SCIENTIA AGRICOLA

Bayesian sequential estimation of the proportion of damage in maize seeds

Isabela da Silva Lima[1]* , Carla Regina Guimarães Brighenti[2] , Raquel Maria de Oliveira Pires[3]

[1]Universidade Federal de Lavras – Depto. de Estatística, C.P. 3037 – 37200-000 – Lavras, MG – Brasil.

[2]Universidade Federal de São João del-Rei – Depto. de Zootecnia, Av. Visconde do Rio Preto, s/n – 36301-360 – São João del-Rei, MG – Brasil.

[3]Universidade Federal de Lavras – Depto. de Agricultura, C.P. 3037 – 37200-000 – Lavras, MG – Brasil.

*Corresponding author <isabela_lima30@hotmail.com>

ABSTRACT: Sampling is an essential step in estimating a parameter: thus, cost and time associated to this step should be minimized. Sequential sampling is characterized by using samples of variable sizes given as a function of observations, and sequential sampling provides a smaller sample size than a fixed-size sample in most cases. In addition, the Bayesian decision theory can be incorporated into sequential sampling to perform parameter estimation because it allows the inclusion of *a priori* information about the parameter of interest, which optimizes the procedure. However, the great challenge to performing the Bayesian sequential estimation in establishing the stopping criteria. Most studies in this area investigate binomial distributions, while few analyze multinomial distributions. This study aimed to define the stopping criteria for the Bayesian sequential estimation of the parameters of multinomial distributions with conjugate Dirichlet priors. The proposed methodology was applied to a set of X-ray test data for quality control of maize seed lots. This test uses conventional sampling techniques in which a sample has a fixed size with 200 seeds. The influence of two priors on the stopping criteria was evaluated, one uniform and one conjugate, with hyperparameters based on reference information from the literature. The results indicated a reduction in the sample size in most lots evaluated.

Keywords: Dirichlet distribution, X-ray testing, multinomial distribution, stopping criteria

Introduction

Reduction of costs in the sampling process means the replacement of samples with fixed size of elements by a process that allows the use of samples with variable size as a function of the observations made. This process is known as sequential sampling, and, in most cases, it reduces the sample size (Schnuerch and Erdfelder, 2020).

The Bayesian decision theory is a solid complement to sequential sampling in parameter estimation because it allows incorporating *a priori* information about the parameter of interest. Thus, it is possible to include information relevant to the sampling plan using priors, which helps decision-making (Berger, 1985).

However, a significant problem with the use of the Bayesian sequential estimation is the difficulty to establish the stopping criteria, as complex mathematics is required due to the dynamic nature of the process and the recursion in the calculations (Fenoy, 2017).

Most studies in this area investigate variables that follow discrete probability distributions in the univariate context for the binomial distribution, such as Plant and Wilson (1985), Karunamuni and Prasad (2003), Brighenti et al. (2019), among others. Few studies have been conducted on multivariate distributions. One example is Jones (1976), which established the stopping criteria for a multinomial distribution; nevertheless, it considered only uniform priors.

Maize is one of Brazil's most economically important crops, and productivity increases in this sector are related to the selection of good quality seeds. X-ray testing evaluates seed quality, deaferenting seeds with good formation from those with some damage. This test uses conventional sampling techniques in which a sample has a fixed size of 200 seeds, according to the Rules for Seed Analysis (RAS), which regulates such procedures. The process of verifying these seeds is performed visually by an analyst, one at a time, which is labor intensive. Sequential sampling may improve these tests, as there is no need to pre-establish the number of seeds to assess and can optimize the procedure, decreasing the required sample size.

In this study, we aimed to develop the stopping criteria for the process of Bayesian sequential estimation of the parameters of a multinomial distribution for conjugate Dirichlet priors because the literature does not report on this development and multinomial distributions have been little explored, despite their great applicability to estimate proportions in problems with more than two response categories, such as estimation of the proportion of seeds with different types of damage in X-ray test. Thus, the developed methodology was applied, and the results were compared with those of the conventional approach to X-ray tests.

Materials and Methods

Multinomial distribution

In complex situations, classification of sample elements in more than two categories can be performed. In seed quality control, there is interest in estimating the proportion of seeds with different types of damage, such as insect damage, density variation, and physical damage, among others (MAPA, 2009).

Thus, the variable of interest is polytomous, and the multinomial distribution is used to estimate the probability of an element belonging to more than two categories, which is a discrete probability distribution and a generalization of the binomial distribution (Najar and Bouguila, 2022).

The multinomial distribution is defined assuming an experiment whose result is one of the events E_1, E_2, ..., E_k with probability $P[E_i] = p_i$, where k is the number of classes of the multinomial distribution. For $i = 1, 2$, ..., k, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$, and let X_i be a random variable that counts the number of occurrences of E_i in m independent replicates of this experiment. Then, the random vector $(X_1, X_2, ..., X_k)$ has a multinomial distribution, with parameters p_1, p_2, ..., p_{k-1}, given by (Najar and Bouguila, 2022):

$$P[X_1 = x_1, X_2 = x_2, ..., X_k = x_k] = \frac{m!}{x_1! x_2! \cdot ... \cdot x_k!} p_1^{x_1} p_2^{x_2} \cdot ... \cdot p_k^{x_k} = m! \prod_{i=1}^{k} \frac{p_i^{x_i}}{x_i!} \quad (1)$$

where: each X_i is a positive integer, p_1, p_2, ..., p_k are population proportions, and $\sum_{i=1}^{k} x_i = m$. There are p_1, p_2, ..., p_{k-1} parameters, and because we have $\sum_{i=1}^{k} p_i = 1$; therefore, $p_k = 1 - \sum_{i=1}^{k-1} p_i$. Thus, $(X \mid m, p) \sim Multinomial$ (m, p).

The expectation, variance, and covariance of the multinomial distribution are given, respectively, by:

$$E(X_i) = mp_i, \quad (2)$$

$$Var(X_i) = mp_i(1 - p_i), \quad (3)$$

$$Cov(X_i, X_j) = -mp_i p_j \quad (4)$$

Bayesian estimation of the parameters of the multinomial distribution

The Dirichlet distribution is a discrete multivariate distribution widely used in the Bayesian context as *a priori* conjugate distribution of the multinomial distribution, a generalization of the beta distribution (Paulino et al., 2018).

Conjugate *a priori* distribution was used because there are closed expressions for *a posteriori* distribution, which facilitates the calculation because it does not require computational effort to implement computational algorithms. Thus, for less experienced researchers, expressions closed for *a posteriori* distribution provide a character to understand that using Bayesian inference implies an update rule with the possibility of generating a historical database.

If $X = (X_1, ..., X_k)^T$ is a vector with k components, then it follows a Dirichlet distribution of order $k \geq 2$ with a vector of parameters $a = (a_1, ..., a_k)^T$, that is, (Paulino et al., 2018): $(X \mid a) \sim Dirichlet(a)$. Its probability density function is given by:

$$f_x(x_1, ..., x_k) = \frac{\Gamma(a_0)}{\prod_{i=1}^{k} \Gamma(a_i)} \prod_{i=1}^{k} p_i^{a_i - 1}, 0 < p_i < 1 \quad (5)$$

where: $a_0 = \sum_{i=1}^{k} a_i$, $\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$ is the gamma function and $\sum_{i=1}^{k} p_i = 1$. The marginal distribution is a beta with parameters a_i and $(a_0 - a_i)$ for each i, from which we have:

$$E(X_i) = \frac{a_i}{a_0}, Var(X_i) = \frac{a_i(a_0 - a_i)}{a_0^2(a_0 + 1)}, Cov(X_i, X_j) = \frac{-a_i a_j}{a_0^2(a_0 + 1)} \quad (6)$$

The probability vectors p, parameters of the multinomial distribution, follow a Dirichlet distribution with parameters a. Thus, if *a priori* distribution is a Dirichlet and the observed variable follows a multinomial distribution, then *a posteriori* distribution is a Dirichlet distribution, with another parameter:

$$p \mid X \sim Dirichlet(a_1^* = x_1 + a_1, ..., a_k^* = x_k + a_k)$$

where: $X = (x_1, ..., x_k, a_1, ..., a_k)^T$.

