→ Start of lecture 4a ←

2.

Conditional Expectation (Section B.5)

formal definition:

Let $\mathcal{G} \subset \mathcal{F}$, $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$: $E|X| < \infty$, $X \in \mathcal{F}$

Thm: "$\exists !$" $\hat{X} \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$ s.t. !

$E \hat{X} 1_G = E X 1_G, \quad \forall G \in \mathcal{G}$ $\left[1_G(\omega) = \begin{cases} 1 & \text{if } \omega \in G \\ 0 & \text{if } \omega \notin G \end{cases} \right]$

uniqueness: another $\hat{X}'$ satisfies
$P(\hat{X} = \hat{X}') = 1$.

Such an $\hat{X}$ is a version of the conditional expectation of X given $\mathcal{G}$, we often write $E[X|\mathcal{G}]$ instead of $\hat{X}$

Conditional Expectation

If $X \in \mathcal{F}$ and if $\mathcal{G} \subseteq \mathcal{F}$ then we write $E[X|\mathcal{G}]$ for the conditional expectation of X given the information contained in $\mathcal{G}$. Sometimes we use the notation $E_{\mathcal{G}}[X]$.

p.137

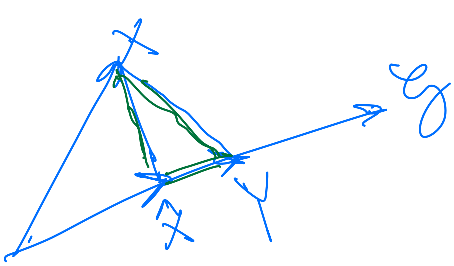
The following proposition contains everything that we will need to know about conditional expectations within this course. Also:

If EX^2, then for all $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ one has

$$E(X-Y)^2 = E(X-\hat{X})^2 + E(\hat{X}-Y)^2$$

[Think of Pythagoras and ordinary projections, you "see" in the picture

$$X-\hat{X} \perp \hat{X}-Y$$
]



Main Results

Proposition 1: Assume that $X \in \mathcal{F}$, and that $\mathcal{G} \subseteq \mathcal{F}$. Then the following hold.

- The random variable $E[X|\mathcal{G}]$ is completely determined by the information in $\mathcal{G}$ so we have

$$E[X|\mathcal{G}] \in \mathcal{G}$$

(by definition)

- If we have $Y \in \mathcal{G}$ then Y is completely determined by $\mathcal{G}$ so we have

$$E[XY|\mathcal{G}] = YE[X|\mathcal{G}]$$

  In particular we have

$$E[Y|\mathcal{G}] = Y$$

- If $\mathcal{H} \subseteq \mathcal{G}$ then we have the "law of iterated expectations"

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$$

(analogy with iterated projections)

- In particular we have

$$E[X] = E[E[X|\mathcal{G}]]$$

3.

Changing Measures

Changing Measures

Section B.6

Consider a probability measure P on $(\Omega, \mathcal{F})$, and assume that $L \in \mathcal{F}$ is a random variable with the properties that

$$L \geq 0$$

and

$$E^P[L] = 1.$$

For every event $A \in \mathcal{F}$ we now define the real number $Q(A)$ by the prescription

$$Q(A) = E^P[L \cdot I_A]$$

where the random variable I_A is the indicator for A, i.e.

$$I_A = \begin{cases} 1 & \text{if} \quad A \text{ occurs} \\ 0 & \text{if} \quad A^c \text{ occurs} \end{cases}$$

[I wrote before 1_A instead of I_A.]

Recall that

$$Q(A) = E^P [L \cdot I_A]$$

We now see that $Q(A) \geq 0$ for all A, and that

$$Q(\Omega) = E^P [L \cdot I_\Omega] = E^P [L \cdot 1] = 1$$

We also see that if $A \cap B = \emptyset$ then

$$\begin{aligned} Q(A \cup B) &= E^P [L \cdot I_{A \cup B}] = E^P [L \cdot (I_A + I_B)] \\ &= E^P [L \cdot I_A] + E^P [L \cdot I_B] \\ &= Q(A) + Q(B) \end{aligned}$$

(extends to finite disjoint unions)

Furthermore we see that

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0$$

We have thus more or less proved the following

Proposition 2: If $L \in \mathcal{F}$ is a nonnegative random variable with $E^P[L] = 1$ and Q is defined by

$$Q(A) = E^P[L \cdot I_A]$$

then Q will be a probability measure on $\mathcal{F}$ with the property that

for which you need countable additivity (true!)

$$\boxed{P(A) = 0 \quad \Rightarrow \quad Q(A) = 0.}$$

I turns out that the property above is a very important one, so we give it a name.

Absolute Continuity

Definition: Given two probability measures P and Q on $\mathcal{F}$ we say that Q is **absolutely continuous** w.r.t. P on $\mathcal{F}$ if, for all $A \in \mathcal{F}$, we have

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0$$

We write this as

$$Q << P.$$

If $Q << P$ and $P << Q$ then we say that P and Q are **equivalent** and write

$$Q \sim P$$

(this does NOT mean $Q = P$!)

