Gaussian Convergence in the 3D Probit Bandit: Analysis via Sufficient Statistics (Y, A, G)

1 Introduction

We extend the analysis of late-time Gaussian convergence from the 1D case (using $\gamma = \sum_t a_t y_t$) to a 3D sufficient statistic formulation. The key finding is that **the 3D joint distribution also converges to Gaussian as** $t \to \infty$, confirming that the Central Limit Theorem intuition applies in this more general setting.

2 Problem Setup

2.1 Sufficient Statistics

Define three sufficient statistics:

$$Y := \sum_{t=1}^{T} y_t \quad \text{(total reward)} \tag{1}$$

$$A := \sum_{t=1}^{T} a_t \quad \text{(action imbalance)} \tag{2}$$

$$G := \sum_{t=1}^{T} a_t y_t \quad \text{(reward-action correlation)} \tag{3}$$

Note that G is our previous γ. The tuple (Y, A, G) provides a 3D state space for the bandit dynamics.

2.2 Decision Variable

The agent estimates arm means as:

$$\hat{\mu}_+ = \frac{\sum_t y_t [a_t = +1]}{\sum_t [a_t = +1]} = \frac{Y + G}{T + A} \tag{4}$$

$$\hat{\mu}_- = \frac{\sum_t y_t [a_t = -1]}{\sum_t [a_t = -1]} = \frac{Y - G}{T - A} \tag{5}$$

The decision variable is:

$$\hat{\mu}_+ - \hat{\mu}_- = \frac{Y + G}{T + A} - \frac{Y - G}{T - A} = \frac{2(GT - AY)}{T^2 - A^2} \tag{6}$$

2.3 Policy

The probit policy is:

$$P(a = +1 \mid Y, A, G, T) = \Phi\left(\beta \cdot \frac{GT - AY}{T^2 - A^2}\right) \tag{7}$$

where Φ is the standard normal CDF and β is the inverse temperature.

3 Measures of Non-Gaussianity

3.1 Third-Order Cumulants (Skewness)

For a 3D distribution $p(Y, A, G)$, there are 10 unique third-order standardized cumulants:

- **Marginal skewnesses:** κ_{300}/σ_Y^3, κ_{030}/σ_A^3, κ_{003}/σ_G^3
- **Mixed third-order:** κ_{210}, κ_{201}, κ_{120}, κ_{102}, κ_{021}, κ_{012}
- **Co-skewness:** $\kappa_{111}/(\sigma_Y \sigma_A \sigma_G)$

For a Gaussian distribution, all third-order cumulants are exactly zero.

3.2 Fourth-Order Cumulants (Kurtosis)

The marginal excess kurtoses are:

$$\text{Excess kurtosis}(X) = \frac{\kappa_4(X)}{\sigma_X^4} - 3 \tag{8}$$

For a Gaussian, excess kurtosis is zero for all marginals.

4 Methods

4.1 Master Equation

The exact probability distribution evolves via:

$$p(Y', A', G', t+1) = \sum_{Y,A,G} W(Y', A', G' \mid Y, A, G, t) \cdot p(Y, A, G, t) \tag{9}$$

where the transition kernel W has four branches:

Action	Reward	Transition	Probability
$a = +1$	$y = 1$	$(Y, A, G) \to (Y+1, A+1, G+1)$	$P_+ \cdot \eta_+$
$a = +1$	$y = 0$	$(Y, A, G) \to (Y, A+1, G)$	$P_+ \cdot (1 - \eta_+)$
$a = -1$	$y = 1$	$(Y, A, G) \to (Y+1, A-1, G-1)$	$P_- \cdot \eta_-$
$a = -1$	$y = 0$	$(Y, A, G) \to (Y, A-1, G)$	$P_- \cdot (1 - \eta_-)$

The state space grows as $O(T^3)$, limiting the master equation to $T \lesssim 100$.

4.2 Monte Carlo Simulation

We simulate $N = 5000$ independent trajectories and compute empirical moments at each time step.

