D I P L O M A R B E I T

Martingale Convergence Avoiding the Upcrossing Inequality

Ausgeführt am Institut für
Statistik und Wahrscheinlichkeitstheorie
der Technischen Universität Wien

unter der Anleitung von
Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Karl Grill

durch
Michael Hofbauer-Tsiflakos

Wien, am 01. März 2010

For my loving parents...

I will never stop saying and overemphasizing how fortunate I am to have parents like I do. For me, it's only right and natural to claim that without my mum and dad I wouldn't be the person I am now. Thank you mum and dad for making me proud of the human being I see when I look in the mirror.

Thank you for all the love. I love you both!

I also want to thank Prof. Grill, who supervised my thesis that I actually suprised him with out of nowhere. His nature somehow took my motivation on a different level. In addition to that I want to seize the opportunity to shortly highlight how important and rare a joint productive atmosphere is in explaining my working relationship with Prof. Grill: Imagine Clint Eastwood in one of his western movies riding with a younger clone of himself. Of course the clone still needs to learn and Eastwood, because of his smooth behaviour, disseminates his wisdom in a way that makes the developing process for the clone even more enjoyable than it already is. Being mentally linked together and picking up each others lines both seem to get the best out of it.

Martingale convergence avoiding the upcrossing inequality

Michael Hofbauer-Tsiflakos

ABSTRACT The covering of martingale convergence in todays probability textbooks begins justifiably with the formulation of the powerful upcrossing inequality. It is *the* tool to prove almost sure convergence for $\mathcal{L}^1$-bounded martingales. Once this is established one may impose stronger porperties, containing $\mathcal{L}^1$-boundedness, such as uniform integrability, on martingales, in order to achieve additional convergence results.
This thesis shows that it is possible to prove almost sure convergence for $\mathcal{L}^1$-bounded martingales in three different ways without using the upcrossing inequality at all. The fact of various intakes unveils the possibility to derive a deeper understanding of martingale theory and its connection with other mathematical branches.

Contents

1 Prelude

The first section is divided in two parts: The first exams convergence of uniform integrable sequences of random variables, the second discusses the issue of stopping times. Both are of great importance for the martingale theory to follow. Most of the information gathered here does not appear in undergradued courses and thus needs to be mentioned explicitly. In order to keep the interested reader focused, we will omit *everydays* measure and probability theory knowledge and, if required, refer to it in the appendix.

1.1 Notation and Preassumptions

Every math writer, acknowledged or not, develops his/her own semantic style while absorbing literature. Even though the majority of mathematicians understand the correspondence between, say, X_n *converges to* X*, a.e. for* $\omega \in \Omega$ and $X_n \xrightarrow{a.s.} X$, it adds a personal note using one or the other. Consequently, we want to outline some expressions and assumptions that will be used throughout this thesis.

Conventions
Unless otherwise stated, all random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Mark the bold letter $\mathbf{E}$ as the expected value operator.

Sometimes a sequence of random variables or a stochastic process $(X_n)_{n\in\mathbb{N}}$ maybe alternately noted as $(f_n)_{n\in\mathbb{N}}$, if, for instance, the theorem using $(f_n)_{n\in\mathbb{N}}$ is of measure theoretic origin.

Let $(f_n)_{n\in\mathbb{N}} \in \mathcal{H}$ be a sequence of random variables having property A. We then say: A holds for f_n, $\forall\ n \in \mathbb{N}$, $\Leftrightarrow$ A holds for f_n, $\forall\ f_n \in \mathcal{H}$.

The notation $\mathcal{L}^p$, $1 \leq p \leq \infty$, is always an abbreviation of $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$.

Denote by $(\mathcal{L}^p, \|\cdot\|_p)$, the seminormed vector space of p^{th} power integrable functions and $(L^p, \|\cdot\|_p)$ as the resulting normed vector space by defining $L^p := \mathcal{L}^p/N$ and $N := \ker(\|\cdot\|_p)$.

If $\mathcal{G}$ is a (sub-)σ-Algebra we write $\mathcal{L}^p(\mathcal{G})$ to emphasize the underlying (sub-)σ-Algebra.

Convergence Distinctions
We say a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ *converges in (the strong sense in)* $\mathcal{L}^p$ to X if, $X_n, X \in \mathcal{L}^p$, $\forall\ n \in \mathbb{N}$, and

$$\lim_{n\to\infty} \|X_n - X\|_p \to 0.$$

We say a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ *converges in the p-norm* if, $X_n, X \in \mathcal{L}^p$, $\forall\ n \in \mathbb{N}$, and

$$\lim_{n\to\infty} \|X_n\|_p = \|X\|_p.$$

1.2 Uniform Integrability

A property of a set of random variables $\mathcal{H}$ that will lead to further proofs is uniform integrability.

Definition 1.1 (Uniform Integrability 1)**.** Let $\mathcal{H}$ be a subset of the space $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$. One says that $\mathcal{H}$ is a *uniformly integrable* collection of random variables, if

$$\lim_{c\to\infty} \sup_{f\in\mathcal{H}} \int_{\{|f|\geq c\}} |f| \ d\mathbf{P} \to 0.$$

Another possible definition of uniform integrability requires the notation of certain parts of a random variable f:

$$f^c(\omega) = \begin{cases} f(\omega) & \text{for} \quad |f(\omega)| \leq c \\ 0 & \text{for} \quad f(\omega) > c \\ 0 & \text{for} \quad f(\omega) < -c \end{cases}$$

We put $f_c = f - f^c$ and achieve

Definition 1.2 (Uniform Integrability 2)**.** $\mathcal{H}$ is *uniformly integrable* if and only if for every $\varepsilon > 0$, exists a number c such that $\|f_c\|_1 < \varepsilon$ for every function $f \in \mathcal{H}$.

Example 1.3. Since $\mathbf{E}[|f_i|] < \infty$, $i = \{1, \ldots, n\}$, each finite sequence $f_1, \ldots, f_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ fulfills the hypothesis of uniform integrability.

Example 1.4. Suppose that $(X_n)_{n\geq 1}$ is a sequence of random variables, with a random variable $Y \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P}), 1 \leq p < \infty$, dominating the X_n, i.e., $|X_n| \leq Y$. Then $\{|X_n|^p\}$ is uniformly integrable as a consequence of the dominating convergence theorem of Lebesgue [A.1].
Alternatively, one could show this result by the following chain:
$|X_n| \leq Y \Rightarrow |X_n|^p \leq Y^p \Rightarrow [|X_n| \leq c] \subseteq [Y \leq c] \Rightarrow \int_{\{|X_n| \leq c\}} |X_n|^p \ d\mathbf{P} \leq \int_{\{Y\leq c\}} Y^p \ d\mathbf{P} < \varepsilon$, for c large enough.

Throughout this section we will derive additional conditions, necessary for later martingale proofs, that will also help us to show **E1.4** in a different way. The next theorem states an equivalence assertion to uniform integrability:

Theorem 1.5. *Let $\mathcal{H}$ be a subset of $\mathcal{L}^1$. Then $\mathcal{H}$ is uniformly integrable if and only if the following conditions are realized:*
(a) The expectations $\mathbf{E}[|f|]$, $f \in \mathcal{H}$, are uniformly bounded.
(b) $\forall\ \varepsilon > 0, \exists\ \delta > 0$, so that $\forall\ A \in \mathcal{F}$ with $\mathbf{P}(A) \leq \delta$

$$\int_A |f| \ d\mathbf{P} \leq \varepsilon, \ \forall\ f \in \mathcal{H}.$$

Proof. $\Rightarrow$:
To establish the necessity of conditions *(a)* and *(b)*, we note that, for every integrable function f and every set $A \in \mathcal{F}$:

$$\int_A |f| \ d\mathbf{P} \leq c\mathbf{P}(A) + \mathbf{E}[|f_c|].$$

Supposing that $\mathcal{H}$ is uniformly integrable, choose c large enough so that

$$\mathbf{E}[|f_c|] < \varepsilon/2, \ \forall \ f \in \mathcal{H}. \tag{1.1}$$

Set $A = \Omega$ in (1.1) and (a) follows. By choosing $\delta = \varepsilon/2c$ we have proven the validity of *(b)*.
$\Leftarrow$:
Conversely, supposing that properties *(a)* and *(b)* are satisfied and given $\varepsilon > 0$, associate with it some $\delta > 0$ satisfying *(b)* and take $c = \sup_{f\in\mathcal{H}} \mathbf{E}[|f|]/\delta$, a finite quantity in view of *(a)*. Apply formula (1.1), taking for A the set $\{|f| \geq c\}$, whose probability is less than δ in view of the inequality

$$\mathbf{P}[|f| \geq c] \leq \frac{1}{c}\mathbf{E}[|f|].$$

We obtain the inequality

$$\int_{\{|f|\geq c\}} |f| \ d\mathbf{P} \leq \varepsilon, \ \forall \ f \in \mathcal{H},$$

and $\mathcal{H}$ is thus uniformly integrable.

□

Remark 1.6. Point *(b)* in **T1.5** is nothing else but the *absolute continuity* property, in a probability space, stated for the function f.

Remark 1.7. One be ensured that $\lim_{c\to\infty} \mathbf{P}[f \geq c] = 0$ does not necessarily yield uniform boundedness of $\mathbf{E}[|f|]$: Suppose $X \geq 1$ and $\mathbf{P}[X \geq x] = \frac{1}{x}$. The distribution function of this example is $F_X(x) = 1 - \frac{1}{x}$ and hence $\mathbf{E}[|X|] = \int_1^\infty \mathbf{P}[X \geq x] \ dx = \int_1^\infty \frac{1}{x} \ dx = \ln x \ |_1^\infty = \infty$, but $\lim_{x\to\infty} \mathbf{P}[X \geq x] = 0$.

Obviously, uniform integrability encapsules uniform boundedness. We can use this information to derive a generalization of Lebesgue's dominating convergence theorem [A.1], namely, the convergence theorem of Vitali. Before, we first state

Lemma 1.8. *Let* $\mathcal{H} \subseteq \mathcal{L}^p$, $1 \leq p < \infty$. *Then* $\sup_{f\in\mathcal{H}} \|f\|_p < \infty$ *if and only if* $\forall \ \varepsilon > 0, \ \exists \ E \in \mathcal{F}$ *with* $\int_{E^c} |f|^p \, d\mathbf{P} < \varepsilon, \ \forall \ f \in \mathcal{H}$.

Proof. $\Rightarrow$:
Choose $E := \{f \le c\}$ with $c > 0$ big enough so that $\int_{\{|f|>c\}} |f|^p \, d\mathbf{P} < \varepsilon$.
$\Leftarrow$:
Take $E^c = \emptyset$, then $\int_{E^c} |f|^p \ d\mathbf{P} = 0$, $\forall\ f \in \mathcal{H}$, and hence $\sup_{f\in\mathcal{H}} \|f\|_p < \infty$.
□

For $p = 1$, $\sup_{f\in\mathcal{H}} \|f\|_p < \infty$, states condition **T1.5***(a)*. Determined to give a neat proof of the Vitali theorem we interwine one of Riesz' theorems.

Theorem 1.9 (Riesz 1928)**.** *Let* $(f_n)_{n\ge1}, f \in \mathcal{L}^p$, $1 \le p < \infty$*. Then the following two assertions are equivalent:*
(a) $\lim_{n\to\infty} \|f_n - f\|_p \to 0$.
(b) $f_n \xrightarrow{\mathbf{P}} f$ and $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$.

Proof. $\Rightarrow$:
By the inequality chain

$$\begin{aligned}\mathbf{P}(\{|f_n - f| \ge \epsilon\}) &\le \int_{\{|f_n - f|\ge\epsilon\}} \left|\frac{f_n - f}{\varepsilon}\right|^p d\mathbf{P} \\ &\le \int_\Omega \left|\frac{f_n - f}{\varepsilon}\right|^p d\mathbf{P} = \varepsilon^{-p}\|f_n - f\|_p \xrightarrow{n\to\infty} 0,\end{aligned}$$

convergence in probability of f_n to f follows.
For $1 \le p < \infty$ the lower triangle inequality gives

$$\big|\|f_n\|_p - \|f\|_p\big| \le \|f_n - f\|_p$$

and therefore $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$.
$\Leftarrow$:
One way to show this is to use *Pratt's theorem* (1960). More on this has, i.e., Elstrodt [13, p. 261]. For our purposes this direction is void.
□

Note that **T1.9***(b)* implies **T1.9***(a)* also if $f_n \xrightarrow{a.s.} f$. The opposite does not hold.

Theorem 1.10 (Vitali 1907)**.** *Let* $(f_n)_{n\ge1}, f \in \mathcal{L}^p$, $1 \le p < \infty$*. Then the following assertions are equivalent:*
(a) $\lim_{n\to\infty} \|f_n - f\|_p \to 0$.
(b) *(I)* $f_n \xrightarrow{\mathbf{P}} f$
(II) $\forall\ \varepsilon > 0, \exists\ E \in \mathcal{F}$ *with*

$$\int_{E^c} |f_n|^p \, d\mathbf{P} < \varepsilon,\ \ \forall\ n \in \mathbb{N}.$$

(III) $\forall\ \varepsilon > 0, \exists\ \delta > 0$*, so that* $\forall\ A \in \mathcal{F}$ *with* $\mathbf{P}(A) \le \delta$,

$$\int_A |f_n|^p \, d\mathbf{P} < \varepsilon,\ \ \forall\ n \in \mathbb{N}.$$

A sequence of functions $(f_n)_{n\geq 1} \in \mathcal{L}^p$ with properties *(II)* and *(III)* is said to be uniformly integrable in the p-norm. According to **T1.5** and **L1.8** we have encountered this case for $p = 1$ already.

Proof. $\Rightarrow$:
Let us suppose first that f_n converges to f in L^p. We take advantage of the inequality $|f_n|^p - |f|^p \leq |f_n - f|^p$ to achieve

$$\|f_n\|_p \leq \|f\|_p + \|f_n - f\|_p, \ \forall \ n \in \mathbb{N}.$$

As it appears $\sup_{n\in\mathbb{N}} \|f_n\|_p < \infty$ and **L1.8** implies *(II)*. On the other hand, let us choose an integer N such that $\|f_n - f\|_p \leq \varepsilon/2$ for every $n > N$, and a number δ such that the inequality $\mathbf{P}(A) \leq \delta$ implies $\int_A |g|^p \ d\mathbf{P} \leq \varepsilon/2$, when g ranges over the finite collection consisting of the functions $\{f_1, \ldots, f_N, f\}$. It follows then that

$$\int_A |f_n|^p \ d\mathbf{P} \leq \int_A |f|^p \ d\mathbf{P} + \int_A |f_n - f|^p \ d\mathbf{P} \leq \varepsilon,$$

for every n when $\mathbf{P}(A)$ is less than δ, and condition *(III)* is verified.
$\Leftarrow$:
Conversely, let at first f_n converge to f almost surely. For every $\varepsilon > 0$ we choose an $E \in \mathcal{F}$ according to *(II)* and a $\delta > 0$ according to *(III)*. Due to Egorov's theorem [A.3] there exists a measurable set $B \subset E$ with $\mathbf{P}(E \setminus B) < \delta$, so that $(f_n|B)_{n\geq 1}$ converges uniformly to $f|B$.
Applying inequality

$$|f + g|^p \leq (2 \max(|f|, |g|))^p \leq 2^p(|f|^p + |g|^p)$$

we estimate

$$\int_\Omega |f_n - f|^p \, d\mathbf{P} \leq$$
$$2^p \int_{E^c} (|f_n|^p + |f|^p) \ d\mathbf{P} + 2^p \int_{E\setminus B} (|f_n|^p + |f|^p) \ d\mathbf{P} + \int_B |f_n - f|^p \ d\mathbf{P}. \tag{1.2}$$

Due to Fatou's lemma [A.4]

$$\int_{E^c} |f|^p \, d\mathbf{P} \leq \liminf_{n\to\infty} \int_{E^c} |f_n|^p \, d\mathbf{P} \leq \varepsilon,$$

$$\int_{E\setminus B} |f|^p \, d\mathbf{P} \leq \liminf_{n\to\infty} \int_{E\setminus B} |f_n|^p \, d\mathbf{P} \leq \varepsilon.$$

The first two members on the right side of (1.2) are together $< 2^{p+1}\varepsilon$. Since $(f_n|B)_{n\geq 1}$ converge uniformly to $f|B$, the third member is also $< \varepsilon$ for every $n \geq n(\varepsilon)$, and so *(a)* follows.

Now let f_n converge to f in probability only. Suppose that there is a $\delta > 0$ and a subsequence $(f_{n_k})_{k\geq 1}$ with

$$\|f_{n_k} - f\|_p \geq \delta, \; \forall \; k \in \mathbb{N}. \tag{1.3}$$

We assume that f_{n_k} converges *a.s.* to f since every subsequence f_{n_k}, for every sequence f_n that converges in probability to f, has an *a.s.* convergent subsequence to the same limit [A.5]. Using the same steps for f_{n_k} as we did before for the *a.s.* convergent sequence f_n we achieve

$$\lim_{k\to\infty} \|f_{n_k} - f\|_p \to 0,$$

in contrary to (1.3) and the theorem is proven.

□

Like indicated before, we can use **T1.10** to reverse Lebesgue's theorem [A.1]: Knowing that a sequence f_n of integrable random variables converges in $\mathcal{L}^p$, leads to the property of uniform integrability for the f_n and thus to uniform boundedness.

1.2.1 The La Valleé Poussin Theorem

Coming up next is the theorem of La Valleé Poussin, which clarifies another equivalent condition for a set of functions $\mathcal{H}$ being uniformly integrable. This theorem comes with a conventional proof given by Meyer in [29, p. 19]. An alternative, shorter proof can be found in [B.1].

