尺度点（目盛り）決定・測定標準作成における一つの統計的考え方について

林　知己夫

(1958 年 10 月受付)

Some Statistical Methods on the Determination of Scale Points

Chikio HAYASHI

In the present paper, we shall treat the following problems.

(i) Suppose that an interval $[x_0, x_{R+1}]$ is given and these points x_0, x_{R+1} are end points of a scale. We should like to make R scale points, $x_1, x_2, \cdots, x_R$, between them. When we want to measure an object, it is, of course, desirable to minimize the error of measurement interpolating it between two scale points. Thus, it is desirable to determine the scale points to minimize the errors in measurement. First, we must take a universe, all elements of which is to be measured by the scale. Let x's be the numerical labels (measured values) of the elements in the universe which has a diffrentiable density function $f(x)$ with respect to x.

x_0 x_1 x_2 $\cdots\cdots\cdots\cdots$ x_R x_{R+1} We assume that the numerical values of an object determined by repeated mesurements, the true value of which exists, are a random variable, the mean of which is the true value and the degree of magnitude of error in measuremt is to be represented by its standard deviation which is proportionate to the distance between the true value and the nearest scale point. Thus we take the following measure as the total errors of measurements of all elements in the universe in question.

$$Q^2 = \sum_{i=1}^{R+1} k^2 \int_{x_{i-1}}^{x_i} \min\{(x-x_i)^2, (x-x_{i+1})^2\} f(x)\,dx$$

where k is a constant.

We determine x_i $(i=1, 2 \cdots R)$ to minimize Q^2. x_i $(i=1, 2, \cdots R)$ must statisfy,

$$x_i = \int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} x \frac{f(x)}{\int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} f(x)\,dx}\,dx$$

If, $f(x)$ is uniformly constant, the scale points of equal distance are optimum. We fancy, scale points of equal distance are conventionally determined under the assumption of $f(x)$ being uniformly constant. It seems to be reasonable to revise the underlying assumption. Especially, this idea is indispensable in the Thurston's scale construction.

(ii) The errors of scale points determined by a bisection method between A and B are discussed. The definition of scale points will be shown in Fig 1. (91P.) These results will be shown in (α), (β), at 91～92P.P.

The Institute of Statistical Mathematics.

ここでは尺度点（目盛り）や測定標準を作成するに当つての統計的考え方についてのべてみよう．これらについては，慣例などによつてきめられることが多いと思われるので，例をあげつつ統計の立場から考えなおしてみることにする．