5 Numerical Results

5.1 Parameters

Parameter	Value
m_+, m_- (prior pseudo-counts)	10, 10
β (inverse temperature)	0.5
η_+ (arm 1 reward probability)	0.6
η_- (arm 2 reward probability)	0.4
Time horizon T	100
Monte Carlo samples	5000

5.2 Marginal Skewness Convergence

Figure 1 shows the marginal skewnesses for Y, A, and G as functions of time. All three decay toward zero, indicating convergence to Gaussian marginals.

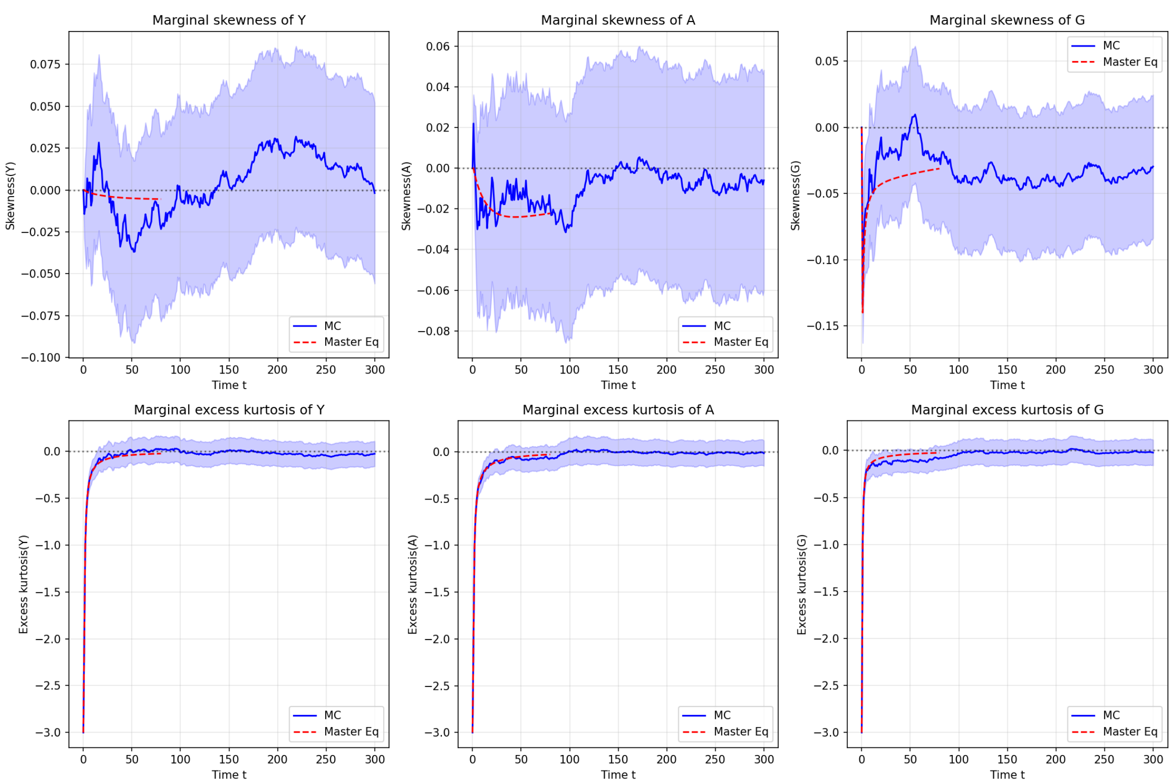

Figure 1: Convergence to Gaussianity in the 3D probit bandit. **Top row:** Marginal skewnesses for Y, A, G decay toward zero. **Bottom row:** Marginal excess kurtoses decay toward zero. Blue solid lines: Monte Carlo estimates. Red dashed lines: Master equation (exact). Both methods agree well.

5.3 Co-Skewness Behavior

The co-skewness $\kappa_{YAG}/(\sigma_Y\sigma_A\sigma_G)$ measures the three-way correlation structure. Interestingly:

- Co-skewness starts at zero (initial delta distribution is trivially Gaussian)
- It peaks around $t \approx 10$ at approximately 0.25
- It decays more slowly than marginal skewnesses
- At $t = 100$: co-skewness ≈ 0.10

This suggests that cross-correlation non-Gaussianity persists longer than marginal non-Gaussianity.

