

Seminar 7: Continuous Random Variables. Normal Distribution

Carlos Buitrago & Anastasiya Chikalova

Table of contents

1	Probability Distributions	2
1.1	Cumulative Distribution Function	2
1.1.1	Properties of the CDF	2
1.2	Discrete Probability Distributions	3
1.3	Example 1: Bernoulli Distribution	3
2	Continuous Probability Distributions	4
2.1	Properties of the Probability Density Function	5
2.2	Examples of Absolutely Continuous Distributions	5
2.2.1	Uniform Distribution	5
2.2.2	Exponential Distribution	6
2.2.3	Normal Distribution	6
2.3	The Standard Normal Table	8
2.3.1	How is the Z-table organized?	8
2.4	Why is the normal distribution so important?	10
3	Simulating Random Variables from Uniform Random Numbers	11
3.1	The Inverse Transform Method	11
3.2	Example: Exponential Distribution	12

A discrete random variable assigns positive probability to countably many values.

A **continuous random variable** distributes its probability over intervals rather than individual points. In this lecture we study the most important class of continuous distributions: **absolutely continuous distributions**.

1 Probability Distributions

A **probability distribution** on $\mathbb{R}$ is a probability measure

$$P : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1],$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ -algebra.

1.1 Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable ξ is defined by

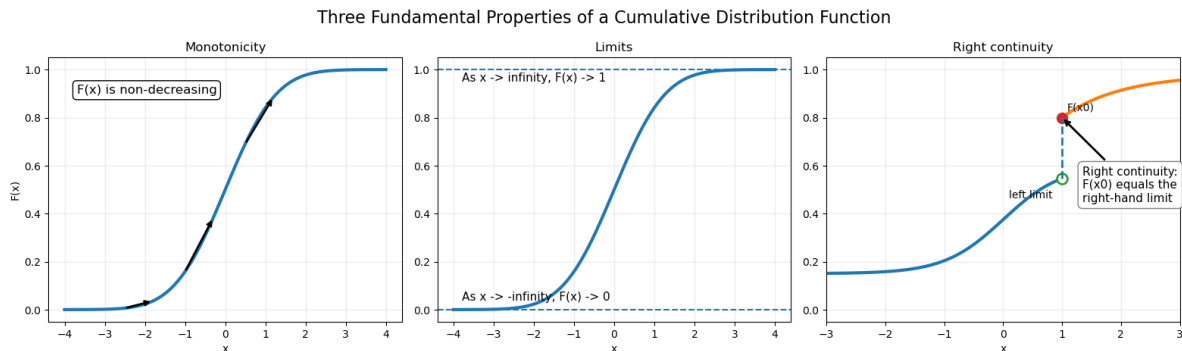
$$F_{\xi}(x) = P(\xi \leq x), \quad x \in \mathbb{R}.$$

Notice that the CDF completely determines the probability distribution.

In other words, if two random variables have the same cumulative distribution function, then they have the same distribution.

1.1.1 Properties of the CDF

1. $F(x)$ is non-decreasing.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.
3. $F(x)$ is right-continuous.



There are three fundamentally different types of probability distributions:

1. **Discrete distributions;**
2. **Singular distributions;**
3. **Absolutely continuous distributions.**

Moreover, every probability distribution can be uniquely decomposed into a mixture of these three types.

A **singular distribution** is concentrated on an uncountable set of Lebesgue measure zero, while assigning probability zero to every individual point.

The most famous example is the **Cantor distribution**, whose cumulative distribution function is known as the *Devil's staircase*.

In this course we will only study **discrete** and **absolutely continuous** distributions.

