A SHORT PROOF OF THE DOOB-MEYER THEOREM

MATHIAS BEIGLBÖCK, WALTER SCHACHERMAYER, BEZIRGEN VELIYEV

Abstract. Every submartingale S of class D has a unique Doob-Meyer decomposition $S = M + A$, where M is a martingale and A is a predictable increasing process starting at 0.

We provide a short and elementary prove of the Doob-Meyer decomposition theorem. Several previously known arguments are included to keep the paper self-contained.

MSC2010: 60G05, 60G07.
keywords: Doob-Meyer decomposition, Komlos Lemma.

1. Introduction

Throughout this article we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a right-continuous complete filtration $(\mathcal{F}_t)_{0\le t\le T}$.

An adapted process $(S_t)_{0\le t\le T}$ is of class D if the family of random variables S_τ where τ ranges through all stopping times is uniformly integrable ([Mey62]).

The purpose of this paper is to give a short and elementary proof of the following

Theorem 1.1 (Doob-Meyer)**.** *Let $S = (S_t)_{0\le t\le T}$ be a càdlàg submartingale of class D. Then, S can be written in a unique way in the form*

$$S = M + A \tag{1}$$

where M is a martingale and A is a predictable increasing process starting at 0.

Doob [Doo53] noticed that in discrete time an integrable process $S = (S_n)_{n=1}^\infty$ can be uniquely represented as the sum of a martingale M and a predictable process A starting at 0; in addition, the process A is increasing iff S is a submartingale. The continuous time analogue, Theorem 1.1, goes back to Meyer [Mey62, Mey63], who introduced the class D and proved that every submartingale $S = (S_t)_{0\le t\le T}$ can be decomposed in the form (1), where M is a martingale and A is a *natural* process. The modern formulation is due to Doléans-Dade [DD67, DD68] who obtained that an increasing process is natural iff it is predictable. Further proofs of Theorem 1.1 were given by Rao [Rao69], Bass [Bas96] and Jakubowski [Jak05].

Rao works with the $\sigma(L^1, L^\infty)$-topology and applies the Dunford-Pettis compactness criterion to obtain the desired continuous time decomposition as a weak-L^1 limit from discrete approximations. To obtain that A is predictable one then invokes the theorem of Doléans-Dade.

Bass gives a more elementary proof based on the dichotomy between predictable and totally inaccessible stopping times.

Jakubowski proceeds as Rao, but notices that predictablity of the process A can also be obtained through an application of Komlos' Lemma [Kom67].

The proof presented subsequently combines ideas from [Jak05] and [BSV10] to construct the continuous time decomposition using a suitable Komlos-type lemma.

All authors affiliated at
Faculty of Mathematics, Vienna University, Nordbergstr. 15, 1090 Wien, Austria.
Corresponding author
mathias.beiglboeck@univie.ac.at, phone: +43 1 4277 50724. fax: +43 1 4277 50727.

2. Proof of Theorem 1.1

The proof of uniqueness is standard and we have nothing to add here; see for instance [Kal02, Lemma 25.11].

For the remainder of this article we work under the assumptions of Theorem 1.1 and fix $T = 1$ for simplicity.

Denote by $\mathcal{D}_n$ and $\mathcal{D}$ the set of n-th resp. all dyadic numbers $j/2^n$ in the interval $[0, 1]$. For each n, we consider the discrete time Doob decomposition of the sampled process $S^n = (S_t)_{t\in\mathcal{D}_n}$, that is, we define A^n, M^n by $A^n_0 := 0$,

$$A^n_t - A^n_{t-1/2^n} := \mathbb{E}[S_t - S_{t-1/2^n}|\mathcal{F}_{t-1/2^n}] \text{ and} \tag{2}$$

$$M^n_t := S_t - A^n_t \tag{3}$$

so that $(M^n_t)_{t\in\mathcal{D}_n}$ is a martingale and $(A^n_t)_{t\in\mathcal{D}_n}$ is predictable with respect to $(\mathcal{F}_t)_{t\in\mathcal{D}_n}$.

The idea of the proof is, of course, to obtain the continuous time decomposition (1) as a limit, or rather, as an accumulation point of the processes $M^n, A^n, n \geq 1$.

Clearly, in infinite dimensional spaces a (bounded) sequence need not have a convergent subsequence. As a substitute for the Bolzano-Weierstrass Theorem we establish the Komlos-type Lemma 2.1 in Section 2.1.

In order to apply this auxiliary result, we require that the sequence $(M^n_1)_{n\geq1}$ is uniformly integrable. This follows from the class D assumption as shown by [Rao69]. To keep the paper self-contained, we provide a proof in Section 2.2.

