情報地質 第28巻 第2号 082-085頁 2017年
Geoinformatics, vol.28, no.2, pp.082-085, 2017

選鉱粒子の品位に対する統計分布の当てはめ

正路　徹也

Fitting Statistical Models for Grade of Mineral Dressing Particles

Tetsuya Shoji*

* 東京大学　The University of Tokyo, Tokyo 113-8639, Japan.
E-mail: t-t_shoji@jcom.home.ne.jp

キーワード：鉱石，岩石学，選鉱，品位，統計，ベータ分布，卍分布
Key words: Ore, Petrography, Mineral processing, Grade, Statistics, Beta distribution, Manji distribution

1. はじめに

金属鉱床の価値を地質学的・鉱物学的に評価する場合，採鉱や選鉱過程での特徴や特性を考慮する必要がある．例えば，鉱石の選鉱性は，構成鉱物の種類と粒度，各鉱物間の関係すなわち組織など岩石学的性質に依存し，鉱石の岩石学的性質と適用される分離条件で決まる．選鉱特性が同じならば，粗鉱(≒給鉱)の品位が低下すると精鉱の品位も低下することが，粗鉱と精鉱・尾鉱の間の物質収支に基づく計算から導かれる(正路，2016)．この場合，「選鉱特性が同じ」をどのような因子で判断できるかが重要である．

ところで，2 つの母集団があって，それらの平均値に差があるという仮説が棄却された場合，それ以後の検討は「2 つの母集団の平均は等しい」という仮定のもとに進められる．同様に，複数種類の選鉱粒子群があって，それらの統計的性質に差があるとは認められない場合，「それらの選鉱粒子群の岩石学的選鉱特性は同じ」と仮定することができる．そのためには，まず選鉱粒子がどのような統計モデルに従っているかを解明するする必要がある．そのための 1 つの試みとして，選鉱粒子の品位データに，統計モデルとしてベータ分布と卍(まんじ)分布(仮称)を当てはめてみた．

解析の対象としたデータは，MLA[注 2]で得られた銅鉱石の選鉱粒子の品位である．データには試料 A～H の 8 種類があり，このうち試料 D はさらに 10 組のランダムで排他的な試料 D0～D9 に分割した．分割試料を作成した理由は，同一試料の部分集合におけるばらつきを知るためである．試料 D を選んだ理由は，粒度に対する Rosin-Rammler 分布の当てはめで，当てはめ精度が最悪だったためである．ただし，本稿では，試料 A～H の結果に限って紹介する．

以下の記述で，品位は銅鉱物の含有率を意味する．理由は，ベータ分布も卍分布も変数が 0 と 1 の間で与えられているのに対し，鉱石の銅品位は銅鉱物の銅含有率を上限とし，それ以上，例えば銅鉱物が黄銅鉱 $CuFeS_2$ の場合 34.63% 以上にはならないためである．また，複数の銅鉱物が存在する場合，銅品位のみでは，それが単体粒子か片刃粒子かの判定できないためでもある．

2. 統計モデル

ベータ分布の密度関数は次の式(1)で定義される．

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{1}$$

ここで，x は区間(0.1)で定義される確率変数，α と β は正の定数，$B(\alpha,\beta)$は次の式(2)で定義されるベータ関数である．

$$B(\alpha,\beta) = \int_0^1 \xi^{\alpha-1}(1-\xi)^{\beta-1} d\xi \tag{2}$$

ベータ分布の平均 mean μ と分散 variance $V(x)$は，それぞれ次の式(3)と(4)で与えられる．

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{3}$$

$$V(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4}$$

当てはめでは，定数の μ と $V(x)$を銅鉱物含有率の重量平均と重量分散で置き換えて，α と β を求めた．なお，累積頻度は MS-Excel の組込み関数, BETA.DIST(x, α, β,−1)を利用した．

