## STAT2201

# Analysis of Engineering \& Scientific Data 

## Unit 3

Slava Vaisman

The University of Queensland School of Mathematics and Physics

## What we learned in Unit 2 (1)

- We defined a sample space of a random experiment $\Omega$.
- We defined events (subset of $\Omega$ ).
- Finally, we assigned a "measure" which tells us how likely it is that a particular event will occur.
- These brought us to the formal definition of probability. Specifically to:


## Definition

A probability $\mathbb{P}: \mathcal{F} \rightarrow[0,1]$, is a rule (or function) which assigns a number between 0 and 1 to each event, and which satisfies the following axioms:

- $0 \leq \mathbb{P}(A) \leq 1$,
- $\mathbb{P}(\Omega)=1$,
- if $\left\{A_{i}\right\}$ are disjoint, then $\mathbb{P}\left(\bigcup_{i} A_{i}\right)=\sum_{i} \mathbb{P}\left(A_{i}\right)$.


## What we learned in Unit 2 (2)

- A measure can be defined on both discrete and continuous state spaces.
- In the discrete case it will (generally) result in solving a counting problem:

- In the continuous case it will (generally) result in a solution of finding an area (an integration) problem:



## What we learned in Unit 2 (3)

- We discussed conditional probability:

$$
\mathbb{P}(A \mid B)=\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}
$$

- the Chain rule:

$$
\mathbb{P}(A \mid B)=\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(A \cap B)=\mathbb{P}(A \mid B) \mathbb{P}(B)
$$

- the Law of Total Probability:

$$
\mathbb{P}(A)=\sum_{i=1}^{n} \mathbb{P}\left(A \cap B_{i}\right)=\sum_{i=1}^{n} \mathbb{P}\left(A \mid B_{i}\right) \mathbb{P}\left(B_{i}\right)
$$

## What we learned in Unit 2 ( (4)

- and, event Independence:

$$
A, B \text { independent iff } \mathbb{P}(A \mid B)=\mathbb{P}(A)
$$

$A, B$ independent iff $\mathbb{P}(A \cap B)=\mathbb{P}(A) \mathbb{P}(B)$.

We next proceed with an important concept of random variables.

## Random Variables

The outcome of a random experiment is often expressed as a number or measurement. For example,

- the number of defective transistors out of 100 inspected ones,
- the number of bugs in a computer program,
- the amount of rain in Brisbane in June,
- the amount of time needed for an operation.

This number is called a random variable.

Definition
A function $X(\omega)$ assigning to every outcome $\omega \in \Omega$ a real number is called a random variable.

## Some notation

- We usually write $X$ instead of $X(\omega)$.
- Instead of fully specifying an event via $\{\omega \in \Omega ; X(\omega) \leq x\}$, we abbreviate and write: $\{X \leq x\}$.
- Similarly, instead of $\{\omega \in \Omega ; X(\omega)=x\}$, write $\{X=x\}$.

The corresponding probabilities of these events are therefore expressed as follows.

- $\mathbb{P}(\{\omega \in \Omega ; X(\omega) \leq x\})=\mathbb{P}(\{X \leq x\})=\mathbb{P}(X \leq x)$.
- $\mathbb{P}(\{\omega \in \Omega ; X(\omega)=x\})=\mathbb{P}(\{X=x\})=\mathbb{P}(X=x)$.


## Example

- We flip a coin $n$ times.
- The sample space is $\Omega=\{0,1\}^{n}$, i.e. sequences of length $n$ of 0 's (failures) and 1's (successes).
- Consider the function $X: \Omega \rightarrow\{0, \ldots, n\}$ which maps $\boldsymbol{\omega}=\left(\omega_{1}, \ldots, \omega_{n}\right)$ to $X(\boldsymbol{\omega})=\omega_{1}+\omega_{2}+\cdots+\omega_{n}$.
- $X$ is a random variable. The set $\{X=k\}$ corresponds to the set of outcomes with exactly $k$ successes. Hence, we can interpret $X$ as the total number of successes in $n$ trials.
$\Rightarrow$ What is $\mathbb{P}(X=k)$ ?

$$
\mathbb{P}(X=k)=\binom{n}{k} p^{k}(1-p)^{n-k}
$$

where $p$ is the probability of success in a single experiment.

## Some important remarks

- Random variables are usually the most convenient way to describe random experiments; they allow us to use intuitive notations for certain events.
- Although mathematically a random variable is neither random nor a variable (it is a function), in practice we may interpret a random variable as the measurement on a random experiment which we will carry out "tomorrow". However, all the thinking about the experiment is done "today".
- We denote random variables by upper case Roman letters, $X, Y, \ldots$.
- Numbers we get when we make the measurement (the outcomes of the random variables) are denoted by the lower case letter, such as $x_{1}, x_{2}, x_{3}$ for three values for $X$.


