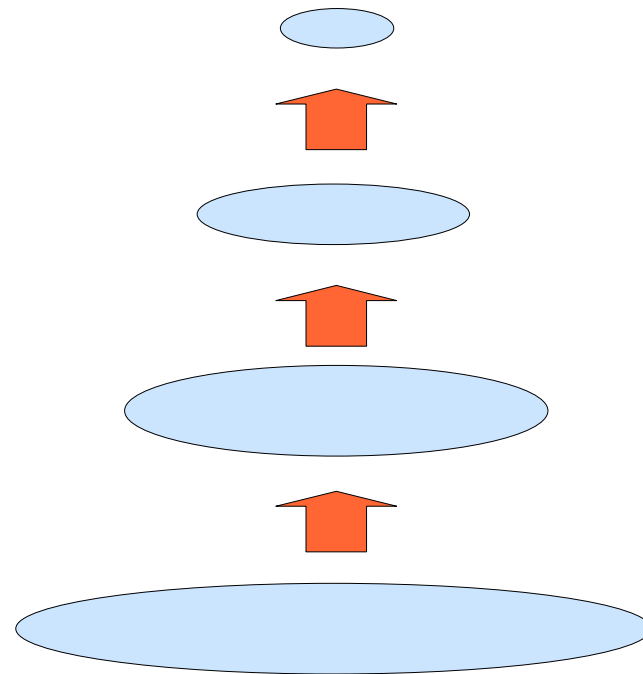


# 比護 貴之

-

- $$q(x) = \frac{e^{-f(x)\beta}}{Z}$$

## $\beta = \infty$ :最適解

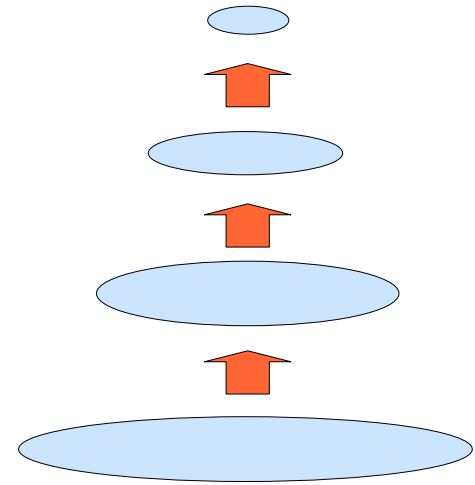
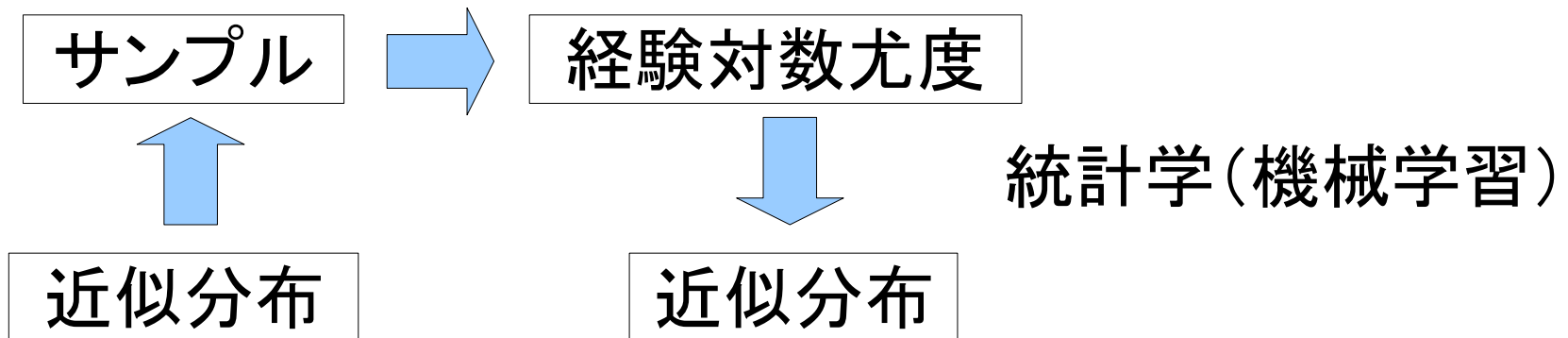


# 近似分布 $p(x)$

Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria A

- 目標分布を近似する
- インポートランスサンプリング
  - 分散が推定精度に影響する
  - $p(x)$ と $q(x)$ が似てる方がよい

$$\int q(x) \log \hat{p}(x) dx \simeq \frac{1}{N} \sum \frac{q(x)}{p(x)} \log \hat{p}(x)$$

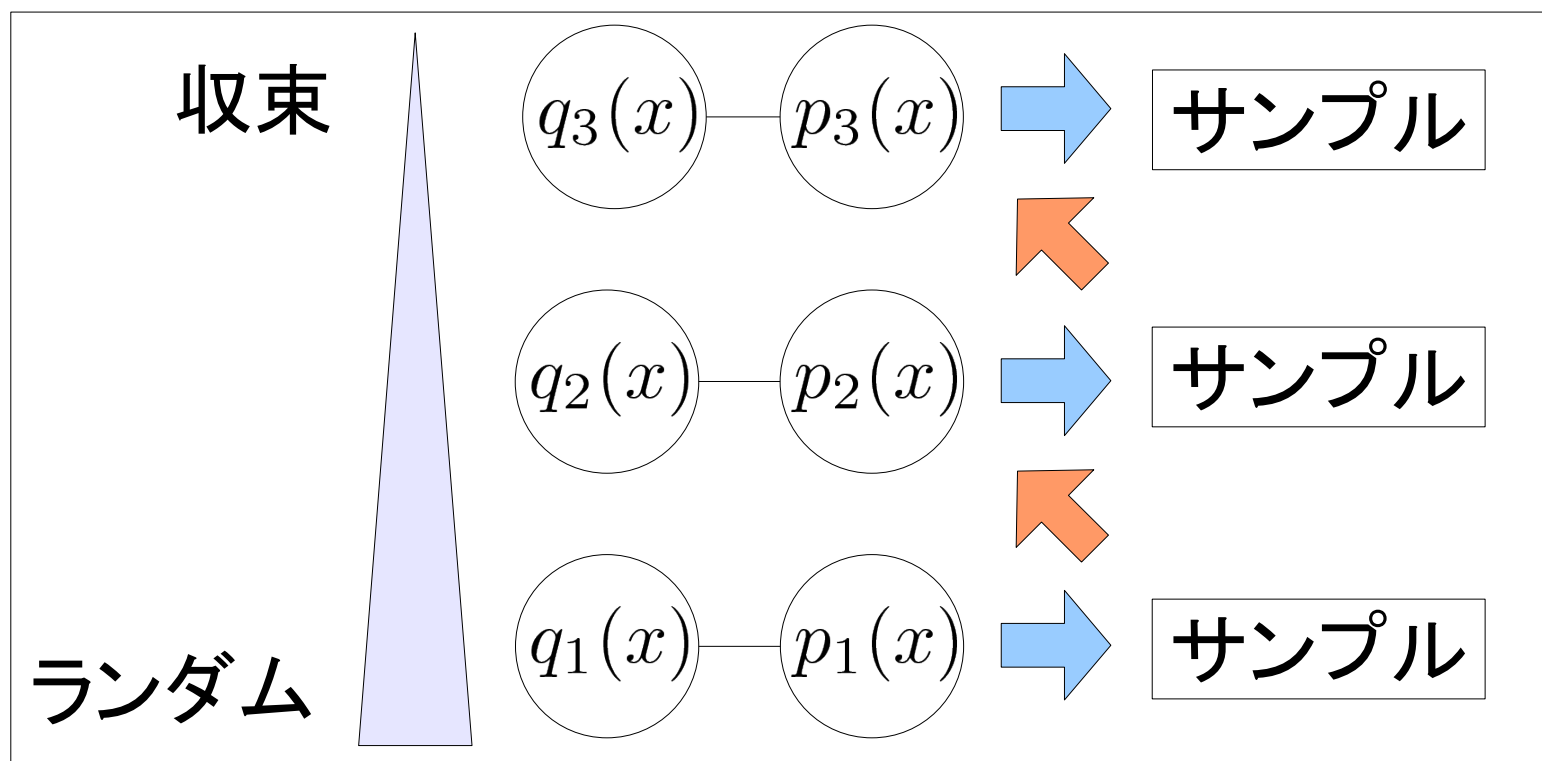


# アニーリング

- アニーリング＝ゆっくり収束させる


- 分散を抑える効果

$$\frac{1}{N} \sum_{p_1(x)} \frac{q_2(x)}{p_1(x)} \log \hat{p}(x|\theta)$$



- $$\int q(x) \log \hat{p}(x) dx \simeq \frac{1}{N} \sum \frac{q(x)}{p(x)} \log \hat{p}(x)$$

- ## ・3つの方法で改良します！！

- 
- ## スケジュール 手法3





- # 例

## GAで良い解だけを集めると多様性が失われ収束する

# 提案手法の概要: RPM

- 集めたサンプル分布の近似計算法の確立
  - インポートランスサンプリングが適用可能になる  
→ 偏りを消去できる
- 基本操作
  - 重み付け
    - 集めた過去サンプルの分布を記憶する
  - リサンプリング
    - 分布を変えずに集めた過去サンプルのサイズを変更

## • 2D イジング模型

$$x = \{0, 1\}$$

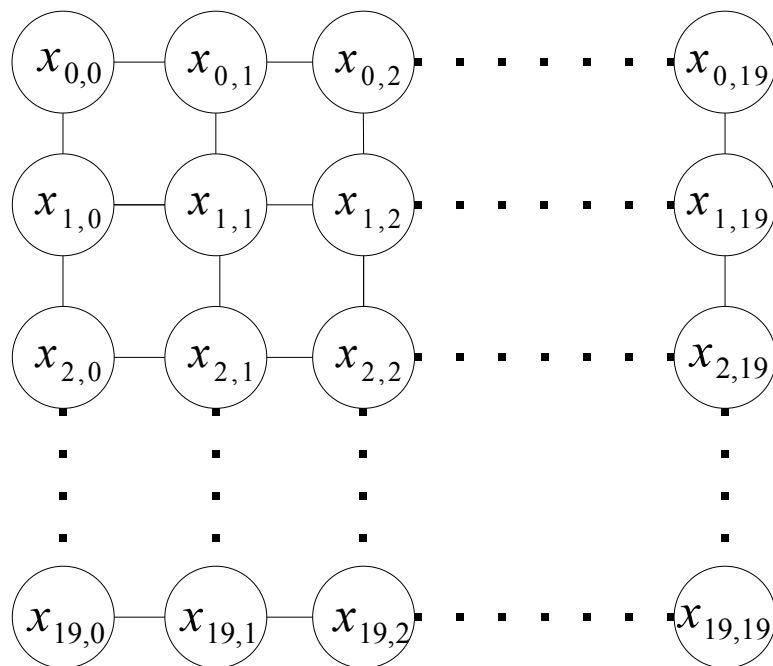
$$f(x) = - \sum_{i=0}^{19} \sum_{j=0}^{19} \{ J(x_{ij}, x_{i+1,j}) + J(x_{ij}, x_{i,j+1}) \}$$

$$J(x_i, x_j) = \begin{cases} 1 & x_i = x_j \\ 0 & x_i \neq x_j \end{cases}$$

– 20×20

## - 周期境界条件

- 最小值: -800



- 提案手法(RPM) vs 従来法(IDEAとEDA)

