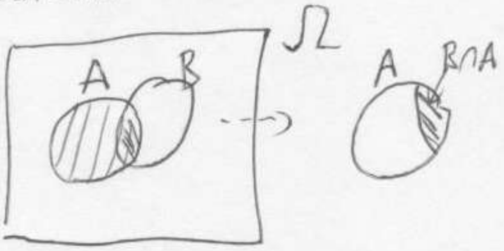
Conditional expectation

WARNING: This is HARD.

Introduction

Def: If $A, B \subset \Omega$ are events, $\mathbb{P}(A) > 0$, then the conditional probability of B under the condition A is

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



[Philosophically: $\frac{\#\text{ favourable outcomes}}{\#\text{ all outcomes}}$]

Problem: What if $\mathbb{P}(A) = 0$?

[**Remark:** A good question is why we are interested in that. We will come to that later.]

Bad news 1: If $\mathbb{P}(A) = 0$, then $\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)}$ formally

makes no sense: division by zero.

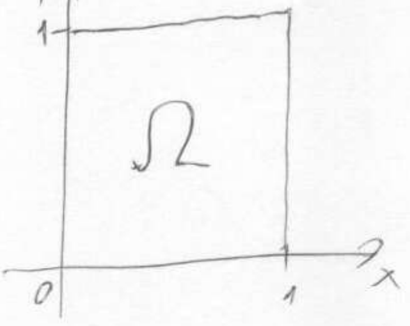
Bad news 2: If $\mathbb{P}(A) = 0$, then $\mathbb{P}(B|A)$ REALLY

makes no sense: $\boxed{\text{not even intuitively}}$!

Scaring example:

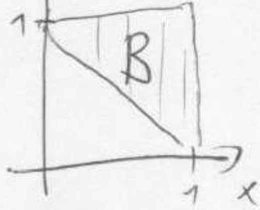
Consider the prob. space $\Omega = [0,1]^2$ with $\mathbb{P} = \text{Leb}$,

so $\omega \in (x,y) \in \Omega$ is uniformly chosen from $[0,1]^2$

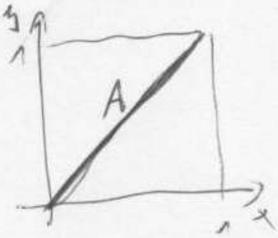


Now consider the events

$B := \{(x,y) \in \Omega \mid x+y \geq 1\}$:



$A := \{(x,y) \in \Omega \mid y = x\}$



How can we give sense to $\mathbb{P}(B|A)$?

Idea: Approximate A with some A_ε such that $\mathbb{P}(A_\varepsilon) > 0$ (although small), so $\mathbb{P}(B|A_\varepsilon)$ makes sense, and $\approx \mathbb{P}(B|A)$ at least intuitively.

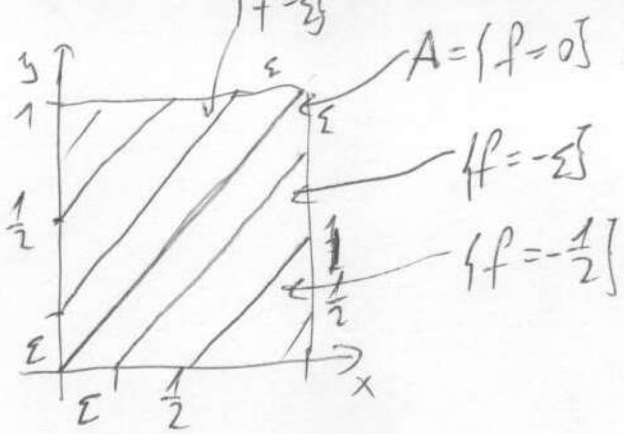
Q: How can we reasonably construct such an A_ε in a natural way?

Spoiler: I will give two equally natural constructions, leading to completely different results.

$\boxed{\text{Idea 1}}$ Let $f: \Omega \to \mathbb{R}$, $f(x,y) := y - x$.

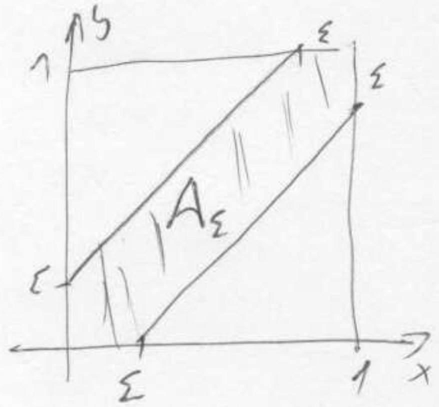
So $A = \{(x,y) \in \Omega \mid f(x,y) = 0\}$ is a level set of f.

