Statistics 212: Lecture 1 (January 27, 2025)

Preview of Topics and Radon-Nikodym Theorem

Instructor: Mark Sellke

Scribes: Kevin Liu, Haozhe (Stephen) Yang

1 Lecture 1

Today we're focusing on a preview of future topics and proof of Radon-Nikodym Theorem. Main topics for today:

- Advanced martingales
- Brownian motion
- Ito (Stochastic) Calculus

See Instructor Website for more info. Also, sign ups for 5 minute meetings with Mark Sellke 1-2:30 Wed, Jan 29 or Mon Feb 3. Form to be sent out!

1.1 Preview of Brownian Motion and Ito Calculus

Definition 1.1 (Brownian Motion)**.**

(a) Einstein's definition. Scaling limit of a simple random walk. An example of a simple random walk is

$$\begin{aligned} x_0 &= 0 \\ x_1 &= \pm 1 \\ x_2 &= x_1 \pm 1 \\ &\vdots \end{aligned}$$

where all the $\pm$ are iid uniform. If we "scale out" the graph of the simple random walk, by the central limit theorem, we have $x_t \approx \mathcal{N}(0,t)$ where t is a very large number (say a googol). We have $B_s = \frac{x_t}{\sqrt{10^{100}}} \sim \mathcal{N}(0,s)$. Graphing out these B's, we obtain a graph that is a random fractal.

(b) Wiener's definition. Brownian motion on $t \in [0,1]$ is a random Fourier series. We have

$$B_t = g_0 t + \sum_{k \geq 1} g_k \sqrt{2} \frac{\sin(\pi k)}{\pi k},$$

where $g_0, g_1, \ldots$ are IID standard Gaussian. (This is called a Karhunen–Loève decomposition.)

(c) Gaussian process. The value at every timestamp is a Gaussian. We have

$$E[B_s] = 0$$
$$\mathrm{Cov}(B_{s_1}, B_{s_2}) = \min(s_1, s_2).$$

Explanation for covariance: If $s_1 < s_2$, then

$$E[B_{s_1}^2] = s_1$$
$$E[B_{s_1}(B_{s_2} - B_{s_1})] = 0,$$

so $E[B_{s_1} B_{s_2}] = s_1$.

Definition 1.2 (Ito Calculus)**.** Can think about Ito Calculus as calculus for Brownian motion or processes that are similar to Brownian motion.

Examples:

- Stock prices. They are continuous, and we can think of it as a martingale. However, a stock price can have a time-changing volatility, which is not quite Brownian (Brownian motion has constant volatility over time, so it looks the same everywhere). Ito Calculus allows us to analyze these quasi-Brownian objects.
- $Z_t = B_t^2$. The process would clearly never go negative, and it oscillates more when at large values. Even though this isn't quite Brownian, we can still use Ito calculus on this process.
- Used for biology, optimal control, PDEs, complex analysis, diffusion sampling, quantum mechanics, etc.

1.2 Radon-Nikodym Theorem

Definition 1.3 (Absolute continuity of finite measures)**.** $\nu \ll \mu$ indicates that: the finite measure ν is absolutely continuous with respect to μ if for every measurable set S such that $\nu(S) = 0$ implies $\mu(S) = 0$. Equivalently, we have $\nu(S) > 0$ implies $\mu(S) > 0$. We say that ν is absolutely continuous with respect to μ.

Theorem 1.4 (Radon-Nikodym Theorem)**.** *Start off with finite measure μ on $(\Omega, \mathscr{F})$. Essentially, RN tells us what kind of measures we can produce starting with μ.*

(a) *If $\nu \ll \mu$ (i.e., ν is absolutely continuous with respect to μ), then there exists a non-negative integrable f such that*

$$\nu(S) = \int_S f(\omega)\, d\mu(\omega) = \int_S f\, d\mu$$

for any measurable set $S \in \mathscr{F}$. We define $f = \frac{d\nu}{d\mu}$ as the Radon-Nikodym derivative.

(b) *(More general) Without assuming absolute continuity, there exists a non-negative integrable f and finite measure Θ such that we can decompose*

$$\nu(S) = \Theta(S) + \int_S f(\omega)\, d\mu(\omega).$$

Furthermore Θ, μ are disjointly supported, i.e. there exists an $S \in \mathscr{F}$ such that $\mu(S) = 0$ and $\Theta(\Omega \setminus S) = 0$. We call $\Theta(S)$ the "singular part" and the second term the "absolute continuous part" of ν.

(c) *Assume $\nu \le \mu$, i.e. $\nu(S) \le \mu(S)$ for all $S \in \mathscr{F}$. Then there exists a measurable $f : \Omega \to [0,1]$ with $f = d\nu / d\mu$.*

Remark. We can show $(b) \implies (a) \implies (c)$. Also note that if we're given (c), then for general finite measures (ν, μ), we have $\nu \le \nu + \mu$ so we can simply apply (c) to the pair $(\nu, \nu+\mu)$. On the first homework, we will show this recovers (a) and (b). Intuitively, all 3 statements have the same core difficulty, that one has to "conjure up" the function f out of thin air.

