Synthetic Probability Theory

Alex Simpson

Faculty of Mathematics and Physics
University of Ljubljana, Slovenia

Categorical Probability and Statistics
8 June 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 731143

Synthetic probability theory?

In the spirit of synthetic differential geometry (Lawvere, Kock, . . .)

Axiomatise contingent facts about probability as it is experienced, rather than deriving probabilistic results as necessary consequences of set-theoretic definitions that have a tenuous relationship to the concepts they are formalising.

A main goal is to provide a single set of axioms that suffices for developing the core constructions and results of probability theory.

I believe the approach has the potential to provide a simplification of textbook probability theory.

Gian-Carlo Rota (1932-1999):

> *" The beginning definitions in any field of mathematics are always misleading, and the basic definitions of probability are perhaps the most misleading of all. "*

Twelve Problems in Probability Theory No One Likes to Bring Up,
The Fubini Lectures, 1998 (published 2001)

The definition of "random variable"

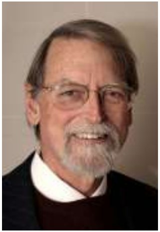
An A-valued random variable is:

$$X : \Omega \to A$$

where:

- the value space A is a measurable space (set with σ-algebra of measurable subsets);
- the sample space Ω is a probability space (measurable space with probability measure $\mathbf{P}_\Omega$); and
- X is a measurable function.

David Mumford:



“ The basic object of study in probability is the random variable and I will argue that it should be treated as a basic construct . . . and it is artificial and unnatural to define it in terms of measure theory. ”

The Dawning of the Age of Stochasticity, 2000

Approach of talk

Present an axiomatisation of random variables in terms of their interface (what one can do with them) rather than by means of a concrete set-theoretic implementation.

General setting:

- We work axiomatically with the category **Set** of sets in one of: set theory (allowing atoms) / type theory / topos theory.
- The underlying logic is classical.
- We assume the axiom of dependent choice (DC) but not the full axiom of choice.

We formulate the axioms in the most convenient form for fuss-free probability theory (e.g., avoiding fussing over measurability).

Functions act on random variables

Axiom:

- For every set A there is a set $\mathrm{RV}(A)$ of A-valued random variables.
- For every function $f\colon A \to B$ and random variable $X \in \mathrm{RV}(A)$ there is an associated

$$f(X) \in \mathrm{RV}(B) .$$

  Moreover,

$$\mathrm{id}(X) = X \qquad (g \circ f)(X) = g(f(X)) .$$

Equivalently: We have a functor $\mathrm{RV}\colon \mathbf{Set} \to \mathbf{Set}$.

Random variables have probability laws

Axiom:

- Every $X \in \mathrm{RV}(A)$ has an associated law $\mathbf{P}_X \in \mathcal{M}_1(A)$, where:

  $$\mathcal{M}_1(A) \;=\; \{\mu \colon \mathcal{P}(A) \to [0,1] \mid \mu \text{ is a probability measure}\}\;\;.$$

  Here $\mathcal{P}(A)$ is the full powerset.

- For every $f \colon A \to B$ and random variable $X \in \mathrm{RV}(A)$ we have $\mathbf{P}_{f(X)} = f_*(\mathbf{P}_X)$, where $f_*(\mu) \in \mathcal{M}_1(B)$ is the pushforward probability measure $f_*(\mu)(B') := \mu(f^{-1}B')\,.$

Equivalently: We have a natural transformation $\mathbf{P} \colon \mathrm{RV} \Rightarrow \mathcal{M}_1$

Probability for individual random variables

The equality in law relation for $X, Y \in \mathrm{RV}(A)$

$$X \sim Y \quad \Leftrightarrow \quad \mathbf{P}_X = \mathbf{P}_Y$$

$X \in \mathrm{RV}(\mathbb{R})$ is said to be integrable if it has finite expectation:

$$\mathbf{E}(X) \; := \; \int_x x \, \mathrm{d}\mathbf{P}_X$$

Similarly, define variance, moments, etc.

