Journal homepage: https://itce.vn.ua/en

Information Technologies and Computer Engineering **Vol. 23, No. 2. 2026**

Article's History: Received: 10.02.2026 Revised: 17.05.2026
Accepted: 25.06.2026 Published: 03.08.2026

ISSN 1999-9941; e-ISSN 2078-6387
UDC 338.45:519.233
DOI: 10.31649/vitce/2.2026.55

A sequential decision framework for Bayesian A/B Tests: Reconciling interval precision, practical significance, and resource constraints

Artur Markov*
Postgraduate Student
Taras Shevchenko National University of Kyiv
01033, 60 Volodymyrska Str., Kyiv, Ukraine
https://orcid.org/0009-0000-5222-4397

Abstract. The study aimed to justify an approach to the planning and completion of Bayesian A/B tests, within which interval certainty criteria are combined with threshold requirements for a practically significant effect and an assessment of the economic feasibility of further data collection. The methodology was based on methods of analytical synthesis, structural-logical classification, logical-algorithmic formalisation, and illustrative case analysis. The study established that in Bayesian A/B tests, "interval precision" has two dimensions: informativeness (the narrowness and stability of the credible interval for supporting an engineering decision) and calibratedness (the proximity of actual coverage to the nominal level, which may deviate for small samples). The study demonstrated that uncertainty decreases with increasing n but with diminishing returns; therefore, further narrowing of intervals requires disproportionately greater expenditure of time and computational or experimental resources and is justified only when such costs are warranted, with the required n depending on the metric and the scale of the effect. The volume of data can be reduced by incorporating prior information into the prior distribution or by defining "sufficiency" based on criteria of evidence or the probability of achieving the objective, whereas high variability of engineering metrics, rare events, low baseline event rates, strict accuracy requirements, and a small effect increase the required n and the duration of the test. To reduce the required volume of data, a sequential framework for concluding a Bayesian A/B test has been proposed, combining an information component, a target component, and an economic feasibility check: "update estimates → check threshold → check interval → weigh up the benefits and costs of continuing". The practical applicability of this approach was confirmed by engineering studies, where Bayesian methods reduced resource consumption by 66.6%, 36.7%, and 26.9% in stream computing and achieved up to 25% energy savings, 15% greater energy efficiency, and about 10 ms tighter latency guarantees in edge computing. The practical significance of the results lies in their applicability to researchers and analysts involved in engineering experiments when making decisions regarding the termination of the experiment and the implementation of the proposed changes

Keywords: solution; experiment; data; sufficiency; certainty

Introduction

The complexity of engineering systems in information technologies and computer engineering is steadily increasing, and even minor changes may affect performance, reliability, and resource efficiency. Under such conditions, comparative experiments are increasingly used to support informed decision-making regarding system configurations, algorithms, software implementations, network protocols, and machine-learning models. However, every experiment requires time, computational resources, and engineering effort: a test that is too short may yield unreliable estimates and increase the risk of an incorrect technical decision, whereas a test that is too long may lead to unnecessary resource consumption and delays in the deployment of improved solutions. This is particularly important for systems characterised by variable workloads, noisy data, rare failures, and expensive testing cycles. In this context, Bayesian comparative experiments are appropriate because they provide direct probabilistic statements

Suggested Citation:
Markov, A. (2026). A sequential decision framework for Bayesian A/B Tests: Reconciling interval precision, practical significance, and resource constraints. *Information Technologies and Computer Engineering*, 23(2), 55-66. doi: 10.31649/vitce/2.2026.55.

*Corresponding author (*ArturMarkov0_1@outlook.com*)

about the effect and naturally support sequential data accumulation. At the same time, it is essential to formalise how the precision of uncertainty intervals (credible intervals) changes as the sample size (n) increases, and how these changes can be reconciled with the practical feasibility of the experiment.

The state of research reveals the existence of distinct methodological approaches to determining sample size and stopping criteria in Bayesian tests. In the paper by Q. Fu *et al.* (2022), an approach is proposed for determining the sample size for Bayesian analysis of variance under informative hypotheses, where "sufficiency of data" is defined by the requirement to exceed the threshold of evidence (Bayes' factor) with high probability. This formulation can be used to plan *n* not only concerning the width of the credibility interval, but also based on the criterion of achieving sufficient evidence for the effect. In a study by F. Quin *et al.* (2024), which synthesised a systematic review of A/B testing research, it was shown that classical schemes predominate in practice, whilst standardised approaches to interpreting uncertainty and stopping rules remain underdeveloped. This highlights the gap between the widespread use of A/B experiments and the need for formalised recommendations regarding the choice of *n* and the requirements for the accuracy of interval estimates.

As shown by V. Sambucini (2024), a hybrid classical-Bayesian procedure for sample size determination in two-arm superiority trials can formally incorporate prior information and uncertainty about unknown parameters while retaining a frequentist analysis framework. For binary outcomes, the author derived hybrid sample-size criteria using three alternative effect measures: the difference between success proportions, the logarithm of the relative risk, and the logarithm of the odds ratio. M. Moerbeek (2021) argued that *n* can be increased iteratively through Bayesian updating, and the decision to stop can be based on a pre-specified criterion of sufficiency of evidence. This provides a theoretical basis for stopping rules that combine requirements for the precision of credible intervals and/or a threshold posterior probability in Bayesian A/B tests. T.K. Wong & J.N. Tendeiro (2025) proposed a universal scheme for selecting *n* for the Bayesian criterion to obtain a Bayesian odds ratio above the threshold with a given probability, that is, to achieve a predefined level of evidence. This makes it possible to plan *n* as an alternative or complement to approaches that optimise *n* based on the width of credible intervals.

T.S. Richardson *et al.* (2022) found that, due to heterogeneity in the timing of entry into an observed process, it is incorrect to assume that the arrival of new observations over consecutive periods is independent and identically distributed; they therefore proposed a Bayesian approach to forecasting the number of additional observations in the next period based on early data. Such a forecast provides a practical basis for planning the experiment horizon and the expected *n*, on which the attainability of the specified precision of credible intervals in Bayesian A/B tests directly depends. Based on computer simulations, I. Douven (2023) compared various Bayesian stopping rules, including the widely used Bayes-factor stopping rule, and showed that there is no single universally best rule; rather, the appropriateness of the stopping criterion depends on the conditions of the problem. This provides a methodological justification for selecting a stopping policy tailored to specific objectives, such as interval precision and/or a posterior probability threshold, which is directly relevant for aligning *n* with the stopping criteria of Bayesian comparative experiments.

