Euphytica 5 (1956): 329–338

SOME APPLICATIONS OF BINOMIAL PROBABILITY PAPER IN GENETIC ANALYSES

J. H. A. FERGUSON
Institute of Horticultural Plant Breeding, Wageningen
Received 31 Aug. 1956

INTRODUCTION

Binomial Probability Paper, designed in 1946 by MOSTELLER and TUKEY, is a useful tool for genetic analyses, especially with reference to the segregation-ratio in a progeny based upon a binomial distribution. The paper is graduated with a square-root scale on both axes: the distance from a point on a coordinate axis to the origin is the square root of the value indicated. CORSTEN has called it: Double Square-root Paper ("Dubbel-Vierkantswortel Papier").

A brief explanation of the mathematical background will be followed by four examples, after which some data on the accuracy are given.

MATHEMATICAL BACKGROUND

The basic idea involved is R. A. FISHER's transformation:

$$\cos^2 \varphi_i = n_i/n \quad (n = \Sigma\, n_i)$$

in which the observed numbers $n_1, n_2, \ldots\ldots n_k$ of a multinomial, discontinuous distribution are transformed into direction angles $\varphi_1, \varphi_2, \ldots\ldots \varphi_k$, which have approximately normal distribution with standard deviation $1 : 2\sqrt{n}$, the angles measured in radians.

Now let k and $n—k$ be the observed numbers in two categories in which k has a binomial distribution with a probability p. Such a pair of observed numbers will be referred to as a paired count. It is plotted in a graph (fig. 1) as the point with the rectangular coordinates (distances) $\sqrt{k}$ and $\sqrt{n—k}$.

For a number of samples of n individuals from the same distribution the points are all situated on a circle with centre in the origin and radius $\sqrt{n}$. In figure 1 three points have been plotted. The pair of expectation values, np, $n(1—p)$ is represented on this circle by the point $[\sqrt{np},\ \sqrt{n(1—p)}]$. The straight line plotted through the origin and this point is referred to as the split. It passes through all points whose coordinates represent the same ratio, $p : (1—p)$. Apparently the lengths of the arcs between the points and the split have approximately normal distributions with zero mean and standard deviation of $1 : 2\sqrt{n}$ radian. Multiplied with the length of the radius, the standard deviation in absolute measure is $\frac{1}{2}$, which appears to be independent of the size of the sample.

This is the reason why Binomial Probability Paper (B.P.-paper) is of great practical value.

A simplification is obtained by replacing the arcs by the distances from the paired counts to the split.

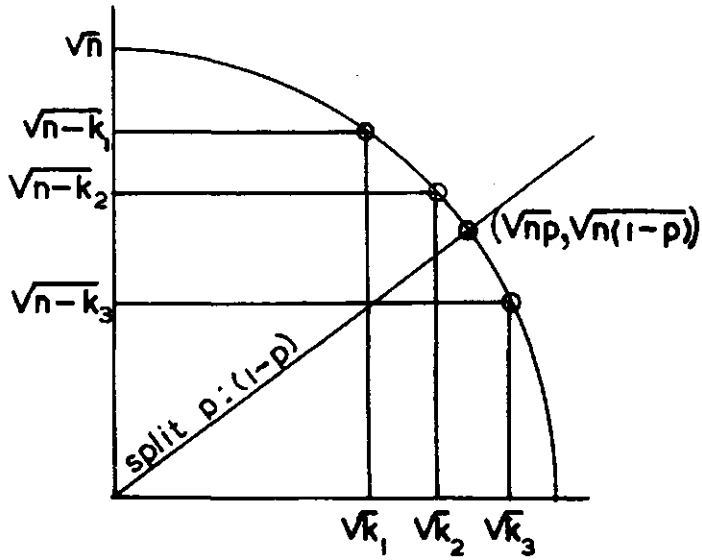

FIG. 1. GRAPH FOR THE SEGREGATION-RATIO $p:(1-p)$

MOSTELLER and TUKEY found empirically a modification, which improved the approximations near $p = 0$ and $p = 1$, namely by adding $\frac{1}{2}$ to each coordinate, thus passing from $(k, n-k)$ to $(k+\frac{1}{2}, n-k+\frac{1}{2})$.

