

Exact Likelihood-Coin Poisson Sampling for Bayesian Inverse Problems with Sharp Complexity Bounds

Zhiliang Deng

dengzh1@uestc.edu.cn

University of Electronic Science and Technology of China

Xiaomei Yang

Southwest Jiaotong University

Research Article

Keywords: Bayesian inverse problems, exact sampling, Poisson thinning, Bernoulli factory, small-ball asymptotics, Feynman–Kac formula, function-valued parameter

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11211426/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Exact Likelihood-Coin Poisson Sampling for Bayesian Inverse Problems with Sharp Complexity Bounds

Zhiliang Deng*

Xiaomei Yang[†]

Abstract

We develop an exact Bayesian computational framework for inverse problems in which selected bounded forward observables can be queried through Bernoulli events. A Bernstein–Poisson construction converts these forward events directly into scaled Gaussian likelihood coins, avoiding an intermediate numerical evaluation of the forward map or likelihood. Coupling the likelihood coins with Poisson thinning yields posterior atoms that are iid conditional on their number. For independent Gaussian observations, we derive an exact mean-work identity for the implemented early-stopped sampler and sharp small-noise complexity laws governed by the local prior-predictive mass near the exact-fit set. This identifies a computational phase transition controlled by an effective predictive dimension. A factorial-moment construction extends the likelihood-coin mechanism to correlated Gaussian errors. The posterior algorithm is model-independent once Bernoulli access is available; a Feynman–Kac construction with Poisson killing and lazy random-series evaluation provides a nontrivial continuum realization for a function-valued elliptic inverse problem. Numerical experiments validate the factories and sharp work regimes, while a matched-accuracy comparison with finite-difference rejection illustrates when removal of deterministic discretization bias can offset the cost of exact stochastic simulation.

Keywords: Bayesian inverse problems; exact sampling; Poisson thinning; Bernoulli factory; small-ball asymptotics; Feynman–Kac formula; function-valued parameter.

1 Introduction

Bayesian computation becomes difficult when the likelihood is unavailable or expensive to evaluate. In PDE-constrained inverse problems this difficulty is especially pronounced because each likelihood evaluation may require a numerical forward solve, and posterior sampling can require many such evaluations [1, 2]. A broad computational literature addresses this bottleneck through deterministic approximation, multilevel sampling, reduced models, and surrogates [3–5]. Probabilistic numerical and randomized-forward approaches provide another route by representing numerical error probabilistically and propagating it through Bayesian inference [6–8]. These methods motivate a complementary computational question: when stochastic simulation of a forward observable is available exactly, can it be used directly to construct the posterior decision rather than first being converted into a numerical forward or likelihood approximation?

Exact Bayesian computation provides the closest methodological context. Pseudo-marginal methods use unbiased stochastic likelihood information while retaining a Markov-chain architecture [9]. Bernoulli factories transform events with unknown success probability into new exact Bernoulli decisions [10–13], and their use in exact Bayesian computation is well established. Flegal and Herbei [14] combine a Bernoulli factory with a regenerative rejection construction to obtain exact samples from an intractable target. Vats et al. [15] use Bernoulli factories to implement

*School of Mathematical Sciences, University of Electronic Science and Technology of China. Email: dengzhl@uestc.edu.cn

[†]School of Mathematics, Southwest Jiaotong University. Email: yangxiaomath@swjtu.edu.cn

exact MCMC acceptance decisions for intractable likelihoods, while recent work develops scalable Bernoulli-factory MCMC constructions as the data size grows [16]. These methods establish that Bernoulli factories can support exact Bayesian computation; the distinction in the present work lies in the computational interface at which the factory is used.

Here the primitive Bernoulli variables encode bounded *forward observables*. For a candidate parameter q , we assume access to independent $p_j(q)$ -coins, where $p_j(q)$ is a selected forward quantity. We use these coins directly to construct, up to a known deterministic scale, the Gaussian likelihood factor associated with that candidate. Thus the factory produces a likelihood coin rather than an MCMC acceptance coin or a coin derived from a target-density bound. Classical Poisson thinning [17, 18] then converts the likelihood coin into a posterior sampling mechanism: an inflated prior point process is thinned to a process with intensity proportional to the posterior, and conditional on the retained count its locations are independent posterior draws. The resulting computation contains neither a numerical likelihood evaluation nor a Markov transition.

The likelihood construction establishes exactness, but exact algorithms are useful only if their computational work can be characterized. This motivates the second part of the paper. For independent Gaussian observations, the implemented sampler reveals the Bernoulli inputs of a Bernstein event sequentially and terminates a candidate as soon as rejection is certified. We derive an exact identity for this stopped work and then study its small-noise behavior. Under a local small-ball law for the prior predictive distribution, the work exhibits three sharp regimes, with the transition determined by the prior mass near the exact-fit set. In particular, the relevant dimension is predictive rather than the nominal dimension of the parameter space. This links a statistical property of the inverse problem directly to the computational complexity of the exact sampler.

The model-facing assumption is only Bernoulli access to bounded forward observables. To show that this interface can arise without a deterministic forward solve, we construct it for a bounded class of elliptic resolvent problems. A Feynman–Kac representation yields the required forward coins through Poisson killing and a terminal Bernoulli mark. For function-valued coefficients represented by absolutely convergent random series, coordinates are generated lazily until a deterministic tail enclosure resolves the required Bernoulli decision. This PDE construction is one realization of the general computational interface rather than a restriction of the posterior algorithm to a particular equation class.

The contributions are therefore threefold. First, we construct exact Gaussian likelihood coins directly from Bernoulli-accessible forward observables, with a scalar Bernstein–Poisson factory for independent observations and a factorial-moment extension for correlated Gaussian errors. Second, we characterize the actual work of the early-stopped independent sampler and obtain sharp small-noise complexity laws governed by prior-predictive small-ball geometry. Third, we provide a continuum Feynman–Kac realization for a function-valued inverse problem and use numerical experiments to test both correctness and computational behavior. The final benchmark compares the exact stochastic sampler with finite-difference prior rejection at matched posterior accuracy; it is designed to identify the accuracy–cost tradeoff rather than to claim a universal speed advantage.

The paper is organized as follows. Section 2 develops posterior Poissonization from likelihood coins. Section 3 constructs the Gaussian likelihood factories and analyzes the work of the independent sampler. Section 4 presents the Feynman–Kac realization and its function-valued extension. Section 5 reports the numerical validation and matched-accuracy comparison, and Section 6 concludes the paper.

2 Posterior Poissonization from likelihood coins

This section isolates the point-process part of the method. On a general measurable parameter space, Bayesian updating weights the prior by the likelihood, while independent thinning weights

a Poisson intensity by a retention probability. The construction below identifies these two weights; the later sections show how to generate the required retention event without evaluating the likelihood.

2.1 Poisson point processes on a general parameter space

Let $(\mathbb{X}, \mathcal{X})$ be a measurable space and let Λ be a finite measure on it. A PPP with intensity measure Λ is a random counting measure $\Xi = \sum_{i=1}^N \delta_{Q_i}$, with the following property: for pairwise disjoint measurable sets $A_1, \dots, A_m \in \mathcal{X}$, the counts $\Xi(A_1), \dots, \Xi(A_m)$ are independent and $\Xi(A_j) \sim \text{Poisson}(\Lambda(A_j))$ for $j = 1, \dots, m$. We write $\Xi \sim \text{PPP}(\Lambda)$. Standard references include [17, 18]. Since $\Lambda(\mathbb{X}) < \infty$, the process may equivalently be generated by

$$N \sim \text{Poisson}(\Lambda(\mathbb{X})), \quad Q_i \mid N \stackrel{\text{iid}}{\sim} \frac{\Lambda}{\Lambda(\mathbb{X})}. \quad (2.1)$$

Thus a finite PPP contains two pieces of information: the total mass of its intensity is encoded by the random cardinality, whereas the normalized intensity is encoded by the conditional location law.

Independent thinning is a fundamental property of PPPs. Let $r : \mathbb{X} \rightarrow [0, 1]$ be measurable and, conditional on Ξ , retain each atom Q_i independently with probability $r(Q_i)$. The retained process is a PPP with intensity measure $r\Lambda$, defined by $(r\Lambda)(A) = \int_A r(z) d\Lambda(z)$, $A \in \mathcal{X}$. Thus thinning multiplies the original intensity by the retention function.

Two further standard properties are useful computationally. If $\Xi_1 \sim \text{PPP}(\Lambda_1)$ and $\Xi_2 \sim \text{PPP}(\Lambda_2)$ are independent, then their superposition $\Xi_1 + \Xi_2$ is $\text{PPP}(\Lambda_1 + \Lambda_2)$. Conversely, restrictions of a PPP to disjoint measurable subsets are independent. These identities justify splitting the master population into independently processed batches without changing its law and make atom-wise processing embarrassingly parallel. More generally, independent marking of the atoms produces a marked PPP, and thinning is the special case in which the mark records only “keep” or “discard.” Thus the posterior algorithm can be viewed as a marked point-process construction whose retention mark is manufactured from Bernoulli forward events.

2.2 Bayesian updating as selective survival

Let $q \in \mathbb{X}$ denote the unknown, and let μ_0 be its prior probability measure. For fixed data y , let $L(q; y) \in [0, 1]$ be a measurable reduced likelihood and assume $0 < Z(y) := \int_{\mathbb{X}} L(q; y) d\mu_0(q) < \infty$. Define

$$d\nu^y(q) = L(q; y) d\mu_0(q), \quad \frac{d\mu^y}{d\mu_0}(q) = \frac{L(q; y)}{Z(y)}.$$

If a PPP with intensity μ_0 could be thinned with retention probability $L(q; y)$, its retained intensity would be ν^y . In a PDE inverse problem the likelihood value may be unavailable without a forward solve, so we replace the numerical value by a Bernoulli decision having the required success probability.

Definition 2.1 (Scaled likelihood coin). A scaled likelihood-coin oracle with known factor $c \in (0, 1]$ returns, when queried at a parameter value $q \in \mathbb{X}$, a Bernoulli variable $C_q \in \{0, 1\}$ satisfying

$$\mathbb{P}(C_q = 1) = cL(q; y). \quad (2.2)$$

Repeated queries use independent auxiliary randomness. The numerical value of $L(q; y)$ need not be available.

We use a positive parameter γ as a Poissonization scale. It controls the overall size of the generated point process; in particular, the retained process will have expected cardinality $\gamma Z(y)$.

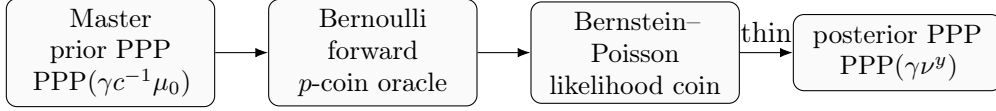


Figure 1: Likelihood-coin Poisson sampling. A Bernoulli forward oracle supplies the input coins for a scaled likelihood coin, which thins the master prior PPP into the target posterior PPP.

Theorem 2.2 (Likelihood-coin posterior Poissonization). *Suppose that the scaled likelihood-coin oracle (2.2) is available. For any Poissonization scale $\gamma > 0$, generate $\Xi_\gamma = \sum_{i=1}^{M_\gamma} \delta_{Q_i} \sim \text{PPP}(\frac{\gamma}{c}\mu_0)$. Conditional on Ξ_γ , query the oracle independently at each atom Q_i and denote the resulting Bernoulli mark by C_i , so that $\mathbb{P}(C_i = 1 \mid Q_i) = cL(Q_i; y)$. Define the retained point process by $\eta_\gamma := \sum_{i=1}^{M_\gamma} C_i \delta_{Q_i}$. Then $\eta_\gamma \sim \text{PPP}(\gamma\nu^y)$. Consequently,*

$$N_\gamma := \eta_\gamma(\mathbb{X}) \sim \text{Poisson}(\gamma Z(y)), \quad (2.3)$$

and, conditional on $N_\gamma = k \geq 1$,

$$\eta_\gamma \mid \{N_\gamma = k\} \stackrel{d}{=} \sum_{i=1}^k \delta_{Q_i^*}, \quad Q_i^* \stackrel{\text{iid}}{\sim} \mu^y.$$

Moreover, $\hat{Z}_\gamma = \frac{N_\gamma}{\gamma}$ is unbiased for $Z(y)$ with $\text{Var}(\hat{Z}_\gamma) = \frac{Z(y)}{\gamma}$.

Proof. Conditional on the master process, the marks C_i are independent and the retention probability of an atom at $z \in \mathbb{X}$ is $cL(z; y)$. Therefore, by the independent-thinning theorem, the point process $\eta_\gamma = \sum_{i=1}^{M_\gamma} C_i \delta_{Q_i}$ has intensity $\frac{\gamma}{c} cL(z; y) \mu_0(dz) = \gamma \nu^y(dz)$. Hence $\eta_\gamma \sim \text{PPP}(\gamma\nu^y)$. Its total intensity is $\gamma Z(y)$; hence (2.3) holds. Conditional on $N_\gamma = k$, the finite-PPP representation (2.1) has iid locations with law $\frac{\nu^y}{Z(y)} = \mu^y$. Finally, since $N_\gamma \sim \text{Poisson}(\gamma Z(y))$, $\mathbb{E}\hat{Z}_\gamma = Z(y)$, $\text{Var}(\hat{Z}_\gamma) = \frac{Z(y)}{\gamma}$. $\square$

Conditional on $N_\gamma = k$, Theorem 2.2 makes the retained locations an ordinary iid posterior sample. Hence, for any square-integrable posterior functional φ , the empirical posterior average has conditional variance

$$\text{Var}\left(\frac{1}{k} \sum_{i=1}^k \varphi(Q_i^*) \mid N_\gamma = k\right) = \frac{\text{Var}_{\mu^y}(\varphi)}{k}.$$

The same theorem also makes the rejection cost explicit: $\mathbb{E}M_\gamma = \gamma/c$ and $\mathbb{E}N_\gamma = \gamma Z(y)$, so the ratio of their means is $[cZ(y)]^{-1}$. This ratio records the average candidate count associated with the outer thinning step; the forward-oracle work per processed candidate is analyzed separately in Section 3.