- EDA: 過去サンプルを利用しない
- IDEA: ヒューリスティックな過去サンプル利用法
- 確率モデル: 全変数独立

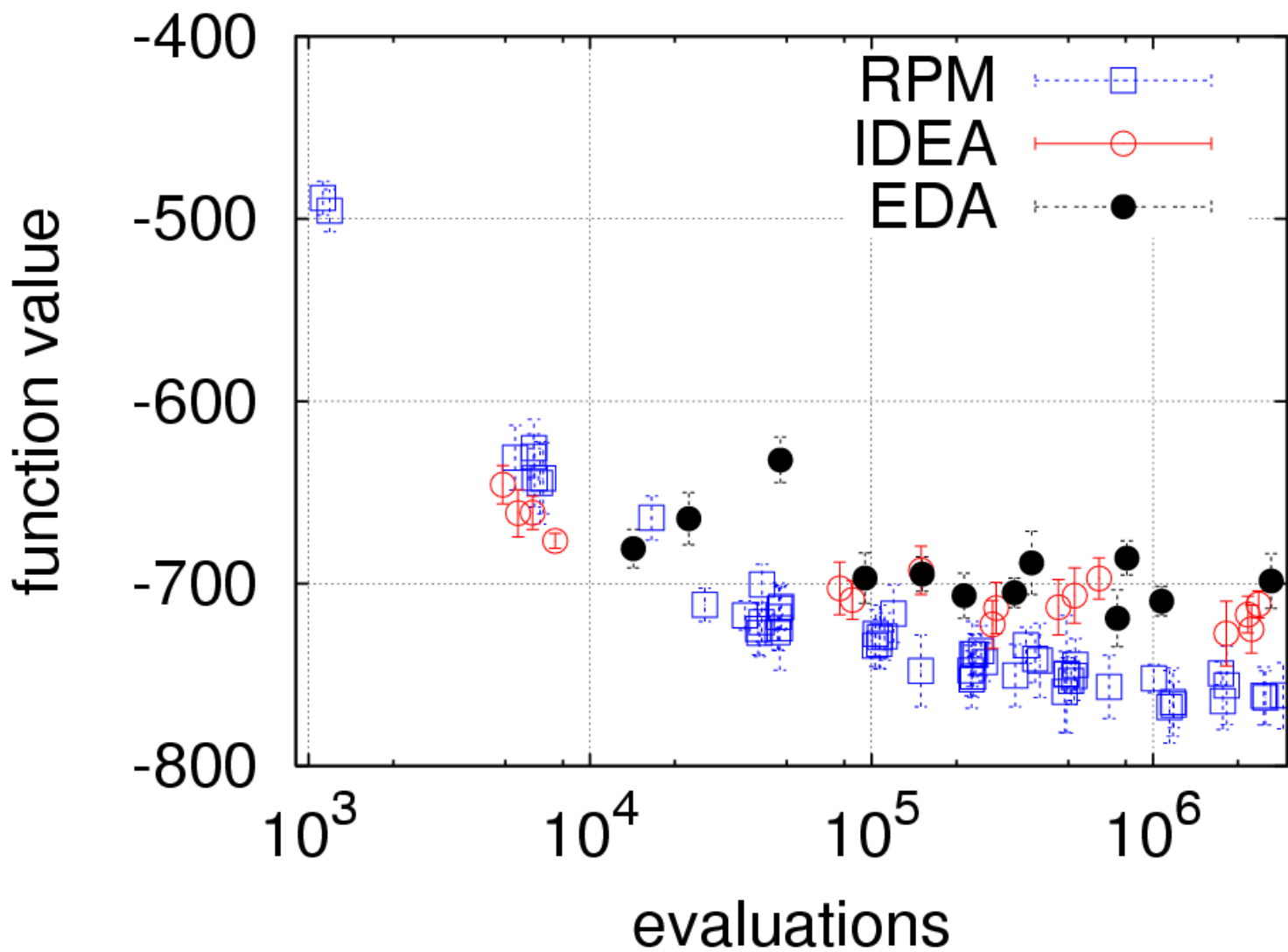
$$p(x|w) = \prod_i p(x_i|w_i)$$

- パラメータ

- 同じ条件になるように設定

それぞれの組合せに対して10試行

# 結果

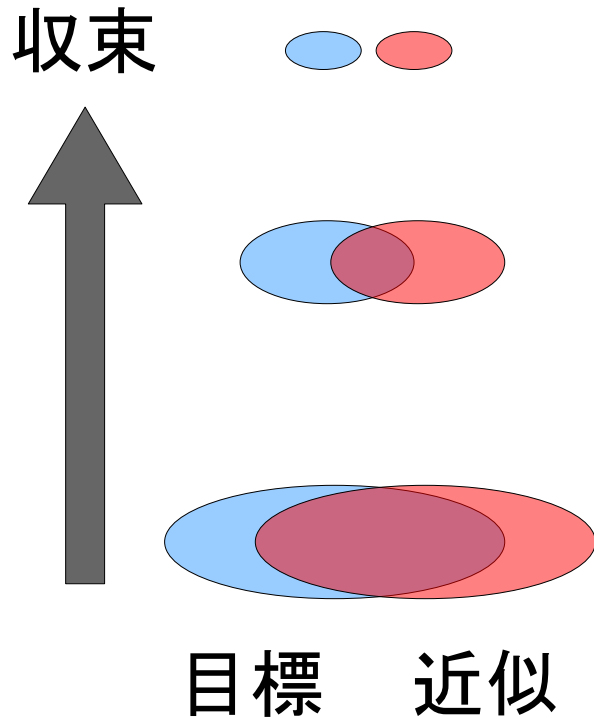






## ● 局所解の問題

- 目標分布と近似分布がずれるとインポートランスサンプリングで有効なサンプルが生成されない



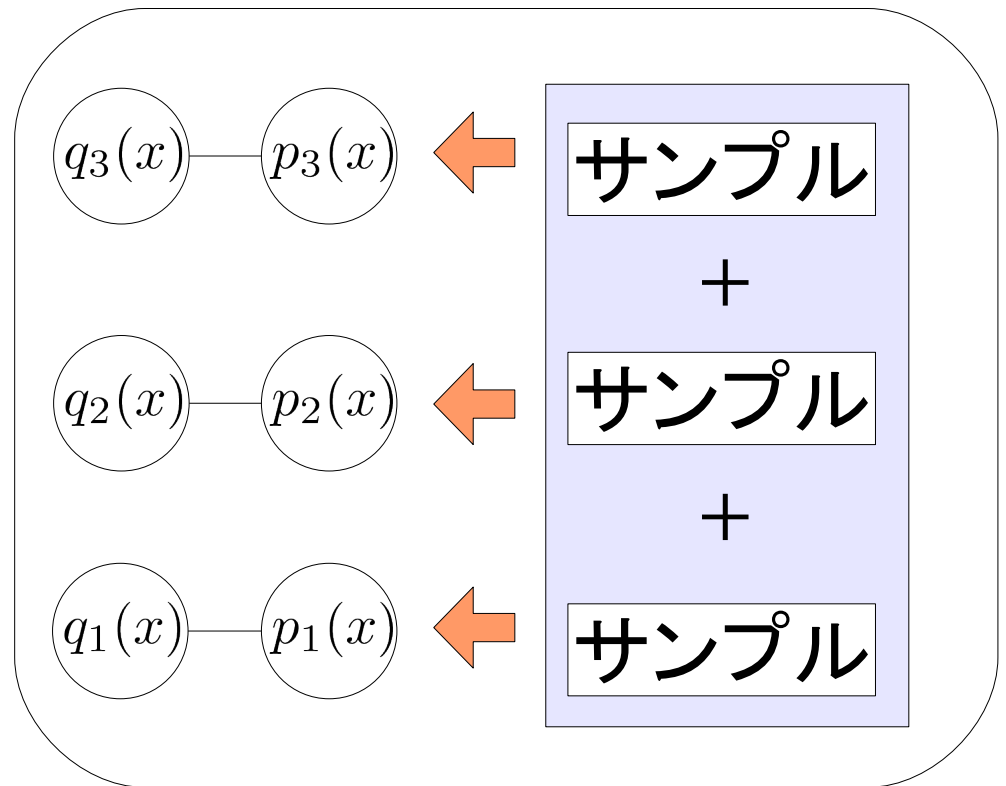
## 重なっている部分が有効なサンプル



- 

$$\frac{1}{2}p_l(x) + \frac{1}{2}p_{l-1}(x)$$

- ## (1) サンプリング



- 2D イジング模型

$$x = \{0, 1\}$$

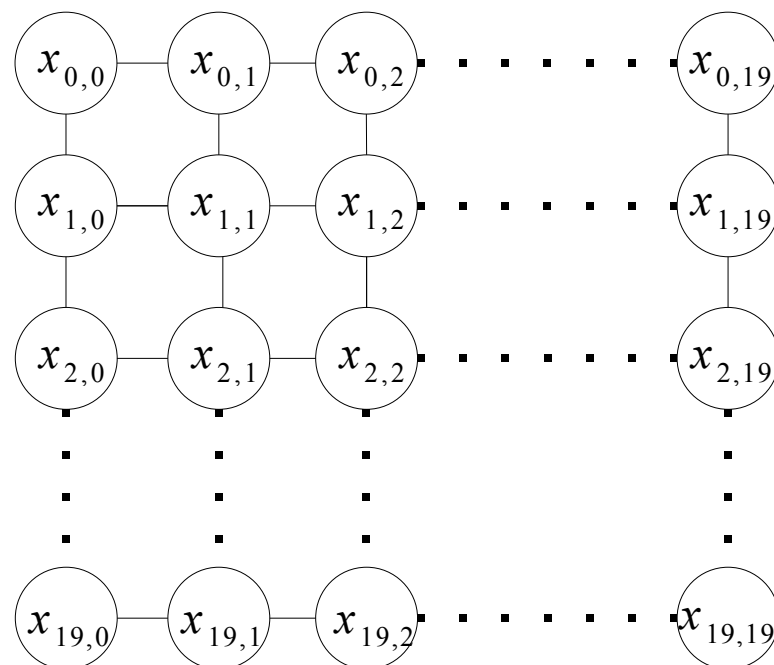
$$f(x) = - \sum_{i=0}^{19} \sum_{j=0}^{19} \{J(x_{ij}, x_{i+1,j}) + J(x_{ij}, x_{i,j+1})\}$$

$$J(x_i, x_j) = \begin{cases} 1 & x_i = x_j \\ 0 & x_i \neq x_j \end{cases}$$

– 20×20

## – 周期境界条件

- 最小值: -800



## 確率モデル: 全変数独立

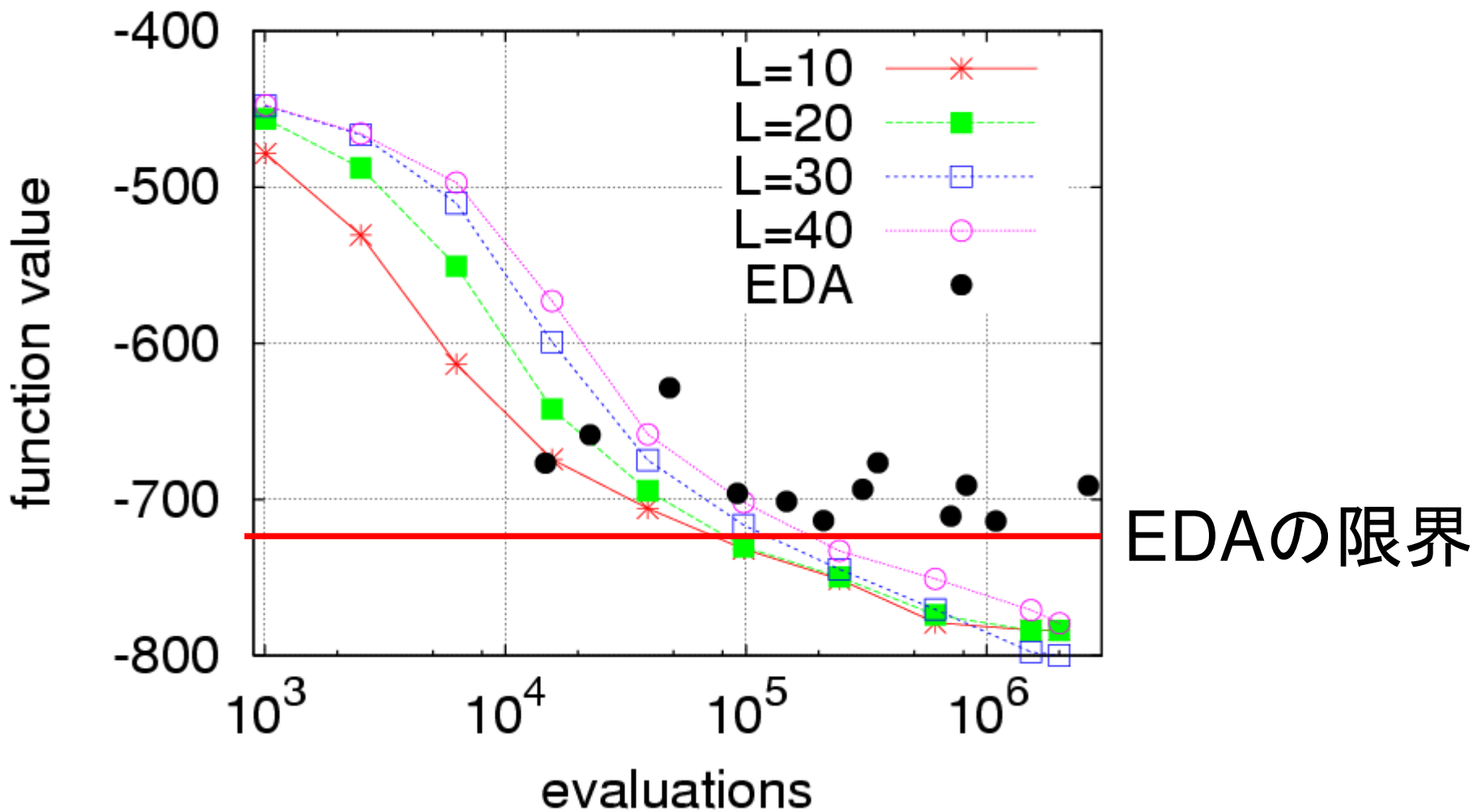
- パラメータ

- EDA: サンプル数=100,500,  
1000,3000

Diagram illustrating a Markov chain structure with three rows, each representing a step in the chain:

- Top row:  $q_3(x)$  and  $p_3(x)$  are connected by a horizontal line, followed by a blue arrow pointing to a box labeled "サンプル".
- Middle row:  $q_2(x)$  and  $p_2(x)$  are connected by a horizontal line, followed by a blue arrow pointing to a box labeled "サンプル".
- Bottom row:  $q_1(x)$  and  $p_1(x)$  are connected by a horizontal line, followed by a blue arrow pointing to a box labeled "サンプル".