Theorem 1.11 (La Valleé Poussin 1937)**.** *Let $\mathcal{H}$ be a subset of $\mathcal{L}^1$. The following properties are equivalent:*
(1) $\mathcal{H}$ is uniformly integrable.
(2) There exists a function $G(t)$ defined on $\mathbb{R}^+$, which is positive, increasing and convex[1]*, such that*

$$\lim_{t\to+\infty} \frac{G(t)}{t} = +\infty,$$

and

$$\sup_{f\in\mathcal{H}} \mathbf{E}[G \circ |f|] < +\infty. \tag{1.4}$$

Proof. (2) ⇒ (1):
Given an $\varepsilon > 0$, put $a = M/\varepsilon$, where $M := \sup_{f\in\mathcal{H}} \mathbf{E}[G \circ |f|]$. Choose a number c so large that $G(t)/t \geq a$ for $t \geq c$.
We then have $|f| \leq (G \circ |f|)/a$ on the set $\{|f| \geq c\}$ and consequently

$$\int_{\{|f|\geq c\}} |f| \; d\mathbf{P} \leq \frac{1}{a} \int_{\{|f|\geq c\}} (G \circ |f|) \; d\mathbf{P} \leq \frac{1}{a} M = \varepsilon$$

[1] Convexity of f is not used in the proof

for each function $f \in \mathcal{H}$. **D1.2** is thus verified.
(1) ⇒ (2):
We establish the converse implication by constructing a function $G(t)$ of the form $\int_0^t g(u)\ du$, where g denotes an increasing function, tending to $+\infty$ with t, which has a constant value g_n on each interval $[n, n+1)$, $n \in \mathbb{N}$. Put, for each $f \in \mathcal{H}$,

$$a_n(f) = \mathbf{P}\left\{|f| > n\right\}$$

Let us take $g_0 = 0$; we have

$$\begin{aligned}\mathbf{E}[G \circ |f|] &= \int_0^\infty G(|f|) d\mathbf{P} \\ &= \sum_{n=1}^{\infty} \int_{\{n \le |f| < n+1\}} g(|f|)\mathbf{P}[n \le |f| < n+1], \text{ set } g_n := \int_{\{n \le |f| < n+1\}} g(|f|), \\ &\le g_1 \mathbf{P}[1 \le |f| < 2] + (g_1 + g_2)\mathbf{P}[2 \le |f| < 3] + \ldots = \sum_{n=1}^{\infty} g_n a_n(f).\end{aligned}$$

It remains to show that one may choose coefficients g_n, which tend to infinity with n, such that the sum $\sum_n g_n a_n(f)$ is uniformly bounded. Choose a sequence c_n, which increases to infinity, such that

$$\sup_{f \in \mathcal{H}} \int_{\{|f| \ge c_n\}} |f|\ \ d\mathbf{P} \le 2^{-n},$$

which is possible by the virtue of uniform integrability. Then

$$\begin{aligned}2^{-n} &\ge \sup_{f \in \mathcal{H}} \int_{\{|f| \ge c_n\}} |f|\ \ d\mathbf{P} \\ &\ge \int_{\{c_n < |f| \le c_n + 1\}} c_n\ d\mathbf{P} + \int_{\{c_n + 1 < |f| \le c_n + 2\}} c_n + 1\ d\mathbf{P} + \ldots \\ &= \sum_{m=c_n}^{\infty} m\mathbf{P}[m < |f| \le m+1] \ge \sum_{m=c_n}^{\infty} \mathbf{P}[|f| > m] = \sum_{m=c_n}^{\infty} a_m(f).\end{aligned}$$

It follows that the sum $\sum_n \sum_{c_n}^\infty a_m(f)$ is uniformly bounded for $f \in \mathcal{H}$. But this sum is of the form $\sum_m g_m a_m(f)$, g_m denoting the number of integers n such that $c_n \le m$. This proves the assertion.

□

Remark 1.12. The fact, that the sums $\sum_n \sum_{c_n}^\infty a_m(f)$ and $\sum_m g_m a_m(f)$, are of the same form might not be obvious for the inexperienced eye.

To verify this relationship we write the addends of the first sum explicitly

$$
\begin{aligned}
&\big(\mathbf{P}[|f| > c_1] + \ldots + \mathbf{P}[|f| > c_1 + i] + \ldots\big) + \ldots \\
&. \\
&. \\
&. \\
&\big(\mathbf{P}[|f| > c_k] + \ldots + \mathbf{P}[|f| > c_k + j] + \ldots\big) + \ldots, \;\; i, k, j \in \mathbb{N}
\end{aligned}
$$

Now we pick out the integers of the sequence c_n that satisfy the equality $c_n = 1, n \in \mathbb{N}$. Let's say there are $g_{1,1}$ of them and thus $\mathbf{P}[|f| > 1]$ appears $g_{1,1}$ times. In addition to this we add all the integers of the sequence c_n that obey the inequality $c_n < 1$, say $g_{1,2}$, to $g_{1,1}$ and set $g_1 := g_{1,1} + g_{1,2}$. The reader might notice that this turn was a bit inaccurately, because although we have g_1 times $\mathbf{P}[|f| > 1]$, substituting every number in the set of integers that fulfills $c_n < 1$ with 1 in the probability brackets, would result altogether in a higher probability. But this is not crucial for the convergence of our important sum. It should just display its setup.

Remark 1.13. A function $G(t)$, that satisfies the condition of **T1.11**, is t^p. Applying it to (1.4), we see that every bounded subset $\mathcal{H} \subset \mathcal{L}^p$, $p > 1$, is uniformly integrable.

Remark 1.14. Reconsider **E1.4**, where a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ was bounded by an element of $\mathcal{L}^p$, $1 \leq p < \infty$, $\forall\, n \in \mathbb{N}$. Due to $|X_n| \leq Y$, $\forall\, n \in \mathbb{N}$, it follows by an easy transformation that $\sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^p] \leq \mathbf{E}[Y^p]$. For $p \geq 2$, with $G(t) = t^p$, this satisfies condition *(2)* of **T1.11** and the sequence $(X_n)_{n\in\mathbb{N}}$ is thus uniformly integrable.

Example 1.15. An example of a $\mathcal{L}^1$-bounded subset of functions that is not uniformly integrable would be $\mathcal{H} := \{f_n = n\mathbf{1}_{[0,\frac{1}{n}]}\}$. For $n > c$, it is

$$\int_{\{|f_n|>c\}} |f_n|\, d\mathbf{P} = 1,$$

and obviously

$$\lim_{c\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|f_n|>c\}} |f_n|\, d\mathbf{P} = 1.$$

A contradiction to definition **D1.2** of uniform integrability.

1.2.2 The Dunford-Pettis Theorem

Theorem **T1.16**, also referred to as *the compactness criterion of Dunford-Pettis*, reveils a convergence property of a subset $\mathcal{H} \subset \mathcal{L}^1$ just by using its property of uniform integrability.

Theorem 1.16 (DUNFORD-PETTIS 1953). *Let $\mathcal{H}$ be a subset of the space L^1. The following properties are equivalent:*
(1) $\mathcal{H}$ is uniformly integrable.
(2) $\mathcal{H}$ is relatively compact in L^1 in the weak topology $\sigma(L^1, L^\infty)$.
(3) Every sequence of elements of $\mathcal{H}$ contains a subsequence that converges in the sense of the topology $\sigma(L^1, L^\infty)$.

We will prove the implications of *(1)* $\Rightarrow$ *(2)* and *(2)* $\Rightarrow$ *(3)*, because the opposite directions are not relevant for the results to come.

Proof. *(1)* $\Rightarrow$ *(2)*:
Let $\mathfrak{A}$ be an ultrafilter on $\mathcal{H}$. For each function $f \in \mathcal{H}$, each set $E \in \mathcal{F}$, put

$$I_f(E) = \int_E f \; d\mathbf{P}.$$

From the relation $|I_f(E)| \leq \mathbf{E}[|f|]$, and condition *(a)* of **T1.5**, the numbers $I_f(E)$ are uniformly bounded. The limit

$$I(E) = \lim_{\mathfrak{A}} I_f(E)$$

thus exists for every $E \in \mathcal{F}$. Since $I(\bigcup_n A_n) = \sum_n I(A_n)$, for disjoint A_n, I is a measure. Condition **T1.5***(b)* says that $\forall\, \varepsilon > 0, \exists\, \delta > 0$, such that $\mathbf{P}(E) \leq \delta$ implies $|I(E)| < \varepsilon$; We obtain an absolutely continuous measure I with respect to $\mathbf{P}$. By the theorem of Radon-Nikodym [A.8], there exists a function $\phi \in L^1$ such that for every measurable subset E:

$$I(E) = \int_E \phi \; d\mathbf{P}.$$

Now

$$\lim_{\mathfrak{A}} \int_E f \; d\mathbf{P} = \int_E \phi \; d\mathbf{P},$$

holds and assertion *(2)* will be established if we show that $\mathfrak{A}$ converges to ϕ in the weak topology. Evidently

$$\lim_{\mathfrak{A}} \mathbf{E}[f \cdot g] = \mathbf{E}[\phi \cdot g], \; \forall \; g \in \mathcal{L}^\infty.$$

The function $g \in \mathcal{L}^\infty$ is a finite linear combination of indicator functions of sets. The assertion is proven if we remark that $g \in \mathcal{L}^\infty$ is a uniform limit of such functions.
(2) $\Rightarrow$ *(3)*:
Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of elements of $\mathcal{L}^1$, whose classes belong to $\mathcal{H}$. Denote by $\mathcal{T}$ the σ-field generated by the functions f_n, and by $\mathcal{T}_0$ the smallest collection of subsets of Ω, closed under $(\bigcup, ^{\mathrm{C}})$, which contain the sets of the form $\{f_n < a : n \in \mathbb{N}, a \in \mathbb{Q}\}$.

It is easily verified that $\mathcal{T}_0$ is a countable collection, which generates the σ-field $\mathcal{T}$, i.e., $\mathcal{T} = \sigma((f_n)_{n\in\mathbb{N}}) = \sigma(\{f_n < a, n \in \mathbb{N}, a \in \mathbb{Q}\}) = \sigma(T_0)$. By means of the diagonal procedure, extract from the sequence (f_n) a subsequence $(f_{n_k})_{k\in\mathbb{N}}$, such that the integrals

$$\int_E f_{n_k}\ d\mathbf{P},\ E \in \mathcal{T}_0$$

have a limit when $k \to \infty$. We shall show that the sequence (f_{n_k}) is weakly convergent. It suffices for this, according to assertion *(2)*, to show that this sequence has a single limit point in $\mathcal{L}^1$. Let ϕ and ϕ^* be two limit points. These two functions are $a.s.$ equal to $\mathcal{T}$-measurable functions and in order to establish their $a.s.$ equality, it thus suffices to show, due to [A.2], that

$$\int_E \phi\ d\mathbf{P} = \int_E \phi^*\ d\mathbf{P},\ \forall\ E \in \mathcal{T}.$$

Now this equality holds for $E \in \mathcal{T}_0$. Denote by $\mathcal{M}$ the collection of subsets $E \in \mathcal{T}$ for which this equality is true. It follows from Lebesgue's theorem that $\mathcal{M}$ is closed under passage to the monotone limits, and from [A.14] that $\mathcal{M} = \mathcal{T}$. Thus $\phi = \phi^*$ $a.s.$, and the theorem is established.

□

Remark 1.17. The original proof of **T1.16** was given by Dunford and Pettis in [12, p. 458] and uses results of weakly convergent Cauchy-sequences in Banach-spaces. To reformulate it here would definitely exceed the frame of this thesis. This much shorter proof here is due to Meyer [29, p. 20]. The appendix includes a more modern proof [B.3], which can be also found in Kallenberg [23, p. 46].

We will conclude this section with a theorem about the convex hull of a uniform integrable set $\mathcal{H}$.

Theorem 1.18. Let $\mathcal{H}$ be a uniformly integrable subset of L^1. Its closed convex hull is also uniformly integrable.

Proof. Let $f, g \in \mathcal{H}$. Set $C := \sup_{f\in\mathcal{H}} \int_\Omega |f|\, d\mathbf{P}$ and derive

$$\mathbf{P}(|f| > c) \leq \frac{C}{c},\ \forall\ f \in \mathcal{H}.$$

Due to the uniform integrability of f, there exists $\forall\ \varepsilon > 0$ a $\delta > 0$, so that $\forall\ A \in \mathcal{F}$ with $\mathbf{P}[A] < \delta$:

$$\int_A |f|\ d\mathbf{P} < \varepsilon.$$

We choose $\mathbf{P}[|f| > c] < \frac{\delta}{2} \wedge \mathbf{P}[|g| > c] < \frac{\delta}{2}$ and observe that

$$[|\alpha f + (1-\alpha)g| > c] \subseteq [\alpha\,|f| + (1-\alpha)\,|g| > c] \subseteq [|f| > c] \cup [|g| > c],$$

and

$$\mathbf{P}([|f| > c] \cup [|g| > c]) < \delta.$$

This implies

$$\int_{\{|\alpha f+(1-\alpha)g>c|\}} |f| \; d\mathbf{P} < \varepsilon \;\wedge\; \int_{\{|\alpha f+(1-\alpha)g>c|\}} |g| \; d\mathbf{P} < \varepsilon \Rightarrow$$
$$\int_{\{|\alpha f+(1-\alpha)g>c|\}} |\alpha f + (1-\alpha)g| \; d\mathbf{P} < 2\varepsilon.$$

□

1.3 Stopping Times

The concept of stopping times gives us the power to filter certain information out of a process $(X_n)_{n\in\mathbb{T}}$. Before we give an initial explanation of what exactly a stopping time is, we need to equip ourselves with a few *new* measure theoretic terms. Our main results in later sections are proven for the discrete index set $\mathbb{N}$, which will also be the index set of choice for stopping times here. Let $(\Omega, \mathcal{F})$ be a measurable space and let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be a family of sub-σ-fields of $\mathcal{F}$ such that the relation $s \leq t$ implies $\mathcal{F}_s \subset \mathcal{F}_t$. We say that $(\mathcal{F}_t)$ is an *increasing family of sub-σ-fields of* $\mathcal{F}$, and we call $\mathcal{F}_t$, for each $t \in \mathbb{N}$, the σ-fields of *events prior to t.* An increasing family of sub-σ-fields $(\mathcal{F}_t)$ is oftenly also called a *filtration* of $\mathcal{F}$ on a given index set $\mathbb{T}$.

Definition 1.19. Let $(X_t)_{t\in\mathbb{N}}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be a filtration of $\mathcal{F}$. The process (X_t) is said to be adapted to the family $(\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable for every $t \in \mathbb{N}$.

The smallest filtration with respect to which X_t can be adapted is the induced filtration $\sigma(\{X_s : s \leq t\})$.
The situation of a gambler gives an intuitively description of stopping times. Suppose now that we interpret the index set $\mathbb{T}$ as a time set and each event $A \in \mathcal{F}$ as a possible event to occur. Say that the gambler awaits a certain event with agony and notes the time $T(\omega)$ when the event appears for the first time. The set $\{\omega : \ T(\omega) \leq t\}$ can then be paraphrased as the event, which occurs if and only if the gamblers wish takes place at least one time at the instant t. From this comes the following definition.

Definition 1.20. Let $(\Omega, \mathcal{F})$ be a measurable space, and let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be a filtration of $\mathcal{F}$. A positive random variable T defined on Ω is said to be a stopping time of the family $(\mathcal{F}_t)$ if T satisfies the following property:

$$\{T = t\} \in \mathcal{F}_t, \ \forall \ t \in \mathbb{N}.$$

Retrieving the gamblers example we could say that the desired events that have to take place before a time instant t do surely belong to the σ-field $\mathcal{F}_t$, which contains all the possible events that have occured prior to time t.

Remark 1.21. Every random variable equal to a constant is a stopping time.

Definition 1.22. Let T be a stopping time relative to the family of σ-fields $(\mathcal{F}_t)_{t\in\mathbb{N}}$. We denote by $\mathcal{F}_T$ the collection of events $A \in \mathcal{F}_\infty$, where $\mathcal{F}_\infty$ is defined as $\sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$, such that

$$A \cap \{T = t\} \in \mathcal{F}_t, \ \forall \ t \in \mathbb{N}.$$

It is easy to verify that $\mathcal{F}_T$ constitutes a σ-field and for $T = t$ the σ-field $\mathcal{F}_t$ is recovered. One could associate $\mathcal{F}_T$ with the set of events known at time T.

Lemma 1.23. *Let S and T be two stopping times. Then*
(a) $S \wedge T$, $S \vee T$ and $S + T$ are stopping times.
(b) T is $\mathcal{F}_T$ measurable.

Proof. Straight forward.

□

The exploration of stopping times continuous with a few lemmatas.

Lemma 1.24. *Let T be a stopping time and S an $\mathcal{F}_T$-measurable random variable such that $S \geq T$. S is then a stopping time.*

Proof. Since $\{S \leq t\} \subseteq \mathcal{F}_T$, $\{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t$. Note that the intersection is equal to $\{S \leq t\}$.

□

Lemma 1.25. *Let S and T be two stopping times and let A be an element of $\mathcal{F}_S$. We then have*

$$A \cap \{S \leq T\} \in \mathcal{F}_T.$$

Proof. In order to verify that

$$A \cap \{S \leq T\} \cap \{T \leq t\} \in \mathcal{F}_T, \ \forall\, t \in \mathbb{N},$$

it suffices to write the left-hand side in the form

$$[A \cap \{S \leq t\}] \cap \{T \leq t\} \cap \{S \wedge t \leq T \wedge t\}\,.$$

Each of these three events belongs to $\mathcal{F}_t$. The first, by reason of the relation $A \in \mathcal{F}_S$; the second, from the fact that T is a stopping time; the third, finally, follows from the fact that $S \wedge t$ and $T \wedge t$ are $\mathcal{F}_t$-measurable.
□

Lemma 1.26. *Let S and T be two stopping times such that $S \leq T$. We then have $\mathcal{F}_S \subset \mathcal{F}_T$*

Proof. Let $A \in \mathcal{F}_S$. From **L1.25** we have

$$A = A \cap \{S \leq T\} \in \mathcal{F}_T.$$

□

A very important example of stopping times is that of first hitting times. A profitable gambler is interested at which time his budget will increase to a desired amount a, rather than asking how often his budget will raise and drop to a in one night.

Theorem 1.27. *Let (X_t) be a real-valued stochastic process and assume that every $\mathcal{F}_t$ is complete. Then, for every Borel-subset A,*

$$D_A(\omega) = \begin{cases} \inf\{t : X_t(\omega) \in A\} \\ +\infty, \ \textit{if the above set is empty} \end{cases}$$

is a stopping time.

Proof. We encounter no difficulties in the discrete case. □

In the first case of **T1.27** D_A is bounded stopping time. If the event A never occurs the stopping time D_A becomes unbounded. Most results on stopping times connected with stochastic processes are stated for bounded stopping times first. It requires further properties of a stochastic process to prove the same results for unbounded stopping times.

Definition 1.28. A stochastic process (X_t) stopped at time T is denoted as X^T. Analogously $X_n^T = X_{n \wedge T}$.

