Probabilities as Shapes

Harry Crane

Department of Statistics
Rutgers

May 3, 2017

Guiding questions

1. What is the essence of probability?
2. Does probability as **number** capture the essence of probability?
3. Is **number** essential for this purpose?

I claim:

- Numbers do not capture the essence of probability.
- Numbers are not necessary for expressing probability as we wish to use it in science and statistics.
- Standard formalism of probability is misaligned with the man on the street's intuition about what it means for something 'to be probable'.
- Possible source of miscommunication and confusion between statisticians and everyone else.

A thought:

Statistical analyses are often misunderstood because they often don't make sense.

Probability as Shape

Probability is the very guide to life. (Joseph Butler)

- **Philosophy**:
  - Experience **shapes** worldview which guides judgments about truth and probability.
  - How probabilities take shape (experience) determines the shape that probabilities take (judgments).
  - Probabilities are shapes.
- **Pre-formal**: A **shape** is a collection of points and paths between points.
- **Formal**: A **shape** is a homotopy type, i.e., a topological space 'up to homotopy'.
  - D. Tsementzis. (2017). A Meaning Explanation for HoTT. PhilSci Archive:12824.
- **Probability**: '*A* is probable' if there is **evidence** supporting *A*.

The **probability of** *A is the body of evidence favoring A.*

A **is probable** *if we possess a piece of such evidence.*

Relations between probabilities and the evidence comprising them give probabilities **shape**. *(worldview interpretation)*

Whether something counts as evidence for A depends on how it 'fits into' this shape (e.g., scientific theory, past experience, etc.)

Outline

How does the man on the street think about probability?

1. Probability as Number
2. Probability as Shape

Chicken and Egg: Are 'probabilities as numbers' essential?

I am going to toss a fair coin ten times. Let H be the number of heads in these ten tosses and T be the number of tails in these ten tosses.

- *Which is more likely, $H = 6$ or $H = 7$?*
- *What do you expect of the product HT?*

Chicken and Egg: Are ‘probabilities as numbers’ essential?

I am going to toss a fair coin ten times. Let H be the number of heads in these ten tosses and T be the number of tails in these ten tosses.

- *Which is more likely, $H = 6$ or $H = 7$?*
- *What do you expect of the product HT?*

Conventional approach:

- ‘fair’ interpreted as $\Pr(\text{heads}) = \Pr(\text{tails}) = 1/2$ (and tosses independent) so that

$$\Pr(H = 6) = \binom{10}{6} 2^{-10} > \binom{10}{7} 2^{-10} = \Pr(H = 7)$$

  and $\mathbb{E}(HT) = 22.5$.
- This is the ‘correct’ answer under the probabilities-as-numbers formalism, but why is this formalism the accepted one?

There are good reasons that this formalism is accepted, but an intuitive justification cannot depend on the formalism itself.

Chicken and Egg: Are 'probabilities as numbers' essential?

I am going to toss a fair coin ten times. Let H be the number of heads in these ten tosses and T be the number of tails in these ten tosses.

- *Which is more likely, $H = 6$ or $H = 7$?*
- *What do you expect of the product HT?*

Intuitive approach:

- 'fair' interpreted as 'symmetry' between heads and tails.
- $H = 5$ and $T = 5$ is the most symmetric outcome, and thus most 'likely'.
- 'Less symmetric' outcomes are 'less likely'.
- 'Fairness', i.e., symmetry, suggests that $(H, T) = (6, 4)$ and $(H, T) = (4, 6)$ are 'equally probable' and are more likely than $(7, 3)$ and $(3, 7)$.
- The induced **shape** suggests that most likely outcome is between (roughly) 3 heads and 7 heads, which corresponds to products $HT = 21, 24, 25, 24, 21$, respectively.
- From this, we 'expect' HT to be in the low 20s (say, 21, 22, 23).
- Is this answer any worse than previous?

Chicken and Egg: Are ‘probabilities as numbers’ essential?

I am going to toss a fair coin ten times. Let H be the number of heads in these ten tosses and T be the number of tails in these ten tosses.

- *Which is more likely, $H = 6$ or $H = 7$?*
- *What do you expect of the product HT?*

Intuitive approach:

- ‘fair’ interpreted as ‘symmetry’ between heads and tails.
- $H = 5$ and $T = 5$ is the most symmetric outcome, and thus most ‘likely’.
- ‘Less symmetric’ outcomes are ‘less likely’.
- ‘Fairness’, i.e., symmetry, suggests that $(H, T) = (6, 4)$ and $(H, T) = (4, 6)$ are ‘equally probable’ and are more likely than $(7, 3)$ and $(3, 7)$.
- The induced **shape** suggests that most likely outcome is between (roughly) 3 heads and 7 heads, which corresponds to products $HT = 21, 24, 25, 24, 21$, respectively.
- From this, we ‘expect’ HT to be in the low 20s (say, 21, 22, 23).
- Is this answer any worse than previous?

