Trustworthy Monte Carlo
Supplement

Juha Harviainen
University of Helsinki
juha.harviainen@helsinki.fi

Petteri Kaski
Aalto University
petteri.kaski@aalto.fi

Mikko Koivisto
University of Helsinki
mikko.koivisto@helsinki.fi

Contents

A Proofs

A.1 Proof of Theorem 1

We use Algorithm V with a univariate polynomial p over the extension field F of $\mathbb{F}$ that satisfies $|F| = |\mathbb{F}|^\ell > 3Kd$ with ℓ as small an integer as possible.

Pick a subset $S \subset F$ of K elements. Associate each point $u^k \in \mathbb{F}^n$, $1 \leq k \leq K$, with a unique element $\xi_k \in S$; denote by u^ξ the point associated with element $\xi \in S$. For each $i = 1, 2, \ldots, n$, let $\lambda_i(x) \in F[x]$ be an interpolation polynomial satisfying $\lambda_i(\xi) := u_i^\xi$ for all $\xi \in S$. Each λ_i has degree at most K.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Finally, define $p(x)$ as the polynomial $C\big(\lambda_1(x), \lambda_2(x), \ldots, \lambda_n(x)\big)$ over F. Note that $p(\xi_k) = C(u^k)$ and the degree of p is at most Kd. Thus, if $\tilde{p} \neq p$, a random $\xi_0 \in F$ satisfies $\tilde{p}(\xi_0) \neq p(\xi_0)$ with probability at least $1 - Kd/|F| \geq 2/3$; indeed, $\tilde{p} - p \neq 0$ has at most Kd roots.

Consider the time requirement of the verifier in each step of Algorithm V. In step V1, the verifier puts $E := Kd+1$, picks $E-K$ additional elements $\xi_{K+1}, \xi_{K+2}, \ldots, \xi_E$ from F, end sents all the E elements to the prover; this takes time $\tilde{O}(E)$. In step V2, the verifier receives the claimed E values; this takes time $\tilde{O}(E)$. In step V3, the verifier computes the degree D and the coefficients $\tilde{p}_0, \tilde{p}_1, \ldots, \tilde{p}_D$ of the claimed polynomial $\tilde{P}$; this takes time $\tilde{O}(E)$ thanks to fast univariate interpolation [7]. In step V4, the verifier evaluates both $\tilde{p}$ and p at a random element $\xi_0 \in F$; the former evaluation takes time $\tilde{O}(E)$, e.g., by Horner's method. The latter evaluation takes time $\tilde{O}(Kn + s)$, since the interpolation polynomials $\lambda_i(x) \in F[x]$ can be constructed in time $\tilde{O}(Kn)$, again using fast univariate interpolation, computing the values $u_i^0 := \lambda_i(\xi_0)$ takes time $\tilde{O}(K)$ for each i, and evaluating the circuit C at the single point $(u_1^0, u_2^0, \ldots, u_n^0)$ takes time $\tilde{O}(s)$ by gate-by-gate evaluation. Step V5 is trivial. In total, the verifier requires time $\tilde{O}(E + Kn + s)$, which is $\tilde{O}(K(n+d) + s)$.

The time requirement of the prover is $\tilde{O}(Es)$, which is $\tilde{O}(Kds)$, provided that the $E - K$ points $(\lambda_i(\xi_k))_{i=1}^n \in F^n$ can be constructed sufficiently fast for all $K < k \leq E$, which can be achieved in time $\tilde{O}(En)$ using fast multipoint evaluation of univariate polynomials [7]; also observe that we can without loss of generality assume that $s = \Omega(n)$, thus leading to time $\tilde{O}(Kds)$ for the prover, since otherwise the s gates of the circuit C would not touch all of its n inputs, and we could assume a smaller n. This completes the proof of Theorem 1.

Remark. The proof above presented a prover that is centralized. To obtain a massively distributed prover, we can perform a minor change of steps V1 and V2 of Algorithm V. In step V1, rather than sending the points $\xi_1, \xi_2, \ldots, \xi_E$ to a centralized prover, the verifier first computes the vectors $\lambda(\xi_1), \lambda(\xi_2), \ldots, \lambda(\xi_E) \in F^n$. This takes time $\tilde{O}(En)$—which is $\tilde{O}(Kdn)$ and thus within the claimed time in Theorem 1—using fast multipoint evaluation of univariate polynomials [7]. Then, the verifier sends each vector $\lambda(\xi_k)$ to a prover. Each prover computes $y_k = C(\lambda(\xi_k)) = p(\xi_k)$ in time $\tilde{O}(s)$ by gate-by-gate evaluation of the circuit C, and then sends the result y_k back to the verifier. Thus, the total work of the provers is $\tilde{O}(Es)$, and there can be up to E provers doing the work in parallel, completely independently of each other, since $k = 1, 2, \ldots, E$. In step V2, the verifier receives the proof as in the centralized setting, only now the proof arrives in one or more parts from one or more distributed provers.

A.2 Proof of Lemma 4

Define

$$C_*(u) := \sum_{v \in \{0,1\}^b} C(y_t^1, y_t^2, \ldots, y_t^n)\,, \quad \text{with} \quad t := (u_1, u_2, \ldots, u_a, v_1, v_2, \ldots, v_b)\,.$$

It remains to show that C_* can be represented as an arithmetic circuit over $\mathbb{F}$ with the claimed size and degree. To this end, construct the circuit C_* from 3 layers.

The first layer maps the input to 2^b tuples $(y_t^1, y_t^2, \ldots, y_t^n)$. The bit vector t is obtained by concatenating the given u with each v. For each $j = 1, 2, \ldots, n$, there is a separate parity circuit that computes y_t^j given x^j and t as input. The size of this layer is $O(2^b \ell n)$ and the degree is ℓ.

The second layer consists of 2^b circuits C_v, each of which takes the n bits y_t^j as input and produces $C(y_t^1, y_t^2, \ldots, y_t^n)$ as output. The size and the degree of each C_v are thus s and d.

The third layer adds up the outputs of the circuits C_v, the size and the degree being $O(2^b)$ and 1.

The total size of C_* is the sum of the sizes of the three layers, yielding $O(2^b(\ell n + s))$. The degree of C_* is obtained as the product of the degrees of the three layers, yielding the degree ℓd.

