hexagon:2611.00663v1

Finite-resolution information about a common clock: alias depth, holding-time loss, and the sampled Ehrenfest chain

Samuil Petkov
Department of Physics, École normale supérieure, Université PSL, Paris, France
samuil.petkov@phys.ens.psl.eu

30 September 2026

Abstract

We study how much Fisher information about a common rate multiplier survives when a reaction system is observed only at discrete times. For a finite known mark law, the first loss at short sampling times is governed by an alias depth h, the smallest length mismatch among reaction-count fibres: the unrecovered fraction of reaction-count information is $1 - F_p(\lambda) = C_p\lambda^{h-1} + O(\lambda^h)$ with an explicit alias coefficient C_p. For a fixed irreducible finite generator, holding times add a separate loss: the missing log-clock information is $A_2t^2 + O(t^3)$, or $(A_3 + H)t^3 + O(t^4)$ when $A_2 = 0$. As a worked example we treat the stationary symmetric Ehrenfest chain of known size Ω observed at its endpoints, with sampling time t and $\lambda = \Omega t$. For every population and every finite positive t, the exact endpoint information lies between an explicit quadratic lower bound and its matched Gaussian benchmark, $B_n \le R_n < g(t)/\lambda$, with relative width at most $1/(2\lambda)$ measured from B_n. The symmetric compound-Poisson benchmark is uniformly relatively accurate on $1 \le \lambda \le L_\Omega$ if and only if $L_\Omega = o(\Omega)$. A formula-level comparison with the binomial AR(1) literature (Cui–Lund; Weiß–Kim) has not been completed.

Mathematics Subject Classification (2020). Primary 62M05; Secondary 60J27, 62B10.
Keywords. Fisher information; missing information; discrete-time sampling; alias depth; holding times; Ehrenfest chain; compound-Poisson approximation.

1 Introduction

Sampling a reaction system at discrete times does not reveal the intervening reaction counts or holding times. Both can carry information about a parameter that multiplies every transition rate. We quantify how much of the complete-trajectory Fisher information about such a common clock survives when only the sampled states are observed, and we locate the first loss exactly. The inferred clock is a fixed parameter; no distribution of random external inputs is assumed, and the physical sampling interval Δ is held fixed under differentiation. With $t = \theta\Delta$ the dimensionless sampling time, the information fraction retained is a function of t, or of $\lambda = \Omega t$ for a population of size Ω.

Results. *Reaction-count aliases* (Section 2). For a finite known mark law with nonzero integer marks, the *alias depth* h is the smallest value of $\max\{|z|, |z'|\}$ over pairs of reaction-count vectors z, z' of different total length with the same increment. Theorem 2.2 gives the loss of the compound-Poisson count experiment as $1 - F_p(\lambda) = C_p\lambda^{h-1} + O_p(\lambda^h)$ with an explicit positive *alias coefficient* C_p, and $F_p \equiv 1$ when $h = \infty$. Corollary 2.3 writes C_p in terms of the h-step mark random walk and evaluates it for nearest-neighbour and simplex mark laws. Two mark laws with identical Gaussian moments can have different exponents (Section 2.1).

Holding-time loss (Section 3). For a fixed finite irreducible generator Q, Theorem 3.2 classifies the missing log-clock information $\mathcal{M}_Q(t)$ between the complete trajectory and its endpoints: $\mathcal{M}_Q(t) = A_2(Q)t^2 + O_Q(t^3)$, and $\mathcal{M}_Q(t) = [A_3(Q) + H(Q)]t^3 + O_Q(t^4)$ when $A_2(Q) = 0$. Here A_2, A_3 are explicit path-alias coefficients

and $H \geq 0$ is the holding-time term, positive exactly when the escape rates are not all equal. Every reversible generator has $A_2 > 0$ (Remark 3.3), so the t^3 regime requires a generator without two-cycles.

The stationary Ehrenfest chain (Sections 4–7) is the worked example, observed at its full endpoint pair for a known population size Ω. Theorem 6.2 gives, for every Ω and finite $t > 0$, the finite Gaussian envelope $B_n(t) \leq R_n(\lambda) < g(t)/\lambda$ for the retained information fraction R_n, sharp as $\Omega \to \infty$ at fixed t. Theorem 7.1 shows that the symmetric compound-Poisson fraction $F(\lambda)$ approximates $R_\Omega(\lambda)$ uniformly in relative error on $1 \leq \lambda \leq L_\Omega$ if and only if $L_\Omega = o(\Omega)$.

Scope. All statements concern Fisher information, not the efficiency of a particular estimator. The alias and holding-time expansions hold for known finite mark laws, finite constant-escape chains and fixed finite irreducible generators; their remainders are not uniform over increasing populations or disappearing edges. For the Ehrenfest chain, the Gaussian envelope bounds the exact endpoint information at every population and positive sampling time, whereas the sharp window concerns the compound-Poisson fraction. Approximation accuracy does not mean that the retained fraction of trajectory information is close to one. The constants in Theorems 5.2 and 7.1 are universal but not optimized, and the Gaussian envelope is proved for the stationary symmetric Ehrenfest experiment only.

Status of the comparison with binomial autoregression. A formula-level comparison of the Ehrenfest results below with Cui and Lund [8] and Weiß and Kim [34, 33] (binomial AR(1)) has not been completed. The paired aggregate counts of the Ehrenfest chain belong to the classical bivariate-binomial and binomial-autoregressive families, so parts of the Ehrenfest results may already follow from moment calculations in that literature; we have not checked which. No priority over those articles is claimed for the Ehrenfest results.

Relation to previous work. Duval and Hoffmann [13, Definition 1 and Theorems 1–3] studied intensity inference for the same limiting compound-Poisson process with symmetric Bernoulli marks in microscopic, intermediate and macroscopic sampling regimes. Their published increment law and score [13, Eqs. (18) and (22)–(23)] give the benchmark F used here. Their estimator comparison [13, Theorems 4–5] is relative to the sampled-data experiment, not to a continuously observed jump trajectory. Further coarse-scale limitations for compound-Poisson inference appear in [12]; high-frequency nonparametric estimation of jump laws [7] and high-frequency Fisher-information asymptotics for Lévy models [1] address different experiments.

Likelihood computation for discretely observed finite-state chains is established [18, 5]; Kalbfleisch and Lawless obtain the score and expected information of panel-observed chains from derivatives of the transition matrix. For one-parameter birth or death processes, the choice of observation times and the comparison of discrete with continuous sampling have been treated as optimal-design problems [4, 25], and Aït-Sahalia and Mykland [2] compare the losses caused by discrete and by random sampling of diffusions. We have not compared their small-interval behaviour with Theorem 3.2. Our use of the word alias refers to unequal-length reaction-count fibres. It is distinct from the aliasing and identification problem for continuous-time models observed at discrete times [27], and from the embedding problem of whether, and how uniquely, a transition matrix is the exponential of a generator [19, 17]. Here the generator is known up to the common clock, and the question is how much Fisher information about that clock the sampled states retain. The conditional-score identity used below is the missing-information principle [22, 28]. Endpoint-conditioned jump counts, occupation-time moments, and their use in observed information are established for parameterized finite chains [15, 16, 32]. For reaction networks, see pathwise information [26] and experimental design [20, 14]. Transition-score additivity and Gaussian Fisher-information calculations are established ingredients for correlated processes [30]; the standard Gaussian mean-and-covariance formula [30, Section 3.1] supplies our Gaussian benchmark. Diaconis and Griffiths [10, Theorem 1 and Section 6] characterize correlated Bernoulli pair representations and the associated Ehrenfest transition kernels, including their normal and Poisson process limits; the stationary endpoint representation used here is a classical special case. The quadratic lower projection uses classical score-projection principles and fourth moments.

Moor et al. [24] already demonstrate Gaussian reaction-versus-state information discrepancies and their connection to multistate sensing. Kinetic and clock uncertainty relations [9, 29, 23] identify dynamical activity, timing information, and sufficient count statistics. Integer fibres and moves are standard in algebraic statistics [11], and exact cancellation of lattice jumps and its effect on compound-Poisson intensity estimation

are explicit in Coca [6, Section 2.3.1, Eq. (2.17), and Lemma 2.2], whose jump law is unknown. Here the finite mark law is known, and we identify the first unequal-length fibre and its parametric Fisher-information loss coefficient.

2 Reaction-count aliases determine the first loss term

Let $\nu_1, \dots, \nu_r \in \mathbb{Z}^d \setminus \{0\}$ be a finite mark set, with $p_i > 0$, $\sum_i p_i = 1$. Repeated marks are allowed. Write S for the matrix with these columns and let

$$Z_i \overset{\text{ind}}{\sim} \text{Pois}(\lambda p_i), \quad K = \sum_i Z_i, \quad D = SZ, \quad F_p(\lambda) = 1 - \lambda^{-1}\mathbb{E}\,\text{Var}(K \mid D). \tag{1}$$

This is the frozen-intensity experiment; it is not automatically the exact one-bin experiment of a state-dependent network.

Definition 2.1 (Unequal-length alias depth)**.** For $d \in \mathbb{Z}^d$ and $k \geq 0$ put

$$b_{d,k} = \sum_{z \in \mathbb{N}^r:\, Sz = d,\, |z| = k} \prod_{i=1}^{r} \frac{p_i^{z_i}}{z_i!}. \tag{2}$$

If the fibre over d supports at least two distinct lengths, denote its smallest and next smallest lengths by m_d and k_d. Let $h = \min_d k_d$, or $h = \infty$ if no such fibre exists.

Equivalently, h is the smallest value of $\max\{|z|, |z'|\}$ for $Sz = Sz'$ and $|z| \neq |z'|$. $\mathbb{N}$ denotes the nonnegative integers. Same-length aliases do not hide the total reaction count and hence do not lose information about its common intensity.

Theorem 2.2 (Alias-depth expansion)**.** *Let F_p be the retained fraction of the frozen experiment* (1)*, and let h, $b_{d,k}$, m_d, k_d be as in the preceding definition. If $h < \infty$, then $h \geq 2$ and*

$$1 - F_p(\lambda) = C_p \lambda^{h-1} + O_p(\lambda^h), \qquad C_p = \sum_{d:\, k_d = h} (h - m_d)^2 b_{d,h} > 0. \tag{3}$$

If $h = \infty$, then $F_p(\lambda) = 1$ for every $\lambda > 0$.

The proof of Theorem 2.2 is in Appendix A.

The coefficient has a random-walk form. Let $\xi_1, \xi_2, \dots$ be independent with $\mathbb{P}(\xi_\ell = \nu) = \sum_{i:\, \nu_i = \nu} p_i$, put $\zeta_k = \xi_1 + \cdots + \xi_k$ ($\zeta_0 = 0$), and let $\mathcal{Z}_k = \{d : \mathbb{P}(\zeta_k = d) > 0\}$.