卍分布の累積頻度は次の式(5)で与えられる[注 1]．

$$M_{\mathrm{H}}(x) = \frac{x^r q}{(1-x)^r(1-q)+x^r q} \tag{5}$$

ここで, r と q はそれぞれ曲率と水平となる縦軸の位置を与える定数である．このうち，q が水平となる位置を与えることは，上式で x = 0.5 とすると $M_{\mathrm{H}}(x) = q$ であることで明らかである．式(5)は次の式(6)のように書き換えられる．

$$\log\left(\frac{1-M_{\mathrm{H}}(x)}{M_{\mathrm{H}}(x)}\right) = \frac{1}{r}\log\left(\frac{1-x}{x}\right) + \log\left(\frac{q}{1-q}\right) \tag{6}$$

$$\log\left(\frac{G(x)}{1-G(x)}\right) = \frac{1}{r}\log\left(\frac{1-x}{x}\right) + \log\left(\frac{q}{1-q}\right) \tag{6'}$$

ここで，$G(x) = 1 - M_{\mathrm{H}}(x)$．式(6)'で，$X = \log\{(1-x)/x\}$，$Y = \log[G(x)/\{1-G(x)\}]$と置くと，次の線形の式(7)が得られる．

$$Y = AX + B \tag{7}$$

ここで，$A = 1/r$，$B = \log\{q/(1-q)\}$．横軸に X，縦軸に Y をとって，変数とその累積頻度の関係を図示したとき，傾きが A，切片が B の直線で近似できると，式(5)あるいは(6)の r と q は次式から得られる．

$$r = 1/A \tag{8}$$

$$q = 10^B/(10^B+1) \tag{9}$$

そこで，X-Y 座標にプロットされた点列に式(7)を当てはめて A と B を求め，式(8)と(9)からそれぞれ r と q を得た．なお，頻度を高品位側から累積する G を使うと，水平となる縦軸の値が平均品位に一致する．

3. 鉱石における鉱石鉱物含有率の統計分布

図 1 に試料 D にベータ分布と卍分布を当てはめた例を示

す．

当てはめ精度の評価は，実測の累積頻度と回帰曲線からの推定値との間の相関係数と平均平方誤差(MSE = Mean Squared Error : 最小二乗法で回帰を行う場合に最小とする値を試料数で除した値)で行った．頻度の値は真数と対数の両方で算出したので，2 種の分布モデル，2 種の評価因子と合わせて，各試料について 8 個の評価値が計算された．真数を使った評価値と対数を使った評価値を較べると，当然対数の方が大きくなる(ここでの真数は 1 以下なので). すなわち，回帰曲線が実測の頻度からより大きくずれている．そこで，ベータ分布の対数に基づく評価で最良を示す試料 A と，卍分布の対数に基づく評価で最悪を示す試料 F の頻度分布図を，それぞれ図 2 左と右に示す．

相関係数と MSE の定義から，両者は当然負の相関を示すと考えられる．しかし，例外的にベータ分布の真数に基づく相関係数と卍分布の真数に基づく MSE の組合せは正の相関を示した(0.335). 一方，予想通り負の相関を示したほとんどの組合せのうち，最も強い相関を示したのは卍分布の対数に基づく相関係数と MSE である(–0.901). 両者をそれぞれ図 3 の左と右に示す．上述の正の相関係数を示す図 3 左で塗りつぶし三角のうち，最大の MSE を示す点は相関係数も大きいため，右上にプロットされている．これに対して，左方の 3 点は相関係数が小さいのに，MSE も小さい．これら 4 点が相関係数を正にしている元凶である．そこで，最大の MSE を与える試料 G と最左端の試料 H の実測頻度と推定頻度の関係をそれぞれ図 4 左と右に示す．実測値と推定値が一致すると，点は図で横軸と縦軸の比が 1:1 の破線上に乗る．左図で，塗りつぶし三角で表されているベータ分布の点列は直線的に並んでいるため，相関係数は高い．しかし，破線と傾斜が違うため破線から次第に遠ざかり，当てはめ精度は低い．一方，中抜き丸の卍分布は原点に近い部分では破線上に並んでいるが，これも次第に離れるため MSE の値が大きくなる．右図のベータ分布の点は途中から破線と離れるので相関係数が小さくなる．一方，卍分布の点は破線の近くにプロットされ，離れ始める品位で実測値がなくなるので，その値は大きくならない．いずれの場合も，ベータ分布の点が破線から次第にずれるのは，回帰曲線が実測値から系統的に離れるためである．これらのことから，実測値と推定値の間の相関係数は，当てはめ精度の評価には必ずしも使えないと結論される．また，このことから，一方が他方の代用にならないことも結論される．したがって，当てはめ精度の評価は MSE で行うべきである．そもそも，最小二乗法とは MSE に試料数を掛けた値を最小する回帰式の係数を求める操作であるから，当然の結論である．

