PRISM: Probabilistic Regime-Integrated Signal Model
A Bayesian Workflow for Event-Probability Inference from L3 Order Flow

Sean Slattery

February 2026

Abstract

PRISM (Probabilistic Regime-Integrated Signal Model) is a Bayesian workflow for posterior inference on an event of interest using granular L3 order-flow information and calibrated machine-learning signals. The canonical pipeline is: L3 order flow $\rightarrow$ anchor beliefs $\rightarrow$ ML signal $\rightarrow$ calibration $\rightarrow$ fusion $\rightarrow$ corrections $\rightarrow$ posterior. The structural anchor prior is fixed at $\text{Beta}(1, 1)$, and machine-learning output is treated as a signal rather than a prior. Calibrated signal uncertainty is mapped conservatively into Beta parameters through moment matching with feasibility constraints and concentration caps. The workflow is designed for interpretable posterior summaries, stress-test diagnostics, and explicit limitation reporting under low-support or regime-instability conditions. The main paper provides the workflow, assumptions, and conservative theory claims; the supplement provides full synthetic design details, stress tests, diagnostics, and extended proofs.

1 Introduction

Purpose. This section motivates PRISM as a practical Bayesian workflow for extracting event-level beliefs from market microstructure data. The emphasis is statistical coherence and uncertainty reporting, not direct pricing output.

What you will learn.

- Why L3 order flow is informative for posterior belief extraction in parimutuel-style settings.
- Why machine-learning output is treated as a calibratable signal instead of a prior.
- How PRISM separates workflow definition from empirical validation claims.

Financial markets and prediction venues generate high-frequency traces of participant behavior. Order arrivals, cancellations, queue placement, and trade aggressiveness contain information about directional beliefs, but the information is noisy and behaviorally distorted. Prior work on information aggregation in parimutuel environments shows that trading flow can encode private information in structured ways [Koessler et al., 2012, Axelrod et al., 2009].

PRISM formalizes that intuition as a Bayesian workflow: it maps raw microstructure to anchor beliefs, calibrates a machine-learning signal, fuses the two information sources, applies explicit correction maps, and outputs a posterior distribution over the event probability. The workflow uses transparent objects and conservative assumptions so that interpretation remains stable under stress scenarios.

The framework is intentionally modular. Each stage can be inspected, replaced, or stress tested without changing the meaning of the final posterior object. This design supports auditing and avoids implicit dependence on hidden black-box assumptions.

What this section established.

- PRISM is motivated by belief aggregation and uncertainty quantification from market microstructure.
- The central output is a posterior distribution for an event probability.
- The workflow is modular, auditable, and designed for conservative interpretation.

2 Scope and Positioning

Purpose. This section defines the paper boundary so claims remain aligned with what is actually modeled and audited. It also states what PRISM is not.

What you will learn.

- The canonical problem solved by PRISM.
- Explicit exclusions that prevent over-interpretation.
- How main-paper claims are separated from supplement-level evidence.

PRISM is a Bayesian workflow for posterior inference on a binary event of interest. It consumes granular order-flow evidence and a calibrated machine-learning signal, then returns posterior summaries such as mean, credible intervals, and calibration diagnostics.

PRISM is not a derivative-pricing engine, not a risk-neutral valuation model, and not a claim of trading-profit superiority. Downstream decision modules may consume the posterior, but those modules are outside this paper's canonical scope.

Claims in the main paper are limited to:

- workflow definition,
- internal mathematical consistency,
- conservative theoretical properties under explicit assumptions,
- high-level synthetic validation summary.

Detailed stress-study outputs, finite-sample diagnostics, and scenario tables are placed in the supplement.

What this section established.

- PRISM is positioned as posterior inference workflow, not pricing model.
- The manuscript scope is explicitly bounded.
- Main and supplement responsibilities are separated.