# 結果





-

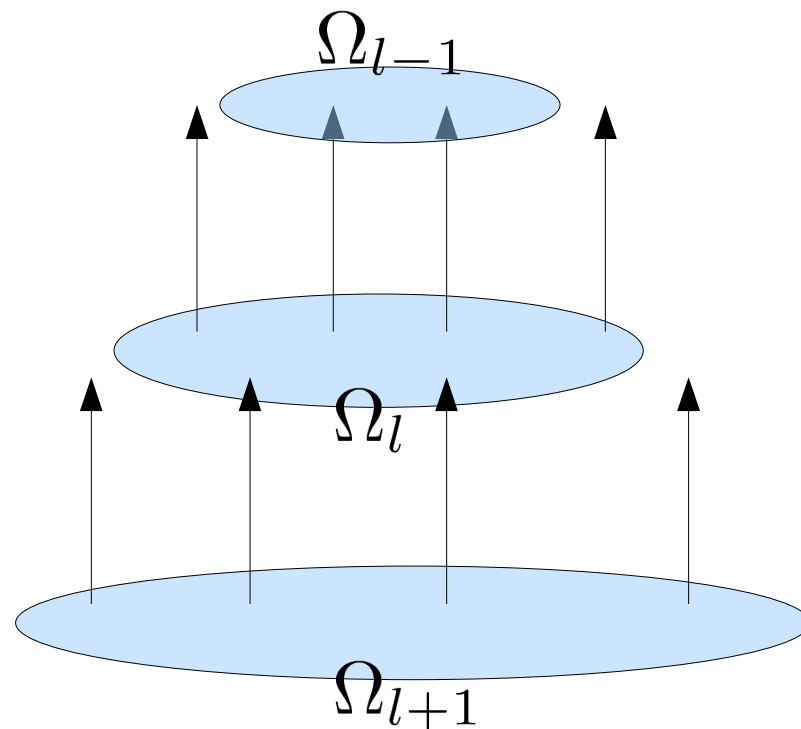
- $$\min \text{Var} \left( \frac{q_{t+1}(x)}{q_t(x)} \log \hat{p}(x) \right)_{q_t(x)}$$



エントロピーを線形で減少させればよい

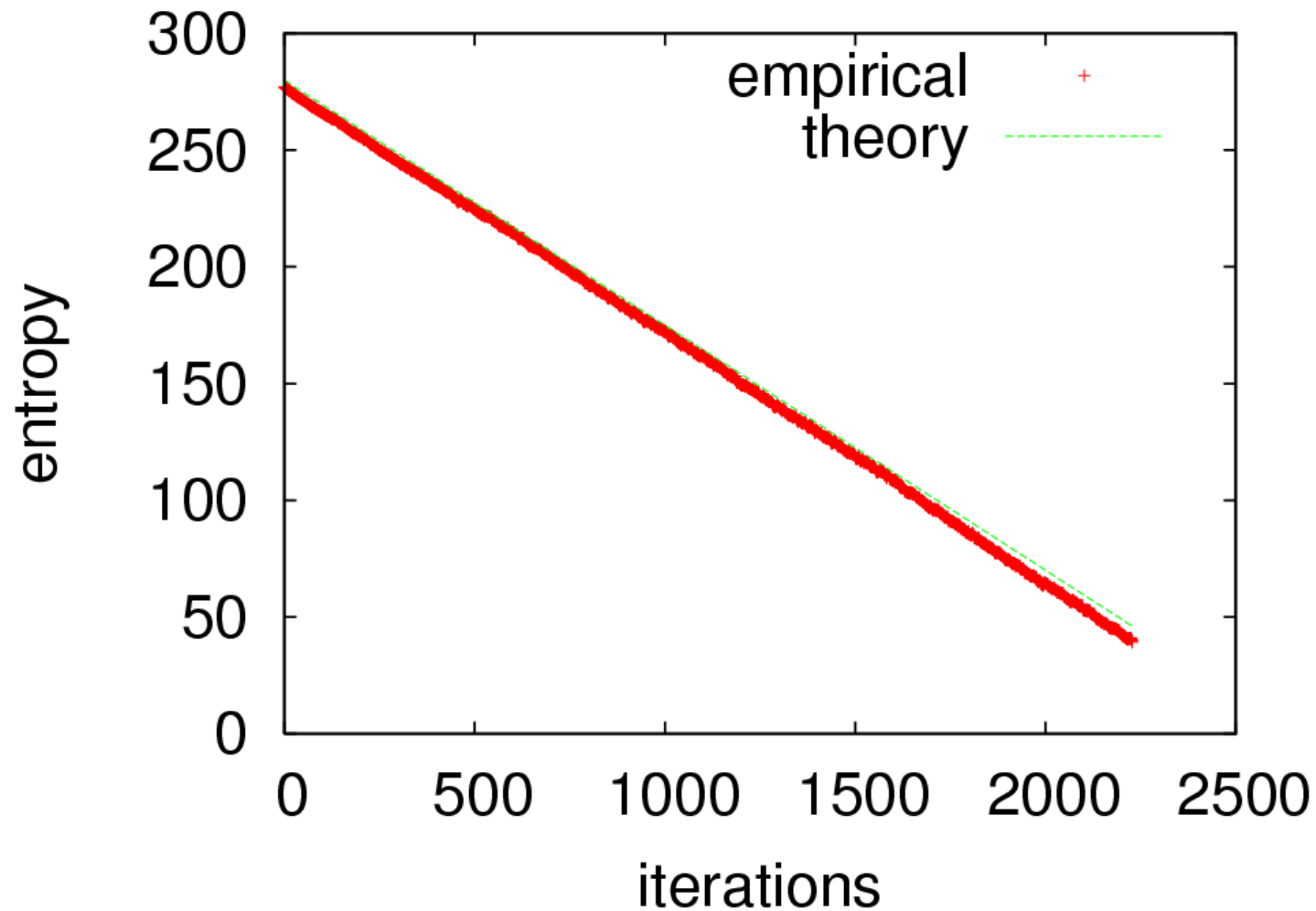


- ## 指数オーダーの解空間に対して 線形時間で収束



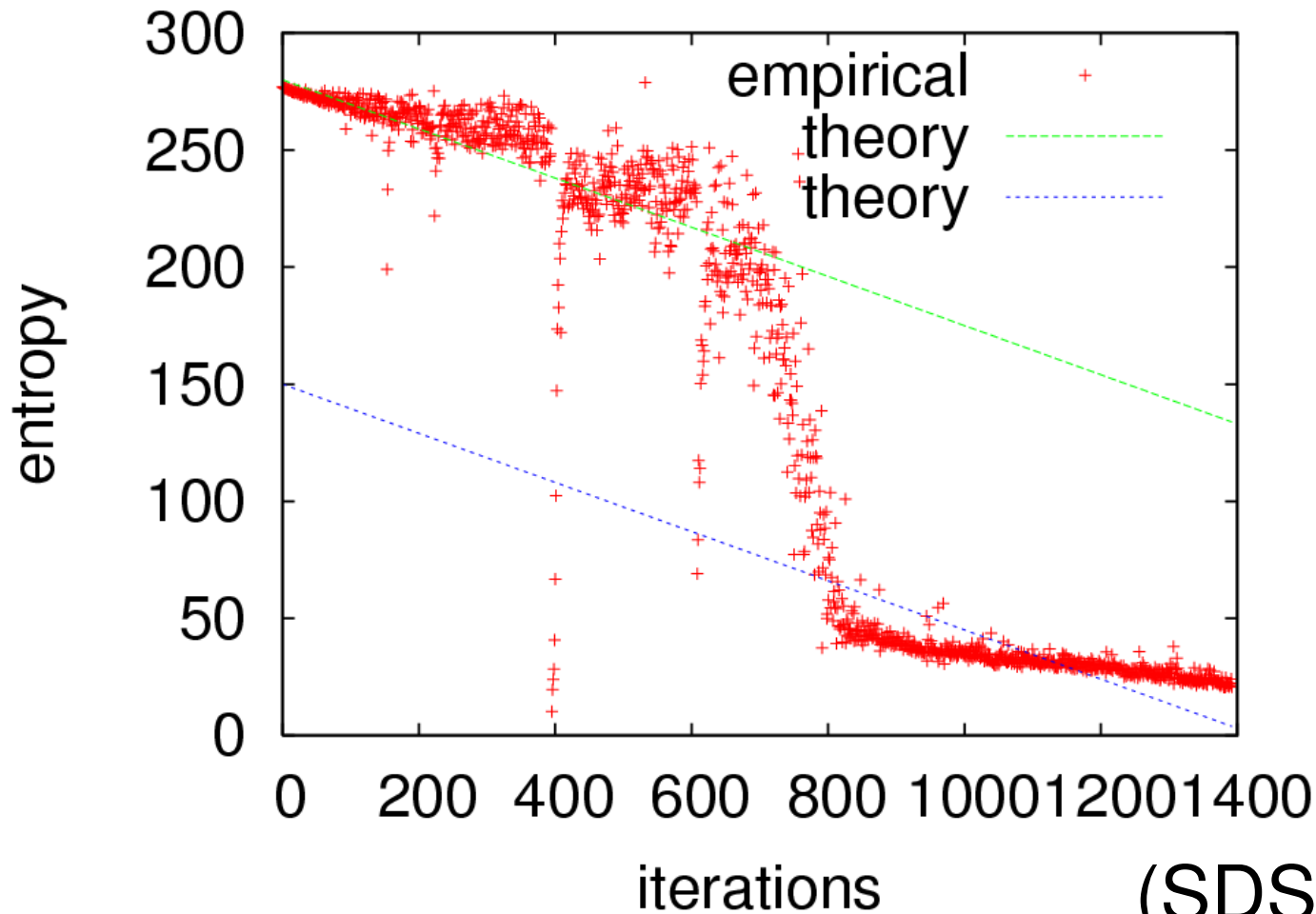
# 実験

- 基本性質調査
    - Onemax: 簡単
    - 2Dイジング: 複雑
  - 先行研究との比較
    - Standard Deviation Schedule
- [Mhanig and Muhlenbein 2001]



# 2D イジング

- 一定の傾きでエントロピーを減少させたい  
(傾きはパラメータとして与えている)
- 近似的にしか実現できない



(SDSもほぼ同じ軌跡)

