中國統計學社叢書之一

統計論叢

目　　錄

上海黎明書局發行

中華民國二十三年六月出版

中國統計學社叢書之一

統計論叢

1934

上海黎明書局發行

序　言

一國之統計事業欲有長足的進步與健全的發展，至少應從三方面努力：一爲統計學理之高深研究及正當宣傳，二爲統計行政之制度完善及組織嚴密，三爲統計事實之妥慎調查及精細整理。吾國統計事業雖有悠久之歷史，但因墨守成法，不求進步，奉行故事，視同具文。於上述三方面皆缺少不斷之努力。故一旦與他國比擬，遂事事相形見絀，瞠乎其後。且國家一切政治設施與夫經濟社會文化教育諸事業，皆缺乏精確可靠之統計以供參考。今後我們欲發奮圖强，勇猛精進。使我們國族的共同生活與國民的個人生活，均更爲安全鞏固，更爲高尚文明。則統計事業當愈趨重要，統計功用亦應愈見昭著。方能適應時代之需求。本社同人平素對於統計學術頗饒興趣。同時復鑒於吾國統計事業之不振。爰有學社之組織及本刊之發行。對於吾國統計前途實抱有極大之宏願。目下全國統計行政正從事根本革新。統計法規亦將分別次第完成。黨政最高當局倘

能努力提倡，切實扶助，我國今後的統計工作必能大放光明。是以本社同人願各本所學，以與國人相見。縱不能爲發明家，固願爲宣傳家與實行家。盡心竭力，爲國家社會服務。至本刊文字責任均由同志著者自負，編輯同人僅略盡催稿之義務而已。

民國二十三年五月編者誌

目　　錄

(以著作人姓之筆畫多寡爲次序)

次數分配之分析——基本的問題

王仲武

〔本文曾由著者在民國二十年本社第一屆年會宣讀〕

次數分配之分析法——各種度量——配合曲線——次數曲線系之比較——用轉矩配合次數曲線法——算術平均數與一次轉矩——標準差與二次轉矩——偏度與轉矩——峯度與轉矩——衆數與轉矩——平均差與轉矩。 II. 原轉矩之意義——原轉矩之假定——薛伯氏校正——皮爾生與派爾蒙(Pairman)之校正法——根據插補法之校正法——平均斜度校正法——實例。 III. 皮氏之轉矩計算法——實例——各種方法之比較。 IV. 中位數計算法——實例。 V. 衆數計算法——改進方法。 VI. 配合曲線之用途——轉矩之重要。

I.

次數分配之分析可大別爲兩種。爲通常目的求得分配之各種度量卽足以敍述分配之性質，其最主要而常用者爲全體分配之總數，中數（平均數，衆數，中位數或其他百分位數）及

離中差度（標準差，平均差，偏態，峯態）等。再進一步的分析法卽爲曲線之配合。至次數曲線之成爲系統者，要不外下列兩類：一類爲皮爾生氏曲線系(Pearson's System of Frequency Curves)，其他爲常態曲線之種種變化。例如西爾(Thiele)，格蘭姆（Gram），愛基華斯（Edgeworth）等之次數函數是。後者一類之次數函數由機率的假定推求而得，從理論的觀察自較皮氏之系統爲滿意。惟其長處亦卽其缺點。蓋一般事實現象之次數分配每有相當的偏度（尤其在經濟，社會，生物等現象中），而此等函數不能適應偏度稍大之分配。卽使並無適應偏態之困難，至少一端受限制之分配（社會，經濟，教育，心理等現象中甚多），或J形U形之分配不能用此種函數配合，皮氏曲線系則無此缺點。

配合曲線之方法通常用最小二乘法。然若曲線之方程式非多項式而爲超越函數（Transcendental Functions），例如皮爾生氏之曲線型，則最小二乘法往往不切實用或竟爲不可能者。關於此點皮爾生氏曾明白指出(見 Biometrika Vol. I, Part 3. p. 267)，并倡轉矩（Moment）方法以代最小二乘法，氏曾證明此法之理論的根據。其實，倘次數曲線爲多項式（拋物線），則此法與最小二乘法完全一樣。

用轉矩配合曲線，無論在理論上或實際上，並非唯一的方法。皮氏自認此法在理論上不能視爲『最佳』（best）之法，蓋

所謂『最佳』無合理的定義。至於事實上特殊的曲線往往有特殊便利的配合法，卽如人壽保險學上之實例是。然此種特殊方法每不能施之於一般曲線，故不能視爲普遍的有系統的配合方法。若論普遍的有系統的配合法，則除轉矩之法外，尙未有所發現。故轉矩之計算，仍爲次數分配之分析的基礎。

