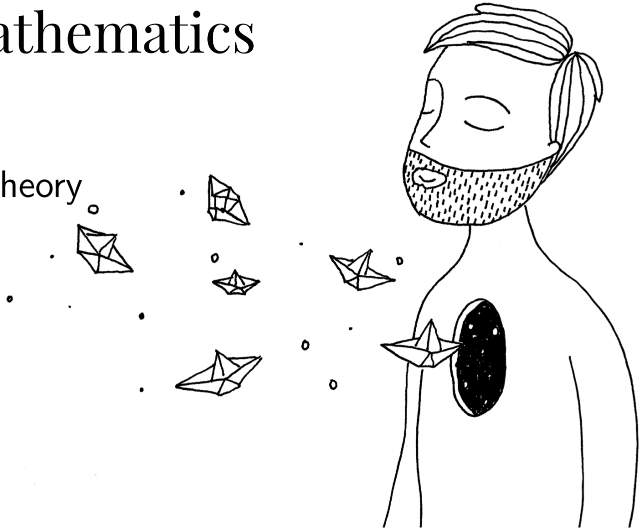
4509 - Bridging Mathematics

Introduction to Probability Theory



PAULO FAGANDINI

NOVA SCHOOL OF BUSINESS & ECONOMICS

Figure: $\Omega :=$ Universe possible outcomes.

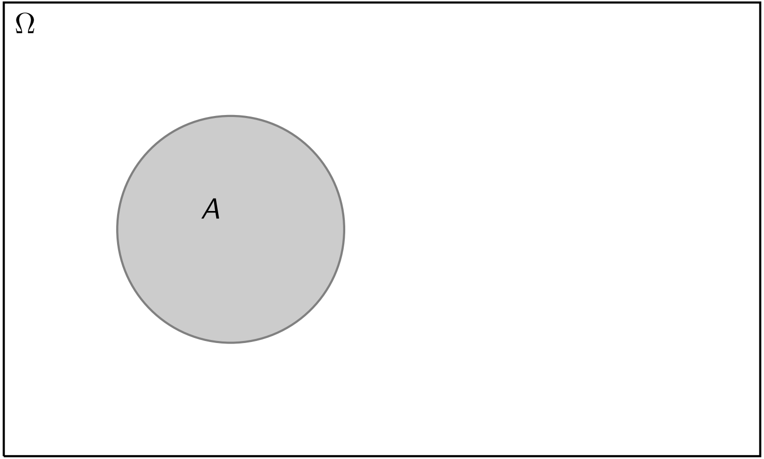


Figure: $A :=$ Event, set of possible outcomes. $A \subseteq \Omega$.

Figure: Outcomes not in A.

So this looks a lot like set theory, and it is.

So this looks a lot like set theory, and it is.

To know how likely is that an event in Ω occurs, we need to have a *measure* of A. That's why **measure theory** is so important in probability.

So this looks a lot like set theory, and it is.

To know how likely is that an event in Ω occurs, we need to have a *measure* of A. That's why **measure theory** is so important in probability.

Call the function $P : \mathcal{P}(\Omega) \to \mathbb{R}$ a measure. For P to be a probability three things must hold:

1. $P(\Omega) = 1$.
2. $P(A \subseteq \Omega) \geq 0$
3. If, for $A, B \subseteq \Omega$, $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

These three conditions are known as the **Kolmogorov Axioms.**

So this looks a lot like set theory, and it is.

To know how likely is that an event in Ω occurs, we need to have a *measure* of A.
That's why **measure theory** is so important in probability.

Call the function $P : \mathcal{P}(\Omega) \to \mathbb{R}$ a measure. For P to be a probability three things must hold:

1. $P(\Omega) = 1$.
2. $P(A \subseteq \Omega) \geq 0$
3. If, for $A, B \subseteq \Omega$, $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

These three conditions are known as the **Kolmogorov Axioms.**
We will not cover measure theory here, just use it in a somehow intuitive way.

Definition

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a measure space with $P(\Omega) = 1$. Then $(\Omega, \mathcal{P}(\Omega), P)$ is called a **probability space** with sample space Ω, event space $\mathcal{P}(\Omega)$, and probability measure P.

