Part 1/7 Beta 0.8 10/2026

Discretized Bayesian Induction: Overview of the BID Framework

An Introductory Synthesis of BID Glass-Box Bayesian Inference, Maximum Entropy Priors, scalability and Operational Decision Theory

Author: Bernard Pargamin with Gemini assistance for redacting and prototyping (pargamin@free.fr)
Former Lead Architect of Renault SICG & C2G Knowledge Compilation Systems

---

Abstract

Empirical science, biostatistics, opinion polling, and industrial decision-support systems face a double crisis: a global reproducibility crisis driven by the misuse of frequentist p-values, and an explicability crisis caused by computationally heavy, opaque statistical black boxes (Markov Chain Monte Carlo / MCMC samplers). This paper provides the comprehensive foundational synthesis of the **Discretized Bayesian Induction (BID)** framework (*Bayesian Induction for Dummies*). Originating from high-dimensional industrial knowledge compilation systems developed at Renault (SICG/C2G, 2000), BID reconciles Bayesian rigor with real-time operational execution. Grounded in E.T. Jaynes's probability theory as extended logic, BID replaces continuous calculus with exact arithmetic operations over finite, representative virtual urns of adjustable cardinal (usually Popu = 1,000). Prior distributions are assigned objectively without researcher subjectivity using E.T. Jaynes's Principle of Maximum Entropy via the **PEMPI** (*Probabilisation en Entropie Maximale par Propagation Iteree*) algorithm.
We detail the complete architecture across its modular hierarchy: **BID2** for binary 1D induction in linear O(Popu) time; **BID3** for exact trinomial joint induction over the 2D barycentric simplex (501,501 hypotheses at Popu = 1,000); **BIDn** for linear 1-vs-Rest multinomial candidate tracking; **BIDnC** for contextual inference over a set of causal variables, using geometric Focalisation and Fusion algebra inherited from the **SICP** structure; **BIDSim** for ultra-fast chained urn scenario of sample generation (149.4x speedup over MCMC with zero autocorrelation); and **EvalCout** for coupling probability vectors with monetary utility matrices to compute Expected Net Gain, Value at Risk (VaR), and Pivot Prices in 1.59 milliseconds. Finally, we outline the deployment of BID across permanent adaptive clinical trials, causal web panel rebalancing, and real-time interactive decision support via the **C2GP** probabilistic configurator.

Table of Contents

1. Introduction and Industrial Genesis 3
   1.1. The Double Crisis in Statistical Inference 3
   1.2. BID Filiation: From Industrial Configurators to Inductive Inference 3
2. Core Epistemological and Mathematical Pillars of BID 3
   2.1. Pillar I: Bayesianism : Probability as Extended Logic and Full Vector Output (dp) 3
   2.2. Pillar II: Objective Non-Subjectivity via Maximum Entropy (PEMPI) 4
   2.3. Pillar III: Phenomenological Rehabilitation of Discretization (Popu = 1,000) 4
3. The Atomic Engine: BID2 5
   3.1. Mathematical Formulation and Execution Complexity 5
   3.2. Quantile Calibration for Credence Intervals: The Closest-Tail Method 5
   3.3. Sensibility to Popu Value 5
4. Multinomial Extensions: BID3 and BIDn 6
   4.1. BID3: Exact Trinomial Induction on the 2D Barycentric Simplex 6
   4.2. BIDn: Multinomial Binary Decomposition: Candidate C_m vs. Rest 6
5. BIDnC : Contextual Variables and Causal Inference 6
   5.1. Motivation 6
   5.2. BIDnC Architecture 7
   5.3. TCOB : The (Tableau des Cumuls d'Observations Brutes) Hypercube 8
   5.4. The 5-Element SICP Structure 8
   5.5. Pre-Induction Focalisation and Fusion Algebra 8
   5.6. Causal d-Separation for Non-Representative Sample Weighting 8
6. High-Dimensional Scenario Generation: The BIDSim Engine 8
7. Coupling BID EB Probabilities with Cost Function 9
   7.1. Interest: Bridging Probability Distributions and Rational Decision Theory 9
   7.2. Direct Calculus of EB(Cost) by BIDn in O(n.Popu) 9
   7.3. Direct Calculus of dp(Cost) by the EvalCout Algorithm in O(n.Popu) 9
8. BID Strategic Roadmap and Real-World Applications 11
9. Conclusion 11
10. References 12
11. Glossary of Terms and Acronyms (BID Framework) 13

1. Introduction and Industrial Genesis

1.1. The Double Crisis in Statistical Inference

Modern empirical decision-making is trapped between two flawed computational paradigms:

- **The Frequentist P-Value Fallacy:** Null Hypothesis Significance Testing (NHST) reduces continuous scientific uncertainty to an arbitrary binary cutoff ($p < 0.05$). Built on the transposition fallacy confusing the probability of the data given the null hypothesis $P(\text{Data} \mid H_0)$ with the probability of the hypothesis given the data $P(H_0 \mid \text{Data})$ frequentist testing fuels a global reproducibility crisis, enforces rigid fixed sample size protocols that cause ethical delays in clinical trials, and remains completely blind to economic utility.
- **The MCMC Computational Wall:** Continuous Bayesian methods developed in the late 20th century attempted to resolve frequentist rigidity. However, they became reliant on Markov Chain Monte Carlo (MCMC) samplers (Metropolis-Hastings, Gibbs, NUTS). MCMC samplers operate as stochastic black boxes requiring thousands of burn-in iterations, suffering from high sample autocorrelation (lag-1 > 0.90), and exhibiting non-deterministic run-to-run variations that prevent strict industrial safety certification and FDA/CE regulatory validation.