Most of the applications of B.P. paper deal with tests of hypotheses. Given p and n a critical region $k \leqslant r = r(p, n)$ is determined by a choice of r. If $k > r$ the hypothesis $(p, 1-p)$ will be accepted, and if $k \leqslant r$ it will be rejected. The number $r + \frac{1}{2}$ can act as an unambiguous frontier between the critical and the non-critical integers, yielding the point $(r+\frac{1}{2}, n-r-\frac{1}{2})$. In accordance with the last paragraph this frontier is scored on the graph as if it were $(r+1, n-r)$. For a test in the other direction the other coordinate is increased by unity, which gives $(r, n-r+1)$.

Therefore MOSTELLER and TUKEY plot a paired count as a triangle, with sides extending one scale unit in the positive directions of the axes from its right angle vertex, which has the paired count for its coordinates. For a test of hypothesis the distance from a paired count to the split is measured as the shortest distance from the triangle to (and perpendicular on) the split. The centre of the hypotenusa is the point $(k+\frac{1}{2}, n-k+\frac{1}{2})$ which, in view of the modification of MOSTELLER and TUKEY, mentioned above, yields a better estimate of $p:(1-p)$. For large samples the triangle will be practically reduced to a line-segment or to a dot.

Now let a split be drawn which represents the given hypothetical binomial distribution (e.g. $p:(1-p) = 1:3$). Let a number of paired counts, samples from the given distribution, be entered in the graph. As mentioned before the standard deviation of the distances from the paired counts to the split is $\frac{1}{2}$, independent of the size of the samples. In the graph this is half the distance from the origin to the point 1 on a coordinate axis. In the original B.P. paper, as printed by Codex Book Cy (see references) σ amounts to 5.08 mm, as stated by the designers. Two straight lines drawn parallel to the split on both sides at distances of $1.96\,\sigma = 10.0$ mm from the split, limit a region within which 95 % of all sampled paired counts from the universe may be expected; the remaining 5 % will be found outside the limiting lines.

Now let there be another sample of which it is not known whether it belongs to the universe. If this paired count, i.e. the whole triangle, falls within the critical region, that is outside the parallel lines, we decide for this sample to reject the given hypothesis at a significance level of 0.05.

In table no. 1 some normal deviates with matching significance levels, for two-sided and one-sided critical regions, have been recorded.

TABLE 1. SIGNIFICANCE LEVELS FOR DEVIATES IN MILLIMETERS FOR CODEX' BINOMIAL PROBABILITY PAPER

Deviate in millimeters	Multiple of σ (= 5.08 mm)	Significance level	
		one-sided	two-sided
0	0	50 %	100 %
1.3	0.253	40	80
2.7	0.524	30	60
3.4	0.674	25	50
4.3	0.842	20	40
5.3	1.036	15	30
6.5	1.282	10	20
8.4	1.645	5	10
10.0	1.960	2.5	5
11.8	2.326	1.0	2
13.1	2.576	0.5	1
15.7	3.090	0.1	0.2

At the left top of the original B.P.-paper a line has been drawn to indicate multiples of the individual standard deviation ("Full Scale"). In fig. 2 the paper has been truncated, the abscis running from 0 to 150 instead of 600.

The quarter circle with centre in the origin on the B.P.-paper indicates the place on the splits where the coordinates add up to $n = 100$, thus facilitating the reading of the value of a split in percentages.

APPLICATIONS

Some of the applications of B.P.-paper, probably of most importance in genetic studies, will be explained by means of three examples, chosen from the 21 given by MOSTELLER and TUKEY. They are followed by a special problem for which B.P.-paper appeared to be useful.

