Probability theory is a very young mathematical discipline, despite the fact that humans have been occupied with the (apparent) randomness of natural events throughout their cultural history. In fact, the ancient Egyptians are known to have played some versions of dice games. Pharaoh Rameses III, who lived in the 12th century B.C., arranged for himself to be portrayed on the high gate of the temple of Medinet Haboo playing a dice game with two of the women of his household.

Dice were not limited to the so-called classical civilisations. Members of cultures from around the world have gambled with dice featuring unusual shapes and markings. Native Americans, Aztecs, Mayas, Polynesians, Eskimos, and sub-Saharan Africans employed plum stones, peach pits, pebbles, seeds, bones, deer horn, pottery, walnut shells, beaver teeth, and seashells to play dice games.

Anyhow, it is unlikely that Julius Caesar, when he said the words "alea iacta est" ("the die is cast"), regarded his situation as a result of chance. Crossing the river Rubicon in 49 B.C., and incidentally dropping the famous quotation, Caesar passed beyond the limits of his province and became an invader in Italy, thus starting a civil war against the Senate of Rome.
Traditionally, randomness takes on an operational meaning in natural science: something is apparently random if its cause cannot be determined or controlled. When an experiment is performed and all the control variables are fixed, the remaining variation is ascribed to uncontrolled (i.e. "random") influences. Whether a patient survives a particular therapeutic intervention is usually dependent upon so many unknown factors that every doctor would subscribe to the view that at least a bit of "luck" is involved.

For most of the history of science, randomness has thus been interpreted as ignorance on the part of the observer. With the advent of quantum mechanics, however, it appears that the world might be irreducibly random. It is indeed possible to set up an experiment in which all relevant parameters are totally controlled, and yet which will still have a perfectly random outcome. Many physical processes resulting from quantum-mechanical effects are random in nature, with the best-known example being the timing of radioactive decay.

In the 1920's, Sir Ronald A. Fisher became the founding father of modern statistics when he developed and inspired a proper, scientific theory of the design and analysis of experiments. Most of his work was carried out at Rothamsted Agricultural Experiment Station, where he worked in close contact with empirical scientists for almost 15 years. Regarding the role of statistics in science, Fisher's views are probably best encapsulated by the following 1938 quotation of his: "To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of."
An event is said to occur at random if its outcome is variable and unpredictable.
Pierre Simon Laplace, a French mathematician and astronomer, was born in Beaumont-en-Auge in Normandy on March 23, 1749, and died in Paris on March 5, 1827. He was the son of a farm-labourer and owed his education mainly to the interest excited in some wealthy neighbours by his abilities and engaging presence. Laplace is remembered as one of the greatest scientists of all time (sometimes referred to as a French Newton) with a natural phenomenal mathematical faculty possessed by none of his contemporaries. One of the major themes of his life's endeavours was probability theory, and the 'classical' interpretation of probability as the ratio of the number of favourable cases to the whole number of possible cases was first indicated by Laplace in 1779. In 1812, Laplace also proved the so-called 'Central Limit Theorem' which provides an explanation for why so many data sets follow a distribution that is bell-shaped.

In his 1812 book 'Théorie analytique des probabilités', Laplace went as far as saying "We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it [...] It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge [...] The most important questions in life are, for the most part, really only problems of probability."
John Venn was an English mathematician, born in Hull on August 4, 1834, who made major contributions to the study of mathematical logic and probability. He is best known, however, for developing the 'Venn diagrams' used to graphically represent sets and their intersections.

Around the mid 19th century, the problems and paradoxes of the classical interpretation of probability had motivated the development of a different concept of probability. In his 1866 book 'Logic of Chance', John Venn was among those who laid the foundation for this so-called 'frequentist theory' of probability which holds that probabilities only make sense in the context of well-defined random experiments. In the view of a frequentist, the relative frequency of occurrence of an experimental outcome, when repeating the experiment, is a measure of the probability of that random event. Frequentism is still, by far, the most commonly held view among natural scientists.

John Venn died in Cambridge on 4 April, 1923.
Frank Plumpton Ramsey (1903-1930) was a British mathematician and philosopher who is best known for his work on the foundations of mathematics. However, Ramsey also made remarkable contributions to epistemology, semantics, logic, philosophy of science, mathematics, statistics, probability and decision theory, economics and metaphysics. His philosophical and scientific work comprises no more than some 15 papers. Although they are on quite disparate subjects, they nevertheless all contain the same view regarding philosophy as a method of analysis.

