Theoretical and Applied Genetics 40, 76—87 (1970)

On the Determination of Relative Fitness from Frequencies of Genotypes in Subsequent Generations*

Part II. Statistical Problems in Estimation of Fitnesses

R. J. LORENZ

Institut für Biologie der Universität Tübingen, Lehrstuhl für Genetik, und Bundesforschungsanstalt für Viruskrankheiten der Tiere, Tübingen (Germany, BRD)

Summary. In part I of this paper (Theoret. Appl. Genetics 40, p. 11—17) an algebraic treatment of reproduction and selection processes in populations of diploids was given. Here the statistical properties of a special fitness estimator, which is due to Hayman, are considered. Confidence intervals and tests of significance concerning hypotheses about fitnesses are established and checked by computer simulation studies in part III.

The genotype frequencies $f_n^{(i)}$ and $f_{n+1}^{(i)}$ are not known exactly in real experiments, but estimated by means of sampling. This is also true for the parameters ϑ_i (rate of outcrossing, recombination frequency etc.) which appear in Φ, the transition operator. Therefore, $f_n^{(i)}$ and $f_{n+1}^{(i)}$ are random variables, while φ_i is a function of random variables (for meaning of symbols see part one). According to (1.44) or (1.44.a) — see part one of this paper — w_i is itself a random variable. What can be said about the precision of the fitness estimates?

This question obliges to distinguish between the true (but unknown) fitness of a genotype and its estimate. We arrive at the following problem: Given a numerical difference of the fitness estimates of two genotypes, as calculated according to (1.44) or (1.44.a) from observed genotype frequencies, would it be correct to conclude that the corresponding true fitnesses differ? And if so, what can be inferred about the amount of the true difference?

There are a lot of examples which furnish evidence that distinct genotypes (or types of gametes) differ in their (true) fitnesses with respect to the same environment. Today this kind of behavior of a population is accepted to be the rule rather than the exception. Nevertheless it seems to the author that in many cases this experience is based more on repeated observations of the trend of this behavior than on stringent tests of significance. The general difficulties of making such a significance test were mentioned in the introduction of part one. This implies that studies of important aspects which depend on judging of single experiments (as for example the dependence of natural selection on locality and season) are limited by the lack of such tests.

Populations which were investigated by Allard and coworkers (lima bean, barley) retain an amount of genotypic variability after many generations which can hardly be explained by outcrossing (2% in barley, 5—8% in lima bean). It was postulated therefore that heterozygotes are favoured by natural selection. Fitness estimates were calculated according to (1.44) and (1.45). The proportion of random outcrossing, t, however, was estimated by means of a separate experiment and the estimate assumed to be equal to the true value. Thus, in (1.45), t is used as a constant parameter instead of a random variable. Under this condition the fitness estimators were presented as maximum likelihood estimators. Furthermore, the variance-covariance matrix was specified (Allard and Workman, 1963).

With respect to the details of experimentation the use of t as a parameter instead of a random variable implies the neglect of a considerable source of sampling fluctuation of the total fitness estimates. Apart from this, information is needed about the joint sampling distribution of the fitness estimates in order to utilize fully the variance-covariance matrix (e.g. for construction of tests of significance). With reference to the m.l.-property this means information about the speed of convergence towards normally distributed random variables. In other words: What size of sampling is required for "sufficient" normality of the joint distribution of fitness estimates? The chief point is that errors of the first and second kind associated with confidence statements and (formally executed) tests of significance may deviate from their nominal values if the condition of normality is not satisfied.

Confidence intervals and tests of significance have not yet been established by Allard. What follows is an attempt to supplement the extensive studies of Allard's group with respect to the statistical aspect of the problem.

---

* Gekürzte Fassung einer der Mathem.-naturwiss. Fakultät der Universität Tübingen eingereichten Dissertationsschrift (Teil II und III).

We refer to the experiments as described by Allard and Workman (1963). It should be added, however, that the sample size N_1 for estimation of the proportion of outcrossing was small as compared with the sample size N for estimation of genotype frequencies in generations n and n + 1. In general, N_1 was 200—500 seeds, and even 150 seeds were used in some cases (Wöhrmann, Jana, pers. commun.), whereas N varied from 2500 to 5000.

The quantity t was then estimated by the ratio

$$\hat{t} = \frac{\hat{H}}{\hat{p}}, \tag{2.1}$$

where $\hat{H}$ is the observed proportion of heterozygote individuals among the progeny of the N_1 recessive seeds, and $\hat{p}$ the estimated genotype frequency in generation n (estimated from a sample of size N). The variance of $\hat{t}$ is

$$\operatorname{var}(\hat{t}) = \left(\frac{1}{p}\right)^2 \frac{H(1-H)}{N} + \left(\frac{H}{p^2}\right)^2 \frac{p(1-p)}{N} - \frac{f_n^{(2)}}{4N}\left(\frac{t}{p}\right)^2. \tag{2.2}$$

The last term on the right-hand must be added because the gene frequency is estimated not by counting of single genes but of genotypes.

The Fitness Estimators

First the estimators corresponding to (1.43) and (1.45) with c = 1 are considered, with all parameters being substituted for by their estimates. Writing (x_n, y_n, z_n) instead of $(\hat{f}_n^{(1)}, \hat{f}_n^{(2)}, \hat{f}_n^{(3)})$ and $(x_{n+1}, y_{n+1}, z_{n+1})$ instead of $(\hat{f}_{n+1}^{(1)}, \hat{f}_{n+1}^{(2)}, \hat{f}_{n+1}^{(3)})$ in order to simplify the notation, we have

$$\left.\begin{aligned}
\hat{w}_x &= \frac{x_{n+1}}{\hat{\varphi}_x} = \frac{x_{n+1}}{\hat{t}\left(x_n + \frac{1}{2}y_n\right)^2 + (1-\hat{t})\left(x_n + \frac{1}{4}y_n\right)},\\
\hat{w}_y &= \frac{y_{n+1}}{\hat{\varphi}_y} = \frac{y_{n+1}}{2\hat{t}\left(x_n + \frac{1}{2}y_n\right)\left(z_n + \frac{1}{2}y_n\right) + \frac{1}{2}(1-\hat{t})\,y_n},\\
\hat{w}_z &= \frac{z_{n+1}}{\hat{\varphi}_z} = \frac{z_{n+1}}{\hat{t}\left(z_n + \frac{1}{2}y_n\right)^2 + (1-\hat{t})\left(z_n + \frac{1}{4}y_n\right)}.
\end{aligned}\right\} \tag{2.3}$$

With reference to (2.1) it is evident that $\hat{t}$ is not independent of (x_n, y_n, z_n), since $\hat{p} = x_n + (1/2)\,y_n$. On the other hand, $\hat{H}$ is independent of (x_n, y_n, z_n). The joint distribution of $x_{n+1}, y_{n+1}, z_{n+1}$ is multinomial, which, for $N \to \infty$, is known to converge to a singular three-dimensional normal distribution. The same holds for x_n, y_n, z_n. The vector $(x_{n+1}, y_{n+1}, z_{n+1})$ is independent of (x_n, y_n, z_n).

The nominator and the denominator in a fitness estimator (2.3) are therefore independent random variables. The denominators are rational functions of random variables, which are asymptotically normal but not independent. The question arises as to whether the fitness estimates (2.3) themselves may be expected to be asymptotically normally distributed under these circumstances.

Ratios and Products of Random Variables

If X and Y have a joint two-dimensional normal distribution with means m_x, m_y, variances σ_x^2, σ_y^2 and coefficient of correlation ϱ, then the density function of $Z = X/Y$ is given by Fieller's formula (1932), which specializes to the density function of Cauchy's distribution for $m_x = m_y = 0$. It is well known that the latter has infinite moments. The same is true for the distribution of Z. According to Fieller this difficulty may be overcome by taking notice of the fact that X as well as Y is restricted to some limited interval of positive values in many applications. This is true in the case of fitness estimates, where X and Y are identified with the nominator and denominator respectively. Here we have $0 \leq X \leq 1$ and $0 \leq Y \leq 1$. Furthermore it may be assumed that a constant $\varepsilon > 0$ exists, such that $0 < \varepsilon \leq Y \leq 1$.

