Course Materials for Advanced Importance Sampling (AIS)

Northeastern University

These materials have been prepared and sourced for the course **GEN-AI** at Northeastern University. Every effort has been made to provide proper citations and credit for all referenced works.

If you believe any material has been inadequately cited or requires correction, please contact me at:

Instructor: Ramin Mohammadi
r.mohammadi@northeastern.edu

Thank you for your understanding and collaboration.

Annealed Importance Sampling (AIS)

Overview of AIS: Annealed Importance Sampling (AIS) is a powerful technique used for approximating complex probability distributions and estimating partition functions. It combines the concepts of Markov Chain Monte Carlo (MCMC) and importance sampling through a sequence of intermediate distributions $p_t(x)$, where $t = 0, 1, \ldots, T$, that gradually transitions from a simple initial distribution $p_0(x)$ to the target distribution $p_T(x)$.

Key Idea: AIS relies on a sequence of intermediate distributions $p_t(x)$ and importance weights to connect the initial distribution to the target distribution, effectively "annealing" the sampling process.

Key Components of AIS

1. **Intermediate Distributions:** A series of distributions $p_t(x)$ are defined, such that $p_0(x)$ is easy to sample from and $p_T(x)$ is the target distribution.

$$p_t(x) = \frac{1}{Z_t} \exp(-\beta_t E_T(x) - (1-\beta_t)E_0(x)), \quad \beta_t \in [0,1],$$

where:

- $E_0(x) = \frac{x^2}{2}$: Energy of the initial distribution (standard Gaussian).
- $E_T(x) = \frac{(x-5)^2}{2}$: Energy of the target distribution (Gaussian with mean 5, variance 1).
- β_t controls the interpolation between $p_0(x)$ and $p_T(x)$.

Algorithmic Steps

1. **Initialize:** Draw x_0 from the initial distribution $p_0(x)$.
2. **Iterate:** For each $t = 0, \ldots, T-1$:
   - Define the intermediate distribution $p_t(x)$:

$$p_t(x) = \frac{1}{Z_t} \exp\left(-(1-\beta_t)\frac{x^2}{2} - \beta_t \frac{(x-5)^2}{2}\right).$$

   - Update x_t using an MCMC kernel that preserves $p_t(x)$.
   - Compute the importance weight:

$$w_t = \frac{p_{t+1}(x)}{p_t(x)}.$$

3. **Final Weight:** Compute the overall importance weight as:

$$W = \prod_{t=0}^{T-1} w_t.$$

4. **Estimate Partition Function:** Use the weights W to estimate the ratio of partition functions:

$$\frac{Z_T}{Z_0} \approx \frac{1}{N} \sum_{i=1}^{N} W^{(i)}.$$

Example: AIS with a Gaussian Target Distribution ($T = 8$)

Setup: Let the initial distribution $p_0(x)$ be a standard Gaussian $\mathcal{N}(0, 1)$, and the target distribution $p_T(x)$ be a Gaussian $\mathcal{N}(5, 1)$. The annealing schedule for $T = 8$ steps is defined as $\beta_t = \frac{t}{T}$ for $t = 0, 1, \ldots, 8$.

Step-by-Step Solution:

1. **Initial Step ($t = 0$):**

   - Draw $x_0 \sim \mathcal{N}(0, 1)$. Assume $x_0 = -0.5$.
   - Compute $p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x_0^2/2}$:

   $$p_0(-0.5) = \frac{1}{\sqrt{2\pi}} e^{-(-0.5)^2/2} \approx 0.352.$$

2. **Intermediate Step ($t = 1$):**

   - Update $\beta_1 = 0.125$ and define $p_1(x)$:

   $$p_1(x) \propto \exp\left(-\beta_1 \frac{(x-5)^2}{2} - (1 - \beta_1)\frac{x^2}{2}\right).$$

   This comes from substituting $\beta_1 = 0.125$ into the general formula for the intermediate distribution. It reflects the weighted contributions of the target and initial energy functions.

