Some Mathematical Preliminaries

Vladimir Feinberg

April 13, 2021

1 Probability Spaces, Random Variables and Stochastic Processes

This section introduces standard measure theoretic structures, all of which are pretty standard. A couple lemmas and theorems are provided, some of which are definition-related, such as Doob-Dynkin for conditional expectation and Kolmogorov Extension to be able to construct stochastic processes with prescribed distributions on infinite spaces.

2 An Important Example: Brownian Motion

Definition 2.1 (Brownian Motion). Brownian Motion B_t is a $\mathbb{R}^n$-valued process over $[0, \infty)$ characterized by

1. origin: $B_0 = 0$,

2. independent increments: for any $t, h \geq 0$, $B_{t+h} - B_t \perp\!\!\!\perp \{B_s \mid s \leq t\}$

3. Gaussian increments: for any $t, h \geq 0$, $B_{t+h} - B_t \sim N(\mathbf{0}, hI)$

4. continuous sample paths $t \mapsto B_t(\omega)$ for almost all $\omega \in \Omega$.

The first three can be checked to yield existence of some process via Kolmogorov extension. Via Kolmogorov Continuity Theorem, one can show that a continuous version of such a process exists.

Theorem 1 (Kolmogorov's Continuity Theorem). *Suppose a stochastic process U_t on a separable, complete metric space with indices $t \in S \subset \mathbb{R}^d$ satisfies*

$$\mathbb{E}\left[d(U_x, U_y)^\alpha\right] \leq C\|x - y\|^{d+\beta} \ .$$

Then, U admits a continuous version over S.

The continuity theorem proof looks essentially like chaining. Many other characterizations exist for Brownian Motion, also called a Wiener process.

Definition 2.2 (Off-origin Brownian Motion). Denote by $\mathbb{P}^x$ and $\mathbb{E}^x$ the laws of off-origin Brownian motion starting at $B_0 = x$.

3 Exercises

Theorem 2 (Law of the Unconscious Statistician, LOTUS). *Consider a rv X with measure μ. Then*

$$\mathbb{E}\, g(x) = \int g \, \mathrm{d}X_* \mathbb{P}$$

Proof. By change of variables,

$$\mathbb{E}\,g(x) = \int_\Omega g(X)\,\mathrm{d}\,\mathbb{P} = \int_X g\,\mathrm{d}X_*\,\mathbb{P}\;,$$

$\square$

Though it seems like a simple one-step formula, LOTUS allows us to actually evaluate expectations. Written as a measure-theoretic expectation, $(X_*\,\mathbb{P})(g)$ is defined as a limit of simple functions. This, in turn, simplifies in most cases for Borel X, which states that $\mu = \mu_\lambda + \mu_\delta + \mu_{\text{singular}}$, where μ_λ is absolutely continuous with respect to the Lebesgue measure, μ_δ is atomic, and the remaining singular distribution μ_{singular} is quite exceptional: it's basically everything else, which is diffuse measures which admit no Borel subsets of positive Lebesgue measure. An example of the latter is the Cantor distribution (recall the Cantor set is Borel!).

Actual computation must be proven to be equivalent to familiar forms: for Riemann integrable functions, Lebesgue integration is equal to the city-style integration, but otherwise must be interpreted as the limit of pancake-style integration (visualization). By Radon-Nikodym, given μ_λ we can compute this analogous integral.

For the discrete part μ_δ using the probability axioms directly for disjoint sets, integrals are countable sums.

I don't really know how to compute integrals μ_{singular}, since they're not well-characterized. For certain examples such as the Cantor distribution, fractal structure can be leveraged to recover recurrences for expectations (since nonzero singular measures are necessarily uncountable, it's difficult to sum over, but in practice these may come up in game-like settings).

3.1

Let X be a real-valued function on ω with corresponding σ-algbra $\mathcal{F}$ taking on only countably many values $\{a_k\}$ in $\mathbb{R}$.

3.1.1

X is a random variable iff $X^{-1}\{a_k\} \in \mathcal{F}$ for all k. ($\implies$) is immediate, as the singleton sets are Borel. ($\impliedby$) as any Borel set contains at most countably many a_k singletons, whose preimage unions must then also be in $\mathcal{F}$.

3.1.2

The rest of the problem follows from LOTUS.

3.2

For a real rv X define its distribution function as $F(x) = \mathbb{P}\{X \leq x\}$.