Corollary 2.3 (Random-walk form of the alias coefficient)**.** *If $h < \infty$, then*

$$C_p = \frac{1}{h!} \sum_{m=0}^{h-1} (h - m)^2\, \mathbb{P}(\zeta_h \in \mathcal{Z}_m). \tag{4}$$

For every finite nonzero mark law, $1 - F_p(\lambda) = \big[2\mathbb{P}(\zeta_2 = 0) + \frac{1}{2}\mathbb{P}(\zeta_2 \in \mathcal{Z}_1)\big]\lambda + O_p(\lambda^2)$, and the bracket vanishes exactly when $h \geq 3$. In particular:

1. *for nearest-neighbour marks $\pm e_j$ in $\mathbb{Z}^d$ with probabilities $p_j^{\pm} \geq 0$ (marks of probability zero omitted), one has $h = 2$ if some $p_j^+ p_j^- > 0$, and then $C_p = 4 \sum_j p_j^+ p_j^-$; otherwise $h = \infty$;*
2. *for the simplex marks $e_1, \dots, e_d, -(e_1 + \cdots + e_d)$ in $\mathbb{Z}^d$, $d \geq 1$, with probabilities $p_1, \dots, p_d, p_0 > 0$, one has $h = d + 1$ and $C_p = (d+1)^2 \prod_{i=0}^{d} p_i$.*

Proof. By the multinomial theorem, $b_{d,k} = \mathbb{P}(\zeta_k = d)/k!$. For $m < h$, the sets $\mathcal{Z}_m$ are pairwise disjoint, since a common point would be an unequal-length alias of maximal length below h. If $d \in \mathcal{Z}_m \cap \mathcal{Z}_h$ with $m < h$, the same argument shows that m is the smallest length of the fibre over d and that no length lies strictly between m and h; hence $m_d = m$ and $k_d = h$. Conversely, $k_d = h$ means $d \in \mathcal{Z}_{m_d} \cap \mathcal{Z}_h$. Summing over m gives (4). If

$h = 2$, only $m = 0, 1$ occur, giving the bracket. If $h \geq 3$, both probabilities in the bracket are zero, because a positive one would be an alias of lengths $0, 2$ or $1, 2$, and Theorem 2.2 gives $1 - F_p = O_p(\lambda^2)$; the case $h = \infty$ is trivial. For nearest-neighbour marks, every coordinate sum of ζ_2 is even and of ζ_1 odd, so $\mathbb{P}(\zeta_2 \in \mathcal{Z}_1) = 0$ and $\mathbb{P}(\zeta_2 = 0) = 2\sum_j p_j^+ p_j^-$. If no opposite pair is present, the increment determines every count. For the simplex marks, the integer kernel of the mark matrix is spanned by $(1, \ldots, 1)$, so distinct-length aliases differ by a positive multiple of that vector. The smallest maximal length is $d + 1$, attained only by the return to zero from the empty word, with $\mathbb{P}(\zeta_{d+1} = 0) = (d + 1)! \prod_i p_i$. □

For the two laws of Section 2.1 below, (4) reproduces $C_A = 9 \cdot \frac{1}{6} \cdot \frac{2}{9} = \frac{1}{3}$ and $C_B = \frac{1}{2}\big(4 \cdot \frac{1}{6} + \frac{1}{3}\big) = \frac{1}{2}$.

Remark 2.4 (A graph version). The same proof applies to every finite, constant-escape-rate chain $q_\theta = \theta a(P - I)$ with $P_{ii} = 0$, started in a positive stationary law. Replace $b_{d,k}$ by $\pi_i (P^k)_{ij}/k!$ and observe (i, j). The alias depth is then the first pair of different-length directed walks with the same starting and ending states. Artificial self-transitions from uniformization must not be counted as observed reactions.

2.1 Identical Gaussian moments do not fix the exponent

Consider the two frozen mark laws

$$\begin{aligned} A &: (1,0),\ (0,1),\ (-1,-1), && p = 1/3, \\ B &: \pm(1,0),\ \pm(0,1),\ \pm(-1,-1), && p = 1/6. \end{aligned}$$

Both have zero mean and second-moment matrix $C = \frac{1}{3}\left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$. Their Gaussian increment experiments are therefore identical: $D_G \sim N(0, \lambda C)$. In law A, the first unequal-length alias is the three-jump return to zero; $b_{0,3} = 1/27$. For law B, the two-jump return has $b_{0,2} = 1/12$ and each of its six marks also has a two-jump representation with $b_{d,2} = 1/36$. Theorem 2.2 yields

$$1 - F_A(\lambda) = \tfrac{1}{3}\lambda^2 + O(\lambda^3), \qquad 1 - F_B(\lambda) = \tfrac{1}{2}\lambda + O(\lambda^2). \tag{5}$$

This is a counterexample at the fixed-intensity level, not a claim that two complete state-dependent networks have the same drift Jacobian.

3 Holding-time information loss

Constant escape is essential to replacing the complete score by a jump count. Let Q be an irreducible finite-state row generator on at least two states, with off-diagonal rates q_{ij}, escape rates $a_i = \sum_{j \neq i} q_{ij}$ and stationary law π. Take $X_0 \sim \pi$; this initial law is independent of the common clock parameter. The unknown parameter multiplies Q. In dimensionless time $t = \theta\Delta$, the complete log-θ score is

$$S_t = K_t - \int_0^t a(X_u)\, du, \qquad \mathbb{E}S_t^2 = \bar{a}t, \quad \bar{a} = \sum_i \pi_i a_i. \tag{6}$$

The endpoint-conditioned count and dwelling-time machinery, including the observed-score and missing-information identities, is established [15, 16]; the following coefficient makes the holding-time contribution to information loss explicit.

Proposition 3.1 (Irreducible holding-time loss)**.** *Let $\mathcal{H}$ contain the initial state and the ordered visited-state sequence, but not the jump times. As $t \downarrow 0$,*

$$\mathbb{E}\,\mathrm{Var}(S_t \mid \mathcal{H}) = \frac{t^3}{12} \sum_{i, j \neq i} \pi_i q_{ij} (a_i - a_j)^2 + O_Q(t^4). \tag{7}$$

The missing information when only the endpoints are observed is at least this quantity.

Proof. For exactly one jump $i \to j$, its time U has conditional density proportional to $e^{-a_i u - a_j(t-u)}$ on $[0,t]$. Rescaling by t and expanding the bounded exponential shows $\mathrm{Var}(U \mid \mathcal{H}) = t^2/12 + O_Q(t^3)$. On this event, $S_t = 1 - a_j t - (a_i - a_j)U$, and the event probability is $\pi_i q_{ij} t + O_Q(t^2)$. Summation gives the displayed leading term. For no jumps there is no conditional variance. On a skeleton with at least two jumps, the integral in (6) lies in an interval of length at most $a_{\max} t$, whereas K_t is fixed by the skeleton. Such skeletons have total probability $O_Q(t^2)$ by Poisson domination, giving an $O_Q(t^4)$ contribution. Finally, the endpoint observation is a function of $\mathcal{H}$, so conditional variance decomposition proves the lower bound. □

3.1 A local classification for every finite irreducible chain

The holding-time term can be combined with path aliases to classify the first loss order for every finite irreducible chain. Write B for the off-diagonal part of Q, so that $B_{ii} = 0$, and let m_{ij} be the shortest directed-walk length from i to j, with $m_{ii} = 0$. Define the graph alias depth h as the smallest second distinct walk length among all ordered endpoint pairs. Irreducibility ensures $h < \infty$. Put

$$A_2(Q) = \frac{1}{2} \sum_i \pi_i \left[4(B^2)_{ii} + \sum_{j \neq i:\, q_{ij} > 0} (B^2)_{ij} \right], \tag{8}$$

$$H(Q) = \frac{1}{12} \sum_{i, j \neq i} \pi_i q_{ij} (a_i - a_j)^2. \tag{9}$$

When $A_2(Q) = 0$, also define

$$A_3(Q) = \frac{1}{6} \sum_{i,j:\, m_{ij} < 3} \pi_i (3 - m_{ij})^2 (B^3)_{ij}. \tag{10}$$

Theorem 3.2 (Endpoint information at small times)**.** *For a fixed finite irreducible Q on at least two states, let*

$$\mathcal{M}_Q(t) = \bar{a} t - t^2 \sum_{i,j} \pi_i \frac{((Qe^{tQ})_{ij})^2}{(e^{tQ})_{ij}}. \tag{11}$$

This is the missing log-clock information between the full trajectory and its endpoints. With A_2, H, A_3 from (8)–(10) *and the graph alias depth h defined above, as $t \downarrow 0$,*

$$\mathcal{M}_Q(t) = A_2(Q) t^2 + O_Q(t^3). \tag{12}$$

If $A_2(Q) = 0$, then

$$\mathcal{M}_Q(t) = [A_3(Q) + H(Q)] t^3 + O_Q(t^4). \tag{13}$$

Consequently, if escape rates are not constant, the fractional information loss has power $\min\{h-1, 2\}$*. If they are constant, its power is $h-1$ with the coefficient given by the graph form of Theorem 2.2. All remainders concern a fixed generator, not a family with vanishing edges or increasing state space.*

The proof of Theorem 3.2 is in Appendix B.

For a heterogeneous three-cycle with rates $(1, 2, 3)$, for example, $\bar{a} = 18/11$, $A_2 = 0$, $A_3 = 9$, and $H = 3/11$. Hence $\mathcal{M}_Q(t) = (102/11)t^3 + O(t^4)$: count and holding-time effects enter at the same order and add. For the Ehrenfest chain with Ω particles, the same formula gives $A_2 = \Omega(\Omega + 1)$ and $H = 0$. Dividing by the full information Ωt, the leading loss of the retained fraction is $(1 + 1/\Omega)\lambda$ with $\lambda = \Omega t$.

Remark 3.3 (Reversible and birth–death generators). If Q is reversible, every allowed edge $i \to j$ has $q_{ji} = \pi_i q_{ij}/\pi_j > 0$, so $(B^2)_{ii} \geq q_{ij} q_{ji} > 0$ and $A_2(Q) > 0$. The endpoint loss of a reversible generator is therefore always of order t^2; the t^3 regime of (13), where H enters, requires a generator without two-cycles. For a birth–death chain with birth rates β_i, death rates δ_i (absent rates set to zero) and $a_i = \beta_i + \delta_i$, there are no two-step shortcuts, and detailed balance $\pi_i \beta_i = \pi_{i+1} \delta_{i+1}$ gives

$$A_2(Q) = 2 \sum_i \pi_i (\beta_i \delta_{i+1} + \delta_i \beta_{i-1}), \qquad H(Q) = \frac{1}{6} \sum_i \pi_i \beta_i (a_{i+1} - a_i)^2.$$

Here H is the explicit holding-time lower bound of Proposition 3.1, not a leading-order term. For the Ehrenfest chain, $a_i \equiv \Omega$ gives $H = 0$ and $A_2 = \Omega(\Omega + 1)$, as above.

4 Worked example: observation protocol and exact information for the Ehrenfest chain

Fix $T > 0$, $\theta \in [\theta_-, \theta_+] \subset (0, \infty)$, and an integer $\Omega \geq 1$. Let N be the stationary chain on $\{0, \dots, \Omega\}$ with

$$q_\theta(n, n+1) = \theta(\Omega - n), \qquad q_\theta(n, n-1) = \theta n. \tag{14}$$

The stationary law $\pi_\Omega(n) = 2^{-\Omega}\binom{\Omega}{n}$ is independent of θ. We observe exact instantaneous counts at $0, \Delta, \dots, M\Delta$, where

$$M = \lfloor T/\Delta \rfloor, \quad \tau = M\Delta, \quad t = \theta\Delta, \quad \lambda = \Omega\theta\Delta. \tag{15}$$

Every parameter derivative holds (Ω, Δ, T) fixed. In particular, λ labels an asymptotic regime and is not a parameter-dependent design.

4.1 Finite-sum likelihood

Each of the Ω independent particles flips between two states at rate θ. Set $p = (1 - e^{-2t})/2$. Conditional on $N_0 = n$, the increment is $U - V$, where $U \sim \mathrm{Bin}(\Omega - n, p)$ and $V \sim \mathrm{Bin}(n, p)$ independently. Writing $d = m - n$ gives

$$P_t(n, m) = \sum_v \binom{\Omega - n}{v + d}\binom{n}{v} p^{2v+d}(1-p)^{\Omega - 2v - d}, \tag{16}$$

with the sum restricted to valid binomial indices.