The complete construction is summarized in Figure 1. A master PPP drawn from the prior provides the candidate atoms. A Bernoulli forward oracle generates the required forward coins, the Bernstein–Poisson factory converts them into a scaled likelihood coin, and independent thinning yields the posterior PPP. Sections 3 and 4 develop the two middle steps.

3 Bernstein–Poisson likelihood coins for Gaussian observations

The posterior construction of Section 2 requires a scaled likelihood coin, while the model-specific constructions developed later provide Bernoulli access only to individual bounded forward observables. We first treat conditionally independent Gaussian observation errors, for which the likelihood factory, the exact stopping rules, and the sharp work analysis share a common

componentwise architecture. Section 3.2 develops a compact correlated-Gaussian extension. Conditional on the unknown q , let

$$y_j = p_j(q) + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \sigma_j^2), \quad j = 1, \dots, J, \quad (3.1)$$

where $p_j(q) \in [0, 1]$, $y_j \in \mathbb{R}$, and the noise components are independent. The quantities $p_j(q)$ need not be independent or otherwise unrelated: conditional on q they are deterministic values, and in the PDE applications they are typically different observations of the same solution. After removing the Gaussian normalizing constants, the reduced likelihood is

$$L(q; y) = \exp \left\{ - \sum_{j=1}^J a_j [p_j(q) - y_j]^2 \right\}, \quad a_j = \frac{1}{2\sigma_j^2}. \quad (3.2)$$

The values $p_j(q)$ are not supplied to the sampler. For a queried parameter value, the sampler can only call an oracle that returns independent Bernoulli variables with the corresponding success probability. Since the vector likelihood factorizes, it is enough to construct a coin for one scalar factor

$$\ell_y(p) = e^{-a(p-y)^2}, \quad p \in [0, 1], \quad y \in \mathbb{R}, \quad a > 0, \quad (3.3)$$

from independent p -coins, without evaluating p .

The centered scalar factor $y = 0$ is unusually simple: two independent p -coins form a p^2 -event, and a Poisson zero-event transform then returns e^{-ap^2} . For a general component y , however, the term $(p - y)^2 = p^2 - 2yp + y^2$ contains a signed cross term and cannot itself be used as an event probability. The construction below resolves precisely this scalar obstruction; the full vector likelihood is recovered by multiplying the resulting independent component coins.

For an integer $n \geq 2$, let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$, $K = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$, and define

$$\beta_{k,n}(y) = y^2 - \frac{2yk}{n} + \frac{k(k-1)}{n(n-1)}, \quad k = 0, \dots, n. \quad (3.4)$$

Lemma 3.1. *For $K \sim \text{Binomial}(n, p)$, $\mathbb{E} \beta_{K,n}(y) = (p - y)^2$.*

Proof. Use $\mathbb{E}K = np$ and $\mathbb{E}[K(K-1)] = n(n-1)p^2$ in (3.4). $\square$

The statistic $\beta_{K,n}(y)$ is unbiased for the squared residual but may take negative values. Moreover, setting $p = y$ in Lemma 3.1 shows that $\mathbb{E} \beta_{K,n}(y) = 0$, $K \sim \text{Binomial}(n, y)$, and hence $\min_{0 \leq k \leq n} \beta_{k,n}(y) \leq 0$. We therefore define the smallest nonnegative upward shift

$$b_n(y) := \left[- \min_{0 \leq k \leq n} \beta_{k,n}(y) \right]_+ \geq 0.$$

For $y \in [0, 1]$ the minimum is nonpositive, so the positive-part operation is inactive; for observations outside this interval it keeps the scaling factor below one while preserving the same construction. Set

$$c_{k,n}(y) = \beta_{k,n}(y) + b_n(y), \quad C_n(y) = \max \left\{ 1, \max_{0 \leq k \leq n} c_{k,n}(y) \right\}. \quad (3.5)$$

The shift makes $c_{k,n}(y)$ nonnegative, while the normalization by $C_n(y)$ places it in $[0, 1]$. We can therefore convert the random statistic $c_{K,n}(y)$ into a Bernoulli event. After generating K , define the auxiliary Bernoulli variable

$$H_{n,y} \mid K = k \sim \text{Bernoulli} \left(\frac{c_{k,n}(y)}{C_n(y)} \right). \quad (3.6)$$

By conditioning on K and using Lemma 3.1,

$$\mathbb{P}(H_{n,y} = 1) = \frac{b_n(y) + (p - y)^2}{C_n(y)}.$$

For $y \in [0, 1]$, one in fact has $0 \leq c_{k,n}(y) \leq 1$, and hence $C_n(y) = 1$; see Appendix A. For observations $y \notin [0, 1]$, the same exact factory remains valid with the generic finite normalization $C_n(y)$ defined in (3.5); only the closed-form compensation and the sharp small-noise analysis below use the interior-data restriction. We retain $C_n(y)$ in the formulas to make this distinction explicit.

Crucially, the Bernstein coefficients $c_{k,n}(y)$ are completely determined by the observed datum y , the order n , and the realized success count k ; they do not require the numerical value of p . The forward model enters the construction only through $K \sim \text{Binomial}(n, p)$, which is generated by repeated calls to the p -coin oracle.

Theorem 3.2 (Bernstein–Poisson Gaussian likelihood coin). *Fix $p, y \in [0, 1]$, $a > 0$, and an integer $n \geq 2$. Suppose independent p -coins and independent auxiliary randomness are available. Let $R \sim \text{Poisson}(aC_n(y))$. Independently generate R copies of $H_{n,y}$ from (3.6), and define the Gaussian likelihood coin $G_{n,y}$ by setting $G_{n,y} = 1$ if and only if all R copies are zero. Then*

$$\mathbb{P}(G_{n,y} = 1) = e^{-ab_n(y)} e^{-a(p-y)^2}. \quad (3.7)$$

Thus $G_{n,y}$ is a scaled likelihood coin for (3.3), with known scaling factor $e^{-ab_n(y)}$.

Proof. Conditioning on $R = r$, we obtain

$$\mathbb{P}(G_{n,y} = 1 \mid R = r) = \left\{ 1 - \frac{b_n(y) + (p - y)^2}{C_n(y)} \right\}^r.$$

Using the probability generating function of a Poisson variable, $\mathbb{E}[s^R] = \exp\{aC_n(y)(s - 1)\}$, with $s = 1 - [b_n(y) + (p - y)^2]/C_n(y)$ proves (3.7). $\square$

The deterministic compensation can be evaluated explicitly.

Proposition 3.3. *Let $k_*(y, n) = y(n - 1) + \frac{1}{2}$, $d_n(y) = \text{dist}(k_*(y, n), \mathbb{Z})$. For $y \in [0, 1]$ and $n \geq 2$,*

$$b_n(y) = \frac{y(1 - y)}{n} + \frac{\frac{1}{4} - d_n(y)^2}{n(n - 1)}. \quad (3.8)$$

Consequently,

$$\frac{y(1 - y)}{n} \leq b_n(y) \leq \frac{y(1 - y)}{n} + \frac{1}{4n(n - 1)}, \quad (3.9)$$

and

$$b_n(y) = \frac{y(1 - y)}{n} + O(n^{-2}).$$

Proof. As a function of a real variable k , (3.4) is a convex quadratic with coefficient $[n(n - 1)]^{-1}$ and minimizer k_* . Completing the square puts this expression in the form

$$\beta_{k,n}(y) = -\frac{y(1 - y)}{n} - \frac{1}{4n(n - 1)} + \frac{(k - k_*)^2}{n(n - 1)}.$$

The nearest integer to $k_* \in [1/2, n - 1/2]$ belongs to $\{0, \dots, n\}$. Taking the discrete minimum and changing sign establishes (3.8); (3.9) follows from $0 \leq d_n \leq 1/2$. $\square$

3.1 Assembly of the vector likelihood

The preceding subsection constructs an exact scaled likelihood coin for a single scalar Gaussian observation. We now assemble these scalar constructions for a vector of conditionally independent observations. Choose an order $n_j \geq 2$ for each observation component and construct independent coins G_{n_j, y_j} by Theorem 3.2. Define $B_{\mathbf{n}}(y) = \sum_{j=1}^J a_j b_{n_j}(y_j)$. Here $a_j = (2\sigma_j^2)^{-1}$ is the componentwise counterpart of the scalar coefficient a used above. Since the Gaussian likelihood factorizes across the observation components, the logical AND of the J component coins has success probability

$$e^{-B_{\mathbf{n}}(y)} L(q; y). \quad (3.10)$$

Corollary 3.4. *Generate $\Xi_\gamma \sim \text{PPP}(\gamma e^{B_{\mathbf{n}}(y)} \mu_0)$, write the master process as $\Xi_\gamma = \sum_i \delta_{Q_i}$ and retain an atom Q_i if all component likelihood coins generated from the $p_j(Q_i)$ -oracles succeed. Then $\eta_\gamma \sim \text{PPP}(\gamma \nu^y)$. No numerical value of $p_j(q)$, $L(q; y)$, or $Z(y)$ is required by the sampler.*

Proof. Equation (3.10) defines a scaled likelihood coin with factor $c = e^{-B_{\mathbf{n}}(y)}$; applying Theorem 2.2 then establishes the posterior point-process statement. $\square$

3.2 A compact extension to correlated Gaussian observations

The componentwise construction above exploits likelihood factorization. This is the architecture used in the sharp work analysis below, but factorization is not essential to the Bernoulli-access principle. To make this point without introducing a second complexity theory, we record a multivariate construction for

$$y = p(q) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 \Sigma), \quad \Omega = \Sigma^{-1} \succ 0,$$

with $p(q) = (p_1(q), \dots, p_J(q))^\top \in [0, 1]^J$ and $a_\sigma = (2\sigma^2)^{-1}$. The reduced likelihood is

$$L_\Sigma(q; y) = \exp\{-a_\sigma [p(q) - y]^\top \Omega [p(q) - y]\}.$$

Choose orders $n_j \geq 2$ and, using independent auxiliary oracle randomness, draw

$$K_j \sim \text{Binomial}(n_j, p_j(q)), \quad \hat{p}_j = \frac{K_j}{n_j}, \quad \hat{p}_j^{[2]} = \frac{K_j(K_j - 1)}{n_j(n_j - 1)}.$$

Define

$$\beta_{\mathbf{n}}^\Omega(K; y) = y^\top \Omega y - 2y^\top \Omega \hat{p} + \sum_{j=1}^J \Omega_{jj} \hat{p}_j^{[2]} + 2 \sum_{1 \leq i < j \leq J} \Omega_{ij} \hat{p}_i \hat{p}_j.$$

The cross-products are unbiased because the auxiliary counts are independent conditional on q .

Theorem 3.5 (Correlated-Gaussian likelihood coin). *For every q ,*

$$\mathbb{E} \beta_{\mathbf{n}}^\Omega(K; y) = [p(q) - y]^\top \Omega [p(q) - y].$$

Let

$$b_{\mathbf{n}}^\Omega(y) = \left[- \min_{0 \leq k_j \leq n_j} \beta_{\mathbf{n}}^\Omega(k; y) \right]_+, \quad C_{\mathbf{n}}^\Omega(y) = \max \left\{ 1, \max_k [\beta_{\mathbf{n}}^\Omega(k; y) + b_{\mathbf{n}}^\Omega(y)] \right\}.$$

Conditional on K , generate

$$H_{\mathbf{n}, y}^\Omega \sim \text{Bernoulli} \left(\frac{\beta_{\mathbf{n}}^\Omega(K; y) + b_{\mathbf{n}}^\Omega(y)}{C_{\mathbf{n}}^\Omega(y)} \right).$$

If $R \sim \text{Poisson}(a_\sigma C_{\mathbf{n}}^\Omega(y))$ and $G_{\mathbf{n}, y}^\Omega = 1$ exactly when all R independent copies of $H_{\mathbf{n}, y}^\Omega$ are zero, then

$$\mathbb{P}(G_{\mathbf{n}, y}^\Omega = 1) = e^{-a_\sigma b_{\mathbf{n}}^\Omega(y)} L_\Sigma(q; y).$$

Hence correlated Gaussian observation errors can be handled without factorizing the likelihood or evaluating the forward probabilities numerically.