Another supposition in Fieller's consideration is that $\sigma_y \ll m_y$, which may be seen also to hold good in our application for sufficiently large sample size. If m_x and m_y are positive and large compared with σ_x and σ_y respectively, then the predominant part of the mass of the joint distribution of (X, Y) is contained in the interior of an ellipse which in turn is situated within the (++)-quadrant of the system of coordinates. If now the joint (normal) densitiy function is taken to be zero for all (X, Y) outside this ellipse, one may ask about the effect of this curtailment upon the density of Z. It is clear that the moments of Z are now finite. Fieller's conclusion is that the area under the density curve is increased near the mode at the expense of the tails, but that for sufficiently large diameters of the surrounding ellipse this has no visible effect on the appearance of the distribution. For an estimation of the amount of distortion, see the original paper.

Now the moments of the curtailed distribution of Z may be approximated by the method of Merrill (1928). Using the decomposition $X = m_x + \xi$, $Y = m_y + \eta$, where $E(\xi) = E(\eta) = 0$, $\operatorname{var}(\xi) = \sigma_x^2$, $\operatorname{var}(\eta) = \sigma_y^2$, and $\varrho(\xi, \eta) = 0$, we have

$$Z = \frac{m_x + \xi}{m_y + \eta} = \frac{m_x}{m_y}\left[1 + \frac{\xi}{m_x}\right]\left[1 + \frac{\eta}{m_y}\right]^{-1}. \tag{2.4}$$

For $|\eta| < m_y$ this may be expanded to

$$Z = \frac{m_x}{m_y}\left[1 + \frac{\xi}{m_x}\right]\left[1 - \frac{\eta}{m_y} + \frac{\eta^2}{m_y^2} - \frac{\eta^3}{m_y^3} + \cdots\right]. \tag{2.5}$$

The expectation $E(Z^n)$ of the n-th power of this expansion is taken as μ_n', the n-th moment about zero of Z. Merrill retains the products $\xi^r \eta^s$ as far as the eighth order and takes for their mean value the product moments of the normal surface. For the justification of this procedure, see Fieller (1932).

Here the expressions for the first five central moments $\mu_i = E\,[(Z - E(Z))^i]$ of Z are reported. For abbreviation

$$\left.\begin{aligned} &Z_0 = \frac{m_x}{m_y}, \quad v_x = \frac{\sigma_x}{m_x}, \quad v_y = \frac{\sigma_y}{m_y}, \\ &p_2 = v_x^2 - 2\,\varrho\, v_x v_y + v_y^2, \\ &\lambda = \frac{v_y - \varrho\, v_x}{\sqrt{v_x^2 - 2\,\varrho\, v_x v_y + v_y^2}} = \frac{v_y - \varrho\, v_x}{\sqrt{p_2}} \quad (\lambda < 1)\,. \end{aligned}\right\} \tag{2.6}$$

Then

$$\left.\begin{aligned} \mu_1' &= E(Z) = Z_0\,[1 + v_y\,(v_y - \varrho\, v_x) \\ &\quad \times (1 + 3\,v_y^2 + 15\,v_y^4 + 105\,v_y^6)] \\ \mu_2 &= Z_0^2\, p_2\,[1 + v_y^2\,(5\,\lambda^2 + 3) + v_y^4\,(54\,\lambda^2 + 15) \\ &\quad + v_y^6\,(591\,\lambda^2 + 105)] \\ \mu_3 &= Z_0^3\, p_2^{3/2}\, v_y \lambda\,[6 + 4\,v_y^2\,(11\,\lambda^2 + 18) \\ &\quad + v_y^4\,(1044\,\lambda^2 + 828)] \\ \mu_4 &= Z_0^4\, p_2^2\,[3 + 30\,v_y^2\,(3\,\lambda^2 + 1) \\ &\quad + v_y^4\,(633\,\lambda^4 + 2106\,\lambda^2 + 315)] \\ \mu_5 &= Z_0^5\, p_2^{5/2}\, v_y \lambda\,[60 + v_y^2\,(1580\,\lambda^2 + 1380)]\,. \end{aligned}\right\} \tag{2.7}$$

Furthermore it is

$$\begin{aligned} \beta_1 = \frac{\mu_3^2}{\mu_2^3} &= v_y^2 \lambda^2\,[36 - v_y^2\,(12\,\lambda^2 - 540) \\ &\quad + 4\,v_y^4\,(-\,146\,\lambda^4 + 450\,\lambda^2 + 1917)] \\ \beta_2 = \frac{\mu_4}{\mu_2^2} &= 3 + 3\,[v_y^2\,(20\,\lambda^2 + 4) \\ &\quad + v_y^4\,(-14\,\lambda^2 + 404\,\lambda^2 + 42)]\,. \end{aligned}$$

From this we obtain

$$\text{skewness} \qquad \gamma_1 = \sqrt{\beta_1} \tag{2.8}$$

$$\text{excess} \qquad \gamma_2 = \beta_2 - 3 \tag{2.9}$$

The expressions for β_1 and β_2 show that skewness and excess of the distribution of Z depend on v_y and λ only, and that if $v_y \to 0$, μ_3 and μ_5 as well as γ_1 and γ_2 converge to zero.

Merrill has given some conditions under which the distribution of Z will itself be close to the normal. In general this is true for X and Y following the normal law closely, and if v_x and v_y as well as ϱ are small. On the other hand, in a footnote in Merrill's paper, K. Pearson called attention to the observation that comparatively slight deviations from the bivariate normal distribution of X and Y may lead to remarkable divergence from normality in the distribution of Z.

In our application we use another statement on quotients of independent random variables, which is due to Curtiss (1941; theorem 6.1):

Let the independent variables X_α and Y_β have respective distribution functions $F_\alpha(x)$ and $G_\beta(y)$ which depend upon the two parameters α and β. Let $H_{\alpha,\beta}(z)$ be the distribution function of the quotient $Z_{\alpha,\beta} = X_\alpha / Y_\beta$. If there exist two chance variables X and Y with respective distribution functions $F(x)$ and $G(y)$ such that $\lim\limits_{\alpha\to\infty} F_\alpha(x) = F(x)$ at all points of continuity of $F(x)$, and $\lim\limits_{\beta\to\infty} G_\beta(y) = G(y)$, at all points of continuity of $G(y)$, then

$$\lim_{\substack{\alpha\to\infty \\ \beta\to\infty}} H_{\alpha,\beta}(z) = \lim_{\alpha\to\infty}\,\lim_{\beta\to\infty} H_{\alpha,\beta}(z) = \lim_{\beta\to\infty}\,\lim_{\alpha\to\infty} H_{\alpha,\beta}(\beta) = H(z)$$

at all points of continuity of $H(z)$, where $H(z)$ is the distribution function of the variable X/Y.

With regard to products of random variables a theorem of Aroian (1947) should be mentioned:

If X and Y have a joint normal distribution and if $-1 + \varepsilon < \varrho \leq 1$ $(\varepsilon > 0)$, then the distribution of $T = X \cdot Y$ approaches normality for $v_x \to 0$, $v_y \to 0$.

According to Aroian, this theorem may be extended to the case that X and Y are asymptotically normal. As a special case we notice that the square of a (asymptotically) normal variate is asymptotically normal if its coefficient of variation converges to zero.

Application to the Fitness Estimator

With respect to the theorems of Curtiss and of Aroian it follows from (2.3), that $\hat{w}_x$, $\hat{w}_y$, $\hat{w}_z$ and also $\hat{w}_x^* = \hat{w}_x/\hat{w}_y$, $\hat{w}_z^* = \hat{w}_z/\hat{w}_y$ are asymptotically normal for $N \to \infty$, $N_1 \to \infty$. Moreover it has been shown by P. M. Geppert (1969, pers. commun.) that $\hat{w}_x$, $\hat{w}_y$, $\hat{w}_z$ are maximum likelihood estimators under the same condition. On the other hand, in Allard's experiments N_1 is bounded and H is very small (.01 to .1), whereas N is very large as compared with N_1. Under this condition the distribution of $\hat{t}$ is asymptotically binomial for $N \to \infty$, and the denominators in (2.3) are no longer asymptotically normal.

This means that if t is estimated from an experiment of small size (N_1), certainly the fitness estimators themselves are not asymptotically normal for $N \to \infty$.

But if $\hat{t}$ is taken as a constant, the corresponding estimate of the fitness variance does not contain that component which results from $\hat{t}$ being a random variable. The component even increases in proportion to the total variability of the fitness estimate if N increases. Thus it may be speculated that the formulas for the variances of fitness estimates which are used by Allard tend to become unreliable at just the point when N becomes large, provided that N_1 remains small.