Conjecture

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space satisfying the Kolmogorov Axioms, then:

1. $P(\emptyset) = 0$.
2. *If* $A \subseteq B$, *then* $P(A) \leq P(B)$.
3. $P(A) \in [0, 1]$, $\forall A \in \mathcal{P}(\Omega)$.

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

1. A and A^c are disjoint.

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

1. A and A^c are disjoint.
2. $P(A) + P(A^c) = P(A \cup A^c)$.

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

1. A and A^c are disjoint.
2. $P(A) + P(A^c) = P(A \cup A^c)$.
3. But $A \cup A^c = \Omega$, so $P(A) + P(A^c) = P(\Omega)$

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

1. A and A^c are disjoint.
2. $P(A) + P(A^c) = P(A \cup A^c)$.
3. But $A \cup A^c = \Omega$, so $P(A) + P(A^c) = P(\Omega)$
4. But we know that $P(\Omega) = 1$, so...

From the example at the start of the class, let's show that $P(A^c) = 1 - P(A)$.

1. A and A^c are disjoint.
2. $P(A) + P(A^c) = P(A \cup A^c)$.
3. But $A \cup A^c = \Omega$, so $P(A) + P(A^c) = P(\Omega)$
4. But we know that $P(\Omega) = 1$, so...
5. $P(A) + P(A^c) = 1 \quad \Rightarrow \quad P(A^c) = 1 - P(A)$

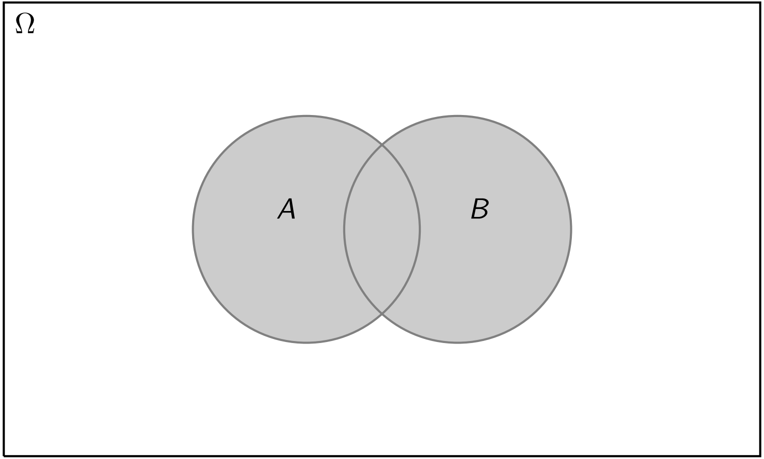


Figure: $A \cup B :=$ Outcomes in A or B happening.

What's the probability of $A \cup B$ when A and B are not disjoint? What's the problem with $P(A) + P(B)$?

What's the probability of $A \cup B$ when A and B are not disjoint? What's the problem with $P(A) + P(B)$?

$P(A \cap B)$ is being counted twice!, so

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

What's the probability of $A \cup B$ when A and B are not disjoint? What's the problem with $P(A) + P(B)$?

$P(A \cap B)$ is being counted twice!, so

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

What happens with the intersection?

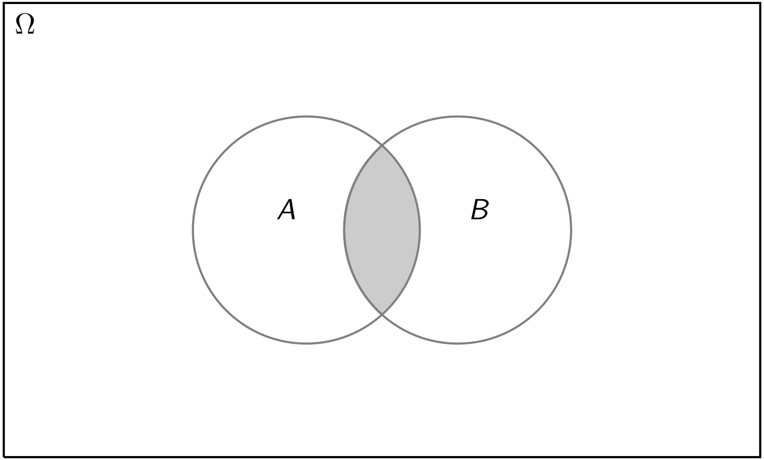


Figure: $A \cap B :=$ Outcomes that belong in A and B.