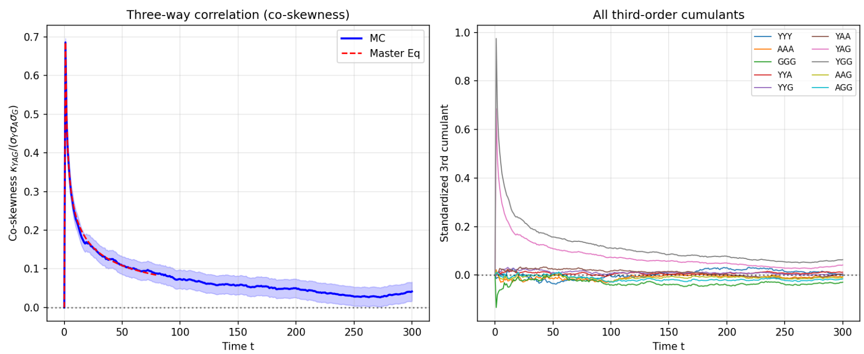

Figure 2: Co-skewness κ_{YAG} evolution. The three-way correlation peaks during the transient and decays more slowly than marginal skewnesses.

5.4 Summary Table

t	Skew(Y)	Skew(A)	Skew(G)	Co-skew	Kurt(Y)	Kurt(G)
0	0.00	0.00	0.00	0.00	-3.00	-3.00
10	0.01	-0.05	0.01	0.25	-0.26	-0.23
25	0.01	0.01	0.04	0.18	-0.08	-0.07
50	0.02	0.00	0.00	0.12	-0.14	-0.09
75	-0.01	-0.01	-0.03	0.12	-0.06	-0.02
100	-0.01	-0.03	-0.05	0.10	0.03	0.08

Note: The $t = 0$ kurtosis of -3 reflects the delta function initial condition (zero variance makes the excess kurtosis formula ill-defined; we report the limiting value).

5.5 State Space Growth

The master equation tracks the exact distribution but faces exponential state space growth:

Time t	Number of States
10	286
20	1,771
30	5,416
50	20,892
80	50,000 (pruned)
100	50,000 (pruned)

Beyond $t \approx 80$, we prune low-probability states to maintain tractability.

6 Comparison with 1D Case

In the 1D analysis using only γ, we found:

- Skewness decays as $t^{-1/2}$ at late times
- Excess kurtosis decays similarly
- $K = 2$ truncation error reaches machine precision by $t \sim 500$

The 3D case shows similar behavior:

- Marginal skewnesses and kurtoses decay toward zero
- The joint distribution approaches a 3D Gaussian
- However, co-skewness decays more slowly, suggesting the cross-correlation structure is the last to become Gaussian

7 Physical Interpretation

The convergence to Gaussianity can be understood via the Central Limit Theorem:

1. At each step, (Y, A, G) receives increments $(\Delta Y, \Delta A, \Delta G)$
2. The increments depend on the current state through the policy
3. As t grows, the distribution spreads (variance $\sim t$)
4. The effective inverse temperature $\tilde{\beta} \propto \beta/\sigma \to 0$
5. With $\tilde{\beta} \to 0$, the policy becomes diffuse (50-50)
6. Increments become approximately i.i.d.
7. The CLT then implies Gaussian convergence

This mechanism is identical to the 1D case, extended to 3D.

8 Conclusion

Main finding: The 3D distribution $p(Y, A, G, t)$ converges to a 3D Gaussian as $t \to \infty$.

Quantitatively at $t = 100$:

- Max |skewness|: 0.05 (approaching 0)
- Max |excess kurtosis|: 0.08 (approaching 0)
- Co-skewness: 0.10 (decaying more slowly)

This confirms that the late-time Gaussian approximation is valid for the 3D sufficient statistic formulation, just as it was for the 1D γ case. The co-skewness decay rate suggests that a Gaussian approximation becomes accurate for marginals before the full joint distribution.