1.2 Discrete Probability Distributions

A probability distribution is called **discrete** if there exists a finite or countable set

$$X = \{x_1, x_2, \dots\} \subset \mathbb{R}$$

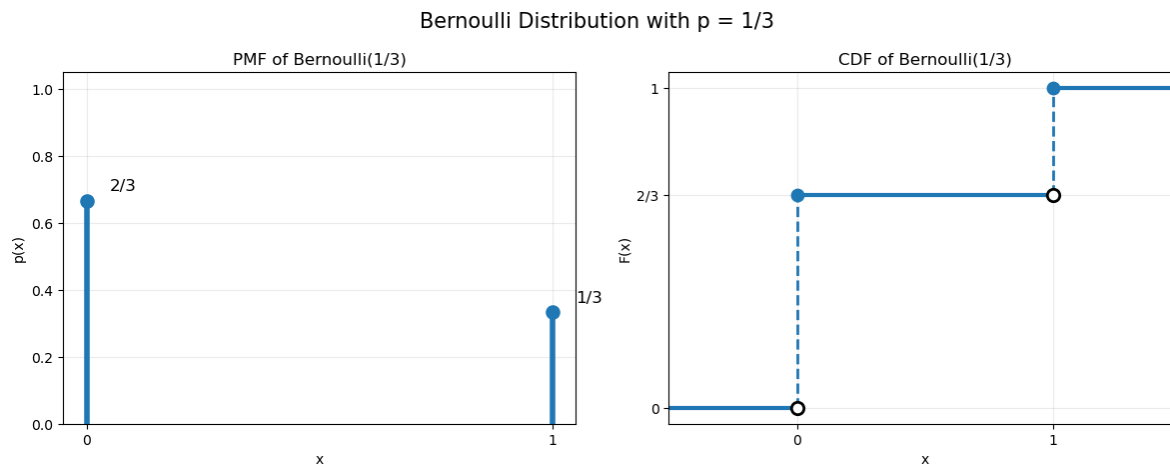
such that $P(X) = 1$.

1.3 Example 1: Bernoulli Distribution

Let $X \sim \text{Bernoulli}(p)$. Its cumulative distribution function is

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Notice that the CDF is a step function with jumps equal to the probabilities of the corresponding outcomes.



2 Continuous Probability Distributions

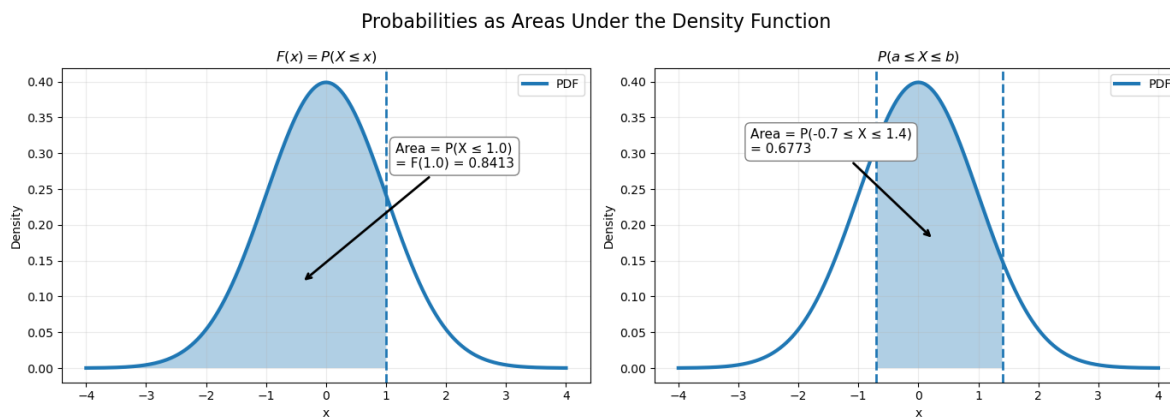
A probability distribution is called **absolutely continuous** if there exists a non-negative function

$$p : \mathbb{R} \rightarrow \mathbb{R}_+$$

such that

$$F(x) = \int_{-\infty}^x p(t) dt, \quad x \in \mathbb{R}.$$

The function $p(x)$ is called the **probability density function (PDF)** of the distribution.



