Uncertainty Quantification: From Weighted Bootstrap to Generators

Hedibert F. Lopes
Insper Institute of Education and Research, São Paulo

Nicholas G. Polson
Booth School of Business
University of Chicago

Vadim Sokolov*
Department of Systems Engineering
and Operations Research
George Mason University

First Draft: April 2026
This Draft: April 17, 2026

Abstract

We reinterpret the bootstrap as a *generator*: a deterministic map from random weights or predictive draws to posterior samples. Introduced by Efron in 1979, the idea of resampling from observed data to approximate sampling distributions launched nearly five decades of methodological innovation. This paper provides a review and computational comparison of that evolution through five developments: Efron's original resampling, Rubin's Bayesian reinterpretation via the Dirichlet process, the Newton–Raftery weighted likelihood bootstrap, the weighted Bayesian bootstrap, and the generative Bayesian bootstrap. The generator view organises all five: random weights or predictive draws provide the base randomness, and functionals, optimisation maps, or learned transports define the deterministic inference mechanism. It also extends naturally to modern nonlinear and deep-learning models, where the same pushforward construction delivers amortised uncertainty quantification via stochastic gradient descent (SGD). We summarise conditions under which the methods coincide or diverge, present a coverage study revealing their distinct finite-sample behaviour, and derive a closed-form asymptotic coverage formula $C(\rho) = 2\Phi(z_{\alpha/2}\sqrt{\rho}) - 1$ that explains the observed undercoverage of the generative Bayesian bootstrap and the overcoverage of the Jeffreys WBB in terms of a single posterior-variance ratio ρ. A corresponding calibrated GBB eliminates this variance deficit, matching the baseline Bayesian bootstrap (and nominal coverage in the large-n limit) at essentially zero

*Corresponding author: vsokolov@gmu.edu

computational cost. We connect the framework to modern generative computation, including transport maps, normalising flows, and score-based methods.

Keywords: bootstrap, generator, Bayesian bootstrap, Dirichlet process, weighted resampling, generative Bayesian computation.

1 Introduction

Uncertainty quantification is a central task in statistical computing, and the bootstrap is among the most widely used computational tools for it. In a 1979 paper in the *Annals of Statistics*, Bradley Efron proposed to approximate the sampling distribution of an estimator by resampling with replacement from the observed data, treating the empirical distribution as a stand-in for the unknown population [Efron, 1979]. The recipe is simple: compute the estimator on many synthetic datasets drawn from the observed one. Efron's original proposal is useful as a first-order procedure, but its discrete multinomial weights, reliance on plug-in functional evaluation, and lack of a natural Bayesian interpretation limit its scope. The contributions that followed (Rubin's Bayesian bootstrap, the weighted likelihood bootstrap, the weighted Bayesian bootstrap, and the generative Bayesian bootstrap) sharpened the idea into a broader computational framework for uncertainty quantification. We reinterpret all five through a common *generator* lens: each method is a deterministic pushforward from a tractable base distribution (Multinomial, Dirichlet, exponential, or the Pólya urn law) through a map G (a functional, an M-estimator, a predictive resampling rule, or a learned transport) to posterior samples. This reinterpretation makes uncertainty quantification portable to the nonlinear and deep-learning models that dominate modern computational statistics: once the point estimate is obtained by stochastic gradient descent (SGD), the same optimisation machinery, applied to a weighted objective, delivers posterior draws at essentially no additional cost. Modern SGD solvers (Adam, RMSProp, and their variants) make this amortised inference instantaneous for generalised linear models, regularised regressions, and moderately sized neural networks.

The development after 1979 reshaped the bootstrap into a Bayesian computational tool. Within two years, Rubin [1981] reinterpreted it as posterior inference under a Dirichlet process prior. Within fifteen years, Newton and Raftery [1994] leveraged it to approximate parametric Bayesian posteriors without Markov chain Monte Carlo. In the ensuing decades, weighted variants were developed that incorporated auxiliary information, moment constraints, and structural restrictions. And in the most recent chapter, Fong et al. [2019, 2023] and Polson and Sokolov [2025] connected the bootstrap to sequential predictive distributions, martingale posteriors, and generative computation.

This paper tells the story of that evolution. Our goals are three. First, we present each of the five major bootstrap developments (Efron, Rubin, Newton–Raftery, the weighted Bayesian bootstrap, and the generative Bayesian bootstrap) together with the generative Bayesian computation (GBC) framework that unifies the last two, all with enough mathematical detail to be genuinely useful to researchers. Second, we highlight the connections among these developments, showing how the generator framework (2) organises them, with method-specific inference mechanisms. Third, we place the generative Bayesian bootstrap within the broader landscape of modern probabilistic machine learning, where its connections to sequential Monte Carlo, score-based generative models, and variational inference are beginning to be understood.

Throughout, our unifying theme is the *generator representation*. At the classical level, every bootstrap variant computes

$$\hat{\theta}^* = G\left(\sum_{i=1}^{n} w_i \, \delta_{x_i}\right), \tag{1}$$

where n is the sample size, δ_{x_i} is the Dirac measure placing unit mass at observation x_i, $(w_1, \ldots, w_n)$ is a random weight vector drawn from a base distribution on the simplex, and G is the functional of interest. The modern perspective recognises this as a special case of the generator equation

$$\hat{\theta}^{*(k)} = G(\mathbf{w}^{(k)}, \lambda), \qquad \mathbf{w}^{(k)} \sim p_{\text{base}}, \quad k = 1, \ldots, B, \tag{2}$$

where G is a deterministic map from random weights to posterior samples and λ is a hyperparameter controlling regularisation or concentration. The base distribution p_{base} (Dirichlet for the Bayesian bootstrap, exponential for the weighted Bayesian bootstrap, Gaussian for diffusion models) is the source of randomness, while the generator G embodies the inference mechanism: the evaluation of a functional (classical bootstrap), the mode of a randomly weighted objective (WBB), or a learned transport such as a quantile neural network [Polson and Sokolov, 2025]. Equation (1) identifies G with a functional of a distribution, while (2) treats G as a map from weight vectors to parameters; the two views are linked by the observation that specifying weights $\mathbf{w}$ determines a discrete distribution $\sum_i w_i \, \delta_{x_i}$, so the functional and the map coincide. Every method in this review shares the structure of (2); they differ in p_{base}, G, and the inference mechanism that defines the map.

The contributions of this paper are as follows. (i) We provide a unified mathematical treatment of five major bootstrap developments spanning 1979–2025, showing that they can all be expressed through the generator representation (2). (ii) We summarise the conditions under which the Bayesian bootstrap, weighted likelihood bootstrap, and generative Bayesian bootstrap coincide asymptotically, and highlight the regimes where they diverge. (iii) We derive a closed-form *asymptotic* coverage formula for weighted-bootstrap generators (Proposition 1), instantiate it for the GBB and the Dirichlet-α WBB (Corollaries 1–2), and propose a calibrated GBB that restores nominal coverage (Proposition 2). The proposition is asymptotic but predicts the empirical coverage patterns observed at sample sizes as small as $n = 50$. (iv) We connect the generative Bayesian bootstrap to martingale posteriors, sequential Monte Carlo, and modern deep generative models, showing that the generator G takes the form of a mode-finding map (WBB) or a learned neural-network transport (GBC), and charting directions for future research.

The paper is organised as follows. Section 2 previews the generator perspective. Section 3 presents Efron's original bootstrap. Section 4 develops Rubin's Bayesian bootstrap, empirical likelihood, and the weighted likelihood bootstrap. Section 5 develops weighted bootstrap variants and the weighted Bayesian bootstrap as a generator, including a worked tutorial and connections to modern generative models. Section 6 presents the generative Bayesian bootstrap. Section 7 provides a numerical comparison. Section 8 discusses extensions, open challenges, and concluding remarks.

2 The Generator Perspective

We call a triple $(p_{\text{base}}, G, \lambda)$ a *generator* for a target posterior or sampling distribution π if draws $\mathbf{w}^{(k)} \sim p_{\text{base}}$ composed with the deterministic map G produce samples $\theta^{(k)} = G(\mathbf{w}^{(k)}, \lambda)$ that approximate π, either exactly, asymptotically, or in a specified finite-sample sense. This perspective builds on Newton et al. [2021], Liu [2001], Newton and Raftery [1994] and Polson et al. [2015]. Every bootstrap variant reviewed in this paper admits a generator representation, but crucially the class of maps G covered is broader than any single one of the classical instances. We distinguish five inference mechanisms:

1. **Functional evaluation** (Efron, Rubin BB): $G(\mathbf{w}) = T(\sum_i w_i \, \delta_{x_i})$ applies a plug-in functional T to a randomly weighted empirical measure. The base p_{base} is Multinomial (Efron) or Dirichlet (Rubin).

2. **Importance resampling** (WLB, SIR variant): G computes the MLE on a resampled dataset and applies a sampling-importance-resampling step (equations 9–10).

3. **Randomly weighted M-estimation** (WBB, equation (13)): $G(\mathbf{w}) = \arg\min_\theta \{\sum_i w_i L_i(\theta) + \lambda w_p \phi(\theta)\}$ returns the mode of a randomly perturbed regularised objective. The base is $\text{Exp}(1)^{n+1}$ (or a Dirichlet variant).

4. **Predictive resampling** (GBB): G augments the observed data with m draws from the Pólya urn (7), then applies a functional. The base is the joint law of the urn trajectory.

5. **Learned transport** (parametric GBC): $G = g_\phi$ is a neural-network quantile map or normalising flow trained to push a Gaussian base onto the target [Polson and Sokolov, 2025].

Items 1, 3, and 4 correspond to special cases routinely studied in the literature (functionals of empirical measures, bootstrap of M-estimators, predictive resampling). The generator framework treats them as specific realisations of a common pushforward and admits item 5 as a natural extension once G is unshackled from any closed-form prescription. The base distribution p_{base} plays the role of the source of randomness, while the map G defines the inference mechanism. The framework thus organises classical and modern bootstrap variants by the smoothness, expressivity, and computational cost of G, rather than by their historical origin. The sections that follow develop each instance in detail.

3 The Multinomial Bootstrap

We begin with the original bootstrap of Efron [1979] [see also Efron and Tibshirani, 1993], presenting the sampling distribution problem it was designed to solve, the algorithm itself, its weight representation, and its theoretical properties.