A.3 Proof of Theorem 5

The verifier uses the median trick: It computes $T := \lceil 12 \ln \delta^{-1} \rceil$ independent $(\epsilon, 1/4)$-approximations of the mean and returns their median. By a standard Chernoff bound for a binomial variable with T trials and success probability $1/4$, the median is an (ϵ, δ)-approximation of the mean.

An $(\epsilon, 1/4)$-approximation is obtained by taking an average of $N \geq 4\epsilon^{-2}R$ *pairwise* independent copies of W. Namely, by Chebyshev's inquality the average deviates from the mean by more than ϵ times the mean with probability at most the critical ratio divided by $\epsilon^2 N$.

To apply Lemma 4, we may assume that N is a power of 2. To obtain an average of N pairwise copies of W, the verifier draws a tuple of bit vectors $X := (x^1, x^2, \ldots, x^n)$ uniformly at random from $\{0,1\}^{n\ell}$, with $2^\ell - 1 \geq N$, and then evaluates the circuit C_* at 2^a points (X, u), one point per $u \in \{0,1\}^a$; for a moment, let $a, b > 0$ be any integers such that $2^a 2^b = N$.

Supposing the values sent by the prover are correct, by Lemma 4 and Fact 2, their sum divided by N is an average of N pairwise independent copies of W. By Lemma 4, C_* takes $n_* := a$ elements of $\mathbb{F}$ as input and is of size $s_* = O(2^b(\ell n + s))$ and degree $d_* := \ell d$.

Now, apply Theorem 1 for $K := \lceil 12 \ln \delta^{-1} \rceil 2^a$ evaluations of the circuit C_*. We get that the prover time is $\tilde{O}(K d_* s_*)$, which is $\tilde{O}(2^a 2^b s d \log \delta^{-1})$, matching the claimed bound. The verifier time is $\tilde{O}(K d_* n_* + s_*)$, which is $\tilde{O}(2^a d \log \delta^{-1} + 2^b s)$. Note that the bounds suppress the insignificant terms $n \leq s$ and $\ell = O(\log N)$.

To optimize the asymptotic verifier time, write $d' = d \ln \delta^{-1}$ and select a and b so as to minimize $2^a d' + 2^b s$ under the constraints $2^a 2^b = N$ and $a, b \geq 0$. An elementary analysis reveals that the optimal values are $a' := (\log_2 Ns/d')/2$ and $b' := (\log_2 Nd'/s)/2$, provided that the values are nonnegative. The verifier time is then $\tilde{O}\big((Nds \log \delta^{-1})^{1/2}\big)$. If $a' < 0$, then $N < d'/s$, yielding a verifier time $\tilde{O}(d')$; If $b' < 0$, then $N < s/d'$, yielding a verifier time $\tilde{O}(s)$.

A.4 Proof of Corollary 6

Let μ and σ be the mean and the standard deviation of W. Let $\epsilon, \delta \in (0,1)$. Let $\alpha \leq \epsilon/3$ and $\beta \leq \delta\epsilon^2/\big(3456R \ln(2/\delta)\big)$. Write $W' := C_\gamma(x)$, and let A denote an acceptance event for W'.

We apply Theorem 5, with (ϵ, δ) replaced by $(\epsilon/2, \delta/2)$ and C replaced by C_γ.

Observe first that, conditionally on the event A, the random variable W' has mean μ' and standard deviation σ' that approximate μ and σ to within a relative error of $\alpha \leq \epsilon/3$ and 1, respectively. This implies that the critical ratio of W' is bounded from above by $R' := [2/(1-\epsilon/3)]^2 R \leq 9R$. Using this bound, the number of copies of W' needed for the estimator is $M := (12 \ln(2/\delta))(4(\epsilon/2)^{-2}R') = 1728R\epsilon^{-2} \ln(2/\delta)$. By the union bound, with probability at least $1 - M\beta \geq 1 - \delta/2$, the event A occurs for all these samples.

Now, since we obtain an $(\epsilon/2, \delta/2)$-approximation μ'' of μ', with probability at least $1 - \delta/2$, we have $1 - \epsilon/2 \leq \mu''/\mu' \leq 1 + \epsilon/2$, whence

$$1 - \epsilon \leq (1 - \epsilon/2)(1 - \epsilon/3) \leq \mu''/\mu \leq (1 + \epsilon/2)(1 + \epsilon/3) \leq 1 + \epsilon\,.$$

The probability of not yielding an estimate within a relative error of ϵ is at most $\delta/2 + \delta/2 = \delta$.

A.5 Proof of Lemma 7

Observe that

$$c(\psi) = |S_1 \cup S_2 \cup \cdots \cup S_m| = \sum_{j=1}^{m} |S_j \setminus (S_1 \cup S_2 \cup \cdots \cup S_{j-1})|\,.$$

Thus the mean of W equals

$$\mu = \Pr(W = 1) = \sum_{j}^{m} \frac{|S_j|}{\sum_{k=1}^{m} |S_k|} \cdot \frac{|S_j \setminus (S_1 \cup S_2 \cup \cdots \cup S_{j-1})|}{|S_j|} = \frac{c(\psi)}{\sum_{j=1}^{m} |S_j|}\,.$$

Since W is a Bernoulli random variable, its variance equals $\mu(1-\mu)$ and the critical ratio equals $(1-\mu)/\mu \leq 1/\mu$. Clearly $\mu \geq 1/m$, completing the proof.

A.6 Proof of Lemma 8

The circuit consists of three layers. First, we select some clause C_j. Then, a satisfying assignment $a \in S_j$ is generated for it. Finally, we either accept or reject the assignment.

Denote $T_k := \sum_{j=1}^{k} |S_j|$. Let $r = r(b)$ be the remainder of 2^b modulo T_m, and choose b to be the smallest nonnegative integer such that $r(b)/2^b < \beta$. Thus, the integer J represented by $b = O(n + \log \beta^{-1})$ independent uniformly distributed bits is between 0 and $T_m \cdot \lfloor 2^b/T_m \rfloor$ with probability at least $1 - \beta$. Denote $d := \lfloor 2^b/T_m \rfloor$. For each j, construct a comparison circuit for computing $B_j(J) := [\![dT_{j-1} \le J < dT_j]\!] \in \{0, 1\}$. The value is 1 for at most one index j.

