Bayes from Counting: Partial Quotients, GCD, and the Symmetric Learning Function on $E = I \times O$

Claude and MJC

February 12, 2026

Abstract

We show that Bayes' theorem falls out of the $E \to N \to Q$ chain when the joint event space $E = I \times O$ is factored via partial quotients. A *partial quotient* divides E by a single atomic event $i_0 \in I$ (not by the entire event space I), yielding a conditional distribution over O scaled by an equivalence class of size $c(i_0)$. The GCD of each row (or column) of the count table separates the *common evidence* (how often the conditioning event occurred) from the *differential evidence* (the shape of the conditional), and the GCD cancels from every conditional. Bayes' theorem is then the *consistency condition* between the I-side and O-side partial quotients: the two GCD-reduced decompositions of the same joint count must agree, and their agreement forces $P(o|i)/P(i|o) = P(o)/P(i)$. The quotient $Q = 1/P$ (luck) decomposes as $Q_{\text{joint}} = Q_{\text{marginal}} \cdot Q_{\text{conditional}}$, and the symmetry of the log contingency table—the output of the standard learning function—is what makes the causal and evidential directions both readable from the same matrix, justifying the CMP claim that correlation captures causation.

1 The Joint Event Space

Let I and O be finite sets (input and output event spaces) with $|I| = m$ and $|O| = n$. The joint event space is $E = I \times O$. A dataset D of N observations is a sequence of joint events $(i_t, o_t) \in E$ for $t = 0, \dots, N-1$.

Definition 1 (Joint count). *The* joint count function $c : E \to \mathbb{N}_{\geq 0}$ *is*

$$c(i,o) = |\{t : (i_t, o_t) = (i,o)\}|.$$

The marginals are $c(i) = \sum_o c(i,o)$ *and* $c(o) = \sum_i c(i,o)$*, with* $\sum_{i,o} c(i,o) = N$*.*

2 The Standard Learning Function: $E \to N$

The standard learning function ω_0 (CMP, §5) records every observed joint event. Its output is the *log contingency table*:

$$s(i,o) = \log_2 c(i,o)$$

for each (i,o) with $c(i,o) > 0$, and $s(i,o) = -\infty$ otherwise.

This is the first step of the $E \to N$ map: events are counted, and counts are the natural numbers assigned to events.

Remark 1 (The matrix is symmetric in its role)**.** *The log contingency table M with $M_{io} = s(i, o)$ can be read in two directions:*

- *$I \to O$: fix a row i, read off the log supports for each o.*
- *$O \to I$: fix a column o, read off the log supports for each i.*

CMP observes that "learning takes the directed causal arrow from rain to wet ground into a bidirectional one; we may infer that it has rained from the ground being wet just as well." This bidirectionality is a consequence of the matrix being a single object read two ways—the same counts *encode both directions.*

3 The Partial Quotient

Definition 2 (Full quotient)**.** *The* full quotient *E/I identifies all events that share the same output: $(i_1, o) \sim (i_2, o)$ for all i_1, i_2. The result is O, the output event space. Similarly $E/O \cong I$.*

Definition 3 (Partial quotient)**.** *The* partial quotient *of E by an atomic event $i_0 \in I$ is the pair:*

1. *The* ***conditional fiber***: *$O_{i_0} = \{(i_0, o) : o \in O\} \cong O$, with counts $c(i_0, o)$ for each o.*
2. *The* ***equivalence class***: *$[i_0] = \{t \in D : i_t = i_0\}$, with $|[i_0]| = c(i_0)$.*

The partial quotient *divides* by a single atom (fixing $I = i_0$) and *multiplies* by the equivalence class (recording how many positions share that input). The reconstruction identity is:

$$c(i_0, o) = P(o \mid i_0) \cdot c(i_0),$$

or in log support:

$$s(i_0, o) = \underbrace{\log_2 P(o \mid i_0)}_{\text{conditional (shape)}} + \underbrace{s_I(i_0)}_{\text{class size (scale)}}, \tag{1}$$

where $s_I(i_0) = \log_2 c(i_0)$.

