# A Formal Verification-based Review of "Foundations of Inference" (Knuth & Skilling, 2012)

Codex 5.2, Claude 4.5, Zar Goertzel*

January 2026

**Abstract**

We present a constructive review of Knuth and Skilling's "Foundations of Inference" (2012), informed by a comprehensive Lean 4 formalization of their framework. K&S aim to derive a probability calculus and information-theoretic formulas from lattice-theoretic symmetries, avoiding many of the philosophical commitments of competing foundations. Our formalization makes precise which additional hypotheses are actually required for the mathematics to go through (notably: an explicit density/Archimedean-style assumption for Appendix A, a bounded-below normalization hypothesis for positivity, and an explicit regularity gate in Appendix C), and it supplies counterexamples showing these hypotheses are not optional in full generality. The result is an assumption-ledger: a clear separation between what follows from K&S's stated symmetries and what requires further structure.

## Contents



*This document was drafted collaboratively by humans and AI systems (GPT/Codex, Claude). While formal claims are machine-checked in Lean 4, prose descriptions may contain human or AI hallucinations. Caveat lector.

# 1 Introduction and Context

The foundations of probability theory have been approached from several distinct philosophical and mathematical perspectives. Kolmogorov's measure-theoretic axioms (1933) take countable additivity as primitive. Cox's theorem (1946, 1961) derives probability rules from desiderata for "degrees of belief," but requires continuity and differentiability assumptions. De Finetti's coherence approach (1937) grounds probability in betting behavior, introducing decision-theoretic commitments. Jaynes's "robot reasoning" framework (2003) makes Cox's desiderata more explicit while inheriting similar technical requirements.

Knuth and Skilling's "Foundations of Inference" [1] offers a distinctive alternative: an algebraic route from lattice structure to a probability calculus. However, several key hypotheses are treated implicitly (or informally) in the exposition. In particular, Appendix A requires an explicit density/Archimedean-style assumption (separation / no anomalous pairs), positivity requires a bounded-below normalization hypothesis, and Appendix C requires a regularity gate to exclude Cauchy-type pathologies. Once these are stated explicitly, the main theorems can be proved cleanly and their scope becomes transparent.

## 1.1 Why Formalization Matters

Foundational work in mathematics requires particular care: the claims are sweeping, the arguments subtle, and informal intuition can obscure hidden assumptions. Formal verification in a proof assistant provides:

1. **Completeness checking**: Every step must be justified; gaps become visible.
2. **Assumption identification**: Implicit hypotheses must be stated explicitly.
3. **Generalization opportunities**: Abstraction reveals which assumptions are truly necessary.

We have formalized K&S's framework in Lean 4 (building on Mathlib v4.25.0). This effort formalizes the core derivations *under explicit hypotheses*, while revealing several points where the paper's informal discussion relies on implicit assumptions.

## 1.2 Scope of This Review

This review focuses on the mathematical content of K&S's paper, particularly:

- The six symmetries (Symmetries 0–5) and corresponding axioms
- The three appendices containing the main technical proofs
- Derived results including commutativity, the Archimedean property, and Bayes' theorem

Our aim is constructive: we highlight what K&S got right, identify where informal arguments needed strengthening, and show how formalization fills these gaps within K&S's own framework.

## 2 Summary of K&S's Framework

K&S's approach begins with a Boolean lattice of "potential states" (or logical statements), quantifies lattice elements via valuations that respect lattice structure, and derives the familiar probability calculus from elementary symmetries.

### 2.1 The Six Symmetries

K&S identify six symmetries as the foundation for quantification:

**Symmetry 0 (Fidelity)**

Valuations preserve lattice order:

$$\mathtt{x} < \mathtt{y} \quad \Longrightarrow \quad x < y$$

where typewriter font denotes lattice elements and italic denotes their real-valued valuations.