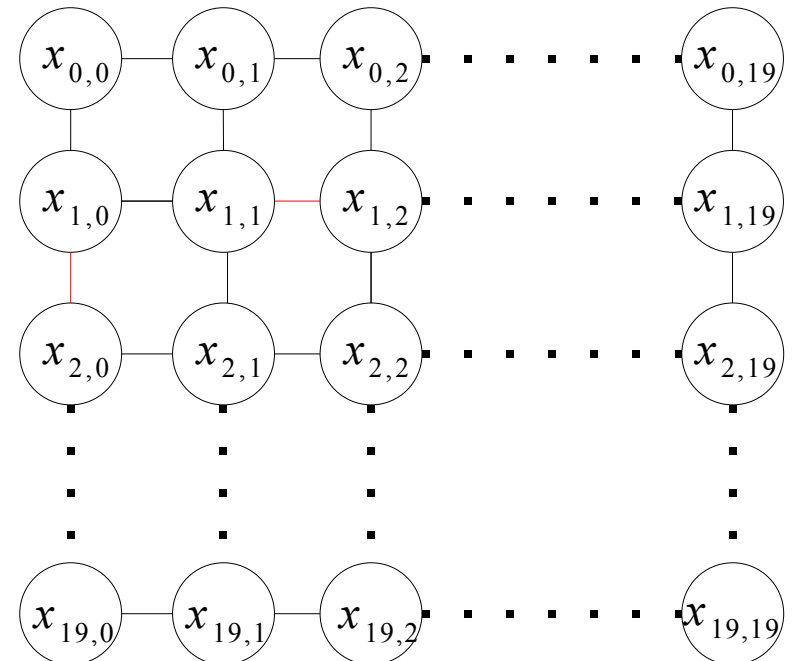
- function value



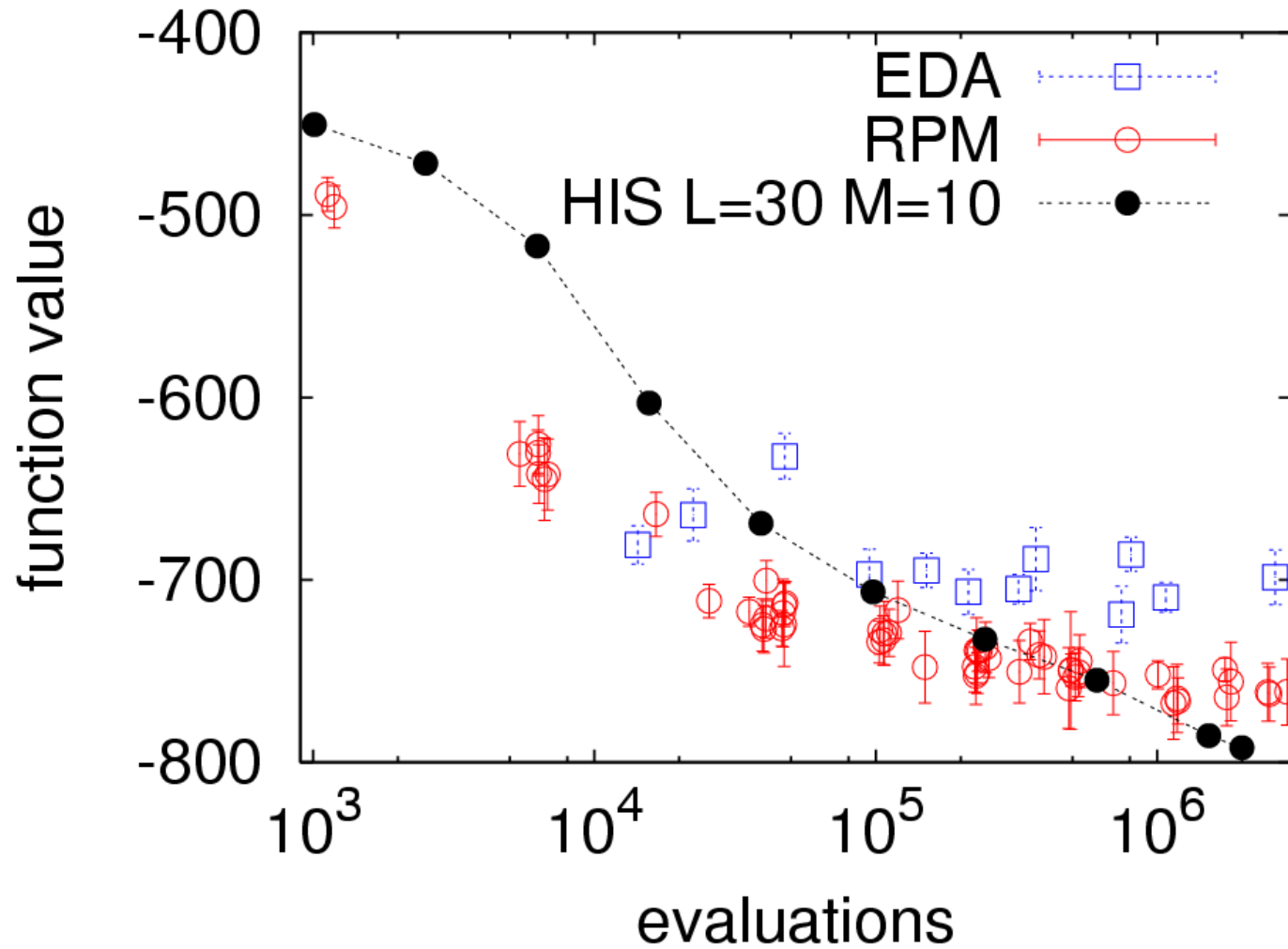




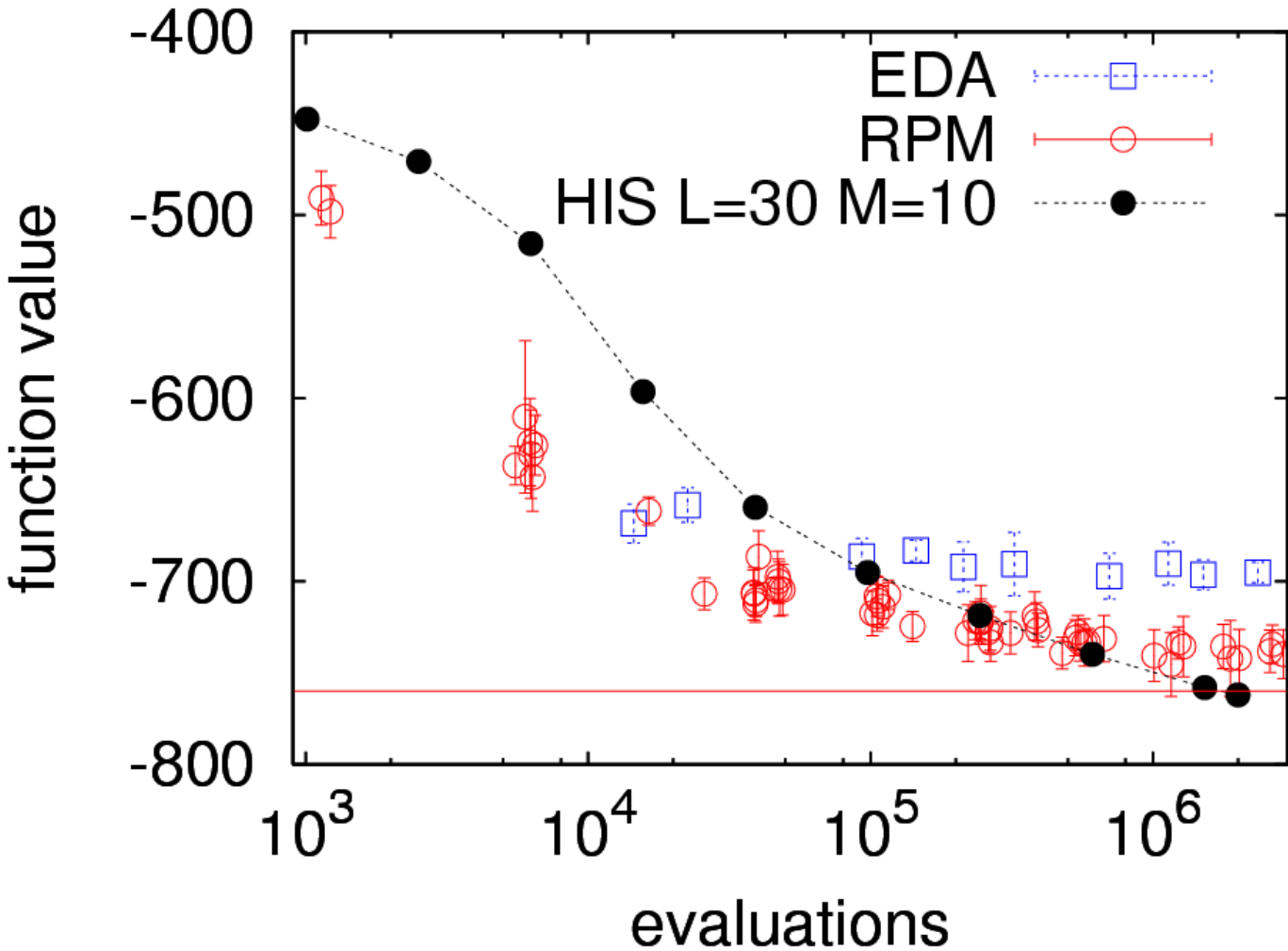
- 実験1: 通常
- 実験2: 5%で制約を反転
- 実験3: 10%で制約を反転



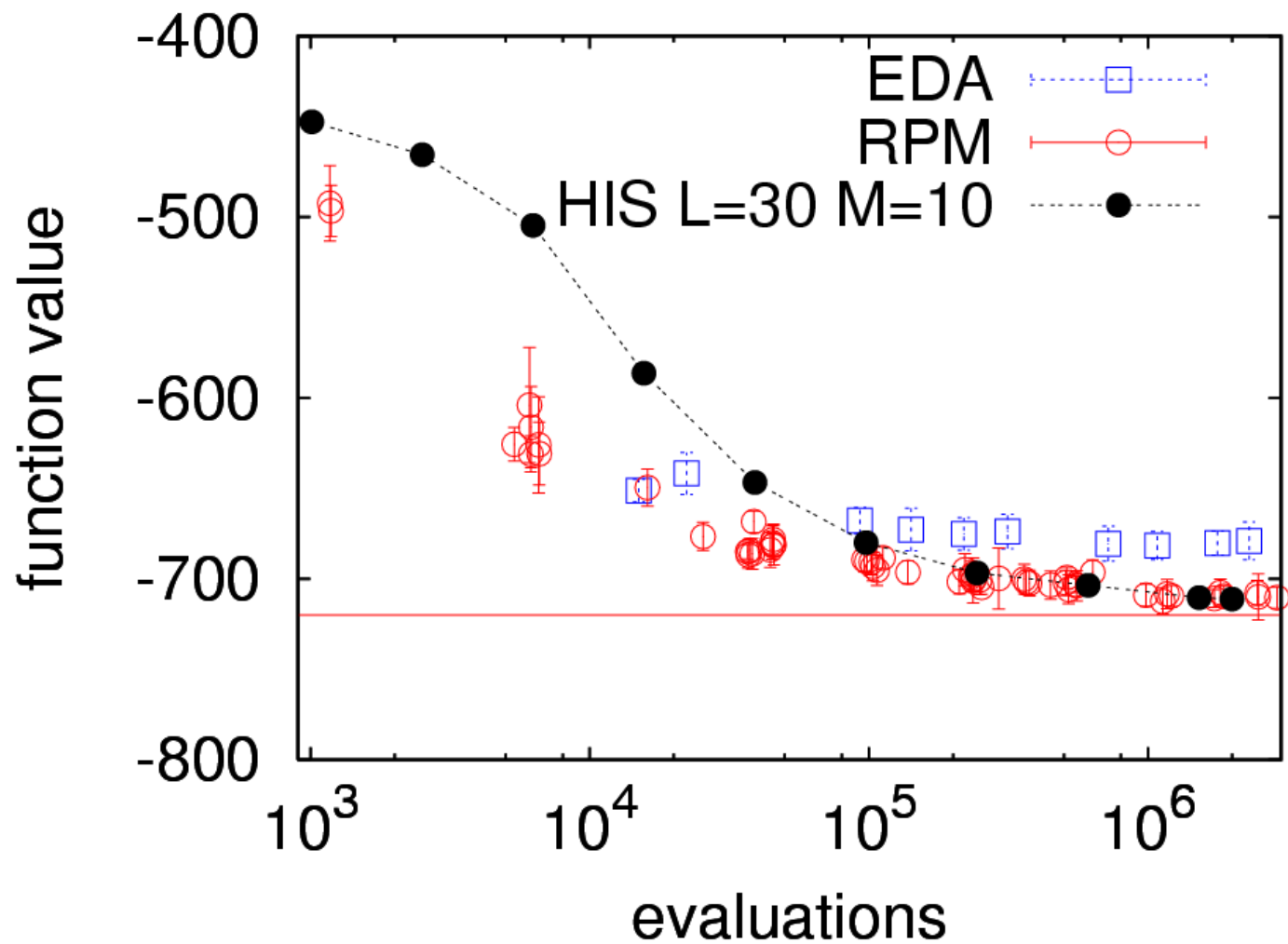




**p(J=-1)=5%**



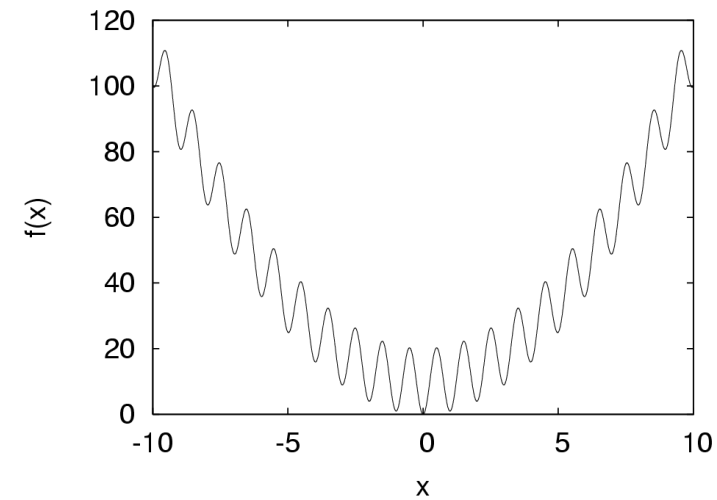
**p(J=-1)=10%**





クラスターは  
反転制約で縮小する  
RPM=HIS





- |             | EDA | RPM | HIS |
|-------------|-----|-----|-----|
| Rosenbrock  | 0   | 10  | 0   |
| Rastrigin   | 10  | 0   | 10  |
| S-Rastrigin | 10  | 0   | 10  |