皮氏曲線配合法有賴轉矩之能否準確求得。倘用不甚準確之轉矩配合曲線，則失却配合曲線之本意。然轉矩之用尙不止爲配合曲線也。次數分配之主要度量，除百分位數外幾乎皆可用轉矩表示之。蓋算術平均數卽一次轉矩，標準差卽二次轉矩之平方根。

設μ_n爲第n次轉矩，則:

$$\sqrt{\beta_1}=\frac{\mu_3}{\mu_2^{\frac{3}{2}}}$$

爲偏態度量之一種。$\sqrt{\beta_1}=0$ 表示對稱（對於算術平均數上之縱坐標而言。）其他各種偏態之度量，除最簡陋之方法用算術平均數與衆數之距離外，皆根據百分位數而定。百分位數通常不易準確求得。卽使其能準確求得，此種偏態之度量對於各次數之變化幷無靈敏之感覺。$\sqrt{\beta_1}$則不然，任何一組之次數，倘有些微變動卽能影響$\sqrt{\beta_1}$之值。又根據百分位數之偏態度量之精度（Precision）視全體分配之總數而定。倘全體分配之總數不大，則此種偏態度量不甚可靠。$\sqrt{\beta_1}$則不受此影響。

峯態則可用

$$\beta_2=\frac{\mu_4}{\mu_2^2}$$

爲度量，在常態曲綫上$\beta_2=3$。倘$\beta_2>3$則分配之中部較常態曲綫爲平坦。反之，倘$\beta_2<3$則較常態曲線爲陡峭。

至於衆數則可分爲實際分配上之衆數與配合次數曲線上之衆數。後者之衆數可用轉矩表示，設配合次數曲線之微分方程式爲

$$\frac{1}{y}\frac{dy}{dx}=\frac{x+a}{b_0+b_1x+b_2x^2}$$

此處a_1及b_0,b_1,b_2爲常數。此系曲線之各次轉矩可直接由此微分方程式求得。若假定曲線之兩端與x軸相遇(不必相切)則

$$a=\frac{\sigma\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}$$

故

$$x=-a=-\frac{\sigma\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}$$

卽次數曲線上之衆數(原點算術平均數)。故求曲線上之衆數不必實際求得曲線之方程式，只須求得轉矩之値卽可。至於實際分配上之衆數當於下節詳述。

倘分配爲常態者，則平均差爲標準差之$\sqrt{\frac{2}{\pi}}$倍，故亦可用轉矩表示之。

綜上所述可知轉矩之計算，實爲次數分配分析之基本問題。

II.

設 f 爲次數，m 爲組中值，則四個轉矩之公式爲

$$v_1' = \frac{\Sigma fm}{\Sigma f}$$

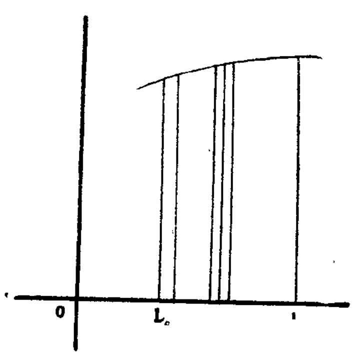
$$v_2' = \frac{\Sigma fm^2}{\Sigma f}$$

$$v_3' = \frac{\Sigma fm^3}{\Sigma f}$$

$$v_4' = \frac{\Sigma fm^4}{\Sigma f}$$

由此四式算出之轉矩通常稱爲原轉矩 (Raw Moments)。卽次數分配之轉矩，但與次數曲線之轉矩不同，

設O爲原點，L_1 與 L_2 爲一組之上下限，m爲組中值，I 爲組距，y=f(x) 爲次數曲線之方程式。假定將數組距 L_1 L_2 分成若干份，在每一分點上作一縱坐標。如是得多個長而狹之四邊形。將每一四邊形之面積乘其底邊範圍內任何一點與原點之距離。將此各乘積相加。如是相加求得之



