M800 Roger Temam 4/29 Report

1 Types of Random Variable Convergence and Skorohod's Theorem

Let $\{X_n\}$ and X be r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a metric space (E, d).

Definition 1.1.

(i) $X_n \longrightarrow X$ **almost surely in $\mathbb{P}$ ($\mathbb{P}$-a.s.)** (also written as $X_n \xrightarrow{\mathbb{P}\text{-a.s.}} X$ or $X_n \xrightarrow{\text{a.s.}} X$ if context is known) if $\mathbb{P}\left[\lim_{n \to +\infty} X_n = X\right] = 1$.

(ii) $X_n \to X$ **in probability** (also written as $X_n \xrightarrow{\mathbb{P}} X$) if $\forall \epsilon > 0$ we have,

$$\lim_{n \to +\infty} \mathbb{P}[d(X_n, X) > \epsilon] = 0.$$

(iii) Let $\{\mu_n\}_{n=1}^{\infty}$ and μ be Borel probability measures (defined on all open sets of E) on a metric space (E, d). We say that $\{\mu_n\}_{n=1}^{\infty}$ **converges weakly to μ** and we write $\mu_n \Rightarrow \mu$ if for every bounded continuous function $f : E \to \mathbb{R}$ we have,

$$\int_E f d\mu_n \to \int_E f d\mu \text{ as } n \to +\infty.$$

For E-valued r.v.'s $\{X_n\}_{n=1}^{\infty}, X$, we say X_n **converges to X in distribution** if $\mu_{X_n} \Rightarrow \mu_X$, and write $X_n \Rightarrow X$.
Source: M647 Lecture 1 (revised) Definition 1.4.1, Definition 1.4.3.

Proposition 1.2. Let $\{X_n\}_{n=1}^{\infty}, X$ be E-valued r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$:

(i) If $X_n \to X$ $\mathbb{P}$-a.s. then $X_n \to X$ in probability.

(ii) If $X_n \to X$ in probability then $X_n \Rightarrow X$.
Source: M647 Lecture 1 (revised) Proposition 1.4.2, Propostion 1.4.4.

Proof.

(i) If $X_n \xrightarrow{\text{a.s.}} X$, observe that given $\epsilon > 0$, since

$$\lim_{n \to +\infty} X_n = X \Longrightarrow d(X_n, X) \leq \epsilon \text{ e.v.},$$

we find that

$$\begin{aligned}\lim_n \mathbb{P}[d(X_n, X) \leq \epsilon] \geq \liminf_n \mathbb{P}[d(X_n, X) \leq \epsilon] &= \mathbb{P}\left[\bigcup_{N\geq 1}\bigcap_{n\geq N}[d(X_n, X) \leq \epsilon]\right] \\ &= \mathbb{P}\Big[d(X_n, X) \leq \epsilon \text{ e.v.}\Big] \geq \mathbb{P}\Big[\lim_{n \to +\infty} X_n = X\Big] = 1.\end{aligned}$$

$$\Longrightarrow \lim_n \mathbb{P}[d(X_n, X) \leq \epsilon] = 1 \Longrightarrow \lim_{n \to +\infty} \mathbb{P}[d(X_n, X) > \epsilon] = 0.$$

(ii) Suppose $X_n \xrightarrow{\mathbb{P}} X$ and let $f \in C_b(E)$. First, note that by continuity of f that $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. Next, let $\epsilon > 0$, and choose $N \geq 1$ such that

$$n \geq N \Longrightarrow \mathbb{P}\left[|f(X_n) - f(X)| > \frac{\epsilon}{2}\right] \leq \frac{\epsilon}{2M},$$

where $M := \sup_{x\in E}|f(x)|$. Then

$$\begin{aligned}n \geq N \Longrightarrow \left|\int_E f d\mu_{X_n} - \int_E f d\mu_X\right| &= \left|\int_E f(X_n) - f(X) d\mathbb{P}\right| \leq \int_\Omega |f(X_n) - f(X)| d\mathbb{P} \\ &= \int_{|f(X_n)-f(X)|>\frac{\epsilon}{2}} |f(X_n) - f(X)| d\mathbb{P} + \int_{|f(X_n)-f(X)|\leq\frac{\epsilon}{2}} |f(X_n) - f(X)| d\mathbb{P} \\ &\leq \int_{|f(X_n)-f(X|)>\frac{\epsilon}{2}} M d\mathbb{P} + \int_{|f(X_n)-f(X)|\leq\frac{\epsilon}{2}} \frac{\epsilon}{2} d\mathbb{P} \\ &= M \cdot \mathbb{P}\left[|f(X_n) - f(X)| > \frac{\epsilon}{2}\right] + \frac{\epsilon}{2} \cdot \mathbb{P}\left[|f(X_n) - f(X)| \leq \frac{\epsilon}{2}\right] \\ &\leq M \cdot \frac{\epsilon}{2M} + \frac{\epsilon}{2} \cdot 1 = \epsilon. \quad \square\end{aligned}$$

Note that the converses of $X_n \Rightarrow X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} X \Longrightarrow X_n \xrightarrow{\text{a.s.}} X$ in general do not hold, though we'll talk about some partial converses in the next proposition (more discussion of this to come in a future draft).

Proposition 1.3.

(i) if $X_n \to X$ in probability, then there exists a subsequence $\{X_{n_k}\}_{k=1}^{\infty}$ such that $X_{n_k} \to X$ $\mathbb{P}$-a.s.

(ii) $X_n \to X$ in probability if and only if for every subsequence $\{X_{n_k}\}_{k=1}^{\infty}$, there exists a further subsequence $\{X_{n_{k_j}}\}_{k=1}^{\infty}$ such that $X_{n_{k_j}} \xrightarrow{\text{a.s.}} X$.
Source: M647 Lecture 1 (revised) Proposition 1.4.2.

Proof.

