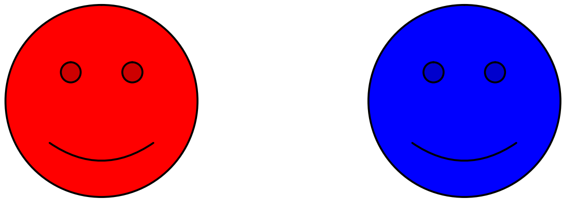
JWEIN08

ネットワークが創発する知能研究会

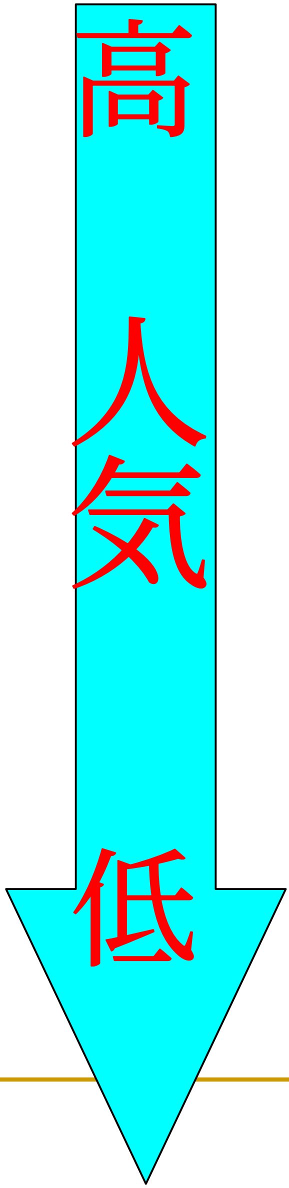
多数決とスケール不変性



守　真太郎@北里大学理学部

久門　正人@スタンダード＆プアーズ

２００６年　有馬記念



人気	馬の名前	オッズ
１	ディープインパクト	１．２
２	ドリームパスポート	１３．１
３	ダイワメジャー	１５．１
４	メイショウサムソン	２１．４
５	スイープトウショウ	２９．６
６	ポップロック	３１．１
７	アドマイヤメイン	３５．４
８	コスモバルク	３６．４
９	デルタブルース	３９
１０	トウショウナイト	７３．９
１１	スウィフトカレント	９２
１２	アドマイヤフジ	１３１．４
１３	ウインジェネラーレ	２０９．９
１４	トーセンシャナオー	２１８．２

２００６年　有馬記念
レース結果

人気	着順
１	１
２	４
３	３
４	５
５	１０
６	２
７	９
８	１１
９	６
１０	７
１１	１２
１２	８
１３	１３
１４	１４

２００７年　有馬記念

高

人気

低

人気	馬の名前	オッズ
１	メイショウサムソン	２．４
２	ポップロック	５
３	ウオッカ	６．９
４	ロックドゥカンブ	７．５
５	ダイワスカーレット	８．１
６	ダイワメジャー	１５．２
７	ドリームパスポート	２０．５
８	インティライミ	２７．４
９	マツリダゴッホ	５２．３
１０	デルタブルース	５９．６
１１	チョウサン	７１．１
１２	コスモバルク	８７．６
１３	ハイアーゲーム	９４．１
１４	サンツェッペリン	１１１．１
１５	レゴラス	１３７．６

２００７年　有馬記念
レース結果

人気	着順
１	８
２	５
３	１１
４	４
５	２
６	３
７	６
８	９
９	１
１０	12
１１	１3
１２	10
１３	14
１４	15
１５	7

人気、オッズ、得票率：投票

VS

着順、勝ち負け：結果

ここに、一般的な原理

「**Physics**」

はあるのか？

答え

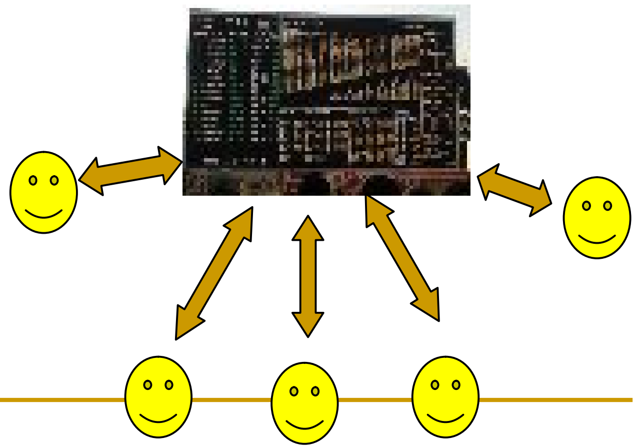
ある。

「スケール不変性」

今回のテーマ

投票モデル

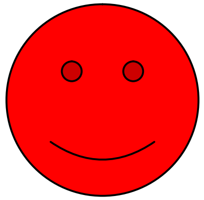
- 「真の情報」と「スコア」
- スケール不変性の創発
- 競馬市場での実証



投票モデル

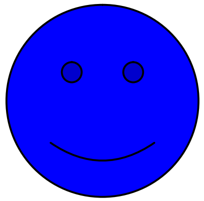
二値の候補者

$$N\text{候補者}\begin{cases} N_s \\ \\ N_w \end{cases}$$



「優秀」

勝ち馬
(着順の高い馬)



「劣等」

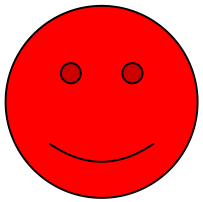
負け馬
(着順の低い馬)

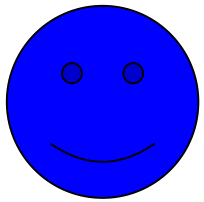
二値の情報とスコア

「二値の情報」はレース後、開票後でないとわからない。

「神のみぞ知る情報」

投票者の知りうる情報＝「スコア」

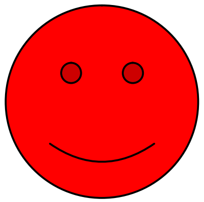
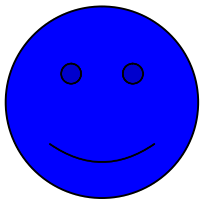
 : s

 : w

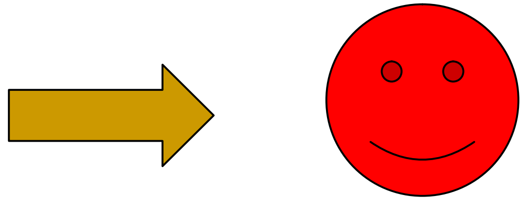
投票モデルI

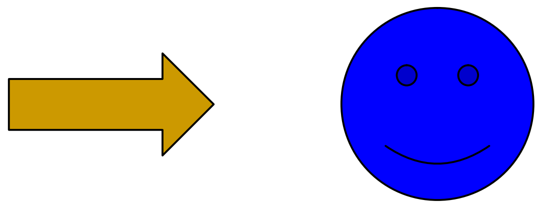
スコアに比例した確率で投票

$$N_s = N_w = 1$$

	スコア	得票確率
	$: s$	$p^s = \dfrac{s}{s+w}$
	$: w$	$p^w = \dfrac{w}{s+w}$

投票結果　$t \to \infty$

$s > w$  が勝つ

$w > s$  が勝つ

投票モデルI　：比率s/wのみ意味をもつ。

投票結果の解釈

●投票者がスコアのみに基づいて投票

●スコアが正しい ➡ 「優秀」な候補者を必ず選ぶ。

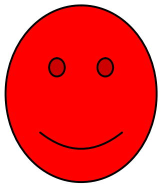
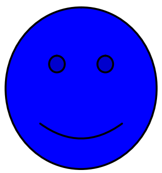
$$s > w$$

●スコアが間違い ➡ 「劣等」な候補者を必ず選ぶ。

$$w > s$$

選別能力は１００％か０％。

投票モデルI　一般の N_s, N_w

	得票数	得票確率
	$\{X^{s}_{i,t}\}_{i=1,\cdots,N_s}$	$p^s_i = \dfrac{s}{N_s\cdot s + N_w\cdot w}$
	$\{X^{w}_{i,t}\}_{i=1,\cdots,N_w}$	$p^w_i = \dfrac{w}{N_s\cdot s + N_w\cdot w}$

