Kanziro SAIO : Some consideration when environmental effects of observations of bio-individual characteristics are dependent upon each other — particularly concerned with variance and covariance (With English summary, p. 140) Japan. J. Breeding Vol. 16, No. 2. 1966

Synopsis. H. F. SMITH has suggested empirically that the variance of the average of n individual environmental effects is the variance of one effect divided by n^b ($0 \leq b \leq 1$), when these effects are statistically dependent. The appropriateness and the limitation of Smith's assumption is pointed out by considering the relation between the parameter b and the correlation of the usual measure of dependency. The result mentioned above for the variance is also suitable for the covariance when two or more characteristics are observed.

各生物個体形質測定値の間に従属性 のある場合のとりあつかい

—— 特に分散および共分散について ——

齋 尾 乾 二 郎

(東京大学 農学部)

I. 緒 言

有用生物の育種においては、いうまでもなく各個体形質の遺伝効果をすることが重要である。しかし、ほとんどすべての場合、それは環境効果と混合された形で表現される。そこで環境効果の性質をみきわめることは育種上でかなり重要な位置をしめる。そこで、ここではその環境効果についての確率論的ないしは数理統計的考察をし、それなりの結論をみちびこうとするものである。

普通の確率模型では、各個体形質の間の環境効果は独立とするのが普通である。そのような場合には、環境効果について特にいうべき問題はほとんどなにもない。ここでは従属性のある場合を問題にする。そして実際問題として、この従属性を問題にしなければならない場合がときどきおきてくる。林木などをあつかう場合は特にそうであるが、そればかりでなく一般作物でもそういった従属性を問題にしなければならない必要性を充分にもっている。

II. 問題のあらましと焦点

遺伝的に均一な集団を考え、個体測定値の個体単位の分散を E とすると、 n 個体の平均の分散は各個体測定値が統計的に独立であれば当然 E/n であることはいうまでもない。ところが、実際にはそれより大きくなることが多く(特に林木等において)、H. F. SMITH (1938) は経験的にこのような場合 E/n^b ($0 \leq b \leq 1$) のような形になるとした。この b は場合場合によつて、すなわち集団がことなることによつて、ことなるとし、その集団すなわ

ち個体測定値の集まりに固有の一種の統計的母数として存在し、実際にはそれを推定する問題にまでたちいたるわけである。はたして、こうした母数を設定することが妥当であるか、そして分散の場合妥当であつたとしても、個体測定値が2形質またはそれ以上ある場合、それらの共分散においても分散と同じような形が許されるかどうかが問題である。

一応これが是認されたとして、こうした問題がどのように適応されるかを念のためのべておくと、今遺伝的に変異のある集団を考え、1個体単位のその遺伝効果の分散を G とすると、この効果に関しては各個体独立(独立でない場合はそれぞれ特殊な場合として別々に考えてゆけばよいからここでは独立な場合しか考えない)であるとしうから、 n 個体の平均のこの効果の分散は G/n である。したがつて、個体測定値の1個体単位の分散は $V_1 = G + E$ 、 n 個体の平均のそれは H. F. SMITH の仮説を適応して、 $V_n = G/n + E/n^b$ 、したがつて $b \neq 1$ すなわち $b < 1$ であつて、 V_1 と V_n が測定してあれば、この2つの式から G と E とを推定することができる。もつと正確にいえば、 G 、 E に関する2つの方程式の G 、 E の係数の行列の階数が2となるから、 b があらかじめなんらかの方法でしられていれば、 G および E は推定でき、したがつて遺伝力等も推定できることになるわけである。 V_1 と V_n だけでなくもつとことなつた n についても推定値がわかつていけば、 b の推定も同時にできるし、また G および E の最小自乗法的推定等さらによい推定値をうることができることになる。しかし、この場合