(i) Suppose $X_n \xrightarrow{\mathbb{P}} X$. Choose a subsequence $\{X_{n_k}\}$ such that

$$\mathbb{P}\left[d(X_{n_k}, X) > 2^{-k}\right] \leq 2^{-k}.$$

Since

$$\mathbb{P}\left[\lim_{n \to +\infty} X_n = X\right] = \inf_{\epsilon>0} \mathbb{P}\left[d(X_{n_k}, X) \leq \epsilon \text{ e.v.}\right] = 1 - \sup_{\epsilon>0} \mathbb{P}\left[d(X_{n_k}, X) > \epsilon \text{ i.o.}\right].$$

It shall suffice to show that given $\epsilon > 0$, we have

$$\mathbb{P}\left[d(X_{n_k}, X) > \epsilon \text{ i.o.}\right] = 0. \qquad (1.1)$$

Noting that $2^{-k_0} \leq \epsilon$, for some $k_0 \geq 1$, we find that eventually we have

$$\mathbb{P}[d(X_{n_k}, X) > \epsilon] \leq \mathbb{P}\left[d(X_{n_k}, X) > 2^{-k}\right] \leq 2^{-k},$$

hence

$$\sum_{k=1}^{\infty} \mathbb{P}[d(X_{n_k}, X) > \epsilon] < +\infty,$$

and (1.1) immediately follows by the *First Borel Cantelli-Lemma*. (refer to *Billingsley, page 59*) *(ii)* $\Longrightarrow$ Suppose $X_n \xrightarrow{\mathbb{P}} X$. For every subsequence $\{X_{n_k}\}_{k=1}^{\infty}$, the existence of a further subsequence $\{X_{n_{k_j}}\}_{k=1}^{\infty}$ such that $X_{n_{k_j}} \xrightarrow{\text{a.s.}} X$. follows immediately by part *(ii)*, since $X_{n_k} \xrightarrow{\mathbb{P}} X$.

$\Longleftarrow$ Conversely, suppose for every subsequence $\{X_{n_k}\}_{k=1}^{\infty}$, there exists a further subsequence $\{X_{n_{k_j}}\}_{k=1}^{\infty}$ such that $X_{n_{k_j}} \xrightarrow{\text{a.s.}} X$. Let $\epsilon > 0$. By part *(i)*, we find that every subsequence $\{\mathbb{P}[d(X_{n_k}, X) \geq \epsilon]\}_{k \in \mathbb{N}}$ of $\{\mathbb{P}[d(X_n, X)]\}_{n \in \mathbb{N}}$ has a further subsequence $\{\mathbb{P}[d(X_{n_{k_j}}, X) \geq \epsilon]\}_{j \in \mathbb{N}}$ such that $\mathbb{P}[d(X_{n_{k_j}}, X) \geq \epsilon] \xrightarrow{j \to +\infty} 0$. Then our conclusion follows by the fact that $\{\mathbb{P}[d(X_n, X) \geq \epsilon]\}_{n \in \mathbb{N}}$ is a bounded sequence since nonconvergence of 0 implies existence of a subsequence $\{\mathbb{P}[d(X_{n_k}, X) \geq \epsilon]\}_{k \in \mathbb{N}}$ such that $\mathbb{P}[d(X_{n_k}, X) \geq \epsilon] \geq \eta$ for some $\eta > 0$, contradicting existence of a further subsequence $\{\mathbb{P}[d(X_{n_{k_j}}, X) \geq \epsilon]\}_{j \in \mathbb{N}}$ such that $\mathbb{P}[d(X_{n_{k_j}}, X) \geq \epsilon] \xrightarrow{j \to +\infty} 0$. $\square$

Proposition 1.4. Let $f : E \to E'$ be a continuous function and $\{X_n\}$ be a sequence of E-random variables such that $X_n \xrightarrow{\mathbb{P}} X$, for some random variable . Then $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.
Source: Billingsley, page 268

Note: The proposition is in the more general setting of metric-spaced valued random variables, whereas the analogous statement of this proposition metioned in Billingsley is in the real-valued random variable setting. However, the proof (which we shall provide in full detail below) is pretty much the same as in this case.

Proof. We shall use **Proposition 1.3** *(ii)* to prove our results. Let $\{f(X_{n_j})\}_{j \in \mathbb{N}}$ be a subsequence of $\{f(X_n)\}_{n \in \mathbb{N}}$. Choose a further subsequence $\{X_{n_{j_k}}\}_{k \in \mathbb{N}}$ of $\{X_{n_j}\}_{j \in \mathbb{N}}$ such that $X_{n_{j_k}} \xrightarrow{\text{a.s.}} X$ as $k \to +\infty$. Since

$$\lim_{k \to +\infty} X_{n_{j_k}}(\omega) = X(\omega) \Longrightarrow \lim_{k \to +\infty} f(X_{n_{j_k}}(\omega)) = f(X(\omega)),$$

we find that $\{f(X_{n_{j_k}})\}_{k \in \mathbb{N}}$ is a subsequence of $\{f(X_{n_j})\}_{j \in \mathbb{N}}$ such that $f(X_{n_{j_k}}) \xrightarrow{\text{a.s.}} f(X)$ as $k \to +\infty$. $\square$

Definition 1.5.

(i) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say events $A_1, A_2, \ldots, A_m \in \mathcal{F}$ are

independent if

$$\mathbb{P}\left[\bigcap_{j\in I} A_j\right] = \prod_{j\in I} \mathbb{P}[A_j]\ \forall I \subset \{1,2,\ldots,m\}$$

(ii) A family $\{\mathcal{F}_i\}_{i\in I}$ of subfields of $\mathcal{F}$ is **independent** if for every finite $\{i_1,\ldots,i_m\} \subset I$ and every choice of $A_{i_k} \in \mathcal{F}_{i_k}, k = 1,\ldots,m$, we have that

$$\mathbb{P}\left[\bigcap_{k=1}^{m} A_{i_k}\right] = \prod_{k=1}^{m} \mathbb{P}[A_{i_k}].$$

(iii) A family $\{X_i\}$ of r.v.'s on $(\Omega,\mathcal{F},\mathbb{P})$ is **independent** if $\{\sigma(X_i)\}_{i\in I}$ is independent (in the sense of the definition above).
Source: M647 Lecture 1 (revised) Definition 1.3.1-1.3.3.