和，在 L_1L_2 間部分之數目無限制的增加，每一份無限制的縮小時之極限值，除以曲線下之全面積，卽此一組之一次轉矩。用積分之符號表示之，次數曲線上專就此一組而言之一次轉矩為

$$\int_{m-\frac{1}{2}}^{m+\frac{1}{2}} x f(x)\,dx \Big/ \int_{A}^{B} f(x)\,dx.$$

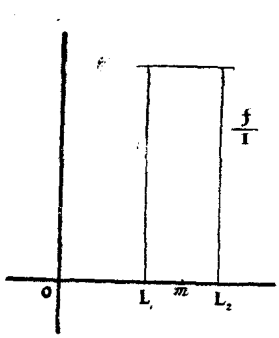
此處 A,B 為曲線兩端之範圍。

原轉矩專就此一組而言之一次轉矩為

$$\frac{fm}{\Sigma f}$$

此式之值等於假定在此一組內次數曲線為與 x 軸平行之直綫。設此直綫與 x 軸之距離為 f/I，組距組中値如前，則

$$\int_{m-\frac{1}{2}}^{m+\frac{1}{2}} (fx/\mathrm{I})\,dx = fm$$



此種假定之缺乏普遍性自不待言。故實際計算時原轉矩求得後尚須加以一種校正，使其更近於次數曲線上之轉矩。

然此處之困難乃求四個轉矩之目的，卽為求配合次數曲線，次數曲線之形式，既尚未求得，則如何在原轉矩上加以校正。

薛伯氏（Sheppard）所用之方法，為假定次數曲線之兩

端與 x 軸爲密切者。用數學的術語述之，卽假定次數函數之一，二，三次微分係數在兩端皆爲零。如是以奧愛勒墨克勞倫定理（Euler-Maclaurin Theorem）用級數表示積分。如是求得之近似値，卽著名之薛伯氏校正（Sheppard Correction 詳見 Biometrika Vol. V）。設 $\nu_1'\ \nu_2'\ \nu_3'\ \nu_4'$ 爲依次四個原轉矩，$\mu_1'\ \mu_2'\ \mu_3'\ \mu_4'$ 爲次數曲線之依次四個轉矩，則薛伯氏之校正爲

$$\mu_1'=\nu_1'$$

$$\mu_2'=\nu_2'-\frac{1}{12}I^2$$

$$\mu_3'=\nu_3'-\frac{1}{4}\nu_1'I^2$$

$$\mu_4'=\nu_4'-\frac{1}{2}\nu_2'I^2+\frac{7}{240}I^4$$

此種校正方法早經世人引用，無須作者介紹。惟本文目的，幷非介紹，乃對於此種校正法，加以批評幷另覓他種更普遍的校正法。區區愚見，不無有相當價值，尙希各位明達指教！

薛伯氏校正，假定次數函數之一，二，三次微分係數在兩端皆爲零。然此種條件往往與次數分佈之形式不合。卽就皮爾生氏之各種次數曲線型而言，其第一，二，三，八，九，十一，十二各種曲線，往往一端雖與 x 軸相遇而並不密切，縱令其一次微分係數在一端爲零，而二次三次微分係數並不等於零。或爲 J 形，或爲 U 形，則薛伯氏校正之不適用自不待言。

故此種校正並非任何分配皆可應用者，倘不顧此種校正之假定而盲用之，所得結果往往反遠不如原轉矩之爲佳。關於此點下節當舉實例以證明之。

皮爾生氏因感此種校正之乏普遍性，曾用其他方法計算轉矩（詳見 Biometrika Vol. XII. P. 231 et seq.），其法爲假定次數函數在起端之鄰近可用與愛勒墨克勞倫定理展開成

$$1+\frac{a_1}{1!}(x-A)+\frac{a_2}{2!}(x-A)^2+\cdots\cdots$$

之形式；在盡端展開成，

$$1+\frac{b_1}{1!}(x-B)+\frac{b_2}{2!}(x-B)^2+\cdots\cdots$$

之形式。此處 A,B 爲曲綫之起點與終點對於原點之距離，a_1, a_2,……, b_1,b_2,……等爲可用起端之各組次數，與盡端之各組次數表示者。

皮氏如是求得與愛勒墨克勞倫公式中所含曲綫兩端之微分係數，因此得

$$a_1=-\frac{1}{60}\left\{137f_1-163f_2+137f_3-63f_4+12f_5\right\}$$

$$a_2=\frac{1}{12}\left\{45f_1-109f_2+105f_3-51f_4+10f_5\right\}$$

$$a_3=-\frac{1}{4}\left\{17f_1-54f_2+64f_3-34f_4+7f_5\right\}$$

$$a_4=\left\{3f_1-11f_2+15f_3-9f_4+2f_5\right\}$$

$$a_5=-\left\{f_1-4f_2+6f_3-4f_4+f_5\right\}$$

其餘 b_1, b_2, b_3, b_4, b_5 之五個公式與此完全相像。以上各式中 f 爲第 r 組之相對次數。如是求得四個轉矩爲

$$\mu_1' = v_1' + \frac{1}{12}\left(a_1 - \frac{1}{60}a_3 + \frac{1}{2520}a_5\right) + \frac{1}{12}\left(b_1 - \frac{1}{60}b_3 + \frac{1}{2520}b_5\right)$$