Other level sets are also line segments:

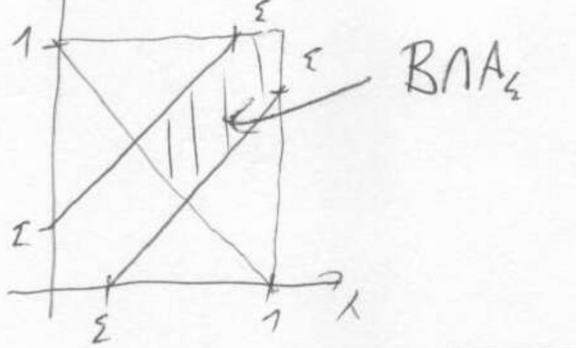


So let ~~$A_\varepsilon := \{(x,y)\in\Omega \mid -\varepsilon \le f$~~

$A_\varepsilon := \{(x,y)\in\Omega \mid -\varepsilon \le f(x,y) \le \varepsilon\}$:



Easy to see:



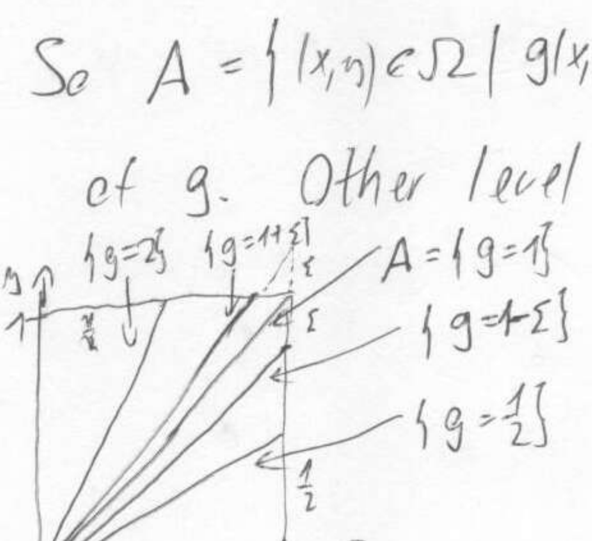
$P(B|A_\varepsilon) = \frac{1}{2}$ for $\forall \varepsilon > 0$,

so

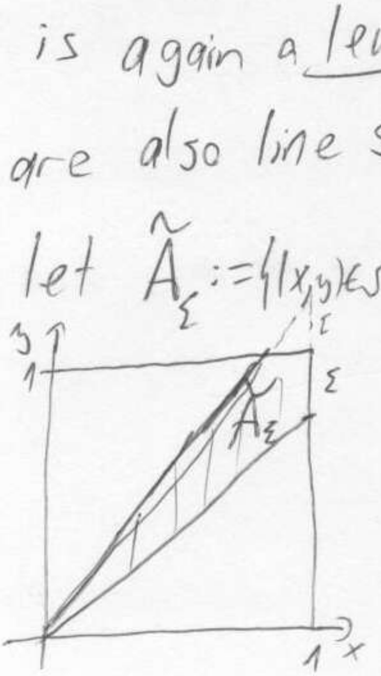
$$P(B|A) := \lim_{\varepsilon \to 0} P(B|A_\varepsilon) = \underline{\underline{\frac{1}{2}}}$$

<u>Idea 2</u>: Let $g: \Omega \to \mathbb{R}$, $g(x,y) := \frac{y}{x}$ (defined $\mathbb{P}$-a.e.)

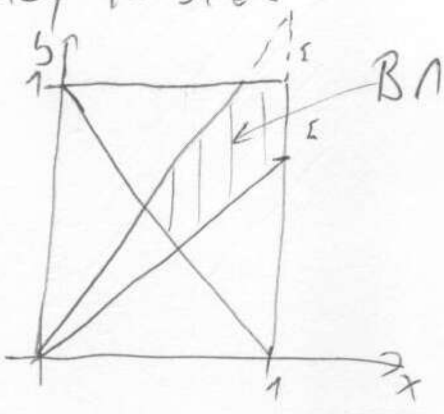
So $A = \{(x,y)\in\Omega \mid g(x,y) = 1\}$ is again a <u>level set</u> of g. Other level sets are also line segments:



So let $\tilde{A}_\varepsilon := \{(x,y)\in\Omega \mid 1-\varepsilon \le g(x,y) \le 1+\varepsilon\}$



Easy to see:



$\mathbb{P}(B|\tilde{A}_\varepsilon) \approx \frac{3}{4}$, so

$\approx$ triangle cut in half

$\tilde{\mathbb{P}}(B|A) := \lim_{\varepsilon \to 0} \mathbb{P}(B|\tilde{A}_\varepsilon) = \underline{\underline{\frac{3}{4}}}$.

