AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(4): 10883–10907.
DOI:10.3934/math.2026447
Received: 05 February 2026
Revised: 27 March 2026
Accepted: 07 April 2026
Published: 20 April 2026

Research article

Bayesian decision making with soft probabilities under statistical regularity

Yuan Zou[1,2,*]

[1] School of Economics, Chongqing Technology and Business University, Chongqing 400067, China
[2] Chongqing Fintech Institute, Chongqing 401300, China

* **Correspondence:** Email: cqzouy@163.com.

Abstract: Soft set theory, originated by Molodtsov in 1999, is a general mathematical technique for modeling uncertainties via parameterization tools. A fundamental branch of this theory, soft probability, serves as an immediate measurement on a statistical base that takes a parameterized family of subintervals of the unit interval as its value. It can handle any stochastic events, including non-stable ones. To ensure soft probabilities accurately capture the statistical characteristics of stochastic information, the minimum sample size for their construction is determined using two statistical regularity hypotheses. Moreover, soft dependency guides attribute selection to enhance decision robustness against parameter perturbations. Hence, a novel Bayesian decision-making model integrating soft probabilities under statistical regularity hypotheses is proposed. Key steps of the method include: First, verifying these hypotheses on the initial database to determine the minimum sample size for soft probability construction; second, designing a soft dependency-based attribute selection procedure; third, calculating each decision alternative's soft posterior risk and identifying the optimal one by comparing interval-valued possibility degrees; and finally a medical diagnosis case study, combined with critical parameter sensitivity analysis and comparisons with Naive Bayes, logistic regressions (without regularization, L1 regularization and L2 regularization), decision tree, and support vector machine, demonstrates the method's feasibility and effectiveness.

Keywords: soft probabilities; statistical regularity; soft dependency; Bayes' rule; decision making
Mathematics Subject Classification: Primary 62C10; Secondary 60A10

1. Introduction

The types of uncertainties that exist in complicated problems in economics, engineering, and the environment have been tackled for a long time by philosophers, logicians, and mathematicians. In recent years, it has also become a crucial issue to address in the domain of decision making, especially when dealing with a small sample size. There are many mathematical tools such as probability theory, fuzzy sets [1], intuitionistic fuzzy sets [2], rough sets [3], and grey sets [4] for understanding and manipulating such uncertainties. The concept of soft set initiated by Molodtsov [5], offers a significant departure from conventional tools for quantifying uncertainties. Molodtsov argued that, without going into mathematical details, the aforementioned concepts are inadequate for handling uncertainties due to inadequate parametrization tools and practical challenges in precise modeling for exact solutions. Soft set theory differs from classical math in that it adopts an approximate description for the initial objects using any parametrization tool (like words, sentences, real numbers, functions, or mappings) and thus requires approximate solutions instead of exact ones. The absence of any restrictions on the approximate description makes soft set theory highly convenient to implement in practical applications. The author concluded several major potential applications, such as game theory, operations research, and soft analysis. Subsequently, the same author published a monograph on soft set theory [6]. This foundational work has led to significant theoretical adavances, as summarized in [7], and widespread practical applications, such as decision making [8,9], data filling [10], clustering [11], fault diagnosis [12], business failure prediction [13], rule mining [14], optimization [15], and assessment methods [16].

Soft probability [6], as a key component of soft set theory, takes the form of parameterized interval values rather than a single real number in the unit interval. Unlike the axiomatic definition in classical probability theory, soft probability relies entirely on statistical information and is defined through immediate measurements based on statistical data. Thus, it offers a dynamic description of stochastic phenomena in a computationally straightforward manner, as it changes with the addition of new statistical data. This adaptability makes it highly effective in practical applications under uncertain scenarios, particularly for those small samples involving non-stochastically stable information, arising in financial portfolio control [17], credit scoring [18], and Bayesian decision making [19]. However, these studies inadequately incorporate the statistical regularity of sample data in their soft probability computations, consequently making it difficult to accurately reflect the statistical characteristics of the database.

The Bayesian model provides an elegant approach to many data science and decision making problems, offering a general framework for integrating prior knowledge with new data to optimize choices under uncertainty. Based on Bayes' rule, which describes the updating of probabilities with new evidence, this approach helps decision makers quantify risks, evaluate alternatives, and determine optimal strategies in fields like economic behavior [20], risk management [21], classification [22], machine learning [23], medical diagnosis [24], and environmental science [25]. Note that practical decision making often relies on time series data, requiring the extraction of statistical regularities from samples of decision alternatives and their attributes. Such assessments involve uncertainty inherent in incomplete data. The classical Bayesian decision model is widely recognized for its logical approach to handling uncertainty but has two key limitations: First, it takes into account the framework of traditional point probabilities that require large samples under a universal statistical stability hypothesis, and thus cannot adequately characterize non-stochastically stable scenarios, potentially leading to inaccurate decisions. Second, the model assumes conditional independence and uses all attributes without selection, thereby increasing decision risk in a noisy database. The researchers in [26] and [27] suggested using interval probabilities instead of point probabilities to describe significant uncertainties and instabilities in stochastic information, and thus developed decision making techniques with interval probabilities. However, a critical unresolved issue lies in how to obtain the values of interval probabilities from stochastic information, especially from small sample sizes and non-stationary data. Fuzzy probability [28] can quantify random phenomena with fuzziness, but a widely discussed limitation of fuzzy probability is that the selection of membership functions for fuzzy sets is often subjective and non-data-driven, which may lead to instability of fuzzy probabilities and further influence the consistency of decision outcomes. A feasible Bayesian decision model with soft probabilities was proposed in [19], marking the first attempt to integrate Bayesian decision making within the framework of soft probabilities. However, the model employed subjective assignment to determine the sample size for calculating soft probabilities, neglecting the statistical regularity of specific samples. Additionally, it failed to consider reducing redundant attributes as a means to mitigate interference with decision outcomes. To make up for these deficiencies, we establish new theoretical connections between statistical regularity hypotheses and the Bayesian decision framework with soft probabilities, and implement attribute selection to enhance decision robustness. Thus, we propose a novel Bayesian model based on soft probability under statistical regularity hypotheses. The proposed model can capture stochastically stable information from large samples and non-stochastically stable information from small samples, while maintaining decision consistency across parameter perturbations.

Therefore, the major highlights of this work can be summarized as follows:

(1) A novel Bayes risk model in the framework of soft probabilities is developed.
(2) A criterion for minimum sample size determination in soft probability computation under the statistical regularity hypotheses is proposed.
(3) A soft dependency-based attribute selection procedure is designed.
(4) A combination strategy for multi-attribute decision making with ordered samples integrating Bayes risk minimization and possibility-based interval ranking is raised.

The remainder of the paper is structured as follows: In Section 2, we review some relevant background knowledge. In Section 3, we propose the Bayesian decision model with soft probabilities under statistical regularity hypotheses, as well as the corresponding implementation algorithm. In Section 4, a medical diagnosis case study validates the proposed algorithm's feasibility and effectiveness, with sensitivity analysis of critical parameter variations and comparisons with Naive Bayes, logistic regressions (without regularization, L1 regularization, and L2 regularization), a decision tree, and a support vector machine. Finally, conclusions and future work are summarized in Section 5.

2. Preliminaries

In this section, we provide a concise overview of foundational concepts, including soft probability, its associated statistical regularity hypotheses for determining minimum sample size, and Bayesian decision theory, to establish the theoretical basis for subsequent research.

2.1. Soft probabilities

The concept of soft set initiated by Russian mathematician Molodtsov, which serves as an adequate parametrization tool to deal with various types of uncertainties, is a starting point for constructing soft probability.

Definition 2.1. *[5] Let U be an initial universe of discourse, E be a set of parameters that can have an arbitrary nature (numbers, functions, sets of words, etc.), and 2^U represent the power set of U. A pair (F, E) is called a soft set over U if F is a mapping from the set E to the power set of U, i.e., $F : E \to 2^U$.*

Apparently, a soft set is not the Cantor set in the ordinary sense but rather a parameterized family of subsets of the set U. For each parameter $e \in E$, the subset $F(e)$ can be viewed as the set of e-elements or e-approximate elements of the soft set (F, E). In other words, it provides an approximate description of the concerned objects by using the parameterized family. To define soft probability, Molodtsov generalizes soft sets to soft mappings and thereby proposes soft random functions via this generalization.

Definition 2.2. *[17] Let X be a given set. A pair (F, E) is called a soft mapping over U if F is a mapping from the Cartesian product $X \times E$ to the power set of U, i.e., $F : X \times E \to 2^U$.*

Definition 2.3. *[17] Let Ω denote the set of possible outcomes, without restrictions regarding its properties or structure. Let $\mathbb{R}$ represent the set of real numbers. A soft random function is defined as any bounded real-valued function defined over Ω; thus, the set of soft random functions can be written in the form*

$$F = \left\{ f | f : \Omega \to \mathbb{R}, \sup_{\omega \in \Omega} |f(\omega)| < +\infty \right\}. \tag{2.1}$$

Subsequently, based on soft random functions, Molodtsov proposed the concept of soft probability, which can handle any stochastic events, including those that are non-stochastically stable. This differs from classical probability, which can deal with only stochastically stable phenomena. A series of specific repeated experiments leads to a sequence of outcomes called the statistical base, which refers to an ordered set denoted as $Base = (\omega_1, \omega_2, \ldots, \omega_n)$, each $Base_i = \omega_i \in \Omega$ and $Base \in \Omega^n$. The size of $Base$ is denoted by $|Base| = n$. Some operators on $Base$ are defined as follows.

Definition 2.4. *[6] Let Ω be a base outcome space, and let $Base = \{\omega_1, \omega_2, \ldots, \omega_n\}$ denote an ordered database.*

(1) The (i, j)-truncation operator retains only elements with indices from i to j, denoted as $Base_{(i,j)}$, and takes the form

$$Base_{(i,j)} = (\omega_i, \ldots, \omega_j), \tag{2.2}$$

if $j > |Base|$ and $i \leq |Base|$, then $Base_{(i,j)} = (\omega_i, \ldots, \omega_{|Base|})$. If $i > |Base|$, then $Base_{(i,j)} = \emptyset$.

(2) The k-sequence operator consists of a series of k consecutive elements from $Base$ without omissions or order changes, denoted as $Base^{(k)}$, and takes the form

$$Base^{(k)} = ((\omega_1, \ldots, \omega_k), (\omega_2, \ldots, \omega_{k+1}), \ldots, (\omega_{n-k+1}, \ldots, \omega_n)), \tag{2.3}$$

if $j > |Base|$, then $Base^{(j)} = Base^{(n)}$.

(3) Applying the (i, j)-truncation operator and the k-sequence operator on $Base$ jointly produce

$$Base^{(k)}{}_{(i,j)} = \Big((\omega_i, \ldots, \omega_{i+k-1}), \ldots, (\omega_j, \ldots, \omega_{j+k-1})\Big), \tag{2.4}$$

$$Base_{(i,j)}{}^{(k)} = \Big((\omega_i, \ldots, \omega_{i+k-1}), \ldots, (\omega_{j-k+1}, \ldots, \omega_j)\Big). \tag{2.5}$$

Example 2.1. *Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ be an initial base outcome space, and let $Base = (\omega_3, \omega_1, \omega_1, \omega_1, \omega_1, \omega_2)$ be an ordered database.*

(1) $Base_{(3,6)} = (\omega_1, \omega_1, \omega_1, \omega_2)$,
(2) $Base^{(4)} = ((\omega_3, \omega_1, \omega_1, \omega_1), (\omega_1, \omega_1, \omega_1, \omega_1), (\omega_1, \omega_1, \omega_1, \omega_2))$,
(3) ${Base^{(4)}}_{(1,2)} = ((\omega_3, \omega_1, \omega_1, \omega_1), (\omega_1, \omega_1, \omega_1, \omega_1))$,
${Base_{(2,6)}}^{(4)} = ((\omega_1, \omega_1, \omega_1, \omega_1), (\omega_1, \omega_1, \omega_1, \omega_2))$.