Thus, the mean, variance, and covariance of *a posteriori* distribution of Dirichlet are given respectively by (Avetisyan and Fox, 2012):

$$E(p \mid X_i) = \frac{x_i + a_i}{\sum_{i=1}^{k}(x_i + a_i)}, \quad (7)$$

$$Var(p \mid X_i) = \frac{(x_i + a_i)\left\{\left[\sum_{i=1}^{k}(x_i + a_i)\right] - (x_i + a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i + a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i + a_i)\right] + 1\right\}}, \quad (8)$$

$$Cov(p \mid X_i, X_j) = \frac{-(x_i + a_i)(x_j + a_j)}{\left[\sum_{i=1}^{k}(x_i + a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i + a_i)\right] + 1\right\}}. \quad (9)$$

Stopping criteria for the Bayesian sequential estimation of the parameters of multinomial distributions

The major challenge of Bayesian sequential estimation is to establish the stopping criteria. This process was initially based on the article by Jones (1976), where the expressions involved in the procedure are presented, such as those for the immediate and expected risks, based on dynamic programming. However, these expressions are developed considering only uniform priors.

Thus, the first step was to understand how Jones (1976) derived the expressions for the risks using a uniform prior, and for this purpose, demonstrations were carried out to enhance the understanding of the process. Subsequently, these risk expressions were generalized for any conjugate Dirichlet prior, following the paths provided by the author, and corresponding demonstrations were also performed.

Some concepts are involved in the Bayesian sequential estimation procedure, such as the loss function, Bayes risk, immediate risk, expected risk, and cost function. It follows that a sequential procedure d is

a decision rule. This is a function defined in the space of the possible results of an experiment that assumes values in the space of possible actions.

Each decision d and each possible value of the parameter p can be associated with a loss, which assumes positive values, in addition to the cost function $C(n)$, indicating the cost of taking n observations.

The loss function is defined as $L(p, d)$. According to Ali (2015), the loss function most used in estimation problems is the quadratic loss function, defined as:

$$L(\hat{p}, p) = (\hat{p} - p)^2. \qquad (10)$$

The risk of a decision rule, denoted by $R(p, d)$, is the expected loss of *a posteriori*, that is (Berger, 1985):

$$R(p, d) = E_{posteiori}[L(p, d)]. \qquad (11)$$

The Bayes risk of a sequential procedure d is defined by:

$$r(\pi, d) = E^\pi [R(p, d)], \qquad (12)$$

that is, the expected risk associated with the procedure to estimate parameter p given a prior π, after n observations.

Therefore, the Bayes estimator of p concerning the loss function has the lowest Bayes risk. In the case of using a quadratic loss function, the Bayes estimator for the parameter p is the mean of its *posterior* distribution (Berger, 1985; Ali, 2015).

The "one-step look ahead" method can be considered one of the most valuable methods to develop the stopping criteria for the Bayesian sequential estimation procedure proposed in the literature. Based on this method, the Bayes risk of making an immediate decision is $r_0(\pi^n, n) = \inf_{a \in A} (\pi^n, a, n)$, where A is the set of available actions and $r_0(\pi^n, a, n) = E^{\pi^n}[L(\theta, a, n)]$ is the expected *a posteriori* loss of action a in n (Berger, 1985).

The lowest *a posteriori* Bayes risk was demonstrated is the variance of the posterior distribution, denoted by $var_{post}(n)$ (Pratt et al., 1964). The *posteriori* risk under quadratic loss function is simply the variance, that is (Ali, 2015, 2013):

$$r_0(\pi^n, n) = var_{post}(n). \qquad (13)$$

Thus, the expected *a posteriori* Bayes risk when another observation is made is the expectation of this variance, that is (Pham-Gia, 1998):

$$r^1(\pi^n, n) = E[var_{post}(n)]. \qquad (14)$$

In this sense, to determine the stopping criteria, it is necessary to calculate the immediate risk, $r_0(\pi^n, n)$, plus the cost of n observations, and the expected risk, $r^1(\pi^n, n)$ with the increase in the cost with one more observation (Berger, 1985).

However, the risks are given by a recurrence relationship, and a closed formula must be found to solve a recurrence, but the function $E[var_{post}(n)]$ is generally not available in closed form, which makes the whole calculation highly complex. Thus, to calculate the risks, an adaptation to the "one-step look ahead" method was performed using dynamic programming equations instead of calculating $E[var_{post}(n)]$.

The After evaluating the nth observation, the procedure consists of comparing $r_0(\pi^n, n)$ with $r^1(\pi^n, n)$. If $r_0(\pi^n, n) > r^1(\pi^n, n)$, sampling continues; if $r_0(\pi^n, n) \leq r^1(\pi^n, n)$, the sampling stops. The Bayes sequential rule is also known as Bayesian learning because *a posteriori* distribution calculated for the current n is used to update *a priori* distribution still to be used in the $(n + 1)$-th inspection (Berger, 1985).

Therefore, we considered the multinomial distribution with $(k+1)$ classes to determine the stopping criteria. The probability of an observation in the i-th class is p_i, with $i = 1, 2, ..., k$, and that in the $(k+1)$-th class is $\left(1 - \sum_{i=1}^{k} p_i\right)$ because $p_{(k+1)} = \left(1 - \sum_{i=1}^{k} p_i\right)$ since $\sum_{i=1}^{k+1} p_i = 1$.

The *a priori* information about the parameter p, given by $p = (p_1, p_2, ..., p_k)^T$, can be adequately represented by a member of the natural conjugate Dirichlet family of distributions with integer parameters $a_0, a_i, i = 1, 2, ..., k$, with density proportional to:

$$\prod_{i=1}^{k} p_i^{a_i - 1} \left(1 - \sum_{i=1}^{k} p_i\right)^{a_0 - \Sigma a_i - 1}.$$

With $p_i \geq 0$ and $\sum_{i=1}^{k} p_i \leq 1$.

The a_i parameter of the Dirichlet distribution is a vector parameter, and the other parameter of this distribution can be written as $a_0 = \sum_{i=1}^{k+1} a_i$. Therefore, $a_{(k+1)} = a_0 - \sum_{i=1}^{k} a_i$, according to the previous definition for $k + 1$ classes.

According to the Bayes theorem, after m observations resulting in x_i in the i-th classes, the posterior density of p is given by a Dirichlet with parameters $a_0 + m$, $a_i + x_i$, where m is the total number of observations or the sample size, and x_i is the number of observations in each of the i-th classes (Jones and Madhi, 1988).

The result of the sampling can be represented as a sample path that begins at point (a, a_0), where $a = (a_1, a_2, ..., a_k)$, in the whole dimensional space $(k + 1)$ and is interrupted when the stopping limit, which must be determined, is reached.