Observation: The descriptions $\{f=0\} = A = \{g=1\}$

can be equally natural, depending on the context

– whether A showed up while you studied $y-x$,

or while you studied y/x.

~~Why we want the~~

Lesson to learn: For a single event $A \subset \Omega$ with $\mathbb{P}(A)=0$,

$\mathbb{P}(B|A)$ can not be made sensible.

So we will have to define $\mathbb{P}(B|A^z)$ simultaneously

for an entire family of events $A^z \subset \Omega$

with $\mathbb{P}(A^z) = 0$

[Say, in the above examples: $A^z := \{f=z\}$ $(z \in [-1,1])$

or $A^z := \{g=z\}$ $(z \in [0,\infty])$]

[

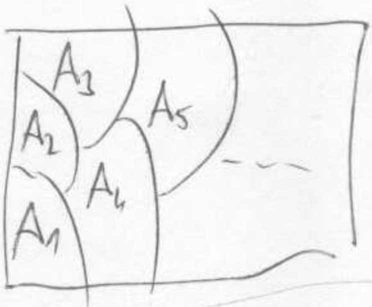
Why we want to define $\mathbb{P}(B|A)$ if $\mathbb{P}(A)=0$

Q: What is conditional prob. good for?

A: It's $\underset{\text{nearly}}{\wedge}$ always used to calculate total probabilities.

~~Thm (Theorem of total~~

Def: The sequence of events $\overbrace{A_1, A_2, A_3 \dots}^{\text{COUNTABLY MANY}} \subset \Omega$ is called a partition if always exactly 1 of them occurs:



$$\bigcup_i A_i = \Omega \quad \text{(covering)}$$

$$A_i \cap A_j = \emptyset \ \forall i \neq j \quad \text{(pairwise disjoint)}.$$

Thm (Theorem of total probability) (countable)

Let $A_1, A_2, A_3 \dots \subset \Omega$ be a partition, and $B \subset \Omega$ an event.

$\boxed{\text{T.T.P.}}$

Then $\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\,\mathbb{P}(B|A_i)$.

$\boxed{\text{Remark}}$ Convention: $0 \cdot \text{undefined} := 0$, so

if $\mathbb{P}(A_i)=0$, then $\mathbb{P}(A_i)\mathbb{P}(B|A_i) \overset{\text{convention}}{=} 0$.

[This is natural for 2 reasons:

1.) $\mathbb{P}(B|A_i)$ makes no sense, but if it would, then it would surely be $\in [0,1]$ and $0 \cdot \text{anything in } [0,1] = 0$.

2.) $\mathbb{P}(A_i)\,\mathbb{P}(B|A_i) \overset{\text{should be}}{=} \mathbb{P}(\underbrace{B \cap A_i}_{\subset A_i}) \le \mathbb{P}(A_i)$, so $=0$ if $\mathbb{P}(A_i)=0$.]

Conclusion 1

It seems that we don't need to define $\mathbb{P}(B|A_i)$ when $\mathbb{P}(A_i)=0$: it will/would be multiplied by 0 anyway.

Another tempting excuse:

$\mathbb{P}(B|A)$ means: ~~what~~ what would be the chance of B ocurring, if A would occur?

But $\boxed{\text{who cares}}$ what $\boxed{\text{would}}$ happen if A would occur?

If $\mathbb{P}(A)=0$, then A $\boxed{\text{never}}$ occurs!

Problem with these:

1) This Theorem / Convention / Conclusion 1 only hold as the partition is countable!

Adding „more than countably many zeroes" is not so simple.

2.) Zero probability events do occur all the time:

If X is a continuous random variable, then $\mathbb{P}(\{X=x\})=0$, but one of the (continuum many) $\{X=x\}$ is sure to happen.

Conclusion 2:

If the theorem is the only important thing, about conditional probability, then

Why not turn it into a definition ?

axiomatic

Indeed: Our def. on conditional probability will essentially be a generalization of the T.T.P. for uncountable partitions

Idea of the construction

[WARNING again: this is a HARD construction/definition, with many ingenious ideas.]