Arbitrarily index the variables. Next, draw an assignment $a = (a_1, a_2, \ldots, a_n)$ from S_j using n additional random input bits $x_1, x_2, \ldots, x_n$, as follows. Let F_i be the set of indices of clauses that contain the ith variable. Additionally, let $y_{ij} \in \{0, 1\}$ be 1 if and only if the ith variable has to be true in the jth clause. Now, we let

$$a_i := \Big(1 - \sum_{j \in F_i} B_j(J)\Big) x_i + \sum_{j \in F_i} B_j(J) y_{ij} \,.$$

On the final step, we check whether a is in S_k for any $k < j$. For S_k, this can be done by comparing the variables contained by the clause against the bits of a. Thus, we get the result as

$$\sum_{j=1}^{m} B_j(J) \prod_{k=1}^{j-1} [\![a \notin S_k]\!] \,.$$

By noting that $B_j(J) = [\![J < dT_j]\!] \big(1 - [\![J < dT_{j-1}]\!]\big)$, we conclude that the size of the circuit is $O\left(m(n + \log \beta^{-1})\right)$ and the degree is $O\left(mn(n + \log \beta^{-1})\right)$.

B Circuits for gradient estimation

This section gives an arithmetic circuit that maps n independent uniformly distributed bit vectors $r_1, r_2, \ldots, r_n \in \{0, 1\}^\ell$ to an approximation of $\ln g(yx \cdot w)$, where each component w_i of w is a discretized sample from a normal distribution $\mathcal{N}(\mu_i, \sigma_i)$. Here $y \in \{-1, 1\}, x_i, \mu_i, \sigma_i$ are constants, which do not depend on the input $(r_i)_i$. We construct the circuit in several phases.

First, we give a circuit that uses the random bits in r_i to generate a discretized sample v_i from the standard normal distribution $\mathcal{N}(0, 1)$. To this end, we begin in Section B.1 by adopting von Neumann's algorithm to generate samples from the exponential distribution, and continue in Section B.2 by adopting Kahn's algorithm to generate normal variables from independent exponential variables.

Second, we give a circuit that maps the samples $v_1, v_2, \ldots, v_n$ to the dot product $z := yx \cdot w$, where $w_i = \mu_i + \sigma_i v_i$. Put otherwise, $z = \tilde{\mu} + \tilde{\sigma} \cdot v$, where $\tilde{\mu} = \sum_i y x_i \mu_i$ and $\tilde{\sigma}_i = y x_i \sigma_i$ are constants.

Finally, we give a circuit that maps z to an approximation of $\ln g(z)$.

Our constructions rely on "circuit primitives" for comparing, adding, and multiplying numbers represented as bit vectors. We give the needed results in Section C.

B.1 Sampling from the exponential distribution

For a positive integer s, denote by $\mathbb{B}_s$ the set of numbers of the form $k \cdot 2^{-s}$, where k is an integer between $-4^s + 1$ and $4^s - 1$. Elements of $\mathbb{B}_s$ are represented as bit vectors of length $2s + 1$, the highest bit being the sign bit.

Definition 5. Let F be a probability distribution function, $\epsilon > 0$, and X a random variable that takes values in $\mathbb{B}_s$. Furthermore, let $a, b \in \mathbb{B}_s$ with $a \le b$. We say that X is an *ϵ-approximate* draw from F over (a, b) if

$$F(b) - F(a) \ge 1 - \epsilon \quad \text{and} \quad \mathrm{e}^{-\epsilon} E(x) \le \Pr(X = x) \le \mathrm{e}^{\epsilon} E(x) \quad \text{for all } x \in [a, b] \cap \mathbb{B}_s \,,$$

where $E(x) := F(x + 2^{-s}) - F(x)$ is the *exact probability*.

Theorem 9. *For all $\epsilon > 0$, there is an arithmetic circuit of size and degree $O(\log^3 \epsilon^{-1})$ whose output, given independent random bits as input, is an ϵ-approximate draw from $\mathrm{Exp}(1)$ over $(0, \ln \epsilon^{-1})$.*

We prove this result in the rest of this section.

Von Neumann's algorithm. The following ingenious algorithm for generating random variables from $\mathrm{Exp}(1)$ is due to von Neumann [6]. The original algorithm assumes the availablity of an arbitrarily long sequence of draws from the uniform distribution on the real interval $[0, 1]$. We modify the algorithm by discretizing the uniform variables to s bits and by bounding their number. Let $\mathcal{U}_s(0, 1)$ denote the uniform distribution on $\{0, 1\}^s$; we interpret the bit vectors drawn from $\mathcal{U}_s(0, 1)$ as numbers in $[0, 1) \cap \mathbb{B}_s$. In what follow, we assume the parameter u is an odd integer.

Algorithm E

E1 Set $l \leftarrow 0$.

E2 Sample x from $\mathcal{U}_s(0, 1)$.

E3 Sample independent $U_1, U_2, \ldots, U_u$ from $\mathcal{U}_s(0, 1)$.

E4 Let n be the largest $n \leq u$ with $x > U_1 > U_2 > \cdots > U_n$; set $n \leftarrow 0$ if $x \leq U_1$.

E5 If n is odd, then increase l by 1 and go back to Step E2.

E6 Return $l + x$.

For an explanation of the orginal algorithm, consider the behavior of the algorithm as s and u tend to infinity. Then $\mathcal{U}_s$ tends to the uniform distribution on the interval $[0, 1]$. For a fixed $x \in [0, 1]$ and $n \geq 0$, the probability that $x > U_1 > U_2 > \cdots > U_n$ is $x^n/n!$. Thus the probability that n is the largest value with this property is $x^n/n! - x^{n+1}/(n + 1)!$. This implies that n is even with probability

$$\sum_{n \text{ even}} \frac{x^n}{n!} - \frac{x^{n+1}}{(n+1)!} = \mathrm{e}^{-x}\,.$$

When integrated over x, we have that n is even with probability $\int_0^1 1 \cdot \mathrm{e}^{-x} \mathrm{d}x = 1 - 1/\mathrm{e}$. Consequently, exactly l trials of Steps E2–E5 are performed with probability $(1 - (1 - 1/\mathrm{e}))^l \mathrm{e}^{-x} = \mathrm{e}^{-(l+x)}$. For more details and variants of the algorithm, we refer to Karney [3].

The circuit. To turn von Neumann's algorithm into an arithmetic circuit, we assume that the number of available random bits is $st(u + 1)$, i.e., the maximum of t trials (or rounds), each consuming $u + 1$ uniform bit vectors of size s. Thus we represent x using s bits and l using $\lceil \log_2 t \rceil$ bits.

