Recall: We consider the special case where X, Y have common density $f_{X,Y}(x,y)$:

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy.$$

We are interested in $\mathbb{E}(X \mid Y)$ and define the conditional density

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx$.

Now set $g(y) = \int_{\mathbb{R}} x\, f_{X|Y}(x|y)\,dx$

("the expected value of X given $Y=y$").

We want to show that $g(y)$ satisfies the conditions of conditional expectations.

By inspection, g is $\sigma(Y)$ measurable and integrable.

It remains to check (3): Let $A \in \Sigma_2$, then

$$\int_{\{Y \in A\}} X \, d\mathbb{P} \overset{\text{WTS}}{=} \int_{\{Y \in A\}} g(Y) \, d\mathbb{P}$$

$$\int_\Omega X \, I_{\{Y \in A\}} \, d\mathbb{P} \qquad\qquad \int_\Omega g(Y) \, I_{\{Y \in A\}} \, d\mathbb{P}$$

$$\iint_{\mathbb{R}\times\mathbb{R}} x \, I_{\{Y\in A\}} f_{X,Y}(x,y) \, dx \, dy \qquad \iint_{\mathbb{R}\times\mathbb{R}} g(y) \, I_{\{Y\in A\}} f_{X,Y}(x,y) \, dx \, dy$$

$$\int_{\mathbb{R}} x \int_A f_{X,Y}(x,y) \, dy \, dx \qquad \int_{\mathbb{R}} \int_A g(y) \, f_{X,Y}(x,y) \, dy \, dx$$

$$(*) \qquad\qquad (+)$$

By definition, $f_{X,Y}(x,y) = f_{X|Y}(x|y)\, f_Y(y)$ for Lebesgue almost every value. Then,

$$(*) = \int_{\mathbb{R}} x \int_A f_{X|Y}(x|y)\, f_Y(y)\, dy \quad dx$$

$$= \int_A f_Y(y) \int_{\mathbb{R}} x\, f_{X|Y}(x|y)\, dx\, dy$$

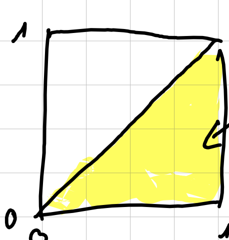
$$= \int_A f_Y(y)\, g(y)\, dy$$

$$= \int_A g(y) \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} \int_A g(y)\, f_{X,Y}(x,y)\, dy\, dx = (+)$$

Which concludes the proof as A was arbitrary $\square$

Example: Consider random variables X, Y on $[0,1]\times[0,1]$ with density $f(x,y) = 2\, I_{\{x \geq y\}}$



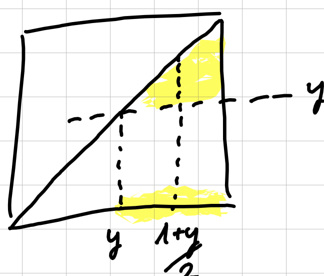
uniform distribution here.

We get $f_Y(y) = \int_0^1 2\, I_{\{x \geq y\}}\, dx = 2 \int_y^1 dx = 2(1-y)$

$$f_{X|Y}(x|y) = \frac{2\, I_{\{x\geq y\}}}{2(1-y)} = \frac{I_{\{x \geq y\}}}{1-y}$$

$$g(y) = \int_0^1 x \frac{I_{\{x\geq y\}}}{1-y}\, dx = \frac{1}{1-y} \int_y^1 x\, dx = \frac{1}{1-y} \cdot \frac{1-y^2}{2}$$

$$= \frac{1+y}{2} \Rightarrow \mathbb{E}(X \mid Y=y) = \frac{1+y}{2}$$



Example: Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables. Let $S_n = X_1 + \dots + X_n$. What is $\mathbb{E}(X_1 \mid S_n)$?

By symmetry $\mathbb{E}(X_1|S_n) = \mathbb{E}(X_2|S_n) = \dots = \mathbb{E}(X_n|S_n)$ a.s.

We must have

$$\int_A \mathbb{E}(X_1|S_n)\, d\mathbb{P} + \dots + \int_A \mathbb{E}(X_n|S_n)\, d\mathbb{P}$$
$$= \int_A X_1\, d\mathbb{P} + \dots + \int_A X_n\, d\mathbb{P} = \int_A S_n\, d\mathbb{P}$$

for all $A \in \sigma(S_n)$. Hence

$$n \int_A \mathbb{E}(X_1|S_n)\, d\mathbb{P} = \int_A S_n\, d\mathbb{P}$$ and

$$\mathbb{E}(X_1|S_n) = \frac{S_n}{n} \quad \text{a.s.}$$

<u>Remark:</u> We define cond. probabilities through cond. expectations:

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}(I_A \mid \mathcal{G}) \quad \left(\text{Now a random variable !!}\right)$$

This is uniquely determined (a.s) and satisfies

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n \,\Big|\, \mathcal{G}\Big) = \sum_{n\geq 1} \mathbb{P}(A_n \mid \mathcal{G}) \quad \text{a.s.}$$

if $A_1, A_2, \ldots$ are disjoint.

