10. Martingale convergence

We will now give another application of maximal inequalities, this time drawing strongly on ideas and techniques from probability theory.

10.1 Conditional expectation.

In one of its typical applications, probability is a study of stochastic processes. These are sequences of random variables — which is the probabilist’s name for measurable functions — parametrized by time (which can be integer or real valued). And, as the process evolves, more and more information is revealed. This information is technically encoded in terms of \( \sigma \)-algebras. We are thus naturally led to the following concept:

**Definition 10.1 (Filtration)** Let \( (\Omega, \mathcal{F}, \mu) \) be a measurable space. A **filtration** is a sequence \( \{ \mathcal{F}_n \}_{n \geq 1} \) of \( \sigma \)-subalgebras of \( \mathcal{F} \) with

\[
\forall n \geq 1: \mathcal{F}_n \subseteq \mathcal{F}_{n+1}.
\]

(10.1)

Many different \( \sigma \)-algebras often come up in the study of stochastic processes. Given a function \( f \), the question is then: What is the best estimator of \( f \) given the “knowledge” contained in the \( \sigma \)-algebra \( G \)? This is the subject of:

**Theorem 10.2 (Conditional expectation)** Let \( (\Omega, \mathcal{F}, \mu) \) be a measure space with \( \mu(\Omega) < \infty \) and let \( G \subseteq \mathcal{F} \) be a \( \sigma \)-algebra. For each \( f \in L^1(\mu) \) there exists a unique \( f_G \in L^1(\mu) \) such that

1. \( f_G \) is \( G \)-measurable, and
2. for all \( A \in G \),

\[
\int_A f \, d\mu = \int_A f_G \, d\mu.
\]

(10.2)

Moreover, we have

\[
\|f_G\|_{L^1(\Omega, G, \mu)} \leq \|f\|_{L^1(\Omega, \mathcal{F}, \mu)}.
\]

(10.3)

**Proof.** Given \( f \in L^1(\Omega, \mathcal{F}, \mu) \), consider the measure \( \nu \) on \( (\Omega, G) \) defined by

\[
\forall A \in G: \quad \nu(A) := \int_A f \, d\mu.
\]

(10.4)

Then \( \nu \ll \mu \) (as measures on \( (\Omega, G) \)) and so, by the Radon-Nikodym Theorem, there exists a unique \( f_G \in L^1(\Omega, G, \mu) \) such that

\[
\forall A \in G: \quad \nu(A) = \int_A f_G \, d\mu.
\]

(10.5)

Comparing with (10.4), this proves (1-2) above. Since

\[
\|f_G\|_{L^1(\Omega, G, \mu)} = \int_{\{f_G \geq 0\}} f_G \, d\mu - \int_{\{f_G < 0\}} f_G \, d\mu
\]

\[
= \int_{\{f_G \geq 0\}} f \, d\mu - \int_{\{f_G < 0\}} f \, d\mu \leq \|f\|_{L^1(\Omega, \mathcal{F}, \mu)},
\]

(10.6)

we get that \( f_G \) is an \( L^1 \)-contraction as well.

\[\Box\]
Remark 10.3. The above construction works even for \( \mu \) that is \( \sigma \)-finite albeit when restricted \( \mathcal{G} \). This restriction is too annoying to carry around through derivations. We will therefore restrict attention to finite measures only.

In probability, where \( \mu(\Omega) = 1 \), we call \( f_\mathcal{G} \) the conditional expectation given \( \mathcal{G} \), with the notation \( E(f|\mathcal{G}) \). (In particular, conditional expectations are still random variables.) It may not be quite apparent from the properties (1-2) above why \( f_\mathcal{G} \) is the best estimator of \( f \) given \( \mathcal{G} \). This is perhaps clearer from:

**Lemma 10.4** (Conditional expectation is a projection) Let \((\Omega, \mathcal{F}, \mu)\) be a measure space with \( \mu(\Omega) < \infty \) and let \( \mathcal{G} \subseteq \mathcal{F} \) be a \( \sigma \)-algebra. Then \( L^2(\Omega, \mathcal{G}, \mu) \) is a closed linear subspace of \( L^2(\Omega, \mathcal{F}, \mu) \) and the map \( f \mapsto f_\mathcal{G} \) is the orthogonal projection of \( L^2(\Omega, \mathcal{F}, \mu) \) onto \( L^2(\Omega, \mathcal{G}, \mu) \).