We call the readers attention not to mix the two notations $X_T \neq X^T$ up.

2 Martingales

The term martingale corresponds to a stochastic process with *certain* properties. Compared to other probability topics, the field of martingale theory, like its name, is rather new. The term martingale first surfaced in a work of Ville 1939 [37] who refers to systems of play as martingales. Almost ten years later, 1948, the mathematical probability community began to use the term martingale consistently. The first people studying sequences with martingale properties were Bernstein, Lévy and Kolmogorov in the 1930s. The so-called *father* of martingale theory is Joseph Leo Doob[2]. He revolutionised probability theory with his book *Stochastic Processes* [11], that was first published in 1953. He took advantage of the martingale process to generalize and reformulate probabilistic results like *Kolmogorov's inequality*, the *strong law of large numbers* or the *lemma of Borel-Cantelli.* In the same manner he invented a martingale framework, which is still the number one reference for any serious person who wants to pursue investigation on martingales. Through the last decades martingale theorey has been immensely developed in theoretic and practical aspects. Though, it shouldn't be a suprise that a huge quantity of probability textbooks have already covered the fundamental principles of martingales. But, what they all have in common is that their approach to main convergence results follow the same path. We will take advantage of the general properties of a martingale and state different proofs for various theorems.
In order to understand them completely we have to get through the basics of martingale theory first.

2.1 Basics

Let $\mathbb{T}$ be an ordered index set by a relation $\leq$. We will use the abbreviation X to denote the process $(X_t)_{t\in\mathbb{T}}$ and X_t for the value of the process X at the instant t. Conventionally, the process will be defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in $\mathbb{R}$.

Definition 2.1 (Martingale)**.** Let X be a stochastic process, adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ of $\mathcal{F}$. Then X is said to be a *martingale* if the following two conditions hold:
(1) Every variable X_t is integrable: $\mathbf{E}[|X_t|] < \infty, \forall\ t \in \mathbb{T}$
(2) For every pair of elements $s,\ t \in \mathbb{T}$ such that $s \leq t$, we have

$$\mathbf{E}[X_t|\mathcal{F}_s] = X_s \quad a.s. \tag{2.1}$$

[2] Cincinnati, Ohio 27.10.1910 – 07.06.2004

Our index set $\mathbb{T}$ will from now on be linked with $\mathbb{N}$. Such a martingale is also called a discrete martingale. It should be mentioned that we don't deal with martingales X with continuous index sets, since if these martingales have the property of right-continuity the proofs of our main theorems translate flawlessly into the continuous case.

If equation (2.1) gets substituted with $\mathbf{E}[X_t|\mathcal{F}_s] \leq X_s$ (resp. $\mathbf{E}[X_t|\mathcal{F}_s] \geq X_s$), the process X is called a *supermartingale* (resp. *submartingale*).

The second condition of **D2.1** leads to alternative martingale definitions. Rewrite (2.1) with

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n. \tag{2.2}$$

Integrating this equation and using the defining property of conditional expectation gives

$$\int_A \mathbf{E}[X_{n+1}|\mathcal{F}_n]\ d\mathbf{P} = \int_A X_{n+1}\ d\mathbf{P} = \int_A X_n\ d\mathbf{P},\ \forall\ A \in \mathcal{F}_n. \tag{2.3}$$

The equation of the second two members of (2.3) is also oftenly replacing condition (2) in **D2.1**. Raising the index n by 1 in (2.2) results in

$$\int_A X_{n+2}\ d\mathbf{P} = \int_A X_{n+1}\ d\mathbf{P},\ \forall\ A \in \mathcal{F}_{n+1}. \tag{2.4}$$

Since $(\mathcal{F}_n)$ are a filtration (2.4) can be modified with (2.3) into

$$\int_A X_{n+k}\ d\mathbf{P} = \ldots = \int_A X_{n+1}\ d\mathbf{P} = \int_A X_n\ d\mathbf{P},\ \forall\ A \in \mathcal{F}_n,\ k \in \mathbb{N}, \tag{2.5}$$

which is the same as

$$\mathbf{E}[X_{n+k}|\mathcal{F}_n] = X_n. \tag{2.6}$$

Setting $A = \Omega$ gives

$$\mathbf{E}[X_1] = \mathbf{E}[X_2] = \ldots\ . \tag{2.7}$$

If we substract $\mathbf{E}[X_n|\mathcal{F}_n]$ on both sides of (2.2), use the measurability and adaptedness of X along with the linearity of the expected value we get

$$\mathbf{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0. \tag{2.8}$$

Setting $X_{n+1} - X_n =: \Delta_n$, one can derive a fourth version of the martingale property **D2.1**(2) just by using the process differences.

Remark 2.2. Due to its integrability and adaptation, a martingale (resp. supermatingale, submartingale) X is an element of $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$. However, $\lim_{n\to\infty} X_n$ may not be included in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$.

Remark 2.3. Since every X_n is $\mathcal{F}_n$-measurable and the $(\mathcal{F}_n)_{n\in\mathbb{N}}$ are nested, X_n is also measurable $\mathcal{F}$. A fact that we used in point **R2.2** tacitly.

An interesting observation refers to the application of certain functions f on a martingale X that maintains most of its properties.

Theorem 2.4. *Let X be a martingale and f a convex function defined on $\mathbb{R}$ such that $f(X)$ is integrable. Then $f(X)$ is a submartingale.*

Proof. The proof comes very easy using Jensen's inequality for conditional expectations [A.10] for convex functions:

$$f(X_n) = f(\mathbf{E}[X_{n+1}|\mathcal{F}_n]) \leq \mathbf{E}[f(X_{n+1})|\mathcal{F}_n].$$

□

For concave f in **T2.4**, $f(X)$ is a supermartingale. If X is already a supermartingale (resp. submartingale) an application of a concave (resp. convex) mapping f on X does not change this. A popular convex mapping is the absolute value $|.|^p$, $p \in [1, \infty)$.

We have come to the point to show some examples.

Example 2.5. A martingale represents the perfect role model for a fair gambling game. Suppose that X_n stands for the gamblers budget at time n. Now, if X obeys the martingale conditions, (2.6) holds and that means that the expected budget at some instant after time n must be the same as the budget at time n. Another interpretation would be that the expected gain from one time instant to the other should be 0 (2.8). The filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ can be seen as the information, for instance, possible outcomes, a gambler at a certain point of time has. Measurability of X ensures the gambler that the information is accessible. Finally, the integrability of X asserts that infinite budget cannot be expected at any time.

Example 2.6. X is a supermartingale if and only if $-X$ is a submartingale.

Example 2.7. Let X and Y be two martingales (resp. supermartingales) with the same filtration and a, b two constants (resp. non-negative constants). Then $aX + bY$ is a martingale (resp. a supermartingale) and $X \wedge Y$ is a supermartingale.

Example 2.8. Let Y be an integrable random variable and set

$$X_n = \mathbf{E}[Y|\mathcal{F}_n].$$

Then X is a martingale. Indeed, due to the law of iterated expectations we have

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] = \mathbf{E}[\mathbf{E}[Y|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbf{E}[Y|\mathcal{F}_n] = X_n.$$

Note, that a martingale is called *closable* or *closed in L^p*, $p \geq 1$, if for $Y \in L^p$, $X_n = \mathbf{E}[Y|\mathcal{F}_n]$ holds $\forall\ n \in \mathbb{N}$.

Example 2.9. Consider the probability space $([0,1), \mathcal{B}_{[0,1)}, \lambda)$, with λ being the Lebesgue measure on the Borel-σ-Algebra of the half-open unit interval. We define the finite σ-Algebra generated by all dyadic intervals of $[0,1)$ of length 2^{-j}, $j \in \mathbb{N}$, by

$$\mathcal{A}_j := \sigma([0,2^{-j}], \ldots, [k2^{-j}, (k+1)2^{-j}), \ldots, [(2^{-j}-1)2^{-j}, 1)]),\ k = 0, 1, \ldots, 2^j-1.$$

Obviously, $(\mathcal{A}_j)_{j\in\mathbb{N}} \subset \mathcal{B}_{[0,1)}$ is a filtration. Then $(X_j)_{j\in\mathbb{N}} := 2^j \mathbf{1}_{[0,2^{-j})}$, is a martingale. Since the sets $[k2^{-j}, (k+1)2^{-j})$ are a disjoint partition of $[0,1)$, every $A \in \mathcal{A}_j$ consists of a finite disjoint union of such sets. If $[0,2^{-j}) \subset A$, we have

$$\begin{aligned}\int_A X_{j+1}\ d\lambda &= \int_\Omega 2^{j+1} \mathbf{1}_{A\cap[0,2^{-(j+1)})}\ d\lambda = 2^{j+1}2^{-(j+1)} = 2^j 2^{-j} \\ &= \int_\Omega 2^j \mathbf{1}_{A\cap[0,2^{-j})}\ d\lambda = \int_A X_j\ d\lambda\end{aligned}$$

and, if $[0.2^{-j}) \not\subset A$,

$$\int_A X_{j+1}\ d\lambda = \int_A 2^{j+1}\mathbf{1}_{[0,2^{-(j+1)})}\ d\lambda = 0 = \int_A 2^j \mathbf{1}_{[0,2^{-j})}\ d\lambda = \int_A X_j\ d\lambda.$$

Example 2.10. Suppose that the stochastic process $(Z_n)_{n\in\mathbb{N}}$ represents the size of a population at time n and let $(p_j)_{j\in\mathbb{N}}$ be a probability distribution, also called reproduction distribution, that gives the chance for each entity of n-th generation having j descendants. Additionally, let each entities life duration be one time instant. If an entity is born at time n, it dies at time $n+1$ and its descendants begin their life at the time $n+1$. The stochastic process $(Z_n)_{n\in\mathbb{N}}$ together with its reproduction distribution $(p_j)_{j\in\mathbb{N}}$ forms a Markov-chain with transition matrix $\mathbf{T} = (p_{ij})_{i,j\in\mathbb{N}}$ and transition probabilites

$$p_{ij} = \mathbf{P}[Z_{n+1} = j | Z_n = i] = p_j^{*(i)}.$$

Such a process is called *Galton-Watson-process*. Let $(X_{nk})_{n,k\geq 1}$, be an independent array of identically distributed random variables standing for the progeny of the k-th entity of n-th generation. Link each X_{nk} with reproduction distribution $(p_j)_{j\in\mathbb{N}}$ and set

$$Z_n = \sum_{k=1}^{Z_{n-1}} X_{nk}. \tag{2.9}$$

Assume now that Z_0 and $(X_{nk})_{n,k\geq 1}$ along with a sequence of probability measures $\mathbf{P}_i$, $i \in \mathbb{N}$, where i denotes the entities at the beginning ($Z_0 = i$), are defined on a measurable space $(\Omega, \mathcal{F})$.

This basic setup manifested in

$$(\Omega, \mathcal{F}, (\mathbf{P}_i)_{i\in\mathbb{N}}, (X_{nk})_{n,k\geq 1}, (Z_n)_{n\in\mathbb{N}})$$

is called the standard Galton-Watson-process modell. For more details see ALSMEYER [1].
Set $\mathbf{E}[X_{nk}] = \mu$, $0 < \mu < \infty$, $\mathcal{F}_0 := \sigma(Z_0)$ and
$\mathcal{F}_n := \sigma(Z_0, X_{jk}, 1 \leq j \leq n, k \geq 1)$. $\mathcal{F}_n$ contains the information of every possible sequence of reproduction of each entity until time n. A commonly investigated stochastic process is Z_n weighted by the reciprocal of its expectation[3], namely

$$W_n = \frac{Z_n}{\mu^n}.$$

We will show that W is a martingale. $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is obviously a filtration and Z_n, defined as is (2.9), is $\mathcal{F}_n$-measurable, which leads to $\mathcal{F}_n$-measurability of the W_n. The expected value of W being 1, $\forall\ n \in \mathbb{N}$, the only thing left to show is the martingale property (2.1): At first notice that the entities $X_{n+1,k}$ of the $(n+1)$-th generation are independent of $\mathcal{F}_n$. Then

$$\begin{aligned}
\mathbf{E}[W_{n+1}|\mathcal{F}_n] &= \frac{1}{\mu^{n+1}}\mathbf{E}[Z_{n+1}|\mathcal{F}_n] \\
&= \frac{1}{\mu^{n+1}}\mathbf{E}[\sum_{k=1}^{Z_n} X_{n+1,k}|\mathcal{F}_n] \\
&= \frac{1}{\mu^{n+1}}\sum_{k=1}^{Z_n}\mathbf{E}[X_{n+1,k}] \\
&= \frac{1}{\mu^{n+1}}\sum_{k=1}^{Z_n}\mu = \frac{Z_n\mu}{\mu^{n+1}} = \frac{Z_n}{\mu^n} = W_n
\end{aligned}$$

and W is martingale.

Example 2.11. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let ν be a finite measure on $\mathcal{F}$ and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a filtration. Suppose that ν is absolutely continuous with respect to $\mathbf{P}$ [A.7] when they are both restricted on $\mathcal{F}_n$. The Radon-Nikodym theorem [A.8] yields that there exists a derivative X_n of ν with respect to $\mathbf{P}$ when both are restricted to $\mathcal{F}_n$. X_n is a function measurable $\mathcal{F}_n$ and integrable with respect to $\mathbf{P}$, and it satisfies

$$\int_A X_n\ d\mathbf{P} = \nu(A),\ \ A \in \mathcal{F}_n.$$

If $A \in \mathcal{F}_n$, then $A \in \mathcal{F}_{n+1}$ as well, so that $\int_A X_n\ d\mathbf{P} = \nu(A) = \int_A X_{n+1}\ d\mathbf{P}$, which gives (2.1) and thus X is a martingale.

[3]To see that $\mathbf{E}[Z_n] = \mu^n$ follows by properties of the generating function of p_j, consult [1, p. 6]

2.2 Transformations

Definition 2.12. A process X is said to be predictable if X_0 is $\mathcal{F}_0$-measurable and for every $t > 0$ each X_t is $\mathcal{F}_{t-1}$-measurable.

This dry definition will become rather juicy when we connect it with gambling. Consider X as the fortune of a gambler with X_n being $\mathcal{F}_n = \sigma(\mathcal{F}_k, k \leq n)$-measurable for every n. That means that the gamblers fortune at point n depends only on the n-plays being played until then. Suppose now that our gambler is cautious and just bets unit stakes at each play. The gain between each gamble would be $\Delta_n := X_n - X_{n-1}$. Now a second gambler, who is known as a specialist in choosing the right stakes, adivces the first gambler how to modify his bets in order to increase his fortune. Say, the random variable W_n is the wager the first gambler bets at the n-th round. The variable W_n depends only on the first $(n-1)$-plays and is thus predictable. It would be unlogical, for us, to bet on the n-th game *and* to be in posession of its outcome. After the consulting of the second gambler, the first one wins $W_n\Delta_n$ per game. The total return for the first gambler after n games is

$$Z_n = W_0\Delta_0 + W_1\Delta_1 + \ldots + W_n\Delta_n.$$

An elementary example to illustrate this formula would be a coin tossing game. Just associate with the Δ_n the usual Bernoulli trials.

Definition 2.13. Let V and X be two processes, the first predictable and the second one adapted. Then the process

$$Z_n = V_0X_0 + V_1(X_1 - X_0) + \ldots + V_n(X_n - X_{n-1}).$$

is called the transform of X by V, and denoted by $Z = V \cdot X$

If T is a stopping time, X an adapted process, the stopped process X_T is the transformation of X by $V_n = \mathbf{1}_{\{n \leq T\}}$.

2.3 The Doob Decomposition

Suppose that X is an adapted process with $\mathbf{E}[|X_n|] < \infty$, $\forall\ n \in \mathbb{N}$. Define the random variables Y_n, A_n, recursively, by induction, in the following manner:

$$\begin{array}{ll} Y_0 = X_0 & A_0 = 0 \\ Y_1 = Y_0 + (X_1 - \mathbf{E}[X_1|\mathcal{F}_0]) & A_1 = X_0 - \mathbf{E}[X_1|\mathcal{F}_0] \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ Y_n = Y_{n-1} + (X_n - \mathbf{E}[X_n|\mathcal{F}_{n-1}]) & A_n = A_{n-1} + (X_{n-1} - \mathbf{E}[X_n|\mathcal{F}_{n-1}]) \end{array} \tag{2.10}$$

We summarize our first observations in the next Lemma.

Lemma 2.14. *Let X be an adapted, integrable process and Y, A two processes defined as in (2.10).*
(a) $X = Y$ - A.
(b) Y is a martingale.
(c) A is a predictable process.

Proof. (a):
This will be proven by induction:

Base case ($n = 0$): $X_0 = X_0 - 0$.
Induction hypothesis: $X_n = Y_n - A_n$, holds for some $n \in \mathbb{N}$.
Induction step:

$$\begin{aligned} Y_{n+1} - A_{n+1} &= Y_n + X_{n+1} - \mathbf{E}[X_{n+1}|\mathcal{F}_n] - A_n - X_n + \mathbf{E}[X_{n+1}|\mathcal{F}_n] \\ &= Y_n - A_n + X_{n+1} - X_n \\ &= X_n + X_{n+1} - X_n = X_{n+1}. \end{aligned}$$

(b):
Adaptation and integrability follow immediately. We add that

$$\mathbf{E}[Y_{n+1}|\mathcal{F}_n] = \mathbf{E}[Y_n|\mathcal{F}_n] + \mathbf{E}[X_{n+1}|\mathcal{F}_n] - \mathbf{E}[\mathbf{E}[X_{n+1}|\mathcal{F}_n]|\mathcal{F}_n] = Y_n.$$

(c):
We use the induction principle again:

Base case ($n = 0$): $A_0 = 0$ is trivially $\mathcal{F}_0$-measurable.
Induction hypothesis: The A_n are $\mathcal{F}_{n-1}$-measurable.
Induction step:

$$A_{n+1} = \underbrace{A_n}_{\mathcal{F}_{n-1}-\text{measurable}} + \underbrace{(X_n - \mathbf{E}[X_{n+1}|\mathcal{F}_n])}_{\mathcal{F}_n\text{-measurable}} \Rightarrow A_{n+1} \text{ is } \mathcal{F}_n - \text{measurable.}$$

And the lemma is proven.

□

Inspired by this result we can state a stronger one, namely, the *Doob decomposition.*

Theorem 2.15 (DOOB DECOMPOSITION)**.** *Let X be an adapted, integrable process. Then X can be decomposed as*

$$X = Y - A$$

Where Y is a martingale, A a predictable process such that $A_0 = 0$. This decomposition is unique. It is called the Doob decomposition of X.

