Foundations of Inference: A Formal Treatment

Formalization of Knuth & Skilling (2012)

January 27, 2026

Abstract

We present a complete formalization in Lean 4 of Knuth & Skilling's axiomatic foundations for inference (2012). Their approach derives the probability calculus, measure theory, and information theory from elementary symmetries, without assuming continuity or differentiability. Our formalization makes explicit the implicit assumptions in their derivation, provides **three independent proof paths** for the main representation theorem (grid-based induction, Dedekind cuts, and Hölder embedding), and extends their framework to handle independence and product measures. We demonstrate K&S's claim that the framework applies to non-Boolean distributive lattices (chains, open sets of $\mathbb{R}$). A key finding: the **NoAnomalousPairs** condition (from classical ordered semigroup theory) is sufficient for the representation theorem, via Eric Paul's formalized Hölder embedding.

Contents

1 Introduction

Knuth & Skilling [1] presented a foundation for inference that unites and extends the approaches of Cox and Kolmogorov. Their key innovation is to derive the standard probability calculus from elementary *symmetries* rather than axiomatizing it directly. The core insight: algebraic symmetries of combination operations uniquely determine quantification rules.

1.1 The Knuth-Skilling Program

The K&S approach proceeds in stages:

1. **Symmetries 0–2** (fidelity, monotonicity, associativity) ⇒ **Measure theory**: The sum rule $x \oplus y = x + y$ for disjoint events.

2. **Symmetries 3–4** (product distributivity, product associativity) ⇒ **Independence**: The product form for independent measures.

3. **Symmetry 5** (chaining associativity) ⇒ **Probability calculus**: Conditional probability and Bayes' theorem.

Each symmetry becomes an *axiom of quantification*, and these axioms uniquely determine the calculus. K&S emphasize:

- **Minimal assumptions**: No continuity, differentiability, or negation required.

- **Constructive proofs**: Build from finite systems, avoiding unobservable infinitary subtleties.

- **Wide applicability**: Foundations cover measure theory, probability, divergence, and entropy.

1.2 This Formalization

We formalize K&S Appendices A and B in Lean 4, providing:

- Precise statement of axioms and symmetries

- **Three independent proof paths** for the representation theorem:

    1. Grid-based induction (following K&S's paper structure)
    2. Dedekind cuts (classical Hölder-style construction)
    3. Hölder embedding via Eric Paul's OrderedSemigroups library

- Explicit treatment of implicit assumptions (linear order, separation property)

- Connection to the classical **NoAnomalousPairs** condition (Alimov 1950)

- Non-Boolean examples demonstrating genuine generality

- Extensions to handle independence and products

- Counterexamples showing necessity of assumptions

The formalization reveals that K&S's informal proofs are essentially correct, but rely on several implicit assumptions that become apparent only when formalizing. These details are documented in Appendix A.

A key finding of this formalization is that the **NoAnomalousPairs** condition from classical ordered semigroup theory (Alimov 1950, Fuchs 1963) is sufficient for the representation theorem, and is *weaker* than K&S's separation property. This connects K&S's work to the century-old Hölder embedding theorem for ordered groups.

2 The Knuth-Skilling Axioms

2.1 Informal Statement

K&S model the world as being in one of finitely many mutually exclusive states. States correspond to *atoms*, combined via logical OR to form elements of a Boolean lattice. A *valuation* assigns real numbers to lattice elements, faithful to the lattice order.

For disjoint elements x, y, z, K&S require:

- **Axiom 0** (Fidelity): $x > 0$ (positive valuations)

- **Axiom 1** (Monotonicity): $x < y \Rightarrow x \oplus z < y \oplus z$ and $z \oplus x < z \oplus y$

- **Axiom 2** (Associativity): $(x \oplus y) \oplus z = x \oplus (y \oplus z)$

These three axioms suffice to derive the sum rule $x \oplus y = x + y$ (Appendix A).
For independent measures on product spaces:

- **Axiom 3** (Product Distributivity): $(x \otimes t) \oplus (y \otimes t) = (x \oplus y) \otimes t$

- **Axiom 4** (Product Associativity): $(u \otimes v) \otimes w = u \otimes (v \otimes w)$

These determine the multiplicative form of $\otimes$ (Appendix B).

2.2 Formal Definition

We formalize the basic K&S algebra structure as a linearly ordered monoid with strict monotonicity:

Listing 1: Core K&S algebra axioms

```
class KnuthSkillingAlgebra (α : Type*) extends LinearOrder α where
  op : α → α → α                                    -- ⊕ combination
  ident : α                                         -- identity element
  op_assoc : ∀ x y z, op (op x y) z = op x (op y z)  -- Axiom 2
  op_ident_right : ∀ x, op x ident = x
  op_ident_left : ∀ x, op ident x = x
  op_strictMono_left : ∀ y, StrictMono (fun x => op x y)    -- Axiom 1a
  op_strictMono_right : ∀ x, StrictMono (fun y => op x y)   -- Axiom 1b
  ident_le : ∀ x, ident ≤ x                         -- Axiom 0
```

Note: K&S never explicitly state that elements are totally ordered, but their proofs rely on trichotomy ($x < y$, $x = y$, or $x > y$). Our formalization makes this explicit via `LinearOrder`.

2.3 Iteration Notation

Following K&S, we define iterated application of the operation:

Listing 2: Iteration definition

```
def iterate_op (x : α) : N → α
  | 0 => ident
  | n + 1 => op x (iterate_op x n)
```

This builds the sequence $\text{ident}, x, x \oplus x, x \oplus (x \oplus x), \ldots$, which K&S write as "0 of x", "1 of x", "2 of x", etc.

3 The Separation Property

3.1 Definition

K&S's Appendix A proof requires finding rational approximations to any pair of distinct values. This is implicit in their "induction to more than one type of atom" construction. Following a suggestion by B. Goertzel, we isolate this as an explicit axiom—the *separation property*:

Listing 3: Separation axiom

```
class KSSeparation (α : Type*) [KnuthSkillingAlgebra α] where
  separation : ∀ {a x y : α}, ident < a → ident < x → ident < y → x < y →
    ∃ n m : N, 0 < m ∧
      iterate_op x m < iterate_op a n ∧
      iterate_op a n ≤ iterate_op y m
```

Intuition: For any base $a > \text{ident}$ and distinct $x < y$, we can find exponents (n, m) such that $x^m < a^n \leq y^m$. This lets us "separate" x and y using rational powers of a.

3.2 Necessity

The separation property is *necessary* for representation. Counterexample: Consider $\mathbb{N} \times \mathbb{N}$ with lexicographic order:

$$(n_1, m_1) < (n_2, m_2) \iff n_1 < n_2 \text{ or } (n_1 = n_2 \text{ and } m_1 < m_2)$$

and operation:

$$(n_1, m_1) \oplus (n_2, m_2) = (n_1 + n_2, m_1 + m_2)$$

This satisfies Axioms 0–2 but *not* separation: $(1, 1) < (2, 0)$, but no powers of $(1, 0)$ can separate them. Consequently, no order embedding into $\mathbb{R}$ exists.

3.3 Derivable Consequences

From separation, we derive:

- **Archimedean property**: For any $x > \text{ident}$ and y, there exists n such that $\text{iterate\_op}\, x\, n > y$.