次に，選鉱粒子の品位分布に対するベータ分布と卍分布の適合性を比較するために，図 5 を作成した．実測値の頻度と回帰式から推定された頻度の間の相関係数を示す左図から次の 2 点が明らかになる．1) 対数を表す中抜き丸は横軸と縦軸の比が 1:1 である破線より上にプロットされている．したがって，頻度の対数に対する当てはめ精度は卍分布の方が格段によい．2) 真数を表す塗りつぶし丸は両軸とも 0.9 以上の範囲にプロットされている．したがって，図では，真数の場合，当てはめ精度はベータ分布の方が多少よいように見えるが，両者の差は小さい．右の実測値の頻度と回帰式から推定された頻度の間の MSE を示す図からは，次の 2 点が明らかとなる．1) 中抜き丸が右上に，塗りつぶし丸が左下にプロットされていることから，当てはめ誤差は真数の方が対数より小さい(真数が 1 以下であることは前述). 2) 真数・対数とも 1:1 の破線より下にプロットされているので，すべての試料について卍分布の方が当てはめ精度が高い．

前述したように，式(6)’の q は高品位側から累積頻度 G が水平となる位置を与え，この値が平均品位にほぼ一致する(予想 1). また，r は卍分布の $x \approx 0$ および $x \approx 1$ 付近における曲率を与えるので，分布が水平となる範囲の幅に関係すると考えられる．もしそうならば，片刃粒子の量が少ないと幅が狭くなるので r は大きく，量が多いと幅が広くなるので r は小さくなると予想される(予想 2). この 2 つの予想を確かめるために，平均品位と q の関係(図 6 左)，および片刃粒子の量と r の関係(図 6 右)を描いた．図 6 左は予想 1 が成り立つことを明確に示している．これに対し，図 6 右によると，予想 2 は成り立たず，片刃粒子の量と r との間には何らの関係も認められない．

4. まとめと今後の課題

銅鉱山の選鉱粒子から得られた MLA データを使って，品位(銅鉱物含有率)の統計分布を調べ，以下の結論を得た．1) 統計モデルとして，ベータ分布と卍分布を当てはめたところ，卍分布の方が優れていた．特に，高品位側からの累積頻度が少ない部分を対数表示すると適合性の差がより一層鮮明になる．2) 当てはめ精度の評価は，実測の累積頻度と回帰曲線からの推定値との間の相関係数より MSE (平均平方誤差)の方が優れている．3) 卍分布で水平位置を与える定数 q の値は鉱石の平均品位に一致する．

ベータ分布の当てはめ精度が悪かった理由の 1 つに，当てはめ方が考えられる．ここでは，式(3)と(4)を使って定数の α と β を求めた．これに対し，評価因子の値が最低となる α と β を繰り返し計算で求めると，当てはめ精度が向上する可能性がある．

今回解析の対象とした試料は銅鉱石である．このため，平均品位が 1%前後だった．今後は，金や白金など平均品位が数 ppm の鉱石に対しても同様の統計モデルの適用が可能か否かを調べる必要がある．

