Cross-Entropy Method for Optimization

Anny Cui

December 12, 2023

1 Introduction

The Cross-Entropy (CE) method, developed by Dr. Reuven Y. Rubinstein in the late 1990s, is a significant optimization and estimation technique, initially designed for estimating probabilities of rare events in complex stochastic networks (Rubinstein, 1997). Rubinstein is an Israeli scientist renowned for his contributions to simulation and optimization techniques. His expertise lies in applied probability and stochastic modeling, and his work has significantly influenced the fields of operations research and computer simulation. The CE method is related to and evolved from earlier work in importance sampling and rare event simulation, techniques used for efficient probability estimation in complex stochastic systems (Rubinstein & Kroese, 2004). These methods were foundational in the development of CE as they also focus on modifying probability distributions to estimate rare events more efficiently.

The CE method is particularly suited for problems like combinatorial optimization, where it has been used in network design and resource allocation, and continuous optimization, applicable in engineering and financial modeling. Its robustness makes it highly effective in simulation-based optimization, especially where analytical models are either infeasible or unavailable.

This paper will unfold with a generic form of the CE method, and then apply the algorithm to two typical optimization problem examples to see its performance and evaluate its applications with pros and cons.

2 Algorithm Technique

2.1 General Form of Optimization Problems

The CE method addresses a wide range of optimization problems, both discrete and continuous, typically formulated as finding an optimal solution x^* that minimizes or maximizes a given objective function $f(x)$.

2.2 Mathematical Framework

- **Objective Function** $f(x)$: The objective function to be optimized, which can be in various forms depending on the problem. If x is an n-dimensional vector, then $f : \mathbb{R}^n \to \mathbb{R}$
- **Solution Space** $\mathcal{X}$: A modifiable distribution from which samples are drawn. The choice of distribution often depends on the nature of the optimization problem. $\mathcal{X} \subset \mathbb{R}^n$ for continuous problems or $\mathcal{X} \subset \mathbb{Z}^n$ for discrete problems
- **Probability Distribution** P: Sampling of solutions updated based on performance.

2.3 Steps of the Method

0. **Initialization**: Guess an initial proposal distribution P_0 over the solution space $\mathcal{X}$.
1. **Sampling**: Generate a set of samples $\{x_i\}$ from the distribution P.
2. **Evaluation**: Evaluate these samples using the objective function $f(x_i)$.
3. **Selection**: Select the best-performing samples (called elite), based on the chosen elite fraction. Only a certain percentage of the best samples, as determined by the elite fraction (ie. top 20%), are selected to influence the next generation.
4. **Update**: Update the distribution P to increase the likelihood of generating similar high-performing samples in the next iteration.
5. **Iteration**: Repeat the sampling, evaluation, selection, and update steps until a convergence criterion is met or reach the max number of iterations.

3 Application to Examples

The application of the Cross-Entropy (CE) method to two different types of optimization problems — Portfolio Optimization and the Traveling Salesman Problem (TSP) — showcases its versatility and efficiency in handling both continuous and discrete problem spaces. Here's a detailed breakdown of each application:

3.1 Example 1: Portfolio Optimization

Maximize the expected return of a financial portfolio under risk constraints, specifically by maximizing the Sharpe Ratio, a common measure in financial portfolio optimization for balancing return and risk. The Sharpe ratio is calculated as the ratio of the excess expected return of the portfolio over the risk-free rate, divided by the standard deviation of the portfolio's excess return (i.e., its risk). For simplicity, we will assume a risk-free rate of 0% in this case.

- **Expected Returns**: Giving a set of expected returns for three assets:

$$\text{expected returns} = \begin{bmatrix} 0.12 \\ 0.10 \\ 0.15 \end{bmatrix}$$

- **Covariance Matrix**: Represents the covariance between the asset returns:

$$\text{covariance matrix} = \begin{bmatrix} 0.10 & 0.02 & 0.04 \\ 0.02 & 0.08 & 0.01 \\ 0.04 & 0.01 & 0.12 \end{bmatrix}$$

3.1.1 Objective Function: Sharpe Ratio

The Sharpe Ratio, S, is defined as:

$$\max \quad S(w) = \frac{R_p(w) - R_f}{\sigma_p(w)}$$

where:

- $R_p(w)$ is the expected return of the portfolio, calculated as $R_p(w) = w^T r$.
- R_f is the risk-free rate, assumed to be 0 for simplicity.
- $\sigma_p(w)$ is the standard deviation of the portfolio's return, representing risk, calculated as $\sigma_p(w) = \sqrt{w^T C w}$.
- w is the vector of portfolio weights.
- r is the vector of expected returns for each asset.
- C is the covariance matrix of asset returns.

3.1.2 CE Method Application (See code)

0. **Initialization**: We start with an initial guess of the portfolio weights, assuming a uniform distribution.
1. **Sampling**: Generate a set of portfolio weights from the current probability distribution.
2. **Evaluation**: Calculate the Sharpe Ratio for each set of weights.
3. **Selection**: Select the top-performing portfolios.
4. **Update**: Adjust the probability distribution to reflect the characteristics of the top-performing portfolios.
5. **Iteration**: Repeat the process until convergence is achieved.

3.1.3 Outcome

The Cross-Entropy method has yielded the following optimized portfolio weights:

$$\text{Optimized Weights} = \begin{bmatrix} 0.24257426 \\ 0.38118812 \\ 0.37623762 \end{bmatrix}$$

- The first asset should constitute approximately 24.26% of the total investment.
- The second asset should account for about 38.12% of the portfolio.
- The third asset should be allocated around 37.62%.