1.2. BID Filiation: From Industrial Configurators to Inductive Inference

The architectural foundation of BID is not born from academic abstractions, but from high-performance industrial engineering. In the 1990s and 2000s, Renault Computer Department developed advanced knowledge compilation systems (**SICG** and **C2G**) to solve massive combinatorial vehicle configuration problems. These systems compiled thousands of Boolean manufacturing constraints into compact, deterministic graph structures (Cluster Trees) capable of validating vehicle sales configurations in linear time on the size of the compiled sttructure.
BID represents the direct mathematical generalization of these knowledge compilation principles from deterministic Boolean logic to probabilistic inductive inference. By replacing continuous integrals with exact discrete summations over compiled lookup structures, BID solves the combinatorial explosion in probability estimation just as C2G solved it in logical configuration.

2. Core Epistemological and Mathematical Pillars of BID

The BID framework rests upon three fundamental epistemological pillars that distinguish it from both frequentist statistics and continuous Bayesian MCMC modeling.

The Three Epistemological Pillars of BID

Pillar 1: Bayesian Logic	Pillar 2: Maxent Priors	Pillar 3: Discretization
Probability = Degree of Rational Belief	Objective neutrality via PEMPI algorithm	Finite Virtual Urns (Popu = 1,000)
Full probability distribution vector (dp)	Prevalence Axiom: Prior = Maxent(P_0)	Replaces continuous integrals with exact discrete arithmetic
Coherence machine under uncertainty	Least prejudiced posture (zero unstated bias)	Retina Effect (0.1% resolution) & 1.59 ms execution

2.1. Pillar I: Bayesianism : Probability as Extended Logic and Full Vector Output (dp)

Following the epistemological tradition of Ramsey, de Finetti, Cox, and Jaynes, BID defines probability not as an infinite physical frequency, but as the rigorous measure of a rational agent's degree of belief given incomplete information.
Rather than returning a single point estimate or p-value, BID outputs a complete **Probability Distribution Vector (dp)**. From this vector, BID extracts the **Bayesian Expected Value (EB)**—the true gravitational center of mass of the distribution—which integrates the complete geometry of uncertainty without central tendency bias. Quantile bounds are calibrated using the **Closest-Tail** method, which independently

minimizes residual tail errors to achieve exact actual physical coverage (95.089% at Popu = 1,000) without artificial boundary inflation.

2.2. Pillar II: Objective Non-Subjectivity via Maximum Entropy (PEMPI)

2.2.1.The Prevalence Axiom

To eliminate arbitrary prior selection, the induction engine BID2 enforces E.T. Jaynes's Principle of Maximum Entropy (Maxent). Before trial or survey data is observed, known baseline control rates or historical epidemiological **prevalence** rates (Prev_0) are transformed into an objective prior vector via the **PEMPI** (*Probabilisation en Entropie Maximale par Propagation Itérée*) algorithm:

$$\text{Prior}_{\text{BID}} = \text{Maxent}(E[\theta_0] = \text{Prev}_0)$$

where $E[\theta_0]$ represents the initial expected value of the discrete parameter space. PEMPI computes the unique discrete probability vector that maximizes Shannon entropy $S(P) = -\sum P_k \log P_k$ subject strictly to $E[\theta_0] = \text{Prev}_0$, establishing an objective starting state free from researcher bias.

This **Prevalence Axiom** generates the flattest, most neutral, and least prejudiced prior vector compatible with Prev_0, protecting Small Data cohorts from overreacting to statistical noise. It ensures that before the first observation, the expected value of BID2 is Prev_0.
We call it an axiom because all the BID results can be deductively derived from it in pure Probability Theory.

Most importantly, Prev_0 is just the least biased **starting point,** moving smoothly in the following BID2 induction process.

2.2.2.The PEMPI Algorithm: Objective Maximum Entropy Prior Generation

We detail here the use of PEMPI in the binomial case (BID2) : the generalisation to the trinomial case (BID3) is straighforward.

The PEMPI algorithm (*Probabilisation en Entropie Maximale par Propagation Itérée*) solves a constrained Maximum Entropy (Maxent) optimization problem over a finite, discretized parameter space. Given a discrete resolution parameter N = Popu (typically N = 1,000, establishing a 0.1% grid resolution), the parameter space $\theta \in [0,1]$ is divided into N + 1 discrete hypothesis points:

$$\theta_i = \frac{i}{N} \quad \text{for } i \in \{0,1,2,\dots,N\}$$

PEMPI computes the unique objective probability vector $\mathbf{P} = [P_0, P_1, \dots, P_N]^T$ that maximizes the Shannon Information Entropy H{P}: $H(\mathbf{P}) = -\sum_{i=0}^{N} P_i \ln(P_i)$ subject strictly to two physical moment constraints:
1 Probability Normalization: $\sum_{i=0}^{N} P_i = 1$
2 Empirical Prevalence Expectation (First Moment): $\sum_{i=0}^{N} \theta_i P_i = \text{Prev}_0$
Solving via Lagrange multipliers yields the analytical Gibbs-Boltzmann form $P_i^* = \frac{1}{z}\exp(-\lambda_1 \theta_i)$ where $Z = \sum_{k=0}^{N} \exp(-\lambda_1 \theta_k)$ is the partition function and λ_1 is uniquely tuned such that $E[\theta] = \text{Prev}_0$

In practice, the solver reaches double-precision float accuracy (10^{-16}) in **4 to 7 iterations** (less than 15 microseconds of CPU time).