Example 1

Given a paired count of numbers in two categories in a sample, it is asked whether these numbers agree with a given theoretical sampling probability. For instance: an F_2 segregates in the numbers 115:25. The expectation was 3:1. It is asked whether the departure from the expected numbers, 105:35, is significant. The paired count is plotted on B.P.-paper as a triangle, (115, 25), (115, 26) and (116, 25). The split 3:1 is drawn, after which the short distance of the paired count to the split appears to amount

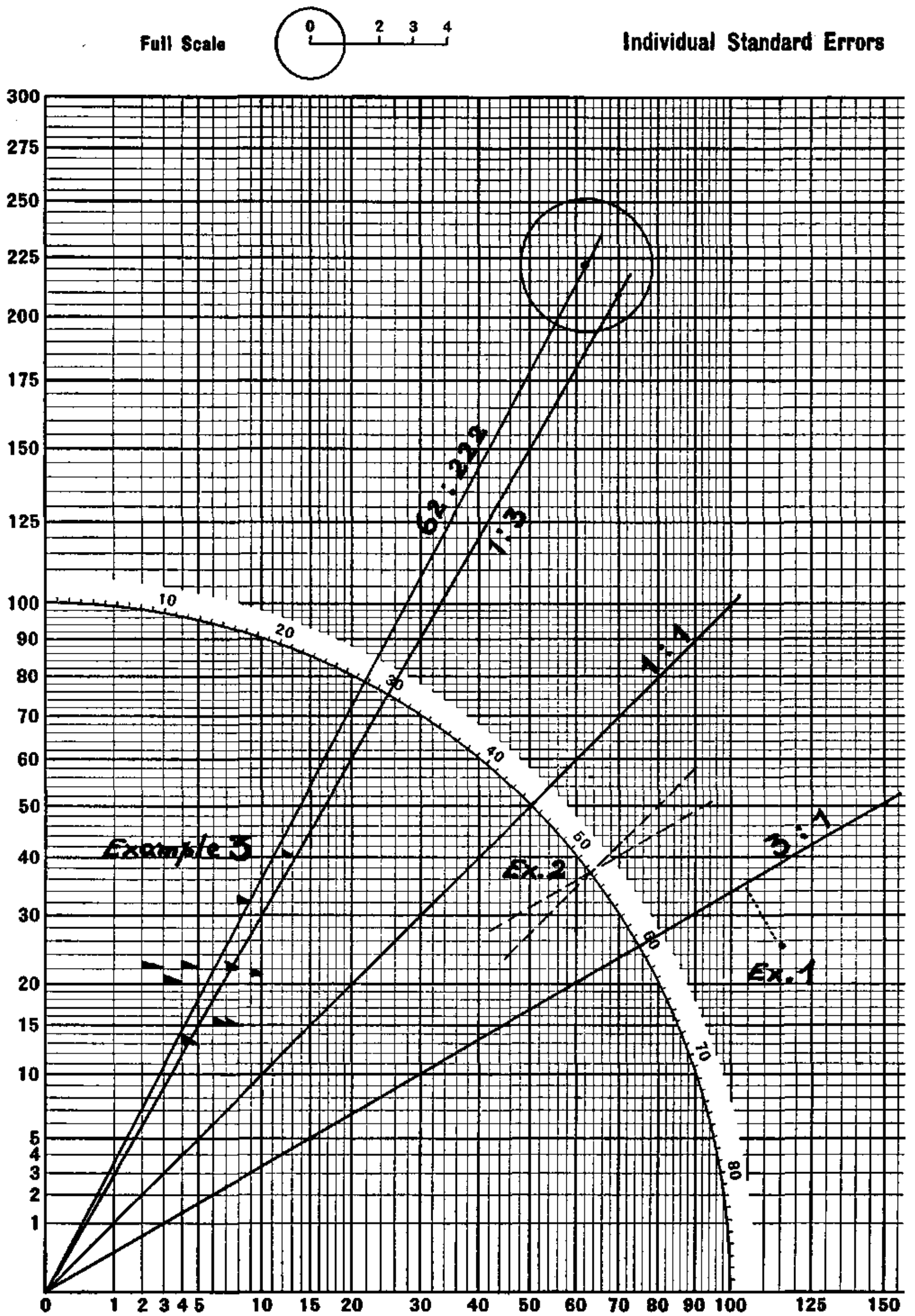

FIG. 2. THE USE OF BINOMIAL PROBABILITY PAPER

to 9.5 mm. From the table it appears that a departure in both directions of at least this size may be expected with a probability of more than 5 %. Therefore the deviation from the expectation is not significant and the hypothesis need not be rejected.

Example 2.

A test will be expected to give significant evidence whether a certain character appears in one half, or in one quarter of a progeny of a certain mating. What is the smallest size of the progeny that will give significant evidence against at least one of the theories at the one-sided $\frac{1}{2}$ per cent level? A straight line is drawn at 13.1 mm distance parallel to and below the split 1:1, and another line at the same distance parallel to and above the split 3:1. These lines correspond with a one-sided $\frac{1}{2}$ % significance level. They intersect at the point (63, 37). This point separates the triangle (63, 36), (63, 37), (64, 36) from the triangle (62, 37), (62, 38), (63, 37). Therefore the point (63, 36) is exactly beyond the 1 % significance level from the 1:1 split, while (62, 37) is beyond the 1 % level from the 3:1 split. Thus a progeny of 99 individuals should be large enough. It should be mentioned that the real significance level is at 0.5 % level, as we test against one alternative hypothesis, one-sided only.