Definition 1.6. The intersection

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1},\ldots)$$

is the tail σ-field associated with a sequence of random variables $X_1, X_2,\ldots$ its elements are **tail events**.
Source: Billingsley, page 287

Here is the important Kolmogorov's zero-one law:

Theorem 1.7. *(Kolmogorov's zero-one law)* Suppose that $\{X_n\}_{n\in\mathbb{N}}$ are independent random variables and A is a tail event of the sequence, i.e.

$$A \in \mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1},\ldots).$$

Then $\mathbb{P}[A] = 0$ or $\mathbb{P}[A] = 1$.
Source: Billingsley, Theorem 22.3 (page 287)

Remark 1.1. Note that **Definition 1.6** and **Theorem 1.7** are given in more general form than in Billingsley in the sense that $\{X_n\}_{n\in\mathbb{N}}$ is a metric space-valued random variable and not necessarily a real-valued random variable. It's worth noting that the proof of Theorem 1.7 is pretty much the same proof, regardless of the increased generality.

Proof. It shall suffice to prove that any $A \in \mathcal{T}$ is independent with itself, since the equation

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2,$$

has the solution $\mathbb{P}[A] = 0, 1$. It shall suffice to show that $\mathcal{T}$ is independent from $\sigma(X_n, X_{n+1}, \dots)$, for all $n \geq 1$. Note that $\sigma(X_n, X_{n+1}, \dots)$ is generated by the π-system

$$\mathcal{P}_n := \left\{ \bigcap_{j=1}^{k} X_{n_j}^{-1}(B_{n_j}) : n \leq n_1 < \dots < n_k \text{ and } B_{n_j} \in \mathcal{B}(E),\ 1 \leq j \leq r \right\},$$

and for any $\bigcap_{j=1}^{k} X_{n_j}^{-1}(B_{n_j}) \in \mathcal{P}_n$, we find for $m := \max(n_1, \dots, n_j)$ we have $A \in \sigma(X_{m+1}, X_{m+2}, \dots)$--since $\mathcal{T} \subset \sigma(X_{m+1}, X_{m+2}, \dots)$--which is independent from $\sigma(X_1, \dots, X_m)$ containing $\bigcap_{j=1}^{k} X_{n_j}^{-1}(B_{n_j})$. $\square$

Definition 1.8. A subset $\mathcal{M}$ of probability measures on $\mathcal{S}$ is said to be

(i) **tight**, if for each $\epsilon > 0$ there exists a compact set K such that

$$\sup_{\mu \in \mathcal{M}} \mu\left(K^c\right) \leq \epsilon.$$

(ii) **relatively (sequentially) weakly compact** if any sequence $\{\mu_n\}_{n\geq 1}$ in $\mathcal{M}$ admits a weakly convergent subsequence $\{\mu_{n_k}\}_{k\geq 1}$.

Theorem 1.9. *(Prohorov)* Suppose that the metric space (E, d) is complete and separable, and let $\mathcal{M}$ be a set of probability measures on $(E, \mathcal{B}(d))$, where $\mathcal{B}(d)$ is the Borel σ-algebra on the metric space (E, d). Then $\mathcal{M}$ is relatively weakly compact if and only if it is tight. *Žitkovic, Theorem 7.14 (page 94)*

The proof will be provided in the next draft.

Theorem 1.10. *(Skorohod Representation Theorem)* Let E be a complete and separable metric space. Let $\{X_n\}_{n=1}^{\infty}$ and X be E-valued r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \Rightarrow X$. Then there exists a probability space $\left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\right)$ and r.v.'s $\left\{\widetilde{X}_n\right\}_{n=1}^{\infty}$ and $\widetilde{X}$ on $\left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\right)$ such that,

(i) $\widetilde{X}_n \overset{\mathfrak{D}}{=} X_n$
(ii) $\widetilde{X} \overset{\mathfrak{D}}{=} X$
(iii) $\widetilde{X}_n \to \widetilde{X}$ $\widetilde{\mathbb{P}}$-a.s.
Source: M647 Lecture 1 (revised) Theorem 1.4.5.

Remark 1.2.

(i) Note that in the general metric space setting, we alas do not have access to cumumulative distribution functions (CDF's), which makes the proof in the real-valued random variable setting rather straightforward. The general strategy of this proof is to do the next best thing to utilizing CDF's, which is to find a probability space $\widetilde{\mathbb{P}}$ where each $\widetilde{X}_n$ are independent. In the situation where $\widetilde{X}_n$ is independent, there's a lot less we have to do, i.e., it suffices to show that $\widetilde{X}_n \xrightarrow{\mathbb{P}} \widetilde{X}$ since the converse of **Proposition 1.2** *(i)*. holds when $\widetilde{X}_n$ is independent (something to be mentioned in a future draft), and moreover we have access to *Kolmogorov's zero-one law* that we proved previously, yet another powerful tool at our disposal.

(iii) This report was made to improve on the proof of the Skorrohod Theorem (*Theorem A.1*) in *Nguyen, Tawri, Temam 2021*. That proof may have an issue with it and there was a citation of another proof of Skorrohod's Thoerem written in the paper by *Z. Brzeźniak, E. Hausenblas, and P. Razafimandimby 2014*. It turns out that the cited theorem of that paper (*Theorem C.1*) is NOT the Skorrohod Theorem, so the citation of that paper is incorrect. (though perhaps the intent was to cite a different source that proves Skorrohod's Theorem)

(iv) As far as how my personal proof of this theorem worked out, I unfortunately found an error in the way I was going about my strategy, and it's kind of a huge setback alas. I may have found a way to remedy my strategy using approximations of simple measures, and then make independent events correspond to finer and finer approximations, allowing us to use the zero-one law as desired, but that is something I will not be able to use until the distant future.