3.2.1

By definition of a measure, F goes from 0 to 1 across the number line and is bounded by the two in between. The limiting properties hold by considering the nested sets and monotone convergence of $\bigcup_{x=1}^\infty \{X \leq x\} = \Omega$.

3.2.2

F is increasing as measures are monotonic over their events.

2

3.2.3

F is rc (right continuous), again because of nested set limits $\bigcap_{y>x} \{X \leq y\} = \{X \leq x\}$. Notice left continuity does *not* hold analogously.

3.3

Let $\{\mathcal{F}_i\}_{i \in I}$ be a family of σ-algebras. Their intersection is then also a σ-algbra. This holds immediately because all σ-algebra properties are of the form $\forall j \in J \ B_j \in \mathcal{F} \implies B \in \mathcal{F}$ for various index sets J. This is invariant under intersection.

3.4

Chebyshev's inequality $\mathbb{P}\{|X| \geq \lambda\} \leq \frac{\mathbb{E}|X|^p}{\lambda^p}$ follows for $p > 0$ from

$$\mathbb{E}|X|^p \geq \int_{X>\lambda} |X|^p \, d\mathbb{P} = \lambda^p \int_{X>\lambda} \frac{|X|^p}{\lambda^p} d\mathbb{P} \geq \lambda^p \int_{X>\lambda} d\mathbb{P} = \lambda^p \, \mathbb{P}\{X > \lambda\}$$

Then an MGF inequality stating that $M = \mathbb{E}\exp(k\,|X|) < \infty$ implies $\mathbb{P}\{X \geq \lambda\} \leq M\exp(-k\lambda)$ may be proven the same way.

3.5

Let $X \perp\!\!\!\perp Y$. Then $\mathbb{E}[XY] = \mathbb{E}X\,\mathbb{E}Y$. As described in the exercise, by monotone convergence it suffice to consider simple functions $X = \sum_i a_i 1_{A_i}, Y = \sum_j b_j 1_{B_j}$ where $\sigma(X) \perp\!\!\!\perp \sigma(Y)$ implies $\mathbb{P}(A \cap B) = \mathbb{P}A\,\mathbb{P}B$. Then the discrete expectation factors as desired.

3.6

The Borel-Cantelli lemma states that for $n \in \mathbb{N}$, events A_n such that $\sum_n \mathbb{P}A_n < \infty$, $\mathbb{P}\bigcap_m B_m = 0$ where $B_m = \bigcup_{n=m}^{\infty} A_n$. To show this, note by measure continuity that $\mathbb{P}\bigcap_m B_m = \lim_m \mathbb{P}B_m$. Then by union bound, $\mathbb{P}B_m \leq \sum_{n=m}^{\infty} \mathbb{P}A_n \to 0$ where the latter tends to 0 in m because the sum of probabilities must converge.

3.7

3.7.1

A Dynkin system made of unions of a finite partition G_i of Ω is a σ-algebra is necessarily a π-system and thus a σ-algebra by the π-λ theorem because the intersection of union of disjoint sets is just the union of the common sets.

3.7.2

Any finite σ-algebra on Ω is a Dynkin system of a finite partition.

Given a σ-algebra $\mathcal{F}$, consider $P = \{G \in \mathcal{F} \mid \nexists H \in \mathcal{F}, H \neq \emptyset, H \subset G\}$. P must be disjoint, since otherwise assume for contradiction it contains nondisjoint G, H. As $G \cap H \in \mathcal{F}$ we have a contradiction, as both G, H are in P yet they do not meet its non-containment property. On the other hand, P is a partition, because any $\omega \in \Omega$ absent from all $G \in P$ is present in $\Omega \setminus \bigcup_{G \in P} G \in \mathcal{F}$, which must itself be P, a contradiction. Finally, P generates $\mathcal{F}$: any $F \in \mathcal{F}$ must be covered by some subset of the partition $J \subset P$. Consider any such cover with the smallest cardinality such that union $U = \bigcup_{j \in J} G_j$ satisfies $F \subset U$. Then consider $U \setminus F$, if this is empty then we have shown that all sets in $\mathcal{F}$ are generated by unions of P. So assume for contradiction it is nonempty. This difference must be covered by some G_j for fixed $j \in J$. But G_j must be disjoint with F, since otherwise it contains the $U \setminus F \in \mathcal{F}$ set, violating its membership with P. This contradicts the minimality of U, which could do without G_j.