Proof. Setting Y and A as in (2.10) only uniqueness is left to prove. Assume there exists a second decomposition so that $X = Y - A = Y' - A'$. According to the definition $M = Y - Y' = A - A'$ is a predictable martingale with $M_0 = 0$. The martingale property means $\mathbf{E}[M_{n+1}|\mathcal{F}_n] = M_n$, the predictability that $\mathbf{E}[M_{n+1}|\mathcal{F}_n] = M_{n+1}$, hence $M_{n+1} = M_n = \ldots = M_0 = 0$ almost surely, which implies $Y = Y'$ and $A = A'$.

□

Remark 2.16. If we replace X in the Doob decomposition with a supermartingale, $\mathbf{E}[X_{n-1} - X_n|\mathcal{F}_n] > 0$ holds, which means that a positive quantity is being added constantely to the A_n of (2.10). We consequently follow that $A_n \leq A_{n+1}$ and a process with this property is called an *increasing process.*

Remark 2.17. The Doob decomposition gives uniqueness even if the predictable process A is non-increasing. To assert uniqueness without predictability would be wrong as predictability was needed to prove the uniqueness of the Doob decomposition.

Actually, the Doob decomposition could have been proven in a later segment of this thesis. Even though the placement seems to interrupt the thought process of the reader, it should be assured that the Doob decomposition will be needed in later proofs and, because of its premise, doesn't require deep martingale theory. This is the main reason why it was put right after the martingale basics.

2.4 Doob's Optional Sampling

A martingale X can cover a wide spectrum of information. An observant shall be interested when a particular event occurs for the first time. We learned to model this with a stopping time T, the desired event being X_T. Another observant might focus on a whole family of events $X_{(T_i)_{i \in \mathcal{I}}}$. The question is, if the martingale (*resp.* supermartingale, submartingale) property (2.1) still holds if we insert two different stopping times, say, $T_i \leq T_j$. For our purposes we will restrict ourselves to bounded stopping times.

Theorem 2.18. *Let X be a martingale (resp. supermartingale, submartingale) and V a non-negative process which is predictable. If the random variables $(V \cdot X)_n$ are integrable, the transformed process $V \cdot X$ is a martingale (resp. supermartingale, submartingale).*

Proof. Examine that

$$\begin{aligned}\mathbf{E}[(V \cdot X)_{n+1} - (V \cdot X)_n | \mathcal{F}_n] &= \mathbf{E}[V_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] \\ &= V_{n+1}\mathbf{E}[X_{n+1} - X_n | \mathcal{F}_n],\end{aligned}$$

since V is predictable. If X is a martingale the second factor in the last member is 0 and $V \cdot X$ is a martingale. If X is a supermartingale the last member is ≤ 0, since $V \geq 0$. Thus $V \cdot X$ is a supermartingale.

□

Remark 2.19. Reconsider the predictable process W, which represented the gamblers wager. Such a sequence is also called a *betting system*. Instantly, W satisfies the conditions of **T2.18** and thus gives it a spicy interpretation: If X is a fair game then an arbitrary change of the stakes at each time n does not change its fairness. A game is called subfair if X is a supermatingale and superfair if X is a submartingale. Equivalently, a subfair game remains a subfair and a superfair game remains a superfair game. The stake advisor from p. 20 obviously becomes a fraud.

Lemma 2.20. *Let X be a martingale (resp. supermartingale, submartingale), and T be a bounded stopping time. Then the stopped process X^T is a martingale (resp. supermartingale, submartingale).*

Proof. In the preceding theorem take $V_n = \mathbf{1}_{\{n \leq T\}}$, so that $X_n^T = (V \cdot X)_n$. Predictability of $\mathbf{1}_{\{n \leq T\}}$ follows by

$$[T \geq n]^c = [T < n] = [T \leq n-1] \in \mathcal{F}_{n-1}.$$

Note for integrability that: $\left|X_n^T\right| \leq |X_0| + \ldots + |X_n|$.

□

As mentioned before, this version of the theorem refers to bounded stopping times only.

The fact, that we can model a stopped process X^T as a transformation helps us to prove Doob's optional sample theorem in an elegant way. We also include the standard proof of the optional sampling theorem as alternate proof in order to see the advantage of shortness the proof using transformations has.

Theorem 2.21 (Optional Sampling)**.** *Let X be a supermartingale (resp. martingale), and let S and T be two bounded stopping times such that $S \leq T$. Then X_S and X_T are integrable and we have*

$$\mathbf{E}[X_T | \mathcal{F}_S] \leq X_S \quad a.s. \tag{2.11}$$

(= in the martingale case).

Proof - 1st version. The martingale case follows from the supermartingale case applied to X and $-X$. Set $V_n = \mathbf{1}_{\{S<n\leq T\}}$, a positive and predictable process, and $Y = V \cdot X$. According to **T2.18** Y is a supermartingale. Suppose now that $S \leq T \leq k$, $k \in \mathbb{N}$, we then have

$$\begin{aligned} Y_k &= (\mathbf{1}_{\{S<T\}} \cdot X)_k \\ &= \mathbf{1}_{\{S<0\leq T\}} X_0 + \mathbf{1}_{\{S<1\leq T\}}(X_1 - X_0) + \ldots + \mathbf{1}_{\{S<k\leq T\}}(X_k - X_{k-1}) \\ &= (X_{S+1} - X_S) + \ldots + (X_T - X_{T-1}) = X_T - X_S. \end{aligned}$$

Applying the expected value on both sides and using the supermartingale property of Y gives $\mathbf{E}[Y_k] = \mathbf{E}[X_T - X_S] \leq 0$. Denote by $S_A(\omega) = S(\omega)$, $\omega \in A$, and $S_A(\omega) = +\infty$, $\omega \in A^c$. Let $A \in \mathcal{F}_S$, and apply the preceding reasoning to $S' = S_A \wedge k$, $T' = T_A \wedge k$. The desired inequality $\int_A (X_T - X_S)\, d\mathbf{P} \leq 0$ results.

□

Proof - 2nd version. Step 1: Assume $0 \leq T - S \leq 1$.
In this case

$$\{S < T\} \cap \{S = j\} = \{T > j\} \cap \{S = j\} = \{T \leq j\}^c \cap \{S = j\} \in \mathcal{F}_j, \; j \in \mathbb{N}.$$

We see, because of (2.1) for supermartingales and $T - S \leq 1$ that

$$\begin{aligned} \int_\Omega X_S \, d\mathbf{P} &= \int_{\{T=S\}} X_S \, d\mathbf{P} + \sum_{j=1}^{N-1} \int_{\{S<T\}\cap\{S=j\}} X_j \, d\mathbf{P} \\ &\geq \int_{\{T=S\}} X_T \, d\mathbf{P} + \sum_{j=1}^{N-1} \int_{\{S<T\}\cap\{S=j\}} X_{j+1} \, d\mathbf{P} \\ &= \int_{\{T=S\}} X_T \, d\mathbf{P} + \int_{\{S<T\}} X_T \, d\mathbf{P} \\ &= \int_\Omega X_T \, d\mathbf{P} \end{aligned}$$

Step 2: If $S \leq T \leq N$ we introduce (at most N) random variables $\rho_j := T \wedge (S + j)$, $j = 0, ..., k \leq N$. By **L1.23** these are stopping times. For some $k \leq N$ we get $S = \rho_0 \leq ... \leq \rho_k = T$, while $\rho_{l+1} - \rho_l \leq 1$, $l = 0, ..., N-1$. Repeating step 1 from above k times yields

$$\int_\Omega X_S \, d\mathbf{P} = \int_\Omega X_{\rho_0} \, d\mathbf{P} \geq ... \geq \int_\Omega X_{\rho_k} \, d\mathbf{P} = \int_\Omega X_T \, d\mathbf{P}.$$

Note that, due to **D1.20** and **L1.25**, for any $A \in \mathcal{F}_S$ the function $\rho := \rho_A := S\mathbf{1}_A + T\mathbf{1}_{A^c}$ is a bounded stopping time:

$$\{\rho \leq j\} = (A \cap \{S \leq j\}) \cup (A^c \cap \{T \leq j\}) \in \mathcal{F}_j.$$

Since $\rho \leq T$,

$$\int_\Omega X_S \mathbf{1}_A + X_T \mathbf{1}_{A^c} \; d\mathbf{P} = \int_\Omega X_\rho \; d\mathbf{P} \geq \int_\Omega X_T \; d\mathbf{P}$$

yields $\int_A X_S \; d\mathbf{P} \geq \int_A X_T \; d\mathbf{P}$, $\forall \; A \in \mathcal{F}_S$.

□

Example 2.22. If the process M is a Martingale and T a stopping time such that M_T is integrable, it is in general not true that $\mathbf{E}[M_T] = \mathbf{E}[M_0]$.

Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. Bernoulli trials, i.e.,
$\mathbf{P}[\xi_0 = -1] = \mathbf{P}[\xi_0 = +1] = \frac{1}{2}$. Let $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_n)$ and
$M_n = \xi_0 + \ldots + \xi_n$. Then $(M_n)_{n\in\mathbb{N}}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.
Define

$$T(\omega) := \min \{ n \in \mathbb{N} : M_n(\omega) = 1 \} .$$

This is a stopping time and one can show that $T < \infty$ almost surely.
Hence, $M_T = 1$ almost surely and M_T is integrable. But
$\mathbf{E}[M_T] = 1 \neq 0 = \mathbf{E}[M_0]$.

The example shows that plain integrability of M_T doesn't imply $\mathbf{E}[M_T|\mathcal{F}_0] = M_0$. Indeed, this would imply $\mathbf{E}[M_T] = \mathbf{E}[\mathbf{E}[M_T|\mathcal{F}_0]] = \mathbf{E}[M_0]$ and we just saw that this is false in general. Truly, under the preassumptions of the optional sampling theorem **T2.21** the boundedness property of the stopping times can not be omitted.

2.5 The Maximal Inequality

Definition 2.23. Let X be any process. Then we set
$\|X\|_p := \sup_{n\in\mathbb{N}} \|X_n\|_p$, $1 \leq p \leq \infty$. A process X such that $\|X\|_p < \infty$ is said to be bounded in $\mathcal{L}^p$.

This definition allows us to use a shorthanded notation for bounded processes.
Through the last decades many textbooks concerned with martingale theory include the maximal inequality. Basically, it is convinient to state the inequality for submartingales, because the inequality chain used in its proof becomes very clear. We will swim against the main stream, by swaping the maximum operator with the supremum one, in order to get a hold of the whole process, and prove the maximal inequality for supermartingales and martingales. This exchange makes it somehow funny for us to still call it maximal inequality. The validity of this inequality remains unbroken for submartingales.

Theorem 2.24 (Maximal Inequality). *Let X be a supermartingale, c a positive constant. Then*

$$\mathbf{P}[\sup_{n\in\mathbb{N}} |X_n| \geq c] \leq A\frac{\|X\|_1}{c}, \text{ with} \tag{2.12}$$

(1) $A = 1$ if $X \leq 0$, $X \geq 0$ or X is a martingale.
(2) $A = 3$ if X is a supermartingale.

Proof. For arbitrary $k \in \mathbb{N}$ set

$$T(\omega) = \inf\{n : n \leq k,\ X_n(\omega) \geq c\}.$$

If no such n exists, we set $T(\omega) = k$. We see that T is a bounded stopping time and apply the optional sampling theorem on X to get $\mathbf{E}[X_0] \geq \mathbf{E}[X_T]$. If $X_n \geq c$ for some $n \leq k$, we have $X_T \geq c$. Otherwise $X_T = X_k$. Accordingly, by splitting $\mathbf{E}[X_T]$ in two integrals, we estimate

$$\mathbf{E}[X_0] \geq \mathbf{E}[X_T] \geq \int_{\left\{\sup_{n\leq k} X_n \geq c\right\}} c\ d\mathbf{P} + \int_{\left\{\sup_{n\leq k} X_n < c\right\}} X_k\ d\mathbf{P},$$

which leads to

$$\mathbf{E}[X_0] + \int_{\left\{\sup_{n\leq k} X_n < c\right\}} (-X_k)\ d\mathbf{P} \geq c\mathbf{P}[\sup_{n\leq k} X_n \geq c]. \tag{2.13}$$

Due to

$$\int_{\left\{\sup_{n\leq k} X_n < c\right\}} (-X_k)\ d\mathbf{P} \leq \mathbf{E}[X_k^-] \leq \sup_{k\in\mathbb{N}} \mathbf{E}[X_k^-],$$

and

$$\begin{aligned} c\mathbf{P}[\sup_{n\leq k} X_n \geq c] &\leq c\mathbf{P}[\sup_{n\in\mathbb{N}} X_n \geq c] \\ &\leq \int_{\left\{\sup_{n\in\mathbb{N}} X_n \geq c\right\}} \sup_{n\in\mathbb{N}} X_n\ d\mathbf{P} \leq \int_{\left\{\sup_{n\in\mathbb{N}} X_n \geq c\right\}} X_0\ d\mathbf{P}, \end{aligned}$$

we derive (2.14)

$$c\mathbf{P}[\sup_{n\in\mathbb{N}} X_n \geq c] \leq \mathbf{E}[X_0] + \sup_{k\in\mathbb{N}} \mathbf{E}[X_k^-] \leq 2\|X\|_1. \tag{2.14}$$

Note that we have proven the theorem with $A = 1$ in case of $X \geq 0$, since this implies disappearence of the member $\sup_{k\in\mathbb{N}} \mathbf{E}[X_k^-]$ in (2.14). To deal with negative values we modify the above stopping time into

$$T(\omega) = \inf\{n : n \leq k,\ X_n(\omega) \leq -c\}.$$

Again, if no such n exists, $T(\omega) = k$. We apply the optional sampling theorem again and get $\mathbf{E}[X_T] \geq \mathbf{E}[X_k]$, which implies, as above

$$\mathbf{E}[X_k] \leq \mathbf{E}[X_T] \leq -c\mathbf{P}[\inf_{n \leq k} X_n \leq -c] + \int_{\{\inf_{n \leq k} X_n > -c\}} X_T \; d\mathbf{P}.$$

Multiplying the inequality with (-1) and substracting
$\int_{\{\inf_{n \leq k} X_n > -c\}} (-X_T) \; d\mathbf{P}$ gives

$$\begin{aligned} c\mathbf{P}[\inf_{n \leq k} X_n \leq -c] &\leq \int_{\{\inf_{n \leq k} X_n \leq -c\}} (-X_T) \; d\mathbf{P} \\ &\leq \int_{\{\inf_{n \leq k} X_n \leq -c\}} (-X_k) \; d\mathbf{P} \leq \mathbf{E}[X_k^-] \leq \|X\|_1. \end{aligned}$$

With the help of inequality

$$c\mathbf{P}[\inf_{n \leq k} X_n \leq -c] \leq c\mathbf{P}[\inf_{n \in \mathbb{N}} X_n \leq -c] \leq \int_{\{\inf_{n \in \mathbb{N}} X_n \leq -c\}} (-X_k) \; d\mathbf{P}$$

we deduce the case for $n \in \mathbb{N}$

$$c\mathbf{P}[\inf_{n \in \mathbb{N}} X_n \leq -c] \leq \int_{\{\inf_{n \in \mathbb{N}} X_n \leq -c\}} (-X_k) \; d\mathbf{P} \leq \sup_{k \in \mathbb{N}} \mathbf{E}[X_k^-] \leq \|X\|_1. \tag{2.15}$$

We get (2.12), with $A = 1$, for the case $X \leq 0$, and also for the martingale case, since $-|X|$ is a negative supermartingale. Adding (2.14) and (2.15) we find in the general case

$$c\mathbf{P}[\sup_{n \in \mathbb{N}} |X_n| \geq c] \leq \mathbf{E}[X_0] + 2 \sup_{k \in \mathbb{N}} \mathbf{E}[X_k^-] \leq 3\|X\|_1. \tag{2.16}$$

□

Note that the proof of the maximal inequality holds the two sharper inequalites (2.14) and (2.15), which are specially interesting and will be used repeatedly. In some particular applications, these inequalities are needed in their conditional form. We derive them by claiming that if X is a supermartingale and $\mathbf{P}^A$ is the conditional probability

$$\mathbf{P}^A[B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[A]}, \quad (A \in \mathcal{F}_0, \; B \in \mathcal{F}, \; \mathbf{P}[A] > 0)$$

on Ω then X is still a supermartingale for $\mathbf{P}^A$. Now, set $X^\star := \sup_{n \in \mathbb{N}} X_n$ and apply (2.14) to $\mathbf{P}^A$ instead of $\mathbf{P}$.
It appears that

$$\begin{aligned} c\mathbf{P}^A[X^\star \geq c] = c\frac{\mathbf{P}[[X^\star \geq c] \cap \; A]}{\mathbf{P}[A]} &\leq \int_\Omega X_0 \; d\mathbf{P}^A + \sup_{k \in \mathbb{N}} \int_\Omega X_k^- \; d\mathbf{P}^A \\ &= \frac{1}{\mathbf{P}[A]} (\int_A X_0 \; d\mathbf{P} + \sup_{k \in \mathbb{N}} \int_A X_k^- \; d\mathbf{P}). \end{aligned}$$

We apply the theorem of conditional expectation [A.9]: Since $\mathbf{1}_{\{X^\star \geq c\}} \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{F}_0$ is a subsigma Algebra of $\mathcal{F}$, there exists a $\mathcal{F}_0$-measurable function $\mathbf{E}[\mathbf{1}_{\{X^\star \geq c\}}|\mathcal{F}_0] = \mathbf{P}[X^\star \geq c|\mathcal{F}_0]$ with

$$\int_A \mathbf{1}_{\{X^\star \geq c\}}\; d\mathbf{P} = \int_A \mathbf{P}[X^\star \geq c|\mathcal{F}_0]\; d\mathbf{P},\; \forall\; A \in \mathcal{F}_0.$$

Using $\int_A \mathbf{1}_{\{X^\star \geq c\}}\; d\mathbf{P} = \mathbf{P}[[X^\star \geq c] \cap\; A],\; \forall\; A \in \mathcal{F}_0$, we estimate

$$\int_A c\mathbf{P}[X^\star \geq c|\mathcal{F}_0]\; d\mathbf{P} \leq \int_A X_0\; d\mathbf{P} + \sup_{k\in\mathbb{N}} \int_A X_k^-\; d\mathbf{P},\; \forall\; A \in \mathcal{F}_0.$$

