Randomness as Pointlessness

Greg Langmead

December 16, 2020

1   Introduction

Mathematicians have been exploring the advantages of locale theory for 60 or so years, along with its related disciplines of topos theory and type theory. When you tie one hand behind your back (such as avoiding the use of the existence of points, the axiom of choice, and/or the law of the excluded middle) then the constructions and proofs you invent can be more illuminating and deeper than before. I want to convince you that this may be true for algorithmic randomness notions. In [3], Alex Simpson has formulated a measure theory for locales and provided a new definition of randomness in this setting. I will describe this work, emphasizing the *point-free* interpretation.

Random real numbers satisfy the following hand-waving characterizations.

1. They cannot be obtained constructively.

2. They obey almost-everywhere properties.

3. They appear tied to the foundations of mathematics.

For a survey of these and other properties, see Rute [1]. I will argue that defining random points in terms of locales is well suited to capturing or even explaining all three of these properties. First of all, locales provide a setting for topology that supports constructive arguments, and randomness is in some sense the opposite of constructivity. Secondly, locales provide a model of topological spaces and continuous functions, so we can connect with classical results in analysis including almost-everywhere theorems like Lebesgue's theorem. And thirdly locales mediate a generalization from set theory to topos theory and type theory, modern constructive settings where

mathematical foundations are being explored. Constructions like the effective topos and assemblies capture computability in this general context. It seems inevitable that randomness will find enhanced definitions in these new foundations.

1.1 Randomness vs determinism: a confused debate

In the Stanford Encyclopedia of Philosophy's article on Chance and Randomness [2], section 5.2 ("Chaotic Dynamics"), the authors describe an iterated transformation on the unit square, somewhat analogous to repeatedly folding dough. Given a starting point $(p, q), 0 \leq p, q \leq 1$, define a function

$$\phi(p, q) = \begin{cases} (2p, q/2), & \text{if } 0 \leq p \leq \frac{1}{2} \\ (2p - 1, (1 + q)/2), & \text{if } \frac{1}{2} \leq p \leq 1 \end{cases}$$

As the authors say, "This corresponds to transforming the unit square to a rectangle twice as wide and half the height, chopping off the right half, and stacking it back on top to fill the unit square again." So, needlessly more violent than folding, but close. If we represent p and q in their binary expansion, we have

$$\phi(0.p_1 p_2 \ldots, 0.q_1 q_2 \ldots) = (0.p_2 p_3 \ldots, 0.p_1 q_1 q_2 \ldots).$$

In the article this example is intended to convey a paradox. On the one hand, the system is ergodic and behaves like a random process. Points will wander throughout the space without any pattern. But on the other hand the system is deterministic. The bit shift formula shows how it works: this is an engine that uses the next bit of the binary expansion to determine where the point moves next.

The resolution of the paradox, I believe, is to worry more about whether you can get your hands on p and q in the first place. Sure, once you have them, the machine behaves deterministically. But you can't have them! They require an infinite amount of information to specify, and to have one is to provide its entire future path through the process. The problem with the paradox is therefore in one of its assumptions. The process is deterministic, but the randomness is in the input data.

Indeed, the most important property of random numbers appears to be their infinite nature, which prevents us from ever having one fully specified

for us, either in a machine, in our minds, or in our theories. They are not computable, and they are not constructible.

Confusingly, there are several different definitions of algorithmic randomness. In fact there is a whole family (or if you're in a teasing mood, a zoo) of them. They go by names such as Martin-Löf randomness, Schnorr randomness, weak-n-randomness, and computable randomness. They are not equivalent but they form a simple chain of inclusions, and smart people have worked out that the differences can be explained and parameterized by a family of choices that relate them. What I will present here is a consideration that is orthogonal to the flavor of randomness, namely the *packaging* of the random numbers inside the reals. The main ingredient is a shift from *set theory* to *locale theory*. A locale is a generalization of a topological space, and I will argue that it is very well suited to a philosophical interpretation of randomness.

