## Probability-the Science of Uncertainty and Data

## Probability

## Probability models and axioms

Definition (Sample space) A sample space $\Omega$ is the set of all possible outcomes. The set's elements must be mutually exclusive, collectively exhaustive and at the right granularity.
Definition (Event) An event is a subset of the sample space. Probability is assigned to events.
Definition (Probability axioms) A probability law $\mathbb{P}$ assigns probabilities to events and satisfies the following axioms:
Nonnegativity $\mathbb{P}(A) \geq 0$ for all events $A$.
Normalization $\mathbb{P}(\Omega)=1$.
(Countable) additivity For every sequence of events $A_{1}, A_{2}, \ldots$ such that $A_{i} \cap A_{j}=\varnothing: \mathbb{P}\left(\underset{i}{\bigcup} A_{i}\right)=\sum_{i} \mathbb{P}\left(A_{i}\right)$.
Corollaries (Consequences of the axioms)

- $\mathbb{P}(\varnothing)=0$.
- For any finite collection of disjoint events $A_{1}, \ldots, A_{n}$, $\mathbb{P}\left(\bigcup_{i=1}^{n} A_{i}\right)=\sum_{i=1}^{n} \mathbb{P}\left(A_{i}\right)$.
- $\mathbb{P}(A)+\mathbb{P}\left(A^{c}\right)=1$.
- $\mathbb{P}(A) \leq 1$.
- If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- $\mathbb{P}(A \cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A \cap B)$.
- $\mathbb{P}(A \cup B) \leq \mathbb{P}(A)+\mathbb{P}(B)$.

Example (Discrete uniform law) Assume $\Omega$ is finite and consists of $n$ equally likely elements. Also, assume that $A \subset \Omega$ with $k$ elements. Then $\mathbb{P}(A)=\frac{k}{n}$.

## Conditioning and Bayes' rule

Definition (Conditional probability) Given that event $B$ has occurred and that $P(B)>0$, the probability that $A$ occurs is

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

Remark (Conditional probabilities properties) They are the same as ordinary probabilities. Assuming $\mathbb{P}(B)>0$ :

- $\mathbb{P}(A \mid B) \geq 0$.
- $\mathbb{P}(\Omega \mid B)=1$
- $\mathbb{P}(B \mid B)=1$.
- If $A \cap C=\varnothing, \mathbb{P}(A \cup C \mid B)=\mathbb{P}(A \mid B)+\mathbb{P}(C \mid B)$.

Proposition (Multiplication rule)
$\mathbb{P}\left(A_{1} \cap A_{2} \cap \cdots \cap A_{n}\right)=\mathbb{P}\left(A_{1}\right) \cdot \mathbb{P}\left(A_{2} \mid A_{1}\right) \cdots \mathbb{P}\left(A_{n} \mid A_{1} \cap A_{2} \cap \cdots \cap A_{n-1}\right)$.
Theorem (Total probability theorem) Given a partition $\left\{A_{1}, A_{2}, \ldots\right\}$ of the sample space, meaning that $\cup A_{i}=\Omega$ and the events are disjoint, and for every event $B$, we have

$$
\mathbb{P}(B)=\sum_{i} \mathbb{P}\left(A_{i}\right) \mathbb{P}\left(B \mid A_{i}\right) .
$$

Theorem (Bayes' rule) Given a partition $\left\{A_{1}, A_{2}, \ldots\right\}$ of the sample space, meaning that $\bigcup_{i} A_{i}=\Omega$ and the events are disjoint, and if $\mathbb{P}\left(A_{i}\right)>0$ for all $i$, then for every event $B$, the conditional probabilities $\mathbb{P}\left(A_{i} \mid B\right)$ can be obtained from the conditional probabilities $\mathbb{P}\left(B \mid A_{i}\right)$ and the initial probabilities $\mathbb{P}\left(A_{i}\right)$ as follows:

$$
\mathbb{P}\left(A_{i} \mid B\right)=\frac{\mathbb{P}\left(A_{i}\right) \mathbb{P}\left(B \mid A_{i}\right)}{\sum_{j} \mathbb{P}\left(A_{j}\right) \mathbb{P}\left(B \mid A_{j}\right)}
$$

## Independence

Definition (Independence of events) Two events are independent if occurrence of one provides no information about the other. We say that $A$ and $B$ are independent if

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

Equivalently, as long as $\mathbb{P}(A)>0$ and $\mathbb{P}(B)>0$,

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

## Remarks

- The definition of independence is symmetric with respect to $A$ and $B$.
- The product definition applies even if $\mathbb{P}(A)=0$ or $\mathbb{P}(B)=0$. Corollary If $A$ and $B$ are independent, then $A$ and $B^{c}$ are independent. Similarly for $A^{c}$ and $B$, or for $A^{c}$ and $B^{c}$.
Definition (Conditional independence) We say that $A$ and $B$ are independent conditioned on $C$, where $\mathbb{P}(C)>0$, if