Now X_0 and X_k^- lie both in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and we can apply the theorem of conditional expectation again, getting

$$\int_A c\mathbf{P}[X^\star \geq c|\mathcal{F}_0]\; d\mathbf{P} \leq \int_A \mathbf{E}[X_0|\mathcal{F}_0]\; d\mathbf{P} + \sup_{k\in\mathbb{N}} \int_A \mathbf{E}[X_k^-|\mathcal{F}_0]\; d\mathbf{P},\; \forall\; A \in \mathcal{F}_0. \tag{2.17}$$

Finally, we apply [A.2] to the left side and to the first member on the right side of (2.17) and simplify to

$$c\mathbf{P}[X^\star \geq c|\mathcal{F}_0] \leq X_0 + \sup_{k\in\mathbb{N}} \mathbf{E}[X_k^-|\mathcal{F}_0].$$

2.6 The $\mathcal{L}^p$ inequality

We already discovered in **R2.2** that a martingale X lies in $\mathcal{L}^1$. If additionally $\mathbf{E}[|X|^p] < \infty$, $1 \leq p \leq \infty$, holds, X is an element of $\mathcal{L}^p$. The theorem this section is dedicated to, simplifies $\mathcal{L}^p$-convergence results of martingales. Oftenly it is referred to as *Doob's inequality.*

Theorem 2.25 (Doob Inequality)**.** *Let* $\frac{1}{p} + \frac{1}{q} = 1$, $1 < p, q < \infty$, *and let* X *be a non-negative submartingale such that*

$$\sup_{n\in\mathbb{N}} \mathbf{E}[(X_n)^p] < \infty. \tag{2.18}$$

The random variable $\sup_{n\in\mathbb{N}} X_n$ *then belongs to* $\mathcal{L}^p$, *and we have*

$$\|\sup_{n\in\mathbb{N}} X_n\|_p \leq q \sup_{n\in\mathbb{N}} \|X_n\|_p. \tag{2.19}$$

Proof. Due to **T1.11** of La Valleé Poussin the random variables (X_n) are uniformly integrable for every $n \in \mathbb{N}$, so that $\lim_{n\to\infty} X_n = X_\infty < \infty$ almost surely.
Let $F : [0, +\infty) \to \mathbb{R}^+$ be an increasing, convex and continuous function, with $F(0) = 0$. Using Jensen's inequality, the continuity of F and Fatou's lemma [A.4] we establish

$$\begin{aligned} F(\mathbf{E}[\liminf_{n\to\infty} X_n]) &\leq \mathbf{E}[F(\liminf_{n\to\infty} X_n)] \\ &= \mathbf{E}[\liminf_{n\to\infty} F(X_n)] \leq \liminf_{n\to\infty} \mathbf{E}[F(X_n)]. \end{aligned}$$

Putting $F(y) = y^p$ yields $\mathbf{E}[(X_\infty)^p] \leq \liminf_{n\to\infty} \mathbf{E}[(X_n)^p] < \infty$ and $X_\infty \in \mathcal{L}^p$.
If we modify the proof of **T2.24** for non-negative submartingales, it holds for (2.14) that

$$\mathbf{P}[\sup_{n\in\mathbb{N}} X_n \geq c] \leq \frac{1}{c} \int_{\{\sup_{n\in\mathbb{N}} X_n \geq c\}} X_k \; d\mathbf{P} \tag{2.20}$$

Set $Y = \sup_{n\in\mathbb{N}} X_n$ and $L(y) = \mathbf{P}[Y \geq y]$. Using a change of variable, we get

$$\mathbf{E}[F(Y)] = \int_\Omega F(Y) \; d\mathbf{P}(\omega) = \int_0^\infty F(y) \; d\mathbf{P}[Y < y] = -\int_0^\infty F(y) \; dL(y).$$

Now we partially integrate the last member, use (2.20) and Fubini's theorem [A.6], so that

$$\begin{aligned}
\mathbf{E}[F(Y)] &= -\int_0^\infty F(y) \; dL(y) = -\lim_{x\to\infty} F(y)L(y)|_0^x + \int_0^\infty L(y) dF(y) \\
&\leq \int_0^\infty L(y) \; dF(y) \leq \int_0^\infty (\frac{1}{y} \int_{\{Y\geq y\}} X_\infty \; d\mathbf{P}) \; dF(y) \\
&= \int_\Omega X_\infty (\int_0^Y \frac{1}{y} \; dF(y)) \; d\mathbf{P} = \mathbf{E}[X_\infty \int_0^Y \frac{1}{y} \; dF(y)]. \qquad (2.20)
\end{aligned}$$

Let us now take $F(y) = y^p$. We get due to Hölder's inequality [A.11]

$$\mathbf{E}[Y^p] \leq \frac{p}{p-1} \mathbf{E}[X_\infty Y^{p-1}] \leq \frac{p}{p-1} \|X_\infty\|_p \|Y^{p-1}\|_q. \tag{2.21}$$

Since $\|Y^{p-1}\|_q = (\mathbf{E}[Y^{\frac{(p-1)p}{p-1}}])^{\frac{1}{q}} = (\mathbf{E}[Y^p])^{\frac{1}{q}} = \|Y\|_p^{\frac{p}{q}}$ and $\frac{p}{p-1} = q$, dividing (2.21) by $\|Y\|_p^{\frac{p}{q}}$, gives

$$\| \sup_{n\in\mathbb{N}} X_n \|_p \leq q \|X_\infty\|_p,$$

and $\sup_{n\in\mathbb{N}} X_n \in \mathcal{L}^p$. From $|X_n - X_\infty| \leq 2 \sup_{n\in\mathbb{N}} X_n$, we get
$\|X_n - X_\infty\|_p \leq 2 \| \sup_{n\in\mathbb{N}} X_n\|_p < \infty$.
By the dominated convergence theorem $\|X_n - X_\infty\|_p \to 0$.
Finally, $\|X_\infty\|_p \leq \|X_\infty - X_n\|_p + \sup_{n\in\mathbb{N}} \|X_n\|_p$ yields
$\|X_\infty\|_p = \sup_{n\in\mathbb{N}} \|X_n\|_p$ and (2.19) is verified.

□

2.6.1 The weak-L^p space

This section provides a different method of deducing a sharper Doob's inequality by stating it as a special case of the famous interpolation theorem of Marcinkiewicz. In order to make it sharper we have to be more generous towards the preassumptions.

A random variable f is in the space weak-$L^p(\Omega, \mathcal{F}, \mathbf{P})$, $1 \leq p < \infty$, denoted by $L^p_w(\Omega, \mathcal{F}, \mathbf{P}) = L^p_w$, if there is a positive constant A such that

$$\mathbf{P}[|f| > t] \leq \frac{A^p}{t^p}, \ \forall\ t > 0. \tag{2.22}$$

The best constant A for this inequality is the L^p_w-norm of f, set as

$$\|f\|_{p,w} = \inf\Big\{A|\ \mathbf{P}[|f| > t] \leq \frac{A^p}{t^p},\ \forall\ t > 0\Big\} = \sup_{t>0} \Big\{t(\mathbf{P}[|f| > t])^{1/p}\Big\}.$$

The space L^∞_w is defined as the collection of random variables for which (2.22) holds for some constant A, for all $p \geq 1$, and for all $t > 0$.
Let us now assert some basic properties of L^p_w and its interaction with L^p.

Proposition 2.26. *(1) L^p_w is a linear space.*
(2) For $1 \leq p < \infty$, the space L^p is a subset of L^p_w, i.e. $L^p \subseteq L^p_w$.
(3) For $p = \infty$, the space L^∞ coincides with L^∞_w, i.e. $L^\infty = L^\infty_w$.

Proof. (1):
Let $f, g \in L^p_w$ and choose arbitrary non-negative constants α, β. Then for all $t > 0$,

$$[|\alpha f + \beta g| > t] \subset \Big[|f| > \frac{t}{2\,|\alpha|}\Big] \cup \Big[|g| > \frac{t}{2\,|\beta|}\Big],$$

and thus L^p_w is a linear space.
(2):
If $f \in L^p$, then for all $t > 0$,

$$\mathbf{P}[|f| > t] \leq \frac{\mathbf{E}[|f|^p]}{t^p} = \frac{\|f\|^p_p}{t^p},$$

by Markov's inequality ([12], p.110). Therefore $f \in L^p_w$ and $\|f\|_p = \|f\|_{p,w}$.
(3):
If $f \in L^\infty_w$ and $t > A$, then $\mathbf{P}[|f| > t] = 0$ and $f \in L^\infty$ follows. On the other hand, if $f \in L^\infty$ then (2.22) holds with $A = \|f\|_\infty$ and $f \in L^\infty_w$. Thus $L^\infty_w = L^\infty$ and $\|\cdot\|_{p,w} = \|\cdot\|_p$.

□

Remark 2.27. Consider $L^1\big((0,1), \mathcal{B}_{(0,1)}, \lambda\big)$. An example for a function that lies in L^1_w, but not in L^1 would be $f(x) := \frac{1}{x}$, $x \in (0, 1)$.

Indeed, check that

$$\int_{(0,1)} f \, d\lambda = +\infty, \text{ but } \|f\|_{p,w} = \sup_{t>0}\{t\mathbf{P}[|f| > t]\} = 1.$$

Examples for $p > 1$ can be constructed in the same manner.

However, the term *norm* for $\| \cdot \|_{p,w}$ does not seem to be accurate as the following example for $p = 1$ shows.

Example 2.28. Consider $L^1\big((0,1), \mathcal{B}_{(0,1)}, \lambda\big)$. Set $f(x) := x$ and $g(x) := 1 - x$, $x \in (0,1)$. Evaluate that

$$\|f + g\|_{p,w} = 1 \not\leq \frac{1}{2} = \|f\|_{p,w} + \|g\|_{p,w},$$

and hence, that the triangle inequality is not satisfied.

2.6.2 Quasi-linear maps

Definition 2.29. (1) A map $T : L^p \to L^q$, $1 \leq p, q \leq \infty$, is quasi-linear if there exists a positive constant C such that for all $f, g \in L^p$

$$|T(f + g)| \leq C(|T(f)| + |T(g)|), \text{ for a.e. } \omega \in \Omega.$$

(2) A quasi-linear map $T : L^p \to L^q$ is of strong type (p, q) if there exists a positive constant $M_{p,q}$ such that

$$\|T(f)\|_q \leq M_{p,q}\|f\|_p, \ \forall \ f \in L^p.$$

(3) A quasi-linear map $T : L^p \to L^q_w$ is of weak type (p, q) if there exists a positive constant $N_{p,q}$ such that

$$\|T(f)\|_{q,w} \leq N_{p,q}\|f\|_p, \ \forall \ f \in L^p.$$

When $p = q$ set $M_{p,q} = M_p$ and $N_{p,q} = N_p$.

Example 2.30. Suppose that $(f_n)_{n\in\mathbb{N}}$ is a non-negative submartingale with $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$. Set $T := \sup_{n\in\mathbb{N}}(\cdot)$. We assert that the map

$$T : L^\infty \to L^\infty \quad T : L^1 \to L^1_w$$

is quasi-linear. Additionally, we want to show that the first one is of strong type (∞, ∞) and the second one is of weak type $(1, 1)$.

Proof. Verify that for $(f_n)_{n\in\mathbb{N}} \in L^\infty \subset L^1$, exists a positive constant C such that

$$\begin{aligned}|\sup_{i,j\in\mathbb{N}}(f_i + f_j)| &\leq |\sup_{i\in\mathbb{N}}(f_i) + \sup_{j\in\mathbb{N}}(f_j)| \\ &\leq C(|\sup_{i\in\mathbb{N}}(f_i)| + |\sup_{j\in\mathbb{N}}(f_j)|), \text{ for a.e. } \omega \in \Omega.\end{aligned}$$

This shows that the above map is quasi-linear.
Now due to our assumption $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$, $T : L^\infty \to L^\infty$ is trivially of type (∞, ∞). To see that $T : L^1 \to L^1_w$ is of weak type $(1, 1)$, consult the help of the maximum inequality (2.12) for the non-negative submartingale case and notice that

$$\|\sup_{n\in\mathbb{N}}(f_n)\|_{1,w} = \sup_{t>0}\Big\{t(\mathbf{P}[\sup_{n\in\mathbb{N}}(f_n) > t])\Big\} \leq \sup_{n\in\mathbb{N}} \|f_n\|_1 = \|f\|_1$$

holds.

□

2.6.3 The Marcinkiewicz Interpolation Theorem

For our purpose, it is sufficient to formulate a certain version of the Marcinkiewicz theorem deduced from its general case.

Theorem 2.31 (Marcinkiewicz 1939)**.** *Define the map T by*

$$T : L^\infty \to L^\infty \quad T : L^1 \to L^1_w.$$

Let T satisfy the following conditions:
(1) T is quasi-linear
(2) T is of strong type (∞, ∞)
(3) T is of weak type $(1, 1)$
Then

$$\|T(f)\|_p \leq A_p\|f\|_p, \ \forall \ f \in L^1 \cap L^p$$

where A_p is a positive constant, depended on p, and $p \in (1, +\infty)$.

Proof. See [9, p. 392].

□

Taking $T = \sup_{n\in\mathbb{N}}(\cdot)$ and with the preassumptions of **E2.30** we derive that T satisfies the conditions of the stated Marcinkiewicz theorem. However, our assumption $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$ is stronger than $\sup_{n\in\mathbb{N}} \mathbf{E}[(X_n)^p] < \infty$, which was made in **T2.25** to proof Doob's inequality. Nevertheless, the inequality of the outcome

$$\|\sup_{n\in\mathbb{N}}(f_n)\|_p \leq A_p\|f_n\|_p, \ \forall \ f_n \in L^1 \cap L^p,$$

is sharper than Doob's inequality (2.19).

3 Martingale Convergence

Through the last decades, since Doobs *Stochastic Processes* [11], the statements of martingale convergence theorems have been unified to a large degree. The majority of math text books covering martingale convergence follow a certain pattern of theorems, which begins with the formulation of the upcrossing inequality. It has become the standard tool to proof the main convergence theorem of sub/supermartingales. The idea behind it is pretty simple: Suppose that $X_n \to X$ in $\mathbb{R}$ and $X \in (a,b)$, $a,b \in \mathbb{R}$. There exists an $N > 0$ such that $X_n \in (a,b)$ for $n \geq N$, meaning that only finitely many fluctuations of X_n between $(-\infty, a]$ and $[b, +\infty)$ can happen. Now assume that X_n doesn't have a limit. Then $\underline{X} = \liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n = \overline{X}$. Let $(\underline{X}, \overline{X}) \supset (a,b)$. In that case exists a subsequence (X_{n_k}) with $X_{n_k} > b$ and a subsequence (X_{m_k}) with $X_{m_k} < a$ for infinitely many k. Hence there are infinitely many fluctuations of X_n over the intervall (a,b). We conclude that convergence of the sequence X_n is equivalent of having only finitely many upcrossings from below value a to above value b, with $a < X < b$, $\forall\ a,b \in \mathbb{R}$.
In a probability context we would deal with a $\mathcal{L}^1$-bounded martingale (resp. sub/supermartingale) X and define a set $D := \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}$. Since X is $\mathcal{L}^1$-bounded the upcrossing theorem says that only finitely many upcrossings can happen which implies that $\mathbf{P}(D) = 0$ so that X converges almost surely.
The usual wide-spreaded convergence chain consists of three theorems and the upcrossing inequality:

Theorem ⓤ (Upcrossing Inequality) *Let X be a submartingale and let U be the number of upcrossing of $[a,b]$, $a,b \in \mathbb{R}$, by a sample sequence $X_1, \ldots, X_n$. Then*

$$\mathbf{E}[U] \leq \frac{\mathbf{E}[|X_n|] + |a|}{b-a}.$$

Theorem ⓝ (Convergence theorem for submartingales) *Let X be a submartingale with $\sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|] < \infty$. Then X converges almost surely to an integrable random variable Y.*

Theorem ⓞ (Convergence theorem for uniformly integrable submartingales) *Let X be a uniformly integrable submartingale. Then X converges almost surely and in L^1.*

Theorem ⓢ (Convergence theorem for closed martingales) *Let X be a martingale closed by a random variable $Y \in L^p$, $p \geq 1$, in L^p. Then X converges almost surely and in L^p.*

Due to **E2.6**, if X is a submartingale, $-X$ is a supermartingale and theorems ⓤ, ⓝ, ⓞ have a valid analogous form for supermartingales.

The order is essential. ⓤ is used to proof ⓝ, ⓝ is used and proof ⓞ and ⓞ is used to proof ⓢ.

The last passage portrayed is:

$$ⓤ \quad \to \quad ⓝ \quad \to \quad ⓞ \quad \to \quad ⓢ$$

A full arrow denotes usage of the theorem to the left of the arrow in the proof of the theorem that is pointed on.
Our main interest is to show how ⓝ can be proved without using ⓤ. For that, we will examine two different approaches: One, being proving ⓝ directly without using ⓤ and the other, by reversing the above stated theorem pattern in beginning with proving ⓢ, following ⓞ and ending with ⓝ. Because the main focus of this thesis is the alternative approach of ⓝ we will deduce it for martingales and easily use an extension to proof the same result for sub/supermartingales.

3.1 Historical facts about martingale convergence theorems

The first proof of convergence theorem ⓝ was given by Doob [10, p. 460] in 1939. Even though the upcrossing inequality ⓤ was not used, the idea of the proof comes very close to the proof with upcrossings. However, the very first proof of martingale convergence concerned closed martingales: A martingale X of the form $X_n = \mathbf{E}[Y|\mathcal{F}_n]$, $\forall\, n \in \mathbb{N}$, where $Y \in L^1$ converges almost surely and in L^1. This was shown in a slightly differnt form by Jessen [20] 1934 and Lévy [27] 1935. Jessen proved the case where $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$ with the components of process Z being mutually independent with a common distribution, each distributed uniformly in the interval $[0, 1]$. By that time, Lévy had read Jessen's theorem and proved a few months later the case with $Y = \mathbf{1}_A$, A being a point set, and the same filtration but with no distributional restrictions towards the conditioning process Z. As for the proof of this theorem given today, it is not really different from the extremely simple one that Lévy proposed to Jessen, in the case of increasing filtrations and of set indexes. A sketch of this correspondence has Bru [5, p. 3]. An execution of this sketch in a modern way proof can be found in Billingsley [2, p. 116].
It is a matter of fact that in the mid 30s probability theory was not yet established as a branch of mathematics but rather under investigation. Probabilistic terms and notation of that time differ from those in the present. Doob's definition of a martingale in his paper of 1940 [10, p. 460] was the following:

Let t vary in any simple ordered set, and let $\{x_t\}$ be a family of chance variables. We shall say that chance variables have the property ε if whenever $t_1 < \cdots < t_{n+1}$

$$E[x_{t_1}, \cdots, x_{t_n}; x_{t_{n+1}}] = x_{t_n}, \tag{3.1}$$

with probability 1 ... the x_t are measurable functions defined on a space Ω, on certain sets of which a measure function is defined.