- **Commutativity**: $\text{op}\, x\, y = \text{op}\, y\, x$ for all x, y.

These are proven in `Separation/SandwichSeparation.lean` and `RepresentationTheorem/Core/Separation`

4 The Representation Theorem (Appendix A)

4.1 Statement

Theorem 4.1 (K&S Appendix A). *Let α be a type with `KnuthSkillingAlgebraBase`. Under appropriate regularity conditions (see below), there exists an order embedding $\Theta : \alpha \to \mathbb{R}$ such that:*

 1. $\Theta(ident) = 0$

 2. $\forall x, y : \alpha,\ \Theta(op\,x\,y) = \Theta(x) + \Theta(y)$

This is K&S's main result: Axioms 1–2 plus regularity uniquely determine that $\oplus$ corresponds to addition in $\mathbb{R}$ (up to order-preserving regrade).

4.2 Three Proof Approaches

Our formalization provides **three independent proof paths**, each with different assumptions and characteristics:

4.2.1 Approach 1: Grid-Based Induction (K&S Style)

Assumptions: `KSSeparation` + `RepresentationGlobalization`
   Technique: K&S's constructive proof proceeds by:

 1. **One type**: For a single atom a, assign iterate_op $a\,n \mapsto n \cdot \Theta(a)$.

 2. **Induction**: For a new atom d, use separation to find (u, v) such that $a^u < d^v \leq a^{u+1}$. Define $\Theta(d) = (u + \delta)/v$ for appropriate $\delta \in [0, 1]$.

 3. **Globalization**: Show this extends to all combinations consistently (the "triple family trick").

   Files: `RepresentationTheorem/Core/Induction/ThetaPrime.lean` ($\sim$2500 lines),
      `RepresentationTheorem/Core/MultiGrid.lean` ($\sim$800 lines)
   Pros: Follows K&S paper closely; self-contained
Cons: Complex grid management; hard to maintain

Listing 4: Grid-based representation theorem

```
1  theorem associativity_representation
2      (α : Type*) [KnuthSkillingAlgebra α] [KSSeparation α]
3      [RepresentationGlobalization α] :
4      ∃ Θ : α → R,
5        (∀ a b : α, a ≤ b ↔ Θ a ≤ Θ b) ∧
6        Θ ident = 0 ∧
7        ∀ x y : α, Θ (op x y) = Θ x + Θ y
```

4.2.2 Approach 2: Dedekind Cuts (Classical Style)

Assumptions: `KSSeparation` + `KSSeparationStrict`
   Technique: Construct Θ directly via Dedekind cuts:

 1. For a base element $a > ident$, define the *cut set*:

$$\text{cutSet}(a, x) = \{q \in \mathbb{Q} : q \geq 0 \wedge \exists m, n.\ a^m \leq x^n \wedge q = m/n\}$$

2. Define $\Theta(x) = \sup(\text{cutSet}(a, x))$

3. Prove this is well-defined, order-preserving, and additive

   Files: `RepresentationTheorem/Alternative/DirectCuts.lean` ($\sim$900 lines)
   Pros: Classical approach; more direct than grid method
Cons: Requires strict separation; still substantial code

Listing 5: Cuts-based representation theorem

```
theorem associativity_representation_cuts
   (α : Type*) [KnuthSkillingAlgebra α] [KSSeparation α]
   [KSSeparationStrict α] :
   ∃ Θ : α → R,
     (∀ a b : α, a ≤ b ↔ Θ a ≤ Θ b) ∧
     Θ ident = 0 ∧
     ∀ x y : α, Θ (op x y) = Θ x + Θ y
```

4.2.3 Approach 3: Hölder Embedding (Classical Route)

Assumptions: `NoAnomalousPairs` only!
   Technique: Leverage Eric Paul's formalized Hölder theorem for ordered semigroups:

1. Show `KnuthSkillingAlgebraBase` provides an `IsOrderedCancelSemigroup` instance

2. Translate our `AnomalousPair` definition to Eric's `anomalous_pair`

3. Apply Eric's `holder_not_anom`: $\neg\text{has\_anomalous\_pair} \Rightarrow \alpha \hookrightarrow (\mathbb{R}, +)$

4. Extract Θ from the multiplicative order isomorphism

   Files: `Separation/HolderEmbedding.lean` ($\sim$320 lines) + Eric's library ($\sim$1500 lines)
   Pros: Classical assumptions; shortest local code; most elegant
Cons: Depends on external library

Listing 6: Hölder-based representation theorem (classical route)

```
theorem representation_from_noAnomalousPairs
   (α : Type*) [KnuthSkillingAlgebraBase α] [NoAnomalousPairs α] :
   ∃ Θ : α → R,
     (∀ a b : α, a ≤ b ↔ Θ a ≤ Θ b) ∧
     Θ ident = 0 ∧
     ∀ x y : α, Θ (op x y) = Θ x + Θ y
```

4.3 The NoAnomalousPairs Condition

Definition 4.2 (Anomalous Pair). Elements a, b form an *anomalous pair* if for every $n > 0$:

$$(a^n < b^n < a^{n+1}) \quad \text{or} \quad (a^n > b^n > a^{n+1})$$

i.e., the iterates stay "squeezed" forever.

Definition 4.3 (NoAnomalousPairs). A K&S algebra has *no anomalous pairs* if no such pair exists:

$$\forall a\, b, \ \neg\text{AnomalousPair}(a, b)$$

7

This condition is from classical ordered semigroup theory (Alimov 1950, Fuchs 1963). It is **weaker** than `KSSeparation`:

Listing 7: KSSeparation implies NoAnomalousPairs (under IdentIsMinimum)

```
theorem noAnomalousPairs_of_KSSeparation_with_IdentMin
    [KSSeparation α] [IdentIsMinimum α] : NoAnomalousPairs α
```

Note: `IdentIsMinimum` ($\forall a$, ident $\leq a$) is natural for the K&S probability setting where `ident` represents the certain event.

4.4 Comparison of Approaches

Metric	Grid	Cuts	Hölder
Local lines of code	$\sim$4000	$\sim$1100	$\sim$420
Total lines (incl. deps)	$\sim$4000	$\sim$1100	$\sim$1900
Assumptions	Medium	Medium+	**Classical**
Self-contained	✓	✓	–
Follows K&S paper	✓	partial	–
Conceptual clarity	Low	Medium	**High**
Maintainability	Low	Medium	**High**

Recommendation: For practical use, the Hölder approach is preferred due to its minimal assumptions and elegant proof structure. For pedagogical purposes or K&S paper alignment, the cuts approach provides a good balance.

4.5 Unified Interface

All three approaches provide instances of a common interface:

Listing 8: Unified representation theorem interface

```
class HasRepresentationTheorem (α : Type*) [KnuthSkillingAlgebraBase α] :
    Prop where
  exists_representation :
    ∃ Θ : α → R,
      (∀ a b : α, a ≤ b ↔ Θ a ≤ Θ b) ∧
      Θ ident = 0 ∧
      ∀ x y : α, Θ (op x y) = Θ x + Θ y

-- Instances from each approach:
instance holder_hasRepresentationTheorem [NoAnomalousPairs α]
instance grid_hasRepresentationTheorem [KSSeparation α] [
    RepresentationGlobalization α]
instance cuts_hasRepresentationTheorem [KSSeparation α] [
    KSSeparationStrict α]
```

File: `RepresentationTheorem/Comparison.lean`

5 The Product Theorem (Appendix B)

Lean status: COMPLETE (no `sorry`, no `axiom`).

