

方向統計学における統計的性質の表現

—分散表現を利用したフォン・ミーゼス分布パラメータの推定法の提案—

松 原 達 郎

1. はじめに

製品開発において、筐体部分に用いられるマテリアルの生成は重要であり、近年は製造工程の進歩により高耐久かつ安価なマテリアルが作れるようになり、携帯電話、家具、家電、車の内装パネルなど、様々な製品への応用が期待できる。

一方で生成されたマテリアルの視覚的な特性を分析するのは開発にとって重要なファクターであり、実物から質感の特徴分析をすることにより、消費者を想定した被験者の主観的評価との関係を明確にすることができ、近年の製品開発の手法として体系化が望まれている。

これまでの研究で提案したバーチャルプロトタイプピングにおいて、質感の性質がツヤの変化に現れるとし、方向統計学とコンピュータグラフィックスを用いてツヤを統計分布のバリエーションとして表現した¹⁾。また、実物からの測定を行い、素材間の違いを凹凸の特徴と光の反射性質を統計的にとらえ、素材間の違いと被験者評価の違いを明確にした²⁾。

これまでの研究では測定結果と分布への当てはめでは尖度推定を機械的に行い、推定の統計的性質に関しては考察を行なわなかった。本論文では、測定データから方向統計学の統計的性質に関して、方向統計学で用いられる分布への当てはめの手法について考察を行い、既存の手法との比較を行なった。

2. 方向統計学における統計的性質

2. 1 分析時のデータの表現について

一般の統計学では、データの次元によって線上、面上、空間上と低次元であれば視覚的に表現することが簡単だが、方向統計学の場合は取り扱うデータは複素的な要素を持ち、根本の性質が異なる。

その一つの理由として、方角データを θ で表す時、 θ が周期 2π で循環する。

計算時には方角データ θ を2方向の成分($\cos\theta$, $\sin\theta$)として表現することがある。また複素空間上のベクトルとして下記のように表現することもできる。

$$\cos\theta + i\sin\theta$$

方向統計学では観測データと分析する際の数値表現が異なるため注意する必要がある。 n 個のデータ群($\theta_1, \theta_2, \dots, \theta_n$)からなるmean resultant length \bar{R} は1次の三角モーメントの振幅成分として置かれているが、実質的にはresultant length R に対し、 $\bar{R} = R/n$ と定義されており、 R は以下の式のように全ての方角データからなる合成ベクトルのノルムとして求められている。

$$R = [(\sum_{i=1}^n \cos\theta_i)^2 + (\sum_{i=1}^n \sin\theta_i)^2]^{\frac{1}{2}}$$

合成ベクトルであるということは、異なる方向をさすデータどうしが合成された場合、ベクトルが取り出せてもノルムが小さくなる。ノルムが0になる場合は一様分布として扱われるが、サンプルデータの個数が少ない場合やはずれ値がデータに含まれる

平成28年1月8日受理

連絡先 〒769-0201 香川県綾歌郡宇多津町浜一番丁10番地

香川短期大学 経営情報科

TEL 0877(49)5523 FAX 0877(49)5252

Email matsubara@kjc.ac.jp

と推定されるモーメントが母集団に対して小さくなりやすく、分布への当てはめが困難になる。このため再サンプリングについて考慮する必要が出てくるため、少ないデータの個数から分析するのは実用的でない。

方向統計学では、通常よりもデータの偏りや誤差による影響に敏感であり、ロバストな手法を得ることが難しい。このためモーメントを用いた計算を行なう際にはサンプルサイズの問題やはずれ値に強く注意を払う必要がある。

2. 2 尖度の推定

尖度推定では、方向統計学で定義されている三角モーメントから推定することは前述の理由もあり実用的でない。そこで方向統計学では分布パラメータの最尤推定法から尖度を推定するのが一般的な手法である³⁾。

式3はフォン・ミーゼス分布に対するパラメータの最尤推定法になる。最尤推定値 $\hat{\kappa}_{ML}$ は \bar{R} に影響を受けるため、以下のように \bar{R} の数値ごとに近似数の定義が異なる

$$\hat{\kappa}_{ML} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5/6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1 - \bar{R}) & 0.53 \leq \bar{R} < 0 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & R \geq 0.85 \end{cases}$$