The notation for conditional expectation was adopted from Kolmogrov's *Foundations of the Theory of Probability* [25]. At that time, Doob, did not recognize the impact such a process could have. Later, in 1948, Doob started to give this concept the name martingale, which he borrowed from Ville's paper [37] of 1939. Jessen, who had by 1946 put his work of 1934 in a fully measure-probabilistic context [21] showed that almost sure convergence of an L^1-bounded martingale could also be achieved by considering a sequence of derivatives f_n, implicitly identified as a martingale, of η_n with respect to μ_n, where η_n is a signed measure and μ_n a measure, both restricted to a sub-σ-algebra $\mathcal{F}_n$. We have covered an identification of a martingale being a derivative of absolute continuous measures in **E2.11**. Since the idea of proving almost sure convergence for a sequence of densities evolved of the convergence of closed martingales, Jessen was able to carry over the closeness property for the density sequence f. Thus, the convergence of f is a special case of the convergence of a closed martingale. A treatment of this will be shown in the next section in **R3.5**. It was possible for Jessen to be less restrictive on his preassumptions of his theorem and reformulated it in [22] 1948, which Doob included and developed in his great treatise of 1953 [11].
After 1953, Moy, 1954, extended the term of conditional expectation to Bochner-integrable random variables [31]. Mathematicians like Chatterji, Scalora, C. Ionescu Tulcea and A. Ionescu Tulcea, began to study convergence behaviour of martingales who take their values in arbitrary Banach spaces and could obtain identical convergence results comparing to the real case [36, 32, 8, 6, 7]. Their approach to convergence was more functional analytic based as they set $\mathbf{E}[\cdot|\mathcal{F}_n]$ to be a bounded linear operator T_n on $L^p(\mathfrak{B})$, $\mathfrak{B}$ being the Banach-space the random variables of $L^p(\mathfrak{B})$ are valued in. A complete collection of these convergence theorems was published by Chatterji in 1960 [7]. Afterwards, 1965, Billingsley rearranged in his book about ergodic theory and information a similar proof of almost sure convergence for L^1-closed martingales in the real case [2, p. 116], which was regiven in a paper of Chatterji two years later [8, p .57]. Finally, 1970, the french probabilist Meyer modified Billingsley's proof to start a different approach to martingale convergence, avoiding the upcrossing inequality, by inverting the convergence pattern, going

$$ⓢ \rightarrow ⓞ \rightarrow ⓝ.$$

At first, we will prove theorem ⓝ directly using Doob's and Jessen's original proofs and follow with Meyer's convergence pattern.

3.2 Doob's proof

In his original paper [10, p. 460], Doob stated that a martingale converges almost surely if $\lim_{n\to\infty} \mathbf{E}[|X_n|] = K < \infty$, K an integer, holds. In newer textbooks, like Billingsley [3, p. 490] or Schilling [33, p. 192], the condition is replaced by $\sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|] < \infty$, which is equivalent to the first one if X is a martingale. Before formulating his martingale convergence theorem Doob postulated the maximal inequality **T2.24** [10, p. 458], which serves as a tool for proofs of the whole paper. The maximal inequality itself is proven exactly the same way as Doob did 13 years later in Stochastic Processes [11, p. 315]. Regarding the convergence theorem of martingales, both proofs of Doob show that the probability of having infinite many upcrossings is zero. In contrast to the proof, which uses the uprossing inequality, Doobs very first proof does not explicitly estimate a martingales sample paths but rather express them in sets with an asigned probability. A sketch of Doob's first proof for closable martingales can be found in Meyer [30, p. 74].

Theorem 3.1 (Doob 1940)**.** *Let X be a martingale. If $\lim_{n\to\infty} \mathbf{E}[|X_n|] = K < \infty$ holds, then X converges almost surely to an integrable random variable Y.*

Proof. Suppose there are two numbers $a < b$ and define a set $D := [\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n]$ assuming a positive probability η. Set $A_{n_1} := \{\sup_{0\leq j\leq n_1} X_j \geq b\}$ and choose the integer n_1 so large that $\mathbf{P}[D \cap A_{n_1}] > \eta(1 - 3^{-1})$ holds. In the same manner define a set $A_{n_2} := \{\inf_{n_1\leq j\leq n_2} X_j \leq a\}$ and choose the integer n_2 so large that $\mathbf{P}[D \cap A_{n_1} \cap A_{n_2}] > \eta(1 - 3^{-1} - 3^{-2})$ holds. Inductively, choose a sequence of integers $(n_j)_{j\in\mathbb{N}}$ such that

$$\mathbf{P}[D \cap A_{n_1} \cap A_{n_2} \cap \cdots \cap A_{n_j}] > \eta(1 - 3^{-1} - 3^{-2} - \cdots - 3^{-j})$$

holds. Obviously, if r is odd $A_{n_r} = \{\sup_{n_{r-1}\leq j\leq n_r} X_j \geq b\}$ and if r is even $A_{n_r} = \{\inf_{n_{r-1}\leq j\leq n_r} X_j \leq a\}$. Define a set $\Lambda_n := \bigcap_{j=1}^{n} A_{n_j}$ and estimate that

$$\mathbf{P}[\Lambda_n] > \mathbf{P}[D \cap \Lambda_n] > \eta\bigl(1 - \sum_{i=1}^{\infty} \frac{1}{3^i}\bigr) = \frac{\eta}{2}.$$

Now, if $m \geq 2r$, $m \in \mathbb{N}$, use the fact that X is a martingale and the conditional expectation property to evaluate

$$
\begin{aligned}
b\mathbf{P}[\Lambda_{2r-1}] &= b\mathbf{P}[A_{n_1} \cap \cdots \cap A_{n_{2r-1}}] \\
&\leq \int_{\Lambda_{2r-1}} X_j \, d\mathbf{P}, \ j \ \in \{n_{2r-2}, ..., n_{2r-1}\} \\
&= \int_{\Lambda_{2r-1}} \mathbf{E}[X_m|\mathcal{F}_j] \, d\mathbf{P} = \int_{\Lambda_{2r-1}} X_m \, d\mathbf{P}.
\end{aligned}
$$

Notice that this is nothing else but an application of inequality (2.15) of the maximal inequality proof. In the same way, one can show that

$$
a\mathbf{P}[\Lambda_{2r}] \geq \int_{\Lambda_{2r}} X_m \, d\mathbf{P}.
$$

For $M_r := \Lambda_{2r-1} \backslash \Lambda_{2r}$

$$
\int_{M_r} X_m \, d\mathbf{P} \geq b\mathbf{P}[\Lambda_{2r-1}] - a\mathbf{P}[\Lambda_{2r}] = (b-a)\mathbf{P}[\Lambda_{2r-1}] + a\mathbf{P}[M_r].
$$

Since $\Lambda_{2r} \subseteq \Lambda_{2r-1}$ the M_r are disjunct, and if m is sufficiently large

$$
\sum_{r=1}^{q} \int_{M_r} X_m \, d\mathbf{P} \geq (b-a) \sum_{r=1}^{q} \mathbf{P}[\Lambda_{2r-1}] + a \sum_{r=1}^{q} \mathbf{P}[M_r] \geq (b-a)q\frac{\eta}{2} - |a|.
$$

But in this case

$$
\begin{aligned}
K = \lim_{n\to\infty} \mathbf{E}[|X_n|] &\geq \int_{\Omega} |X_m| \, d\mathbf{P} \\
&\geq \sum_{r=1}^{q} \int_{M_r} X_m \, d\mathbf{P} \geq (b-a)q\frac{\eta}{2} - |a| \,,
\end{aligned}
$$

which is impossible to hold $\forall \ q \in \mathbb{N}$, due to our assumption $K < \infty$, unless $\mathbf{P}[D] = \eta = 0$. This means exactly that $X_n \xrightarrow{a.s.} Y$. By Fatou's lemma $\mathbf{E}[\liminf_{n\to\infty} |X_n|] \leq \liminf_{n\to\infty} \mathbf{E}[|X_n|] < K$ and Y is integrable.

□

3.3 Jessen's proof

Consider the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and let η be a σ-finite signed measure on this space. Write for $\mathbf{P}_n$ respectively η_n the restriction of $\mathbf{P}$ respectively η to the sub-σ-algebra $\mathcal{F}_n$. Assume that $\mathbf{P}_n$ is absolutely continuous on $\mathcal{F}_n$ with respect to η_n and let X_n be the density of η_n relative to $\mathbf{P}_n$. Additionally, define $\mathcal{F}_\infty := \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$. We have shown in **E2.11** that in this case the densities X_n form a martingale. The theorem of Jessen states that the limit of the density sequence X_n exists almost surely and hence the almost sure convergence of the martingale X. In Jessens original proof the probability measure $\mathbf{P}$ was replaced by a σ-finite measure μ. This proof here follows HEWITT-STROMBERG [18, p. 369].

Theorem 3.2 (JESSEN 1948)**.** *Let $\mathbf{P}_\infty$ and η_∞ be $\mathbf{P}$ and η restricted to $\mathcal{F}_\infty$. Then both functions*

$$\overline{X} = \limsup_{n\to\infty} X_n \qquad \underline{X} = \liminf_{n\to\infty} X_n$$

are derivatives of η_∞ with respect to $\mathbf{P}_\infty$. Thus: $\lim_{n\to\infty} X_n$ *exists almost surely and the $\mathbf{P}_\infty$-singular part of η_∞ is confined to the set* $\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = \pm\infty\}$.

In perspective of the preassumption of having a density X_∞ for $A \in \mathcal{F}_\infty$ we can use **E2.11**, the theorem of conditional expectation [A.9] and [A.2] to assert that X_∞ closes the martingale X, ergo, $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$, $\forall\ n \in \mathbb{N}$. This last sentence is here for plain information and will be shown with all necessary details in **R3.5**.

Proof. Let a be a real number. In this proof write $L_a := \{\omega \in \Omega : \underline{X} \leq a\}$ and $G_a := \left\{\omega \in \Omega : \overline{X} \geq a\right\}$. In view of [A.15], it suffices to prove

$$\eta\left\{L_a \cap A\right\} \leq a\mathbf{P}\left\{L_a \cap A\right\} \tag{3.2}$$

and

$$\eta\left\{G_a \cap A\right\} \geq a\mathbf{P}\left\{G_a \cap A\right\} \tag{3.3}$$

for all $A \in \mathcal{F}_\infty$. Note, that the conditions of [A.15] are already obtained since

$$\{\omega \in \Omega : \underline{X} \geq a\} \subset \left\{\omega \in \Omega : \overline{X} \geq a\right\}$$

and

$$\left\{\omega \in \Omega : \overline{X} \leq a\right\} \subset \{\omega \in \Omega : \underline{X} \leq a\}.$$

Let $(a_n)_{n\geq 1}$ be a strictly decreasing sequence of real numbers with limit a. For $n \in \mathbb{N}$ set

$$H_n := \Big\{\omega \in \Omega : \inf_{n\in\mathbb{N}} \{X_{n+1}, X_{n+2}, ...\} < a_n\Big\}$$

$$H_{n,1} := \Big\{\omega \in \Omega : X_{n+1} < a_n\Big\}$$

and

$$H_{n,p} := \Big\{\omega \in \Omega : \min_{n\in\mathbb{N}} \{X_{n+1}, ..., X_{n+p-1}\} \geq a_n, X_{n+p} < a_n\Big\}.$$

It is clear that $H_{n,p} \in \mathcal{F}_{n+p}$, that $(H_{n,p})_{p\geq 1}$ is a pairwise disjoint family, that $H_n = \bigcup_{p=1}^{\infty} H_{n,p}$ and that $L_a = \bigcap_{n=1}^{\infty} H_n$. Let A be any set in the algebra $\bigcup_{n=1}^{\infty} \mathcal{F}_n$, so that $A \in \bigcap_{n=n_0}^{\infty} \mathcal{F}_n$ for some n_0. The set $H_{n,p} \cap A$ is in $\mathcal{F}_{n+p}$ for $n \geq n_0$ and $p \geq 1$. We assemble all of these facts to write

$$\begin{aligned} \eta(H_n \cap A) &= \sum_{p=1}^{\infty} \eta(H_{n,p} \cap A) = \sum_{p=1}^{\infty} \eta_{n+p}(H_{n,p} \cap A) \\ &\leq \sum_{p=1}^{\infty} a_n \mathbf{P}_{n+p}(H_{n,p} \cap A) = a_n \mathbf{P}(H_n \cap A), \end{aligned} \tag{3.4}$$

for all $n \geq n_0$. Since $(H_1 \cap A) \supset (H_2 \cap A) \supset \cdots$ and $L_a \cap A = \bigcap_{n=1}^{\infty}(H_n \cap A)$, we can take the limit in (3.4) to write

$$\eta(L_a \cap A) = \lim_{n\to\infty} \eta(H_n \cap A) \leq \lim_{n\to\infty} a_n \mathbf{P}(H_n \cap A) = a\mathbf{P}(L_a \cap A).$$

and due to the fact that η is σ-finite on $\mathcal{F}_1$ we have obtained inequality (3.2) for all $A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n$.
To see that (3.2) is valid for all $A \in \mathcal{F}_\infty$ let $(G_n)_{n\geq 1}$ be a disjoint family of sets in $\mathcal{F}_1$ such that $\Omega = \bigcup_{n=1}^{\infty} G_n$. Let ν_n be the set function on $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ defined by

$$\nu_n(A) = a\mathbf{P}(G_n \cap L_a \cap A) - \eta(G_n \cap L_a \cap A).$$

A routine computation shows that ν_n is a countably additive, non-negative, finite-valued measure on the algebra $\bigcup_{n=1}^{\infty} \mathcal{F}_n$. Let ν be the set function $\sum_{n=1}^{\infty} \nu_n$, which is also defined only on $\bigcup_{n=1}^{\infty} \mathcal{F}_n$. It is easy to see that ν is a non-negative, countably additive, finite measure on $\bigcup_{n=1}^{\infty} \mathcal{F}_n$. Now, ν admits a unique countably additive extension over the σ-algebra $\mathcal{F}_\infty$. All this implies that (3.2) holds not only for $A \in \bigcup_{n=1}^{\infty} \mathcal{F}_n$ but for all $A \in \mathcal{F}_\infty$. Inequality (3.3) is proven the exact same way. Hence X converges almost surely.

$\square$

3.4 Meyer's proof

The third way to proof theorem ⓝ is by inversion of the usual convergence pattern, so instead of going ⓝ → ⓞ → ⓢ we will go ⓢ → ⓞ → ⓝ. Meyer sketched it, 1970, nonpedagogically in [30, p. 30].

3.4.1 Convergence of closable martingales

Definition 3.3. A martingale X is said to be closed in $\mathcal{L}^p$, $1 \leq p < \infty$, by a random variable Y if $Y \in \mathcal{L}^p$ and $X_n = \mathbf{E}[Y|\mathcal{F}_n]$, for every $n \in \mathbb{N}$.

If Y closes X so does $\mathbf{E}[Y|\mathcal{F}_\infty]$.

In **E2.8** we reckoned that a process X defined by $X_n = \mathbf{E}[Y|\mathcal{F}_n]$ for an integrable variable Y is a martingale.

For a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ we define once again $\sigma(\bigcup_{n=0}^{\infty} \mathcal{F}_n) := \mathcal{F}_\infty$. Obviously $\mathcal{F}_n \subset \mathcal{F}_\infty$ for all $n \in \mathbb{N}$.

Theorem 3.4. *Let X be a martingale, closed in $\mathcal{L}^p$, $1 \leq p < \infty$, by a random variable Y. Then X converges a.s. and in $\mathcal{L}^p$ to $\mathbf{E}[Y|\mathcal{F}_\infty]$.*
$X_\infty = \lim_{n\to\infty} X_n$ is the only $\mathcal{F}_\infty$-measurable random variable, which closes X in $\mathcal{L}^1$, so that the extended martingale property $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$ holds.

This is an extension of theorem ⓢ.