3.1 The sampling distribution problem

Consider independent observations $x_1, \ldots, x_n$ drawn from an unknown distribution F. We wish to estimate a parameter $\theta = G(F)$ (a functional of F) using the plug-in estimator $\hat{\theta} = G(\hat{F}_n)$, where $\hat{F}_n = n^{-1} \sum_{i=1}^{n} \delta_{x_i}$ is the empirical distribution function. The inferential challenge is to characterise the sampling distribution of $\hat{\theta}$, from which confidence intervals, standard errors, and hypothesis tests can be derived.

For the sample mean, the central limit theorem provides an asymptotic answer. For the sample variance, a slightly more involved calculation is required. For more complex functionals (the trimmed mean, the ratio of two means, the largest eigenvalue of a sample covariance matrix, the parameters of a fitted parametric model, or quantities arising in survival analysis, time series, or spatial statistics), analytic results are often unavailable, prohibitively complex, or rest on approximations whose accuracy is unknown in finite samples.

Before Efron, statisticians had three main alternatives: analytic asymptotic theory (delta method, Edgeworth expansions), the jackknife, and simulation-based methods restricted to known parametric models. The delta method required differentiability and produced only first-order approximations. The jackknife was inconsistent for some statistics (notably the sample median) and provided only variance estimates, not full distributional approximations. Specifically, the jackknife standard error of $\hat{\theta} = G(x_1, \ldots, x_n)$ is

$$\widehat{\text{se}}_{\text{jack}} = \left[\frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{(-i)} - \bar{\hat{\theta}}_{(\cdot)} \right)^2 \right]^{1/2},$$

where $\hat{\theta}_{(-i)}$ is the statistic computed with observation i deleted and $\bar{\hat{\theta}}_{(\cdot)} = n^{-1} \sum_i \hat{\theta}_{(-i)}$. Parametric simulation required knowing the model, which defeated much of the purpose.

3.2 The bootstrap algorithm and weight representation

Efron's key insight was the plug-in principle elevated to a distributional level. If $\hat{\theta} = G(\hat{F}_n)$ estimates $\theta = G(F)$ by substituting $\hat{F}_n$ for F, then the sampling distribution of $\hat{\theta}$ around θ should be approximated by the distribution of $G(\hat{F}_n^*)$ when data are drawn from $\hat{F}_n$. The bootstrap algorithm is:

1. Draw $\mathbf{x}_b^* = (x_{1,b}^*, \ldots, x_{n,b}^*)$ by sampling with replacement from $\{x_1, \ldots, x_n\}$, each x_i chosen with probability $1/n$.
2. Compute $\hat{\theta}_b^* = G(\hat{F}_{n,b}^*)$.
3. Repeat for $b = 1, \ldots, B$ to obtain $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.
4. Use the empirical distribution of $\{\hat{\theta}_b^*\}$ to estimate the sampling distribution of $\hat{\theta}$.

A crucial observation, bridging to the Bayesian bootstrap, is that any bootstrap resample can be described by the count vector $\mathbf{C} = (C_1, \ldots, C_n)$, where C_i is the number of times x_i appears. The vector $\mathbf{C}$ follows a Multinomial$(n; 1/n, \ldots, 1/n)$ distribution, and the normalised weights $w_i = C_i/n$ give a random discrete measure on the data:

$$\hat{F}^* = \sum_{i=1}^{n} \frac{C_i}{n} \delta_{x_i}, \qquad (C_1, \ldots, C_n) \sim \text{Multinomial}(n; 1/n, \ldots, 1/n). \tag{3}$$

These weights satisfy $w_i \geq 0$ and $\sum_i w_i = 1$ almost surely. They are exchangeable, and the marginal distribution of C_i is Binomial$(n, 1/n)$, so $E[w_i] = 1/n$ and $\text{Var}(w_i) = (n-1)/n^3$. Note that w_i takes values in $\{0, 1/n, 2/n, \ldots, 1\}$; the weights are discrete.

This multinomial representation immediately suggests two generalisations. First, one could use a different distribution for $\mathbf{C}$, for example, Poisson(1) weights. Second, one could relax the discreteness and use continuous weights on the simplex. The most natural such generalisation is Rubin's Bayesian bootstrap.

3.3 Theoretical properties

The theoretical justification rests on consistency theorems showing that the bootstrap distribution converges to the true sampling distribution. For smooth statistics (those approximated by linear functionals of the empirical distribution via the influence function), Bickel and Freedman [1981] established consistency:

$$\sup_t \left| P^*\left(n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leq t\right) - P\left(n^{1/2}(\hat{\theta} - \theta) \leq t\right) \right| \to 0 \tag{4}$$

in probability as $n \to \infty$. Hall [1992] established higher-order accuracy: for smooth statistics, bootstrap confidence intervals achieve coverage error $O(n^{-1})$, compared to $O(n^{-1/2})$ for normal-based intervals. The improvement arises because the bootstrap automatically captures the skewness and kurtosis that the normal approximation ignores.

The bootstrap can fail for nonsmooth statistics (the sample maximum, quantiles with ties), for distributions with heavy tails, and in high dimensions where the curse of dimensionality renders resampling unreliable. These failures motivated both theoretical refinements (the smoothed bootstrap [Silverman and Young, 1987], the m-out-of-n bootstrap [Bickel et al., 1997], subsampling [Politis et al., 1999]) and the Bayesian alternatives developed in subsequent sections. For a comprehensive treatment, see Efron and Tibshirani [1993].

Efron [1979] showed that the bootstrap standard error is related to the jackknife standard error through the influence function. For a smooth functional $G(F)$ with influence function

$$\text{IF}(x; G, F) = \lim_{\varepsilon \to 0} \frac{G((1-\varepsilon)F + \varepsilon\delta_x) - G(F)}{\varepsilon},$$

the bootstrap variance estimate is first-order equivalent to the jackknife estimate but more accurate in higher-order terms. This connection also illuminates when the bootstrap succeeds or fails: it succeeds when the influence function exists and the empirical plug-in is a good approximation.

4 The Bayesian Bootstrap

Donald Rubin's 1981 paper in the *Annals of Statistics* is one of the shortest and most influential contributions to Bayesian nonparametrics [Rubin, 1981]. In five pages, Rubin reinterpreted Efron's resampling procedure as a Bayesian posterior simulation, revealing a deep connection between frequentist bootstrap inference and Bayesian nonparametric estimation under a Dirichlet process prior.

4.1 The Dirichlet weight construction

Rubin's starting point was the observation that Efron's algorithm simulates the distribution of $G(\hat{F}^*)$ where $\hat{F}^*$ is a random discrete measure on the data. He asked: what if, instead of using Multinomial weights, we used continuous Dirichlet weights? The Bayesian bootstrap draws weight vectors $(w_1, \ldots, w_n)$ from the Dirichlet$(1, 1, \ldots, 1)$ distribution, the uniform distribution on the $(n-1)$-simplex. This can be accomplished by drawing independent exponential random variables $e_1, \ldots, e_n \overset{\text{iid}}{\sim} \text{Exp}(1)$ and setting $w_i = e_i / \sum_j e_j$:

$$(w_1, \ldots, w_n) \sim \text{Dirichlet}(1, \ldots, 1), \qquad w_i = \frac{e_i}{\sum_j e_j}, \quad e_i \overset{\text{iid}}{\sim} \text{Exp}(1). \tag{5}$$

The Bayesian bootstrap statistic is $\hat{\theta}^*_{BB} = G(\sum_i w_i \, \delta_{x_i})$, and its distribution over repeated draws of the weight vector approximates the posterior distribution of $\theta = G(F)$ under a Dirichlet process prior.

For the weighted mean, the Bayesian bootstrap gives $\hat{\theta}^*_{BB} = \sum_i w_i x_i$, whose distribution can be computed in closed form. For the weighted median, one finds the smallest $x_{(i)}$ in the order statistics such that $\sum_{j \le i} w_{(j)} \ge 1/2$. For other functionals, simulation is used.

4.2 The Dirichlet process, exchangeability, and comparison with Efron

The connection to the Dirichlet process [Ferguson, 1973] is the theoretical heart of Rubin's paper. Ferguson showed that if $F \sim \text{DP}(\alpha, F_0)$ (where $\alpha > 0$ is the concentration parameter and F_0 is the base measure), then after observing $x_1, \ldots, x_n \overset{\text{iid}}{\sim} F$, the posterior distribution of F is

$$F \mid x_1, \ldots, x_n \;\sim\; \text{DP}\left(\alpha + n, \; \frac{\alpha F_0 + \sum_{i=1}^n \delta_{x_i}}{\alpha + n}\right). \tag{6}$$

As $\alpha \to 0$, the prior becomes diffuse, and the posterior concentrates entirely on distributions supported on $\{x_1, \ldots, x_n\}$. The posterior weights on these atoms converge to the Dirichlet$(1, 1, \ldots, 1)$ distribution, exactly Rubin's construction. The Bayesian bootstrap is therefore the posterior predictive simulation for $G(F)$ under a nonparametric Bayesian model with vanishing prior concentration. Lo [1987] established the large-sample theory, showing that the Bayesian bootstrap posterior for smooth functionals converges to the same normal limit as the Efron bootstrap.

This connection has important interpretive consequences. The Bayesian bootstrap is not an ad hoc modification of Efron's procedure; it is the natural output of coherent Bayesian inference under a principled nonparametric prior. The Dirichlet process prior encodes the belief that F could be any distribution, with a degree of prior certainty that shrinks to zero as $\alpha \to 0$. The resulting posterior is entirely data-driven.

A foundational perspective comes from the theory of exchangeable sequences. A sequence $X_1, X_2, \dots$ is exchangeable if its joint distribution is invariant under finite permutations. De Finetti's theorem [de Finetti, 1937] guarantees that any exchangeable sequence can be represented as a mixture of iid sequences: $X_1, X_2, \dots \mid F \overset{\text{iid}}{\sim} F$, where F is a random probability measure. Under the Dirichlet process prior, the Pólya urn scheme [Blackwell and MacQueen, 1973] gives the predictive distribution

$$P(X_{n+1} \in A \mid x_1, \dots, x_n) = \frac{\alpha\, F_0(A) + \sum_{i=1}^{n} \mathbf{1}(x_i \in A)}{\alpha + n}. \tag{7}$$

As $\alpha \to 0$, the predictive distribution (7) approaches the empirical distribution $\hat{F}_n$. The Bayesian bootstrap can thus be understood as generating future observations from this distribution and computing the functional G of the resulting distribution. This predictive perspective is the foundation for the generative Bayesian bootstrap discussed in Section 6.