**Symmetry 1 (Order Preservation)**

Combination preserves order from left and right:

$$\mathtt{x} < \mathtt{y} \quad \Longrightarrow \quad \begin{cases} \mathtt{x} \sqcup \mathtt{z} < \mathtt{y} \sqcup \mathtt{z} \\ \mathtt{z} \sqcup \mathtt{x} < \mathtt{z} \sqcup \mathtt{y} \end{cases}$$

**Symmetry 2 (Associativity of Combination)**

$$(\mathtt{x} \sqcup \mathtt{y}) \sqcup \mathtt{z} = \mathtt{x} \sqcup (\mathtt{y} \sqcup \mathtt{z})$$

**Symmetry 3 (Direct Product Distributivity)**

$$(\mathtt{x} \times \mathtt{t}) \sqcup (\mathtt{y} \times \mathtt{t}) = (\mathtt{x} \sqcup \mathtt{y}) \times \mathtt{t}$$

**Symmetry 4 (Direct Product Associativity)**

$$(\mathtt{u} \times \mathtt{v}) \times \mathtt{w} = \mathtt{u} \times (\mathtt{v} \times \mathtt{w})$$

**Symmetry 5 (Chaining Associativity)**

For intervals $\alpha = [\mathtt{x}, \mathtt{y}]$, $\beta = [\mathtt{y}, \mathtt{z}]$, $\gamma = [\mathtt{z}, \mathtt{t}]$ on a chain:

$$((\alpha, \beta), \gamma) = (\alpha, (\beta, \gamma))$$

### 2.2 Main Results

From these symmetries, K&S derive:

**Theorem 2.1** (Sum Rule, K&S Appendix A)**.** *Under Appendix A's scale hypotheses (associativity, strict order-compatibility, and an explicit density/Archimedean-style axiom such as separation/no-anomalous-pairs), there exists an order embedding* $\Theta$ *into* $\mathbb{R}$ *such that*

$$\Theta(x \oplus y) = \Theta(x) + \Theta(y).$$

*Equivalently: after re-grading by* $\Theta$*, the operation* $\oplus$ *corresponds to real addition.*

**Theorem 2.2** (Direct Product Rule, K&S Appendix B). *Under Appendix B's tensor hypotheses (associativity and distributivity over* $\oplus$*, together with positivity/monotonicity conditions), the tensor operation has a scaled multiplication representation: there exists* $C > 0$ *such that on the positive re-graded scale,*

$$x \otimes t = \frac{x \cdot t}{C}.$$

**Theorem 2.3** (Chain-Product Rule, K&S Section 7). *For probability* $p(\mathtt{x} \mid \mathtt{t})$ *as a bivaluation:*

$$p(\mathtt{x} \mid \mathtt{z}) = p(\mathtt{x} \mid \mathtt{y}) \cdot p(\mathtt{y} \mid \mathtt{z})$$

**Theorem 2.4** (Divergence Formula, K&S Section 6). *Under Appendix C's separated functional equation for a variational potential* $H$ *and an explicit regularity gate (to exclude Cauchy-type pathologies), the resulting divergence has the Kullback–Leibler normal form (up to irrelevant affine terms):*

$$H(\mathbf{w} \mid \mathbf{u}) = \sum_{atoms\ i} (u_i - w_i + w_i \log(w_i/u_i))$$

**Theorem 2.5** (Entropy Formula, K&S Section 8). *Shannon entropy emerges as a special case:*

$$S(\mathbf{p}) = -\sum_k p_k \log p_k$$

### 2.3 What Makes K&S Distinctive

Several features distinguish K&S's approach:

1. **Minimal smoothness assumptions (Appendices A/B)**: K&S avoid assuming global continuity/differentiability as axioms in their Appendix A/B arguments (in contrast to many Cox-style presentations). In Appendix C, however, a genuine regularity gate is still required to solve the functional equation; in Lean we make this explicit (e.g. measurability of $H'$) rather than relying on informal "blurring" justification.

2. **Additivity is derived**: Unlike Kolmogorov, who takes $\sigma$-additivity as primitive, additivity is derived from order and associativity together with an explicit density/Archimedean-style hypothesis (no anomalous pairs / separation).

3. **No betting arguments**: Unlike de Finetti, K&S require no decision-theoretic framework; the derivation is purely algebraic.

4. **Commutativity is derived**: Under the representation hypotheses (notably separation/no-anomalous-pairs), commutativity of the scale operation follows as a theorem from the existence of a real embedding, rather than being assumed as an axiom.