After Appendix A regrades $\oplus$ to real addition, K&S introduce a second scalar operation $\otimes$ for the direct product of independent lattices. Appendix B shows that $\otimes$ must be multiplication up to a single global scale constant.

In Lean we formalize this conclusion in three complementary ways (all sorry-free):

1. **Product-equation route** (closest to K&S's presentation): assume an additive order-isomorphism representation of $\otimes$ on $\mathbb{R}_{>0}$, i.e. $\Theta(x \otimes y) = \Theta x + \Theta y$, together with distributivity over $+$ (Axiom 3). This yields the Appendix B product equation for $\Psi := \Theta^{-1}$, which we solve to obtain $\Psi(x) = Ce^{Ax}$ and hence $(x \otimes y) = (xy)/C$. (Lean: `ProductTheorem/Main.lean`.)

2. **Lean-friendly "avoid Appendix A again" route**: do *not* assume a representation theorem for $\otimes$. Instead, package the regularity properties as `TensorRegularity` (associativity + injectivity of $t \mapsto 1 \otimes t$) and combine with distributivity (Axiom 3) to derive $(x \otimes y) = (xy)/C$ directly. (Lean: `ProductTheorem/AczelTheorem.lean`.)

3. **WARNING: Circular Reasoning Anti-Pattern**. A previous file `EventBridge.lean` was **deleted** because it contained circular reasoning: it defined `mulTensor := multiplication` and then "proved" multiplication equals scaled multiplication. **This proves nothing!** The correct approach (what `AczelTheorem.lean` does) is to start with an *abstract* tensor $\otimes$ satisfying Axioms 3–4 and prove that ANY such tensor equals scaled multiplication.

Key point: the Appendix A axiom `ident_le` (identity is the bottom element) is appropriate for $\oplus$ but structurally incompatible with $\otimes$ on $(0, \infty)$, where the identity is 1 and is *not* a minimum element. Our Appendix B development avoids this mismatch by never trying to model $\otimes$ as a "bottomed" additive monoid.

Axiom audit: We checked all `ProductTheorem/*.lean` files for `axiom`, `sorry`, or `admit`— none found. The injectivity requirement in `TensorRegularity` is *derived* from distributivity + commutativity via `TensorRegularity.of_assoc_of_distrib_of_comm`, not assumed.

5.1 K&S's Approach: Fibonacci Recurrence

For independent random variables, K&S derive that the measure function Ψ must satisfy:

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

where τ, ξ, η are independent real variables.

K&S prove the unique solution $\Psi(x) = C \cdot e^{Ax}$ via:

1. **2-term recurrence** (special case $\xi = \eta$): Shows $\Psi(\theta + na) = 2^n \Psi(\theta)$

2. **3-term recurrence** (Fibonacci-like): Produces golden ratio behavior

3. **Density argument**: Since b/a is irrational, offsets $mb - na$ approximate any real

This clever proof avoids continuity assumptions and works purely combinatorially.

5.2 Formalization Notes

- **K&S's Fibonacci proof**: COMPLETE. The full Fibonacci recurrence argument is formalized in `FibonacciProof.lean`, including:

  1. Golden ratio irrationality (`goldenRatio_irrational`)

9

2. Dense offsets $\{m - n\phi : m, n \in \mathbb{N}, n \geq 1\}$ in $\mathbb{R}$ (`dense_fib_linear_combinations`)

3. The 2-term and 3-term recurrences that produce Fibonacci behavior

- **KEY ACHIEVEMENT: Continuity derivation**. We prove that continuity follows from the ProductEquation + strict monotonicity + positivity, avoiding the need to assume continuity:

Listing 9: Deriving continuity from ProductEquation

```
theorem productEquation_strictMono_pos_continuous
    {Ψ : R → R} {ζ : R → R → R}
    (hProd : ProductEquation Ψ ζ)
    (hPos : ∀ x, 0 < Ψ x)
    (hMono : StrictMono Ψ) : Continuous Ψ
```

Proof idea: ProductEquation implies $\Psi(\theta + na) = 2^n \Psi(\theta)$ (doubling). Combined with the dense range $\{m \cdot C/2^k : m \geq 1, k \in \mathbb{N}\}$ where $C = \Psi(0)$, strict monotonicity forces Ψ to be continuous (no jumps possible).

- **Main theorem** (K&S Appendix B):

Listing 10: K&S Appendix B: StrictMono case

```
theorem ks_appendix_b_fibonacci_strictMono
    (Ψ : R → R) (ζ : R → R → R)
    (hProd : ProductEquation Ψ ζ)
    (hPos : ∀ x, 0 < Ψ x)
    (hMono : StrictMono Ψ) :
    ∃ C A : R, 0 < C ∧ ∀ x : R, Ψ x = C * Real.exp (A * x)
```

This theorem is **sorry-free** and combines the continuity derivation with `productEquation_solution_of_cont`

- **Mathlib support**: we use `Real.log_mul` and related properties of exp / log for the log-coordinate arguments.

- **Archived alternatives**: Two incomplete approaches (deriving StrictMono from Continuous, and deriving StrictMono from weak Monotone) are preserved in `FibonacciProofScratch.lean` for reference. These were not needed because K&S Appendix A already provides StrictMono from the order isomorphism representation.

6 The Variational Theorem (Appendix C)

K&S's Appendix C addresses entropy and divergence via a *variational approach*. This section contains a subtle but important error in the original paper that our formalization corrects.

6.1 The Variational Functional Equation

K&S consider a potential function $H(m)$ satisfying the functional equation:

$$H'(m_x \cdot m_y) = \lambda(m_x) + \mu(m_y) \tag{1}$$

where $H' = dH/dm$ and $m_x, m_y > 0$.

6.2 Path B: Lagrange Multipliers (Deriving the Functional Equation)

In the paper, Equation (1) is motivated by a Lagrange multiplier argument: K&S begin with the variational constraint form (their Eq. `vary2`), introduce multipliers (`vary3`), and then specialize to a two-dimensional x-by-y direct-product application. There, the direct-product rule gives a target value $m(x \times y) = m_x m_y$, and taking the coordinate derivative with respect to $m(x \times y)$ yields their Eq. `Hproduct`:

$$H'_{xy}(m_x m_y) = \lambda_1 f_1(m_x) + \lambda_2 f_2(m_y).$$

In Lean we formalize the *coordinate-derivative calculation* explicitly, without smuggling it in as an axiom. The key lemma is `lagrange_coordinate_deriv_eq_product` (Mettapedia/ProbabilityTheory/KnuthS which models "$\partial / \partial \, m(x \times y)$" as the 1D derivative along the update map $t \mapsto$ update $m \ (x, y) \ t$. It assumes only local differentiability of the relevant terms and a stationarity hypothesis stated as a `HasDerivAt ... 0 ...` condition (so this is a real derivative, not the default `deriv` value at non-differentiable points).