と置かれる。サンプルサイズが少ない場合では、上記の方法では $\hat{\kappa}_{ML}$ の取る値が実際の κ よりも大きくなりやすいため、 $n < 15$ では

$$\hat{\kappa} = \begin{cases} \max(\hat{\kappa}_{ML} - 2(n\hat{\kappa}_{ML})^{-1}, 0) & \hat{\kappa}_{ML} < 2 \\ (n-1)^3\hat{\kappa}_{ML}/(n^3+n) & \hat{\kappa}_{ML} \geq 2 \end{cases}$$

と推定するのが推奨される。ただし、この方法では数値的近似に重みを置いているため、直感的でない。

3. 分散表現と推定

3. 1 方向統計学における分散表現

方向統計学では、定義的に分散という概念で捉えにくいいためvarianceという単語は使われず、

dispersionという単語が使われている。平均方向 $\bar{\theta}$ が求められている時、サンプルデータによるdispersion $D(\theta)$ は下記の式から求められる⁴⁾。

$$D(\bar{\theta}) = \frac{1}{n} \sum_{i=1}^n [1 - \cos(\theta_i - \bar{\theta})]$$

3. 2 分散表現から尖度推定

フォン・ミーゼス分布の確率密度関数からもdispersionを求めることができる。分布の確率密度関数を $p(\theta; \mu, \kappa)$ とし $I_0(\kappa)$ を0次の変形ベッセル関数とすると

$$p(\theta; \mu, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\theta - \mu)]$$

このときdispersionを求めるための関数を $d(\bar{\theta})$ と置くと

$$d(\bar{\theta}) = \int_{-\pi}^{\pi} (1 - \cos\theta) p(\theta; \mu, \kappa) d\theta$$

$\mu = \bar{\theta}$ とおき、確率密度関数からdispersionを求めた場合、 κ に従属する。また、 n が十分に大きい時、 $D(\bar{\theta}) \sim d(\bar{\theta})$ になる。

この $d(\bar{\theta})$ が逆関数の近似式を持つと考える時、近似式を使って κ の近似値をもとめる関数

$$d_{inv}[d(\bar{\theta})] \simeq \kappa$$

を定義する。またサンプルデータのdispersionに対して行なった場合、尖度の推定値として表現することができる。

$$\hat{\kappa}_D = d_{inv}[D(\bar{\theta})]$$

4. サンプルデータに対する尖度推定の評価

4. 1 サンプルデータについて

既存の尖度推定の方法と上記の方法を用いて尖度の当てはまり具合について評価を行った。サンプルデータは $n = 15, 30, 60, 120$ とサンプルサイズが

異なるデータで、 κ は1から10バリエーションを作成した。これらのデータから尖度推定のシミュレーションを行なった。

4. 2 シミュレーション結果について

表1は既存の尖度推定としてML推定、今回の尖度推定手法にD推定という名称にしている。サンプルデータのシミュレーションは50回試行し、試行によってえられた尖度推定値に対し、平均と標準偏差を求めた。D推定ではML推定よりも推定値が大きくなる傾向にある。 $\kappa < 7$ まではML推定のほうが推定値は安定しているが、サンプルサイズが大きくなるにつれ、差が無くなっている。

いずれの推定値も $n=15$ の平均は κ が大きくなるにつれ標準偏差が非常に大きくなっており、実用的なデータでない。どの尖度でも $n=60$ 以降のサンプルが実用的な数値になっている。

また、尖度が大きく、 $\kappa > 9$ ではD推定のほうが若干安定する傾向が見れる。これは図1でのdispersionの変化が $\kappa > 6$ で緩やかになっており、十分にサンプルサイズがある状況では $D(\bar{\theta})$ と $d(\bar{\theta})$ の差が小さくなるので、尖度推定が安定していると考えられる。

5. おわりに

測定データから方向統計学の統計的性質に関して、dispersionを使用してフォン・ミーゼス分布の尖度を推定する手法を提案し、既存の方法と比較し

た。シミュレーション結果から見ても少ないサンプルサイズから分布推定を行なうのは困難であり、最低でもサンプルサイズが30は必要で、実際には60程度のサンプルサイズがないと尖度推定について実用的でないことが分かった。これまでの研究で、方向統計学の分布に対して当てはめを行なった際、前回の研究²⁾では測定した角度データは60~75個程度あったため、測定手法や分析に用いたデータの個数は適切であったと考えられる。

