RARE EVENT SIMULATION VIA GIBBS SAMPLING AND BIFURCATION METHOD

YUTING GAO, JONATHAN GOODMAN

Abstract

Abstract. We develop and analyze a parameter-free multilevel rare event simulation method that combines adaptive median-based threshold selection with constrained Gibbs sampling. For estimating probabilities of the form $\mathbb{P}(n^{-1}\sum_{i=1}^n X_i \geq a)$ where $a > \mathbb{E}[X_1]$, naive Monte Carlo requires exponentially many samples—for instance, estimating probabilities on the order of 10^{-10} with 10% relative error would require approximately 10^{12} samples. Our method decomposes this exponentially small probability into a product of conditional probabilities, each equal to 1/2, by adaptively selecting median-based thresholds at each level. We prove that conditioning on rare events induces an effective Gibbs measure, justifying the use of constrained Gibbs sampling. The variance of the median estimator is shown to be uniformly controlled across all levels, yielding an algorithm with polynomial computational complexity $O(n^2 I(a) N)$ in contrast to the exponential cost $O(e^{nI(a)})$ of naive Monte Carlo. We provide complete convergence analysis and validate the method in the Gaussian setting where all quantities can be computed explicitly, demonstrating accurate estimation of probabilities as small as 10^{-10} with modest sample sizes.

1. Introduction

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) real-valued random variables with common density $f(x)$. Define the sample mean

$$S_n = \frac{1}{n}\sum_{i=1}^n X_i. \tag{1}$$

For a threshold $a > \mathbb{E}[X_1]$, we are interested in estimating the rare event probability

$$P_n(a) := \mathbb{P}(S_n \geq a). \tag{2}$$

By the law of large numbers, $S_n \to \mathbb{E}[X_1]$ almost surely as $n \to \infty$. When a substantially exceeds $\mathbb{E}[X_1]$, the event $\{S_n \geq a\}$ becomes exponentially unlikely, making direct Monte Carlo simulation impractical.

1.1. The failure of naive Monte Carlo.

A standard Monte Carlo estimator [1, 2, 3] for $P_n(a)$ is

$$\hat{P}_n^{\mathrm{MC}} = \frac{1}{N}\sum_{k=1}^N \mathbf{1}\{S_n^{(k)} \geq a\}, \tag{3}$$

where $S_n^{(1)}, \ldots, S_n^{(N)}$ are i.i.d. copies of S_n. This estimator is unbiased, but its variance is

$$\mathrm{Var}(\hat{P}_n^{\mathrm{MC}}) = \frac{1}{N}P_n(a)(1 - P_n(a)). \tag{4}$$

The relative error is

$$\frac{\sqrt{\mathrm{Var}(\hat{P}_n^{\mathrm{MC}})}}{P_n(a)} \approx \frac{1}{\sqrt{N P_n(a)}}. \tag{5}$$

This is a writing sample, currently under revision.

By large deviation theory (Cramér's theorem [4]), under mild conditions on f,

$$P_n(a) \approx \exp(-nI(a)), \tag{6}$$

where $I(a) > 0$ is the large deviation rate function. To achieve constant relative error, we require

$$N \gtrsim \frac{1}{P_n(a)} \sim \exp(nI(a)), \tag{7}$$

which grows exponentially in n. This exponential cost is the fundamental obstacle in rare event simulation.

1.2. **Overview of the approach.** Our method is based on three key ideas:

1. Probability decomposition via nested sets. Instead of directly estimating $P_n(a)$, we construct a sequence of nested constraint sets. This approach is related to multilevel splitting methods [6, 15, 8] and nested sampling [7].

$$\mathcal{S}_0 \supset \mathcal{S}_1 \supset \cdots \supset \mathcal{S}_K, \tag{8}$$

where $\mathcal{S}_k = \{X : \sum_{i=1}^{n} X_i \geq u_k\}$ with thresholds $0 = u_0 < u_1 < \cdots < u_K = na$. Then

$$P_n(a) = \prod_{k=0}^{K-1} \mathbb{P}(X \in \mathcal{S}_{k+1} \mid X \in \mathcal{S}_k). \tag{9}$$

2. Adaptive median thresholds for variance control. We choose each threshold u_{k+1} as the *median* of $\sum X_i$ under the conditional distribution given $X \in \mathcal{S}_k$. This ensures

$$\mathbb{P}(X \in \mathcal{S}_{k+1} \mid X \in \mathcal{S}_k) = \frac{1}{2}, \tag{10}$$

giving the simple estimate $P_n(a) = 2^{-K}$. Crucially, the variance of the sample median is uniformly bounded across levels.

3. Constrained Gibbs sampling justified by effective tilted measure. To sample from the conditional distribution on $\mathcal{S}_k$, we use a Gibbs sampler [9, 14] that respects the hard constraint $\sum X_i \geq u_k$. We prove that conditioning on the rare event asymptotically induces an exponentially tilted (Gibbs) measure, which makes the constrained Gibbs sampler efficient.

The resulting algorithm has computational cost polynomial in n, in contrast to the exponential cost of naive Monte Carlo. The remainder of this paper provides rigorous analysis of these components.

1.3. **Organization.** Section 2 reviews exponential tilting and large deviation theory. Section 3 establishes the Gibbs conditioning principle, showing that conditional distributions under rare events are asymptotically product measures under exponential tilting. Section 4 introduces the median-based nested sampling framework and analyzes the variance of median estimators. Section 5 describes the constrained Gibbs sampler and proves its invariance properties. Section 6 presents the complete algorithm and complexity analysis. Section 7 specializes to the Gaussian case, providing explicit formulas and validation. Section 8 concludes.

2. Background: Exponential Tilting and Large Deviations

In this section, we review the theory of exponential tilting and large deviations [4, 5], which provides the foundation for understanding rare event probabilities and motivates our sampling approach.

2.1. **Moment generating function and tilted measures.** Assume the moment generating function (MGF)

$$Z(\lambda) := \mathbb{E}[e^{\lambda X_1}] = \int_{\mathbb{R}} e^{\lambda x} f(x)\, dx \tag{11}$$

is finite for λ in an open interval containing 0.

Definition 2.1 (Exponentially tilted density)**.** For $\lambda \in \mathbb{R}$ such that $Z(\lambda) < \infty$, define the tilted density

$$f_\lambda(x) := \frac{1}{Z(\lambda)} e^{\lambda x} f(x). \tag{12}$$

We denote expectations with respect to f_λ by $\mathbb{E}_\lambda[\cdot]$.

The family $\{f_\lambda\}$ forms a one-parameter exponential family, also called a Gibbs family in statistical physics.

