Acta Scientiarum
http://www.periodicos.uem.br/ojs/
ISSN on-line: 1807-8621
Doi: 10.4025/actasciagron.v47i1.72116

CROP PRODUCTION

Vigor and damage of soybean seeds estimated by sequential Bayesian techniques in the tetrazolium test

Maria Luiza Capellari Leite da Silva[1], Isabela da Silva Lima[1*], Bruna Cardoso Braga[2] and Carla Regina Guimarães Brighenti[1,3]

[1]Departamento de Estatística, Universidade Federal de Lavras, Campus Universitário, Cx. Postal 3037, 37200-000, Lavras, Minas Gerais, Brazil. [2]Programa de Pós-Graduação em Zootecnia, Escola de Veterinária e Zootecnia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil. [3]Departamento de Zootecnia, Universidade Federal de São João del-Rei, Campus Tancredo Neves, São João del-Rei, Minas Gerais, Brazil. *Author for correspondence. E-mail: isabela_lima30@hotmail.com

ABSTRACT. Soybean is an oilseed of great relevance to Brazil because it is one of the main crops produced in the country. The success of this crop depends on several factors, the main one being seed quality. The tetrazolium test has been used for quality control because in addition to evaluating the viability of germination and the vigor of the seed lots, it facilitates classification of the types of damage. Thus, the use of sequential sampling, which takes into account the immediate evaluation of each seed, may facilitate an earlier decision on viability and vigor without the need to evaluate all seeds as in the classic test as a formality of the method. In addition, the estimates obtained through Bayesian techniques can be improved by including prior information. The objective of this study was to study the application of Bayesian sequential tests to estimate the proportion of vigor in lots (binomial modeling) and the proportion of damage per category in soybean seeds (multinomial modeling) through the tetrazolium test. In each case, conjugate priors were used, and the parameters were elicited. It can be concluded that the four approaches, frequentist, Bayesian, sequential, and Bayesian sequential, were efficient for estimation of and decision-making about the parameters, with a reduction in the sample size. Moisture damage was present in 20.17% of the soybean seeds evaluated, damage by stink bugs in 3.50%, and mechanical damage, in the case of manual harvesting, in only 1.92%. In addition, 1.00% of seeds presented more than two types of damage.

Keywords: *Glycine max* (L.) Merrill; multinomial distribution; dirichlet; prior elicitation.

Received on May 4, 2024.
Accepted on August 31, 2024.

Introduction

Soybeans (*Glycine max* (L.) Merrill) are the main agricultural product in Brazilian exports and the main product responsible for the increase in the domestic grain harvest. The success of the crop depends on several factors, but undoubtedly, the most important of these is the use of high-quality seeds that generate high-vigor plants and, therefore, superior performance in the field (Toloi et al., 2021). To ensure this requirement, it is necessary to employ an agile, versatile, and reliable quality control system, offering fast and accurate results (Lopes et al., 2020).

Among the main factors compromising seed quality are those that reduce the physiological quality of soybean seeds, causing their deterioration, such as moisture damage, stink bugs, and mechanical damage. Excessive moisture results in wrinkles on the seed coat, while pests like stink bugs cause tissue necrosis. Mechanical damage significantly impacts seed quality, affecting germination and seedling emergence. Additionally, seeds can be simultaneously affected by two or more types of damage. Furthermore, recent climate changes are also a factor that can alter seed quality (Abati et al., 2021).

It is important to ensure the physical integrity of seeds, as it plays a significant role in achieving high productivity levels. Thus, it is essential for seeds to undergo rigorous quality control before reaching the farmer (Jesus et al., 2021).

To evaluate the viability of germination and the vigor of the seed lots and to identify the different types of damage in the seeds, the tetrazolium test has been used in the seed industry in Brazil for quality control, mainly for soybean crops (Marcos-Filho, 2015). In the tetrazolium test, seeds are classified according to different vigor classes (vigorous or nonvigorous) and the types of damage they present (moisture, stink bug, and/or mechanical) (França-Neto & Krzyzanowski, 2022).

Thus, each evaluated seed can be classified as vigorous or nonvigorous, characterizing a Bernoulli experiment, in which *p* corresponds to the proportion of vigor and there is, in a lot with sample size *n*, a binomial distribution. In the case of grain damage assessment, in which there are more than two categories, the modeling follows a multinomial distribution, in which each type of damage corresponds to a category. However, multinomial modeling has rarely been used for seed analysis. However, it can be a useful tool for analyzing the problems that affect seed quality and for categorizing them, contributing to the advancement of decision-making tests of soybean seed lots, as well as any other product with more than two categories (Lima et al., 2024a).

Obtaining representative samples of a lot, with possible cost reduction, was always the objective of the sampling process. One of the ways to reduce these costs is to replace classic sampling, with a fixed number of evaluated seeds, by sequential sampling that allows the use of a variable number of sampling units as a function of the quality of the lot. Sequential sampling has been shown to be faster than classical sampling and is recommended by the International Seed Testing Association (ISTA) (Mulgan, 2021). However, sequential sampling is still rarely used due to the regular need for computational support to assist in the calculations during the decision-making process. Regional and seasonal aspects can also be included in the sampling plan, incorporating information related to these aspects through Bayesian techniques, using a prior distribution (Schnuerch & Erdfelder, 2020).

Thus, the main objective of this work is to study the application of Bayesian sequential tests to evaluate the vigor and damage proportions per category in soybean seeds using the tetrazolium test. This proposal opens the possibility of incorporating information for the discrimination of seed quality.

Material and methods

The data from the tetrazolium test were obtained from the Central Seed Laboratory of the Department of Agriculture of the Federal University of Lavras from the evaluation of soybean seeds from producers in the southern region of Minas Gerais State located in the municipality of Lavras, Minas Gerais State, Brazil, at 21°14'43" south and 44°59'59" west with an altitude of 919 meters above sea level.

To validate the procedure proposed in this study, the traditional analysis of the test was performed in twelve lots analyzed with two replicates of 50 seeds, totaling 100 seeds per lot, according to the standards established in the Seed Analysis Rules (Brasil, 2009).