The uniform distribution is a particular case of the Dirichlet distribution, corresponding to the case in which $a_1 = a_2 = ... = a_k = 1$. A uniform prior is noninformative because all possible values of the parameter of interest are equally probable. The fact that the Dirichlet class includes these "noninformative" natural antecedents is a reason to use it as *a priori* distribution of the multinomial distribution (Paulino et al., 2018).

To obtain the stopping limits, the quadratic loss is considered in the estimate of p by $d = (d_1, d_2, ..., d_k)^T$,

and this has the general quadratic form $(p − d)^T K(p - d)$, where K is a positive symmetric $I \times I$ matrix of constant loss (Chen, 1988; Jones, 1976). Then, according to Ali (2015) using a quadratic loss, the Bayes estimator d^* is the mean of *a posteriori* distribution of (x, m), that is, is the mean of *posteriori* Dirichlet distribution, given by:

$$d_i^* = E(p \mid X_i) = \frac{a_i + x_i}{\Sigma (a_i + x_i)} = \frac{a_i + x_i}{a_0 + \Sigma x_i} = \frac{a_i + x_i}{a_0 + m}. \quad (15)$$

The above equalities come from the fact that $\sum_{i=1}^{k+1} a_i = a_0$ from the Dirichlet distribution and x_i comes from the likelihood, in this case, multinomial, which represents the number of observations in each class, and therefore, it is known that $\sum_{i=1}^{k+1} x_i = m$.

In the case of the use of a uniform prior, then the parameters are $a_i = 1$ and $a_0 = k + 1$. Replacing them in expression Eq. (15), the Bayes estimator using a uniform prior is given by (Jones, 1976):

$$d_i^* = \frac{(x_i + 1)}{(m + k + 1)}. \quad (16)$$

The expression given by d_i^* is also *a posteriori* marginal probability of the following observation within the *i*-th class.

The Bayes risk or immediate risk, or even the risk of the stopping decision-making, is given by (Jones, 1976):

$$S((x_1, ..., x_k), m) = S(x, m) = trace\ (K\Sigma), \quad (17)$$

where: K is a positive symmetric square matrix of dimension $I \times I$ of constant loss and Σ is the dispersion matrix of *a posteriori* Dirichlet distribution of dimension $I \times I$.

Therefore, considering a uniform prior, the immediate risk is given by:

$$S(x, m) = \frac{\left(\sum_{i=1}^{k} K_{ii} d_i^* - \sum_{i,j=1}^{k} K_{ij} d_i^* d_j^* \right)}{m + k + 2}. \quad (18)$$

The dynamic programming equations providing the partition for the stop and continuation points are given by (Jones, 1976; Jones and Madhi, 1988):

$$D(x, m) = min[S(x, m), B(x, m)], \quad (19)$$

$$B(x,m) = c + \sum_{i=1}^{k} \left[D\left(x + e_i, m+1\right) d_i^* \right] + D(x, m+1)\left(1 - \sum_{i=1}^{k} d_i^*\right), \quad (20)$$

where: $B(x, m)$ is the risk of making an additional observation at a cost c (expected risk). $D(x, m)$ is the minimum risk or also known as ideal risk.

Therefore, $S(x, m)$ is the immediate risk, $B(x, m)$ is the expected risk, and $D(x, m)$ is the minimum between the immediate and the expected risks. The ideal risk $D(x, m)$ is equal to that expected when the decision is to continue sampling; otherwise, when the decision is to stop sampling, it is equal to the immediate risk.

Therefore, the dynamic programming Eq. (19) and

(20) are used successively for $m \leq N^*$ to find the smallest integer m that satisfies (Jones, 1976):

$$B(x, m) > S(x, m), D(x, m + 1) = S(x, m + 1) \text{ for all } x \quad (21)$$

and this m provides the maximum sample size. Thereby, the sequential inspection procedure known as "one-step look ahead", in which the inspection ends in the smallest integer m that satisfies the two conditions in Eq. (21), is characterized.

Therefore, the expression of the expected risk $B(x, m)$ that establishes the stopping criterion, for uniform priors, is given by:

$$B(x,m) = c + S(x,m)\left(\frac{m + k + 1}{m + k + 2} \right). \quad (22)$$

The following flowchart (Figure 1) summarizes the steps of the Bayesian sequential estimation procedure.

Example and application to maize seed damage submitted to the X-ray test

Finally, a hypothetical example is presented to explain the procedure and the recursion involved thoroughly. In addition, after the theory for the consolidated Bayesian sequential estimation is presented, it is applied to the data of maize seeds subjected to X-ray testing for quality control to verify and discuss the results.

The X-ray testing for the maize seed analysis was conducted in the city of Lavras, Minas Gerais State, Brazil. The radiographic images were generated by a Faxitron MX-20 device (Faxitron X-ray Corp) connected to a computer and monitor. It was configured at 26 kV, and the seeds were exposed to radiation for 20 s.

A total of 100 lots of maize seeds were analyzed, with four replicates of 50 seeds for each lot, totaling 200 seeds per lot, fixed in an orderly manner on acrylic plates (21 × 15 cm) with double-sided transparent tape adequately labeled with the lot number, the replicate, and the position of each seed to allow individual identification in subsequent analyses. Each seed was analyzed individually.

In the X-ray test, digital radiographs were generated that were visually analyzed to classify the presence or absence of damage. Intact seeds were defined as those
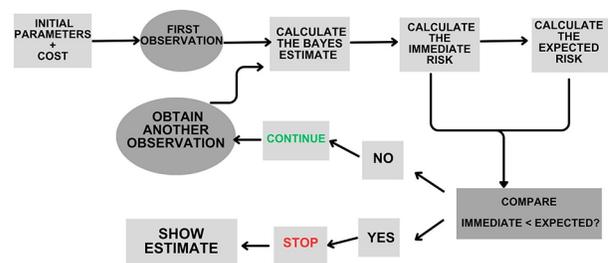
Figure 1 – Flowchart of the Bayesian sequential estimation procedure.

without any type of damage, such as insect damage, physical damage, or damage due to density variations.

Therefore, the proportion of seeds was estimated for the three classes: seeds without damage, seeds with variations in density, and seeds with other types of damage (insect damage + physical damage), considering two priors, a uniform prior, where the values of the hyperparameters were equally probable (all equal to one). Another prior derived from the literature, constructed through elicitation based on the results of Javorski and Cicero (2017), who evaluated the damage to sorghum seeds and grasses, such as maize, using X-ray tests.

Results

Development of the stopping criterion

Expressions involved in the Bayesian sequential estimation procedure for the parameters of the multinomial distribution are presented in Jones (1976); however, these are developed considering only uniform priors.

Jones (1976) used a uniform prior to obtain the stop limits to facilitate the calculations and find the immediate and expected risks. This is a particular case of the Dirichlet distribution, where the parameters all assume value one.

The expressions contained in Jones (1976) for uniform priors are demonstrated as follows for a better understanding and generalization to the use of any other conjugate Dirichlet priors:

Considering a uniform prior, the dispersion matrix of the posterior distribution has the elements:

$$var(p_i) = \frac{d_i^*(1 - d_i^*)}{m + k + 2} \quad \text{and} \quad cov(p_i, p_j) = -\frac{d_i^* d_j^*}{m + k + 2}. \qquad (23)$$

Given that the matrix $\boldsymbol{K}$ is a definite positive symmetric matrix of constant loss, then it is known that $\boldsymbol{K}$ is a square matrix and $K_{ij} = K_{ji}$.