Example 3

The progenies of ten plants segregate in a certain character, probably monofactorial, into the following numbers of alternative types:

Plant nr.	1	2	3	4	5	6	7	8	9	10	Total
Number type A	2	3	6	7	4	7	4	9	8	12	62
Number type B	22	20	15	22	22	15	13	21	32	40	222

It is asked whether the segregations are homogeneous and upon which ratio they might be based. The 10 points are plotted as triangles. The totals give the paired count (62, 222). With this point as a centre a circle is drawn with a radius of 10 mm. The two tangents through the origin mark the confidence limits of the segregation ratio upon which the result may be based. It may be presumed that the true ratio is 1:3, which split is now drawn. The distance from this split to the remotest paired count, measured as short distance, appears to be 9 mm. Therefore there is no objection to accepting a segregation ratio 1:3 for the entire experiment.

Special problem

This problem refers to data from a pilot-investigation, not yet published, of Dr. Banga, which he was kind enough to allow me to use. It concerns male-sterility of *Daucus carota*, which normally bears bisexual flowers. In some cases, however, individuals referred to as "male-sterile" are found, whose stamina have not developed. In this investigation male-sterile plants grown from imported seed were mated freely with normal bisexual plants from different varieties of Dutch origin. Twenty seeds from each of the 124 male-sterile mother plants were planted to obtain progenies, which varied in number from 17 to zero, owing to bad germination. These progenies may segregate into normal bisexual and male-sterile plants, each giving a paired