Finally, in Section 2.3, we obtain the desired decomposition by passing to a limit of the discrete time versions. As the Komlos-approach guarantees convergence in a strong sense, predictability of the process A follows rather directly from the predictability of the approximating processes. This idea is taken from [Jak05].

2.1. Komlos' Lemma.

Following Komlos [Kom67][1], it is sometimes possible to obtain an accumulation point of a bounded sequence in an infinite dimensional space if appropriate convex combinations are taken into account.

A particularly simple result of this kind holds true if $(f_n)_{n\geq1}$ is a bounded sequence in a Hilbert space. In this case

$$A = \sup_{n\geq1} \inf\{\|g\|_2 : g \in \operatorname{conv}\{f_n, f_{n+1}, \ldots\}\}$$

is finite and for each n we may pick some $g_n \in \operatorname{conv}\{f_n, f_{n+1}, \ldots\}$ such that $\|g_n\|_2 \leq A+1/n$. If n is sufficiently large with respect to $\varepsilon > 0$, then $\|(g_k+g_m)/2\|_2 > A-\varepsilon$ for all $m, k \geq n$ and hence

$$\|g_k - g_m\|_2^2 = 2\|g_k\|_2^2 + 2\|g_m\|_2^2 - \|g_k + g_m\|_2^2 \leq 4(A + \tfrac{1}{n})^2 - 4(A - \varepsilon)^2.$$

By completeness, $(g_n)_{n\geq1}$ converges in $\|.\|_2$.

By a straight forward truncation procedure this Hilbertian Komlos-Lemma yields an L^1-version which we will need subsequently.[2]

Lemma 2.1. *Let $(f_n)_{n\geq1}$ be a uniformly integrable sequence of functions on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exist functions $g_n \in \operatorname{conv}(f_n, f_{n+1}, \ldots)$ such that $(g_n)_{n\geq1}$ converges in $\|.\|_{L^1(\Omega)}$.*

Proof. For $i, n \in \mathbb{N}$ set $f^{(i)}_n := f_n \mathbb{1}_{\{|f_n|\leq i\}}$ such that $f^{(i)}_n \in L^2(\Omega)$.

We claim that there exist for every n convex weights $\lambda^n_n, \ldots, \lambda^n_{N_n}$ such that the functions $\lambda^n_n f^{(i)}_n + \ldots + \lambda^n_{N_n} f^{(i)}_{N_n}$ converge in $L^2(\Omega)$ for every $i \in \mathbb{N}$.

[1] Indeed, [Kom67] considers Cesaro sums along subsequences rather then arbitrary convex combinations. But for our purposes, the more modest conclusion of Lemma 2.1 is sufficient.

[2] Lemma 2.1 is also a trivial consequence of Komlos' original result [Kom67] or other related results that have been established through the years. Cf. [KS09, Chapter 5.2] for an overview.

To see this, one first uses the Hilbertian lemma to find convex weights $\lambda_n^n, \dots, \lambda_{N_n}^n$ such that $(\lambda_n^n f_n^{(1)} + \dots + \lambda_{N_n}^n f_{N_n}^{(1)})_{n\geq 1}$ converges. In the second step, one applies the lemma to the sequence $(\lambda_n^n f_n^{(2)} + \dots + \lambda_{N_n}^n f_{N_n}^{(2)})_{n\geq 1}$, to obtain convex weights which work for the first two sequences. Repeating this procedure inductively we obtain sequences of convex weights which work for the first m sequences. Then a standard diagonalization argument yields the claim.

By uniform integrability, $\lim_{i\to\infty} \|f_n^{(i)} - f_n\|_1 = 0$, uniformly with respect to n. Hence, once again, uniformly with respect to n,

$$\lim_{i\to\infty} \|(\lambda_n^n f_n^{(i)} + \dots + \lambda_{N_n}^n f_{N_n}^{(i)}) - (\lambda_n^n f_n + \dots + \lambda_{N_n}^n f_{N_n})\|_1 = 0.$$

Thus $(\lambda_n^n f_n + \dots + \lambda_{N_n}^n f_{N_n})_{n\geq 1}$ is a Cauchy sequence in $L^1(\Omega)$. □