In his paper 'Truth and Probability', written in 1926, Ramsey laid the foundations for the modern theory of subjective probability. He showed how people's beliefs and desires can be measured by use of traditional betting methods. Thus, he claimed that a person's belief can be measured by proposing a bet to "see what are the lowest odds which he will accept." Ramsey took this method to be "fundamentally sound", but saw that it suffered from "being insufficiently general, and from being necessarily inexact [...] partly because of the diminishing marginal utility of money, partly because the person may have a special eagerness or reluctance to bet." To avoid these difficulties he laid the foundations for the modern theory of utility. He showed that if people obey a set of axioms or rules in their behaviour then the measure of their 'degrees of belief' will satisfy the laws of probability.
Andrei Nicolaevich Kolmogorov (born April 25, 1903, in Tambov, died October 20, 1987, in Moscow) was a Russian mathematician who made major advances in the fields of probability theory and topology. He is considered to be one of the greatest mathematicians of all time, with influences not only in mathematics, but also in many engineering disciplines.

Kolmogorov graduated from Moscow State University in 1925 when, remarkably enough, he had already published eight scientific papers, all written while still an undergraduate. He completed his doctorate in 1929 and was appointed a professor at Moscow University in 1931.

In his monograph 'Grundbegriffe der Wahrscheinlichkeitsrechnung' ('Foundations of the Theory of Probability'), published in 1933, Kolmogorov built up probability theory in a rigorous way from fundamental axioms in a way comparable to Euclid's treatment of geometry. Kolmogorov's axiomatization, which will be presented briefly on the next slide, has since achieved the status of orthodoxy, and it is typically what mathematicians and statisticians have in mind when they think of 'probability theory'.
Now, hold your breath ...
Kolmogorov's Probability Calculus

Let $\Omega$ be a non-empty set and let $\Delta$ be a family of subsets of $\Omega$ that has $\Omega$ as a member, and that is closed under complementation (with respect to $\Omega$) and union. Let $P$ be a function from $\Delta$ to the real numbers that fulfils

1. **Non-negativity** $P(A) \geq 0$ for all $A \in \Delta$
2. **Normalization** $P(\Omega) = 1$
3. **Additivity** $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$
   for all $A_1, A_2, ... \in \Delta$ with $A_i \cap A_j = \emptyset$

$P$ is called a 'probability' and $(\Omega, \Delta, P)$ is called a 'probability space'.

*Foundations of the Theory of Probability (1933)*

Don't worry, this is a formalised version of Kolmogorov’s axioms as it would appear in any probability theory text book. Remember however that the axioms represent the foundation of a whole mathematical discipline although, admittedly, their full comprehension would go far beyond the standard of a lecture on medical research methodology. Nevertheless, it is interesting to see that Kolmogorov’s axioms encapsulate ideas about probability that become immediately clear when illustrated by an appropriate example.

Think of $\Omega$ as a box with a large number of marbles of different colours. A straightforward choice of $\Delta$ would be the family of all subsets of marbles. This family contains $\Omega$ itself (because all marbles in the box are, of course, a subset of marbles), it is closed under complementation (the marbles not included in a subset, $A$, form a subset themselves, called '$A^C$') and union (two subsets A and B taken together form another subset, called '$A \cup B$').

Now, think of $P(A)$ as the relative number of marbles in a subset $A$. This is a mathematical function because it assigns a real number to a set. Furthermore, you can easily verify that it is a probability, too, since it fulfils all three of Kolmogorov’s axioms. By the way, many quantities that have nothing to do with probability do satisfy Kolmogorov’s axioms as well: normalized mass, length, area, and volume.
For a dice game, i.e. the standard example used to illustrate the principles of probability theory, $\Omega$ equals the set of numbers from 1 to 6. $\Delta$ is again the family of all subsets of $\Omega$, and the members of $\Delta$, i.e. the subsets of $\Omega$, will henceforth be called 'events'. For example, the subset $\{1,3,5\}$ equals the event 'uneven number', subset $\{5,6\}$ is the event 'at least five points'. For a given subset $A$, we define $P(A)$ again as the relative number of elements in $A$. This implies $P(\{1,3,5\})=1/2$ and $P(\{5,6\})=1/3$. 

