

Article

From a Hierarchical Dirichlet-Type Construction to the Informative Bayesian Double Bootstrap

Guadalupe Eunice Campirán García 

Business School, Tecnológico de Monterrey, Campus Estado de México, Atizapán de Zaragoza 52926, Mexico; eunice.campirán@tec.mx

Abstract

Efron's double bootstrap and hierarchical Bayesian nonparametric methods have largely developed along separate paths. This paper connects them and uses that connection to motivate a new resampling procedure. We show that a suitable two-level Dirichlet construction, with base measures matched to the data, can approach the classical double bootstrap when its concentration parameters become large, while the hierarchical Dirichlet process of Teh et al. does not share this limit. This distinction identifies the double bootstrap as the endpoint of a construction of hierarchical Dirichlet type—though not of the hierarchical Dirichlet process itself—and as the boundary of a broader family of two-level resampling methods. Moving away from that boundary leads to the Informative Bayesian Double Bootstrap (IBDB). The method introduces prior information at the first stage while keeping the second stage focused on calibration, as in the classical double bootstrap. We also establish finite-concentration bounds describing how the proposed construction differs from its classical counterpart and when it can move beyond the support of the observed data. In simulations against four competing methods, the IBDB performs best for tail-sensitive quantities and heavy-tailed settings, while it tends to over-cover simple location parameters. Similar patterns appear in the Danish fire-insurance and Siemens equity-loss examples. Its main advantage is improved calibration through interval repositioning rather than simply wider intervals. The gains are most relevant when sample information is limited.

Keywords: double bootstrap; Bayesian bootstrap; two-level Dirichlet resampling; hierarchical Dirichlet process; informative resampling; empirical Bayes; Kullback–Leibler divergence; coverage calibration; tail quantiles

MSC: 62F40; 62G09; 60G57; 62C12



Academic Editors: Nuttanan Wichitaksorn and Mengheng Li

Received: 29 July 2026

Revised: 23 August 2026

Accepted: 26 August 2026

Published: 4 September 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The bootstrap and the Dirichlet process are two of the most influential devices for quantifying uncertainty in modern statistics. Their literatures share a reliance on empirical support and random discrete distributions, yet the relationship between hierarchical Dirichlet resampling and the classical double bootstrap has remained unexamined. That relationship is the subject of this paper, and the hierarchical Dirichlet process of Teh et al. [1] enters it as the natural candidate that does not quite work: understanding why it fails is what shows which construction does.

The bootstrap, introduced by Efron [2], approximates sampling variability by resampling from the empirical distribution. The Bayesian bootstrap of Rubin [3] replaces

multinomial resampling weights by Dirichlet weights and gives a Bayesian nonparametric interpretation to empirical resampling. The bridge between the two devices was consolidated by Lo [4], who studied the Bayesian bootstrap as the Dirichlet process posterior with vanishing prior mass and established its large-sample equivalence with Efron's bootstrap, and by Muliere and Secchi [5], whose proper Bayesian bootstrap retains an informative base measure in the one-level posterior of the Dirichlet process of Ferguson [6]. In a different tradition, the hierarchical Dirichlet process of Teh et al. [1] generates random probability measures through multi-level nonparametric priors, with methodological goals that have usually been distinct from those of the resampling literature. Beyond resampling, uncertainty quantification for tail-sensitive quantities under limited data arises across contemporary statistical modeling—for instance, in reliability and prognostics, where recent work models degradation data with initiation–growth correlations [7], predicts remaining useful life through exponential dispersion processes with random drifts [8], and surveys the passage from statistical to AI-integrated inverse Gaussian process models [9]—and the present paper addresses the resampling side of that broader inferential landscape.

The double bootstrap of Efron [10] and related iterated-bootstrap procedures add a second resampling level to the ordinary bootstrap. The first level produces bootstrap samples; the second level is used to calibrate inferential quantities such as coverage probabilities or error rates. This second level is computationally expensive but conceptually important: it is not merely a second estimate of the same object. It is a calibration device. Any Bayesian or Bayesian-nonparametric reinterpretation of the double bootstrap must therefore preserve two separate roles: an outer uncertainty-generating layer and an inner calibration layer.

This paper begins from the observation that the hierarchical Dirichlet process also has a two-level structure. A recent limit theorem for hierarchical Dirichlet constructions due to Feng and Paguyo [11] describes the behavior of such structures as their concentration parameters diverge. Read only as an asymptotic result, that theorem belongs to Bayesian nonparametrics. Read constructively, however, it suggests a methodological question: can the classical double bootstrap be understood as an infinite-concentration endpoint of a two-level hierarchical Dirichlet resampling scheme?

The answer is affirmative, and its precise form is what this paper exploits. We identify the classical double bootstrap as the *non-informative endpoint* of a two-level Dirichlet construction, and we use that identification to build a method. The correspondence between a single Dirichlet level and the bootstrap is of course not new—it is the substance of Lo [4] and Muliere and Secchi [5]—and the limit theory of the two-level case is due to Feng and Paguyo [11]. Their results motivate the large-concentration regime considered here, but they do not identify the classical double bootstrap as its resampling endpoint; that identification, and the finite-concentration total-variation and Kullback–Leibler bounds of Section 2.2, are established independently below. What is new is the constructive use to which the two-level identification is put. Once the double bootstrap is seen as an endpoint rather than an isolated algorithm, it becomes the boundary of a family of finite-concentration hierarchical procedures whose interior has never been explored as a methodology. At the boundary both levels collapse to ordinary resampling; away from it, prior information enters through the base measure and the concentration parameters.

This observation leads to a new methodology: the *Informative Bayesian Double Bootstrap* (IBDB). The IBDB keeps the two-level logic of the double bootstrap but replaces the non-informative resampling mechanism by a finite-concentration hierarchical Dirichlet construction. The outer level incorporates prior information through a base measure, while the inner level retains the calibration role that distinguishes the double bootstrap from

a one-level Bayesian bootstrap. In this sense, the IBDB is not simply a proper Bayesian bootstrap with an informative prior. It is an informative *double* bootstrap.

The results below are of five different kinds, and it is worth stating at the outset how they depend on one another so that the paper is read as one argument rather than as a collection.

- **Principal contribution (the identification).** The classical double bootstrap is the non-informative endpoint of a two-level Dirichlet construction—of the construction taken conditionally on its realized upper level, and of no regime of the hierarchical Dirichlet process itself. With one concentration diverging, the franchise keeps a tie probability of $1/(1 + \alpha)$; with both diverging, it degenerates to a *single* bootstrap stage rather than two (Section 2.3). The two-level Dirichlet form has appeared before, in Shin et al. [12], as a device for computational amortization; it is not identified as a limit nor quantified there. The identification is what licenses the IBDB: the method is not a heuristic addition to the double bootstrap but the interior of a family on whose boundary the double bootstrap already sits (Sections 2.1 and 2.2, and the object itself in Section 3).
- **Key conceptual result.** What the second level buys is calibration, not support extension and not interval width. A correctly implemented one-level informative bootstrap is a strong competitor—it crosses the support ceiling and reaches high coverage in moderate cells—but it does so by inflating the interval, over-covering at about 1.2 times the frequentist width for the value-at-risk. On the expected shortfall, where the quantile ceiling has been lifted by construction, the two-level method attains the highest coverage in every cell at widths within ten percent of the strongest one-level competitor, and its gain appears as a fall in the fraction of intervals lying entirely below the truth rather than as widening—repositioning, which no one-level competitor in our panel reproduces (Sections 3.4, 4.6 and 4.11).
- **Theoretical result.** A total-variation coupling bound and a Kullback–Leibler bound with an exact second-order constant. These serve two purposes at once: they anchor the family, guaranteeing that the classical procedure is recovered at the boundary, and they quantify the rate at which departing from the boundary can begin to buy anything (Section 2.2).
- **Predictive result.** A competence map across statistics of increasing moment order—upper-tail quantiles, the standard deviation, and the skewness—together with a support threshold that yields a falsifiable prediction: coverage of any resampling method *restricted to the observed sample support* must be exactly zero while the largest attainable sample value of the statistic lies below the population value (Section 4.5). Across populations, sample sizes, and quantile levels, the classical double bootstrap does not merely respect the support ceiling $1 - p^n$ but *attains* it, while the informative method crosses it (Section 4.4).
- **Validation.** A uniform factorial study against a fixed panel of four competitors—the single and double bootstrap, the proper Bayesian bootstrap of Muliere and Secchi [5] in its two-step form, and the posterior bootstrap of Fong et al. [13]—across six statistics, together with a misspecification study, a directional diagnostic, and two real-data applications, to the Danish fire losses and the Siemens daily equity losses (Sections 4–4.10).

Figure 1 summarizes the conceptual point. The classical double bootstrap appears at the purely empirical end of a continuum. Moving away from that end lets prior information enter through the base measure; at the opposite end the prior fully drives the inference. The contribution of the paper is to study the methodological consequences of the entire continuum rather than only its endpoint.

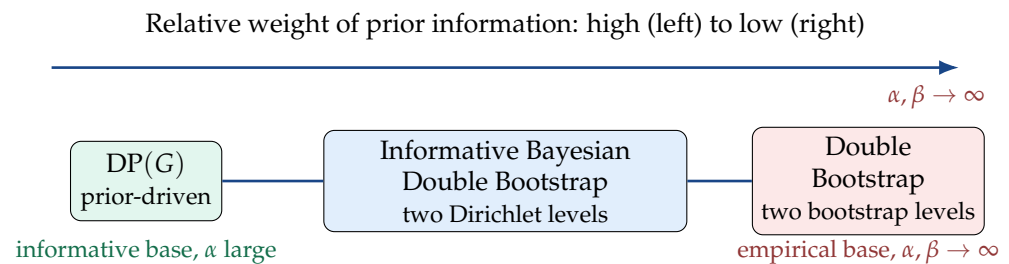


Figure 1. The continuum underlying the paper, drawn on the conceptual axis of prior influence. At one end the prior fully drives the inference; at the other sits the purely empirical double bootstrap. The Informative Bayesian Double Bootstrap occupies the middle, preserving two hierarchical levels. The precise dials behind this schematic are made explicit in the text: the concentration of the empirical-base predictive construction for the bootstrap endpoint (Proposition 1), and the prior mass relative to the sample for the influence of G (Section 3.4). The axis is a resultant of those two dials, not a single one: they are distinct coordinates and are not interchangeable (Section 3.1).

The paper is organized as follows. Section 2.1 revisits a finite example with five observations and explains why the double bootstrap and the hierarchical Dirichlet process share a two-level structure, and where that shared structure stops. Section 2.2 fixes the notation convention, defines the two-level predictive construction, and proves the elementary total-variation bound that identifies the double bootstrap as its infinite-concentration endpoint. A companion Kullback–Leibler bound, of second order in the inverse concentrations, sharpens the total-variation rate by a factor of order $\sqrt{n}$ through Pinsker’s inequality and admits an exact second-order constant; both results are situated within the limit theory of Feng and Paguyo [11]. Section 2.3 works out the hierarchical Dirichlet process at $J = 1$ in a five-point example and shows that it does not reach the endpoint: the reinforcement carried by its return path to the shared base measure survives a single diverging concentration. That failure is diagnostic, since it locates the one feature that has to be given up, and the section then states the resulting two-level model in both its Dirichlet-process and finite forms. Section 3 defines the IBDB methodology. Section 3.3 describes the empirical Bayes regulation of prior influence and its limitations. Section 3.4 explains why a two-level construction is preferable to a highly informative one-level Dirichlet process, and Section 3.5 provides analytical estimates for the proposed scheme itself at finite concentration: total-variation proximity bounds with explicit constants, an exact support-escape formula that formalizes its separation from the support-confined methods, and a first-order validity result for the mean. Section 4 supports that argument with a uniform factorial study organized by target statistic: location, scale, upper quantiles, skewness, and the risk functionals, followed by a misspecification study (Section 4.7), the directional diagnostic that interprets it (Section The Directional Artifact, Diagnosed), a sensitivity study for the resolution q and the concentration coupling (Section 4.8), and two real-data applications: the Danish fire losses (Section 4.9) and the Siemens daily equity losses (Section 4.10). Section 4.11 states what the simulations establish. Section 5 situates the method within the literature, interprets what was established, gives practical guidance on when the method should and should not be used (Section 5.3), and states the limitations and the directions they leave open.

2. From the Double Bootstrap to a Two-Level Dirichlet Construction

2.1. Motivating Example

Let $\theta_1, \dots, \theta_5$ be five distinct observations from a continuous distribution. A first-level bootstrap sample might be

$$S_1 = (\theta_1, \theta_3, \theta_5, \theta_5, \theta_3),$$

and a second-level bootstrap sample drawn from S_1 might be

$$S_{11} = (\theta_1, \theta_3, \theta_5, \theta_5, \theta_5).$$

Similarly, two other first-level bootstrap samples and their second-level descendants could be

$$S_2 = (\theta_4, \theta_2, \theta_1, \theta_5, \theta_5), \quad S_{21} = (\theta_5, \theta_5, \theta_4, \theta_5, \theta_1),$$

and

$$S_3 = (\theta_1, \theta_5, \theta_4, \theta_2, \theta_3), \quad S_{31} = (\theta_1, \theta_2, \theta_2, \theta_3, \theta_5).$$

The double bootstrap therefore creates a two-level tree:

$$\{\theta_1, \dots, \theta_5\} \longrightarrow S_j \longrightarrow S_{jk}.$$

The key feature is not only that the data are resampled, but that resampling is performed twice, with the second level conditional on the first.

The Dirichlet process has a different language but a related mechanism. Suppose

$$P \mid \alpha, G \sim \text{DP}(\alpha, G), \quad \phi_i \mid P \sim P.$$

The Blackwell–MacQueen predictive distribution [14] states that, after observing $i - 1$ generated atoms $\phi_1, \dots, \phi_{i-1}$,

$$\phi_i \mid \phi_1, \dots, \phi_{i-1}, \alpha, G \sim \frac{1}{\alpha + i - 1} \sum_{r=1}^{i-1} \delta_{\phi_r} + \frac{\alpha}{\alpha + i - 1} G.$$

Thus, each draw either repeats a previously generated atom or innovates from the base measure G . When G is chosen to be the empirical distribution on $\theta_1, \dots, \theta_5$, innovation from G means drawing one of the original observations.

The hierarchical Dirichlet process of Teh et al. [1] makes this two-level idea precise. It stacks two Dirichlet processes: a *global* random measure

$$G_0 \mid \alpha, H \sim \text{DP}(\alpha, H),$$

drawn once from a base measure H , and then, for each group j , a *group-level* measure

$$G_j \mid \beta, G_0 \sim \text{DP}(\beta, G_0)$$

that takes the realized G_0 as its own base. Because G_0 is almost surely discrete, its atoms are exactly the innovations produced at the global level—in our running example, the original observations $\theta_1, \dots, \theta_5$ selected by the first-level urn. Every group-level measure G_j can only reuse those atoms, so the groups are coupled through the shared global menu. The two stages are therefore genuinely distinct: the first (global) stage decides *which* atoms are available and with what weights; the second (group) stage resamples *from those realized atoms* with its own concentration β . It is this second stage—resampling conditionally on the atoms fixed by the first—that plays the role of the double bootstrap’s inner level.

For the present paper, the relevant object is not the full mixture-model machinery of the HDP but the two-level predictive resampling mechanism it induces. The analogy with the double bootstrap is then immediate: the first hierarchical level generates a resampling distribution, and the second generates another, conditionally on the first. One structural contrast should be kept in view from the start. In Teh et al. [1] the hierarchical Dirichlet process arises from finite two-level Dirichlet models by sending the *dimension* of the global

support to infinity with both concentrations held fixed (their Section 4.1); the limit studied in this paper turns the orthogonal knob, holding the support fixed at the data and prior atoms and sending both *concentrations* to infinity. The first limit builds the process; the second collapses it onto the double bootstrap.

The motivating question is therefore simple:

Do the two-level resampling probabilities of the double bootstrap arise as a limiting case of a two-level hierarchical Dirichlet construction?

The answer depends crucially on which hierarchical Dirichlet construction is used. Some constructions retain reinforcement in a way that prevents collapse to independent bootstrap resampling. The construction relevant for the present paper is the one whose large-concentration behavior is covered by the limit theory discussed in Section 2.2.

2.2. Connection with the Hierarchical Dirichlet Process

This section formulates the conceptual connection between the double bootstrap and the large-concentration behavior of hierarchical Dirichlet constructions. The purpose is not to restate the limit theory of Feng and Paguyo [11], but to make explicit the bootstrap interpretation of the relevant limiting regime, and to record an elementary, self-contained bound that identifies the double bootstrap as the endpoint.

Notation Convention

Throughout the paper, α denotes the level-one (outer) concentration and β the level-two (inner) concentration. Both are concentration parameters in the standard sense: larger values give more weight to innovation from the base measure, and the bootstrap endpoint is $\alpha, \beta \rightarrow \infty$. This convention agrees with Feng and Paguyo [11], whose level-one and level-two parameters are also denoted α and β ; in the notation of Teh et al. [1], the corresponding pair is (γ, α_0) . Two warnings for readers of related material: First, some early working notes on this project used the reversed assignment (top-level β , group-level α); the present convention supersedes it. Second, concentrations are always totals: when a base measure is represented by q atoms, each atom carries concentration α/q , and statements about the influence of the prior must be read at the level of the total α .

Let P_{DB} denote the law of a double-bootstrap pair $(\Phi^{(1)}, \Phi^{(2)})$, where $\Phi^{(1)}$ is a first bootstrap sample drawn with replacement from the empirical distribution of $\theta_1, \dots, \theta_n$, and $\Phi^{(2)}$ is a second bootstrap sample drawn with replacement from the empirical distribution of $\Phi^{(1)}$. Throughout, the observations $\theta_1, \dots, \theta_n$ are assumed distinct, as happens almost surely when they are drawn from a continuous distribution; the empirical base measure then places mass $1/n$ on n distinct atoms, which is the setting the propositions below use.

Let $P_{\alpha, \beta}$ denote the law of the pair generated by the two-level predictive construction with the base measures matched to the empirical measures at the corresponding levels: $\Phi^{(1)} = (\varphi_1^{(1)}, \dots, \varphi_n^{(1)})$ is drawn sequentially from the Blackwell–MacQueen urn with concentration α and base $\hat{F}_n$ so that

$$\varphi_i^{(1)} \mid \varphi_1^{(1)}, \dots, \varphi_{i-1}^{(1)} \sim \frac{1}{\alpha + i - 1} \sum_{r=1}^{i-1} \delta_{\varphi_r^{(1)}} + \frac{\alpha}{\alpha + i - 1} \hat{F}_n,$$

and, conditionally on $\Phi^{(1)}$, the second sample $\Phi^{(2)}$ is drawn from the urn with concentration β and base the empirical distribution of $\Phi^{(1)}$. We never place one Dirichlet process inside another's base measure: throughout this manuscript, we do not write $\text{DP}(\alpha, \text{DP}(\cdot))$.

Proposition 1 (Bootstrap endpoint of the two-level construction). *For all $\alpha, \beta > 0$ and $n \geq 1$,*

$$d_{\text{TV}}(P_{\alpha, \beta}, P_{\text{DB}}) \leq \frac{n(n-1)}{2} \left(\frac{1}{\alpha} + \frac{1}{\beta} \right).$$

In particular, $P_{\alpha, \beta} \rightarrow P_{\text{DB}}$ in total variation whenever both concentration parameters diverge, at rate $O(n^2/\alpha + n^2/\beta)$, and at rate $O(n^2/\alpha)$ in the symmetric regime $\alpha = \beta$.

Proof. We construct an explicit coupling in three steps.

Step 1 (couple level one). Generate the urn draws sequentially. At step i , flip an innovation coin with success probability $\alpha/(\alpha + i - 1)$: on success, draw a fresh atom from $\hat{F}_n$, on failure reuse a uniformly chosen previous atom. Give the bootstrap sample the *same* innovation draws from $\hat{F}_n$.