2.1 Properties of the Probability Density Function

1. If the cumulative distribution function is differentiable, then

$$p(x) = F'(x).$$

2. Every probability density function integrates to one,

$$\int_{-\infty}^{\infty} p(x) dx = 1.$$

3. The probability of any Borel set $B \subseteq \mathbb{R}$ is obtained by integrating the density over that set,

$$P(B) = \int_B p(t) dt.$$

In particular,

$$P(a \leq X \leq b) = \int_a^b p(t) dt.$$

2.2 Examples of Absolutely Continuous Distributions

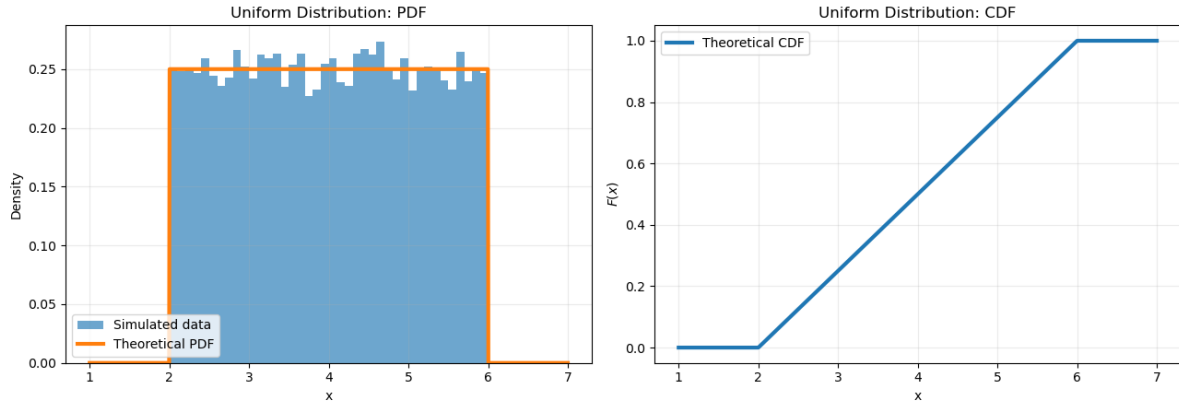
2.2.1 Uniform Distribution

Let $X \sim U([a, b])$, where $a < b$. The probability density function is

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

Every interval of the same length inside $[a, b]$ has the same probability.

Uniform Distribution on [2, 6]



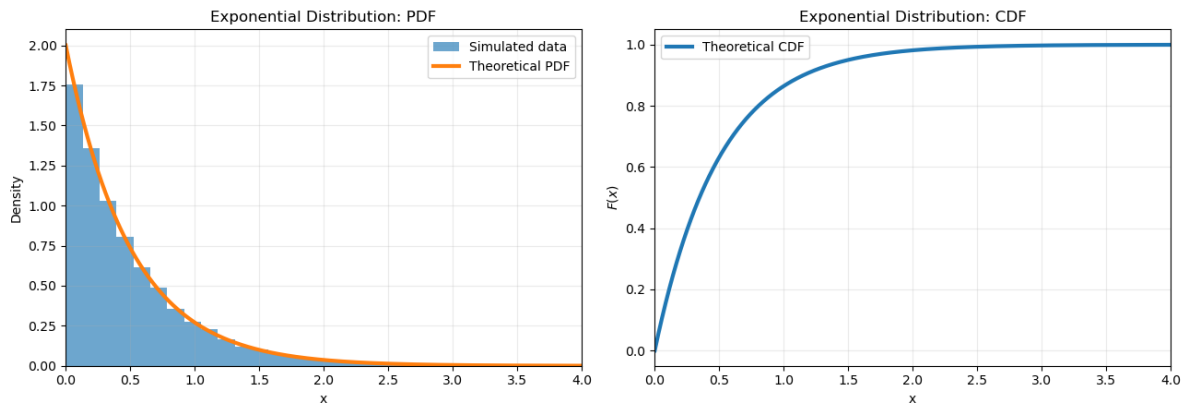
2.2.2 Exponential Distribution

Let $X \sim \text{Exp}(\lambda)$, where $\lambda > 0$. Its probability density function is

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The exponential distribution is commonly used to model **waiting times**.