In order to study the implications of these restrictions in more detail, a system of computer simulations of the estimation procedure was performed by means of pseudo-random numbers. The program was built up with regard to the following special problems:

1. Evaluation of empirical distribution functions (d.f.) of $\hat{w}_x$, $\hat{w}_y$, $\hat{w}_z$, and also for $\hat{w}_x^* = \hat{w}_x/\hat{w}_y$, $\hat{w}_z^* = \hat{w}_z/\hat{w}_y$ and of the corresponding statistics (mean, variance, skewness, excess).

2. Evaluation of the d.f. s of those normal distributions, which are expected to be the asymptotic distributions.

3. Evaluation of confidence intervals for the true fitness (under normality) and estimation of corresponding probabilities of error. These may differ from the theoretical ones if the normality condition is not satisfied.

4. Evaluation of tests of significance for detecting of fitness differences and calculation of the power functions of these tests for a number of different combinations of genotype frequencies and sample sizes.

Part III. Simulation Experiments

1. Empirical Distribution Functions of $\hat{w}_x, \hat{w}_y, \hat{w}_z, \hat{w}_x^*, \hat{w}_z^*$

Two samples, each of size N, from multinomial distributions $(f_n^{(x)}, f_n^{(y)}, f_n^{(z)})$ and $(f_{n+1}^{(x)}, f_{n+1}^{(y)}, f_{n+1}^{(z)})$ were generated by means of equally distributed random numbers. These samples represent empirical genotype frequencies (x_n, y_n, z_n) and $(x_{n+1}, y_{n+1}, z_{n+1})$ in generations n and $n+1$. Another set of N_1 equally distributed random numbers was used to produce a sample from a binomial distribution with mean $N_1 H$, where H is determined such that $t = H/p_n$ equals a fixed value given in advance; p_n is predetermined to be $p_n = f_n^{(x)} + (1/2) f_n^{(y)}$. The simulated estimate of t is then given by $\hat{t} = \hat{H}/\hat{p}_n = \hat{H}/(x_n + (1/2)\, y_n)$.

From these numbers the quantities $\hat{\varphi}_x, \hat{\varphi}_y, \hat{\varphi}_z$ and $\hat{w}_x, \hat{w}_y, \hat{w}_z$ are calculated according to (2.3), and the simulation of an individual field experiment is completed. In order to gain a good approximation of the distributions of $\hat{w}_x$, $\hat{w}_y$, and $\hat{w}_z$, 1000 simulations, each with a new set of random numbers, were performed for the same basic genotype frequencies and ratio of outcrossing. The mean and the central moments of order 2, 3, and 4 were estimated for $\hat{w}_x, \hat{w}_y, \hat{w}_z$ from the corresponding sample of 1000 simulated estimates. At the same time a set of 1000 vectors of "reduced" estimates $(\hat{w}_x^*, 1, \hat{w}_z^*)$ were obtained, but no moments were calculated from them.

Different combinations of the parameters $\{f_n^{(i)}\}$ and $\{f_{n+1}^{(i)}\}$ and of t were used to get an impression of how the empirical d.f. s depend upon the parameters. These combinations were selected in the following manner: Different sets of p_n and t are given in advance. For any given p_n, one of the three genotype frequencies in generation n may be taken at will, for example $f_n^{(x)}$. Furthermore, only populations in equilibrium were considered, i.e. $f_{n+1}^{(i)} = f_n^{(i)}$. Under the equilibrium condition the quantities w_x^* and w_z^* (as well as w_x, w_y, w_z) are uniquely determined by $(p_n, t, f_n^{(x)})$.

Four different rates of outcrossing $t = .10$, $.08$, $.06$, and $.04$ were used at two different levels of gene frequency $p_n = .55$ and $.85$. For any of these eight combinations five to eight different genotype frequencies $f_n^{(x)}$ were selected in such a manner that the corresponding values of w_x^* (and w_z^*) form a sequence of decreasing levels. Thus a totality of 51 combinations was obtained (see Table 1).

Furthermore N and N_1 were varied to demonstrate the dependence of the statistics and the tests upon sample size. In the main group of simulations N_1 was taken to equal N, and five different levels were employed for each of the 51 combinations mentioned above: $N = N_1 = 400$, 800, 1200, 1600, 2000. For populations no. 21, 29, and 49 (Table 1) an additional size of 4000 was used.

The assumption $N_1 = N$ was made in order to simulate the asymptotic behavior of the fitness estimates under normality condition. In contrast to this more theoretical case a number of populations (No. 1, 2, 3, 4, 5) were considered also under the condition that N_1 remains fixed at $N_1 = 400$, whereas N varies as before.

To summarize, there are 283 different combinations in all of the parameters $p_n, t, f_n^{(x)}$ and $N(N_1)$. For each of these designs a run of 1000 individual simulations was accomplished.

After accomplishing of any run, from the resulting sample of 1000 vectors $(\hat{w}_x, \hat{w}_y, \hat{w}_z)$ the following statistics were calculated for each component: mean, mean square deviation, skewness, and excess. For the same parameter combination, the corresponding "true" parameters according to Merrill's formulas (2.7) to (2.9) were also calculated. In addition, the Merrill-parameters of $\hat{w}_x^*$ and $\hat{w}_z^*$ were calculated for each parameter combination.

It should be noticed that the comparison of the empirical statistics with their corresponding Merrill-parameters is asymptotically correct only because of the supposition in Merrill's formulas that a quotient's nominator and denominator are strictly normal variates, whereas for small values of N and N_1 the nominators and denominators of the fitness estimators are certainly not. It is to be expected, however, that the statistics approach the parameters for increasing N and N_1.

To calculate the Merrill-parameters, the coefficients of variation of the nominators and denominators are needed. In addition, for $\hat{w}_x^*$ and $\hat{w}_z^*$ the coefficients of correlation between $\hat{w}_x$ and $\hat{w}_y$ and between $\hat{w}_z$ and $\hat{w}_y$ respectively are needed. Variances and covariances of functions of random variables were calculated by the approximate formulae (Cramér 1946)

$$\operatorname{var}(g) \sim \sum_i \sum_j \frac{\partial g}{\partial \mu_i} \frac{\partial g}{\partial \mu_j} \operatorname{cov}(\xi_i, \xi_j) \qquad (2.10)$$

where $g = g(\xi_1, \ldots, \xi_k)$, $\mu_i = E(\xi_i)$, and $\frac{\partial g}{\partial \mu_i} = \left(\frac{\partial g}{\partial \xi_i}\right)_{\mu_i}$,

$$\operatorname{cov}(g, h) \sim \sum_i \sum_j \frac{\partial g}{\partial \mu_i} \frac{\partial h}{\partial \mu_j} \operatorname{cov}(\xi_i, \xi_j) \qquad (2.11)$$

where $h = h(\xi_1, \ldots, \xi_k)$.

Table 1. *Equilibrium populations used for simulation.*
p_n: gene frequency, t: proportion of outcrossing