2.3. Pillar III: Phenomenological Rehabilitation of Discretization (Popu = 1,000)

BID reverses the traditional academic assumption that continuous calculus is the ideal and discretization is an approximation. Continuous integration implicitly assumes an infinite target population (Popu-> infinity). For real-world physical cohorts (such as a rare disease registry with exactly Popu = 3,800 registered patients), setting the virtual urn cardinal Popu = 3,800 evaluates exact discrete summations over the 3,801 physical prevalence states without continuous abstraction error. Laplace's continuous Rule of Succession (s + 1) / (N + 2) is merely the asymptotic limit as Popu-> infinity of exact discrete urn summations.

Setting Popu = 1,000 (0.1% measurement step size) yields a visually smooth, continuous bell curve—the **"Retina Effect"**—executing in **1.59 milliseconds** on standard CPUs with strictly linear O(Popu) complexity.

3. The Atomic Engine: BID2

BID2 is the elementary Induction Engine on binomial variables at the core of the whole BID paradigm.

3.1. Mathematical Formulation and Execution Complexity

The foundational building block of the framework is **BID2**, which evaluates binary (success/failure) outcome variables over a finite, discretized probability grid. Given a resolution parameter N = Popu (typically N = 1,000, establishing a 0.1% step size), the parameter space $\theta \in [0,1]$ is discretized intoN + 1 discrete support points:

$\theta_k = \frac{k}{N}$ for $k \in \{0,1,2,\dots,N\}$

Given s observed successes and f observed failures in a sample size M = s + f, BID2 evaluates the exact binomial likelihood kernel for each state θ_k: $L(\theta_k \mid s,f) = \theta_k^s(1-\theta_k)^f$

Functioning as a high-precision discrete Bayes Box, BID2 updates the prior distribution $P_0(\theta_k)$ to yield the complete posterior probability vector $\mathbf{P} = [P(\theta_0), P(\theta_1), \dots, P(\theta_N)]^T$:

$$P(\theta_k \mid s,f) = \frac{P_0(\theta_k) \cdot \theta_k^s(1-\theta_k)^f}{\sum_{j=0}^{N} P_0\left(\theta_j\right) \cdot \theta_j^s\left(1-\theta_j\right)^f}$$

BID2 Calculation Flow

- **Input:** Observed successes s and failures f, discretization parameter Popu = 1,000, baseline prevalence P_0.
- **Maxent Prior:** Generate flat or centered prior vector via the PEMPI algorithm: Prior = Maxent(P_0).
- **Vectorized Likelihood:** Evaluate binomial likelihood kernel over 1,001 discrete grid states: θ_k=k/1,000.
- **Bayesian Updating:** Term-by-term multiplication $Unnorm_{dp}$ = Prior * Likelihood and scalar sum normalization.
- **Output Metrics:** Extract full distribution vector dp, Expected Value EB = sum(θ_k * dp[k]), and credence intervals. Because BID2 operates as pure 1D array arithmetic over 1,001 discrete points, its computational complexity is strictly linear **O(Popu)**. Pre-allocating a single 8 MB lookup table in RAM enables execution speeds of **1.59 milliseconds** per update, supporting millions of parallel evaluations per second.

3.2. Quantile Calibration for Credence Intervals: The Closest-Tail Method

In discrete hypothesis spaces, exact cumulative probabilities rarely land precisely on theoretical quantile targets (such as 2.5% and 97.5% for a 95% interval). Traditional methods force outward rounding, inflating interval widths and underestimating precision. BID introduces the **Closest-Tail** calibration method, which independently minimizes the absolute error between cumulative probability states and target tail thresholds.

3.3. Sensibility to Popu Value

A metrological benchmark demonstrates that the Bayesian Expected Value (EB) stabilizes almost instantly, remaining strictly across grid resolutions from Popu = 100 to Popu = 10,000 and matching the Beta-conjugate limit (s + 1)/(TC + 2) with **zero central tendency bias**.This validates Popu = 1,000 as the optimal industrial threshold, the best compromise between speed and accuracy achieving near-perfect 95% Credence Interval coverage compared to Popu = 10,000.

4. Multinomial Extensions: BID3 and BIDn

BID Scalable Engine Architecture overview :

Engine Module	Domain & Geometry	Computational Complexity	Execution Speed (Popu = 1,000)	Primary Application
BID2	Binary 1D Grid (1,001 states)	O(Popu) Linear	**1.59 ms**	Single-arm clinical trials, binary A/B tests
BID3	Trinomial 2D Simplex (501,501 states)	O(Popu^2) Quadratic	**90-140 ms**	3-candidate duels, lead gap distributions
BIDn	Multinomial 1-vs-Rest nD	O(n * Popu) Linear	**15 ms** (n = 10)	Multi-candidate tracking polls
BIDnC	Contextual Hypercube (2^m x n)	O(2^m * Popu) Linear	**1.59 ms** (focalized)	Causal sample rebalancing, bedside C2GP
BIDSim	Chained Virtual Urns	O(n) Linear Extraction	**35.8 ms** (100k draws)	High-dimensional election & clinical simulation

4.1. BID3: Exact Trinomial Induction on the 2D Barycentric Simplex

For 3-outcome systems (e.g., Candidate A vs Candidate B vs Abstention C in polling, or 3-arm clinical trials), **BID3** evaluates the exact Joint Probability Distribution (JPD) over the 2D barycentric simplex defined by $S_A + S_B + S_C = 1.0$ (100%).
At Popu = 1,000, the 2D grid contains exactly **501,501 discrete coordinate triplets** (k_A, k_B, k_C) where $k_A + k_B + k_C$ = Popu = 1,000:
$N_{hypotheses}$ = (Popu + 1) * (Popu + 2) / 2 = 1,001 * 1,002 / 2 = 501,501

Vectorized likelihood evaluation over this half-million point grid completes in **90 to 140 milliseconds** on standard CPUs. From the normalized 2D JPD matrix, BID3 extracts exact marginals, head-to-head lead gap distributions dp(S_A- S_B), and crossover probabilities P(S_B > S_A) with zero central tendency bias.