Idea 1: For given $B \subset \Omega$ and partition $\{A_i\}_{i \in I}$ of Ω

we want to define all $\mathbb{P}(B|A_i)$ simultaneously, such that the T.T.P. holds:

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i)\mathbb{P}(B|A_i).$$

However, this does not characterize $\{\mathbb{P}(B|A_i)\}_{i \in I}$, even when I is finite:

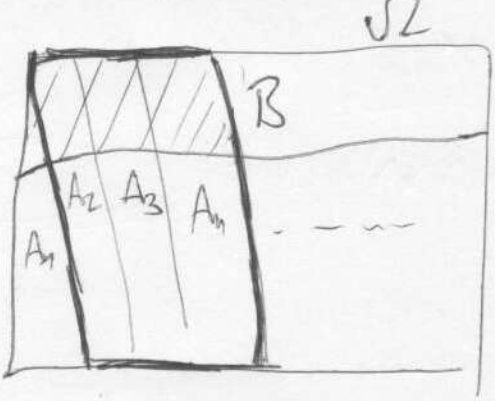
the equation $\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(A_i)\, x_i$ has <u>many</u>

solutions: a single equation for many variables $\{x_i\}_{i\in I}$.

So consider the following <u>easy corollary</u>:

<u>Thm</u> (L.T.T.P. Extra): Let ~~$A_1, A_2, \ldots \subset \Omega$~~ $\{A_i\}_{i\in I}$ be a countable partition. Let $B \subset \Omega$ be an event.

Furthermore, let $C \subset \Omega$ be an event like this: $\longrightarrow$



$C = \bigcup_{i\in J} A_i$ for some $J \subset I$

$C = A_2 \cup A_3 \cup A_4$

or, equivalently,

for any $i \in I$ either $A_i \subset C$ or $A_i \cap C = \phi$.

Then $\mathbb{P}(B \cap C) = \sum_{\underline{\underline{i\in J}}} \mathbb{P}(A_i)\,\mathbb{P}(B|A_i)$

[weight of the striped region above]

<u>This is what we will generalize.</u>

<u>Idea 2</u>: Defining $\mathbb{P}(B|A_i)$ <u>simultaneously</u> for all i means that the conditional probability is <u>not a number</u>, but a <u>function</u>:

Cond. prob: $\cancel{A_i \mapsto}$

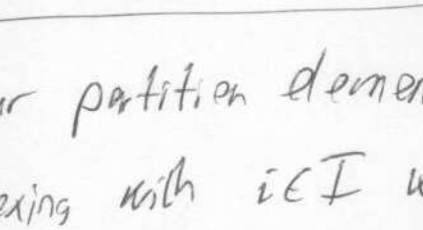
$$\{\text{partition elements}\} \longrightarrow \cancel{\mathbb{R}}\ [0,1]$$
$$A_i \longmapsto \mathbb{P}(B|A_i)$$

<u>Q</u>: How can we formulate such a function?

<u>Answer 1</u>: $\boxed{\text{measure theorist's approach:}}$

- Index your partition elements in your favourite way: say, indexing with $i \in I$ will do, and then cond. prob. is a function
$$Y : I \to [0,1]$$
$$i \mapsto \mathbb{P}(B|A_i)$$

Equivalently: the A_i are <u>level sets</u> of the function $f : \Omega \to I$ $\quad f(\omega) := i$ for $\omega \in A_i$



So $A_i = \{f = i\}$

Then the T.T.P. Extra reads: For $J \subset I$

$$\mathbb{P}(B \cap \{f \in J\}) = \sum_{i \in J} \mathbb{P}(A_i)\, Y(i).$$

This can then be generalized such that

$$\sum_{i \in J} \mathbb{P}(A_i) Y(i) \text{ becomes } \int_J Y(i)\, dg(i)$$

some measure on I.

$\boxed{\text{This is doable.}}$

$\boxed{\text{This is a standard construction in measure theory, called conditional measure.}}$

$\boxed{\text{This is } \boxed{\text{NOT}} \text{ what we will do.}}$

Disadvantage: The indexing is quite arbitrary:

- ~~s~~ elements of the partition can be reordered
- equivalently: <u>many functions</u> $f: \Omega \to$ some set I produce the <u>same level sets</u>
- brings in the extra object I, which has no real meaning.

