Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chemometrics and Intelligent Laboratory Systems 75 (2005) 189–200

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

Minimum cost acceptance sampling plans for grain control, with application to GMO detection

A. Kobilinsky[a,*], Y. Bertheau[b]

[a]INRA, Biométrie (BIA), Domaine de Vilvert, 78352 Jouy-en Josas Cedex, France
[b]INRA, PMDV/MDO, RD 10, Route de Saint-Cyr. 78026 Versailles Cedex, France

Received 13 October 2003; received in revised form 29 June 2004; accepted 7 July 2004
Available online 19 September 2004

Abstract

Quality control by attribute [A. Hald, Statistical Theory of Sampling Inspection by Attributes, Academic Press, New York, 1981; E.G. Schilling, Acceptance sampling in quality control. In: Statistics: Textbooks and Monographs, Vol. 42, Dekker, 1982; J.J. Daudin, C.S. Tapiero. Les outils et le contrôle de la qualité. Economica (1996).] may be used with grain lots to control their purity. But usually the control cannot be made on each grain separately. The presence of an impurity is rather assayed in groups of grains the size of which is an important parameter which can be used to find a cost optimal acceptance sampling plan among those which give acceptable consumer's and producer's risks. This group control has been studied for virus or bacterium detection in grains by the Elisa method [Y. Maury, C. Duby, J.M. Bossenec, G. Boudazin, Group analysis using ELISA: determination of the level of transmission of Soybean Mosaic Virus in soybean seed, Agronomie 5, 1985, 405–415; Y. Maury, C. Duby, R.K. Khetarpal, Seed certification for viruses. In: Plant Virus, Disease Control, A. Hadidi, R.K. Khetarpal, H. Koganezawa, eds., APS Press, Chap. 18, 1998, 237–248.] and is advocated by Remund et al. [K. Remund, D. Dixon, D. Wright, L. Holden, Statistical considerations in seed purity testing for transgenic traits, Seed Sci. Res. 11, 2001, 101–119.] for genetically modified organism (GMO) detection. But no optimization method to select the cheapest acceptance single- or double-sampling plan has yet been described.

Given a control cost function depending on the number of groups to analyse and on the total number of grains, we describe in this paper a practical way to get the least expensive acceptance sampling plan keeping both the consumer's and the producer's risks below a predetermined threshold. The method is more specially illustrated by examples in GMO detection.

Keywords: Seed acceptance sampling; Quality control; Operating characteristic curve; Single sampling by attributes; double sampling by attributes; Group analysis; GMO detection

1. Introduction

In grain production and commercialisation, the purity and absence of defect of a grain lot processed has to be checked for various criteria. For instance, its ability to germinate, its belonging to the indicated variety, its good health state, the absence of toxics like mycotoxin, etc. Recent advances in agricultural biotechnology have produced new transgenic varieties whose characteristics have to be thoroughly studied before they can be authorized for commercialisation. Hence, there is also an important need to control that no such unauthorized genetically modified organism (GMO) exists in the commercialised lots of grains, or when legislation makes the labelling compulsory that it has been adequately labelled.

In some case, the analysed criteria can be evaluated by a quantitative response. For instance, the optical density in an ELISA method, the cycle threshold (Ct) in a real time polymerase chain reaction (RtPCR), etc. can be calibrated to predict microorganism or GMO concentration. But doing that rigorously implies a good knowledge of the distribution of the response and its uncertainty.

* Corresponding author. Tel.: +33 1 34 65 22 30; fax: +33 1 34 65 22 17.
E-mail address: kobi@jouy.inra.fr (A. Kobilinsky).

0169-7439/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2004.07.005

An easier and more robust way of dealing with such a measurement is to transform it into a binary response such as presence or absence. In quality control, this binary response is called an *attribute* and a control based on it is a *control by attributes* (Refs. [1–3]). The sampling scheme then consists in collecting an *initial sample*, which is sent to the laboratory for analyses. From this initial sample, the laboratory extracts one or several subsamples and determines the presence or absence for each of them. It is from these binary responses that the control is made. The procedure used to get the initial sample is called here the *initial sampling protocol*, while the procedure used by laboratories to control this initial sample is called the *acceptance sampling plan*. A somewhat different terminology appears in Ref. [6] who distinguish the *testing plan* from the *sampling plan*, but we have chosen here a terminology more in accordance with the classical statistical one (see Ref. [2] for instance). Note that grain is used here as a substitute for *seeds and grains*.

Standardised initial sampling protocol for grains are available [7] and so they will not be taken into consideration here. Their aim is to get a first initial sample as representative as possible of the whole lot. This is important as the acceptance sampling plan further proceeds only from this initial sample and bases the acceptance entirely on it. If the initial sample is representative, it can be assumed that each subsample behaves as a simple random sample from the whole lot.

Several acceptance sampling plans using the control by attribute of the subsamples are proposed in Ref. [6] especially for detection of transgenic traits (see also Ref. [8]). Similar acceptance sampling plans had been described previously for the detection of viruses or bacteria by Refs. [4,5]. The subsamples are called *pools* in one case [6], *groups* in the other [4], but the approaches described in these articles are essentially the same. Following Ref. [4], we shall refer to this kind of control as *group analysis*. From the classical quality control point of view, each subsample is a *sampling unit* [9]. It is however important to notice that the sizes of the different subsamples need not be identical and are important to take into account.

Acceptance sampling plans using the control by attributes are also recommended by GIPSA and used by numerous seed producers and grain exporters. Consider for instance the control described and recommended by Ref. [10] for the detection of StarLink, a GMO maize developed by Aventis CropScience, which produces a protein Cry9C with insecticidal properties effective in controlling the European corn borer and which was approved for use as animal feed only. In that case, the GIPSA proposal to check that the lot does not include StarLink grains is to ground and assay enough grains to give the consumer, if no GMO is found in them, a very low probability of accepting a lot with an unacceptable GMO rate. For instance, if a proportion above 0.3% is considered unacceptable, 2300 grains have to be used to get a 0.1% probability of accepting a lot reaching this 0.3% GMO rate. Since detection is made in that case through the expressed protein Cry9C, it is recommended to perform the analysis on separate subsamples of 400 grains insuring that even 1 GM grain in the subsample is reliably detected. The lot is accepted if none of the subsamples contains GMO. Thus, the GIPSA procedure also uses subsamples and their attributes, but the size 400 of these subsamples is only dictated by the limit of the analytical method of detection and does not take into account the producer's risk to see his lot rejected when there is a slight fortuitous GMO presence.

Multiple sampling plans with more than two stages would be unpractical in this context, so we have limited the investigation to the single- and double-sampling plans [2] (called single-stage or double-stage testing plan by Ref. [6]).

For them, we describe a practical algorithm to get the cheapest sampling plan keeping both the consumer's and the producer's risks below predetermined small levels.

For ordinary quality control, practical algorithms to find suitable single- or double-sampling plans have been proposed ([11,12]) and computer programs are available (see the Journal of Quality Technology or Ref. [13]). But the approach is here complicated by the fact that the number of subsamples and their size are simultaneously optimized. Moreover, in double-sampling plans, the size needs not be the same in the first and second stage. Modifying it between the two stages can offer a lot more flexibility and consequently decrease costs.

As indicated by its name, double sampling proceeds in two stages. A few subsamples are first assayed. A second more important set of subsamples is then assayed only if the first results leave some doubt. Since in many cases the first assays are conclusive, this double sampling may achieve the same risk control with a cost far smaller than the single sampling which assays all subsamples simultaneously [14].