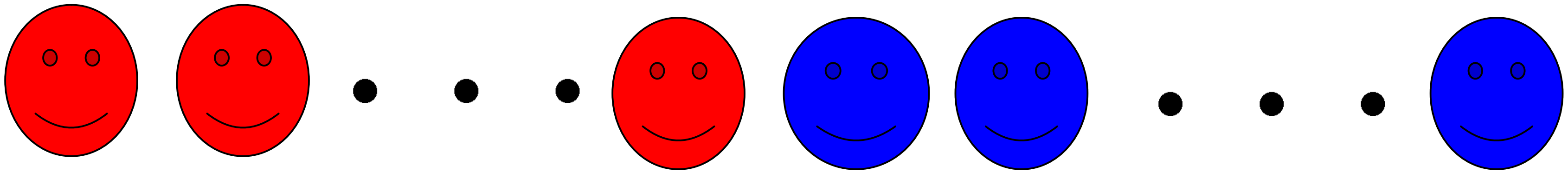
$X^{\mu}_{i,t}$:時刻tでのi番目のμ候補者の得票数

$X^{\mu}_{i,0} = 0$　　$\mu \in \{s, w\}$

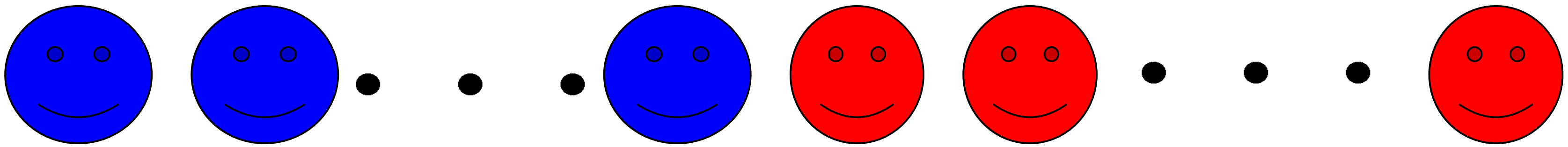
投票結果

得票数順に候補者を並べると

$$s > w \Longrightarrow \mathrm{Prob.}(X^{s}_{i,t} > X^{w}_{j,t}) = 100\%$$



$$w > s \Longrightarrow \mathrm{Prob.}(X^{s}_{i,t} > X^{w}_{j,t}) = 0\%$$



濃淡のない「白」「黒」ハッキリした世界。

投票者はインテリジェント？

スコアのみに基づいて投票？

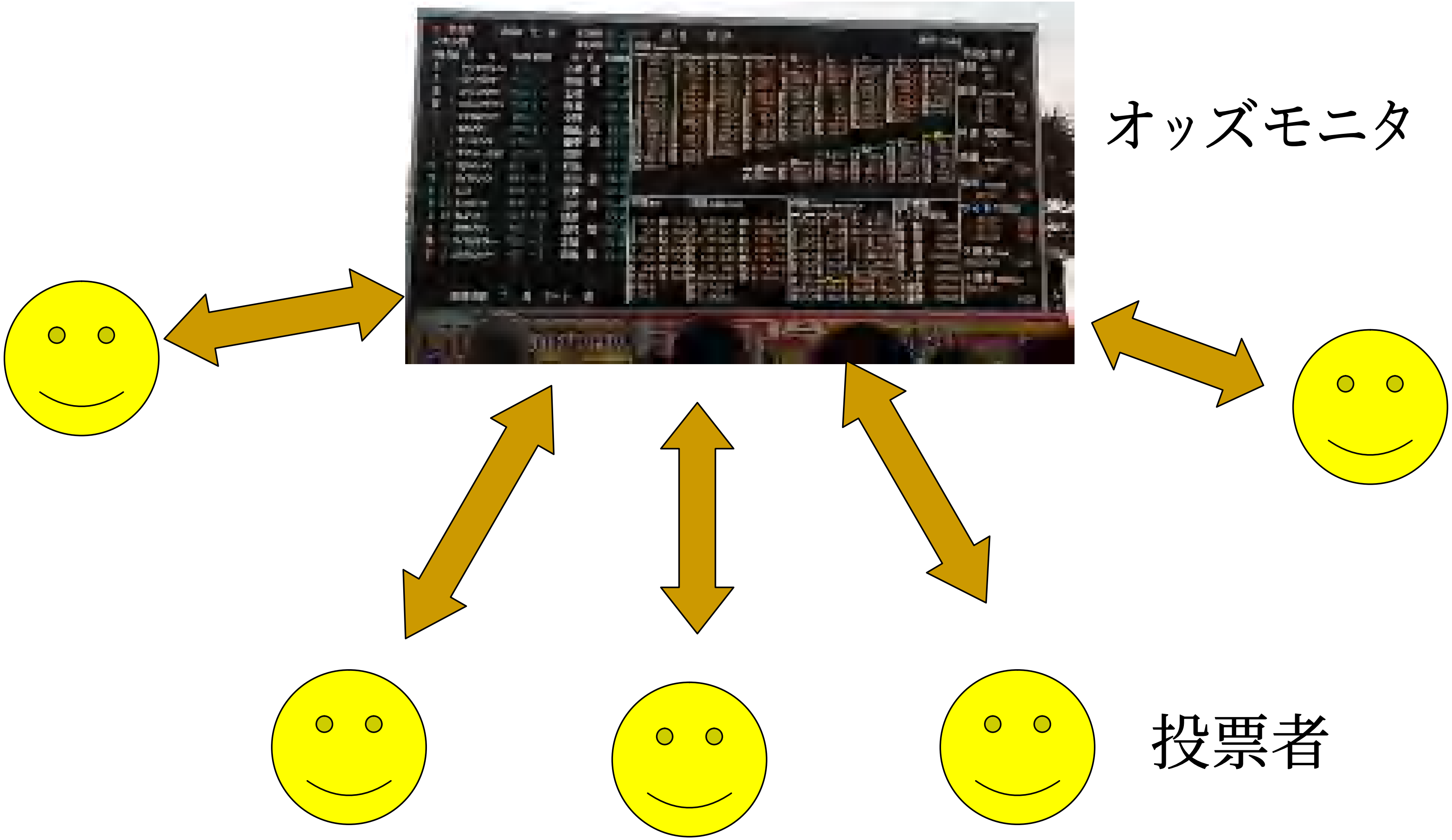
答え：Ｎｏ　！

ニュース、風評、他の投票者の動向に左右される

コピーキャット

競馬では、投票結果（オッズ）がリアルタイムに表示され、それをもとに買う馬券を決定。

ネットワーク構造



画像:http://www.flickr.com/photos/aquasaga/331836798/

インテリジェント＋コピーキャット

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

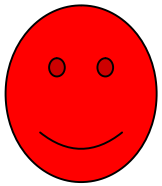
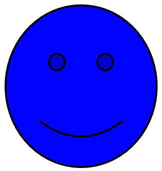
$$p_{i,t}^\mu = \frac{\mu + X_{i,t}^\mu}{N_s \cdot s + N_w \cdot w + t}$$

投票モデルII

一票の価値を基準に、s、wがスケールされる。

s/wと**|s|**が意味をもつ。

投票モデルII

	得票数	得票確率
	$\{X^s_{i,t}\}_{i=1,\cdots,N_s}$	$p^s_{i,t} = \dfrac{s + X^s_{i,t}}{N_s \cdot s + N_w \cdot w + t}$
	$\{X^w_{i,t}\}_{i=1,\cdots,N_w}$	$p^w_{i,t} = \dfrac{w + X^w_{i,t}}{N_s \cdot s + N_w \cdot w + t}$

投票モデルII　$N_s = N_w = 1$

初期条件　t＝0

$$\left\{ \begin{array}{l} p_1^s = \dfrac{s}{s+w}, X_0^s = 0 \\[2ex] p_0^w = \dfrac{w}{s+w}, X_0^w = 0 \end{array} \right.$$

$$\left\{ \begin{array}{l} p_1^s = \dfrac{s+1}{s+1+w}, X_1^s = 1 \\[2ex] p_1^w = \dfrac{w}{s+1+w}, X_1^w = 0 \end{array} \right.$$

t＝**1**

$$\left\{ \begin{array}{l} p_1^s = \dfrac{s}{s+w+1}, X_1^s = 0 \\[2ex] p_1^w = \dfrac{w+1}{s+w+1}, X_1^w = 1 \end{array} \right.$$

インテリジェント＋コピーキャット

$$p_{i,t}^{\mu} = \frac{\mu}{N_s \cdot s + N_w \cdot w + t} + \frac{X_{i,t}^{\mu}}{N_s \cdot s + N_w \cdot w + t}$$

$$\begin{cases} \to \dfrac{\mu}{N_s \cdot s + N_w \cdot w + t} & t \to 0 \\ \text{初期はインテリジェント} & \\ \to \dfrac{X_{i,t}^{\mu}}{N_s \cdot s + N_w \cdot w + t} & t \to \infty \end{cases}$$

