統計数理研究所彙報
第32巻　第1号（1984）

順位のグラフ表現法とその応用

統計数理研究所　馬　場　康　維
岡山大学　脇　本　和　昌
中央大学　遠　藤　紀　久　雄
〃　金　子　武
〃　飯　田　豪
〃　野　中　敏　雄*

（1984年3月　受付）

1. は じ め に

データの構造を理解する上で，データのグラフ化は極めて有用な手段である（脇本・後藤・松原（1979））．グラフ表現法の持つ最大の利点は，人間のパターン認識の能力を用いることによって，数値的な表現では得られない多くの情報を得るところにある．このような観点からここでは，順位データに限って，そのグラフ表現法と応用例について述べる．

m 種類のもの（以下アイテムと呼ぶ）を何らかの判定基準によって順位づけを行うものとする．たとえば官能検査（日科技連（1973））の問題などはこれにあたる．n 人の人によって順位づけされた m 個のアイテムの順位を

$$\begin{array}{cccc} R_{11}, & R_{12}, & \cdots, & R_{1m} \\ R_{21}, & R_{22}, & \cdots, & R_{2m} \\ \vdots & \vdots & & \vdots \\ R_{n1}, & R_{n2}, & \cdots, & R_{nm} \end{array}$$

とする．ここで $R_{\alpha j}$ は α 番目の人が j アイテムに付けた順位を表わし，$(R_{\alpha 1}, R_{\alpha 2}, \cdots, R_{\alpha m})$ は $(1, 2, \cdots, m)$ からなる順列の一つである．

こういうデータを扱う係数の代表的なものの一つとして Kendall の一致度係数がある（Kendall (1948)）．最近，グラフ表現という立場から脇本，白旗によって Kendall の一致度係数と相関度の高い係数が提案されている（脇本・白旗 (1983)）．この一致度係数は連結ベクトルによって表現される多角形〝一致度グラフ〟によって物理的な意味づけができ，その結果一致度の検定が出来ることに加えて順位の持つ構造が表現できるところに特徴がある．

ところで順位データの解析の際，我々は各アイテムに付けられた順位が一致するかどうかの他に，どのアイテムがどういう順位を与えられているかということに関心がある．本稿では順位データの構造をより詳細に表現するものとして〝順位連結グラフ〟と〝順位グラフ〟の二つの表現法を提案する．〝順位連結グラフ〟は各アイテムに与えられる順位の構成比を表現するものである．また〝順位グラフ〟は各アイテムに与えられる順位の平均と分散を同時に表現できるものである．上記のグラフは脇本，白旗の方法で一致度グラフを描くための前処理として用いられるグラフの描き方を少し換えたものに対応している．次節でその表現法を述べ，角度デー

* 現在　東京理科大学．

タ解析の観点からその意味づけを行う．第3節ではその応用例として女子大学生に対するアンケートの分析例を示す．

2. 順位のグラフ表現法

m 個のアイテム，n 標本に対する順位データが与えられているものとする．

$$\begin{array}{cccc} R_{11}, & R_{12}, & \cdots, & R_{1m} \\ R_{21}, & R_{22}, & \cdots, & R_{2m} \\ \vdots & \vdots & & \vdots \\ R_{n1}, & R_{n2}, & \cdots, & R_{nm} \end{array}$$

$(R_{\alpha 1}, R_{\alpha 2}, \cdots, R_{\alpha m})$ は α 番目の標本の m 個のアイテムにつけた順位である．j アイテムの順位データの中で l 順位であるものの標本数を f_{jl} とする．