## The range of a random variable

The set of all possible values a random variable $X$ can take is called the range of $X$. We further distinguish between discrete and continuous random variables:

- Discrete random variables can only take isolated values.

For example: a count can only take non-negative integer values.

- Continuous random variables can take values in an interval.

For example: rainfall measurements, lifetimes of components, lengths, ...; these are (at least in principle) continuous.

## A Function That Is Not a Random Variable

- Recall that a probability space is defined via a triple $(\Omega, \mathcal{F}, \mathbb{P})$.
- Suppose that $\Omega=\{1,2,3\}$.
- Let $\mathcal{F}=\{\emptyset,\{1,2\},\{3\},\{1,2,3\}\}$.
- Finally, define

$$
\left\{\begin{array}{l}
\mathbb{P}(\emptyset)=\mathbb{P}(\{1,2,3\})=0 \\
\mathbb{P}(\{1,2\})=0.6 \\
\mathbb{P}(\{3\})=0.4 .
\end{array}\right.
$$

- Now, define a function $X$ from $\Omega$ to $\mathbb{R}$. For example,

$$
X(1)=1, X(2)=2, X(3)=3 .
$$

- However, when trying to find $\mathbb{P}(X=1)$, we fail, since $\mathbb{P}(X=1)=\mathbb{P}(1)$; but, $\{1\} \notin \mathcal{F}$.
- We conclude that $X$ is not a random variable!


## Probability mass function

Let $X$ be a random variable. We would like to specify the probabilities of events such as

- $\{X=x\}$,
- $\{X \geq x\}$ or $\{X \leq x\}$, and
- $\left\{x_{1} \leq X \leq x_{2}\right\}$, where $x_{1}<x_{2}$.


## Definition (Probability mass function)

For discrete random variable $X$, the function $\mathbb{P}(X=x)$ is called the probability mass function (pmf) of $X$. We have for any
$B \subseteq\left\{x_{1}, \ldots, x_{n}\right\}$,

$$
\mathbb{P}(X \in B)=\sum_{x \in B} \mathbb{P}(X=x)
$$

Note that two additional properties are implied:

1. $\mathbb{P}(X=x) \geq 0$ for all $x$.
2. $\sum_{i=1}^{n} \mathbb{P}\left(X=x_{i}\right)=1$.

## Example (Fair die)

Toss a die and let $X$ be its face value. $X$ is discrete with range $\{1,2,3,4,5,6\}$. If the die is fair the probability mass function is given by

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | $\Sigma$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $p(x)=\mathbb{P}(X=x)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

Example (Maximum of two dice)
Toss two dice and let $Y$ be the largest face value showing. The pmf of $Y$ satisfies:

$$
\begin{array}{l|llllll|l}
y & 1 & 2 & 3 & 4 & 5 & 6 & \Sigma \\
\hline p(y)=\mathbb{P}(Y=y) & \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} & 1
\end{array}
$$

We can now work out the probability of any event defined by $Y$ so we know the distribution of $Y$. For example: $\mathbb{P}(Y>4)=\mathbb{P}(Y=5)+\mathbb{P}(Y=6)=\frac{9+11}{36}=\frac{5}{9}$.

## Probability density function

- Similarly, we would like to define a probability mass function for a continuous random variable.
- However, in the continuous case, $\mathbb{P}(X=x)=0$.
- Hence, we cannot characterize the distribution of $X$ via the probability mass function.


## Definition (Probability density function)

A random variable $X$ is said to have a continuous distribution if $X$ is a continuous random variable for which there exists a positive function $f$ with total integral 1 , such that for all $a, b$ :

$$
\mathbb{P}(a \leq X \leq b)=\int_{a}^{b} f(x) \mathrm{d} x
$$

## Probability density function

The function $f$ is called the probability density function (pdf) of $X$.


## Properties of probability density function

1. $f(x) \geq 0$ for all $x$.
2. $\int_{-\infty}^{\infty} f(x) \mathrm{d} x=1$.
3. Any function $f$ satisfying these conditions, and for which $\int_{a}^{b} f(x) \mathrm{d} x$ is well-defined, can be a pdf.
$f(x)$ can be interpreted as the "infinitesimal" probability, that $X=$ $x$; that is:

$$
\mathbb{P}(x \leq X \leq x+h)=\int_{x}^{x+h} f(x) \mathrm{d} x \approx h f(x)
$$

However, it is important to realise that $f(x)$ is not a probability. Recall that $\mathbb{P}(X=x)=0$.

## Example (Drawing a random number from continuous interval)

Draw a random number from the interval of real numbers [0, 2].
Each number is equally possible. Let $X$ represent the number. What is the probability density function $f$ ?