後期はコピーキャット

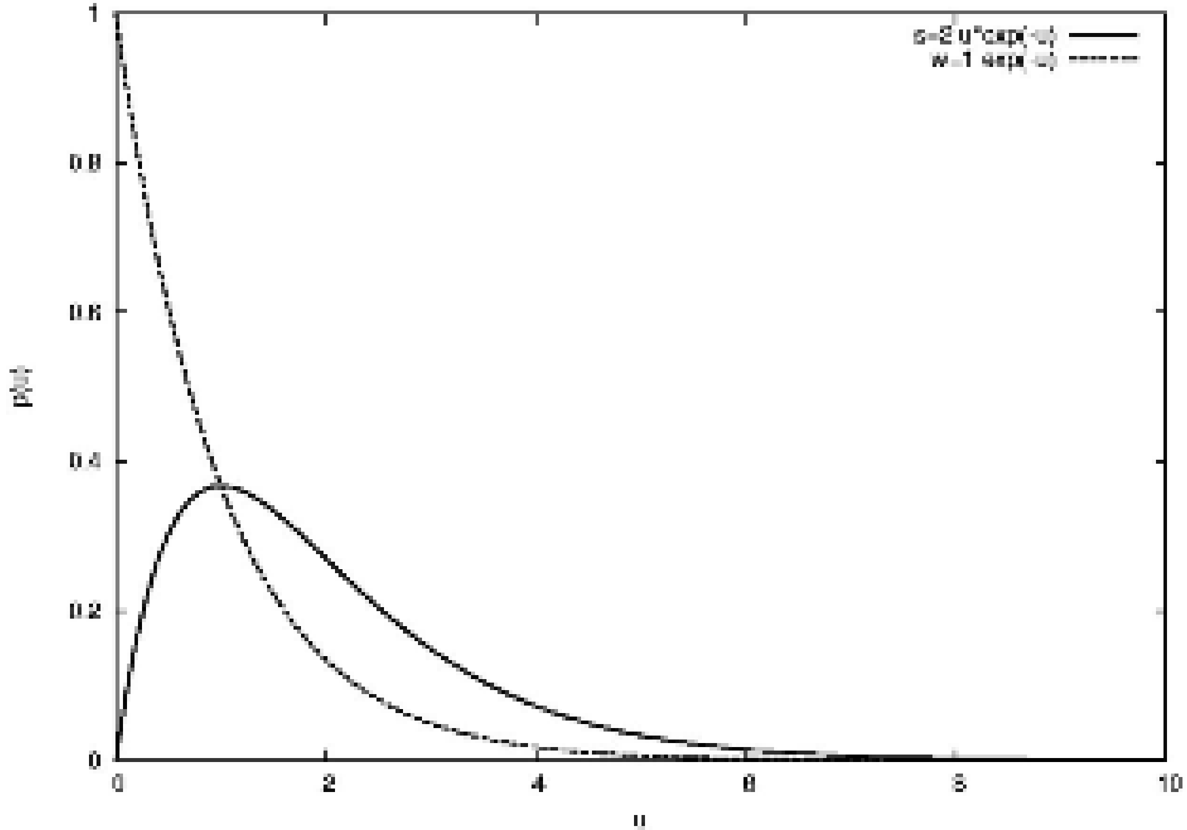
投票結果　$t \to \infty$

$X_{i,t}^{\mu}$ はガンマ分布に従う。

$$X_{i,t}^{\mu} \propto p_{\mu}(u) = \frac{1}{\Gamma(\mu)} e^{-u} \cdot u^{\mu-1}$$

分岐過程 (連続時間モデル) にマップすると厳密に解ける。

投票結果　$s = 2, w = 1$



s > w　でも、「優秀」な候補者が必ず勝つわけではない。

選別能力

$$\mathrm{Prob.}(X^{s}_{i,t} > X^{w}_{j,t}) = \mathrm{B}(0.5, s, w)$$

$$t \to \infty$$

$$\mathrm{B}(x, s, w) = \frac{1}{B(s, w)} \int_0^x t^{s-1} \cdot (1-t)^{w-1} dx$$

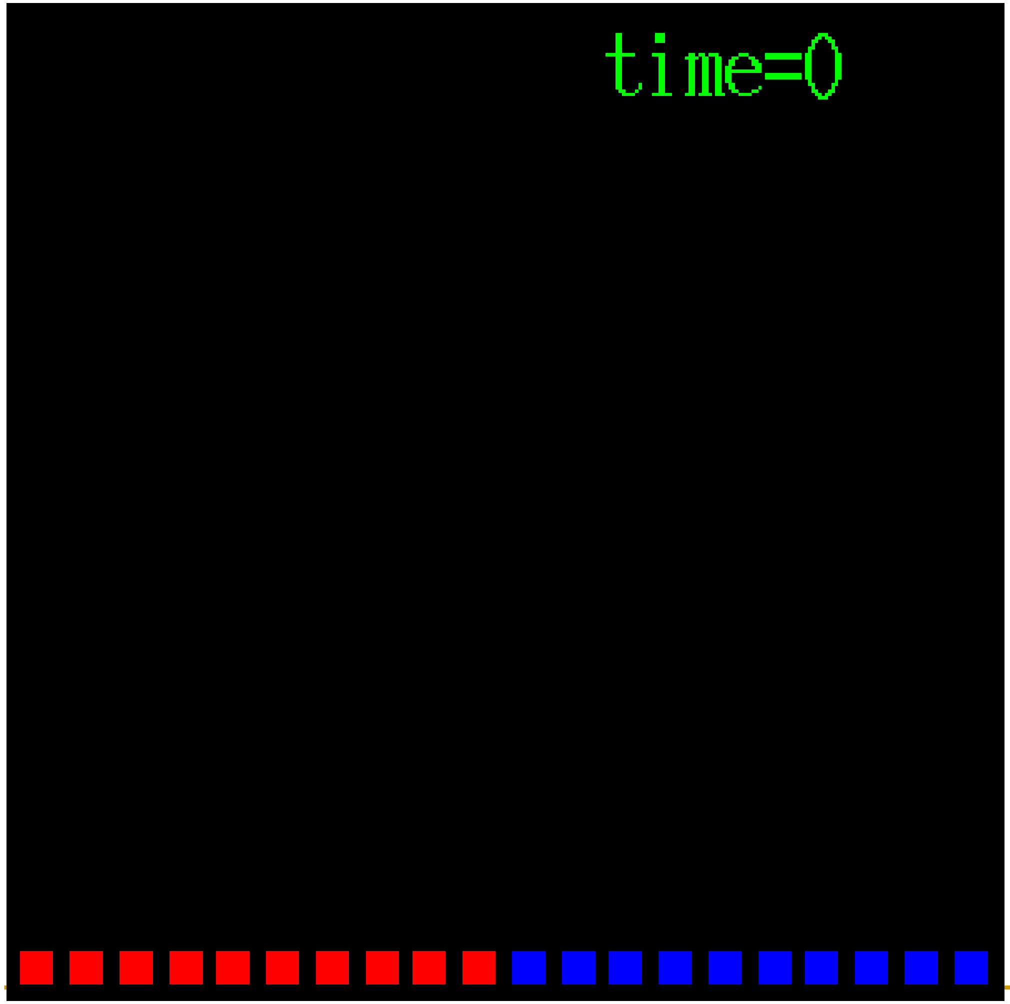
$$s = 2, w = 1$$

$$\mathrm{Prob.}(X^{s}_{i,t} > X^{w}_{j,t}) = 75\%$$

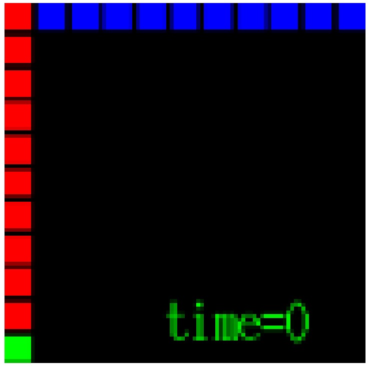
コピーキャットにより「選別能力」は低下した。

シミュレーション

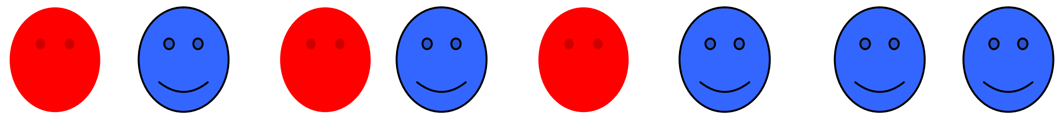
$$N = 20, N_s = 10, N_w = 10, s = 2, w = 1$$



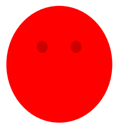
ROC カーブ

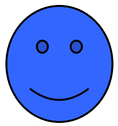


混合の可視化＝ROC　カーブ

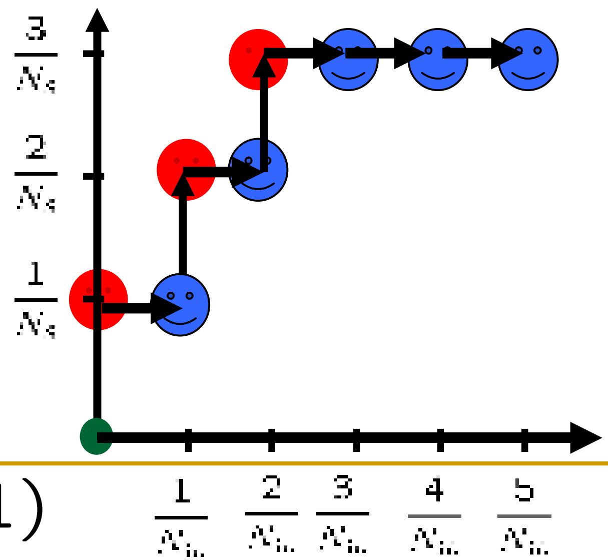


混合　$\{\mu_k\}_{k=1,\cdots,N} \in \{s, w\}$ ⟺ 経路 (x_k, y_k)