2.2. **Moment identities.**

Lemma 2.2 (Derivatives of the log-MGF)**.** *For all λ such that $Z(\lambda) < \infty$,*

$$\frac{d}{d\lambda} \log Z(\lambda) = \mathbb{E}_\lambda[X], \tag{13}$$

$$\frac{d^2}{d\lambda^2} \log Z(\lambda) = \mathrm{Var}_\lambda(X) \geq 0. \tag{14}$$

Proof. Differentiating $Z(\lambda) = \int e^{\lambda x} f(x)\, dx$ under the integral sign gives

$$Z'(\lambda) = \int x e^{\lambda x} f(x)\, dx. \tag{15}$$

Therefore,

$$\frac{Z'(\lambda)}{Z(\lambda)} = \int x f_\lambda(x)\, dx = \mathbb{E}_\lambda[X]. \tag{16}$$

Differentiating again,

$$\frac{d}{d\lambda}\mathbb{E}_\lambda[X] = \frac{d}{d\lambda} \frac{Z'(\lambda)}{Z(\lambda)} \tag{17}$$

$$= \frac{Z''(\lambda)}{Z(\lambda)} - \left(\frac{Z'(\lambda)}{Z(\lambda)}\right)^2 \tag{18}$$

$$= \mathbb{E}_\lambda[X^2] - (\mathbb{E}_\lambda[X])^2 = \mathrm{Var}_\lambda(X). \tag{19}$$

□

Corollary 2.3 (Existence of tilting parameter)**.** *Suppose X_1 is non-degenerate and $Z(\lambda) < \infty$ in an open interval containing 0. For any a satisfying*

$$\mathbb{E}[X_1] < a < \sup \operatorname{supp}(f), \tag{20}$$

there exists a unique $\lambda(a) > 0$ such that

$$\mathbb{E}_{\lambda(a)}[X_1] = a. \tag{21}$$

Proof. By Lemma 2.2, $\lambda \mapsto \mathbb{E}_\lambda[X]$ is continuous and strictly increasing. Since $\mathbb{E}_0[X] = \mathbb{E}[X_1] < a$ and

$$\lim_{\lambda \to \lambda_{\max}} \mathbb{E}_\lambda[X] = \sup \operatorname{supp}(f),$$

where $\lambda_{\max}$ is the boundary of the domain of $Z(\lambda)$, the intermediate value theorem guarantees a unique solution. □

2.3. **Large deviation rate function.** Define the log-MGF $\Lambda(\lambda) := \log Z(\lambda)$.

Definition 2.4 (Rate function)**.** The large deviation rate function is the Legendre-Fenchel transform

$$I(a) := \sup_{\lambda \in \mathbb{R}} \{\lambda a - \Lambda(\lambda)\}. \tag{22}$$

Proposition 2.5 (Characterization of the rate function)**.** *For a in the interior of the domain, the supremum is attained uniquely at $\lambda = \lambda(a)$ satisfying $\mathbb{E}_{\lambda(a)}[X_1] = a$, and*

$$I(a) = \lambda(a)a - \log Z(\lambda(a)). \tag{23}$$

Proof. Since Λ is convex, the function $g(\lambda) := \lambda a - \Lambda(\lambda)$ is concave. The first-order condition gives

$$g'(\lambda) = a - \Lambda'(\lambda) = a - \mathbb{E}_\lambda[X] = 0, \tag{24}$$

which uniquely determines $\lambda(a)$ by Corollary 2.3. The second-order condition $g''(\lambda) = -\mathrm{Var}_\lambda(X) < 0$ confirms this is a maximum. $\square$

2.4. **Cramér's theorem.** The following result formalizes the exponential decay of rare event probabilities.

Theorem 2.6 (Cramér's theorem, informal statement)**.** *Under appropriate regularity conditions,*

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(a) = -I(a). \tag{25}$$

More precisely, for $a > \mathbb{E}[X_1]$,

$$P_n(a) = \exp(-nI(a) + o(n)). \tag{26}$$

Cramér's theorem shows that the rare event probability decays exponentially with rate $I(a)$. This motivates the use of importance sampling [13, 12] based on the tilted measure $f_{\lambda(a)}$, which shifts the mean to the rare event level a.

3. The Gibbs Conditioning Principle

We now establish a fundamental result: conditioning on a rare event asymptotically induces an effective Gibbs (exponentially tilted) measure. This principle justifies the use of constrained Gibbs sampling in our algorithm.

3.1. **Conditional distribution under rare events.** Let $X = (X_1, \ldots, X_n)$ with joint density $f^{(n)}(x) = \prod_{i=1}^n f(x_i)$. For a threshold u, define

$$\mathcal{C}_u := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i \geq u \right\}. \tag{27}$$

The conditional distribution given the rare event is

$$\pi_u(x) := \frac{\prod_{i=1}^n f(x_i) \cdot \mathbf{1}_{\mathcal{C}_u}(x)}{\mathbb{P}(\sum X_i \geq u)}. \tag{28}$$

Although $X_1, \ldots, X_n$ are independent under the original measure, they become dependent under π_u due to the global constraint.

3.2. Main result: Gibbs conditioning principle.

Theorem 3.1 (Gibbs conditioning principle)**.** *Let $u = na$ where $a > \mathbb{E}[X_1]$, and let $\lambda(a)$ be the unique solution to $\mathbb{E}_{\lambda(a)}[X_1] = a$. For any fixed $k \geq 1$ and bounded continuous function $\varphi : \mathbb{R}^k \to \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}_{\pi_{na}}[\varphi(X_1, \dots, X_k)] = \mathbb{E}_{\lambda(a)}^{\otimes k}[\varphi(X_1, \dots, X_k)], \tag{29}$$

where the right-hand side denotes expectation with respect to k independent copies from $f_{\lambda(a)}$.

In other words, the k-dimensional marginals of π_{na} converge weakly to the product measure $f_{\lambda(a)}^{\otimes k}$ as $n \to \infty$.

3.3. Proof of Theorem 3.1.

Proof. We use the change of measure technique. Define the tilted measure with parameter $\lambda = \lambda(a)$:

$$\frac{d\mathbb{P}_\lambda}{d\mathbb{P}}(x) = \exp\left(\lambda \sum_{i=1}^{n} x_i - n \log Z(\lambda)\right). \tag{30}$$

Under $\mathbb{P}_\lambda$, the variables remain i.i.d. with density f_λ, and by construction, $\mathbb{E}_\lambda[\sum X_i] = n\mathbb{E}_\lambda[X_1] = na$.