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

Definition (Independence of a collection of events) We say that events $A_{1}, A_{2}, \ldots, A_{n}$ are independent if for every collection of distinct indices $i_{1}, i_{2}, \ldots, i_{k}$, we have

$$
\mathbb{P}\left(A_{i_{1}} \cap \ldots \cap A_{i_{k}}\right)=\mathbb{P}\left(A_{i_{1}}\right) \cdot \mathbb{P}\left(A_{i_{2}}\right) \cdots \mathbb{P}\left(A_{i_{k}}\right)
$$

## Counting

This section deals with finite sets with uniform probability law. In this case, to calculate $\mathbb{P}(A)$, we need to count the number of elements in $A$ and in $\Omega$.
Remark (Basic counting principle) For a selection that can be done in $r$ stages, with $n_{i}$ choices at each stage $i$, the number of possible selections is $n_{1} \cdot n_{2} \cdots n_{r}$.
Definition (Permutations) The number of permutations (orderings) of $n$ different elements is

$$
n!=1 \cdot 2 \cdot 3 \cdots n
$$

Definition (Combinations) Given a set of $n$ elements, the number of subsets with exactly $k$ elements is

$$
\binom{n}{k}=\frac{n!}{k!(n-k)!} .
$$

Definition (Partitions) We are given an $n$-element set and nonnegative integers $n_{1}, n_{2}, \ldots, n_{r}$, whose sum is equal to $n$. The number of partitions of the set into $r$ disjoint subsets, with the $i^{\text {th }}$ subset containing exactly $n_{i}$ elements, is equal to

$$
\binom{n}{n_{1}, \ldots, n_{r}}=\frac{n!}{n_{1}!n_{2}!\cdots n_{r}!}
$$

Remark This is the same as counting how to assign $n$ distinct elements to $r$ people, giving each person $i$ exactly $n_{i}$ elements.

## Discrete random variables

Probability mass function and expectation
Definition (Random variable) A random variable $X$ is a function of the sample space $\Omega$ into the real numbers (or $\mathbb{R}^{n}$ ). Its range can be discrete or continuous
Definition (Probability mass funtion (PMF)) The probability law of a discrete random variable $X$ is called its PMF. It is defined as

$$
p_{X}(x)=\mathbb{P}(X=x)=\mathbb{P}(\{\omega \in \Omega: X(\omega)=x\})
$$

Properties
$p_{X}(x) \geq 0, \forall x$.
$\sum_{x} p_{X}(x)=1$.
Example (Bernoulli random variable) A Bernoulli random variable $X$ with parameter $0 \leq p \leq 1(X \sim \operatorname{Ber}(p))$ takes the following values:

$$
X= \begin{cases}1 & \text { w.p. } p \\ 0 & \text { w.p. } 1-p .\end{cases}
$$

An indicator random variable of an event ( $I_{A}=1$ if $A$ occurs) is an example of a Bernoulli random variable.
Example (Discrete uniform random variable) A Discrete uniform random variable $X$ between $a$ and $b$ with $a \leq b(X \sim \operatorname{Uni}[a, b])$ takes any of the values in $\{a, a+1, \ldots, b\}$ with probability $\frac{1}{b-a+1}$. Example (Binomial random variable) A Binomial random variable $X$ with parameters $n$ (natural number) and $0 \leq p \leq 1$ $(X \sim \operatorname{Bin}(n, p))$ takes values in the set $\{0,1, \ldots, n\}$ with probabilities $p_{X}(i)=\binom{n}{i} p^{i}(1-p)^{n-i}$.
It represents the number of successes in $n$ independent trials where each trial has a probability of success $p$. Therefore, it can also be seen as the sum of $n$ independent Bernoulli random variables, each with parameter $p$.
Example (Geometric random variable) A Geometric random variable $X$ with parameter $0 \leq p \leq 1(X \sim \operatorname{Geo}(p))$ takes values in the set $\{1,2, \ldots\}$ with probabilities $p_{X}(i)=(1-p)^{i-1} p$.
It represents the number of independent trials until (and including) the first success, when the probability of success in each trial is $p$. Definition (Expectation/mean of a random variable) The expectation of a discrete random variable is defined as

$$
\mathbb{E}[X] \triangleq \sum_{x} x p_{X}(x)
$$

assuming $\sum_{x}|x| p_{X}(x)<\infty$.
Properties (Properties of expectation)

- If $X \geq 0$ then $\mathbb{E}[X] \geq 0$.
- If $a \leq X \leq b$ then $a \leq \mathbb{E}[X] \leq b$.
- If $X=c$ then $\mathbb{E}[X]=c$.

Example Expected value of know r.v.

- If $X \sim \operatorname{Ber}(p)$ then $\mathbb{E}[X]=p$.
- If $X=I_{A}$ then $\mathbb{E}[X]=\mathbb{P}(A)$.
- If $X \sim \operatorname{Uni}[a, b]$ then $\mathbb{E}[X]=\frac{a+b}{2}$.
- If $X \sim \operatorname{Bin}(n, p)$ then $\mathbb{E}[X]=n p$.
- If $X \sim \operatorname{Geo}(p)$ then $\mathbb{E}[X]=\frac{1}{p}$.