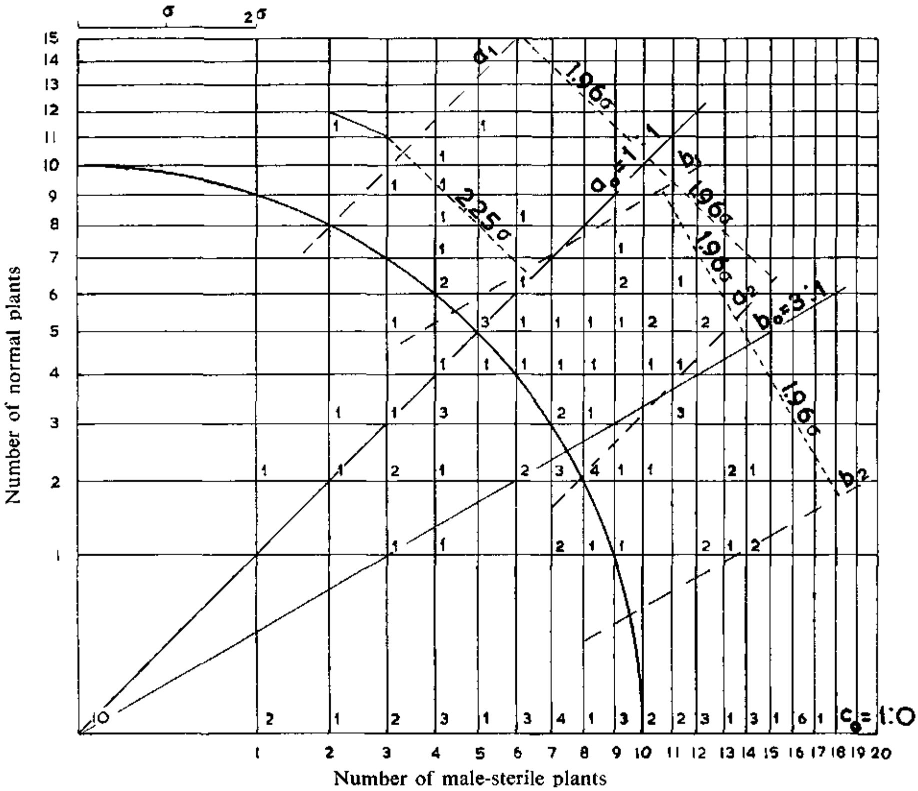

FIG. 3. SEGREGATION OF PROGENIES FROM DAUCUS CAROTA, MALE-STERILE × NORMAL PLANTS

count which was entered on binomial probability paper. As the progenies were small in number a special large-scale graph was drawn (fig. 3). As many paired counts coincide, a number indicating the frequency was noted in the lower left angle of the triangle, which stands for the paired count. For the sake of clearness of the chart the hypotenusae of the triangles were not drawn, with the exception of the hypotenusa belonging to (2, 11) which was drawn.

Apparently progenies entirely consisting of normal plants do not occur.

There appears to be a slight tendency of concentration around the splits 1:1, 3:1 and 1:0, which were drawn and indicated by a_0, b_0 and c_0 (the abscis). Working on these data Dr. BANGA could draw a preliminary genetical hypothesis (not to be explained here) which would be most acceptable if the segregation b, 3:1, would occur with a chance twice as large as the segregation a, 1:1 and c, 1:0 respectively. Now we want to test this hypothesis.

In the first place the short distance of the paired count (2, 11) from the split a_0 was measured and appeared to be 2,25σ; this deviation is not significant on the two-tailed 1 % level and therefore can not upset the hypothesis. The other paired counts are all nearer to any drawn split; therefore it may be concluded that all samples may be explained by at least one of the hypothetical segregations.

To check whether the frequencies of the three possible segregations, a, b and c agree with the odds $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively, only progenies with at least 10 plants have been used, which gives a number of 71 progenies. The expected values of the frequencies are $17\frac{3}{4}$, $35\frac{1}{2}$ and $17\frac{3}{4}$ respectively. However, the normal ranges in the three categories overlap considerably.

Parallel to, and at distances of 1,96 σ from the splits a_0 and b_0, four lines were drawn, indicated by a_1, a_2, b_1 and b_2. Within the region between a_1 and a_2 95 % of the a-distribution, that is the paired counts based on the a-segregation, are expected. The same holds for the b-segregation between the lines b_1 and b_2. All progenies that can have only male-sterile plants, must lie in the lowest row of paired counts. Above the split a_0, which is expected to be the median of the a-distribution, we expect one half of all paired counts from this distribution plus a very small percentage (circa $2\frac{1}{2}$ %) of the b-distribution, which we neglect.