The bootstrap always draws from $\hat{F}_n$, whereas the urn does so only when it innovates. A single reinforcement therefore breaks the agreement: at that step the urn reuses a previously generated atom while the bootstrap draws afresh from $\hat{F}_n$, and the two samples need not coincide thereafter. The coupling therefore succeeds at level one whenever no reinforcement occurs, so

$$\begin{aligned} \mathbb{P}(\text{coupling fails at level one}) &\leq \mathbb{P}(\text{some reinforcement occurs}) \\ &\leq \sum_{i=1}^n \frac{i-1}{\alpha + i - 1} \leq \frac{1}{\alpha} \sum_{i=1}^n (i-1) = \frac{n(n-1)}{2\alpha}, \end{aligned}$$

where the first inequality is a union bound over the n steps and the second uses $\alpha + i - 1 \geq \alpha$. Write A for the event that the two level-one samples coincide.

Step 2 (couple level two, conditionally). On A , both constructions have produced the *same* first-level sample. Repeat the coupling of Step 1 at level two, conditionally on that common sample, with concentration β and base its empirical distribution. The same computation gives

$$P(\text{coupling fails at level two} \mid \Phi^{(1)}) \leq \frac{n(n-1)}{2\beta} \quad \text{for every realization of } \Phi^{(1)}.$$

The bound does not depend on which first-level sample was realized; this uniformity is what lets the conditional bound be used without any independence between the two levels.

Step 3 (combine). The coupling fails only if it fails at level one or succeeds there and fails at level two:

$$P(\text{coupling fails}) \leq P(A^c) + P(\text{fails at level two} \mid A) P(A) \leq \frac{n(n-1)}{2\alpha} + \frac{n(n-1)}{2\beta},$$

using $P(A) \leq 1$ and the uniform bound of Step 2. Since the total variation distance between two laws is bounded by the probability that any coupling of them fails, the claim follows. $\square$

Three remarks calibrate the statement. First, the bound is elementary but exact in its dependence: both concentrations must diverge, and the slower of the two governs the error—sending a single parameter to infinity is not sufficient, as Section 2.3 demonstrates with a closed-form counterexample in the franchise construction. Second, the regime in which both parameters diverge at comparable rates, $\alpha/\beta \rightarrow c \in (0, \infty)$, is precisely the regime in which Feng and Paguyo [11] develop the fluctuation theory of the hierarchical Dirichlet process: laws of large numbers, central limit theorems, and a large deviation principle for power-sum functionals of the weights. Proposition 1 adds the resampling

reading of that regime; the Gaussian theory of Feng and Paguyo [11] describes the finer structure around the endpoint. Third, Feng and Paguyo [11] also treat the finite-dimensional hierarchical Dirichlet process, in which the level-one weights are $\text{Dirichlet}(\alpha/n, \dots, \alpha/n)$ over n atoms, under the additional scaling $\alpha/n \rightarrow d \in (0, \infty)$. The parameter d quantifies the impact of the finite-dimensional approximation, and the same ratio of total concentration to number of atoms reappears in Section 3.3 as the quantity governing how much influence a finitely represented prior can exert.

The coupling argument measures the discrepancy in total variation. The discrepancy can also be measured in information units, and the result is stronger: the Kullback–Leibler divergence between the two laws is of *second* order in the inverse concentrations, and converting it back through Pinsker’s inequality improves the total-variation rate of Proposition 1 by a factor of order $\sqrt{n}$.

Proposition 2 (Kullback–Leibler bound and improved total-variation rate). *Under the assumptions of Proposition 1, for all $\alpha, \beta > 0$ and $n \geq 1$,*

$$\text{KL}(P_{\text{DB}} \parallel P_{\alpha, \beta}) \leq \frac{n(n-1)^2}{4} \left(\frac{1}{\alpha^2} + \frac{1}{\beta^2} \right),$$

and consequently, by Pinsker’s inequality,

$$d_{\text{TV}}(P_{\alpha, \beta}, P_{\text{DB}}) \leq (n-1) \sqrt{\frac{n}{8}} \sqrt{\frac{1}{\alpha^2} + \frac{1}{\beta^2}}.$$

Proof. We prove the divergence bound in five steps.

Step 1 (chain rule, and the measure it is taken under). Both laws are laws of the $2n$ coordinates of the pair $(\Phi^{(1)}, \Phi^{(2)})$, so the chain rule for the Kullback–Leibler divergence applies coordinate by coordinate:

$$\text{KL}(P_{\text{DB}} \parallel P_{\alpha, \beta}) = \sum_{\text{steps } i} \mathbb{E}_{P_{\text{DB}}} \left[\text{KL}(P_{\text{DB}}(\cdot \mid \text{past}) \parallel P_{\alpha, \beta}(\cdot \mid \text{past})) \right].$$

The outer expectation is taken *under* P_{DB} , the law in the first argument. This is what makes the counts below binomial: under P_{DB} , the draws at a level are i.i.d., whereas under $P_{\alpha, \beta}$ they reinforce and are not.

Step 2 (one level at a time). Each level has the same structure: a concentration $\kappa \in \{\alpha, \beta\}$ and a base pmf p supported on K atoms. At level one, p is uniform over the n data atoms and $K = n$. At level two, conditionally on $\Phi^{(1)}$, p is the empirical pmf of $\Phi^{(1)}$ and K is its number of distinct values. Fix a level and write $c_i(a)$ for the number of occurrences of atom a among the first $i-1$ draws. Then

$$P_{\text{DB}}(\text{draw } i = a \mid \text{past}) = p_a, \quad P_{\alpha, \beta}(\text{draw } i = a \mid \text{past}) = \frac{c_i(a) + \kappa p_a}{\kappa + i - 1}.$$

Step 3 (the step- i divergence). Substituting the two conditionals into the definition of the divergence and simplifying the ratio—the $\log p_a$ contributed by the numerator cancels exactly against the $-\log p_a$ obtained by factoring p_a out of the denominator $c_i(a) + \kappa p_a = p_a(\kappa + c_i(a)/p_a)$ —

$$\text{KL}_i = \sum_a p_a \mathbb{E}_{P_{\text{DB}}} \left[\log(\kappa + i - 1) - \log\left(\kappa + \frac{c_i(a)}{p_a}\right) \right], \quad c_i(a) \stackrel{P_{\text{DB}}}{\sim} \text{Binomial}(i-1, p_a).$$

The binomial law is the law of the counts *under* P_{DB} , by Step 1; no independence is assumed of the urn.

Step 4 (the first-order term vanishes). Because $\mathbb{E}_{P_{\text{DB}}}[c_i(a)] = (i-1)p_a$,

$$\mathbb{E}_{P_{\text{DB}}}\left[\kappa + \frac{c_i(a)}{p_a}\right] = \kappa + i - 1,$$

so the argument of the logarithm has mean exactly equal to the constant it is compared against: the expected reinforcement mass under i.i.d. sampling matches the urn's normalization without residue. Write $X = \kappa + c_i(a)/p_a$ and $\mu = \kappa + i - 1$ so that $\mathbb{E}_{P_{\text{DB}}}[X] = \mu$. A second-order Taylor expansion of $\log$ about μ , with Lagrange remainder at a point ξ between X and μ , gives after taking expectations (the linear term vanishes)

$$\mathbb{E}_{P_{\text{DB}}}[\log X] = \log \mu - \frac{1}{2} \mathbb{E}_{P_{\text{DB}}}\left[\frac{(X - \mu)^2}{\xi^2}\right].$$

Since $c_i(a)/p_a \geq 0$, we have $X \geq \kappa$, hence $\xi \geq \kappa$ and $\xi^{-2} \leq \kappa^{-2}$, so

$$0 \leq \mathbb{E}_{P_{\text{DB}}}\left[\frac{(X - \mu)^2}{\xi^2}\right] \leq \frac{\text{Var}(X)}{\kappa^2}.$$

Subtracting a quantity bounded *above* leaves the difference bounded *below*, which is the direction required to bound KL_i from above:

$$\mathbb{E}_{P_{\text{DB}}}\left[\log\left(\kappa + \frac{c_i(a)}{p_a}\right)\right] \geq \log(\kappa + i - 1) - \frac{\text{Var}(c_i(a)/p_a)}{2\kappa^2}, \quad \text{Var}\left(\frac{c_i(a)}{p_a}\right) = \frac{(i-1)(1-p_a)}{p_a},$$

the variance being that of $c_i(a) \sim \text{Binomial}(i-1, p_a)$ divided by p_a^2 .

Step 5 (sum over atoms, steps, and levels). Weighting by p_a and summing over a gives $\sum_a p_a \text{Var}(c_i(a)/p_a) = (i-1)(K-1)$, so $\text{KL}_i \leq (i-1)(K-1)/(2\kappa^2)$. Summing over $i = 1, \dots, n$ bounds the level by

$$\frac{n(n-1)(K-1)}{4\kappa^2} \leq \frac{n(n-1)^2}{4\kappa^2}, \quad \text{since } K-1 \leq n-1$$

at both levels and for every realization of $\Phi^{(1)}$ —which is what allows the level-two bound to be taken out of the expectation over $\Phi^{(1)}$ without any independence between levels. Adding the two levels proves the divergence bound, and Pinsker's inequality $d_{\text{TV}} \leq \sqrt{\text{KL}/2}$ gives the total-variation statement. $\square$

Three remarks are in order. First, the rate improvement is uniform in n : in the symmetric regime $\alpha = \beta$ the new bound reads $d_{\text{TV}} \leq (n-1)\sqrt{n}/(2\alpha)$, against $n(n-1)/\alpha$ from the coupling—a gain of exactly $2\sqrt{n}$ —so Proposition 2 supersedes the coupling bound quantitatively, while the coupling proof retains its value as the constructive mechanism behind the limit. Second, the divergence and the distance live on different scales: KL is $O(n^3(\alpha^{-2} + \beta^{-2}))$ while d_{TV} is $O(n^{3/2}(\alpha^{-1} + \beta^{-1}))$, and the second-order behavior of the divergence is not an artifact of the bound. Proposition 3 below identifies the exact second-order constant, and exact enumeration of both laws at $n \in \{3, 4, 5\}$ confirms the bound at all concentrations and the convergence to that constant (for instance, at $n = 5$ and $\alpha = \beta = 1000$, $\alpha^2 \text{KL} = 31.56$ against the limit 31.81, with exact $d_{\text{TV}} = 0.0030$ against the bound 0.0045). Third, the information-theoretic reading is the natural one for the methodology that follows: the divergence measures, in information units, how far a finite-concentration hierarchy sits from its purely empirical endpoint, and it is exactly this distance that the informative construction of Section 3 exploits rather than sends to zero.

Proposition 3 (Exact second-order constant). *Let K_n denote the number of distinct values in an i.i.d. uniform sample of size n from n atoms so that $\mathbb{E}[K_n] = n\{1 - (1 - 1/n)^n\}$. Then, as $\alpha, \beta \rightarrow \infty$ at arbitrary relative rates,*

$$\text{KL}(P_{\text{DB}} \| P_{\alpha, \beta}) = \frac{C_1}{\alpha^2} + \frac{C_2}{\beta^2} + o\left(\frac{1}{\alpha^2} + \frac{1}{\beta^2}\right), \quad C_1 = \frac{n(n-1)^2}{4}, \quad C_2 = \frac{n(n-1)}{4} \mathbb{E}[K_n - 1].$$

Proof. By the chain rule, $\text{KL}(P_{\text{DB}} \| P_{\alpha, \beta}) = \text{KL}_1(\alpha) + \mathbb{E}_{\Phi^{(1)} \sim P_{\text{DB}}} [\text{KL}_2(\beta; \Phi^{(1)})]$, and the mixing law of $\Phi^{(1)}$ under P_{DB} does not depend on either concentration, so the two levels can be treated separately and the joint statement follows by adding the two error terms.

Fix a level with concentration κ and base pmf p on K atoms, and recall from the proof of Proposition 2 that the step- i term is

$$T_i(\kappa) = \sum_a p_a \mathbb{E}[\log(\kappa + i - 1) - \log(\kappa + X_{i,a})], \quad X_{i,a} = \frac{c_i(a)}{p_a}, \quad \mathbb{E}[X_{i,a}] = i - 1.$$

The Lagrange form of the second-order expansion around the mean gives

$$T_i(\kappa) = \sum_a p_a \mathbb{E}\left[\frac{(X_{i,a} - (i - 1))^2}{2\tilde{\xi}_{i,a}^2}\right],$$

with $\tilde{\xi}_{i,a}$ between $\kappa + X_{i,a}$ and $\kappa + i - 1$; the linear term has vanished exactly because $\mathbb{E}[X_{i,a}] = i - 1$. Now bound $\tilde{\xi}_{i,a}$ on both sides. Since $X_{i,a} \geq 0$, we have $\tilde{\xi}_{i,a} \geq \kappa$. Since $p_a \geq 1/n$ at both levels (the base is uniform over n atoms at level one, and an empirical pmf with atoms of mass at least $1/n$ at level two), $X_{i,a} \leq (i - 1)/p_a \leq n(n - 1)$, so $\tilde{\xi}_{i,a} \leq \kappa + n(n - 1)$. Writing $V_i = \sum_a p_a \text{Var}(X_{i,a}) = (i - 1)(K - 1)$, the sandwich

$$\frac{V_i}{2(\kappa + n(n - 1))^2} \leq T_i(\kappa) \leq \frac{V_i}{2\kappa^2}$$

holds for every i , every base pmf with atoms of mass at least $1/n$, and every $\kappa > 0$; both sides are deterministic, so no integrability argument is needed. Summing over i gives, for the whole level,

$$\frac{n(n - 1)(K - 1)}{4(\kappa + n(n - 1))^2} \leq \text{KL}_{\text{level}}(\kappa) \leq \frac{n(n - 1)(K - 1)}{4\kappa^2},$$

and both sides are asymptotic to $n(n - 1)(K - 1)/(4\kappa^2)$ as $\kappa \rightarrow \infty$. At level one, $K = n$ yields C_1 . At level two, the sandwich holds conditionally on every realization of $\Phi^{(1)}$ with $K = K(\Phi^{(1)})$; taking the expectation over the fixed finite mixture $\Phi^{(1)} \sim P_{\text{DB}}$ preserves it, and $\mathbb{E}[K(\Phi^{(1)})] = \mathbb{E}[K_n]$ is the expected number of distinct values in an i.i.d. uniform sample. Its closed form follows from linearity: each atom is missed with probability $(1 - 1/n)^n$, so $\mathbb{E}[K_n] = n\{1 - (1 - 1/n)^n\}$. $\square$

Figure 2 gives the intended interpretation of Proposition 1. The two concentration parameters are explicitly different because the two hierarchical levels need not be regulated by the same parameter.

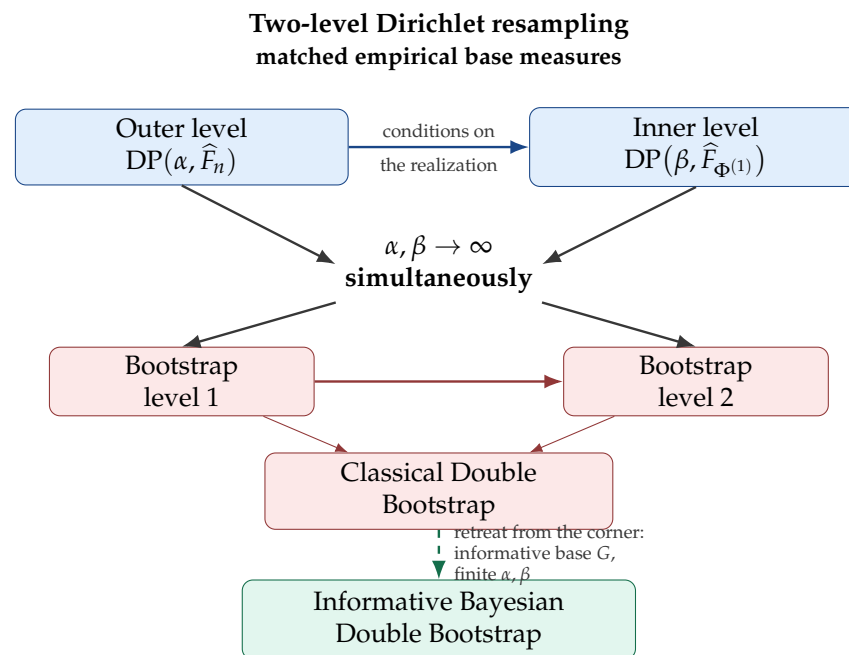


Figure 2. The two-level structure behind the connection, and what it is a connection *between*. The outer and inner levels have distinct concentration parameters, here denoted by α and β . The object drawn at the top is the *matched empirical* family, in which both base measures are empirical at their respective levels—the hypothesis of Proposition 1—and it is that family, not the IBDB, whose two hierarchical resampling levels collapse to the two bootstrap levels of the classical double bootstrap as both concentrations diverge. The IBDB is obtained by moving in the opposite direction, replacing the outer empirical base by an informative G and keeping the concentrations finite; with G held fixed, sending $\alpha, \beta \rightarrow \infty$ leads to the prior rather than to the double bootstrap, so the two coordinates are not interchangeable (Section 3.1). The non-informative calibration loop that fixes the interval level is a device attached to the two levels and is not drawn here (Algorithm 1). The figure is conceptual and does not place one Dirichlet process inside another’s base measure.

Remark 1 (Why the construction matters). *The connection is not true for every construction that uses the words “hierarchical Dirichlet process.” In particular, one must distinguish the predictive construction that converges to independent resampling from constructions that retain reinforcement across groups. A reader attempting to verify the result through the wrong representation may obtain a counterexample and conclude that the bootstrap connection fails. The correct statement is therefore not that any HDP collapses to the double bootstrap, but that the relevant large-concentration hierarchical Dirichlet construction admits the double bootstrap as its endpoint.*

Section 2.3 makes Remark 1 concrete: a worked five-point example shows the reinforcement mechanism that survives large concentration in the franchise construction, computes its probability in closed form, and exhibits the structural cut that removes it in the construction used by the IBDB.

2.3. The $J = 1$ Construction

This section revisits the five observations of Section 2.1 under the hierarchical Dirichlet process itself, specialized to a single group ($J = 1$), and contrasts it with the two-level finite Dirichlet construction on which the IBDB is built. The purpose is threefold: to show that the defining coupling of the HDP survives at $J = 1$; to prove, in a two-line calculation, that a single diverging concentration does not collapse the HDP to independent resampling (the counterexample anticipated by Remark 1); and to display the exact structural feature whose removal makes the bootstrap limit true.

Algorithm 1 Informative Bayesian Double Bootstrap (IBDB)

Require: Data $x_1, \dots, x_n$; base measure G represented by q atoms $g_1, \dots, g_q$; outer concentration α ; inner concentration β ; statistic T ; nominal level $1 - \gamma$; outer replicates B , calibration replicates C ; calibration band Γ around γ .

Support. Set the joint atom vector $a = (x_1, \dots, x_n, g_1, \dots, g_q)$ of length $n + q$.

Blocks 1–2—prior layer F_1 and data-anchored layer F_2 (repeated B times).

- 1: **for** $b = 1, \dots, B$ **do**
- 2: Draw outer atom weights $w_1^{(b)} \sim \text{Dirichlet}(\alpha/q, \dots, \alpha/q)$ over the q atoms of G .
- 3: Form concentration vector $c^{(b)} = (\underbrace{1, \dots, 1}_n, \beta w_{1,1}^{(b)}, \dots, \beta w_{1,q}^{(b)})$ over the $n + q$ atoms: unit mass on each datum, prior mass $\beta w_{1k}^{(b)}$ on prior atom k .
- 4: Draw joint weights $W^{(b)} \sim \text{Dirichlet}(c^{(b)})$ and set the outer statistic $T_b^* = T(\sum_i W_i^{(b)} \delta_{a_i})$.
- 5: If $b \leq M$ (see Block 3), draw a synthetic sample $y_1^{(b)}, \dots, y_n^{(b)}$ i.i.d. from $\sum_i W_i^{(b)} \delta_{a_i}$.
- 6: **end for**
- Block 3—non-informative calibration loop** (over the first $M = \min(B, 60)$ replicates).
- 7: **for** $b = 1, \dots, M$ **do**
- 8: Draw C flat-Dirichlet (Rubin) reweightings of $y^{(b)}$: $V^{(b,c)} \sim \text{Dirichlet}(1, \dots, 1)$, and record $T_{b,c}^* = T(\sum_i V_i^{(b,c)} \delta_{y_i^{(b)}})$.
- 9: **end for**
- 10: **For** each candidate level $\gamma' \in \Gamma$, form the interval from the $\gamma'/2$ and $1 - \gamma'/2$ quantiles of $\{T_{b,c}^*\}_{c=1}^C$ within each b , and estimate its coverage of the replicate's own target T_b^* across $b = 1, \dots, M$: within synthetic world b the sample $y^{(b)}$ was drawn from $\sum_i W_i^{(b)} \delta_{a_i}$, whose statistic is T_b^* , so T_b^* is the truth that the inner interval must cover.
- 11: Choose $\gamma^* = \arg \min_{\gamma' \in \Gamma} |\widehat{\text{cov}}(\gamma') - (1 - \gamma)|$.
- 12: **return** the interval given by the $\gamma^*/2$ and $1 - \gamma^*/2$ quantiles of the outer statistics $\{T_b^*\}_{b=1}^B$.