An interesting question, what's the probability that an outcome in B is also an outcome that could happen in A?

An interesting question, what's the probability that an outcome in B is also an outcome that could happen in A?

This is what "conditional probability" is all about. Given that we are looking only at elements in B, our universe is B and not Ω, so the probability will be $P(A \cap B)$ divided by the size of B, that is, $P(B)$.

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

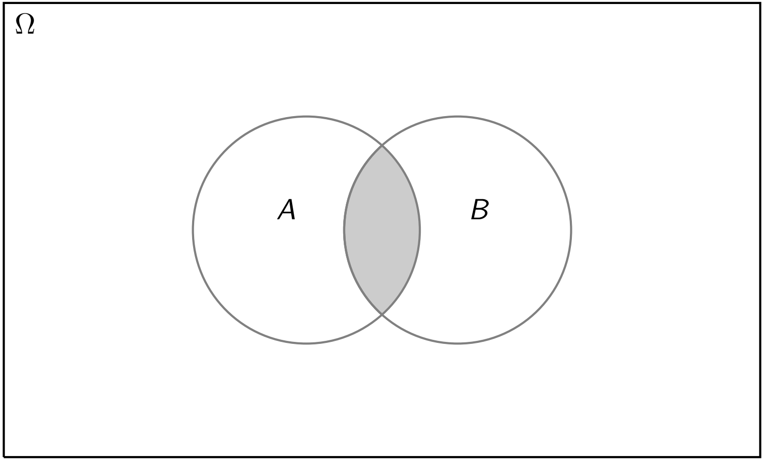


Figure: $A \cap B :=$ Outcomes that belong in A <u>and</u> B.

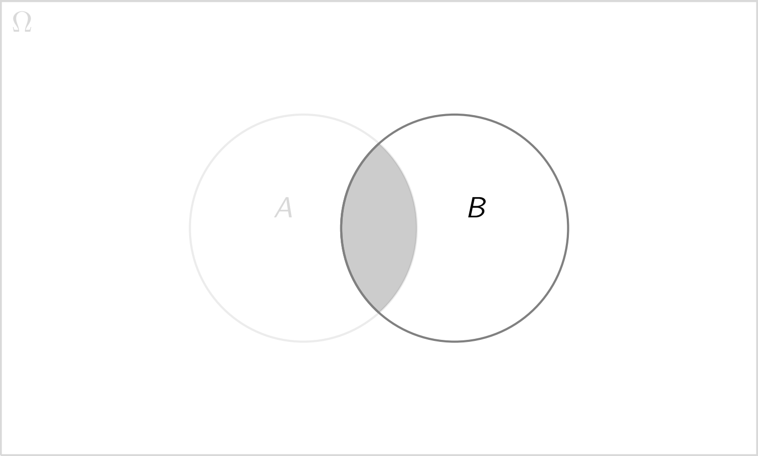


Figure: $A|B$, Outcomes that belong to A given that they are also in B.

Note that as $A \cap B \subseteq B$, then $P(A \cap B) \leq P(B)$, and therefore $P(A|B) \leq 1$.

It should be clear also that $P(A|B) = P(B|A)$ if and only if $P(A) = P(B)$.

Note that if we know $P(A|B)$ then we can obtain $P(A \cap B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(B)P(A|B)$$

This is known as the **product rule**.