For each value $l = 0, 1, \ldots, t - 1$, we have a separate subcircuit. The input of the subscircuit consist of $u + 1$ bit vectors $x, U_1, U_2, \ldots, U_u$ (we omit the dependence on l in the notation). The output is

$$y_l := [\![x \leq U_1]\!] + [\![x > U_1]\!] \sum_{\substack{1 \leq n < u \\ n \text{ even}}} [\![U_n \leq U_{n+1}]\!] \prod_{i=0}^{n-1} [\![U_i > U_{i+1}]\!]\,.$$

In other words, the output is 1 if $x > U_1 > U_2 > \cdots > U_n \leq U_{n+1}$ for some even $n < u$, and otherwise 0. Each subcircuit is of size and degree $O(us)$.

We run the t subcircuits in parallel and select the smallest valid output $l + x$, i.e.,

$$\sum_{l=0}^{t-1} \big(l + x(l)\big) \cdot [\![y_l = 1]\!] \prod_{j=0}^{l-1} [\![y_j = 0]\!]\,.$$

Here we write $x(l)$ for the value x in round l. The size and degree of the whole circuit is $O(tus)$.

Note that that the circuit outputs 0 if all t rounds fail to produce an even n.

Finite-precision analysis. To prove Theorem 11, it suffices to show that the output the circuit is an ϵ-approximate draw from the exponential distribution when $s, t, u = O(\log \epsilon^{-1})$.

Consider the following events for a single trial:

$$\begin{aligned} A_1 &: U_i \le U_{i+1} \text{ for some } 1 \le i < u\,, \\ A_2 &: U_i \ne U_j \text{ whenever } 1 \le i \ne j \le u\,, \\ B_{x,n} &: x > U_1 > U_2 > \cdots > U_n\,. \end{aligned}$$

The trial fails if either of the complement events $\bar{A}_1$ or $\bar{A}_2$ occurs. By the union bound, the probability of failing is at most

$$\Pr(\bar{A}_1) + \Pr(\bar{A}_2) = \frac{\binom{2^s}{u}}{2^{su}} + \left(1 - \frac{\binom{2^s}{u} u!}{2^{su}}\right) \le \frac{1}{u!} + \frac{u(u-1)}{2^s} =: p(s,u)\,,$$

where the inequality follows from the following bound:

Lemma 10. *Let n and k be positive integers. Denote $n^{(k)} := n(n-1)\cdots(n-k+1)$. Then*

$$\frac{n^{(k)}}{n^k} \ge 1 - \frac{k(k-1)}{n}\,.$$

Proof. We have

$$\frac{n^{(k)}}{n^k} \ge \Big(\frac{n-k+1}{n}\Big)^k = \Big(1 - \frac{k-1}{n}\Big)^k \ge 1 - \frac{k(k-1)}{n}\,,$$

where the last step is an application of Bernoulli's inequality. □

The probability that a specific $n < u$ is obtained in Step E4 for a fixed x is given by

$$\Pr(B_{x,n} \cap A_1 \cap A_2) = 1 - \Pr(\bar{B}_{x,n} \cup \bar{A}_1 \cup \bar{A}_2) \ge \Pr(B_{x,n}) - \Pr(\bar{A}_1 \cup \bar{A}_2)\,.$$

An upper and lower bound for $\Pr(B_{x,n} \cap A_1 \cap A_2)$ is now obtained by applying the bounds

$$\frac{x^n}{n!} \ge \Pr(B_{x,n}) = \frac{k^{(n)}}{2^{sn} n!} \ge \frac{x^n}{n!}\Big(1 - \frac{n(n-1)}{k}\Big)\,,$$

where in in the latter we used again Lemma 10.

The probability that an even n is obtained is bounded from above by

$$\begin{aligned} \Pr(n \text{ is even, given } x) &= \sum_{\substack{n \text{ even} \\ n<u}} \Pr(A_1 \cap A_2 \cap B_{x,n} \cap \bar{B}_{x,n+1}) \\ &\le \sum_{\substack{n \text{ even} \\ n<u}} \frac{x^n}{n!} - \left(\frac{x^{n+1}}{(n+1)!}\Big(1 - \frac{(n+1)n}{k}\Big) - p(s,u)\right) \\ &\le \sum_{n \le u} \frac{(-x)^n}{n!} + \left(\frac{x}{k} \sum_{\substack{n \text{ odd} \\ n \le u}} \frac{n(n-1)}{n!}\right) + (u-1)\cdot p(s,u) \\ &\le \mathrm{e}^{-x} + \frac{2}{2^s} + (u-1)\cdot p(s,u) \\ &\le \mathrm{e}^{-x} + u \cdot p(s,u)\,. \end{aligned}$$

Similarly, we bound the probability from below by

$$\begin{aligned} \Pr(n \text{ is even, given } x) &\ge \sum_{\substack{n \text{ even} \\ n<u}} \left(\frac{x^n}{n!}\Big(1 - \frac{n(n-1)}{k}\Big) - p(s,u)\right) - \frac{x^{n+1}}{(n+1)!} \\ &\ge \sum_{n \le u} \frac{(-x)^n}{n!} - \left(\frac{x^2}{k} \sum_{\substack{n \text{ even} \\ n \le u}} \frac{n(n-1)}{n!}\right) - (u-1)\cdot p(s,u) \\ &\ge \mathrm{e}^{-x} - \frac{2}{2^s} - (u-1)\cdot p(s,u) - \frac{1}{u!} \\ &\ge \mathrm{e}^{-x} - u \cdot p(s,u)\,. \end{aligned}$$