$\mu_k = s$  ⟺ 上に移動

$\mu_k = w$  ⟺ 右に移動

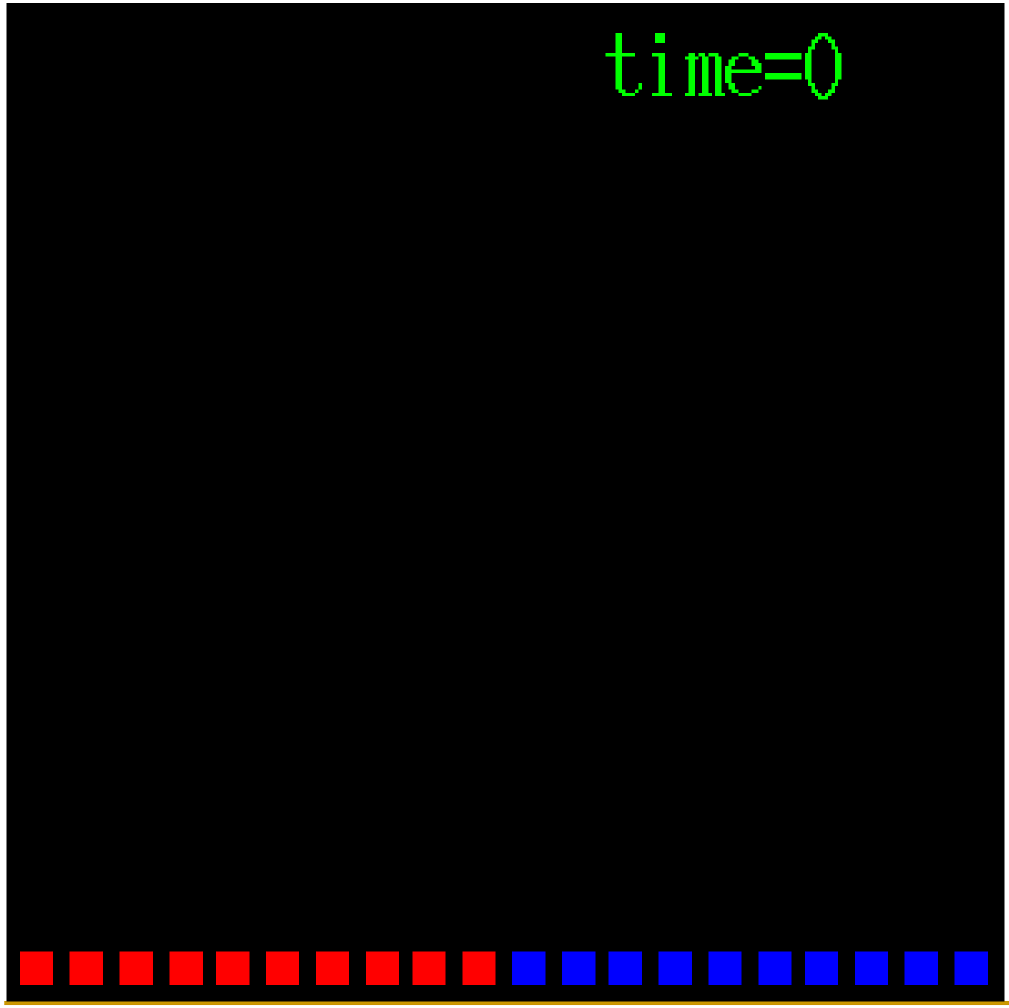
$$\begin{cases} x_k = \dfrac{1}{N_w} \displaystyle\sum_{l=1}^{k} \delta_{\mu_l, s} \\ y_k = \dfrac{1}{N_s} \displaystyle\sum_{l=1}^{k} \delta_{\mu_l, w} \end{cases}$$



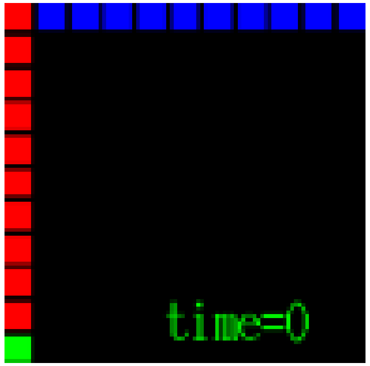
$(x_0, y_0) = (0, 0), (x_N, y_N) = (1, 1)$

シミュレーション

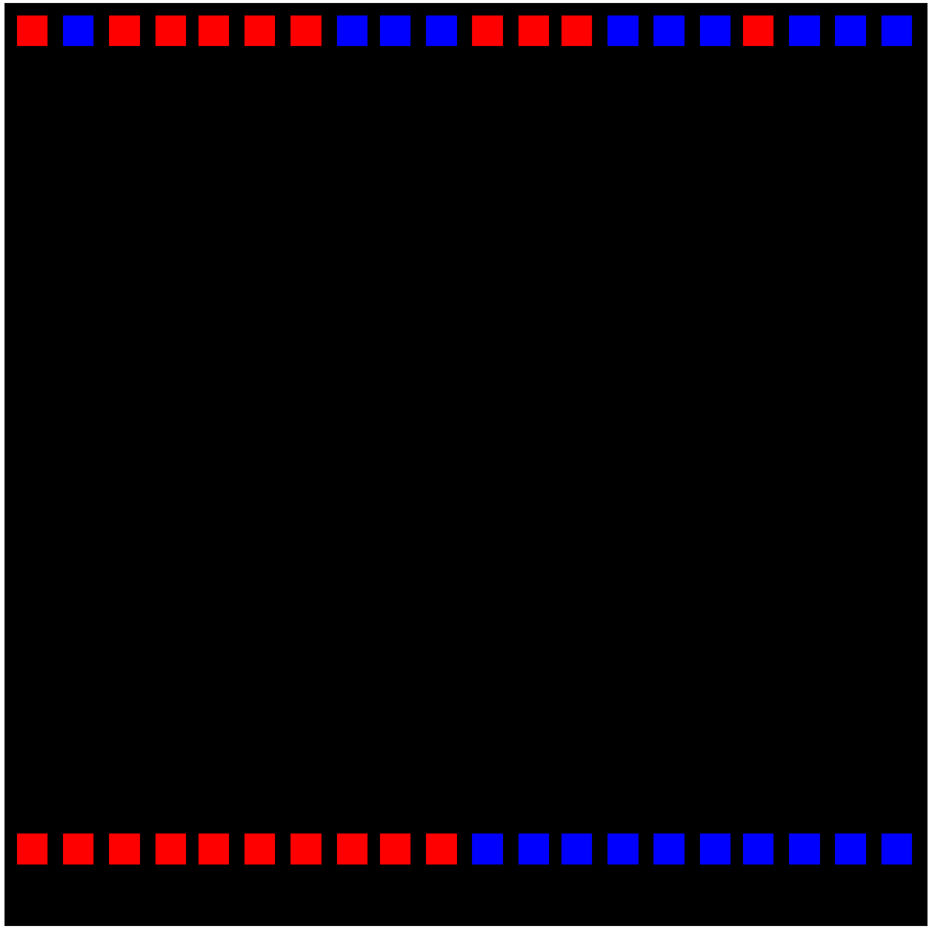
$$N = 20, N_s = 10, N_w = 10, s = 2, w = 1$$



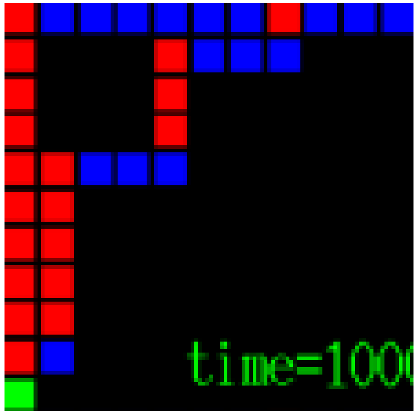
ROC カーブ



投票結果



ROC カーブ

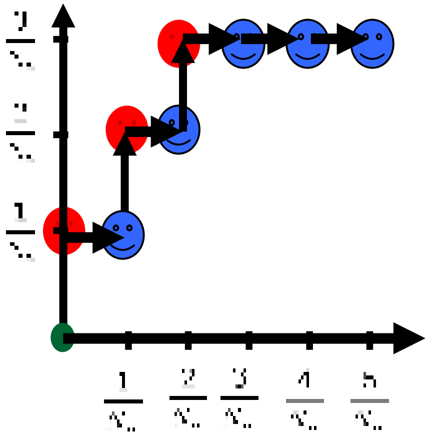


得票数
124
110
94
92
88
79
71
67
56
43
43
33
21
20
17
16
15
6
5
0

ROCカーブの連続極限

$N_s, N_w \to \infty, \quad r_s = \frac{N_s}{N}, r_w = \frac{N_w}{N}$ fixed

$$\begin{cases} x(t) = \int_t^{\infty} p_w(u)du \\ y(t) = \int_t^{\infty} p_s(u)du \end{cases}$$