§1. ものを測定しようとするとき，測定のための物差しを必要とする．物差しには尺度点（目盛り）がつけられてあり，これと測定しようとするものと比較することによつて，測定値がきまつてくるものである．長さや重さを測る事を頭におけば容易に理解できることである．また態度のようなものを測定するとき，（サーストンの方法を用いるときには，ある事に対し非常に好意的なものから非好意的なものに到る間いくつかの短文を配列するのである．これが物差しの目盛りに相当するものとなる．測定されるものは自分の意見に最も近いものを見つけて反応することになる，これは物差しで長さを測るのと全く同じである．このとき，どのような目盛りの短文を用意するかが問題になるし，短文の目盛りはしかるべき手続きによつて決定されるのであるがいまは詳しくは論じない——例えば高木貞二編，心理学における数量化の研究，東大出版，昭和 30 年参照．さてそれではこれらには）どのような目盛りをつけるべきであるかが問題になるが，一般には等間隔目盛り（等間隔の標準具の採用）が用いられてあり，ある場合には対数による目盛り（対数尺度の標準具の採用）が用いられてある．後者のような場合には，潜在的にはウェーバー—フェッヒネルの法測が考えられて居るものと思われる．即ち物理尺度の対数が感覚尺度においてある意味で通常の長さをはかるときの物差しの様な——厳密ないい方ではないが——機能をもつと考えられるからであろう．つまり感覚尺度の意味で等間隔の目盛りを得たいためには，物理尺度では変換された尺度で，目盛りがつくられねばならないと考えられているものと思われる．いずれにしても最後の測定において等間隔目盛りということが中心になつていることが多いのである．一般にこのような考え方でよいであろうか，更に態度測定のような場合，この考え方でよいであろうか，考えてみる必要がある．物差し——すべてをひつくるめてこう言い現わしておく——でものを測定しようとするとき，尺度点（目盛り）にぴつたり合うようなものはなく，尺度点の間に測定しようとするものが落ちるのが普通であるから，われわれはそれを尺度点（目盛り）を利用して測定値を読みとる，即ち内挿することを行つて測定値をつくることをしているのである．サーストンの場合でも，調査されるものは自分の意見にぴつたり一致する短文はないために，自分の意見にもつとも近い短文を選んで反応しなければならないので事情は全く同様である．一方は測定者が，一方では被測定者が読みとりを行つて，測定値を出しているのである．したがつて，測定値には必ずそのための誤差を伴なつているものと考えてよい．このように考えてくるならば測定の誤差が最小になるように尺度点（目盛り）をきめてくる必要がおこつてくる．このような考え方の下に尺度点をきめる事を以下にのべてみる．この場合，その物差しを用いて，測定しなければならないすべてのものの集団を考えに入れなくてはならない．これがユニヴァース（調査すべき対象の集り——いままで調査対象と呼んでいたがここで名づけたようにする方が解り易くてよいので今後からこう呼ぶことにする）となる．このユニヴァースの性質が異るならば尺度点のきめ方も異つてくるのである．ユニヴァースとしていかなるものをとるべきかを考えることが，尺度点決定の第一の問題点である．次に，そこから抽出されるべき「測定しようとするもの」を考えるのであるが，抽出される確率が等しい確率であるとして母集団を構成するものとする．したがつて，母集団から抽出されたサムプルが「測定すべきもの」となつてくるのである．ここまで考えておいて，ユニヴァースを決定しておくことが大事な事

である．このきめ方によつて，ある特定のものに対して有効妥当な尺度点の決定もあれば，ひろい範囲を考えて一般的に妥当な尺度点の決定——特定のものに対しては特によくないが，ひろい範囲で平均的によいものになる——もあり得る．要はここまで考えて，実際の必要に応じて尺度点をきめればよいのであつて，大事な事は尺度点はユニヴァースと相対的な関係にあるものと考えておく事にある．

いま，物差しは数量で表現されているものとし，その端点を x_0, x_{R+1} としこの間に R 個の尺度点をもうける事を考えてみる．これを $x_1, x_2, \cdots x_R$ とし $x_0<x_1<x_2<\cdots<x_R<x_{R+1}$ とする．測定標準をいかにつくるかもこれと同じである．例えば比較法による測定のための「色の標準」を作くろうとするとき——色は無限にある——R をきめれば，いかなる目盛りを附与された標準を選んで作くるべきか等もこの考え方になつてくるのである．——通常は等色度の標準をつくるべく努力されているようであるが，いままで述べた考え方によれば必ずしも妥当ではない，測色すべきユニヴァースを考えて標準はつくるべきものと考えられる．

簡単のため一次元で考えを進めてみる．多次元であつて全く同様であり拡張は容易である．前述の x_0, x_{R+1} は二次元以上のヴェクトルではないとする．またこの x_0 と x_{R+1} との間はすべて同質である，つまり内挿するときの誤差は x_0 と x_{R+1} との間の各点で等価におこりうる（どこどこでは特にちがつたおこり方をすることがあるというようなことはない）ものとしておく．

x_0　　x_1　　x_2　x_3 $\cdots\cdots\cdots\cdots\cdots\cdots$ x_R　　x_{R+1}

（$x_0<x_1<x_2<x_3<\cdots\cdots<x_R<x_{R+1}$ とする）

測ろうとする集団の要素は，ある微分可能な密度函数 $f(x)$ をもつものとする——分布函数だけでもよいのであるが簡単のため微分可能な密度函数をもつものとしておく——．このように測ろうとする集団を意識的に考え，その分布の状態を考慮する所が大切なのである．