Let f be a soft random function, and $Base_{(i,i+k-1)} = (\omega_i, \ldots, \omega_{i+k-1})$ be a series of order k for the database $Base$. Then, the average $\langle f, Base_{(i,i+k-1)} \rangle$ of a soft random function f over $Base_{(i,i+k-1)}$ is defined as the arithmetic mean, taking the form

$$\langle f, Base_{(i,i+k-1)} \rangle = \frac{1}{k} \sum_{j=i}^{i+k-1} f(\omega_j). \tag{2.6}$$

For any natural numbers m and k such that $|Base_{(1,m)}| \geq k$, define two functionals

$$\underline{\mu}^k(Base_{(1,m)}, f) = \min_{1 \leq i \leq |Base_{(1,m)}|-k+1} \langle f, {Base_{(1,m)}}^{(k)} i \rangle, \tag{2.7}$$

$$\overline{\mu}^k(Base_{(1,m)}, f) = \max_{1 \leq i \leq |Base_{(1,m)}|-k+1} \langle f, {Base_{(1,m)}}^{(k)} i \rangle, \tag{2.8}$$

where ${Base_{(1,m)}}^{(k)} i$ indicates the i-th element in ${Base_{(1,m)}}^{(k)}$. The pair of functionals $(\underline{\mu}^k, \overline{\mu}^k)$ defines the boundary for the average values of a soft random function over the statistical base $Base_{(1,m)}$ for a given sample size k. In essence, this pair plays a role analogous to the expectation in classical probability theory. For a soft random function χ that is an indicator function of an event A ($A \subseteq \Omega$), i.e.,

$$\chi_A(\omega_i) = \begin{cases} 1, & \omega_i \in A, \\ 0, & \omega_i \notin A. \end{cases} \tag{2.9}$$

The pair of functionals can be considered as the soft probability of event A over $Base_{(1,m)}$.

Definition 2.5. *[6] Let $\mathbb{N}$ denote the set of natural numbers. A soft mapping (μ, E) over the set of real numbers $\mathbb{R}$ is called soft probability of event A if μ is a mapping $\mu : F \times Base \times E \to 2^{\mathbb{R}}$, where $\chi_A \in F$, the set of parameters E consists of pairs of (k, m), $k, m \in \mathbb{N}$, $|Base_{(1,m)}| \geq k$, and the mapping μ is given by*

$$\mu(Base, \chi_A, k, m) = \begin{cases} \left[\underline{\mu}^k(Base_{(1,m)}, \chi_A), \overline{\mu}^k(Base_{(1,m)}, \chi_A)\right], & |Base_{(1,m)}| \geq k, \\ \emptyset, & \textit{otherwise}. \end{cases} \tag{2.10}$$

Intuitively, soft probability is a mapping that associates each random function, database size, and series size (not exceeding the database size) with an interval whose endpoints are the minimum and maximum average values obtained by the random function over any series of the given length. It is worth noting that soft probability is a parametric family of sub-intervals within $[0, 1]$, with interval values depending on the sample size and the size of the statistical base, as well as providing a detailed description of average values. Moreover, it is defined directly through statistical data, eliminating the need for axiomatic foundations and exhibiting dynamic changes as new statistical data emerge.

For stochastically stable events given sufficiently large samples, soft probability refers to a narrow interval with a high confidence level. Conversely, it exhibits a wide interval for non-stochastically stable events. These features enable it to be applied to any types of events, including non-stochastically stable scenarios.

Example 2.2. *Consider the Example 2.1. Let $k = 3$ and $m = 5$. For the random event $A = \{\omega_1\} \subset \Omega$. One can compute*

$$\begin{aligned}
\underline{\mu}^3(Base_{(1,5)}, \chi_A) &= \min_{1\le i\le 3}\left(\langle \chi_A, Base_{(1,5)}{}^{(3)}i\rangle\right) \\
&= \min\left(\langle \chi_A, (\omega_3, \omega_1, \omega_1)\rangle, \langle \chi_A, (\omega_1, \omega_1, \omega_1)\rangle, \langle \chi_A, (\omega_1, \omega_1, \omega_1)\rangle\right) \\
&= \min\left((0+1+1)/3, (1+1+1)/3, (1+1+1)/3\right) = 2/3, \\
\overline{\mu}^3(Base_{(1,5)}, \chi_A) &= \max_{1\le i\le 3}\left(\langle \chi_A, Base_{(1,5)}{}^{(3)}i\rangle\right) \\
&= \max\left(\langle \chi_A, (\omega_3, \omega_1, \omega_1)\rangle, \langle \chi_A, (\omega_1, \omega_1, \omega_1)\rangle, \langle \chi_A, (\omega_1, \omega_1, \omega_1)\rangle\right) \\
&= \max\left((0+1+1)/3, (1+1+1)/3, (1+1+1)/3\right) = 1.
\end{aligned}$$

Then the soft probability of event $A = \{\omega_1\}$ over Base is $\mu(\chi_A, Base, 3, 5) = [2/3, 1]$.

Let "$Base|A$" denote the ordered collection of sample elements in *Base* given the occurrence of a random event A. The soft conditional probability describes statistical averages over *Base*, under the condition that outcomes belonging to only a certain set of outcomes $A \subseteq \Omega$ are considered. A natural way to formalize this description is to remove all elements not belonging to A from *Base* and then calculate the soft probability over the resulting database $Base \cap A$. The formal concept of soft conditional probability is defined as follows:

Definition 2.6. *[6] Let Ψ be another base outcome space, and let $Y \subseteq \Psi$, f be a mapping from Ω to Ψ. The soft conditional probability of an event Y given $A \subseteq \Omega$ is defined as a soft mapping (μ, E), in which $\mu : \chi_Y \times Base|A \times E \to 2^{\mathbb{R}}$, and the set of parameters E is composed of pairs of natural numbers (k, m) such that $|Base_{(1,m)} \cap A| \ge k$. The mapping μ is given by*

$$\mu(\chi_Y, Base|A, k, m) = \begin{cases} \left[\underline{\mu}^k(Base_{(1,m)} \cap A, \chi_Y), \overline{\mu}^k(Base_{(1,m)} \cap A, \chi_Y)\right], & |Base_{(1,m)} \cap A| \ge k, \\ \emptyset, & otherwise. \end{cases} \tag{2.11}$$

Note that the symbol χ_Y denotes an indicator function of Y. For $\forall \omega_i \in Base_{(1,m)} \cap A$, satisfying

$$\chi_Y(\omega_i) = \begin{cases} 1, & f(\omega_i) \in Y, \\ 0, & f(\omega_i) \notin Y. \end{cases} \tag{2.12}$$

Example 2.3. *Reconsider Example 2.2. Suppose that another base outcome space $\Psi = \{\psi_1, \psi_2\}$ and the random event $Y = \{\psi_1\} \subset \Psi$. Let a statistical database $Base' = (\psi_1, \psi_1, \psi_2, \psi_1, \psi_2, \psi_1)$, in which each trial is associated with an element of Ψ. Two statistical databases, Base and $Base'$, are represented in Table 1.*

Table 1. The tabular representation of *Base* and $Base'$.

	1	2	3	4	5	6
Base	ω_3	ω_1	ω_1	ω_1	ω_1	ω_2
$Base'$	ψ_1	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1

It is straightforward to compute

$$\begin{aligned}
\underline{\mu}^3(Base_{(1,5)} \cap A, \chi_Y) &= \min_{1\le i\le 2}\left(\langle \chi_Y, (Base_{(1,5)} \cap A)^{(3)} i\rangle\right)\\
&= \min_{1\le i\le 2}\left(\langle \chi_Y, (Base_2, Base_3, Base_4, Base_5)^{(3)} i\rangle\right)\\
&= \min_{1\le i\le 2}\left(\langle \chi_Y, (Base_2, Base_3, Base_4)\rangle, \langle \chi_Y, (Base_3, Base_4, Base_5)\rangle\right)\\
&= \min\left((1+0+1)/3, (0+1+0)/3\right) = 1/3,\\
\overline{\mu}^3(Base_{(1,5)} \cap A, \chi_Y) &= \max_{1\le i\le 2}\left(\langle \chi_Y, (Base_{(1,5)} \cap A)^{(3)} i\rangle\right)\\
&= \max_{1\le i\le 2}\left(\langle \chi_Y, (Base_2, Base_3, Base_4, Base_5)^{(3)} i\rangle\right)\\
&= \max_{1\le i\le 2}\left(\langle \chi_Y, (Base_2, Base_3, Base_4)\rangle, \langle \chi_Y, (Base_3, Base_4, Base_5)\rangle\right)\\
&= \max\left((1+0+1)/3, (0+1+0)/3\right) = 2/3.
\end{aligned}$$

Then the soft conditional probability of $Y = \{\psi_1\}$ *given* $A = \{\omega_3\}$ *over Base is* $\mu(\chi_Y, Base|A, 3, 5) = [1/3, 2/3]$.

2.2. Statistical regularity hypotheses on soft probability

All phenomena can be categorized into three types: The first type is characterized by deterministic regularity, meaning identical conditions always produce the same outcome. Phenomena of the second and third types exhibit non-deterministic regularity. The second type comprises statistically regular phenomena characterized by the statistical stability of outcome frequencies in large samples, while the third type includes all remaining phenomena that cannot be described by classical probability theory. It should be noted that the statistical database comprises outcomes of events that have occurred. Thus, the only significant aspect of statistical regularity lies in its role as a hypothesis about the future behavior of trial outcomes. Therefore, verification of the statistical regularity hypotheses for a statistical database of finite size is of paramount importance for constructing soft probability.

Definition 2.7. *[29] A random function* f *is called to be statistically* (k, ϵ)*-regular on a database* $Base_{(1,m)}$ *if for all samples* $I = {Base_{(1,m)}}^{(k)} i, J = {Base_{(1,m)}}^{(k)} j \in {Base_{(1,m)}}^{(k)}$ *such that*

$$|\langle f, I\rangle - \langle f, J\rangle| \le \epsilon, \tag{2.13}$$

where ϵ *is a positive real number that represents a predetermined bias threshold for* f *on* $Base_{(1,m)}$*. Clearly, a low bias is associated with high stability, and vice versa.*

For clarity, an equivalent formulation of statistical regularity for a random function is presented as follows: Let

$$2a = \max_{I\in {Base_{(1,m)}}^{(k)}} \langle f, I\rangle + \min_{I\in {Base_{(1,m)}}^{(k)}} \langle f, I\rangle, \tag{2.14}$$

$$2b = \max_{I\in {Base_{(1,m)}}^{(k)}} \langle f, I\rangle - \min_{I\in {Base_{(1,m)}}^{(k)}} \langle f, I\rangle. \tag{2.15}$$

The statistical regularity of f holds if and only if $2b \le \epsilon$. For any sample $I \in \mathrm{Base}_{(1,m)}^{(k)}$, this implies $|\langle f, I\rangle - a| \le b \le \epsilon/2$. This inequality is also equivalent to $\mu(Base, f, k, m) \subseteq [a - \epsilon/2, a + \epsilon/2]$.

