Kenziro SAIO: The expression by means of a stochastic model on genetic structures of quantitative characters and its extension. Part 1. The stochastic model without the linkage and the epistasis and its extension. (With English summary, p.186) Jap. J. Breeding Vol.14, No.3. 1964

量的形質の遺伝機構における確率模型 による表現とそのとりあつかい

第1報 連鎖および上位性のない場合の確率模型とそのとりあつかい

斎 尾 乾 二 郎
(東京大学農学部)

I. 緒言および摘要

有用植物の育種においてはほとんどの場合収量に代表されるような量的形質をあつかわなければならない。したがって量的形質の遺伝機構を確実にあきらかにすることが非常に重要である。そのためには量的形質を確率変量と考えることに基礎をおく統計遺伝学的手法によることが少くとも現段階のところ最良の方法であつて、実際にいろいろな結果がえられている。しかしそれらの結果は多分に思いつきのもので、基礎があいまいであるものも少なくない。すなわちその結果がどんな仮説をおいて出発しているかが明確でなかつたり、結果のみちびきかたに飛躍があつたりする場合がある。われわれが実際にその結果をつかつて育種に役立てようと思つても、そうした危惧があつてはせつかく結果があつてもそれを役立てるわけにはゆかない。そこで量的形質が確率変量であると考え、もとにもどつてこれらの結果を吟味したり新しい結果をうみだすように努める必要がある。ここではつねに確率模型を基礎において、簡単なものから複雑なものへと発展させていろいろな結果をだしてゆく、したがって新しい結果を生ずることもあるが、この第1報では統計遺伝学ではもつとも普通流布されている仮説、連鎖および上位性のない場合をあつかうので、今までえられている結果を確率論的に整理しなおすのが主眼になる。しかしその過程において新しい結果もえられた。たとえば、遺伝共分散が遺伝分散と同様に、相加的部分と優性による部分とに分れることがあきらかになつた。

II. 遺伝機構をあらわす確率模型

Malécot (1948) の考え方に大体ならうことにするが、必ずしもそのままを採用するとはかぎらない。次の遺伝模型を考える。

$$Y = X + E \quad (1)$$

ここに Y はある集団に属するある個体のある形質の観測値であつて、勿論確率変量である。 X はその個体のその形質に関する遺伝的な構造による効果を表現する確率変量であり、 E はその個体がそうぐうする環境の同じ形質にたいする効果をあらわす確率変量である。この他に X と E との交互作用を考えてもよいが、それをつけ加えてもつけ加えなくてもこれから考えてゆくことに本質的な影響はないから、模型を簡単にするためにしばらくの間その交互作用は考えない。いいかえると E の中から特に交互作用をとりださないで E の中に含めておく。さてここで注意しなければならないのは、 Y 、 E のみならず、 X も確率変量であるということである。一般に集団からある個体を抽出することは、その個体がそうぐうしている環境を抽出するばかりでなく、遺伝構造の集合体から1つの遺伝構造を抽出するのであるから、 X を確率変量と考えることは当然である。それにもかかわらず、必ずしもこの点が明確でなく意識せずに確率変量としてあつかつたり単なる変量(running variable)としてあつかつたりするために論理的混乱をおこしている場合が多い。また確率変量としてあつかつてもそれを意識しないで確率の場を明確にしないでおこる混乱はさらに多い。ひどい場合には、統計的母数(parameter)なのかその推定値なのかさえはつきりしない場合がある。

個体がさらされる環境はそれのもつている遺伝構造には従属しないと考えられるかぎり、 X と E は独立である。そしてこれはほとんどすべての場合真である。なぜならば、個体のもつ遺伝的な要因が積極的に環境を作りだすというようないいまわしが適當である場合でも、遺伝的な要因が作りだした環境の効果というものは遺伝構造にもとづく効果であつて決してここでいう環境の効果ではないからである。したがって、この模型でははじめから X と E とは独立であることを仮定しておく。

1963年12月27日受領

次に E は他の個体のこの効果とは独立であり、またこの期待値は0であることを仮定する。はじめの独立性の仮定が妥当であることは、個体のそうぐうする環境が全く偶然的なものであることによる。期待値0の仮定は変量の適当な平行移動をすれば期待値は0となるから一般性を損傷しない。実はもつと現実的な意味があつて、もともと E は Y を X が表現しきれない残りの部分と考えうるから、期待値に関することは全部 X にもたせて、 E の期待値を0とすることができるのである。