$$\mu_2' = v_2' - \frac{1}{12} - \frac{1}{120}\left(a_2 - \frac{5}{126}a_4\right) + \frac{1}{6}A\left(a_1 - \frac{1}{60}a_3 + \frac{1}{2520}a_5\right)$$
$$- \frac{1}{120}\left(b_2 - \frac{5}{126}b_4\right) + \frac{1}{6}B\left(b_1 - \frac{1}{60}b_3 + \frac{1}{2520}b_5\right)$$

$$\mu_3' = v_3' - \frac{1}{4}v_1' - \frac{1}{40}\left(a_1 - \frac{5}{63}a_3 + \frac{1}{240}a_5\right) - \frac{1}{40}A\left(a_2 - \frac{5}{126}a_4\right)$$
$$+ \frac{1}{4}A^2\left(a_1 - \frac{1}{60}a_3 + \frac{1}{2520}a_5\right) - \frac{1}{40}\left(b_1 - \frac{5}{63}b_3 + \frac{1}{240}b_5\right)$$
$$- \frac{1}{40}B\left(b_2 - \frac{5}{126}b_4\right) + \frac{1}{4}B^2\left(b_1 - \frac{1}{60}b_3 + \frac{1}{2520}b_5\right)$$

$$\mu_4' = v_4' - \frac{1}{2}v_2' + \frac{7}{240} + \frac{1}{126}\left(a_2 - \frac{7}{80}a_4\right) - \frac{1}{10}A\left(a_1 - \frac{5}{63}a_3 + \frac{1}{240}a_5\right)$$
$$- \frac{1}{20}A^2\left(a_2 - \frac{5}{126}a_4\right) + \frac{1}{3}A^3\left(a_1 - \frac{1}{60}a_3 + \frac{1}{2520}a_5\right)$$
$$+ \frac{1}{126}\left(b_2 - \frac{7}{80}b_4\right) - \frac{1}{10}B\left(b_1 - \frac{5}{63}b_3 + \frac{1}{240}b_5\right)$$
$$- \frac{1}{20}B^2\left(b_2 - \frac{5}{126}b_4\right) + \frac{1}{3}B^3\left(b_1 - \frac{1}{60}b_3 + \frac{1}{2520}b_5\right)$$

至於 J 形曲綫之轉矩計算法，皮氏假定曲線之起端爲向上之一端，並假定其作

$$N\left\{1 + \chi^q(A + B\chi + \cdots\cdots)\right\} \qquad q < 1$$

之形式。如是求得起端第一組之轉矩(以曲線起點在 x 軸上之射影作原點）爲

$$f_1\mu_1'=-.8127818f_1+.6770691f_2-.6605497f_3$$
$$+.3471889f_4-.0737827f_5$$
$$f_1\mu_2'=.7067407f_1-.8241137f_2+.8305586f_3$$
$$-.4415218f_4+.0943572f_5$$
$$f_1\mu_3'=-.6347502f_1+.8572689f_2-.8807900f_3$$
$$+.4714747f_4-.1011083f_5$$
$$f_1\mu_4'=.5814517f_1-.8541149f_2+.8888688f_3$$
$$-.4780407f_4+.1027607f_5$$

如是求得之第一組之各個轉矩須加於其餘各組之各轉矩（亦以起端爲原點)。J 形曲線上之$a_1,a_2,\cdots\cdots$之值爲

$$a_1=-.0679063f_1-1.6512369f_2+1.1771875f_3$$
$$-.5548634f_4+.1118034f_5$$
$$a_2=.3322458f_1+.9155792f_2-2.3842525f_3$$
$$+1.3685243f_4-.2981424f_5$$
$$a_3=-.9887796f_1+2.8237204f_2-2.1260271f_3$$
$$+.4720491f_4-.0279508f_5$$
$$a_4=2.4976496f_1-9.9398504f_2+13.3946734f_3$$
$$-7.8229490f_4+1.6770510f_5$$
$$a_5=-7.4134958f_1+25.3208762f_2-33.8993138f_3$$
$$+20.7685399f_4-4.8564601f_5$$