p_n	t	$f_n^{(x)}$	$f_n^{(y)}$	$f_n^{(z)}$	w_x	w_y	w_z	w_x^*	w_z^*	Pop No.
.55	.10	.500	.100	.400	.9945	1.0582	.9932	.9398	.9385	1
		.480	.140	.380	.9722	1.2444	.9651	.7812	.7755	2
		.450	.200	.350	.9370	1.4337	.9204	.6536	.6420	3
		.400	.300	.300	.8738	1.6260	.8386	.5374	.5157	4
		.300	.500	.200	.7268	1.8215	.6395	.3990	.3511	5
.55	.08	.510	.080	.410	.9965	1.0471	.9956	.9516	.9508	6
		.500	.100	.400	.9858	1.1682	.9823	.8438	.8409	7
		.480	.140	.380	.9639	1.3462	.9548	.7160	.7093	8
		.450	.200	.350	.9294	1.5198	.9110	.6115	.5994	9
		.400	.300	.300	.8673	1.6892	.8306	.5134	.4917	10
.55	.06	.520	.060	.420	.9980	1.0363	.9975	.9631	.9626	11
		.510	.080	.410	.9877	1.1887	.9847	.8309	.8284	12
		.500	.100	.400	.9772	1.3038	.9717	.7495	.7453	13
		.480	.140	.380	.9557	1.4660	.9447	.6519	.6444	14
		.450	.200	.350	.9218	1.6168	.9017	.5702	.5577	15
.55	.04	.530	.040	.430	.9991	1.0256	.9988	.9741	.9739	16
		.525	.050	.425	.9941	1.1416	.9928	.8709	.8697	17
		.520	.060	.420	.9892	1.2346	.9866	.8012	.7992	18
		.515	.070	.415	.9841	1.3109	.9804	.7508	.7479	19
		.510	.080	.410	.9791	1.3746	.9741	.7123	.7087	20
		.500	.100	.400	.9688	1.4749	.9613	.6568	.6518	21
		.490	.120	.390	.9583	1.5504	.9482	.6181	.6116	22
.85	.10	.825	.050	.125	.9988	1.0471	.9921	.9588	.9524	23
		.822	.056	.122	.9968	1.1045	.9787	.9024	.8861	24
		.820	.060	.120	.9954	1.1429	.9697	.8710	.8485	25
		.810	.080	.110	.9887	1.3008	.9224	.7601	.7091	26
		.800	.100	.100	.9815	1.4209	.8730	.6907	.6144	27
		.790	.120	.090	.9750	1.5094	.8163	.6459	.5408	28
		.770	.160	.070	.9610	1.6410	.6914	.5856	.4213	29
.85	.08	.830	.040	.130	.9993	1.0309	.9954	.9693	.9655	30
		.825	.050	.125	.9960	1.1521	.9743	.8645	.8457	31
		.820	.060	.120	.9927	1.2500	.9524	.7942	.7619	32
		.815	.070	.115	.9894	1.3308	.9296	.7435	.6986	33
		.810	.080	.110	.9861	1.3986	.9061	.7051	.6479	34
		.800	.100	.100	.9794	1.5060	.8562	.6503	.5685	35
		.790	.120	.090	.9727	1.5873	.8021	.6128	.5053	36
.85	.06	.833	.034	.133	.9984	1.0870	.9899	.9185	.9107	37
		.830	.040	.130	.9965	1.1730	.9778	.8495	.8336	38
		.825	.050	.125	.9933	1.2887	.9571	.7708	.7427	39
		.820	.060	.120	.9900	1.3793	.9357	.7178	.6784	40
		.815	.070	.115	.9868	1.4523	.9134	.6795	.6290	41
		.810	.080	.110	.9835	1.5123	.8903	.6504	.5887	42
		.800	.100	.100	.9770	1.6051	.8414	.6087	.5242	43
.85	.04	.838	.024	.138	.9986	1.1050	.9918	.9038	.8976	44
		.837	.026	.137	.9980	1.1463	.9879	.8707	.8619	45
		.835	.030	.135	.9968	1.2195	.9804	.8174	.8039	46
		.833	.034	.133	.9955	1.2821	.9726	.7765	.7587	47
		.831	.038	.131	.9943	1.3362	.9648	.7441	.7220	48
		.827	.046	.127	.9918	1.4250	.9487	.6960	.6658	49
		.820	.060	.120	.9874	1.5384	.9195	.6418	.5977	50
		.800	.100	.100	.9745	1.7182	.8271	.5672	.4814	51

Theoretical and empirical values are shown for the runs of population 21 in Table 2. Instand of the mean, the bias is noted there, this being the difference between the expected and the true value in the Merrill case, and the difference between the observed sample mean and the true value in the empirical case.

From Table 2 it is seen that the theoretical and empirical values of the bias and of the standard deviation (s.d.) agree very well, whereas empirical skewness and excess show remarkable fluctuation about their respective parameter values. It is difficult to judge these deviations, since tolerance limits for these statistics are not known under conditions of non-normality. If a comparison is made with the standard error of skewness and excess in samples of normal variates (Cramér 1946), the deviations are found to be of an acceptable order of magnitude. Apart from this, for all statistics a tendency of convergence

Table 2. *Population No. 21: Theoretical and empirical distribution parameters of fitness estimates*

	N	Bias		Standard dev.		Skewness		Excess	
		Merrill	empir.	Merrill	empir.	Merrill	empir.	Merrill	empir.
$\hat{w}_x$	400	.00222	.00138	.06739	.06884	.20118	.32146	.09612	.43739
	800	.00111	.00105	.04750	.04846	.14103	.26937	.04716	.11971
	1200	.00074	.00054	.03874	.04070	.11482	.00594	.03124	−0.24499
	1600	.00055	.00084	.03354	.03442	.09930	−0.03382	.02335	.01125
	2000	.00044	.00043	.02999	.02755	.08874	.02934	.01865	−0.04205
	4000	.00022	.00030	.02119	.02142	.06265	.16765	.00929	−0.01955
$\hat{w}_y$	400	.03022	.03622	.31890	.30971	.66224	.51801	1.01437	.49378
	800	.01463	.01450	.21890	.21657	.43140	.44870	.44454	.15878
	1200	.00966	.01325	.17709	.17663	.34310	.36351	.28244	.30307
	1600	.00721	.00750	.15268	.15233	.29334	.38374	.20661	.21379
	2000	.00575	.00621	.13619	.14495	.26039	.15397	.16278	−0.43242
	4000	.00286	.00291	.09581	.09503	.18142	.14683	.07890	−0.23115
$\hat{w}_z$	400	.00308	.00351	.08069	.08273	.23459	.19559	.13230	.28179
	800	.00153	.00153	.05681	.05731	.16386	.11913	.06445	.02883
	1200	.00102	.00104	.04632	.04812	.13326	.15231	.04259	−0.11952
	1600	.00077	.00028	.04009	.03931	.11518	.20840	.03180	.27217
	2000	.00061	.00052	.03584	.03162	.10289	.17359	.02537	−0.20748
	4000	.00031	.00018	.02532	.02486	.07259	−0.05940	.01262	−0.10134
$\hat{w}_x^*$	400	.03991		.20260		1.86312		5.81713	
	800	.01716		.12289		1.02889		2.07518	
	1200	.01096		.09622		.77843		1.22437	
	1600	.00806		.08172		.65015		.86175	
	2000	.00637		.07228		.56941		.66302	
	4000	.00311		.05003		.38681		.30602	
$\hat{w}_z^*$	400	.03843		.19855		1.81997		5.59516	
	800	.01650		.12072		1.00420		1.99312	
	1200	.01053		.09459		.75952		1.17549	
	1600	.00774		.08037		.63426		.82820	
	2000	.00612		.07109		.55544		.63638	
	4000	.00299		.04923		.37724		.29367	

to the parameter values is recognized with increasing N $(= N_1)$.

The statements made about population 21 hold good in all other cases. For illustration some of the empirical cumulative d.f. s (c.d.f.) of $\hat{w}_x$ and of $\hat{w}_x^*$ from populations 4, 21, 29, and 49 are shown in Fig. 1 for three different sample sizes together with the c.d.f. of the limit normal distribution (with w_x and w_x^* respectively as mean and with s.d. as derived from Merrill's formula). The figures confirm what one would expect from theory: the speed of convergence towards the limit normal distribution is larger in general for $p_n = .55$ as compared with .85, and within these groups in turn it is larger for $t = .10$ than for $t = .04$. On the other hand, the convergence is better for $\hat{w}_x$ than for $\hat{w}_x^*$. The figure for population 49 shows in addition the c.d.f. of $\hat{w}_y$ as an example where the genotype frequency $f_n^{(y)} = f_{n+1}^{(y)}$ is extremely small (.046).

What is of more importance than pure measures of "non-normality" is the possible effect of these deviations on the value of confidence statements and tests of significance, a problem with which we will be concerned in the next sections.

2. *Confidence Intervals*

Approximate confidence limits for the true fitness values were determined according to the procedure of Fieller (1944): Let again X and Y be normal variates with $E(X) = \xi$, $\text{var}(X) = \sigma_x^2$ $E(Y) = \eta$, $\text{var}\,(Y) = \sigma_y^2$, $\text{cov}\,(X, Y) = \sigma_{x,y}$, and $\alpha = \xi/\eta$. Then $a = X - \alpha\, Y$ is normal with $E(a) = 0$ and $\text{var}\,(a) = \sigma_x^2 - 2\,\alpha\,\sigma_{x,y} + \alpha^2\,\sigma_y^2$. Furthermore

$$U = \frac{X - \alpha\, Y}{\sqrt{\sigma_x^2 - 2\,\alpha\,\sigma_{x,y} + \alpha^2\,\sigma_y^2}} \tag{2.12}$$

is normal with $E(U) = 0$, $\text{var}\,(U) = 1$. Thus a constant $u_\varepsilon > 0$ may be selected such that