Let the dispersion matrix also square of dimension $I \times I$:

$$\Sigma = \begin{bmatrix} \frac{d_i^*(1 - d_i^*)}{m+k+2} & \frac{-d_i^* d_j^*}{m+k+2} & \cdots & \frac{-d_i^* d_j^*}{m+k+2} \\ \frac{-d_i^* d_j^*}{m+k+2} & \frac{d_i^*(1 - d_i^*)}{m+k+2} & \cdots & \frac{-d_i^* d_j^*}{m+k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-d_i^* d_j^*}{m+k+2} & \frac{-d_i^* d_j^*}{m+k+2} & \cdots & \frac{d_i^*(1 - d_i^*)}{m+k+2} \end{bmatrix}$$

Therefore,

$$S(\boldsymbol{x}, m) = trace(\boldsymbol{K}\Sigma) = trace \begin{bmatrix} K_{ii} & K_{ij} & \cdots & K_{ij} \\ K_{ij} & K_{ii} & \cdots & K_{ij} \\ \vdots & \vdots & \ddots & K_{ij} \\ K_{ij} & K_{ij} & \cdots & K_{ii} \end{bmatrix}$$

$$\begin{bmatrix} \frac{d_i^*(1 - d_i^*)}{m+k+2} & \frac{-d_i^* d_j^*}{m+k+2} & \cdots & \frac{-d_i^* d_j^*}{m+k+2} \\ \frac{-d_i^* d_j^*}{m+k+2} & \frac{d_i^*(1 - d_i^*)}{m+k+2} & \cdots & \frac{-d_i^* d_j^*}{m+k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-d_i^* d_j^*}{m+k+2} & \frac{-d_i^* d_j^*}{m+k+2} & \cdots & \frac{d_i^*(1 - d_i^*)}{m+k+2} \end{bmatrix} =$$

$$= \frac{K_{ii}d_i^*(1 - d_i^*)}{m+k+2} - \frac{K_{ij}d_i^* d_j^*}{m+k+2} + \cdots - \frac{K_{ij}d_i^* d_j^*}{m+k+2} - \frac{K_{ij}d_i^* d_j^*}{m+k+2} + \frac{K_{ii}d_i^*(1 - d_i^*)}{m+k+2} + \cdots$$
$$- \frac{K_{ij}d_i^* d_j^*}{m+k+2} - \frac{K_{ij}d_i^* d_j^*}{m+k+2} - \frac{K_{ij}d_i^* d_j^*}{m+k+2} + \cdots + \frac{K_{ii}d_i^*(1 - d_i^*)}{m+k+2} =$$

$$= \frac{K_{ii}d_i^* - K_{ii}d_i^* d_i^* - K_{ij}d_i^* d_j^* + \cdots - K_{ij}d_i^* d_j^* - K_{ij}d_i^* d_j^* + K_{ii}d_i^* - K_{ii}d_i^* d_i^* + \cdots - K_{ij}d_i^* d_j^* -}{m+k+2}$$

$$\frac{-K_{ij}d_i^* d_j^* - K_{ij}d_i^* d_j^* + \cdots + K_{ii}d_i^* - K_{ii}d_i^* d_i^*}{m+k+2} =$$

$$= \frac{1}{m+k+2}\left(\sum_{i=1}^{k} K_{ii}d_i^* - \sum_{i=1}^{k} K_{ii}d_i^* d_i^* - \sum_{i,j=1}^{k} K_{ij}d_i^* d_j^*\right). \qquad (24)$$

For the case where $i = j$, $\sum_{i=1}^{k} K_{ij}d_i^* d_j^* = \sum_{i=1}^{k} K_{ii}d_i^* d_i^*$. Equation (24) is reduced to:

$$S(\boldsymbol{x}, m) = \frac{\left(\sum_{i=1}^{k} K_{ii}d_i^* - \sum_{i,j=1}^{k} K_{ij}d_i^* d_j^*\right)}{m + k + 2}. \qquad (25)$$

The dynamic programming equations are used to estimate the unknown probabilities p_i, in which $i = 1, 2, …, k$, from sequential sampling because they provide an optimal decision at each point. The stopping criterion for the multinomial distribution is obtained from these equations, as shown below.

For this purpose, a point $(\boldsymbol{x}, m) = (x_1, x_2, …, x_k, m)$ is considered. If c is the sampling cost of an observation, $B(\boldsymbol{x}, m)$ the risk of making an additional observation at a cost c (expected risk), and $D(\boldsymbol{x}, m)$ is the minimum risk or ideal risk, then the dynamic programming equations providing the partition for the stop and continuation points are (Jones, 1976; Jones and Madhi, 1988):

$$D(\boldsymbol{x}, m) = min[S(\boldsymbol{x}, m), B(\boldsymbol{x}, m)], \qquad (26)$$

$$B(\boldsymbol{x}, m) = c + \sum_{i=1}^{k}\left[D(\boldsymbol{x} + \boldsymbol{e}_i, m+1)d_i^*\right] + D(\boldsymbol{x}, m+1)\left(1 - \sum_{i=1}^{k}d_i^*\right), \qquad (27)$$

for each point in the integer space $(k + 1)$, there are $(k + 1)$ possible transitions, $(\boldsymbol{x} + \boldsymbol{e}_i, m + 1)$, with probability d_i^*, where $\boldsymbol{e}_i$ is the line vector with one in the i-th position and zero in the other positions: $\boldsymbol{e}_i = (0, …, 1, 0, 0)$.

Since $S(\boldsymbol{x}, m) \to 0$ and $B(\boldsymbol{x}, m) \to c$ as $m \to \infty$, a large value of $m = N^*$ is obtained, so all points $(\boldsymbol{x}, N^*)$ are stop points for $\sum_{i=1}^{k} x_i \leq N^*$. The dynamic programming Eq. (26) and (27) can now be used successively for $m \leq N^*$ to find the smallest integer m that satisfies (Jones, 1976):

$$B(\boldsymbol{x}, m) = c + \sum_{i=1}^{k}\left[D(\boldsymbol{x} + \boldsymbol{e}_i, m+1)d_i^*\right] + D(\boldsymbol{x}, m+1)\left(1 - \sum_{i=1}^{k}d_i^*\right), \qquad (28)$$

and this m is the maximum sample size.

Therefore, to find the expression of the expected risk $B(x, m)$ that establishes the stopping criterion for uniform priors, we have:

$$B(x,m) = c + \sum_{i=1}^{k} \left[D(x+e_i, m+1) d_i^* \right] + D(x,m+1)\left(1 - \sum_{i=1}^{k} d_i^*\right)$$

$$B(x,m) = c + \sum_{i=1}^{k} \left[D(x+e_i, m+1)\left(\frac{x_i+1}{m+k+1}\right) \right] + D(x,m+1)\left(1 - \sum_{i=1}^{k} \frac{x_i+1}{m+k+1}\right)$$

$$B(x,m) = c + \sum_{i=1}^{k} \left[D(x+e_i, m+1)\left(\frac{x_i+1}{m+k+1}\right) \right] + D(x,m+1)\left(\frac{m+k+1-\sum_{i=1}^{k}(x_i+1)}{m+k+1}\right)$$

Using the fact that when $S(x, m) < B(x, m)$, we decide to stop the analysis, and thus $D(x, m+1) = S(x, m+1)$ since the goal is to find the $B(x, m)$ for which we have the stop rule. Then, $D(x, m+1)$ can be replaced by $S(x, m+1)$:

$$B(x,m) = c + \sum_{i=1}^{k} \left[S(x+e_i, m+1)\left(\frac{x_i+1}{m+k+1}\right) \right] + S(x,m+1)\left(\frac{m+k+1-\sum_{i=1}^{k}(x_i+1)}{m+k+1}\right)$$

$$B(x,m) = c + \sum_{i=1}^{k} \left[S(x,m)\left(\frac{x_i+1}{(m+1)+k+1}\right) \right] + S(x,m)\left(\frac{m+k+1-\sum_{i=1}^{k}(x_i+1)}{(m+1)+k+1}\right)$$

$$B(x,m) = c + S(x,m)\left\{ \left[\sum_{i=1}^{k}\left(\frac{x_i+1}{(m+1)+k+1}\right) \right] + \left(\frac{m+k+1-\sum_{i=1}^{k}(x_i+1)}{(m+1)+k+1}\right) \right\}$$

$$B(x,m) = c + S(x,m)\left(\frac{\sum_{i=1}^{k}x_i + k + m + k + 1 - \sum_{i=1}^{k}x_i - k}{(m+1)+k+1} \right)$$

$$B(x,m) = c + S(x,m)\left(\frac{m+k+1}{m+k+2} \right). \tag{29}$$

Therefore, Eq. (29) is the expected risk when using uniform priors.