Soft probability, as parameterized families, depends on the selection of parameter pairs (k, m). The base size m is typically quantified as the total cardinality of the initial *Base*; alternatively, it can be flexibly specified based on needs. The natural number k satisfies $1 \leq k \leq m$. Note that k cannot be too small, such as $k = 1$, the soft probability of event A in Example 2.2 is the unit interval $[0, 1]$, indicating complete uncertainty about A. Conversely, if k is too large, for instance $k = m$, the soft probability is a single point, i.e., $[\underline{\mu}^5, \overline{\mu}^5] = \left[\frac{4}{5}, \frac{4}{5}\right]$, which matches classical probability calculations but ignores the future possibility of trial outcome. By employing the statistical regularity hypothesis on soft probability, one can determine an appropriate value for k as the minimal positive integer such that for all $k' \geq k$, taking the form

$$k^* = \min_{1 \leq k \leq m} \{k | \mu(Base, f, k', m) \subseteq [a - \epsilon/2, a + \epsilon/2], \forall k' = k, k+1, \ldots, m\}, \tag{2.16}$$

where $\mu(Base, f, k', m) = \left[\underline{\mu}^{k'}(Base_{(1,m)}, f), \overline{\mu}^{k'}(Base_{(1,m)}, f)\right]$ denotes the *k'-approximate mean value* of f on $Base_{(1,m)}$.

Example 2.4. *Continue to consider the Example 2.2. Given $\epsilon = 1/4$. It is straightforward to compute*

$\mu(Base, \chi_{\omega_1}, 1, 5) = [0, 1] \nsubseteq [3/8, 5/8]$, $\mu(Base, \chi_{\omega_1}, 2, 5) = [1/2, 1] \nsubseteq [5/8, 7/8]$,
$\mu(Base, \chi_{\omega_1}, 3, 5) = [2/3, 1] \nsubseteq [17/24, 23/24]$, $\mu(Base, \chi_{\omega_1}, 4, 5) = [3/4, 1] \subseteq [3/4, 1]$,
$\mu(Base, \chi_{\omega_1}, 5, 5) = [4/5, 4/5] \subseteq [27/40, 37/40]$.

Then $k^* = 4$, and it is concluded that χ_{ω_1} is statistically $(4, 1/4)$-regular on the database $Base_{(1,5)}$.

It can be readily seen that the statistical regularity hypothesis of a random function depends on parameters and therefore requires dynamic verification and adjustment when new statistical data arise. This idea differs significantly from the hypothesis of classical probability theory, which is accepted universally for all possible events and are consequently neither adjusted nor re-verified. As a result, it can be widely applied in diverse scenarios, including cases where some events are statistically regular and others are not.

Using the statistical regularity of a random function, two types of hypotheses about its future values are proposed. These hypotheses focus on the future values of the statistical database.

Definition 2.8. *[29] A database Base is called to be statistically (k, ϵ)-regular with respect to a random function f if for all samples $I = {Base_{(1,m)}}^{(k)} i, J = {Base_{(1,m)}}^{(k)} j \in {Base_{(1,m)}}^{(k)}$ such that*

$$|\langle f, I \rangle - \langle f, J \rangle| \leq \epsilon. \tag{2.17}$$

Definition 2.9. *[29] A database Base is called to be statistically significantly $(k, \epsilon, \bar{a})$-regular with respect to a random function f if for all samples $I = {Base_{(1,m)}}^{(k)} i, J = {Base_{(1,m)}}^{(k)} j \in {Base_{(1,m)}}^{(k)}$ such that*

$$|\langle f, I \rangle - \bar{a}| \leq \epsilon, \tag{2.18}$$

where $\bar{a}$ is a real number that represents the analogue of the mean value of f over the database Base.

Note that the difference between Definitions 2.8 and 2.9 is that the former specifies only the length of an interval containing the approximate mean, while the latter determines the exact range boundaries for the approximate mean with a given accuracy. Once a type of hypothesis is selected, then it must be verified or adjusted as new statistical data appear.

Assume that the statistical regularity hypothesis universally holds in Base. Then, one can compute the minimum sample size required to satisfy this hypothesis for constructing soft probability. Parameter k, i.e., the minimum sample size, can be formally determined using the following criteria:

(1) If *Base* is *statistically* (k^*, ϵ)*-regular* with respect to f, then the parameter k^* is quantified as the minimal positive integer such that

$$k^* = \min_{1 \le k \le |Base|} \{k | \mu(Base, f, k', |Base|) \subseteq [a - \epsilon/2, a + \epsilon/2], \forall k' = k, k+1, \ldots, |Base|\}, \quad (2.19)$$

where $a = \left(\max_{I \in Base^{(k)}} \langle f, I \rangle + \min_{I \in Base^{(k)}} \langle f, I \rangle\right)/2$.

(2) If *Base* is *statistically significantly* $(k^*, \epsilon, \bar{a})$*-regular* with respect to f, then the parameter k^* is quantified as the minimal positive integer such that

$$k^* = \min_{1 \le k \le |Base|} \{k | \mu(Base, f, k', |Base|) \subseteq [\bar{a} - \epsilon, \bar{a} + \epsilon], \forall k' = k, k+1, \ldots, |Base|\}. \quad (2.20)$$

2.3. Bayesian decision theory

Bayesian decision theory is widely regarded as a powerful technique for handling uncertainty in a logical and consistent manner, which aims to achieve optimal decision(s) based on the principle of Bayes risk minimization. Its fundamental assumption is that prior knowledge follows a specific prior distribution, derived from subjective experience or historical data. By utilizing observed samples and Bayes' rule, it updates prior distribution to obtain posterior distribution, thereby identifying the decision option with the minimum Bayes risk as the optimal choice. The Bayesian decision process can be formalized as follows: Let $\mathcal{D} = \{d_1, d_2, \ldots, d_L\}$ denote the space of all possible decisions that can be chosen by the decision maker (DM), called the decision space, and $\Theta = \{\theta_1, \theta_2, \ldots, \theta_S\}$ denote the space of all possible states of nature. A probability mass function $P(\theta)$ on $\theta \in \Theta$ denotes the probabilities of the different states of nature θ based on prior knowledge, called the prior distribution. Note that if the spaces $\mathcal{D}$ and Θ are finite with respective dimensions L and S, then $P(\theta)$ is a vector of S probabilities. A loss function $\lambda(d_l, \theta_s) = \lambda_{ls}$ ($l = 1, 2, \ldots, L, s = 1, 2, \ldots, S$), which quantifies the consequences of choosing each decision $d_l \in \mathcal{D}$ for each possible outcome $\theta_s \in \Theta$, can be expressed as an $L \times S$ decision-loss matrix, all of whose components are real numbers, as shown below:

$$[\lambda(d_l, \theta_s)] = \begin{pmatrix} & \theta_1 & \theta_2 & \cdots & \theta_S \\ d_1 & \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1S} \\ d_2 & \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2S} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_L & \lambda_{L1} & \lambda_{L2} & \cdots & \lambda_{LS} \end{pmatrix}. \quad (2.21)$$

The Bayesian decision strategy establishes that an optimal decision should be selected by minimizing the posterior expected loss for the DM. Let $\mathcal{X}$ denote a set of Γ-dimensional attribute vectors $\mathbf{X} = (X_1, X_2, \ldots, X_\Gamma)$, where each component X_γ denotes an observed attribute value that is conditionally generated given the state $\theta \in \Theta$. During an experimental trial, the DM acquires a specific realization $\mathbf{X} = \mathbf{x}$. Given the prior probability of each state θ_s, denoted as $P(\theta_s)$, and the likelihood function $P(\mathbf{x}|\theta_s)$ quantifying the probability of observation $\mathbf{x}$ under state θ_s, under the Naive Bayes assumption (i.e., attribute conditional independence assumption), the posterior probability of state θ_s given a specific observation $\mathbf{x}$ can be derived using Bayes' rule

$$P(\theta_s|\mathbf{x}) = \frac{P(\theta_s)P(\mathbf{x}|\theta_s)}{\sum_{s=1}^{S} P(\theta_s)P(\mathbf{x}|\theta_s)} = \frac{P(\theta_s)\Pi_{\gamma=1}^{\Gamma} P(x_\gamma|\theta_s)}{\sum_{s=1}^{S} P(\theta_s)\Pi_{\gamma=1}^{\Gamma} P(x_\gamma|\theta_s)}. \quad (2.22)$$

Given an observed attribute vector $\mathbf{x}$, the posterior risk $R(d_l|\mathbf{x})$ associated with decision d_l is formally defined as the expected value of the loss function $\lambda(d_l, \theta_s)$ taking all $\theta_s \in \Theta$, given by

$$R(d_l|\mathbf{x}) = \sum_{s=1}^{S} \lambda(d_l, \theta_s) P(\theta_s|\mathbf{x}). \tag{2.23}$$

The Bayes risk adopting d_l, denoted as $R(d_l(\mathbf{x}))$ and defined as the posterior risk over all observations $\mathbf{x}$, is expressed as

$$R(d_l(\mathbf{x})) = \mathbb{E}_{\mathbf{x}} \left[R(d_l(\mathbf{x})|\mathbf{x}) \right]. \tag{2.24}$$

For each observation $\mathbf{x}$, if d^* minimizes the posterior risk $R(d_l|\mathbf{x})$, then the Bayes risk $R(d_l(\mathbf{x}))$ is also minimized. Thus, the Bayes decision $d_l \in \mathcal{D}$ is defined as the minimizer of posterior risk

$$d^*(\mathbf{x}) = \arg\min_{d_l \in D} R(d_l|\mathbf{x}). \tag{2.25}$$

If the loss function $\lambda(d_l, \theta_s)$ is the commonly used "$0 - 1$" model χ, defined as

$$\chi(d_l, \theta_s) = \begin{cases} 0, & \text{if } d_l = \theta_s, \\ 1, & \text{otherwise.} \end{cases} \tag{2.26}$$

Therefore, the posterior risk simplifies to $R(\theta_s|\mathbf{x}) = 1 - P(\theta_s|\mathbf{x})$. In this case, the minimization of Bayes risk reduces to maximum a posteriori (MAP) estimation

$$\theta^*(\mathbf{x}) = \arg\max_{\theta_s \in \Theta} P(\theta_s|\mathbf{x}). \tag{2.27}$$

The basic process of the Bayesian decision making model is illustrated in Figure 1. One can consult [30] for further details on Bayesian decision theory.