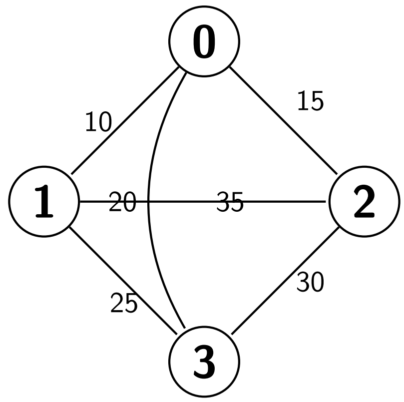
These weights were determined by the CE method to maximize the Sharpe Ratio, which is a measure of the excess return per unit of risk. The higher the Sharpe Ratio, the better the risk-adjusted return of the portfolio. The distribution of investment across the assets reflects a balance between maximizing expected returns and minimizing risk, as quantified by the covariance matrix and the expected returns of individual assets. The allocation suggests a diversified portfolio, with a significant proportion of the investment spread across all three assets. Such diversification is a common strategy to mitigate risk. The specific weight distribution is influenced by both the expected returns and the relationships between the assets as indicated by the covariance matrix. Assets with higher expected returns or those contributing to lower overall portfolio risk might receive higher allocations.

Real-world Implications: In a practical setting, these weights would guide how an investor or portfolio manager allocates capital among the given assets. For instance, if the total investment is \$100,000, then approximately \$24,257 would be invested in the first asset, $\$38,119$ in the second, and $\$37,624$ in the third.It's important to note that these results are based on the provided data and assumptions (like a 0% risk-free rate). In a real-world application, other factors such as transaction costs, taxes, liquidity, and changing market conditions would also need to be considered.

In summary, the output from the CE optimization process provides a data-driven recommendation for asset allocation that aims to maximize the risk-adjusted returns of the portfolio, considering the expected returns and covariance among the assets.

3.2 Example 2: Traveling Salesman Problem (TSP)

The Traveling Salesman Problem (TSP) is a classic problem in combinatorial optimization. The goal is to find the shortest possible route that visits a given set of cities and returns to the origin city. We represent the problem using a distance matrix where each element indicates the distance between two cities.



- **Cities**: Consider a set of four cities, labeled as 0, 1, 2, and 3.
- **Distance Matrix**: Represents the distances between each pair of cities:

$$\text{distance matrix} = \begin{bmatrix} 0 & 10 & 15 & 20 \\ 10 & 0 & 35 & 25 \\ 15 & 35 & 0 & 30 \\ 20 & 25 & 30 & 0 \end{bmatrix}$$

3.2.1 Objective Function: Total Route Distance

The objective is to minimize the total distance of the route, defined as:

$$\min \quad D(r) = \sum_{i=0}^{n-1} d(r_i, r_{i+1})$$

where:

- $D(r)$ is the total distance of route r.
- r_i represents the i^{th} city in the route.
- $d(r_i, r_{i+1})$ is the distance from city r_i to city r_{i+1}.
- n is the total number of cities.

3.2.2 CE Method Application (See code)

0. **Initialization**: Begin with a uniform probability distribution for each city transition.
1. **Sampling**: Generate a set of routes based on the current probability distribution.
2. **Evaluation**: Calculate the total distance for each route.
3. **Selection**: Select the routes with the shortest total distances.
4. **Update**: Adjust the probability distribution to favor transitions found in shorter routes.
5. **Iteration**: Repeat the process until convergence is achieved.

3.2.3 Outcome

The Cross-Entropy method has yielded the following optimized route:

$$\text{Optimized Route} = [0, 2, 3, 1, 0]$$

The route starts and ends at city 0 with total distance of 80, visiting all other cities exactly once. This route was determined by the CE method to minimize the total distance traveled, providing an efficient solution to the TSP. The specific route order is influenced by the distances between cities, with the algorithm favoring transitions that collectively lead to a shorter travel distance.

Real-world Implications: The TSP has practical applications in logistics and transportation, where finding the most efficient route can lead to significant time and cost savings. For instance, in delivery services or route planning for sales representatives, an optimized route minimizes travel time and distance, leading to reduced fuel costs and improved operational efficiency.

In summary, the TSP is a fundamental problem in optimization, and the CE method provides a robust approach to finding a near-optimal solution, showcasing its applicability in solving complex combinatorial problems.

4 Implementation Insights

The key advantages of the CE method include its simplicity, making it relatively easy to implement and understand, and its robustness, allowing effective performance in noisy and complex environments. However, the method is not without its disadvantages; it can be sensitive to the choice of parameters, such as the selection percentile of elite fraction, and might struggle with convergence in extremely high-dimensional spaces or in scenarios with extremely rugged landscapes.

Examples demonstrate CE's versatility in handling both continuous (portfolio optimization) and discrete problems (TSP). CE efficiently identified near-optimal solutions, showcasing its effectiveness. Despite its effectiveness, the CE method has certain limitations, such as scalability issues in very large or complex problem spaces and a tendency to get trapped in local optima. We can try to combine CE with other algorithms (like genetic algorithms or local search) to escape local optima. Such hybrid approaches could help in overcoming the drawbacks of CE, particularly in dealing with local optima and enhancing its scalability and applicability to even more diverse problem domains.

References

[1] Rubinstein, R. Y. (1997). Optimization of computer simulation models with rare events. European Journal of Operational Research, 99(1), 89–112. Retrieved on Dec. 12, 2023, from https://doi.org/10.1016/S0377-2217(96)00385-2

[2] Rubinstein, R. Y., & Kroese, D. P. (2004). The Cross-Entropy Method. Springer. https://doi.org/10.1007/978-1-4757-4321-0

[3] Kochenderfer, M. J., & Wheeler, T. A. (2018). Algorithms for Optimization. MIT Press. Retrieved Dec. 12, 2023, from https://mitpress.mit.edu/9780262039420/algorithms-for-optimization/