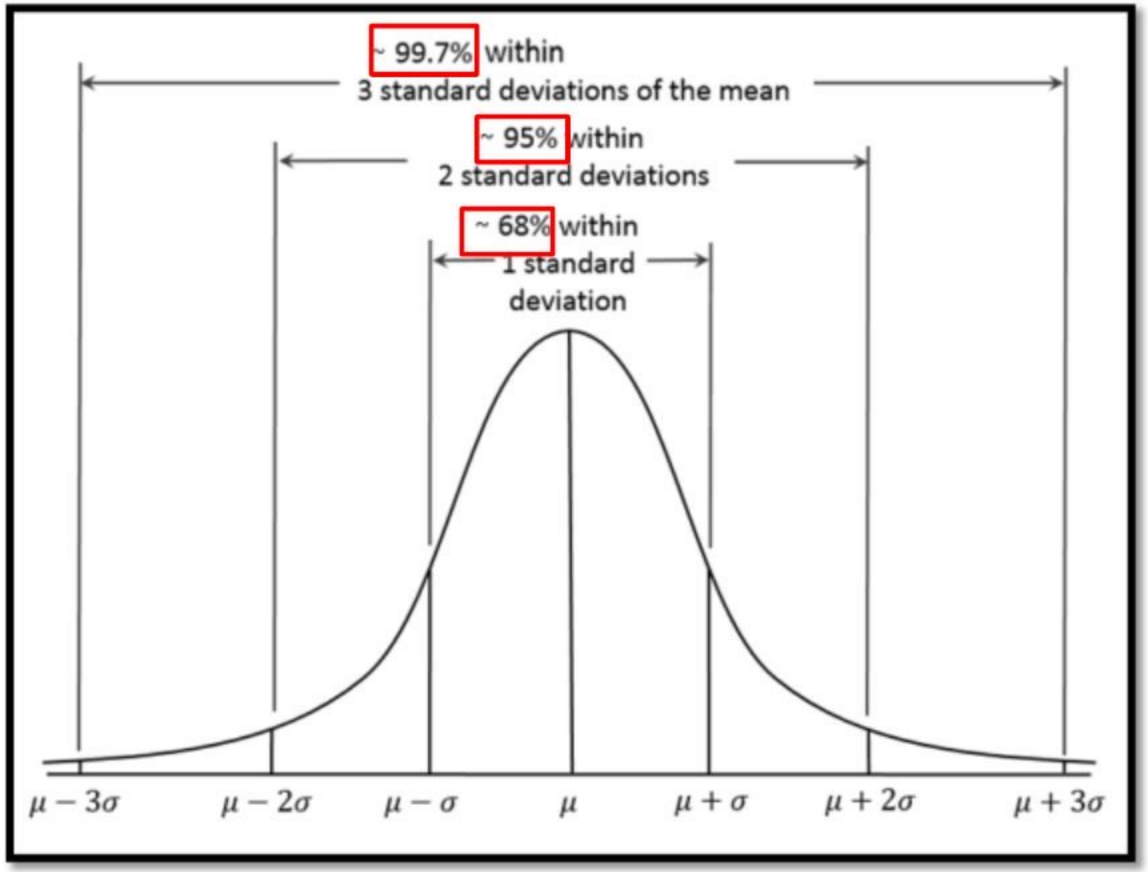
$$E(X) = a,\ Var(X) = b.$$

This is the function for the **normal (bell-shaped) curve**.

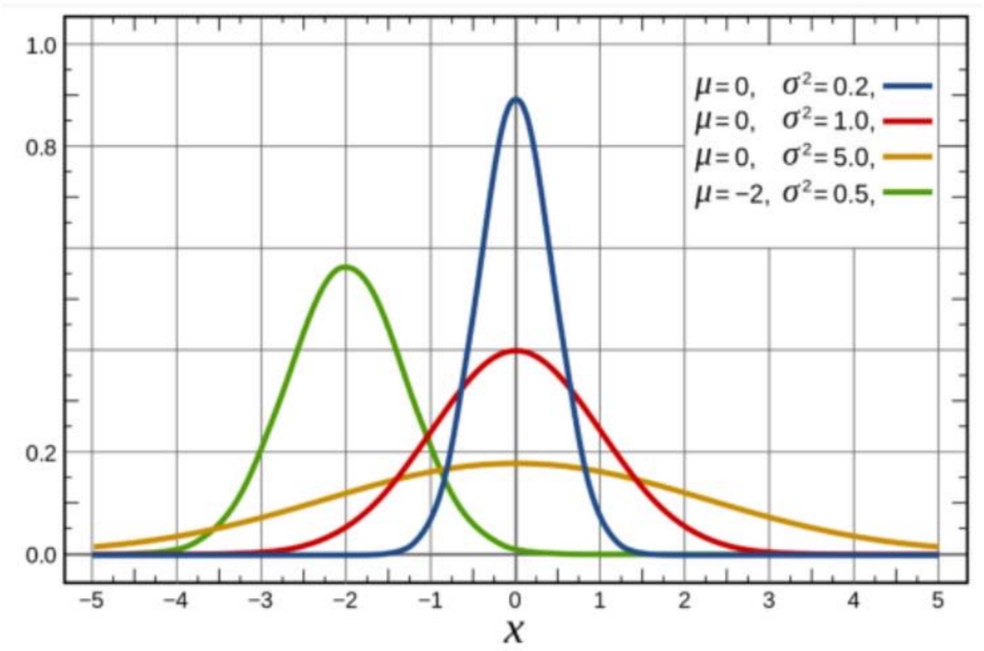


Normal Distribution

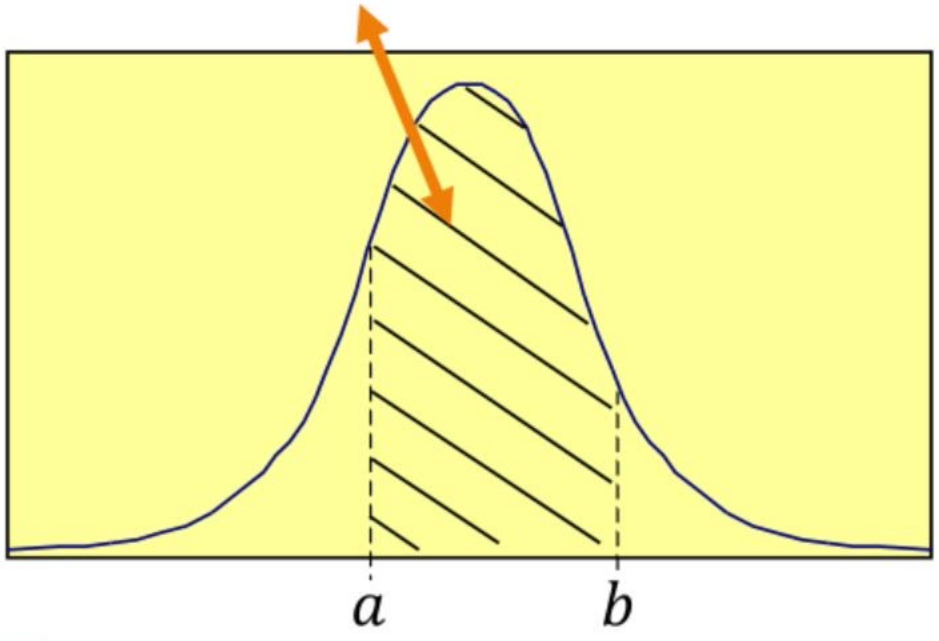
- In particular, we will have the following results in terms of μ and σ:



Normal Distribution



- Note that all normal distributions have the bell-shaped density curve regardless of the values of μ and σ.
- The area under the normal density curve over the interval from *a* to *b* represents the probability that the normal-distributed rv *X* falls into that interval, i.e., P($a \leq X \leq b$).



Standard Normal Distribution

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

$$=\frac{1}{1\times\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} \qquad -\infty\le x\le\infty$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z)^2} \qquad -\infty\le z\le\infty \qquad \left(\frac{x-0}{1}=z\right)$$

Standard normal distribution

The normal distribution with mean 0 and variance 1, i.e. N(0,1).

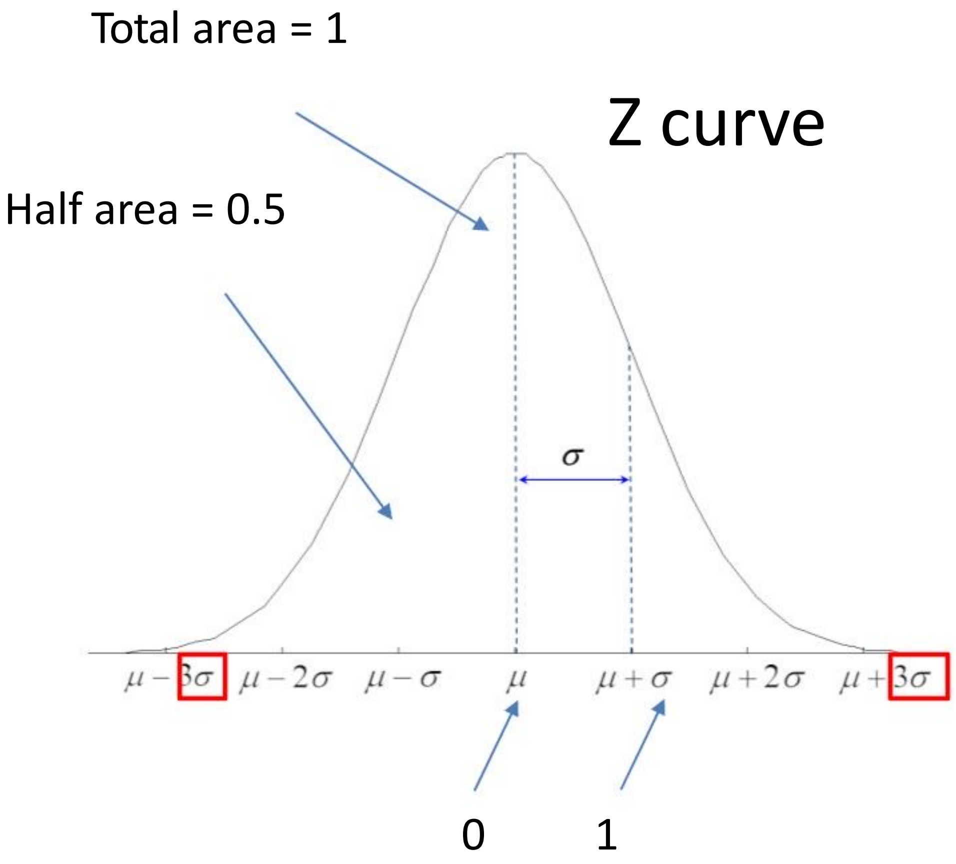
The random variable following the standard normal distribution is often denoted by *Z* in probability and statistics

Any normal probability can be reduced to the probability of N(0,1). If we want to find an area over an interval ***for any normal curve*** by just finding the corresponding area ***under a standard normal density curve***.

Standard Normal Distribution

Properties of the *Standard* Normal Distribution:

- The total area under the curve is 1. The distribution is symmetric.
- The distribution has a mean of 0 and standard deviation of 1.
- The mean divides the area in half, 0.5 on each side.
- Nearly all the area is between the standard variable $z = -3$ and $z = 3$.



Standardization

- Standardization is a process used to transform any normal-distributed r.v., say X~N(μ, σ^2) , to a standard normal-distributed r.v.