Solution: In general, a pdf of uniform random variable $a \leq X \leq b$ (for which each number is equally possible), is given by:

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

- Note that $\int_{a}^{b} \frac{1}{b-a} \mathrm{~d} x=1$, and
- $f(x) \geq 0$ for all $x$.

In or example, $a=0$ and $b=2$, so $f(x)=1 / 2$ for $0 \leq x \leq 2$.
And, we can calculate probabilities, such as

$$
\mathbb{P}(X \geq 1)=\int_{1}^{2} \frac{1}{2} \mathrm{~d} x=\frac{1}{2} \times 2-\frac{1}{2} \times 1=\frac{1}{2} .
$$

## Cumulative Distribution Function

Although we will usually work with pmfs for discrete and pdfs for continuous random variables, the following function is defined for both continuous and discrete random variables.
Definition (Cumulative Distribution Function)
The cumulative distribution function (cdf) of $X$ is the function $F: \mathbb{R} \rightarrow[0,1]$ defined by:

$$
F(x)=\mathbb{P}(X \leq x)
$$



## Properties of cumulative distribution function

$$
F(x)=\mathbb{P}(X \leq x)
$$

1. $0 \leq F(x) \leq 1$.
2. $F$ is increasing: $x \leq y \quad \Rightarrow \quad F(x) \leq F(y)$.
3. It holds that $\lim _{x \rightarrow \infty} F(x)=1$, and $\lim _{x \rightarrow-\infty} F(x)=0$.
4. $F$ is right-continuous: $\lim _{h \rightarrow 0} F(x+h)=F(x)$.

- For a discrete random variable $X$ the cdf $F$ is a step function with jumps of size $\mathbb{P}(X=x)$ at all the points $x \in\left\{x_{1}, \ldots, x_{n}\right.$.
- For a continuous random variable $X$ with pdf $f$, the $\operatorname{cdf} F$ is continuous, and satisfies $F(x)=\int_{-\infty}^{x} f(u) \mathrm{d} u$. Therefore,

$$
f(x)=\frac{\mathrm{d}}{\mathrm{~d} x} F(x) .
$$

Example (Back to drawing a random number from continuous interval)
Draw a random number from the interval of real numbers [0, 2].
Each number is equally possible. Let $X$ represent the number.
Find the cdf $F$ of $X$ and derive the probability density function $f$.

## Solution:

- Take an $x \in[0,2]$.
- Drawing a number $X$ "uniformly" in [0,2] means that $\mathbb{P}(X \leq x)=x /|\Omega|=x / 2$, for all such $x$. In particular, the cdf of $X$ satisfies:

$$
F(x)= \begin{cases}\frac{x}{2} & 0 \leq x \leq 2 \\ 0 & \text { otherwise }\end{cases}
$$

$\Rightarrow$ Since $f(x)=\frac{\mathrm{d}}{\mathrm{d} x} F(x)$,

$$
f(x)= \begin{cases}\frac{1}{2} & 0 \leq x \leq 2 \\ 0 & \text { otherwise }\end{cases}
$$

## Examples

$$
F(x)=\left\{\begin{array}{ll}
0 & x<-1 \\
0.3 & -1 \leq x<1 \\
1 & x \geq 1,
\end{array} \quad F(x)= \begin{cases}0 & x<0 \\
x & 0 \leq x<1 \\
1 & x \geq 1 .\end{cases}\right.
$$

## What next?

- If we can specify all probabilities involving $X$, we say that we have specified the probability distribution of $X$.
- This can be done via cdf or, via pmf/pdf.
- Describing an experiment via a random variable and its pdf, pmf or cdf seems much easier than describing the experiment by giving the probability space. In fact, we have not used a probability space in the above examples.
- Although all the probability information of a random variable is contained in its cdf or pmf/pdf, it is often useful to consider various numerical characteristics.
- Specifically, one such number is the expectation of a random variable; it is a sort of "weighted average" of the values that $X$ can take.


## Expectation and Variance

## Definition (Expectation)

Let $X$ be a discrete random variable with pmf $f$. The expectation (or expected value) of $X$, denoted by $\mathbb{E}[X]$, is defined by

$$
\mathbb{E}[X]=\sum_{x} x \mathbb{P}(X=x)
$$

This number, sometimes written as $\mu_{X}$, is an indication of the "mean" of the distribution.

Example (Mean toss of a fair die)
Find $\mathbb{E}[X]$ if $X$ is the outcome of a toss of a fair die. Since

$$
\mathbb{P}(X=1)=\cdots=\mathbb{P}(X=6)=1 / 6
$$

we have $\mathbb{E}[X]=\sum_{x} x \mathbb{P}(X=x)=\sum_{i=1}^{6} i \times 1 / 6=7 / 2$.