2.2. **Uniform integrability of the discrete approximations.**

Lemma 2.2. *The sequence $(M_1^n)_{n\geq 1}$ is uniformly integrable.*

Proof. Subtracting $\mathbb{E}[S_1|\mathcal{F}_t]$ from S_t we may assume that $S_1 = 0$ and $S_t \leq 0$ for all $0 \leq t \leq 1$. Then $M_1^n = -A_1^n$, and for every $(\mathcal{F}_t)_{t\in\mathcal{D}_n}$-stopping time τ

$$S_\tau^n = -\mathbb{E}[A_1^n|\mathcal{F}_\tau] + A_\tau^n. \tag{4}$$

We claim that $(A_1^n)_{n=1}^\infty$ is uniformly integrable. For $c > 0$, $n \geq 1$ define

$$\tau_n(c) = \inf\big\{(j-1)/2^n : A_{j/2^n}^n > c\big\} \wedge 1.$$

From $A_{\tau_n(c)}^n \leq c$ and (4) we obtain $S_{\tau_n(c)} \leq -E[A_1^n|\mathcal{F}_{\tau_n(c)}] + c$. Thus,

$$\int_{\{A_1^n>c\}} A_1^n \, d\mathbb{P} = \int_{\{\tau_n(c)<1\}} \mathbb{E}[A_1^n|\mathcal{F}_{\tau_n(c)}] \, d\mathbb{P} \leq c\,\mathbb{P}\big[\tau_n(c) < 1\big] - \int_{\{\tau_n(c)<1\}} S_{\tau_n(c)} \, d\mathbb{P}.$$

Note $\{\tau_n(c) < 1\} \subseteq \{\tau_n(\frac{c}{2}) < 1\}$, hence, by (4)

$$\begin{aligned}\int_{\{\tau_n(\frac{c}{2})<1\}} -S_{\tau_n(\frac{c}{2})} \, d\mathbb{P} &= \int_{\{\tau_n(\frac{c}{2})<1\}} A_1^n - A_{\tau_n(\frac{c}{2})}^n \, d\mathbb{P} \\ &\geq \int_{\{\tau_n(c)<1\}} A_1^n - A_{\tau_n(\frac{c}{2})}^n \, d\mathbb{P} \geq \frac{c}{2}\,\mathbb{P}[\tau_n(c) < 1].\end{aligned}$$

Combining the above inequalities we obtain

$$\int_{\{A_1^n>c\}} A_1^n \, d\mathbb{P} \leq -2 \int_{\{\tau_n(\frac{c}{2})<1\}} S_{\tau_n(\frac{c}{2})} \, d\mathbb{P} - \int_{\{\tau_n(c)<1\}} S_{\tau_n(c)} \, d\mathbb{P}. \tag{5}$$

On the other hand

$$\mathbb{P}[\tau_n(c) < 1] = \mathbb{P}[A_1^n > c] \leq \mathbb{E}[A_1^n]/c = -\mathbb{E}[M_1^n]/c = -\mathbb{E}[S_0]/c,$$

hence, as $c \to \infty$, $\mathbb{P}[\tau_n(c) < 1]$ goes to 0, uniformly in n. As S is of class D, (5) implies that the sequence $(A_1^n)_{n\geq 1}$ is uniformly integrable and hence $(M_1^n)_{n\geq 1} = (S_1 - A_1^n)_{n\geq 1}$ is uniformly integrable as well. □

2.3. **The limiting procedure.** For each n, extend M^n to a (càdlàg) martingale on $[0,1]$ by setting $M_t^n := \mathbb{E}[M_1^n|\mathcal{F}_t]$. By Lemma 2.1 and Lemma 2.2 there exist $M \in L^1(\Omega)$ and for each n convex weights $\lambda_n^n, \dots, \lambda_{N_n}^n$ such that with

$$\mathcal{M}^n := \lambda_n^n M^n + \dots + \lambda_{N_n}^n M^{N_n} \tag{6}$$

we have $\mathcal{M}_1^n \to M$ in $L^1(\Omega)$. Then, by Jensen's inequality, $\mathcal{M}_t^n \to M_t := \mathbb{E}[M|\mathcal{F}_t]$ for all $t \in [0,1]$. For each $n \geq 1$ we extend A^n to $[0,1]$ by

$$A^n := \textstyle\sum_{t\in\mathcal{D}_n} A_t^n \mathbb{1}_{(t-1/2^n,t]} \tag{7}$$

$$\text{and set} \quad \mathcal{A}^n := \lambda_n^n A^n + \dots + \lambda_{N_n}^n A^{N_n}, \tag{8}$$

where we use the same convex weights as in (6). Then the càdlàg process

$$(A_t)_{0\le t\le 1} := (S_t)_{0\le t\le 1} - (M_t)_{0\le t\le 1}$$

satisfies for every $t \in \mathcal{D}$

$$\mathcal{A}^n_t = (S_t - \mathcal{M}^n_t) \;\to\; (S_t - M_t) = A_t \quad \text{in } L^1(\Omega).$$

Passing to a subsequence which we denote again by n, we obtain that convergence holds also almost surely. Consequently, A is almost surely increasing on $\mathcal{D}$ and, by right continuity, also on $[0,1]$.

