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　理工学研究科　物理学専攻
千代延研究室　漆原 勉

講演の内容：

- モンテカルロ法における重点サンプリング法, 特に Cross-Entropy 法とは
- Cross-Entropy 法を用いたレアイベント (めったにおきない事象) のシミュレーションに対するアルゴリズム
- 具体的問題へのシミュレーション (Coin flipping, finding max)

モンテカルロ法

- 定積分などの解析的な計算を，サンプル (乱数) を用いて確率論的に近似する手法.

以下の設定を与える.

- $\mathrm{X} = (X_1, \cdots, X_n)$: R^n-値の確率変数
- H : $\mathrm{R}^n \to \mathrm{R}$ のある関数
- f : X の結合密度関数であるとする.

このとき,

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

の値を知りたいとする. ただし, E_f : f に従う確率変数 X に対する期待値とする.

このとき最も素朴な (CMC) 推定量は算術平均, すなわち,
$X_1, \cdots, X_N$, $i.i.d$, $\sim f$ として,

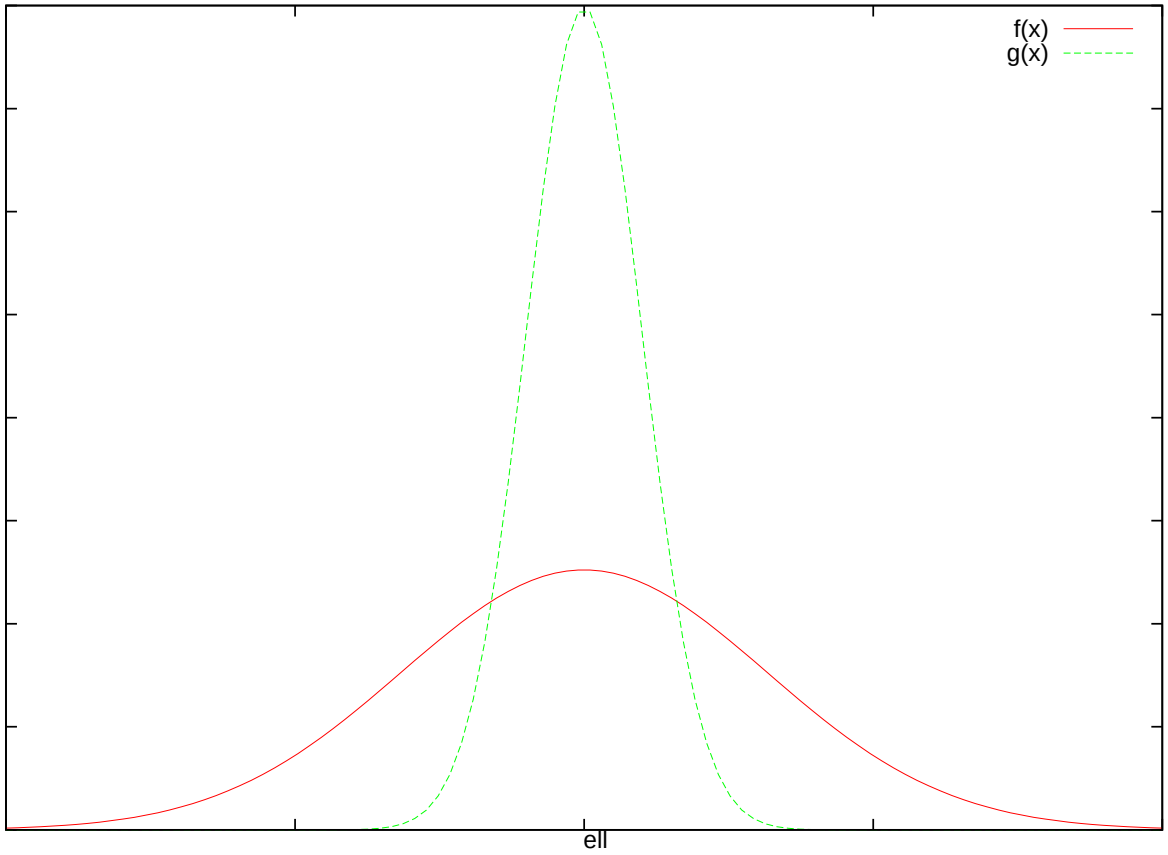
$$\hat{\ell}_{CMC} = \frac{1}{N} \sum_{i=1}^{N} H(X_i)$$

と考えられる. これは大数の法則より,

$$\lim_{N \to \infty} \hat{\ell}_{CMC} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} H(X_i) = E_f[H(\mathrm{X})] = \ell \quad (P - a.s.)$$

が保証される.

$\Longrightarrow$ 乱数 X_N を大量に発生させることで, 推定量 $\hat{\ell}$ で ℓ を近似
することができる!!


f(x)
g(x)
ell

$\Longrightarrow$ **効率のよい (CMC と同じぐらいの近似を行う際に用いる サンプル数 N が少なくすむ) 推定をするために, 分散の小さい 分布からサンプリングをする事を考えたい.**

$\Longrightarrow$ **重点サンプリング法**

重点サンプリング法 (Importance Sampling)

- $\mathrm{X} = (X_1, \cdots, X_n)$: R^n-値の確率変数
- H : $\mathrm{R}^n \to \mathrm{R}$ のある関数
- f : X の結合密度関数であるとする.

このとき, $\ell = E_f[H(\mathrm{X})]$ を推定したいとする.

IS のアイデアは, f とは別の確率密度 g であって, $g(x) = 0$ のとき $H(\mathrm{x})f(\mathrm{x}) = 0$ となる g を導入して, ℓ を

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

と考えるものである. ただし, E_g : g に従う X についての期待値とする.

