Young Scientists Meeting on Statistical Physics and Information Processing 2011

機械学習に隠されたJarzynski等式

大関　真之
京都大学大学院情報学研究科　システム科学専攻

非平衡統計力学というと，何から手をつけてよいのかが分からない．筆者もその印象の持ち主であった．何がゴールなのかを明確にする事を忘れさせるくらいに非常に多岐に渡ったトピックが存在する．その中でも極めて異彩を放って，何かがありそうと魅了してくれるのが Jarzynski 等式である [1, 2]．非平衡状態での微視的な状態変化に関する平均と，平衡状態の物理量が関係付けられる厳密な等式である．荒く言えば，この等式は熱力学第二法則の拡張と言われる．熱力学第二法則は，外部からの操作としての仕事と操作前後の自由エネルギーの変化との関係を与える．最近ではこの操作の選択が，微視的状態の情報を利用したフィードバック制御に基づくものとした拡張も議論されるようになってきた [3]．他にも効率的なサンプリング手法としての発展や [4]，スピングラスにおける対称性との組み合わせによる理論的考察も重ねられてきている [5, 6]．本講演では，この Jarzynski 等式と意外な分野との接点について紹介する．

今日ではデジタルカメラを利用する場面がよく散見される．顔の表情を捉えて，シャッターチャンスを逃さないなどといった高度な情報処理に基づいた撮影補助システム等が標準的に搭載されている．この場合，内部では例えば笑顔であるかそうでないかを判別するような情報処理を行っている．他にも撮影時に，動いている被写体が標的物体であるかどうかを瞬時に判別して，それを追うような枠を動画内に表示させる技術等も開発されている．やや前時代的な例では，電子メールの文章内容からスパムメールであるかどうかを判別するなどの処理も同種の技術に分類される．このように，与えられたデータのなかから，ある基準に収まるかどうかを判別する問題を判別問題と呼び，機械学習の分野でのテーマのひとつとなっている [7]．事前にいくつかの事例を与えて，学習を行う事によって，今後来る未知の情報に対しても適切に判別が行えるような優れた判別器を構成するのが目標となる．優れた判別器を構成する学習アルゴリズムのひとつに Adaptive Boosting(AdaBoost) と呼ばれる手法がある [8]．この AdaBoost は，事例集合の判別を通して，優れた判別器を高速に構成できる事が知られている．実はこの AdaBoost の内部処理方法が，Jarzynski 等式の構成方法と全く同じである事が分かった．興味深い事に，Jarzynski 等式の成立とほぼ同時，正確には少し前に AdaBoost が提案されている．

AdaBoost の概要を簡単に説明しよう．多次元空間上で2値を取る事例集合 $\mathbf{x}_k \in \{-1,1\}^N$ に対して，判別の答えに相当する $y_k \in \{-1,1\}$ が割り当てられている．ここで k は N_D 個の事例集合の添え字とする．また各事例には判別の難しさに相当する重み $D_k(t)$ を設定する．初期段階では，この重みは一様なものとする．我々の目的は，$f(\mathrm{x}_k) = y_k$ となるような判別関数を構成することである．AdaBoost ではいくつかの判別器 $h_j(\mathbf{x}_k) \in \{-1,1\}$ の以下のような線形和によって優れた判別器を構成する．

$$f_t(\mathrm{x}_k) = \sum_{j=0}^{t} \alpha_j h_j(\mathrm{x}_k). \tag{1}$$

ここで α_j は判別器 $h_j(\mathrm{x}_k)$ に対する信頼度の指標を表す．この信頼度と選択する判別器 $h_j(\mathbf{x}_k)$ は段階的に決定していく．そのため時刻に相当する t という添え字が上記の線形和による判別器にはつく．各段階で判別器は，標準的には，以下のように最小の誤り確率を与えるものが選ばれる．

$$h_j = \arg\min \epsilon(h_j; \{D_k(j)\}) \tag{2}$$

ここで ϵ は誤り確率を表し，

$$\epsilon(h_j; \{D_k(j)\}) = \sum_{k=1}^{N_D} I[h_j(\mathbf{x}_k) \neq y_k] \frac{D_k(j)}{C(j)} \tag{3}$$

と定義する．分母に現れるのは規格化定数 $C(j) = \sum_k D_k(j)$ であり，$I[\cdots]$ はインジケータ関数である．この誤り確率を最小化する判別器に対しての信頼度を以下のように与える．

$$\mathrm{e}^{-2\alpha_j} = \frac{1-\epsilon(h_j; \{D_k(j)\})}{\epsilon(h_j; \{D_k(j)\})}. \tag{4}$$

この信頼度と選択した判別器を用いた線形和によって最終的に得られる関数は，非常に良好な判別性能を持つことが知られている．そのため AdaBoost は，汎用的な性質もあって，広く有効な判別アルゴリズムとして利用されている．

上記の AdaBoost における過程を物理のプロセスであると読み解く事によって，AdaBoost の数理構造が Jarzynski 等式と非常に似ている事がわかる．さらに Jarzynski 等式の成立には直接関係はないものの，効率的に判別問題を解くために適応的に判別器を選択するプロセスが AdaBoost では含まれている．この対応関係はその意味で，新しいタイプの Jarzynski 等式の発見であるともいえる．この進展は例えば生物の進化などの適応的変化における熱力学第二法則の拡張へと繋がる可能性を秘めている [9]．極めて萌芽的で挑戦的な内容で，まとまりがない部分もあるが，議論の種になる事を期待する．

参考文献

[1] C. Jarzynski, Phys. Rev. Lett. **78** (1997) 2691.

[2] C. Jarzynski, Phys. Rev. E **56** (1997) 5018.

[3] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104** (2010) 090602.

[4] K. Hukushima and Y. Iba, AIP. Conf. Proc. **690** (2003) 200.

[5] M. Ohzeki and H. Nishimori, J. Phys. Soc. Jpn. **79** (2010) 084003.

[6] M. Ohzeki, H. Katsuda and H. Nishimori, to appear soon.

[7] C. M. Bishop *Pattern Recognition And Machine Learning* (2006) (Berlin: Springer-Verlag).

[8] Y. Freund, and R. Schapire, J. Comp. and Sys. Sci. **55** (1997) 119.

[9] M. Ohzeki and A. Sakata, to appear soon.