Families of random variables

Giving a finite or countably infinite family of random variables is equivalent to giving a random family.

Axiom: For every $(X_i \in \mathrm{RV}(A_i))_{i \in I}$ with I countable, there exists a unique $Z \in \mathrm{RV}(\prod_{i \in I} A_i)$ such that $X_k = \pi_i(Z)$ for every $k \in I$, where $\pi_k \colon (\prod_{i \in I} A_i) \to A_k$ is the projection.

Equivalently: RV preserves countable (including finite) products.

Notation: For notational convenience we work as if the canonical isomorphism $\mathrm{RV}(\prod_{i \in I} A_i) \cong \prod_{i \in I} \mathrm{RV}(A_i)$ is equality.
(E.g., we write $(X_i)_i$ for Z above.)

Independence

Independence between $X \in \mathrm{RV}(A)$ and $Y \in \mathrm{RV}(B)$:

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad \forall A' \subseteq A, B' \subseteq B \quad \mathbf{P}_{(X,Y)}(A' \times B') \; = \; \mathbf{P}_X(A') \, . \, \mathbf{P}_Y(B')$$

Mutual independence

$$\perp\!\!\!\perp X_1, \ldots, X_n \quad \Leftrightarrow \quad \perp\!\!\!\perp X_1, \ldots, X_{n-1} \text{ and } (X_1, \ldots, X_{n-1}) \perp\!\!\!\perp X_n$$

Infinite mutual independence

$$\perp\!\!\!\perp (X_i)_{i \geq 1} \quad \Leftrightarrow \quad \forall n \geq 1. \; \perp\!\!\!\perp X_1, \ldots, X_n$$

Restriction of random variables

Random variables restrict to probability-1 subsets.

Restriction axiom:

Given $Y \in \mathrm{RV}(B)$ and $A \subseteq B$ with $\mathbf{P}_Y(A) = 1$, there exists (a necessarily unique) $X \in \mathrm{RV}(A)$ such that $Y = i(X)$, where $i \colon A \to B$ is the inclusion function.

An extensionality principle

Equality of random variables is almost sure equality.

Proposition (Extensionality)

For $X, Y \in \mathrm{RV}(A)$:

$$X = Y \quad \Leftrightarrow \quad \begin{array}{ll} \mathbf{P}_{(X,Y)}\{(x, y) \mid x = y\} = 1 & \text{(official notation)} \\ \mathbf{P}(X = Y) = 1 & \text{(informal notation)} \end{array}$$

Corollary Given $X, X' \in \mathrm{RV}(A)$ and $A \subseteq B$, $i(X) = i(X')$ implies $X = X'$.

The uniqueness of the random variable X whose existence is postulated in the restriction axiom follows.

Proof of extensionality

Proof of interesting (right-to-left) implication

Suppose $X, Y \in \mathrm{RV}(A)$ satisfy

$$\mathbf{P}_{(X,Y)}(D) \;=\; 1 \;,$$

where $D := \{(x,y) \in A \times A \mid x = y\}$.

By restriction, there exists $Z \in \mathrm{RV}(D)$ such that $i(Z) = (X, Y)$, where $i\colon D \to A \times A$ is the inclusion function.

Then

$$\begin{aligned} (\pi_1 \circ i)(Z) &= \pi_1(X,Y) &= X \\ (\pi_2 \circ i)(Z) &= \pi_2(X,Y) &= Y \end{aligned}$$

Since $\pi_1 \circ i = \pi_2 \circ i\colon D \to A$, it follows that $X = Y$. $\square$

Categrory-theoretic formulation of restriction

Restriction category-theoretically:

If $m\colon A \to B$ is a monomorphism then the naturality square below is a pullback.

$$\begin{tikzcd} \mathrm{RV}(A) \arrow[r, "X \mapsto \mathbf{P}_X"] \arrow[d, "\mathrm{RV}(m)"'] & \mathcal{M}_1(A) \arrow[d, "\mathcal{M}_1(m)"] \\ \mathrm{RV}(B) \arrow[r, "Y \mapsto \mathbf{P}_Y"] & \mathcal{M}_1(B) \end{tikzcd}$$

Proposition: The functor $\mathrm{RV}\colon \mathbf{Set} \to \mathbf{Set}$ preserves equalisers.