(v) After some browsing on the internet, there has been some work with ranging from 2010 to even as late as 2021 with proving stronger versions (and even necessary and sufficient conditions) of the Skorrohod Theorem. First is the paper by *Lo, Niang, and Okereke* about a new proof of the Skorrohod Thoerem--or what this paper calls the Skorrohod-Winchura Theorem since the widely used proof of the theorem was done by Winchura. The paper notes that the old proof by Winchura is not very accessible to graudate students who learn about the thoerem, which gives us a motivation to coming up with more elegant proofs of Skorrohod Theorem

Next are the papers by a [Pratelli, Rigo 2022] and [Berti, Pratelli, Rigo] which both give (different) necessary and sufficient conditions for the outcome of Skorrohod's Theorem (i.e., the existence of a sequence of random variables with distributions corresponding to the original sequence of measures converging almost surely a rando variable with its distribution as the weak* limit random variable with the distribution as the limit measure of the sequence of measures.

(vi) For now, I am working with the additional assumption that E is a Banach space, since most of the results in this area of research involve Banach spaces, and more generally Hilbert spaces. Also we potentially have better tools at our disposal, since E being a Banach space E means that we can conceive of the limit we are trying to find explicitly as a random E-valued series of $\widetilde{Y}_n := \widetilde{X}_n - \widetilde{X}_{n-1}$ and more easily utilize tools such as Kolmogorov's 0-1 Law--and maybe even the (more general analogue in multidimensional spaces of the) *3 Series Theorem* --to get our result.

Here is a weaker version of the *Skorrohod Thoerem* where the strategy I am trying to utilize ends up sufficing. The goal of my proof of Skorrohod's Theorem would be to take an arbitrary sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of probability measures that converge in distribution and figure out a sample space that would lead to a scenario similar to the case highlighted in **Theorem 1.11**.

Theorem 1.11. Let E be a Banach space. Let $\{X_n\}_{n=1}^{\infty}$ and be independent E-valued r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$, and let μ_{S_n} be the distribution of $S_n := X_1 + \cdots + X_n$. If $\mu_{S_n} \Rightarrow \mu$, then the random series $\sum_{k=1}^{\infty} X_k$ converges almost surely.

Outline of Proof.

$\Longrightarrow$ Suppose $\mu_{S_n} \Rightarrow \mu$ and towards contradiction $\sum_{k=1}^{\infty} X_k$ does not converge almost surely.

First, we prove the claim that $\sum_{k=1}^{\infty} X_k$ converging is a tail event, and it immediately follows by

Theorem 1.7 (i.e., *Kolmogorov's 0-1 Law*) that

$$\mathbb{P}\left[\sum_{k=1}^{\infty} X_k \text{ converges}\right] = 0 \Longrightarrow \sum_{k=1}^{\infty} ||X_k|| = +\infty \text{ almost surely}$$

We find that given $x \in E$ and $\epsilon_x > 0$ sufficiently small, we find

$$\mu(B_E(x,\epsilon_x)) = \lim_{n\to+\infty} \mu_{S_n}(B_E(x,\epsilon_x)) = \lim_{n\to+\infty} \mathbb{P}[S_n \in B_E(x,\epsilon_x)] = \mathbb{P}\left[\sum_{k=1}^{\infty} X_k \in B_E(x,\epsilon_x)\right] = 0,$$

hence

$$\mu = 0 \Longrightarrow \lim_{n\to+\infty} \mu_{S_n}(E) = \mu(E) = 0,$$

which contradicts the fact that $\mu_{S_n}(E) = 1$, for every $n \geq 1$.

$\Longleftarrow$ Immediate by **Proposition 1.2** *(ii)*. □

Sources:

Spring 2022-M647 Lecture 1 (revised)

Probability and Measure, 3rd edition, § 4, § 22
Billingsley

Theory of Probability Lecture Notes, § 7.1
Gordan Žitkovic

Nonlinear Stochastic Parabolic Partial Differential Equations with a Monotone Operator of the Ladyzenskaya-Smagorinsky Type, Driven by a Levy Noise
By Nguyen, Tawri, Temam 2021

Z. Brzeźniak, E. Hausenblas, and P. Razafimandimby. Martingale solutions for stochastic equation of reaction diffusion type driven by Lévy noise or Poisson random measure. ArXiv e-prints, *February 2014*

Pratelli, Rigo. A Strong Version of the Skorohod Representation Theorem.
2022

Berti, Pratelli, Rigo. A Skorohod Representation Theorem Without Separability.

2 Stochastic Differentials

Note that **deterministic differentials**, over a topological space $(S, \mathcal{T})$ can be looked at as more or less a Borel (signed-)measure μ on S. However, we look at differentials in the context of cadlag functions $f : I \to X$ for some interval (bounded or unbounded) I and Banach Space X, and we reconcile this by looking at the Borel measure ν^f generated by the premeasure ν_0^f on the open intervals (a, b) defined by

$$\nu_0^f(a, b) := f(b) - f(a)$$

This gives meaning to the differential phrase "df" in the following way: Assume (for now) that f is complex-valued, and hence that ν^f is a complex-valued measure; integrating over the "differential" "df" equates to integrating over ν^f, i.e., for $(a, b) \subset I$ we have

$$\int_a^b df = \nu^f(a, b) \; (= f(b) - f(a)),$$

or more generally, for ν^f-measurable function $g : I \to X$ and $(a, b) \subset I$, we have

$$\int_a^b g df = \int_{(a,b)} g d\nu^f$$

Another way to define differentials, in terms of measures involves absolute continuity, with respect to another measure (usually the Lebesgue measure), and then utilizng the Radon Nikodym Theorem to get a measure in terms of that other measure. Recall that the definition of absolute continuity and the Radon Nikodym Theorem (for real-valued measures) tell us the following:

Definition 2.1. Let μ and ν be $\mathbb{R}$-valued (deterministic and signed) measures on some measurable space $(\Sigma, \mathcal{A})$. Then ν is **absolutely continuous** with respect to μ, which we denote $\nu \ll \mu$, if for all $A \in \mathcal{A}$, we have

$$\mu(A) = 0 \implies \nu(A) = 0$$

Folland, page 88

Theorem 2.2. *(Real-valued Radon Nikodym Theorem)* Let μ and ν be $\mathbb{R}$-valued (deterministic and signed) measures such that $\mu \ll \nu$. Then there exists some μ-a.e. unique function $h \in L^1(\mu)$ such that

$$\nu = \int_{(-)} h d\mu, \quad (2.1)$$

i.e., for all $A \in \mathcal{A}$, we have $\nu(A) = \int_A h d\mu$.