Let $\epsilon_0 := 2\mathrm{e}u \cdot p(s,u)$. We have $\mathrm{e}^{-x} - \epsilon_0/(2\mathrm{e}) \geq \mathrm{e}^{-x}(1-\epsilon_0/2) \geq \mathrm{e}^{-x}\mathrm{e}^{-\epsilon_0}$ and $\mathrm{e}^{-x} + \epsilon_0/(2\mathrm{e}) \leq \mathrm{e}^{-x}(1+\epsilon_0/2) \leq \mathrm{e}^{-x}\mathrm{e}^{\epsilon_0/2} \leq \mathrm{e}^{-x}\mathrm{e}^{\epsilon_0}$. Thus, by integrating over x, we get

$$\begin{aligned}
\Pr(n \text{ is even}) &\lessgtr \sum_{k=0}^{2^s-1} 2^{-s} \exp(-k2^{-s} \pm \epsilon_0) \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)} \sum_{k=0}^{2^s-1} \int_{k2^{-s}}^{(k+1)2^{-s}} \mathrm{e}^{-z}\mathrm{d}z \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)}(1-1/\mathrm{e})\,,
\end{aligned}$$

where $\epsilon_1 := 2^{-s}$; here we use the shorthands $\lessgtr$ and $\pm$ to give a symmetric upper and lower bound. Now, for any integer $0 \leq l < t$ and $x \in (0,1) \cap \mathbb{B}_s$ we have

$$\begin{aligned}
\Pr(\text{Algorithm E outputs } l+x) &\lessgtr \big(1 - \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)}(1-1/\mathrm{e})\big)^l \mathrm{e}^{-x\pm\epsilon_0} 2^{-s} \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)l}\big(1-(1-1/\mathrm{e})\big)^l \mathrm{e}^{-x\pm\epsilon_0} 2^{-s} \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)l}\mathrm{e}^{-l}\mathrm{e}^{-x\pm\epsilon_0} 2^{-s} \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)l\pm\epsilon_0}\mathrm{e}^{-(l+x)} 2^{-s} \\
&\lessgtr \mathrm{e}^{\pm(\epsilon_0+\epsilon_1)(l+1)} E(l+x)\,.
\end{aligned}$$

It remains to select the parameters s, t, u such that $(\epsilon_0+\epsilon_1)t \leq \epsilon$ and $\mathrm{e}^{-t} \leq \epsilon$. We put $t := \lceil \ln \epsilon^{-1} \rceil$ to satisfy the latter. To satisfy the former, we let $u := t$ and $s := 3t$. Assuming $\ln t \geq 3$, we have

$$(\epsilon_0+\epsilon_1)t \leq (2\mathrm{e})\Big(\frac{ut}{u!} + \frac{u^3 t}{2^s}\Big) \leq 6\Big(t^2\mathrm{e}^{-t\ln t + t} + t^4 2^{-3t}\mathrm{e}^t \cdot \mathrm{e}^{-t}\Big) \leq 6\big(t^2\mathrm{e}^{-t} + t^4(8/\mathrm{e})^{-t}\big)\mathrm{e}^{-t}\,.$$

One can verify that this is less than $\mathrm{e}^{-t} \leq \epsilon$ when $t \geq 15$, which is implied by $\ln t \geq 3$.

This completes the proof of Theorem 11.

B.2 Sampling from the normal distribution

Theorem 11. *For all $\epsilon \in (0, 1/\mathrm{e})$, there is an arithmetic circuit of size and degree $\ln^{O(1)}\epsilon^{-1}$ whose output, given independent random bits as input, is an ϵ-approximate draw from $\mathcal{N}(0,1)$ over $(\ln\epsilon, \ln\epsilon^{-1})$.*

We prove this result in the rest of this section.

Kahn's algorithm. Karney [3] attributes the following algorithm to Herman Kahn. In our modification, we let the exponential random variables be ϵ-approximate draws; we denote this discrete distribution by $\mathrm{Exp}(1;\epsilon)$.

Algorithm N

N1 Sample independent Y and Z from $\mathrm{Exp}(1;\epsilon)$ using Algorithm E.

N2 If $2Z \leq (Y-1)^2$, go back to Step N1.

N3 Sample X uniformly at random from $\{-Y, Y\}$.

N4 Return X.

For an explanation of the orginal algorithm, consider the behavior of the algorithm as ϵ tends to 0. Then Y and Z are exact draws from $\mathrm{Exp}(1)$. The probability that Steps N1 and N2 result in a value

$Y < v$ is given by

$$\begin{aligned}\Pr\left(Y < v \text{ and } 2Z > (Y-1)^2\right) &= \int_0^v \int_0^\infty \mathrm{e}^{-y}\mathrm{e}^{-z}\,[\![2z > (y-1)^2]\!]\,\mathrm{d}z\,\mathrm{d}y \\ &= \int_0^v \mathrm{e}^{-y}\Big(-\mathrm{e}^{-z}\big|_{(y-1)^2/2}^{\infty}\Big)\mathrm{d}y \\ &= \int_0^v \mathrm{e}^{-y}\mathrm{e}^{-(y-1)^2/2}\mathrm{d}y \\ &= \mathrm{e}^{-1/2}\int_0^v \mathrm{e}^{-y^2/2}\mathrm{d}y\,.\end{aligned}$$

Thus the marginal probability of proceeding to Step N3 is $\mathrm{e}^{-1/2}\sqrt{\pi/2} = \sqrt{\pi/(2\mathrm{e})}$. Conditionally on that, the probability of $Y < v$ is $2 \cdot \int_0^v (2\pi)^{-1/2}\mathrm{e}^{-y^2/2}\mathrm{d}y$, as desired. The expected number of trials (Steps N1 and N2) needed is $\sqrt{2\mathrm{e}/\pi} \approx 1.3$. For more details and variants of the algorithm, we refer to Karney [3].

The circuit. To turn Kahn's algorithm into an arithmetic circuit, we assume that the inputs consists of t triplets $(Y_l, Z_l, R_l) \in \mathbb{B}_s \times \mathbb{B}_s \times \{0,1\}$, $l = 1, 2, \ldots t$, where $s = O\log\epsilon^{-1})$ is specified by Algorithm E.

For each value $l = 1, 2, \ldots, t$, we have a separate subcircuit. The input of the subcircuit is the triplet (Y_l, Z_l, R_l). The output is

$$X_l := (1 - 2R_l) \cdot Y_l \cdot [\![2Z_l > (Y_l - 1)^2]\!]\,.$$

We run the t subcircuits in parallel and select X_l with the smallest l such that $X_l \neq 0$:

$$\sum_{l=1}^{t} X_l \cdot [\![X_l \neq 0]\!] \prod_{j=1}^{l-1} [\![X_l = 0]\!]\,.$$

Note that that the circuit outputs 0 if all t rounds satisfy the if-condition in Step N2.