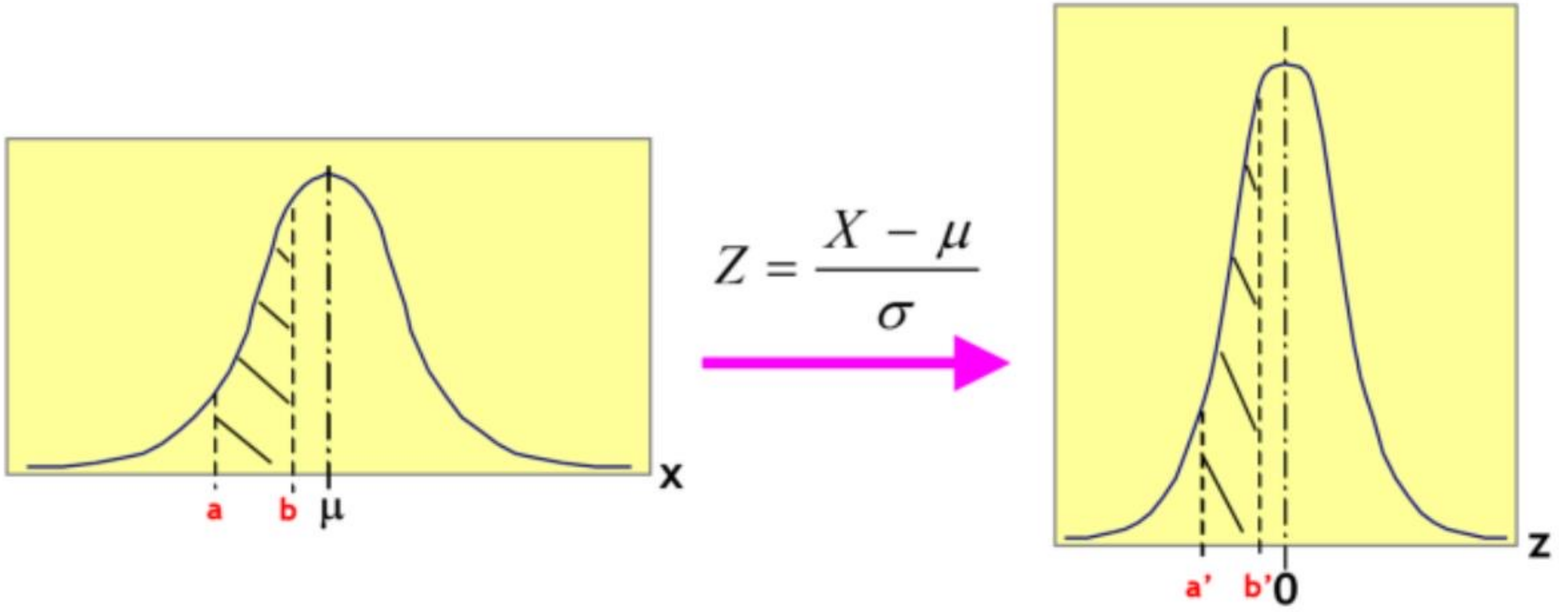
- To be more precise, we have the following result:

If $X \sim N(\mu, \sigma^2)$, and then we standardize X, i.e. subtract μ from X and then divide by σ then

$$\frac{X-\mu}{\sigma} \sim N(0,1) .$$

Thus, notationally, we have $\boxed{Z = \frac{X-\mu}{\sigma}}$.

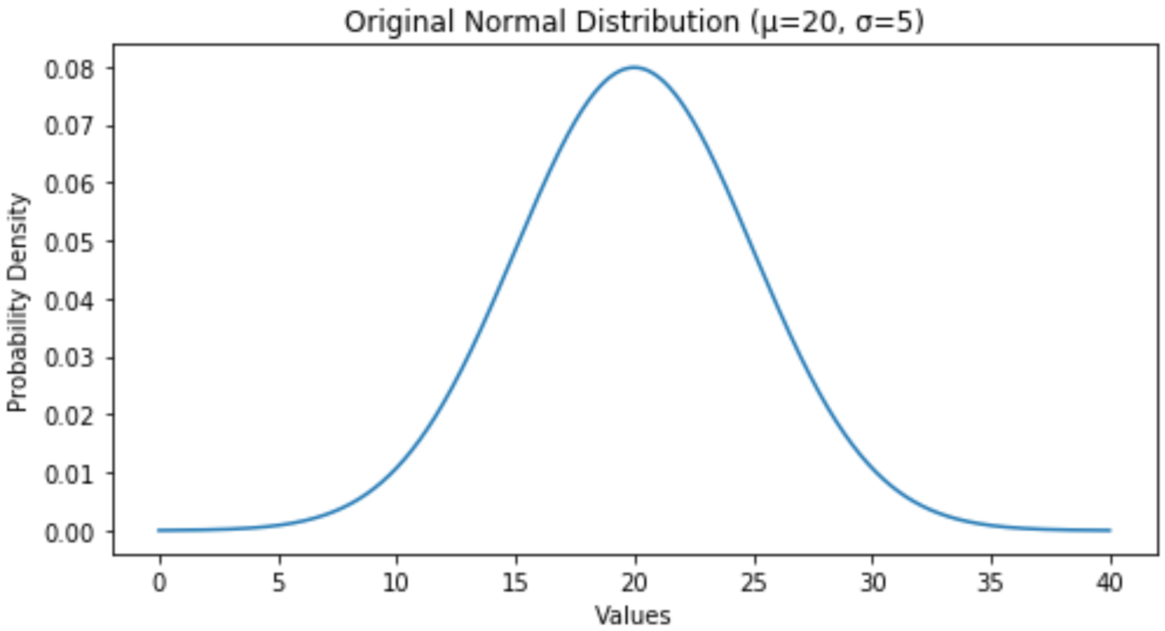
Standardization

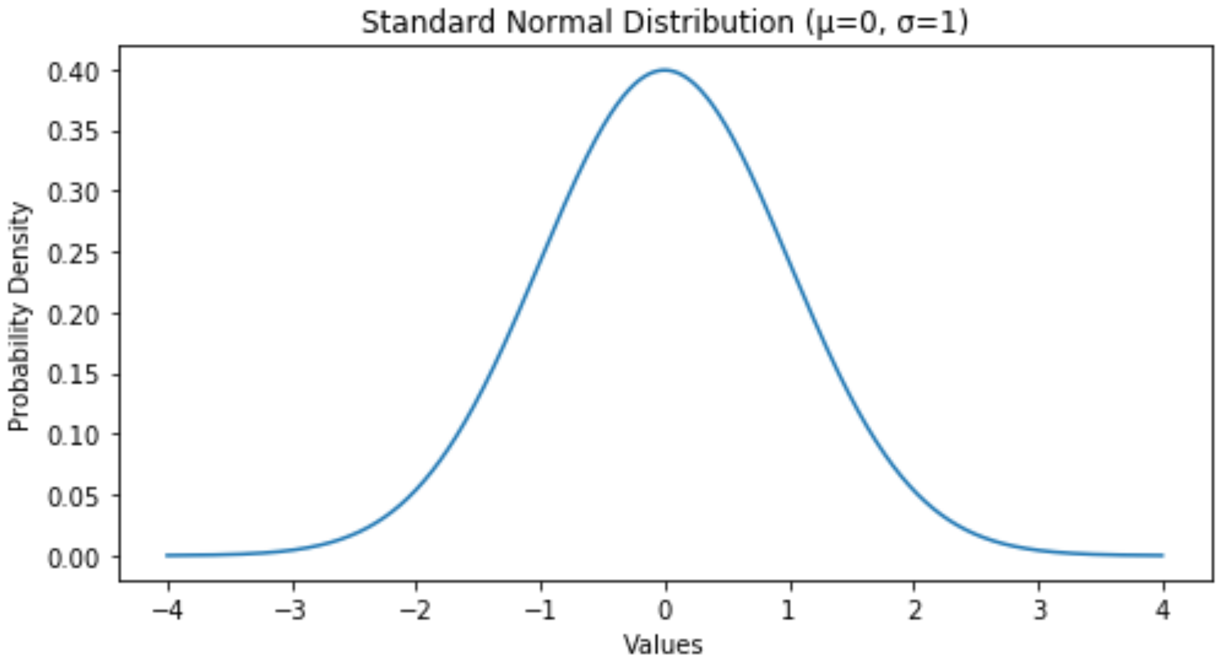


$$P(a \le X \le b) = P(a' \le Z \le b')$$

According to the standardization, the probability of any normal-distributed r.v. can be converted to the probability of a standard normal-distributed.

Standardization





Standardization

Suppose $X \sim N(\mu, \sigma^2)$. Let's define

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

It is clear that $\mathrm{Z} \sim N(0,1)$. Compute

$$P\{X < b\} = P\left\{\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\}$$
$$= P\left\{Z < \boxed{\frac{b-\mu}{\sigma}}\right\}$$
$$= \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Pick another number a, then

$$P\{X < a\} = \Phi\left(\boxed{\frac{a-\mu}{\sigma}}\right)$$

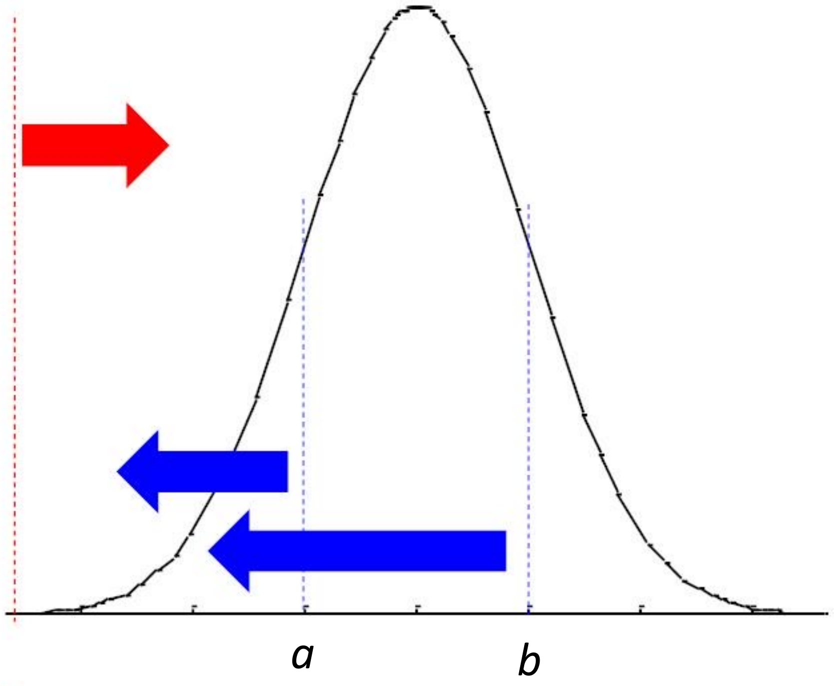
Note that

$$\{X < b\} = \{X \le a\} \cup \{a < X < b\}$$

And $\{X \le a\} \cap \{a < X < b\} = \emptyset$

So

$$\boxed{\begin{aligned} P\{a < X < b\} &= P\{X < b\} - P\{X < a\} \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}}$$



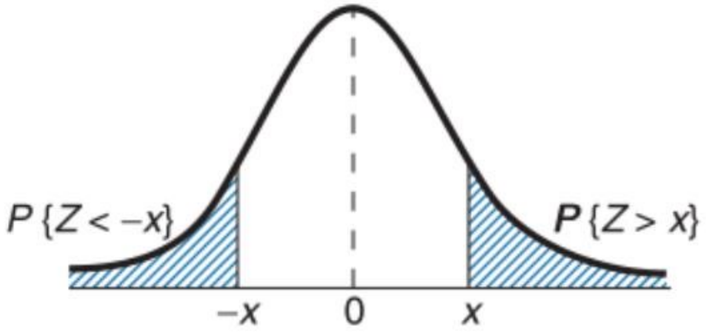
A Symmetric Property of Φ

For any number x,

$$P\{Z > x\} = 1 - P\{Z \leq x\} = 1 - \Phi(x)$$

And

$$P\{Z < -x\} = \Phi(-x)$$



SYMMETRIC PROPERTY

$$\Phi(-x) = 1 - \Phi(x)$$

Table of Standard Normal (z) Curve Areas

- For any number z^*, from -3.49 to 3.49 and rounded to two decimal places, the Z table gives **the area under the z curve and to the left of z^*.**

$$P(z < z^*) = P(z \leq z^*)$$

Where

the letter z is used to represent a random variable whose distribution is the standard normal distribution.