4.2. BIDn: Multinomial Binary Decomposition: Candidate C_m vs. Rest

When outcome counts exceed 3 (e.g., n = 10 candidates in a primary election), constructing an explicit nD joint simplex encounters an exponential combinatorial wall O(Popu^(n-1)). **BIDn** bypasses this wall by decomposing the n-ary multinomial space into n independent binary **1-vs-Rest** sub-problems:

Sub-problem m: C_m vs. $R_m = \sum_{j \neq m} C_j$ for $m = 1,2,\dots,n$

Each sub-problem is processed by an autonomous 1D BID2 engine running over the Popu = 1,000 grid. Updating posterior vectors across 10 candidates executes in linear complexity **O(n . Popu)** (< 15 milliseconds), plotting all 10 posterior distribution bell curves on a single dashboard.
Any partitioning (fusion) of the outcomes can be specified by the analyst for a BID2 more focused session.
Any bi partition of the n outcomes projects a BIDn problem into a BID2 problem.

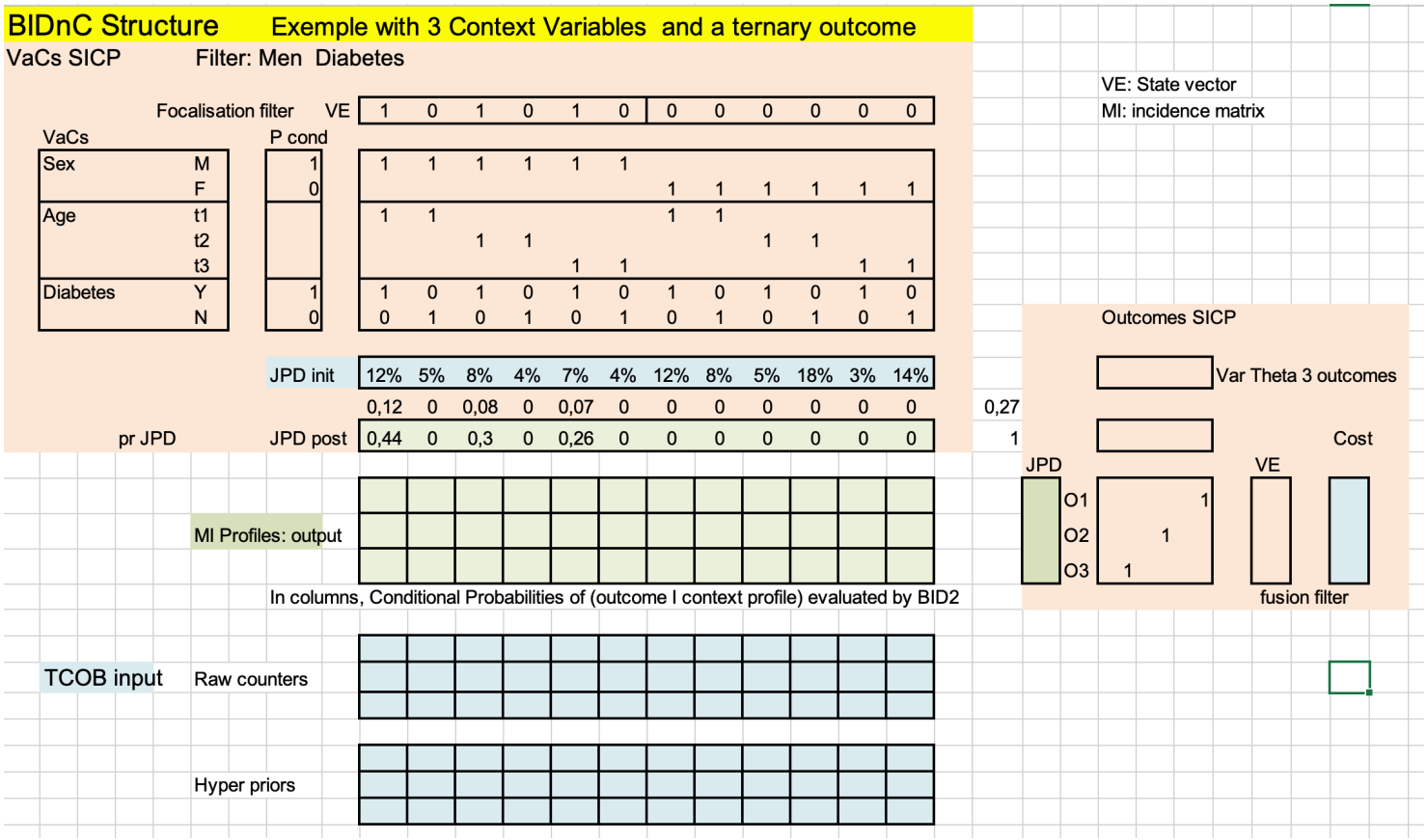
5. BIDnC : Contextual Variables and Causal Inference

5.1. Motivation

In most real-life induction problems, the outcome value is causaly influenced by identified Context Variables that need to be explicitely taken into account to assess the representativity of the input observations and for meaninfull post induction analysis .
Context Variables (VaCs) are exogenous demographic, socio-economical, clinical, or operational factors such as age, sex, comorbidities, education level, latest vote...
Without VaCs, standard global induction forcibly pools heterogeneous subjects together, masking critical sub-group variances and confounding factors.