Theorem (Expected value rule) Given a random variable $X$ and a Properties (Properties of joint PMF)
function $g: \mathbb{R} \rightarrow \mathbb{R}$, we construct the random variable $Y=g(X)$.
Then

$$
\sum_{y} y p_{Y}(y)=\mathbb{E}[Y]=\mathbb{E}[g(X)]=\sum_{x} g(x) p_{X}(x)
$$

Remark (PMF of $Y=g(X))$ The PMF of $Y=g(X)$ is $p_{Y}(y)=\sum_{x: g(x)=y} p_{X}(x)$.
Remark In general $g(\mathbb{E}[X]) \neq \mathbb{E}[g(X)]$. They are equal if $g(x)=a x+b$.
Variance, conditioning on an event, multiple r.v.
Definition (Variance of a random variable) Given a random variable $X$ with $\mu=\mathbb{E}[X]$, its variance is a measure of the spread of the random variable and is defined as

$$
\operatorname{Var}(X) \triangleq \mathbb{E}\left[(X-\mu)^{2}\right]=\sum_{x}(x-\mu)^{2} p_{X}(x)
$$

Definition (Standard deviation)

$$
\sigma_{X}=\sqrt{\operatorname{Var}(X)}
$$

Properties (Properties of the variance)

- $\operatorname{Var}(a X)=a^{2} \operatorname{Var}(X)$, for all $a \in \mathbb{R}$.
- $\operatorname{Var}(X+b)=\operatorname{Var}(X)$, for all $b \in \mathbb{R}$.
- $\operatorname{Var}(a X+b)=a^{2} \operatorname{Var}(X)$.
- $\operatorname{Var}(X)=\mathbb{E}\left[X^{2}\right]-(\mathbb{E}[X])^{2}$.

Example (Variance of known r.v.)

- If $X \sim \operatorname{Ber}(p)$, then $\operatorname{Var}(X)=p(1-p)$.
- If $X \sim \operatorname{Uni}[a, b]$, then $\operatorname{Var}(X)=\frac{(b-a)(b-a+2)}{12}$.
- If $X \sim \operatorname{Bin}(n, p)$, then $\operatorname{Var}(X)=n p(1-p)$.
- If $X \sim \operatorname{Geo}(p)$, then $\operatorname{Var}(X)=\frac{1-p}{p^{2}}$

Proposition (Conditional PMF and expectation, given an event) Given the event $A$, with $\mathbb{P}(A)>0$, we have the following

- $p_{X \mid A}(x)=\mathbb{P}(X=x \mid A)$.
- If $A$ is a subset of the range of $X$, then:

$$
p_{X \mid A}(x) \triangleq p_{X \mid\{X \in A\}}(x)= \begin{cases}\frac{1}{\mathrm{P}(A)} p_{X}(x), & \text { if } x \in A \\ 0, & \text { otherwise }\end{cases}
$$

- $\sum_{x} p_{X \mid A}(x)=1$.
- $\mathbb{E}[X \mid A]=\sum_{x} x p_{X \mid A}(x)$.
- $\mathbb{E}[g(X) \mid A]=\sum_{x} g(x) p_{X \mid A}(x)$.

Proposition (Total expectation rule) Given a partition of disjoint events $A_{1}, \ldots, A_{n}$ such that $\sum_{i} \mathbb{P}\left(A_{i}\right)=1$, and $\mathbb{P}\left(A_{i}\right)>0$,

$$
\mathbb{E}[X]=\mathbb{P}\left(A_{1}\right) \mathbb{E}\left[X \mid A_{1}\right]+\cdots+\mathbb{P}\left(A_{n}\right) \mathbb{E}\left[X \mid A_{n}\right]
$$

Definition (Memorylessness of the geometric random variable) When we condition a geometric random variable $X$ on the event $X>n$ we have memorylessness, meaning that the "remaining time" $X-n$, given that $X>n$, is also geometric with the same parameter. Formally,

$$
p_{X-n \mid X>n}(i)=p_{X}(i)
$$

Definition (Joint PMF) The joint PMF of random variables $X_{1}, X_{2}, \ldots, X_{n}$ is $p_{X_{1}, X_{2}, \ldots, X_{n}}\left(x_{1}, \ldots, x_{n}\right)=\mathbb{P}\left(X_{1}=x_{1}, \ldots, X_{n}=x_{n}\right)$.

- $\sum_{x_{1}} \cdots \sum_{x_{n}} p_{X_{1}, \ldots, X_{n}}\left(x_{1}, \ldots, x_{n}\right)=1$.
- $p_{X_{1}}\left(x_{1}\right)=\sum_{x_{2}} \cdots \sum_{x_{n}} p_{X_{1}, \ldots, X_{n}}\left(x_{1}, x_{2}, \ldots, x_{n}\right)$.
- $p_{X_{2}, \ldots, X_{n}}\left(x_{2}, \ldots, x_{n}\right)=\sum_{x_{1}} p_{X_{1}, X_{2}, \ldots, X_{n}}\left(x_{1}, x_{2}, \ldots, x_{n}\right)$.