謝辞 : 解析の対象とした MLA データは，早稲田大学，創造理工学研究科の所研究室から提供された．このため，同研究室の所千晴教授を始めとする多くの方から協力と助言を得た．なかでも，堀内健吾博士と大学院学生の福田宏樹氏からは特に有用な協力と有益な助言を得た．記して感謝の意を表する．

引用文献

正路徹也(2016) : 発見された鉱床の選鉱特性の評価は地球情報学の役目．情報地質，**27**(*2*)，42-47.

注 1 : 式(5)は，(–∞,∞)で定義されるロジスティック関数を[0,1]での定義に変換したのち，その逆関数として得られる．また，その密度関数 $m_{\mathrm{H}}(x)$は，式(5)を微分して得られる次式である．

$$m_{\mathrm{H}}(x)=\frac{x^{r-1}q(1-x)^{r-1}(1-q)}{\{(1-x)^{r}(1-q)+x^{r}q\}^{2}}$$

注 2 : MLA は，Mineral Liberation Analysis (鉱物単体分離解析)あるいは Mineral Liberation Analyzer (鉱物単体分離解析装置)の略．

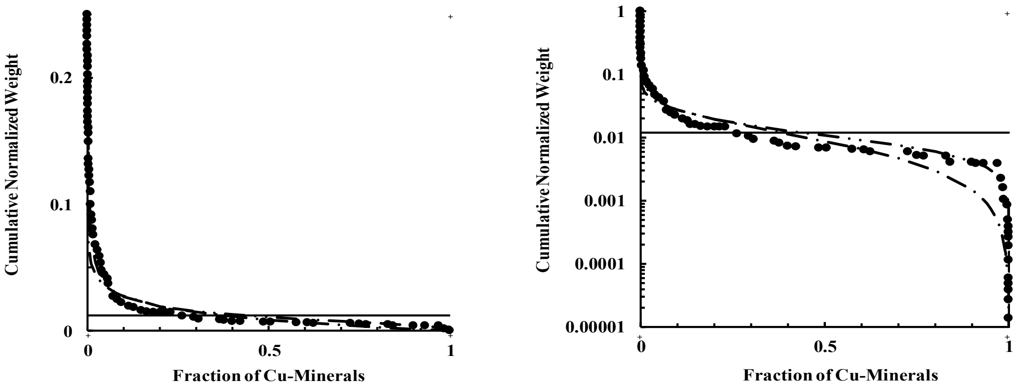

図 1．試料 D にベータ分布と卍分布を当てはめた結果．横軸は品位(銅鉱物含有率)，縦軸は高品位側からの累積頻度．鉱石の銅鉱石含有率は 1.21%．黒丸が実測値，1 点鎖線がベータ分布の，2 点鎖線が卍分布の回帰曲線．縦軸の値が 0.01 付近の水平な実線は卍分布の q の値．左は頻度分布図で品位が 0 となる頻度より下の部分，右は頻度が対数表示．頻度の表示が線形で上限が 1 の図では，実測値，回帰直線とも横軸とほぼ重なって，それぞれの識別が困難である．