参考文献

- 1) 松原達郎：感性工学におけるバーチャルプロトタイプینگの研究（1）物体表面の質感表現における基礎研究，香川短期大学紀要（39），pp.39-45, 2011.
- 2) 松原達郎，松原行宏，石原茂和：加工紙の表面形状および光の反射特性と表面質感の感性評価，日本感性工学会論文誌 14（1），pp.107-117, 2015.
- 3) N.I. Fisher, T. Lewis, B.J.J Embleton: Statistical analysis of Spherical data, Cambridge University Press, pp.88-89, 1993.
- 4) K.V. Mardia, Peter E. Jupp: Directional Statistics, Wiley, pp.17-20, 1999.

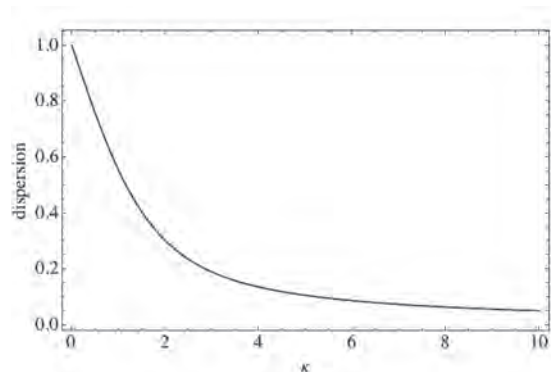


図1 κ とdispersionの対応関係

表 1

 $\kappa = 1$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	0.9723	0.5203	1.0382	0.5327
n=30	0.9428	0.3258	0.9749	0.3114
n=60	1.0114	0.2194	1.0295	0.2139
n=120	1.0191	0.1465	1.0233	0.1452

 $\kappa = 2$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	1.9959	0.6838	2.2082	0.7983
n=30	1.9161	0.4369	1.9867	0.4692
n=60	1.9818	0.3451	2.0248	0.3680
n=120	1.9966	0.2039	1.9972	0.1966

 $\kappa = 3$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	3.0013	1.3107	3.5309	1.5270
n=30	3.0419	0.8074	3.2805	0.8308
n=60	2.8577	0.5166	2.9854	0.3680
n=120	3.0089	0.3621	3.1025	0.3861

 $\kappa = 4$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	4.3702	3.0365	4.7422	2.5070
n=30	4.1311	1.1257	4.4652	1.1981
n=60	3.9819	0.7600	4.1487	0.7531
n=120	3.9874	0.5035	4.0922	0.5003

 $\kappa = 5$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	4.8122	1.9172	5.5542	1.9075
n=30	4.9649	1.3857	5.1880	1.2453
n=60	5.1191	0.9613	5.2726	0.9688
n=120	5.0497	0.6977	5.1188	0.7034

 $\kappa = 6$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	5.8631	2.1018	6.8097	2.2783
n=30	6.1678	1.5892	6.6369	1.7349
n=60	6.1871	1.0973	6.3883	1.1176
n=120	5.9778	0.8014	6.0854	0.8308

 $\kappa = 7$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	6.9763	3.1280	7.6202	2.8123
n=30	6.6673	1.4151	7.1883	1.5461
n=60	6.8720	1.4989	7.1088	1.4495
n=120	7.2389	1.1130	7.3667	1.1016

 $\kappa = 8$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	7.9385	3.9685	8.6707	2.9148
n=30	7.6791	2.1406	8.2171	1.9978
n=60	7.8056	1.1442	8.1085	1.1310
n=120	7.9040	0.9442	8.0619	0.9486

 $\kappa = 9$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	10.4919	5.1877	10.8514	3.3939
n=30	9.4161	2.6343	9.7237	2.1836
n=60	9.2132	1.9821	9.3857	1.7206
n=120	8.9699	1.3040	9.1054	1.2147

 $\kappa = 10$

	ML推定		D推定	
	平均	標準偏差	平均	標準偏差
n=15	10.3189	4.6438	10.8903	3.3702
n=30	10.1429	3.0921	10.3655	2.4218
n=60	10.4020	2.0147	10.4443	1.6917
n=120	9.8318	1.2699	9.8815	1.1274