Definition (Functions of multiple r.v.) If $Z=g\left(X_{1}, \ldots, X_{n}\right)$, where $g: \mathbb{R}^{n} \rightarrow \mathbb{R}$, then $p_{Z}(z)=\mathbb{P}\left(g\left(X_{1}, \ldots, X_{n}\right)=z\right)$.
Proposition (Expected value rule for multiple r.v.) Given $g: \mathbb{R}^{n} \rightarrow \mathbb{R}$,
$\mathbb{E}\left[g\left(X_{1}, \ldots, X_{n}\right)\right]=\sum_{x_{1}, \ldots, x_{n}} g\left(x_{1}, \ldots, x_{n}\right) p_{X_{1}, \ldots, X_{n}}\left(x_{1}, \ldots, x_{n}\right)$.
Properties (Linearity of expectations)

- $\mathbb{E}[a X+b]=a \mathbb{E}[X]+b$.
- $\mathbb{E}\left[X_{1}+\cdots+X_{n}\right]=\mathbb{E}\left[X_{1}\right]+\cdots+\mathbb{E}\left[X_{n}\right]$.


## Conditioning on a random variable, independence

Definition (Conditional PMF given another random variable)
Given discrete random variables $X, Y$ and $y$ such that $p_{Y}(y)>0$ we define

$$
p_{X \mid Y}(x \mid y) \triangleq \frac{p_{X, Y}(x, y)}{p_{Y}(y)}
$$

Proposition (Multiplication rule) Given jointly discrete random variables $X, Y$, and whenever the conditional probabilities are defined,

$$
p_{X, Y}(x, y)=p_{X}(x) p_{Y \mid X}(y \mid x)=p_{Y}(y) p_{X \mid Y}(x \mid y)
$$

Definition (Conditional expectation) Given discrete random variables $X, Y$ and $y$ such that $p_{Y}(y)>0$ we define

$$
\mathbb{E}[X \mid Y=y]=\sum_{x} x p_{X \mid Y}(x \mid y)
$$

Additionally we have

$$
\mathbb{E}[g(X) \mid Y=y]=\sum_{x} g(x) p_{X \mid Y}(x \mid y)
$$

Theorem (Total probability and expectation theorems) If $p_{Y}(y)>0$, then

$$
\begin{aligned}
& p_{X}(x)=\sum_{y} p_{Y}(y) p_{X \mid Y}(x \mid y) \\
& \mathbb{E}[X]=\sum_{y} p_{Y}(y) \mathbb{E}[X \mid Y=y]
\end{aligned}
$$

Definition (Independence of a random variable and an event) A discrete random variable $X$ and an event $A$ are independent if $\mathbb{P}(X=x$ and $A)=p_{X}(x) \mathbb{P}(A)$, for all $x$.
Definition (Independence of two random variables) Two discrete random variables $X$ and $Y$ are independent if $p_{X, Y}(x, y)=p_{X}(x) p_{Y}(y)$ for all $x, y$.
Remark (Independence of a collection of random variables) A collection $X_{1}, X_{2}, \ldots, X_{n}$ of random variables are independent if

$$
p_{X_{1}, \ldots, X_{n}}\left(x_{1}, \ldots, x_{n}\right)=p_{X_{1}}\left(x_{1}\right) \cdots p_{X_{n}}\left(x_{n}\right), \forall x_{1}, \ldots, x_{n}
$$

Remark (Independence and expectation) In general,
$\mathbb{E}[g(X, Y)] \neq g(\mathbb{E}[X], \mathbb{E}[Y])$. An exception is for linear functions: $\mathbb{E}[a X+b Y]=a \mathbb{E}[X]+b \mathbb{E}[Y]$.

Proposition (Expectation of product of independent r.v.) If $X$ and $Y$ are discrete independent random variables,

$$
\mathbb{E}[X Y]=\mathbb{E}[X] \mathbb{E}[Y]
$$

Remark If $X$ and $Y$ are independent,
$\mathbb{E}[g(X) h(Y)]=\mathbb{E}[g(X)] \mathbb{E}[h(Y)]$.
Proposition (Variance of sum of independent random variables) IF $X$ and $Y$ are discrete independent random variables,

$$
\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)
$$

## Continuous random variables

PDF, Expectation, Variance, $C D F$
Definition (Probability density function (PDF)) A probability density function of a r.v. $X$ is a non-negative real valued function $f_{X}$ that satisfies the following

- $\int_{-\infty}^{\infty} f_{X}(x) \mathrm{d} x=1$.
- $\mathbb{P}(a \leq X \leq b)=\int_{a}^{b} f_{X}(x) \mathrm{d} x$ for some random variable $X$. Definition (Continuous random variable) A random variable $X$ is continuous if its probability law can be described by a PDF $f_{X}$. Remark Continuous random variables satisfy:
- For small $\delta>0, \mathbb{P}(a \leq X \leq a+\delta) \approx f_{X}(a) \delta$.
- $\mathbb{P}(X=a)=0, \forall a \in \mathbb{R}$.