The stopping criterion is summarized by comparing the values of the immediate and expected risks for each observation. When $S(x, m) < B(x, m)$, the sampling stops, and the parameters of interest are estimated. Otherwise, if $S(x, m) > B(x, m)$, the sampling continues, making one more observation before another decision is made.

From the previous demonstrations and the paths given by Jones (1976), it was possible to establish general expressions for any other conjugate Dirichlet priors to be used, the demonstrations of which are presented below.

Using the general expression for any conjugate Dirichlet priors, without restriction to only uniform priors to find the immediate risk, the dispersion matrix of *a posteriori* distribution has the following elements:

$$Var\left(p|X_i\right) = \frac{(x_i+a_i)\left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right] - (x_i+a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}}, \tag{30}$$

$$Cov\left(p|X_i, X_j\right) = \frac{-(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}}. \tag{31}$$

Therefore,

$$\Sigma = \begin{bmatrix} \frac{(x_i+a_i)\left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]-(x_i+a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} & \cdots & \frac{-(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} \\ \vdots & \ddots & \vdots \\ \frac{-(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} & \cdots & \frac{(x_i+a_i)\left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]-(x_i+a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} \end{bmatrix}$$

Let $K\Sigma$ square matrices of dimensions $I \times I$, we have: $S(x, m) = $ trace $(K\Sigma)$

$$= trace \begin{bmatrix} K_{ii} & \cdots & K_{ij} \\ \vdots & \ddots & \vdots \\ K_{ij} & \cdots & K_{ii} \end{bmatrix}$$

$$\begin{bmatrix} \frac{(x_i+a_i)\left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]-(x_i+a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} & \cdots & \frac{-(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} \\ \vdots & \ddots & \vdots \\ \frac{-(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} & \cdots & \frac{(x_i+a_i)\left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]-(x_i+a_i)\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} \end{bmatrix} =$$

$$= \frac{K_{ii}\left\{(x_i+a_i)\left[\sum_{i=1}^{k}(x_i+a_i)-(x_i+a_i)\right]\right\} + \cdots - K_{ij}\left[(x_i+a_i)(x_j+a_j)\right] + \cdots -}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}}$$

$$\frac{-K_{ij}\left[(x_i+a_i)(x_j+a_j)\right] + \ldots + K_{ii}\left\{(x_i+a_i)\left[\sum_{i=1}^{k}(x_i+a_i)-(x_i+a_i)\right]\right\}}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} =$$

$$= \frac{\sum_{i=1}^{k} K_{ii}\left\{(x_i+a_i)\left[\left(\sum_{i=1}^{k}(x_i+a_i)\right)-(x_i+a_i)\right]\right\} - \sum_{i,j=1}^{k} K_{ij}\left[(x_i+a_i)(x_j+a_j)\right]}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}} =$$

$$= \frac{\sum_{i=1}^{k} K_{ii}(x_i+a_i)\left(\sum_{i=1}^{k}(x_i+a_i)\right) - \sum_{i=1}^{k} K_{ii}(x_i+a_i)(x_i+a_i) - \sum_{i,j=1}^{k} K_{ij}(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}}. \tag{32}$$

For the case where $i = j$, we have $\sum_{i,j=1}^{k} K_{ij}(x_i+a_i)(x_j+a_j) = \sum_{i=1}^{k} K_{ii}(x_i+a_i)(x_i+a_i)$.

Therefore, Eq. (32) is reduced to:

$$S(x,m) = \frac{\sum_{i=1}^{k} K_{ii}(x_i+a_i)\left(\sum_{i=1}^{k}(x_i+a_i)\right) - \sum_{i,j=1}^{k} K_{ij}(x_i+a_i)(x_j+a_j)}{\left[\sum_{i=1}^{k}(x_i+a_i)\right]^2 \left\{\left[\sum_{i=1}^{k}(x_i+a_i)\right]+1\right\}}. \tag{33}$$

Therefore, expression Eq. (33) is the general expression for the immediate risk for any conjugate Dirichlet prior, not just uniform priors.

To find the expression of the expected risk $B(x, m)$ that establishes the stopping criterion for any Dirichlet prior:

$$B(x,m) = c + \sum_{i=1}^{k} \left[D(x+e_i, m+1) d_i^* \right] + D(x,m+1)\left(1 - \sum_{i=1}^{k} d_i^*\right)$$

$$B(\boldsymbol{x},m)=c+\sum_{i=1}^{k}\left[D(\boldsymbol{x}+\boldsymbol{e}_i,m+1)\left(\frac{a_i+x_i}{a_0+m}\right)\right]+D(\boldsymbol{x},m+1)\left(1-\sum_{i=1}^{k}\frac{a_i+x_i}{a_0+m}\right)$$

$$B(\boldsymbol{x},m)=c+\sum_{i=1}^{k}\left[D(\boldsymbol{x}+\boldsymbol{e}_i,m+1)\left(\frac{a_i+x_i}{a_0+m}\right)\right]+D(\boldsymbol{x},m+1)\left(\frac{a_0+m-\sum_{i=1}^{k}a_i-\sum_{i=1}^{k}x_i}{a_0+m}\right).$$

Using the fact that when $S(\boldsymbol{x},\ m)\ <\ B(\boldsymbol{x},\ m)$, we decide to stop the analysis and thus $D(\boldsymbol{x},\ m\ +\ 1)\ =\ S(\boldsymbol{x},\ m\ +\ 1)$ since the goal is to find $B(\boldsymbol{x},\ m)$ for which we have the stop rule, $D(\boldsymbol{x},\ m\ +\ 1)$ can be replaced by $S(\boldsymbol{x},\ m\ +\ 1)$:

$$B(\boldsymbol{x},m)=c+\sum_{i=1}^{k}\left[S(\boldsymbol{x}+\boldsymbol{e}_i,m+1)\left(\frac{a_i+x_i}{a_0+m}\right)\right]+S(\boldsymbol{x},m+1)\left(\frac{a_0+m-\sum_{i=1}^{k}a_i-\sum_{i=1}^{k}x_i}{a_0+m}\right)$$

$$B(\boldsymbol{x},m)=c+\sum_{i=1}^{k}\left[S(\boldsymbol{x},m)\left(\frac{a_i+x_i}{a_0+(m+1)}\right)\right]+S(\boldsymbol{x},m)\left(\frac{a_0+m-\sum_{i=1}^{k}a_i-\sum_{i=1}^{k}x_i}{a_0+(m+1)}\right)$$

$$B(\boldsymbol{x},m)=c+\frac{S(\boldsymbol{x},m)}{a_0+(m+1)}\left(\sum_{i=1}^{k}a_i+\sum_{i=1}^{k}x_i+a_0+m-\sum_{i=1}^{k}a_i-\sum_{i=1}^{k}x_i\right)$$

$$B(\boldsymbol{x},m)=c+\frac{S(\boldsymbol{x},m)}{a_0+(m+1)}(a_0+m).$$

Therefore, the general expression of the expected risk for any conjugate Dirichlet priors, not just uniform priors, is given by (34):

$$B(\boldsymbol{x},m)=c+S(\boldsymbol{x},m)\left(\frac{a_0+m}{a_0+m+1}\right). \tag{34}$$

A hypothetical example is presented to better explain the procedure. A partial data set from batch one of seeds, where the first column indicates the seeds, and the other columns are the possible classifications. Where one indicates that the seed has that characteristic and zero otherwise. For example, when analyzed, it was concluded that seed one had no damage; thus, it was filled in by one in the column without damage and this was done successively for the 200 seeds for the 100 lots analyzed (Table 1).