The seeds were preprimed on moistened paper towels, remaining for 16h in the germinator at 25°C. After this period, the seeds were immersed in a salt solution of 2,3,5 triphenyltetrazolium chloride and placed in incubators at 40°C for 3h. In this evaluation, the seed coats were removed, and then the seeds were sectioned longitudinally through the center of the embryonic axis. Then, the cotyledons were separated, leaving the vascular region exposed for evaluation of all types of damage.

From the recorded data, the percentage of vigorous seeds was evaluated under three approaches:

i) percentage of vigorous (p*)* and nonvigorous seeds, following a binomial model;

ii) classification of the seeds into three categories, considering only the nonvigorous seeds and evaluation of the percentages of the three types of damage, moisture, stink bug, and mechanical; the seeds that showed two or more types of damage were included in the damage category that led to the death of the embryo;

iii) classification into five classes, considering vigorous seeds, seeds with one of the three types of damage: moisture, stink bug or mechanical and seeds with two or more type of damage.

The traditional analysis of the tetrazolium test was performed to estimate, considering the binomial distribution, the proportion of vigorous (p) and nonvigorous seeds, in which $S \sim Binomial\ (n, p)$ is written as (Hamura, 2023):

$$f(S|p) = f_S(S) = \binom{n}{S} p^S (1-p)^{n-S} \tag{1}$$

Where $0 \leqslant p \leqslant 1$ and S corresponds to the sum of the vigorous seeds.

The estimated proportions of damaged seeds follow a multinomial distribution consisting of n attempts, and each attempt can result in any of the k possible outcomes: d_1, d_2, ..., d_k. Suppose, in addition, that each possible result can occur with probabilities p_1, p_2, ..., p_k. Therefore, the probability p_1 that d_1 occurs n_1 times, d_2 occurs n_2 times..., and d_k occurs n_k times is (Najar & Bouguila, 2022):

$$P(n_k|p_k) = \left[\frac{n!}{n_1!n_2!\ldots n_k!}\right]\left(p_1^{n_1}.p_2^{n_2}.\ldots.p_k^{n_k}\right) \tag{2}$$

where $n = n_1 + n_2 + \cdots + n_k$

For the Bayesian analysis, in the case of random variables with a domain between 0 and 1, such as proportions, the beta distribution is often used as a conjugate prior distribution of binomial probabilities (Avci, 2021). The probability density function of the beta prior distribution with hyperparameters α and β is given by:

$$P(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad (3)$$

where $0 \leqslant p \leqslant 1$, $\alpha > 0$ and $\beta > 0$.

In Bayesian analysis, the difficulty of obtaining and quantifying a prior is a recurrent issue, and elicitation is a very useful process for prior construction (Stefan et al., 2022). The procedure for constructing the priors begins with the supply of elicited mean and variance values given by:

$$\mu_{prior} = \mu_{beta} = \frac{\alpha}{(\alpha+\beta)}, var_{prior} = \frac{(\alpha\beta)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Knowing the mean and variance for the proportion, the estimates of α and β through the method of moments are given by:

$$\alpha = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \qquad \beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right)$$

To obtain the beta posterior distribution (α', β') or $(\alpha + s, \beta + n - s)$, where s is the sum of vigorous seeds, the mean and variance are given by:

$$\mu_{posterior} = \frac{\alpha+s}{(\alpha+s+\beta+n-s)}, \qquad var_{posterior} = \frac{(\alpha+s)(\beta+n-s)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}.$$

In the case of the multinomial distribution, the conjugate prior distribution is the multivariate generalization of the beta distribution, known as the Dirichlet distribution (Yi et al., 2020).

The Dirichlet distribution is applied to each set of probabilities k, where the probability vectors p_k follow a Dirichlet distribution with parameters α_k

$$P(p_k|\alpha_k) \propto \prod_{k=1}^{k} p_k{}^{\alpha_k - 1} \quad (4)$$

Similarly, the prior construction procedure begins with the provision of the elicited values of the mean (proportion) and variance (variation) for each type of damage given by:

$$\mu_{prior} = \frac{\alpha_k}{\sum \alpha_k},$$

$$var_{prior} = \frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(\alpha_0+1)},$$

where $\alpha_0 = \alpha_1 + \cdots + \alpha_k$, so that the higher the value of α is, the smaller the variance.

Santos et al. (2021), Batista et al. (2022), and Ribeiro et al. (2024) also evaluated the vigor of soybean seeds. However, in this study, we chose to use the mean and variance of the proportion of vigorous seeds presented in the paper by Costa et al. (2005), Schuab et al. (2007), Lopes et al. (2011) and data for the 2016/2017 soybean harvest in Lorini (2018) as values for construction of the priors, as they have the same number of categories worked on in this article.

Through the Bayes theorem, the equations $P(n_k \mid p_k)$ and $P(p_k \mid \alpha_k)$ produce a posterior Dirichlet distribution for p_k with parameters $\alpha_k + n_k$, with (Boukabour & Masmoudi, 2021):

$$\mu_{posterior} = \frac{\alpha_k + n_k}{\sum_k(\alpha_k + n_k)},$$

$$var_{posterior} = \frac{p_k(1-p_k)}{[1+\sum_k(\alpha_k+p_k)]}.$$

To estimate the proportion parameter considering the sample size as a random variable in the sequential estimation, elements of a sample are extracted one at a time, and after each seed is observed, a stopping criterion is used, and the decision is made, or the observation of samples continues.

To obtain the stopping criterion in the binomial distribution, the quadratic loss function $L = (p - \hat{p})^2 + C(n)$ is considered, where $C(n) > 0$ is the cost of taking a sample of size n to estimate p. The expected posterior variance is calculated, which allows estimation of the amount $r(\pi^n, n)$ that represents the lowest Bayes risk that can be achieved since X^n was observed (Hwang, 2020).

To decide whether to stop sampling, the risk values are compared to each seed evaluated, and a new observation is performed until $r_0(\pi^n, n) \leqslant r^1(\pi^n, n)$, when the estimate of the viability proportion parameter in the tested lot is then calculated from the mean of the posterior obtained so far.