これまでのところはすべての基本だからここにまとめておく。

模型: $Y = X + E$

Y : 個体の形質観測値をあらわす確率変量

X : 遺伝確率変量

E : 環境確率変量

X と E は独立

E は他の個体のそれと独立

$\mathcal{E}(E) = 0$

ここに $\mathcal{E}()$ はかつこ内の確率変量の期待値をあらわす

次に X について考える。簡単であり、それ故にもつとも流布されている模型から吟味をはじめるために、当分の間 X は常染色体遺伝子だけの効果で、性染色体上の遺伝子とか細胞質とかの効果は考えないことにする。上位性がなければ

$$X = \sum_{i=1}^n H_i \quad (2)$$

のようにあらわせる。 H_i はここで考えている形質に与する i 番目の遺伝子座の効果であらわす確率変量で、 i 遺伝子座の遺伝子型確率変量とよぶ。また、この形質は n 個の遺伝子座が関係しているとして、 Σ はこのすべての遺伝子座についての和である。上位性はことなつた遺伝子座間の特殊な効果すなわち交互作用として考えることができるが、これは後にまわしてここではもつとも単純な場合だけを考える。同じ理由から連鎖も存在しないとする。いいかえると連鎖というのは、ことなつた遺伝子座の間の従属性の程度である。これも後に考える。更に H_i を分解してみる。ここでは2倍体だけを考えるが、そうすると

$$H_i = G_i + G_i' + D_i \quad (i=1, 2, \dots, n) \quad (3)$$

とかける。 G_i および G_i' は遺伝子座における対立遺伝子の効果をあらわす確率変量で遺伝子確率変量とよぶ。さらに G_i は雌親からきた遺伝子、 G_i' 雄親からきた遺伝子に関するものとしておく。 D_i は優性効果にとりもなう確率変量で、いいかえると H_i を G_i および G_i' ではあらわしきれない残りの部分すなわち偏差であり、あるいは G_i と G_i' の交互作用とも考えられる。こ

では D_i を優性偏差とよんでおく。

III. 各確率変量の分布

さて、 $Y, X, E, H_i, G_i, G_i', D_i$ はすべて確率変量であるから分布をもつ。そしてその分布がわかればその確率変量のすべての性格を知ることができる。

Y の分布は、 X と E の分布できまるが、 E の分布は多くの場合正規分布と考えてよいであろうが、 E についてはここでは考えない。 X の分布は H_i ($i=1, 2, \dots, n$) のそれできまり、 H_i のそれは G_i, G_i', D_i のそれできまる。あとでわかるようにその中でも G_i, G_i' の分布が基本であるからそれをまず考える。ここに確率変量 G_i, G_i' のとりうる値を遺伝子価 (genic value) とよび、 H_i のそれを遺伝子型価 (genotypic value) とよぶことをつけ加えておく。

G_i の分布は一般に値 g_{ik} を確率 p_{ik} でとる ($k=1, \dots, s_i$) ことができる。ここに $p_{i1} + p_{i2} + \dots + p_{is_i} = 1$ で s_i は遺伝子座における対立遺伝子の数である。これの実際的意味は次の通りである。 i 遺伝子座に関する遺伝子を $A_{i1}, A_{i2}, \dots, A_{is_i}$ としてそれらが集団内に $p_{i1}, p_{i2}, \dots, p_{is_i}$ の確率で存在するとする。そしてある個体が遺伝子 A_{ik} をもつたとき (雌親から遺伝子 A_{ik} をえたとき) 平均的にその形質に対して g_{ik} の効果をもたらすと考えることである。それは、ある個体が雌親から遺伝子 A_{ik} をうけることは G_i が g_{ik} という値をとることと同等で、その確率は p_{ik} だからである。ここではすべての遺伝子座について対立遺伝子が2つの場合すなわち $s_i=2$ ($i=1, \dots, n$) の場合だけを考える。すなわち3またはそれ以上の対立遺伝子に拡張することは比較的簡単であるし、もつとも単純な場合からはじめるという原則にしたがうからである。したがつて G_i の分布は g_{i1}, g_{i2} をそれぞれ $p_{i1}, p_{i2} (=1-p_{i1})$ の確率でとることになる。これを第1表にまとめておく。

Table 1. The distribution of genic random variable G_i (or G_i')

the value of G_i (or G_i')	probability
g_{i1}	p_{i1}
g_{i2}	$p_{i2} = 1 - p_{i1}$

G_i と G_i' はただ雌親からきたか、雄親からきたかのちがいであつて、もとなつて集団は同じであり、したがつてこの遺伝子の存在確率も同じであるから、同じ分布をもつ。 G_i と G_i' の同時分布を考える。今まで見てきたように、 G_i も G_i' も i 遺伝子座にどんな遺伝子が座をしめるかによつてその値がきまるわけであるが、その G_i に関する遺伝子と G_i' に関する遺伝子が何代か前に1つの遺伝子から由来する確率を f と