Proposition 4.1 (Exact sampled information)**.** *The information about θ in the sampled counts is*

$$\mathcal{I}_{\Omega,\Delta,T}(\theta) = M\Delta^2 \sum_{n,m=0}^{\Omega} \pi_\Omega(n) \frac{(\partial_t P_t(n, m))^2}{P_t(n, m)}. \tag{17}$$

If K is the actual number of jumps in one interval, define

$$R_\Omega(\lambda) = \frac{\mathrm{Var}(\mathbb{E}[K \mid N_0, N_\Delta])}{\lambda}. \tag{18}$$

Then

$$R_\Omega(\lambda) = 1 - \frac{\mathbb{E}\,\mathrm{Var}(K \mid N_0, N_\Delta)}{\lambda}, \qquad \frac{\mathcal{I}_{\Omega,\Delta,T}(\theta)}{\Omega T/\theta} = \frac{\tau}{T} R_\Omega(\lambda). \tag{19}$$

Proof. The log likelihood is a sum of transition log likelihoods; its initial term has zero score. Each transition score has conditional mean zero given the past. Cross terms therefore vanish, giving (17) by stationarity. The total escape rate is $\Omega\theta$, independently of the state, so $K \sim \mathrm{Pois}(\lambda)$ independently of N_0. Conditional on K, the embedded chain and jump times have no further dependence on θ. The complete score is $(K - \lambda)/\theta$. Differentiation under the finite-state Poisson mixture gives the observed score $\mathbb{E}[K - \lambda \mid N_0, N_\Delta]/\theta$. Conditional variance decomposition proves (19). Over $[0, T]$ the full score has variance $\Omega T/\theta$. □

The transition-score additivity used here is the stationary first-order case of the Markov formula in [30, Eqs. (30) and (34)], with zero initial-law information. It follows from conditional zero scores, not from independence of successive observations. The observed score is the conditional expectation of the complete score, an instance of the missing-information principle [22, 28]; here direct differentiation establishes the identity for this finite-state Poisson mixture.

5 The compound-Poisson benchmark

Let U, V be independent $\mathrm{Pois}(\lambda/2)$ variables, and put $K = U + V$, $D = U - V$. Define

$$F(\lambda) = \frac{\mathrm{Var}(\mathbb{E}[K \mid D])}{\lambda} = \lambda \sum_{d \in \mathbb{Z}} \frac{(p'_\lambda(d))^2}{p_\lambda(d)}, \qquad p_\lambda(d) = e^{-\lambda} I_{|d|}(\lambda). \tag{20}$$

This is the symmetric compound-Poisson intensity experiment of [13, Eqs. (18), (22)–(23) and (29)], normalized by complete-jump information. We record elementary estimates needed for the finite-population reduction, independently of a Gaussian limit theorem.

Lemma 5.1. *For every $\lambda > 0$,*

$$0 < F(\lambda) < 1, \qquad \frac{1}{1+2\lambda} \le F(\lambda) \le \min\{1, 2/\lambda\}. \tag{21}$$

Moreover F is continuous on $(0,\infty)$, $F(\lambda) \to 1$ as $\lambda \downarrow 0$, and $F(\lambda) \to 0$ as $\lambda \to \infty$.

The proof of Lemma 5.1 is in Appendix C.

5.1 Uniform additive and relative comparisons

Theorem 5.2 (Uniform additive crossover)**.** *Let R_Ω be the retained fraction* (18) *of the stationary Ehrenfest experiment and F the compound-Poisson fraction* (20)*. There is a universal constant C such that, for every integer $\Omega \ge 1$,*

$$\sup_{\lambda>0} |R_\Omega(\lambda) - F(\lambda)| \le C\Omega^{-1/4}\sqrt{\log(e\Omega)}. \tag{22}$$

Consequently, uniformly for $0 < \Delta \le T$ and $\theta \in [\theta_-, \theta_+]$,

$$\left| \frac{\mathcal{I}_{\Omega,\Delta,T}(\theta)}{\Omega T/\theta} - F(\Omega\theta\Delta) \right| \le C\Omega^{-1/4}\sqrt{\log(e\Omega)} + \frac{2}{\Omega\theta_- T}. \tag{23}$$

The clipped-score and tail estimates, and the complete proof of Theorem 5.2, are given in Appendix C.

Theorem 5.2 gives additive error even when the information fraction vanishes. The relative approximation in Section 7 identifies its exact window of validity. At finite population, the fraction is not a function independent of Ω: for example,

$$R_1(\lambda) = \frac{4\lambda}{e^{4\lambda}-1} \ne F(\lambda). \tag{24}$$

6 A finite Gaussian envelope for the Ehrenfest chain

Write $n = \Omega$, $t = \theta\Delta$ and $\lambda = nt$. Take the exact endpoint pair $Y = (N_0, N_\Delta)$, actual reaction count K, and endpoint log-clock score $S = \mathbb{E}[K - \lambda \mid Y]$. Thus $\mathbb{E}S^2 = \lambda R_n(\lambda)$. For the quadratic certificates, define

$$A = N_0 + N_\Delta - n, \quad D = N_\Delta - N_0, \quad p = (1-e^{-2t})/2, \quad u = n(1-p), \quad q = np,$$

and $W_1 = A^2 - u$, $W_2 = D^2 - q$. Let Π_2 be orthogonal projection in endpoint L^2 onto $\operatorname{span}\{W_1, W_2\}$. The centering constants are evaluated at the parameter of interest and held fixed when differentiating score moments. The complete projection calculation is in Appendix D. Define also the explicit scalar reference

$$g(t) = 4t^2 \frac{e^{-4t}(1+e^{-4t})}{(1-e^{-4t})^2} = \tfrac{1}{2} - t + O(t^2). \tag{25}$$

6.1 All-population information bounds

The following bounds hold at every population size. They use the classical conditional-score principle and a model-specific quadratic projection, together with the prediction-risk form of missing information.

Define

$$H_0(t) = \frac{4t}{e^{4t}-1}, \qquad B_n(t) = \frac{4t(e^{4t}+1)}{(e^{4t}-1)\{n(e^{4t}-1)+2\}}, \tag{26}$$

$$c_n = 1 + \frac{1}{n}, \qquad \ell_n(\lambda) = \frac{c_n \lambda e^{-\lambda}}{1+\lambda^2/2},$$

$$U_n(\lambda) = 1 + \lambda - \left[1 + \left(1 - \frac{1}{n}\right)\lambda\right] e^{-\lambda}. \tag{27}$$

Both H_0 and B_n have continuous value one at zero, while $\ell_n(0)$ and $U_n(0)$ are zero. Whenever a zero-time information fraction is used, it means the limit $R_n(0) = 1$, not the quotient in (18) at zero.

Theorem 6.1 (All-population information bounds)**.** *In the stationary experiment of Proposition 4.1, for every integer $n \geq 1$ and $t > 0$,*

$$\max\{B_n(t), 1 - U_n(nt)\} \leq R_n(nt) \leq \min\{H_0(t), 1 - \ell_n(nt)\}. \tag{28}$$

In particular,

$$\ell_n(\lambda) \leq 1 - R_n(\lambda) \leq U_n(\lambda), \qquad \lambda > 0. \tag{29}$$

For $n = 1$, the labelled and aggregate endpoint experiments coincide, and $R_1(t) = B_1(t) = H_0(t)$.

The proof of Theorem 6.1 is in Appendix D.

Relation to the sharp window. The quadratic lower bound has a direct connection to the matched Gaussian information $g(t)$ in (25):

$$ntB_n(t) = g(t)\frac{n(e^{4t} - 1)}{n(e^{4t} - 1) + 2}. \tag{30}$$

It has the correct leading information when $\lambda \to \infty$ and $\lambda = o(n)$, since $\lambda B_n(\lambda/n) \to 1/2$, and at fixed $0 < t \leq (\log 2)/2$, where $ntB_n(t) \to g(t)$ and the finite residual certificate below gives $ntR_n(nt) \to g(t)$. In contrast, at fixed $\lambda > 0$, $B_n(\lambda/n) \to 1/(1 + 2\lambda)$; the lower bound does not recover the whole function $F(\lambda)$. For $n \geq 2$, the quadratic projection defined above satisfies

$$\lambda\{R_n(\lambda) - B_n(\lambda/n)\} = \mathbb{E}[(S - \Pi_2 S)^2]. \tag{31}$$

Section 6.3 gives an explicit upper certificate for this residual. It encloses information relative to the quadratic reference B_n. Together with the clipped-score coupling, it also proves the sharp relative comparison with F in Section 7.

6.2 A finite Gaussian envelope

The matched Gaussian pair has means $(n/2, n/2)$, marginal variances $n/4$, and correlation $\rho = e^{-2t}$, as do the exact counts. Its log-clock information is $g(t)$ from (25). The proof uses the posterior parity recurrence to construct three latent polynomials with zero conditional mean; their Gram matrix bounds the conditional second Krawtchouk score. It includes every endpoint cell and uses no Gaussian approximation.

Theorem 6.2 (A finite Gaussian information envelope)**.** *In the stationary Ehrenfest experiment of Proposition 4.1, let R_n be the retained information fraction* (18)*, B_n the quadratic reference* (26) *and g the matched Gaussian information* (25)*. For every integer $n \geq 1$ and finite $t > 0$, with $\lambda = nt$,*

$$B_n(t) \leq R_n(\lambda) < \frac{g(t)}{\lambda}. \tag{32}$$

In particular,

$$1 \leq \frac{R_n(\lambda)}{B_n(t)} < 1 + \frac{2}{n(e^{4t} - 1)} \leq 1 + \frac{1}{2\lambda}. \tag{33}$$

Thus $\lambda \geq 5$ suffices for a relative interval of width at most ten percent around B_n, at every population size and sampling time. For each fixed $t > 0$, the envelope is asymptotically sharp:

$$ntR_n(nt) \longrightarrow g(t), \qquad n \longrightarrow \infty. \tag{34}$$

The complete proof, including the singular populations $n = 1, 2, 3$ and an explicit strict correction for $n \geq 4$, is in Appendix D.4.

6.3 An explicit certificate for the endpoint-score residual

The quadratic bound B_n can also be supplied with a finite upper companion. The certificate concerns the exact stationary endpoint experiment. It identifies the residual with a conditional binomial score, then bounds that score by a polynomial with an exactly computable second moment. The conditional-score principle is classical [22], as is the Krawtchouk–Ehrenfest framework [10]; the required normalization and identities are derived here.