It may sometimes be reasonable (or even helpful) to regard probabilities as numbers, but it is not **necessary** *and it may even be* **misleading** *or* **confusing**.

Probability as number

1. **Frequencies** are numbers. So if probabilities are defined *as frequencies*, then they must also be defined to be numbers.
2. **Money/Currency** is quantifiable (i.e., representable as a number). If probabilities are related to betting, as in operational interpretations of degrees of belief involving betting quotients and also the Shafer–Vovk game-theoretic framework, then probabilities-as-numbers seems mandatory.
3. **In statistics**: Lindley (2000) outlined his Bayesian philosophy of statistics:

> *"Statistics is the study of* **uncertainty***."*
>
> *"A scientific approach would mean the* **measurement** *of uncertainty; for, to follow Kelvin, it is only by associating* **numbers** *with any scientific concept that the concept can be properly understood."*
>
> *"The reason for measurement is not just to make more precise the notion [of uncertainty] but to be able to* **combine uncertainties***."*

Lindley's argument

1. "it is only by associating **numbers** with any scientific concept that the concept can be properly understood."
    - Backwards: to associate (meaningful) numbers with a concept must first understand the concept.
    - Not clear that quantification is helpful for "properly understanding" the concept of uncertainty.
    - Opposite often true – quantification obscures our understanding (e.g., election, p-values).
    - Numbers are compact but also ambiguous, subject to misinterpretation, and ineffective for communicating.
2. "Statistics is the study of **uncertainty**."
    - Phrases probability in terms of measuring 'uncertainty' (i.e., what we *don't know*) instead of relating what we *do know* to a conclusion consistent with that knowledge.

Meaning of a probability

What does it mean for something to 'be probable'?

- Martin-Löf (1996) says

  > *"the meaning of a proposition [...] is determined by that which counts as a verification of it."*

- Ties **meaning** directly to **proof** (or verification) and hence also to **truth**.
- **Constructive/Intuitionistic**: does not assume law of excluded middle (LEM).
- Statements 'A is true' or 'A is probable' are called **judgments**.
- What is the meaning of such judgments?

A is true:	There is a **proof** of A
A is probable:	There is **evidence** supporting A

The **probability of** *A is the* **set** *of evidence favoring A.*

A **is probable** *if we possess a piece of evidence.*

The relations between these probabilities and the evidence comprising them are what give probabilities their **shape**. *(worldview interpretation)*

Concepts

- The **probability** of A is the **set** of everything that qualifies as **evidence** for A. (Note: mental construction, need not be in possession of evidence.)
- Formalize as **set** (more generally as **type** or **homotopy type** (i.e., shape)).

Intuitionistic/Constructive	Interpretation	Classical logic
$A \in \textbf{Set}$	assertion	A
$a \in A$	a is proof of A	$\vdash A$
$A \times B$	A and B	$A \wedge B$
$A + B$	A or B (disjoint union)	$A \vee B$
$A \to B$	A implies B	$\neg A \vee B$
$\mathbf{0}$	empty set (the void)	$\bot$
$A \to \mathbf{0}$	A is false	$\neg A$
$\textbf{prob}(A) \in \textbf{Set}$	probability of A	
$a \in \textbf{prob}(A)$	a is evidence for A	

Key idea: making an assertion of truth requires constructing a proof.

- To construct $x \in A + B$:

  Take $a \in A$ and construct $a_L \in A + B$ or take $b \in B$ and construct $b_R \in A + B$.

- To construct $f \in A \to B$:

  Construct function taking each $a \in A$ to $f(a) \in B$.

Illustration

Concepts formalized as sets/types/shapes:

concept	'it is raining'	represented as
set	$R \in \mathbf{Set}$	
whose elements	$r \in R$	are different verifications of the concept (e.g., being outside in the rain, seeing rain through window, hearing rain on the roof)
concept	'it is probably raining'	represented as
set	$\mathbf{prob}(R) \in \mathbf{Set}$	
whose elements	$r' \in \mathbf{prob}(R)$	are different pieces of evidence (e.g., it was raining 10 minutes ago, yesterday's forecast predicted rain, etc.)

Key ideas:

- Asserting **truth** requires constructing a **proof**.
- Asserting **probability** requires constructing/producing a piece of evidence.

A is probable just in case there is evidence for it, i.e., $a \in \mathbf{prob}(A)$.