Equivalent measures

It is easy to see that P and Q are equivalent if and only if

Other equivalent statements are

$$P(A) = 0 \quad \Leftrightarrow \quad Q(A) = 0$$

$P(A) > 0 \Leftrightarrow Q(A) > 0$,

or, equivalently,

$P(A) < 1 \Leftrightarrow Q(A) < 1$

$$P(A) = 1 \quad \Leftrightarrow \quad Q(A) = 1$$

(look at complements)

Two equivalent measures thus agree on all certain events and on all impossible events, but can disagree on all other events.

Simple examples:

- All non degenerate Gaussian distributions on R are equivalent.

- If P is Gaussian on R and Q is exponential then $Q << P$ but not the other way around. (why?)

Absolute Continuity ct'd

We have seen that if we are given P and **define** Q by

$$Q(A) = E^P [L \cdot I_A] \quad (*)$$

for $L \geq 0$ with $E^P [L] = 1$, then Q is a probability measure and $Q << P$. .

A natural question is now if **all** measures $Q << P$ are obtained in this way. The answer is yes, and the precise (quite deep) result is as follows. The proof is difficult and therefore omitted.

that is, by
formula $(*)$
for some L.

→ End of lecture 4a ←

The Radon Nikodym Theorem

Consider two probability measures P and Q on $(\Omega, \mathcal{F})$, and assume that $Q << P$ on $\mathcal{F}$. Then there exists a unique random variable L with the following properties

unique means that for another L' satisfying 1-4, one has $P(L = L') = 1$.

1. $Q(A) = E^P[L \cdot I_A], \quad \forall A \in \mathcal{F}$
2. $L \geq 0, \quad P - a.s.$
3. $E^P[L] = 1,$
4. $L \in \mathcal{F}$

The random variable L is denoted as

$$L = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}$$

and it is called the **Radon-Nikodym derivative** of Q w.r.t. P on $\mathcal{F}$, or the **likelihood ratio** between Q and P on $\mathcal{F}$.

A simple example

The Radon-Nikodym derivative L is intuitively the local scale factor between P and Q. If the sample space Ω is finite so $\Omega = \{\omega_1, \ldots, \omega_n\}$ then P is determined by the probabilities $p_1, \ldots, p_n$ where

$$p_i = P(\omega_i) \quad i = 1, \ldots, n$$

Now consider a measure Q with probabilities

$$q_i = Q(\omega_i) \quad i = 1, \ldots, n$$

If $Q << P$ this simply says that

$$p_i = 0 \quad \Rightarrow \quad q_i = 0$$

and it is easy to see that the Radon-Nikodym derivative $L = dQ/dP$ is given by

$$L(\omega_i) = \frac{q_i}{p_i} \quad i = 1, \ldots, n \qquad (\text{if } p_i > 0),$$

If $p_i = 0$ then we also have $q_i = 0$ and we can define the ratio q_i/p_i arbitrarily.

If $p_1, \ldots, p_n$ as well as $q_1, \ldots, q_n$ are all positive, then we see that $Q \sim P$ and in fact

$$\frac{dP}{dQ} = \frac{1}{L} = \left(\frac{dQ}{dP}\right)^{-1}$$

as could be expected.

Note on notation: $E_P X$ is ~~often~~ written as the Lebesgue integral $\int X\,dP$

Then $E_P(L I_A) = \int L I_A \, dP$ (1)

But $Q(A) \stackrel{!}{=} E_Q[I_A] = \int I_A \, dQ$ (2)

Now divide formally by dP (and multiply):

$$dQ = \frac{dQ}{dP} dP = L\,dP$$

and we understand (1) and (2) are equal.

mathematical nonsense

Computing expected values

A main use of Radon-Nikodym derivatives is for the computation of expected values.

Suppose therefore that $Q << P$ on $\mathcal{F}$ and that X is a random variable with $X \in \mathcal{F}$. With $L = dQ/dP$ on $\mathcal{F}$ then have the following result.

Proposition 3: With notation as above we have

$$E^Q[X] = E^P[L \cdot X],$$

$\int X \, dQ = \int LX \, dP$ (write $L = \frac{dQ}{dP}$ etc.....)

Proof: We only give a proof for the simple example above where $\Omega = \{\omega_1, \ldots, \omega_n\}$. We then have

$$\begin{aligned} E^Q[X] &= \sum_{i=1}^{n} X(\omega_i) q_i = \sum_{i=1}^{n} X(\omega_i) \frac{q_i}{p_i} p_i \\ &= \sum_{i=1}^{n} X(\omega_i) L(\omega_i) p_i = E^P[X \cdot L] \end{aligned}$$

The Abstract Bayes' Formula

We can also use Radon-Nikodym derivatives in order to compute conditional expectations. The result, known as the abstract **Bayes' Formula**, is as follows.