If we allot half of the four paired counts in (5,5) and (6,6) to the upper half of the a-distribution, this part has 12 paired counts.

As the split b_0 happens to cross the same paired counts as the line a_2, we may again consider the total frequency in the region between a_0 and b_0 as a rough estimate of the lower half of the a-distribution plus the upper half of the b-distribution. The frequencies in the paired counts (8,2) and (11,3) are allotted to the upper half and lower half of the b-distribution proportional to the segments in which the hypotenusae are cut by the split b_0; this yields two points for the upper half and 5 for the lower half. Thus we find 24 paired counts between the splits a_0 and b_0.

The remaining 35 paired counts under the split b_0, belong to the lower half of the b-distribution plus the c-distribution. The line b_2 again happens to cross the ordinate-line for the value 1, which is the upper limit for the c-distribution. Below this line we find 19 paired counts, of which one may be allotted to the lower half of the b-distribution, according to the same principle as mentioned before. Thus 18 paired counts remain for the c-distribution, leaving 35–18 = 17 paired counts in the lower half of the b-distribution.

Thus we found:

12 paired counts in the upper half of the a-distribution
24 „ „ „ „ lower half of the a-distribution plus the upper half of the b-distribution
17 „ „ „ „ lower half of the b-distribution
18 „ „ „ „ c-distribution

Now let m_1, m_2, m_3 and m_4 be the total numbers of paired counts present in the four regions mentioned above. Let μ_1, μ_2, μ_3 and μ_4 be the expected numbers in these regions, depending on the hypothetical proportions p_a, p_b and p_c of the numbers of progenies segregating on the bases of the ratios $a = 1:1$, $b = 3:1$ and $c = 1:0$. Then $m_1 + m_2 + m_3 + m_4 = \mu_1 + \mu_2 + \mu_3 + \mu_4 = N$

Hence we can write:

$$\mu_1 = \tfrac{1}{2} p_a N$$
$$\mu_2 = \tfrac{1}{2} (p_a + p_b) N$$
$$\mu_3 = \tfrac{1}{2} p_b N$$
$$\mu_4 = p_c N$$

RASSENLIJSTEN [1])
UITGEGEVEN DOOR HET INSTITUUT VOOR DE VEREDELING VAN TUINBOUWGEWASSEN

Achtste Beschrijvende Rassenlijst voor Fruit. 1957. . . *f* 1,75

Negende Beschrijvende Rassenlijst voor Groentegewassen. 1957. Redacteur Dr. O. Banga . . . . *f* 1,75

PUBLIKATIES VAN HET INSTITUUT VOOR DE VEREDELING VAN TUINBOUWGEWASSEN IN ANDERE ORGANEN OF IN BOEKVORM EVENTUEEL IN SAMENWERKING MET ANDERE INSTELLINGEN [2])

De publikaties, waarvan prijs èn uitgever worden vermeld zijn verkrijgbaar in de boekhandel. Overigens wende men zich tot de opgegeven bronnen of tot de bibliotheek van het I.V.T.