Listing 11: Path B: coordinate-derivative separation at a stationary point

```
theorem lagrange_coordinate_deriv_eq_product
    (H : Real -> Real) (g : X -> Real -> Real) (h : Y -> Real -> Real)
    (alpha : X -> Real) (beta : Y -> Real) (m : Prod X Y -> Real) (x0 : X)
        (y0 : Y)
    (hprod : m (x0, y0) = rowSum m x0 * colSum m y0)
    (hcrit : HasDerivAt (fun t => productLagrangian H g h alpha beta
            (Function.update m (x0, y0) t)) 0 (m (x0, y0))) :
    deriv H (rowSum m x0 * colSum m y0)
      = alpha x0 * deriv (g x0) (rowSum m x0)
        + beta y0 * deriv (h y0) (colSum m y0)
```

This establishes K&S's separated form *at a stationary point* of a separable Lagrangian. The global functional-equation (1) is then treated by Path A (below), which is the classification step that yields the logarithmic solution.

Fully honest stationarity. To avoid assuming stationarity as an axiom, we also provide a variant that derives the `HasDerivAt ... 0 ...` hypothesis from a local extremum via Fermat's theorem: `lagrange_coordinate_deriv_eq_product_of_isLocalExtr` (Mettapedia/ProbabilityTheory/KnuthSkilling/V We then package the result in the exact "K&S `Hproduct`" shape as `lagrange_Hproduct_eq_of_isLocalExtr`.

Universality/richness (Path B → Path A). Path B establishes a separated coordinate equation at a *particular* stationary point of a separable constrained problem. To treat this as a functional equation valid for all positive pairs (m_x, m_y) (the input required by Path A), K&S rely on an informal "generality across applications" premise: the same potential is intended to apply uniformly across a sufficiently rich class of inference problems. In Lean we make this bridge assumption explicit as the structure `KSVariationalGenerality` (Mettapedia/ProbabilityTheory/KnuthSkilling/VariationalThe so nothing is smuggled: either you assume it, or you do not get a global functional equation.

6.3 K&S's Differentiability Claim (and Why It's Wrong)

K&S assert that "*any variational potential must ... be differentiable at least once*" because it applies to arbitrarily small perturbations. **This claim is false.**

Example 6.1 (Variational minimum without differentiability). Consider $H(x) = |x|$ on $\mathbb{R}$. The point $x_\star = 0$ is a global minimizer: for every small perturbation h,

$$H(0 + h) - H(0) = |h| \geq 0.$$

This is exactly the variational inequality "no small perturbation decreases H". But H is *not differentiable* at 0: the one-sided derivatives are -1 and $+1$.

Example 6.2 (K&S-shaped example on $(0, \infty)$). Consider $H(m) = |\log m|$ on $(0, \infty)$. It has a unique minimum at $m_\star = 1$, but is not differentiable there. This directly refutes K&S's claim in their specific modeling domain.

Nonsmooth variational problems are not exotic: absolute value, max functions, and ℓ_1 penalties are core tools in modern optimization. Rockafellar–Wets [7] explicitly stress that max/min operations often fail to preserve smoothness. These counterexamples are formalized in `Mettapedia/ProbabilityTheory/`

6.4 The Correct Approach: Measurability

6.4.1 Minimal assumptions (sound formal statement)

It is useful to separate two logically distinct issues:

1. *What hypotheses are needed for the functional-equation* classification $H'(m) = B + C \log m$?

2. *What hypotheses are needed to interpret H' as an actual derivative of some potential H?*

Our Lean formalization provides both formulations, and makes each assumption explicit. The minimal package for the *classification step* is:

(C0) Domain / log-coordinates. We work on positive masses $m > 0$ so that $\log m$ is defined and $\log(m_x m_y) = \log m_x + \log m_y$ can be used.

(C1) Variational functional equation. The separation-of-variables equation $H'(m_x m_y) = \lambda(m_x) + \mu(m_y)$ holds for all $m_x, m_y > 0$ (Equation (1)).

(C2) Anti-pathology regularity. Some regularity is required to rule out Hamel-basis pathologies of Cauchy's equation after the $u = \log m$ change of variables. We take **Borel measurability** of H' as the core hypothesis.

This is exactly the statement proved by our main Lean theorem `variationalEquation_solution_measurable` (`Mettapedia/ProbabilityTheory/KnuthSkilling/VariationalTheorem.lean`). It *does not* assume that H' comes from a differentiable potential.

If one also wants a theorem phrased in terms of an underlying potential H, one adds an optional hypothesis:

(C3) Derivative interpretation (optional). Either set $H' = \text{deriv}\, H$, or assume $\text{HasDerivAt}\, H\, (H'(m))\, m$ for all $m > 0$.

In Lean, (C3) is captured by the corollary `variationalEquation_solution` (the `deriv` version), which automatically discharges (C2) via `measurable_deriv`.

Our formalization makes the regularity assumption *explicit*. The key insight is that the functional equation (1), after the log-coordinate change $u = \log m$, reduces to **Cauchy's additive functional equation**:

$$f(u + v) = f(u) + f(v)$$

Without regularity, Cauchy's equation has pathological non-linear solutions (via Hamel basis constructions, assuming the Axiom of Choice). To obtain the unique linear solution $f(t) = Ct$, *some* regularity is required. The existence of non-linear additive solutions is shown by a standard Hamel-basis construction; we formalize this as a counterexample generator in `Mettapedia/ProbabilityTheory/KnuthSki`

The standard hierarchy of sufficient conditions includes:

- Measurable (weakest commonly used)

- Continuous at one point

- Bounded on any interval

- Monotone on any interval

Regularity gate for Cauchy	Forces linearity?	Automatic for deriv?	Comment
Borel measurable	yes	yes	Our chosen explicit replac for K&S "blurring"; discharg `measurable_deriv`.
Continuous at one point	yes	no	Classical Shannon/Faddeev-sty pothesis; also implies measural
Monotone on an interval	yes	no	Order-theoretic alternative; in measurability (hence works wi proof).
Bounded on an interval	yes	no	Another standard gate exc Hamel-basis pathologies.

Our choice: We use **Borel measurability** as the regularity assumption because:

1. It is among the weakest standard conditions

2. It is **automatic** for derivatives: if H is differentiable, then $H' = \operatorname{deriv} H$ is Borel measurable (via `measurable_deriv` in mathlib)

3. It matches what K&S's convolution "blurring" argument implicitly requires (one cannot even write the convolution integral without measurability/integrability hypotheses)

Listing 12: Main Variational Theorem (measurable version)

```
theorem variationalEquation_solution_measurable
    (H' : R → R) (lam mu : R → R)
    (hMeas : Measurable H')
    (hV : VariationalEquation H' lam mu) :
    ∃ B C : R, ∀ m : R, 0 < m →
      H' m = B + C * log m
```

Listing 13: Corollary for actual derivatives (regularity automatic)

```
theorem variationalEquation_solution
    (H : R → R) (lam mu : R → R)
    (hV : VariationalEquation (deriv H) lam mu) :
    ∃ B C : R, ∀ m : R, 0 < m →
      deriv H m = B + C * log m
```

6.5 Alternative: Monotonicity (K&S-Spirited)

An alternative regularity assumption that mirrors Appendix A's order-theoretic style is **monotonicity**:

- Monotone $\Rightarrow$ Measurable $\Rightarrow$ log form

- Monotonicity is physically motivated: entropy has decreasing marginal value

- This parallels Appendix A where order-preservation is central

However, monotonicity is *stronger* than measurability (not all measurable functions are monotone). Our core theorem uses the weaker assumption; monotonicity can be stated as a corollary.