1964年4月11日受領

いかなる n に対してもその n に無関係な集団固有の母数としての前述のような形で b が存在し、かつその大きさは $0 \leq b < 1$ であるという仮定のもとにおいてである。

さらに 2 形質が測定されているとき、同じような仮定が許されるとすれば、個体単位での共分散は $W_1 = G_{12} + E_{12}$ 、ここに G_{12} および E_{12} はそれぞれ形質 1 と形質 2 との遺伝共分散および環境共分散である。また、 n 個体平均値の間の共分散は $W_n = G_{12}/n + E_{12}/n^{b_{12}}$ となるはずである。ここに b_{12} はその集団における形質 1 と形質 2 との間の環境共分散に固有な分散のときと同様な意味での母数である。そしてこの場合 $0 \leq b_{12} < 1$ であれば、 G_{12} および E_{12} が推定でき、したがって遺伝相関等も推定できることになるわけである。

さてここでの問題は、環境分散および環境共分散だけを考えればよく、遺伝的なものを考える必要はないことは自明である。そして、平均値の分散が、 E/n にならないのは各個体測定値の間に従属性のあることによるのがあきらかであるから（独立性を仮定しなければ E/n にならないからである）、この従属性を直接表現する概念“相関”とここにおける母数 b とがいかなる関係にあるかをみ、さらに共分散についても同様のことをみるのである。

また、このような環境効果の従属性の問題は、これからのべる方法以外に、助変数“時”をユークリッド空間の“位置”におきかえて、確率過程または時系列の問題として解くことも考えられることをつけくわえておく。

III. 環境分散について

n 個の個体がある。その形質の i 番目の個体の測定値を x_i とする。それはつぎのようにかけるとする。

$$x_i = \mu + e_i$$

そして x_i の期待値 $\mathcal{E}(x_i) = \mu$ であるとする。ここに $\mathcal{E}(\)$ はかつこ内の確率変数の期待値をあらわす。したがって e_i は $\mathcal{E}(e_i) = 0$ であるような確率変数である。そして、

$$\text{var}(x_i) = \mathcal{E}(e_i^2) = E \quad (i=1, \dots, n)$$

とする。ここに $\text{var}(\)$ はかつこ内の確率変数の分散をあらわす。そのとき、標本平均 $\frac{1}{n} \sum x_i$ の分散は x_i が i に関して独立であれば、いうまでもなく E/n であるが、 x_i が独立でない、すなわち e_i が独立でないときはどうであろうか。それはつぎのようになる。

$$\begin{aligned} \text{var}\left(\frac{1}{n} \sum x_i\right) &= \mathcal{E}\left[\left(\frac{\sum e_i}{n}\right)^2\right] \\ &= \frac{1}{n^2} \left[\sum_i \mathcal{E}(e_i^2) + \sum_{i \neq i'} \mathcal{E}(e_i e_{i'}) \right] \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n^2} \left(nE + \sum_{i \neq i'} \rho_{ii'} E \right) = \frac{1}{n^2} [nE + n(n-1)\rho E] \\ &= \frac{1}{n} [1 + (n-1)\rho] \end{aligned}$$

ここに、 $\rho_{ii'}$ は e_i と $e_{i'}$ の相関で（隣接または近くの環境の相似性によるものだから負とはならない）、 ρ は $\rho_{ii'}$ の平均、すなわち

$$\rho = \frac{1}{n(n-1)} \sum_{i \neq i'} \rho_{ii'}$$

である。したがって勿論、 $0 \leq \rho \leq 1$ である。

さて、一方 H. F. SMITH の仮定を適応して、

$$\text{var}\left(\frac{1}{n} \sum x_i\right) = \frac{E}{n^b} \quad (0 \leq b \leq 1) \quad \text{とおけば、}$$