2   Frames and Locales

Locale theory is point-free topology. See Johnstone [4] for an accessible survey, or Chapter IX of Mac Lane and Moerdijk [5] for a pleasantly breezy but more thorough introduction, or Johnstone [6] chapter C1 for a detailed treatment. We can also call it *synthetic* topology in that it takes open sets as basic, undefined objects, instead of defining them as collections of points. Instead it is points that will become a derived concept. Topological spaces with their actual open sets of actual points provide examples of locales, but there are more general ones as well. This will give us a new point of view on randomness.

A locale is a *frame*. A frame is a partially ordered set (poset) with finite meets $\wedge$ and arbitrary joins $\vee$, with a maximal element $\top$ and minimal element $\bot$, and where the infinite distributive law holds:

$$x \wedge (\bigvee_i y_i) = \bigvee_i (x \wedge y_i).$$

This definition is clearly abstracted from the axioms for a topology: we are to imagine the elements of the poset are open sets of a generalized space. If we require only countable joins then it is called a σ-frame. The open sets of a topological space form a frame (or a σ-frame). A frame map is a function that preserves finite meets and arbitrary joins. A σ-frame map preserves

countable joins. We collect this data into the category Frm of frames and σFrm of σ-frames.

A continuous map between topological spaces induces a frame map on the frames of open sets, but in the opposite direction because the *inverse* image of an open is open. This motivates the definition of the category Loc of locales to be simply $\mathsf{Frm}^{\mathsf{op}}$ and $\sigma\mathsf{Loc} = \sigma\mathsf{Frm}^{\mathsf{op}}$. (Recall that the opposite of a category is a category with all the same objects and morphisms, but with the direction of all the morphism arrows reversed.) We can define a functor $L : \mathsf{Top} \to \mathsf{Loc}$ on objects by mapping a space X to its frame of open sets $\mathcal{O}(X)$ (so on objects, $L = \mathcal{O}$), and on morphisms by mapping a function $f : X \to Y$ to the opposite arrow of the induced morphism on frames, so that's two inversions giving us an arrow $L(f) : L(X) \to L(Y)$.

We'll use notation like $X, Y, \ldots$ to denote locales, and $\mathcal{O}(X), \mathcal{O}(Y), \ldots$ to denote frames of open sets.

3   Points

Consider the topological space $\star$ consisting of a single point. Its topology is just the poset $\{\bot \leq \top\}$ (the empty set and the whole space). To treat points as a derived concept, note that a point in a space X is equivalent to a map $\mathrm{pt} : \star \to X$, which at the level of frames is a map $\mathcal{O}(X) \to \{\top, \bot\}$. The inverse image of $\top$ is the collection of all open sets that contain the point $\mathrm{pt}(\star)$. This forms what is known as a *completely prime filter* in the frame, which is a collection P of opens such that if $x \vee y \in P$ then $x \in P$ or $y \in P$. (This is analogous to a prime ideal, or the ideal generated by a prime number.) We can then move forward with locales and completely prime filters instead of spaces and points!

Note how compatible this is with a constructive point of view. To supply a point you must supply a function $\star \to X$. The points aren't given, you have to do work to produce one.

Given a locale X, the set of completely prime filters constitutes the *spatial part* of X, and this set inherits a topology from the structure of X. Locale maps preserve prime filters so we obtain a set function between the spatial parts of the locales. We can package this as a functor $S : \mathsf{Loc} \to \mathsf{Top}$ which turns out to be right adjoint to L. If we go back and forth with $S \circ L$, starting from an honest topological space, then we get back an equivalent space with an equivalent topology so long as the original space is *sober*, which is a

definition that is almost circularly defined to meet this condition. Hausdorff spaces are sober. (The mathematicians who developed locale theory were very self-conscious about the names according to Peter Johnstone in [7] and it shows — they chose great names like "sober.")

We won't make use of S, but it's important to know about it to have a global picture of how spaces are embedded in the larger category of locales.