To make things clearer, the methodology is illustrated throughout the paper by the case of GMO detection. But it can be easily adapted to any other detection provided there is a way of determining the presence or absence of the incriminated analyte, microorganism, etc in a subsample of grains. In the case of GMO, this determination is generally made either through a protein specific of the GMO trait, or by PCR amplification and subsequent detection of the corresponding DNA sequence.

Note that GMO control is compulsory in numerous countries including those of the European Union. Since the introduction of the novel food and ingredient 258/97/EC directive, all foods containing GMO or derived compounds should be labelled above a 1% threshold of fortuitous presence in each ingredient (directive 49/2000/EC).

2. Single and double sampling

In a single-sampling plan, N groups of n grains are separately ground and analysed to determine if they are

GMO-positive or not. If there are X positive out of the N groups the GMO proportion may be estimated by X/N and the lot is accepted if $X \leq A$, rejected if $X > A$, where A is a predetermined acceptance threshold. Thus, the set of parameters defining the plan is (N,n,A).

In the case of grains, the double sampling is defined as follows. First, N_1 groups of n_1 grains are assayed. The lot is accepted if $X_1 \leq A_1$ where X_1 is the number of positive groups and A_1 a predetermined threshold. If $X_1 \geq R_1$, where R_1 is a predetermined rejection threshold, the lot is rejected. In between, that is if $A_1 < X_1 < R_1$, N_2 new groups are assayed. The lot is then accepted if $X_2 \leq A_2(X_1)$, where X_2 is the number of positive among the N_2 new groups and $A_2(X_1)$ a predetermined threshold. The function $A_2: X_1 \mapsto A_2(X_1)$ giving for each value of X_1 the acceptance threshold at the second step must be a decreasing function of X_1 because the bigger is X_1, the smaller must be X_2 to compensate. In classical quality control, this function has the form $A_2 - X_1$ where A_2 is a fixed number. The lot is therefore accepted at the second step if $X_2 + X_1 \leq A_2$ and rejected if $X_2 + X_1 \geq R_2 = A_2 + 1$. Since we allow here the subsamples examined at steps 1 and 2 to have possibly different numbers n_1 and n_2 of grains, it seems natural as a consequence to adopt a larger frame making possible the use of something less symmetric than the sum $X_1 + X_2$ to base the acceptance at the second step. So the set of parameters to determine in that case is $T = (N_1, n_1, A_1, R_1, N_2, n_2, A_2())$. The brackets following A_2 are put to remind that A_2 is a function.

3. Consumer's and producer's risks

The aim of a control is to insure the consumer that the lot is rejected with high probability whenever the proportion of GM grains is above a *non-tolerable* threshold p_{nt}, called *Limiting Quality Level* (LQL) or *Consumer Quality Level* (CQL) in quality control. But for the producer's sake, there must be also a high probability to accept if the GMO proportion is below a *tolerable* threshold p_t called *Acceptable Quality Level* (AQL) or *Producer Quality Level* (PQL) in quality control. Of course, the tolerable threshold must be smaller than the non-tolerable one, that is $p_t < p_{nt}$; otherwise, there cannot be any agreement between the consumer and the producer.

The consumer's requirement can thus be formalised by the inequality

$$\text{Prob}(\text{acceptance} | p_{nt}) \leq \beta \quad (1)$$

which means that the *risk* for the consumer to accept the lot if the GMO proportion reaches the non-tolerable threshold must be smaller than β. Similarly, the producer's requirement is expressed by the inequality

$$Prob(\text{rejection} | p_t) \leq \alpha \quad (2)$$

meaning that his risk that the lot is rejected if the GMO proportion is the tolerable one p_t does not exceed α. These maximum tolerated risks β and α are called *consumer's* and *producer's risks*, respectively.

The choice of p_{nt}, β, p_t, α depends on many considerations: risk for humans, risk for environment, degree of purity that can be obtained, nature of the grains (commercial seeds, basic seeds, breeder's seeds, grain for human food, for animal feed, etc.). Once there is an agreement between contractors on their choice, the problem is to find the acceptance sampling plan leading to a minimal inspection cost among those which satisfies the constraints (1) and (2).

Note that taking into account the producer's risk appears essential in many cases. First, some GM plants are very scattered today (e.g. Round-up Ready Soja for instance) and it is almost impossible to avoid any trace of them, e.g., in big cargos. This small fortuitous presence of GMO has been often observed. It is taken into account by local regulations, which have generally adopt low but positive thresholds when there are no safety issues. One reason for it is that the dissemination of pollen by the wind or the insects cannot be completely controlled. This is why the producer's point of view has to be taken into account. As outlined by Ref. [15], plans requiring no defects for acceptance can be used only if the state of the arts permits near perfect quality level, and this is not the case here.

Even for GMO free supply chains, the producer's risk has to be taken into account. Otherwise, the fortuitous presence of GMO would lead to many lot rejections, hence to a considerable price increase that would finally entail the suppression of these chains, hence the exact opposite of what is wished by consumers wanting to avoid new GMO products.

The control cost, in which we include sampling, mainly depends on the number of assays and grains. In the next section, its expression is formalised for the two kinds of sampling plans considered, the single and the double.

4. Cost function

4.1. Single-sampling plan

We let C_a be the cost of assaying one group, that is the cost of the whole sequence of operations used to detect GMO in this group (grinding, protein or DNA extraction, immunological or PCR method). Since in single sampling there are N such assays, the global cost of the assays is NC_a. The cost of the grains used for these assays must be added. This cost is $C_g Nn$, where C_g denotes the cost of one grain, which strongly depends on the nature of the grain (ordinary grains, commercial seeds, breeder's seed, etc.). The whole cost in single sampling therefore takes the form

$$C(N,n) = C_a N + C_g N n. \quad (3)$$

It is always possible to assume that C_a=1, that is to take the assay cost as unit price. The cost C_g is then relative to this assay cost. For instance, C_g=0.001 if the price of one grain is 1/1000 of the assay price, or equivalently if the cost of an assay is equivalent to the cost of 1000 grains.

4.2. Double-sampling plan

In double sampling, the cost is random since the second step only occurs if $A_1<X_1<R_1$, hence with a probability denoted by Prob(step 2|p), which depends on the real GMO proportion p. It is the expectation of this cost for a given p, which has to be optimized. This expectation is

$$\begin{aligned} &C(N_1,n_1,A_1,R_1,N_2,n_2,A_2())\\ &\quad = C(N_1,n_1) + \text{Prob(step 2}|p) \quad C(N_2,n_2)\\ &\quad = C_aN_1 + C_gN_1n_1 + \text{Prob(step 2}|p)\ \left(C_aN_2 + C_gN_2n_2\right) \end{aligned} \tag{4}$$

where C is defined as in single sampling by Eq. (3). In this form, it is assumed that the N_2n_2 grains needed for the second step have to be paid only if this second step occurs. In practice, it may be necessary to get and pay the N_2n_2 grains of the second stage even if they remain unused, in order to avoid a second sampling in the lot and because it may be difficult to return them. The cost then takes the following form

$$\begin{aligned} &C(N_1,n_1,A_1,R_1,N_2,n_2,A_2())\\ &\quad = C_g(N_1n_1 + N_2n_2) + C_a(N_1 + \text{Prob(step 2}|p)N_2) \end{aligned} \tag{5}$$

Though we consider here only the first form (Eq. (4)) of the cost, it is not difficult to adapt the methodology to the second form.

