# レイティングのベイズ統計モデリング [1]

## Bayesian Statistical Modeling of Ratings

ネットワーク情報学部　石鎚英也

School of Network and Information　Hideya ISHIZUCHI

**Keywords:** Bayesian, statistical modeling, rating, Elo, Massey, RStan

**Abstract**

In the world of games and sports, various rating methods have been developed and used in order to rank players or teams according to their real abilities - Elo rating in chess and Massey rating in college football, for instance. Players and teams are re-rated according to the results of matches, and rating scores and rankings can be updated instantly. However, the followings may be pointed out as drawbacks of prevailing rating methods: It possibly takes time to get stable rating values. It is difficult to properly set parameters included in the updating formulae. And rating values tend to be inflationary.

In this study, traditional and standard rating methods are extended by Bayesian statistical modeling so as to mitigate such drawbacks. Then, they are applied to real records of games and sports.

## はじめに

ブログや動画等の人気ランキングからホテルやレストランの星の数，そして企業や国債の信用格付けに至るまで，私たちの身の回りには，諸種のランキングやレイティングの情報が溢れている．映画鑑賞の際の年齢制限の規定のようなものもレイティングと呼ばれるが，本論文では，主にゲームやスポーツ分野を想定し，プレーヤーやチームの順位付け（ランキング）を行うために用いられる評価指標（レイティング）を取り扱う．

勿論，レイティングの問題は，単に評価対象のランキングへの興味に限られるものではなく，広範な分野と関連性がある．例えば，web 上の検索サービス，広告への商品の掲載，通販サイトのレコメンデーションといったビジネスの場面では，提示する情報の順序が重要な意味を持ち，その意味では， PageRank や協調フィルタリングなど，そこで用いられるアルゴリズムは，レイティングのための方法とみなすこともできる[15]．レイティングは，また，社会選択理論のようなより理論的な分野とも関連が深い[8]．社会選択理論では，個人の選好と集団的な決定との関係を扱うが，個々の試合結果をできるだけ反映するようなレイティング方法を考えることと，個人の選好をできるだけ反映するように集団的決定を行うこととは，よく似ているからである．

ゲームやスポーツの世界では，試合の勝敗データからランキングを求めるのに様々なレイティング手法が開発され使われてきた．例えば，チェスにおけるイロ・レイティング（Elo rating）[10]や大学フットボールでのマシー・レイティング（Massey rating）[9]などである．そして，試合の結果に従って，参加者，チームが再レイティングされ，レイティングのスコアやランキングの更新が日々繰り返されるようになっている．

しかし，以下のような事項が，レイティングの欠点として挙げられることがある（[5]p.188, [2]）．

- 初期値から落ち着くまでに適正な評価が難しい，あるいは，収束までに時間がかかる．
- 評価のタイミングに左右されやすい．
- レイティングの更新のためのパラメータの適切な設定が困難である．
- インフレしやすい [2].

本研究では，こうした欠点を緩和するため，伝統的，標準的なレイティング手法を，ベイズ統計モデリング（例えば[1], [4], [5]）により拡張し，現実のデータに適用することを試みるものである．

---

[1] 本研究は，2019 年度専修大学長期在外研究での成果公開の１つである．

[2] １プレーヤー（チーム）あたりのレイティングの平均値が経時的に安定せず，上昇（インフレ）や下降（デフレ）のトレンドがしばしば生じる．

スポーツ分野でのビッグデータ活用ということもあり，例えば，一般化線形モデルなどの回帰分析でトーナメントの結果を予測するといった研究[6]や，階層ベイズモデルによるレイティングを扱った研究[17]など，統計モデリングの応用は非常に多くあるようだが，本研究のような，従来からあるレイティング手法を直接的に拡張するアプローチは見かけない．

各章の内容は以下のとおりである：第 2 章では，伝統的，確定的なレイティング方法の例として，Elo レイティングと Massey レイティングを取り上げ，それぞれの数理的な概要を述べる．第 3 章では，Elo レイティングと Massey レイティングの基本的なアイデアをベイズ統計モデル化する．第 4 章では，3 章のモデルを現実の試合結果（将棋棋士の勝敗データとラグビーチームの得点データ）に適用し，簡単な評価を加える．第 5 章はまとめである．

## 伝統的なレイティングについて [3]

例えば，試合 [4]の勝敗データからランキングを求める場合の自然なレイティング方法の 1 つは，プレーヤーの勝率を直接用いることであろう．総当り戦であれば，勝率は，プレーヤーの実力を反映する妥当な指標と考えられる（ただし，引き分けのあるゲームでは，その取扱い方法によって，勝率の数値に微妙な違いが生じ得る）．

しかしながら，勝ち抜き戦（ノックアウト方式）など，プレーヤーの組み合わせによっては対戦機会がない（あるいは対戦回数が異なる）場合だと，勝率は必ずしも妥当な指標とはならない．例えば，同じ勝率のプレーヤーでも，各対戦相手の強さによって評価は異なるであろう．また，プレーヤー数や対戦回数にもよるが，複数のプレーヤーの勝敗数が同じ（従って勝率が同じ）になるケースが多いと，ランキングのための評価指標として好ましくない．

こうしたことから，勝率以外の様々なレイティング方法が開発されてきた．この章では，伝統的，標準的と思われるレイティング方法のうち Elo レイティングと Massey レイティングを取り上げ，ごく簡単にその数理的な概要を示す（[8] [9] [10]等参照）．

### 2.1. Eloレイティングの概要

Elo レイティング（Elo rating）には様々な定義（変種）があるようだが，ここでは，「平均的」プレーヤーを相手にしたときの，あるプレーヤーの勝率を$q$とするとき，対数オッズ

$$r = \log(q/(1-q)) \stackrel{\text{def}}{=} \text{logit}(q)$$

をそのプレーヤーの（基本的な）Elo レイティングと考える．これは，logit の逆関数（logistic 関数 [5]）

$$\text{logistic}(r) \stackrel{\text{def}}{=} 1/(1+\exp(-\text{r}))$$

を使えば等価的に

$$q = \text{logistic}(r)$$

とも書ける．このとき，平均的プレーヤー同士の勝率を$q = 0.5$として，平均的プレーヤーのレイティングは$r = 0$である．

より一般的には，$R = a + br$（$b > 0$）なる$R$を Elo レイティングという．特に，対数の底を 10 とした対数オッズを使って Elo レイティングを

$$R = 1500 + 400\log_{10}(q/(1-q))$$

と定義することがあるが，この場合，

$$\log_{10}(q/(1-q)) = \log(q/(1-q))/\log(10) \fallingdotseq 0.434\, r$$

より，

$$R = 1500 + 400/\log(10)r$$

となる．つまり，

$$a = 1500, b = 400/\log(10) \fallingdotseq 173.7$$

である．この定義の場合，平均的プレーヤーのレイティングは$R = 1500$である．

プレーヤー$A$がプレーヤー$B$を相手にしたときの勝率を$q_{AB}$と記し，勝敗比は積によって推移する：

$$q_{AC}/q_{CA} = q_{AB}/q_{BA} \times q_{BC}/q_{CB}$$

と仮定する．この時，プレーヤー$A$と$B$のレイティングの差は，$A$の勝率の対数オッズに等しいことが示される：

$$r_A - r_B = \log(q_{AB}/(1-q_{AB}))$$

すなわち，プレーヤー$A$と$B$のレイティングの差$d_{AB}$は，logit 関数により

$$d_{AB} = \text{logit}(q_{AB})$$

と書け，逆に，$A$の勝率$q_{AB}$は logistic 関数により

$$q_{AB} = \text{logistic}(d_{AB}) = 1/(1+\exp(-d_{AB}))$$

と書ける．なお，上記$R$の定義による$A, B$のレイティングを$R_A, R_B$と書くと，$d_{AB} = r_A - r_B = \log(10)\cdot(R_A - R_B)/400$なので，$D_{AB} = (R_A - R_B)/400$として，

$$q_{AB} = 1/(1+10^{-D_{AB}})$$

とも書けることに注意する．

実際のレイティング計算においては，適当な初期値から始めて，例えば，以下の更新式のように試合のたびに数値を変更していくことが普通である：

$$r_A(t) = r_A(t-1) + K(s_{AB}(t) - p_{AB}(t-1))$$

ただし，プレーヤー$A$がプレーヤー$B$と（離散時刻）$t$で対

---

[3] レイティングに関する以下の記述では，主に，スポーツや対戦ゲームの状況を想定した用語を用いる（勝敗，得点など）．なお，対戦する主体を「プレーヤー」と呼ぶが，団体スポーツでは「チーム」の意味である．

[4] 以下では，特に言及しなければ，各試合は 2 人，あるいは 2 チームで争われる 2 人ゲームを想定している．

[5] ロジスティック関数は，より一般的に定義されることがあり，ここはその特殊なケースである．標準シグモイド関数（standard sigmoid function）とも呼ばれる．

戦したとし，

- $r_A(t)$：$t$での$A$のレイティング
- $K$：定数
- $s_{AB}(t)$：$A$と$B$の試合の$t$での結果（$A$の勝ち，$B$の勝ちに応じて$1, 0$）．引き分けを$0.5$としてカウントすることもある．
- $p_{AB}(t)$：$A$の$B$に対する勝率の$t$での見積もり．前述の通り，$p_{AB}(t) = \mathrm{logistic}(r_A(t) - r_B(t))$.