Thus, it is necessary to calculate, at each seed evaluation, the immediate risk given by the posterior variance plus the cost of n observations and the expected risk given by the expected posterior variance and the increase in the cost of another observation (Brighenti et al., 2011); thus:

$$r_0(\pi^n, n) = \frac{(\alpha+s)(\beta+n-s)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} + C(n), \qquad r^1(\pi^n, n) = E[var_{posterior}] + C(n+1).$$

Then, the expected risk is given by:

$$r^1(\pi^n, n) = \left(\frac{\alpha+\beta+n}{\alpha+\beta+n+1}\right)\frac{(\alpha+X)(\beta+n-s)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} + C(n+1) \tag{5}$$

The procedure for constructing the *priors,* as well as obtaining the hyperparameter estimates, was developed in the software *R* (R Core Team, 2024).

For a multinomial distribution with $(k + 1)$ classes, according to Jones (1976), the probability of an observation in the i-th class, for $i = 1, 2, \dots, k$, and in $(k + 1)$ can be $\left(1 - \sum_{i=1}^{k} p_i\right)$. Suppose that the prior information about $p = (p_1, p_2, \dots, p_k)^T$ can be adequately represented by a member of the natural conjugate Dirichlet, a family of distributions with integer parameters m_0, n_{0i}, $i = 1, 2, \dots, k$, with density proportional to:

$$\prod_{i=1}^{n} p_i^{n_{0i}-1}\left(1 - \sum_{i=1}^{k} p_i\right)^{m_0 - \sum n_{0i} - 1} \tag{6}$$

with $p_i \geqslant 0$ and $\sum_{i=1}^{k} p_i \leq 1$.

By the Bayes theorem, after m observations resulting in n_i in the i-th classes, the posterior density of p is given by a Dirichlet distribution with parameters $m + m_0$ and $n_i + n_{0i}$.

The result of the sampling, because it is represented as a sample path, starts at the point (n_0, m_0), where $n_0 = (n_{01}, n_{02}, \dots, n_{0k})$, in the $(k + 1)$-dimensional whole space, and is interrupted when the stop limit—which has to be determined—is reached. If $m_0 = k + 1$ and $n_{0i} = 1$ (uniform prior for p) are taken as the origin, then any other appropriate prior with integer parameters gives sample paths starting at $(n_0 - 1, m_0 - k - 1)$, where 1 is the unit line vector.

Consequently, a uniform prior is used to obtain stop limits. Suppose that the loss in the estimate of p by $d = (d_1, d_2, \dots, d_k)^T$ has the general quadratic form $(p - d)^T K(p - d)$, where K is a positive symmetric matrix of constant loss, and the Bayes estimator d^* is the mean of the posterior distribution of (n, m) with elements $d_i^* = (n_i + 1)/(m + k)$ this is also the marginal posterior probability that the next observation falls within the i-th class. The risk of stopping is given by (Jones, 1976):

$$S(n, m) = \text{traço } K\Sigma \tag{7}$$

where Σ is the dispersion matrix of the posterior distribution with the elements: $var\,(p_i) = \frac{d_i^*(1-d_i^*)}{m+k+2}$, $cov\,(p_i, p_j) = -\frac{d_i^* d_j^*}{m+k+2}$.
consequently,

$$S(n, m) = \frac{\left\{\sum_{i=1}^{k} K_{ii} d_i^* - \sum_{ij=1}^{k} K_{ij} d_i^* d_j^*\right\}}{m+k+2} \tag{8}$$

For more details on the quadratic loss function see Tan (2020). To obtain the stopping criterion in the multinomial distribution, one point (n, m) is considered, where $B(n, m)$ is the risk of an additional observation at a cost c to continue thereafter (risk of continuing) and $D(n, m)$ is the minimum risk. Then, the dynamic programming equations giving the partition into stopping and continuation points are:

$$D\,(n, m) = min\,[S(n, m), B(n, m)] \tag{9}$$

$$B\,(n, m) = c + \sum_{i=1}^{k}[D(n + e_i, m + 1) d_i^*] + D(n, m + 1)\left(1 - \sum_{i=1}^{k} d_i^*\right) \tag{10}$$

e_i is the vector line with 1 to the i-th position and zeros everywhere else.

Since $S(n,m) \to 0$, $B(n,m) \to c$ and $m \to \infty$, there is a large value of $m = N^*$ such that all points (n, N^*) are stopping points for $\sum_{i=1}^{k} n_i \leq N^*$. The dynamic programming equations above can now be successively used for $m \leqslant N^*$ to find the smallest integer m satisfying:

$B(n, m) > S(n, m)$, $D(n, m + 1) = S(n, m + 1)$

for all n, and this is the maximum sample size.

The two conditions imply that:

$$B\ (n, m) = c + \left[\frac{\mathrm{m+k+1}}{\mathrm{m+k+2}}\right] S(n, m) \tag{11}$$

For more details on obtaining the stopping criterion, see Lima et al. (2024a).

Results and discussion

In the frequentist estimation of vigor optimization using the binomial distribution with classification of vigorous and nonvigorous seeds (Table 1), vigorous seeds are those that do not present any type of damage, and nonvigorous seeds are those that have some damage that may have affected the embryo or not.

Table 1. Percentages of vigorous and nonvigorous seeds.

Lot	1	2	3	4	5	6	7	8	9	10	11	12	Mean (%)	Standard deviation (%)	CV (%)
Vigorous	81	54	64	62	80	78	84	75	83	60	84	76	73.42	10.54	14.36
Nonvigorous	19	46	36	38	20	22	16	25	17	40	16	24	26.58	10.54	39.67

Thus, the mean vigor of the lots is 73.42% with a standard deviation of 10.54%. To assess the relevance of the two classifications, the chi-square test was performed, which shows a test statistic of 62.66 with a p-value < 0.001. This shows that there are differences between the percentages of vigorous and nonvigorous seeds of the lots.