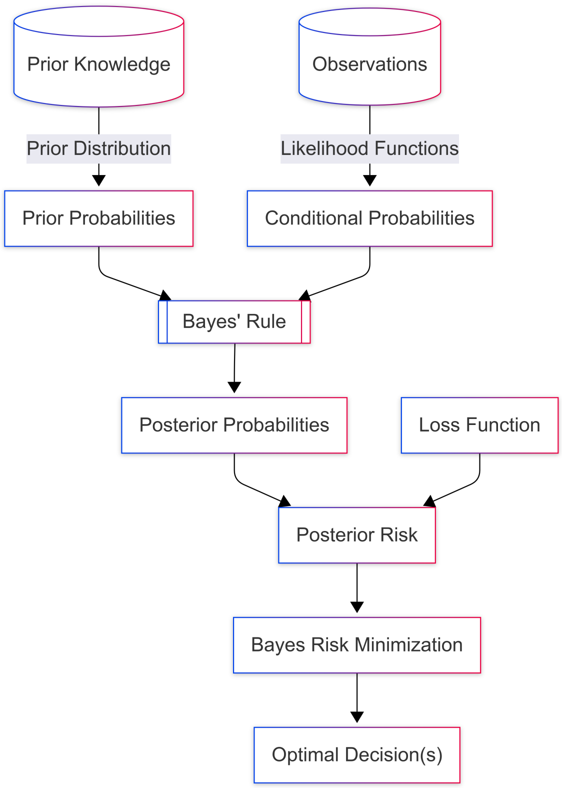

Figure 1. The schematic flowchart of Bayesian decision making.

3. Bayesian decision model based on soft probabilities under statistical regularity

In this section, we develop a Bayesian decision model that integrates soft probability theory with the Bayesian decision framework under statistical regularity hypotheses, addressing the challenges of multi-attribute decision making under a finite sequence of observations. The proposed methodology rigorously incorporates soft probability with Bayes risk minimization under statistical regularity to enhance decision robustness, thus being formally named the Bayesian decision model with soft probabilities under statistical regularity hypotheses (BDM-SP-SR).

A typical formulation of Bayesian decision making with soft probabilities involves a set of L discrete decision alternatives $\mathcal{D} = \{d_1, d_2, \ldots, d_L\}$ and a set of S mutually exclusive states of nature $\Theta = \{\theta_1, \theta_2, \ldots, \theta_S\}$. Let $\mathcal{X} = \{X_1, X_2, \ldots, X_\Gamma\}$ be a set of attributes, and define a Γ-dimensional attribute vector $\mathbf{X}$ as a complete assignment of attributes values to characterize each state, i.e., $\mathbf{X} = \mathbf{x}$, where $\mathbf{x} = (X_1 = x_1, \ldots, X_\Gamma = x_\Gamma)$. Specifically, each vector $\mathbf{x}$ uniquely corresponds to a specific state θ. Alternatives are evaluated using a decision-loss matrix $\lambda = [\lambda(d_l, \theta_s)]$, where $\lambda(d_l, \theta_s)$ represents the loss incurred by adopting alternative d_l when the actual state is θ_s. The prior probability $P(\theta_s)$ and the conditional probability $P(\mathbf{x}|\theta_s)$ are taken into account as soft probabilities that are represented as real sub-intervals of the unit interval $[0, 1]$. The primary objective is to select the optimal decision(s) that minimizes the Bayes risk. The detailed procedure of the proposed method is outlined in the following four major steps.

3.1. Ascertaining the minimum sample size and constructing prior soft probabilities

The prior probabilities are generally assumed to be precise point values in Bayesian decision making. However, in reality, prior information often exhibits inherent imprecision, subjectivity, or incompleteness, while sample data may not satisfy the stochastic stability assumption required for traditional probability theory. A more appropriate tool is to employ soft probabilities, which are interval-valued measures directly dependent on stochastic information. Thus, by modeling Bayesian decision making problems within soft probability frameworks, it becomes feasible to construct prior soft probabilities using available historical sample data.

Suppose a certain set of trial outcomes leads to an ordered database. For each attribute X_γ, there exists a sequence of attribute values $\{x_j^\gamma\}(j = 1, 2, \ldots, n)$. Correspondingly, there also exists a sequence of states $\{\theta_j\}(j = 1, 2, \ldots, n)$ taking values in the state space Θ. Thus, an ordered database denoted by $Base_{(1,n)} = (Base_1, Base_2, \ldots, Base_n)$ contains n records $Base_j$, each of which is expressed as $Base_j = \left((x_j^1, x_j^2, \ldots, x_j^\Gamma), \theta_j\right)$.

Given a state $\theta_s \in \Theta$, an indicator function χ is defined as

$$\chi_j^s = \begin{cases} 1, & \theta_j = \theta_s, \\ 0, & \theta_j \neq \theta_s. \end{cases} \tag{3.1}$$

By adopting the aforementioned strategy, multi-class classification for the state space can be simplified into multiple binary classifications.

As a result, the j-th record of $Base_{(1,n)}$ containing attribute values and specific state is obtained and denoted as $Base_j = (\mathbf{x}_j, \chi_j^s) = \left((x_j^1, x_j^2, \ldots, x_j^\Gamma), \chi_j^s\right)$, where $j \in \{1, 2, \ldots, n\}$. Let $E = \{(k, m)\}$ be a set of parameters. Note that m denotes the size of the sub-database $Base_{(1,m)}$ of the initial $Base_{(1,n)}$. Thus,

$m \le n$. However, typically, m is taken as the total size of $Base_{(1,n)}$, i.e., $m = n$ [29]. Given $\epsilon > 0$ defined as the predefined bias threshold, the value of k is calculated according to the criterion for determining the parameter k based on statistical regularity proposed above. In accordance with Definition 2.5, a soft mapping (μ, E) is formulated as $\mu(\chi_{\theta_s}, Base_{(1,n)}, k, m)$, where $\mu : \chi_{\theta_s} \times Base_{(1,n)} \times E \to 2^{\mathbb{R}}$. Then the prior soft probability of each $\theta_s \in \Theta$ is denoted as $\left[\underline{\mu}^k(Base_{(1,m)}, \chi_{\theta_s}), \overline{\mu}^k(Base_{(1,m)}, \chi_{\theta_s})\right]$, abbreviated as $\left[\underline{\mu}^k_m(\theta_s), \overline{\mu}^k_m(\theta_s)\right]$. Similarly, by Definition 2.6, the soft conditional probability of $X_\gamma = x_\gamma$ given $\theta_s \in \Theta$ is denoted as $\left[\underline{\mu}^k(Base_{(1,m)} \cap \theta_s, \chi_{x^\gamma}), \overline{\mu}^k(Base_{(1,m)} \cap \theta_s, \chi_{x^\gamma})\right]$, abbreviated as $\left[\underline{\mu}^k_m(x_\gamma \mid \theta_s), \overline{\mu}^k_m(x_\gamma \mid \theta_s)\right]$.

3.2. *Reduction of conditional attributes using soft dependency*

It is widely recognized that an excessive number of attributes in Bayesian decision making significantly increases computational complexity. Additionally, the states exhibit low dependency on some attributes, and thus removing these attributes does not substantially affect decision consistency. Thus, attribute selection is needed to remove low-dependency attributes. As such, this dependency is not a fixed value but varies with the parameter pair (m, k); and thus, within the theoretical framework of soft probability, it is essential to use Molodtsov's soft dependency concept [6] to accurately measure the association between states and individual attributes, as well as propose a concrete implementation procedure.

For each attribute $X_\gamma(\gamma = 1, 2, \ldots, \Gamma)$, one can consider a set $\Phi_{X_\gamma} = \{\rho^i_{X_\gamma}(\Theta)\}$ of real numbers, which are the results of dependency denoting that Θ depends in degree $\rho^i_{X_\gamma}(\Theta)$ $(0 \le \rho^i_{X_\gamma}(\Theta) \le 1)$ on X_γ. Based on partitioning the initial universe of discourse by parameters, the degree of dependency $\rho^i_{X_\gamma}(\Theta)$ can be defined as

$$\rho^i_{X_\gamma}(\Theta) = \frac{\left|\left\{(\mathbf{x}_j, \chi^s_j) \in Base_{(i,i+k-1)} \middle| [x^\gamma_j]_{X_\gamma} \subseteq [\chi^s_j]_\Theta\right\}\right|}{|Base_{(i,i+k-1)}|}, i = 1, 2, \ldots, m - k + 1, \tag{3.2}$$

where $[x^\gamma_j]_{X_\gamma} = \left\{t | x^\gamma_t = x^\gamma_j, j, t = i, i+1, \ldots, i+k-1\right\}$ and $[\chi^s_j]_\Theta = \left\{t | \chi^s_t = \chi^s_j, j, t = i, i+1, \ldots, i+k-1\right\}$.

The soft dependency of the dependency set Φ_{X_γ} is defined by intervals of minimum length that contain almost all values from Φ_{X_γ}. Formally, let $r \in [0, 1]$ be a parameter called risk. Define the set of admissible intervals Seg as

$$\text{Seg}(\Phi_{X_\gamma}, r) = \left\{[\alpha, \beta] \subseteq [0, 1] \middle| |S([\alpha, \beta])| \ge (1 - r)|\Phi_{X_\gamma}|\right\}, \tag{3.3}$$

where $|\cdot|$ denotes the cardinality of a set $\cdot$ and

$$S([\alpha, \beta]) = \left\{i \mid \rho^i_{X_\gamma}(\Theta) \in [\alpha, \beta] \subseteq [0, 1], i = 1, 2, \ldots, m - k + 1\right\}. \tag{3.4}$$

Note that for intervals $[\alpha, \beta]$, intuitively, the number of points in the set Φ_{X_γ} and within the interval $[\alpha, \beta]$ is greater than or equal to $(1 - r)|\Phi_{X_\gamma}|$ (where $|\Phi_{X_\gamma}| = (m - k + 1)$).

Thus, $[\alpha, \beta]$ is formulated as the solution to the following optimization problem:

$$\min_{[\alpha, \beta] \in \text{Seg}(\Phi_{X_\gamma}, r)} (\beta - \alpha). \tag{3.5}$$

The interval achieving this minimum constitutes the soft dependency of Φ_{X_γ}.

In what follows, given a dependency threshold ρ, one can compare the midpoint of the soft dependency interval $[\alpha,\beta]$, i.e., $(\alpha+\beta)/2$, with ρ. If $(\alpha+\beta)/2 < \rho$, this indicates that the dependency degree of the states on the attribute X_γ is below the threshold value, and X_γ is removed; otherwise, X_γ is retained. Applying this procedure to all attributes in $\mathcal{X}$ produces a reduced attribute set denoted as $Red(\mathcal{X})$, completing the attribute selection. This operation greatly reduces the computational complexity of the Bayes risk model, enhancing the practicality of the proposed method.