• Step 1: Determine the priors and cost:

Uniform priors: $a_1\ =\ 1$, $a_2\ =\ 1$, and $a_3\ =\ 1$ and Cost: $0.00001\ =\ 10^{-5}$

Table 1 – Data set example.

Seeds	Without damage	Density variations	Other types of damage
1	0	1	0
2	0	1	0
3	1	0	0
4	0	0	1
⋮	⋮	⋮	⋮
199	1	0	0
200	1	0	0

Assuming a multinomial with three classes, in which each observation belongs to one of the classes, we have:

• Step 2: Obtain the first observation.

With the observation in the second class: $x_1\ =\ 0$, $x_2\ =\ 1$, $x_3\ =\ 0$.

• Step 3: Estimate the proportion (mean of the posterior distribution):

$$d_i^*=\frac{(x_i+1)}{(m+k+1)}\Rightarrow d_1^*=d_3^*=\frac{(0+1)}{(1+2+1)}=\frac{1}{4}=0.25. d_2^*=\frac{(1+1)}{(1+2+1)}=\frac{2}{4}=0.50.$$

• Step 4: Calculate the immediate risk:

$$S(\boldsymbol{x},m)=\frac{1}{m+k+2}\left(\sum_{i=1}^{k}K_{ii}d_i^*-\sum_{i=1}^{k}K_{ii}d_i^*d_i^*-\sum_{i,j=1}^{k}K_{ij}d_i^*d_j^*\right)=\frac{10}{80}=0.125.$$

• Step 5: Calculate the expected risk:

$$B(\boldsymbol{x},m)=c+S(\boldsymbol{x},m)\left(\frac{m+k+1}{m+k+2}\right)=0.10001.$$

• The expected risk without using the expression, for a better explanation of the recursive procedure, is as follows:

$$B(\boldsymbol{x},m)=c+\sum_{i=1}^{k}\left[D(\boldsymbol{x}+\boldsymbol{e}_i,m+1)d_i^*\right]+D(\boldsymbol{x},m+1)\left(1-\sum_{i=1}^{k}d_i^*\right).$$

For the first observation, the following occurs: (0, 1, 0, 1).

For the second observation: (1, 1, 0, 2), (0, 2, 0, 2), (0, 1, 1, 2). Then:

$B(\boldsymbol{x},\ \mathrm{m})\ =\ 0.00001\ +\ [D(1,\ 1,\ 0,\ 2)\ d_1^*\ +\ D(0,\ 2,\ 0,\ 2)d_2^*]\ +\ D(0,\ 1,\ 1,\ 2)[1\ -\ (d_1^*\ +\ d_2^*)]$.

• As the goal is to find the B for which the procedure should stop, then $D\ =\ S$:

$B(\boldsymbol{x},\ \mathrm{m})\ =\ 0.00001\ +\ [S(1,\ 1,\ 0,\ 2)\ d_1^*\ +\ S(0,\ 2,\ 0,\ 2)d_2^*]\ +\ S(0,\ 1,\ 1,\ 2)[1\ -\ (d_1^*\ +\ d_2^*)]$.

• Therefore, S:

$$S(1,1,0,2)=\frac{16}{150}, S(0,2,0,2)=\frac{14}{150} \text{ and } S(0,1,1,2)=\frac{16}{150}.$$

• Therefore,

$$B(\boldsymbol{x},m)=0.00001+\frac{16}{150}0.25+\frac{14}{150}0.50+\frac{16}{150}0.25=0.10001.$$

• Step 6: Compare immediate and expected risk:

Because $0.125\ >\ 0.10001\ \rightarrow$ *Continue*

• Make a new observation:

The prior is the previous estimate: $a_1\ =\ 0.25$; $a_2\ =\ 0.50$; and $a_3\ =\ 0.25$.

The observation belongs to the second class: $x_1 = 0$; $x_2 = 2$; $x_3 = 0$.
- The ratio of the immediate and expected risk is calculated iteratively until the immediate risk is lower than the expected risk.

- Estimate the proportion: $d_i^* = \dfrac{a_i + x_i}{a_0 + m}$.

- Immediate risk:

$$S(\boldsymbol{x},m) = \frac{\sum_{i=1}^{k} K_{ii}(x_i + a_i)\left(\sum_{i=1}^{k}(x_i + a_i)\right) - \sum_{i,j=1}^{k} K_{ij}(x_i + a_i)(x_j + a_j)}{\left[\sum_{i=1}^{k}(x_i + a_i)\right]^2 \left[\sum_{i=1}^{k}(x_i + a_i)\right] + 1}.$$

- Expected risk:

$$B(\boldsymbol{x},m) = c + S(\boldsymbol{x},m)\left(\frac{a_0 + m}{a_0 + m + 1}\right).$$

Application to maize seed damage

First, the priors were constructed to apply the Bayesian sequential estimation technique to the maize seed dataset. The prior of the literature was based on elicitation, where the hyperparameters were obtained from the mean and variance values extracted from the article by Javorski and Cicero (2017), replacing in the expression:

$$E(X_i) = \frac{a_i}{a_0}. \tag{35}$$

A uniform prior is a particular case of Dirichlet distribution; therefore, *a priori* parameters, denoted by hyperparameters, were $a_1 = 1$, $a_2 = 1$, $a_3 = 1$, with all possible parameter values equally probable. The values of *a priori* means (0.3333), variances (0.0556), and covariance (– 0.0278) calculated from the expressions are given in Eq. (6).

Table 2 shows the values of the hyperparameters found for the Dirichlet prior based on the article by Javorski and Cicero (2017), with their respective *a priori* means, variances, and covariances:

A cost of 10^{-5} was selected according to Bach (2015) because it has an order of magnitude similar to the order of magnitude of the loss function $(\boldsymbol{p} - \boldsymbol{d})^T \boldsymbol{K}(\boldsymbol{p} - \boldsymbol{d})$; thus ensuring that the risk function is not dominated exclusively by cost. As the loss is the square of a difference between the actual and estimated proportion values, which are included in the interval [0, 1], the results are always close to zero and, therefore, the cost should also be close to zero. However, the cost is not restricted to these conditions and, in other applications, the use of the correlation matrix can be recommended.

Thus, all necessary calculations were performed for each of the 100 lots using a pivot table built in Microsoft Excel®. A uniform prior was used to begin the Bayesian sequential estimation process; thus, the risks were calculated based on the expressions given in Eq. (18) and (22), according to Jones (1976). From the second seed evaluated, the previous estimates were used as priors to update the information. Thus, the priors followed a Dirichlet distribution and the risks were calculated from the expressions Eq. (33) and (34) developed.

