Understanding the Quotient

An Explainer for MCP Graf 13

Claude and MJC

March 12, 2026

Abstract

Graf 13 of the MCP records five equations from the early paper *Quotient Renormalization and Maximum Entropy*. They are compact but they carry a complete picture: a fine-grained distribution can be factored into a quotient distribution over equivalence classes and a conditional distribution inside each class. In log-space this becomes an additive decomposition of code lengths. If the within-class detail is unknown, the principled default reconstruction is the maximum-entropy lift, which spreads each class mass uniformly across its members. This note explains those five equations with plain definitions and a worked example.

1 Graf 13 Verbatim

The five equations of MCP graf 13 are reproduced here verbatim:

$$t_{[e]} = \log_2 \left(\sum_{e' \sim e} 2^{t_{e'}} \right) \tag{1}$$

$$t_{e|es} = t_e - t_{[es]} = \log_2 P(e|es) \tag{2}$$

$$t_e = t_{[es(e)]} + t_{e|es} \tag{3}$$

$$P(e) = P(es) \times P(e|es) \tag{4}$$

$$\tilde{P}(e) = P(es) \times \frac{1}{|es|} \tag{5}$$

2 The Setup

Let E be a finite set of fine-grained events. We choose an equivalence relation $\sim$ on E and let $[e]$ denote the equivalence class containing e. Following the source paper, we also write $es(e)$ for the class of e when we want to stress the event space interpretation.

Definition 1 (Quotient and fiber)**.** *The* quotient *is the coarse space $E/\sim$ of equivalence classes. The* fiber *over a class $c \in E/\sim$ is the set of fine events whose class is c:*

$$\pi^{-1}(c) = \{e \in E : [e] = c\}.$$

Assume a probability distribution P on E. MCP writes

$$t_e = \log_2 P(e).$$

These are negative code lengths up to sign convention: multiplying probabilities becomes addition in t-space.

3 The Quotient Pushforward

The first equation of graf 13 is:

$$t_{[e]} = \log_2 \left(\sum_{e' \sim e} 2^{t_{e'}} \right).$$

Since $2^{t_{e'}} = P(e')$, this is just

$$P([e]) = \sum_{e' \sim e} P(e').$$

Proposition 1. *The quotient pushforward replaces each fine event by its equivalence class and sums the probability mass within that class.*

This is the coarse distribution. It forgets which member of a class occurred and keeps only which class occurred.

4 The Fiber Coordinate

The second equation is:

$$t_{e|es} = t_e - t_{[es]} = \log_2 P(e|es).$$

This is the residual information needed to identify e once its class es is known. In ordinary probability notation,

$$P(e|es) = \frac{P(e)}{P(es)}.$$

The fiber coordinate is therefore the within-class distribution.

The third and fourth equations simply restate the same factorization in log-space and probability-space:

$$t_e = t_{[es(e)]} + t_{e|es},$$

$$P(e) = P(es) \times P(e|es).$$

These are the key structural statements of the graf:

1. first choose the coarse class;
2. then choose the fine event inside that class.

In compression language: total code length splits into a class-description term and a within-class residual term.

5 Worked Example

Example 1 (Two classes)**.** *Let*

$$E = \{a_1, a_2, b_1, b_2, b_3\},$$

with classes

$$A = \{a_1, a_2\}, \qquad B = \{b_1, b_2, b_3\}.$$

Suppose

$$P(a_1) = \tfrac{1}{8}, \quad P(a_2) = \tfrac{3}{8}, \quad P(b_1) = P(b_2) = P(b_3) = \tfrac{1}{6}.$$

Then

$$P(A) = \tfrac{1}{2}, \qquad P(B) = \tfrac{1}{2}.$$

Inside the fibers we have

$$P(a_1|A) = \tfrac{1}{4}, \quad P(a_2|A) = \tfrac{3}{4},$$

and

$$P(b_i|B) = \tfrac{1}{3}.$$

This example shows the two-stage factorization explicitly:

$$P(a_2) = P(A)\,P(a_2|A) = \tfrac{1}{2} \cdot \tfrac{3}{4} = \tfrac{3}{8}.$$

The same logic applies to every event in E.

6 The Maximum-Entropy Lift

The last equation of graf 13 is:

$$\tilde{P}(e) = P(es) \times \frac{1}{|es|}.$$

This answers a different question. Suppose we know the quotient distribution $P(es)$ but not the within-class details. Which fine distribution should we reconstruct?

The source paper chooses the maximum-entropy answer: make the distribution uniform inside each class. Then each member of a class gets an equal share of that class's mass.

In the worked example above, if we forget the internal splits in A and B but keep $P(A) = P(B) = 1/2$, the lift becomes

$$\tilde{P}(a_1) = \tilde{P}(a_2) = \tfrac{1}{4},$$
$$\tilde{P}(b_1) = \tilde{P}(b_2) = \tilde{P}(b_3) = \tfrac{1}{6}.$$

So the max-entropy lift preserves the coarse structure while refusing to invent unsupported within-fiber asymmetry.

7 Why Graf 13 Matters

Graf 13 matters because it isolates three distinct objects:

1. the quotient distribution over coarse classes;
2. the fiber distributions within each class;
3. the max-entropy reconstruction when the fiber distributions are unknown.

That gives a clean hierarchy:

1. coarse structure can be modeled separately from fine detail;
2. the fine detail is exactly the conditional distribution on the fibers;
3. if the fine detail is discarded, the least-committal lift is uniform within each fiber.

This is why the graf belongs in the MCP. It is not tied to the original ES experiment. It is a general statement about quotienting a distribution, decomposing description length, and reconstructing a fine distribution from a coarse one.

8 Relation to the Source Paper

The source paper *Quotient Renormalization and Maximum Entropy* contains these equations directly: the quotient pushforward, the within-ES fiber coordinate, the multiplicative factorization $P(e) = P(es)P(e|es)$, and the max-entropy lift. What MCP graf 13 does is strip away the experiment-specific table and keep the reusable mathematical core.

References

[1] Claude and MJC. *Quotient Renormalization and Maximum Entropy*. Hutter archive, 31 January 2026. `docs/archive/20260131_2/quotient-maxent.tex`

[2] Claude and MJC. *MCP*. Working manuscript, March 2026.