By symmetry, the partial quotient of E by $o_0 \in O$ gives:

$$s(i, o_0) = \log_2 P(i \mid o_0) + s_O(o_0). \tag{2}$$

4 The GCD Decomposition

Definition 4 (Row and column GCD)**.** *For each $i \in I$, define the* row GCD*:*

$$g_I(i) = \gcd_{o : c(i,o) > 0} c(i, o).$$

For each $o \in O$, define the column GCD*:*

$$g_O(o) = \gcd_{i : c(i,o) > 0} c(i, o).$$

Since $g_I(i)$ divides every entry in row i, we can write:

$$c(i, o) = g_I(i) \cdot r_I(i, o), \qquad \gcd_o r_I(i, o) = 1,$$

where $r_I(i, o) \in \mathbb{N}$ are the *reduced counts* (coprime across the row).

Definition 5 (Reduced total)**.** $R_I(i) = \sum_o r_I(i,o) = c(i)/g_I(i)$.

Proposition 1 (GCD cancels from the conditional)**.**

$$P(o \mid i) = \frac{c(i,o)}{c(i)} = \frac{g_I(i) \cdot r_I(i,o)}{g_I(i) \cdot R_I(i)} = \frac{r_I(i,o)}{R_I(i)}.$$

The conditional probability depends only on the reduced counts r_I and reduced total R_I. The GCD $g_I(i)$ cancels completely.

In log support, the joint count decomposes as:

$$s(i,o) = \underbrace{\log_2 g_I(i)}_{\text{common evidence}} + \underbrace{\log_2 r_I(i,o)}_{\text{differential evidence}}. \tag{3}$$

The **common evidence** $\log_2 g_I(i)$ is the part of the log support that is shared by *all* outputs given input i. It tells us about the *prevalence* of input i without distinguishing which output accompanied it.

The **differential evidence** $\log_2 r_I(i,o)$ is the part that distinguishes output o from the other outputs. It is the *irreducible* evidence for this specific joint event.

By symmetry, the column-side decomposition gives:

$$s(i,o) = \log_2 g_O(o) + \log_2 r_O(i,o), \qquad \gcd_i r_O(i,o) = 1. \tag{4}$$

5 Bayes as Consistency of Partial Quotients

The same joint count $c(i,o)$ admits two GCD decompositions:

$$c(i,o) = g_I(i) \cdot r_I(i,o) \qquad \text{(row decomposition)} \tag{5}$$

$$c(i,o) = g_O(o) \cdot r_O(i,o) \qquad \text{(column decomposition)} \tag{6}$$

Theorem 1 (Bayes from GCD consistency)**.** *Equating* (**??**) *and* (**??**)*:*

$$g_I(i) \cdot r_I(i,o) = g_O(o) \cdot r_O(i,o),$$

*which gives the **GCD bridge**:*

$$\frac{r_I(i,o)}{r_O(i,o)} = \frac{g_O(o)}{g_I(i)}. \tag{7}$$

Now compute the ratio of conditionals:

$$\frac{P(o|i)}{P(i|o)} = \frac{r_I(i,o)/R_I(i)}{r_O(i,o)/S_O(o)} = \frac{r_I(i,o) \cdot S_O(o)}{r_O(i,o) \cdot R_I(i)}$$

$$= \frac{g_O(o)}{g_I(i)} \cdot \frac{S_O(o)}{R_I(i)} \qquad \textit{(by (??))}$$

$$= \frac{g_O(o) \cdot S_O(o)}{g_I(i) \cdot R_I(i)} = \frac{c(o)}{c(i)} = \frac{P(o)}{P(i)}.$$