**Proof.** Since \( f \mapsto f \) is an isometry of \( L^2(\Omega, \mathcal{G}, \mu) \) into \( L^2(\Omega, \mathcal{F}, \mu) \), the fact that \( L^2(\Omega, \mathcal{G}, \mu) \) is closed in itself implies that it is closed as a subspace of \( L^2(\Omega, \mathcal{F}, \mu) \). The orthogonal projection \( Pf \) of \( f \in L^2(\Omega, \mathcal{F}, \mu) \) onto \( L^2(\Omega, \mathcal{G}, \mu) \) is then characterized by

\[
\forall g \in L^2(\Omega, \mathcal{G}, \mu) : \quad \int g Pf \, d\mu = \int g f \, d\mu \quad (10.7)
\]

Specializing this to \( g := 1_A \) for \( A \in \mathcal{G} \) (where \( \mu(\Omega) < \infty \) ensures that \( 1_A \in L^2(\Omega, \mathcal{G}, \mu) \)), the uniqueness of \( f_\mathcal{G} \) then shows \( Pf = f_\mathcal{G} \) \( \mu \)-a.e. \( \square \)

The conditional expectation has a number of natural properties. Assume henceforth that \((\Omega, \mathcal{F}, \mu)\) is a measure space with \( \mu \) finite.

**Lemma 10.5** (Additivity of conditional expectation) For each \( \sigma \)-algebra \( \mathcal{G} \subseteq \mathcal{F} \), \( f \mapsto f_\mathcal{G} \) is a continuous linear map \( L^1(\Omega, \mathcal{F}, \mu) \to L^1(\Omega, \mathcal{G}, \mu) \). For all \( p \in [1, \infty) \), the map extends to a continuous linear map of \( L^p(\Omega, \mathcal{F}, \mu) \to L^p(\Omega, \mathcal{G}, \mu) \) with

\[
\|f_\mathcal{G}\|_{L^p(\Omega, \mathcal{G}, \mu)} \leq \|f\|_{L^p(\Omega, \mathcal{F}, \mu)} \quad (10.8)
\]

**Proof.** Linearity follows from Theorem 10.2. The continuity as the map of \( L^1 \)-space follows from (10.3). From (10.2) we get, for all \( A \in \mathcal{G} \),

\[
\int_A f_\mathcal{G} \, d\mu = \int_A f \, d\mu \leq \int_A |f| \, d\mu = \int_A |f|_\mathcal{G} \, d\mu. \quad (10.9)
\]

Specializing to \( A := \{f_\mathcal{G} > |f|_\mathcal{G}\} \) and invoking a sign reversal for \( f \) we get

\[
|f_\mathcal{G}| \leq |f|_\mathcal{G}, \quad \mu\text{-a.e.}. \quad (10.10)
\]

Let \( p, q \in [1, \infty] \) be Hölder conjugate indices and let \( h \in L^q(\Omega, \mathcal{G}, \mu) \). Then (10.2) and approximations by simple functions show

\[
\int f_\mathcal{G} h \, d\mu = \int f h \, d\mu \leq \|f\|_{L^p(\Omega, \mathcal{F}, \mu)} \|h\|_{L^q(\Omega, \mathcal{G}, \mu)} \quad (10.11)
\]

where we noticed that \( \|h\|_{L^1(\Omega, \mathcal{F}, \mu)} = \|h\|_{L^1(\Omega, \mathcal{G}, \mu)} \). The standard duality argument then implies (10.8). \( \square \)

