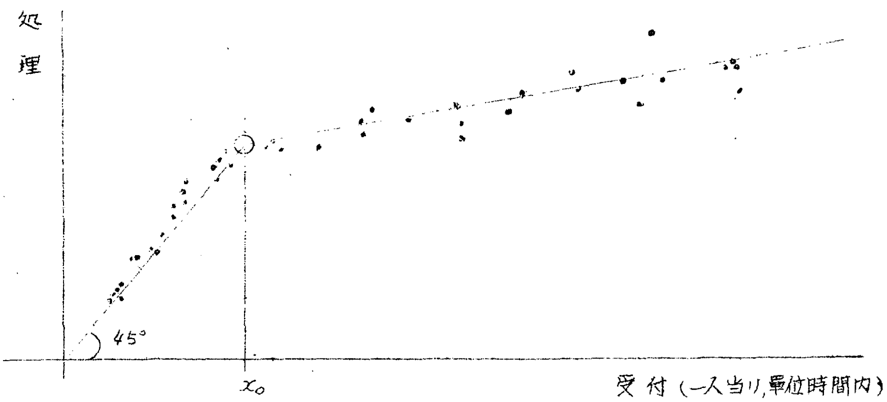
(25) 推定された二直線の交点の信頼巾について

林　知己夫

この問題は次の様な実際問題を取扱ふ時計算したものである。

準長距離電話を取扱ふ電話局で「出現する Call 」を受付けたとき，これを各 operator が処理するのであるが，単位時間における operator 一人当りの call がある限度以上に達すると受付けた call すべてを処理出来なくり，処理率，処理数は減少してくるのである。それまではまづ受付けた call はすべて処理できるものと観察せられた。これを実際に調査してみたところ（全 Operator のグループからランダムにサムプルされた operator について行ふ）



の様なグラフを得てゐる。

この時処理と受付との関係を示す Curve の切線の角度の急に変る点 x_0 を求めることが問題となる。

この点 x_0 をみつけておけば平均的に云つて単位時間内に現れるべき call の operator 一人当りの数が求められた点 x_0 になる様に Operator を配置しておけば現実的に最も能率のよい配置法となると考へられるからである．

x_0 以下になる様に配置すれば無駄に operator をつかつてゐることになり，x_0 以上になる様に配置しておけば service の点で不十分なことになるのであり，又電話料の総収入に減少を来すことになるからである。

観察された結果から x_0 をある信頼度の下に得ようとすることになる。

このために観察された各点はある母集団からのランダム．サムプルと考へることにする。 curve の切線の急にかわる点をみつけるのは困難であり，且観察の結果 curve は二つの直線から成立つてゐると見做せるのでこの立場から x_0 を求めることにした。 即ち二つの推定された回帰直線の交点の信頼度を求めることにした。 このことは事の処理に対して有用な information をあたへることと思ふ。

さて，観察の点はそれぞれの母集団（夫々回帰直線をもつ）からのランダム．サムプルと考へるのである。 この問題は昭和25年11月の講究会に於て樋口伊佐夫氏によつて考へられたものに近くなる。 しかし，ここでは考へ方をかへ，サムプリング調査の Regression Estimate につかふ考へで求めてみようと思ふ。 まづ問題となるのはどの観察点までを一つの回帰線上のものと見做すかの問題である。 我々の場合一まづある点までを含め回帰直線をつくり次の点がこの回帰線をもつ母集団からの一つのランダム．サムプル（等しい確率）と見做せるか否かの検定を行ひ棄却せられればもう一つの母集団のものと一応考へてみる。

この様な検定を行つて観察点の帰属問題を決定した。 幸にして我々の場合この立場からすべての点の帰属はきめられた。

(註)、孰れに属するともきめられない時はその点を一応除外するのも一案である。

さて，各二つの回帰直線をもつと仮定された母集団についてさ

らに次の仮定を設けよう。

① 母集団は一応無限母集団とする。

② 回帰直線を

$$\tilde{y} = \beta x + \alpha, \quad \alpha = \overline{Y} - \beta\overline{X}, \text{（ここに}\overline{X}, \overline{Y}\text{は母集団平均）}$$