すると, Importance Sampling(IS) を用いた推定量として

$$\hat{\ell}_{IS} = \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}$$

を得る. ただし, サンプリングは $\mathbf{X}_1, \cdots, \mathbf{X}_N \sim g$ とする.
IS 推定量も大数の法則から

$$\lim_{N\to\infty} \hat{\ell}_{IS} = E_g[H(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}] = \ell \quad (P-a.s.)$$

が成り立っている.
よって, $Var_f[\hat{\ell}_{CMC}] \geq Var_g[\hat{\ell}_{IS}]$ なる g が存在すれば重点サンプリング法によって推定するほうが効率がよいと言える.

$\Longrightarrow g^*(\mathrm{x}) = \frac{|H(\mathrm{x})|f(\mathrm{x})}{\ell}$ ととれば分散が最小化できる.
すなわち, $g = g^*$ ととると $\hat{\ell}_{IS}$ の分散が 0 になる.

実際 $H(\mathrm{x}) \geq 0$ とすると, $H(\mathrm{x})\frac{f(\mathrm{x})}{g^*(\mathrm{x})} = \ell$(定数) であるから,

$$Var_{g^*}[\hat{\ell}_{IS}] = Var_{g^*}[\frac{1}{N}\sum_{i=1}^{N} H(\mathrm{X}_i)\frac{f(\mathrm{X}_i)}{g(\mathrm{X}_i)}] = \frac{1}{N}Var_{g^*}[\ell] = 0$$

という理想的な結果を得る.
$\Longrightarrow$ g^* の式に今推定したい ℓ 自体が含まれているから, このままでは役に立たない!!
$\Longrightarrow$ よく分かっている分布の中から, この g^* と最も近いものを探そう.
$\Longrightarrow$ Cross-Entropy 法

Cross-Entropy(CE) 法

- 重点サンプリング法で得た最適分布 g^* と何らかの意味で近い分布をよく分かっている分布の中から探すことを考えたい.

その 1 つのアプローチが Cross-Entropy 法によるものである．CE 法のアイデアは, 密度 g と h の近さを測る尺度として,

- Cross-entropy

$$D(g,h) = \int_{\mathbf{R}^n} \log \frac{g(\mathrm{x})}{h(\mathrm{x})} g(\mathrm{x}) d\mathrm{x}$$

を導入し, $D(g^*,h)$ を最小にする h を求めれば g^* に近い分布, つまり低分散を実現する分布であると考えるものである.

ここで分布のクラスを以下のように制限する.

- $V \subset \mathrm{R}^n$: パラメータ空間
- $F = \{f(\cdot; \mathrm{v}), \mathrm{v} \in V\}$: よくわかっている分布のクラス
- 特に, 最初に考えていた分布 $f(\cdot)$ を $f(\cdot) = f(\cdot; \mathrm{u}) \in F$ とする.

このようなクラスの中から,

$$\min_{\mathrm{v}} D(g^*, f(\cdot; \mathrm{v}))$$

を attain する v を探す問題を考える. すると, D の定義から

$$D(g^*, f(\mathrm{x}; \mathrm{v})) = \int_{\mathrm{R}^n} g^*(\mathrm{x}) \log g^*(\mathrm{x}) d\mathrm{x} - \int_{\mathrm{R}^n} g^*(\mathrm{x}) \log f(\mathrm{x}; \mathrm{v}) d\mathrm{x}$$

となる.

ここで, 第 1 項は v に無関係なので,

$$\max_{\mathrm{v}\in V}\int_{\mathbf{R}^n} g^*(\mathrm{x})\log f(\mathrm{x};\mathrm{v})d\mathrm{x}$$

を attein する v を求めればいいことになる.

さらに $H(\mathrm{x})\geq 0$ とすると, $g^*=\frac{H(\mathrm{x})f(\mathrm{x};\mathrm{u})}{\ell}$ であったから,

$$\begin{aligned}
&\max_{\mathrm{v}\in V}\int_{\mathbf{R}^n} g^*(\mathrm{x})\log f(\mathrm{x};\mathrm{v})d\mathrm{x}\\
&\qquad=\frac{1}{\ell}\max_{\mathrm{v}\in V}\int_{\mathbf{R}^n} H(\mathrm{x})\log f(\mathrm{x};\mathrm{v})f(\mathrm{x};\mathrm{u})d\mathrm{x}\\
&\qquad=\max_{\mathrm{v}\in V} E_u[H(\mathrm{X})\log f(\mathrm{X};\mathrm{v})]
\end{aligned}$$

と帰着できる. ただし, E_u : $f(\cdot;\mathrm{u})$ に従う X についての期待値とする.

$\Longrightarrow$ **微分法より,**

$$E_u[H(\mathrm{X})\frac{\partial}{\partial \mathrm{v}}\log f(\mathrm{X};\mathrm{v})] = 0$$

をなる v **を求めれば** $\min_{\mathrm{v}} D(g^*, f(\cdot;\mathrm{v}))$ **を** attein **する** v**,**
つまりは低分散を実現する分布を得ることができる.

ここで, さらに解析を進めるために指数族に分布のクラスを制限する. 簡単のために 1 次元の話をする.

- 1 次元指数族とは, $v \in \mathbf{R}$ に対して,

$$f(x;v) = \frac{1}{z} e^{vx} f(x)$$

によって作ることができる確率密度の集合のことである. ただし, z は $z(v) = \int_{\mathbf{R}} e^{vx} f(x) dx = E[e^{vX}]$ とする.

すると確かに, $f(x;v)$ は

$$\int_{\mathbf{R}} f(x;v) dx = \frac{1}{z} \int_{\mathbf{R}} e^{vx} f(x) dx = 1$$

を満たすので新たな確率密度になっている.