With the endpoint pair Y, the statistics A, D, the endpoint log-clock score S, the centered quadratics W_1, W_2 and the projection Π_2 of Section 6, define the endpoint-score residual

$$E_n(t) = \mathbb{E}[(S - \Pi_2 S)^2] = \lambda\{R_n(\lambda) - B_n(t)\}, \qquad \lambda = nt. \tag{35}$$

For the following formulas, put

$$\begin{aligned} s &= 1 - p, & r &= 1 - 2p = e^{-2t}, \\ w &= ps, & \nu &= nw, \\ h &= p^2 + s^2 = 1 - 2w, & \gamma &= \frac{tr}{w} = \frac{2t}{\sinh(2t)}, \\ d &= 2\nu + r^2, & e &= 2\nu + 3 - 14w, \end{aligned} \tag{36}$$

where $p = (1 - e^{-2t})/2$. Define the polynomial

$$\mathcal{V}_n(p) = 2n(n-1)w^2\left[h^2 + 4hr^2 + (n-2)wh + (11n-14)wr^2\right] \tag{37}$$

and the explicit certificate

$$T_n(t) = \frac{\gamma^2 r^2 \mathcal{V}_n(p)}{d^2 e^2}. \tag{38}$$

Theorem 6.3 (Finite residual and relative information bounds)**.** *Let R_n, B_n, E_n and T_n be as in* (18), (26), (35) *and* (38)*, with $\lambda = nt$. For every integer $n \geq 2$ and $0 < t \leq a_0 = (\log 2)/2$,*

$$0 \leq E_n(t) \leq T_n(t) \leq \frac{3}{2\lambda} + \frac{5}{8\lambda^2}. \tag{39}$$

Consequently,

$$B_n(t) \leq R_n(\lambda) \leq B_n(t) + \frac{3}{2\lambda^2} + \frac{5}{8\lambda^3}, \tag{40}$$

$$B_n(t) \leq R_n(\lambda) \leq B_n(t)\left(1 + \frac{3}{\lambda} + \frac{2}{\lambda^2}\right). \tag{41}$$

Only when $\lambda \geq 1$ may the upper terms be simplified to $E_n(t) \leq 17/(8\lambda)$ and $R_n(\lambda) \leq B_n(t) + 17/(8\lambda^2)$. The sharper relative interval width supplied by T_n is

$$\frac{T_n(t)}{\lambda B_n(t)} = \frac{r^2 \mathcal{V}_n(p)}{de^2 \nu h}. \tag{42}$$

For $n = 1$, Theorem 6.1 gives $R_1 = B_1$ exactly. The complete conditional-polynomial and global moment proof is in Appendix D.

7 Relative agreement and the sharp sampling window

The stationary particle representation permits control of the observed score at intensities where additive convergence does not determine relative error.

Theorem 7.1 (Relative crossover and sharp window)**.** *Let R_Ω and F be as in* (18) *and* (20)*, and put $t_0 = (\log 2)/2$. For every integer $\Omega \geq 16$ and real $1 \leq L \leq t_0\Omega$,*

$$\sup_{1\leq\lambda\leq L}\left|\frac{R_\Omega(\lambda)}{F(\lambda)}-1\right| \leq \frac{9L}{\Omega} + 2000\Omega^{-1/6}\log(e\Omega)^{1/3}. \tag{43}$$

For every deterministic sequence $L_\Omega \geq 1$,

$$\sup_{1\leq\lambda\leq L_\Omega}\left|\frac{R_\Omega(\lambda)}{F(\lambda)}-1\right| \longrightarrow 0 \quad\Longleftrightarrow\quad L_\Omega/\Omega \longrightarrow 0. \tag{44}$$

More precisely, for fixed $0 < t \leq t_0$, writing $p = (1-e^{-2t})/2$,

$$\frac{R_\Omega(\Omega t)}{F(\Omega t)} \longrightarrow \left(\frac{2t}{\sinh(2t)}\right)^2\left((1-p)^2+p^2\right) < 1. \tag{45}$$

The proof of Theorem 7.1 is in Appendix C. It joins the compact-intensity coupling to Corollary D.1; it does not use a local limit theorem.

8 Open questions

A natural wider scaling is $q_\theta^\Omega(x, x+\nu) = \theta\Omega a_\nu(x/\Omega)$ near a concentrated stationary equilibrium [21, 3]. A uniform Fisher-information approximation there is not proved here: for nonconstant total propensity one must couple both the endpoint and the compensated score, control its tails, and handle intensities beyond the local frozen approximation. Open problems include such an all-scales theorem, sharper quantitative intensity windows, relative coarse-scale bounds for general balanced marks, noisy or time-averaged observations, and the comparison with the binomial AR(1) literature mentioned in the introduction.

AI assistance disclosure

OpenAI GPT-5.6 Sol, accessed through ChatGPT and Codex, was used to develop and audit the proofs and symbolic computations, and for drafting. Anthropic Claude Opus 5.5, used through Claude Code, was used to condense and edit the paper, verify the Gram-determinant algebra symbolically, and derive Corollary 2.3 and Remark 3.3. The author reviewed and takes responsibility for every statement and proof. AI systems are not authors of this work, and the work has not been peer reviewed.

References

[1] Y. Aït-Sahalia and J. Jacod. "Fisher's Information for Discretely Sampled Lévy Processes". In: *Econometrica* 76.4 (2008), pp. 727–761. DOI: 10.1111/j.1468-0262.2008.00858.x.

[2] Y. Aït-Sahalia and P. A. Mykland. "The effects of random and discrete sampling when estimating continuous-time diffusions". In: *Econometrica* 71.2 (2003), pp. 483–549. DOI: 10.1111/1468-0262.t01-1-00416.

[3] D. F. Anderson, G. Craciun, and T. G. Kurtz. "Product-form stationary distributions for deficiency zero chemical reaction networks". In: *Bulletin of Mathematical Biology* 72.8 (2010), pp. 1947–1970. DOI: 10.1007/s11538-010-9517-4. arXiv: 0803.3042.

[4] G. Becker and G. Kersting. "Design problems for the pure birth process". In: *Advances in Applied Probability* 15.2 (1983), pp. 255–273. DOI: 10.2307/1426436.

[5] M. Bladt and M. Sørensen. "Statistical inference for discretely observed Markov jump processes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.3 (2005), pp. 395–410. DOI: 10.1111/j.1467-9868.2005.00508.x.

[6] A. J. Coca. "Efficient nonparametric inference for discretely observed compound Poisson processes". In: *Probability Theory and Related Fields* 170 (2018), pp. 475–523. DOI: 10.1007/s00440-017-0761-5. arXiv: 1512.08472.

[7] F. Comte and V. Genon-Catalot. "Nonparametric adaptive estimation for pure jump Lévy processes". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 46.3 (2010), pp. 595–617. DOI: 10.1214/09-AIHP323.

[8] Y. Cui and R. Lund. "Inference in binomial AR(1) models". In: *Statistics & Probability Letters* 80.23–24 (2010), pp. 1985–1990. DOI: 10.1016/j.spl.2010.09.003.

[9] I. Di Terlizzi and M. Baiesi. "Kinetic uncertainty relation". In: *Journal of Physics A: Mathematical and Theoretical* 52, 02LT03 (2019). DOI: 10.1088/1751-8121/aaee34. arXiv: 1809.06410.

[10] P. Diaconis and R. Griffiths. "Exchangeable pairs of Bernoulli random variables, Krawtchouck polynomials, and Ehrenfest urns". In: *Australian & New Zealand Journal of Statistics* 54.1 (2012), pp. 81–101. DOI: 10.1111/j.1467-842X.2012.00654.x.

[11] P. Diaconis and B. Sturmfels. "Algebraic algorithms for sampling from conditional distributions". In: *The Annals of Statistics* 26.1 (1998), pp. 363–397. DOI: 10.1214/aos/1030563990.

[12] C. Duval. "When is it no longer possible to estimate a compound Poisson process?" In: *Electronic Journal of Statistics* 8.1 (2014), pp. 274–301. DOI: 10.1214/14-EJS885.

[13] C. Duval and M. Hoffmann. "Statistical inference across time scales". In: *Electronic Journal of Statistics* 5 (2011), pp. 2004–2030. DOI: 10.1214/11-EJS660. arXiv: 1106.1031.

[14] Z. R. Fox and B. Munsky. "The finite state projection based Fisher information matrix approach to estimate information and optimize single-cell experiments". In: *PLOS Computational Biology* 15.1, e1006365 (2019). DOI: 10.1371/journal.pcbi.1006365.

[15] A. Hobolth and J. L. Jensen. "Statistical inference in evolutionary models of DNA sequences via the EM algorithm". In: *Statistical Applications in Genetics and Molecular Biology* 4.1, Article 18 (2005). DOI: 10.2202/1544-6115.1127.

[16] A. Hobolth and J. L. Jensen. "Summary statistics for endpoint-conditioned continuous-time Markov chains". In: *Journal of Applied Probability* 48.4 (2011), pp. 911–924. DOI: 10.1239/jap/1324046009.

[17] R. B. Israel, J. S. Rosenthal, and J. Z. Wei. "Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings". In: *Mathematical Finance* 11.2 (2001), pp. 245–265. DOI: 10.1111/1467-9965.00114.

[18] J. D. Kalbfleisch and J. F. Lawless. "The analysis of panel data under a Markov assumption". In: *Journal of the American Statistical Association* 80.392 (1985), pp. 863–871. DOI: 10.1080/01621459.1985.10478195.

[19] J. F. C. Kingman. "The imbedding problem for finite Markov chains". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 1.1 (1962), pp. 14–24. DOI: 10.1007/BF00531768.

[20] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf. "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models". In: *Proceedings of the National Academy of Sciences* 108.21 (2011), pp. 8645–8650. DOI: 10.1073/pnas.1015814108.

[21] T. G. Kurtz. "Solutions of ordinary differential equations as limits of pure jump Markov processes". In: *Journal of Applied Probability* 7.1 (1970), pp. 49–58. DOI: 10.2307/3212147.

[22] T. A. Louis. "Finding the observed information matrix when using the EM algorithm". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 44.2 (1982), pp. 226–233. DOI: 10.1111/j.2517-6161.1982.tb01203.x.

[23] K. Macieszczak. *Ultimate Kinetic Uncertainty Relation and Optimal Performance of Stochastic Clocks*. Preprint. 2024. DOI: 10.48550/arXiv.2407.09839. arXiv: 2407.09839.

[24] A.-L. Moor, A. Tjalma, M. Reinhardt, P. R. ten Wolde, and C. Zechner. "State- versus Reaction-Based Information Processing in Biochemical Networks". In: *Physical Review Letters* 136, 248401 (2026). DOI: 10.1103/nfk6-8x5s. arXiv: 2505.13373.

[25] D. E. Pagendam and P. K. Pollett. "Locally optimal designs for the simple death process". In: *Journal of Statistical Planning and Inference* 140.11 (2010), pp. 3096–3105. DOI: 10.1016/j.jspi.2010.04.017.

[26] Y. Pantazis, M. A. Katsoulakis, and D. G. Vlachos. "Parametric sensitivity analysis for biochemical reaction networks based on pathwise information theory". In: *BMC Bioinformatics* 14, 311 (2013). DOI: 10.1186/1471-2105-14-311.

[27] P. C. B. Phillips. "The problem of identification in finite parameter continuous time models". In: *Journal of Econometrics* 1.4 (1973), pp. 351–362. DOI: 10.1016/0304-4076(73)90021-3.

[28] D. Pollard. "A note on insufficiency and the preservation of Fisher information". In: *From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner*. Vol. 9. Institute of Mathematical Statistics Collections. Institute of Mathematical Statistics, 2013, pp. 266–275. DOI: 10.1214/12-IMSCOLL919. arXiv: 1107.3797. URL: https://arxiv.org/abs/1107.3797.

[29] K. Prech, G. T. Landi, F. Meier, N. Nurgalieva, P. P. Potts, R. Silva, and M. T. Mitchison. "Optimal Time Estimation and the Clock Uncertainty Relation for Stochastic Processes". In: *Physical Review X* 15, 031068 (2025). DOI: 10.1103/rpls-mp8z. arXiv: 2406.19450.

[30] M. Radaelli, G. T. Landi, K. Modi, and F. C. Binder. "Fisher information of correlated stochastic processes". In: *New Journal of Physics* 25, 053037 (2023). DOI: 10.1088/1367-2630/acd321.

[31] K. Takeuchi and A. Takemura. "On sum of 0–1 random variables. I. Univariate case". In: *Annals of the Institute of Statistical Mathematics* 39.1 (1987), pp. 85–102. DOI: 10.1007/BF02491451. URL: https://www.ism.ac.jp/editsec/aism/pdf/039_1_0085.pdf.