とする。 各点は

$$y_i = \beta x + \alpha + e_i$$ であらはされると考へる。

$$\text{ここに } e_i \text{ は } \begin{cases} E(e_i) = 0 \\ E(e_i^2) = \sigma^2 \quad (i, x_i \text{ に関係せぬ}) \\ E(e_i e_j) = 0 \quad (i \neq j) \end{cases} \text{ なる}$$

確率変数である。

なほ二つの母集団では夫々

$$\tilde{y}(1) = \beta_1 x(1) + \alpha_1, \quad \alpha_1 = \overline{Y}_1 - \beta_1\overline{X}_1$$

$$\tilde{y}(2) = \beta_2 x(2) + \alpha_2 \quad \alpha_2 = \overline{Y}_2 - \beta_2\overline{X}_2$$

確率変数は $e(1)$, $e(2)$ とする。

かうすると回帰直線の交点 x_0 は

$$x_0 = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2}$$

として求められる。

さて我々の場合 β_1, β_2, α_1, α_2 はわからない。 サムプルの値から推定された夫々の値 b_1, b_2, a_1, a_2 を知るのみである。 この時 x_0 の推定値 x_0'，その信頼度は如何に求められるであらうか。 これを次に考へてみることにする。

二つの母集団での議論は全く平行に行はれるのでサフィックスを落して論ずることにする。

サムプルの値から回帰係数，常数を推定する問題を考へる。

$$\tilde{y} = bx + a$$

こゝに

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

である。 nはサムプル数 x_i, y_i は各サムプルの受付数，処理数をあらはす。

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

さてこゝで

$$y_i = \beta x_i + \alpha + e_i$$

βは母集団回帰係数
$\alpha = \bar{Y} - \beta\bar{X}$,
e_i は $E(e_i) = 0$, $E(e_i^2) = \sigma^2$,（x_iによらぬ）
$E(e_i e_j) = 0 \ (i \neq j)$
なる確率変数

と考へ，まづ $x_1, \cdots\cdots, x_n$ を固定して待望値をとり，次に $x_1, \cdots\cdots, x_n$ をうごかして待望値をとると

$$E(b) = \beta, \qquad E(a) = \alpha$$

なることがわかる。

b及びaはβ，α の偏りのない推定値となつてゐる。

さて x_0' を求める前に，b 及び a の分散及び相関係数を求めておく必要がある。

上と同様の方法によつて計算をすると容易に次のことがわかる

$$\sigma_b^2 = E(b - \beta)^2$$

$$= E\left(\frac{\sum_{i=1}^{n}(e_i-\bar{e})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^2$$

$$= E\,\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \fallingdotseq \frac{\sigma^2}{(n-1)\,\sigma_x^{\,2}}$$

$$= \frac{(1-\rho^2)\,\sigma_y^{\,2}}{\sigma_x^{\,2}}\cdot\frac{1}{n-1} \fallingdotseq \frac{1}{n}(1-\rho^2)\,\frac{\sigma_y^{\,2}}{\sigma_x^{\,2}}$$

但し $\bar{e} = \dfrac{\sum_{i=1}^{n} e_i}{n}$

又 $E(e^2) = \sigma^2 = (1-\rho^2)\,\sigma_y^2$ は明らか

$$\sigma_a^{\,2} = E\,(a-\alpha)^2$$

$$= E\left(\bar{e} - \bar{x}\,\frac{\sum_{i=1}^{n}(e_i-\bar{e})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^2$$

$$= E\,\frac{\sigma^2}{n}\left(1+\frac{n\cdot\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$\fallingdotseq \frac{(1-\rho^2)\,\sigma_y^{\,2}}{n}\left(1+\frac{n\,E\,\bar{x}^2}{E\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$\fallingdotseq \frac{(1-\rho^2)\,\sigma_y^{\,2}}{n}\left(1+\frac{\frac{1}{n}\sigma_x^{\,2}+\bar{X}^2}{\sigma_x^{\,2}}\right)$$

$$= \frac{(1-\rho^2)\,\sigma_y^{\,2}}{n}\left(1+\frac{\bar{X}^2}{\sigma_x^2}+\frac{1}{n}\right)$$

$$\sigma_a\,\sigma_b\,\rho_{ab} = E\,(b-\beta)(a-\alpha)$$

$$\fallingdotseq -\frac{(1-\rho^2)\,\sigma_y^{\,2}}{n\cdot\sigma_x^{\,2}}\,\bar{X}$$