2.3.1. The Model and Its Two Urns

With one group, the hierarchical model reads

$$G_0 \mid \alpha, H \sim \text{DP}(\alpha, H), \quad G_1 \mid \beta, G_0 \sim \text{DP}(\beta, G_0), \quad \theta_1, \dots, \theta_5 \mid G_1 \stackrel{\text{iid}}{\sim} G_1. \quad (1)$$

In the notation of Teh et al. [1], the pair (α, β) used here corresponds to their (γ, α_0) ; we keep the convention of Section 2.2, with α at level one. Integrating out G_1 and G_0 in (1) yields the two coupled Pólya urns of Teh et al. [1] (their Equations (24) and (25)), which at $J = 1$ read as follows. Customer i joins table t with probability $n_{1t}/(i - 1 + \beta)$ or opens a new table with probability $\beta/(i - 1 + \beta)$. Every new table orders a dish, repeating an existing dish ϕ_k with probability $m_{\cdot k}/(m_{\cdot\cdot} + \alpha)$ or innovating from H with probability $\alpha/(m_{\cdot\cdot} + \alpha)$. The global menu remains even though there is a single restaurant: it couples the tables of that restaurant to one another. Figure 3 shows the structure.

Both urns obey a rich-get-richer (preferential attachment) law: a table is joined, and a dish repeated, with probability proportional to its current count, so mass concentrates on atoms that are already popular, and the hierarchical stacking makes the cluster sizes more overdispersed than those of a single Dirichlet process of the same total concentration.

The structural cut described below removes this menu-level reinforcement in the construction used by the IBDB: its lower level has no dish urn. Once the upper level has realized its atoms, the lower level resamples only from that realized empirical measure, with no path back to a shared menu through which distinct draws could be forced to coincide. The only reinforcement that remains is the ordinary single-urn concentration already present in the one-level Bayesian bootstrap, so no reinforcement is added by the second level. This is exactly what allows both levels to collapse to independent resampling as the concentrations diverge: with the menu-level coupling absent, matching the base

measures to the empirical distributions sends the tie probability to zero—the bootstrap limit that the franchise cannot reach (Equation (3)).

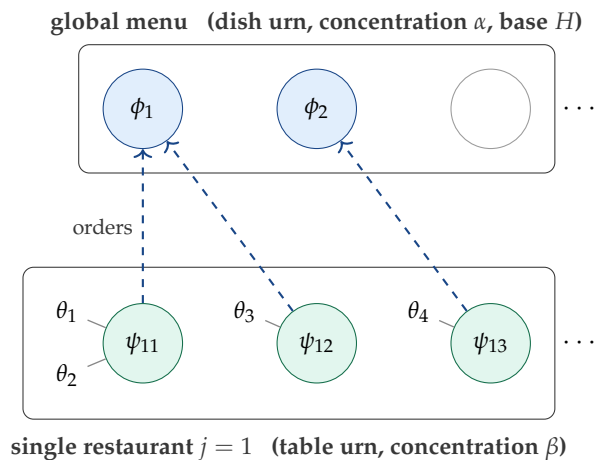


Figure 3. The hierarchical Dirichlet process at $J = 1$. One restaurant, but two urns: customers choose tables (concentration β), and tables order dishes from a global menu (concentration α , base H). Distinct tables may serve the same dish—here, $\psi_{11} = \psi_{12} = \phi_1$ —which is the coupling that a plain Dirichlet process does not have. The unfilled node on the menu is a dish not yet ordered: the innovation channel remains open at every step.

2.3.2. A Realization and the Signature Event

Take $\beta = \alpha = 1$ and H continuous. Customer 1 necessarily opens table 1, which orders a new dish $\phi_1 \sim H$; hence $\theta_1 = \phi_1$. Customer 2 then faces the tree of Figure 4: join table 1 with probability $\frac{1}{2}$ (so $\theta_2 = \phi_1$), or open table 2 with probability $\frac{1}{2}$, in which case the dish urn decides—repeat ϕ_1 with probability $\frac{1}{2}$, or innovate $\phi_2 \sim H$ with probability $\frac{1}{2}$.

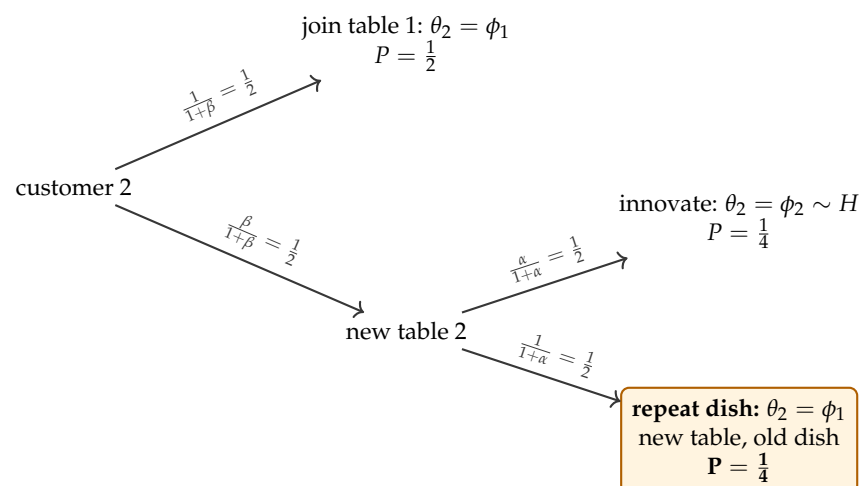


Figure 4. The second customer at $J = 1$ with $\beta = \alpha = 1$. The highlighted branch—a new table serving an old dish—has probability $\frac{1}{4}$ and is the signature of the hierarchical coupling: it requires the global menu and has no counterpart in a plain Dirichlet process, where $\theta_2 = \theta_1$ only if customer 2 physically joins Table 1. The unfilled node is the menu's open slot for a dish not yet drawn, and the ellipses indicate that the sequences of tables and dishes continue indefinitely; neither affects the construction.

Collecting branches,

$$P(\theta_2 = \theta_1) = \underbrace{\frac{1}{1+\beta}}_{\text{join the table}} + \underbrace{\frac{\beta}{1+\beta} \cdot \frac{1}{1+\alpha}}_{\text{new table, old dish}}, \quad (2)$$

which equals $\frac{3}{4}$ at $\beta = \alpha = 1$, against $\frac{1}{2}$ for a plain $\text{DP}(\beta, H)$ with $\beta = 1$. The menu is doing real probabilistic work even with a single restaurant: the coupling property of the HDP does not require multiple groups to operate, it only needs to be shared across them.

Table 1. The two constructions, level by level. The generative levels are not merely parallel: conditional on the upper level they are the *same* object, a Dirichlet process whose base measure is the realization of another. The difference is what happens with that upper level. The franchise marginalizes it *analytically*: the predictive rule integrates G_0 out in closed form, and it is that integration which creates the dish urn—the return path to the shared base measure, along which a fresh atom can still enter at any point. The IBDB instead draws the upper level afresh at each replicate and conditions on the realization so that, within a replicate, the lower level sees a fixed atom set and no return path exists, while across replicates the upper level is marginalized *by replication* rather than replaced by its mean (Section 2.3.5). The two senses of marginalization should be kept apart: the first is what the franchise does inside its predictive rule, the second is what the IBDB does across the outer loop of Algorithm 1. Two further differences follow: the base measure of the lower level is an empirical Bayes construction, part elicited and part estimated from the data, whose mixing proportions conjugacy fixes but does not motivate; and the interval level is chosen by a non-informative calibration loop with no counterpart in the franchise—a device attached to the two levels, not a third one. In the implementation, both concentrations are set by the empirical Bayes rule of Section 3.3, coupled as $\alpha = \beta = \hat{\alpha}$. The right-hand column describes the IBDB, whose outer base measure is informative; were both base measures instead matched to the empirical distributions, that same construction would reach the double-bootstrap endpoint as $\alpha, \beta \rightarrow \infty$, which is the family of Proposition 1 and the boundary from which the IBDB retreats (Section 3.1).

	HDP Franchise at $J = 1$	IBDB Construction (Cut)
Level 1	$G_0 \mid \alpha, H \sim \text{DP}(\alpha, H)$, with α fixed by the analyst. Marginalized: the predictive rule integrates G_0 out, and marginalization creates the dish urn—the return path to the shared base, along which a fresh atom can enter at any point	$F_1 \mid \hat{\alpha}, G \sim \text{DP}(\hat{\alpha}, G)$, with $\hat{\alpha}$ set from the data by the rule of Section 3.3. Drawn afresh at each replicate and conditioned on: within a replicate the lower level sees the realized atoms and nothing else; no dish urn, no return path
Level 2	$G_1 \mid \beta, G_0 \sim \text{DP}(\beta, G_0)$; the output is $\theta_i \mid G_1 \stackrel{iid}{\sim} G_1$, with no calibration. As $\beta \rightarrow \infty$ with α fixed, $P(\theta_2 = \theta_1) \rightarrow 1/(1 + \alpha) > 0$; with $\alpha, \beta \rightarrow \infty$ one stage collapses—no regime reaches the double-bootstrap endpoint	$F_2 \mid \hat{\beta}, F_1, x_{1:n} \sim \text{DP}(\hat{\beta} + n, \frac{\hat{\beta}}{\hat{\beta} + n} F_1 + \frac{n}{\hat{\beta} + n} \hat{F}_n)$: the empirical enters here, as the estimated component of an empirical Bayes base measure—part elicited (F_1), part estimated ($\hat{F}_n$)—with proportions fixed by conjugacy. The output is $T(F_2)$ over the outer draws, at a level γ^* chosen by a non-informative calibration loop (a device, not a level)

2.3.3. Why One Diverging Concentration Is Not Enough

Now send $\beta \rightarrow \infty$ in (2) with α fixed. Every customer opens their own table—the table-level reinforcement dies—yet

$$P(\theta_2 = \theta_1) \longrightarrow \frac{1}{1 + \alpha} > 0. \quad (3)$$

The draws become i.i.d. from G_0 , but $G_0 \sim \text{DP}(\alpha, H)$ is a discrete random measure: ties persist through the dish urn no matter how large β is. With $\alpha = 1$, the tie probability remains $\frac{1}{2}$ at $\beta = \infty$, even though i.i.d. sampling from the continuous H would give a tie probability of zero. This is Remark 1 in miniature, now with a closed-form witness: in the franchise, a single diverging concentration does not yield independent resampling. Nor does sending both to infinity rescue it, for the failure then changes sign: with $\alpha, \beta \rightarrow \infty$, every customer opens a table and every table innovates its own dish from H , so the draws become i.i.d. from H directly—one resampling stage, not two. The franchise thus misses the double bootstrap in both regimes, keeping ties in one and losing a level in the other, and no setting of (α, β) repairs it. The conditions of Proposition 1—both levels conditioned on realizations, with base measures matched to the empirical distributions—belong to the

scheme of Section 2.2, not to the franchise. A reader who tests the connection by sending parameters to infinity in the full HDP will manufacture exactly the counterexamples this section quantifies.

2.3.4. The Structural Cut in the IBDB Construction

Before describing the cut in words, Table 1 places the two constructions side by side. The left column is the hierarchical Dirichlet process at $J = 1$; the right column is the construction on which the IBDB is built. The table has one row per level. In the first row the constructions part over what happens with the upper level: the franchise integrates it out—and it is marginalization that produces the dish urn and with it the return path—while the IBDB draws it afresh at each replicate and conditions on the realization. In the second row, both are a Dirichlet process based on the realization of the level above; that shared skeleton is not where they differ. They differ in the base measure, which for the IBDB is empirical Bayes, part elicited and part estimated from the data—this is where the empirical distribution enters—and in what follows the level: the franchise outputs draws, the IBDB an interval at a level chosen by a non-informative calibration loop it attaches to the two levels.

The two-level construction on which the IBDB is built—the finite hierarchical Dirichlet model in the spirit of Equation (29) of Teh et al. [1], as implemented in our simulations—resamples differently. Level 1 draws a weight vector over the q atoms of the prior G with its own concentration and generates a sample s_1 i.i.d. given those weights; level 2 places a Dirichlet distribution anchored on the data together with the realized s_1 —the empirical Bayes base of Table 1—and resamples from it. In the five-point notation of Section 2.1: once $S_1 = (\theta_1, \theta_3, \theta_5, \theta_5, \theta_3)$ is realized, the second level sees only these realized atoms together with the data—a fixed, closed support. There is no dish urn and no path back to a global menu: the event “new table, old dish” has no counterpart, by construction. Figure 5 displays the contrast.

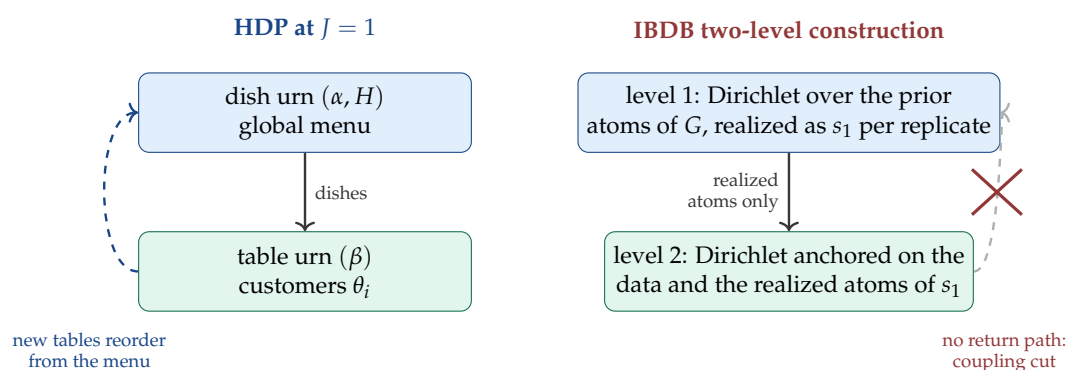


Figure 5. The structural difference that decides the limit. Left: at $J = 1$, the HDP retains the return path to the global menu, so reinforcement survives large β (Equation (3)). Right: the two-level construction underlying the IBDB conditions level 2 on the data and the realized atoms of level 1 and on nothing else; the return path is absent by construction, and with base measures matched to the empirical distributions, both levels collapse to bootstrap resampling as their concentrations diverge. Box colors distinguish the two constructions (the franchise on the left, the IBDB levels on the right), and the cross marks the return path to the shared menu that the IBDB construction removes.

Two consequences follow. First, the correct reading of Proposition 1: the double-bootstrap endpoint belongs to the two-level construction without the return path, not to the full franchise. Second, a clarification promised in Section 2.1: the small- α and large- α regimes both have bootstrap-like readings, but through different mechanisms. At small concentration the innovation channel dies and resampling concentrates on previously generated atoms; at large concentration, with bases matched

to empiricals, reinforcement dies and resampling becomes i.i.d. from the empirical distribution. The bootstrap endpoint of Figure 1—the purely empirical end of the conceptual continuum—is reached through the second mechanism, which is the one covered by Proposition 1 and by the limit theory of Feng and Paguyo [11].

A third consequence is computational, and it is what keeps the construction inside the resampling family rather than moving it into Markov chain Monte Carlo. The posterior of a hierarchical Dirichlet process given the data is not available in closed form. Writing $\mathbf{z}$ for the latent assignment of each observation to an atom of the shared menu, the joint posterior factorizes as

$$p(\beta, \phi, \pi, \mathbf{z} \mid X) \propto p(\beta \mid \alpha) \prod_k p(\phi_k \mid H) \prod_j p(\pi_j \mid \beta, \alpha) \prod_j \prod_i \pi_{j,z_{ji}} p(x_{ji} \mid \phi_{z_{ji}}), \quad (4)$$

and the assignments $\mathbf{z}$ cannot be eliminated: only the values x_{ji} are observed, never the atom that produced each one, and in the franchise representation, the table counts m_k that update the global menu are latent for the same reason. Even after integrating out the atoms ϕ_k under a conjugate base measure, the assignments remain, and drawing from (4) requires a Gibbs sampler alternating between the assignments and the table configuration. A method built on that posterior would be an MCMC procedure, not a resampling one.

The cut removes the need for it. Because the lower level is anchored on the realized atoms of the upper level and on the empirical distribution of the data, the update is performed by the empirical anchor rather than by conditioning on latent assignments. The second level is an ordinary Dirichlet process posterior given a fixed base measure, which is available in closed form and can be sampled directly. This is the same device that makes the Bayesian bootstrap of Rubin [3] a resampling scheme rather than a posterior computation, and it is what allows the IBDB to inherit the double bootstrap's algorithmic form. The price is stated plainly: the IBDB does not sample the posterior of a hierarchical Dirichlet process, and it does not claim to. It samples a two-level construction whose upper level carries prior information and whose lower level is anchored on the data, and the propositions of Section 2.2 describe where that construction sits relative to the classical double bootstrap.

2.3.5. The Genuinely Two-Level Model

The preceding subsections identified which construction collapses to the double bootstrap. It remains to state the genuinely two-level model itself—the object referred to in our simulations as the hierarchical variant—in its two equivalent forms: the idealized Dirichlet-process form, and the finite Dirichlet form actually implemented.

Dirichlet-Process Form

Let G be the informative base measure (the analyst's belief about the data distribution) and let α measure the confidence placed in it. The model stacks two Dirichlet processes:

$$\begin{aligned} \text{level 1 (knowledge): } F_1 \mid \alpha, G &\sim \text{DP}(\alpha, G), \\ \text{level 2 (data/resampling): } F_2 \mid \beta, F_1, x_1, \dots, x_n &\sim \text{DP}\left(\beta + n, \frac{\beta}{\beta + n} F_1 + \frac{n}{\beta + n} \hat{F}_n\right). \end{aligned} \quad (5)$$

The upper level carries the prior: G enters as the base of F_1 , with α as the confidence dial. The lower level is built on the realization of the upper one: its base measure mixes the random F_1 with the empirical distribution, and β governs the resampling. Inference targets $T(F_2)$ through draws of F_2 , with the same calibration layer as in the main method (synthetic samples from each outer draw, a non-informative inner loop targeting the outer statistic

of the same replicate, and a calibrated level γ^* chosen within a bounded band around the nominal level).

Finite Dirichlet Form (as Implemented)

Represent G by q atoms $g_1, \dots, g_q$. Then,

$$\begin{aligned} \text{level 1: } w_1 &\sim \text{Dirichlet}(\alpha/q, \dots, \alpha/q), & F_1 &= \sum_{k=1}^q w_{1k} \delta_{g_k}, \\ \text{level 2: } (v_1, \dots, v_q, u_1, \dots, u_n) &\sim \text{Dirichlet}(\beta w_{11}, \dots, \beta w_{1q}, 1, \dots, 1), & (6) \\ F_2 &= \sum_{k=1}^q v_k \delta_{g_k} + \sum_{i=1}^n u_i \delta_{x_i}. \end{aligned}$$

Equation (6) is the definition of the method, not an approximation to (5) adopted for convenience. The analyst's belief G is represented by a fixed set of q atoms, and the resampling distribution is supported on those atoms together with the data. The modeling statement is that the prior contributes this finite collection of possible values, held fixed across replicates, rather than a fresh set of values drawn anew at each iteration. Two properties make the choice self-contained. First, q controls fidelity, not strength: each atom carries concentration α/q , so the total prior mass is α , whatever q may be, and raising q describes the same belief more finely without giving it more influence. Second, the construction has the right limit: as $q \rightarrow \infty$ level 1 converges to $\text{DP}(\alpha, G)$, the symmetric-Dirichlet limit of Ishwaran and Zarepour [15], and the same finite two-level mechanism underlies Section 4.1 of Teh et al. [1]. The limit anchors (6) in the Dirichlet-process family and licenses the propositions of Section 2.2, which are stated for the idealized object; but the procedure defined and studied in this paper is the finite one, and its properties are established at the finite q actually used. In the implementation, both concentrations are regulated by the empirical Bayes rule of Section 3.3, coupled as $\alpha = \beta = \hat{\alpha}$; a decoupled two-dimensional empirical Bayes rule is left as an open refinement.