Exponential Distribution Exp(2)



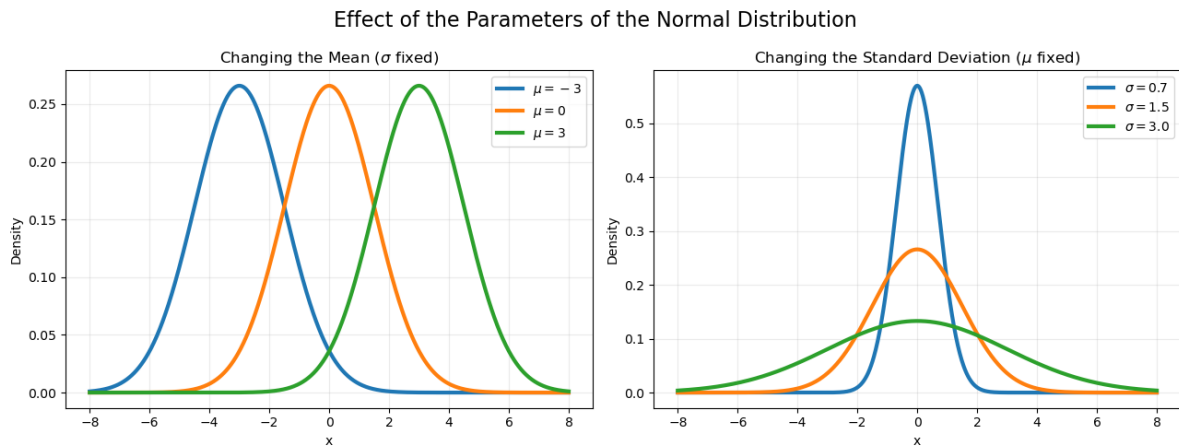
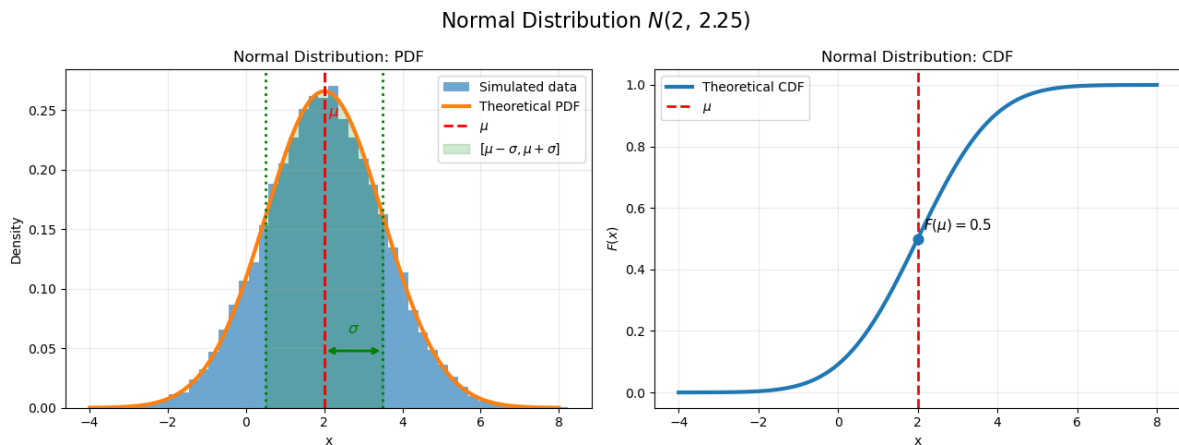
2.2.3 Normal Distribution

A random variable is said to follow a **normal distribution** with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, if its probability density function is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We write

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



The normal distribution is one of the most important probability distributions in statistics and data science. Many natural phenomena are approximately normally distributed, and many statistical procedures are based on this distribution.

A particularly important special case is the **standard normal distribution**

$$N(0, 1),$$

which will play a central role throughout the remainder of the course.