The parameter p appearing in Eqs. (4) and (5) is unknown. To get an idea of how the optimal plan depends on it, two values p_t (the tolerable threshold) and p_0=p_t/10 will be given to it in this paper. Similarly, the ratio C_g/C_a giving the cost of a grain relatively to the cost of assay of a subsample will also be given two values 0.001 and 0.01. To know which values of p and C_g/C_a are used, the cost defined by Eq. (4) with given p, C_g and C_a=1 is denoted by $\mathcal{C}$ (p,C_g) instead of $\mathcal{C}$ alone. Thus

$$\begin{aligned} &C\left(p,C_g\right)(N_1,n_1,A_1,R_1,N_2,n_2,A_2())\\ &\quad = N_1 + C_gN_1n_1 + \text{Prob(step 2}|p)\left(N_2 + C_gN_2n_2\right) \end{aligned} \tag{6}$$

In the sequel, the set of parameters T=(N_1, n_1,A_1, R_1,N_2,n_2,A_2()) will sometimes be omitted if the context clearly shows which one is used.

5. Cost minimization of group analysis control by attributes

5.1. Case of the single sampling

Let p be the proportion of GM grains in the lot. As mentioned in the introduction, it is assumed that each grain in the initial sample has probability p to be GMO, independently of the other grains. The probability that no grain is GMO in a group of n grains is $(1-p)^n$ and, consequently, if the GMO detection procedure is error-free, the probability to find a group positive, that is containing at least one GM grain is

$$P = 1 - (1-p)^n. \tag{7}$$

Section 5.2 explains how to take into account possible errors of detection by replacing P by the probability $\tilde{P}$ to detect, either rightly or wrongly.

The number X of groups, among the N, containing at least one GMO grain follows a binomial law $\mathcal{B}$(N,P). Assuming that any GMO grain is detected with certainty by the procedure, the probability of acceptance and rejection are respectively

$$\begin{aligned} &\text{Prob}(X\le A|p) = \sum_{j=0}^{A}\binom{N}{j}P^j(1-P)^{N-j},\\ &\text{Prob}(X>A|p) = 1 - \text{Prob}(X\le A|p). \end{aligned} \tag{8}$$

The consumer's requirement (Eq. (1)) is that the acceptance probability is smaller then β when p=p_{nt}, the producer's one (Eq. (2)) that the rejection probability is smaller than α when p=p_t:

$$\text{Prob}(X\le A|p_{\text{nt}})\le\beta, \quad \text{Prob}(X>A|p_t)\le\alpha. \tag{9}$$

These probabilities are given by Eq. (8) where P takes the corresponding values P_{nt} and P_t:

$$P_{\text{nt}} = 1 - (1-p_{\text{nt}})^n, \quad P_t = 1 - (1-p_t)^n.$$

The problem is to find N, n and A minimizing the quantity (3) and satisfying the constraints (9).

For n fixed, Guenther's algorithm [11] provides the solution, that is the minimum acceptable N and corresponding value of A such that the constraints (9) are satisfied. Fig. 1 represents these values in function of n when

$$p_{\text{nt}} = 1\%,\ \beta = 5\%,\ p_t = 0.2\%,\ \alpha = 5\%. \tag{10}$$

These parameters will be used to illustrate the methodology throughout this article. Note that the 1% value for p_{nt} is the current European threshold for food labelling.

The minimum acceptable number of groups N is always greater or equal to 7. The smallest value of n leading to this minimum is 204. The corresponding acceptance threshold is A=4.

The corresponding acceptance sampling plan assays the presence of GMO in 7 groups of 204 grains and accept the

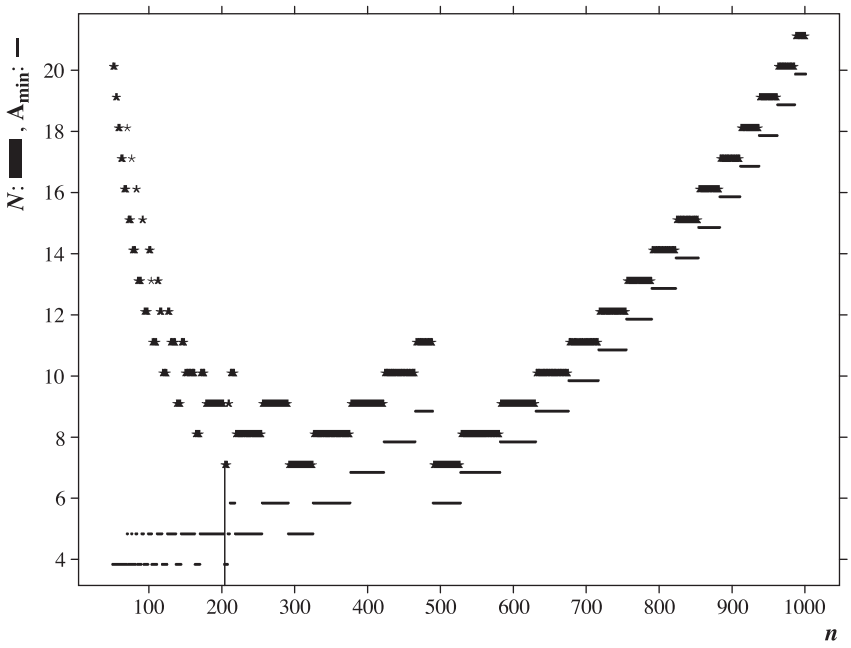

Fig. 1. Minimum N and corresponding A for each n, when the constraint parameters are given by the equalities (10). The graphic gives an information similar to that of Table 1. The constraints remain those defined by the equalities (10). The thick line gives for each n the minimum acceptable N, while the thin line gives the corresponding A. The minimum N of 7 is obtained when n=204 as in Table 1. The corresponding value of A is 4. A vertical line has been drawn at this abscissa.

lot if the number X of groups where GMO is detected is below 4: $X \leq 4$. This plan guarantees that the consumer's risk to accept a lot with more than 1% GMO and the producer's risk to reject a lot with less than 0.2% GMO are both below 5%.

Note that this acceptance sampling plan minimizes N but not necessarily the control cost (3). Consider for instance the two values of C_g introduced in Section 4. If $C_g = C_a/1000$, this plan minimizes the cost function (3) but if $C_g = C_a/100$ then the cheapest choice is N=8, n=164, A=4. Table 1 illustrates this point.

5.2. False positive and false negative

The assay method sometimes gives a false result, that is find GMO when there is no GMO grain in the group –false positive– or do not find GMO though there is some in the group –false negative–. The origin of such errors and the way to reduce them is thoroughly discussed in Ref. [6]. To take them into account, one has to replace the probability P that a group is GMO positive by the probability $\tilde{P}$ that the group is found GMO positive by the assay.