3.7.3

Suppose X is a real-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{F}$ is finite then then show X has a finite image.

By Sec. 3.7.2, we can find a finite partition $P = \{F_i\}$ which generates $\mathcal{F}$ with Dynkin completion. Consider for any fixed i two $\omega_1, \omega_2 \in F_i$. If $X(\omega_1) \neq X(\omega_2)$ then $A_1 = f_i \cap X^{-1}X\{\omega_1\}$, $A_2 = F_i \cap X^{-1}X\{\omega_2\}$ are disjoint and in $\mathcal{F}$. Yet as subsets of F_i they cannot be recovered by a Dynkin completion of P. Thus, $|XF_i| = 1$ for all i, so $|X\Omega| \leq \sum_i |XF_i| < \infty$.

3.8

Let W_t be 1-dimensional Brownian motion from the origin.

3.8.1

Recall that the multivariate CF for x-origin d-dimensional Brownian motion at a collection of times $t_1, \cdots, t_k$ is given by the following function of k vectors $u_j \in \mathbb{R}^d$:

$$\mathbb{E}\exp\left(\imath \sum_{j=1}^k u_j^\top B_{t_j}\right) = \exp\left(\imath \sum_j x^\top u_j - \frac{1}{2}\sum_{j_1=1}^d \sum_{j_2=1}^d t_{j_1 \wedge j_2} u_{j_1}^\top u_{j_2}\right)$$

Thus in our 1-dimensional case, $\mathbb{E}\exp\left(\imath u W_t\right) = \exp\left(-\frac{1}{2}u^2 t\right)$

3.8.2

As exp is entire, it is perfectly formal to rewrite the power series like so:

$$\mathbb{E}\exp\left(\imath u W_t\right) = \mathbb{E}\sum_{k=0}^\infty \frac{(\imath u W_t)^k}{k!} \ ,$$

where strictly speaking the measure integration performed by $\mathbb{E}$ is really performing two separate real integrals, and the extension to $\mathbb{C}$ can be viewed like sugar for $\mathbb{E} = \mathbb{E}\operatorname{Re} + \imath \mathbb{E}\operatorname{Im}$. Then we may do some limit exchanges,

$$\partial_u \mathbb{E}\sum_{k=0}^\infty \frac{(\imath u W_t)^k}{k!} = \mathbb{E}\,\partial_u \sum_{k=0}^\infty \frac{(\imath u W_t)^k}{k!} = \mathbb{E}\sum_{k=0}^\infty \partial_u \frac{(\imath u W_t)^k}{k!} \ ,$$

justified by Dominated Convergence Theorem (as for real arguments uW_t, $\exp(\imath u W_t)$ is on the unit circle) and then absolute convergence, which is really itself a form of dominated convergence.

Given the setup above, we may take several derivatives:

$$\partial_u^{2k} \mathbb{E}\exp\left(\imath u W_t\right)\big|_{u=0} = \mathbb{E}\left[\imath^{2k} W_t^{2k}\right] \ ,$$

which setting equal to $\partial_u^{2k}\exp\left(-\frac{1}{2}u^2 t\right)\big|_{u=0} = (-1)^k \frac{(2k)! t^k}{2^k \cdot k!}$ yields a formula for the $(2k)$-th moment (and thus satisfies the condition for Kolmogorov Continuity).

3.8.3

The above was completely formal, so the non-characteristic proof is superfluous.

3.9

Kolmogorov Extension may give rise to non-continuous versions, since even agreement in distribution (and therefore any finite-dimensional distribution of a subset of a process' indices) can be equivalent while path properties differ.

Consider a probability space on $\Omega = [0, \infty)$ with Borel σ-algebra and any diffuse measure μ. We can define two processes on $[0, \infty)$, $Y_t = 0$ and $X_t(\omega) = 1\{t = \omega\}$. Notice the processes are equal in law. By disjointness and measure properties,

$$\mathbb{P}\{X_{t_1} \in A_1, \cdots, X_{t_k} \in A_k\} = \mathbb{P}\{X_{t_1}^{-1} A_1 \setminus t_1, \cdots, X_{t_k} \in A_k\} + \mathbb{P}\{\omega = t_1, \cdots, X_{t_k} \in A_k\} \ .$$

But $\mathbb{P}\{\omega = t_1, \cdots, X_{t_k} \in A_k\} \leq \mathbb{P}\{\omega = t_1\} = 0$ as μ is diffuse, so by repeating the same reasoning across the other indices, the original expression

$$\mathbb{P}\{X_{t_i} \in A_i \,|\, i \in [k]\} = \mathbb{P}\{X_{t_i}^{-1} A_i \setminus t_i \,|\, i \in [k]\} = 1\{\forall i \in [k], \ 0 \in A_i\} = \mathbb{P}\{Y_{t_i} \in A_i \,|\, i \in [k]\}$$

3.10

Brownian motion has stationary increments, which follows immediately from translation invariance of Brownian motion.