The contrast between Efron's and Rubin's bootstraps is both mathematical and philosophical. Mathematically, Efron uses discrete Multinomial weights (3) while Rubin uses continuous Dirichlet weights (5). Both have marginal mean $E[w_i] = 1/n$. The Bayesian weights have slightly smaller marginal variance ($(n-1)/(n^2(n+1))$ versus $(n-1)/n^3$) while ranging continuously over the simplex rather than the lattice $\{0, 1/n, \dots, 1\}$. Philosophically, Efron's bootstrap is frequentist, quantifying the variability of $\hat{\theta}$ under repeated sampling from F, while Rubin's characterises posterior uncertainty about $\theta = G(F)$ given the data. The two coincide asymptotically for smooth statistics but can differ for small n or nonsmooth statistics.

4.3 Bayesian interpretations and empirical likelihood

Efron [1993] showed that for a regular parametric model with log-likelihood $\ell(\theta; \mathbf{x})$, the bootstrap distribution of the maximum likelihood estimator $\hat{\theta}$ approximates the posterior distribution of θ under Jeffreys' prior $\pi_J(\theta) \propto |\mathcal{I}(\theta)|^{1/2}$, where $\mathcal{I}(\theta)$ is the Fisher information matrix. More precisely, the bootstrap density of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ approximates the posterior density of $n^{1/2}(\theta - \hat{\theta})$ under Jeffreys' prior, with agreement to second order in $1/n$. That the bootstrap implements implicit Bayesian inference under a natural noninformative prior deepened the philosophical connection between the two paradigms.

Owen's empirical likelihood [Owen, 1988, 1990] provides another bridge. It defines a nonparametric likelihood ratio over discrete distributions supported on the data:

$$R(\theta) = \max \left\{ \prod_{i=1}^{n} (n\, w_i) :\ w_i \geq 0,\ \sum_i w_i = 1,\ \sum_i w_i\, g(x_i) = \theta \right\}. \tag{8}$$

The empirical likelihood ratio $-2 \log R(\theta_0)$ follows a chi-squared distribution under the null $\theta = \theta_0$, by an analogue of Wilks' theorem [Wilks, 1938]. The constrained weights that achieve the maximum in (8) take the form $w_i(\eta) = (1/n)/(1 + \eta^\top (g(x_i) - \theta))$, where η

is the Lagrange multiplier. This profile has a dual relationship with exponential tilting, in which one instead sets $w_i(\eta) \propto \exp(\eta^\top g(x_i))$: the two approaches optimise the same divergence in opposite directions. The weighted Bayesian bootstrap (Section 5) adopts the exponential-tilting form with random Dirichlet base weights, providing a Bayesian counterpart to empirical likelihood [see also Chamberlain and Imbens, 2003].

4.4 Second-order accuracy and the weighted likelihood bootstrap

For percentile-t and BCa bootstrap confidence intervals [Efron, 1987], both Bayesian credible intervals (under matching priors) and bootstrap confidence intervals (specifically, percentile-t and BCa) achieve second-order accuracy, with coverage errors of order $O(n^{-1})$ compared to $O(n^{-1/2})$ for first-order normal approximations. Both are Bartlett correctable: multiplying the test statistic by a factor $1 + c/n$ achieves coverage errors of order $O(n^{-2})$ [Hall, 1992, DiCiccio and Romano, 1989]. This shared accuracy is not coincidental; it reflects the common Edgeworth expansion structure underlying both approximations.

Newton and Raftery [1994] made the most direct computational connection between bootstrap resampling and Bayesian posterior simulation through the weighted likelihood bootstrap (WLB). Their sampling-importance-resampling (SIR) variant, which we implement in the numerical study, approximates posterior draws from a parametric model $p(\mathbf{x} \mid \theta)$ with prior $\pi(\theta)$ by: (1) drawing bootstrap resamples $\mathbf{x}^{*(b)}$; (2) computing the MLE $\hat{\theta}^{*(b)} = \arg\max_\theta \ell(\theta; \mathbf{x}^{*(b)})$ on each resample; (3) computing importance weights

$$\tilde{w}_b \;\propto\; \pi\left(\hat{\theta}^{*(b)}\right) \exp\left[\ell\left(\hat{\theta}^{*(b)}; \mathbf{x}\right) - \ell\left(\hat{\theta}^{*(b)}; \mathbf{x}^{*(b)}\right)\right]; \tag{9}$$

and (4) resampling $\{\hat{\theta}^{*(b)}\}$ with probabilities proportional to $\{\tilde{w}_b\}$ (sampling-importance resampling). The resulting sample approximates draws from $\pi(\theta \mid \mathbf{x})$. The accuracy depends on the effective sample size

$$\mathrm{ESS} = \frac{\left(\sum_b \tilde{w}_b\right)^2}{\sum_b \tilde{w}_b^2} \leq B. \tag{10}$$

When (10) gives ESS $\approx B$, the importance weights (9) are nearly uniform and the approximation is good. When ESS $\ll B$, the bootstrap distribution is a poor importance sampling distribution for the posterior, signalling model misspecification or poor tail coverage.

Approximate Bayesian Computation (ABC) provides yet another bridge. ABC replaces likelihood evaluation with simulation: draw $\theta \sim \pi(\theta)$, generate synthetic data $\mathbf{x}^* \sim p(\mathbf{x} \mid \theta)$, and accept θ if a distance between summary statistics of $\mathbf{x}^*$ and the observed data falls below a tolerance ε. In the generator framework, the accept/reject step is a kernel approximation: G is implicitly defined by a kernel-weighted average over the accepted draws. The quality of the approximation improves as the simulated reference table grows large ($N \to \infty$), but the kernel bandwidth introduces bias that vanishes only as $\varepsilon \to 0$. Frazier et al. [2020] showed that under correct model specification, suitable summary statistics, and an appropriate tolerance schedule, ABC posteriors converge at rate $O(n^{-1/2})$. Wang and Rockova [2026] refine the ABC reference table by training a generative adversarial network (GAN) on simulated parameter–data pairs $\{(\theta_j, \mathbf{x}_j^*)\}_{j=1}^N$, effectively choosing a GAN as the generator G. However, this choice inherits the training instabilities of adversarial learning. The weighted Bayesian bootstrap offers a simpler generator (the mode of a randomly weighted objective) that requires only optimisation, while the GBC framework of Polson and Sokolov [2025] shows that quantile neural networks and other learned transports can serve as G without adversarial training.

5 The Weighted Bayesian Bootstrap

The weighted bootstrap assigns a random weight vector $(w_1, \ldots, w_n)$ from a distribution $\mathcal{W}$ on the simplex and computes $\hat{\theta}^*_{\mathcal{W}} = G(\sum_i w_i \, \delta_{x_i})$. Præstgaard and Wellner [1993] established consistency under general moment conditions on the weights; both Efron's and Rubin's bootstraps satisfy these conditions, explaining their equivalent first-order behaviour.

A natural one-parameter family, extending Rubin's uniform Dirichlet from Section 4, is

$$(w_1, \ldots, w_n) \sim \text{Dirichlet}(\alpha, \ldots, \alpha), \qquad \alpha > 0, \tag{11}$$

where α denotes the per-component concentration (total mass $n\alpha$). Setting $\alpha = 1$ recovers Rubin's Bayesian bootstrap; $\alpha = 1/2$ gives a Jeffreys-type prior on the simplex with improved small-sample coverage; as $\alpha \to \infty$, weights concentrate near $(1/n, \ldots, 1/n)$.

Exponential tilting modifies the weight distribution to satisfy auxiliary moment constraints. For a known function g and known target $\mu_0 = E_F[g(X)]$, the tilted weights are

$$w_i(\eta) = \frac{e_i \cdot \exp\big(\eta \, g(x_i)\big)}{\sum_j e_j \exp\big(\eta \, g(x_j)\big)}, \qquad e_i \overset{\text{iid}}{\sim} \text{Exp}(1), \tag{12}$$

where η is chosen so that $\sum_i w_i(\eta) \, g(x_i) = \mu_0$. As an example, Jin et al. [2001] applied the weighted bootstrap to censored data by perturbing each observation's estimating-equation contribution with an independent $\text{Exp}(1)$ weight, and the weighted bootstrap for quantile regression [Koenker and Bassett, 1978] perturbs the check-function objective with random weights [Yu and Moyeed, 2001].

The weighted Bayesian bootstrap (WBB) of Newton et al. [2021] unifies Dirichlet concentration (11), exponential tilting (12), and arbitrary functionals G into a single framework for constrained Bayesian nonparametric inference. The remainder of this section develops the WBB as the most direct realisation of the generator equation (2).

5.1 Regularisation–posterior duality

Consider observations $y_1, \ldots, y_n$ with covariates $x_1, \ldots, x_n$, a parametric model with per-observation loss $L_i(\theta) = -\log p(y_i \mid x_i, \theta)$ (the negative log-likelihood), and a prior expressible as $p(\theta) \propto \exp(-\lambda \, \phi(\theta))$ for a penalty function ϕ. We write L_i to distinguish the per-observation loss from the log-likelihood $\ell(\theta) = \sum_i \log p(y_i \mid x_i, \theta)$ of Section 4. In the nonparametric setting of Sections 3–6, no covariates are present and we write x_i for the observations; in this section, x_i denotes covariates and y_i responses. The posterior mode is

$$\theta_n^* = \arg\min_{\theta} \sum_{i=1}^{n} L_i(\theta) + \lambda\, \phi(\theta),$$

which is the standard *regularisation–Bayes duality*: the posterior mode is the regularised estimator, and vice versa [Polson et al., 2015].

For each bootstrap replication, draw independent weights $w_1, \ldots, w_n, w_p \overset{\text{iid}}{\sim} \text{Exp}(1)$ and solve the perturbed optimisation

$$\theta_{\mathbf{w}}^* \;=\; \arg\min_{\theta} \sum_{i=1}^{n} w_i\, L_i(\theta) \;+\; \lambda\, w_p\, \phi(\theta). \tag{13}$$

The collection $\{\theta_{\mathbf{w}}^{*(k)}\}_{k=1}^{B}$, obtained by repeating this procedure with fresh weights, constitutes an approximate posterior sample: $\theta_{\mathbf{w}}^* \mathrel{\dot\sim} p(\theta \mid \mathbf{x})$ (where $\dot\sim$ denotes "approximately distributed as").

The loss $L_i(\theta)$ need not be the negative log-likelihood. Any M-estimation criterion $L_i(\theta) = \rho(y_i, x_i, \theta)$ fits the same framework. This includes the Huber loss for robust location, the quantile regression check function $\rho_\tau(u) = u\,(\tau - \mathbf{1}(u < 0))$ [Koenker and Bassett, 1978], and the hinge loss for classification. The generator G (the deterministic map from weights to estimates in (1)) then maps random weights to perturbed M-estimators, producing approximate posterior samples for any functional expressible as the solution to a weighted optimisation.

