PRINTED IN POLAND

Nakład 1500 egz. Papier bezdrz. sat. 100 g 70×100 cm.
Druk ukończono w czerwcu 1951 r.

13/51. — DRUKARNIA WROCŁAWSKIEGO TOW. NAUK. WE WROCŁAWIU. — F. 1-11084

COLLOQUIUM MATHEMATICUM

VOL. II WROCŁAW 1951 FASC. 2

C O M M U N I C A T I O N S

ON THE DUAL ASPECT OF SAMPLING PLANS

BY

J. ODERFELD (WARSAW)

1. Remarks on sampling problems in industry. Control plays an important part in the process of manufacture. The scope of this paper will be restricted to the final acceptance control of the finished product supplied in *lots.* Let us further assume that each piece of the lot may be classified into one of two classes only, i. e. *good* or *defective.* This is called *classification by attributes.* If a 100% inspection is possible, control activities are reduced to counting and screening of all defective pieces. However, in many instances only a statistical approach to the problem of control is possible.

The simplest method of sampling consists in taking a sample of n pieces from the lot. According to the agreement between the producer and the consumer the lot will be accepted if in the sample of n pieces $0, 1, 2, \ldots$ or m defective pieces are found; the lot will be rejected if, out of the n pieces, more than m defective pieces are found. We call such an agreement a *single sampling plan.* This plan being characterized by two numbers n and m only, we denote it briefly by $m\|n$.

2. Two points of view. Both the producer and the consumer speak in terms of quality. For the purposes of this paper *quality* will be defined as the ratio of the number of all defective pieces in the lot to the number of all the pieces in the lot. Each of the two partners interested in the acceptance procedure puts a different question to the statistician.

The producer asks: "If I do constantly offer a product of a given quality and if the sampling plan $m\|n$ is specified, what shall be my risk of rejection?"

The consumer's problem is entirely different. What he really wants to know is this: "What is the probability of my accepting a product of a lower quality than required if I do adopt the sampling plan $m\|n$?"

Other, but less important types of the consumer's questions may be asked and answered by applying direct probabilities or the principle of maximum likelihood.

Statistics answers the producer's question by means of *direct probabilities*, and the consumer's question by means of *inverse probabilities*. The former method is also called *prospective*, and the later — *retrospective*. In the retrospective method a recourse has to be made to some hypothesis as to the prior probability of the qualities before the experiment. This problem was solved as early as 1763 by Bayes (*Bayes' theorem*), who in addition stated that the distribution of the prior probabilities was to be assumed as uniform if there was no reason known to suppose them to be different (*Bayes' postulate*). It is not intended here to criticize or to defend Bayes' postulate; too much has been written in its favour or against it.

On the other hand one cannot accept without some criticism the wording of the producer's prospective question. What is the real meaning of the condition: "if the quality of the product is permanently constant", or even of the more cautious formulation: "if the average quality of the product is constant"? How is the producer to know that he really supplies a product of a constant quality? If the 100% control is not possible, he is compelled to perform some sort of statistical control for his own guidance, and in this case he is necessarily obliged to reason from the events observed to the hypothesis which may explain them. This brings him back to the retrospective method.

Therefore it is evident that some way of solving the retrospective problem is essential from the industrial point of view.

3. Prospective and retrospective interpretation of the sampling plan. Both in the producer's question and in that of the consumer a pair of statistical values is mentioned: a quality and a probability. When comparing different sampling plans from the producer's point of view, it is sufficient to know, for any given plan $m\|n$, two such pairs, (α_1,β_1) and (α_2,β_2), the letter α being adopted for qualities, and the letter β — for probabilities. In general use, fixed values β_1 and β_2 are agreed upon to be applied in all sampling plans; they are fairly high, say 0,90, 0,95 or 0,99. It is convenient to choose $\beta_1=\beta_2$, although this is not necessary. The qualities α_1 and α_2 are then defined so as to verify the following two statements:

3.1. If a lot of quality α_1 is submitted for inspection according to the plan $m\|n$, it will be rejected with the probability β_1. We briefly denote this statement by $P_1(\alpha_1,\beta_1,m,n)$.

3.2. If a lot of quality α_2 is submitted for inspection according to the plan $m\|n$, it will be accepted with the probability β_2. We briefly denote this statement by $P_2(\alpha_2,\beta_2,m,n)$.

When we compare sampling plans from the consumer's point of view, the pairs of statistical values proper to characterize any plan $m\|n$ should verify the following four statements:

3.3. If a lot submitted for inspection according to the plan $m\|n$ has been accepted, its quality is at least α_1 with a probability *at least* β_1. This statement may be denoted by $R_1(\alpha_1,\beta_1,m,n)$.