いま, 分布 $f(\cdot)$ の平均値を $E[X] = \int_{\mathbf{R}} xf(x)dx = u$,
$f(\cdot; v)$ の平均値を $E_v[X] = \int_{\mathbf{R}} xf(x; v)dx = v$ とする.
ここで, キュムラント母関数を

$$\psi(v) = \log z(v) = \log \int_{\mathbf{R}} e^{vx} f(x)dx$$

とおく. すると, $z = e^{\log \int_{\mathbf{R}} e^{vx} f(x)dx} = e^{\psi(v)}$ より,

$$f(x; v) = e^{vx - \psi(v)} f(x)$$

である. よって,

$$\begin{aligned} \frac{\partial}{\partial v} \log f(x; v) &= \frac{\partial}{\partial v} \log(e^{vx - \psi(v)} f(x)) \\ &= x - \psi'(v) \end{aligned}$$

となる.

さらに,

$$\begin{aligned}\psi'(v) &= \frac{E_u[Xe^{vX}]}{E_u[e^{vX}]} = E_u[Xe^{vX}]e^{-\log E_u[e^{vX}]} \\ &= \int_{\mathbf{R}} xe^{vx-\psi(v)} f(x)dx \\ &= \int_{\mathbf{R}} x\ f(x;v)dx = E_v[X] = v\end{aligned}$$

となる. 以上より,

$$\frac{\partial}{\partial v}\log f(x;v) = x - \psi'(v) = x - v$$

と求められる.

従って, この結果を

$$E_u[H(X)\frac{\partial}{\partial v}\log f(X;v)] = 0$$

に代入することで, 指数族のクラスの中においては

$$E_u[H(X)(X-v)] = 0$$

なる v を求めれば低分散を実現する分布が決まることがわかる. これを解くと, 最適パラメータ v^* として

$$v^* = \frac{E_u[H(X)X]}{E_u[H(X)]}$$

を得る.

従って, 大数の法則から分散の意味で最適な v^* は,
$X_1, \cdots, X_N$,i.i.d, $\sim f(\cdot; u)$ として,

$$\hat{v}^* = \frac{\sum_{i=1}^{N} H(X_i) X_i}{\sum_{i=1}^{N} H(X_i)}$$

によって推定できる.

v,u,X を多次元化しても同様に, $\mathrm{v}^* = (v_1, \cdots, v_n)$ に対して,

$$v_j^* = \frac{E_{\mathrm{u}}[H(\mathrm{X})X^{(j)}]}{E_{\mathrm{u}}[H(\mathrm{X})]} \quad (j = 1, \cdots, n)$$

が成り立つ. 実際の推定においては v^* の推定量 $\hat{\mathrm{v}}^*$ として,

$$\hat{v}_j^* = \frac{\sum_{i=1}^{N} H(\mathrm{X}_i)X_i^{(j)}}{\sum_{i=1}^{N} H(\mathrm{X}_i)}$$

を用いる.

レアイベントに対する CE 法

- $\mathbf{X} = (X_1, \cdots, X_n)$: R^n-値の確率変数
- H : $\mathrm{R}^n \to \mathrm{R}$ のある関数 (sample performance)
- $f(\cdot; \mathrm{u}) \in F = \{f(\cdot; \mathrm{v}), \mathrm{v} \in V\}$: $\mathbf{X}$ の結合密度関数
- $\{S(\mathbf{X}) \geq \gamma\}$: レアイベントであるとする.

このとき, ある定めた γ に対して

$$\ell = P_u(S(\mathbf{X}) \geq \gamma) = E_u[1_{\{S(\mathbf{X}) \geq \gamma\}}] << 1$$

を推定することを考える.

先程の議論から, $H(X) = 1_{\{S(\mathrm{X}) \geq \gamma\}}$ のとき, 最適パラメータ v^* は $X_1, \cdots, X_N, i.i.d., \sim f(\cdot; \mathrm{u})$ とし,

$$\hat{v}^* = \frac{\sum_{i=1}^{N} 1_{\{S(\mathrm{X}_i) \geq \gamma\}} X_i}{\sum_{i=1}^{N} 1_{\{S(\mathrm{X}_i) \geq \gamma\}}}$$

と推定できる. これは, 分母分子がそれぞれ大数の法則によって,

$$\begin{aligned} \lim_{N \to \infty} \hat{v}^* &= \lim_{N \to \infty} \frac{\sum_{i=1}^{N} 1_{\{S(\mathrm{X}_i) \geq \gamma\}} X_i}{\sum_{i=1}^{N} 1_{\{S(\mathrm{X}_i) \geq \gamma\}}} \\ &= \frac{E_u[1_{\{S(\mathrm{X}_i) \geq \gamma\}} X_i]}{E_u[1_{\{S(\mathrm{X}_i) \geq \gamma\}}]} = v^* \quad (P - a.s.) \end{aligned}$$

が成り立っていることから, 定められたものであった.

しかし今, $\{S(\mathrm{X}) \geq \gamma\}$ はレアイベントであるから,
$1_{\{S(\mathrm{X}) \geq \gamma\}}$ が 1 をとる確率はほとんどない.

$\Longrightarrow$ 大数の法則が働くのに著しく多くのサンプルが必要になってしまう.

$\Longrightarrow$ これを解決するために, γ と v の 2 つのパラメータを
$\mathrm{Adaptive}$ に update する方法を述べる.