To estimate the proportion of vigorous seeds of the lots with the Bayesian approach, a beta prior was used (α, β), knowing that after observing the sample, the posterior distribution is beta $(\alpha + s, \beta + n - s)$, where s is the sum of vigorous seeds.

The two priors used were based on the following information: prior 1 using a mean of 71% vigorous seeds and a coefficient of variation (CV) of 7.22% presented by Schuab et al. (2007) when studying the germination test under water stress to evaluate the vigor of soybean seeds; and prior 2 of 83% vigorous seeds with a CV of 8.65% proposed by Lopes et al. (2011) when studying the effects of mechanical and physiological damage on soybean seed harvesting and processing. From the cited references, the hyperparameters and percentages of the means and variances of the priors were elicited using the method of moments are shown in Table 2.

Table 2. Values of hyperparameters, means (%) and variances (%) of priors.

Prior	Hyperparameters		Mean (%)	Variance (%)
	α	β		
Schuab et al. (2007)	0.7044	0.2877	71.00	10.34
Lopes et al. (2011)	0.8277	0.1695	83.00	7.06

Table 3 shows the results of different posterior distributions. Using a size n = 100 and different priors, similar results are obtained in the posterior, indicating consistency in the analysis.

Table 3. Percentages of means and variances of the posteriors.

Posterior	Mean (%)	Variance (%)
Schuab et al. (2007)	73.40	0.19
Lopes et al. (2011)	73.51	0.19

The frequentist approach to damage estimation was performed with a multinomial distribution in three classes (Table 4) and five classes (Table 5).

Of the 319 nonvigorous seeds obtained in the 12 lots, a mean of 77.08% and a standard deviation of 13.03 show moisture damage, compared to a mean of 14.84% and a standard deviation of 11.80 for damage by stink bugs and a mean of 8.08% and standard deviation of 9.42 for mechanical damage.

Table 4. Percentages of seeds according to the three classifications.

Lot	$\hat{p}_1$: moisture damage	$\hat{p}_2$: stink bugs damage	$\hat{p}_3$: mechanical damage
1	47.37	42.10	10.53
2	78.26	13.04	8.70
3	86.11	8.33	5.56
4	92.11	5.26	2.63
5	85.00	10.00	5.00
6	72.73	27.27	0.00
7	93.75	6.25	0.00
8	72.00	20.00	8.00
9	64.71	0.00	35.29
10	85.00	12.50	2.50
11	68.75	25.00	6.25
12	79.17	8.33	12.50
Mean (%)	77.08	14.84	8.08
Standard deviation (%)	13.03	11.80	9.42

Table 5. Percentages of seeds according to the five classifications.

Lot	$\hat{p}_1$: vigorous	$\hat{p}_2$: moisture damage	$\hat{p}_3$: stink bugs damage	$\hat{p}_4$: mechanical damage	$\hat{p}_5$: more than one type of damage
1	81	7	8	2	2
2	54	35	6	4	1
3	64	29	3	2	2
4	62	35	1	1	1
5	80	14	2	1	3
6	78	16	5	0	1
7	84	15	1	0	0
8	75	18	5	2	0
9	83	11	0	6	0
10	60	32	5	1	2
11	84	11	4	1	0
12	76	19	2	3	0
Mean (%)	73.42	20.17	3.50	1.92	1.00
Standard deviation (%)	10.54	9.94	2.39	1.73	1.04

In the evaluation of vigor with the multinomial distribution in five classes, the moisture damage stands out, with a mean of 20.17%, compared to the presence of stink bugs, with a mean of 3.50%, mechanical damage, with a mean of 1.92%, and the presence of more than one type of damage, with 1.0% (Table 5).

The results found in the two forms of classifications, with three or five classes, for moisture damage in the analyzed lots is in agreement with the data found in the literature (Souza et al., 2024).

The mechanical damage in the analyzed lots is lower than that found in the literature due to the manual harvesting of the grains instead of the use of a harvester (Martins et al., 2016).

The impact of stink bug damage on the analyzed lots is not considerable. Costa et al. (2005), when studying the main factors responsible for the reduction in soybean seed quality, found that an incidence of stink bug damage up to 5% does not affect seed vigor and that only after 6% is there a significant decrease in the vigor rate.

For the Bayesian approach with the three types of damage, moisture, stink bug, and mechanical, considering the multinomial distribution, a conjugate Dirichlet prior was used (α_1, α_2, α_3), knowing that after observing the sample, the posterior distribution is a Dirichlet (+ n_1, ..., $\alpha_k + n_k$), where n_k is the frequency of each α_k.

For the prior, the parameters applied were obtained by Costa et al. (2005) when studying the physical, physiological, and chemical aspects of soybean seeds produced in six regions of Brazil, where 11.17% of seeds had mechanical damage, 4.77% had stink bug damage, and 6.8% had moisture damage. Thus, the prior is the Dirichlet prior (0.2456; 0.1048; 0.1496), with the mean and variance values shown in Table 6.

Table 6. Values of the prior of multinomial with three parameters.

Type of damage	α_k	Mean (%)	Variance (%)
Mechanical	0.2456	49.12	16.66
Stink bugs	0.1048	20.97	11.05
Moisture	0.1496	29.91	13.98

In the analyzed lots, the frequency of mechanical damage is 23 seeds, that of stink bug damage is 42 seeds, and that of moisture damage is 242. These values were used to obtain the posterior, which is the Dirichlet posterior (23.2456; 42.1048; 242.1496). Table 7 shows the means and variances of the posterior for each parameter.

Table 7. Values of the posterior of the multinomial with three parameters.

Type of damage	$\alpha_k + y_k$	Mean (%)	Variance (%)
Mechanical	23.2456	7.54	0.02
Stink bugs	42.1048	13.67	11.80
Moisture	242.1496	78.79	16.81

To better explain the impact of the parameter α_k for the prior and posterior, the simplex of dimension k-1 is presented in Figure 1, and the points indicate the occurrence of damage; it is possible to observe that the prior is uniform and that the posterior tends to have a higher occurrence of moisture damage (Inacio et al., 2022).