4   Sublocales

Mathematicians have been careful to choose the correct notion of subspace in this context: the sublocale. There are three equivalent ways to define sublocales. Normally I'd expect two since we are bouncing between two opposite categories. But there is a super interesting third definition as well, and the interplay of all three sheds light on the application to randomness.

4.1   Definition 1: Embedded subspaces

To me the most intuitive definition is the one that lets us think of a sublocale as being like an *embedded subspace*, i.e. a monomorphic map with the extra requirement that the subspace is embedded. In the category of topological spaces this means the domain is homeomorphic to its image. For example, for a map from $[0, 1]$ into a space X to be an embedding the curve must not self-intersect or do anything pathological. In categorical terms embeddings and sublocales are *regular monomorphisms* and these have special factorization properties. You can maybe already tell that this is a less permissive notion of subspace than subspaces in Top, where any old subset of a topological space (i.e. any monomorphism) can be equipped with the subspace topology and considered a subobject.

4.2   Definition 2: Quotients of frames

In the category of frames the arrows are reversed and so a sublocale is in fact a quotient frame, i.e. an equivalence relation. Intuitively, two open sets are equivalent if they agree on the sublocale. Precisely, a sublocale $=_Y$ is an

equivalence relation that preserves finite meets and arbitrary joins:

$$U =_Y V \text{ and } U' =_Y V' \implies U \wedge U' =_Y V \wedge V'$$

$$U_i =_Y V_i \text{ for all } i \geq 0 \implies \bigvee_{i \geq 0} U_i =_Y \bigvee_{i \geq 0} V_i.$$

4.3 Definition 3: Nuclei

The nucleus definition of a sublocale also takes place in the category of frames. It is just the surjective map corresponding to the quotient. A nucleus is a function $j : \mathcal{O}(X) \to \mathcal{O}(X)$ satisfying

$$U \leq jU$$
$$jjU = jU$$
$$j(U \wedge V) = j(U) \wedge j(V)$$

The intuition is that j takes an open set to some largest version of itself satisfying some property, which is why the first condition has the containment, and why the second condition is idempotence – doing the operation twice yields the same largest open. The third condition is that forming the "largest version" preserves finite intersection.

Denote the fixed points of j (i.e. the collection of largest sets) by $\mathcal{O}_j(X)$ and denote the dual object by X_j. Then X_j is the sublocale. The map j is a surjective frame map onto the set of fixed points, and so gives an embedding $i : X_j \hookrightarrow X$. This is basically performing the quotient from the previous definition, mapping every open onto the largest representative in the equivalence class.

4.4 The collection of all sublocales

The collection of all sublocales of a fixed locale X is partially ordered by inclusion, just as we have in any category where we can define subobjects. It's worth mentioning that even more is true: the collection of sublocales forms a *coframe*, which is duel to a frame, so is a poset with finite joins and arbitrary meets, and where the joins distribute over the meets. This collection is nowhere near as straightforward to contemplate as the collection of subsets of a set, but we will keep thinking through some of the consequences to find how it connects with randomness.

4.5   A pointless sublocale

We're ready to define a sublocale with no points! Let X be a Hausdorff topological space without isolated points and let $\mathcal{O}(X)$ be the frame of its open sets. Consider the double-negation operation $\neg\neg : \mathcal{O}(X) \to \mathcal{O}(X)$. What does this do? The negation of an open U is the largest open disjoint from U (or equivalently, the union of all opens disjoint from U). In an actual space such as X this is the same as the interior of the complement, but in a locale we don't have complements since they are not open. Repeating the operation, $\neg\neg U$ is thus the largest open disjoint from the largest open disjoint from U. For a lot of sets this is just U again but sometimes it erases information. For any $p \in X$ we have that $\neg\neg(X - \{p\}) = X$, because $\neg(X - \{p\}) = \emptyset$. So this operation blurs out the points.