*This is **Bayes' theorem**:*

$$\boxed{P(o \mid i) = \frac{P(i \mid o) \cdot P(o)}{P(i)}.}$$

Remark 2 (What the proof says)**.** *Bayes' theorem is the* consistency condition *between the two partial quotients of $E = I \times O$. Dividing by $I = i$ and dividing by $O = o$ must give compatible conditionals, because they decompose the* same *joint count. The GCD mediates: the ratio $g_O(o)/g_I(i)$ converts between the two reduced-count representations, and when multiplied by the reduced totals, recovers the marginal ratio $P(o)/P(i)$.*

Corollary 1 (Log-support form of Bayes)**.** *In the log domain:*

$$\log_2 r_I(i,o) - \log_2 r_O(i,o) = \log_2 g_O(o) - \log_2 g_I(i). \tag{8}$$

The left side is the difference between the I-reduced and O-reduced log supports for a specific joint event (i,o). The right side depends only on the marginal GCDs.

Equivalently:

$$s(o|i) - s(i|o) = s_O(o) - s_I(i).$$

The asymmetry of the conditionals equals the asymmetry of the priors.

6 The $E \to N \to Q$ Chain

We now connect the partial quotient to the quotient chain from the thermodynamic framework [**?**].

Definition 6 (The three-step map)**.** *For a joint event (i,o):*

1. *E: Identify the event $(i,o) \in I \times O$.*
2. *N: Count it. $c(i,o) \in \mathbb{N}$.*
3. *Q: Compute the quotient. $Q(i,o) = N/c(i,o) = \lambda(i,o)$ (the joint luck).*

The joint luck $\lambda(i,o) = 1/P(i,o)$ decomposes via the partial quotient:

Proposition 2 (Luck decomposition)**.**

$$\log_2 Q(i,o) = \underbrace{\log_2 Q_I(i)}_{\textit{marginal luck}} + \underbrace{\log_2 Q(o|i)}_{\textit{conditional luck}} \tag{9}$$

where $Q_I(i) = N/c(i) = 1/P(i)$ and $Q(o|i) = c(i)/c(i,o) = 1/P(o|i) = R_I(i)/r_I(i,o)$.

The conditional luck depends only on the reduced counts:

$$Q(o|i) = \frac{R_I(i)}{r_I(i,o)}.$$

The GCD $g_I(i)$ cancels from the conditional luck, just as it cancels from the conditional probability.

Theorem 2 (Bayes in quotient form)**.**

$$\frac{Q(o|i)}{Q(i|o)} = \frac{Q_O(o)}{Q_I(i)} = \frac{P(i)}{P(o)}.$$

The ratio of conditional lucks equals the inverse *ratio of the priors. An event that is "lucky" from one direction is "ordinary" from the other, balanced by the marginal frequencies.*

Proof. $Q(o|i)/Q(i|o) = P(i|o)/P(o|i) = P(i)/P(o)$ by Theorem **??**. □

Remark 3 (Luck is relative)**.** *The joint luck $Q(i,o) = N/c(i,o)$ is an absolute quantity. But it decomposes into marginal and conditional luck in* two *ways:*

$$\begin{aligned}\log_2 Q(i,o) &= \log_2 Q_I(i) + \log_2 Q(o|i) \qquad \textit{(I-first)}\\ &= \log_2 Q_O(o) + \log_2 Q(i|o) \qquad \textit{(O-first)}\end{aligned}$$

The two decompositions are consistent (they give the same joint luck) because Bayes holds. The "luck" of seeing output o after input i is not intrinsic to the pair—it depends on which direction you condition. The GCD structure (Eq. **??***) is what makes the two directions compatible.*

7 Why Correlation Captures Causation

CMP makes the claim that "we may infer that it has rained from the ground being wet just as well as that the ground must be wet from the fact that it is raining." The partial quotient framework makes this precise.