## Expectation

Note: $\mathbb{E}[X]$ is not necessarily a possible outcome of the random experiment.

- One way to interpret the expectation is as a type of "expected profit".
- Specifically, suppose we play a game where you throw two dice, and $I$ pay you out, in dollars, the sum of the dice, $X$ say.
- However, to enter the game you must pay me $d$ dollars.
- You can play the game as many times as you like. What would be a "fair" amount for $d$ ? Answer

$$
\begin{aligned}
d & =\mathbb{E}[X]=\sum_{i=2}^{12} i \times \mathbb{P}(X=i)= \\
& =2 \times \mathbb{P}(X=2)+\cdots+12 \times \mathbb{P}(X=2)= \\
& =2 \times \frac{1}{36}+3 \times \frac{3}{36}+\cdots+12 \times \frac{1}{36}=7 .
\end{aligned}
$$

## Expectation

- Namely, in the long run the fractions of times the sum is equal to $2,3,4, \ldots$ are $1 / 36,2 / 36,3 / 36, \cdots$,
- so the average pay-out per game is the weighted sum of $2,3,4, \ldots$ with the weights being the probabilities/fractions.
- Thus the game is "fair" if the average profit (pay-out $-d$ ) is zero.

Another interpretation of expectation is as a centre of mass. Imagine that point masses with weights $p_{1}, p_{2}, \ldots, p_{n}$ are placed at positions $x_{1}, x_{2}, \ldots, x_{n}$ on the real line


## Expectation



Then there centre of mass, the place where we can "balance" the weights, is

$$
\text { centre of mass }=x_{1} p_{1}+\cdots+x_{n} p_{n}
$$

which is exactly the expectation of the discrete variable $X$ taking values $x_{1}, \ldots, x_{n}$ with probabilities $p_{1}, \ldots, p_{n}$.

## Expectation of a function of $X$

## Definition (Expectation of a function of $X$ )

If $X$ is a discrete random variable, then, for any real-valued function $g$,

$$
\mathbb{E}[g(X)]=\sum_{x} g(x) \mathbb{P}(X=x)
$$

Example
Toss of a fair die Find $\mathbb{E}\left[X^{2}\right]$ if $X$ is the outcome of the toss of a fair die. We have

$$
\mathbb{E}\left[X^{2}\right]=1^{2} \times \frac{1}{6}+2^{2} \times \frac{1}{6}+\cdots+6^{2} \times \frac{1}{6}=\frac{91}{6}
$$

Note that

$$
\frac{91}{6}=\mathbb{E}\left[X^{2}\right] \neq(\mathbb{E}[X])^{2}=\left(\frac{7}{2}\right)^{2}=\frac{49}{4}
$$

## Expectation in the continuous case

- By simply replacing the probability mass function with the probability density function and the summation with an integration,
- we find the expectation of a (function of a) continuous random variable. In particular, if the continuous variable has density function $f$, then

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

Example (Expectation of a signal)
Let $Y=a \cos (\omega t+X)$ be the value of a sinusoid signal at time $t$ with uniform random phase $X \in(0,2 \pi]$.

## Random signal example

1. The expected value $\mathbb{E}[Y]$ is

$$
\begin{aligned}
\mathbb{E}[Y] & =\mathbb{E}[a \cos (\omega t+X)]=\int_{0}^{2 \pi} \cos (\omega t+x) \frac{1}{2 \pi} \mathrm{~d} x \\
& =\left.a \frac{1}{2 \pi} \sin (\omega t+x)\right|_{0} ^{2 \pi}=0
\end{aligned}
$$

as is to be expected.
2. More important is the average power of the signal, i.e. $\mathbb{E}\left[Y^{2}\right]$. Since $\cos ^{2}(x)=1+\cos (2 x)$, we have,

$$
\begin{aligned}
\mathbb{E}\left[Y^{2}\right] & =a^{2} \mathbb{E}\left[\cos ^{2}(\omega t+X)\right]=a^{2} \mathbb{E}[1+\cos (2 \omega t+2 X)]= \\
& =\frac{a^{2}}{2}+\frac{a^{2}}{4 \pi} \int_{0}^{2 \pi} \cos (2 \omega t+2 x) \mathrm{d} x=\frac{a^{2}}{2} .
\end{aligned}
$$

## The variance of a random variable

## Definition (Variance)

The variance of a random variable $X$, denoted by $\operatorname{Var}(X)$ is defined by

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

- This number, sometimes written as $\sigma_{X}^{2}$, measures the spread or dispersion of the distribution of $X$.
- It may be regarded as a measure of the consistency of outcome: a smaller value of $\operatorname{Var}(X)$ implies that $X$ is more often near $\mathbb{E}[X]$ than for a larger value of $\operatorname{Var}(X)$.
- The square root of the variance is called the standard deviation.