Proof. The factorial-moment identities satisfy $\mathbb{E}\widehat{p}_j = p_j(q)$ and $\mathbb{E}\widehat{p}_j^{[2]} = p_j(q)^2$, while independence of the auxiliary counts implies $\mathbb{E}(\widehat{p}_i\widehat{p}_j) = p_i(q)p_j(q)$ for $i \neq j$. Thus the displayed quadratic statistic is unbiased. The shift and normalization make its affine rescaling a valid Bernoulli probability, so

$$\mathbb{P}(H_{\mathbf{n},y}^\Omega = 1) = \frac{b_{\mathbf{n}}^\Omega(y) + [p(q) - y]^\top \Omega [p(q) - y]}{C_{\mathbf{n}}^\Omega(y)}.$$

The Poisson probability-generating function then evaluates to the stated success probability. $\square$

For large J , the discrete minimum need not be enumerated. With $\delta_j = \Omega_{jj}/(n_j - 1)$, define

$$\widetilde{\beta}_{\mathbf{n}}^\Omega(z; y) = (z - y)^\top \Omega (z - y) - \sum_{j=1}^J \delta_j z_j (1 - z_j), \quad z \in [0, 1]^J,$$

and let $\bar{b}_{\mathbf{n}}^\Omega(y) = [-\min_{z \in [0,1]^J} \widetilde{\beta}_{\mathbf{n}}^\Omega(z; y)]_+$. Because the count grid is contained in $[0, 1]^J$, $\bar{b}_{\mathbf{n}}^\Omega(y) \geq b_{\mathbf{n}}^\Omega(y)$, so this continuous shift is safe. The minimization is strictly convex since its Hessian is $2\Omega + 2\text{diag}(\delta_1, \dots, \delta_J) \succ 0$; moreover, for fixed $y \in (0, 1)^J$ and $n_{\min} = \min_j n_j \rightarrow \infty$,

$$b_{\mathbf{n}}^\Omega(y), \bar{b}_{\mathbf{n}}^\Omega(y) = \sum_{j=1}^J \frac{\Omega_{jj} y_j (1 - y_j)}{n_j} + O(n_{\min}^{-2}).$$

Appendix B contains the proof.

This extension shows that the likelihood-coin mechanism is not tied to diagonal observation covariance. We do not develop a parallel sharp-work theory for the multivariate factory. The remainder of this section returns to the independent factorized architecture, for which scalar range stopping and componentwise Poisson streams expose the actual work exactly.

3.3 Early stopping and order selection

A direct implementation of $H_{n,y}$ draws all n forward coins, but many outcomes can be decided earlier. Write $\widetilde{c}_{k,n}(y) = \frac{c_{k,n}(y)}{C_n(y)} \in [0, 1]$, draw $V \sim \text{Unif}(0, 1)$, and note that $H_{n,y} = \mathbf{1}_{\{V < \widetilde{c}_{K,n}(y)\}}$. After s forward flips with k successes, the final total K can only lie in $\{k, \dots, k + n - s\}$. With $\ell = n - s$, define

$$\widetilde{c}_{\min}(k, \ell) = \min_{j=k, \dots, k+\ell} \widetilde{c}_{j,n}(y), \quad \widetilde{c}_{\max}(k, \ell) = \max_{j=k, \dots, k+\ell} \widetilde{c}_{j,n}(y).$$

If $V < \widetilde{c}_{\min}(k, \ell)$, the final event is already true for every possible continuation; if $V \geq \widetilde{c}_{\max}(k, \ell)$, it is already false.

Lemma 3.6. *The range test above has exactly the same output as first drawing all n forward coins, computing K , and returning $\mathbf{1}_{\{V < \widetilde{c}_{K,n}(y)\}}$. It therefore leaves the law of $H_{n,y}$ unchanged.*

Proof. Early stopping occurs only when the indicator has the same value for every final success count that remains reachable. No admissible continuation can therefore change the result. $\square$

The Bernstein order trades master-process inflation against the number of forward-oracle calls used by each factory event. Proposition 3.3 shows that the compensation is of order n^{-1} . The next subsection specializes to a common small-noise scale and common order in order to derive both a full-batch benchmark and the sharp work of the actual range-stopped implementation.

Combining the Bernstein likelihood coin, the range-stopped implementation, and Poisson thinning completes the exact sampler summarized in Algorithm 1. For independent Gaussian observations its output is the posterior point process $\eta_\gamma \sim \text{PPP}(\gamma \nu^y)$.

Algorithm 1 Likelihood-coin posterior PPP for independent Gaussian observations

- 1: Choose orders n_j and compute $b_{n_j}(y_j)$ and $B_{\mathbf{n}}(y) = \sum_j a_j b_{n_j}(y_j)$.
 - 2: Set $A_{\mathbf{n}} := \sum_{j=1}^J a_j C_{n_j}(y_j)$ and $\pi_j := a_j C_{n_j}(y_j) / A_{\mathbf{n}}$.
 - 3: Draw $M_\gamma \sim \text{Poisson}(\gamma e^{B_{\mathbf{n}}(y)})$ and candidates $Q_i \stackrel{\text{iid}}{\sim} \mu_0$, $i = 1, \dots, M_\gamma$.
 - 4: **for** $i = 1, \dots, M_\gamma$ **do**
 - 5: Draw $R \sim \text{Poisson}(A_{\mathbf{n}})$ and set keep = 1.
 - 6: **for** $h = 1, \dots, R$ **do**
 - 7: Independently draw a component label j with probability π_j .
 - 8: Generate an early-stopped Bernstein event H_{n_j, y_j} from stochastic $p_j(Q_i)$ -coins.
 - 9: **if** $H_{n_j, y_j} = 1$ **then**
 - 10: Set keep = 0 and terminate the remaining likelihood tests for candidate i .
 - 11: **end if**
 - 12: **end for**
 - 13: **if** keep = 1 **then**
 - 14: retain Q_i .
 - 15: **end if**
 - 16: **end for**
 - 17: Return the retained point process.
-

The marked Poisson stream in Algorithm 1 is equivalent, before stopping, to independent component streams with counts $R_j \sim \text{Poisson}(a_j C_{n_j}(y_j))$. Its randomized component order preserves the likelihood-coin law and is used in the exact work identity below. Algorithm 1 contains two distinct forms of early termination. Within each Bernstein event, the range test of Lemma 3.6 may decide the event before all n_j forward coins are generated. At the candidate level, the first rejected auxiliary event terminates all remaining likelihood tests for that candidate. Neither mechanism changes the retained point-process law; both affect only the amount of forward-oracle work. The choice of the orders n_j , and hence the balance between master-process inflation and range-stopped work, is analyzed in Section 3.4.

3.4 Common-order benchmark and sharp stopped-work complexity

Algorithm 1 saves work in two different ways. Within a single Bernstein event, the range rule may determine the event before all forward coins are revealed. At the candidate level, the first rejecting auxiliary event terminates all remaining likelihood tests for that candidate. The goal of this subsection is to quantify these two savings and to identify the small-noise regime in which each one matters.

The exact likelihood factory above permits arbitrary fixed real observations. For the sharp complexity analysis we impose the additional interior-data condition $y \in (0, 1)^J$, which yields $C_n(y_j) = 1$ and the explicit compensation asymptotics of Proposition 3.3. To keep the complexity analysis focused on the stopping mechanism rather than on componentwise allocation, we specialize to a common observational scale $\sigma_1 = \dots = \sigma_J = \sigma$, write $a_\sigma = (2\sigma^2)^{-1}$, and use the same Bernstein order n for every component. The scaled order $\vartheta_\sigma = n/a_\sigma$ compares the Bernstein order with the inverse-noise scale. We also set

$$Q_y(q) := \sum_{j=1}^J [p_j(q) - y_j]^2, \quad b_n^{\text{tot}}(y) := \sum_{j=1}^J b_n(y_j), \quad D_y := \sum_{j=1}^J y_j(1 - y_j).$$

Here $Q_y(q)$ is the predictive squared misfit. The quantity $b_n^{\text{tot}}(y)$ is the total deterministic compensation introduced by the Bernstein factories, and therefore controls the inflation of the

master PPP. By Proposition 3.3,

$$b_n^{\text{tot}}(y) = \frac{D_y}{n} + O(n^{-2}). \quad (3.11)$$

We begin with a benchmark in which both stopping mechanisms are disabled. In this full-batch implementation every auxiliary event uses all n forward coins, and every Poisson event is processed even if the candidate could already have been rejected.

Proposition 3.7. *Assume $y \in (0, 1)^J$ and define the scaled Bernstein order $\vartheta_\sigma := n/a_\sigma$. Suppose that $\vartheta_\sigma \rightarrow \vartheta \in (0, \infty)$ as $\sigma \downarrow 0$. Equivalently, $n = n(\sigma) \sim \vartheta a_\sigma$, so ϑ measures the asymptotic Bernstein order per unit small-noise scale. If W_{full} denotes the total number of forward Bernoulli calls in the full-batch implementation, then*

$$\frac{\mathbb{E}W_{\text{full}}}{\gamma a_\sigma^2} \longrightarrow J\vartheta \exp\left(\frac{D_y}{\vartheta}\right).$$

Among orders with $n/a_\sigma \rightarrow \vartheta \in (0, \infty)$, the limit is minimized at $\vartheta = D_y$. Hence the minimizing full-batch scale satisfies

$$n \sim a_\sigma D_y, \quad a_\sigma b_n^{\text{tot}}(y) \rightarrow 1, \quad e^{a_\sigma b_n^{\text{tot}}(y)} \rightarrow e.$$

Proof. The master PPP has mean size $\gamma e^{a_\sigma b_n^{\text{tot}}(y)}$. Because $C_n(y_j) = 1$, each of the J component factories has a $\text{Poisson}(a_\sigma)$ auxiliary count, and each completed event costs exactly n forward calls. Therefore $\mathbb{E}W_{\text{full}} = \gamma e^{a_\sigma b_n^{\text{tot}}(y)} J a_\sigma n$. Equation (3.11), together with $\vartheta_\sigma \rightarrow \vartheta$, establishes the stated limit. Minimizing $\vartheta e^{D_y/\vartheta}$ yields $\vartheta = D_y$. $\square$

Proposition 3.7 identifies the full-batch benchmark: without early stopping, the natural complexity scale is $\gamma a_\sigma^2 = \Theta(\gamma \sigma^{-4})$. We next quantify the first saving mechanism: stopping within a single Bernstein event.

For a scalar component with forward probability $p \in [0, 1]$, let $\tau_{n,y}(p)$ be the number of p -coin calls used by the range rule in Lemma 3.6, and let $\bar{\tau}_{n,y}(p) = \mathbb{E}\tau_{n,y}(p)$. Suppose that a fraction t of the n forward coins has been revealed. For large n , the revealed successes contribute approximately tp to the final normalized count, while the unrevealed fraction $1 - t$ may still contribute anything between 0 and $1 - t$. This motivates the interval $I_t(p) := [tp, tp + 1 - t]$. Define

$$\kappa_y(p) := \int_0^1 \left(\sup_{z \in I_t(p)} (z - y)^2 - \inf_{z \in I_t(p)} (z - y)^2 \right) dt.$$

The integrand measures how much uncertainty about the final normalized count can still change the squared residual after a fraction t of the forward coins has been revealed. Accordingly, $\kappa_y(p)$ will be the asymptotic fraction of the n calls that remain necessary under the range rule.

Proposition 3.8. *For every fixed $y \in [0, 1]$,*

$$\sup_{p \in [0, 1]} \left| \frac{\bar{\tau}_{n,y}(p)}{n} - \kappa_y(p) \right| \longrightarrow 0. \quad (3.12)$$

Moreover, κ_y is continuous and $1/12 \leq \kappa_y(p) \leq 1$ for all $p \in [0, 1]$.

Proof. Let K_m be the number of successes among the first m input coins and put $t_m = m/n$. After these m calls, the reachable normalized final counts form the grid

$$\mathcal{I}_{m,n} := \left\{ \frac{K_m + r}{n} : 0 \leq r \leq n - m \right\} \subset \left[\frac{K_m}{n}, \frac{K_m + n - m}{n} \right].$$

For $z = k/n$, a direct calculation gives

$$\beta_{k,n}(y) = (z - y)^2 - \frac{z(1 - z)}{n - 1},$$

so

$$\sup_{0 \leq k \leq n} |\beta_{k,n}(y) - (k/n - y)^2| \leq \frac{1}{4(n - 1)}.$$

The compensation $b_n(y)$ is additive and therefore does not affect the range width. Since $z \mapsto (z - y)^2$ is 2-Lipschitz on $[0, 1]$, the width of the reachable threshold range differs, by at most a universal constant times

$$\left| \frac{K_m}{n} - t_m p \right| + \frac{1}{n},$$

from

$$\sup_{z \in I_{t_m}(p)} (z - y)^2 - \inf_{z \in I_{t_m}(p)} (z - y)^2.$$

Here the additional n^{-1} term accounts for the count-grid spacing and the $O(n^{-1})$ factorial-moment correction above. Uniformly in $p \in [0, 1]$,

$$\mathbb{E} \left| \frac{K_m}{n} - t_m p \right| \leq \frac{\sqrt{mp(1-p)}}{n} \leq \frac{1}{2\sqrt{n}}.$$

Consequently the expected reachable-range width converges uniformly in p and m to the corresponding deterministic width, up to an $O(n^{-1/2})$ error.