する。これが Malécot の定義した *coefficient de consanguinité* (血縁係数) である。遺伝子座の 2 つの遺伝子が同じ遺伝子たとえば A_{ik} である確率、すなわち G_i と G_i' がいずれも g_{ik} である確率は、それが 1 つの遺伝子から由来したことによつて同じであつて (これを Kempthorne (1957) は “identity by descent” といっている。ここでは “由来的に同一” と訳しておく) それが A_{ik} である確率は fp_{ik} である。はじめの 1 つの遺伝子が A_{ik} である確率が p_{ik} で由来的に同一である確率ははじめに f ときめ、そしてこれらは独立な事象だからである。次にその 2 つの遺伝子が由来的にではなく、単に偶然機会的に同一であつて (“identity in state” これを “機会的に同一” と訳しておく) それが A_{ik} である確率は $(1-f)p_{ik}^2$ である。由来的に同じでない確率は $1-f$ であつて、2 つとも A_{ik} である確率は p_{ik}^2 だからである。したがつて G_i および G_i' がいずれも g_{ik} である確率は、これらの和すなわち $fp_{ik} + (1-f)p_{ik}^2 = p_{ik}^2 + fp_{ik}(1-p_{ik})$ である。次に遺伝子座の 2 つの遺伝子が一方 (定つた一方) が A_{ik} で他方が A_{il} ($k \neq l$) である確率、すなわち G_i が g_{ik} で G_i' が g_{il} である確率 (あるいは G_i が g_{il} で G_i' が g_{ik} である確率) は由来的に同一であれば勿論異なつた遺伝子ではないのだから、由来的に同一でなくて機会的に A_{ik} と A_{il} である確率だけを考えればよい。由来的に同一でない確率は $1-f$ 、機会的に A_{ik} と A_{il} である確率は $p_{ik}p_{il}$ したがつて $(1-f)p_{ik}p_{il}$ が求める確率である。したがつて G_i , G_i' の区別をしないでどちらか一方が A_{ik} で他の一方が A_{il} である確率はこの 2 倍 $2(1-f)p_{ik}p_{il}$ である。いずれにせよ前にきめたようにすべての遺伝子座に 2 つの遺伝子座しかない場合について G_i と G_i' の同時分布をまとめると第 2 表のようになる。ところで G_i と G_i' の相関係数を計算しよう。 G_i の分散は第 1 表から

$$\text{var}(G_i) = \mathcal{E}(G_i^2) - \mathcal{E}^2(G_i) = p_{i1}g_{i1}^2 + p_{i2}g_{i2}^2 - (p_{i1}g_{i1} + p_{i2}g_{i2})^2 = p_{i1}p_{i2}(g_{i1} - g_{i2})^2 \quad (4)$$

G_i' の分散は G_i のそれと等しい、 G_i と G_i' の共分散は第 2 表から

 Table 2. The joint distribution of G_i and G_i'

the value of (G_i, G_i')	probability
(g_{i1}, g_{i1})	$p_{i1}^2 + fp_{i1}p_{i2}$
(g_{i1}, g_{i2})	$(1-f)p_{i1}p_{i2}$
(g_{i2}, g_{i1})	$(1-f)p_{i1}p_{i2}$
(g_{i2}, g_{i2})	$p_{i2}^2 + fp_{i1}p_{i2}$

$$\begin{aligned} \text{cov}(G_i, G_i') &= \mathcal{E}(G_i G_i') - \mathcal{E}(G_i) \mathcal{E}(G_i') \\ &= (p_{i1}^2 + fp_{i1}p_{i2}) g_{i1}^2 + 2(1-f)p_{i1}p_{i2}g_{i1}g_{i2} \\ &\quad + (p_{i2}^2 + fp_{i1}p_{i2}) g_{i2}^2 - (p_{i1}g_{i1} + p_{i2}g_{i2})^2 \end{aligned}$$

$$= fp_{i1}p_{i2}(g_{i1} - g_{i2})^2 = f \text{var}(G_i) \quad (5)$$

したがつて G_i と G_i' の相関は f である。また G_i と G_i' の相関が f であり、それらの周辺分布 (単独分布) はいずれも第 1 表のようであるとすれば、 G_i と G_i' の同時分布は第 2 表のようになる。ところで、Wright はこれほど確率の場をはつきりさせてはいないが G_i と G_i' の相関を *inbreeding coefficient* (近交係数) と定義したが、この値は実は Malécot の *coefficient de consanguinité* f に他ならない。