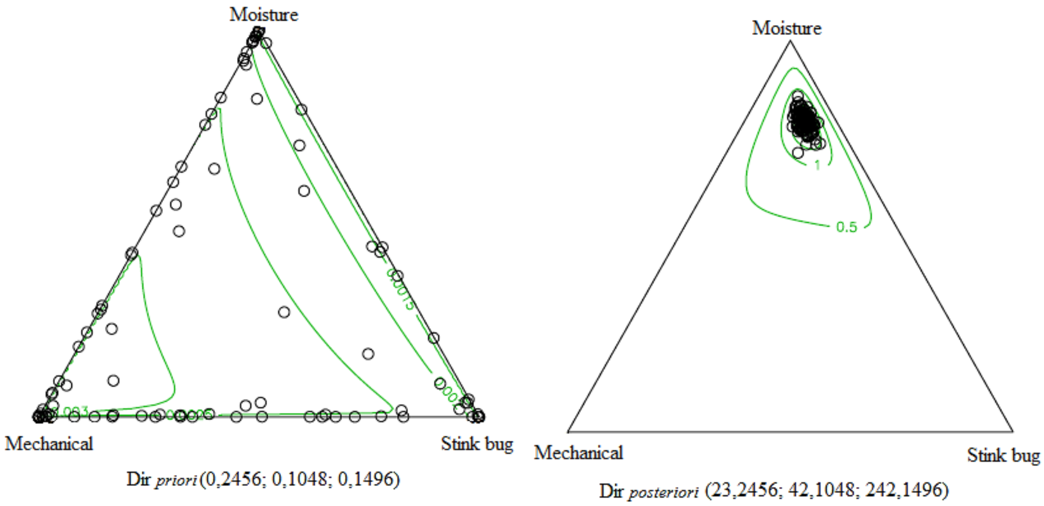

Figure 1. Simplex of the Dirichlet prior with parameters (0.2456; 0.1048; 0.1496) and posterior with parameters (23.2456; 42.1048; 242.1496). Source: Prepared by the author.

The results found by the Bayesian estimation of the three types of damage are consistent with the values found in the experiment, which show means of 8.08% for mechanical damage, 14.84% for stink bug damage, and 77.08% for moisture damage.

To estimate the proportion of seeds in the five classes shown in Table 8 with a multinomial distribution, data from the 2016/2017 soybean harvest were used Lorini (2018) to obtain the parameters. Therefore, the proportion for obtaining the prior for vigorous seeds was 82%, that for moisture damage was 3.1%, that for stink bug damage was 0.7%, that for mechanical damage was 4.9%, and that for the incidence of two or more types of damages was unspecified, so 0% was used. Thus, the conjugate prior is a Dirichlet prior (0.4520; 0.0171; 0.0039; 0.0270; 0.0000), with the mean and variance values shown in Table 8.

Table 8. Values of the prior of the multinomial with five parameters.

$\hat{p}$	α	Mean (%)	Variance (%)
$\hat{p}_1$: vigorous	0.4520	90.41	5.78
$\hat{p}_2$: moisture damage	0.0171	3.42	2.20
$\hat{p}_3$: stink bugs damage	0.0039	0.77	0.51
$\hat{p}_4$: mechanical damage	0.0270	5.40	3.41
$\hat{p}_5$: more than one type of damage	0.0000	0.00	0.00

To calculate the Dirichlet posterior, the frequencies of each parameter were used, with 881 vigorous seeds, 242 with moisture damage, 42 with stink bug damage, 23 with mechanical damage, and 12 with two or more types of damage. Therefore, the posterior is a Dirichlet posterior (881.4520; 242.0171; 42.0039; 23.0270; 12.0000), with means and variances shown in Table 9.

Table 9. Values of the posterior of the multinomial with five parameters.

$\hat{p}$	$\alpha_k + y_k$	Mean (%)	Variance (%)
$\hat{p}_1$: vigorous	881.4520	73.42	0.02
$\hat{p}_2$: moisture damage	242.0171	20.16	16.09
$\hat{p}_3$: stink bugs damage	42.0039	3.50	3.38
$\hat{p}_4$: mechanical damage	23.0270	1.92	1.88
$\hat{p}_5$: more than one type of damage	12.0000	1.00	0.99

The results obtained in the multinomial Bayesian analysis with five classifications are also consistent with those of the experiment, which show means of 73.42% for vigorous seeds, 20.17% for those with moisture damage, 3.50% for those with stink bug damage, 1.92% for those with mechanical damage, and 1.00% for those with two or more types of damage.

Bayesian sequential estimation was performed using a Delphi application developed by Brighenti et al. (2011) in the twelve lots. Table 10 shows the results of these reports with the individual results of the lots.

Table 10. Percentage vigor of the seed lots estimated by the Bayesian sequential method, with their respective final sample sizes (N_{seq}).

Lot	Vigor			
	60%		75%	
	N_{seq}	$\hat{p}_{seq}$ (%)	N_{seq}	$\hat{p}_{seq}$ (%)
1	12	63.88	12	71.60*
2	12	41.49*	13	37.61*
3	12	63.88	12	63.83*
4	12	63.88	12	63.83*
5	12	63.88	12	63.83*
6	8	91.06	8	92.61
7	8	91.06	8	92.61
8	12	63.88	12	63.83*
9	11	77.10	11	77.63
10	12	71.34	12	71.60*
11	10	83.86	10	84.77
12	11	77.10	11	77.63
Mean (%)	11	71.03	11.08	71.78
Standard deviation (%)	1.54	14.04	1.62	15.15

*: Lots were rejected.