2.3 The Standard Normal Table

The cumulative distribution function of the standard normal distribution, $Z \sim N(0, 1)$, is denoted by

$$\Phi(z) = P(Z \leq z).$$

Unlike the cumulative distribution functions of the uniform or exponential distributions, the integral

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

cannot be evaluated using elementary functions.

For this reason, values of $\Phi(z)$ are usually obtained from a **standard normal table**, often called the **Z-table**.

2.3.1 How is the Z-table organized?

The Z-table contains numerical values of $\Phi(z) = P(Z \leq z)$ for many values of z .

Typically,

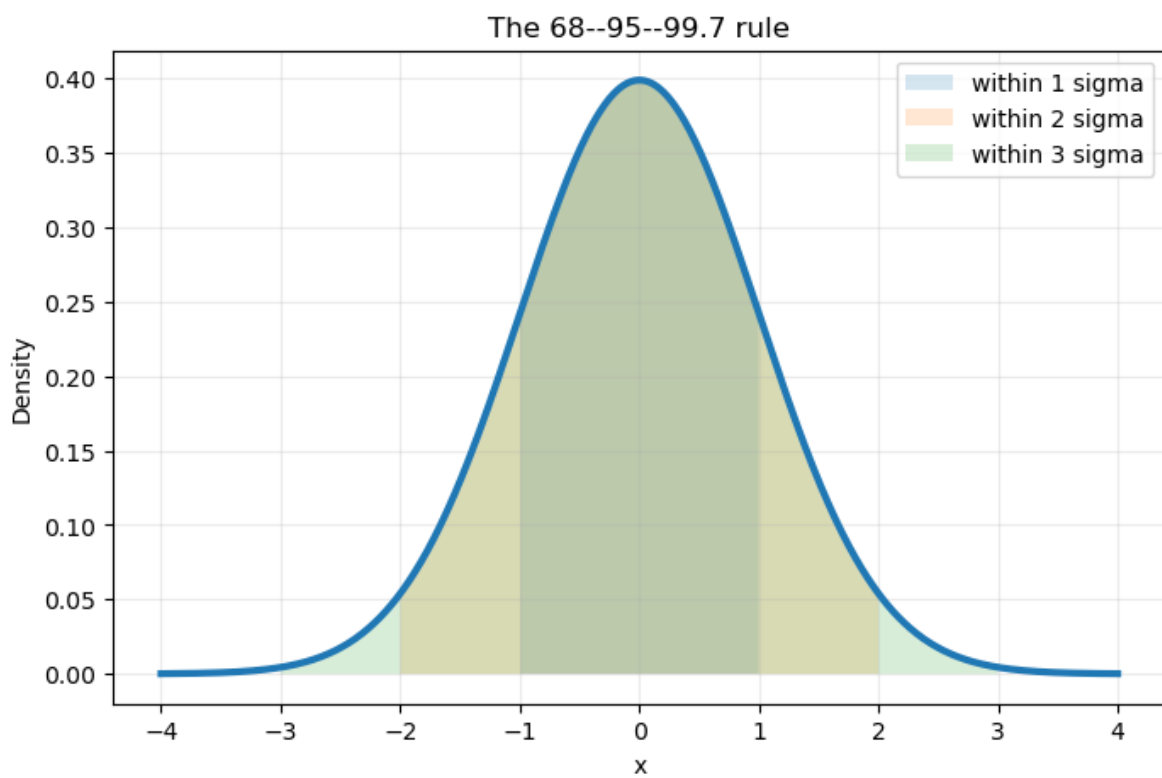
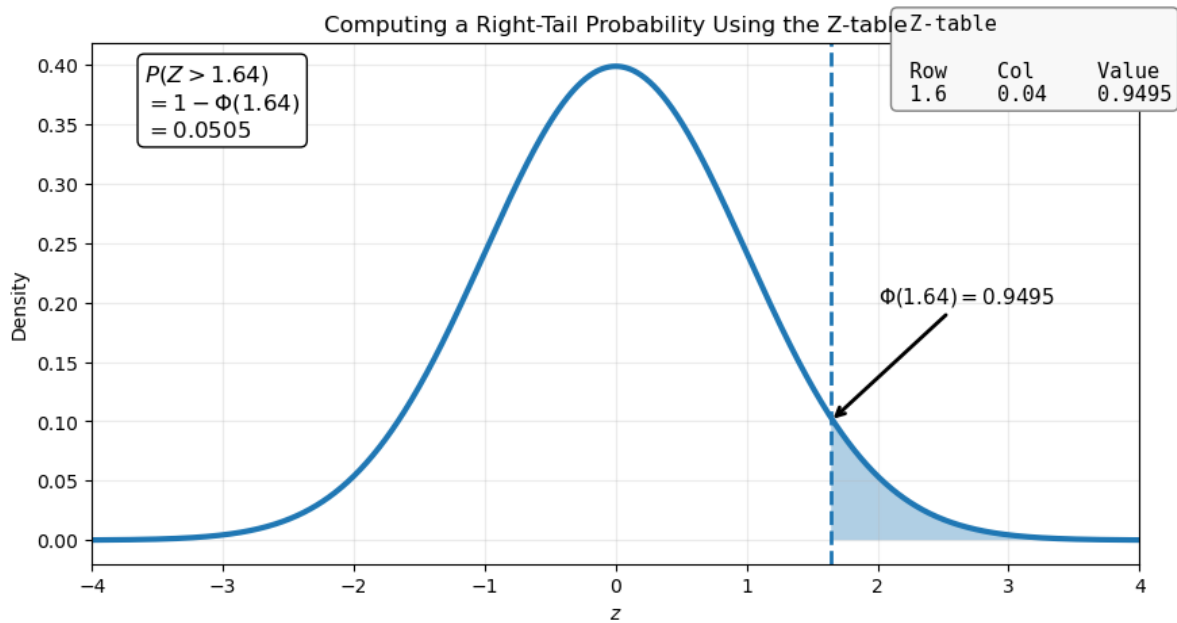
- the **row** contains the first decimal digit of z ;
- the **column** contains the second decimal digit.