とする．

なお，ここまでは，試合の結果を勝敗とし，勝率を説明するものとしてレイティングを考えてきたが，試合の結果が，0以上1以下である何らかの別のスコアであったとしても，勝率の代わりにスコアの期待値を考えれば同様の議論となる．

## 2.2. Masseyレイティングの概要

Massey レイティング（Massey rating）は，試合の結果が得点（失点）である場合，あるいは得点差（margin of victory）である場合を扱うことができる（勝敗も得点とみなせるので扱える）．その基本的なアイデアは，試合の得点差が，その試合で対戦したプレーヤーのレイティングの差で近似できる，あるいは，そうなるようにレイティングを定義せよというものである．従って，理想的には以下の数式が成立する：

$$r_A - r_B = y_g$$

ただし，試合$g$でプレーヤー$A, B$が対戦したとし，

- $r_A, r_B$：プレーヤー$A, B$のレイティング
- $\mathrm{y}_g$：試合$g$における$A$の（$B$に対する）得点差

とする．$g, A, B$を自然数のインデックスとして表現すると，上式は，ベクトル・行列記法で

$$\mathrm{Xr} = \mathrm{y}$$

と書ける．試合$g$でプレーヤー$A, B$が対戦した上記のような場合だと，係数行列Xは，その$g$行目の$A, B$列の値がそれぞれ$1, -1$でそれ以外の列の値が$0$であるスパースな行列である．rはプレーヤーのレイティングを示すベクトル，yは試合での得点差を示すベクトルである．

この方程式$\mathrm{Xr} = \mathrm{y}$は，一般に解が存在しないが，通常の最小二乗法と同じく，両辺にXの転置行列$\mathrm{X}^T$を乗じて，正規方程式

$$\mathrm{X}^T\mathrm{Xr} = \mathrm{X}^T\mathrm{y}$$

を得ることができる．$\mathrm{M} \stackrel{\text{def}}{=} \mathrm{X}^T\mathrm{X}, \mathrm{p} \stackrel{\text{def}}{=} \mathrm{X}^T\mathrm{y}$と置けば，これは

$$\mathrm{Mr} = \mathrm{p}$$

と書ける．Mは Massey 行列と呼ばれる．Xの定義から，Mの対角成分$\mathrm{M}_{ii}$は第$i$プレーヤーの試合総数であり，$\mathrm{M}_{ij}\ (i \neq j)$はプレーヤー$i$と$j$の対戦回数に$-1$を乗じた値となる．また，pは累積の得点差を意味し，累積得点をf，累積失点をaとすれば，$\mathrm{p} = \mathrm{f} - \mathrm{a}$である．

Massey 行列Mはフルランクではなく，正規方程式には一意の解が存在しない[6]．そこで，例えば，全プレーヤーのレイティングの合計を0とするといった制限を加えて，方程式を修正することが行われる（例えば，Mの最後の行の要素を全て1に変えて$\bar{\mathrm{M}}$とし，右辺のベクトルpの最後の要素を0に変えて$\bar{\mathrm{p}}$とする）．その上で修正された正規方程式

$$\bar{\mathrm{M}}\mathrm{r} = \bar{\mathrm{p}}$$

を解けば，レイティングベクトルrを得ることができる．

Massey レイティングでは，そのようにして得られた総合的なレイティングに加えて，攻撃のレイティングと守備のレイティングも以下のように求めることができる：Massey 行列Mは，対角成分がMと同じ対角行列Tと，対角成分が全て0で非対角成分がMの対応する成分に$-1$を乗じた値である行列Pの差として$\mathrm{M} = \mathrm{T} - \mathrm{P}$と表現できる．このとき，Tの対角要素は各プレーヤーの試合総数を示し，Pの要素は各プレーヤーペア[7]の試合総数を示している．攻撃のレイティングをoとし，守備のレイティングをdとして，

$$\mathrm{r} = \mathrm{o} + \mathrm{d}$$

を仮定し，正規方程式を変形すると，

$$(\mathrm{T} - \mathrm{P})(o + \mathrm{d}) = \mathrm{f} - \mathrm{a}$$

から

$$(\mathrm{To} - \mathrm{Pd}) - (\mathrm{Po} - \mathrm{Td}) = \mathrm{f} - \mathrm{a}$$

が得られるが，左辺の各項の意味から，

$$\mathrm{To} - \mathrm{Pd} = \mathrm{f}$$
$$\mathrm{Po} - \mathrm{Td} = \mathrm{a}$$

のように分けることができる．

上の式に$\mathrm{o} = \mathrm{r} - \mathrm{d}$を代入して整理すれば，

$$(\mathrm{T} + \mathrm{P})\mathrm{d} = \mathrm{Tr} - \mathrm{f}$$

が得られるので，まず，$\bar{\mathrm{M}}\mathrm{r} = \bar{\mathrm{p}}$から rを求め，次にこの式を解けば，o, dを求めることができる．

# レイティングの統計モデリング

ここでは，前章で述べたような確定的なレイティングモデルを統計モデリングにより確率論的なモデルに変換する．また，試合結果の生成メカニズムを想像することで，レイティングの統計モデルを直接的に得るケースなどを簡単に紹介する．

## 3.1. Eloレイティングの統計モデリング

対戦成績が，累積勝敗回数としてペア別に与えられる

[6] 各プレーヤーのレイティング値に同じ値を加減してもプレーヤー間のレイティングの差は変化しないので，レイティング差が得点差を近似するという条件だけでは，各プレーヤーのレイティング値は定まらないことが分かる．

[7] 2プレーヤーの組み合わせ（順序対）をプレーヤーペアと呼ぶ．

場合と，勝者と敗者を試合毎に指定する場合に分ける．

【プレーヤーペア別の対戦成績】

第$i$番目のプレーヤーペアを$(A,B)$とする（$i$はプレーヤーペア$(A,B)$を示すインデックス）．$A$と$B$の総試合数（対戦回数）$M[i]$が与えられているとする．$B$との対戦における$A$の勝率$q[i]$が分かれば，そのうちで$A$が勝った（$B$が負けた[8]）回数$Y[i]$は，$M[i]$と$q[i]$をパラメータとする二項分布に従う確率変数の実現値であると考えることができる．また，「2.1. Elo レイティングの概要」で述べたように，$A$の勝率$q[i]$は，$A$と$B$のレイティングの差にロジスティック関数を適用して求められる[9]

ただし，既述したように，プレーヤーのレイティングに定数を加えても，プレーヤーペアのレイティングの差は変わらないことから，スケールを決めなければユニークなレイティングが求められない．推定の不定性を避けるため，緩い制約を課す．ここでは，レイティングは平均0，標準偏差$\sigma_r$の正規分布に従うと考える[10]．

Stan ライクな記法を使うと，このモデルは以下のように記すことができる[11]．

■ EloModel-1 ■

$$q[i] = \text{inv\_logit}\big(r\big[P[i,1]\big] - r\big[P[i,2]\big]\big) \quad i = 1,\dots,I$$
$$Y[i] \sim \text{Binomial}(M[i], q[i]) \quad i = 1,\dots,I$$
$$r[n] \sim \text{Normal}(0, \sigma_r) \quad n = 1,\dots,N$$

ここで，$I$は対戦のあったプレーヤーペアの数，$N$はプレーヤーの数を示し，$i, n$は，それぞれプレーヤーペアとプレーヤーのインデックスである．inv_logitはロジット関数の逆関数（ロジスティック関数）である．$r$はプレーヤーのレイティング，$P$はプレーヤーペアを示し，$P[i,1], P[i,2]$は第$i$ペアの第 1 プレーヤー，第 2 プレーヤーである．また，Binomialは二項分布，Normal は正規分布を示す．なお，Normal の第 2 引数は分散ではなく標準偏差である．

【試合別の対戦成績】

試合$g$ごとに敗者$A$と勝者$B$を記した記録$LW$が与えられているとする．$B$との対戦における$A$の勝率$q[g]$が分かれば，$A$の勝敗はベルヌーイ分布に従う（が，実際の結果は負けだった）と考えることができる．勝率とレイティングの生成に関しては前と同じように考えると，このモデルは以下のようになる．

■ EloModel-2 ■

$$q[g] = \text{inv\_logit}\big(r\big[LW[g,1]\big] - r\big[LW[g,2]\big]\big) \quad g = 1,\dots,G$$
$$0 \sim \text{Bernoulli}(q[g]) \quad g = 1,\dots,G$$
$$r[n] \sim \text{Normal}(0, \sigma_r) \quad n = 1,\dots,N$$

ここで，$G$は全試合数を示し，$g$は，試合のインデックスである．$LW[g,1], LW[g,2]$は，それぞれ試合$g$における敗者と勝者である．また，Bernoulliはベルヌーイ分布を示し，入力データの形式から，実現値はすべて0である．

## 3.2. Masseyレイティングの統計モデリング

ここでは，試合毎の対戦成績が，2 プレーヤーの得点差として与えられる場合と，各プレーヤーの得点として与えられる場合を想定し，またレイティングの種類を変えた 2 種類のモデル（基本モデルと拡張モデル）を示す．

試合毎の得点差のデータからプレーヤーの総合的なレイティングを求めるモデルを基本モデルとする．各プレーヤーの試合毎の得点データから，総合的なレイティングと共に，攻撃，守備のレイティングも求めるモデルを拡張モデルと呼ぶ．

【基本モデル】

試合の得点差が，その試合で対戦したプレーヤーのレイティングの差で近似されるという Massey レイティングの基本的なアイデアに従い，「得点差は，レイティングの差を平均とする正規分布に従う[12]」と考えると，以下のようなモデルが得られる．

■ MessayModel-1 ■

$$Y[g] \sim \text{Normal}\big(r\big[P[g,1]\big] - r\big[P[g,2]\big], \sigma_Y\big) \quad g = 1,\dots,G$$
$$r[n] \sim \text{Normal}(0, \sigma_r) \quad n = 1,\dots,N$$

ここでの$P[g,1], P[g,2]$は，それぞれ試合$g$のプレーヤーペアの第 1 プレーヤーと第 2 プレーヤーである．また，$Y[g]$は，試合$g$の得点差（第 1 プレーヤーの得点と第 2 プレーヤーの得点の差）である．