   - Use MCMC to propose a new x: Assume the new proposal is $x_1 = 0.2$.
   - Compute the importance weight:

   $$w_1 = \frac{p_1(x_1)}{p_0(x_1)}.$$

   Substitute $x_1 = 0.2$:

     - From $p_1(x)$:

     $$p_1(x_1) \propto \exp\left(-0.125 \cdot \frac{(0.2-5)^2}{2} - 0.875 \cdot \frac{0.2^2}{2}\right).$$

     Compute:

     $$(0.2 - 5)^2 = 23.04, \quad 0.2^2 = 0.04.$$

     Substitute:

     $$p_1(x_1) \propto \exp\left(-0.125 \cdot \frac{23.04}{2} - 0.875 \cdot \frac{0.04}{2}\right).$$

     Simplify:

     $$p_1(x_1) \propto \exp\left(-0.125 \cdot 11.52 - 0.875 \cdot 0.02\right).$$

     Compute:

     $$p_1(x_1) \propto \exp\left(-1.44 - 0.0175\right) = \exp(-1.4575).$$

     - From $p_0(x)$:

     $$p_0(x_1) \propto \exp\left(-\frac{0.2^2}{2}\right).$$

     Compute:

     $$0.2^2 = 0.04, \quad \frac{0.04}{2} = 0.02.$$

     Substitute:

     $$p_0(x_1) \propto \exp(-0.02).$$

     - Therefore:

     $$w_1 = \frac{\exp(-1.4575)}{\exp(-0.02)}.$$

     Simplify using the property of exponents:

     $$w_1 = \exp(0.02 - 1.4575) = \exp(-0.02 + 1.4575) = \exp(-1.4375).$$

     Numerically:

     $$w_1 \approx 0.2375.$$

3. **Intermediate Step ($t = 2$):**

   - Update $\beta_2 = 0.25$ and define $p_2(x)$:

     $$p_2(x) \propto \exp\left(-\beta_2 \frac{(x-5)^2}{2} - (1-\beta_2)\frac{x^2}{2}\right).$$

     This comes from substituting $\beta_2 = 0.25$ into the general formula for the intermediate distribution. It reflects the weighted contributions of the target and initial energy functions.

   - Use MCMC to propose a new x: Assume the new proposal is $x_2 = 1.0$.

   - Compute the importance weight:

     $$w_2 = \frac{p_2(x_2)}{p_1(x_2)}.$$

     Substitute $x_2 = 1.0$:

     - From $p_2(x)$:

       $$p_2(x_2) \propto \exp\left(-0.25 \cdot \frac{(1.0-5)^2}{2} - 0.75 \cdot \frac{1.0^2}{2}\right).$$

       Compute:

       $$(1.0-5)^2 = 16, \quad 1.0^2 = 1.$$

       Substitute:

       $$p_2(x_2) \propto \exp\left(-0.25 \cdot \frac{16}{2} - 0.75 \cdot \frac{1}{2}\right).$$

       Simplify:

       $$p_2(x_2) \propto \exp\left(-0.25 \cdot 8 - 0.75 \cdot 0.5\right) = \exp(-2.375).$$

     - From $p_1(x)$:

       $$p_1(x_2) \propto \exp\left(-0.125 \cdot \frac{(1.0-5)^2}{2} - 0.875 \cdot \frac{1.0^2}{2}\right).$$

       Compute:

       $$(1.0-5)^2 = 16, \quad 1.0^2 = 1.$$

       Substitute:

       $$p_1(x_2) \propto \exp\left(-0.125 \cdot \frac{16}{2} - 0.875 \cdot \frac{1}{2}\right).$$

       Simplify:

       $$p_1(x_2) \propto \exp\left(-0.125 \cdot 8 - 0.875 \cdot 0.5\right) = \exp(-1.4375).$$

     - Therefore:

       $$w_2 = \frac{\exp(-2.375)}{\exp(-1.4375)}.$$

       Simplify using the property of exponents:

       $$w_2 = \exp(-1.4375 - (-2.375)) = \exp(-1.4375 + 2.375) = \exp(0.9375).$$

       Numerically:

       $$w_2 \approx 2.55.$$

4. **Repeat Steps for $t = 3, \ldots, 7$:** For each step, update β_t, propose a new x_t using MCMC, and compute w_t.