Answer 2: Probabilist's approach

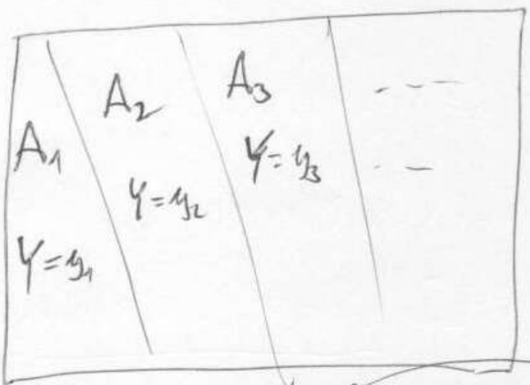
[This is what we will do] [Slightly more general than the previous]

Assigning numbers to partition elements $A_i \subset \Omega$

~~is equivalent to~~ $\quad Y: A_i \longmapsto \mathbb{P}(B|A_i)$

is equivalent to ~~def~~ giving a function Y

- defined on Ω, so $Y: \Omega \to [0,1]$
- but constant on each A_i: $Y\big|_{A_i} = y_i := \mathbb{P}(B|A_i)$



Now Thm T.T.P. Extra reads: for $C = \bigcup_{i \in J} A_i$

$$\mathbb{P}(B \cap C) = \sum_{i \in J} \mathbb{P}(A_i)\, \underset{\substack{\uparrow \\ Y = y_i \text{ on } A_i \subset \Omega}}{y_i} = \sum_{i \in J} \int_{A_i} Y(\omega)\, d\mathbb{P}(\omega) = \int_C Y \, d\mathbb{P}$$

Moreover: ~~A~~ GREAT OBSERVATION:

Let $\boxed{\mathcal{G} = \sigma(\{A_i\}_{i \in I}) \text{ be the } \sigma\text{-algebra generated by the partition}}$

Then • "Y is constant on each A_i" means $\boxed{Y \text{ is measurable w.r.t. } \mathcal{G}}$

• "$C = \bigcup_{i\in J} A_i$ for some $J \subset I$" means $\boxed{C \in \mathcal{G}}$

So Thm T.T.P. Extra reads:

<u>Thm</u> (<u>Theorem of total probability Extra, equivalent 2nd version</u>)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space.

Let $\{A_i\}_{i\in I}$ be a <u>countable</u> partition of Ω. (into events $A_i \in \mathcal{F}$)

Let $\mathcal{G} = \sigma(\{A_i\}_{i\in I}) \subset \mathcal{F}$ be the sub-σ-algebra generated ~~by the partition~~ by the partition.

Let $B \in \mathcal{F}$ be any event.

Define $Y: \Omega \to [0,1]$ as $Y(\omega) := \mathbb{P}(B \mid A_i)$ when $\omega \in A_i$

[<u>Notice</u>: The good old $\mathbb{P}(B \mid A_i) := \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(A_i)}$ is well defined whenever $\mathbb{P}(A_i) > 0$. Since the partition is <u>countable</u> (by assumption), this $Y(\omega)$ is well defined for $\mathbb{P}$-almost every ω.]

<u>Then</u>
- Y is $\mathcal{G}$-measurable, and
- for any $C \in \mathcal{G}$ $\quad \mathbb{P}(B \cap C) = \int_C Y \, d\mathbb{P}$

This statement can be generalized to <u>define</u> the conditional probability for <u>non-countable</u> partitions.

This is ALMOST what we will do.

Last idea (~~p~~ quite obvious):

~~It's no extra diff~~ Without extra difficulty, we can immediately define the conditional ~~probability~~ expectation.
then the cond. prob. will be a special case:

$$\mathbb{P}(B|A) = \mathbb{E}(\mathbb{1}_B|A).$$

Def: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space, $X: \Omega \to \mathbb{R}$ measurable, and assume that $\mathbb{E}X$ exists. Let $A \in \mathcal{F}$, $\mathbb{P}(A) > 0$.
Then the conditional expectation of X under the condition A is

$$\mathbb{E}(X|A) := \frac{\int_A X \, d\mathbb{P}}{\mathbb{P}(A)}$$

[weighted average of values of X on A.]

Thm (Theorem of total expectation) (T.T.E.)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space, $X: \Omega \to \mathbb{R}$ measurable s.t. $\mathbb{E}X$ exists.
Let $\{A_i\}_{i \in I}$ be a countable partition of Ω.

Then $$\mathbb{E}X = \sum_{i \in I} \mathbb{P}(A_i)\, \mathbb{E}(X|A_i)$$

[with the convention $0 \cdot \text{undefined} := 0$]

Proof: trivial from the definition.