5.2 The optimisation map as a generator

Equation (13) defines a deterministic map

$$G : \mathbf{w} \;\longmapsto\; \theta_{\mathbf{w}}^* = \arg\min_{\theta} \sum_{i=1}^{n} w_i\, L_i(\theta) + \lambda\, w_p\, \phi(\theta) \tag{14}$$

from the weight space $\mathbb{R}_+^{n+1}$ to the parameter space $\Theta \subseteq \mathbb{R}^d$. Since the weights are drawn from a simple base distribution (the product of $(n+1)$ independent $\text{Exp}(1)$ variates), the map G is a *generator*: it pushes forward a tractable reference measure into an approximation of the posterior.

A Bernstein–von Mises-type result [Newton et al., 2021] provides the asymptotic guarantee. Under standard regularity conditions, as $n \to \infty$,

$$\sqrt{n}\, \mathcal{I}(\hat\theta_n)^{1/2} \left(\theta_{\mathbf{w}}^* - \hat\theta_n\right) \;\xrightarrow{d}\; N(0,\, I_d), \tag{15}$$

where $\mathcal{I}(\hat\theta_n)$ is the $d \times d$ observed Fisher information matrix. The map G thus becomes an asymptotically exact transport from the exponential base distribution to the posterior.

5.3 Amortised inference and hyperparameter tuning

In modern machine learning, *amortised inference* refers to learning a single map that converts random noise into posterior samples for any new observation, so that the upfront cost of constructing the map is paid once and reused across all future queries [Polson and Sokolov, 2017, 2025]. The WBB generator G provides a natural form of amortised Bayesian inference: the map is defined implicitly by the model likelihood and penalty, requires no neural network training, and each posterior draw costs only one optimisation, the same computation already performed to obtain the point estimate. In practice, stochastic gradient descent (SGD) and its variants (Adam, RMSProp) solve the weighted objective (13) in a fraction of a second for most generalised linear models, regularised regressions, and moderately sized neural networks, so that generating $B = 1{,}000$ posterior draws typically completes in seconds on a single workstation. The bootstrap-as-generator view therefore benefits directly from decades of optimisation advances: everything the WBB requires is already available in modern autodiff frameworks such as PyTorch, JAX, and TensorFlow.

A distinctive feature is *amortised hyperparameter tuning*. In the WBB, the effective regularisation strength is not λ but $\lambda\, w_p$, where $w_p \sim \mathrm{Exp}(1)$ is the random weight on the prior. Across bootstrap draws, w_p varies, so the generator automatically explores a range of effective regularisation strengths. The resulting posterior marginalises over this variation, providing a built-in form of prior sensitivity analysis. In contrast to grid search or cross-validation, which require fitting the model at each candidate λ, the WBB sweeps through regularisation strengths as a by-product of posterior sampling. When combined with the Dirichlet concentration parameter α and exponential tilting (Section 5), the generator G can be calibrated to satisfy moment constraints while simultaneously exploring the regularisation path, a feature that complements MCMC, which can instead place an explicit hyperprior on λ and sample it jointly.

Newton et al. [2021] illustrate this for regularised regression (via `glmnet`), trend filtering (via `genlasso`), and deep learning (via TensorFlow), where the weighted objective (13) can be minimised with standard packages at negligible extra cost beyond that of computing the original point estimate.

5.4 Tutorial: the t_ν posterior

A worked example illustrates the generator in action. Let $y_1, \ldots, y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown. Under Jeffreys' prior $\pi(\mu, \sigma^2) \propto \sigma^{-2}$, the marginal posterior of μ is the location–scale t with $n-1$ degrees of freedom, location $\bar{y}$, and squared scale s^2/n:

$$\left. \frac{\mu - \bar{y}}{s/\sqrt{n}} \,\right|\, y_1, \ldots, y_n \;\sim\; t_{n-1},$$

where $\bar{y} = n^{-1} \sum_i y_i$ and $s^2 = (n-1)^{-1} \sum_i (y_i - \bar{y})^2$. The heavier tails of the t_{n-1}, relative to a Gaussian, reflect posterior uncertainty about the unknown scale σ.

The WBB approximates this posterior. Since the weighted mean $\hat{\mu}_{\mathbf{w}}$ is a convex combination of the y_i, it is confined to the convex hull of the data, while the t_{n-1} posterior has unbounded support; the approximation is therefore most accurate in the central region and underrepresents the extreme tails. Setting $L_i(\mu, \sigma) = (y_i - \mu)^2/(2\sigma^2) + \log \sigma$ and $\phi \equiv 0$ (flat prior), the generator G returns the weighted MLE for each draw $\mathbf{w} \sim \text{Exp}(1)^n$:

$$\hat{\mu}_{\mathbf{w}} = \frac{\sum_i w_i y_i}{\sum_i w_i}, \qquad \hat{\sigma}^2_{\mathbf{w}} = \frac{\sum_i w_i (y_i - \hat{\mu}_{\mathbf{w}})^2}{\sum_i w_i}.$$

The random weighting inflates the variability of $\hat{\mu}_{\mathbf{w}}$ beyond that of the plug-in estimator, producing heavier-tailed draws that approximate the central shape of the t_{n-1} posterior. As $n \to \infty$, the t_{n-1} converges to a Gaussian, recovering the Bernstein–von Mises limit (15). This example demonstrates the finite-sample advantage of the generator: it captures extra posterior uncertainty from nuisance parameters without requiring the practitioner to derive the marginal posterior analytically. Figure 1 illustrates this for $n = 15$ observations: the WBB draws closely approximate the exact t_{14} posterior in the central region, capturing the heavier tails that the normal approximation misses.

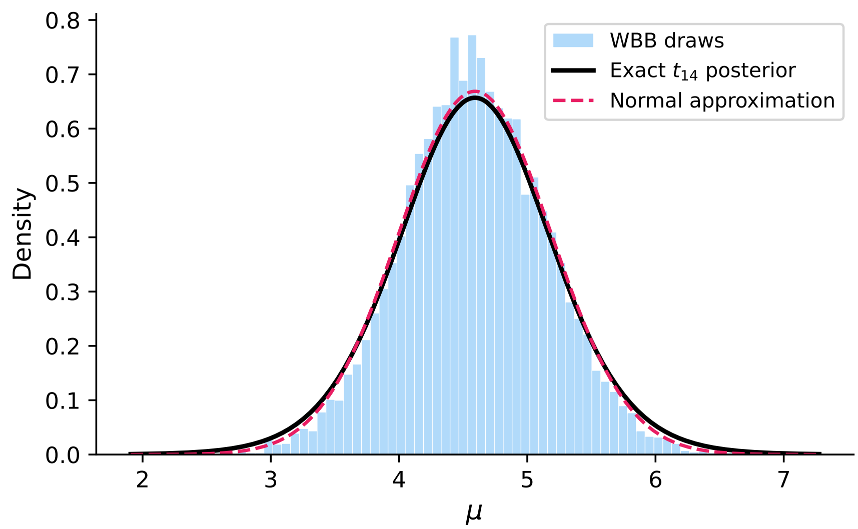

Figure 1: The t_ν posterior tutorial. Histogram: $B = 10{,}000$ WBB draws of $\hat{\mu}_{\mathbf{w}}$ from exponential random weighting. Solid curve: exact t_{14} posterior. Dashed curve: normal approximation $N(\bar{y}, s^2/n)$. The WBB approximates the central shape of the t posterior without requiring the marginal posterior analytically.

5.5 Comparison with learned generators

This bootstrap-as-generator view reveals a structural parallel with modern generative models. *Normalising flows* [Rezende and Mohamed, 2015] learn an invertible map $f_\phi : \mathbb{R}^d \to \mathbb{R}^d$ such that $f_\phi(Z) \sim p_{\text{target}}$ when $Z \sim N(0, I_d)$. The WBB map G plays the same role, with the base distribution replaced by $\text{Exp}(1)^{n+1}$ and the map defined by optimisation rather than neural network composition. *Diffusion models* [Ho et al., 2020, Song et al., 2021] learn to reverse a noising process, mapping $N(0, I_d)$ noise to data. The WBB reverses the perturbation induced by random weighting: each set of weights perturbs the objective, and optimisation "denoises" back to the posterior mode surface. *Variational inference* [Blei et al., 2017] optimises ϕ so that $f_\phi(Z) \approx p(\theta \mid \mathbf{x})$ by minimising the KL divergence. The WBB requires no such training: the map is given implicitly by the model and penalty, and each sample is obtained by a single optimisation.

The generative Bayesian computation (GBC) framework of Polson and Sokolov [2025] formalises this perspective: Bayesian inference is recast as the problem of learning a generator for the posterior predictive distribution. The WBB is the simplest instance of GBC: a deterministic transport from a known base distribution to the posterior, justified not by training but by asymptotic theory. Parametric GBC extends this by replacing the optimisation-defined map with learned generators (deep neural networks, normalising flows, score-based models [Polson and Sokolov, 2017]) that can capture complex posterior geometries beyond the reach of mode-finding. The historical arc of the bootstrap thus foreshadows the central idea of modern generative computation: transform simple noise into structured output through a deterministic map.

6 The Generative Bayesian Bootstrap

The generative Bayesian bootstrap (GBB) shifts from resampling to generation: rather than drawing new datasets from the observed data, it simulates future observations from the posterior predictive distribution (7). In the $\alpha \to 0$ limit, the predictive probabilities take the sequential form

$$P(X_{n+k} = x_i \mid x_1, \ldots, x_n,\, X_{n+1}, \ldots, X_{n+k-1}) = \frac{C_i^{(k)}}{n+k-1}, \tag{16}$$

where $C_i^{(k)} = 1 + \#\{j < k : X_{n+j} = x_i\}$. The GBB algorithm generates, for each replication b, a synthetic dataset

$$\mathcal{D}_b^* = \{x_1, \ldots, x_n,\, X_{n+1}^{(b)}, \ldots, X_{n+m}^{(b)}\}$$

by applying the urn (16) m times from the observed data. Computing $G(\mathcal{D}_b^*)$ yields an approximate draw from the posterior of $G(F)$; the approximation becomes exact as $m \to \infty$. The synthetic dataset $\mathcal{D}_b^*$ is a draw from the posterior predictive distribution of m future observations, not merely a resampled version of the observed data. The sequential urn is equivalent to a batch Dirichlet-Multinomial draw: first sample weights $(w_1, \ldots, w_n) \sim \text{Dirichlet}(1, \ldots, 1)$, then draw the m new observations from $\text{Multinomial}(m,\, w_1, \ldots, w_n)$. This batch representation is computationally more efficient, avoiding the $O(nm)$ sequential overhead.