As the processes A^n and $\mathcal{A}^n$ are left-continuous and adapted, they are predictable. To obtain that A is predictable, we show that for a.e. ω and every $t \in [0,1]$

$$\limsup_n \mathcal{A}^n_t(\omega) = A_t(\omega). \tag{9}$$

If $f_n, f : [0,1] \to \mathbb{R}$ are increasing functions such that f is right continuous and $\lim_n f_n(t) = f(t)$ for $t \in \mathcal{D}$, then

$$\limsup_n f_n(t) \le f(t) \text{ for all } t \in [0,1] \text{ and} \tag{10}$$

$$\lim_n f_n(t) = f(t) \text{ if } f \text{ is continuous at } t. \tag{11}$$

Consequently, (9) can only be violated at discontinuity points of A. As A is càdlàg, every path of A can have only finitely many jumps larger than $1/k$ for $k \in \mathbb{N}$. It follows that the points of discontinuity of A can be exhausted by a countable sequence of stopping times, and therefore it is sufficient to prove $\limsup_n \mathcal{A}^n_\tau = A_\tau$ for every stopping time τ.

By (10), $\limsup_n \mathcal{A}^n_\tau \le A_\tau$ and as $\mathcal{A}^n_\tau \le \mathcal{A}^n_1 \to A_1$ in $L^1(\Omega)$ we deduce from Fatou's Lemma that

$$\liminf_n \mathbb{E}\big[A^n_\tau\big] \le \limsup_n \mathbb{E}\big[\mathcal{A}^n_\tau\big] \le \mathbb{E}\big[\limsup_n \mathcal{A}^n_\tau\big] \le \mathbb{E}\big[A_\tau\big].$$

Therefore it suffices to prove $\lim_n \mathbb{E}[A^n_\tau] = \mathbb{E}[A_\tau]$. For $n \ge 1$ set

$$\sigma_n := \inf\{t \in \mathcal{D}_n : t \ge \tau\}.$$

Then $A^n_\tau = A^n_{\sigma_n}$ and $\sigma_n \downarrow \tau$. Using that S is of class D, we obtain

$$\mathbb{E}[A^n_\tau] = \mathbb{E}[A^n_{\sigma_n}] = \mathbb{E}[S_{\sigma_n}] - \mathbb{E}[M_0] \to \mathbb{E}[S_\tau] - \mathbb{E}[M_0] = \mathbb{E}[A_\tau].$$

References

[Bas96] R. F. Bass. The Doob-Meyer decomposition revisited. *Canad. Math. Bull.*, 39(2):138–150, 1996.

[BSV10] M. Beiglböck, W. Schachermayer, and B. Veliyev. A direct proof of the Bichteler-Dellacherie theorem and connections to arbitrage. *Ann. Probab.*, 2010. to appear.

[DD67] C. Doléans-Dade. Processus croissants naturels et processus croissants très-bien-mesurables. *C. R. Acad. Sci. Paris Sér. A-B*, 264:A874–A876, 1967.

[DD68] C. Doléans-Dade. Existence du processus croissant natural associé à un potentiel de la classe (D). *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 9:309–314, 1968.

[Doo53] J. L. Doob. *Stochastic processes*. John Wiley & Sons Inc., New York, 1953.

[Jak05] A. Jakubowski. An almost sure approximation for the predictable process in the Doob-Meyer decomposition theorem. In *Séminaire de Probabilités XXXVIII*, volume 1857 of *Lecture Notes in Math.*, pages 158–164. Springer, Berlin, 2005.

[Kal02] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kom67] J. Komlós. A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.*, 18:217–229, 1967.

[KS09] Y. Kabanov and M. Safarian. *Markets with transaction costs*. Springer Finance. Springer-Verlag, Berlin, 2009. Mathematical theory.

[Mey62] P.-A. Meyer. A decomposition theorem for supermartingales. *Illinois J. Math.*, 6:193–205, 1962.

[Mey63] P.-A. Meyer. Decomposition of supermartingales: the uniqueness theorem. *Illinois J. Math.*, 7:1–17, 1963.

[Rao69] K. M. Rao. On decomposition theorems of Meyer. *Math. Scand.*, 24:66–78, 1969.