To use the table:

Find the correct row and column (see the following example)

The number at the intersection of that row and column is the probability

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

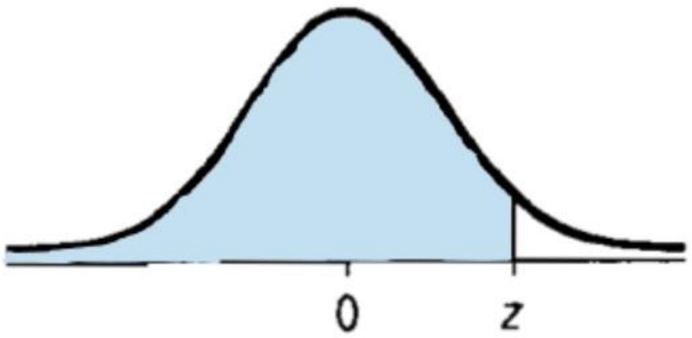
STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Standard Normal Table

Suppose we are interested in the probability that z* is less than 0.12

- Find the row labeled 0.1
- Find the column labeled 0.02
- Find the intersection of the row and column



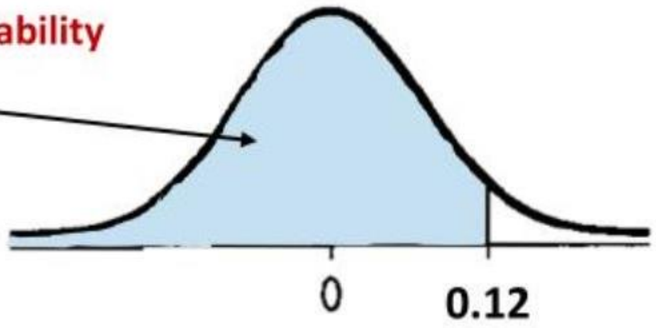
More precisely, we would have a table below

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	**0.5478**	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700

$P(Z \leq 0.12) = 0.5478$

Location of line

probability

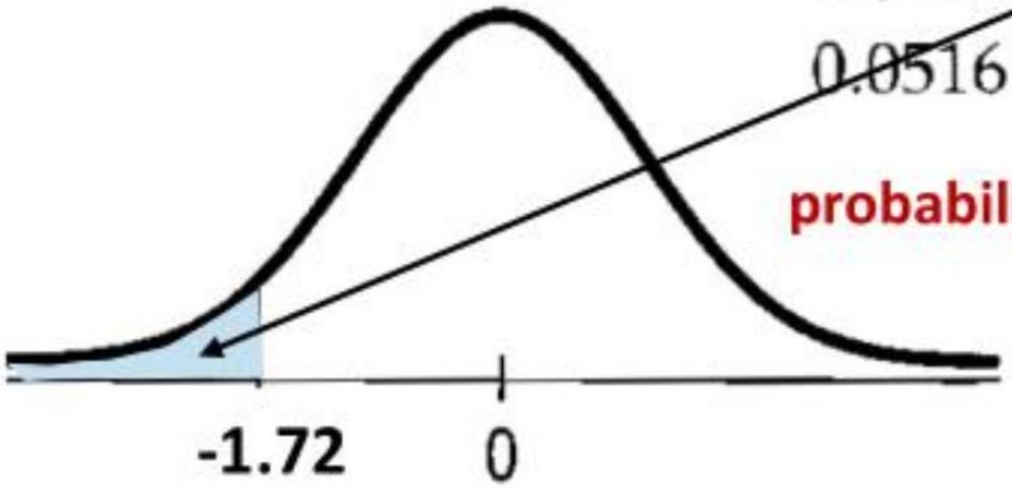


Standard Normal Table

Suppose we are interested in the probability that z* is less than -1.72

Location of line

0.03	0.02	0.01	0.00	z
0.0268	0.0274	0.0281	0.0287	−1.9
0.0336	0.0344	0.0351	0.0359	−1.8
0.0418	**0.0427**	0.0436	0.0446	−1.7
0.0516	0.0526	0.0537	0.0548	−1.6

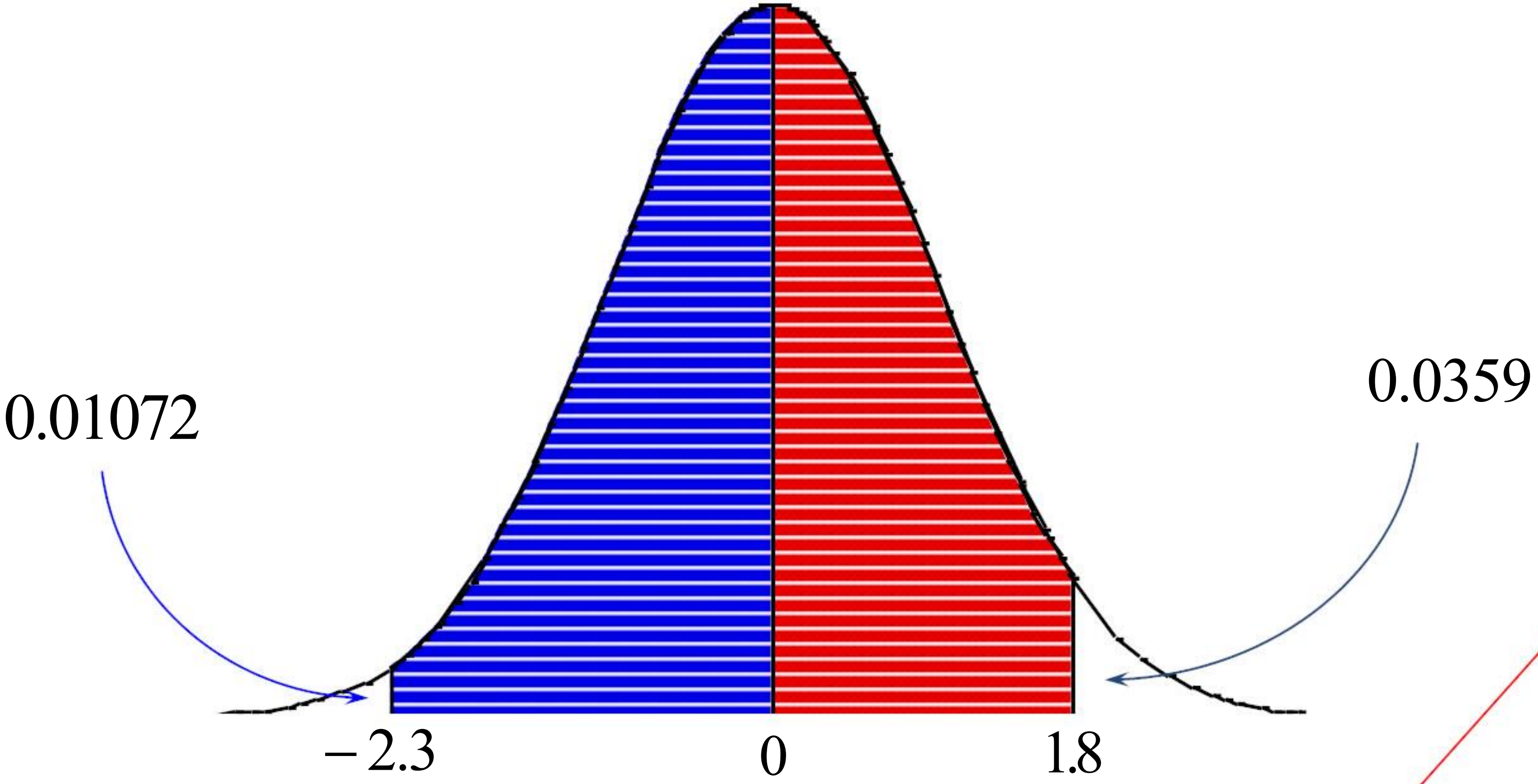


probability

$P(Z \leq -1.72) = 0.0427$

Example

Find the area between $z = -2.3$ and $z = 1.8$.



$$P(-2.3 < z < 1.8) = 0.96407 - 0.01072$$
$$= 0.95338$$

STANDARD NORMAL DISTRIBUTION: Table Values R

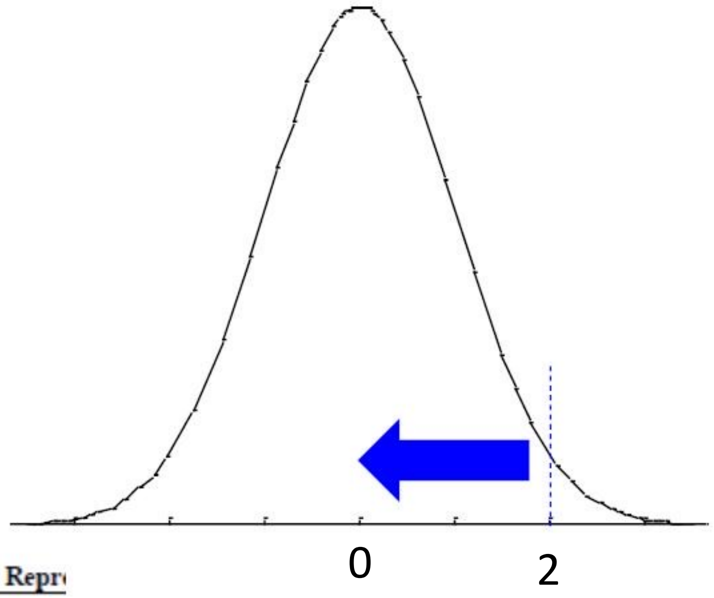
Z	.00	.01	.02	.03	.04
0.0	.50000	.50399	.50798	.51197	.51595
0.1	.53983	.54380	.54776	.55172	.55567
0.2	.57926	.58317	.58706	.59095	.59483
0.3	.61791	.62172	.62552	.62930	.63307
0.4	.65542	.65910	.66276	.66640	.67003
0.5	.69146	.69497	.69847	.70194	.70540
0.6	.72575	.72907	.73237	.73565	.73891
0.7	.75804	.76115	.76424	.76730	.77035
0.8	.78814	.79103	.79389	.79673	.79955
0.9	.81594	.81859	.82121	.82381	.82639
1.0	.84134	.84375	.84614	.84849	.85083
1.1	.86433	.86650	.86864	.87076	.87286
1.2	.88493	.88686	.88877	.89065	.89251
1.3	.90320	.90490	.90658	.90824	.90988
1.4	.91924	.92073	.92220	.92364	.92507
1.5	.93319	.93448	.93574	.93699	.93822
1.6	.94520	.94630	.94738	.94845	.94950
1.7	.95543	.95637	.95728	.95818	.95907
1.8	.96407	.96485	.96562	.96638	.96712
1.9	.97128	.97193	.97257	.97320	.97381

STANDARD NORMAL DISTRIBUTION: Table V

Z	.00	.01	.02	.03
-3.9	.00005	.00005	.00004	.00004
-3.8	.00007	.00007	.00007	.00006
-3.7	.00011	.00010	.00010	.00010
-3.6	.00016	.00015	.00015	.00014
-3.5	.00023	.00022	.00022	.00021
-3.4	.00034	.00032	.00031	.00030
-3.3	.00048	.00047	.00045	.00043
-3.2	.00069	.00066	.00064	.00062
-3.1	.00097	.00094	.00090	.00087
-3.0	.00135	.00131	.00126	.00122
-2.9	.00187	.00181	.00175	.00169
-2.8	.00256	.00248	.00240	.00233
-2.7	.00347	.00336	.00326	.00317
-2.6	.00466	.00453	.00440	.00427
-2.5	.00621	.00604	.00587	.00570
-2.4	.00820	.00798	.00776	.00755
-2.3	.01072	.01044	.01017	.00990
-2.2	.01390	.01355	.01321	.01287

Example

If X is a normal random variable with mean μ = 3 and variance σ^2 = 16, find $P\{X < 11\}$



Solution:

$$P\{X < 11\} = \Phi\left(\boxed{\frac{11-3}{4}}\right) = \Phi(2) \approx .9772$$

Standardization

STANDARD NORMAL DISTRIBUTION: Table Values Repr

Z	.00	.01	.02	.03	.04	
0.0	.50000	.50399	.50798	.51197	.51595	.5
0.1	.53983	.54380	.54776	.55172	.55567	.5
0.2	.57926	.58317	.58706	.59095	.59483	.5
0.3	.61791	.62172	.62552	.62930	.63307	.6
0.4	.65542	.65910	.66276	.66640	.67003	.6
			⋮			
$\boxed{2.0}$	$\boxed{.97725}$	.97778	.97831	.97882	.97932	.9
2.1	.98214	.98257	.98300	.98341	.98382	.9
2.2	.98610	.98645	.98679	.98713	.98745	.9