As before, a trivial (equivalent) generalization:

<u>Thm</u> (T.T.E. Extra): Under the same assumptions, if $C = \bigcup_{i\in J} A_i$ for some $J \subset I$, then

$$\int_C X \, d\mathbb{P} = \sum_{i\in J} \mathbb{P}(A_i)\,\mathbb{E}(X|A_i).$$

And, just as before, $\sum_{i\in J} \mathbb{P}(A_i)\cdot(\dots) \rightsquigarrow \int_C (\dots)\, d\mathbb{P}$

<u>Thm</u> (<u>Theorem of total expectation Extra, equivalent 2nd Version</u>)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space.

Let $\{A_i\}_{i\in I}$ be a <u>countable</u> partition of Ω.

Let $\mathcal{G} = \sigma(\{A_i\}_{i\in I})$ be the generated σ-algebra.

Let $X: \Omega \to \mathbb{R}$ and assume that $\mathbb{E}X$ exists.

Define $Y: \Omega \to [-\infty, \infty]$ as $Y(\omega) := \mathbb{E}(X|A_i)$ when $\omega \in A_i$.

[Again: this is well-defined for $\mathbb{P}$-a.e. ω.]

Then
- Y is $\mathcal{G}$-measurable, and
- for any $C \in \mathcal{G}$ $\quad \int_C X\, d\mathbb{P} = \int_C Y\, d\mathbb{P}$

[<u>Notice!!</u> If $\int_C X d\mathbb{P} = \int_C Y d\mathbb{P}$ would hold for every $C \in \mathcal{F}$, that would mean $Y = X$ ~~a.s.~~ almost surely.
<u>But</u> we only claim it for $\boxed{C \in \mathcal{G}}$!!]

We are (finally) ready to define the conditional expectation ☺☺☺

Def: (Conditional expectation with respect to a σ-algebra)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. Let $X: \Omega \to \mathbb{R}$ integrable

The function $Y: \Omega \to \mathbb{R}$ is called the conditional expectation of Y w.r.t. $\mathcal{G}$, if

1.) Y is $\mathcal{G}$-measurable, and

2.) for any $A \in \mathcal{G}$ $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$.

Notation: $Y = \mathbb{E}(X \mid \mathcal{G})$

[Remark: Instead of „the conditional expectation" it would be fair to say „a version of the conditional expectation", since we haven't (yet) checked uniqueness.]

[Remark': If $Y = Z$ a.s, then $\int_A Y \, d\mathbb{P} = \int_A Z \, d\mathbb{P}$, so $\mathbb{E}(X \mid \mathcal{G})$ can only be unique up to modification on 0-measure sets (respecting $\mathcal{G}$-measurability)]

Example:

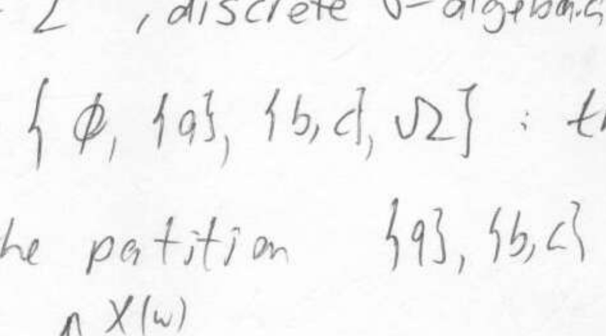
Let $\Omega = \{a, b, c\}$, $X = \mathbb{1}_{\{a,b\}}$, so $X(a) = 1$
$X(b) = 1$
$X(c) = 0$

$\mathbb{P} = \frac{1}{3} \chi$: uniform prob.

$\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \frac{1}{3}$

$\mathcal{F} = 2^{\Omega}$, discrete σ-algebra

$\mathcal{G} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$: this is generated by the patition $\{a\}, \{b, c\}$



Set $A = \{a\}$
$B = \{b, c\}$

Then $Y : \Omega \to \mathbb{R}$ defined as

$$Y|_A := 1 \qquad Y|_B := \frac{1}{2}$$

so $Y(a) = 1$, $Y(b) = \frac{1}{2}$, $Y(c) = \frac{1}{2}$

will do for $Y = \mathbb{E}(X|\mathcal{G})$:

- $\int_\emptyset X \, d\mathbb{P} = 0 = \int_\emptyset Y \, d\mathbb{P}$ ✓
- $\int_A X \, d\mathbb{P} = \cancel{A}\, \mathbb{P}(\{a\}) X(a) = \frac{1}{3} \cdot 1 = \mathbb{P}(\{a\}) Y(a) = \int_A Y \, d\mathbb{P}$ ✓
- $\int_B X \, d\mathbb{P} = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0$; $\int_B Y \, d\mathbb{P} = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}$ EQUAL ✓
- $\int_\Omega X \, d\mathbb{P} = \_\_ = \frac{2}{3} = \_\_ \int_\Omega Y \, d\mathbb{P}$ ✓

Of course: we chose

$$Y|_A := \text{average of } X \text{ on } A = \{a\} \quad (\text{boring})$$
$$Y|_B := \text{average of } Y \text{ on } B = \{b,c\} = \frac{1+0}{2}.$$

good old conditional expectation.

Thm (Theorem of total expectation extra, 3rd equivalent version)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space.

Let $\{A_i\}_{i\in I}$ be a countable partition of Ω.

Let $\mathcal{G} = \sigma(\{A_i\}_{i\in I}) \subset \mathcal{F}$ be the generated sub-σ-algebra.

Let $X : \Omega \to \mathbb{R}$ integrable

Define $Y : \Omega \to \mathbb{R}$ as $Y(\omega) := \mathbb{E}(X \mid A_i)$ when $\omega \in A_i$

[Where $\mathbb{E}(X|A_i)$ is the good old $\mathbb{E}(X|A_i) := \dfrac{\int_{A_i} X \, d\mathbb{P}}{\mathbb{P}(A_i)}$

This is well defined a.s.]

Then $\mathbb{E}(X|\mathcal{G}) = Y$.

So, indeed, $\mathbb{E}(X|\mathcal{G})$ is the generalization of the old notion.

very Special case: If $A \in \mathcal{F}$, $\mathbb{P}(A) \neq 0$ and

$\mathcal{G} = \{\emptyset, A, A^c, \Omega\}$, then $\mathbb{E}(X|\mathcal{G})(\omega) = \begin{cases} \mathbb{E}(X|A) & \text{if } \omega \in A \\ \mathbb{E}(X|A^c) & \text{if } \omega \in A^c \end{cases}$

<u>Thm</u>: The conditional expectation is <u>unique</u> up to changes on a set of measure 0:
If $Y:\Omega\to\mathbb{R}$ and $Z:\Omega\to\mathbb{R}$ are both versions of $\mathbb{E}(X|\mathcal{G})$, then $\mathbb{P}(Z=Y)=1$.

<u>Proof</u>: Y, Z are both $\mathcal{G}$-measurable, so
$Y-Z$ is also $\mathcal{G}$-measurable, so
$A:=\{Y-Z>0\}\in\mathcal{G}$, so by definition of the conditional expectation

$$\int_A Y\,d\mathbb{P}=\int_A X\,d\mathbb{P}=\int_A Z\,d\mathbb{P} \Longrightarrow \int_A \underbrace{Y-Z}_{>0 \text{ on } A}\,d\mathbb{P}=0$$

$$\mathbb{P}(Y>Z) = \mathbb{P}(A)=0$$

Similarly $\mathbb{P}(Y<Z)=0$ $\square$

<u>Existence of $\mathbb{E}(X|\mathcal{G})$</u>

<u>Thm</u>: <u>The conditional expectation exists</u>:
For any $(\Omega,\mathcal{F},\mathbb{P})$ prob-space, $\mathcal{G}\subset\mathcal{F}$, $X:\Omega\to\mathbb{R}$ integrable
$\exists\, Y:\Omega\to\mathbb{R}$ such that $Y=\mathbb{E}(X|\mathcal{G})$

$$\left[\begin{array}{l}\text{meaning } Y\in m\mathcal{G} \ (\mathcal{G}\text{-measurable}) \\ \text{and } \int_A X\,dP=\int_A Y\,d\mathbb{P} \quad \forall A\in\mathcal{G}\end{array}\right].$$

<u>Proof:</u> Assume 1st that $X \geq 0$.

Recall the <u>Radon-Nikodym theorem</u>: (with unusual notation)

~~If $(\Omega, \mathcal{G})$ is a meas~~

If $(\Omega, \mathcal{G})$ is a measurable space,

$\mu, \nu : \mathcal{G} \to \mathbb{R}$ σ-finite measures

and $\nu \ll \mu$,

then $\exists\, Y : \Omega \to \mathbb{R}^+$ $\boxed{\mathcal{G}\text{-measurable}}$

Such that $\nu(A) = \int_A Y \, d\mu$ for all $\underline{\underline{A \in \mathcal{G}}}$.