Proposition 3 (The symmetric matrix)**.** *The log contingency table $M_{io} = s(i,o) = \log_2 c(i,o)$ is a single matrix that encodes:*

- *The causal direction: $P(o|i) = 2^{s(i,o)} / \sum_{o'} 2^{s(i,o')}$.*
- *The evidential direction: $P(i|o) = 2^{s(i,o)} / \sum_{i'} 2^{s(i',o)}$.*
- *Both partial quotients and their GCD decompositions.*
- *All lucks, marginals, and conditionals.*

The matrix M is the **sufficient statistic** for the function $I \to O$ (CMP, §5.2). But it is equally the sufficient statistic for $O \to I$. *No additional data is needed to reverse the direction.*

Why? Because the GCD decomposition is stored implicitly. The row-GCDs $g_I(i)$ and column-GCDs $g_O(o)$ are computable from M, and they mediate Bayes via the bridge equation (**??**). The causal direction (the world's arrow from rain to wet ground) determined which counts landed where, but the counts *themselves* are direction-agnostic.

Remark 4 (Correlation and the GCD)**.** *Consider the pointwise mutual information:*

$$\mathrm{PMI}(i,o) = \log_2 \frac{c(i,o) \cdot N}{c(i) \cdot c(o)} = \log_2 \frac{r_I(i,o) \cdot N}{R_I(i) \cdot c(o)}.$$

The PMI measures correlation*: how much the joint count exceeds what independence would predict. Using $c(o) = g_O(o) \cdot S_O(o)$ and the GCD bridge:*

$$\begin{aligned}\mathrm{PMI}(i,o) &= \log_2 r_I(i,o) - \log_2 R_I(i) - \log_2 S_O(o) + \log_2 N - \log_2 g_O(o)\\ &= \log_2 r_O(i,o) - \log_2 R_I(i) - \log_2 S_O(o) + \log_2 N - \log_2 g_I(i).\end{aligned}$$

Both forms give the same PMI (as they must). The global GCD $g = \gcd_{i,o} c(i,o)$ cancels entirely from the PMI. The row and column GCDs contribute through the reduced totals R_I and S_O. The PMI depends on the shape *of the counts (the reduced parts), not their* scale *(the GCDs).*

Correlation (PMI) captures causation (the directional relationship $I \to O$) because the log contingency table is the sufficient statistic for both, and the GCD structure guarantees that the causal and evidential readings are consistent via Bayes.

8 The Quotient as the Thermodynamic Bridge

The $E \to N \to Q$ chain from the February 11 archive [?, ?] identifies:

Step	Object	Meaning
E	Joint event (i, o)	What happened
N	Count $c(i, o)$	How often it happened
Q	Quotient $N/c(i, o) = \lambda$	How surprising (luck)

The partial quotient adds the decomposition at the Q level:

Component	Formula	Meaning
$Q_{\text{joint}}(i, o)$	$N/c(i, o)$	Total luck of the joint event
$Q_I(i)$	$N/c(i)$	Luck of the input (marginal)
$Q(o\|i)$	$R_I(i)/r_I(i, o)$	Luck of the output given input
$g_I(i)$	$\gcd_o c(i, o)$	Common evidence (direction-free)
$r_I(i, o)$	$c(i, o)/g_I(i)$	Differential evidence (directional)

The GCD $g_I(i)$ is the "thermodynamic floor"—the minimum evidence that every output shares given input i. In the microstate–macrostate framework [?], it corresponds to the number of microstates that are assigned to the input macrostate i uniformly across all output classes.

Proposition 4 (GCD and the entropy budget)**.** *The row entropy decomposes as:*

$$H(O|I = i) = \log_2 R_I(i) - \sum_o \frac{r_I(i, o)}{R_I(i)} \log_2 r_I(i, o).$$

The GCD $g_I(i)$ does not appear. The conditional entropy—and therefore the conditional luck $\log_2 Q(o|i) = -\log_2 P(o|i)$—is determined entirely by the reduced counts.