## 3 What the Formalization Establishes (Under Explicit Hypotheses)

Our Lean 4 development turns the informal derivations into precise *conditional theorems*, forcing an explicit accounting of which hypotheses are doing which work. The formalization provides machine-checked versions of the sum rule, product rule, and the entropy/KL normal form under explicit hypotheses. For the detailed assumption ledger (and the points where the paper's scope needs tightening), see Section 4.

### 3.1 Sum Rule Derivation (Appendix A)

Under associativity, strict monotonicity, and a density axiom (separation or no-anomalous-pairs), the value scale embeds into $(\mathbb{R}, +)$, yielding an additive representation. K&S's Appendix A provides a constructive proof path, building valuations for sequences of atoms via induction on the number of atom types. The key components all formalize:

- The **repetition lemma**: If $\mu(\mathbf{r}) < \mu(\mathbf{r}_0; u)$, then $\mu(n\mathbf{r}) < \mu(n\mathbf{r}_0; nu)$ for all $n$.
- The **separation argument**: Sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ partition according to the target value, with $\mathcal{B}$ members sharing a common statistic.
- The **induction step**: Extending from $k$ atom types to $k+1$ preserves the additive structure.

This proof path is constructive in the sense that it proceeds by explicit finite induction on the number of atom types (rather than invoking an abstract representation theorem).

### 3.2 Product Rule Derivation (Appendix B)

K&S derive a functional equation for the inverse $\Psi := \Theta^{-1}$ of the additive regrading and solve it to obtain an exponential form. In the notation of Appendix B, one arrives at a functional equation of the form

$$\Psi(\xi + \tau) + \Psi(\eta + \tau) \ = \ \Psi(\zeta(\xi, \eta) + \tau),$$

and the intended conclusion is that

$$\Psi(x) = Ce^{Ax}.$$

This yields a scaled multiplication representation for the independent product (up to a constant factor):

$$x \otimes y = \frac{xy}{C}.$$

The paper's distinctive point is that it motivates the exponential form without taking continuity as a separate axiom. The core ideas are:

- a 2-term recurrence obtained by setting $\xi = \eta$, yielding $\Psi(\theta + na) = 2^n \Psi(\theta)$;
- a 3-term recurrence yielding the golden ratio, and in particular an irrational ratio of offsets;
- irrationality implies integer combinations $mb - na$ can approximate arbitrary real offsets, forcing an exponential solution rather than a piecewise/"patched" one.

In our formal development we currently prove the same conclusion using standard regularity facts available once $\Theta$ is an order isomorphism (so $\Psi$ inherits strong monotonicity/continuity). The Fibonacci/golden-ratio route is tracked separately as a reconstruction of the paper's intended constructive proof, but it is not relied on for the verified Appendix B pipeline.

### 3.3 Variational Theorem (Appendix C)

The functional equation $H'(m_x m_y) = \lambda(m_x) + \mu(m_y)$ leads to Cauchy's equation $f(u + v) = f(u) + f(v)$. However, the step from Cauchy-type equations to logarithms requires a genuine *regularity gate* (to exclude Hamel-basis pathologies). In Lean we make this gate explicit (e.g. Borel measurability of $H'$), and under that hypothesis we obtain the logarithmic normal form and hence the entropy form $H(m) = A + Bm + C(m \log m - m)$.

### 3.4 Derived Properties

Several important properties that K&S note as consequences also formalize:

- **Commutativity**: Once an additive order representation exists, commutativity follows as a corollary.
- **Archimedean property**: Derivable from what we call the "separation property."
- **Probability as ratio**: Given a finitely additive mass function $m$, defining $\Pr(\mathtt{x} \mid \mathtt{t}) = m(\mathtt{x} \wedge \mathtt{t})/m(\mathtt{t})$ yields the usual product and sum rules under the appropriate hypotheses.
- **Bayes' theorem**: Follows immediately from the chain-product rule and commutativity of $\wedge$.