Suppose we want to compute $P(Z \leq 1.23)$. Looking up the entry corresponding to row **1.2**, and column **0.03**, we obtain approximately

$\Phi(1.23) \approx 0.8907.$

Therefore,

$$P(N(0, 1) \leq 1.23) \approx 0.8907.$$



2.4 Why is the normal distribution so important?

Many real-world variables are approximately normally distributed. Human heights, standardized test scores, measurement errors, IQ scores, and many biological characteristics often follow a distribution that is remarkably close to a normal distribution. Even when the data are not perfectly normal, the normal distribution frequently provides an excellent approximation.

In the examples below, we compare two real datasets:

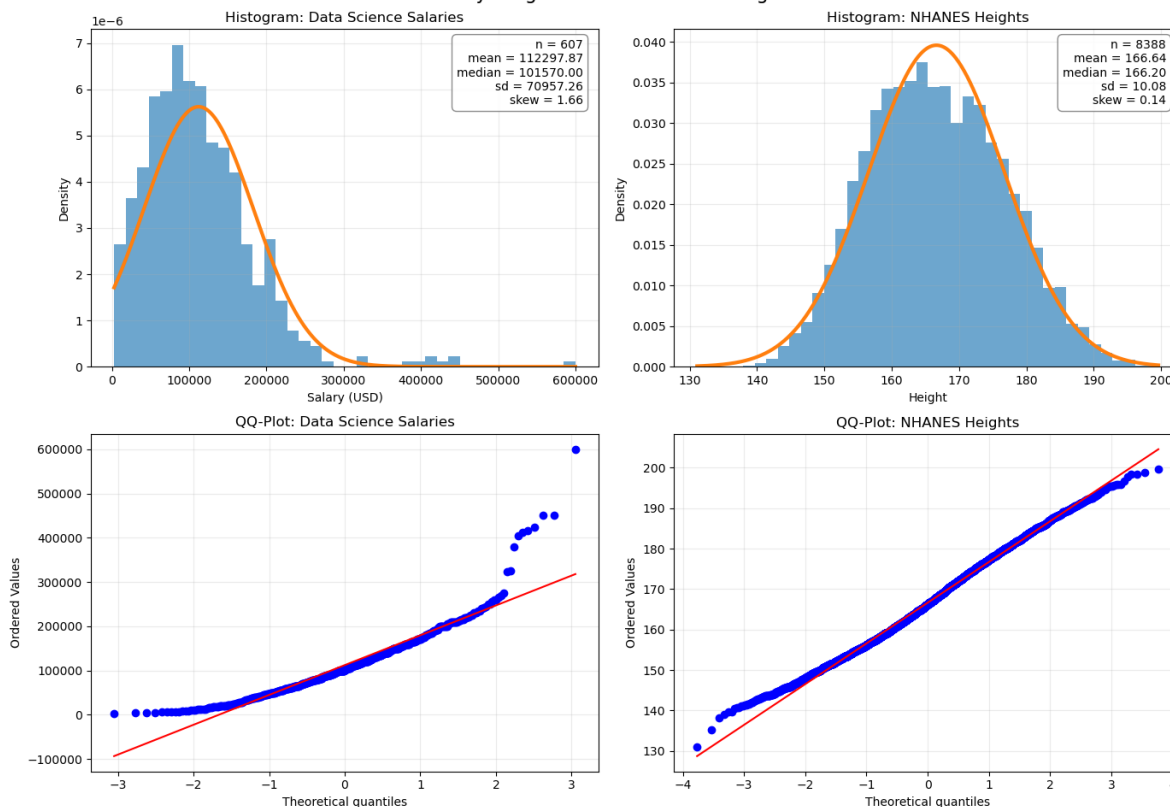
- **Data Science salaries (Worldwide, 2023)** — the distribution is noticeably right-skewed because a small number of very highly paid professionals earn much more than the majority.
- **NHANES adult heights (United States, 2017–2020)** — the distribution is very close to a normal distribution, as predicted by many biological models.

The histograms and QQ-plots clearly illustrate this contrast. While salaries deviate substantially from normality, human heights closely follow a bell-shaped curve, making the normal distribution an excellent model.

This is one of the main reasons why the normal distribution plays such a central role in probability and statistics. In later seminars, we will see that many statistical methods—including confidence intervals, hypothesis tests, linear regression, and the Central Limit Theorem—are built upon the remarkable mathematical properties of the normal distribution.

	Dataset	Year	Country	Sample size	Mean	Median	Std. deviation	S
0	Data science salaries	2023	Worldwide	607	112297.87	101570.0	70957.26	1
1	NHANES heights	2017–2020	United States	8388	166.64	166.2	10.08	0

Normality Diagnostics: Salaries vs Heights



3 Simulating Random Variables from Uniform Random Numbers

Computers do not directly generate random variables from every possible distribution. In practice, the basic object generated by a computer is usually a number that behaves like

$$U \sim U(0, 1).$$

From this uniform random variable, we can generate random variables with many other distributions by applying a suitable function.

The key idea is the **inverse transform method**.

3.1 The Inverse Transform Method

Let $U \sim U(0, 1)$ and let F be the CDF of a continuous random variable. Define

$$X = F^{-1}(U).$$

Then X has CDF F . Indeed, for any x ,

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x),$$

because U is uniformly distributed on $[0, 1]$.

Thus, $X = F^{-1}(U)$ transforms uniform random numbers into random numbers with distribution function F .

3.2 Example: Exponential Distribution

Let $X \sim \text{Exp}(\lambda)$. Its cumulative distribution function is

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

To generate X , we solve $u = 1 - e^{-\lambda x}$ for x .

This gives

$$x = -\frac{1}{\lambda} \ln(1 - u).$$

Hence, if $U \sim U(0, 1)$, then

$$X = -\frac{1}{\lambda} \ln(1 - U) \sim \text{Exp}(\lambda).$$

Since $1 - U$ has the same distribution as U , we often write simply $X = -\frac{1}{\lambda} \ln U$.