3.3. Calculating joints and soft posterior probabilities

Under the Naive Bayes assumption, i.e., given the state $\theta = \theta_s$, the attributes in $Red(\mathcal{X})$ are assumed mutually independent, and the joint soft conditional probabilities of attributes $\{\mathbf{X}_{Red(\mathcal{X})} = \mathbf{x}_{Red(\mathcal{X})} \mid \theta = \theta_s\}$ $(\theta_s \in \Theta)$ are formulated as

$$\begin{aligned} \underline{\mu}^k_m(\mathbf{X}_{Red(\mathcal{X})} = \mathbf{x}_{Red(\mathcal{X})} \mid \theta = \theta_s) &= \prod_{X_\gamma \in Red(\mathcal{X})} \underline{\mu}^k_m(X_\gamma = x_\gamma \mid \theta = \theta_s), \\ \overline{\mu}^k_m(\mathbf{X}_{Red(\mathcal{X})} = \mathbf{x}_{Red(\mathcal{X})} \mid \theta = \theta_s) &= \prod_{X_\gamma \in Red(\mathcal{X})} \overline{\mu}^k_m(X_\gamma = x_\gamma \mid \theta = \theta_s). \end{aligned} \tag{3.6}$$

Thus, the prior soft probabilities and the joint soft conditional probabilities can be regarded as special cases of interval probabilities [19]. By applying Bayes' rule for interval probabilities [31], the soft posterior probabilities are calculated as

$$\begin{aligned} \underline{\mu}^k_m(\theta = \theta_s \mid \mathbf{X}_{Red(\mathcal{X})} = \mathbf{x}_{Red(\mathcal{X})}) &= \frac{\underline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_s) \cdot \underline{\mu}^k_m(\theta_s)}{\underline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_s) \cdot \underline{\mu}^k_m(\theta_s) + \sum_{\theta \neq \theta_s} \overline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta) \cdot \overline{\mu}^k_m(\theta)}, \\ \overline{\mu}^k_m(\theta = \theta_s \mid \mathbf{X}_{Red(\mathcal{X})} = \mathbf{x}_{Red(\mathcal{X})}) &= \frac{\overline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_s) \cdot \overline{\mu}^k_m(\theta_s)}{\overline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_s) \cdot \overline{\mu}^k_m(\theta_s) + \sum_{\theta \neq \theta_s} \underline{\mu}^k_m(\mathbf{x}_{Red(\mathcal{X})} \mid \theta) \cdot \underline{\mu}^k_m(\theta)}. \end{aligned} \tag{3.7}$$

3.4. Optimal decision making via Bayes risk minimization

Bayesian decision making employs a loss function $\lambda(d_l, \theta_s)$ to quantify the cost of making the decision d_l when the actual state is θ_s. Given an attribute vector $\mathbf{X} = \mathbf{x}$ and the loss function $\lambda(d_l, \theta_s)$, the Bayes risk of making decision d_l given $\mathbf{x}$ is defined as

$$R(d_l|\mathbf{x}_{Red(\mathcal{X})}) = \mathbb{E}_{\theta_s|\mathbf{x}_{Red(\mathcal{X})}}\left[\lambda(d_l, \theta_s)\right] = \left[\underline{R}(d_l|\mathbf{x}_{Red(\mathcal{X})}), \overline{R}(d_l|\mathbf{x}_{Red(\mathcal{X})})\right], \tag{3.8}$$

where

$$\underline{R}(d_l|\mathbf{x}_{Red(\mathcal{X})}) = \sum_{s=1}^{S} \lambda(d_l, \theta_s) \underline{\mu}^k_m(\theta_s|\mathbf{x}_{Red(\mathcal{X})}), \tag{3.9}$$

$$\overline{R}(d_l|\mathbf{x}_{Red(\mathcal{X})}) = \sum_{s=1}^{S} \lambda(d_l, \theta_s) \overline{\mu}^k_m(\theta_s|\mathbf{x}_{Red(\mathcal{X})}). \tag{3.10}$$

Since the soft probabilities used in calculating Bayes risk are interval-valued, the resulting Bayes risk is inherently an interval. To compare interval-valued Bayes risk, a common interval ranking method from [32] is adopted to define the possibility degree (pd) for intervals $a = [\underline{a}, \overline{a}]$ and $b = [\underline{b}, \overline{b}]$ as

$$\mathrm{pd}(a \geq b) = \frac{\min\left\{(\overline{a} - \underline{a}) + (\overline{b} - \underline{b}), \max(\overline{a} - \underline{b}, 0)\right\}}{(\overline{a} - \underline{a}) + (\overline{b} - \underline{b})}, \tag{3.11}$$

where $\mathrm{pd}(a \geq b)$ measures the degree to which a is *superior*, *equivalent*, or *inferior* to b, and higher values indicate stronger evidence for $a \geq b$. Note that $\mathrm{pd}(a \geq b) > 1/2 \Rightarrow a > b$, $\mathrm{pd}(a \geq b) = 1/2 \Rightarrow a = b$, and $\mathrm{pd}(a \geq b) < 1/2 \Rightarrow a < b$. This measure satisfies $\mathrm{pd}(a \geq b) \in [0, 1]$, with straightforward properties as follows:

(1) $\mathrm{pd}(a \geq b) + \mathrm{pd}(b \geq a) = 1$;
(2) If $\overline{a} \leq \underline{b}$, then $\mathrm{pd}(a \geq b) = 0$, if $\overline{b} \leq \underline{a}$, then $\mathrm{pd}(a \geq b) = 1$;
(3) If $\mathrm{pd}(a \geq b) \geq 1/2$ and $\mathrm{pd}(b \geq c) \geq 1/2$, then $\mathrm{pd}(a \geq c) \geq 1/2$.

The pairwise comparison of soft posterior risks under different decisions yields a judgment matrix $[\mathrm{pd}_{lt}]_{L\times L}$, where $\mathrm{pd}_{lt} = \mathrm{pd}\left(R(d_l|\mathbf{x}) \geq R(d_t|\mathbf{x})\right)$. Without loss of generality, if $\mathrm{pd}_{lt} \geq 1/2$, the alternative d_t dominates d_l due to its smaller Bayes risk. Based on the matrix $[\mathrm{pd}_{lt}]_{L\times L}$, all alternatives are ranked to identify the most desirable one. The optimal decision $d_{l^*} \in \mathcal{D}$ is determined by minimizing the soft posterior risk, formally expressed as

$$d_{l^*} = \left\{ l^* \in \{1, 2, \ldots, L\} \mid \min_{l \in \{1,2,\ldots,L\}} \mathrm{pd}_{ll^*} = \mathrm{pd}_{l^* l^*} = 1/2 \right\}, \tag{3.12}$$

where $\mathrm{pd}_{ll^*} = \mathrm{pd}\left(R(d_l|\mathbf{x}) \geq R(d_{l^*}|\mathbf{x})\right)$ quantifies the likelihood that decision d_{l^*} is superior to d_l with a smaller soft posterior risk.

It is worth noting that for the “0–1” loss function ($\lambda_{ls} \in \{0, 1\}$), choosing d_{l^*} equivalently maximizes the soft posterior probability. Therefore, the implementation algorithm for the Bayesian decision model with soft probabilities under statistical regularity hypotheses is outlined below.

Step 1. Input the decision space $\mathcal{D} = \{d_1, d_2, \ldots, d_L\}$, the state space $\Theta = \{\theta_1, \theta_2, \ldots, \theta_S\}$, and the set of attributes $\mathcal{X} = \{X_1, X_2, \ldots, X_\Gamma\}$ for describing each state. As a result of repeated trials, a sequence of outcomes containing state data and attribute data for each state has been collected, and thus an ordered database $Base_{(1,n)}$ is established.

Step 2. Given a bias threshold $\epsilon > 0$, determine the parameter pair (m, k) based on the *statistically* (k, ϵ)*-regular* hypothesis on $Base_{(1,n)}$. Alternatively, given an estimate of the mean of f denoted by $\bar{a}$ and the bias threshold ϵ, determine (m, k) based on the *statistically significantly* $(k, \epsilon, \bar{a})$-*regular* hypothesis on $Base_{(1,n)}$.

Step 3. Given a risk level $0 \leq r \leq 1$, calculate the soft dependency $[\alpha_\gamma, \beta_\gamma]$ for each $X_\gamma \in \mathcal{X}$. Set $\rho \in [0, 1]$ as the dependency threshold, and reduce the attributes by comparing midpoints $\frac{\alpha_\gamma + \beta_\gamma}{2}$ ($\gamma \in \{1, 2, \ldots, \Gamma\}$) with ρ to obtain the reduced attribute set $Red(\mathcal{X})$.

Step 4. Calculate the prior soft probabilities $\left[\underline{\mu}^k_m(\theta_s), \overline{\mu}^k_m(\theta_s)\right]$ for each state $\theta_s \in \Theta$ and soft conditional probabilities $\left[\underline{\mu}^k_m(x_\gamma \mid \theta_s), \overline{\mu}^k_m(x_\gamma \mid \theta_s)\right]$ for $X_\gamma \in Red(\mathcal{X})$.

Step 5. Observe a new attribute vector $\mathbf{x}_{Red(X)}$ and derive the joint soft conditional probabilities of attributes given θ_s denoted by $\left[\underline{\mu}_m^k(\mathbf{x}_{Red(X)} \mid \theta_s), \overline{\mu}_m^k(\mathbf{x}_{Red(X)} \mid \theta_s)\right]$ under the Naive Bayes assumption. Apply Bayes' rule with interval probabilities to compute the soft posterior probabilities $\left[\underline{\mu}_m^k(\theta_s \mid \mathbf{x}_{Red(X)}), \overline{\mu}_m^k(\theta_s \mid \mathbf{x}_{Red(X)})\right] (\theta_s \in \Theta)$.

Step 6. Define the loss associated with each decision $d_l \in \mathcal{D}$ when the true state is $\theta_s \in \Theta$, and compute the soft posterior risk $R(d_l|\mathbf{x}_{Red(X)})$ for each decision d_l.

Step 7. Rank all decision alternatives using the interval number ranking principle, and select the optimal decision(s) by minimizing the soft posterior risk.

The flowchart of the algorithm above is shown in Figure 2.

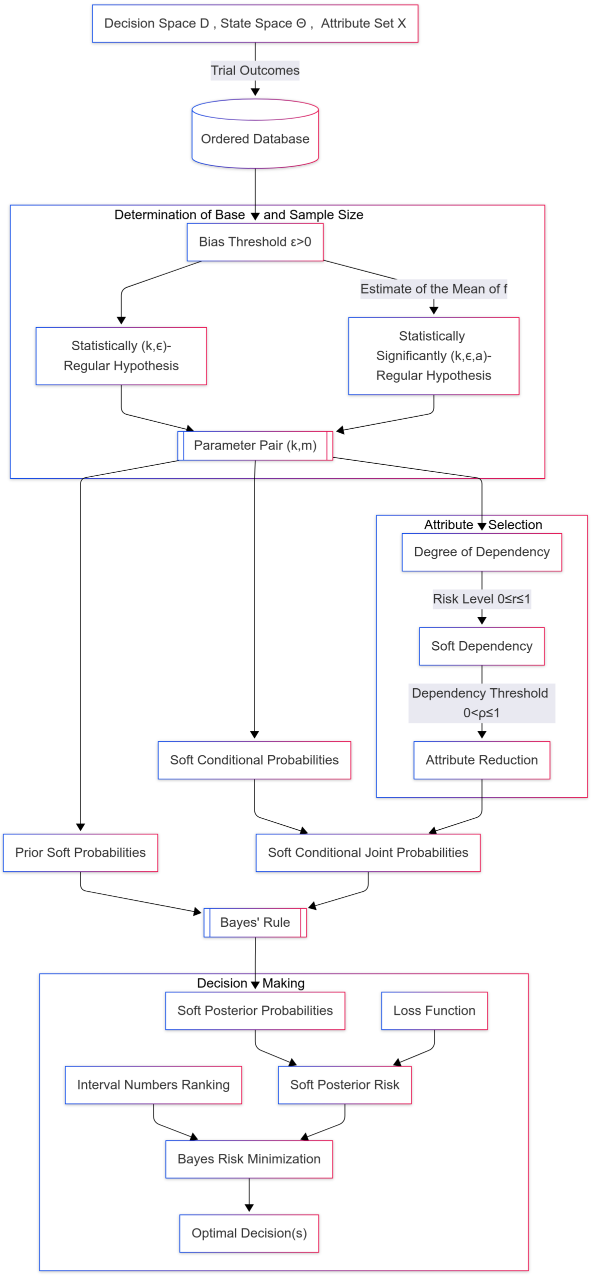

Figure 2. The implementation process of the proposed model.