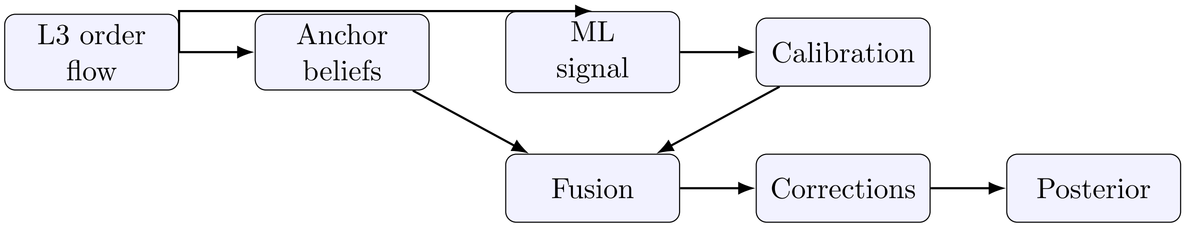

Figure 1: PRISM workflow DAG used throughout the manuscript.

3 Notation and Problem Setup

Purpose. This section defines the event, data objects, and the stage-by-stage variables used across the paper. The same symbols are reused in theory, diagnostics, and supplement tables.

What you will learn.

- Canonical notation for anchor, ML, fused, corrected, and posterior quantities.
- The data objects provided at each workflow stage.
- The exact stage ordering used throughout the manuscript.

Let $Y \in \{0,1\}$ denote the event of interest and let $p = \mathbb{P}(Y = 1)$ denote the unknown event probability. For each decision window t, let $\mathcal{D}_t^{\mathrm{L3}}$ denote raw L3 order-flow records and let X_t denote non-order-flow covariates available to an ML model.

We use the following stage-specific probability objects:

$p_{\mathrm{anchor},t}$	anchor belief from raw order flow,
$p_{\mathrm{ml,raw},t}$	raw ML score mapped to $(0,1)$,
$p_{\mathrm{ml,cal},t}$	calibrated ML probability,
$p_{\mathrm{fused},t}$	fused probability before corrections,
$p_{\mathrm{corr},t}$	probability after correction maps,
$p_{\mathrm{post},t}$	posterior mean after updated evidence.

The canonical flow used in both text and figures is:

$$\mathcal{D}_t^{\mathrm{L3}} \to \text{anchor beliefs} \to \text{ML signal} \to \text{calibration} \to \text{fusion} \to \text{corrections} \to \text{posterior}.$$

What this section established.

- All core variables are defined with stable notation.
- Stage ordering is explicit and fixed.
- Workflow figure is aligned with the mathematical objects used later.

4 Data and Anchor Beliefs

Purpose. This section defines how raw L3 order flow is transformed into anchor beliefs under a flat structural prior. The anchor object is the first probabilistic input to fusion.

What you will learn.

- The exact L3 fields used for anchor extraction.
- How effective yes/no evidence is computed from raw flow.
- Why the structural prior is fixed at $\mathrm{Beta}(1, 1)$.

For each observed message j in window t, let:

- $s_j \in \{+1, -1\}$ be side indicator (yes or no pressure),
- $q_j > 0$ be size,
- $d_j \geq 0$ be depth-relative distance,
- $\ell_j \geq 0$ be order lifetime proxy.

Define a bounded reliability weight

$$\omega_j = \min\{\omega_{\max}, \max\{\omega_{\min}, \ (1 + d_j)^{-1} \min(1, \ell_j/\ell_0)\}\},$$

with fixed constants $0 < \omega_{\min} \leq \omega_{\max}$ and scale $\ell_0 > 0$.

Effective evidence totals are

$$N_t^+ = \sum_{j \in \mathcal{D}_t^{\mathrm{L3}}} \omega_j q_j \, \mathbf{1}\{s_j = +1\}, \quad N_t^- = \sum_{j \in \mathcal{D}_t^{\mathrm{L3}}} \omega_j q_j \, \mathbf{1}\{s_j = -1\}.$$

The canonical structural prior is fixed at $\mathrm{Beta}(1, 1)$. The anchor belief is then

$$\pi_{\mathrm{anchor},t}(p) = \mathrm{Beta}(\alpha_{\mathrm{anchor},t}, \beta_{\mathrm{anchor},t}), \quad \alpha_{\mathrm{anchor},t} = 1 + N_t^+, \quad \beta_{\mathrm{anchor},t} = 1 + N_t^-.$$

This construction makes the prior-to-anchor transition explicit: PRISM starts neutral and lets order-flow evidence determine concentration.