## The variance of a random variable



## Properties of expectation

The expectation is "linear"; specifically, it holds that:

1. $\mathbb{E}[a X+b]=a \mathbb{E}[X]+b$.
2. $\mathbb{E}[g(X)+h(X)]=\mathbb{E}[g(X)]+\mathbb{E}[h(X)]$.

## Proof.

1. Suppose $X$ has a pmf $f(x)$. Then (the proof is similar for the discrete case)

$$
\begin{aligned}
\mathbb{E}[a X+b] & =\int(a x+b) f(x) \mathrm{d} x=a \int x f(x) \mathrm{d} x+b \int f(x) \mathrm{d} x \\
& =a \mathbb{E}[X]+b \times 1=a \mathbb{E}[X]+b
\end{aligned}
$$

2. Similarly,

$$
\begin{aligned}
\mathbb{E}[g(X)+h(X)] & =\int(g(x)+h(x)) \mathrm{d} x \\
& =\int g(x) \mathrm{d} x+\int h(x) \mathrm{d} x=\mathbb{E}[g(X)]+\mathbb{E}[h(X)] .
\end{aligned}
$$

## Properties of variance

1. $\operatorname{Var}(X)=\mathbb{E}\left[X^{2}\right]-(\mathbb{E}[X])^{2}$.
2. $\operatorname{Var}(a X+b)=a^{2} \operatorname{Var}(X)$.

## Proof.

1. For convenience, write $\mathbb{E}[X]=\mu$; then,

$$
\begin{aligned}
\operatorname{Var}(X) & =\mathbb{E}\left[(X-\mu)^{2}\right]=\mathbb{E}\left[X^{2}-2 X \mu+\mu^{2}\right] \\
& =\mathbb{E}\left[X^{2}\right]-2(\mathbb{E}[X])^{2}+(\mathbb{E}[X])^{2}=\mathbb{E}\left[X^{2}\right]-(\mathbb{E}[X])^{2}
\end{aligned}
$$

2. Similarly, since $\mathbb{E}[a X+b]=a \mu+b$, we have:

$$
\begin{aligned}
\operatorname{Var}(a X+b) & =\mathbb{E}\left[(a X+b-(a \mu+b))^{2}\right] \\
& =\mathbb{E}\left[a^{2}(X-\mu)^{2}\right]=a^{2} \operatorname{Var}(X)
\end{aligned}
$$

## Moments of a random variable

- Note that

$$
\mathbb{E}[X]=\mathbb{E}\left[X^{1}\right]
$$

- Similarly,

$$
\operatorname{Var}(X)=\mathbb{E}\left[X^{2}\right]-\left(\mathbb{E}\left[X^{1}\right]\right)^{2}
$$

- It can be useful to know

$$
\mathbb{E}\left[X^{r}\right]
$$

which is called the $r$-th moment of $X$, because many quantities of interest are a function of these moments.

## Remark:

However, note that the expectation, or any moment of a random variable, need not always exist or can be $\pm \infty$.

## Important Discrete Distributions

A random variable is said the have a discrete distribution if $S$ is countable, and for any subset $B \subset S$,

$$
\mathbb{P}(X \in B)=\sum_{x \in B} \mathbb{P}(X=x)
$$

- Think of $X$ as the measurement of a random experiment that will be carried out tomorrow.
- However, all the "thinking" is done today.
- The behaviour of the experiment is summarized by the probability mass function.


## Important Discrete Distributions (Bernoulli)

## Definition

We say that $X$ has a Bernoulli distribution with success probability $p$ if $X$ can only assume the values 0 and 1 , with probabilities

$$
P(X=1)=p=1-P(X=0) .
$$

- We write $X \sim \operatorname{Ber}(p)$.
- Despite its simplicity, this is one of the most important distributions in probability!
- It models for example:
- a single coin toss experiment,
- a success or a failure of message passing,
- a success of a certain drug,
- or, randomly selecting a person from a large population, and ask if she votes for a certain political party.