Let δ be the probability to find positive a group without GM grain and λ the probability to find negative a group including at least one GM grain. Then

$$\tilde{\mathbf{P}} = P(1-\lambda) + (1-P)\delta = (1-(1-p)^n)(1-\lambda) + (1-p)^{n\delta}$$

It is natural to assume δ constant. As to λ, it may also be assumed independent from the GMO fraction in the subsample if this fraction is not too small, which will systematically be the case when n is not too big. For instance, it is known that StarLink detection through the protein Cry9C correctly works as far as the GMO rate is not lower than 1 GMO grain in 400. So λ may be assumed constant for group sizes below 400, but for bigger values it should be considered as a function of the GMO fraction. The probability $P(1-\lambda)$ that a group

Table 1
Minimum N and associated costs for some values of n, when equalities (10) are satisfied

n	60	100	140	164	180	200	203	204	208	209
N	18	14	9	8	9	9	9	7	7	9
A	4	5	4	4	5	5	5	4	4	5
Nn	1080	1400	1260	1312	1620	1800	1827	1428	1456	1881
$C(N,n)=N+0.001Nn$	19.1	15.4	10.3	9.3	10.6	10.8	10.8	**8.4**	8.5	10.9
$C(N,n)=N+0.01Nn$	28.8	28.0	21.6	**21.1**	25.2	27.0	27.3	21.3	21.6	27.8

Example: if each group has n=204 grains, the consumer's and producer's requirements can be simultaneously satisfied only if $N \geq 7$, and for N=7, A must be equal to 4. However, if the grain cost is, e.g., 1% of a one group assay cost, hence the global cost proportional to $N+0.01Nn$, then it is cheaper to use N=8 groups of 164 grains than N=7 groups of 204 grains.

includes GMO and that the assay correctly detects it should then be replaced by

$$(1-\lambda(1/n))\binom{n}{1}p(1-p)^{n-1} + (1-\lambda(2/n))\binom{n}{2}p^2(1-p)^{n-2}+\cdots.$$

To take into account the assay deficiencies, δ and λ have first to be evaluated, for instance by assaying groups with known GMO fraction. Once they are known, the relation between p and $\tilde{P}$ can be used to compute the risks for given p_{nt} and p_t and therefore to find a suitable sampling plan as previously.

5.3. Case of the double sampling

The double sampling allows to reach the same risk targets with a much lower cost. Let

$$P_1 = 1-(1-p)^{n_1}, \quad P_2 = 1-(1-p)^{n_2}. \tag{11}$$

The probability to accept the lot at the first step is

$$\mathrm{Prob}(X_1 \le A_1 | p) = \sum_{j=0}^{A_1}\binom{N_1}{j}P_1^j(1-P_1)^{N_1-j} \tag{12}$$

There is a second step only if X_1 takes a value j such that $A_1<j<R_1$. In that case, the lot is accepted if $X_2 \le A_2(j)$, hence with probability

$$\mathrm{Prob}(X_2 \le A_2(j)|p) = \sum_{k=0}^{A_2(j)}\binom{N_2}{k}P_2^k(1-P_2)^{N_2-k}. \tag{13}$$

As the probability to get $X_1=j$ and $X_2 \le A_2(j)$ is $\mathrm{Prob}(X_1=j|p)\mathrm{Prob}(X_2 \le A_2(j)|p)$, the probability of acceptance when $A_1 < X_1 < R_1$ is

$$\sum_{j=A_1+1}^{R_1-1}\mathrm{Prob}(X_1=j|p)\mathrm{Prob}(X_2 \le A_2(j)|p)$$

and finally the global probability of acceptance, denoted by $P_a(p)$, is

$$\begin{aligned} P_a(p) &= \mathrm{Prob}(X_1 \le A_1|p) + \sum_{j=A_1+1}^{R_1-1}\mathrm{Prob}(X_1=j|p)\mathrm{Prob}(X_2 \le A_2(j)|p) \\ &= \sum_{j=0}^{A_1}\binom{N_1}{j}P_1^j(1-P_1)^{N_1-j} + \sum_{j=A_1+1}^{R_1-1}\binom{N_1}{j}P_1^j(1-P_1)^{N_1-j} \\ &\quad \times \sum_{k=0}^{A_2(j)}\binom{N_2}{k}P_2^k(1-P_2)^{N_2-k}. \end{aligned} \tag{14}$$

5.4. Finding an optimal double-sampling plan

To determine an optimal double-sampling plan in classical quality control, Daudin et al. [16] first vary the parameters of the acceptance sampling plan continuously in $\mathbb{R}$, then select integer values in the vicinity of the real optimal solution. In GMO control, the cost is strongly dependent on the number of assayed groups. Therefore only small values of N_1 and N_2 have to be examined and it is possible to work directly on sets of integer parameters. The number of possible values for n_1, n_2 is however too important to examine all of them. So the optimization is carried out in two steps. First, for each possible set of parameters $T=(N_1,A_1,R_1,N_2,A_2())$, the optimal couple (n_1,n_2) satisfying the constraints, if any, is looked for and the associated cost $\mathcal{C}(T)$ computed. Then the optimal set T is obtained by comparison between these $\mathcal{C}(T)$ for all possible T. This may be done for several values of p to get an idea of the sensibility of the optimum to the percentage of GMO grains in the lot.

Fig. 3 illustrates the constraints on (n_1,n_2) for two choices of $T=(N_1,A_1,R_1,N_2,A_2())$, with the constraint parameters (10). To satisfy the consumer's and producer's constraints

$$P_a(p_{nt}) \le \beta, \tag{15}$$

$$1-P_a(p_t) \le \alpha, \tag{16}$$

the point (n_1,n_2) must be above the curve B (Buyer's=consumer's curve) defined by $P_a(p_{nt})=\beta$ and below the curve S (Seller's=producer's curve) defined by $1-P_a(p_t)=\alpha$ (the results of Appendix A.1 explain why). If the area thus defined is empty as in Fig. 3b, there is no couple (n_1,n_2) satisfying the constraints. When it is not empty as in Fig. 3a, the following step is to find the couple giving the minimum cost.

Fig. 3 suggests the following specific algorithm to determine if the constraint area is empty or not. The minimum of the difference curve $D=B-S$ is looked for. This difference D appears to be first decreasing, then increasing in function of n_1. It is easy to find its minimum as the point where its derivative in function of n_1 is zero (the derivative is given in Appendix A.3). If B is below S at this minimum, there is a couple (n_1,n_2) satisfying the constraints and the constraint area, bordered at its top by S and at its bottom by B, is not empty. In that case, the value of n_1 corresponding to the point at the extreme left of the constraint area can be easily found as the value such that $B-S=0$.

For n_1 fixed, it appears from Eq. (6) that the optimum n_2 is the smallest one compatible with the constraints, that is the one associated to n_1 on the B curve. So the following optimization step is to find the minimum cost along the consumer's curve B.

In most cases, the cost $\mathcal{C}$ given by Eq. (6) appears to be first decreasing, then increasing when moving from the left to the right on the consumer's curve B. It is then easy to find

the zero of its derivative $d\mathcal{C}/dn_1$ with respect to n_1 (which is given in Appendix A.3). The corresponding point on the B curve gives the searched optimum. In some case, however, it gives only a local optimum but it was observed that this occurs only for parameters N_1, A_1, R_1, N_2, A_2 which are far from the optimum. Fig. 2 shows the variation of the cost $\mathcal{C}(p,C_g)(T)$ in function of n_1 on the B curve for some values of p, C_g and set of parameters $T=(N_1,A_1,R_1,N_2,A_2())$.

When $C_g=0$ and p is small, the derivative $d\mathcal{C}/dn_1$ is already positive at the leftest point of the constraint area. This point is then optimal. This is for instance the case in Fig. 3a. The optimum is obtained for $n_1=393$, $n_2=330$ when $C_g=0$ and $p=p_0$. To explain this, note that when $C_g=0$, the cost (6) becomes $\mathcal{C}(p,0) = N_1 + \text{Prob}(\text{step } 2|p)N_2$. It does not depends on n_2 and its dependence on n_1 is only through Prob(step 2|p) which can be written

$$\text{Prob}(\text{step } 2|p) = \sum_{j=A_1+1}^{R_1-1} \binom{N_1}{j} P^j (1-P)^{N_1-j},$$

$$\text{where } P = 1-(1-p)^{n_1}. \qquad (17)$$

Lemma A.1 shows that this quantity is an increasing function of P if P is sufficiently small. Hence, the positivity of $d\mathcal{C}/dn_1$ when n_1 is the abscissa of the leftest point of the constraint area and p small enough.