For the binomial sequential Bayesian estimation of the proportion of vigorous seeds in soybean lots, this study utilized two different prior distributions for the purpose of comparing the criterion. The first prior distribution considered has a mean of 60%, corresponding to a lot with a low vigor level, with only 60% of vigorous seeds. The second prior distribution has a mean of 75%, corresponding to a lot with a higher vigor level, with 75% of vigorous seeds. Thus, based on these two prior distributions, the sequential Bayesian estimation of the proportion of vigorous seeds was performed. It was observed that with the first prior distribution, only lots with an estimated vigor above 60% were accepted, thus only rejecting lot 2. With the second prior distribution, only lots with estimated vigor greater than 75% were accepted, resulting in a stricter criterion that rejected lots 1, 2, 3, 4, 5, 8, and 10 (Table 10). Still in Table 10, it can be observed that the maximum sample size per lot was 12 seeds. The case presented here demonstrates the influence that the choice of the prior distribution can have on the results of parameter estimates. The priors used reflect soybean seed lots of low vigor, especially in the first case (60%). Consequently, the mean vigor estimates were 71.03% for the first prior and 71.78% for the second prior, slightly lower compared to the results obtained by the frequentist approach (73.42%) and the Bayesian approach (73.40% and 73.51%), but still close, indicating the consistency of the analyses.

To establish whether sampling should be continued, it is necessary for the risk values to be compared for each seed evaluated and for a new observation to be performed until the immediate risk ($S(n,m)$) is less than or equal to the expected risk ($B(n,m)$).

However, to find the immediate and expected risk values, we must first find the values of the elements d_i^* of the posterior distribution of (n,m), given by $d_i^* = \frac{n_i+1}{m+k}$, where n_i is the cumulative values of defects, m is the observations, and k is the number of classes. Subsequently, the calculation of the immediate risk is obtained.

The next step was to calculate the expected risk, for which it was necessary to know the cost of another observation. The cost adopted was that used in the Bayesian sequential sampling of 10^{-4} vigor.

The multinomial Bayesian sequential estimation was performed using a dynamic table in Microsoft Excel® in the twelve lots proposed for this study, using a prior with three classes presented by Costa et al. (2005) and the five classes of data from the 2016/2017 soybean harvest presented by Lorini (2018).

As an example, we can cite the evaluation of the first seed of lot 4 with a cost of 10^{-3}, because when using the cost of 10^{-4}, sampling was not interrupted in any of the twelve lots. For the prior, the parameters of Costa et al. (2005) obtained in the study of the physical, physiological and chemical aspects of soybean seeds produced in six regions of Brazil, with 11.17% of seeds with mechanical damage, 4.77% with stink bug damage, and 6.8% with moisture damage, the immediate risk ($S(n,m)$) and the expected risk ($B(0000000000,m)$) for the multinomial with three parameters were calculated from the Dirichlet prior (0.1496; 0.1048; 0.2456) whose posterior is a Dirichlet posterior (1.1496; 0.1048; 0.2456). Considering the seeds included in class 1, we have $S(0.1667) > B(0.1390)$, and sampling should continue. The risk calculations must be performed until the expected risk is greater than the immediate risk, thus ending the sampling and providing an estimation.

The results found by the multinomial Bayesian sequential estimation, presented in Table 11, with three parameters, with the cost of one more observation of 10^{-3}, are consistent with the values of the frequentist approach, which shows means of 8.08% in mechanical damage, 14.84% in stink bug damage, and 77.08% in moisture damage, and consequently, with the results also found with the Bayesian estimation.

Table 11. Damage percentages of seed lots estimated by the multinomial Bayesian sequential method, with their respective final sample sizes (N_{seq}), according to three parameters.

Lot	N_{seq}	Damage					
		$\hat{p}_1$: moisture damage		$\hat{p}_2$: stink bugs damage		$\hat{p}_3$: mechanical damage	
		N_1	$\hat{p}_{1\ seq}$ (%)	N_2	$\hat{p}_{2\ seq}$ (%)	N_3	$\hat{p}_{3\ seq}$ (%)
1*	19	9	47.37	8	42.10	2	10.53
2	25	20	80.00	3	12.00	2	8.00
3	25	22	88.00	2	8.00	1	4.00
4	25	24	96.00	0	0.00	1	4.00
5*	20	17	85.00	2	10.00	1	5.00
6*	22	16	72.73	5	22.72	1	4.55
7*	16	15	93.75	1	6.25	0	0.00
8	25	18	72.00	5	20.00	2	8.00
9*	17	11	64.71	0	0.00	6	35.29
10	26	21	80.77	4	15.38	1	3.85
11*	16	11	68.75	4	25.00	1	6.25
12*	24	19	79.17	2	8.33	3	12.5
Mean (%)		77.35		14.15		8.50	
Standard deviation (%)		10.20		9.08		5.47	

*: Sampling was not interrupted.

Note that for lots 1, 5, 6, 7, 9, 11, and 12, it was not possible to perform the multinomial sequential analysis because the numbers of samples were small. Thus, the minimum number of samples to interrupt the sampling of the lots was 25 seeds.

Table 12 presents the results for the multinomial with five classes, with the cost of one more observation of 10^{-4}, whose values are in agreement with those found in the experiment, with means of 73.42% for vigorous seeds, 20.17% for those with moisture damage, 3.50% for those with stink bug damage, 1.92% for seeds with mechanical damage, and 1% for seeds with two or more types of damage.

In addition, the maximum number of samples per lot is 94 seeds to stop sampling and thus to make a decision. This indicates how the multinomial Bayesian sequential method considering seed damage reduces the sampling time required to make a decision about a lot compared to the traditional method of 100 seeds.

It is important to highlight that the major limitation of the sequential Bayesian method lies in the existence of complex calculations during the decision-making process, specifically in determining the stopping criterion that establishes the optimal moment to stop and estimate the parameters of interest. This is due to the recurrence relationship between the observations, which determines the final sample size. Thus, the procedure may require computational support to assist in the execution of the method. However, for data from the binomial distribution, the Delphi application developed by Brighenti et al. (2011) can be used, and for the multinomial distribution, a dynamic table in Microsoft Excel®. Both are easy to implement and friendly to producers.

Table 12. Damage percentages of seed lots estimated by the multinomial Bayesian sequential method, with their respective final sample sizes (N_{seq}), according to five parameters.

