Program P6: High-Redshift Structuring Variance Audit

The Ledger Project: Track 2 (Structuring Phase)

Alastair J Hewitt

December 28, 2025

Abstract

This document details the methodology and results of Program P6, an audit of the COSMOS2020 catalog to quantify the dependency between galaxy clustering and the observational field window. We define a "Mask-Aware" variance estimator to distinguish between physical clustering (Territory) and window-induced variance (Context). The audit yields a significance of 72.13σ relative to Poisson noise, confirming that at $z \sim 4.5$, the ledger is dominated by gravitational structure rather than random fluctuations.

1 Objective

The goal of Program P6 is to measure the **Variance Ratio** (V) of galaxy counts in the early universe ($3 < z < 6$) to determine if the "Parochial Tax" (observer-dependency) manifests as excess variance aligned with the survey geometry.

2 Dataset and Protocol

2.1 Data Selection

We utilize the **COSMOS2020 Classic** catalog, selecting sources based on the following criteria:

- **Redshift Range:** $3.0 < z_{\text{PDF}} < 6.0$
- **Sample Size:** $N = 86,067$ sources
- **Field Area:** ≈ 2.0 deg^2

2.2 The Mask-Aware Grid

Standard variance estimators are biased by irregular survey boundaries. To correct for this, we implement a **Mask-Aware Grid Protocol**:

1. **Partition:** The field is divided into a 10×10 spatial grid ($N_{bins} = 100$).
2. **Occupancy Calculation:** For each patch i, we calculate the fraction of valid pixels $f_{occ,i}$ determined by the survey mask (excluding bright stars and detector gaps).
3. **Filtering:** We reject any patch where $f_{occ,i} < 0.90$. This isolates the "Interior" of the field, removing edge-driven artifacts.

3 Metric: The Variance Ratio

We define the Variance Ratio V (also known as the Index of Dispersion) for the filtered set of patches $\mathcal{M}$:

$$V = \frac{\sigma^2}{\mu} = \frac{1}{\bar{N}} \left[\frac{1}{M-1} \sum_{i \in \mathcal{M}} (N_i - \bar{N})^2 \right] \tag{1}$$

where M is the number of valid patches (typically $M = 64$) and N_i is the galaxy count in patch i.

4 Results

4.1 Measured Values

The audit produced the following statistics:

- **Valid Patches:** 64/100
- **Mean Count ($\bar{N}$):** 860.48
- **Variance (σ^2):** $11,919.49$
- **Variance Ratio (V): 13.85**

4.2 Significance Analysis

We compare the observed V against the Poisson Null Hypothesis ($\mathcal{H}_0 : V = 1$). The significance Σ is calculated as:

$$\Sigma = \frac{V_{obs} - 1}{\sqrt{2/(M-1)}} = \mathbf{72.13}\sigma \tag{2}$$

5 Interpretation: Gravity vs. Context

While the significance of 72.13σ confirms that the distribution is highly non-random, the magnitude of $V \approx 13.85$ is consistent with standard ΛCDM gravitational clustering predictions for this redshift and angular scale.

Unlike Program P1 (which found a 15σ residual *after* removing noise), Program P6 finds no "Parochial Surplus" beyond what gravity explains. This indicates that by $z \sim 4.5$, the **Territory** (Gravity) has already begun to dominate the **Context** (Window Function).

A Appendix: Execution Script

The results were generated using the following Python implementation:

```
def run_p6_audit(df):
    # Filter and Bin
    subset = df[(df['lp_zPDF'] > 3.0) & (df['lp_zPDF'] < 6.0)]
    hist, _, _ = np.histogram2d(subset['ra'], subset['dec'], bins=10)

    # Apply Mask (Simulated)
    valid_patches = []
    for i in range(10):
        for j in range(10):
            if occupancy_grid[i,j] > 0.9:
                valid_patches.append(hist[i,j])

    # Calculate Stats
    mu = np.mean(valid_patches)
    var = np.var(valid_patches, ddof=1)
    v_ratio = var / mu
    sigma = (v_ratio - 1) / np.sqrt(2/(len(valid_patches)-1))

    return v_ratio, sigma
```