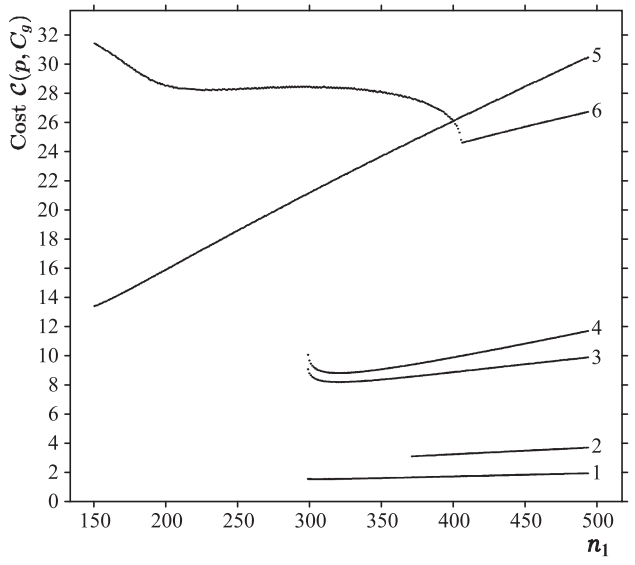

Fig. 2. Variation of the cost $\mathcal{C}(p, C_g)$ along the B curve for some double-sampling plans

Curve	p	C_g	N_1	A_1	R_1	N_2	A_2
1*	0.0002	0.001	1,	0,	2,	6,	$A_2(1)=4$
2	0.0020	0.000	2,	1,	3,	5,	$A_2(2)=4$
3	0.0020	0.001	1,	0,	2,	9,	$A_2(1)=8$
4	0.0002	0.010	1,	0,	2,	10,	$A_2(1)=9$
5*	0.0020	0.010	4,	1,	5,	5,	$A_2(2,3,4)=2,1,0$
6	0.0020	0.010	3,	0,	3,	10,	$A_2(1,2)=9,6$

The conditions marked by a * are optimal for the corresponding p and C_g (see Table 2). Note that curve 6 exhibits a local optimum around $n_1=220$, but that the corresponding parameters give a much more expensive plan than those associated with curve 5.

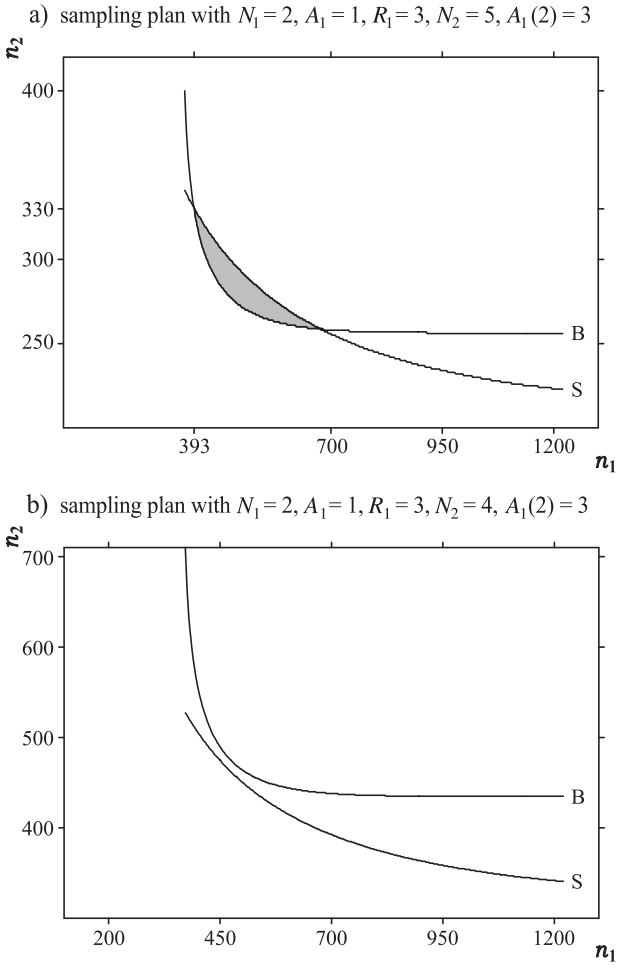

Fig. 3. Constraints for n_1 and n_2 associated with the parameters in (10)

B: buyer's or consumer's curve defined by Prob(acceptation|p_{nt})=β
S: seller's or producer's curve defined by Prob(rejection|p_t)=α

To satisfy the consumer's constraint, the couple (n_1,n_2) must be above curve B. It must be below curve S to satisfy the producer's constraint. Thus, B and S are respectively the bottom and top borders of the constraint area. Figure a illustrates a situation where this admissible area, below S and above B, is not empty, while figure b illustrates on the contrary a situation where the consumer's and producer's constraints cannot be simultaneously satisfied. The admissible area in figure a is colored in light grey.

It is crucial in the sequence of operations leading to the optimum choice of (n_1,n_2) to have an efficient way of determining the n_2 associated to a given n_1 on the S and B curves. These values are found as the solutions of the equations $1-P_a(p_t)=\alpha$ and $P_a(p_{nt})=\beta$. It is shown in Appendix A.1 that for fixed values of the other parameters, the acceptance probability $P_a(p)$ is a decreasing function of n_2. This result can be used to solve efficiently these equations in n_2.

Example. The algorithm is used with the parameters defined by the equalities (10). The number of groups N_1 is varied between 1 and 4 and A_1, R_1 among all possible compatible sets of values. Then, N_2 is varied between 1 and 10 and A_2 among all compatible decreasing acceptance function A_2.

There are 6017 possible combinations (N_1,A_1,R_1,N_2,A_2) to explore. For 3899 of them, $D=B-S$ is negative at its minimum and so there exists a couple (n_1,n_2) satisfying the consumer's and producer's constraints. For each of these 3899 combinations, the couple (n_1,n_2) at the left of the constraint area, i.e., with minimum n_1, is first looked for. It appears that this couple minimizes the cost function $\mathcal{C}(p,0)$ given by Eq. (6) for the two values p_t and $p_0=p_t/10$ of p introduced in Section 4.

Then, optima are searched along the B curve for the cost functions $\mathcal{C}(p,C_g)$ where p takes the values p_0=0.02%, p_t=0.2% and C_g the values 1/1000, 1/100 (C_g is expressed as in Eq. (6) with the price of the group assay as unit price, i.e., C_a=1).

The whole search takes less than half an hour on modern microcomputers. The bigger part of this time is used to get the 3899 admissible conditions with the couple (n_1,n_2) at the left of the constraint area, and the rest to get the corresponding optimal n_1 for the four costs with $C_g\neq 0$. In many cases, it can be immediately checked that these last ones coincide with the former value n_1. Table 2 gives the optimal plans thus obtained. Note that most of the time is spent on condition N_1=4 which appears interesting only to minimize the cost computed with $p=p_t$, C_g=0.01.

The statistical properties of an acceptance sampling plan are well summed up by the operating characteristic curve representing the acceptance probability $P_a(p)$ in function of p. This probability is given by Eq. (14) in the double sampling, by Eq. (8) in the single one. Fig. 4 gives this curve for some single- and double-sampling plans. The double-sampling plan 4 in this figure uses only one assay in the first step and goes to step 2 with probability 0.06 if $p=p_0$=0.0002 and 0.49 if $p=p_t$. It is far less expensive than the single-sampling optimal plan 3 with N=7 which gives a similar operating characteristic curve. The single-sampling plan 1 would be adopted to limit to 5% the risk of accepting a lot with 1% GMO content if, like in Ref. [10], one does not care about the producer's risk.