[32] P. Tataru and A. Hobolth. "Comparison of methods for calculating conditional expectations of sufficient statistics for continuous time Markov chains". In: *BMC Bioinformatics* 12, 465 (2011). DOI: 10.1186/1471-2105-12-465.

[33] C. H. Weiß and H.-Y. Kim. "Binomial AR(1) processes: moments, cumulants, and estimation". In: *Statistics* 47.3 (2013), pp. 494–510. DOI: 10.1080/02331888.2011.605893.

[34] C. H. Weiß and H.-Y. Kim. "Parameter estimation for binomial AR(1) models with applications in finance and industry". In: *Statistical Papers* 54 (2013), pp. 563–590. DOI: 10.1007/s00362-012-0449-y.

A Proof of the alias-depth expansion

Take the finite nonzero mark model of Theorem 2.2.

Proof. The joint mass of $(D, K) = (d, k)$ is $e^{-\lambda} b_{d,k} \lambda^k$. For nonnegative weights w_k, their unnormalized conditional variance is

$$\sum_k k^2 w_k - \frac{(\sum_k k w_k)^2}{\sum_k w_k} = \frac{\sum_{k<l} (k-l)^2 w_k w_l}{\sum_k w_k}. \tag{46}$$

First restrict the experiment to $K \le h$, without renormalizing its mass. By the definition of h, every nontrivial fibre in this restriction has just two lengths, $m_d < h$ and h. Its contribution to missing information is

$$e^{-\lambda} \frac{(h-m_d)^2 b_{d,m_d} b_{d,h} \lambda^{m_d+h}}{b_{d,m_d} \lambda^{m_d} + b_{d,h} \lambda^h} = (h-m_d)^2 b_{d,h} \lambda^h + O_p(\lambda^{h+1}).$$

There are finitely many such fibres. Adding the omitted mass can only increase the minimum squared prediction risk. Conversely, use the truncated conditional mean, which lies in $[0, h]$, on supported fibres, and use zero elsewhere. The additional risk is at most

$$\mathbb{E}[(K+h)^2 \mathbf{1}_{\{K>h\}}] = O(\lambda^{h+1}).$$

This proves the missing-information expansion and its remainder. Divide by λ. Nonzero marks exclude an alias of lengths zero and one. If no unequal lengths occur, K is determined by D. □

B Fixed-generator details and other time scales

B.1 Proof of the local classification

Take the fixed finite irreducible row-generator experiment and notation of Theorem 3.2.

Proof. Let $\mathcal{H}$ be the ordered-state skeleton of Proposition 3.1 and set $Z_t = \mathbb{E}[S_t \mid \mathcal{H}]$. The endpoint observation $Y = (X_0, X_t)$ is a function of $\mathcal{H}$, and the variance decomposition gives

$$\mathcal{M}_Q(t) = \mathbb{E}\,\mathrm{Var}(S_t \mid \mathcal{H}) + \mathbb{E}\,\mathrm{Var}(Z_t \mid Y). \tag{47}$$

For a fixed k-jump skeleton $i = i_0, \dots, i_k = j$, integration over its jump times gives mass

$$\pi_i \frac{t^k}{k!} \prod_{\ell=0}^{k-1} q_{i_\ell i_{\ell+1}} (1 + O_Q(t)),$$

and $Z_t = k + O_Q(t)$. Both expansions are uniform among skeletons with a fixed bounded length. Indeed the holding factor lies between $e^{-a_{\max} t}$ and 1, and $0 \le \int_0^t a(X_u)\, du \le a_{\max} t$.

Truncate at $K_t \le 2$. The only different-length skeletons with the same endpoints have lengths $0, 2$ on a return or $1, 2$ on a shortcut. The pairwise variance formula gives the coefficient (8). Same-length one-jump skeletons with fixed endpoints are unique, and a same-length two-jump discrepancy in Z_t is $O_Q(t)$, contributing only $O_Q(t^4)$. Poisson domination by $\mathrm{Pois}(a_{\max} t)$ bounds the squared risk on omitted skeletons by $O_Q(t^3)$. The first term of (47) is $O_Q(t^3)$ by Proposition 3.1; this proves (12).

If $A_2 = 0$, there is at most one supported length below three at each endpoint pair. Truncate instead at $K_t \le 3$. Every different-length pair now contributing at order t^3 has lengths $m_{ij} < 3$ and 3. Their masses sum to $\pi_i (B^3)_{ij} t^3/6$, and their score differences have leading value $3 - m_{ij}$. Thus the second term of (47) is $A_3 t^3 + O_Q(t^4)$. Same-length differences contribute $O_Q(t^4)$ or less; the omitted squared risk is $O_Q(t^4)$. Restoring the first term supplies $H t^3$ and proves (13). These risk bounds follow by using the retained conditional predictor on supported fibres and zero on unsupported fibres; its absolute value is bounded by $3 + a_{\max} t$.

Positivity of all stationary masses gives $A_2 > 0$ exactly when $h = 2$, and, conditional on $A_2 = 0$, $A_3 > 0$ exactly when $h = 3$. If rates are not constant, some allowed edge joins two states of different escape rate; otherwise irreducibility would make every a_i equal. Hence $H > 0$. Divide the missing information by $\bar{a} t$ to obtain the asserted powers. At constant escape the score is a centered Poisson count, so the previous graph alias theorem applies without a holding-time term. □

C Population-comparison proofs

C.1 Bounds for the limiting compound-Poisson experiment

The following is the proof of Lemma 5.1.

Proof. The strict upper inequality follows because $D = 0$ leaves both $K = 0$ and $K = 2$ possible. For the lower bound, $\mathrm{Cov}(K, D^2) = \lambda$ and $\mathrm{Var}(D^2) = \lambda + 2\lambda^2$; projection onto the centered variable D^2 gives the result and strict positivity.

For completeness we prove the coarse upper bound. Conditional on $|D| = d$, $K = d + 2Z$ with weights proportional to $(\lambda^2/4)^j/[j!(j+d)!]$. Put $a(j) = \lambda^2/[4(j+d+1)]$. Index shifting gives $\mathbb{E}[Z f(Z)] = \mathbb{E}[a(Z) f(Z+1)]$. Since a is decreasing,

$$\mathrm{Var}\, Z = \mathrm{Cov}(a(Z), Z) + \mathbb{E} a(Z) \le \mathbb{E} Z.$$

The same weights give $\mathbb{E}[Z(Z+d)] = \lambda^2/4$, so $\mathbb{E}[K^2 \mid |D| = d] = \lambda^2 + d^2$. With $\mu_d = \mathbb{E}[K \mid |D| = d]$, Jensen's inequality and $\mathrm{Var}\, K \le 2(\mu_d - d)$ imply

$$\sqrt{\lambda^2 + (d+1)^2} - 1 \le \mu_d \le \sqrt{\lambda^2 + d^2}.$$

Thus $s(d) = \mu_d - \lambda$ satisfies $-1 \le s(d) \le d^2/(2\lambda)$. Using $\mathbb{E} D^4 = \lambda + 3\lambda^2$ gives $\mathbb{E} s(D)^2 \le 7/4 + 1/(4\lambda) \le 2$ for $\lambda \ge 1$.

Finally, $|D|$ recovers K whenever $K \le 1$, whence $1 - F(\lambda) \le \lambda + 1 - e^{-\lambda} \le 2\lambda$. Continuity follows by truncating the complete score at a fixed K cutoff: on a compact intensity interval the Poisson second-moment tail is uniformly small, while each truncated finite sum is continuous. The coarse bound proves the remaining limit. □

C.2 Uniform additive comparison

We use the exact fixed-grid stationary experiment of Proposition 4.1. The two estimates below supply the proof of Theorem 5.2.

Lemma C.1 (Clipped-score comparison)**.** *Suppose $K \sim \mathrm{Pois}(\lambda)$ and two observations Y, Y_* are coupled using the same K, with $\mathbb{P}(Y \neq Y_*) \leq \lambda\delta$. If $J_Y = \mathbb{E}(\mathbb{E}[K - \lambda \mid Y])^2$, then for every $h > 0$,*

$$\frac{|J_Y - J_{Y_*}|}{\lambda} \leq 24h^2\delta + 4\sqrt{\frac{\mathbb{E}[(|K-\lambda| - h)_+^2]}{\lambda}}. \tag{48}$$

Proof. Write $S = K - \lambda$, let C_h clip a real number to $[-h, h]$, and put $S_h = C_h(S) - \mathbb{E}C_h(S)$. Then $|S_h| \leq 2h$ and $\operatorname{Var} S_h \leq \operatorname{Var} S = \lambda$. The last inequality follows by applying the Lipschitz property of C_h to two independent copies of S. Also $\|S - S_h\|_2^2 \leq \mathbb{E}[(|S| - h)_+^2]$. For either observation, the contraction property of conditional expectation gives

$$|\|\mathbb{E}[S \mid Y]\|_2^2 - \|\mathbb{E}[S_h \mid Y]\|_2^2| \leq 2\sqrt{\lambda}\,\|S - S_h\|_2.$$

For the clipped score use

$$\|\mathbb{E}[S_h \mid Y]\|_2^2 = \sup_{|f| \leq 2h} \mathbb{E}[2S_h f(Y) - f(Y)^2].$$

On the coupling event $Y = Y_*$ the objectives agree for each f; off it their absolute difference is at most $24h^2$. Take the supremum and restore S in both experiments. □

Lemma C.2 (Uniform Poisson score tail)**.** *For $0 < \lambda \leq L$, $x \geq 1$, and $h(L, x) = \sqrt{2eLx} + 2x$,*

$$\frac{\mathbb{E}[(K-\lambda)^2 \mathbf{1}_{\{|K-\lambda| > h(L,x)\}}]}{\lambda} \leq 2e(1 + 2x)e^{-x}. \tag{49}$$

Proof. For $S = K - \lambda$ an exponential tilt gives exactly

$$\frac{\mathbb{E}[S^2 e^{uS}]}{\lambda} = [e^u + \lambda(e^u - 1)^2] \exp\{\lambda(e^u - 1 - u)\}.$$

For $|u| \leq 1$, the right side is at most $(e + e^2 Lu^2) \exp(eLu^2/2)$. Apply the exponential bound to the positive and negative tails with $u = \min\{1, \sqrt{2x/(eL)}\}$. In either case $eLu^2/2 - uh(L, x) \leq -x$ and $e + e^2 Lu^2 \leq e(1 + 2x)$. Summing the two tails proves the claim. □

Proof of Theorem 5.2. Construct the actual embedded jump chain and a fair-sign walk with common uniform random variables at each event, and the same independent $K \sim \mathrm{Pois}(\lambda)$. Start the actual chain in π_Ω. At state n, the actual upward probability differs from $1/2$ by $|n/\Omega - 1/2|$. The embedded chain is stationary because its total escape rate is constant. A union bound at the first disagreement therefore yields

$$\mathbb{P}(D_{\text{actual}} \neq D_{\text{fair}}) \leq \lambda \mathbb{E}_{\pi_\Omega} |N/\Omega - 1/2| \leq \frac{\lambda}{2\sqrt{\Omega}}. \tag{50}$$

Include the same N_0 in both observations. It is independent of the fair walk, so its inclusion does not alter F. Lemmas C.1 and C.2 imply

$$\sup_{0 < \lambda \leq L} |R_\Omega(\lambda) - F(\lambda)| \leq \mathcal{E}_\Omega(L, x), \tag{51}$$

$$\mathcal{E}_\Omega(L, x) = \frac{12h(L, x)^2}{\sqrt{\Omega}} + 4\sqrt{2e(1 + 2x)e^{-x}}.$$

To treat larger intensities, subdivide a bin into m equal parts. Information cannot decrease when these extra observations are retained. Stationarity and orthogonality of the transition scores give

$$R_\Omega(\lambda) \leq R_\Omega(\lambda/m), \qquad m \in \mathbb{N},\ m \geq 1. \tag{52}$$

For $\lambda \geq L$, choose $m = \lfloor \lambda/L \rfloor$; then $L \leq \lambda/m \leq 2L$. Both $F(\lambda)$ and $R_\Omega(\lambda)$ are bounded above by $2/L + \mathcal{E}_\Omega(2L, x)$, using Lemma 5.1. Thus

$$\sup_{\lambda>0} |R_\Omega(\lambda) - F(\lambda)| \leq \frac{2}{L} + \mathcal{E}_\Omega(2L, x). \tag{53}$$

Take $x = \log(e\Omega)$ and $L = \Omega^{1/4}/\sqrt{x}$. The elementary inequality $h(L, x)^2 \leq 4eLx + 8x^2$ proves (22); the additional $x^2/\sqrt{\Omega}$ term is absorbed in its universal constant. Finally, $1 - \tau/T \leq \Delta/T$ and $F(\lambda) \leq 2/\lambda$, so the unobserved terminal remainder costs at most $2/(\Omega\theta T)$. □

C.3 The relative window from finite residual control

The following is the complete proof of Theorem 7.1.