さて，あるもの x をもち来つて，この値をはかろうとしたとき，それが x_i と x_{i+1} の間のあるところにおちたとする．その時われわれは内挿によつてその数値をきめるのであるが，このための誤差——われわれは測定によつて決める値が確率変数であり，その平均は所謂測ろうとするものの妥当な値であり，誤差の程度はその標準偏差として計量されるものとする——がおこるものとする．そうしてこの標準偏差は端点からの距離に比例するものとする，即ちもし $x_i \leqq x \leqq \dfrac{x_i+x_{i+1}}{2}$

$x_i \overset{\longleftrightarrow}{\quad} x$　　x_{i+1}　（$x_i<x_{i+1}$）　ならば $x-x_i$ に比例し，もし $\dfrac{x_i+x_{i+1}}{2}<x\leqq x_{i+1}$ ならば $x_{i+1}-x$ に比例するものとする．

つまり x_i と x_{i+1} の中央附近のとき誤差が最もなくなるのである．

また x_i と x_{i+1} の間におちるような x の母集団での相対頻度は $\int_{x_i}^{x_{i+1}} f(x)dx$ となつている．

さて，全体的にみた誤差の総量として確率変数たる測定値の分散の重みつき（相対頻度を重みとする，こうすれば分散の総和が出る）和というようなものを考えると

$$Q^2=\sum_{i=1}^{R+1} k^2 \int_{x_{i-1}}^{x_i} \min\{(x-x_i)^2,\ (x-x_{i-1})^2\} f(x)dx \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathrm{A})$$

となる．ここに k は比例常数とする．

書きなおすと

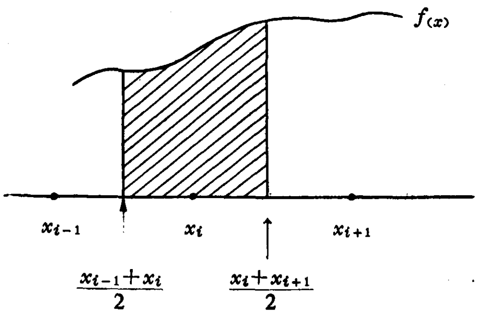
$$Q^2=k^2\Big\{\cdots\cdots+\int_{x_{i-1}}^{\frac{x_{i-1}+x_i}{2}}(x-x_{i-1})^2 f(x)\,dx+\int_{\frac{x_{i-1}+x_i}{2}}^{x_i}(x_i-x)^2 f(x)dx+$$
$$\int_{x_i}^{\frac{x_i+x_{i+1}}{2}}(x-x_i)^2 f(x)dx+\int_{\frac{x_i+x_{i+1}}{2}}^{x_{i+1}}(x_{i+1}-x)^2 f(x)dx+\cdots\cdots\Big\}$$

となる．一応ここで微分して極小値を求めてみることにする．$\dfrac{\partial Q^2}{\partial x_i}=0,(i=1,2,\cdots R)$ をつくると

$$x_i\int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}}f(x)dx=\int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}}xf(x)dx$$

即ち
$$x_i=\int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}}x\frac{f(x)}{\int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}}f(x)dx}dx,\quad (i=1,2,\cdots R)\cdots\cdots\cdots\cdots\cdots\text{(B)}$$

となり，斜線の部分の平均値が分点となつているのである．Q^2 の形（x_i に関し微分可能，一価函数）と x_i に対する条件から，このような解は一般に存在し，それが存在するならば，また $f(x)\neq 0$（端点を除く）であるならば，(*B*) を満足するものは，極小を与えることがわかる．かくして我々の目的とする最小のものを求めることが出来る．