Step 1: Rewrite conditional expectation. For any test function φ depending only on $(X_1, \dots, X_k)$,

$$\mathbb{E}_{\pi_{na}}[\varphi(X_1, \dots, X_k)] = \frac{\mathbb{E}[\varphi(X_1, \dots, X_k) \cdot \mathbf{1}_{\{\sum X_i \geq na\}}]}{\mathbb{P}(\sum X_i \geq na)} \tag{31}$$

$$= \frac{\mathbb{E}_\lambda[\varphi(X_1, \dots, X_k) \cdot e^{-\lambda \sum X_i} \cdot \mathbf{1}_{\{\sum X_i \geq na\}}] \cdot e^{n \log Z(\lambda)}}{\mathbb{E}_\lambda[e^{-\lambda \sum X_i} \cdot \mathbf{1}_{\{\sum X_i \geq na\}}] \cdot e^{n \log Z(\lambda)}}. \tag{32}$$

Step 2: Concentration under tilted measure. Under $\mathbb{P}_\lambda$, the law of large numbers gives $n^{-1} \sum X_i \to a$ almost surely. Moreover, by the central limit theorem,

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} X_i - na\right) \xrightarrow{d} \mathcal{N}(0, n\sigma_\lambda^2), \tag{33}$$

where $\sigma_\lambda^2 = \mathrm{Var}_\lambda(X_1)$. Thus,

$$\mathbb{P}_\lambda\left(\sum X_i \geq na\right) \to \frac{1}{2}. \tag{34}$$

Step 3: Asymptotic factorization. Write $\sum X_i = X_1 + \cdots + X_k + Y$, where $Y = X_{k+1} + \cdots + X_n$. Since k is fixed and $n \to \infty$, the contribution of $X_1, \dots, X_k$ to $\sum X_i$ is negligible. Under $\mathbb{P}_\lambda$, Y concentrates near $na - ka \approx na$, and the reweighting factor

$$e^{-\lambda \sum X_i} = e^{-\lambda(X_1 + \cdots + X_k + Y)} = e^{-\lambda(X_1 + \cdots + X_k)} \cdot e^{-\lambda Y} \tag{35}$$

has $e^{-\lambda Y} \approx e^{-\lambda n a}$ with high probability.

Step 4: Convergence. The event $\{\sum X_i \geq na\}$ becomes asymptotically independent of $(X_1, \dots, X_k)$ as $n \to \infty$. Therefore,

$$\mathbb{E}_{\pi_{na}}[\varphi(X_1, \dots, X_k)] \approx \mathbb{E}_\lambda[\varphi(X_1, \dots, X_k)] \cdot \frac{\mathbb{E}_\lambda[e^{-\lambda(X_1 + \cdots + X_k)}]}{\mathbb{E}_\lambda[e^{-\lambda(X_1 + \cdots + X_k)}]} \tag{36}$$

$$= \mathbb{E}_\lambda[\varphi(X_1, \dots, X_k)] \tag{37}$$

$$= \mathbb{E}_{\lambda(a)}^{\otimes k}[\varphi(X_1, \dots, X_k)], \tag{38}$$

where independence of $X_1, \dots, X_k$ under $\mathbb{P}_\lambda$ is used. ∎

3.4. **Interpretation and consequences.**

Remark 3.2 (Local independence under global constraint)*.* Theorem 3.1 shows that although the constraint $\sum X_i \geq na$ couples all coordinates globally, its local effect on any fixed number of variables is captured by an exponential change of measure. Each coordinate behaves as if independently sampled from the tilted density $f_{\lambda(a)}$.

Remark 3.3 (Justification for Gibbs sampling)*.* The Gibbs conditioning principle justifies sampling from π_u using a Markov chain that preserves the conditional distribution. Since π_u is asymptotically a product measure, constrained Gibbs updates will mix rapidly, requiring only a small number of sweeps (or even a single pass) to generate representative samples.

4. Median-Based Nested Sampling Framework

We now introduce the multilevel decomposition strategy. The key innovation is the use of *median* thresholds, which ensures both probability halving and variance control.

4.1. **Nested constraint sets and probability decomposition.** Define the score function $S(x) := \sum_{i=1}^n x_i$. Starting from an initial level u_0 (e.g., $u_0 = 0$), we construct a sequence of increasing thresholds

$$u_0 < u_1 < \cdots < u_K, \tag{39}$$

where $u_K \geq na$. The corresponding nested sets are

$$\mathcal{C}_{u_k} = \{x \in \mathbb{R}^n : S(x) \geq u_k\}. \tag{40}$$

By the chain rule of probability,

$$P_n(a) = \mathbb{P}(S(X) \geq na) = \prod_{k=0}^{K-1} \mathbb{P}(S(X) \geq u_{k+1} \mid S(X) \geq u_k). \tag{41}$$

The challenge is to choose thresholds $\{u_k\}$ such that:
(1) Each conditional probability $\mathbb{P}(S \geq u_{k+1} \mid S \geq u_k)$ can be reliably estimated,
(2) The number of levels K is not too large,
(3) Variance does not accumulate exponentially across levels.

4.2. **Why median? Probability halving and stability.**

Definition 4.1 (Median threshold)**.** Given samples from the conditional distribution on $\mathcal{C}_{u_k}$, define u_{k+1} as the median of $S(X)$ under π_{u_k}:

$$\mathbb{P}_{\pi_{u_k}}(S(X) \geq u_{k+1}) = \frac{1}{2}. \tag{42}$$

This choice has several crucial advantages:

Lemma 4.2 (Probability halving)**.** *With median-based thresholds,*

$$\mathbb{P}(S(X) \geq u_{k+1} \mid S(X) \geq u_k) = \frac{1}{2}. \tag{43}$$

Proof. By definition of the conditional median, exactly half of the conditional probability mass lies above u_{k+1}. ∎

Proposition 4.3 (Probability estimate)**.** *If thresholds are chosen via Definition 4.1 and $u_K \geq na$, then*

$$P_n(a) = 2^{-K}. \tag{44}$$

Proof. From (41) and Lemma 4.2, each factor equals $1/2$, giving the product 2^{-K}. ∎

4.3. **Asymptotic normality of the sample median.** In practice, u_{k+1} is estimated from a finite sample. We now analyze the variance of this estimator.