$(x(0), y(0)) = (1, 1) \quad (x(\infty), y(\infty)) = (0, 0)$

$$\begin{cases} 1 - x(t) = \int_0^{t} p_w(u)du = \frac{1}{\Gamma(w)} \cdot \gamma(w, t) \\ 1 - y(t) = \int_0^{t} p_s(u)du = \frac{1}{\Gamma(s)} \cdot \gamma(s, t) \end{cases}$$

$\gamma(a, t) \equiv \int_0^{t} e^{-u} \cdot u^{a-1} du$ **: Incomplete Gamma Function of the 1st Kind**

スケール不変性

(x,y)=(1,1) 近傍でのスケール不変性

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

$$(s, w) \to (0, 0) \quad \text{with} \quad \alpha = \frac{s}{w} \quad \text{fixed}$$ (一票の価値固定)

$$\boxed{1 - y = (1 - x)^{\alpha}}$$

選別能力 $\text{Prob.}(X_{i,t}^{s} > X_{j,t}^{w}) = \dfrac{\alpha}{\alpha + 1}$

スケール不変性の意味

$(x, y) \rightarrow (1, 1)$ ＝得票数が少ない極限

$$1 - y \simeq (1 - x)^{\alpha}$$

二値の候補者の相対確率

$$\rho_{s/w} = \frac{dy}{dx} \propto (1 - x)^{\alpha - 1}$$

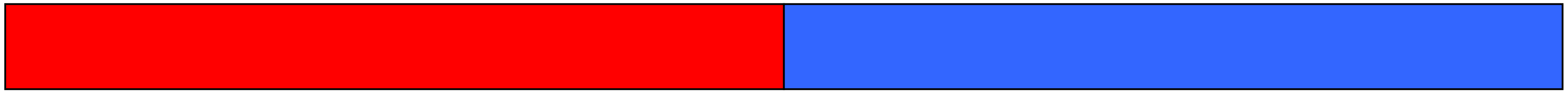
「優秀」な候補者は、どんな低得票数のところにも必ず存在する。

濃淡のあるグレーの世界。

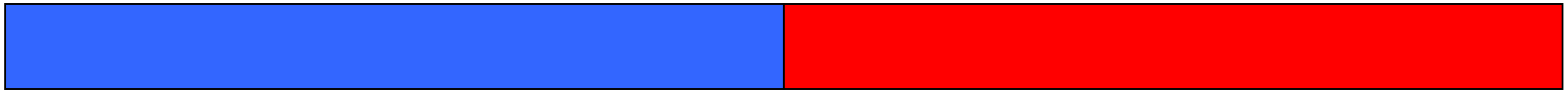
投票モデルの性質のまとめ

インテリジェントのみ： **投票モデルI**

$s > w$ ⇒ 選別能力 100%



$w > s$ ⇒ 選別能力 0%



濃淡のない「白」「黒」ハッキリした世界。

投票モデルの性質のまとめ

インテリジェント＋コピーキャット:　**投票モデルII**

選別能力 $= \mathrm{B}(0.5, s, w)$

濃淡のあるグレーの世界。

コピーキャットの存在の効果

$s > w$ ⇒ $100\% > \mathrm{B}(0.5, s, w)$

モデル Ⅰ　　　　モデル Ⅱ

コピーキャットの存在により選別能力は低下！

$s < w$ ⇒ $0\% < \mathrm{B}(0.5, s, w)$

コピーキャットがいないと、完全に間違った選択を行う。

コピーキャットの存在の意義

「インテリジェント」のみ。スコアが正しいなら
「優秀」な候補者を多数決で選別可能。

「コピーキャット」が存在すると、選別能力が低下。

しかし、「優秀」か「劣等」かは神のみぞ知る。

投票者は、断片的な情報＝「スコア」しか知りえない。

スコアは間違っている可能性がある。

コピーキャットにより、スケール不変が創発。
完全に間違った選択をさけられる。

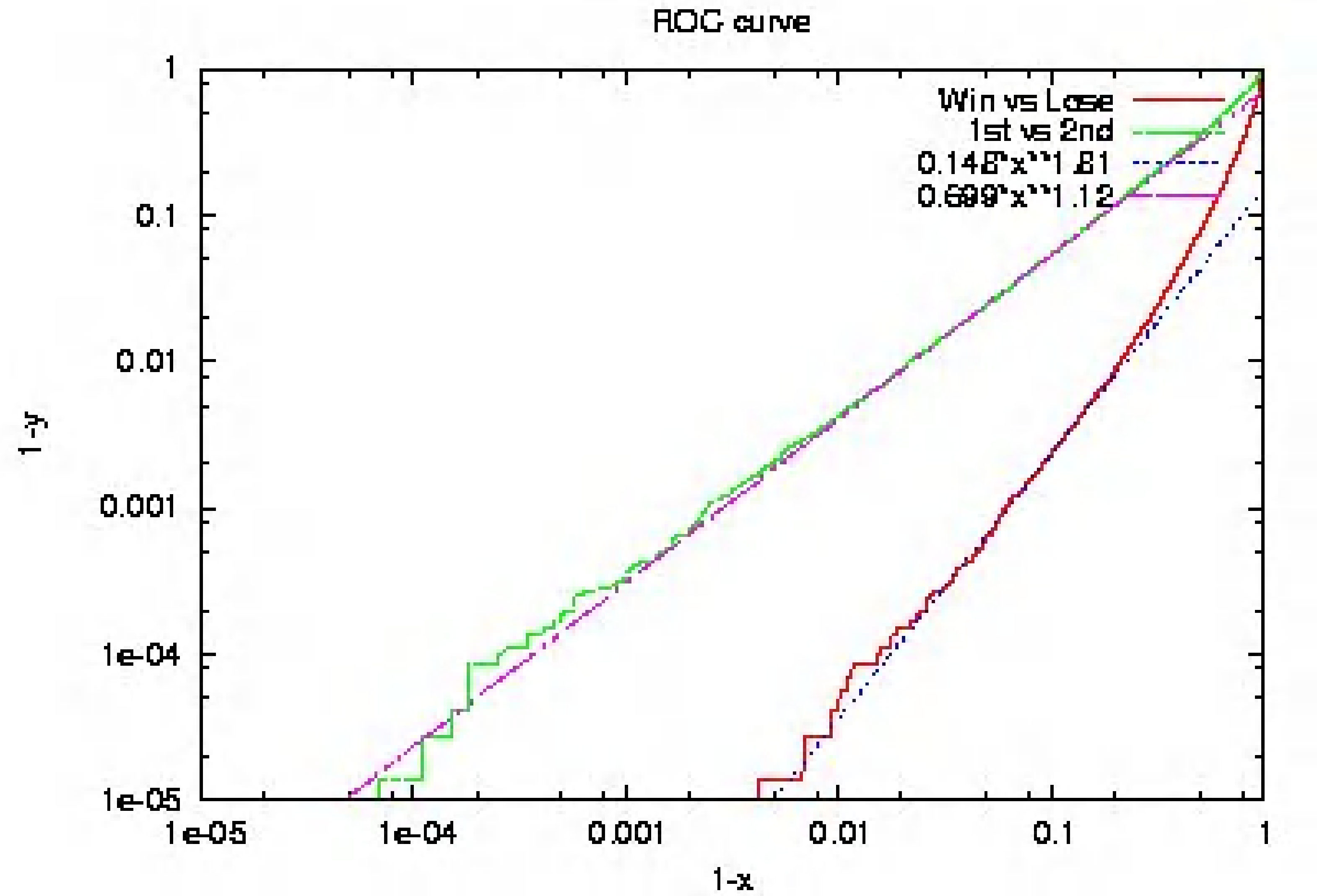
競馬のデータ解析

使用データ

- **JRA 1986-2006,**単勝市場の得票率と着順
- **901366**頭**, 71549**レース.