## Properties of the Bernoulli Distribution

- The pmf is $f(x, p)=p^{x}(1-p)^{1-x}$.
- The expected value is

$$
\mathbb{E}[X]=1 \times p+0 \times(1-p)=p
$$

- The variance is

$$
\operatorname{Var}(X)=\mathbb{E}\left[X^{2}\right]-(\mathbb{E}[X])^{2}=1^{2} \times p+0^{2} \times(1-p)-p^{2}=p(1-p)
$$



Figure: The cdf of the Bernoulli distribution

## A sequence of independent Bernoulli trials (1)

- Often, we have a sequence of independent Bernoulli trials.
- That is, we sequentially perform Bernoulli experiments, such that the outcome (success or failure) of each experiment does not depend on the other experiments.
- Here is a way to graphically show the outcomes (white failure, green - success):



## A sequence of independent Bernoulli trials (2)

- Let the sample space for each trial be $\{0,1\}$ ( 1 for a success, 0 for a failure).
- The sample space $\Omega$ of a sequence of $n$ trials is therefore the set of all binary vectors of length $n$ :

$$
\Omega=\{(0,0, \ldots, 0), \ldots(1,1, \ldots, 1)\}
$$

- To specify $\mathbb{P}$, let $A_{i}$ denote the event of "success" during the $i$-th trial.
- By definition,

$$
\mathbb{P}\left(A_{i}\right)=p, \quad i=1,2, \ldots
$$

- Finally, $\mathbb{P}$ must be such that $A_{1}, A_{2}, \ldots$ are independent.
- These two rules completely specify $\mathbb{P}$.


## Binomial Distribution

- Consider a sequence of $n$ coin tosses.
- If $X$ is the random variable which counts the total number of heads and the probability of "head" is $p$ then we say $X$ has a binomial distribution with parameters $n$ and $p$
- and write $X \sim \operatorname{Bin}(n, p)$.
- The probability mass function $X$ is given by

$$
f(x, p)=\binom{n}{x} p^{x}(1-p)^{n-x}, \quad x=0,1,2, \ldots, n .
$$

- The expected value of $X \sim \operatorname{Bin}(n, p)$ is equal to $n p$. This is sort of intuitive, since our success probability in a single trial is $p$ and we perform $n$ experiments overall.
- However, one need to show formally that

$$
\mathbb{E}[X]=\sum_{x=0}^{n} x\binom{n}{x} p^{x}(1-p)^{n-x}=n p
$$

## Properties of Binomial Distribution (Expectation)

- Instead of evaluating the sum

$$
\mathbb{E}[X]=\sum_{x=0}^{n} x\binom{n}{x} p^{x}(1-p)^{n-x}=n p
$$

we can express $X$ as

$$
X=X_{1}+X_{2}+\cdots+X_{n}
$$

where $X_{i} \sim \operatorname{Ber}(p)$.

- By linearity of expectation

$$
\mathbb{E}[X]=\mathbb{E}\left[X_{1}+X_{2}+\cdots+X_{n}\right]=\sum_{i=1}^{n} \mathbb{E}\left[X_{i}\right]=n p
$$

## Properties of Binomial Distribution (Variance)

- The variance of $X$ is $\operatorname{Var}(X)=n p(1-p)$.
- This is proved in a similar way to the expectation (since Bernoulli variables are uncorrelated):

$$
\begin{aligned}
\operatorname{Var}(X) & =\operatorname{Var}\left(X_{1}+X_{2}+\cdots+X_{n}\right) \\
& =\sum_{i=1}^{n} \operatorname{Var}\left(X_{i}\right)=n p(1-p) .
\end{aligned}
$$

## Binomial Distribution



## Binomial Distribution



## Binomial Distribution

$$
\mathrm{n}=10, \mathrm{p}=0.7
$$



## Binomial Distribution - example

- In a large country, $51 \%$ favours party A and $49 \%$ favours party B.
- We randomly select 200 people from this population.
- What is the probability that of this group more people vote for $B$ than for $A$ ?


## Solution:

1. Let a vote for A be a "success".
2. Selecting the 200 people is equivalent to performing a sequence of 200 independent Bernoulli trials with success probability 0.51 .
3. We are looking for the probability that we have less than 100 successes, which is (use computer)

$$
\sum_{x=0}^{99}\binom{200}{x} 0.51^{x} 0.49^{200-x} \approx 0.36
$$

## Geometric Distribution

- Again we look at a sequence of coin tosses but count a different thing.
- Let $X$ be the number of tosses needed before the first head occurs. Then:

$$
\mathbb{P}(X=x)=(1-p)^{x-1} p, \quad x=1,2,3, \ldots
$$

- since the only string that has the required form is

$$
\underbrace{T T T T \ldots T}_{t-1 \text { times }} H
$$

and this has probability $(1-p)^{x-1} p$.

- Such a random variable $X$ is said to have a geometric distribution with parameter $p$. We write $X \sim \mathrm{G}(p)$.


## Properties of Geometric Distribution

- It can be shown that

$$
\mathbb{E}[X]=\frac{1}{p},
$$

- and that

$$
\operatorname{Var}(X)=\frac{1-p}{p^{2}}
$$

## Geometric Distribution



## Geometric Distribution



## Geometric Distribution - example

- In the game Ludo you have to throw a six before you can put your token onto the board.
- What is the probability that you need more than 6 throws of the die before this happens?