The marginal entropy includes the GCD:

$$H(I) = \log_2 N - \sum_i \frac{c(i)}{N} \log_2 c(i) = \log_2 N - \sum_i \frac{c(i)}{N} [\log_2 g_I(i) + \log_2 R_I(i)].$$

The GCD contributes to the marginal entropy but not the conditional. Factoring out the GCD is "accounting for the prior."

9 The Tropical Approximation

The UM's standard update function operates in the $(\max, \min)$ tropical semiring:

$$(f_p(t))_j = \max_i \min(t_i, p_{ij}).$$

Definition 7 (Tropical GCD)**.** *In the tropical semiring $(\max, +)$, the GCD of a set of values $\{a_1, \dots, a_n\}$ is $\min(a_1, \dots, a_n)$.*

The tropical GCD of the log-support row $\{s(i, o)\}_{o \in O}$ is:

$$\min_o s(i, o) = \log_2 \min_o c(i, o).$$

The integer GCD gives $\log_2 \gcd_o c(i,o)$. Since gcd ≤ min always:

$$\log_2 \gcd_o c(i,o) \le \min_o \log_2 c(i,o),$$

with equality when $\min_o c(i,o)$ divides every $c(i,o)$ in the row.

Remark 5 (The UM approximates the GCD)**.** *The* min *operation in the standard update is the tropical version of the integer GCD. When the UM computes* $\min(t_i, p_{ij})$*, it is performing a tropical GCD that* upper bounds *the true integer GCD. The gap* $\log_2(\min / \gcd)$ *measures how far the tropical computation is from exact Bayesian inference.*

For integer counts where the smallest count divides all others (common when counts arise from a generative process with integer rates), the gap is zero and the UM's tropical computation is exact.

10 Worked Example

Let $I = \{a, b\}$, $O = \{x, y\}$, with count table:

	x	y	$c(i)$
a	6	2	8
b	3	9	12
$c(o)$	9	11	$N = 20$

Row GCDs and reduced counts. $g_I(a) = \gcd(6,2) = 2$, so $r_I(a,\cdot) = (3,1)$, $R_I(a) = 4$. Check: $c(a) = 2 \cdot 4 = 8$. ✓

$g_I(b) = \gcd(3,9) = 3$, so $r_I(b,\cdot) = (1,3)$, $R_I(b) = 4$. Check: $c(b) = 3 \cdot 4 = 12$. ✓

Column GCDs and reduced counts. $g_O(x) = \gcd(6,3) = 3$, so $r_O(\cdot,x) = (2,1)$, $S_O(x) = 3$. Check: $c(x) = 3 \cdot 3 = 9$. ✓

$g_O(y) = \gcd(2,9) = 1$, so $r_O(\cdot,y) = (2,9)$, $S_O(y) = 11$. Check: $c(y) = 1 \cdot 11 = 11$. ✓

Conditionals (GCD-free). $P(x|a) = r_I(a,x)/R_I(a) = 3/4$.

$P(x|b) = r_I(b,x)/R_I(b) = 1/4$.

$P(a|x) = r_O(a,x)/S_O(x) = 2/3$.

GCD bridge (Eq. ??). $r_I(a,x)/r_O(a,x) = 3/2 = g_O(x)/g_I(a) = 3/2$. ✓

$r_I(b,y)/r_O(b,y) = 3/9 = 1/3 = g_O(y)/g_I(b) = 1/3$. ✓

Bayes. $P(x|a) = P(a|x) \cdot P(x)/P(a) = (2/3)(9/20)/(8/20) = (2/3)(9/8) = 3/4$. ✓

Luck. $Q(x|a) = R_I(a)/r_I(a,x) = 4/3 \approx 1.33$.

$Q(a|x) = S_O(x)/r_O(a,x) = 3/2 = 1.5$.