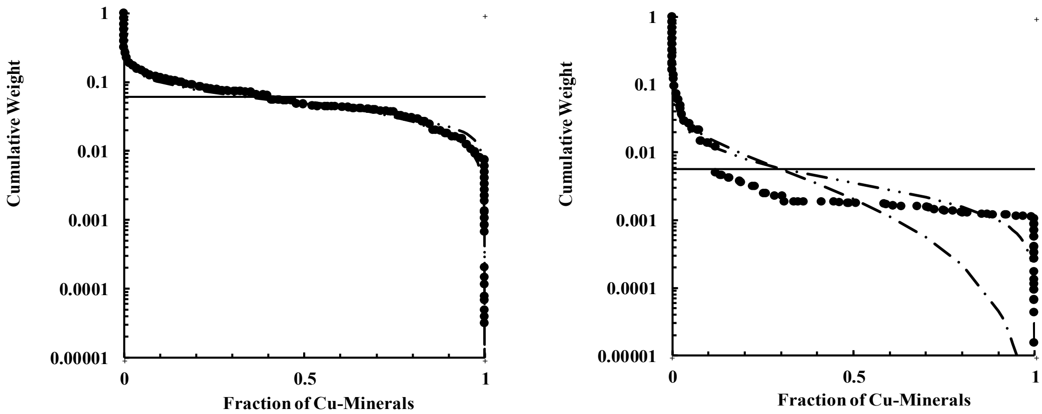

図 2．対数に基づく評価値で最良を示した試料 A (左)と最悪を示した試料 F(右)の頻度分布図．銅鉱物の含有率は，試料 A (左)が 6.12%，試料 F (右)が 0.57%．黒丸が実測値，1 点鎖線がベータ分布の，2 点鎖線が卍分布の回帰曲線．縦軸の値が 0.1 (左)と 0.01 (右)付近の水平な直線は卍分布の q の値．

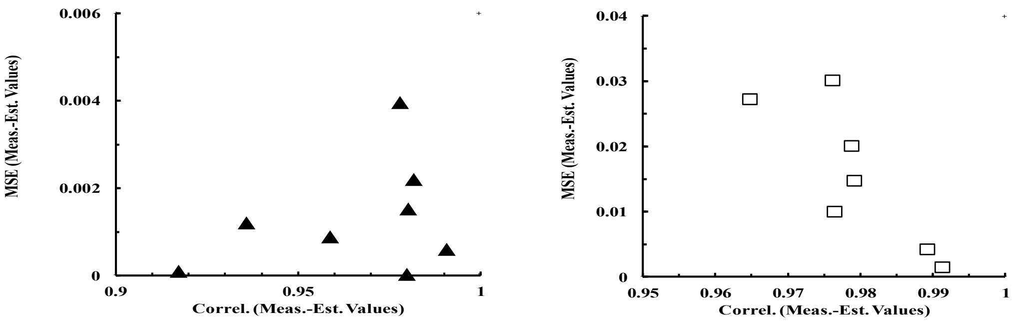

図 3．ベータ分布および卍分布の当てはめにおける実測頻度と推定頻度の間の相関係数と MSE との散布図で，最悪の相関(左：0.335)と最良の相関(右：–0.901；計算で頻度は対数に変換)の例．

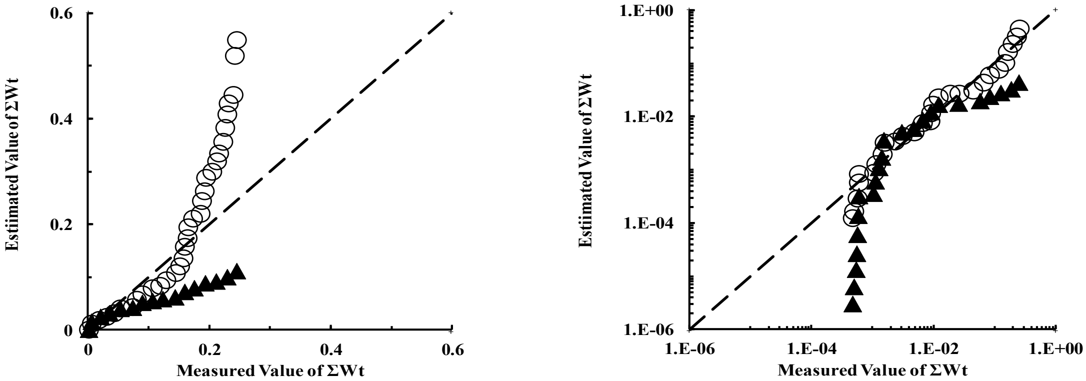

図 4．図 3 右に示されている散布図(最悪の相関を示すベータ分布の真数に基づく相関係数と卍分布の真数に基づく MSE の組合せ)にプロットされている点のうち，最大の MSE を与える試料 G (左 : r = 0.978, MSE = 0.0040)と最小の相関係数を与える試料 H (右 : r = 0.836, MSE = 0.001)の実測頻度(横軸)と推定頻度(縦軸)の関係．両図とも，塗りつぶし三角がベータ分布，中抜き丸が卍分布．