$$n=\sum_{l=1}^{m} f_{jl} \qquad (j=1, 2, \cdots, m)$$

である．

(1) $$p_{jl}=f_{jl}/n \qquad (j=1, 2, \cdots, m\ ;\ l=1, 2, \cdots, m)$$

と置く．p_{jl} は j アイテムが l 番目とされた相対頻度を表わす．次に

(2) $$\theta_l=\frac{l-1}{m-1}\pi \qquad (l=1, 2, \cdots, m)$$

と置き半径1の半円上に角度θ_l の目盛をつけ，順位 l にθ_l を対応させる（第1図）．
j アイテム l 順位に対応するベクトルを

(3) $$\boldsymbol{x}_{jl}=(p_{jl}\cos\theta_l,\quad p_{jl}\sin\theta_l)$$

とする．$l=1, 2, \cdots, m$ の順にベクトルを結合してできる折れ線をグラフ上に描く．

$$\sum_{l=1}^{m} p_{jl}=1$$

であるから折れ線全体の長さは1である．したがって折れ線の終点は半円の内側にある．折れ線を形成するベクトル $\boldsymbol{x}_{jl}$ の長さが，j アイテムが l 順位となる比率を表わす（第1図）．このようにして描かれるグラフを〝順位連結グラフ〟と呼ぶことにする．次に折れ線の終点と原点とを結ぶ線分を引く（第1図）．これは合成ベクトル

(4) $$\boldsymbol{x}_j=\sum_{l=1}^{m}\boldsymbol{x}_{jl}$$

を表わす．

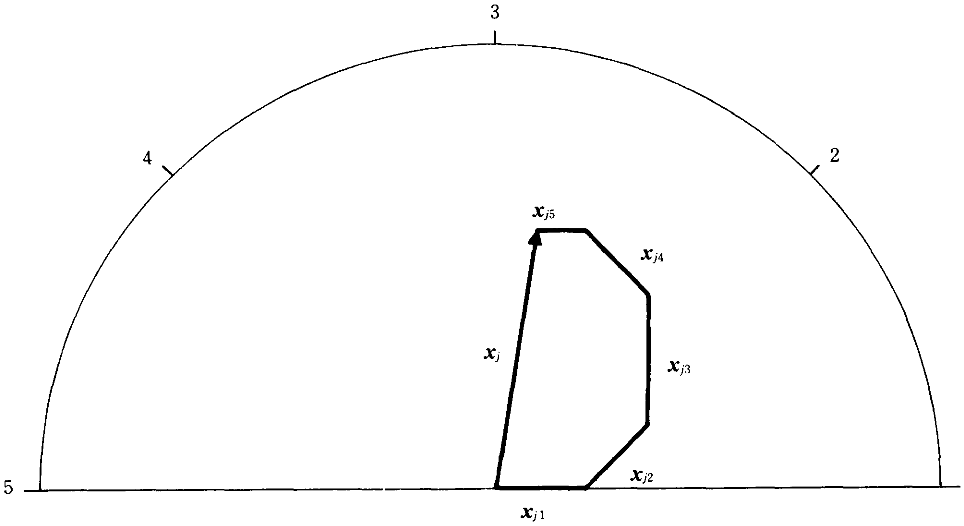
(5) $$r_j=|\boldsymbol{x}_j|$$

(6) $$\phi_j=\arg(\boldsymbol{x}_j)$$

とすると，ベクトルの方向 ϕ_j は j アイテムにつけられた平均的な順位を表わすものと考えられる．したがって折れ線および合成ベクトルからあるアイテムの平均順位と各順位の比率が読みとれる．

合成ベクトル $\boldsymbol{x}_j$ の長さ r_j について考えてみよう．もし j アイテムにつけられた順位 R_{1j}, $R_{2j}, \cdots, R_{nj}$ が全て同じ順位であれば，$r_j=1$ であり，そうでなければ $r_j<1$ となる．したがって r_j は順位のつけ方のバラツキの尺度になる．即ち合成ベクトルが円周に近いほど順位の安定性を示すことになる．各アイテムの合成ベクトルを描くことにより，アイテムの平均順位と順位のバラツキを同時に表示することが出来る．以下では半径1の半円と各アイテムに対応する合成ベクトルを描いたグラフを〝順位グラフ〟と呼ぶことにする．