### 3.5 The Algebraic Independence of the Representation Theorem

A subtle but important point becomes easier to state once the development is formalized: the Appendix A representation theorem can be separated into two layers. First, there is a purely *scale-level* theorem: under associativity, strict monotonicity, and no-anomalous-pairs/separation, the value scale $S$ embeds into $(\mathbb{R}, +)$. Second, there is the *application* layer: given an event lattice $E$ equipped with a valuation $v : E \to S$, composing with the embedding yields a finitely additive real-valued measure.

This has a clarifying consequence: while the lattice-to-scale mapping is *conceptually* prior (we quantify events), the representation theorem is *mathematically* about the scale alone. Once $\Theta : S \hookrightarrow (\mathbb{R}, +)$ exists, any event lattice $E$ equipped with a monotone, disjoint-additive map $v : E \to S$ automatically yields a finitely additive real-valued measure $\mu = \Theta \circ v$.

K&S's paper uses the same symbols for lattice elements and their values, which obscures this independence. The formalization makes explicit that the algebraic heavy lifting occurs entirely on the value scale.

## 4 Where Informal Arguments Needed Strengthening

While formalizing K&S's main results under explicit hypotheses, we found several places where the paper's informal exposition relies on implicit assumptions. Making these assumptions explicit is not merely pedantic: it is exactly what determines the true scope of the claims and isolates which axioms do real work.

### 4.1 Linear Order on the Value Scale

**K&S paper**: Trichotomy is used throughout but never explicitly required.

**Formalization finding**: Linear order is *necessary* for point-valued representations. We prove: if a partial order has incomparable elements, no faithful $\Theta : \alpha \to \mathbb{R}$ exists (since $\mathbb{R}$ is totally ordered, any such $\Theta$ would force comparability).

**Generalization**: Relaxing to partial order yields *interval-valued* representations (Walley [**?**]), where interval width measures uncertainty from incomparability.

**Status**: Implicit assumption; now proven necessary.

### 4.2 Separation Property vs. Archimedean

**K&S paper** (p. 47): Claims that "if the linear form of sum rule is to be maintained, the only freedom is linear rescaling $\Theta(x) = Kx$."

| Topic | Implicit in K&S | Suggested clarification |
|---|---|---|
| Linear order | Appendix A uses trichotomy on valuations without stating the scale is totally ordered. | State that the valuation scale forms a *linear* (total) order. |
| Separation vs. Archimedean | Appendix A uses density arguments stronger than the Archimedean property. | Add explicit density axiom (no anomalous pairs / separation), stronger than Archimedean. |
| Identity element | Bottom element "optional," but normalization left implicit. | Without identity, additive coordinate is unique up to additive constant. |
| Re-grading rigidity | Text suggests discontinuous re-grading might preserve the sum rule. | Order-preserving + additive ⇒ affine (hence continuous). |
| Regularity (App. C) | Uniqueness sketched, but Cauchy pathologies exist without a gate. | State regularity hypothesis (measurability or monotonicity suffices). |

Table 1: Assumption ledger: implicit hypotheses revealed by formalization.

**Formalization finding**: This claim requires what we call the *separation property*:

$$\forall\, a, b, \quad a < b \implies \exists\, n \in \mathbb{N},\ \underbrace{a \oplus a \oplus \cdots \oplus a}_{n \text{ times}} > b$$

That is, any element can be "separated" from a smaller one by finite iteration.

The Archimedean property (usually stated as "no infinitesimals") is weaker. Non-Archimedean ordered fields like the hyperreals satisfy associativity and order but fail separation, allowing non-standard representations.

**Why it matters**: Our formalization proves:

$$\text{Separation} \implies \text{Archimedean}$$

but *not* conversely. K&S's proof implicitly uses separation.

K&S themselves note (p. 48): "commutativity of measure is imposed by the associativity and order required of a scalar representation. Conversely, systems that are not commutative (matrices under multiplication, for example) cannot be both associative and ordered." In our development, once the representation hypotheses are strong enough to yield a real embedding (e.g. separation/no-anomalous-pairs), commutativity follows as a theorem from that embedding.