[<u>Of course</u> Y is $\mathcal{G}$-measurable;

Of course $A \in \mathcal{G}$:

there's no other σ-algebra in the assumptions.

But this <u>measurability is of key importance</u> for us now.]

[The first time during the course]

So consider the measurable space $(\Omega, \mathcal{G})$ with the measures μ, ν on <u>this</u> space : $\mu, \nu : \underline{\underline{\mathcal{G}}} \to \mathbb{R}^+$

$\mu(A) := \mathbb{P}(A)$ for all $A \in \mathcal{G}$ – so $\mu = \mathbb{P}|_{\mathcal{G}}$

$\nu(A) := \int_A X \, d\mathbb{P}$ for $A \in \mathcal{G}$.

<u>Lemma1</u> These are both finite measures on $(\Omega, \mathcal{G})$.

<u>Proof</u>: easy, HW - use the def. of the measure and basic properties of the integral.

In particular, $\mu(\Omega) = \mathbb{P}(\Omega) = 1$

$$\nu(\Omega) = \int_\Omega X \, d\mathbb{P} = \mathbb{E}X < \infty.$$

Lemma2 $\nu << \mu$.

Proof: ~~If~~ Obvious: If $\mu(A) = \mathbb{P}(A) = 0$, then

$$\nu(A) = \int_A X \, d\mathbb{P} = 0 \quad \checkmark$$

So applying the Radon-Nikody thm gives:

$\exists Y: \Omega \to \mathbb{R}$ <u>$\mathcal{G}$-measurable</u> s.t. $\forall A \in \mathcal{G}$

$$\nu(A) = \int_A Y \, d\mu$$

$$\overset{\text{def}}{\|}$$

$$\int_A X \, d\mathbb{P}$$

$\checkmark\checkmark$

<u>Lemma3</u>: For $A \in \mathcal{G} \subset \mathcal{F}$, $\mu = \mathbb{P}|_{\mathcal{G}}$ $\quad \int_A Y \, d\mu = \int_A Y \, d\mathbb{P}$ $\blacksquare$

integral on $(\Omega, \mathcal{G})$ — integral on $(\Omega, \mathcal{F})$

[Formally they are not the same, so we would need to check that they are equal — DON'T DO IT!]

We are done when $X \geq 0$.

In the general case, write $X = X_+ - X_-$

positive part ↑ ↑ negative part,

so $X_+ \geq 0$, $X_- \geq 0$,

So $Y_+ := \mathbb{E}(X_+ | \mathcal{G})$ and $Y_- := \mathbb{E}(X_- | \mathcal{G})$ exist,

So $Y := Y_+ - Y_-$ will do (easy).

$\square$

<u>Def.</u> (Conditional probability w.r.t. a σ-algebra)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a prob. space, $A \in \mathcal{F}$, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. Then $\mathbb{P}(A|\mathcal{G}) \overset{\text{def}}{=} \mathbb{E}(\mathbb{1}_A | \mathcal{G})$

<u>Properties of conditional expectation</u>

1.) If $\mathcal{G} = \mathcal{F}$, then $\mathbb{E}(X|\mathcal{G}) = X$.

2.) More generally: If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$

<u>Proof</u>: obvious – check the definition.

3.) If $\mathcal{G} = \{\emptyset, \Omega\}$ (the indiscrete σ-algebra), then

$$\mathbb{E}(X|\mathcal{G}) \equiv \mathbb{E}X \quad \text{(constant function)}$$

Proof: check the def.

4.) If $a, b \in \mathbb{R}$ ~~then~~, X, Y are integrable then

$$\mathbb{E}(aX+bY \mid \mathcal{G}) = a\,\mathbb{E}(X \mid \mathcal{G}) + b\,\mathbb{E}(Y \mid \mathcal{G})$$: <u>linearity</u>

Proof: check the def.

5.) $\boxed{\text{Tower rule}}$: Let $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{F}$, so

$\mathcal{G}_2$ is a smaller / rougher σ-algebra than $\mathcal{G}_1$.

Then $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}_1) \mid \mathcal{G}_2) = \mathbb{E}(X \mid \mathcal{G}_2)$,

but also $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) = \mathbb{E}(X \mid \mathcal{G}_2)$

No typo: the <u>smaller/ rougher</u> σ-algebra wins.

<u>Phenomenon:</u> ~~a~~ smaller σ-algebra = <u>less information</u>.