

Theorem (Jensen's inequality for conditional expectations)

Assume $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ is a convex function, X is a random variable on $(\Omega, \mathcal{B}, \mathbb{P})$ s.t. $X, \varphi(X) \in L^1(\Omega, \mathcal{B}, d\mathbb{P})$. Then

$$\varphi(\mathbb{E}(X|F)) \leq \mathbb{E}(\varphi(X)|F) \quad (\text{a.e.})$$

Fact

Assume $U \subset \mathbb{R}^n$ is a convex open set, $\varphi: U \rightarrow \mathbb{R}$ is a convex function. Then φ is continuous and $\forall x_0 \in U$ there is an affine function l (so $l(x) = a_1 x_1 + \dots + a_n x_n + b$) s.t. $\varphi \geq l$ on U and $\varphi(x_0) = l(x_0)$.

We have proved this for $n=1$. We will use this only for $n=2$ and functions of the form $\varphi(x) = \psi(|x|)$, where $\psi: (0, \infty) \rightarrow \mathbb{R}$ is convex, for which the result can be deduced from the case $n=1$.

Corollary

There is a countable collection M of affine functions s.t. $\varphi = \sup_{l \in M} l$.

Proof

For every $x_0 \in U \cap \mathbb{Q}^n$ choose l s.t. $\varphi \geq l$ and $\varphi(x_0) = l(x_0)$. Let M be the collection of the functions l we thus get.

Consider $\widehat{\varphi} \stackrel{\text{def}}{=} \sup_{\ell \in M} \ell$. Then $\widehat{\varphi}$ is convex and $\varphi(x) = \widehat{\varphi}(x) \quad \forall x \in U \cap \mathbb{Q}^n$. Since φ and $\widehat{\varphi}$ are continuous by Fact, we conclude that $\varphi = \widehat{\varphi}$. $\square$

Proof of Jensen's inequality

We identify $\mathbb{C}$ with $\mathbb{R}^2$, $s+it \mapsto (s, t)$. Let M be a countable collection of affine functions on $\mathbb{R}^2$ s.t. $\varphi = \sup_{\ell \in M} \ell$. Write $X = S + iT$, $S, T: \Omega \rightarrow \mathbb{R}$.

For every $\ell \in M$, $\ell(s, t) = as + bt + c$, we have

$$\begin{aligned} \ell(\mathbb{E}(X | \mathcal{F})) &= \ell(\mathbb{E}(S | \mathcal{F}) + i \mathbb{E}(T | \mathcal{F})) \\ &= a \mathbb{E}(S | \mathcal{F}) + b \mathbb{E}(T | \mathcal{F}) + c \\ &= \mathbb{E}(aS + bT + c | \mathcal{F}) \\ &= \mathbb{E}(\ell(X) | \mathcal{F}) \\ &\leq \mathbb{E}(\varphi(X) | \mathcal{F}). \end{aligned}$$

If we view $\mathbb{E}(X | \mathcal{F})$, $\mathbb{E}(\varphi(X) | \mathcal{F})$ as functions, this means that

$$\ell(\mathbb{E}(X | \mathcal{F})) \leq \mathbb{E}(\varphi(X) | \mathcal{F}) \quad \text{a.e.}$$

Since M is countable, taking the supremum over $\ell \in M$, we get that

$$\varphi(\mathbb{E}(X | \mathcal{F})) \leq \mathbb{E}(\varphi(X) | \mathcal{F}) \quad \text{a.e.} \quad \square$$

Recall that since $P(\Omega) = 1$, we have
 $L^p(\Omega, \mathcal{B}, dP) \subset L^1(\Omega, \mathcal{B}, dP)$ for all $1 \leq p \leq \infty$.
 (If $1 \leq p < \infty$, then by Hölder's inequality

$$\|f\|_1 = \int_{\Omega} |f| dP \leq \left(\int_{\Omega} |f|^p dP \right)^{1/p} \left(\int_{\Omega} 1^q dP \right)^{1/q}$$

$$= \|f\|_p P(\Omega)^{1/q} = \|f\|_p.$$

We also have $\|f\|_1 \leq \|f\|_{\infty} \int_{\Omega} 1 dP = \|f\|_{\infty}$.

Corollary

For every $1 \leq p \leq \infty$, $\mathbb{E}(\cdot | \mathcal{F})$ defines a contraction $L^p(\Omega, \mathcal{B}, dP) \rightarrow L^p(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$.