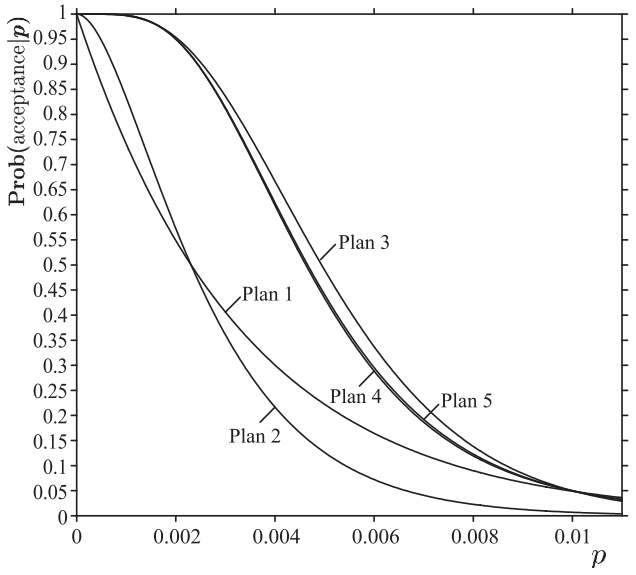

Fig. 4. Some acceptance sampling plans operating characteristic curves respecting the constraints associated with the equalities (10)

Plan 1: single sampling N=1, n=300, A=0
Plan 2: single sampling N=3, n=300, A=1
Plan 3: single sampling N=7, n=204, A=4
Plan 4: double sampling N_1=1, n_1=333, A_1=0, R_1=2, N_2=6, n_2=339, $A_2(1)$=4
Plan 5: double sampling N_1=2, n_1=393, A_1=1, R_1=3, N_2=5, n_2=330, $A_2(2)$=3

Plans 4 and 5 are the double-sampling plans respectively minimizing the costs $\mathcal{C}$(0.02%,0.001) and $\mathcal{C}$(0.2%,0.001) under the consumer and producer constraints defined by equalities (10) (see Table 2). Plan 3 gives the cost-optimal single-sampling plan when C_g/C_a=1/1000 under the same constraints (see Table 1). The single-sampling plan 1 takes only the consumer constraint into account, while the plan 2 shows what can be achieved with a single sampling if no more than three groups are to be assayed.

Thus, an appropriate choice of an acceptance sampling plan by attribute can lead to a quite economical control taking into account both the consumer's and producer's requirements. In some cases, this plan can lead to acceptance even if

Table 2
Optimal parameters for some cost function $\mathcal{C}(p,C_g)$ in double sampling

Cost $\mathcal{C}(p,C_g)$	Parameters of the plan							(1)	Cost	(2)
	N_1	n_1	A_1	R_1	N_2	n_2	A_2			
$\mathcal{C}(0.02\%,0)$	1	333	0	2	6	339	(4)	0.06	1.39	
$\mathcal{C}(0.02\%,0.001)$	1	333	0	2	6	339	(4)	0.06	1.85	(0.49, 5.24 for p=0.2%)
$\mathcal{C}(0.02\%,0.01)$	1	314	0	2	9	182	(4)	0.06	5.68	
$\mathcal{C}(0.2\%,0)$	2	392	1	3	5	601	(4)	0.30	3.48	
$\mathcal{C}(0.2\%,0.001)$	2	393	1	3	5	330	(3)	0.30	4.76	(0.01, 2.82 for p=0.02%)
$\mathcal{C}(0.2\%,0.01)$	4	155	1	5	5	147	(2,1,0)	0.29	13.79	

col (1): Prob(step 2|p).
col (2): gives the probability to go to step 2 and the cost for the other value of p (the one not used for the optimization).
Example: if the mean GMO rate is p=0.02% and the cost of a grain about 1% of the cost of one group assay, the more economical double-sampling plan assays first 1 group of 314 grains. The lot is accepted if the result is negative, otherwise 9 other groups of 182 grains are assayed and the lot accepted only if less than 4 of them are GMO positive.

some assayed subsamples are GMO positive. This fact must be explained to users as perfectly coherent with the agreement reached by producers and consumers on the choice of a non-zero tolerable threshold p_t. These plans have still to be compared to procedures using the very sensitive quantitative response provided by quantitative real time PCR, which could be more adequate in some cases but need more work to know which factors are influencing the distribution of the quantitative response.

Acknowledgment

The authors thank Professor Max Feinberg for his very efficient help in revising this paper.

Appendix A. Appendices on double-sampling optimization

A.1. Some results about the acceptance probability in double sampling

The description of the constraint area for (n_1,n_2) made after Eqs. (15) and (16) relies on the fact that the probability of acceptance $P_a(p_{nt})$ given by Eq. (14) is a decreasing function of n_2. This fact is also used to find for each n_1 the corresponding n_2 on the curves B and S, i.e., $B(n_1)$ and $S(n_1)$.

Indeed, the greater is n_2, the greater is the probability $P_2=1-(1-p_{nt})^{n_2}$ that GMO grains are detected in a group of the second stage, hence the greater is the probability of rejection at step 2 and the lower is the probability of acceptance. In fact, the next proposition gives a little more general result which will be used to prove that for given values of the other parameters, the probability of acceptance $P_a(p_{nt})$ is also a decreasing function of n_1.

Proposition A.1. *Let* $\phi(P)=\sum_{j=0}^{N} q_j \binom{N}{j} P^j(1-P)^{N-j}$ *where* $q_0\geq q_1\cdots\geq q_N$ *are non-increasing weights. Then,* ϕ *is a decreasing function of P.*

Proof. Let $\phi_j(P)$ be the probability that $X=j$ where X follows the binomial law $\mathcal{B}(N,P)$:

$$\varphi_j(P)=\binom{N}{j}P^j(1-P)^{N-j}$$

Then, $\phi(P)=\sum_{j=0}^{N} q_j\phi_j(P)$. The derivative of $\phi_j(P)$ is

$$\begin{aligned}\phi_j'(P)&=\binom{N}{j}\left(jP^{j-1}(1-P)^{N-j}-(N-j)P^j(1-P)^{N-j-1}\right)\\&=\binom{N}{j}P^{j-1}(1-P)^{N-j-1}(j(1-P)-(N-j)P)\\&=\binom{N}{j}NP^{j-1}(1-P)^{N-j-1}\left(\frac{j}{N}-P\right).\end{aligned}$$

If $0<P<1$, the formula clearly holds for $j=0$ or $j=N$. This derivative is positive if $P\leq j/N$, negative otherwise. Let i be the smallest index j such that $P\leq j/N$. Then, for $j<i$, $\phi_j'(P)<0$ and $q_j\geq q_i$ so that $q_j\phi_j'(P)\leq q_i\phi_j'(P)$. For $i\leq j$, $\phi_j'(P)\geq 0$, $q_j\leq q_i$ and the same inequality $q_j\phi_j'(P)\leq q_i\phi_j'(P)$ is consequently satisfied. Hence

$$\phi'(P)=\sum_{j=0}^{N} q_j\phi_j'(P)\leq\sum_j q_i\phi_j'(P)=q_i\sum_j\phi_j'(P)$$

But the latter sum is 0 since $\sum_j\phi_j(P)=1$. Hence, $\phi'(P)$ is negative and the proposition is proved.

We restate as a lemma a result of the proof which will be used later.