As there are 100 lots, it is not feasible to present all estimates; however, some results of the Bayesian sequential estimates for the uniform prior are shown in Table 3.

The results with the prior from the literature are shown in Table 4.

Dimension 3 simplex was constructed for to better understand the priors (Figure 2).

A frequency distribution of the sample sizes of the lots considering a uniform prior and literature-based prior was obtained (Table 5).

Table 2 – Hyperparameters and values of the mean, variance, and covariance of the prior from the literature.

Classes	a_i	Mean	Variance	Covariance
Without damage	0.45000	0.9000	0.0600	– 0.0507
Density variations	0.04225	0.0845	0.0516	– 0.0093
Other types	0.00775	0.0155	0.0102	– 0.0009

Table 3 – Bayesian sequential estimates of the proportions according to three classifications: $\widehat{p}_1$: seeds without damage; $\widehat{p}_2$: density variations; and $\widehat{p}_3$: other types of damage, using the uniform prior.

Lots	m_{seq}	Types of damage					
		$\widehat{p}_1$:Without damage		$\widehat{p}_2$:Density variations		$\widehat{p}_3$:Other types of damage	
		m_1	$\widehat{p_{1\,seq}}$ (%)	m_2	$\widehat{p_{2\,seq}}$ (%)	m_3	$\widehat{p_{3\,seq}}$ (%)
1	155	134	86.45	13	8.39	8	5.16
2	6	6	99.98	0	0.01	0	0.01
3	6	6	99.98	0	0.01	0	0.01
4	141	125	88.65	15	10.64	1	0.71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	6	6	99.98	0	0.01	0	0.01
99	193	149	77.20	29	15.03	15	7.77
100	170	141	82.94	21	12.35	8	4.71
Mean (%)			87.34		8.70		3.97
Standard deviation (%)			8.70		6.79		3.86

Table 4 – Bayesian sequential estimates of the proportions according to three classifications: $\widehat{p}_1$: seeds without damage; $\widehat{p}_2$: density variations; $\widehat{p}_3$: other types of damage, with the literature-based prior.

Lots	m_{seq}	Types of damage					
		$\widehat{p}_1$:Without damage		$\widehat{p}_2$:Density variations		$\widehat{p}_3$:Other types of damage	
		m_1	$\widehat{p_{1\,seq}}$(%)	m_2	$\widehat{p_{2\,seq}}$(%)	m_3	$\widehat{p_{3\,seq}}$(%)
1	5	5	99.99	13	0.005	0	0.005
2	5	5	99.99	0	0.005	0	0.005
3	5	5	99.99	0	0.005	0	0.005
4	141	125	88.65	15	10.64	1	0.71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	5	5	99.99	0	0.005	0	0.005
99	5	5	99.99	0	0.005	0	0.005
100	170	141	82.94	21	12.35	8	4.71
Mean (%)			89.33		7.43		3.24
Standard deviation (%)			9.27		7.09		3.84

The X-ray test performed by the conventional method with 200 seeds resulted in average estimates of seeds without damage of 83.82 %, with a standard deviation of 5.18 %, 11.38 % with variations in density and a standard deviation of 5.07 %, and 4.80 % with other types of damage, and a standard deviation of 3.49 %. Thus, compared with the Bayesian sequential approach, it can be concluded that the estimates were closer when a uniform *a priori* was used.

Descriptive statistics were obtained for the sample sizes of all lots and the lots that did not stop quickly (selected lots) (Table 6).

Discussion

For some lots, the sampling was interrupted very quickly because there was a very discrepant class in relation to the others, in this case, the class of seeds without damage. The number of cases in which this happened was larger for the literature-based prior.

In the other lots, there was a considerable reduction in seeds evaluated for the estimation of the proportion, which indicates how the Bayesian sequential method for the multinomial distribution, considering the three classes extensively, reduced the sampling time necessary to judge a lot in comparison to the traditional method with 200 seeds.

It is noticed that the mean proportion and the standard deviation for p_2 (Density variations) and p_3 (other types of damages) are practically similar. At p_3, the standard deviation is greater than the mean proportion. It is essential to highlight this result, as it reveals a characteristic that allows detecting the degree of flattening of the distribution, that is, the type of kurtosis. For p_1, there is a leptokurtic distribution; however, for the others, it suggests a platycurtic format in a descriptive way. It should be noted that when considering p_1 through the application proposed in this study, the estimate of this proportion is more accurate (Table 4).
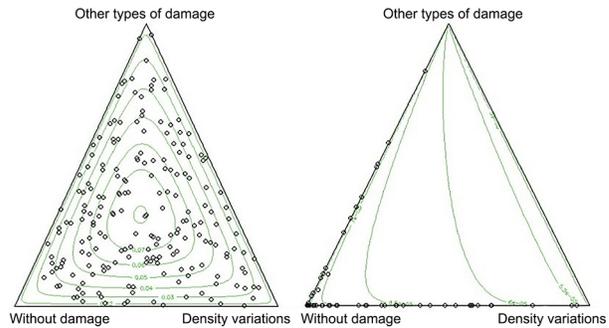
Figure 2 – Uniform prior simplex (left) and the literature prior simplex (right).

Table 5 – Frequency distribution of the sample sizes of the lots considering a uniform prior and literature-based prior.

Sample size	Uniform prior			Literature prior		
	f_i	fr_i	F_i%	f_i	fr_i	F_i%
0 ⊢ 40	25	0.25	25	37	0.37	37
40 ⊢ 80	1	0.01	26	1	0.01	38
80 ⊢ 120	1	0.01	27	1	0.01	39
120 ⊢ 160	28	0.28	55	22	0.22	61
160 ⊢ 200	45	0.45	100	39	0.39	100
Total	100	1.00		100	1.00	

Table 6 – Descriptive statistics of the sample sizes with the different priors adopted for All lots and Selected lots (lots that did not stop quickly).

Prior	Lots	Min	Mean	Median	Standard deviation
Uniform	All	6	125.67	155	72.42
	Selected	72	165.56	169	23.79
Literature	All	5	106.15	145	80.21
	Selected	72	165.56	169	24.13

Many lots stopped with few seeds, and then, the process stabilized and stopped in a region where the estimates were better, but still with a smaller sample size. This shows the need to perform truncation (Table 5).

Thus, when removing the lots that stopped quickly, considering *a priori* uniform, 75 lots remained. The average estimates were 83.12 % with a standard deviation of 5.42 % for seeds without damage, 11.59 % and a standard deviation 5.26 % for seeds with density variations and 5.29 % and a standard deviation 3.59 % for seeds that showed other types.

Considering the literature *priori*, 63 lots remained, and the average estimates were 83.07 %, a standard deviation of 5.45 % for seeds without damage, 11.79 %, and a standard deviation of 5.31 % for seeds with variations in density and 5.14 % and standard deviation 3.69 % for seeds with other types of damage. Evidence that truncation is necessary, as the estimates were very close to the conventional method, but with the advantage of decreasing the sample size.

It can be concluded that the average sample size to estimate the parameters of interest was 165 (Table 6).

Similar results were obtained by Brighenti et al. (2019) when estimating the viability of coffee seeds by the Bayesian sequential method, in which the average percentage of viability using the conventional frequentist method was 88 %, whereas the viability obtained with the Bayesian method with both priors was 89 %. However, on average, the Bayesian method required only 89 samples to reach this value, while the traditional estimation method needed as many as 200 samples.