:Lot	N_{seq}	Damage									
		$\hat{p}_1$: vigorous		$\hat{p}_2$: moisture damage		$\hat{p}_3$: stink bugs damage		$\hat{p}_4$: mechanical damage		$\hat{p}_5$: more than one type of damage	
		N_1	$\hat{p}_{1\ seq}$ (%)	N_2	$\hat{p}_{2\ seq}$ (%)	N_3	$\hat{p}_{3\ seq}$ (%)	N_4	$\hat{p}_{4\ seq}$ (%)	N_5	$\hat{p}_{5\ seq}$ (%)
1	94	77	81.91	7	7.45	6	6.38	2	2.13	2	2.13
2	94	49	52.13	35	37.23	6	6.38	3	3.19	1	1.06
3	93	57	60.64	29	31.18	3	3.23	2	2.15	1	1.08
4	93	58	61.70	32	34.41	1	1.08	1	1.08	0	0.00
5	93	74	79.72	14	15.05	2	2.15	0	0.00	2	2.15
6	93	71	75.53	16	17.20	5	5.38	0	0.00	0	0.00
7	94	79	84.04	14	14.89	1	1.06	0	0.00	0	0.00
8	94	71	75.53	16	17.02	5	5.32	2	2.13	0	0.00
9	94	78	82.98	10	10.64	0	0.00	6	6.38	0	0.00
10	94	57	60.64	30	31.91	5	5.32	1	1.06	1	1.06
11	94	78	82.98	11	11.70	4	4.26	1	1.06	0	0.00
12	94	73	77.66	18	19.15	1	1.06	2	2.13	0	0.00
Mean (%)		73.27		20.69		3.49		1.88		0.67	
Standard deviation (%)		9.40		8.69		2.04		1.24		0.73	

Comparing the four approaches - frequentist, Bayesian, sequential analysis and Bayesian sequential analysis - used to analyze the twelve lots, there is potential in their use, and the choice of the prior distribution is an important step. In the frequentist and Bayesian approaches, there is agreement between the results because the prior distribution elicited in articles is consistent with the sample of lots used. However, when the Bayesian method is added to the sequential analysis, a decrease in the sample size for decision-making about the lot is observed, resulting in a reduction in the sampling time and cost; however, the prior distribution has a strong influence that is not compatible with the frequentist results. Thus, the importance of the choice of the prior is reinforced. The fact that it is possible to incorporate information from producers or even scientific articles with regional data may reduce the sampling time and cost; however, such choice must have a consistent basis.

Furthermore, it can be observed that Bayesian sequential analysis, when used, has been achieving good results, Brighenti et al. (2019) used this approach applied to binomial distribution to estimate the proportion of viable coffee seeds. Lima et al. (2024b), used the multinomial model to estimate the allelic and genotypic proportions. Lima et al. (2024a) also used multinomial distribution, estimating different types of damage to corn seeds from the X-ray test. All of these studies achieved a reduction in sample size and consequently a reduction in costs. Thus, it is noted that Bayesian sequential analysis can be used in various areas of science to estimate parameters with the aim of reducing costs.

Conclusion

In the decision-making during the tetrazolium test, for the modeling of vigor, the use of a binomial distribution is adequate, and in the case of the analysis of damage types, in which there are several categories, the use of a multinomial distribution is promising. It can be concluded that in the Bayesian sequential process, both in the estimation of vigor (use of the binomial distribution) and damage in seeds (use of the multinomial distribution), there was a significant reduction in the required sample size. This is an advantage, as it reduces time and cost in tetrazolium testing. The estimates of interest obtained by the sequential Bayesian method were consistent when compared with the traditional method.

Acknowledgements

The authors thank the Seed Laboratory of the Department of Agriculture of the Federal University of Lavras for the data provided and the Brazilian Federal Agency for Support and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) for financial support for this study.

Data availability

We inform you that the data used in the research were made publicly available and can be accessed via the link https://ufsj.edu.br/extensao_brighenti/conjunto_dados_estatisticos.php

References

Abati, J., Zucareli, C., Brzezinski, C. R., Krzyzanowski, F. C., França-Neto, J. B., & Henning, F. A. (2021). Metabolites of the phenylpropanoid pathway and physiological quality of soybean seeds in storage. *Journal of Seed Science*, *43*, 1-14. https://doi.org/10.1590/2317-1545v43253585

Avci, E. (2021). Using systematic review and meta-analysis in order to obtain different empirical-informative prior for bayesian binomial proportion. *Sigma Journal of Engineering and Natural Sciences*, *39*(2), 195-202. https://doi.org/10.14744/sigma.2021.00009

Batista, E. C., Villela, G. M., Pires, R. M. O., Santos, H. O., Carvalho, E. R., & Bruzi, A. T. (2022). Physiological quality of soybean seeds and the influence of maturity group. *Journal of Seed Science*, *44*, 1-11. https://doi.org/10.1590/2317-1545v44261325

Boukabour, S., & Masmoudi, A. (2021). On the consistency of Bayes estimates for the infinite continuous mixture of Dirichlet distributions. *Hacettepe Journal of Mathematics and Statistics*, *50*(5), 1534-1547. https://doi.org/10.15672/hujms.774732

Brasil. (2009). *Regras para análise de sementes.* SNDA/MAPA.

Brighenti, C. R. G., Resende, M., & Brighenti, D. M. (2011). Estimação sequencial Bayesiana aplicada à proporção de infestação de psilídeos em alecrim do campo. *Revista Brasileira de Biometria, 29*(2), 342-354.