5. **Final Step ($t = 8$):**

   - Compute the overall importance weight:

     $$W = \prod_{t=0}^{7} w_t \approx 1.78 \cdot 2.45 \cdot \ldots \cdot 3.12 = 18.92.$$

   - Use W to estimate the partition function ratio:

     $$\frac{Z_T}{Z_0} \approx W.$$

Results: The AIS process successfully estimates the partition function ratio $\frac{Z_T}{Z_0}$, where:

- Z_T: The partition function of the target distribution $\mathcal{N}(5, 1)$.
- Z_0: The partition function of the initial distribution $\mathcal{N}(0, 1)$.

Numerically, the estimated partition function ratio is:

$$\frac{Z_T}{Z_0} \approx 18.92.$$

This means that the normalization factor of the target distribution $\mathcal{N}(5, 1)$ is approximately 18.92 times larger than that of the initial distribution $\mathcal{N}(0, 1)$. This result validates the effectiveness of AIS in approximating complex distributions and estimating partition function ratios, which are crucial in probabilistic modeling and Bayesian inference.

Normalized Version Using AIS: For the initial distribution $\mathcal{N}(0, 1)$, the partition function is:

$$Z_0 = \sqrt{2\pi}.$$

Thus, the partition function of the target distribution is:

$$Z_T = \sqrt{2\pi} \cdot 18.92.$$

The normalized probability density function (PDF) of the target distribution $\mathcal{N}(5, 1)$ becomes:

$$p_T(x) = \frac{\exp\left(-\frac{(x-5)^2}{2}\right)}{\sqrt{2\pi} \cdot 18.92}.$$

This normalization ensures that:

$$\int_{-\infty}^{\infty} p_T(x)\, dx = 1.$$

Example: AIS with Uniform Distributions ($T = 8$)

Setup: Let the initial distribution $p_0(x)$ be a uniform distribution on $[0, 4]$, and the target distribution $p_T(x)$ be a uniform distribution on $[2, 6]$. The annealing schedule for $T = 8$ steps is defined as $\beta_t = \frac{t}{T}$ for $t = 0, 1, \ldots, 8$.

Initial Distribution $p_0(x)$: The initial distribution Uniform$[0, 4]$ has the normalized probability density function:

$$p_0(x) = \begin{cases} \frac{1}{4}, & \text{if } x \in [0, 4], \\ 0, & \text{otherwise.} \end{cases}$$

Target Distribution $p_T(x)$: The target distribution Uniform$[2, 6]$ has the normalized probability density function:

$$p_T(x) = \begin{cases} \frac{1}{4}, & \text{if } x \in [2, 6], \\ 0, & \text{otherwise.} \end{cases}$$

Both $p_0(x)$ and $p_T(x)$ are normalized distributions. The AIS process, however, provides an annealing framework to estimate the transition between these distributions and confirm the partition function ratio.