Definition (Expectation of a continuous random variable) The expectation of a continuous random variable is

$$
\mathbb{E}[X] \triangleq \int_{-\infty}^{\infty} x f_{X}(x) \mathrm{d} x
$$

assuming $\int_{-\infty}^{\infty}|x| f_{X}(x) \mathrm{d} x<\infty$.
Properties (Properties of expectation)

- If $X \geq 0$ then $\mathbb{E}[X] \geq 0$.
- If $a \leq X \leq b$ then $a \leq \mathbb{E}[X] \leq b$.
- $\mathbb{E}[g(X)]=\int_{-\infty}^{\infty} g(x) f_{X}(x) \mathrm{d} x$.
- $\mathbb{E}[a X+b]=a \mathbb{E}[X]+b$.

Definition (Variance of a continuous random variable) Given a continuous random variable $X$ with $\mu=\mathbb{E}[X]$, its variance is

$$
\operatorname{Var}(X)=\mathbb{E}\left[(X-\mu)^{2}\right]=\int_{-\infty}^{\infty}(x-\mu)^{2} f_{X}(x) \mathrm{d} x
$$

It has the same properties as the variance of a discrete random variable.
Example (Uniform continuous random variable) A Uniform continuous random variable $X$ between $a$ and $b$, with $a<b$, ( $X \sim \operatorname{Uni}(a, b)$ ) has PDF

$$
f_{X}(x)= \begin{cases}\frac{1}{b-a}, & \text { if } a<x<b \\ 0, & \text { otherwise }\end{cases}
$$

We have $\mathbb{E}[X]=\frac{a+b}{2}$ and $\operatorname{Var}(X)=\frac{(b-a)^{2}}{12}$.

Example (Exponential random variable) An Exponential random variable $X$ with parameter $\lambda>0(X \sim \operatorname{Exp}(\lambda))$ has PDF

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

We have $E[X]=\frac{1}{\lambda}$ and $\operatorname{Var}(X)=\frac{1}{\lambda^{2}}$.
Definition (Cumulative Distribution Function (CDF)) The CDF of a random variable $X$ is $F_{X}(x)=\mathbb{P}(X \leq x)$.
In particular, for a continuous random variable, we have

$$
\begin{aligned}
F_{X}(x) & =\int_{-\infty}^{x} f_{X}(x) \mathrm{d} x \\
f_{X}(x) & =\frac{\mathrm{d} F_{X}(x)}{\mathrm{d} x}
\end{aligned}
$$

Properties (Properties of CDF)

- If $y \geq x$, then $F_{X}(y) \geq F_{X}(x)$.
- $\lim _{x \rightarrow-\infty} F_{X}(x)=0$.
- $\lim _{x \rightarrow \infty} F_{X}(x)=1$.

Definition (Normal/Gaussian random variable) A Normal random variable $X$ with mean $\mu$ and variance $\sigma^{2}>0\left(X \sim \mathcal{N}\left(\mu, \sigma^{2}\right)\right)$ has PDF

$$
f_{X}(x)=\frac{1}{\sqrt{2 \pi \sigma^{2}}} \mathrm{e}^{-\frac{1}{2 \sigma^{2}}(x-\mu)^{2}} .
$$

We have $E[X]=\mu$ and $\operatorname{Var}(X)=\sigma^{2}$.
Remark (Standard Normal) The standard Normal is $\mathcal{N}(0,1)$.
Proposition (Linearity of Gaussians) Given $X \sim \mathcal{N}\left(\mu, \sigma^{2}\right)$, and if $a \neq 0$, then $a X+b \sim \mathcal{N}\left(a \mu+b, a^{2} \sigma^{2}\right)$.
Using this $Y=\frac{X-\mu}{\sigma}$ is a standard gaussian.
Conditioning on an event, and multiple continuous r.v.
Definition (Conditional PDF given an event) Given a continuous random variable $X$ and event $A$ with $P(A)>0$, we define the conditional PDF as the function that satisfies

$$
\mathbb{P}(X \in B \mid A)=\int_{B} f_{X \mid A}(x) \mathrm{d} x
$$

Definition (Conditional PDF given $X \in A$ ) Given a continuous random variable $X$ and an $A \subset \mathbb{R}$, with $P(A)>0$ :

$$
f_{X \mid X \in A}(x)= \begin{cases}\frac{1}{\mathrm{P}(A)} f_{X}(x), & x \in A, \\ 0, & x \notin A .\end{cases}
$$