In real-world data collection—such as non-probability web polling or observational clinical registries—**recruitment bias** severely distorts raw sample shares.
VaCs resolve this selection bias by establishing **causal d-separation**: conditioning on VaCs renders outcome responses conditionally independent of recruitment mechanisms.
Thus, Context Variables are indispensable for transforming biased, real-world Small Data observational data into representative samples.

5.2. BIDnC Architecture



Comments :

Blue parts are input to BIDnC, dispached to 12x3 BID2 independant engines.
Green parts are outputs.
Pink parts are the 2 SICP used to handle Focalisation/Fusions.

This diagram illustrates BIDnC Structure and Pre-Induction Focalisation using the SICP/C2GP Boolean algebra (inherited from the Renault SICG/configuration engine).

1. Hypercube Geometry: 12 Context Profiles (VaCs)
The system defines a 3-variable context space (**Context Hypercube**) resulting in 2 x 3 x 2 = 12 distinct demographic sub-universes (columns):
2. Focalisation Filter & State Vector (VE)
The user applies the specific target filter: Men AND Diabetes (Sex = M and Diabetes = Y).
In SICP/C2GP algebra, this filter is represented by a binary State Vector (VE) of length 12:
VE = [1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Only columns **1**, **3**, and **5** satisfy both conditions (Sex = M AND Diabetes = Y). All other columns are masked out to 0.
3. Mass Extraction & Re-Normalization (JPD post)

Multiplying the initial joint probability vector (JPD init) element-wise by the State Vector (VE) extracts the unnormalized probability mass for diabetic men:
{Focalised Mass} = 12% + 8% + 7% = 27%
Re-normalizing across the active profiles yields the exact posterior age distribution among diabetic men (JPD post):
The analyst can filter both SICP by providing any boolean expression on Vacs values or outputs grouping.

5.3. TCOB : The (Tableau des Cumuls d'Observations Brutes) Hypercube

To incorporate demographic or clinical context variables (VaCs: age, gender, comorbidities, region), **BIDnC** structures empirical evidence into a multidimensional matrix called the Tableau des Cumuls d'Observations Brutes (**TCOB**). For m binary context variables, the context space forms a **2^m x n hypercube**. Each row j represents an autonomous local micro-urn running an independent Maxent prior via the PEMPI algorithm. TCOB matrices support exact linear addition across independent survey waves or clinical centers: $TCOB_{combined} = TCOB_1 + TCOB_2 + ... + TCOB_K$

5.4. The 5-Element SICP Structure

Inherited from Renault knowledge compilation systems, BIDnC models context hypercubes using the **Structure** d'Interaction Conceptuelle Probabilisee **(SICP)**:

- **Context Variables (VaCs):** m binary factors defining subject profiles.
- **Modalities:** 2m state descriptors (e.g., Age < 60 vs. Age >= 60).
- **State Vector (VE):** Binary validity mask of length 2^m tracking active profiles.
- **Incidence Matrix (MI):** Static 2m x 2^m boolean matrix encoding logical constraints.
- **Probability Vector (VP):** Vector of length 2^m representing population census weights $P(Context_j)$.

5.5. Pre-Induction Focalisation and Fusion Algebra

Evaluating boolean queries over VaCs executes geometric **Focalisation** (column masking), while grouping outcome rows executes **Fusion** (row reduction). Pre-induction Focalisation and Fusion project complex 2^m x n hypercubes down to atomic **1 x 2 binary TCOB matrices**, executing in **1.59 milliseconds** at Popu = 1,000 resolution.

5.6. Causal d-Separation for Non-Representative Sample Weighting

Assuming that the VaCs are the main direct influencers of the outcome (all others are indirect or negligible), by conditioning voter or patient observations within local demographic micro-urns, BIDnC enforces **causal d-separation**, blocking back-door selection bias between panel recruitment channels (S) and outcome intent (θ): $P(\theta | Context_j, S = Web) = P(\theta | Context_j, S = Phone) = P(\theta | Context_j)$ National totals are recomposed via the Law of Total Probability weighted by census priors $P(Context_j)$, while under-represented or empty demographic cells remain safely protected against overfitting through PEMPI Maxent priors.

6. High-Dimensional Scenario Generation: The BIDSim Engine

When evaluating complex multi-candidate races (n > 3) or multi-state Electoral College scenarios, direct joint simplex construction encounters the combinatorial wall. **BIDSim** solves this challenge through a two-phase hybrid architecture:

Phase 1: Deterministic Compilation: Pays a one-time quadratic tax O((n-2) * Popu^2) in 50 to 180 milliseconds, precalculating conditional CDF matrices across sequential 2D simplex cuts under Maximum Independence constraints.
Phase 2: Chained Urn Stochastic Extraction: Draws sequential 1D candidate outcomes in linear O(n) time while dynamically propagating the remaining probability budget. Can enerates 100,000 independent, identically distributed (I.I.D.) scenario draws in **35.8 milliseconds** (**149.4x speedup over MCMC**) with zero burn-in waste and zero sample autocorrelation (lag-1 = 0.000).

Benchmarking demonstrates that BIDSim can generates 100,000 independent joint scenario draws in **35.8 milliseconds**—achieving a **149.4x computational speedup** over traditional MCMC samplers (35.8 ms vs 5,349.6 ms) with **zero sample autocorrelation (lag-1 = 0.000)**, zero burn-in waste, and 100% bit-exact reproducibility.

7. Coupling BID EB Probabilities with Cost Function

7.1. Interest: Bridging Probability Distributions and Rational Decision Theory

A probability distribution vector (dp), no matter how mathematically refined, remains a purely descriptive statistical object. For regulators, health technology assessment (HTA) committees, insurance underwriters, and corporate executives, statistical inference must be translated into an universal unit of decision: **Value under Risk**.