<table>
<thead>
<tr>
<th>$P=6/6$</th>
<th>$P=4/6$</th>
<th>$P=2/6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${1,2,3,4,5,6}$</td>
<td>${1,2,3,4} {1,2,3,5} {1,2,3,6}$</td>
<td>${1,2} {1,3} {1,4} {1,5}$</td>
</tr>
<tr>
<td>${1,2,4,5} {1,2,4,6} {1,2,5,6}$</td>
<td>${1,6} {2,3} {2,4} {2,5}$</td>
<td>${1,6} {2,3} {2,4} {2,5}$</td>
</tr>
<tr>
<td>${1,3,4,5} {1,3,4,6} {1,3,5,6}$</td>
<td>${2,6} {3,4} {3,5} {3,6}$</td>
<td>${2,6} {3,4} {3,5} {3,6}$</td>
</tr>
<tr>
<td>${1,4,5,6} {2,3,4,5} {2,3,4,6}$</td>
<td>${4,5} {4,6} {5,6}$</td>
<td>${4,5} {4,6} {5,6}$</td>
</tr>
<tr>
<td>${2,3,5,6} {2,4,5,6} {3,4,5,6}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P=5/6$</th>
<th>$P=3/6$</th>
<th>$P=1/6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${1,2,3,4,5} {1,2,3,4,6}$</td>
<td>${1,2,3} {4,5,6} {1,2,4} {3,5,6}$</td>
<td>${1} {2} {3}$</td>
</tr>
<tr>
<td>${1,2,3,5,6} {1,2,4,5,6}$</td>
<td>${1,2,5} {3,4,6} {1,2,6} {3,4,5}$</td>
<td>${4} {5} {6}$</td>
</tr>
<tr>
<td>${1,3,4,5,6} {2,3,4,5,6}$</td>
<td>${1,3,4} {2,5,6} {1,3,5} {2,4,6}$</td>
<td>$P=0/6$</td>
</tr>
<tr>
<td>${1,3,6} {2,4,5} {1,4,5} {2,3,6}$</td>
<td>${1,3,6} {2,4,5} {1,4,5} {2,3,6}$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>${1,4,6} {2,3,5} {1,5,6} {2,3,4}$</td>
<td>${1,4,6} {2,3,5} {1,5,6} {2,3,4}$</td>
<td></td>
</tr>
</tbody>
</table>
Understanding probability theory is made easy by associating events with areas, and probabilities with the content of areas. This represents the measure theoretical view of probability. Thus, if events A and B are thought of as two areas of a Venn diagram, then the occurrence of at least one of the two events is the event that is represented by the union of the two areas, i.e. $A \cup B$. The simultaneous occurrence of A and B corresponds to the area that overlaps with both areas, i.e. $A \cap B$. Finally, $A^C$ (the 'complement' of A) is that part of the whole (i.e. $\Omega$) that does not overlap with A.

With these interpretations in mind, the two little theorems on this slide are easy to understand. First, the content of $A \cup B$ equals the sum of the contents of A and B, minus the content of $A \cap B$, which has been counted twice. Second, if the content of $\Omega$ is unity (as required by the 2$^{nd}$ Kolmogorov axiom), then the content of what lies outside of A is unity minus the content of A itself.
Two events are called 'stochastically independent' if the probability of their simultaneous occurrence equals the product of the individual event probabilities. For example, let us assume that the sex ratio in a population is 1:1, and that student birthdays are equally distributed over week days. Then we may conclude that the probability of the next student entering the lecture hall being a female Sunday's child equals 1/2 times 1/7, or 1/14.

This is one way of making use of the concept of independence, namely to assume independence on the basis of scientific evidence or common sense, and to calculate the probability of coincidence by multiplication. In scientific research, however, this logical relationship is often being used in reverse. Many experiments specifically address the question whether two random events are connected with one another or not, and use the experimental results to answer that question. In other words, independence may also be concluded from the fact that the probability of coincidence of two events equals the product of the individual event probabilities.
The association between hyperlipidemia and hypertension can be inferred from the fact that more than three times as many adult US Americans suffer from both conditions as expected "by chance alone". More formally, the estimated probability of a randomly drawn person being both hypertensive and hyperlipidemic (0.17) is three times larger than the product of the individual disease probabilities (0.25·0.20=0.05).
The original concept of random variables is mathematically complex and difficult to understand without a solid background in probability theory. In the context of discussing scientific research methods, however, it suffices to regard random variables as a way of simplifying the formal reference to random events that have not yet occurred. In a way, random variables are substitutes for experimental results that are either anticipated, or that will never be observed but are worth thinking about.