Folland, Theorem 3.8 (page 90)

Important Note: The Theorem given here is less general than the theorem given in page 90 of
Folland. The more general version deals with the case where ν and μ are any measures, in which case, there exists a decomposition λ, ρ such that $\nu = \lambda + \rho$ and we have $\rho \ll \mu$ and $\lambda \perp \mu$, and we moreover have some $h \in L^1(\mu)$ such that $d\rho = h d\mu$. As you can see, the notation of the more general version is cumbersome, while not offering any useful insight to what we're studying, and so we present the less general version.

Remark 2.1. Whenever we have $\mu \ll \nu$, and hence some function f such that (2.1) holds, we call h the Radon-Nikodym derivative and denote h as $d\nu / d\mu$. We note that this generalizes the notion of derivative of a continuous function $f : I \to \mathbb{C}$, i.e., ν^f is absolutely continuous with respect to the Lebesgue measure m if and only if f is differentiable m-a.e. and for all $A \in \mathcal{B}(I)$, we have

$$\nu^f(A) = \int_A f' dm,$$

i.e., we have $f' = d\nu^f / dm$. This is more or less the statement of the Fundamental Theorem of Calculus, and what this tells us, is in the case where f is differentiable m-a.e., the differential "df" is defined by "$f'dm$".

Remark 2.2. This allows us to generalize the notion of "differentials" specifically for an almost everywhere continuously n-times differentiable function $f : B_n(x, \epsilon) \to \mathbb{R}$. In particular, we find for all $A \in \mathcal{B}(B_n(x, \epsilon))$, we have a well-defined measure

$$\nu^f(A) := \int_A \frac{\partial_n f}{\partial x_1 \partial x_2 \cdots \partial x_n} dm^n$$

where $m^n := \underbrace{m \otimes \cdots \otimes m}_{n \text{ times}}$ is the nth dimensional Legesgue measure. Note in this situation that

$$\frac{d\nu^f}{dm^n} = \frac{\partial_n f}{\partial x_1 \partial x_2 \cdots \partial x_n},$$

and we can define the differential df as ν^f.

Remark 2.3. Note that in this current draft, the setting of the measures, and later in this text random measures, are in the $\mathbb{R}$-valued setting. In this setting, we have the **Radon Nikodym Property**, i.e., an anologue of the *Radon Nikodym Theorem* (mentioned here for $\mathbb{R}$-valued measures as **Theorem 2.2**) holding for said banach space X. In a more general draft, we hope to present everything in the more general case of X-valued measures for a banach space X with the Radon Nikodym Property. In the next report after this one, I hope to talk about the Radon Nikodym Property in greater detail.

What we will look at, now, is stochastic differentials, which we shall simply define as the random measure generated by the corresponding premeasure.

Defintion 2.2. Given an algebra $(E, \mathcal{P})$, we define a **random premeasure** $\nu_0 : \mathcal{P} \to \mathbb{R}$ on $(E, \mathcal{B}(\mathcal{P}))$ to be a function such that

(i) $\nu_0(\cdot, F)$ is $\mathbb{P}$-measurable for any fixed $F \in \mathcal{P}$.

(ii) $\nu_0(\omega, \cdot)$ is a (signed-)premeasure on $\mathcal{P}$, almost surely for $\omega \in \Omega$, i.e., $\nu_0(\omega, \varnothing) = 0$ and given a countable/finite family $\mathcal{F} \subset \mathcal{P}$ of disjoint sets, we have

$\nu_0(\omega, \cup \mathcal{F}) = \sum_{F \in \mathcal{F}} \nu_0(\omega, F)$ (i.e., if $\mathcal{F}$ is infinite, any arrangement $F_1, \ldots, F_n, \ldots$ has $\sum_{n=1}^{\infty} \nu_0(\omega, F_n)$ converging absolutely to $\nu_0(\omega, \cup \mathcal{F})$, almost surely for $\omega \in \Omega$.

Folland, page 30

Remark 2.4.

(i) It's immediate from *Cartheodory's Theorem* (Found in *Folland* § *1.4 (page 27-32)*) applied to $\nu_0(\omega, \cdot)$, for each $\omega \in \Omega$ that ν_0 gives rise to a random measure ν that is $\mathbb{P}$-a.s. uniquely

determined, i.e., any other random measure ν' that also extends ν_0, we have $\nu(\omega, \cdot) = \nu'(\omega, \cdot)$ almost surely for $\omega \in \Omega$.

(ii) We find that **Definition 2.2** is equivalent to the notion of an $\mathcal{L}^0(I; \mathbb{R})$-valued premeasures. This shall be proved in a later draft. Moreover, a future report will try to define more generally premeasures in the context of Banach-valued measures.

Definition 2.3. Given a stochastic process X on some (bounded or unbounded) interval I with a.s. Cadlag paths, we define the **stochastic differential** dX of X to be the random measure ν^X generated by the random premeasure

$$\nu_0^X(a, b) := X(b) - X(a)$$

defined on the algebra of open intervals.
Folland, page 89

Remark 2.5.