A bound for the degree of each subcircuit is obtained as $O(s^{6.13})$, since subtraction contributes a factor of $O(s)$, multiplication contributes another factor of $O(s^{4.13})$ by Lemma 17, and comparison contributes yet another factor of $O(s)$. After taking a product of the results of the t comparisons, we obtain the bound $O(ts^{6.13})$ for the degree of the whole circuit.

The size of the circuit is dominated by the sizes of the subcircuits for the multiplications $(Y_l - 1)^2$. By Lemma 17, the size of each subcircuit is $O(s^{3.13})$. The total size is thus $O(ts^{3.13})$.

Finally, we combine the constructed circuit with one for the exponential distribution. We get that the combined circuit has size $O(ts^{3.13})$ and degree $O(ts^{9.13})$. We show next that $t = O(s)$ is sufficient for obtaining the claimed approximation guararantees.

Finite-precision analysis. Consider one round of Algorithm N. We drop the index l and, by Theorem 11 and its proof, assume that Y and Z are independent ϵ-approximate draws from $\mathrm{Exp}(1)$ over $(0, b)$, with $b = \lceil \ln \epsilon^{-1} \rceil$, taking values in $\mathbb{B}_s \cap [0, b)$ with $s = 3b$. We set the number of trials to $t = b$.

Let $y \in \mathbb{B}_s \cap (0, b')$ where $b' := b^{1/2}$. Put $a := (y-1)^2/2$. We have

$$\Pr(Y = y, Z \in [a, b)) = \sum_{z \in [a,b) \cap \mathbb{B}_s} \Pr(Y = y)\Pr(Z = z) \lesseqgtr \sum_{z \in [a,b) \cap \mathbb{B}_s} \mathrm{e}^{\pm 2\epsilon} E(y)E(z)\,,$$

where E is the exact probability for the exponential distribution.

Our goal is to show (i) that a successful round results in an approximate sample,

$$\Pr(Y = y \,|\, Y \in (0, b)), Z \in [a, b)) \quad \lesseqgtr \quad \mathrm{e}^{\pm\epsilon_0} E_0(y)\,,$$

where E_0 is the exact probability for the standard one-sided normal distribution, and $\epsilon_0 = O(\sqrt{\epsilon})$, and (ii) that each round has a good success probability,

$$\Pr(Y \in (0, b)), Z \in [a, b)) \quad \geq \quad 1 - 1/\mathrm{e}\,,$$

when $\epsilon_0 \in (0, 1/\mathrm{e})$, so that all t trials fail with probability at most $\mathrm{e}^{-t} \le \epsilon$. This is sufficient for the theorem (with ϵ replaced by ϵ_0), since the size and degree of the circuit are polylogarithmic in ϵ^{-1}

We have

$$\sum_{z \in [a,b) \cap \mathbb{B}_\ell} E(z) = F(b) - F(a) = \mathrm{e}^{-a} - \mathrm{e}^{-b}\,.$$

Observe that $\mathrm{e}^{-a} - \mathrm{e}^{-b} \ge \mathrm{e}^{-a}(1-\epsilon_1) \ge \mathrm{e}^{-a}\mathrm{e}^{-2\epsilon_1}$, where $\epsilon_1 := \mathrm{e}^{-b/2} \le \sqrt{\epsilon}$; this follows since $b \ge 2a$ (by our choice of b') and $1 - x \ge \mathrm{e}^{-2x}$ for all $0 \le x < 1/2$.

This gives us

$$\begin{aligned}
\Pr(Y = y, Z \in [a,b)) &\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1)} E(y)\,\mathrm{e}^{-a} \\
&\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1)\pm\epsilon_2}\mathrm{e}^{-a-y}2^{-s} \\
&\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1)\pm\epsilon_2}\mathrm{e}^{-1/2}\mathrm{e}^{-y^2/2}2^{-s} \\
&\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1+\epsilon_2)}\mathrm{e}^{-1/2} E_0(y)\sqrt{\frac{\pi}{2}}\,,
\end{aligned}$$

where $\epsilon_2 := 2^{-s} \le \epsilon^3$. Likewise

$$\begin{aligned}
\Pr\big(Y \in (0,b'), Z \in [a,b)\big) &\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1+\epsilon_2)}\mathrm{e}^{-1/2}\int_{2^{-s}}^{b'} \mathrm{e}^{-y^2/2}\mathrm{d}y \\
&\lesseqgtr \mathrm{e}^{\pm 2(\epsilon+\epsilon_1+\epsilon_2+\epsilon_3)}\mathrm{e}^{-1/2}\sqrt{\frac{\pi}{2}}\,.
\end{aligned}$$

where $\epsilon_3 := 2\sqrt{\epsilon}$; here we used $b' \ge 1$ and the bound

$$\int_{2^{-s}}^{b'} \mathrm{e}^{-y^2/2}\mathrm{d}y \ge \int_0^{\infty} \mathrm{e}^{-y^2/2}\mathrm{d}y - 2^{-s} - \mathrm{e}^{-b/2} \ge \sqrt{\frac{\pi}{2}}\Big(1 - 2\sqrt{\epsilon}\Big) \ge \sqrt{\frac{\pi}{2}}\mathrm{e}^{-4\sqrt{\epsilon}}$$

Now, taking the ratio yields the desired result for the conditional probability, as long as we put $\epsilon_0 \ge 4(\epsilon + \epsilon_1 + \epsilon_2 + \epsilon_3)$. In fact, it is convenient to put $\epsilon_0 := 6(\epsilon + \epsilon_1 + \epsilon_2 + \epsilon_3)$, so that we can bound the success probability of one round, as follows. Observe that $\mathrm{e}^{-\epsilon_0/3} \ge 1 - 1/(3\mathrm{e})$ and $\sqrt{\pi/(2\mathrm{e})} \ge 1 - 2/(3\mathrm{e})$, implying $\Pr\big(Y \in (0,b'), Z \in [a,b)\big) \ge (1 - 1/3\mathrm{e})(1 - 2/3\mathrm{e}) \ge 1 - 1/\mathrm{e}$.

B.3 Computing the dot product

Now, assume we have generated independent ϵ-approximate draws $v_1, v_2, \ldots, v_n \in \mathbb{B}_s$ from $\mathcal{N}(0,1)$, as described in the previous section. Recall that $s = 3\lceil \ln \epsilon^{-1} \rceil$. We will assume that $n < 4s$; if this does not hold, we increase s accordingly.