Existence of random variables

Proposition (Deterministic RVs)
For every $x \in A$ there exists a unique random variable $\delta_x \in \mathrm{RV}(A)$ satisfying, for every $A' \subseteq A$:

$$\mathbf{P}_{\delta_x}(A') \;=\; \begin{cases} 1 & \text{if } x \in A' \\ 0 & \text{otherwise} \end{cases}$$

We write δ for the function $x \mapsto \delta_x \colon A \to \mathrm{RV}(A)$.

Axiom (Fair coin)
There exists $K \in \mathrm{RV}\{0,1\}$ with $\mathbf{P}_K\{0\} = \frac{1}{2} = \mathbf{P}_K\{1\}$.

Existence of independent random variables

The independence axiom

For every $X \in \mathsf{RV}(A)$ and $Y \in \mathsf{RV}(B)$, there exists $X' \in \mathsf{RV}(A)$ such that:

$$X' \sim X \quad \text{and} \quad X' \perp\!\!\!\perp Y \; .$$

Proposition For every random variable $X \in \mathrm{RV}(A)$ there exists an infinite sequence $(X_i)_{i \geq 0}$ of mutually independent random variables with $X_i \sim X$ for every X_i.

Proof

Let $X_0 = X$.

Given $X_0, \ldots, X_{i-1}$, the independence axiom gives us X_i with $X \sim X_i$ such that $X_i \perp\!\!\!\perp (X_0, \ldots, X_{i-1})$.

This defines the required sequence $(X_i)_{i \geq 0}$ by DC. $\square$

By the proposition there exists an infinite sequence $(K_i)_{i \geq 0}$ of independent random variables identically distributed to the fair coin K.

Laws of large numbers

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbf{P}\left(\left| \left(\frac{\sum_{i=0}^{n-1} K_i}{n} \right) - \frac{1}{2} \right| < \epsilon \right) \;=\; 1 \qquad \text{(weak)}$$

$$\mathbf{P}\left(\lim_{n\to\infty} \left(\frac{\sum_{i=0}^{n-1} K_i}{n} \right) = \frac{1}{2} \right) \;=\; 1 \qquad \text{(strong)}$$

Everything thus far, up to and including the formulation of the weak law, only uses the preservation of finite products by RV. The formulation of the strong law, however, makes essential use of the preservation of countably infinite products to define:

$$\lambda \;:=\; \mathbf{P}_{(K_i)_i} \;\in\; \mathcal{M}_1(\{0,1\}^{\mathbb{N}})$$

The near-Borel axiom

A standard Borel space is a set A together with a σ-algebra $\mathcal{B} \subseteq \mathcal{P}(A)$ that arises as the σ-algebra of Borel sets with respect to some complete separable metric space structure on A.

Let $(A, \mathcal{B})$ be a standard Borel space. We say that a probability measure $\mu \in \mathcal{M}_1(A)$ is near Borel if: for every $A' \subseteq A$ there exists $B \in \mathcal{B}$ such that $\mu(A' \Delta B) = 0$.

We say that $\mu \in \mathcal{M}_1(A)$ is an RV-measure if there exists $X \in \mathrm{RV}(A)$ with $\mathbf{P}_X = \mu$.

Axiom Every RV-measure on a standard Borel space is near Borel.

(If one assumes all subsets of $\mathbb{R}$ are Lebesgue measurable then *every* $\mu \in \mathcal{M}_1(A)$ is near Borel. I prefer the axiom above, as I believe its consistency does not require an inaccessible cardinal.)

Relating RV and Borel measures

Proposition (Raič & S.) Suppose μ, ν are RV-measures on a standard Borel space $(A, \mathcal{B})$. The following are equivalent.