What this section established.

- Anchor beliefs are computed directly from raw L3 signals.
- The baseline prior is exactly $\mathrm{Beta}(1, 1)$.
- Anchor concentration scales with reliability-weighted evidence, not heuristic phase labels.

5 ML Signal and Calibration

Purpose. This section formalizes the ML component as a signal generator, then calibrates that signal before any Bayesian fusion step. Uncertainty is propagated through conservative Beta moment matching.

What you will learn.

- Why ML output is not treated as a prior object.
- A transparent calibration map from raw score to calibrated probability.
- A defensible conversion from calibrated mean/variance to Beta parameters.

Let an ML model produce raw score $r_t \in \mathbb{R}$ or raw probability $\tilde{p}_t \in (0,1)$. If a score is emitted, map to probability by

$$p_{\mathrm{ml,raw},t} = \mathrm{sigmoid}(r_t).$$

If probability is emitted directly, set $p_{\mathrm{ml,raw},t} = \tilde{p}_t$.

Calibration is performed on a held-out calibration set using logistic calibration:

$$p_{\mathrm{ml,cal},t} = \mathrm{sigmoid}(a + b \ \mathrm{logit}(\mathrm{clip}(p_{\mathrm{ml,raw},t}, \varepsilon, 1-\varepsilon))),$$

with fitted (a, b) and small $\varepsilon > 0$.

Let $m_t = p_{\mathrm{ml,cal},t}$ and let v_t denote an uncertainty estimate from calibration-resampling (bootstrap on calibration folds [Efron and Tibshirani, 1994]) or a conservative floor when resampling support is weak. Beta moment matching is feasible when $0 < v_t < m_t(1-m_t)$. In that regime,

$$\kappa_t = \frac{m_t(1-m_t)}{v_t} - 1, \quad \alpha_{\mathrm{ml},t} = m_t\kappa_t, \quad \beta_{\mathrm{ml},t} = (1-m_t)\kappa_t.$$

To avoid overconfidence in low-support calibration regimes, PRISM applies a conservative cap and floor:

$$\kappa_t \leftarrow \min\{\kappa_{\max}, \max\{\kappa_{\min}, \kappa_t\}\}.$$

When feasibility fails, PRISM uses fallback variance $v_t \leftarrow m_t(1-m_t)/(1+\kappa_{\min})$ and reports the fallback in diagnostics.

Remark 1. This mapping is a modeling choice for uncertainty transport, not a claim that ML output is itself a Bayesian prior. The calibration-quality dependence must be reported alongside posterior summaries.

What this section established.

- ML output enters PRISM as a calibratable signal.
- Calibration is explicit and auditable.
- ML uncertainty enters fusion through conservative Beta moment matching with fallback rules.

6 Fusion and Corrections

Purpose. This section combines anchor and calibrated ML evidence into a single conjugate representation, then applies behavioral and structural correction maps before posterior updating.

What you will learn.

- The additive-evidence fusion rule used by PRISM.
- How correction maps are parameterized and constrained.
- Which quantities are transformed and which remain invariant.