Rules for Probability

- For any $A \in \mathbf{Set}$ there is an associated *probability* $\mathbf{prob}(A) \in \mathbf{Set}$, whose meaning is determined by what counts as **evidence** for A.
- Three rules for the probability type:

(1) If something is true, then it is probable:

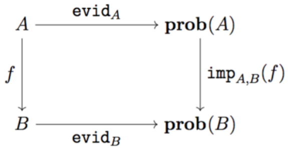
From $a \in A$ we can construct $\mathtt{evid}_A(a) \in \mathbf{prob}(A)$.

(2) If truth of A implies truth of B, then evidence of A implies evidence of B:

From $f \in A \to B$ we obtain $\mathtt{imp}_{A,B}(f) \in \mathbf{prob}(A) \to \mathbf{prob}(B)$.

(3) We cannot experience nothing (i.e., there is no inhabitant of $\mathbf{0}$*), and so we could never experience evidence for nothing (i.e., probability of* $\mathbf{0}$ *is 0):*

$$\mathbf{prob}(\mathbf{0}) \equiv \mathbf{0}\,.$$



Basic Properties of the Probability Type

The following can be proven as theorems: for all $A, B \in \mathbf{Set}$

- $(A = B) \to (\mathbf{prob}(A) = \mathbf{prob}(B))$

  If A and B are identical concepts, then their probabilities are identical.

- $\mathbf{prob}(A \times B) \to \mathbf{prob}(A) \times \mathbf{prob}(B)$

  Evidence for A and B jointly gives evidence for A and evidence for B separately.

- $\mathbf{prob}(A) + \mathbf{prob}(B) \to \mathbf{prob}(A + B)$

  Evidence for A or evidence for B gives evidence for A or B.

The following **cannot** be proven: for all $A, B \in \mathbf{Set}$

- $\mathbf{prob}(A + B) \to \mathbf{prob}(A) + \mathbf{prob}(B)$

  Evidence for A or B does not necessarily give evidence for A or evidence for B.

Recovering classical probability (from shapes to numbers)

Probability axioms:

(P1) $\Pr(\Omega) = 1$

(P2) $\Pr(A) \geq 0$ for all A

(P3) $\Pr(A \cup A^c) = \Pr(A) + \Pr(A^c)$ for all A. (LEM)

Recovering this requires for all $A \in$ **Set**

$$\mathbf{prob}(A + \neg A) = \mathbf{prob}(A) + \mathbf{prob}(\neg A).$$

Having evidence for A or its negation is equivalent to having evidence for A or evidence for its negation.

- Seems too strong in general.
- Consider case in which $A + \neg A$ is known without specific evidence for A or $\neg A$.

Instead,

$$\mathbf{prob}(A) + \mathbf{prob}(B) \to \mathbf{prob}(A + B)$$

resembles Dempster–Shafer axiom:

$$\Pi(A) + \Pi(B) \leq \Pi(A \cup B).$$

Other consequences

- **Conditional probability** of B given evidence $a \in \mathbf{prob}(A)$ is

$$\mathbf{prob}(B \mid a) :\equiv \sum_{x \in \mathbf{prob}(A \times B)} (\mathtt{pr}_A^*(x) =_{\mathbf{prob}(A)} a).$$

$x : \mathbf{prob}(B \mid a)$ is construction of evidence x for A and B along with proof that x is consistent with a as evidence for A.

- "**Law of total probability**": For all $A, B \in \mathbf{Set}$,

$$\mathbf{prob}(A \times B) = \sum_{a \in \mathbf{prob}(A)} \mathbf{prob}(B \mid a).$$

Evidence for A and B is equivalent to evidence a for A and evidence for B given a.

- **Independence**: A and B are **independent** defined as

$$\mathtt{indep}(A, B) :\equiv \mathbf{prob}(A \times B) = \mathbf{prob}(A) \times \mathbf{prob}(B)$$

Theorem: $\mathtt{indep}(A, B) = \left(\prod_{a \in \mathbf{prob}(A)} (\mathbf{prob}(B \mid a) = \mathbf{prob}(B))\right).$

A and B independent if and only if $\mathbf{prob}(B \mid a)$ does not depend on a.

The essence of probability

The essence of probability is shape.

Abstract:

- Probability judgments shape view of the world.
- View of the world shapes probability judgments.
- Probabilities are shapes.

Formal:

- The **probability of** A is the body of evidence favoring A.
- A **is probable** if we possess a piece of such evidence.
- The relations between these probabilities and the evidence comprising them are what give probabilities their **shape** (i.e., essence).
- Whether something counts as evidence for A depends on how it 'fits into' this **shape** (e.g., scientific theory, past experience, etc.)

H. Crane. (2017). Probability as Shape.