6.6 Shore–Johnson (1980) comparison

Shore–Johnson's axioms for maximum entropy inference (IEEE Trans. IT, 1980) include a "system independence" requirement that can be viewed as another route to the same logarithmic structure. In a discrete, summand-level setting:

- Assume an atomic divergence $d(p, q)$ satisfies $d(0, q) = 0$ and that the sum $\sum_i d(p_i, q_i)$ is additive over product distributions (an SJ4-style hypothesis).

- Dirac-delta test distributions then force a multiplicative Cauchy equation for $g(q) := d(1, q)$ on the probability domain $0 < q \leq 1$.

- With a regularity gate (we again take measurability), this implies $g(q) = C \log q$.

We formalize this "Dirac extraction $\rightarrow$ log" lemma as `d_one_eq_const_mul_log_of_measurable` in `Mettapedia/ProbabilityTheory/KnuthSkilling/ShoreJohnson.lean`. This provides a clean comparison point between Shore–Johnson's system-independence intuition and K&S Appendix C: both rely on the same functional-equation rigidity, and both require an explicit anti-pathology hypothesis to exclude Hamel-basis solutions.

We also formalize a small but useful *KL uniqueness* lemma ("uniqueness up to scale, within a ratio-form class"):

- If $g : (0, \infty) \rightarrow \mathbb{R}$ satisfies $g(xy) = g(x) + g(y)$ and is Borel measurable, then $g(x) = C \log x$.

- Hence any ratio-form atom divergence $d(w, u) = w\, g(w/u)$ (for $w, u > 0$) must be a constant multiple of the KL atom $w \log(w/u)$.

- Without the regularity gate, uniqueness fails: we exhibit a multiplicative-additive (Hamel-basis) counterexample and the induced ratio-form d is *not* a constant multiple of KL.

These results are proved in `Mettapedia/ProbabilityTheory/KnuthSkilling/ShoreJohnsonKL.lean`. For convenience, the project re-exports the key Shore–Johnson lemmas we use in `Mettapedia/ProbabilityTheory/` We then connect the resulting atom-level KL identity to the project's finite KL divergence definition on `ProbDist` (Section 8 layer) in `Mettapedia/ProbabilityTheory/KnuthSkilling/ShoreJohnsonBridge.lean`.

6.7 The Entropy/Divergence Form

From the variational theorem, integrating $H'(m) = B + C \log m$ yields:

$$H(m) = A + Bm + C(m \log m - m)$$

With appropriate choices of constants ($C = 1$, $B = -\log u$, $A = u$), this specializes to the **Kullback–Leibler divergence**:

$$D(w\|u) = \sum_i \left(u_i - w_i + w_i \log \frac{w_i}{u_i} \right)$$

This is formalized in `Mettapedia/ProbabilityTheory/KnuthSkilling/Divergence.lean` with proofs of non-negativity, zero iff equal, and asymmetry. For probability distributions (equal total mass), the linear terms $\sum_i (u_i - w_i)$ cancel, leaving the usual $\sum_i w_i \log(w_i/u_i)$.

6.8 Summary: What K&S Should Have Said

K&S informally assert that a variational potential must be C^1 because it applies to arbitrarily small perturbations. This is not a logical consequence of local optimality. Standard nonsmooth variational analysis replaces gradients with subgradients (Rockafellar–Wets; Clarke).

In our formalization, we **state measurability explicitly** as an analytic assumption. This is also exactly what is needed to rule out pathological Cauchy solutions and obtain the unique logarithmic form. No circularity: we assume (functional equation + measurability), and *derive* the log form.

7 Probability and Conditional Inference

K&S's Axiom 5 (chaining associativity) extends the framework to conditional probability. We formalize the chain product rule, Bayes' theorem, and the ratio-of-measures representation in `Mettapedia/ProbabilityTheory/KnuthSkilling/ConditionalProbability/Basic.lean`.

8 Information and Entropy (Section 8)

K&S Section 8 specializes the variational potential to (i) information/divergence for probability distributions and (ii) Shannon entropy. We formalize this in `Mettapedia/ProbabilityTheory/KnuthSkilling/Infor`

For comparison (and to honor the historical development), we also formalize classical axiomatic characterizations of Shannon entropy *independently of K&S*:

- Shannon's original 1948 axioms + uniqueness (Appendix 2 of Shannon): `Mettapedia/InformationTheory/Sh` (sorry-free).

- Faddeev (1956) and Shannon–Khinchin (1957) axiom systems: `Mettapedia/InformationTheory/ShannonEn` and `Mettapedia/InformationTheory/ShannonEntropy/ShannonKhinchin.lean` (both WIP).

- The equivalence *glue* is kept in a separate file so the Shannon–Khinchin development does not depend on Faddeev: `Mettapedia/InformationTheory/ShannonEntropy/Equivalence.lean`.

Highlights:

- **Finite and extended KL**: a finite KL definition under "support positivity" and an extended `ENNReal`-valued version returning $\top$ when $p_i > 0$ but $q_i = 0$ (`klDivergence`, `klDivergenceTop`).

- **Mathlib KL bridge (finite → general)**: we show our `klDivergenceTop` on `Fin n` agrees with mathlib's measure-theoretic `InformationTheory.klDiv` on the associated Dirac-sum measures (`klDiv_toMeasureTop_eq_klDivergenceTop` in `Mettapedia/ProbabilityTheory/KnuthSkilling/`). This provides a clean "infinite/continuum" variant by reusing mathlib's general KL divergence.

- **Mathlib KL bridge (countable discrete)**: for sequences on $\mathbb{N}$ we connect the K&S countable divergence sum `divergenceInf` to `InformationTheory.klDiv` for Dirac-sum measures (`Seq.klDiv_toMeasureTop_eq_divergenceInfTop` and `Seq.klDiv_toMeasureTop_eq_ofReal_divergenceInf` in `Mettapedia/ProbabilityTheory/KnuthSkilling/DivergenceMathlib.lean`).

- **Shannon properties**: entropy of the uniform distribution is $\log n$ (`shannonEntropy_uniform`), strict monotonicity in n for uniform distributions (`shannonEntropy_uniform_strictMono`), grouping/additivity in a binary refinement (`shannonEntropy_grouping_probDist`), and continuity on the interior of the simplex (`shannonEntropy_continuous_on_interior`).

9 Conclusion

The K&S program successfully derives probability calculus from symmetries. Our formalization confirms their approach is sound, while making explicit:

- Linear order (implicit in their trichotomy arguments)

- Separation property (implicit in rational approximation)

- Archimedean property (derivable from separation)

- Commutativity (derivable from NoAnomalousPairs)

Key contributions of this formalization:

1. **Three independent proof paths**: Grid-based (K&S style), Dedekind cuts (classical), and Hölder embedding (via Eric Paul's library).

2. **Discovery of minimal assumptions**: The `NoAnomalousPairs` condition from classical ordered semigroup theory (Alimov 1950) is *sufficient* for the representation theorem and is *weaker* than K&S's separation property.