$\neg\neg$ is idempotent, and in fact defines a nucleus hence an embedding $nn : X_{\neg\neg} \hookrightarrow X$, where $nn^{-1}U = \neg\neg U$. It's this sublocale $X_{\neg\neg}$ that has no points. For suppose $p : \star \to X_{\neg\neg}$ is a point. The composition $nn \circ p$ is thus a point of X. Consider the open set given by deleting this point from X, i.e. $X - \{nn(p(\star))\}$. Taking the inverse image in frames we have $p^{-1}(\neg\neg(X - \{nn(p(\star))\})) = \bot$ because only opens that contain the image of $\star$ map back to $\top$. But as we saw in the previous paragraph $\neg\neg(X - \{nn(p(\star))\}) = X$, so the above expression $p^{-1}(\neg\neg(X - \{nn(p(\star))\})$ becomes $p^{-1}(X)$ which is $\top$, a contradiction! You can't have points if you blurred away the points.

Furthermore, $X_{\neg\neg}$ is *dense* in X. For a sublocale $A \hookrightarrow X$ to be dense means that A intersects every open sublocale. (An open sublocale is one of the original open sets in X, promoted to a sublocale by embedding it in X.) This follows from being a nucleus: $\neg\neg A \supseteq A$. The definitions are piling up, aren't they?

4.6   Locale morphisms

A fact that makes the category of locales different from the category of spaces is that any two dense sublocales intersect to form another dense sublocale! If we intersect all dense sublocales we form a *smallest dense sublocale* of any locale X and it turns out to be exactly $X_{\neg\neg}$. By smallest I mean inside the poset of sublocales. None of this is the case with topological spaces such as $\mathbb{R}$ which can be decomposed into disjoint dense sub*sets* such as $\mathbb{Q}$ and $\mathbb{R}\backslash\mathbb{Q}$. This is confusing! What does the sublocale $\mathbb{Q} \hookrightarrow \mathbb{R}$ look like then? The consequences for measure theory are significant, since defining measures on

sets leads to some pathological examples involving cosets of the rationals.

Let's look at the *image* of a sublocale inclusion. For this we need the "push-forward" map $f_* : \mathcal{O}(X) \to \mathcal{O}(Y)$, which is actually right adjoint to the inverse image map $f^{-1} : \mathcal{O}(Y) \to \mathcal{O}(X)$. It is defined as follows:

$$f_*(U) = \bigcup_{f^{-1}(V) \subseteq U} V.$$

In other words, an open set $U \subseteq X$ maps under f_* to the *largest* open set in Y that pulls back to some subset of U.

This is an elementary operation but not one that comes up in a standard undergraduate topology course (at least a 20th century one!). Here's an example that may help: imagine $f : [0,1] \to \mathbb{R}^2$ is some smooth curve in the plane. Given an open interval $(a,b) \subseteq [0,1]$, its image in the plane is a one-dimensional subset, so won't itself be open. Instead to compute $f_*((a,b))$ we have to find the largest open set in the plane whose inverse image is (a,b), which will look like *all of* $\mathbb{R}^2$ minus two skinny pieces of the curve that are the images of $[0,a]$ and $[b,1]$. So it has the image in it, plus all the rest of the plane that isn't in the image of the rest of $[0,1]$.

The requirement that f_* give the *largest* open set such that some condition holds should remind us of a nucleus. In fact given a sublocale $f : X \to Y$ then $f_* \circ f^{-1}$ is the nucleus for the sublocale! It maps an open set to the direct image of its inverse image, which will be the largest open that contains the one you started with, and which itself also maps back to the same open in X.

Now let's get back to considering the inclusion $i : \mathbb{Q} \hookrightarrow \mathbb{R}$. The push-forward map i_* sends the open set "all of $\mathbb{Q}$" in $\mathbb{Q}$ to the largest open in $\mathbb{R}$ whose inverse image is contained in $\mathbb{Q}$. But $\mathbb{Q}$ is dense so this is all of $\mathbb{R}$! There's no smaller open that contains the rationals! The requirement that this push-forward be an open set is forbidding us from carving out these pathological dense subsets in the category of locales. The topological glue of the real numbers is constraining what kind of subobjects we have access to.