Proof. We start with the case $p = 1$, and the assumption that $\mathcal{F} = \mathcal{F}_\infty$. Denote by $\mathcal{H}$ the set

$$\mathcal{H} := \left\{ Y \in \mathcal{L}^1 : X_n = \mathbf{E}[Y|\mathcal{F}_n] \overset{n\to\infty}{\to} Y \ a.s. \text{ and in } \mathcal{L}^1 \right\}.$$

Our goal is to identify $\mathcal{H}$ with $\mathcal{L}^1$ so that no matter which $Y \in \mathcal{L}^1$ closes X, the convergence *a.s.* and in $\mathcal{L}^1$ for X holds.
If $n \geq k$, due to (2.1), **D3.3** and $\mathcal{F}_k \subseteq \mathcal{F}_n$-measurablility of Y

$$X_n = \mathbf{E}[X_{n+1}|\mathcal{F}_n] = \mathbf{E}[\mathbf{E}[Y|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbf{E}[Y|\mathcal{F}_n] = Y,$$

and $\mathcal{L}^1(\mathcal{F}_k) \in \mathcal{H}$ for every finite k. Now $\bigcup_{k\in\mathbb{N}} \mathcal{L}^1(\mathcal{F}_k)$ is dense in $\mathcal{L}^1(\mathcal{F}_\infty) = \mathcal{L}^1(\mathcal{F})$ and in order to prove that $\mathcal{H} = \mathcal{L}^1(\mathcal{F})$ we just have to show that $\mathcal{H}$ is closed.
Say $Z \in \overline{\mathcal{H}}$. Then there exists a sequence $(Z_n) \in \mathcal{H}$ with $\|Z - Z_n\|_1 \leq 2^{-n}$.
Set $X_{n,k} := \mathbf{E}[Z_k|\mathcal{F}_n]$ and let Z close the martingale X.
Verify by

$$\mathbf{E}[X_{n+1} - X_{n+1,k}|\mathcal{F}_n] = \mathbf{E}[\mathbf{E}[Z - Z_k|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbf{E}[Z - Z_k|\mathcal{F}_n] = X_n - X_{n,k},$$

that $X_n - X_{n,k}$ is a martingale (this would also follow by **E2.7**, since the difference of two martingales is again a martingale). We can now apply the maximal inequality **T2.24** on $X_n - X_{n,k}$ and estimate, with Jensen's inequality for conditional expectations [A.10],

$$\begin{aligned}
\mathbf{P}[\sup_{n\in\mathbb{N}} |X_n - X_{n,k}| > \frac{1}{k}] &\leq k \sup_{n\in\mathbb{N}} \|X_n - X_{n,k}\|_1 \\
&= k \sup_{n\in\mathbb{N}} \|\mathbf{E}[Z - Z_k|\mathcal{F}_n]\|_1 \\
&= k \sup_{n\in\mathbb{N}} \int_\Omega |\mathbf{E}[Z - Z_k|\mathcal{F}_n]| \; d\mathbf{P} \\
&\leq k \sup_{n\in\mathbb{N}} \int_\Omega \mathbf{E}[|Z - Z_k|\,|\mathcal{F}_n] \; d\mathbf{P} \\
&= k \sup_{n\in\mathbb{N}} \mathbf{E}[\mathbf{E}[|Z - Z_k|\,|\mathcal{F}_n]] = k \sup_{n\in\mathbb{N}} \mathbf{E}[|Z - Z_k|] \\
&= k\|Z - Z_k\|_1 \leq \frac{k}{2^k}.
\end{aligned}$$

If we set $A_k := \left\{\sup_{n\in\mathbb{N}} |X_n - X_{n,k}| > \frac{1}{k}\right\}$ and sum over $\mathbf{P}[A_k]$ we see, by applying the ratio test, that

$$\sum_{k=0}^{\infty} \mathbf{P}[A_k] < \sum_{k=0}^{\infty} \frac{k}{2^k} < \infty$$

and the first Borel-Cantelli lemma [A.13] states that $\mathbf{P}[\limsup_{k\to\infty} A_k] = 0$, which implies that $X_{n,k}$ converges uniformly in n to X_n, for almost every $\omega \in \Omega$. In that matter, the limit of X is preserved under uniform convergence in n and X converges almost surely to Z. Next we have to prove the $\mathcal{L}^1$ convergence of X to Z and thus we estimate

$$\begin{aligned}
\|Z - X_n\|_1 &\leq \|Z - Z_k\|_1 + \|Z_k - X_{n,k}\|_1 + \|X_{n,k} - X_n\|_1 \\
&\leq 2^{-k} + \|Z_k - X_{n,k}\|_1 + \|X_{n,k} - X_n\|_1.
\end{aligned}$$

Now, the second member of the last inequality is definitely smaller than 2^{-k} for n large enough since $Z_k \in \mathcal{H}$ and this implies $\|Z_k - X_{n,k}\|_1 \to 0$. For the last member follows

$$\begin{aligned}
\|X_{n,k} - X_n\|_1 &= \int_\Omega |\mathbf{E}[Z_k - Z|\mathcal{F}_n]| \; d\mathbf{P} \\
&\leq \int_\Omega \mathbf{E}[|Z_k - Z|\,|\mathcal{F}_n] \; d\mathbf{P} = \mathbf{E}[\mathbf{E}[|Z_k - Z|\,|\mathcal{F}_n]] \\
&= \mathbf{E}[|Z_k - Z|] = \|Z_k - Z\|_1 \leq 2^{-k}
\end{aligned}$$

and so $\|Z - X_n\|_1 \leq 3 \cdot 2^{-k}$, for n large enough. Thus convergence of X to Z in $\mathcal{L}^1$ follows and we have proven that $\mathcal{H}$ is closed. This settles the case $p = 1$, $\mathcal{F} = \mathcal{F}_\infty$. If $p = 1$, $\mathcal{F} \neq \mathcal{F}_\infty$, and $Y \in \mathcal{L}^1(\mathcal{F})$ closes X, apply the above results to $(\Omega, \mathcal{F}_\infty, \mathbf{P})$, and the random variable $\mathbf{E}[Y|\mathcal{F}_\infty]$ which closes X_n on this space.

Finally, if $Y \in \mathcal{L}^p$, $p > 1$, we have

$$\begin{aligned}
\|X\|_p &= \sup_{n\in\mathbb{N}} \|X_n\|_p = \sup_{n\in\mathbb{N}} \|\mathbf{E}[Y|\mathcal{F}_n]\|_p \\
&= \sup_{n\in\mathbb{N}} (\int_\Omega |\mathbf{E}[Y|\mathcal{F}_n]|^p \ d\mathbf{P})^{\frac{1}{p}} \leq \sup_{n\in\mathbb{N}} (\int_\Omega \mathbf{E}[|Y|^p \, |\mathcal{F}_n] \ d\mathbf{P})^{\frac{1}{p}} \\
&= \sup_{n\in\mathbb{N}} (\mathbf{E}[\mathbf{E}[|Y|^p \, |\mathcal{F}_n]])^{\frac{1}{p}} = \sup_{n\in\mathbb{N}} (\mathbf{E}[|Y|^p])^{\frac{1}{p}} = \|Y\|_p.
\end{aligned}$$

and due to **T2.25** we have $\| \sup_{n\in\mathbb{N}} |X_n| \, \|_p \leq q\|Y\|_p$. Therefore dominating convergence of X takes place in $\mathcal{L}^p$ and the limit X_∞ exists almost surely since X is uniformly integrable by Vitali's **T1.10** and hence uniformly bounded in $\mathcal{L}^1$. The last sentence of the theorem is obvious: If Y is $\mathcal{F}_\infty$-measurable, then $X_\infty = \mathbf{E}[Y|\mathcal{F}_\infty] = Y$. Now $X_\infty \in \mathcal{L}^1$ closes X so that $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$, for every $n \in \mathbb{N}$.

□

Remark 3.5. Assume a probability measure $\mathbf{P}$ on $\mathcal{F}_\infty$ and let η be a signed measure, which is absolutely continuous with respect to $\mathbf{P}$ on $\mathcal{F}_\infty$. Let X_n be the density of η_n with respect to $\mathbf{P}_n$, where η_n and $\mathbf{P}_n$ are η and $\mathbf{P}$ both restricted to $\mathcal{F}_n$. Recalling **E2.11** the density sequence X is a martingale and under the present hypothesis

$$\eta_n(A) = \int_A X_n \ d\mathbf{P} = \int_A X_\infty \ d\mathbf{P} = \eta_\infty(A), \ \forall \ A \in \mathcal{F}_n.$$

Due to the existence of conditional expectation $\mathbf{E}[X_\infty|\mathcal{F}_n]$ and [A.2]

$$X_n = \mathbf{E}[X_\infty|\mathcal{F}_n], \ \forall \ n \in \mathbb{N}.$$

According to the previous **T3.4**, $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$ converges almost surely to $\mathbf{E}[X_\infty|\mathcal{F}_\infty] = X_\infty$ as n tends to infinity and Jessen's theorem becomes a special case of **T3.4**.

3.4.2 Convergence of uniformly integrable martingales

Closability of a martingale X implies strong convergence properties and because of the resulting boundedness of X in $\mathcal{L}^p$, $1 \leq p < \infty$, we can link it directly with uniform integrability, which we discussed detailed in section 1.

Theorem 3.6. *(1) A martingale X is closable in $\mathcal{L}^1$ if and only if it is uniformly integrable.*
(2) A martingale X is closable in $\mathcal{L}^p$ ($p > 1$) if and only if it is bounded in $\mathcal{L}^p$.

This corresponds extensively to theorem ⓞ.

Proof. *(1)* $\Rightarrow$:
If a martingale X is closable in $\mathcal{L}^1$, then it converges $a.s.$ and in $\mathcal{L}^1$ and due to **T1.10** is uniformly integrable.
(1) $\Leftarrow$:
Conversely, if X is uniformly integrable, the compactness criterion of Dunford-Pettis **T1.16** reveils that every sequence of the martingale X contains a subsequence X_{n_k}, which converges weakly in $\mathcal{L}^1$ to some random variable $Z \in \mathcal{L}^1$. Otherwisely stated, for $Z \in \mathcal{L}^1$:

$$\lim_{k\to\infty} \int_A X_{n_k} \; d\mathbf{P} = \int_A Z \; d\mathbf{P}, \; \forall \; A \in \mathcal{F}.$$

Now take $A \in \mathcal{F}_m$ and $n_k > m$, then

$$\begin{aligned}
\mathbf{E}[X_{n_k}|\mathcal{F}_m] = X_m &\Leftrightarrow \int_A X_{n_k} \; d\mathbf{P} = \int_A X_m \; d\mathbf{P} \\
&\Leftrightarrow \lim_{k\to\infty} \int_A X_{n_k} \; d\mathbf{P} = \int_A Z \; d\mathbf{P} = \int_A X_m \; d\mathbf{P} \\
&\Leftrightarrow \mathbf{E}[Z|\mathcal{F}_m] = X_m,
\end{aligned}$$

and thus X is closed by Z in $\mathcal{L}^1$.
(2) $\Rightarrow$:
Suppose now that X is closed in $\mathcal{L}^p$, then X is obviously bounded in $\mathcal{L}^p$.
(2) $\Leftarrow$:
If X is bounded in $\mathcal{L}^p$ the theorem of la Valleé Poussin **T1.11** yields that X is uniformly integrable and due to the previous statement *(1)* in this theorem it is also closable in $\mathcal{L}^1$. It follows that X converges $a.s.$ to a closable variable X_∞. In order, for X, to be closable in $\mathcal{L}^p$ we still have to show that $X_\infty \in \mathcal{L}^p$. Since X is bounded n $\mathcal{L}^p$, this is easily deduced by Fatou's lemma [A.4] $\|X_\infty\|_p \leq \liminf_{n\to\infty} \|X_n\|_p < \infty$ and it follows that X_∞ closes the martingale.

□

3.4.3 Convergence of $\mathcal{L}^1$-bounded martingales

3.4.4 The Krickeberg Decomposition Lemma

The following lemma shortens the proof for a much deeper result, namely, $a.s.$ convergence for $\mathcal{L}^1$-bounded martingales. It is called the *Krickeberg decomposition lemma.*

Lemma 3.7 (Krickeberg Decomposition)**.** *A martingale X is bounded in $\mathcal{L}^1$ if and only if it can be written as a difference of two positive martingales $M_n^{(1)}$ and $M_n^{(2)}$*

$$X_n = M_n^{(1)} - M_n^{(2)}.$$

Proof. $\Leftarrow$:
If $M_n^{(1)}$ and $M_n^{(2)}$ are two positive martingales they are obviously $\mathcal{L}^1$-bounded and so are their differences.
$\Rightarrow$:
Set X^+ the positive and X^- the negative part of the process X. We then decompse X as:

$$X_n = X_n^+ - X_n^-.$$

Set $Y_{m,n} := \mathbf{E}[X_m^+|\mathcal{F}_n]$ for $m \geq n$ and check that

$$Y_{m+1,n} = \mathbf{E}[X_{m+1}^+|\mathcal{F}_n] = \mathbf{E}[\mathbf{E}[X_{m+1}^+|\mathcal{F}_m]|\mathcal{F}_n] \geq \mathbf{E}[X_m^+|\mathcal{F}_n] = Y_{m,n},$$

which implies that $Y_{m,n}$ is monotonically increasing in m.
Take $M_n^{(1)} := \lim_{m\to\infty} Y_{m,n}$. We will show that $M_n^{(1)}$ is a martingale. First, verify that due to the monotone convergence theorem [A.16] and the boundedness assumption on X that

$$\begin{aligned}\mathbf{E}[M_n^{(1)}] &= \mathbf{E}[\lim_{m\to\infty} Y_{m,n}] = \lim_{m\to\infty} \mathbf{E}[Y_{m,n}] \\ &= \lim_{m\to\infty} \mathbf{E}[\mathbf{E}[X_m^+|\mathcal{F}_n]] = \lim_{m\to\infty} \mathbf{E}[X_m^+] < \infty.\end{aligned}$$

Thus $Y_{m,n}$ converges almost surely to an integrable random variable $M_n^{(1)}$. Because of its definition $M_n^{(1)}$ is $\mathcal{F}_n$-measurable and the martingale condition (2.1) is proven by using the monotone convergence theorem for conditional expectations

$$\begin{aligned}\mathbf{E}[M_{n+1}^{(1)}|\mathcal{F}_n] &= \mathbf{E}[\lim_{m\to\infty} \mathbf{E}[X_m^+|\mathcal{F}_{n+1}]|\mathcal{F}_n] \\ &= \lim_{m\to\infty} \mathbf{E}[X_m^+|\mathcal{F}_{n+1}|\mathcal{F}_n] = \lim_{m\to\infty} Y_{m,n} = M_n^{(1)}.\end{aligned}$$

Setting $Z_{m,n} := \mathbf{E}[X_m^-|\mathcal{F}_n]$, $M_n^{(2)} := \lim_{m\to\infty} Z_{m,n}$ and following the same steps as above yields that $M_n^{(2)}$ is a martingale too. The last step will show that X is the difference of $M_n^{(1)}$ and $M_n^{(2)}$:

$$\begin{aligned}M_n^{(1)} - M_n^{(2)} &= \lim_{m\to\infty} Y_{m,n} - \lim_{m\to\infty} Z_{m,n} = \lim_{m\to\infty} Y_{m,n} - Z_{m,n} \\ &= \lim_{m\to\infty} \mathbf{E}[X_m^+ - X_m^-|\mathcal{F}_n] = \lim_{m\to\infty} \mathbf{E}[X_m|\mathcal{F}_n] = X_n.\end{aligned}$$

□

Our efforts have paid-off as we are about to witness the highly anticipated convergence theorem for $\mathcal{L}^1$-bounded martingales. The proof of this theorem doesn't appear in standard math books. However, we are more than happy to be able to formulate it here.

Theorem 3.8. *Any $\mathcal{L}^1$-bounded martingale X converges a.s. to an a.s. random variable.*

This is the desired theorem ⓝ.

Proof. According to the Krickeberg decomposition it suffices to prove this theorem for a positive martingale X.
Let N be an integer and set

$$T = \inf_{m \in \mathbb{N}} \{m : X_m \geq N\}$$

and denote by Y the martingale X stopped at time T, i.e. $Y = X^T$.
We will observe X_T on the sets $\{T < \infty\}$ and $\{T = \infty\}$.
Suppose that X_T is only defined on $\{T < \infty\}$ and set $X_T = 0$ on $\{T = \infty\}$.
We remark that X_T is integrable, since
$\mathbf{E}[X_T] \leq \liminf_{n \to \infty} \mathbf{E}[X_{T \wedge n}] = \mathbf{E}[X_0]$, due to Fatou's lemma and the optional sampling theorem **T2.21**. Y is dominated by the integrable random variable $N \vee X_T$, since $Y_n \leq N$ for $n < T$, $Y_n = X_T$ for $n \geq T$. Therefore Y is uniformly integrable and **T3.6** and **T3.4** yield that Y converges *a.s.* to an integrable random variable.
Let's focus on X being defined on the set $\{T = \infty\}$. Fix an $\omega \in \{T = \infty\}$: Then $X_n(\omega) = Y_n(\omega)$, for every $n \in \mathbb{N}$ and every fixed $\omega \in \{T = \infty\}$, and thus X also converges *a.s.* on this set. Note that if X is $\mathcal{L}^1$-bounded, X_∞ is not only finite, but also integrable due Fatou's lemma [A.4].

□

This concludes Meyers proof.

3.5 Convergence of $\mathcal{L}^1$-bounded supermartingales

It is possible to extend **T3.8** to sub/supermartingales not interfering, again, with the upcrossing inequality ⓤ.

Theorem 3.9. *Let X be a supermartingale such that* $\sup_{n \in \mathbb{N}} \mathbf{E}[X_n^-] < \infty$. *Then X converges a.s. to an a.s. random variable.*

Proof. Without restriction of generality we assume that $X_0 = 0$. Since $\mathbf{E}[X_0] \geq \mathbf{E}[X_n] \geq -\mathbf{E}[X_n^-]$, $\mathbf{E}[X_n]$ remains bounded, and so does $\mathbf{E}[|X_n|] = \mathbf{E}[X_n] + 2\mathbf{E}[X_n^-]$. Consider Doob's decomposition $X = M - A$, where A is increasing with $A_0 = 0$, $M = X + A$ is a martingale. We have $M_n^- \leq X_n^-$, therefore $\mathbf{E}[M_n^-]$ remains bounded. Since $\mathbf{E}[M_n] = \mathbf{E}[M_0] = 0$, $\mathbf{E}[|M_n|] = 2\mathbf{E}[M_n^-]$, and M is $\mathcal{L}^1$-bounded. According to **T3.8**, M converges *a.s.* to an integrable random variable. On the other hand, A_n increases, and $\mathbf{E}[A_n] = \mathbf{E}[M_n] - \mathbf{E}[X_n]$ remains bounded, therefore A_n converges *a.s.* to an integrable random variable.

□

To prove the $\mathcal{L}^1$ and *a.s.* convergence of an uniformly integrable supermartingale comes very easy now. Due to **T1.10** we can extend our result to convergence in $\mathcal{L}^p$, $1 \leq p < \infty$.

Theorem 3.10. *Let X be an uniformly integrable supermartingale. Then X converges in $\mathcal{L}^p$ and a.s.*

Proof. Combine **T1.5** with **T3.9** and **T1.10**.

□

3.6 Comparison of the proofs

If someone reads a comparison test about distinct electronic devices, the reader may not be necessarily interested in a product specific detailed description but rather in an overview or rating of each tested product. Due to that, the products that were tested here are the various proofs of the martingale convergence theorem ⓝ. Our testing criteria are

I Expenditure.

II Common ground of the proofs without upcrossings with the standard proof using upcrossings.

We are going to review Doob's, Jessen's and Meyer's proof.

I Expenditure

(a) **Doob's proof**
For a complete understanding it is sufficient to study the martingale section 2 of this thesis. None of the appendix or section 1 related results on uniformly integrability is needed.

(b) **Jessen's proof**
This proof requires a certain flexibility towards measure theory, especially for density functions of two absolute continuous measures. A mathematician with a profound measure theoretic background won't have trouble getting through the details without reading any of the preceding sections. Without this knowledge but with comprehension of section 1 and 2 this proof should be outfigurable.