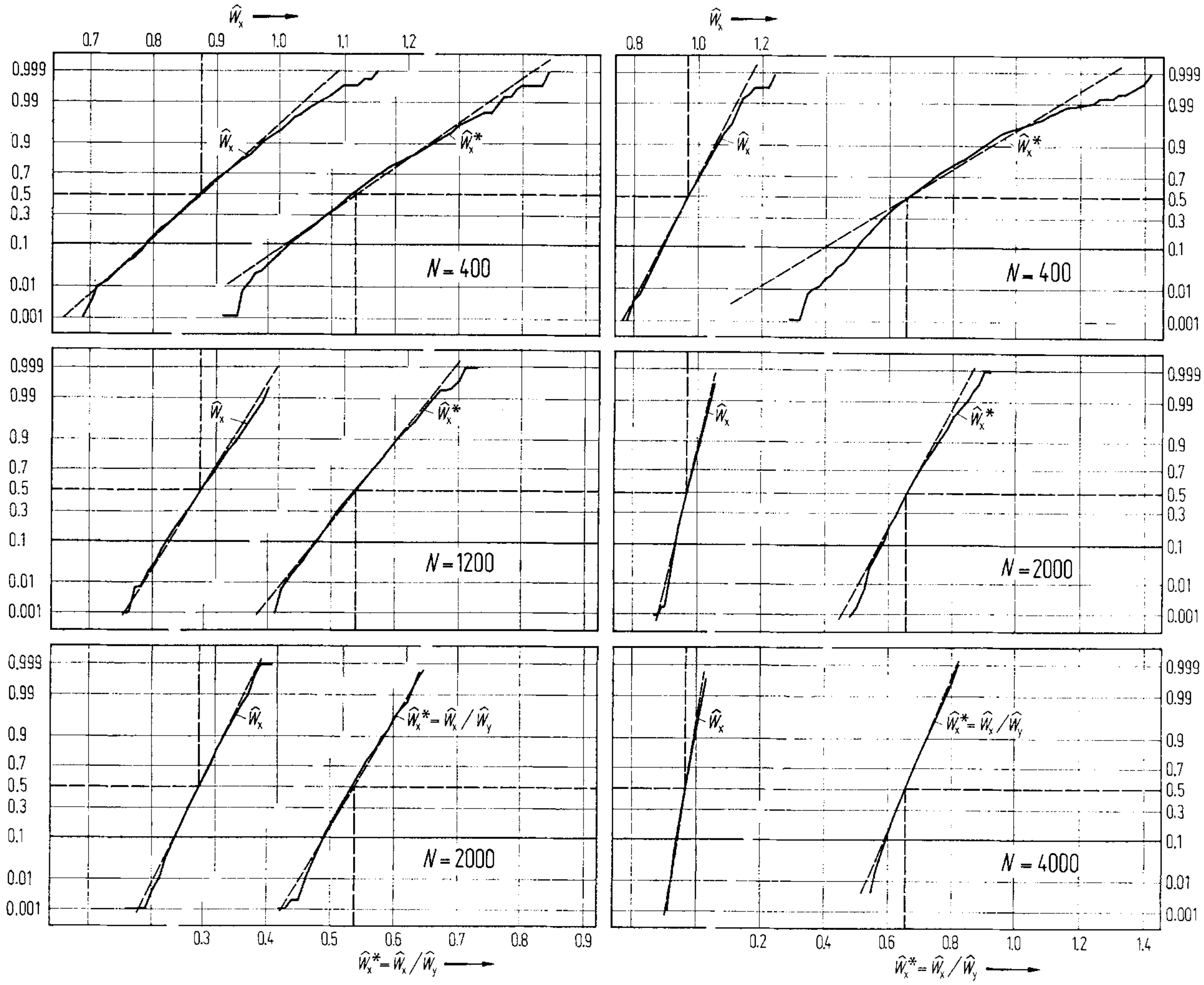
$$\text{Prob}\,\{|u| \leq u_\varepsilon\} = 1 - \varepsilon\,. \tag{2.13}$$

The set of all those values of α which satisfy this equation for given x and y constitute a confidence interval for α with confidence coefficient $1 - \varepsilon$. The upper and lower confidence limits α_u and α_l are found as solutions of

$$(u_\varepsilon^2\,\sigma_y^2 - y^2)\,\alpha^2 - 2\,(u_\varepsilon^2\,\sigma_{x,y} - x\,y)\,\alpha + (u_\varepsilon^2\,\sigma_\alpha^2 - x^2) = 0\,, \tag{2.14}$$

which results from the inequality on the left hand side of (2.13).

To applicate Fieller's theorem to w_x, w_y, and w_z, x is to be replaced by x_{n+1} (y_{n+1}, z_{n+1} resp.), whereas y must be replaced by $\hat{\varphi}_x$($\hat{\varphi}_y$, $\hat{\varphi}_z$ resp.); w_x(w_y, w_z resp.) is to be used for α. The corresponding variances are to be applied accordingly. The covariances vanish in these cases, since nominators and denominators are



Popul. No. 4 ($p_n = .55$; $t = .10$)

Popul. No. 21 ($p_n = .55$; $t = .04$)

Fig. 1. Empirical cumulative frequency curves of w_x and w_x^* from

independent. The variances are obtained by application of formulae (2.10) and (2.11), where the estimates are substituted for the true parameters. For example we have

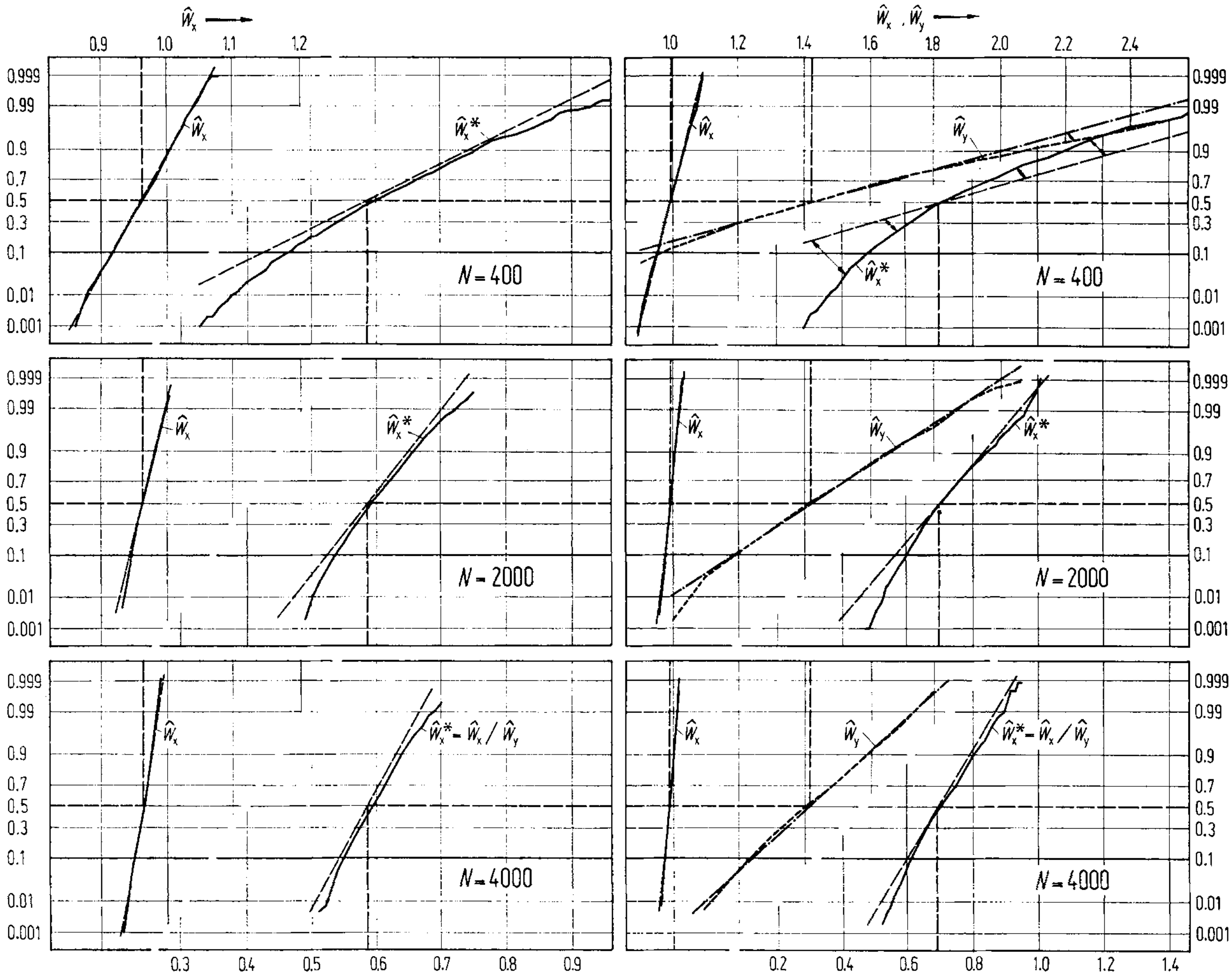
$$\operatorname{var}(\hat{\varphi}_i) = \left[\frac{\partial\varphi_i}{\partial t}, \frac{\partial\varphi_i}{\partial f_n^{(x)}}, \frac{\partial\varphi_i}{\partial f_n^{(y)}}\right] \times$$