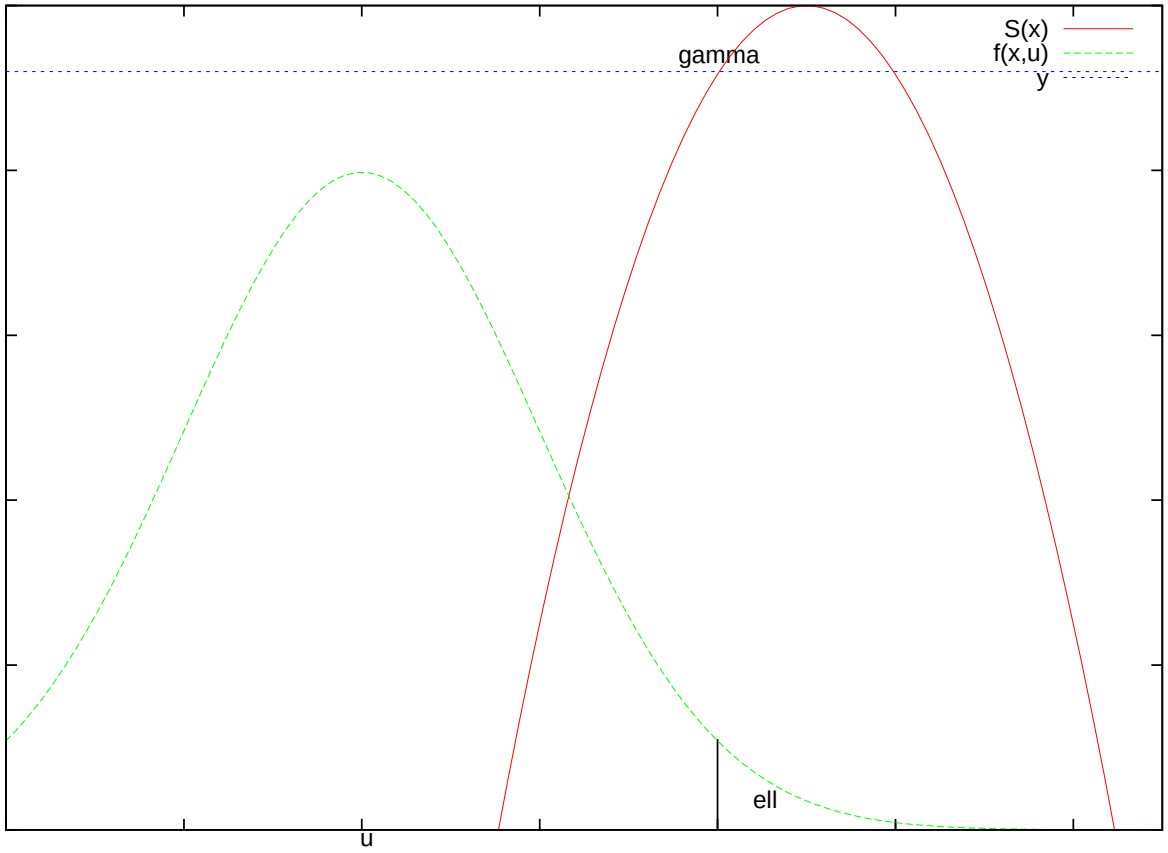


Figure 1: $\ell = P_u(S(\mathrm{X}) \geq \gamma)$ のイメージ

1 次元指数族に対する Adaptive IS のアルゴリズム:

1. **小さすぎない数 ρ を $\frac{1}{100} \leq \rho \leq \frac{1}{10}$ と決め, u を勝手に定める.**

2. **γ_1 を,**

$$P_u(S(X) \geq \gamma_1) = \rho$$

をみたすものとして決める.

実際には, γ_1 の推定量 $\hat{\gamma}_1$ を代わりに用いる. その方法としては十分大きな N に対して, $X_1, \cdots, X_N$,i.i.d, $\sim f(\cdot; u)$ とし, 全ての i について $S(X_i)$ を計算し, $S_{(1)}, \cdots, S_{(N)}$ と小さい順に並べ直す.

そして, 小さい方から $(1-\rho)N$ 番目の値,

$$\hat{\gamma}_1 = S_{(\lceil (1-\rho)N \rceil)}$$

を推定量 $\hat{\gamma}_1$ を γ_1 としてとる. これは大数の法則より, N が大きいとき, $\hat{\gamma}_1 \sim \gamma_1$ $(P-a.s.)$ となるから正当化される.