The reachable threshold intervals are nested as m increases. Because the auxiliary $V \sim \text{Unif}(0, 1)$ is independent of the input coins, $\mathbb{P}(\tau_{n,y}(p) > m \mid K_m)$ equals the width of the reachable normalized threshold interval after m calls. The tail-sum identity therefore yields

$$\frac{\bar{\tau}_{n,y}(p)}{n} = \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{E}[\text{reachable threshold width at } m].$$

The preceding uniform estimate and the uniform Riemann convergence of the continuous range-width function establish (3.12). Continuity of κ_y follows from continuity of the extrema of a continuous function over the interval $I_t(p)$. Since $(z - y)^2 \in [0, 1]$, the upper bound is immediate. Finally, any interval of length $1 - t$ has squared-residual range at least $(1 - t)^2/4$; integrating over $t \in [0, 1]$ gives the lower bound $1/12$. $\square$

For the vector likelihood write

$$\bar{\tau}_n^{\text{tot}}(q) := \sum_{j=1}^J \bar{\tau}_{n,y_j}(p_j(q)), \quad \kappa_y^{\text{tot}}(p) := \sum_{j=1}^J \kappa_{y_j}(p_j).$$

The previous proposition describes the cost of one auxiliary event. The second, and more important, saving comes from terminating a candidate as soon as the first auxiliary event rejects it. The next identity combines both levels of stopping.

Theorem 3.9 (Exact mean actual-work identity). *Let W_{act} be the total number of forward Bernoulli calls made by Algorithm 1 with common order n , exact scalar range stopping, and candidate termination at the first auxiliary event equal to one. Then*

$$\mathbb{E} W_{\text{act}} = \gamma e^{a_\sigma b_n^{\text{tot}}(y)} \int_{\mathbb{X}} \bar{\tau}_n^{\text{tot}}(q) \frac{1 - e^{-a_\sigma [b_n^{\text{tot}}(y) + Q_y(q)]}}{b_n^{\text{tot}}(y) + Q_y(q)} \mu_0(dq). \quad (3.13)$$

Proof. In the common-order setting $C_n(y_j) = 1$, so the marked stream in Algorithm 1 has $R \sim \text{Poisson}(Ja_\sigma)$ events and independent uniform component labels in $\{1, \dots, J\}$. For fixed q , a randomly labelled event rejects with probability $h_n(q) = [b_n^{\text{tot}}(y) + Q_y(q)]/J$, while its mean forward-call cost is $J^{-1}\bar{\tau}_n^{\text{tot}}(q)$. Conditional on $R = r$, the expected number of events that are started before the first rejection is $[1 - (1 - h_n(q))^r]/h_n(q)$. Averaging over the Poisson count evaluates to

$$J \frac{1 - e^{-a_\sigma[b_n^{\text{tot}}(y) + Q_y(q)]}}{b_n^{\text{tot}}(y) + Q_y(q)}.$$

Multiplying by the mean cost of a started event cancels the factor J . Finally, averaging over the master PPP leads to (3.13). $\square$

The identity separates the two sources of computational cost and makes the small-noise mechanism transparent. Apart from the master-process inflation and the cost of one started event, the decisive factor is

$$g_{a_\sigma}(s) := \frac{1 - e^{-a_\sigma s}}{s}, \quad s > 0.$$

If $a_\sigma s \ll 1$, then $g_{a_\sigma}(s) \sim a_\sigma$; if $a_\sigma s \gg 1$, then $g_{a_\sigma}(s) \sim s^{-1}$. Thus a parameter value with large predictive misfit $Q_y(q)$ is rejected after only a few auxiliary events, whereas a nearly exact fit with $Q_y(q) \approx 0$ survives much longer. The total work is therefore determined by how much prior mass lies near the exact-fit set $\{q : Q_y(q) = 0\}$.

We encode this local geometry through a small-ball condition.

Assumption 3.10 (Prior-predictive small-ball law). For the fixed observation vector y , there exist $c_y > 0$ and $\alpha > 0$ such that

$$F_y(t) := \mu_0\{q : Q_y(q) \leq t\} = c_y t^\alpha \{1 + o(1)\}, \quad t \downarrow 0.$$

The exponent α measures the local amount of prior mass that can produce an almost exact fit. If the prior predictive law of $p(q)$ has a nondegenerate d_{pred} -dimensional density near y , then $Q_y(q) = \|p(q) - y\|_2^2$ and a ball of squared radius t has mass of order $t^{d_{\text{pred}}/2}$. Hence $\alpha = d_{\text{pred}}/2$, so $d_{\text{pred}} = 2\alpha$ is the effective local predictive dimension. This dimension concerns the image of the prior under the observation map and need not agree with the dimension of the unknown q .

For $0 < \alpha < 1$, define

$$\Psi_\alpha(\beta) := \alpha \int_0^\infty t^{\alpha-1} \frac{1 - e^{-(\beta+t)}}{\beta+t} dt = \Gamma(\alpha+1) \int_0^1 e^{-\beta s} s^{-\alpha} ds. \quad (3.14)$$

Theorem 3.11 (Sharp small-noise complexity). Assume $y \in (0, 1)^J$, Assumption 3.10, and $\vartheta_\sigma = n/a_\sigma \rightarrow \vartheta > 0$. Put $\beta = D_y/\vartheta$. Then

$$\mathbb{E}W_{\text{act}} \sim \begin{cases} \gamma e^\beta \vartheta c_y \kappa_y^{\text{tot}}(y) \Psi_\alpha(\beta) a_\sigma^{2-\alpha}, & 0 < \alpha < 1, \\ \gamma e^\beta \vartheta c_y \kappa_y^{\text{tot}}(y) a_\sigma \log a_\sigma, & \alpha = 1, \\ \gamma e^\beta \vartheta \mathcal{I}_y a_\sigma, & \alpha > 1, \end{cases} \quad (3.15)$$

where $\mathcal{I}_y := \int_{\mathbb{X}} \frac{\kappa_y^{\text{tot}}(p(q))}{Q_y(q)} \mu_0(dq) < \infty$, ($\alpha > 1$). Consequently,

$$\mathbb{E}W_{\text{act}} = \begin{cases} \Theta(\gamma \sigma^{-4+2\alpha}), & 0 < \alpha < 1, \\ \Theta(\gamma \sigma^{-2} \log(1/\sigma)), & \alpha = 1, \\ \Theta(\gamma \sigma^{-2}), & \alpha > 1. \end{cases}$$

Proof. Equation (3.11) and $n/a_\sigma \rightarrow \vartheta$ imply $a_\sigma b_n^{\text{tot}}(y) \rightarrow \beta$. Proposition 3.8 gives the uniform expansion

$$\sup_{q \in \mathbb{X}} \left| \frac{\bar{\tau}_n^{\text{tot}}(q)}{a_\sigma \vartheta} - \kappa_y^{\text{tot}}(p(q)) \right| \rightarrow 0.$$

Because $Q_y(q) = \|p(q) - y\|_2^2$, the condition $Q_y(q) \leq t$ forces $\|p(q) - y\|_2 \leq \sqrt{t}$. Continuity of the functions κ_{y_j} from Proposition 3.8 therefore implies

$$\sup_{Q_y(q) \leq t} |\kappa_y^{\text{tot}}(p(q)) - \kappa_y^{\text{tot}}(y)| \rightarrow 0, \quad t \downarrow 0.$$

Moreover $J/12 \leq \kappa_y^{\text{tot}} \leq J$, so the uniform $o(1)$ term above is negligible relative to the weighted integrals appearing in (3.13). The regular-variation lemma in Appendix C, applied with $w(q) = \kappa_y^{\text{tot}}(p(q))$, now yields the cases $0 < \alpha < 1$ and $\alpha = 1$ with the displayed leading constants.

If $\alpha > 1$, Assumption 3.10 implies $\mu_0\{Q_y = 0\} = 0$ and local integrability of Q_y^{-1} near the exact-fit set. Away from that set Q_y^{-1} is bounded. Since κ_y^{tot} is bounded, dominated convergence gives the finite constant $\mathcal{I}_y$ and the third asymptotic regime. Finally, $a_\sigma = (2\sigma^2)^{-1}$ gives the equivalent rates in σ . $\square$

Theorem 3.11 is the main asymptotic complexity result of the paper: it provides leading-order constants as well as the phase transition in the small-noise exponent. The theorem shows precisely what early candidate rejection changes. The full-batch benchmark is always of order $\gamma\sigma^{-4}$. With exact stopping, the rate is governed by the local predictive exponent. In particular, if $d_{\text{pred}} = 2\alpha$, then

$$d_{\text{pred}} = 1 : \Theta(\gamma\sigma^{-3}), \quad d_{\text{pred}} = 2 : \Theta(\gamma\sigma^{-2} \log(1/\sigma)), \quad d_{\text{pred}} > 2 : \Theta(\gamma\sigma^{-2}).$$

The transition at $d_{\text{pred}} = 2$ comes from the borderline integrability of $Q_y(q)^{-1}$ near the exact-fit set.

For fixed γ , however, the expected number of retained posterior atoms also decreases as the noise level shrinks. Indeed, the same small-ball law implies

$$Z_\sigma(y) := \int e^{-a_\sigma Q_y(q)} \mu_0(dq) \sim c_y \Gamma(\alpha + 1) a_\sigma^{-\alpha},$$

so $\mathbb{E}N_\gamma = \gamma Z_\sigma(y)$. Dividing the total work by the expected retained count is the cost per posterior atom.

Corollary 3.12. *Under Assumption 3.10,*

$$\frac{\mathbb{E}W_{\text{act}}}{\mathbb{E}N_\gamma} = \begin{cases} \Theta(a_\sigma^2), & 0 < \alpha < 1, \\ \Theta(a_\sigma^2 \log a_\sigma), & \alpha = 1, \\ \Theta(a_\sigma^{\alpha+1}), & \alpha > 1. \end{cases}$$

Equivalently, in terms of $d_{\text{pred}} = 2\alpha$, the cost per expected posterior atom is $\Theta(\sigma^{-4})$ for $d_{\text{pred}} = 1$, $\Theta(\sigma^{-4} \log(1/\sigma))$ for $d_{\text{pred}} = 2$, and $\Theta(\sigma^{-(d_{\text{pred}}+2)})$ for $d_{\text{pred}} > 2$.

Proof. Use the same regular-variation lemma for $Z_\sigma(y)$ and divide (3.15) by $\gamma Z_\sigma(y)$. $\square$

The same asymptotics determine how the Bernstein order should scale with the noise level. Because $\vartheta_\sigma = n/a_\sigma \rightarrow \vartheta$ and $\beta = D_y/\vartheta$, the parameter β is exactly the limiting compensation exponent: $a_\sigma b_n^{\text{tot}}(y) \rightarrow \beta$ and the master-process inflation tends to e^β . Optimizing the order is therefore equivalent to optimizing the leading constant with respect to β .

Theorem 3.13 (Asymptotically optimal Bernstein order). *Under the assumptions of Theorem 3.11, consider orders satisfying $n/a_\sigma \rightarrow \vartheta \in (0, \infty)$. The leading asymptotic work is minimized by*

$$n^*(\sigma) \sim \frac{a_\sigma D_y}{\beta_\alpha},$$

where

$$\beta_\alpha = \begin{cases} \arg \min_{\beta > 0} \frac{e^\beta \Psi_\alpha(\beta)}{\beta}, & 0 < \alpha < 1, \\ 1, & \alpha \geq 1. \end{cases}$$

For $0 < \alpha < 1$, the minimizer is unique and is characterized by

$$1 - \frac{1}{\beta_\alpha} + \frac{\Psi'_\alpha(\beta_\alpha)}{\Psi_\alpha(\beta_\alpha)} = 0.$$

Proof. Since $y \in (0, 1)^J$, one has $D_y > 0$, and the change of variables $\vartheta = D_y/\beta$ is one-to-one on $(0, \infty)$. The β -dependent factor in (3.15) is $D_y e^\beta \Psi_\alpha(\beta)/\beta$ for $0 < \alpha < 1$ and $D_y e^\beta/\beta$ for $\alpha \geq 1$. For $0 < \alpha < 1$, (3.14) writes Ψ_α as the Laplace transform of a nondegenerate positive measure on $(0, 1)$, hence $\log \Psi_\alpha$ is convex. Therefore

$$\beta + \log \Psi_\alpha(\beta) - \log \beta$$

is strictly convex because $-\log \beta$ is strictly convex. The objective diverges as $\beta \downarrow 0$ because of the factor β^{-1} and also as $\beta \rightarrow \infty$ because the exponential factor dominates the polynomial decay of $\Psi_\alpha(\beta)$. A unique minimizer therefore exists, and differentiation gives the stated first-order equation. For $\alpha \geq 1$, e^β/β has the unique minimizer $\beta = 1$. $\square$

Thus early stopping does more than reduce a constant. The within-event range rule changes the average cost of each started Bernstein event, while candidate-level termination changes the small-noise exponent itself. The latter effect is controlled by the local prior-predictive geometry through α , not by the nominal dimension of q . The only remaining model-specific requirement is exact Bernoulli access to the bounded forward observables; Section 4 constructs such access for a class of elliptic PDEs.