$$\times \begin{bmatrix} \operatorname{var}(\hat{t}) & \operatorname{cov}(\hat{t}, x_n) & \operatorname{cov}(\hat{t}, y_n) \\ \operatorname{cov}(\hat{t}, x_n) & \operatorname{var}(x_n) & \operatorname{cov}(x_n, y_n) \\ \operatorname{cov}(\hat{t}, y_n) & \operatorname{cov}(x_n, y_n) & \operatorname{var}(y_n) \end{bmatrix} \begin{bmatrix} \frac{\partial\varphi_i}{\partial t} \\ \frac{\partial\varphi_i}{\partial f_n^{(x)}} \\ \frac{\partial\varphi_i}{\partial f_n^{(y)}} \end{bmatrix}. \tag{2.15}$$

The terms $\frac{\partial\varphi_i}{\partial t}$ etc. are functions of p_n, $f_n^{(x)}$, and t, whereas the figures in the 3×3-matrix are other functions of the same parameters and of H. In all these expressions the involved parameters are replaced by their estimates.

Application of Fieller's theorem to w_x^* and w_z^* means to replace x by $\hat{w}_x$ ($\hat{w}_z$ resp.), y by $\hat{w}_y$, and α by w_x^* (w_z^* resp.). For $\sigma_{x,y}$ the covariances cov $(\hat{w}_x, \hat{w}_y)$ and cov $(\hat{w}_z, \hat{w}_y)$ respectively must be substituted, and var $(\hat{w}_x)$ (var $(\hat{w}_z)$ resp.) is to be used for σ_x^2 and var $(\hat{w}_y)$ for σ_y^2, these all being determined according to (2.10) and (2.11). These expressions are functions of the population parameters, and again these are substituted by their estimates, so that, for example, we have the expression $\widehat{\operatorname{cov}}(\hat{w}_x, \hat{w}_y)$ instead of cov $(\hat{w}_x, \hat{w}_y)$.

With this procedure, however, some of the conditions of Fieller's method are infringed upon. The estimates of variances and covariances, as obtained here, are not independent of the estimates of the ratios, as is required. Furthermore, the nominators and denominators are only approximately normal. The simulation experiments tell us how robust against these deviations Fieller's method for determining confidence limits is. For each individual experiment the 95%-confidence limits for w_x, w_y, w_z, w_x^*, and w_z^* were calculated by this "modified" Fieller-procedure. It was then decided whether or not the true fitnesses



Popul. No. 29 ($p_n = .85$; $t = .10$)

Popul. No. 49 ($p_n = .85$; $t = .04$)

populations 4, 21, 29, 49. Broken lines: Expected limit normal curves

are included. After a run of 1000 experiments an empirical percentage of correct statements is obtained which is compared with the theoretical value of 95%. In Table 3 the relative frequencies of correct statements are shown for the runs of populations no. 4, 21, 29, 49.

Another method for calculating approximate confidence limits for w_x^* and w_z^* was applied at the same time. Estimates of the variance of w_x^* and w_z^* are obtained in the same way as described above. For example

$$\widehat{\mathrm{var}}\,(\hat{w}_x^*) = \left(\frac{\hat{w}_x}{\hat{w}_y}\right)^2 \left[\frac{\widehat{\mathrm{var}}\,(\hat{w}_x)}{\hat{w}_x^2} + \frac{\widehat{\mathrm{var}}\,(\hat{w}_y)}{\hat{w}_y^2} - \frac{\widehat{\mathrm{cov}}\,(\hat{w}_x,\hat{w}_y)}{\hat{w}_x\,\hat{w}_y}\right]. \tag{2.16}$$

All values of w_x^* which now fit the inequality

$$\hat{w}_x^* - 1.96\sqrt{\widehat{\mathrm{var}}\,(w_x^*)} \leq w_x^* \leq \hat{w}_x^* + 1.96\sqrt{\widehat{\mathrm{var}}\,(\hat{w}_x^*)} \tag{2.17}$$

form an approximate 95%-confidence interval for w_x^* in an individual experiment. Again for each run the frequency of correct statements was evaluated. In order to distinguish this kind of limit from the Fieller confidence limit they will be denominated as the Gauss confidence limits.

From Table 3 it is seen that the Fieller limits for w_x, w_y, w_z produce a frequency of errors which in most cases is higher than the admitted level of 5%. Although a certain convergence to the admitted level is observed, which is more pronounced for w_x and w_z than for w_y, even with $N = 2000$ this limit is not attained in all cases. On the contrary, the error frequency for w_x^* and w_z^* agrees well with 5%. This is true for the Fieller limits as well as for the Gauss limits.

3. *Tests of Significance*

Different tests can be constructed for the class of null hypotheses H_0: $w_i = w_j$ $(i, j = x, y, z;\ j \neq i)$ against the alternatives H_1: $w_i \neq w_j$ and then be compared with respect to their power. $\hat{w}_i$ is asymptotically normal for $N = N_1 \to \infty$, whereas $\widehat{\mathrm{var}}\,(\hat{w}_i)$

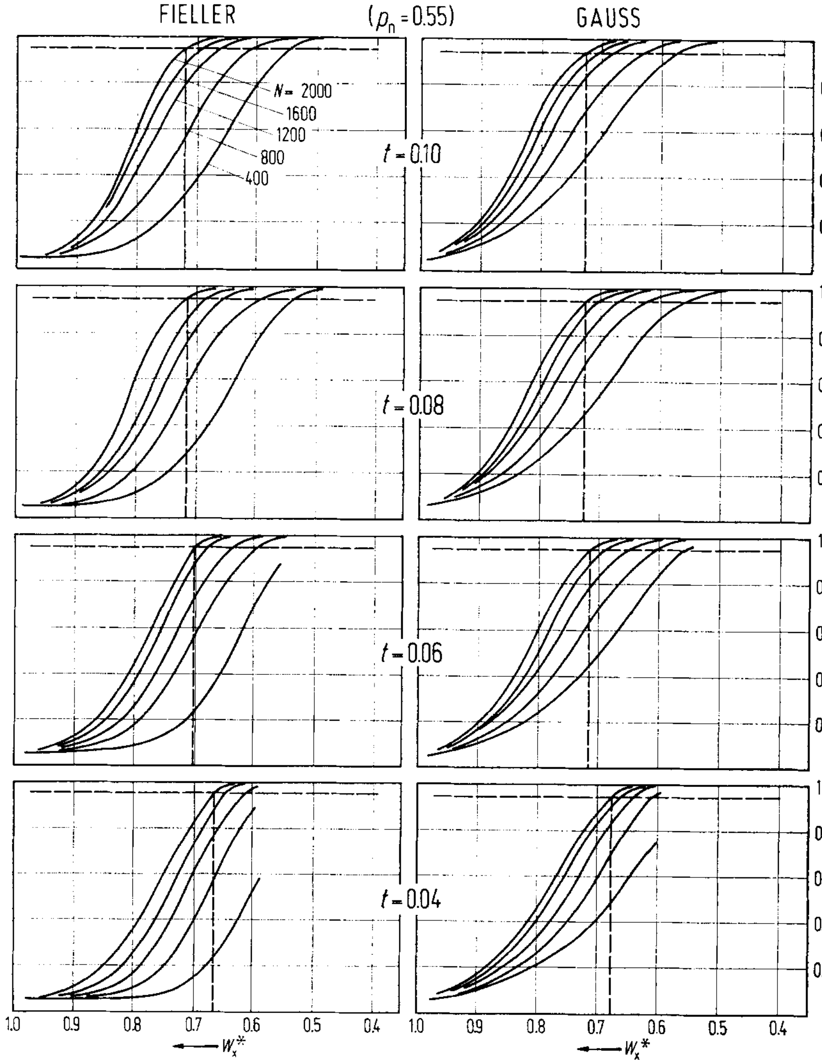


Fig. 2. Power function graphs for quotient tests "due to Fieller" and

and $\widehat{\mathrm{cov}}\,(\hat{w}_i, \hat{w}_j)$ — as calculated according to the rule given above — converge stochastically towards var $(\hat{w}_i)$ and cov $(\hat{w}_i, \hat{w}_j)$ respectively. Therefore a critical region R for H_0 may be chosen with reference to the u-test for normally distributed variates. If

$$\delta = \frac{x - y}{\sqrt{\sigma_x^2 - 2\,\sigma_{x,y} + \sigma_y^2}} = \frac{d}{\sigma_d} \tag{2.18}$$

is used as the test statistic, then the critical region R consists of all those pairs w_i, w_j, for which

$$\mathrm{Prob}\,\{|\delta| > \delta_\varepsilon\} = \varepsilon\,, \tag{2.19}$$

where ε is the admitted error of the first kind and $\delta_\varepsilon = 1.96$ for $\varepsilon = .05$. It is expected that (2.19) holds true approximately under H_0 for N and N_1 being large. On the other hand, it is not expected that under H_0 the test statistic δ has Student's distribution, for the following reasons: $\hat{w}_i$ and $\hat{w}_j$ are not normal; the expression under the root in the denominator is neither distributed as a χ^2 nor are degrees of freedom defined for it; and this expression is not independent of the nominator.