Random Versus Fixed Prior Weights

A single line separates (6) from a simpler construction in which the prior atoms enter the resampling level with fixed concentration α/q each; in (6) they enter with random concentration βw_{1k} , re-drawn at every outer replicate so that the upper-level realization of F_1 is genuinely marginalized rather than replaced by its mean. Only the second is a hierarchy of two Dirichlet processes in the sense of (5); the first collapses the upper level to a deterministic weight vector and is best read as a mean-field approximation to it. The two constructions share the same total prior mass α and therefore the same first-order behavior in the bounds of Section 2.2; they differ only at second order, and a paired comparison at 1000 replications across the design of Section 4.1 does not resolve that difference. We nevertheless adopt (5) and (6) as the IBDB proper, on the structural grounds that only it marginalizes the upper level rather than replacing it by its mean, and we record the fixed-weight simplification as the mean-field variant of the family.

3. The Informative Bayesian Double Bootstrap

3.1. Definition of the Method

The connection in Proposition 1 is asymptotic, and a natural question is why an asymptotic identification should motivate a finite procedure at all. The answer is that the proposition does more than describe a limit: it locates the classical double bootstrap on the boundary of a family.

The double bootstrap sits at a corner in two coordinates: base measures matched to the empirical distributions, and infinite concentration. The coordinates are distinct dials—the base decides *what* can enter the resampling, the concentration decides *how much* urn

reinforcement deforms it—and they are not interchangeable: sending both concentrations to infinity with an informative base held fixed does not approach the double bootstrap at all, since the total weight of the data in the second level is $n/(n + \beta)$ and the law concentrates on the prior instead. Working in the interior is therefore a deliberate retreat from the corner along both coordinates: informative atoms in the base, and finite concentrations that leave them influence, while the two-level architecture is retained. We refer to the resulting methodology as the *Informative Bayesian Double Bootstrap* (IBDB).

Structurally, the proposal consists of two Dirichlet levels stacked hierarchically, joined by the cut of Section 2.3. The outer level is a Dirichlet process $DP(\alpha, G)$ on the analyst's base measure, realized in the finite form (6) as a weight vector w_1 over the q atoms of G . A second Dirichlet process $DP(\beta, F_1)$, anchored on the observed data, takes that realization as its base, so the prior atoms enter the resampling distribution with random concentration βw_{1k} , re-drawn at every replicate. From the resulting weights, an informative synthetic sample is generated. A separate calibration loop places a flat Dirichlet distribution over the empirical distribution of that realized synthetic sample, with no return path to the outer atoms.

The two generative levels are genuine Dirichlet resampling steps, of the same two-level form as the family of Proposition 1, whose matched empirical members have the classical double bootstrap as their joint infinite-concentration endpoint. But the levels play asymmetric roles by design. Information enters only at the outer level, through G and the empirical-Bayes-regulated concentration of Section 3.3; the calibration loop is kept at its non-informative endpoint and preserves intact the calibration logic of the double bootstrap. The remainder of this section makes the asymmetry operational and records the evidence that settles it.

One point of scope should be stated plainly, to avoid claiming more than is proved. The endpoint propositions of Section 2.2 concern the *matched empirical* two-level family, in which both base measures are empirical, and identify the classical double bootstrap as its joint infinite-concentration endpoint. The IBDB is obtained from that family by replacing the first empirical base measure with the informative G while retaining the same second-level architecture. The propositions therefore *motivate* the construction—they establish the boundary it retreats from and the rate at which it does so—but they are not themselves a coverage theorem for the IBDB at any fixed finite concentration. That the retreat buys coverage, and where, is a matter for the simulations of Section 4 and the grid of Section 4.4, not for the bounds of Section 2.2. One further distinction belongs here. The propositions are stated for the sequential Blackwell–MacQueen scheme, which is the representation the coupling argument acts on; the procedure implemented and studied below is the finite Dirichlet construction (6). At matched empirical bases the two share the same infinite-concentration endpoint, which is what ties the method to the family the propositions describe, but they are distinct representations and we do not claim that the bounds transfer verbatim to the finite form at fixed q . What the propositions establish is that the family is well defined and that the classical double bootstrap sits on its boundary—the reason for looking in its interior at all. The finite-concentration results of Section 3.5 extend the analytical picture to the IBDB itself, but the same distinction applies to them: they are distributional proximity and support statements, not coverage guarantees—with the single exception of the first-order validity result for the mean (Proposition 6), which certifies asymptotic correctness precisely where the method claims no advantage—and the nominal-coverage evidence at finite concentration remains the factorial study of Section 4.

The design of the inner level rests on the calibration reading of the double bootstrap, which is worth stating in intuitive terms because it dictates why that level must stay non-informative. A confidence procedure built from a single resampling level rarely

attains its nominal coverage: an interval advertised as 95% may cover somewhat less. The role of a second, inner level is to measure that discrepancy and correct it. In the prepivoting formulation of Beran [16], the inner level transforms the confidence root by its own bootstrap distribution function, pushing it toward the uniform law it would follow if it were exactly pivotal. Hall and Martin [17] recast the same operation as the empirical solution of an equation in the quantity of interest so that iterating the resampling step drives coverage error down by a further factor at each pass. Read this way, the outer level estimates and the inner level gauges how far that estimate's coverage has strayed and adjusts accordingly. This is precisely why the informative construction that follows loads prior information onto the outer level alone: the inner level performs calibration rather than information incorporation, and an informative base measure there would corrupt the very function it exists to serve.

The IBDB is designed to preserve two features simultaneously. First, it incorporates prior information through a base measure G and finite concentration parameters. Second, it retains the inner calibration layer that distinguishes the double bootstrap from a one-level bootstrap or Bayesian bootstrap.

Let $x_1, \dots, x_n$ denote the observed data and let $T(P)$ be the target functional, such as a quantile, standard deviation, skewness, or value-at-risk. Let G be an informative base distribution or discrete approximation to prior knowledge. In applications to tail inference, G may encode information about tail shape, tail location, or extreme quantiles obtained from external data, theory, or regulation.

A practical version of the IBDB is stated as Algorithm 1.

The design question that the algorithm leaves visually open—whether the inner level is a genuine finite-concentration Dirichlet step or a non-informative calibration loop—is settled, empirically and deliberately, in favor of the latter. The inner level is a flat Dirichlet (Rubin) resampling loop over the synthetic sample generated from each outer draw, and the level γ^* is searched within a bounded band around the nominal level. The coverage target of that loop is the outer statistic of the same replicate: the synthetic sample $y^{(b)}$ is drawn from the realized $F_2^{(b)}$, so within replicate b the quantity $T(F_2^{(b)})$ plays the role that the unknown $T(F)$ plays in the real problem, and it is that quantity the inner interval must cover. Calibrating instead against a single sample-level proxy $T(\hat{F}_n)$ would ask the inner interval to cover a value that is not the truth of the world it was generated from. In the language of the family of Proposition 1, only the outer level moves away from the infinite-concentration endpoint; the inner level stays at its non-informative endpoint because its job is calibration, not information. Two independent implementations of the alternative—an informative inner level, whether drawing its own Dirichlet weights over the prior atoms or carrying a second informative base measure—degraded coverage in the tail-quantile benchmark (Section 4). What is at issue there is the *informativeness* of the inner level, not the choice of calibration target described above. The mechanism is that a synthetic world built from the prior is systematically less variable than the sampling process it stands in for, so calibration performed inside it is optimistic and the resulting intervals are too narrow. The asymmetry is therefore between the two *roles*, not between two kinds of objects: the two generative levels (F_1 and F_2) are both genuine Dirichlet steps, and it is the separate calibration loop that stays at the non-informative endpoint.

The algorithm above intentionally separates prior incorporation from calibration. The outer level introduces information through G and α . The inner level estimates the amount of calibration required, as in the classical double bootstrap. This separation is the methodological reason for using a two-level construction rather than merely a one-level informative Bayesian bootstrap.

3.2. The Role of the Empirical Distribution

The empirical distribution appears inside the base measure of the second level of (5), and it is worth being explicit about why it is there, since this is the point at which a reader may suspect that the data have been given the standing of prior belief. It enters there as the estimated component of a data-dependent base measure—the empirical Bayes reading developed below—and its mixing proportions are not free because the same law is produced by conjugacy. If $F_2 \mid F_1 \sim \text{DP}(\beta, F_1)$ and $x_1, \dots, x_n$ are conditionally independent draws from F_2 , then the conjugacy of the Dirichlet process [6] gives

$$F_2 \mid F_1, x_1, \dots, x_n \sim \text{DP}\left(\beta + n, \frac{\beta}{\beta + n} F_1 + \frac{n}{\beta + n} \hat{F}_n\right), \quad (7)$$

which is the second line of (5). The mixture is therefore forced rather than posited: given β , the weights $\beta/(\beta + n)$ and $n/(\beta + n)$ are determined by the conjugacy formula, so the method carries no separate mixing parameter for the balance between the elicited and the empirical part of that base measure—the analyst's only dial is β itself, set here by the empirical Bayes rule. The finite form (6) inherits the same structure, with unit concentration on each observation against concentration βw_{1k} on the prior atoms.

The same law admits a second, equivalent reading, and it is this reading that fixes the status of the construction: taken directly rather than as an update, Equation (7) is a Dirichlet process whose base measure is built partly from the data—a data-dependent prior, which is the defining move of empirical Bayes. This is precisely how Lo [4] reads the Bayesian bootstrap of Rubin: not as an update of anything, but as a Dirichlet process with empirical base measure. In a Dirichlet process $\text{DP}(\alpha, H)$, both arguments are hyperparameters of the prior: the concentration α and the base measure H . Estimating hyperparameters from the data is the defining move of empirical Bayes, and it does not become something else when the hyperparameter happens to be a measure rather than a scalar. The IBDB estimates two. The concentration is set by the discrepancy rule of Section 3.3. The base measure of the second level is the mixture $\frac{\beta}{\beta + n} F_1 + \frac{n}{\beta + n} \hat{F}_n$, in which the informative part is elicited—it descends from the analyst's G through the realization of the upper level—and the remaining part is estimated by the empirical distribution of the sample. The method is therefore an empirical Bayes procedure with a partly elicited, partly estimated base measure. The two readings describe the same distribution: conjugacy explains why the weights of the mixture are not free, and the empirical Bayes reading names what kind of object the procedure is.

The precedent is direct rather than analogical. The Bayesian bootstrap of Rubin [3] is recovered as a Dirichlet process with empirical base measure and vanishing concentration, a correspondence made precise by Lo [4]. The proper Bayesian bootstrap of Muliere and Secchi [5] combines an elicited base measure with the empirical distribution in exactly the proportion $\frac{k}{n+k}$ against $\frac{n}{n+k}$ that appears at the second level here. And the posterior bootstrap of Fong et al. [13] draws pseudo-samples from a centering measure alongside the observed data. In none of these is the empirical distribution treated as prior belief elicited before seeing the data; in all of them, it enters as an estimate of the part of the base measure the analyst does not wish to specify.

What we do not claim is exactness. The construction is not presented as a sampler for the posterior of a hierarchical Dirichlet model, and the reader should not read the conditioning bar in (5) as asserting one. Even with the structural cut of Section 2.3 in place, exact posterior inference for the pair (F_1, F_2) would require marginalizing the assignment of each observation to an atom of F_1 , a latent configuration whose summation does not collapse and which would return the procedure to Markov chain Monte Carlo.

The data are therefore used twice: once to anchor the second level through (7), and once to select the concentration through the rule of Section 3.3. This duplication is intrinsic to

empirical Bayes and is the classical objection raised against it; we record it rather than argue it away. It is also the reason the claims defended in this paper are frequentist coverage properties, established by the simulations of Section 4, and not the exactness of a posterior.

The concern is not new, and neither is the remedy. Laird and Louis [18] observed that intervals read off a prior estimated from the data are systematically too short because they ignore the uncertainty in that estimate and proposed bootstrap schemes that resample from the compound model to repair the coverage; Carlin and Gelfand [19] developed a conditional bias correction of the same intervals. Their setting is a compound model with many units, where the object of inference is a unit-specific parameter and the estimated prior governs shrinkage across units. The present setting is different—a single sample, a functional of one unknown distribution, and a prior that enters as a base measure rather than as a shrinkage target—but the lineage is the same: a data-estimated prior does not deliver nominal coverage on its own, and resampling is the instrument that restores it. In the IBDB, that instrument is the inner Dirichlet level.

3.3. Empirical Bayes Specification

In practice, the concentration parameters determine how much the base measure influences the procedure. If the base measure G is well specified, larger concentration can improve tail inference by supplying information that the sample may not contain. If G is badly specified, excessive concentration can poison the resampling distribution. The method therefore requires a principled or at least transparent rule for regulating prior influence.

The problem is not new, and the priority belongs to Hjort [20]. His treatment appeared as an Oslo technical report in May 1991 and circulated only informally for three decades; it resurfaced in 2026, when it was posted to arXiv, and the present paper is written with the benefit of that reappearance. In a report that anticipates much of the modern literature on the Bayesian bootstrap with a non-degenerate Dirichlet prior, Hjort poses exactly the question addressed here: one may have a candidate base measure from previous similar data without knowing that the new data come from the same distribution. One then wants a small concentration if the data disagree with that candidate and a larger one if they appear to fit. He also supplies a solution that is more principled than the heuristic below. Using the identity $\mathbb{E} \int (\hat{F}_n - F_0)^2 dW = n^{-1} \{1 + (n-1)/(a+1)\} \int F_0(1-F_0) dW$, valid for every weight measure W , the concentration a can be fitted by moment matching; a maximum likelihood estimator is also available, whose sufficient statistic is, tellingly for the present paper, the number of distinct data points—the same quantity K_n that governs the second-order constant of Proposition 3. We state at the outset, therefore, that the empirical Bayes component of the IBDB is not where its novelty lies, and that a reader seeking a principled rule for a *one-level* informative Dirichlet bootstrap should consult Hjort [20] first.

The empirical Bayes rule used throughout the experiments of this paper is

$$\hat{\alpha} = \text{clip} \left(\frac{10}{1 + \|d(x) - d(G)\|}, 0.3, 20 \right), \quad d = (\text{median}, \text{IQR}).$$

In every implementation reported in this paper, the two concentrations are coupled, $\beta = \hat{\alpha}$, so this single fit regulates both levels; the decoupled two-dimensional rule is listed among the refinements of Section 5.4. Here, $d(x)$ is computed from the data and $d(G)$ from the atoms or summaries of the base measure. The rule is intentionally simple: it increases the prior weight when the base measure agrees with the observed sample and decreases it when the base measure is incompatible with the sample. It is honest in the sense that it uses no oracle information about the true sampling distribution.

This rule should be understood as a working heuristic rather than a final empirical Bayes theory, and it is deliberately of the same qualitative type as the moment rule of Hjort [20]: both shrink the concentration in proportion to a discrepancy between the base measure and the sample. Ours substitutes a robust two-number summary—median and interquartile range—for Hjort’s L_2 functional, which buys resistance to the heavy tails that are the home ground of this paper at the price of theoretical grounding. Its strengths and limitations are clear. It reacts well to location discrepancies and is therefore useful for some tail-quantile problems. It is less effective for scale or skewness targets because median and interquartile range are not tailored to second- or third-moment behavior.

Two refinements follow, and they are of different orders. The first is to import Hjort’s moment rule, or his maximum likelihood estimator, into the present two-level setting. This is not immediate, because both were derived for a single Dirichlet level, and the object to be regulated here is a pair (α, β) in which the second concentration governs a process whose base is itself random. Establishing what the moment identity becomes when F_0 is replaced by a realization of $DP(\alpha, G)$ is the natural first step, and we do not take it here. The second refinement is target matching: a rule whose discrepancy measure is tuned to the functional being estimated, which neither Hjort’s rule nor ours attempts. The competence map of Section 4.11 shows why this matters, since the blind spots of our rule are exactly the statistics whose behavior median and interquartile range cannot see.

A second limitation is structural and connects to the finite-dimensional theory of Feng and Paguyo [11]. The influence a finitely represented prior can exert on a tail functional is governed not by α against n globally, but by the prior mass placed in the region that matters—for an upper quantile at level p , roughly αp effective prior atoms against np tail observations. A clipping range for $\hat{\alpha}$ that is fixed once and for all therefore caps the achievable tail influence regardless of how good G is; a target-matched rule should scale its upper limit with the local information np rather than with a universal constant. Deriving a principled empirical Bayes criterion—from a marginal likelihood or a predictive scoring rule, possibly with separate rules for the two concentrations—is left as the main methodological refinement for future work.

3.4. Why the Hierarchical Construction?

A natural objection is that one might not need a hierarchical construction at all. Why not use a one-level Dirichlet process with a highly informative base measure? If the goal is simply to incorporate prior information, a one-level proper Bayesian bootstrap [5] already seems to provide a route:

$$P \mid x_1, \dots, x_n \sim DP\left(\alpha + n, \frac{\alpha}{\alpha + n}G + \frac{n}{\alpha + n}\hat{F}_n\right).$$

This posterior has prior weight $\alpha/(\alpha + n)$ and empirical weight $n/(\alpha + n)$. Thus, a sufficiently large α can make G influential.

The question is therefore not whether a one-level Dirichlet process can incorporate prior information. It can. The question is whether it can reproduce the two roles of the double bootstrap. A one-level DP has one resampling layer. It can provide an informative posterior or predictive distribution, but it does not naturally separate prior incorporation from coverage calibration. The double bootstrap, in contrast, has two levels precisely because the second level is used for calibration.

The argument for the hierarchical construction is therefore as follows. The outer level controls the information entering through G . The inner level controls the calibration behavior inherited from the double bootstrap. With two concentration parameters, α and β ,

these roles can in principle be separated. In a one-level DP, both roles are collapsed into a single concentration parameter and a single random probability measure.

Construction	Prior information	Calibration layer
One-level DP/proper Bayesian bootstrap	Yes	No calibration layer
Classical double bootstrap	No informative prior	Yes
IBDB/two-level hierarchical construction	Yes	Yes

A related claim is sometimes made informally: that the base measure of a Dirichlet process has limited influence unless the concentration parameter is large, because the posterior base measure weights G by $\alpha/(\alpha + n)$. This is true as a statement about posterior weighting, but it should not be overstated: G can be highly influential when α is large, and its influence is target-dependent. The stronger and more defensible argument for the present paper is therefore not that G is always weak in a one-level DP, but that a one-level DP does not provide the two-stage calibration architecture of the double bootstrap.

Section 4 carries out exactly this comparison, and the answer is more interesting than a simple failure of the one-level procedure. Given the same base measure and the same empirical Bayes concentration, a correctly implemented one-level informative bootstrap is a strong competitor: it crosses the support ceiling and reaches high coverage in many cells. What it cannot do is calibrate. It reaches that coverage by widening the interval—over-covering at about 1.2 times the frequentist width for the value-at-risk (see the value-at-risk study of Section 4.6)—whereas the two-level construction attains uniformly higher coverage at comparable width on the expected shortfall, where the quantile support ceiling has been lifted by design. The separation of roles, not the informative base measure, is what the second level supplies.

3.5. Analytical Estimates at Finite Concentration

Propositions 1–3 quantify the boundary of the family: they bound the distance between the two-level scheme with *empirical* base measures and the classical double bootstrap. It is natural to ask what can be established analytically about the IBDB itself—the scheme with an informative base measure and finite, empirically regulated concentrations. Two results answer this at the level of the generative predictive scheme, with the same elementary tools and explicit constants as Proposition 1.