3.4. If a lot submitted for inspection according to the plan $m\|n$ has been rejected, its quality is at most α_2 with a probability *at least* β_2. This statement may be denoted by $R_2(\alpha_2,\beta_2,m,n)$.

3.5. In both statements R_1 and R_2 the hypothesis of uniform distribution a priori of the qualities has been tacitly assumed as valid.

3.6. The italicized *"at least"* in R_1 means that the probability is exactly β_1, if the inspection has shown m defectives, and that it is greater in other cases of acceptance. The italicized *"at least"* in R_2 means that the probability spoken of is exactly β_2, if the inspection has revealed $m+1$ defectives, and that it is greater in other cases of rejection.

4. The theorem and its proof. There are very simple relations connecting the statements of the kind P with those of the kind R. Denoting by $\equiv$ the equivalence of propositions we can write

$$P_1(\alpha,\beta,m,n)\equiv R_1(\alpha,\beta,m,n-1), \tag{1}$$

$$P_2(\alpha,\beta,m,n)\equiv R_2(\alpha,\beta,m-1,n-1). \tag{2}$$

4.1 [1]). To prove (1) and (2) let us consider Taylor's expansion of $f(x)=x^n$. It gives

$$(\alpha+\omega)^n=\sum_{j=0}^{m}\binom{n}{j}\alpha^{n-j}\omega^j+R_{m+1} \qquad (m<n). \tag{3}$$

[1]) Section 4.1 is worked out along the lines given by M. G. Kendall in his *Advanced Theory of Statistics*, London, 1947.

Since

(4) $$f^{(m+1)}(x) = n(n-1)\dots(n-m)x^{n-m-1} \qquad (m<n),$$

Cauchy's expression

5) $$\frac{\omega^{m+1}}{m!}\int_0^1 f^{(m+1)}(\alpha+t\omega)(1-t)^m dt$$

for the remainder R_{m+1} assumes in our case the form

(6) $$R_{m+1} = \frac{n!\,\omega^{m+1}}{m!(n-m-1)!}\int_0^1 (\alpha+t\omega)^{n-m-1}(1-t)^m dt.$$

If $0<\alpha<1$, and $\omega=1-\alpha$, the transformation $t=1-x/\omega$ reduces $\alpha+t\omega$ to $1-x$, and formula (6) to

(7) $$R_{m+1} = \frac{n!}{m!(n-m-1)!}\int_0^\omega x^m(1-x)^{n-m-1}dx.$$

It is evident by (3) that R_{m+1} is the sum of all the terms of the Newton-Moivre expansion

$$(\alpha+\omega)^n = \sum_{j=0}^{n}\binom{n}{j}\alpha^{n-j}\omega^j,$$

with $j=m+1,\ m+2,\ \dots,\ n$.

4.2. Now assume that α is the quality of the product submitted for inspection, and that n is the size of the sample. As $\omega=1-\alpha$ is the probability that a piece chosen at random from the lot is defective, R_{m+1} is the probability that more than m defective pieces will be found in the sample. Thus, if the plan is $m\|n$, R_{m+1} is the probability that a product of quality α will be rejected.

4.3. Suppose now that in a sample of $n-1$ pieces m are found to be defective. Furthermore, assume the hypothesis of uniform distribution of the qualities a priori. Then, according to Bayes' rule, R_{m+1}, as given by expression (7), is the probability a posteriori that in the inspected product the ratio of defective pieces in the lot is not greater than $\omega=1-\alpha$; in other words, that the quality of the product is not lower than α.

4.4. Putting β for R_{m+1}, and taking account of 3.1, 3.3, 3.5, 3.6, 4.2 and 4.3, we get the equivalence (1).

4.5. As in the plan $m\|n$ rejection occurs when more than m defectives have been found in the sample, and acceptance means the complementary case of no more than m defectives, the context of 3.1 and 3.2 leads immediately to the equivalence

(8) $$P_1(\alpha,\beta,m,n) \equiv P_2(\alpha,1-\beta,m,n).$$

On the other hand, the statement $R_1(\alpha,\beta,m,n-1)$, as defined by 3.3 and 3.6, gives the probability β for a quality at least α, if there were exactly m defectives in a sample of $n-1$ pieces; the statement $R_2(\alpha,1-\beta,m-1,n-1)$, as defined by 3.4 and 3.6, gives the probability $1-\beta$ for a quality at most α, if there were exactly m defectives in a sample of $n-1$ pieces. The two statements being equivalent, we write

(9) $$R_1(\alpha,\beta,m,n-1) \equiv R_2(\alpha,1-\beta,m-1,n-1).$$

Comparing (1), (8) and (9) we get

$$P_2(\alpha,1-\beta,m,n) \equiv R_2(\alpha,1-\beta,m-1,n-1),$$

and writing β for $1-\beta$ we get equivalence (2).