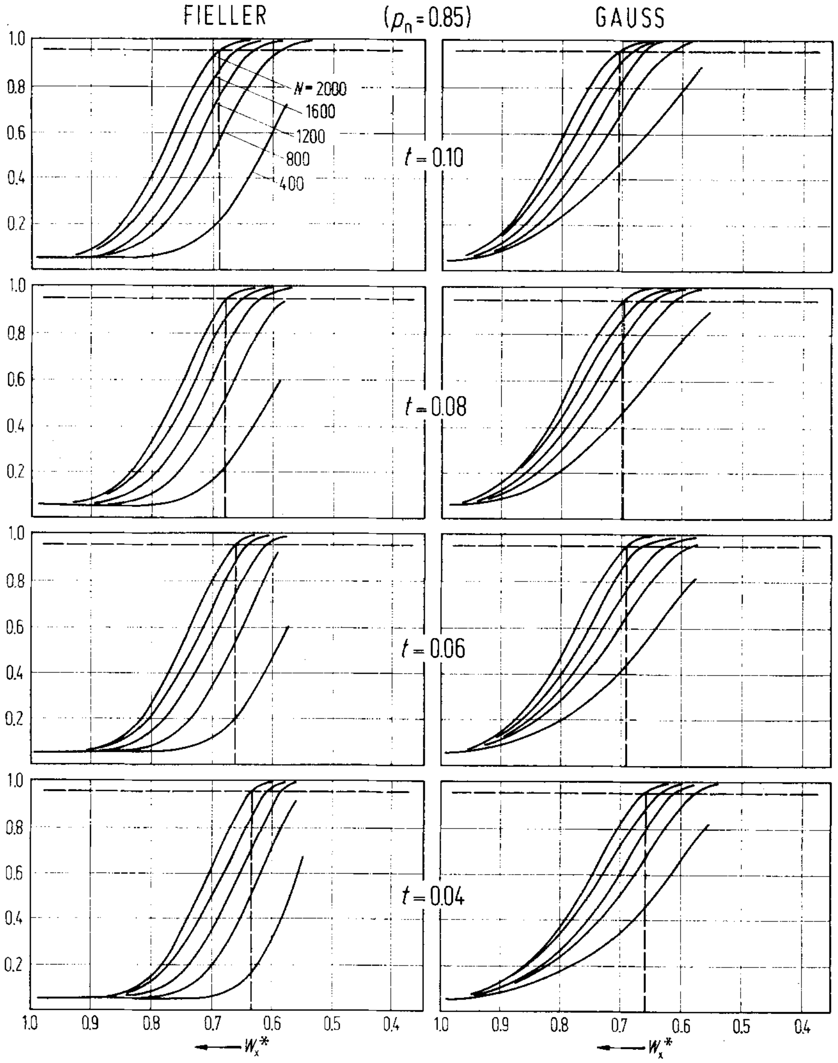
In order to examine the properties of this test, it was asked for each run how frequently the test statistic δ falls into R if H_0 is true. In this way an estimate of the true error of the first kind ε^* is obtained from each run and this is hoped to be equal or very near to ε $(= .05)$.

To do this, for any of the 8 combinations of p_n and t the equilibrium population under the condition $w_x = w_y = w_z = 1$ was ascertained. For each population N and N_1 $(N_1 = N)$ were graduated from 400 to 2000 as before, and for each of the resulting combinations of p_n, t and N a run of 1000 simulations was performed. From each run the estimates $\hat{\varepsilon}^*$ were determined for the three null hypotheses $w_x - w_y = 0$, $w_z - w_y = 0$, $w_x - w_z = 0$ and then compared with .05. The agreement was satisfactory in all cases except for $p_n = .85$, $t = .04$, $N = 400$ and $N = 800$, where the largest deviations were obtained $(\hat{\varepsilon}^* = .075$ and $.077)$.

Afterwards the frequency of δ falling into R was evaluated in cases where H_0 in fact is not true, i.e., where w_i and w_j are different. This frequency is now an estimate of the probability of rejecting H_0, when indeed H_0 is false, such that $D = w_i - w_j$ has a certain non-zero value. Taking into account a series of different values of D, the corresponding frequencies respresent points which are near to the power function $g(D)$ of the test under consideration.

This test, which is based on D, shall be denoted as "difference test". Its power function $g(D)$ will be represented as a function of D, although the power function depends not only on D, but also on the parameters w_i and w_j themselves. To speak in geometrical terms, one has to imagine a power function surface, each point of which has three coordinates w_i, w_j, g, which are non-negative numbers: $w_i > 0$, $w_j > 0$, $0 < g \leq 1$. One of the properties of this surface is that $g = \varepsilon^*$ for all w_i, w_j with $w_i = w_j$, i.e. $g(0) = \varepsilon^*$. The power functions considered here arise from intersections of this surface with cylindrical surfaces, which are erected on a curve in the w_i, w_j-plane, starting at a point where $w_i = w_j$.

Another possibility to compare $\hat{w}_x$ (and $\hat{w}_z$ resp.) with $\hat{w}_y$ arises from confidence intervals, inasmuch



"due to Gauss" for different combinations of p_n, t, and N ($= N_1$)

as a confidence interval may be considered as a test of significance. In our case the null hypothesis H_0: $w_i - w_y = 0$ ($i = x, z$) is to be converted into the form $(w_i/w_y) - 1 = 0$ or H_0^*: $w_i^* - 1 = 0$. H_0^* will be accepted or rejected according to whether or not the number 1.0 is contained within the confidence limits which are determined for w_i^* from the ratio of estimates $\hat{w}_i$, $\hat{w}_y$, i.e., from $\hat{w}_i^* = \hat{w}_i/\hat{w}_y$. In other words, the critical region R^* consists of the set of all those pairs $(\hat{w}_i, \hat{w}_y)$, for which the confidence interval does not contain 1.0. This test will be called a "quotient test". The confidence limits may be determined either "due to Fieller" or "due to Gauss". Both methods were applied in each individual simulation experiment.

Whenever $w_i^* = 1$ is true, the number 1.0 should be outside the confidence limits with probability ε ($= .05$). This requirement was proven to be satisfied for the 8 equilibrium population with $w_x = w_y = w_z = 1$, as described above.

If the quotient test for w_x^* and w_z^* is based on Fieller's limits with respect to the alternative hypotheses in Table 1, it is found that the empirical frequencies $\hat{g}$ of rejecting H_0 are exactly the same as those of the corresponding difference tests for $w_x - w_y$ and $w_z - w_y$ respectively in almost all runs of simulation. Inspection of individual experiments confirmed that $(\hat{w}_i, \hat{w}_y)$ falls into R, if and only if the number 1.0 is included in the critical region of the quotient test "due to Fieller". Only a few exceptions to this rule were observed when $N = 400$. It seems, thus, that both tests are asymptotically equivalent.

To summarize, the following tests were performed in each experiment and the estimates $\hat{g}$ of power determined from each run:

$$H_0\colon w_x - w_y = 0\,,$$
$$w_z - w_y = 0\,,\quad w_x - w_z = 0\,;$$
$$H_0^*\colon w_x^* - 1 = 0\,,\quad w_z^* - 1 = 0$$
(due to Fieller),
$$H_0^*\colon w_x^* - 1 = 0\,,\quad w_z^* - 1 = 0$$
(due to Gauss).