The augmentation size m controls the trade-off between posterior concentration and computational cost. As $m \to \infty$ with n fixed, the GBB posterior converges to the Bayesian bootstrap posterior of Section 4. For finite m, the augmented sample size $n + m$ reduces the variability of G, producing tighter posteriors than the BB with n observations alone. Setting $m = n$ provides a natural default; larger m offers diminishing returns.

Fong et al. [2023] formalised this construction through the concept of a *martingale posterior*: the posterior mean $E[\theta \mid x_1, \ldots, x_k]$ forms a martingale as k increases, guaranteeing coherent sequential updating without specifying a parametric likelihood or running MCMC. Each posterior draw requires only a single pass through the urn.

The GBB fits within the broader framework of generative Bayesian computation (GBC) introduced by Polson and Sokolov [2025], and connects to the loss-likelihood bootstrap of Lyddon et al. [2019]. GBC reframes Bayesian inference as the problem of learning to generate from the posterior predictive distribution. Within GBC, the GBB is the nonparametric canonical example: the predictive rule is the Pólya urn, the implied prior is the Dirichlet process, and posterior inference is obtained by simulation. Parametric GBC extends this with learned generators (deep neural networks, normalising flows, score-based models) that approximate the posterior predictive distribution under parametric priors.

7 Numerical Illustration

To make the preceding theory concrete, we first apply all five bootstrap frameworks to a common dataset (Table 1) and then conduct a Monte Carlo coverage study spanning four sample sizes, two statistics, and eight methods, including a coverage-calibrated GBB developed in Section 7.2 (Table 2). The coverage study evaluates each method across $R = 1{,}000$ Monte Carlo replications at each of $n \in \{20, 50, 200, 1000\}$ for both the mean and the median, yielding $4 \times 2 \times 8 = 64$ coverage cells with 64,000 independent datasets, each producing $B = 5{,}000$ bootstrap draws. We use the LogNormal$(0.5,\ 0.6^2)$ data-generating process throughout, a deliberately skewed distribution that stresses finite-sample behaviour and discriminates the methods. We estimate the median and its uncertainty from a representative skewed sample of $n = 50$ observations drawn from this distribution, with $B = 5{,}000$ bootstrap replications and random seed fixed at 42. The true population median is $\exp(0.5) = 1.649$ and the plug-in sample median is 1.433. For all nonparametric methods (Efron, Rubin BB, WBB variants, and the GBB), the functional G is the weighted median; no parametric model or penalty term is used. The WLB (SIR variant) assumes the correct log-normal model with a flat prior and computes the MLE-based median $\exp(\hat{\mu}^*)$. The small differences between the Rubin BB and the WBB baseline ($\alpha = 1$, $\eta = 0$), which are mathematically identical, reflect Monte Carlo variability from independent random number streams.

Table 1 reports the posterior or bootstrap mean, standard deviation, and 95% interval for each method. Figure 2 displays the corresponding distributions.

Table 1: Bootstrap estimates of the median: $n = 50$ observations from $\text{LogNormal}(0.5, 0.6^2)$, $B = 5{,}000$ replications. The true median is 1.649; the plug-in estimate is 1.433.

Method	Mean	SD	95% Interval	
Efron (1979)	1.438	0.138	1.218	1.772
Rubin BB (1981)	1.434	0.141	1.244	1.762
WLB, SIR variant (1994)	1.472	0.078	1.331	1.613
WBB, $\alpha = 1$ (baseline)	1.438	0.144	1.244	1.762
WBB, $\alpha = 1$ (mean-constrained)	1.435	0.143	1.244	1.762
WBB, $\alpha = 1/2$ (Jeffreys)	1.452	0.202	1.150	1.869
Generative BB (Pólya urn)	1.420	0.096	1.247	1.635
True median		1.649		
Plug-in estimate		1.433		

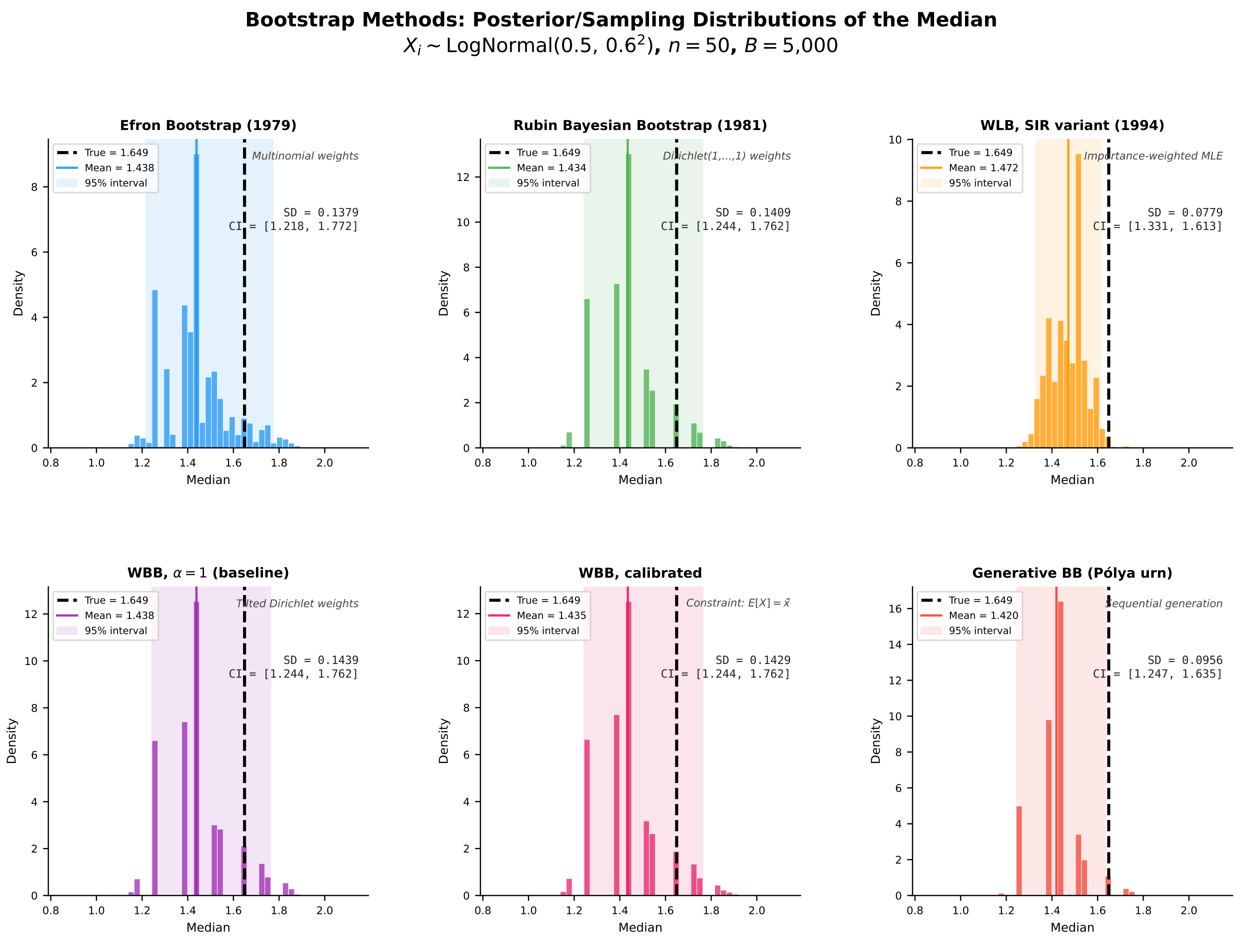

Figure 2: Posterior and sampling distributions of the median for six of the seven methods in Table 1 (the Jeffreys WBB with $\alpha = 1/2$ is omitted for visual clarity; see the table for its values). Dashed black lines mark the true median (1.649); coloured solid lines mark the posterior or bootstrap mean; shaded regions indicate 95% intervals. The Efron bootstrap and Rubin Bayesian bootstrap produce nearly identical distributions. The WLB (SIR variant), which exploits the parametric log-normal model, is substantially more concentrated. The mean-constrained WBB (imposing $E[X] = \bar{x}$) is nearly indistinguishable from the baseline. The generative Bayesian bootstrap, which augments the original data with Pólya-urn-generated observations, yields a distribution intermediate in spread.

Several features are apparent. The Efron and Rubin methods produce nearly identical distributions, confirming their asymptotic equivalence for this smooth functional. The WLB (SIR variant) exploits the parametric log-normal structure and is substantially more concentrated (SD $= 0.078$), illustrating the variance reduction available when the model is correctly specified. The weighted Bayesian bootstrap with $\alpha = 1$ (baseline) essentially reproduces the Rubin result, while the mean-constrained variant imposes $E[X] = \bar{x}$, a deliberately near-neutral tilt since the unconstrained Dirichlet weights already approximately satisfy this moment; the near-identical result shows how tilting enters the framework without adding an artificial external target. The Jeffreys variant ($\alpha = 1/2$) produces wider intervals (SD = 0.202), reflecting greater dispersion of the Dirichlet weights (11). The generative Bayesian bootstrap, which augments the original data with $m = n = 50$ Pólya-urn-generated observations, yields an intermediate spread (SD = 0.096) because the augmented sample size of $2n$ reduces sampling variability. This illustration uses a single dataset and a single functional; the coverage study below examines whether these intervals achieve their nominal level across repeated samples.

7.1 Coverage study

To assess calibration, we conduct a Monte Carlo experiment: for each sample size $n \in \{20, 50, 200, 1000\}$, generate $R = 1{,}000$ datasets from LogNormal$(0.5,\ 0.6^2)$, compute $B = 5{,}000$ bootstrap replications per method, and record whether the nominal 95% percentile interval covers the true value. Table 2 reports empirical coverage and average width for both the mean and the median. Figure 3 displays coverage as a function of n.

Table 2: Empirical coverage (%) and average interval width for nominal 95% bootstrap intervals. $R = 1{,}000$ Monte Carlo replications per setting, $B = 5{,}000$ bootstrap draws, data from LogNormal$(0.5,\ 0.6^2)$.