3.5. The computational complexity of the proposed algorithm

The computational complexity of the proposed BDM-SP-SR algorithm is evaluated with respect to the number of instances n and attributes Γ. The dominant factor is the soft dependency calculation across W sliding windows, where $W = n - k + 1$ (assuming $m = n$). Computing the dependency degree for each attribute within a window of size k requires $O(k)$ operations, resulting in a total of $O(\Gamma W k)$. Subsequently, constructing the admissible intervals requires sorting the dependency sets, which contributes $O(\Gamma W \log W)$. Thus, the overall time complexity is $O(\Gamma n(k + \log n))$, which reduces to $O(\Gamma n \log n)$ when k is constant. Regarding storage, the space complexity is $O(n\Gamma)$, which is mostly determined by the database and temporary dependency arrays. Consequently, the proposed method achieves linear space and near-linear time complexities, demonstrating its suitability for large-scale applications.

4. Application of the proposed model

To demonstrate the validity and effectiveness of the proposed method, an illustrative case study is presented as follows:

In medical diagnosis, patients with a disease often exhibit multiple visible symptoms. It is important to note that some symptoms may appear in more than one disease, while other symptoms have limited diagnostic value in ruling out specific diseases, thereby creating diagnostic challenges for clinicians. Thus, prior information is often absent or unavailable due to inconsistency in subjective experience or lack of homogeneous historical diagnostic data. Doctors commonly analyze sample data to identify the most probable disease. Additionally, misdiagnosis risks exist in similar cases, and incorrect judgments can lead to different outcomes or losses. Consequently, this problem can be formulated as a Bayesian decision making problem, and the proposed approach addresses diagnostic uncertainties when prior knowledge and sample data are insufficient.

4.1. A numerical case study in medical diagnosis

Consider a typical medical diagnosis problem involving five symptoms, *headache*, *chest pain*, *abnormal blood pressure*, *cough*, and *fever*, and two diseases, *flu* (F) and *pneumonia* (P). Define the attribute set as $\mathcal{X} = \{X_1 : \textit{headache}, X_2 : \textit{chest pain}, X_3 : \textit{abnormal blood pressure}, X_4 : \textit{cough}, X_5 : \textit{fever}\}$ and the state space as $\Theta = \{\theta_1 = F, \theta_2 = P\}$. A statistical case database consisting of 20 sequentially diagnosed patients $\{\mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_{20}\}$ has been established (see Table 2), where the binary variable $x_j^\gamma = 1$ indicates that patient p_j suffers from X_γ, and 0 otherwise.

Suppose that a new patient p_{21} suffers from *abnormal blood pressure*, *cough*, and *fever*. The goal is to determine how the doctor can identify the patient's disease using the statistical case base and the patient's observed symptoms, distinguishing the two diseases. To achieve this goal, apply the proposed method to determine which disease is most consistent with the symptoms. The procedures are described as follows:

Step 1. In the decision making process, the state space is constituted by two diseases, *flu* (F) and *pneumonia* (P), and is formally denoted by $\Theta = \{F, P\}$. Two classes of diagnostic decisions are considered: Diagnosing the illness as *flu* (d_F) or *pneumonia* (d_P). These categories form a decision space denoted by $\mathcal{D} = \{d_F, d_P\}$, and the attribute set comprising the observed symptoms is denoted by $\mathcal{X} = \{X_1 = \textit{headache}, X_2 = \textit{chest pain}, X_3 = \textit{abnormal blood pressure}, X_4 = \textit{cough}, X_5 = \textit{fever}\}$.

Step 2. Given a bias threshold $\epsilon = 0.07$, the frequency of pneumonia occurrence in $Base_{(1,20)}$ is calculated as the mean estimate of χ_{θ_F}, yielding $\bar{a} = 0.4$. Based on the statistical regularity hypotheses, the *statistically* (k, ϵ)*-regular* hypothesis and the *statistically significantly* $(k, \epsilon, \bar{a})$*-regular* hypothesis, the parameter k is determined to be 15 under the bias threshold $\epsilon = 0.07$. Thus, the parameter pair is established as $(m, k) = (20, 15)$. Likewise, the statistical regularity hypotheses established for χ_{θ_F} also apply to χ_{θ_P}.

Step 3. Given a risk level $r = 0.2$ and a dependency threshold $\rho = 0.1$, the soft dependency of states on each attribute is calculated. By comparing the midpoint of the soft dependency interval with $\rho = 0.1$, the reduced attribute set $Red(\mathcal{X}) = \{X_1, X_2, X_4, X_5\}$ is obtained. Detailed results are presented in Table 3.

Step 4. The prior soft probabilities of suffering from *flu* is calculated as

$$\mu(Base_{(1,20)}, \chi_{\theta_F}, 15, 20) = \left[\underline{\mu}_{20}^{15}(\theta_F), \overline{\mu}_{20}^{15}(\theta_F)\right], \tag{4.1}$$

where

$$\underline{\mu}_{20}^{15}(\theta_F) = \min_{1\le i\le 6} \frac{1}{15} \sum_{j=i}^{i+k-1} \chi_{\theta_F}(\theta_j) = 0.533, \quad \overline{\mu}_{20}^{15}(\theta_F) = \max_{1\le i\le 6} \frac{1}{15} \sum_{j=i}^{i+k-1} \chi_{\theta_F}(\theta_j) = 0.600. \tag{4.2}$$

Moreover, the prior soft probabilities of suffering from *pneumonia* is calculated as

$$\mu(Base_{(1,20)}, \chi_{\theta_P}, 15, 20) = \left[\underline{\mu}_{20}^{15}(\theta_P), \overline{\mu}_{20}^{15}(\theta_P)\right], \tag{4.3}$$

where

$$\underline{\mu}_{20}^{15}(\theta_P) = \min_{1\le i\le 6} \frac{1}{15} \sum_{j=i}^{i+k-1} \chi_{\theta_P}(\theta_j) = 0.400, \quad \overline{\mu}_{20}^{15}(\theta_P) = \max_{1\le i\le 6} \frac{1}{15} \sum_{j=i}^{i+k-1} \chi_{\theta_P}(\theta_j) = 0.467. \tag{4.4}$$

The results of soft conditional probabilities for suffering from *flu* or *pneumonia* are calculated in a similar manner and presented in Table 4 below.

Step 5. Given a new patient p_{21} exhibiting symptoms of *abnormal blood pressure*, *cough*, and *fever*, the corresponding attribute vector is $\mathbf{x}_{Red(\mathcal{X})} = (X_1 = 0, X_2 = 0, X_4 = 1, X_5 = 1)$ (where X_3 is removed). Consequently, the joint soft conditional probabilities of attributes given $\mathbf{x}_{Red(\mathcal{X})}$ are computed as

$$\left[\underline{\mu}_{20}^{15}(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_F), \overline{\mu}_{20}^{15}(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_F)\right] = [0.031, 0.139], \tag{4.5}$$

$$\left[\underline{\mu}_{20}^{15}(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_P), \overline{\mu}_{20}^{15}(\mathbf{x}_{Red(\mathcal{X})} \mid \theta_P)\right] = [0.000, 0.072]. \tag{4.6}$$

Thus, the soft posterior probabilities are calculated using Bayes' rule with interval probabilities as follows:

$$\left[\underline{\mu}_{20}^{15}(\theta_F \mid \mathbf{x}_{Red(\mathcal{X})}), \overline{\mu}_{20}^{15}(\theta_F \mid \mathbf{x}_{Red(\mathcal{X})})\right] = [0.329, 1], \tag{4.7}$$

$$\left[\underline{\mu}_{20}^{15}(\theta_P \mid \mathbf{x}_{Red(\mathcal{X})}), \overline{\mu}_{20}^{15}(\theta_P \mid \mathbf{x}_{Red(\mathcal{X})})\right] = [0, 0.671]. \tag{4.8}$$

Step 6. Define the loss function $\lambda(d_l, \theta_s)$ to measure the loss of adopting decision d_l when patient p_i has disease θ_s. Note that different loss functions yield different decision losses. The "0 − 1" model, a commonly used loss function, effectively assesses decision correctness. Thus, using the "0 − 1" model, the decision-loss matrix is given by

$$[\lambda(d_l, \theta_s)] = \begin{pmatrix} & \theta_1 = F & \theta_2 = P \\ d_1 = d_F & 0 & 1 \\ d_2 = d_P & 1 & 0 \end{pmatrix}. \tag{4.9}$$

Consequently, the soft posterior risks of the two decisions given the observed symptoms are computed as

$$R(d_F|\mathbf{x}_{Red(X)}) = \left[\underline{R}(d_F|\mathbf{x}_{Red(X)}), \overline{R}(d_F|\mathbf{x}_{Red(X)})\right] = [0, 0.671], \tag{4.10}$$

$$R(d_P|\mathbf{x}_{Red(X)}) = \left[\underline{R}(d_P|\mathbf{x}_{Red(X)}), \overline{R}(d_P|\mathbf{x}_{Red(X)})\right] = [0.329, 1]. \tag{4.11}$$

Step 7. Based on the possible degree for interval number ranking, a minimizer of $R(d_l|\mathbf{x}_{Red(X)})$ with respect to $d_l \in \{d_F, d_P\}$ is determined as below:

$$\mathrm{pd}(R(d_F|\mathbf{x}_{Red(X)}) \geq R(d_P|\mathbf{x}_{Red(X)})) = 0.255 < 0.5. \tag{4.12}$$

Hence, $R(d_F|\mathbf{x}_{Red(X)}) < R(d_P|\mathbf{x}_{Red(X)})$, implying that the optimal decision d_F (diagnosing *flu*) is preferable to d_P (diagnosing *pneumonia*) for patient p_{21}.

Table 2. The tabular representation of an ordered sample database for *flu* (F) and *pneumonia* (P).

Patient	X_1	X_2	X_3	X_4	X_5	$\Theta = \{F, P\}$
p_1	0	1	0	1	0	F
p_2	0	1	1	1	1	P
p_3	0	1	1	0	0	F
p_4	1	1	0	1	1	P
p_5	0	0	0	0	0	F
p_6	0	0	1	1	0	F
p_7	0	1	0	1	1	P
p_8	0	0	0	0	0	F
p_9	1	1	1	1	1	P
p_{10}	0	1	0	1	1	P
p_{11}	0	0	1	0	0	F
p_{12}	1	1	0	1	1	P
p_{13}	0	0	1	1	1	F
p_{14}	0	0	0	1	0	F
p_{15}	1	1	0	1	1	P
p_{16}	0	0	0	0	0	F
p_{17}	1	0	0	1	1	F
p_{18}	1	0	0	1	1	P
p_{19}	0	0	0	0	0	F
p_{20}	0	0	0	1	0	F

Table 3. The results of attribute selection using soft dependency with $(r, \rho) = (0.2, 0.1)$.