(c) **Meyer's proof**
Now this proof is definitely the hardest and longest of the three, considering the number of theorems needed that are proved previously. Knowledge about section 1, 2 and the appendix is unavoidable. Its kind of a bee, which knows the easiest way in the garden to access the blossom with the sweetest pollen but instead of approaching it directly the bee chooses a path full of wild and exotic blossoms, which seem to have magical taste, that provoke its appetite to fly on such unknown paths even more. Needless to say that the sweetest pollen became even sweeter. Expenditure reaches its maximum.

II Common ground of the proofs without upcrossings with the standard proof using upcrossings.

(a) **Doob's proof**
Certainly the one with the highest similarity to the standard proof using upcrossings.

(b) **Jessen's proof**
An access that does not explicitly point to the proof with upcrossings but more implicitly and from a set-theoretic point of view.

(c) **Meyer's proof**
Probably the one with the smallest intersection.

By reviewing the three proofs it seems that Doob's and Jessen's proof oppose, in terms of length and expenditure, the one Meyer postponed. Deciding which proof one prefers as an alternative choice one with an affinity towards probability theory should go with Doob. If someone belongs more to the measure theoretic camp the proof of Jessen will serve well. It has to be mentioned that the standard proof of theorem ⓝ is by far the one with the least amount of difficulty comparing the rest. For a quick understanding of martingale almost sure convergence the proof with upcrossings is definitely the best choice. Though, for mathematicians eager to expand their knowledge of tools used to show martingale convergence, the proof of Jessen and especially the proof of Meyer cannot be neglected. The arsenal of math needed to proof theorem ⓝ the way Meyer did is of great importance for any mathematician.

A Measure and Probability Prequisites

This appendix provides basic statements about measure and probability theory. Most of them are formulated for random variables being defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, the rest for measurable variables on an arbitrary measure space $(X, \mathcal{A}, \mu)$. Since a probability space is a special case of a measure space, the results below, which are stated for measure spaces also hold for probability spaces.
To emphasize the measure theoretic deduction of assertions given for probability spaces we sometimes write $\int_\Omega X \, d\mathbf{P}$ instead of $\mathbf{E}[X]$.
All results listed here accompanied with lots of deeper and exciting information can be found in [13, 3, 9, 14, 17, 24, 23, 33].

Theorem A.1 (Lebesgue 1910). *Let $f_n, f : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be measurable und let $\lim_{n\to\infty} f_n = f$ almost surely. Further assume that there exists a function $g \in \mathcal{M}^+$, such that $|f_n| \leq g$ almost surely, for all $n \in \mathbb{N}$. Then f and f_n are integrable, $\forall\, n \in \mathbb{N}$, and*

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mathbf{P} = \int_\Omega f \, d\mathbf{P}$$

and

$$\lim_{n\to\infty} \int_\Omega |f_n - f| \, d\mathbf{P} = 0.$$

Lemma A.2. *Two random variables f, g both integrable and measurable on a sub-σ-algebra $\mathcal{G}$ are almost surely equal, $f = g$, if and only if*

$$\int_A f \, d\mathbf{P} = \int_A g \, d\mathbf{P}, \ \forall \ A \in \mathcal{G}.$$

Theorem A.3 (Egorov 1911). *Let $(f_n)_{n\geq 1}$ be a sequence of random variables and let $\lim_{n\to\infty} f_n = f$ almost surely, where f is a random variable. Then $(f_n)_{n\geq 1}$ converges almost uniformly to f.*

Lemma A.4 (Fatou 1911). *If $(f_n)_{n\geq 1}$ is a sequence of measurable functions then*

$$\int_\Omega \liminf_{n\to\infty} f_n \, d\mathbf{P} \leq \liminf_{n\to\infty} \int_\Omega f_n \, d\mathbf{P}.$$

Theorem A.5 (Riesz 1910). *If a sequence of random variables $(f_n)_{n\geq 1}$ converges in probability to a random variable f, then there exists a subsequence $(f_{n_k})_{k\geq 1}$, that converges almost surely to f.*

Theorem A.6 (Fubini 1907). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two σ-finite measure spaces and denote with $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ the product measure space. If $f : X \times Y \to \mathbb{R}$ is $\mu \otimes \nu$-integrable, then*

(1) $f(x, \cdot)$ is ν-integrable for μ-a.e. $y \in Y$.

(2) $f(\cdot, y)$ is μ-integrable for ν-a.e. $x \in X$.

(3) The functions

$$x \to \int_Y f(x, y)\ d\nu \quad \textit{resp.} \quad y \to \int_X f(x, y)\ d\mu$$

are μ-integrable resp. ν-integrable.

(4)

$$\begin{aligned}\int_{X\times Y} f(x, y)\ d(\mu \otimes \nu) &= \int_X \left(\int_Y f(x, y)\ d\nu \right) d\mu \\ &= \int_Y \left(\int_X f(x, y)\ d\mu \right) d\nu.\end{aligned}$$

Definition A.7 (Absolute Continuity of Measures). It μ, ν are two measures on the same measurable space, then μ is said to be absolutely continuous with respect to ν, or dominated by ν, if for every $A \in \mathcal{A}$: $\mu(A) = 0 \Rightarrow \nu(A) = 0$. This is written as $\mu \ll \nu$.

Theorem A.8 (Radon-Nikodým). *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, ν a signed measure on $\mathcal{A}$ and $\nu \ll \mu$. Then ν has a density in respect to μ, thus, there exists a quasiintegrable function $f : X \to \mathbb{R}$, such that*

$$\nu(A) = \int_A f\ d\mu,\ \forall\ A \in \mathcal{A},$$

and f is μ-a.e. uniquely determined. If ν is a measure f can be choosen ≥ 0.

Theorem A.9 (Conditional Expectation). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, Y a random variable with $\mathbf{E}[|Y|] < \infty$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Then there exists a $\mathcal{G}$-measurable function $\mathbf{E}[Y|\mathcal{G}]$, such that*

$$\int_A Y\ d\mathbf{P} = \int_A \mathbf{E}[Y|\mathcal{G}]\ d\mathbf{P},\ \forall\ A \in \mathcal{G},$$

holds almost surely.

Theorem A.10 (Jensen Inequality for conditional expectation). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, Y an integrable random variable and $\phi : \mathbb{R} \to \mathbb{R}$ a convex function. Then*

$$\phi\left(\mathbf{E}[Y|\mathcal{G}]\right) \leq \mathbf{E}[\phi\left(Y\right)|\mathcal{G}],$$

holds almost surely.

Theorem A.11 (Hölder Inequality 1889). *Let $(X, \mathcal{A}, \mu)$ be a measure space, $f, g : X \to \mathbb{R}$ two measurable functions and $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\int_X |fg| \; d\mu \leq \left(\int_X |f|^p \; d\mu \right)^{1/p} \left(\int_X |g|^q \, d\mu \right)^{1/q} .$$

Theorem A.12 (Markov Inequality). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, Y a random variable and $k \in \mathbb{N}$. Then*

$$\mathbf{P}[|Y| \geq \alpha] \leq \frac{1}{\alpha^k} \mathbf{E}[|Y|^k].$$

Lemma A.13 (Borell-Cantelli). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. If $\sum_{n=1}^{\infty} \mathbf{P}(A_n)$ converges, then $\mathbf{P}(\limsup_{n\to\infty} A_n) = 0$.*

Theorem A.14 (Halmos 1950). *Let $\mathcal{C}$ be a field and $\mathcal{M}$ a monotone class that contains $\mathcal{C}$. Then this implies $\sigma(\mathcal{C}) \subset \mathcal{M}$.*

Lemma A.15. *Let g be an extended real-valued, $\mathcal{F}$-measurable function on Ω. For each $a \in \mathbb{R}$, let $G_a = \{\omega \in \Omega : g(\omega) \geq a\}$ and $L_a = \{\omega \in \Omega : g(\omega) \leq a\}$. The function g is a derivative of η with respect to μ if and only if the following conditions obtain:*
(1) for every $a \in \mathbb{R}$ and every $A \in \mathcal{F}$, we have

$$\eta(G_a \cap A) \geq a\mu(G_a \cap A),$$

(2) for every $a \in \mathbb{R}$ and every $A \in \mathcal{F}$, we have

$$\eta(L_a \cap A) \leq a\mu(L_a \cap A).$$

Theorem A.16 (Levi 1906). *For every increasing sequence $(f_n)_{n\in\mathbb{N}}$ of positive measurable functions holds*

$$\int_\Omega (\lim_{n\to\infty} f_n) \; d\mathbf{P} = \lim_{n\to\infty} \int_\Omega f_n \; d\mathbf{P}.$$

B Alternative Proofs

The interested reader will find here some alternative proofs to theorems of the text.

Theorem B.1 (La Valleé Poussin 1937). *Let $\mathcal{H}$ be a subset of $\mathcal{L}^1$. The following properties are equivalent:*
(1) $\mathcal{H}$ is uniformly integrable.
(2) There exists a function $G(t)$ defined on $\mathbb{R}^+$, which is positive, increasing and convex[4]*, such that*

$$\lim_{t\to+\infty}\frac{G(t)}{t}=+\infty,$$

and

$$\sup_{f\in\mathcal{H}}\mathbf{E}[G\circ|f|]<+\infty.$$

Proof. (1) ⇒ (2):
Without loss of generality assume that $f\geq 0$. Uniform integrability of f means that

$$\forall\, n\in\mathbb{N},\exists\ \widetilde{c_n}\colon\ \sup_{f\in\mathcal{H}}\int_{\{f\geq\widetilde{c_n}\}}f\ d\mathbf{P}\leq 2^{-n}.$$

Set $c_n:=\widetilde{c_n}\vee n$. This sequence c_n diverges to infinity and we get

$$\begin{aligned}\sum_{n=0}^{\infty}\int_{\{f\geq c_n\}}f\ d\mathbf{P}&=\sum_{n=0}^{\infty}\sum_{m=c_n}^{\infty}\int_{\{m\leq f<m+1\}}f\ d\mathbf{P}\\&\leq\sum_{n=0}^{\infty}\sum_{m=c_n}^{\infty}m\mathbf{P}[m\leq f<m+1]\\&=\sum_{m=0}^{\infty}m\mathbf{P}[m\leq f<m+1]\sum_{n:c_n\leq m}^{\infty}1,\ \text{set } g(m):=\sum_{n:c_n\leq m}^{\infty}1,\\&=\sum_{m=0}^{\infty}mg(m)\mathbf{P}[m\leq f<m+1].\end{aligned}$$

Of course $\lim_{m\to\infty}g(m)=\infty$ and this implies $\lim_{m\to\infty}\frac{mg(m)}{m}=\infty$. Now set $G(m):=mg(m)$, then

$$\begin{aligned}\infty&>\sum_{m=0}^{\infty}G(m)\mathbf{P}[m\leq f<m+1]\\&=\int_{\Omega}\sum_{m=0}^{\infty}G(m)\mathbf{1}_{[m,m+1)}f(\omega)\ d\mathbf{P}=\mathbf{E}[G(f)],\end{aligned}$$

[4]Convexity of f is not used in the proof

with $G(z) = \sum_{m=0}^{\infty} G(m)\mathbf{1}_{[m,m+1)}(z)$.

□

An even shorter proof of implication (1) ⇒ (3) of Dunford-Pettis' theorem **T1.16** can be shown by using the following theorem:

Theorem B.2. *Any L^2-bounded sequence has a subsequence that converges weakly in L^2.*

Theorem B.3 (Dunford-Pettis 1953)**.** *Every uniformly integrable sequence of random variables has a subsequence that converges weakly in L^1.*

Proof. Let X be uniformly integrable. Define $X_n^k := X_n \mathbf{1}_{\{|X_n| \leq k\}}$ and note by

$$\int_\Omega \left| X_n \mathbf{1}_{\{|X_n| \leq k\}} \right|^2 \, d\mathbf{P} = \int_{\{|X_n| \leq k\}} |X_n|^2 \, d\mathbf{P} \leq k^2 \mathbf{P}[|X_n| \leq k] < \infty,$$

that $(X_n^k)_{n \in \mathbb{N}}$ is L^2-bounded for each $k \in \mathbb{N}$. By **B.2** and a diagonalization argument, there exists a subsequence and some random variables $\eta_1, \eta_2, \eta_3, \ldots$ such that for each k, $X_{n_r}^k \to \eta_k$, holds weakly in L^2 and because of inclusion also in L^1, as r tends to infinity.
Now, by Fatou's lemma [A.4], $\|\eta_k - \eta_l\|_1 \leq \liminf_{r \to \infty} \|X_{n_r}^k - X_{n_r}^l\|_1$, and due to uniform integrability **D1.1**, the right-hand side tends to 0 as $k, l \to \infty$. Thus the sequence $(\eta_n)_{n \in \mathbb{N}}$ is Cauchy in L^1 and so, it converges in L^1 towards some ξ. Take $\gamma \in L^\infty$, let μ be the weak limit of X_{n_r} and note that for any $n, k \in \mathbb{N}$

$$\mathbf{E}[\gamma(X_n - \mu)] \leq \|\gamma\|_\infty \sup_n \|X_n - X_{n_r}^k\|_1 + \mathbf{E}[\gamma(X_{n_r}^k - \eta_k)] + \|\gamma\|_\infty \|\eta_k - \mu\|_1.$$

Taking lim sup on both sides as $n \to \infty$ the second term on the right-hand side vanishes. Then taking $k \to \infty$ lets the first term on the right-hand side vanish because of uniform integrability of the sequence X. It also lets the last term disappear: $\forall\, \varepsilon > 0$ exists an integer $N(\varepsilon)$ such that $\|X_n - X_n^k\|_1$ and $\|\xi - \eta_k\|_1$ are both smaller than ε for $k \geq N$ and thus $\mu = \xi$, which leads to the assertion.

□

References

[1] Alsmeyer, Gerold: *Verzweigungsprozesse*, WWU Münster, 2007

[2] Billingsley, Patrick: *Ergodic Theory and Information*, John Wiley & Sons, 1965

[3] Billingsley, Patrick: *Probability and Measure*, John Wiley & Sons, 1986

[4] Blümlinger, Martin; Woracek, Harald; Kaltenbäck, Michael: *Funktionalanalysis*, Technical University of Vienna, 2006

[5] Bru, Bernard: *Jessen's Theorem and Levy's Lemma*, Electronic Journal for History of Probability and Statistics, Vol. 5, No. 1, 2009

[6] Chatterji, S.D.: *Martingales of Banach-Valued Random Variables*, Bulletin of the American Mathematical Society, Vol. 56, No. 5, 1960

[7] Chatterji, S.D.: *A Note on the Convergence of Banach-Space Valued Martingales*, Math. Annalen 153, 1964

[8] Chatterji, S.D.: *Symposium on Probability Methods*, SpringerVerlag, 1967

[9] DiBenedetto, Emmanuele: *Real Analysis*, Birkhäuser, 2002

[10] Doob, Joseph L.: *Regularity Properties of Certain Families of Chance Variables*, Transactions of the American Mathematical Society, Vol. 47, No. 3, 1940

[11] Doob, Joseph L.: *Stochastic Processes*, John Wiley & Sons, 1953

[12] Dunford, Nelson; Schwartz, Jacob T.: *Linear Operators I*, John Wiley & Sons, 1988

[13] Elstrodt, Jürgen: *Maß- und Integrationstheorie*, SpringerVerlag, 2009

[14] Halmos, Paul R.: *Measure Theory*, Van Nostrand, 1950

[15] Helms, Lester L.: *Mean Convergence of Martingales*, Transactions of the American Mathematical Society Vol. 87, No. 2, 1956

[16] Hesse, Christian: *Angewandte Wahrscheinlichkeitstheorie*, Vieweg, 2003

[17] Heuser, Harro: *Lehrbuch der Analysis - Teil 2*, Teubner, 2004

[18] Hewitt, Edwin; Stromberg, Karl: *Real and Abstract Analysis*, SpringerVerlag, 1965

[19] Höpfner, Reinhard: *Vorlesung Stochastik II*, University of Mainz, 2006

[20] Jessen, Borge: *The Theory of Integration in a Space of an Infinite Number of Dimensions*, Acta Math. 63, 1934

[21] Jessen, Borge; Andersen, Erik S.: *Some Limit Theorems on Integral on an Abstract Set*, Det Kongelige Danske Videnskabernes Selskab, Matematisk-fysiske Meddelelser 14, 1946

[22] Jessen, Borge; Andersen, Erik S.: *Some Limit Theorems on Set-Functions*, Det Kongelige Danske Videnskabernes Selskab, Matematisk-fysiske Meddelelser 5, 1948

[23] Kallenberg, Olav: *Foundations of Modern Probability*, SpringerVerlag, 1997

[24] Klenke, Achim: *Wahrscheinlichkeitstheorie*, SpringerVerlag, 2008

[25] Kolmogorov, Andrej N.: *Foundations of the Theory of Probability*, Chelsea Publishing Company, 1956

[26] Krieger, Henry: *Martingales*, Cambrige University Press, 2008

[27] Lévy, Paul P.: *Proprits asymptotiques des sommes de variables alatoires enchanes*, Sci. Math., Vol. 59, No. 2, 1935

[28] Mansuy, Roger: *The Origins of the Word 'Martingale'*, Electronic Journal for History of Probability and Statistics, Vol. 5, No. 1, 2009

[29] Meyer, Paul-André: *Probability and Potentials*, Blaisdell Publishing Company, 1966

[30] Meyer, Paul-André: *Martingales and Stochastic Integrals I*, SpringerVerlag, 1970

[31] Moy, Shu-Teh Chen: *Characterizations of conditional expectation as a transformation on function spaces*, Pacific J. Math., Vol. 4, No. 1, 1954

[32] Scalora, Frank S.: *Abstract Martingale Convergence Theorems*, Pacific Journal of Mathematics, Vol. 11, No. 1, 1961

[33] Schilling, Rene L.: *Measure, Integrals and Martingales*, Cambrige University Press, 2005

[34] Snell, Laurie: *A Conversation with Joe Doob*, Stat. Sci. Volume 24, No. 1, 1997

[35] Spreij, Peter: *Measure Theoretic Probability*, University of Amsterdam, 2005

[36] Ionescu Tulcea, Alexandra; Ionescu Tulcea, Cassius: *Abstract Ergodic Theorems*, Transactions of the American Mathematical Society, Vol. 107, No. 1, 1963

[37] Ville, Jean: *Etude Critique de la Notion de Collectif*, Monographies des Probabilits, No. 3, 1939