Theorem 4: Consider two measures P and Q with $Q << P$ on $\mathcal{F}$ and with

$$L^{\mathcal{F}} = \frac{dQ}{dP} \quad \text{on } \mathcal{F}$$

Assume that $\mathcal{G} \subseteq \mathcal{F}$ and let X be a random variable with $X \in \mathcal{F}$. Then the following holds

$$E^{Q}\left[X \middle| \mathcal{G}\right] = \frac{E^{P}\left[L^{\mathcal{F}} X \middle| \mathcal{G}\right]}{E^{P}\left[L^{\mathcal{F}} \middle| \mathcal{G}\right]}$$

note the denominator,

different from $E^Q X = E^P[LX]$ ----

(see book Proposition B.41)

Dependence of the σ-algebra

Suppose that we have $Q << P$ on $\mathcal{F}$ with

$$L^{\mathcal{F}} = \frac{dQ}{dP} \quad \text{on } \mathcal{F}$$

Now consider smaller σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Our problem is to find the R-N derivative

$$L^{\mathcal{G}} = \frac{dQ}{dP} \quad \text{on } \mathcal{G}$$

Note that also $Q \ll P$ on $\mathcal{G}$!

We recall that $L^{\mathcal{G}}$ is characterized by the following properties

1. $Q(A) = E^P\left[L^{\mathcal{G}} \cdot I_A\right] \quad \forall A \in \mathcal{G}$

2. $L^{\mathcal{G}} \geq 0$

3. $E^P\left[L^{\mathcal{G}}\right] = 1$

4. $L^{\mathcal{G}} \in \mathcal{G}$ crucial!

A natural guess would perhaps be that $L^{\mathcal{G}} = L^{\mathcal{F}}$, so let us check if $L^{\mathcal{F}}$ satisfies points 1-4 above.

By assumption we have

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] \quad \forall A \in \mathcal{F}$$

Since $\mathcal{G} \subseteq \mathcal{F}$ we then have

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] \quad \forall A \in \mathcal{G}$$

so point 1 above is certainly satisfied by $L^{\mathcal{F}}$. It is also clear that $L^{\mathcal{F}}$ satisfies points 2 and 3. It thus seems that $L^{\mathcal{F}}$ is also a natural candidate for the R-N derivative $L^{\mathcal{G}}$, but the problem is that we do not in general have $L^{\mathcal{F}} \in \mathcal{G}$. So $L^{\mathcal{G}} \neq L^{\mathcal{F}}$ in general.

This problem can, however, be fixed. By iterated expectations we have, for all $A \in \mathcal{G}$,

$$Q(A) = E^P\left[L^{\mathcal{F}} \cdot I_A\right] = E^P\underbrace{\left[E^P\left[L^{\mathcal{F}} \cdot I_A \middle| \mathcal{G}\right]\right]}_{\text{Consider this}}$$

Then previous formula becomes

$Q(A) = E^P \left[E^P [L^{\mathcal{F}} | \mathcal{G}] I_A \right]$

Since $A \in \mathcal{G}$ we have

$$E^P \left[L^{\mathcal{F}} \cdot I_A \middle| \mathcal{G} \right] = E^P \left[L^{\mathcal{F}} \middle| \mathcal{G} \right] I_A$$

Let us now define $L^{\mathcal{G}}$ by

$$L^{\mathcal{G}} = E^P \left[L^{\mathcal{F}} \middle| \mathcal{G} \right]$$

We then obviously have $L^{\mathcal{G}} \in \mathcal{G}$ and

by a defining property of conditional expectation given $\mathcal{G}$

$$Q(A) = E^P \left[L^{\mathcal{G}} \cdot I_A \right] \quad \forall A \in \mathcal{G}$$

It is easy to see that also points 2-3 are satisfied so we have proved the following result.

A formula for $L^{\mathcal{G}}$

Proposition 5: If $Q << P$ on $\mathcal{F}$ and $\mathcal{G} \subseteq \mathcal{F}$ then, with notation as above, we have

$$L^{\mathcal{G}} = E^P \left[L^{\mathcal{F}} \middle| \mathcal{G} \right]$$!

Here, the point was that we wanted

$L^{\mathcal{G}}$ to be $\mathcal{G}$-measurable

Another example is that P and Q are two possible distributions of a random variable with densities f_P and f_Q. Assume $f_P > 0$ and $f_Q > 0$ on $\mathbb{R}$.

Then $P \sim Q$ (seen as probability measures on $\mathcal{F} = \mathcal{B}$, Borel sets, and

$$\frac{dQ}{dP}(x) = \frac{f_Q(x)}{f_P(x)}, \quad \frac{dP}{dQ}(x) = \frac{f_P(x)}{f_Q(x)}.$$

Think of two normal densities f_P, f_Q.

→ End of lecture 4b ←