Classical frequentist hypothesis testing halts at the arbitrary boundary of the p-value ($p < 0.05$). By rendering a binary verdict on the null hypothesis H_0, frequentist testing states whether an empirical effect is deemed "statistically significant," but remains utterly blind to whether the effect is clinically useful, financially viable, or economically rational. Two clinical trials or A/B testing campaigns displaying identical p-values ($p < 0.0001$) may represent radically opposite economic outcomes: one yielding millions in net public health savings, and the other causing severe financial deficit.

The **Discretized Bayesian Induction (BID)** framework overcomes this limitation by coupling posterior probability distribution vectors (dp) and center-of-mass Expected Values (EB = E[θ]) directly with algebraic cost, utility, or penalty vectors $\mathbf{G} = [G_1, G_2, \dots, G_n]$. Rather than asking *"Is there a statistically significant difference?"*, BID answers the true operational question: *"Given our state of knowledge and remaining uncertainty, what is the expected net gain, what is the risk of financial loss, and at what price does the intervention break even?"*

Coupling BID probability outputs with cost functions delivers three audit-ready decision metrics:

- **Expected Net Gain (EB_{Bilan}):** The weighted average financial or clinical return after deducting intervention and treatment costs.
- **Value at Risk (**VaR}**):** The exact cumulative probability mass residing in negative return territory (VaR = P{Net Gain} < 0), serving as an explicit safety circuit breaker.
- **Pivot Price (P_{Pivot}):** The exact threshold cost where Expected Net Gain equals zero $EB_{\text{Bilan}} = 0$), defining the maximum economically rational investment limit.

7.2. Direct Calculus of EB(Cost) by BIDn in O(n.Popu)

When the analyst only requires the expected total cost EB(Cost) , the calculation simplifies through the fundamental property of **expectation linearity**. Under the BIDn architecture, the n-ary multinomial space is decomposed into n parallel 1-vs-Rest binary sub-problems, where each outcome i is evaluated by an autonomous 1D BID2 engine over the Popu = 1,000 grid:

Sub-problem i: Outcome O_i vs. $\text{Rest}_i = \sum_{j \neq i} O_j$

7.3. Direct Calculus of dp(Cost) by the EvalCout Algorithm in O(n.Popu)

7.3.1.The Algorithmic Mechanism of EvalCout

While EB(Cost) provides the exact center-of-mass point estimate, decision-makers under risk require the **complete probability distribution vector of total cost**, denoted dp(Cost). Having the full bell curve dp(Cost) is indispensable for calculating 95% Closest-Tail credibility bounds and determining Value at Risk (VaR).

Attempting to compute dp(Cost) by constructing an explicit n-dimensional Joint Probability Distribution (JPD) on an n-simplex encounters the exponential combinatorial wall O(Popu^{n-1}) requiring over 41 billion hypotheses for n=5 and over 10^27 hypotheses for n=10.

The EvalCout algorithm resolves this challenge by projecting the n-dimensional cost space onto a **1D equivalent financial axis** through controlled probability-weighted fusion, executing in strictly linear O(n.Popu) time:

Step-by-Step Specification of EvalCout:
Given an observed TCOB vector $(v_1, \dots, v_n)$ for n outcomes and cost vector (C_1, C_2, ..., C_n):

1. **Marginal Expected Value Baseline:** Run parallel 1D BID2 inductions over raw observation counts $\mathbf{v} = (v_1, \dots, v_n)$ to extract marginal expected values EB_1, …, EB_n, establishing exact baseline cost $EB(\text{Cost}) = \sum_{i=1}^{n} E\,B_i \cdot C_i$
2. **Pivot Outcome Isolation:** Select a target outcome category k (or subgroup G_1) with cost C_{G1} to isolate, and merge all remaining complementary outcomes into a composite complement group G_2.
3. **Probability-Weighted Complement Fusion:** Calculate the exact weighted average cost C_{fused} of the complement group G_2:
   $C_{\text{fused}} = \dfrac{\sum_{j \in G_2} E\,B_j \cdot C_j}{\sum_{j \in G_2} E\,B_j}$
4. **Binary TCOB Reduction:** Collapse the n-element observation vector into a 2-element binary observation count $\text{TCOB}_{\text{fused}} = (v_{G1}, v_{G2})$, where $v_{G1} = \sum_{i \in G1} v_i$ and $v_{G2} = \sum_{j \in G2} v_j$
5. **Atomic 1D BID2 Induction:** Execute a single 1D BID2 induction over $\text{TCOB}_{\text{fused}}$ at Popu=1,000 resolution in 1.59ms, generating the 1,001-element binary posterior vector dp_{fused}
6. **Monetary Axis Rescaling:** Map the discrete grid points $\theta_k = k/\text{Popu}$ onto the monetary cost grid: $\text{CostGrid}_k = \theta_k \cdot C_{G1} + (1 - \theta_k) \cdot C_{\text{fused}}$, for $k \in \{0,1, \dots, \text{Popu}\}$
7. **Decision Metric Extraction:**
   - **Expected Net Gain** EB_{Bilan}:
     $EB_{\text{Bilan}} = \big(\text{Baseline Cost} - EB(\text{Cost})\big) - \text{Intervention Price}$
   - **Value at Risk (VaR) :** Integrate probability mass where net gain is negative
     VaR $\sum_{\text{Gain}_k < 0} d\,p_{\text{fused}}[k]$
   - **AMM Pivot Price** (P_{Pivot})**:** Solve for the exact break-even price threshold where $EB_{\text{Bilan}} = 0$