4 A PDE realization of the Bernoulli-access interface

The likelihood-coin framework developed in Section 3 is model-agnostic once exact Bernoulli access to the forward observables is available. The likelihood factory does not otherwise depend on the governing equation. This section provides one concrete realization of that interface for a bounded class of elliptic PDEs. A Feynman–Kac representation expresses each observable as the expectation of a bounded path functional, which can be realized directly as a Bernoulli event rather than estimated by Monte Carlo averaging. For the elliptic resolvent problem below, the reaction term is implemented through Poisson killing and the source term through a terminal Bernoulli mark. Function-valued coefficients are handled by lazy evaluation of convergent random series, so the posterior sampler requires neither a deterministic PDE solve nor a fixed truncation of the unknown.

4.1 Feynman–Kac representation and Bernoulli access

Let $D \subset \mathbb{R}^d$ be bounded, let $(B_t)_{t \geq 0}$ be Brownian motion with generator $\frac{1}{2}\Delta$, and let τ_D be its first exit time. For

$$\left(q - \frac{1}{2}\Delta\right)u = f \quad \text{in } D, \quad u = g \quad \text{on } \partial D,$$

with $q \geq 0$, under the standard regularity assumptions ensuring the classical Feynman–Kac representation, the solution is represented as

$$u(x) = \mathbb{E}_x \left[e^{-\int_0^{\tau_D} q(\mathbf{B}_s) ds} g(\mathbf{B}_{\tau_D}) + \int_0^{\tau_D} e^{-\int_0^t q(\mathbf{B}_s) ds} f(\mathbf{B}_t) dt \right].$$

See, for example, [19, 20]. The exponential factor is the survival probability under the pathwise hazard $q(\mathbf{B}_t)$. Whenever a stochastic representation can be written as $u_q(x) = \mathbb{E}[Y_q(x) \mid q]$ with $0 \leq Y_q(x) \leq 1$, an independent $U \sim \text{Unif}(0, 1)$ turns it directly into the exact forward coin $\mathbf{1}_{\{U < Y_q(x)\}} \sim \text{Bernoulli}(u_q(x))$. Thus the method uses a single randomized path event rather than first estimating $u_q(x)$ numerically:

$$\text{path event} \longrightarrow u_q(x)\text{-coin} \longrightarrow \text{likelihood coin} \longrightarrow \text{PPP thinning}.$$

4.2 A bounded Feynman–Kac resolvent class

On the two-dimensional torus $\mathbb{T}^2$, let $\lambda > 0$ and $q_{\max} > 0$, and assume $0 \leq q(x) \leq q_{\max}$, $0 \leq h(x) \leq \lambda$. Consider

$$\left(\lambda + q(x) - \frac{1}{2} \Delta \right) u(x) = h(x) \quad \text{on } \mathbb{T}^2. \quad (4.1)$$

If $T \sim \text{Exp}(\lambda)$ is independent of Brownian motion on $\mathbb{T}^2$, the resolvent form of Feynman–Kac is

$$u(s) = \mathbb{E}_s \left[\exp \left\{ - \int_0^T q(\mathbf{B}_t) dt \right\} \frac{h(\mathbf{B}_T)}{\lambda} \right]. \quad (4.2)$$

The bounds on q and h imply $0 \leq u \leq 1$. Here λ is both the resolvent parameter and the rate of the exponential clock; the choice $h \equiv \lambda$ used later is only a specialization for which the terminal source coin is identically one.

Assume that at a queried location x we can generate independent coins with success probabilities $q(x)/q_{\max}$ and $h(x)/\lambda$. The first coin marks potential killing events and the second determines the terminal payoff.

Theorem 4.1 (Poisson-killing forward coin). *Fix $s \in \mathbb{T}^2$ and a coefficient pair (q, h) satisfying the bounds above. For each oracle query use fresh auxiliary randomness, conditionally independent given (q, h) . Draw $T \sim \text{Exp}(\lambda)$ and generate a homogeneous Poisson process of rate $q_{\max}$ on $[0, T]$. At each event time τ_r , simulate the Brownian location $\mathbf{B}_{\tau_r}$ and generate a $q(\mathbf{B}_{\tau_r})/q_{\max}$ -coin. Return zero at the first successful killing mark. If no killing mark succeeds before T , generate an $h(\mathbf{B}_T)/\lambda$ -coin and return its outcome. Then*

$$J_u(s) \sim \text{Bernoulli}(u(s)),$$

where u is defined by (4.2).

Proof. Condition on T and the Brownian path. Thinning the rate- $q_{\max}$ Poisson process with acceptance probability $q(\mathbf{B}_t)/q_{\max}$ induces an inhomogeneous killing process with rate $q(\mathbf{B}_t)$. The conditional probability of no accepted killing event before T is therefore $\exp \left\{ - \int_0^T q(\mathbf{B}_t) dt \right\}$. Conditional on survival, the terminal coin succeeds with probability $h(\mathbf{B}_T)/\lambda$. Taking expectation with respect to the Brownian path and the independent exponential time T recovers (4.2). $\square$

4.3 Lazy Bernoulli access to function-valued unknowns

Let $\{\phi_\ell\}_{\ell \geq 1}$ be measurable functions on a domain $\mathcal{D}$ with $|\phi_\ell(x)| \leq 1$, and let

$$f_\xi(x) = b + \sum_{\ell=1}^{\infty} \omega_\ell \xi_\ell \phi_\ell(x), \quad |\xi_\ell| \leq 1,$$

where

$$A := \sum_{\ell=1}^{\infty} |\omega_\ell| < \min\{b, 1 - b\}. \quad (4.3)$$

The coordinate space is $\mathcal{X}_\xi = [-1, 1]^\mathbb{N}$ with its product σ -algebra and a product prior π_0 . If the ϕ_ℓ are continuous, absolute summability implies uniform convergence and measurability of $\xi \mapsto f_\xi$ into $C(\mathcal{D})$; the coordinate prior therefore induces a function-space prior by pushforward.

By (4.3), $f_\xi(x) \in (0, 1)$ uniformly. Define $S_r(x) = b + \sum_{\ell=1}^r \omega_\ell \xi_\ell \phi_\ell(x)$, $R_r = \sum_{\ell > r} |\omega_\ell|$. Then

$$S_r(x) - R_r \leq f_\xi(x) \leq S_r(x) + R_r. \quad (4.4)$$

For a single $U \sim \text{Unif}(0, 1)$, reveal coefficients sequentially until either $U < S_r - R_r$ or $U > S_r + R_r$.

Theorem 4.2 (Exact lazy function-value coin). *Assume (4.3). For every fixed x , the lazy decision stops after finitely many coefficient reveals almost surely and returns $J_f(x) = \mathbf{1}_{\{U < f_\xi(x)\}}$. Hence, conditional on f_ξ , $J_f(x) \sim \text{Bernoulli}(f_\xi(x))$. If τ is the number of revealed coefficients, then*

$$\mathbb{P}(\tau > r \mid \xi) \leq \min\{1, 2R_r\}, \quad \mathbb{E}(\tau \mid \xi) \leq \sum_{r=0}^{\infty} \min\{1, 2R_r\}. \quad (4.5)$$

In particular, if $|\omega_\ell| = A(1 - \rho)\rho^{\ell-1}$, $0 < \rho < 1$, then $R_r = A\rho^r$ and, when $2A \leq 1$, $\mathbb{E}\tau \leq \frac{2A}{1-\rho}$.

Proof. Since $R_r \rightarrow 0$, the intervals in (4.4) shrink to $f_\xi(x)$. Because U is continuous and independent, $\mathbb{P}(U = f_\xi(x) \mid \xi) = 0$, so the decision resolves almost surely. The event $\{\tau > r\}$ requires $U \in [S_r - R_r, S_r + R_r] \cap [0, 1]$, whose length is at most $2R_r$. Summing the tail probabilities establishes (4.5); the geometric case follows directly. $\square$

A retained atom may therefore be represented by a lazily revealed infinite coefficient sequence. Each resolved field-value decision is measurable with respect to the revealed prefix and its auxiliary uniform variable and is unchanged for every admissible completion of the tail. Because the coordinate prior is a product measure, coefficients beyond the largest revealed index retain their product-prior law conditional on the realized algorithmic history and may be generated later when a fuller realization is needed.

We use two specializations of (4.1). If $q \equiv 0$ and $h = \lambda f_\xi$, then

$$\left(\lambda - \frac{1}{2}\Delta\right) u_{f_\xi} = \lambda f_\xi, \quad u_{f_\xi}(s) = \mathbb{E}_s[f_\xi(\mathbf{B}_T)],$$

and Theorem 4.2 implements the terminal $f_\xi(\mathbf{B}_T)$ -coin. This furnishes an exact inverse-source specialization of the same oracle construction.

If $h \equiv \lambda$ and

$$q_\xi(x) = q_{\max} f_\xi(x), \quad 0 < f_\xi(x) < 1, \quad (4.6)$$

then

$$\left(\lambda + q_\xi(x) - \frac{1}{2}\Delta\right) u_{q_\xi} = \lambda, \quad (4.7)$$

and

$$p_s(q_\xi) := u_{q_\xi}(s) = \mathbb{E}_s \left[\exp \left\{ - \int_0^T q_\xi(\mathbf{B}_t) dt \right\} \right].$$

Here the terminal coin is one and each killing mark is generated by the lazy f_ξ -coin.

For this reaction specialization, define $\mathcal{R}_f := \sum_{r=0}^{\infty} \min\{1, 2R_r\}$. For a common Bernstein order n , write $B_n(y) := \sum_{j=1}^J a_j b_n(y_j)$.

Proposition 4.3. *Let N_{kill} be the number of candidate killing events inspected by one forward coin for (4.7), and let N_{coef} be the total number of random-series coefficient terms inspected by its lazy field-value coins, counting repeated inspections across distinct field queries. Then*

$$\mathbb{E}N_{\text{kill}} \leq \frac{q_{\max}}{\lambda}, \quad \mathbb{E}N_{\text{coef}} \leq \frac{q_{\max}}{\lambda} \mathcal{R}_f.$$

For a common Bernstein order n and Gaussian observations (3.2),

$$\mathbb{E}W_{\text{coef}} \leq \gamma e^{B_n(y)n} \left(\sum_{j=1}^J a_j C_n(y_j) \right) \frac{q_{\max}}{\lambda} \mathcal{R}_f. \quad (4.8)$$

The bound contains neither a spatial discretization parameter nor a fixed truncation dimension for the random field.

Proof. If every rate- $q_{\max}$ event on $[0, T]$ were inspected, the conditional mean count would be $q_{\max}T$. Early termination at the first accepted killing event can only reduce the count, so $\mathbb{E}N_{\text{kill}} \leq q_{\max}\mathbb{E}T = q_{\max}/\lambda$. Each inspected killing location invokes a lazy field-value coin with expected inspection count at most $\mathcal{R}_f$. A Bernstein event uses at most n forward coins, sensor j generates a Poisson number of such events with mean $a_j C_n(y_j)$, and the inflated master PPP has mean cardinality $\gamma e^{B_n(y)}$. Ignoring all early terminations leads to (4.8). $\square$

Proposition 4.3 is a model-side work bound for one Bernoulli-accessible PDE class. When the prior predictive law of the observable vector also satisfies Assumption 3.10, Theorem 3.11 supplies the corresponding sharp likelihood-factory rate; the factor $q_{\max}\mathcal{R}_f/\lambda$ remains a conservative bound for the average coefficient-query cost of one forward Bernoulli call.

Corollary 4.4. *Let π_0 be a product prior on $\mathcal{X}_\xi$, let $\mathcal{Q}(\xi) = q_\xi$, and let $\mu_0 = \mathcal{Q}_\# \pi_0$ be the induced prior on the reaction fields (4.6). Use Theorem 4.1 to generate the forward coins in the Bernstein–Poisson likelihood factory at every sensor, keeping the same sampled field q_ξ for a candidate while using fresh path and marking randomness for repeated oracle calls. Then*

$$\eta_\gamma \sim \text{PPP}(\gamma \nu^y). \quad (4.9)$$

Conditional on its cardinality, the retained fields are independent draws from the continuum Bayesian posterior. No deterministic spatial discretization or fixed parameter truncation enters the mathematical target.

Proof. Theorem 4.1 establishes exact Bernoulli access to each $p_j(q_\xi)$. The Gaussian factory therefore constructs a coin with success probability $e^{-B_n(y)} L(q_\xi; y)$, and Corollary 3.4 establishes (4.9). The conditional-iid statement follows from Theorem 2.2. $\square$

The continuum oracle is deliberately restricted to bounded Bernoulli-accessible resolvent observables. In the present construction the reaction term is nonnegative and bounded by a known $q_{\max}$, and the terminal source mark lies in $[0, 1]$. Unknown diffusion coefficients, sign-changing or unbounded potentials, and oscillatory or complex-valued observables require different stochastic representations and are outside the current theory.

Table 1: Two-parameter benchmark: continuum forward-coin validation.

sensor	exact probability	coin estimate	observed data	$2\widehat{\text{SE}}$
(0, 1/4)	0.520226	0.521000	0.532226	0.005768
(1/4, 0)	0.396220	0.396200	0.386220	0.005648

5 Numerical experiments

The numerical study proceeds from validation to computational assessment. We first verify the forward and likelihood coins against analytically available reference quantities and test the sharp work regimes predicted by the small-ball theory. We then move to a function-valued inverse problem, where the continuum Bernoulli interface is implemented through Feynman–Kac sampling and lazy coefficient revelation. The final experiment compares the resulting stochastic sampler with finite-difference rejection at matched posterior accuracy. Deterministic PDE solves enter only through reference calculations and the comparative benchmark, not through the continuum sampler itself.