着順	出走数	平均得票率
1	**71650**	**21.23%**
2	**71590**	**15.40%**
負け	**829716**	**6.79%**
トータル	**901366**	**7.94%**

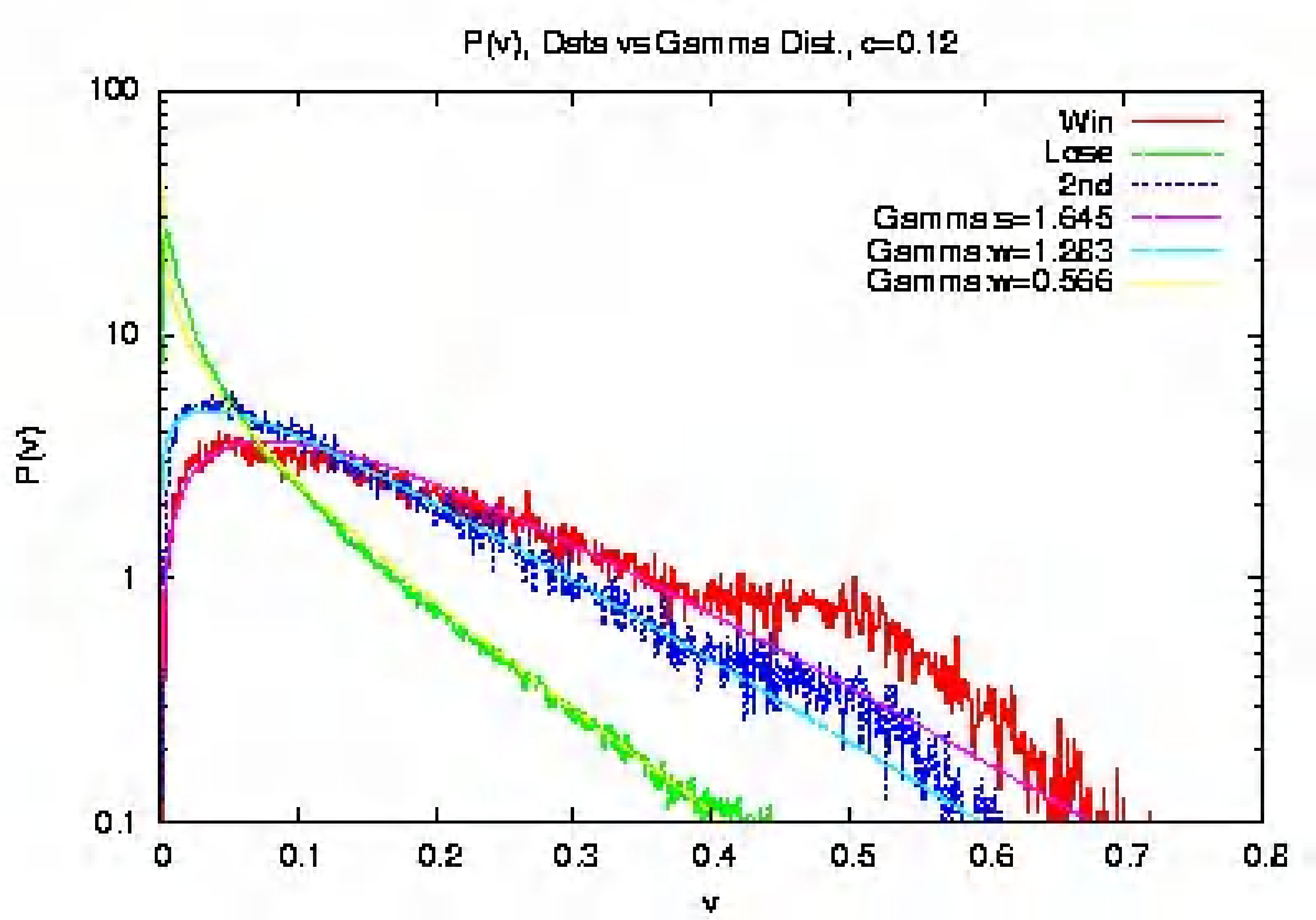
スケール不変性



ROC curve
Win vs Lose
1st vs 2nd
0.148*x**1.81
0.699*x**1.12
1
0.1
0.01
0.001
1e-04
1e-05
1e-05
1e-04
0.001
0.01
0.1
1
1-y
1-x

得票率の分布はガンマ分布？

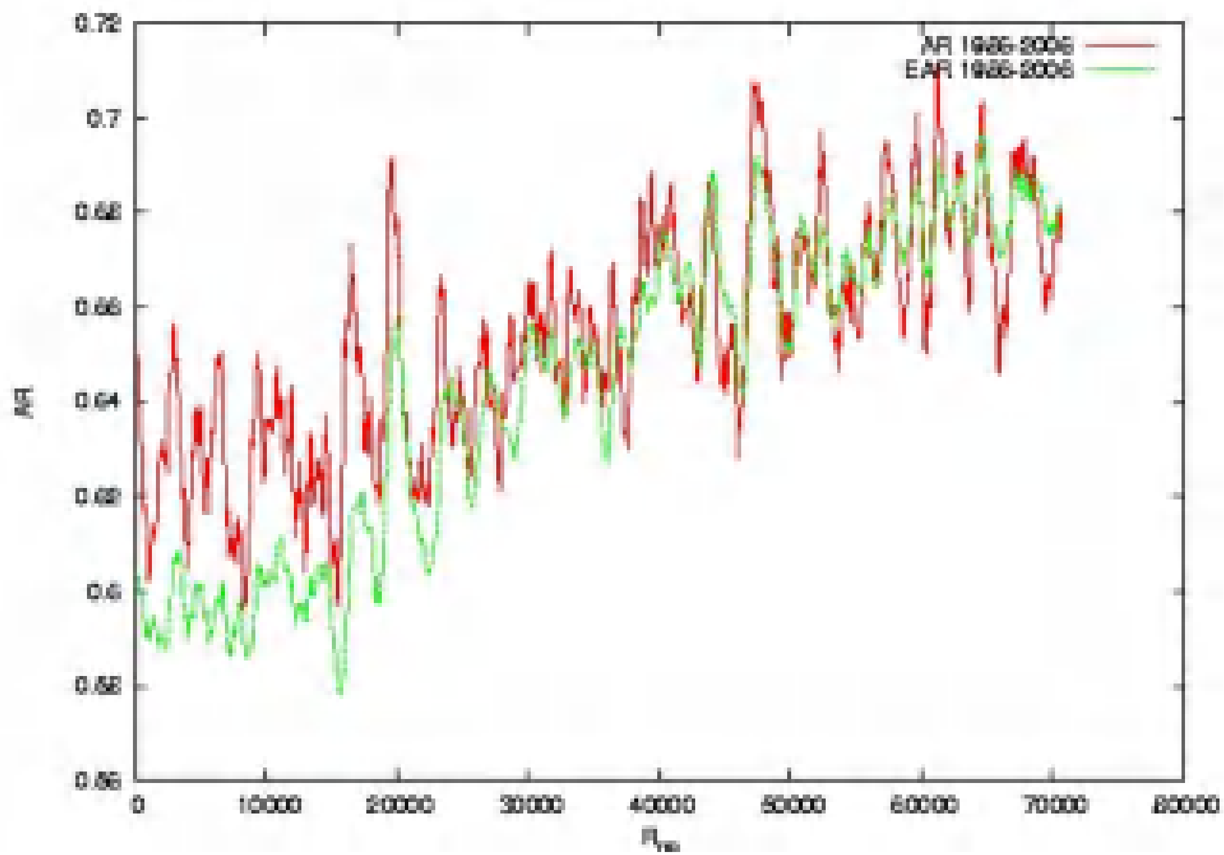
P(v)



選別能力の時間変化

勝ち負けの1000レース分のデータ

$$\mathsf{AR} \equiv 2 \times (\mathsf{Prob.}(X^{s}_{i,\infty} > X^{w}_{j,\infty}) - 0.5)$$