Attribute	Φ_{X_γ}	Soft dependency	Midpoint VS $\rho = 0.1$	Selection
X_1	$\{0.267, 0.267, 0, 0, 0, 0\}$	$[0, 0.267]$	$0.134 > 0.1$	Retained
X_2	$\{0.4, 0.467, 0.533, 0.4, 0.333, 0.333\}$	$[0.333, 0.467]$	$0.400 > 0.1$	Retained
X_3	$\{0, 0, 0, 0, 0, 0\}$	$[0, 0]$	$0 < 0.1$	Removed
X_4	$\{0.267, 0.333, 0.333, 0.267, 0.333, 0.267\}$	$[0.267, 0.333]$	$0.300 > 0.1$	Retained
X_5	$\{0.467, 0.467, 0.467, 0.4, 0.467, 0.467\}$	$[0.467, 0.467]$	$0.467 > 0.1$	Retained

Table 4. The soft conditional probabilities of attributes with respect to *flu* (F) or *pneumonia* (P).

$\mu(x_1 = 1 \mid \theta_1 = F) = [0.000, 0.125]$	$\mu(x_1 = 0 \mid \theta_1 = F) = [0.875, 1.000]$	$\mu(x_2 = 1 \mid \theta_1 = F) = [0.000, 0.250]$	$\mu(x_2 = 0 \mid \theta_1 = F) = [0.750, 1.000]$
$\mu(x_1 = 1 \mid \theta_2 = P) = [0.571, 0.714]$	$\mu(x_1 = 0 \mid \theta_2 = P) = [0.286, 0.429]$	$\mu(x_2 = 1 \mid \theta_2 = P) = [0.833, 1.000]$	$\mu(x_2 = 0 \mid \theta_2 = P) = [0.000, 0.167]$
$\mu(x_4 = 1 \mid \theta_1 = F) = [0.375, 0.556]$	$\mu(x_4 = 0 \mid \theta_1 = F) = [0.444, 0.625]$	$\mu(x_5 = 1 \mid \theta_1 = F) = [0.125, 0.250]$	$\mu(x_5 = 0 \mid \theta_1 = F) = [0.750, 0.875]$
$\mu(x_4 = 1 \mid \theta_2 = P) = [1, 1]$	$\mu(x_4 = 0 \mid \theta_2 = P) = [0, 0]$	$\mu(x_5 = 1 \mid \theta_2 = P) = [1, 1]$	$\mu(x_5 = 0 \mid \theta_2 = P) = [0, 0]$

4.2. Sensitivity analysis

To assess the robustness of the BDM-SP-SR method, a sensitivity analysis is conducted by varying three critical parameters: bias threshold ϵ, risk level r, and dependency threshold ρ.

(1) **bias threshold** ϵ. Note that ϵ controls the stringency of statistical regularity determination, establishes a trade-off between "estimation precision" and "minimum sample size requirement", and plays a key role in verifying statistical regularity on the dataset and deciding the minimum sample size. To investigate the effect on the minimum sample size k^*, one can choose the threshold ϵ from 0.01 to 0.1 in steps of 0.01, as illustrated in Figure 3.

As ϵ increases, k^* generally decreases, indicating that relaxed constraints enable smaller minimum sample size to meet the requirements. When $0 \leq \epsilon \leq 0.05$, $k^* \equiv 19$ consistently holds. This is because the prior soft probabilities of *flu* and *pneumonia* are $\mu(Base_{(1,20)}, \chi_{\theta_F}, 19, 20) = 0.579$ and $\mu(Base_{(1,20)}, \chi_{\theta_P}, 19, 20) = 0.421$, respectively. Both degenerate into exact point probabilities and thus result in $2b = 0 \leq \epsilon$ holding constantly. Since point probabilities cannot capture non-stochastic stable information, one can restrict ϵ to $[0.06, 0.1]$. The dynamic impact of changes in ϵ on decision results is shown in the Table 5.

As we can see, a large ϵ (e.g., $\epsilon = 0.1$ or 0.09) implies high uncertainty in the statistical regularity hypothesis, resulting in significant uncertainty in decision results that cannot distinguish *flu* and *pneumonia*. By contrast, a smaller ϵ reduces the uncertainty of the hypothesis, leading to more precise decision results that consistently support the diagnosis of *flu* for patient p_{21}. It is worth emphasizing that ϵ should not be too small (e.g., $\epsilon \leq 0.05$), as the statistical regularity assumption will hold constantly, thereby causing biased decision results when dealing with non-stochastically stable information.

(2) **The parameter pair** (r, ρ). Parameter r determines the coverage proportion for constructing soft dependency intervals and establishes a trade-off between "interval width" and "coverage reliability". As a dependency threshold, ρ determines if the dependency between attributes and states is strong enough by comparing it to a criterion, establishes a trade-off between "attribute relevance" and "model complexity" to reduce attributes while keeping key attributes. Let $\epsilon = 0.07$ be fixed; thus, $(m, k) = (20, 15)$. It follows that the set Φ_{X_γ} contains six elements. Therefore, one can set the range of r as $[0, 1]$ with a step of $1/6$, and select the parameter ρ that varies from 0 to 0.5 with a 0.1 step to generate the reduced attribute set. The dynamic impact of changes in (r, ρ) on decision results is shown in Table 6.

The results in Table 6 show that the patient p_{21} should be diagnosed as *flu* under different parameter pairs (r, ρ) with $\rho \in \{0, 0.1, 0.2, 0.3\}$. It is important to note that the dependency threshold ρ should not be set too high. For instance, setting $\rho = 0.4$ would eliminate the key attribute X_2 (*chest pain*), which would result in a misdiagnosis of *pneumonia*. From a practical perspective, the selection of (r, ρ) should be guided by domain knowledge and the acceptable level of risk. A lower r ($r \leq 0.2$) ensures high coverage ($\geq 80\%$) of dependency values, yielding more reliable intervals. The threshold ρ should be chosen based on the relative importance of attributes: If attribute reduction is critical for interpretability or computational efficiency, a moderate ρ ($\rho \in [0.1, 0.2]$) can be used; if preserving all potentially relevant attributes is paramount, a low ρ ($\rho = 0$) is advisable. In medical diagnosis, where each symptom may be clinically significant, a conservative ρ ($\rho \leq 0.1$) is recommended to avoid discarding important diagnostic indicators.

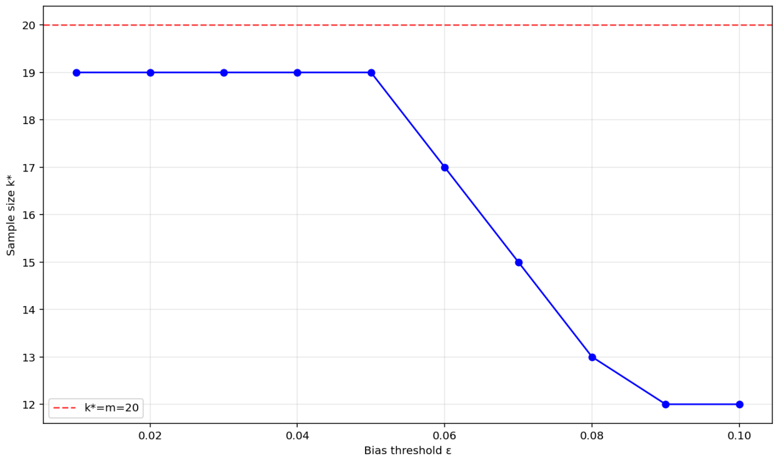

Figure 3. The impact of bias threshold ϵ on the minimum sample size k^*.

Table 5. The impact of ϵ on minimum sample size k^*, reduced attribute set $Red(\mathcal{X})$, and decision possibility degree ($(r, \rho) = (0.2, 0.1)$).

ϵ	k^*	$Red(\mathcal{X})$	$\mathrm{pd}(R(d_F \mid \mathbf{x}_{Red(\mathcal{X})}) \geq R(d_P \mid \mathbf{x}_{Red(\mathcal{X})}))$
0.10	12	$\{X_1, X_2, X_4, X_5\}$	0.5
0.09	12	$\{X_1, X_2, X_4, X_5\}$	0.5
0.08	13	$\{X_1, X_2, X_4, X_5\}$	0.282
0.07	15	$\{X_1, X_2, X_4, X_5\}$	0.255
0.06	17	$\{X_2, X_4, X_5\}$	0.248

Table 6. The reduced attribute sets and decision possibility degrees under different (r, ρ) with $\epsilon = 0.07$.

(r, ρ)	$Red(X)$	$\text{pd}(R(d_F\|\mathbf{x}_{Red(X)}) \geq R(d_P\|\mathbf{x}_{Red(X)}))$
$(i/6, 0)$, $i = \{0, 1, 2, 3, 4, 5, 6\}$	$\{X_1, X_2, X_3, X_4, X_5\}$	0.214
$(i/6, 0.1), i = \{0, 1\}$	$\{X_1, X_2, X_4, X_5\}$	0.255
$(i/6, 0.1), i = \{2, 3, 4, 5, 6\}$	$\{X_2, X_4, X_5\}$	0.380
$(i/6, 0.2), i = \{0, 1, 2, 3, 4, 5, 6\}$	$\{X_2, X_4, X_5\}$	0.380
$(i/6, 0.3), i = \{0, 1, 2\}$	$\{X_2, X_4, X_5\}$	0.380
$(i/6, 0.3), i = \{3, 4, 5, 6\}$	$\{X_2, X_5\}$	0.179
$(i/6, 0.4), i = \{0, 1, 3\}$	$\{X_2, X_5\}$	0.179
$(i/6, 0.4), i = \{2, 4, 5, 6\}$	$\{X_5\}$	1
$(i/6, 0.5), i = \{0, 1, 2, 3, 4, 5, 6\}$	$\emptyset$	–

4.3. Comparative analysis

To evaluate the efficiency of the proposed BDM-SP-SR, comparative experiments are conducted using well-established classification methods as baseline models: Naive Bayes (NB), logistic regression (LR) without regularization, L1-regularized logistic regression (L1-LR), L2-regularized logistic regression (L2-LR), decision tree (DT), and support vector machine (SVM). The decision outcomes for the diagnostic classification of patient p_{21} using the above-mentioned methods are summarized in Table 7.

Table 7. Comparison of decision outcomes across methods.

Methods	Decision basis	Decisions for p_{21}
NB	$R(d_F\|\mathbf{x}) \approx 0.266 < 0.734 \approx R(d_P\|\mathbf{x})$	*flu* (F)
LR (without regularization)	$P(F\|\mathbf{x}) \approx 0.9999 > 0.0001 \approx P(P\|\mathbf{x})$	*flu* (F)
L1-LR	$P(F\|\mathbf{x}) \approx 0.4430 < 0.5570 \approx P(P\|\mathbf{x})$	*pneumonia* (P)
L2-LR	$P(F\|\mathbf{x}) \approx 0.6565 > 0.3435 \approx P(P\|\mathbf{x})$	*flu* (F)
DT	Decision path: $X_5 = 1 \rightarrow X_2 = 0 \rightarrow X_1 = 0 \rightarrow$ Terminal node: F	*flu* (F)
SVM	$X_2 - X_3 < 0.5\ (F), X_2 - X_3 > 0.5\ (P); X_2(0) - X_3(1) < 0.5$	*flu* (F)
BDM-SP-SR	$\text{pd}(R(d_F\|\mathbf{x}_{Red(X)}) \geq R(d_P\|\mathbf{x}_{Red(X)})) = 0.255 < 0.5$	*flu* (F)

(1) **Comparison with the NB model.** The proposed BDM-SP-SR model contrasts with the NB model mostly in its treatment of uncertainty and attribute dependency. While NB assumes strict attribute independence and uses point probabilities, BDM-SP-SR employs interval probabilities $[\underline{\mu}, \overline{\mu}]$ under statistical regularity hypotheses, explicitly modeling uncertainty through parameters such as risk level r and dependency threshold ρ. BDM-SP-SR further incorporates data-driven attribute reduction based on soft dependency degrees, whereas NB uses all attributes without selection, which may lead to overfitting when dealing with noisy data. In decision making, BDM-SP-SR minimizes soft posterior risk with interval arithmetic, adapting to risk preferences, while NB relies on point-wise maximum posterior probability and is sensitive to inaccuracies in prior probability specification. As such, BDM-SP-SR can be seen as a generalization of the NB model. To be precise, NB corresponds to BDM-SP-SR when the parameter triplet (ϵ, r, ρ) is set to $(0, 0, 0)$. In this scenario, the posterior risks are quantified as $R(d_F|\mathbf{x}) \approx 0.266$ and $R(d_P|\mathbf{x}) \approx 0.734$, with

$R(d_F) \leq R(d_P)$. This results in the diagnosis of patient p_{21} with the *flu*, illustrating the consistency of decision outcomes across the models. These advances enable BDM-SP-SR to outperform NB in domains requiring robust uncertainty handling, particularly with small samples or redundant attributes.