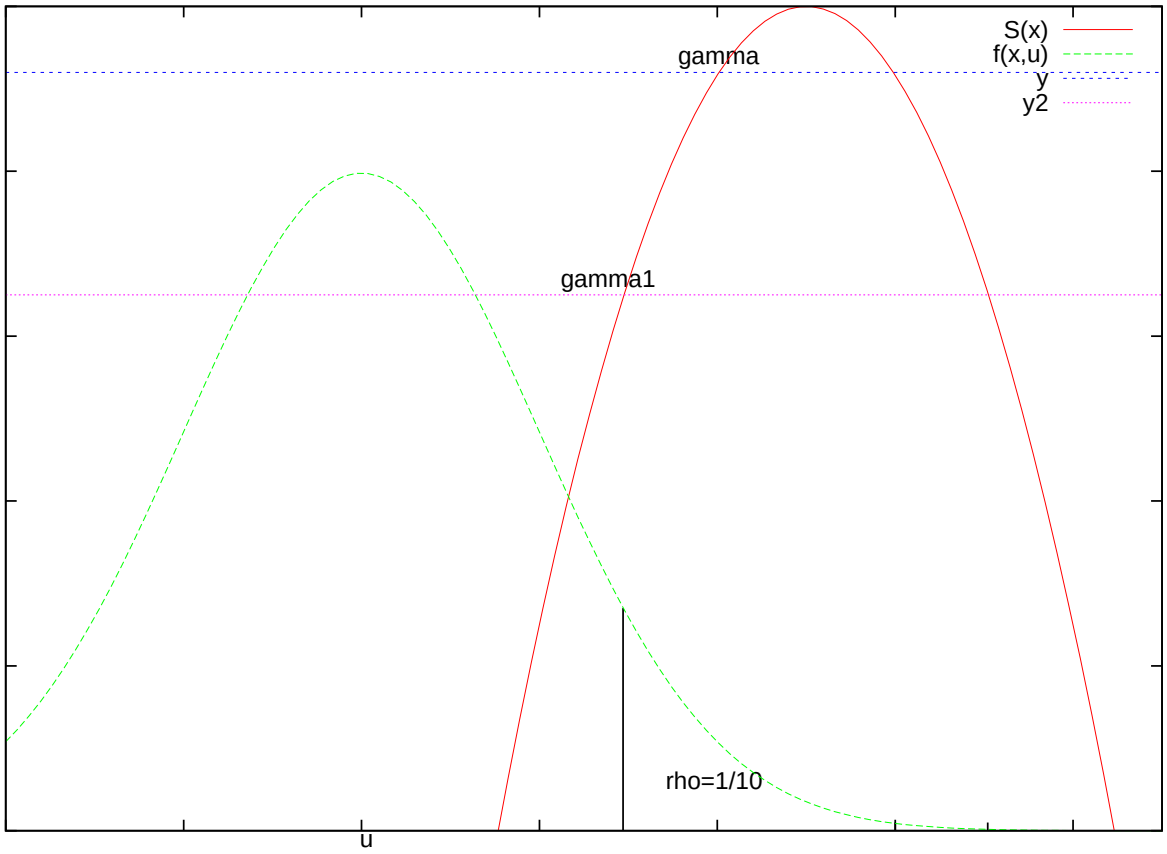


Figure 2: $P_u(S(X) \geq \gamma_1) = \rho = \frac{1}{10}$ **の取り方**

3. この $\hat{\gamma}_1$ の取り方から極端に大きくない N に対しても大数の法則が効くので CE 法によって v_1 が定めることができる. つまり, $X_1, \cdots, X_N, i.i.d, \sim f(\cdot; u)$(2 で生成したサンプル) として,

$$\hat{v}_1 = \frac{\sum_{i=1}^{N} \mathbf{1}_{\{S(\mathbf{X}_i) \geq \gamma_1\}} X_i}{\sum_{i=1}^{N} \mathbf{1}_{\{S(\mathbf{X}_i) \geq \gamma_1\}}}$$

を推定量 $\hat{v}_1$ を決め v_1 としてとる.

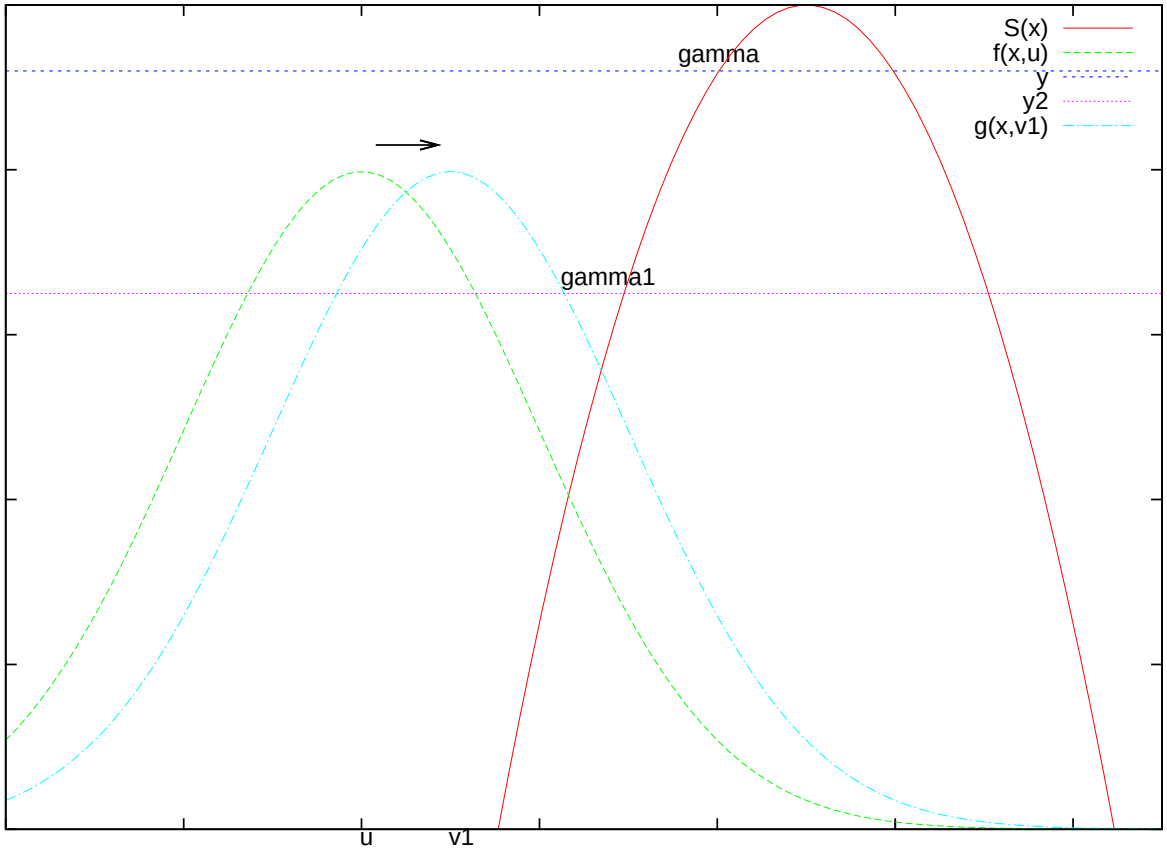


Figure 3: $v_1 = \frac{E_u[1_{\{S(\mathbf{x}_i)\geq\gamma_1\}}\mathbf{X}_i]}{E_u[1_{\{S(\mathbf{x}_i)\geq\gamma_1\}}]}$ **の取り方**

4. さらに, γ_2 を, 2 と同様に,

$$P_{v_1}(S(X) \geq \gamma_2) = \rho$$

をみたすものとして, $X_1, \cdots, X_N, i.i.d, \sim f(\cdot; v_1)$ として, $\hat{\gamma}_2 = S_{(\lceil (1-\rho)N \rceil)}$ から決める.

5. $X_1, \cdots, X_N, i.i.d, \sim f(\cdot; v_1)$(4 で生成したサンプル) として各要素を,

$$\hat{v}_2 = \frac{\sum_{i=1}^{N} 1_{\{S(X_i) \geq \gamma_2\}} W(X_i; u, v_1) X_i}{\sum_{i=1}^{N} 1_{\{S(X_i) \geq \gamma_2\}} W(X_i; u, v_1)}$$

と決め推定量 $\hat{v}_2$ を v_2 としてとる. ただし, W は likelihood ratio とよばれるものであって,

$W(X_i; u, v_1) = \dfrac{f(X_i; u)}{f(X_i; v_1)}$ である.