3.11

d-dimensional Brownian motion B_t is equal to $B_t = (B_t^{(1)}, \cdots, B_t^{(d)})$ d components of independent 1-dimensional Brownian motion. This follows from the CF of B_t, which factors componentwise for any finite set of indices $t_1, \cdots, t_k$, proving independence, and that components have the same Brownian process law.

3.12

For any $t_0 \geq 0$, $B_{t_0+t} - B_{t_0}$ is Brownian motion. This follows immediately from checking the Brownian motion characterization above.

3.13

Let D_ρ be the origin-centered disc of radius ρ in $\mathbb{R}^2$. then $\mathbb{P}^0\{B_t \in D_\rho\} = \mathbb{P}\{\chi_2^2 \leq \rho^2/t\} = 1 - \exp\frac{-\rho^2}{2t}$

3.14

Prove that the Green measure on $\mathbb{R}^n$ with respect to B_t, $K \mapsto \mathbb{E}\int 1\{B_t \in K\}\,dt$ is absolutely continuous with respect to the Lebesgue measure.

The interior integral cannot be easily dominated, so interchange is difficult, but Fubini's Theorem merely requires σ-finite measures, which the Lebesgue measure dt is. Thus by Fubini,

$$\mathbb{E}\int 1\{B_t \in K\}\,dt = \int dt \int d\mathbb{P}\, 1\{B_t \in K\} = \int dt \int_K d\,((B_t)_* \mathbb{P}) = 0 \ ,$$

since the pushforward measure $((B_t)_* \mathbb{P})$ is diffuse.

3.15

UB_t is Brownian motion for unitary U, most easily shown here by the fact that all components remain independent normals.

3.16

For $c > 0$, $\frac{1}{c}B_{c^2 t}$ is Brownian motion, which follows immediately from checking the properties of Brownian motion. The scaling $1/c$ is squared so variance is preserved.

3.17

Define the p-th variation process for a continuous stochastic process X_t as

$$[X]_t^{(p)}(\omega) = \lim_{n \to \infty} \sum_{t_k^{(n)}} \left| X_{t_{k+1}^{(n)}}(\omega) - X_{t_k^{(n)}} \right|^p ,$$

where the limit is taken over any increasingly fine splits with $\delta_k^{(n)} = t_{k+1}^{(n)} - t_k^{(n)}$ where $0 = t_1^{(n)} \le t_2^{(n)} \le \cdots \le t_n^{(n)} = t$. Note that this construction intentionally permits duplicates. Call such sequences split sequences. $[X]_t^{(1)}$ is referred to as the total variation process and $[X]_t^{(2)} = [X]_t$ as the quadratic variation process.

Definition 3.1 (Shrinking Property). A split sequence $\left\{ t_k^{(2)} \right\}_{k=1}^2, \cdots, \left\{ t_k^{(n)} \right\}_{k=1}^n, \cdots$ is *shrinking* if

$$\lim_n \max_k \delta_k^{(n)} = 0 .$$

Note that by continuity on a compact interval $[0, t]$, $t \mapsto X_t(\omega)$ is uniformly continuous.

Lemma 1 (Partition Invariance). Fix ω and let $\Delta_k^{(n)}(t) = \left| X_{t_{k+1}}^{(n)}(\omega) - X_{t_k}^{(n)}(\omega) \right|$. Suppose a shrinking split sequence t exists such that $a_n = \sum_k \left| \Delta_k^{(n)}(t) \right|^p$ converges. Then any other shrinking split sequence s with $b_m = \sum_k \left| \Delta_k^{(m)}(s) \right|^p$ will also converge and $\lim_n a_n = \lim_m b_m$, so long as $p \ge 1$.

Proof. First, we fix any $\epsilon > 0$ and consider n large enough such that a_n is within its limit by ϵ.