いま x_i $(i=1,2,\cdots R)$ をすべて少しうごかしてみることにし，x_i の代りに x_i に近い $x_i+\varepsilon_i$（ε_i は十分小さいとしておく）をとりあげ，これを Q^2 に代入してみる．これを $Q^2{}_i$ としておく．上述の x_i を代入した Q^2 を $Q_0{}^2$としておく．これを単純にほごして，比較してみると $f(x_i)\neq 0,(i=1,2,\cdots R)$ であるならば常に

$$Q_0{}^2<Q^2{}_i,\quad(\varepsilon_i\neq 0,\ \varepsilon \text{ は十分小},\ i=1,2,\cdots R)$$

となる．つまり常に $Q_0{}^2$ が小さくなつてくる．即ち $Q_0{}^2$ がその附近で最も小さくなり，求めたものが，われわれの目的に適することがわかる．$f(x)=0$ のある場合でも場合を分ければ議論をつくすことが容易に出来る．分点が $f(x)$ の形に依存する所が面白い．これによると密度の多いところに分点が多くできる傾向があり，$f(x)$ が小さい所では分点が粗くなつてよいのである．

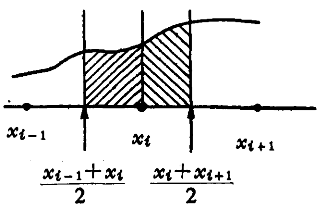
少し誤差の計量のし方をかえて

$$Q^2=\sum_{i=1}^{R+1}k\int_{x_{i-1}}^{x_i}\min\{|x-x_i|,\ |x-x_{i+1}|\}f(x)dx\cdots\cdots\cdots\cdots\text{(A)*}$$
$$k>0$$

と考えるならば分点は，

$$\int_{\frac{x_{i-1}+x_i}{2}}^{x_i}f(x)dx=\int_{x_i}^{\frac{x_i+x_{i+1}}{2}}f(x)dx,\quad(i=1,2,\cdots R)\cdots\cdots\cdots\text{(B)*}$$

となる．

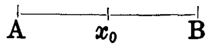


つまり等面積分割となる．

いずれにしても $f(x)$ が一様分布の時には等間隔分割が一番よくなるのである．ここが大事な点である．一般の尺度を構成するときは，暗黙のうちに一様分布の考えが，底流しているものと考えてよい．小さなものさしで長さを測ろうとする場合などその性質上一様分布が仮定されるのもおかしくないが，特殊目的のもの，うごかすのに困難なものであれば $f(x)$ の形を考えねばならず，前にものべた態度測定のため——サーストンの方法による——にいかなる短文をどんな間隔に配置す

べきか等については，$f(x)$ の概略を予備調査などからしらべ，ここでのべたような考察を行うのが妥当であろうと思われる．

§**2.**　割り込みによる折半法を用いて，尺度点をきめる場合がある．標準の点が A, B と両端に



あり，その中央の点を実験によつてきめる．これを x_0 としておく．次に（A と x_0）の中央の点及び（x_0 と B）の中央の点を実験によつてきめる．こうして中央の点が実験によりきめられると，また次にそれぞれの中央の点をきめる，こうして等間隔の尺度点をきめる方法がある．このようなとき，その誤差はどうなるであろうかという問題を考えてみる．この時も誤差の程度（確率変数たる測定値の標準偏差で表現されるものとしておく，勿論測定値の平均値は正しい値を示すものとする）は中央の値をきめようとしている時の両端の間隔（尺度の意味は §1 参照）に比例するものとしておく．