---

[8] 議論の簡単化のため，引き分けはないものとする．
[9] その意味で，二項ロジスティック回帰の特殊なケースと類似している．
[10] $\sigma_r$自体の事前分布については，幅の広い一様分布など無情報事前分布を仮定する．従って，レイティングを個体差とする階層モデルである．
[11] $\sigma_r$のように，特に分布の指定をしていないパラメータについては，前述のように，無情報事前分布を仮定する（これ以降の統計モデルについても同様である）．
[12] 得点や得点差は整数値であることが多いので，近似的な仮定である．

【拡張モデル】

「2.2. Massey レイティングの概要」で述べたように，総合的なレイティングは，攻撃レイティングと守備レイティングとの和である（$\mathrm{r} = \mathrm{o} + \mathrm{d}$）．

また，自分の攻撃レイティングの累積から対戦相手の守備レイティングの和を差し引いたものを累積得点とみなした（$\mathrm{To} - \mathrm{Pd} = \mathrm{f}$）．このことから，各試合でも，自分の攻撃レイティングから相手の守備レイティングを引いたものは，近似的に得点と等しく，相手の攻撃レイティングから自分の守備レイティングを引いたものは，近似的に失点に等しいと考えるのは自然であろう．ここでは，得点，失点は，攻撃レイティングと守備レイティングの差を平均とする正規分布に従うと考える [13]．

■ MessayModel-2 ■

$$
\begin{aligned}
&r[n] = o[n] + d[n] && n = 1,\dots,N\\
&Y[g,1] \sim \mathrm{Normal}(o[P[g,1]] - d[P[g,2]], \sigma_Y) && g = 1,\dots,G\\
&Y[g,2] \sim \mathrm{Normal}(o[P[g,2]] - d[P[g,1]], \sigma_Y) && g = 1,\dots,G\\
&o[n] \sim \mathrm{Normal}(0, \sigma_o) && n = 1,\dots,N\\
&d[n] \sim \mathrm{Normal}(0, \sigma_d) && n = 1,\dots,N
\end{aligned}
$$

ここで，$o[n], d[n]$は，それぞれプレーヤー$n$の攻撃レイティングと守備レイティングである．$Y[g,1], Y[g,2]$は，それぞれ試合$g$における第 1，第 2 プレーヤーの得点である．

## 3.3. その他の統計モデリング

前の 2 節のように，オーソドックスなレイティング方法は，その計算の根拠が明らかであれば，適当な統計モデルに拡張することができるが，ここでは，試合結果の生成メカニズムを想像することで，レイティングの統計モデルを直接的に得るケースを 2 例示す．

【例 1】（[5]より [14]）

「3.1. Elo レイティングの統計モデリング」の【試合別の対戦成績】と同じく，試合ごとに敗者と勝者を記録したデータが与えられている状況を想定する．

試合の勝敗は，対戦者がその試合で発揮できた力（パフォーマンス）の大小で決まると考える．そして，各プレーヤーのパフォーマンスは，真の実力を平均とする正規分布に従うとする．また，その正規分布の標準偏差は，プレーヤーごとに異なり「勝負ムラ」と解釈する．

■ Example-1 ■

$$
\begin{aligned}
&pf[g,1] < pf[g,2] && g = 1,\dots,G\\
&pf[g,i] \sim \mathrm{Normal}\left(\mu[LW[g,i]], \sigma_{pf}[LW[g,i]]\right) && i = 1,2\\
&\mu[n] \sim \mathrm{Normal}(0, \sigma_\mu) && n = 1,\dots,N\\
&\sigma_{pf}[n] \sim \mathrm{Gamma}(10,10) && n = 1,\dots,N
\end{aligned}
$$

ここで，$LW[g,1], LW[g,2]$は，それぞれ試合$g$における敗者と勝者であり，$pf[g,1], pf[g,2]$は，敗者と勝者のパフォーマンスである．$\mu[n]$はプレーヤー$n$の実力を示し，平均0，標準偏差$\sigma_\mu$の正規分布に従うとしている．$\mu$をプレーヤーのレイティングと考えることができる．パフォーマンスの標準偏差$\sigma_{pf}[n]$はプレーヤー$n$の勝負ムラを示し，参考文献[5]と同じく，弱情報事前分布（weakly informative prior）として，形状母数（shape parameter）10，逆尺度母数（rate parameter）10のガンマ分布$\mathrm{Gamma}(10,10)$を仮定している．ガンマ分布は指数分布を一般化した確率分布で，確率変数は正の実数値をとる．$\mathrm{Gamma}(\alpha,\beta)$の平均は$\alpha/\beta$，標準偏差は$\sqrt{\alpha}/\beta$なので，このモデルでは，勝負ムラ$\sigma_{pf}[n]$の平均は1，標準偏差は$\sqrt{10}/10 \cong 0.316$である．

【例 2】（[3]より [15]）

これまで示したモデルと異なり，$N(\geq 2)$人ゲームにおけるレイティングの例である [16]．試合ごとに，$N$人の固定メンバーの得点を記録したデータが与えられている状況を想定する．

[3]では，4 人のメンバーによる麻雀を対象として，実力の比較を行っている．麻雀の各試合（半荘）は零和ゲームで，得点合計が0という制約がある．そこで，各試合の 4 人の得点を，合計して 1 となる正の実数に変換し，ディリクレ分布（Dirichlet [17]）を用いたモデル化を行っている． この変換には，4 人の元の得点を 300 で割った値に softmax 関数を適用するという方法を用いている．

■ Example-2 ■

$$
\begin{aligned}
&\mathbf{Point}[g] = \mathrm{softmax}(\mathbf{Y}[g] \cdot \mathrm{D}) && g = 1,\dots,G\\
&\mathbf{Point}[g] \sim \mathrm{Dirichlet}(\mathbf{janryoku}) && g = 1,\dots,G
\end{aligned}
$$

[13] 得点，失点は，非負数と考えるのが普通なので，正規分布の仮定は厳密には正しくないが，攻撃レイティングに対して，守備レイティングがかなり小さく推定されるようなら，問題ないであろう．もしそうでなければ，ガンマ分布や対数正規分布のような非負値をとらない分布を仮定する必要があるかも知れない．

[14] 文献の 10 章に示されているモデルの 1 つ（モデル式 10-4）である．ただし，表示を若干変更している．

[15] 文献の表示を若干変更している．

[16] 3 人以上のプレーヤーからなる試合用のレイティングの例としては，ゲーミングサービス Xbox Live でのマッチメイキング用にマイクロソフトが開発した TrueSkill というシステムがある[7]．

[17] ディリクレ分布に従う確率変数は，2 個以上の要素からなるベクトルで，要素は正の実数で合計すると 1 となる．

ここで，$\mathbf{Y}[g]$は，試合$g$における全プレーヤーの得点を要素とするベクトルであり，Dは定数（文献では$1/300$）である [18]．$\mathbf{Point}[g]$は変換後の得点ベクトルであり，**janryoku**ベクトルをパラメータとするディリクレ分布に従うと仮定している．

この方法は，元の得点をDによりスケール変換してソフトマックス関数を適用しているが，**janryoku**ベクトルの要素は，変数変換の影響を強く受ける．元の点数をより小さくするほど，実力（＝レイティング）を示すパラメータの推定値が大きくなるため，プレイヤー間の差も大きくなるが，比は逆に小さくなる．

$n(\geq 2)$人ゲームにおけるレイティングをモデル化には，例 1 を拡張するのも自然な方法の 1 つであろう．

## モデルの適用例

ここでは，実際のデータに対して，前章までに述べたレイティングのモデルを適用した例を示す．主に取り上げるのは，将棋棋士の勝敗データによる Elo レイティングと，ラグビーチームの得点データによる Massey レイティングである．

### 4.1. 将棋棋士の勝敗データ

「将棋棋士成績 DB」[12]というウェブページにある対局ごとの結果データを使用させていただいた [19]．

**Table 1　対局結果データ**

| loss | win | date |
|---|---|---|
| 森下卓 | 郷田真隆 | 2001/4/2 |
| 久保利明 | 三浦弘行 | 2001/4/2 |
| 北浜健介 | 松尾歩 | 2001/4/2 |
| … | … | … |
| 船江恒平 | 中村亮介 | 2019/11/15 |
| 糸谷哲郎 | 豊島将之 | 2019/11/15 |
| 広瀬章人 | 渡辺明 | 2019/11/17 |

2001 年 4 月 2 日から 2019 年 11 月 17 日までの 52996 対局のデータを取得し，そのうちで女流棋士や不戦勝敗等を除く現役棋士 168 名（12 月 2 日現在）の 34752 対局データを得た [20]．具体的には，対局ごとに敗者と勝者をまとめた Table 1 のような形式のデータである．統計モデルのパラメータ推定には，そのうち 2019 年 9 月末までのデータを用いた．

### 4.2. 将棋棋士のEloレイティング

標準的な従来の Elo レイティングについては, 例えば,「棋士別成績一覧」[13]や「将棋棋士レーティングランキング」[16]といったサイトで具体的な数値が日々求められている．

「棋士別成績一覧」では，2001 年 4 月時点で全員のレートを 1500 とし，また新規のプレーヤー（新四段）についても初期値を 1500 とするレイティングの更新式を用いて計算している．2.1 節の記述に合わせると，同サイトでは，

$$R = 1500 + 400\log_{10}(q/(1-q))$$

をレイティングととらえ，また，具体的な計算では，それに対応する更新式

$$R_A(t) = R_A(t-1) + K(s_{AB}(t) - P_{AB}(t-1))$$

を用いていることになる．ただし，

- $R_A(t)$：$t$での$A$のレイティング
- $K$：定数（同サイトでは 16 に設定）
- $P_{AB}(t)$：$A$の$B$に対する勝率の$t$での見積もり．

これは，$D_{AB}(t) = (R_A(t) - R_B(t))/400$として $P_{AB}(t) = 1/(1 + 10^{-D_{AB}(t)})$である．

「将棋棋士レーティングランキング」[16]では，2005 年 4 月 1 日をレイティングの開始時として全員のレートを 1500 としている．同サイトの説明によると，新規のプレーヤーのレイティングは，奨励会員（女流の場合はアマチュア）のレイティングを初期値にしており，当初の 30 対局のレイティング変動は通常より大きく設定しているとのことであるが詳細は不明である.「棋士別成績一覧」と比較すると，レイティングの数値は小さめ [21]だが，現役棋士のランキングには差異は少ない．

次に，統計モデリングによる Elo レイティングの例を示す．3.1 節の EloModel-2 の$LW$は，Table 1 のデータの 1，2 列に相当する入力データなので，Table 1 のデータに対して， EloModel-2 を自然な形で適用することがで

[18] Dによるスケールの変換を行わない場合（D = 1）だと，[3]のデータで試算すると，**janryoku**ベクトルの要素は 0.0139 といったオーダーの値になる．

[19] 同ウェブページ[12]によれば，サイトのデータは，日本将棋連盟による公式戦の結果と将棋年鑑に基づいているとのことである．

[20] レイティングの評価等のために，それ以外の対局データも若干用いている．

[21] レイティング 1500 が平均的プレーヤーを意味することになっているが，現役棋士のレイティングの平均値はそうなっておらず，インフレ（あるいはデフレ）と呼ばれる．各棋士の初対局時のレイティングを 1500 としても，例えば，引退時のレイティングがそれより低ければ，その棋士以外のレイティングの平均は上昇することになる．2001 年スタートの「棋士別成績一覧」[13]の方が 2005 年スタートの「将棋棋士レーティングランキング」[16]より数値が大きくなっていることから，経年的にインフレ傾向があるようである．

きる.