5 Measure theory

We have prepared all the abstract background. We have a new setting for topology called locale theory, which has broken away from the concept of a point. We have seen that the monomorphisms are different than in the

8

category of topological spaces, and we have examples of sublocales that have no points. Let's bring measure theory into the mix and define randomness in this setting.

Technically we are back to requiring the "$\sigma-$" on the words "frame", "locale" and "sublocale" so that we can reconnect with the countability requirements of measures. On the other hand many spaces we care about are "strongly Lindelöf" and we can drop the $\sigma-$. Simpson works out the details in [3]. To be strongly Lindelöf means that for any collection of open sets, their union is contained in a *countable* union of open sets. This includes $\mathbb{R}$. For the rest of the essay we'll assume we are dealing with strongly Lindelöf spaces and we'll drop the $\sigma-$, but of course the full theory can be more general if we want.

We will define a measure on a locale/frame and then extend it to all the sublocales of the frame.

A measure on a frame is a function $\mu : F \to [0, \infty]$ satisfying

$$\mu(\bot) = 0$$

$$x \leq y \Longrightarrow \mu(x) \leq \mu(y)$$

$$\mu(x) + \mu(y) = \mu(x \vee y) + \mu(x \wedge y) \quad (\text{"modularity"})$$

$$(x_i)_{i \geq 0} \text{ ascending} \Longrightarrow \mu(\bigvee_{i \geq 0} x_i) = \sup_{i \geq 0} \mu(x_i) \quad (\text{"continuity"}).$$

We say μ is a probability measure if $\mu(\top) = 1$. We will also call μ a measure on a locale if it is a measure on the corresponding frame of opens.

To bring the concept of measure to sublocales we will generalize the concept of outer measure. We say a locale is *fitted* if for every sublocale $Y \subseteq X$ it holds that $Y = \bigwedge \mathcal{N}(Y)$ where $\mathcal{N}(Y) = \{U \in \mathcal{O}(X) | Y \subseteq U\}$, the family of opens containing Y. We extend the measure μ to a measure μ^* on any sublocale by $\mu^*(Y) = \inf_{U \in \mathcal{N}(Y)} \mu(U)$, i.e. the limit of the measure of the open sets that intersect to Y. There is work to show that spaces we care about are fitted, and that this definition really defines a measure (see section 5 of [3]).

We can now cash in the fact that we cannot decompose locales into two dense sublocales. Simpson [3] proves in Example 4.7 that the standard measure on open subsets of $\mathbb{R}^n$, which extends to a measure on sublocales in the way we just defined, is preserved under Euclidean translations. This measure agrees with Lebesgue measure whenever the sublocale agrees with a measurable set. In standard measure theory there is no translation invariant measure

9

on arbitrary subsets of $\mathbb{R}^n$ due to pathological examples like cosets of rational numbers, which of course are automatically avoided in locale theory.

5.1   The random sublocale

We are ready to define the random sublocale Ran_X of an arbitrary locale X equipped with a probability measure μ. We'll do it three times, once for each definition of sublocale.

The spatial definition of Ran_X is the smallest sublocale of measure 1, or equivalently the intersection of all sublocales of measure 1. This tracks exactly with the definition of Kurtz randomness seen in the algorithmic randomness literature. It's morally similar to $X_{\neg\neg}$ but making use of measure. $X_{\neg\neg}$ is the smallest dense sublocale of X, whereas Ran_X is the smallest measure-1 sublocale.

The quotient definition of Ran_X is the equivalence relation $=_{\mathrm{Ran}_X}$ given by $U =_{\mathrm{Ran}_X} V \Longleftrightarrow \mu(U) = \mu(U \cap V) = \mu(V)$. Opens are equivalent if they cover the same piece of space with equal measure.

The nucleus definition is not mentioned by Simpson, but obviously it's given by $j_{\mathrm{Ran}_X}(U) = \bigcup_{V \supseteq U, \mu(U)=\mu(V)} V$. In other words, it maps U to the maximal open that contains U and has the same measure as U.