Proof. Write $n = \Omega$, $\ell = \log(en)$, $\rho = n^{-1/6}\ell^{1/3}$, and $M = 1/\rho = n^{1/6}\ell^{-1/3}$. For $n \geq 16$, $M \geq 1$: the inequality $\sqrt{n} \geq \log(en)$ holds at 16 and its difference increases for $n > 4$. Also $M/n \leq n^{-5/6} < 1/4 < t_0$.

The coupling behind (51) uses the full observations $Y = (N_0, D_{\text{actual}})$ and $Y_* = (N_0, D_{\text{fair}})$, using the same initial count, Poisson event count and event uniforms. The first is equivalent to (N_0, N_Δ); N_0 is independent of the fair walk, so the information fraction of the second is F. For $1 \leq \lambda \leq M$, (51) with $x = \ell$ and Lemma 5.1 give

$$|R_n/F - 1| \leq 3M\,\mathcal{E}_n(M, \ell).$$

Since $h(M, \ell)^2 \leq 4eM\ell + 8\ell^2$ and $e^{-\ell} = (en)^{-1}$, the right side is at most

$$\frac{144eM^2\ell}{\sqrt{n}} + \frac{288M\ell^2}{\sqrt{n}} + \frac{12\sqrt{2}M\sqrt{1+2\ell}}{\sqrt{n}}. \tag{54}$$

The first term is $144e\rho$. Dividing the second by ρ gives $288\ell^{4/3}e^{-(\ell-1)/6}$, whose maximum for $\ell \geq 1$ is $4608e^{-7/6}$, attained at $\ell = 8$. As $1 + 2\ell \leq 3\ell$, the third term divided by ρ is at most $12\sqrt{6}\,n^{-1/6}\ell^{-1/6} \leq 12\sqrt{6}$. Hence

$$\sup_{1\leq\lambda\leq M} |R_n/F - 1| \leq (144e + 4608e^{-7/6} + 12\sqrt{6})\rho < 2000\rho. \tag{55}$$

For the last inequality, $e < 11/4$, $e^{7/6} > 3$ by the cubic Taylor lower bound, and $12\sqrt{6} < 30$ bound the constant by 1962.

For $M \leq \lambda \leq L$, Corollary D.1 instead gives

$$|R_n/F - 1| \leq 9L/n + 57/(8M) = 9L/n + (57/8)\rho.$$

Join the two intervals; if $L < M$, only the first is needed. This proves (43). The corollary uses only the finite residual certificate and the additive theorem, so this argument has no dependence on the relative theorem itself.

For fixed $0 < t \leq t_0$, (30) gives $ntB_n(t) \to g(t)$, and (39) gives $nt\{R_n(nt) - B_n(t)\} \to 0$. The one-sided compound-Poisson bounds in (73) imply $\lambda F(\lambda) \to 1/2$. Thus $R_n(nt)/F(nt) \to 2g(t)$, which equals the expression in (45). Both $2t/\sinh(2t)$ and $(1-p)^2 + p^2$ are strictly below one for $t > 0$.

If $L_n/n \to 0$, (43) applies eventually and its right side vanishes. Conversely, if this ratio does not vanish, choose $\delta > 0$ and a subsequence with $L_n/n \geq \delta$; fix $0 < t < \min\{\delta, t_0\}$. Then $nt \in [1, L_n]$ eventually on that subsequence, and (45) keeps the relative error away from zero. This proves (44). □

D Finite information certificates

D.1 All-population endpoint bounds

The following is the complete proof of Theorem 6.1.

Proof. The complete log-clock score is $C = K - \lambda$, where $K \sim \mathrm{Pois}(\lambda)$. For $Y = (N_0, N_\Delta)$, its conditional expectation $S = \mathbb{E}[C \mid Y]$ is the endpoint log-clock score. Hence

$$\mathbb{E}S^2 = \lambda R_n(\lambda), \qquad \lambda(1 - R_n(\lambda)) = \inf_f \mathbb{E}[(K - f(Y))^2]. \tag{56}$$

The infimum is over square-integrable endpoint predictors.

Quadratic projection. Put $p = (1 - e^{-2t})/2$, $u = n(1-p)$, $q = np$, and

$$A = N_0 + N_\Delta - n, \qquad D = N_\Delta - N_0.$$

In the independent stationary-particle representation, each particle contributes $(\pm 1, 0)$ with probabilities $(1-p)/2$, or $(0, \pm 1)$ with probabilities $p/2$. The n particle contributions are independent; successive observation intervals are not. Their second and fourth moments give

$$\mathbb{E}A^2 = u, \quad \mathbb{E}D^2 = q, \quad \mathbb{E}A^4 = 3u^2 + u(-2+3p), \quad \mathbb{E}D^4 = 3q^2 + q(1-3p),$$

and $\mathbb{E}A^2D^2 = n(n-1)p(1-p)$. For example, the last identity follows by expanding the sums and using that a single particle cannot contribute to both axes. Therefore the centered quadratics $W_1 = A^2 - u$ and $W_2 = D^2 - q$ have covariance matrix

$$\Gamma = \begin{pmatrix} 2u^2 + u(-2+3p) & -np(1-p) \\ -np(1-p) & 2q^2 + q(1-3p) \end{pmatrix}. \tag{57}$$

Freeze the centering constants u, q at the parameter value being evaluated. The score identity differentiates the raw moments, at fixed physical spacing Δ, and gives

$$v := \begin{pmatrix} \mathbb{E}[SW_1] \\ \mathbb{E}[SW_2] \end{pmatrix} = nte^{-2t} \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{58}$$

Here the log-clock derivative is $t\partial_t$, and $p'(t) = e^{-2t}$. For $n \geq 2$,

$$\det \Gamma = 2n^2 p(1-p)(n-1)\{1 + 2(n-2)p(1-p)\} > 0.$$

The orthogonal projection of S onto $\mathrm{span}\{W_1, W_2\}$ has squared norm $v^\top \Gamma^{-1} v$. Direct inversion gives

$$\frac{v^\top \Gamma^{-1} v}{nt} = H_0(t) \frac{(1-p)^2 + p^2}{1 + 2(n-2)p(1-p)} = B_n(t). \tag{59}$$

Thus $R_n(nt) \geq B_n(t)$. For $n = 1$, this covariance matrix is singular; the direct parity calculation (24) instead yields $R_1(t) = H_0(t) = B_1(t)$.

Labelled endpoints. Revealing every particle's initial and final state refines the aggregate observation. Each parity-flip indicator is Bernoulli with probability p, and its log-clock information is $(tp'(t))^2/[p(1-p)]$. The labelled experiment therefore has information $n(tp'(t))^2/[p(1-p)]$. Dividing by nt and applying data processing gives

$$R_n(nt) \leq \frac{te^{-4t}}{p(1-p)} = H_0(t).$$

These indicators count particles with different endpoint states, not the actual number of reactions K.

Prediction by the endpoint displacement. The predictor $f(Y) = |D|$ recovers K when $K \leq 1$. At $K = 2$ its squared error is four exactly on a return to the initial aggregate count. The embedded chain selects a uniformly random particle at each event. With probability $1/n$ the same particle is selected twice; otherwise a return requires opposite initial states, with probability $1/2$. Thus the stationary two-step return probability is

$$\bar{r}_n = \frac{1}{n} + \frac{1}{2}\left(1 - \frac{1}{n}\right) = \frac{c_n}{2}. \tag{60}$$

For $K \geq 3$, $|D| \leq K$ gives $(K - |D|)^2 \leq K^2$. Consequently

$$\lambda(1 - R_n(\lambda)) \leq c_n \lambda^2 e^{-\lambda} + \mathbb{E}[K^2 \mathbf{1}_{\{K \geq 3\}}].$$

Since $\mathbb{E}K^2 = \lambda + \lambda^2$, removal of the $K = 1, 2$ terms gives

$$\frac{\mathbb{E}[K^2 \mathbf{1}_{\{K \geq 3\}}]}{\lambda} = 1 + \lambda - (1 + 2\lambda)e^{-\lambda}.$$

This proves the upper loss bound $1 - R_n \leq U_n$.

Retaining only zero jumps and two-jump returns. Deleting other histories decreases every squared prediction risk, and therefore decreases its infimum in (56). Given $N_0 = i$, write $r_i = [2i(n - i) + n]/n^2$ for the embedded two-step return probability. The retained diagonal weights are

$$w_0 = e^{-\lambda}, \qquad w_2 = e^{-\lambda} \lambda^2 r_i / 2.$$

The minimum unnormalized squared risk for these two masses is

$$\frac{4 w_0 w_2}{w_0 + w_2} = \frac{2\lambda^2 e^{-\lambda} r_i}{1 + \lambda^2 r_i / 2}.$$

Average over the stationary initial count. Since $0 \leq r_i \leq 1$ and $\mathbb{E} r_i = \bar{r}_n = c_n/2$, bounding the denominator above by $1 + \lambda^2/2$ yields

$$\lambda(1 - R_n(\lambda)) \geq \frac{c_n \lambda^2 e^{-\lambda}}{1 + \lambda^2/2}.$$

This proves the lower loss bound and completes (28). After cancellation of the common Poisson weight, the rational denominator $1 + \lambda^2 r_i/2$ is at least one. No rare endpoint probability needs to be bounded below in this argument. □

D.2 The exact residual and its global norm

Take the model and parameters of Theorem 6.3. The following is its complete proof.

Proof. **Exact conditional polynomials and parity.** Let M be the number of particles with different initial and terminal states. Then $M \sim \mathrm{Bin}(n, p)$, and, conditional on M,

$$A = \sum_{j=1}^{n-M} \varepsilon_j, \qquad D = \sum_{j=1}^{M} \eta_j,$$

where the two sets of signs are independent symmetric Rademacher variables. The correlated-Bernoulli construction is classical [10, Theorem 1 and Section 6]. This is an exact endpoint representation; M is not the actual jump count K. With $X = M - np$, differentiation at fixed physical grid gives

$$S = \gamma \mathbb{E}[X \mid Y]. \tag{61}$$

Indeed the complete score for the endpoint representation in coordinate p is X/w, and the log-clock derivative is $tr\partial_p$.