- $\mu(B) = \nu(B)$ for all $B \in \mathcal{B}$.
- $\mu = \nu$.

Corollary The measure $\lambda \in \mathcal{M}_{\mathsf{RV}}(\{0,1\}^{\mathbb{N}})$ is translation invariant.
(We write $\mathcal{M}_{\mathsf{RV}}(A)$ for the set of RV-measures on A.)

Proposition Every Borel probability measure $\mu_{\mathcal{B}} \colon \mathcal{B} \to [0,1]$ on a standard Borel space $(A, \mathcal{B})$ extends to a unique $\mu \in \mathcal{M}_{\mathsf{RV}}(A)$.

Towards conditional expectation

In standard probability theory, conditional expectation takes the form $\mathbf{E}(X \mid \mathcal{F})$, where

- $\mathcal{F}$ is a sub-σ-algebra of the underlying σ-algebra on the sample space Ω.
- The characterising (up to almost sure equality) properties of $\mathbf{E}(X \mid \mathcal{F})$ include $\mathcal{F}$-measurability.

We have no sample space Ω!

- We condition with respect to other random variables $\mathbf{E}(X \mid Y)$. (In our setting, this is general enough.)
- The measurability condition is replaced by functional dependency.

Conditional expectation

We say that $Z \in \mathrm{RV}(B)$ is functionally dependent on $Y \in \mathrm{RV}(A)$ (notation $Z \leftarrow Y$) if there exists $f : A \to B$ such that $Z = f(Y)$.

Proposition

For $Y \in \mathrm{RV}(A)$ and integrable $X \in \mathrm{RV}(\mathbb{R})$, there exists a unique integrable random variable $Z \in \mathrm{RV}(\mathbb{R})$ satisfying:

- $Z \leftarrow Y$, and
- for all $A' \subseteq A$

$$\mathbf{E}(Z \,.\, \mathbf{1}_{A'}(Y)) = \mathbf{E}(X \,.\, \mathbf{1}_{A'}(Y))$$

The unique such Z defines the conditional expectation $\mathbf{E}(X \mid Y)$.

Conditional probability

For $X \in \mathrm{RV}(A)$, $Y \in \mathrm{RV}(B)$ and $A' \subseteq A$ define:

$$\mathbf{P}(X \in A' \mid Y) \;:=\; \mathbf{E}(\mathbf{1}_{A'}(X) \mid Y)\;.$$

Conditional independence

For $X \in \mathrm{RV}(A)$, $Y \in \mathrm{RV}(B)$ and $Z \in \mathrm{RV}(C)$ define:

$$X \perp\!\!\!\perp Y \mid Z \;\Leftrightarrow\; \text{for all } A' \subseteq A,\ B' \subseteq B$$
$$\mathbf{P}((X, Y) \in A' \times B' \mid Z) = \mathbf{P}(X \in A' \mid Z)\,.\,\mathbf{P}\,(Y \in B' \mid Z)\;.$$

Universality of λ RVs

Every random variable is functionally dependent on some $\{0,1\}^{\mathbb{N}}$-valued random variable with law λ .

Axiom: For every $Y \in \mathrm{RV}(A)$ there exist a random variable $X \in \mathrm{RV}(\{0,1\}^{\mathbb{N}})$ with $\mathbf{P}_X = \lambda$ such that $Y \leftarrow X$.

God tosses coins!

Regular conditional probabilities

For $X \in \mathrm{RV}(A)$ and $Y \in \mathrm{RV}(B)$ a regular conditional probability (rcp) for Y conditioned on X is a random variable $Z \in \mathrm{RV}(\mathcal{M}_{\mathrm{RV}}(B))$ such that:

- $Z \leftarrow X$
  (so Z is induced from X by an RV-kernel $A \to \mathcal{M}_{\mathrm{RV}}(B)$)
- For every $B' \subseteq B$,

$$Z(B') \;=\; \mathbf{P}(Y \in B' \,|\, X) \;,$$

  where $Z(B') \in \mathrm{RV}[0,1]$ abbreviates $(\mu \mapsto \mu(B'))(Z)$.