In Ukrainian scientific and applied discourse, A/B testing and corresponding stopping criteria are increasingly used to evaluate the effectiveness of digital solutions and to support data-driven decisions in applied IT contexts. The choice of a stopping rule determines the point at which the experiment ends, the expected sample size n, and consequently influences the attainable interval precision of the effect estimate. As an example of the practical application of A/B testing in Ukrainian applied research, V.O. Tkachuk & A.V. Shestakova (2022) demonstrated the selection of a more effective version of a digital solution based on real-world data. The use of experiments in environments constrained by time and the availability of observations reinforces the need for approaches that reconcile interval-certainty requirements with the necessary data volume. Despite progress in research, there is still a lack of a coherent approach that simultaneously links the narrowing of credible intervals with an increase in n, the attainment of a threshold posterior probability (PP), and the practical feasibility of continuing the A/B test. Therefore, the study aimed to develop a methodological framework for determining data sufficiency in Bayesian A/B tests, which reconciles interval certainty requirements with system performance criteria and resource constraints.

Materials and Methods

To conduct the study, methods of analytical synthesis, structural-logical classification, logical-algorithmic formalisation and illustrative case analysis were combined to formalise the criteria for interval certainty, the planning of *n* and the stopping rules in Bayesian A/B tests. Using the method of analytical synthesis, the approach to evaluating the "precision of intervals" in Bayesian A/B tests was systematised and methodically formalised concerning confidence intervals, sequential monitoring, stopping rules and sample planning, after which the concept of "precision" was operationalised in two dimensions – informativeness (narrowness and stability of the interval) and calibratedness (frequency coverage) (Kamalbasha & Eugster, 2021; De Santis & Gubbiotti, 2021; de Paula Castro *et al.*, 2022). This was done to reasonably determine data sufficiency and the stopping point of the experiment.

Using a structural-logical classification method, the factors that increase the required sample size in Bayesian A/B tests were systematised: high variability of engineering metrics (e.g., latency, throughput, error rate, precision,

recall, accuracy, energy consumption, failure probability, or Mean Time Between Failures (MTBF)), rare events and low baseline event rates, stringent accuracy requirements (a narrow target interval), and an expected small effect size (Pan & Banerjee, 2023; Shen *et al.*, 2023; Ali, 2025). The selection of these specific factors is due to the fact that they represent the key "root causes" of the increase in the required *n* and cover the fundamental mechanisms underlying the deterioration of the interval certainty of the effect estimate (increased noise, event scarcity, weak signal, and heightened certainty requirements). The classification was conducted based on following criteria: factor, typical feature in the data/metric, mechanism of influence on interval certainty (width and stability of the credibility interval), and consequence for the required n/duration of the experiment. This was done to systematise the determinants of the increase in the required *n* in Bayesian A/B tests and to link them to the mechanisms of reduced interval certainty and the implications for the duration of the experiment.

The structural-logical classification method was also used to compare approaches to reducing the required data volume for achieving the target interval of confidence in Bayesian A/B tests – the use of historical data via a well-founded prior distribution and the shift from a "narrow interval" objective to a "probability of exceeding the predefined threshold" objective, based on the following criteria: approach, applicability assumptions, mechanism for reducing the required n, implications for *n* planning and test stopping rules (Zheng *et al.*, 2023). The choice of these specific approaches is determined by the fact that they represent two fundamental levers for reducing the required n: increasing information content in the early stages through prior information and shifting the formulation of the data sufficiency criterion. This was used to compare the basic approaches to reducing the required *n* in Bayesian A/B tests.

Using the method of logical-algorithmic formalisation, the procedure for sequential decision-making in a Bayesian A/B test was modelled as a repetitive cycle: "arrival of new data → updating of posterior estimates (posterior distribution of the effect, credible interval, probability that effect ⩾ threshold) → assessment of practical significance → assessment of interval certainty → assessment of the feasibility of continuation", with prior specification of sufficiency criteria (PAI/zero threshold, minimum posterior confidence, requirement for the width of the credible interval, and cost of continuation) (Pananos, 2024; Kelter & Pawel, 2025). This was done to standardise the rules for continuous monitoring and experiment termination based on agreed criteria of practical significance, interval certainty, and resource constraints.

Using an illustrative case-study approach, examples of the application of Bayesian methods to sequential decision-making and optimisation in engineering experiments were selected, including Bayesian-driven autoscaling in stream computing (Zhang *et al.*, 2024), adaptive resource scheduling in edge computing (Sahu *et al.*, 2025), and parameter tuning for IEEE 802.11 VANET safety messaging in vehicular communication systems (Wright *et al.*, 2026). These examples were selected because they represent complementary aspects of the practical applicability of the proposed framework, namely the achievement of target Quality of Service (QoS) indicators with minimal resource expenditure, the balance between energy consumption and latency under adaptive resource allocation, and decisions on whether further testing is justified in systems with strict requirements for latency, reliability, and computational cost. This was done to provide an illustrative justification for the need to consider both the decision-oriented logic of experiment termination and the practical feasibility of Bayesian procedures in engineering applications. On this basis, recommendations were formulated for the preparation, conduct, and completion of Bayesian comparative experiments in information technologies and computer engineering.

Results and Discussion

Interval precision and sample planning in Bayesian A/B tests. In the context of Bayesian A/B tests, it is necessary to distinguish between two related but distinct aspects of "interval precision": the informativeness of the interval (narrowness and robustness) and its calibration (frequency coverage). In an applied engineering context, "precision" refers to how narrow the Bayesian credible interval is for estimating the effect and how consistently it changes as more data are added during sequential monitoring. A narrower and more stable interval implies less uncertainty regarding the magnitude of the effect, which facilitates technical decision-making, such as selecting between alternative configurations or assessing whether the observed improvement satisfies the required performance threshold. In this sense, intervals serve as a decision-support tool in comparative engineering experiments and are therefore directly relevant to experiment duration, stopping rules, and practical implementation under resource constraints (Kamalbasha & Eugster, 2021; Pananos, 2024). In a stricter statistical sense, "precision" is associated with the extent to which the interval corresponds to the stated level when the experiment is repeated multiple times, i.e., how often the true value of the parameter falls, for example, within the 95% interval in a series of repetitions. Although Bayesian intervals have a probabilistic interpretation under a correctly specified model and prior distribution, their frequency coverage is not guaranteed to coincide with the nominal level, particularly for approximate intervals and/or small samples. In terms of proportions, the study demonstrated that requirements for *n* can be formulated to ensure calibrability; for small values of *n*, deviations may be significant, whilst coverage improves as *n* increases (De Santis & Gubbiotti, 2021).

For most metrics in Bayesian A/B tests, increasing the data volume reduces uncertainty with diminishing returns: initially, the credible interval narrows rapidly, and then more slowly. As a rule, halving the interval width typically requires substantially more data than simply doubling the sample size. For engineering experiments, this means that a larger sample size almost always leads to a longer experiment duration, greater use of computational, testing,

or laboratory resources, and higher costs associated with delaying the deployment of an improved configuration, algorithm, or system setting. Therefore, extending the test is justified only when the additional reduction in uncertainty is warranted in terms of technical relevance and resource expenditure (Kamalbasha & Eugster, 2021). The precision–n relationship can be optimised by incorporating prior information: historical data or results from previous experiments may be formally included in the prior distribution, thereby reducing the required *n* needed to achieve a given level of accuracy or a given probability of the target effect (Zheng *et al*., 2023).