We remark that (10.10) is a special case (for \( \varphi(x) := |x| \)) of the following lemma whose proof we leave to the reader:

---

Preliminary version (subject to change anytime!) 
Typeset: March 3, 2020
Lemma 10.6 (Conditional Jensen’s inequality) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex with at most linear growth. Then
\[ \varphi \circ (f_{\mathcal{G}}) \leq (\varphi \circ f)_{\mathcal{G}} \quad \mu\text{-a.e.} \quad (10.12) \]

Of particular interest in this lecture will be also the dependence of $f_{\mathcal{G}}$ on the underlying $\sigma$-algebra. This is the content of:

Lemma 10.7 (“Smaller always wins”) Suppose $\Omega, \mathcal{F}, \mu$ is a measure space with $\mu$ finite and let $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ be $\sigma$-algebras. Then for all $f \in L^1(\mu)$,
\[ (f_{\mathcal{G}_2})_{\mathcal{G}_1} = f_{\mathcal{G}_1} = (f_{\mathcal{G}_1})_{\mathcal{G}_2}. \quad (10.13) \]

Proof. Since $\mathcal{G}_1 \subseteq \mathcal{G}_2$, from (10.2) we get
\[ \int_A (f_{\mathcal{G}_2})_{\mathcal{G}_1} \, d\mu = \int_A f_{\mathcal{G}_2} \, d\mu = \int_A f \, d\mu. \quad (10.14) \]
Since $(f_{\mathcal{G}_2})_{\mathcal{G}_1}$ is $\mathcal{G}_1$-measurable, we get $(f_{\mathcal{G}_2})_{\mathcal{G}_1} = f_{\mathcal{G}_1}$, by the uniqueness of the conditional expectation. The second equality follows from the fact that if $f \in L^1(\mu)$ is $\mathcal{G}$-measurable, then $f = f_{\mathcal{G}}$ by the uniqueness of the conditional expectation. By $\mathcal{G}_1 \subseteq \mathcal{G}_2$, a $\mathcal{G}_1$-measurable function is automatically $\mathcal{G}_2$-measurable. \hfill $\square$

10.2 Martingales.

Having state the basic properties of the conditional expectation, we now move to the study of stochastic processes. The following concept is central to the present lecture:

Definition 10.8 (Martingale) Given a filtration $\{\mathcal{F}_n\}_{n \geq 1}$ on a measure space $(\Omega, \mathcal{F}, \mu)$, where $\mu$ is finite, a sequence $\{f_n\}_{n \geq 1}$ of functions is said to be a martingale with respect to $\{\mathcal{F}_n\}_{n \geq 1}$ if for all $n \geq 1$,
\begin{itemize}
  \item[(1)] $f_n$ is $\mathcal{F}_n$-measurable,
  \item[(2)] $f_n \in L^1(\mu)$ and
  \end{itemize}

\[ (f_{n+1})_{\mathcal{F}_n} = f_n. \quad (10.15) \]

We note that, in one of the standard interpretation of a martingale, $f_n$ denotes the wealth of a player in $n$-th round of a game. Property (1) ensures that the wealth at “time” $n$ is determined by the first $n$ rounds of the game. Property (2) in turns shows that, if measured by expectation, the game is fair.

There are many examples of martingales that come up in probability. The one we will care for is the subject of:

Lemma 10.9 Let $\{\mathcal{F}_n\}_{n \geq 1}$ be a filtration on a measure space $(\Omega, \mathcal{F}, \mu)$, where $\mu$ is finite. Then for all $f \in L^1(\mu)$
\[ f_n := f_{\mathcal{F}_n} \quad (10.16) \]
defines a martingale with respect to $\{\mathcal{F}_n\}_{n \geq 1}$.