Brighenti, C. R. G., Cirillo, M. Â., Costa, A. L. A., Rosa, S. D. V. F., & Guimarães, R. M. (2019). Bayesian sequential procedure to estimate the viability of seeds *Coffea arabica* L. in tetrazolium test. *Scientia Agricola*, *76*(3), 198-207. https://doi.org/10.1590/1678-992X-2017-0123

Costa, N. P., Mesquita, C. M., Maurina, A. C., França Neto, J. B., Krzyzanowski, F. C., Oliveira, M. C. N., & Henning, A. A. (2005). Perfil dos aspectos físicos, fisiológicos e químicos de sementes de soja produzidas em seis regiões do Brasil. *Revista Brasileira de Sementes*, *27*(2), 172-181. https://doi.org/10.1590/S0101-31222005000200025

França-Neto, J. B., & Krzyzanowski, F. C. (2022). Use of the tetrazolium test for estimating the physiological quality of seeds. *Seed Science and Technology*, *50*(2), 31-44. https://doi.org/10.15258/sst.2022.50.1.s.03

Hamura, Y. (2023). Bayesian point estimation and predictive density estimation for the binomial distribution with a restricted probability parameter. *Communications in Statistics-Theory and Methods*, *52*(11), 3767-3794. https://doi.org/10.1080/03610926.2021.1980046

Hwang, L.-C. (2020). Second-order approximations of a robust sequential procedure in Bayes sequential estimation. *Sequential Analysis*, *39*(4), 467-483. https://doi.org/10.1080/07474946.2020.1826785

Jesus, M. A., Reis, V. M. A., Sampaio, F. R., Posse, F. L., & Barbosa, R. M. (2021). Quality control charts in the processing of soybean seeds. *Journal of Seed Science*, *43*, 1-11. https://doi.org/10.1590/2317-1545v43245104

Inacio, M., Izbicki, R., Lopes, D., Diniz, M. A., Salasar, L. E., & Ferreira, J. C. P. (2022). What if the forecaster knew? Assessing forecast reliability via simulation. *Trends in Computational and Applied Mathematics*, *23*(1), 175-192. https://doi.org/10.5540/tcam.2022.023.01.00175

Jones, P. W. (1976). Bayes sequential estimation of multinomial parameters. *Mathematische Operationsforschung und Statistik*, *7*, 123-127. https://doi.org/10.1080/02331887608801283

Lima, I. S., Brighenti, C. R. G., & Pires, R. M. O. (2024a). Bayesian sequential estimation of the proportion of damage in maize seeds. *Scientia Agricola*, *81*, 1-11. https://doi.org/10.1590/1678-992X-2023-0094

Lima, I. S., Brighenti, C. R. G., & Yazbeck, G. M. (2024b). Sequential Bayesian approach for genetic diversity analysis of the piracanjuba fish (*Brycon orbignyanus*). *Brazilian Journal of Biometrics*, *42*(2), 171-181. https://doi.org/10.28951/bjb.v42i2.695

Lopes, M. M., Prado, M. O. D., Sader, R., & Barbosa, R. M. (2011). Efeitos dos danos mecânicos e fisiológicos na colheita e beneficiamento de sementes de soja. *Bioscience Journal*, *27*(2), 230-238.

Lopes, K. B. A., Pípolo, A. E., Bazzo, J. H. B., & Zucareli, C. (2020). Intracultivar selection for seed quality of soybeans in an ultra-low-density selection model (Honeycomb Selection Designs). *Acta Scientiarum. Agronomy*, *42*(1), 1-10. https://doi.org/10.4025/actasciagron.v42i1.44299

Lorini, I. (2018). *Qualidade de sementes e grãos comerciais de soja no Brasil - safra 2016/17*. Embrapa Soja.

Marcos-Filho, J. (2015). Seed vigor testing: an overview of the past, present and future perspective. *Scientia Agricola*, *72*(4), 363-374. https://doi.org/10.1590/0103-9016-2015-0007

Martins, C. C., Unêda-Trevisoli, S. H., Môro, G. V., & Vieira, R. D. (2016). Metodologia para seleção de linhagens de soja visando germinação, vigor e emergência em campo. *Revista Ciência Agronômica*, *47*(3), 455-461. https://doi.org/10.5935/1806-6690.20160055

Mulgan, N. (2021). Binomial early stopping times. *Sequential Analysis*, *40*(4), 501-517. https://doi.org/10.1080/07474946.2021.2010409

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

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ribeiro, G. F. R., Carmo-Filho, A. S., Alves, R. M., França-Neto, J. B., Marcos-Filho, J., & Gomes-Junior, F. G. (2024). The Vigor-S system as an auxiliary component of the accelerated aging test procedure in soybean seeds. *Journal of Seed Science*, *46*, 1-10. https://doi.org/10.1590/2317-1545v46280795

Santos, R. F., Placido, H. F., Bosche, L. L., Zeni Neto, H., Henning, F. A., & Braccini, A. L. (2021). Accelerated aging methodologies for evaluating physiological potential of treated soybean seeds. *Journal of Seed Science*, *43*, 1-10. https://doi.org/10.1590/2317-1545v43250605

Schnuerch, M., & Erdfelder, E. (2020). Controlling decision errors with minimal costs: The sequential probability ratio t test. *Psychological Methods*, *25*(2), 206-226. https://doi.org/10.1037/met0000234

Schuab, S. R. P., Braccini, A. L., Scapim, C. A., França-Neto, J. B., Meschede, D. K., & Ávila, M. R. (2007). Germination test under water stress to evaluate soybean seed vigour. *Seed Science and Technology*, *35*(1), 187-199. https://doi.org/10.15258/sst.2007.35.1.17

Souza, M., Horikoshi, W. C., Saito, P. T., & Bugatti, P. H. (2024). Soybean seed vigor classification through an effective image learning-based approach. *Multimedia Tools and Applications*, *83*, 13113-13136. https://doi.org/10.1007/s11042-023-15804-0

Stefan, A. M., Evans, N. J., & Wagenmakers, E. J. (2022). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*, *27*(2), 177-197. https://doi.org/10.1037/met0000354

Toloi, M. N. V., Bonilla, S. H., Toloi, R. C., Silva, H. R. O., & Nääs, I. A. (2021). Development indicators and soybean production in Brazil. *Agriculture*, *11*(11), 1-15. https://doi.org/10.3390/agriculture11111164

Tan, M. H. (2020). Bayesian optimization of expected quadratic loss for multiresponse computer experiments with internal noise. *SIAM/ASA Journal on Uncertainty Quantification*, *8*(3), 891-925. https://doi.org/10.1137/19M1272676

Yi, D. H., Kim, D. W., & Park, C. S. (2020). Prior selection method using likelihood confidence region and Dirichlet process Gaussian mixture model for Bayesian inference of building energy models. *Energy and Buildings*, *224*, 1-17. https://doi.org/10.1016/j.enbuild.2020.110293