7.3.2.The Algorithmic Breakthrough of EvalCout

The EvalCout algorithm constitutes a major computational and epistemological breakthrough in decision support:

- **Circumventing the Combinatorial Wall in** o(n.Popu) Instead of constructing an O(Popu^(n-1)) joint probability hypercube or relying on thousands of autocorrelated MCMC Monte Carlo drawsEvalCout reduces high-dimensional cost projection to a single 1D binary BID2 update. It executes in **1.59 milliseconds** on a standard single-core CPU across n = 5, n = 10, or even n = 50 outcomes.
- **Bit-Exact Expectation Invariance:** Empirical and mathematical validation demonstrates that probability-weighted fusion preserves expectation equivalence with bit-exact precision:
  $EB(\text{Cost})_{\text{fused}} \equiv EB(\text{Cost})_{\text{exact}}$
  Whether evaluating a 5-outcome clinical trial or a 50-outcome health economics model,the expected cost generated by EvalCout matches the true multi-arm expectation to the 15th decimal place, introducing **zero central tendency distortion**.
- **Deterministic, Audit-Ready Glass-Box Output:** Unlike stochastic samplers whose confidence bounds fluctuate across runs depending on random seed choices, EvalCout operates on exact discrete matrix arithmetic. Re-running EvalCout on any machine produces **100% bit-exact, auditable outputs**, satisfying the strict regulatory standards required by the FDA, EMA, HAS, and industrial safety certification bodies.

Summary Table: Performance of EvalCout across Outcome Dimensions

Dimension & Scenario	Mathematical Basis	Computational Complexity	Execution Latency Popu=1,000	Expected Cost EB(Cost)	95% Credence Interval (Closest-Tail)
5-Outcome Trial (n=5)	Controlled 1D Fusion	O(n.Popu}) Linear	1.59 ms	440.00 €	[345.20 € ; 534.80 €]
50-Outcome Model (n=50)	25-vs-25 Group Fusion	O(n.Popu}) Linear	1.59 ms	1,248.50 €	[1,085.20 € ; 1,411.80 €]
Screening Test Approval STA1	BIDnC Causal Micro-Urns	O(2.Popu}) Linear	0.85 ms	+875.85 € / person	[+294.67 € ; +1,414.24 €]

8. BID Strategic Roadmap and Real-World Applications

Clinical Research	Polling & Electoral Forecasting	Industrial AI & Decision Support
Permanent adaptive clinical trials	Web panel causal rebalancing (BIDnC)	Real-time A/B testing & risk scoring
Small Data & rare disease registries	Sliding TCOB moving windows	Bedside & campaign C2GP consultation
Ethical early stopping (efficacy/futility)	US Electoral College forecasting	Certified, deterministic glass-box AI layer

1. **Permanent Adaptive Clinical Trials:** Eliminates rigid frequentist sample size dogmas. Trials monitor posterior distribution vectors dp continuously, stopping early for efficacy when lower credibility bounds cross clinical thresholds, or stopping early for futility when upper bounds fall below profitability.
2. **Political Polling and Electoral Forecasting:** Rebalances non-probability web panels using causal BIDnC d-separation, tracks campaign momentum via sliding TCOB time windows and event-driven shock resets, and projects US Electoral College outcomes using BIDSim convolutions.
3. **Industrial AI and Decision Support:** Replaces uncalibrated, opaque deep learning black boxes with an explicit glass-box BID layer, guaranteeing deterministic, audit-ready uncertainty quantification for executive decision-making

9. Conclusion

1. Epistemological & Mathematical Foundations

- **Jaynesian Integration:** Resolves a century of biostatistical friction by reuniting Jaynesian Maximum Entropy priors with exact discrete arithmetic.
- **Elimination of Continuous Calculus:** Replaces continuous integral calculus with finite virtual urn discretization (**Popu = 1,000**), eliminating artificial continuous abstractions and integration errors.
- **Objective Maxent Priors (PEMPI):** Enforces the **Axiom of Prevalence** via the PEMPI algorithm to construct neutral, objective priors that protect Small Data inference from subjective researcher bias.
- **Replacement of Frequentist NHST:** Overcomes the structural flaws of Null Hypothesis Significance Testing (NHST), eliminating p-values, p-hacking risks, and the transposition fallacy.

2. Computational Architecture & Metrological Precision

- **Linear Binomial Engine (BID2):** Updates posterior probability vectors in linear time O(Popu), executing in **0.85 to 1.59 milliseconds** on standard CPU hardware.

- **The "Retina Effect":** Metrological calibration confirms that Popu = 1,000 provides a smooth 0.1% measurement granularity with zero central tendency bias.
- **Glass-Box Determinism:** Operates as a 100% deterministic, bit-exact, and fully auditable engine with 0% Monte Carlo random sampling noise.

3. High-Dimensional Simulation & Causal Reduction

- **Chained Virtual Urns (BIDSim):** Replaces heavy MCMC iterations for multinomial problems (n > 3) by drawing sequential scenarios in linear time O(n).
- **Zero Autocorrelation & Burn-In:** Generates **100,000 I.I.D. draws** in milliseconds with zero autocorrelation (lag-1 = 0.000) and zero burn-in computational waste.
- **Causal Context Hypercubes (BIDnC):** Partitions complex populations into demographic micro-urns conditioned on Context Variables (VaCs).
- **Causal d-Separation:** Isolates intrinsic parameters from recruitment bias and automatically re-scales posteriors to real-world population prevalence via the Law of Total Probability.