In addition to approaches based on interval width, the Bayesian literature determines sample size using evidence criteria (in particular, Bayesian factors) and approaches to assessing the probability of achieving the objective (assurance). This changes the way the question "how much data is needed" is framed, as well as the method of formalising "sufficient accuracy" for a decision. In applied problems, such approaches are implemented as methodologies and software tools for calculating sample size and assurance (Pan & Banerjee, 2023; Kelter & Pawel, 2025). In general, an A/B test in a Bayesian framework is viewed as a decision-making problem under uncertainty. Experimental data reduce uncertainty regarding the magnitude of the effect, and the Bayesian posterior/credible interval serves as a compact representation of this uncertainty. Since the narrowing of the Bayesian posterior/credible interval depends on the properties of the metric, the baseline event rates, and the expected effect size, the same accuracy requirements may necessitate different amounts of data across different tasks. Table 1 summarises the factors under which a larger *n* or a longer experiment duration is required to achieve a given interval certainty.

Table 1. Factors affecting the required sample size to achieve a specified interval certainty in Bayesian A/B tests

Factor	Typical attribute in the data/ metrics	Mechanism influencing interval certainty	Result for the required *n* of duration
High variability of engineering metrics	High variance, skewness, heavy tails, or outliers in latency, throughput, energy consumption, or related engineering metrics	Uncertainty surrounding the average effect decreases slowly; the interval may remain unstable because of large fluctuations or isolated extreme observations	A larger sample size or a longer experiment is required to achieve stable interval estimates
Rare events and low baseline event rates	A small number of observed failures, errors, or other low-frequency events in the sample	Information about the parameter accumulates slowly, and the interval estimates remain wide for longer	A larger *n* is required, especially for rare failures, low error rates, or reliability-related metrics
Stringent accuracy requirements	A narrowly specified target interval width or a strict tolerance for uncertainty	A stricter interval-width requirement increases the amount of data needed to reach the desired level of certainty	The required *n* increases substantially as the target interval becomes narrower
Expected small effect size	The difference between the compared configurations is close to zero or to the predefined engineering threshold	The signal is weak relative to the noise, and the interval continues to overlap zero or the target threshold	A larger *n* is required to obtain sufficient certainty regarding the effect

Source: compiled by the author based on F. De Santis & S. Gubbiotti (2021), S. Chennu *et al*. (2023), Y. Shen *et al*. (2023), J. Pan & S. Banerjee (2023), I. Ali (2025)

These factors share a common characteristic: they reduce the amount of information provided by a single observation regarding the magnitude of the effect, or they lower the signal-to-noise ratio. High variability of engineering metrics and rare events slow down the convergence of the posterior distribution, meaning that the credibility interval remains wide for longer. High precision requirements and an expectedly small effect, in turn, raise the "bar" for certainty – namely, even with a correct model, a larger volume of data is needed for the interval to stop overlapping zero or the predefined engineering threshold. The practical implication is that sample size *n* should be determined not only by the target interval width, but also by considering the properties of the metric and the expected magnitude of the effect; otherwise, there is a systematic risk of underestimating the duration of the experiment and the cost of achieving the desired decision certainty. Thus, the problems of high variability, rare events and a small, expected effect not only slow down the narrowing of the credible interval but also create a risk of systematically under-planning experiments. Therefore, alongside assessing the "complexity of the metric", it is advisable to immediately consider approaches that accelerate the convergence of the posterior distribution or shift the formulation of the problem from "maximum estimation accuracy" to "sufficient certainty for an engineering decision". Such approaches are summarised in Table 2.

Table 2. Approaches to reducing the required data volume to achieve the target interval certainty in Bayesian A/B tests

Approach	Conditions for applicability	Mechanism for reducing the required sample size	Consequences for planning *n* and stopping the test
Use of historical data via a well-founded prior distribution	Historical data are consistent with the current operating conditions, workload characteristics, and system configuration	Provides information at an early stage and accelerates the concentration of the effect	Narrower intervals for small *n*, the possibility of achieving the target certainty sooner

Table 2. Continued

Approach	Conditions for applicability	Mechanism for reducing the required sample size	Consequences for planning *n* and stopping the test
Shift from a "narrow interval" target to a probability of exceeding the predefined threshold" target	A predefined threshold for the minimum practically significant effect is specified	The decision is based on the probability of meeting the criterion rather than on maximising the precision of the effect size	The experiment may be completed earlier while maintaining relevance to the engineering decision

Source: compiled by the author based on S. Kamalbasha & M.J.A. Eugster (2021), H. Zheng *et al*. (2023), J. Pan & S. Banerjee (2023)

Table 2 summarises two different approaches to reducing the required data volume in Bayesian A/B testing. The information-based approach involves incorporating relevant historical data via a well-founded prior distribution, which accelerates the convergence of the posterior distribution of the effect and makes it possible to achieve the target interval certainty with a smaller n. The target-based approach involves shifting from the requirement of a "narrow interval" to the criterion of a "sufficient probability of exceeding the minimum significance threshold", which may allow earlier termination of the experiment without compromising relevance for the engineering decision. Therefore, optimisation of *n* in the Bayesian framework can be achieved both by increasing the information content through the use of prior information and by correctly formulating the decision criterion through a threshold-based specification.

An informative prior distribution can accelerate the convergence of Bayesian posterior/credibility intervals; however, the validity of this effect depends on a key condition: the prior assumptions must be consistent with the current data and experimental conditions. If the prior distribution is incompatible with the current operating conditions, for example because of differences in workload characteristics, system configuration, or the hardware and software environment, there is a risk of false certainty: the interval may appear narrow, but the posterior estimate of the effect may be systematically biased, increasing the probability of an erroneous engineering decision. Within the theoretical framework, it is worth emphasising two principles: an informative prior distribution should be viewed as a tool for accelerating inferences, which requires substantive justification and verification of compatibility with current data; the correctness of interval inferences is determined by the adequacy of the statistical model and assumptions; in particular, the properties of the intervals and their reliability in repeated scenarios may depend on the application regime and the chosen approximations (De Santis & Gubbiotti, 2021; Zheng *et al.*, 2023).

The results of the study showed that, in Bayesian A/B tests, the concept of "interval precision" should be interpreted not as a single characteristic, but as two distinct dimensions: interval informativeness (the narrowness and stability of the posterior/credible interval as data accumulates) and calibratedness (frequency coverage, i.e. how often the true value falls within the interval in repeated runs of the experiment). This formulation is consistent with T. de Paula Castro *et al.* (2022), who examined the comparison of credible and confidence intervals within a common (consistent) methodological framework for intervals using the example of binomial proportions. The authors found that a correct comparison is not possible using a single criterion, and that coverage and interval length provide different but complementary "measures of quality" for interval estimates. For comparative experiments in engineering, "precision" understood as the narrowness and stability of the interval should not be equated with "precision" understood as calibratedness, because conflating these two properties may lead to misleading conclusions about the reliability of the results, particularly when *n* is small or when aggressive stopping rules are applied.