3. **Connection to classical mathematics**: K&S's work fits within the century-old tradition of Hölder embedding theorems for ordered algebraic structures. Eric Paul's Lean 4 formalization of this classical theory provides a clean, minimal proof path.

4. **Unified interface**: All three proof approaches provide instances of a common `HasRepresentationTheorem` typeclass, enabling modular use.

Practical recommendation: For new formalizations building on K&S, the Hölder approach is preferred:

- Uses classical assumptions (`NoAnomalousPairs` only)

- Shortest local codebase ($\sim$420 lines vs. $\sim$4000 for grid approach)

- Most mathematically elegant

- Leverages Eric Paul's battle-tested library

The formalized representation theorem provides a constructive, computationally meaningful foundation for inference, now available via multiple independent proof paths.

A Corrections to the K&S Axioms

Formalization revealed several implicit assumptions and derivable properties in K&S. We document these findings with a generous reading of their intent.

A.1 The Implicit Linear Order

K&S never explicitly state that elements are totally ordered, yet their proofs rely on trichotomy: for any x, y, exactly one of $x < y$, $x = y$, or $x > y$ holds.

Evidence: K&S line 1339:

"By interchanging x and y in axiom 1, the same relationship holds when '$<$' is replaced throughout by '$>$', and replacement by '$=$' holds trivially. So, in effect, the axiom makes a three-fold assertion"

This implicitly assumes trichotomy, which we make explicit via `LinearOrder`.

A.2 Commutativity: Derivable, But Only With Separation

K&S state (Section 3.1):

"We also confirm that commutativity is not a necessary assumption. Rather, commutativity of measure is imposed by the associativity and order required of a scalar representation."

Our finding: Commutativity is *not* derivable from Axioms 0–2 alone. It requires the *separation property*.

Proof: Theorem in `RepresentationTheorem/Core/SeparationImpliesCommutative.lean`:

```
theorem op_comm_of_associativity
    (α : Type*) [KnuthSkillingAlgebra α] [KSSeparation α] :
    ∀ x y : α, op x y = op y x
```

Generous reading: K&S likely intended "associativity and order" to include the implicit separation property that enables their rational approximation construction.

A.3 Archimedean: Derivable from Separation

K&S's proofs implicitly use the Archimedean property: for any $a > $ ident and x, there exists n such that $a^n > x$.

Our finding: The Archimedean property is *derivable* from separation, not an independent axiom.

Proof: Theorem in `Separation/SandwichSeparation.lean`:

```
1  theorem op_archimedean_of_separation
2      [KSSeparation α] (a x : α) (ha : ident < a) :
3      ∃ n : N, x < iterate_op a n
```

This is formalization-revealed structure, not a correction to K&S.

A.4   The Separation Property Itself

K&S's Appendix A construction implicitly assumes we can "separate" any two distinct values using rational powers of a base. We make this explicit as KSSeparation.

Evidence: K&S's "induction to more than one type of atom" (Appendix A, starting line 1437) constructs rationals u/v to approximate new atom values. This requires separation.

A.5   Summary of Assumptions (Three Tiers)

For the representation theorem, we have three tiers of assumptions:

Tier 1 (Base axioms—always required):

1. **LinearOrder** (implicit in K&S)

2. **Associativity, Identity, Strict Monotonicity** (K&S Axioms 1–2)

Tier 2 (Regularity—one of the following):

- **NoAnomalousPairs** (classical, used by Hölder approach)

- **KSSeparation** (K&S's implicit assumption, used by grid/cuts approaches)

- **KSSeparation + KSSeparationStrict** (strongest, used by cuts approach)

Relationship:

$$\text{KSSeparation} + \text{IdentIsMinimum} \Rightarrow \text{NoAnomalousPairs}$$

From NoAnomalousPairs alone, we *derive*:

- Commutativity (via Eric Paul's not_anomalous_pair_commutative)

- Hölder embedding into $(\mathbb{R}, +)$ (via Eric's holder_not_anom)

- Additive representation $\Theta : \alpha \to \mathbb{R}$

Key finding: NoAnomalousPairs is sufficient for the representation theorem and connects to classical ordered semigroup theory.

B   Philosophical Context: Cox, Kolmogorov, and K&S

We compare three foundational approaches to probability, all formalized in Mettapedia.

B.1 Kolmogorov (1933): Measure-Theoretic

Kolmogorov's axioms define probability as a normalized measure on a σ-algebra. Formalized in Mathlib as `IsProbabilityMeasure`:

Listing 14: Kolmogorov probability in Mathlib

```
class IsProbabilityMeasure (μ : Measure Ω) : Prop where
  measure_univ : μ Set.univ = 1
```

Key properties (all in Mathlib):

- σ-additivity: $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i

- Monotonicity: $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

- Normalization: $\mu(\Omega) = 1$, $\mu(\emptyset) = 0$

Comparison with K&S: Kolmogorov *axiomatizes* additivity; K&S *derives* it from associativity symmetry.

B.2 Cox (1946, 1961): Logical Plausibility

Formalization: See `Cox/Basic.lean`

B.2.1 Philosophical Setting

Cox takes an *epistemological* starting point: How should a rational agent assign degrees of belief to propositions? Unlike Kolmogorov (who starts with sets and measures), Cox starts with propositions and asks what constraints logic imposes on plausibility assignments.

The key philosophical move: rather than axiomatizing what probability *is*, Cox axiomatizes how plausibility *behaves* and proves probability emerges uniquely.

B.2.2 The Cox Axioms

1. **Real-valued plausibility**: Each proposition A given background B has a plausibility $p(A|B) \in [0, 1]$.

2. **Order preservation**: If A is more plausible than A' given B, then $p(A|B) > p(A'|B)$.

3. **Product rule** (the key axiom): There exists *some* function F such that:

$$p(A \wedge B|C) = F(p(A|C), p(B|A \wedge C))$$

4. **Negation rule**: There exists *some* function G such that:

$$p(\neg A|B) = G(p(A|B))$$

The crucial point: Cox does *not* assume $F(x, y) = xy$ or $G(x) = 1 - x$. He proves these are the *only* possibilities (up to reparametrization).

B.2.3 The Derivation

The associativity of conjunction—$(A \wedge B) \wedge C \equiv A \wedge (B \wedge C)$—forces:

$$F(F(x, y), z) = F(x, F(y, z))$$

This is the same functional equation as K&S Axiom 2! Combined with **continuity** of F, it forces F to be multiplication after a monotone reparametrization Θ:

$$F(x, y) = \Theta^{-1}(\Theta(x) \cdot \Theta(y))$$

The standard choice $\Theta = \text{id}$ gives the probability product rule $p(A \wedge B | C) = p(A | C) \cdot p(B | A \wedge C)$.

B.2.4 Continuity vs. Separation: A Key Contrast

Both Cox and K&S face the same challenge: the associativity equation $F(F(x, y), z) = F(x, F(y, z))$ has pathological solutions (Cauchy's discontinuous additive functions). They rule these out differently:

	Cox	**K&S**
Setting	$F : \mathbb{R} \to \mathbb{R} \to \mathbb{R}$	Abstract algebra α
Regularity	F is **continuous**	**Separation** property
Domain	Already on $\mathbb{R}$	Must embed into $\mathbb{R}$

What they share: Both conditions rule out pathological non-measurable solutions.
How they differ:

- **Continuity** (Cox): A topological condition on $\mathbb{R}$. Directly invokes the topology of the reals.