$$\frac{E}{n} [1 + (n-1)\rho] = \frac{E}{n^b}$$

$$\text{したがって、} n^b = \frac{n}{1 + (n-1)\rho}$$

あるいは、

$$b = 1 - \frac{\log [1 + (n-1)\rho]}{\log n} = 1 - O(1)$$

したがって n がある程度以上大きければ b は n に独立な値と考えてもさしつかえなからう。また、

$$n^b \rightarrow \frac{1}{\rho} \quad (n \rightarrow \infty)$$

IV. 環境共分散について

2 つの形質の測定値を考える。前同様それはつぎのようにおける。

$$x_{1i} = \mu_1 + e_{1i}$$

$$x_{2i} = \mu_2 + e_{2i}$$

ここで $x_{\alpha i}$ ($\alpha=1, 2$) は i 番目の個体の形質 α の測定値である。また $x_{\alpha i}$ ($\alpha=1, 2$) および $e_{\alpha i}$ ($\alpha=1, 2$) 等の期待値についての仮定は分散のときと同様であることはいうまでもない。そこで n 個の形質 1 の平均値 $\sum_i x_{1i}/n$ と形質 2 の平均値 $\sum_i x_{2i}/n$ との間の共分散を考える。それはつぎのようになる。

$$\begin{aligned} \text{cov}\left(\frac{1}{n} \sum_i x_{1i}, \frac{1}{n} \sum_i x_{2i}\right) &= \mathcal{E}\left(\frac{\sum_i e_{1i}}{n} \cdot \frac{\sum_i e_{2i}}{n}\right) \\ &= \frac{1}{n^2} \left[\sum_i \mathcal{E}(e_{1i} e_{2i}) + \sum_{i \neq i'} \mathcal{E}(e_{1i} e_{2i'}) \right] \\ &= \frac{1}{n^2} (nE_{12} + \sum_{i \neq i'} \rho_{ii'} E_{12}) \\ &= \frac{1}{n^2} (nE_{12} + n(n-1) E_{12} \rho_{12}) \\ &= \frac{E_{12}}{n} [1 + (n-1) \rho_{12}] \end{aligned}$$

ここに、 $\text{cov}(\ , \)$ はかつこ内の 2 つの変数の共分散をあらわし、

$$E_{12} = \mathcal{E}(e_{1i} e_{2i}) \quad (i=1, \dots, n)$$

また, e_{1i} と e_{2i} の相関を ρ^{12} , e_{1i} と $e_{2i'}$ の相関を $\rho_{ii'}^{12}$ とし, $\rho_{ii'}^{12} = \rho_{ii'/12} \rho^{12}$ のようにおくことにより $\rho_{ii'/12}$ を定義すれば,

$$\begin{aligned}\mathcal{C}(e_{1i}e_{2i'}) &= \rho_{ii'}^{12} \sqrt{E_1 E_2} = \rho_{ii'/12} \rho^{12} \sqrt{E_1 E_2} \\ &= \rho_{ii'/12} E_{12}\end{aligned}$$

ただし, $E_\alpha = \text{var}(x_{\alpha i}) = \mathcal{C}(e_{\alpha i}^2)$ ($\alpha=1, 2$)

さらにまた,

$$\rho_{12} = \frac{1}{n(n-1)} \sum_{i \neq i'} \rho_{ii'/12}$$

そしてこの場合1つの個体の形質1と2との相関の絶対値の方が, 相異なる個体のその相関の絶対値より小さいことはないと考えられるから, $|\rho_{ii'}^{12}| \leq |\rho^{12}|$ であつて, またこの両者 $\rho_{ii'}^{12}$ と ρ^{12} とは同符号であると考えられる。したがつて, $\rho_{ii'/12} = \rho_{ii'}^{12} / \rho^{12}$ ($\rho^{12} \neq 0$ のときは $|\rho_{ii'}^{12}| \leq |\rho^{12}|$ だから $\rho_{ii'}^{12}$ も0となるが, そのときは $\rho_{ii'/12} = 0$ と考える)。そしてまたその平均 ρ_{12} は閉区間 $[0, 1]$ の中のいずれかの値をとる。故に分散の場合と同じように

$$\begin{aligned}\text{cov}\left(\frac{1}{n} \sum_i x_{1i}, \frac{1}{n} \sum_i x_{2i}\right) &= \frac{E_{12}}{n^{b+12}} \\ (0 < b_{12} < 1)\end{aligned}$$