角度データ解析の観点から順位グラフの意味を考えてみる．(3), (4), (5), (6) より



第１図　順位のグラフ表現の例．各アイテムに対応する折れ線を描いて出来るグラフが"順位連結グラフ"．各アイテムに対応する合成ベクトルを描いて出来るグラフが"順位グラフ"．図は $m=5$ の場合．円周上の目盛は順位に対応する方向を表わす．合成ベクトル $\boldsymbol{x}_j$ の方向が j アイテムの平均順位に対応する．

$$r_j \cos\phi_j = \sum_{l=1}^{m} p_{jl}\cos\theta_l$$

$$r_j \sin\phi_j = \sum_{l=1}^{m} p_{jl}\sin\theta_l$$

(1) より

$$r_j \cos\phi_j = \frac{1}{n}\sum_{l=1}^{m} f_{jl}\cos\theta_l$$

$$r_j \sin\phi_j = \frac{1}{n}\sum_{l=1}^{m} f_{jl}\sin\theta_l$$

したがって ϕ_j は角度データ θ の平均方向 (mean direction)，$1-r_j$ は円分散 (circular variance) に対応する (馬場 (1981))．即ち順位グラフは (2) 式により順位を角度に変換したときの平均方向と円分散を同時に描いたものである．なお順位のバラツキは r_j から便宜的な標準偏差

$$s_j = \sqrt{-2\log_e r_j}$$

によって求められる (馬場 (1981))．

最後に，脇本，白旗の一致度グラフ (脇本・白旗 (1983)) と，ここで述べた順位グラフおよび順位連結グラフ (第３図) との関係に触れておこう．j アイテムに対応する合成ベクトルを $\boldsymbol{x}_j$ その方向を ϕ_j とする ((4), (6) 式)．$\phi_1, \phi_2, \cdots, \phi_m$ の小さい順に並べ換えた合成ベクトルを $\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(2)}, \cdots, \boldsymbol{x}_{(m)}$ とすると，$\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(2)}, \cdots, \boldsymbol{x}_{(m)}, -\boldsymbol{x}_{(1)}, -\boldsymbol{x}_{(2)}, \cdots, -\boldsymbol{x}_{(m)}$ の順にベクトルを連結することによって描かれるものが一致度グラフになっている．一方本稿で示した順位グラフは半円内に $x_1, x_2, \cdots, x_m$ を描いたものである．したがって順位グラフに用いたベクトルから一致度グラフによる検定も行える．

ところで順位連結グラフは合成ベクトル $\boldsymbol{x}_j$ の順位構造を表現している．したがってデータの構造に関して

[データ] → [順位連結グラフ] → [順位グラフ] → [一致度グラフ]

というような図式が成り立っている.

ここでは順位データの表現法について述べてきた. ここで述べたグラフの数理については引き続く課題として現在研究中である.

3. 応 用 例

応用例として結婚に関するアンケートの結果の分析例を示す. 対象としたのは都内7大学(短大)および神奈川県内1大学の女子学生である. このアンケートの中に以下のような順位づけの質問がある. ここではこの質問を中心にした解析例について述べる.

問 結婚相手を決めるのに, 次の要素で選ぶとして, 優先させる順に並びかえて下さい.

1. 親 の 意 見　　2. 相手の体格
3. 相手の顔だち　　4. 相手の家柄
5. 相手の経済力　　6. 相手の性格
7. 趣 味 の 一 致　　8. 相手の学歴

優先する← →優先しない

〔 〕—〔 〕—〔 〕—〔 〕—〔 〕—〔 〕—〔 〕—〔 〕

注意 8項目すべてに順位をつけて下さい.

上記の質問に対する有効回答数は636である. これらの標本をもとに順位の頻度および比率を求めたものを表1に示す. これをもとに平均方向, 円分散, 平均順位, s_j を求めた結果を表2に示す. 第2図には順位グラフ, 第3図には順位連結グラフを示す. 第2図よりほとんどの人が相手の性格を第1位にあげていることが読みとれる. 平均順位が次に高いのが経済力, その次が趣味の一致である. 注目すべきは趣味の一致を表わすベクトルの長さが小さいことである. これは趣味の一致に対する順位は個人差が大きいことを示している.