ここで新たに W の項が出てきたのは, $\hat{v}$ が N が大きいとき最適パラメータ v^* に近づくように決めたいからである.
すなわち, 大数の法則により,
$X_1, \cdots, X_N, i.i.d., \sim f(\cdot; v_1)$ とするとき,

$$\begin{aligned}
\hat{v}_2 &= \frac{\sum_{i=1}^{N} 1_{\{S(X_i)\geq\gamma_2\}} W(X_i; u, v_1) X_i}{\sum_{i=1}^{N} 1_{\{S(X_i)\geq\gamma_2\}} W(X_i; u, v_1)} \\
&\to \frac{E_{v_1}[1_{\{S(X_i)\geq\gamma_2\}} \frac{f(X_i;u)}{f(X_i;v_1)} X_i]}{E_{v_1}[1_{\{S(X_i)\geq\gamma_2\}} \frac{f(X_i;u)}{f(X_i;v_1)}]} \quad (N \to \infty) \\
&= \frac{E_u[1_{\{S(X_i)\geq\gamma_2\}} X_i]}{E_u[1_{\{S(X_i)\geq\gamma_2\}}]} = v^* \quad (P-a.s.)
\end{aligned}$$

を満たすように決めたいからである.

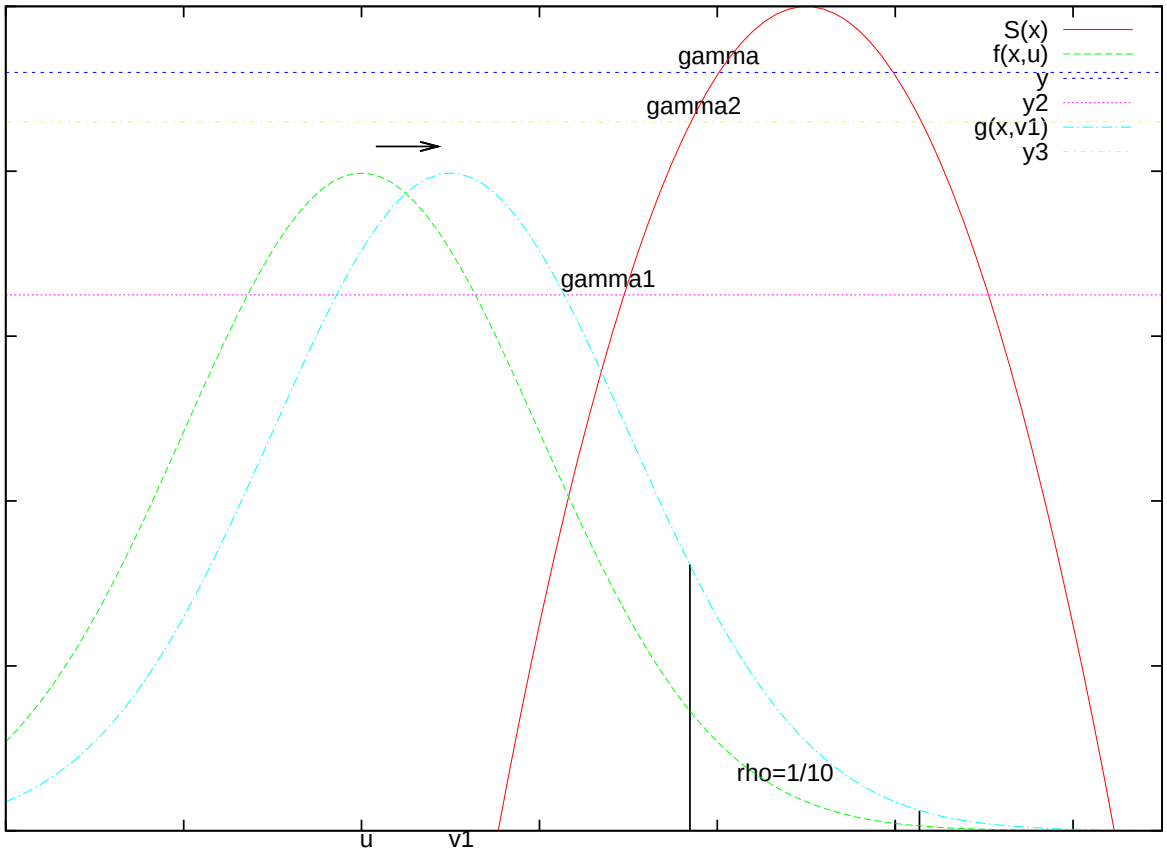


Figure 4: $P_{v_1}(S(X) \geq \gamma_2) = \rho = \frac{1}{10}$ のイメージ

6. これを $\gamma_t \geq \gamma$ となるまで t 回繰り返す.