For example, if we are interested in studying the rate at which five dice show 6, we may refer to the possible outcome of a single game as "... the first die, which comes to a halt 50 milliseconds before the second, does not show a 6, whereas the second die, which stops 7.5 mm from the first die, shows a 6, which is not however the case for the third die which ...". Why not make life simpler? Let us denote the number of dice showing 6 by "X" and the possible outcome by "X=2". The complex pattern of red wood dice that we will later see on the table is then called a 'realisation of X'.
Realisations of random variables occur at random. Therefore it makes sense to ask questions such as "What is the probability that X will equal two?" or, more formally, "What is $P(X=2)$". The function $f(a)$ that assigns $P(X=a)$ to any possible outcome "a" of a realisation of X is called the 'probability function' of X. The probability function fully specifies the so-called 'distribution', and thereby the random nature, of X.
The binomial distribution is the most widely used model for binary random variables, i.e. for random variables that can only take two values (e.g. "male", "female"). In the definition of the probability function, the term $\pi^k (1-\pi)^{n-k}$ is short for the product $\pi \cdot \pi \cdot \pi \ldots \{k \text{ times}\} \cdot (1-\pi) \cdot (1-\pi) \ldots \{n-k \text{ times}\}$. Here, product formation reflects the independence of the $n$ replications. In each replication, we will either observe a success (with probability $\pi$) or a failure (with probability $1-\pi$), and we ask for the probability that there will be $k$ successes and $n-k$ failures. The obscure binomial coefficient is required to allow for the fact that the order of successes and failures is irrelevant.
Let us assume, for example, that \( n=5 \) and \( k=3 \). In addition to term \( \pi \cdot (1-\pi) \cdot \pi = \pi^3 (1-\pi)^2 \), calculation of \( f(3) \) also has to take into consideration term \( (1-\pi) \cdot \pi \cdot \pi = \pi^3 (1-\pi)^2 \) and, in fact, all products of three \( \pi \)'s and two \( (1-\pi) \)'s. But how many such products are there? The answer is given by the binomial coefficient.

Obviously, there are 5 possibilities to choose the first position, then 4 possibilities for the second and 3 possibilities for the third. This gives a total of \( 5 \cdot 4 \cdot 3 = 60 \) possibilities, but many of these differ only by the order in which particular positions were chosen. For example, the 60 possibilities include 1-2-3 as well as 3-2-1.

In order to correct for this multiplicity, the total number of choices (i.e. 60) must be divided by the number of equivalent choices, defined as choices of the same positions, but in different order. This last number obviously equals the number of orders (i.e. permutations) of 3 different positions, which in turn equals \( 3! = 3 \cdot 2 \cdot 1 = 6 \) (or \( k! \) in general).
Effectiveness of Antibiotics

An antibiotic is effective in 85% of patients with a particular disease. What is the probability that 8 or more out of 10 patients given the drug will be cured?

\[ n=10, \ \pi=0.85, \ k=8, \ 9 \ or \ 10 \]

\[
P(X \geq 8) = f(8) + f(9) + f(10) =
\]
\[
= \binom{10}{8}0.85^80.15^2 + \binom{10}{9}0.85^90.15^1 + \binom{10}{10}0.85^{10}0.15^0
\]
\[
= 45 \cdot 0.272 \cdot 0.023 + 10 \cdot 0.232 \cdot 0.150 + 1 \cdot 0.197 \cdot 1.000
\]
\[
= 0.820
\]
Probability Function

Bin(10,0.85)
Continuous random variables are conceptually similar to discrete random variables. With continuous random variables, however, it does not usually make sense to derive a probability function. For example, if $X$ denotes the BMI of a randomly chosen male from your home town population, it would be meaningless to ask for the probability that the BMI will be (exactly) equal to 22.5. The answer is "zero" and, for most continuous random variables $X$ of practical interest, it will be the same.

The distribution of continuous random variables is specified by the 'distribution function' instead of a probability function. For every real number $b$, the distribution function $F(b)$ equals the probability that $X$ takes values smaller or equal to $b$, i.e. $F(b)=P(X\leq b)$. 