付録 a.2 に EloModel-2 の Stan コードの例を示す [22]. 実装には, R 言語(version 3.6.1), rstan(version 2.19.2), gcc（version 8.1.0）を用いた．なお，rstan のデフォルトの設定 [23]では，事後確率を求める sampling 関数の実行時に，treedepth の設定値に関する警告が生じたため，max_treedepth を 10 から 15 に変更して再実行した．

MCMC の収束診断として，トレースプロットの一部を Fig. 1（右半分）に例示する．上段と中段の図は，それぞれ藤井聡太七段 [24]のレイティング r[137]と渡辺明三冠 [25]のレイティング r[130]に対応している．下段の図は，レイティングの標準偏差 s_r に対応している．いずれのトレースプロットも 4 本のチェーンがランダムに入り混ざり，収束していそうだと判断できる．

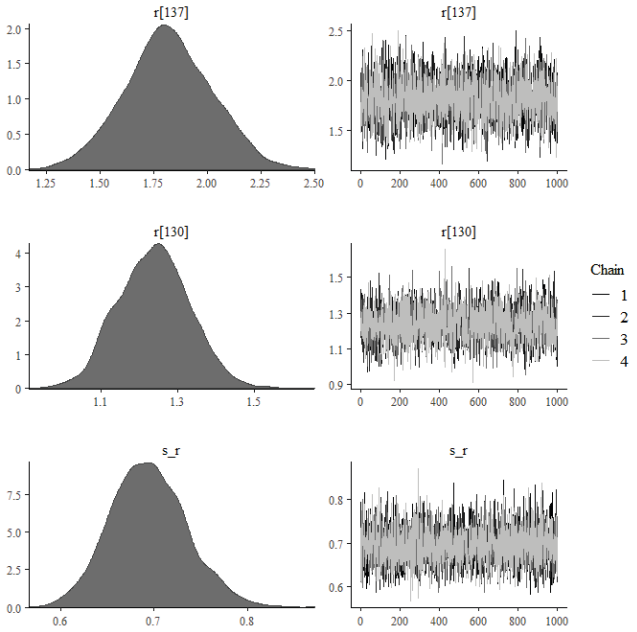


**Fig. 1　事後分布と trace plot（EloModel-2）**

図の左半分は，パラメータの事後分布を示している．また，それらの要約統計量の例を Table 2 に示す．

表の 2 列目から順に，事後平均（posterior mean）あるいは事後期待値（EAP: expected a posteriori），その標準誤差，標準偏差，2.5 パーセント点，事後中央値（MED: posterior median），97.5 パーセント点，収束性を示す指標$\hat{R}$が示されている．表は，ごく一部のパラメータのみ例示しているが，$\hat{R}$ の最大値は 1.003，従って，すべてのパラメータについて$\hat{R}$の値は 1.1 より小さく，MCMC は収束しているとみなすことができる [26].

**Table 2　要約統計量（EloModel-2）**

| para-meter | mean | se mean | sd | 2.5% | 50% | 97.5% | Rhat |
|---|---|---|---|---|---|---|---|
| r[137] | 1.819 | 0.003 | 0.204 | 1.418 | 1.815 | 2.216 | 0.999 |
| r[130] | 1.236 | 0.003 | 0.094 | 1.061 | 1.238 | 1.419 | 1.000 |
| s_r | 0.694 | 0.001 | 0.041 | 0.618 | 0.693 | 0.778 | 0.999 |

Fig. 2 は,「棋士別成績一覧」による通常の標準的な Elo レイティング（以下 EloRating と呼ぶ）の値分布（右図）と統計モデリングによるレイティング（以下 EloModel-2 と呼ぶ）の値分布（左図）を対比した（頻度分布付きの）バイオリンプロット（例えば[14]参照）である．

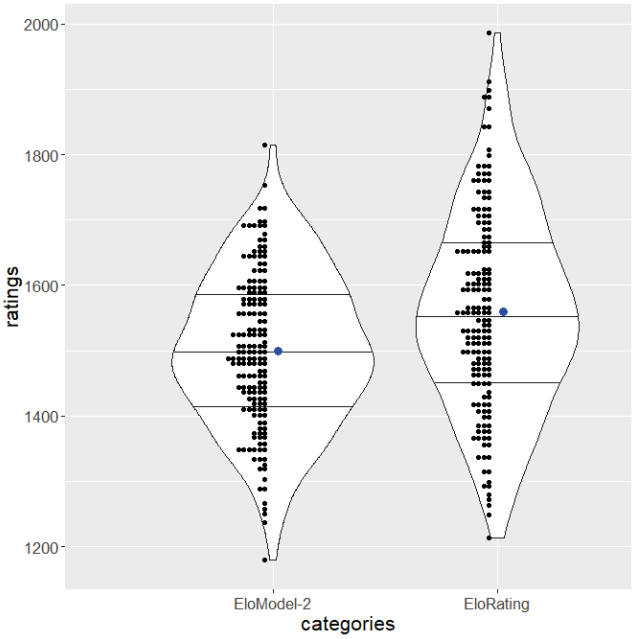


**Fig. 2　Elo レイティングの分布比較**

縦軸はレイティングの値であり，２つの図（バイオリン）の中の水平線は, それぞれの四分位数を示している．また，小さい点は個別のデータで，各カテゴリに１つずつあるやや大きい点は平均値を示している. この図より，

[22] 付録の Stan コードでは，敢えてベクトル化は用いていない．
[23] デフォルトでは，4 本のシミュレーション（chains=4）を 2000 回ずつ（iter=2000）実行し，その半分をバーンイン（ウォームアップ（warmup=1000））として捨てる等の設定となっている．
[24] 段位や称号は，執筆時点でのものである（以下も同様）．
[25] 棋王・王将・棋聖の三冠．
[26] チェイン数が 3 以上というのも収束の判定条件の 1 つに加えられることがあるが，本論文では，チェイン数はすべてデフォルトの 4 としているので満たされる．

EloModel-2 の方が EloRating よりも平均値が小さく，バラツキも少ないことが分かる．特に，EloModel-2 の平均値も中央値もほぼ 1500 なのに対して，EloRating では，いずれも 1550 より大きく，やはりインフレ傾向にあるといえる．

次に，トップクラスの棋士のレイティングとランキングをより具体的に対比する（Table 3）．表の 2, 3 列目は，EloRating による棋士のランキングとレイティング（2019 年 9 月末時点）を示し，4, 5 列目は， EloModel-2 によるランキングとレイティングを示す [27]. EloModel-2 のレイティングは，パラメータ事後分布の中央値を用いて計算している．例えば，ランキング 1 位，藤井七段のレイティング 1815 は，パラメータ r[137]の事後分布の中央値 1.815（Table 2 参照）により，

$$R = 1500 + 400/\log(10) \times 1.815 \approx 1815$$

と計算したものである．

それぞれのランキング 10 位までの棋士のうち両方に名が挙がっている棋士は 7 名であり，片方にだけ名が挙がっている棋士は 6 名である．この表は，この合わせて 13 名の棋士の順位とレイティングを示している．

**Table 3　棋士のレイティング比較**

| rank | player1 | EloRating | player2 | EloModel-2 |
|---|---|---|---|---|
| 1 | 渡辺明 | 1986 | 藤井聡太 | 1815 |
| 2 | 豊島将之 | 1912 | 羽生善治 | 1753 |
| 3 | 永瀬拓矢 | 1898 | 豊島将之 | 1721 |
| 4 | 広瀬章人 | 1891 | 渡辺明 | 1715 |
| 5 | 藤井聡太 | 1884 | 永瀬拓矢 | 1699 |
| 6 | 羽生善治 | 1871 | 千田翔太 | 1697 |
| 7 | 千田翔太 | 1846 | 菅井竜也 | 1695 |
| 8 | 木村一基 | 1840 | 近藤誠也 | 1692 |
| 9 | 菅井竜也 | 1807 | 大橋貴洸 | 1691 |
| 10 | 久保利明 | 1799 | 増田康宏 | 1688 |
| … | … | … | … | … |
| 14 | 増田康宏 | 1773 | 広瀬章人 | 1668 |
| 17 | … | … | 久保利明 | 1655 |
| 18 | 近藤誠也 | 1763 | … | … |
| 19 | … | … | 木村一基 | 1650 |
| 22 | 大橋貴洸 | 1740 | … | … |

この結果から，これらのランキングにはかなりの差異があることが分かる．EloRating で 1 位の渡辺 3 冠は EloModel-2 では 4 位で，EloModel-2 で 1，2 位の藤井七段と羽生九段は EloRating では 5，6 位である．10 位以降に名前のある棋士については，概ねさらに大きな順位差がある [28].