Example

Let $X \sim N(1, 4)$.
Find $P(X > 1.5)$

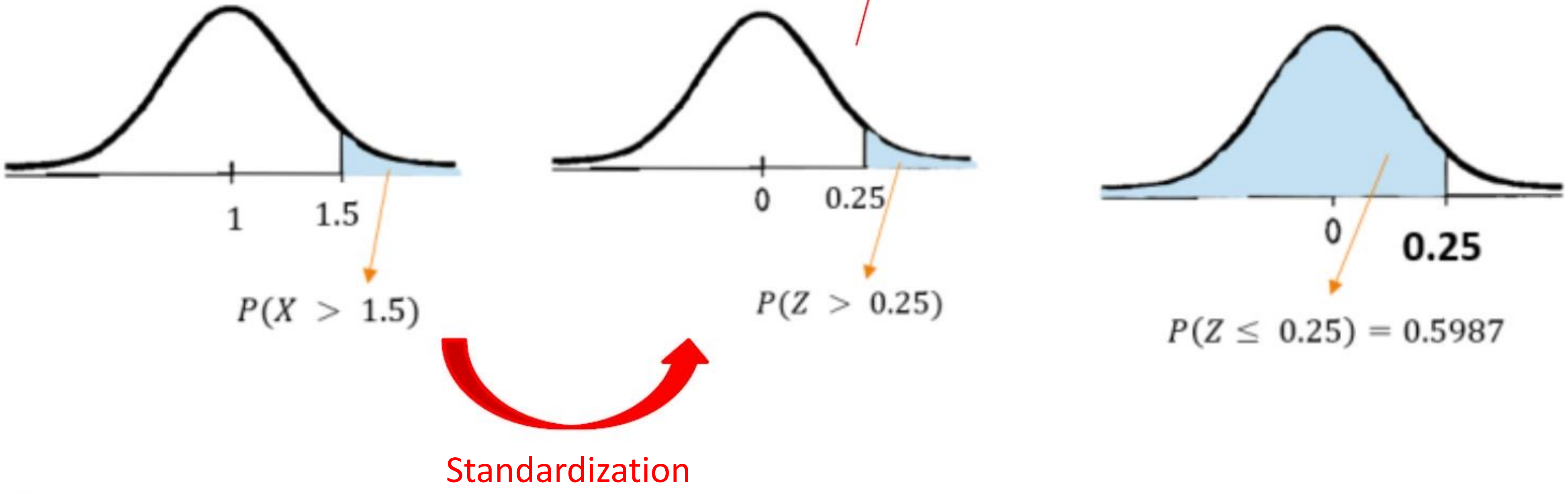
STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
$\boxed{0.2}$	.57926	.58317	.58706	.59095	.59483	$\boxed{.59871}$	.60257	.60642	.61026	.61409

$$P(X > 1.5) = P\left(\frac{X-1}{2} > \frac{1.5-1}{2}\right) = P(Z > 0.25) = 1 - P(Z \leq 0.25)$$

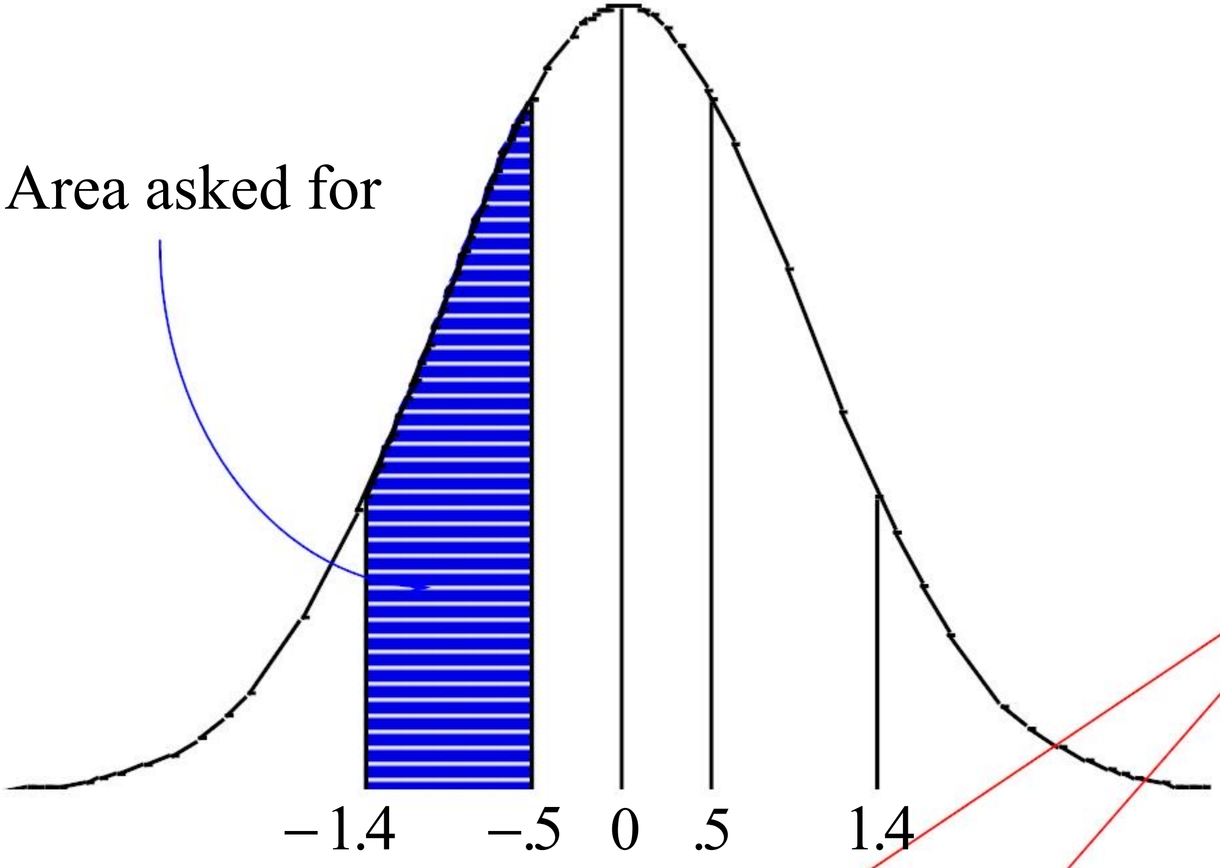
$$= 1 - 0.5987$$

$$= 0.4013$$



Example - Normal Distribution

Find the area between $z = -1.4$ and $z = -0.5$.



STANDARD NORMAL DISTRIBUTION: Table Values R

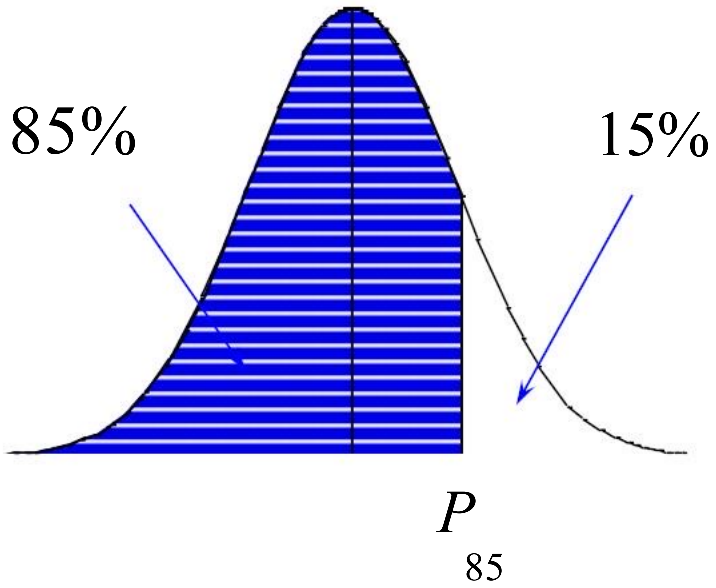
Z	.00	.01	.02	.03	.04
-3.9	.00005	.00005	.00004	.00004	.00004
-3.8	.00007	.00007	.00007	.00006	.00006
-3.7	.00011	.00010	.00010	.00010	.00009
-3.6	.00016	.00015	.00015	.00014	.00014
-3.5	.00023	.00022	.00022	.00021	.00020
⋮					
-1.6	.05480	.05370	.05262	.05155	.05050
-1.5	.06681	.06552	.06426	.06301	.06178
$\boxed{-1.4}$	$\boxed{.08076}$	.07927	.07780	.07636	.07493
⋮					
$\boxed{-0.5}$	$\boxed{.30854}$	.30503	.30153	.29806	.29460
-0.4	.34458	.34090	.33724	.33360	.32997
-0.3	.38209	.37828	.37448	.37070	.36693
-0.2	.42074	.41683	.41294	.40905	.40517
-0.1	.46017	.45620	.45224	.44828	.44433

$$P(-1.4 < z < -.5) = 0.30854 - 0.08076$$
$$= 0.2277$$

Normal Distribution

Note: The normal distribution table may also be used to determine a *z*-score if we are given the area (*to work backwards*).

Example: What is the z-score associated with the 85th percentile?



From the Table.

STANDARD NORMAL DISTRIBUTION: Table Values Re

Z	.00	.01	.02	.03	.04
0.0	.50000	.50399	.50798	.51197	.51595
0.1	.53983	.54380	.54776	.55172	.55567
0.2	.57926	.58317	.58706	.59095	.59483
0.3	.61791	.62172	.62552	.62930	.63307
0.4	.65542	.65910	.66276	.66640	.67003
0.5	.69146	.69497	.69847	.70194	.70540
0.6	.72575	.72907	.73237	.73565	.73891
0.7	.75804	.76115	.76424	.76730	.77035
0.8	.78814	.79103	.79389	.79673	.79955
0.9	.81594	.81859	.82121	.82381	.82639
$\boxed{1.0}$	.84134	.84375	.84614	$\boxed{.84849}$	$\boxed{.85083}$

0.85 → ?

Linear Interpolation – What and Why?

- Linear interpolation is obtained by passing a straight line between 2 data points.