Let

$$\pi_{\text{anchor},t} = \text{Beta}(\alpha_{\text{anchor},t}, \beta_{\text{anchor},t}), \quad \pi_{\text{ml},t} = \text{Beta}(\alpha_{\text{ml},t}, \beta_{\text{ml},t}).$$

Define reliability weights $w_{a,t}, w_{m,t} \in [0, 1]$ with $w_{a,t} + w_{m,t} = 1$. The fused evidence parameters are

$$\alpha_{\text{fused},t} = 1 + w_{a,t}(\alpha_{\text{anchor},t} - 1) + w_{m,t}(\alpha_{\text{ml},t} - 1),$$
$$\beta_{\text{fused},t} = 1 + w_{a,t}(\beta_{\text{anchor},t} - 1) + w_{m,t}(\beta_{\text{ml},t} - 1).$$

Behavioral correction is a bounded log-odds transform:

$$\phi_{\text{beh},t}(p) = \text{sigmoid}\big(\,\text{logit}(p) + \delta_0 + \delta_1 z_{\text{herd},t} + \delta_2 z_{\text{longshot},t}\big),$$

with covariates $(z_{\text{herd},t}, z_{\text{longshot},t})$ and coefficients estimated out of sample.

Structural correction is a shrink map toward the anchor mean:

$$\phi_{\text{str},t}(p) = \lambda_t\, p + (1 - \lambda_t)\, \frac{\alpha_{\text{anchor},t}}{\alpha_{\text{anchor},t} + \beta_{\text{anchor},t}}, \quad \lambda_t \in [0, 1].$$

Define

$$p_{\text{fused},t} = \frac{\alpha_{\text{fused},t}}{\alpha_{\text{fused},t} + \beta_{\text{fused},t}}, \quad p_{\text{corr},t} = \phi_{\text{str},t}(\phi_{\text{beh},t}(p_{\text{fused},t})).$$

To preserve conjugate updating, PRISM reconstructs corrected Beta parameters with fused concentration

$$\kappa_{\text{fused},t} = \alpha_{\text{fused},t} + \beta_{\text{fused},t} - 2,$$

then

$$\alpha_{\text{corr},t} = 1 + \kappa_{\text{fused},t} p_{\text{corr},t}, \quad \beta_{\text{corr},t} = 1 + \kappa_{\text{fused},t}(1 - p_{\text{corr},t}).$$

What this section established.

- Fusion is implemented as weighted additive evidence relative to a flat baseline.
- Corrections are explicit maps with bounded domains and clear interpretation.
- Corrected parameters remain in a conjugate Beta family for downstream updating.

7 Posterior and Outputs

Purpose. This section defines the final posterior update and the reported output objects. The output layer is centered on probabilistic inference and diagnostics.

What you will learn.

- How corrected parameters combine with adjusted evidence.
- Closed-form posterior summaries used in PRISM reports.
- Which diagnostics must accompany posterior means.

Let y_t^* denote corrected yes-count evidence and let n_t^* denote corrected total count for window t (with $0 \le y_t^* \le n_t^*$). Given corrected prior parameters $(\alpha_{\mathrm{corr},t}, \beta_{\mathrm{corr},t})$, PRISM updates through

$$\pi_{\mathrm{post},t}(p) = \mathrm{Beta}(\alpha_{\mathrm{corr},t} + y_t^*, \beta_{\mathrm{corr},t} + n_t^* - y_t^*) \,.$$

The posterior mean is

$$p_{\mathrm{post},t} = \frac{\alpha_{\mathrm{corr},t} + y_t^*}{\alpha_{\mathrm{corr},t} + \beta_{\mathrm{corr},t} + n_t^*}.$$

The $(1-\eta)$ credible interval is obtained from Beta quantiles.

For every experiment scenario, posterior reporting includes:

- posterior mean and interval width,
- reliability-curve diagnostics,
- Brier score and log score,
- support flags (effective sample size, empty-bin checks, and calibration fallbacks).

Remark 2. PRISM outputs posterior beliefs about an event probability. Any downstream contract settlement or policy layer is external and must not be conflated with the posterior construction itself.

What this section established.

- Posterior updating remains closed-form under corrected evidence.
- Reported outputs are probabilistic summaries plus diagnostics.
- The output contract is inference-first and explicitly separated from downstream decisions.

8 Theory

Purpose. This section keeps only theory statements that follow from the revised PRISM workflow: flat structural prior, calibrated-signal fusion, and correction maps with explicit regularity assumptions.

What you will learn.