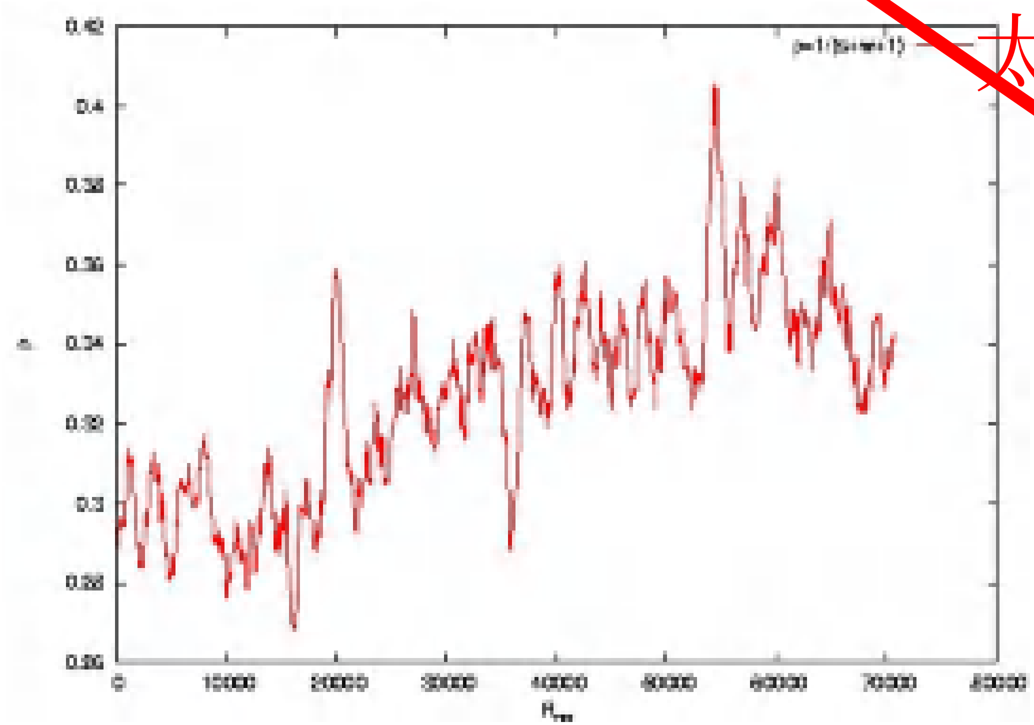


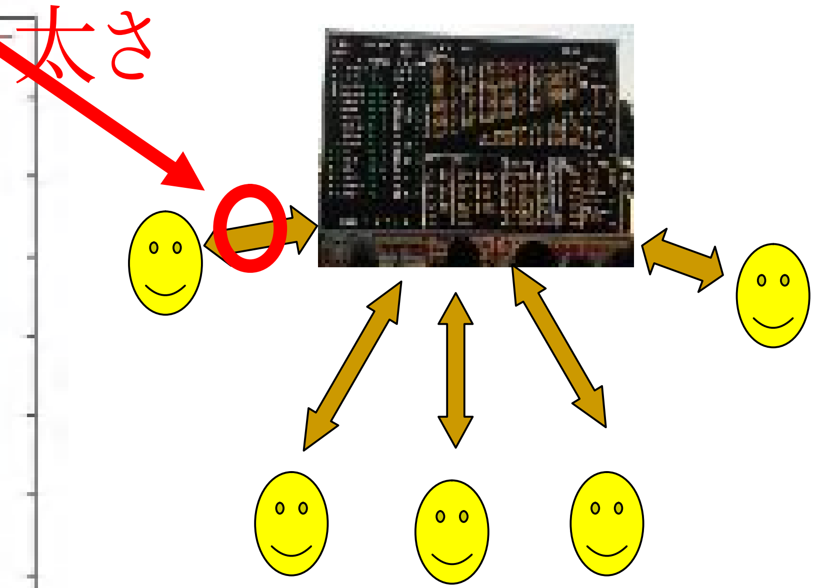
1986年　　　　　　　　　　　　2006年

コピーキャット度の時間変化

勝ち負けのデータ(選別能力)と平均得票率からスコアs、wを推定し、**相関係数**を計算。

$$\text{コピーキャット度}\ \rho = \frac{1}{s + w + 1}$$





1986年　　　　　　　　2006年

今回のテーマ

投票モデル

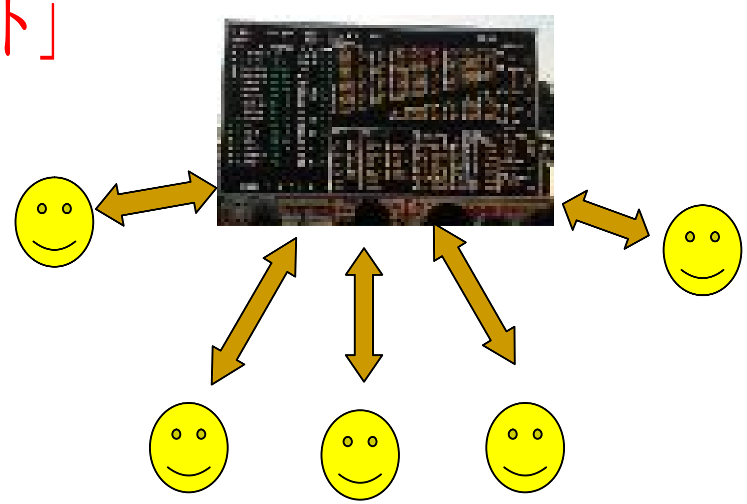
- 「真の情報」と「スコア」
- スケール不変性の創発
- 競馬データでの実証

結論

「コピーキャット」による「スケール不変性」の創発

「競馬ファンはコピーキャット」

「コピーキャットは有用」



Exact Scale Invariance in the Mixing of Binary Candidates in a Voting System
Shintaro Mori, Masato Hisakado, preprint **arXiv:physics-0806.0185**

ご清聴ありがとうございました。

ディープインパクト

現役期間	2004年〜2006年
英字表記	Deep Impact
品種	サラブレッド
性別	牡
毛色	鹿毛
生誕	2002年3月25日
抹消日	2006年12月25日
父	サンデーサイレンス
母	ウインドインハーヘア
生涯成績	14戦12勝 （中央競馬13戦12勝） （フランス1戦0勝）
獲得賞金	14億5455万1000円

画像：http://agasa2018.blogzine.jp/agasa2018/2006/05/post_d718.html

（Wikipediaより）

メイショウサムソン

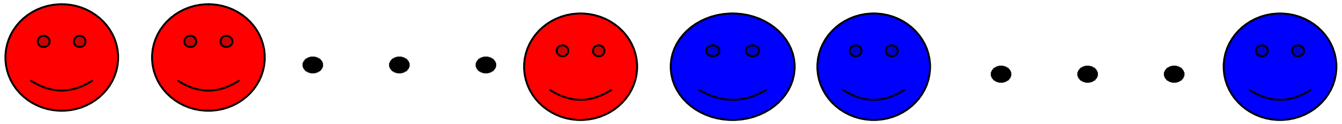
品種　サラブレッド
性別　牡
毛色　鹿毛
生誕　2003年3月7日（5歳）
父　オペラハウス
母　マイヴィヴィアン

生涯成績　24戦9勝
獲得賞金　10億6594万9000円

（Wikipediaより）

画像：http://keiba.vis.ne.jp/768/showimage.php?hid=379&size=0

厳密なスケール不変性の直感的な証明



$$(s, w) \to (0, 0)$$ ⟺ 一票の価値が無限大

(μ', j) が得票

$$p^{\mu}_{i,0} = \frac{\mu}{N_s \cdot s + N_w \cdot w}$$

$$p^{\mu'}_{j,1} = \frac{\mu' + 1}{N_s \cdot s + N_w \cdot w + 1} \to 1$$

$$p^{\mu''}_{k,1} = \frac{\mu''}{N_s \cdot s + N_w \cdot w + 1} \to 0$$

$$\mu'' \neq \mu' \text{ or } k \neq j$$

(μ', j) が他の候補者を圧倒し、得票順位が1番と確定。

厳密なスケール不変性の直感的な証明

１番の候補者を除外し、残り(N－1)人の候補者で考える。

次に選ばれた候補者の得票確率が1になり、順位が2番で確定。

投票による選択＝相対確率s、wでのランダムな選択

$$dx \propto (1-x) \cdot w \qquad dy \propto (1-y) \cdot s$$

$$\Rightarrow \frac{dy}{dx} = \alpha \frac{1-y}{1-x}$$

$$\Rightarrow 1-y = (1-x)^{\alpha}$$

Proof　厳密なスケール不変性の数学的な証明

$$\gamma(s,t) = \frac{1}{s} t^s \cdot e^{-t} \cdot M(1, s+1, t)$$

$$M(1, a+1, t) = 1 + \frac{1}{a+1} t + \frac{1}{(a+1)(a+2)} t^2 + \frac{1}{(a+1)(a+2)(a+3)} t^3 \cdots$$

Kummer's confluent hypergeometric function

$$M(a,b,t) \equiv \sum_{n=0}^{\infty} \frac{(a)_n \cdot t^n}{(b)_n \cdot n!}$$

$(a)_n$: rising factorial

$$M(a,b,t) = e^t \cdot M(b-a, b, -t)$$

$$e^{-t} \cdot M(1, a+1, t) = M(a, a+1, -t)$$

$$\gamma(s,t) = \frac{1}{s} t^s \cdot M(s, s+1, -t)$$

$$1 - x(t) = \frac{t^w}{\Gamma(w+1)} \cdot M(w, w+1, -t) \qquad 1 - y(t) = \frac{t^s}{\Gamma(s+1)} \cdot M(s, s+1, -t)$$

$$(1-x)^{\alpha} = (1-y) \cdot \frac{\Gamma(s+1)}{M(s, s+1, -y)} \cdot \left(\frac{M(w, w+1, -t)}{\Gamma(w+1)} \right)^{\alpha} \qquad \alpha = \frac{s}{w}$$