Our aim is to compute the dot product $z := yx \cdot w$, where $w_i = \mu_i + \sigma_i v_i$ so that w_i follows approximately a discretized normal distribution $\mathcal{N}(\mu_i, \sigma_i^2)$. We can simplify by writing $z = \tilde{\mu} + \tilde{\sigma} \cdot v$, where the sum $\tilde{\mu} = \sum_i yx_i\mu_i$ and the components $\tilde{\sigma}_i = yx_i\sigma_i$ are constants. We assume that these constants are encoded as elements of $\mathbb{B}_s$.

Our arithmetic circuit computes the dot product exactly. Each multiplication $\tilde{\sigma}_i v_i$ is computed by a multiplication circuit and the result is represented as an element of $\mathbb{B}_{2s}$. By Lemma 17, each of the n circuits is of size $O(s^{3.13})$ and degree $O(s^{4.13})$. The sum of the n products, plus the constant $\tilde{\mu}$, is represented as an element of $\mathbb{B}_{3s}$. Since $n < 4s$, we can compute it by a circuit for adding $4s$ numbers, each represented using $4s$ bits. By the proof of Lemma 17, this circuit of size $O(s^{3.13})$ and degree $O(s^{4.13})$. Combining the two layers yields a circuit of size $O(ns^{3.13})$ and degree $O(s^{6.26})$.

Since each v_i converges in distribution to a standard normal random variable as ϵ tends to 0, the dot product z converges in distribution to a random variable that follows $\mathcal{N}\big(\tilde{\mu}, \sum_{i=1}^n \tilde{\sigma}_i^2\big)$. Clarly, the mean of z is $\tilde{\mu}$ (with no error) and the variance of z is $\sum_{i=1}^n \tilde{\sigma}_i^2 c_i$, where c_i is the variance of v_i. An analysis, similar to those in Section B.2, shows that $c_i \lesseqgtr \mathrm{e}^{\pm O(\epsilon)}$, i.e., $|1 - c_i| \le O(\epsilon)$, thus guaranteeing that the variance of z is has a relative error of $O(\epsilon)$.

B.4 Approximating the log-sigmoid

The *sigmoid* is the function $\sigma : \mathbb{R} \to \mathbb{R}$ given by $\sigma(x) := 1/(1 + \mathrm{e}^{-x})$. The *log-sigmoid* is the function $\lambda : \mathbb{R} \to \mathbb{R}$ given by $\lambda(x) := \ln \sigma(x)$.

Lemma 12. *For all $n = 1, 2, \ldots$, the log-sigmoid λ is n times continuously differentiable in $\mathbb{R}$ and $\|\lambda^{(n)}\|_\infty \leq (n-1)!$.*

Proof. We have $\lambda^{(1)}(x) = \mathrm{e}^{-x}/(1 + \mathrm{e}^{-x}) = 1 - \sigma(x)$. Thus, $\lambda^{(n)} = -\sigma^{(n-1)}$. Since $\sigma^{(1)} = \sigma(1-\sigma)$, we have $\sigma^{(k+1)} = \big(\sigma(1-\sigma)\big)^{(k)} = \sum_{i=0}^{k} \binom{k}{i} \sigma^{(i)} \sigma^{(k-i)}$. As $\sigma(x) \leq 1$ for all $x \in \mathbb{R}$, by induction, $\|\sigma^{(k)}\|_\infty \leq k!$. □

Definition 6. The *natural cubic spline interpolant* of a function f at points $x_0 < x_1 < \ldots < x_n$ is a 2 times continuously differentiable function s that is a degree-3 polynomial in each $[x_{i-1}, x_i]$ for $i = 1, 2, \ldots, n$, satisfying $s(x_i) = f(x_i)$ for $i = 0, 1, \ldots, n$, and $s'(x_j) = f'(x_j)$ for $j = 0, n$.

Lemma 13. *Let s be a natural cubic spline interpolant of λ with $n+1$ points $x_0 < x_1 < \cdots < x_n$ with spacing length ϵ. Furthermore, assume that $|x_0|, |x_n| = O(\log \epsilon^{-1})$. Then, the coefficients of each polynomial are of size $O(\epsilon^{-2})$.*

Proof. We can write the polynomial for the interval $[x_i, x_{i+1}]$ as

$$f_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i.$$

First, note that $\lambda(x_i) = f_i(x_i) = d_i$. Since the absolute value of the derivative of λ never exceeds 1 and λ is increasing, we have $|\lambda(x_i)| = O(\log \epsilon^{-1})$,

$$\lambda(x_{i+1}) - \lambda(x_i) < \epsilon$$

and

$$|\lambda(x_i) - 2\lambda(x_{i+1}) + \lambda(x_{i+2})| < \epsilon.$$

These inequalities are used to bound the coefficients of the polynomials.

Coefficients b_i can be solved from a matrix equation, and the other coefficients can be written as a function of b_i's, d_i's and ϵ (see, for example, [4] for details). With some effort one can see that

$$|a_i| = O(\epsilon^{-2}), \quad |b_i| = O(\epsilon^{-1}), \quad |c_i| = O(1), \quad \text{and} \quad |d_i| = O(\log \epsilon^{-1}).$$

Thus, it follows that the coefficients of f_i are of magnitude $O(\epsilon^{-2})$ as well. □

Theorem 14 ([1, 2])**.** *Let $f : [a, b] \to \mathbb{R}$ be 4 times continuously differentiable in $[a, b]$. Let s be the natural cubic spline interpolant of f at $n+1$ evenly spaced points from a to b. Then*

$$\|f - s\|_\infty \leq \frac{5}{384} \|f^{(4)}\|_\infty n^{-4}.$$

Theorem 15 (Approximating circuit for the log-sigmoid)**.** *For any $\epsilon > 0$, there exists an arithmetic circuit $C : \{0,1\}^{2m+1} \to \mathbb{F}$ of size $O(\epsilon^{-1/4} \log \epsilon^{-1})$ and degree $O(\log \epsilon^{-1})$ such that*

$$\big|\lambda(z) - C(z2^m) \cdot 2^{-M}\big| \leq \epsilon \qquad \text{for all} \quad z \in \big\{r \cdot 2^{-m}, -r \cdot 2^{-m} : r = 0, 1, \ldots, 4^m - 1\big\},$$

where $m = O(\log \epsilon^{-1})$ and $M = O(\log \epsilon^{-1})$.