符号のつけ方をきめておく．左側を 1，右側を 2 とつけておく．そして常に新しくできた点を基準にして次の点の左右をきめてゆくことにする．

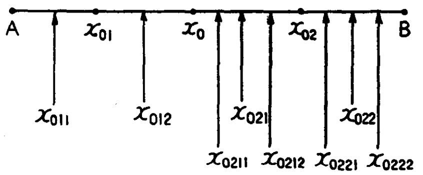


Fig 1

A と B との間の距離を R とする．また原点を A にとつて尺度点の位置を座標によつてあらわすことにする．また各点でのみおこる誤差は独立としておく．

誤差をあらわす確率変数を総称して ε とかく．この下つきは定めようとする点の各段階でおこるものを示すとする．各段階での ε はそれぞれ独立であり，その平均値は 0 としておく．

$$\left.\begin{aligned}
&x_0=\frac{\mathrm{R}}{2}+\varepsilon_0,\quad E(\varepsilon_0)=0,\quad E(\varepsilon_0{}^2)=\sigma_0{}^2=k^2R^2,\quad k\ \text{はある比例常数},\\
&x_{02}=x_0+\frac{1}{2}(R-x_0)+\varepsilon_{02}=\frac{R}{2}+\frac{x_0}{2}+\varepsilon_{02}=\frac{3}{4}R+\frac{\varepsilon_0}{2}+\varepsilon_{02},\\
&E(x_{02})=\frac{3}{4}R,\quad \sigma_{02}{}^2=E\left(x_{02}-\frac{3}{4}R\right)^2\\
&\qquad\qquad\qquad\qquad=\frac{1}{4}k^2R^2+\frac{1}{4}k^2R^2=\frac{1}{2}k^2R^2\\
&x_{022}=x_{02}+\frac{1}{2}(R-x_{02})+\varepsilon_{022}\\
&\qquad=\frac{\mathrm{R}}{2}+\frac{1}{2}x_{02}+\varepsilon_{022}\\
&\text{したがつてこの分散}\quad \sigma_{022}{}^2=\frac{1}{4}\sigma_{02}{}^2+\frac{1}{16}k^2R^2=\frac{3}{16}k^2R^2\\
&x_{021}=x_0+\frac{1}{2}(x_{02}-x_0)+\varepsilon_{021}\\
&\qquad=\frac{1}{2}(x_0+x_{02})+\varepsilon_{021}\\
&\qquad=\frac{1}{2}\left(\frac{\mathrm{R}}{2}+\frac{3}{2}x_0+\varepsilon_{02}\right)+\varepsilon_{021}\\
&\qquad=\frac{5}{8}R+\frac{3}{4}\varepsilon_0+\frac{1}{2}\varepsilon_{02}+\varepsilon_{021}\\
&\sigma_{021}{}^2=\frac{1}{4}(\sigma_0{}^2+\sigma_{02}{}^2+2\,C_{0,02})+\sigma_{021}{}^2=\left(\frac{9}{16}+\frac{1}{16}+\frac{1}{16}\right)k^2R^2=\frac{11}{16}k^2R^2
\end{aligned}\right\}(\alpha)$$

ただし $C_{0,02}$ は x_0 と x_{02} とのコヴァリアンスを示す．

明らかに端点が固定されている場合の分散は小さくなつて居り中央の方が端にくらべて分散が大きいことに注意すべきである．左側も同様である．

高次のものは同様にくりかえせばよいが，もう一回だけ示しておく．一般式をつくることも容易である．

$$x_{0211}=x_0+\frac{1}{2}(x_{021}-x_0)+\varepsilon_{0211}$$

$$=\frac{1}{2}(x_{021}+x_0)+\varepsilon_{0211}$$

この分散　$\sigma_{0211}{}^2=\frac{1}{4}(\sigma_{021}{}^2+\sigma_0{}^2+2\ C_{0,021})+\sigma_{0211}{}^2$

$$=\frac{k^2R^2}{4}\left\{\left(\frac{11}{16}+1+1+\frac{1}{2}\right)+\frac{1}{16}\right\}$$

$$=\frac{13}{16}k^2R^2$$

$$x_{0212}=x_{021}+\frac{1}{2}(x_{02}-x_{021})+\varepsilon_{0212}$$

$$=\frac{1}{2}(x_{02}+x_{021})+\varepsilon_{0212}$$

したがつて　$\sigma_{0212}{}^2=\frac{1}{4}(\sigma_{02}{}^2+\sigma_{021}{}^2+2\ C_{02,021})+\sigma_{0212}{}^2$

$$=\frac{9}{16}k^2R^2$$

等々である．

次に間隔はどうなるであろうか，下つきが3個までのものについてみよう．

$$(x_{021}-x_0),\quad (x_{02}-x_{021}),\quad (x_{022}-x_{02}),\quad (R-x_{022})$$