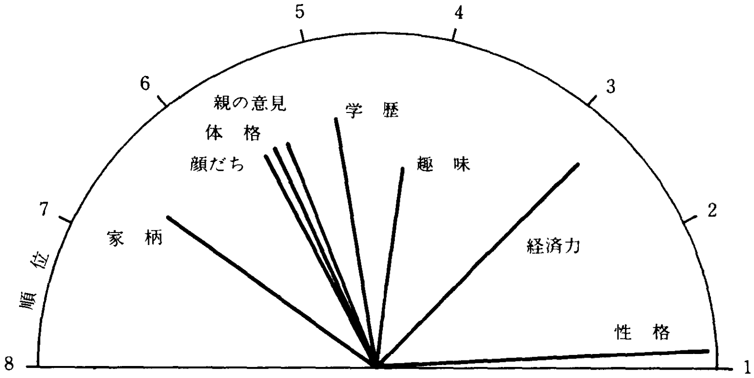
各アイテムごとの順位の比率は順位連結グラフ（第3図）を観察することによって読みとれる. 第3図から約9割の人が相手の性格を第1位にあげていることが読みとれる. 他の質問とのクロス集計をしてもこの比率はほとんど変わらない. 標本を層別して, 各層ごとの質問に対

表1. 各アイテムにつけられた順位の頻度

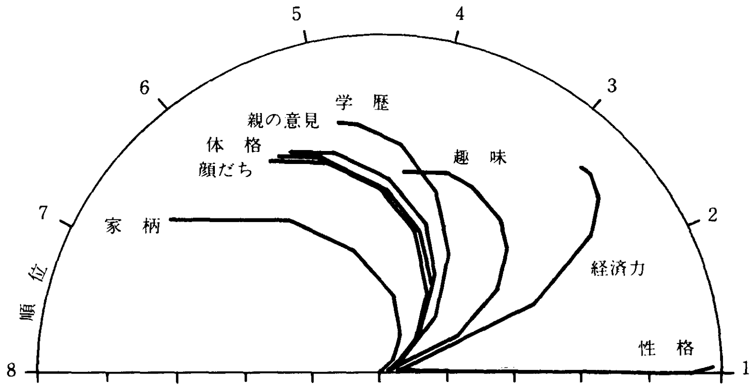
アイテム ＼ 順位	1	2	3	4	5	6	7	8
親 の 意 見	5	36	69	120	101	104	118	83
相 手 の 体 格	2	45	53	108	102	110	132	84
相 手 の 顔 だ ち	8	21	72	89	122	103	112	109
相 手 の 家 柄	1	12	26	51	72	113	125	236
相 手 の 経 済 力	30	290	171	69	46	16	9	5
相 手 の 性 格	575	43	10	5	3	0	0	0
趣 味 の 一 致	8	160	118	73	60	81	52	84
相 手 の 学 歴	7	29	117	121	130	109	88	35

表2．平均方向，円分散，平均順位，標準偏差の表。円分散を v_j としたとき標準偏差は $s_j=\sqrt{-2\log_e(1-v_j)}$ によって求めた

アイテム	平均方向(度)	円分散	平均順位	標準偏差
親の意見	112.1	0.2915	5.36	1.85
相手の体格	115.4	0.2879	5.49	1.84
相手の顔だち	117.6	0.2896	5.58	1.84
相手の家柄	144.0	0.2304	6.60	1.61
相手の経済力	45.6	0.1461	2.77	1.25
相手の性格	3.2	0.0230	1.12	0.48
趣味の一致	82.9	0.4022	4.22	2.26
相手の学歴	99.4	0.2475	4.87	1.68