In Tabular Form:

x_0	$f(x_0)$
x	$g(x)$
x_1	$f(x_1)$

If $g(x)$ is a linear function then

$$g(x) = Ax + B$$

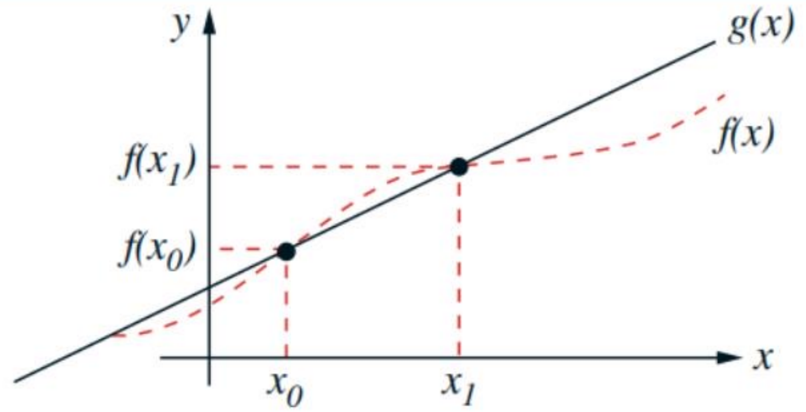
where A and B are unknown coefficients

To pass through points $(x_0, f(x_0))$ and $(x_1, f(x_1))$, we must have:

$$Ax_0 + B = f(x_0) \rightarrow B = f(x_0) - Ax_0 \quad (1)$$

$Ax_1 + B = f(x_1)$ (2), Put (1) into (2), $Ax_1 + f(x_0) - Ax_0 = f(x_1)$

$$\rightarrow A = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \qquad B = \frac{f(x_0)x_1 - f(x_1)x_0}{x_1 - x_0}$$



f(x) = the exact function for which values are known only at a discrete set of data points

g(x) = the interpolated approximation to *f(x)*

x_0, x_1 = the data points (also referred to as interpolation points or nodes)

Substituting A and B into equation $g(x) = Ax + B$

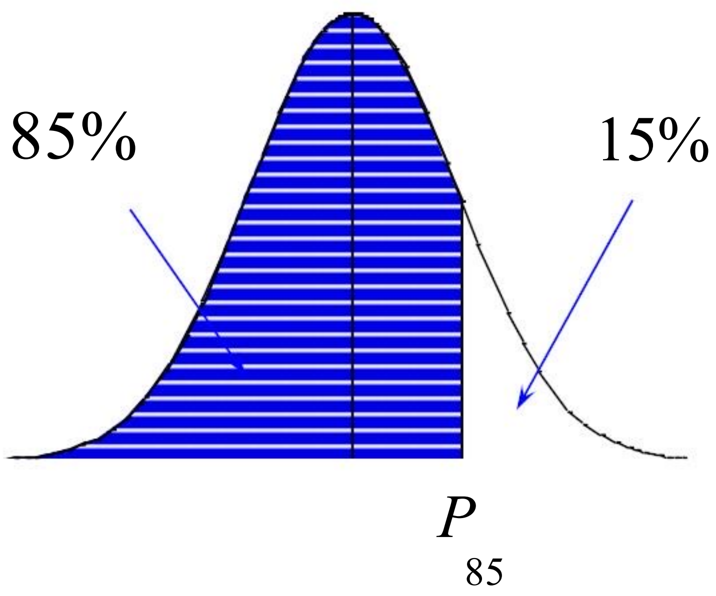
$$g(x) = f(x_0)\frac{(x_1 - x)}{(x_1 - x_0)} + f(x_1)\frac{(x - x_0)}{(x_1 - x_0)}$$

This is the formula for linear interpolation

Normal Distribution

Note: The normal distribution table may also be used to determine a *z*-score if we are given the area (*to work backwards*).

Example: What is the z-score associated with the 85th percentile?



Linear interpolation

$$g(x) = f(x_0)\frac{(x_1 - x)}{(x_1 - x_0)} + f(x_1)\frac{(x - x_0)}{(x_1 - x_0)}$$

$$0.85 = 0.84849\frac{(1.04 - x)}{(1.04 - 1.03)} + 0.85083\frac{(x - 1.03)}{(1.04 - 1.03)}$$

$$x = 1.0365$$

From the Table.

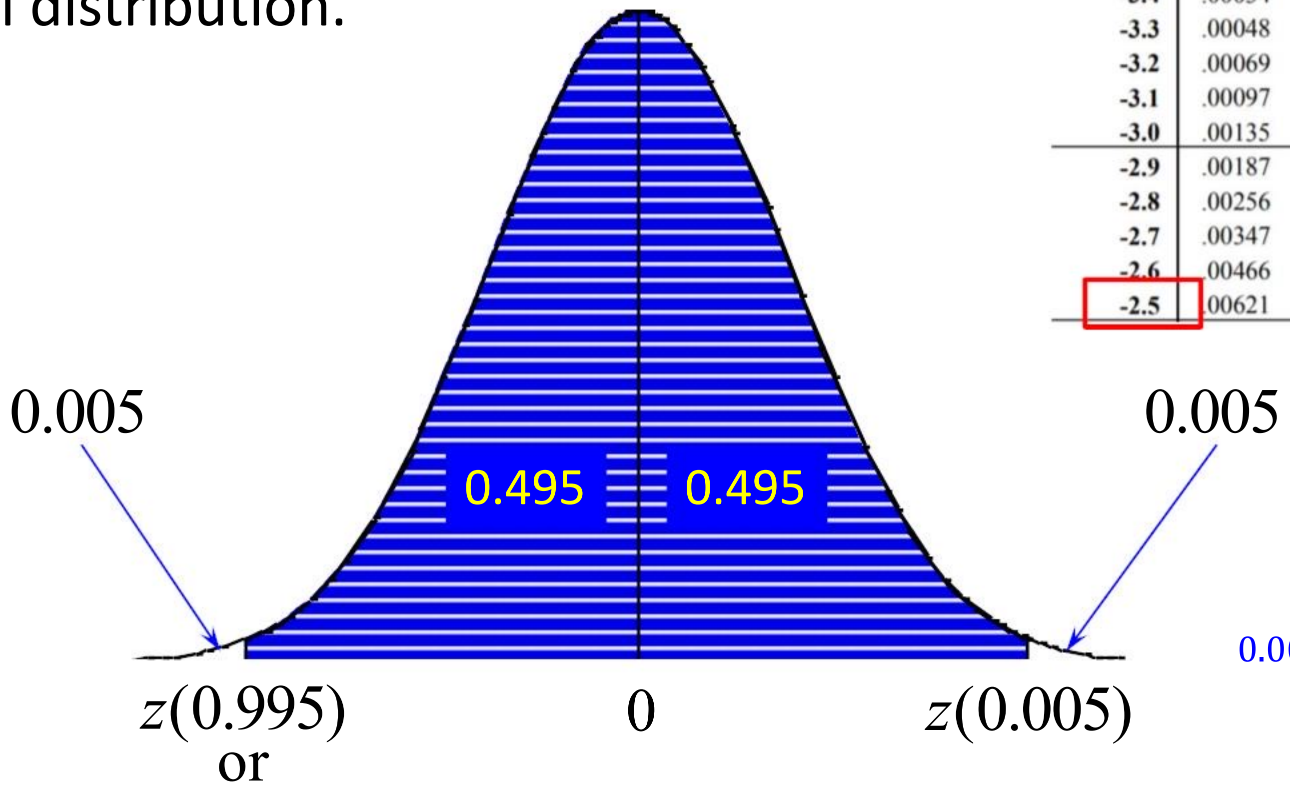
STANDARD NORMAL DISTRIBUTION: Table Values Re

Z	.00	.01	.02	.03	.04
0.0	.50000	.50399	.50798	.51197	.51595
0.1	.53983	.54380	.54776	.55172	.55567
0.2	.57926	.58317	.58706	.59095	.59483
0.3	.61791	.62172	.62552	.62930	.63307
0.4	.65542	.65910	.66276	.66640	.67003
0.5	.69146	.69497	.69847	.70194	.70540
0.6	.72575	.72907	.73237	.73565	.73891
0.7	.75804	.76115	.76424	.76730	.77035
0.8	.78814	.79103	.79389	.79673	.79955
0.9	.81594	.81859	.82121	.82381	.82639
1.0	.84134	.84375	.84614	.84849	.85083

0.85 → ?

Example

Find the *z*-scores that bound the middle 0.99 of the normal distribution.



STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z sc

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494

0.005 → ?

Linear interpolation

$$g(x) = f(x_0)\frac{(x_1 - x)}{(x_1 - x_0)} + f(x_1)\frac{(x - x_0)}{(x_1 - x_0)}$$

$$0.005 = 0.00508\frac{(-2.58 - x)}{(-2.58 + 2.57)} + 0.00494\frac{(x + 2.57)}{(-2.58 + 2.57)}$$

$$x = -2.576$$

Use the Table: $z(0.005) = 2.576$ and $z(0.995) = -z(0.005) = -2.576$

CityU

How to use python to calculate the probability of Normal distribution

```python
import scipy.stats as stats

# Define the mean and standard deviation of the normal distribution
mean = 20
std_dev = 5

# Calculate the probability of a value being less than or equal to a specific value
x = 25
probability = stats.norm.cdf(x, loc=mean, scale=std_dev)

print("Probability:", probability)
```

```
Probability: 0.8413447460685429
```

How to use python to calculate the probability of Normal distribution

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409

Let $X \sim N(1, 4)$.
Find $P(X > 1.5)$

$$P(X > 1.5) = P\left(\frac{X-1}{2} > \frac{1.5-1}{2}\right) = P(Z > 0.25) = 1 - P(Z \leq 0.25)$$

$$= 1 - 0.5987$$

$$= 0.4013$$

```python
import scipy.stats as stats

# Define the mean and standard deviation of the normal distribution
mean = 1
std_dev = 2

# Calculate the probability of a value being less than or equal to a specific value
x = 1.5
probability =stats.norm.cdf(x, loc=mean, scale=std_dev)

print("Probability:", 1-probability)
```

```
Probability: 0.4012936743170763
```

How to use python to calculate the probability of Normal distribution

Methods

rvs(loc=0, scale=1, size=1, random_state=None)	Random variates.
pdf(x, loc=0, scale=1)	Probability density function.
logpdf(x, loc=0, scale=1)	Log of the probability density function.
cdf(x, loc=0, scale=1)	Cumulative distribution function.
logcdf(x, loc=0, scale=1)	Log of the cumulative distribution function.
sf(x, loc=0, scale=1)	Survival function (also defined as `1 - cdf`, but *sf* is sometimes more accurate).
logsf(x, loc=0, scale=1)	Log of the survival function.
ppf(q, loc=0, scale=1)	Percent point function (inverse of `cdf` — percentiles).
isf(q, loc=0, scale=1)	Inverse survival function (inverse of `sf`).
moment(order, loc=0, scale=1)	Non-central moment of the specified order.
stats(loc=0, scale=1, moments='mv')	Mean('m'), variance('v'), skew('s'), and/or kurtosis('k').

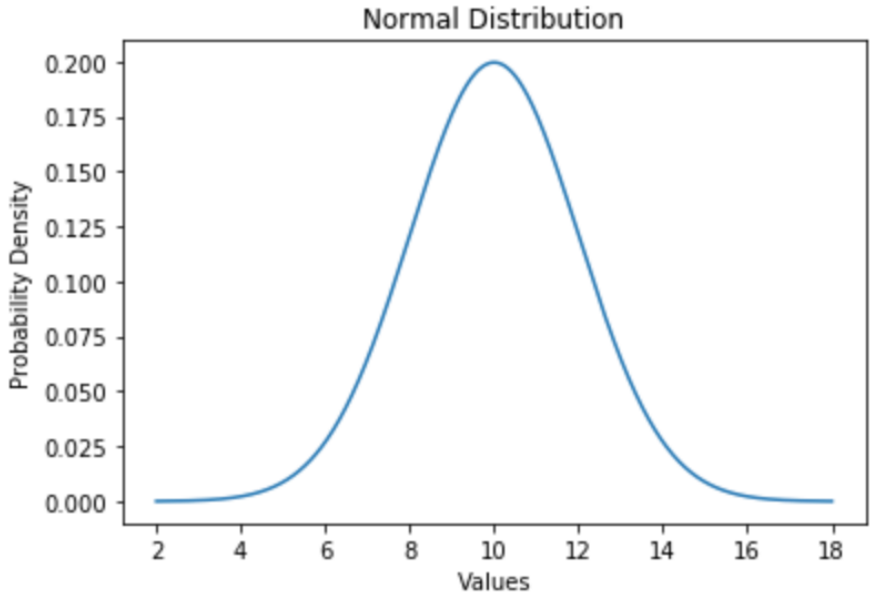
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Define the mean and standard deviation of the normal distribution
mean = 10
std_dev = 2

# Generate a range of values
x = np.linspace(mean - 4*std_dev, mean + 4*std_dev, 1000)

# Calculate the probability density function (PDF) of the normal distribution
pdf = norm.pdf(x, loc=mean, scale=std_dev)

# Plot the normal distribution
plt.plot(x, pdf)
plt.xlabel('Values')
plt.ylabel('Probability Density')
plt.title('Normal Distribution')
plt.show()
```



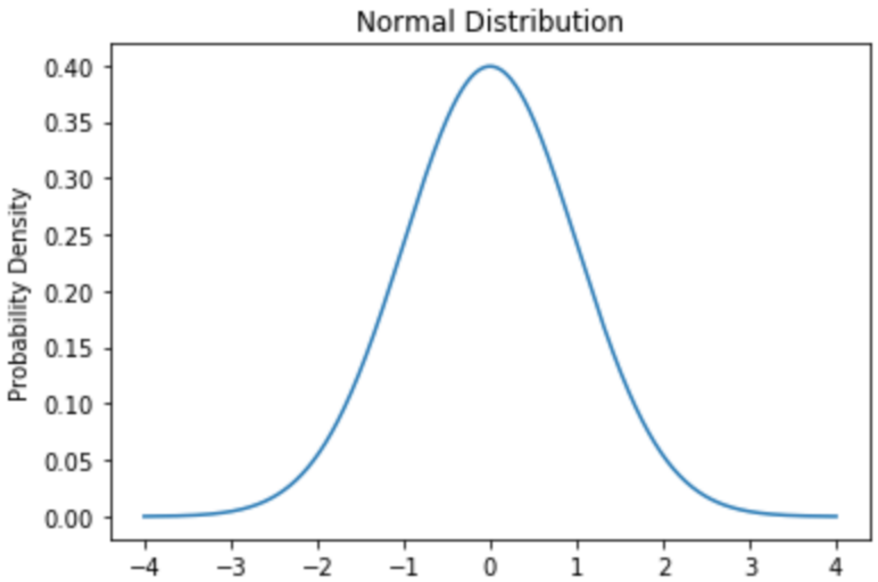
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Define the mean and standard deviation of the normal distribution
mean = 10
std_dev = 2

# Generate a range of values
x = np.linspace(mean - 4*std_dev, mean + 4*std_dev, 1000)
x_standard = np.linspace(-4, 4, 1000)

# Calculate the probability density function (PDF) of the normal distribution
pdf = norm.pdf(x_standard, loc=0, scale=1)

# Plot the normal distribution
plt.plot(x_standard, pdf)
plt.xlabel('Values')
plt.ylabel('Probability Density')
plt.title('Normal Distribution')
plt.show()
```



Normal Distribution

Suppose X_1, X_2,…, X_n are independent, and

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

What will $X = \sum_I X_I$ be?