Proof. Assume C takes as input a rational number of the form $\pm r \cdot 2^{-m}$, where $r = 0, 1, \ldots, 4^m - 1$ represented by $2m + 1$ bits. Observe that it suffices to have $2^{-m} \leq \epsilon/3$, or, $m = O(\log \epsilon^{-1})$, since the derivative of λ is everywhere positiv and at most 1—higher resolution of the input numbers is not needed for the target accuracy.

Small and large arguments x are not challenging for the approximation of $\lambda(x)$. Indeed, let $b \geq 1$ be the smallest multiple of 2^{-m} such that $\lambda(b) > -\epsilon/3$ and $x - \lambda(x) < \epsilon/3$ for all $x < -b$. Denote $a := -b$. One finds that $b - a = 2b = O(\log \epsilon^{-1})$. Outside the interval $[a, b]$ we now have simple approximations to λ, namely, the functions x and 0.

It remains to approximate λ inside $[a, b]$. This is accomplished by using a natural cubic spline s and n_ϵ points with spacing length $O(\epsilon^{1/4})$. It follows from Lemma 12 and Theorem 14 that this suffices for absolute error $\epsilon/2$ in $[a, b]$.

We round each coefficient c of the spline polynomials to a rational number r/q with integers r and q; we take a common $q := 2^\ell$ for all these coefficients, thus the absolute error of r/q to c is at most $2^{-\ell}$. By Lemma 13, $O(\log \epsilon^{-1})$ bits are enough for representing the coefficients of the polynomials. Denote the rounded cubic spline by $\tilde{s}$. We get that $\left|s(x) - \tilde{s}(x)\right| \leq 4b^3 2^{-\ell}$, which is at most $\epsilon/2$ when we put $\ell := \lceil \log_2(8b^3\epsilon^{-1}) \rceil = O(\log \epsilon^{-1})$.

For evaluations of $\tilde{s}$ at rational points $x := r \cdot 2^{-m}$ with a varying r, let C be the (piecewise) polynomial of degree 3 where in each piece the kth coefficient is obtained by multiplying the kth coefficient of $\tilde{s}$ by $2^\ell 2^{(3-k)m}$. Observe that C has integer coefficients and $\tilde{s}(r \cdot 2^{-m}) = C(r) \cdot 2^{-\ell-3m}$. Therefore, using the triangle inequality, we conclude that $\left|\lambda(r \cdot 2^{-m}) - C(r) \cdot 2^{-\ell-3m}\right| \leq \epsilon$ for all $r \cdot 2^{-m} \in [a, b]$. Finally, C can be implemented as an arithmetic circuit that takes $O(m)$ bits as input and is of size $O(m + \epsilon^{-1/4} \log \epsilon^{-1})$ and degree $O(m)$. Note that the bit representation of the input number z is only used for selecting the piece to which z belongs—the spline polynomial itself is evaluated over an appropriately large prime field, adding only 3 to the total degree of the circuit. □

C Circuit primitives for binary number representation

Lemma 16 (Addition)**.** *There is an arithmetic circuit $C : \{0,1\}^{2n} \to \{0,1\}^{n+1}$ of size $O(n)$ and degree at most $2n$ for adding two n-bit numbers together for any positive integer n.*

Proof. We compute the sum of $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ by using full adders: For each bit i compute a carry bit c_i and an output bit z_i such that

$$z_i = (1 - 2c_{i+1})(x_i + y_i - 2x_i y_i) + c_{i+1}$$

and

$$c_i = x_i y_i + x_i c_{i+1} + y_i c_{i+1} - 2x_i y_i c_{i+1}$$

with $c_{n+1} = 0$. □

Lemma 17 (Multiplication)**.** *There is an arithmetic circuit $C : \{0,1\}^{2n} \to \{0,1\}^{2n+1}$ of size and degree $O(n^{4.13})$ for multiplying two n-bit numbers for any positive integer n.*

Proof. There exists a formula of size $O(n^{3.13})$ per output bit that uses operators $\{\wedge, \oplus, \neg\}$ to add together n integers [5]. Each of these operators can be implemented as an arithmetic circuit of degree at most 2. Thus, the degree of the polynomial produced by a gate is the sum of degrees of its inputs.

The formula actually outputs two numbers whose sum equals the sum of all n numbers, so one additional addition is required. This can be computed with, for example, the addition circuit of Lemma 16, yielding the total size $O(n^{4.13})$ and degree $O(n^{4.13})$ for the circuit.

We obtain the same complexities for the multiplication circuit: Multiplying two n-bit numbers x and y is equivalent to summing x copies of y, and 2^k times y equals shifting y by k bits. □

Lemma 18 (Comparison)**.** *There is an arithmetic circuit $C : \{0,1\}^{2n} \to \{0,1\}$ of size $O(n)$ and degree $2n$ such that $C(x, y) = [x < y]$ for all $x, y \in \{0,1\}^n$ for any positive integer n.*

Proof. Let

$$C(x, y) := \sum_{i=1}^{n} [\![x_i < y_i]\!] \prod_{j=1}^{i-1} [\![x_j = x_j]\!] ,$$

where $[\![x_i < y_i]\!]$ is a shorthand for $(1-x_i)y_i$ and $[\![x_j = y_j]\!]$ is a shorthand for $x_j y_j + (1-x_j)(1-y_j)$. Observe that $C(x, y) = 1$ if $x < y$, and $C(x, y) = 0$ otherwise. □

References

[1] Charles A. Hall. On error bounds for cubic spline interpolation. *Journal of Approximation Theory*, 1:209–218, 1968.

[2] Charles A. Hall and W. Weston Meyer. Optimal error bounds for cubic spline interpolation. *Journal of Approximation Theory*, 16(2):105–122, 1976.

[3] Charles F. F. Karney. Sampling exactly from the normal distribution. *ACM Trans. Math. Softw.*, 42(1):3:1–3:14, 2016.

[4] Sky McKinley and Megan Levine. Cubic spline interpolation. *College of the Redwoods*, 45(1):1049–1060, 1998.

[5] Mike Paterson and Uri Zwick. Shallow circuits and concise formulae for multiple addition and multiplication. *Comput. Complex.*, 3:262–291, 1993.

[6] John von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, chapter 13, pages 36–38. US Government Printing Office, Washington, DC, 1951.

[7] Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. Cambridge University Press, Cambridge, third edition, 2013.