- When Beta moment matching from calibrated ML uncertainty is well-defined.
- Why conjugate updating is preserved after fusion and correction reconstruction.
- A conservative finite-sample sensitivity bound for posterior means.

Proposition 1 (Feasibility of Beta moment matching)**.** *For any $m \in (0,1)$ and v satisfying $0 < v < m(1-m)$, there exists a unique $(\alpha, \beta) \in (0, \infty)^2$ such that a* $\mathrm{Beta}(\alpha, \beta)$ *random variable has mean m and variance v.*

Proof. Set $\kappa = m(1-m)/v - 1$. The variance condition implies $\kappa > 0$. Then define $\alpha = m\kappa$ and $\beta = (1-m)\kappa$. Substitution gives the desired mean and variance. □

Proposition 2 (Conjugacy is preserved by PRISM reconstruction)**.** *If $\alpha_{\mathrm{corr},t} > 0$ and $\beta_{\mathrm{corr},t} > 0$, and if corrected evidence is represented by (y_t^*, n_t^*) with $0 \le y_t^* \le n_t^*$, then the PRISM posterior is Beta with parameters*

$$\alpha_{\mathrm{post},t} = \alpha_{\mathrm{corr},t} + y_t^*, \quad \beta_{\mathrm{post},t} = \beta_{\mathrm{corr},t} + n_t^* - y_t^*.$$

Proof. Immediate from the Beta-Binomial conjugate update identity. □

Assumption 1 (Lipschitz correction map)**.** *For each window t, the composition $\phi_t = \phi_{\mathrm{str},t} \circ \phi_{\mathrm{beh},t}$ is L_t-Lipschitz on $(0,1)$.*

Theorem 1 (Posterior-mean stability under bounded perturbations)**.** *Under Assumption 1, consider two data realizations producing fused probabilities $p_{\mathrm{fused},t}$ and $p'_{\mathrm{fused},t}$ and corrected counts (y_t^*, n_t^*), $(y_t^{*\prime}, n_t^{*\prime})$. If*

$$|p_{\mathrm{fused},t} - p'_{\mathrm{fused},t}| \le \Delta_p, \quad |y_t^* - y_t^{*\prime}| \le \Delta_y, \quad |n_t^* - n_t^{*\prime}| \le \Delta_n,$$

then there is a finite constant C_t depending on concentration bounds such that

$$|p_{\mathrm{post},t} - p'_{\mathrm{post},t}| \le C_t \left(L_t \Delta_p + \Delta_y + \Delta_n \right).$$

Proof. A direct decomposition into corrected-mean perturbation and count perturbation terms yields the bound. Details are in Appendix B. □

Remark 3. Asymptotic concentration claims require effective sample growth and dependence control assumptions. When those assumptions are weak in finite synthetic runs, PRISM treats large-sample claims as directional rather than fully validated [van der Vaart, 1998, Doukhan, 1994].

What this section established.

- Moment matching is mathematically well-posed on the feasible variance domain.
- Conjugate updating remains intact after correction-parameter reconstruction.
- Posterior means are Lipschitz-stable under bounded perturbations when correction maps are regular.

9 Validation Summary

Purpose. This section summarizes synthetic validation outcomes at a high level and points to the supplement for full diagnostics. The summary intentionally follows audit boundaries and does not convert indeterminate scenarios into broad claims.

What you will learn.

- Which experiment families were executed.
- Where validation is determinate versus indeterminate under current audit gates.
- Why current evidence supports a constrained interpretation.

Synthetic experiments cover:

- information-efficiency curves (A1),
- convergence timing (A2),
- strategic-timing attacks (A3),
- correction no-regret checks (B1),
- asymptotic-rate checks (B2),
- misspecification-regret grids (B3),
- regime concentration tests (B4),
- projection impact tests (B5).

Scenario-level status in 'results/**/audit.json' is mostly indeterminate for families A1, A3, B1, B2, B3, and B4 under current criteria evaluation and coverage gating. Determinate outcomes are concentrated in B5, with partial determinate support in A2.