Proof

For $1 \leq p < \infty$, we apply Jensen's inequality to $\varphi(z) = |z|^p$:

$$\|\mathbb{E}(X | \mathcal{F})\|_p^p \leq \mathbb{E}(|X|^p | \mathcal{F}). \quad (*)$$

By integrating over Ω we get

$$\begin{aligned} \|\mathbb{E}(X | \mathcal{F})\|_p^p &\leq \int_{\Omega} \mathbb{E}(|X|^p | \mathcal{F}) dP \\ &= \int_{\Omega} |X|^p dP = \|X\|_p^p. \end{aligned}$$

For $p = \infty$ we have by (*) (for $p=1$) that

$$\begin{aligned} |\mathbb{E}(X | \mathcal{F})| &\leq \mathbb{E}(|X| | \mathcal{F}) \leq \mathbb{E}(\|X\|_{\infty} \mathbb{1} | \mathcal{F}) \\ &= \|X\|_{\infty} \mathbb{1}, \end{aligned}$$

hence $\|\mathbb{E}(X | \mathcal{F})\|_{\infty} \leq \|X\|_{\infty}$. □

We also have the following important property.

Proposition

For $X \in L^1(\Omega, \mathcal{B}, dP)$ and $Y \in L^\infty(\Omega, \mathcal{F}, P|_{\mathcal{F}})$,
we have $E(XY | \mathcal{F}) = E(X | \mathcal{F})Y$.

Proof

Let us first check this for $Y = \pi_A$ ($A \in \mathcal{F}$).

By definition of $E(\cdot | \mathcal{F})$, for all $B \in \mathcal{F}$,
we have

$$\begin{aligned} \int_B E(X \pi_A | \mathcal{F}) dP &= \int_B X \pi_A dP = \int_{A \cap B} X dP \\ &= \int_{A \cap B} E(X | \mathcal{F}) dP = \int_B E(X | \mathcal{F}) \pi_A dP. \end{aligned}$$

Hence $E(X \pi_A | \mathcal{F}) = E(X | \mathcal{F}) \pi_A$.

By linearity we then get $E(XY | \mathcal{F}) = E(X | \mathcal{F})Y$

for all simple $\mathcal{F}$ -measurable Y .

For general $Y \in L^\infty(\Omega, \mathcal{F}, P|_{\mathcal{F}})$ we can

find simple $\mathcal{F}$ -measurable ~~Y_n~~ Y_n s.t.

~~$Y_n \rightarrow Y$~~ $Y_n \rightarrow Y$ in L^∞ . Then $XY_n \rightarrow XY$ in L^1 ,

so $E(XY_n | \mathcal{F}) \rightarrow E(XY | \mathcal{F})$ in L^1 . On the

other hand, $E(XY_n | \mathcal{F}) = E(X | \mathcal{F})Y_n$

$\rightarrow E(X | \mathcal{F})Y$. Hence $E(XY | \mathcal{F}) = E(X | \mathcal{F})Y$. □

Using this we get the following
interpretation of conditional expectations
on L^2 -spaces.

Proposition

$E(\cdot | \mathcal{F}) : L^2(\Omega, \mathcal{B}, dP) \rightarrow L^2(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$
is the orthogonal projection.

Proof

We need to check that

$$\int_{\Omega} X \bar{Y} dP = \int_{\Omega} E(X | \mathcal{F}) \bar{Y} dP$$

for all $X \in L^2(\Omega, \mathcal{B}, dP)$ and $Y \in L^2(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$.
For $Y \in L^{\infty}(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$ this follows from
the previous proposition. Since $L^{\infty}(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$
is dense in $L^2(\Omega, \mathcal{F}, dP|_{\mathcal{F}})$, we then
get the result for all Y . $\square$

Theorem (Levy's martingale convergence theorem)

Assume $(\Omega, \mathcal{B}, P)$ is a probability space
and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ is a sequence of σ -sub-
algebras of $\mathcal{B}$. Put $\mathcal{F}_{\infty} = \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$.

Then, for every $X \in L^1(\Omega, \mathcal{B}, dP)$, we have

$$E(X | \mathcal{F}_n) \xrightarrow{n \rightarrow \infty} E(X | \mathcal{F}_{\infty}) \text{ a.e. and in } L^1.$$

Proof

We will only show convergence in L^1
(the same proof shows a similar result
for $L^p, 1 \leq p < \infty$), a.e. convergence is
significantly harder.