For illustration the power function graphs of the quotient tests "due to Fieller" and "due to Gauss" are given for all 8 combinations of p_n and t, and for different sizes of N ($= N_1$) in Fig. 2.

As a marker for comparison we may use that value of w_x^* for which the (estimated) probability of an error of the second kind is just .05. This means that, if the true value w_x^* in fact deviates from 1.0, then the probability of rejecting the null hypothesis is .95. For $N = 2000$ these critical values are marked by vertical broken lines. It is evident that a test is the more powerful, the smaller the deviation from 1.0, which it can detect at the same level of error of second kind. From the figures it is seen that both quotient tests have more power at $p_n = .55$ than at $p_n = .85$, and that within each p_n-group the power decreases as t decreases. On the other hand the power naturally depends on N, especially if N is not large.

As an example we assume that the errors of the first and second kind shall each be .05. With $N = N_1 = 2000$ a fitness reduction of the homozygotes can be detected with probability .95, if the true fitness

Table 3. *Relative frequencies of correct statements with approximate 95% confidence intervals due to Fieller and Gauss*

N	w_x	w_y	w_z	w_x^*	w_z^*	w_x^*	w_z^*
	(Fieller)			(Fieller)		(Gauss)	
Population No. 4							
400	.896	.863	.888	.950	.948	.944	.942
800	.893	.842	.912	.947	.956	.931	.941
1200	.904	.887	.918	.948	.957	.943	.953
1600	.893	.874	.920	.941	.963	.949	.955
2000	.932	.884	.936	.960	.956	.956	.948
Population No. 21							
400	.909	.822	.886	.951	.953	.944	.941
800	.912	.851	.902	.947	.953	.945	.954
1200	.900	.851	.895	.950	.950	.942	.948
1600	.896	.857	.908	.947	.962	.946	.956
2000	.940	.792	.948	.944	.940	.952	.940
Population No. 29							
400	.915	.849	.893	.939	.952	.947	.937
800	.946	.874	.907	.949	.959	.965	.954
1200	.952	.871	.907	.958	.952	.961	.959
1600	.956	.871	.941	.955	.950	.955	.961
2000	.952	.876	.912	.948	.948	.956	.960
Population No. 49							
400	.937	.849	.850	.950	.967	.952	.943
800	.946	.873	.860	.937	.969	.971	.950
1200	.959	.865	.887	.946	.950	.968	.953
1600	.962	.888	.913	.948	.955	.961	.951
2000	.956	.892	.908	.944	.972	.968	.964

w_x^* is .73 or less. In order to make a statement at the same level of security if the true fitness w_x^* is about .8, the sample size must be increased tremendeously.

From the figures it is seen furthermore that generally the quotient test due to Gauss is more powerful than the test due to Fieller. Especially for values of w_x^* near to 1.0 the power function of the Gauss-test increases more rapidly, and the 95%-threshold is reached for values which are nearer to 1.0 than in case of the Fieller-test. This means that under the same experimental conditions it is easier to detect a certain fitness difference with the Gauss-test than with Fieller's.

The Case $N_1 < N$. Under natural conditions N_1 is generally much smaller than N. As is expected, the approximation of the empirical d.f.s of the fitness estimates by normal d.f.s is less satisfactory than in the case $N_1 = N$. We may ask now for the effect on tests of significance. In Fig. 3 (lower part) some power function graphs are shown for $p_n = .55$, $t = .10$ (population 4, Table 1). For comparison the corresponding graphs under $N_1 = N$ from Fig. 2 are given again. As may be observed, the effect is a considerable one. The critical fitness value w_x^*, which can just be detected with probability .95 at $N = 2000$, $N_1 = 400$, is now .68 (.73 with $N = N_1 = 2000$).

Discussion

It is difficult to draw direct and concrete conclusions from this body of simulation data about the practice of field experiments. Natural populations in general are not in equilibrium, as was assumed here. But this argument has no relevancy to the stochastic properties of fitness estimators. Of practical interest is mainly the detection of true deviations of w_x^* and w_z^* from 1.0, if they exist. At the same time one must avoid the assertion that such deviates exist when they do not.

Among the procedures described above the most simple one denoted as "quotient test due to Gauss" turned out also to be the best one with respect to this purpose. Within the range of population parameters (p_n, t, $f_n^{(x)}$) which was under consideration here, the power of this test depends more on the sample size than on the parameters. However, in order to keep

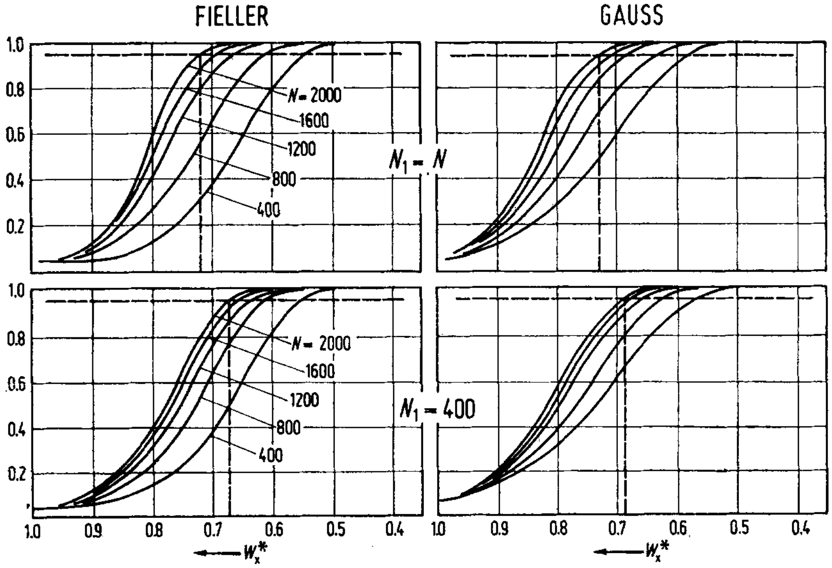


Fig. 3. Power function graphs of quotient tests "due to Fieller" and "due to Gauss" for $p_n = .55$, $t = .10$. Lower part: $N_1 = 400$, upper part: $N_1 = N$ for comparison (from Fig. 2)

the nominal probabilities of an error of first kind, p_n as well as $f_n^{(x)}$ (and $f_{n+1}^{(x)}$) should not have extremely large or small values. Furthermore, t should not be too small if usual sample sizes are employed.

Because of the relatively small sample size which is available for estimating the outcrossing rate t, the only variables which are quite at the experimenter's disposal are the sizes of genotype samples in generations n and $n+1$. Those amounts of the sizes which are necessary in order to guarantee a certain probability of errors of the second kind may be evaluated by means of simulation experiments. In a step-by-step procedure from provisional estimates and confidence intervals for w_x^* and w_z^*, the power of the test can be evaluated for a set of compatible hypotheses about w_x^* and w_z^*. Then it must be decided if more individuals are to be included in the sample.

The program for the estimation of the power, which is written in Fortran IV, may be obtained from the author.

Zusammenfassung

Im ersten Teil dieser Arbeit* wurde eine algebraische Darstellung von Reproduktions- und Selektionsvorgängen in Populationen diploider Organismen gegeben. Im zweiten Teil werden die statistischen Eigenschaften einer speziellen, von Hayman stammenden Schätzfunktion für die Fitness betrachtet. Es werden verschiedene Möglichkeiten zur Berechnung von Konfidenzintervallen und zur Durchführung von Signifikanztesten angegeben und mit Hilfe von Simulationsstudien geprüft.

Acknowledgements

I would like to thank Frau Prof. Dr. M. P. Geppert for valuable critical comments made about an earlier draft of this paper. To Prof. Dr. W. Seyffert I am indebted for his support of the paper. To Dr. K. Wöhrmann I am grateful for many stimulating and critical discussions. The simulation studies were performed on the CDC 3300 computer at the Zentrum für Datenverarbeitung der Universität Tübingen. I am indebted to Dipl.-Mathem. G. Schubring for his help in making the computer program run.

Literature references

of this part of the paper are found at the end of part one, which is published in this journal, Vol. 40, p. 11–17.

---

* Vol. 40 (1970), No. 1, S. 11–17, dieser Zeitschrift.

Received December 19, 1969
Communicated by W. Seyffert

Dipl.-Math. R. J. Lorenz
Bundesforschungsanstalt für
Viruskrankheiten der Tiere
Postfach 1149
74 Tübingen (Germany, BRD)