This sublocale has no points! Given a point $p \in X$, the set $X - \{p\}$ has measure 1 and so $\mathrm{Ran}_X \subseteq X - \{p\}$ and so p is not present in the random sublocale. Equating sets blurs away the points like $\neg\neg$ does.

6   Conclusion: open sets vs randomness

Topology is about open sets, and open sets correspond to semideciadable properties, i.e. properties that can be verified in finite time, but might need infinite time to refute. This is philosophically adjacent to being computable and constructive. In the Cantor space of binary sequences, the open sets are unions of cylinders, which are finite prefixes of sequences. We can verify propositions about infinite sequences just when we can do it in finite time. The finiteness is the openness and vice versa.

Locales take just the open sets from topology and build from there, without allowing us to have points unless we can construct them.

We've seen that locales behave very differently! There have smallest dense subsets so you can't decompose the reals into a disjoint union of rational and

irrational. The topological glue constrains you from doing this.

Measure theory can be done on locales and it avoids some pathological problems that set theory causes. And we can form a smallest full-measure sublocale, "random sublocale." This sublocale has no points.

So if we revisit our set theoretic foundations and consider open sets to be primary, then we arrive at a definition of randomness where the randoms constitute a space with full measure but no points. This makes perfect sense once you appreciate that having points requires work to produce one, but random points can never be fully produced.

This simple presentation gives the appearance, at least to me, that locale theory permits only one definition of randomness, which if true would be a major disadvantage. Algorithmic randomness contains a few moving parts such as whether certain measures can be computably obtained, or whether convergence of a limit is computable or can be computably bounded. Can we gain control over analogous parts of the theory within locale theory? Or are we stuck with Kurtz randomness? There are indications that the Borel hierarchy of sets can be imported into locale theory[1] from which we could obtain weak n-randomness, so that would be the logical next step to flesh out.

Perhaps there is another option, which is that locale theory may not by itself offer control over all the extra choices contained in the theory of algorithmic randomness, but that passing from locales to the theory of assemblies in a realizability topos may offer them instead. Topos theory is built on top of locale theory, so this would be like continuing the narrative. In such a case locales would offer the unified big picture of the origins of randomness, which are then exported to different implementations.

Lastly let me speculate on the role of almost-everywhere properties in locale theory. I imagine we could replace statements like

> $x \in [0, 1]$ is Schnorr random if and only if every computable function f of bounded variation with effectively integrable derivative f' is differentiable at x

with

> If $f : [0, 1] \to \mathbb{R}$ is a morphism in Loc, then f is differentiable on $\mathrm{Ran}_{[0,1]}$.

---

[1]See slides by Simpson at `https://ccc2017.loria.fr/slides/ccc_simpson.pdf`.

We could perhaps drop the explicit requirement of computability because we'd be in a constructive framework already, and instead of stating the result pointwise we'd have a global statement. It remains to properly define smoothness in the locale framework and to work out many details!

References

[1] Jason Rute. On the close interaction between algorithmic randomness and constructive/computable measure theory, 2019.

[2] Antony Eagle. Chance versus Randomness. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition, 2019.

[3] Alex Simpson. Measure, randomness and sublocales. *Annals of Pure and Applied Logic*, 163(11):1642 – 1659, 2012. Kurt Goedel Research Prize Fellowships 2010.

[4] Peter T. Johnstone. The point of pointless topology. *Bull. Amer. Math. Soc. (N.S.)*, 8(1):41–53, 01 1983.

[5] Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: a first introduction to topos theory*. Universitext. Springer, Berlin, 1992.

[6] Peter T Johnstone. *Sketches of an elephant: a Topos theory compendium*. Oxford logic guides. Oxford Univ. Press, New York, NY, 2002.

[7] Peter Johnstone. *Elements of the History of Locale Theory*, pages 835–851. Springer Netherlands, Dordrecht, 2001.