## Solution:

We are again dealing with a sequence of independent Bernoulli trials, with success probability $1 / 6$ of throwing a 6 . Hence the required probability is

$$
\begin{aligned}
\sum_{k=7}^{\infty}\left(\frac{5}{6}\right)^{m-1} \frac{1}{6} & =\left(\frac{5}{6}\right)^{6} \frac{1}{6} \sum_{k=0}^{\infty}\left(\frac{5}{6}\right)^{k} \\
& =\left(\frac{5}{6}\right)^{6} \frac{1}{6} \frac{1}{1-\frac{5}{6}}=\left(\frac{5}{6}\right)^{6} \approx 0.33
\end{aligned}
$$

## Geometric Distribution - the memoryless property

- A property of the geometric distribution which deserves extra attention is the memoryless property.
- Think again of the coin toss experiment. Suppose we have tossed the coin $k$ times without a success (Heads).
- What is the probability that we need more than $x$ additional tosses before getting a success?
- The answer is, obviously, the same as the probability that we require more than $x$ tosses if we start from scratch, that is,

$$
\mathbb{P}(X>x)=(1-p)^{x}
$$

irrespective of $k$.

- The fact that we have already had $k$ failures does not make the event of getting a success in the next trial(s) any more likely.


## Geometric Distribution - the memoryless property (2)

- In other words, the coin does not have a memory of what happened, hence the word memoryless property.
- Mathematically, it means that for any $x, k=1,2, \ldots$,

$$
\mathbb{P}(X>k+x \mid X>k)=\mathbb{P}(X>x)
$$

Proof: By the definition of conditional probability

$$
\mathbb{P}(X>k+x \mid X>k)=\frac{\mathbb{P}(\{X>k+x\} \cap\{X>k\})}{\mathbb{P}(X>k)}
$$

- Note that the event $\{X>k+x\}$ is a subset of $\{X>k\}$.
$\Rightarrow$ In addition, the probability of $\{X>k+x\}$ and $\{X>k\}$ is $(1-p)^{k+x}$ and $(1-p)^{k}$, respectively, so:

$$
\begin{aligned}
\mathbb{P}(X>k+x \mid X>k) & =\frac{\mathbb{P}(X>k+x)}{\mathbb{P}(X>k)}=\frac{(1-p)^{k+x}}{(1-p)^{k}}=(1-p)^{x} \\
& =\mathbb{P}(X>x)
\end{aligned}
$$

## Geometric Distribution - a real life application

- Suppose that a person is looking for a job.
- However, the market is not in a very good condition, so the probability that the person is accepted (after a job interview) is rather small, say $p=0.05$.
- Let $X$ be the number of interviews in which the person is rejected until he eventually finds a job.
- $X$ is clearly a geometric random variable.
- In particular $\mathbb{E}[X]=1 / p=20$.
- In addition, the probability that a person is not accepted after $k$ tries drops rapidly, specifically, it is equal to $(1-p)^{k}$.
- For example, for $k=50$, this probability is smaller than $8 \%$ !
- What is your conclusion?


## Important Continuous Distributions

A random variable has a continuous distribution with probability density function (pdf) $f$ if, for all $[a, b]$,

$$
\mathbb{P}(a \leq X \leq b)=\int_{a}^{b} f(x) \mathrm{d} x
$$

The cumulative distribution function (cdf) $F$ is given by

$$
F(x)=\mathbb{P}(X \leq x)=\int_{-\infty}^{x} f(u) \mathrm{d} u
$$

- Think of $X$ as the result of a random experiment that will be carried out tomorrow, but all the "thinking" is done today.
- The behaviour of the experiment is summarized by $f$ or $F$.


## Some Important Continuous Distributions - Uniform Distribution

We say that a random variable $X$ has a uniform distribution on the interval $[a, b]$, if it has density function $f$, given by

$$
f(x)=\frac{1}{b-a}, \quad a \leq x \leq b
$$

- We write $X \sim \mathrm{U}[a, b]$.
- $X$ can model a randomly chosen point from the interval $[a, b]$, where each choice is equally likely.


Figure: The pdf of the uniform distribution on $[a, b]$.