	EDA	RPM	HIS
Rosenbrock	0	10	10
Rastrigin	10	0	3
S-Rastrigin	0	0	3



- 関数評価回数(単位:1E5)

	EDA	RPM	HIS
Rosenbrock	8	11	x
Rastrigin	8	11	18
S-Rastrigin	9	14	1

=打ち切り時間

=幾何平均で4

	EDA	RPM	HIS
Rosenbrock	7	6	2
Rastrigin	9	9	4
S-Rastrigin	11	9	4

打ち切り時間の影響あり  
(成功ケースは2〜3)

	偏り消去	局所解	分布推定への依存
EDA	×	△	△
RPM	○	×	悪い分布でもOK
HIS	○	○	推定が悪いとダメ



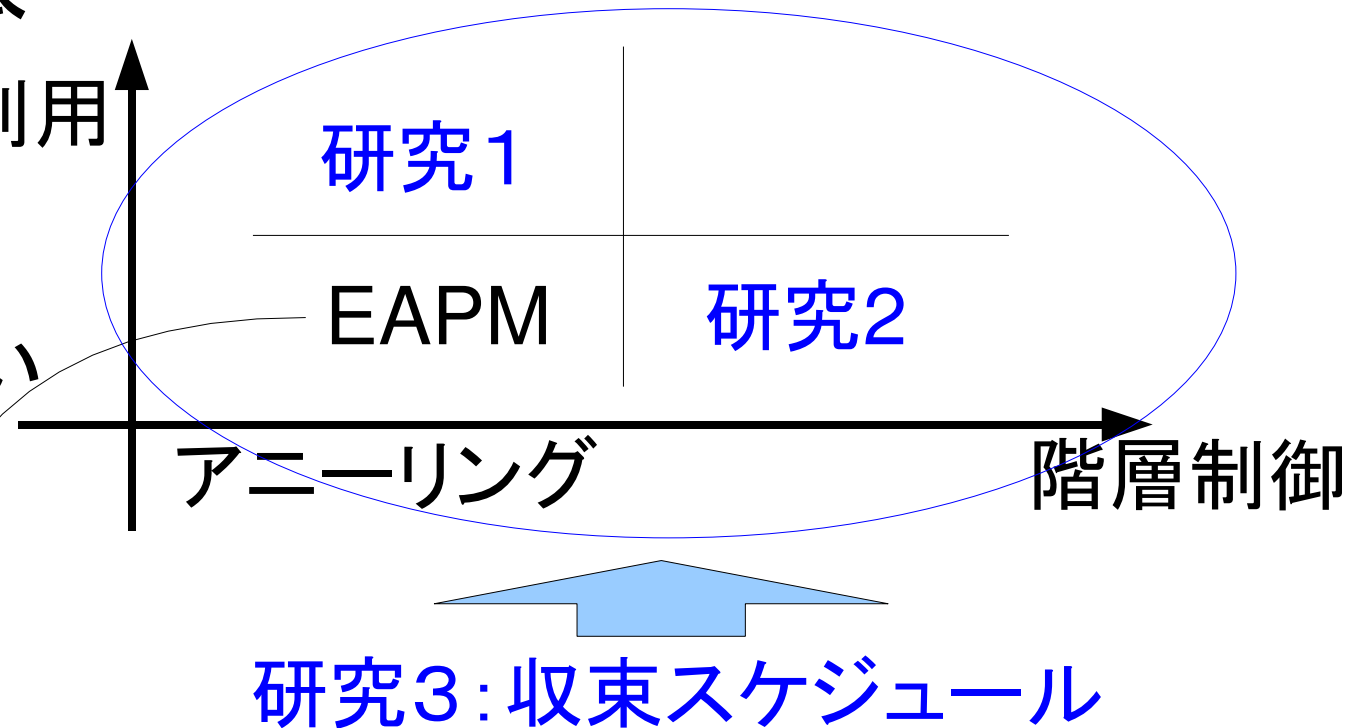
- 2つの拡張

# 過去サンプル利用

利用しない

# EDA

## インポートانسサンプリングで 近似を使っている方法









## 実験の妥当性





- # entropy



- $$\Delta S \simeq -\sigma^2 \int \beta d\beta$$

- ## - ZをISで推定

- 目標分布

$$q(x) = \frac{\tilde{q}(x)}{Z}$$

$$Z = \int \tilde{q}(x) dx \simeq \frac{1}{N} \sum \frac{\tilde{q}(x)}{p(x)}$$

- $$\frac{\partial S}{\partial \beta} = -\sigma^2 \beta$$

- $$\begin{aligned}\Delta S &= \int -\sigma^2 \beta \, \mathrm{d}\beta \\ &\simeq -\sigma^2 \int \beta \, \mathrm{d}\beta = \frac{-\sigma^2}{2} \{\beta_1^2 - \beta_0^2\}\end{aligned}$$

- $$r = -\sigma^2(\beta_1 - \beta_0)^2$$
- $$\beta_1 = \frac{\sqrt{-r}}{\sigma} + \beta_0$$

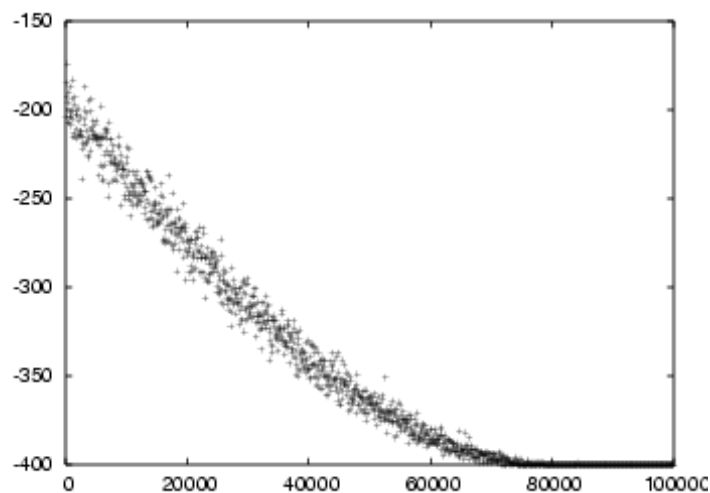
## 一 提案手法

$$r = -\sigma^2(\beta_1^2 - \beta_0^2)$$

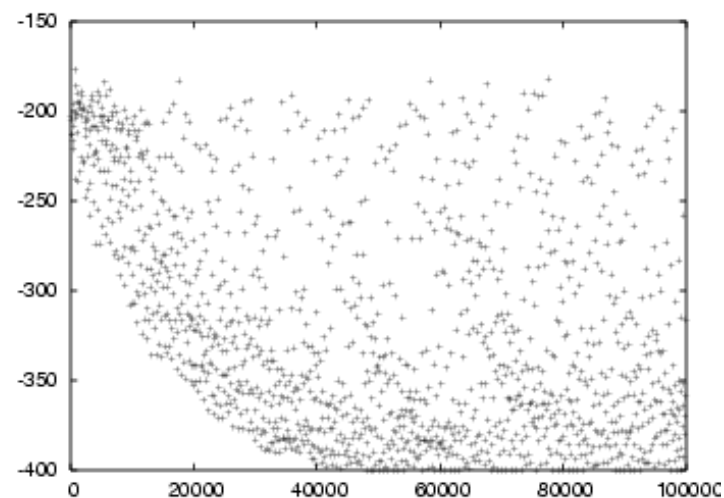
$$\beta_1 = \sqrt{\beta_0^2 - \frac{r}{\sigma^2}}$$

# 提案手法によるサンプリングの様子

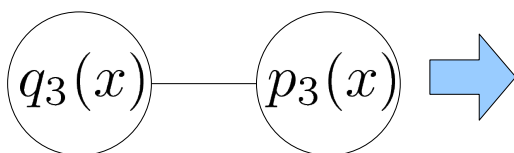
- 常に乱雑度をキープ



## 従来法 (EDA)



- 
- A diagram showing two nodes,  $q_4(x)$  and  $p_4(x)$ , connected by a horizontal line. An orange arrow points from the right towards the  $p_4(x)$  node.

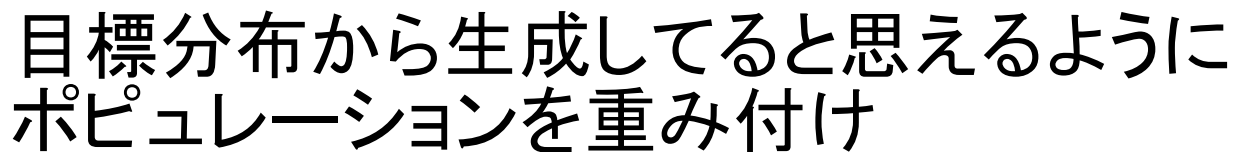


## 对数尤度

$$L = \frac{1}{N} \sum_{p_3(x) \cup p'_3(x)} \frac{q_4(x)}{\alpha p_3(x) + (1 - \alpha) p'_3(x)} \log \hat{p}(x|\theta)$$



## 2. リサンプリング



- $$X_{pop} = \{(x_i, w_i)\}_{i=1}^N$$

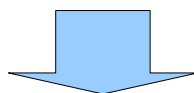
A graph of a discrete function  $\hat{q}(x)$  versus  $x$ . The x-axis is horizontal and the y-axis is vertical. The function is represented by six vertical lines of increasing height, then decreasing, starting from a positive x-value. The y-axis is labeled  $\hat{q}(x)$  and the x-axis is labeled  $x$ .

- $$\hat{q}(x) = \frac{1}{\sum w_i} \sum_X w_i \delta(x - x_i)$$

$$\int \underbrace{\hat{q}(x)}_{\text{proposal}} f(x) dx = \frac{1}{\sum \frac{q(x)}{p(x)}} \sum \frac{q(x)}{p(x)} f(x) \simeq \int \underbrace{q(x)}_{\text{proposal}} f(x) dx$$

# サンプル

$$X_{samp} = \{(x_i, 1)\}_{i=1}^M$$



$$w^{(t+1)} = w^{(t)} \frac{q_{t+1}(x)}{p_m^{(t)}(x)}$$

$$= w^{(t)} \frac{q_{t+1}(x)}{\alpha p_t(x) + (1 - \alpha)q_t(x)}$$

# 更新式

$$w_{t+1} = w_t \frac{q_{t+1}(x)}{q_t(x)}$$

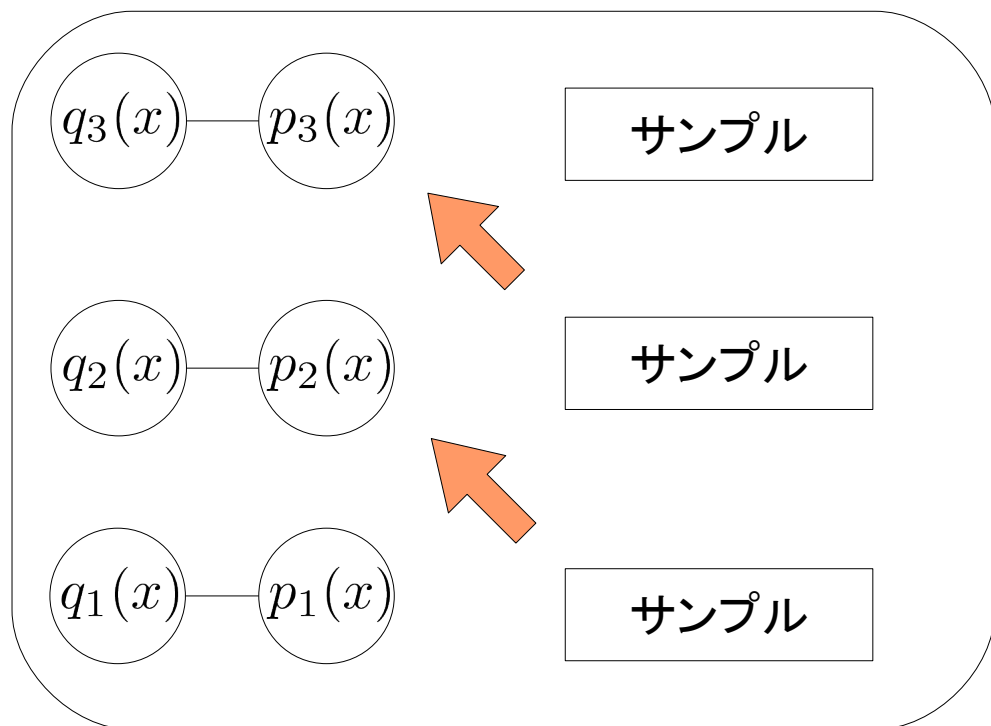
- $$\hat{q}(x) = \frac{1}{\sum w_i} \sum_{x_i \in X} w_i \delta(x - x_i)$$