Theorem For every pair of random variables X, Y, there exists a unique rcp for Y conditioned on X. We write $P_{Y|X}$ for this rcp.

From kernels to RVs

The previous theorem takes us from pairs of random variables to RV-kernels. Conversely we have:

Theorem

Suppose $k\colon A \to \mathcal{M}_{\mathrm{RV}}(B)$ is an RV-kernel where $|B| \leq 2^{\aleph_0}$. Then, for any $X \in \mathrm{RV}(A)$, there exists $Y \in \mathrm{RV}(B)$ such that:

$$P_{Y|X} \ = \ k(X) \ .$$

Simple illustrative application:

Using the RV-kernel $(\mu, \sigma) \mapsto \mathcal{N}_{\mu,\sigma^2}\colon \mathbb{R}^2 \to \mathcal{M}_{\mathrm{RV}}(\mathbb{R})$, we obtain for any $M, S \in \mathrm{RV}(\mathbb{R})$ a random variable Z such that

$$P_{Z\,|\,M,S} = \mathcal{N}_{M,S^2} \quad (\text{in statistician's notation } Z \sim \mathcal{N}_{M,S^2})$$

Existence of conditionally independent RVs

Proposition

For every $X \in \mathrm{RV}(A)$, $Y \in \mathrm{RV}(B)$ and $Z \in \mathrm{RV}(C)$, there exists $X' \in \mathrm{RV}(A)$ such that:

$$(X', Z) \sim (X, Z) \quad \text{and} \quad X' \perp\!\!\!\perp Y \mid Z\ .$$

Towards stochastic processes: a myth

David Williams:

> " At the level of this book, the theory would be more elegant if we regarded a random variable as an *equivalence class* of measurable functions, two functions belonging to the same equivalence class if and only if they are equal almost everywhere. . . . [In the] more interesting, and more important, theory where the parameter set of our process is uncountable . . . the equivalence class formulation just will not work . . . it loses the subtlety which is essential even for formulating the fundamental results on the existence of continuous modifications, etc. "

Probability with Martingales, 1990

Stochastic processes

Traditional probability theory

For $T \subseteq \mathbb{R}$, a T-indexed stochastic process is given by

$$\Omega \times T \longrightarrow \mathbb{R}$$

(measurable in the first argument)

Synthetic probability theory

We have no Ω, and we have $\mathrm{RV}(A)$ as a replacement for A^Ω.

There are thus two natural options for T-indexed stochastic processes:

$$\mathrm{RV}(\mathbb{R})^T \qquad\qquad \mathrm{RV}(\mathbb{R}^T)$$

The second is the useful choice!

For $T \subseteq \mathbb{R}$, a T-indexed stochastic process is a random variable

$$X_T \in \mathrm{RV}(\mathbb{R}^T) \ .$$

If $S \subseteq T$ then we use

$$(f \mapsto \lambda s.\, f(s)) \colon \mathbb{R}^T \to \mathbb{R}^S$$

to define

$$X_S \ := \ (f \mapsto \lambda s.\, f(s))\,(X_T) \in \mathrm{RV}(\mathbb{R}^S) \ .$$

For $t \in T$ we define

$$X_t \ := \ (f \mapsto f(t))\,(X_T) \in \mathrm{RV}(\mathbb{R}) \ .$$

Consider the map.

$$\mathrm{RV}\left(\mathbb{R}^T\right) \xrightarrow{X_T \mapsto (X_t)_{t\in T}} (\mathrm{RV}(\mathbb{R}))^T$$

Given X_T, Y_T we have, by extensionality,

$$X_T = Y_T \;\Leftrightarrow\; \mathbf{P}(X_T = Y_T) = 1$$

This says that X_T and Y_T are indistinguishable. Similarly,

$$(X_t)_{t\in T} = (Y_t)_{t\in T} \;\Leftrightarrow\; \forall t\ \mathbf{P}(X_t = Y_t) = 1$$

This says that X_T and Y_T are modifications of each other.