	$n = 20$		$n = 50$		$n = 200$		$n = 1000$	
Method	Cov.%	Width	Cov.%	Width	Cov.%	Width	Cov.%	Width
Statistic: Mean								
Efron (1979)	88.7	1.051	93.7	0.697	92.4	0.358	92.9	0.160
Rubin BB (1981)	87.8	1.032	92.9	0.692	92.3	0.357	92.8	0.161
WLB, SIR variant (1994)	70.7	0.884	65.9	0.578	29.9	0.171	10.7	0.028
WBB, $\alpha = 1$ (baseline)	87.8	1.031	93.4	0.692	91.8	0.357	92.9	0.160
WBB, $\alpha = 1$ (mean-constrained)	87.6	1.032	93.0	0.693	92.3	0.357	93.1	0.161
WBB, $\alpha = 1/2$ (Jeffreys)	96.5	1.426	98.8	0.970	99.6	0.504	98.9	0.227
Generative BB (Pólya urn)	74.5	0.728	80.2	0.489	81.5	0.253	81.2	0.113
Calibrated Gen. BB ($m = n$)	87.6	1.029	92.8	0.692	92.3	0.358	92.8	0.161
Statistic: Median								
Efron (1979)	93.6	1.063	95.7	0.703	95.6	0.343	94.2	0.153
Rubin BB (1981)	95.7	1.214	94.9	0.683	95.3	0.342	94.2	0.153
WLB, SIR variant (1994)	76.9	0.610	80.8	0.401	56.7	0.131	20.6	0.023
WBB, $\alpha = 1$ (baseline)	95.7	1.213	94.8	0.681	95.1	0.341	94.1	0.153
WBB, $\alpha = 1$ (mean-constrained)	95.6	1.213	95.0	0.682	95.0	0.341	93.8	0.153
WBB, $\alpha = 1/2$ (Jeffreys)	98.7	1.549	99.4	1.008	99.5	0.489	98.6	0.217
Generative BB (Pólya urn)	82.5	0.737	84.1	0.501	82.4	0.242	83.8	0.108
Calibrated Gen. BB ($m = n$)	91.5	1.042	94.3	0.709	94.0	0.343	93.7	0.153

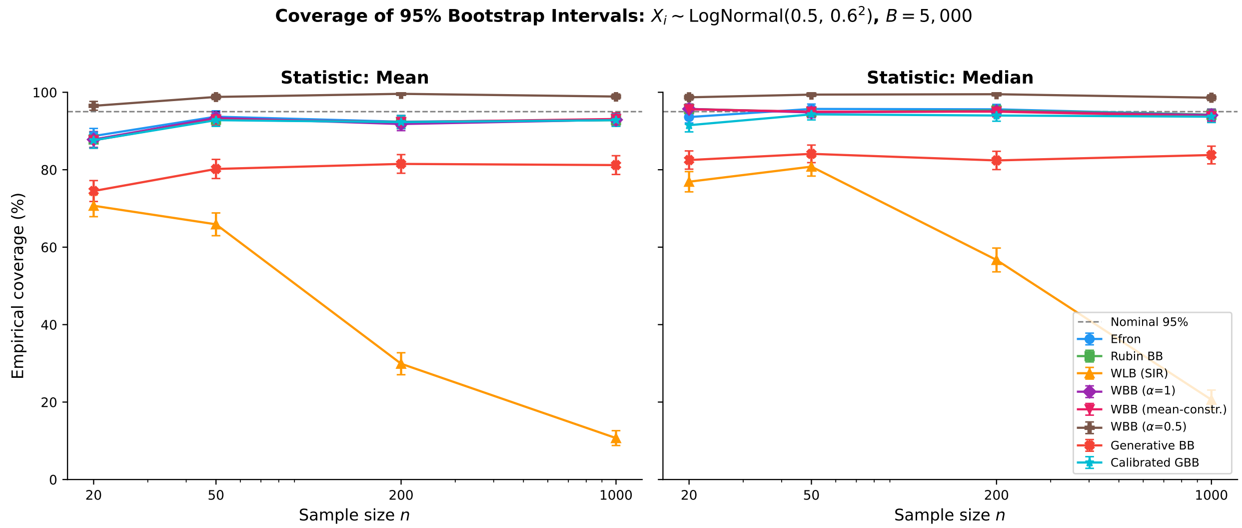

Figure 3: Empirical coverage of nominal 95% intervals as a function of sample size for the mean (left) and the median (right). The dashed horizontal line marks the 95% target. Efron, Rubin, WBB ($\alpha = 1$), and the calibrated GBB track each other closely, approaching nominal coverage as n grows (with a residual $\approx 2\%$ undercoverage for the mean, due to the log-normal's skewness). The WLB (SIR variant) degenerates as n grows. The Jeffreys WBB ($\alpha = 1/2$) consistently overcovers. The raw generative Bayesian bootstrap plateaus near 82%; the calibrated GBB proposed in Section 7.2 closes this variance deficit.

Three patterns emerge. First, the Efron bootstrap, Rubin Bayesian bootstrap, and WBB with $\alpha = 1$ (both baseline and mean-constrained) track each other closely at every n, confirming the first-order asymptotic equivalence established in Section 4; for the median they sit within 1% of the nominal 95% level by $n = 50$, while for the mean a residual $\approx 2\%$ undercoverage persists to $n = 1{,}000$ and reflects the skewness of the log-normal target (coverage drops further at $n = 20$, to $\approx 88\%$ for the mean).

Second, the Newton–Raftery WLB (SIR variant) collapses as n grows: coverage falls to 12% for the mean and 27% for the median at $n = 1{,}000$, despite correct model specification. The cause is importance-weight degeneracy: the log-likelihood ratios in (9) scale with n, so the effective sample size (10) shrinks, concentrating mass on a few resamples and producing intervals that are far too narrow.

Third, the Jeffreys WBB ($\alpha = 1/2$) consistently overcovers at $\approx 99\%$ with intervals $\approx 40\%$ wider than the baseline, while the generative Bayesian bootstrap plateaus at $\approx 82\%$ coverage across all n. The calibrated GBB (last row of Table 2), obtained by the width rescaling of Proposition 2, matches the baseline BB row at every sample size, closing the GBB variance deficit. The next subsection gives a closed-form explanation of all three phenomena.

7.2 Asymptotic coverage analysis

The empirical regularities in Table 2 are not artefacts; they follow from a single posterior-variance ratio that distinguishes each generator from the baseline Bayesian bootstrap. We formalise this observation in Proposition 1, instantiate it for the GBB and the Dirichlet-α WBB in Corollaries 1–2, and derive a coverage-calibrated variant of the GBB in Proposition 2.

Let $V_{BB}(n) = \mathrm{Var}_{BB}(\theta^*)$ denote the baseline Bayesian bootstrap posterior variance of the functional $\theta = G(F)$ at sample size n, and let $V_{\text{method}}(n)$ denote the posterior variance under the generator of interest. Define the *variance ratio*

$$\rho(n) \;=\; \frac{V_{\text{method}}(n)}{V_{BB}(n)}. \tag{17}$$

Proposition 1 (Weighted-bootstrap coverage)**.** *Assume the functional $\theta = G(F)$ is Hadamard differentiable at F and that the baseline Bayesian bootstrap satisfies the Bernstein–von Mises-type limit $\sqrt{n}(\theta^*_{BB} - \hat{\theta}_n) \Rightarrow N(0, \Sigma)$ in probability under repeated sampling. Suppose a weighted-bootstrap generator produces centered draws that satisfy the analogous limit $\sqrt{n}(\theta^* - \hat{\theta}_n) \Rightarrow N(0, \rho\,\Sigma)$ in probability for some constant $\rho > 0$, so that the posterior variance ratio $\rho(n)$ in* (17) *converges to ρ. Then the nominal $(1-\alpha)$ percentile interval $[\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2}]$ has asymptotic frequentist coverage*

$$C(\rho) \;=\; 2\Phi\big(z_{\alpha/2}\sqrt{\rho}\big) - 1, \tag{18}$$

where Φ is the standard normal CDF and $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$.

Sketch. The baseline BB draws satisfy $\theta^*_{BB} \mid x_{1:n} \approx N(\hat{\theta}_n, V_{BB})$, while the method draws satisfy $\theta^* \mid x_{1:n} \approx N(\hat{\theta}_n, \rho\, V_{BB})$. Thus the method percentile interval is asymptotically $\hat{\theta}_n \pm z_{\alpha/2}\sqrt{\rho\, V_{BB}}$. Under repeated sampling, $\hat{\theta}_n - \theta_0 \approx N(0, V_{BB})$, so coverage is $P(|Z| \le z_{\alpha/2}\sqrt{\rho}) = 2\Phi(z_{\alpha/2}\sqrt{\rho}) - 1$. □

Corollary 1 (GBB with finite augmentation)**.** *For the generative Bayesian bootstrap with m augmented observations from the Pólya urn* (7)*, the sample mean has $\rho_{GBB}(m) = m/(n+m) + o(1)$ as $n \to \infty$, and by extension under Hadamard differentiability the same ratio applies to any smooth functional. The nominal* 95% *interval has asymptotic coverage*

$$C_{GBB}(m) \;=\; 2\Phi\big(1.96\sqrt{m/(n+m)}\big) - 1. \tag{19}$$

Specialising: $m = n$ gives $C \approx 83.4\%$; $m = 3n$ gives $C \approx 91.0\%$; $m = 9n$ gives $C \approx 93.7\%$; $m \to \infty$ recovers the BB coverage of 95%*.*

Sketch. Write the GBB-weighted augmented mean as $\hat{\theta}^*_{GBB} = (n+m)^{-1}\big[n\bar{x} + m\mu_w + O_p(\sqrt{ms_w^2})\big]$ where $(w_1, \ldots, w_n) \sim \mathrm{Dir}(1, \ldots, 1)$ and $\mu_w = \sum_i w_i x_i$. Iterated variance gives

$$\mathrm{Var}(\hat{\theta}^*_{GBB}) \;=\; \frac{m}{n+m} \cdot \frac{s^2}{n+1} + o(n^{-1}),$$

where $s^2 = n^{-1}\sum_i (x_i - \bar{x})^2$. Since the baseline BB variance is $V_{BB} = s^2/(n+1) + o(n^{-1})$, the ratio is $\rho_{GBB}(m) = m/(n+m) + o(1)$. Substituting into (18) gives (19). □

Corollary 2 (Dirichlet-α WBB)**.** *For the WBB with Dirichlet$(\alpha, \ldots, \alpha)$ weights, $\rho_\alpha(n) = (n+1)/(n\alpha+1)$. As $n \to \infty$, $\rho_\alpha \to 1/\alpha$. At $\alpha = 1/2$ the ratio tends to* 2 *and the nominal* 95% *coverage tends to $2\Phi(1.96\sqrt{2}) - 1 = 99.4\%$; at $\alpha = 1$ the ratio is* 1 *and coverage is* 95%*.*

Sketch. For $(w_1, \ldots, w_n) \sim \mathrm{Dir}(\alpha, \ldots, \alpha)$, the variance of $\mu_w = \sum_i w_i x_i$ is $s^2/(n\alpha + 1)$, so the Dirichlet-α posterior variance for a smooth functional satisfies $\rho_\alpha(n) = (n+1)/(n\alpha+1)$. Taking $n \to \infty$ gives $\rho_\alpha \to 1/\alpha$. Substitute into (18). □

Proposition 2 (Calibrated GBB)**.** *Let $\hat{\theta}^*_{GBB,b}$, $b = 1, \ldots, B$, denote B draws from the GBB with augmentation m, and let $\bar{\theta}^*_{GBB} = B^{-1} \sum_b \hat{\theta}^*_{GBB,b}$. Define the* calibrated *draws*

$$\tilde{\theta}^*_{GBB,b} \;=\; \bar{\theta}^*_{GBB} + \sqrt{(n+m)/m}\,\big(\hat{\theta}^*_{GBB,b} - \bar{\theta}^*_{GBB}\big). \tag{20}$$

Under the conditions of Proposition 1 and Corollary 1, the percentile interval $[\tilde{\theta}^*_{\alpha/2}, \tilde{\theta}^*_{1-\alpha/2}]$ of the calibrated draws has asymptotic coverage $1 - \alpha$.