Lemma A.1. *The derivative of* $\phi_j(P)=\binom{N}{j}P^j(1-P)^{N-j}$ *is*

$$\phi_j'(P)=\binom{N}{j}NP^{j-1}(1-P)^{N-j-1}\left(\frac{j}{N}-P\right).$$

This proposition can be used to prove that the acceptance function $P_a(p)$ given by Eq. (14) is a decreasing function of n_2. Note first that the sum on the right of Eq. (13) can be written

$$\sum_{k=0}^{N_2} q_k\binom{N_2}{k}P_2^k(1-P_2)^{N_2-k}$$

where $q_k=1$ for $k\leq A_2(j)$ and $q_k=0$ otherwise. By the proposition, it is a decreasing function of P_2, hence of n_2 by Eq. (11). Then, since $P_a(p)$ is a linear combination of such sums with positive weights independent of n_2, it is also a decreasing function of n_2.

From formula (14), it also follows that $P_a(p)$ is a decreasing function of n_1 when the other parameters, including n_2, are fixed. Indeed, let $q_j=1$ if $j\leq A_1$, $q_j=\mathrm{Prob}(X_2\leq A_2(j)|p)$ if $A_1<j<R_1$ and $q_j=0$ otherwise. Then

$$P_a(p)=\sum_{j=0}^{N_1} q_j\binom{N_1}{j}P_1^j(1-P_1)^{N_1-j}.$$

The distribution of X_2 does not involve n_1, hence the weights q_j do not depend on n_1. Moreover, since $j\mapsto A_2(j)$ is decreasing, so is the sequence q_j for $A_1<j<R_1$ and therefore the global sequence $q_0,\ldots,q_{N_1}$. So the proposition can be applied to show that $P_a(p)$ is a decreasing function of $P_1=1-(1-p)^{n_1}$, hence of n_1.

A.2. Inequalities used to accelerate the search in double sampling

A.2.1. Bounds for n_1 and n_2

For $T=(N_1,A_1,R_1,N_2,A_2)$ fixed, the search on (n_1,n_2) can use the following bounds.

• **Lower bound for** n_1. The size n_1 must be big enough to insure that the probability of acceptance at the first step is lower than β when $p=p_{nt}$. In particular, we must have $\text{Prob}(X_1=0|p_{nt})\leq\beta$, that is $(1-p_{nt})^{n_1 N}1\leq\beta$ or equivalently

$$n_1 \geq \frac{\ln(\beta)}{N_1 \ln(1-p_{nt})}$$

This gives a lower bound for n_1. If $A_1>0$, a better lower bound can be found by solving numerically the equality $\text{Prob}(X_1\leq A_1|p_{nt})=\beta$ in n_1. This is easily done as the probability of acceptance $\text{Prob}(X_1\leq A_1|p_{nt})$ is a decreasing function of n_1.

• **Upper bound for** n_1. If the proportion p is very small, the lot should be accepted at the first step with a good probability to avoid an expensive second step. This can be formalized by requiring that

$$\text{Prob}(X_1>A_1|p_0)\leq\beta_0 \tag{18}$$

where p_0 is a fixed small proportion, for instance the value $p_0=p_{nt}/10$ introduced in the end of Section 4, and β_0 a predetermined risk to get no acceptance at the first step when the proportion of GMO is only p_0. We could for instance take $\beta_0=10\%$ to accept at first step with probability 90% when the proportion of GMO is only p_0.

Since $\text{Prob}(X_1(A_1|p_0)$ is an increasing function of n_1, the maximal n_1 satisfying Eq. (18) is easily found. An initial upper bound for n_1 can be deduced from the inequality

$$\text{Prob}(X_1 = N_1|p_0)\leq\beta_0 \tag{19}$$

which is clearly implied by the preceding one since $N_1>A_1$. The equality $X_1=N_1$ means that all N_1 groups assayed at step 1 contain GMO grains. Since the probability under p_0 for a group to include at least one GMO grain is $1-(1-p_0)^{n_1}$, the probability that there is at least one GMO grain in each of the n_1 groups is

$$(1-(1-p_0)^{n_1})^{N_1}.$$

Hence

$$\begin{aligned}\text{Prob}(X_1 = N_1|p_0)\leq\beta_0 &\Leftrightarrow (1-(1-p_0)^{n_1})^{N_1}\leq\beta_0\\ &\Leftrightarrow N_1\ln(1-(1-p_0)^{n_1})\leq\ln(\beta_0).\end{aligned}$$

It is easy to check that $\ln(1-\delta)>-2\delta$ if $\delta\leq 0.79$. Hence, if $(1-p_0)^{n_1}\leq 0.79$, that is $n_1\geq\ln(0.79)/\ln(1-p_0)$, then $N_1\ln(1-(1-p_0)^{n_1})>-2N_1(1-p_0)^{n_1}$ and this is greater $\ln(\beta_0)$ iff $n_1\geq\ln(-\ln(\beta_0)/2N_1)/\ln(1-p_0)$. Therefore, if

$$n_1\geq\max\left(\frac{\ln 0.79}{\ln(1-p_0)}, \frac{\ln(-\ln(\beta_0)/2N_1)}{\ln(1-p_0)}\right) \tag{20}$$

then $\text{Prob}(X_1=N_1|p_0)>\beta_0$. Consequently, Eq. (19), hence Eq. (18), can be satisfied only if n_1 is smaller than the bound on the right of Eq. (20).

• **Lower bound for** n_2 **for a fixed** n_1. For each fixed value of n_1, one has to find the minimum value of n_2 satisfying Eq. (15) (i.e., $n_2=B(n_1)$). This inequality implies

$$\text{Prob}(X_2=0|p_{nt})=(1-p_{nt})^{n_2 N_2}\leq\beta_2=\beta-\text{Prob}(X_1\leq A_1|p_{nt})$$

hence

$$n_2\geq\frac{\ln(\beta_2)}{N_2\ln(1-p_{nt}).}$$

The term on the right provides a lower bound for n_2.

A.2.2. Restrictions in the choice of A_1, R_1, A_2

As previously mentioned, since the group assays are very expensive, only a few values of N_1 and N_2 have to be used. We give below the constraints on the other parameters A_1, R_1, A_2.

• We must have $0\leq A_1\leq N_1-1$, then $A_1+2\leq R_1\leq N_1+1$. If $R_1=N_1+1$, this means that there is no possibility for a rejection at the first step. Note that, if R_1 were equal to A_1+1, there would not be any possibility to get a second step, hence the equality $A_1+2\leq R_1$.

• The acceptance function A_2 varies among the decreasing functions from the set $\{A_1+1,\ldots,R_1-1\}$ into the set $\{0, 1, \ldots, N_2-1\}$.

A.3. Some useful derivatives

The optimization algorithm uses the derivatives of the B and S function dB/dn_1, dS/dn_1, then the derivative of the cost in function of n_1 : $d\mathcal{C}/dn_1$. We show below how to compute exactly these derivatives.