R. Johari *et al.* (2022) found that under conditions of continuous monitoring of A/B tests, where the researcher repeatedly reviews interim results and may halt the experiment at a point determined by the observed data, standard frequency-based conclusions and confidence intervals may lose their validity because stopping is based on interim analyses. In response, the authors formalised an approach to always-valid inference, specifically always-valid p-values and always-valid confidence intervals, which remain valid under any data-dependent stopping rule and provide inferential support even when decisions are made during the course of the experiment. These results align with the present study insofar as the "precision" of intervals is also viewed as the stability of the interval as data accumulate during sequential testing. Accordingly, the interval in comparative engineering experiments is interpreted as a decision-support tool not only at the end of the experiment, but throughout the process of data accumulation. Furthermore, this correlates with A.M. Stefan *et al.* (2022), who found that, in the context of sequential testing, the effectiveness and behaviour of procedures over time depend on the test formulation and the choice of thresholds; that is, the stopping rule is not only an operational but also a methodologically defining element of the experimental design. It is therefore not enough simply to state that "the interval has narrowed"; it is also necessary to verify whether the monitoring and stopping method itself introduces systematic errors into the conclusions, and whether the chosen inference procedure is compatible with real-time engineering decision-making.

The results of the study showed that the required sample size for achieving a given interval confidence level in Bayesian A/B tests depends on the data structure and model (in particular, in the presence of operating conditions, dependency structure of observations, and variability); therefore, universal rules for planning *n* may systematically underestimate the duration of the experiment. This is confirmed by S. Vasishth *et al.* (2023), demonstrating that

for Bayesian hierarchical models, it is advisable to determine the sample size through simulation analysis of the design: the authors select n such that, in repeated simulations under given prior assumptions and a chosen criterion (e.g., evidence), the required certainty of the conclusion is achieved. Consequently, planning n in Bayesian A/B tests should be conducted for a specific model and data structure, and for segmented and hierarchical setups – based on simulation design analysis – to avoid systematic under-planning of the experiment duration.

T. Rainforth *et al.* (2024) systematised the Bayesian experimental design approach, within which experimental planning is viewed as the problem of selecting a design that maximises expected utility or information gain, particularly in adaptive and sequential scenarios. These findings are consistent with the results of the current study, which shows that as n increases, uncertainty decreases with diminishing returns: initially, the interval narrows rapidly, then more slowly, and achieving further narrowing requires a disproportionately larger volume of data, which increases the duration of the experiment and data consumption. Accordingly, continuing the test is justified only when the additional reduction in uncertainty economically outweighs the cost of the decision delay and resource expenditure. Thus, the "precision of intervals" in Bayesian A/B tests should be assessed simultaneously in terms of informativeness (width/stability) and calibration (frequency coverage), as these measures depend differently on the volume of data and the properties of the metric. Consequently, sample planning and stopping rules should be formulated as a decision-making problem under uncertainty, considering the economic cost of additional data and the possibility of reducing it through prior information or threshold criteria.

Sequential decision-making in Bayesian A/B tests. For the practical application of Bayesian A/B testing, it is advisable to establish not only criteria for data "sufficiency" but also a protocol for sequential decision-making as observations accumulate. Figure 1 illustrates the logic of such a protocol and formalises the experiment as a repeatable procedure for evaluating results as data becomes available. Its purpose is to reconcile the three components that determine the practical usefulness of the experiment: practical significance (exceeding the null threshold), interval certainty (sufficient narrowness of the credibility interval) and economic feasibility (the ratio of the value of additional information to the cost of continuing the experiment).

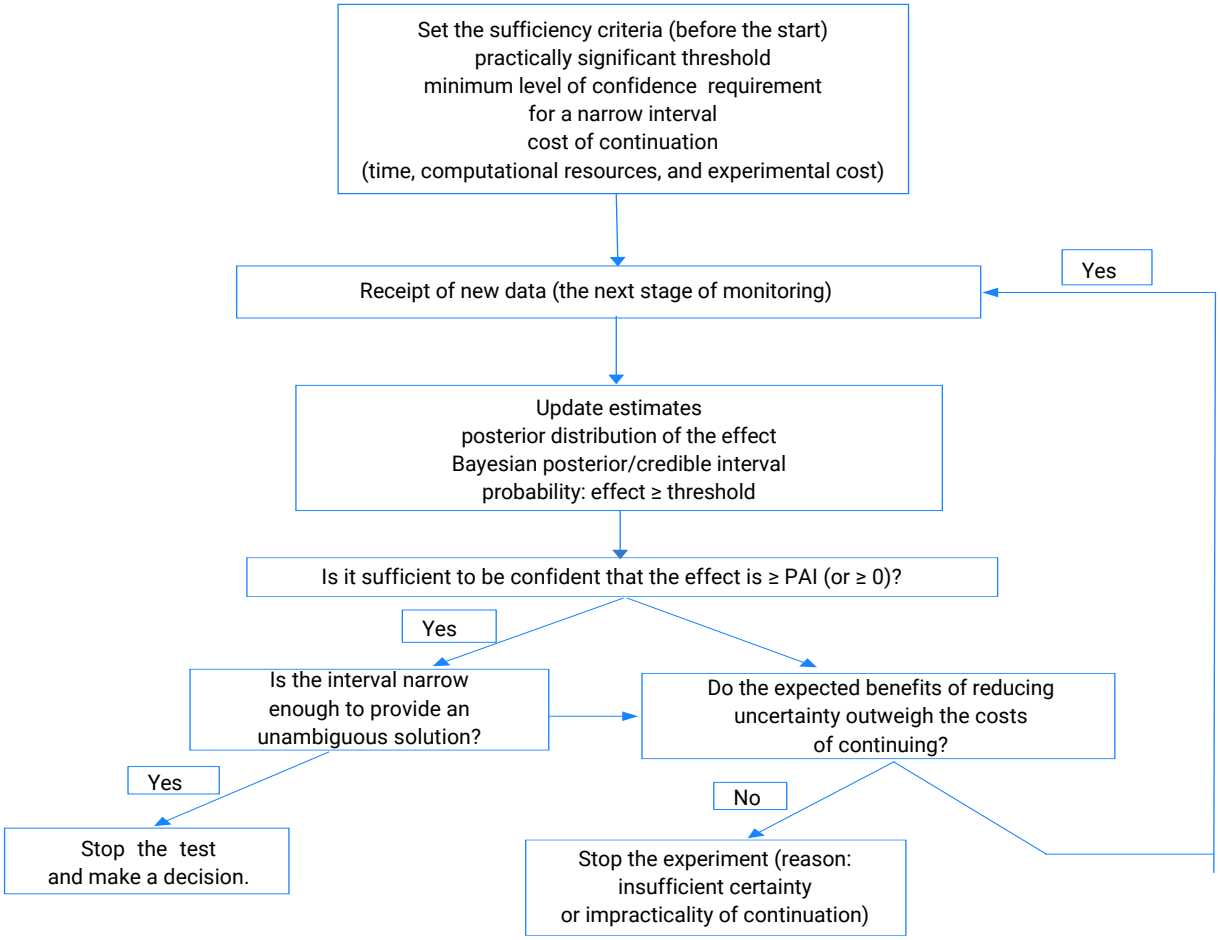

Figure 1. Flowchart for sequential decision-making and termination of a comparative engineering experiment
Source: compiled by the author based on S. Kamalbasha & M.J.A. Eugster (2021), F. De Santis & S. Gubbiotti (2021), H. Zheng *et al.* (2023), J. Pan & S. Banerjee (2023), D. Pananos (2024), R. Kelter & S. Pawel (2025)