At a high level:

- A2 currently has 2 determinate passes and 8 indeterminate scenarios.
- B5 currently has 6 determinate passes (with 3 marked paper-ready).
- A1, A3, B1, B2, B3, and B4 are currently indeterminate at scenario level.
- Paper-ready flags and determinacy do not coincide uniformly; both must be checked before any empirical claim.

These outcomes justify a cautious conclusion: the workflow is implementable and auditable, but broad empirical validation claims remain unsupported at family level under current determinacy patterns.

What this section established.

- Validation evidence is mixed and must be interpreted conservatively.
- Main text reports high-level status only; detailed diagnostics are deferred to supplement.
- Current empirical evidence supports workflow viability, not universal claim validation.

10 Discussion, Limitations, and Future Work

Purpose. This section records practical and statistical limitations so that workflow interpretation remains aligned with observed evidence and finite-sample realities.

Family	Scenario Status Pattern	Main-Paper Interpretation
A1	9 indeterminate / 9 total	exploratory only
A2	2 pass, 8 indeterminate / 10 total	supportive but limited
A3	8 indeterminate / 8 total	stress-sensitive and unresolved
B1	20 indeterminate / 20 total	correction robustness unresolved
B2	2 indeterminate / 2 total	asymptotic rate unresolved empirically
B3	8 indeterminate / 8 total	misspecification robustness unresolved
B4	4 indeterminate / 4 total	concentration claim unresolved
B5	6 pass / 6 total (3 paper-ready)	projection behavior conditionally supported

Table 1: High-level summary computed from scenario-level ‘audit.json‘ fields (‘criteria_evaluation.overall_pass‘, ‘seed_grid_coverage.paper_ready‘).

What you will learn.

- Which limitations are structural versus finite-sample.
- Where sensitivity enters through calibration and correction steps.
- Which extensions are natural and still within PRISM positioning.

The dominant current limitation is support quality in stress scenarios. Many scenarios in the synthetic suite carry low-support flags and off-nominal coverage diagnostics, which weakens direct empirical backing for broad robustness claims.

Calibration quality is a second limitation. When calibration data are sparse or shifted, uncertainty estimates widen and Beta concentration must be capped aggressively. This makes posterior intervals wider by design, which is appropriate but can reduce decisiveness.

Correction-map estimation is a third limitation. Behavioral and structural transforms are intentionally simple and bounded, but parameter error in these transforms can dominate posterior movement in adversarial regimes.

Future work that remains within PRISM scope includes:

- richer but still auditable anchor extraction from L3 message dynamics,
- stronger uncertainty quantification for calibration transport,
- regime-adaptive correction maps with explicit identifiability diagnostics,
- broader synthetic support before any stronger empirical claims.

What this section established.

- Current limits are explicit and tied to concrete pipeline stages.
- Conservative posterior interpretation is appropriate under present diagnostics.
- A clear future-work path exists without changing PRISM’s core scope.

11 Conclusion

Purpose. This section summarizes the refactored manuscript contribution in one statement: PRISM is a coherent Bayesian workflow for event-probability inference from L3 order flow and calibrated ML signals.

What you will learn.

- What the paper now contributes.
- What is supported versus still open.
- Why the revised workflow is auditable end-to-end.

PRISM delivers a transparent probabilistic pipeline with explicit stage contracts, flat structural anchoring, calibrated ML uncertainty transport, and conjugate posterior outputs. The manuscript removes unsupported structural machinery and keeps theory aligned with the revised workflow.

Empirical evidence from current synthetic artifacts is mixed; therefore conclusions remain conservative. The contribution is a rigorous and auditable inference workflow, not a claim of universal empirical dominance.

What this section established.

- The paper now presents a coherent PRISM workflow with consistent notation and scope.
- Theoretical and empirical claims are bounded to what the current artifacts support.
- Main and supplement jointly provide reproducible documentation for further extension.

A Project Lineage

This appendix records historical lineage only.

