Unification: Bayes, Thermo, Quotient, and Factor Maps

Hutter RNN Project

January 31, 2026

1 The Five Pieces

1.1 Bayes (Probability)

Luck as the reciprocal of probability:

$$\lambda(e) := \frac{1}{p(e)}, \qquad \Lambda(e) := \log_2 \lambda(e) = -\log_2 p(e)$$

Additivity of log-luck:

$$\Lambda(e_1, \ldots, e_n) = \sum_{i=1}^{n} \Lambda(e_i \mid e_1, \ldots, e_{i-1})$$

1.2 Thermo (Microstates)

Event space cardinality:

$$|E| = \prod_{i=1}^{k} |E_i|$$

Information content:

$$I(E) = \log_2 |E| = \sum_{i=1}^{k} \log_2 |E_i|$$

Remaining microstates after observation:

$$\text{Remaining} = |E| \cdot p = \frac{|E|}{\lambda}$$

1.3 Quotient (Equivalence Classes)

Layer quotient:

$$q_i := \frac{|\text{distinguishable inputs}|}{|\text{distinguishable outputs}|}$$

Cumulative quotient:

$$Q_n := \prod_{i=1}^{n} q_i$$

After n layers: $|E|/Q_n$ distinguishable states remain.

1.4 E → N Embedding (Mixed-Radix)

Bijection from product space to natural numbers:

$$(e_1, \ldots, e_k) \mapsto \sum_{i=1}^{k} e_i \cdot \prod_{j>i} |E_j|$$

Inverse via repeated div/mod.

1.5 Factor Maps

Factoring map $\phi_i : E \to A_i \times B_i$, decoder $d_i : A_i \times B_i \to T$.

Translator between factorizations:

$$\tau_{1\to 2} := \rho_2 \circ d_1$$

Commuting law:

$$d_2 \circ \tau_{1\to 2} = d_1$$

2 The Unifying Diagram

```
        E (events)
        |
   phi  |  factor map
        v
  E_1 x E_2 x ... x E_k
        |
 encode |  mixed-radix
        v
        N
        |
   Q    |  quotient = luck
        v
      N/Q (remaining microstates)
```

Reading top to bottom:

1. **Factor**: Decompose event space into product of factors
2. **Encode**: Map product to natural number (mixed-radix polynomial)
3. **Quotient**: Collapse by equivalence relation (layer compression)

3 The Key Identity

The quotient IS the luck.

$$Q = \lambda$$

This unifies:

- Bayesian inference (updating beliefs via likelihood)
- Thermodynamics (counting remaining microstates)
- Neural network layers (dimensionality reduction)
- Arithmetic coding (interval narrowing)

4 Bayes ↔ Thermo Duality

Bayes	**Thermo**	**Relation**
p	$1/\lambda$	probability = 1/luck
$-\log p$	$\Lambda = \log \lambda$	entropy = log-luck
$\|E\| \cdot p$	$\|E\|/\lambda$	remaining microstates

They are the same thing in different notation.

5 Factor Map Commutation

When two factorizations ϕ_1, ϕ_2 both approximate the true behavior t^*:

```
       E
      / \
phi_1/   \phi_2
    /     \
   v       v
  Z_1 --> Z_2
    \     /
  d_1\   /d_2
      \ /
       v
       T
```

The triangle commutes: $d_2 \circ \tau_{1\to 2} = d_1$.
Agreement bound:

$$\hat{t}_1 =_{\varepsilon_1+\varepsilon_2} \hat{t}_2$$

6 Connection to Pattern Injection

SVD factorization of the bigram matrix P:

$$P^T = U \cdot S \cdot V^T$$

This is a factor map: $P^T \to$ (output directions) $\times$ (input directions).

The singular values S are the "luck" of each direction—how much variance (microstates) each component captures.

Truncating to rank-k is a quotient operation:

$$Q = \frac{\sum S_i^2}{\sum_{i \leq k} S_i^2}$$

Component 0 captures 97.5% of variance (low luck, high probability). Components 1–5 capture interpretable structure in the remaining 2.5%.

7 The Depth Limit

Maximum pattern depth in float32:

$$d_{\max} = \left\lfloor \frac{24}{-\log_2 p_{\text{avg}}} \right\rfloor$$

This is where all four views converge:

- **Bayes**: Can't represent probabilities below 2^{-24}
- **Thermo**: Can't distinguish more than 2^{24} microstates
- **Quotient**: Cumulative Q overflows after $d_{\max}$ layers
- **Factor maps**: Can't translate patterns deeper than precision allows

Q carries unbounded information. `float32` doesn't.