- **Separation** (K&S): An algebraic condition on abstract α. For any $x < y$, there exist n, m with $x^m < a^n \leq y^m$. This is *density of rational powers*—no topology needed!

Key insight: Separation and continuity are *similar conditions in different settings*.
The precise relationship:

1. **From α to $\mathbb{R}$**: Separation + base axioms $\Rightarrow$ embedding $\Theta : \alpha \to \mathbb{R}$ exists. Once embedded, the operation becomes addition on $\text{Im}(\Theta) \subseteq \mathbb{R}$, which is automatically continuous.

2. **On $\mathbb{R}$**: Any continuous, monotone, associative $F : \mathbb{R} \to \mathbb{R} \to \mathbb{R}$ satisfies separation. Why? Separation on $(\mathbb{R}, +)$ says: for $x < y$, $\exists n, m$ with $mx < na \leq my$. This is just $x < (n/m) \cdot a \leq y$ for some rational—density of $\mathbb{Q} \cdot a$ in $\mathbb{R}$.

3. **Failure mode**: What if separation *fails*? Then no embedding into $\mathbb{R}$ exists at all.

Counterexample: `NatProdLex` ($\mathbb{N} \times \mathbb{N}$ with lexicographic order and componentwise addition) satisfies the base K&S axioms but *fails* separation. The gap between $(0, k)$ and $(1, 0)$ cannot be bridged by rational powers. Consequently, no order-preserving additive embedding into $\mathbb{R}$ exists.
  See: `RepresentationTheorem/Counterexamples/`
    `ProductFailsSeparation.lean`
  Connection theorem: `cox_comm_of_additiveOrderIsoRep` in
    `Literature/CoxTheorem.lean`

B.3 K&S (2012): Lattice Symmetries

K&S unify both approaches via lattice symmetries:

- Like Kolmogorov: works with lattices (Boolean for sets, more general for logic)

- Like Cox: derives rules from functional equations, not axiomatizes them

- Beyond both: minimal assumptions (no continuity, no differentiability)

Summary of relationships:

	Kolmogorov	Cox	K&S
Additivity	Axiom	Derived	Derived
Commutativity	Assumed	Derived	Derived
Continuity	Needed	Assumed	Not needed
Domain	σ-algebra	Propositions	Lattice

C The Direct Cuts Approach

K&S prove the representation theorem via **inductive extension**: given a representation for n atoms, extend it to $n + 1$ atoms by finding the unique real value for the new atom.

Our formalization provides **both** proofs:

1. **K&S's approach** (grid/induction): `RepresentationTheorem/Main.lean`
   Uses `RepresentationGlobalization` typeclass machinery.

2. **Direct cuts** (our alternative): `RepresentationTheorem/Alternative/Main.lean`
   Constructs Θ via Dedekind cuts explicitly. Smaller and more direct.

Both prove the same theorem; the cuts version is more concise.

C.1 Connection to Hölder and Alimov

The representation theorem fits within a classical tradition of embedding theorems for ordered algebraic structures:

Theorem	Structure	Result
Hölder (1901)	Archimedean ordered abelian groups	Embed in $(\mathbb{R}, +)$
Alimov (1950)	Cancellative ordered semigroups with NAP	Commutative + embed in $(\mathbb{R}, +)$
Fuchs (1963)	Textbook treatment of the above	Classical reference
Paul (2024)	**Lean 4 formalization of the above**	**OrderedSemigroups**
K&S + Separation	Ordered monoids with separation	Embed in $(\mathbb{R}, +)$

Key insight: Eric Paul's OrderedSemigroups library (github.com/ericluap/OrderedSemigroups) provides a **complete Lean 4 formalization** of the Alimov-Hölder embedding theorem:

Listing 15: Eric Paul's Hölder theorem (from OrderedSemigroups)

```
theorem holder_not_anom [LinearOrder α] [Semigroup α]
    [IsOrderedCancelSemigroup α] [Pow α N+] [PNatPowAssoc α]
    (h : ¬has_anomalous_pair (α := α)) :
    ∃ G : Subsemigroup (Multiplicative R), Nonempty (α ≃*o G)
```

This theorem states: if a linearly ordered cancellative semigroup has no anomalous pairs, it embeds (as an ordered multiplicative isomorphism) into $(\mathbb{R}, +)$.

Our contribution: We show that `KnuthSkillingAlgebraBase` satisfies all the hypotheses of Eric's theorem:

- `Semigroup`: mul := op, mul_assoc := op_assoc

- `IsOrderedCancelSemigroup`: From `op_strictMono_left/right`

- `Pow` $\mathbb{N}^+$: $x^n := \text{iterate\_op}\, x\, n$

- `PNatPowAssoc`: Requires commutativity, which Eric proves follows from ¬has_anomalous_pair!

Thus, `NoAnomalousPairs` alone suffices for the representation theorem—no `KSSeparation` needed!
Files:

- `Separation/HolderEmbedding.lean`: Bridge to Eric's library

- `Separation/AnomalousPairs.lean`: NoAnomalousPairs definition

- `RepresentationTheorem/Comparison.lean`: Unified interface

C.2 The A/B/C Partition

K&S partition grid points when extending the representation to a new atom d:

- **Set A**: Points below some power of d: $\mu(F, r) < d^u$

- **Set B**: Points equal to some power of d: $\mu(F, r) = d^u$

- **Set C**: Points above some power of d: $\mu(F, r) > d^u$

C.3 Dedekind Cut Property (Our Reformulation)

When $B = \emptyset$ (the generic case), we show that $\delta := \Theta(d)$ is determined as a Dedekind cut:

Listing 16: Dedekind cut tightness lemma

```
lemma delta_cut_tight (hB_empty : ∀ r u, r ∉ extensionSetB F d u) :
    ∀ ε > 0, ∃ rA rC uA uC,
      rA ∈ extensionSetA F d uA ∧
      rC ∈ extensionSetC F d uC ∧
      |stat(rC, uC) - δ| < ε ∧
      |δ - stat(rA, uA)| < ε
```

Meaning: $\delta = \sup(\text{A-statistics}) = \inf(\text{C-statistics})$, a tight Dedekind cut.

C.4 Why This Works

The separation property (`KSSeparation`) guarantees that for any $x < y$, we can find exponents that "separate" them. This provides the density needed for the Dedekind cut to uniquely determine δ.

Key insight: The K&S proof is essentially constructing Dedekind cuts to embed the abstract algebra into $\mathbb{R}$, one atom at a time.

D   The Probability Hypercube

We formalize a multi-axis classification of probability theories, placing K&S among its neighbors.

D.1   Hypercube Axes

Each axis represents a binary or multi-valued property of probability theories.
   Formalization: See `Hypercube/Basic.lean`

- **Commutativity**: $P(A \cap B) = P(B \cap A)$? (Classical: yes; Quantum: no)

- **Precision**: Point-valued or interval-valued plausibilities?

- **Additivity**: $P(A \cup B) = P(A) + P(B)$ for disjoint A, B?

