2008 年度　修士論文要旨

モンテカルロシミュレーションにおける重点サンプリング法に対する大偏差理論の適用について

関西学院大学大学院　理工学研究科

物理学専攻　千代延研究室　漆原 勉

モンテカルロ法とは通信工学や金融工学など様々な分野で用いられている手法であって, 定積分などの決定論的な計算を, 乱数を用いて確率論的に近似 (推定) するものである. その際, 特にレアイベントシミュレーションにおいて推定が十分な精度を保つために用いるサンプル (乱数) の数は莫大なものになる. そこには推定量の分散が大きく関わっている. 以下の状況について考える.

- $\mathbf{X} = (X_1, \cdots, X_n)$: $\mathbf{R}^n$-値の確率変数
- H : $\mathbf{R}^n \to \mathbf{R}$ のある関数 (sample performance)
- f : $\mathbf{X}$ の結合密度関数

とするとき,

$$\ell = E_f[H(\mathbf{X})] = \int_{\mathbf{R}^n} H(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

を推定することを考える.

このとき ℓ に対する最も素朴な (Crude Monte Carlo) 推定量は大数の法則より, $\hat{\ell} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{X}_k$ と得られる. ただし, $\mathbf{X}_1, \cdots, \mathbf{X}_N$ は $i.i.d.$ で密度 f をもつ分布に従うとする. しかし, この CMC 推定量よりもっと分散の低い推定量が得られればより少ないサンプル数で信頼性の高い推定が可能になる.

重点サンプリング法のアイデアとして, f とは別のある確率密度であって, $g(\mathbf{x}) = 0$ のとき $H(\mathbf{x}) f(\mathbf{x}) = 0$ である (Importance Sampling density)g を導入する. すると, ℓ は

$$\ell = \int_{\mathbf{R}^n} H(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{R}^n} H(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} = E_g[H(\mathbf{X}) \frac{f(\mathbf{X})}{g(\mathbf{X})}]$$

と変形できる. 従って, 重点サンプリング法に基づいた (Importance Sampling) 推定量としては同様に大数の法則より, $\hat{\ell} = \frac{1}{N}\sum_{k=1}^{N} H(\mathbf{X}_k)\frac{f(\mathbf{X}_k)}{g(\mathbf{X}_k)}$ が得られる. ただし, $\mathbf{X}_1, \cdots, \mathbf{X}_N$ は $i.i.d.$ で密度 g をもつ分布に従うとする. 従って, この IS 推定量の分散が CMC 推定量の分散よりも小さくなるように密度 g^* をとることができるのならば, IS 推定量に従って ℓ を推定する方が効率的であるといえる.

すると, 解析的な計算から $g^* = \frac{H(x)f(x)}{\ell}$ と得られるが, g^* は ℓ によって決まっているので, 未知である ℓ に対しては役立たない!!

本論文では, この重点サンプリング法で得られた最適分布 g^* に近い分布を定める方法として, (1) 大偏差原理による方法, (2)Cross-Entropy 法による方法を述べた. (1) の方法に対しては, 具体的な例として標本平均に対する大偏差確率, 保険会社の倒産確率を考えた. また, (2) の方法に対して, レアイベントシミュレーションに対するアルゴリズムを構築し, またそれを Traveling Salesman Problem などの最適化問題に対して適用することで最適解が得られることについて論じた.