そこで，これらのレイティングについて，ごく大雑把ではあるが，2 通りの評価を試みた．1 つは，10 月以降のデータによる 13 人の対戦成績を各レイティングから予測し，実際の結果と対比することである．2 つ目の評価は，これまでの 13 人の対戦成績を各レイティングから推定し，実際の結果と対比することである．

【評価 1】

13 人の棋士による，10 月 3 日から 12 月 9 日までの全 21 対局を取り上げる（Table 4 参照）．表の 2, 3 列目は勝者と敗者の名前，4, 5 列目は，EloRating による勝者と敗者のレイティング, そして 6, 7 列目は, EloModel-2 による勝者と敗者のレイティングを示す.

**Table 4　10 月以降の対局データ**

| date | win | loss | r11 | r12 | r21 | r22 |
|---|---|---|---|---|---|---|
| 10/3 | 渡辺明 | 羽生善治 | 1986 | 1871 | 1715 | 1753 |
| … | … | … | … | … | … | … |
| 12/9 | 羽生善治 | 千田翔太 | 1871 | 1846 | 1753 | 1697 |

このデータから，各対局の勝者の勝率をそれぞれのレイティングに基づいて予測する．勝者$A$の敗者$B$に対する勝率は，例えば，10 月 3 日の r11, r12 を使うと，$D_{AB} = (1986 - 1871)/400 = 0.2875$なので,

$$q_{AB} = 1/(1 + 10^{-D_{AB}}) \approx 0.66$$

と予測される．EloModel-2 の r21, r22 を使うと, 約 0.45 と予測される．他の対局についても同様に計算し，その平均値を求めると，EloRating によるレイティングでは約 0.55，EloModel-2 によるレイティングでは約 0.51 となった．

また，各対局において，レイティングの大きい方が勝つと予測する場合，正解率を比較すると，EloRating では約 0.81，EloModel-2 では約 0.67 となった．いずれの場合も，このデータについては，EloRating の方が EloModel-2 よりも予測性に優れている．

【評価 2】

まず，当該の棋士 13 人のこれまでの対戦成績を「将棋棋士成績 DB」より取得し整理した [29]（Table 5）．この表は，表側の棋士の勝数（表頭の棋士の負け数）を示して

[27] パラメータの推定は，2001 年 4 月初めから 2019 年 9 月末のデータによる.

[28] 3.3 節の Example-1 のモデルに基づいたレイティングも試みた．EloRating よりも EloModel-2 にやや近いランキングとなったが，6 位が佐藤天彦九段，10 位が佐藤康光九段となっていた．このモデルは，非常に実行時間がかかり，警告も出たので，結果の詳細は省略する.

[29] 既述の通り，不戦勝敗はカウントしていない.

いる．例えば，渡辺三冠 vs 豊島名人は 24 戦して，渡辺三冠の 15 勝 9 敗であることを意味している．

この勝敗数の分布と，レイティングから見積もられる勝敗数の分布を確率分布として比較する．すなわち，Table 5 の各度数を全度数で除して同時相対度数を求め，レイティングと対戦数から見積もられる分布の確率関数とを比較する．例えば，1 行 2 列目の同時相対度数は $15/679$である．

**Table 5　対戦成績**

| | 渡辺 | 豊島 | 永瀬 | … | 計 | 勝率 |
|---|---|---|---|---|---|---|
| 渡辺 | | 15 | 10 | … | 124 | 0.577 |
| 豊島 | 9 | | 2 | … | 79 | 0.503 |
| 永瀬 | 3 | 1 | | … | 30 | 0.556 |
| … | … | … | … | … | … | … |
| 計 | 91 | 78 | 24 | … | 679 | |

確率関数値については，以下のように求められる：【評価 1】と同様の計算で，Table 3 のレイティングから，対戦ペア毎に勝率を見積もることができる．例えば，渡辺三冠 vs.豊島名人での渡辺三冠の勝率は，EloRating だと約 0.60 である．対戦数が 24 なので，14.5 程度の勝数が期待され，1 行 2 列目の確率関数値を$14.5/679$とする．同様にして，EloModel-2 についても確率関数を見積もることができる．

どちらの確率関数がデータの相対度数に近いだろうか．ここでは，２つの分布の差を計る指標である KL ダイバージェンス（Kullback-Leibler divergence），ヒストグラム間類似度（Histogram Intersection），そして，$\chi^2$値を比較してみる．ただし，２つの離散確率分布$p, q$について，

- KL ダイバージェンス：$D_{KL}(p \parallel q) = \sum_i p_i \log(p_i/q_i)$
- ヒストグラム間類似度：$D_{HI}(p,q) = \sum_i \min(p_i, q_i)$
- $\chi^2$値：$\chi^2(p \parallel q) = \sum_i (p_i - q_i)^2/q_i$

とする．

**Table 6　適合度比較**

| index | EloRating | EloModel-2 |
|---|---|---|
| KL | 0.00887 | 0.00338 |
| HI | 0.869 | 0.902 |
| Chi sq. | 64.2 | 48.5 |

KL ダイバージェンスと$\chi^2$値は非負の値をとり，$p = q$の時，値が 0 となる．ヒストグラム間類似度は，文字通りには，2 つのヒストグラムの交わりの意であるが，2 つの確率関数の（面積 1 の）ヒストグラムの交わりの面積を示し，$[0,1]$の値をとる．$p = q$の時，値が 1 となる．

データの相対度数を$q$とし，EloRating（あるいは EloModel-2）のレイティングから見積もられる確率分布を$p$として上記の指標を計算した結果を Table 6 に示す．

KL ダイバージェンスと$\chi^2$値は，EloRating よりも EloModel-2 の値が小さく，ヒストグラム間類似度は EloRating よりも EloModel-2 の値が大きい．いずれの指標についても，このデータについては，EloModel-2 の方が EloRating よりもこれまでの対戦成績との適合性に優れている．

## 4.3. ラグビーの勝敗データ

ラグビーについては，Statsguru[18]というウェブページにある試合結果のデータを使用させていただいた．2009 年 1 月 31 日から 2019 年 12 月 1 日までの 2906 試合のデータを取得し整形した．具体的には，試合ごとに対戦チームと得点をまとめた Table 7 のような形式のデータである．team1, team2 は対戦チーム名，score1, score2 はそれぞれ team1, team2 の得点，date は試合日である [30]．

なお，パラメータの推定は，ラグビーワールドカップ 2019 日本大会の開催日（2019 年 9 月 20 日）以前のデータ（172 チーム [31]，2821 試合）を使って行った．

**Table 7　対戦データ [32]**

| team1 | team2 | score1 | score2 | date |
|---|---|---|---|---|
| Monaco | Azerbaijan | 38 | 12 | 2009/1/31 |
| Georgia | Germany | 38 | 5 | 2009/2/7 |
| England | Italy | 36 | 11 | 2009/2/7 |
| … | … | … | … | … |
| USA | Canada | 20 | 15 | 2019/9/7 |
| Dominican | Virgin | 15 | 0 | 2019/9/14 |
| Zimbabwe | Zambia | 41 | 5 | 2019/9/14 |
| … | … | … | … | … |
| Madagascar | Nigeria | 63 | 3 | 2019/12/1 |

## 4.4. ラグビー・チームのMasseyレイティング

まず，公式的な記録としては，ラグビーユニオンの国際競技連盟である World Rugby のウェブページに，各種

[30] 引き分けでない試合では，team1 は勝ちチーム，team2 は負けチームである．
[31] ナショナルチームだけでなく，JAPAN XV などの代表チーム，French Barbarians などの選抜チーム，Classic All Blacks などの特別チームを含む．
[32] 参考サイトのデータは，必ずしも国名が統一されていないので，注意が必要である（例えば，Ivory Coast と Cote d'Iv など）．表の Virgin とは British Virgin Islands の意である（ちなみに，US Virgin Islands は US Virgin とした）．

ラグビーチームのランキングとレイティングのデータが掲載されている[20]．ラグビーでのレイティングはポイントと呼ばれ，ポイントは試合ごとに更新される．Elo レイティングと同じく，番狂わせの場合（ポイントの少ないチームがポイントの多いチームに勝った時）には，両チームともにポイントの変動が大きくなる．また，試合における大きな得点差やホームチームの優位性（ホームアドバンテージ），大会の重要性（ワールドカップ）なども考慮した独自の計算方法によるレイティングである（例えば，[19]に解説がある）．

次に，2.2 節で示した標準的，確定的な Massey レイティングについては，付録 a.5 のような関数で計算できる．既述の通り，総合的なレイティングに加え，攻撃と守備のレイティングも求めることができるのが Massey レイティングの特長であるが，総合的なレイティングだけでも，World Rugby のポイントシステムに対して優位な点がある．それは，2 チームのレイティングの差が意味を持つ（得点差を近似する）ということである．

最後に，統計モデリングによる Massey レイティングの例を示す．3.2 節の MasseyModel-2 の$P$と$Y$は，Table 7 のデータの 1, 2 列と 3, 4 列にそれぞれ相当するので，Table 7 のデータに対して，MasseyModel-2 を自然な形で適用することができる．付録 a.4 に MasseyModel-2 の Stan コードの例を示す．なお，rstan のデフォルトの設定では，事後確率を求める sampling 関数の実行時に，iteration に関する警告が生じたため，iter の値を 2000 から 20000 に変更して再実行した．