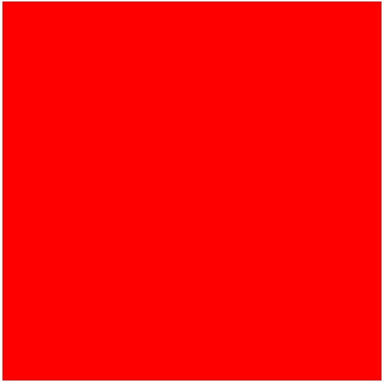
Assume now that $B = C \cap D$. Then we have $P(A \cap (C \cap D)) = P(C \cap D)P(A|(C \cap D))$, but $P(C \cap D) = P(D)P(C|D)$, so we have

$$P(A \cap C \cap D) = P(D)P(C|D)P(A|C \cap D)$$

that can be generalized as follows,

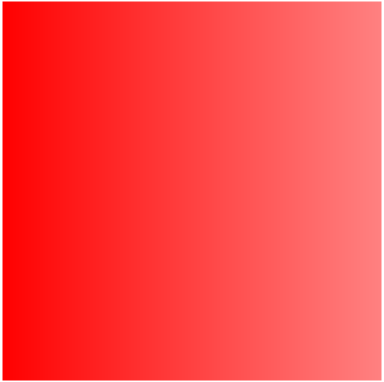
$$P\left(\bigcap_{i=1}^{k} A_i\right) = P(A_1)\prod_{i=2}^{k} P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right)$$

Assume I throw randomly a ball into a box, but you cannot see it:



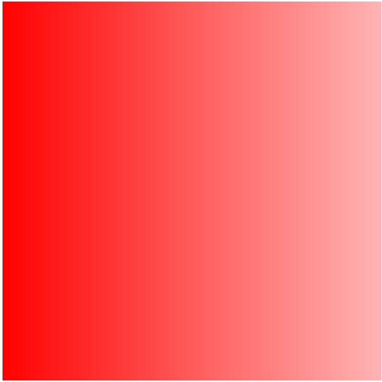
(it's there)...

Now I tell you, I throw randomly another ball, but you cannot see it either.



However, I tell you (I can see them), this second ball is to the right of the initial ball.

Again, I tell you, I throw randomly another ball, and tell you that it is to the right of the initial ball..



Again, I tell you, I throw randomly another ball, and tell you that it is to the right of the initial ball..

If this happens many times, with the updated information, you will believe that the ball is to left with more and more confidence.

This is Bayesian updating, with the new data, you update your beliefs about something, and actually this was similar to Bayes experiment to reach this conclusion.

So the question is:

1. Where is the ball?
2. Where is the ball given the 1 ball randomly fell to its right?
3. Where is the ball given that 2 balls randomly (and indep) fell to its right?
4. Where is the ball given that 3 balls randomly (and indep) fell to its right?
5. ...

Baye's rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Example

Assume a car shows the warning signal (Event B) which indicates that it might have a problem (Event A). $P(A|B)$ is the probability that the car has a problem given that the light turned on. $P(B|A)$ displays the probability that the car turns the light on when there is a problem, $P(A)$ is the unconditional probability that the car has a problem and $P(B)$ is the unconditional probability that the warning light turns on.

Random Variables

Definition

A **random variable** is a measurable function $X : \Omega \rightarrow \mathbb{X}$, where $\mathbb{X}$ is a measurable space, such that for $S \subseteq \mathbb{X}$

$$Pr(X \in S) = P(\{\omega \in \Omega | X(\omega) \in S\})$$

So the probability that X belongs to S is the measure of the set of outcomes that make X to have some particular characteristic (that makes it belong to S).

S could be $\mathbb{R}^n$, true/false, $\mathbb{N}$, a color, a decease, etc.

S could be $\mathbb{R}^n$, true/false, $\mathbb{N}$, a color, a decease, etc.

Definition

Let $X : \Omega \to \mathbb{X}$ be a r.v. X is said to be **discrete** if $\mathbb{X}$ is finite or countable infinite, and **continuous** if $\mathbb{X}$ is uncountable.

The true question is, how likely is that X belongs to S? (how likely is that $x = 2$, $x = \mathit{true}$, $x = \mathit{yellow}$, etc.?)

Definition

Let $X : \Omega \to \mathbb{X}$ be a r.v. The function $f_X : \mathbb{X} \to \mathbb{R}$ is the **density** of X if it holds that:

- for any $x \in \mathbb{X}$, $f_X(x) = P(\{\omega | X(\omega) = x\})$ when X is a discrete r.v., and
- for any $E \subseteq \mathbb{X}$, $\int_E f_X(x)dx = P(\{\omega | X(\omega) \in E\})$ when X is a continuous r.v.