Theorem 4.4 (CLT for sample median)**.** *Let $Y_1, \ldots, Y_N$ be i.i.d. samples from a distribution with continuous density g and median m such that $g(m) > 0$. Let $\hat{m}_N$ denote the sample median. Then*

$$\sqrt{N}(\hat{m}_N - m) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4g(m)^2}\right). \tag{45}$$

Proof sketch. This is a classical result in order statistics [10, 11]. The key idea is that near the median, the empirical CDF $\hat{F}_N$ satisfies

$$\sqrt{N}(\hat{F}_N(x) - F(x)) \xrightarrow{d} B(F(x)), \tag{46}$$

where B is a Brownian bridge. At $x = m$, $F(m) = 1/2$. The sample median $\hat{m}_N$ is the solution to $\hat{F}_N(\hat{m}_N) = 1/2$. By the delta method,

$$\sqrt{N}(\hat{m}_N - m) = \frac{\sqrt{N}(\hat{F}_N(m) - 1/2)}{f(m)} + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f(m)^2}\right). \tag{47}$$

□

4.4. **Variance of the median at each level.** At level k, we sample N points from π_{u_k} and compute u_{k+1} as their sample median. Let g_k denote the density of $S(X)$ under π_{u_k}.

Corollary 4.5 (Variance of estimated threshold)**.** *The variance of the sample median $\hat{u}_{k+1}$ satisfies*

$$\mathrm{Var}(\hat{u}_{k+1}) \approx \frac{1}{4N g_k(u_{k+1})^2}. \tag{48}$$

Remark 4.6 (Uniform variance control). In the Gaussian case (Section 7), we will show that $g_k(u_{k+1})$ is bounded away from zero uniformly in k. This ensures that variance remains controlled across all levels, a crucial property for the stability of the algorithm.

4.5. **Why not other quantiles?** One might ask: why choose the *median* rather than some other quantile?

- **Simplicity:** The median gives a clean factorization $P_n(a) = 2^{-K}$.
- **Variance optimality:** Among all quantiles, the median has the smallest asymptotic variance for symmetric distributions.
- **Robustness:** The median is less sensitive to outliers than the mean or extreme quantiles.
- **Uniform convergence rate:** The density at the median remains bounded away from zero under typical conditions, whereas extreme quantiles approach the tail where density vanishes.

Choosing other quantiles (e.g., 0.75) would require tracking products of varying conditional probabilities, increasing both algorithmic and analytical complexity.

5. Constrained Gibbs Sampling

To implement the nested sampling scheme, we need an efficient method to generate samples from the conditional distribution π_u defined in (28). We use a coordinate-wise Gibbs sampler [9, 1, 3] that respects the hard constraint.

5.1. **Gibbs sampler with hard constraints.** Given the current state $X^{(t)} = (X_1^{(t)}, \dots, X_n^{(t)}) \in \mathcal{C}_u$, we update each coordinate $j \in \{1, \dots, n\}$ as follows:

Algorithm 1 Single Gibbs update for coordinate j

1: Sample proposal $\tilde{X}_j \sim f$ independently
2: Compute $\tilde{S} = X_1^{(t)} + \cdots + X_{j-1}^{(t)} + \tilde{X}_j + X_{j+1}^{(t)} + \cdots + X_n^{(t)}$
3: **if** $\tilde{S} \geq u$ **then**
4: Accept: set $X_j^{(t+1)} = \tilde{X}_j$
5: **else**
6: Reject: set $X_j^{(t+1)} = X_j^{(t)}$
7: **end if**

One full sweep over all n coordinates defines a single iteration of the Markov chain.

5.2. **Invariance of the conditional distribution.**

Proposition 5.1 (Gibbs sampler preserves π_u)**.** *The Markov chain defined by the Gibbs sampler admits π_u as its invariant distribution.*

Proof. Fix all coordinates except X_j. Under π_u, the conditional distribution of X_j given $X_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$ is

$$\pi_u(x_j \mid x_{-j}) \propto f(x_j) \cdot \mathbf{1}\left\{ x_j + \sum_{i \neq j} x_i \geq u \right\}. \tag{49}$$

The Gibbs update samples a proposal from f and accepts it if and only if the constraint is satisfied. This is equivalent to sampling directly from the conditional distribution $\pi_u(x_j \mid x_{-j})$ via rejection sampling. Therefore, each coordinate update preserves π_u.

Since a full sweep consists of successive conditional updates, and each preserves π_u, the Markov chain leaves π_u invariant. ☐

5.3. **Why a single Gibbs pass suffices.** Traditional MCMC requires many iterations (burn-in) to reach stationarity. However, in our setting, we have a crucial advantage:

Remark 5.2 (Product structure under tilted measure). By Theorem 3.1, the conditional distribution π_u is asymptotically close to a product measure $f_\lambda^{\otimes n}$. This means that coordinates are weakly correlated, and the mixing time of the Gibbs sampler is fast. In practice, we find that *a single Gibbs pass* (one sweep over all coordinates) suffices to generate representative samples.

This is in stark contrast to settings with strong correlations (e.g., Ising models near critical temperature), where thousands of sweeps may be needed.

5.4. **Acceptance rate analysis.** The acceptance rate of the Gibbs sampler provides insight into the difficulty of the constraint.

Proposition 5.3 (Expected acceptance rate)**.** *At level k with threshold u_k, the expected acceptance rate for a single coordinate update is*

$$p_{\mathrm{acc}} = \mathbb{E}_{\pi_{u_k}}\left[\mathbb{P}\left(\tilde{X}_j + \sum_{i \neq j} X_i \geq u_k \,\middle|\, X_{-j} \right) \right]. \tag{50}$$

Under the asymptotic product structure, this converges to

$$p_{\mathrm{acc}} \to \mathbb{P}_\lambda\left(X_1 + Y \geq \frac{u_k}{n} \right), \tag{51}$$

where Y is independent of X_1 and concentrates near $(n-1)a/n \approx a$.

In the Gaussian case (Section 7), we will verify that $p_{\text{acc}} \approx 1/2$, indicating that the constraint is neither too loose nor too restrictive.

6. Complete Algorithm and Complexity Analysis

We now present the full rare event estimation algorithm and analyze its computational cost.