Distribution Function of $X$

$F(b)=P(X\leq b)$ for real numbers $b$
Since distribution functions yield probabilities, their values are between 0 and 1. This immediately follows from the first and second of Kolmogorov’s axioms.

If a real number $b_1$ is smaller than another real number $b_2$, then event $X \leq b_1$ implies event $X \leq b_2$. Therefore, $P(X \leq b_1) \leq P(X \leq b_2)$ for any pair of real numbers with $b_1 < b_2$, i.e. every distribution function is monotonically increasing.
In practice, the distribution function of a continuous random variable is most often given by the integral of a so-called 'density function'. A density $f(x)$ is a non-negative function (non-negativity means $f(x) \geq 0$ for all $x$, thereby ensuring the validity of the 1st Kolmogorov axiom) for which the total area under the curve equals unity (2nd Kolmogorov axiom).

The integral of $f(x)$, taken from $-\infty$ ('minus infinity') to $b$, equals the area enclosed by the curve, the x axis, and a vertical line at $b$. In colloquial terms, this means that the probability of $X$ taking values smaller or equal to $b$, i.e. the distribution function $F(b)$, equals the "area under the density up to $b". Similarly, the probability of $X$ taking values between two real numbers $a$ and $b$ equals "the area under the density between $a$ and $b". 

$$F(b) = \int_{-\infty}^{b} f(x) \, dx$$

$$P(a < X \leq b) = F(b) - F(a)$$
Two random variables $X$ and $Y$ are called 'stochastically independent' if

$$P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$$

for any choice of real numbers $a$ and $b$. 

Two random variables $X$ and $Y$ are stochastically independent if the two events "$X$ takes a value smaller or equal to $a$" and "$Y$ takes a value smaller or equal to $b$" are stochastically independent for any choice of $a$ and $b$. 
### Independence of Random Variables

#### Examples

<table>
<thead>
<tr>
<th>Independent</th>
<th>Not Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>X: body height</td>
<td>X: body mass index</td>
</tr>
<tr>
<td>Y: time of consultation</td>
<td>Y: age</td>
</tr>
<tr>
<td>X: gender</td>
<td>X: blood pressure</td>
</tr>
<tr>
<td>Y: hair colour</td>
<td>Y: blood lipid level</td>
</tr>
</tbody>
</table>
The expected value, $E(X)$, of a quantitative random variable $X$ is the value that is expected, on average, to occur in a realisation. To this end, each possible value of $X$ is weighted by its probability of occurrence (or the density), and the weighted sum (discrete variable) or integral (continuous variable) is taken over the whole range of possible values. Similar to the sample mean in descriptive statistics, which measures the location of the sample data, the expected value measures the expected location of a single (future) data point (“where, on average, will my next observation lie?”).
This example nicely illustrates that the expected value of a random variable is not necessarily a possible value. No proper die can show a 3.5, but 3.5 is where the count will approximately be located (and not at 1256.6, or -5.9). Remember that the sample mean in descriptive statistics may also return a value that has never been observed, or that is even impossible to observe.
The 'Law of Large Numbers', an important theorem in probability theory, provides yet another, even more illustrative interpretation of the expected value of a random variable $X$. Let us assume that we are not only anticipating a single realisation of $X$, but a large number of realisations, say $n$. Then, the average of these realisations would also be the realisation of a (new) random variable. The Law of Large Numbers says that, if the number of realisations $n$ gets very large, the average of these realisations will be nearly constant, and not variable or random anymore. The value that the sample mean will take for large $n$ is exactly the expected value of $X$.

Therefore, the expected value of a random variable $X$ can be interpreted as the sample average if one and the same $X$ gets realised ("observed") a very large number of times. For example, the expected BMI of a single random male from your home town is the average BMI you would get if all males were recruited and analysed.
This figure illustrates the Law of Large Numbers, using a dice game as an example. Each dot represents the average count obtained in a game in which the die was thrown either n=10 (left column), n=100 (central column), or n=500 times (right column). As you can see, the average counts are located around the expected value of a single throw, i.e. 3.5, when the game is replicated 100 times for each n. With increasing n, however, the averages vary less over the 100 replicates. For n=500, they are almost constant at 3.5.
The variance, Var(X), of a random variable X is the expected value of a transformation of X, namely $[X-E(X)]^2$. First, the random variable is reduced by its expected value. This subtraction of $E(X)$ from $X$ is called 'centralisation', which ensures that the realisations of the ensuing transformation scatter around zero. Then, the difference is squared so that more emphasis is put upon large deviations of $X-E(X)$ from zero or, equivalently, of $X$ from $E(X)$.