Definition (Conditional expectation) Given a continuous random variable $X$ and an event $A$, with $P(A)>0$ :

$$
\mathbb{E}[X \mid A]=\int_{-\infty}^{\infty} f_{X \mid A}(x) \mathrm{d} x
$$

Definition (Memorylessness of the exponential random variable) When we condition an exponential random variable $X$ on the event $X>t$ we have memorylessness, meaning that the "remaining time" $X-t$ given that $X>t$ is also geometric with the same parameter i.e.,

$$
\mathbb{P}(X-t>x \mid X>t)=\mathbb{P}(X>x)
$$

Theorem (Total probability and expectation theorems) Given a partition of the space into disjoint events $A_{1}, A_{2}, \ldots, A_{n}$ such that $\sum_{i} \mathbb{P}\left(A_{i}\right)=1$ we have the following:

$$
\begin{aligned}
F_{X}(x) & =\mathbb{P}\left(A_{1}\right) F_{X \mid A_{1}}(x)+\cdots+\mathbb{P}\left(A_{n}\right) F_{X \mid A_{n}}(x), \\
f_{X}(x) & =\mathbb{P}\left(A_{1}\right) f_{X \mid A_{1}}(x)+\cdots+\mathbb{P}\left(A_{n}\right) f_{X \mid A_{n}}(x), \\
\mathbb{E}[X] & =\mathbb{P}\left(A_{1}\right) \mathbb{E}\left[X \mid A_{1}\right]+\cdots+\mathbb{P}\left(A_{n}\right) \mathbb{E}\left[X \mid A_{n}\right] .
\end{aligned}
$$

Definition (Jointly continuous random variables) A pair
(collection) of random variables is jointly continuous if there exists a joint PDF $f_{X, Y}$ that describes them, that is, for every set $B \subset \mathbb{R}^{n}$

$$
\mathbb{P}((X, Y) \in B)=\iint_{B} f_{X, Y}(x, y) \mathrm{d} x \mathrm{~d} y
$$

Properties (Properties of joint PDFs)

- $f_{X}(x)=\int_{-\infty}^{\infty} f_{X, Y}(x, y) \mathrm{d} y$.
- $F_{X, Y}(x, y)=\mathbb{P}(X \leq x, Y \leq y)=\int_{-\infty}^{x}\left[\int_{-\infty}^{y} f_{X, Y}(u, v) \mathrm{d} v\right] \mathrm{d} u$.
- $f_{X, Y}(x)=\frac{\partial^{2} F_{X, Y}(x, y)}{\partial x \partial y}$.

Example (Uniform joint PDF on a set $S$ ) Let $S \subset \mathbb{R}^{2}$ with area $s>0$, then the random variable $(X, Y)$ is uniform over $S$ if it has PDF

$$
f_{X, Y}(x, y)= \begin{cases}\frac{1}{s}, & (x, y) \in S \\ 0, & (x, y) \notin S\end{cases}
$$

Conditioning on a random variable, independence, Bayes' rule Definition (Conditional PDF given another random variable) Given jointly continuous random variables $X, Y$ and a value $y$ such that $f_{Y}(y)>0$, we define the conditional PDF as

$$
f_{X \mid Y}(x \mid y) \triangleq \frac{f_{X, Y}(x, y)}{f_{Y}(y)}
$$

Additionally we define $\mathbb{P}(X \in A \mid Y=y) \int_{A} f_{X \mid Y}(x \mid y) \mathrm{d} x$ Proposition (Multiplication rule) Given jointly continuous random variables $X, Y$, whenever possible we have

$$
f_{X, Y}(x, y)=f_{X}(x) f_{Y \mid X}(y \mid x)=f_{Y}(y) f_{X \mid Y}(x \mid y)
$$

Definition (Conditional expectation) Given jointly continuous random variables $X, Y$, and $y$ such that $f_{Y}(y)>0$, we define the conditional expected value as

$$
\mathbb{E}[X \mid Y=y]=\int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y) \mathrm{d} x
$$

Additionally we have

$$
\mathbb{E}[g(X) \mid Y=y]=\int_{-\infty}^{\infty} g(x) f_{X \mid Y}(x \mid y) \mathrm{d} x .
$$

Theorem (Total probability and total expectation theorems)

$$
\begin{aligned}
& f_{X}(x)=\int_{-\infty}^{\infty} f_{Y}(y) f_{X \mid Y}(x \mid y) \mathrm{d} y \\
& \mathbb{E}[X]=\int_{-\infty}^{\infty} f_{Y}(y) \mathbb{E}[X \mid Y=y] \mathrm{d} y
\end{aligned}
$$

## Definition (Independence) Jointly continuous random variables

 $X, Y$ are independent if $f_{X, Y}(x, y)=f_{X}(x) f_{Y}(y)$ for all $x, y$.Proposition (Expectation of product of independent r.v.) If $X$ and $Y$ are independent continuous random variables,

## $\mathbb{E}[X Y]=\mathbb{E}[X] \mathbb{E}[Y]$

Remark If $X$ and $Y$ are independent,
$\mathbb{E}[g(X) h(Y)]=\mathbb{E}[g(X)] \mathbb{E}[h(Y)]$.
Proposition (Variance of sum of independent random variables) If $X$ and $Y$ are independent continuous random variables,

$$
\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)
$$

Proposition (Bayes' rule summary)