Let $P_{\alpha,\beta}^G$ denote the law of the pair $(\Phi^{(1)}, \Phi^{(2)})$ generated by the predictive form of the IBDB construction: $\Phi^{(1)}$ is drawn from the Blackwell–MacQueen urn with concentration α and base G , and, conditionally on $\Phi^{(1)}$, the second sample $\Phi^{(2)}$ is drawn from the urn with concentration $\beta + n$ and base

$$H(\Phi^{(1)}) = \frac{\beta \hat{F}_{\Phi^{(1)}} + n \hat{F}_n}{\beta + n},$$

the empirical Bayes base measure of Section 3.3 with the realized level-one distribution represented by its empirical measure $\hat{F}_{\Phi^{(1)}}$. Let Q^G denote its independence skeleton: $\Phi^{(1)}$ i.i.d. from G , then $\Phi^{(2)}$ i.i.d. from $H(\Phi^{(1)})$.

Proposition 4 (Finite-concentration proximity bounds for the IBDB scheme). *For all $\alpha, \beta > 0$ and every base measure G ,*

$$\begin{aligned} \text{(i)} \quad d_{\text{TV}}(P_{\alpha, \beta}^G, Q^G) &\leq \frac{n(n-1)}{2} \left(\frac{1}{\alpha} + \frac{1}{\beta+n} \right); \\ \text{(ii)} \quad d_{\text{TV}}(P_{\alpha, \beta}^G, P_{\text{DB}}) &\leq \underbrace{\frac{n(n-1)}{2} \left(\frac{1}{\alpha} + \frac{1}{\beta+n} \right)}_{\text{finite concentration}} + \underbrace{n d_{\text{TV}}(G, \hat{F}_n)}_{\text{informativeness gap}} + \underbrace{\frac{n^2}{\beta+n}}_{\text{EB anchoring}}. \end{aligned}$$

Proof. (i) The coupling in the proof of Proposition 1 nowhere uses the empirical character of the base measures: at step i of a level with concentration κ and base B , the urn conditional is a mixture that gives the base weight $\kappa/(\kappa+i-1)$, so a draw can be coupled with an i.i.d. draw from B except on an event of probability at most $(i-1)/(\kappa+i-1) \leq (i-1)/\kappa$. Applying this with $\kappa = \alpha$ at level one and $\kappa = \beta+n$ at level two and summing over steps gives the bound.

(ii) Chain through intermediate laws and apply the conditional coupling argument of Proposition 1. At level one, $d_{\text{TV}}(\text{urn}(\alpha, G), G^{\otimes n}) \leq n(n-1)/(2\alpha)$ by part (i), and $d_{\text{TV}}(G^{\otimes n}, \hat{F}_n^{\otimes n}) \leq n d_{\text{TV}}(G, \hat{F}_n)$. At level two, conditionally on the first-level outcome, $d_{\text{TV}}(\text{urn}(\beta+n, H), H^{\otimes n}) \leq n(n-1)/(2(\beta+n))$, and, since $H - \hat{F}_{\Phi(1)} = \frac{n}{\beta+n}(\hat{F}_n - \hat{F}_{\Phi(1)})$,

$$d_{\text{TV}}(H^{\otimes n}, \hat{F}_{\Phi(1)}^{\otimes n}) \leq n d_{\text{TV}}(H, \hat{F}_{\Phi(1)}) = \frac{n^2}{\beta+n} d_{\text{TV}}(\hat{F}_n, \hat{F}_{\Phi(1)}) \leq \frac{n^2}{\beta+n},$$

for every realization of $\Phi^{(1)}$. Combining the level-one and conditional level-two bounds yields (ii). $\square$

Part (i) is the exact analogue, for the proposed scheme, of the estimate that Proposition 1 provides for the classical one: the entire distance from the independence skeleton is a pure finite-concentration cost, with the same constants. Part (ii) makes the two-coordinate corner of Section 2.2 quantitative for the method itself, and the empirical Bayes rule of Section 3.3 acts on its first two terms in opposite directions: it raises $\hat{\alpha}$ —shrinking the finite-concentration cost—precisely when the summaries of G agree with the sample, which is when the informativeness gap is small.

The second result formalizes the separation that organizes the simulation study: an analytical upper bound on the coverage of the support-confined methods, paired with an exact finite-concentration formula showing that the IBDB is not subject to it. Let $x_{(n)} = \max_i x_i$ and let q_p denote the p -quantile of the continuous population F .

Proposition 5 (Support-ceiling separation). (i) *Let $[L, U]$ be any interval whose endpoints lie almost surely in $[\min_i x_i, x_{(n)}]$, as do the percentile intervals of the single and double bootstrap. Then its coverage of q_p satisfies*

$$P(L \leq q_p \leq U) \leq P(x_{(n)} \geq q_p) = 1 - p^n.$$

(ii) *For the IBDB with any finite $\alpha, \beta > 0$: conditionally on the level-one weights w_1 of Equation (6), the F_2 -mass above $x_{(n)}$ has the exact distribution*

$$F_2((x_{(n)}, \infty)) \mid w_1 \sim \text{Beta}(\beta V, n + \beta(1 - V)), \quad V = \sum_{k: g_k > x_{(n)}} w_{1k},$$

so that $\mathbb{E}[F_2((x_{(n)}, \infty))] = \frac{\beta}{\beta+n} \pi_G$ with $\pi_G = q^{-1\#}\{k : g_k > x_{(n)}\}$, and, for the p -quantile functional T_p ,

$$P(T_p(F_2) > x_{(n)} | w_1) = 1 - I_{1-p}(\beta V, n + \beta(1 - V)),$$

where I is the regularized incomplete beta function; this probability is strictly positive whenever $V > 0$. Consequently the IBDB is not support-confined, and the bound in (i) does not apply to it.

Proof. (i) The coverage event is contained in $\{U \geq q_p\} \subseteq \{x_{(n)} \geq q_p\}$, whose probability is $1 - F(q_p)^n = 1 - p^n$ for continuous F . (ii) Conditionally on w_1 , the joint weight vector of Equation (6) is Dirichlet with parameters $(1, \dots, 1, \beta w_{11}, \dots, \beta w_{1q})$; by the aggregation property of the Dirichlet distribution, the total mass of the atom subset $\{g_k > x_{(n)}\}$ is Beta with the stated parameters, whose mean is $\beta V / (\beta + n)$. Taking expectations and using $\mathbb{E}[V] = \pi_G$, which follows from the symmetry of $\text{Dirichlet}(\alpha/q, \dots, \alpha/q)$, gives the mean identity. Finally $\{T_p(F_2) > x_{(n)}\} = \{F_2((x_{(n)}, \infty)) > 1 - p\}$, and the Beta distribution has full support on $(0, 1)$ when both parameters are positive. $\square$

Corollary 1 (The ceiling for general functionals, and the expected shortfall). *Let $\theta = \theta(F)$ be any target and let $[L, U]$ be an interval whose upper endpoint satisfies $U \leq x_{(n)}$ almost surely, as do the percentile intervals of the single and double bootstrap for the mean, the quantiles and the expected shortfall (the statistic of an equal-weight resample never exceeds the largest resampled point). Then*

$$P(L \leq \theta \leq U) \leq P(x_{(n)} \geq \theta) = 1 - F(\theta^-)^n.$$

For the expected shortfall at level p , since $\text{ES}_p > q_p$, the ceiling $1 - F(\text{ES}_p)^n$ is strictly smaller than the quantile ceiling $1 - p^n$.

Proof. The argument of Proposition 5(i) applies verbatim with θ in place of q_p . $\square$

The corollary sharpens the reading of the risk study of Section 4.6. The design of Section 4.6 lifts the quantile ceiling, but the expected-shortfall ceiling is not lifted: at $n = 100$ it equals 0.784, 0.844, 0.799 and 0.754 for the t_3 , exponential, lognormal and Pareto(2.5) populations. Both Efron methods sit below it in all twelve expected-shortfall cells, as the corollary requires—part of their undercoverage on this functional is analytic necessity—while the IBDB crosses it in all four populations at $n = 100$ (0.813, 0.861, 0.826, 0.808): an exact bound obeyed by the support-confined methods cell by cell and exceeded by the proposed method exactly where its escape mass operates. For the mean the ceiling $1 - F(\mu)^n$ exists but rarely binds, since $F(\mu)$ is far from one except under extremely heavy tails—which is why no ceiling phenomenon appears in the location study of Section 4.2.

For one functional the analytical picture closes all the way to coverage, and it is instructive that it is the one where the method claims no advantage. The mean is linear in the resampling weights, which makes a first-order statement provable with classical tools.

Proposition 6 (First-order validity for the mean). *Let F have mean μ and finite variance $\sigma^2 > 0$, let G be fixed with finite mean and variance, and let $\alpha = \beta = \hat{\alpha}_n$ be bounded in probability, as the empirical Bayes rule of Section 3.3 guarantees by construction. Consider the IBDB interval for the mean formed by the percentiles, at fixed level $1 - \gamma$, of $T(F_2) = \sum_i W_i x_i + \sum_k W_{n+k} g_k$. Then, conditionally on the data, $\sqrt{n}(T(F_2) - \bar{x}_n) \Rightarrow N(0, \sigma^2)$ in probability, and the coverage of μ by that interval converges to $1 - \gamma$.*

Proof. Let $M = \sum_k W_{n+k}$ be the total prior mass. Conditionally on w_1 , the aggregation and neutrality properties of the Dirichlet distribution give $M \sim \text{Beta}(\beta, n)$, independent of the renormalized data weights $\tilde{W} \sim \text{Dirichlet}(1, \dots, 1)$, so $M = O_p(\hat{\alpha}_n/n) = O_p(1/n)$. Decompose $\sqrt{n}(T - \bar{x}_n) = (1 - M)\sqrt{n}(\sum_i \tilde{W}_i x_i - \bar{x}_n) + \sqrt{n}M(\bar{g}_w - \bar{x}_n)$, where $\bar{g}_w$ is the prior-atom average, bounded in L^1 since G has finite mean. The second term is

$O_p(n^{-1/2})$; the first converges conditionally to $N(0, \sigma^2)$ by the Bayesian-bootstrap central limit theorem of Lo [4], and Slutsky's theorem completes the conditional limit. The coverage statement follows by the standard bootstrap argument: the conditional quantiles equal $\bar{x}_n + \sigma z. / \sqrt{n} + o_p(n^{-1/2})$, and $\sqrt{n}(\bar{x}_n - \mu) \Rightarrow N(0, \sigma^2)$. $\square$

The proposition is a validity statement, not an advantage: at first order the prior washes out at rate $\hat{\alpha}_n / \sqrt{n}$, so for the mean the IBDB is asymptotically equivalent to the Bayesian and single bootstrap—the analytical face of the paper's message that the method has nothing to win for location. It is stated for the interval at fixed level; the calibrated procedure searches within a bounded band around the nominal level, and its finite-sample over-coverage in the location study of Section 4.2 dissolves toward nominal as n grows, consistent with the limit. Numerically, with a base measure displaced by 0.3 population standard deviations, the fixed-level coverage stays within Monte Carlo error of 0.90 for n from 25 to 1600. The same argument extends by the delta method to smooth functionals such as the variance under a finite fourth moment—a condition that fails precisely for the t_3 and Pareto populations where the finite-sample action of the scale study of Section 4.3 lives—and the systematic treatment of that regime is left to future work.

Analytical Versus Numerical Estimates

The statements above meet the simulations at three further points. The ceiling $1 - p^n$ appears as the "Ceil." column of the quantile study of Section 4.4, where the frequentist methods do not merely respect it but attain it; the same support logic yields the exact zeros of the skewness study of Section 4.5; and the exact formula of Proposition 5(ii) was verified against the implementation, reproducing the Monte Carlo frequency of the escape event to three decimal places (0.199 against 0.201, and 0.014 against 0.014, at 40,000 level-one draws in the lognormal cell with $n = 25$). Apart from Proposition 6, these are proximity and support statements, not coverage guarantees at finite concentration; the coverage evidence remains the factorial study of Section 4, and the distinction is taken up again in Section 5.

4. Simulation Study

This section evaluates the IBDB against a fixed panel of competitors across statistics of increasing demand on the tail of the distribution. The design is uniform: every table compares the same five methods, and only the target statistic, the population and the sample size vary. This uniformity is what allows the tables to be read against one another.

4.1. Design

Methods

Five procedures appear in every table.

- **Efron single** and **Efron double**: The percentile bootstrap of Efron [2] and its calibrated iterate [10], the frequentist baseline, carrying no prior information.
- **PBB**: The proper Bayesian bootstrap of Muliere and Secchi [5], implemented as they specify—a two-step plan that first resamples $m = n$ points from the posterior mixture $(kF_0 + nF_n)/(n + k)$ and then reweights them by a symmetric Dirichlet $((n + k)/m)$ —with G and the concentration set by the empirical Bayes rule of Section 3.3.
- **Posterior bootstrap**: The procedure of Fong et al. [13], the modern one-level informative method, given the same G and concentration.
- **IBDB**: The two-level hierarchical Dirichlet construction of Section 3.1, with both concentrations coupled and regulated by the same empirical Bayes rule.

The two Efron methods carry no prior. The PBB and the posterior bootstrap carry the same prior information as the IBDB but at a single Dirichlet level, so the contrast between them and the IBDB isolates the contribution of the second level.

Common Settings

All cells use $q = 100$ atoms for G , 1000 replications, nominal 90% central intervals, and seed 20260701. Every method uses $B = 250$ outer draws and $C = 25$ inner draws; the calibration search runs over a band Γ of nine equally spaced candidate levels in $[0.3\gamma, 1.7\gamma]$, and for the IBDB the calibration loop is evaluated on the first $M = \min(B, 60)$ outer replicates, which bounds its cost without measurable effect on the selected level. The maximum Monte Carlo standard error on a coverage of 0.9 is 0.009, and differences below about 0.02 should not be read as real. Statistics are presented in order of increasing demand on the tail: mean, median, standard deviation, upper quantiles, skewness, and the risk functionals.

4.2. Location: The Mean and the Median

We begin where informative resampling should help least. A sample mean is an efficient estimator with no support obstruction and nothing for a prior to supply, so a large advantage here would signal an artifact rather than a method.

Table 2 shows none. The Efron double bootstrap calibrates well, sitting between 0.90 and 0.96 in most cells, and the posterior bootstrap is the best calibrated and the narrowest method in these cells: with nothing for the prior to supply, its pseudo-sample construction behaves like an efficient one-level scheme. The PBB over-covers throughout, reaching 0.96–1.00, because its resampling step widens the interval. The IBDB over-covers as well, and for the same reason: its inner loop widens rather than sharpens when the proxy is already close to the truth. The practical conclusion is the one the introduction announced: the method is not for these statistics, and we report the cells rather than omit them.

4.3. Scale: The Standard Deviation

The standard deviation is the first statistic for which the tail matters, and Table 3 orders the populations by tail weight. The gradient is immediate. Under a normal population all methods are near or above nominal; as the tail thickens the frequentist methods fail badly, to 0.476 for the Efron double bootstrap under a lognormal at $n = 15$.

In the eight heavy-tailed cells (t_3 and lognormal) the IBDB attains the highest coverage, and the margin widens where the problem is hardest: 0.919 against the corrected PBB's 0.813 under a t_3 at $n = 15$, and 0.900 against 0.743 under a lognormal. Under the normal and the t_5 , where the sample lacks little, the IBDB over-covers like the other informative methods and the best-calibrated choices are, in most cells, the frequentist double bootstrap and the PBB—consistent with the gradient the design was built to expose. The properly implemented one-level PBB is a much stronger competitor under heavy tails than a naive version would be—it reaches 0.72–0.81 rather than collapsing—but it does not close the gap. The second level does.

4.4. Upper Quantiles

Upper quantiles are the home ground of the method, for an arithmetic reason. An interval confined to the convex hull of the sample cannot cover a population quantile that exceeds the largest observation, an event of probability p^n , so the coverage of any such method is bounded by

$$1 - p^n. \quad (8)$$

Table 4 reports this ceiling.

Table 2. Coverage of nominal 90% intervals for the mean and the median: coverage (cov) and mean interval width (w). 1000 replications per cell, $q = 100$, seed 20260701. The proper Bayesian bootstrap over-covers throughout, because its resampling step widens the interval where the sample already suffices; the posterior bootstrap, by contrast, is the best calibrated and the narrowest method in these cells. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Statistic	Population	n	Efron 1		Efron 2		PBB		Post. Boot.		IBDB	
			cov	w	cov	w	cov	w	cov	w	cov	w
Mean	Normal	10	0.841	0.96	0.918	1.24	0.996	1.24	0.931	0.75	0.997	1.16
		15	0.846	0.80	0.929	1.04	0.996	1.04	0.904	0.66	0.987	0.98
		25	0.863	0.63	0.938	0.80	0.987	0.83	0.898	0.55	0.981	0.77
		50	0.888	0.45	0.954	0.58	0.984	0.61	0.909	0.42	0.968	0.56
		100	0.876	0.32	0.934	0.41	0.971	0.45	0.899	0.31	0.950	0.40
t_3		10	0.833	1.53	0.925	1.97	0.997	1.98	0.931	1.23	0.993	2.02
		15	0.860	1.22	0.924	1.58	0.990	1.62	0.924	1.04	0.983	1.64
		25	0.863	1.00	0.941	1.29	0.987	1.33	0.890	0.88	0.985	1.33
		50	0.867	0.73	0.943	0.94	0.984	1.00	0.898	0.68	0.970	0.94
		100	0.869	0.54	0.944	0.69	0.972	0.75	0.878	0.52	0.948	0.70
	Lognormal	10	0.760	1.63	0.834	2.06	0.992	2.21	0.924	1.40	0.994	2.30
		15	0.775	1.41	0.857	1.81	0.981	1.89	0.887	1.23	0.989	1.94
		25	0.784	1.18	0.873	1.52	0.970	1.57	0.848	1.06	0.972	1.59
		50	0.847	0.88	0.920	1.13	0.965	1.19	0.883	0.81	0.962	1.14
		100	0.861	0.66	0.930	0.85	0.961	0.91	0.882	0.63	0.952	0.84
	Exponential	10	0.803	0.90	0.880	1.15	0.998	1.18	0.927	0.72	0.998	1.14
		15	0.848	0.77	0.914	0.99	0.995	1.01	0.922	0.64	0.994	0.97
		25	0.864	0.61	0.935	0.79	0.992	0.81	0.905	0.54	0.978	0.77
		50	0.876	0.45	0.942	0.58	0.982	0.61	0.898	0.41	0.966	0.56
		100	0.868	0.32	0.932	0.41	0.968	0.44	0.872	0.31	0.949	0.39
Median	Normal	10	0.875	1.20	0.955	1.56	0.999	1.63	0.944	1.01	0.980	1.34
		15	0.888	1.03	0.912	1.17	0.992	1.34	0.933	0.87	0.966	1.12
		25	0.881	0.80	0.926	0.94	0.990	1.07	0.918	0.71	0.966	0.90
		50	0.884	0.57	0.951	0.71	0.982	0.78	0.910	0.53	0.945	0.66
		100	0.883	0.40	0.947	0.50	0.983	0.56	0.901	0.39	0.949	0.47
t_3		10	0.860	1.40	0.936	1.93	0.995	1.92	0.932	1.14	0.972	1.54
		15	0.882	1.19	0.911	1.35	0.994	1.57	0.937	0.98	0.974	1.28
		25	0.874	0.90	0.929	1.05	0.991	1.23	0.916	0.79	0.957	1.00
		50	0.890	0.63	0.953	0.78	0.986	0.87	0.911	0.59	0.957	0.73
		100	0.890	0.44	0.944	0.55	0.982	0.62	0.902	0.43	0.943	0.52
	Lognormal	10	0.885	1.41	0.959	1.92	0.995	1.92	0.944	1.11	0.983	1.52
		15	0.875	1.14	0.901	1.32	0.993	1.53	0.913	0.95	0.965	1.23
		25	0.884	0.85	0.922	0.99	0.996	1.16	0.920	0.74	0.961	0.95
		50	0.877	0.59	0.932	0.74	0.978	0.82	0.894	0.55	0.941	0.68
		100	0.891	0.41	0.963	0.51	0.987	0.57	0.905	0.39	0.952	0.48
	Exponential	10	0.866	0.99	0.944	1.32	0.997	1.34	0.945	0.80	0.988	1.10
		15	0.877	0.84	0.910	0.96	0.992	1.11	0.927	0.69	0.971	0.91
		25	0.882	0.66	0.934	0.77	0.986	0.89	0.917	0.58	0.959	0.73
		50	0.905	0.46	0.953	0.58	0.987	0.64	0.923	0.43	0.960	0.53
		100	0.874	0.32	0.944	0.40	0.986	0.45	0.892	0.31	0.950	0.38

Table 3. Coverage of nominal 90% intervals for the population standard deviation, ordered by tail weight; coverage (cov) and mean width (w). 1000 replications, $q = 100$. In the eight heavy-tailed cells the IBDB attains the highest coverage, by margins that widen as the tail thickens; under the normal and the t_5 it over-covers like the other informative methods. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Population	n	Efron 1		Efron 2		PBB		Post. Boot.		IBDB	
		cov	w	cov	w	cov	w	cov	w	cov	w
Normal	15	0.795	0.53	0.870	0.68	0.969	0.66	0.889	0.42	0.991	0.62
	25	0.794	0.42	0.865	0.54	0.944	0.54	0.835	0.36	0.968	0.52
	50	0.819	0.31	0.901	0.40	0.942	0.41	0.845	0.28	0.961	0.38
	100	0.845	0.22	0.922	0.29	0.952	0.30	0.867	0.21	0.957	0.28

Table 3. Cont.