H_i および D_i の分布を考える。 G_i , G_i' を固定すると、 H_i および D_i はその値が完全にきまつてしまう。2 倍体だから遺伝子座に入る 2 つの遺伝子がきまつてしまえば、 i 遺伝子座に関するすべてはきまつてしまうからである。そこで i 遺伝子座の 2 つの遺伝子のうちどちらか一方が A_{ik} で他方が A_{il} のとき H_i は $h_{i,(k,l)} = h_{i,(l,k)}$ で D_i は $d_{i,(k,l)} = d_{i,(l,k)}$ をとるとする。 $k \neq l$ なら H_i が $h_{i,(k,l)} [=h_{i,(l,k)}]$ である確率は $2(1-f)p_{ik}p_{il}$ であり、 D_i が $d_{i,(k,l)} [=d_{i,(l,k)} = h_{i,(k,l)} - g_{ik} - g_{il}]$ である確率と等しい。 $k=l$ のときは H_i が $h_{i,(k,k)}$ また D_i が $d_{i,(k,k)}$ である確率は $p_{ik}^2 + fp_{ik}(1-p_{ik})$ である。ここに実際には $h_{i,(k,k)}$ は遺伝子 A_{ik} , A_{il} を同時にもつた個体とその集団で平均的にもつ効果であり、 $d_{i,(k,l)}$ は $h_{i,(k,l)} - g_{ik} - g_{il}$ としてきめることができる。いずれせよこれによつて H_i および D_i の分布がきまる。前のように $s_i = 2$ ($i=1, \dots, n$) のときは、記号を簡単にするために $(1, 1)=1$, $(1, 2)=(2, 1)=2$, $(2, 2)=3$ のように h および d につける脚符を書きあらためれば、分布は第 3 表のようになる。

 Table 3. The distribution of (G_i, G_i') , H_i , and D_i

(G_i, G_i')	H_i	D_i	probability
(g_{i1}, g_{i1})	h_{i1}	$d_{i1} = h_{i1} - 2g_{i1}$	$p_{i1}^2 + fp_{i1}p_{i2}$
(g_{i1}, g_{i2})	h_{i2}	$d_{i2} = h_{i2}$	$2(1-f)p_{i1}p_{i2}$
(g_{i2}, g_{i1})		$-g_{i1} - g_{i2}$	
(g_{i2}, g_{i2})	h_{i3}	$d_{i3} = h_{i3} - g_{i3}$	$p_{i2}^2 + fp_{i1}p_{i2}$

X の分布はここでは連鎖を考えないから互いに独立に分布する H_i ($i=1, \dots, n$) の和の分布として考えることができる。したがつて X が $\sum_i h_{i,(k_i, l_i)}$ である確率は、 H_i が $h_{i,(k_i, l_i)}$ ($i=1, \dots, n$) である確率の積である。

以上 H_i がきまればしたがつて G_i , G_i' がきまれば X はきまるから、大部分の場合には 1 つの遺伝子座だけ考えれば全体のことがいえることがわかる。そこで今後混乱のおきないかぎり遺伝子座の i を省略してかくことにする。

IV. 確率変量 H, G, G', D のとる値の関係

適当な平行移動をすれば

$$\mathcal{E}(H) = (p_1^2 + fp_1p_2)h_1 + 2(1-f)p_1p_2h_2 + (p_2^2 + fp_1p_2)h_3 \quad (6)$$

を0とすることができるから、はじめから $\mathcal{E}(H)=0$ としておいても一般性を損傷しない。このようにまず h_1, h_2, h_3 をきめる。多くの場合実際にこの値を知ることはできないし、また知る必要はない。ただここで考えた模型を基礎にしてその遺伝子座の遺伝子の組合せに対応して H が何らかの値をとることはたしかで、しかも観測値の適当な平行移動でその期待値が0であるようにすることができる。そのようにしてきめた値が h_1, h_2, h_3 であるとはいうものの H は他の G, G', D のように純粋に理論的な存在ではなく、現実の変量として考える。それは、2倍体では1つの遺伝子座に2つの遺伝子が入れば、そしてまたそのときにはじめて、形質を表現しうるからである。だから H のとる値はそれを知ることができるかできないかにかかわらず現実と直接に対処してきめられるべきものである。

このようにして H のとりうる値がきめられたとすると、 G, G', D のとりうる値はこれに関係してきめられる。しかし、 H の値と、ここに与えられた模型からでは一義的にはきまらない。ただこの模型をつくるとき、 H をできるだけ、1つ1つの遺伝子の効果すなわち G および G' だけで表現したく、それでは表現しきれない残りを D とする考え方が、前にふれたとおりその底にあるのである。だからできるだけ D を小さくおさえた。そこでここでは Fisher にしたがって、そしてまたもつとも普通のやり方にしたがって、この D の平方の期待値 $\mathcal{E}(D^2)$ を最小にするように、 G および G' のとりうる値をきめる。すなわち、 g_1, g_2 を h_1, h_2, h_3 で表現する。そして D の値は、 $d_1 = h_1 - 2g_1, d_2 = h_2 - g_1 - g_2, d_3 = h_3 - 2g_2$ としてあらわせる。そこで、

$$\mathcal{E}(D^2) = (p_1^2 + fp_1p_2)(h_1 - 2g_1)^2 + 2(1-f)p_1p_2(h_2 - g_1 - g_2)^2 + (p_2^2 + fp_1p_2)(h_3 - 2g_2)^2 \quad (7)$$