Proof. By Theorem 10.2, $f_{\mathcal{F}_n}$ is well defined and $\mathcal{F}_n$-measurable for all $n \geq 1$. The property (10.15) then follows from Lemma 10.7. \hfill $\square$
A rather useful consequence of the martingale property is the fact that martingales converge a.e. under very mild integrability assumptions. The following is a version of this fact (proved by P. Levy) for the martingale defined in Lemma 10.9:

**Theorem 10.10 (Levy’s Forward Martingale Convergence)** Let \( \{F_n\}_{n \geq 1} \) be a filtration on a measure space \((\Omega, \mathcal{F}, \mu)\) with \(\mu\) finite. Set
\[
F_\infty := \sigma \left( \bigcup_{n \geq 1} F_n \right). \tag{10.17}
\]
Then for all \(f \in L^1(\mu),\)
\[
f_{F_n} \overset{n \to \infty}{\to} f_{F_\infty}, \quad \mu\text{-a.e. and in } L^1(\mu). \tag{10.18}
\]

We will again rely on a maximal inequality. The starting observation is:

**Lemma 10.11 (Doob’s maximal inequality)** Let \(\{f_n\}_{n \geq 1}\) be a martingale with respect to a filtration \(\{F_n\}_{n \geq 1}\). Then for all \(n \geq 1\) and all \(t \in \mathbb{R},\)
\[
t\mu\left( \max_{1 \leq k \leq n} f_k > t \right) \leq \int_{\{\max_{1 \leq k \leq n} f_k > t\}} f_n \, d\mu \tag{10.19}
\]

**Proof.** Let \(A_1 := \{f_1 > t\}\) and for \(k = 2, \ldots, n\) set
\[
A_k := \{ \max_{1 \leq \ell < k} f_\ell \leq t < f_k \} \tag{10.20}
\]
Then \(\{A_k\}_{k=1}^n\) are disjoint with
\[
\{ \max_{1 \leq k \leq n} f_k > t \} = \bigcup_{k=1}^n A_k. \tag{10.21}
\]
Since also (by (1) in the definition of the martingale), \(A_k \in F_k \subseteq F_n,\) from (2) in the definition of the martingale we get
\[
\int_{A_k} f_n \, d\mu = \int_{A_k \cap F_n} f \, d\mu = \int_{A_k \cap F_k} f_k \, d\mu \geq t\mu(A_k) \tag{10.22}
\]
where we used that \(f_k \geq t\) on \(A_k.\) Hence we get
\[
\int_{\{\max_{1 \leq k \leq n} f_k > t\}} f_n \, d\mu = \sum_{k=1}^n \int_{A_k} f_n \, d\mu \geq \sum_{k=1}^n t\mu(A_k) = t\mu\left( \max_{1 \leq k \leq n} f_k > t \right) \tag{10.23}
\]
which is the desired claim. \(\square\)

We note that Doob’s inequality is usually stated with \(f_n\) replaced by \(|f_n|\) on the right-hand side (and \(t\) non-negative), but we need the above version in the proof of Theorem 10.10. Enhancing slightly the argument from the proof of Lemma 10.11, we obtain a limit version thereof:
Lemma 10.12 (Maximal inequality — limit version) Given a filtration \{\mathcal{F}_n\}_{n \geq 1}, for all \( f \in L^1(\mu) \), all \( A \in \mathcal{F}_\infty \) where \( \mathcal{F}_\infty \) is as in (10.17), and all \( t \in \mathbb{R} \),

\[
 t\mu\left( A \cap \left\{ \limsup_{n \to \infty} f_{\mathcal{F}_n} > t \right\} \right) \leq \int_{A \cap \left\{ \limsup_{n \to \infty} f_{\mathcal{F}_n} > t \right\}} f_{\mathcal{F}_\infty} \, d\mu. \tag{10.24}
\]