(2) **Comparison with the other models.** As illustrated, the remaining baseline models, except L1-LR, are consistent with BDM-SP-SR in diagnosing the *flu* (F), but they exhibit structural weaknesses in small-sample scenarios. LR without regularization suffers from severe overfitting and numerical instability, as shown by its extreme probability ($P \approx 0.9999$). L1-LR is the only model that predicts *pneumonia* (P); its attribute selection mechanism may eliminate critical diagnostic attributes when historical data are sparse, resulting in biased inference. While L2-LR stabilizes parameter estimation, it does not perform attribute reduction, thereby retaining redundant features that increase model complexity and may obscure underlying causal relationships. DT and SVM yield results consistent with the proposed method in this case, and their limitations are evident: DT relies only on a subset of samples for determining its decision path, and SVM uses partial features (X_2 and X_3) to construct the classification hyperplane. This reliance on incomplete information leads to insufficient learning of data patterns and thus may result in poor generalization and low decision accuracy.

5. Conclusions and future work

To minimize decision risk in the Bayesian model with ordered samples, we propose a novel Bayesian decision model with soft probabilities under statistical regularity hypotheses (BDM-SP-SR), addressing non-stochastic stability in soft probability computation and attribute selection for small sample scenarios. Note that soft probability, as a special case of interval probability, is defined by immediate measurements on a statistical base and changes dynamically with new data. It offers a more straightforward representation to reflect stochastic and non-stochastic stability, which widely exists in practical problems. Thus, BDM-SP-SR can be regarded as a practical version of the Bayesian decision model with interval probabilities. By assuming that the statistical regularity hypothesis holds universally across the initial database under a given bias threshold, the minimum sample size for constructing soft probabilities is correspondingly quantified. This adaptability is data-driven and enhances decision reliability, especially when statistical regularities in data change over time. A soft dependency-based procedure is designed for attribute reduction, and a decision making strategy with ordered samples is proposed based on the principle of Bayes risk minimization in the soft probability framework. Finally, a numerical case study regarding medical diagnosis is conducted to verify the reasonableness and effectiveness of the proposed methodology.

The major contributions are summarized as follows:

(1) The minimum sample size for constructing soft probabilities is determined with the aid of statistical regularity hypotheses, which directly determines the balance between uncertainty (wider intervals) and precision (narrower intervals) in soft probabilities.
(2) A soft dependency-based procedure for attribute selection reduces the impact of redundant attributes on decisions, which can guarantee the optimality of results.
(3) The BDM-SP-SR model extends the classic NB model by reducing redundant attributes and handling changes in statistical regularities as new data appear, rather than merely using all

attributes and handling stochastic stability under large sample conditions.

At first glance the proposed method, as an integration of soft probabilities and the Bayesian decision model, involves a relatively complex computational process. However, it is primarily data-driven and easy to design implementation procedures, to obtain optimal results automatically. It can be applied to various domains such as supplier selection, rule extraction in information systems, and investment management. Despite its advantages, this study has several limitations. The selection of the bias threshold ϵ remains partially subjective, and the assumption of conditional independence may oversimplify complex attribute interdependencies. Additionally, while the model is computationally efficient, its reliance on discretization for continuous variables potentially induces information loss. In future research, one could focus on integrating more sophisticated dependency structures, such as Bayesian networks, into the soft probability framework to relax independence assumptions. Furthermore, developing data-driven, adaptive mechanisms for parameter selection remains a priority. From an empirical perspective, extending the BDM-SP-SR model to continuous domains and diverse applications, such as credit scoring and investment management with different dataset sizes, will further validate its robustness in non-stochastically stable environments.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The work is sponsored by Chongqing Social Science Planning Project, China (2023NDYB86) and Scientific Research Foundation for Advanced Talents of Chongqing Technology and Business University, China (2155011). I thank the referee for their constructive comments and assistance.

Conflict of interest

The author declares no competing interests to disclose.

References

1. L. Zadeh, Fuzzy sets, *Information and Control*, **8** (1965), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
2. K. Atanassov, *Intuitionistic fuzzy sets: theory and applications*, Berlin: Springer-Verlag, 1999. https://doi.org/10.1007/978-3-7908-1870-3
3. Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences*, **11** (1982), 341–356. https://doi.org/10.1007/BF01001956
4. Y. Yang, R. John, Grey sets and greyness, *Inform. Sciences*, **185** (2012), 249–264. https://doi.org//10.1016/j.ins.2011.09.029
5. D. Molodtsov, Soft set theory—first results, *Comput. Math. Appl.*, **37** (1999), 19–31. https://doi.org/10.1016/S0898-1221(99)00056-5

6. D. Molodtsov, *The theory of soft sets (Russian)*, Moscow: URSS Publishers, 2004.
7. J. Alcantud, A. Khameneh, G. Santos-García, M. Akram, A systematic literature review of soft set theory, *Neural Comput. Applic.*, **36** (2024), 8951–8975. https://doi.org/10.1007/s00521-024-09552-x
8. N. Çağman, S. Enginoğlu, Soft set theory and uni-int decision making, *Eur. J. Oper. Res.*, **207** (2010), 848–855. https://doi.org/10.1016/j.ejor.2010.05.004
9. G. Santos-García, J. Alcantud, Ranked soft sets, *Expert Syst.*, **40** (2023), e13231. https://doi.org/10.1111/exsy.13231
10. Z. Kong, Q. Lu, L. Wang, G. Guo, A simplified approach for data filling in incomplete soft sets, *Expert Syst. Appl.*, **213** (2023), 119248. https://doi.org/10.1016/j.eswa.2022.119248
11. Suhirman, T. Herawan, I. Yanto, J. Zain, Q. Hongwu, Z. Abdullah, A soft set approach for clustering student assessment datasets, *J. Comput. Theor. Nanos.*, **12** (2015), 5928–5939. https://doi.org/10.1166/jctn.2015.4738
12. X. Sun, A new method of fault diagnosis with soft set theory, *Advanced Materials Research*, **383–390** (2012), 7470–7474. https://doi.org/10.4028/www.scientific.net/AMR.383-390.7470
13. W. Xu, Z. Xiao, X. Dang, D. Yang, X. Yang, Financial ratio selection for business failure prediction using soft set theory, *Knowl.-Based Syst.*, **63** (2014), 59–67. https://doi.org/10.1016/j.knosys.2014.03.007
14. F. Feng, J. Cho, W. Pedrycz, H. Fujita, T. Herawan, Soft set based association rule mining, *Knowl.-Based Syst.*, **111** (2016), 268–282. https://doi.org/10.1016/j.knosys.2016.08.020
15. D. Kovkov, V. Kolbanov, D. Molodtsov, Soft sets theory-based optimization, *J. Comput. Syst. Sci. Int.*, **46** (2007), 872–880. https://doi.org/10.1134/S1064230707060032
16. K. Chang, A more general risk assessment methodology using a soft set-based ranking technique, *Soft Comput.*, **18** (2014), 169–183. https://doi.org/10.1007/s00500-013-1045-3
17. D. Molodtsov, Soft portfolio control, *Autom. Remote Control*, **72** (2011), 1705–1717. https://doi.org/10.1134/S000511791108008X
18. X. Feng, Z. Xiao, B. Zhong, J. Qiu, Y. Dong, Dynamic ensemble classification for credit scoring using soft probability, *Appl. Soft Comput.*, **65** (2018), 139–151. https://doi.org/10.1016/j.asoc.2018.01.021
19. Y. Zou, Bayesian decision making under soft probabilities, *Journal of Intelligent & Fuzzy Systems*, **44** (2023), 10661–10673. https://doi.org/10.3233/JIFS-223020
20. E. Karni, Bayesian decision theory with action-dependent probabilities and risk attitudes, *Econ. Theory*, **53** (2013), 335–356. https://doi.org/10.1007/s00199-012-0692-4
21. Y. Guo, Research on innovation risk management based on Bayesian risk decision-making, *International Journal of Business Administration*, **3** (2012), 21–30. https://doi.org/10.5430/ijba.v3n1p21
22. Y. Wang, I. Hussein, Bayesian-based decision-making for object search and classification, *IEEE Trans. Contr. Syst. Tech.*, **19** (2011), 1639–1647. https://doi.org/10.1109/TCST.2010.2087760

23. L. Filstroff, I. Sundin, P. Mikkola, A. Tiulpin, J. KylmÃĪoja, S. Kaski, Targeted active learning for bayesian decision-making, arXiv: 2106.04193. https://doi.org/10.48550/arXiv.2106.04193

24. V. Verma, A. Mishra, R. Narang, Application of Bayesian analysis in medical diagnosis, *Journal of the Practice of Cardiovascular Sciences*, **5** (2019), 136–141. https://doi.org/10.4103/jpcs.jpcs_51_19

25. C. Torres, R. Jones, F. Boelter, J. Poole, L. Dell, P. Harper, A model to systematically employ professional judgment in the Bayesian decision analysis for a semiconductor industry exposure assessment, *J. Occup. Environ. Hyg.*, **11** (2014), 343–353. https://doi.org/10.1080/15459624.2013.866713

26. R. Yager, V. Kreinovich, Decision making under interval probabilities, *Int. J. Approx. Reason.*, **22** (1999), 195–215. https://doi.org/10.1016/S0888-613X(99)00028-6

27. P. Guo, H. Tanaka, Decision making with interval probabilities, *Eur. J. Oper. Res.*, **203** (2010), 444–454. https://doi.org/10.1016/j.ejor.2009.07.020

28. M. Beer, Fuzzy probability theory, In: *Encyclopedia of complexity and systems science*, Berlin: Springer, 2021, 1–25. https://doi.org/10.1007/978-3-642-27737-5_237-2

29. D. Molodtsov, Soft probability of large deviations, *Advances in Systems Science and Applications*, **13** (2013), 53–67.

30. J. Smith, *Bayesian decision analysis: principles and practice*, Cambridge: Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511779237

31. K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, *Int. J. Approx. Reason.*, **24** (2000), 149–170. https://doi.org/10.1016/S0888-613X(00)00032-3

32. Z. Xu, Q. Da, Possibility degree method for ranking interval numbers and its application (Chinese), *Journal of Systems Engineering*, **18** (2003), 67–70.