Write $M = \sum_{i=1}^{n} B_i$ with independent Bernoulli indicators B_i. The subset polynomials below normalize the classical binomial Krawtchouk likelihood-derivative polynomials [31, Eq. (3.2)]. We derive their norms and conditional endpoint identities in the present notation. For $1 \leq j \leq n$, define

$$H_j = j! \sum_{|I|=j} \prod_{i \in I} (B_i - p),$$

and set $H_j = 0$ for $j > n$. Expanding gives

$$\begin{aligned} H_1 &= X, & H_2 &= X^2 - rX - \nu, \\ H_3 &= X^3 - 3rX^2 + \{2r^2 - (3n-2)w\}X + 2nwr. \end{aligned} \tag{62}$$

The last polynomial is zero on the binomial support when $n = 2$, consistent with the empty-subset definition. Independence and centering show

$$\mathbb{E}[H_j H_k] = 0 \ (j \neq k), \qquad \mathbb{E}H_j^2 = j!(n)_j w^j. \tag{63}$$

For distinct subsets some centered factor occurs once, so its expectation is zero; identical subsets give w^j. These observations prove both identities, where $(n)_j = n(n-1)\cdots(n-j+1)$ is the falling factorial, with $(n)_j = 0$ for $j > n$.

Set $\widehat{H}_j = \mathbb{E}[H_j \mid Y]$ and $L = s^2 W_2 - p^2 W_1$. For fixed admissible endpoints $(A, D) = (a, z)$, the conditional weights of $M = m$ are proportional to

$$\frac{s^{n-m} p^m}{\left(\frac{n-m+a}{2}\right)! \left(\frac{n-m-a}{2}\right)! \left(\frac{m+z}{2}\right)! \left(\frac{m-z}{2}\right)!}.$$

Only nonnegative factorial arguments are admitted: $|D| \leq m \leq n - |A|$. Feasible values of m have fixed parity and are spaced by two. Their successive-weight identity is

$$s^2\{(m+2)^2 - z^2\} \Pr(M = m+2 \mid Y) = p^2\{(n-m)^2 - a^2\} \Pr(M = m \mid Y).$$

The omitted endpoint terms vanish at the feasible lower and upper bounds. Summing along this parity class therefore gives

$$p^2 \mathbb{E}[(n-M)^2 - A^2 \mid Y] = s^2 \mathbb{E}[M^2 - D^2 \mid Y].$$

Substitution of $M = np + X$ and (62) yields

$$d\widehat{H}_1 = L - r\widehat{H}_2. \tag{64}$$

Now $\mathbb{E}[W_1 \mid M] = -X$ and $\mathbb{E}[W_2 \mid M] = X$. Since H_2 is orthogonal to X, $\mathbb{E}[\widehat{H}_2 W_1] = \mathbb{E}[\widehat{H}_2 W_2] = 0$. Thus (64) is an orthogonal decomposition:

$$\begin{aligned} \Pi_2 S &= \frac{\gamma L}{d}, & S - \Pi_2 S &= -\frac{\gamma r}{d}\widehat{H}_2, \\ E_n(t) &= \frac{\gamma^2 r^2}{d^2}\mathbb{E}\widehat{H}_2^2. \end{aligned} \tag{65}$$

A differentiated likelihood certificate. Let $f_y(p)$ denote the positive joint mass of an admissible endpoint pair. Its finite Bernoulli mixture, whose conditional signs have no p dependence, gives

$$\widehat{H}_j = w^j \frac{f_y^{(j)}}{f_y}, \qquad j = 1, 2, 3.$$

To see this directly, differentiate the product of n linear Bernoulli probabilities. The jth derivative sums ordered choices of j distinct factors and hence gives precisely the subset polynomial H_j / w^j. For $n = 2$, both the third derivative and H_3 are zero. Equation (64) becomes

$$rw^2 f_y'' + dw f_y' - L f_y = 0.$$

Differentiate this identity at fixed endpoint pair, and put $\beta = (dw)' = r\{1 + 4(n-2)w\}$. Since $dw + (rw^2)' = we$, the result is

$$\begin{aligned} e\widehat{H}_2 &= (L - \beta)\widehat{H}_1 + wL' - r\widehat{H}_3, \\ L' &= -2sD^2 - 2pA^2 - n(r^2 - 2w). \end{aligned} \tag{66}$$

The projection statistics in the previous paragraph are evaluated at the given parameter with their centers held fixed. Here $L(p)$ is a moving coefficient of an identity valid throughout the parameter family; its derivative, including that of its centers, must be kept.

Define the latent polynomial $Z = (L - \beta)X + wL' - rH_3$. Then (66) says $\mathbb{E}[Z \mid Y] = e\widehat{H}_2$. Conditional Jensen gives the global estimate

$$\mathbb{E}\widehat{H}_2^2 \le \frac{\mathbb{E}Z^2}{e^2}. \tag{67}$$

Exact second moment. Conditional on M, $\mathbb{E}[L \mid M] = hX$ and $\mathbb{E}[L' \mid M] = -nh - 2rX$. Consequently

$$\mathbb{E}[Z \mid M] = hH_2 - 4(n-1)wrX - rH_3.$$

Its three terms are orthogonal by (63). For a sum of k independent Rademacher signs, the variance of its square is $2k(k-1)$. The coefficients of D^2 and A^2 in Z are respectively $s^2(X - 2p)$ and $-p^2(X + 2s)$. Conditional independence therefore gives

$$\begin{aligned}\operatorname{Var}(Z \mid M) = {} & 2s^4(X-2p)^2 M(M-1) \\ & + 2p^4(X+2s)^2(n-M)(n-M-1).\end{aligned}$$

The binomial factorial weighting identity $\mathbb{E}[(M)_2 g(M)] = (n)_2 p^2 \mathbb{E}[g(M'+2)]$, with $M' \sim \operatorname{Bin}(n-2, p)$, yields

$$\mathbb{E}[M(M-1)(X-2p)^2] = n(n-1)p^2\{(n-2)w + 4r^2\}.$$

Selecting two zero indicators gives the horizontal counterpart with s^2 in place of p^2. Thus

$$\mathbb{E}\operatorname{Var}(Z \mid M) = 2n(n-1)w^2 h\{(n-2)w + 4r^2\}.$$

The squared norm of the conditional mean is

$$\mathbb{E}\big(\mathbb{E}[Z \mid M]\big)^2 = 2n(n-1)w^2\{h^2 + (11n-14)wr^2\}.$$

Variance decomposition proves $\mathbb{E}Z^2 = \mathcal{V}_n(p)$, including $n = 2$. Together with (65) and (67), this proves $E_n(t) \le T_n(t)$. This calculation includes every endpoint cell. It requires neither a Gaussian approximation nor a lower bound on a rare endpoint probability.

Uniform constants and relative width. On $0 < t \le a_0$, $p \le 1/4$, $w \le 3/16$, $h \ge 5/8$, and $0 < \gamma \le 1$. Also $d \ge 2\nu$ and $e \ge 2\nu + 3/8 > 2\nu$. For $n \ge 2$, all terms in the bracket defining $\mathcal{V}_n$ are nonnegative and their sum is at most $5 + 12\nu$. Hence

$$\mathcal{V}_n \le 2\nu^2(5 + 12\nu), \qquad T_n(t) \le \gamma^2 r^2 \left(\frac{3}{2\nu} + \frac{5}{8\nu^2}\right).$$

Using $\nu = \lambda r/\gamma$ gives

$$T_n(t) \le \frac{3\gamma^3 r}{2\lambda} + \frac{5\gamma^4}{8\lambda^2} \le \frac{3}{2\lambda} + \frac{5}{8\lambda^2}.$$

For the relative interval, the quadratic information formula is $\lambda B_n(t) = \gamma^2 \nu h/d$. Thus

$$\frac{T_n(t)}{\lambda B_n(t)} \le \frac{r^2}{h}\left(\frac{3}{\nu} + \frac{5}{4\nu^2}\right) = \frac{3\gamma r}{h\lambda} + \frac{5\gamma^2}{4h\lambda^2} \le \frac{3}{\lambda} + \frac{2}{\lambda^2},$$

where $r \le h$, $h \ge 5/8$, and $\gamma \le 1$. Equation (35) proves both information bounds; the simplifications with constant $17/8$ use $\lambda \ge 1$. □

D.3 A finite comparison with the compound-Poisson experiment

Corollary D.1 (A finite compound-Poisson comparison)**.** *For $n \geq 2$, $\lambda \geq 1$, and $t = \lambda/n \leq a_0$,*

$$\left|\lambda R_n(\lambda) - \frac{1}{2}\right| \leq 3t + \frac{17}{8\lambda}. \tag{68}$$

For every $\lambda \geq 1$,

$$\left|\lambda F(\lambda) - \frac{1}{2}\right| \leq \frac{17}{8\lambda}. \tag{69}$$

Consequently, on the preceding n, t, λ domain,

$$\left|\frac{R_n(\lambda)}{F(\lambda)} - 1\right| \leq \frac{9\lambda}{n} + \frac{57}{8\lambda}. \tag{70}$$

In particular, this relative error is at most $1/10$ throughout the nonempty continuous region

$$n \geq 25650, \qquad 285/2 \leq \lambda \leq n/180. \tag{71}$$

The following is the proof of Corollary D.1.

Proof. Let $x = n(e^{4t} - 1)$ and $a = x/(x + 2)$. Then (30) gives $\lambda B_n(t) = g(t)a$, where $g(t) = \gamma^2 h/2$. Since $\gamma = 4tr/(1 - r^2) \geq r$ by $1 - e^{-4t} \leq 4t$, and $h = (1 + r^2)/2$,

$$0 \leq \frac{1}{2} - g(t) \leq \frac{(1 - \gamma^2) + (1 - h)}{2} \leq 3t.$$

Also $x \geq 4\lambda$ and $g(t) \leq 1/2$, so

$$0 \leq \frac{1}{2} - \lambda B_n(t) \leq 3t + \frac{1}{2(2\lambda + 1)} \leq 3t + \frac{1}{4\lambda}.$$

Use the sign and bound $0 \leq E_n(t) = \lambda(R_n - B_n) \leq 17/(8\lambda)$ to obtain

$$\frac{1}{2} - 3t - \frac{1}{4\lambda} \leq \lambda B_n(t) \leq \lambda R_n(\lambda) \leq \frac{1}{2} + \frac{17}{8\lambda}. \tag{72}$$

This proves (68).

For each fixed $\lambda \geq 1$, let $n \to \infty$ with $t = \lambda/n$; the preceding domain is then satisfied eventually. The already proved uniform additive convergence in Theorem 5.2 gives $R_n(\lambda) \to F(\lambda)$. Passing to this limit in (72) gives

$$\frac{1}{2} - \frac{1}{4\lambda} \leq \lambda F(\lambda) \leq \frac{1}{2} + \frac{17}{8\lambda}, \tag{73}$$

which also proves (69). This is a limit of the information fractions already defined with fixed physical designs, not a derivative of a parameter-dependent grid.

Subtracting opposite endpoints of (72) and (73) in both directions gives

$$|\lambda\{R_n(\lambda) - F(\lambda)\}| \leq 3t + \frac{19}{8\lambda}.$$

For $\lambda \geq 1$, Lemma 5.1 gives $\lambda F(\lambda) \geq \lambda/(1 + 2\lambda) \geq 1/3$; division proves (70). In (71), $t \leq 1/180 < a_0$ and

$$9t + \frac{57}{8\lambda} \leq \frac{9}{180} + \frac{57}{8 \cdot (285/2)} = \frac{1}{10}.$$

The condition $n \geq 25650 = 180 \cdot (285/2)$ makes the interval nonempty. □

D.4 Posterior controls for the finite Gaussian envelope

The following is the complete proof of Theorem 6.2. Take $s, r, w, h, \nu, d, \gamma$ from (36). First suppose $n \geq 2$ and $t > 0$ is finite, so $0 < p < 1/2$, $0 < w < 1/4$, and $r > 0$. Let $M \sim \mathrm{Bin}(n, p)$ be the parity-change particle count, and put $X = M - np$ and $H_2 = X^2 - rX - \nu$. This M is not the actual reaction count K. Write $\widehat{H}_2 = \mathbb{E}[H_2 \mid A, D]$.