X will still be normal,

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Where

$$\mu = \sum_i \mu_i, \quad \text{and} \quad \sigma^2 = \sum_i \sigma_i^2$$

Example: if $X \sim \mathcal{N}(2, 4)$ and$Y \sim \mathcal{N}(4, 16)$, and they are independent, then

$$X + Y \sim \mathcal{N}(6, 20)$$

Example

Data from the national oceanic and atmospheric administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3 inches. Assume it is independent of other years.

Find the probability that next year's precipitation X_1 will exceed that of the following year X_2 by more than 3 inches.

Solution: we need to find P{X_1-X_2 > 3}.
Note that -X_2 is still normal with mean -12.08 and variance $(-1)^2 3^2$. So

The X1-X2 is still a normal distribution. Mean= mean1 - mean2 = 0; Variance = Variance1 + Variance 2 = 18

Example

Data from the national oceanic and atmospheric administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3 inches. Assume it is independent of other years.

Find the probability that next year's precipitation X_1 will exceed that of the following year X_2 by more than 3 inches.

Solution: we need to find P{X_1-X_2 > 3}.

Note that -X_2 is still normal with mean -12.08 and variance $(-1)^2 3^2$. So

The X1-X2 is still a normal distribution. Mean= mean1 - mean2 = 0; Variance = Variance1 + Variance 2 = 18

$$P\{X_1 - X_2 > 3\} = P\{Z_1 - Z_2 > \frac{3-0}{\sqrt{18}}\} = P\{Z_1 - Z_2 > \frac{1}{\sqrt{2}}\} = 1 - P\{Z_1 - Z_2 < \frac{1}{\sqrt{2}}\} = 1 - 0.7602 = 0.2398$$

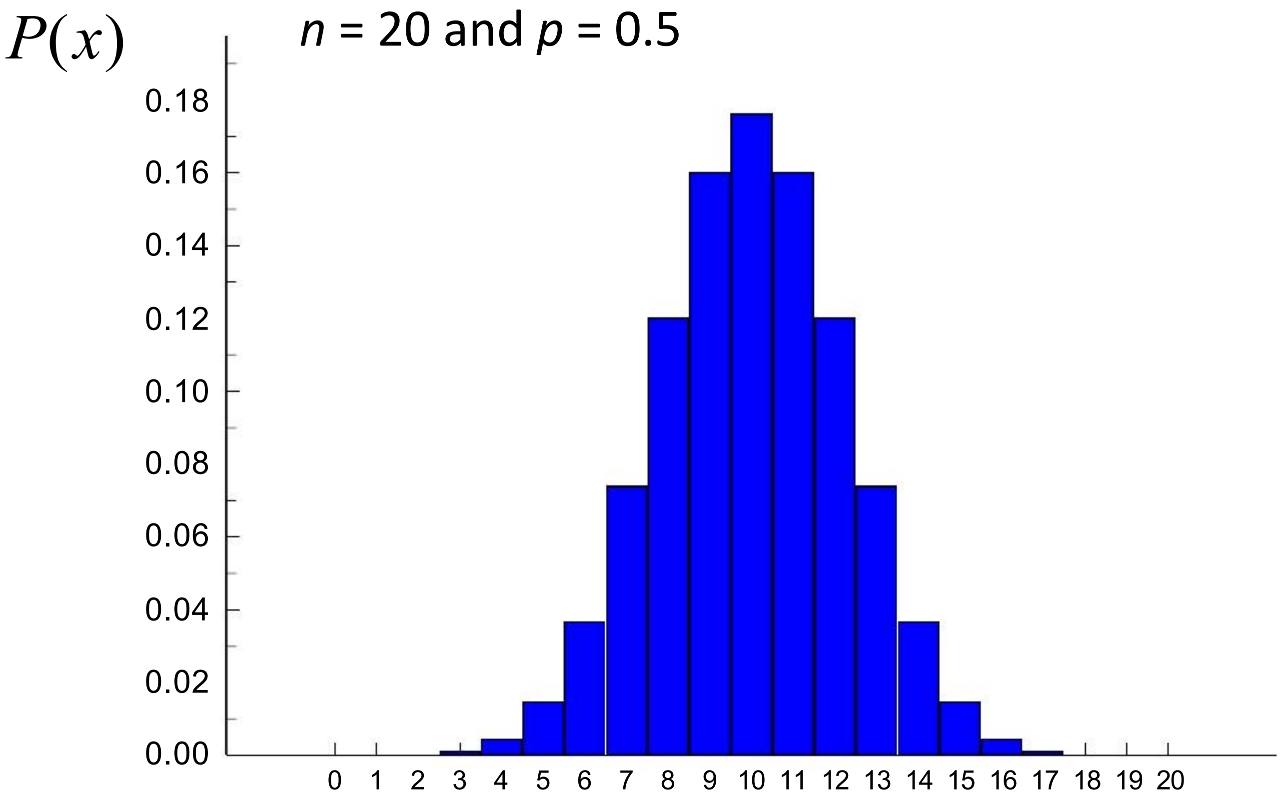
Normal Approximation of the Binomial

- Recall: the binomial distribution is a probability distribution of the discrete random variable x, the number of successes observed in n repeated independent trials.

- Binomial probabilities can be reasonably estimated by using the normal probability distribution.

Normal Approximation of the Binomial

Background: Consider the distribution of the binomial variable *x* when



The histogram may be approximated by a *normal* curve.

Normal Approximation of the Binomial

Note:

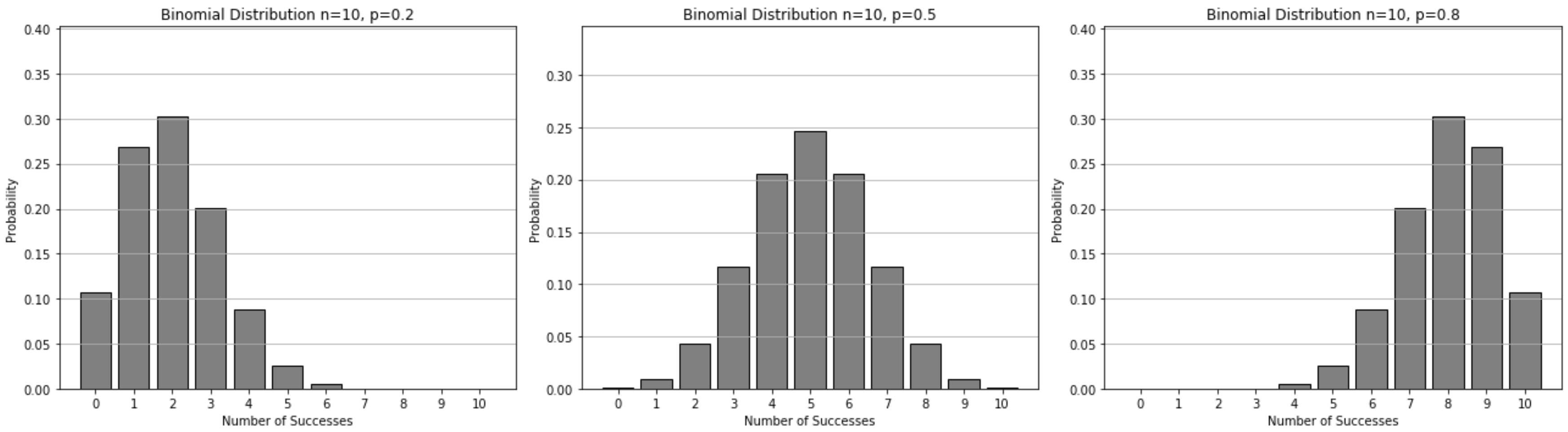
1. The normal curve has mean and standard deviation from the binomial distribution.

$$\boxed{\begin{aligned}\mu &= np = (20)(0.5) = 10 \\ \sigma &= \sqrt{npq} = \sqrt{(20)(0.5)(0.5)} = \sqrt{5} \approx 2.236\end{aligned}}$$

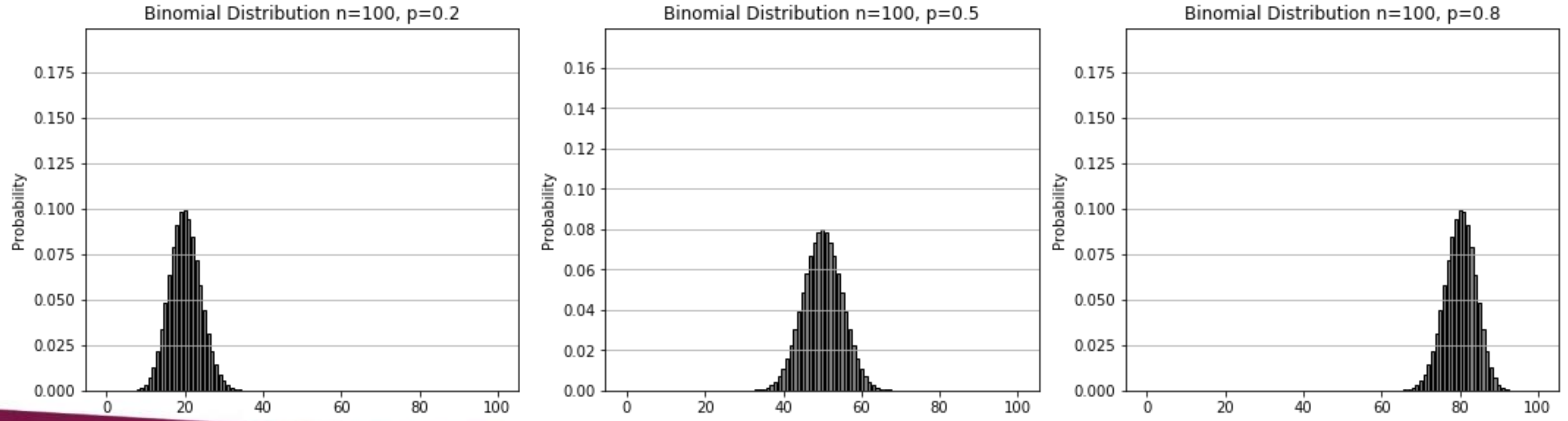
2. Can approximate the area of the rectangles with the area under the normal curve.
3. The approximation becomes more accurate as *n* becomes larger.