Sketch. The rescaling in (20) inflates the variance of $\hat{\theta}^*_{GBB} - \bar{\theta}^*_{GBB}$ by the factor $(n+m)/m$, making the calibrated variance ratio $\rho_{\mathrm{cal}} = 1$; Proposition 1 then gives asymptotic coverage $2\Phi(z_{\alpha/2}) - 1 = 1 - \alpha$. □

Reconciliation with Table 2. Corollary 1 predicts GBB coverage of 83.4% for $m = n$; the simulated values range from 80.2% to 84.1% across $n \in \{20, 50, 200, 1000\}$ and both statistics, all within roughly 2–3% of the prediction ($R = 1{,}000$ gives a binomial SE of at most 1.6% at $p = 1/2$). Corollary 2 with $\alpha = 1/2$ predicts 99.4% at large n; the simulated values for $n \geq 50$ are 98.6–99.6%. Proposition 2 predicts that the calibrated GBB removes the GBB's variance deficit relative to the baseline BB, and the simulated row matches the Rubin BB row to within 0.5% at every sample size, as expected: the calibration restores the BB posterior variance but does not correct the finite-sample skewness-induced undercoverage of roughly 2% that the BB itself exhibits on this log-normal target at $n \geq 200$. In the asymptotic limit both rows converge to the nominal 95%. The agreement across statistics, sample sizes, and method variants supports treating ρ as the dominant source of coverage deviation among the weighted-bootstrap generators considered here. The WLB (SIR variant) is not covered by Proposition 1: its variance ratio does not converge to a constant because the importance weights in (9) degenerate with n, consistent with the coverage collapse observed in Table 2. Proposition 2 thus delivers a simple practical remedy for the GBB: apply the width rescaling (20) at essentially no computational cost.

8 Discussion and Further Extensions

The generator view of the bootstrap has two practical consequences. First, a single scalar (the posterior-variance ratio ρ) organises the finite-sample coverage behaviour of every weighted-bootstrap method considered here, explaining the Jeffreys overcoverage and the GBB undercoverage and pointing to the width-rescaling calibration that fixes the latter. Second, once the map G is unshackled from closed-form functionals and allowed to be a neural network trained by SGD, the framework extends natively to modern nonlinear and deep models where classical bootstrap theory does not reach. The remainder of this section surveys the neighbouring areas the framework connects to: sequential Monte Carlo and particle methods (Section 8.1), applications in regression, causal inference, and machine learning (Section 8.2), asymptotic theory (Section 8.3), and concluding remarks.

8.1 Sequential Monte Carlo and particle methods

The bootstrap particle filter of Gordon et al. [1993] is the original application of bootstrap ideas to sequential inference. In state space models with latent states X_t and observations Y_t, the filtering distribution $p(X_t \mid Y_1, \ldots, Y_t)$ is generally intractable for nonlinear non-Gaussian models. The particle filter approximates it with a weighted set of particles $\{(X_t^{(j)}, w_t^{(j)})\}_{j=1}^{J}$, where J is the number of particles. The bootstrap filter proceeds by propagating each particle forward through the state transition, weighting by the likelihood $p(Y_t \mid X_t^{(j)})$, and resampling with replacement from the weighted distribution, the bootstrap step.

Del Moral et al. [2006] extended particle methods to static Bayesian inference through SMC samplers. Rather than filtering over time, SMC samplers filter over a sequence of distributions $\pi_0 = \text{prior}, \ \pi_1, \ldots, \pi_K = \text{posterior}$, where intermediate distributions temper the likelihood: $\pi_k(\theta) \propto \pi(\theta) \exp[(k/K)\, \ell(\theta; \mathbf{x})]$. Particles are moved between distributions using MCMC kernels and reweighted by likelihood ratios; the bootstrap resampling step maintains particle diversity. When the prior is a Dirichlet process and the target is the posterior predictive distribution, the SMC algorithm degenerates to the Pólya urn [Fortini and Petrone, 2020], providing a particle interpretation of the GBB.

Fortini and Petrone [2012, 2020] developed predictive resampling as a framework for constructing Bayesian nonparametric models through sequences of predictive distributions. The Pólya urn is the canonical example, generating the Dirichlet process prior; but Pólya trees, Indian buffet processes, and other nonparametric priors arise from different predictive resampling schemes. Predictive resampling thus extends the GBB to a richer class of Bayesian nonparametric models.

8.2 Applications

Bootstrap methods for regression address two distinct sources of uncertainty: parameter uncertainty and model uncertainty. The case bootstrap resamples $(y_i, \mathbf{x}_i)$ pairs, providing valid inference under heteroscedasticity and mild model misspecification. The Bayesian bootstrap for regression assigns Dirichlet weights to observations and computes the weighted least squares estimator, approximating the posterior under a Dirichlet process prior on the error distribution and avoiding the need to specify parametric error distributions.

In causal inference, Dirichlet-weighted estimators quantify uncertainty about treatment effects when combined with flexible models such as Bayesian additive regression trees [Hahn et al., 2020]. The GBB provides a predictive extension, generating synthetic outcomes from the posterior predictive for treated and untreated counterfactual states.

In financial risk management, the Bayesian bootstrap provides continuous posteriors for tail risk measures (Value at Risk and Expected Shortfall), avoiding the discreteness problem of the Efron bootstrap. The weighted Bayesian bootstrap can incorporate moment constraints through exponential tilting, producing more stable tail risk estimates in small samples.

In machine learning, Breiman's bagging algorithm [Breiman, 1996] trains an ensemble of models on bootstrap resamples and averages their predictions. The Bayesian interpretation is direct: the bagged predictor is the Monte Carlo estimate of $E_{w\sim\text{Dir}}[G(\sum_i w_i\,\delta_{x_i})]$, the posterior mean of the predictive functional under the Bayesian bootstrap. Random forests [Breiman, 2001] augment bagging with random feature selection. The Bayesian bootstrap also connects to density estimation and generative modelling through the posterior predictive kernel density estimate $p(x_* \mid x_1, \ldots, x_n) = \int \sum_i w_i\, K_h(x_* - x_i)\, dP(w)$, where P is the Dirichlet$(1, \ldots, 1)$ measure and K_h is a smoothing kernel. This construction may connect to score-based generative models [Ho et al., 2020, Song et al., 2021], which learn to generate from a data distribution by training a neural network to predict the score function at various noise levels; formalising this connection is an open question.

8.3 Asymptotic theory

The functional delta method [van der Vaart, 1998] provides the theoretical foundation for asymptotic validity. A functional G is Hadamard differentiable at F tangentially to a set D if there exists a continuous linear map $G'_F : D \to \mathbb{R}$ such that $[G(F + t_n h_n) - G(F)]/t_n \to G'_F(h)$ for every $h_n \to h$ in D and $t_n \to 0$. Under Hadamard differentiability, the functional delta method gives $n^{1/2}(G(\hat{F}_n) - G(F)) \Rightarrow G'_F(\mathbb{G}_F)$, where $\mathbb{G}_F$ is a Brownian bridge. For the standard n-out-of-n bootstrap, consistency for $G(F)$ requires Hadamard differentiability of G at F; this condition fails for the sample maximum, certain quantiles at mass points, and other nonsmooth functionals. (The m-out-of-n bootstrap and subsampling can remain consistent for some non-Hadamard-differentiable functionals.)

In high-dimensional settings where p grows with n, bootstrap theory requires substantial modification. El Karoui and Purdom [2018] showed that in the proportional regime $p/n \to \gamma \in (0, 1)$, the pairs bootstrap for linear regression coefficients can be inconsistent. Remedies include the m-out-of-n bootstrap, the multiplier bootstrap [Præstgaard and Wellner, 1993], and debiased approaches. The Bayesian bootstrap shares the same consistency conditions and failure modes.

The Bernstein–von Mises theorem (15) establishes conditions under which Bayesian posteriors converge to a normal distribution centred at the MLE with covariance $\mathcal{I}(\hat{\theta})^{-1}$. An analogue holds for the Bayesian bootstrap applied to smooth functionals: the posterior of $G(F)$ under the Dirichlet process prior (6) converges to the same normal distribution as the bootstrap sampling distribution (4). This provides a consistency guarantee for the GBB: the martingale posterior generated by the Pólya urn converges to the correct limiting distribution as $n \to \infty$.

8.4 Practical recommendations

The coverage study and theoretical development suggest the following guidelines for practitioners. For smooth statistics (means, regression coefficients) with moderate to large n, the WBB with $\alpha = 1$ and no tilting is recommended: it achieves nominal coverage, requires only one optimisation per draw, and automatically explores the regularisation path. For nonsmooth statistics (medians, quantiles), the Efron or Rubin Bayesian bootstrap is preferable, as these do not require a differentiable objective. The Jeffreys variant ($\alpha = 1/2$)

should be used when conservative coverage is desired, at the cost of wider intervals. The SIR-based weighted likelihood bootstrap should be avoided for large n, as the importance weights degenerate; if a parametric model is available, the WBB with the parametric loss provides the same information without the SIR step. The generative Bayesian bootstrap with $m = n$ undercovers at approximately 83% (Corollary 1); practitioners should apply the width calibration of Proposition 2, which restores baseline BB coverage at essentially zero computational cost, or equivalently increase m substantially ($m \geq 9n$ brings predicted coverage above 93%).