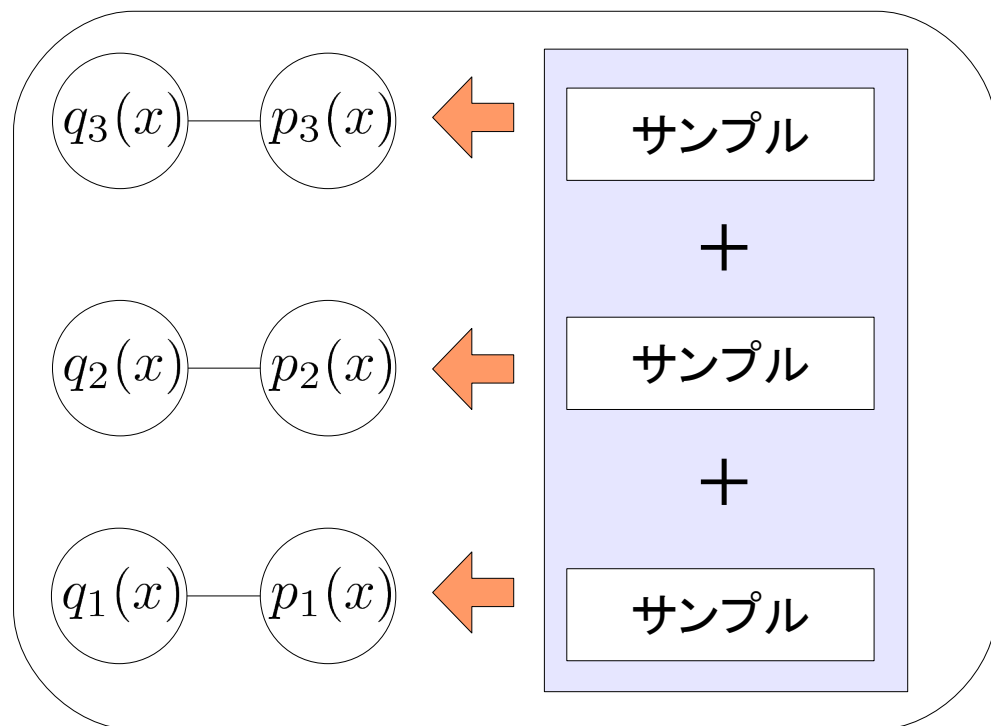
- 本研究では重みを保存した非復元抽出
- 重複したサンプルを保存するのは無意味

# 考察：ポピュレーション収束の一般化

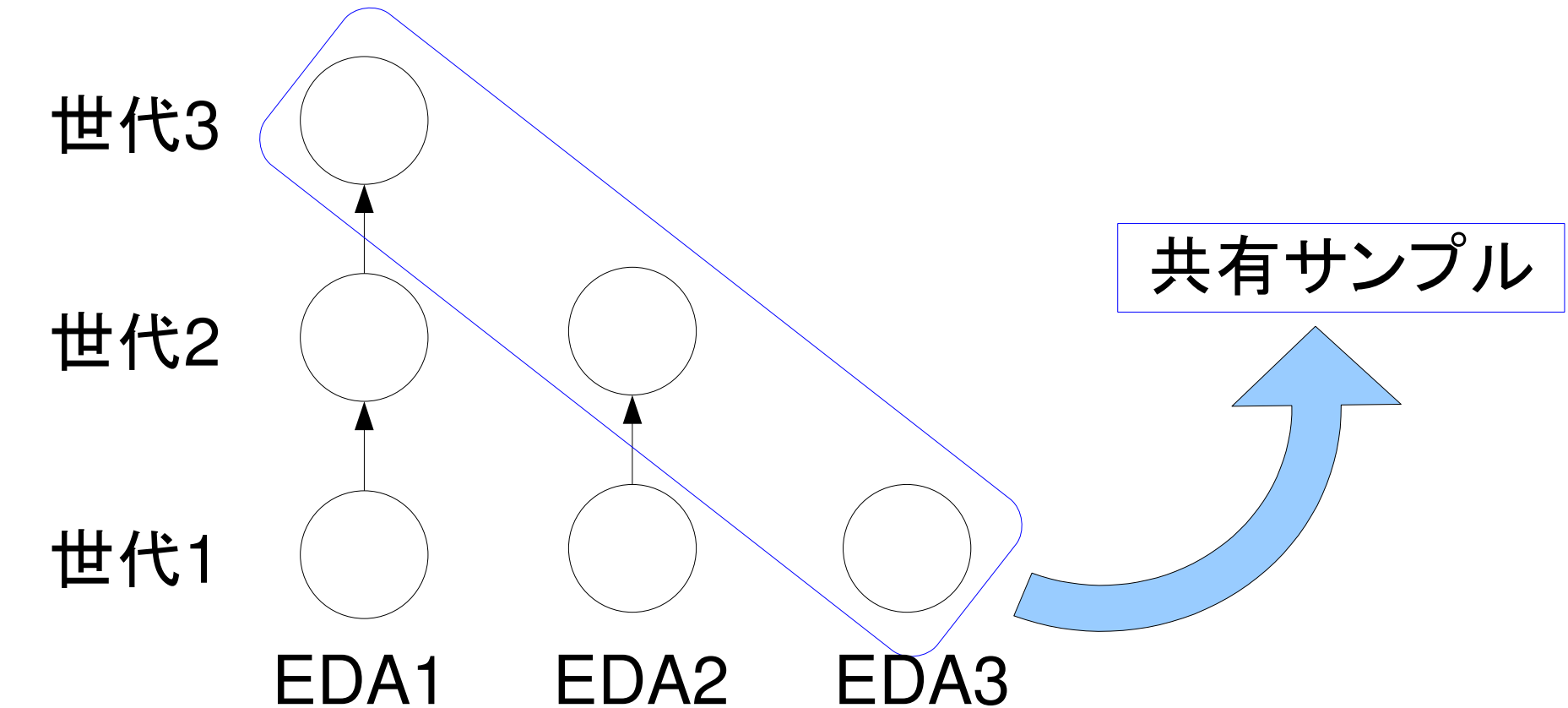
- HIS = 「繰り返しEDA」の一般化
  - $p(x)$  の推定が繰り返しEDAに比べて必ず良い



# 繰り返しEDA



# HIS





# Comparison with Iterative EDA

- EDA (M=3000, C=0.5)
  - The 10 best results in 100 trials: -746, -736, -732, -732, -730, -730, -730, -728, -726, -726
  - Average: -719 and Standard deviation: 15.68
  - HIS is -800 > -746
- Function evaluations
  - HIS is 2e6 vs. EDA is 7e7



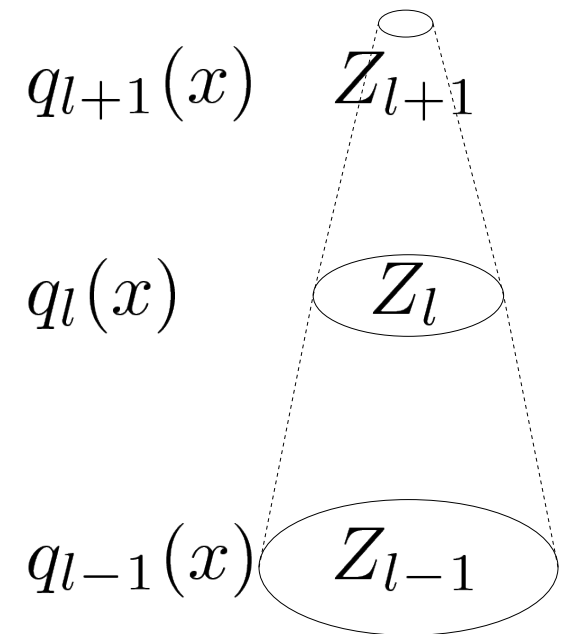
# 目標分布スケジュール(1)

Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria A

- 部分的一様分布
  - 制御パラメータ: 閾値  $\tilde{f}$
  - 正規化定数 → 生成候補解の数

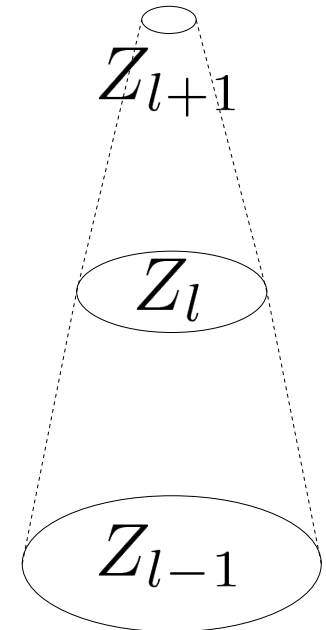
$$q(x) = \frac{\tilde{q}(x|\tilde{f})}{Z}$$

$$\begin{aligned}\tilde{q}(x|\tilde{f}) &= I(f(x) < \tilde{f}) \\ &= \begin{cases} 1 & f(x) < \tilde{f} \\ 0 & \text{else} \end{cases}\end{aligned}$$



- $$\int q_l(x)r(x)dx = \frac{1}{M_{l-1}} \sum_{q_{l-1}(x)} \frac{q_l(x)}{q_{l-1}(x)} r(x)$$

- 無駄サンプル生成確率:  $\frac{Z_l}{Z_{l-1}}$



# 目標分布スケジュール(3)

Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria A

- 無駄サンプルの和の期待値

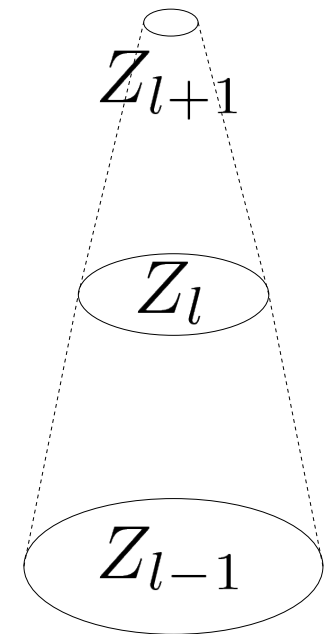
$$\sum_l M_{l-1} \frac{Z_l}{Z_{l-1}}$$

- 極値条件

$$M_{l-1} \frac{Z_l}{Z_{l-1}} = M_l \frac{Z_{l+1}}{Z_l}$$

- 探索空間を指数的に縮小

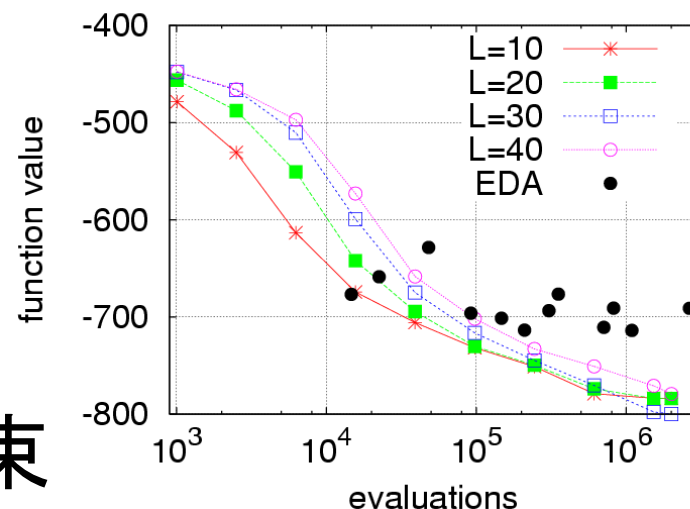
$$1 : 10 = 10 : 100$$



$$\begin{aligned} Z_l &= \int \tilde{q}_l(x) dx \\ &\simeq \frac{1}{M_m} \sum \frac{\tilde{q}_l(x)}{p_m(x)} \end{aligned}$$

$$\begin{aligned}\tilde{q}(x|\tilde{f}) &= I(f(x) < \tilde{f}) \\ &= \begin{cases} 1 & f(x) < \tilde{f} \\ 0 & else \end{cases}\end{aligned}$$