If $\mathcal{G} = \sigma(B)$ is generated by an event B, then $\mathbb{P}(A \mid \mathcal{G})$ is a random variable Y with

$$Y(\omega) = \begin{cases} a & \omega \in B \\ b & \omega \in B^c \end{cases} \quad (\mathcal{G}\text{-measurability}).$$

We get $a\,\mathbb{P}(B) = \int_B Y \, d\mathbb{P} = \int_B I_A \, d\mathbb{P} = \int I_{A\cap B} \, d\mathbb{P} = \mathbb{P}(A \cap B)$

$$\Rightarrow a = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B) \quad \text{and similarly,}$$

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

Independent Random Variables

If $X_1, X_2, \dots, X_n$ are independent, then

$$\mathbb{E}\left(h(X_1, X_2, \dots, X_n) \mid X_1\right) = g(X_1)$$

with $g(x) = \mathbb{E}\left(h(x, X_2, \dots, X_n)\right)$.

This follows from Fubini's theorem:
Since $\sigma(X_1)$ is generated by $\{X \in A\}$, $A \in \mathcal{B}(\mathbb{R})$
we only need to consider $\int_{\{X \in A\}} d\mathbb{P}$.

$$\int_{\{X_1 \in A\}} h(X_1, X_2, \dots, X_n)\, d\mathbb{P}$$

laws

$$= \int_{X_1 \in A} \int \cdots \int_{\mathbb{R}^{n-1}} h(X_1, X_2, \dots, X_n)\, d(\Lambda_1 \times \Lambda_2 \times \dots \times \Lambda_n)$$

(by independence)

$$= \int \cdots \int_{\mathbb{R}^n} I_{\{X_1 \in A\}}\, h(X_1, \dots, X_n)\, d(\Lambda_2 \times \Lambda_3 \times \dots \times \Lambda_n \times \Lambda_1) \quad \text{(Fubini)}$$

$$= \int_{\mathbb{R}} I_{\{X_1 \in A\}} \int \cdots \int_{\mathbb{R}^{n-1}} h(X_1, \dots, X_n)\, d(\Lambda_2 \times \dots \times \Lambda_n)\, d\Lambda_1$$

$= \int_{\{X_1 \in A\}} g(X_1)\, d\Lambda_1$. So $g(X_1)$ satisfies (3).

Example: Let $X_1, \dots, X_n$ be independent and write $\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$. What is $\mathbb{E}(\overline{X} | X_1)$?

Take $h(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}$, then,

$$g(x) = \mathbb{E}\left(\frac{x + X_2 + \dots + X_n}{n}\right) = \frac{x + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)}{n}$$

and $\mathbb{E}(\overline{X} | X_1) = \frac{X_1 + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)}{n}$.

Martingales

Stochastic processes & filtrations

A (discrete) stochastic process is a sequence $X_0, X_1, X_2, \dots$ of random variables.

A filtration is a sequence of σ-algebras
$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$$

We write $\mathcal{F}_\infty = \sigma\left(\bigcup_{n=0}^{\infty} \mathcal{F}_n\right) \subseteq \mathcal{F}$.

The process $X_0, X_1, \dots$ is said to be adapted to the filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$ measurable.

super-

A martingale is a sequence $X_0, X_1, X_2, \dots$

sub-

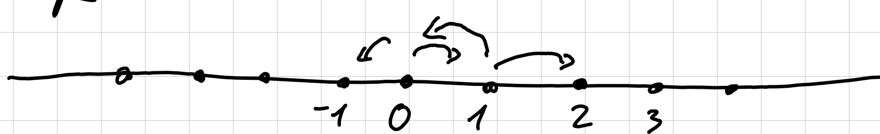
adapted to the filtration $\mathcal{F}_0, \mathcal{F}_1, \dots$ such that

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \overset{\le}{\underset{\ge}{=}} X_{n-1}.$$

Equivalently, in terms of increments,

$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \overset{\le}{\underset{\ge}{=}} 0$$

Example: Standard random walk



Let $Y_1, Y_2, \dots$ be independent random variables with

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = 1/2$$

and set $X_0 = 0$, $X_n = Y_1 + Y_2 + \dots + Y_n$

With $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$, the sequence $X_0, X_1, \dots$ is adapted and

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1} + Y_n \mid \mathcal{F}_{n-1})$$

$$= \mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-1}) + \mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) = X_{n-1} + \mathbb{E}(Y_n) = X_{n-1}.$$

This also works for any other Y_i with $\mathbb{E}(Y_i) = 0$.

Example: Let $Y_1, Y_2, \ldots$ be independent random variables with $\mathbb{E}(Y_i) = 1$.

Set $X_n = X_0 \prod_{i=1}^{n} Y_i$. This is a martingale (w.r.t. $\tilde{\mathcal{F}}_n = \sigma(Y_1, \ldots, Y_n)$)

since $\mathbb{E}(X_n \mid \tilde{\mathcal{F}}_{n-1}) = \mathbb{E}(X_{n-1} \cdot Y_n \mid \tilde{\mathcal{F}}_{n-1})$

$= X_{n-1}\, \mathbb{E}(Y_n \mid \tilde{\mathcal{F}}_{n-1}) = X_{n-1}\, \mathbb{E}(Y_n) = X_{n-1}$.

Example: Let X be a fixed $\tilde{\mathcal{F}}$-meas. random variable. Fix a filtration $(\tilde{\mathcal{F}}_n)$ and set $X_n = \mathbb{E}(X \mid \tilde{\mathcal{F}}_n)$. This is a martingale.

$$\mathbb{E}(X_n \mid \tilde{\mathcal{F}}_{n-1}) = \mathbb{E}(\mathbb{E}(X \mid \tilde{\mathcal{F}}_n) \mid \tilde{\mathcal{F}}_{n-1})$$
$$= \mathbb{E}(X \mid \tilde{\mathcal{F}}_{n-1}) = X_{n-1}$$