8.5 Summary and concluding remarks

We have traced the bootstrap's evolution from Efron's 1979 resampling algorithm through Rubin's Bayesian bootstrap, Newton and Raftery's weighted likelihood bootstrap, the weighted Bayesian bootstrap of Newton et al. [2021], and the generative Bayesian bootstrap of Fong, Holmes, and Walker. One structure recurs: all five methods can be viewed as computing a functional of a random weighted measure on the observed data, as in (1). They differ in the weight distribution and, more broadly, in the inference mechanism: the WLB adds importance resampling, the WBB replaces functional evaluation with optimisation, and the GBB augments the data via posterior-predictive simulation. Table 3 summarises these distributions.

Table 3: Weight distributions for the five bootstrap methods.

Method	Weight distribution	Character
Efron (1979)	Multinomial$(n; 1/n, \ldots, 1/n)$ / n	Discrete
Rubin (1981)	Dirichlet$(1, \ldots, 1)$	Continuous
WLB, SIR variant (1994)	SIR from bootstrap MLEs, eq. (9)	Parametric
Newton–Polson–Xu (2021)	$w_i \propto \xi_i \, e^{\eta \, g(x_i)}$, $\xi_i \sim \text{Ga}(\alpha, 1)$	Generalised
Generative BB	Pólya urn counts $/(n+m)$	Predictive

This unifying representation reveals that frequentist versus Bayesian, nonparametric versus parametric, and resampling versus generation can be viewed as instances of one mathematical idea.

Table 4 summarises the conditions under which the five methods coincide or diverge asymptotically, fulfilling the second goal stated in the introduction.

Table 4: Asymptotic equivalence conditions. The nonparametric methods (Efron, BB, WBB, GBB) agree to first order for smooth G as $n \to \infty$ under Hadamard differentiability; the parametric WLB targets a different (narrower) limit. Methods diverge in the regimes listed.

Pair	Coincide when	Diverge when
Efron / Rubin BB	Smooth G, fixed p	Nonsmooth G (e.g., sample max); small n (BB has continuous weights)
BB / WLB	Same centering under correct model, smooth G	Different variance (WLB narrower); model misspecification
BB / WBB (α, η)	$\alpha = 1, \eta = 0$	$\alpha \neq 1$ (changes weight dispersion); $\eta \neq 0$ (tilting shifts mass)
BB / GBB	$m \to \infty$ (GBB converges to BB)	Finite m (GBB has reduced variance from augmented sample)
All methods / high-p	Fixed p	$p/n \to \gamma > 0$ (standard bootstrap inconsistent; debiasing needed)

Despite its remarkable theoretical development, the bootstrap faces important open challenges. High-dimensional theory (where p grows proportionally with n) remains incomplete; the conditions for frequentist coverage of the Bayesian bootstrap in high dimensions, and the optimal choice of α, are active research areas. Extending the GBB to non-exchangeable data (longitudinal, spatial, and network data) requires new predictive rules beyond the Pólya urn. Some work uses spatial Pólya trees and graph-based exchangeability, but a theory analogous to de Finetti's theorem does not yet exist for these settings.

Our scope is the weight-distribution thread from Efron to the generative Bayesian bootstrap. Methods for dependent data, such as the block bootstrap [Künsch, 1989] and wild bootstrap [Wu, 1986], and for computational scaling [Kleiner et al., 2014], address orthogonal concerns and fall outside this thread.

As developed in Section 2, the weighted Bayesian bootstrap is already a generator (14) in the modern sense: a deterministic transport map from a simple base distribution to the posterior, with amortised hyperparameter tuning as a by-product. The deeper connection between the GBB and learned generative models (normalising flows, diffusion models, variational autoencoders) remains incompletely understood. These models can be interpreted as learning to generate from the posterior predictive distribution under implicit priors, and the GBB provides a nonparametric Bayesian benchmark for evaluating their statistical properties. Developing a rigorous theoretical framework for this connection is an important direction for future research. The computational scaling of the GBB to very large datasets is another open problem. Approximate algorithms based on mini-batch resampling, stratified Pólya urns, or neural network approximations to the predictive rule are needed for large-scale applications.

Bradley Efron's 1979 bootstrap paper promised something that seemed too good to be true: a general-purpose method for uncertainty quantification that required no mathematical analysis beyond the ability to re-run the estimator on new datasets. The deeper significance of the bootstrap, as Rubin, Newton, and their successors have shown, is not computational convenience but theoretical depth. The bootstrap is a window onto the nonparametric Bayesian framework, connecting empirical evidence to probabilistic beliefs through the machinery of the Dirichlet process, exchangeability, and predictive distributions. In its most modern incarnation (the generative Bayesian bootstrap), it is a foundation for a new style of statistical computation that places predictive simulation at the centre of inference. We expect this evolution to continue, as the connections between the bootstrap and modern machine learning, sequential Monte Carlo, and generative computation continue to be explored and formalised. Key open problems include extending the generator framework to high-dimensional and non-exchangeable settings, establishing finite-sample coverage guarantees for the WBB, and developing scalable predictive resampling algorithms for large datasets.

Acknowledgements

The authors thank the Editor, Associate Editor, and anonymous referees for their constructive feedback.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

No empirical data were used in this study. All numerical results are produced from simulated data under the $\text{LogNormal}(0.5, 0.6^2)$ generating process described in Section 7. Code to reproduce every table and figure (including the Monte Carlo coverage study and the t_ν tutorial) is provided as supplementary material and will be released in a public repository upon acceptance.

Artificial Intelligence Disclosure

The authors used ChatGPT as a prose checking and coding assistant during manuscript preparation and revision. Specifically, the tool was used to (a) copy-edit prose under author direction; (b) debug the Python simulation code (`coverage_sim.py`, `aggregate_coverage.py`); and (c) assist with internal-consistency checks of Proposition 1, Corollaries 1–2, and Proposition 2. All mathematical results, methodological decisions, and interpretations were developed, verified, and are the responsibility of the authors; every AI-generated passage was reviewed and edited before inclusion. The tool was not used to generate data, fabricate results, or conduct literature review.

References

Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6):1196–1217, 1981.

Peter J. Bickel, Friedrich Götze, and Willem R. van Zwet. Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31, 1997.

David Blackwell and James B. MacQueen. Ferguson distributions via Pólya urn schemes. *Annals of Statistics*, 1(2):353–355, 1973.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Gary Chamberlain and Guido W. Imbens. Nonparametric applications of Bayesian inference. *Journal of Business & Economic Statistics*, 21(1):12–18, 2003.

Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1):1–68, 1937.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, 2006.

Thomas J. DiCiccio and Joseph P. Romano. The automatic percentile method: Accurate confidence limits in parametric models. *Canadian Journal of Statistics*, 17(2):155–169, 1989.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1): 1–26, 1979.

Bradley Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987.

Bradley Efron. Bayes and likelihood calculations from confidence intervals. *Biometrika*, 80 (1):3–26, 1993.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

Noureddine El Karoui and Elizabeth Purdom. Can we trust the bootstrap in high-dimensions? The case of linear models. *Journal of Machine Learning Research*, 19(1): 170–235, 2018.

Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1(2):209–230, 1973.

Edwin Fong, Chris C. Holmes, and Stephen G. Walker. Scalable nonparametric sampling from multimodal posteriors with the posterior bootstrap. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1952–1962, 2019.

Edwin Fong, Chris C. Holmes, and Stephen G. Walker. Martingale posterior distributions. *Journal of the Royal Statistical Society: Series B*, 85(2):417–472, 2023.

Sandra Fortini and Sonia Petrone. Predictive construction of priors in Bayesian nonparametrics. *Brazilian Journal of Probability and Statistics*, 26(4):423–449, 2012.

Sandra Fortini and Sonia Petrone. Quasi-Bayes properties of a procedure for sequential learning in mixture models. *Journal of the Royal Statistical Society: Series B*, 82(4):1087–1114, 2020.

David T. Frazier, Gael M. Martin, and Christian P. Robert. Approximate Bayesian computation: A regression perspective. *Biometrika*, 107(2):289–309, 2020.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, 140(2):107–113, 1993.

P. Richard Hahn, Jared S. Murray, and Carlos M. Carvalho. Bayesian regression tree models for causal inference. *Bayesian Analysis*, 15(3):965–1056, 2020.

Peter Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

Zhezhen Jin, Zhiliang Ying, and Lee J. Wei. A simple resampling method by perturbing the minimand. *Biometrika*, 88(2):381–390, 2001.

Ariel Kleiner, Ameet Talwalkar, Purnamrita Sarkar, and Michael I. Jordan. A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B*, 76(4):795–816, 2014.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3):1217–1241, 1989.

Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.

Albert Y. Lo. A large sample study of the Bayesian bootstrap. *Annals of Statistics*, 15(1): 360–375, 1987.

Simon P. Lyddon, Chris C. Holmes, and Stephen G. Walker. General Bayesian updating and the loss-likelihood bootstrap. *Biometrika*, 106(2):465–478, 2019.

Michael A. Newton and Adrian E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B*, 56(1):3–26, 1994.

Michael A. Newton, Nicholas G. Polson, and Jianeng Xu. Weighted Bayesian bootstrap for scalable posterior distributions. *Canadian Journal of Statistics*, 49(2):421–437, 2021.

Art B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

Art B. Owen. Empirical likelihood ratio confidence regions. *Annals of Statistics*, 18(1): 90–120, 1990.

Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer, New York, 1999.

Nicholas G. Polson and Vadim Sokolov. Deep learning: A Bayesian perspective. *Bayesian Analysis*, 12(4):1275–1304, 2017.

Nicholas G. Polson and Vadim Sokolov. Generative Bayesian computation. *Bayesian Analysis*, 2025. In press.

Nicholas G. Polson, James G. Scott, and Brandon T. Willard. Proximal algorithms in statistics and machine learning. *Statistical Science*, 30(4):559–581, 2015.

Jens Præstgaard and Jon A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Annals of Probability*, 21(4):2053–2086, 1993.

Danilo J. Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1530–1538, 2015.

Donald B. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9(1):130–134, 1981.

B. W. Silverman and G. A. Young. The bootstrap: To smooth or not to smooth? *Biometrika*, 74(3):469–479, 1987.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

Yuexi Wang and Veronika Rockova. Generative Bayesian inference with GANs. *Journal of Machine Learning Research*, 27(29):1–48, 2026.

Samuel S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1):60–62, 1938.

C. F. Jeff Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295, 1986.

Keming Yu and Rana A. Moyeed. Bayesian quantile regression. *Statistics and Probability Letters*, 54(4):437–447, 2001.