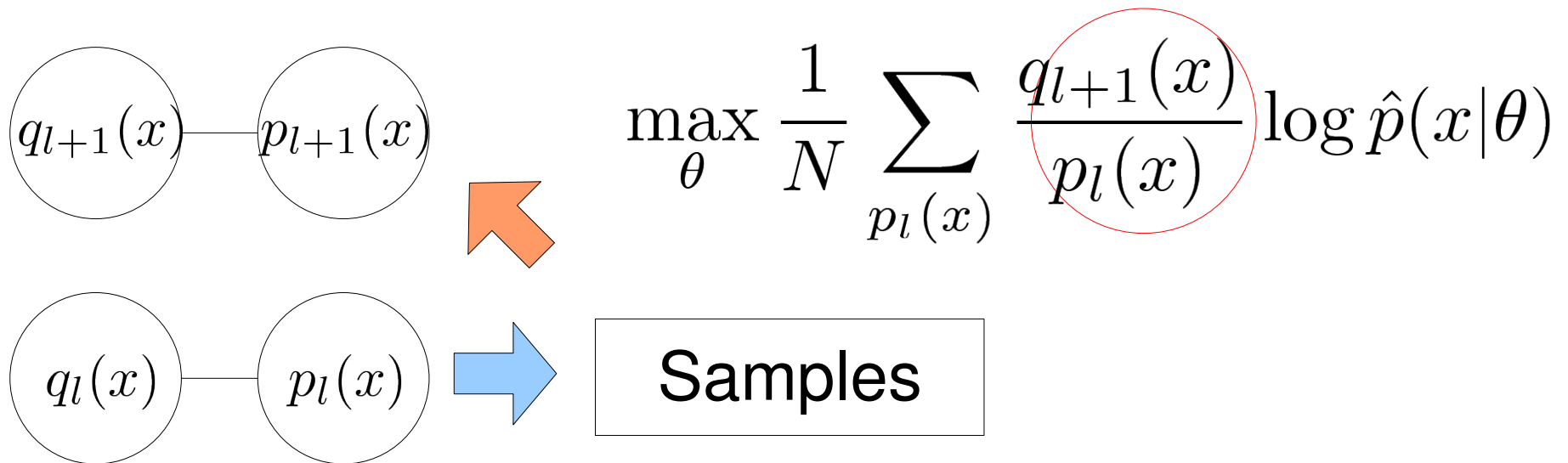
- 
- 悪い解
- 良い解
- 速い収束
- 遅い収束
- function value
- パラメータA
- パラメータB
- パラメータC





# インポートランスサンプリング (IS) の効率

- 悪いISは悪い確率モデルを生成する
- 悪い確率モデルは悪いISの原因になる





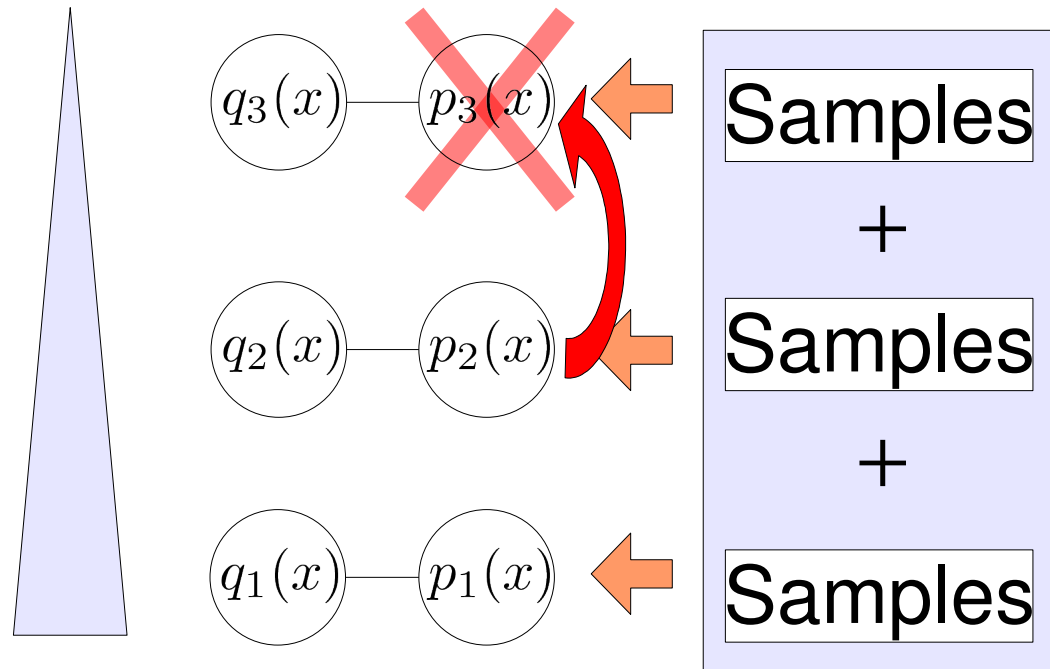


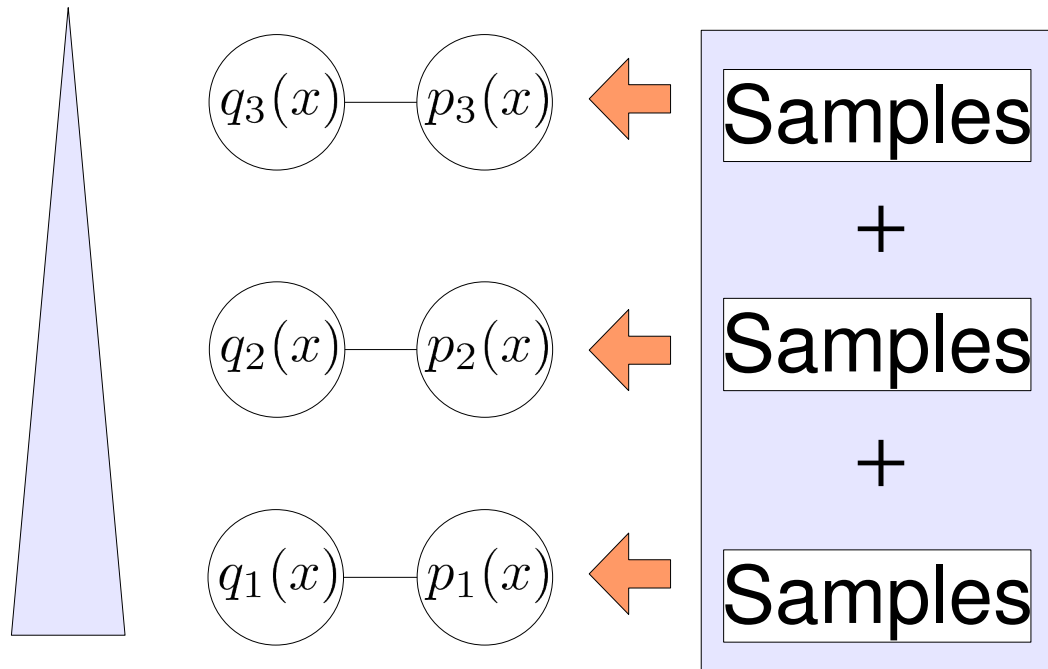


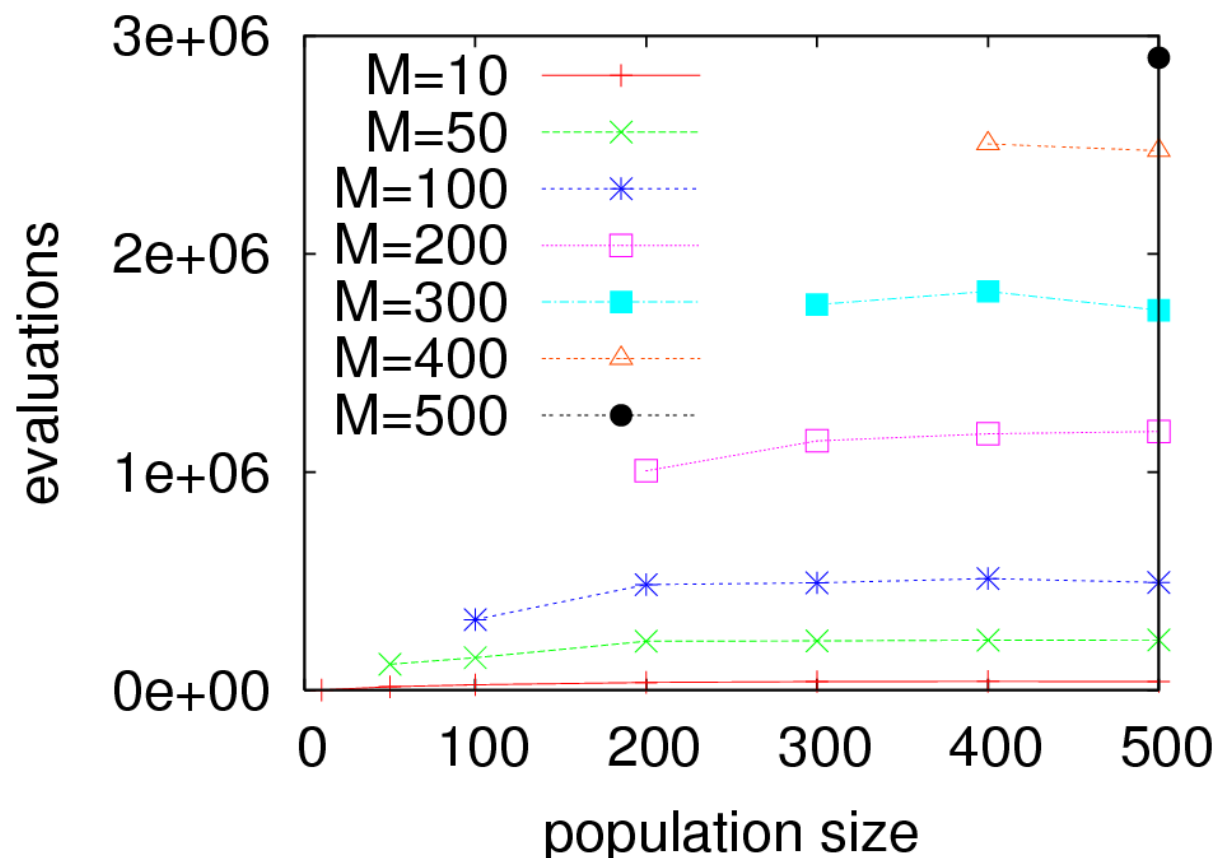




-







- $$q(x) = \frac{\tilde{q}(x|\tilde{f})}{Z} \quad (Z \text{は正規化定数})$$
- $$\begin{aligned} \tilde{q}(x|\tilde{f}) &= I(f(x) < \tilde{f}) \\ &= \begin{cases} 1 & f(x) < \tilde{f} \\ 0 & else \end{cases} \end{aligned}$$

$$q(x) = \frac{\tilde{q}(x|\tilde{f})}{Z} \quad (Z \text{は正規化定数})$$

$$\begin{aligned}\tilde{q}(x|\tilde{f}) &= I(f(x) < \tilde{f}) \\ &= \begin{cases} 1 & f(x) < \tilde{f} \\ 0 & else \end{cases}\end{aligned}$$