のように $\mathcal{E}(D^2)$ を g_1, g_2 の函数として表現し、 $\partial \mathcal{E}(D^2)/\partial g_1$ および $\partial \mathcal{E}(D^2)/\partial g_2$ の両方とも0とおけば

$$(p_1 + fp_2)(h_1 - 2g_1) + (1-f)p_2(h_2 - g_1 - g_2) = 0 \quad (8)$$

$$(1-f)p_1(h_2 - g_1 - g_2) + (p_2 + fp_1)(h_3 - 2g_2) = 0 \quad (9)$$

これをとけば一つのあらわし方として

$$g_1 = \frac{(p_1 + fp_2)h_1 + (1-f)p_2h_2}{1+f} \quad (10)$$

$$g_2 = \frac{(1-f)p_1h_2 + (p_2 + fp_1)h_3}{1+f} \quad (11)$$

のように g_1, g_2 は h_1, h_2, h_3 で表現することができる。

この際、(8)式の左辺を p_1 倍し、(9)式の左辺を p_2 倍して、それらを加えあわせると $\mathcal{E}(D)$ になり、加えあわされたいずれもが0であることから $\mathcal{E}(D)=0$ がわかる。ところが

$$\mathcal{E}(D) = \mathcal{E}(H) - \mathcal{E}(G) - \mathcal{E}(G') \quad (12)$$

であつて、 $\mathcal{E}(H)=0, \mathcal{E}(G)=\mathcal{E}(G')$ であるから $\mathcal{E}(G)=\mathcal{E}(G')=0$ となることに注意して、(8), (9)をとけば、(10), (11)がえられる。

ここでさらに次のことをつけ加えておく、すなわち、(8)式の左辺は $\mathcal{E}(D|G=g_1)$ で、(9)式の左辺は $\mathcal{E}(D|G=g_2)$ だから、(8), (9)は、

$$\mathcal{E}(D|G=g_1)=0, \mathcal{E}(D|G=g_2)=0 \quad (13)$$

と同等である。ここに記号 $\mathcal{E}(Y|X=x)$ は、 $X=x$ という条件付きの Y の期待値をあらわす。たとえば $\mathcal{E}(D|G=g_1)$ は、 $G=g_1$ を与えると、ベクトル変量 (G, G') は (g_1, g_1) および (g_1, g_2) の値しかとりえない。その確率はいわゆる条件確率で同時確率を $G=g_1$ である確率 p_1 でわればよいから、 (g_1, g_1) である条件確率は $(p_1^2 + fp_1p_2)/p_1 = p_1 + fp_2$ 、 (g_1, g_2) である確率は、 $(1-f)p_1p_2/p_1 = (1-f)p_2$ である。このとき D は、それぞれ $h_1 - 2g_1$ および $h_2 - g_1 - g_2$ をとるから、 D の条件分布は第4表のようである。したがって D の条件期待値は(8)式の左辺となる。(9)式についても同様のことがいえる。

Table 4. The conditional distribution of D given $G=g_1$

(G, G')	D	probability
(g_1, g_1)	$h_1 - 2g_1$	$p_1 + fp_2$
(g_1, g_2)	$h_2 - g_1 - g_2$	$(1-f)p_2$

V. 分散

まず X の分散を考える。ここでは H_i は他の H_j と独立に分布しているとしているから、

$$\text{var}(X) = \text{var}(\sum_i H_i) = \sum_i \text{var}(H_i) \quad (14)$$

したがって H の分散を考えればよい。ところが前節から $\mathcal{E}(H)=0$ だから、 $\text{var}(H^2) = \mathcal{E}(H^2)$ である。そこで $\mathcal{E}(H^2) = \mathcal{E}[(G+G'+D)^2]$

$$= \mathcal{E}(G^2) + \mathcal{E}(G'^2) + 2\mathcal{E}(GG') + 2\mathcal{E}(GD) + 2\mathcal{E}(G'D) + \mathcal{E}(D^2) \quad (15)$$