Proof. Abbreviate \( f_n := f_{\mathcal{F}_n} \). Let \( m \geq n \geq 1 \) and pick \( A \in \mathcal{F}_n \). By the same argument as in the proof of Lemma 10.11 we then get

\[
 t\mu\left( A \cap \left\{ \max_{n \leq k \leq m} f_k > t \right\} \right) \leq \int_{A \cap \left\{ \max_{n \leq k \leq m} f_k > t \right\}} f_m \, d\mu \tag{10.25}
\]

Since we can replace \( f_m \) by \( f_{\mathcal{F}_\infty} \) in the integrand on the right-hand side, taking the limit \( m \to \infty \) and \( n \to \infty \) with the help of the Dominated Convergence on the right and the Monotone Convergence for measures on the left (remember that \( \mu \) is finite) we conclude

\[
 t\mu\left( A \cap \left\{ \limsup_{n \to \infty} f_m > t \right\} \right) \leq \int_{A \cap \left\{ \limsup_{n \to \infty} f_m > t \right\}} f_{\mathcal{F}_\infty} \, d\mu \tag{10.26}
\]

holds for all \( A \in \bigcup_{n \geq 1} \mathcal{F}_n \). But both sides are \( \sigma \)-additive in \( A \) and are thus (signed) measures. It is thus readily checked that the set of \( A \in \mathcal{F} \) for which the inequality holds is a \( \sigma \)-algebra. It follows that the inequality holds for all \( A \in \mathcal{F}_\infty \). \( \Box \)

We are now ready to give:

Proof of Theorem 10.10. We will proceed very much like in the proof of Birkhoff’s Pointwise Ergodic Theorem. Abbreviate \( f_n := f_{\mathcal{F}_n} \). For each \( a < b \), define

\[
 E_{a,b} := \left\{ \liminf_{n \to \infty} f_n < a < b < \limsup_{n \to \infty} f_n \right\} \tag{10.27}
\]

Then \( E_{a,b} \in \mathcal{F}_\infty \) and (10.24) shows

\[
 \beta \mu(E_{a,b}) \leq \int_{E_{a,b}} f_\infty \, d\mu. \tag{10.28}
\]

Applying (10.24) to \( \{-f_n\}_{n \geq 1} \) (which arises from \( -f \) instead of \( f \)) in turns gives

\[
 -\alpha \mu(E_{a,b}) \leq \int_{E_{a,b}} -f_\infty \, d\mu. \tag{10.29}
\]

Putting (10.28–10.29) together yields \( \beta \mu(E_{a,b}) \leq \alpha \mu(E_{a,b}) \) which by finiteness of \( \mu \) forces \( \mu(E_{a,b}) = 0 \) for all \( a < b \). It follows that

\[
 \Omega_0 := \bigcap_{a, \beta \in \mathbb{Q}, a < b} (E_{a,b})^c \tag{10.30}
\]

has full measure under \( \mu \).

Since the limit of \( f_n \) exists with values in extended reals on \( \Omega_0 \), we can set

\[
 \bar{f} := \begin{cases} \lim_{n \to \infty} f_n, & \text{on } \Omega_0, \\ 0, & \text{on } \Omega_0^c. \end{cases} \tag{10.31}
\]
Fatou’s lemma then implies
\[ \int |f| \, d\mu \leq \liminf_{n \to \infty} \int |f_n| \, d\mu \leq \int |f| \, d\mu \]  
(10.32)
and so \( f \) is finite \( \mu \)-a.e. and, in fact, \( f \in L^1(\mu) \).