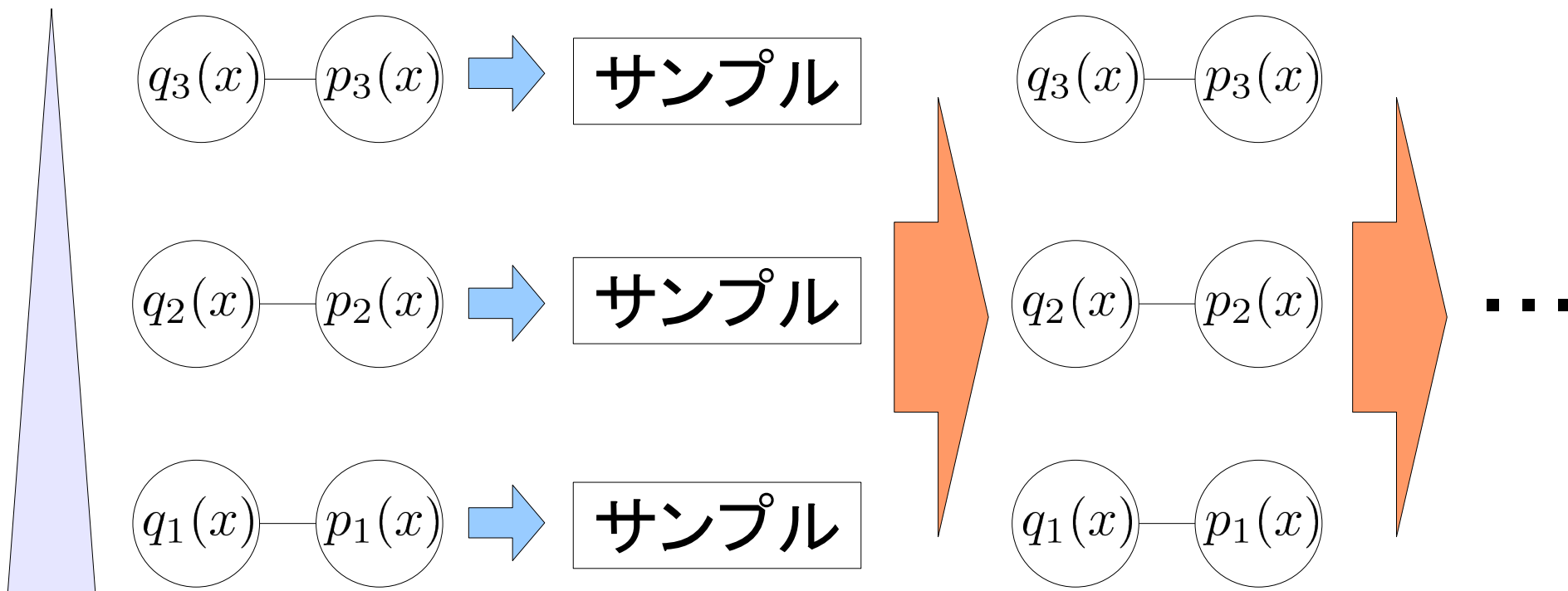




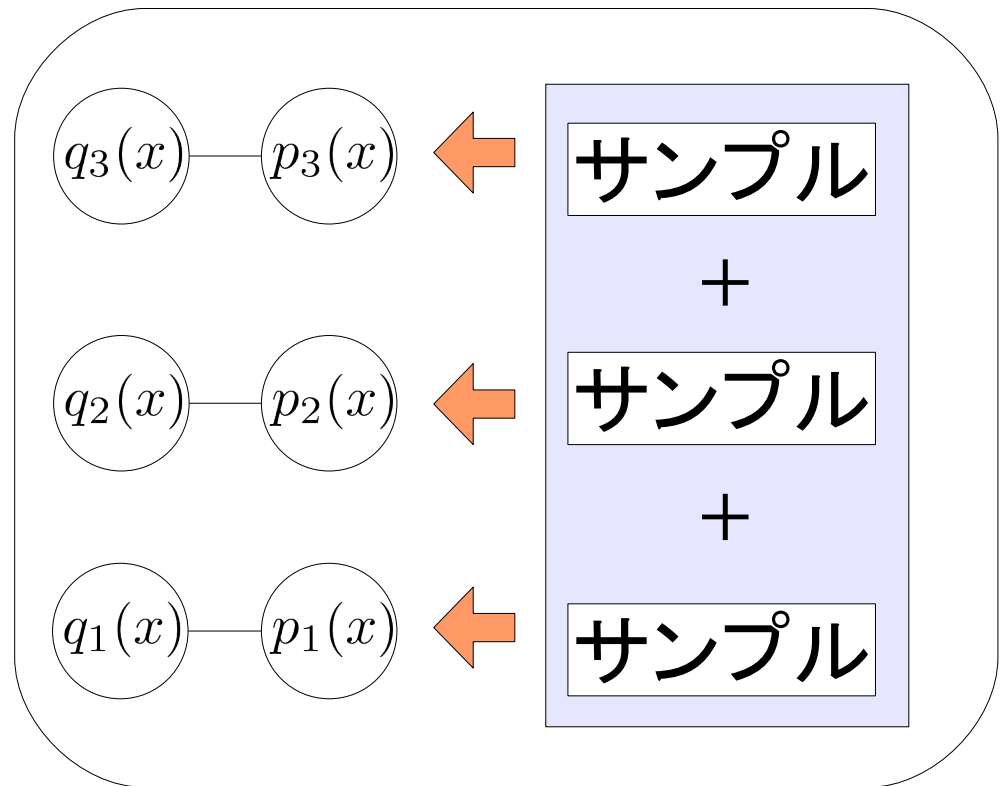
# 提案: 階層型インポートانسサンプリング Hierarchical Importance Sampling (HIS)

Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria Algorithm Maria A

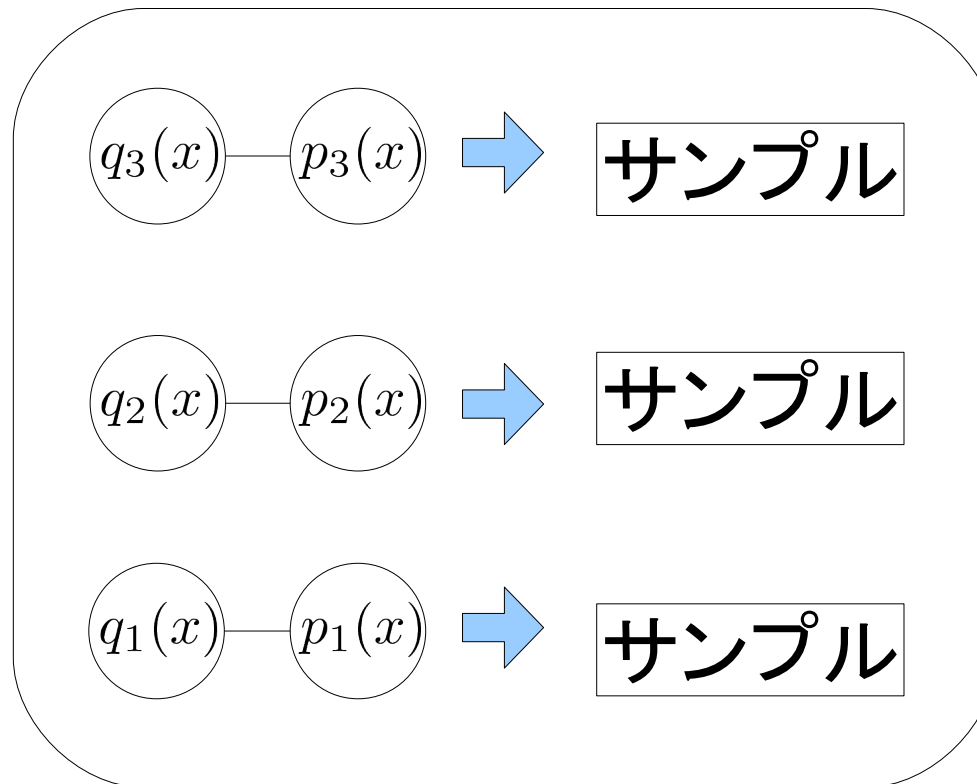
- 全ての確率モデルを同時に繰り返し更新
  - エントロピーの異なるサンプルを混ぜる



- ## (1) サンプリング



- 1度にサンプリング

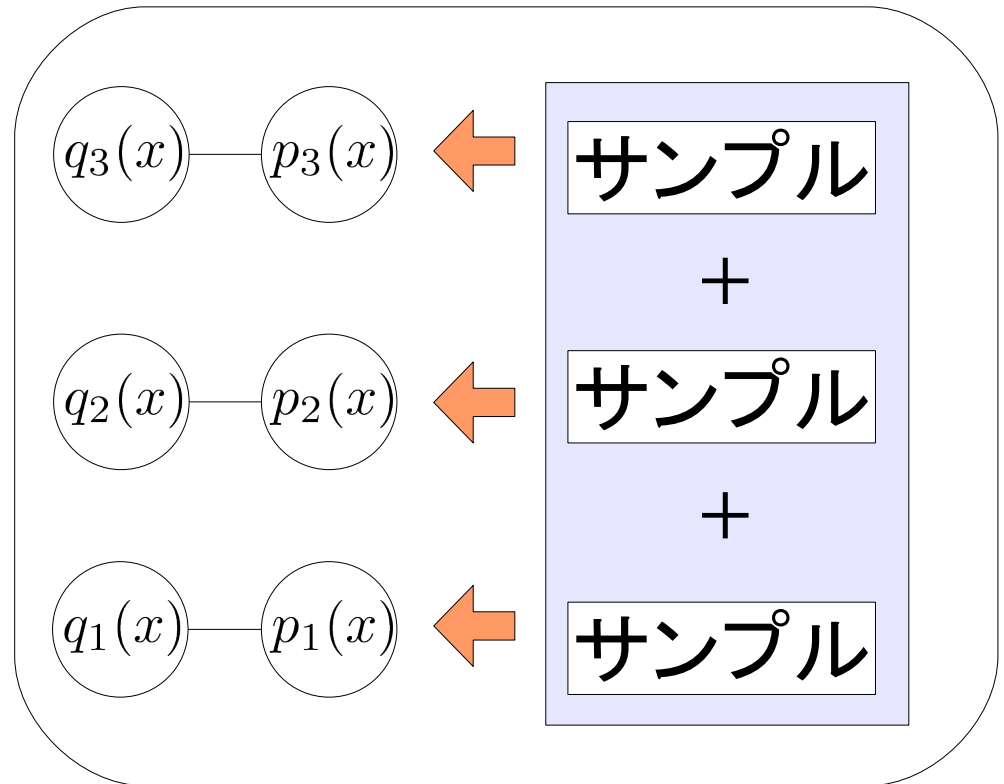


## (2)推定

- 確率分布推定に全てのサンプルを使う
  - サンプル生成分布は混合分布

混合分布

$$\frac{1}{3}p_1(x) + \frac{1}{3}p_2(x) + \frac{1}{3}p_3(x)$$



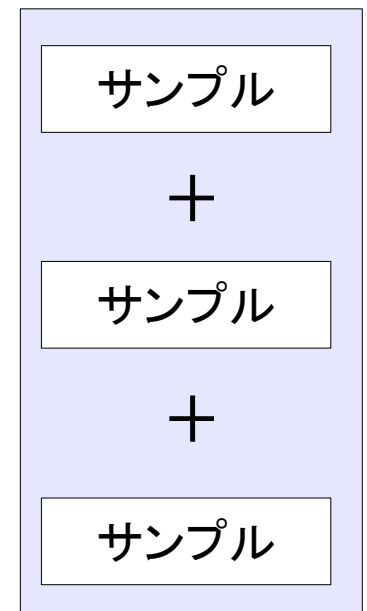
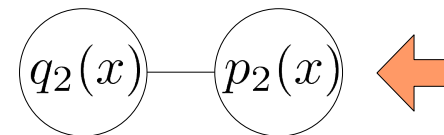
# 混合分布によるインポートانسサンプリング

- サンプルが従う分布:

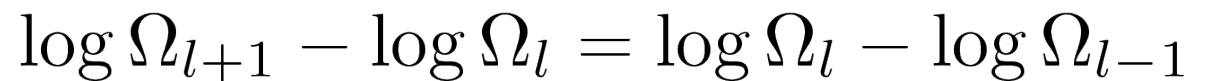
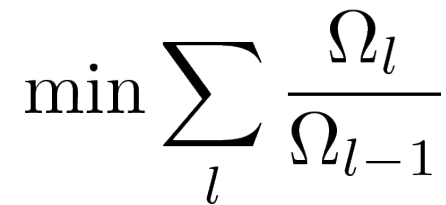
$$p_m(x) = \frac{1}{3}p_1(x) + \frac{1}{3}p_2(x) + \frac{1}{3}p_3(x)$$

- 経験対数尤度

$$\max_{\theta} L_i(\theta) = \frac{1}{N} \sum \frac{q_i(x)}{p_m(x)} \log \hat{p}(x|\theta)$$



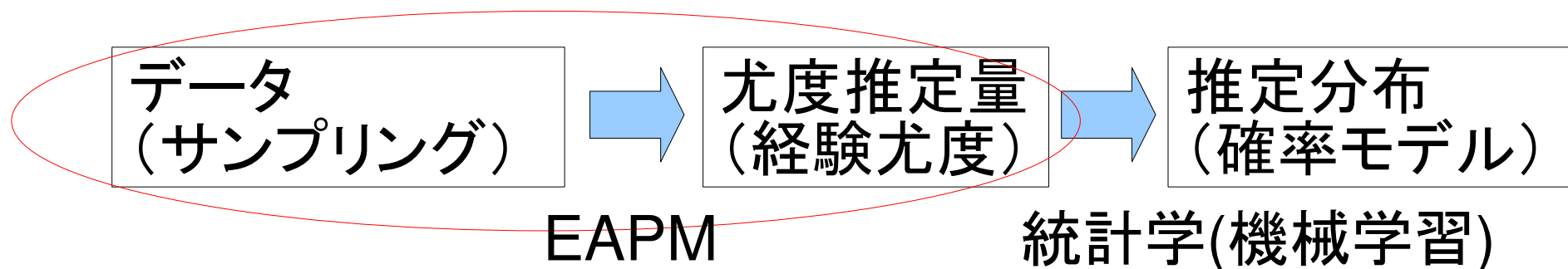
- $$\text{Var} \left( \frac{q_{t+1}(x)}{q_t(x)} \log \hat{p}(x) \right)_{q_t(x)} = \frac{\Omega_l}{\Omega_{l+1}} \text{Var} (\log \hat{p}(x))_{q_t(x)}$$



## 探索空間を指数的に減少

# 研究の目的: 尤度推定の改良

- 統計的推定の3つの要素



- $$\int q(x) \log \hat{p}(x) dx \simeq \sum \frac{q(x)}{p(x)} \log \hat{p}(x)$$
- 目標: モンテカルロ積分の改良
  - サンプル数を増やす
  - 分散を減らす →  $q(x)$  に近い  $p(x)$  を使う



- 
- The diagram illustrates the process of merging two overlapping ellipses into a single ellipse. It is organized into two columns, with a large blue arrow pointing from the left column to the right column. Each column contains three rows of ellipses. In the left column, the top row shows two small, separate ellipses (one blue, one red). The middle row shows two overlapping ellipses of moderate size. The bottom row shows two large, overlapping ellipses. In the right column, the top row shows two overlapping ellipses of moderate size. The middle row shows two overlapping ellipses of moderate size, but the overlapping region is shaded in a darker purple color. The bottom row shows two large, overlapping ellipses, with the overlapping region shaded in a darker purple color. This visualizes the concept of merging two overlapping regions into a single, unified region.