(i) Note that the reason Folland is cited for this definition is the way a stochastic differential is defined mirrors the way that differentials are notationally regarded interchangeably with measures in this text, as is indicated by the following statement from page 89:

"We shall use the following notation to express the relationship: $\nu(E) = \int_E f d\mu$:

$d\nu = f d\mu$"

Looking at stochastic differentials, let alone deterministic differentials, in this way will turn out to be useful in our study of stochastic differential equations.

(ii) Note that ν^X is different from μ_X (which denotes the distribution of X on the space $\mathcal{B}\left(\mathbb{R}^I\right)$, where $\mathbb{R}^I$ denotes the product topology on the space of functions $f: I \to \mathbb{R}$), and the superscript notation plus the use of the letter "ν" instead of "μ" clarifies that distinction

(iii) Note that ν^X is indeed a (well-defined) random measure since we can verify that ν_0^X is $\mathbb{P}$-a.s. a premeasure as follows: *property (i)* of **Definition 2.2** is immediately satisfied by definition, and *property (ii)* is verfied from the well-known fact that $f: I \to \mathbb{R}$ being Cadlag implies ν^f is a well-defined premeasure, hence it follows that X having a.s. Cadlag sample

paths inplies that $\nu^X(\omega, \cdot)$ is a premeasure alomst surely for $\omega \in \Omega$.

Ideally, we'd like to think of stochastic differentials as an algebra (similar to that of deterministic 1-forms) with a notion of sum $dX + dY$ and product $dX \cdot dY$, so we need to define the "stochastic differential sum" and "differential product". We again turn to how the deterministic stochastic differential is defined (measure theoretically), and then create an analogous definition in the context of random measures.

From this point on, the definitions and theorems are based on my own creation and conjectures, so no sources will be cited from this point forward.

Remark 2.6. Recall that deterministically for Cadlag $f, g : I \to \mathbb{R}$ the stochastic differentials $df + dg$ and $df \cdot dg$ are defined as the following premeasures $\nu_0^{f,g,+}$ and $\nu_0^{f,g,\cdot}$

$$\begin{aligned}
&\nu_0^{f,g,+}(a,b) := [f(b) - f(a)] + [g(b) - f(a)] \\
&\nu_0^{f,g,\cdot}(a,b) := [f,g]_{(a,b)} \qquad (2.2)
\end{aligned}$$

on the algebra of subintervals of I, where $[f,g]_{(a,b)}$ denotes the covariation of f and g on the interval (a,b), i.e.

$$\begin{aligned}
[f,g]_{(a,b)} := \lim_{||P|| \to 0} &\sum_{k=1}^{n_P} (f(t_k) - f(t_{k-1}))(g(t_k) - g(t_{k-1})), \\
&(P := \{t_0, \ldots, t_{n_P}\},\ t_0 := a < t_1 < \cdots < t_{n_P - 1} < t_{n_P} := b), \quad (2.3) \\
&||P|| := \sup_{1 \le k \le n_+} |t_k - t_{k-1}|.
\end{aligned}$$

We shall also define the differential hdf to be the measure μ on $(I, \mathcal{B}(I))$ defined by

$$\mu(A) = \int_A h d\nu^f,$$

for all $A \in \mathcal{B}(I)$.

Note that these definitions are better motivated than one otherwise might concieve. With $[f,g]_{(a,b)}$ being the basis for the definition $df \cdot dg$ it's worth thinking about it from a Riemann sum perspective, where if we find the product differential of $df \cdot dg$ it makes sense that the Riemann Sum behind it is a product Riemann Sum in the way it is given. This idea shall hopefully be highlighted in more elaborate examples in a future draft.

Definition 2.4. Given stochastic processes X, Y on some (bounded or unbounded) interval I with a.s. Cadlag paths, we define the stochastic differentials $dX + dY$ and $dX \cdot dY$ to be the random measure generated by the random premeasures $\nu_0^{f,g,+}$ and $\nu_0^{f,g,\cdot}$ defined analogously to (2.2) and (2.3), except for X in place of f and Y in place of g, and the limit in (2.3), being in $\mathbb{P}$, i.e., $[X, Y]_{(a,b)}$ is defined to be the random variable z (if it exists) such that for every $\epsilon > 0$, we have

$$\lim_{||P|| \to 0} \mathbb{P}\left[\left|\left|\sum_{k=1}^{n_P} [(X(t_k) - X(t_{k-1}))(Y(t_k) - Y(t_{k-1}))] - z\right|\right| \geq \epsilon\right] = 0,$$

$$(P := \{t_0, \dots, t_{n_p}\},\ t_0 := a < t_1 < \cdots < t_{n_p - 1} < t_{n_P} := b),$$

Remark 2.7. In the deterministic case, we know that $(df)^2 = 0$, since

$$\begin{aligned}
&[f, f]_{(a,b)} := \lim_{||P|| \to 0} \sum_{k=1}^{n_P} (f(t_k) - f(t_{k-1}))^2 \\
&= \lim_{n \to +\infty} \sum_{k=1}^{n} \left(f\left(a + \frac{k}{b-a}\right) - f\left(a + \frac{k-1}{b-a}\right)\right)^2 \\
&\leq \lim_{n \to +\infty} \left(\sup_{1 \leq k \leq n}\left[f\left(a + \frac{k}{b-a}\right) - f\left(a + \frac{k-1}{b-a}\right)\right] \cdot \sum_{k=1}^{n} \left|f\left(a + \frac{k}{b-a}\right) - f\left(a + \frac{k-1}{b-a}\right)\right|\right) \\
&= \lim_{n \to +\infty} \left(\sup_{1 \leq k \leq n}\left[f\left(a + \frac{k}{b-a}\right) - f\left(a + \frac{k-1}{b-a}\right)\right]\right) \cdot \ell(f, a, b) \cdot \\
&= 0,
\end{aligned}$$

where $\ell(f, a, b)$ is the arclength of the image $f[(a, b)]$.