5.1 Two-parameter PDE benchmark

We first use a two-dimensional problem with an analytically available reference to validate separately the forward Bernoulli oracle, the likelihood factories, and the resulting posterior Poisson thinning.

On $\mathbb{T}^2$, let $\theta = (\theta_1, \theta_2) \in [-1, 1]^2$ and consider

$$\left(36 - \frac{1}{2}\Delta\right) u_\theta(x) = 32 \left\{ \frac{1}{2} + 0.24\theta_1 \cos(2\pi x_1) + 0.24\theta_2 \cos(2\pi x_2) \right\}.$$

The prior is uniform on $[-1, 1]^2$. At $s_1 = (0, 1/4)$ and $s_2 = (1/4, 0)$,

$$p_j(\theta) = \frac{16}{36} + \frac{32(0.24)}{36 + 2\pi^2} \theta_j, \quad j = 1, 2.$$

The closed form is used only for deterministic reference calculations and calibration; the continuum PDE oracle uses the Feynman–Kac construction of Section 4.

We take $\theta^\dagger = (0.55, -0.35)$ and $\sigma = 0.055$, and generate the observations by adding the fixed perturbation $(0.012, -0.010)$ to the corresponding forward values. Table 1 reports the resulting data and compares the exact forward probabilities with estimates based on 30000 continuum forward coins.

Figure 2 provides a direct calibration of both likelihood factories. At each of four fixed parameter values, we run 10000 independent likelihood coins and divide the raw success frequency by the known compensation scale. The compensated frequencies are plotted against the exact reduced likelihoods, which are available analytically in this benchmark. In all eight experiments, the exact likelihood lies within the reported two-standard-error Monte Carlo interval around the compensated coin estimate.

For posterior sampling we use $n = 128$ and $\gamma = 3000$. Under independent errors the scalar compensations result in master inflation 1.8786. For the correlated experiment we take $\Sigma = \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}$, and apply Theorem 3.5. The discrete multivariate compensation is $b_n^\Omega(y) = 0.008717$, the auxiliary normalization is $C_n^\Omega(y) = 2.6373$, and the corresponding master inflation is 4.2241. In both panels of Figure 3, retention is decided by the actual likelihood coin. For the correlated coin the reachable threshold range is propagated backward over the balanced component-reveal schedule, so the multivariate Bernstein event can also terminate before all $2n$ input coins are revealed.

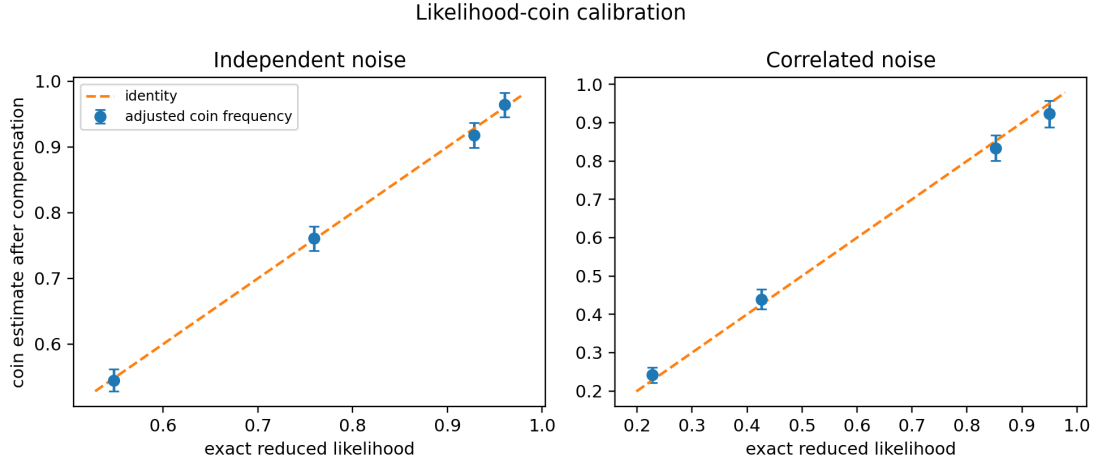


Figure 2: Likelihood-coin calibration. Points are compensated coin frequencies with two-standard-error bars; the dashed line is the identity.

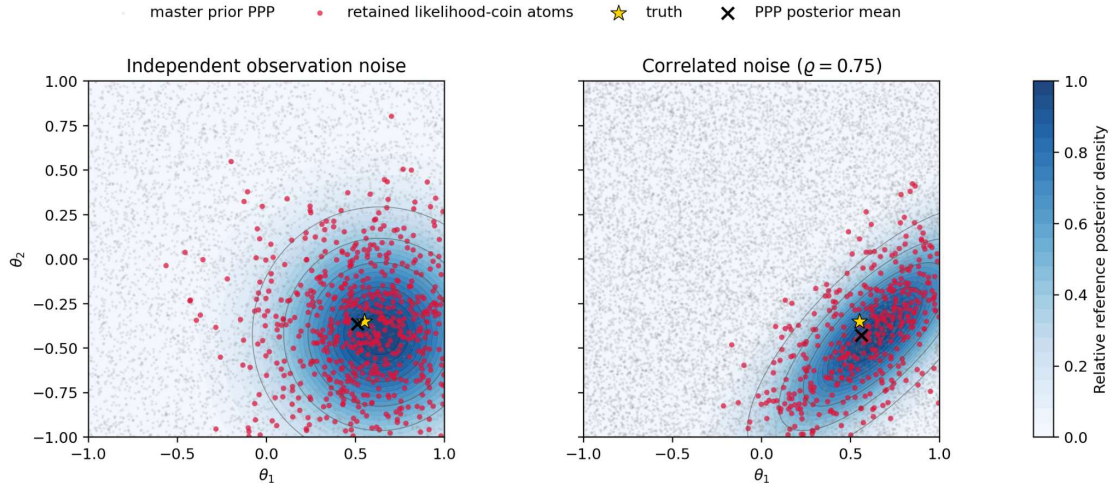


Figure 3: Posterior PPPs for independent noise (left) and correlated noise with $\rho = 0.75$ (right). Gray points are master candidates, red points are retained atoms, and background contours show the deterministic reference posterior.

Table 2: Sharp-complexity experiment: predicted rates and empirical tail exponents. Tail fits use the five largest values of a_σ . For $d_{\text{pred}} = 2$, the fitted quantity is $W/\log a_\sigma$.

d_{pred}	α	sharp scale	tail fitted exponent	predicted exponent
1	1/2	$a_\sigma^{3/2}$	1.418	1.500
2	1	$a_\sigma \log a_\sigma$	0.965	1.000
3	3/2	a_σ	1.006	1.000

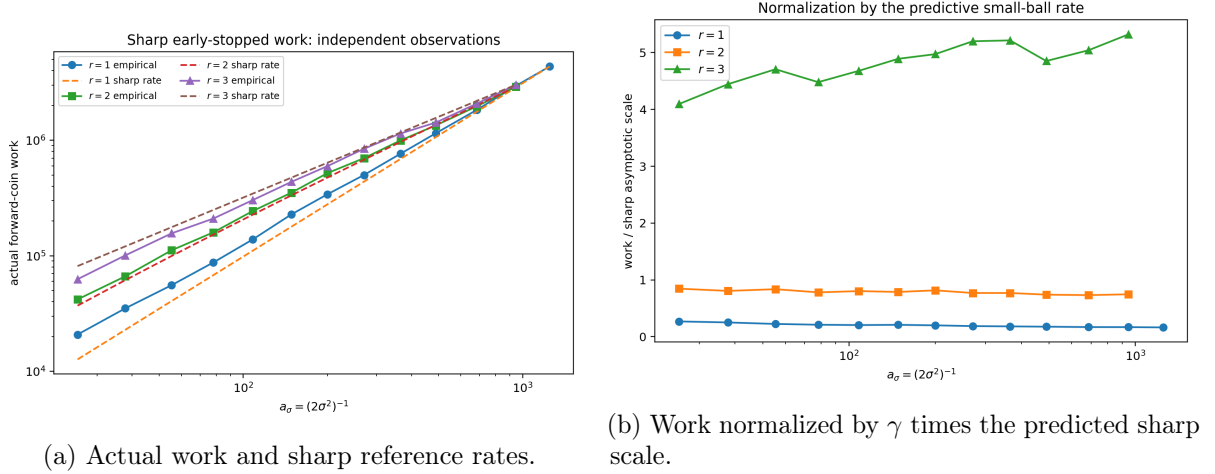


Figure 4: Sharp actual-work scaling under independent observations.

For independent noise, the reference evidence is 0.18943, while the displayed realization gives $N_\gamma/\gamma = 0.19367$. Its posterior mean, $(0.51030, -0.36354)$, is close to the quadrature value $(0.50891, -0.36276)$. For correlated noise, the reference evidence is 0.12313 and the PPP estimate is $N_\gamma/\gamma = 0.12600$; the corresponding posterior means are $(0.55225, -0.45358)$ from quadrature and $(0.56071, -0.42746)$ from the PPP. The rotation of the correlated posterior reflects the off-diagonal precision terms rather than a change in the forward model.

Thus the benchmark validates the forward oracle, likelihood factory, and posterior thinning against independently available reference quantities. We next turn to the work scaling predicted by the sharp complexity theory.

5.2 Small-ball dimension and sharp actual-work scaling

For $d_{\text{pred}} \in \{1, 2, 3\}$, take a uniform prior on $[0, 1]^{d_{\text{pred}}}$, the identity forward map, and observation vector $(1/2, \dots, 1/2)$ with independent errors of common variance σ^2 . Then $\alpha = d_{\text{pred}}/2$. The implementation uses the marked Poisson stream of Algorithm 1, as in Theorem 3.9: each event independently selects a scalar component and uses its range-stopped Bernstein event; a candidate terminates at the first rejection.

We use $\gamma = 600$ and four independent runs at each noise level. The common order follows Theorem 3.13; here $D_y = d_{\text{pred}}/4$. Figure 4 compares measured work with the sharp rates, and Table 2 reports tail fits after removing the logarithmic factor for $d_{\text{pred}} = 2$.

The empirical scaling distinguishes the three regimes predicted by the sharp work theory. For $d_{\text{pred}} = 3$, the tail exponent 1.006 is essentially the predicted linear rate. At the critical dimension $d_{\text{pred}} = 2$, removing the logarithmic factor produces a tail exponent 0.965, close to the predicted value one. Convergence is slower for $d_{\text{pred}} = 1$: the fitted tail exponent 1.418 remains below the asymptotic value $3/2$, and the normalized work retains a mild downward drift over the simulated range. The experiment measures the work required to process the master PPP rather

Table 3: Function-valued reaction-coefficient example: observation data and continuum forward-oracle validation.

sensor x_j	FD64	perturbation	y_j	coin estimate	potential-killing events
(0.5, 0.5)	0.65030	0.008	0.65830	0.65053	0.811
(0, 0.5)	0.52586	-0.006	0.51986	0.52457	0.658
(0, 0)	0.57319	0.006	0.57919	0.57190	0.722
(0.5, 0)	0.75309	-0.007	0.74609	0.75033	0.956
(0.5, 0.72)	0.68927	0.006	0.69527	0.68933	0.863
(0.5, 0.28)	0.68927	-0.005	0.68427	0.68857	0.863
(0, 0.28)	0.54389	0.004	0.54789	0.54167	0.675
(0, 0.72)	0.54389	-0.006	0.53789	0.54407	0.675

than posterior Monte Carlo accuracy; consequently, the small number of retained atoms at the smallest noise levels does not enter the work estimate. This experiment deliberately isolates the independent scalar-factor architecture for which the sharp actual-work theorem is proved.

5.3 A function-valued heterogeneous reaction coefficient

We next place the function-valued unknown directly in the elliptic operator. Let

$$q_\xi(x) = 40 \left\{ \frac{1}{2} + \sum_{\ell=1}^{\infty} \omega_\ell \xi_\ell \cos(2\pi k_\ell \cdot x) \right\}, \quad \xi_\ell \stackrel{\text{iid}}{\sim} \text{Unif}[-1, 1],$$

where $\omega_\ell = A(1 - \rho)\rho^{\ell-1}$, $A = 0.48$, $\rho = 0.5$. Then $0.8 \leq q_\xi(x) \leq 39.2$ almost surely and

$$\left(32 + q_\xi(x) - \frac{1}{2}\Delta \right) u_{q_\xi}(x) = 32, \quad x \in \mathbb{T}^2.$$

Eight observations are used with $\sigma = 0.1$, and a fixed low-mode field generates synthetic data. Table 3 reports the sensor data, continuum forward-coin estimates, and mean numbers of inspected potential-killing events per forward coin. The deterministic posterior is used as a finite-resolution validation reference; the experiment is not intended to demonstrate recovery of the full field from only eight observations.

Figure 5 compares the same continuum coin estimates with deterministic FD refinement.

The deterministic reference uses the first 24 modes; the omitted pointwise field amplitude is at most 2.86×10^{-8} . This truncation is used only for validation. The likelihood-coin sampler has no fixed coefficient cap. Each prior candidate starts with a cache of 16 iid coefficients. If the deterministic geometric tail enclosure does not resolve a field-value Bernoulli decision, fresh iid coordinates are appended one at a time from the product prior while the same auxiliary uniform decision is continued. Revelation stops as soon as the shrinking tail enclosure certifies that decision. Thus the implementation follows the infinite-series construction of Theorem 4.2: the stored prefix is a dynamically extensible cache, not a truncation of the posterior parameter. In the reported posterior run 290 new coefficients were generated on demand across 9446 master candidates, and the largest prefix required by any candidate was 28 coefficients. The average number of inspected coefficients per potential-killing event was about 1.92.