のように分解できる。 G, G', D の期待値も0であることに注意すれば、2乗の期待値は分散、積の期待値は共分散を与える。そこで $\mathcal{E}(G^2) = \mathcal{E}(G'^2) = \text{var}(G)$ 、(4)(5)式(これは $\mathcal{E}(G)$ がいかなる値であつても成立つ一般式であるから、 $\mathcal{E}(G)=0$ の場合にも勿論成立する)から、

$$\mathcal{E}(GG') = \text{cov}(G, G') = f\text{var}(G) \quad (16)$$

したがって

$$\mathcal{E}(G^2) + \mathcal{E}(G'^2) + 2\mathcal{E}(GG') = 2(1+f)\text{var}(G) \quad (17)$$

次に $\mathcal{E}(GD)$ を考えると,

$$\mathcal{E}(GD) = \mathcal{E}[G\mathcal{E}(D|G)] = \mathcal{E}(0) = 0 \quad (18)$$

ここに $\mathcal{E}(D|G)$ は G が任意特定の値たとえば g であるときの D の条件期待値で前のあらわし方をすれば $\mathcal{E}(D|G=g)$ で一般に g の函数になる。ところが実際には g は g_1 または g_2 で (13) からいずれの場合にも 0 すなわち $\mathcal{E}(D|G)=0$ であることがわかる。したがって (18) が成り立つ。 $\mathcal{E}(G'D)$ も同様に 0 であるから, (15) は次のようになる。

$$\text{var}(H) = 2(1+f) + \text{var}(G) + \text{var}(D) \quad (19)$$

したがって

$$\text{var}(X) = 2(1+f) \sum_i \text{var}(G_i) + \sum_i \text{var}(D_i) \quad (20)$$

をうる。 $\text{var}(X)$ を遺伝分散, $\text{var}(H)$ を遺伝子型分散, そして (20) の右辺の第 1 項を遺伝子分散 (相加的分散) その第 2 項を優性分散と名付ける。(20) 式は遺伝分散が遺伝子分散と優性分散に分割されるというよく知られた結果を示している。さらに X と E は独立であるから

$$\text{var}(Y) = \text{var}(X) + \text{var}(E) \quad (21)$$

で, この $\text{var}(E)$ を環境分散と名付け, 比 $\text{var}(X)/\text{var}(E)$, $2(1+f) \sum_i \text{var}(G_i)/\text{var}(E)$ をそれぞれ広義の遺伝力および狭義の遺伝力といったり, 特に後者だけを遺伝力といったりしていることは周知のことであろうが, これらの推定の問題はここでは考えない。ただこのような定数が少くともここに採用した条件 (上位性, 連鎖がない) においては確率論的厳密さをもつて意味のあることがわかる。

VI. 共分散

2つの形質の共分散を考える。ここでは連鎖を考えていないから, 2つの形質の遺伝確率変量の共分散の原因となるものは多面発現だけである。この2つの形質に関与するすべての遺伝子座を考え, それぞれの H を $H_i^{(1)}$, $H_i^{(2)}$ とかく。すなわち $H_i^{(1)}$ は形質 (1) の i 番目の遺伝子座に関与する確率変量で, $H_i^{(2)}$ は形質 (2) のそれである。同様に遺伝確率変量に対しても $X^{(1)}$, $X^{(2)}$, G その他についてもそれにならう。すなわち形質 (1) の観測値を $Y^{(1)}$, (2) のそれを $Y^{(2)}$ とすると,

$$\left. \begin{aligned} Y^{(1)} &= X^{(1)} + E^{(1)}, & Y^{(2)} &= X^{(2)} + E^{(2)} \\ X^{(1)} &= \sum_i H_i^{(1)}, & X^{(2)} &= \sum_i H_i^{(2)} \\ H_i^{(1)} &= G_i^{(1)} + G_i^{(1)'} + D_i^{(1)} \\ H_i^{(2)} &= G_i^{(2)} + G_i^{(2)'} + D_i^{(2)} \end{aligned} \right\} \quad (22)$$

のようにかく。ここでも前と同様混乱のないかぎり脚符の i を省略する。さてここで $G^{(1)}$ と $G^{(2)}$ (または $G^{(1)'}$ と $G^{(2)'}$) とは同じ遺伝子が異つた2つの形質 (1) および (2) に対しての効果に関する確率変量だから, どちらかが固定されれば片方の条件分布は退化してしまう。すなわち唯一つの値を確率1でとるようになってしまふ。それからまた, ここでは2つの形質に関与するすべての遺伝子座を考えているからある, 遺伝子座は一方だけの形質に関与して他には無関係であるかもしれない。たとえば遺伝子座は形質 (1) にだけ関与して形質 (2) に関与しないとすれば, $H_i^{(2)} = G_i^{(2)} = G_i^{(2)'} = D_i^{(2)} = \delta(0)$ と考えればよい。ここに $\delta(0)$ はいわゆるディラクのデルタで0を確率1でとる確率変量である。