Figure 1 formalises the policy of continuous monitoring of a Bayesian A/B experiment as a framework for engineering decision-making under resource constraints. At the pre-experimental stage (first block), the parameters defining the "sufficiency" of the result are established: PAI, the minimum level of posterior confidence for asserting that the threshold has been exceeded, the requirement for the width of the credible interval, and an assessment of the cost of continuation in terms of time, computational resources, testing effort, and losses associated with delayed implementation. The prior specification of these parameters ensures the consistency of the criteria throughout the experiment and reduces the risk of them being "tweaked" to fit interim results. The process is then implemented as a repetitive cycle. At each control step (second block), following the arrival of new data, the posterior estimates are updated (third block): a posterior distribution of the effect is formed, the credible interval is calculated as a compact measure of uncertainty, and the posterior probability of exceeding the threshold (zero or PAI) is estimated. This update distinguishes between two fundamentally different aspects of interpreting the results: whether the data support the claim of a practically significant effect, and whether the estimate of the effect is sufficiently precise for an unambiguous engineering interpretation.

The first decision node checks the condition of practical significance: is there a posteriori confidence that the effect is not less than the PAI (or not less than zero) at a given level. If the result is positive, the evaluation does not end automatically but proceeds to the second node – the precision condition: whether the credible interval is narrow enough to avoid substantial ambiguity regarding the magnitude of the effect and its engineering implications, for example in the assessment of latency, throughput, error rate, energy consumption, reliability, or resource planning. If the conditions of evidence and precision are met, the experiment concludes with a decision being made. If, however, at least one of the conditions (evidence or precision) is not met, the process moves to the third node – the assessment of the economic feasibility of continuing. Continuation is considered justified only when the expected benefit from reducing uncertainty, that is, from a potential refinement of the technical decision, exceeds the cost of additional data collection. If the conclusion is positive, the experiment continues, and the cycle repeats; if negative, the experiment is halted as one where further refinement is not rational within the limits of available resources. In general, the diagram illustrates that the point at which the experiment is halted is not the technical completion of data collection, but the point at which further observations cease to improve the quality in a statistically and economically sound manner. To operationalise the scheme shown in Figure 1, it is advisable to specify experiment-termination rules in the form of a decision matrix that combines the posterior probability that configuration B is superior to configuration A with the expected technical loss as criteria for stopping, continuing, or terminating the experiment (Table 3).

Table 3. Bayesian decision matrix for completing a comparative engineering experiment

Scenario	Condition	Interpretation	Solution
Positive result	$P(B > A) > 95\%$ and $eL(B) < 0.01\%$	There is a high posterior probability that configuration B outperforms configuration A, and the expected technical loss from selecting configuration B is below the acceptable threshold	End the experiment
A positive result with an increased risk	$P(B > A) > 95\%$ and $eL(B) > 0.01\%$	Configuration B is likely to be superior to configuration A; however, the expected technical loss exceeds the acceptable threshold, so the remaining uncertainty is still practically significant for the engineering decision	Continue the experiment or end it with a higher level of risk
The difference has not yet been determined	$P(B > A) < 95\%$ and $eL(B) > 0.01\%$	There is insufficient evidence to conclude that configuration B is superior; no practically significant difference has yet been established	Continue the experiment
Lack of difference	$P(B > A) \approx 50\%$ and $eL(B) < 0.01\%$	There appears to be no practically significant difference between the configurations, and for this metric the conclusion may be considered final	End the experiment. However, the proposed modification may still be implemented if it represents a logical step towards subsequent planned functional changes
There is likely to be no difference	$P(B > A) \approx 50\%$ and $0.01\% < eL(B) < 0.1\%$	The result suggests the absence of a practically significant difference, although additional data could still be collected to reduce uncertainty; however, the overall engineering conclusion is unlikely to change	Data collection may continue, but the overall engineering conclusion is likely to remain unchanged
Negative result	$P(B > A) < 5\%$ and $eL(A) < 0.01\%$	There are strong grounds for concluding that configuration A is superior, and the expected technical loss associated with retaining configuration A does not exceed the acceptable threshold. If P(A>B) or eL(A) is analysed, the result should be interpreted in the opposite direction	End the experiment

Table 3. Continued

Scenario	Condition	Interpretation	Solution
Strong negative signal	P(B > A) < 5% and eL(A) > 0.01%	There are strong indications that configuration B performs worse; however, the expected technical loss criterion has not yet reached the final stopping threshold	The experiment should normally be stopped; if additional confirmation of the negative effect is required, data collection may continue until eL(A) < 0.01% is reached

Source: compiled by the author based on C. Stucchio (2015), S. Kamalbasha & M.J.A. Eugster (2021), F. De Santis & S. Gubbiotti (2021), H. Zheng *et al*. (2023), J. Pan & S. Banerjee (2023), D. Pananos (2024), R. Kelter & S. Pawel (2025), A. Markov & O. Zaritskyi (2025)

In this study, the calculation of expected loss and relative expected loss is presented in accordance with the approach outlined by C. Stucchio (2015). A threshold value of 0.01% for the expected loss is used based on the results of an experimental study conducted by A. Markov & O. Zaritskyi (2025), which show that this level provides a practical trade-off between the risk of an incorrect decision and the duration of the experiment. Examples from engineering practice illustrate that experiment stopping criteria are formulated as a balance between the quality of the technical decision, the attainment of target performance indicators, and resource expenditure.In the field of stream computing, the Bayesian-driven solution AuTraScale+ for operator-level autoscaling on the Flink platform was proposed to achieve target values of throughput, processing-time latency, and event-time latency. Experimental results obtained on three representative workloads showed that the proposed approach reduced resource consumption by 66.6% in scale-down scenarios, by 36.7% in scale-up scenarios, and by 26.9% on average in event-time latency optimisation compared with state-of-the-art methods (Zhang *et al*., 2024). This example demonstrates that the practical value of the Bayesian approach lies not in achieving maximum accuracy at any cost, but in identifying, within a reasonable time, a configuration that ensures the required QoS level with minimal resource expenditure. In the field of edge computing, a Bayesian framework for adaptive resource scheduling was proposed for the joint optimisation of energy consumption and latency. The results showed that the proposed approach reduced energy consumption and average delay compared with baseline algorithms, while cumulative energy savings reached 25% relative to baseline methods; moreover, in trace-driven simulations, the framework demonstrated up to 15% greater energy efficiency and approximately 10 ms stricter latency guarantees compared with pure Bayesian optimisation or ad hoc schedulers (Sahu *et al*., 2025). Thus, in engineering experiments, the optimisation of data volume and test duration is determined not by the pursuit of maximum accuracy, but by the attainment of sufficient technical certainty at the minimum justified cost in terms of time, computational resources, and experimental infrastructure.