The Bernstein order is selected from $n \in \{48, 64, 80, 96, 112, 128\}$. Figure 6 shows the tradeoff between master-process inflation and coefficient-inspection work; the empirical score is minimized at $n = 80$. The compensated likelihood estimates fluctuate around the FD64 reference value 0.88665 without systematic drift in n .

A 20×20 finite-difference importance calculation based on 5000 prior functions is compared with the posterior PPP in Table 4. The PPP uses $\gamma = 3000$ and $n = 80$; 9446 master candidates produce 658 retained atoms.

Figure 7 makes the validation criterion explicit. The coefficient plot includes truth markers only for context. The field plot compares the four-mode FD-reference posterior mean directly with

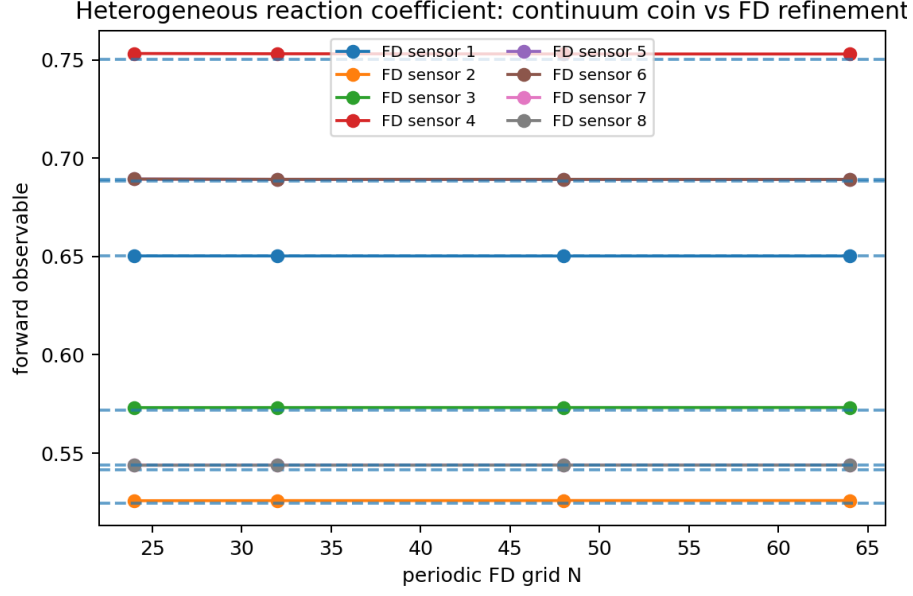
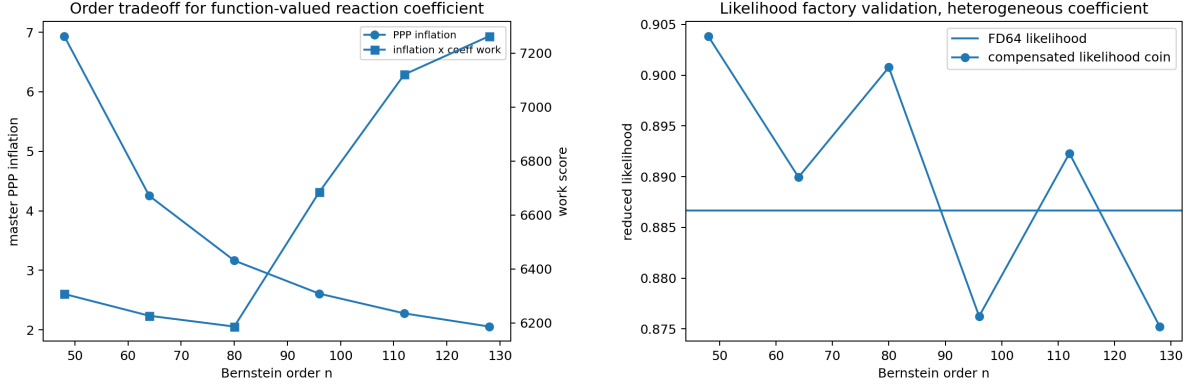


Figure 5: Continuum forward-coin estimates (dashed levels) compared with deterministic FD refinement at the eight sensors.



(a) Master inflation and empirical work score. (b) Compensated likelihood estimates across orders.

Figure 6: Bernstein-order diagnostics for the function-valued example.

Table 4: Function-valued reaction-coefficient example: posterior validation against the finite-difference importance reference.

method	Z	$E[\xi_1 y]$	$E[\xi_2 y]$	$E[\xi_3 y]$	$E[\xi_4 y]$
FD importance reference	0.216810	0.62107	-0.17511	0.00531	-0.02136
posterior PPP	0.219333	0.60203	-0.13922	-0.01602	-0.01271

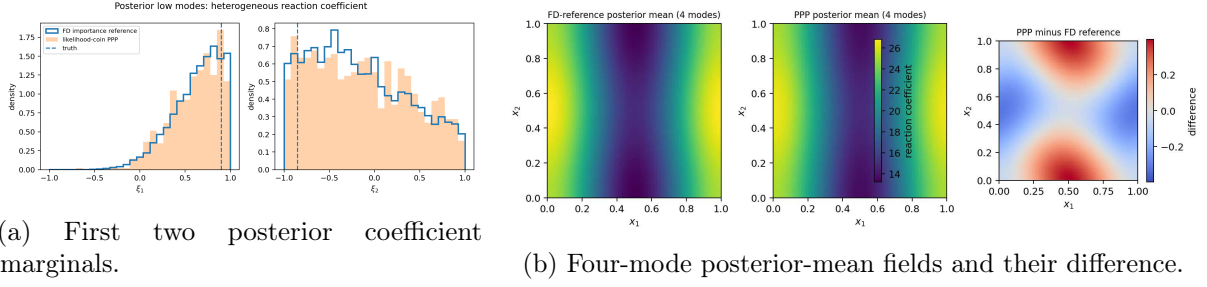


Figure 7: Posterior validation for the heterogeneous reaction coefficient.

the four-mode PPP posterior mean and their difference. The relative L^2 difference between these two projected posterior means is approximately 8.87×10^{-3} , with maximum pointwise difference about 0.396. Their discrepancy from the synthetic truth instead reflects limited identifiability under the eight-observation design.

Taken together, these results show that the continuum likelihood-coin construction is implementable for a genuinely function-valued unknown. The stochastic posterior estimates are consistent with the deterministic reference at the resolution and Monte Carlo accuracy used here, while neither a spatial PDE discretization nor a fixed truncation of the random field enters the posterior target. The next experiment isolates the computational consequence of the first distinction by comparing continuum sampling with finite-difference computation at matched posterior accuracy.

5.4 Accuracy–cost comparison with deterministic PDE rejection

To assess the computational effect of removing deterministic forward discretization, we compare the continuum likelihood-coin sampler with deterministic FD prior rejection at the *same target RMS error* for the posterior mean of the first four coefficients in the heterogeneous reaction-coefficient problem. This comparison uses the same observation locations, the same data, and bilinear periodic interpolation at every FD resolution. A 64×64 FD posterior is used as a high-resolution numerical reference for the benchmark; it is not treated as an exact continuum posterior. The benchmark uses a common pool of 5000 prior draws across FD grids and ten continuum timing replicates. Its purpose is to compare accuracy–cost mechanisms on a common implementation.

Let μ_{64} and Σ_{64} denote the posterior mean and covariance estimated from the 64×64 FD reference. For the continuum sampler, Σ_{64} is used only as a numerical proxy for the continuum posterior covariance when projecting Monte Carlo error. Thus, for m independent continuum posterior atoms, the benchmark uses

$$\widehat{\text{RMSE}}_{\text{coin}}^2(m) \approx \frac{\text{tr}(\Sigma_{64})}{m}. \quad (5.1)$$

For an FD grid N , let μ_N , Σ_N , and Z_N be the corresponding discretized posterior mean, covariance, and reduced evidence. Taking μ_{64} as the high-resolution reference mean, its projected rejection error and cost are

$$\text{RMSE}_N^2(m) \approx \|\mu_N - \mu_{64}\|_2^2 + \frac{\text{tr}(\Sigma_N)}{m}, \quad C_N(m) = m \frac{t_N}{Z_N}, \quad (5.2)$$

where t_N is the measured mean time for one deterministic PDE likelihood evaluation. Thus a grid is infeasible for a requested tolerance ε whenever its estimated discretization bias $\|\mu_N - \mu_{64}\|_2$ already exceeds ε . The factor Z_N^{-1} makes the prior-rejection penalty explicit on the deterministic side as well.

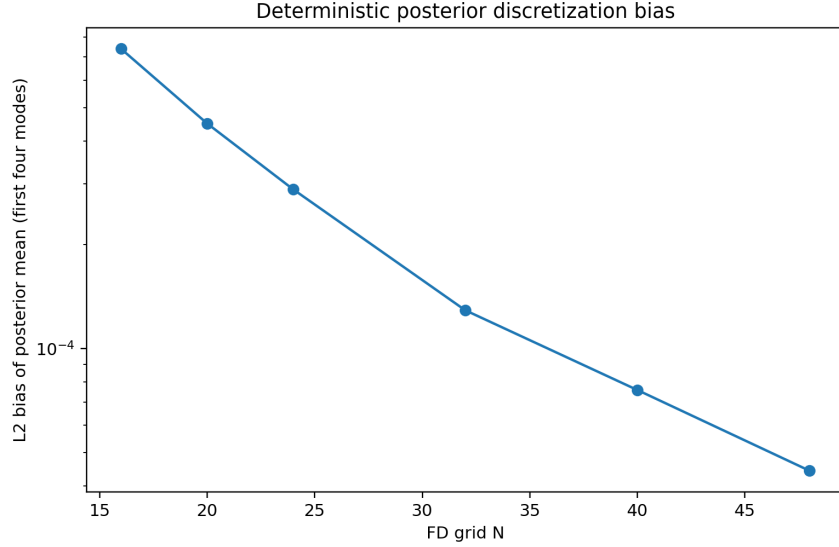


Figure 8: Estimated FD posterior-mean discretization bias relative to the 64×64 reference.

The continuum timing is measured from ten independent likelihood-coin runs after an untimed warm-up call, with each run using PPP scale $\gamma = 3000$. The elapsed wall time is pooled across runs and reported per retained atom. The deterministic benchmark uses a common pool of 5000 prior draws across grids $N \in \{16, 20, 24, 32, 40, 48, 64\}$, reducing Monte Carlo noise in differences between grid-dependent posterior means.

The enlarged common pool exhibits a monotone decrease in the estimated four-mode posterior-mean bias, from 7.39×10^{-4} at $N = 16$ to 4.4×10^{-5} at $N = 48$. Across the ten continuum timing replicates, the pooled cost is 3.376×10^{-3} seconds per retained atom; the replicate wall times have a coefficient of variation of about 1.2%. These diagnostics indicate that the qualitative accuracy–cost comparison is not driven by a single timing realization.

Figure 8 shows the deterministic bias component that was absent from a comparison based only on time per retained draw. At coarse resolution this bias is small enough for moderate accuracy targets, so inexpensive deterministic solves dominate. At tighter tolerances, however, a fixed grid ceases to be admissible regardless of how many Monte Carlo samples are drawn.

The matched-accuracy comparison in Figure 9 shows a regime-dependent tradeoff. For moderate tolerances, the inexpensive $N = 16$ FD rejection calculation is cheaper. As the target approaches the $N = 16$ discretization-bias floor, however, the number of posterior draws required by (5.2) grows rapidly even before that grid becomes infeasible. The projected cost curves cross at approximately 1.31×10^{-3} . At the tightest target shown, 10^{-3} , the cost-minimizing admissible grid has refined to $N = 20$. The continuum sampler itself introduces no spatial-discretization bias into its posterior target; the FD64 quantities enter only as numerical proxies for evaluating the benchmark. We therefore use the crossing only to identify the operating regime of this implementation, not as evidence of a universal speed advantage: its location depends on the PDE solver, implementation, hardware, prior, data, and accuracy functional.

The comparison is intentionally focused: it contrasts two independent-sampling constructions while accounting for deterministic discretization bias in the accuracy budget. It is not a benchmark against optimized multilevel, surrogate, transport, SMC, or MCMC methods. The 64×64 reference and the estimated bias/covariance quantities also carry finite-sample and residual discretization error. Within this controlled comparison, the experiment shows how the absence of a spatial-discretization bias term changes the accuracy–cost balance as the requested tolerance is tightened.