- **Lattice structure**: Boolean, orthomodular, distributive?

- **Separation**: Can distinct values be separated by rationals?

D.2   Named Vertices

Key probability theories correspond to vertices of this hypercube:

Framework	Result	Comm.	Precise	Add.	Lattice	Method
Kolmogorov	Classical prob.	Yes	Yes	Axiom	Boolean	Axiom
Cox	Classical prob.	Yes	Yes	Derived	Boolean	Continuity
K&S (Boolean)	Classical prob.	Yes	Yes	Derived	Boolean	Separation
K&S (general)	Lattice plaus.	Yes	Yes	Derived	Distrib.	Separation
de Finetti	Subjective prob.	Yes	Yes	Finite	Boolean	Coherence
Dempster-Shafer	Belief functions	Yes	No	Sub	Boolean	Combination
Walley	Imprecise prob.	Yes	No	Sub	Boolean	Coherence
von Neumann	Quantum prob.	No	Yes	Axiom	Orthomod.	Axiom

Key insight: Kolmogorov, Cox, and K&S are *different vertices* (different derivation methods) that *converge* to the same theory on Boolean lattices. They differ in what they assume vs. derive:

- **Kolmogorov**: Axiomatizes additivity directly (no derivation)

- **Cox**: Derives additivity from continuity of plausibility functions

- **K&S**: Derives additivity from separation (no continuity needed)

D.3   Distributive vs. Boolean: What's Different?

K&S on general distributive lattices differs from classical probability in one key way: **no negation required**.

- **Boolean lattice**: Every element a has a complement $\bar{a}$ with $a \vee \bar{a} = \top$ and $a \wedge \bar{a} = \bot$. This gives the "negation" rule $P(\bar{A}) = 1 - P(A)$.

- **Distributive lattice**: No complements required. K&S axioms never mention negation—this is the "first fruit of our minimalist approach" (K&S).

When does this matter? Problems where not all combinations of states are allowed. Rather than "padding" a distributive lattice to Boolean, K&S works directly on the natural structure.

Literature: Valuations on distributive lattices appear in Klain & Rota's *Introduction to Geometric Probability* (cited by K&S) and the theory of lattice valuations in combinatorics.

Legend:

- **Comm.**: Is the conjunction operation commutative? ($P(A \cap B) = P(B \cap A)$)

- **Precise**: Single-valued (Yes) vs. interval-valued credal sets (No)

- **Additive**: Full σ-additivity (Yes), finite additivity (Finite), or subadditivity (Sub)

- **Lattice**: Boolean (σ-algebra), distributive (K&S), or orthomodular (quantum)

D.4   Novel Theories

The hypercube suggests unexplored combinations:

- **Imprecise K&S**: K&S with interval-valued plausibilities (credal sets)

- **Quantum D-S**: Dempster-Shafer on orthomodular lattices

- **Non-commutative classical**: Classical probability without $P(A \cap B) = P(B \cap A)$

  Formalization: See `Hypercube/NovelTheories.lean` and `Hypercube/NeighborTheories.lean`

E   To-Do: Open Gaps in Formalization

E.1   Event Lattice to Plausibility Scale

Status: resolved.

The bridge is implemented in `Model.lean` as:

```
structure KSModel (E S : Type*)
    [PlausibilitySpace E] [KnuthSkillingAlgebraBase S] where
  v : E → S
  mono : Monotone v
  v_bot : v ⊥ = ident
  v_sup_of_disjoint : ∀ {a b}, Disjoint a b →
    v (a ⊔ b) = op (v a) (v b)
```

The scale S is equipped with its K&S algebra structure from the start. The representation theorem tells us this algebra embeds into $(\mathbb{R}, +)$, so using $\mathbb{R}_{\geq 0}$ with addition as our scale is canonical.

`KSModelWithRepresentation` extends this by composing with the representation $\Theta : S \to \mathbb{R}$, yielding the full pipeline:

$$E \xrightarrow{v} S \xrightarrow{\Theta} \mathbb{R}$$

Concrete example: `Examples/CoinFlip.lean` demonstrates a coin flip (Bernoulli trial) with 4-element Boolean algebra mapping through $\mathbb{R}_{\geq 0}$ to classical probability in $\mathbb{R}$.

E.2 Appendix B: Product Theorem

Status: COMPLETE (no `sorry`, no `axiom` smuggling).

Lean proves that any $\otimes : \mathbb{R}_{>0} \to \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfying distributivity over $+$ (Axiom 3) and mild regularity (associativity, plus injectivity of $t \mapsto 1 \otimes t$) is multiplication up to a global scale constant. This is formalized as `TensorRegularity` in `ProductTheorem/AczelTheorem.lean`.

WARNING: Circular Reasoning Anti-Pattern. A previous file `EventBridge.lean` was **deleted** because it contained circular reasoning. It claimed to "bridge" event-level to scalar-level by defining `mulTensor := multiplication` and then "proving" multiplication equals scaled multiplication. **This proves nothing!**

The error was confusing "multiplication satisfies these axioms" with "any tensor satisfying these axioms equals multiplication." The former is trivial field theory; the latter is the actual theorem.

The **correct approach** (what `AczelTheorem.lean` does):

1. Start with an *abstract* tensor $\otimes$ satisfying K&S Axioms 3–4

2. Prove that ANY such tensor must equal scaled multiplication

3. The tensor's identity is *derived*, not assumed

E.3 Cauchy's Functional Equation

Status: proved and used.

We prove the needed Cauchy lemma as `continuous_additive_eq_mul` in `Mettapedia/ProbabilityTheory/Knu` using Mathlib's continuous additive $\Rightarrow \mathbb{R}$-linear infrastructure.

E.4 Concrete Examples

Status: demonstrated.

See `Mettapedia/ProbabilityTheory/KnuthSkilling/Examples/` for:

- `CoinFlip.lean`: Full K&S pipeline on a coin flip (Bernoulli trial, Durrett Example 1.6.12). Demonstrates: 4-element Boolean algebra $\to \mathbb{R}_{\geq 0} \to \mathbb{R}$. Verifies: $P(\emptyset) = 0$, $P(\Omega) = 1$, additivity $P(H \cup T) = P(H) + P(T)$.

- `ThreeElementChain.lean`: Minimal non-Boolean distributive lattice (3-element chain).

- `OpenSets.lean`: Open sets of $\mathbb{R}$ (distributive, typically not complemented).

References

[1] Kevin H. Knuth and John Skilling. *Foundations of Inference.* Axioms, 1(1):38–73, 2012.

[2] Eric Paul. *OrderedSemigroups: Formalization in Lean 4.* GitHub repository, 2024. `https://github.com/ericluap/OrderedSemigroups` See also: `https://ericluap.github.io/`

[3] N. G. Alimov. On ordered semigroups. *Izv. Akad. Nauk SSSR Ser. Mat.*, 14:569–576, 1950.

[4] L. Fuchs. *Partially Ordered Algebraic Systems.* Pergamon Press, 1963.

[5] O. Hölder. Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe*, 53:1–64, 1901.

[6] D. Binder. Non-Anomalous Semigroups and Real Numbers. arXiv:1607.05997, 2016.

[7] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, 1998.

[8] F. H. Clarke. *Optimization and Nonsmooth Analysis.* SIAM, 1990.