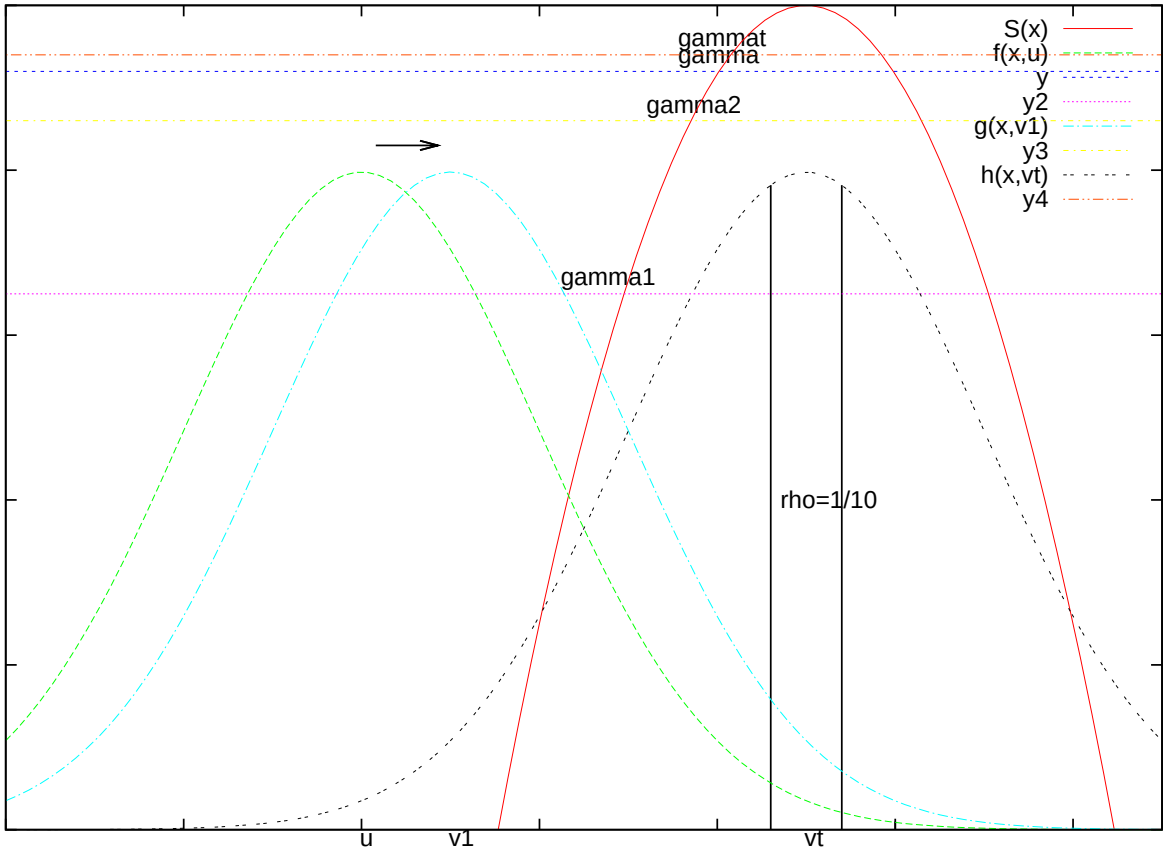


Figure 5: v_t の収束のイメージ

7. ある $t = T$ で $\gamma_t \geq \gamma$ となったとする．あとは，$\gamma_T = \gamma$ ととり，$X_1, \cdots, X_N, i.i.d, \sim f(\cdot; v_T)$ として，

$$\hat{\ell}_{IS} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{S(X_i) \geq \gamma\}} \frac{f(X_i; u)}{f(X_i; v_T)}$$

に従って ℓ を推定すればよい．

Adaptive Sampling のまとめ

- 小さい確率に対する推定においても, 常に大数の法則が働く形でサンプリングを行うことができる.
- サンプリングのみによって, 最適なパラメータを探していく方法であり, 期待値が理論的に計算できない, つまり最適パラメータ v^* が全く分からなくても使う事ができる.

Coin Flipping

公正なコインを 100 回振って表が出る数が 65 回以上となる確率 P を推定したい.

$\Longrightarrow X_1, \cdots, X_N, i.i.d, \sim Bin(100, u)$, $u = \frac{1}{2}$ とするとき,

$$P = P(X_i \geq 65) = E_u[1_{\{X_i \geq 65\}}]$$

を推定する問題を考えたい. これは, 解析的に導くと,

$$P = P(X_i \geq 65) = \sum_{k=65}^{100} P(X_i = k) \sim 0.0017..$$

と求められる.

この時, CMC, IS, Adaptive IS の 3 つの方法について同じサンプル数 N でどれだけよい近似をしてくれるかをシミュレートしてみよう.

- CMC による方法

$X_1, \cdots, X_N, i.i.d, \sim Bin(100, u), u = \frac{1}{2}$ として,

$$\hat{P}_{CMC} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i \geq 65\}}$$

と推定する.

- **IS による方法**

$X_1, \cdots, X_N, i.i.d, \sim Bin(100, v^*), v^* = 0.65$ **として,**

$$\hat{P}_{IS} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i \geq 65\}} \frac{f(X_i; u)}{f(X_i; v^*)}$$

と推定する.

ただし,

$f(X_i; u) = \frac{1}{2^{100}}$, $f(X_i; v^*) = (v^*)^{X_i}(1 - v^*)^{(100 - X_i)}$

である.

- Adaptive IS による方法

1, $\rho = \frac{1}{10}$, $u = \frac{1}{2}$, $v_0 = u$ とする.

2, γ_t を $X_1, \cdots, X_N, i.i.d, \sim Bin(100, v_{t-1})$ として, 各複製 N に関して, 100 回のコイン投げで表が出た回数を求めて小さいもの順にならべる. そして, 下から $(1-\rho)N$ 番目の値としてとる.

3, v_t を, 2 で生成したサンプルから,

$$\hat{v}_t = \frac{\sum_{i=1}^{N} 1_{\{X_i \geq \gamma_t\}} \frac{f(X_i;u)}{f(X_i;v_{t-1})} X_i}{\sum_{i=1}^{N} 1_{\{X_i \geq \gamma_t\}} \frac{f(X_i;u)}{f(X_i;v_{t-1})}}$$

と推定する. ただし, $f(X_i; u) = \frac{1}{2^{100}}$,
$f(X_i; v_{t-1}) = (v_{t-1})^{X_i}(1 - v_{t-1})^{(100-X_i)}$ である.

4, 各 $t = 1, 2, ..$ について以上を繰り返して, $\gamma_t \geq 65$ を満たしたら $t = T$ において, $X_1, \cdots, X_N, i.i.d, \sim Bin(100, v_T)$ として,

$$\hat{P}_{AdaIS} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i \geq 65\}} \frac{f(X_i; u)}{f(X_i; v_T)}$$

と P を推定する.