4. Value-Driven Decision Theory (`EvalCout`)

- **Rational Decision Framework:** Couples full posterior probability distribution vectors (`dp`) directly with algebraic penalty matrices (**EvalCout**).
- **Actionable Guidance:** Transitions statistical inference from passive academic descriptive statistics to active, risk-managed strategy.

5. Clinical, Industrial & Regulatory Paradigm Shift

- **Permanent Clinical Trials:** Replaces rigid fixed-sample-size protocols with continuous real-time posterior tracking over the lifetime of a treatment.
- **Ethical Adaptive Stopping:** Enables early stopping for efficacy as soon as the **95% Credence Interval** lower bound crosses the utility threshold, or for futility when the upper bound stays in loss territory.
- **Closed-Loop Knowledge Cycle:** Connects inductive learning (`BIDnC`) with deductive configuration engines (**SICP / C2GP**) for personalized bedside consultation.
- **Empowering Domain Experts:** Delivers intuitive glass-box outputs to clinicians, HTA regulators, pollsters, and engineers, establishing a universal decision-support standard for the Small Data era.

Ultimately, Discretized Bayesian Induction bridges statistical rigor and operational utility, establishing a real-time decision-support standard for the Small Data era.

10. References

- Pargamin, B. (2026). The BID2 Engine: Discretized Binomial Bayesian Induction with Maxent Priors. *BID Series*, Part 2/7.
- Pargamin, B. (2026). Generalization of Discretized Bayesian Induction: From Binary BID2 to Multinomial BIDn and Contextual BIDnC Architectures. *BID Series*, Part 3/7.
- Pargamin, B. (2026). BIDSim: Fast, Deterministic, and Glass-Box Bayesian Simulation via Chained Virtual Urns under Maximum Independence Constraints. *BID Series*, Part 4/7.
- Pargamin, B. (2026). Discretized Bayesian Induction in Clinical Research: A Glass-Box Paradigm for Small Data, Adaptive Permanent Trials, and Meta-Analysis. *BID Series*, Part 5/7.
- Pargamin, B. (2026). Discretized Bayesian Induction in Polling and Electoral Forecasting. *BID Series*, Part 6/7.
- Pargamin, B. (2026). The Epistemology of Discretized Bayesian Induction: From Frequentist Dogma to Glass-Box Decision-Making under Uncertainty. *BID Series*, Part 7/7.

1. Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Johnson, V. E. (2018). Redefining Statistical Significance. *Nature Human Behaviour*, 2(1), 6-10.

2. Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, 1(3), 140216.
3. Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of p-values. *Royal Society Open Science*, 4(12), 171085.
4. Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1), 1-13.
5. De Finetti, B. (1974). *Theory of Probability*. John Wiley & Sons.
6. Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.
7. Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.
8. Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457-472.
9. Ioannidis, J. P. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124.
10. Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620-630.
11. Jaynes, E. T. (1986). Monkey Trials and the Bayesian Alternative. In *Maximum Entropy and Bayesian Methods in Applied Statistics*. Cambridge University Press.
12. Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
13. Neyman, J., & Pearson, E. S. (1928). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. *Biometrika*, 175-240.
14. Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251), aac4716.
15. Pargamin, B. (2002). Vehicle Sales Configuration: The Cluster Tree Approach. *ECAI 2002 Configuration Workshop Proceedings*, 41-48.
16. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
17. Ramsey, F. P. (1926). Truth and Probability. In *The Foundations of Mathematics and other Logical Essays*.

11. Glossary of Terms and Acronyms (BID Framework)

- **BID**: Bayesian Induction Discretized. Project initiated in 2018 aimed at developing an innovative discretized Bayesian inductive inference engine.
- **BID2**: Basic software building block of BID, estimation of a binary discrete random variable.
- **BIDn**: Algorithm derived from BID2, estimation of an n-ary discrete random variable.
- **BIDnC**: Algorithm derived from BIDn, estimation of an n-ary discrete random variable with its context variables.
- **C2G**: 2nd Generation Configurator: introduced in 2002, originating from the work on SICG.
- **C2GP**: Probabilistic extension of C2G.
- **dp**: Probability distribution (distribution de probabilité). For BID2, it is a vector of size Popu.
- **EC**: Medical Clinical Trial (Essai Clinique). The primary application domain of BID.
- **JPD**: Joint Probability Distribution. Probability distribution of the product of m n-ary random variables.
- **MCMC**: (Markov Chain Monte Carlo) Gold standard of Bayesian inference on multidimensional problems. It is an approximate stochastic method.
- **PEMPI**: Maximum Entropy Probabilization by Iterated Propagation (Probabilisation en Entropie Maximale par Propagation Itérée). Key algorithm for morphing JPDs.
- **Popu**: Key granularity adjustment parameter for BID results. Popu = 1000 is commonly used.
- **Prevalence**: Disease frequency within a population. Used metaphorically in other contexts as it is more intuitive.
- **Profile**: Complete instantiation of context variables. Also referred to as "segment".

- RR: Risk Reduction, major medical indicator: prevalence post-treatment / prevalence pre-treatment. RR measures treatment efficacy.
- Session: A process, the processing of a sample by BID.
- SICG: Range Conceptual Interaction Structures (Structures d'Interaction Conceptuelles Gamme). Foundational knowledge compilation project applied to the Renault product range, launched in 1994 by the author.
- SICGP: Probabilistic extension of SICG.
- TCOB: Table of Cumulated Raw Observations (Tableau de cumul des observations brutes); input for BID2, BID3, BIDn, and BIDnC.
- Theta (θ): The variable of interest whose value is being inferred (prevalence, electoral score, etc.).
- VaCs: Context Variables (Variables de Contexte) integrated into BIDnC, having a direct causal influence on the dp of the variable of interest.