When T is a continuum, there exist distinguishable processes that are modifications of each other.

$\mathrm{RV}\colon \mathbf{Set} \to \mathbf{Set}$ does not preserve arbitrary products!

Example definitions (martingale, Markov process)

$X_T \in \mathrm{RV}(\mathbb{R}^T)$ is a martingale if for every $s < t \in T$

$$\mathbf{E}(X_t \mid X_{\leq s}) \; = \; X_s \;\; ,$$

where $\leq s := \{s' \in T \mid s' \leq s\}$

$X_T \in \mathrm{RV}(\mathbb{R}^T)$ has the Markov property if for every $s \in T$

$$P_{X_{>s} \mid X_{\leq s}} \leftarrow X_s \;\; ,$$

where $> s := \{s' \in T \mid s' > s\}$.

Brownian motion — completely standard!

$B_T \in \mathrm{RV}(\mathbb{R}^T)$, where $T = [0, \infty)$, is a Brownian motion if:

- $B_0 = 0$;
- B_T has independent increments; i.e., for all $0 \leq t_0 < \cdots < t_n$

$$\mathop{\perp\!\!\!\perp}_{1 \leq i \leq n} B_{t_i} - B_{t_{i-1}} \;\; ;$$

- B_T has stationary normal increments; i.e., for all $s, t \geq 0$

$$(B_{s+t} - B_s) \sim \mathcal{N}_{0,t} \;\; ;$$

- $\mathbf{P}(B_T \text{ is continuous}) = 1\,.$

Construction of Brownian motion

Theorem A Brownian motion $B_{[0,\infty)}$ exists.

Proof outline

Use the existence of conditionally independent RVs and DC to iteratively construct a process $B' \in \mathrm{RV}(\mathbb{R}^{[0,\infty)\cap\mathbb{Q}_d})$ satisfying the conditions of Brownian motion, but indexed by dyadic rationals.

Prove that this dyadic-rational-indexed process is almost surely continuous at all real $t \in [0,\infty)$. Thus B' restricts to a random variable on the set

$$\{f \in \mathbb{R}^{[0,\infty)\cap\mathbb{Q}_d} \mid f \text{ is continuous at all } t \in [0,\infty)\} \ .$$

Now apply the function that maps each such f to its unique continuous extension in $\mathbb{R}^{[0,\infty)}$. $\square$

Equality and equivalence

There are two equivalence relations of interest on random variables.

- Almost sure equality — in our setting this is just equality.
  This satisfies the usual (internal) substitutivity laws.

- The weaker equivalence relation: equality in law $\sim$.
  This satisfies a meta-theoretic substitutivity law.

The invariance axiom

All definable properties are equidistribution invariant.

Axiom (schema)

Every sentence of the form

$$\forall X, Y \in \mathrm{RV}(A), \quad \Phi(X) \;\wedge\; X \sim Y \;\rightarrow\; \Phi(Y)$$

is true.

There is no evil!

Ongoing and future work

Prove consistency of the axioms. (I have a candidate sheaf model.)

Develop substantial portions of probability theory in detail.

Transfer theorems.

Constructive and (hence) computable versions.

Type-theoretic formalised probability theory.

"Bayesian variables" instead of random variables?

A convenient category for higher-order probability theory: **Set** !

Where are the monads?

RV is not a monad (I believe)

$\mathcal{M}_1$ is a monad, but I don't know if it is commutative.

Integration w.r.t. RV-measures satisfies the Fubini property. But I don't know if $\mathcal{M}_{\mathsf{RV}}$ forms a monad.

Challenge: Find a model combining:

- cartesian closed with countable limits and colimits;
- Fubini's theorem for integration w.r.t. probability measures;
- infinite product measures $\bigotimes : (\prod_{n \geq 0} \mathsf{M} X_n) \to \mathsf{M}(\prod_{n \geq 0} X_n)$, where MX is the object of "probability measures":;
- M is a monad.