第2図 順位グラフ．円周上の目盛は順位に対応する方向を表わす．



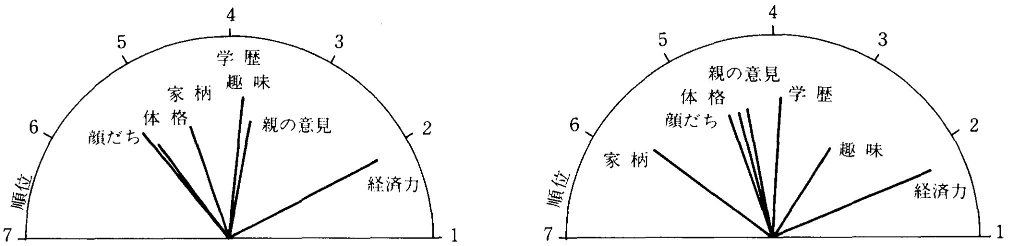
第3図 順位連結グラフ． 半円の径に添った目盛は1目盛が20%の割合に対応する長さを表わす．たとえば"性格"は約90%の人が第1位に挙げている．

する回答パターンの特徴を抽出するのに性格というアイテムはほとんど役に立たない．それ故以下では第1位として相手の性格をあげた標本に対する分析を行う．該当する標本は575標本である．

以下第1位の性格を除き，第2位以下の7アイテムによる結果である．全ての調査結果を示すことが目的ではないので，質問に対する回答によって標本を層別したときの順位グラフのう