De Moura et al. (2017) developed sequential sampling plans for *Empoasca kraemeri* (Ross & Moore) (Homoptera: Cicadellidae), a bean crop pest, and determined the levels of economic injury for common bean at different technological levels of cultivation. The results indicated that the sequential sampling plan and the standardized design produced similar decisions. However, in these decisions, there was a time saving of more than 60 % in relation to the standardized plan, showing that the sequential approach also optimized the process.

According to Ali (2019), the Bayesian methodology provides a natural solution for sequential sampling; thus, the Bayesian estimation with a sequential approach has been used in quality control, which justifies the application of this study in seed quality control. The control charts to monitor the process quality, proposed by Ali (2019) and Riaz and Ali (2015), are related to the results of this study.

Therefore, it is concluded that it is possible to determine the stopping criteria for the Bayesian sequential estimation procedure of the multinomial distribution parameters for conjugate Dirichlet priors using dynamic programming equations.

In addition, it is possible to apply the technique addressed in the quality control of maize seeds, obtaining consistent results, with a reduction in the sample size. Sample sizes were much smaller than the conventional approach applied in X-ray testing.

The Bayesian sequential method has advantages over the traditional method. The Bayesian inference stems naturally from the probability theory by treating the parameters as random. This has many advantages, and it means that all inferential issues can be addressed as probability statements about the parameters, which derive directly from *a posteriori* distribution obtained for each lot and offer more information on the proportions of damages estimated in maize seeds.

This method can be applied in several areas to optimize a procedure and minimize costs and operational time. An expert may use the Bayesian sequential procedure to follow yearly crops or to establish the profile in the region. The Bayesian sequential procedure may compare and identify patterns and outliers to yearly crops in a region. The procedure in this study is not restricted to rating damages in maize seeds but it may be adjusted for any experiment in which the population's variable of interest of has more than two response categories.

Therefore, the main advantage of the Bayesian sequential method is the reduction in the sample size, reducing the time and operational costs of a process. However, the disadvantage of this method is the complexity of the calculations involved. Thus, a proposal for future work is to implement this methodology in software with a more user-friendly interface and create applications for easy use, inserting only a few pieces of information while facilitating the decision-making process.

Acknowledgments

Authors' Contributions

Conceptualization: Lima IS, Brighenti CRG. **Data curation:** Lima IS, Brighenti CRG, Pires RMO. **Data acquisition:** Pires RMO. **Formal analysis:** Lima IS. **Methodology:** Lima IS, Brighenti CRG, Pires RMO. **Project administration:** Lima IS, Brighenti CRG. **Supervision:** Brighenti CRG. **Writing-original draft:** Lima IS, Brighenti CRG, Pires RMO. **Writing-review & editing:** Lima IS, Brighenti CRG, Pires RMO.

References

Ali S. 2013. On the Bayesian estimation of the weighted Lindley distribution. Journal of Statistical Computation and Simulation 85: 855-880. https://doi.org/10.1080/00949655.2013.847442

Ali S. 2015. Mixture of the inverse Rayleigh distribution: properties and estimation in a Bayesian framework. Applied Mathematical Modelling 39: 515-530. https://doi.org/10.1016/j.apm.2014.05.039

Ali S. 2019. A predictive Bayesian approach to sequential time-between-events monitoring. Quality and Reliability Engineering International 36: 365-387. https://doi.org/10.1002/qre.2580

Avetisyan M, Fox JP. 2012. The Dirichlet-Multinomial Model for Multivariate Randomized Response Data and Small Samples. Psicológica 33: 362-390. https://files.eric.ed.gov/fulltext/EJ973385.pdf

Bach DR. 2015. A cost minimisation and Bayesian inference model predicts startle reflex modulation across species. Journal of Theoretical Biology 370: 53-60. https://doi.org/10.1016/j.jtbi.2015.01.031

Berger JO. 1985. Statistical Decision Theory and Bayesian Analysis. 2ed. Springer, New York, NY, USA. https://doi.org/10.1007/978-1-4757-4286-2

Brighenti CRG, Cirillo MÂ, Costa ALA, Rosa SDVF, Guimarães RM. 2019. Bayesian sequential procedure to estimate the viability of seeds Coffea arabica L. in tetrazolium test. Scientia Agricola 76: 198-207. http://dx.doi.org/10.1590/1678-992X-2017-0123

Chen SY. 1988. Restricted risk Bayes estimation for the mean of the multivariate normal distribution. Journal of Multivariate Analysis 24: 207-217. https://doi.org/10.1016/0047-259X(88)90036-X

Fenoy MM. 2017. The invariant optimal sampling plan in a sequentially planned decision procedure. Sequential Analysis 36: 194-209. https://doi.org/10.1080/07474946.2017.1319680

Javorski M, Cicero SM. 2017. Use of x-rays in the evaluation of the internal morphology of sorghum seeds. Brazilian Journal of Maize and Sorghum 16: 310-318 (in Portuguese, with abstract in English).

Jones PW. 1976. Bayes Sequential Estimation of Multinomial Parameters. Mathematische Operationsforschung und Statistik 7: 123-127. https://doi.org/10.1080/02331887608801283

Jones PW, Madhi SA. 1988. Bayesian sequential methods for choosing the best multinomial cell: some simulation results. Statistical Papers 29: 125-132. https://doi.org/10.1007/BF02924517

Karunamuni RJ, Prasad NGN. 2003. Empirical bayes sequential estimation of binomial probabilities. Communications in Statistics-Simulation and Computation 32: 61-71. https://doi.org/10.1081/SAC-120013111

Ministério da Agricultura, Pecuária e Abastecimento [MAPA]. 2009. Rules for Seed Testing = Regras para Análise de Sementes. MAPA, Brasília, DF, Brazil (in Portuguese).

Moura MF, Lopes MC, Pereira RR, Parish JB, Chediak M, Arcanjo LP, et al. 2017. Sequential sampling plans and economic injury levels for Empoasca kraemeri on common bean crops at different technological levels. Pest Management Science 74: 398-405. https://doi.org/10.1002/ps.4720

Najar F, Bouguila N. 2022. Exact fisher information of generalized Dirichlet multinomial distribution for count data modeling. Information Sciences 586: 688-703. https://doi.org/10.1016/j.ins.2021.11.083

Paulino CD, Turkman MAA, Murteira B, Silva GL. 2018. Bayesian Statistics = Estatística Bayesiana. 2 ed. Fundação Calouste Gulbenkian, Lisboa, Portugal (in Portuguese).

Pham-Gia T. 1998. Distribution of the Stopping Time in Bayesian Sequential Sampling. Australian e New Zealand Journal of Statistics 40: 221-227. https://doi.org/10.1111/1467-842X.00025

Plant RE, Wilson LT. 1985. A Bayesian Method for Sequential Sampling and Forecasting in Agricultural Pest Management. Biometrics 41: 203-214. https://doi.org/10.2307/2530655

Pratt JW, Raiffa H, Schlaifer R. 1964. The Foundations of Decision under Uncertainty: An Elementary Exposition. Journal of the American Statistical Association 59: 353-375. https://doi.org/10.1080/01621459.1964.10482164

Riaz M, Ali S. 2015. On process monitoring using location control charts under different loss functions. Transactions of the Institute of Measurement and Control 38: 1107-1119. https://doi.org/10.1177/0142331215583325

Schnuerch M, Erdfelder E, 2020. Controlling decision errors with minimal costs: The sequential probability ratio t Test. Psychological Methods 25: 206-226. https://doi.org/10.1037/met0000234