Population	<i>n</i>	Efron 1		Efron 2		PBB		Post. Boot.		IBDB	
		cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>
t_5	15	0.639	0.79	0.726	0.99	0.907	1.04	0.803	0.69	0.968	1.08
	25	0.730	0.70	0.816	0.88	0.919	0.90	0.793	0.61	0.964	0.90
	50	0.772	0.55	0.863	0.70	0.903	0.72	0.794	0.50	0.932	0.71
	100	0.816	0.43	0.888	0.54	0.919	0.57	0.822	0.40	0.912	0.54
t_3	15	0.512	1.20	0.581	1.46	0.813	1.62	0.693	1.16	0.919	1.86
	25	0.591	1.06	0.660	1.32	0.787	1.42	0.661	1.03	0.887	1.59
	50	0.610	0.94	0.694	1.16	0.777	1.23	0.651	0.88	0.846	1.28
	100	0.644	0.80	0.727	1.00	0.777	1.04	0.666	0.76	0.811	1.06
Lognormal	15	0.442	1.72	0.476	2.02	0.743	2.47	0.673	1.80	0.900	2.88
	25	0.454	1.61	0.505	1.92	0.721	2.20	0.611	1.61	0.862	2.47
	50	0.557	1.52	0.625	1.87	0.732	1.98	0.611	1.43	0.818	2.08
	100	0.642	1.40	0.719	1.73	0.770	1.80	0.674	1.29	0.824	1.78

Table 4. Coverage of nominal 90% intervals for upper quantiles. “Ceiling” is the distribution-free bound $1 - p^n$. The frequentist methods attain the ceiling; the informative methods cross it, and the IBDB crosses furthest in the extreme cells ($p = 0.99$). Coverage (cov) and mean width (*w*). 1000 replications, $q = 100$. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Population	<i>n</i>	<i>p</i>	Efron 1		Efron 2		PBB		Post. Boot.		IBDB		Ceil.
			cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	
Normal	20	0.9	0.857	1.22	0.881	1.40	0.979	1.43	0.887	1.07	0.975	1.35	0.878
	20	0.95	0.625	1.06	0.628	1.21	0.944	1.41	0.832	0.95	0.987	1.49	0.642
	20	0.99	0.184	0.80	0.185	1.00	0.507	1.35	0.645	1.01	0.688	1.21	0.182
	40	0.9	0.819	0.80	0.913	1.06	0.978	1.14	0.884	0.80	0.938	0.98	0.985
	40	0.95	0.842	1.09	0.861	1.24	0.923	1.23	0.845	0.97	0.940	1.16	0.871
	40	0.99	0.328	0.78	0.328	0.93	0.589	1.14	0.586	0.82	0.738	1.12	0.331
	80	0.9	0.879	0.61	0.950	0.77	0.980	0.85	0.902	0.59	0.951	0.71	1.000
	80	0.95	0.835	0.72	0.920	0.93	0.967	1.02	0.893	0.74	0.928	0.87	0.983
	80	0.99	0.567	0.81	0.569	0.93	0.681	0.97	0.605	0.68	0.811	0.99	0.552
	80	0.99	0.567	0.81	0.569	0.93	0.681	0.97	0.605	0.68	0.811	0.99	0.552
t_3	20	0.9	0.843	3.00	0.867	3.27	0.980	3.37	0.869	2.69	0.971	3.32	0.878
	20	0.95	0.636	2.82	0.640	3.02	0.948	3.69	0.830	2.80	0.978	4.73	0.642
	20	0.99	0.187	2.43	0.188	2.73	0.494	4.38	0.629	4.68	0.701	5.36	0.182
	40	0.9	0.847	1.50	0.918	1.99	0.979	2.24	0.886	1.56	0.947	1.93	0.985
	40	0.95	0.844	2.97	0.868	3.25	0.927	3.28	0.843	2.85	0.939	3.31	0.871
	40	0.99	0.324	2.46	0.324	2.77	0.573	3.93	0.578	3.44	0.754	5.10	0.331
	80	0.9	0.884	1.10	0.934	1.40	0.973	1.58	0.906	1.08	0.957	1.33	1.000
	80	0.95	0.833	1.75	0.917	2.35	0.959	2.82	0.889	1.97	0.925	2.43	0.983
	80	0.99	0.553	3.54	0.558	3.82	0.660	4.15	0.595	3.32	0.812	5.14	0.552
	80	0.99	0.553	3.54	0.558	3.82	0.660	4.15	0.595	3.32	0.812	5.14	0.552
Lognormal	20	0.9	0.865	5.51	0.888	5.99	0.975	6.21	0.888	5.29	0.966	6.31	0.878
	20	0.95	0.643	5.30	0.647	5.63	0.949	7.09	0.830	5.36	0.985	9.62	0.642
	20	0.99	0.190	4.60	0.190	5.15	0.472	8.58	0.614	8.85	0.693	10.47	0.182
	40	0.9	0.813	3.04	0.901	4.11	0.971	4.77	0.867	3.26	0.937	4.01	0.985
	40	0.95	0.853	6.60	0.869	7.16	0.917	7.22	0.848	6.34	0.937	7.25	0.871
	40	0.99	0.319	5.59	0.320	6.21	0.565	8.77	0.560	6.95	0.746	10.45	0.331
	80	0.9	0.890	2.25	0.950	2.85	0.977	3.27	0.911	2.24	0.952	2.76	1.000
	80	0.95	0.845	3.88	0.916	5.08	0.958	5.97	0.893	4.30	0.939	5.23	0.983
	80	0.99	0.541	7.07	0.543	7.71	0.650	8.49	0.570	6.49	0.793	10.34	0.552
	80	0.99	0.541	7.07	0.543	7.71	0.650	8.49	0.570	6.49	0.793	10.34	0.552

The frequentist methods do not merely respect it; they attain it. At $p = 0.99$, $n = 20$, under t_3 , the ceiling is 0.182 and both Efron methods cover 0.18–0.19; coverage estimates can sit slightly above the ceiling by Monte Carlo error alone, whose standard error at this level is 0.013. The informative methods cross the ceiling, because the atoms of G carry mass beyond $\max_i x_i$. Two patterns then separate the informative methods. In the extreme cells ($p = 0.99$) the IBDB crosses furthest—0.69–0.81 against the PBB’s 0.47–0.68 and the posterior bootstrap’s 0.56–0.65. In the moderate cells ($p = 0.90, 0.95$) the corrected PBB is strong but over-covers by inflation: 0.975 against a ceiling of 0.878 at $p = 0.90$, $n = 20$. The IBDB stays closer to nominal there. The distinction that matters is therefore not who

covers more but who calibrates: in the moderate cells a one-level informative method reaches high coverage by widening the interval, while in the extreme cells it cannot reach the IBDB at all.

4.5. Skewness

The skewness sharpens the support argument into a falsifiable prediction. The sample skewness is bounded by $(n - 2) / \sqrt{n - 1}$; when that bound lies below the population value, coverage of every resample-based method must be exactly zero. Table 5 reports the bound. For a lognormal whose skewness is 6.185, the bound is 3.47 at $n = 15$ and 4.69 at $n = 25$, and both Efron methods cover exactly 0.000. At $n = 50$ the bound clears the truth and coverage becomes positive. The transition occurs where the arithmetic says it must.

Table 5. Coverage of nominal 90% intervals for the population skewness. “Bound” is the largest attainable sample skewness $(n - 2) / \sqrt{n - 1}$; when it falls below the truth, no resample-based interval can cover. 1000 replications, $q = 100$. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Population	<i>n</i>	Truth	Efron 1		Efron 2		PBB		Post. Boot.		IBDB		Bound
			cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	cov	<i>w</i>	
Normal	15	0.000	0.945	1.51	0.957	1.73	1.000	1.93	0.936	1.22	0.992	1.78	3.47
	25	0.000	0.886	1.20	0.936	1.47	0.996	1.57	0.887	1.03	0.973	1.44	4.69
Gamma(4)	15	1.000	0.679	1.60	0.718	1.90	0.943	2.11	0.821	1.42	0.929	2.01	3.47
	25	1.000	0.692	1.32	0.778	1.63	0.930	1.73	0.799	1.17	0.922	1.64	4.69
	50	1.000	0.707	1.01	0.794	1.27	0.891	1.32	0.762	0.91	0.886	1.25	6.86
	100	1.000	0.742	0.84	0.838	1.05	0.887	1.07	0.735	0.75	0.869	0.99	9.85
Exponential	15	2.000	0.433	1.82	0.515	2.17	0.803	2.50	0.742	1.90	0.876	2.66	3.47
	25	2.000	0.491	1.63	0.587	2.02	0.790	2.19	0.693	1.61	0.845	2.23	4.69
	50	2.000	0.520	1.37	0.645	1.73	0.750	1.78	0.609	1.26	0.770	1.70	6.86
	100	2.000	0.568	1.16	0.672	1.46	0.743	1.49	0.614	1.05	0.767	1.41	9.85
Lognormal	15	6.185	0.000	2.03	0.000	2.41	0.006	3.18	0.149	3.51	0.199	4.46	3.47
	25	6.185	0.000	2.01	0.000	2.43	0.035	3.07	0.142	3.04	0.181	3.85	4.69
	50	6.185	0.011	2.11	0.024	2.61	0.072	2.95	0.125	2.55	0.166	3.26	6.86
	100	6.185	0.086	2.24	0.120	2.74	0.108	2.90	0.122	2.20	0.166	2.87	9.85

Away from that cell the gradient reappears, ordered by asymmetry. The IBDB attains the highest coverage in every exponential and lognormal cell—0.845 against the PBB’s 0.790 for the exponential at $n = 25$ —while under the milder asymmetry of the Gamma(4) the two are indistinguishable, the PBB holding a margin of 0.005 to 0.018 that is at or below the resolution of the design. The informative methods reach positive coverage in the lognormal cell only because the atoms of G carry mass beyond the sample.

4.6. Risk Functionals

The risk functionals are computed at $n \geq 100$, where the quantile support ceiling no longer binds, so an advantage here cannot be quantile-support extension; the tighter ceiling specific to the expected shortfall (Corollary 1) does still bind at $n = 100$, and it is examined there. Table 6 separates two mechanisms.

For the value-at-risk the corrected PBB reaches the highest coverage, 0.96–0.98, but at about 1.2 times the Efron width in every cell: it over-covers by inflation. The IBDB sits at 0.92–0.94, closer to nominal. For the expected shortfall the picture inverts and gives the paper its strongest result. The IBDB attains the highest coverage in all twelve cells—0.81–0.90 against every competitor—at widths within ten percent of the PBB’s, and smaller in four cells. Under a Pareto population at $n = 100$ the IBDB covers 0.808 against the PBB’s 0.701 with an interval ten percent wider—far from the inflation of fifteen to twenty percent by which the PBB buys its value-at-risk coverage. The gain appears where repositioning shows: in the correct-prior rows of the directional diagnostic (Section The Directional Artifact, Diagnosed), the fraction of intervals lying entirely below the truth

falls from the PBB's 0.243 and 0.273 to the IBDB's 0.174 and 0.193, and on the Siemens data (Section 4.10), the IBDB covers more at strictly smaller width than the calibrated double bootstrap. Coverage that rises while the below-truth failures fall, at nearly unchanged width, is repositioning, not inflation. With the ceiling lifted, what remains scarce for the expected shortfall is not the support but the information—roughly $n(1 - p)$ effective tail observations—and the prior supplies it exactly there while the calibration loop converts prudence into coverage.

Table 6. Coverage of nominal 90% intervals for value-at-risk and expected shortfall at level 0.95, with the quantile support ceiling lifted ($1 - 0.95^n > 0.99$). The tighter expected-shortfall ceiling of Corollary 1 is not lifted at $n = 100$: both Efron methods remain below it in every cell and the IBDB crosses it in all four populations. For VaR the PBB over-covers by inflation (intervals about 1.2 times the Efron width); for ES the IBDB attains the highest coverage in every cell at widths within ten percent of the PBB's, and its gain shows as a fall in below-truth failures rather than as widening. Coverage (cov) and mean width (w). 1000 replications, $q = 100$. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Statistic	Population	n	Truth	Efron 1		Efron 2		PBB		Post. Boot.		IBDB	
				cov	w	cov	w	cov	w	cov	w	cov	w
VaR _{0.95}	t_3	100	2.35	0.874	1.58	0.922	2.02	0.972	2.47	0.888	1.55	0.937	2.00
		150	2.35	0.867	1.26	0.932	1.62	0.974	1.91	0.894	1.32	0.942	1.59
		200	2.35	0.861	1.07	0.933	1.40	0.973	1.63	0.888	1.14	0.933	1.37
	Exponential	100	3.00	0.878	1.41	0.923	1.79	0.968	2.04	0.898	1.36	0.941	1.70
		150	3.00	0.861	1.11	0.925	1.43	0.969	1.65	0.882	1.14	0.935	1.39
		200	3.00	0.863	0.96	0.938	1.23	0.971	1.40	0.886	0.98	0.931	1.17
	Lognormal	100	5.18	0.842	3.56	0.896	4.48	0.963	5.52	0.867	3.55	0.926	4.61
		150	5.18	0.861	2.90	0.925	3.73	0.959	4.45	0.886	3.06	0.919	3.73
		200	5.18	0.847	2.41	0.926	3.12	0.967	3.67	0.884	2.52	0.926	3.07
	Pareto(2.5)	100	3.31	0.849	1.91	0.910	2.50	0.962	3.07	0.861	1.92	0.929	2.55
		150	3.31	0.871	1.55	0.924	2.02	0.964	2.45	0.884	1.66	0.932	2.02
		200	3.31	0.864	1.35	0.936	1.78	0.966	2.10	0.885	1.41	0.922	1.72
ES _{0.95}	t_3	100	3.88	0.669	2.83	0.712	3.44	0.728	3.39	0.665	2.51	0.813	3.57
		150	3.88	0.695	2.34	0.769	2.97	0.784	3.06	0.714	2.25	0.840	3.16
		200	3.88	0.724	2.22	0.810	2.86	0.810	2.85	0.724	2.08	0.842	2.89
	Exponential	100	4.00	0.711	1.62	0.777	2.03	0.790	2.05	0.711	1.48	0.861	2.03
		150	4.00	0.773	1.43	0.844	1.83	0.864	1.88	0.803	1.35	0.902	1.83
		200	4.00	0.800	1.29	0.871	1.67	0.872	1.69	0.802	1.21	0.902	1.63
	Lognormal	100	8.56	0.656	5.83	0.718	7.17	0.743	7.12	0.669	5.26	0.826	7.39
		150	8.56	0.716	5.07	0.797	6.44	0.823	6.60	0.745	4.86	0.859	6.73
		200	8.56	0.761	4.77	0.841	6.09	0.850	6.12	0.776	4.44	0.884	6.09
	Pareto(2.5)	100	5.53	0.625	4.09	0.687	5.01	0.701	4.87	0.645	3.65	0.808	5.33
		150	5.53	0.662	3.51	0.727	4.42	0.754	4.51	0.695	3.36	0.814	4.80
		200	5.53	0.702	3.20	0.790	4.10	0.790	4.05	0.696	2.98	0.822	4.18

Two facts about the competitors belong here. The posterior bootstrap of 2019 crosses the support ceiling but undercovers throughout, and in the heaviest cells falls below even the Efron double bootstrap (0.669 against 0.718 for the lognormal expected shortfall at $n = 100$): its single calibration level cannot correct the coverage its generative level delivers. And the corrected PBB, far from being a weak straw man, is the strongest one-level competitor—which is precisely why the IBDB's dominance on the expected shortfall, at smaller width, is the cleanest evidence that the second level does work that no one-level method reproduces.

4.7. Misspecification of the Base Measure

Every table so far gave the informative methods a correct base measure. Table 7 displaces G by six interquartile ranges above and below the data and reports width alongside coverage. The width column is essential: for the standard deviation and the quantile, an upward displacement raises coverage while inflating the interval by factors of up to four, which is enlargement, not inference. Read through the width, the adversarial

direction for an upper quantile is downward, where a wrong prior returns the IBDB to the neighborhood of Efron at a width below one. For the skewness a downward displacement is catastrophic, and the empirical Bayes rule does not detect it because atoms compressed near zero are less discrepant in median and interquartile range than atoms shifted upward. This is the sharpest failure mode of the method, and we state it as such.

Table 7. Misspecification: coverage (cov) and width relative to the Efron double bootstrap (w) when G is displaced six interquartile ranges above (+6) or below (−6) the data. Coverage must be read with width: an apparent gain at large w is inflation. 1000 replications, $q = 100$. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

Statistic	Population	n	G	Efron 1		Efron 2		PBB		Post. Boot.		IBDB	
				cov	w	cov	w	cov	w	cov	w	cov	w
mean	Lognormal	25	+6	0.828	0.78	0.895	1.00	0.988	1.46	0.880	1.03	0.945	1.38
			−6	0.805	0.78	0.870	1.00	0.852	1.48	0.713	1.03	0.824	1.45
sd	Normal	25	+6	0.833	0.78	0.902	1.00	0.984	4.25	0.427	3.21	0.751	4.07
			−6	0.803	0.78	0.872	1.00	0.986	4.36	0.439	3.29	0.774	4.16
sd	Lognormal	25	+6	0.477	0.84	0.538	1.00	0.996	1.50	0.923	1.09	0.968	1.47
			−6	0.511	0.84	0.555	1.00	0.999	1.54	0.940	1.11	0.982	1.42
quantile	t_3	40	+6	0.855	0.91	0.882	1.00	0.992	2.53	0.889	2.21	0.927	2.47
			−6	0.845	0.92	0.864	1.00	0.863	1.00	0.832	0.89	0.843	0.92
quantile	Lognormal	40	+6	0.847	0.92	0.873	1.00	0.990	1.28	0.906	1.10	0.942	1.24
			−6	0.850	0.92	0.870	1.00	0.870	1.00	0.839	0.90	0.849	0.93
skew	Exponential	25	+6	0.497	0.81	0.595	1.00	1.000	1.80	0.943	1.27	1.000	1.69
			−6	0.466	0.81	0.570	1.00	0.293	2.67	0.012	1.45	0.041	2.09
es	Lognormal	100	+6	0.718	0.81	0.768	1.00	0.960	1.06	0.879	0.80	0.988	1.06
			−6	0.677	0.81	0.730	1.00	0.712	0.98	0.655	0.73	0.735	0.97

The Directional Artifact, Diagnosed

One pattern in Table 7 needs more than a width column to dispose of. For the expected shortfall, a base measure displaced upward raises the coverage of the IBDB from 0.825 to 0.985 under a lognormal population and from 0.802 to 0.952 under a Pareto. Taken at face value this would be a robustness claim: a badly specified prior improving inference. It is not, and Table 8 shows why by three diagnostics.

The first is a displacement sweep. Were the gain robustness, coverage would be flat in a neighborhood of the correct base measure; were it a directional accident, coverage would trend with the displacement. It trends: under the lognormal, IBDB coverage runs 0.721, 0.757, 0.772, 0.825, 0.829, 0.814, 0.985 as the displacement moves from −6 to +6 interquartile ranges, rising steadily apart from a dip at +3 that is within the resolution of the design, and rising sharply at +6. There is no plateau around the correct base measure.