Note that it also must hold that $\int_{\mathbb{X}} f(x)dx = 1$.

Definition

Let the continuous r.v. $X : \Omega \to \mathbb{X}$, where $\mathbb{X} \subseteq \mathbb{R}$. The **cumulative** function is:

$$F(x) = Pr(X \leq x) = \int_{\mathbb{X} \cap (-\infty, x]} f_X(z) dz$$

Definition

The **expected value** of a continuous r.v. $X : \Omega \to \mathbb{X}$ is defined as:

$$E[X] = \int_{\mathbb{X}} x f_X(x) dx$$

Note:

1. This is often found as $E[x] = \mu_X$.
2. If $g(x)$ is a function, then $E[g(x)] = \int_{\mathbb{X}} g(x) f(x) dx$ is the expected value of the function.

Definition

The **variance** of a continuous r.v. $X : \Omega \to \mathbb{X}$ is defined as:

$$var(X) = E[(X - E[x])^2] = \int_{\mathbb{X}} (x - E[X])^2 f_X(x) dx$$

You can often find this as $var(X) = \sigma_X^2$.

Definition

A continuous r.v. $X : \Omega \to \mathbb{R}$ is said to be **Normal** or **Gaussian** if its density is:

$$f_X(x) = \frac{\exp(-\frac{(x-\mu_X)^2}{2\sigma_X^2})}{\sqrt{2\pi\sigma_X^2}}$$

Definition

If X is Normal, and satisfies that $\mu_X = 0$ and $\sigma^2 = 1$ it is called **Standard Normal**.

Definition

The **conditional density** $f_{X|B}$ of a continuous r.v. X, given an event B, with $P(B) > 0$ must satisfy that:

$$P(X \in A|B) = \int_A f_{X|B}(x)dx$$

Note: To find, for example, the conditional expectation, you should compute the expectation as usual but using this density instead.

Definition

Let $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ be two jointly continuous r.v. The **joint density** $f_{XY} : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ is a function such that:

$$\int_E \int_I f_{X,Y}(x,y)dxdy = Pr(\{\omega \in \Omega | X(\omega) \in E, Y(\omega) \in I\})$$

for any $E \subseteq \mathbb{X}$, $I \subseteq \mathbb{Y}$.

Note: If $g(x,y)$ is a function, then

$$E[g(x,y)] = \int_E \int_I g(x,y) f_{X,Y}(x,y)dxdy$$

Definition

Let the continuous r.v. X and Y have a joint density $f_{X,Y} : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$. The **marginal** density is the function $f_X(x) : \mathbb{X} \to \mathbb{R}$ such that

$$f_X(x) = \int_{\mathbb{Y}} f_{XY}(x, y) dy$$

What this is trying to do, is to isolate the effect of x.

Definition

The **conditional density** of X, given $Y = y$, is a function $f_{X|Y}(x, y)$ such that $f_{X|Y}(x, y) = P(\{\omega \in \Omega | X(\omega) = x\} | \{\omega \in \Omega | Y(\omega) = y\})$ or,

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Again, if you want to find conditional expectation you should use this density function.

Definition

Two continuous r.v. X and Y are **independent** if their joint density is:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall x, y$$

Note that as a consequence of independence, for all x, y

$$f_{X,Y}(x|y) = f_X(x) \quad f_Y(y) > 0$$
$$f_{X,Y}(y|x) = f_Y(y) \quad f_X(x) > 0$$

And, also for two functions g and h

$$E[g(x)h(Y)] = E[g(X)]E[h(Y)]$$

Finally $var(X+Y) = var(X) + var(Y)$.

Definition

The **covariance** of two r.v., denoted as $cov(X, Y)$, is defined by:

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

If $cov(X, Y) = 0$, X and Y are said to be uncorrelated.

Definition

The **correlation coefficient** ρ of two r.v. X and Y that have strictly positive variances is defined as:

$$\rho = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$