Boom, B. K. Cotoneaster wardii en verwante soorten. De Boomkwekerij **11**, 1955: 3.
Gerritsen, C. J. Zit er wat in de teelt van hazelnoten? De Fruitteelt **45**, 1955: 865.
Kronenberg, H. G. Aardbeien. Wat moeten we toch planten? De Fruitteelt **45**, 1955: 866–867.
Gerritsen, C. J. Gaat U kersen planten? De Fruitteelt **45**, 1955: 909–910.
Burg, J. P. L. L. A. en **G. Elzenga.** Rapport over een studiereis aangaande de teelt en verwerking van geneeskrachtige en aromatische gewassen in Duitsland en Frankrijk (16 t/m 31 augustus 1955). V.N.K.-Nieuws, september 1955: 92–99.
Boom, B. K. Sorbus prattii en S. koeneana. De Boomkwekerij **11**, 1955: 27.
Elzenga G. Het rooien van de wortels van Angelica en Valeriaan. V.N.K.-Nieuws, november 1955: 110–111
Elzenga, G. Pepermunt opnieuw inplanten. V.N.K.-Nieuws, november 1955: 112.
Boom, B. K. Vraagstukken rondom het Cotoneaster-sortiment. De Boomkwekerij **11**, 1955, 41–42.
Jensma, J. R. Rassenkeuze bij bloemkool. Groenten en Fruit **11**, 1956: 721.
Bruyne, A. S. de. Nieuwe appelrassen tot James Grieve. De Fruitwereld **1**, 1956; no 4; 8–9.
Boom, B. K. Cercidiphyllum. De Boomkwekerij **11**, 1955: 27.
Broertjes, C. Reactie op vraagstukken rondom het Cotoneastersortiment. De Boomkwekerij **11**, 1956: 67–68.
Broertjes, C. Veredeling op ziekteresistentie bij rozen. De Boomkwekerij **11**, 1956: 73.
Boom, B. K. Acer platanoides ,reitenbach' en ,rubrum' De Boomkwekerij **11**, 1956: 74.
Bruyn, J. W. de. De exportcontrole van kruiden in 1955. V.N.K.-Nieuws, januari 1956: 134–135.
Boom, B. K. Buxus, buksus of buks. De Boomkwekerij **11**, 1956: 80–81.
Boom, B. K. Drie nieuwe wilgen. De Boomkwekerij **11**, 1956: 81–82.
Boom, B. K. Enkele bontbladige bomen. De Boomkwekerij **11**, 1956: 88.
Boom, B. K. Een nieuwe monographie over het geslacht Philadelphus. De Boomkwekerij **11**, 1956: 96–97.
Gerritsen, C. J. Zal de noot een deugd worden? De Fruitwereld **1**, 1956: no. 14: 5.
Banga, O. Enkele opmerkingen naar aanleiding van een internationale conferentie. Zaadbelangen **10**, 1956: 101–102.
Kronenberg, H. G. Strawberry growing in the Netherlands. American Fruit Grower **76**, 1956; no. 4: 77.
Elzenga, G. Lobelia inflata. V.N.K.-Nieuws, maart 1956: 163–166.
Boom, B. K. Variëteit en cultivar. De Boomkwekerij **11**, 1956: 112–113.
Andeweg, J. M. *Vroegrijpende moneymaker's.* Zaadbelangen **10**, 1956: 145.
Boom, B. K. Verwarring over de plantennamen. Vakblad voor de Bloemisterij **11**, 1956: 130–131.
Gijsbers, J. W. Ruimtebesparing bij de opberging van dia's en negatieven. Meded. Dir. Tuinbouw **19**, 1956: 298–300
Boom, B. K. Over een verzameling prijscouranten. De Boomkwekerij **11**, 1956: 128-129.
Boom, B. K. Een Amerikaan over Boskoop. De Boomkwekerij **11**, 1956: 130.
Huyskes, J. A. Klauwenselectie bij asperges geeft goede resultaten. Boer en Tuinder (Land en Vee) **10**, 1956; no. 482: 17.
Koot, Y. v. en **J. M. Andeweg.** De groenteteelt in Amerika. 's-Gravenhage, C.O.P., 1956. 149 blz. f 7,00.
Banga, O. Kweker en overheid in de sector groentezaden. Zaadbelangen **10**, 1956: 189-190.