最適化問題へのアプローチ

ある集合 χ の中での関数 $S(\mathbf{x})$ の最大化,

$$\gamma^* = \max_{\mathbf{x} \in \boldsymbol{\chi}} S(\mathbf{x})$$

という問題を解きたいとする. つまり, γ^* を求める.

このとき, この問題のかわりに,

$$\ell(\gamma) = P_{\mathbf{u}}(S(\mathbf{X}) \geq \gamma) = E_{\mathbf{u}}[1_{\{S(\mathbf{X}) \geq \gamma\}}]$$

を考えることで γ^* を決めようというアイデアである.

finding max

関数の最大値を見つけるプログラムを $X_i, i.i.d, \sim N(0, 10)$
から始まる Adaptive IS **を用いて行う.**

$\Longrightarrow \gamma^* = \max_{x} S(x)\ (= 500)$ **なる** γ^* **を探す.**

- $S(x) = 500|\sin \frac{x}{100}|$

についてシミュレートを行う.

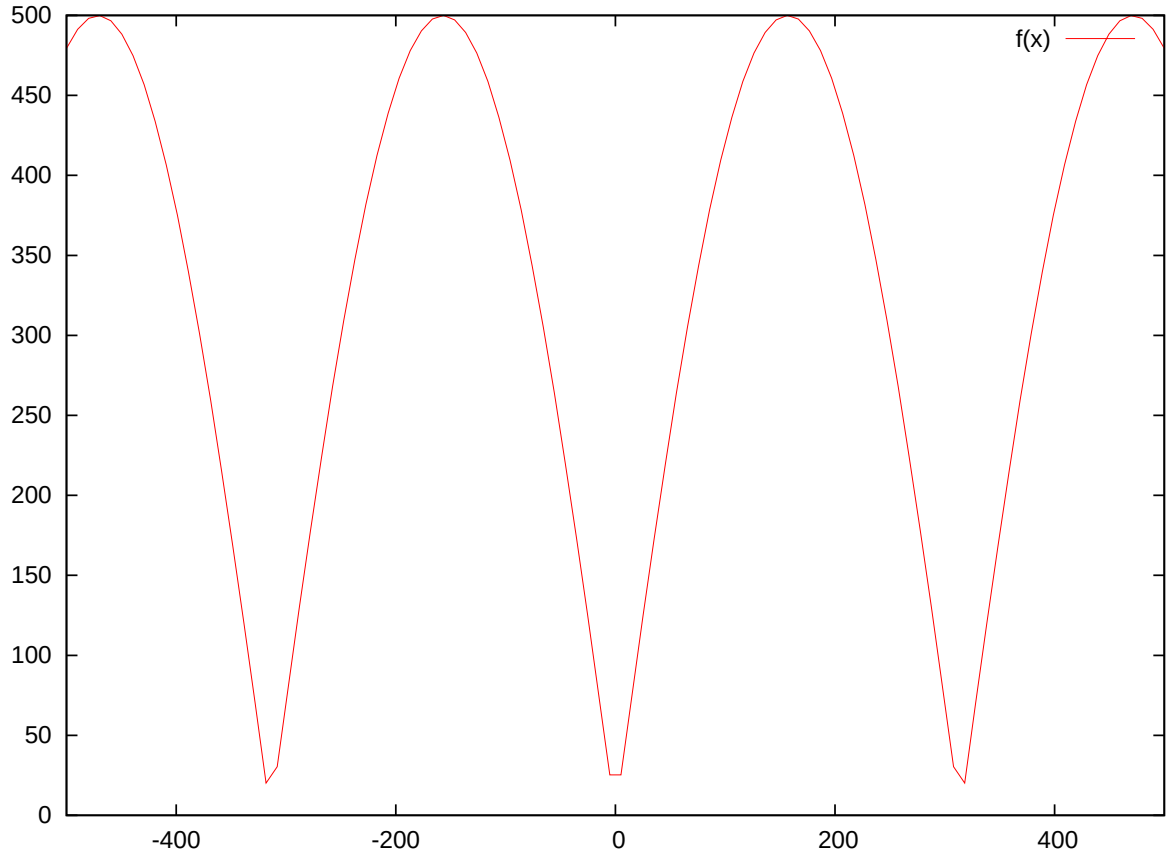


Figure 6: $S(x) = 500|\sin\frac{x}{100}|$ のグラフ

まとめ

実際の社会においてシステムを作る時, レアイベントの確率が問題になってくる. また解析も難しい. 例, 通信におけるエラー率, 倒産のリスク問題 etc.

$\Longrightarrow$ モンテカルロ法によって推定することを考える. しかし, レアイベントの推定において莫大なサンプル数が必要になる. それには, 推定量の分散が大きくなることが原因であった.

$\Longrightarrow$ 推定量の分散を下げる方法を取り入れたい. そのひとつの解決策が, 重点サンプリング法である. しかし, 重点サンプリング法において解析的に導いた理想的な分布 $g*$ はそのままでは役に立たない.

$\Longrightarrow$ **CE 法を用いて, 理想分布 g^* との Cross-Entropy 距離が小さい $h = h^*$ を取れば低分散を実現できているのではないか. 特に分布のクラスが指数族の場合にはさらなる解析が可能になった.**

$\Longrightarrow$ **特に推定したい確率が小さいときには, CE 法を v と γ を Adaptive に update する方法で極端に大きくないサンプル数でも大数の法則が働くように改良した. すると確かに, 少ないサンプリング数で CMC と比べてよい近似が得られた.**

References

[1] **R.Y.Rubinstein and D.P.Kroese, Simulation and the Monte Carlo Method second edition, WILEY, (2007)**

[2] **S.Asmussen and P.W.Glynn, Stochastic Simulation, Springer, (2007)**

ご静聴ありがとうございました。