こゝに $\dfrac{\sigma_x^2}{\bar{x}} \gg 1$ であれば近似的に $\rho=-1$ となる！

である。

さて，以上の考察を用ひ x_0 の推定として

$$x_0' = \frac{a_2-a_1}{b_1-b_2} = \frac{A}{B}$$

をとることにしよう。 suffix は第一，第二の母集団の推定量である。（意味は上述の通り）。この時

$$E(B) = \beta_1 - \beta_2$$

$$E(A) = \alpha_2 - \alpha_1$$

となるが勿論 $E(x_0') \neq x_0$ ではない。 x_0' は x_0 の ratio estimate の形になつてゐるのである。

x_0' の平均自乗誤差をとつてみると

$$\tau^2 = E(x_0' - x_0)^2$$
$$\fallingdotseq \frac{E(A)^2}{E(B)^2}\left(\frac{\sigma_A^2}{E(A)^2} + \frac{\sigma_B^2}{E(B)^2} - 2\rho_{AB}\frac{\sigma_A\sigma_B}{E(A)E(B)}\right)$$

となる。

$$\sigma_A^2 = \sigma_{a_2}^2 + \sigma_{a_1}^2, \quad \sigma_B^2 = \sigma_{b_1}^2 + \sigma_{b_2}^2,$$

$$\rho_{AB}\sigma_A\sigma_B = E(A-\bar{A})(B-\bar{B}) = -\sigma_{a_1}\sigma_{b_1}\rho_{a_1b_1} - \sigma_{a_2}\sigma_{b_2}\rho_{a_2b_2}$$

であるから

$$\tau^2 = \frac{E(A)^2}{E(B)^2}\left\{\left(\frac{\sigma_{a_1}^2}{E(A)^2} + \frac{\sigma_{b_1}^2}{E(B)^2} + 2\rho_{a_1b_1}\frac{\sigma_{a_1}\sigma_{b_1}}{E(A)E(B)}\right)\right.$$
$$\left. + \left(\frac{\sigma_{a_2}^2}{E(A)^2} + \frac{\sigma_{b_2}^2}{E(B)^2} + 2\rho_{a_2b_2}\frac{\sigma_{a_2}\sigma_{b_2}}{E(A)E(B)}\right)\right\}$$

これに計算した値を suffix をつけて入れ、ば平均自乗誤差をうる。これを念のため書き込んでみると

$$\tau^2 = \frac{(\alpha_2-\alpha_1)^2}{(\beta_1-\beta_2)^2} \sum_{i=1}^{2} \left\{ \frac{\frac{1}{n_i} \frac{\sigma_i^2}{\sigma_{x_i}^2}}{(\beta_1-\beta_2)^2} + \frac{\frac{\sigma_i^2}{n_i}\left(1+\frac{\overline{X}_i^2}{\sigma_{x_i}^2}+\frac{1}{n_i}\right)}{(\alpha_2-\alpha_1)^2} \right.$$

$$\left. -2\,\frac{\frac{1}{n_i}\frac{\sigma_i^2}{\sigma_{x_i}^2}\overline{X}_i}{(\beta_1-\beta_2)(\alpha_2-\alpha_1)} \right\}$$

$$= \frac{(\alpha_2-\alpha_1)^2}{(\beta_1-\beta_2)^2} \sum_{i=1}^{2} \frac{\sigma_i^2}{n_i} \left(\frac{\frac{1}{\sigma_{x_i}^2}}{(\beta_1-\beta_2)^2} + \frac{\left(1+\frac{\overline{X}_i^2}{\sigma_{x_i}^2}+\frac{1}{n_i}\right)}{(\alpha_2-\alpha_1)^2} - 2\,\frac{\frac{\overline{X}_i}{\sigma_{x_i}^2}}{(\beta_1-\beta_2)(\alpha_2-\alpha_1)} \right)$$