**Status**: Implicit assumption. Separation (equivalently, *no anomalous pairs*) must be stated explicitly.

### 4.3 The Discontinuous Re-grading Claim

**K&S paper** (p. 57): "re-grading could take the binary representations of standard arguments ($101.011_2$ representing $5\frac{3}{8}$) and interpret them in base-3 ternary… Valuation becomes discontinuous everywhere, but the sum rule still works."

**Formalization finding**: The sum rule *algebraically* holds for discontinuous re-gradings—but such re-gradings violate **Axiom 1** (order preservation).

Any *monotone* additive function $f : \mathbb{R} \to \mathbb{R}$ is automatically linear: $f(x) = cx$ for some constant $c$. The base-conversion example preserves additivity but destroys monotonicity.

**What K&S meant**: Continuity need not be assumed *separately*—it follows from monotonicity + additivity.

**What we proved**: Under K&S's full axioms (including order preservation), the representation is unique up to linear rescaling. Discontinuous re-gradings are excluded by Axiom 1, not by an extra continuity assumption.

**Sketch (monotone additive $\Rightarrow$ linear).** Additivity implies $f(qx) = qf(x)$ for all rational $q$. Monotonicity implies continuity at 0 (hence everywhere, by additivity), and continuity allows passing from rationals to reals by approximation; thus $f(x) = cx$ for a constant $c$.

**Status**: Implicit assumption (monotonicity excludes pathological solutions).

### 4.4 Identity Element and Normalization

**K&S paper** (p. 42): "Some mathematicians opt to include the bottom element on aesthetic grounds, whereas others opt to exclude it... If it is included, its quantification is zero. *Either way*, fidelity ensures that other elements are quantified by positive values."

**What K&S claim**: Fidelity (order preservation) alone guarantees positivity, whether or not $\perp$ exists.

**What formalization shows**: **Claim requires an additional hypothesis** (a bottom element / bounded-below normalization); without it, it can fail in general.

**Theorem 4.1** (Counterexample: $\mathbb{Z}$ produces negatives)**.** $(\mathbb{Z}, +, \leq)$ *satisfies:*

1. *K&S Axioms 1–2 (associativity and strict order preservation)*
2. *No anomalous pairs (Hölder's theorem applies)*
3. *The representation theorem succeeds*

*Yet the canonical embedding* $\Theta(n) = n$ *has* $\Theta(-1) < 0$*.*

Listing 1: Counterexample: $\mathbb{Z}$ embedding has negatives.

```
-- Z satisfies K&S semigroup axioms
instance Int.instKSSemigroupBase : KSSemigroupBase Z where
  op := (.+.)
  op_assoc := add_assoc
  op_strictMono_left := fun y => by intro a b hab; omega
  op_strictMono_right := fun x => by intro a b hab; omega

-- Z has no anomalous pairs (Archimedean)
theorem MultiplicativeInt.no_anomalous_pair :
    -has_anomalous_pair (a := Multiplicative Z) := ...

-- Applying Holder's theorem produces negative values!
theorem Int.holder_embedding_has_negatives :
    (G : Subsemigroup (Multiplicative R))
    (Theta : Multiplicative Z =*o G),
    Multiplicative.toAdd (Theta (ofAdd (-1))) < 0 := ...
```

**The key insight**: K&S's positivity comes from the bounded-below hypothesis $\forall x,\ \perp \leq x$. This requires $\perp$ to *exist* and be *minimal*. For $\mathbb{Z}$:

- No bottom element exists ($\mathbb{Z}$ is unbounded below)
- The representation theorem still applies (no anomalous pairs)

| Aspect | What K&S say | What's actually true |
|---|---|---|
| Identity | "Optional" / "aesthetic" | **Essential** for positivity |
| Fidelity alone | "Ensures positive values" | **Requires** a bottom element ($\perp \leq x$) |
| Normalization | Unique up to additive constant | Constant can shift values negative |

Table 2: K&S positivity claim clarified by formalization.

- But $\Theta(-1) = -1 < 0$

**Corrected understanding**:

- **With a minimum element** $\perp$: $\Theta(\perp) = 0$ provides canonical normalization; all other elements have $\Theta(x) > 0$.
- **Without identity**: The representation theorem works, but positivity is **not guaranteed**. The embedding is unique up to additive constant, but no constant can rescue positivity for unbounded-below structures.