However, it's not necessarily the case that for a stochastic a.s. cadlag process X that $(dX)^2 = 0$. A common example is of course the fact that for a standard one-dimensional Weiner Process W_t, we have $(dW_t)^2 = dt$, which I have shown in my current draft of the 4/22 report, which is an important step of the proof of *Ito's Formula*.

Next, let's get into the Radon-Nikodym Theorem for random measures. Note that this theorem at this moment is basically a conjecture that I haven't gotten around to fully vetting at this moment, but I hope to do so in a later draft.

Definition 2.5. Let μ and ν be $\mathbb{R}$-valued random measures on some measurable space $(E, \mathcal{E})$. Then ν is **absolutely continuous** with respect to μ, which we denote $\nu \ll \mu$ (as we do for deterministic measures), if almost surely, for all $E \in \mathcal{E}$, we have

$$\mu(E) = 0 \Longrightarrow \nu(E) = 0.$$

Theorem 2.6. *(Real-valued Radon Nikodym Theorem on Random Measures)* Let μ and ν be $\mathbb{R}$-valued random measures such that $\mu \ll \nu$. Then there exists some μ-a.e. unique random function $H \in L^1(\mu; \mathbb{R})$ such that almost surely, we have

$$\nu = \int_{(-)} H d\mu, \quad (2.4)$$

i.e., for all $A \in \mathcal{E}$, we almost surely have $\nu(E) = \int_E H d\mu$.

Definition 2.6.

1. Given two real-valued measures μ and ν such that $\mu \ll \nu$, we shall denote the random function (or rather the μ-a.e. equivalence class of random of functions) such that (2.3) holds as $d\mu / d\nu$. This is consistent with the notation in the deterministic case.

2. Let F be a random function $B^n(x, \epsilon) \to \mathbb{R}$ that is amost surely $C^n\left(B^n(x, \epsilon)\right)$ m-almost everywhere. We can define the differential dF of F to be the random measure ν^F defined by

$$\nu^F(A) := \int_E \frac{\partial_n F}{\partial x_1 \partial x_2 \cdots \partial x_n} dm^n,$$

for all $A \in \mathcal{E}$.

Remark 2.8.

(i) We note that since almost surely we have $\dfrac{d\nu^F}{dm^n} = \dfrac{\partial_n F}{\partial x_1 \partial x_2 \cdots \partial x_n}$, this definition of a random differential is an n-dimensional generalization of a random differential of the random function $F : B^n(x, \epsilon) \to \mathbb{R}$ in the situation that F is almost surely n-times differentiable.

(ii) In a future draft, we shall generalize the definition of a random differential of a random function $F : B^n(x,\epsilon) \to \mathbb{R}$ to all almost sure continuous functions, instead of just the ones that are n-times differentiable, using the notion of a (random) premeasure in the n-dimensional case, which can be done via a (random) premeasure (which I shall call the "slice differentials") dF^i on each of the ith slices defined by the change in the value ΔF when there is a change in the ith coordinate input, and then using the notion of the differential product to define $dF := dF^1 \cdot dF^2 \ \cdots \ dF^n$

Note that more work is needed to establish this generalized definition than meets the eye: The *Stochastic Fubini Thoerem* (as given in the last *4/15 draft*) will be needed to show that this differential is almost surely defined by the product measure of each of these slices, and a generalized definition of the differential product (or covariation) applied in the context of slice differentials in the n-variable setting is also needed.

Sources:

Real Analysis, Modern Techniques § 1.4, § 3.2
Folland

3 Definitions of SDE's and Solutions

Generally, we think of (partial) differential equations as equations involving derivatives, i.e., it is of the form

$$g\big(f(y), Df(y), D^2 f(y), \ \ldots, D^m f(y)\big) = 0,$$

where $f \in C^m\big(B^n(x,\epsilon)\big)$ and $m, n \geq 1$.

Stochastic (partial) differential equations (SDE's or SPDE's) are often given in the form

$$g\big(f(Y), Df(Y), D^2 f(Y), \ \ldots, D^m f(Y)\big) = 0,$$

where $f \in C^m\big(B^n(x,\epsilon)\big)$ and $m \geq 1$, and $Y := (Y_1, \ \ldots \ Y_n)$ is an $\mathbb{R}^n$-valued stochastic process and $m, n \geq 1$.

The goal involving equations such as these is to (a) Find the solutions to them, and (b) Represent mathematical phenomenon to them. For this section, we shall give some examples of SPDE's.

First, we shall formally define Stochastic Partial Differential Equation.

Definition 3.1. A **Stochastic Partial Differential Equation (SPDE)** is a linear equation of differentials (as defined in **Definition 2.3**), i.e., a linear equation of measures of the form

$$0 = \nu^{X_1} + \cdots + \nu^{X_n},$$

where as stated before, $\nu^{X_1}, \ldots, \nu^{X_n}$ refer to the random measure generated by the random premeasure $\nu_0^{X_k}$ defined by

$$\nu_0^{X_k}(a, b) := X_k(b) - X_k(a)$$

which we write out in differential form as

$$0 = f_1 dX_1 + \cdots + f_n dX_n$$

almost surely, for processes $f_1, \ldots, f_n$, each of which we assume to be $\mathcal{L}^1\left(\nu^{X_k}\right)$, for each $1 \leq k \leq n$.

We usually take some differentials/random functions of this equation as given, and others as unkowns. The collection of differentials/random functions that when plugged into the set of unkowns satisfy the SPDE--i.e. if we have

$$0 = \int_0^t f_1(s) dX_1(s) + \cdots + \int_0^t f_n(s) dX_n(s),$$

satisfied almost surely--are known as **solutions**.
Source: Gard, page 64-68

Ideally, we start with a fixed stochastic basis on which u_0 is defined and we find a process u satisfying the same properties with respect to the fixed stochastic basis.

In this case u is called a **pathwise solution**.
Source: M647 Lecture 5 (revised) Definition 5.1.1.

Remark 3.1.

