COMP 480/580 — Probabilistic Algorithms and Data Structure Oct 02, 2025

Lecture 12

Lecturer: Anshumali Shrivastava Scribe By: Saj P. sp258, Zachary T. zt31, Reid S. rs273

1 Introduction

These notes summarize the core ideas behind *rejection sampling* (also known as *accept–reject sampling*). The goal is to generate exact samples from a **target** distribution π even when it is hard to sample from directly, by using a tractable **proposal** distribution p and an envelope constant M such that

$$\pi(x) \leq M\, p(x) \qquad \text{for all } x.$$

We focus on the discrete and continuous settings, give a short correctness proof, quantify efficiency, and work through examples (Beta target with Uniform proposal; Gaussian target with Laplace proposal).

2 Important Prior Knowledge

2.1 Conditional Probability

Definition: The conditional probability of an event A given that another event B has occurred is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{where } P(B) > 0.$$

It represents the probability that A occurs under the condition that B is known to have occurred.

Example: Suppose we roll a fair six-sided die. Let:

$$A = \{\text{even number}\} = \{2, 4, 6\}, \quad B = \{\text{number greater than 3}\} = \{4, 5, 6\}.$$

Then:

$$P(A) = \frac{3}{6}, \quad P(B) = \frac{3}{6}, \quad P(A \cap B) = \frac{2}{6}.$$

Hence,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{(2/6)}{(3/6)} = \frac{2}{3}.$$

2.2 Law of Total Probability

Definition: If $\{A_1, A_2, \ldots, A_n\}$ is a partition of the sample space S (i.e., A_i are mutually exclusive and $\bigcup_i A_i = S$), then for any event B:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

Using the definition of conditional probability, we can also write:

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

Example: A factory has three machines M_1, M_2, and M_3 producing 30%, 45%, and 25% of the total items, respectively. The probabilities that a machine produces a defective item are:

$$P(D \mid M_1) = 0.01, \quad P(D \mid M_2) = 0.03, \quad P(D \mid M_3) = 0.02.$$

Then, by the law of total probability:

$$P(D) = \sum_{i=1}^{3} P(D \cap M_i) = \sum_{i=1}^{3} P(D \mid M_i) P(M_i).$$

Substituting values:

$$P(D) = (0.01)(0.30) + (0.03)(0.45) + (0.02)(0.25) = 0.0215.$$

Thus, the overall probability that a randomly chosen item is defective is 2.15%.

3 Algorithm (Accept–Reject/ Rejection Sampling)

Assume we can sample from p and evaluate unnormalized densities $\tilde{\pi}(x) \propto \pi(x)$ and $\tilde{p}(x) \propto p(x)$. Choose $M \geq \sup_x \frac{\tilde{\pi}(x)}{\tilde{p}(x)}$ (any valid upper bound works).

1. Draw $X \sim p(\cdot)$.
2. Draw $U \sim \text{Unif}(0, 1)$ independently.

3. **Accept** X if

$$U \le \frac{\pi(X)}{M\,p(X)} \quad \left(\text{equivalently } U \le \frac{\tilde{\pi}(X)}{M\,\tilde{p}(X)}\right);$$

otherwise **reject** and repeat from step 1.

When π and p are probability *mass* functions, replace densities by masses. In either case, the algorithm returns i.i.d. draws from π.

4 Correctness (Why accepted samples follow π)

Write A for the event "accepted." By the law of total probability,

$$\mathbb{P}(A) = \int \mathbb{P}(A \mid X = x)\,p(x)\,dx = \int \frac{\pi(x)}{M\,p(x)}\,p(x)\,dx = \frac{1}{M}\int \pi(x)\,dx = \frac{1}{M}.$$

(For a discrete space, replace integrals by sums; since $\sum_x \pi(x) = 1$, the last equality still holds.) For any measurable set B,

$$\mathbb{P}(X \in B \mid A) = \frac{\int_B \mathbb{P}(A \mid X = x)\,p(x)\,dx}{\mathbb{P}(A)} = \frac{\int_B \frac{\pi(x)}{M\,p(x)} p(x)\,dx}{1/M} = \int_B \pi(x)\,dx,$$

so the accepted X has distribution π.

5 Acceptance Probability and Efficiency

From the calculation above,

$$\boxed{\mathbb{P}(\text{accept}) = \frac{1}{M}}$$

and if T is the number of proposals needed to get the first acceptance, then

$$T \sim \text{Geom}\left(\frac{1}{M}\right) \qquad \Longrightarrow \qquad \boxed{\mathbb{E}[T] = M}$$

(Geometric with "success" probability $1/M$ on each independent trial.) Thus, smaller M means higher acceptance rate and fewer proposals per accepted sample.