Table 8. The directional artifact, diagnosed. For each displacement of the base measure (in population interquartile ranges) the table reports coverage, mean width, and the fraction of intervals lying entirely below the truth. The upper-tail and lower-tail expected shortfall are mirror functionals: were the gain under an upward displacement robustness, it would appear in both; it appears in one and reverses in the other. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange). 1000 replications per cell, $q = 100$, seed 20260701.

Tail	Population	Shift	Efron 2			PBB			IBDB		
			cov	w	Below	cov	w	Below	cov	w	Below
Upper ES _{0.95}	Lognormal	−6	0.715	7.12	0.284	0.692	6.96	0.308	0.721	6.85	0.276
		−3	0.743	7.29	0.257	0.727	7.14	0.273	0.757	7.06	0.238
		−1	0.743	7.25	0.257	0.722	7.04	0.278	0.772	7.12	0.227
		correct	0.745	7.40	0.255	0.757	7.29	0.243	0.825	7.58	0.174
		+1	0.756	7.41	0.243	0.760	7.25	0.240	0.829	7.42	0.168
		+3	0.735	7.59	0.265	0.755	7.30	0.245	0.814	7.44	0.184
		+6	0.753	7.45	0.247	0.969	7.90	0.031	0.985	7.89	0.005

Table 8. Cont.

Tail	Population	Shift	Efron 2			PBB			IBDB		
			cov	w	Below	cov	w	Below	cov	w	Below
Pareto(2.5)		−6	0.710	4.81	0.290	0.674	4.68	0.326	0.717	4.79	0.283
		−3	0.663	4.59	0.336	0.644	4.45	0.356	0.678	4.55	0.317
		−1	0.682	5.35	0.316	0.671	5.12	0.327	0.744	5.40	0.249
		correct	0.701	5.08	0.298	0.727	4.94	0.273	0.802	5.32	0.193
		+1	0.696	4.99	0.304	0.729	4.83	0.271	0.823	5.21	0.177
		+3	0.684	4.99	0.316	0.706	4.71	0.294	0.799	5.08	0.199
		+6	0.704	5.22	0.296	0.857	5.20	0.143	0.952	5.35	0.039
Lower ES (mirror)	Lognormal	−6	0.858	0.13	0.001	0.991	4.82	0.009	0.599	5.43	0.401
		−3	0.834	0.13	0.002	0.968	3.12	0.032	0.394	3.34	0.606
		−1	0.829	0.13	0.000	0.890	1.10	0.110	0.644	1.20	0.356
		correct	0.844	0.13	0.000	0.857	0.14	0.000	0.913	0.13	0.001
		+1	0.829	0.13	0.002	0.801	0.15	0.000	0.824	0.13	0.002
		+3	0.862	0.13	0.004	0.841	0.14	0.000	0.856	0.12	0.005
		+6	0.844	0.13	0.000	0.813	0.14	0.000	0.842	0.12	0.001
Pareto(2.5)		−6	0.872	0.03	0.002	0.888	3.26	0.112	0.001	3.29	0.999
		−3	0.875	0.03	0.002	0.168	1.68	0.832	0.000	1.65	1.000
		−1	0.863	0.03	0.004	0.089	0.53	0.911	0.013	0.54	0.987
		correct	0.870	0.03	0.002	0.872	0.03	0.000	0.913	0.03	0.003
		+1	0.853	0.03	0.000	0.826	0.03	0.000	0.846	0.03	0.000
		+3	0.887	0.03	0.004	0.853	0.03	0.000	0.884	0.03	0.003
		+6	0.869	0.03	0.005	0.854	0.03	0.000	0.871	0.03	0.006

The second is a surgery on the position of the interval rather than its width. The upper-tail expected shortfall is estimated from a handful of tail observations and is biased downward in finite samples; a prior displaced upward translates the interval toward the truth rather than widening it, which the *below* columns record directly.

The third is decisive, and it is a mirror. If the mechanism is orientation, the same upward displacement applied to a lower-tail functional must stop helping, because the lower-tail expected shortfall $\mathbb{E}[X \mid X \leq q_{1-p}]$ is biased in the opposite direction. The lower block of Table 8 confirms it. Under a Pareto population, a base measure displaced downward gives IBDB coverage of 0.001, 0.000 and 0.013 at −6, −3 and −1 interquartile ranges, with essentially every interval lying entirely below the truth, while an upward displacement leaves coverage between 0.846 and 0.884. The effect has reversed sign, as predicted.

Coverage under a misspecified base measure therefore improves only when the direction of the error happens to oppose the finite-sample bias of the statistic, and collapses when it aligns with it. An analyst cannot rely on this, because the direction of the error is unknown by construction: if it were known, it would have been corrected.

4.8. Sensitivity to the Resolution and to the Concentration Coupling

Two specification choices of Section 3.1 were held fixed throughout the study above: the resolution $q = 100$ of the base-measure representation, and the coupling $\alpha = \beta = \hat{\alpha}$ of the two concentrations. Table 9 varies both on four representative cells—the standard deviation of a lognormal at $n = 25$, the 0.95-quantile of a t_3 at $n = 40$, and the expected shortfall of a lognormal and of a Pareto(2.5) at $n = 100$ —with 1000 replications per cell and the seed discipline of Section 4.1.

The resolution axis moves coverage in one direction: raising q from 50 to 200 raises coverage in every cell, most visibly under the heaviest tail (0.772 to 0.830 for the Pareto expected shortfall), at essentially constant width and with below-truth failures falling (0.224 to 0.166 in the same cell)—the mechanism made exact in Proposition 5(ii), since a finer G places more atoms above the sample maximum. The default $q = 100$ is therefore reasonable; a larger q buys a few points of tail coverage at cost linear in q , and resolutions below 100 give away tail coverage measurably.

Table 9. Sensitivity of the IBDB to the resolution q (at $\alpha = \beta = \hat{\alpha}$) and to the concentration coupling (at $q = 100$ and fixed geometric mean $\sqrt{\alpha\beta} = \hat{\alpha}$). Each entry is the coverage of the nominal 90% interval, with mean width in parentheses. 1000 replications per cell.

Statistic	Population	n	Resolution q			Coupling β/α		
			50	100	200	1/4	1	4
sd	Lognormal	25	0.790 (2.37)	0.857 (2.59)	0.914 (2.60)	0.770 (2.33)	0.886 (2.63)	0.929 (2.97)
$Q_{0.95}$	t_3	40	0.928 (3.54)	0.932 (3.33)	0.944 (3.33)	0.864 (2.99)	0.949 (3.48)	0.979 (4.26)
$ES_{0.95}$	Lognormal	100	0.809 (7.57)	0.835 (7.32)	0.842 (7.58)	0.811 (7.40)	0.828 (7.71)	0.896 (8.66)
$ES_{0.95}$	Pareto(2.5)	100	0.772 (5.18)	0.780 (5.13)	0.830 (5.41)	0.747 (4.81)	0.813 (5.45)	0.862 (6.02)

The coupling axis, explored at fixed geometric mean $\sqrt{\alpha\beta} = \hat{\alpha}$, re-traverses a coverage–width trade-off rather than revealing free coverage: $\beta/\alpha = 4$ raises coverage in all four cells but widens the intervals by 10–23 percent—enlargement, not inference—while $\beta/\alpha = 1/4$ narrows and loses coverage. No decoupled setting dominates the coupled choice, which supports $\alpha = \beta = \hat{\alpha}$ as a defensible default and leaves a two-dimensional empirical Bayes rule as an open refinement.

4.9. A Real-Data Illustration: Danish Fire Losses

The same phenomena appear outside simulation. We use the Danish fire insurance losses [21], 2167 losses exceeding one million Danish kroner between 1980 and 1990, a canonical heavy-tailed benchmark in risk analysis. The design mimics the practitioner’s situation with genuinely external prior information. The chronologically first half of the record (1083 losses) plays the role of the external source, from which $q = 100$ prior atoms are drawn in each replication; the second half is the target population, and the ground truth is its empirical 0.95 quantile, 10.26 million kroner. Samples of size n are drawn without replacement from the target half and the five methods are applied with the settings of Section 4.1.

Table 10 reproduces the structure of the synthetic study on real data. At $n = 25$ the support ceiling is 0.723 and both frequentist methods sit against it, at 0.731 and 0.743, with about a quarter of all intervals lying entirely below the truth; the calibrated double bootstrap cannot improve on the single bootstrap because the limitation is the support, not the level. The informative methods cross the ceiling, and the IBDB crosses it furthest, at 0.959, cutting the entirely below failures from 25.8% to 2.7%. As the sample grows, the ceiling rises and the advantage decays exactly as it should: at $n = 80$, where the ceiling is 0.983 and no longer binds, the IBDB and the frequentist double bootstrap converge to 0.936 and 0.930. The prior has been diluted by the arrival of data, which is the behavior the method should exhibit and the reason its value is largest at a small n .

The difference is not only numerical but practical. At $n = 25$, one in four of the frequentist intervals lies *entirely below* the true 10.26 million kroner quantile—a reserve set anywhere inside them understates the exposure with certainty—against fewer than three percent for the IBDB, at the visible price of an interval roughly sixty percent wider; by $n = 80$ the two methods support essentially the same reserve at comparable width. Read as risk assessment, the informative construction changes the conclusion exactly where a practitioner with a short record and a relevant external source needs it to, and stops mattering when the data suffice.

Two limitations of this exercise should be stated plainly. The ground truth is itself an empirical quantile of a finite record, so it carries its own sampling error. And the chronological split makes the external source well specified by construction, which is the favorable case for the prior; the misspecified cases are the province of Section 4.7.

Table 10. Danish fire insurance losses: coverage of the 0.95 quantile of the target half (10.26 million DKK) by nominal 90% intervals, with mean width and the fraction of intervals lying entirely below the truth. Prior atoms are drawn from the chronologically earlier half of the record; *ceiling* is the distribution-free bound $1 - 0.95^n$. Cells are shaded by proximity of the coverage to nominal within each sample size. 1000 replications, $q = 100$. Cells are shaded by proximity of the coverage to the nominal 0.90 within each row, from best (dark green, bold) through light green and yellow to worst (orange).

n	Ceiling	Method	Coverage	Width	Entirely Below
25	0.723	Efron 1	0.731	19.61	0.258
		Efron 2	0.743	20.42	0.256
		PBB	0.908	22.46	0.091
		Post. boot.	0.778	19.45	0.190
		IBDB	0.959	32.24	0.027
40	0.871	Efron 1	0.863	23.77	0.127
		Efron 2	0.884	25.14	0.115
		PBB	0.915	25.30	0.083
		Post. boot.	0.856	23.72	0.114
		IBDB	0.935	26.15	0.050
80	0.983	Efron 1	0.865	12.54	0.126
		Efron 2	0.930	16.56	0.069
		PBB	0.963	21.79	0.036
		Post. boot.	0.903	14.58	0.075
		IBDB	0.936	18.51	0.047

4.10. A Second Real-Data Illustration: Siemens Daily Losses

The Danish study places the prior in a favorable position by construction: the two halves of the record are draws from essentially the same loss regime. A second illustration removes that favor. We use the daily log-returns of Siemens stock [22], 6146 trading days from 1973 to 1996 and another canonical benchmark of heavy-tailed risk analysis, with losses defined as negative returns. The design is the chronological mirror of Section 4.9: the first half of the record (3073 days, roughly 1973–1985) supplies the $q = 100$ prior atoms in each replication, the second half is the target population, and the ground truth is its empirical expected shortfall at level 0.95, a daily loss of 3.15%. The target functional is the expected shortfall because it is the functional on which Section 4.6 rests its strongest claim, and the sample sizes $n \in \{50, 100, 200\}$ are those at which the quantile support ceiling no longer binds.

The reason this illustration earns its place is that the chronological split is adversarial here rather than favorable. The first half of the record is a calmer regime: its expected shortfall is 2.14% and its largest daily loss is 5.0%, whereas the target half contains the October 1987 crash, with a largest daily loss of 12.0% and an expected shortfall about 1.5 times that of the prior source. The prior therefore understates the tail, in exactly the direction that Section The Directional Artifact, Diagnosed identifies as the harmful one, and no method can be expected to reach nominal coverage: the truth is driven by a handful of crash days that most samples of size $n \leq 200$ never contain.

Table 11 reports what happens. All methods undercover, as they must, and coverage rises with n for all of them. Within that honest picture the ordering of the paper reappears. At $n = 50$ the IBDB covers 0.557 against the double bootstrap's 0.506, a real difference at this replication count, and does so at smaller width. At $n = 100$ and $n = 200$ its coverage matches the double bootstrap's within Monte Carlo error—0.793 against 0.792 at $n = 200$ —but at a width roughly 5% smaller, which is the repositioning signature of Section 4.6 reproduced on market data under a prior that understates the tail. The one-level competitors do not reproduce it: the proper Bayesian bootstrap covers less at comparable width, and the posterior bootstrap is the narrowest and the lowest. The two real-data

studies therefore bracket the method: the Danish study shows what the prior buys when the external source is well specified, and the Siemens study shows that the two-level architecture keeps its edge and degrades no faster than its competitors when the external source understates the tail. In practical terms the exercise delivers two conclusions an analyst can act on: at equal reliability the IBDB supports a sharper capital figure—at $n = 200$ it matches the double bootstrap’s coverage with an interval about five percent narrower—and no resampling method, informative or not, rescues a tail estimate when the relevant extremes are absent from both the sample and the prior source; the crash-driven undercoverage shared by all five methods is a limit of the information, and reporting it plainly is part of honest risk assessment.

Table 11. Siemens daily losses: coverage of the expected shortfall at level 0.95 of the target half (a daily loss of 3.15%) by nominal 90% intervals, with mean width (in percentage points of daily loss) and the fraction of intervals lying entirely below the truth. Prior atoms are drawn from the chronologically earlier half of the record, a calmer regime whose expected shortfall is about 1/1.5 of the target’s, so the prior understates the tail. Cells are shaded by proximity of the coverage to nominal within each sample size. 1000 replications, $q = 100$, seed 20260701.

n	Method	cov	w (%)	Below
50	Efron 1	0.503	2.47	0.495
	Efron 2	0.506	2.71	0.493
	PBB	0.510	2.79	0.490
	Post. boot.	0.487	2.16	0.508
	IBDB	0.557	2.61	0.441
100	Efron 1	0.616	2.11	0.381
	Efron 2	0.676	2.63	0.324
	PBB	0.645	2.51	0.355
	Post. boot.	0.589	1.82	0.405
	IBDB	0.678	2.47	0.321
200	Efron 1	0.745	1.70	0.251
	Efron 2	0.792	2.19	0.208
	PBB	0.775	2.15	0.225
	Post. boot.	0.712	1.55	0.275
	IBDB	0.793	2.09	0.207

4.11. What the Simulations Establish

Four conclusions enter the record. First, coverage is ordered by the demand the statistic places on the tail: the IBDB over-covers for location, where the sample already suffices, and dominates for tail and risk functionals, and for the scale of heavy-tailed populations. Second, the frequentist double bootstrap saturates the support ceiling rather than merely respecting it. Third, a properly implemented one-level informative method is a strong competitor—the corrected PBB reaches high coverage in moderate cells—but it does so by inflation, over-covering at 1.2 times the Efron width for the value-at-risk, and it cannot match the IBDB on the expected shortfall, where the IBDB covers more in every cell at comparable width. The advantage of the second level is therefore calibration, not mere support extension or interval widening. Fourth, the method has a diagnosed failure mode under a base measure misplaced in the direction of the statistic’s finite-sample bias, most severe for the skewness (see Tables 2–7).

5. Discussion and Related Work

5.1. Related Work

The bootstrap was introduced by Efron [2], and double or iterated bootstrap procedures were developed by Efron [10], Hall [23], Beran [16], whose pre pivoting argument identifies coverage calibration as the purpose of the second level. Hall and

Martin [17] unified iterated resampling under a single principle and quantified its effect: each iteration reduces two-sided coverage error by a factor n^{-1} (and one-sided error by $n^{-1/2}$), sharpening the n^{-2} two-sided rate of Beran [16]. The calibration reading of the inner level adopted throughout this paper is theirs. The Bayesian bootstrap is due to Rubin [3], and Lo [4] established its large-sample theory as a Dirichlet process posterior. Hjort [20] developed the version with a non-degenerate Dirichlet prior, proved the first-order equivalence of five procedures—exact Bayesian, Bayesian bootstrap, Rubin’s degenerate-prior bootstrap, Efron’s bootstrap, and the delta method—as the sample grows with the prior held fixed, and gave the first empirical Bayes rules for the concentration. Muliere and Secchi [5] developed the informative one-level version against which Section 4 is benchmarked, and the weighted likelihood bootstrap of Newton and Raftery [24]—which already contemplated incorporating prior information by using the WLB as the proposal in sampling importance resampling—and later weighted Bayesian bootstrap developments provide related one-level Bayesian resampling ideas.

A disambiguation is needed, because superficially similar names denote different objects. The fast double bootstrap of Davidson and MacKinnon [25] accelerates the calibrated double bootstrap by thinning the inner level; the subsampled double bootstrap of Sengupta et al. [26], built on that idea, is a frequentist device for massive data whose two levels exist to reduce computational cost through subsampling, not to calibrate coverage or to carry prior information. Barrientos and Peña [27] constructed Bayesian bootstrap analogues of that scheme and of the bag of little bootstraps, again with scalability as the goal, so that the phrase “double Bayesian bootstrap” in that line of work refers to a computational architecture for large samples; and the Bayesian fast double bootstrap of Matdoan et al. [28] places Bayesian estimators inside a fast-double-bootstrap loop to set control-chart limits, another different object. The present paper sits on an orthogonal axis: both of its levels operate on the full sample, the outer level exists to carry prior information through G and a regulated concentration, and the inner level exists to calibrate coverage in the sense of Beran [16]. Neither the subsampled constructions nor their Bayesian analogues incorporate an informative base measure or an empirical Bayes rule for the concentration, and the IBDB gains nothing computationally: it is an information device, not a scalability device.

The Dirichlet process was introduced by Ferguson [6]. The hierarchical Dirichlet process of Teh et al. [1] provides the two-level Bayesian nonparametric construction that motivates the present one, which is of the same type—a Dirichlet process based on the realization of another—and is separated from it by the conditioning of Section 2.3; its distribution theory is developed by Camerlenghi et al. [29]. The large-concentration regime in which this paper works is the one studied by Feng and Paguyo [11], whose central limit theorems concern symmetric functionals of the weights; the companion large deviation principles of Feng [30], whose rate function is one relative entropy per level, are the closest limit-theoretic relatives of the Kullback–Leibler bounds of Section 2.2. The propositions of this paper are established independently rather than derived from those results; what the paper adds is the bootstrap reading of the regime and the methodology that follows from it. Within the same asymptotic tradition, Feng [31] obtains unconditional large deviation principles for the Dirichlet posterior and the Bayesian bootstrap—the point at which that literature does reach the bootstrap, at a single level. No result in this line concerns the *double* bootstrap or identifies it as the endpoint of the two-level scheme, which is the identification made here.

The closest active family is the one built on general Bayesian updating. Lyddon et al. [32] revisit the weighted likelihood bootstrap and show that it samples exactly from a Bayesian nonparametric posterior for a parameter defined as the minimizer of an expected loss, giving the loss-likelihood bootstrap. Fong et al. [13] turn the same idea

into the *posterior bootstrap*, in which a Dirichlet process prior is placed on the unknown data distribution, centered on a parametric model, and posterior samples are drawn by randomizing an objective function over the data together with pseudo-samples generated from the centering measure. The last point deserves emphasis: because the centering measure contributes atoms of its own, the posterior bootstrap is not confined to the convex hull of the sample, and it can therefore cross the support ceiling of Section 4. So can the proper Bayesian bootstrap of Muliere and Secchi [5], whose original motivation was precisely that the predictive distribution should not be forced to concentrate on the observed values. Support extension is not what distinguishes the present paper.

What distinguishes it is that all of these are one-level procedures, and Section 4 shows what one level costs. A single Dirichlet level has a single dial, and that dial controls both how much the prior contributes and how wide the resulting interval is; the two cannot be set separately. The consequence is visible in Table 6: for the value-at-risk the proper Bayesian bootstrap reaches coverage of 0.96–0.98, but at roughly 1.2 times the width of the frequentist double bootstrap, and for the expected shortfall it cannot match the two-level construction at any width. The posterior bootstrap inherits the same limitation: it fixes the concentration by user choice and has no second level with which to correct the coverage that the first level delivers. The IBDB separates the two functions—the outer Dirichlet process carries the information and extends the support, the inner one calibrates—and regulates the outer concentration by an empirical Bayes rule rather than by fiat. That separation, not the informative base measure, is the contribution.