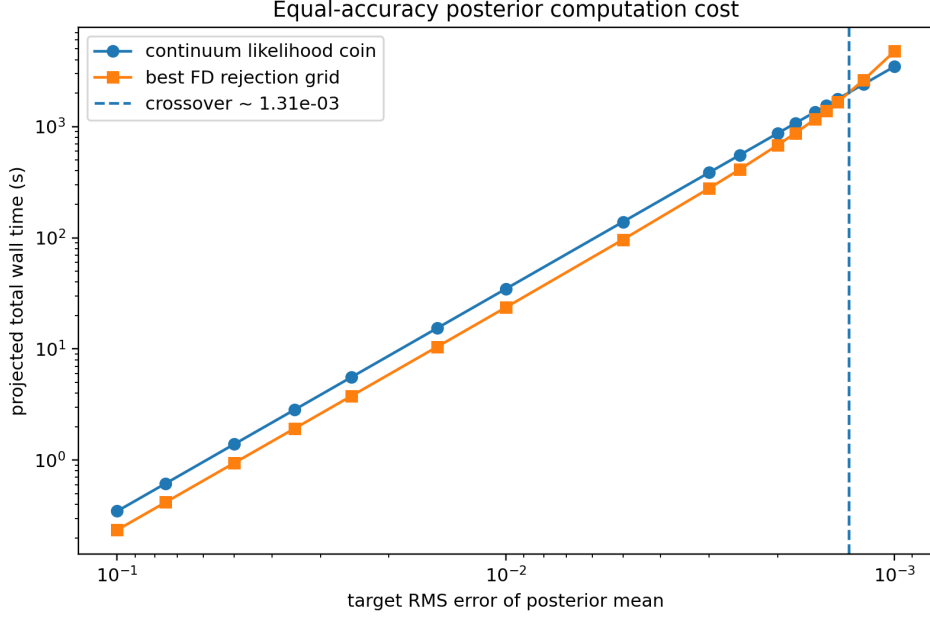


Figure 9: Projected wall time at matched posterior-mean RMS error. The dashed line marks the implementation-specific crossover near 1.31×10^{-3} .

6 Conclusion

We developed an exact Bayesian computational framework that acts directly on stochastic forward information. The key interface is Bernoulli access to selected bounded forward observables: a Bernstein–Poisson construction converts these events into scaled Gaussian likelihood coins, and Poisson thinning converts the likelihood decisions into posterior atoms that are iid conditional on the retained count. This separates the proposed construction from Bernoulli-factory methods whose primitive coins arise from regeneration, target-density bounds, or MCMC acceptance rules.

For independent Gaussian observations, exactness is accompanied by a sharp analysis of the work actually performed by the early-stopped implementation. The mean-work identity shows explicitly how sequential revelation and candidate-level rejection reduce the full-batch cost. Under a local prior-predictive small-ball law, the small-noise complexity undergoes a phase transition determined by the mass near the exact-fit set. The resulting predictive-dimension dependence provides a direct connection between posterior concentration geometry and computational work. The factorial-moment construction for correlated Gaussian errors extends the likelihood-coin mechanism, while the sharp work theory is intentionally confined to the independent architecture analyzed here.

The Feynman–Kac example demonstrates that the required forward oracle can be realized without first computing a deterministic PDE solution. For the bounded elliptic class considered here, Poisson killing and lazy random-series decisions provide Bernoulli access even for a function-valued coefficient. The numerical experiments support the calibration of the forward and likelihood coins, the predicted stopped-work regimes, and the function-valued implementation. In the matched-accuracy benchmark, coarse finite-difference rejection is cheaper at moderate tolerances, whereas deterministic posterior discretization bias becomes increasingly restrictive as the target accuracy is tightened. The observed crossover is therefore an implementation-specific illustration of the tradeoff between exact stochastic computation and deterministic discretization, not a universal efficiency claim.

The framework leaves several computational questions open. Of particular interest are likelihood models admitting practical factories beyond the Gaussian setting, stochastic forward

representations beyond the Feynman–Kac class used here, and adaptive order-selection rules that exploit information accumulated during sampling. More broadly, the small-ball analysis suggests that the usefulness of exact stochastic computation should be assessed jointly through the available probabilistic primitive and the local predictive geometry of the statistical model.

A Bernstein compensation details

For the normalized scalar factors used in Section 3, the auxiliary normalization in (3.5) is in fact inactive. Since $\beta_{k,n}(y)$ is convex in k , the maximum of $c_{k,n}(y) = \beta_{k,n}(y) - \min_j \beta_{j,n}(y)$ is attained at $k = 0$ or $k = n$. For every $k \in \{0, \dots, n\}$,

$$\beta_{0,n}(y) - \beta_{k,n}(y) = \frac{k\{2y(n-1) - (k-1)\}}{n(n-1)} \leq \frac{k(2n-k-1)}{n(n-1)} \leq 1.$$

Hence $c_{0,n}(y) \leq 1$. The symmetry $\beta_{n-k,n}(1-y) = \beta_{k,n}(y)$ implies $c_{n,n}(y) \leq 1$ as well. Therefore

$$0 \leq c_{k,n}(y) \leq 1, \quad C_n(y) = 1, \quad y \in [0, 1].$$

The factor C_n is retained in the main formulas because it makes the Bernoulli admissibility explicit and allows harmless rescalings of the scalar construction without changing the derivation.

The centered scalar factor is a useful special case of the same construction. If a particular component has $y = 0$, then $b_n(0) = 0$, and already at $n = 2$,

$$\beta_{K,2}(0) = \frac{K(K-1)}{2}.$$

This variable equals one if and only if both input p -coins are one. The Poisson zero-event transform therefore reduces to a Poisson number of independent p^2 -events and has success probability e^{-ap^2} . This simplification concerns one scalar likelihood factor; the observation model in the paper remains the vector model (3.1).

B Continuous safe shift for correlated Gaussian errors

Let $\delta_j = \Omega_{jj}/(n_j - 1)$ and

$$\tilde{\beta}_{\mathbf{n}}^{\Omega}(z; y) = (z - y)^{\top} \Omega (z - y) - \sum_{j=1}^J \delta_j z_j (1 - z_j), \quad z \in [0, 1]^J.$$

The count grid is contained in $[0, 1]^J$, so the continuous minimum is no larger than the discrete one and hence $\bar{b}_{\mathbf{n}}^{\Omega}(y) \geq b_{\mathbf{n}}^{\Omega}(y)$. Moreover,

$$\nabla^2 \tilde{\beta}_{\mathbf{n}}^{\Omega} = 2\Omega + 2 \operatorname{diag}(\delta_1, \dots, \delta_J) \succ 0,$$

so the box-constrained problem is strictly convex. Put $S_{\mathbf{n}}(z) = \sum_j \delta_j z_j (1 - z_j)$. Since $t(1-t)$ is one-Lipschitz on $[0, 1]$, $|S_{\mathbf{n}}(z) - S_{\mathbf{n}}(y)| \leq O(n_{\min}^{-1}) \|z - y\|_2$, while $(z - y)^{\top} \Omega (z - y) \geq \lambda_{\min}(\Omega) \|z - y\|_2^2$. Optimizing this quadratic bound leads to

$$\bar{b}_{\mathbf{n}}^{\Omega}(y) = \sum_{j=1}^J \frac{\Omega_{jj} y_j (1 - y_j)}{n_j} + O(n_{\min}^{-2}).$$

Evaluation at a nearest count-grid point has the same expansion for $b_{\mathbf{n}}^{\Omega}(y)$.

C Small-ball integral asymptotics

The following standard regular-variation lemma provides the analytic step used in Theorem 3.11 and Corollary 3.12; see, for example, [21].

Lemma C.1. *Let $Q : \mathbb{X} \rightarrow [0, \infty)$ be measurable and suppose $F(t) := \mu_0\{Q \leq t\} = ct^\alpha\{1 + o(1)\}$ as $t \downarrow 0$, with $c > 0$ and $\alpha > 0$. Let $w : \mathbb{X} \rightarrow [0, \infty)$ be bounded and assume that for some $w_0 > 0$,*

$$\sup_{Q(q) \leq t} |w(q) - w_0| \longrightarrow 0 \quad (t \downarrow 0).$$

If $a \rightarrow \infty$, $b_a \geq 0$, and $ab_a \rightarrow \beta \in (0, \infty)$, then

$$\int w(q) \frac{1 - e^{-a(b_a + Q(q))}}{b_a + Q(q)} \mu_0(dq) \sim \begin{cases} cw_0 \Psi_\alpha(\beta) a^{1-\alpha}, & 0 < \alpha < 1, \\ cw_0 \log a, & \alpha = 1, \end{cases}$$

where Ψ_α is defined in (3.14). If $\alpha > 1$, then Q^{-1} is locally integrable and

$$\int w(q) \frac{1 - e^{-a(b_a + Q(q))}}{b_a + Q(q)} \mu_0(dq) \longrightarrow \int \frac{w(q)}{Q(q)} \mu_0(dq) < \infty.$$

Moreover,

$$\int e^{-aQ(q)} \mu_0(dq) \sim c\Gamma(\alpha + 1)a^{-\alpha}.$$

Proof. Push μ_0 forward by Q . The assumption on w implies that the weighted sublevel measure

$$F_w(t) := \int_{\{Q \leq t\}} w(q) \mu_0(dq)$$

satisfies $F_w(t) \sim cw_0 t^\alpha$ under the stated assumption $w_0 > 0$. For $0 < \alpha < 1$, Karamata's theorem for Stieltjes transforms, applied after the rescaling $t = aQ$ yields

$$a^{\alpha-1} \int \frac{1 - e^{-a(b_a + Q)}}{b_a + Q} dF_w(Q) \longrightarrow cw_0 \alpha \int_0^\infty t^{\alpha-1} \frac{1 - e^{-(\beta+t)}}{\beta + t} dt,$$

which is the first displayed asymptotic. At $\alpha = 1$, split the integral at a fixed small δ . On $a^{-1} \ll Q \leq \delta$ the kernel is $Q^{-1}\{1 + o(1)\}$, and Stieltjes integration by parts with $F_w(t) \sim cw_0 t$ implies $cw_0 \log a\{1 + o(1)\}$; the regions $Q = O(a^{-1})$ and $Q \geq \delta$ contribute only $O(1)$.

If $\alpha > 1$, then $F(t) \rightarrow 0$ implies $\mu_0\{Q = 0\} = 0$, and Stieltjes integration by parts shows that

$$\int_{\{0 < Q \leq \delta\}} Q^{-1} \mu_0(dq) = \frac{F(\delta)}{\delta} + \int_0^\delta \frac{F(t)}{t^2} dt < \infty.$$

On $\{Q > \delta\}$ the function Q^{-1} is bounded by δ^{-1} ; hence Q^{-1} is integrable on all of $\mathbb{X}$. Since w is bounded, the kernel multiplied by w is dominated by an integrable multiple of Q^{-1} and converges pointwise to w/Q . Dominated convergence therefore establishes the high-dimensional limit. Finally, the Laplace asymptotic is the standard Karamata Laplace-transform asymptotic for a distribution function regularly varying at zero with index α . $\square$

Acknowledgments

Correspondence may be addressed to Zhiliang Deng.

Competing Interests: The authors declare that they have no competing interests.

References

- [1] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, New York, 2005.
- [2] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.
- [3] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM Journal on Numerical Analysis*, 48(1):322–345, 2010.
- [4] V. H. Hoang, J. H. Quek, and C. Schwab. Multilevel Markov Chain Monte Carlo for Bayesian inversion of parabolic partial differential equations under Gaussian prior. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):384–419, 2021.
- [5] T. Cui, Y. Marzouk, and K. Willcox. Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction. *Journal of Computational Physics*, 315:363–387, 2016.
- [6] J. Cockayne, C. J. Oates, T. J. Sullivan, and M. Girolami. Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1853 of *AIP Conference Proceedings*, page 060001, 2017.
- [7] J. Cockayne, C. J. Oates, T. J. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789, 2019.
- [8] H. C. Lie, T. J. Sullivan, and A. L. Teckentrup. Random forward models and log-likelihoods in Bayesian inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1600–1629, 2018.
- [9] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.
- [10] M. S. Keane and G. L. O’Brien. A Bernoulli factory. *ACM Transactions on Modeling and Computer Simulation*, 4(2):213–219, 1994.
- [11] S. Nacu and Y. Peres. Fast simulation of new coins from old. *The Annals of Applied Probability*, 15(1A):93–115, 2005.
- [12] K. Łatuszyński, I. Kosmidis, O. Papaspiliopoulos, and G. O. Roberts. Simulating events of unknown probabilities via reverse time martingales. *Random Structures & Algorithms*, 38(4):441–452, 2011.
- [13] M. Huber. Nearly optimal Bernoulli factories for linear functions. *Combinatorics, Probability and Computing*, 25(4):577–591, 2016.
- [14] J. M. Flegal and R. Herbei. Exact sampling for intractable probability distributions via a Bernoulli factory. *Electronic Journal of Statistics*, 6:10–37, 2012.
- [15] D. Vats, F. B. Gonçalves, K. Łatuszyński, and G. P. Roberts. Efficient Bernoulli factory Markov Chain Monte Carlo for intractable posteriors. *Biometrika*, 109(2):369–385, 2022.
- [16] T. Stumpf-Fétizon and F. B. Gonçalves. Scalable Bernoulli factories for Bayesian inference with intractable likelihoods, 2025.
- [17] J. F. C. Kingman. *Poisson Processes*. Clarendon Press, Oxford, 1993.

- [18] G. Last and M. Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017.
- [19] L. C. Evans. *An Introduction to Stochastic Differential Equations*, volume 82 of *Miscellaneous Books*. American Mathematical Society, Providence, RI, 2013.
- [20] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2nd edition, 1991.
- [21] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.