Earlier manuscript generations were written under the project name **CAPOPM**. Those drafts used phase-based organization and included structural machinery that is not part of the current canonical workflow. The present paper refactors that legacy into PRISM with a narrower and cleaner scope: posterior inference for event probabilities from L3 order flow and calibrated machine-learning signals.

Lineage notes:

- Legacy naming was replaced in the active manuscript with PRISM terminology.
- Legacy phase framing was replaced by section-based workflow narrative.
- Legacy structural components not required by the current workflow were removed from active text.
- Original legacy sources are preserved unchanged in `paper/prism/_legacy/` for audit trail.

This appendix is the only location in the active manuscript where the legacy acronym appears.

B Proof Details

This appendix provides details omitted from Section 8.

Details for Proposition 1

The mapping $(m, v) \mapsto (\alpha, \beta)$ uses

$$\kappa = \frac{m(1-m)}{v} - 1, \quad \alpha = m\kappa, \quad \beta = (1-m)\kappa.$$

The strict inequality $v < m(1-m)$ implies $\kappa > 0$, so $\alpha, \beta > 0$. Uniqueness follows because Beta mean and variance identify (α, β) uniquely on $(0, \infty)^2$.

Details for Proposition 2

For likelihood term

$$L(p \,|\, y^*, n^*) \propto p^{y^*}(1-p)^{n^*-y^*}$$

and prior

$$\pi(p) \propto p^{\alpha_{\mathrm{corr}}-1}(1-p)^{\beta_{\mathrm{corr}}-1},$$

the posterior kernel is

$$p^{\alpha_{\mathrm{corr}}+y^*-1}(1-p)^{\beta_{\mathrm{corr}}+n^*-y^*-1},$$

which is Beta.

Details for Theorem 1

Let

$$\mu_t = \frac{\alpha_t + y_t^*}{\alpha_t + \beta_t + n_t^*}, \quad \mu_t' = \frac{\alpha_t' + y_t^{*\prime}}{\alpha_t' + \beta_t' + n_t^{*\prime}}.$$

Use triangle inequality with intermediate $(\alpha_t', \beta_t', y_t^*, n_t^*)$. Parameter perturbations $(\alpha_t, \beta_t) \leftrightarrow p_{\mathrm{corr},t}$ are controlled by Lipschitz correction and bounded concentration. Count perturbations are controlled by bounded derivatives:

$$\left|\frac{\partial \mu_t}{\partial y_t^*}\right| \le \frac{1}{\underline{c}_t}, \qquad \left|\frac{\partial \mu_t}{\partial n_t^*}\right| \le \frac{1}{\underline{c}_t},$$

for lower concentration bound $\underline{c}_t > 0$. Combining yields the stated bound.

C Additional Results Notes

This appendix records concise notes on artifact interpretation.

- Several scenario outputs include low-support and coverage-deviation flags. These flags are treated as first-order qualifiers, not secondary annotations.
- Mixed pass/fail patterns within a family indicate regime sensitivity and do not support unconditional family-level validation claims.
- Where calibration fallback rules were triggered, posterior concentration is intentionally reduced to prevent overconfident inference.

For exact scenario-level values, refer to:

- `results/**/audit.json`
- `results/**/summary.json`
- `results/**/tests.csv`

References

Boris S. Axelrod, Ben J. Kulick, Charles R. Plott, and Kevin A. Roust. The design of improved parimutuel-type information aggregation mechanisms: Inaccuracies and the long-shot bias as disequilibrium phenomena. *Journal of Economic Behavior and Organization*, 69:170–181, 2009. doi: 10.1016/j.jebo.2007.09.010.

Paul Doukhan. *Mixing: Properties and Examples*. Springer, New York, 1994.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1994.

Frederic Koessler, Charles Noussair, and Anthony Ziegelmeyer. Information aggregation and belief elicitation in experimental parimutuel betting markets. *Journal of Economic Behavior and Organization*, 83:195–208, 2012. doi: 10.1016/j.jebo.2012.02.017.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.