問　将来，結婚への準備として，何かしていますか？

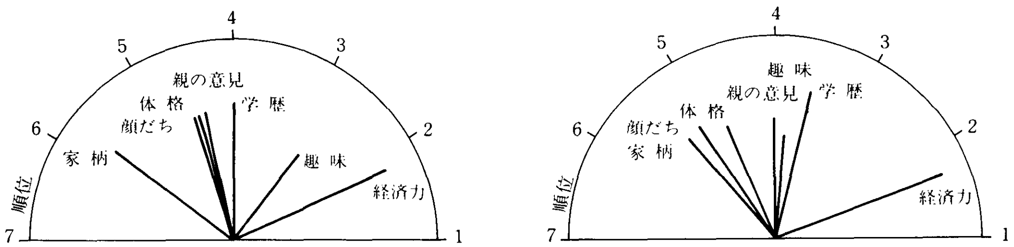
1．はい（44）　　2．いいえ（531）



第4図　結婚の準備をしているかどうかで分類したときの順位グラフ．
括弧の中の数値は該当する標本数．

問　結婚後働く意志はありますか？

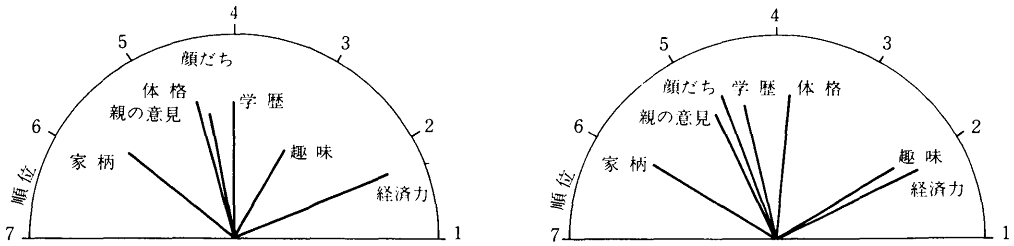
1．はい（417）　　2．いいえ（121）



第5図　結婚後働く意志があるかどうかで分類したときの順位グラフ．
括弧の中の数値は該当する標本数．（無回答37）

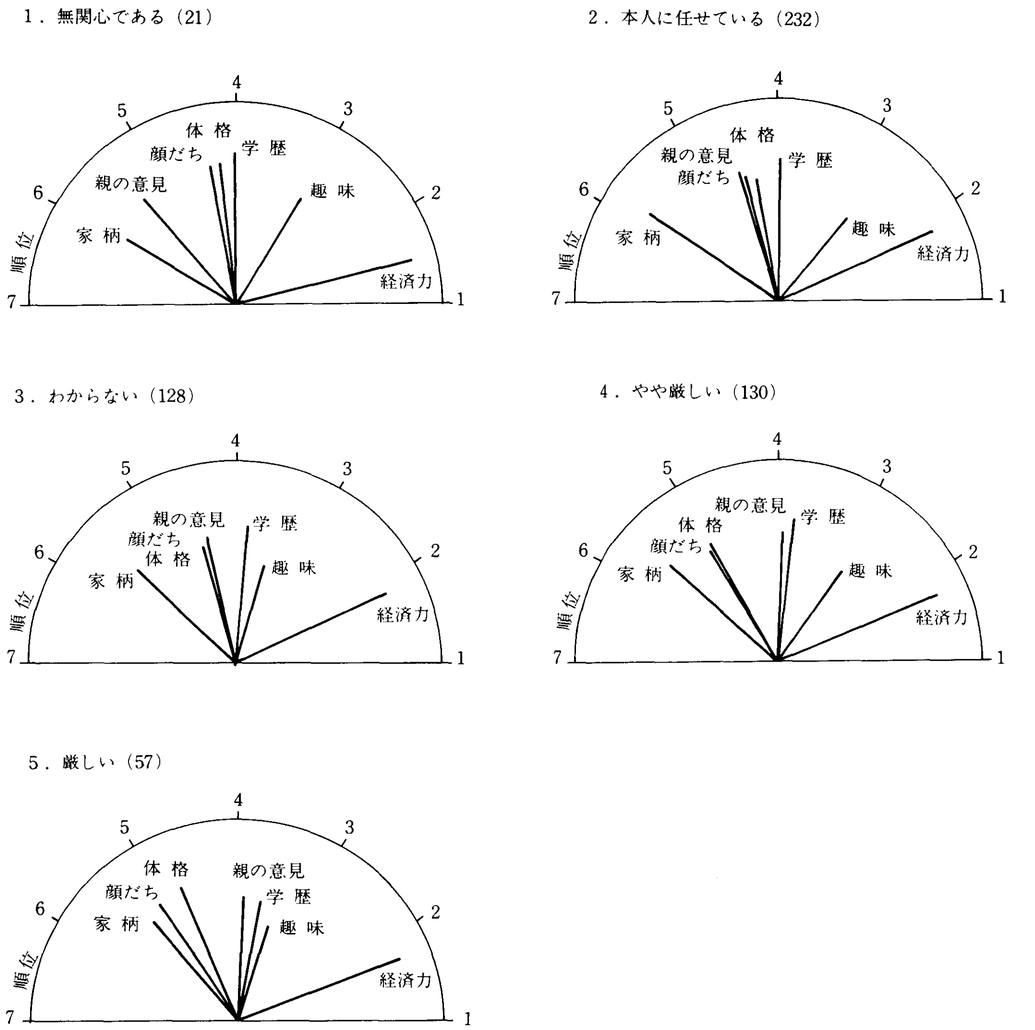
問　結婚は何歳位でしたいと考えていますか？

25才（166）　　30才（22）



第6図　結婚したい年齢別順位グラフ．括弧の中の数値は該当する標本数．
この質問に対する回答内訳は：20歳（1），21歳（1），22歳（12），23歳（63），24歳（136），25歳（166），26歳（74），27歳（44），28歳（24），29歳（3），30歳（22），35歳（1）である．