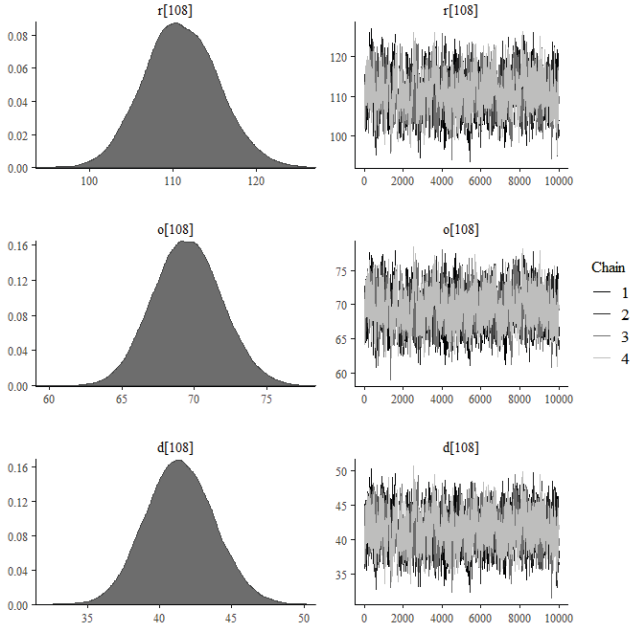


**Fig. 3　事後分布と trace plot（MasseyModel-2）**

MCMC の収束診断として，トレースプロットの一部を Fig. 3（右半分）に例示する．上，中，下段の図は，それぞれ，ニュージーランドの総合的なレイティング r[108]，攻撃レイティング o[108]，そして守備レイティング d[108]に対応している．レイティングの標準偏差（s_o, s_d）や得点の標準偏差（s_Y）といったパラメータについての図は省略する．いずれのトレースプロットも 4 本のチェーンがランダムに入り混ざり，収束していそうだと判断できる．

図の左半分は，パラメータの事後分布であり，それらの要約統計量の例を Table 8 に示す．表は，ごく一部のパラメータのみ例示しているが，$\hat{R}$ の最大値は 1.004，従ってすべてのパラメータについて$\hat{R}$ の値は 1.1 より小さく，MCMC は収束しているとみなすことができる．

**Table 8　要約統計量（MasseyModel-2）**

| para-meter | mean | se mean | sd | 2.5% | 50% | 97.5% | Rhat |
|---|---|---|---|---|---|---|---|
| r[108] | 111.0 | 0.1768 | 4.45 | 102.5 | 110.9 | 119.9 | 1.003 |
| o[108] | 69.5 | 0.0886 | 2.37 | 65.0 | 69.5 | 74.3 | 1.003 |
| d[108] | 41.4 | 0.0883 | 2.35 | 36.9 | 41.4 | 46.1 | 1.003 |
| s_o | 26.0 | 0.0408 | 1.74 | 22.9 | 26.0 | 29.7 | 1.002 |
| s_d | 27.3 | 0.0353 | 1.80 | 24.0 | 27.2 | 31.0 | 1.001 |
| s_Y | 12.7 | 0.0006 | 0.12 | 12.4 | 12.7 | 12.9 | 1.000 |

次に，World Rugby[20]による公式的なレイティング（以下公式レイティングと呼ぶ），およびそれによるランキングと，2.2 節で示した標準的な計算による Massey レイティング（以下 MasseyRating と呼ぶ）の結果，そして，3.2 節の統計モデリング（以下 MasseyModel-2 と呼ぶ）による結果とを一部対比する（Table 9）．

ここでのランキングは，ナショナルチームのみを対象とし，そのうち，各レイティングの上位 23 チームを順位づけたものである．この 23 チームは，どのレイティングでも上位 23 位までのチームとなっており，また，ラグビーワールドカップ 2019 日本大会出場の 20 チームを全て含んでいる．また，いわゆる Tier 1 と Tier 2 のチーム[11]だけからなっている．

表の 2, 3 列目は，公式レイティングによるチームのランキングとレイティング（2019 年 9 月 19 日現在）を示し，4, 5 列目は MessayRating によるランキングとレイティング，6, 7 列目は MasseyModel-2 によるランキングとレイティングを示す．なお，MasseyModel-2 のレイティングは，パラメータ事後分布の中央値を用いている．

MessayRating の値は，MasseyModel-2 のレイティング値より 10～13 程度高いだけで，それらによるランキングは非常によく似ている．また，公式レイティングでは，アイルランドのランクが，他のレイティングとは大

きく異なっている（ワールドカップ終了時点では 5 位であり，タイミングに左右されやすい弱点が出ているようである）．しかし，ベスト 10 までのランキングは比較的似通っており，概ね Tier 1 のチームが占めている．ただし，10 位前後では，Tier 1 と Tier 2 のチーム（イタリア，アルゼンチン，日本，フィジー）が混在している．

**Table 9　ラグビーのレイティング比較** [33]

| rank | team1 | r1 | team2 | r2 | team3 | r3 |
|---|---|---|---|---|---|---|
| 1 | Ireland | 89 | NZL | 124 | NZL | 111 |
| 2 | NZL | 89 | England | 113 | England | 100 |
| 3 | England | 88 | RSA | 113 | RSA | 100 |
| 4 | RSA | 87 | Australia | 110 | Australia | 98 |
| 5 | Wales | 87 | Ireland | 109 | Ireland | 96 |
| 6 | Australia | 84 | Wales | 107 | Wales | 94 |
| 7 | Scotland | 81 | France | 105 | France | 92 |
| 8 | France | 80 | Argentina | 100 | Argentina | 88 |
| 9 | Fiji | 77 | Scotland | 100 | Scotland | 87 |
| 10 | Japan | 77 | Fiji | 88 | Japan | 77 |
| 11 | Argentina | 76 | Italy | 87 | Fiji | 76 |
| 12 | Georgia | 73 | Japan | 87 | Italy | 75 |
| 13 | USA | 72 | Samoa | 85 | Samoa | 73 |
| 14 | Italy | 72 | Tonga | 82 | Tonga | 70 |
| 15 | Tonga | 71 | Georgia | 80 | Georgia | 69 |
| 16 | Samoa | 69 | USA | 77 | USA | 66 |
| 17 | Spain | 68 | Romania | 74 | Romania | 64 |
| 18 | Romania | 67 | Canada | 74 | Canada | 63 |
| 19 | Uruguay | 65 | Russia | 62 | Uruguay | 52 |
| 20 | Russia | 65 | Uruguay | 62 | Russia | 52 |
| 21 | Portugal | 61 | Spain | 62 | Spain | 51 |
| 22 | Canada | 61 | Namibia | 60 | Namibia | 50 |
| 23 | Namibia | 61 | Portugal | 58 | Portugal | 48 |

MessayRating と MasseyModel-2 については，攻撃，守備のレイティングも求められるので，比較してみる（Table 10）．oは攻撃レイティング，dは守備レイティングで，ok, dkは，それらによる順位である．2 行にわたる数値は，上が MessayRating によるもの，下が MasseyModel-2 によるものである．

**Table 10　攻撃，守備レイティングの比較** [34]

| No. | team | o | ok | d | dk |
|---|---|---|---|---|---|
| 1 | New Zealand | 76 | 1 | 48 | 1 |
| | | 70 | 1 | 41 | 1 |
| 2 | England | 67 | 3 | 46 | 2 |
| | | 61 | 3 | 40 | 2 |
| 3 | South Africa | 68 | 2 | 45 | 3 |
| | | 61 | 2 | 39 | 3 |
| 4 | Australia | 67 | 4 | 43 | 6 |
| | | 61 | 4 | 37 | 6 |
| 5 | Ireland | 64 | 5 | 44 | 4 |
| | | 58 | 5 | 38 | 4 |
| 6 | Wales | 63 | 7 | 44 | 5 |
| | | 57 | 7 | 38 | 5 |
| 7 | France | 62 | 8 | 43 | 7 |
| | | 56 | 8 | 36 | 7 |
| 8 | Argentina | 64 | 6 | 36 | 9 |
| | | 58 | 6 | 30 | 9 |
| 9 | Scotland | 60 | 10 | 40 | 8 |
| | | 54 | 10 | 34 | 8 |
| 10 | Fiji | 57 | 11 | 31 | 13 |
| | | 51 | 11 | 25 | 13 |
| 11 | Italy | 56 | 12 | 32 | 11 |
| | | 50 | 12 | 26 | 12 |
| 12 | Japan | 60 | 9 | 27 | 16 |
| | | 55 | 9 | 22 | 16 |
| 13 | Samoa | 53 | 13 | 32 | 10 |
| | | 47 | 13 | 26 | 10 |
| 14 | Tonga | 52 | 14 | 30 | 14 |
| | | 46 | 14 | 24 | 14 |
| 15 | Georgia | 48 | 17 | 31 | 12 |
| | | 43 | 17 | 26 | 11 |
| 16 | USA | 51 | 15 | 26 | 17 |
| | | 46 | 15 | 21 | 17 |
| 17 | Romania | 46 | 19 | 28 | 15 |
| | | 41 | 19 | 23 | 15 |
| 18 | Canada | 50 | 16 | 24 | 18 |
| | | 44 | 16 | 19 | 18 |
| 19 | Russia | 43 | 21 | 19 | 20 |
| | | 38 | 21 | 14 | 21 |
| 20 | Uruguay | 44 | 20 | 18 | 22 |
| | | 40 | 20 | 13 | 22 |
| 21 | Spain | 41 | 22 | 20 | 19 |
| | | 36 | 22 | 15 | 19 |
| 22 | Namibia | 47 | 18 | 12 | 23 |
| | | 43 | 18 | 8 | 23 |
| 23 | Portugal | 40 | 23 | 19 | 21 |
| | | 35 | 23 | 14 | 20 |

[33] すべてのレイティングの数値は整数に丸めている．チーム名の NZL, RSA はそれぞれ，New Zealand と South Africa を示す略号である．
[34] すべてのレイティングの数値は整数に丸めている．

MessayRating によるレイティングと MasseyModel-2 によるレイティングとの差は 5〜6 と，ほぼ一定である．Massey モデルは，元々得点差とレイティング差を等置するものであることからすると，MessayRating と MasseyModel-2 によるレイティングは，ほぼ等価であるといえる [35].