Ratio: $Q(x|a)/Q(a|x) = (4/3)/(3/2) = 8/9 = c(a)/c(x) = 8/9$. ✓

The output x is "less lucky" given input a (luck 1.33) than input a is given output x (luck 1.5), because a is rarer than x.

Log support decomposition.

(i,o)	$s(i,o)$	$\log_2 g_I$	$\log_2 r_I$	$\log_2 g_O$	$\log_2 r_O$
(a,x)	2.585	1.000	1.585	1.585	1.000
(a,y)	1.000	1.000	0.000	0.000	1.000
(b,x)	1.585	1.585	0.000	1.585	0.000
(b,y)	3.170	1.585	1.585	0.000	3.170

Each row confirms $s = \log_2 g_I + \log_2 r_I = \log_2 g_O + \log_2 r_O$.

The GCD bridge in log: $\log_2 r_I - \log_2 r_O = \log_2 g_O - \log_2 g_I$ is verified for every cell (e.g., for (a,x): $1.585 - 1.000 = 1.585 - 1.000$).

11 Discussion

11.1 What the GCD "accounts for"

The GCD of a row $g_I(i) = \gcd_o c(i,o)$ is the largest integer that divides every count in the row. It represents the "granularity" of the evidence: the data about input i comes in packets of size $g_I(i)$. Dividing by the GCD gives the *irreducible* evidence for each output.

In the UM framework, the GCD is "accounted for" by separating it out of the conditional. The SN strength $s(i,o)$ encodes the full joint evidence. The partial quotient decomposes it into common evidence (the GCD, which contributes to the prior) and differential evidence (the reduced counts, which determine the conditional).

Bayes falls out because the joint count has a *unique* value $c(i,o)$ that admits two GCD decompositions (row-side and column-side). Their consistency, mediated by the GCD bridge, IS Bayes' theorem.

11.2 Justifying the "wilder" claims of CMP

Correlation implies causation (in the UM). CMP claims that the log contingency table—pure correlation—captures causal structure. The partial quotient framework explains why: the causal direction ($I \to O$) and the evidential direction ($O \to I$) are *both* readable from the same matrix because Bayes guarantees their consistency. No additional causal knowledge is needed beyond the counts. The *learning function does not know which direction is causal*—it records joint events symmetrically—but the *partial quotient* extracts the directed conditional in either direction.

$Q = \lambda$ (luck). The quotient $Q(o|i) = 1/P(o|i)$ is the conditional luck: how "surprising" the output is given the input. The GCD decomposition shows that luck depends only on the reduced counts. The common evidence (GCD) is "already priced in"—it is the part of the count that carries no surprise. Luck measures only the differential evidence.

The Bayesian identity in luck form (Theorem **??**) says that luck is *relative*: the same event has different luck from different conditioning directions, balanced by the priors. This is the precise content of CMP's "$Q = \lambda$": the quotient is the luck, and the luck is direction-dependent.

The standard learning function is complete. The log contingency table, the output of ω_0, is the sufficient statistic for both $I \to O$ and $O \to I$. No other learning is needed. The GCD decomposition and Bayes are not additional algorithms applied to the data—they are *structural properties* of the integer count table that hold automatically. "To learn everything that can be statistically learned about a function from I to O, we can simply record every observed joint event" (CMP, §5.2). The partial quotient shows that "everything" includes both causal directions, all conditionals, all lucks, and Bayes' theorem itself.

References

[1] Michaeljohn Clement. *CMP*. `https://cmpr.ai/cmp.pdf`, 2026.

[2] Claude and MJC. *The Quotient Chain: $E \to N \to Q$ at Every Operation*. Hutter archive, 11 Feb 2026.

[3] Claude and MJC. *E onto N: The Bi-Embedding of Events and Numbers*. Hutter archive, 11 Feb 2026.

[4] Claude and MJC. *Microstates, Macrostates, and the Partition Function*. Hutter archive, 11 Feb 2026.