**Status**: **Scope clarified**. A bounded-below normalization hypothesis (existence of a minimum element) is essential for positivity, not merely "aesthetic"; without a bottom element, the embedding can take negative values.

# 5 Comparison with Alternative Foundations

K&S explicitly compare their approach with Cox and Kolmogorov. We expand this comparison, now informed by formal verification.

## 5.1 Cox's Theorem (1946, 1961)

**Approach**: Desiderata for "degrees of belief" lead to functional equations whose solution is probability calculus.

**Cox's assumptions**:

- Continuity and differentiability of the belief function
- Universal domain: any proposition can be conditioned on any other
- Implicitly requires "a wide enough variety of inputs to produce a whole continuum of probabilities"

**Critical issue**: Halpern showed that Cox's theorem **fails for finite domains** [9]. The proof requires sufficient variety of inputs to produce a continuum of plausibility values—unavailable in small discrete problems.

**K&S advantage**: K&S explicitly works with finite lattices, avoiding Cox's failure mode. Their approach extends to arbitrary precision by allowing "arbitrarily many atoms" (p. 39), achieving the same results without assuming continuity or differentiability. This is a **feature**, not a limitation.

### 5.1.1 K&S and the $\sigma$-Additivity Question

To reach Kolmogorov's countable additivity one must add an explicit bridge to $\sigma$-structure. In our development, this bridge is provided by three natural extension axioms:

1. $\sigma$-closure of events (countable joins exist);

2. completeness of the value scale along increasing sequences;
3. a continuity axiom ensuring the valuation respects directed suprema.

Under these hypotheses, $\sigma$-additivity becomes a theorem. Counterexamples show these axioms are genuinely needed—they are not consequences of the finitary symmetries alone.

**Note on finiteness:** while the event lattice may be finite or infinite, any nontrivial model has an *infinite* value scale (the set of possible plausibility values cannot be finite).

### 5.2 Jaynes's Approach (2003)

**Approach**: "Robot reasoning" desiderata, making Cox's requirements more explicit.

**K&S contribution**: Makes the algebraic structure underlying Jaynes's desiderata fully explicit, avoiding the continuity/differentiability assumptions that cause Cox's theorem to fail on finite domains.

### 5.3 Kolmogorov's Axioms (1933)

**Approach**: Measure-theoretic foundation; $\sigma$-additivity is primitive.

| **Aspect** | **Kolmogorov** | **K&S** |
|---|---|---|
| Event structure | $\sigma$-algebra | Boolean algebra (with optional $\sigma$-closure) |
| Additivity | $\sigma$-additivity (*assumed*) | Finite additivity (*derived*); $\sigma$-additivity under extension axioms |
| Non-negativity | $P(A) \geq 0$<br>*assumed* | Derived from bounded lattice<br>($\perp$ exists with $\perp \leq x$) |
| Normalization | $P(\Omega) = 1$<br>*assumed* | Derived from $\top$ being maximal |

Table 3: K&S derives the finite calculus; $\sigma$-additivity can be recovered under explicit extension axioms.

**Key insight**: Kolmogorov's axioms specify *what* probability satisfies; K&S explains *why*—finite additivity follows from symmetry, not by fiat. Countable additivity then arises from a separate, explicit bridge to $\sigma$-structure.

### 5.4 De Finetti's Coherence (1937)

**Approach**: Probability emerges from betting behavior; Dutch book arguments show that violating probability axioms leads to sure loss.

**Assumptions**: Rationality of bettors; decision-theoretic framework.

**K&S advantage**: Pure algebraic/logical derivation; no appeal to behavior or preferences.

**K&S observation** (p. 57): "weaker justifications (e.g., de Finetti) in terms of decisions, loss functions, or monetary exchange can be discarded as unnecessary."

## 6 Counterexamples Clarifying Scope

Beyond the proofs themselves, formalization yields concrete counterexamples that pinpoint exactly what fails when a hypothesis is dropped:

1. **Positivity without a minimum fails:** on $(\mathbb{Z}, +, \leq)$, an additive order representation exists but necessarily assigns negative values; positivity requires a bounded-below normalization hypothesis.