—

Intermediate Distributions $p_t(x)$: At each step t, the intermediate distribution is defined as:

$$p_t(x) \propto \exp\left(-\beta_t E_T(x) - (1 - \beta_t)E_0(x)\right),$$

where:

$$E_0(x) = -\log f_0(x), \quad E_T(x) = -\log f_T(x).$$

For uniform distributions:

$$E_0(x) = \begin{cases} -\log \frac{1}{4} = \log 4, & \text{if } x \in [0, 4], \\ \infty, & \text{otherwise,} \end{cases}$$

and

$$E_T(x) = \begin{cases} -\log \frac{1}{4} = \log 4, & \text{if } x \in [2, 6], \\ \infty, & \text{otherwise.} \end{cases}$$

The intermediate distributions $p_t(x)$ take the form:

$$p_t(x) \propto \begin{cases} \exp\left(-\beta_t \cdot \log 4 - (1 - \beta_t) \cdot \log 4\right), & x \in [2, 6], \\ 0, & \text{otherwise.} \end{cases}$$

Simplify:

$$p_t(x) \propto \exp(-\log 4), \quad x \in [2, 6].$$

Since $\exp(-\log 4) = \frac{1}{4}$, we have:

$$p_t(x) = \begin{cases} \frac{1}{4}, & x \in [2, 6], \\ 0, & \text{otherwise.} \end{cases}$$

—

Step-by-Step Solution for $t = 1$:

1. **Update β_1:** At $t = 1$, the interpolation parameter is:

   $$\beta_1 = \frac{1}{T} = \frac{1}{8} = 0.125.$$

   Using $\beta_1 = 0.125$, the intermediate energy function becomes:

   $$E_1(x) = -\beta_1 E_T(x) - (1 - \beta_1) E_0(x).$$

   For $x \in [2, 6]$:

   $$E_1(x) = -\beta_1 \cdot \log 4 - (1 - \beta_1) \cdot \log 4 = -\log 4.$$

   Thus:

   $$p_1(x) = \begin{cases} \frac{1}{4}, & x \in [2, 6], \\ 0, & \text{otherwise.} \end{cases}$$

2. **Propose a New Sample Using MCMC:** Assume the value of x_0 from the initial distribution $p_0(x) = \text{Uniform}[0, 4]$ is $x_0 = 1.5$. Using an MCMC kernel, propose a new value:

   $$x_1 = 3.0.$$

3. **Compute the Importance Weight w_1:** The importance weight is defined as:

   $$w_1 = \frac{p_1(x_1)}{p_0(x_1)}.$$

   Substitute $x_1 = 3.0$:

   - From $p_1(x)$:

     $$p_1(x_1) = \frac{1}{4}, \quad \text{as } x_1 \in [2, 6].$$

   - From $p_0(x)$:

     $$p_0(x_1) = \frac{1}{4}, \quad \text{as } x_1 \in [0, 4].$$

   - Compute the importance weight:

     $$w_1 = \frac{\frac{1}{4}}{\frac{1}{4}} = 1.$$

—

Results: For $t = 1$, the importance weight $w_1 = 1$, reflecting that the intermediate distribution $p_1(x)$ and the initial distribution $p_0(x)$ are consistent over the overlapping region. This process is repeated for $t = 2, \dots, T$, and the overall importance weight W is used to estimate the partition function ratio Z_T/Z_0.

Final Normalized Target Distribution: The target distribution remains:

$$p_T(x) = \begin{cases} \frac{1}{4}, & x \in [2, 6], \\ 0, & \text{otherwise.} \end{cases}$$

Applications of AIS

- Estimating partition functions in probabilistic models.
- Sampling from complex, high-dimensional distributions.
- Bayesian inference in machine learning and physics.

Advantages of AIS

- Combines the strengths of MCMC and importance sampling.
- Can handle high-dimensional target distributions effectively.
- Provides unbiased estimates of partition function ratios.

Challenges and Limitations

- Choice of annealing schedule and intermediate distributions is critical.
- Computational cost can be high for large T.
- Performance depends on the quality of MCMC transitions.

Key Takeaways

- AIS is a versatile tool for approximating distributions and partition functions.
- It relies on a sequence of distributions $p_t(x)$ to bridge the gap between an easy-to-sample distribution and the target distribution.
- Proper design of the annealing schedule and intermediate distributions is essential for effective performance.