Another illustrative example is provided by vehicular communication systems, where Bayesian approaches are used to tune the parameters of IEEE 802.11 VANET safety messaging under strict QoS requirements. In such tasks, the practical value of sequential Bayesian decision-making lies in the fact that further testing is justified only when it leads to a real improvement in the balance between transmission latency, reliability, and the computational complexity of identifying the optimal configuration. This example clearly illustrates the engineering logic of experiment termination: the decision should be based not on an abstract pursuit of maximum accuracy, but on the attainment of sufficient technical certainty within acceptable time and computational costs (Wright *et al*., 2026).

When designing an engineering experiment, the primary task is to clearly define the decision to be made, such as selecting between alternative system configurations or validating a performance improvement. This requires specifying a minimum practically significant effect (δ), corresponding to a technical tolerance or performance requirement. Instead of asking whether any difference exists, the objective is to determine whether the observed effect exceeds this predefined threshold. In a Bayesian framework, this is naturally expressed as a required posterior probability that the effect is greater than zero or δ, ensuring a controlled risk of incorrect decisions. To ensure consistency and avoid arbitrary adjustments during the experiment, both the performance threshold and the acceptable level of uncertainty should be defined in advance. Stopping criteria must balance two components: (i) sufficient posterior evidence that the effect meets the required threshold, and (ii) sufficient interval precision to allow unambiguous interpretation of the effect magnitude. The relative importance of these criteria depends on the task: binary decisions (e.g., accept/reject a configuration) may rely primarily on threshold exceedance, whereas quantitative system assessment requires narrower intervals and, consequently, larger sample sizes.

In engineering contexts, variability is often driven by noise, operating conditions, or system instability. High variability slows the reduction of uncertainty and increases the required data volume. Therefore, results should be interpreted not only through point estimates but also through the range of plausible outcomes. Where appropriate, complex performance metrics can be decomposed into components to improve interpretability and identify sources of variation. Prior information (e.g., historical data or previous experiments) can reduce the required sample size by accelerating convergence, but only if it is consistent with current operating conditions. Otherwise, it may introduce bias and lead to false certainty, so conservative or partially informative priors are recommended when compatibility is uncertain. Common pitfalls include testing for negligible effects without predefined thresholds, equating

narrow intervals with correctness, and extending experiments without considering resource constraints. In engineering practice, sample size should be treated as a tool for managing uncertainty under limited resources (time, computation, or experimental cost). The experiment should be terminated when additional data no longer meaningfully improves the reliability of the decision relative to its cost.

The results of the study have shown that practically applicable Bayesian A/B testing requires not only a calculation of data "sufficiency", but also a formalised protocol for sequential decision-making as observations accumulate. The proposed logic of the protocol treats the experiment as a repetitive cycle of updating posterior estimates and testing conditions. This is consistent with the study by T. Zhou & Y. Ji (2024), which systematises sequential Bayesian designs as decision-making procedures as data arrives, which naturally corresponds to the interpretation of the experiment as a cycle of monitoring and stopping. Furthermore, the work by G. Mulier *et al.* (2024) demonstrates the practice of constructing sequential monitoring strategies and emphasises the need to evaluate the properties of such strategies in applied scenarios, i.e., treating monitoring as part of the design, rather than merely as "metric observation". Thus, the framework of the study, in which stopping criteria are specified in advance and tested repeatedly, is methodologically consistent with the contemporary interpretation of sequential Bayesian procedures of the 21st century.

J.G. Liao *et al*. (2021) examined the testing of interval null hypotheses and compared two approaches to formalising a "practically zero" effect: the Bayes factor and the modified Region of Practical Equivalence (ROPE) procedure. The authors emphasised that, in a range of tasks, it is appropriate to test not for a difference from the point null, but for the effect falling within an interval of practical insignificance or, conversely, for the presence of sufficient evidence against such an interval. These provisions are consistent with the results of the present study, which show that for applied engineering problems, the prior specification of the practical significance threshold and the rules for interpreting "small" effects are critical, since an engineering decision usually depends not on the mere fact of a "difference from zero", but on the practical significance of the effect. In line with the logic of Figure 1, the first decision node is associated with exceeding zero or a predefined engineering threshold; this threshold can be formalised as an interval of practical insignificance, analogous to ROPE, or as an evidence criterion based on the Bayes factor. Consequently, the threshold formulation (PAI / interval "zero") is not only a practical requirement of engineering decision-making but also a statistically sound method for avoiding decisions based on trivial and technically insignificant deviations

The study results indicate that the sample size n should be determined not with the aim of "achieving maximum precision", but rather to attain sufficient certainty for decision-making within resource constraints. Within this approach, n is viewed as a tool for managing uncertainty: too small n leaves intervals too wide for practical interpretation, whilst too large n increases the costs of decision delay. This is consistent with the review by K. Kunzmann *et al.* (2021), which systematised Bayesian approaches to deriving n for confirmatory studies and emphasises the relevance of the definition of "sufficiency" criteria (in particular through probabilistic target values such as the probability of success or achieving the objective). The study by S.F. Williamson *et al.* (2023) is also consistent with this line; using the example of determining n for a Bayesian estimation of diagnostic test characteristics, it emphasises that the choice of n involves a practical trade-off: it determines whether the estimate will be sufficiently precise for decision-making and whether data collection is excessive. Both sources support the conclusion of the present study that the volume of data should be planned based on criteria that reflect engineering decision requirements (threshold + acceptable uncertainty), rather than on the abstract concept of "the highest possible accuracy".

The paper by M. Mezzetti *et al.* (2023) examined Bayesian hierarchical models and procedures for eliciting prior information, with the authors emphasising that prior information is a critical component of the model and can influence the uncertainty of posterior estimates. These findings correlate with the results of the current study, which specifically highlights the role of prior assumptions as a tool for accelerating the attainment of interval certainty, provided they are substantively justified and compatible with the current context. Accordingly, an informative prior can increase certainty in the early stages of data collection, but at the same time raises the bar for procedural discipline: prior assumptions must be established in advance, justified and consistent with the experimental conditions, otherwise there is a risk of false certainty. Thus, practically effective Bayesian A/B testing requires a predefined protocol for sequential decision-making that reconciles practical significance (PAI or Bayesian factor), interval certainty and the economic feasibility of continuing. It is advisable to plan the sample size n as a tool for managing uncertainty under resource constraints (based on model and prior) and to terminate the experiment when additional information no longer justifies the expenditure of time and computational resources.