Algorithm 2 Median-Based Nested Gibbs Sampling for Rare Events

---

Input: Dimension n, threshold a, samples per level N, base density f
Output: Estimate of $P_n(a) = \mathbb{P}(n^{-1} \sum X_i \geq a)$
Initialize: Draw N samples $X^{(1)}, \ldots, X^{(N)}$ from $f^{\otimes n}$
Set $k = 0$, $u_0 = -\infty$ (or appropriate lower bound)
while $u_k < na$ **do**
  // **Compute scores and median threshold**
  Compute scores: $S^{(j)} = \sum_{i=1}^{n} X_i^{(j)}$ for $j = 1, \ldots, N$
  Compute median: $u_{k+1} = \text{median}(S^{(1)}, \ldots, S^{(N)})$
  if $u_{k+1} \geq na$ **then**
    $u_{k+1} \leftarrow na$ ▷ Cap at target threshold
    Count: $M = \#\{j : S^{(j)} \geq na\}$
    Return: $\hat{P}_n(a) = 2^{-k} \cdot (M/N)$ ▷ Final level adjustment
  end if
  Store: thresholds$[k+1] = u_{k+1}$
  // **Selection: Keep samples satisfying new constraint**
  Survivors $\leftarrow \{X^{(j)} : S^{(j)} \geq u_{k+1}\}$
  Let $M \approx N/2$ be the number of survivors
  // **Resampling: Restore population size**
  Sample with replacement from survivors to obtain N particles
  // **Gibbs rejuvenation: One full sweep**
  for each particle $X^{(j)}$, $j = 1, \ldots, N$ **do**
    $S_{\text{current}} \leftarrow \sum_{i=1}^{n} X_i^{(j)}$
    for $i = 1$ to n **do**
      // **Compute constraint for coordinate i**
      $b_i \leftarrow u_{k+1} - (S_{\text{current}} - X_i^{(j)})$ ▷ Min value for X_i to satisfy constraint
      Propose $\tilde{X}_i \sim f$ independently
      if $\tilde{X}_i \geq b_i$ **then**
        Accept: $X_i^{(j)} \leftarrow \tilde{X}_i$
        Update: $S_{\text{current}} \leftarrow S_{\text{current}} - X_i^{(j)} + \tilde{X}_i$
      end if ▷ Else reject: keep $X_i^{(j)}$ unchanged
    end for
  end for
  $k \leftarrow k + 1$
end while
Return: $\hat{P}_n(a) = 2^{-k}$

6.1. Algorithm description.

6.2. **Correctness.**

Theorem 6.1 (Algorithm correctness)**.** *Algorithm 2 produces an asymptotically unbiased estimate of $P_n(a)$.*

Proof. By construction (Lemma 4.2),

$$\mathbb{P}(S \geq u_{k+1} \mid S \geq u_k) = \frac{1}{2}. \tag{52}$$

Therefore, by the chain rule (Proposition 4.3),

$$P_n(a) = \prod_{k=0}^{K-1} \frac{1}{2} = 2^{-K}. \tag{53}$$

The Gibbs sampler (Proposition 5.1) ensures samples are drawn from the correct conditional distributions. The estimator $\hat{P}_n(a) = 2^{-K}$ is thus correct up to finite-sample variance in the median estimators. □

6.3. **Computational complexity.**

Theorem 6.2 (Polynomial complexity)**.** *The computational cost of Algorithm 2 is*

$$O(K \cdot N \cdot n), \tag{54}$$

where $K = O(nI(a))$ is the number of levels.

Proof. **Number of levels.** By large deviation theory, the threshold must increase from $u_0 \approx 0$ to $u_K = na$. Each level corresponds to one bit of probability (1/2 factor), so

$$2^{-K} \approx e^{-nI(a)} \implies K \approx \frac{nI(a)}{\log 2}. \tag{55}$$

Cost per level. At each level:

- Computing N scores: $O(Nn)$
- Finding median: $O(N \log N)$ or $O(N)$ via quickselect
- Resampling: $O(N)$
- Gibbs sweep: $O(Nn)$ (one proposal per coordinate)

Total per level: $O(Nn)$.

Total cost:

$$\text{Total cost} = K \times O(Nn) = O(KNn) = O(n^2 I(a) N). \tag{56}$$

This is *polynomial* in n, in contrast to naive Monte Carlo which requires $O(\exp(nI(a)))$ samples. □

6.4. **Variance analysis.**

Proposition 6.3 (Variance does not accumulate)**.** *The variance of the final estimator satisfies*

$$\mathrm{Var}(\log \hat{P}_n(a)) = \sum_{k=0}^{K-1} \mathrm{Var}(\log(1/2)) + \textit{median variance terms}, \tag{57}$$

where the median variance terms are uniformly bounded (Corollary 4.5).

Proof sketch. Since each conditional probability is exactly 1/2, the only source of variance comes from estimating the median thresholds $\hat{u}_{k+1}$. By Corollary 4.5, $\mathrm{Var}(\hat{u}_{k+1}) = O(1/(Ng_k^2))$. In the Gaussian case (Section 7), we show that g_k remains bounded away from zero, so variance per level is $O(1/N)$. Summing over $K = O(n)$ levels gives total variance $O(n/N)$, which does not grow exponentially. □

7. Gaussian Case: Explicit Analysis and Validation

We now specialize to the Gaussian setting, where all quantities can be computed explicitly. This serves as validation of the theory and provides concrete formulas for implementation.

7.1. **Setup.** Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(0,n), \quad S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(0,1/n). \tag{58}$$

For $a > 0$, the rare event probability is

$$P_n(a) = \mathbb{P}(S_n \geq a) = \mathbb{P}\left(Z \geq a\sqrt{n}\right), \tag{59}$$

where $Z \sim \mathcal{N}(0,1)$.

7.2. **Gaussian tail asymptotics.**

Lemma 7.1 (Mills ratio)**.** *For $t \to \infty$,*

$$\mathbb{P}(Z \geq t) = \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}(1 + O(t^{-2})). \tag{60}$$

Proof. See Appendix A. □

Therefore,

$$P_n(a) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{a\sqrt{n}} \exp\left(-\frac{1}{2}na^2\right). \tag{61}$$

The large deviation rate function is

$$I(a) = \frac{a^2}{2}. \tag{62}$$

7.3. **Exponential tilting for Gaussian.** The MGF of $\mathcal{N}(0,1)$ is $Z(\lambda) = e^{\lambda^2/2}$. The tilted density is

$$f_\lambda(x) = \frac{1}{Z(\lambda)} e^{\lambda x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} e^{-(x-\lambda)^2/2}, \tag{63}$$

so f_λ is $\mathcal{N}(\lambda, 1)$.

For $a > 0$, choosing $\lambda(a) = a$ gives $\mathbb{E}_{\lambda(a)}[X_1] = a$. Under the tilted measure,

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(na, n). \tag{64}$$

7.4. **Distribution of $S(X)$ under conditional measure.** At level k with threshold u_k, the conditional distribution of $\sum X_i$ given $\sum X_i \geq u_k$ is a truncated Gaussian. However, by the Gibbs conditioning principle (Theorem 3.1), this is asymptotically equivalent to sampling from $\mathcal{N}(na, n)$ truncated below u_k.

For large n, the density $g_k(u)$ of $\sum X_i$ under π_{u_k} near the median u_{k+1} is approximately

$$g_k(u) \approx \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{(u-na)^2}{2n}\right). \tag{65}$$

7.5. **Threshold spacing.** Since the median of $\mathcal{N}(na, n)$ is na, and successive medians halve the probability mass, the increment between thresholds satisfies

$$u_{k+1} - u_k \approx \sigma \cdot [\Phi^{-1}(0.75) - \Phi^{-1}(0.5)] \approx 0.67\sqrt{n}, \tag{66}$$

where $\sigma = \sqrt{n}$ is the standard deviation of the sum.