Proof. Simplest proof that Mark was able to find, by Anton Schep (2003). We prove the RN Theorem in the form (c). The idea is to find the largest f such that $f \le d\nu/d\mu$ and show equality holds. Define

$$H = \left\{ f : \Omega \to [0,1]; \forall S \in \mathscr{F}, \int_S f \, d\mu \le \nu(S) \right\}. \tag{1}$$

We want to find the maximum of H. For intuition, one can consider what happens **assuming** a Radon-Nikodym derivative $f_* = \frac{d\nu}{d\mu}$ exists. Then for arbitrary measurable f_1, we have $f_1 \in H$ if and only if $f_1 \le f_*$ holds almost everywhere.

Our first claim that H is closed under maximum, i.e. if $f_1, f_2 \in H$ then $\max(f_1, f_2) \in H$. Assuming a Radon-Nikodym derivative exists, this is just because if $f_1, f_2 \le f_*$ almost everywhere, then $\max(f_1, f_2) \le f_*$. However we can prove it just from the given condition. Let $A = \{\omega \in \Omega : f_1 \ge f_2\}$ and $B = \{\omega \in \Omega : f_1 < f_2\}$ (the complement of A). We have

$$\begin{aligned}
\int_S \max(f_1, f_2) d\mu &= \int_{S\cap A} f_1 \, d\mu + \int_{S \cap B} f_2 d\mu \\
&\le \nu(S \cap A) + \nu(S \cap B) \\
&= \nu(S).
\end{aligned}$$

Thus, $\max(f_1, f_2) \in H$ as well.

Following this observation, we will aim to demonstrate $f_* \in H$ by taking repeated maximums. We attempt to define $f_*(\omega) = \max_{f \in H} f(\omega)$. But this is a faulty definition. Suppose that $\mu(\{\omega\}) = 0 \forall \omega \in \Omega$, i.e. μ has no atoms. Then $f_*(\omega) = 1$ for all ω because $f_\omega(x) = \mathbb{1}_{x=\omega} \in H$.

Instead, we have to take the max of finitely or countably many functions. For $k = 1, 2, \ldots$, define $g_k : \Omega \to [0,1]$. as follows. Let

$$M = \sup_{f \in H} \int_\Omega f \, d\mu.$$

We require $g_k \in H$ with $\int_\Omega g_k d\mu \ge M - \frac{1}{k}$. We can repeatedly take maximums as so:

$$\begin{aligned}
f_1 &= g_1 \in H \\
f_2 &= \max(g_1, g_2) \in H \\
f_3 &= \max(g_1, g_2, g_3) = \max(f_2, g_3) \in H. \\
&\vdots
\end{aligned}$$

Note that $0 \le f_1 \le f_2 \le \cdots \le 1$. By the monotone convergence theorem, there exists an $f_* = \lim_{k\to\infty} f_k$. We want to show that f_* is the RN-derivative.

We can see $f_* : \Omega \to [0,1]$. Less obvious but crucial is that

$$\int_\Omega f_* \, d\mu = M. \tag{2}$$

As

$$\int_\Omega f_* \, d\mu \ge \int_\Omega g_k d\mu \ge M - \frac{1}{k}$$

for all $k \in \mathbb{N}$, we see $\int_\Omega f_* \, d\mu \geq M$. By Fatou's Lemma, we see $\int_\Omega f_* \, d\mu \leq M$ as $f_k \in H$ for all k. We also see that $f_* \in H$ as by Fatou's Lemma, we have $\int_S f_* \, d\mu \leq \liminf_{k\to\infty} \int_S f_k \, d\mu \leq \nu(S)$.

To show that $\int_S f_* \, d\mu = \nu(S)$ for all S, we can proceed with proof by contradiction. Assume that there exists some S such that $\int_S f_* \, d\mu < \nu(S)$, so intuitively there is a "deficit" in S that we have not yet exhausted. We will try to increase f_* while remaining in H, which contradicts maximality of M (due to (2)).

Define $E_1 = \{\omega : f_*(\omega) = 1\}$. We first show that the "deficit" does not come from the part of S in E_1. Recall the initial assumption that $\nu \geq \mu$. But we also have

$$\mu(S \cap E_1) = \int_{S\cap E_1} f_* \, d\mu \leq \nu(S \cap E_1)$$

as $f_* \in H$. Then

$$\mu(S \cap E_1) = \int_{S\cap E_1} f_* \, d\mu = \nu(S \cap E_1).$$

Therefore we can replace S by $S \backslash E_1 = S \cap E_0$, where $E_0 = \Omega \backslash E_1$ is the complement of E_1.