問　異性との交際について，御両親はどのように考えていると思いますか？（お父さんの考え）．



第7図　異性との交際についての父親の考え別順位グラフ．（無回答 7）

ちの主なものを以下第4図から第7図に示す．

第4図からは結婚の準備をしているかどうかで，家柄，趣味の項の順位が動くことがわかる．なおこの順位グラフのパターンは4年生だけ抜き出しても変わらない．

結婚後働く意志があるかどうかによって学歴と趣味の順位が変わる(第5図)．働く意志のある人は相手との趣味の一致を重視するが，働く意志のない人はそうでもない．

第6図は結婚したい年令によって層別した場合の順位グラフである．標本数の少い層のグラフは除いてある．

第7図は異性との交際についての父親の考えはどうかということによって分類した標本に対する順位グラフである．親の意見の項目が動く．

謝　　辞

ここで取り上げたアンケートは中央大学理工学部管理工学科野中研究室の遠藤，金子，飯田の3名の卒業研究の一部分である．アンケート調査は，北里大学，中央大学，鶴見大学，お茶の水女子大学，上智大学，実践女子大学，専修大学，日本女子大学の8大学で行った．調査にご協力いただいた各大学の先生方，学生諸君に感謝する．また査読者の方には有用な御意見をいただいたことを感謝する．

参 考 文 献

馬場康維（1981）．角度データの統計，統数研彙報，第28巻　第1号，41-54．

Kendall, M.G. (1948). *Rank correlation methods*, Charles Griffin and Company Limited.

日科技連官能検査委員会編（1973）．新版官能検査ハンドブック，日科技連出版社．

脇本和昌，後藤昌司，松原義弘（1979）．多変量グラフ解析法，朝倉書店．

脇本和昌，白旗慎吾（1983）．*m* rankingsの一致度のグラフ表現とその応用，第51回日本統計学会講演報告集．

Graphical Representation of Ranks and Its Application

Yasumasa Baba
(The Institute of Statistical Mathematics)

Kazumasa Wakimoto
(Okayama University)

Kikuo Endo, Takeshi Kaneko, Tsuyoshi Iida
and Toshio Nonaka
(Chuo University)

In this paper we show a graphical method for representing a set of rank data and illustrate this method through a couple of examples.

Let us assume n sets of rankings of m-objects (items) are given by n-observers as follows:

$$\begin{matrix} R_{11}, & R_{12}, & \cdots, & R_{1m} \\ R_{21}, & R_{22}, & \cdots, & R_{2m} \\ \vdots & \vdots & & \vdots \\ R_{n1}, & R_{n2}, & \cdots, & R_{nm} \end{matrix}$$

where $R_{\alpha j}$ denotes the rank given by α-th observer to j-th object. Then $R_{\alpha 1}, R_{\alpha 2}, \cdots, R_{\alpha m}$ is a permutation of the first m positive integers.

Let the frequency of the rank l given by n observers to j-th object be f_{jl}, and put

$$p_{jl} = f_{jl}/n \qquad (j=1, 2, \cdots, m\,;\ l=1, 2, \cdots, m)$$

then p_{jl} denote the relative frequency of the l-th rank to j-th object. We define the angle corresponding to rank l by

$$\theta_l = \frac{l-1}{m-1}\pi \qquad (l=1, 2, \cdots, m)$$

and put

$$\boldsymbol{x}_{jl} = (p_{jl}\cos\theta_l, \quad p_{jl}\sin\theta_l).$$

Then the direction and length of the vector $\boldsymbol{x}_{jl}$ correspond to the rank l and the relative frequency p_{jl}, respectively.

Let us consider the linked vector of $\boldsymbol{x}_{j1}, \cdots, \boldsymbol{x}_{jm}$ to represent rank given to j-th object as follows:

$$\boldsymbol{x}_j = \sum_{l=1}^{m} \boldsymbol{x}_{jl}.$$

This representation corresponds to the constellation graph for the ranked data. We may call this type of graph as "linked rank graph". It should be noted that the direction of the resultant vector is the mean direction of θ_l's and that $1-|\boldsymbol{x}_j|$ is the circular variance of θ_l's. Here $|\boldsymbol{x}_j|$ denotes the length of the vector $\boldsymbol{x}_j$. Therefore, the average and deviation of ranks given to each object may be represented visually by drawing the vector $\boldsymbol{x}_j$'s. We call this type of graph as "rank graph".