## Properties of Uniform Distribution

- The expected value of $X \sim \mathrm{U}[a, b]$ is

$$
\mathbb{E}[X]=\int_{a}^{b} x \frac{1}{b-a} \mathrm{~d} x=\frac{1}{b-a} \frac{b^{2}-a^{2}}{2}=\frac{a+b}{2}
$$

- The variance of $X \sim \mathrm{U}[a, b]$ is

$$
\begin{aligned}
\operatorname{Var}(X) & =\mathbb{E}\left[X^{2}\right]-(\mathbb{E}[X])^{2}=\int_{a}^{b} x^{2} \frac{1}{b-a} \mathrm{~d} x-\left(\frac{a+b}{2}\right)^{2} \\
& =\cdots=\frac{(a-b)^{2}}{12}
\end{aligned}
$$

## Exponential Distribution

A random variable $X$ with probability density function $f$, given by

$$
f(x)=\lambda \mathrm{e}^{-\lambda x}, \quad x \geq 0
$$

is said to have an exponential distribution with parameter $\lambda$.

- We write $X \sim \operatorname{Exp}(\lambda)$.
- The exponential distribution can be viewed as a continuous version of the geometric distribution.



## Properties of Exponential Distribution

- $\mathbb{E}[X]=\frac{1}{\lambda}$.
$\Rightarrow \operatorname{Var}[X]=\frac{1}{\lambda^{2}}$.
- The cdf of $X$ is given by:

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

- The Exponential Distribution is the only continuous distribution that has the memoryless property:

$$
\begin{aligned}
\mathbb{P}(X>s+t \mid X>s) & =\frac{\mathbb{P}(X>s+t, X>s)}{\mathbb{P}(X>s)}=\frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>s)} \\
& =\frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}}=\mathrm{e}^{-\lambda t}=\mathbb{P}(X>t) .
\end{aligned}
$$

## Usage of Exponential Distribution

- Lifetime of an electrical component.
- Time between arrivals of calls at a telephone exchange.
- Time elapsed until a Geiger counter registers a radio-active particle.
- Many more...


## Normal, or Gaussian, Distribution

The normal (or Gaussian) distribution is the most important distribution in the study of statistics, engineering, and biology.

We say that a random variable has a normal distribution with parameters $\mu$ and $\sigma^{2}$ if its density function $f$ is given by

$$
f(x)=\frac{1}{\sigma \sqrt{2 \pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}, \quad x \in \mathbb{R}
$$

- We write $X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right)$.
- The parameters $\mu$ and $\sigma^{2}$ turn out to be the expectation and variance of the distribution, respectively.
- If $\mu=0$ and $\sigma=1$ then

$$
f(x)=\frac{1}{\sqrt{2 \pi}} \mathrm{e}^{-\frac{1}{2} x^{2}}, \quad x \in \mathbb{R}
$$

and the distribution is known as a standard normal distribution.

## Properties of Normal Distribution

- If $X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right)$, then

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

Thus by subtracting the mean and dividing by the standard deviation we obtain a standard normal distribution. This procedure is called standardisation.

- Standardisation enables us to express the cdf of any normal distribution in terms of the cdf of the standard normal distribution.
- A trivial rewriting of the standardisation formula gives the following important result: If $X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right)$, then

$$
X=\mu+\sigma Z, \quad Z \sim \mathrm{~N}(0,1)
$$

- In other words, any Gaussian (normal) random variable can be viewed as a so-called affine (linear + constant) transformation of a standard normal random variable.


## Normal Distribution



## Normal Distribution - Calculation

It is very common to compute $\mathbb{P}(a<X<b)$ for $X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right)$ via standardization as follows.

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

That is,

$$
\mathbb{P}(a<X<b)=F_{X}(b)-F_{X}(a)=F_{Z}\left(\frac{b-\mu}{\sigma}\right)-F_{Z}\left(-\frac{a-\mu}{\sigma}\right)
$$

## Transformation method

- We have seen how we can generate (pseudo) random numbers from a $U[0,1]$ distribution.
- How can we generate random numbers from another distribution? One approach is given by the following result.
- Suppose $F$ is a cdf with inverse $F^{-1}$, such that

$$
F^{-1}(F(x))=x
$$

- Let $U \sim U[0,1]$ and define

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

- Then,

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

and hence $X$ has cdf $F$.

## Transformation method



Thus, we can generate random numbers from the cdf $F$ as follows:

1. Generate $U$ from the uniform random generator.
2. Output $X=F^{-1}(U)$. This is called the transformation method.

## Inverse-Transformation method for exponential r.v.

## Example (Exponential r.v generation)

For the exponential distribution, we have

$$
F(x)=1-\mathrm{e}^{-\lambda x} .
$$

so that, for $y \in(0,1)$,

$$
F^{-1}(y)=-\frac{1}{\lambda} \ln (1-y) .
$$

Hence, output

$$
X:=-\frac{1}{\lambda} \ln (1-U)
$$

We could also output

$$
X:=-\frac{1}{\lambda} \ln (U)
$$

## Exponential distribution

Below a sample of size 30 from an $\operatorname{Exp}(1)$ distribution is plotted.