Normal Approximation of the Binomial

There are 10 (small number of trials) independent trials



There are 100 (large number of trials) independent trials



Normal Approximation of the Binomial

Two Problems:

1. As p moves away from 0.5, the binomial distribution is less symmetric, less normal-looking.

Solution: The normal distribution provides a reasonable approximation to a binomial probability distribution whenever the values of np and $n(1 - p)$ both exceed 5.

Normal Approximation of the Binomial

Two Problems:

1. As p moves away from 0.5, the binomial distribution is less symmetric, less normal-looking.

Solution: The normal distribution provides a reasonable approximation to a binomial probability distribution whenever the values of np and $n(1 – p)$ both exceed 5.

2. The binomial distribution is *discrete*, and the normal distribution is *continuous*.

Solution: Use the **continuity correction factor**. Add or subtract 0.5 to account for the width of each rectangle.

If $P(X=n)$ use $P(n - 0.5 < X < n + 0.5)$
If $P(X>n)$ use $P(X > n + 0.5)$
If $P(X \leq n)$ use $P(X < n + 0.5)$
If $P(X<n)$ use $P(X < n - 0.5)$
If $P(X \geq n)$ use $P(X > n - 0.5)$

Example

Research indicates 40% of all students entering a certain university withdraw from a course during their first year. What is the probability that fewer than 650 of this year's entering class of 1800 will withdraw from a class?

Example

Research indicates 40% of all students entering a certain university withdraw from a course during their first year. What is the probability that fewer than 650 of this year's entering class of 1800 will withdraw from a class?

If	P(X=n) use	P(n − 0.5 < X < n + 0.5)
If	P(X>n) use	P(X > n + 0.5)
If	P(X≤n) use	P(X < n + 0.5)
If	P (X<n) use	P(X < n − 0.5)
If	P(X ≥ n) use	P(X > n − 0.5)

Let *x* be the number of students that withdraw from a course during their first year.
x has a binomial distribution: *n* = 1800, *p* = 0.4
The probability function is given by:

$$P(x) = \binom{1800}{x}(0.4)^x(0.6)^{1800-x} \quad \text{for} \quad x = 0, 1, 2, \ldots, 1800$$

Solution:
Since $\boxed{np = (1800)(0.4) = 720 > 5 \text{ and } nq = (1800)(0.6) = 1080 > 5,}$ Checking
we use the normal approximation method.

$$\mu = np = (1800)(0.4) = 720$$

$$\sigma = \sqrt{npq} = \sqrt{(1800)(0.4)(0.6)} = \sqrt{432} \approx 20.78$$

$$P(x \text{ is fewer than } 650) = P(x < 650) \quad (\text{for discrete variable } x)$$

Correction

$$\boxed{= P(x \le 649.5)} \quad (\text{for a continuous variable } x)$$

$$= P\left(\frac{x-720}{20.78} \le \frac{649.5-720}{20.78}\right)$$

$$= P(z \le -3.39)$$

Check the table

$$= 0.00035$$

Introduction

We will introduce some distributions arising from the Normal Distribution (**Chi-square distribution, t-distribution, F distribution**), which are widely used in statistics for hypothesis testing and constructing confidence intervals.

- How they are defined.
- Basic property
- Graphic view

The Chi-Square Distribution

If $Z_1, Z_2, \ldots, Z_n$ are *independent* standard normal random variables, then

$$\boxed{X = Z_1^2 + Z_2^2 + \cdots + Z_n^2}$$

Is said to have *chi-square distribution with n degrees of freedom*.

Notation

$$\boxed{X \sim \chi_n^2}$$

Additive property: if

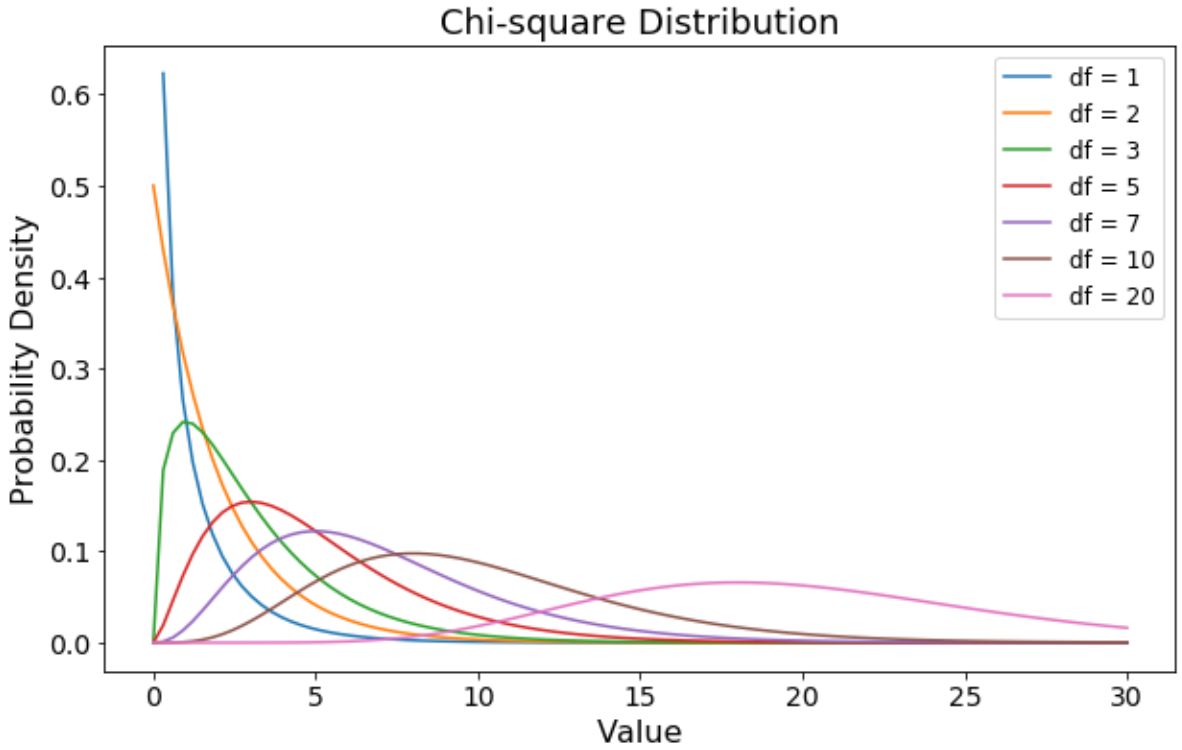
$$X \sim \chi_n^2, \quad Y \sim \chi_k^2$$

then

$$X + Y \sim \chi_{n+k}^2$$

The Chi-Square Distribution

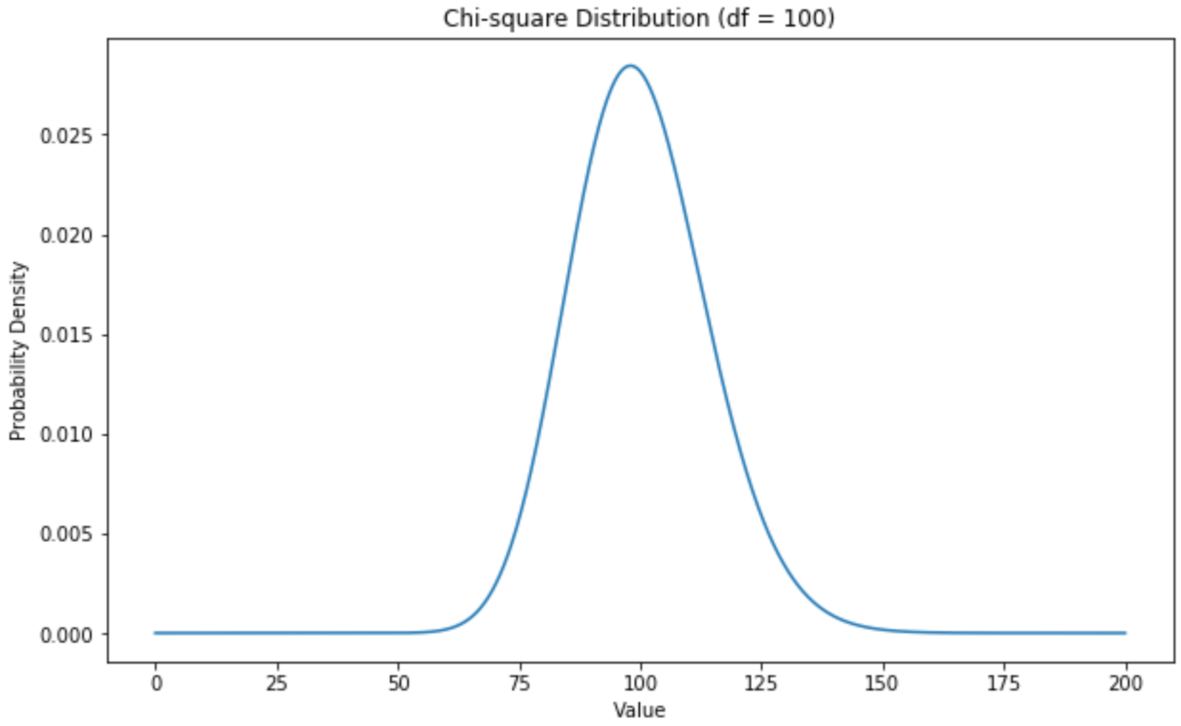
The chi-square distribution is characterized by a single parameter called the degrees of freedom (df). The degrees of freedom determine the shape of the distribution.



The chi-square distribution is positively skewed, and its shape becomes more symmetrical as the degrees of freedom increase.

The Chi-Square Distribution

The chi-square distribution is characterized by a single parameter called the degrees of freedom (df). The degrees of freedom determine the shape of the distribution.



As the degrees of freedom increase, the chi-square distribution becomes increasingly bell-shaped and approaches a normal distribution.

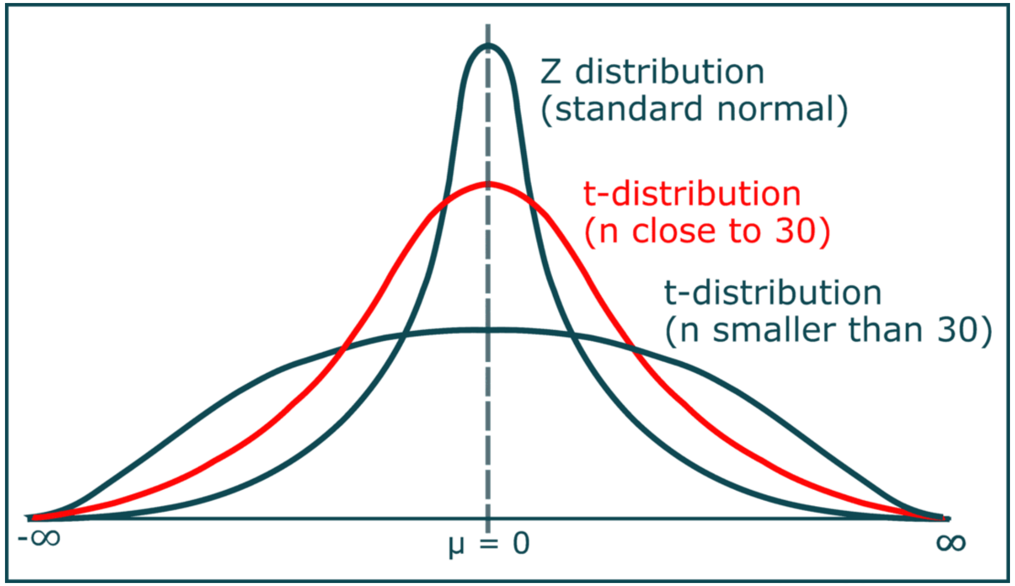
The t-Distribution

If $Z \sim N(0,1)$ and $Y \sim \chi_n^2$, then the random variable

$$\boxed{T_n = \frac{Z}{\sqrt{Y/n}}}$$

It is used in situations where the sample size is relatively small, and the population standard deviation is unknown.