(i) I cited *Introduction to Stochastic Diffferential Equations* by Thomas Gard for **Definition 3.1**

because the ideas most closely correspond to his ideas--expressed in particular in (0.9) and (1.1). But the hope is to generalize these ideas in a future draft to the case where the differentials are Hilbert-space, or even Banach-space, valued.

(ii) This distinction between given variables and unknown variables I will work on for the next draft, and then using that I will define solutions, though in the context of **Example 3.4**, I'll briefly discuss the notion of a **martingale solution**.

(iii) In the next draft, I hope to use *Stochastic Partial Differential Equations* by Sergy V. Lototsky and Borris L. Rozovsky (Ch. 2 for this section in particular) to talk about other kinds of solutions in broader detail. For the time being, the notions of solutions discussed here more or less fall in line with a **classical solution**.

Example 3.2. It is worth noting that *Ito's Formula* is the first stochastic differential equation that we come across. We can write it out as

$$d(F \circ X) = \left(\frac{\partial(F \circ X)}{\partial t} + \frac{\partial^2(F \circ X)}{\partial x^2}\right)dt + \frac{\partial(F \circ X)}{\partial x}dW, \quad (3.1)$$

where $F(t) := f(X(t), t)$, X is the (unknown) stochastic process satisfying the SPDE

$$dX = \varphi dt + \Psi dW, \quad (3.2)$$

for some processes φ, Ψ and $f \in C^2(\mathbb{R} \times [0, +\infty))$. So what *Ito's Formula* tells us in terms of SPDE's is that any process X that is a solution to (3.2) is also a solution to (3.1).
Source: Brownian Motion Notes (linked here)

Example 3.3. The next SPDE is the Levy-Kitchen decompostion, which consists of

$$dL = adt + dW + d\pi + d\widehat{\pi},$$

where L is a Levy process, π is a compound Poisson process, and $\widehat{\pi}$ is a series of compensated compound Poisson processes.

Note that this SPDE happens to have the stochastic equation form of

$$L(t) = at + W(t) + \mathcal{P}_0(t) + \underbrace{\textstyle\sum_{n=1}^{\infty} \widehat{P}_n(t)}_{\mathcal{P}}.$$

Source: M647 Lecture 4 (revised), page 4

Example 3.4. And finally, we have the Stochastic Partial Differential Equations with a Nonlinear Monotone Operator Driven by a Levy Noise

$$
\begin{aligned}
du + [A(u(s)) + B(u(s), u(s))]ds &= G(u(s-))dW(s) + \mathbb{1}_{E_0}K(u(s-),\xi)d\widehat{\pi}(s,\xi) \\
&\quad + \mathbb{1}_{E\setminus E_0}\mathcal{K}(u(s-),\xi)d\pi, \qquad (3.3) \\
u(0) &= u_0,
\end{aligned}
$$

where $u \in L_2(\mathcal{U}_0, \mathcal{H})$, A and B are linear and bilinear operators respectively, and $G, K, \mathcal{K}$ are noise operators.
Source: M647 Lecture 5 (revised) (5.1) *(page 1)*

Remark 3.2. Note the slight change in notation from (3.3) to (5.1) of the cited *M647 Lecture 5* notes cited, which has the equation written

$$
(3.4)\begin{cases}
du + [A(u) + B(u(s), u(s))]ds = G(u(s-))dW(s) + \displaystyle\int_{E_0} K(u(s-),\xi)d\widehat{\pi}(s,\xi) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \displaystyle\int_{E\setminus E_0} \mathcal{K}(u(s-),\xi)d\pi(s,\xi) \\
u(0) = u_0
\end{cases}
$$

To make the differential equation notation of these notes consistent with **Definition 3.1**. wrote the stochastic differential terms $\mathbb{1}_{E_0}K(u(s-),\xi)d\widehat{\pi}(s,\xi)$ and $\mathbb{1}_{E\setminus E_0}\mathcal{K}(u(s-),\xi)d\pi$ in place of the integrals $\int_{E_0} K(u(s-),\xi)d\widehat{\pi}(s,\xi)$ and $\int_{E\setminus E_0}\mathcal{K}(u(s-),\xi)d\pi(s,\xi)$, respectively.

Also concerning this example, it's worth mentioning the notion of a martingale solution
Consider an initial condistion $u_0 \in L^2(\Omega, \mathcal{F}, P; \mathcal{V})$ with laws u_0. A pair $\left(\widetilde{S}, \widetilde{u}\right)$ is a **(strong, global) martingale solution** to (3.3) if we have

- A stochastic basis $\widetilde{S} = \left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \left(\widetilde{\mathcal{F}}_t\right)_{t\geq 0}, \widetilde{P}, \widetilde{W}, \widetilde{\pi}\right)$
- $\widetilde{u} \in L^2\left(\widetilde{\Omega}, L^\infty(0,T;\mathcal{V})\right) \cap L^2\left(\widetilde{\Omega}, L^2(0,T;D(A))\right)$
- $\widetilde{u}$ has cadlag sample paths in $\mathcal{H}$ a.s. and $\widetilde{u}$ is $\widetilde{\mathcal{F}}_t$-adapted.

- and $\forall t \in [0, T]$, $\widetilde{\mathbb{P}}$-a.s. we have,

$$\widetilde{u}(t) + \int_0^t \left[A\widetilde{u}(s) + B\left(\widetilde{u}(s), \widetilde{u}(s)\right) \right] ds = \widetilde{u}(0) + \int_0^t G\left(\widetilde{u}(s-)\right) d\widetilde{W}(s)$$
$$+ \int_{(0,t]} \int_{E_0} K\left(\widetilde{u}(s-), \xi\right) d\widehat{\widetilde{\pi}}(s, \xi).$$

and $\widetilde{u}(0)$ is $\widetilde{\mathcal{F}}_0$-measurable and has law μ_0.
Source: M647 Lecture 5 (revised), Definition 5.1.1.

Sources:

M647 Lecture 4-5 (revised)

Introduction to Stochastic Differential Equations § *3.0,* § *3.1*
Gard

Brownian Motion Notes (linked here)