Next we want a positive amount of space to increase f_*, so we "exhaust" E_0. For each $n \geq 1$, define

$$F_n = \{\omega : f_*(\omega) \leq 1 - \frac{1}{n}\}.$$

These F_n exhaust E_0 in that $F_1 \subseteq F_2 \subseteq \cdots$, and

$$\bigcup_{n=1}^{\infty} F_n = E_0.$$

Then $\int_{S\cap E_0} f_* \, d\mu < \nu(S \cap E_0)$, which implies $\int_{S\cap F_n} f_* \, d\mu < \nu(S \cap F_n)$ for large n. Define $\bar{S} = S \cap F_n$. For $\epsilon > 0$ sufficiently small, we have

$$\int_{\bar{S}} (f_* + \epsilon\chi_{\bar{S}}) d\mu < \nu(\bar{S}). \tag{3}$$

We want to show $f_* + \epsilon\chi_{\bar{S}} \in H$ to contradict the maximality of f_*. Although (3) appears like the condition to be within H, it is only for a specific set $\bar{S}$, while H is a condition on *all* measurable sets. So a bit more is still needed.

To finish the proof, we need one more exhaustion argument. First, the condition for $f_* + \epsilon\chi_{\bar{S}}$ to be in H holds on any set disjoint from $\bar{S}$, since the extra $\epsilon\chi_{\bar{S}}$ term doesn't matter. So it remains to handle subsets $\tilde{S} \subseteq \bar{S}$ (since in general we can decompose $\tilde{S}$ into $\tilde{S} \cap \bar{S}$ and $\tilde{S} \backslash \bar{S}$). It will be convenient to define the "ϵ-deficit"

$$\mathrm{Def}_\epsilon(A) = \nu(A) - \int_A f_* \, d\mu - \epsilon\mu(A).$$

Note that this function is additive. Further, $\mathrm{Def}_\epsilon(\bar{S}) > 0$ by (3). The idea is that if $S_1 \subseteq \bar{S}$ violates the H-condition, i.e.

$$\int_{S_1} (f_* + \epsilon\chi_{\bar{S}}) d\mu = \int_{S_1} (f_* + \epsilon\chi_{S_1}) d\mu > \nu(S_1),$$

this means $\mathrm{Def}_\epsilon(S_1) < 0$ is negative. Then we can simply remove S_1 and use additivity of this functional to find that

$$\mathrm{Def}_\epsilon(\bar{S} \backslash S_1) = \mathrm{Def}_\epsilon(\bar{S}) - \mathrm{Def}_\epsilon(S_1) \geq \mathrm{Def}_\epsilon(\bar{S}) > 0. \tag{4}$$

Thus intuitively, removing a violating set S_1 (with negative deficit) only widens the deficit. So by removing "all possible violating sets", there will be no more room for violations. To be precise, we construct a sequence of disjoint sets $S_1, S_2, \cdots \subseteq \bar{S}$, each of which attains an "almost maximal violation" (subject to the restriction of disjointness with previous sets). Namely define

$$a_k = \inf_{S_k} \mathrm{Def}_\epsilon(S_k)$$

with the infimum being over $S_k \subseteq \bar{S}$ disjoint with $S_1, S_2, \ldots, S_{k-1}$. We have $a_k \leq 0$ since the empty set is always an option. We'll choose S_k to approximately optimize this infimum up to margin $1/k$, i.e.:

$$\mathrm{Def}_\epsilon(S_k) \leq a_k + \frac{1}{k}. \tag{5}$$

Then we can define

$$\hat{S} = \bar{S} \setminus (\bigcup_{k \geq 1} S_k)$$

to be "$\bar{S}$ with all the deficit removed". Now let's verify that $\hat{f} = f + \epsilon \chi_{\hat{S}} \in H$, and that this contradicts maximality of M to finish the proof:

- First, similarly to (4), it follows that

$$\int_{\hat{S}} \hat{f}\, d\mu - \nu(\hat{S}) = \int_{\bar{S}} (f_* + \epsilon \chi_{\bar{S}})\, d\mu - \nu(\bar{S}) < 0.$$

  In particular, this means that $\nu(\bar{S}) > 0$ so $\bar{S}$ is non-empty. Additionally, $\nu(\bar{S}) \leq \mu(S)$. Therefore **if** we can verify that $\hat{f} \in H$ **then** we will contradict maximality of M.

- As argued before, the condition (1) holds for $f + \epsilon \chi_{\hat{S}}$ automatically on sets disjoint from $\hat{S}$, since $f \in H$. By additivity, it suffices to check (1) for any remaining subset $S_0 \subseteq \hat{S}$. (I.e. for a general set E, we can check for both $E \cap \hat{S}$ and $E \setminus \hat{S}$ and add as before.)

- So, suppose for contradiction that $S_0 \subseteq \hat{S}$ violates (1), i.e.

$$\int_{S_0} \hat{f}\, d\mu > \nu(S_0) + \delta$$

  for some positive δ, or equivalently $\mathrm{Def}_\epsilon(S_0) < -\delta$. Then for $k > 1/\delta$, we see that S_k was chosen wrong: we could have used $S_k \cup S_0$ instead, and since they are disjoint we have

$$\mathrm{Def}_\epsilon(S_k \cup S_0) = \mathrm{Def}_\epsilon(S_k) + \mathrm{Def}_\epsilon(S_0) \leq \mathrm{Def}_\epsilon(S_k) - \delta \leq \mathrm{Def}_\epsilon(S_k) - 1/k.$$

  This shows that S_k does not obey the approximate-optimality condition (5). This gives the desired contradiction and concludes the proof.

$\square$