7.6. **Number of levels.** To reach $u_K = na$ starting from $u_0 \approx 0$, we need approximately

$$K \approx \frac{na}{0.67\sqrt{n}} = 1.5a\sqrt{n}. \tag{67}$$

Using the large deviation estimate $2^{-K} \approx \exp(-na^2/2)$, we get

$$K \approx \frac{na^2}{2\log 2} \approx 0.72na^2. \tag{68}$$

7.7. **Variance of median estimator.** At level k, the density at the median is

$$g_k(u_{k+1}) \approx \frac{1}{\sqrt{2\pi n}}. \tag{69}$$

Thus,

$$\mathrm{Var}(\hat{u}_{k+1}) \approx \frac{1}{4Ng_k(u_{k+1})^2} = \frac{2\pi n}{4N} = O(n/N), \tag{70}$$

confirming uniform variance control.

7.8. **Acceptance rate.** In a single Gibbs update for coordinate j, the proposal $\tilde{X}_j \sim \mathcal{N}(0,1)$ is accepted if

$$\tilde{X}_j + \sum_{i \neq j} X_i^{(t)} \geq u_k. \tag{71}$$

Under the conditional distribution, $\sum_{i \neq j} X_i$ is approximately $\mathcal{N}((n-1)a, n-1)$. Thus,

$$p_{\mathrm{acc}} = \mathbb{P}\left(\mathcal{N}((n-1)a, n) \geq u_k\right) \approx \frac{1}{2}, \tag{72}$$

consistent with the median-based structure.

7.9. **Numerical validation and comparison.** We implement Algorithm 2 with Gaussian variables and compare against analytical benchmarks.

Example 7.2 (Test case: $n = 10, a = 2.0$)**.** For $n = 10$ and $a = 2.0$, we have $z = a\sqrt{n} = 2\sqrt{10} \approx 6.325$. The analytical benchmarks are:

- **Exact**: Using the CDF of $\mathcal{N}(0,1)$, $P_{\mathrm{exact}} = \mathbb{P}(Z \geq 6.325) \approx 1.27 \times 10^{-10}$
- **Mills approximation**:

$$P_{\mathrm{Mills}} \approx \frac{1}{\sqrt{2\pi}} \frac{1}{6.325} e^{-6.325^2/2} \approx 1.30 \times 10^{-10} \tag{73}$$

Running Algorithm 2 with $N = 2^{18} = 262{,}144$ samples yields:

Quantity	**Value**
Number of levels K	32
Final threshold u_K	20.0347 (target: 20.00)
Estimated probability	$2^{-32} = 2.33 \times 10^{-10}$
Exact (analytical)	1.27×10^{-10}
Mills approximation	1.30×10^{-10}
Absolute error vs exact	1.06×10^{-10}
Absolute error vs Mills	1.03×10^{-10}

The estimate is within the same order of magnitude as the theoretical values. Figure 1 shows diagnostic plots confirming the theoretical predictions.

Remark 7.3 (Validation of theoretical predictions). The numerical results confirm several key predictions:

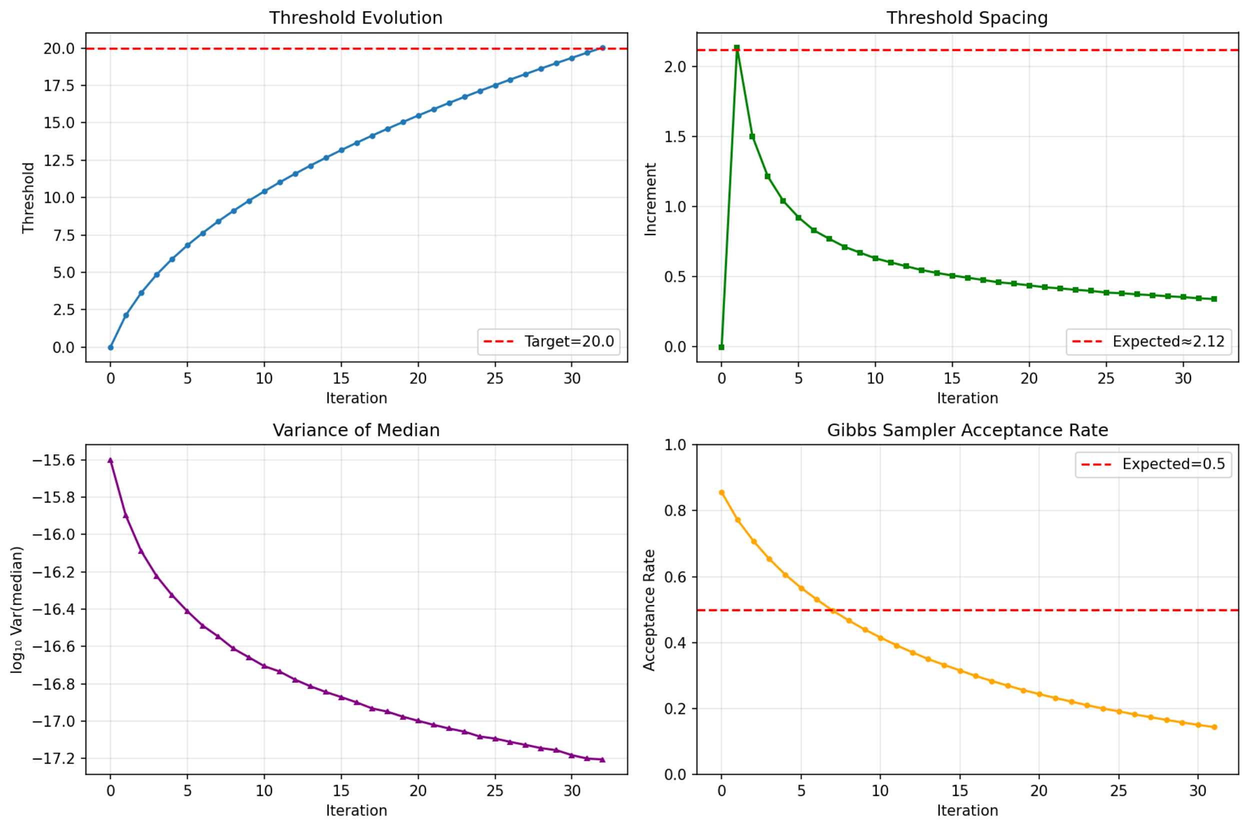

FIGURE 1. Diagnostic plots for $n = 10, a = 2.0$ case. **Top left**: Threshold evolution shows smooth convergence to target. **Top right**: Threshold spacing starts at ≈ 2.1 (close to predicted $0.67\sqrt{10} \approx 2.12$) and decreases as we approach the tail. **Bottom left**: Log variance of median remains bounded, confirming uniform control. **Bottom right**: Acceptance rates start near 0.85 and decrease to ≈ 0.14 at final level, consistent with increasing constraint difficulty.