Kho, Y. O. Opbrengstvermindering en kiemkrachtverlaging van wortelzaad als gevolg van aantasting door wantsen. Zaadbelangen **10**, 1956: 193-194.
Elzenga, G. *Digitalis lanata Ehr.* V.N.K.-*Nieuws 1956:* 167-170, 193-199.
Andeweg, J. M. Rationalisatie en rassenkeuze. Groenten en Fruit **12**, 1956: 111.
Kho, Y. O. en **J. P. Braak.** Opbrengstvermindering en kiemkrachtverlaging van wortelzaad als gevolg van aantasting door wantsen. Meded. Dir. Tuinb. **19**, 1956: 440-445.
Kronenberg, H. G. Praktijkproeven met aardbeien in 1956. De Tuinderij **36**, 1956, no 33: 1-3. Groenten en Fruit **12**, 1956: 177.
Floor, J. en **P. A. Wezelenburg.** Stekken onder plastic. De Boomkwekerij, **11**, 1956: 174-175.
Terpstra, W. Some factors influencing the abscission of debladed leaf petioles. Acta Botanica Neerlandica **5**, 1956: 157-170
Bruyne, A. S. de. Trends and developments in Dutch varieties. The Commercial Grower 1956, no 3165: 419-422.
Smeets, L. A note on the shortening of the juvenile phase in cherry seedlings. Euphytica **5**, 1956: 117-118.
Boertjes, C. Vorstschade aan Rhododendronvariëteiten in 1956. De Boomkwekerij **11**, 1956: 187-189.
Rodenburg, C. M. Het kweken van wolfresistente spinazierassen. Zaadbelangen **10**, 1956: 325-326.
Andeweg, J. M. Een waardevolle vroege kruisingsouder. Zaadbelangen **10**, 1956: 344.
Huyskes, J. A. en **C. M. Rodenburg.** Internationale samenwerking bij het onderzoek van slarassen. Meded. Dir. Tuinb. **19**, 1956: 823:826.
Gerritsen, C. J. De teelt van buitenperziken I, II, III, IV. Groenten en Fruit **12**, 1956: 537-538; 569-570; 603; 628-629.
Gerritsen, C. J. De Feyoa, aan nieuw cultuurgewas? Meded. Dir. Tuinb. **19**, 1956: 889-894.
Jensma, J. R. *Sluitkoolrassen. Wageningen,* I.V.T. *1956:* 150 blz. f 13.50.
Floor, J. Planten in plastic. Wageningen, I.V.T., 1956. f 0.35.
Elzenga, G. De teelt van Valeriaan. V.N.K.-Nieuws **4**, 1956: 234-236.
Elzenga, G. Het mechanisch rooien van Valeriaan. V.N.K.-Nieuws **4**, 1956: 246.
Elzenga, G. Het opkweken van plantmateriaal van Levisticum en Rheum. V.N.K.-Nieuws **4**, 1956: 246-247.
Elzenga, G. Roest in munt. Zou gier helpen? V.N.K.-Nieuws **4**, 1956: 247.
Elzenga, G. Mechanisch planten van Valeriaan blijkt zeer goed mogelijk. V.N.K.-Nieuws **4**, 1956: 248-249.
Elzenga, G. De oogstdatum van Digitalis lanata. V.N.K.-Nieuws **4**, 1956: 249-250.
Giessen, A. C. v. d. en **A. v. Steenbergen.** Een nieuwe methode voor de toetsing van bonen op resistentie tegen vlekziekte. Zaadbelangen **11**, 1957: 26-27.

[1]) Zolang de voorraad strekt kunnen deze publikaties franco worden toegezonden, na ontvangst van het vermelde bedrag op giro no. 425340 van het Instituut voor de Veredeling van Tuinbouwgewassen, S. L. Mansholtlaan 15 te Wageningen onder vermelding van wat verlangd wordt; ook bestaat de mogelijkheid deze publikaties uit de bibliotheek van het I.V.T. te lenen.
[2]) Eerder verschenen publikaties zijn vermeld achterin in de Mededelingen nos 1 t/m 89.

DRUK: H. VEENMAN EN ZONEN, WAGENINGEN