Proof. **A posterior recurrence with no boundary remainder.** At an admissible endpoint $(A, D) = (a, z)$, the posterior weights q_m are proportional to

$$\frac{p^m s^{n-m}}{\left(\frac{n-m+a}{2}\right)!\left(\frac{n-m-a}{2}\right)!\left(\frac{m+z}{2}\right)!\left(\frac{m-z}{2}\right)!}.$$

The complete feasible parity class has $m \geq |z|$ and $n - m \geq |a|$. Its consecutive terms obey

$$p^2\{(n-m)^2 - a^2\}q_m = s^2\{(m+2)^2 - z^2\}q_{m+2}.$$

Define the random polynomials

$$\alpha = p^2\{(n-M)^2 - A^2\}, \qquad \beta = s^2(M^2 - D^2).$$

At the highest posterior support point $\alpha = 0$, and at the lowest point $\beta = 0$. Summing the recurrence gives, for every polynomial f,

$$\mathbb{E}[\alpha f(M+2) \mid A, D] = \mathbb{E}[\beta f(M) \mid A, D]. \tag{74}$$

This includes singleton supports. Hence

$$T_f = (\alpha - \beta)f(M) + \alpha\{f(M+2) - f(M)\}$$

has conditional mean zero. Use the three polynomials

$$f_0 = 1, \qquad f_1 = X - 1, \qquad f_2 = H_2 - 2X + 2 + r, \qquad T = (T_{f_0}, T_{f_1}, T_{f_2})^\top. \tag{75}$$

Their coefficients are evaluated at the fixed parameter of interest.

The exact Gram calculation. Put $\kappa = \mathbb{E}H_2^2 = 2n(n-1)w^2$, and define $G = \mathbb{E}[TT^\top]/\kappa$ and $v = \mathbb{E}[H_2 T]/\kappa$. Direct expansion gives

$$\begin{aligned}
v_0 &= -r, \qquad v_1 = -2nw + 8w - 1, & v_2 &= -r(4nw - 10w + 1), \\
G_{00} &= 2(nw - 4w + 1), & G_{01} &= 2r(2nw - 6w + 1), \\
G_{02} &= -2(10nw^2 - 3nw - 24w^2 + 10w - 1), \\
G_{11} &= 2(2n^2w^2 - 27nw^2 + 5nw + 54w^2 - 16w + 1), \\
G_{12} &= 2r(7n^2w^2 - 57nw^2 + 8nw + 90w^2 - 20w + 1), \\
G_{22} &= 2(6n^3w^3 - 154n^2w^3 + 27n^2w^2 + 772nw^3 \\
&\qquad - 215nw^2 + 13nw - 984w^3 + 334w^2 - 32w + 1).
\end{aligned} \tag{76, 77}$$

The matrix is symmetric. For completeness, all entries follow by substituting the controls in (75), conditioning first on M, and using

$$\begin{aligned}
\mathbb{E}[A^2 \mid M] &= n - M, & \mathbb{E}[A^4 \mid M] &= 3(n-M)^2 - 2(n-M), \\
\mathbb{E}[D^2 \mid M] &= M, & \mathbb{E}[D^4 \mid M] &= 3M^2 - 2M, \\
\mathbb{E}[A^2 D^2 \mid M] &= (n-M)M.
\end{aligned}$$

The signs on the two axes are conditionally independent. The remaining binomial moments, through degree eight, are obtained from

$$\mathbb{E}M^j = \sum_{\ell=0}^{j} \left\{ {j \atop \ell} \right\} (n)_\ell p^\ell.$$

This identity expands powers in the falling-factorial basis and counts ordered distinct Bernoulli coordinates. As one entry check, $T_{f_0} = L - dX - rH_2$ with the L of the residual proof. Orthogonality gives $\mathbb{E}[H_2 T_{f_0}] = -r\kappa$, and the conditional fourth moments give $\mathbb{E}T_{f_0}^2 = 2\kappa(nw + r^2)$, the displayed v_0, G_{00}.

A positive rational majorant for $n \geq 4$. Define

$$\begin{aligned} Q(n,w) &= 12n^3w^3 - 142n^2w^3 + 23n^2w^2 + 624nw^3 - 240nw^2 + 24nw \\ &\quad - 1152w^3 + 784w^2 - 180w + 14, \end{aligned} \tag{78}$$

$$P(n,w) = -3w^2hn^2 + w(76w^2 - 28w + 4)n + 576w^3 - 536w^2 + 152w - 14. \tag{79}$$

Expansion of the determinant gives

$$\det G = 8w^3(n-3)(n-2)^2 Q(n,w).$$

For $k = n - 4 \geq 0$ and $v_* = 1/4 - w \geq 0$,

$$\begin{aligned} Q(n,w) &= 12w^3k^3 + (2w^3 + 23w^2)k^2 + w(64v_*^2 + 24v_* + 14)k \\ &\quad + 160v_*^3 + 72v_*^2 + 18v_* + 5/2 > 0. \end{aligned}$$

Since G is a Gram matrix, it is positive semidefinite; its nonzero determinant makes it positive definite. Its inverse is therefore valid. For every deterministic vector c, conditional Jensen and $\mathbb{E}[T \mid A, D] = 0$ give

$$\mathbb{E}\widehat{H}_2^2 \leq \mathbb{E}(H_2 - c^\top T)^2.$$

Minimize the right side at $c = G^{-1}v$. Substitution of (76)–(77), or the adjugate identity, gives

$$\mathbb{E}\widehat{H}_2^2 \leq C_n(p) := \kappa(1 - v^\top G^{-1} v) = \frac{n(n-1)w^2 d\{3nwh + 200w^2 - 110w + 14\}}{Q(n,w)}. \tag{80}$$

The exact gap is

$$\frac{hd}{2} - C_n(p) = -\frac{d^2 P(n,w)}{2Q(n,w)}. \tag{81}$$

As a quadratic in n, P has strictly negative leading coefficient $-3w^2h$ and discriminant $-8w^2Z(w)$, where

$$Z(w) = 1006w^4 - 1940w^3 + 1086w^2 - 242w + 19.$$

On $0 < w \leq 3/16$, put $u = 3/16 - w \geq 0$. Then

$$Z(3/16 - u) = 1006u^4 + \frac{2371}{2}u^3 + \frac{13245}{64}u^2 + \frac{6571}{512}u + \frac{8519}{32768} > 0.$$

For $0 < w \leq 3/16$, this proves $P(n,w) < 0$ for every real n. For $3/16 \leq w < 1/4$, put $v_* = 1/4 - w \in (0, 1/16]$. The quadratic P is strictly concave in n, and direct substitution gives

$$\partial_n P(4,w) = w(124v_*^2 - 10v_* - 5/4), \qquad P(4,w) = -976v_*^3 + 36v_*^2 - 3v_* - 1/4.$$

Since $124v_*^2 \leq (31/4)v_*$ and $36v_*^2 \leq (9/4)v_*$,

$$\partial_n P(4,w) \leq w\{-9v_*/4 - 5/4\} < 0, \qquad P(4,w) \leq -976v_*^3 - 3v_*/4 - 1/4 < 0.$$

Concavity makes the derivative decrease for $n \geq 4$, so $P(n,w) < 0$ throughout the remaining parameter interval as well. All denominators in (81) are positive, proving $\mathbb{E}\widehat{H}_2^2 < hd/2$ for $n \geq 4$.

The singular populations $n = 2, 3$. No Gram inverse is used for these populations. The full joint weights are

$$b_m(a,z) = \frac{n! 2^{-n} p^m s^{n-m}}{\left(\frac{n-m+a}{2}\right)! \left(\frac{n-m-a}{2}\right)! \left(\frac{m+z}{2}\right)! \left(\frac{m-z}{2}\right)!},$$

with only nonnegative integer factorial arguments admitted. Evaluating the finite sum

$$\mathbb{E}\widehat{H}_2^2 = \sum_{a,z} \frac{\{\sum_m b_m(a,z) H_2(m)\}^2}{\sum_m b_m(a,z)}$$

over every feasible cell gives

$$n = 2: \quad \mathbb{E}\widehat{H}_2^2 = 2w^2/h, \qquad \frac{hd}{2} - \mathbb{E}\widehat{H}_2^2 = \frac{r^2}{2h} > 0,$$

$$n = 3: \quad \mathbb{E}\widehat{H}_2^2 = \frac{6w^2(1+2w)(2-7w)}{2-8w+9w^2},$$

$$\frac{hd}{2} - \mathbb{E}\widehat{H}_2^2 = \frac{(1+2w)^2(33w^2-16w+2)}{2(2-8w+9w^2)} > 0.$$

At $n = 2$, $d = 1$. At $n = 3$, the denominator is decreasing on $[0, 1/4]$ and has minimum $9/16 > 0$. The numerator quadratic has positive leading coefficient and discriminant -8, proving its sign.

Translation to clock information. The conditional-score and parity identities in the proof of (65) are exact for $0 < p < 1/2$; their algebra does not use the small-time restriction on the residual remainder certificate. In this full parameter range, $d = 1 + 2(n-2)w$ and the quadratic covariance matrix satisfies

$$\det \Gamma = 2n^2 w(n-1)d > 0, \qquad n \geq 2.$$

Thus its projection is well-defined, with exact squared norm $\lambda B_n(t) = \gamma^2 \nu h/d$. The parity identity $d\widehat{H}_1 = L - r\widehat{H}_2$ and orthogonality of $\widehat{H}_2$ to W_1, W_2 then give

$$\lambda R_n(\lambda) = \gamma^2 \left\{ \frac{\nu h}{d} + \frac{r^2}{d^2} \mathbb{E}\widehat{H}_2^2 \right\} < \gamma^2 \left\{ \frac{\nu h}{d} + \frac{r^2 h}{2d} \right\} = \frac{\gamma^2 h}{2} = g(t).$$

This is the exact endpoint experiment, without a Gaussian approximation or a removed set of rare cells. For $n \geq 4$, the more precise certificate (80) and its gap yield

$$\lambda R_n(\lambda) \leq g(t) + \frac{\gamma^2 r^2 P(n,w)}{2Q(n,w)} < g(t).$$

For $n = 1$, the exact parity score is $S = \gamma(M - p)$, with $M \sim \text{Bernoulli}(p)$. Thus $\lambda R_1 = \gamma^2 w$ and $g(t) - \lambda R_1 = \gamma^2 r^2/2 > 0$ for finite $t > 0$. This uses no Gram inverse and keeps $R_1 = B_1$. Finally, put $x = n(e^{4t} - 1) \geq 4\lambda$. The exact quadratic formula (30) gives $\lambda B_n(t) = g(t)x/(x+2)$, hence $1 \leq R_n/B_n < 1 + 2/x \leq 1 + 1/(2\lambda)$. The ten-percent claim follows when $\lambda \geq 5$. At every fixed $t > 0$, $x = n(e^{4t} - 1) \to \infty$. The bounds $g(t)x/(x+2) \leq ntR_n(nt) < g(t)$ therefore prove (34) by squeezing, without a differentiated local limit theorem. □