最後に，ラグビーワールドカップ 2019 日本大会の各対戦成績のデータを用いて，公式レイティングと MasseyModel-2 のレイティングのごく大雑把な評価を 2 通り試みる．1 つ目の評価は，各対戦の勝敗をそれぞれのレイティングから予測し，実際の結果と対比することである．2 つ目は，MessayRating と MasseyModel-2 のレイティングから各対戦の得点差等を予測し，実際の結果と対比することである．

【評価 1】

ワールドカップの 45 試合（Table 11）について，勝敗の予測を行う．ここでは，対戦チームのうち，（総合）レイティング値の大きい方が勝利するという単純な予測法を用いた．

例えば，9 月 20 日の日本 vs.ロシア戦だと，公式レイティング（Table 9 の r1）では 77：65，MasseyModel-2（Table 9 の r3）では 77：52 でいずれも日本の勝利予想となり，実際もそうなっている．11 月 2 日の南アフリカ vs.イングランド戦だと，公式レイティングでは 87：88，MasseyModel-2 では 100：100（より正確には，100.17：100.37）でいずれもイングランドの勝利予想となり，実際と異なっている．

全 45 試合の正解率は，公式レイティングが$38/45 = 0.844$，MasseyModel-2 が$39/45 = 0.867$でほぼ互角である．なお，MasseyRating による各試合の勝敗予測は MasseyModel-2 のそれと全く同じであり，よって正解率も同じである．

**Table 11　ワールドカップの試合**

| team1 | team2 | score1 | score2 | date |
|---|---|---|---|---|
| Japan | Russia | 30 | 10 | 2019/9/20 |
| … | … | … | … | … |
| South Africa | England | 32 | 12 | 2019/11/2 |

【評価 2】

MessayRating と MasseyModel-2 のレイティングから各対戦の得点，得点差を予測し，実際の結果と対比する．

9 月 20 日の日本 vs.ロシア戦だと，日本の攻撃，守備の MessayRating（と MasseyModel-2）のレイティング値は，それぞれ 60，27（55，22）で，ロシアが 43，19（38, 14）なので，日本の得点は$60 - 19 = 41$（$55 - 14 = 41$），ロシアの得点は$43 - 27 = 16$（$38 - 22 = 16$）と予測され，得点差の予測はいずれのレイティングでも25点となる．実際には，30 対 10 で得点差は 20 である．

11 月 2 日の南アフリカ vs.イングランド戦だと，南アフリカの攻撃，守備レイティングがそれぞれ 68, 45（61, 39）で，イングランドが 67，46（61，40）なので，南アフリカの得点は$68 - 46 = 22$（$61 - 40 = 21$），イングランドの得点は$67 - 45 = 22$（$61 - 39 = 22$）と予測され，得点差の予測は 0（$-1$）点となる．なお，MessayRating のより正確な数値 [36]を用いると，南アフリカの得点は 21.8，イングランドの得点は 22.0 と予測され，得点差は$-0.2$なので，整数に丸めると（$-1$ではなく）0 となる．このように，MasseyRating による各試合の得点，得点差の予測と MasseyModel-2 による得点，得点差の予測とはほぼ同じである．

Fig. 4 は，実際の得点と MasseyModel-2 による得点の予測値の分布を対比した（頻度分布付きの）バイオリンプロットである．横軸の a と p は，実際の得点と予測値のカテゴリをそれぞれ示しており，縦軸は点数である．この図より，実際の得点と MasseyModel-2 による得点の予測値は，平均や中央値はほぼ等しいが，予測値の方がバラツキが少ないことが分かる．実際，得点の標準偏差は約 16.6 なのに対して，予測値の標準偏差は約 13.2 である．

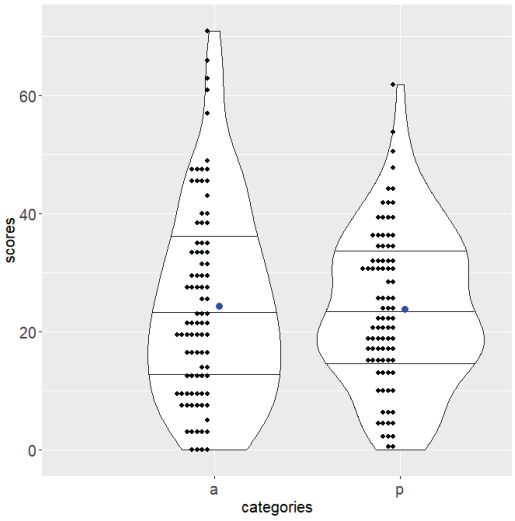


**Fig. 4　得点の予測と実際**

勝敗別に，実際の得点と得点の予測値を対比すると，Fig. 5 のような図が得られる．

[35] MasseyModel-2 ではパラメータのベイズ信頼区間等が得られるので，例えば，レイティング差から得点差を見積もる場合でも，点推定ではなく区間推定ができるなど，MessayRating よりも深い分析が可能になる．
[36] 小数第一位までの数値では，南アフリカの攻撃，守備レイティングが 67.7，45.1 で，イングランド 67.1，45.9 である．

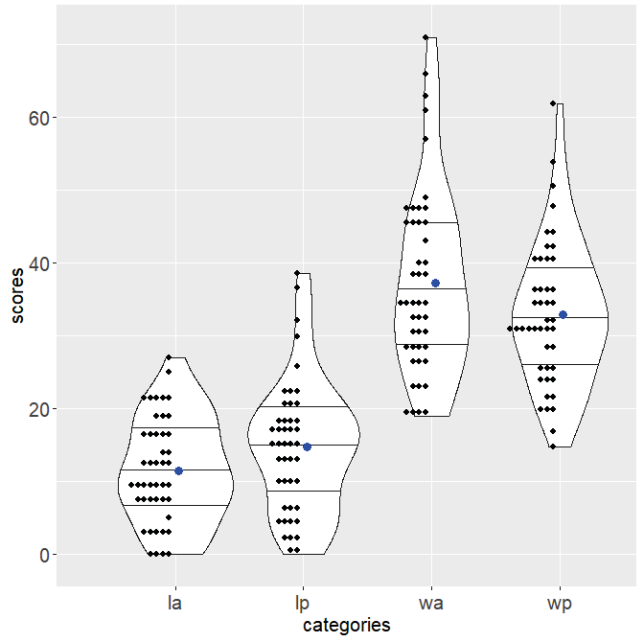


**Fig. 5　得点の予測と実際（勝敗別）**

ここで，横軸の la と lp は，敗れたチームの実際の得点と予測値のカテゴリをそれぞれ示し，wa と wp は，勝ったチームの実際の得点と予測値のカテゴリをそれぞれ示している．この図より，MasseyModel-2 による得点の予測は，得点の低い負けチームについては高めの予測，得点の高い勝ちチームについては低めの予測となっていることが分かる．

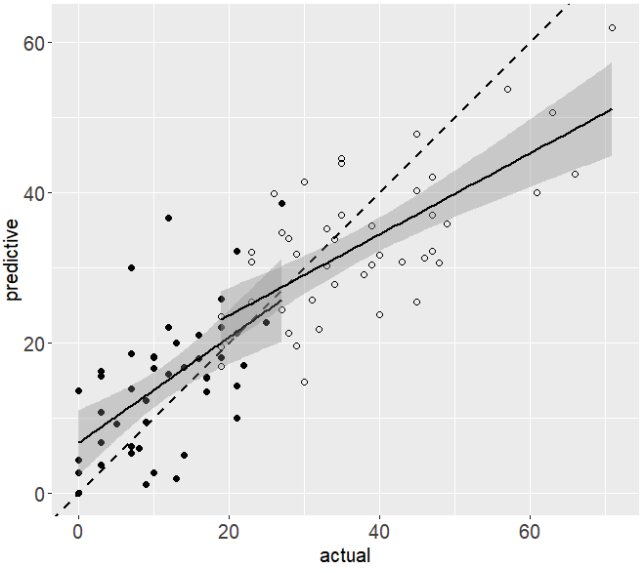


**Fig. 6　予測と実際の関係**

実際の得点と得点の予測値の関係を散布図で描くと Fig. 6 のようになる．横軸が実際の得点，縦軸が予測値であり，白丸は勝ちチーム，黒丸は負けチームのデータである．破線は原点を通る45°の直線で，２つの実線は，勝ちチーム，負けチームの回帰直線を示す．この図からも，MasseyModel-2 による得点の予測は，得点の低い負けチームについては高めの予測，得点の高い勝ちチームについては低めの予測となっていることが分かる．なお，実際と予測の相関係数は約 0.83 である．

## おわりに

本研究では，伝統的，標準的なレイティング手法を，ベイズ統計モデリングにより確率・統計的なモデルに拡張し，現実のデータに適用することを試みた．

第 2 章では，標準的なレイティング方法の例として，Elo レイティングと Massey レイティングを取り上げ，それぞれの数理的な概要を述べた．Elo レイティングの値は，平均的プレイヤーに対する勝率の対数オッズであり，勝敗比の推移性の仮定の下では，任意の 2 プレーヤーのレイティングの差から各プレーヤーの勝率が定まる．Massey レイティングは，試合の得点差がプレーヤーのレイティングの差で（最小二乗）近似されるようにレイティングを定めるというものである．総合的なレイティングに加え，攻撃，守備のレイティングが得られるという特長がある．

第 3 章では，Elo レイティングと Massey レイティングをベイズ統計モデル化した．Elo レイティングについては，対戦結果のデータのタイプに応じて 2 種類のモデルを求めた．すなわち，プレーヤーの対ごとに勝敗回数を集計したデータを入力とする EloModel-1 と，試合毎に勝者と敗者を特定したデータを入力とする EloModel-2 である．これらは，それぞれ二項分布とベルヌーイ分布によるロジスティック回帰の階層モデルの一種であり，標準的な Elo レイティングのベイズ統計モデル化になる．Massey レイティングについては，各プレーヤーの試合毎の得点データから，総合的なレイティングを得る基本モデル（MessayModel-1）と，攻撃，守備のレイティングも得られる拡張モデル（MessayModel-2）を求めた．これらは，標準的な Massey レイティングのベイズ統計モデル化であり，正規分布を仮定した線形の階層モデルである．