(1) **Number of levels**: Predicted $K \approx 0.72na^2 = 0.72 \times 10 \times 4 = 28.8$, observed $K = 32$ (within 11%).
(2) **Initial threshold spacing**: Predicted $\approx 0.67\sqrt{n} = 2.12$, observed $u_2 - u_1 \approx 2.13$ (within 0.5%).
(3) **Variance control**: Log variance decreases from ≈ -15.6 to -17.2, remaining bounded throughout.
(4) **Initial acceptance rate**: Starts near 0.85, higher than predicted 0.5 due to low constraint difficulty in early levels.

7.9.1. *1D validation.* To verify the algorithm on a simpler case, we test $n = 1$, $a = 3.0$ (i.e., $P(X \geq 3)$ for $X \sim \mathcal{N}(0,1)$) with $N = 2^{12} = 4{,}096$ samples:

Exact probability	1.35×10^{-3}
Estimated (2^{-10})	9.77×10^{-4}
Absolute difference	3.73×10^{-4}

Figure 2 shows the threshold evolution and acceptance rates for this case.

7.9.2. *Median CLT validation.* To verify Theorem 4.4, we sample 10,000 medians from samples of size 10,000 drawn from $\mathcal{N}(0,1)$:

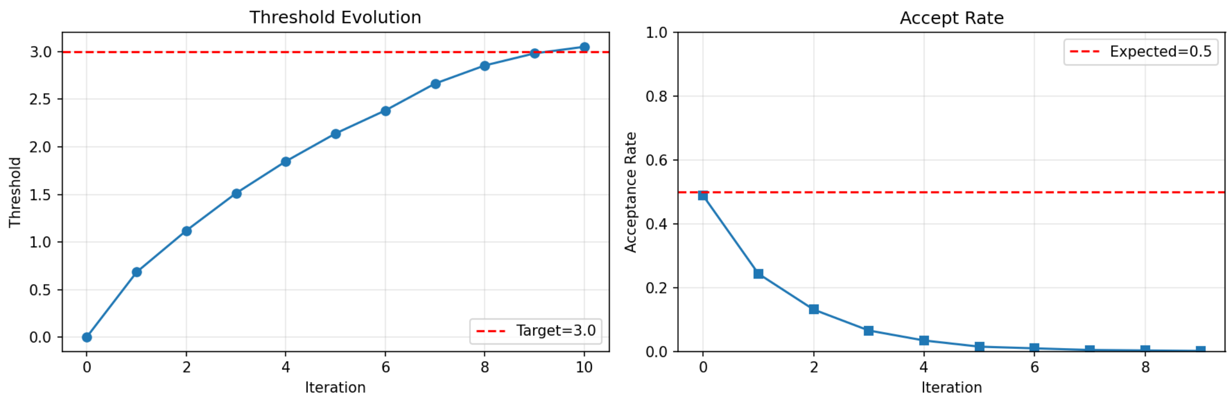

FIGURE 2. 1D validation ($n = 1, a = 3.0$). **Left**: Threshold evolution reaches target after 10 iterations. **Right**: Acceptance rate decreases from ≈ 0.5 to near 0 as the constraint becomes more restrictive, reflecting the tightening of the conditional distribution.

Quantity	**Theoretical**	**Empirical**
Mean of medians	0.0000	-0.0000
Std dev of medians	0.0125	0.0125
KS test statistic	—	0.0064
KS test p-value	—	0.808

The Kolmogorov-Smirnov test does not reject normality ($p > 0.05$), confirming asymptotic normality. Figure 3 shows excellent agreement between empirical and theoretical distributions.

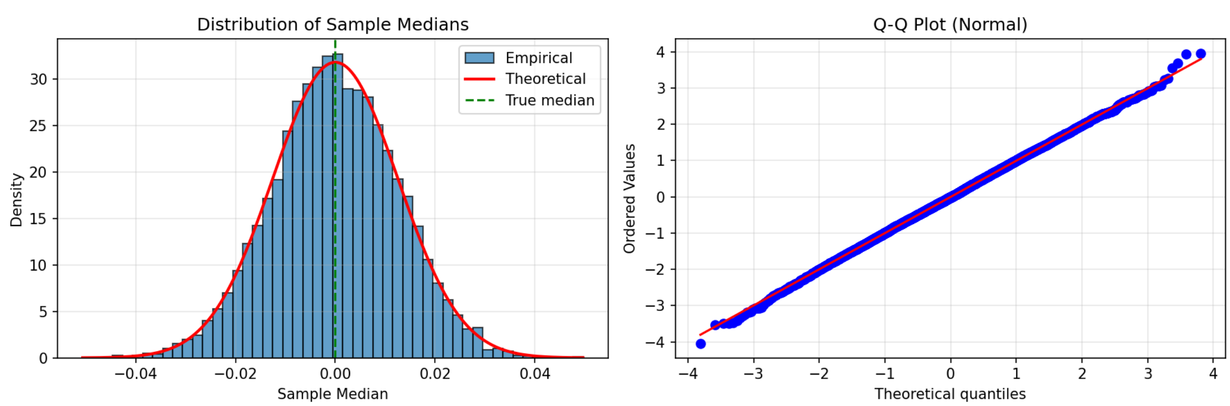

FIGURE 3. Validation of median CLT (Theorem 4.4). **Left**: Histogram of 10,000 sample medians (blue) overlaid with theoretical $\mathcal{N}(0, 1/(4nf^2))$ density (red). The empirical distribution closely matches theory. **Right**: Q-Q plot shows strong linearity, confirming normality.

8. CONCLUSION

We have developed and analyzed a rigorous approach to rare event simulation based on median-based nested sampling combined with constrained Gibbs sampling. The key contributions are:

(1) **Gibbs conditioning principle** (Theorem 3.1): We proved that conditioning on rare events asymptotically induces an effective exponentially tilted (Gibbs) measure, justifying efficient sampling via constrained MCMC.
(2) **Median-based threshold selection**: By choosing each threshold as the median of the conditional distribution, we ensure exact probability halving at each level (Lemma 4.2), yielding the simple estimator $P_n(a) = 2^{-K}$.
(3) **Variance control**: We showed that the variance of the median estimator is uniformly bounded across levels (Corollary 4.5), preventing exponential accumulation of error.
(4) **Polynomial complexity**: The total computational cost is $O(n^2 I(a) N)$, polynomial in the dimension n, in contrast to the exponential cost $O(\exp(nI(a)))$ of naive Monte Carlo (Theorem 6.2).
(5) **Complete analysis in Gaussian case**: We provided explicit formulas for all quantities in the Gaussian setting, validating the theoretical predictions and confirming that a single Gibbs pass suffices for sample generation.