We consider fixed values of N_1, A_1, R_1, N_2 and of the acceptance function A_2. With each couple (n_1,n_2) is then associated the probability $P_a(p)$ of acceptance whenever p is the GMO rate in the lot. We denote it by $F(p, n_1, n_2)$ to stress its dependence on the couple (n_1,n_2). With this notation, the equations defining the B and S curves, derived from the inequalities (15) and (16), become

$$F(p_{nt}, n_1, n_2) = \beta \text{ for } B,$$

$$1 - F(p_t, n_1, n_2) = \alpha \text{ for } S.$$

They are both of the form $F(p, n_1, n_2)=\text{cte}$, which gives by differentiation

$$\frac{\partial F}{\partial n_1}dn_1 + \frac{\partial F}{\partial n_2}dn_2 = 0$$

hence

$$\frac{dn_2}{dn_1} = -\frac{\partial F/\partial n_1}{\partial F/\partial n_2}. \quad (21)$$

Let

$$\phi_{1j}(P_1) = \binom{N_1}{j} P_1^j (1-P_1)^{N_1-j},$$

$$\phi_{2k}(P_2) = \binom{N_2}{k} P_2^k (1-P_2)^{N_2-k}$$

where P_1 and P_2 are defined as in Eq. (11). With these notations, Eq. (14) becomes

$$F(p,n_1,n_2) = \sum_{j=0}^{A_1} \phi_{1j}(P_1) + \sum_{j=A_1+1}^{R_1-1} \phi_{1j}(P_1) \sum_{k=0}^{A_2(j)} \phi_{2k}(P_2).$$

F depends on n_1 through P_1 and on n_2 through P_2. Hence

$$\frac{\partial F}{\partial n_i} = \frac{\partial F}{\partial P_i}\frac{dP_i}{dn_i}.$$

From Eq. (11), we have

$$\frac{dP_i}{dn_i} = -(1-p)^{n_i}\ln(1-p).$$

Consequently

$$\frac{\partial F}{\partial n_1} = -(1-p)^{n_1}\ln(1-p)\left[\sum_{j=0}^{A_1} \phi_{1j}{}'(P_1) + \sum_{j=A_1+1}^{R_1-1} \phi_{1j}'(P_1) \sum_{k=0}^{A_2(j)} \phi_{2k}(P_2)\right]$$

$$\frac{\partial F}{\partial n_2} = -(1-p)^{n_2}\ln(1-p)\left[\sum_{j=A_1+1}^{R_1-1} \phi_{1j}(P_1) \sum_{k=0}^{A_2(j)} \phi_{2k}'(P_2)\right]$$

and Eq. (21) takes the form

$$\frac{dn_2}{dn_1} = -(1-p)^{n_1-n_2}\frac{\sum_{j=0}^{A_1}\phi_{1j}{}'(P_1) + \sum_{j=A_1+1}^{R_1-1}\phi_{1j}{}'(P_1)\sum_{k=0}^{A_2(j)}\phi_{2k}(P_2)}{\sum_{j=A_1+1}^{R_1-1}\phi_{1j}(P_1)\sum_{k=0}^{A_2(j)}\phi_{2k}{}'(P_2)} \quad (22)$$

Lemma A.1 gives

$$\phi_{1j}{}'(P_1) = \binom{N_1}{j} N_1 P_1^{j-1}(1-P_1)^{N_1-j-1}\left(\frac{j}{N_1} - P_1\right) \quad (23)$$

$$\phi_{2k}{}'(P_2) = \binom{N_2}{k} N_2 P_2^{k-1}(1-P_2)^{N_2-k-1}\left(\frac{k}{N_2} - P_2\right). \quad (24)$$

The equality (22) can then be used with $p=p_{nt}$ and $n_2=B(n_1)$ to get the derivative dB/dn_1 at any point n_1, and similarly with $p=p_t$ and $n_2=S(n_1)$ to get dS/dn_1.

We go on assuming N_1, A_1, R_1, N_2, A_2 fixed. For each n_1, the less expansive choice of n_2 is the smaller compatible with the constraints, that is the corresponding point $n_2=B(n_1)$ on the consumer's curve. Recall that the optimal n_1 is found by studying the variation of the cost $\mathcal{C}$ given by Eq. (6) when n_1 varies and n_2 varies accordingly on the B curve.

The derivative of this cost with respect to n_1 is

$$\frac{d}{dn_1}\mathcal{C} = C_g N_1 + \frac{d\ \mathrm{Prob}(\text{step } 2|p)}{dn_1}\left(N_2 + C_g N_2 n_2\right) + \mathrm{Prob}(\text{step } 2|p) C_g N_2 \frac{dn_2}{dn_1}.$$

If the terms depending on C_g are put together, it takes the form

$$\frac{d}{dn_1}\mathcal{C} = C_g\left(N_1 + N_2 n_2 \frac{dP(\text{step } 2|p)}{dn_1} + \mathrm{Prob}(\text{step } 2|p) N_2 \frac{dn_2}{dn_1}\right) + N_2 \frac{d\ \mathrm{Prob}(\text{step } 2|p)}{dn_1}$$

The probability Prob(step 2|p) to go to step 2 when the GMO rate is p is given by Eq. (17). It follows from Lemma A.1 that its derivative with respect to n_1 is

$$\frac{d\ \mathrm{Prob}(\text{step } 2|p)}{dn_1} = \frac{dP}{dn_1}\sum_{j=A_1+1}^{R_1-1}\binom{N_1}{j} N_1 P^{j-1}(1-P)^{N_1-j-1} \times \left(\frac{j}{N_1} - P\right)$$

where

$$\frac{dP}{dn_1} = -\ln(1-p)(1-p)^{n_1} > 0.$$

Using these formula and $n_2=B(n_1)$, it is easy to compute the derivative $d\mathcal{C}/dn_1$ for each n_1.

References

[1] A. Hald, Statistical Theory of Sampling Inspection by Attributes, Academic Press, New York, 1981.

[2] E.G. Schilling, Acceptance sampling in quality control, Statistics: Textbooks and Monographs, vol. 42, Dekker, 1982.

[3] J.J. Daudin, C.S. Tapiero, Les outils et le contrôle de la qualité, Economica (1996).

[4] Y. Maury, C. Duby, J.M. Bossennec, G. Boudazin, Group analysis using ELISA: determination of the level of transmission of Soybean Mosaic Virus in soybean seed, Agronomie 5 (1985) 405–415.

[5] Y. Maury, C. Duby, R.K. Khetarpal, Seed certification for viruses, in: A. Hadidi, R.K. Khetarpal, H. Koganezawa (Eds.), Plant Virus, Disease Control, APS Press, 1998, pp. 237–248 (Chap. 18).

[6] K. Remund, D. Dixon, D. Wright, L. Holden, Statistical considerations in seed purity testing for transgenic traits, Seed Science Research 11 (2001) 101–119.

[7] GIPSA. Grain Inspection Handbook, Book 1, Grain Sampling. http://www.usda.gov/gipsa/pubs/pubs.htm.
[8] GIPSA. Sampling for the Detection of Biotech Grains. http://www.usda.gov/gipsa/biotech/sample2.htm.
[9] NF ISO 3534-2 (1993). Statistics—vocabulary and symbols: Part 2. Statistical quality control.
[10] GIPSA. Sampling and Testing Recommendations for the Detection of Cry9C Protein in Hybrid Seed Corn. http://www.usda.gov/gipsa/biotech/starlink/cry9cdetection.htm.
[11] W.C. Guenther, Use of the binomial, hypergeometric and Poisson tables to obtain sampling plans, Journal of Quality Technology 1 (2) (1969) 105–109.
[12] W.C. Guenther, A procedure for finding double sampling plans for attributes, Journal of Quality Technology 2 (4) (1970) 219–225.
[13] Qalstat, "Logiciel pour PC, 1992", ADEPRINA, 16 rue Claude Bernard 75231, Paris Cedex 05.
[14] H.C. Hamaker, R. van Strik, The efficiency of double sampling by attributes, Journal of the American Statistical Association 50 (1955) 830–849.
[15] E.G. Schilling, A lot sensitive sampling plan for compliance testing and acceptance inspection, Journal of Quality Technology 10 (2) (1978) 47–51.
[16] J.J. Daudin, C. Duby, P. Trécourt, Plans de controle double optimaux, Revue de Statistique Appliquée xxxviii (4) (1990) 45–59.