Conclusions

The results of the study showed that, before the start of the experiment, it is advisable to define the PAI, the minimum posterior confidence level for exceeding the threshold, the requirement for the width of the credible interval, and the cost of continuation in terms of time, computational resources, testing effort, or losses due to delayed implementation, which ensures the consistency of the criteria and reduces the risk of adjusting them to fit interim results. A sequential Bayesian A/B test should be implemented as a repetitive cycle of updating posterior estimates, in which the posterior distribution of the effect, the credible interval,

and the probability of exceeding the null or predefined engineering threshold are calculated at each step. The decision must separately address the evidence of practical significance and interval precision, that is, whether the interval is narrow enough for an unambiguous interpretation of the effect size. If these conditions are not met, continuing the test is justified only when additional data are expected to improve the engineering decision in a statistically and practically sound manner. Engineering metrics characterised by high variability or rare events require a larger margin in n, because interval uncertainty decreases more slowly under such conditions. It is also useful, where appropriate, to decompose complex engineering metrics into interpretable components in order to better explain the source of observed changes.

Recent engineering applications confirm the practical value of Bayesian approaches under performance and resource constraints. In stream computing, Bayesian-driven autoscaling reduced resource consumption by 66.6%, 36.7%, and 26.9% across QoS scenarios, while in edge computing Bayesian adaptive scheduling achieved up to 25% energy savings, 15% greater energy efficiency, and about 10 ms tighter latency guarantees. In vehicular communication systems, Bayesian optimization likewise supports experiment termination based on balancing transmission latency, reliability, and computational cost under strict QoS requirements rather than on pursuing maximum accuracy. Practically viable Bayesian A/B testing must be based on a predefined, reproducible and computationally feasible sequential decision protocol at scale, which reconciles the PAI, interval certainty and economic feasibility of continuation, considering the specifics of the metrics, minimising the risk of common errors and the unnecessary prolongation of experiments. A limitation of the study is its theoretical nature; no analysis of real empirical data or simulation experiments was carried out. Future research should focus on the empirical and simulation validation of the proposed protocol for the sequential termination of Bayesian A/B tests (PAI/Bayesian factor, interval certainty, economic feasibility), in particular using Google Colab, to assess its reliability and practical effectiveness.

Acknowledgements
None.

Funding
None.

Conflict of Interest
None.

References

[1] Ali, I. (2025). *Probabilistic A/B testing with rstanarm*. Retrieved from https://cran.r-project.org.

[2] Chennu, S., Maher, A., Pangerl, C., Prabanantham, S., Bae, J.H., Martin, J., & Goswami, B. (2023). Rapid and scalable Bayesian AB testing. *arXiv*. doi: 10.48550/arXiv.2307.14628.

[3] de Paula Castro, T., Paulino, C.D., & Singer, J.M. (2022). A fair comparison of credible and confidence intervals: An example with binomial proportions. *METRON*, 80, 371-382. doi: 10.1007/s40300-021-00225-6.

[4] De Santis, F., & Gubbiotti, S. (2021). Sample size requirements for calibrated approximate credible intervals for proportions in clinical trials. *International Journal of Environmental Research and Public Health*, 18(2), article number 595. doi: 10.3390/ijerph18020595.

[5] Douven, I. (2023). Bayesian stopping. *Journal of Mathematical Psychology*, 116, article number 102794. doi: 10.1016/j.jmp.2023.102794.

[6] Fu, Q., Moerbeek, M., & Hoijtink, H. (2022). Sample size determination for Bayesian ANOVAs with informative hypotheses. *Frontiers in Psychology*, 13, article number 947768. doi: 10.3389/fpsyg.2022.947768.

[7] Johari, R., Koomen, P., Pekelis, L., & Walsh, D. (2022). Always valid inference: Continuous monitoring of A/B tests. *Operations Research*, 70(3), 1806-1821. doi: 10.1287/opre.2021.2135.

[8] Kamalbasha, S., & Eugster, M.J.A. (2021). Bayesian A/B testing for business decisions. In P. Haber, T. Lampoltshammer, M. Mayr & K. Plankensteiner (Eds.), *Proceedings of the 3rd international data science conference: Data science – analytics and applications* (pp. 50-57). Vieweg Wiesbaden: Springer. doi: 10.1007/978-3-658-32182-6_9.

[9] Kelter, R., & Pawel, S. (2025). Bayesian power and sample size calculations for Bayes factors in the binomial setting. *arXiv*. doi: 10.48550/arXiv.2502.02914.

[10] Kunzmann, K., Grayling, M.J., Lee, K.M., Robertson, D.S., Rufibach, K., & Wason, J.M.S. (2021). A review of Bayesian perspectives on sample size derivation for confirmatory trials. *American Statistician*, 75(4), 424-432. doi: 10.1080/00031305.2021.1901782.

[11] Liao, J.G., Midya, V., & Berg, A. (2021). Connecting and contrasting the bayes factor and a modified ROPE procedure for testing interval null hypotheses. *American Statistician*, 75(3), 256-264. doi: 10.1080/00031305.2019.1701550.

[12] Markov, A., & Zaritskyi, O. (2025). Improvements to sequential approaches in A/B testing. In *Proceedings of the IEEE 5th international conference on smart information systems and technologies* (pp. 1-7). Astana: IEEE. doi: 10.1109/SIST61657.2025.11139233.

[13] Mezzetti, M., Ryan, C.P., Balestrucci, P., Lacquaniti, F., & Moscatelli, A. (2023). Bayesian hierarchical models and prior elicitation for fitting psychometric functions. *Frontiers in Computational Neuroscience*, 17, article number 1108311. doi: 10.3389/fncom.2023.1108311.