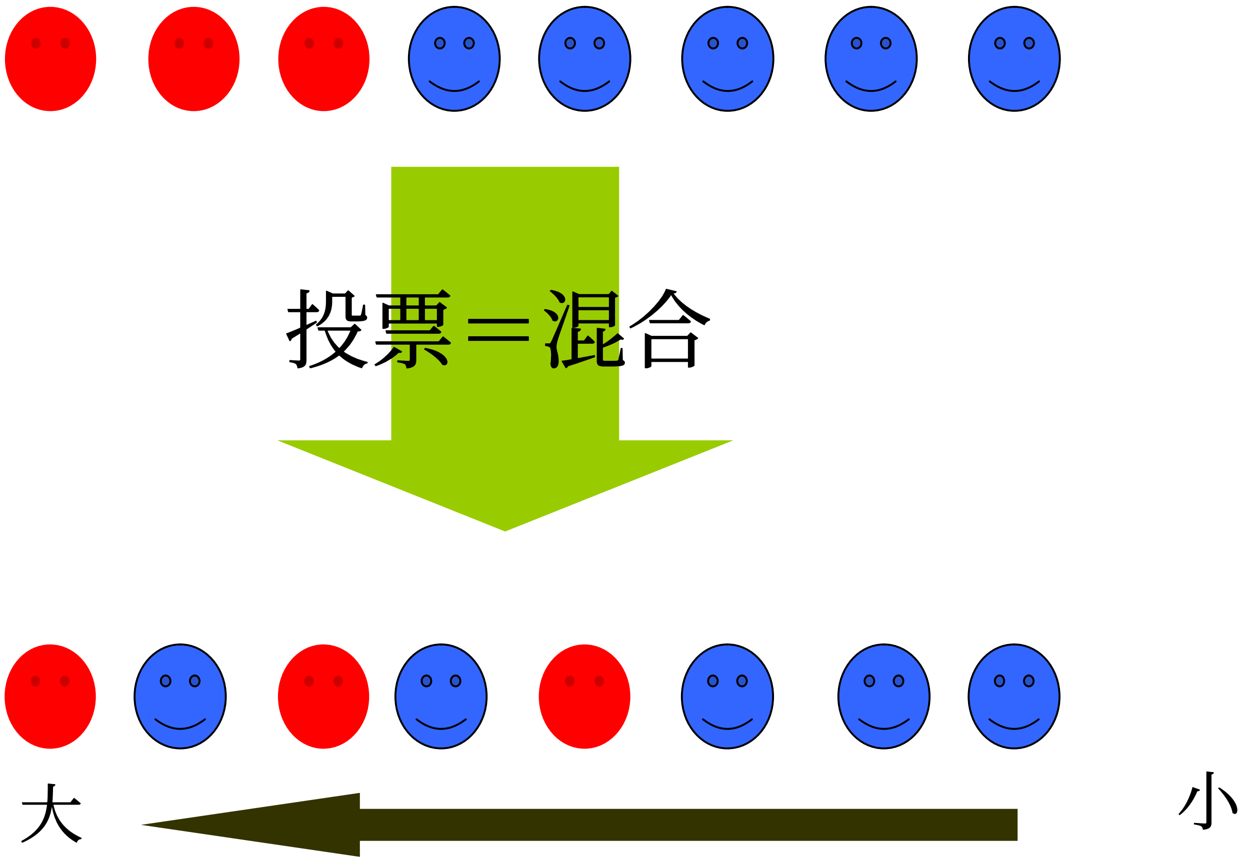
$$s \to 0 \qquad \Gamma(s+1), M(s, s+1, -t) \to 1$$

$$\Rightarrow \quad 1 - y = (1-x)^{\alpha}$$

投票と混合

$N_s = 3, N_w = 5$

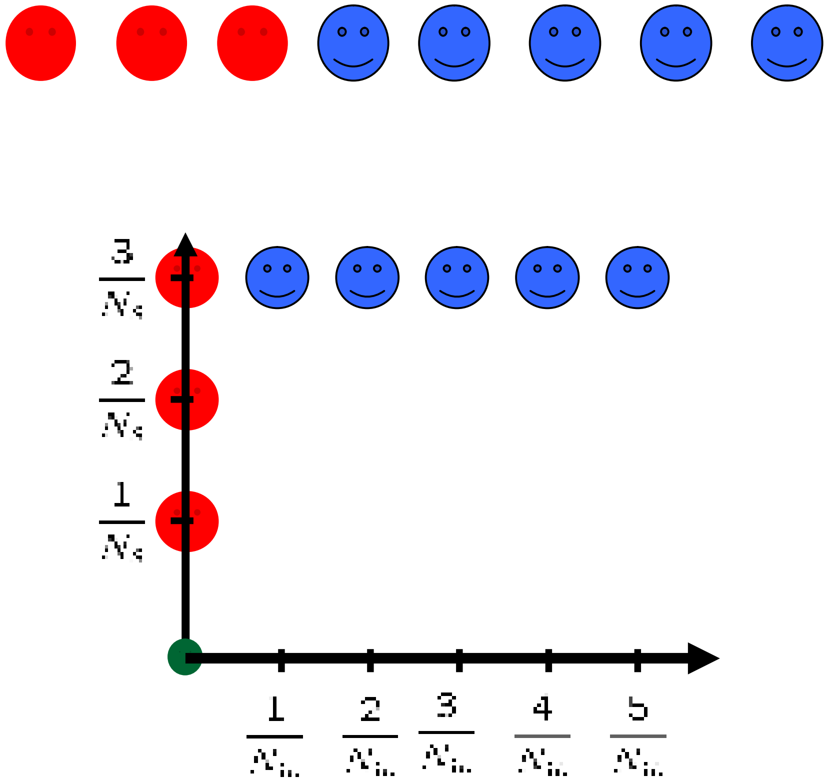
「優秀」な候補者　「劣等」な候補者



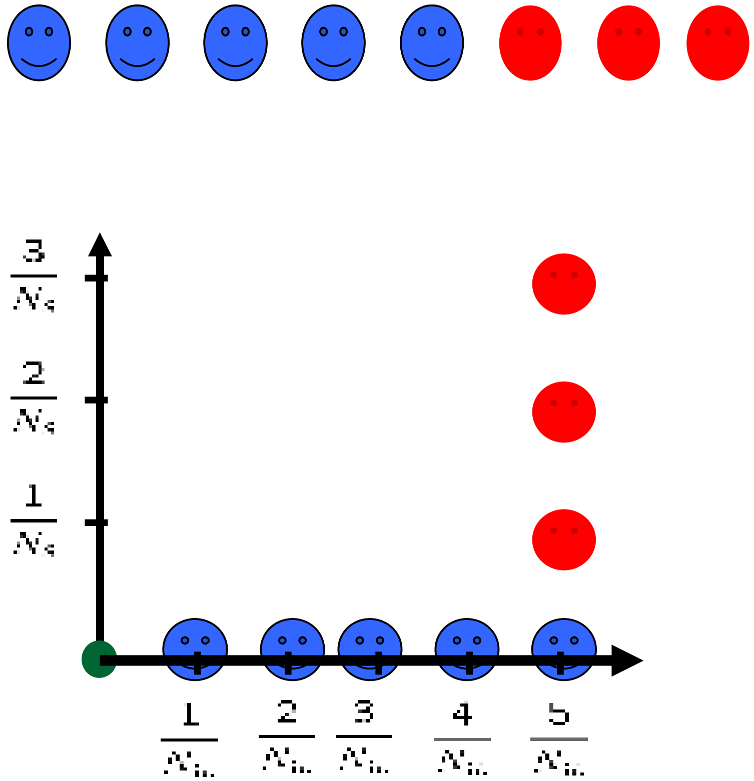
得票数

混合とROCカーブ

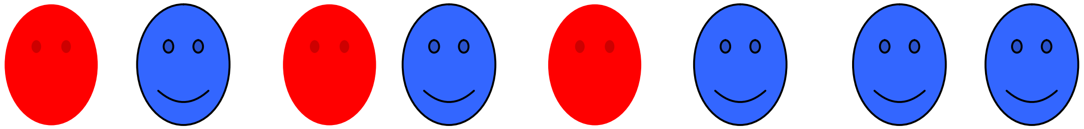
選別能力１００％

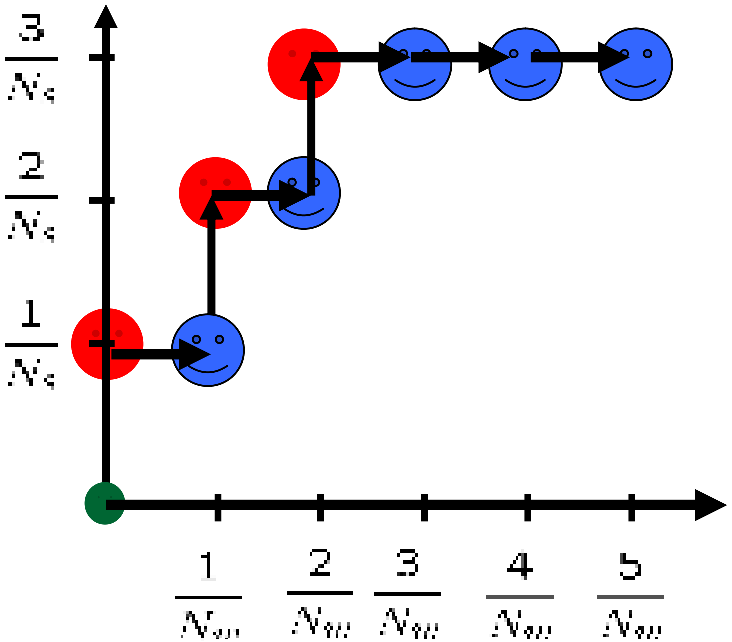


選別能力　０％



混合とROCカーブ





選別能力

$$\text{Prob.}(X_{i,t}^{s} > X_{j,t}^{w}) = 12/15 = 80\%$$

$$\text{Prob.}(X_{i,t}^{s} > X_{j,t}^{w}) = \sum_{j=1}^{N} y_j \cdot (x_j - x_{j-1})$$