$$\begin{aligned} \text{さて } \mathcal{E}(H^{(1)}) &= \mathcal{E}(H^{(2)}) = 0 \text{ としてあるから,} \\ \text{cov}(H^{(1)}, H^{(2)}) &= \mathcal{E}(H^{(1)}H^{(2)}) = \mathcal{E}[(G^{(1)} + G^{(1)'} + D^{(1)})(G^{(2)} + G^{(2)'} + D^{(2)})] \\ &= \mathcal{E}(G^{(1)}G^{(2)}) + \mathcal{E}(G^{(1)}G^{(2)'}) \\ &\quad + \mathcal{E}(G^{(1)'}G^{(2)}) + \mathcal{E}(G^{(1)'}G^{(2)'}) + \mathcal{E}(G^{(1)}D^{(2)}) \\ &\quad + \mathcal{E}(G^{(1)'}D^{(2)}) + \mathcal{E}(D^{(1)}G^{(2)}) + \mathcal{E}(D^{(1)}G^{(2)'}) \\ &\quad + \mathcal{E}(D^{(1)}D^{(2)}) \end{aligned} \quad (23)$$

そしてこれら確率変量のとる値を前のあらわし方の右肩に (1) または (2) をつけて形質をあらわすとすれば, $G^{(1)}$ と $G^{(2)}$ の同時分布は前に注意したことから, 第5表のようである。ところで $\mathcal{E}(G^{(1)}G^{(2)}) = p_1g_1^{(1)}g_1^{(2)} + p_2g_2^{(1)}g_2^{(2)}$ であるが $\mathcal{E}(G^{(1)})=0$, $\mathcal{E}(G^{(2)})=0$ を入れて計算すると

$$\mathcal{E}(G^{(1)}G^{(2)}) = sg(g_1^{(1)}g_1^{(2)})\sqrt{\text{var}(G^{(1)})\text{var}(G^{(2)})}$$

となる。

ここに $sg(g_1^{(1)}g_1^{(2)})$ は $g_1^{(1)}g_1^{(2)}$ の符号である。次に $G^{(1)}$ と $G^{(2)'}$ の同時分布は, やはり前に注意した $G^{(1)}$ と $G^{(2)}$ との関係 $G^{(1)'}$ と $G^{(2)'}$ との関係を考慮すれば, 第6表のようである。これから, やはり $\mathcal{E}(G^{(1)})=0$, $\mathcal{E}(G^{(2)'})=0$ に注意して計算すれば, $\mathcal{E}(G^{(1)}G^{(2)'}) = sg(g_1^{(1)}g_1^{(2)})f\sqrt{\text{var}(G^{(1)})\text{var}(G^{(2)})}$ となる。したがって $\mathcal{E}(G^{(1)}G^{(2)'}) = sg(g_1^{(1)}g_1^{(2)})f\sqrt{\text{var}(G^{(1)})\text{var}(G^{(2)})} = f\mathcal{E}(G^{(1)}G^{(2)})$ (24)

Table 5. The joint distribution of $G^{(1)}$ and $G^{(2)}$

$(G^{(1)}, G^{(2)})$	probability
$(g_1^{(1)}, g_1^{(2)})$	p_1
$(g_2^{(1)}, g_2^{(2)})$	$p_2 = 1 - p_1$

Table 6. The joint distribution of $G^{(1)}$ and $G^{(2)'}$

$(G^{(1)}, G^{(2)'})$	probability
$(g_1^{(1)}, g_1^{(2)})$	$p_1^2 + fp_1p_2$
$(g_1^{(1)}, g_2^{(2)})$	$(1-f)p_1p_2$
$(g_2^{(1)}, g_1^{(2)})$	$(1-f)p_1p_2$
$(g_2^{(1)}, g_2^{(2)})$	$p_2^2 + fp_1p_2$

$\mathcal{E}(G^{(1)}G^{(2)}) = \mathcal{E}(G^{(1)'G^{(2)'})$, $\mathcal{E}(G^{(1)}G^{(2)'}) = \mathcal{E}(G^{(1)'G^{(2)})}$
であるから (23) 式のはじめの4つの項の和は

$$2(1+f)\mathcal{E}(G^{(1)}G^{(2)}) \\ = 2(1+f)\text{cov}(G^{(1)}, G^{(2)}) \quad (25)$$