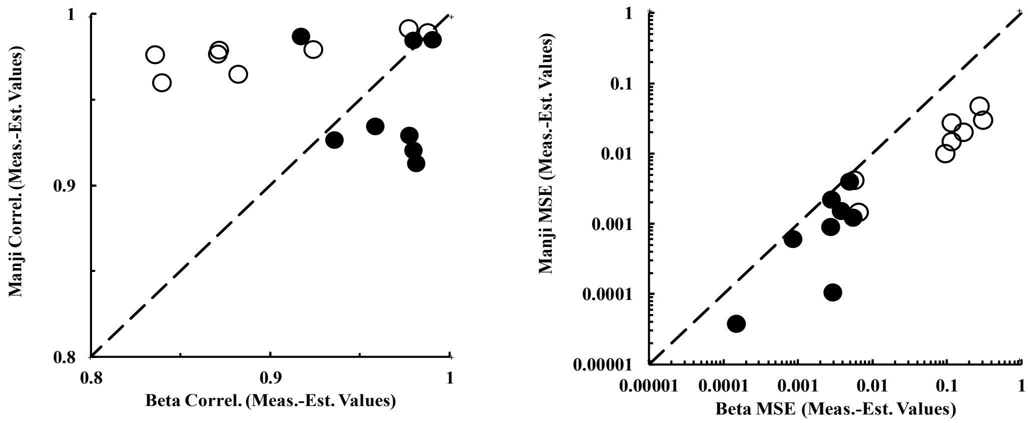

図 5．実測頻度と推定頻度の間の相関係数(左)と MSE (右)のベータ分布(横軸)と卍分布(縦軸)の関係．塗りつぶし丸は真数，中抜き丸は対数．破線上は縦軸と横軸の比が 1:1．

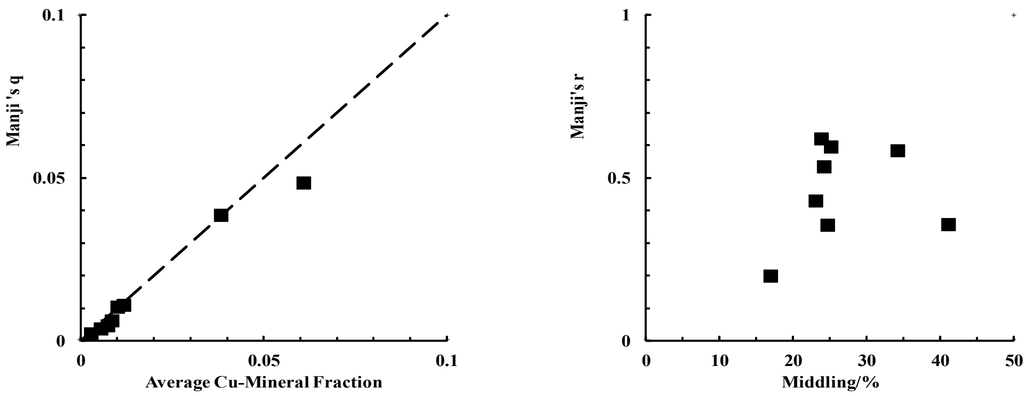

図 6．左) 試料の平均銅鉱物含有率と卍分布の水平部の位置を与える定数 q との関係．右) 試料の片刃粒子の量と卍分布の曲率を与える定数 r との関係．