今もし $\dfrac{\overline{X}_i}{\sigma_{x_i}} \gg 1$ であれば

$$\fallingdotseq \frac{(\alpha_2-\alpha_1)^2}{(\beta_1-\beta_2)^2} \sum_{i=1}^{2} \frac{(1-\rho_i^2)\sigma_{y_i}^2}{n_i\,\sigma_{x_i}^2} \left(\frac{1}{(\beta_1-\beta_2)} - \frac{\overline{X}_i}{(\alpha_2-\alpha_1)} \right)^2$$

となる。

又 $E(A) = E(B)$ になる様に Scale が與へてあれば

$$\tau^2 = \sum_{i=1}^{2} \frac{(1-\rho_i)^2\sigma_{y_i}^2}{n_i} \left\{ \frac{1}{\sigma_{x_i}^2} + \left(1+\frac{\overline{X}_i^2}{\sigma_{x_i}^2}+\frac{1}{n_i}\right) - 2\,\frac{\overline{X}_i}{\sigma_{x_i}^2} \right\}$$

$$n_1 = n_2 = n$$
$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$
$$\sigma_{x_1}^2 = \sigma_{x_2}^2 = \sigma_x^2$$

$$\tau^2 \fallingdotseq \frac{\sigma^2}{n} \left\{ \sum_{i=1}^{2} \left(\frac{1}{\sigma_x^2} + \left(1+\frac{\overline{X}_i^2}{\sigma_x^2}\right) - 2\,\frac{\overline{X}_i}{\sigma_x^2} \right) \right\}$$

$\dfrac{\overline{X}_i}{\sigma_x} \gg 1$ ならば

$$\tau^2 \fallingdotseq \frac{\sigma^2}{n\sigma_x^2} \sum_{i=1}^{2} (1-\overline{X}_i)^2$$

$$\frac{\overline{X}_i}{\sigma_x} \ll 1 \quad ならば$$

$$\tau^2 \doteqdot \frac{\sigma^2}{n}\sum_{i=1}^{2}\left(1+\frac{1}{\sigma_x^2}-2\frac{\overline{X}_i}{\sigma_x^2}\right)$$

となる。

さて x_0' の信頼巾は

$$P_r\left\{\,|x_0'-x_0|>(k+1)\tau\,\right\}\leqq 1-w(k)$$

但し　信頼度　$1-w(k)=95\%$　以上で、$k=3$ でまづ安全に與へられるであらう。

又調査を企画する場合は τ^2 の式から

$$\frac{\tau}{x_0} < \varepsilon \qquad (εはある定めた整数)$$

をつくりサムプル数 n_1, n_2 を決定すればよい。

なほこの理論に設けた仮定は果して妥当なものであらうか。この方法によつて定められた x_0 は意味のあるものであらうか。このことは理論式ではきまらぬものである。如上の validity 性は実際に移されて後始めてたしかめられるものである。

以上は類似の問題の参考としてのべたものである。

ON "MR. OGAWA'S CRITICISM ON HAYASHI'S NORMALITY TEST"

CHIKIO HAYASHI

Let x_i be a independent random variable ($i = 1, 2, \cdots n$) which has a normally distributed density function $N(m, \sigma)$.

Then if the following tranformation is used, the parts he pointed out to be improper will be solved.

$y_1, y_2, \cdots y_{n-1}$ are transformed random variables.

$$
\left.
\begin{aligned}
y_1 &= \frac{1}{\sqrt{2}}(x_2 - x_1) \\
y_2 &= \sqrt{\frac{2}{3}}\left(x_3 - \frac{x_1+x_2}{2}\right) \\
&\vdots \\
y_i &= \sqrt{\frac{i}{i+1}}\left(x_{i+1} - \frac{x_1+\cdots+x_i}{i}\right) \\
&\vdots \\
y_{n-1} &= \sqrt{\frac{n-1}{n}}\left(x_n - \frac{x_1+\cdots+x_{n-1}}{n-1}\right)
\end{aligned}
\right\}
$$

Then y_i is easily proved to be normally distributed and mutually in dependent.

Of course the mean is 0 and the variance is σ^2.