- For $X, Y$ discrete: $p_{X \mid Y}(x \mid y)=\frac{p_{X}(x) p_{Y \mid X}(y \mid x)}{p_{Y}(y)}$.
- For $X, Y$ continuous: $f_{X \mid Y}(x \mid y)=\frac{f_{X}(x) f_{Y \mid X}(y \mid x)}{f_{Y}(y)}$.
- For $X$ discrete, $Y$ continuous: $p_{X \mid Y}(x \mid y)=\frac{p_{X}(x) f_{Y \mid X}(y \mid x)}{f_{Y}(y)}$.
- For $X$ continuous, $Y$ discrete: $f_{X \mid Y}(x \mid y)=\frac{f_{X}(x) p_{Y \mid X}(y \mid x)}{p_{Y}(y)}$.


## Derived distributions

Proposition (Discrete case) Given a discrete random variable $X$ and a function $g$, the r.v. $Y=g(X)$ has PMF

$$
p_{Y}(y)=\sum_{x: g(x)=y} p_{X}(x) .
$$

Remark (Linear function of discrete random variable) If $g(x)=a x+b$, then $p_{Y}(y)=p_{X}\left(\frac{y-b}{a}\right)$.

Proposition (Linear function of continuous r.v.) Given a continuous random variable $X$ and $Y=a X+b$, with $a \neq 0$, we have

$$
f_{Y}(y)=\frac{1}{|a|} f_{X}\left(\frac{y-b}{a}\right) .
$$

Corollary (Linear function of normal r.v.) If $X \sim \mathcal{N}\left(\mu, \sigma^{2}\right)$ and $Y=a X+b$, with $a \neq 0$, then $Y \sim \mathcal{N}\left(a \mu+b, a^{2} \sigma^{2}\right)$.

Example (General function of a continuous r.v.) If $X$ is a continuous random variable and $g$ is any function, to obtain the pdf of $Y=g(X)$ we follow the two-step procedure:

1. Find the CDF of $Y: F_{Y}(y)=\mathbb{P}(Y \leq y)=\mathbb{P}(g(X) \leq y)$.
2. Differentiate the CDF of $Y$ to obtain the PDF:

$$
f_{Y}(y)=\frac{\mathrm{d} F_{Y}(y)}{\mathrm{d} y}
$$

Proposition (General formula for monotonic $g$ ) Let $X$ be a continuous random variable and $g$ a function that is monotonic wherever $f_{X}(x)>0$. The PDF of $Y=g(X)$ is given by

$$
f_{Y}(y)=f_{X}(h(y))\left|\frac{\mathrm{d} h}{\mathrm{~d} y}(y)\right| .
$$

where $h=g^{-1}$ in the interval where g is monotonic.

## Sums of independent r.v., covariance and correlation

Proposition (Discrete case) Let $X, Y$ be discrete independent random variables and $Z=X+Y$, then the PMF of $Z$ is

$$
p_{Z}(z)=\sum_{x} p_{X}(x) p_{Y}(z-x) .
$$

Proposition (Continuous case) Let $X, Y$ be continuous independent random variables and $Z=X+Y$, then the PDF of $Z$ is

$$
f_{Z}(z)=\int_{-\infty}^{\infty} f_{X}(x) f_{Y}(z-x) \mathrm{d} x
$$

Proposition (Sum of independent normal r.v.) Let $X \sim \mathcal{N}\left(\mu_{x}, \sigma_{x}^{2}\right)$ and $Y \sim \mathcal{N}\left(\mu_{y}, \sigma_{y}^{2}\right)$ independent. Then $Z=X+Y \sim \mathcal{N}\left(\mu_{x}+\mu_{y}, \sigma_{x}^{2}+\sigma_{y}^{2}\right)$.
Definition (Covariance) We define the covariance of random variables $X, Y$ as

$$
\operatorname{Cov}(X, Y) \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]
$$

Properties (Properties of covariance)

- If $X, Y$ are independent, then $\operatorname{Cov}(X, Y)=0$.
- $\operatorname{Cov}(X, X)=\operatorname{Var}(X)$.
- $\operatorname{Cov}(a X+b, Y)=a \operatorname{Cov}(X, Y)$.
- $\operatorname{Cov}(X, Y+Z)=\operatorname{Cov}(X, Y)+\operatorname{Cov}(X, Z)$.
- $\operatorname{Cov}(X, Y)=\mathbb{E}[X Y]-\mathbb{E}[X] \mathbb{E}[Y]$.