次に $\mathcal{E}(G^{(1)}D^{(2)})$ は勿論 $G^{(1)}$ と $D^{(2)}$ との同時分布から求めることができるが、もつと簡単に、 $G^{(2)}$ を固定すれば $G^{(1)}$ は退化した確率変量になるから、 $G^{(2)}$ を固定した $G^{(1)}$ と $D^{(2)}$ との部分相関 $\rho(G^{(1)}, D^{(2)}|G^{(2)}) = 0$ である。ところがよく知られているように、

$$\rho(G^{(1)}, D^{(2)}|G^{(2)}) = \frac{\sqrt{1-\rho(G^{(1)}, G^{(2)})} \sqrt{1-\rho(D^{(2)}, G^{(2)})}}{\rho(G^{(1)}, D^{(2)}) - \rho(G^{(1)}, G^{(2)}) \rho(D^{(2)}, G^{(2)})} \quad (26)$$

したがって

$$\rho(G^{(1)}, D^{(2)}) = \rho(G^{(1)}, G^{(2)}) \rho(D^{(2)}, G^{(2)}) \quad (27)$$

ところが (18) 式から $\mathcal{E}(D^{(2)}G^{(2)}) = \text{cov}(D^{(2)}, G^{(2)}) = 0$ したがって $\rho(D^{(2)}, G^{(2)}) = 0$ だから $\rho(G^{(1)}, D^{(2)}) = 0$, 結局 $\mathcal{E}(G^{(1)}D^{(2)}) = 0$, 同様にして $\mathcal{E}(G^{(1)'D^{(2)'}) = \mathcal{E}(D^{(1)}G^{(2)}) = \mathcal{E}(D^{(1)}G^{(2)}) = 0$ である。結局

$$\text{cov}(H^{(1)}, H^{(2)}) = 2(1+f)\text{cov}(G^{(1)}, G^{(2)}) \\ + \text{cov}(D^{(1)}, D^{(2)}) \quad (28)$$

となつて分散の (19) 式に対応する結果がでる。したがって

$$\text{cov}(X^{(1)}, X^{(2)}) = 2(1+f) \sum_i \text{cov}(G_i^{(1)}, G_i^{(2)}) \\ + \sum_i \text{cov}(D_i^{(1)}, D_i^{(2)}) \quad (29)$$

また

$$\text{cov}(Y^{(1)}, Y^{(2)}) = \text{cov}(X^{(1)}, X^{(2)}) + \text{cov}(E^{(1)}, E^{(2)}) \\ = 2(1+f) \sum_i \text{cov}(G_i^{(1)}, G_i^{(2)}) + \sum_i \text{cov}(D_i^{(1)}, D_i^{(2)})$$

$$+ \text{cov}(E^{(1)}, E^{(2)}) \quad (30)$$

となる。ここで大切なことは共分散の場合も $\text{cov}(G^{(1)}, D^{(2)})$ 等 G と D に関する共分散が 0 となることである。

$\text{cov}(X^{(1)}, X^{(2)})$ は遺伝共分散 $\text{cov}(H^{(1)}, H^{(2)})$ を遺伝子型共分散、そして (29) 式の第1項を遺伝子共分散 (相加的共分散) 第2項を優性共分散と名付けることは分散と同様である。また、 $\text{cov}(E^{(1)}, E^{(2)})$ を環境共分散と名付ける。 $\text{cov}(X^{(1)}, X^{(2)}) / \sqrt{\text{var}(X^{(1)}) \text{var}(X^{(2)})}$ を遺伝相関、

$$\sum_i \text{cov}(G_i^{(1)}, G_i^{(2)}) / \sqrt{\sum_i \text{var}(G_i^{(1)}) \sum_i \text{var}(G_i^{(2)})}$$

を遺伝子相関 (相加的遺伝相関)、

$$\text{cov}(E^{(1)}, E^{(2)}) / \sqrt{\text{var}(E^{(1)}) \text{var}(E^{(2)})}$$

を環境相関等と名付けることは、よく知られたことである。遺伝力と同様ここではこれらの推定ことは問題にしない。

なお第2報以下では主として連鎖を問題にするはずである。

引用文献

- KEMPTHORNE, O. 1957. An introduction to genetic statistics: Wiley, New York.
MALÉCOT, G. 1948. Les mathématiques de l'hérédité: Masson et Cie, Paris.
WRIGHT, S. 1921. Mating system, I~V: Genetics 6: 111-178.

Summary

In order to apply statistical genetics to actual breeding, we should describe strictly genetic structures of quantitative characteristics based on the probability theory, which is the purpose of this report. But the starting point must be from a single stochastic model. Then, in Part 1 of the series of reports, the linkage and the epistasis are

excluded from consideration. Consequently, some results which have been well known are proved perfectly. In addition some new results are given, for instance the genetic covariance is partitioned into the genic (additive) and the dominant one, just as the case of the variance.