Is said to have a *t-Distribution with n degrees of freedom.*



The F-Distribution

If

$$X \sim \chi^2_n, \quad Y \sim \chi^2_m,$$

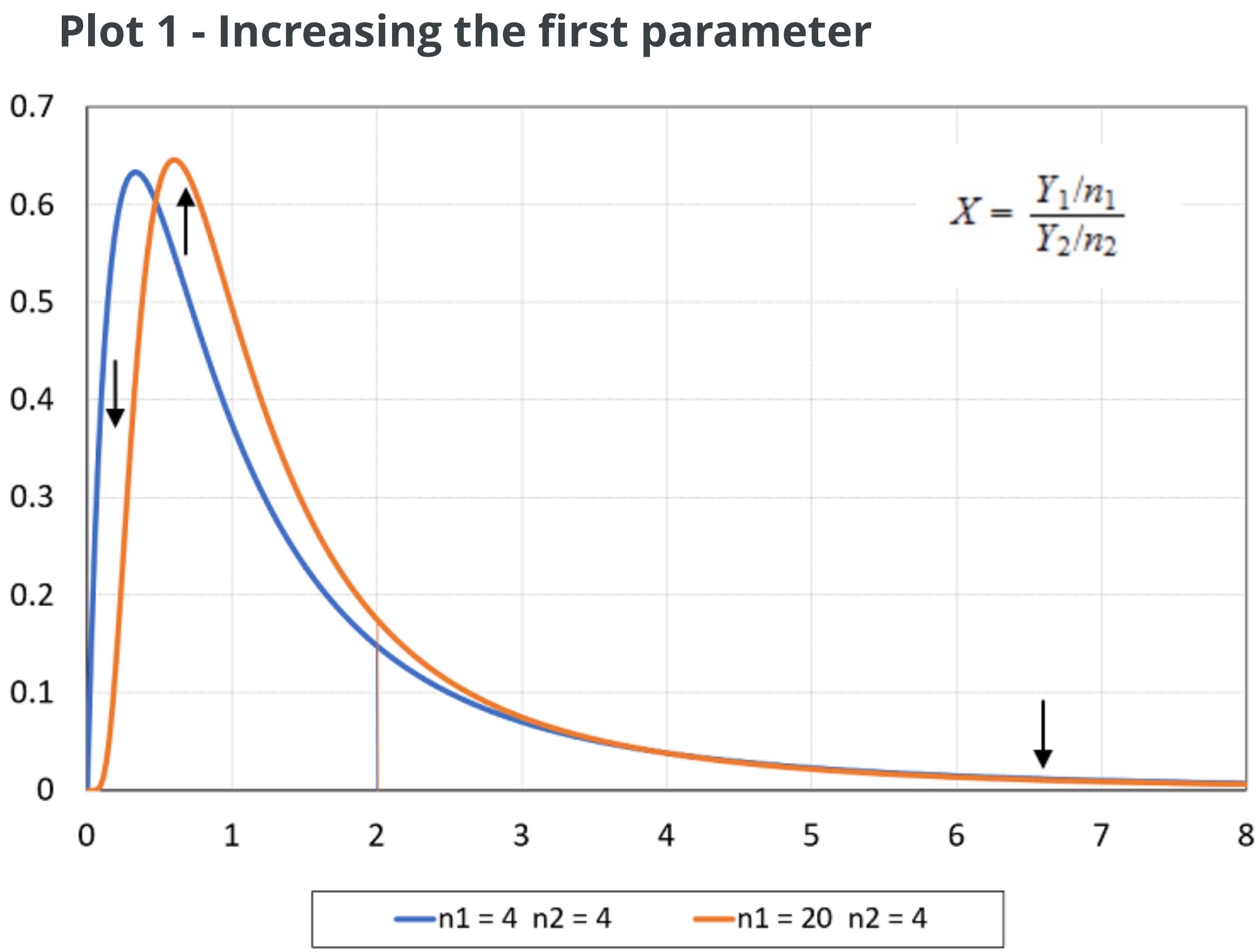
Then the random variable

$$\boxed{F_{n,m} = \frac{X/n}{Y/m}}$$

Is said to have *F-Distribution* *with* *n and m degrees of freedom*

The degrees of freedom determine the shape of the distribution.

The F-Distribution

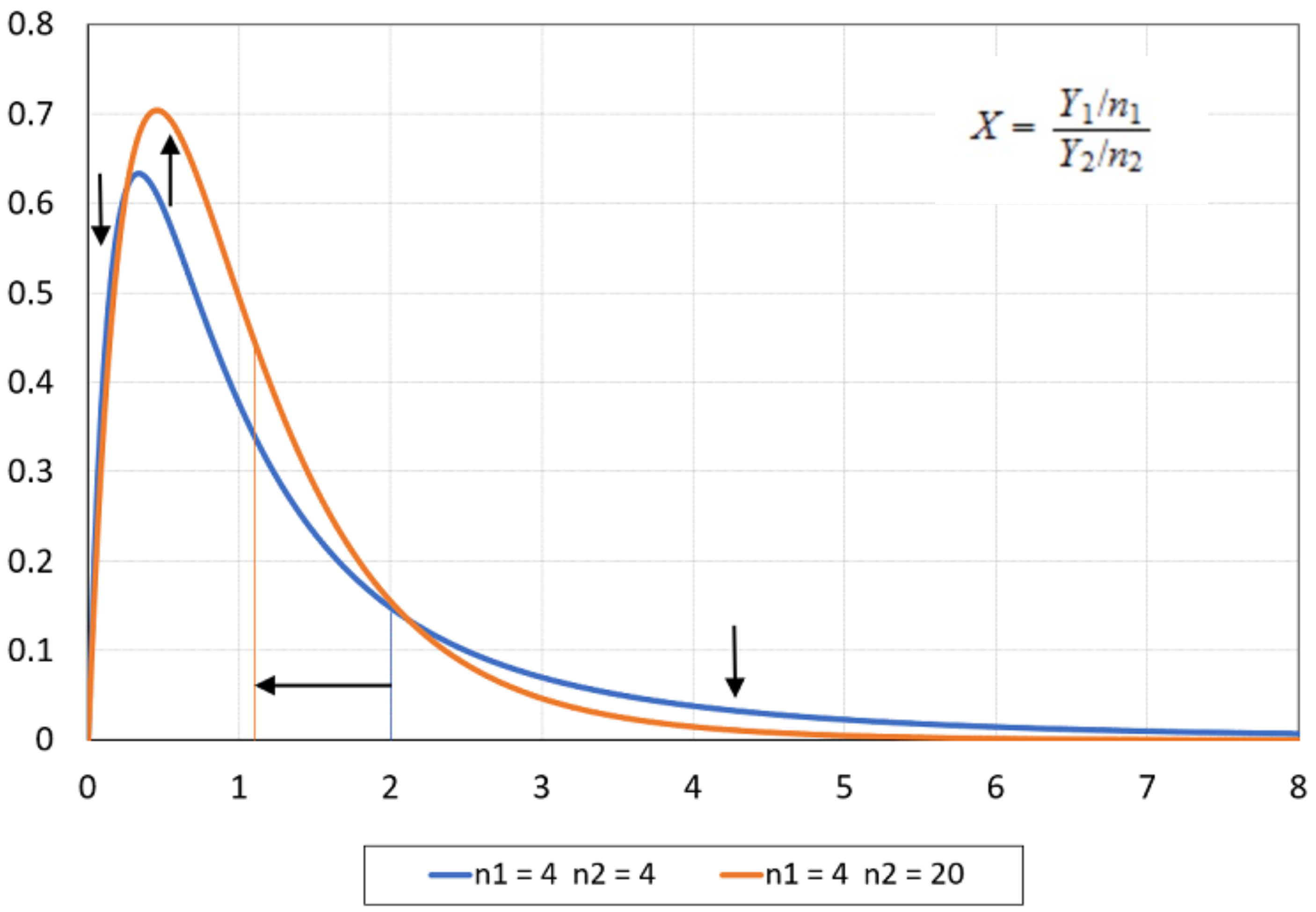


By increasing the first parameter, the mean of the distribution (vertical line) does not change.

However, part of the density is shifted from the tails to the center of the distribution.

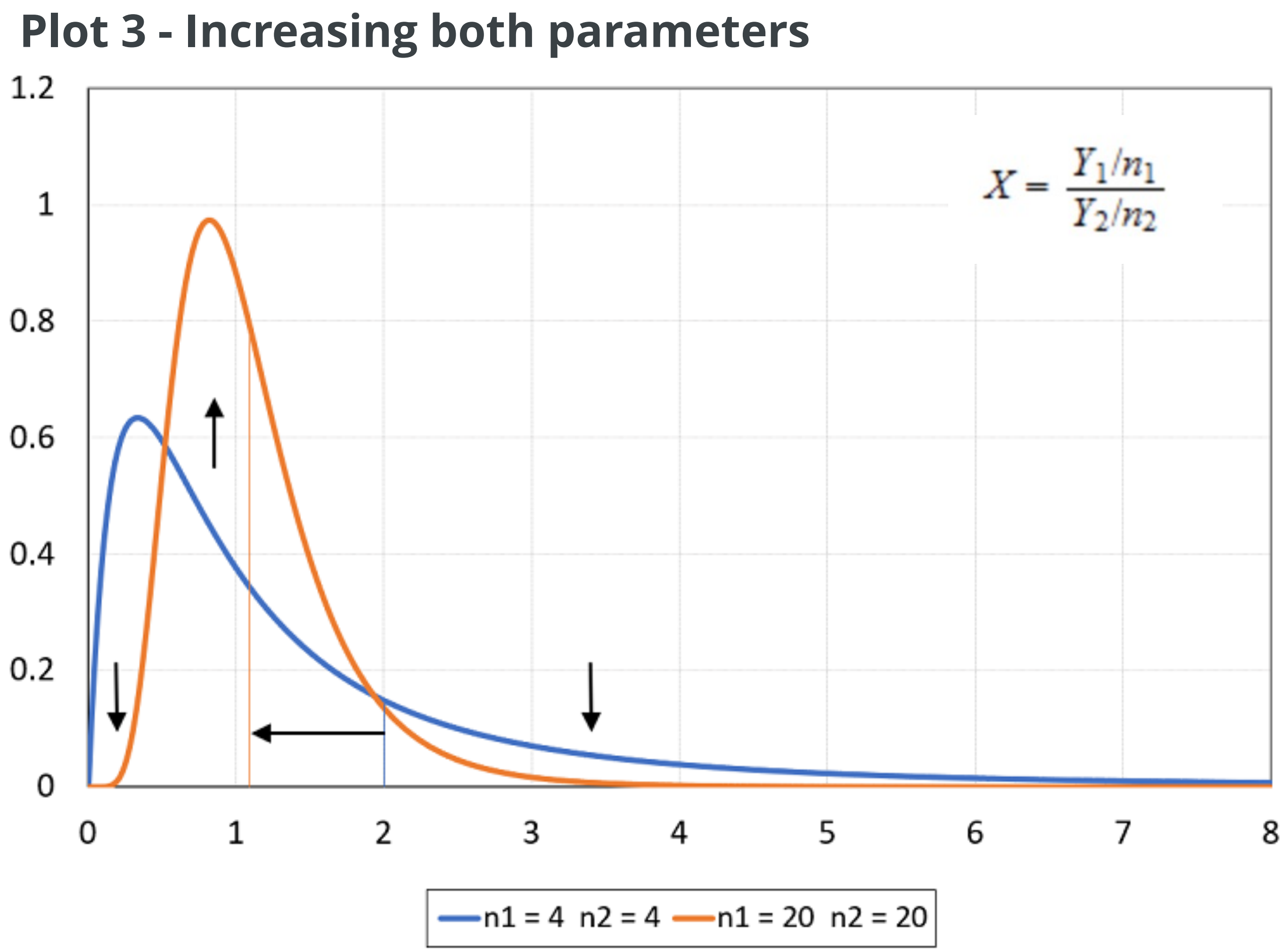
The F-Distribution

Plot 2 - Increasing the second parameter



- The mean of the distribution (vertical line) decreases
- Some density is shifted from the tails (mostly from the right tail) to the center of the distribution.

The F-Distribution



- The mean of the distribution decreases.
- The density is shifted from the tails to the center of the distribution.
- As a result, the distribution has a bell shape similar to the shape of the normal distribution.