第 4 章では，3 章で求めた統計モデル（EloModel-2 と MasseyModel-2）を現実のデータ（将棋棋士の勝敗データとラグビーナショナルチームの得点データ）に適用し，それぞれのレイティングとランキングを求め，標準的な Elo レイティング，Massey レイティングの結果や公式的なレイティング記録と比較した．

将棋のデータでは，ランキングサイトの標準的な Elo レイティング（EloRating）と EloModel-2 のレイティングを比較したが，EloRating は平均値についてインフレ

気味であり，バラツキも EloModel-2 より大きかった．レイティングによる勝率に基づく勝敗予測では，EloRating の方が EloModel-2 よりも正解率が高かった．2001 年からの対局データによる勝敗数の分布とレイティングから見積もられる分布の比較では，逆に，EloModel-2 の方が EloRating よりも適合性に優れていた．これらは，更新式による EloRating のリアルタイムなアップデートが機能している（モデルはオーバーフィッティングしている）とも解せるが，データ数が少ない（21 対局）ことから，より詳細な評価が必要であり，今後の課題である．

ラグビーのデータでは，公式的なレイティングと標準的な Massey レイティング（MessayRating），そして，統計モデリング MasseyModel-2 のレイティングを比較した．総合的なレイティングについても攻撃・守備のレイティングでも，MessayRating と MasseyModel-2 とは，差がほぼ一定となっており，それらによるランキングは非常に似ている．ワールドカップ日本大会（45 試合）の勝敗予測では，３つのレーティング法の正解率はほぼ同じであった．MessayRating と MasseyModel-2 のレイティングは，各対戦の得点，得点差の予測でもほぼ同じ結果となった．また，実際の得点（失点）と MasseyModel-2 での予測値を比較すると，得点の低い負けチームについては高めの予測，得点の高い勝ちチームについては低めの予測となっているが，実際と予測の得点間の相関はかなり高い．

標準的なレイティングの問題点として，本文で以下の事項を述べた．

- 初期値から落ち着くまでに適正な評価が難しい，あるいは，収束までに時間がかかる．
- 評価のタイミングに左右されやすい．
- レイティングの更新のためのパラメータの適切な設定が困難である．
- インフレしやすい．

このうち，最初の 2 つの項目に対しては，リアルタイムにデータ処理するならば，程度の多少はあれ，どのような手法でも生じ得ることであり，統計モデリングについても同様であるが，ベイズ信頼区間が求められるので問題を緩和できると考えられる．後の 2 つについては，提案したような統計モデリングでは，Elo レイティングのような更新式による計算ではないため，そのようなパラメータ設定の問題はないし，インフレもしない（ただし，レイティングの更新は，簡単な計算ではできない）．

本論文では，スタティックなモデルしか扱わなかったが，時系列分析的な取り扱いを含めれば，統計モデリングでの予測性を高められそうである．伝統的なレイティング手法をこうしたアプローチでさらに拡張することは，より詳細な評価と共に，今後の課題である．

## 付録

### a.1. STANコード例（EloModel-1）

```
data {
  int N;
  int I;
  int<lower=0> P[I, 2];
  int<lower=0> M[I];
  int<lower=0> Y[I];
}

parameters {
  real r[N];
  real<lower=0> s_r;
}

transformed parameters {
  real<lower=0, upper=1> q[I];
  for (i in 1:I)
    q[i] = inv_logit(r[P[i, 1]] - r[P[i, 2]]);
}

model {
  for (i in 1:I)
    Y[i] ~ binomial(M[i], q[i]);

  for (n in 1:N)
    r[n] ~ normal(0, s_r);
}
```

### a.2. STANコード例（EloModel-2）

```
data {
  int N;
  int G;
  int<lower=1, upper=N> LW[G, 2];
}

parameters {
  real r[N];
  real<lower=0> s_r;
}

transformed parameters {
  real<lower=0, upper=1> q[G];
  for (g in 1:G)
    q[g] = inv_logit(r[LW[g, 1]] - r[LW[g, 2]]);
}

model {
  for (g in 1:G)
    0 ~ bernoulli(q[g]);

  for (n in 1:N)
    r[n] ~ normal(0, s_r);
}
```

### a.3. STANコード例（MasseyModel-1）

```
data {
   int N;
   int G;
   int<lower=1, upper=N> P[G,2];
   real Y[G];
}

parameters {
   real r[N];
   real<lower=0> s_r;
   real<lower=0> s_Y;
}

model {
   for (g in 1:G)
         Y[g] ~ normal(r[P[g,1]]-r[P[g,2]], s_Y);
   for (n in 1:N)
         r[n] ~ normal(0, s_r);
}
```

### a.4. STANコード例（MasseyModel-2）

```
Messey <- function(P,f,a){
   n <- nrow(P)
   T <- diag(rowSums(P))
   Ma <- T-P
   Ma[n,] <- rep(1,n)
   pa <- f-a
   pa[n] <- 0
   r <- solve(Ma, pa)
   rk <- r2rk(r)
   d <- solve(T+P, T%*%r-f)
   dk <- r2rk(d)
   o <- r-d
   ok <- r2rk(o)
   res <- list(r=r, rk=rk, o=o, ok=ok, d=d, dk=dk)
   return(res)
}

r2rk <- function(r)
   (1:length(r))[order(order(r, decreasing=TRUE))]
```

### a.5. Rコード例（自作関数：Messey）

```
data {
   int N;
   int G;
   int<lower=1, upper=N> P[G, 2];
   real Y[G, 2];
}

parameters {
   real o[N];
   real d[N];
   real<lower=0> s_o;
   real<lower=0> s_d;
   real<lower=0> s_Y;
}

transformed parameters {
   real r[N];
   for (n in 1:N)
         r[n] = o[n]+d[n];
}

model {
   for (g in 1:G) {
         Y[g, 1] ~ normal(o[P[g,1]]-d[P[g,2]], s_Y);
         Y[g, 2] ~ normal(o[P[g,2]]-d[P[g,1]], s_Y);
   }
   for (n in 1:N) {
         o[n] ~ normal(0, s_o);
         d[n] ~ normal(0, s_d);
   }
}
```

ただし，関数の引数の意味は以下の通りである：

- P : 対角要素が 0 の対称行列で，$(i,j)$要素はチーム$i$とチーム$j$の対戦数を示す.
- f : チーム数を要素数とするベクトルで，その第$i$要素はチーム$i$の累積得点を示す.
- a : チーム数を要素数とするベクトルで，その第$i$要素はチーム$i$の累積失点を示す.

また，r2rk はレイティングのベクトル r からそのランキング（順位）を求める関数である.

## 参考文献


[1] 久保拓弥，データ解析のための統計モデリング入門，岩波書店，2012.

[2] 小中英嗣，“バレーボール各国代表チームのレーティング手法の提案および結果予測・大会形式評価への応用，”統計数理, Vol.65, No.2, pp.251-269, 2017.

[3] 杣取恵太，“本当に麻雀が強いのはだれか？”（豊田秀樹他，たのしいベイズモデリング，18 章），北大路書房，2018.

[4] 馬場真哉，ベイズ統計モデリングによるデータ分析入門，講談社，2019.

[5] 松浦健太郎，Stan と R でベイズ統計モデリング，共立出版，2016.

[6] Groll A., Schauberger G. and Tutz G., “Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: An application to the FIFA World Cup 2014, *Journal of Quantitative Analysis in Sports*, 11, pp.97-115, 2015.

[7] Herbrich, R., Minka, T.,and Graepel, T., "True Skill™ : A Bayesian Skill Rating System" (PDF), Advances in Neural Information Processing Systems 19, MIT Press, pp. 569–576, 2007.

[8] Langville, A. and Meyer C. *Who's #1?: The Science of Rating and Ranking*, Princeton University Press, 2013.（邦訳：レイティング・ランキングの数理－No.1 は誰か？，共立出版, 2015.）

[9] Massey, K. “Statistical Models Applied to the Rating of Sports Teams,” *Bluefield College: Bluefield.*, 1997.

**ウェブページ（URL 順）**

[10] Elo Rating System, https://en.wikipedia.org/wiki/Elo_rating_system（2019/12/23 閲覧）

[11] List of international rugby union teams, https://en.wikipedia.org/wiki/List_of_international_rugby_union_teams（2019/12/13 閲覧）

[12] 将棋棋士成績 DB, http://kenyu1234.php.xdomain.jp/ （2019/12/2 閲覧）

[13] 将棋連盟 棋士別成績一覧（レーティング），http://kishibetsu.com/rating.html（2019/12/2 閲覧）

[14] Project Cabinet Blog　進化系？箱ひげ図, https://k-metrics.netlify.com/post/2018-09/violinplot/ （2020/1/28 閲覧）

[15] レイティングアルゴリズム入門, https://medium.com/eureka-engineering/レイティングアルゴリズム入門-cae5d7dd0db2 （2020/1/28 閲覧）

[16] shogidata.info －将棋のデータベースサイトー，http://shogidata.info/list/rateranking.html（2019/12/2 閲覧）

[17] 松浦健太郎，“階層ベイズモデルで勝敗データからプロ棋士の強さを推定する，” http://statmodeling.hatenablog.com/entry/kishi-rating（2019/11/21 閲覧）

[18] Statsguru, http://stats.espnscrum.com/statsguru/rugby/stats/index.html（2019/12/12 閲覧）

[19] ラグビー世界ランキングの決め方を解説、週次タイムリーな順位変動, https://www.aulii.net/rugby-worldranking/（2019/12/16 閲覧）

[20] World Rugby, “Men’s World Rugby Rankings,” https://www.world.rugby/rankings/mru（2019/12/14 閲覧）