The standard deviation of X equals the square root of Var(X). As in the case of the expected value, the theoretical standard deviation defined here equals the sample value that would be obtained if one and the same X gets realised a large number of times. The standard deviation of the BMI of a single random male from your home town therefore equals the standard deviation of the BMI among all these males.
Variance $\text{Var}(X)$

$E(X_1)$

$E(X_2)$

$\text{Var}(X_1) < \text{Var}(X_2)$
Game of Dice

X: number of points in a single roll

\[
\text{Var}(X) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} = 2.9
\]

Y: sum of points in two rolls

\[
\text{Var}(Y) = (2-7)^2 \cdot \frac{1}{36} + (3-7)^2 \cdot \frac{2}{36} + (4-7)^2 \cdot \frac{3}{36} + (5-7)^2 \cdot \frac{4}{36} + (6-7)^2 \cdot \frac{5}{36} + (7-7)^2 \cdot \frac{6}{36} + (8-7)^2 \cdot \frac{5}{36} + (9-7)^2 \cdot \frac{4}{36} + (10-7)^2 \cdot \frac{3}{36} + (11-7)^2 \cdot \frac{2}{36} + (12-7)^2 \cdot \frac{1}{36} = 5.8
\]
If X and Y are two random variables, then X+Y is of course also a random variable (as is X·Y, X/Y, Y/X etc.). For example, if X is the annual income of the male partner in a marriage, and Y is the income of the female partner, then X+Y denotes the annual family income. Similarly, if X is a random variable and \( \alpha \) is a real number, then \( \alpha \cdot X \) is also a random variable. Think of X as the weight of a random male in kilograms, which implies that \( 2.2 \cdot X \) would be the weight of a random male in pounds.

Some Computational Rules

\[
E(X+Y) = E(X) + E(Y) \\
E(\alpha \cdot X) = \alpha \cdot E(X) \\
\text{Var}(\alpha \cdot X) = \alpha^2 \cdot \text{Var}(X)
\]

if X and Y are independent

\[
E(X \cdot Y) = E(X) \cdot E(Y) \\
\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)
\]
The most important distribution for continuous random variables is the so-called 'normal' distribution. A random variable is said to follow a normal distribution if its distribution is specified by the density function $f(x)$ shown on this slide. The normal distribution is often called the 'Gaussian distribution', after Carl Friedrich Gauss (1777-1855), who discovered many of its properties (he did not, however, detect the distribution; this achievement was made around 1720 by the English mathematician Abraham de Moivre). Gauss, commonly viewed as one of the greatest mathematicians of all time (if not "the" greatest), was once properly honoured by Germany on their 10 Deutschmark bill.
A very special normal distribution is the so-called 'standard normal distribution' N(0,1), i.e. the normal distribution for which $\mu=0$ and $\sigma^2=1$. Tables summarizing the distribution function $\Phi(z)$ of the standard normal distribution can be found in any statistics text book. As we shall see later in the course, the value of $\Phi(z)$ highlighted on the present slide, $\Phi(1.96)=0.975$, is of particular practical relevance.
The distribution function \( F(b) \) of a normal distribution cannot be calculated simply "by hand", that is, by direct integration of the density function. On the other hand, it would not be possible to tabulate \( F(b) \) for all possible combinations of expected value \( \mu \) and variance \( \sigma^2 \). Fortunately, this turns out to be unnecessary because the distribution function of any normal distribution can be derived easily from the distribution function \( \Phi(z) \) of the standard normal distribution.