Further methods in this family include the weighted Bayesian bootstrap of Newton et al. [33], the Bayesian bootstrap spike-and-slab lasso of Nie and Ročková [34], the generalized bootstrap for estimating equations of Chatterjee and Bose [35], and other nonparametric resampling procedures; their common objective is scalable posterior sampling under model misspecification, not coverage calibration for a difficult functional.

The family has since been generalized. Fong et al. [36] propose the *martingale posterior*, in which Bayesian uncertainty is located in the missing observations rather than in a prior over parameters: the analyst elicits a sequence of one-step-ahead predictive distributions, posterior sampling becomes predictive imputation, and Doob's martingale convergence theorem recovers the conventional posterior when the predictive is the Bayesian one. The framework subsumes the posterior bootstrap as a special case and applies to any statistic. A specialization to quantiles is developed by Fong and Yiu [37], with explicit connections to the Bayesian bootstrap, and a semiparametric moment-based branch appears in Yung et al. [38].

This is the closest active line of work to the present paper, and one point of contact is worth naming. Locating uncertainty in the observations one did not see is the same intuition that Section 4 makes quantitative, since the support ceiling $1 - 0.95^n$ is exactly the probability that the sample fails to bracket the quantity of interest. The objectives nevertheless differ. The martingale posterior is a foundational reconstruction of Bayesian inference that dispenses with the likelihood–prior pair and buys tractable, parallelizable computation. It does not regulate the influence of an informative base measure by an empirical Bayes rule, and it does not carry a second resampling level whose purpose is to calibrate the coverage of the first. A systematic comparison of the IBDB against the quantile martingale posterior on the benchmark of Section 4.7 is a natural next step, and we do not attempt it here.

A different route to the same objective is generalized posterior calibration [39], in which the coverage of a generalized posterior—an object that does carry prior information—is corrected by a bootstrap loop that solves a coverage equation for a scalar learning rate; learning-rate selection rules are compared by Wu and Martin [40]. That construction

shares the two-layer logic of the present paper, with the informative object outside and the calibration device inside. It differs in the calibration instrument: a single scalar tempering exponent rescales the credible set globally, whereas the inner Dirichlet level here acts on the resampling distribution itself, which is what permits the repositioning rather than widening documented in Section 4.6. In the same spirit at a single level, Gibson [41] tunes the concentration of a non-informative Bayesian bootstrap from the data to meet a conformal coverage target; that shares the idea of regulating a concentration for coverage, but with no informative base measure and no second resampling level.

The literature above falls into two mature branches, and the gap between them locates this paper. One branch makes a resampling scheme *informative*—a line opened by Boos and Monahan [42] and developed by Hjort [20], Muliere and Secchi [5], and Fong et al. [13]—carrying prior information and often extending the support, typically at a single Dirichlet level and, within this resampling line, without a mechanism whose purpose is to correct the coverage that the informative level delivers. Two-level Bayesian bootstrap constructions do exist: the hierarchical Bayesian bootstrap of Oganisian et al. [43] places a hierarchical Dirichlet process over stratum-specific distributions so that sparse strata borrow support from the pooled sample. But there the second level exists to share information across groups, not to calibrate coverage, so the distinction that locates the present paper is not the number of Dirichlet levels but the function of the inner one. The other makes it *calibrated*—Efron [10], Beran [16], Beran [44], Hall and Martin [17], Martin [45]—adding an inner stage whose base measure is the empirical distribution and nothing more. Analytic corrections such as the bias-corrected and accelerated intervals of Efron [46] sharpen the endpoints of a given resampling distribution without a second level, but they give prior information no place to enter, which is the role the outer level plays here. Each branch lacks what the other has, and Proposition 1 shows why the two were never combined: they are separated by a limit, the informative construction being a retreat from the infinite-concentration endpoint at which the calibrated double bootstrap sits. Closest in mechanics is the Generative Multi-purpose Sampler of Shin et al. [12], whose bootstrap specialization applies a two-level Dirichlet smoothing to the iterated bootstrap for computational amortization, with non-informative weights at both levels. Here, by contrast, the two Dirichlet levels are the statistical object itself, with an informative outer base measure and a regulated concentration. The distinctive feature of the present paper is the combination of informative prior incorporation with a second calibration level in the sense of the double bootstrap.

Table 12 summarizes the landscape. Each neighbor holds some of the components; the conjunction of the three—an informative outer base measure, a concentration regulated from the data, and a second resampling level whose purpose is coverage calibration—is the cell this paper occupies.

5.2. Interpretation

This paper places the double bootstrap and the hierarchical Dirichlet process within a common probabilistic framework and uses that connection to construct the Informative Bayesian Double Bootstrap. Five results were established. First, the classical double bootstrap is the non-informative endpoint of a two-level Dirichlet construction—of the construction conditional on its realized upper level, and not of the hierarchical Dirichlet process, whose return path keeps the tie probability bounded away from zero. Proposition 1 bounds the total variation between the two-level predictive scheme and the double bootstrap by $\frac{n(n-1)}{2}(\alpha^{-1} + \beta^{-1})$, and Propositions 2 and 3 sharpen this to a Kullback–Leibler statement with an exact second-order constant, verified by exact enumeration.

Table 12. The location of the IBDB among its nearest neighbors. Columns: an informative outer base measure; a concentration parameter regulated from the data; a second resampling level; and the purpose that second level serves. For generalized posterior calibration the second layer is a bootstrap loop that tunes a scalar learning rate rather than a resampling level with its own base measure (see text); for the classical double bootstrap there is no concentration parameter to regulate. A check mark indicates that the feature is present in the corresponding method; a cross that it is absent.

Method	Informative Base	Regulated Conc.	Second Level	Purpose of Second Level
Double bootstrap [10]	×	—	✓	coverage calibration
Hjort [20]	✓	✓	×	—
Proper Bayesian bootstrap [5]	✓	×	×	—
Posterior bootstrap [13]	✓	×	×	—
GMS [12]	×	×	✓	cost amortization
Hierarchical Bayesian bootstrap [43]	×	×	✓	partial pooling
Generalized posterior calibration [39]	✓	×	✓	coverage calibration
IBDB (this paper)	✓	✓	✓	coverage calibration

Second, moving away from that endpoint at the outer level while holding the inner calibration loop at its non-informative endpoint defines the IBDB, a practical resampling procedure rather than merely an asymptotic construction.

Third, the analytical picture extends to the method itself: Proposition 4 bounds its distance from the independence skeleton with the classical constants, and Propositions 5 and Corollary 1 pair the distribution-free ceilings of the support-confined methods with an exact escape formula showing that the IBDB is not subject to them—a separation confirmed against the simulations to three decimal places.

Fourth, the role of the second level was isolated: its contribution is calibration rather than support extension. A correctly implemented one-level informative bootstrap also crosses the support ceiling and also reaches high coverage, but it does so by widening the interval. On the expected shortfall, by contrast, the two-level construction covers more in every cell at comparable width, with the below-truth failures reduced rather than the interval enlarged.

Fifth, the domain of applicability was characterized by a uniform factorial design across six statistics: the method over-covers for location, where the sample already suffices, and dominates for tail and risk functionals, and for the scale of heavy-tailed populations. Section 4.4 shows the classical double bootstrap *attaining* the distribution-free support ceiling $1 - p^n$ rather than merely respecting it, while the informative construction crosses it, and Section The Directional Artifact, Diagnosed diagnoses, by a displacement sweep and a mirrored functional, the directional artifact that makes an apparently favorable misspecification untrustworthy.

The central methodological claim is deliberately narrow. It is not that prior information improves every bootstrap problem; it is that the hierarchical construction gives prior information a place to enter without discarding the calibration role of the second level. The advantage is real where the sample cannot see the target and negligible where it can.

5.3. When to Use the IBDB: Practical Guidance

The factorial study supports a compact prescription. The IBDB is the right tool when three conditions hold at once: the target is a tail-sensitive functional—an upper quantile, the expected shortfall, the scale or skewness of a heavy-tailed population—so that the sample alone cannot see the quantity of interest; the sample is small to moderate, since the advantage decays as the support ceiling lifts (Tables 4 and 10); and an external source exists that is at least vaguely informative about the right region—historical records, a longer related series, engineering knowledge—from which the q atoms of G can be drawn. When these conditions hold, the method’s gain takes the specific form documented

above: intervals repositioned toward the truth at comparable width, with below-truth failures reduced.

The prescription has an equally compact negative side. For location statistics the method should not be preferred: the sample suffices and the informative machinery only over-covers (Table 2). The principal danger is not a weak prior but a *directionally* wrong one—a base measure displaced in the direction of the statistic’s finite-sample bias can silently destroy coverage (the skewness rows of Table 7 are the worst case), and the empirical Bayes rule, which reacts to location and dispersion, does not detect a tail-only misspecification (the Siemens study). As a routine safeguard, the IBDB interval should be compared with the frequentist double bootstrap interval: agreement indicates the prior is adding little, and a large one-sided displacement flags exactly the situation in which the prior, rather than the data, is speaking.

5.4. Limitations and Future Work

The limitations are concrete, and none of them is cosmetic. The empirical Bayes rule of Section 3.3 is a heuristic: it regulates the outer concentration sensibly but is not derived from a marginal-likelihood or risk criterion, and a principled replacement—extending the moment and likelihood rules of Hjort [20] from one Dirichlet level to the pair (α, β) —remains open. The resolution q of the base measure is at present chosen by hand rather than by a selection rule, although Section 4.8 now maps the sensitivity: coverage improves mildly and monotonically in q at essentially constant width, so the choice is a computational trade-off rather than a fragile one. The two concentrations are coupled, $\alpha = \beta$, for convenience rather than from any optimality argument; the decoupling grid of Section 4.8 shows that the coupled choice is not dominated within the explored family, and that tilting mass to the inner level buys coverage only through inflation. The method depends on G being at least mildly informative in the right region; a base measure misplaced in the direction that matters returns the procedure to the double bootstrap rather than improving on it, as the adversarial column of Table 7 shows. The advantage decreases with sample size. And there is no general coverage theorem: the endpoint propositions motivate the construction and pin down its boundary behavior, but they do not by themselves certify nominal coverage of the IBDB at any fixed finite concentration, which is established here by simulation.

Two directions follow naturally. A marginal-likelihood rule for α would replace the current heuristic and, with it, a selection rule for q ; and a decoupling of (α, β) with an accompanying criterion would let the calibration level and the informative level be tuned separately. Neither is required for the method to be used as it stands, but both would move it from a demonstrably useful instrument toward a fully automatic one.

This work does not reinterpret the double bootstrap through Bayesian nonparametrics merely for conceptual interest. The hierarchical construction identifies the classical procedure as one endpoint of a broader family of resampling schemes, and that identification is what allows prior information to enter without abandoning the calibration role of the second resampling stage. What the experiments then teach is compact: the second level buys calibration rather than support extension, and entering the interior of the family pays precisely where the sample cannot see the target—in the tail.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/math14173192/s1>, Supplementary Code and Data S1: Python scripts, data and CSV result files reproducing all simulations, tables and numerical verifications (engine.py, ibdb.py, the run_*.py drivers, verify_meanc1t.py and verify_posterior.py, auxiliary scripts, the results_*.csv files, and the danish.rda, siemens.rda and bmw.rda data files used in the real-data applications).

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the Supplementary Material. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the author used Claude (Anthropic) for assistance with LaTeX preparation, literature search, simulation code, and the drafting and numerical verification of mathematical arguments, and ChatGPT (OpenAI) for assistance with writing. The author has reviewed, verified, and edited all output and takes full responsibility for the content of this publication. The author thanks Eduardo Gutiérrez-Peña for early discussions on the bootstrap–Dirichlet connection.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Teh, Y.W.; Jordan, M.I.; Beal, M.J.; Blei, D.M. Hierarchical Dirichlet processes. *J. Am. Stat. Assoc.* **2006**, *101*, 1566–1581. [\[CrossRef\]](#)
2. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [\[CrossRef\]](#)
3. Rubin, D.B. The Bayesian bootstrap. *Ann. Stat.* **1981**, *9*, 130–134. [\[CrossRef\]](#)
4. Lo, A.Y. A large sample study of the Bayesian bootstrap. *Ann. Stat.* **1987**, *15*, 360–375. [\[CrossRef\]](#)
5. Muliere, P.; Secchi, P. Bayesian nonparametric predictive inference and bootstrap techniques. *Ann. Inst. Stat. Math.* **1996**, *48*, 663–673. [\[CrossRef\]](#)
6. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, *1*, 209–230. [\[CrossRef\]](#)
7. Wu, W.; Wang, B.; Chen, P. Modeling and estimation for degradation data with initiation–growth correlations. *IEEE Trans. Reliab.* **2026**, *75*, 1377–1391. [\[CrossRef\]](#)
8. Wang, P.; Tang, Y. Remaining useful life prediction based on exponential dispersion process with random drifts. *Stat. Theory Relat. Fields* **2025**, *9*, 404–433. [\[CrossRef\]](#)
9. Zhuang, L.; Ma, Y.; Wang, J.; Pan, R.; Xu, A. From statistical modeling to AI-integrated inverse Gaussian process: A comprehensive review for prognostics and health management. *Eng. Manag.* **2026**, *13*, 65–84. [\[CrossRef\]](#)
10. Efron, B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Stat. Assoc.* **1983**, *78*, 316–331. [\[CrossRef\]](#)
11. Feng, S.; Paguyo, J.E. Central limit theorems associated with the hierarchical Dirichlet process. *Stoch. Process. Their Appl.* **2025**, *190*, 104767. [\[CrossRef\]](#)
12. Shin, M.; Wang, S.; Liu, J.S. Generative multi-purpose sampler for weighted M-estimation. *J. Comput. Graph. Stat.* **2024**, *33*, 1084–1097. [\[CrossRef\]](#)
13. Fong, E.; Lyddon, S.; Holmes, C. Scalable nonparametric sampling from multimodal posteriors with the posterior bootstrap. In *Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019*; PMLR: Cambridge, MA, USA, 2019; Volume 97, pp. 1952–1962.
14. Blackwell, D.; MacQueen, J.B. Ferguson distributions via Pólya urn schemes. *Ann. Stat.* **1973**, *1*, 353–355. [\[CrossRef\]](#)
15. Ishwaran, H.; Zarepour, M. Exact and approximate sum representations for the Dirichlet process. *Can. J. Stat.* **2002**, *30*, 269–283. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Beran, R. Prepivoting to reduce level error of confidence sets. *Biometrika* **1987**, *74*, 457–468. [\[CrossRef\]](#)
17. Hall, P.; Martin, M.A. On bootstrap resampling and iteration. *Biometrika* **1988**, *75*, 661–671. [\[CrossRef\]](#)
18. Laird, N.M.; Louis, T.A. Empirical Bayes confidence intervals based on bootstrap samples. *J. Am. Stat. Assoc.* **1987**, *82*, 739–750. [\[CrossRef\]](#)
19. Carlin, B.P.; Gelfand, A.E. Approaches for empirical Bayes confidence intervals. *J. Am. Stat. Assoc.* **1990**, *85*, 105–114. [\[CrossRef\]](#)
20. Hjort, N.L. *Bayesian and Empirical Bayesian Bootstrapping*; Technical Report, Posted to arXiv in 2026: arXiv:2605.11677; University of Oslo and Norwegian Computing Centre: Oslo, Norway, 1991.
21. McNeil, A.J. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bull.* **1997**, *27*, 117–137. [\[CrossRef\]](#)
22. McNeil, A.J.; Frey, R. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *J. Empir. Financ.* **2000**, *7*, 271–300. [\[CrossRef\]](#)
23. Hall, P. On the bootstrap and confidence intervals. *Ann. Stat.* **1986**, *14*, 1431–1452. [\[CrossRef\]](#)
24. Newton, M.A.; Raftery, A.E. Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *J. R. Stat. Soc. Ser. B* **1994**, *56*, 3–48. [\[CrossRef\]](#)
25. Davidson, R.; MacKinnon, J.G. Improving the reliability of bootstrap tests with the fast double bootstrap. *Comput. Stat. Data Anal.* **2007**, *51*, 3259–3281. [\[CrossRef\]](#)

26. Sengupta, S.; Volgushev, S.; Shao, X. A subsampled double bootstrap for massive data. *J. Am. Stat. Assoc.* **2016**, *111*, 1222–1232. [\[CrossRef\]](#)
27. Barrientos, A.F.; Peña, V. Bayesian bootstraps for massive data. *Bayesian Anal.* **2020**, *15*, 363–388. [\[CrossRef\]](#)
28. Matdoan, M.Y.; Mashuri, M.; Ahsan, M. Control chart for geometrically distributed data based on Bayesian fast double bootstrap. *MethodsX* **2025**, *14*, 103307. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Camerlenghi, F.; Lijoi, A.; Orbanz, P.; Prünster, I. Distribution theory for hierarchical processes. *Ann. Stat.* **2019**, *47*, 67–92. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Feng, S. Hierarchical Dirichlet process and relative entropy. *Electron. Commun. Probab.* **2023**, *28*, 1–12. [\[CrossRef\]](#)
31. Feng, S. Unconditional large deviation principles for Dirichlet posterior and Bayesian bootstrap. *Ann. Appl. Probab.* **2025**, *35*, 2967–2982. [\[CrossRef\]](#)
32. Lyddon, S.P.; Holmes, C.C.; Walker, S.G. General Bayesian updating and the loss-likelihood bootstrap. *Biometrika* **2019**, *106*, 465–478. [\[CrossRef\]](#)
33. Newton, M.A.; Polson, N.G.; Xu, J. Weighted Bayesian bootstrap for scalable posterior distributions. *Can. J. Stat.* **2021**, *49*, 421–437. [\[CrossRef\]](#)
34. Nie, L.; Ročková, V. Bayesian bootstrap spike-and-slab LASSO. *J. Am. Stat. Assoc.* **2023**, *118*, 2013–2028. [\[CrossRef\]](#)
35. Chatterjee, S.; Bose, A. Generalized bootstrap for estimating equations. *Ann. Stat.* **2005**, *33*, 414–436. [\[CrossRef\]](#)
36. Fong, E.; Holmes, C.; Walker, S.G. Martingale posterior distributions (with discussion). *J. R. Stat. Soc. Ser. B* **2023**, *85*, 1357–1391. [\[CrossRef\]](#)
37. Fong, E.; Yiu, A. Bayesian quantile estimation and regression with martingale posteriors. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2025**, qkaf080. [\[CrossRef\]](#)
38. Yung, Y.Y.; Lee, S.M.S.; Fong, E. Moment martingale posteriors for semiparametric predictive Bayes. *arXiv* **2025**, arXiv:2507.18148.
39. Syring, N.; Martin, R. Calibrating general posterior credible regions. *Biometrika* **2019**, *106*, 479–486. [\[CrossRef\]](#)
40. Wu, P.-S.; Martin, R. A comparison of learning rate selection methods in generalized Bayesian inference. *Bayesian Anal.* **2023**, *18*, 105–132. [\[CrossRef\]](#)
41. Gibson, G. Bayesian conformal prediction via the Bayesian bootstrap. *arXiv* **2025**, arXiv:2508.01418.
42. Boos, D.D.; Monahan, J.F. Bootstrap methods using prior information. *Biometrika* **1986**, *73*, 77–83. [\[CrossRef\]](#)
43. Oganisian, A.; Mitra, N.; Roy, J.A. Hierarchical Bayesian bootstrap for heterogeneous treatment effect estimation. *Int. J. Biostat.* **2024**, *20*, 93–106. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Beran, R. Pivoting test statistics: A bootstrap view of asymptotic refinements. *J. Am. Stat. Assoc.* **1988**, *83*, 687–697. [\[CrossRef\]](#)
45. Martin, M.A. On bootstrap iteration for coverage correction in confidence intervals. *J. Am. Stat. Assoc.* **1990**, *85*, 1105–1118. [\[CrossRef\]](#)
46. Efron, B. Better bootstrap confidence intervals. *J. Am. Stat. Assoc.* **1987**, *82*, 171–185. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.