Now, replacing m by $m+1$, we get

(10) $$P_2(\alpha,\beta,m+1,n) \equiv R_2(\alpha,\beta,m,n-1).$$

Equivalences (1) and (10) constitute the *rule of dualism*, which can be expressed in the following verbal form:

If a lot of quality α_1 submitted to the plan $m\|n$ has the probability β_1 of being rejected, then a lot accepted according to the plan $m\|n-1$ has a quality at least α_1 with a probability at least β_1.

This is a consequence of (1), considering 3.1, 3.3, 3.5 and 3.6.

If a lot of quality α_2 submitted to the plan $m+1\|n$ has the probability β_2 of being accepted, then a lot rejected according to the plan $m\|n-1$ has a quality at most α_2 with a probability at least β_2.

This is a consequence of (10), considering 3.2, 3.4, 3.5, and 3.6.

Both rules can be combined into one in a rather descriptive way:

Any sampling plan $m\|n-1$ corresponding to the retrospective parameters $\alpha_1,\alpha_2,\beta_1,\beta_2$ can be split into two sampling plans: $m\|n$ with prospective parameters α_1,β_1, and $m+1\|n$ with prospective parameters α_2,β_2.

5. Applications. When computing the prospective probabilities, it is possible to make use of the tables of the Incomplete Beta-function. For this purpose only the expression for R_{m+1} given in section 4.1 is needed. It is in this way that the Incomplete Beta-function seems to have been used by the Columbia Statistical Research Group in the excellent book *Sampling Inspection* [2]).

5.1. This application, however, is limited to a rather narrow field of small samples, as the maximum size of the sample depends on the range of the tables of the Incomplete Beta-function. For instance, the tables of K. Pearson make it possible to compute probabilities in all sampling plans for sample sizes $n \leqslant 49$. Strictly speaking, it is possible to increase slightly the range of application of the tables, the ultimate limit of sample size being $n=49+m$, where m is the maximum allowable number of defective pieces.

5.2. The inverse application of the rule of dualism, i. e. getting retrospective information from the prospective data, is of far greater importance. This procedure is particularly easy if the expected quality of the lot is high, say $\alpha>0,9$. In this case the expected fraction defective $\omega=1-\alpha$ is low ($\omega<0,1$), this permitting of free application of Poisson's distribution as the limiting form of the binomial. Any table giving summation terms of Poisson's formula is suitable for this purpose. In order to minimize the necessary computation work, Table I has been prepared [3]).

In this table C_1 and C_2 are respectively the expected numbers of defective pieces in the sample, which yield the probability 0,05 and 0,95 respectively that no more than m' defective pieces will occur in the sample.

TABLE I. 0,05 and 0,95 points of Poisson's distribution.

m'	C_1	C_2
0	3,00	0,051
1	4,74	0,36
2	6,30	0,84
3	7,75	1,37
4	9,15	1,97
5	10,51	2,61
6	11,84	3.28
7	13,15	3,98
8	14,14	4,70
9	15,72	5,43
10	16,98	6,17

[2]) McGraw and Hill, New York 1948.

[3]) Based on Molina's tables *Poisson's Exponential Binomial Limit*.

We see from the numerical values of the probabilities that Table I is valid only when $\beta_1=\beta_2=0{,}95$ (standard value adopted by the Polish Standards Committee) — yet other tables may be easily drawn up if required. In Poisson's distribution the probability β depends on the expected number C_1 or C_2 only and on observed numbers m' of defective pieces in the sample. If, then, $\beta=$ const. and $m'=$ const., the expected number must also be constant.

Using index $_p$ for prospective and index $_r$ for retrospective interpretation of the sampling plan, the expected number C_1 or C_2 may be considered as the product of the sample size $n+1$ by the fraction defective ω_p (prospective interpretation of the sampling plan $m\|n+1$).

As given by the rule of dualism, ω_p is at the same time the retrospective fraction defective ω_r in the plan $m\|n+1-1$, i. e. in the plan $m\|n$, provided that the probabilities are the same. Hence

$$\omega_{1r}=C_1/(n+1) \quad \text{and} \quad \alpha_{1r}=1-C_1/(n+1), \tag{11}$$

$$\omega_{2r}=C_2/(n+1) \quad \text{and} \quad \alpha_{2r}=1-C_2/(n+1). \tag{12}$$

Example 1. 3 defective pieces were found in a sample of 150 pieces. Find the lower and the upper limit of the quality of the lot.