について考える．平均は勿論 $\frac{R}{8}$ である．この分散をそれぞれ $\tau_1{}^2$, $\tau_2{}^2$, $\tau_3{}^2$, $\tau_4{}^2$ とすると

$$\left.\begin{aligned}
\tau_1{}^2&=E\left\{(x_{021}-x_0)-\frac{R}{8}\right\}^2=E\left\{\frac{1}{2}(x_{02}-x_0)+\varepsilon_{021}-\frac{R}{8}\right\}^2\\
&=E\left\{-\frac{\varepsilon_0}{4}+\frac{\varepsilon_{02}}{2}+\varepsilon_{021}\right\}^2=\frac{3}{16}k^2R^2\\
\tau_2{}^2&=E\left\{(x_{02}-x_{021})-\frac{R}{8}\right\}^2=E\left\{\frac{1}{2}(x_{02}-x_0)-\varepsilon_{021}-\frac{R}{8}\right\}^2\\
&=E\left\{-\frac{\varepsilon_0}{4}+\frac{\varepsilon_{02}}{2}-\varepsilon_{021}\right\}^2=\tau_1{}^2\\
\tau_3{}^2&=E\left\{(x_{022}-x_{02})-\frac{R}{8}\right\}^2=E\left\{-\frac{x_{02}}{2}+\varepsilon_{022}+\frac{3}{8}R\right\}^2\\
&=E\left\{-\frac{\varepsilon_0}{4}-\frac{\varepsilon_{02}}{2}+\varepsilon_{022}\right\}^2=\tau_1{}^2\\
\tau_4{}^2&=E\left\{-\frac{\varepsilon_0}{4}-\frac{\varepsilon_{02}}{2}-\varepsilon_{022}\right\}^2=\tau_1{}^2
\end{aligned}\right\}(\beta)$$

となりいたる所で値が同一であるのは興味深い．

x_0 より左側についても全く同様の式をかくことができる．これによれば間隔の分散はすべて同一となつていることが解る．なお間隔相互間の相関係数は上記の式が示す通り 0 ではない事は勿論である．より高次のものでも全く同様である．即ち

$(x_{0211}-x_0)$ の分散は $\left\{\frac{1}{2}(x_{021}-x_0)+\varepsilon_{0211}\right\}$ の分散となり $\frac{1}{4}\tau_1{}^2+\frac{k^2R^2}{64}$ となる．

また（$x_{0212}-x_{021}$）の分散は $\left\{\frac{1}{2}(x_{02}-x_{021})+\varepsilon_{0212}\right\}$ の分散で $\frac{1}{4}\tau_1{}^2+\frac{k^2R^2}{64}$ となり，

また（$R-x_{0222}$）の分散は $\frac{1}{4}\sigma_{022}{}^2+\frac{k^2R^2}{64}=\frac{1}{4}\tau_1{}^2+\frac{k^2R^2}{64}$，となり全く同様に等しくなる．

なおこのような間隔の相対誤差は，間隔の数が大になるにしたがつて大きくなつてきていることは容易にわかる．

一般式を書くのは容易である．尺度点そのものの誤差は中央に到るほど増大するが，間隔の誤差は同一なのである．

こうして作られた目盛りをもつものさしを用いてあるものを測定しようとするときは，内挿を考える前§の考え方と，ここでのべた尺度点そのものの誤差とを結合させればよいのである．

（統計数理研究所）