[14] Moerbeek, M. (2021). Bayesian updating: Increasing sample size during the course of a study. *BMC Medical Research Methodology*, 21, article number 137. doi: 10.1186/s12874-021-01334-6.
[15] Mulier, G., Lin, R., Aparicio, T., & Biard, L. (2024). Bayesian sequential monitoring strategies for trials of digestive cancer therapeutics. *BMC Medical Research Methodology*, 24, article number 154. doi: 10.1186/s12874-024-02278-3.
[16] Pan, J., & Banerjee, S. (2023). Bayesassurance: An R package for calculating sample size and Bayesian assurance. *The R Journal*, 15(3), 138-158. doi: 10.32614/RJ-2023-066.
[17] Pananos, D. (2024). *Calibration of mind changing for Bayesian AB tests*. Retrieved from https://dpananos.github.io.
[18] Quin, F., Weyns, D., Galster, M., & Silva, C.C. (2024). A/B testing: A systematic literature review. *Journal of Systems and Software*, 211, article number 112011. doi: 10.1016/j.jss.2024.112011.
[19] Rainforth, T., Foster, A., Ivanova, D.R., & Smith, F.B. (2024). Modern Bayesian experimental design. *Statistical Science*, 39(1), 100-114. doi: 10.1214/23-STS915.
[20] Richardson, T.S., Liu, Y., McQueen, J., & Hains, D. (2022). A Bayesian model for online activity sample sizes. In *Proceedings of the 25th international conference on artificial intelligence and statistics* (pp. 1775-1785). Valencia: PMLR.
[21] Sahu, D., Nidhi, Chaturvedi, R., Prakash, S., Yang, T., Rathore, R.S., & Alsolbi, I. (2025). Optimizing energy and latency in edge computing through a Boltzmann driven Bayesian framework for adaptive resource scheduling. *Scientific Reports*, 15, article number 30452. doi: 10.1038/s41598-025-16317-6.
[22] Shen, Y., Psioda, M.A., & Ibrahim, J.G. (2023). BayesPPD: An R package for Bayesian sample size determination using the power and normalized power prior for generalized linear models. *The R Journal*, 14(4), 335-351. doi: 10.32614/RJ-2023-016.
[23] Stefan, A.M., Schönbrodt, F.D., Evans, N.J., & Wagenmakers, E.-J. (2022). Efficiency in sequential testing: Comparing the sequential probability ratio test and the sequential Bayes factor test. *Behavior Research Methods*, 54(6), 3100-3117. doi: 10.3758/s13428-021-01754-8.
[24] Stucchio, C. (2015). *Bayesian A/B testing at VWO*. Retrieved from https://www.chrisstucchio.com.
[25] Sambucini, V. (2024). Hybrid classical-Bayesian approach to sample size determination for two-arm superiority clinical trials. *The International Journal of Biostatistics*, 20(2), 553-570. doi: 10.1515/ijb-2023-0050.
[26] Tkachuk, V.O., & Shestakova, A.V. (2022). Prospects for the use of mobile applications for digitalization of the main business processes in the restaurant business. *Economics, Management and Administration*, 4(102), 28-34. doi: 10.26642/ema-2022-4(102)-28-34.
[27] Turner, R.M., Clements, M.N., Quartagno, M., Cornelius, V., Cro, S., Ford, D., Tweed, C.D., Walker, A.S., & White, I.R. (2023). Practical approaches to Bayesian sample size determination in non-inferiority trials with binary outcomes. *Statistics in Medicine*, 42(8), 1127-1138. doi: 10.1002/sim.9661.
[28] Vasishth, S., Yadav, H., Schad, D.J., & Nicenboim, B. (2023). Sample size determination for Bayesian hierarchical models commonly used in psycholinguistics. *Computational Brain & Behavior*, 6, 102-126. doi: 10.1007/s42113-021-00125-y.
[29] Williamson, S.F., Williams, C.J., Lendrem, B.C., & Wilson, K.J. (2023). Sample size determination for point-of-care COVID-19 diagnostic tests: A Bayesian approach. *Diagnostic and Prognostic Research*, 7, article number 17. doi: 10.1186/s41512-023-00153-1.
[30] Wong, T.K., & Tendeiro, J.N. (2025). On a generalizable approach for sample size determination in Bayesian t tests. *Behavior Research Methods*, 57, article number 130. doi: 10.3758/s13428-025-02654-x.
[31] Wright, A., Ding, S., Philips, S.J., Matthew, R.A., & Ma, X. (2026). Model-based constrained Bayesian optimization of IEEE 802.11 VANET safety messaging. *Wireless Networks*, 32, 751-766. doi: 10.1007/s11276-025-04080-5.
[32] Zhang, L., Zheng, W., Zheng, K., Zhu, H., Li, C., & Guo, M. (2024). Bayesian-driven automated scaling in stream computing with multiple QoS targets. *IEEE Transactions on Parallel and Distributed Systems*, 35(7), 1251-1267.
[33] Zheng, H., Jaki, T., & Wason, J.M.S. (2023). Bayesian sample size determination using commensurate priors to leverage preexperimental data. *Biometrics*, 79(2), 669-683. doi: 10.1111/biom.13649.
[34] Zhou, T., & Ji, Y. (2024). On Bayesian sequential clinical trial designs. *The New England Journal of Statistics in Data Science*, 2(1), 136-151. doi: 10.51387/23-NEJSDS24.

Послідовна структура рішень для баєсівських A/B-тестів: узгодження точності інтервалу, практичної значущості та обмежень ресурсів

Артур Марков
Аспірант
Київський національний університет імені Тараса Шевченка
01033, вул. Володимирська, 60, м. Київ, Україна
https://orcid.org/0009-0000-5222-4397

Анотація. Дослідження мало на меті обґрунтувати підхід до планування та проведення баєсівських A/B-тестів, у рамках якого критерії інтервальної достовірності поєднуються з пороговими вимогами для практично значущого ефекту та оцінкою економічної доцільності подальшого збору даних. Методологія базувалася на методах аналітичного синтезу, структурно-логічної класифікації, логіко-алгоритмічної формалізації та ілюстративного аналізу випадків. У дослідженні було встановлено, що в баєсівських A/B-тестах «інтервальна точність» має два виміри: інформативність (вузькість та стабільність достовірного інтервалу для підтримки інженерного рішення) та каліброваність (близькість фактичного покриття до номінального рівня, який може відхилятися для малих вибірок). Дослідження продемонструвало, що невизначеність зменшується зі збільшенням n, але зі зменшенням віддачі; тому подальше звуження інтервалів вимагає непропорційно більших витрат часу та обчислювальних або експериментальних ресурсів і виправдане лише тоді, коли такі витрати є виправданими, причому необхідне n залежить від метрики та масштабу ефекту. Обсяг даних можна зменшити, включивши апріорну інформацію до апріорного розподілу або визначивши «достатність» на основі критеріїв доказів або ймовірності досягнення мети, тоді як висока мінливість інженерних метрик, рідкісні події, низькі базові показники подій, суворі вимоги до точності та невеликий ефект збільшують необхідне n та тривалість тесту. Щоб зменшити необхідний обсяг даних, було запропоновано послідовну структуру для завершення байєсівського A/B-тесту, що поєднує інформаційний компонент, цільовий компонент та перевірку економічної доцільності: «оновлення оцінок → поріг перевірки → інтервал перевірки → зважте переваги та витрати на продовження». Практичну застосовність цього підходу підтвердили інженерні дослідження, де байєсівські методи зменшили споживання ресурсів на 66,6 %, 36,7 % та 26,9 % у потокових обчисленнях та досягли економії енергії до 25 %, більшої енергоефективності на 15 % та приблизно на 10 мс менших гарантій затримки в периферійних обчисленнях. Практичне значення результатів полягає в їх застосовності дослідниками та аналітиками, які беруть участь у інженерних експериментах, під час прийняття рішень щодо завершення експерименту та впровадження запропонованих змін

Ключові слова: рішення; експеримент; дані; достатність; визначеність