ROC カーブとスケール不変性
ROC curve and scale invariance

守 真太郎 [1]、久門 正人 [2]
Shintaro Mori and Masato Hisakado

北里大学 理学部 物理学科 [1]
神奈川県 相模原市 北里1ー15ー1：mori@sci.kitasato-u.ac.jp
スタンダード & プアーズ [2]
東京都 千代田区 丸の内1ー6ー5：masato_hisakado@standardandpoors
Department of Physics, School of Science, Kitasato University[1]
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan.
Standard & Poor's[2]
1-6-5 Marunouchi, Chiyoda-ku, Tokyo 100-0005, Japan.

概 要

We discuss the relation between the mixing property of binary objects and the ROC curve. We propose a voting model to binary candidates and show that the mixing of them in the axis of the number of votes becomes scale invariant. The scale invariance holds over the entire range in a limit, which is a remarkable feature. We also find that the mixing of winning horses and losing horses is scale invariant.

Keywords
voting, scale invariance, ROC curve, stochastic process, accuracy ratio

1 情報の非対称性と多数決

二頭の馬のレースを考え人々が自分が勝つと思う馬に対して投票を行うとする。一方の馬を〇、もう一方の馬を×とし、レースを行うと馬〇が勝つとする。つまり、馬〇は真に強い馬であり、馬×は真に弱い馬である。しかし、「真に強い」、とか、「真に弱い」という情報はレース後になって判明することであり、投票時点ではどちらの馬がレースに勝つのか分らない。こうした場合、人々は投票時点で得られる情報をもとに判断し投票することになる。その情報が正しいものならば、人々は馬〇に多く投票し、馬〇は得票数で馬×に勝つことになる。多数決とレース結果が一致するわけである。しかし、投票時点で得られる情報は必ずしも正しくはない。そこには、馬の強さに対する正しい情報に加えて嘘の情報や馬に対する思い入れも入ってくるだろうし、また、他の人々の投票行動にも左右されるであろう。そうした場合、必ずしも馬〇が馬×よりも多数の票を集めるとは限らない。神のみぞしる情報〇、×と人々が知る情報には非対称性があり、多数決での勝者とレースでの勝者は異なることが起きうる。

では、このような投票を多数回繰り返したとき、一体どのような現象が見えてくるのだろうか?我々は、簡単な投票モデルを考え、投票行動が他の投票者の行動に左右されるコピーキャット的な場合は、馬〇と馬×の混合がスケール不変性を示すことを示した。さらに、コピーキャットの度合を最大限に強くすると、スケール不変性が得票数軸の全領域で成立すること、また、実際の競馬のデータでもこうしたスケール不変性が見られることを発見した [1]。

2 投票モデル

N 頭の馬のうち N_s 頭が強い馬、N_w 頭が弱い馬とし、時刻 $t = 0, 1, \cdots$ での得票数を $X_{i,t}^{\mu}$ で記述する。$\mu = s$ が優秀、$\mu = w$ が劣等を表し、添え字 i は $i = 1, 2, \cdots, N_\mu$ の値をとる。初期条件は $X_{i,0}^{\mu} = 0$ である。投票者は、馬の強さについて部分的な情報を持ち、また、他の投票者の行動にも影響されると考えて、投票数と馬の持つスコアの和に比例した確率で投票するとする。強い馬のスコアは s、弱い馬のスコアは w とし、馬の各時刻 t での投票確率 $p_{i,t}^{\mu}$ は、

$$p_{i,t}^{\mu} \propto \mu + X_{i,t}^{\mu}$$

とする。このモデルでは、投票の初期段階では各馬の持つスコアに比例して投票を行うが、投票が進むにつれて得票数の得票確率への寄与が大きくなり、その結果、投票行動がコピーキャット的になる。

この確率モデルは分岐過程にマップすることで厳密に解くことができ、得票数の分布関数はガンマ分布になる。強い馬と弱い馬の混合を表示する手段として ROC（Receiver Operating Charactersitic）カーブを用いる。これは、二種類の馬の得票数軸での累積分布を二次元空間のカーブで表したものである。$N_s, N_w \to \infty$ の極限では、ROC カーブの媒介変数表示 $(x(t), y(t))$ は、

$$x(t) = 1 - \frac{\gamma(w,t)}{\Gamma(w)}, y(t) = 1 - \frac{\gamma(s,t)}{\Gamma(s)}$$

となる。ここで、$\gamma(\mu,t)$、$\Gamma(\mu)$ は不完全ガンマ関数、ガンマ関数である。$(x, y) = (1, 1)$ の近傍では、得票数が低い二種類の馬の累積 $1 - x$ と $1 - y$ は次の関係式を満たすことが分かる。

$$1 - y \propto (1 - x)^{\alpha} \quad \text{with} \quad \alpha \equiv \frac{s}{w} \tag{1}$$

さらに興味深いのは、$\alpha = s/w$ を固定したまま $s, w \to 0$ の極限をとった場合である。このとき、累積 $1 - x$ と $1 - y$ は $0 \le x, y \le 1$ の全領域で巾乗則に従う [1]。

$$1 - y = (1 - x)^{\alpha} \quad \text{with} \quad \alpha \equiv \frac{s}{w} \tag{2}$$

3 競馬のデータとスケール不変性

投票における勝ち馬、負け馬の混合のスケール不変性を確認するために、１９８６年から２００６年までの約９０万頭の馬のデータを用いて ROC カーブを描き、両対数プロットしたものが次の図 1 である。$1 - y$ 軸に着目すれば、二桁（勝ち馬 VS 負け馬）から４桁（１着 VS ２着）の範囲でスケール不変となっていることが分かる。

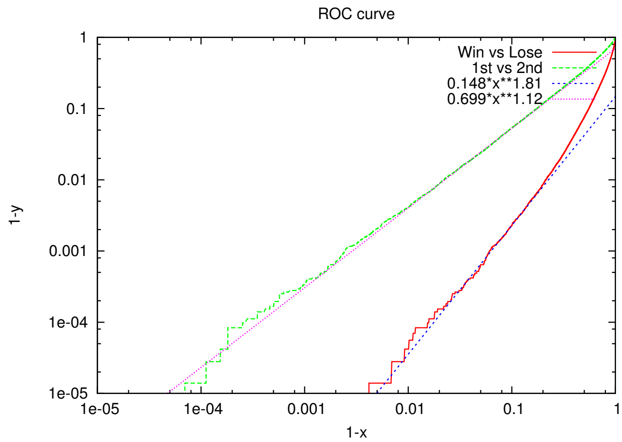

図 1: 勝ち馬と負け馬のペア（赤）、１着馬と２着馬のペア（緑）に対して ROC カーブを両対数プロットした。

スケール不変性が多数決というシステムで持つ意義 [2]、また、投票モデルの ROC カーブの $(s, w) \to 0$ の極限での形 $1 - y = (1 - x)^{\alpha}$ がヤング図形のランダム成長の厳密解を与えることついても発表する。

参考文献

[1] S.Mori and M.Hisakado, *Exact Scale Invariance in the Mixing of Binary Candidates in a Voting System* , preprint arXiv-physics:0806.0185.

[2] 守、久門, 多数決とスケール不変性,「ネットワークが創発する知能研究会」(JWEIN2008) 予稿集,pp.106-pp.113.