**Standardization of \( N(\mu, \sigma^2) \)**

If \( X \) has a normal distribution with expected value \( \mu \) and variance \( \sigma^2 \), then the random variable

\[
Z = \frac{X - \mu}{\sigma}
\]

has a standard normal distribution. The distribution function \( F(b) \) of \( X \) fulfils

\[
F(b) = \Phi\left(\frac{b - \mu}{\sigma}\right)
\]
**Blood Pressure**

If the diastolic blood pressure of healthy individuals has a normal distribution with expected value $\mu = 80$ mmHg and standard deviation $\sigma = 10$ mmHg, what is the probability that a randomly chosen individual has a blood pressure between 70 mmHg and 85 mmHg?

\[
P(70 \leq X \leq 85) = F(85) - F(70) \\
= \Phi\left(\frac{85 - 80}{10}\right) - \Phi\left(\frac{70 - 80}{10}\right) \\
= \Phi(0.5) - \Phi(-1) = 0.6915 - 0.1587 \\
= 0.5328
\]
This is a graphical illustration of the standard normal distribution. The density of any normal distribution is symmetric around its expected value (in this case zero) and exhibits the typical bell shape. The area highlighted in blue equals 0.68, which means that the probability of a random variable with \(N(0,1)\) distribution taking values between -1 and +1 is 0.68.

Values \(x=+1\) and \(x=-1\) correspond to the points of inflection of the density. If you were to drive along \(f(x)\) in a car, these are the two points where the steering wheel would be back in straight position. The points of inflection of the density of a general normal distribution, \(N(\mu, \sigma^2)\), are at \(\mu-\sigma\) and \(\mu+\sigma\), and the area under the density between the two points is again 0.68.
Owing to the symmetry of the normal density, the probability that a normally distributed random variable takes a value below its expected value is exactly 0.5.
Some 95% of the "mass" of the N(0,1) distribution falls between -1.96 and +1.96. In other words, the probability of a realisation falling outside this particular interval equals 1-0.95=0.05.
Some 95% of the mass of the $N(0,1)$ distribution falls below 1.65 and, owing to symmetry, 95% of the mass falls above -1.65 as well.
The blue vertical line marks the expected value $\mu$, the red lines are located at $\mu-\sigma$ and $\mu+\sigma$. As you can see, changing $\mu$ means shifting the density to the left (decreasing $\mu$) or right (increasing $\mu$), but leaving the shape unchanged. Changing $\sigma^2$ either stretches (increasing $\sigma^2$) or compresses (decreasing $\sigma^2$) the density, but leaves the position unchanged.
The 'Central Limit Theorem' is one of the most remarkable results of probability theory. It states that the mean of a large number of independent realisations from the same distribution has, after appropriate centralisation and standardisation, an approximate standard normal distribution. Moreover, the approximation steadily improves as the number of realisations increases. The Central Limit Theorem is considered the heart of probability theory.
Galton's board, also known as a 'quincunx' or 'bean machine', is a device for statistical experiments named after English scientist Sir Francis Galton. It consists of a lattice of pins that generate random walks for balls falling from the top to the bottom row. Each time a ball hits one of the pins, it may turn right or left. If a glass box is placed under each pin in the bottom row, and the number of balls is sufficiently large, the bar diagram formed by the heaps of balls will approximate a normal distribution.
The histogram depicts a sample of 100 average counts, each taken over 500 throws of a single die. The histogram somehow approximates a bell shaped curve, i.e. a normal distribution, but the approximation is still not very good.
The Central Limit Theorem explains why so many natural phenomena that represent the aggregation of small independent effects exhibit a bell shaped distribution in the population.

However, the distribution of IQ test scores cannot be expected to follow a bell curve unless it is constructed by the tester to do so. The shape of the distribution of IQ test scores will depend on the average difficulty of the test items as well as their correlation. The Central Limit Theorem does not apply to random variables which are correlated, and the high inter-item correlation in most IQ tests imply that the IQ distribution can take a variety of shapes.

Herrnstein and Murray's 1994 'The Bell Curve' argues that the IQ is a powerful predictor of many social ills, including crime. It uses this "scientific reality" to oppose social welfare policies and, in particular, justify the punishment of offenders. By portraying offenders as driven into crime predominantly by cognitive disadvantage, the authors mask the reality that stronger risk factors not only exist but also are amenable to effective correctional intervention.
Summary

- Probability theory, as a scientific discipline, is based upon Kolmogorov's axiomatic definition of probability.
- Random variables map complex (real) events onto simple numerical scales; they can be discrete or continuous.
- The distribution of a random variable is characterised by its probability function (discrete) or density (continuous).
- Random variables are independent if the joint distribution equals the product of their individual distributions.
- Important summary values of the distribution of random variables include the expected value and the variance.
- The normal distribution is a universal, large-scale approximation of the "average" of other random variables.