Here $m'=3$ and $n=150$. From Table I we get $C_1=7{,}75$ and $C_2=1{,}37$. From (11) we get the lower limit of the quality of the inspected product

$$\alpha_{1r}=1-7{,}75/151=0{,}9487=94{,}87\,\%.$$

From (12) we get the upper limit of the quality of the inspected product

$$\alpha_{2r}=1-1{,}37/151=0{,}9910=99{,}10\,\%.$$

Each of the results $\alpha>\alpha_{1r}$ and $\alpha<\alpha_{2r}$ is valid with the probability 0,95, assuming a uniform distribution of prior qualities in the universe.

5.4. Any one endeavouring to compute directly the probability given by (7), especially with higher values of n (say $n=200$) and of m (say $m=15$), will appreciate the economy of time brought about by the rule of dualism. The integral expands into a series of positive and negative terms, and the result is obtained as a difference of two sums of nearly the same absolute values. In consequence the terms have to be computed exactly up to 20 and more decimal places. Compare it with the simplicity of Example 1!

Even in the cases where Poisson's distribution does not apply because of the high fraction defective, it is simpler to compute

$$\sum_{j=0}^{m}\binom{n}{j}\alpha^{n-j}\omega^{j}.$$

Since $\alpha+\omega=1$, the expression (3) gives immediately

$$R_{m+1}=1-\sum_{j=0}^{m}\binom{n}{j}\alpha^{n-j}\omega^{j}.$$

A similar procedure [4]) may be easily adopted for drawing up tables of the Incomplete Beta-function. It is much simpler than the method of integrating and of recurrence formula, used by K. Pearson in his classical *Tables*.

5.4. As the next application consider the retrospective interpretation of the prospective tables of single sampling plans. In Table II an extract from the collection of such plans made by the Polish Standards Committee is reproduced.

TABLE II. Sample size n. Maximum number of defectives m (ω_1 and ω_2 in $^0/_0$).

n	ω_2	m	ω_1	ω_2	m	ω_1	ω_2	m	ω_1	ω_2	m	ω_1
150	0,24	1	3,16	0,56	2	4,20	0,91	3	5,16	1,31	4	6,10
250	0,34	2	2,52	0,55	3	3,10	1,04	5	4,20	1,59	7	5,26

Thus, for instance, the sampling plan 3‖150 is characterized by two values:

$\omega_1=0{,}91\,^0/_0$ (i. e. $\alpha_1=98{,}19\,^0/_0$) and $\omega_2=5{,}16\,^0/_0$ (i. e. $\alpha_2=94{,}84\,^0/_0$).

These prospective data may be readily used for retrospective information.

[4]) A different approach to this problem may be found in the paper by E. C. Molina, *Application to the Binomial Summation of a Laplacian Method for the Evaluation of Definite Integrals*, The Bell System Technical Journal 8 (1929), p. 99-108, New York.

Example 2. Find the estimated qualities of the lot accepted and of the lot rejected in the sampling plan 3‖150.

Starting from the data of the plan 3‖150 we get for the product accepted in this plan

$$\omega_{1r}=5{,}16\frac{150}{151}=5{,}13\,^0/_0,\qquad \alpha_{1r}=94{,}87\,^0/_0.$$

The quality of the product accepted in the sampling plan 3‖150 is higher than $94{,}87\,^0/_0$, with the probability 0,95 at least.

By computing retrospective data concerning rejection in the sampling plan 3‖150 we must start from the data of the plan 4‖150. We then get

$$\omega_{2r}=1{,}31\frac{150}{151}=1{,}30\,^0/_0,\qquad \alpha_{2r}=98{,}70\,^0/_0.$$

Thus the quality of the product rejected in the sampling plan 3‖150 lies beneath $98{,}70\,^0/_0$, with the probability 0,95 at least.

All the results are valid assuming uniform distribution of prior qualities in the universe.

5.5. As shown by Example 2, the numerical differences between the retrospective and prospective characteristics of the sampling plans are practically negligible if only the sample size is not too low, say, not lower than 25 pieces. Since the sample size ranges up to $n=1500$, it is evident that for most practical applications the retrospective characteristics of sampling plans may be read immediately from the prospective tables.

This is perhaps the most important practical conclusion from the rule of dualism. The requirements of the producer and of the consumer may be specified in prospective or retrospective terms. The meaning of the probabilities is different in each case. Yet if numerical values of the qualities are the same, and if the same probabilities are required, the plans will be the same in both cases. And for the final decision — to accept or to reject — the sampling plan is the only thing that counts.

Example 3. Find the estimated fraction defective of the lot, if in the sample of 250 pieces 7 were classified as defective.

From Table II we immediately find that the fraction defective lies between $5{,}26\,^0/_0$ and $1{,}59\,^0/_0$ (under usual assumptions and neglecting the factor 250/251).

Państwowy Instytut Matematyczny
The State Institute of Mathematics