6 Choosing the Proposal p and the Envelope M

Good proposals satisfy:

- **Easy to sample:** direct sampler or cheap inverse CDF.
- **Shape match:** p approximates π in location, spread, and tails.
- **Small bound:** find a tight $M \geq \sup_x \frac{\tilde{\pi}(x)}{\tilde{p}(x)}$.

There is a trade-off: heavier-tailed proposals (vs. the target) make it easier to cover the tails uniformly (valid for all x) but typically increase M.

7 Example A: Beta(a, b) target on $(0, 1)$ with Uniform proposal

Let $a \geq 1$, $b \geq 1$. The Beta density is

$$\pi(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1.$$

Choose $p(x) = \text{Unif}(0, 1)$, so $p(x) = 1$ for $0 < x < 1$. Then

$$\frac{\pi(x)}{p(x)} = \pi(x) \leq \max_{x \in (0,1)} \pi(x) =: M.$$

A convenient bound is $M = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \max_{x\in(0,1)} x^{a-1}(1-x)^{b-1}$; for $a, b > 1$ the maximum occurs at $x^* = \frac{a-1}{a+b-2}$ with value

$$M = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{a-1}{a+b-2}\right)^{a-1} \left(\frac{b-1}{a+b-2}\right)^{b-1}.$$

Sampler.

1. Draw $X \sim \text{Unif}(0, 1)$ and $U \sim \text{Unif}(0, 1)$.
2. Accept X if $U \leq \dfrac{x^{a-1}(1-x)^{b-1}}{M}$; else reject and repeat.

8 Example B: Standard Normal target with Laplace proposal

Target: $\pi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ on $\mathbb{R}$. Proposal (Laplace/Double Exponential with rate $\lambda > 0$):

$$p(x) = \frac{\lambda}{2} e^{-\lambda |x|}, \qquad x \in \mathbb{R}.$$

Both are symmetric and centered; Laplace has a sharper peak and heavier tails.

Bounding M

For $x \geq 0$ (by symmetry),

$$\frac{\pi(x)}{p(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{\lambda}{2} e^{-\lambda x}} = \frac{1}{\lambda} \sqrt{\frac{2}{\pi}} \exp\Big(\lambda x - \tfrac{x^2}{2}\Big).$$

Maximize $f(x) = \lambda x - \frac{x^2}{2}$ over $x \geq 0$. The maximizer is $x = \lambda$, giving

$$\boxed{M = \frac{1}{\lambda} \sqrt{\frac{2}{\pi}}\, e^{\lambda^2/2}}$$

and acceptance probability $1/M$. (Choosing $\lambda = 1$ is a common, simple setting.)

Sampler using inverse CDF of Laplace

1. Draw $U \sim \mathrm{Unif}(0, 1)$ and set

$$Y = \begin{cases} -\dfrac{1}{\lambda} \ln(2U), & U \leq \frac{1}{2}, \\ \dfrac{1}{\lambda} \ln\big(2(1 - U)\big), & U > \frac{1}{2}. \end{cases}$$

   (This yields $Y \sim \mathrm{Laplace}(0, 1/\lambda)$.)

2. Draw $V \sim \mathrm{Unif}(0, 1)$.
3. Accept Y as a $N(0, 1)$ sample if

$$V \leq \frac{\pi(Y)}{M\, p(Y)} = \exp\Big(-\frac{(|Y| - \lambda)^2}{2}\Big);$$

   otherwise reject and repeat.

Visual intuition

The envelope $Mp(x)$ lies just above $\pi(x)$ everywhere. The area ratio $\int \pi / \int (Mp) = 1/M$ is the acceptance probability. A closer shape match (smaller M) yields fewer rejections.

9 Comparison to Importance Sampling (brief)

Both methods use a proposal p:

- **Rejection sampling** produces *exact* draws from π; efficiency is governed by M (expected proposals per acceptance is M).
- **Importance sampling** draws from p and forms weighted estimates with weights $\frac{\pi(x)}{p(x)}$; variance depends on weight stability. No rejections, but samples are not from π directly.

When a tight bound M is easy to obtain and p is close to π, rejection sampling is attractive; when exact sampling is less critical or a good bound is hard, importance sampling may be preferable.

10 Practical Notes

- If π is known only up to a constant, use unnormalized densities; the constant cancels in the acceptance ratio.
- In high dimensions, finding a good p and small M can be difficult (acceptance rates deteriorate).
- Always verify $\pi(x) \leq Mp(x)$ for all x (or compute a provable bound for M).

11 References

- Standard Monte Carlo texts (e.g., Robert & Casella) for accept–reject methods and envelope bounds.