By the shrinking property of b_m, for any $\delta > 0$, there exists M such that for all $m > M$, the split $\left\{ s_k^{(m)} \right\}$ is a δ-cover of $[0, t]$. Let the relabeling function $c : [n] \to [m]$ map the index k of each split point $t_k^{(n)}$ to the closest one k' of $s_{k'}^{(m)}$, choose c such that it is weakly increasing, which must be possible since both sequences are increasing. Then, for any $0 < \eta \le 1$, by choice of δ,

$$a_n = \sum_k \left| X_{t_{k+1}}^{(n)} - X_{t_k}^{(n)} \right|^p \le n\eta + \sum_k \left| X_{s_{c(k+1)}}^{(n)} - X_{s_{c(k)}}^{(n)} \right|^p \le n\eta + \sum_k \sum_{j=c(k)}^{c(k+1)-1} \left| X_{s_{j+1}}^{(n)} - X_{s_j}^{(n)} \right|^p \le n\eta + b_m ,$$

where the first inequality follows by choosing δ small enough such that the modulus of continuity of $X_t(\omega)$ is below $(\eta/2)^{1/p}$, the second inequality follows from triangle inequality (requiring $p \ge 1$), and the third adds in missing positive terms.

Choose $\eta \le \epsilon/n$. Thus by choice of M through δ we have for all $m > M$ that $a_n - \epsilon \le b_m$.

Since the sequence $\{a_{n'}\}_{n'}$ is itself shrinking, we may repeat this logic above starting from any b_m to find that for any $\epsilon' > 0$, for all sufficiently large $n' = n'(\epsilon')$, $b_m - \epsilon' \le a_{n'}$. Taking limits we find that $b_m \le \lim_{n'} a_{n'}$. Combining with the original property of n, we have that b_m must be within 2ϵ of $\lim_n a_n$ for all sufficiently large m.

$\square$

Given the above, we need only consider some fixed shrinking split sequence to analyze variation of a stochastic process. One surely exists: just use the unique evenly spaced split sequence.

Theorem 3 (Quadratic Variation of Brownian Motion). *For Brownian motion X_t,*

$$[X]_t^{(2)} = t .$$

Proof. Let $\Delta_k = \Delta_k^{(n)}(t)$ for $t_k^{(n)}$ to be the sequence with distance $\frac{t}{n-1}$ between each point. Then we will show $\sum_k \Delta_k \xrightarrow{L^2(\mathbb{P})} t$. Recall that convergence in L^p implies convergence in probability by Markov's inequality, and convergence in probability implies almost sure convergence of a subsequence by decreasing the probability of deviation exponentially and applying Borel-Cantelli. Then, some subsequence $\{n_m\}_m$ yields $\sum_k \Delta_k^{(n_m)}(t)$ which converges almost surely to both t and $[X]_t^{(2)}$, implying the latter are equal.

It remains to show the original $L^2(\mathbb{P})$ convergence. Indeed, leveraging independent increments, and letting $\sigma^2 = \frac{t}{n-1}$,

$$\mathbb{E}\left(\sum_k \Delta_k - t\right)^2 = \sum_k \mathbb{E}\,\Delta_k^4 + \sum_k \sum_{j \neq k} \mathbb{E}\,\Delta_k^2\,\mathbb{E}\,\Delta_j^2 - 2t \sum_k \mathbb{E}\,\Delta_k^2 + t^2$$
$$= 3n\sigma^4 + n(n-1)\sigma^4 - 2tn\sigma^2 + t^2$$
$$= O\left(\text{poly}(t)n^{-1}\right)\ ,$$

where critically all nonvanishing t^2 terms cancel. $\qquad\square$

Theorem 4 (Total Variation of Brownian Motion). *For Brownian motion X_t,*

$$[X]_t^{(1)} = \infty\ .$$

Proof. It suffices to show that $[X]_t^{(1)} < \infty$ implies $[X]_t^{(2)} = 0$. Then by the contrapositive the conclusion follows since Brownian motion has a positive quadratic variation.

Indeed, for any $\epsilon > 0$ if n is sufficiently large then $\left|\Delta_k^{(n)}(t)\right| \leq \epsilon$, holding across all k, by uniform continuity, since $\delta_k^{(n)}$ is sufficiently small. In turn $\sum_k \left|\Delta_k^{(n)}\right|^2 \leq \epsilon \sum_k \left|\Delta_k^{(n)}\right| = O(\epsilon)$ since the sum is itself $O(1)$ by assumption. Taking limits we have that the quadratic variation must be 0. $\qquad\square$