In order to prove convergence in \( L^1(\mu) \), we use a uniform integrability argument. Fix \( t > 0 \). Then
\[ \|f_n - \bar{f}\|_{L^1(\mu)} \leq \int_{\{|f_n| \leq t\}} |f_n - \bar{f}| \, d\mu + \int_{\{|f_n| > t\}} |f_n| \, d\mu + \int_{\{|f_n| > t\}} |\bar{f}| \, d\mu. \]  
(10.33)
Since \( |f_n - \bar{f}| \leq t + |\bar{f}| \in L^1(\mu) \), Dominated Convergence ensures that the first integral on the right-tends to zero as \( n \to \infty \). By the same token, the third integral converges to \( \int_{\{|f_n| > t\}} |\bar{f}| \, d\mu \) which tends to zero as \( t \to \infty \). Concerning the middle integral, here we note that (10.10) implies
\[ \int_{\{|f_n| > t\}} |f_n| \, d\mu \leq \int_{\{f_n > t\}} |f| \, d\mu \leq \int_{\{sup_{n \geq n} |f_n| > t\}} |f| \, d\mu. \]  
(10.34)
Dominated Convergence then shows
\[ \limsup_{n \to \infty} \int_{\{|f_n| > t\}} |f_n| \, d\mu \leq \int_{\{f > t\}} |f| \, d\mu. \]  
(10.35)
Since \( f \in L^1(\mu) \), the right-hand side tends to zero as \( t \to \infty \).

We thus have \( f_n \to \bar{f} \) in \( L^1(\mu) \). But then for each \( A \in \mathcal{F}_\infty \)
\[ \int_A f_\infty \, d\mu = \lim_{n \to \infty} \int_A f_n \, d\mu = \int_A \bar{f} \, d\mu. \]  
(10.36)
Since \( \bar{f} \) is \( \mathcal{F}_\infty \)-measurable, checking this for \( A := \{|f > f_\infty\} \) and \( A := \{|\bar{f} > f_\infty\} \) then gives \( \bar{f} = f_\infty \) \( \mu \)-a.e., thus finishing the proof.

We finish by noting that the above martingale convergence gives us another version of the differentiation theorem for functions \( f : \mathbb{R} \to \mathbb{R} \). Indeed, we have:

**Theorem 10.13** (Differentiation along dyadic intervals)  Given \( x \in \mathbb{R} \), let \( I_n(x) \) be the unique interval of the form \([k2^{-n}, (k + 1)2^{-n}) \) (for some \( k \in \mathbb{Z} \)) containing \( x \). Then for any Radon measure \( \mu \) on \( \mathbb{R} \), all \( f \in L^1(\mu) \) and \( \mu \)-a.e. \( x \in \mathbb{R} \),
\[ \lim_{n \to \infty} \frac{1}{\mu(I_n(x))} \int_{I_n(x)} f \, d\mu = f(x). \]  
(10.37)

**Proof.** By localizing \( \mu \) suitably, we may assume that \( \mu \) is finite. Consider the \( \sigma \) algebras
\[ \mathcal{F}_n := \sigma \left( \left\{ [k2^{-n}, (k + 1)2^{-n}) : k \in \mathbb{Z} \right\} \right). \]  
(10.38)
Then \( \{\mathcal{F}_n\}_{n \geq 1} \) is a filtration. Moreover,
\[ \frac{1}{\mu(I_n(x))} \int_{I_n(x)} f \, d\mu = f_{\mathcal{F}_n}(x) \]  
(10.39)

Preliminary version (subject to change anytime!)  
Typeset: March 3, 2020
whenever $\mu(I_n(x)) > 0$. (Otherwise, we may interpret both sides as zero.) Theorem 10.10 thus shows that, for $\mu$-a.e. $x \in \mathbb{R}$,

$$\lim_{n \to \infty} \frac{1}{\mu(I_n(x))} \int_{I_n(x)} f \, d\mu = f_{\mathcal{F}_\infty}(x)$$

(10.40)

It remains to prove that the right-hand side equals $f$ $\mu$-a.e. This follows by noting that the dyadic intervals generate all open sets and so all Borel sets. Hence $\mathcal{F}_\infty = \mathcal{B}(\mathbb{R})$ and so $f_{\mathcal{F}_\infty} = f$ $\mu$-a.e. $\square$