倘爲U形則當然兩端皆如是。用此法求轉矩甚爲精確(見下節實例。)然此法極爲繁複,誠有令人望洋興嘆之感。

況此法並非唯一的方法。倘以插補法爲基礎,次數分配之

轉矩未始不可求得。例如在相鄰每三組上配二次拋物綫或在相鄰每五組上配四次拋物綫，卽根據二次或四次拋物線，求各次轉矩。惟此處所宜注意者，卽各組之次數通常不能視爲高度，而必須當作次數曲線下各組內之面積。故普通之插補方法並不適用。

設 f_1, f_2, f_3, f_4, f_5 爲相鄰五組之次數。求依次各組之面積等於 f_1, f_2, f_3, f_4, f_5 之四次拋物線。設此拋物線之方程式爲

$$y_{x+1}=a+bx+cx^2+dx^3+ex^4$$

以組距爲單位則

$$f_1=\int_{-\frac{1}{2}}^{\frac{1}{2}} y_{x+1}dx=\left[ax+\frac{b}{2}x^2+\frac{c}{3}x^3+\frac{d}{4}x^4+\frac{e}{5}x^5\right]_{-\frac{1}{2}}^{\frac{1}{2}}$$

$$=a+\frac{1}{12}c+\frac{1}{80}e \cdots\cdots\cdots\cdots (1)$$

$$f_2=\int_{\frac{1}{2}}^{\frac{3}{2}} y_{x+1}dx=\left[ax+\frac{b}{2}x^2+\frac{c}{3}x^3+\frac{d}{4}x^4+\frac{e}{5}x^5\right]_{\frac{1}{2}}^{\frac{3}{2}}$$

$$=a+b+\frac{13}{12}c+\frac{5}{4}d+\frac{121}{80}e \cdots\cdots\cdots\cdots (2)$$

$$f_3=\int_{\frac{3}{2}}^{\frac{5}{2}} y_{x+1}dx=\left[ax+\frac{b}{2}x^2+\frac{c}{3}x^3+\frac{d}{4}x^4+\frac{e}{5}x^5\right]_{\frac{3}{2}}^{\frac{5}{2}}$$

$$=a+2b+\frac{49}{12}c+\frac{34}{4}d+\frac{1441}{80}e \cdots\cdots\cdots\cdots (3)$$

$$f_4=\int_{\frac{5}{2}}^{\frac{7}{2}} y_{x+1}dx=\left[ax+\frac{b}{2}x^2+\frac{c}{3}x^3+\frac{d}{4}x^4+\frac{e}{5}x^5\right]_{\frac{5}{2}}^{\frac{7}{2}}$$

$$=a+3b+\frac{109}{12}c+\frac{111}{4}d+\frac{6841}{80}e \cdots\cdots\cdots (4)$$

$$f_5=\int_{\frac{7}{2}}^{\frac{9}{2}} y_{x+1}dx=\left[ax+\frac{b}{2}x^2+\frac{c}{3}x^3+\frac{d}{4}x^4+\frac{e}{5}x^5\right]_{\frac{7}{2}}^{\frac{9}{2}}$$

$$=a+4b+\frac{193}{12}c+\frac{260}{4}d+\frac{21121}{80}e \cdots\cdots(5)$$

解此五式

$$a=\frac{1}{1920}(1689f_1+684f_2-746f_3+364f_4-71f_5)$$

$$b=\frac{1}{1920}(-3800f_1+6960f_2-4800f_3+2000f_4-360f_5)$$

$$c=\frac{1}{1920}(2760f_1-8160f_2+8880f_3-4320f_4+840f_5)$$

$$d=\frac{1}{1920}(-800f_1+2880f_2-3840f_3+2240f_4-480f_5)$$

$$e=\frac{1}{1920}(80f_1-320f_2+480f_3-320f_4+80f_5)$$

在方程式

$$y_{x+1}=a+bx+cx^2+dx^3+ex^4$$

中令 $x=0,1,2,3,4$，即得此曲線在相鄰五組組中值上之高度，換言之即組中值上之縱坐標。

$$y_1=a$$

$$y_2=a+b+c+d+e$$

$$y_3=a+2b+4c+8d+16e$$

$$y_4=a+3b+9c+27d+81e$$

$$y_5=a+4b+16c+64d+256e$$

以上列 a,b,c,d,e 之值代入得

$$y_1=\frac{1}{1920}(1689f_1+684f_2-746f_3+364f_4-71f_5) \cdots\cdots\cdots\cdots (6)$$