Proposition (Variance of a sum of r.v.)

$$
\operatorname{Var}\left(X_{1}+\cdots+X_{n}\right)=\sum_{i} \operatorname{Var}\left(X_{i}\right)+\sum_{i \neq j} \operatorname{Cov}\left(X_{i}, X_{j}\right) .
$$

Definition (Correlation coefficient) We define the correlation coefficient of random variables $X, Y$, with $\sigma_{X}, \sigma_{Y}>0$, as

$$
\rho(X, Y) \triangleq \frac{\operatorname{Cov}(X, Y)}{\sigma_{X} \sigma_{Y}}
$$

Properties (Properties of the correlation coefficient)

- $-1 \leq \rho \leq 1$.
- If $X, Y$ are independent, then $\rho=0$.
- $|\rho|=1$ if and only if $X-\mathbb{E}[X]=c(Y-\mathbb{E}[Y])$.
- $\rho(a X+b, Y)=\operatorname{sign}(a) \rho(X, Y)$.


## Conditional expectation and variance, sum of

 random number of r.v.Definition (Conditional expectation as a random variable) Given random variables $X, Y$ the conditional expectation $\mathbb{E}[X \mid Y]$ is the random variable that takes the value $\mathbb{E}[X \mid Y=y]$ whenever $Y=y$. Theorem (Law of iterated expectations)

$$
\mathbb{E}[\mathbb{E}[X \mid Y]]=\mathbb{E}[X]
$$

Definition (Conditional variance as a random variable) Given random variables $X, Y$ the conditional variance $\operatorname{Var}(X \mid Y)$ is the random variable that takes the value $\operatorname{Var}(X \mid Y=y)$ whenever $Y=y$.
Theorem (Law of total variance)

$$
\operatorname{Var}(X)=\mathbb{E}[\operatorname{Var}(X \mid Y)]+\operatorname{Var}(\mathbb{E}[X \mid Y])
$$

Proposition (Sum of a random number of independent r.v.) Let $N$ be a nonnegative integer random variable. Let $X, X_{1}, X_{2}, \ldots, X_{N}$ be i.i.d. random variables. Let $Y=\sum_{i} X_{i}$. Then

$$
\mathbb{E}[Y]=\mathbb{E}[N] \mathbb{E}[X],
$$

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

## Convergence of Random variables

## Inequalities, convergence, and the Weak Law of

## Large Numbers

Theorem (Markov inequality) Given a random variable $X \geq 0$ and, for every $a>0$ we have

$$
\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}
$$

Theorem (Chebyshev inequality) Given a random variable $X$ with $\mathbb{E}[X]=\mu$ and $\operatorname{Var}(X)=\sigma^{2}$, for every $\epsilon>0$ we have

$$
\mathbb{P}(|X-\mu| \geq \epsilon) \leq \frac{\sigma^{2}}{\epsilon^{2}}
$$

Theorem (Weak Law of Large Number (WLLN)) Given a sequence of i.i.d. random variables $\left\{X_{1}, X_{2}, \ldots\right\}$ with $\mathbb{E}\left[X_{i}\right]=\mu$ and $\operatorname{Var}\left(X_{i}\right)=\sigma^{2}$, we define

$$
M_{n}=\frac{1}{n} \sum_{i=1}^{n} X_{i}
$$

for every $\epsilon>0$ we have

$$
\lim _{n \rightarrow \infty} \mathbb{P}\left(\left|M_{n}-\mu\right| \geq \epsilon\right)=0
$$

Definition (Convergence in probability) A sequence of random variables $\left\{Y_{i}\right\}$ converges in probability to the random variable $Y$ if

$$
\lim _{n \rightarrow \infty} \mathbb{P}\left(\left|Y_{i}-Y\right| \geq \epsilon\right)=0
$$

for every $\epsilon>0$.
Properties (Properties of convergence in probability) If $X_{n} \rightarrow a$ and $Y_{n} \rightarrow b$ in probability, then

- $X_{n}+Y_{n} \rightarrow a+b$.
- If $g$ is a continuous function, then $g\left(X_{n}\right) \rightarrow g(a)$.
- $\mathbb{E}\left[X_{n}\right]$ does not always converge to $a$.


## The Central Limit Theorem

Theorem (Central Limit Theorem (CLT)) Given a sequence of independent random variables $\left\{X_{1}, X_{2}, \ldots\right\}$ with $\mathbb{E}\left[X_{i}\right]=\mu$ and $\operatorname{Var}\left(X_{i}\right)=\sigma^{2}$, we define

$$
Z_{n}=\frac{1}{\sigma \sqrt{n}} \sum_{i=1}^{n}\left(X_{i}-\mu\right)
$$

Then, for every $z$, we have

$$
\lim _{n \rightarrow \infty} \mathbb{P}\left(Z_{n} \leq z\right)=\mathbb{P}(Z \leq z)
$$

where $Z \sim \mathcal{N}(0,1)$.
Corollary (Normal approximation of a binomial) Let $X \sim \operatorname{Bin}(n, p)$ with $n$ large. Then $S_{n}$ can be approximated by $Z \sim \mathcal{N}(n p, n p(1-p))$.
Remark (De Moivre-Laplace $1 / 2$ approximation) Let $X \sim$ Bin, then $\mathbb{P}(X=i)=\mathbb{P}\left(i-\frac{1}{2} \leq X \leq i+\frac{1}{2}\right)$ and we can use the CLT to approximate the PMF of $X$.