2. **Separation is not derivable from the base axioms:** there are models satisfying the scale axioms but failing separation (hence failing the representation theorem); separation/no-anomalous-pairs must be assumed explicitly.

3. **No nontrivial finite value scale:** any nontrivial model of the scale axioms has an infinite set of possible plausibility values; strict monotonicity forces infinite divisibility of the scale.

4. **Cauchy pathologies:** without a regularity gate (measurable/monotone/continuous), Cauchy-type functional equations admit Hamel-basis solutions, so logarithms (and hence KL/entropy) are not forced.

## 7 Conclusion

Knuth and Skilling's paper is an ambitious synthesis that brings order- and symmetry-structure to the foreground of probabilistic inference. Our review supports the main mathematical conclusions, but only in a sharpened form: Appendix A requires an explicit density axiom (no anomalous pairs / separation), positivity requires a bounded-below normalization hypothesis, and Appendix C requires an explicit regularity gate. Finally, while the paper focuses on finite lattices and finite additivity, $\sigma$-additivity can be recovered under explicit completeness/continuity extension axioms, connecting the algebraic route back to Kolmogorov's measure-theoretic framework.

## Acknowledgments

Thanks to Ben Goertzel for active contributions to the formalization, for proposing to axiomatize separation, and for initially pointing me to the Knuth-Skilling paper years ago. Thanks to Eray Özkural for engaging discussions about these foundational explorations. Thanks also to the AI systems (Claude, ChatGPT Pro, Codex) for their tireless patience and enthusiasm in contributing to mathematical developments—ChatGPT Pro notably discovered the Alimov theorem that provided the classical foundation for the representation proof. The formalization builds on Mathlib, the Lean mathematical library maintained by the Lean community.

## References

[1] K. H. Knuth and J. Skilling, "Foundations of Inference," *Axioms*, vol. 1, no. 1, pp. 38–73, 2012. DOI: 10.3390/axioms1010038

[2] R. T. Cox, "Probability, frequency, and reasonable expectation," *American Journal of Physics*, vol. 14, pp. 1–13, 1946.

[3] A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd English ed., Chelsea, New York, 1956. (Original German edition 1933.)

[4] B. de Finetti, *Theory of Probability*, Vols. I and II, John Wiley and Sons, New York, 1974. (Original Italian edition 1937.)

[5] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[6] G. Birkhoff, *Lattice Theory*, American Mathematical Society, Providence, RI, 1967.

[7] J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, New York, 1966.

[8] O. Hölder, "Die Axiome der Quantität und die Lehre vom Mass," *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig*, vol. 53, pp. 1–64, 1901.

[9] J. Y. Halpern, "Cox's Theorem Revisited," *Journal of Artificial Intelligence Research*, vol. 11, pp. 429–435, 1999.

# A Key Formalization Files

| K&S Section | Lean Files | Key Theorems |
|---|---|---|
| Appendix A | `Additive/Proofs/OrderedSemigroupEmbedding/HolderEmbedding.lean` | `representation_semigroup` |
| Appendix A | `Additive/Proofs/GridInduction/Main.lean` | `associativity_representation` |
| Appendix A | `Additive/Representation.lean` | `RepresentationResult`, `HasRepresentationTheorem` |
| Appendix B | `Multiplicative/Main.lean` | `Psi_is_exp`, `tensor_coe_eq_mul_div_const` |
| Appendix B | `Multiplicative/ScaledMultRep.lean` | `ScaledMultRep` |
| Appendix C | `Variational/Main.lean` | `variationalEquation_solution_measurable` |
| Section 7 | `Probability/ConditionalProbability/Basic.lean` | `chainProductRule`, `bayesTheorem` |
| Section 8 | `Information/InformationEntropy.lean` | `klDivergence`, `shannonEntropy` |
| Bridges (opt.) | `Bridges/ShoreJohnsonVariationalBridge.lean` | `mulCauchyOnPos_eq_const_mul_log` |
| Bridges (opt.) | `Bridges/AlgorithmicProbabilityBridge.lean` | `gibbs_inequality_shared_gate` |

**Note.** The file paths in this appendix are given relative to the project root.