If $X \in L^1(\Omega, \mathcal{F}_{n_0}, dP|_{\mathcal{F}_{n_0}})$ for some n_0 ,
 then $E(X|\mathcal{F}_n) = E(X|\mathcal{F}_{n_0})$ for all $n \geq n_0$,
 including $n = \infty$, and there is nothing to prove.
 Thus the result is true for $X \in \bigcup_{n=1}^{\infty} L^1(\Omega, \mathcal{F}_n, dP|_{\mathcal{F}_n})$.

Since the maps $E(\cdot|\mathcal{F}_n)$ are contracting,
 the result is then true for $X \in \overline{\bigcup_{n=1}^{\infty} L^1(\Omega, \mathcal{F}_n, dP|_{\mathcal{F}_n})}$.

We claim that this closure coincides
 with $L^1(\Omega, \mathcal{F}_{\infty}, dP|_{\mathcal{F}_{\infty}})$. Using as usual

density of simple functions, for this it
 suffices to show that $\forall A \in \mathcal{F}_{\infty} \forall \varepsilon > 0$

there is $B \in \mathcal{F}_n$ for n large enough s.t.

$P(A \Delta B) = \| \pi_A - \pi_B \|_1 < \varepsilon$. This, in turn,

is proved by showing that the collection
 of sets $A \in \mathcal{F}_{\infty}$ satisfying this property $\forall \varepsilon > 0$
 is a σ -algebra containing $\bigcup_{n=1}^{\infty} \mathcal{F}_n$, hence
 it coincides with $\mathcal{F}_{\infty}$.

Finally, for general $X \in L^1(\Omega, \mathcal{B}, dP)$
 we have $E(X|\mathcal{F}_n) = E(E(X|\mathcal{F}_{\infty})|\mathcal{F}_n)$
 (since if $A \in \mathcal{F}_n$, then $\int_A E(E(X|\mathcal{F}_{\infty})|\mathcal{F}_n) dP$
 $= \int_A E(X|\mathcal{F}_{\infty}) dP = \int_A X dP$). Hence
 $E(X|\mathcal{F}_n) = E(E(X|\mathcal{F}_{\infty})|\mathcal{F}_n) \xrightarrow[n \rightarrow \infty]{L^1} E(X|\mathcal{F}_{\infty})$ □

Example

Assume Ω_n ($n \geq 1$) is a finite or countable set. We view it as a metric space with metric $d_n(x, y) = 1$ for $x \neq y$. Consider

$$\Omega = \prod_{n=1}^{\infty} \Omega_n, \quad \mathcal{B} = \prod_{n=1}^{\infty} \mathcal{B}(\Omega_n) = \prod_{n=1}^{\infty} \mathcal{P}(\Omega_n).$$

Let $\mathcal{F}_n$ be the σ -algebra generated by the sets $A_1 \times \dots \times A_n \times \Omega_{n+1} \times \dots$ ($A_i \subset \Omega_i$).

$$\text{Then } \mathcal{B}\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right) = \mathcal{B}.$$

Assume P is a probability measure on $(\Omega, \mathcal{B})$.

For $\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n$, define

$$\Omega(\omega_1, \dots, \omega_n) = \{x \in \Omega : x_1 = \omega_1, \dots, x_n = \omega_n\}.$$

$$\text{Then } E(X | \mathcal{F}_n) = \sum_{\omega_1, \dots, \omega_n} \frac{\sum_{\Omega(\omega_1, \dots, \omega_n)} X dP}{P(\Omega(\omega_1, \dots, \omega_n))} \mathbb{1}_{\Omega(\omega_1, \dots, \omega_n)}$$

By Levy's theorem we conclude that for P -a.e. $\omega \in \Omega$ we have

$$\frac{\sum_{\Omega(\omega_1, \dots, \omega_n)} X dP}{P(\Omega(\omega_1, \dots, \omega_n))} \xrightarrow{n \rightarrow \infty} X(\omega). \quad (*)$$

In particular, if μ is a finite measure on $(\Omega, \mathcal{B})$, $\mu \ll P$, then

$$\frac{\mu(\Omega(\omega_1, \dots, \omega_n))}{P(\Omega(\omega_1, \dots, \omega_n))} \xrightarrow{n \rightarrow \infty} \frac{d\mu}{dP}(\omega) \text{ for } P\text{-a.e. } \omega.$$

For $d\mu = \mathbb{1}_A dP$ ($A \in \mathcal{B}$) this gives

$$\frac{P(A \cap \Omega(\omega_1, \dots, \omega_n))}{P(\Omega(\omega_1, \dots, \omega_n))} \xrightarrow{n \rightarrow \infty} \mathbb{1}_A(\omega) \text{ for } P\text{-a.e. } \omega.$$