とおけば, b_{12} と ρ_{12} との間の関係は, 分散の場合の b と ρ との間の関係と全く同様で, H. F. SMITH の仮定は分散において妥当であると同様な程度に共分散においても妥当である。

ここで, b_{12} ないしは ρ_{12} が, b_1 と b_2 ないしは ρ_1 と ρ_2 (b_α ないしは ρ_α は形質 α の分散に関連した b ないしは ρ である) と何らかの関係があるかについて考える。一般にいつて, これらの性質からわかるように, たとえば $b_{12} = f(b_1, b_2)$ のようなはつきりした関数関係はない。要するに, b_1 , b_2 や ρ_1 , ρ_2 の値にかかわらず, b_{12} や ρ_{12} の値は存在する。ただ, ρ_1 や ρ_2 の値よりも ρ_{12} の値の方が多くの場合小さいことが予想される。それは, ことなつた形質のことなつた個体の間の従属性は, 同じ形質のことなつた個体の間の従属性より小さいと考えられるからである。そしてまた, 一般に関数関係はないと

いつたが, 勿論条件がくわれば関数関係が生じてくるのは当然である。たとえば, e_{1i} と $e_{2i'}$ との e_{2i} を固定したときの部分相関を0だとおく (e_{1i} は $e_{2i'}$ と e_{2i} をとおしてだけしか関係していない) と ρ_{12} は ρ_1 と ρ_2 によつて, したがつて b_{12} は b_1 と b_2 によつて表現できる (実際には3次方程式をとくことになる)。しかし, こうした条件がいつも成立するかどうかは何もいえず一例としてあげたにすぎないのであつて, その場合により適当な条件をいれれば ρ_{12} は ρ_1 と ρ_2 によつて表現できるような場合も生じてくる。

V. 摘 要

各生物個体測定値の環境効果に従属性のある場合には, 個体単位のその分散を E としたとき, n 個体平均値の分散は E/n にはならない。H. F. SMITH (1938) は経験的に E/n^b ($0 \leq b \leq 1$) となるとしたが, 従属性を直接表現する概念相関とこの b との関係をあきらかにし, H. F. SMITH の仮定の妥当性あるいは妥当性の限界を示した。また2形質あるいはそれ以上の形質が測定してある場合の共分散についても分散におけることと同様のことを考察し共分散においても分散と同じことがいえることを証明した。

最後に, この問題提起に関しては国立遺伝学研究所酒井寛一博士におうところが多く, また内容の検討に関しては東京大学農学部畑村又好教授におうところが多い。よつて両氏に厚くお礼いたします。

参 考 文 献

- FREEMAN, G. H. 1963. The combined effect of environmental and plant variation. *Biometrics* 19: 273~277.
- SMITH, H. F. 1938. An empirical law describing heterogeneity in the yields of the agricultural crop. *J. Agric. Sci.* 28: 1~23.
- WHITTLE, P. 1962. Topographic correlation, power-law covariance functions, and diffusion. *Biometrika* 49: 305~314.

Summary

When there is dependency among environmental effects, if the variance of an individual unit is E , the variance of n individuals is not to be E/n . H. F. SMITH has suggested empirically that it should be E/n^b ($0 \leq b \leq 1$). In this paper, the author clears up the relation between a parameter of b and the

concept of a correlation directly showing dependency and indicates the appropriateness, or the limit of the appropriateness, of SMITH's assumption. Besides, it also proves that the above-mentioned fact for the variance is fit for the covariance when two or more characteristics are observed.