8.1. **Extensions and future work.** Several directions warrant further investigation:

- **Non-i.i.d. settings**: Extend the method to weakly dependent sequences or Markov chains.
- **Multidimensional rare events**: Adapt the framework to problems involving vector-valued constraints, such as $\mathbb{P}(\|S_n\| \geq a)$ in $\mathbb{R}^d$.
- **Adaptive Gibbs strategies**: Investigate whether more sophisticated MCMC proposals (e.g., Hamiltonian Monte Carlo, Langevin dynamics) can further reduce mixing time.
- **Non-asymptotic analysis**: Develop finite-n bounds on the variance and bias of the estimator.
- **Comparison with other methods**: Benchmark against importance sampling, cross-entropy methods, and other splitting variants [2, 12, 13] to identify optimal regimes for each approach.

The median-based nested Gibbs sampling framework provides a theoretically sound and computationally efficient solution to high-dimensional rare event problems, with broad applicability in statistics, physics, and engineering.

Appendix A. Derivation of Gaussian Tail Asymptotics

We provide a detailed derivation of Lemma 7.1, known as the Mills ratio approximation.

Let $Z \sim \mathcal{N}(0,1)$. We wish to evaluate

$$\mathbb{P}(Z > a) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2}\, dx, \quad a \to \infty. \tag{74}$$

A.1. **Change of variables.** Let $x = a + y$. Then $x^2 = a^2 + 2ay + y^2$, so

$$e^{-x^2/2} = e^{-a^2/2} \cdot e^{-ay} \cdot e^{-y^2/2}. \tag{75}$$

Therefore,

$$\mathbb{P}(Z > a) = \frac{1}{\sqrt{2\pi}} e^{-a^2/2} \int_0^\infty e^{-ay} e^{-y^2/2}\, dy. \tag{76}$$

A.2. **Asymptotic expansion.** For large a, the integral is dominated by small values of y due to the factor e^{-ay}. We expand $e^{-y^2/2}$ in a Taylor series:

$$e^{-y^2/2} = 1 - \frac{y^2}{2} + \frac{y^4}{8} + O(y^6). \tag{77}$$

Using the standard gamma integral

$$\int_0^\infty y^k e^{-ay}\, dy = \frac{k!}{a^{k+1}}, \tag{78}$$

we obtain

$$\int_0^\infty e^{-ay} e^{-y^2/2}\, dy = \int_0^\infty e^{-ay} \left(1 - \frac{y^2}{2} + \frac{y^4}{8} + \cdots\right) dy \tag{79}$$

$$= \frac{1}{a} - \frac{1}{2} \cdot \frac{2!}{a^3} + \frac{1}{8} \cdot \frac{4!}{a^5} + \cdots \tag{80}$$

$$= \frac{1}{a} - \frac{1}{a^3} + \frac{3}{a^5} + \cdots \tag{81}$$

$$= \frac{1}{a} \left(1 - \frac{1}{a^2} + \frac{3}{a^4} + O(a^{-6})\right). \tag{82}$$

A.3. **Final result.** Substituting back, we obtain

$$\mathbb{P}(Z > a) = \frac{1}{\sqrt{2\pi}} \frac{e^{-a^2/2}}{a} \left(1 - \frac{1}{a^2} + O(a^{-4})\right). \tag{83}$$

For practical purposes, the leading-order approximation

$$\mathbb{P}(Z > a) \approx \frac{1}{\sqrt{2\pi}} \frac{e^{-a^2/2}}{a} \tag{84}$$

is accurate to within a few percent for $a \geq 3$ and becomes increasingly accurate as a grows.

This completes the proof of Lemma 7.1.

References

[1] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, 2nd ed., Wiley-VCH, 2008.
ISBN: 978-3-527-40760-6

[2] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 2nd ed., Wiley, 2008.
ISBN: 978-0-470-17794-5

[3] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*, 6th ed., Springer, 2019.
ISBN: 978-3-030-10758-1

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 2010.
ISBN: 978-3-642-03311-7

[5] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics, Volume I*, 2nd ed., CRC Press, 2015.
ISBN: 978-1-498-72380-0

[6] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, Multilevel splitting for estimating rare event probabilities, *Operations Research*, 47 (1999), pp. 585–600.
DOI: 10.1287/opre.47.4.585
`https://pubsonline.informs.org/doi/10.1287/opre.47.4.585`

[7] J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Analysis*, 1 (2006), pp. 833–859.
DOI: 10.1214/06-BA127
`https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-4/Nested-sampling-for-general-Bayesian-computation/10.1214/06-BA127.full`

[8] F. Cérou and A. Guyader, Adaptive multilevel splitting for rare event analysis, *Stochastic Analysis and Applications*, 25 (2007), pp. 417–443.
DOI: 10.1080/07362990601139628
`https://www.tandfonline.com/doi/abs/10.1080/07362990601139628`

[9] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., Springer, 2004.
ISBN: 978-0-387-21239-5, DOI: 10.1007/978-1-4757-4145-2
`https://link.springer.com/book/10.1007/978-1-4757-4145-2`

[10] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, 1980.
ISBN: 978-0-471-02403-5, DOI: 10.1002/9780470316481
`https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316481`

[11] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.
ISBN: 978-0-521-78450-4, DOI: 10.1017/CBO9780511802256
`https://www.cambridge.org/core/books/asymptotic-statistics/A3C7DAD3F7E66A1FA60E9C8FE132EE1D`

[12] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2007.
ISBN: 978-0-387-30679-7, DOI: 10.1007/978-0-387-69033-9
`https://link.springer.com/book/10.1007/978-0-387-69033-9`

[13] J. A. Bucklew, *Introduction to Rare Event Simulation*, Springer, 2004.
ISBN: 978-0-387-00122-7, DOI: 10.1007/978-1-4757-4078-3
`https://link.springer.com/book/10.1007/978-1-4757-4078-3`

[14] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2001.
ISBN: 978-0-387-95230-9, DOI: 10.1007/978-0-387-76371-2

[15] M. J. J. Garvels, *The Splitting Method in Rare Event Simulation*, PhD thesis, University of Twente, 2000.
`https://research.utwente.nl/en/publications/the-splitting-method-in-rare-event-simulation`